Kapitel 6

Schließende Statistik

Die bisher betrachteten Wahrscheinlichkeitsverteilungen folgten aus Modellen, etwa dem des Binomialexperiments oder des Ziehens aus einer Urne; wir haben daher die Verteilung als bekannt angenommen. In der Praxis ist die Zielrichtung meist umgekehrt: Man kennt die Verteilung nicht, hat aber eine Stichprobe und möchte daraus Aussagen über die unbekannte Verteilung gewinnen. Solche Probleme sollen jetzt untersucht werden.

6.1 Mittelwert, empirische Varianz und Punktschätzung

6.1.1 Erwartungswert und Varianz von Summen unabhängiger Zufallsvariablen

In diesem Kapitel wird die Unabhängigkeit sehr wichtig werden. Insbesondere wird stets vorausgesetzt, dass Elemente einer Stichprobe voneinander unabhängig sind. Es wird dann immer wieder benutzt werden, dass Erwartungswert[1] bzw. Varianz einer Summe unabhängiger Zufallsvariablen gleich der Summe der Erwartunswerte bzw. der Varianzen der einzelnen Zufallsvariablen ist. Genauer: Für unabhängige Zufallsvariablen $X_1, \dots, X_n$ gilt

$$\mathbb{E}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathbb{E}(X_i)\,, \quad \mathbb{V}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathbb{V}(X_i)\,.$$

6.1.2 Eigenschaften des Mittelwerts

Gegeben sei eine Stichprobe $(x_1, x_2, \dots, x_n)$. Die Verteilung der zugrundeliegenden Zufallsvariablen X kennen wir nicht, insbesondere sind der Erwartungswert[2] μ und die Varianz σ^2 unbekannt und sollen aus der Stichprobe geschätzt werden.

[1] Für den Erwartungswert gilt die Additivität sogar ganz allgemein. Wir werden sie aber nur im Zusammenhang mit unabhängigen Zufallsvariablen benutzen.

[2] Der Erwartungswert hieß im ersten Teil der Vorlesung durchweg $\underline{m}$, ab jetzt nennen wir ihn μ, wie es in der Statistik üblich ist.

Der uns schon bekannte Stichprobenmittelwert

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

ist ein Näherungswert für μ; in der Sprache der Statistik spricht man von einem *Schätzwert* oder *Schätzer* (der Hut $\widehat{}$ ist zu lesen als "Schätzer für").

Im Gegensatz zu μ hängt $\widehat{\mu}$ von der zufällig gewählten Stichprobe ab. Um diese Abhängigkeit herauszuarbeiten, betrachten wir zwei Möglichkeiten, die Stichprobe $(x_1, x_2, \ldots, x_n)$ zu interpretieren:

i) als n Realisierungen der Zufallsvariablen X (z.B. Augenzahl beim einfachen Würfelexperiment)

ii) als *eine* Realisierung der Zufallsvariablen $(X_1, \ldots, X_n)$, der sogenannten *mathematischen Stichprobe*. Dabei ist X_i das Ergebnis des i-ten (Teil-) Experiments (z.B. die Augenzahl beim i-ten Wurf in einem n-fachen Würfelexperiment). Die X_i sind unabhängig und *identisch verteilt*. Letzteres heißt, sie haben alle dieselbe Wahrscheinlichkeitsverteilung (nämlich die der Größe X aus i). "Unabhängig und identisch verteilt" wird als i.i.d. ("independent and identically distributed") abgekürzt.

Mit der zweiten Interpretation ist $\overline{x}$ eine Realisierung der Zufallsvariablen

$$\overline{X} := \frac{1}{n}\sum_i X_i .$$

Ihr Erwartungswert ist gleich dem Erwartungswert von X, denn

$$\begin{aligned} \mathbb{E}(\overline{X}) &= \mathbb{E}\left(\frac{1}{n}\sum_i X_i\right) = \frac{1}{n}\mathbb{E}\left(\sum_i X_i\right) = \frac{1}{n}\sum_i \mathbb{E}(X_i) \\ &= \frac{1}{n}\sum_i \mu = \frac{1}{n} n\,\mu = \mu\,, \end{aligned} \tag{6.1}$$

wobei wir die Rechenregeln für den Erwartungswert, die Additivität des Erwartungswerts und die identische Verteilung der X_i verwendet haben.

Den Erwartungswert $\mathbb{E}(\overline{X})$ kann man — salopp ausgedruückt — als Mittelwert über unendlich viele Stichprobenmittelwerte verstehen. Im Mittel liefert der Stichprobenmittelwert also den "wahren" Erwartungswert μ der zugrundeliegenden Verteilung, man bezeichnet ihn deshalb als **erwartungstreuen** (oder auch **unverzerrten**) Schätzer des Erwartungswerts. Allgemein heißt eine Zufallsvariable U erwartungstreuer Schätzer für eine Größe u, wenn $\mathbb{E}(U) = u$.

Für die Varianz des Mittelwertes erhält man

$$
\begin{aligned}
\mathbb{V}(\overline{X}) &= \mathbb{V}\left(\frac{1}{n}\sum_i X_i\right) = \frac{1}{n^2}\mathbb{V}\left(\sum_i X_i\right) = \frac{1}{n^2}\sum_i \mathbb{V}(X_i) \\
&= \frac{1}{n^2}\sum_i \sigma^2 = \frac{1}{n^2} n\,\sigma^2 = \frac{1}{n}\sigma^2\,, \qquad (6.2)
\end{aligned}
$$

wobei wir die Rechenregeln für die Varianz, die Additivität der Varianz bei unabhängigen Zufallsvariablen, sowie wieder die identische Verteilung der X_i verwendet haben.

Die Varianz geht also mit wachsendem Stichprobenumfang gegen 0; man sagt, der Mittelwert ist ein **konsistenter** Schätzer für den Erwartungswert. Allgemein heißt ein Schätzer U konsistent, wenn $\lim_{n\to\infty} \mathbb{V}(U) = 0$. Eigenschaften (6.1) und (6.2) zusammen besagen, dass die Schätzung mit zunehmendem Stichprobenumfang immer "besser" wird in dem Sinne, dass sie den wahren Wert immer besser trifft.

6.1.3 Zentraler Grenzwertsatz und Normalapproximation

Im vorigen Abschnitt haben wir den Mittelwert als Zufallsvariable betrachtet und deren Erwartungswert und Varianz untersucht. Für große n kann man sogar eine Aussage über die *Verteilung* des Mittelwerts machen; wir kennen sie schon in der Gestalt des zentralen Grenzwertsatzes, an den wir hier nochmal erinnern:

Satz 6.1 (Zentraler Grenzwertsatz). Seien $X_1, X_2, \dots, X_n$ unabhängige und identisch verteilte ZV mit Erwartungswert μ mit $|\mu| < \infty$ und Varianz $0 < \sigma^2 < \infty$. Dann ist ihr Mittelwert, $\overline{X} := \frac{1}{n}\sum_{i=1}^n X_i$, für großes n näherungsweise normalverteilt mit Erwartungswert μ und Varianz σ^2/n, genauer:

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma} \le z\right) = \Phi(z),$$

wobei $\Phi(z)$ die Verteilungsfunktion der Standardnormalverteilung bezeichnet. □

Bemerkung. Man beachte, dass die Zufallsgröße in der Aussage des zentralen Grenzwertsatzes gerade der sogenannte **standardisierte** Stichprobenmittelwert ist:

$$\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma} = \frac{\overline{X} - \mathbb{E}(\overline{X})}{\sqrt{\mathbb{V}(\overline{X})}}$$

mit Erwartungswert 0 und Varianz 1. Die analoge Transformation für normalverteilte Zufallsvariablen haben Sie schon in Gleichung (1.50) kennen gelernt.

Wie in Kapitel 4.2 bereits ausgeführt, lässt sich die Aussage des ZGWS durch Umkehrung der Standardisierung auch unmittelbar für den Stichprobenmittelwert umformulieren,

$$\overline{X} \approx \mathcal{N}(\mu, \frac{\sigma^2}{n}) \qquad (n \text{ groß}), \qquad (6.3)$$

sowie für die Summe der X_i:

$$\sum_{i=1}^{n} X_i \;=\; n\overline{X} \;\approx\; \mathcal{N}(n\mu, n\sigma^2) \qquad (n \text{ groß}). \tag{6.4}$$

Insbesondere kann für hinreichend großes n der linke Ausdruck durch den auf der rechten Seite sehr gut angenähert werden. Eine Illustration findet sich in Abb. 6.1. Sie bezieht sich auf die Approximation von $\overline{X}$ gemäß Gleichung (6.3). In anderen Quellen, insbesondere im Internet, finden sich sehr viel häufiger andere Darestllungen, etwa für die Approximation der *Summe* gemäß Gleichung (6.4). Lassen Sie sich dadurch nicht verwirren!

Die meisten Größen, die man in der Natur beobachtet, werden von einer großen Zahl zufälliger Faktoren beeinflusst. Der zentrale Grenzwertsatz sorgt dann dafür, dass deren Summe (bzw. deren Mittelwert) in guter Näherung normalverteilt ist, auch wenn das für die einzelnen Faktoren keineswegs zutrifft. (Dies gilt sogar dann, wenn die X_i nicht identisch verteilt sind, vergleiche die Bemerkungen aus Kapitel 4.) Der zentrale Grenzwertsatz ist somit die Ursache für die fundamentale Bedeutung und weite Verbreitung der Normalverteilung.

Außerdem erlaubt er die Approximation anderer Verteilungen durch die Normalverteilung, und zwar immer dann, wenn die Ereignisse aus hinreichend vielen unabhängigen Einzelereignissen zusammengesetzt sind. Darüber gibt die folgende Liste Auskunft (siehe auch Kap. 4.2, wo die Approximation genauer begründet wurde).

Beispiele und "Faustregeln" für die Normalapproximation:

i) Ist X binomialverteilt, gilt also $X \sim \mathcal{B}(n,p)$, mit $\sigma^2 = np(1-p) \geq 9$, so ist X näherungsweise normalverteilt nach $\mathcal{N}(np, np(1-p))$.

ii) Ist X Poisson-verteilt mit Parameter $\lambda = \mu = \sigma^2 \geq 10$, so ist X näherungsweise normalverteilt nach $\mathcal{N}(\lambda, \lambda)$.

iii) Ist X hypergeometrisch verteilt mit Parametern W, S, n, und gilt $n/(W + S) \leq 0.05$ sowie $np(1-p) \geq 9$ mit $p = W/(W+S)$, so ist X näherungsweise normalverteilt nach $\mathcal{N}(\mu, \sigma^2)$ mit $\mu = np$ und $\sigma^2 = np(1-p)$. (Man kann in diesem Fall die hypergeometrische Verteilung gut durch die Binomialverteilung annähern und dann die Normalapproximation aus i) benutzen.)

Bemerkung. Sind X_i für $i = 1, \ldots, n$ unabhängige, normalverteilte ZV mit $X_i \sim \mathcal{N}\big(\mu_i, \sigma_i^2\big)$, so ist ihre Summe *sogar exakt* normalverteilt, genauer:

$$\sum_i X_i \;\sim\; \mathcal{N}\Big(\sum_i \mu_i, \sum_i \sigma_i^2\Big)\,.$$

Dies ist, wie schon weiter oben erwähnt, der sogenannte *Additionssatz der Normalverteilung*.

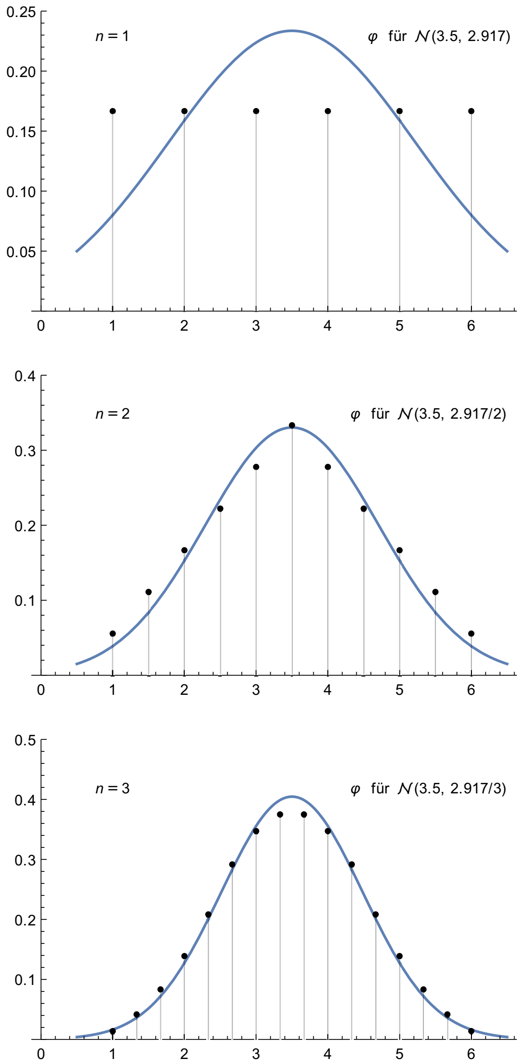

Abbildung 6.1: Exakte Verteilung der mittleren Augenzahl bei $n = 1, 2, 3$ Ausspielungen eines idealen Würfels, und deren Normalapproximation nach der lokalen Version des ZGWS, vgl. (6.3). Durchgezogene Kurve: Dichte $\varphi(x)$ der zugehörigen Normalverteilung. Beim einmaligen Würfeln hat die Augenzahl X den Erwartungswert $\mu = 3.5$ und die Varianz $\sigma^2 = 35/12 = 2.917$, der Stichprobenmittelwert $\overline{X}$ ist also gemäß (6.3) approximativ nach $\mathcal{N}(3.5, 2.917/n)$ verteilt.

6.1.4 Eigenschaften der empirischen Varianz

Die uns ebenfalls schon bekannte *empirische Varianz*

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \widehat{\sigma^2} \tag{6.5}$$

kann man – analog zur Vorgehensweise beim Mittelwert – als Realisierung der Zufallsgröße $S^2 := (1/(n-1))\sum_i (X_i - \overline{X})^2$ auffassen; für sie gilt $\mathbb{E}(S^2) = \sigma^2$, und $\lim_{n\to\infty} \mathbb{V}(S^2) = 0$. Die empirische Varianz ist somit eine erwartungstreue und konsistente Schätzung der Varianz σ^2 der der Stichprobe zugrundeliegenden Verteilung – also wieder eine Schätzung mit "guten" Eigenschaften. Ganz so selbstverständlich, wie es scheint, sind diese Eigenschaften aber nicht.

Beispiel. Die Zufallsvariable $\widetilde{S^2} := (1/n)\sum_i (X_i - \overline{X})^2$ hat Erwartungswert

$$\mathbb{E}(\widetilde{S^2}) = \mathbb{V}(X) - \mathbb{V}(\overline{X}) = (1 - (1/n))\,\sigma^2\,, \tag{6.6}$$

schätzt die Varianz also zu klein! Die Ursache ist, dass $\widetilde{S^2}$ die Schwankung um den *Stichprobenmittelwert* misst, während eigentlich die Schwankung um den wahren (aber unbekannten) Erwartungswert μ geschätzt werden soll. Da der Mittelwert aus der Stichprobe selber ermittelt wurde, fällt die Schwankung um ihn etwas geringer aus als die um μ. $\widetilde{S^2}$ liefert zwar auch einen Schätzer für die Varianz, aber dieser ist nicht erwartungstreu. Wir haben also den Grund für den Faktor $1/(n-1)$ in der Definition der empirischen Varianz gefunden.

Erwartungswert und Varianz sind die beiden wichtigsten Momente einer Verteilung; Mittelwert und empirische Varianz werden entsprechend als **Momentenschätzer** bezeichnet.

6.1.5 Schätzung weiterer Parameter der Verteilung

Momentenschätzer kann man auch benutzen, um weitere unbekannte Parameter von Verteilungen zu schätzen, etwa das p eines Bernoulliexperiments, oder den Intensitätsparameter der Poisson-Verteilung. Zu diesem Zweck werden die Beziehungen zwischen dem zu schätzenden Parameter und den Momenten der Verteilung benutzt, um den Parameter als Funktion der Momente auszudrücken. Die Momente werden dann durch die *Momentenschätzungen* ersetzt; so ergibt sich ein Schätzwert für den Parameter. Dieses Vorgehen nennt man die **Momentenmethode**.

Beispiel. i) Ist X binomialverteilt mit bekanntem n und unbekanntem p (Wahrscheinlichkeit für das Eintreffen des Ereignisses A), so ist $\mu = np$, also $p = \mu/n$. Ersetzt man nun μ durch den zugehörigen Schätzwert $\widehat{\mu}$, so erhält man einen Schätzwert $\widehat{p}$ für p:

$$\widehat{p} = \frac{\widehat{\mu}}{n} = \frac{\overline{x}}{n} = \frac{n_A}{n}, \tag{6.7}$$

wobei n die Gesamtzahl der Experimente und n_A die Zahl der Experimente mit Ausgang A ist. (Wir haben hier das n-fache Bernoulliexperiment als einfaches Binomialexperiment mit Stichprobenumfang 1 aufgefasst, also $\widehat{\mu} = \overline{x} = n_A$.) Gleichung (6.7) ähnelt verdächtig der Häufigkeitsdefinition der Wahrscheinlichkeit, siehe das Beispiel des Münzwurfs im ersten Kapitel! $\widehat{p}$ ist erwartungstreuer und konsistenter Schätzer für p.
ii) Sei X eine Poisson-verteilte ZV und $(x_1, \ldots, x_n)$ eine Stichprobe, wobei wir den Parameter λ nicht kennen. Dann ist $\mu = \lambda$ und somit

$$\widehat{\lambda} = \widehat{\mu} = \overline{x} \tag{6.8}$$

erwartungstreuer, konsistenter Schätzer des Parameters λ. Achtung: Es ist nicht zu empfehlen, anstelle des Mittelwerts die empirische Varianz zur Schätzung von λ zu verwenden, denn der resultierende Schätzer hat eine höhere Varianz als (6.8).

Es gibt noch weitere Methoden zur Parameterschätzung, wie zum Beispiel die sogenannte *Maximum-Likelihood-Methode* oder auch die *Methode der kleinsten Quadrate*. Dies werden wir hier nicht vertiefen, siehe die entsprechenden Lehrbücher. All diese Methoden liefern für den unbekannten Parameter jeweils einen einzigen Wert; eine sogenannte **Punktschätzung**. Dieser Wert entspricht i.a. nicht dem wahren Wert. Wir werden im nächsten Abschnitt eine Methode kennenlernen, anstelle eines einzigen Schätzwertes obere und untere Grenzen anzugeben, innerhalb derer wir den Wert des Parameters vermuten dürfen.

6.2 Schätzung von Konfidenzintervallen

Im letzten Abschnitt haben wir für unbekannte Parameter der Verteilung einzelne Schätzwerte ermittelt. Basierend auf solchen *Punktschätzungen* sollen nun Intervalle ermittelt werden, innerhalb derer wir den Parameter mit einer gewissen Sicherheit vermuten dürfen – sogenannte **Konfidenzintervalle**.

Wir betrachten zunächst die Wahrscheinlichkeit, eine standardnormalverteilte Zufallsvariable Z in einem *vorgegebenen symmetrischen Intervall* $[-c, c]$ anzutreffen: Das ist

$$\mathbb{P}(Z \in [-c, c]) = \mathbb{P}(Z \le c) - \mathbb{P}(Z \le -c) = \Phi(c) - \Phi(-c) = 2\Phi(c) - 1,$$

da wegen der Symmetrie der Standardnormalverteilung $\Phi(-c) = 1 - \Phi(c)$. Es soll nun umgekehrt ein symmetrisches Intervall so bestimmt werden, dass die Werte von Z mit *vorgegebener Wahrscheinlichkeit* $1 - \alpha$ hineinfallen. Dazu muss die Gleichung

$$2\Phi(c) - 1 = 1 - \alpha \iff \Phi(c) = 1 - \frac{\alpha}{2}$$

nach c aufgelöst werden. Da wir hierfür keine expliziten Formeln zur Verfügung haben, entnehmen wir den Wert der Normalverteilungstabelle (rückwärts benutzt).

Allgemein bezeichnet man denjenigen Wert z_q, für den

$$\Phi(z_q) = q \tag{6.9}$$

gilt, als das q-**Quantil** der Standardnormalverteilung, siehe Abbildung 6.2. Anders ausgedrückt ist $c = z_q = \Phi^{-1}(q)$, wobei Φ^{-1} die Umkehrfunktion von Φ bezeichnet. Die Umkehrfunktion existiert, da Φ streng monoton steigend ist.

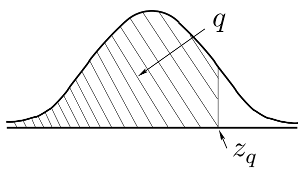

Abbildung 6.2: Der Wert z_q ist das q-Quantil der Verteilung. Die Masse der Verteilung bis z_q ist gerade q.

Mit Quantilen kann man jetzt das gesuchte Intervall explizit angeben. Wir haben

$$\mathbb{P}(Z \in [-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}]) = 1 - \alpha,$$

vergleiche Abbildung 6.3.

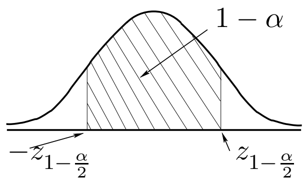

Abbildung 6.3: Die Werte der ZV Z sollen mit Wahrscheinlichkeit $1 - \alpha$ in ein symmetrisches Intervall um 0 fallen. Es sind die entsprechenden Quantile eingetragen.

Beispiel. Sei $\alpha = 0.05$ vorgegeben. Die Gleichung $\Phi(c) = 0.975$ ergibt $c = 1.96 = z_{0.975}$; diesen Wert findet man, indem man die Normalverteilungstabelle 'rückwärts' benutzt. Das gesuchte Intervall ist also $[-1.96, 1.96]$. In diesem Intervall liegen also 95% der Werte von Z.

Sei nun X eine nach $\mathcal{N}(\mu, \sigma^2)$ verteilte ZV. Als "Umkehrung der Standardisierung" schreiben wir X als $X = \mu + \sigma Z$, also gilt $Z \sim \mathcal{N}(0, 1)$. Damit erhalten wir

$$\mathbb{P}(X \in [\mu \; \pm z_{1-\frac{\alpha}{2}}\sigma]) \; = 1 - \alpha\,. \tag{6.10}$$

Wir betrachten nun – wie im letzten Abschnitt – unabhängige und identisch verteilte Zufallsvariablen $X_1, \ldots, X_n$ mit $\mathbb{E}(X_i) = \mu$ und $\mathbb{V}(X_i) = \sigma^2 > 0$ und fragen nach dem symmetrischen Intervall, das den Mittelwert $\overline{X} = \frac{1}{n}\sum_i X_i$ mit