Mathematik für Maschinenbauer, Bauingenieure und Umwelttechniker III

Vorlesungsskriptum WS 2001/02 - WS 2002/03
überarbeitet September 2008

R. Verfürth

Fakultät für Mathematik, Ruhr-Universität Bochum

Inhaltsverzeichnis

KAPITEL XI

Stochastik I
Diskrete Modelle

XI.1. Modelle für Zufallsexperimente

XI.1.1. Endliche Wahrscheinlichkeitsräume. Wir betrachten Zufallsexperimente mit endlich vielen möglichen Ausgängen. Diese werden beschrieben durch eine endliche, nicht leere Menge Ω, deren Elemente ω die Versuchsausgänge bezeichnen. Sie heißen ERGEBNISSE oder ELEMENTAREREIGNISSE. Ω heißt ERGEBNISMENGE.

Die Teilmengen von Ω sind die EREIGNISSE, die in dem Modell in Betracht gezogen werden. Genauer:

> Wir identifizieren $A \subset \Omega$ mit dem Ereignis, dass ein $\omega \in A$ der beobachtete Versuchsausgang ist.

Dementsprechend bezeichnen $A \cap B$ bzw. $A \cup B$ die Ereignisse, dass A und B bzw. A oder B eintreten. Die leere Menge $\emptyset$ heißt auch das UNMÖGLICHE EREIGNIS; Ω ist das SICHERE EREIGNIS.

Jedem Ereignis ordnen wir eine Wahrscheinlichkeit zu. Die Menge aller möglichen Ereignisse ist die POTENZMENGE $\mathcal{P}(\Omega)$, d.h. die Menge aller Teilmengen von Ω. Eine Abbildung $P : \mathcal{P}(\Omega) \to [0, 1]$ heißt WAHRSCHEINLICHKEITSVERTEILUNG oder WAHRSCHEINLICHKEITSMASS, wenn sie folgende Eigenschaften hat:

$P(\Omega) = 1$	(NORMIERUNG)
$P(A) \geq 0$	für alle A (POSITIVITÄT)
$P(A \cup B) = P(A) + P(B)$	für alle disjunkten A, B (ADDITIVITÄT)

$P(A)$ heißt die WAHRSCHEINLICHKEIT des Ereignisses A. Das Paar (Ω, P) heißt der dem Zufallsexperiment zugeordnete WAHRSCHEINLICHKEITSRAUM.

BEISPIEL XI.1.1. Wir wollen die Wahrscheinlichkeit bestimmen, dass die Summe der bei zwei Würfen eines Würfels erhaltenen Augenzahlen mindestens 10 ist. Wir können die Ergebnisse des Zufallsexperimentes „Zweimaliges Werfen eines Würfels“ durch die Paare (i, k)

der beobachteten Augenzahlen beschreiben. Daher ist

$$\Omega = \{(i, k) : 1 \leq i, k \leq 6\}.$$

Ω hat 36 Elemente. Aus Symmetriegründen ist es naheliegend, sie alle als gleich wahrscheinlich zu betrachten. Jedes (i, k) hat also die Wahrscheinlichkeit $\frac{1}{36}$, Die Menge der Ergebnisse, für die die Summe $i + k$ mindestens 10 ist, ist

$$A = \{(6, 6), (6, 5), (5, 6), (6, 4), (4, 6), (5, 5)\}.$$

Da A genau 6 Elemente hat, ist die gesuchte Wahrscheinlichkeit $\frac{6}{36} = \frac{1}{6}$.

Aus den obigen Eigenschaften eines Wahrscheinlichkeitsmaßes folgen leicht weitere Eigenschaften. Für $A, B, A_i \in \mathcal{P}(\Omega)$ gilt

$$\boxed{\begin{aligned} P(\Omega \backslash A) &= 1 - P(A) \\ P(\emptyset) &= 0 \\ A \subset B &\Rightarrow P(A) \leq P(B) \\ P(A \backslash B) &= P(A) - P(B) \end{aligned}}$$

und

$$\boxed{\begin{aligned} P(\bigcup_{i=1}^{n} A_i) &= \sum_{i=1}^{n} P(A_i) \quad \text{falls } A_1, \ldots, A_n \text{ paarweise disjunkt} \\ P(\bigcup_{i=1}^{n} A_i) &\leq \sum_{i=1}^{n} P(A_i) \quad \text{für beliebige } A_1, \ldots, A_n \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \end{aligned}}$$

Aus diesen Eigenschaften folgt insbesondere

$$P(A) = \sum_{\omega \in A} P(\{\omega\}).$$

Die Wahrscheinlichkeit von A ist also die Summe der Wahrscheinlichkeiten der Ergebnisse, bei denen A eintritt. P ist also eindeutig bestimmt durch die Werte aller $P(\{\omega\})$ mit $\omega \in \Omega$. Man schreibt auch $P(\omega)$ statt $P(\{\omega\})$. Die Abbildung $\omega \mapsto P(\omega)$ heißt Wahrscheinlichkeitsfunktion. Die Wahrscheinlichkeitsverteilung kann also auch durch Angabe der zugehörigen Wahrscheinlichkeitsfunktion beschrieben werden. Offensichtlich gilt für jede Wahrscheinlichkeitsfunktion:

$$\begin{aligned} P(\omega) &\geq 0 \quad \text{für alle } \omega \in \Omega \\ \sum_{\omega \in \Omega} P(\omega) &= 1. \end{aligned}$$

Ein besonders wichtiger Spezialfall einer Wahrscheinlichkeitsfunktion ist die Gleichverteilung auf Ω, bei der alle Ergebnisse $\omega \in \Omega$ gleich wahrscheinlich sind. In diesem Fall wird (Ω, P) als Laplacescher Wahrscheinlichkeitsraum bezeichnet; das zugehörige Zufallsexperiment heißt Laplace-Experiment. Offensichtlich gilt dann:

$$\boxed{\begin{aligned} P(\omega) &= \frac{1}{\operatorname{card}(\Omega)} \qquad \text{für alle } \omega \in \Omega, \\ P(A) &= \frac{\operatorname{card}(A)}{\operatorname{card}(\Omega)} \qquad \text{für alle } A \subset \Omega. \end{aligned}}$$

XI.1.2. Urnenmodelle. Viele Laplace-Experimente lassen sich durch Urnenmodelle beschreiben. Dabei stellen wir uns vor, dass wir aus einer Urne mit N Kugeln, die von 1 bis N nummeriert sind, sukzessive n Kugeln zufällig ziehen.

Offensichtlich sind zwei Arten des Ziehens zu unterscheiden:

(1) mit Rücklegen: Nach jedem Zug wird die gerade gezogene Kugel wieder in die Urne zurückgelegt; jede Kugel kann mehrmals gezogen werden.
(2) ohne Rücklegen: eine gezogene Kugel wird nicht zurückgelegt; jede Kugel kann nur einmal gezogen werden.

Ein Beispiel für (1) ist das mehrmalige Werfen eines Würfels; ein Beispiel für (2) ist die Ziehung der Lottozahlen.

Man kann das Ergebnis der Folge der Ziehungen dadurch beschreiben, dass man das n-Tupel $(\omega_1, \ldots, \omega_n)$ angibt, in dem ω_i die Nummer der im i-ten Zug gezogenen Kugel ist. Offensichtlich kann man hier wiederum zwei Arten von Ziehen unterscheiden:

(i) mit Reihenfolge: Es kommt auf die Reihenfolge des Erscheinens an; $(1, 2)$ und $(2, 1)$ bezeichnen verschiedene Ergebnisse.
(ii) ohne Reihenfolge: Es kommt nicht auf die Reihenfolge des Erscheinens an; $(1, 2)$ und $(2, 1)$ bezeichnen gleiche Ergebnisse.

Ein Beispiel für (i) ist das Bestimmen einer Losnummer durch sukzessives Ziehen ihrer Ziffern; ein Beispiel für (ii) ist wieder die Ziehung der Lottozahlen.

Durch diese Unterscheidungen ergeben sich insgesamt vier verschiedene Ergebnismengen. Zu ihrer Beschreibung setzen wir

$$\mathbb{A} = \{1, 2, \ldots, N\}.$$

(I) Stichproben in Reihenfolge mit Rücklegen (1i):

$$\Omega_I = \mathbb{A}^n,$$
$$\operatorname{card}(\Omega_I) = N^n.$$

(II) Stichproben in Reihenfolge ohne Rücklegen (2i):

$$\Omega_{II} = \{(\omega_1, \dots, \omega_n) \in \mathbb{A}^n : \omega_i \neq \omega_j \text{ für } i \neq j\}$$

$$\begin{aligned}\operatorname{card}(\Omega_{II}) &= \frac{N!}{(N-n)!}\\ &= N \cdot (N-1) \cdot \ldots \cdot (N-n+1).\end{aligned}$$

(III) Stichproben ohne Reihenfolge ohne Rücklegen (2ii):

$$\Omega_{III} = \{(\omega_1, \dots, \omega_n) \in \mathbb{A}^n : \omega_1 < \omega_2 < \ldots < \omega_n\}$$

$$\begin{aligned}\operatorname{card}(\Omega_{III}) &= \binom{N}{n}\\ &= \frac{N!}{(N-n)!n!}.\end{aligned}$$

(IV) Stichproben ohne Reihenfolge mit Rücklegen (1ii):

$$\Omega_{IV} = \{(\omega_1, \dots, \omega_n) \in \mathbb{A}^n : \omega_1 \leq \omega_2 \leq \ldots \leq \omega_n\}$$

$$\operatorname{card}(\Omega_{IV}) = \binom{N+n-1}{n}.$$

Man kann diese Urnenmodelle auch alternativ wie folgt interpretieren:

> Gefragt ist nach der Anzahl der Möglichkeiten, n Murmeln auf N Plätze zu verteilen. Dabei entspricht die Nummer der Ziehung der Nummer der Murmel und die Nummer der Kugel der Nummer des Platzes.

XI.1.3. Anwendungsbeispiele.

Beispiel XI.1.2. Es werden vier völlig gleich aussehende Würfel gleichzeitig geworfen. Welches ist die Wahrscheinlichkeit p, dass die vier erscheinenden Augenzahlen verschieden sind?
Da ein gleichzeitiger Wurf von vier Würfeln dem viermaligen Werfen eines Würfels entspricht, handelt es sich um ein Experiment mit Rücklegen. Da die Würfel nicht unterscheidbar sind, könnte man versucht sein, ein Modell ohne Reihenfolge zu betrachten. Dies ist aber *falsch*: Dem Resultat, dass die Augenzahlen 1, 2, 3 und 4 auftreten, entspricht in diesem Modell nur ein Ergeignis, da es nicht auf die Reihenfolge ankommt. Tatsächlich entsprechen diesem Resultat jedoch die $4! = 24$ Permutationen der Menge $\{1, 2, 3, 4\}$. Das *richtige* Modell ist also dasjenige mit Berücksichtigung der Reihenfolge. Daher sind alle möglichen Fälle Ω_I mit $N = 6$ und $n = 4$. Die günstigen Fälle (alle Augenzahlen verschieden) sind Ω_{II} mit $N = 6$ und $n = 4$. Daher ist die gesuchte Wahrscheinlichkeit

$$\begin{aligned}p &= \frac{\frac{6!}{2!}}{6^4}\\ &= \frac{3 \cdot 4 \cdot 5 \cdot 6}{6^4}\end{aligned}$$

$$= \frac{5}{18}.$$

Beispiel XI.1.3. Wie groß ist die Wahrscheinlichkeit p, dass mindestens 2 von 25 Schülern einer Klasse am gleichen Tag Geburtstag haben?

Als Ereignisraum können wir Ω_I mit $n = 25$ und $N = 365$ wählen. Das Ergebnis $(\omega_1, \ldots, \omega_{25})$ bedeutet, dass Schüler Nummer 1 am ω_1-ten Tag Geburtstag hat, Schüler Nummer 2 am ω_2-ten Tag usw. Das interessierende Ereignis ist das Komplement des Ereignisses Ω_{II}, das alle Schüler an verschiedenen Tagen Geburtstag haben. Also ist

$$p = 1 - P(\Omega_{II}).$$

Für $P(\Omega_{II})$ erhalten wir

$$\begin{aligned} P(\Omega_{II}) &= \frac{\frac{N!}{(N-n)!}}{N^n} \\ &= \frac{N}{N} \cdot \frac{N-1}{N} \cdot \frac{N-2}{N} \cdot \ldots \cdot \frac{N-n+1}{N} \\ &= 1 \cdot (1 - \frac{1}{N}) \cdot (1 - \frac{2}{N}) \cdot \ldots \cdot (1 - \frac{n-1}{N}). \end{aligned}$$

Wenn wir jeden Ausdruck $(1 - \frac{k}{N})$ durch $\exp(-\frac{k}{N})$ approximieren, erhalten wir

$$\begin{aligned} P(\Omega_{II}) &\approx \exp\left(-\sum_{k=1}^{n-1} \frac{k}{N}\right) \\ &= \exp\left(-\frac{n(n-1)}{2N}\right) \\ &= \exp\left(-\frac{600}{730}\right) \\ &\approx 0.44. \end{aligned}$$

Damit ergibt sich $p \approx 0.56$. Der exakte Wert ohne diese Approximation ist $p = 0.568$.

Beispiel XI.1.4. Beim Zahlenlotto „6 aus 49“ werden $n = 6$ Kugeln aus $N = 49$ Kugeln ohne Rücklegen gezogen. Dabei kommt es nicht auf die Reihenfolge an. Also ist die Ereignismenge Ω_{III} mit

$$\begin{aligned} \mathrm{card}(\Omega_{III}) &= \binom{49}{6} \\ &= 13983816. \end{aligned}$$

Daher ist die Wahrscheinlichkeit p_6 für „6 Richtige“

$$\begin{aligned} p_6 &= \frac{1}{\mathrm{card}(\Omega_{III})} \\ &\approx 7.1511 \cdot 10^{-8}. \end{aligned}$$

Wie groß ist die Wahrscheinlichkeit p_4 für „genau 4 Richtige“?
Seien $\tilde{\omega}_1, \dots \tilde{\omega}_6$ die gezogenen Zahlen. Dann können wir die Menge aller gesuchten Ergebnisse wie folgt erzeugen: Wir ziehen zunächst 4 Kugeln aus $\{\tilde{\omega}_1, \dots, \tilde{\omega}_6\}$ und dann zwei Kugeln aus den 43 Kugeln $\{1, \dots, 49\} \backslash \{\tilde{\omega}_1, \dots, \tilde{\omega}_6\}$. Für das erste Ziehen haben wir

$$\binom{6}{4} = 15$$

Möglichkeiten, für das zweite Ziehen

$$\binom{43}{2} = 21 \cdot 43 = 903$$

Möglichkeiten. Daher ist

$$\begin{aligned} p_4 &= \frac{15 \cdot 903}{13983816} \\ &= \frac{3 \cdot 5 \cdot 43}{7 \cdot 44 \cdot 46 \cdot 47} \\ &\approx 9.68619 \cdot 10^{-4}. \end{aligned}$$

Analog kann man die Wahrscheinlichkeit p_5 für „genau 5 Richtige“ berechnen. Die Wahrscheinlichkeit für „mindestens 4 Richtige“ ist dann

$$p_4 + p_5 + p_6 \approx 9.8714 \cdot 10^{-4}.$$

Beispiel XI.1.5. Auf wie viele Arten können sich zwei nicht unterscheidbare Spatzen auf vier Telegraphenleitungen verteilen?
Dies ist Ω_{IV} mit $N = 4$ und $n = 2$. Dementsprechend gibt es

$$\binom{5}{2} = 10$$

Möglichkeiten.

XI.1.4. Die hypergeometrische Verteilung. In Beispiel XI.1.4 haben wir einen Spezialfall dieser wichtigen Verteilung kennen gelernt. Für den allgemeinen Fall betrachten wir eine Urne, die S schwarze und W weiße Kugeln enthält, insgesamt $N = S + W$ Kugeln. Es werden $n \leq S + W$ Kugeln ohne Zurücklegen gezogen. Die Wahrscheinlichkeit, dass dabei s schwarze und $w = n - s$ weiße Kugeln gezogen werden, beträgt

$$\boxed{h(s; n, N, S) = \frac{\binom{S}{s}\binom{W}{w}}{\binom{S+W}{n}} \quad (0 \leq s \leq n).}$$

Diese Verteilung heißt hypergeometrische Verteilung.

Beispiel XI.1.6. Wie groß ist die Wahrscheinlichkeit, dass beim Skatspiel ein Spieler genau 3 Asse erhält?
Dies entspricht obiger Situation mit $S = 4$ (Zahl aller Asse), $W =$

28 (Zahler aller anderen Karten), $s = 3$ (Zahl der Asse des Spielers) und $n = 10$ (Zahl der Karten des Spielers). Daher ist die gesuchte Wahrscheinlichkeit

$$\begin{aligned} p &= \frac{\binom{4}{3}\binom{28}{7}}{\binom{32}{10}} \\ &= \frac{4 \cdot 22 \cdot 8 \cdot 9 \cdot 10}{32 \cdot 31 \cdot 30 \cdot 29} \\ &= \frac{66}{899} \\ &\approx 7.34\%. \end{aligned}$$

Die Wahrscheinlichkeit, dass mindestens ein Spieler 3 Asse erhält, ist dreimal so groß, da keine zwei Spieler gleichzeitig drei Asse erhalten können.

XI.1.5. Multinomialkoeffizienten. Die Binomialkoeffizienten $\binom{n}{k}$ beschreiben, auf wie viele Arten man eine Menge von n nummerierten Kugeln so in zwei Gruppen einteilen kann, dass die erste Gruppe k Kugeln enthält. Wie viele Möglichkeiten gibt es, die Zahlen $1, \ldots, n$ so in r Gruppen einzuteilen, dass die erste Gruppe k_1 Elemente hat, die zweite k_2 Elemente usw.? Dabei muss man natürlich $k_1 + \ldots + k_r = n$ voraussetzen. Man kann zuerst auf $\binom{n}{k_1}$ Arten die erste Gruppe auswählen, dann auf $\binom{n-k_1}{k_2}$ Arten die zweite Gruppe usw. Insgesamt erhält man

$$\begin{aligned} &\binom{n}{k_1} \cdot \binom{n-k_1}{k_2} \cdot \binom{n-k_1-k_2}{k_3} \cdot \ldots \cdot \binom{n-k_1-\ldots-k_{r-1}}{k_r} \\ &= \frac{n!}{k_1!k_2!\ldots k_r!} \end{aligned}$$

Möglichkeiten. Diesen Ausdruck bezeichnet man als Multinomialkoeffizienten und kürzt ihn mit $\binom{n}{k_1,k_2,\ldots,k_r}$ ab:

$$\boxed{\binom{n}{k_1, k_2, \ldots, k_r} = \frac{n!}{k_1!k_2!\ldots k_r!} \qquad \text{mit } k_1 + \ldots + k_r = n.}$$

Beispiel XI.1.7. 26 Schulkinder haben einen Fußball, vier Tennisschläger, ein Fußballfeld und einen Tennisplatz. Die Zahl der Einteilungen in zwei Fußballmannschaften mit je 11 Spielern und in zwei Tennisteams mit je 2 Spielern beträgt

$$\begin{aligned} \binom{26}{11, 11, 2, 2} &= \frac{26!}{11!11!2!2!} \\ &= 2 \cdot 5 \cdot 12 \cdot 13 \cdot 14 \cdot 15 \cdot 17 \cdot 19 \cdot 23 \cdot 26 \\ &\approx 6.327 \cdot 10^{10}. \end{aligned}$$

XI.1.6. Identitäten für Binomialkoeffizienten. Aus der Binomischen Formel (vgl. Abschnitt I.2.8 (S. 19, Teil I))

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

ergibt sich durch Einsetzen von $x = 1$, $y = 1$ bzw. $x = -1$, $y = 1$ bzw. durch Ableiten nach x und anschließendes Einsetzen von $x = 1$, $y = 1$:

$$\boxed{\begin{aligned} \sum_{k=0}^{n} \binom{n}{k} &= 2^n, \\ \sum_{k=0}^{n} (-1)^k \binom{n}{k} &= 0, \\ \sum_{k=1}^{n} k \binom{n}{k} &= n 2^{n-1}. \end{aligned}}$$

In Abschnitt I.2.8 (S. 19, Teil I) haben wir bereits die folgenden Identitäten kennen gelernt:

$$\boxed{\begin{aligned} \binom{n}{k} &= \binom{n}{n-k}, \\ \binom{n}{k} &= \binom{n-1}{k} + \binom{n-1}{k-1}. \end{aligned}}$$

Da die Summe der Wahrscheinlichkeiten $h(s; n, N, S)$ aus Abschnitt XI.1.4 (S. 10) gleich 1 ist, ergibt sich schließlich

$$\boxed{\sum_{s=0}^{n} \binom{S}{s} \binom{W}{n-s} = \binom{S+W}{n} \quad \text{mit } n \le S + W.}$$

XI.2. Bedingte Wahrscheinlichkeit und Unabhängigkeit

XI.2.1. Definition bedingter Wahrscheinlichkeiten. Häufig steht, bevor das Ergebnis eines Zufallsexperimentes bekannt ist, schon die Information zur Verfügung, dass das Ergebnis zu einer gewissen Teilmenge des Ereignisraumes gehört. Z.B. sieht ein Spieler beim Skat seine eigenen Karten. Interessiert sich Spieler 1 für das Ereignis A, dass Spieler 2 zwei Asse hat, so wird er zunächst seine eigenen Asse zählen. Hat er selbst drei oder vier Asse, so ist für ihn die Wahrscheinlichkeit von A natürlich 0; hat er maximal zwei Asse, so ist die Wahrscheinlichkeit von A für ihn positiv.

Für Laplace-Experimente ist der Ansatz für die Definition bedingter Wahrscheinlichkeiten sehr nahe liegend. Waren ursprünglich alle Ergebnisse $\omega \in \Omega$ gleich wahrscheinlich und erhält man nun die Information, dass $\omega \in B$ liegt, so ordnen wir den Ergebnissen in $\Omega \backslash B$ die bedingte Wahrscheinlichkeit 0 zu und betrachten die Ergebnisse aus B als gleich wahrscheinlich unter der bedingten Wahrscheinlichkeit. Dies bedeutet, dass für $A \subset \Omega$ die bedingte Wahrscheinlichkeit von A bei gegebenem B den Wert

$$P(A|B) = \frac{\operatorname{card}(A \cap B)}{\operatorname{card}(B)}$$

erhält. Aus

$$P(A \cap B) = \frac{\operatorname{card}(A \cap B)}{\operatorname{card}(\Omega)}$$

und

$$P(B) = \frac{\operatorname{card}(B)}{\operatorname{card}(\Omega)}$$

ergibt sich daher in diesem Fall

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}.}$$

Wir definieren nun auch für beliebige Wahrscheinlichkeitsräume (Ω, P) und für beliebige Ereignisse B mit $P(B) > 0$ die BEDINGTE WAHRSCHEINLICHKEIT $P(A|B)$ von A bei gegebenem B durch diese Formel.

BEISPIEL XI.2.1. Aus einer Urne, die zwei schwarze und drei weiße Kugeln enthält, werden nacheinander zwei Kugeln ohne Zurücklegen gezogen. Wir interessieren uns für die bedingte Wahrscheinlichkeit des Ereignisses A, dass die zweite gezogene Kugel schwarz ist, bei gegebenem Ereignis B, dass die erste gezogene Kugel weiß ist. Da nach dem ersten Ziehen drei von vier Kugeln schwarz sind, sollte diese Wahrscheinlichkeit $\frac{3}{4}$ betragen. Um die Richtigkeit unserer Definition nachzuprüfen, nummerieren wir die weißen Kugeln mit 1 und 2 und die schwarzen Kugeln mit 3, 4 und 5. Dann haben die interessierenden Ereignisse die Form

$$\begin{aligned} A \cap B &= \{(1,3),(1,4),(1,5),(2,3),(2,4),(2,5)\}, \\ B &= \{(1,2),(1,3),(1,4),(1,5),(2,1),(2,3),(2,4),(2,5)\}. \end{aligned}$$

Es ergibt sich

$$P(A|B) = \frac{\operatorname{card}(A \cap B)}{\operatorname{card}(B)} = \frac{6}{8} = \frac{3}{4}.$$

XI.2.2. Eigenschaften. In der Praxis wird häufig nicht $P(A|B)$ aus $P(A \cap B)$ und $P(B)$ berechnet, sondern umgekehrt $P(A \cap B)$ aus $P(A|B)$ und $P(B)$. Aus der Definition von $P(A|B)$ ergibt sich nämlich die Produktformel für Wahrscheinlichkeiten

$$\boxed{P(A \cap B) = P(B)P(A|B).}$$

Beispiel XI.2.2. In Beispiel XI.2.1 können wir zur Berechnung von $P(A \cap B)$ wie folgt argumentieren: Da zu Beginn 2 von 5 Kugeln weiß sind, ist $P(B) = \frac{2}{5}$. Da nach Eintreten von B 3 von 4 Kugeln schwarz sind, ist $P(A|B) = \frac{3}{4}$. Also muss

$$P(A \cap B) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10}$$

sein.

Die obige Produktformel lässt sich wie folgt verallgemeinern: Sind $A_1, \dots, A_k$ Ereignisse mit $P(A_1 \cap \dots \cap A_k) > 0$, so ist

$$\boxed{P(A_1 \cap \dots \cap A_k) = P(A_1) \cdot P(A_2|A_1) \cdot \ldots \cdot P(A_k|A_1 \cap \dots \cap A_{k-1}).}$$

Beispiel XI.2.3. Wir wollen die Wahrscheinlichkeit bestimmen, dass beim Skat jeder der drei Spieler genau ein Ass hat. Bezeichne dazu mit A_i das Ereignis, dass Spieler i genau ein Ass erhält. Aus Symmetriegründen können wir annehmen, dass Spieler 1 die ersten 10 ausgeteilten Karten erhält, Spieler 2 die nächsten 10, dann Spieler 3 zehn, und die letzten 2 Karten in den Stock kommen. Dann ist die gesuchte Wahrscheinlichkeit

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2).$$

Es ist

$$P(A_1) = \frac{\binom{4}{1}\binom{28}{9}}{\binom{32}{10}}.$$

Da nach Austeilen der Karten an Spieler 1 noch 22 Karten incl. 3 Asse zu verteilen sind, ist

$$P(A_2|A_1) = \frac{\binom{3}{1}\binom{19}{9}}{\binom{22}{10}}.$$

Analog ist

$$P(A_3|A_1 \cap A_2) = \frac{\binom{2}{1}\binom{10}{9}}{\binom{12}{10}}.$$

Damit ergibt sich

$$P(A_1 \cap A_2 \cap A_3) = \frac{50}{899} \\ \approx 5.56\%.$$

Folgende Eigenschaften bedingter Wahrscheinlichkeiten sind sehr nützlich:

(1) Sei $P(B) > 0$. Durch $P_B(A) = P(A|B)$ wird ein Wahrscheinlichkeitsmaß auf Ω definiert. Ist $A \subset \Omega \backslash B$ oder $P(A) = 0$, so ist $P(A|B) = 0$.

(2) (Formel von der totalen Wahrscheinlichkeit) Sind $B_1, B_2, \ldots$ paarweise disjunkte Teilmengen von Ω, deren Vereinigung Ω ist, so gilt für jedes Ereignis A

$$P(A) = \sum_k P(B_k) P(A|B_k).$$

(3) (Formel von Bayes) Ist $P(A) > 0$ und gelten die Voraussetzungen von (2), so gilt für jedes i

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_k P(B_k)P(A|B_k)}.$$

(4) Ist C die Vereinigung der paarweise disjunkten Ereignisse $C_1, C_2, \ldots$ mit $P(C_i) > 0$ für alle i und sind die $P(A|C_i)$ alle gleich, so ist $P(A|C) = P(A|C_1)$.

Beispiel XI.2.4. Wir greifen Beispiel XI.2.1 (S. 13) auf. Wie groß ist die Wahrscheinlichkeit für das Ereignis C, dass beide ohne Zurücklegen gezogenen Kugeln die gleiche Farbe haben?
Sei B_1 das Ereignis, dass die erste gezogene Kugel weiß ist, und B_2 das Ereignis, dass sie schwarz ist. Wenn die erste gezogene Kugel weiß ist, ist beim zweiten Zug nur eine von vier Kugeln weiß. Damit ist $P(C|B_1) = \frac{1}{4}$. Analog ergibt sich $P(C|B_2) = \frac{1}{2}$. Wegen $P(B_1) = \frac{2}{5}$ und $P(B_2) = \frac{3}{5}$ ergibt sich mit der Formel von der totalen Wahrscheinlichkeit

$$\begin{aligned} P(C) &= P(B_1)P(C|B_1) + P(B_2)P(C|B_2) \\ &= \frac{2}{5} \cdot \frac{1}{4} + \frac{3}{5} \cdot \frac{1}{2} \\ &= \frac{4}{10} \\ &= \frac{2}{5}. \end{aligned}$$

Beispiel XI.2.5. Eine Krankheit komme bei ca. 0.5% der Bevölkerung vor. Ein Test zur Erkennung der Krankheit führt bei 99% der Kranken zu einer Reaktion, aber auch bei 2% der Gesunden. Wie groß ist die Wahrscheinlichkeit, dass eine Person, bei der die Reaktion auftritt, tatsächlich krank ist?
Zur Lösung denken wir uns die Bevölkerung von 1 bis N nummeriert. B_1 sei die Menge der Kranken und B_2 diejenige der Gesunden. Also ist $\text{card}(B_1) \approx 0.005N$ und $\text{card}(B_2) \approx 0.995N$. A bezeichne die Menge der Personen, bei denen der Test zur Reaktion führt. Dann ist $\text{card}(A \cap B_1) \approx 0.99\,\text{card}(B_1)$ und $\text{card}(A \cap B_2) \approx 0.02\,\text{card}(B_2)$. Gesucht ist $P(B_1|A)$.
Wir wenden die Formel von Bayes an und ordnen jeder Person bei zufälliger Auswahl die Wahrscheinlichkeit $\frac{1}{N}$ zu. Dann ist

$$\begin{aligned} P(B_1) &= 0.005, \\ P(B_2) &= 0.995, \\ P(A \cap B_1) &= 0.99 \cdot 0.005, \\ P(A \cap B_2) &= 0.02 \cdot 0.995. \end{aligned}$$

Damit ergibt sich mit der Formel von Bayes

$$\begin{aligned} P(B_1|A) &= \frac{P(A \cap B_1)}{P(A)} \\ &= \frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.02 \cdot 0.995} \\ &= \frac{495}{2485} \\ &\approx 0.2. \end{aligned}$$

Von allen Personen, bei denen der Test eine Reaktion zeigt, sind also nur 20% tatsächlich erkrankt.

XI.2.3. Unabhängigkeit. Zwei Ereignisse A und B heißen unabhängig, wenn $P(A \cap B) = P(A)P(B)$ ist. Ist $P(B) > 0$, ist dies äquivalent zu $P(A) = P(A|B)$.

Unabhängigkeit von A und B drückt aus, dass A und B wahrscheinlichkeitstheoretisch in dem Sinn keinerlei Einfluss aufeinander haben, dass die Information „B tritt ein“ keinen Einfluss auf die Wahrscheinlichkeit von A hat. Dies muss man deutlich von der realen Beeinflussung unterscheiden.

Beispiel XI.2.6. Wir werfen einen Würfel zweimal. Sei A das Ereignis, dass die Summe der beiden Augenzahlen gerade ist, und B das Ereignis, dass die zweite Augenzahl gerade ist. Offensichtlich ist

$$P(A) = P(B) = \frac{1}{2}$$

und

$$P(A \cap B) = \frac{1}{4} = P(A)P(B).$$

Also sind die beiden Ereignisse unabhängig, obwohl B mitbestimmt, ob A eintritt.

Wir müssen den Begriff der Unabhängigkeit auf Familien von Ereignissen übertragen. Dazu sagen wir, dass für die endliche Familie $\{A_i : i \in J\}$ von Ereignissen die PRODUKTFORMEL gilt, wenn

$$P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i)$$

ist. Eine beliebige Familie $\{A_i : i \in J\}$ von Ereignissen heißt UNABHÄNGIG, wenn für jede endliche Teilfamilie die Produktformel gilt.

Damit ergeben sich folgende Eigenschaften:

(1) Jede Teilfamilie einer unabhängigen Familie von Ereignissen ist unabhängig. Eine Familie ist genau dann unabhängig, wenn jede endliche Teilfamilie unabhängig ist.
(2) Ist $\{A_i : i \in J\}$ eine Familie unabhängiger Ereignisse, k ein nicht zu J gehörender Index und $P(A_k) = 0$ oder $P(A_k) = 1$, so ist auch $\{A_i : i \in J \cup \{k\}\}$ unabhängig.
(3) Ist $\{A_i : i \in J\}$ unabhängig und B_i für jedes i eines der Ereignisse A_i, $\Omega \backslash A_i$, $\emptyset$ oder Ω, so ist $\{B_i : i \in J\}$ unabhängig.
(4) Ist $J = \{1, \ldots, n\}$ endlich, so ist $\{A_i : i \in J\}$ genau dann unabhängig, wenn für jede Wahl von $B_i \in \{A_i, \Omega \backslash A_i\}$ die Produktformel für $B_1, \ldots, B_n$ gilt.

XI.2.4. Produktexperimente. Wir nehmen an, dass wir schon Modelle (Ω_1, P_1), ..., (Ω_n, P_n) für gewisse Zufallsexperimente kennen, und wollen nun ein Modell für das Experiment konstruieren, das in der unabhängigen Hintereinanderausführung aller dieser Teilexperimente besteht. Dazu liegt es nahe, als Ereignisraum das kartesische Produkt

$$\Omega_1 \times \ldots \times \Omega_n = \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i \in \Omega_i, 1 \leq i \leq n\}$$

zu wählen und als Wahrscheinlichkeitsfunktion

$$P(\omega) = P_1(\omega_1) \cdot \ldots \cdot P_n(\omega_n).$$

Man bezeichnet (Ω, P) als PRODUKT DER WAHRSCHEINLICHKEITSRÄUME (Ω_i, P_i) und schreibt

$$\Omega = \Omega_1 \times \ldots \times \Omega_n,$$

$$P = P_1 \times \ldots \times P_n.$$

Ist A_i ein Ereignis im i-ten Experiment, so bezeichnet das kartesische Produkt $A_1 \times \ldots \times A_n$ das Ereignis in Ω, dass für alle i im i-ten Teilexperiment A_i eintritt. Die Wahrscheinlichkeit für dieses Ereignis ist

$$\boxed{P(A_1 \times \ldots \times A_n) = P(A_1) \cdot \ldots \cdot P(A_n).}$$

XI.2.5. Binomialverteilung. Wir betrachten ein Experiment, das in der unabhängigen n-fachen Wiederholung eines Einzelexperimentes mit nur zwei verschiedenen möglichen Ausgängen besteht. Wir bezeichnen diese beiden Ausgänge mit 0 und 1. Dann ist $\Omega_i = \{0, 1\}$ und $\Omega = \{0, 1\}^n$. Da die Teilexperimente Wiederholungen des gleichen Experimentes sind, sollen in allen Teilexperimenten die gleichen Wahrscheinlichkeiten auftreten. $p = P_i(1)$ soll also nicht von i abhängen. p heißt oft Erfolgswahrscheinlichkeit. Natürlich ist dann $P_i(0) = 1 - p$. Dann ist im Produktmodell $P(\omega) = p^k(1-p)^{n-k}$, wenn k die Anzahl der Einsen in ω ist. Ein Experiment dieser Form nennt man Bernoulli-Experiment und P heißt Bernoulli-Verteilung.

Das Ereignis, dass insgesamt k Einsen auftreten wird durch

$$E_k = \{\omega \in \Omega : \sum_{i=1}^{n} \omega_i = k\}$$

beschrieben. Die Zahl der Elemente von E_k ist gleich der Zahl der Möglichkeiten, die k Zeitpunkte in $\{1, \ldots, n\}$ festzulegen, an denen die Einsen auftreten, also $\binom{n}{k}$. Es folgt

$$P(E_k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Die Terme

$$\boxed{b_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{mit } 0 \le k \le n}$$

bestimmen die Wahrscheinlichkeitsverteilung. Sie sind nicht negativ und haben die Summe 1. Man nennt sie Binomialverteilung mit Parametern n und p oder kurz $b_{n,p}$-Verteilung.

Beispiel XI.2.7. Wie groß ist die Wahrscheinlichkeit, bei 10 Würfen eines Würfels 3 Sechsen zu erhalten?
Wir können die gewofenen Sechsen als Erfolge auffassen und ihnen den Wert 1 zuordnen; die anderen Ergebnisse sind Misserfolge und erhalten den Wert 0. Dann ist $p = \frac{1}{6}$. Die gesuchte Wahrscheinlichkeit ist

$$\binom{10}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^7 \approx 15.5\%.$$

Beispiel XI.2.8. Die Wahrscheinlichkeit für die Geburt eines Jungen sei $p = 0.51$. Aufeinanderfolgende Geburten seien unabhängig. Dann ist die Wahrscheinlichkeit, dass in einer Familie mit 4 Kindern 2 Jungen und 2 Mädchen vorkommen, gleich

$$\binom{4}{2} 0.51^2 0.49^2 \approx 37.47\%.$$

XI.2.6. Multinomialverteilung. Hier hat man wieder n unabhängige, identische Teilversuche, aber jeder Teilversuch hat nun r mögliche Ausgänge. Die Teilexperimente sind beschreibbar durch $P_i(j) = p_j$, $j = 1, \ldots, r$, wobei die Wahrscheinlichkeiten $p_1, \ldots, p_r$ beliebig vorgegeben sind mit $p_j \geq 0$, $j = 1, \ldots, r$, und $p_1 + \ldots + p_r = 1$. Die Wahrscheinlichkeit, in den n Teilversuchen k_1-mal das Ergebnis 1, k_2-mal das Ergebnis 2 usw. zu erreichen, ist

$$\binom{n}{k_1, \ldots, k_r} p_1^{k_1} \cdot \ldots \cdot p_r^{k_r}.$$

XI.2.7. Geometrische Verteilung. Die Wahrscheinlichkeit, dass in einem Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p der erste Erfolg genau im k-ten Versuch eintritt, ist $p(1-p)^{k-1}$, $k \geq 1$. Man k als Ergebnis eines Experimentes auffassen, das darin besteht, zu beobachten, wann in einer Folge von Bernoulli-Experimenten der erste Erfolg eintritt. Die zugehörige Ereignismenge ist dann $\Omega = \mathbb{N} \setminus \{0\}$. Die zugehörige Wahrscheinlichkeitsfunktion ist $P(k) = p(1-p)^{k-1}$. Die hierdurch definierte Verteilung heißt geometrische Verteilung. Man beachte, dass im Gegensatz zu allen bisherigen Beispielen der Ereignisraum jetzt nicht mehr endlich ist.

XI.2.8. Negative Binomialverteilung. Dies ist eine Verallgemeinerung der geometrischen Verteilung. Sei $f(k; r, p)$ die Wahrscheinlichkeit, dass bei $n \geq r + k$ Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit p genau k Misserfolge dem r-ten Erfolg vorangehen. Man kann nachrechnen, dass

$$f(k; r, p) = \binom{k + r - 1}{k} p^r (1-p)^k$$

ist. Die hierdurch bei festem r auf $\Omega = \mathbb{N}$ definierte Verteilung nennt man negative Binomialverteilung oder Pascal-Verteilung.

XI.3. Zufallsvariable, Erwartungswert, Varianz

XI.3.1. Zufallsvariable. Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum und $\mathcal{X}$ eine beliebige Menge. Eine Abbildung $X : \Omega \to \mathcal{X}$ heißt ($\mathcal{X}$-wertige) Zufallsvariable.

Bemerkung XI.3.1. Da Ω endlich ist, ist die Bildmenge $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$ einer Zufallsvariablen X auf Ω stets endlich.

Beispiel XI.3.2. Vor einer Wahl werden zufällig ausgewählte Bürger gefragt, welche Partei sie wählen wollen. Hier ist Ω die Menge der Bürger, P die Gleichverteilung auf Ω, $\mathcal{X}$ die Menge der wählbaren Parteien vereinigt mit der „Partei“ der Nichtwähler und $X(\omega)$ die Partei, die Bürger ω angeblich wählen will.

Beispiel XI.3.3. Wir interessieren uns für die Summe der Augenzahlen beim Würfeln mit zwei Würfeln. Hier ist $\Omega = \{(i,j) : 1 \leq i,j \leq 6\}$, P die Gleichverteilung auf Ω, $\mathcal{X} = \{2,\ldots,12\}$ und $X((i,j)) = i+j$.

An Zufallsvariablen interessiert uns vor allem ihre Verteilungsfunktion. Diese gibt an, wie wahrscheinlich die einzelnen Ergebnisse von X sind. Genauer versteht man unter der Verteilungsfunktion von $X : \Omega \to \mathcal{X}$ das Wahrscheinlichkeitsmaß P_X auf $\mathcal{X}_X = X(\Omega)$, das definiert ist durch

$$\boxed{P_X(x) = P(\{\omega \in \Omega : X(\omega) = x\}) \quad \text{mit } x \in \mathcal{X}_X.}$$

$P_X(x)$ ist also die Wahrscheinlichkeit, dass X den Wert x annimmt.

Die Verteilungsfunktion einer Zufallsvariablen X kann durch ein Stabdiagramm veranschaulicht werden. Dazu erstellt man zunächst eine Liste der möglichen Werte $x_k \in \mathcal{X}$ von X, berechnet für jedes x_k die Wahrscheinlichkeit $P_X(x_k)$ und zeichnet dann senkrecht über den Punkten x_k der x-Achse Striche der Länge $P_X(x_k)$.

Beispiel XI.3.4. Die Verteilungsfunktion der Zufallsvariablen aus Beispiel XI.3.3 ist durch Tabelle XI.3.1 bestimmt. Das zugehörige Stabdiagramm ist in Abbildung XI.3.1 wiedergegeben.

Tabelle XI.3.1. Verteilungsfunktion der Zufallsvariablen aus Beispiel XI.3.3

x_k	2	3	4	5	6	7	8	9	10	11	12
$P_X(x_k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

XI.3.2. Gemeinsame Verteilung mehrerer Zufallsvariabler. Sind auf einem Wahrscheinlichkeitsraum (Ω, P) mehrere Zufallsvariable $X_1, \ldots, X_n$ mit evtl. verschiedenen Wertebereichen $\mathcal{X}_1, \ldots, \mathcal{X}_n$ definiert, kann man sie zu einer einzigen Zufallsvariablen X mit Wertebereich

$$\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$$

zusammenfassen, indem man

$$X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$$

setzt. Die Verteilungsfunktion von X nennt man die gemeinsame Verteilungsfunktion von $X_1, \ldots, X_n$.

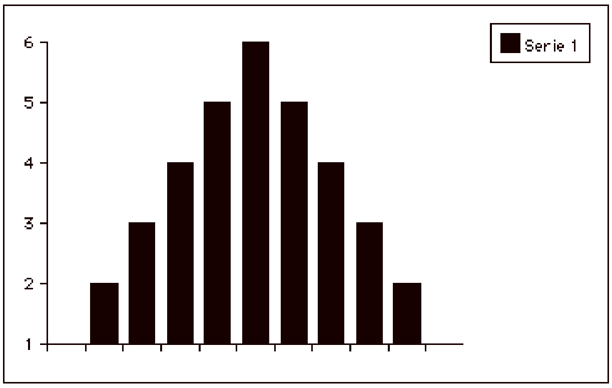

ABBILDUNG XI.3.1. Stabdiagramm der Zufallsvariablen aus Beispiel XI.3.3

BEISPIEL XI.3.5. Wir betrachten die Bernoulli-Verteilung P zu $p \in (0,1)$ auf $\Omega = \{0,1\}^n$ (vgl. Abschnitt XI.2.5 (S. 18)). Für $\omega \in \Omega$ sei

$$S(\omega) = \omega_1 + \ldots + \omega_n$$

die Zahl der Erfolge. Für ω mit $S(\omega) \geq 1$ setzen wir

$$N(\omega) = \min\{j \geq 1 : \omega_j = 1\}.$$

Dies ist die Wartezeit bis zum ersten Erfolg. Ist $S(\omega) = 0$, setzen wir $N(\omega) = n + 1$. S und N sind zwei Zufallsvariable auf (Ω, P). Ihre gemeinsame Verteilung wird beschrieben durch die Angabe der Wahrscheinlichkeiten

$$p(k,h) = P(S = k, N = h)$$

mit $0 \leq k \leq n$ und $1 \leq h \leq n+1$. Wir wollen diese Verteilung bestimmen.
Offenbar ist

$$p(0, n+1) = (1-p)^n$$

und

$$p(0,h) = 0$$

für $1 \leq h \leq n$. Ist $S(\omega) = k \geq 1$ und $N(\omega) = h$, so muss gelten $\omega_i = 0$ für $i < h$ und $\omega_h = 1$, und es müssen genau $k-1$ Einsen unter $\omega_{h+1}, \ldots, \omega_n$ vorkommen. Es gibt $\binom{n-h}{k-1}$ solche Elemente, und jedes hat die Wahrscheinlichkeit $p^k(1-p)^{n-k}$. Also ist

$$p(k,h) = \binom{n-h}{k-1} p^k (1-p)^{n-k}$$

für $k \geq 1$. Ist $k - 1 > n - h$, so gibt es kein solches ω, und es ist $\binom{n-h}{k-1} = 0$. Die obige Formel bleibt also richtig.

XI.3.3. Unabhängigkeit. Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum. Eine Familie X_i, $i \in I$, von Zufallsvariablen $X_i : \Omega \to \mathcal{X}_i$ heißt UNABHÄNGIG, wenn für jede Wahl von $A_i \subset \mathcal{X}_i$, $i \in I$, die Ereignisse $\{X_i \in A_i\}$, $i \in I$, unabhängig sind.

Sei nun $X_1, \ldots, X_n$ eine unabhängige Familie von Zufallsvariablen. Dann gilt gemäß Abschnitt XI.2.3 (S. 16) speziell für alle $(x_1, \ldots, x_n) \in \mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$

$$\boxed{P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i).}$$

Die gemeinsame Verteilung von $X_1, \ldots, X_n$ ist also das Produkt der Verteilungen der X_i.

Die Unabhängigkeit von Zufallsvariablen bleibt unter der Komposition von Abbildungen erhalten: Sind X_i, $i \in I$, unabhängige Zufallsvariable und $F_i : \mathcal{X}_i \to \mathcal{Y}_i$ beliebige Abbildungen mit beliebigen Wertebereichen, so sind die Zufallsvariablen $F_i \circ X_i$, $i \in I$, auch unabhängig.

XI.3.4. Erwartungswert. Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum und X eine reellwertige Zufallsvariable auf Ω. Dann ist der ERWARTUNGSWERT von X definiert durch

$$\boxed{EX = E(X) = \sum_{\omega \in \Omega} X(\omega) P(\omega).}$$

BEISPIEL XI.3.6. Für die Zufallsvariable „Summe der Augenzahlen beim Werfen von zwei Würfeln“ aus Beispiel XI.3.4 erhalten wir den Erwartungswert

$$\begin{aligned} EX &= \frac{1}{36}\{1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + 4 \cdot 5 + 5 \cdot 6 + 6 \cdot 7 \\ &\qquad + 5 \cdot 8 + 4 \cdot 9 + 3 \cdot 10 + 2 \cdot 11 + 1 \cdot 12\} \\ &= \frac{252}{36} \\ &= 7. \end{aligned}$$

Sind X, Y reellwertige Zufallsvariable auf Ω und ist $\lambda \in \mathbb{R}$, so gilt

$$\boxed{\begin{aligned} E(\lambda X) &= \lambda E(X) \\ E(X+Y) &= E(X) + E(Y) \\ X, Y \text{ unabhängig} \implies E(XY) &= E(X)E(Y). \end{aligned}}$$

Die ersten beiden Eigenschaften folgen direkt aus der Definition des Erwartungswertes. Zur Nachweis der dritten Eigenschaft bezeichnen wir mit x_i, $i \in I$, und y_j, $j \in J$, die (endlich vielen!) möglichen Werte von X bzw. Y. Dann folgt mit dem vorigen Abschnitt

$$\begin{aligned}
E(XY) &= \sum_{\omega \in \Omega} X(\omega) Y(\omega) P(\omega) \\
&= \sum_{i \in I} \sum_{j \in J} x_i y_j \underbrace{P(X = x_i, Y = y_j)}_{P(X=x_i)P(Y=y_j)} \\
&= \sum_{i \in I} \sum_{j \in J} x_i y_j P(X = x_i) P(Y = y_j) \\
&= \left\{ \sum_{i \in I} x_i P(X = x_i) \right\} \left\{ \sum_{j \in J} y_j P(Y = y_j) \right\} \\
&= E(X) E(Y).
\end{aligned}$$

Beispiel XI.3.7 (Erwartungswert der Binomialverteilung). Für die Zufallsvariable S aus Beispiel XI.3.5 erhalten wir

$$\begin{aligned}
E(S) &= \sum_{k=0}^{n} k \underbrace{P(S = k)}_{=\binom{n}{k} p^k (1-p)^{n-k}} \\
&= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \underbrace{k \binom{n}{k}}_{=n\binom{n-1}{k-1}} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n \binom{n-1}{k-1} p p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell} (1-p)^{n-1-\ell} \\
&= np.
\end{aligned}$$

Das gleiche Ergebnis erhalten wir einfacher mit folgendem „Trick", der auch in anderen Situationen hilfreich ist: Definiere die Zufallsvariablen $S_1, \ldots, S_n$ durch die Vorschrift $S_i(\omega) = 1$, falls das i-te Teilexperiment ein Erfolg ist, $S_i(\omega) = 0$ sonst. Dann ist $S = S_1 + \ldots + S_n$. Für jedes i ist aber offensichtlich $E(S_i) = p$. Daher ist

$$E(S) = \sum_{i=1}^{n} E(S_i) = np.$$

Beispiel XI.3.8 (Erwartungswert der hypergeometrischen Verteilung). In einer Urne befinden sich S schwarze und W weiße Kugeln, insgesamt $N = S+W$ Kugeln. Wir ziehen daraus $n \leq N$ Kugeln ohne Zurücklegen (vgl. Abschnitt XI.1.4 (S. 10)). Die Zufallsvariable X sei die Zahl der dabei gezogenen schwarzen Kugeln. Um $E(X)$ zu berechnen, bezeichnen wir mit X_i die Zufallsvariable, die den Wert 1 liefert, wenn im i-ten Zug eine schwarze Kugel gezogen wird, und die sonst den Wert 0 gibt. Dann ist $X = X_1 + \ldots + X_n$, und für jedes i gilt $E(X_i) = \frac{S}{N}$. Damit folgt

$$E(X) = n\frac{S}{N}.$$

XI.3.5. Varianz und Kovarianz. Seien X und Y zwei reellwertige Zufallsvariable auf einem endlichen Wahrscheinlichkeitsraum (Ω, P). Die Varianz von X ist definiert durch

$$\boxed{\operatorname{Var}(X) = E((X - E(X))^2).}$$

Die Größe

$$\boxed{\sigma_X = \sqrt{\operatorname{Var}(X)}}$$

heißt die Streuung oder Standardabweichung von X.

Die Kovarianz von X und Y ist definiert durch

$$\boxed{\operatorname{Cov}(X,Y) = E((X - E(X))(Y - E(Y))).}$$

Die Größe

$$\boxed{\rho_{XY} = \frac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y}}$$

heißt der Korrelationskoeffizient von X und Y. Die Zufallsvariablen X und Y heißen unkorreliert, wenn $\operatorname{Cov}(X,Y) = 0$ ist.

Die Varianz ist ein Maß für die Streuung einer Zufallsvariablen um den Erwartungswert EX, der ein Maß für den Mittelwert ist. Ist $EX = 0$ und haben die Werte $x_1, \ldots, x_n$ von X alle die gleiche Wahrscheinlichkeit $\frac{1}{n}$, so ist

$$\sigma_X = \sqrt{\frac{1}{n}\sum x_i^2}.$$

Bis auf den Skalierungsfaktor $\frac{1}{\sqrt{n}}$ ist also σ_X der Euklidische Abstand des Punktes $(x_1, \ldots, x_n)$ vom Ursprung.

Eine positive Kovarianz von X und Y bedeutet, dass X die Tendenz hat, dort groß zu sein, wo auch Y groß ist.

Sind X, Y, X_i reellwertige Zufallsvariable und a, b, c, d reelle Zahlen, so gelten folgende Eigenschaften:

$$\boxed{\begin{aligned}
\operatorname{Var}(X) &= E(X^2) - E(X)^2 \\
\operatorname{Var}(aX + b) &= a^2 \operatorname{Var}(X) \\
\operatorname{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\
\operatorname{Cov}(aX + b, cY + d) &= ac \operatorname{Cov}(X, Y) \\
\operatorname{Cov}(X, Y) &= \operatorname{Cov}(Y, X) \\
|\operatorname{Cov}(X, Y)| &\le \sigma_X \sigma_Y \\
\operatorname{Var}(\sum_{i=1}^{n} X_i) &= \sum_{i=1}^{n} \operatorname{Var}(X_i) + \sum_{i \neq j} \operatorname{Cov}(X_i, X_j) \\
X, Y \text{ unabhängig } &\Longrightarrow \operatorname{Cov}(X, Y) = 0 \\
X_1, \ldots, X_n \text{ unabhängig } &\Longrightarrow \operatorname{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \operatorname{Var}(X_i)
\end{aligned}}$$

Zum Nachweis dieser Eigenschaften beginnen wir mit der dritten. Da der Erwartungswert einer konstanten Zufallsvariablen diese Konstante ist, erhalten wir

$$\begin{aligned}
\operatorname{Cov}(X, Y) &= E((X - E(X))(Y - E(Y))) \\
&= E(XY - XE(Y) - YE(X) + E(X)E(Y)) \\
&= E(XY) - E(XE(Y)) - E(YE(X)) + E(E(X)E(Y)) \\
&= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\
&= E(XY) - E(X)E(Y).
\end{aligned}$$

Die erste Eigenschaft ist der Spezialfall $X = Y$ der dritten Eigenschaft. Aus der dritten Eigenschaft folgt $\operatorname{Cov}(X, Y) = 0$, wenn eine der Zufallsvariablen X oder Y konstant ist. Aus der Definition der Kovarianz folgt zudem, dass die Abbildung $X, Y \mapsto \operatorname{Cov}(X, Y)$ bilinear ist. Damit folgt

$$\begin{aligned}
\operatorname{Cov}(aX + b, cY + d) &= \operatorname{Cov}(aX, cY) + \operatorname{Cov}(aX, d) \\
&\quad + \operatorname{Cov}(b, cY) + \operatorname{Cov}(b, d) \\
&= ac \operatorname{Cov}(X, Y).
\end{aligned}$$

Die zweite Eigenschaft ist der Spezialfall $X = Y$, $a = c$ der soeben bewiesenen vierten Eigenschaft.
Die fünfte Eigenschaft folgt direkt aus der Definition.
Die sechste Eigenschaft ist eine Konsequenz der Cauchy-Schwarzschen

Ungleichung (vgl. Abschnitt II.6.6 (S. 103, Teil I)).
Zum Nachweis der siebten Eigenschaft setze $\overline{X}_i = X_i - E(X_i)$. Dann folgt

$$\begin{aligned}\operatorname{Var}(\sum_{i=1}^n X_i) &= E((\sum_{i=1}^n \overline{X}_i)^2)\\ &= E(\sum_{i=1}^n \overline{X}_i^2 + \sum_{i\neq j} \overline{X}_i\overline{X}_j)\\ &= \sum_{i=1}^n E(\overline{X}_i^2) + \sum_{i\neq j} E(\overline{X}_i\overline{X}_j)\\ &= \sum_{i=1}^n \operatorname{Var}(X_i) + \sum_{i\neq j} \operatorname{Cov}(X_i, X_j).\end{aligned}$$

Sind X und Y unabhängig, so auch $X - E(X)$ und $Y - E(Y)$. Damit folgt die achte Eigenschaft:

$$\begin{aligned}\operatorname{Cov}(X,Y) &= E((X-E(X))(Y-E(Y)))\\ &= E(X-E(X))E(Y-E(Y)) \quad \text{(wg. Unabhängigkeit)}\\ &= 0.\end{aligned}$$

Die letzte Eigenschaft schließlich folgt aus der siebten und achten Identität.

Beispiel XI.3.9 (Varianz der Binomialverteilung). Wir betrachten die Zufallsvariable S aus Beispiel XI.3.5 (S. 21) und definieren die Zufallsvariablen S_i wie in Beispiel XI.3.7 (S. 23). Da sich S_i nur auf das i-te Teilexperiment bezieht und die Teilexperimente unabhängig sind, sind die S_i unabhängig. Daher ist

$$\begin{aligned}\operatorname{Var}(S) &= \operatorname{Var}(S_1 + \ldots + S_n)\\ &= \operatorname{Var}(S_1) + \ldots + \operatorname{Var}(S_n)\\ &= n \operatorname{Var}(S_1),\end{aligned}$$

da jedes S_i die gleiche Varianz hat. Weiter ist

$$\begin{aligned}\operatorname{Var}(S_1) &= E(S_1^2) - E(S_1)^2\\ &= p - p^2\\ &= p(1-p).\end{aligned}$$

Insgesamt folgt

$$\operatorname{Var}(S) = np(1-p).$$

Beispiel XI.3.10 (Varianz der hypergeometrischen Verteilung). Wir betrachten die Zufallsvariablen X und $X_1, \dots, X_n$ aus Beispiel XI.3.8 (S. 24). Setze $p = \frac{S}{N}$. Alle X_i haben die gleiche Varianz:

$$\begin{aligned}\operatorname{Var}(X_i) &= \operatorname{Var}(X_1) \\ &= E(X_1^2) - E(X_1)^2 \\ &= p - p^2 \\ &= p(1-p).\end{aligned}$$

Ebenso gilt aus Symmetriegründen für alle $i \neq j$

$$\begin{aligned}\operatorname{Cov}(X_i, X_j) &= \operatorname{Cov}(X_1, X_2) \\ &= E(X_1 X_2) - E(X_1)E(X_2).\end{aligned}$$

Weiter ist

$$\begin{aligned}E(X_1 X_2) &= P(\{X_1 = 1\} \cap \{X_2 = 1\}) \\ &= P(\{X_1 = 1\})P(\{X_2 = 1\} | \{X_1 = 1\}) \\ &= \frac{S}{N}\frac{S-1}{N-1} \\ &= p\frac{S-1}{N-1}.\end{aligned}$$

Damit folgt

$$\begin{aligned}\operatorname{Cov}(X_i, X_j) &= p\frac{S-1}{N-1} - p^2 \\ &= p(\frac{S-1}{N-1} - p)\end{aligned}$$

für $i \neq j$ und

$$\begin{aligned}\operatorname{Var}(X) &= \sum_{i=1}^{n} \operatorname{Var}(X_i) + \sum_{i \neq j} \operatorname{Cov}(X_i, X_j) \\ &= np(1-p) + (n^2 - n)p \underbrace{(\frac{S-1}{N-1} - p)}_{= \frac{S-1}{N-1} - \frac{S}{N} = \frac{S-N}{N(N-1)}} \\ &= np(1-p) + n(n-1)p \underbrace{\frac{S-N}{N(N-1)}}_{= \frac{1}{N-1}(p-1)} \\ &= np(1-p)(1 - \frac{n-1}{N-1}) \\ &= np(1-p)\frac{N-n}{N-1}.\end{aligned}$$

Tabelle XI.3.2 fasst die Ergebnisse der vorigen Beispiele zusammen.

Tabelle XI.3.2. Erwartungswert und Varianz einiger Verteilungen

Verteilung	Erwartungswert	Varianz
binomial	np	$np(1-p)$
hypergeometrisch $h(s;n,N,S)$, $p=\frac{S}{N}$	np	$np(1-p)\frac{N-n}{N-1}$

XI.3.6. Das schwache Gesetz der großen Zahlen. Seien X_1, $\ldots, X_n$ unabhängige, reellwertige Zufallsvariable mit gleichem Erwartungswert und $\mathrm{Var}(X_i) \le M$ für alle i. Dann gilt für alle $\varepsilon > 0$:

Schwaches Gesetz der grossen Zahlen:

$$P\left(|\frac{1}{n}(X_1+\ldots+X_n)-E(X_1)| \ge \varepsilon\right) \le \frac{M}{\varepsilon^2 n} \underset{n\to\infty}{\longrightarrow} 0.$$

Es besagt, dass bei immer häufigerer, unabhängiger Wiederholung eines Experimentes das Ergebnis im Mittel fast sicher gleich dem Erwartungswert ist.

Beispiel XI.3.11. Für die Zufallsvariablen S und $S_1, \ldots, S_n$ des Bernoulli-Experimentes aus den Beispielen XI.3.5 (S. 21) und XI.3.9 haben wir für alle i

$$E(S_i) = p,$$
$$\mathrm{Var}(S_i) = p(1-p) \le \frac{1}{4}.$$

$h_n = \frac{1}{n}S$ ist die relative Häufigkeit, mit der ein Erfolg eintritt. Die Wahrscheinlichkeit, dass diese relative Häufigkeit um mehr als ε von dem Erwartungswert p abweicht, ist

$$P(|h_n - p| \ge \varepsilon) \le \frac{1}{4\varepsilon^2 n}.$$

Beispiel XI.3.12. Wir betrachten ein für uns vorteilhaftes Spiel, bei dem wir auf Dauer fast sicher verlieren.
Die Spielregeln lauten:

- Der Anfangseinsatz beträgt 1 Euro.
- Es wird immer wieder eine Münze geworfen. Fällt Zahl geht die Hälfte des aktuellen Einsatzes verloren, fällt Kopf gewinnen wir $\frac{2}{3}$ des aktuellen Einsatzes hinzu.

Da der Gewinn höher ist als der Verlust, ist dieses Spiel vorteilhaft für uns. Dennoch werden wir auf Dauer fast sicher verlieren.
Um dies einzusehen, bezeichnen wir mit X_{n-1} den Einsatz vor dem n-ten Wurf, $n \ge 1$. Dann ist $X_0 = 1$ Euro und $X_n = \frac{1}{2}X_{n-1}$, wenn im n-ten Wurf Zahl kommt, und $X_n = \frac{5}{3}X_{n-1}$, wenn im n-ten Wurf Kopf fällt. Setze $Y_n = \frac{1}{2}$, falls im n-ten Wurf Zahl kommt, und $Y_n = \frac{5}{3}$,

falls im n-ten Wurf Kopf fällt. Dann ist $X_n = Y_1 \cdot \ldots \cdot Y_n$. Die Y_i sind unabhängig. Es ist

$$\begin{aligned} E(Y_i) &= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{5}{3} \\ &= \frac{13}{12} \\ &> 1. \end{aligned}$$

Daher ist

$$\begin{aligned} E(X_n) &= \prod_{i=1}^{n} E(Y_i) \\ &= \left(\frac{13}{12}\right)^n \\ &\underset{n\to\infty}{\longrightarrow} \infty. \end{aligned}$$

Setze

$$\mu = E(\log Y_i).$$

Dann ist

$$\begin{aligned} \mu &= \frac{1}{2}\log(\frac{1}{2}) + \frac{1}{2}\log(\frac{5}{3}) \\ &< \frac{1}{2}\log(\frac{1}{2}) + \frac{1}{2}\log(2) \\ &= 0. \end{aligned}$$

Mit $\varepsilon = \frac{1}{2}|\mu| = -\frac{1}{2}\mu$ gilt nach dem schwachen Gesetz der großen Zahlen

$$P\left(|\frac{1}{n}(\log(Y_1) + \ldots + \log(Y_n)) - \mu| \le \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 1.$$

Wegen

$$\log(X_n) = \log(Y_1) + \ldots + \log(Y_n)$$

und

$$\left\{|\frac{1}{n}\log(X_n) - \mu| \le -\frac{\mu}{2}\right\} \subset \left\{\frac{1}{n}\log(X_n) \le \frac{\mu}{2}\right\}$$

folgt

$$P(\frac{1}{n}\log(X_n) \le \frac{\mu}{2}) \underset{n\to\infty}{\longrightarrow} 1.$$

Für große n gilt also mit einer Wahrscheinlichkeit nahe bei 1

$$X_n \le e^{n\frac{\mu}{2}}.$$

Wegen $\mu < 0$ strebt dieser Ausdruck exponentiell schnell gegen Null.
Dieses scheinbar paradoxe Verhalten liegt daran, dass man mit sehr kleiner Wahrscheinlichkeit sehr große Gewinne machen kann.
Wir können uns das Ergebnis heuristisch auch leicht wie folgt klar

machen: Nach $2n$ Spielrunden sollten wir etwa n-mal gewonnen und n-mal verloren haben. Dann beträgt unser Kapital

$$X_{2n} \approx \left(\frac{1}{2}\right)^n \left(\frac{5}{3}\right)^n = \left(\frac{5}{6}\right)^n.$$

Wegen $\frac{5}{6} < 1$ strebt dieser Ausdruck für $n \to \infty$ aber gegen Null.

XI.4. Grundbegriffe der Schätztheorie

XI.4.1. Motivation. Ein Teich enthält eine unbekannte Zahl N von Fischen, die geschätzt werden soll. Um dies zu erreichen, fangen wir W Fische, markieren sie mit einem weißen Fleck und setzen sie wieder aus. Wir warten eine Weile und fangen nun n Fische. Sei x die Zahl der markierten Fische in diesem Fang. Eine von dieser Zahl x abhängige Schätzung $\widehat{N}(x)$ für N erhalten wir mit folgender Plausibilitätsbetrachtung: Wenn x nicht zu klein ist, sollte der Anteil $\frac{x}{n}$ der markierten Fische im zweiten Fang etwa dem Anteil $\frac{W}{N}$ der markierten Fische am Gesamtbestand entsprechen, d.h. $\frac{x}{n} \approx \frac{W}{N}$. Da wir x, n und W kennen, können wir diese Näherungsgleichung nach N auflösen und erhalten die Schätzung

$$N \approx \widehat{N}(x) = \left\lceil \frac{Wn}{x} \right\rceil.$$

Dabei bezeichnet $\lceil a \rceil$ die größte ganze Zahl $\leq a$.

Wir können diese Plausibilitätsbetrachtung auf ein etwas sichereres mathematisches Fundament stellen. Dazu beschreiben wir den zweiten Fang durch das Modell des Ziehens ohne Rücklegen von n Kugeln aus einer Urne mit W weißen und $N - W$ schwarzen Kugeln. Die Wahrscheinlichkeit, dass x weiße Kugeln gezogen werden, ist dann

$$P_N(x) = \frac{\binom{W}{x}\binom{N-W}{n-x}}{\binom{N}{n}} \quad \text{mit } 0 \leq x \leq n.$$

Man beachte, dass die Abhängigkeit von N hier eine ganz andere ist als bisher: N ist nicht ein Ereignis, sondern ein unbekannter, zu bestimmender Parameter.

Der Maximum-Likelihood Ansatz zur Schätzung von N besagt, dass man bei gegebenem W und x den Wert von N so bestimmt, dass $P_N(x)$ maximal wird. Um N so zu bestimmen, betrachten wir

$$\begin{aligned}\frac{P_N(x)}{P_{N-1}(x)} &= \frac{\binom{W}{x}\binom{N-W}{n-x}\binom{N-1}{n}}{\binom{N}{n}\binom{W}{x}\binom{N-1-W}{n-x}} \\ &= \frac{(N-W)(N-n)}{N(N-W-n+x)}.\end{aligned}$$

Offensichtlich gilt

$$P_N(x) > P_{N-1}(x)$$

$$\iff \quad (N-W)(N-n) > N(N-W-n+x)$$
$$\iff \quad Wn > Nx.$$

Analoge Äquivalenzen gelten auch für die Beziehungen $<$ und $=$. Daher liefert der Maximum-Likelihood Ansatz die gleiche Schätzung wie unser Plausibilitätsargument.

XI.4.2. Der allgemeine Rahmen von Schätzproblemen. Zur Beschreibung eines Schätzproblemes mit endlichem Stichprobenraum benötigen wir:

- eine nicht leere, endliche Menge $\mathcal{X}$, den STICHPROBENRAUM,
- eine Familie $\{P_\vartheta : \vartheta \in \Theta\}$ von Wahrscheinlichkeitsmaßen auf $\mathcal{X}$ und
- eine zu schätzende Funktion $g(\vartheta)$.

$\mathcal{X}$ ist die Menge der möglichen Beobachtungsergebnisse. Durch die unterschiedliche Notation zu den vorigen Abschnitten – $\mathcal{X}$ statt Ω – soll betont werden, dass jedes $x \in \mathcal{X}$ tatsächlich beobachtbar sein muss. Bei wahrscheinlichkeitstheoretischen Modellen treten dagegen auch Stichproben- bzw. Ereignisräume Ω auf, deren Elemente ω nicht beobachtbar sind.

BEISPIEL XI.4.1. Im Rahmen des vorigen Abschnittes ist $\mathcal{X} = \{1, 2, \ldots, n\}$, wenn n die Zahl der Fische im zweiten Fang ist. Weiter ist $\vartheta = N$ die unbekannte Zahl der Fische im Teich. $P_\vartheta = P_N$ ist die hypergeometrische Verteilung $h(.; n, N, W)$ und die Funktion g ist gegeben durch $g(\vartheta) = \vartheta = N$.

In obigem Beispiel ist die zu schätzende Funktion $g(\vartheta) = \vartheta$. Ist dagegen z.B. die Varianz einer Binomialverteilung zu schätzen, ist $\vartheta = p$ und $g(p) = np(1-p)$.

Bezeichne mit $\mathcal{Y}$ den Wertebereich der zu schätzenden Funktion g. Jede Abbildung $T : \mathcal{X} \to \mathcal{Y}$ heißt ein SCHÄTZER von g. Diese Sprechweise lässt natürlich auch völlig unsinnige Schätzer zu.

Häufig deutet man in der Notation an, dass geschätzt wird, und setzt ein „Dach“ über die zu schätzende Größe. $\widehat{N}$ wäre also ein Schätzer für N, $\widehat{p}$ ein Schätzer für p und $\widehat{g}$ ein Schätzer für g.

XI.4.3. Maximum-Likelihood Schätzer. Sei $x \in \mathcal{X}$ eine feste Beobachtung. Die Funktion

$$L_x : \Theta \to [0, 1]$$

mit

$$L_x(\vartheta) = P_\vartheta(x)$$

nennt man LIKELIHOOD-FUNKTION. Wenn L_x einen Maximalwert in $\widehat{\vartheta} = \widehat{\vartheta}(x)$ annimmt, d.h.

$$L_x(\widehat{\vartheta}) = \sup\{L_x(\vartheta) : \vartheta \in \Theta\},$$

nennt man $\widehat{\vartheta}(x)$ eine MAXIMUM-LIKELIHOOD SCHÄTZUNG von ϑ und $g(\widehat{\vartheta})$ eine Maximum-Likelihood Schätzung von $g(\vartheta)$.

Häufig ist Θ ein Intervall in $\mathbb{R}$, und eine Maximum-Likelihood Schätzung kann wie in Abschnitt IV.2.1 (S. 136, Teil I) durch Differentiation gefunden werden. Dabei ist es oft einfacher mit der Funktion $\mathcal{L}_x = \ln L_x$ zu arbeiten. Wegen der Monotonie des Logarithmus haben $\mathcal{L}_x$ und L_x ihre Maxima an der gleichen Stelle.

BEISPIEL XI.4.2. In n Bernoulli-Experimenten soll die Erfolgswahrscheinlichkeit p aus der Zahl x der Erfolge geschätzt werden. Es ist

$$\begin{aligned}
L_x(p) &= \binom{n}{x} p^x (1-p)^{n-x} \\
\mathcal{L}_x(p) &= \ln L_x(p) \\
&= \ln \binom{n}{x} + x \ln p + (n-x)\ln(1-p) \\
\frac{d}{dp}\mathcal{L}_x(p) &= \frac{x}{p} - \frac{n-x}{1-p}.
\end{aligned}$$

Also liegt das eindeutige Extremum bei $\frac{x}{n}$. Wie man leicht nachprüft, ist dies auch ein Maximum. $\frac{x}{n}$ ist die Maximum-Likelihood Schätzung für die Erfolgswahrscheinlichkeit p.

XI.4.4. Erwartungstreue. Ist $T : \mathcal{X} \to \mathbb{R}$ ein Schätzer, so bezeichnen wir den Erwartungswert von T bzgl. P_ϑ mit $E_\vartheta(T)$:

$$E_\vartheta(T) = \sum_{x \in \mathcal{X}} T(x) P_\vartheta(x).$$

Die Beobachtung des Experimentes, das dem Schätzproblem zugrunde liegt, können wir als eine Zufallsvariable X auffassen, d.h. $X(x) = x$. Mit dieser Notation können wir den Begriff der Erwartungstreue definieren:

Ein Schätzer $\widehat{g}$ für $g(\vartheta)$ heißt ERWARTUNGSTREU, wenn für alle $\vartheta \in \Theta$ gilt

$$E_\vartheta(\widehat{g}(X)) = g(\vartheta).$$

Speziell heißt $\widehat{\vartheta}$ ein ERWARTUNGSTREUER SCHÄTZER von ϑ, wenn für alle $\vartheta \in \Theta$ gilt

$$E_\vartheta(\widehat{\vartheta}(X)) = \vartheta.$$

Die Differenz

$$b(\vartheta, \widehat{g}) = E_\vartheta(\widehat{g}(X)) - g(\vartheta)$$

heißt der BIAS (wörtlich: Vorurteil) des Schätzers $\widehat{g}$. Ein Schätzer ist also genau dann erwartungstreu (engl.: UNBIASED), wenn sein Bias gleich Null ist.

Beispiel XI.4.3. Ist X binomial verteilt mit den Parametern n und p, so ist $E(\frac{X}{n}) = p$ (vgl. Abschnitt XI.3.5 (S. 24)). Daher ist die Schätzung $\widehat{p}(X) = \frac{X}{n}$ aus Beispiel XI.4.2 erwartungstreu.

Beispiel XI.4.4. Häufig erfolgt die Messung einer unbekannten Größe μ durch Ausführen von n unabhängigen Zufallsexperimenten, die μ als Erwartungswert haben. Mathematisch bedeutet dies, dass wir n unabhängige Zufallsvariable $X_1, \ldots, X_n$ und eine unbekannte Familie P_ϑ von Wahrscheinlichkeiten haben, so dass $E_\vartheta(X_i) = \mu$ ist für alle i und ϑ. Sei

$$\begin{aligned} g_1(\vartheta) &= E_\vartheta(X_i) \\ &= \mu \end{aligned}$$

und

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

der Mittelwert. Dann ist für jedes ϑ

$$\begin{aligned} E_\vartheta(\overline{X}) &= \frac{1}{n} \sum_{i=1}^{n} E_\vartheta(X_i) \\ &= \mu. \end{aligned}$$

Also ist der Mittelwert ein erwartungstreuer Schätzer für g_1.
Haben die X_i alle die gleiche unbekannte Varianz σ^2, so bestimmt die Familie P_ϑ auch diese Größe. Setze in diesem Fall

$$\begin{aligned} g_2(\vartheta) &= \sigma^2 \\ &= \mathrm{Var}(X_i). \end{aligned}$$

Wir wählen

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

als Schätzer für σ^2. (Beachte, dass durch $n-1$ geteilt wird!) Da die X_i unabhängig sind, erhalten wir für jedes ϑ

$$\begin{aligned} E_\vartheta((\overline{X} - \mu)^2) &= \mathrm{Var}_\vartheta(\overline{X}) \\ &= \frac{1}{n^2} \mathrm{Var}_\vartheta(X_1 + \ldots + X_n) \\ &= \frac{1}{n^2} (\mathrm{Var}_\vartheta(X_1) + \ldots + \mathrm{Var}_\vartheta(X_n)) \\ &= \frac{\sigma^2}{n} \end{aligned}$$

und

$$\begin{aligned}
E_\vartheta((X_i - \overline{X})^2) &= E_\vartheta(((X_i - \mu) - (\overline{X} - \mu))^2) \\
&= E_\vartheta((X_i - \mu)^2) - 2E_\vartheta((X_i - \mu)(\overline{X} - \mu)) \\
&\quad + E_\vartheta((\overline{X} - \mu)^2) \\
&= \sigma^2 - \frac{2}{n} \sum_{j=1}^{n} \underbrace{E_\vartheta((X_i - \mu)(X_j - \mu))}_{=0 \text{ für} j \neq i} + \frac{\sigma^2}{n} \\
&= \sigma^2 - \frac{2}{n} E_\vartheta((X_i - \mu)^2) + \frac{\sigma^2}{n} \\
&= \sigma^2 - \frac{\sigma^2}{n} \\
&= \sigma^2 (1 - \frac{1}{n}).
\end{aligned}$$

Hieraus folgt

$$\begin{aligned}
E_\vartheta(s^2) &= \frac{1}{n-1} \sum_{i=1}^{n} E_\vartheta((X_i - \overline{X})^2) \\
&= \frac{1}{n-1} n\sigma^2 (1 - \frac{1}{n}) \\
&= \sigma^2.
\end{aligned}$$

Also ist s^2 ein erwartungstreuer Schätzer für die Unbekannte Varianz σ^2.

XI.4.5. Der mittlere quadratische Fehler. Die wohl wichtigste Forderung an einen Schätzer T von $g(\vartheta)$ ist wohl diejenige, dass der Fehler $T(X) - g(\vartheta)$ der Schätzwerte „klein" sein sollte. Um diese Größe quantitativ zu messen, betrachtet man den MITTLEREN QUADRATISCHEN FEHLER

$$\boxed{R(\vartheta, T) = E_\vartheta((T(X) - g(\vartheta))^2).}$$

Man kann den mittleren quadratischen Fehler auch durch die Varianz und den Bias ausdrücken. Berücksichtigt man, dass $E_\vartheta(T) - g(\vartheta)$ eine Konstante ist, erhält man nämlich

$$\begin{aligned}
&E_\vartheta\left(\Big(T(X) - g(\vartheta)\Big)^2\right) \\
&= E_\vartheta\left(\Big(\big(T(X) - E_\vartheta(T)\big) - \big(g(\vartheta) - E_\vartheta(T)\big)\Big)^2\right)
\end{aligned}$$

$$
\begin{aligned}
&= \underbrace{E_\vartheta\left(\left(T(X) - E_\vartheta(T)\right)^2\right)}_{=\mathrm{Var}_\vartheta(T)} \\
&\quad - 2\underbrace{E_\vartheta\left(\left(T(X) - E_\vartheta(T)\right)\left(g(\vartheta) - E_\vartheta(T)\right)\right)}_{=(g(\vartheta)-E_\vartheta(T))E_\vartheta(T(X)-E_\vartheta(T))=0} \\
&\quad + \underbrace{E_\vartheta\left(\left(g(\vartheta) - E_\vartheta(T)\right)^2\right)}_{=(g(\vartheta)-E_\vartheta(T))^2} \\
&= \mathrm{Var}_\vartheta(T) + b(\vartheta, T)^2.
\end{aligned}
$$

Also ist der mittlere quadratische Fehler

$$\boxed{R(\vartheta, T) = \mathrm{Var}_\vartheta(T) + b(\vartheta, T)^2.}$$

Beispiel XI.4.5. Für den Schätzer $T(X) = \frac{X}{n}$ aus Beispiel XI.4.3 (S. 33) für die binomial verteilte Zufallsvariable X gilt gemäß Beispiel XI.4.3 und Abschnitt XI.3.5 (S. 24)

$$
\begin{aligned}
b(\vartheta, T) &= 0, \\
\mathrm{Var}_\vartheta(T) &= \frac{1}{n^2}\,\mathrm{Var}_\vartheta(X) \\
&= \frac{1}{n}\vartheta(1-\vartheta).
\end{aligned}
$$

Also ist der mittlere quadratische Fehler

$$R(\vartheta, T) = \frac{1}{n}\vartheta(1-\vartheta).$$

Für alle $\vartheta \in \Theta = (0,1)$ gilt

$$R(\vartheta, T) \le \frac{1}{4n}.$$

Beispiel XI.4.6. Wir greifen Beispiel XI.4.4 (S. 33) auf. Der Bias des Mittelwertes $\overline{X}$ ist Null; seine Varianz ist gemäß Beispiel XI.4.4 gleich $\frac{\sigma^2}{n}$, wenn alle X_i die gleiche Varianz σ^2 haben. Also ist der mittlere quadratische Fehler

$$R(\vartheta, \overline{X}) = \frac{\sigma^2}{n}.$$

XI.5. Approximationen der Binomialverteilung

Für große Werte von n ist die exakte Berechnung der Wahrscheinlichkeit

$$b_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

äußerst mühselig und anfällig für Rundungsfehler. Dies gilt vermehrt für Summen dieser Ausdrücke. Wir wollen daher in diesem Abschnitt Approximationen für $b_{n.p}(k)$ bestimmen, die leichter und mit geringeren Rundungsfehlern berechnet werden können.

Zur Vereinfachung der Sprache führen wir die Notation

$$\boxed{a_n \sim b_n \quad \Longleftrightarrow \quad \lim_{n\to\infty} \frac{a_n}{b_n} = 1}$$

ein. $a_n \sim b_n$ bedeutet, dass es eine Nullfolge $(\theta_n)_{n\in\mathbb{N}}$ gibt mit

$$b_n(1-\theta_n) \le a_n \le b_n(1+\theta_n)$$

für alle bis auf endlich viele n.

XI.5.1. Approximation von $n!$ und $b_{n,p}(k)$. Unser wichtigstes Hilfsmittel ist die Stirlingsche Formel

$$\boxed{n! \sim \sqrt{2\pi n}\, n^n e^{-n}.}$$

Für die eingangs erwähnte Folge $(\theta_n)_{n\in\mathbb{N}}$ kann man in diesem Fall zeigen:

$$\frac{1}{12n+1} \le \theta_n \le \frac{1}{12n}$$

für alle n.

Tabelle XI.5.1 gibt einen Eindruck von der Genauigkeit der Stirlingschen Formel.

Tabelle XI.5.1. Genauigkeit der Stirlingschen Formel

n	$n!$	$\sqrt{2\pi n}\, n^n e^{-n}$	$\frac{\lvert n!-\sqrt{2\pi n}\, n^n e^{-n}\rvert}{n!}$
2	2	1.919	4.05%
5	120	118.019	1.65%
10	3628800	3598690	0.83%

Mit der Stirlingschen Formel folgt

$$\begin{aligned} b_{n,p}(k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &\sim \frac{\sqrt{2\pi n}\, n^n e^{-n}}{\sqrt{2\pi k}\, k^k e^{-k} \sqrt{2\pi(n-k)}\, (n-k)^{(n-k)} e^{-(n-k)}}\, p^k (1-p)^{n-k} \\ &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{n(1-p)}{n-k}\right)^{n-k}. \end{aligned}$$

Wir betrachten nun eine Folge $(k_n)_{n\in\mathbb{N}}$ mit $k_n \sim np$. Dann ist $n - k_n \sim n(1-p)$ und

$$\begin{aligned}\sqrt{\frac{n}{k_n(n-k_n)}} &\sim \frac{1}{\sqrt{np(1-p)}} \\ &= \frac{1}{\sigma_n},\end{aligned}$$

wobei

$$\sigma_n = \sqrt{np(1-p)}$$

die Standardabweichung der $b_{n,p}$-Verteilung ist (vgl. Abschnitt XI.3.5 (S. 24)). Wir wollen nun das Verhalten von

$$\chi(n, k_n) = \left(\frac{np}{k_n}\right)^{k_n} \left(\frac{n(1-p)}{n-k_n}\right)^{n-k_n}$$

studieren. Dazu bilden wir den Logarithmus und setzen

$$t_n = \frac{k_n}{n}.$$

Dann gilt $t_n \to p$ für $n \to \infty$ und

$$\begin{aligned}-\ln\chi(n, k_n) &= -k_n \ln(\frac{np}{k_n}) - (n-k_n)\ln(\frac{n(1-p)}{n-k_n}) \\ &= n\left\{\frac{k_n}{n}\ln(\frac{t_n}{p}) + (1-\frac{k_n}{n})\ln(\frac{1-\frac{k_n}{n}}{1-p})\right\} \\ &= n\left\{t_n\ln(\frac{t_n}{p}) + (1-t_n)\ln(\frac{1-t_n}{1-p})\right\}.\end{aligned}$$

Definiere die Funktion $g : (0,1) \to \mathbb{R}$ durch

$$g(t) = t\ln(\frac{t}{p}) + (1-t)\ln(\frac{1-t}{1-p}).$$

Dann ist

$$\begin{aligned}g(p) &= 0, \\ g'(p) &= 0, \\ g''(p) &= \frac{1}{p} + \frac{1}{1-p} \\ &= \frac{1}{p(1-p)}.\end{aligned}$$

Nach der Taylor Formel ist daher

$$g(t) = \frac{1}{2p(1-p)}(t-p)^2 + r(t-p) \quad \text{mit } |r(t-p)| \le c|t-p|^3$$

in einer Umgebung von p.

Wir nehmen nun an, dass nicht nur $t_n \to p$ gilt, sondern dass sogar $n(t_n - p)^3 \to 0$ gilt. Dann folgt $n|r(t_n - p)| \to 0$ und daher

$$\left| -\ln \chi(n, k_n) - \frac{n(t_n - p)^2}{2p(1-p)} \right| \underset{n\to\infty}{\longrightarrow} 0.$$

Setzen wir

$$x(n, k_n) = \frac{k_n - np}{\sigma_n},$$

so ist

$$\frac{n(t_n - p)^2}{2p(1-p)} = \frac{x(n, k_n)^2}{2}.$$

Damit erhalten wir

$$\chi(n, k_n) \sim \exp\left(-\frac{x(n, k_n)^2}{2}\right).$$

Insgesamt haben wir somit unter obigen Voraussetzungen die Beziehung

$$b_{n,p}(k_n) \sim \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_n} \exp\left(-\frac{x(n, k_n)^2}{2}\right)$$

gezeigt.

Die Funktion

$$\boxed{\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)}$$

heißt Dichte der Standard-Normalverteilung. Damit lässt sich unsere Approximation der $b_{n,p}$-Verteilung schreiben als

$$b_{n,p}(k) \sim \frac{1}{\sigma_n} \varphi(x(n, k))$$

mit

$$\begin{aligned} \sigma_n &= \sqrt{np(1-p)}, \\ x(n, k) &= \frac{k - pn}{\sigma_n}. \end{aligned}$$

Mit Hilfe von Ergebnissen der Funktionentheorie folgt

$$\boxed{\int_{-\infty}^{\infty} \varphi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 1.}$$

XI.5.2. Der Satz von Moivre-Laplace. Wir wollen Summen von Wahrscheinlichkeiten $b_{n,p}(k)$ für große n näherungsweise berechnen. Dazu definieren wir

$$\boxed{\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.}$$

Die Funktion Φ heißt die Verteilungsfunktion der Standard-Normalverteilung. Sie hat folgende Eigenschaften

$$\boxed{\begin{aligned} \lim_{x\to-\infty} \Phi(x) &= 0 \\ \lim_{x\to+\infty} \Phi(x) &= 1 \\ \Phi(0) &= \frac{1}{2} \\ \Phi(-x) &= 1 - \Phi(x) && \text{für } x \geq 0 \\ \Phi(x) &\leq \Phi(y) && \text{für } x \leq y \end{aligned}}$$

Sie ist ausführlich tabelliert (z.B. in U. Krengel: Einführung in die Wahrscheinlichkeitstheorie und Statistik, Vieweg 2000, S. 248).

Sei nun S_n eine $b_{n,p}$-verteilte Zufallsvariable. Setze

$$\begin{aligned} S_n^* &= \frac{S_n - np}{\sigma_n} \\ &= \frac{S_n - E(S_n)}{\sqrt{\mathrm{Var}(S_n)}}. \end{aligned}$$

Dann ist $E(S_n^*) = 0$ und $\mathrm{Var}(S_n^*) = 1$. S_n^* heißt daher die standardisierte oder normierte Form von S_n. Nimmt S_n den Wert k an, so hat S_n^* den Wert $x(n,k)$ mit $x(n,k)$ aus dem vorigen Abschnitt.

Der Satz von Moivre-Laplace besagt, dass Φ eine gute Approximation für die Verteilung von S_n^* ist:

$$\boxed{\lim_{n\to\infty} P(a \leq S_n^* \leq b) = \Phi(b) - \Phi(a).}$$

Beispiel XI.5.1. Wie groß ist näherungsweise die Wahrscheinlichkeit, bei 600 Würfen mit einem Würfel mindestens 90 und höchstens 100 Sechsen zu erhalten?
Es ist

$$n = 600$$

und

$$p = \frac{1}{6},$$

also

$$np = 100$$

und

$$\begin{aligned}\sigma_n &= \sqrt{600 \cdot \frac{1}{6} \cdot \frac{5}{6}} \\ &= \sqrt{\frac{250}{3}} \\ &\approx 9.13\end{aligned}$$

Damit folgt

$$\begin{aligned}P(90 \le S_n \le 100) &= P(\frac{90-100}{\sigma_n} \le S_n^* \le \frac{100-100}{\sigma_n}) \\ &\approx \Phi(0) - \Phi(-\frac{10}{9.13}) \\ &= 0.5 - \Phi(-1.095) \\ &= 0.5 - (1 - \Phi(1.095)) \\ &\approx 0.36.\end{aligned}$$

Beispiel XI.5.2. Wir wollen den Prozentsatz der Wähler der Partei A schätzen. Werden n Wähler befragt und sind darunter S_n Wähler der Partei A, so sei $\frac{S_n}{n}$ der Schätzer für die Wahrscheinlichkeit p, dass ein zufällig ausgewählter Wähler die Partei A wählt. Wie groß muss n sein, damit die Wahrscheinlichkeit eines Irrtums um 1% nicht größer ist als 0.05?

Es muss also gelten

$$P(-0.01 \le \frac{S_n}{n} - p \le 0.01) \approx 0.95.$$

Mit

$$\sigma_n = \sqrt{np(1-p)}$$

erhalten wir

$$\begin{aligned}0.95 &\approx P(-\frac{0.01n}{\sigma_n} \le S_n^* \le \frac{0.01n}{\sigma_n}) \\ &\approx \Phi(\frac{0.01n}{\sigma_n}) - \Phi(-\frac{0.01n}{\sigma_n}) \\ &= 2\Phi(\frac{0.01n}{\sigma_n}) - 1\end{aligned}$$

also

$$\Phi(\frac{0.01n}{\sigma_n}) \approx 0.975.$$

Da Φ streng monoton ist, existiert die Umkehrfunktion Φ^{-1}. Es muss also gelten

$$\frac{0.01n}{\sigma_n} \approx \Phi^{-1}(0.975).$$

Aus einer Tabelle von Φ entnehmen wir

$$\Phi^{-1}(0.975) \approx 1.96.$$

Also muss gelten

$$\begin{aligned} 1.96 &\approx \frac{0.01}{\sigma_n} \\ &= \frac{0.01\sqrt{n}}{\sqrt{p(1-p)}} \\ \implies \quad \sqrt{n} &\approx 196\sqrt{p(1-p)} \\ \implies \quad n &\approx 196^2 p(1-p). \end{aligned}$$

Wir kennen aber p nicht! Aber wir wissen, dass für alle möglichen p, nämlich $0 \le p \le 1$, gilt

$$p(1-p) \le \frac{1}{4}.$$

Damit erhalten wir die Schätzung

$$\begin{aligned} n &\approx \frac{196^2}{4} \\ &= 98^2 \\ &= 9604. \end{aligned}$$

Wir müssen also ca. 9600 Wähler befragen.
Hätten wir z.B. die Zusatzinformation $p \le 0.1$ zur Verfügung, kämen wir wegen

$$\max_{0\le p\le 0.1} p(1-p) = 0.1 \cdot 0.9 = 0.09$$

mit ca. 3450 Befragungen aus.

XI.5.3. Die Poisson-Approximation. In diesem Abschnitt geben wir eine andere Approximation der $b_{n,p}$-Verteilung an, die für kleine Werte von p besser ist als die Normalverteilung aus Abschnitt XI.5.1.

Eine Zufallsvariable X heißt POISSON-VERTEILT mit Parameter $\lambda \ge 0$, wenn gilt

$$\boxed{P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{mit } k = 0, 1, \dots}$$

Ist $(p_n)_{n\in\mathbb{N}}$ eine Folge in $[0, 1]$ mit $np_n \to \lambda$, so gilt für jedes k die Poisson-Approximation

$$\boxed{\begin{aligned} b_{n,p_n}(k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &\underset{n\to\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^k}{k!}. \end{aligned}}$$

Um dies einzusehen, setze $\lambda_n = np_n$. Dann gilt für festes k

$$\begin{aligned} b_{n,p_n}(k) &= \frac{n!}{k!(n-k)!}\left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\ &= \frac{1}{k!}\frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{n^k}\lambda_n^k\left(1 - \frac{\lambda_n}{n}\right)^n\left(1 - \frac{\lambda_n}{n}\right)^{-k} \\ &= \frac{1}{k!}\frac{n}{n}\cdot\frac{n-1}{n}\cdot\ldots\cdot\frac{n-k+1}{n}\lambda_n^k\left(1 - \frac{\lambda_n}{n}\right)^n\left(1 - \frac{\lambda_n}{n}\right)^{-k}. \end{aligned}$$

Für $n \to \infty$ streben die Quotienten $\frac{n}{n}$, $\frac{n-1}{n}$, $\dots$, $\frac{n-k+1}{n}$ alle gegen 1; nach Voraussetzung gilt

$$\lambda_n^k \to \lambda^k$$

und

$$(1 - \frac{\lambda_n}{n})^{-k} \to 1.$$

Wegen der Stetigkeit der Exponentialfunktion gilt

$$(1 - \frac{\lambda_n}{n})^n \to e^{-\lambda}.$$

Damit folgt die Poisson-Approximation.

Tabelle XI.5.2. Poisson-Approximation für die Wahrscheinlichkeit, dass unter 91 Personen genau k am heutigen Tage Geburtstag haben

k	$P(X = k) \approx \frac{0.25^k}{k!}e^{-0.25}$
1	0.1947
2	0.0243
5	$6.337 \cdot 10^{-6}$

Beispiel XI.5.3. In einem Hörsaal befinden sich 91 Studierende. Die Wahrscheinlichkeit p, heute Geburtstag zu haben, ist $p = \frac{1}{365}$. Die Zahl derer, die heute Geburtstag haben, ist praktisch Poisson-verteilt mit $\lambda = \frac{91}{365} \approx 0.25$. Tabelle XI.5.2 gibt die mit der Poisson-Approximation genäherte Wahrscheinlichkeit $P(X = k)$ an, dass heute genau k Personen aus dem Hörsaal Geburtstag haben.

Beispiel XI.5.4. Bei der Produktion von Blitzlichtlampen ist mit der Wahrscheinlichkeit $p = 0.015$ eine Lampe schon bei der Produktion defekt. Wie groß muss man n wählen, damit in einem Karton mit n unabhängig gewählten Lampen mit Wahrscheinlichkeit ≥ 0.8 mindestens 100 funktionstüchtige Exemplare sind?
Die Zahl n ist minimal zu wählen unter der Bedingung

$$\begin{aligned} 0.8 &\leq P(\text{„höchstens } n-100 \text{ fehlerhafte Lampen“}) \\ &= \sum_{k=0}^{n-100} b_{n,p}(k). \end{aligned}$$

Approximieren wir die rechte Seite mit der Poisson-Verteilung zu $\lambda_n = np$, erhalten wir die Bedingung

$$0.8 \leq r_n = e^{-\lambda_n} \sum_{k=0}^{n-100} \frac{\lambda_n^k}{k!}.$$

Mit Hilfe eines Taschenrechners prüfen wir nach, dass $n = 102$ die kleinste Zahl ist, die diese Bedingung erfüllt. Wir müssen also 102 Lampen in den Karton legen.
Würden wir die Wahrscheinlichkeit auf 0.9 erhöhen, müssten wir 103 Lampen in den Karton legen.

XI.6. Tests

XI.6.1. Motivation. Eine Person behauptet, dass sie bei Milchkaffee schmecken kann, ob zuerst der Kaffee oder die Milch in die Tasse gegeben wurde. Wie können wir die Stichhaltigkeit dieser Behauptung nachprüfen?

Ein erster Ansatz geht auf Fisher (1935) zurück. Er schlägt folgende „Versuchsanordnung“ vor:

> In vier Tassen wird zuerst Kaffee und dann Milch gegossen. Diese Tassen nennen wir Tassen vom Typ 1. In weitere vier Tassen wird zuerst Milch und dann Kaffee gegossen. Diese Tassen nennen wir vom Typ 2. Die Kaffee- und Milchmenge ist dabei natürlich bei allen Tassen die gleiche, und es wird gut umgerührt. Nun werden der Person die acht Tassen in beliebiger Reihenfolge (8! Möglichkeiten) vorgesetzt, und sie soll die vier Tassen vom Typ 1

> identifizieren. Falls ihr das gelingt, wollen wir ihre Behauptung als wahr akzeptieren.

Die Wahrscheinlichkeit, dass die Person rein zufällig richtig rät, ist aufgrund der hypergeometrischen Verteilung aus Abschnitt XI.1.4 (S. 10) gleich

$$\binom{8}{4}^{-1} = \frac{1}{70}.$$

Wir geben der Person also nur mit dieser Wahrscheinlichkeit Recht, wenn ihre Behauptung nicht stimmt.

Schwieriger wird die Angelegenheit, wenn die Person behauptet, nicht unfehlbar zu sein, aber mit großer Sicherheit die Unterscheidung treffen zu können. Wenn wir ihr die Behauptung in diesem Fall schon dann glauben, wenn sie mindestens drei der vier Tassen vom Typ 1 richtig identifiziert, ist die Wahrscheinlichkeit für einen zufälligen Treffer schon

$$\begin{aligned} \frac{\binom{4}{4}\binom{4}{0} + \binom{4}{3}\binom{4}{1}}{\binom{8}{4}} &= \frac{1 + 4 \cdot 4}{70} \\ &= \frac{17}{70} \\ &\approx 0.24. \end{aligned}$$

Wie kann man in diesem Fall die Gewissheit, der Person auf die Schliche zu kommen, erhöhen, ohne sie dabei zu benachteiligen?

Neyman (1950) schlägt dazu die folgende verbesserte „Versuchsanordnung“ vor:

> Der Person wird n-mal die Aufgabe gestellt, zwei Tassen, von denen eine vom Typ 1 und eine vom Typ 2 ist, korrekt zu klassifizieren. Die beiden Tassen werden ihr jeweils in einer zufälligen, durch Münzwurf bestimmten Reihenfolge vorgesetzt. Damit die Person unabhängig von früheren Entscheidungen urteilen kann, wird jeder Teilversuch an einem anderen Tag ausgeführt.

X sei die Zahl der Tage, an denen die Person die Tassen richtig klassifiziert.

Ein mathematisches Modell für diese Versuchsanordnung ist die Annahme, X sei binomial verteilt mit Parametern n und p. Die Hypothese „die Person schwindelt und trifft ihre Entscheidung rein zufällig“ entspricht dem Fall $p = \frac{1}{2}$. Diese Hypothese nennt man die „Nullhypothese“. Die Alternative ist, dass die Person tatsächlich die behauptete Fähigkeit hat, und entspricht dem Fall $p > \frac{1}{2}$. Es wird nun eine Zahl t festgelegt, so dass $P(X \geq t) \leq \alpha$ ist. Dabei ist α eine vorgegebene, kleine, positive Schranke, z.B. $\alpha = 0.05$. Dies ist die „Sicherheitsmarge“, mit der wir ausschließen wollen, dass wir der Person ungerechtfertigterweise auf den Leim gehen. Falls die Person nun mindestens t-mal die Tassen richtig klassifiziert, wollen wir ihr glauben.

Der Unterschied des Neymanschen Ansatzes zu dem von Fisher liegt darin, dass er präzise Aussagen über die Wahrscheinlichkeit erlaubt, die Nullhypothese zu akzeptieren, wenn die Alternative zutrifft.

XI.6.2. Grundbegriffe der Testtheorie. Von einem Testproblem spricht man, wenn eine zufällige Größe X mit einer unbekannten Verteilung P_ϑ beobachtet wird, und man aufgrund des beobachteten Wertes x von X entscheiden soll, ob P_ϑ einer bestimmten Menge von Verteilungen angehört oder nicht.

Zur genaueren mathematischen Beschreibung sei $\mathcal{X}$ die Menge der möglichen Werte der Zufallsvariablen X und $\{P_\vartheta : \vartheta \in \Theta\}$ die Menge der in Betracht gezogenen Verteilungen von X. Unter diesen möglichen Verteilungen ist eine echte Teilmenge $\{P_\vartheta : \vartheta \in H\}$ durch zusätzliche Bedingungen ausgezeichnet. Ein TEST ist eine Entscheidungsregel, die für jeden möglichen Wert von X festlegt, ob man sich für die HYPOTHESE „$\vartheta \in H$“ oder für die ALTERNATIVE „$\vartheta \in \Theta\backslash H$“ entscheiden soll. Man nennt auch kurz H die Hypothese und $K = \Theta\backslash H$ die Alternative. Die Entscheidung für die Hypothese nennt man ANNAHME DER HYPOTHESE, und die Entscheidung für die Alternative nennt man VERWERFEN DER HYPOTHESE.

Ein Test ist also beschrieben durch Angabe der Menge R derjenigen Werte x von X, für die die Hypothese verworfen werden soll. R heißt VERWERFUNGSBEREICH oder KRITISCHER BEREICH des Tests.

Innerhalb des gewählten Modells sind zwei Arten von Fehlern möglich:

- Ist $\vartheta \in H$ und wird die Hypothese verworfen, spricht man von einem FEHLER ERSTER ART.
- Ist $\vartheta \in K$ und wird die Hypothese angenommen, spricht man von einem FEHLER ZWEITER ART.

Praktisch beschreibt man den Verwerfungsbereich R durch eine Funktion $T(x)$, die so gewählt wird, dass große Werte gegen die Hypothese sprechen. T heißt TESTSTATISTIK. Man wählt dann einen KRITISCHEN WERT t und verwirft die Hypothese, wenn $T(x) \geq t$ ist. Es ist also

$$R = \{x : T(x) \geq t\}.$$

Bisher ist das Testproblem so formuliert, dass H und K symmetrische Rollen spielen. Dies ist aber in der Praxis in der Regel nicht der Fall. Vielmehr wird man die Hypothese so wählen, dass sie der etablierten Hypothese oder der bisherigen Erfahrung entspricht. Bei unserem einführenden Beispiel bedeutet dies, dass man als Hypothese

„die Person schwindelt“ nehmen wird. Bei der Einführung eines neuen Medikamentes würde man z.B. die Hypothese „das Medikament ist nicht wirksamer als die bekannten Mittel“ wählen.

Man zieht nur Verwerfungsbereiche in Betracht, für die die Wahrscheinlichkeit eines Fehlers erster Art durch eine vorgegeben kleine Zahl $\alpha > 0$ begrenzt ist. Quantitative Aussagen erhält man durch die GÜTEFUNKTION

$$\beta(\vartheta) = P_\vartheta(X \in R).$$

Sie ordnet jedem ϑ die Verwerfungswahrscheinlichkeit unter P_ϑ zu. Der Test hat das Niveau α, wenn für alle $\vartheta \in H$ die Ungleichung $\beta(\vartheta) \le \alpha$ gilt. Die Wahrscheinlichkeit eines Fehlers erster Art ist dann also maximal α. Gebräuchliche Werte für α sind 0.05, 0.02 oder 0.01.

XI.6.3. Zurück zur Motivation. In unserem motivierenden Beispiel des ersten Abschnittes ist das gesuchte Modell beschrieben durch

$$\begin{aligned}
\mathcal{X} &= \{0, 1, \ldots, n\}, \\
\Theta &= [\frac{1}{2}, 1], \\
\vartheta &= p, \\
P_\vartheta(X = x) &= b_{n,p}(x) \\
&= \binom{n}{x} p^x (1-p)^{n-x}.
\end{aligned}$$

Die Hypothese ist

$$H = \{\frac{1}{2}\}$$

und die Alternative ist

$$K = (\frac{1}{2}, 1].$$

Der Verwerfungsbereich ist von der Form

$$R = \{x : x \ge t\}.$$

Sei

$$\beta(p|t, n) = P_p(X \ge t)$$

die Gütefunktion dieses Tests. Als Niveau wollen wir

$$\alpha = 0.05$$

wählen. Wir müssen dann für gegebenes n den Parameter t so wählen, dass

$$\beta(\frac{1}{2}|t, n) \le 0.05$$

ist.

Für $n = 5$ kommt nur $t = 5$ in Frage. Denn für $t = 4$ ist

$$\begin{aligned}\beta(\frac{1}{2}|4,5) &= (\frac{1}{2})^5 + \binom{5}{1}(\frac{1}{2})^5 \\ &= \frac{6}{32} \\ &\approx 0.187 \\ &> 0.05.\end{aligned}$$

Die Person müsste also in jedem der 5 Versuche die Tassen richtig klassifizieren, damit wir ihr glauben.

Für $n = 5$ ist

$$\beta(p|5,5) = p^5.$$

Es ist

$$\beta(0.6|5,5) \approx 0.08$$

und

$$\beta(0.9|5,5) \approx 0.59.$$

Hätte also die Person eine Erfolgswahrscheinlichkeit von 0.6 pro Klassifizierung, würde sie doch nur mit Wahrscheinlichkeit 0.08 ihre Fähigkeit beweisen können. Selbst bei einer Erfolgswahrscheinlichkeit von 0.9, würde ihre Behauptung nur mit Wahrscheinlichkeit 0.59 akzeptiert. Die Gütefunktion zeigt also, ob der Test überhaupt in der Lage ist, eine Abweichung von der Hypothese anzuzeigen.

Ist z.B. $p = 0.6$, ist erst bei $n = 42$ zu klassifizierenden Tassenpaaren die Wahrscheinlichkeit wenigstens $\frac{1}{3}$, dass die Behauptung der Person akzeptiert wird. Der kleinste Wert von t mit $\beta(\frac{1}{2}|t, 42) \leq 0.05$ ist $t = 27$. Die Person müsste also mindestens 27von 42 Tassenpaaren richtig klassifizieren.

KAPITEL XII

Stochastik II
Allgemeine Modelle

XII.1. Wahrscheinlichkeitsmaße mit Dichten

XII.1.1. Ergebnismengen. Im vorigen Kapitel haben wir stets endliche Ergebnismengen Ω betrachtet. Die Bedingung $\operatorname{card}(\Omega) < \infty$ lassen wir jetzt fallen und betrachten allgemeine Ergebnismengen Ω. Die wichtigsten Beispiele für dieses Kapitel sind $\Omega = \mathbb{R}$ und $\Omega = \mathbb{R}^n$.

XII.1.2. σ-Algebren. Bisher waren die Wahrscheinlichkeitsmaße stets auf der ganzen Potenzmenge $\mathcal{P}(\Omega)$ definiert, d.h. jede Teilmenge von Ω war ein mögliches Ereignis. Man kann zeigen, dass dies für unendliche Mengen wie $\Omega = \mathbb{N}$ oder $\Omega = \mathbb{R}$ zu Widersprüchen führt. Daher können wir nur noch Teilmengen $\mathcal{A}$ von $\mathcal{P}(\Omega)$ als mögliche Ereignismengen betrachten. Diese müssen aber gewisse Eigenschaften haben.

Der richtige Begriff ist hier derjenige der σ-Algebra. Eine Menge $\mathcal{A} \subset \mathcal{P}(\Omega)$ von Teilmengen von Ω heißt σ-ALGEBRA (über Ω), wenn folgende drei Eigenschaften erfüllt sind:

$$\boxed{\begin{aligned} &\Omega \in \mathcal{A} \\ &A \in \mathcal{A} \implies \Omega\backslash A \in \mathcal{A} \\ &A_1, A_2, \ldots \in \mathcal{A} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{A} \end{aligned}}$$

Sei nun $\mathcal{B} \subset \mathcal{P}(\Omega)$. Dann kann man zeigen, dass es genau eine σ-Algebra $\mathcal{A}$ mit den folgenden beiden Eigenschaften gibt:

- $\mathcal{B} \subset \mathcal{A}$.
- Für jede σ-Algebra $\tilde{\mathcal{A}}$ mit $\mathcal{B} \subset \tilde{\mathcal{A}}$ gilt $\mathcal{A} \subset \tilde{\mathcal{A}}$.

$\mathcal{A}$ ist gewissermaßen die kleinste σ-Algebra, die das Mengensystem $\mathcal{B}$ enthält. $\mathcal{A}$ heißt die von $\mathcal{B}$ ERZEUGTE σ-ALGEBRA.

BEISPIEL XII.1.1. Sei $\Omega = \mathbb{R}$. Betrachte folgende Mengensysteme

$$\begin{aligned} \mathcal{B}_1 &= \{(-\infty, a] : a \in \mathbb{R}\}, \\ \mathcal{B}_2 &= \{(-\infty, a) : a \in \mathbb{R}\}, \\ \mathcal{B}_3 &= \{(a, b) : a, b \in \mathbb{R}, a < b\}, \\ \mathcal{B}_4 &= \{[a, b] : a, b \in \mathbb{R}, a \leq b\}. \end{aligned}$$

Alle vier Mengensysteme erzeugen die gleiche σ-Algebra $\mathcal{B}$. Diese heißt die σ-Algebra der BORELMENGEN von $\mathbb{R}$. Alle Intervalle, egal ob offen oder nicht, egal ob beschränkt oder nicht, sind in $\mathcal{B}$ enthalten.

BEISPIEL XII.1.2. Sei $\Omega = \mathbb{R}^n$. Wir definieren die Relation $\leq$ auf $\mathbb{R}^n$ durch

$$x \leq y \quad \Longleftrightarrow \quad x_i \leq y_i \text{ für alle } 1 \leq i \leq n.$$

Mit dieser Relation können wir Intervalle verallgemeinern:

$$[a, b] = \{x \in \mathbb{R}^n : a \leq x \leq b\}.$$

Für $n = 2$ ist z.B. $[a, b]$ das Rechteck, dessen linke untere Ecke die Koordinaten (a_1, a_2) hat, und dessen rechte obere Ecke die Koordinaten (b_1, b_2) hat. Das Mengensystem $\{[a, b] : a, b \in \mathbb{R}^n\}$ aller dieser Intervalle erzeugt eine σ-Algebra $\mathcal{B}$. Diese heißt wieder σ-Algebra der BORELMENGEN von $\mathbb{R}^n$. Zu ihr gehören z.B. alle offenen und alle abgeschlossenen Mengen (vgl. Abschnitt VIII.2.1 (S. 35, Teil II)).

XII.1.3. Wahrscheinlichkeitsmaße. Sei nun Ω eine beliebige Menge und $\mathcal{A}$ irgendeine σ-Algebra über Ω. Eine Abbildung $P : \mathcal{A} \to [0, 1]$ heißt ein WAHRSCHEINLICHKEITSMASS auf Ω, wenn folgende drei Bedingungen erfüllt sind:

$$\boxed{\begin{aligned} &P(\Omega) = 1 \\ &P(A) \geq 0 \quad \text{für alle } A \in \mathcal{A} \\ &A_1, A_2, \ldots \in \mathcal{A} \text{ paarweise disjunkt} \Longrightarrow P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \end{aligned}}$$

Ein WAHRSCHEINLICHKEITSRAUM ist ein Tripel $(\Omega, \mathcal{A}, P)$ mit einer Menge Ω, einer σ-Algebra $\mathcal{A}$ und einem Wahrscheinlichkeitsmaß P. Die Elemente von $\mathcal{A}$ heißen EREIGNISSE.

Im vorigen Kapitel konnten wir auf die Angabe der σ-Algebra $\mathcal{A}$ verzichten, weil diese „per ordre de mufti“ stets die Potenzmenge $\mathcal{P}(\Omega)$ war.

Eine wichtige Konsequenz aus den obigen Eigenschaften eines Wahrscheinlichkeitsmaßes ist:

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum.

- Ist $B_1 \subset B_2 \subset \ldots$ eine wachsende Folge von Ereignissen und B deren Vereinigung, so ist

$$P(B) = \lim_{i \to \infty} P(B_i).$$

- Ist $C_1 \supset C_2 \supset \ldots$ eine fallende Folge von Ereignissen und C deren Durchschnitt, so ist
$$P(C) = \lim_{i\to\infty} P(C_i).$$

XII.1.4. Wahrscheinlichkeitsmaße mit Dichten. Wir betrachten nun den Fall $\Omega = \mathbb{R}$ und $\mathcal{A} = \mathcal{B}$ die Menge der Borelmengen auf $\mathbb{R}$. Sei P ein Wahrscheinlichkeitsmaß auf $\mathbb{R}$. Dann existiert für jedes $x \in \mathbb{R}$ die Zahl
$$F(x) = P((-\infty, x]).$$
Die Funktion $F : \mathbb{R} \to [0, 1]$ ist monoton wachsend mit
$$\lim_{x\to-\infty} F(x) = 0,$$
$$\lim_{x\to+\infty} F(x) = 1.$$
Außerdem existiert für jedes $x \in \mathbb{R}$ der rechtsseitige Grenzwert $F(x+0)$ (vgl. Abschnitt III.3.4 (S. 121, Teil I)) und stimmt mit $F(x)$ überein. Die Funktion F heißt die zu P gehörende VERTEILUNGSFUNKTION.

Ist umgekehrt $F : \mathbb{R} \to [0, 1]$ eine Funktion mit obigen Eigenschaften, so wird durch
$$P([a, b]) = F(b) - F(a)$$
ein Wahrscheinlichkeitsmaß auf $\mathbb{R}$ definiert. Die gegebene Funktion F ist dann die zugehörige Verteilungsfunktion. In diesem Sinne besteht eine eindeutige Beziehung zwischen Wahrscheinlichkeitsmaßen auf $\mathbb{R}$ und Verteilungsfunktionen.

Eine Funktion $f : \mathbb{R} \to [0, \infty)$ mit
$$\int_{-\infty}^{\infty} f(x)dx = 1$$
heißt DICHTE. Sie definiert durch
$$F(x) = \int_{-\infty}^{x} f(t)dt$$
eine Verteilungsfunktion und damit ein Wahrscheinlichkeitsmaß P auf $\mathbb{R}$. Die Funktion f heißt dann die zu P gehörende Dichte.

Nicht jedes Wahrscheinlichkeitsmaß auf $\mathbb{R}$ besitzt eine Dichte. Aber Wahrscheinlichkeitsmaße mit Dichte sind für die Anwendungen besonders wichtig.

XII.1.5. Gleichverteilung auf einem Intervall. Seien $a, b \in \mathbb{R}$ mit $a < b$. Die Funktion $f : \mathbb{R} \to [0, \infty)$ mit
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{falls } x \in [a, b] \\ 0 & \text{sonst} \end{cases}$$

ist eine Dichte. Sie heißt Dichte der GLEICHVERTEILUNG. Das zugehörige Wahrscheinlichkeitsmaß ordnet jedem Intervall I seinen relativen Anteil an $[a, b]$ zu.

XII.1.6. Exponentialverteilung. Für jedes $\lambda > 0$ wird durch

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

eine Dichte definiert. Sie heißt Dichte der EXPONENTIALVERTEILUNG.

XII.1.7. Normalverteilung. Die Funktion

$$\boxed{\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

heißt Dichte der NORMALVERTEILUNG mit Mittelwert μ und Varianz σ^2. Ist

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$$

die Funktion aus Abschnitt XI.5.1 (S. 36), so ist

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma}).$$

Mit

$$\boxed{\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt}$$

und der Substitutionsregel (vgl. Abschnitt V.2.3 (S. 174, Teil I)) folgt daher

$$\boxed{\int_a^b \varphi_{\mu,\sigma^2}(x) dx = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma}).}$$

XII.1.8. Produktdichten. Werden n Teilexperimente durch Dichten $f_1, \ldots, f_n$ beschrieben, verwendet man

$$f(x_1, \ldots, x_n) = f_1(x_1) \cdot \ldots \cdot f_n(x_n)$$

als Dichte für die Verteilung auf $\mathbb{R}^n$, die die unabhängige Hintereinanderausführung der Teilexperimente beschreibt.

XII.2. Zufallsvariable und ihre Momente

XII.2.1. Messbare Funktionen. Seien Ω und Ω' zwei beliebige Mengen und $\mathcal{A}$ und $\mathcal{A}'$ zwei beliebige σ-Algebren auf Ω bzw. Ω'. Eine Abbildung $f : \Omega \to \Omega'$ heißt MESSBAR, wenn für jede Menge $A' \in \mathcal{A}'$ das Urbild

$$f^{-1}(A') = \{x \in \Omega : f(x) \in A'\}$$

unter f in $\mathcal{A}$ enthalten ist.

Man beachte, dass der Begriff der Messbarkeit von den σ-Algebren $\mathcal{A}$ und $\mathcal{A}'$ abhängt.

BEISPIEL XII.2.1. Sei Ω eine endliche Menge und $\mathcal{A} = \mathcal{P}(\Omega)$. Dann ist für jede Menge Ω' und jede σ-Algebra $\mathcal{A}'$ auf Ω' jede Abbildung $f : \Omega \to \Omega'$ messbar. Aus diesem Grunde haben wir den Begriff der Messbarkeit in Kapitel XI nicht benötigt.

BEISPIEL XII.2.2. Sei $\Omega = \mathbb{R}^m$, $\Omega' = \mathbb{R}^n$ und $\mathcal{A}$ und $\mathcal{A}'$ die σ-Algebren der Borelmengen. Dann ist jede stetige Abbildung $f : \Omega \to \Omega'$ messbar. Ist speziell $m = 1$, so ist jede stückweise stetige Funktion (vgl. Abschnitt V.1.1 (S. 165, Teil I)) messbar.

Sind $f : \Omega \to \Omega'$ und $g : \Omega' \to \Omega''$ messbar, dann ist auch die Komposition $g \circ f : \Omega \to \Omega''$ messbar. Sind $f_1, f_2 : \Omega \to \mathbb{R}$ messbar und $\alpha_1, \alpha_2 \in \mathbb{R}$, so sind auch $\alpha_1 f_1 + \alpha_2 f_2$, $f_1 \cdot f_2$, $\min\{f_1, f_2\}$ und $\max\{f_1, f_2\}$ messbar.

XII.2.2. Zufallsvariable. Seien $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, Ω' eine beliebige Menge und $\mathcal{A}'$ eine σ-Algebra auf Ω'. Eine messbare Funktion $X : \Omega \to \Omega'$ heißt (Ω'-wertige) ZUFALLSVARIABLE. Jede Zufallsvariable $X : \Omega \to \Omega'$ definiert durch

$$P_X(A') = P(X \in A')$$

für $A' \in \mathcal{A}'$ ein Wahrscheinlichkeitsmaß auf Ω'. Ist speziell $\Omega' = \mathbb{R}$, so lässt sich dieses Wahrscheinlichkeitsmaß durch seine Verteilungsfunktion (vgl. Abschnitt XII.1.4 (S. 51))

$$\boxed{F_X(x) = P_X((-\infty, x]) = P(X \leq x)}$$

beschreiben. F_X heißt die VERTEILUNGSFUNKTION der reellwertigen Zufallsvariablen X. Besitzt F_X eine Dichte f, so heißt diese die DICHTE der Zufallsvariablen.

Seien nun $X : \Omega \to \mathbb{R}$ eine reellwertige Zufallsvariable und $\varphi : \mathbb{R} \to \mathbb{R}$ eine streng monoton wachsende, differenzierbare Funktion. Dann ist $Y = \varphi \circ X$ auch eine reellwertige Zufallsvariable. X habe die Dichte f. Dann folgt aus der Substitutionsregel (vgl. Abschnitt V.2.3 (S. 174, Teil I)) für jedes $y \in \mathbb{R}$

$$F_Y(y) = P(\varphi \circ X \leq y)$$

$$= P(X \leq \varphi^{-1}(y)) \quad \text{wegen der Monotonie}$$
$$= \int_{-\infty}^{\varphi^{-1}(y)} f(x)dx \quad \big| x = \varphi^{-1}(t)\ ,\ dx = \frac{1}{\varphi'(\varphi^{-1}(t))}dt$$
$$= \int_{-\infty}^{y} f(\varphi^{-1}(t))\frac{1}{\varphi'(\varphi^{-1}(t))}dt.$$

Also hat $\varphi \circ X$ die Dichte

$$g(y) = \frac{f(\varphi^{-1}(y))}{\varphi'(\varphi^{-1}(y))}.$$

Mit den gleichen Argumenten folgt, dass für eine streng monoton fallende, differenzierbare Funktion $\psi : \mathbb{R} \to \mathbb{R}$ die Zufallsvariable $\psi \circ X$ die Dichte

$$g(y) = \frac{f(\psi^{-1}(y))}{|\psi'(\psi^{-1}(y))|}$$

hat. Speziell folgt:

- Hat X die Dichte f, so hat $X + a$ die Dichte
$$g(y) = f(y - a).$$
- Ist $c \neq 0$, hat cX die Dichte
$$g(y) = \frac{1}{|c|}f(\frac{y}{c})$$

XII.2.3. Unabhängigkeit. Seien Ω_i, $i \in I$, beliebige Mengen und $\mathcal{A}_i$ σ-Algebren auf den Ω_i, sowie $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X_i : \Omega \to \Omega_i$ Zufallsvariable. Dann heißen die X_i UNABHÄNGIG, wenn für alle $A_i \in \mathcal{A}_i$ die Ereignisse $\{X_i \in A_i\}$ unabhängig sind.

Diese Definition ist eine direkte Verallgemeinerung derjenigen aus Abschnitt XI.3.3 (S. 22). Für die praktische Rechnung ist sie aber wenig geeignet. Für den Spezialfall reellwertiger Zufallsvariabler mit Dichten hilft folgendes Ergebnis:

> Die reellwertigen Zufallsvariablen $X_1, \ldots, X_n$ mit Dichten $f_1, \ldots, f_n$ sind genau dann unabhängig, wenn $X = (X_1, \ldots, X_n)$ die Produktdichte
> $$f(x_1, \ldots, x_n) = f_1(x_1) \cdot \ldots \cdot f_n(x_n)$$
> besitzt.

Seien X_1 und X_2 zwei reellwertige Zufallsvariable mit Dichten f_1 und f_2. Dann folgt aus Obigem, dass $X_1 + X_2$ die Dichte

$$(f_1 * f_2)(u) = \int_{-\infty}^{\infty} f_1(u - v)f_2(v)dv$$

besitzt. Der Ausdruck $f_1 * f_2$ heißt FALTUNG der Dichten f_1 und f_2.

Beispiel XII.2.3. Betrachte zwei unabhängige, reellwertige Zufallsvariable X_1 und X_2, die $N(\mu_i, \sigma_i^2)$-verteilt sind, $i = 1, 2$. Wir behaupten, dass ihre Summe $X_1 + X_2$ dann $N(\mu, \sigma^2)$-verteilt ist mit

$$\mu = \mu_1 + \mu_2$$

und

$$\sigma^2 = \sigma_1^2 + \sigma_2^2.$$

Wir müssen also für alle $u \in \mathbb{R}$ die Gleichung

$$\begin{aligned}&\frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}} \exp\left(-\frac{(u-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right)\\ &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} \exp\left(-\frac{(u-v-\mu_1)^2}{2\sigma_1^2}\right) \exp\left(-\frac{(v-\mu_2)^2}{2\sigma_2^2}\right) dv\end{aligned}$$

beweisen. Diese ist äquivalent zu

$$\begin{aligned}&\frac{1}{\sqrt{2\pi}} \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left[\frac{(u-v-\mu_1)^2}{\sigma_1^2} + \frac{(v-\mu_2)^2}{\sigma_2^2}\right.\right.\\ &\qquad\qquad\left.\left. - \frac{(u-\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}\right]\right) dv\\ &= 1.\end{aligned}$$

Wir betrachten den Term in eckigen Klammern und erhalten mit der Abkürzung

$$\begin{aligned}x &= u - \mu_1 - \mu_2\\ y &= v - \mu_2\end{aligned}$$

die Beziehung

$$\begin{aligned}&\frac{(u-v-\mu_1)^2}{\sigma_1^2} + \frac{(v-\mu_2)^2}{\sigma_2^2} - \frac{(u-\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}\\ &= \frac{(x-y)^2}{\sigma_1^2} + \frac{y^2}{\sigma_2^2} - \frac{x^2}{\sigma_1^2+\sigma_2^2}\\ &= \frac{x^2}{\sigma_1^2} - \frac{2xy}{\sigma_1^2} + \frac{y^2}{\sigma_1^2} + \frac{y^2}{\sigma_2^2} - \frac{x^2}{\sigma_1^2+\sigma_2^2}\\ &= \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}x^2 - \frac{1}{\sigma_1^2}2xy + \frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}y^2\\ &= \left(\frac{\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}x - \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}y\right)^2.\end{aligned}$$

Daher liefert die Variablentransformation

$$t = \frac{\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}(u-\mu_1-\mu_2) - \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}(v-\mu_2)$$

$$dt = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1 \sigma_2} dv$$

für alle $a < b$ und alle $u \in \mathbb{R}$

$$\frac{1}{\sqrt{2\pi}} \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1 \sigma_2} \int_a^b \exp\left(-\frac{1}{2} \left[\frac{(u - v - \mu_1)^2}{\sigma_1^2} + \frac{(v - \mu_2)^2}{\sigma_2^2} - \frac{(u - \mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \right] \right) dv$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\frac{\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}(u-\mu_1-\mu_2) - \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}(a-\mu_2)}^{\frac{\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}(u-\mu_1-\mu_2) - \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}(b-\mu_2)} \exp(-\frac{1}{2}t^2) dt$$

$$\underset{\substack{a \to -\infty \\ b \to \infty}}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{1}{2}t^2) dt$$

$$= 1.$$

Dies beweist die Behauptung.

XII.2.4. Erwartungswert. Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine reellwertige Zufallsvariable.

Falls X nur endlich viele Werte $x_1, \ldots, x_n$ annimmt, können wir X wie in Abschnitt XI.3.4 (S. 22) einen Erwartungswert durch

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i)$$

zuordnen.

Wir wollen diese Einschränkung gerne fallen lassen. Dazu nehmen wir zunächst an, dass X beschränkt ist, d.h. es gibt ein $R > 0$ mit $|X(\omega)| \leq R$ für alle $\omega \in \Omega$. Für $k \in \mathbb{Z}$ und $n \in \mathbb{N}^*$ definieren wir dann

$$A_{k,n} = \left\{ \omega \in \Omega : \frac{k}{n} \leq X(\omega) < \frac{k+1}{n} \right\},$$

$$\chi_{A_{k,n}}(\omega) = \begin{cases} 1 & \text{falls } \omega \in A_{k,n}, \\ 0 & \text{sonst,} \end{cases}$$

$$X_n = \sum_{k \in \mathbb{Z}} \frac{k}{n} \chi_{A_{k,n}}.$$

Da X beschränkt ist, nimmt X_n nur endlich viele Werte an und es ist

$$E(X_n) = \sum_{k \in \mathbb{Z}} \frac{k}{n} P(A_{k,n}).$$

Man kann zeigen, dass die Folge $(E(X_n))_{n \in \mathbb{N}}$ konvergiert. Den Grenzwert nennen wir Erwartungswert von X und bezeichnen ihn mit

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

Falls X die stückweise stetige Dichte f besitzt, erhalten wir für jedes n

$$\begin{aligned} E(X_n) &= \sum_{k\in\mathbb{Z}} \frac{k}{n} P(A_{k,n}) \\ &= \sum_{k\in\mathbb{Z}} \frac{k}{n} \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x)dx. \end{aligned}$$

Man kann zeigen, dass dieser Ausdruck für $n \to \infty$ gegen

$$\int_{-\infty}^{\infty} xf(x)dx$$

konvergiert. Daher erhalten wir in diesem Fall

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

Wenn X nicht beschränkt ist, benutzen wir diese Identität als Definition des Erwartungswertes:

> Die reellwertige Zufallsvariable X auf Ω besitze die stückweise stetige Dichte f. Dann ist der Erwartungswert von X definiert durch
>
> $$E(X) = \int_{-\infty}^{\infty} xf(x)dx,$$
>
> sofern das uneigentliche Integral existiert.

Beispiel XII.2.4. X sei $N(\mu, \sigma^2)$-verteilt. Für $a < 0 < b$ folgt

$$\begin{aligned} &\int_a^b x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad \Big| t = \frac{x-\mu}{\sigma} \text{ , } dt = \frac{1}{\sigma} dx \\ &= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} (\sigma t + \mu) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \underbrace{te^{-\frac{1}{2}t^2}}_{=-\frac{d}{dt}e^{-\frac{1}{2}t^2}} dt + \frac{\mu}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-\frac{1}{2}t^2} dt \\ &= \frac{\sigma}{\sqrt{2\pi}} \left[e^{-\frac{1}{2}(\frac{a-\mu}{\sigma})^2} - e^{-\frac{1}{2}(\frac{b-\mu}{\sigma})^2} \right] + \frac{\mu}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-\frac{1}{2}t^2} dt \\ &\underset{\substack{a\to-\infty \\ b\to\infty}}{\longrightarrow} \mu. \end{aligned}$$

Also hat X den Erwartungswert μ.

Beispiel XII.2.5. X sei exponentiell verteilt mit Parameter λ. Für $a > 0$ gilt

$$\begin{aligned}\int_0^a x \underbrace{\lambda e^{-\lambda x}}_{=-\frac{d}{dx}e^{-\lambda x}} dx &= -xe^{-\lambda x}\Big|_{x=0}^{x=a} + \int_0^a e^{-\lambda x} dx \\ &= -ae^{-\lambda a} - \frac{1}{\lambda}e^{-\lambda a} + \frac{1}{\lambda} \\ &\xrightarrow[a\to\infty]{} \frac{1}{\lambda}.\end{aligned}$$

Da $X(x) = 0$ ist für $x < 0$, folgt

$$E(X) = \frac{1}{\lambda}.$$

Offensichtlich gilt für den Erwartungswert

$$\boxed{E(aX + bY) = aE(X) + bE(Y),}$$

wobei X, Y reellwertige Zufallsvariablen und a, b reelle Zahlen sind.

XII.2.5. Varianz. Sei X eine reellwertige Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit einem endlichem Erwartungswert $E(X)$. Die Varianz von X ist definiert durch

$$\boxed{\mathrm{Var}(X) = E((X - E(X))^2),}$$

sofern dieser Erwartungswert existiert. Die Standardabweichung ist dann definiert durch

$$\sigma = \sqrt{\mathrm{Var}(X)}.$$

Wie in Abschnitt XI.3.5 (S. 24) gilt

$$\boxed{\mathrm{Var}(X) = E(X^2) - E(X)^2.}$$

Hat daher X die stückweise stetige Dichte f, ergibt sich

$$\boxed{\mathrm{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - \left[\int_{-\infty}^{\infty} xf(x)dx\right]^2.}$$

Beispiel XII.2.6. Sei X $N(\mu, \sigma^2)$-verteilt. Für $a < 0 < b$ erhalten wir wie in Beispiel XII.2.4 (S. 57)

$$\int_a^b x^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad \Big| t = \frac{x-\mu}{\sigma}\ ,\ dt = \frac{1}{\sigma}dx$$

$$
\begin{aligned}
&= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} (\sigma t + \mu)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} t^2 e^{-\frac{1}{2}t^2} dt \\
&\quad + \frac{2\sigma\mu}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} t e^{-\frac{1}{2}t^2} dt \\
&\quad + \frac{\mu^2}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-\frac{1}{2}t^2} dt \\
&= \frac{\sigma^2}{\sqrt{2\pi}} t(-e^{-\frac{1}{2}t^2})\Big|_{t=\frac{a-\mu}{\sigma}}^{t=\frac{b-\mu}{\sigma}} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-\frac{1}{2}t^2} dt \\
&\quad + \frac{2\sigma\mu}{\sqrt{2\pi}} (-e^{-\frac{1}{2}t^2})\Big|_{t=\frac{a-\mu}{\sigma}}^{t=\frac{b-\mu}{\sigma}} \\
&\quad + \frac{\mu^2}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-\frac{1}{2}t^2} dt \\
&\underset{\substack{a\to-\infty \\ b\to\infty}}{\longrightarrow} \sigma^2 + \mu^2.
\end{aligned}
$$

Da $E(X) = \mu$ ist, folgt

$$
\operatorname{Var}(X) = \sigma^2.
$$

Beispiel XII.2.7. X sei exponentiell verteilt mit Parameter λ. Für $a > 0$ erhalten wir

$$
\begin{aligned}
\int_0^a x^2 \lambda e^{-\lambda x} dx &= x^2(-e^{-\lambda x})\Big|_{x=0}^{x=a} + 2 \int_0^a x e^{-\lambda x} dx \\
&= -a^2 e^{-\lambda a} + \frac{2}{\lambda} x(-e^{-\lambda x})\Big|_{x=0}^{x=a} + \frac{2}{\lambda} \int_0^a e^{-\lambda x} dx \\
&= -a^2 e^{-\lambda a} - \frac{2}{\lambda} a e^{-\lambda a} - \frac{2}{\lambda^2} e^{-\lambda a} + \frac{2}{\lambda^2} \\
&\underset{a\to\infty}{\longrightarrow} \frac{2}{\lambda^2}.
\end{aligned}
$$

Wegen $E(X) = \frac{1}{\lambda}$ ergibt sich

$$
\operatorname{Var}(X) = \frac{1}{\lambda^2}.
$$

Tabelle XII.2.1 fasst die Ergebnisse der vorigen Beispiele zusammen.

Wie in Abschnitt XI.3.5 (S. 24) gilt die Rechenregel

$$
\boxed{\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X),}
$$

Tabelle XII.2.1. Erwartungswert und Varianz der Normal- und Exponentialverteilung

Verteilung	Erwartungswert	Varianz
$N(\mu, \sigma^2)$	μ	σ^2
Exponential λ	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

wobei X eine reellwertige Zufallsvariable und a, b reelle Zahlen sind. Ebenso gilt

$$\boxed{\begin{aligned} &X_1, \ldots, X_n \text{ unabhängig} \\ \Longrightarrow &\operatorname{Var}(X_1 + \ldots + X_n) = \operatorname{Var}(X_1) + \ldots + \operatorname{Var}(X_n). \end{aligned}}$$

XII.3. Schätzverfahren

XII.3.1. Maximum-Likelihood Schätzung. Es werde eine Zufallsvariable X mit Werten in $\mathbb{R}^n$ beobachtet. Die Verteilung von X hänge von einem unbekannten Parameter $\vartheta \in \Theta$ ab. Wir nennen sie P_ϑ. P_ϑ habe für jedes $\vartheta \in \Theta$ eine Dichte $f(\cdot; \vartheta)$. Dann ist für jedes $x \in \mathbb{R}^n$ $P_\vartheta(x) = 0$. Wir können also nicht, wie im diskreten Fall, aus der Betrachtung von $P_\vartheta(x)$ Schätzer ableiten. Statt dessen definieren wir nun die Likelihood-Funktion durch die Dichte

$$L_x(\vartheta) = f(x; \vartheta).$$

Wie in Abschnitt XI.4.3 (S. 31) setzen wir

$$\mathcal{L}_x(\vartheta) = \ln L_x(\vartheta).$$

Beispiel XII.3.1. Die Zufallsvariablen $X_1, \ldots, X_n$ seien unabhängig und jeweils $N(\mu, \sigma^2)$-verteilt. Dann ist $\vartheta = (\mu, \sigma)$. Die Dichte von X_i ist

$$f_i(x; \vartheta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

Daher hat $X = (X_1, \ldots, X_n)$ an der Stelle $x = (x_1, \ldots, x_n)$ die Produktdichte

$$\begin{aligned} f(x; \vartheta) &= \prod_{i=1}^n f_i(x; \vartheta) \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right). \end{aligned}$$

Der Maximum-Likelihood Schätzer $\widehat{\vartheta} = \widehat{\vartheta}(x)$ ist wieder der Parameterwert, der $L_x(\vartheta)$ bzw. $\mathcal{L}_x(\vartheta)$ maximiert. Aus obiger Darstellung der Dichte erhalten wir

$$\mathcal{L}_x(\vartheta) = -n\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

und somit

$$\begin{aligned}\frac{\partial}{\partial\mu}\mathcal{L}_x(\vartheta) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)\\ &= -\frac{n\mu}{\sigma^2}+\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i,\\ \frac{\partial}{\partial\sigma}\mathcal{L}_x(\vartheta) &= -\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2.\end{aligned}$$

Wir müssen nun drei Fälle unterscheiden:
1. μ IST UNBEKANNT, ABER $\sigma^2 = \sigma_0^2$ IST BEKANNT: Dann ist $\Theta = \{(\mu,\sigma_0) : \mu \in \mathbb{R}\}$, und wir müssen eine Nullstelle $\widehat{\mu}$ von $\frac{\partial}{\mu}\mathcal{L}_x(\mu,\sigma_0)$ finden. Aus der Formel für $\frac{\partial}{\partial\mu}\mathcal{L}_x$ folgt sofort

$$\boxed{\widehat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n}x_i.}$$

Indem wir die zweite Ableitung bilden, sehen wir, dass $\widehat{\mu}$ wirklich ein Maximum ist.
2. $\mu = \mu_0$ IST BEKANNT, ABER σ^2 IST UNBEKANNT: Jetzt ist $\Theta = \{(\mu_0,\sigma) : \sigma > 0\}$, und wir müssen eine Nullstelle $\widehat{\sigma}$ von $\frac{\partial}{\partial\sigma}\mathcal{L}_x(\mu_0,\sigma)$ finden. Aus der Formel für $\frac{\partial}{\partial\sigma}\mathcal{L}_x$ ergibt sich

$$\boxed{\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2.}$$

Wieder folgt durch Bilden der zweiten Ableitung, dass $\widehat{\sigma}^2$ tatsächlich ein Maximum ist.
3. μ UND σ^2 SIND BEIDE UNBEKANNT: Die Gleichungen

$$\begin{aligned}\frac{\partial}{\partial\mu}\mathcal{L}_x(\mu,\sigma) &= -\frac{1}{\sigma^2}\left[n\mu-\sum_{i=1}^{n}x_i\right]\\ &= 0,\\ \frac{\partial}{\partial\sigma}\mathcal{L}_x(\mu,\sigma) &= \frac{1}{\sigma^3}\left[-n\sigma^2+\sum_{i=1}^{n}(x_i-\mu)^2\right]\end{aligned}$$

$$= 0$$

für einen kritischen Punkt liefern die Lösungen

$$\boxed{\begin{aligned}\widehat{\mu} &= \overline{x},\\ \widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.\end{aligned}}$$

Durch Bilden der Hesse-Matrix sehen wir, dass diese Werte wirklich das Maximum liefern.
Der Schätzer $\widehat{\mu}$ und der Schätzer $\widehat{\sigma}^2$ bei bekanntem Erwartungswert μ_0 sind erwartungstreu. Der Schätzer $\widehat{\sigma}^2$ im Fall 3 dagegen ist nicht erwartungstreu. Analog zu Abschnitt XI.4.4 (S. 32) liefert in diesem Fall

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

einen erwartungstreuen Schätzer für die Varianz.

XII.3.2. Die Methode der kleinsten Quadrate. Oft stellt sich das Problem, eine Gerade, eine Parabel oder eine andere „einfache“ Funktion einer gegebenen Menge von Messwerten anzupassen. Zum Beispiel kann eine Größe y in Abhängigkeit von einer Größe x gemessen worden sein und es liegen nun n Messergebnisse (x_1, y_1), …, (x_n, y_n) vor. Wenn die Messergebnisse relativ gut auf einer Geraden liegen, können wir einen linearen Zusammenhang der beobachteten Größen vermuten, der durch Messfehler z_i gestört ist. Dann wäre

$$y_i = \alpha + \beta x_i + z_i \quad , i = 1, \ldots, n,$$

mit zu bestimmenden Parametern α und β.

Allgemeiner nehmen wir an, dass $\vartheta_1, \ldots, \vartheta_p$ unbekannte Parameter sind und dass für bekannte Funktionen φ_i der gemessene Wert bei der i-ten Messung von der Form ist

$$y_i = \varphi_i(\vartheta_1, \ldots, \vartheta_p) + z_i$$

mit einem Messfehler z_i. Im Beispiel der Geraden ist

$$\begin{aligned} p &= 2,\\ \vartheta_1 &= \alpha,\\ \vartheta_2 &= \beta,\\ \varphi_i(\vartheta_1, \vartheta_2) &= \vartheta_1 + \vartheta_2 x_i. \end{aligned}$$

Die Methode der kleinsten Quadrate besteht darin, die Parameter $\vartheta_1, \dots, \vartheta_p$ so zu bestimmen, dass die Größe

$$Q = \sum_{i=1}^{n} [y_i - \varphi_i(\vartheta_1, \dots, \vartheta_p)]^2$$

minimal wird.

Dieser Ansatz kann ad hoc ohne Statistik formuliert werden und wird häufig auch so angewandt. Wir wollen nun zeigen, dass dieser Ansatz statistisch fundiert ist. Dazu nehmen wir an, dass die Messfehler z_i Realisierungen von unabhängigen $N(0, \sigma^2)$-verteilten Zufallsvariablen $Z_1, \dots, Z_n$ sind. Dann sind die y_i Realisierungen von unabhängigen Zufallsvariablen $Y_1, \dots, Y_n$, die $N(\varphi_i(\vartheta_1, \dots, \vartheta_p), \sigma^2)$-verteilt sind. Daher hat $Y = (Y_1, \dots, Y_n)$ die Dichte

$$\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - \varphi_i(\vartheta_1, \dots, \vartheta_p)]^2\right).$$

Diese Dichte ist genau dann maximal, wenn obige Größe Q minimal ist. Die Methode der kleinsten Quadrate ist also gerade der Maximum-Likelihood Schätzer für die Parameter $\vartheta_1, \dots, \vartheta_p$.

Beispiel XII.3.2 (Regressionsgerade). Wir wollen eine Gerade $y = \alpha + \beta x$ mit der Methode der kleinsten Quadrate an Messdaten (x_1, y_1), …, (x_n, y_n) anpassen. Dann ist

$$Q = Q(\alpha, \beta) = \sum_{i=1}^{n} (y_i - \beta x_i - \alpha)^2.$$

Die Gleichungen für einen kritischen Punkt $\widehat{\alpha}$, $\widehat{\beta}$ lauten

$$\begin{aligned} 0 &= \frac{\partial}{\partial \alpha} Q(\widehat{\alpha}, \widehat{\beta}) = -2\sum_{i=1}^{n}(y_i - \widehat{\beta} x_i - \widehat{\alpha}) \\ 0 &= \frac{\partial}{\partial \beta} Q(\widehat{\alpha}, \widehat{\beta}) = -2\sum_{i=1}^{n}(y_i - \widehat{\beta} x_i - \widehat{\alpha}) x_i. \end{aligned}$$

Zur Abkürzung setzen wir

$$\boxed{\begin{aligned} \overline{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \quad , \qquad & \overline{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i, \\ \overline{xx} &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 \quad , \qquad & \overline{xy} &= \frac{1}{n}\sum_{i=1}^{n} x_i y_i. \end{aligned}}$$

Nach Multiplikation mit $-\frac{1}{2n}$ nehmen dann obige Bestimmungsgleichungen die Form an

$$0 = \overline{y} - \widehat{\beta}\overline{x} - \widehat{\alpha}$$
$$0 = \overline{xy} - \widehat{\beta}\overline{xx} - \widehat{\alpha}\overline{x}.$$

Dies liefert die Lösung

$$\boxed{\begin{aligned}\widehat{\beta} &= \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{xx} - \overline{x}\,\overline{x}},\\ \widehat{\alpha} &= \overline{y} - \widehat{\beta}\overline{x}.\end{aligned}}$$

Die Gerade $y = \widehat{\alpha} + \widehat{\beta}x$ heißt REGRESSIONSGERADE.
Mit den Abkürzungen

$$\boxed{\begin{aligned}s_{xy} &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})\\ s_{xx} &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2\end{aligned}}$$

erhält man wegen

$$s_{xy} = \frac{n}{n-1}\overline{xy} - \frac{n}{n-1}\overline{x}\,\overline{y},$$
$$s_{xx} = \frac{n}{n-1}\overline{xx} - \frac{n}{n-1}\overline{x}\,\overline{x}$$

die alternative Darstellung

$$\boxed{\widehat{\beta} = \frac{s_{xy}}{s_{xx}}.}$$

XII.3.3. Median. Wenn in einem schweizer Bergdorf fünfzig Einheimische und fünf zugezogene Millionäre leben, ist es für die Einheimischen wenig befriedigend, wenn man ihnen erklärt, das durchschnittliche Einkommen in diesem Dorf sei hoch. Dieses Beispiel zeigt, dass der Erwartungswert manchmal kein guter Maßstab für das „Zentrum“ einer Verteilung ist. Einige wenige „Ausreißer“ können ihn stark verfälschen. Man betrachtet daher auch andere Maßzahlen. Die bekannteste davon ist der Median.

Ist Z eine reellwertige Zufallsvariable, so heißt *jede* Zahl μ_m mit

$$P(Z \geq \mu_m) \geq \frac{1}{2}$$

und

$$P(Z \le \mu_m) \ge \frac{1}{2}$$

ein MEDIAN von Z.

Man beachte, dass der Median μ_m nicht notwendig eindeutig bestimmt ist. Eine Mehrdeutigkeit tritt genau dann auf, wenn es ein Intervall $[a, b]$ gibt mit $a < b$ und $P(Z \ge b) = \frac{1}{2}$ und $P(Z \le a) = \frac{1}{2}$. Dann ist jede Zahl in dem Intervall $[a, b]$ ein Median. Gibt es dagegen eine Zahl c mit $P(Z \ge c + t) = P(Z \le c - t)$ für alle $t > 0$, so ist c der eindeutige Median von Z. In diesem Fall heißt Z symmetrisch, und es ist $c = E(Z)$.

Für n Messwerte $x_1, \ldots, x_n$ können wir den EMPIRISCHEN MEDIAN $\widehat{\mu}_m$ wie folgt ermitteln:
Bezeichne mit $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$ die der Größe nach umgeordneten x_i. Dann ist

$$\boxed{\widehat{\mu}_m = \begin{cases} x_{(\frac{n+1}{2})} & \text{falls } n \text{ ungerade,} \\ \frac{1}{2}[x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}] & \text{falls } n \text{ gerade.} \end{cases}}$$

BEISPIEL XII.3.3. Für $x_1 = 3.1$, $x_2 = 4.5$, $x_3 = 4.5$, $x_4 = 2.8$, $x_5 = 2.6$, $x_6 = 9.8$ erhalten wir $x_{(1)} = 2.6$, $x_{(2)} = 2.8$, $x_{(3)} = 3.1$, $x_{(4)} = 4.5$, $x_{(5)} = 4.5$, $x_{(6)} = 9.8$ und $\widehat{\mu}_m = \frac{1}{2}[x_{(3)} + x_{(4)}] = \frac{1}{2}[3.1 + 4.5] = 3.8$. Der Mittelwert ist $\overline{x} = 4.55$.

XII.4. Tests

XII.4.1. Vorbemerkungen. Wir beobachten eine, im allgemeinen vektorwertige, Zufallsvariable X, deren Verteilung einer Familie $\{P_\vartheta : \vartheta \in \Theta\}$ angehört. Θ ist die disjunkte Vereinigung der Mengen H und K, der Hypothese und der Alternative. Aufgrund des beobachteten Wertes x von X soll entschieden werden, ob der Parameter ϑ der zu X gehörenden Verteilung in der Menge H liegt oder nicht.

Hierzu bilden wir den LIKELIHOOD-QUOTIENTEN

$$q(x) = \frac{\sup\{L_x(\vartheta) : \theta \in K\}}{\sup\{L_x(\vartheta) : \theta \in H\}}.$$

Offensichtlich ist $0 \le q(x)$ für alle x. $q(x) \approx 0$ spricht für die Hypothese, $q(x) \gg 0$ spricht für die Alternative.

Ein TEST ist nun eine messbare Abbildung φ des Wertebereiches $\mathcal{X}$ von X in $[0, 1]$. Wird x beobachtet, besagt $\varphi(x) = 1$, dass die Hypothese verworfen werden soll, und $\varphi(x) = 0$, dass sie angenommen werden soll. Im Fall $0 < \varphi(x) < 1$ soll ein zusätzliches Zufallsexperiment mit Wahrscheinlichkeit $\varphi(x)$ zur Verwerfung führen.

Ein Test φ heißt LIKELIHOOD-QUOTIENTEN TEST, wenn es ein $c > 0$ gibt mit den Eigenschaften

- $q(x) > c \Longrightarrow \varphi(x) = 1$ und
- $q(x) < c \Longrightarrow \varphi(x) = 0$.

XII.4.2. Der t-Test. Wir beobachten Zufallsvariable $X_1, \ldots, X_n$, die unabhängig und $N(\mu, \sigma^2)$-verteilt seien mit unbekanntem (μ, σ^2). Für ein gegebenes μ_0 sei zu testen, ob $\mu = \mu_0$ ist oder nicht. Dann ist

$$\begin{aligned}
\vartheta &= (\mu, \sigma) \\
\Theta &= \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\} \\
H &= \{(\mu_0, \sigma) : \sigma > 0\} \\
K &= \{(\mu, \sigma) : \mu \neq \mu_0, \sigma > 0\}.
\end{aligned}$$

Die Dichte $f(x; \vartheta)$ von $X = (X_1, \ldots, X_n)$ ist gemäß Abschnitt XII.3.1 (S. 60)

$$f(x; \vartheta) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right).$$

Aus Stetigkeitsgründen ist

$$\sup\{f(x; \vartheta) : \vartheta \in K\} = \sup\{f(x; \vartheta) : \vartheta \in \Theta\}.$$

Damit ergibt sich aus Abschnitt XII.3.1 (S. 60), dass das Supremum an der Stelle

$$\begin{aligned}
\widehat{\mu} &= \overline{x} \\
&= \frac{1}{n} \sum_{i=1}^n x_i, \\
\widehat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \overline{x})^2
\end{aligned}$$

angenommen wird. Ebenso folgt, dass das Supremum $\sup\{f(x; \vartheta) : \vartheta \in H\}$ an der Stelle

$$\widetilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2$$

angenommen wird. Daher ist der Likelihood-Quotient

$$\begin{aligned}
q(x) &= \frac{f(x; \widehat{\mu}, \widehat{\sigma})}{f(x; \mu_0, \widetilde{\sigma})} \\
&= \left(\frac{\widetilde{\sigma}}{\widehat{\sigma}} \right)^n \exp\Bigl(-\frac{1}{2\widehat{\sigma}^2} \underbrace{\sum_{i=1}^n (x_i - \widehat{\mu})^2}_{=n\widehat{\sigma}^2} + \frac{1}{2\widetilde{\sigma}^2} \underbrace{\sum_{i=1}^n (x_i - \mu_0)^2}_{=n\widetilde{\sigma}^2} \Bigr) \\
&= \left(\frac{\widetilde{\sigma}}{\widehat{\sigma}} \right)^n.
\end{aligned}$$

Ist φ irgendein Likelihood-Quotienten Test, gibt es daher ein $c > 0$ mit

$$\begin{aligned}\varphi(x) = 1 \quad \text{auf } \left\{\left(\frac{\widetilde{\sigma}}{\widehat{\sigma}}\right)^n > c\right\}\\ \varphi(x) = 0 \quad \text{auf } \left\{\left(\frac{\widetilde{\sigma}}{\widehat{\sigma}}\right)^n < c\right\}.\end{aligned}$$

Mit

$$c' = c^{2/n}$$

ist dies äquivalent zu

$$\begin{aligned}\varphi(x) = 1 \quad \text{auf } \left\{\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2} > c'\right\}\\ \varphi(x) = 0 \quad \text{auf } \left\{\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2} < c'\right\}.\end{aligned}$$

Es ist

$$\begin{aligned}\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2} &= \frac{\sum\limits_{i=1}^n (x_i - \mu_0)^2}{\sum\limits_{i=1}^n (x_i - \overline{x})^2}\\ &= \frac{\sum\limits_{i=1}^n (x_i - \overline{x} + \overline{x} - \mu_0)^2}{\sum\limits_{i=1}^n (x_i - \overline{x})^2}\\ &= \frac{\sum\limits_{i=1}^n (x_i - \overline{x})^2 + n(\overline{x} - \mu_0)^2}{\sum\limits_{i=1}^n (x_i - \overline{x})^2}\\ &= 1 + \frac{(\overline{x} - \mu_0)^2}{\widehat{\sigma}^2}.\end{aligned}$$

Wir definieren daher

$$T(x) = \frac{\sqrt{n}(\overline{x} - \mu_0)}{s(x)}$$

mit

$$s(x)^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2.$$

Dann folgt

$$\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2} = 1 + \frac{1}{n-1}|T(x)|^2.$$

Also gibt es für jeden Likelihood-Quotienten Test φ ein $t > 0$ mit

$$\varphi(x) = \begin{cases} 1 & \text{falls } |T(x)| > t, \\ 0 & \text{falls } |T(x)| < t. \end{cases}$$

Unser Testproblem ist damit auf die Bestimmung der Verteilung von $T(x)$ zurückgeführt. Diese Verteilung heißt t-Verteilung mit $n-1$ Freiheitsgraden oder kurz t_{n-1}-Verteilung. Die zugehörige Dichte ist gegeben durch

$$\boxed{h_{n-1}(x) = \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})\sqrt{n-1}} \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}.}$$

Dabei ist Γ die Eulersche Gammafunktion aus Abschnitt V.4.3 (S. 192, Teil I). Für $n \to \infty$ konvergiert $h_{n-1}(x)$ gegen die Dichte $\frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$ der Standard-Normalverteilung.

Damit lautet der t-Test zum Niveau α auf die Hypothese einer $N(\mu_0, \sigma^2)$-verteilten Stichprobe:

> Berechne die Größen
>
> $$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
>
> $$s(x) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
>
> $$T(x) = \frac{\sqrt{n}(\overline{x} - \mu_0)}{s(x)}.$$
>
> Bestimme aus einer Tabelle der t_{n-1}-Verteilung das $(1-\frac{\alpha}{2})$-Quantil $t_{n-1,1-\frac{\alpha}{2}}$ aus der Bedingung
>
> $$\int_{-\infty}^{t_{n-1,1-\frac{\alpha}{2}}} h_{n-1}(x)dx = 1 - \frac{\alpha}{2}.$$
>
> Falls $|T(x)| \le t_{n-1,1-\frac{\alpha}{2}}$ ist, wird die Hypothese angenommen, sonst wird sie verworfen.

Beispiel XII.4.1. An 15 Glasfaserplatten einer bestimmten Stärke wird die Wärmeleitfähigkeit gemessen. Es ergibt sich der Mittelwert $\overline{x} = 17.1$ und der Wert $s^2 = 0.36$. Zu testen ist die Hypothese $\mu = 17$ zu dem Niveau $\alpha = 0.1$. Aus einer Tabelle der t_{14}-Verteilung lesen wir das 0.95-Quantil $t_{14,0.95} = 1.76$ ab. Aus den gegebenen Daten ergibt sich

$$T(x) = \frac{\sqrt{14}(17.1 - 17)}{\sqrt{0.36}} \approx 0.624.$$

Also kann die Hypothese akzeptiert werden.

XII.4.3. Der χ^2-Test. Wir betrachten folgendes Testproblem: Es werden n unabhängige, gleichartige Teilexperimente ausgeführt. Jedes hat $r \geq 2$ mögliche Ausgänge, und der i-te Ausgang hat die Wahrscheinlichkeit p_i. Der Parameter $\vartheta = (p_1, \ldots, p_r)$ ist unbekannt. Für einen gegebenen Wahrscheinlichkeitsvektor $\pi = (\pi_1, \ldots, \pi_r)$ ist zu testen, ob $\vartheta = \pi$ ist.

Die Wahrscheinlichkeit, die Häufigkeiten $x_1, \ldots, x_r$ zu beobachten, ist nach der Multinomialverteilung

$$P_\vartheta(x) = \binom{n}{x_1, \ldots, x_r} p_1^{x_1} \cdot \ldots \cdot p_r^{x_r}.$$

Der Vektor $x = (x_1, \ldots, x_r)$ muss dabei natürlich der Bedingung $x_1 + \ldots + x_r = n$ genügen. Die Likelihood-Funktion ist $L_x(\vartheta) = P_\vartheta(x)$. Bei der Ermittlung des Maximums muss die Nebenbedingung $p_1 + \ldots + p_r = 1$ berücksichtigt werden. Dies führt auf das Gleichungssystem

$$\frac{\partial}{\partial p_1} \ln L_x(\vartheta) = \frac{\partial}{\partial p_2} \ln L_x(\vartheta) = \ldots = \frac{\partial}{\partial p_r} \ln L_x(\vartheta).$$

Als Lösung ergibt sich der Schätzer $\widehat{p}_i(x) = \frac{x_i}{n}$. Damit folgt für den Likelihood-Quotienten

$$q(x) = \frac{P_{\widehat{\vartheta}}(x)}{P_\pi(x)} \approx \left(\frac{\widehat{p}_1}{\pi_1} \cdot \ldots \cdot \frac{\widehat{p}_r}{\pi_r} \right)^{-\frac{1}{2}} \exp\left(\frac{1}{2} \sum_{i=1}^{r} \frac{(x_i - n\pi_i)^2}{n\pi_i} \right).$$

Wir definieren daher

$$\boxed{V^2(x) = \sum_{i=1}^{r} \frac{(x_i - n\pi_i)^2}{n\pi_i}.}$$

Gilt die Hypothese $\vartheta = \pi$, so ist nach dem Gesetz der großen Zahlen $\widehat{p}_i(x) = \frac{x_i}{n}$ mit Wahrscheinlichkeit nahe 1 gleich π_i. Daher ist in diesem Fall

$$q(x) \approx \exp(\frac{1}{2} V^2(x)).$$

Man kann zeigen, dass $V^2(x)$ durch die sog. χ^2-Verteilung mit $r - 1$ Parametern, kurz χ^2_{n-1}-Verteilung approximiert wird. Diese hat die Dichte

$$g_{r-1}(x) = \begin{cases} \frac{1}{2^{\frac{r-1}{2}}\Gamma(\frac{r-1}{2})} x^{\frac{r-1}{2}-1} e^{-\frac{x}{2}} & \text{falls } x > 0, \\ 0 & \text{sonst.} \end{cases}$$

Damit lautet der χ^2-Test zum Niveau α:

Berechne die Größe
$$V^2(x) = \sum_{i=1}^{r} \frac{(x_i - n\pi_i)^2}{n\pi_i}.$$
Bestimme aus einer Tabelle der χ^2_{r-1}-Verteilung das $(1-\alpha)$-Quantil $\chi^2_{r-1,1-\alpha}$ aus der Bedingung
$$\int_0^{\chi^2_{r-1,1-\alpha}} g_{r-1}(x)dx = 1 - \alpha.$$
Falls $V^2(x) < \chi^2_{r-1,1-\alpha}$ ist, wird die Hypothese angenommen, sonst wird sie verworfen.

Beispiel XII.4.2. Man vermutet, dass die Blütenfarben rot, rosa und weiß einer Rosenart mit den Wahrscheinlichkeiten $\frac{1}{4}$, $\frac{1}{2}$ und $\frac{1}{4}$ vererbt werden. Unter einer Auswahl von 320 Nachkommen beobachtet man 102 rote, 156 rosa und 62 weiße Blüten. Die Hypothese soll mit dem χ^2-Test zum Niveau $\alpha = 0.1$ getestet werden. Es ist
$$\pi = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$$
und
$$(\widehat{p}_1, \widehat{p}_2, \widehat{p}_3) = (\frac{102}{320}, \frac{156}{320}, \frac{62}{320}).$$
Wir erhalten den Wert
$$\begin{aligned} V^2(x) &= \frac{(102-80)^2}{80} + \frac{(156-160)^2}{160} + \frac{(62-80)^2}{80} \\ &\approx 10.2. \end{aligned}$$
Aus einer Tabelle der χ^2-Verteilung mit 2 Parametern erhalten wir das 0.9-Quantil
$$\chi^2_{2,0.9} = 4.61.$$
Daher muss die Hypothese verworfen werden.

Beispiel XII.4.3. Es wird behauptet, dass auf einer Pferderennbahn die Startposition einen Einfluss auf die Gewinnwahrscheinlichkeit hat. In 144 Rennen hatten die Sieger die Startposition 1, 2, …, 8 mit den Häufigkeiten 29, 19, 18, 25, 17, 10, 15, 11. Wir wollen die Hypothese, dass alle Startpositionen die gleiche Gewinnwahrscheinlichkeit haben, zu dem Niveau $\alpha = 0.05$ testen. Aus einer Tabelle der χ^2-Verteilung mit 7 Freiheitsgraden lesen wir das 0.95-Quantil

$$\chi^2_{7,0.95} = 20.28$$

ab. Da $n\pi_i = 16$ ist für alle i, erhalten wir mit den beobachteten Werten

$$\begin{aligned} V^2(x) &= \frac{1}{16}\{(29-16)^2 + (19-16)^2 + (18-16)^2 + (25-16)^2 \\ &\qquad + (17-16)^2 + (10-16)^2 + (15-16)^2 + (11-16)^2\} \\ &\approx 20.375. \end{aligned}$$

Also wird die Hypothese verworfen.

Die χ^2-Verteilung wird auch für das Testen der Varianz einer Normalverteilung genutzt. Der χ^2-Test zum Niveau α auf die Hypothese einer $N(\mu, \sigma_0^2)$-verteilten Stichprobe lautet:

Berechne die Größen

$$\begin{aligned} \overline{x} &= \frac{1}{n}\sum_{i=1}^n x_i \\ s(x)^2 &= \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2 \\ T(x) &= \frac{(n-1)s(x)^2}{\sigma_0^2}. \end{aligned}$$

Bestimme aus einer Tabelle der χ^2_{n-1}-Verteilung die $(1-\frac{\alpha}{2})$- und $\frac{\alpha}{2}$- Quantile $\chi^2_{n-1,1-\frac{\alpha}{2}}$ und $\chi^2_{n-1,\frac{\alpha}{2}}$ aus den Bedingungen

$$\begin{aligned} \int_{-\infty}^{\chi^2_{n-1,1-\frac{\alpha}{2}}} g_{n-1}(x)dx &= 1 - \frac{\alpha}{2}, \\ \int_{-\infty}^{\chi^2_{n-1,\frac{\alpha}{2}}} g_{n-1}(x)dx &= \frac{\alpha}{2}. \end{aligned}$$

Falls

$$\chi^2_{n-1,\frac{\alpha}{2}} \le T(x) \le \chi^2_{n-1,1-\frac{\alpha}{2}}$$

ist, wird die Hypothese angenommen, sonst wird sie verworfen.

Bemerkung XII.4.4. Bei dem oben dargestellten Test wird auf $\sigma^2 = \sigma_0^2$ getestet. Falls auf $\sigma^2 < \sigma_0^2$ getestet werden soll, lautet das Testkriterium

$$T(x) \le \chi^2_{n-1,1-\alpha}.$$

Falls auf $\sigma^2 > \sigma_0^2$ getestet werden soll, lautet das Testkriterium

$$T(x) \geq \chi^2_{n-1,\alpha}.$$

Beispiel XII.4.5. Nach Angaben des Herstellers eines bestimmten PKW-Typs ist der Benzinverbrauch im Stadtvekehr annähernd normalverteilt mit Erwartungswert 9.5 `l/100km` und Streuung 2.5 `l/100km`. Zur Überprüfung dieser Angaben testet eine Verbraucherorganisation 25 PKWs und misst einen Durchschnittverbrauch von 9.9 `l/100km` mit einer Streuung von 3.5 `l/100km`. Das Niveau des Tests soll $\alpha = 0.05$ sein.
Für den Erwartungswert ergeben diese Daten

$$\begin{aligned} T(x) &= \frac{\sqrt{25}(9.9 - 9.5)}{3.5} \\ &\approx 0.571 \\ t_{24,0.975} &= 2.064. \end{aligned}$$

Daher ist die Aussage über den erwarteten Durchschnittsverbrauch zu akzeptieren.
Für die Varianz liefern diese Daten

$$\begin{aligned} T(x) &= \frac{24 \cdot 3.5^2}{2.5^2} \\ &\approx 47.04 \\ \chi^2_{24,0.025} &= 12.40 \\ \chi^2_{24,0.975} &= 39.36. \end{aligned}$$

Wegen $T(x) > \chi^2_{24,0.975}$ ist die Aussage über die Streuung der Verbrauchswerte zu verwerfen.

Zusammenfassung

XI Stochastik I: Diskrete Modelle

1. Modelle für Zufallsexperimente

   Ergebnisse oder Elementarereignisse; Ergebnismenge; Ereignisse; Potenzmenge; Wahrscheinlichkeitsverteilung oder -maß; Wahrscheinlichkeit; Wahrscheinlichkeitsraum; Wahrscheinlichkeitsfunktion; Laplace-Experimente; Urnenmodelle: Stichproben in Reihenfolge mit Rücklegen, Stichproben in Reihenfolge ohne Rücklegen, Stichproben ohne Reihenfolge ohne Rücklegen, Stichproben ohne Reihenfolge mit Rücklegen; Anwendungsbeispiele; hypergeometrische Verteilung; Multinomialkoeffizienten; Identitäten für Binomialkoeffizienten

2. Bedingte Wahrscheinlichkeit und Unabhängigkeit

   Bedingte Wahrscheinlichkeiten; Beispiele; Produktformel; Formel der totalen Wahrscheinlichkeit; Formel von Bayes; Anwendungen; Unabhängigkeit; Produktexperimente; Produkte von Wahrscheinlich-
   keitsräumen; Bernoulliexperiment und -verteilung; Binomialverteilung; Multinomialverteilung; geometrische Verteilung; negative Binomialverteilung oder Pascalverteilung

3. Zufallsvariable, Erwartungswert, Varianz

   Zufallsvariable; Beispiele; Stabdiagramme; gemeinsame Verteilungsfunktion; Unabhängigkeit; Erwartungswert; Erwartungswert der Binomialverteilung; Erwartungswert der hypergeometrischen Verteilung; Varianz; Kovarianz; Streuung; Standardabweichung; Korrelationskoeffizient; unkorrelierte Zufallsvariable; Varianz der Binomialverteilung; Varianz der hypergeometrischen Verteilung; schwaches Gesetz der großen Zahl; Anwendungen

4. Grundbegriffe der Schätztheorie

   Motivation; allgemeiner Rahmen; Schätzer; Maximum-Likelihood Schätzer; Erwartungstreue; Mittelwert; mittlerer quadratischer Fehler

5. Approximationen der Binomialverteilung

   Stirlingsche Formel; Dichte der Standard-Normalverteilung; Satz von Moivre-Laplace; Verteilungsfunktion der Standard-Normalverteilung; Anwendungen; Poissonapproximation; Poissonverteilung

6. Tests

   Motivation; Hypothese; Annehmen und Verwerfen einer Hypothese; Verwerfungsbereich oder kritischer Bereich; Fehler erster und zweiter Art; Teststatistik; kritischer Wert; Gütefunktion

XII Stochastik II: Allgemeine Modelle

1. Wahrscheinlichkeitsmaße mit Dichten
   Ergebnismengen; σ-Algebren; Borelmengen; Wahrscheinlichkeitsmaße; Wahrscheinlichkeitsräume; Dichten; Gleichverteilung auf einem Intervall; Exponentialverteilung; Normalverteilung; Produktdichten
2. Zufallsvariable und ihre Momente
   Messbare Funktionen; Zufallsvariable; Verteilungsfunktion; Dichte; Unabhängigkeit; Faltung; Faltung von Normalverteilungen; Erwartungswert; Eigenschaften des Erwartungswertes; Varianz; Eigenschaften der Varianz
3. Schätzverfahren
   Maximum-Likelihood Schätzung; Methode der kleinsten Quadrate; Regressionsgerade; Median
4. Tests
   Likelihood-Quotienten; Likelihood-Quotienten Test; t-Test; χ^2-Test; Anwendungen

Index