Die Steinsche Methode

Peter Eichelsbacher

Fakultät für Mathematik
Ruhr-Universität Bochum

meiner Frau

Helene

gewidmet

Vorwort

In der ersten Hälfte des 20. Jahrhunderts beschäftigte man sich in der Wahrscheinlichkeitstheorie vornehmlich mit klassischen Grenzwertsätzen, wie den *starken Gesetzen der großen Zahlen*, den *zentralen Grenzwertsätzen* und den *Gesetzen vom iterierten Logarithmus*, und dies für Summen unabhängiger Zufallsgrößen. Der zentrale Grenzwertsatz fand seither viele Anwendungen in der Statistik, insbesondere beim Studium der Verteilungen von Teststatistiken. Methodisch ist die Theorie der *Fourier-Transformierten* das verwendete Handwerkszeug, welches gewöhnlich nicht dazu dient, auch die Konvergenzgüte des Grenzwertsatzes abschätzen zu können. 1940 haben Berry und Esséen mittels einer sogenannten Glättungsungleichung eine erste Rate der Konvergenz in Form einer exakten Schranke geliefert. 1966 hat Le Cam ein vergleichbares Resultat für das Poissonsche Gesetz der kleinen Zahlen geliefert. Die Resultate beziehen sich auf den Fall unabhängiger Zufallsgrößen. In Anwendungen liegt jedoch meist Abhängigkeit vor, wo die Methode der Fourier-Transformation nur schwerlich zu Konvergenzgüte-Abschätzungen führen kann.

1972 hat Charles Stein für den Fall der Normal-Approximation eine neue Methode entwickelt, die heute *Steinsche Methode* genannt wird. Sie basiert auf einem Differentialoperator und einer Wahl eines austauschbaren Paares von Zufallsgrößen; die Kombination dieser Techniken ist noch heute schwer verständlich, sie führt aber zu expliziten Abschätzungen des Approximationsfehlers auch im Fall abhängiger Zufallsgrößen.

Stein wendete seine Methode zunächst für den zentralen Grenzwertsatz (ZGW) für Summen stationärer, mischender Zufallsgrößen an. Poisson-Approximation wurde 1975 von Louis Chen studiert. 1988 wurde die Methode für stochastische Prozesse entwickelt, der multivariate ZGW wurde 1991 mittels der Steinschen Methode untersucht. Stein's Idee, mittels Konzentrationsungleichungen seine Methode fortzuentwickeln, wurde 2001 verwirklicht. Dies sind nur einige Stationen der Entwicklung der Steinschen Methode in den zurückliegenden 30 Jahren.

Der Grund für die Stärke der Steinschen Methode bleibt bis heute mysteriös. Viele Probleme sind offen. Die Kontrolle des Verhaltens der Lösungen der Steinschen Gleichungen kann sehr schwer sein. Multivariate Verteilungen wurden bisher kaum untersucht. Punktprozess-Approximationen und ihre Güte sind noch recht schlecht verstanden. Trotzdem fand die Steinsche Methode bereits viele Anwendungen, wie etwa in den Bereichen räumlicher Statistiken, der Theorie zufälliger Graphen, der „computational biology“, bei interagierenden Partikelsystemen, beim Bootstrap-Verfahren, bei der mathematischen Untersuchung von Epidemien, sowie in der Versicherungs- und Finanzmathematik.

Viele Varianten des eigentlichen Zugangs von Stein sind entstanden. Methodisch ist die der Kopplung hervorzuheben.

Das vorliegende Skript entstand zu einer Vorlesung, die ich in Bielefeld und in Bochum zu diesem Thema gehalten habe. Es behandelt vorwiegend den Fall der Normal- und Poisson-Approximation. Unter den vielen Beispielen treten *Zufallsgraphen* kontinuierlich auf. Aktuelle Entwicklungen werden in *Ausblicken* am Ende der Kapitel beschrieben.

Die Steinsche Methode begegnete mir bei Friedrich Götze in Bielefeld, vertieft später in Zürich bei Andrew Barbour und Erwin Bolthausen. In vielen Diskussionen mit ihnen und mit Gesine Reinert und Goscha Roos begann ich zu verstehen, was in diesem Teilgebiet der Stochastik fasziniert und staune immer wieder, wie mysteriös sie, die Steinsche Methode, doch bleiben kann.

Mein Dank gilt den geduldigen Hörerinnen und Hörern meiner Vorlesungen, die Ungenauigkeiten und Schreibfehler entdeckten. Mein Dank gilt auch Hanna Döring, die das Skript sehr engagiert überarbeitet hat.

Bochum, im Oktober 2003

Peter Eichelsbacher

Inhaltsverzeichnis

1. Die Poisson-Verteilung

In diesem Kapitel stellen wir einen Satz von Le Cam aus dem Jahr 1960 vor. Der Satz gibt eine Abschätzung des Totalvariations-Abstands zwischen der Binomial-Verteilung und einer Poisson-Verteilung an und behandelt somit die Konvergenzgüte des klassischen Grenzwertsatzes von Poisson. Der Beweis stellt eindrücklich die sogenannte Kopplungsmethode vor. Le Cams Resultat, sprich die Schranke der Abschätzung, zu verbessern ist ein natürlicher Sport in der Mathematik. Eine eindrucksvolle Verbesserung gelingt mit Hilfe der Steinschen Methode, die in Kapitel 2 für den Fall der Poisson-Approximation entwickelt wird.

Zunächst erinnern wir an einfache Begriffe der Wahrscheinlichkeitstheorie und Eigenschaften der Poissonverteilung. Wir führen den Totalvariations-Abstand ein und formulieren den Satz von Le Cam. Als Vorbereitung für den Beweis des Satzes stellen wir ein Kopplungslemma bereit. Wir schliessen mit einer Diskussion des Results von Le Cam.

Vorerst kommen wir mit den Grundgrößen der diskreten Wahrscheinlichkeitstheorie zurecht, als da sind:

Definition 1.1. (i) Ein *diskreter Wahrscheinlichkeitsraum* (Ω, p) besteht aus einer endlichen oder abzählbar unendlichen Menge Ω und einer Abbildung $p : \Omega \to [0,1]$, für die $\sum_{\omega \in \Omega} p(\omega) = 1$ gilt.

(ii) *Ereignisse* sind Teilmengen von Ω. Für $A \subset \Omega$ ist $P(A) := \sum_{\omega \in A} p(\omega)$.

(iii) $X : \Omega \to \mathbb{R}$ heißt (diskrete) *Zufallsvariable / Zufallsgröße*. Es sei $f(z) := P(X = z)$. Dann ist $(X(\Omega), f)$ ein diskreter Wahrscheinlichkeitsraum. f heißt die *Verteilung (Wahrscheinlichkeitsverteilung)* der Zufallsgröße X.

(iv) Es sei X eine Zufallsgröße. Man sagt, dass der *Erwartungswert von X existiert*, falls $\sum_{z \in X(\Omega)} |z| \, P(X = z) < \infty$ ist. Der *Erwartungswert* von X ist dann definiert durch

$$\mathbb{E}(X) = \sum_{z \in X(\Omega)} z \, P(X = z).$$

Wir schreiben oft $\mathbb{E}X$ statt $\mathbb{E}(X)$.

(v) Es sei X eine Zufallsgröße mit existierendem Erwartungswert $\mathbb{E}X$. Dann heißt

$$V(X) := \sum_{z \in X(\Omega)} (z - \mathbb{E}X)^2 \, P(X = z)$$

die *Varianz* von X und $S(X) := +\sqrt{V(X)}$ die *Standardabweichung* von X, falls die auftretende Reihe konvergiert.

(vi) Sind X und Y zwei Zufallsgrößen, so wird die *Kovarianz* zwischen X und Y definiert durch $\operatorname{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$, falls alle in diesem Ausdruck vorkommenden Erwartungswerte existieren.

(vii) n diskrete Zufallsgrößen $X_1, \ldots, X_n$ heißen *stochastisch unabhängig* , wenn

$$P(X_1 = z_1, \ldots, X_n = z_n) = \prod_{i=1}^{n} P(X_i = z_i)$$

für alle $z_i \in X_i(\Omega)$, $i \in \{1, \ldots, n\}$, gilt.

Bemerkung 1.2. Beispiele und Rechenregeln zu all diesen Größen entnehme man einer einführenden Vorlesung zu den Stochastischen Methoden (siehe zum Beispiel das Buch von U. Krengel [Kre88]).

Es sei daran erinnert, dass eine Zufallsgröße X mit der Verteilung

$$P(X = k) = b(k; n, p) = \binom{n}{k} p^k \, (1-p)^{n-k}$$

mit $k = 0, 1, \ldots, n$ *binomialverteilt* mit Parametern p und n heißt.

Die Bedeutung dieser Verteilung liegt in Folgendem: man betrachte eine *Bernoulli-Kette* von Experimenten der Länge n mit Erfolgswahrscheinlichkeit p (n-maliges Durchführen ein und des gleichen Experiments, und zwar stochastisch unabhängig; Produktraum). Dann ist die Anzahl der Erfolge binomialverteilt.

Bemerkung 1.3. Bei aller Bedeutung dieser experimentellen Situation sei schon jetzt bemerkt, dass gleiche Bedingungen für die Einzelexperimente und die stochastische Unabhängigkeit der Einzelexperimente in der Praxis kaum realisierbar sind. Wir wollen daher in dieser Vorlesung *nicht-identisch verteilte* und (oder) *abhängige* Experiment-Ketten in den Vordergrund stellen. Trotzdem erinnern wir zunächst an den i.i.d. (*independent and identically distributed*) Fall.

Eine Zufallsgröße X mit $X(\Omega) = \mathbb{N}_0$ und Verteilung $p(k|\lambda) := e^{-\lambda}\frac{\lambda^k}{k!}$ heißt *Poisson-verteilt* mit Parameter $\lambda > 0$. Es handelt sich um eine Verteilung, denn

$$\sum_{k \geq 0} p(k|\lambda) = e^{-\lambda} \sum_{k \geq 0} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Der Erwartungswert dieser Verteilung ist leicht auszurechnen:

$$\sum_{k=0}^{\infty} k\, p(k|\lambda) = e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda} \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \lambda e^{+\lambda} = \lambda.$$

Eine Poisson-verteilte Zufallsgröße hat also Erwartungswert λ. Als nächstes wollen wir die Varianz ausrechnen:

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{k=0}^{\infty} k^2 p(k|\lambda) = e^{-\lambda}\sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} \\ &= e^{-\lambda}\sum_{k=1}^{\infty}(k(k-1)+k)\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{k+2}}{k!} + \lambda = \lambda^2+\lambda.\end{aligned}$$

Somit gilt

$$V(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \lambda^2+\lambda-\lambda^2 = \lambda.$$

Erwartungswert und Varianz einer Poisson-verteilten Zufallsgröße sind gleich dem Parameter λ.

Einer der ältesten Grenzwertsätze in der Wahrscheinlichkeitstheorie ist Poissons *Gesetz der kleinen Zahlen* , wonach $b(k;n,p)$ gegen $p(k|\lambda)$ für $n \to \infty$ konvergiert, wenn $p = p(n) = \frac{\lambda}{n}$ für $\lambda > 0$ vorausgesetzt wird (genauer: wenn $\lim_{n\to\infty} n\,p(n) = \lambda > 0$). Dies bewies Poisson im Jahre 1837. Hier ist $p = P($das k-te Teilexperiment liefert Erfolg$)$. Diese Wahrscheinlichkeit muß also klein werden in n. Dies ist bei der Beobachtung einer *festen* Folge zufälliger Einzelexperimente (Zufallsgrößen) $X_1, X_2, \dots$ mathematisch aber schon nicht möglich. Zur Präzisierung des klassischen Grenzwertsatzes benötigt man eine Folge wachsender Serien von Zufallsgrößen:

$$\begin{array}{l} X_1^{(1)}; \\ X_1^{(2)}, X_2^{(2)}; \\ X_1^{(3)}, X_2^{(3)}, X_3^{(3)}; \\ \quad\dots \\ X_1^{(n)}, \dots, X_n^{(n)}; \end{array}$$

(man nennt dies ein Dreiecks-Schema), wobei $P(X_i^{(n)} = 1) = p_n$ und $P(X_i^{(n)} = 0)$ $= 1 - p_n$, $i \in \{1, \dots, n\}$, sei. Wir betrachten dies nicht weiter (vergleiche trotzdem Kapitel 5, §4, in Borovkov [Bor76]). In einem konkreten Beispiel liegen aber p (konstant) und n vor. Naheliegend ist nun der Ansatz $\lambda := n\,p$. Man möchte in dieser Situation Poisson-Approximation studieren, also der Frage nachgehen, wie gut die Poisson-Wahrscheinlichkeiten $p(\cdot|\lambda)$ die vorliegenden Wahrscheinlichkeiten annähern. Bevor wir ein Resultat von Le Cam aus dem Jahr 1960 formulieren, führen wir einen speziellen Abstand auf der Menge der Wahrscheinlichkeitsverteilungen auf $\mathbb{N}_0$ ein:

Gegeben seien zwei Wahrscheinlichkeiten (dies sagen wir für Wahrscheinlichkeitsverteilungen) P und Q auf $\mathbb{N}_0$.

Definition 1.4. Die Größe

$$d_{TV}(P,Q) := \sup\{|P(A) - Q(A)| : A \subseteq \mathbb{N}_0\}$$

heißt *Totalvariations-Abstand* oder kurz *Totalvariation* von P und Q.

Der Inhalt der folgenden Bemerkungen ist jeweils eine einfache Übung.

Bemerkungen 1.5.

(i) Es gilt $d_{TV}(P,Q) = \frac{1}{2}\sum_{k\geq 0}|P(k)-Q(k)|$, wobei $P(k) := P(\{k\})$.

(ii) $d_{TV}(\cdot,\cdot)$ ist eine vollständige Metrik auf dem Raum aller Wahrscheinlichkeiten auf $\mathbb{N}_0$.

(iii) Die Definition ist einfach übertragbar, wenn $\mathbb{N}_0$ ersetzt wird durch einen beliebigen meßbaren Raum $(\Omega, \mathcal{A})$.

(iv) $0 \leq d_{TV}(\cdot,\cdot) \leq 1$.

(v) Eine Folge reellwertiger Zufallsgrößen $(Y_n)_{n\in\mathbb{N}}$ konvergiert gegen eine Zufallsgröße Y *in Verteilung*, wenn

$$\lim_{n\to\infty} P(Y_n \leq x) = P(Y \leq x)$$

gilt für alle Stetigkeitspunkte x von $F(x) := P(Y \leq x)$. In Zeichen schreibt man dafür: $Y_n \xrightarrow{d} Y$. Sind die Zufallsgrößen $\mathbb{N}_0$-wertig, ist dies gleichbedeutend mit

$$\lim_{n\to\infty} P(Y_n = k) = P(Y = k)$$

für alle $k \in \mathbb{N}_0$. Sei nun $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsgrößen mit Werten in $\mathbb{N}_0$. Diese konvergiert in Verteilung gegen eine Wahrscheinlichkeit P genau dann wenn

$$\lim_{n\to\infty} d_{TV}(\mathcal{L}(X_n), P) = 0.$$

Hier bezeichnet $\mathcal{L}(X_n)$ die Verteilung von X_n. Ist Ω nicht diskret (zum Beispiel $\mathbb{R}$), so ist die Konvergenz in Totalvariation stärker als die Konvergenz in Verteilung (und manchmal einfach zu stark, um nützlich zu sein).

Es gilt nun (siehe [LC60]):

Satz 1.6. *(Le Cam, 1960) Es seien $I_1,\ldots,I_n$ unabhängige Zufallsgrößen, definiert auf einem gemeinsamen Wahrscheinlichkeitsraum, mit $P(I_i = 1) = p_i$ und $P(I_i = 0) = 1 - p_i$ mit $0 < p_i < 1$ für alle $i = 1,\ldots,n$. Sei $W := I_1 + \cdots + I_n$ und $\lambda := \mathbb{E}W = p_1 + \cdots + p_n$. Dann gilt:*

$$d_{TV}(\mathcal{L}(W), \mathrm{Po}(\lambda)) \leq \sum_{i=1}^{n} p_i^2,$$

wobei $\mathrm{Po}(\lambda)$ *die Poisson-Verteilung mit Parameter* λ *bezeichne.*

Es folgt dann unmittelbar:

Satz 1.7.

$$d_{TV}(B(n,p), \mathrm{Po}(n\,p)) \leq n\,p^2,$$

wobei $B(n,p)$ die Binomial-Verteilung zu den Parametern n und p bezeichne.

Bemerkung 1.8. (i) Für $B(n,p)$ ist der Erwartungswert $n\,p$, daher der Ansatz $\mathrm{Po}(n\,p)$ gewählt.

(ii) Es folgt der klassische Grenzwertsatz von Poisson: Aus $\lim_{n\to\infty}(n\,p(n)) = \lambda$ folgt $\lim_{n\to\infty} p(n) = 0$ und somit $\lim_{n\to\infty} n\,p(n)^2 = 0$. Da

$$|b(k; n, p(n)) - p(k|n\,p(n))| \leq 2\, d_{TV}\big(B(n,p(n)), \mathrm{Po}(n\,p(n))\big),$$

folgt

$$\lim_{n\to\infty} |b(k; n, p(n)) - p(k|n\,p(n))| = 0,$$

und mit $\lim_{n\to\infty} p(k|n\,p(n)) = p(k|\lambda)$ folgt dann der Grenzwertsatz von Poisson.

(iii) Es darf also auch $n\,p(n) \to \infty$ gelten, wenn nur $n\,p^2(n) \to 0$ erfüllt ist (zum Beispiel kommt somit auch $p(n) = 1/n^{2/3}$ in Frage).

Wir geben den Beweis von Satz 1.6. Hierbei werden wir eine *Kopplungsmethode* anwenden. Außerdem können wir erkennen, dass noch Platz in der von Le Cam gefundenen Abschätzung ist. Wir benötigen die folgenden Vorbereitungen:

Proposition 1.9. *X und Y seien unabhängig und Poisson-verteilt mit Parametern λ beziehungsweise $\mu > 0$. Dann ist $X + Y$ Poisson-verteilt mit Parameter $\lambda + \mu$.*

Beweis: Für $n \in \mathbb{N}_0$ gilt:

$$\begin{aligned}
P(X+Y=n) &= \sum_{k=0}^{n} P(X=k, Y=n-k) \\
&= \sum_{k=0}^{n} P(X=k)P(Y=n-k) \quad \text{(Unabhängigkeit)} \\
&= \sum_{k=0}^{n} \frac{\lambda^k}{k!}\frac{\mu^{n-k}}{(n-k)!} e^{-\lambda} e^{-\mu} = \frac{1}{n!}\left(\sum_{k=0}^{n} \binom{n}{k} \lambda^k \mu^{n-k}\right) e^{-(\lambda+\mu)} \\
&= \frac{1}{n!}(\lambda+\mu)^n e^{-(\lambda+\mu)} = p(n|\lambda+\mu).
\end{aligned}$$

□

Per Induktion folgt sofort, dass die Summe von endlich vielen unabhängigen Poisson-verteilten Zufallsgrößen wieder Poisson-verteilt ist, wobei der Parameter sich als Summe der Einzelparameter ergibt.

Wir verwenden eine Technik, die man *Kopplung* nennt. Nehmen wir an, f und g seien zwei Wahrscheinlichkeitsverteilungen auf $\mathbb{N}_0$: $f, g : \mathbb{N}_0 \to [0,1]$, $\sum_k f(k) = \sum_k g(k) = 1$. Wir wollen zeigen, dass $\sum_{k=0}^{\infty} |f(k) - g(k)|$ klein ist. Wir werden das tun, indem wir Zufallsgrößen X, Y auf einem gemeinsamen Wahrscheinlichkeitsraum Ω konstruieren, die die Verteilung f beziehungsweise g haben, und die ,,möglichst weitgehend" übereinstimmen. Es soll also die folgende Situation vorliegen:

$$f(k) = P(X = k), \quad g(k) = P(Y = k).$$

A sei ein Ereignis mit $X(\omega) = Y(\omega)$ für $\omega \in A$, das heißt $A \subset \{\omega : X(\omega) = Y(\omega)\}$. Man sagt, X und Y seien auf A ,,gekoppelt".

Lemma 1.10. *Unter den obigen Bedingungen gilt*

$$\frac{1}{2} \sum_{k=0}^{\infty} |f(k) - g(k)| \leq P(A^c).$$

Beweis: Sei $M := \{k \in \mathbb{N}_0 : f(k) > g(k)\}$. Dann ist

$$\begin{aligned}
\sum_{k=0}^{\infty} |f(k) - g(k)| &= \sum_{k \in M} (f(k) - g(k)) - \sum_{k \notin M} (f(k) - g(k)) \\
&= 2 \sum_{k \in M} (f(k) - g(k)) - \sum_{k=0}^{\infty} (f(k) - g(k)) \\
&= 2(P(X \in M) - P(Y \in M)) - (1 - 1) \\
&= 2(P(X \in M, A) + P(X \in M, A^c) - P(Y \in M)) \\
&\leq 2(P(Y \in M, A) + P(A^c) - P(Y \in M)) \\
&\leq 2P(A^c).
\end{aligned}$$

□

Wir wenden nun dieses Kopplungsargument an, um Satz 1.6 zu beweisen. Der Hauptteil des Beweises besteht in einer geeigneten Wahl des zugrundeliegenden Wahrscheinlichkeitsraumes. Da wir nur die Verteilung von W berechnen müssen, ist es quasi egal, auf welchem Wahrscheinlichkeitsraum die Zufallsgrößen I_i definiert werden. Es ist für uns nur wichtig, dass die Zufallsgrößen unabhängig sind und $P(I_i = 1) = p_i$ sowie $P(I_i = 0) = 1 - p_i$ gilt. Diese Freiheit nutzen wir für eine Wahl derart, dass eine Poisson-verteilte Zufallsgröße zum Parameter λ möglichst weitgehend mit W in Verteilung übereinstimmt. Dazu sei $\Omega_i = \{-1, 0, 1, 2, \ldots\}$, $P_i(0) = 1 - p_i$ und $P_i(k) = \frac{e^{-p_i}}{k!} p_i^k$ für $k \geq 1$ sowie $P_i(-1) = 1 - P_i(0) - \sum_{k \geq 1} P_i(k) = e^{-p_i} - (1 - p_i)$. Nach Konstruktion sind somit (Ω_i, P_i) W.-Räume. Betrachte dann den Produktraum (Ω, P) der (Ω_i, P_i). (Erinnerung: $\Omega = \Omega_1 \times \cdots \times \Omega_n$. Für jedes $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$ definieren wir $P(\omega) = P_1(\omega_1) P_2(\omega_2) \cdots P_n(\omega_n)$. Offensichtlich gilt $\sum_{\omega \in \Omega} P(\omega) = 1$.) Wir setzen für $\omega \in \Omega$

$$I_i(\omega) := \begin{cases} 0, & \text{falls } \omega_i = 0, \\ 1, & \text{sonst }, \end{cases}$$

und

$$Y_i(\omega) := \begin{cases} k, & \text{falls } \omega_i = k, k \geq 1, \\ 0, & \text{sonst }. \end{cases}$$

Dann haben nach Definition die Zufallsgrößen I_i die geforderte Verteilung: $P(I_i = 1) = p_i$ und $P(I_i = 0) = 1 - p_i$. Sie sind weiter nach Definition des Produktraumes unabhängig. Die Y_i sind nach Definition Poisson-verteilt zum Parameter p_i und ebenfalls unabhängig. Also folgt mit Proposition 1.9, dass $Y = Y_1 + \cdots + Y_n$ Poisson-verteilt ist zum Parameter λ. Nun stimmen die Zufallsgrößen in den Werten 0 und 1 überein, und es ist $P(I_i = Y_i) = P_i(0) + P_i(1) = (1 - p_i) + e^{-p_i} p_i$, und somit

$$P(I_i \neq Y_i) = p_i(1 - e^{-p_i}) \leq p_i^2,$$

denn für $x > 0$ gilt $1 - e^{-x} \leq x$. Nach Lemma 1.10 folgt dann

$$\frac{1}{2} \sum_{k=0}^{\infty} |P(W = k) - p(k|\lambda)| \leq P(W \neq Y) \leq \sum_{i=1}^{n} P(I_i \neq Y_i) \leq \sum_{i=1}^{n} p_i^2.$$

Damit ist Satz 1.6 bewiesen.

Bemerkungen 1.11. (i) Wie gut ist die Abschätzung $\sum_{i=1}^n p_i^2$? Betrachte den Fall $p = p_1 = \cdots = p_n$, also Satz 1.7, so gilt für $k = 0$ via Potenzreihenentwicklung:

$$P(W = 0) - p(0|n\,p) = e^{n \log(1-p)} - e^{-n\,p} \sim e^{-n\,p} \left(\frac{(n\,p)^2}{2n} \right) = \frac{e^{-n\,p}}{2}\, n\,p^2.$$

Ist also $\frac{e^{-n\,p}}{2}$ nicht zu klein, ist für $k = 0$ die gegebene Abschätzung kaum noch zu verbessern. Für $n\,p$ groß ist $e^{-n\,p}$ allerdings von Bedeutung. In solchen Fällen scheint also noch Luft in den Abschätzungen zu sein. Das abstrakte Verfahren läßt uns natürlich beim Kopplungslemma für jedes $k \in \mathbb{N}_0$ noch Luft: Hier werden die Wahrscheinlichkeiten seltener Ereignisse $P(X \in M)$ und $P(Y \in M)$ rausgeworfen. Dies ist tatsächlich in vielen Fällen eine grobe Abschätzung.

(ii) Wir rechnen ein Beispiel: Gegeben seien 10^4 Säcke mit Korn. Es sei bekannt, dass sich in diesen Säcken 5000 markierte Körner befinden. Wie groß ist die Wahrscheinlichkeit, dass in irgendeinem fixierten Sack wenigstens ein markiertes Korn ist? Es ist $p = \frac{1}{10^4}$ und $n = 5000$ (Körner zufällig auf Säcke verteilen). Also ist $n\,p = 0.5$. Setze $I_k = 1$, wenn das k-te Korn in dem ausgewiesenen Sack ist. Dann ist

$$P\left(\sum_{k=1}^{5000} I_k = 0\right) \approx e^{-n\,p} = e^{-0.5},$$

also ist die gesuchte Wahrscheinlichkeit $\approx 1 - e^{-0,5}$. Der relative Fehler ist kleiner oder gleich $\frac{1}{4 \cdot 10^4}$, also erstaunlich klein.

Das Resultat von Le Cam ist eine konkrete Schranke für den Aproximationsfehler. Naürlich fragt man sich, wie gut diese Schranke eigentlich ist. Kann man ganz allgemein diese Schranke noch verbessern? Wir wollen in den folgenden Kapiteln in Bezug auf die Poisson-Approximation vorstellen, dass man

(i) die Schranken verbessern kann,
(ii) weiterhin nicht-identisch verteilte I_i untersuchen kann,
(iii) auch für abhängige Zufallsgrößen I_i Schranken finden kann,
(iv) die Methode der Konstruktion von Kopplungen weiterhin verwenden kann.

Dazu werden wir in die *Steinsche Methode* einführen. Wir werden viele Beispiele abhängiger Zufallsgrößen vorstellen.

In Bezug auf eine mögliche Poisson-Approximation fragen wir: wie stark ist das *Gesetz der kleinen Zahlen* von Poisson von der Annahme der identischen Verteilung der I_i und der Unabhängigkeit der I_i abhängig?

2. Die Steinsche Methode für Poisson-Approximation

Charles Stein [Ste72] hat 1972 eine Methode entwickelt, mit der er den Fehler in der Approximation einer geeignet normierten Summe von abhängigen Zufallsvariablen durch die *Normalverteilung* abschätzen konnte. Die Methode ist enorm effektiv und ist heute nach ihm benannt. In den Folgejahren zeigte sich, dass die Methode nicht auf den Fall der Normal-Approximation beschränkt bleibt. 1975 hat Louis Chen [Che76] (ein Schüler von Stein) die Steinsche Methode für die Poisson-Approximation eingeführt. Im Falle der Poisson-Verteilung spricht man in der Literatur daher auch von der Stein-Chen-Methode oder Chen-Stein-Methode. Wir wollen die Methode hier für den Fall der Poisson-Approximation vorstellen. Wir nennen sie fortan einfach die *Steinsche Methode*.

Wir stellen die Grundlagen der Steinschen Methode für Poisson-Approximation vor. Unser erstes Beispiel wird eine Verbesserung des Satzes von Le Cam sein.

Zunächst sammeln wir zwei Beobachtungen zur Poisson-Verteilung.

(i) Es sei $g : \mathbb{N}_0 \to \mathbb{R}$ eine beschränkte Abbildung und es sei Z eine Zufallsgröße, die Poisson-verteilt ist zum Parameter $\lambda > 0$. Wir schreiben dafür im Folgenden immer $Z \sim \mathrm{Po}(\lambda)$. Dann gilt:

$$\mathbb{E}(\lambda g(Z+1)) = \sum_{j\geq 0} \lambda g(j+1) \frac{\lambda^j}{j!} e^{-\lambda}$$

und

$$\mathbb{E}(Z\, g(Z)) = \sum_{l\geq 1} l\, g(l) \frac{\lambda^l}{l!} e^{-\lambda} = \sum_{j\geq 0} \lambda\, g(j+1) \frac{\lambda^j}{j!} e^{-\lambda}.$$

Wir erhalten also insgesamt:

$$\mathbb{E}\Big\{ \lambda\, g(Z+1) - Z\, g(Z) \Big\} = 0$$

für jede beschränkte Abbildung $g : \mathbb{N}_0 \to \mathbb{R}$ und jedes $Z \sim \mathrm{Po}(\lambda)$.

(ii) Es gilt weiter: Ist $f : \mathbb{N}_0 \to \mathbb{R}$ eine beschränkte Abbildung und gilt $\mathbb{E}(f(Z)) = 0$ für $Z \sim \mathrm{Po}(\lambda)$, so kann f immer in der Form

$$f(j) = \lambda\, g_{f,\lambda}(j+1) - j\, g_{f,\lambda}(j), \quad j \geq 0$$

dargestellt werden. Dabei ist $g_{f,\lambda}(\cdot)$ eine beschränkte Abbildung, die definiert ist durch

$$g_{f,\lambda}(j+1) = \frac{j!}{\lambda^{j+1}} \sum_{k=0}^{j} p(k|\lambda) e^{\lambda} f(k) = -\frac{j!}{\lambda^{j+1}} \sum_{k=j+1}^{\infty} p(k|\lambda) e^{\lambda} f(k). \tag{2.1}$$

Der Nachweis erfolgt via einfacher Rekursion (Induktion): Ohne Einschränkung setzen wir $g_{f,\lambda}(0) := 0$. Dann ist für $j = 0$

$$g_{f,\lambda}(1) = \frac{f(0)}{\lambda} = \frac{f(0)}{\lambda}\, e^{\lambda}\, p(0|\lambda)$$

und

$$\begin{aligned} g_{f,\lambda}(j+1) &= \left(f(j) + j\, g_{f,\lambda}(j)\right)\frac{1}{\lambda} \\ &= \frac{1}{\lambda} f(j)p(j|\lambda)\frac{1}{p(j|\lambda)} + \frac{j}{\lambda}\frac{(j-1)!}{\lambda^j}\sum_{k=0}^{j-1} p(k|\lambda)\, e^{\lambda} f(k) \\ &= \frac{j!}{\lambda^{j+1}} f(j)p(j|\lambda)e^{\lambda} + \frac{j!}{\lambda^{j+1}}\sum_{k=0}^{j-1} p(k|\lambda)\, e^{\lambda} f(k) \\ &= \frac{j!}{\lambda^{j+1}}\sum_{k=0}^{j} p(k|\lambda)e^{\lambda} f(k). \end{aligned}$$

Diese Rechnung gilt immer, man benötigt bisher $\mathbb{E}(f(Z)) = 0$ nicht. Dies bedeutet, dass die bisher gefundene Lösung $g_{f,\lambda}$ immer existiert. Es gilt weiter

$$\begin{aligned} 0 = \mathbb{E}(f(Z)) = \sum_{k=0}^{j} \frac{e^{-\lambda}\lambda^k}{k!} f(k) \quad &+ \quad \sum_{k=j+1}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} f(k) \\ \Rightarrow \sum_{k=0}^{j} p(k|\lambda)\, f(k)\, e^{\lambda} \quad &= \quad - \sum_{k=j+1}^{\infty} p(k|\lambda)\, f(k)\, e^{\lambda}, \end{aligned}$$

womit die zweite Darstellung folgt (hier verwenden wir $E(f(Z)) = 0$). Bezeichnen wir mit $||g_{f,\lambda}|| := \sup_{j\geq 0} |g_{f,\lambda}(j)|$ und $||f|| := \sup_{j\geq 0} |f(j)|$, dann folgt mit (2.1)

$$||g_{f,\lambda}|| \leq ||f|| \sum_{k=j+1}^{\infty} \frac{\lambda^{k-(j+1)}}{k!} j! \leq ||f||\, e^{\lambda},$$

also ist $g_{f,\lambda}$ beschränkt.

Wir fassen die obigen Rechnungen zusammen (siehe [Che76]):

Satz 2.2. *(Chen, 1975) Sei $f : \mathbb{N}_0 \to \mathbb{R}$ eine beschränkte Funktion. Dann existiert eine beschränkte Lösung $g_{f,\lambda} : \mathbb{N}_0 \to \mathbb{R}$ von*

$$\lambda\, g_{f,\lambda}(j+1) - j\, g_{f,\lambda}(j) = f(j), \quad j \geq 0$$

genau dann, wenn $\mathbb{E}(f(Z)) = 0$ mit $Z \sim \mathrm{Po}(\lambda)$.

Bemerkungen 2.3. (i) Ist $Z \sim \text{Po}(\lambda)$ und $f : \mathbb{N} \to \mathbb{R}$ eine beschränkte Funktion mit $\mathbb{E}(f(Z)) \neq 0$. Dann folgt

$$|g_{f,\lambda}(j)| \to \infty \text{ für } j \to \infty.$$

(Übungsaufgabe)

(ii) Es sei X eine Zufallsgröße mit Werten in $\mathbb{N}_0$. Dann gilt sogar: X ist Poisson-verteilt mit Parameter $\lambda > 0$ genau dann wenn für jede beschränkte Funktion $g : \mathbb{N}_0 \to \mathbb{R}$ gilt:

$$\mathbb{E}\Big\{ \lambda g(X+1) - X\, g(X) \Big\} = 0.$$

Die Poisson-Verteilung wird also durch diese Gleichheit charakterisiert. Der Beweis ist recht elementar: Betrachte

$$f(j) := h(j) - e^{-\lambda} \sum_{k=0}^{\infty} h(k) \frac{\lambda^k}{k!}$$

für jedes $j \in \mathbb{N}_0$, wobei $h : \mathbb{N}_0 \to \mathbb{R}$ eine beschränkte und reellwertige Funktion ist. f ist beschränkt und erfüllt $\mathbb{E}(f(Z)) = 0$ für $Z \sim \text{Po}(\lambda)$ nach Definition. Nach Satz 2.2 gibt es eine beschränkte Lösung $g_{f,\lambda}$ mit

$$0 = \mathbb{E}\Big\{ \lambda g_{f,\lambda}(X+1) - X\, g_{f,\lambda}(X) \Big\} = \mathbb{E}\Big\{ h(X) - e^{-\lambda} \sum_{k=0}^{\infty} h(k) \frac{\lambda^k}{k!} \Big\},$$

wobei wir für die linke Gleichheit die Voraussetzung verwendet haben. Wählen wir nun speziell die beschränkte Funktion $h_A(j) := I[j \in A]$ für $A \subseteq \mathbb{N}_0$ ($I[j \in A]$ bezeichne die Indikator-Zufallsgröße, die bei Eintritt des Ereignisses $\{j \in A\}$ den Wert 1 annimmt, sonst den Wert 0), so folgt aus der rechten Gleichheit $P(X \in A) - \text{Po}(\lambda)\{A\} = 0$ für jedes $A \subseteq \mathbb{N}_0$, wobei wir die Bezeichnung $\text{Po}(\lambda)\{A\} := \sum_{k \in A} p(k|\lambda)$ verwenden. Somit ist $X \sim \text{Po}(\lambda)$. Einen schönen alternativen Beweis dieser Charakterisierung der Poisson-Verteilung findet man in [Als00]. Dort wird mittels erzeugender Funktionen argumentiert.

Wir schreiben kurz $g_f := g_{f,\lambda}$.

Wir betrachten nun die für das Weitere wesentliche Anwendung: es sei im Folgenden Z immer eine Poisson-verteilte Zufallsgröße zum Parameter λ. Es sei $A \subseteq \mathbb{N}_0$ und $I[j \in A]$ die Indikator-Zufallsgröße. Weiter sei

$$f_A(j) := I[j \in A] - \text{Po}(\lambda)\{A\}, \quad j \geq 0.$$

Wie in obiger Bemerkung verwenden wir die Bezeichnung $\mathrm{Po}(\lambda)\{A\} := \sum_{k\in A} p(k|\lambda)$. Nun ist $f_A(\cdot)$ beschränkt und $\mathbb{E}(f_A(Z)) = 0$ für jedes $A \subseteq \mathbb{N}_0$. Also existiert nach Satz 2.2 eine Lösung der Gleichung

$$\lambda\, g_{A,\lambda}(j+1) - j\, g_{A,\lambda}(j) = I[j \in A] - \mathrm{Po}(\lambda)\{A\}, \quad j \geq 0 \quad \textbf{(STEIN-Gleichungen)} \tag{2.4}$$

für jedes $A \subseteq \mathbb{N}_0$. Mit der Bezeichnung $U_m := \{0, 1, 2, \ldots, m\}$ folgt aus der allgemeinen Lösungsformel (2.1) für $g_{f,\lambda}$:

$$g_{A,\lambda}(j+1) = \frac{j!}{\lambda^{j+1}} e^{\lambda} \big(\mathrm{Po}(\lambda)\{A \cap U_j\} - \mathrm{Po}(\lambda)\{A\}\mathrm{Po}(\lambda)\{U_j\}\big) \ \textbf{(STEIN-Lösungen)}. \tag{2.5}$$

Wieder setzen wir (ohne Einschränkung) $g_{A,\lambda}(0) = 0$.

Den Nutzen dieser Stein-Gleichungen und ihrer Lösungen kann man nun schnell erkennen: Setze für j eine Zufallsgröße W mit Werten in $\mathbb{N}_0$ ein, so folgt aus (2.4), wenn man auf beiden Seiten der Gleichung den Erwartungswert bildet:

$$\mathbb{E}\big(\lambda\, g_{A,\lambda}(W+1) - W g_{A,\lambda}(W)\big) = P(W \in A) - \mathrm{Po}(\lambda)\{A\}.$$

Wir sind an Poisson-Approximationen interessiert. Somit wollen wir also die rechte Seite der Gleichung gleichmäßig in A abschätzen. Die Stein-Gleichungen sagen uns nun, dass wir die linke Seite gleichmäßig in A analysieren sollten. Wir hoffen also auf eine gute Berechenbarkeit der linken Seite unter Ausnutzung der Stein-Lösungen $g_{A,\lambda}(\cdot), A \subseteq \mathbb{N}_0$, und möglicher Schranken dieser Lösungen, gleichmäßig in A. Tatsächlich kann man gute Schranken für die Lösungen und für die Zuwächse der Lösungen finden, siehe Satz 2.9 und Satz 2.13. Diese gelten universell in dem Sinne, als dass sie mit der Zufallsgröße W, deren Verteilung auf Poisson-Appoximierbarkeit hin untersucht wird, nichts zu tun haben. Die Steinschen Methode besteht also aus zwei Schritten:

(i) Zunächst stellt man die charakterisierende Steinsche Gleichung für die zu untersuchende Verteilung auf (hier die Poisson-Verteilung) und versucht die Lösungen der korrespondierenden Gleichungen gut abzuschätzen. Dieser Schritt betrachtet also ausschließlich die *Target-Verteilung*, losgelöst von der Frage, ob eine Zufallsgröße W mittels der Target-Verteilung approximiert werden kann. Natürlich ist dieser Schritt auch das theoretische Herzstück der Methode, denn eine möglichst optimale Abschätzung der Lösungen und der Zuwächse der Lösungen kann universell verwendet werden.

(ii) Dann wird man die Eigenschaften der Zufallsgröße W des stochastischen Modells ins Spiel bringen müssen, hier um den Ausdruck $\mathbb{E}\big(\lambda\, g_{A,\lambda}(W+1) - W g_{A,\lambda}(W)\big)$ abschätzen zu können. In der Anwendung der Steinschen Methode ist dieser Schritt der eigentlich aktive Schritt.

Wir betrachten (als Beispiel) die Zufallsgröße W aus Satz 1.6 (Le Cam):

Es seien $I_1, \dots, I_n$ unabhängige Zufallsgrößen, definiert auf einem gemeinsamen Wahrscheinlichkeitsraum, mit $P(I_j = 1) = p_j = 1 - P(I_j = 0)$, $0 < p_j < 1$, $j = 1, \dots, n$ und

$$W = \sum_{j=1}^{n} I_j; \quad \lambda := \mathbb{E}(W) = \sum_{j=1}^{n} p_j.$$

Sei weiter

$$W_i := \sum_{j=1, j \neq i}^{n} I_j.$$

Wir kürzen wieder ab: $g_A := g_{A,\lambda}$. Es folgt nun

$$\mathbb{E}\big(I_i g_A(W)\big) = \mathbb{E}\big(I_i\, g_A(W_i + 1)\big) = p_i\, \mathbb{E}\big(g_A(W_i + 1)\big), \tag{2.6}$$

denn nach Definition ist I_i stochastisch unabhängig von W_i. Also ist

$$\mathbb{E}\big(\lambda g_A(W+1) - W g_A(W)\big) = \sum_{i=1}^{n} p_i \big(\mathbb{E}(g_A(W+1)) - \mathbb{E}(g_A(W_i+1))\big).$$

Nun stimmen die Zufallsgrößen W und W_i überein, es sei denn $I_i = 1$. Dieses Ereignis hat Wahrscheinlichkeit p_i. Es folgt somit:

$$|P(W \in A) - \mathrm{Po}(\lambda)\{A\}| \leq 2 \sup_{j \geq 0} |g_A(j)| \sum_{i=1}^{n} p_i^2$$

bzw.

$$|P(W \in A) - \mathrm{Po}(\lambda)\{A\}| \leq \sup_{j \geq 0} |g_A(j+1) - g_A(j)| \sum_{i=1}^{n} p_i^2.$$

Wenn wir nun noch die Größen

$$\begin{aligned} ||g|| &:= ||g_{A,\lambda}|| := \sup_{j\geq 0} |g_{A,\lambda}(j+1) - g_{A,\lambda}(j)| \qquad \text{und} \\ \Delta g &:= \Delta g_{A,\lambda} := \sup_{j \geq 0} |g_{A,\lambda}(j+1) - g_{A,\lambda}(j)| \end{aligned}$$

gleichmäßig in $A \subseteq \mathbb{N}_0$ beschränken können, haben wir für den unabhängigen Fall die Funktionsweise der Steinschen Methode bereits vollständig erfaßt.

Bemerkung 2.7. Nach Kapitel 1 hoffen wir, dass entweder $2\,||g||$ oder Δg kleiner 1 ist (sonst hätte sich der Aufwand nicht gelohnt).

Nun wollen wir Δg, also die Zuwächse von $g_{A,\lambda}$, abschätzen. Sei $Z \sim \mathrm{Po}(\lambda)$. Dann ist

$$f_A(j) = \sum_{k\in A}\big(I[j=k] - P(Z=k)\big),$$

denn Wahrscheinlichkeiten sind abzählbar-additiv und die Indikatorvariable zu einer disjunkten Vereinigung von Ereignissen ist gleich der Summe der einzelnen Indikatoren. Sei nun g_k die zugehörige Lösung der STEIN-Gleichung zu

$$f_k(j) := I[j=k] - P(Z=k), \quad j \geq 0,$$

so vermutet man, dass die folgende Identität gilt:

$$g_A(j) = \sum_{k\in A} g_k(j), \tag{2.8}$$

denn bei endlich vielen Summanden folgt dies sofort aus der („Linearität"der) Stein-Gleichung. Um (2.8) allgemein einzusehen, betrachten wir eine beliebige Menge $A \subset \mathbb{N}_0$ und führen diesen Fall via Abschneidetechnik auf den endlichen Fall zurück. Es ist für ein $M > 0$

$$f_A = \sum_{\substack{k\in A\\ k\leq M}} f_k + f_{A\cap\{k:k>M\}}.$$

Im Folgenden kürzen wir ab: $\{k : k > M\} =: \{> M\}$. Da f_A so als eine endliche Summe dargestellt wird, gilt für die zu f_A gehörige Stein-Lösung:

$$g_A = \sum_{\substack{k\in A\\ k\leq M}} g_k + g_{A\cap\{>M\}}.$$

Nun ist nach Definition $f_{A\cap\{>M\}}(j) = -P\big(Z \in A\cap\{>M\}\big)$ für alle $j \leq M$. Es sei $j < M$, dann gilt für die Stein-Lösung nach (2.5):

$$g_{A\cap\{>M\}}(j+1) = \frac{-P\big(Z\in A\cap\{>M\}\big)}{\lambda}\left(1 + \frac{j}{\lambda} + \frac{j(j-1)}{\lambda^2} + \cdots + \frac{j!}{\lambda^j}\right),$$

also

$$|g_A(j+1) - \sum_{\substack{k\in A\\ k\leq M}} g_k(j+1)| \leq \mathrm{const}(\lambda,j)\, P\big(Z\in A\cap\{>M\}\big) \leq \mathrm{const}(\lambda,j) P\big(Z>M\big).$$

Nun konvergiert die rechte Seite für $M \to \infty$ gegen Null, womit (2.8) gezeigt ist. Wir verwenden diese Vorbereitung, indem wir uns nun auf die Zuwächse der g_k konzentrieren.

Wir erinnern uns an (2.1):

$$g_k(j+1) = \frac{1}{\lambda}\left(f_k(j) + \frac{j}{\lambda} f_k(j-1) + \frac{j(j-1)}{\lambda^2} f_k(j-2) + \cdots + \frac{j!}{\lambda^j} f_k(0)\right).$$

Nach Definition ist

$$f_k(j) = \begin{cases} -P(Z=k), & j<k, \\ 1-P(Z=k), & j=k, \\ -P(Z=k), & j>k. \end{cases}$$

Also folgt für $j<k$:

$$g_k(j+1) = -\frac{P(Z=k)}{\lambda}\left(1+\frac{j}{\lambda}+\cdots+\frac{j!}{\lambda^j}\right) \le 0,$$

und daher ist hier $g_k(j+1)-g_k(j) \le 0$, also g_k monoton fallend. Es sei nun $j>k$. Mit (2.5) ist

$$\begin{aligned} g_k(j+1) &= e^\lambda\, j!\, \lambda^{-j-1}\left(p(k|\lambda)I[k\le j] - p(k|\lambda)\,\mathrm{Po}(\lambda)\{U_j\}\right) \\ &= e^\lambda\, j!\, \lambda^{-j-1} p(k|\lambda)\big(1-\mathrm{Po}(\lambda)\{U_j\}\big) \\ &\ge 0. \end{aligned}$$

Wir verwenden diese Darstellung und $k \le j-1$ (da $k<j$):

$$\begin{aligned} g_k(j+1)-g_k(j) &= \lambda^{-j}\,(j-1)!\,p(k|\lambda)\left(\frac{j}{\lambda}\sum_{l=j+1}^{\infty}\frac{\lambda^l}{l!} - \sum_{l=j}^{\infty}\frac{\lambda^l}{l!}\right) \\ &\le \lambda^{-j}\,(j-1)!\,p(k|\lambda)\left(\sum_{l=j}^{\infty}\frac{\lambda^l}{l!} - \sum_{l=j}^{\infty}\frac{\lambda^l}{l!}\right) \\ &= 0. \end{aligned}$$

Somit ist g_k in diesem Bereich ebenfalls monoton fallend, womit nur der Zuwachs $g_k(k+1)-g_k(k)$ positiv ist, und es gilt:

$$\begin{aligned} g_k(k+1)-g_k(k) &= \frac{k!}{\lambda^{k+1}}P(Z=k)\left(\sum_{j=k+1}^{\infty}\frac{\lambda^j}{j!}\right) + \frac{(k-1)!}{\lambda^k}P(Z=k)\left(\sum_{j=0}^{k-1}\frac{\lambda^j}{j!}\right) \\ &= e^{-\lambda}\left(\frac{1}{\lambda}\sum_{j=k+1}^{\infty}\frac{\lambda^j}{j!} + \frac{1}{k}\sum_{j=0}^{k-1}\frac{\lambda^j}{j!}\right) \\ &\le e^{-\lambda}\left(\frac{1}{\lambda}\sum_{j=k+1}^{\infty}\frac{\lambda^j}{j!} + \frac{1}{\lambda}\sum_{j=1}^{k}\frac{\lambda^j}{j!}\right) \\ &= \frac{e^{-\lambda}}{\lambda}\big(e^\lambda-1\big) = \frac{1}{\lambda}\big(1-e^{-\lambda}\big) \\ &\le \min\left(1,\frac{1}{\lambda}\right). \end{aligned}$$

Bemerke, dass die Ungleichung in der zweiten Zeile im Fall $k = 1$ eine Gleichung ist. Somit haben wir die Zuwächse g_k analysiert. Es folgt nun mit (2.8):

$$g_A(j+1) - g_A(j) = \sum_{k \in A} \big(g_k(j+1) - g_k(j)\big);$$

hier sind alle Summanden negativ außer im Fall $j = k$. Also ist

$$g_A(j+1) - g_A(j) \leq g_j(j+1) - g_j(j),$$

falls $j \in A$, sonst ist $g_A(j+1) - g_A(j) \leq 0$. Es gilt also

$$g_A(j+1) - g_A(j) \leq \frac{1}{\lambda}\big(1 - e^{-\lambda}\big)$$

für alle $A \subseteq \mathbb{N}_0$ und für alle $j \geq 0$. Wenn $g_A(j+1) - g_A(j) \geq 0$ für alle $j \geq 0$, so folgt

$$\Delta g \leq \frac{1}{\lambda}\big(1 - e^{-\lambda}\big).$$

Im Fall $g_A(j+1) - g_A(j) < 0$ ist

$$0 < -g_A(j+1) + g_A(j) = g_{A^c}(j+1) - g_{A^c}(j) \leq \frac{1}{\lambda}\big(1 - e^{-\lambda}\big),$$

denn $f_A + f_{A^c} = f_{\mathbb{Z}_+} \equiv 0$, und somit folgt für die korrespondierenden Stein-Lösungen: $g_A = -g_{A^c}$. Wir haben somit gezeigt (siehe [BE83]):

Satz 2.9. *(Barbour und Eagleson, 1983) Für die Zuwächse der Stein-Lösungen finden wir gleichmäßig in $A \subset \mathbb{N}_0$:*

$$\Delta g \leq \frac{1}{\lambda}\big(1 - e^{-\lambda}\big) \leq \min\left(1, \frac{1}{\lambda}\right).$$

Bemerkung 2.10. Ist $A = \{1\}$, so zeigte die obige Rechnung, dass die Schranke exakt ist: $\Delta g = \frac{1}{\lambda}\big(1 - e^{-\lambda}\big)$.

Wir haben nun im Falle der unabhängigen Indikatoren den Satz von Le Cam verbessert (siehe [BH84]):

Satz 2.11. *(Barbour und Hall, 1984) In der Notation von Satz 1.6 gilt:*

$$d_{TV}\big(\mathcal{L}(W), \mathrm{Po}(\mathbb{E}W)\big) \leq \frac{1}{\lambda}\left(1 - e^{-\lambda}\right) \sum_{i=1}^{n} p_i^2.$$

Le Cam hat in [LC60] mittels der Fourier-Transformierten gezeigt, dass

$$d_{TV}\big(\mathcal{L}(W), \mathrm{Po}(\mathbb{E}W)\big) \leq 4.5 \max_i p_i$$

gilt. Weiter zeigte er unter der Bedingung $\max_i p_i \leq \frac{1}{4}$:

$$d_{TV}\big(\mathcal{L}(W), \mathrm{Po}(\mathbb{E}W)\big) \leq \min\big(1, \frac{8}{\lambda}\big) \sum_{i=1}^{n} p_i^2.$$

Die Konstante 8 wurde in [Ker64] von Kerstan verbessert zu 1.05 und in [DVJ88] von Daley und Vere-Jones zu 0.71. Barbour und Hall haben in [BH84] die folgende untere Schranke bewiesen, mit der wir uns in einem späteren Kapitel beschäftigen werden:

$$d_{TV}\big(\mathcal{L}(W), \mathrm{Po}(\mathbb{E}W)\big) \geq \frac{1}{32} \min\big(1, \frac{1}{\lambda}\big) \sum_{i=1}^{n} p_i^2.$$

Für $\lambda > 1$ und insbesondere für große λ ist Satz 2.11 eine deutliche Verbesserung von Satz 1.6. Es folgt im Falle der Binomialverteilung:

Satz 2.12.

$$d_{TV}\big(B(n,p), \mathrm{Po}(n\,p)\big) \leq n\,p^2 \min\left(1, \frac{1}{n\,p}\right) = \mathcal{O}(p).$$

Wir geben eine Schranke für $\|g\|$ ohne Beweis an (siehe [BE83]):

Satz 2.13. *(Barbour und Eagleson, 1983) Für die durch* (2.1) *gegebene Funktion* $g_{f,\lambda}$ *gilt:*

$$||g_{f,\lambda}|| \leq 2\,||f||\, \min\left(1, \frac{1}{\sqrt{\lambda}}\right).$$

Für die Lösungen der Stein-Gleichungen gilt die etwas feinere Abschätzung (für alle $A \subset \mathbb{N}_0$*):*

$$||g|| \leq \min\left(1, \sqrt{\frac{2}{e\,\lambda}}\right).$$

(siehe [BHJ92, (10.2.4)]).

Der eigentliche Erfolg der Steinschen Methode liegt nun darin, dass die Annahme der Unabhängigkeit sehr einfach abgeschwächt werden kann. Im Beispiel (Modell aus Satz 1.6 und 2.11) haben wir die Unabhängigkeit nur für die Identität

$$\mathbb{E}\big(I_i\, g(W)\big) = p_i\, \mathbb{E}\big(g(W_i + 1)\big)$$

benötigt. In den folgenden zwei Kapiteln lernen wir zwei Ansätze kennen, die in abhängigen Modellen die fehlende Gleichheit quasi auffangen.

Zum Thema Poisson-Approximation ist die Monographie [BHJ92] von Barbour, Holst und Janson empfehlenswert. Dort findet man eine große Anzahl von Beispielen, die wir in den folgenden Kapiteln teilweise kennenlernen. Das Buch von M.

Penrose [Pen03] stellt ebenfalls die Steinsche Methode für Poisson-Approximation vor und wendet sie für die Modellklasse der sogenannten *Random geometic graphs* an. Wir werden uns in einigen Kapiteln mit der Beispielklasse der *Zufallsgraphen* im Sinne von Erdös und Rényi beschäftigen.

Ausblick: Die zentralen Fragen zur Steinschen Methode sind:

(i) Warum funktioniert diese Methode in so vielen Beispielen (bei denen zuvor nur mühsam oder kein Approximations-Resultat hergeleitet werden konnte)?

(ii) Wie kann man die Charakterisierung der Poisson-Verteilung interpretieren (um auch für andere Verteilungen eine Steinsche Methode herleiten zu können)?

Dazu berichten wir kurz von den Sichtweisen von Charles Stein und Andrew Barbour: Ganz allgemein stellt sich das folgende Problem: Es sei $(S, \mathcal{S}, \mu)$ ein Wahrscheinlichkeitsraum, und χ eine Menge von meßbaren Funktionen $f : S \to \mathbb{R}$. Weiter sei $\chi_0 \subset \chi$ so gewählt, dass alle $f \in \chi_0$ μ-integrierbar sind. Man möchte die Integrale $\int_S f\, d\mu$ für alle $f \in \chi_0$ berechnen. Die Steinsche Methode kann man nun so verstehen: Man wähle ein Wahrscheinlichkeitsmaß μ_0 so dass alle $f \in \chi_0$ μ_0-integrierbar sind, und alle Integrale $\int_S f\, d\mu_0$ "einfach" auszurechnen sind. Nun bestimme man eine Menge von Funktionen $\mathcal{F}_0$ and eine Abbildung $T_0 : \mathcal{F}_0 \to \chi$, so dass für alle $f \in \chi_0$ die Gleichung

$$T_0 g = f - \int_S f\, d\mu_0 \tag{2.14}$$

eine Lösung $g \in \mathcal{F}_0$ hat. Dann gilt

$$\Big|\int_S f\, d\mu - \int_S f\, d\mu_0\Big| \leq \int_S |(T_0 g)|\, d\mu.$$

Man nennt hier (2.14) die *Steinsche Gleichung*, T_0 heißt der *Stein-Operator* und g die *Stein-Transformierte* oder die *Stein-Lösung* . Stein schlägt in [Ste86] den folgenden Weg vor, einen geeigneten Stein-Operator T_0 für das Maß μ_0 zu finden.

(i) Man wähle ein *austauschbares Paar* von Zufallsgrößen (X, Y) mit Marginalverteilung μ_0.

(ii) Wähle eine Abbildung $\alpha : \mathcal{F}_0 \to \mathcal{F}$, wobei $\mathcal{F}$ der Raum der meßbaren antisymmetrischen Funktionen $F : S^2 \to \mathbb{R}$ mit $\mathbb{E}\big(|F(X,Y)|\big) < \infty$ ist.

(iii) Wähle nun $T_0 = T \circ \alpha$ mit $T : \mathcal{F} \to \chi$ gegeben durch

$$(TF)(x) := \mathbb{E}\big(F(X,Y) | X = x\big).$$

Diese Prozedur garantiert nicht, einen geeigneten Stein-Operator zu finden, aber die Erfahrung zeigt, dass es doch häufig klappt. Wir betrachten diesen Ansatz im Fall der Poisson-Verteilung: Hier ist $(S, \mathcal{S}, \mu) = (\mathbb{Z}_+, \mathcal{B}_{\mathbb{Z}_+}, \mu)$, und $\mu_0 = \mathrm{Po}(\lambda)$. Wir setzen $\mathcal{F}_0 = \chi$ and definieren $T_0 \colon \chi \to \chi$ mittels

$$(T_0 g)(k) = \lambda g(k+1) - k g(k), \quad k \in \mathbb{Z}_+.$$

Dabei lassen wir uns im Moment einfach davon leiten, dass wir die Steinsche Gleichung oben schon gesehen haben. Im allgemeinen stellt sich natürlich die Frage, wie man auf T_0 kommt. Es folgt hier, dass die Steinsche Gleichung

$$T_0 g = f - \int_{\mathbb{Z}_+} f \, d\mu_0$$

eine eindeutige Lösung g für jede μ_0-integrierbare Funktion f hat (ausgenommen $g(0)$, was beliebig gewählt werden kann). Wir haben diese Lösung zu Beginn des Kapitels schon gesehen:

$$\begin{aligned} g(k) &= \frac{(k-1)!}{\lambda^k} \sum_{i=0}^{k-1} \big(f(i) - \int_{\mathbb{Z}_+} f \, d\mu_0\big) \frac{\lambda^i}{i!} \\ &= -\frac{(k-1)!}{\lambda^k} \sum_{i=k}^{\infty} \big(f(i) - \int_{\mathbb{Z}_+} f \, d\mu_0\big) \frac{\lambda^i}{i!}. \end{aligned}$$

Weiter gilt die schon gezeigte Charakterizierung: Ein Wahrscheinlichkeitsmaß μ auf $\mathbb{Z}_+$ ist identisch $\mu_0 = \mathrm{Po}(\lambda)$ genau dann wenn

$$\int_{Z_+} \big(T_0 g\big) \, d\mu = 0$$

für alle beschränkten $g \colon \mathbb{Z}_+ \to \mathbb{R}$.

Um die von Stein vorgeschlagene Prozedur etwas zu verstehen, betrachten wir die sogenannte *Generator-Methode* , die von Barbour 1988 in [Bar88] vorgeschlagen wurde: Es sei $\{Z_t\}_{t \in \mathbb{R}_+}$ ein stationärer *Geburts-Todes-Prozess* auf $\mathbb{Z}_+$ mit konstanter Todesrate λ und Geburtsrate $\mu_k = k$. Dieser Prozess ist reversibel mit der stationären Verteilung $\mu_0 = \mathrm{Po}(\lambda)$. Damit ist (Z_0, Z_t) ein austauschbares Paar mit Marginalverteilung μ_0. Nun wählen wir die Abbildung $\alpha \colon \chi \to \mathcal{F}$ durch

$$(\alpha g)(x, y) = g(y) - g(x)$$

(definiert für Funktionen g, die nicht zu schnell wachsen). Die Abbildung $T \colon \mathcal{F} \to \chi$ definieren wir durch

$$(TF)(k) := \mathbb{E}\big(F(Z_0, Z_t) | Z_0 = k\big).$$

Dann folgt

$$\begin{aligned}
\lim_{t\downarrow 0}\big(T\circ\alpha g\big)(k) &= \lim_{t\downarrow 0}\frac{\mathbb{E}\big(g(Z_t)|Z_0=k\big)-g(k)}{t}\\
&= \lambda g(k+1)+kg(k-1)-(\lambda+k)g(k)\\
&= \big(\mathcal{A}g\big)(k)=\big(T_0\nabla g\big)(k).
\end{aligned}$$

Hierbei bezeichnet $\nabla g(k)=g(k)-g(k-1)$ und $\mathcal{A}$ den *Generator* des Stochastischen Prozesses $\{Z_t\}_{t\in\mathbb{R}_+}$. Die korrespondierende Steinsche Gleichung (eine Poisson Gleichung!) ist

$$-\mathcal{A}g=f-\int_{\mathbb{Z}_+} f\,d\mu_0. \tag{2.15}$$

Nun liefert abschliessend die Steinsche Gleichung für jedes $A\subset\mathbb{Z}_+$

$$\mu(A)-\mu_0(A)=\int_{\mathbb{Z}_+}\big(T_0 g_A\big)\,d\mu=\mathbb{E}\big(\lambda g_A(W+1)-Wg_A(W)\big),$$

wobei g_A die Lösung der Steinschen Gleichung zu $f_A(\cdot)=I[\cdot\in A]$ bezeichnet und μ die Verteilung der Zufallsgröße W.

Dieser Weg wirkt in der ersten Begegnung nicht unbedingt einleuchtend. Er zeigt, wie Steins Idee der austauschbaren Paare im Beispiel der Poisson-Verteilung verläuft. Einleuchtender erscheint die Idee von Barbour: im Falle der Poisson-Verteilung wähle man einen geeigneten Markov-Prozess, dessen invariante Limesverteilung $\mathrm{Po}(\lambda)$ ist. Der Generator dieses Prozesses liefert die korrespondierende Steinsche Gleichung! Dies lässt den allgemeinen Weg der Generator-Methode erkennen: Zur Targetverteilung μ_0 wähle man einen geeigneten (!) Markov-Prozess mit invarianter Limesverteilung μ_0. Der Generator des Prozesses liefert eine Steinsche Gleichung. In der Literatur findet man Steins Zugang mittels austauschbarer Paare und die Generatormethode von Barbour gleichermaßen. In weiteren Ausblicken wird uns zumindest die Generator-Methode wieder begegnen.

In diesem Kapitel entwickeln wir einen speziellen Ansatz der Steinschen Methode, der in der Literatur der *lokale Ansatz* genannt wird. Er geht auf die Arbeit von Chen [Che76] zurück. Grob gesprochen handelt es sich hier um die Analyse von Modellen, bei denen die Zufallsgrößen lokale Abhängigkeiten aufweisen. Zunächst betrachten wir Schranken für die Totalvariation der Verteilung einer Zufallsgröße W und der Poisson-Verteilung. Anschliessend betrachten wir drei Beispiele von Summen abhäniger Zufallsgrößen und untersuchen deren Poisson-Approximation. Es sind dies k-Runs, das Geburtstagsproblem und ein Beispiel aus der Zuverlässigkeits-Theorie. Weitere Anwendungen des lokalen Ansatzes findet man in den Monographien [BHJ92] und [Pen03]. In den vergangenen 15 Jahren wurde der lokale Ansatz insbesondere beim Studium biomolekularer Probleme verwendet. In [Jia02] findet man eine gute Übersicht über die Publikationen dazu. Erste Resultate findet man in der Monographie [Wat95] von Waterman.

Die Theorie:

Sei fortan Γ eine endliche Indexmenge und $(I_\alpha)_{\alpha\in\Gamma}$ eine endliche Menge von Indikatorvariablen, definiert auf einem gemeinsamen Wahrscheinlichkeitsraum. Betrachte $W := \sum_{\alpha\in\Gamma} I_\alpha$. Zu jedem $\alpha \in \Gamma$ sei Γ_α die Menge der $\beta \in \Gamma \setminus \{\alpha\}$, für die gilt: I_β ist abhängig von I_α. Es soll also gelten: für jedes $\beta \in (\Gamma_\alpha \cup \{\alpha\})^c$ ist I_β unabhängig von I_α. Es sei weiter

$$Z_\alpha := \sum_{\beta\in\Gamma_\alpha} I_\beta \quad \text{und} \quad W'_\alpha := \sum_{\beta\in(\Gamma_\alpha\cup\{\alpha\})^c} I_\beta.$$

Dann gilt für jedes $\alpha \in \Gamma$:

$$W = I_\alpha + Z_\alpha + W'_\alpha.$$

Weiter bezeichnen wir mit $\pi_\alpha := \mathbb{E}I_\alpha$ und $\lambda := \mathbb{E}W = \sum_{\alpha\in\Gamma} \pi_\alpha$. Dann gilt der folgende in [AGG89] bewiesene Satz:

Satz 3.1. *(Lokaler Ansatz I, Arratia, Goldstein, Gordon, 1989) Mit den obigen Notationen gilt ganz allgemein:*

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \le \min\left(1, \frac{1}{\lambda}\right)\left\{\sum_{\alpha\in\Gamma}\pi_\alpha^2 + \sum_{\alpha\in\Gamma}\pi_\alpha\,\mathbb{E}Z_\alpha + \sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha}\mathbb{E}(I_\alpha\, I_\beta)\right\}.$$

Bemerkung 3.2. Zunächst besagt dieses abstrakte Resultat nur so viel: sind die π_α klein und die $\mathbb{E}(I_\alpha\, I_\beta)$ ebenfalls (keine "starke" Abhängigkeit), so ist die gefundene Schranke gut. In diesen Fällen kann man guten Gewissens die Verteilung von

W durch die Poissonverteilung zum Parameter λ approximieren. Die Beispiele im zweiten Teil des Kapitels füllen diese vage Interpretation mit Leben.

Beweis von Satz 3.1: Wir erinnern uns an den Ansatz von Stein, nach dem für jede Teilmenge $A \subset \mathbb{N}_0$ die Identität

$$P(W \in A) - Po(\lambda)\{A\} = \mathbb{E}\big(\lambda\, g_A(W+1) - W g_A(W)\big)$$

gilt, wobei g_A (genauer $g_{A,\lambda}$) die STEIN-Lösung - siehe (2.5) - bezeichnet. Wir bestimmen für jedes $A \subset \mathbb{N}_0$ die rechte Seite dieser Identität. Zunächst berechnen wir $\mathbb{E}\big(I_\alpha\, g_A(W)\big)$:

Es gilt die folgende Identität (zwischen Zufallsgrößen):

$$I_\alpha\, g_A(W) = I_\alpha\, g_A(Z_\alpha + W'_\alpha + 1),$$

da $I_\alpha \in \{0,1\}$. Wir schreiben die rechte Seite als

$$I_\alpha\, g_A(W'_\alpha + 1) + I_\alpha\big\{g_A(Z_\alpha + W'_\alpha + 1) - g_A(W'_\alpha + 1)\big\}.$$

Somit folgt mit der Definition Δg der Zuwächse von g_A (vergleiche Kapitel 2):

$$|I_\alpha\, g_A(W) - I_\alpha\, g_A(W'_\alpha + 1)| \le I_\alpha\, Z_\alpha\, \Delta g,$$

also bei Bildung des Erwartungswertes:

$$|\mathbb{E}\big(I_\alpha\, g_A(W)\big) - \mathbb{E}\big(I_\alpha\, g_A(W'_\alpha + 1)\big)| \le \mathbb{E}\big(I_\alpha\, Z_\alpha\big)\Delta g.$$

Nun ist nach Definition von W'_α diese Zufallsgröße unabhängig von I_α, also ist $\mathbb{E}\big(I_\alpha\, g_A(W'_\alpha + 1)\big) = \pi_\alpha \mathbb{E}\big(g_A(W'_\alpha + 1)\big)$. Bilden wir nun die Summe über alle $\alpha \in \Gamma$, haben wir insgesamt gezeigt:

$$\left|\mathbb{E}\big(W\, g_A(W)\big) - \sum_{\alpha\in\Gamma} \pi_\alpha \mathbb{E}\big(g_A(W'_\alpha + 1)\big)\right| \le \Delta g \sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha} \mathbb{E}\big(I_\alpha\, I_\beta\big).$$

Weiter gilt

$$\mathbb{E}\big(\lambda\, g_A(W+1)\big) = \sum_{\alpha\in\Gamma} \pi_\alpha \mathbb{E}\big(g_A(W+1)\big)$$

und mit $W - W'_\alpha = I_\alpha + Z_\alpha$ folgt

$$|g_A(W+1) - g_A(W'_\alpha + 1)| \le \Delta g\,(I_\alpha + Z_\alpha).$$

Also

$$\begin{aligned}\left|\mathbb{E}\big(\lambda g_A(W+1)\big) - \sum_{\alpha\in\Gamma} \pi_\alpha \mathbb{E}\big(g_A(W'_\alpha + 1)\big)\right| &\le \Delta g \sum_{\alpha\in\Gamma} \pi_\alpha\, \mathbb{E}(I_\alpha + Z_\alpha) \qquad (3.3)\\ &= \Delta g \sum_{\alpha\in\Gamma}\big(\pi_\alpha^2 + \pi_\alpha\, \mathbb{E}Z_\alpha\big).\end{aligned}$$

Zusammen folgt somit unsere Behauptung durch Anwendung der Dreiecksungleichung:

$$\begin{aligned} \big|P(W \in A) - Po(\lambda)\{A\}\big| &= \big|\mathbb{E}\big(\lambda g_A(W+1) - W\, g_A(W)\big)\big| \\ &\leq \Delta g \Big\{ \sum_{\alpha\in\Gamma} \pi_\alpha^2 + \sum_{\alpha\in\Gamma} \pi_\alpha\, \mathbb{E} Z_\alpha + \sum_{\alpha\in\Gamma} \sum_{\beta\in\Gamma_\alpha} \mathbb{E}(I_\alpha\, I_\beta) \Big\}. \end{aligned}$$

Die Abschätzung für Δg nach Satz 2.9 vervollständigt den Beweis. □

Der Satz sollte im Fall von *unabhängigen* Zufallsgrößen $(I_\alpha)_{\alpha\in\Gamma}$ das Resultat von Satz 2.11 reproduzieren. Da zu festem $\alpha \in \Gamma$ hier nun alle anderen Indikatorvariablen I_β unabhängig von I_α sind, setzen wir $\Gamma_\alpha = \emptyset$ für jedes $\alpha \in \Gamma$. Dann ist insbesondere $Z_\alpha = 0$ für jedes $\alpha \in \Gamma$. Somit folgt unmittelbar

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq \min\Big(1, \frac{1}{\lambda}\Big) \sum_{\alpha\in\Gamma} \pi_\alpha^2.$$

Bevor wir Beispiele diskutieren, bei denen die auftretenden I_α abhängig sind, diskutieren wir noch eine Verallgemeinerung von Satz 3.1.

Man läßt nun zu, dass die Indikatorvariablen in der Summe W'_α auch noch von I_α abhängen, aber schwächer als die Summanden in Z_α. Es sei fortan zu jedem $\alpha \in \Gamma$ eine Zerlegung der Form

$$\Gamma = \{\alpha\} \cup \Gamma_\alpha \cup \Gamma_\alpha^w$$

gegeben. Hierbei steht Γ_α^w für Indizes, deren Indikatorvariable schwach (weak) von I_α abhängt. Wir verwenden nun die Notation

$$Z_\alpha := \sum_{\beta\in\Gamma_\alpha} I_\beta, \qquad W'_\alpha := \sum_{\beta\in\Gamma_\alpha^w} I_\beta.$$

Weiter sei

$$\eta_\alpha := \mathbb{E}\big|\mathbb{E}(I_\alpha | W'_\alpha) - \pi_\alpha\big|.$$

Dabei bezeichnet $\mathbb{E}(I_\alpha | W'_\alpha)$ den bedingten Erwartungswert von I_α, gegeben W'_α. Es gilt nun der folgende Satz, der ebenfalls in [AGG89] bewiesen wurde:

Satz 3.4. *(Lokaler Ansatz II, Arratia, Goldstein, Gordon, 1989) Mit den obigen Notationen gilt:*

$$\begin{aligned} d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) &\leq \min\Big(1, \frac{1}{\lambda}\Big) \Big\{ \sum_{\alpha\in\Gamma} \pi_\alpha^2 + \sum_{\alpha\in\Gamma} \pi_\alpha\, \mathbb{E} Z_\alpha + \sum_{\alpha\in\Gamma} \sum_{\beta\in\Gamma_\alpha} \mathbb{E}(I_\alpha\, I_\beta) \Big\} \\ &\quad + \min\Big(1, \frac{1}{\sqrt{\lambda}}\Big) \sum_{\alpha\in\Gamma} \eta_\alpha. \end{aligned}$$

Bemerkung 3.5. Bei diesem Resultat spielt die Schranke von $||g||$ auch eine Rolle. Mit $1/\sqrt{\lambda}$ wird der zweite Summand in der Größenordnung dominieren. Die hier (zunächst abstrakt) angenommene stärkere Abhängigkeit der Indikatorvariablen kostet also den Preis einer im allgemeinen schlechteren Schranke. Weiterhin folgt Satz 3.1 aus Satz 3.4, wenn W'_α und I_α unabhängig sind für jedes $\alpha \in \Gamma$.

Beweis von Satz 3.4: Der wesentliche Unterschied zum Beweis von Satz 3.1 ist, dass nun nicht mehr die Identität

$$\mathbb{E}\big(I_\alpha\, g_A(W'_\alpha + 1)\big) = \pi_\alpha\, \mathbb{E}\big(g_A(W'_\alpha + 1)\big)$$

gilt. Daher kümmern wir uns um die Differenz der linken und rechten Seite. Es gilt:

$$\begin{aligned}&\Big|\sum_{\alpha\in\Gamma} \pi_\alpha \mathbb{E}\big(g_A(W'_\alpha + 1)\big) - \sum_{\alpha\in\Gamma} \mathbb{E}\big(I_\alpha\, g_A(W'_\alpha + 1)\big)\Big| \\ &\leq \sum_{\alpha\in\Gamma} \Big|\mathbb{E}\big((I_\alpha - \pi_\alpha)\, g_A(W'_\alpha + 1)\big)\Big|.\end{aligned}$$

Nun gilt nach Definition des bedingten Erwartungswertes:

$$\mathbb{E}\big((I_\alpha - \pi_\alpha)\, g_A(W'_\alpha + 1)\big) = \mathbb{E}\Big(\mathbb{E}\big((I_\alpha - \pi_\alpha)|W'_\alpha\big)\, g_A(W'_\alpha + 1)\Big).$$

Somit können wir die rechte Seite in 3.6 weiter abschätzen:

$$\begin{aligned}&\leq \sum_{\alpha\in\Gamma} \mathbb{E}\Big(\big|\mathbb{E}\big((I_\alpha - \pi_\alpha)|W'_\alpha\big)\big|\big|g_A(W'_\alpha + 1)\big|\Big) \\ &\leq ||g|| \sum_{\alpha\in\Gamma} \mathbb{E}\big|\mathbb{E}(I_\alpha|W'_\alpha) - \pi_\alpha\big| \\ &= ||g|| \sum_{\alpha\in\Gamma} \eta_\alpha.\end{aligned}$$

Wir verwenden nun noch die Schranke für $||g||$ aus Satz 2.13 und haben den Satz bewiesen. □

Bemerkung 3.6. Es gibt Beispiele, bei denen die optimale Größenordnung durch $\frac{1}{\sqrt{\lambda}}$ gegeben ist. Wir diskutieren dies allerdings nicht. In den hier betrachteten Beispielen wird immer $\eta_\alpha = 0$ für alle $\alpha \in \Gamma$ sein.

Korollar 3.7.

$$|P(W = 0) - e^{-\lambda}| \leq \frac{1 - e^{-\lambda}}{\lambda} \left\{ \sum_{\alpha\in\Gamma} \pi_\alpha^2 + \sum_{\alpha\in\Gamma} \pi_\alpha \mathbb{E} Z_\alpha + \sum_{\alpha\in\Gamma} \mathbb{E}(I_\alpha\, Z_\alpha) + \sum_{\alpha\in\Gamma} \eta_\alpha \right\}.$$

Beweis: Für $f_{\{0\}}(j) := 1_{\{0\}}(j) - e^{-\lambda}$ gilt $||g|| = g_{\{0\}}(1) = \frac{1-e^{-\lambda}}{\lambda}$. Dies folgt unmittelbar aus der Definition der Stein-Lösung $g_{\{0\}}$. □

Bemerkung 3.8. Was ist eine andere Methode, Poisson-Approximation zu untersuchen? Eine Methode ist die der *faktoriellen Momente* . Für $x \in \mathbb{R}$ und ein $r \in \mathbb{N}$ bezeichne $(x)_r = x(x-1)\cdots(x-r+1)$. Ist X eine diskrete Zufallsvariable auf einem Wahrscheinlichkeitsraum (Ω, P), dann bezeichnet $\mathbb{E}_r(X) := \mathbb{E}((X)_r)$ das r-te faktorielle Moment von X. Es seien nun $X_1, X_2, \ldots$ weitere Zufallsgrößen mit Werten in $\mathbb{N}_0$. Es gelte

$$\lim_{n\to\infty} \mathbb{E}_k(X_n) = \mathbb{E}_k(X)$$

für alle $k \in \mathbb{N}_0$ und $\lim_{k\to\infty} \mathbb{E}_k(X)k^m/k! = 0$ für alle $m \in \mathbb{N}_0$, dann folgt

$$\lim_{n\to\infty} d_{TV}(X_n, X) = 0.$$

Im Fall eine Poisson-verteilten Zufallsvariable X gilt: Es seien $\lambda = \lambda(n)$ eine nichtnegative beschränkte Funktion auf $\mathbb{N}$ und $X_1, X_2, \ldots$ Zufallsgrößen mit Werten in $\mathbb{N}_0$, für die

$$\lim_{n\to\infty} \left(\mathbb{E}_k(X_n) - \lambda^k\right) = 0$$

für alle $k \in \mathbb{N}_0$ gilt. Dann folgt

$$\lim_{n\to\infty} d_{TV}(X_n, Po(\lambda)) = 0.$$

Allgemein wird man in den Beispielen zum lokalen Ansatz der Steinschen Methode feststellen: Die Steinsche Methode

(i) ist einfacher anzuwenden; man muß nur Erwartungswerte und Kovarianzen berechnen.
(ii) liefert auch Konvergenzraten.
(iii) kann auch dann noch funktionieren, wenn höhere Momente (> 2) wachsen.

Diese Vorteile beziehen sich auf die Methode ganz allgemein und nicht spezifisch auf den lokalen Ansatz.

Beispiele:

Beispiel 3.9. (k-runs, Roos, 1993, siehe [Roo93]) Es seien $X_1, \ldots, X_n$ i.i.d. Zufallsgrößen mit $P(X_i = 1) = p = 1 - P(X_i = 0)$ für $i = 1, \ldots, n$, und $k \in \mathbb{N}$ sei fest gewählt. Wir vereinbaren, dass für alle $1 \le i \le n$ und $l \in \mathbb{N}_0$ der Index $i + n\,l$ mit i identifiziert wird. Es sei weiter $\Gamma = \{1, \ldots, n\}$ und $I_\alpha := \prod_{i=\alpha}^{\alpha+k-1} X_i$ und $W = \sum_{\alpha\in\Gamma} I_\alpha$. Dann ist $\mathbb{E}(I_\alpha) = p^k =: \pi$ für alle $\alpha \in \Gamma$ und somit $\mathbb{E}W = n\,p^k = n\pi$. Somit beschreibt die Zufallsgröße W die Anzahl der *runs* von Länge mindestens k (ein k-run ist also eine Folge von k aufeinanderfolgenden Einsen). Wir wählen in diesem Beispiel

$$\Gamma_\alpha = \{\alpha - (k-1), \ldots, \alpha - 1, \alpha + 1, \ldots, \alpha + k - 1\},$$

denn außerhalb dieser Menge sind die zugehörigen Indikatorvariablen I_β unabhängig von I_α, da die X_i unabhängig gewählt sind. Also sind die Summanden in W'_α unabhängig von I_α und wir können Satz 3.1 anwenden. Es gilt

$$\sum_{\alpha\in\Gamma}\big(\pi_\alpha^2+\pi_\alpha\,\mathbb{E}Z_\alpha\big)=\sum_{\alpha\in\Gamma}\big(\pi^2+\pi^2(2k-2)\big)=n\,\pi^2\,(2k-1).$$

Weiter ist einfach zu sehen, dass

$$\sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha}\mathbb{E}(I_\alpha\,I_\beta)=2n\,\pi\sum_{i=1}^{k-1}p^i$$

gilt. Ist zum Beispiel $\beta=\alpha+1$, so müssen wir $\mathbb{E}(I_\alpha\,I_\beta)=P(I_\alpha=I_\beta=1)$ bestimmen: Es soll also an der Position α und an der folgenden $\alpha+1$ ein k-run starten. Dies geschieht mit Wahrscheinlichkeit $p^k\,p=\pi\,p$. Analog gilt für $\beta=\alpha+2$: $P(I_\alpha=I_\beta=1)=\pi\,p^2$, und sofort. Mit $\lambda:=\mathbb{E}W=n\,\pi$ und Satz 3.1 folgt:

$$\begin{aligned}
d_{TV}\big(\mathcal{L}(W),Po(n\,\pi)\big) &\le \frac{1}{n\,\pi}\big(1-e^{-\lambda}\big)\left((2k-1)n\,\pi^2+2n\,\pi\sum_{i=1}^{k-1}p^i\right)\\
&= \big(1-e^{-\lambda}\big)\left((2k-1)\,p^k+2\sum_{i=1}^{k-1}p^i\right)\\
&= \big(1-e^{-\lambda}\big)\left((2k-1)\,p^k+\frac{2p(1-p^{k-1})}{1-p}\right)\\
&= \mathcal{O}(p).
\end{aligned}$$

Die Ordnung stimmt also mit dem $B(n,p)$-Resultat aus Satz 2.12 überein. Tatsächlich kann dieses Resultat noch verbessert werden: Mittels einer sogenannten Compound-Poisson Approximation erhält man eine Ordnung $\mathcal{O}(p^k)$. Dies stellen wir in einem späteren Kapitel vor.

Beispiel 3.10. (Geburtstagsproblem (Arratia, Goldstein und Gordon, 1990), siehe [AGG90]) Gegeben seien n Personen. Ihr jeweiliger Geburtstag sei unabhängig über d Tage des Jahres verteilt. Wie groß ist die Wahrscheinlichkeit, dass mindestens zwei Personen an einem Tag Geburtstag haben (unter der Annahme, dass Geburtstage gleichmäßig über das Jahr verteilt seien)? Es sei W die Zufallsgröße, die die Anzahl von Personenpaaren mit identischem Geburtstag angibt. Aus einer einführenden Vorlesung zur Stochastik ist bekannt:

$$P(W=0)=\prod_{i=1}^{n-1}\left(1-\frac{i}{d}\right).$$

Wie groß ist aber zum Beispiel $P(W = m)$, oder die Wahrscheinlichkeit für das Ereignis, dass drei Personen an einem Tag Geburtstag haben? Wie verhält sich P für den Fall einer nicht-gleichmäßigen Verteilung der Geburtstage? Exakte Formeln für diese Wahrscheinlichkeiten sind recht kompliziert und teils auch schwer zu bestimmen. Daher sucht man gute Approximationen. Bevor wir Poisson-Approximationen untersuchen, kleiden wir das Geburtstagsproblem noch ein wenig anders ein: Gegeben seien n *Bälle* (Personen) mit den Nummern $\{1, \ldots, n\}$, die gleichmäßig und unabhängig über d *Urnen* (Tage) verteilt seien. Man bestimme die Wahrscheinlichkeit, dass in mindestens einer Urne k oder mehr Bälle enthalten sind für ein festes $k \in \{2, 3, \ldots\}$. Setze $\Gamma := \{\alpha \subset \{1, 2, \ldots, n\} : |\alpha| = k\}$. Weiter sei I_α die Indikatorvariable des Ereignisses, dass die Bälle, die durch α indiziert sind, alle gemeinsam in einer Urne liegen. Mit $W = \sum_{\alpha \in \Gamma} I_\alpha$ und $\pi_\alpha = \mathbb{E} I_\alpha = d^{1-k}$ folgt

$$\lambda = \mathbb{E} W = \binom{n}{k} d^{1-k}.$$

Ist Z Poisson-verteilt zum Parameter λ, so folgt

$$P(Z = 0) = e^{-\lambda} = \exp\left\{ - \binom{n}{k} d^{1-k} \right\}.$$

Wir wollen $P(W = 0) = P(\text{keine Urne hat } k \text{ oder mehr Bälle})$ mittels $P(Z = 0)$ approximieren. Hier ein Zahlenbeispiel: es sei $d = 365$ und $k = 2$. Wir betrachten also das klassische Geburtstagsproblem. Ist $n = 23$ (dies ist der Wert der Trendwende), so ist $\lambda = \binom{23}{2}/365 \approx \log(2)$ (die ersten 4 Stellen stimmen überein!) und

$$P(W = 0) = 0.492 < 0.499998 = \exp(-\lambda) = P(Z = 0)$$

und

$$\begin{aligned} P(W = 0) = \prod_{i=1}^{n-1} \left(1 - \frac{i}{d}\right) \quad &< \quad \exp\left(-\frac{1}{d} \sum_{i=1}^{n-1} i\right) \\ &= \quad \exp(-\lambda) \qquad (3.11) \\ &= \quad P(Z = 0). \end{aligned}$$

Man spricht hier davon, dass die Approximation für $k = 2$ bei gleichmäßiger Verteilung immer *konservativ* ist.

Wir betrachten nun die Poisson-Approximation allgemein: Dazu müssen wir Γ_α finden. Da $\alpha \cap \beta = \emptyset$ impliziert, dass I_α und I_β unabhängig sind, wählen wir

$$\Gamma_\alpha = \{\beta \in \Gamma : \alpha \cap \beta \neq \emptyset\}.$$

Nun wollen wir Satz 3.1 anwenden. Es gilt:

$$\begin{aligned}\sum_{\alpha\in\Gamma}\pi_\alpha^2+\sum_{\alpha\in\Gamma}\pi_\alpha\,\mathbb{E}Z_\alpha &= |\Gamma|\,|\Gamma_\alpha\cup\{\alpha\}|\,\pi_\alpha^2\\ &= \binom{n}{k}\left(\binom{n}{k}-\binom{n-k}{k}\right)d^{2-2k}\\ &= \lambda^2\left(\binom{n}{k}-\binom{n-k}{k}\right)\Big/\binom{n}{k}<\frac{\lambda^2\,k^2}{n}.\end{aligned}$$

Für alle $\alpha\in\Gamma$ und $\beta\in\Gamma_\alpha$ gilt nun im Fall $k=2$:

$$\mathbb{E}(I_\alpha\,I_\beta)=\pi_\alpha\,\pi_\beta$$

(für $k=2$ liegt also paarweise Unabhängigkeit vor). Also folgt in diesem Spezialfall

$$\begin{aligned}\sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha}\mathbb{E}(I_\alpha\,I_\beta) &= \binom{n}{2}\left(\binom{n}{2}-\binom{n-2}{2}-1\right)d^{-2}\\ &= d^{-2}\binom{n}{2}\frac{(4n-8)}{2}<\frac{4\,\lambda^2}{n}.\end{aligned}$$

Also erhalten wir insgesamt (im Fall $k=2$):

$$d_{TV}\big(\mathcal{L}(W),Po(\lambda)\big)\le\frac{1}{d^2}\binom{n}{2}(4n-7)\,\frac{1-e^{-\lambda}}{\lambda}\le\frac{8\lambda(1-e^{-\lambda})}{n},$$

und die gleiche Schranke für $|P(W=0)-e^{-\lambda}|$. Der Fall $n=50$ und $k=3$ wird in einer Übungsaufgabe genauer untersucht.

Abschließend betrachten wir nun noch den allgemeinen Fall, in dem $k\in\mathbb{N}$ beliebig ist. Hier gilt nun

$$\sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha}\mathbb{E}(I_\alpha\,I_\beta)=\sum_{j=1}^{k-1}\binom{n}{k}\binom{k}{j}\binom{n-k}{k-j}d^{1+j-2k};$$

hier beschreibt der j-te Summand den Beitrag von Paaren (α,β) mit $|\alpha\cap\beta|=j$. Hier gilt $\mathbb{E}(I_\alpha\,I_\beta)=d^{1+j-2k}$. Hier ist $|\Gamma|=\binom{n}{k}$ und $\binom{k}{j}\binom{n-k}{k-j}$ die Anzahl der $\beta\in\Gamma_\alpha$ mit $|\alpha\cap\beta|=j$.

Wir wollen dieses Resultat ein wenig diskutieren. Zunächst kann man sehen, dass in dem Fall, in dem d/n groß ist, der dominierende Term in der obigen Summe der Summand zu $j=k-1$ ist. Weiter seien im Folgenden n und d wachsend gegen ∞ derart, dass $\frac{\lambda}{1}$ beschränkt ist weg von 0 und ∞. Wir schreiben dafür

$$\lambda\asymp 1.$$

Dann ist $n^k \asymp d^{k-1}$ und

$$\sum_{\alpha\in\Gamma} \pi_\alpha^2 + \sum_{\alpha\in\Gamma} \pi_\alpha \, \mathbb{E} Z_\alpha \asymp n^{-1}$$

sowie

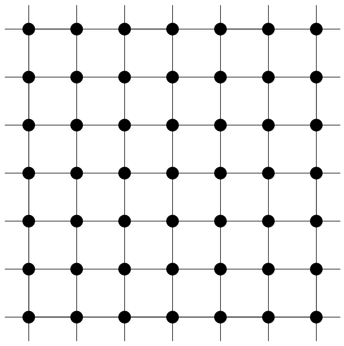
$$\sum_{\alpha\in\Gamma} \sum_{\beta\in\Gamma_\alpha} \mathbb{E}(I_\alpha \, I_\beta) \asymp n^{1+k} \, d^{-k} \asymp \frac{n}{d} \asymp n^{-1/(k-1)}.$$

Insgesamt folgt also für wachsende n und d mit $\lambda \asymp 1$

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq \mathcal{O}\big(n^{-1/(k-1)}\big).$$

Also liefert der Summand $\sum_{\alpha\in\Gamma} \sum_{\beta\in\Gamma_\alpha} \mathbb{E}(I_\alpha \, I_\beta)$ in Größenordnung den Hauptbeitrag.

Beispiel 3.12. Gegeben sei ein Gitter auf dem Torus mit n Knoten und $N = 2n$ Kanten. Wir vereinbaren die folgende Regel: *Eine Kante kann mit Wahrscheinlichkeit $1 - p = q$ entfernt werden, unabhängig von allen anderen Kanten.*



Dies ist ein Modell für einen Multiprozessor, bei dem die Knoten für Prozessoren und die Kanten für Verbindungen (Beziehungen zwischen Prozessoren) stehen. Aus Erfahrung wisse man, dass das System zusammenbricht, wenn m Knoten isoliert sind. Allgemein sagt man nun, dass die Zuverlässigkeit des Systems die Wahrscheinlichkeit ist, dass das System (der Multiprozessor) arbeitet. Dies sei die Wahrscheinlichkeit, dass weniger als m isolierte Knoten auftreten. Man interessiert sich nun für den Wert dieser Wahrscheinlichkeit, den man die *Zuverlässigkeit* oder die *Zuverlässigkeits-Wahrscheinlichkeit* nennt (reliability-probability). Der genaue Wert ist bei komplexen Systemen schwer zu bestimmen. Daher möchte man eine gute Approximation finden. Wir untersuchen hier die Poisson-Approximation.

Es sei $\Gamma = \{1, \ldots, n\}$ und P(eine Kante zwischen α und β wird entfernt) $= q$ für alle α und $\beta \in \Gamma$, die Nachbar-Gitterpunkte sind. Setze nun

$$I_\alpha := I(\text{Knoten } \alpha \text{ ist isoliert})$$

für $\alpha \in \Gamma$ und $W = \sum_{\alpha \in \Gamma} I_\alpha$. Dann ist

$$P(\text{Knoten } \alpha \text{ ist isoliert}) = q^4 = \pi_\alpha$$

und $\mathbb{E}W = n\, q^4 = \lambda$. Wie groß ist $P(W \leq m-1)$? Setze

$$\Gamma_\alpha = \{\beta \neq \alpha : \beta \text{ und } \alpha \text{ sind Nachbar-Gitterpunkte}\}.$$

Dann ist $|\Gamma_\alpha| = 4$ und

$$\sum_{\alpha \in \Gamma} \pi_\alpha^2 = n\, q^8$$

und

$$\sum_{\alpha \in \Gamma} \pi_\alpha\, \mathbb{E}Z_\alpha = \sum_{\alpha \in \Gamma} \sum_{\beta \in \Gamma_\alpha} \pi_\alpha\, \pi_\beta = 4\, n\, q^8$$

sowie

$$\sum_{\alpha \in \Gamma} \sum_{\beta \in \Gamma_\alpha} \mathbb{E}(I_\alpha\, I_\beta) = 4\, n\, q^7.$$

Mit Hilfe von Satz 3.1 finden wir

$$d_{TV}\big(\mathcal{L}(W), Po(n\, q^4)\big) \leq \big(5\, n\, q^8 + 4\, n\, q^7\big) \left(\frac{1-e^{-\lambda}}{n\, q^4}\right) \leq 5\, q^4 + 4\, q^3 = \mathcal{O}(q^3).$$

Wenn also zum Beispiel $\lim_{n\to\infty} n\, q(n)^4 = \mu_\infty > 1$ gilt, so folgt

$$d_{TV}\big(\mathcal{L}(W), Po(n\, q(n)^4)\big) \leq \mathcal{O}(1/n^{3/4}).$$

(Hinweis: wieder kann man mittels einer Compound-Poisson Approximation dieses Resultat verbessern.) Die Zuverlässigkeit des Systems kann mit Hilfe einer Poisson-verteilten Zufallsgröße Z zum Parameter $\lambda = n\, q^4$ mittels

$$P(W \leq m-1) \approx \sum_{i=0}^{m-1} P(Z = i) = \sum_{i=0}^{m-1} e^{-n\, q^4} \frac{(n\, q^4)^i}{i!}$$

approximiert werden.

Wir werden den lokalen Ansatz noch in weiteren Beispielen anwenden.

4. Der Kopplungs-Ansatz (Theorie und Beispiele)

In diesem Kapitel entwickeln wir einen Ansatz der Steinschen Methode, bei dem *Kopplungen* im Zentrum stehen. Kopplungen werden seit Steins Monographie [Ste86] aus dem Jahre 1986 intensiv untersucht. Eine spezielle Kopplung war uns bereits im Kapitel 1 begegnet.

Wir leiten zunächst mittels allgemeiner Kopplungen Schranken für die Totalvariation im Fall der Poisson-Approximation her und betrachten anschließend einige Beispiele. Die sogenannten *monotonen Kopplungen* stehen im Vordergrund. Sie werden in einer Arbeit von Barbour und Holst [BH89] sowie in der Monographie [BHJ92] hergeleitet.

Bei den Beispielen betrachten wir Besetzungszahlen in Urnenmodellen und zufällige Permutationen. Darunter fällt ein sogenanntes Matching- und das Ménage-Problem.

Die Theorie:

Im Fall unabhängiger Indikatoren $(I_\alpha)_{\alpha\in\Gamma}$ war Gleichung (2.6) im zweiten Kapitel von entscheidendem Nutzen:

$$\mathbb{E}\big(I_\alpha\, g_A(W)\big) = \pi_\alpha\, \mathbb{E}\big(g_A(W_\alpha + 1)\big),$$

wobei wir die Notation aus Kapitel 3 verwenden mit

$$W_\alpha = \sum_{\beta\in\Gamma,\beta\neq\alpha} I_\beta.$$

Wir wollen diese Gleichung nun ein wenig umschreiben. Dazu führen wir (relativ inoffiziell) den Begriff des *bedingten Erwartungswertes* für diskrete Zufallsgrößen ein. Es sei X eine $\mathbb{N}_0$-wertige Zufallsgröße. Dann bezeichnet

$$\mathbb{E}(X|I_\alpha = 1) := \sum_{k\geq 0} k\, P(X = k|I_\alpha = 1)$$

den bedingten Erwartungswert von X gegeben das Ereignis $\{I_\alpha = 1\}$. Hier ist der bedingte Erwartungswert also eine Zahl (man kläre für sich, dass dieser Begriff in den allgemeinen Rahmen einer Einführung des bedingten Erwartungswertes als eine Zufallsvariable paßt). Wir schreiben nun für die linke Seite unserer Ausgangsgleichung

$$\mathbb{E}\big(I_\alpha g_A(W)\big) = \pi_\alpha \mathbb{E}\big(g_A(W)|I_\alpha = 1\big),$$

denn

$$\pi_\alpha \mathbb{E}\big(g_A(W)|I_\alpha = 1\big) = \sum_{k\geq 0} g_A(k)\, P(W = k, I_\alpha = 1) = \mathbb{E}\big(I_\alpha g_A(W)\big).$$

Somit gilt für den Steinschen Ansatz

$$P(W \in A) - Po(\lambda)\{A\} = \sum_{\alpha\in\Gamma} \pi_\alpha \Big(\mathbb{E} g_A(W+1) - \mathbb{E}\big(g_A(W)|I_\alpha = 1\big)\Big).$$

Aus dieser neuen Schreibweise ziehen wir nun den folgenden Nutzen: Wir nehmen an, dass für jedes $\alpha \in \Gamma$ Zufallsgrößen U_α und V_α existieren, definiert auf einem gemeinsamen Wahrscheinlichkeitsraum, mit

$$\mathcal{L}(U_\alpha) = \mathcal{L}(W)$$

und

$$\mathcal{L}(V_\alpha + 1) = \mathcal{L}(W|I_\alpha = 1).$$

Wir nennen fortan das Paar (U_α, V_α) *Kopplung*. Es gilt dann nach Definition von V_α

$$\mathbb{E}\big(g_A(W)|I_\alpha = 1\big) = \mathbb{E}(g_A(V_\alpha + 1)).$$

Mit Hilfe dieser einfachen Beobachtungen zeigen wir nun den folgenden, für das weitere Vorgehen sehr wichtigen Satz (siehe [BH89]):

Satz 4.1. *(Kopplungsansatz, Barbour, Holst, 1989) Wenn für jedes $\alpha \in \Gamma$ die Kopplungen (U_α, V_α) mit den oben beschriebenen Eigenschaften existieren, folgt:*

$$d_{TV}\big(\mathcal{L}(W), Po(\mathbb{E}W)\big) \le \Delta g \sum_{\alpha\in\Gamma} \pi_\alpha \mathbb{E}|U_\alpha - V_\alpha|.$$

Beweis: Nach den vorbereitenden Bemerkungen wissen wir bereits, dass gilt:

$$\begin{aligned}\mathbb{E}\big(\lambda g_A(W+1) - W\, g_A(W)\big) &= \sum_{\alpha\in\Gamma} \pi_\alpha \Big(\mathbb{E}\big(g_A(W+1)\big) - \mathbb{E}\big(g_A(V_\alpha + 1)\big)\Big) \\ &= \sum_{\alpha\in\Gamma} \pi_\alpha \Big(\mathbb{E}\big(g_A(U_\alpha + 1) - g_A(V_\alpha + 1)\big)\Big),\end{aligned}$$

also folgt unmittelbar

$$\big|\mathbb{E}\big(\lambda g_A(W+1) - W\, g_A(W)\big)\big| \le \Delta g \sum_{\alpha\in\Gamma} \pi_\alpha\, \mathbb{E}|U_\alpha - V_\alpha|.$$

□

Bemerkungen 4.2. (i) Das Ziel in den Anwendungen wird es sein, die Kopplungen (U_α, V_α) so zu konstruieren, dass $\mathbb{E}|U_\alpha - V_\alpha|$ klein wird. Diese Konstruktionsfreiheit wird sehr effektiv genutzt werden können.

(ii) Im Fall von unabhängigen Summanden I_α setzt man $U_\alpha = W$ und $V_\alpha = W_\alpha = \sum_{\beta\neq\alpha} I_\beta$ für jedes $\alpha \in \Gamma$. Dann erhält man unmittelbar Satz 2.11.

(iii) Wir werden sehen, dass der Kopplungsansatz häufig dann benutzt wird, wenn der Bereich der lokalen Abhängigkeit Γ_α, definiert wie in Kapitel 3, nicht gut beschrieben oder gar nicht angegeben werden kann. Es handelt sich bei den Anwendungen um Situationen, bei denen die Summanden I_α alle untereinander ein symmetrisches, globales Verhalten in ihrer Abhängigkeit aufweisen. Schwierig bei der Anwendung von Satz 4.1 ist, dass das Auffinden einer geeigneten Kopplung manchmal sehr trickreich sein kann. Andererseits taucht bei der Abschätzung immer nur $\Delta g \leq \min\bigl(1, \frac{1}{\lambda}\bigr)$ auf.

Wir diskutieren nun die folgende Situation. Angenommen, man kann (U_α, V_α) für jedes $\alpha \in \Gamma$ so konstruieren, dass $U_\alpha \geq V_\alpha$ (fast sicher) für alle $\alpha \in \Gamma$ gilt. Dann folgt für die Schranke in Satz 4.1:

$$\sum_{\alpha\in\Gamma} \pi_\alpha \mathbb{E}|U_\alpha - V_\alpha| = \sum_{\alpha\in\Gamma} \pi_\alpha \mathbb{E}(U_\alpha - V_\alpha)$$

und

$$\sum_{\alpha\in\Gamma} \pi_\alpha \, \mathbb{E}U_\alpha = \sum_{\alpha\in\Gamma} \pi_\alpha \, \mathbb{E}W = \lambda \, \mathbb{E}W = (\mathbb{E}W)^2$$

sowie

$$\begin{aligned} \sum_{\alpha\in\Gamma} \pi_\alpha \, \mathbb{E}V_\alpha &= \sum_{\alpha\in\Gamma} \pi_\alpha \, \mathbb{E}\bigl(W - I_\alpha | I_\alpha = 1\bigr) = \sum_{\alpha\in\Gamma} \mathbb{E}\bigl(I_\alpha(W - I_\alpha)\bigr) \\ &= \mathbb{E}\bigl(W \sum_{\alpha\in\Gamma} I_\alpha\bigr) - \mathbb{E}\bigl(\sum_{\alpha\in\Gamma} I_\alpha\bigr) = \mathbb{E}(W^2) - \mathbb{E}W. \end{aligned}$$

Also folgt

$$\sum_{\alpha\in\Gamma} \pi_\alpha \mathbb{E}|U_\alpha - V_\alpha| = \lambda - V(W).$$

Dies führt zu dem folgenden Resultat, welches in [BH89] bewiesen wurde. Man spricht von *monotonen Kopplungen.*

Korollar 4.3. *(Barbour, Holst 1989) Die Kopplungen (U_α, V_α) aus Satz 4.1 seien so gewählt, dass $U_\alpha \geq V_\alpha$ (fast sicher) für alle $\alpha \in \Gamma$ gilt. Dann folgt:*

$$d_{TV}\bigl(\mathcal{L}(W), Po(\lambda)\bigr) \leq \Delta g \, (\lambda - V(W)) \leq \min(1, \lambda)\left(1 - \frac{V(W)}{\lambda}\right).$$

Bemerkung 4.4. Ist $W \sim Po(\lambda)$, so ist $\mathbb{E}W = V(W) = \lambda$. Somit beschreibt $\frac{V(W)}{\lambda}$, wie dicht man bei einer Poisson-Verteilung ist. Die Berechnung der Schranke reduziert sich hier darauf, $\mathbb{E}W$ und $V(W)$ zu bestimmen.

Bisher haben wir noch keine konkrete Konstruktion einer Kopplung betrachtet. Wir zeigen im letzten Teil der theoretischen Ausführungen, dass man die Kopplungen (U_α, V_α) mit Hilfe von *Kopplungen an den Indikatoren* I_α konstruieren kann. Für jedes $\alpha \in \Gamma$ seien Zufallsgrößen $(J_{\beta\,\alpha}, \beta \in \Gamma)$ gegeben, die auf demselben Wahrscheinlichkeitsraum wie die $(I_\beta, \beta \in \Gamma)$ definiert seien, und die die folgende Eigenschaft haben: für jedes $\beta \in \Gamma$ sei

$$\mathcal{L}(J_{\beta\,\alpha}) = \mathcal{L}(I_\beta | I_\alpha = 1). \tag{4.5}$$

Wir setzen $J_{\alpha\,\alpha} \equiv 1$. Es sei nun $U_\alpha := W$ und

$$V_\alpha := \sum_{\beta \neq \alpha} J_{\beta\,\alpha}$$

für jedes $\alpha \in \Gamma$. Dann ist $\mathcal{L}(U_\alpha) = \mathcal{L}(W)$ und

$$\mathcal{L}(V_\alpha + 1) = \mathcal{L}(W | I_\alpha = 1)$$

für jedes $\alpha \in \Gamma$. Ist nun weiter diese Kopplung so konstruierbar, dass

$$J_{\beta\,\alpha} \leq I_\beta \tag{4.6}$$

für alle $\beta \in \Gamma \setminus \{\alpha\}$ gilt, so ist $V_\alpha \leq U_\alpha$ für alle $\alpha \in \Gamma$ erreicht. Somit ist Korollar 4.3 anwendbar. Man sagt, dass die $(I_\beta, \beta \in \Gamma)$ *negativ zusammenhängen*, wenn (4.6) erfüllt ist (bei $\geq$ in (4.6) spricht man von positiv zusammenhängend). Wann kann man Kopplungen $J_{\beta\,\alpha}$ mit (4.5) und (4.6) konstruieren? Wir machen die folgende Beobachtung: gilt (4.5), so folgt für alle $\alpha, \beta \in \Gamma$:

$$\operatorname{cov}(I_\alpha, I_\beta) \leq 0.$$

Notwendig für die Konstruktion einer Kopplung wie oben beschrieben, die (4.5) erfüllt, ist also die Eigenschaft, dass die Indikatoren $(I_\alpha)_{\alpha \in \Gamma}$ *nicht-positiv korreliert* sind. Wir wollen den Beweis dieses notwendigen Kriteriums noch führen: Nach Definition ist $\operatorname{cov}(I_\alpha, I_\beta) = \mathbb{E}(I_\alpha\, I_\beta) - \pi_\alpha\, \pi_\beta$. Nun gilt

$$\mathbb{E}(I_\alpha I_\beta) = P(I_\alpha = 1, I_\beta = 1) = \pi_\alpha\, P(I_\beta = 1 | I_\alpha = 1) = \pi_\alpha P(J_{\beta\,\alpha} = 1) = \pi_\alpha\, \mathbb{E}(J_{\beta\,\alpha})$$

und

$$\mathbb{E}(J_{\beta\,\alpha}) = P(J_{\beta\alpha} = 1) = \pi_\beta\, P(J_{\beta\alpha} = 1 | I_\beta = 1).$$

Also ist

$$\mathbb{E}(I_\alpha\, I_\beta) - \pi_\alpha\, \pi_\beta = \pi_\alpha\, \pi_\beta \big(P(J_{\beta\,\alpha} = 1 | I_\beta = 1) - 1\big) \leq 0.$$

Bemerkung 4.7. Im Fall unabhängiger $(I_\alpha)_{\alpha\in\Gamma}$ folgt mittels Korollar 4.3 erneut Satz 2.11: Wähle dazu $J_{\beta\,\alpha} = I_\beta$ für alle $\beta \neq \alpha$. Dann sind (4.5) und (4.6) erfüllt. Weiter ist $V(W) = \sum_{\alpha\in\Gamma} \pi_\alpha\,(1-\pi_\alpha)$ und $\mathbb{E}W = \sum_{\alpha\in\Gamma}\pi_\alpha$, also

$$\lambda - V(W) = \sum_{\alpha\in\Gamma} \pi_\alpha^2.$$

Beispiele:

Wir betrachten zunächst Beispiele, bei denen Kopplungen der obigen Art Anwendung finden:

Beispiel 4.8. (Barbour, Holst, 1989, [BH89]) Gegeben seien $r \in \mathbb{N}$ Boxen und man werfe n Bälle unabhängig hinein, wobei p_k die Wahrscheinlichkeit sei, dass Box Nr.k getroffen wird. Es sei $I_j, j = 1,\dots,r$, der Indikator für das Ereignis, dass Box Nr.j nach den n Würfen *leer* geblieben ist. Setze $W := I_1+\cdots+I_r$. Also gibt W die Anzahl der leeren Boxen an. Es ist $\pi_k = P(I_k = 1) = (1-p_k)^n$ und $\lambda = \mathbb{E}W = \sum_{k=1}^r \pi_k$. Weiter gilt

$$\mathbb{E}(I_i\,I_k) = P(I_i = 1 = I_k) = (1-p_i-p_k)^n$$

und

$$V(W) = \sum_{k=1}^r \pi_k\,(1-\pi_k) + \sum_{i,k:i\neq k} \big(\mathbb{E}(I_i\,I_k) - \pi_i\,\pi_k\big).$$

Es gilt in diesem Modell:

Satz 4.9.

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq (1\wedge\lambda)\left(1 - \frac{V(W)}{\lambda}\right).$$

Hierbei bezeichnet $x \wedge y$ das Minimum von x und y.

Beweis: Wir müssen eine Kopplung finden, die (4.5) und (4.6) erfüllt. Dazu werfe jeden Ball, der in Box Nr. k gefallen ist, unabhängig in eine der anderen Boxen mit Wahrscheinlichkeit

$$\frac{p_i}{(1-p_k)}$$

für Box Nr. i. Es sei dann $J_{i\,k} = 1$, wenn Box Nr. i leer ist, und $J_{i\,k} = 0$ sonst. Es folgt nach Konstruktion unmittelbar $J_{i\,k} \leq I_i$ für alle $i \neq k$. Weiter ist

$$\mathcal{L}(J_{1\,k},\dots,J_{r\,k}) = \mathcal{L}(I_1,\dots,I_r | I_k = 1).$$

Dies betrachten wir ausführlich:

$$P(J_{i\,k} = 1) = \left(1 - \frac{p_i}{1-p_k}\right)^n = \left(\frac{1-p_i-p_k}{1-p_k}\right)^n = P(I_i = 1 | I_k = 1).$$

Der Satz ist somit bewiesen. $\square$

Wir vergleichen unser Resultat mit einem Resultat von Sevast'yanov (1972), welches mittels der *Methode der faktoriellen Momente* gewonnen wurde. Es sei zu $r \in \mathbb{N}$ nun $n = n_r$, $p_k = p_{k\,r}$ und $W = W_r$, dann gilt:

Wenn $\max_{1 \le k \le r} \pi_k \to 0$ für $r \to \infty$ und

$$\lim_{r\to\infty} \mathbb{E}W = \lambda_\infty < \infty,$$

so folgt

$$\lim_{r\to\infty} \mathbb{E}_j(W) = \lambda_\infty^j,$$

wobei $\mathbb{E}_j(W)$ das j-te faktorielle Moment von W bezeichnet. Die Methode der faktoriellen Momente liefert dann:

$$\lim_{r\to\infty} d_{TV}\big(\mathcal{L}(W), Po(\lambda_\infty)\big) = 0.$$

Was liefert Satz 4.9 mehr?

(i) Wir benötigen nur $\mathbb{E}_2(W) \to \lambda_\infty^2$ und bekommen Konvergenzraten.
(ii) $\mathbb{E}W \to \infty$ ist auch zugelassen, sobald nur $\lim_{r\to\infty} \frac{V(W)}{\mathbb{E}W} = 1$ gilt. Bemerke, dass $V(W) < \mathbb{E}W$ für alle $n \ge 1$ gilt.

Im Spezialfall $p_k = \frac{1}{r}$ für $k = 1, \ldots, r$ hat bereits de Moivre eine exakte Form der Verteilung $\mathcal{L}(W)$ gegeben. Von Mises hat in diesem Fall einen Poissonschen Grenzwertsatz bewiesen, wenn $\mathbb{E}W \to \lambda_\infty$ gilt. In einer Arbeit von Vatutin und Mikhailov (1982) wurde gezeigt, dass W die gleiche Verteilung wie eine Summe von unabhängigen Bernoulli Zufallsgrößen hat. Wir verfolgen diesen Spezialfall nicht weiter (siehe Literaturangaben in der Arbeit von Barbour und Holst, 1989).

Gegeben sei eine Urne mit N Bällen mit r verschiedenen Farben mit Wahrscheinlichkeit $p_1, \ldots, p_r$. Ziehe n Bälle mit Zurücklegen. W bezeichne die Anzahl der Farben, die unter den n Bällen nicht vorkommen. Dann ist dieses Modell offenbar nur eine andere Interpretation des gerade Betrachteten. Nun modifizieren wir das Beispiel: wir legen nicht zurück. Setze $I_k = 1$, wenn kein Ball der Farbe k dabei ist, $I_k = 0$ sonst. Dann ist

$$\pi_k = P(I_k = 1) = \frac{\binom{N(1-p_k)}{n}}{\binom{N}{n}}$$

und

$$\mathbb{E}(I_i\, I_k) = P(I_i = I_k = 1) = \frac{\binom{N(1-p_k-p_i)}{n}}{\binom{N}{n}}.$$

Es gilt nun:

Satz 4.10.
$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq (1 \wedge \lambda)\left(1 - \frac{V(W)}{\lambda}\right).$$

Beweis: Finde erneut eine geeignete Kopplung: Ziehe die n Bälle ohne Zurücklegen aus der Urne. Nehme dann alle Bälle der Farbe k aus der Urne. Ersetze weiter alle Bälle der Farbe k in der Ziehung durch Bälle aus der "neuen" Urne (ohne Zurücklegen). Nun setze $J_{i\,k} = 1$, wenn in der neuen Verteilung der Bälle kein Ball der Farbe i ist und $J_{i\,k} = 0$ sonst. Es folgt: $J_{i\,k} \leq I_i$ für alle $k \neq i$ und

$$\mathcal{L}(J_{1\,k}, \ldots, J_{r\,k}) = \mathcal{L}(I_1, \ldots, I_r | I_k = 1),$$

denn:
$$P(I_i = 1 | I_k = 1) = \frac{\binom{N(1-p_i-p_k)}{n}}{\binom{N(1-p_k)}{n}} = P(J_{i\,k} = 1).$$
Der Satz ist somit gezeigt. □

Beispiel 4.11. (zufällige Permutationen, Barbour, Holst 1989, [BH89]) Gegeben sei eine $r \times r$ Matrix $(c_{ij})_{i,j=1}^r$ mit Elementen in $\{0,1\}$. Weiter sei $I_i := c_{i\sigma(i)} =: c(i, \sigma(i))$, wobei σ eine zufällige Permutation ist: versehe die Menge aller Permutationen von $\{1, \ldots, r\}$ mit der Gleichverteilung. Sei $W = \sum_{i=1}^r I_i$. Es gilt dann

$$\pi_i = \frac{1}{r}\sum_{j=1}^r c_{ij} = \mathbb{E}(I_i) = P(I_i = 1).$$

Weiter sei
$$\varrho_j := \frac{1}{r}\sum_{i=1}^r c_{ij}$$

und
$$\lambda = \mathbb{E}W = \frac{1}{r}\sum_{i,j=1}^r c_{ij}.$$

Dann gilt der folgende Satz:

Satz 4.12. *Für das obige Modell zufälliger Permutationen gilt*
$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq \left(1 \wedge \frac{1}{\lambda}\right)\frac{3}{2}\left(\sum_{i=1}^r \pi_i^2 + \sum_{j=1}^r \varrho_j^2 - \frac{\lambda}{3r}\right).$$

Bemerkung 4.13. Die Summe $\sum_{i=1}^r \pi_i^2$ tritt in dieser Form im Fall der unabhängigen Indikatorvariablen auf. Die Summe $\sum_{j=1}^r \varrho_j^2$ kann die gleiche Interpretation erhalten, denn es gilt $W = \sum_{j=1}^r I'_j$ mit $I'_j := c(\sigma^{-1}(j), j)$.

Der Beweis dieses Satzes ist etwas aufwendiger. Wir diskutieren zunächst Anwendungen, also Beispiele, in denen zufällige Permutationen eine Rolle spielen.

Matchings: Es seien A und B zwei Kartenspiele mit r Karten: A habe a_1 Karten vom Typ 1, a_2 Karten vom Typ 2, ..., a_n Karten vom Typ n. B habe b_1 Karten vom Typ 1, b_2 Karten vom Typ 2, ..., b_n Karten vom Typ n. Es sei

$$r = \sum_{j=1}^{n} a_j = \sum_{j=1}^{n} b_j.$$

Jede Karte aus A wird nun an eine Karte aus B gelegt, wobei alle $r!$ Möglichkeiten, dies zu tun, je Wahrscheinlichkeit $1/r!$ haben. Ein Paar liefert ein *Match*, wenn beide Karten vom gleichen Typ sind. W zähle die Matches. Wir modellieren dies genauer: Es sei $A_0 = B_0 = 0$ und

$$A_s := \sum_{j=1}^{s} a_j, \qquad B_s := \sum_{j=1}^{s} b_j, \quad 1 \le s \le n.$$

Setze nun

$$\alpha(i) := s, \quad \text{wenn} \quad A_{s-1} < i \le A_s$$

und analog

$$\beta(j) := s, \quad \text{wenn} \quad B_{s-1} < j \le B_s.$$

Nun definiere die $r \times r$-Matrix durch

$$c_{ij} := \begin{cases} 1, & \text{falls } \alpha(i) = \beta(j), \\ 0, & \text{sonst.} \end{cases}$$

Dann liefern obige $I_i := c(i, \sigma(i))$ Indikatoren für Matches und $W = \sum_{i=1}^{r} I_i$ zählt die Matches. Es gilt hier

$$\lambda = \mathbb{E}W = \sum_{j=1}^{n} \frac{a_j\, b_j}{r}$$

und

$$\sum_{k=1}^{r} \pi_k^2 = \sum_{j=1}^{n} \frac{a_j\, b_j^2}{r^2}$$

sowie

$$\sum_{j=1}^{r} \varrho_j^2 = \sum_{j=1}^{n} \frac{a_j^2\, b_j}{r^2}.$$

Wir wenden nun Satz 4.12 an und erhalten:

Korollar 4.14. *(Matching-Problem)*

$$d_{TV}\big(\mathcal{L}(W), Po(\mathbb{E}W)\big) \le \left(1 \wedge \frac{1}{\lambda}\right) \frac{3}{2} \sum_{j=1}^{n} \frac{a_j\, b_j (a_j + b_j)}{r^2}.$$

Im Spezialfall $r = nc$; $a_j = b_j = c$ für alle j folgt: $\mathbb{E}W = c$ und somit

$$d_{TV}\big(\mathcal{L}(W), Po(c)\big) \le \frac{3c}{r}.$$

Wir betrachten ein weiteres Beispiel, in dem zufällige Permutationen auftreten:

Das Ménage Problem:

Es sei $(\sigma_1, \dots, \sigma_r)$ eine zufällige Permutation σ von $\{1, \dots, r\}$ und sei $1 \le c \le r$ fest gegeben. Wir setzen $I_k = 1$, wenn $k \in \{\sigma_k, \sigma_{k+1}, \dots, \sigma_{k+c-1}\}$ und $I_k = 0$ sonst. Hierbei setzen wir $\sigma_{r+j} = \sigma_j$. Wieder sei $W = I_1 + \dots + I_r$. Es ist $P(I_k = 1) = \frac{c}{r}$ und $\lambda = \mathbb{E}W = c$.

Im Fall $c = 1$ handelt es sich um das "klassische" Matching Problem.

Im Fall $c = 2$ handelt es sich um das klassische Ménage-Problem. Hier hat $P(W = 0)$ die folgende Interpretation: setze r Mann-Frau-Paare an einen Runden Tisch, so dass Männer und Frauen alternierend sitzen, und frage nach der Wahrscheinlichkeit, dass niemand neben seiner Partnerin oder seinem Partner sitzt.

Es gilt nun:

Korollar 4.15. *(das allgemeine Ménage-Problem)*

$$d_{TV}\big(\mathcal{L}(W), Po(c)\big) \le \frac{3\,c}{r}.$$

Also gilt für festes c:

$$\lim_{r \to \infty} d_{TV}\big(\mathcal{L}(W), Po(c)\big) = 0.$$

Beweis: Wir setzen

$$c_{ij} := \begin{cases} 1, & \text{falls } i \le j \le i + c - 1 \text{ oder } \quad 1 \le j \le i + c - 1 - r, \\ 0, & \text{sonst.} \end{cases}$$

Dann ist $\mathcal{L}(W) = \mathcal{L}\big(\sum_{i=1}^{r} c(i, \sigma(i))\big)$ und $\pi_k = \varrho_j = \frac{c}{r}$ und alles folgt aus den obigen Ausführungen. $\square$

Hier nun der Beweis von Satz 4.12: Wir benötigen eine sogenannte *detaillierte Kopplung*. Detailliert soll heißen, dass man nicht nur auf Ereignisse der Form $\{I_\alpha = 1\}$ bedingen wird. Zu $\alpha \in \Gamma$ sei X_α eine Zufallsgröße und $U_{\alpha,x}$ und $V_{\alpha,x}$ seien auf einem gemeinsamen Wahrscheinlichkeitsraum definierte Zufallsgrößen, so dass

$$\begin{aligned} \mathcal{L}(U_{\alpha,x}) &= \mathcal{L}(W) \text{ und} \\ \mathcal{L}(V_{\alpha,x}+1) &= \mathcal{L}(W|I_\alpha = 1, X_\alpha = x). \end{aligned}$$

Dann gilt analog zu Satz 4.1 (siehe [BH89]):

Satz 4.16. *Mit den obigen Notationen gilt:*

$$\begin{aligned} d_{TV}\big(\mathcal{L}(W), Po(\mathbb{E}W)\big) &\leq \frac{1}{\lambda}\big(1-e^{-\lambda}\big)\sum_{\alpha\in\Gamma} \pi_\alpha\, \mathbb{E}\big(\theta_\alpha(X_\alpha)|I_\alpha = 1\big) \\ &= \frac{1}{\lambda}\big(1-e^{-\lambda}\big)\sum_{\alpha\in\Gamma} \mathbb{E}\big(I_\alpha\, \theta_\alpha(X_\alpha)\big) \end{aligned}$$

mit

$$\theta_\alpha(x) := \mathbb{E}|U_{\alpha,x} - V_{\alpha,x}|.$$

Der Beweis ist eine einfache Übung entlang des Beweises von Satz 4.1.

Wir wenden diesen Satz an, um Satz 4.12 zu beweisen:

Die Zufallsgröße W wird nun an die folgende Zufallsgröße $V_{k\,j}$ gekoppelt: zu k suche ein j mit $c_{kj} = 1$. Dann modifiziere eine zufällig gewählte Permutation σ zu einer Permutation σ', so dass $\sigma'(k) = j$ gilt. Dies erreicht man durch das Anfügen einer Transposition τ, so dass $\sigma'(k) := \tau \circ \sigma(k) = j$, also $\sigma'(\sigma^{-1}(j)) := \sigma(k)$, und $\sigma'(i) := \sigma(i)$ für alle anderen i gilt. Definiere nun

$$V_{k\,j} := \sum_{\substack{i=1\\ i\neq k}}^{r} c(i, \sigma'(i)).$$

Dann gilt

$$|W - V_{k\,j}| = |c(k,\sigma(k)) + c(\sigma^{-1}(j), j) - c(\sigma^{-1}(j), \sigma(k))|,$$

und somit folgt

$$\mathcal{L}(1 + V_{k\,j}) = \mathcal{L}(W|I_k = 1, \sigma(k) = j).$$

Nach Satz 4.16 bleibt $\mathbb{E}|W - V_{k\,j}|$ abzuschätzen. Es gilt

$$\begin{aligned} &\mathbb{E}|c(k,\sigma(k)) + c(\sigma^{-1}(j), j) - c(\sigma^{-1}(j), \sigma(k))| \\ = {}& \frac{1}{r}c_{kj} + \frac{1}{r(r-1)}\sum_{l\neq j}\sum_{m\neq k}|c_{kl} + c_{mj} - c_{ml}|. \end{aligned} \tag{4.17}$$

Wir bezeichnen nun mit $c_{i\cdot} := \sum_{j=1}^r c_{ij}$, $c_{\cdot j}$ analog und $c_{\cdot\cdot} = \sum_{i=1}^r \sum_{j=1}^r c_{ij}$. Dann können wir wie folgt abschätzen:

$$\begin{aligned} &\leq \frac{1}{r}c_{kj} + \frac{1}{r}(c_{k\cdot} - c_{kj}) + \frac{1}{r}(c_{\cdot j} - c_{kj}) \\ &\quad + \frac{1}{r(r-1)}(c_{\cdot\cdot} - c_{\cdot j} - c_{k\cdot} + c_{kj}). \end{aligned}$$

Wir erinnern daran, dass $\lambda = \frac{c_{\cdot\cdot}}{r}$, $\pi_k = \frac{c_{k\cdot}}{r}$ und $\varrho_j = \frac{c_{\cdot j}}{r}$ gilt. Somit folgt, dass (4.9.7) identisch gleich

$$\begin{aligned} &= \frac{1}{r}c_{kj} + \pi_k - \frac{c_{kj}}{r} + \varrho_j - \frac{c_{kj}}{r} + \frac{1}{r-1}\big(\lambda - \varrho_j - \pi_k + \frac{c_{kj}}{r}\big) \\ &= \frac{1}{r-1}\Big(\lambda + (r-2)\big(\pi_k + \varrho_j - \frac{c_{kj}}{r}\big)\Big) \end{aligned}$$

ist. Mit Satz 4.16 folgt nun

$$\begin{aligned} d_{TV}\big(\mathcal{L}(W), Po(\mathbb{E}W)\big) &\leq \frac{1-e^{-\lambda}}{\lambda}\sum_{k=1}^r \mathbb{E}\big(I_k\,\theta_k(\sigma(k))\big) \\ &\leq \frac{1-e^{-\lambda}}{\lambda}\sum_{k=1}^r\sum_{j=1}^r \frac{c_{kj}}{r}\frac{1}{r-1}\Big(\lambda + (r-2)\big(\pi_k + \varrho_j - \frac{c_{kj}}{r}\big)\Big) \\ &= \frac{1-e^{-\lambda}}{\lambda(r-1)}\left(\lambda^2 + (r-2)\left(\sum_{k=1}^r \pi_k^2 + \sum_{j=1}^r \varrho_j^2 - \frac{\lambda}{r}\right)\right). \end{aligned}$$

Mit Hilfe der Cauchy-Schwarz Ungleichung folgt

$$r^2\,\lambda^2 = \left(\sum_{j=1}^r c_{\cdot j}\right)^2 \leq r\,\sum_{j=1}^r c_{\cdot j}^2,$$

also

$$\lambda^2 \leq r\,\sum_{j=1}^r \varrho_j^2,$$

und analog

$$\lambda^2 \leq r\,\sum_{k=1}^r \pi_k^2.$$

Zusammenfassend ergibt dies

$$2\,\lambda^2 \leq r\left(\sum_{k=1}^r \pi_k^2 + \sum_{j=1}^r \varrho_j^2\right).$$

Somit ergibt sich abschließend die folgende Abschätzung:

$$\begin{aligned}
d_{TV}\big(\mathcal{L}(W), Po(\mathbb{E}W)\big) &\leq \Delta g \left(\left(\frac{r}{2(r-1)} + \frac{r-2}{r-1} \right) \left(\sum_{k=1}^{r} \pi_k^2 + \sum_{j=1}^{r} \varrho_j^2 \right) - \frac{r-2}{r-1}\frac{\lambda}{r} \right) \\
&\leq \frac{1-e^{-\lambda}}{\lambda} \left(\frac{3}{2} \left(\sum_{k=1}^{r} \pi_k^2 + \sum_{j=1}^{r} \varrho_j^2 \right) - \frac{\lambda}{2\,r} \right) \\
&= \frac{1-e^{-\lambda}}{\lambda}\, \frac{3}{2} \left(\sum_{k=1}^{r} \pi_k^2 + \sum_{j=1}^{r} \varrho_j^2 - \frac{\lambda}{3\,r} \right).
\end{aligned}$$

Damit ist der Satz bewiesen.

Ausblick: Der lokale Ansatz und der Kopplungsansatz wurden hier für Poisson-Approximation vorgestellt. Natürlich werden diese Ansätze für Approximationen mittels anderer Verteilungen ebenfalls untersucht, und dies auch im Rahmen der Entwicklung der Steinschen Methode. Schliesslich rührt der jeweilige Ansatz von der Abhängigkeitsstruktur der Zufallsvariablen, die W bilden, her. Wir werden beide Ansätze in späteren Kapiteln, die die Approximation mittels einer Compound-Poisson-Verteilung bzw. mittels der Normalverteilung untersuchen, wiedertreffen, wobei der Kopplungsansatz nur in den Ausblicken Platz finden wird.

In einer Arbeit von Barbour [Bar88] wurde der Startschuss für die Entwicklung der Steinschen Methode für die Approximation mittels eines *Poissonschen Punktprozesses* gegeben. Weitere wichtige grundlegende Entwicklungen findet man in [BB92]. Wir werden den Fall der Punktprozess-Approximationen hier nicht behandeln. Er befindet sich aktuell in einer raschen Entwicklung, setzt aber Grundkenntnisse in der Theorie der Punktprozesse voraus, siehe etwa die Monographie [DVJ88]. Wir belassen es hier bei den folgenden Andeutungen:

Es sei ξ ein Punkt-Prozess auf $(S, \mathcal{S})$, wo S ein lokal kompakter Hausdorffscher topologischer Raum ist, der dem 2. Abzählbarkeitsaxiom genüge. ν sei das lokal endliche Mittelwert-Maß. Für jeden Punkt $x \in S$ sei ξ^x der assoziierte Palmsche Prozess, definiert auf dem gleichen Wahrscheinlichkeitsraum wie ξ. Dann kann man für jede relativ kompakte Menge $B \in \mathcal{S}$ die folgende Rechnung aus unseren bisherigen

Herleitungen (Kopplungsmethode) ableiten:

$$\begin{aligned}
&d_{TV}\big(\mathcal{L}(\xi(B)), \mathrm{Po}(\nu(B))\big) \\
&= \sup_{A\subset\mathbb{Z}_+} \big|\mathbb{E}\big(\xi(B)\, g_A(\xi(B)) - \nu(B)\, g_A(\xi(B)+1)\big)\big| \\
&= \sup_{A\subset\mathbb{Z}_+} \Big|\int_B \mathbb{E}\big(g_A(\xi^x(B)) - g_A(\xi(B)+1)\big) d\nu(x)\Big| \\
&\leq \frac{1-e^{-\nu(B)}}{\nu(B)} \int_B \mathbb{E}\big|\xi(B) - \xi^x(B) + 1\big|\, d\nu(x).
\end{aligned}$$

Weiter ist $\xi(B)$ eine meßbare Funktion auf dem Raum der lokal endlichen Zählmaße auf $(S, \mathcal{S})$. Daher folgt

$$d_{TV}\big(\mathcal{L}(\xi(B)), \mathrm{Po}(\nu(B))\big) \leq d_{TV}\big(\mathcal{L}(\xi), \mathcal{L}(\xi_0)\big),$$

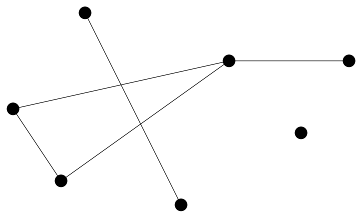
wobei ξ_0 einen Poissonschen Punkt-Prozess auf $(S, \mathcal{S})$ mit Intensitätsmaß ν bezeichnet. Die rechte Seite wird nun mittels der Steinschen Methode abgeschätzt, wobei die Schranken für die Lösungen der korrespondierenden Steinschen Gleichungen im Allgemeinen nicht so gut sind wie im Fall der diskreten Verteilung. Man erreicht hier eine deutliche Verbesserung, wenn man den Totalvariations-Abstand durch die sogenannte Wasserstein-Metrik ersetzt. Dies ist der Gegenstand aktueller Forschung, siehe etwa [BBX98] und [BWX00].

5. Zufallsgraphen

In diesem Kapitel wollen wir eine Einführung in die Theorie der zufälligen Graphen geben und in Kapitel 6 werden wir Poisson-Approximation diverser Zählstatistiken von Zufallsgraphen mittels der Steinschen Methode untersuchen.

Wir betrachten ein paar Grundbegriffe der Graphentheorie, beschreiben das Modell eines Zufallsgraphen und führen sogenannte Schwellenfunktionen ein. Dieser Begriff wird an Hand einiger Grapheneigenschaften untersucht. Historisch sind die zu beschreibenden Objekte 1959 von Paul Erdős eingeführt worden, und zwar zur Beantwortung einiger Fragen der Graphentheorie, siehe unter anderem [Erd59], [Erd61] sowie die Arbeit von Erdős und Rényi [ER60]. Wir schliessen das Kapitel mit dem Beweis eines fundamentalen Satzes von Erdős. Die Theorie der Zufallsgraphen ist mittlerweile sehr weit ausgebaut. Man betrachte zum Beispiel die Monographien von Bollobás [Bol01], von Alon, Spencer und Erdős [AS00] und von Janson, Łuczak und Rucinski [JŁR00].

Ein Graph ist eine Ansammlung von Punkten (Knoten), wobei manche Punkte durch eine Kante verbunden sind. Etwa



Definition 5.1. Ein *Graph* ist ein Paar $G = (V, E)$ disjunkter Mengen mit $E \subseteq [V]^2$ (bezeichne die Menge aller 2-elementigen Teilmengen von V). Elemente von V nennt man *Ecken* oder *Knoten* (engl. *vertices*) des Graphen, die Elemente von E heißen *Kanten* (engl. *edges*). Der Graph G heißt endlich bzw. unendlich je nachdem, ob V endlich oder unendlich ist. Für $|V|$ schreiben wir auch $|G|$ und nennen es die *Ordnung* von G. Zwei Knoten von G sind *benachbart* in G, wenn $xy \in E$ ist. xy bezeichnet kurz $\{x, y\}$; $x, y \in V$.

Zur Historie: Es ist manchmal schwer – warum auch immer –, einen Graphen mit einer speziellen Eigenschaft zu konstruieren: Gibt es Graphen, die gleichzeitig beliebig große *Taillenweite* und beliebig hohe *chromatische Zahl* haben? Erdős definierte für jedes $n \in \mathbb{N}$ auf der Menge aller Graphen der Ordnung n einen Wahrscheinlichkeitsraum und zeigte, dass bei geeigneter Wahl der Wahrscheinlichkeitsmaße die

Wahrscheinlichkeit, dass der gesuchte Graph existiert, für hinreichend große n positiv wird! Man spricht von der *probabilistischen Methode*. Wir werden diese historisch bedeutende Frage später genau untersuchen. Um die Frage aber zumindest formal verstehen zu können, hier die Definition der Begriffe:

Definition 5.2. (i) Ein *Weg* ist ein nicht leerer Graph $P = (V, E)$ mit

$$V = \{x_0, x_1, \ldots, x_k\} \quad \text{und} \quad E = \{x_0x_1, x_1x_2, \ldots, x_{k-1}x_k\},$$

wobei die x_i paarweise verschieden sind. Die Anzahl der Kanten eines Weges ist seine *Länge*. Ist $P = x_0 \cdots x_{k-1}$ (Kurzschreibweise) ein Weg und $k \geq 3$, so ist $C := P + x_{k-1}x_0$ ein *Kreis* (wobei + bedeutet: wir vereinen die Kantenmenge bei gleichen $V = \{x_0, \ldots, x_{k-1}\}$). $C = x_0 \cdots x_{k-1}x_0$ (Kurzschreibweise). Ein Kreis der Länge k wird mit C^k bezeichnet.

(ii) Die Länge eines kürzesten Kreises in einem Graphen G (bezüglich $\subseteq$: ein Graph $G'(V', E')$ ist ein Teilgraph von G, wenn $V' \subseteq V$ und $E' \subseteq E$ zu $G = (V, E)$) ist die *Taillenweite* $g(G)$ von G.

(iii) Eine *Eckenfärbung* eines Graphen $G = (V, E)$ ist eine Abbildung $c : V \to S$ mit $c(v) \neq c(w)$ für je zwei benachbarte Ecken v, w. S ist die Farbenmenge. Das kleinste $k \in \mathbb{N}$, so dass G eine Eckenfärbung hat ($c : V \to \{1, \ldots, k\}$), nennt man die (Ecken-) *chromatische Zahl* von G, $\chi(G)$ in Bezeichnung.

Die oben genannte Frage, dessen Antwort Erdős fand, kann nun wie folgt gestellt werden: Existiert zu jedem $k \in \mathbb{N}$ ein Graph G mit $g(G) > k$ *und* $\chi(G) > k$? Man benötigt Kenntnisse aus der Graphentheorie, um genauer zu verstehen, warum beide Forderungen gleichzeitig einer typischen Konstruktion entgegenstehen. Aus dem Bauch heraus verstehen wir jetzt aber, warum sich die beiden Forderungen an einen zu konstruierenden Graphen in die Quere kommen: Kreise (wie groß sie auch immer sind) benötigen zur Eckenfärbung nur 2 oder 3 Farben. Sie stören also, wenn man eine große Mindestzahl von Farben haben möchte.

Wir wenden uns den Zufallsgraphen zu: Zu $V = \{0, 1, \ldots, n-1\}$ wollen wir die Menge $\mathcal{G}$ der Graphen auf V in einen Wahrscheinlichkeitsraum verwandeln. Dann machen Fragen der Art "mit welcher Wahrscheinlichkeit hat ein Graph $G \in \mathcal{G}$ eine spezielle Eigenschaft" Sinn. Wir betrachten das sogenannte $\mathcal{G}(n, p)$-Modell:

$\mathcal{G}(n, p)$-Modell:

Für alle Eckenpaare $e \in [V]^2$ entscheidet man unabhängig voneinander, ob e eine Kante von G sein soll oder nicht. Die Wahrscheinlichkeit, dass e eine Kante sei, soll $0 < p < 1$ sein (zeitliche Interpretation: starte mit einem leeren Graphen auf

n Knoten. Jede der $\binom{n}{2}$ Kanten wird mit Wahrscheinlichkeit p eingefügt). Ist G_0 ein konkreter Graph auf V mit m Kanten, so hat das Elementarereignis $\{G_0\}$ die Wahrscheinlichkeit

$$p^m\, q^{\binom{n}{2}-m}.$$

Der Wahrscheinlichkeitsraum $\mathcal{G} = \mathcal{G}(n,p)$ ist nun ganz einfach zu konstruieren: zu $e \in [V]^2$ sei $\Omega_e := \{0,1\}$ und $P_e(0) = 1-p$ und $P_e(1) = p$. Dann bezeichne $\mathcal{G}(n,p)$ den Produktraum (Ω, P) mit

$$\Omega := \prod_{e\in[V]^2} \Omega_e, \quad P \text{ Produktmaß.}$$

Ein $\omega \in \Omega$ identifiziert man mit einem Graphen G auf V und Kantenmenge $E = \{e : \omega(e) = 1\}$. Man sagt nun, G ist ein *Zufallsgraph* auf V mit Kantenwahrscheinlichkeit p. Es sei weiter

$$A_e := \{\omega : \omega(e) = 1\}$$

(e ist eine Kante in G).

Korollar 5.3. *Die Ereignisse $(A_e)_e$ sind unabhängige Ereignisse mit Wahrscheinlichkeit p.*

Dies ist nach Konstruktion des Produktraumes klar.

Es sei $X : \mathcal{G}(n,p) \to \mathbb{N}$ die Zufallsgröße, die jedem Zufallsgraphen G die Anzahl seiner zu C^k isomorphen Teilgraphen zuordnet. Wir sagen genauer, was dies bedeutet:

Definition 5.4. $G = (V,E)$ und $G' = (V',E')$ seien zwei Graphen. Gilt $V' \subseteq V$ und $E' \subseteq E$, so ist G' ein *Teilgraph* von G. G heißt *isomorph* zu G', $G \simeq G'$ in Zeichen, wenn es eine Bijektion $\varphi : V \to V'$ gibt mit

$$xy \in E \Leftrightarrow \varphi(x)\varphi(y) \in E' \quad \text{für alle } x,y \in V.$$

φ heißt *Isomorphismus*.

Um mit den neuen Begriffen ein wenig Übung zu bekommen, zeigen wir

Lemma 5.5. *Die mittlere Anzahl von Kreisen der Länge $k(\geq 3)$ in $G \in \mathcal{G}(n,p)$ beträgt*

$$\mathbb{E}(X) = \frac{(n)_k}{2k} p^k,$$

wobei wir $(n)_k := n(n-1)(n-2)\dots(n-k+1)$ setzen.

Beweis: Für $C := C^k$ mit Ecken aus $V = \{0, \dots, n-1\}$ sei

$$X_C : G \mapsto \begin{cases} 1, & \text{falls } C \subseteq G \text{ ist,} \\ 0, & \text{sonst.} \end{cases}$$

Dann ist

$$\mathbb{E}(X_C) = P(X_C = 1) = P(\text{ Menge aller Graphen aus } \mathcal{G}(n,p), \text{ die} \, C \text{ enthalten}).$$

Dies ist nichts anderes als die Wahrscheinlichkeit, dass C ein Kreis in G ist, also $P(C \subseteq G) = p^k$ (da C ein Kreis der Länge k ist). Wieviele dieser Kreise $C = v_0 \cdots v_{k-1} v_0$ gibt es? Es gibt $(n)_k$ Folgen $v_0 \cdots v_{k-1}$ mit (unterschiedlichen) Ecken aus V. Jeder Kreis wird durch $2\,k$ dieser Folgen beschrieben: also gibt es $\frac{(n)_k}{2k}$ solche Kreise. Nun ist X die Summe aller $X_C(G)$, wobei C die $\frac{(n)_k}{2k}$ Kreise der Länge k mit Ecken aus V durchläuft, also

$$\mathbb{E}(X) = \sum_C \mathbb{E}(X_C) = \frac{(n)_k}{2k} p^k.$$

Damit ist das Lemma bewiesen. $\square$

Schwellenfunktionen:

Man sagt, dass ein Graph G eine *Eigenschaft* $\mathcal{A} \subseteq \mathcal{G}(n,p)$ hat, wenn $G \in \mathcal{A} \subseteq \mathcal{G}(n,p)$ gilt. Erdős und Rényi entdeckten 1960, dass viele Grapheneigenschaften ab einem gewissen $p(n)$ eintreten. Wir lassen nun also p in n variieren. Geht $P(G \in \mathcal{A}) \to 1$ für $n \to \infty$, so sagt man, dass *fast alle* G in $\mathcal{G}(n,p)$ die Eigenschaft $\mathcal{A}$ haben. Strebt die Wahrscheinlichkeit gegen 0, so hat *fast kein* G die Eigenschaft $\mathcal{A}$. (Vorsicht im Vergleich zur üblichen Definition von “fast alle” (fast sicher).)

Ein kleiner Eindruck der *Evolution* eines Zufallsgraphen: ist $p(n)$ wesentlich unterhalb von $1/n^2$, hat fast jeder Graph nur isolierte Ecken. Ab $p(n) = 1/(n\sqrt{n})$ hat fast jeder Graph die ersten Komponenten mit mehr als zwei Ecken (dabei ist eine Komponente ein maximal zusammenhängender nicht leerer Teilgraph, ohne Details). Ab $p(n) = 1/n$ tauchen die ersten Kreise auf. Eine spezielle Komponente beginnt stärker zu wachsen als die anderen und bei etwa $p(n) = \log n/n$ “verschlingt” sie andere und die Graphen werden zusammenhängend.

Die Entwicklung vollzieht sich in Schüben: die genannten $p(n)$-Werte sind Schwellenwerte, unterhalb derer fast kein Graph und oberhalb derer fast jeder Graph die betreffende Eigenschaft hat.

Definition 5.6. Eine reelle Funktion $t = t(n)$ mit $t(n) \neq 0$ für alle $n \in \mathbb{N}$ heißt *Schwellenfunktion* für eine Grapheneigenschaft $\mathcal{A}$, wenn für jedes $p = p(n)$ und $G \in \mathcal{G}(n, p)$ gilt:

$$\lim_{n\to\infty} P(G \in \mathcal{A}) = \begin{cases} 0, & \text{falls } p/t \underset{n\to\infty}{\to} 0, \\ 1, & \text{falls } p/t \underset{n\to\infty}{\to} \infty. \end{cases}$$

Ist t Schwellenfunktion, so natürlich auch $c\,t$ mit $c \in \mathbb{R}_+$. Schwellenfunktionen sind also nur bis auf einen konstanten Faktor eindeutig bestimmt (wenn es sie gibt).

Wir wollen nun eine sehr zentrale Schwellenfunktion bestimmen: Zu einem gegebenen Graphen H bezeichnen wir mit $\mathcal{A}_H$ die Grapheneigenschaft, eine Kopie von H als Teilgraphen zu enthalten:

$$\mathcal{A}_H := \{G : H \subseteq G\}.$$

Definition 5.7. (i) Die *Dichte* eines Graphen $G = (V, E)$ ist definiert durch

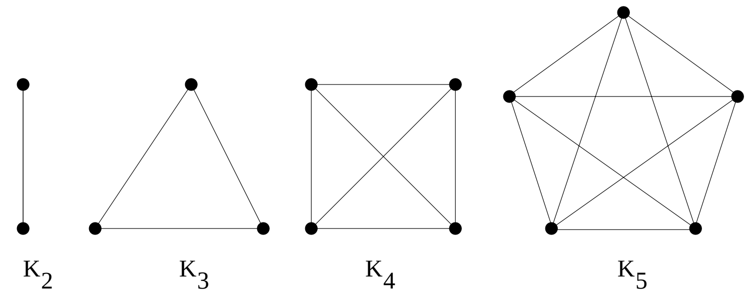
$$\varrho(G) = \frac{|E|}{|V|}.$$

(ii) Man nennt G *ausgewogen* oder *balanciert*, wenn für alle Teilgraphen G' von G gilt:

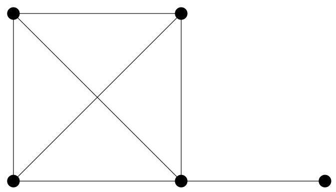
$$\varrho(G') \leq \varrho(G)$$

(*strikt ausgewogen, strikt balanciert*, wenn $<$ gilt).

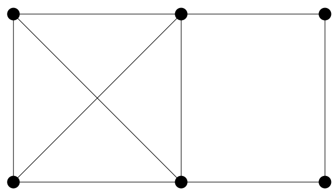
Beispiele 5.8. (i) Sind je zwei Ecken von G benachbart, so heißt G *vollständig*. Hat er k Ecken, bezeichnen wir diesen Graphen mit K_k. K_k ist strikt balanciert.



(ii)



nicht balanciert



balanciert

Der erste Graph ist nicht balanciert, denn die Dichte ist $7/5$ und K_4 hat Dichte $\frac{3}{2}$. Der zweite ist balanciert, aber nicht strikt (liegt an K_4).

Satz 5.9. *Es sei H ein balancierter Graph mit k Ecken und $l \geq 1$ Kanten. Dann ist $t(n) := n^{-k/l}$ eine Schwellenfunktion für $\mathcal{A}_H$.*

Der Beweis ist recht umfangreich und wird eine sehr wichtige, vielfach angewendete Methode präsentieren. Zuvor sammeln wir ein paar Folgerungen aus diesem Satz:

Korollar 5.10. *Ist $k \geq 3$ und $\mathcal{A}$ die Eigenschaft, einen Kreis der Länge k als Teilgraph zu enthalten, so ist $t(n) = 1/n$ eine Schwellenfunktion für $\mathcal{A}$.*

Beweis: Hier ist $l = k$ und C^k ist balanciert. □

Definition 5.11. Ein Graph heißt *zusammenhängend*, wenn er für je zwei seiner Ecken x, y einen Weg von x nach y enthält. Ein Graph, der keinen Kreis enthält und zusammenhängend ist, heißt *Baum*. Eine alternative Definition (ohne Beweis): Ein zusammenhängender Graph mit k Ecken ist genau dann ein Baum, wenn er $k-1$ Kanten hat.

Korollar 5.12. *Ist T ein Baum der Ordnung $k \geq 2$ und $\mathcal{A}$ die Eigenschaft, eine Kopie von T als Teilgraph zu enthalten, so ist $t(n) = n^{-k/k-1}$ eine Schwellenfunktion für $\mathcal{A}$.*

Beweis: $k/k-1$ folgt aus der Definition eines Baums. Ein Baum ist balanciert. □

Bemerke, dass die Schwelle beim Kreis unabhängig ist von der Länge des Kreises. Beim Baum ist sie unabhängig von der Gestalt des Baumes.

Korollar 5.13. *Ist $k \geq 2$ und $\mathcal{A}$ die Eigenschaft, einen K_k als Teilgraph zu enthalten, so ist $t(n) = n^{-2/k-1}$ eine Schwellenfunktion.*

Beweis: Es gilt immer

$$\varrho(K_i) = \frac{1}{2}(i-1) < \frac{1}{2}(k-1) = \varrho(K_k)$$

für alle $i < k$. Nun ist $l = \binom{k}{2}$ für K^k. Es ist $l = \frac{1}{2}k(k-1)$ und daher ist die Schwelle $n^{-k/l} = n^{-2/k-1}$. $\square$

Die *Cliquenzahl* eines Graphen G ist definiert durch

$$w(G) := \max\{k : \text{ es existiert ein } K_k \text{ in } G\}.$$

Damit ist zum Beispiel die Schwellenfunktion der Eigenschaft

$$\mathcal{A} = \{\text{Existenz einer Clique der Größe mindestens } 4\} = \{w(G) \geq 4\}$$

gleich $n^{-2/3}$. Dies ist eine einfache Monotonie-Überlegung.

Es folgt nun der Beweis von Satz 5.9:

Vorbereitung: Es sei $X(G)$ die Anzahl der zu H isomorphen Teilgraphen von $G \in \mathcal{G}(n,p)$. Weiter sei $\mathcal{H}$ die Menge aller zu H isomorphen Graphen auf Teilmengen von $\{0,\ldots,n-1\}$ (Eckenmenge der Graphen aus $\mathcal{G}(n,p)$). Formalisiert liefert dies:

$$\mathcal{H} = \{H' : H' \simeq H, V(H') \subseteq \{0,\ldots,n-1\}\}$$

Es sei nun $H' \in \mathcal{H}$ und $G \in \mathcal{G}(n,p)$. Ist $H' \subseteq G$, so bezeichnet dies die Tatsache, dass H' ein Teilgraph von G ist.

Sei h die Anzahl der zu H isomorphen Graphen auf einer festen Menge von k Ecken. Dann ist $h \leq k!$ und somit

$$|\mathcal{H}| = \binom{n}{k} h \leq \binom{n}{k} k! \leq n^k.$$

Zu $p = p(n)$ sei $\gamma = \frac{p}{t}$, also

$$p = \gamma\, n^{-k/l}.$$

Wir müssen nun zeigen:

(i) fast kein $G \in \mathcal{G}(n,p)$ liegt in $\mathcal{A}_H$, falls $\gamma \to 0$ für $n \to \infty$ und
(ii) fast alle $G \in \mathcal{G}(n,p)$ liegen in $\mathcal{A}_H$, falls $\gamma \to \infty$ für $n \to \infty$.

Beweis von Punkt (1): Hier führen wir den Beweis mit Hilfe der *Methode der ersten Momente*: Dazu eine Erinnerung:

Lemma 5.14. *Ist X eine nicht-negative, ganzzahlige Zufallsgröße, so ist*

$$P(X > 0) \leq \mathbb{E}(X).$$

Beweis: Es gilt

$$\mathbb{E}(X) = \sum_{i \geq 0} i\, P(X = i) \geq \sum_{i \geq 1} P(X = i) = P(X > 0).$$

□

Gilt für eine Folge von Zufallsgrößen $(X_n)_n$, dass $\lim_{n\to\infty} \mathbb{E}(X_n) = 0$, so folgt

$$\lim_{n\to\infty} P(X_n = 0) = 1,$$

und man sagt hierzu, X_n ist fast sicher gleich Null. Wir wollen also $\mathbb{E}(X)$ bestimmen. Es gilt

$$\mathbb{E}(X) = \sum_{H' \in \mathcal{H}} P(H' \subseteq G) = |\mathcal{H}|\, p^l \leq n^k (\gamma n^{-k/l})^l = \gamma^l$$

und mit γ geht also auch $\mathbb{E}(X)$ für wachsende n gegen Null. Also liegt fast kein G aus $\mathcal{G}(n,p)$ in $\mathcal{A}_H$.

Beweis von Punkt (2): Hier wird uns die *Methode der zweiten Momente* helfen:

Lemma 5.15. *Ist X eine Zufallsgröße mit Werten in $\mathbb{N}_0$, so gilt*

$$P(X = 0) \leq \frac{V(X)}{(\mathbb{E}X)^2}.$$

Beweis: Für $\lambda := \frac{\mathbb{E}(X)}{\sqrt{V(X)}}$ ist

$$P(X = 0) \leq P\big(|X - \mathbb{E}(X)| \geq \lambda\sqrt{V(X)}\big) = P\big(|X - \mathbb{E}(X)| \geq \mathbb{E}(X)\big) \leq \frac{V(X)}{(\mathbb{E}X)^2}$$

nach der Tschebyschev-Ungleichung. □

Ist nun $V(X) = o((\mathbb{E}X)^2)$, so ist $X > 0$ fast sicher. Wir untersuchen also $V(X)$: zunächst ist

$$\mathbb{E}(X^2) = \sum_{(H',H'') \in \mathcal{H}^2} P(H' \cup H'' \subseteq G) = \sum_{(H',H'') \in \mathcal{H}^2} p^{2l - |E(H' \cap H'')|}.$$

Nun ist $\varrho(H' \cap H'') \leq \varrho(H) = l/k$ (balanciert!). Ist $|V(H' \cap H'')| = i$, so folgt $|E(H' \cap H'')| \leq \frac{il}{k}$. Also ist wegen $p < 1$

$$P(H' \cup H'' \subseteq G) \leq p^{2l - \frac{il}{k}}.$$

Nun betrachten wir

$$\mathcal{H}_i^2 := \{(H', H'') \in \mathcal{H}^2 : |V(H' \cap H'')| = i\}$$

für $i = 0, \ldots, k$. Wir bestimmen nun

$$\sum_i P(H' \cup H'' \subseteq G) =: A_i,$$

wobei $\sum_i$ die Summe über alle $(H', H'') \in \mathcal{H}_i^2$ bezeichne. Im Fall $i = 0$ sind H' und H'' disjunkt und $\{H' \subseteq G\}$ ist somit unabhängig von $\{H'' \subseteq G\}$. Also

$$A_0 \leq \sum_{(H',H'') \in \mathcal{H}^2} P(H' \subseteq G)\, P(H'' \subseteq G) = (\mathbb{E}X)^2.$$

Für $i \geq 1$ ist

$$\sum_i = \sum_{H',H'' \in \mathcal{H} : |V(H' \cap H'')| = i}.$$

Ist H' fest gewählt, so hat die Summe über die H''

$$\binom{k}{i}\binom{n-k}{k-i} h$$

Summanden. Damit ist

$$\begin{aligned}
A_i &\leq \sum_{H' \in \mathcal{H}} \binom{k}{i}\binom{n-k}{k-i}\, h\, p^{2l} p^{-il/k} \\
&= |\mathcal{H}| \binom{k}{i}\binom{n-k}{k-i}\, h\, p^{2l} \left(\gamma\, n^{-k/l}\right)^{-il/k} \\
&\leq |\mathcal{H}|\, p^l\, c_1\, n^{k-i}\, h\, p^l\, \gamma^{-il/k}\, n^i \\
&= \mathbb{E}(X)\, c_1\, n^k\, h\, p^l\, \gamma^{-il/k} \\
&\leq \mathbb{E}(X)\, c_2 \binom{n}{k}\, h\, p^l\, \gamma^{-il/k} \\
&= \mathbb{E}(X)^2\, c_2\, \gamma^{-il/k} \\
&\leq \mathbb{E}(X)^2\, c_2\, \gamma^{-l/k}.
\end{aligned}$$

Also folgt mit $c_3 = k\, c_2$:

$$\frac{\mathbb{E}(X^2)}{(\mathbb{E}X)^2} = \left(\frac{A_0}{(\mathbb{E}X)^2} + \sum_{i=1}^{k} A_i \frac{1}{(\mathbb{E}X)^2} \right) \leq 1 + c_3\, \gamma^{-l/k},$$

also

$$\frac{\mathbb{E}(X^2) - (\mathbb{E}X)^2}{(\mathbb{E}X)^2} \leq c_3\, \gamma^{-l/k},$$

und die rechte Seite geht für $\gamma \to \infty$ gegen Null. Somit ist $X(G) > 0$ für fast alle $G \in \mathcal{G}(n,p)$, das heißt fast alle $G \in \mathcal{G}(n,p)$ enthalten eine Kopie von H als Teilgraphen und liegen somit in $\mathcal{A}_H$. Damit ist der Satz bewiesen.

Für die Anzahl $X(G)$ der zu H isomorphen Teilgraphen von $G \in \mathcal{G}(n,p)$ hatten wir $\mathbb{E}(X) = |\mathcal{H}|\, p^l \sim n^k\, p^l$ gesehen. Somit folgt bei der Wahl $\lim_{n\to\infty} p/n^{-k/l} = 0$ für $p = p(n)$, dass $\lim_{n\to\infty} \mathbb{E}(X) = 0$, und für die Wahl $\lim_{n\to\infty} p/n^{-k/l} = \infty$, dass $\lim_{n\to\infty} \mathbb{E}(X) = \infty$ gilt. Dies gilt für alle H mit k Ecken und l Kanten. Trotzdem gilt der folgende Satz:

Satz 5.16. *Es sei H nicht balanciert mit k Ecken und $l \geq 1$ Kanten. Dann ist $t(n) = n^{-k/l}$ keine Schwellenfunktion für $\mathcal{A}_H$.*

Beweis: Es sei $H' \subseteq H$ so gewählt, dass $\varrho(H') > \frac{l}{k}$ (H nicht balanciert). Dann ist

$$\begin{aligned}\mathbb{E}(X') &\sim n^{|V(H')|}\, p^{|E(H')|} \\ &= (n^{|V(H')|/|E(H')|}\, p)^{|E(H')|},\end{aligned}$$

wobei $X'(G)$ die Anzahl der zu H' isomorphen Teilgraphen von $G \in \mathcal{G}(n,p)$ bezeichnet und $|V(H')|$ ($|E(H')|$) die Knotenanzahl (Kantenanzahl) von H'. Wir wählen nun $p = p(n)$ so, dass

$$p(n)/n^{-k/l} \to \infty \quad \text{und} \quad p(n)/n^{-|V(H')|/|E(H')|} \to 0$$

(dies ist möglich, da $\frac{|E(H')|}{|V(H')|} > \frac{l}{k}$). Damit folgt aber $\lim\limits_{n\to\infty} \mathbb{E}(X') = 0$, also $\lim\limits_{n\to\infty} P(X' = 0) = 1$. Da weiter $\{X' = 0\} \subseteq \{X = 0\}$, kann $t(n)$ keine Schwellenfunktion für $\mathcal{A}_H$ sein. □

Wir stellen nun die zu Beginn des Kapitels erwähnte *probabilistische Methode* exemplarisch mittels des Beweises des Satzes von Erdős über die Existenz von Graphen mit hoher Taillenweite und chromatischer Zahl dar.

Satz 5.17. *(Erdős, 1959) Zu jedem $k \in \mathbb{N}$ gibt es einen Graphen H mit Taillenweite $g(H) > k$ und chromatischer Zahl $\chi(H) > k$.*

Wir bereiten den Beweis vor. Zu einem Graph $G = (V, E)$ heißt eine Teilmenge von V (Eckenmenge) *unabhängig*, wenn ihre Elemente paarweise nicht benachbart sind. Die größte Mächtigkeit einer unabhängigen Eckenmenge in G ist die *Unabhängigkeitszahl* $\alpha(G)$ von G.

Proposition 5.18. *Für jedes $k \geq 2$ gilt für die Wahrscheinlichkeit, dass $G \in \mathcal{G}(n,p)$ eine unabhängige Eckenmenge der Mächtigkeit k enthält,*

$$P\big(\alpha(G) \geq k\big) \leq \binom{n}{k}(1-p)^{\binom{k}{2}}.$$

Beweis: Die Wahrscheinlichkeit, dass eine fest gewählte Eckenmenge $U \in [V]^k$ in G unabhängig ist, beträgt $(1-p)^{\binom{k}{2}}$. Es gibt nur $\binom{n}{k}$ solche Mengen U. □

Bemerkung 5.19. (analog zur Proposition zu beweisen)

$$P(\omega(G) \geq k) \leq \binom{n}{k} p^{\binom{k}{2}}.$$

(Wahrscheinlichkeit, dass G in $\mathcal{G}(n,p)$ einen K_k enthält; Erinnerung: $\omega(G)$ ist die Cliquenzahl).

Wir nennen Kreise der Länge $\leq k$ *kurz* und Mengen von $\frac{|G|}{k}$ oder mehr Knoten von G *groß*. Wir suchen (Erdős suchte) einen Graphen G *ohne kurze Kreise* und *ohne große unabhängige Eckenmengen*: die letzt genannte Eigenschaft impliziert, dass wir mehr als k Farben brauchen, um G zu färben (denn: $\alpha(G) < \frac{|G|}{k}$, also $k < \frac{|G|}{\alpha(G)} \leq \chi(G)$). Wählt man p, die Kantenwahrscheinlichkeit, klein genug, so werden die Zufallsgraphen aus $\mathcal{G}(n,p)$ mit hoher Wahrscheinlichkeit keine (kurzen) Kreise haben (wir kennen bereits die Schwelle $p(n) = 1/n$). Intuitiv erwartet man bei p groß genug, dass kaum große unabhängige Eckenmengen auftreten. Es ergibt sich somit die folgende Frage: Kann man p so wählen, dass es für große n gleichzeitig klein genug für $P(g \leq k) < \frac{1}{2}$ und groß genug für $P(\alpha \geq n/k) < \frac{1}{2}$ ist? Wenn ja, so enthält $\mathcal{G}(n,p)$ mindestens einen Graphen ohne kurze Kreise und ohne große unabhängige Eckenmengen. Tatsächlich kann man p so nicht wählen, da $1/n$ auch eine Schwellenfunktion für große unabhängige Mengen ist. Der Beweis des Satzes verfolgt eine andere Idee.

Beweis von Satz 5.17: Es sei $\varepsilon > 0$ und $\varepsilon < 1/k$ fest gewählt. Sei $p := p(n) = \frac{n^\varepsilon}{n} = \frac{1}{n^{1-\varepsilon}}$. Wir liegen oberhalb der Schwelle $1/n$ und kümmern uns um die Anzahl $X(G)$ der kurzen Kreise in einem Zufallsgraphen $G \in \mathcal{G}(n,p)$ (Anzahl der Kreise der Länge $\leq k$). Es gilt mit Lemma 5.5:

$$\mathbb{E}(X) = \sum_{i=3}^{k} \frac{(n)_i}{2i} p^i \leq \frac{1}{2} \sum_{i=3}^{k} p^i\, n^i \leq \frac{1}{2}(k-2) n^k\, p^k,$$

denn $(n\,p)^i \leq (n\,p)^k$ wegen $n\,p = n^\varepsilon \geq 1$. Daraus folgt

$$P\left(X \geq \frac{n}{2}\right) \leq \frac{\mathbb{E}(X)}{n/2} \leq (k-2)\, n^{k-1}\, p^k = (k-2)\, n^{k-1}\, n^{(\varepsilon-1)k} = (k-2)\, n^{k\,\varepsilon - 1}.$$

Nun ist $k\,\varepsilon - 1 < 0$ nach obiger Wahl, also

$$\lim_{n\to\infty} P\left(X \geq \frac{n}{2}\right) = 0.$$

Wir wählen n so groß, dass $P\big(X \geq n/2\big) < 1/2$. Wir finden also einen Graphen $G \in \mathcal{G}(n,p)$ mit weniger als $n/2$ kurzen Kreisen.

Behauptung: n kann so groß gewählt werden, dass bei obiger Wahl von $p(n) = \frac{1}{n^{1-\varepsilon}}$ $P\big(\alpha \geq \frac{n}{2k}\big) < 1/2$ gilt. Wir finden dann einen Graphen $G \in \mathcal{G}(n,p)$ mit $\alpha(G) < \frac{n}{2k}$ und mit weniger als $n/2$ kurzen Kreisen. Dann sind wir aber bereits am Ziel:

Aus *jedem* kurzen Kreis von G entfernen wir eine Ecke. Der entstehende Graph H hat dann noch mindestens $n/2$ Ecken. Er enthält keine kurzen Kreise mehr, also ist $g(H) > k$ und weiter gilt:

$$\chi(H) \geq \frac{|H|}{\alpha(H)} \geq \frac{n/2}{\alpha(G)} > k$$

(die Unabhängigkeitszahl ist nach Konstruktion nicht größer als die von G). Damit ist der Satz bewiesen, wenn obige Behauptung nachgeliefert wird.

Lemma 5.20. *Es seien $k > 0$ und $p = p(n)$ gegeben. Ist $p \geq \frac{6k \log n}{n}$ für große n, so gilt*

$$\lim_{n\to\infty} P\left(\alpha \geq \frac{n}{2k}\right) = 0.$$

Mit Hilfe dieses Lemmas folgt dann für $p = n^{\varepsilon - 1}$ für beliebig kleines $\varepsilon > 0$, dass ein $G \in \mathcal{G}(n,p)$ existiert mit $\alpha(G) < n/(2k)$.

Beweis von Lemma 5.20: Es seien $n, r \in \mathbb{N}$ mit $n \geq r \geq 2$. Weiter sei $G \in \mathcal{G}(n,p)$. Mit $q := 1 - p$ gilt nach Proposition 5.18

$$\begin{aligned} P(\alpha \geq r) &\leq \binom{n}{r} q^{\binom{r}{2}} \leq n^r\, q^{\binom{r}{2}} \\ &= \big(n\, q^{\frac{r-1}{2}}\big)^r \leq \big(n\, e^{-\frac{p(r-1)}{2}}\big)^r, \end{aligned}$$

wobei wir $(1-p) \leq e^{-p}$, $0 < p < 1$, verwendet haben. Es sei nun $p \geq \frac{6k \log n}{n}$ und $r \geq \frac{n}{2k}$, so folgt

$$\begin{aligned} n\, e^{-p(r-1)/2} &= n\, e^{-pr/2 + p/2} \leq n\, e^{-\frac{3}{2}\log n + p/2} \\ &\leq n\, n^{-3/2} e^{1/2} = \frac{\sqrt{e}}{\sqrt{n}} \underset{n\to\infty}{\to} 0. \end{aligned}$$

Mit $r := [\frac{n}{2k}]$ ist

$$\lim_{n\to\infty} P\left(\alpha \geq \frac{n}{2k}\right) = \lim_{n\to\infty} P(\alpha \geq r) = 0,$$

womit das Lemma bewiesen ist. $\square$

6. Poisson-Approximation für Zufallsgraphen

Mit Hilfe von Satz 5.9 konnten wir das asymptotische Verhalten der Wahrscheinlichkeit des Ereignisses, dass ein Zufallsgraph einen *balancierten* Teilgraphen enthält, bestimmen. An der Schwelle $t(n) = n^{-k/l}$ springt diese Wahrscheinlichkeit im Limes von 0 auf 1(die genaue Bedeutung liefert die Definition der Schwellenfunktion, siehe Definition 5.6). Wir wollen in diesem Kapitel genauere Approximationen der Größenordnung der genannten Wahrscheinlichkeit untersuchen. Dabei wenden wir die *Steinsche Methode* an.

Analog zur Notation im Beweis von Satz 5.9 sei fortan $X(G)$ die Anzahl der zu H isomorphen Teilgraphen von $G \in \mathcal{G}(n,p)$. Wir beweisen, dass oberhalb der Schwelle – im Fall, dass H balanciert ist – sich X stochastisch wie der Erwartungswert $\mathbb{E}(X)$ verhält. Unterhalb der Schwelle erhält man eine in Null konzentrierte Verteilung. Im Schwellenbereich untersuchen wir anschließend die Poisson-Approximation. Diese können wir für die Teilklasse der *strikt balancierten* Graphen herleiten. Die Steinsche Methode liefert erneut Informationen über die Konvergenzgeschwindigkeit. Abschließend betrachten wir die Anzahl isolierter Knoten in einem Zufallsgraph.

Nach der Tschebyschev–Ungleichung gilt immer

$$P\big(|X_n - \mathbb{E}(X_n)| > \varepsilon\, \mathbb{E}(X_n)\big) \le \frac{V(X_n)}{\varepsilon^2(\mathbb{E}(X_n))^2},$$

sobald die Größen existieren. Wenn also $V(X_n) = o((\mathbb{E}(X_n))^2)$ gilt, so konvergiert $X_n/\mathbb{E}(X_n)$ stochastisch gegen 1, womit wir

$$\lim_{n\to\infty} P\left(\left|\frac{X_n}{\mathbb{E}(X_n)} - 1\right| > \varepsilon\right) = 0 \quad \text{für alle } \varepsilon > 0$$

meinen. Wir können zeigen:

Satz 6.1. *Es sei H ein balancierter Graph mit k Ecken und $l \ge 1$ Kanten. Wenn $\lim_{n\to\infty} p(n)\, n^{k/l} = \infty$ gilt, so folgt*

$$\lim_{n\to\infty} P\left(\left|\frac{X}{\mathbb{E}(X)} - 1\right| > \varepsilon\right) = 0 \quad \textit{für alle } \varepsilon > 0.$$

Beweis: Der Kern des Beweises von Satz 5.9 zeigte $V(X) = o((\mathbb{E}(X))^2)$, womit wir Satz 6.1 quasi geschenkt bekommen. □

Wir wollen nun die Verteilung von X analysieren. Ist $p(n) \ll n^{-k/l}$, so erwarten wir im Limes eine Verteilung von X, die in der Null konzentriert ist.

Bemerkung 6.2. Es gilt, dass X in Verteilung gegen 0 für $n \to \infty$ konvergiert, wenn $p(n) \ll n^{-k/l}$. Wir formalisieren dies nicht weiter.

Was passiert, wenn

$$\lim_{n\to\infty} p(n)\, n^{k/l} = c > 0$$

gilt ($p(n)$ im Schwellenbereich)? Wir untersuchen die Poisson-Approximation mittels der Steinschen Methode. Zunächst betrachten wir den lokalen Ansatz (siehe Kapitel 3). Wir gleichen die Notation an:

Sei H ein fester Graph mit k Ecken und $l \geq 1$ Kanten. Γ sei die Menge aller Kopien von H in K_n (alle Teilgraphen von K_n isomorph zu H). Zu $\alpha \in \Gamma$ ist $I_\alpha := I(\alpha \subset \mathcal{G}(n,p))$ und

$$X = W = \sum_{\alpha\in\Gamma} I_\alpha.$$

Dann ist $\mathbb{E}(I_\alpha) = \pi_\alpha = p^l$. Weiter ist $|\Gamma| = \binom{n}{k}|\mathcal{H}|$ und $\lambda = \mathbb{E}W = \binom{n}{k}|\mathcal{H}|p^l \sim c^l$ (laut Voraussetzung). Zu $\alpha \in \Gamma$ ist Γ_α die Menge der $\beta \in \Gamma\setminus\{\alpha\}$, für die I_β abhängig ist von I_α. β muß also mindestens eine gemeinsame Kante, deshalb mindestens 2 gemeinsame Knoten mit α haben. Mit der Definition

$$Z_\alpha := \sum_{\beta\in\Gamma_\alpha} I_\beta = \sum_{s=2}^{k} \sum_{|V(\alpha\cap\beta)|=s} I_\beta$$

folgt die Abschätzung

$$\begin{aligned}
\mathbb{E}Z_\alpha &\leq \sum_{s=2}^{k} \binom{k}{s}\binom{n-k}{k-s} p^l\, k! \\
&= \mathcal{O}\Big(\sum_{s=2}^{k} p^l\, n^{k-s}\Big) = \mathcal{O}\big(n^{k-2}\, p^l\big) = \mathcal{O}\big(n^{-2}\big).
\end{aligned}$$

Weiter folgt analog zu den Rechnungen auf Seite 54 für ein balanciertes H: $\mathbb{E}(I_\alpha I_\beta) \leq p^{2\,l - s\,\frac{l}{k}} = p^{\frac{l}{k}(2k-s)}$, wenn α und β genau s gemeinsame Knoten haben. Also folgt

$$\begin{aligned}
\sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha} \mathbb{E}(I_\alpha I_\beta) &\leq |\Gamma| \sum_{s=2}^{k} \binom{k}{s}\binom{n-k}{k-s} k!\, p^{\frac{l}{k}(2k-s)} \\
&= \mathcal{O}\Big(\sum_{s=2}^{k} n^k\, n^{k-s} p^{\frac{l}{k}\,(2k-s)}\Big) = \mathcal{O}\Big(\sum_{s=2}^{k} (n\, p^{l/k})^{2\,k-s}\Big) = \mathcal{O}\,(\text{const.}).
\end{aligned}$$

Mit $\sum_{\alpha\in\Gamma} \pi_\alpha^2 = |\Gamma| p^{2l}$ und $\lambda \sim c^l$ liefert Satz 3.1 insgesamt

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq p^l + \mathcal{O}(n^{-2}) + \mathcal{O}\,(\text{const.}).$$

Diese Rechnung zeigt, dass die lokale Steinsche Methode im allgemeinen im Schwellenbereich für $p(n)$ keine verwendbare Poisson-Approximation liefert. Es gilt aber (siehe [Bar82]):

Satz 6.3. *(Barbour, 1982) Es sei H ein strikt balancierter Graph mit k Ecken und $l \geq 1$ Kanten. Ist $p(n)$ so gewählt, dass $\lim_{n\to\infty} p(n)\, n^{k/l} = c > 0$ gilt für ein $c > 0$, so folgt*

$$\lim_{n\to\infty} d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) = 0.$$

(Im Schwellenbereich liegt für strikt balancierte Teilgraphen stets Poisson-Approximation für die Anzahl W vor.)

Wir haben in den obigen Rechnungen nur die Balanciertheit benötigt. Bei der Analyse von $\sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha} \mathbb{E}(I_\alpha I_\beta)$ werden wir nun die strengere Voraussetzung der strikten Balanciertheit geeignet verwenden. Dazu führen wir den *Balanceindex* ein. Sei H strikt balanciert, so definieren wir

(i) $m_t(H) = \max\{\varrho(F); F \subsetneq H, V(F) = t\}$. Also ist $m_t(H) < \varrho(H)$ für $1 \leq t \leq k$. $m_1(H) = 0$.
(ii) $i_t(H) = \varrho(H) - m_t(H) > 0$ für $1 \leq t \leq k$. Es gilt $i_1(H) = \varrho(H)$.
(iii) (*Balanceindex*) $\varepsilon := \varepsilon(H) := \min_{1\leq t\leq k} i_t(H)$. Es gilt: $\varepsilon \leq i_1(H) = \varrho(H)$ und $\varepsilon > 0$.

Lemma 6.4. *Sei H strikt balanciert, G beliebig aber $G \not\supseteq H$ und $|V(G\cap H)| \geq 1$. Dann gilt*

$$|E(H\cup G)| \geq |E(H)| + |E(G)| - \varrho(H)|V(H\cap G)| + \varepsilon.$$

Bemerkung 6.5. Ist $G \supseteq H$, so ist

$$|E(H)| + |E(G)| - \varrho(H)\,|V(H\cap G)| = |E(H)| + |E(G)| - |E(H)| = |E(H\cup G)|.$$

Hier stimmt also das $+\varepsilon$-Resultat nicht.

Beweis von Lemma 6.4: Es ist

$$\begin{aligned} |E(H\cap G)| &= \varrho(H\cap G)\,|V(H\cap G)| \leq m_{|V(H\cap G)|}\,|V(H\cap G)| \\ &= \big(\varrho(H) - i_{|V(H\cap G)|}\big)\,|V(H\cap G)| \\ &\leq \varrho(H)\,|V(H\cap G)| - \varepsilon. \end{aligned}$$

Mit

$$|E(H\cup G)| = |E(H)| + |E(G)| - |E(H\cap G)|$$

folgt dann die Behauptung. □

Lemma 6.6. *Es seien H_1 und H_2 verschiedene Kopien von H (wie in Lemma 6.4 gewählt) in K_n mit $|V(H_1 \cap H_2)| > 0$. Sei $F := H_1 \cup H_2$, dann gilt*

$$|E(F)| \geq \varrho(H)\,|V(F)| + \varepsilon.$$

Beweis: Setze in 6.4 $H := H_1$ und $G := H_2$, also $H_1 \not\subseteq H_2$ nach Voraussetzung. Damit folgt mit Lemma 6.4:

$$\begin{aligned} |E(F)| &= |E(H_1 \cup H_2)| \geq 2|E(H)| - \varrho(H)\,|V(H_1 \cap H_2)| + \varepsilon \\ &= 2|E(H)| - \varrho(H)\,(2|V(H)| - |V(H_1 \cup H_2)|) + \varepsilon \\ &= \varrho(H)\,|V(H_1 \cup H_2)| + \varepsilon. \end{aligned}$$

□

Beweis von Satz 6.3: Seien $\alpha, \beta \in \Gamma$ so, dass $|V(\alpha \cap \beta)| = s$ gilt. Also hat $\alpha \cup \beta$ genau $2k - s$ Eckpunkte. Es gilt

$$\mathbb{E}(I_\alpha I_\beta) \leq p^{\varrho(H)\,(2k-s)+\varepsilon} = p^{2|E(H)|-s\,\varrho(H)+\varepsilon}$$

(da $0 < p < 1$) und somit

$$\begin{aligned} \sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha} \mathbb{E}(I_\alpha I_\beta) &\leq |\Gamma| \sum_{s=2}^{k} \binom{k}{s}\binom{n-k}{k-s} k!\, p^{2l-\frac{sl}{k}+\varepsilon} \\ &= \mathcal{O}\Big(\sum_{s=2}^{k} n^k\, n^{k-s}\, p^{2l-\frac{sl}{k}+\varepsilon}\Big) = \mathcal{O}\Big(\sum_{s=2}^{k} (n\,p^{l/k})^{2k-s}\, p^{\varepsilon}\Big) \\ &= \mathcal{O}\big(p^{\varepsilon}(n\,p^{l/k})^{2k-2}\big) = \mathcal{O}\big(p^{\varepsilon}\big). \end{aligned}$$

Satz 3.1 liefert nun $d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq p^l + \mathcal{O}\big(n^{-2}\big) + \mathcal{O}\big(p^{\varepsilon}\big)$. □

Bemerkung 6.7. Unsere Rechnungen zeigen

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq p^l + (n^k\,p^l)\,n^{-2} + (p^l\,n^k)^{1-2/k}\,n^{-(k/l)\varepsilon}.$$

Hier kann λ auch groß werden. Dies diskutieren wir aber nicht genauer.

Für die gleiche Zufallsgröße W betrachten wir nun einen *Kopplungsansatz*. Wir erhalten eine schöne Darstellung der Schranke. Für $\beta \in \Gamma$ sei nun

$$J_{\beta\,\alpha} := I(\beta \subset \mathcal{G}(n,p) \cup \alpha) \geq I_\beta.$$

Es gilt $\mathcal{L}(J_{\beta\,\alpha}) = \mathcal{L}(I_\beta | I_\alpha = 1)$ für alle $\beta \in \Gamma$. Also ist (4.5) in Kapitel 4 erfüllt. Aber in (4.6) gilt nun die umgekehrte Anordnung. Wir nannten dies *positiv zusammenhängend*. Mit $U_\alpha := W$ und $V_\alpha := \sum_{\beta\neq\alpha} J_{\beta\,\alpha}$, also

$$\mathcal{L}(U_\alpha) = \mathcal{L}(W) \;\text{ und }\; \mathcal{L}(V_\alpha + 1) = \mathcal{L}(W | I_\alpha = 1)$$

erhalten wir nun (siehe [BHJ92, Section 5]):

Satz 6.8. *(Barbour, Holst, Janson, 1992) Für die obige Kopplung gilt:*

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq \frac{1-e^{-\lambda}}{\lambda}\left(V(W) - \lambda + 2\sum_{\alpha\in\Gamma}\pi_\alpha^2\right).$$

Beweis: Nach Definition gilt für jedes $\alpha \in \Gamma$: $V_\alpha + I_\alpha \geq U_\alpha$. Nach Satz 4.1 müssen wir $\mathbb{E}|U_\alpha - V_\alpha|$ abschätzen. Es gilt

$$\begin{aligned}
|U_\alpha - V_\alpha| &= |(U_\alpha - I_\alpha - V_\alpha) + I_\alpha| \\
&\leq |I_\alpha| + |U_\alpha - I_\alpha - V_\alpha| \\
&= |I_\alpha| + V_\alpha - U_\alpha + I_\alpha, \quad \text{da } U_\alpha - I_\alpha \leq V_\alpha \\
&\leq 2|I_\alpha| + V_\alpha - U_\alpha.
\end{aligned}$$

Nach dem Beweis von Korollar 4.3 wissen wir bereits, dass

$$\sum_{\alpha\in\Gamma}\pi_\alpha\mathbb{E}U_\alpha = (\mathbb{E}W)^2 \;\text{ und }\; \sum_{\alpha\in\Gamma}\pi_\alpha\mathbb{E}V_\alpha = \mathbb{E}(W^2) - \mathbb{E}W,$$

also

$$\sum_{\alpha\in\Gamma}\pi_\alpha\mathbb{E}|U_\alpha - V_\alpha| \leq 2\sum_{\alpha\in\Gamma}\pi_\alpha^2 + V(W) - \lambda.$$

□

Für die Zählvariable W des Teilgraphenproblems ist

$$\frac{2}{\lambda}\sum_{\alpha\in\Gamma}\pi_\alpha^2 = \frac{2}{p^l}(p^l)^2 = 2\,p^l,$$

also

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq \big(1-e^{-\lambda}\big)\left(\frac{V(W)}{\lambda} - 1 + 2\,p^l\right).$$

Wir betrachten die Varianz (siehe [BHJ92, Section 5]):

Lemma 6.9. *(Barbour, Holst, Janson, 1992) Zu H sei $c(H',H)$ die Anzahl der Kopien von H' in H. Dann gilt:*

$$\frac{V(W)}{\lambda} - 1 + 2\,p^l \leq \sum_{H'}\frac{a(H')}{a(H)}\,c(H',H)\,n^{|V(H)|-|V(H')|}\,p^{|E(H)|-|E(H')|},$$

wobei die Summe über alle $H' \subset H$ gebildet wird und H' keine isolierten Knoten habe. Desweiteren bezeichnet $a(H)$ die Anzahl der Elemente der Automorphismengruppe von H (Permutationen der Ecken von H, die Nachbarschaften erhalten).

Beweis: Es ist $V(W) = \sum_\alpha V(I_\alpha) + \sum_\alpha \sum_{\beta \neq \alpha} \operatorname{cov}(I_\alpha, I_\beta)$. Mit $V(I_\alpha) = \pi_\alpha(1-\pi_\alpha)$ $= (1-p^l)\,\pi_\alpha$ ist $V(W) = \lambda(1-p^l) + \sum_\alpha \sum_{t=1}^{l} \sum_{\beta \in \Gamma_\alpha^t} \operatorname{cov}(I_\alpha, I_\beta)$, wobei $\Gamma_\alpha^t \subset \Gamma$ die Menge aller Teilgraphen bezeichnet, die isomorph zu H sind und in t Kanten nicht mit α übereinstimmen. Da $\mathbb{E}(I_\alpha I_\beta) = \pi_\alpha\,\mathbb{E}(I_\beta | I_\alpha = 1)$, ist $\operatorname{cov}(I_\alpha, I_\beta) = \pi_\alpha\big(\mathbb{E}(I_\beta | I_\alpha = 1) - \pi_\beta\big)$. Nun ist $\mathbb{E}(I_\beta | I_\alpha = 1)$ für ein $\beta \in \Gamma_\alpha^t$ zu bestimmen. Da die Kanten unabhängig gewählt werden, ist diese bedingte Erwartung jeweils p^t. Also folgt

$$V(W) = \lambda\Bigg(1 - p^l + \sum_{t=1}^{l-1} \sum_{\beta \in \Gamma_\alpha^t} (p^t - p^l)\Bigg).$$

Sei N_H und $N_{H'}$ die Anzahl der Kopien von H und H' in K_n. Dann gibt es $N_H\, c(H', H)$ Paare $\tilde{H}', \tilde{H}$ mit $\tilde{H}' \subseteq \tilde{H} \subset K_n$ und $\tilde{H}' \simeq H'$ und $\tilde{H} \simeq H$. Somit ist jede Kopie von H' in $N_H\, c(H', H)/N_{H'}$ Kopien von H enthalten. Es folgt für $0 \leq t \leq l-1$:

$$|\Gamma_\alpha^t| < \sum_{H' : l-t = |E(H')|} \frac{c(H', H)\, N_H}{N_{H'}} \leq \sum_{H' : l-t = |E(H')|} c(H', H)\, \frac{a(H')}{a(H)}\, n^{k - |V(H')|}.$$

Damit bekommen wir

$$\begin{aligned} &\sum_{\beta \in \Gamma_\alpha^t} (p^t - p^l) \leq |\Gamma_\alpha^t| (p^t - p^l) \\ \leq &\sum_{H' : l-t = |E(H')|} c(H', H)\, \frac{a(H')}{a(H)}\, n^{|V(H)| - |V(H')|}\, (p^{|E(H)| - |E(H')|} - p^l). \end{aligned}$$

Insgesamt folgt

$$\frac{V(W)}{\lambda} - 1 + 2\,p^l \leq p^l + \sum_{H'} \frac{a(H')}{a(H)} c(H', H)\, n^{|V(H)| - |V(H')|}\, p^{|E(H)| - |E(H')|} - |\Gamma| p^l$$

und somit die Behauptung. □

Diskussion: Betrachte erneut den Fall $\lambda \sim n^{|V(H)|}\, p^{|E(H)|} = n^k\, p^l$. Angenommen es existiert ein $H' \subset H$ mit $\varrho(H') \geq \varrho(H)$, also $|V(H')| \leq \frac{k\,|E(H')|}{l}$, dann ist $n^{-|V(H')|} \geq n^{-\frac{k\,|E(H')|}{l}}$, also die obige Abschätzung mindestens von der Ordnung

$$\big(n^k\, p^l\big)^{1 - \frac{|E(H')|}{|E(H)|}} \sim \lambda^{1 - \frac{|E(H')|}{|E(H)|}}.$$

Für $\lambda \geq c > 0$ liefert die Schranke im Lemma keinen kleinen Fehler in der Poisson-Approximation. Wenn so ein H' nicht existiert, ist H strikt balanciert. Wir betrachten eine zweite Variante der Poisson-Approximation für diesen Fall. Es sei H strikt balanciert mit k Ecken und $l \geq 1$ Kanten. Weiter sei

$$\alpha(H) := \min_{H'} \frac{k - |E(H')|}{l - |V(H')|}, \quad H' \subset H,\ |E(H')| > 0,$$

und

$$\gamma(H) := \min\left(\frac{k}{l}|V(H')| - |E(H')|\right) = \min_{H'} |V(H')| \, (\varrho(H) - \varrho(H'))$$

(H ist also strikt balanciert, wenn $\gamma(H) > 0$ oder $\alpha(H) > \varrho(H)$ gilt).

Dann gilt (siehe [BHJ92, Section 5]):

Satz 6.10. *(Barbour, Holst, Janson, 1992) Sei H strikt balanciert mit k Ecken und $l \geq 1$ Kanten. Dann existiert eine Konstante C, die nur von H abhängt, so dass*

$$\begin{aligned} & d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \\ \leq & \begin{cases} C \min\{\lambda^{1-2/k} \, p^{\gamma(H)}, \lambda^{1-1/l} \, n^{-\gamma(H)/\varrho(H)}, n \, p^{\alpha(H)} + p\} & \textit{für } \lambda \geq 1, \\ C \min\{\lambda \, p^{\gamma(H)}, \lambda \, n^{-\gamma(H)/\varrho(H)}, n \, p^{\alpha(H)} + p\} & \textit{für } \lambda < 1. \end{cases} \end{aligned}$$

Beweis: Erinnere $(1 - e^{-\lambda}) \leq \min(1, \lambda)$, so folgt mit Lemma 6.9

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq C \min(1, \lambda) \sum_{H'} n^{|V(H)| - |V(H')|} \, p^{|E(H)| - |E(H')|}.$$

Mit $\lambda \sim n^{|V(H)|} \, p^l = n^k \, p^l$ folgt, mit unterschiedlichen Konstanten C,

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq C \min(1, \lambda) \sum_{H'} \lambda^{1 - |V(H')|/|V(H)|} \, p^{|E(H)| \, |V(H')|/|V(H)| - |E(H)|}$$

$$\leq \begin{cases} C \, \lambda^{1-2/|V(H)|} \, p^{\gamma(H)} & \text{für } \lambda \geq 1, \\ C \, \lambda \, p^{\gamma(H)} & \text{für } \lambda < 1. \end{cases}$$

Dies ist der erste Teil der Abschätzung, den zweiten Teil bekommt man analog. Der dritte Teil ist klar für $n \, p^{\alpha(H)} > 1$. Ist $n \, p^{\alpha(H)} \leq 1$, so betrachten wir

$$\begin{aligned} & d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \\ \leq \ & C\Bigg(\sum_{H' : |V(H')| < |V(H)|} (n \, p^{\alpha(H)})^{|V(H)| - |V(H')|} + \sum_{H' : |V(H')| = |V(H)|} p^{|E(H)| - |E(H')|}\Bigg) \\ \leq \ & C\big(n \, p^{\alpha(H)} + p\big). \end{aligned}$$

□

Beispiel 6.11. Es sei H ein Kreis der Länge v (v Knoten, v Kanten). Dann ist H strikt balanciert und $\lambda = \mathbb{E}W = \binom{n}{v} p^v \frac{(v-1)!}{2}$. Weiter ist $\varrho(H) = 1$, $\alpha(H) \stackrel{!}{=} \frac{v-1}{v-2}$ und $\gamma(H) = 1$, also

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq C\, n^{v-2}\, p^{v-1}.$$

Sei $p(n)$ so, dass $\lim_{n\to\infty} n\, p(n) = c > 0$ mit $c < \infty$. Dann folgt $\lambda \to \lambda_\infty = \frac{1}{2v}\, c^v$ und $W \xrightarrow{\mathcal{D}} Po(\lambda_\infty)$ mit Rate $\mathcal{O}(p) = \mathcal{O}\big(n^{-1}\big)$.

Eine natürliche Frage ist nun, was im Bereich $p(n) \gg n^{-k/l}$ passiert. Es sei $\tilde{W} := \frac{W - \mathbb{E}W}{\sqrt{V(W)}}$. Wir werden später sehen, dass $\tilde{W} \xrightarrow{\mathcal{D}} N(0,1)$, wenn $1 - p \gg n^{-2}$. Die Steinsche Methode wird erneut Konvergenzraten liefern.

Die Steinsche Methode ist bei Zufallsgraphen nicht nur für Subgraph-counting nützlich. Wir betrachten hier noch eine weitere Anwendung. Es sei W_n die Anzahl der *isolierten Knoten* in $G \in \mathcal{G}(n,p)$.

Satz 6.12. *Ist $p = p(n)$ so, dass*

$$\lim_{n\to\infty} n(1-p)^{n-1} = \lambda$$

für ein $0 < \lambda < \infty$, so gilt

$$\lim_{n\to\infty} d_{TV}\big(\mathcal{L}(W_n), Po(\mathbb{E}W_n)\big) = 0.$$

Beweis: Wir wenden Satz 6.8 (Kopplung) an: Es sei $\Gamma = \{1, \ldots, n\}$ und

$$I_\alpha = I(\alpha \text{ ist isoliert}).$$

Dann ist $\mathbb{E}(I_\alpha) = (1-p)^{n-1}$ und $W_n = \sum_{\alpha\in\Gamma} I_\alpha$. Es gilt nach Voraussetzung $\lim_{n\to\infty} \mathbb{E}(W_n) = \lambda$. Setze $U_\alpha = W_n = \sum_{\alpha\in\Gamma} I_\alpha$. V_α bestimmen wir wie folgt: entferne α aus Γ und alle Kanten, die an α hängen. Setze $\Gamma' = \{1, \ldots, n\} \setminus \{\alpha\}$ und setze $I'_\beta = I(\beta \text{ ist isoliert in} \, \mathcal{G}(n-1,p))$. Setze nun $V_\alpha = \sum_{\beta\in\Gamma'} I'_\beta$. Dann ist $I'_\beta \geq I_\beta$, denn jedes isolierte β in $\mathcal{G}(n,p)$, ist auch in $\mathcal{G}(n-1,p)$ isoliert. Also ist $V_\alpha \geq \sum\limits_{\beta\in\Gamma\setminus\{\alpha\}} I_\beta$ und somit $U_\alpha - I_\alpha \leq V_\alpha$ für alle $\alpha \in \Gamma$. Wir können also Satz 6.8 anwenden. Es ist

$$\begin{aligned}
\mathbb{E}(W_n^2) &= \sum_\alpha \mathbb{E}(I_\alpha I_\alpha) + \sum_\alpha \sum_{\beta\neq\alpha} \mathbb{E}(I_\alpha I_\beta) = \sum_\alpha \pi_\alpha + \sum_\alpha \sum_{\beta\neq\alpha} \pi_\alpha \mathbb{E}(I_\beta | I_\alpha = 1) \\
&= \mathbb{E}W_n + \sum_\alpha \sum_{\beta\neq\alpha} \pi_\alpha (1-p)^{n-2} = \mathbb{E}W_n + \mathbb{E}W_n (1-p)^{n-2}(n-1) \\
&\approx \mathbb{E}W_n + (\mathbb{E}W_n)^2 (1-p)^{-1}.
\end{aligned}$$

Somit folgt

$$d_{TV}\big(\mathcal{L}(W_n), Po(\mathbb{E}W_n)\big) \leq$$
$$\min\left(1, \frac{1}{\mathbb{E}(W_n)}\right)\left\{\mathbb{E}W_n + \frac{(\mathbb{E}W_n)^2}{1-p} - (\mathbb{E}W_n)^2 - \mathbb{E}W_n + 2\,\mathbb{E}W_n(1-p)^{n-1}\right\},$$

woraus unsere Behauptung unmittelbar folgt.

□

Für weitere Anwendungen der Steinschen Methode für Zufallsgraphen siehe die Monographien von Barbour, Holst und Janson [BHJ92] und von Janson, Łuczak und Ruciński [JŁR00].

Ausblick: Wir hatten in Kapitel 2 bereits darauf hingewiesen, dass man neben Zufallsgraphen im Sinne von Erdős in jüngster Zeit sogenannte *random geometric graphs* untersucht. Eine sehr schöne Darstellung findet man in der Monographie [Pen03]. Hier eine kurze Andeutung: es sei f eine gegbenen Wahrscheinlichkeitsdichte im $\mathbb{R}^d$ und $X_1, X_2, \ldots$ seien unabhängige und identisch verteilte Zufallsgrößen mit Werten in $\mathbb{R}^d$ und gemeinsamer Dichte f. Es sei $\chi_n := \{X_1, \ldots, X_n\}$. Weiter sei $\|\cdot\|$ eine Norm im $\mathbb{R}^d$ und $r > 0$. Für jede endliche Teilmenge $\chi \subset \mathbb{R}^d$ bezeichnen wir mit $G(\chi, r)$ den ungerichteten Graphen mit Knotenmenge χ und mit der Menge der ungerichteten Kanten, die alle Paare $\{X, Y\}$ mit $\|X - Y\| \leq r$ verbinden. Dann heißt $G(\chi_n, r)$ *zufälliger geometrischer Graph*. Der Hauptunterschied zu den Erdős-Zufallsgraphen ist, dass verschiedene Kanten dort unabhängig oder beinahe unabhängig sind, was im Fall des geometrischen Modells nicht stimmt. Wenn X_i in der Nähe von X_j und X_j in der Nähe von X_k liegt, so liegt X_i in der Regel auch nahe bei X_k. Diese Dreiecksungleichung scheint durchaus realistischer als Modellannahme. Weitere historische Dinge und Motivationen zu dieser Klasse von Zufallsgraphen findet man in [Pen03]. Die Poisson-Approximation wird in [Pen03, Section 2.1, 2.2 und 3.2] untersucht.

7. Die Janson-Ungleichung

In den Kapiteln 3, 4 und 6 haben wir die Verteilung einer Zufallsgröße der Form $W = \sum_{\alpha \in \Gamma} I_\alpha$ untersucht. Häufig ist es von Interesse, "lediglich" die Wahrscheinlichkeit $P(W = 0)$ genau zu approximieren. In Korollar 3.7 hatten wir mit Hilfe des lokalen Ansatzes der Steinschen Methode $|P(W = 0) - e^{-\lambda}|$ abgeschätzt. In Beispiel 3.10 (Geburtstagsproblem) war $P(W = 0)$ untersucht worden. Auch in Beispiel 3.12 (Zuverlässigkeitstheorie) war $P(W = 0)$ von Interesse. Dort war es die Wahrscheinlichkeit, daß ein Prozessor „in Funktion"ist. Ist λ sehr groß und somit $e^{-\lambda}$ sehr klein, hat man mit der Schranke für die Totalvariation Probleme: häufig ist die „beste Ordnung"in der Abschätzung via der Steinschen Methode größer als die (kleinen) zu approximierenden Wahrscheinlichkeiten. Man benötigt andere Methoden, um kleine Wahrscheinlichkeiten gut approximieren zu können. Wir stellen hier die *Janson-Ungleichung* vor, die Janson 1990 in [Jan90] bewies.

In der Vorbereitung hierzu befassen wir uns mit der sogenannten *FKG-Ungleichung*, benannt nach Fortuin, Kasteleyn und Ginibre, siehe [FKG71]. Nach der Formulierung und der Beweisführung der Janson-Ungleichung stellen wir zwei Anwendungen aus dem Bereich der Zufallsgraphen vor. Zur Behandlung des zweiten Beispiels benötigen wir ein klassisches Resultat über Kopplungen von Strassen, [Str65].

Für $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ sei $x \geq y$ durch $x_i \geq y_i$ für alle $1 \leq i \leq n$ definiert. Eine Abbildung $f : \mathbb{R}^n \to \mathbb{R}$ heißt *steigend*, wenn für alle $x \geq y$ gilt: $f(x) \geq f(y)$. Analog heißt f *fallend*, wenn für alle $x \geq y$ gilt: $f(x) \leq f(y)$. Zentral für das Weitere ist nun der folgende Satz (siehe [FKG71] und die wichtige Vorarbeit von Harris [Har60]):

Satz 7.1. *(FKG-Ungleichung, Harris 1960) Es seien $X_1, \ldots, X_n$ unabhängige, reellwertige Zufallsgrößen, definiert auf einen gemeinsamen Wahrscheinlichkeitsraum. Sind f und g beschränkte und steigende (fallende) Funktionen von $\mathbb{R}^n$ nach $\mathbb{R}$, so gilt*

$$\mathbb{E}\big(f(X_1, \ldots, X_n)\, g(X_1, \ldots, X_n)\big) \geq \mathbb{E}(f(X_1, \ldots, X_n))\, \mathbb{E}(g(X_1, \ldots, X_n))$$

Man nennt die Ungleichung auch eine Korrelationsungleichung.

Beweis: Wir betrachten einen Induktionsbeweis. Es sei $X := (X_1, \ldots, X_n)$.

$n = 1$: Es sei X' stochastisch unabhängig von X so gewählt, dass X' in Verteilung mit X übereinstimmt. Es gilt für alle $x, x' \in \mathbb{R}$

$$\big(f(x) - f(x')\big)\,\big(g(x) - g(x')\big) \geq 0.$$

Also folgt

$$0 \leq \mathbb{E}\Big(\big(f(X) - f(X')\big) \, \big(g(X) - g(X')\big) \Big) = 2 \Big(\mathbb{E}\big(f(X)\, g(X)\big) - \mathbb{E}(f(X))\, \mathbb{E}(g(X)) \Big),$$

womit die Aussage des Satzes für $n = 1$ bewiesen ist.

Es sei nun $n > 1$ und die Aussage gelte für alle $1 \leq r \leq n-1$. Nach Definition der bedingten Erwartung gilt:

$$\mathbb{E}\big(f(X)\, g(X)\big) = \mathbb{E}\big(\mathbb{E}(f(X)\, g(X) | X_{1,\ldots,} X_{n-1})\big).$$

Ist $X_{1,\ldots,} X_{n-1}$ gegeben, so ist nur X_n zufällig. f und g sind in der letzten Komponente X_n bei festgehaltenem $X_{1,\ldots,} X_{n-1}$ steigende Funktionen. Damit ist

$$\mathbb{E}\big(f(X)\, g(X) | X_1, \ldots, X_{n-1}\big) \geq \mathbb{E}\big(f(X) | X_1, \ldots, X_{n-1}\big)\, \mathbb{E}\big(g(X) | X_1, \ldots, X_{n-1}\big).$$

Dies folgt aus dem Resultat für $n = 1$. Es sei

$$f_1(x_1, \ldots, x_{n-1}) := \mathbb{E}\big(f(x_1, \ldots, x_{n-1}, X_n)\big)$$

und

$$g_1(x_1, \ldots, x_{n-1}) := \mathbb{E}\big(g(x_1, \ldots, x_{n-1}, X_n)\big).$$

Dann ist $f_1(X_1, \ldots, X_{n-1}) = \mathbb{E}\big(f(X) | X_1, \ldots, X_{n-1}\big)$ und analog $g_1(X_1, \ldots, X_{n-1}) = \mathbb{E}\big(g(X) | X_1, \ldots, X_{n-1}\big)$. Da f_1 und g_1 steigend sind, folgt

$$\begin{aligned}
\mathbb{E}\big(f(X)\, g(X)\big) &\geq \mathbb{E}\Big(\mathbb{E}\big(f(X) | X_1, \ldots, X_{n-1}\big)\, \mathbb{E}\big(g(X) | X_1, \ldots, X_{n-1}\big) \Big) \\
&= \mathbb{E}\big(f_1(X_1, \ldots, X_{n-1})\, g(X_1, \ldots, X_{n-1})\big) \\
&\geq \mathbb{E}\big(f_1(X_1, \ldots, X_{n-1})\big)\, \mathbb{E}\big(g(X_1, \ldots, X_{n-1})\big) \\
&= \mathbb{E}\Big(\mathbb{E}\big(f(X) | X_1, \ldots, X_{n-1}\big) \Big)\, \mathbb{E}\Big(\mathbb{E}\big(g(X) | X_1, \ldots, X_{n-1}\big) \Big) \\
&= \mathbb{E}(f(X))\, \mathbb{E}(g(X)).
\end{aligned}$$

Damit ist der Satz bewiesen. □

Wir betrachten nun die folgende Situation: Es sei $(J_i, i \in Q)$ eine Familie unabhängiger Indikatorvariablen und $(Q(\alpha), \alpha \in \Gamma)$ eine endliche Familie von Teilmengen der Indexmenge Q. Wir setzen $I_\alpha := \prod_{i \in Q(\alpha)} J_i$ und $W := \sum_{\alpha \in \Gamma} I_\alpha$. W bestimmt die Anzahl der Mengen $Q(\alpha)$, die in der Zufallsmenge $\{i \in Q : J_i = 1\}$ enthalten sind. Die I_α sind steigende Funktionen von $(J_i, i \in Q)$. Sei weiter

$$\Gamma_\alpha^+ := \{\beta \neq \alpha : Q(\alpha) \cap Q(\beta) \neq \emptyset\}.$$

Dann ist I_α unabhängig von I_β für $\beta \in \Gamma \setminus \Gamma_\alpha^+$ und $\beta \neq \alpha$. Wir definieren

$$\pi_\alpha = \mathbb{E} I_\alpha, \lambda = \mathbb{E} W = \sum_{\alpha\in\Gamma} \pi_\alpha,\ \delta = \frac{1}{\lambda}\Delta$$

mit $\Delta = \sum_\alpha \sum_{\beta\in\Gamma_\alpha^+} \mathbb{E}(I_\alpha\, I_\beta)$ sowie $\eta = \max_\alpha \pi_\alpha$. Es gilt der Satz, die *Ungleichung von Janson* [Jan90]:

Satz 7.2. *(Janson, 1990) In der obigen Notation gilt immer:*

$$\exp\Big(-\frac{\lambda}{1-\eta}\Big) \leq P(W=0) \leq \exp\Big(-\frac{\lambda}{1+\delta}\Big).$$

Ist $\lambda\,\eta \leq 1$ *und* $\lambda\,\delta \leq 1$ *so gilt*

$$P(W=0) = e^{-\lambda}\big(1 + \mathcal{O}(\lambda\,\eta + \lambda\,\delta)\big).$$

Bemerkung 7.3. Die Janson-Ungleichung liefert eine gute Annäherung an die Poisson-Wahrscheinlichkeit $e^{-\lambda}$, wenn $\lambda\,\eta$ und $\lambda\,\delta$ klein sind.

Beweis: 1. Schritt: Es gilt

$$P(W=0) = P(\text{alle } I_\alpha = 0) \geq \prod_{\alpha\in\Gamma}(1-\pi_\alpha) = \prod_{\alpha\in\Gamma} P(I_\alpha = 0).$$

Per Induktion folgt: ist $|\Gamma| = 2$, so gilt mit der FKG-Ungleichung

$$\begin{aligned} P(I_{\alpha_1} = I_{\alpha_2} = 0) &= \mathbb{E}\big((-(1-I_{\alpha_1}))(-(1-I_{\alpha_2}))\big) \\ &\geq \mathbb{E}(1-I_{\alpha_1})\mathbb{E}(1-I_{\alpha_2}) = (1-\pi_{\alpha_1})\,(1-\pi_{\alpha_2}). \end{aligned}$$

Ist $\Gamma = \Gamma_1 \cup \Gamma_2$ mit $\Gamma_1 \cap \Gamma_2 = \emptyset$, so folgt mit der FKG-Ungleichung

$$\begin{aligned} P\Big(\bigcap_{\alpha\in\Gamma}\{I_\alpha = 0\}\Big) &= \mathbb{E}\Big(\big(-\prod_{\alpha\in\Gamma_1}(1-I_\alpha)\big)\,\big(-\prod_{\alpha\in\Gamma_2}(1-I_\alpha)\big)\Big) \\ &\geq \mathbb{P}\Big(\bigcap_{\alpha\in\Gamma_1}\{I_\alpha = 0\}\Big)\,P\Big(\bigcap_{\alpha\in\Gamma_2}\{I_\alpha = 0\}\Big) \end{aligned}$$

(hier benötigt man $|\Gamma_1| \geq 1$, $|\Gamma_2| \geq 1$). Da $|\Gamma_1| < |\Gamma|$ und $|\Gamma_2| < |\Gamma|$, liefert Induktion das Resultat. Wir erhalten somit

$$\begin{aligned} P(W=0) &\geq \exp\Big(\sum_{\alpha\in\Gamma}\log(1-\pi_\alpha)\Big) \\ &= \exp\Big(\sum_{\alpha\in\Gamma}(-\pi_\alpha - \frac{\pi_\alpha^2}{2} - \cdots)\Big) \geq \exp\Big(-\sum_{\alpha\in\Gamma}\frac{\pi_\alpha}{1-\pi_\alpha}\Big), \end{aligned}$$

denn $\frac{\pi_\alpha}{1-\pi_\alpha} = \pi_\alpha + \pi_\alpha^2 + \pi_\alpha^3 + \cdots$. Nach Definition von η folgt

$$P(W=0) \geq \exp\Big(-\sum_{\alpha\in\Gamma} \frac{\pi_\alpha}{1-\eta}\Big) = \exp\Big(-\frac{\lambda}{1-\eta}\Big).$$

Also ist die untere Abschätzung gezeigt.

2. Schritt: Es gilt immer

$$P(W=0) \leq \psi(t) := \mathbb{E}\big(\exp(-t\,W)\big),\ t \geq 0.$$

Die Funktion $\psi(t)$ heißt die *Laplace-Transformierte* von W. Wir wollen $\psi(t)$ abschätzen. Es gilt

$$-\frac{d}{dt}\,\psi(t) = \mathbb{E}\Big(W\,e^{-t\,W}\Big) = \sum_\alpha \mathbb{E}\Big(I_\alpha\,e^{-t\,W}\Big).$$

Für jedes α schreiben wir $W = I_\alpha + Z_\alpha + W'_\alpha$ (siehe Kapitel 3),wobei

$$Z_\alpha = \sum_{\beta\in\Gamma_\alpha^+} I_\beta \text{ und } W'_\alpha = W - I_\alpha - Z_\alpha.$$

Damit ist W'_α unabhängig von I_α. Es gilt

$$\mathbb{E}\Big(I_\alpha\,e^{-t\,W}\Big) = \pi_\alpha\,\mathbb{E}\Big(e^{-t\,Z'_\alpha - t\,W'_\alpha}|I_\alpha = 1\Big)$$

mit $Z'_\alpha := Z_\alpha + I_\alpha$. Die Bedingung $I_\alpha = 1$ legt die $(J_i, i \in Q(\alpha))$ fest. Z'_α und W'_α sind steigende Funktionen der verbleibenden $(J_i, i \notin Q(\alpha))$. Also sind $e^{-tZ'_\alpha}$ und $e^{-tW'_\alpha}$ fallende Funktionen dieser unabhängigen Zufallsgrößen. Satz 7.1 (FKG-Ungleichung) liefert

$$\begin{aligned}
\mathbb{E}\Big(I_\alpha\,e^{-t\,W}\Big) &\geq \pi_\alpha\mathbb{E}\Big(e^{-t\,Z'_\alpha}|I_\alpha = 1\Big)\mathbb{E}\Big(e^{-t\,W'_\alpha}|I_\alpha = 1\Big) \\
&= \pi_\alpha\,\mathbb{E}\Big(e^{-t\,Z'_\alpha}|I_\alpha = 1\Big)\mathbb{E}\Big(e^{-t\,W'_\alpha}\Big) \\
&\geq \pi_\alpha\mathbb{E}\Big(e^{-t\,Z'_\alpha}|I_\alpha = 1\Big)\,\psi(t),
\end{aligned}$$

denn $W \geq W'_\alpha$, also $e^{-tW'_\alpha} \geq e^{-tW}$. Wir erhalten

$$\begin{aligned}
-\frac{d}{dt}\big(\log\psi(t)\big) &= \frac{1}{\psi(t)}\sum_\alpha \mathbb{E}\Big(I_\alpha\,e^{-t\,W}\Big) \\
&\geq \sum_\alpha \pi_\alpha\,\mathbb{E}\Big(e^{-t\,Z'_\alpha}|I_\alpha = 1\Big).
\end{aligned}$$

Nun verwenden wir die Jensen-Ungleichung: ist $f(x)$ eine konvexe Funktion, so gilt $\mathbb{E}(f(X)) \geq f(\mathbb{E}(X))$ (siehe zum Beispiel die Monographie von Bauer, [Bau02, Satz 3.9]). Wir wenden Jensen für $f(x) = e^{-t\,x}$, $t > 0$, an. Es gilt dann

$$\begin{aligned}
-\frac{d}{dt}\big(\log\psi(t)\big) &\overset{Jensen}{\geq} \sum_\alpha \pi_\alpha \exp\big(-\mathbb{E}(t\,Z'_\alpha|I_\alpha = 1)\big) \\
&= \lambda\sum_\alpha \frac{\pi_\alpha}{\lambda} \exp\big(-t\,\mathbb{E}(Z'_\alpha|I_\alpha = 1)\big) \\
&\overset{Jensen}{\geq} \lambda\exp\Bigg(-\sum_\alpha \frac{\pi_\alpha}{\lambda}\, t\,\mathbb{E}\big(Z'_\alpha|I_\alpha = 1\big)\Bigg) \\
&= \lambda\exp\Bigg(-\frac{t}{\lambda}\sum_\alpha \mathbb{E}(I_\alpha\,Z'_\alpha)\Bigg) \\
&= \lambda\exp\Bigg(-\frac{t}{\lambda}\Bigg(\sum_\alpha \mathbb{E}I_\alpha^2 + \sum_\alpha\sum_{\beta\in\Gamma_\alpha^+}\mathbb{E}(I_\alpha\,I_\beta)\Bigg)\Bigg) \\
&= \lambda\exp\Bigg(-\frac{t}{\lambda}(\lambda+\Delta)\Bigg) = \lambda\exp(-t(1+\delta)).
\end{aligned}$$

Integration und $\psi(0) = 1$ liefern

$$\begin{aligned}
-\log\psi(T) + 0 &\geq \int_0^T \lambda\,e^{-t(1+\delta)}\,dt \\
&= \frac{\lambda}{1+\delta}\Bigg(1 - e^{-T(1+\delta)}\Bigg)
\end{aligned}$$

und somit

$$P(W=0) \leq \psi(T) \leq \exp\Bigg(-\frac{\lambda}{1+\delta}\big(1 - e^{-T(1+\delta)}\big)\Bigg)$$

für alle $T \geq 0$. Wir betrachten nun $T \to \infty$:

$$P(W=0) \leq \exp\Bigg(-\frac{\lambda}{1+\delta}\Bigg).$$

3. Schritt: Es gilt

$$e^\lambda\,e^{-\frac{\lambda}{1+\delta}} = e^{\frac{\lambda\delta}{1+\delta}} \leq e^{\lambda\delta} = 1 + \mathcal{O}(\lambda\delta)$$

für $\lambda\,\delta \leq 1$ und

$$e^\lambda\,e^{-\frac{\lambda}{1-\eta}} = e^{-\frac{\lambda\eta}{1-\eta}} \geq 1 - \frac{\lambda\eta}{1-\eta} = 1 + \mathcal{O}(\lambda\,\eta)$$

für $\lambda\,\eta \leq 1$. □

Anwendungen: Wir betrachten zunächst das Beispiel des Teilgraphenzählens bei Zufallsgraphen. $X(G)$ sei die Anzahl der zu H isomorphen Teilgraphen in $G \in$

$\mathcal{G}(n,p)$. H sei strikt balanciert mit k Knoten und $l \geq 1$ Kanten, Γ wie in Kapitel 6. Wenn $J_1, \dots, J_{\binom{n}{2}}$ die Kantenindikatoren bezeichnen und

$$Q(\alpha) := \{\text{Kanten in der Kopie } \alpha \text{ von } H\},$$

so ist $I_\alpha := \prod_{i \in Q(\alpha)} J_i$ die vertraute Zufallsgröße. Wir betrachten den kritischen Fall $\lim_{n\to\infty} p(n)\, n^{k/l} = c > 0$. Wir wissen bereits, dass

$$\lambda = \mathbb{E}W = \mathbb{E}\Big(\sum_{\alpha\in\Gamma} I_\alpha\Big) \sim \frac{c^l}{a}$$

gilt. Weiter ist $\Delta = \mathcal{O}(p^\varepsilon)$, wobei ε der Balanceindex von H ist. Weiter ist $\eta = p^l$. Also ist $\lambda\,\delta = \Delta = \mathcal{O}(p^\varepsilon)$ klein und $\lambda\,\eta = \lambda\, p^l = \mathcal{O}(p^l)$ ebenfalls klein. Die Janson-Ungleichung, Satz 7.2, liefert dann

$$P(X=0) = P(W=0) = e^{-\lambda}\,(1 + \mathcal{O}(p^\varepsilon)),$$

denn $p^l < p^\varepsilon$ nach Definition von ε. Wir erhalten $P(X=0) = \exp\big(-c^l/a\big)(1+\mathcal{O}(p^\varepsilon))$. Dies ist vergleichbar mit den Resultaten mittels der Steinschen Methode (Satz 6.3).

Wir wollen nun die Situation untersuchen, in der $\lambda \to \infty$. Wir wissen $\lambda \sim \frac{n^k p^l}{a}$. Wir wählen $p(n) = n^{-k/l+\gamma}$ mit $\gamma > 0$. Dann ist $\lambda \asymp n^{\gamma\, l}$, also wächst λ schnell. Es folgt, dass $\eta = p^l = n^{-k+\gamma\, l}$ und somit $\lambda\,\eta \asymp n^{-k+2\gamma\, l}$. Die rechte Seite ist klein, wenn $\gamma < \frac{k}{2l}$ gilt. Weiter ist

$$\Delta = \lambda\,\delta = \mathcal{O}\big((n^k\, p^l)^{2-2/k}\, p^\varepsilon\big) = \mathcal{O}\big(n^{\gamma\, l (2-2/k)}\, n^{-\frac{\varepsilon\, k}{l}+\gamma\,\varepsilon}\big).$$

Nach Satz 7.2 impliziert Δ und $\lambda\,\eta$ klein:

$$P(X=0) \sim e^{-\lambda}.$$

Wir wollen den zugehörigen γ-Bereich bestimmen:

(i) $\lambda\,\eta$ klein benötigt $\gamma < \frac{k}{2l}$.
(ii) Δ klein benötigt $\gamma\big(l\,(2-\frac{2}{k}) + \varepsilon\big) < \frac{\varepsilon\, k}{l}$, also $\gamma < \frac{\varepsilon\, k}{l\big(l\,(2-\frac{2}{k})+\varepsilon\big)}$.

In aller Regel ist die zweite Schranke für γ maßgebend, sprich kleiner.

Bemerkung 7.4. Ist $\delta \leq 1$, so liefert Satz 7.2:

$$\begin{aligned} P(W=0) &\leq e^{-\lambda}\, \exp\Big(-\lambda\big(\frac{1}{1+\delta} - 1\big)\Big) \\ &= e^{-\lambda}\, \exp\Big(\frac{\lambda\,\delta}{1+\delta}\Big) \leq e^{-\lambda}\, e^{\lambda\,\delta}. \end{aligned}$$

Ist $\delta \geq 1$, so folgt

$$P(W=0) \leq \exp\Big(-\frac{\lambda}{2\delta}\Big) = \exp\Big(-\frac{\lambda^2}{2\Delta}\Big).$$

Wir wollen diese Beobachtung für das obige Beispiel diskutieren: δ klein bedeutet $\frac{\Delta}{\lambda}$ klein oder $(n^k\, p^l)^{1-\frac{2}{k}}\, p^\varepsilon$ klein bzw. $n^{\gamma\, l\,(1-\frac{2}{k})}\, n^{-\frac{\varepsilon\, k}{l}+\varepsilon\,\gamma}$ klein. Es genügt

$$\gamma\Big(l\,\big(1-\frac{2}{k}\big)+\varepsilon\Big) < \frac{\varepsilon\, k}{l},$$

also

$$\gamma < \frac{\varepsilon\, k}{l\Big(l(1-\frac{2}{k})+\varepsilon\Big)}.$$

Wir vergleichen dies mit den Resultaten, die die Steinschen Methode lieferte:

$$e^{-\lambda} \sim e^{-n^{\gamma\, l}\, const.}.$$

Für $\gamma > 0$ ist die rechte Seite, also $e^{-\lambda}$, sehr klein. Stein lieferte

$$p^\varepsilon \asymp n^{-\left(\frac{k}{l}-\gamma\right)\,\varepsilon},$$

was viel größer ist.

Im Fall $\delta \geq 1$ betrachte

$$\frac{2\Delta}{\lambda^2} = \mathcal{O}\big((n^k\, p^l)^{-2/k}\, p^\varepsilon\big),$$

da $n^k\, p^l \sim n^{\gamma\, l}$ und $p^\varepsilon \sim n^{-\frac{k\,\varepsilon}{l}+\varepsilon\,\gamma}$, folgt hier

$$P(W=0) \leq e^{-\frac{\lambda^2}{2\Delta}} \sim \exp(-n^{\frac{2\gamma l}{k}+\frac{k\varepsilon}{l}-\varepsilon\gamma}).$$

Vergleiche dies mit

$$\begin{aligned}(1-p)^{\binom{n}{2}} &\sim \exp(-\frac{pn(n-1)}{2}) \sim \exp\Big(-\frac{p\, n^2}{2}\Big)\\ &\sim \exp\Big(-\frac{n^2}{2}\, n^{-k/l+\gamma}\Big) = \exp\Big(-\frac{1}{2} n^{2-\frac{k}{l}+\gamma}\Big).\end{aligned}$$

Der Exponent ist größer als $n^{2\gamma\, l/k+k\,\varepsilon/l-\varepsilon\,\gamma}$.

Wir betrachten eine weitere Anwendung der Janson Ungleichung. Wir gehen der folgenden Frage nach:

Wie groß ist die Wahrscheinlichkeit des Ereignisses, dass in einem $G \in \mathcal{G}(n,p)$ jeder Eckpunkt zu einem Dreieck gehört?

Es sei $\Gamma = \{1, \dots, n\}$ und $J_1, \dots, J_{\binom{n}{2}}$ seien wieder die Kantenindikatoren. Wir schreiben J_{ij}, wenn der Kantenindikator zu den Knoten i und j gemeint ist. Zu $i \in \{1, \dots, n\}$ sei $I_i = I(\text{Eckpunkt } i \text{ gehört zu keinem Dreieck})$. Formal bedeutet dies mit $J_{ijk} := J_{ij}J_{jk}J_{ik}$:

$$I_i = \prod_{\substack{j<k \\ j,k\neq i}} (1 - J_{ijk}).$$

Sei weiter $W = \sum_{i=1}^n I_i$, dann ist $\{W = 0\}$ das Ereignis, dass jeder Eckpunkt zu einem Dreieck gehört.

Wir versuchen, die folgende Frage zu beantworten: Gilt $P(W = 0) \sim e^{-\lambda}$ mit $\lambda = \mathbb{E}W = n\,\tilde{p}$, wobei $\tilde{p} = P(I_i = 1)$? Erneut versuchen wir mittels einer Kopplung die Steinsche Methode anzuwenden.

Zu $j \in \{1, \dots, n\}$ seien U_j, V_j so gewählt, dass $\mathcal{L}(U_j) = \mathcal{L}(W)$ und $\mathcal{L}(V_j + 1) = \mathcal{L}(W|I_j = 1)$ gilt. In unserer Situation sind die I_j fallende Funktionen der unabhängigen Kantenindikatoren $J_1, \dots, J_{\binom{n}{2}}$. Sei $\phi: \mathbb{R} \to \mathbb{R}$ eine steigende Funktion. Dann ist

$$\begin{aligned}
\mathbb{E}(\phi(V_j)) &= \mathbb{E}(\phi(W - I_j)|I_j = 1) \\
&= \mathbb{E}\Big(\phi\Big(\sum_{k\neq j} I_k\Big)\, I_j\Big)/\mathbb{E}(I_j).
\end{aligned}$$

Dabei ist $\phi(\sum_{k\neq j} I_k)$ ebenfalls eine fallende Funktion der unabhängigen Kantenindika-toren. Satz 7.1, die FKG-Ungleichung, liefert $\mathbb{E}(\phi(V_j)) \geq \mathbb{E}(\phi(W - I_j))$ für jede steigende Funktion ϕ. Dann existiert ein Wahrscheinlichkeitsmaß auf dem Produkt der Räume, auf denen V_j und $W - I_j$ definiert sind, mit Marginalverteilungen $\mathcal{L}(V_j)$, $\mathcal{L}(W - I_j)$ und mit

$$\mu(V_j \geq W - I_j) = 1.$$

Also ist fast sicher $V_j \geq U_j - I_j$ erfüllt und somit können wir Satz 6.8 anwenden:

$$d_{TV}\big(\mathcal{L}(W), Po(\mathbb{E}W)\big) \leq \min\Big(1, \frac{1}{\mathbb{E}W}\Big) \Big(V(W) - \mathbb{E}W + 2\sum_{j=1}^n (\mathbb{E}I_j)^2\Big).$$

Insbesondere ist dies eine Schranke für $|P(W = 0) - e^{-\lambda}|$. Die Existenz von μ entnehmen wir einem Resultat von Strassen, siehe [Str65]:

Satz 7.5. *(Strassen, 1965) Eine Kopplung μ wie beschrieben existiert genau dann, wenn für jede steigende Funktion $\phi: \mathbb{R} \to \mathbb{R}$*

$$\mathbb{E}(\phi(V_j)) \geq \mathbb{E}(\phi(W - I_j))$$

erfüllt ist (wir geben den Beweis hier nicht).

Wir müssen $\mathbb{E}W$ und $V(W)$ bestimmen. Dabei hilft die Janson-Ungleichung! Wir bestimmen $P(I_j = 1)$ und $P(I_j = I_k = 1)$.

Es gilt $I_1 = 1$ genau dann, wenn $J_{1jk} = 0$ für alle $2 \leq j < k \leq n$. Also ist $P(I_1 = 1) = P(W_1 = 0)$ mit

$$W_1 := \sum_{2 \leq j < k \leq n} J_{1jk}$$

(erinnere: $J_{1jk} = J_{1j} J_{jk} J_{1k}$). Wir wenden die Janson-Ungleichung für W_1 an. Es ist $\eta_1 := \max \mathbb{E}J_{1jk} = p^3$, $\lambda_1 := |\Gamma_1| p^3 = \binom{n-1}{2} p^3$ und

$$\Delta_1 = \sum_{\Gamma_1} \sum_{(j',k'):|\{j,k\} \cap \{j',k'\}|=1} \mathbb{E}(J_{1kj}\, J_{1j'k'}).$$

Eine kleine Überlegung zeigt $\Delta_1 = \binom{n-1}{2} p^5\, 2(n-3)$. Wir betrachten den Fall, dass $P(I_1 = 1) \sim \frac{c}{n}$, also $\lambda \sim c$ wird. Wenn wir $p = p(n)$ so wählen, dass $\binom{n-1}{2} p^3 = \log n - \log c$, also $e^{-\lambda_1} = \frac{c}{n}$, so ist

$$\Delta_1 \asymp n^3 p^5 \asymp n^3 (\log n)^{5/3} n^{-10/3} = n^{-1/3} (\log n)^{5/3} =: \varepsilon_n$$

(denn $p \asymp \big(\log n / (n^2)\big)^{1/3} \sim n^{-2/3}$, vgl. Schwelle $\frac{1}{n}$). Es ist also Δ_1 klein und $\lambda_1 \eta_1 = \mathcal{O}(n^{-2} (\log n)^2)$ ebenfalls klein. Also ist nach Satz 7.2

$$P(W_1 = 0) = \mathbb{E}I_1 = e^{-\lambda_1} \big(1 + \mathcal{O}(\varepsilon_n)\big) = \frac{c}{n} \big(1 + \mathcal{O}(\varepsilon_n)\big).$$

Somit ist $\lambda = \mathbb{E}W = n\, P(I_1 = 1) = c\big(1 + \mathcal{O}(\varepsilon_n)\big)$.

Wir betrachten nun $q := P(I_1 = I_2 = 1)$ (weder 1 noch 2 gehören zu einem Dreieck). Es sei

$$W_2 = \sum_{3 \leq j \leq k \leq n} (J_{1jk} + J_{2jk}) + \sum_{3 \leq j \leq n} J_{12j},$$

dann ist $P(I_1 = I_2 = 1) = P(W_2 = 0)$. Erneut wenden wir die Janson-Ungleichung an. Es ist $\eta_2 = p^3 = \mathcal{O}(n^{-2} \log n)$, also klein, und $\lambda_2 = \mathbb{E}W_2 = p^3 \left(2\binom{n-2}{2} + (n-2)\right)$ sowie

$$\Delta_2 = \left(2\binom{n-2}{2}(2(n-3)+1) + 3(n-2)(n-3)\right) p^5 = \mathcal{O}(n^3 p^5) = \mathcal{O}(n\, p^2\, \lambda_1).$$

Wir erhalten somit $\Delta_2 = \mathcal{O}\big((\log n)\, n^{-1/3}\, (\log n)^{2/3}\big) = \mathcal{O}\big(n^{-1/3}\, (\log n)^{5/3}\big)$. Da also sowohl Δ_2 als auch $\lambda_2\, \eta_2$ klein sind, liefert Satz 7.2

$$\begin{aligned} P(W_2 = 0) &= e^{-\lambda_2} \big(1 + \mathcal{O}(\varepsilon_n)\big) \\ &= \left(\frac{c}{n}\right)^2 \big(1 + \mathcal{O}(\varepsilon_n)\big). \end{aligned}$$

Somit fogt für die Varianz

$$V(W) = n\,\tilde{p}(1-\tilde{p}) + n(n-1)(q-p^2) = \lambda + \mathcal{O}(\varepsilon_n),$$

und somit

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \leq \mathcal{O}(\varepsilon_n).$$

Insbesondere ist

$$|P(W=0) - e^{-\mathbb{E}W}| = \mathcal{O}(\varepsilon_n).$$

In anderer Formulierung erhalten wir also insgesamt

$$P(W=0) = e^{-c} + \mathcal{O}\bigg(\big(\log n^5/n\big)^{1/3}\bigg).$$

Ausblick: Einige Kommentare zur Janson-Ungleichung: Zunächst kann man feststellen, dass im Falle unabhängiger Zufallsvariablen die Schranke der Janson-Ungleichung auch auf direktem Wege (Tailabschätzungen für eine Binomial-verteilte Zufallsgröße) erhalten werden kann: siehe etwa [JŁR00, Theorem 2.1]. Der Nachteil der Janson-Ungleichung ist, dass sie nur im Fall einer Summe von Indikator-Variablen mit einer sehr speziellen Struktur angewendet werden kann. Darunter fallen die Zählstatistiken bei Zufallsgraphen. Suen hat in [Sue90] eine wesentlich größere Klasse von Indikatorvariablen zugelassen und dabei eine nur leicht schwächere Ungleichung herleiten können. Janson wiederum hat in [Jan98] neue Versionen der Ungleichungen von Suen hergeleitet. Man findet dies auch in Spencer [Spe98], sowie in der Monographie [JŁR00, Section 2.3].

8. Untere Abschätzungen bei Poisson-Approximationen

Wie gut sind die betrachteten oberen Schranken, die wir mittels der Steinschen Methode gefunden haben? Haben wir nicht eventuell in den Beispielen schlechte Kopplungen (U_α, V_α) gefunden? Häufig war bisher die obere Schranke im wesentlichen von der Ordnung

$$\varepsilon := \big(1 \wedge \mathbb{E}W\big)\left|\frac{V(W)}{\mathbb{E}W} - 1\right|$$

(siehe Kopplungsansatz, Korollar 4.3 und Satz 6.8). Wir wollen in diesem Kapitel kurz auf untere Schranken eingehen.

In der Literatur kann man untere Schranken für eine recht große Klasse von Indikatorvariablen $(I_\alpha, \alpha \in \Gamma)$ finden. In diesen hier nicht betrachteten Beispielen findet man eine Konstante c, so dass

$$d_{TV}\big(\mathcal{L}(W), Po(\mathbb{E}W)\big) \geq c\,\varepsilon\,|\log\varepsilon|^{-2}$$

gilt. Dieses allgemeine Resultat wollen wir nicht beweisen. Man findet es in Section 3.1 im Buch von Barbour, Holst, und Janson, [BHJ92].

In der Situation von Korollar 4.3 (monotones Koppeln) ist dann ε die korrekte Größenordnung, abgesehen vom $|\log\varepsilon|^{-2}$ Faktor. Poissonkonvergenz liegt in dieser Situation genau dann vor, wenn $\varepsilon \to 0$ gilt.

Wir wollen für die Fälle der monotonen Kopplung genauere untere Schranken herleiten.

Für die Totalvariation zwischen P und Q, zwei Wahrscheinlichkeiten auf $\mathbb{N}_0$, gilt

$$\begin{aligned} d_{TV}(P,Q) &= \sup_{0\leq f\leq 1}\left|\int f\,dP - \int f\,dQ\right| \\ &= \frac{1}{2}\sup_{|f|\leq 1}\left|\int f\,dP - \int f\,dQ\right|. \end{aligned}$$

Hier bezeichnet $\int f\,dP$ den Erwartungswert von f bezüglich der Verteilung P.

Es sei $g:\mathbb{N}_0 \to \mathbb{R}$ eine beschränkte Funktion und $Z \sim Po(\lambda)$. Erinnere:

$$\mathbb{E}\{\lambda g(Z+1) - Z\,g(Z)\} = 0.$$

Setzt man nun

$$h(j) := \lambda g(j+1) - j\,g(j), \;\; j \in \mathbb{N},$$

so folgt

$$\begin{aligned}|\mathbb{E}(\lambda g(W+1) - W g(W))| &= |\mathbb{E}(h(W) - h(Z))| \\ &\leq 2\, d_{TV}\big(\mathcal{L}(W), \mathcal{L}(Z)\big)\, ||h||_\infty \\ &= 2\, d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \sup_{j\geq 0} |\lambda g(j+1) - j\, g(j)|.\end{aligned}$$

Somit folgt

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \geq \frac{|\mathbb{E}(\lambda g(W+1) - W g(W))|}{2 \sup_{j\geq 0} |\lambda g(j+1) - j\, g(j)|}$$

für jede beschränkte Funktion $g : \mathbb{N}_0 \to \mathbb{R}$. Es bezeichne $\kappa_4(W) := \mathbb{E}(W^4) - 3(\mathbb{E}(W^2))^2$. Dann gilt (siehe [BHJ92, Section 3]):

Satz 8.1. *Es seien $(I_\alpha, \alpha \in \Gamma)$ negativ zusammenhängend (siehe Seite 37). Bezeichne $\lambda := \mathbb{E}W, \varepsilon_- = 1 - \frac{V(W)}{\mathbb{E}W}$ und $\gamma_- = 1 - \frac{\kappa_4(W)}{\mathbb{E}W}$. Dann gilt:*

$$(1 - e^{-\lambda})\, \varepsilon_- \geq d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \geq \frac{\varepsilon_-}{11 + 3\, \max(0, \gamma - /\lambda\, \varepsilon_-)}.$$

Beweis: Die obere Schranke folgt nach Korollar 4.3. Nun betrachten wir eine spezielle beschränkte Funktion g: Es sei $f(z) = (z - \lambda)\, e^{-(z-\lambda)^2/(\theta\,\lambda)}$ mit $\theta, \lambda > 0$. Setze nun $g(j) := f(j)$ für jedes $j = 0, 1, 2, \ldots$. Es gilt:

Lemma 8.2. *Für $j = 0, 1, 2, \ldots$ und $y \geq x$ gilt*

(i) $-2e^{-3/2} \leq f(j+1) - f(j) \leq 1$
(ii) $f(y+\lambda) - f(x+\lambda) \geq y - x - \frac{y^3 - x^3}{\theta\,\lambda}$
(iii) $|\lambda\, f(j+1) - j\, f(j)| \leq \lambda \max\big(1, 2e^{-3/2} + \theta e^{-1}\big)$.

Wir beweisen das Lemma etwas später und fahren zunächst im Beweis des Satzes fort:

Es gilt mit Hilfe von (iii) des Lemmas 8.2:

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \geq \frac{|\mathbb{E}(\lambda g(W+1) - W g(W))|}{2\lambda(2\, e^{-3/2} + \theta\, e^{-1})}$$

für alle $\theta \geq e$. Wir verwenden die Notation $V_\alpha := \sum_{\beta \neq \alpha} J_{\beta\alpha}$ mit den $J_{\beta\alpha}$ wie in Kapitel 4. Dann gilt

$$\begin{aligned}
\mathbb{E}(\lambda g(W+1) - Wg(W)) &= \sum_{\alpha \in \Gamma} \pi_\alpha \left(\mathbb{E}\, g(W+1) - \mathbb{E}(g(W)|I_{\alpha=1})\right) \\
&= \sum_{\alpha\in\Gamma} \pi_\alpha \mathbb{E}(g(W+1) - g(V_{\alpha+1})) \\
&\overset{(W \geq V_\alpha)}{\geq} \sum_{\alpha\in\Gamma} \pi_\alpha \mathbb{E}\left((W+1-\lambda) - (V_\alpha + 1 - \lambda) - \frac{(W+1-\lambda)^3 - (V_\alpha+1-\lambda)^3}{\theta\,\lambda}\right) \\
&= \lambda - \sum_{\alpha\in\Gamma} \mathbb{E}\big(I_\alpha(W-\lambda)\big) - \frac{1}{\theta\,\lambda}\left(\lambda\,\mathbb{E}(W+1-\lambda)^3 - \sum_{\alpha\in\Gamma}\mathbb{E}\big(I_\alpha(W-\lambda)^3\big)\right) \\
&= \lambda - V(W) - \frac{1}{\theta\,\lambda}\Big\{\lambda\mathbb{E}(W+1-\lambda)^3 - \mathbb{E}(W(W-\lambda)^3)\Big\} \\
&= \lambda - V(W) - \frac{1}{\theta\,\lambda}\Big\{\lambda - \Big(\mathbb{E}(W-\lambda)^4 - 3\,V(W)^2\Big) + 3(\lambda - V(W))V(W)\Big\} \\
&= \lambda\,\varepsilon_- \left(1 - \frac{\gamma_-/\lambda\varepsilon_- + 3(1-\varepsilon_-)}{\theta}\right).
\end{aligned}$$

Damit folgt

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \geq \varepsilon_- \frac{1 + \frac{3\varepsilon_-}{\theta} - \frac{3+\gamma_-/\lambda\varepsilon_-}{\theta}}{2(2e^{-3/2} + \theta\, e^{-1})}$$

für $\theta \geq e$. Wähle nun $\theta = 6 + 2\,\max(0, \gamma_-/(\lambda\,\varepsilon_-))$. Man erhält nach ein paar Rechenschritten das Resultat. □

Beweis von Lemma 8.2: Der erste Teil folgt mit

$$-2e^{-3/2} \leq \frac{d}{dt}\, t\, e^{-t^2/(\theta\,\lambda)} = \left(1 - \frac{2t^2}{\theta\,\lambda}\right) e^{-t^2/(\theta\,\lambda)} \leq 1.$$

Da weiter $\left(1 - \frac{2t^2}{\theta\,\lambda}\right) e^{-t^2/(\theta\,\lambda)} \geq 1 - 3t^2/(\theta\,\lambda)$, bekommen wir für alle $y \geq x$:

$$\begin{aligned}
& ye^{-y^2/(\theta\,\lambda)} - x\, e^{-x^2/(\theta\,\lambda)} \\
&\qquad = \int_x^y \left(1 - \frac{2t^2}{\theta\,\lambda}\right) e^{-t^2/(\theta\,\lambda)}\, dt \\
&\qquad \geq \int_x^y \left(1 - \frac{3t^2}{\theta\,\lambda}\right) dt = y - x - \frac{y^3 - x^3}{\theta\,\lambda},
\end{aligned}$$

also folgt (ii). (i) impliziert

$$\begin{aligned}
|\lambda\, f(j+1) - j\, f(j)| &= |\lambda\,(f(j+1) - f(j)) - (j-\lambda)^2\, e^{-(j-\lambda)^2/(\theta\,\lambda)}| \\
&\leq \lambda \max\big(1, 2e^{-3/2} + \theta\, e^{-1}\big).
\end{aligned}$$

Damit ist das Lemma bewiesen. □

Ist in Satz 8.1 γ_-/ε_- beschränkt, so ist die Schranke $(1-e^{-\lambda})\,\varepsilon_-$ aus Korollar 4.3 scharf bis auf eine Konstante. Es gilt insbesondere (siehe [BH84]):

Korollar 8.3. *(Barbour, Hall, 1984) Sind die $(I_\alpha, \alpha \in \Gamma)$ unabhängig, so gilt*

$$\big(1-e^{-\lambda}\big)\frac{1}{\lambda}\sum_{\alpha\in\Gamma}\pi_\alpha^2 \geq d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \geq \frac{1}{32}\min\left(1, \frac{1}{\lambda}\right)\sum_{\alpha\in\Gamma}\pi_\alpha^2.$$

Beweis: Die Unabhängigkeit der I_α impliziert

$$\gamma_- = 1 - \frac{1}{\lambda}\sum_{\alpha\in\Gamma}\pi_\alpha(1-\pi_\alpha)(1-6\,\pi_\alpha(1-\pi_\alpha))$$

und

$$\varepsilon_- = 1 - \frac{1}{\lambda}\sum_{\alpha\in\Gamma}\pi_\alpha(1-\pi_\alpha) = \frac{1}{\lambda}\sum_{\alpha\in\Gamma}\pi_\alpha^2.$$

Es folgt somit

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \geq \frac{\varepsilon_-}{11+21/\lambda} \geq \frac{\varepsilon_-}{32}\min(1,\lambda).$$

Die obere Schranke sahen wir in Satz 2.11. □

Ein ähnliches Resultat für positiv zusammenhängende $(I_\alpha, \alpha \in \Gamma)$ (siehe Satz 6.8) wollen wir hier nur angeben. Den Beweis findet man in [BHJ92, Theorem 3.E].

Satz 8.4. *Es seien $(I_\alpha, \alpha \in \Gamma)$ positiv zusammenhängend, $\lambda = \mathbb{E}W, \varepsilon_+ = \frac{V(W)}{\mathbb{E}W} - 1$, $\gamma_+ = \frac{\kappa_4(W)}{\lambda} - 1$ sowie*

$$\psi = \max\left(0, \frac{\gamma_+}{\lambda\varepsilon_+}\right) + 3\,\varepsilon_+ + \frac{\sum\pi_\alpha^2}{\lambda^2\varepsilon_+} + 3\frac{\mathbb{E}\{((W-\lambda)^2 + W - \lambda)\sum\pi_\alpha I_\alpha\}}{\lambda^2\varepsilon_+}.$$

Sei $\varepsilon_+ > 0$. Dann gilt:

$$\big(1-e^{-\lambda}\big)\,\big(\varepsilon_+ + 2\sum_{\alpha\in\Gamma}\pi_\alpha^2/\lambda\big) \geq d_{TV}\big(\big(\mathcal{L}(W), Po(\lambda)\big) \geq \frac{\varepsilon_+}{11+3\psi}.$$

9. Compound Poisson-Approximation

Der lokale Ansatz in Kapitel 3 lieferte die Abschätzung

$$d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \le \min\Big(1, \frac{1}{\lambda}\Big)\Big\{\sum_{\alpha\in\Gamma}\pi_\alpha^2 + \sum_{\alpha\in\Gamma}\pi_\alpha\,\mathbb{E}Z_\alpha + \sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha}\mathbb{E}(I_\alpha\, I_\beta)\Big\},$$

wobei $Z_\alpha := \sum_{\beta\in\Gamma_\alpha} I_\beta$ und zu jedem $\alpha \in \Gamma$ ist Γ_α die Menge der $\beta \in \Gamma \setminus \{\alpha\}$, für die gilt: I_β ist abhängig von I_α. Es soll also gelten: für jedes $\beta \in (\Gamma_\alpha \cup \{\alpha\})^c$ ist I_β unabhängig von I_α.

Der Ausdruck $\lambda^{-1}\sum_{\alpha\in\Gamma}\pi_\alpha(\pi_\alpha + \mathbb{E}Z_\alpha)$ beschreibt die erwartete Anzahl von 1'sen, einschließlich der 1 in α (wenn dort eine ist), in einer *typischen* Nachbarschaft Γ_α von $\alpha \in \Gamma$. Der zweite Summand $\lambda^{-1}\sum_{\alpha\in\Gamma}\mathbb{E}(I_\alpha Z_\alpha) = \lambda^{-1}\sum_{\alpha\in\Gamma}\pi_\alpha\,\mathbb{E}(Z_\alpha|I_\alpha = 1)$ beschreibt die erwartete Anzahl von 1'sen, ausschließlich der 1 in α, in einer typischen Nachbarschaft Γ_α von $\alpha \in \Gamma$, für die $I_\alpha = 1$ gilt. Er beschreibt also die Tendenz von lokalen *Klumpen* von 1'sen. Diese Tendenz kann auch für kleine Γ_α stark sein. Man sieht schnell ein, dass lokales Klumpen die Poisson-Approximation schlecht machen kann: Im Fall $I_{2j} = I_{2j+1} = I'_j$ für jedes j mit $P(I'_j = 1) = p = 1 - P(I'_j = 0)$, unabhängig gewählt, und $W := \sum_{i=1}^n I_i$ folgt: $\mathcal{L}(W)$ ist nicht nahe der Poisson-Verteilung für $p \sim \lambda/n$, wobei wieder $\lambda = \mathbb{E}W$ bezeichnet. Dies erkennt man unmittelbar an dem zweiten Summanden $\lambda^{-1}\sum_{i=1}^n p\,\mathbb{E}(Z_i|I_i = 1)$, sobald Γ_i die nächsten Nachbarn von i enthält! Wir halten fest: Mit der Tendenz lokaler Klumpungen erwarten wir keine oder eine nicht mehr gute Poisson-Approximation. Eine natürliche Frage ist nun: kann bei Klumpungs-Tendenzen eine Verteilung gefunden werden, die in geeigneter Weise diese Klumpung darstellt und durch die dann eine bessere Approximation im Vergleich zur Poisson-Approximation erreicht werden kann?

Wir werden in diesem Kapitel die Approximation mittels einer sogenannten *Compound Poisson-Verteilung* studieren und dabei die Steinsche Methode für diese neue Klasse von Limesverteilungen entwickeln. Dazu gehört analog zu den Ausführungen in Kapitel 2 zunächst das Aufstellen geeigneter Steinscher Gleichungen und dann das Aufsuchen von guten Abschätzungen für die Lösungen der Gleichungen sowie für die Zuwächse dieser Lösungen. Als dann wird man versuchen, den lokalen Ansatz und auch den Kopplungsansatz für die neue Situation zu entwickeln. Wir beschränken uns hier auf den lokalen Ansatz und schliessen mit zwei eindrücklichen Beispielen.

Die erste Arbeit zur Steinschen Methode für Compound Poisson-Approximation ist die von Barbour, Chen und Loh [BCL92]. Eine ganz hervorragende Übersicht über den beinahe aktuellen Stand der Forschung zu diesem Gebiet ist die Arbeit von Barbour und Chryssaphinou [BC01].

Wir betrachten nun die folgende spezielle Klasse von sogenannten Compound Poisson-Verteilungen $CP(\lambda,\mu)$: Es sei μ eine Wahrscheinlichkeitsverteilung auf $\mathbb{N}$ und $(Y_j, j \geq 1)$ eine unabhängige Folge von Zufallsgrößen mit Werten in $\mathbb{N}$, verteilt nach μ. Weiter sei M Poisson-verteilt mit Parameter λ und unabhängig von der Folge $(Y_j, j \geq 1)$ gewählt. Dann definieren wir

$$CP(\lambda,\mu) := \mathcal{L}\Big(\sum_{j=1}^{M} Y_j\Big).$$

Wir setzen $\mu_i := \mu(\{i\})$ für alle $i \in \mathbb{N}$.

Lemma 9.1. *Sind $(M_i, i \geq 1)$ unabhängige $Po(\lambda\,\mu_i)$-verteilte Zufallsgrößen, so gilt*

$$\mathcal{L}\Big(\sum_{j=1}^{M} Y_j\Big) = \mathcal{L}\Big(\sum_{i\geq 1} i\, M_i\Big).$$

Mit $\lambda_i := \lambda\,\mu_i$ ist also $\mathbb{E}(W) = \sum_{i\geq 1} i\,\lambda_i$ für $W \sim CP(\lambda,\mu)$.

Bemerkung 9.2. Man interpretiert die Verteilung wie folgt: man hat eine Poisson-Anzahl M von Klumpen (also eine zufällige Anzahl), deren Größe Y_j unabhängig jeweils nach der Verteilung μ bestimmt wird.

Beweis von Lemma 9.1: Man betrachte die erzeugende Funktion

$$\begin{aligned}
\mathbb{E}\Big(s^{\sum_{j=1}^{M} Y_j}\Big) &= \sum_{m=0}^{\infty} \mathbb{E}\big(s^{Y_1}\big)^m \,\frac{\lambda^m e^{-\lambda}}{m!} \\
&= \exp\Big(-\lambda\big(1-\mathbb{E}\big(s^{Y_1}\big)\big)\Big) = \exp\Big(-\sum_{k=1}^{\infty} \lambda_k\big(1-s^k\big)\Big) \\
&= \prod_{k=1}^{\infty} \exp\big(\lambda_k(s^k-1)\big) = \prod_{k=1}^{\infty} \mathbb{E}\big(s^{k\,M_k}\big) = \mathbb{E}\Big(s^{\sum_{k=1}^{\infty} k\,M_k}\Big)
\end{aligned}$$

für alle $s \in (0,1)$. Damit ist das Lemma bewiesen, denn ist $G(s) := \mathbb{E}\big(s^X\big)$, so ist

$$P(X=n) = \frac{G^{(n)}(0)}{n!}.$$

Also bedeutet die Übereinstimmung der erzeugenden Funktion von $\sum_{j=1}^{M} Y_j$ und $\sum_{i\geq 1} i\,M_i$, dass diese Zufallsgrößen in Verteilung übereinstimmen. $\square$

Mit der Setzung $\lambda_i = \lambda\,\mu_i$ wollen wir im Folgenden

$$CP(\lambda) := CP(\lambda,\mu)$$

schreiben. λ steht dabei für $(\lambda_1, \lambda_2, \dots)$ und wir betrachten die Situation $\sum_{i\geq 1} i\,\lambda_i < \infty$. Für $\lambda_1 > 0$ und $\lambda_i = 0$ für $i \geq 2$ folgt $CP(\lambda) = Po(\lambda_1)$.

Wir entwickeln nun die Steinsche Methode für die Klasse $CP(\lambda)$. Wir stellen fest (und rechnen dies nicht im Einzelnen nach, siehe [BCL92] und [BC01]):

(i) Ist $g : \mathbb{N}_0 \to \mathbb{R}$ eine beschränkte Abbildung und $Z \sim CP(\lambda)$ mit $\mathbb{E}Z = \sum_{i\geq 1} i\,\lambda_i < \infty$, so gilt

$$\mathbb{E}\Big\{ Z\,g(Z) - \sum_{i\geq 1} i\,\lambda_i\, g(Z+i) \Big\} = 0.$$

(ii) Ist $f : \mathbb{N}_0 \to \mathbb{R}$ eine beschränkte Abbildung und gilt $\mathbb{E}(f(Z)) = 0$ für $Z \sim CP(\lambda)$, so kann f immer in der Form

$$f(j) = j\,g(j) - \sum_{i\geq 1} i\,\lambda_i\, g(j+i)$$

dargestellt werden. Wir verzichten auf die Angabe der Formel für $g(j) = g_{f,\lambda}(j)$.

(iii) Erneut wählen wir $f_{A,\lambda}(j) := I(j \in A) - CP(\lambda)\{A\}$ mit $j \geq 0$ und $A \subseteq \mathbb{N}_0$. Dann existiert eine beschränkte Lösung $g_{A,\lambda}$ der **Stein-Gleichungen**

$$j\,g_{A,\lambda}(j) - \sum_{i\geq 1} i\,\lambda_i\, g_{A,\lambda}(j+i) = I(j \in A) - CP(\lambda)\{A\}$$

für alle $j \geq 0$ und $A \subseteq \mathbb{N}_0$.

(iv) Es sei W eine $\mathbb{N}_0$-wertige Zufallsgröße, so folgt aus der Stein-Gleichung

$$\mathbb{E}\bigg(W\,g_{A,\lambda}(W) - \sum_{i=1}^{\infty} i\,\lambda_i\, g_{A,\lambda}(W+i) \bigg) = P(W \in A) - CP(\lambda)\{A\} \tag{9.3}$$

und somit

$$d_{TV}\big(\mathcal{L}(W), CP(\lambda)\big) = \sup_{A\subseteq \mathbb{N}_0} \bigg\{ \bigg| \mathbb{E}\bigg(W\,g_{A,\lambda}(W) - \sum_{i=1}^{\infty} i\,\lambda_i\, g_{A,\lambda}(W+i) \bigg) \bigg| \bigg\}.$$

Analog zur Poisson-Situation erhoffen wir nun für die linke Seite in (9.3) Schranken der Form

$$||g_{A,\lambda}||\, H_1 + \Delta g_{A,\lambda}\, H_2$$

mit

$$\Delta g_{A,\lambda} := \sup_{j\geq 0} |g_{A,\lambda}(j+1) - g_{A,\lambda}(j)|$$

und

$$||g_{A,\lambda}|| := \sup_{j\geq 0} |g_{A,\lambda}(j)|.$$

Mit

$$c_1(\lambda) := \sup_{A\subseteq\mathbb{N}_0} ||g_{A,\lambda}||$$

und

$$c_2(\lambda) := \sup_{A\subseteq\mathbb{N}_0} \Delta g_{A,\lambda}$$

wollen wir Schranken der Form

$$d_{TV}\big(\mathcal{L}(W), CP(\lambda)\big) \leq c_1(\lambda)\, H_1 + c_2(\lambda)\, H_2$$

herleiten.

Tatsächlich kann man für die H_i Schranken via eines *lokalen Ansatzes* oder eines Ansatzes mit *Kopplungsmethoden* erreichen. Schranken für die $c_i(\lambda)$ folgen universell aus der Form der Lösungen der Steingleichungen für $CP(\lambda)$.

Nun geraten wir bereits in tiefe Wasser aktueller Forschungsgegenstände. In den letzten Jahren wurde sowohl der lokale Ansatz als auch der Kopplungsansatz entwickelt. Eine gute Übersicht ist [BC01]. Wir stellen hier nur den lokalen Ansatz und somit auch nur Beispiele mit lokaler Abhängigkeitsstruktur vor. Auf die vielen eindrücklichen Beispiele zu beiden Ansätzen in [BC01] sei an dieser Stelle nur hingewiesen.

Enorm problematisch ist die Analyse universeller Schranken für die $c_i(\lambda)$. Dies mathematisch genauer auszuführen, führt an dieser Stelle zu weit. Wir geben ohne Beweis die in [BCL92] hergeleiteten Schranken für die $c_i(\lambda)$ an. Der Beweis des folgenden Satzes verwendet die Generator-Methode von Barbour, wie sie im Ausblick in Kapitel 2 kurz vorgestellt wurde.

Satz 9.4. *(Schranken für die Steinlösungen; Barbour, Chen, Loh, 1992)*

(i) *Es gilt immer*

$$c_2(\lambda) \leq \left(1 \wedge \frac{c}{\lambda_1}\right) \exp\left(\sum_{i\geq 1} \lambda_i\right).$$

(ii) *Sind die Werte $(i\,\lambda_i)_i$ in i monoton fallend, gilt also*

$$i\,\lambda_i - (i+1)\lambda_{i+1} \geq 0$$

für alle $i \in \mathbb{N}$, so folgt

$$c_1(\lambda) \leq \begin{cases} 1, & \textit{wenn } \lambda_1 - 2\,\lambda_2 \leq 1, \\ \frac{1}{\sqrt{\lambda_1 - 2\,\lambda_2}}\left(2 - \frac{1}{\sqrt{\lambda_1 - 2\,\lambda_2}}\right), & \textit{für } \lambda_1 - 2\,\lambda_2 > 1 \end{cases}$$

und

$$c_2(\lambda) \leq 1 \wedge \frac{1}{\lambda_1 - 2\lambda_2}\left(\frac{1}{4(\lambda_1 - 2\lambda_2)} + \log^+(2(\lambda_1 - 2\lambda_2))\right).$$

Bemerkung 9.5. Wenn $\sum_{i\geq 1} \lambda_i$ groß ist, so ist die erste Schranke nicht zu gebrauchen. Gilt $i\lambda_i \searrow 0$, so ist für $\lambda_2 \ll \lambda_1$ die Limesverteilung dicht bei der Poisson-Verteilung und die Schranke für $c_2(\lambda)$ ist vergleichbar mit der im Poisson-Fall gefundenen Schranke. Aber es gibt noch einen log-Term.

Bemerkung 9.6. In den letzten Jahren wurde in einigen Arbeiten an Verbesserungen für die Schranken in obigem Satz gearbeitet. Wir geben im Ausblick am Ende dieses Kapitels eine kurze Übersicht. Es sei aber jetzt schon darauf hingewiesen, dass es λ gibt, für die die Bedingung

$$i\,\lambda_i - (i+1)\lambda_{i+1} \geq 0, \quad \text{für alle} \quad i \in \mathbb{N},$$

nicht erfüllt ist, und für die ein $\beta > 0$ und eine Konstante $K(\lambda)$ existieren, so dass

$$c_2(\lambda) \geq K(\lambda)\, \exp\left(\beta \sum_{i\geq 1} \lambda_i\right)$$

gilt.

Wir entwickeln nun den *lokalen Ansatz*, der für Beispiele mit lokaler Abhängigkeitsstruktur anwendbar ist (vgl. Kapitel 3).

Es sei Γ eine endliche Menge und $(I_\alpha)_{\alpha\in\Gamma}$ eine endliche Menge von Indikatorvariablen. Sei $W := \sum_{\alpha\in\Gamma} I_\alpha$. Zerlege Γ für jedes α in

$$\Gamma = \{\alpha\} \cup \Gamma_\alpha^{vs} \cup \Gamma_\alpha^b \cup \Gamma_\alpha^{vw},$$

wobei

$$\begin{aligned}
\Gamma_\alpha^{vs} :=& \quad \{\beta \in \Gamma \setminus \{\alpha\} : I_\beta \text{ ist stark abhängig von} \, I_\alpha\},\\
\Gamma_\alpha^{vw} :=& \quad \big\{\beta \in \Gamma \setminus \{\alpha\} : I_\beta \text{ ist schwach abhängig von } \{I_\gamma, \gamma \in \{\alpha\} \cup \Gamma_\alpha^{vs}\}\big\}.
\end{aligned}$$

und

$$\Gamma_\alpha^b := \Gamma \setminus \{\{\alpha\} \cup \Gamma_\alpha^{vs} \cup \Gamma_\alpha^{vw}\}.$$

Bei der Wahl $\Gamma_\alpha^{vs} = \emptyset$ und $\Gamma_\alpha^b = \Gamma_\alpha$, definiert wie in Kapitel 3, erhalten wir die Struktur des lokalen Ansatz bei der Poisson-Approximation. Wir definieren weiter $Z_\alpha := I_\alpha + U_\alpha$ mit

$$U_\alpha := \sum_{\beta\in\Gamma_\alpha^{vs}} I_\beta,$$

$$X_\alpha := \sum_{\beta\in\Gamma_\alpha^b} I_\beta \text{ und } Y_\alpha := \sum_{\beta\in\Gamma_\alpha^{vw}} I_\beta = W - I_\alpha - U_\alpha - X_\alpha.$$

Es gilt der folgende Satz, bewiesen von Roos in [Roo94], siehe auch [Roo93]:

Satz 9.7. *(Roos, 1993) Für jede Wahl von Γ_α^{vs} und Γ_α^{vw} und für jedes beschränkte $g:\mathbb{N}\to\mathbb{R}$ gilt:*

$$\begin{aligned}
&\left|\mathbb{E}\left\{W\,g(W) - \sum_{i\geq 1} i\,\lambda_i\,g(W+i)\right\}\right| \\
&\leq\ \Delta g \sum_{\alpha\in\Gamma}\Big((\mathbb{E}I_\alpha)^2 + \mathbb{E}I_\alpha\,\mathbb{E}(U_\alpha+X_\alpha) + \mathbb{E}(I_\alpha X_\alpha)\Big) + ||g||\,\phi,
\end{aligned}$$

wobei

$$\begin{aligned}
\lambda_i &:= \frac{1}{i}\sum_{\alpha\in\Gamma}\mathbb{E}(I_\alpha\,I(Z_\alpha=i)) \ \textit{für}\ i=1,\ldots,G+1,\\
\mathbb{E}W &= \sum_{i=1}^{G+1} i\,\lambda_i,\ \ G:=\max_{\alpha\in\Gamma}(|\Gamma_\alpha^{vs}|),\\
\phi &:= \sum_{\alpha\in\Gamma}\sum_{i=1}^{|\Gamma_\alpha^{vs}|+1}\phi_{\alpha\,i},\\
\phi_{\alpha\,i} &:= |\mathbb{E}(I_\alpha\,I(Z_\alpha=i)\,|\,(I_\beta;\beta\in\Gamma_\alpha^{vw})) - \mathbb{E}(I_\alpha\,I(Z_\alpha=i))|
\end{aligned}$$

(in Analogie zu Satz 3.4).

Korollar 9.8. *Es gilt in der Situation wie in Satz 9.7*

$$d_{TV}\big(\mathcal{L}(W), CP(\lambda)\big) \leq c_2(\lambda)\sum_{\alpha\in\Gamma}\Big((\mathbb{E}I_\alpha)^2 + \mathbb{E}I_\alpha\,\mathbb{E}(U_\alpha+X_\alpha) + \mathbb{E}(I_\alpha X_\alpha)\Big) + c_1(\lambda)\,\phi.$$

In der praktischen Anwendung muß man zunächst Γ_α^{vs} und Γ_α^{vw} für jedes $\alpha\in\Gamma$ geeignet wählen. Die Wahl wird durch die jeweilige Abhängigkeitsstruktur bestimmt. Dann muß man die Terme $\sum_{\alpha\in\Gamma}(\mathbb{E}I_\alpha)^2$ sowie

$$\sum_{\alpha\in\Gamma}\ \sum_{\beta\in\Gamma_\alpha^{vs}\cup\Gamma_\alpha^{b}} \mathbb{E}I_\alpha\,\mathbb{E}I_\beta$$

und

$$\sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha^{b}}\mathbb{E}(I_\alpha I_\beta)$$

berechnen. Sind die $\{I_\beta,\beta\in\Gamma_\alpha^{vs}\cup\{\alpha\}\}$ unabhängig zu $\{I_\gamma,\gamma\in\Gamma_\alpha^{vw}\}$, so ist $\phi=0$. Weiter muß man die λ_i's für $i=1,\ldots,G$ berechnen! Dies ist nicht so leicht, aber natürlich wichtig, da man nur so die Limesverteilung wirklich kennt. Abschließend verwendet man die Schranken für $c_1(\lambda)$ und $c_2(\lambda)$ gemäß Satz 9.4 (oder aktuellerer Arbeiten, siehe Ausblick).

Beweis von Satz 9.7: Es sei $W_\alpha := W - I_\alpha$ und $W_{\alpha,U} = W_\alpha - U_\alpha = W - I_\alpha - U_\alpha = X_\alpha + Y_\alpha$. Dann gilt

$$\begin{aligned}
\mathbb{E}\big(I_\alpha\, g(W)\big) &= \mathbb{E}\big(I_\alpha\, g(W_\alpha+1)\big) \\
&= \sum_{i=1}^{|\Gamma_\alpha^{vs}|+1} \mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\, g(W_{\alpha,U}+i)\big) \\
&= \sum_{i=1}^{|\Gamma_\alpha^{vs}|+1} \mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\, g(Y_\alpha+i)\big) \\
&\quad + \sum_{i=1}^{|\Gamma_\alpha^{vs}|+1} \mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\, \big(g(Y_\alpha + X_\alpha + i) - g(Y_\alpha+i)\big)\big).
\end{aligned}$$

Weiter gilt

$$\begin{aligned}
&\mathbb{E}\left(W\, g(W) - \sum_{i=1}^{G+1} i\, \lambda_i\, g(W+i)\right) \\
&= \sum_{\alpha\in\Gamma} \mathbb{E}\big(I_\alpha\, g(W)\big) - \sum_{i=1}^{G+1} i\,\lambda_i\, \mathbb{E}\big(g(W+i)\big) \\
&= \sum_{\alpha\in\Gamma} \sum_{i=1}^{|\Gamma_\alpha^{vs}|+1} \mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\, \big(g(Y_\alpha + X_\alpha + i) - g(Y_\alpha+i)\big)\big) \\
&\quad + \sum_{\alpha\in\Gamma} \sum_{i=1}^{|\Gamma_\alpha^{vs}|+1} \mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\, g(Y_\alpha+i)\big) - \mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\big)\mathbb{E}\big(g(Y_\alpha+i)\big) \\
&\quad + \sum_{\alpha\in\Gamma} \sum_{i=1}^{|\Gamma_\alpha^{vs}|+1} \mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\big)\mathbb{E}\big(g(Y_\alpha+i) - g(W+i)\big).
\end{aligned}$$

Bemerke, dass

$$\Big|\mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\big(g(Y_\alpha + X_\alpha + i) - g(Y_\alpha+i)\big)\big)\Big| \le \Delta g\, \mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\, X_\alpha\big)$$

und

$$\Big|\mathbb{E}\big(g(Y_\alpha+i) - g(W+i)\big)\Big| \le \Delta g\, \mathbb{E}(I_\alpha + U_\alpha + X_\alpha)$$

sowie

$$\mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\, g(Y_\alpha+i)\big) - \mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\big)\mathbb{E}\big(g(Y_\alpha+i)\big) \le ||g||\, \phi_{\alpha i}$$

gilt. Also folgt:

$$
\begin{aligned}
&\left|\mathbb{E}\left\{W\,g(W) - \sum_{i=1}^{G+1} i\,\lambda_i\,g(W+i)\right\}\right| \\
\leq\;& \Delta g \sum_{\alpha\in\Gamma}\sum_{i=1}^{|\Gamma_\alpha^{vs}|+1} \mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\, X_\alpha\big) \\
&+ \Delta g \sum_{\alpha\in\Gamma}\sum_{i=1}^{|\Gamma_\alpha^{vs}|+1} \mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\big)\mathbb{E}\big(I_\alpha + U_\alpha + X_\alpha\big) \\
&+ ||g|| \sum_{\alpha\in\Gamma}\sum_{i=1}^{|\Gamma_\alpha^{vs}|+1} \phi_{\alpha\, i} \\
=\;& \Delta g \sum_{\alpha\in\Gamma}\Big((\mathbb{E}I_\alpha)^2 + \mathbb{E}I_\alpha\, \mathbb{E}(U_\alpha + X_\alpha) + \mathbb{E}\big(I_\alpha\, X_\alpha\big)\Big) + ||g||\,\phi.
\end{aligned}
$$

Damit ist der Satz bewiesen. □

Der Vorteil der Compound Poisson-Approximation liegt darin, dass der Ausdruck $\sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha^{vs}} \mathbb{E}(I_\alpha I_\beta)$ nicht vorkommt. Lokale Klumpen beeinflussen die Ordnung der Abschätzung nicht, sind aber in der Limesverteilung berücksichtigt. Wir betrachten zwei Beispiele:

Beispiel 9.9. (k-runs, Roos 1993). Das Modell ist gewählt wie in Beispiel 3.9. Nun wählen wir

$$
\begin{aligned}
\Gamma_\alpha^{vs} &= \{\alpha-(k-1),\ldots,\alpha-1,\alpha+1,\ldots,\alpha+k-1\} \\
\Gamma_\alpha^{b} &= \{\alpha-2(k-1),\ldots,\alpha-k,\alpha+k,\ldots,\alpha+2(k-1)\}
\end{aligned}
$$

und $\Gamma_\alpha^{vw} = \Gamma\setminus\{\{\alpha\}\cup\Gamma_\alpha^{vs}\cup\Gamma_\alpha^{b}\}$. Dann ist $|\Gamma_\alpha^{vs}| = 2(k-1)$ und $|\Gamma_\alpha^{b}| = 2(k-1)$ für jedes $\alpha\in\Gamma$. Es ist mit $\pi := p^k$

$$
\sum_{\alpha\in\Gamma}\big((\mathbb{E}I_\alpha)^2 + \mathbb{E}I_\alpha\,\mathbb{E}(U_\alpha + X_\alpha)\big) = (4k-3)\,n\,\pi^2
$$

und $\sum_{\alpha\in\Gamma}\mathbb{E}(I_\alpha\, X_\alpha) = 2(k-1)\,n\,\pi^2$ sowie $\phi = 0$. Also ist

$$
d_{TV}\big(\mathcal{L}(W), CP(\lambda)\big) \leq c_2(\lambda)\,(6k-5)\,n\,\pi^2.
$$

Wir bestimmen nun die λ_i für $i = 1,\ldots,2k-1$.

Lemma 9.10. *Für obiges k-run Beispiel gilt*

$$\lambda_i = \begin{cases} n\pi\, p^{i-1}(1-p)^2, & i = 1,\dots,k-1, \\ \frac{n\pi}{i}(2p^{i-1}(1-p) + (2k-i-2)p^{i-1}(1-p)^2), & i = k,\dots,2k-2, \\ \frac{n\pi}{2k-1}\, p^{2k-2}, & i = 2k-1. \end{cases}$$

Beweis: Es gilt

$$\begin{aligned} \lambda_i &= \frac{1}{i}\sum_{\alpha\in\Gamma} \mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\big) = \frac{n}{i}\mathbb{E}\big(I_k\, I(Z_k = i)\big) \\ &= \frac{n\,\pi}{i} P(I_1 + \dots + I_{k-1} + I_{k+1} + \dots + I_{2k-1} = i-1 | I_k = 1) \\ &= \frac{n\,\pi}{i} P\big(V_{k-1} + V'_{k-1} = i-1\big), \end{aligned}$$

wobei V_{k-1} und V'_{k-1} unabhängige und *geometrisch verteilte* Zufallsgrößen zum Parameter p bezeichnen, die bei $k-1$ abgeschnitten sind:

$$P(V_{k-1} = l) = \begin{cases} p^l\,(1-p), & l = 1,\dots,k-2, \\ p^{k-1}, & l = k-1. \end{cases}$$

Somit folgt

$$\begin{aligned} & P\big(V_{k-1} + V'_{k-1} = i-1\big) \\ &\quad = \sum_{l=0}^{i-1} P\big(V_{k-1} = l\big)\, P\big(V'_{k-1} = i-1-l\big) \\ &\quad = \sum_{l=0}^{i-1} p^l\,(1-p)\, p^{i-1-l}\,(1-p) = i\, p^{i-1}\,(1-p)^2 \end{aligned}$$

für $i = 1,\dots,k-1$. Für $i = k,\dots,2k-2$ ist

$$\begin{aligned} & P\big(V_{k-1} + V'_{k-1} = i-1\big) \\ &\quad = \sum_{l=i-1-(k-1)}^{k-1} P\big(V_{k-1} = l\big)\, P\big(V'_{k-1} = i-1-l\big) \\ &\quad = 2\,p^{k-1}\,p^{i-k}\,(1-p) + \sum_{l=i-k+1}^{k-2} p^l\,(1-p)\,p^{i-1-l}\,(1-p) \\ &\quad = 2\,p^{i-1}\,(1-p) + (2k-i-2)\,p^{i-1}\,(1-p)^2 \end{aligned}$$

und für $i = 2k-1$ ist

$$P\big(V_{k-1} + V'_{k-1} = 2k-2\big) = p^{2k-2}.$$

Dies zusammen liefert das Lemma. □

Um die Schranken aus Satz 9.4 anwenden zu können, müssen wir analysieren, ob $i\lambda_i \searrow 0$ gilt. Wir erhalten:

Lemma 9.11. *Es gilt*

$$i\,\lambda_i \searrow 0 \;\; \textit{falls} \;\; \begin{cases} p < 1/3, & \textit{für } k = 2, \\ p < \frac{5-\sqrt{17}}{2}, & \textit{für } k = 3, \\ p < \frac{1}{2}, & \textit{für } k \geq 4. \end{cases}$$

Weiter ist $\lambda_1 - 2\,\lambda_2 = n\,\pi(1-p)^2(1-2p)$.

Der Beweis geht via einfacher Fallunterscheidungen und soll hier nicht geführt werden, siehe aber [Roo93]. Wir erhalten in Abhängigkeit von k nur für spezielle p-Werte eine Verbesserung durch Compound Poisson-Approximation mittels der Verwendung der Schranken in Satz 9.4:

Satz 9.12. *(k-run, Compound Poisson-Approximation) Wenn* $i\,\lambda_i \searrow 0$ *gilt, so ist*

$$\begin{aligned} & c_2(\lambda) \\ & \leq \left\{ 1 \wedge \frac{1}{n\,\pi\,(1-p)^2(1-2p)} \left\{ \frac{1}{4\,n\,\pi\,(1-p)^2(1-2p)} + \log^+\big(2\,n\,\pi\,(1-p)^2(1-2p)\big) \right\} \right\}. \end{aligned}$$

Die Schranke hat für die Werte p und k, für die $i\,\lambda_i \searrow 0$ gilt, die Ordnung $k\,p^k$ und ist somit eine deutliche Verbesserung gegenüber $\mathcal{O}(p)$ bei der Poisson-Approximation.

Beispiel 9.13. Wir betrachten das Modell *isolierter Knoten* wie in Beispiel 3.12.

Hier wählen wir

$$\Gamma_\alpha^{vs} = \big\{\beta \neq \alpha : \beta \text{ ist Gitternachbarpunkt von } \alpha\big\}$$

und

$$\Gamma_\alpha^{b} = \big\{\gamma \neq \alpha : \gamma \text{ ist Nachbar von } \beta \text{ mit } \beta \in \Gamma_\alpha^{vs}\big\}.$$

Dann ist $|\Gamma_\alpha^{vs}| = 4$ und $|\Gamma_\alpha^{b}| = 8$. Weiter gilt $\sum_{\alpha\in\Gamma}(\mathbb{E}I_\alpha)^2 = n\,q^8$ und

$$\sum_{\alpha\in\Gamma} \mathbb{E}I_\alpha\,\mathbb{E}\big(U_\alpha + X_\alpha\big) = 12\,n\,q^8$$

und

$$\sum_{\alpha\in\Gamma} \mathbb{E}(I_\alpha\,X_\alpha) = 8\,n\,q^8.$$

Wieder ist leicht zu sehen, dass $\phi = 0$ ist. Also ist

$$d_{TV}\big(\mathcal{L}(W), CP(\lambda)\big) \leq c_2(\lambda)\,21\,n\,q^8.$$

Nun bestimmen wir die Parameter λ_i: Es ist

$$\begin{aligned}\lambda_i &= \frac{1}{i}\sum_{\alpha\in\Gamma}\mathbb{E}\big(I_\alpha\, I(Z_\alpha = i)\big)\\ &= \frac{q^4}{i}\sum_{\alpha\in\Gamma}P\bigg(\sum_{\beta\in\Gamma_\alpha^{vs}} I_\beta = i-1 | I_\alpha = 1\bigg)\end{aligned}$$

für $i = 1,\ldots,5$. Nun ist $\sum_{\beta\in\Gamma_\alpha^{vs}} I_\beta$ verteilt nach $b(\cdot,4,q^3)$ bei Bedingung auf das Ereignis $I_\alpha = 1$. Daher folgt:

$$\begin{aligned}\lambda_i &= \frac{n}{i}q^4\binom{4}{i-1}q^{3(i-1)}(1-q^3)^{5-i}\\ &= \frac{n}{i}\binom{4}{i-1}q^{3i+1}(1-q^3)^{5-i}\end{aligned}$$

Es ist

$$\begin{aligned}\lambda_1 &= nq^4(1-q^3)^4,\\ \lambda_2 &= 2nq^7(1-q^3)^3 \quad \text{und}\\ \lambda_i &= \mathcal{O}(n\,q^{10}) \;\text{ für }\; 3\le i\le 5.\end{aligned}$$

Nun wenden wir einen kleinen Trick an: man kann die Anzahl der zu bestimmenden Parameter λ_i mit der folgenden Konstruktion reduzieren: Wieder bezeichne $G = \max_{\alpha\in\Gamma}\big(|\Gamma_\alpha^{vs}|\big)$.

Satz 9.14. *Definiere*

$$\begin{array}{lll}\lambda_1^\star &= \lambda_1 + \sum_{i=l+1}^{G+1} i\,\lambda_i, & \textit{für } l < G+1,\\ \lambda_j^\star &= \lambda_j, & \textit{für } j = 2,\ldots,l, \;\; \textit{und}\\ \lambda_j^\star &= 0, & \textit{für } j \ge l+1.\end{array}$$

Dann gilt in der obigen Notation:

$$\begin{aligned}&d_{TV}\big(\mathcal{L}(W), CP(\lambda^\star)\big)\\ &\le\; c_2(\lambda^\star)\bigg\{\sum_{\alpha\in\Gamma}\Big((\mathbb{E}I_\alpha)^2 + \mathbb{E}I_\alpha\,\mathbb{E}(U_\alpha + X_\alpha) + \mathbb{E}(I_\alpha\, X_\alpha)\Big)\\ &\qquad + \sum_{i=l+1}^{G+1} i(i-1)\lambda_i\bigg\} + c_1(\lambda^\star)\,\phi,\end{aligned}$$

wobei $\lambda^\star := (\lambda_1^\star,\ldots,\lambda_l^\star)$.

Der Beweis ist nicht all zu schwer. Wir lassen ihn weg, wollen den Satz aber für Beispiel 9.13 anwenden:

Wir wählen $l = 2$. Dann ist

$$\begin{aligned}
\lambda_1^\star &= \lambda_1 + \sum_{i=3}^{5} i\lambda_i = \mathbb{E}W - 2\lambda_2 \\
\lambda_2^\star &= \lambda_2 = 2nq^7(1-q^3)^3.
\end{aligned}$$

Da hier $\sum_{i=3}^{5} i(i-1)\lambda_i = 6\lambda_3 + 12\lambda_4 + 20\lambda_5$ eine Ordnung hat, die kleiner als $21nq^8$ ist, liefert der Zusatzterm in Satz 9.14 keine Probleme in Bezug auf die Größenordnung. Weiter gilt $i\lambda_i^\star \searrow 0$, denn

$$\frac{2\lambda_2^\star}{\lambda_1^\star} = \frac{4q^3(1-q^3)^3}{1 - 4\,q^3(1-q^3)^3} < 1 \text{ für } 0 \le q < 1.$$

Es ist $\lambda_1^\star - 2\lambda_2^\star = \mathbb{E}W - 4\lambda_2 = nq^4(1 - 8\,q^3(1-q^3)^3)$. Weiter ist

$$\sum_{i=3}^{5} i(i-1)\lambda_i = 12\,n\,q^{10}(1 + \mathcal{O}(q^3)).$$

Wir erhalten mit Satz 9.14 unter Verwendung der Schranken für $c_2(\lambda^\star)$ aus Satz 9.4:

$$\begin{aligned}
&d_{TV}\big(\mathcal{L}(W), CP(\lambda^\star)\big) \\
&\le \frac{1}{nq^4(1 - 8\,q^3(1-q^3)^3)}\big(1 + \log^+ 2nq^4\big)\big(21nq^8 + 12nq^{10}(1 + \mathcal{O}(q^3))\big) \\
&\le \big(1 + \log^+ 2\,nq^4\big)21\,q^4\,\big(1 + \mathcal{O}(q^2)\big).
\end{aligned}$$

Für $\lim_{n\to\infty} n\,q_n^4 = \mu_\infty > 1$ gilt insbesondere $d_{TV}\big(\mathcal{L}(W), CP(\lambda^\star)\big) = \mathcal{O}(1/n)$. Vergleiche mit Beispiel 3.12. Dort hatten wir $d_{TV}\big(\mathcal{L}(W), Po(\lambda)\big) \le \mathcal{O}(1/n^{3/4})$ für $\lim_{n\to\infty} n\,q_n^4 = \mu_\infty > 1$ gezeigt. Analog zu Beispiel 3.12 ist die Zuverlässigkeit $P(W \le m-1)$ des Systems approximiert durch

$$P(W \le m-1) \approx \sum_{i=0}^{m-1} P(X = i) \text{ mit } X \sim CP(\lambda^\star).$$

Ist $N_1 \sim Po(\lambda_1^\star)$ und $N_2 \sim Po(\lambda_2^\star)$, unabhängig von N_1, so gilt

$$P(X = i) = P(N_1 + 2\,N_2 = i) = \sum_{j=0}^{[i/2]} \frac{(\lambda_1^\star)^{i-2j}}{(i-2j)!}\frac{(\lambda_2^\star)^j}{j!}e^{-(\lambda_1^\star + \lambda_2^\star)}$$

mit $\lambda_1^\star = n\,q^4(1 - 4\,q^3(1-q^3)^3$ und $\lambda_2^\star = 2n\,q^7(1-q^3)^3$.

Ausblick: Wir haben bereits darauf hingewiesen, dass es bei der Compound Poisson-Approximation via der Steinschen Methode interessante Entwicklungen gibt. Wir stellen hier kurz einige vor.

Wir haben nur aus der Gewohnheit der Poisson-Approximation heraus und wegen der ins Auge gefassten Beispiele Summen von Indikatorvariablen betrachtet. Die Methode kann für die Beispiel-Klasse $W = \sum_{\alpha \in \Gamma} Y_\alpha$ für Zufallsgrößen Y_α mit Werten in $\mathbb{Z}_+$ entwickelt werden. Weiter bestimmt man die Distanz zwischen der Verteilung von W und der $CP(\lambda, \mu)$-Verteilung nicht immer in der Totalvariation. Man untersucht auch den *Kolmogorov-Abstand*, der für zwei Verteilungen P und Q auf $\mathbb{Z}_+$ durch

$$d_{\mathrm{K}}(P, Q) := \sup_{m \in \mathbb{Z}_+} |P\{[m, \infty)\} - Q\{[m, \infty)\}|$$

definiert ist. Entsprechend der Steinschen Gleichung hat man für den Kolmogorov-Abstand nun die folgende Größe abzuschätzen:

$$d_K\big(\mathcal{L}(W), CP(\lambda)\big) = \sup_{m \in \mathbb{Z}_+} \left| \mathbb{E}\bigg(-W\, g_{[m,\infty),\lambda}(W) + \sum_{i=1}^{\infty} i\, \lambda_i\, g_{[m,\infty),\lambda}(W+i) \bigg) \right| .$$

Barbour und Xia haben 1999 in [BX99] das folgende Resultat bewiesen. Wir betrachten die Verteilung $CP(\lambda, \mu)$. Es sei $m_1 := \sum_{i \geq 1} i \mu_i$ und $m_2 := \sum_{i \geq 1} i^2\, \mu_i$: wenn die Bedingung

$$\theta := m_1^{-1}(m_2 - m_1) < 1/2 \tag{9.15}$$

gilt, so folgt

$$c_1(\lambda, \mu) \leq \frac{1}{(1-2\theta)\sqrt{\lambda m_1}}$$

und

$$c_2(\lambda, \mu) \leq \frac{1}{(1-2\theta)\, \lambda\, m_1}.$$

Diese Schranken sind in λ von der gleichen Ordnung wie die Schranken in Kapitel 2. Der Parameter θ kann einfach ausgerechnet werden. In der Notation des lokalen Ansatzes gilt

$$\theta = \sum_{\alpha \in \Gamma} \mathbb{E}\big(Y_\alpha\, U_\alpha\big) / \sum_{\alpha \in \Gamma} \mathbb{E} Y_\alpha .$$

Für die Kolmogorov-Distanz haben Barbour und Xia in [BX00] die folgenden Schranken hergeleitet: Gilt die Monotonie-Bedingung

$$i \mu_i \geq (i+1) \mu_{i+1}, \quad i \geq 1, \tag{9.16}$$

so folgt

$$c_1^K(\lambda, \mu) \leq \min\left(1, \sqrt{\frac{2}{e\, \lambda\, \mu_1}}\right)$$

und

$$c_2^K(\lambda, \mu) \leq \min\left(\frac{1}{2}, \frac{1}{\lambda\, \mu_1 + 1}\right).$$

Es gibt eine Reihe von Beispielen, in denen die Monotonie-Bedingung (9.16) aber auch die Bedingung (9.15) nicht erfüllt sind. In diesen Fällen sind die Schranken für die Stein-Lösungen sehr unbefriedigend. Barbour und Utev haben in [BU98] und in [BU99] Verbesserungen der Schranken unter anderen, technisch recht aufwendigen Bedingungen hergeleitet. Eine kurze Darstellung dieser Resultate findet man in [ER99] sowie in [BC01].

Nun geben wir für das Beispiel des k-Runs den aktuellen Stand der Forschung an: Die Schranke in Satz 9.12 gilt im Fall $p \leq 1/3$, denn in diesem Fall ist die Monotonie-Bedingung (9.16) erfüllt. Die Bedingung (9.15) der Arbeit von Barbour und Xia ist für alle $p \leq 1/5$ erfüllt, denn es gilt die Abschätzung $\theta \leq 2p/(1-p)$. Dann liefert die obige Abschätzung

$$d_{\mathrm{TV}}\big(\mathcal{L}(W), CP(\lambda,\mu)\big) \leq (6k-5)\, p^k\, (1-p)/(1-5p),$$

was also sogar besser ist als $\mathcal{O}(k\, p^k)$. Ist keine Bedingung an p gestellt, so haben Eichelsbacher und Roos in [ER99] gezeigt, dass die Abschätzung von Barbour und Utev in der Totalvariation einen Fehler der Ordnung

$$\mathcal{O}\big(k\, p^k + e^{-const.(k)\, n\, p^k}\big)$$

liefern. Hierbei ist die Konstante nicht näher bestimmt. Schliesslich haben Barbour und Xia in [BX99] den Fall $k=2$ genauer untersucht. Mittels eines recht aufwendigen Arguments und unter Verwendung einer modifizierten approximierenden Compound Poisson-Verteilung bekommen sie eine Fehlerschranke der Ordnung $\mathcal{O}\big(p^k (np^k)^{-1/2}\big)$. Die Schranke ist besser als alle anderen.

Schliesslich sei auf die Arbeit von Barbour und Månsson [BM02] hingewiesen, in der die Steinsche Methode für die Compound Poisson-Approximation auf der Ebene der *Punkt-Prozesse* entwickelt wird.

10. Normal-Approximation via Steinscher Methode

Charles Stein entwickelte 1972 in [Ste72] die nach ihm benannte Methode ursprünglich zur Untersuchung von Normal-Approximation bei abhängigen Zufallsgrößen. Er untersuchte speziell stationäre, mischende Zufallsgrößen. Wir wollen in diesem Kapitel die Normal-Approximation betrachten.

Zunächst führen wir den Begriff der schwachen Konvergenz ein. Dann skizzieren wir einen klassischen Beweis des Zentralen Grenzwertsatzes von Markov aus dem Jahre 1898 für Summen von unabhängigen und identisch verteilten Zufallsgrößen. Klassisch heißt, dass wir Fourier-Transformierte und ihre Eigenschaften verwenden. Im Anschluss erfolgt die Steinsche Charakterisierung der Normalverteilung, die Formulierung der Steinschen Gleichung sowie die Abschätzung der Lösungen der Gleichung und die Abschätzung der ersten und zweiten Ableitung dieser Lösungen. Wir beweisen dann den Satz von Markov via Steinscher Methode und diskutieren die Konvergenzrate, die ohne Aufwand aus dieser Beweis-Methode gewonnen werden kann. Diese Rate ist von der Ordnung $n^{-1/4}$. Wir skizzieren im Ausblick, wie Bolthausen [Bol84] die optimale Rate $n^{-1/2}$ von Berry [Ber41] und Esséen [Ess45] via der Steinschen Methode gewonnen hat.

Für Wahrscheinlichkeitsmaße auf $\mathbb{N}$ lieferte der Totalvariationsabstand einen geeigneten Konvergenzbegriff. Was ist ein geeigneter Begriff für Maße auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$?

Wir erinnern an den Satz von de Moivre und Laplace: Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit $P(X_i = 1) = p$ und $P(X_i = 0) = 1 - p$ für alle i mit einem $p \in (0,1)$. Dann ist $S_n := X_1 + \cdots + X_n$ $b(\cdot; n, p)$ verteilt mit Parametern n und p und es gilt für alle $t \in \mathbb{R}$

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \le t\right) = \Phi(t) := \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx.$$

Konvergiert die Folge dieser Zufallsgrößen? In welchem Sinne meinen wir Konvergenz? Eine inoffizielle Mitteilung ist: aus dem Kolmogorov'schen 0-1-Gesetz folgt, dass $\limsup_{n\to\infty}(S_n - np)/\sqrt{np(1-p)} = \infty$ fast sicher und $\liminf_{n\to\infty}(S_n - np)/\sqrt{np(1-p)} = -\infty$ fast sicher ist. Die Zufallsgrößen konvergieren also *nicht* fast sicher! Tatsächlich konvergiert die Folge der *Verteilungen* von $(S_n - np)/\sqrt{np(1-p)}$. Dies wollen wir nun präzisieren. Sei μ_n die Folge der Verteilungen von $(S_n - np)/\sqrt{np(1-p)}$ und μ sei die Standard-Normalverteilung. Was bedeutet μ_n konvergiert gegen μ? Gilt $\mu_n(A) \to \mu(A)$ für alle $A \in \mathcal{B}(\mathbb{R})$? Nein, denn für jedes n existiert eine endliche Menge A_n mit $\mu_n(A_n) = 1$. Setzen wir

$A := \cup_{n=1}^{\infty} A_n$, so ist A abzählbar, und es gilt $\mu_n(A) = 1$ für alle n, aber es ist $\mu(A) = 0$.

Wir geben nun eine Definition eines geeigneten Konvergenzbegriffs: Es sei (S, d) ein metrischer Raum und $\mathcal{B}_S$ die kleinste σ-Algebra, die die offenen Mengen enthält. $\mathcal{M}_1(S)$ bezeichne die Menge der Wahrscheinlichkeitsmaße auf $(S, \mathcal{B}_S)$. $C(S)$ bezeichne die Menge der reellwertigen, beschränkten und stetigen Funktionen auf S.

Bemerkung 10.1. Es seien $\nu, \mu \in \mathcal{M}_1(S)$. Gilt $\int f d\mu = \int f d\nu$ für alle $f \in C(S)$, so gilt $\mu = \nu$ (ohne Beweis).

Definition 10.2. (i) Seien $\mu_n, \mu \in \mathcal{M}_1(S)$ für $n \in \mathbb{N}$. Die Folge $(\mu_n)_{n\in\mathbb{N}}$ *konvergiert schwach* gegen μ (Notation $\mu_n \xrightarrow{w} \mu$), wenn

$$\lim_{n\to\infty} \int f \, d\mu_n = \int f \, d\mu \ \text{ für alle } \ f \in C(S) \ \text{ gilt.}$$

(ii) Es seien X_n und X $(S, \mathcal{B}_S)$-wertige Zufallsgrößen, die auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ definiert sind. Falls $P \circ X_n^{-1} \xrightarrow{w} P \circ X^{-1}$, so sagt man, die Folge $(X_n)_{n\in\mathbb{N}}$ *konvergiert in Verteilung* gegen X, und schreibt

$$\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(X).$$

Es gilt:

Satz 10.3. *(Zentraler Grenzwertsatz, Markov, 1898)*
Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger, reeller, quadratisch integrierbarer, identisch verteilter Zufallsvariablen mit $\mathbb{E}X_n = 0$ und $V(X_n) > 0$. Es bezeichne $\sigma^2 := \sigma^2(X_n) := V(X_n)$ und $S_n := \sum_{k=1}^n X_k$. Dann konvergiert $\mathcal{L}\big(\frac{S_n}{\sqrt{n}\sigma}\big)$ schwach gegen die Standardnormalverteilung.

Ein typischer Beweis verwendet *charakteristische Funktionen*. Für $\mu \in \mathcal{M}_1(\mathbb{R})$ ist $\int e^{ixy}\mu(dy) =: \hat{\mu}(x)$ die Fourier-Transformierte (komplexe Funktion). Für eine reellwertige Zufallsgröße X heißt $\mathbb{E}(e^{ixX}) = \varphi_X(x)$ charakteristische Funktion. Es gilt: Die charakteristische Funktion der Summe unabhängiger $\mathbb{R}$-wertiger Zufallsgrößen ist gleich dem Produkt der einzelnen charakteristischen Funktionen (siehe zum Beispiel Satz 22.5 in [Bau02]). Der *Stetigkeitssatz von Lévy* besagt: Ist die Folge $(\hat{\mu}_n)_{n\in\mathbb{N}}$ der Fourier-Transformierten einer Folge $(\mu_n)_{n\in\mathbb{N}}$ von Wahrscheinlichkeitsmaßen auf $\mathbb{R}$ punktweise konvergent gegen eine in $x = 0$ stetige komplexe Funktion φ auf $\mathbb{R}$, so ist φ die Fourier-Transformierte eines eindeutig bestimmten Maßes $\mu \in \mathcal{M}_1(\mathbb{R})$ und die Folge $(\mu_n)_{n\in\mathbb{N}}$ konvergiert schwach gegen μ (siehe Satz 23.8 in [Bau02]). Mittels dieser Informationen geht der Beweis von Satz 10.3 etwa so: $x \mapsto x/(\sigma\sqrt{n})$

ist linear, somit folgt nach elementaren Rechenregeln (Satz 22.4 (d) in [Bau02]) für die charakteristische Funktion von $S_n/\sigma\sqrt{n}$:

$$\varphi_{S_n/\sigma\sqrt{n}}(x) = \hat{\mu}\left(\frac{x}{\sigma\sqrt{n}}\right)^n, \quad x \in \mathbb{R},$$

wenn μ die Verteilung der X_i bezeichnet. Es ist also zu zeigen (Anwendung des Stetigkeitssatzes):

$$\text{für alle } x \in \mathbb{R} \text{ gilt: } \lim_{n\to\infty} \hat{\mu}\left(\frac{x}{\sigma\sqrt{n}}\right)^n = \exp\bigl(-x^2/2\bigr).$$

Die rechte Seite ist die Fourier-Transformierte der $N(0,1)$-Verteilung (Übung).

Da X_n quadrat-integrierbar ist, existieren für μ die *Momente*:

$$M_0 := \mathbb{E}X_n^0 = 1; \quad M_1 := \mathbb{E}X_n = 0; \quad M_2 := V(X_n) = \sigma^2.$$

Für die Taylorentwicklung von $\hat{\mu}(x)$ gilt allgemein:

$$\begin{aligned} \hat{\mu}(x) &= \sum_{\kappa=0}^{k} \frac{(i\,x)^\kappa}{\kappa!} M_\kappa + \Theta(x)\frac{x^k}{k!}, \text{ wobei} \\ |\Theta(x)| &\le \sup_{0\le\vartheta\le 1} |\hat{\mu}^{(k)}(\vartheta\, x) - \hat{\mu}^{(k)}(0)|, \end{aligned}$$

sobald alle auftretenden Größen existieren (siehe Kapitel 25 in [Bau02]).

Hier ist nun:

$$\hat{\mu}(x) = 1 - \frac{1}{2}\sigma^2 x^2 + \Theta(x)\frac{x^2}{2}$$

(Θ ist eine stetige komplexe Funktion auf $\mathbb{R}$ mit $\Theta(0) = 0$). Also folgt

$$\hat{\mu}\left(\frac{x}{\sigma\sqrt{n}}\right)^n = \left(1 + \frac{-x^2/2 + \Theta_n\, x^2}{n}\right)^n$$

mit

$$\Theta_n := \frac{1}{2\sigma^2}\Theta\left(\frac{x}{\sigma\sqrt{n}}\right).$$

Es gilt $\lim_{n\to\infty} \Theta_n = \Theta(0) = 0$. Da $|a^n - b^n| \le n|a-b|$ für $a, b \in \mathbb{C}$ mit $|a| \le 1, |b| \le 1$ (via Induktion), folgt für hinreichend große n

$$\left| \left(1 + \frac{-x^2/2 + \Theta_n x^2}{n}\right)^n - \left(1 + \frac{-x^2/2}{n}\right)^n \right| \le |\Theta_n\, x^2|,$$

und somit

$$\exp(-x^2/2) = \lim_{n\to\infty}\left(1 + \frac{-x^2/2}{n}\right)^n = \lim_{n\to\infty}\left(1 + \frac{-x^2/2 + \Theta_n\, x^2}{n}\right)^n,$$

also ist $e^{-x^2/2}$ tatsächlich der Limes für $n \to \infty$ und der zentrale Grenzwertsatz ist bewiesen. Die Theorie der charakteristischen Funktion muß allerdings vollständig entwickelt werden. Ein zentraler Aspekt der Beweisskizze ist, dass bei unabhängigen Zufallsgrößen die charakteristische Funktion faktorisiert. So hat man lange Zeit unabhängige $(X_n)_{n\in\mathbb{N}}$ studiert. Ausführlich wird der Fall *unabhängiger Zufallsgrößen* in [Bau02, Kapitel 28] behandelt. Die Annahme, die Zufallsgrößen seien identisch verteilt, kann deutlich abgeschwächt werden. Dies geht zurück auf Lindeberg und Feller. Der Satz von Lindeberg und Feller gibt notwendige und hinreichende Bedingungen für die Gültigkeit des zentralen Grenzwertsatzes für Folgen von unabhängigen quadratisch integrierbaren Zufallsgrößen mit positiver Varianz an, siehe Satz 28.3 in [Bau02].

Die Steinsche Methode hat nun den Vorteil, dass *schwach abhängige* Zufallsgrößen studiert werden können. Dies betrachten wir im Kapitel 11. Wesentlich ist weiterhin, dass die Methode *Konvergenzraten* liefert. Für den in Satz 10.3 betrachteten Fall ist bekannt (siehe [Ber41], [Ess45]):

Satz 10.4. *(Berry, Esséen, 1945) Gegeben sei die Situation aus Satz 10.3. Gilt zusätzlich $\xi := \mathbb{E}|X_i|^3 < \infty$, so ist*

$$\sup_{t\in\mathbb{R}} \left| P\left(\frac{S_n}{\sqrt{n}\sigma} \le t \right) - \Phi(t) \right| \le \frac{K\xi}{\sigma^3\sqrt{n}},$$

wobei K eine universelle Konstante ist.

Dieser Satz gilt auch im Fall nicht identisch verteilter Zufallsgrößen und lautet:

Satz 10.5. *(Berry, Esséen, 1945) Gegeben sei eine Folge $(X_n)_{n\in\mathbb{N}}$ unabhängiger quadratisch integrierbarer, reeller Zufallsvariablen mit positiver Varianz $\sigma_n^2 := \mathrm{Var}(X_n)$. Es sei*

$$s_n := \left(\sigma_1^2 + \cdots + \sigma_n^2\right)^{1/2}$$

und $\eta_n := \mathbb{E}(X_n)$. Dann gilt

$$\sup_{t\in\mathbb{R}} \left| P\left(\frac{1}{s_n} \sum_{j=1}^n (X_j - \eta_j) \le t \right) - \Phi(t) \right| \le 6 s_n^{-3} \sum_{j=1}^n \mathbb{E}\big(|X_j - \eta_j|^3\big),$$

sobald die rechte Seite existiert.

Wir betrachten den originalen (und technischen) Beweis von Satz 10.4 bzw. Satz 10.5 nicht, sondern beginnen, die Steinsche Methode zu entwickeln:

Es gilt die folgende Charakterisierung der Standardnormalverteilung (siehe [Ste72]):

Lemma 10.6. *(Stein, 1972) Es sei Z eine Zufallsgröße mit Werten in $\mathbb{R}$. Z ist Standardnormalverteilt genau dann, wenn für jede stetige und stückweise stetig differenzierbare Funktion $f:\mathbb{R}\to\mathbb{R}$ mit $\int_{\mathbb{R}}|f'(x)|e^{-x^2/2}dx<\infty$ gilt:*

$$\mathbb{E}\{f'(Z)-Z\,f(Z)\}=0.$$

Beweis: Es sei Z Standardnormalverteilt und $\int_{\mathbb{R}}|f'(x)|e^{-x^2/2}dx<\infty$. Dann gilt mittels partieller Integration

$$\begin{aligned}\mathbb{E}(f'(Z)) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f'(w)\,e^{-w^2/2}\,dw\\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(z)\,z\,e^{-z^2/2}\,dz=\mathbb{E}(Z\,f(Z)).\end{aligned}$$

Nun nehmen wir an, dass für jede stetige und stückweise stetig differenzierbare Funktion f mit $\int_{\mathbb{R}}|f'(x)|e^{-x^2/2}\,dx<\infty$ gilt $\mathbb{E}\{f'(Z)-Z\,f(Z)\}=0$. Dann gilt es insbesondere für

$$f_{w_0}(y):=e^{y^2/2}\int_{-\infty}^{y}\big(h_{w_0}(x)-N(h_{w_0})\big)\,e^{-x^2/2}\,dx$$

mit

$$h_{w_0}(x)=\begin{cases}1, & \text{für } x\le w_0,\\ 0, & \text{für } x>w_0\end{cases}$$

und

$$N(h_{w_0}):=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} h_{w_0}(x)\,e^{-x^2/2}\,dx=\Phi(w_0).$$

Es gilt

$$\begin{aligned}f'_{w_0}(y) &= e^{y^2/2}\big(h_{w_0}(y)-N(h_{w_0})\big)\,e^{-y^2/2}\\ &\quad +y\,e^{y^2/2}\int_{-\infty}^{y}\big(h_{w_0}(x)-N(h_{w_0})\big)\,e^{-x^2/2}\,dx\\ &= h_{w_0}(y)-N(h_{w_0})+y\,f_{w_0}(y)\end{aligned}$$

(also löst $f_{w_0}(\cdot)$ die Differentialgleichung $f'(y)-y\,f(y)=h_{w_0}(y)-\Phi(w_0)$). Somit folgt

$$\begin{aligned}0 &= \mathbb{E}\{f'_{w_0}(Z)-Z\,f_{w_0}(Z)\}\\ &= \mathbb{E}\{h_{w_0}(Z)-N(h_{w_0})\}\\ &= P\big(Z\le w_0\big)-\Phi(w_0).\end{aligned}$$

Es bleibt zu zeigen, dass f_{w_0} stetig und stückweise stetig differenzierbar ist sowie $\int_{\mathbb{R}}|f'_{w_0}(x)|\,e^{-x^2/2}\,dx<\infty$ erfüllt. Wir stellen eine Bemerkung voran:

Bemerkung 10.7. Ist $h : \mathbb{R} \to \mathbb{R}$ eine stückweise stetige Funktion mit $\int_{\mathbb{R}} |h(x)|\, e^{-x^2/2}\, dx < \infty$ und ist

$$U_N(h)(y) := e^{y^2/2} \int_{-\infty}^{y} (h(x) - N(h))\, e^{-x^2/2}\, dx,$$

so gilt:

$$U_N'(h)(y) - y\, U_N(h)(y) = h(y) - N(h).$$

Man sagt, $U_N(h)(\cdot)$ ist Lösung der *Stein-Gleichung*

$$f'(y) - y\, f(y) = h(y) - N(h),$$

wobei

$$N(h) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} h(x)\, e^{-x^2/2}\, dx.$$

Da f_{w_0}' existiert, ist f_{w_0} stetig. Da $h(y) - N(h)$ und $y\, U_N(h)(y)$ stückweise stetige Funktionen sind, ist $U_N(h)(\cdot)$ stückweise stetig differenzierbar, also insbesondere f_{w_0}. Wenn wir noch zeigen, dass

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |U_N'(h)(x)|\, e^{-x^2/2}\, dx < \infty,$$

ist das Lemma bewiesen. $\int_{\mathbb{R}} |h(x) - N(h)|\, e^{-x^2/2}\, dx < \infty$ gilt nach Voraussetzung. Es gilt $\int_{\mathbb{R}} (h(x) - N(h))\, e^{-x^2/2}\, dx = 0$ nach Definition von $N(h)$. Damit ist

$$\begin{aligned} U_N(h)(w) &= e^{w^2/2} \int_{-\infty}^{w} (h(x) - N(h))\, e^{-x^2/2}\, dx \\ &= -e^{w^2/2} \int_{w}^{\infty} (h(x) - N(h))\, e^{-x^2/2}\, dx. \end{aligned}$$

Mittels partieller Integration folgt hieraus:

$$\begin{aligned} &\int_0^\infty |x\, U_N(h)(x)|\, e^{-x^2/2}\, dx \\ &\leq \int_0^\infty x \Big(\int_x^\infty |h(y) - N(h)|\, e^{-y^2/2}\, dy \Big)\, dx \\ &= \int_0^\infty |h(y) - N(h)|\, e^{-y^2/2}\, \frac{y^2}{2}\, dy. \end{aligned}$$

Das letzte Integral ist für $h = h_{w_0}$ endlich. Eine analoge Abschätzung gilt für das Integral über $\mathbb{R}_-$ und somit folgt das Lemma. $\square$

Das Lemma läßt vermuten: wenn $\mathbb{E}\{f'(W) - W\, f(W)\}$ klein ist für eine große Klasse von Funktionen f, so ist die Verteilung von W dicht bei der Standardnormalverteilung. Zu einer *glatten* Funktion h finden wir nach Bemerkung 10.7 eine Lösung

$U_N h(\cdot)$ der Steingleichung $f'(x) - x\, f(x) = h(x) - N(h)$ und es gilt

$$\mathbb{E}\{U_N(h)'(W) - W\, U_N(h)(W)\} = \mathbb{E}h(W) - N(h).$$

Wir suchen Schranken für die linke Seite und werden dabei die spezielle Struktur der Zufallsgröße W sowie Schranken für die Lösung $U_N(h)$ verwenden. Das folgende Lemma stammt aus [Ste86].

Lemma 10.8. *(Stein, 1986) Angenommen $h : \mathbb{R} \to \mathbb{R}$ ist eine beschränkte und differenzierbare Funktion und $U_N(h)$ die Lösung der Steingleichung. Dann gilt:*

$$\begin{aligned} ||U_N(h)|| &\leq \sqrt{\pi/2}\, ||h - N(h)|| && (10.9)\\ ||U_N(h)'|| &\leq 2||h - N(h)|| \quad \textit{und} && (10.10)\\ ||U_N(h)''|| &\leq 2||h'||, && (10.11) \end{aligned}$$

wobei $||f||$ die Supremumsnorm bezeichne: $||f|| := \sup_{x\in\mathbb{R}} |f(x)|$.

Wir geben den Beweis des Lemmas später und wollen zunächst die Funktionsweise der Steinschen Methode für die Situation aus Satz 10.3 vorführen: es sei $\sigma = 1$ und $\xi = \mathbb{E}|X_i|^3 < \infty$. Setze $W_n := S_n/\sqrt{n}$ und sei $h : \mathbb{R} \to \mathbb{R}$ stetig differenzierbar mit $||h'|| < \infty$. Wir bezeichnen die Steinlösung $U_N(h)$ mit f. Weiter sei

$$W_n^i = W_n - \frac{X_i}{\sqrt{n}}.$$

Wir entwickeln f via der Taylor-Formel um W_n^i:

$$f(W_n) - \big(f(W_n^i) + (W_n - W_n^i)\, f'(W_n^i)\big) = \int_{W_n^i}^{W_n} (W_n - t)\, f''(t)\, dt.$$

Daraus folgt

$$\mathbb{E}\bigg\{\frac{X_i}{\sqrt{n}} f(W_n) - \frac{X_i}{\sqrt{n}} f(W_n^i) - \frac{X_i^2}{n} f'(W_n^i)\bigg\} = \mathbb{E}\bigg\{\frac{X_i}{\sqrt{n}} \int_{W_n^i}^{W_n} (W_n - t)\, f''(t)\, dt\bigg\}.$$

Nun kann die rechte Seite nach oben abgeschätzt werden durch

$$\mathbb{E}\bigg|\frac{X_i}{\sqrt{n}} \int_{W_n^i}^{W_n} |W_n - W_n^i|\, ||f''||\, dt\bigg| = ||f''|| \frac{\mathbb{E}|X_i|^3}{n^{3/2}}.$$

Die Zufallsgrößen X_i und W_n^i sind stochastisch unabhängig, also ist $\mathbb{E}(X_i\, f(W_n^i)) = \mathbb{E}X_i\, \mathbb{E}f(W_n^i) = 0$ und $\mathbb{E}(X_i^2\, f'(W_n^i)) = \mathbb{E}X_i^2\, \mathbb{E}f'(W_n^i) = \mathbb{E}f'(W_n^i)$. Somit erhalten wir:

$$\bigg|\mathbb{E}\bigg(\frac{X_i}{\sqrt{n}} f(W_n)\bigg) - \frac{1}{n}\mathbb{E}f'(W_n^i)\bigg| \leq ||f''|| \frac{\xi}{n^{3/2}}.$$

Wir können nun $|\mathbb{E}\, h(W_n) - N(h)|$ wie folgt abschätzen:

$$
\begin{aligned}
& |\mathbb{E}(f'(W_n) - W_n\, f(W_n))| \\
\leq \quad & \left|\mathbb{E}\left\{ f'(W_n) - \tfrac{1}{n}\textstyle\sum_{i=1}^n f'(W_n^i)\right\}\right| + \left|\mathbb{E}\left\{ \tfrac{1}{n}\textstyle\sum_{i=1}^n f'(W_n^i) - \sum_{i=1}^n \tfrac{X_i}{\sqrt{n}} f(W_n)\right\}\right| \\
\leq \quad & \tfrac{1}{n}\textstyle\sum_{i=1}^n \mathbb{E}|f'(W_n) - f'(W_n^i)| + \sum_{i=1}^n \left| \tfrac{1}{n}\mathbb{E}f'(W_n^i) - \mathbb{E}\left(\tfrac{X_i}{\sqrt{n}} f(W_n)\right)\right|.
\end{aligned}
$$

Der erste Ausdruck ist mittels des Mittelwertsatzes durch

$$
\frac{1}{n}\sum_{i=1}^n ||f''|| \frac{\mathbb{E}|X_i|}{\sqrt{n}}
$$

abschätzbar. Jensen's Ungleichung liefert $\mathbb{E}|X_i| \leq \sqrt{\mathbb{E}X_i^2} = 1$ und $1 = (\mathbb{E}X_i^2)^{1/2} \leq \big(\mathbb{E}|X_i|^3\big)^{1/3}$, also

$$
\frac{1}{n}\sum_{i=1}^n ||f''|| \frac{\mathbb{E}|X_i|}{\sqrt{n}} \leq ||f''|| \frac{\xi}{\sqrt{n}}.
$$

Wir verwenden nun die Abschätzung $||f''|| \leq 2||h'||$ aus Lemma 10.8 und erhalten insgesamt

$$
|\mathbb{E}h(W_n) - N(h)| \leq 4\, ||h'|| \frac{\xi}{\sqrt{n}}
$$

als Schranke vom *Berry-Esséen-Typ*.

Man beachte bei den Rechnungen, dass wir tatsächlich nur von der Unabhängigkeit der Zufallsgrößen Gebrauch gemacht haben. Zwar war die Rechnung übersichtlich bei der Annahme gleicher Erwartungswerte und gleicher Varianzen $\sigma = 1$, doch identische Verteilungen wurden nicht benötigt. Die obige Herleitung kann leicht verallgemeinert werden für den Fall unterschiedlicher Erwartungswerte, Varianzen und dritter Momente.

Das so elementar gefundene Resultat muß noch eingeordnet werden:

Das klassische Berry-Esséen Resultat (Satz 10.4) erhalten wir nicht ganz: Zunächst können wir aber aus der via Stein gezeigten schwachen Konvergenz die Konvergenz der Verteilungsfunktionen folgern: Es gilt allgemein (siehe Satz 10.16 am Ende dieses Kapitels): Es seien $\mu_n, \mu \in \mathcal{M}_1(\mathbb{R})$ für $n \in \mathbb{N}$ mit Verteilungsfunktionen F_n (von μ_n) beziehungsweise F (von μ). Es gilt genau dann $\mu_n \xrightarrow{w} \mu$, wenn für jedes $t \in \mathbb{R}$, in dem F stetig ist, $\lim_{n\to\infty} F_n(t) = F(t)$ ist. Satz 10.3 (hier via der Steinschen Methode bewiesen) besagt somit:

Satz 10.12. *Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsgrößen, die die Bedingungen von Satz 10.3 erfüllen. Dann gilt für alle $t \in \mathbb{R}$:*

$$\lim_{n\to\infty} P\left(\frac{S_n}{\sqrt{n}\sigma} \le t\right) = \Phi(t)$$

(denn $\Phi(\cdot)$ ist überall stetig).

Wie kann nun die mittels Steinscher Methode gewonnene Abschätzung der Größenordnung $\frac{1}{\sqrt{n}}$ auf eine Abschätzung für $\left|P\big(S_n/(\sqrt{n}\sigma) \le t\big) - \Phi(t)\right|$ übertragen werden? Hier sind die Testfunktionen Indikatoren auf Halbgeraden, also keine geeigneten Testfunktionen, um die Stein-Technik anwenden zu können. Wir wenden $|\mathbb{E}h(W_n) - N(h)| \le 4\,||h'||\frac{\xi}{\sqrt{n}}$ auf spezielle Funktionen $h = h_{w_0,\Delta}$ an:

$$h_{w_0,\Delta}(w) := \begin{cases} 1, & w \le w_0, \\ 1 - \frac{w-w_0}{\Delta}, & w_0 \le w \le w_0 + \Delta, \\ 0, & w \ge w_0 + \Delta. \end{cases} \tag{10.13}$$

Dann gilt $P(W \le w_0) \le \mathbb{E}h_{w_0,\Delta}(W)$ und $\mathbb{E}h_{w_0-\Delta,\Delta}(W) \le P(W \le w_0)$. Nun folgt

$$\begin{aligned} P(W \le w_0) &\le \mathbb{E}h_{w_0,\Delta}(W) \le N(h_{w_0,\Delta}) + 4\,||h'_{w_0,\Delta}||\frac{\xi}{\sqrt{n}} \\ &\le \Phi(w_0) + \frac{\Delta}{\sqrt{2\pi}} + 4\,||h'_{w_0,\Delta}||\frac{\xi}{\sqrt{n}}. \end{aligned}$$

Da $\sup_{x\in\mathbb{R}} |h'_{w_0,\Delta}(x)| \le \frac{1}{\Delta}$, erhalten wir

$$P(W \le w_0) - \Phi(w_0) \le \frac{\Delta}{\sqrt{2\pi}} + \frac{4}{\Delta}\frac{\xi}{\sqrt{n}}.$$

Diese Abschätzung können wir nun in Δ minimieren: Wähle dazu $\Delta = 2\sqrt{\xi}\left(\frac{2\pi}{n}\right)^{1/4}$. Es folgt

$$P(W \le w_0) - \Phi(w_0) \le \frac{2\sqrt{\xi}(2\pi)^{1/4}}{\sqrt{2\pi}n^{1/4}} + \frac{4\xi}{2\sqrt{\xi}(2\pi)^{1/4}n^{1/4}} = \mathcal{O}(1/n^{1/4}).$$

Wir erhalten die analoge Ordnung der Schranke für $\Phi(w_0) - P(W \le w_0)$. Somit kommen wir an die optimale Ordnung $1/\sqrt{n}$ im Berry-Esséen Theorem nicht heran. Tatsächlich gibt es modifizierte Techniken, die im Fall unabhängiger Zufallsgrößen via der Steinschen Methode $1/\sqrt{n}$ als Konvergenzrate liefern. Bolthausen hat in [Bol84] einen hübschen induktiven Beweis vorgestellt, der im Ausblick skizziert wird. Dieser Beweis liefert direkt den allgemeinen Fall aus Satz 10.5, also den Zentralen Grenzwertsatz für unabhängige, aber nicht notwendig identisch verteilte, Zufallsgrößen und die Berry-Esséen Schranken! Wir beschränken uns hier auf die obige Diskussion und liefern nun den Beweis für die Schranken in Lemma 10.8:

Beweis von Lemma 10.8: Es gilt $\int_{\mathbb{R}}(h(x) - N(h))\, e^{-x^2/2}\, dx = 0$ nach Definition von $N(h)$. Damit ist

$$\begin{aligned} U_N(h)(w) &= e^{w^2/2} \int_{-\infty}^{w} (h(x) - N(h))\, e^{-x^2/2}\, dx \\ &= -\, e^{w^2/2} \int_{w}^{\infty} (h(x) - N(h))\, e^{-x^2/2}\, dx. \end{aligned}$$

Zum Beweis der ersten Abschätzung $||U_N(h)|| \leq \sqrt{\pi/2}\, ||h - N(h)||$: für $w \leq 0$ ist

$$|U_N(h)(w)| \leq \sup_{x \leq 0} |h(x) - N(h)| \left(\int_{-\infty}^{w} e^{-x^2/2}\, dx \right) e^{w^2/2}$$

und für $w \geq 0$ ist

$$|U_N(h)(w)| \leq \sup_{x \geq 0} |h(x) - N(h)|\, e^{w^2/2} \int_{w}^{\infty} e^{-x^2/2}\, dx.$$

Da $\frac{d}{dw}\left(e^{w^2/2} \int_{-\infty}^{w} e^{-x^2/2}\, dx \right) = 1 + w\, e^{w^2/2} \int_{-\infty}^{w} e^{-x^2/2}\, dx$ und $\Phi(w) < \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{w} -\frac{x}{|w|}\, e^{-x^2/2}\, dx = \frac{e^{-w^2/2}}{|w|\sqrt{2\pi}}$, folgt $\frac{d}{d\omega}\left(e^{w^2/2} \int_{-\infty}^{w} e^{-x^2/2}\, dx \right) > 0$. Somit wird das Maximum der Abschätzung für $|U_N(h)(w)|$ bei $w = 0$ angenommen. Ausgerechnet liefert dies

$$||U_N(h)|| \leq \sqrt{\frac{\pi}{2}} ||h - N(h)||.$$

Zum Beweis der zweiten Abschätzung $||U_N(h)'|| \leq 2||h - N(h)||$: Mittels der Steingleichung für $U_N(h)$ folgt für $w \geq 0$:

$$\sup_{w \geq 0} |(U_N(h))'(w)| \leq ||h - N(h)|| \left(1 + \sup_{w \geq 0} w\, e^{w^2/2} \int_{w}^{\infty} e^{-x^2/2}\, dx \right).$$

Da $1 - \Phi(w) < \frac{1}{\sqrt{2\pi}} \int_{w}^{\infty} \frac{x}{w}\, e^{-x^2/2}\, dx = \frac{e^{-w^2/2}}{w\,\sqrt{2\pi}}$, folgt

$$\sup_{w \geq 0} |(U_N(h))'(w)| \leq 2\, ||h - N(h)||.$$

Mit $h^\star(w) := h(-w)$ folgt $U_N(h^\star)(w) = U_N(h)(-w)$ und somit folgt die zweite Ungleichung.

Zum Beweis der dritten Ungleichung $||U_N(h)''|| \leq 2||h'||$:

Wir leiten $(U_N(h))'(w) - w\, U_N(h)(w) = h(w) - N(h)$ ab:

$$\begin{aligned} (U_N(h))''(w) &= U_N(h)(w) + w(U_N(h))'(w) + h'(w) \\ &= (1 + w^2)\, U_N(h)(w) + w\,(h(w) - N(h)) + h'(w). \end{aligned}$$

Wir drücken nun $(U_N(h))''$ in Termen von h' aus. Mittels partieller Integration folgt

$$
\begin{aligned}
& h(x) - N(h) \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (h(x) - h(y))\, e^{-y^2/2}\, dy \\
&= \frac{1}{\sqrt{2\pi}} \left\{ \int_{-\infty}^{x} \left(\int_{y}^{x} h'(z)\, dz \right) e^{-y^2/2}\, dy - \int_{x}^{\infty} \left(\int_{x}^{y} h'(z)\, dz \right) e^{-y^2/2} dy \right\} \\
&= \frac{1}{\sqrt{2\pi}} \left\{ \int_{-\infty}^{x} h'(z) \left(\int_{-\infty}^{z} e^{-y^2/2}\, dy \right) - \int_{x}^{\infty} h'(z) \left(\int_{z}^{\infty} e^{-y^2/2}\, dy \right) dz \right\} \\
&= \int_{-\infty}^{x} h'(z)\, \Phi(z)\, dz - \int_{x}^{\infty} h'(z)(1 - \Phi(z))\, dz.
\end{aligned}
$$

Somit folgt

$$
\begin{aligned}
& (U_N(h))(w) \\
&= e^{w^2/2} \int_{-\infty}^{w} (h(x) - N(h))\, e^{-x^2/2}\, dx \\
&= e^{w^2/2} \int_{-\infty}^{w} \left(\int_{-\infty}^{x} h'(z)\, \Phi(z)\, dz - \int_{x}^{\infty} h'(z)(1 - \Phi(z))\, dz \right) e^{-x^2/2}\, dx \\
&= e^{w^2/2} \Bigg(\int_{-\infty}^{w} h'(z)\, \Phi(z) \left(\int_{z}^{w} e^{-x^2/2}\, dx \right) dz \\
&\qquad - \int_{-\infty}^{w} h'(z)(1 - \Phi(z)) \left(\int_{-\infty}^{z} e^{-x^2/2}\, dx \right) dz \\
&\qquad - \int_{w}^{\infty} h'(z)(1 - \Phi(z)) \left(\int_{-\infty}^{w} e^{-x^2/2}\, dx \right) dz \Bigg) \\
&= -\sqrt{2\pi}\, e^{w^2/2} \left\{ (1 - \Phi(w)) \int_{-\infty}^{w} h'(z)\, \Phi(z)\, dz + \Phi(w) \int_{w}^{\infty} h'(z)(1 - \Phi(z))\, dz \right\}.
\end{aligned}
$$

Wir erhalten

$$
\begin{aligned}
(U_N(h))''(w) &= h'(w) + (w - \sqrt{2\pi}(1 + w^2)\, e^{w^2/2} (1 - \Phi(w)) \int_{-\infty}^{w} h'(z)\, \Phi(z)\, dz \\
&\quad + (-w - \sqrt{2\pi}(1 + w^2)\, e^{w^2/2}\, \Phi(w)) \int_{w}^{\infty} h'(z)(1 - \Phi(z))\, dz.
\end{aligned}
$$

Nun verwenden wir

$$
w + \sqrt{2\pi}(1 + w^2)\, e^{w^2/2}\, \Phi(w) > 0 \tag{10.14}
$$

und

$$
-w + \sqrt{2\pi}(1 + w^2)\, e^{w^2/2} (1 - \Phi(w)) > 0 \tag{10.15}
$$

Die erste Ungleichung gilt für alle $w \geq 0$. Man nehme also an, dass $w < 0$ ist. Mittels partieller Integration stimmt dann die linke Seite in (10.14) überein mit

$$\begin{aligned} & w + e^{w^2/2} \int_{-\infty}^{w} e^{-x^2/2}\, dx + w^2 e^{w^2/2} \left(-w^{-1} e^{-w^2/2} - \int_{-\infty}^{w} x^{-2} e^{-x^2/2} dx \right) \\ & = e^{w^2/2} \int_{-\infty}^{w} \left(1 - \frac{w^2}{x^2} \right) e^{-x^2/2} dx > 0. \end{aligned}$$

Wendet man nun (10.14) auf $-w$ an, so erhält man (10.15).

Weiter ist

$$\int_{-\infty}^{w} \Phi(z)\, dz = w\, \Phi(w) + \frac{e^{-w^2/2}}{\sqrt{2\pi}}$$

und

$$\int_{w}^{\infty} (1 - \Phi(z))\, dz = -w(1 - \Phi(w)) + \frac{e^{-w^2/2}}{\sqrt{2\pi}}.$$

Mit der Positivität all dieser Ausdrücke folgt nun:

$$\begin{aligned} & ||(U_N(h))''|| \\ \leq \; & \Bigg(1 + \sup_w \Bigg(\left[-w + \sqrt{2\pi}(1 + w^2)\, e^{w^2/2} (1 - \Phi(w)) \right] \left[w\, \Phi(w) + \frac{e^{-w^2/2}}{\sqrt{2\pi}} \right] \\ & + \left[w + \sqrt{2\pi}(1 + w^2)\, e^{w^2/2}\, \Phi(w) \right] \left[-w(1 - \Phi(w)) + \frac{e^{-w^2/2}}{\sqrt{2\pi}} \right] \Bigg) \Bigg) \, ||h'|| \\ = \; & 2\, ||h'||, \end{aligned}$$

womit die dritte Abschätzung bewiesen ist. □

Wie angekündigt geben wir nun einen zentralen Satz aus der Theorie der schwachen Konvergenz von Wahrscheinlichkeitsmaßen mit Beweis an.

Satz 10.16. *Es seien $\mu_n, \mu \in \mathcal{M}_1(\mathbb{R}), n \in \mathbb{N}$. Es bezeichnen $F_n, F, n \in \mathbb{N}$, die zu μ_n und μ zugehörigen Verteilungsfunktionen. Dann sind die folgenden Aussagen zueinander äquivalent:*

(i) *μ_n konvergiert schwach gegen μ.*

(ii) *Für alle unendlich differenzierbaren Funktionen ϕ, die außerhalb eines endlichen Intervalls verschwinden, gilt*

$$\lim_{n\to\infty} \int_{\mathbb{R}} \phi\, d\mu_n = \int_{\mathbb{R}} \phi\, d\mu. \tag{10.17}$$

(iii) *Es gilt*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

für alle x, in denen F stetig ist.

Beweis: Zunächst zeigen wir, dass $\mu_n \xrightarrow{w} \mu$, wenn (10.17) für alle stetigen Funktionen ϕ, die außerhalb eines endlichen Intervalls verschwinden, gilt. Sei dazu ϕ eine beliebige, stetig beschränkte Funktion, $|\phi(x)| \leq c$ für alle $x \in \mathbb{R}$. Zu $\varepsilon > 0$ existiert ein $N \in \mathbb{N}$ mit

$$\mu(\{|x| \geq N\}) < \frac{\varepsilon}{2c}.$$

Sei Θ_N und $\Theta_{N'}$ gegeben durch

$$\Theta_N(x) = \varphi(|x, [-N, N]|)$$

und

$$\Theta_{N'}(x) = \varphi(|x, [-N-1, N+1]|),$$

wobei $|x, [a,b]|$ den gewöhnlichen Abstand von x zum abgeschlossenen Intervall $[a,b]$ bezeichnet und

$$\varphi(t) := \begin{cases} 1, & t \leq 0, \\ 1-t, & 0 \leq t \leq 1, \\ 0, & 1 \leq t. \end{cases}$$

Dann gilt

$$\begin{aligned} 1 - \frac{\varepsilon}{2c} &< \mu(\{|x| \leq N\}) \leq \int_{\mathbb{R}} \Theta_N(x)\, d\mu(x) \\ &= \liminf_{n\to\infty} \int_{\mathbb{R}} \Theta_N(x)\, d\mu_n(x) \leq \liminf_{n\to\infty} \mu_n(\{|x| \leq N+1\}), \end{aligned}$$

also

$$\limsup_{n\to\infty} \mu_n(\{|x| > N+1\}) = 1 - \liminf_{n\to\infty} \mu_n(\{|x| \leq N+1\}) < \frac{\varepsilon}{2c}.$$

Betrachte $\phi_N := \phi\, \Theta'_N$, so ist $\phi_N = \phi$ auf der Menge $\{|x| \leq N+1\}$ und $|\phi(x)| \leq c$ auf $\{|x| > N+1\}$. Somit erhalten wir:

$$\begin{aligned} &\limsup_{n\to\infty} \left| \int_{\mathbb{R}} \phi\, d\mu_n - \int_{\mathbb{R}} \phi\, d\mu \right| \\ &\leq \limsup_{n\to\infty} \left| \int_{\mathbb{R}} \phi_N\, d\mu_n - \int_{\mathbb{R}} \phi_N\, d\mu \right| \\ &\quad + \limsup_{n\to\infty} \Big(c\mu_n(\{|x| > N+1\}) + c\mu(\{|x| > N+1\}) \Big) \\ &= \limsup_{n\to\infty} c\, \mu_n(\{|x| > N+1\}) + c\mu(\{|x| > N+1\}) \\ &< c\frac{\varepsilon}{2c} + c\frac{\varepsilon}{2c} = \varepsilon, \end{aligned}$$

also folgt $\lim_{n\to\infty} \int_{\mathbb{R}} \phi\, d\mu_n = \int_{\mathbb{R}} \phi\, d\mu$. Nun zeigen wir mit der gerade bewiesenen Hilfsaussage (ii) $\Rightarrow$ (i):

Sei dazu $\varepsilon > 0$ und

$$\varrho_\varepsilon(x) := \begin{cases} d(\varepsilon)\exp\left\{-\frac{1}{(1-x^2/\varepsilon^2)}\right\}, & \text{für } |x| < \varepsilon, \\ 0, & \text{sonst.} \end{cases}$$

Dabei ist $d(\varepsilon)$ so gewählt, dass $\int \varrho_\varepsilon(x)\,dx = 1$ ist. $\varrho_\varepsilon(x)$ ist unendlich oft differenzierbar (Übung). Sei ϕ eine stetige Funktion, die außerhalb eines endlichen Intervalls verschwindet. So ist ϕ gleichmäßig stetig, also $\delta(\varepsilon) = \sup\{|\phi(x)-\phi(y)| : |x-y| \le \varepsilon\} \to 0$ für $\varepsilon \searrow 0$.

Definiere nun

$$\phi^\varepsilon(x) := \int_{-\varepsilon}^{\varepsilon} \phi(x-y)\,\varrho_\varepsilon(y)\,dy.$$

$\phi^\varepsilon(x)$ ist ein „Durchschnitt"für Werte von ϕ innerhalb des Intervalls $(x-\varepsilon, x+\varepsilon)$, also $|\phi^\varepsilon(x) - \phi(x)| \le \delta(\varepsilon)$ für alle ε. Daher ist $|\int_{\mathbb{R}} \phi\,d\mu_n - \int_{\mathbb{R}} \phi^\varepsilon\,d\mu_n| \le \delta(\varepsilon)$ für alle $n \in \mathbb{N}$ und $|\int_{\mathbb{R}} \phi\,d\mu - \int_{\mathbb{R}} \phi^\varepsilon\,d\mu| \le \delta(\varepsilon)$. Also

$$\begin{aligned} \left|\int_{\mathbb{R}} \phi\,d\mu_n - \int_{\mathbb{R}} \phi\,d\mu\right| &\le \left|\int_{\mathbb{R}} \phi\,d\mu_n - \int_{\mathbb{R}} \phi^\varepsilon\,d\mu_n\right| \\ &\quad + \left|\int_{\mathbb{R}} \phi^\varepsilon\,d\mu_n - \int_{\mathbb{R}} \phi^\varepsilon\,d\mu\right| + \left|\int_{\mathbb{R}} \phi^\varepsilon\,d\mu - \int_{\mathbb{R}} \phi\,d\mu\right| \\ &\le 2\,\delta(\varepsilon) + \left|\int_{\mathbb{R}} \phi^\varepsilon\,d\mu_n - \int_{\mathbb{R}} \phi^\varepsilon\,d\mu\right| \to 2\,\delta(\varepsilon) \end{aligned}$$

für $n \to \infty$. Da $\delta(\varepsilon) \to 0$ für $\varepsilon \searrow 0$, folgt die Behauptung.

(i) $\Rightarrow$ (ii) ist unmittelbar klar.

Nun zeigen wir (i) $\Rightarrow$ (iii):

Sei x_0 ein Stetigkeitspunkt von F. Zu $\varepsilon > 0$ existiert ein $\eta(\varepsilon) > 0$ mit $|F(x) - F(x_0)| < \varepsilon$ für $|x - x_0| \le \eta(\varepsilon)$. Sei nun $\psi_\varepsilon^+(x) = 1$ für $x \le x_0$ und $= 0$ für $x \ge x_0 + \eta(\varepsilon)$ und $\psi_\varepsilon^+(x)$ für $x_0 < x < x_0 + \eta(\varepsilon)$ die lineare Interpolation. Sei analog $\psi_\varepsilon^-(x) = 1$ für $x \le x_0 - \eta(\varepsilon)$, $\psi_\varepsilon'(x) = 0$ für $x \ge x_0$ und die lineare Interpolation im Intervall $(x_0 - \eta(\varepsilon), x_0)$. Dann gilt

$$\begin{aligned} \limsup_{n\to\infty} F_n(x_0) &\le \limsup_{n\to\infty} \int \psi_\varepsilon^+(x)\,d\mu_n(x) \\ &= \int \psi_\varepsilon^+(x)\,d\mu(x) \le F(x_0 + \eta(\varepsilon)) < F(x_0) + \varepsilon \end{aligned}$$

und

$$\begin{aligned}\liminf_{n\to\infty} F_n(x_0) &\geq \liminf_{n\to\infty} \int \psi_\varepsilon^-(x)\, d\mu_n(x)\\ &= \int \psi_\varepsilon^-(x)\, d\mu(x) \geq F(x_0-\eta(\varepsilon)) > F(x_0)-\varepsilon.\end{aligned}$$

Da $\varepsilon > 0$ beliebig, folgt

$$\limsup_{n\to\infty} F_n(x_0) \leq F(x_0) \leq \liminf_{n\to\infty} F_n(x_0),$$

also $\lim_{n\to\infty} F_n(x_0) = F(x_0)$.

Bleibt (iii) $\Rightarrow$ (i) zu zeigen: Sei f eine stetige Funktion, die außerhalb von $[a,b]$ verschwindet und a,b seien Stetigkeitspunkte von F. Zerlege $[a,b]$ in endlich viele Teilintervalle, deren Endpunkte alle Stetigkeitspunkte von F sind und approximiere f durch eine Treppenabbildung, konstant in jedem Teilintervall. Dann folgt nach Voraussetzung, dass (10.17) für die Treppenabbildungen gilt. Via Dreiecksungleichung folgt dann wie gewohnt der Satz. □

Ausblick: Wir hatten bereits erwähnt, dass Bolthausen in [Bol84] mittels der Steinschen Methode die optimale Konvergenzrate $n^{-1/2}$ bekommt. Tatsächlich ist der Aufwand seines Beweises für uns an dieser Stelle gar nicht mehr so groß. Die wichtigsten Vorbereitungen sind getroffen. Wir beschränken uns trotzdem auf eine Skizze und verweisen auf ein im WS 2003/04 entstehendes Vorlesungsskript zur Wahrscheinlichkeitstheorie, in dem der Beweis von Bolthausen ausgeführt wird. In der Monographie von Stroock [Str93] ist Bolthausens Beweisidee dargestellt, jedoch nur schwerlich zu verdauen!

Wie bereits erwähnt, wird in [Bol84] der Fall betrachtet, in dem die Zufallsgrößen zwar unabhängig, aber nicht notwendig identisch verteilt sind. Man bezeichne mit $L(n,\gamma)$ die Menge der n-Tupel $X = (X_1,\dots,X_n)$ von *unabhängigen* reellwertigen Zufallsvariablen mit $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$ und $\mathbb{E}|X_i|^3 \leq \gamma$. Mit der Jensenschen Ungleichung gilt $\mathbb{E}|X_i|^3 = \mathbb{E}(X_i^2)^{3/2} \geq (\mathbb{E}X_i^2)^{3/2}$. Also muß man $\gamma \geq 1$ annehmen, damit die Menge $L(n,\gamma)$ nichtleer ist. Die Zufallsgrößen müssen *nicht identisch verteilt* sein. Wir setzen wieder $S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ und wählen die Funktion $h_{z,\lambda}$ für $z \in \mathbb{R}$ und $\lambda \geq 0$ wie in (10.13). Weiter sei $\delta(\lambda,\gamma,n) := \sup\{|\mathbb{E}h_{z,\lambda}(S_n) - \Phi(h_{z,\lambda})| : z \in \mathbb{R}, X \in L(n,\gamma)\}$ und $\delta(\gamma,n) := \delta(0,\gamma,n)$. Bolthausen beweist in [Bol84]:

$$\text{Wenn} \quad \gamma, n \geq 1, \quad \text{so gilt} \quad \delta(\gamma,n) \leq 25\frac{\gamma}{\sqrt{n}}.$$

Wir setzen $h := h_{z,\lambda}$ und analog für die Lösung der korrespondierenden Steinschen Gleichung $f := f_{z,\lambda}$. Weiter setzten wir

$$S_{n,i} := \frac{1}{\sqrt{n}} \sum_{j \neq i} X_j.$$

Wie oben müssen wir die folgenden Größen abschätzen:

$$\begin{aligned} \mathbb{E}h(S_n) - \Phi(h) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}\big(f'(S_n) - f'(S_{n,i})\big) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \mathbb{E}\bigg(X_i^2 \int_0^1 \Big(f'\big(S_{n,i} + t\frac{X_i}{\sqrt{n}}\big) - f'(S_{n,i})\Big)\, dt\bigg). \end{aligned} \tag{10.18}$$

Wieder wurde nur die Steinsche Gleichung, die Bedingungen $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$ und die Tatsache, dass X_i unabhängig von $S_{n,i}$ ist, verwendet. Weiter wurde die Differenz $f(S_n) - f(S_{n,i})$ als Integral der Ableitung von f in den Grenzen $S_{n,i}$ und S_n dargestellt. Bolthausen verwendet nun Teile von Lemma 10.8: Es gilt $|f(x)| \leq 3$ und $|f'(x)| \leq 2$. Anstelle einer Abschätzung der zweiten Ableitung von f leitet er her:

$$|f'(x+\varepsilon) - f'(x)| \leq |\varepsilon| \bigg(3 + 2|x| + \frac{1}{\lambda} \int_0^1 I_{(z,z+\lambda]}(x + s\varepsilon)\, ds \bigg). \tag{10.19}$$

Der Beweis ist ein Zweizeiler. Weiter zeigt er (ebenfalls ein Zweizeiler)

$$|\mathbb{E}I_{(a,b]}(S_{n,i} + c)| \leq \sqrt{\frac{n}{2\pi(n-1)}}(b-a) + 2\delta(\gamma, n-1) \tag{10.20}$$

für jedes Intervall $(a, b]$ und jede Konstante c. Dies folgt einfach aus der Definition von δ, aus der die Schranke

$$\big|\mathbb{E}h_{z,0}\big(\sqrt{\frac{n}{n-1}} S_{n,i}\big) - \Phi h_{z,0}\big| \leq \delta(\gamma, n-1)$$

folgt. Die Abschätzung (10.20) ist ganz zentral, wird sie doch eine Abschätzung von $\delta(\lambda, \gamma, n)$ mittels $\delta(\gamma, n-1)$ liefern. Dies führt dann zu einem Induktions-Argument. Weiter benötigt man noch $\delta(\gamma, n) \leq \delta(\lambda, \gamma, n) + \frac{\lambda}{\sqrt{2\pi}}$, was unmittelbar aus der Definition der Abbildung $h_{z,\lambda}$ und $h_{z,0}$ folgt. Für den ersten Ausdruck in (10.18) ergibt sich somit unter Verwendung von (10.19) und (10.20)

$$\mathbb{E}|f'(S_n) - f(S_{n,i})| \leq \frac{1}{\sqrt{n}} \bigg(6 + \frac{2}{\lambda} \delta(\gamma, n-1) \bigg), \quad n \geq 2.$$

Für den zweiten Ausdruck erfolgt analog für alle $n \geq 2$

$$\mathbb{E}\big| X_i^2 \int_0^1 \big(f'\big(S_{n,i} + t\frac{X_i}{\sqrt{n}}\big) - f'(S_{n,i})\big)\, dt \big| \leq \frac{\gamma}{\sqrt{n}} \bigg(3 + \frac{1}{\lambda} \delta(\gamma, n-1) \bigg).$$

Dies liefert insgesamt für $n \geq 2$

$$\delta(\lambda, \gamma, n) \leq \frac{\gamma}{\sqrt{n}}\Big(9 + \frac{3}{\lambda}\delta(\gamma, n-1)\Big).$$

Somit folgt

$$\delta(\gamma, n) \leq \frac{\gamma}{\sqrt{n}}\Big(9 + \frac{3}{\lambda}\delta(\gamma, n-1)\Big) + \frac{\lambda}{\sqrt{2\pi}}.$$

Dabei ist λ noch frei wählbar. Setzt man $\lambda = 6\sqrt{2}\gamma n^{-1/2}$, so erhält man

$$\delta(\gamma, n) \leq 12.5\frac{\gamma}{\sqrt{n}} + \frac{1}{2\sqrt{2}}\delta(\gamma, n-1).$$

Der Satz folgt nun einfach via Induktion: Es ist $\delta(\gamma, 1) \leq 1$ im Fall $n = 1$. Es gelte der Satz für $n-1$, dann folgt aus der letzten Ungleichung

$$\delta(\gamma, n) \leq 12.5\frac{\gamma}{\sqrt{n}} + \frac{25\gamma}{2\sqrt{2}\sqrt{n-1}} \leq 12.5\frac{\gamma}{\sqrt{n}} + 12.5\frac{\gamma}{\sqrt{n}}.$$

Wie bei der Poisson-Approximation erwartet man auch bei der Normal-Approximation die Unterscheidung des lokalen Ansatzes und des Kopplungsansatzes. Wir haben hier den lokalen Ansatz vorgestellt. In der Übersichtsarbeit von Reinert [Rei98] findet man im zweiten Kapitel einen Überblick zum lokalen Ansatz. Zentrales Anliegen der Arbeit von Reinert ist die Vorstellung unterschiedlicher *Kopplungsansätze* bei Normal-Approximation und deren Anwendungen. Man betrachte dazu die Kapitel 3, 4 und 5 in [Rei98] und die vielen Referenzen.

Barbour [Bar90] und Götze [Göt91] haben allgemeinere Resultate bewiesen. Barbour untersucht Diffusions-Approximationen während Götze den multivariaten Fall der Normal-Approximation mittels der Steinschen Methode untersucht. In beiden Arbeiten wird die Generator-Methode von Barbour verwendet. Dabei wird die linke Seite der Steinschen Gleichung, also $xf(x) - f'(x)$, umgeschrieben zu $xf'(x) - f''(x)$, und dies kann als $\mathcal{A}f$ geschrieben werden, wobei $\mathcal{A}$ der Generator des Ornstein-Uhlenbeck Prozesses ist.

Jüngste Arbeiten von Chen und Shao [CS01], [CS03] beschäftigen sich mit der Herleitung *nicht-uniformer* Berry-Esséen Schranken, sowie mit uniformen und nicht-uniformen Schranken für Normal-Approximation unter lokaler Abhängigkeit. Nicht-uniforme Berry-Esséen Schranken in der Situation des Satzes von Lindeberg und Feller sind: Für alle $z \in \mathbb{R}$ gilt

$$\Big|P\Big(\frac{1}{s_n}\sum_{j=1}^n (X_j - \eta_j) \leq z\Big) - \Phi(z)\Big| \leq C(1+|z|)^{-3} s_n^{-3} \sum_{j=1}^n \mathbb{E}\big(|X_j - \eta_j|^3\big).$$

11. Normal-Approximation für Zufallsgraphen

Für das Modell eines Zufallsgraphen $\mathcal{G}(n,p)$ ergab sich in den Kapiteln 5 und 6 das folgende Bild:

Es sei H ein (strikt) *balancierter* Graph mit k Ecken und $l \geq 1$ Kanten und $X(G)$ die Anzahl der zu H isomorphen Teilgraphen von $G \in \mathcal{G}(n,p)$. Dann hatten wir bewiesen:

(i) Die Grapheneigenschaft $\mathcal{A}_H$, eine Kopie von H als Teilgraphen zu enthalten, hat $t(n) = n^{-k/l}$ als Schwellenfunktion (gilt für alle H, die balanciert sind).

(ii) Ist $p(n) \ll n^{-k/l}$, so ist $\mathcal{L}(X) \xrightarrow{w} \delta_0$. Dies gilt für alle balancierten Graphen H.

(iii) Ist H *nicht balanciert*, so ist $t(n) = n^{-k/l}$ *keine Schwellenfunktion* für $\mathcal{A}_H$.

(iv) Ist H *strikt balanciert* und $p(n)$ so gewählt, dass $\lim_{n\to\infty} p(n)\, n^{k/l} = c > 0$ gilt, so folgt
$$\lim_{n\to\infty} d_{TV}\big(\mathcal{L}(X), Po(E(X))\big) = 0.$$
Die Konvergenzrate hierzu wurde in Satz 6.10 im Detail beschrieben.

Was passiert nun oberhalb der Schwellenfunktion?

Mit Hilfe der Janson-Ungleichung hatten wir gesehen, dass sich in einem genauer spezifizierbaren Bereich oberhalb der Schwelle die Poisson-Approximation erhalten läßt. Zumindest für die Approximation der Wahrscheinlichkeit $P(X=0)$ hatten wir dies gesehen.

Wir betrachten jetzt
$$W := \frac{X - \mathbb{E}X}{\sqrt{V(X)}}$$
und wollen die Normal-Approximation beweisen. Ziel dieses Kapitels ist der Beweis des folgenden Satzes, siehe [BKR89]:

Satz 11.1. *(Normal-Approximation, Barbour, Karoński und Ruciński, 1989) Für jede beschränkte Funktion $f:\mathbb{R}\to\mathbb{R}$ mit beschränkter erster und zweiter Ableitung gilt:*
$$\big|\mathbb{E}(W\, f(W) - f'(W))\big| \leq C\, k(H) \begin{cases} \psi^{-1/2}, & \textit{für } p \leq \frac{1}{2}, \\ n^{-1}(1-p)^{-1/2}, & \textit{für } p > \frac{1}{2} \end{cases}$$
mit
$$\psi := \min_{G\subseteq H, |E(G)|>0} \{n^{|V(G)|}\, p^{|E(G)|}\}$$
und $k(H)$ eine Konstante, die nur von H abhängt; $C := ||f''||$.

Wir werden zunächst die Aussage dieses Satzes genauer diskutieren. Im Anschluß stellen wir eine spezielle Zerlegung von Zufallsgrößen vor, die einem lokalen Ansatz entspricht. Wir leiten ein allgemeines Resultat für sogenannte *zerlegbare* Zufallsgrößen her und folgern Satz 11.1.

Zunächst betrachten wir Folgerungen aus Satz 11.1. Wir untersuchen den Fall, wenn H balanciert ist. Dann ist $\varrho(H) = \max\{\varrho(G) : G \subseteq H\}$. Also ist

$$n^{|V(G)|/|E(G)|}\, p \geq n^{k/l}\, p$$

für alle $G \subset H$ und $|E(G)| > 0$ und

$$\min_{G \subset H, |E(G)|>0} \{n^{|V(G)|/|E(G)|}\, p\} \; = n^{k/l}\, p,$$

und somit $\psi = n^k\, p^l = (n^{k/l} p)^l$. Ist $p(n)$ nun so gewählt, dass $\lim_{n\to\infty} p(n)\, n^{k/l} = \infty$, so folgt $\psi \to +\infty$. Für $p \leq \frac{1}{2}$ liefert der Satz Konvergenz: Zu geeigneten Testfunktionen $h : \mathbb{R} \to \mathbb{R}$ (z. B. die Klasse der stetig differenzierbaren h, die außerhalb eines endlichen Intervalls verschwinden, mit $||h'|| < \infty$) ist die Steinlösung $U_N(h)$ beschränkt und die erste und zweite Ableitung ebenfalls (siehe Lemma 10.8). Also folgt für diese Konvergenz-determinierende Klasse (vgl. Satz 10.16):

Korollar 11.2. *Ist H balanciert und $p(n) \gg n^{-k/l}$ und $1 - p(n) \gg n^{-2}$, so konvergiert W schwach gegen die Standardnormalverteilung.*

In Kapitel 10 hatten wir bereits diskutiert, wie man nun auf eine Konvergenzrate für $|P(W \leq t) - \Phi(t)|$ kommt. Wir diskutieren dies nicht weiter, sondern bemerken ohne Beweis die folgende Konsequenz aus Satz 11.1: Es sei für $h : \mathbb{R} \to \mathbb{R}$

$$||h|| := \sup_{x\in\mathbb{R}} |h(x)| + \sup_{x\in\mathbb{R}} |h'(x)|$$

und

$$d_1\big(\mathcal{L}(W_n), N(0,1)\big) := \sup_h ||h||^{-1} \left| \int h(x)\, dF_n(x) - \int h(x)\, d\Phi(x) \right|,$$

wobei F_n die Verteilungsfunktion von W_n bezeichne. Das Supremum wird über alle beschränkten Funktionen mit beschränkter erster Ableitung gebildet. Es folgt aus Satz 11.1:

Satz 11.3. *Sei $W = \frac{X - \mathbb{E}X}{\sqrt{V(X)}}$. So ist*

$$d_1\big(\mathcal{L}(W), N(0,1)\big) \leq k(H) \begin{cases} \psi^{-1/2}, & \textit{für } p \leq 1/2, \\ n^{-1}(1-p)^{-1/2}, & \textit{für } p > 1/2 \end{cases}$$

mit den Bezeichnungen aus Satz 11.1

Für einen Beweis siehe [BKR89]. $d_1(\cdot,\cdot)$ ist eine Metrik auf $\mathcal{M}_1(\mathbb{R})$ (ohne Beweis) und es gilt: Gilt $d_1\big(\mathcal{L}(W), N(0,1)\big) \to 0$, so konvergiert W schwach gegen die Standardnormalverteilung (ebenfalls ohne Beweis).

Wir wollen nun Satz 11.1 beweisen. Dazu betrachten wir allgemeiner die folgende Situation sogenannter *zerlegbarer* Zufallsgrößen. Wir schießen dabei etwas über das Ziel hinaus; wir wollen „nur“für unsere Zählstatistik X einen zentralen Grenzwertsatz beweisen. Die Struktur, die hier betrachtet wird, hat vielseitige Anwendungsmöglichkeiten. Dies kann man bei Interesse im Detail in der Arbeit [BKR89] nachlesen.

Sei I eine endliche Menge und

$$W = \sum_{i \in I} X_i, \quad X_i \text{ jeweils quadrat-integrierbar.}$$

Es sei weiter $\mathbb{E}X_i = 0$, $i \in I$, und $\mathbb{E}W^2 = 1$. Weiter sei $W = W_i + Z_i$, wobei W_i von X_i unabhängig ist, und

$$Z_i := \sum_{k \in K_i} Z_{ik}, i \in I, K_i \subset I,$$

$W_i := W_{ik} + V_{ik}, i \in I, k \in K_i$, und W_{ik} sei jeweils unabhängig vom Paar (X_i, Z_{ik}). Dann gilt:

Lemma 11.4. *Ist W zerlegbar wie oben und alle Zufallsgrößen der Zerlegung seien quadrat-integrierbar. Dann gilt für jede beschränkte Funktion $f : \mathbb{R} \to \mathbb{R}$ mit beschränkter erster und zweiter Ableitung:*

$$\big|\mathbb{E}(f'(W) - W\,f(W))\big| \le C\,\varepsilon,$$

wobei $C := \sup_x |f''(x)|$ und

$$\varepsilon := \frac{1}{2}\sum_{i \in I} \mathbb{E}(|X_i|Z_i^2) + \sum_{i \in I}\sum_{k \in K_i}\big(\mathbb{E}|X_i Z_{ik} V_{ik}| + \mathbb{E}|X_i Z_{ik}|\,\mathbb{E}|Z_i + V_{ik}|\big).$$

Wieder kann, wie oben angedeutet, gefolgert werden:

Satz 11.5. *Ist W zerlegbar wie oben, so gilt*

$$d_1\big(\mathcal{L}(W), N(0,1)\big) \le K\,\varepsilon$$

mit einer universellen Konstante K und ε definiert wie in Lemma 11.4

Wir werden Satz 11.1 aus Lemma 11.4 folgern und dann folgt Satz 11.3 aus 11.5.

Beweis von Lemma 11.4: Es gilt

$$\begin{aligned}\sum_{i\in I}\sum_{k\in K_i}\mathbb{E}(X_i Z_{ik}) &= \sum_{i\in I}\mathbb{E}(X_i\, Z_i) = \sum_{i\in I}\big(\, E(X_i W) - \ E(X_i W_i)\big)\\ &= \sum_{i\in I}\mathbb{E}(X_i W) = \mathbb{E}W^2 = 1.\end{aligned}$$

So kann man schreiben:

$$\begin{aligned}\mathbb{E}(W\,f(W) - f'(W)) &= \Big\{\mathbb{E}(W\,f(W)) - \sum_{i\in I}\mathbb{E}(X_i Z_i\, f'(W_i))\Big\}\\ &\quad + \Big\{\sum_{i\in I}\mathbb{E}(X_i Z_i\, f'(W_i)) - \sum_{i\in I}\sum_{k\in K_i}\mathbb{E}(X_i Z_{ik})\,\mathbb{E}(f'(W_{ik}))\Big\}\\ &\quad + \sum_{i\in I}\sum_{k\in K_i}\Big\{\mathbb{E}(X_i Z_{ik})\big[\mathbb{E}(f'(W_{ik}) - \mathbb{E}(f'(W))\big]\Big\}.\end{aligned}$$

Taylor liefert

$$\begin{aligned}W\,f(W) &= \sum_{i\in I} X_i\, f(W)\\ &= \sum_{i\in I} X_i\Big\{f(W_i) + Z_i\, f'(W_i) + \frac{1}{2}Z_i^2\, f''(W_i + \theta_i\, Z_i)\Big\}\end{aligned}$$

für $\theta_i \in [0,1]$. Nun ist nach Zerlegungs-Konstruktion

$$\begin{aligned}&\Big|\mathbb{E}(W\,f(W)) - \sum_{i\in I}\mathbb{E}\big(X_i Z_i\, f'(W_i)\big)\Big|\\ &\quad\le \Big|\mathbb{E}\Big(\sum_{i\in I} X_i\, f(W_i)\Big)\Big| + \frac{1}{2}\,C\,\sum_{i\in I}\mathbb{E}\big(|X_i| Z_i^2\big).\end{aligned}$$

Der erste Summand auf der rechten Seite verschwindet, da X_i und W_i stochastisch unabhängig sind. Erneut Taylor angewandt liefert

$$\begin{aligned}X_i Z_i\, f'(W_i) &= \sum_{k\in K_i} X_i Z_{ik}\, f'(W_i)\\ &= \sum_{k\in K_i} X_i Z_{ik}\big\{f'(W_{ik}) + V_{ik}\, f''(W_{ik} + \theta_{ik}\, V_{ik})\big\}.\end{aligned}$$

Da W_{ik} unabhängig vom Paar (X_i, Z_{ik}) ist, folgt

$$\begin{aligned}&\Big| \sum_{i\in I} \mathbb{E}(X_i Z_i\, f'(W_i)) - \sum_{i\in I}\sum_{k\in K_i} \mathbb{E}(X_i Z_{ik})\, \mathbb{E}(f'(W_{ik}))\Big| \\ &\le C \sum_{i\in I}\sum_{k\in K_i} E|X_i Z_{ik} V_{ik}|.\end{aligned}$$

Da $W_{ik} = W_i - V_{ik} = W - Z_i - V_{ik}$, folgt nach Taylor

$$f'(W_{ik}) = f'(W) - (Z_i + V_{ik})\, f''(W - \theta_{ik}(Z_i + V_{ik})),$$

also

$$\big|\mathbb{E}f'(W_{ik}) - \mathbb{E}f'(W)\big| \le C\,\mathbb{E}|Z_i + V_{ik}|,$$

womit das Lemma bewiesen ist. □

Sei speziell $I = \{1,\dots,n\}^r$ und

$$K_i = \{k \in I : \{k_1,\dots,k_r\} \cap \{i_1,\dots,i_r\} \neq \emptyset\}.$$

Wähle für die obige Zerlegung

$$Z_{ik} = X_k \;\text{ und }\; V_{ik} = \sum_{l\in K_k\setminus K_i} X_l,\;\; i\in I, k\in K_i.$$

Dann ist

$$\frac{1}{2}\sum_{i\in I}\mathbb{E}\Big(|X_i|\Big(\sum_{k\in K_i} X_k\Big)\Big(\sum_{l\in K_i} X_l\Big)\Big) \le \frac{1}{2}\sum_{i\in I}\sum_{k,l\in K_i}\mathbb{E}\big(|X_i X_k X_l|\big)$$

und

$$\sum_{i\in I}\sum_{k\in K_i}\mathbb{E}\Big(\Big|X_i X_k\Big(\sum_{l\in K_k\setminus K_i} X_l\Big)\Big|\Big) \le \sum_{i\in I}\sum_{k\in K_i, l\in K_i}\mathbb{E}\big(|X_i X_k X_l|\big)$$

sowie

$$\sum_{i\in I}\sum_{k\in K_i}\mathbb{E}|X_i X_k|\;\; E\Big|\sum_{l\in K_i} X_l + \sum_{l\in K_k\setminus K_i} X_l\Big| \le \sum_{i\in I}\sum_{k\in K_i, l\in K_i}\Big(\mathbb{E}|X_i X_k|\mathbb{E}|X_l|\Big),$$

also gilt in diesem Fall:

$$\varepsilon \le 2\Big\{\sum_{i\in I}\sum_{k,l\in K_i}\mathbb{E}\big(|X_i X_k X_l|\big) + \sum_{i\in I}\sum_{k,l\in K_i}\mathbb{E}|X_i X_k|\,\mathbb{E}|X_l|\Big\}.$$

Diese Schranke wollen wir nun für die Zählstatistik im Modell des Zufallsgraphen anwenden. Hier liegt die folgende Zerlegung vor:

Sei Γ_n die Menge der Kopien von H in K_n, kantenweise beschrieben:

$$\Gamma_n = \Big\{ i = (i_1,\dots,i_l) : 1 \le i_1 < \cdots < i_l \le \binom{n}{2} : \{e_{i_1},\dots,e_{i_l}\} \text{ Kopie von } H\Big\},$$

wobei e_j die Kante Nr. j bezeichnet. Es ist $X = \sum_{i \in \Gamma_n} I_i$ mit $I_i = I$(die Kanten zu i bilden eine Kopie von H). Bezeichnen $J_1^{(n)}, \ldots, J_{\binom{n}{2}}^{(n)}$ die *Kantenindikatoren*: $P(J_k^{(n)} = 1) = p(n)$ für alle $k = 1, \ldots, \binom{n}{2}$, so ist $I_i = \prod_{k \in i} J_k^{(n)}$. Mit der Funktion $\phi(x_1, \ldots, x_l) := (\prod_{k=1}^l x_k - p^l) \frac{1}{\sigma}$ und $\sigma^2 := V(X)$ ist dann

$$W = \frac{X - \mathbb{E}X}{\sqrt{V(X)}} = \sum_{i \in \Gamma_n} \phi(J_{i_1}^{(n)}, \ldots, J_{i_l}^{(n)}).$$

Diese Form nennt man *unvollständige* oder *gewichtete* U-Statistik. Dazu am Ende des Kapitels etwas nähere Informationen.

Um σ^2 hatten wir uns in Kapitel 6 schon gekümmert:

$$\sigma^2 = \sum_{i \in \Gamma_n} \sum_{j \in K_i} \operatorname{cov}(Y_i, Y_j)$$

mit $Y_i := \prod_{k=1}^l J_{i_k}^{(n)}$. Es gilt mit $G_i := \{e_{i_1}, \ldots, e_{i_l}\}$ und für $i \in \Gamma_n$

$$\begin{aligned} \sigma^2 &= \sum_{G \subseteq H, |E(G)| \geq 1} \quad \sum_{i,j \in \Gamma_n, G_i \cap G_j = G} p^{2l - |E(G)|} (1 - p^{|E(G)|}) \\ &\sim \sum_{G \subseteq H, |E(G)| \geq 1} C_G \, n^{2k - |V(G)|} \, p^{2l - |E(G)|} \, (1 - p^{|E(G)|}), \end{aligned}$$

wobei C_G eine kombinatorische Konstante ist, die von H und G abhängt (siehe Beweis von Lemma 6.9). Wir erhalten nun für den letzten Ausdruck eine Abschätzung nach unten:

$$\begin{aligned} \sigma &\geq (1 - p) \, n^{2k} \, p^{2l} \sum_{G \subseteq H, |E(G)| \geq 1} C_G \, n^{-|V(G)|} \, p^{-|E(G)|} \\ &\geq (1 - p) \, n^{2k} \, p^{2l} C_{G^\star} \, \psi^{-1}, \end{aligned}$$

wenn $G^\star$ der Graph ist, an dem das Minimum in der Definition von ψ angenommen wird. Wir wollen wie in Lemma 11.4 vorgegeben abschätzen:

$$\varepsilon \leq C \left(\frac{16}{\sigma^3} \sum_{i \in \Gamma_n} \sum_{k,l \in K_i} \mathbb{E}(Y_i Y_k Y_l) + 2 \sum_{i \in \Gamma_n} \sum_{k,l \in K_i} \mathbb{E}\big(|X_i X_k|\big) \, \mathbb{E}|X_l| \right).$$

Nun gilt: $|X_i| = \left| \frac{Y_i - \mathbb{E}Y_i}{\sigma} \right|$, also

$$\sigma |X_i| = |Y_i - \mathbb{E}Y_i| = |\prod_{k=1}^l J_{i_k}^{(n)} - p^l| \leq 1$$

und

$$\sigma \mathbb{E}|X_i| = \mathbb{E}|Y_i - \mathbb{E}Y_i| = \mathbb{E}|1 - Y_i - \mathbb{E}(1 - Y_i)| \leq 2\mathbb{E}(1 - Y_i).$$

Also ist $\mathbb{E}|X_i X_k|\,\mathbb{E}|X_l| \le \frac{2\mathbb{E}(1-Y_i)}{\sigma^3}$ und somit

$$\varepsilon \le \left\{ \frac{32}{\sigma^3} \sum_{i\in\Gamma_n} \sum_{k,l\in K_i} \mathbb{E}(Y_i Y_k Y_l) \right\} \wedge \left\{ \frac{8}{\sigma^3} \sum_{i\in\Gamma_n} \sum_{k,l\in K_i} \mathbb{E}(1-Y_i) \right\}.$$

Der zweite Ausdruck ist für $p \ge 1/2$

$$\begin{aligned} &\le \text{ const.}\, \sigma^{-3}\, n^{3k-4}\, (1-p^l) \\ &\le \text{ const.}\, n^{3k-4}\, (1-p^l) \left(n^{2k-2}(1-p)\right)^{-3/2} \asymp n^{-1}(1-p)^{-1/2} \end{aligned}$$

(die Konstanten sind uniform in $p \ge 1/2$). Der erste Ausdruck liefert

$$\begin{aligned} \varepsilon &\le \frac{32}{\sigma^3} \sum_{\substack{G\subseteq H \\ |E(G)|\ge 1}} \sum_{\substack{i,k\in\Gamma_n \\ G_i\cap G_k = G}} \sum_{\substack{K\subset(G_i\cup G_k) \\ |E(K)|\ge 1}} \sum_{\substack{l\in\Gamma_n: \\ G_l\cap(G_i\cup G_K)=K}} p^{3l-|E(G)|-|E(K)|} \\ &\le \frac{32}{\sigma^3} \sum_{\text{“}} \sum_{\text{“}} \sum_{\substack{K\subset(G_i\cup G_k): K\subseteq G_1 \text{ für ein } l \\ |E(K)|\ge 1}} c_K\, p^{2l-|E(G)|}\, n^{k-|V(K)|}\, p^{l-|E(K)|} \\ &\le \text{ const.}\, \sigma^{-1}\, \psi^{-1}\, n^k\, p^l \end{aligned}$$

mit const. uniform in $p \le 1/2$. Es folgt somit insgesamt $\varepsilon \le$ const. $\psi^{-1/2}$ für $p \le 1/2$ und Satz 11.1 ist somit bewiesen.

Ausblick:

Die Darstellung

$$W = \sum_{i\in\Gamma_n} \phi(J_{i_1}^{(n)}, \ldots, J_{i_l}^{(n)})$$

führt uns zu U-*Statistiken*. Es sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von unabhängigen und identisch verteilten Zufallsgrößen mit Werten in $\mathbb{R}$ und $h\colon\mathbb{R}^l \to \mathbb{R}$ eine Funktion. Dann heißt

$$U_n^l(h) := \frac{1}{\binom{n}{l}} \sum_{1\le i_1<\cdots<i_l\le n} h(X_{i_1}, \ldots, X_{i_l})$$

U-Statistik zur Kernfunktion h vom Grad l. Das Wort U-Statistik kommt dabei von der Bedeutung der Zufallsgröße $U_n^l(h)$ in der Statistik. Für $l = 1$ und $h(x) = x$ ist $U_n^1(h) = \frac{1}{n}\sum_{i=1}^n X_i$ der Mittelwert, ein erwartungstreuer Schätzer des Erwartungswertes der Verteilung der $(X_i)_{i\ge 1}$. Dies ist der Inhalt des starken Gesetzes der großen Zahlen. Nun gibt es, denkbar, noch andere erwartungstreue Schätzer für $\vartheta = \mathbb{E}X_i$. $U_n^1(h)$ zeichnet sich dadurch aus, dass unter *allen* erwartungstreuen Schätzen von ϑ $U_n^1(h)$ optimal in seinem Streuungsverhalten ist, also $V(U_n^1(h))$ minimal ist. Allgemein gilt dies für $U_n^l(h)$ mit jedem h, welches $\mathbb{E}h(X_1, \ldots, X_l) < +\infty$ erfüllt. Unbiased heißt erwartungstreu, daher U-Statistik. Für $h(x,y) = \frac{1}{2}(x-y)^2$ erhalten wir in $U_n^2(h)$ einen erwartungstreuen Schätzer für die Varianz der Stichprobe, optimal in seiner Varianz $V(U_n^2(h))$ unter allen erwartungstreuen Schätzern von σ^2. Es ist nun naheliegend, dass man allgemeine Gesetzmäßigkeiten für U-Statistiken hergeleitet hat wie etwa starke Gesetze, zentrale Grenzwertsätze, Poisson-Grenzwertsätze usw..

Die Zählstatistik für $X(G)$ bei fest vorgegebenen Graphen H im Zufallsgraphenmodell $\mathcal{G}(n,p)$ ist nun eine gewichtete U-Statistik mit $h = \phi$. Gewichtet bedeutet formal einfach, das die Summanden $\phi(J_{i_1}^{(n)}, \dots, J_{i_l}^{(n)})$ das Gewicht 1 bekommen, wenn $i = (i_1, \dots, i_l) \in \Gamma_n \subseteq \big\{ i = (i_1, \dots, i_l) : 1 \le i_1 < \cdots < i_l \le \binom{n}{2} \big\}$ und das Gewicht 0 sonst.

Allgemeine gewichtete U-Statistiken haben die Gestalt

$$\sum_{1 \le i_1 < \cdots < i_l \le n} w(i_1, \dots, i_l)\, h(X_{i_1}, \dots, X_{i_l})$$

mit $w(i_1, \dots, i_l) \in \{0, 1\}$ (oder der normierte Ausdruck).

Für eine spezielle Klasse von gewichteten U-Statistiken, den Zählstatistiken im Zufallsgraphen-Modell, haben wir in dieser Vorlesung gesehen:

Unter Anwendung der Steinschen Methode für Poisson-Approximation (lokaler Ansatz, Kopplungsansatz, Kapitel 3 und 4): Ist H strikt balanciert, so liegt im Schwellenbereich $p(n)\, n^{k/l} \to c$ die Poisson-Approximation der gewichteten U-Statistik vor (siehe Kapitel 6).

Naheliegend sind zum Beispiel die Fragen:

Was passiert im Schwellenbereich für H balanciert, aber *nicht strikt balanciert*? Was bedeutet die Eigenschaft, strikt balanciert zu sein, in Bezug auf die Gewichte $w(i_1, \dots, i_l)$? Gibt es für andere Gewichte eine Poisson-Approximation?

Via der Janson-Ungleichung (siehe Kapitel 7) konnten wir für strikt balancierte H auch oberhalb der Schwelle $t(n) = n^{-k/l}$ Poisson-Approximation beobachten. „Oberhalb“wurde dort genauer beschrieben. Da dort der Erwartungswert wächst, waren die Wahrscheinlichkeiten sehr klein. Die Janson-Ungleichung lieferte eine gute Abschätzung sehr kleiner Wahrscheinlichkeiten. Man bekam eine Abschätzung nach oben

$$P(W = 0) \le \exp\left(-\frac{\lambda}{1 + \delta} \right).$$

und eine nach unten:

$$\exp\left(-\frac{\lambda}{1 - \eta} \right) \le P(W = 0).$$

Bei exponentiellen Abschätzungen spricht man in der Literatur auch von Wahrscheinlichkeiten *großer Abweichungen*. Eine Frage hier ist, was für H balanciert aber *nicht strikt balanciert* passiert? Wir erhalten für gewichtete U-Statistiken Information über große Abweichungen via der Janson-Ungleichung.

Nicht betrachtet haben wir den Aspekt, dass für eine spezielle Klasse H von balancierten, *nicht strikt balancierten* Graphen X im Schwellenbereich Compound Poisson-verteilt ist für $n \to \infty$ (Kapitel 9). Es handelt sich dabei um sogenannte lokale Graphen H, siehe die Arbeit von Stark [Sta01].

Es ergeben sich die Fragen: Was passiert oberhalb der Schwelle für lokale H? Hilft wieder die Janson-Ungleichung? Was geschieht mit nicht-strikt balancierten und *nicht lokalen* H für $n \to \infty$? Hilft die Steinsche Methode?

Immer sind die Zählstatistiken X gewichtete U-Statistiken. Welche Gewichte korrespondieren zu welcher Klasse von Graphen? Kann man aus der Interpretation der gewichteten U-Statistiken als Zählstatistiken im Zufallsgraphenmodell $\mathcal{G}(n,p)$ auf Limesverteilungen für gewichtete U-Statistiken schließen?

Wir belassen es bei diesen offenen Fragen.

Neben der in Lemma 11.4 betrachteten Abhängigkeitsstruktur betrachtet man häufig den sogenanten *Abhängigkeitsgraph*: Angenommen (I,E) sei ein Graph mit endlich oder abzählbar vielen Ecken. Für $i,j \in I$ schreiben wir $i \sim j$, wenn $\{i,j\} \in E$. Für $i \in I$ sei $\mathcal{N}_i$ die Nachbarschaft von i, womit die Menge $\{i\} \cup \{j \in I : j \sim i\}$ gemeint ist. Wir sagen, dass der Graph $(I,\sim)$ ein Abhängigkeitsgraph für eine Familie von Zufallsgrößen $(\xi_i)_{i\in I}$ ist, wenn für je zwei disjunkte Teilmengen I_1, I_2 von Knoten aus I, die durch keine Kante miteinander verbunden sind, die Familie $(\xi_i)_{i\in I_1}$ unabhängig von $(\xi_i)_{i\in I_2}$ ist.

Nun kann man zum Beispiel das folgende Resultat zeigen: (es geht auf Baldi und Rinott [BR89] zurück und hat seinen Ursprung in einer allgemeineren Form in Stein [Ste86], siehe auch [Pen03, Section 2.3]). Es sei eine endliche Kollektion von Zufallsgrößen $(\xi_i)_{i\in I}$ mit Abhängigkeitsgraph $(I,\sim)$ gegeben. Der maximale Grad des Graphen sei $D-1$. Ferner sei $\mathbb{E}\xi_i = 0$ für jedes $i \in I$. Für $W = \sum_{i\in I} \xi_i$ nehmen wir $\mathbb{E}W^2 = 1$ an. Dann gilt

$$\left|P(W \le t) - \Phi(t)\right| \le 2(2\pi)^{-1/4} \sqrt{D^2 \sum_{i\in I} \mathbb{E}|\xi_i|^3} + 6\sqrt{D^3 \sum_{i\in I} E|\xi_i|^4}.$$

Viele Anwendungen dieses Resultats findet man in der Monographie [Pen03]. Neuste Arbeiten und Diskussionen zeigen Anwendungen dieses Resultats für Probleme der geometrischen Wahrscheinlichkeitstheorie.

12. Weitere Entwicklungen

In diesem Kapitel werden kurz *weitere Entwicklungen* im Rahmen der Steinschen Methode angedeutet, die in keinem der Ausblicke der bisherigen Kapitel Erwähnung fanden.

Eine natürliche Frage ist, ob sich die Steinsche Methode nicht auch für *andere bekannte diskrete und kontinuierliche Verteilungen* herleiten lässt.

Im Bereich der diskreten Verteilungen ist ein einheitlicher Blickwinkel zur Aufstellung der Steinschen Gleichung und der Bestimmung der Lösungen der jeweiligen Gleichungen mittels der Generator-Methode entwickelt worden. Man muss nur einen geeigneten *Geburts-Todes-Prozess* wählen, der die in Betracht gezogene diskrete Verteilung als invariante Verteilung hat. Probabilistische Methoden, aber auch der rekursive Ansatz aus Kapitel 2 führen dann zu Abschätzungen der Lösungen und der Zuwächse der Lösungen. Diese Abschätzungen sind nicht immer gut oder gar optimal. Probleme hiermit zeigte das Kapitel zur Compound Poisson-Approximation eindrücklich. Eine Übersicht über Arbeiten zu verschiedenen Verteilungen findet man in [Rei98].

Der Gesichtspunkt der Geburts-Todes-Prozesse ist in Arbeiten von Brown und Xia [BX01], von Holmes [Hol00] und von Eichelsbacher und Reinert [ER03] zu finden. Die Arbeit von Holmes stellt die Steinsche Methode auch für Verteilungen vor, die nicht einer bekannten Klasse entstammen, sondern deren diskrete Gewichte zum Beispiel von einer Simulation stammen. Dahinter steckt der Simulationsaspekt der Markov chain Monte Charlo Methode, die Diaconis, Holmes und Reinert mit der Steinschen Methode, insbesondere mit der Generator-Methode in Verbindung bringen. In der Arbeit von Eichelsbacher und Reinert werden Gibbs-Maße betrachtet und es wird die Steinsche Methode für diese Limesverteilungen hergeleitet. Dies ermöglicht die Untersuchung erster Beispiele aus der statistischen Mechanik.

Zu kontinuierlichen Verteilungen mit einer Dichte gibt es den sogenannten Dichte-Ansatz von Stein, der bisher unveröffentlicht ist. Dabei wird angenommen, dass die Dichte p strikt positiv ist und die Ableitung von p überall existert. Weiter sei

$$\int_{\mathbb{R}} |p'(x)|\, dx < \infty.$$

Es sei

$$\psi(x) = \frac{p'(x)}{p(x)}.$$

Dann hat eine Zufallsgröße Z genau dann eine Verteilung mit Dichte p, wenn für alle Funktionen f, deren Ableitung f' die Bedingung

$$\int_{\mathbb{R}} |f'(z)|\, p(z)\, dz < \infty$$

erfüllt, gilt:

$$\mathbb{E}\big(f'(Z) + \psi(Z)\, f(Z)\big) = 0.$$

Abschließend sei nur noch exemplarische aufgezählt, dass neue Arbeiten zur Steinschen Methode zu den folgenden Themen entstanden sind oder gerade geschrieben werden: Steinsche Methode für die Normal-Approximation von hierarchischen Folgen, Steinsche Methode für Diffusionen auf Manigfaltigkeiten, Steinsche Methode und die Theorie Palmscher Maße, Steinsche Methode für das Wignersche Halbkreisgesetz, Steinsche Methode für Gibbsche Punktprozesse, Steinsche Methode und Edgeworth-Entwicklungen von Summen abhängiger Zufallsgrößen.

Literatur

[AGG89] R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.*, 17(1):9–25, 1989.

[AGG90] R. Arratia, L. Goldstein, and L. Gordon. Poisson approximation and the Chen-Stein method. *Stat. Science*, 5:403–434, 1990.

[Als00] G. Alsmeyer. *Wahrscheinlichkeitstheorie.* Skripten zur Mathematischen Statsitik Nr.30, Münster, 2000.

[AS00] N. Alon and J. H. Spencer. *The probabilistic method.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience [John Wiley & Sons], New York, second edition, 2000. With an appendix on the life and work of Paul Erdős.

[Bar82] A. D. Barbour. Poisson convergence and random graphs. *Math. Proc. Cambridge Philos. Soc.*, 92(2):349–359, 1982.

[Bar88] A. D. Barbour. Stein's method and Poisson process convergence. *J. Appl. Probab.*, (Special Vol. 25A):175–184, 1988. A celebration of applied probability.

[Bar90] A. D. Barbour. Stein's method for diffusion approximations. *Probab. Theory Related Fields*, 84(3):297–322, 1990.

[Bau02] H. Bauer. *Wahrscheinlichkeitstheorie.* de Gruyter Lehrbuch. [de Gruyter Textbook]. Walter de Gruyter & Co., Berlin, fifth edition, 2002.

[BB92] A. D. Barbour and T. C. Brown. Stein's method and point process approximation. *Stochastic Process. Appl.*, 43(1):9–31, 1992.

[BBX98] A. D. Barbour, T. C. Brown, and A. Xia. Point processes in time and Stein's method. *Stochastics Stochastics Rep.*, 65(1-2):127–151, 1998.

[BC01] A. D. Barbour and O. Chryssaphinou. Compound Poisson approximation: a user's guide. *Ann. Appl. Probab.*, 11(3):964–1002, 2001.

[BCL92] A. D. Barbour, L. H. Chen, and W.-L. Loh. Compound Poisson approximation for nonnegative random variables via Stein's method. *Ann. Probab.*, 20:1843–1866, 1992.

[BE83] A. D. Barbour and G. K. Eagleson. Poisson approximation for some statistics based on exchangeable trials. *Adv. in Appl. Probab.*, 15(3):585–600, 1983.

[Ber41] A. C. Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.*, 49:122–136, 1941.

[BH84] A. D. Barbour and P. Hall. On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.*, 95(3):473–480, 1984.

[BH89] A. D. Barbour and L. Holst. Some applications of the Stein-Chen method for proving Poisson convergence. *Adv. in Appl. Probab.*, 21(1):74–90, 1989.

[BHJ92] A. D. Barbour, L. Holst, and S. Janson. *Poisson Approximation.* Oxford University Press, Oxford, 1992.

[BKR89] A. D. Barbour, M. Karoński, and A. Ruciński. A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B*, 47(2):125–145, 1989.

[BM02] A. D. Barbour and M. Månsson. Compound Poisson process approximation. *Ann. Probab.*, 30(3):1492–1537, 2002.

[Bol84] E. Bolthausen. An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete*, 66(3):379–386, 1984.

[Bol01] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[Bor76] A. A. Borovkov. *Wahrscheinlichkeitstheorie*. Birkhäuser Verlag, Basel, 1976. Eine Einführung, Aus dem Russischen übersetzt von Egmar Rödel und Hans Kühne, Herausgegeben von Peter Franken, Lehrbücher und Monographien aus dem Gebiete der Exakten Wissenschaften, Mathematische Reihe, Band 53.

[BR89] P. Baldi and Y. Rinott. On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.*, 17(4):1646–1650, 1989.

[BU98] A. D. Barbour and S. Utev. Solving the Stein equation in compound Poisson approximation. *Adv. in Appl. Probab.*, 30(2):449–475, 1998.

[BU99] A. D. Barbour and S. Utev. Compound Poisson approximation in total variation. *Stochastic Process. Appl.*, 82(1):89–125, 1999.

[BWX00] T. C. Brown, G. V. Weinberg, and A. Xia. Removing logarithms from Poisson process error bounds. *Stochastic Process. Appl.*, 87(1):149–165, 2000.

[BX99] A. D. Barbour and A. Xia. Poisson perturbations. *ESAIM Probab. Statist.*, 3:131–150 (electronic), 1999.

[BX00] A. D. Barbour and A. Xia. Estimating Stein's constants for compound Poisson approximation. *Bernoulli*, 6(4):581–590, 2000.

[BX01] T. C. Brown and A. Xia. Stein's method and birth-death processes. *Ann. Probab.*, 29(3):1373–1403, 2001.

[Che76] L. H. Y. Chen. An approximation theorem for sums of certain randomly selected indicators. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 33(1):69–74, 1975/76.

[CS01] L. H. Y. Chen and Q.-M. Shao. A non-uniform Berry-Esseen bound via Stein's method. *Probab. Theory Related Fields*, 120(2):236–254, 2001.

[CS03] L. H. Y. Chen and Q.-M. Shao. Normal approximation under local dependence. preprint, 2003.

[DVJ88] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes*. Springer Series in Statistics. Springer-Verlag, New York, 1988.

[ER60] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[ER99] P. Eichelsbacher and M. Roos. Compound Poisson approximation for dissociated random variables via Stein's method. *Combin. Probab. Comput.*, 8(4):335–346, 1999. Random graphs and combinatorial structures (Oberwolfach, 1997).

[ER03] P. Eichelsbacher and G. Reinert. Stein methode for Gibbs measures. preprint, submitted, 2003.

[Erd59] P. Erdős. Graph theory and probability. *Canad. J. Math.*, 11:34–38, 1959.

[Erd61] P. Erdős. Graph theory and probability. II. *Canad. J. Math.*, 13:346–352, 1961.

[Ess45] C.-G. Esséen. Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Math.*, 77:1–125, 1945.

[FKG71] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971.

[Göt91] F. Götze. On the rate of convergence in the multivariate CLT. *Ann. Probab.*, 19(2):724–739, 1991.

[Har60] T. E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.*, 56:13–20, 1960.

[Hol00] S. Holmes. Stein's method for birth and death chains. preprint, 2000.

[Jan90] S. Janson. Poisson approximation for large deviations. *Random Structures Algorithms*, 1(2):221–229, 1990.

[Jan98] S. Janson. New versions of Suen's correlation inequality. In *Proceedings of the Eighth International Conference "Random Structures and Algorithms" (Poznan, 1997)*, volume 13, pages 467–483, 1998.

[Jia02] T. Jiang. A comparison of scores of two protein structures with foldings. *Ann. Probab.*, 30(4):1893–1912, 2002.

[JŁR00] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[Ker64] J. Kerstan. Verallgemeinerung eines Satzes von Prochorow und Le Cam. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 2:173–179 (1964), 1964.

[Kre88] U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, volume 59 of *Vieweg Studium: Aufbaukurs Mathematik [Vieweg Studies: Mathematics Course]*. Friedr. Vieweg & Sohn, Braunschweig, 1988.

[LC60] L. M. Le Cam. An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.*, 10:1181–1197, 1960.

[Pen03] M. Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.

[Rei98] G. Reinert. Couplings for normal approximations with Stein's method. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 193–207. Amer. Math. Soc., Providence, RI, 1998.

[Roo93] M. Roos. *Stein-Chen method for Compound Poisson approximation*. PhD thesis, University of Zürich, Switzerland, 1993.

[Roo94] M. Roos. Stein's method for compound Poisson approximation: the local approach. *Ann. Appl. Probab.*, 4(4):1177–1187, 1994.

[Spe98] J. Spencer. A useful elementary correlation inequality. II. *J. Combin. Theory Ser. A*, 84(1):95–98, 1998.

[Sta01] D. Stark. Compound Poisson approximations of subgraph counts in random graphs. *Random Structures Algorithms*, 18(1):39–60, 2001.

[Ste72] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602, Berkeley, Calif., 1972. Univ. California Press.

[Ste86] C. Stein. *Approximate computation of expectations*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA, 1986.

[Str65] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist*, 36:423–439, 1965.

[Str93] D.W. Stroock. *Probability Theorey, an analytic view*. Cambridge University Press, Cambridge, 1993.

[Sue90] W.-C. S. Suen. A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph. *Random Structures Algorithms*, 1(2):231–242, 1990.

[Wat95] M. S. Waterman. *Introduction to Computational Biology.* Chapman and Hall, CRC Press, 1995.

Index