Kapitel 4

Eindimensionale Zufallsgrößen

4.5 Wichtige Kennzahlen von Zufallsgrößen

Die wichtigsten Kennzahlen von Zufallsgrößen sind der **Erwartungswert** und die **Varianz**. In der frequentistischen Deutung der Wahrscheinlichkeit sollen diese dem Stichprobenmittel und der Stichprobenvarianz entsprechen. Wir erinnern an die Definition des Stichprobenmittels

$$\overline{x} = \frac{1}{n} \sum_{j=1}^{k} x_j H_j = \sum_{j=1}^{k} x_j h_j$$

mit den relativen Häufigkeiten h_j Für **diskrete Zufallsgrößen** X definieren wir den **Erwartungswert** $\mathrm{E}(X)$ gemäß der Entprechung $h_j \approx P(X = x_j)$:

$$\boxed{\mathrm{E}(X) = \sum_j x_j P(X = x_j)}$$

Beispiel (Alternativverteilung):

Hier ist $P(X = 1) = p$ und $P(X = 0) = 1 - p$, daher

$$\mathrm{E}(X) = 0 \cdot (1 - p) + 1 \cdot p = p,$$

und wir erkennen die wichtige Tatsache, dass der Erwartungswert der Alternativverteilung gerade gleich der Wahrscheinlichkeit des Ereignisses $X = 1$ ist.

Um den Erwartungswert für **kontinuierliche Zufallsgrößen** zu definieren, gehen wir von der Beobachtung aus, dass

$$\sum x_j h_j = \frac{\sum x_j h_j}{\sum h_j}$$

im Histogramm den Schwerpunkt der Rechtecksflächen darstellt, wenn man für x_j die Klassenmitten wählt. Bei Verfeinerung der Klasseneinteilung nähert sich der so berechnete Flächenschwerpunkt dem Stichprobenmittel $\overline{x}$.

Dies legt nahe, den Erwartungswert als den Schwerpunkt der Fläche unter der Dichtekurve zu definieren:

$$\boxed{\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x.}$$

Da $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$ ist, ist das erste Moment tatsächlich gleich dem Schwerpunkt.

Beispiel (Normalverteilung $\mathcal{N}(\mu, \sigma^2)$):

Es ergibt sich

$$\mathrm{E}(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \Big((x-\mu)\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} + \mu\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\Big)\,\mathrm{d}x = 0 + \mu\cdot 1 = \mu;$$

der Erwartungswert stimmt mit dem Modalwert der Dichtekurve überein.

In derselben Weise wollen wir nun die *Varianz* als Gegenstück zur Stichpobenvarianz

$$s^2 = \frac{n}{n-1}\sum_{j=1}^{k}(x_j - \overline{x})^2 h_j \approx \sum_{j=1}^{k}(x_j - \overline{x})^2 h_j$$

einführen. Die Stichprobenvarianz ist die mittlere quadratischen Abweichung vom Mittelwert. Entsprechend definieren wir die **Varianz** $\mathrm{V}(X)$ einer Zufallsgröße X als den Erwartungswert der quadratischen Abweichung von $\mathrm{E}(X)$, also:

$$\mathrm{V}(X) = \mathrm{E}\Big(\big(X - \mathrm{E}(X)\big)^2\Big)$$

Für **diskrete Zufallsgrößen** ergibt sich demnach

$$\boxed{\mathrm{V}(X) = \sum_j \big(x_j - \mathrm{E}(X)\big)^2 P(X = x_j)}$$

für **kontinuierliche Zufallsgrößen**:

$$\boxed{\mathrm{V}(X) = \int_{-\infty}^{\infty} \big(x - E(X)\big)^2 f(x)\,\mathrm{d}x.}$$

Wir fassen zusammen: Der *Erwartungswert* einer Zufallsgröße X ist das erste Moment oder der Massenmittelpunkt der Verteilung:

$$\mathrm{E}(X) = \mu_X = \begin{cases} \sum_j x_j\, p_j\,, & X \text{ diskret,} \\ \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x\,, & X \text{ kontinuierlich.} \end{cases}$$

Die *Varianz* ist das zentrierte zweite Moment oder der Erwartungswert der quadratischen Abweichung von μ_X:

$$\mathrm{V}(X) = \sigma_X^2 = \begin{cases} \sum_j (x_j - \mu_X)^2\, p_j\,, & X \text{ diskret,} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\,\mathrm{d}x\,, & X \text{ kontinuierlich.} \end{cases}$$

Die *Standardabweichung* oder *Streuung* σ_X ist die Quadratwurzel der Varianz, der *Variationskoeffizient* v_X der Quotient aus Standardabweichung und Erwartungswert:

$$\sigma_X = \sqrt{\mathrm{V}(X)}, \quad v_X = \frac{\sigma_X}{\mu_X}.$$

Beispiel 4.13 (Alternativverteilung) Es gilt $P(X = 1) = p$, $P(X = 0) = 1 - p$ und

$$\mathrm{E}(X) = 0 \cdot (1-p) + 1 \cdot p = p, \qquad \mathrm{V}(X) = (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p = p(1-p).$$

Beispiel 4.14 (Diskrete Gleichverteilung) Falls die Werte einer gleichverteilten diskreten Zufallsgröße die natürlichen Zahlen $x_1 = 1, \ldots, x_k = k$ sind (mit Wahrscheinlichkeiten $P(X = j) = 1/k$), kann man den Erwartungswert und die Varianz explizit berechnen:

$$\mathrm{E}(X) = \sum_{j=0}^{k} j \cdot \frac{1}{k} = \frac{(k+1)k}{2} \cdot \frac{1}{k} = \frac{k+1}{2},$$

$$\mathrm{V}(X) = \sum_{j=0}^{k} j^2 \cdot \frac{1}{k} - \frac{(k+1)^2}{4} = \frac{k(k+1)(2k+1)}{6} \cdot \frac{1}{k} - \frac{(k+1)^2}{4} = \frac{k^2-1}{12},$$

letzteres unter Verwendung der Formel $\mathrm{V}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$ (Anwendung 4.26).

Beispiel 4.15 (Binomialverteilung) Ihr Erwartungswert ergibt sich aus

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{j=0}^{n} jP(X=j) = \sum_{j=0}^{n} j\binom{n}{j}p^j(1-p)^{n-j} \\
&= \sum_{j=0}^{n} j \cdot \tfrac{n(n-1)\cdots(n-j+1)}{1\cdot 2\cdots(j-1)\cdot j}\, p^j(1-p)^{n-j} \\
&= np \cdot \sum_{j=1}^{n} \binom{n-1}{j-1}p^{j-1}(1-p)^{n-j} \\
&= np\sum_{l=0}^{n-1} \binom{n-1}{l}p^l(1-p)^{n-1-l} = np(p+1-p)^{n-1} \\
&= np.
\end{aligned}$$

Analog erhält man für die Varianz

$$\mathrm{V}(X) = \sum_{j=0}^{n} (j-np)^2\binom{n}{j}p^j(1-p)^{n-j} = np(1-p).$$

Beispiel 4.16 (Poissonverteilung) Es gilt:

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{j=0}^{\infty} j\tfrac{\mu^j}{j!}\,\mathrm{e}^{-\mu} = \mu\sum_{j=1}^{\infty} j\tfrac{\mu^{j-1}}{j!}\,\mathrm{e}^{-\mu} = \mu\sum_{i=0}^{\infty} \tfrac{\mu^i}{i!}\,\mathrm{e}^{-\mu} = \mu \\
\mathrm{V}(X) &= \sum_{j=0}^{\infty} (j-\mu)^2\tfrac{\mu^j}{j!}\,\mathrm{e}^{-\mu} = \mu
\end{aligned}$$

Bei der Poissonverteilung stimmen also Erwartungswert und Varianz überein und sind gleich dem Parameter μ.

Beispiel 4.17 (Normalverteilung $\mathcal{N}(\mu, \sigma^2)$) Den Erwartungswert haben wir oben zu

$$\mathrm{E}(X) = \mu$$

berechnet. Für die Varianz ergibt sich

$$\mathrm{V}(X) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x = \sigma^2$$

Die Parameter μ, σ^2 der Normalverteilung bedeuten also gerade Erwartungswert und Varianz.

Beispiel 4.18 (Kontinuierliche Gleichverteilung) Simple Integrationen ergeben die folgenden Formeln:

$$\begin{aligned}
\mathrm{E}(X) &= \int_a^b x \cdot \tfrac{1}{b-a}\,\mathrm{d}x = \tfrac{x^2}{2}\Big|_a^b \cdot \tfrac{1}{b-a} = \tfrac{b^2-a^2}{2} \cdot \tfrac{1}{b-a} = \tfrac{a+b}{2} \\
\mathrm{V}(X) &= \int_a^b \big(x - \tfrac{a+b}{2}\big)^2 \cdot \tfrac{1}{b-a}\,\mathrm{d}x = \tfrac{1}{(b-a)} \cdot \tfrac{1}{3}\big(x - \tfrac{a+b}{2}\big)^3 \Big|_a^b \\
&= \tfrac{1}{3(b-a)}\Big(\big(\tfrac{b-a}{2}\big)^3 - \big(\tfrac{a-b}{2}\big)^3\Big) = \tfrac{1}{12} \cdot \tfrac{(b-a)^3}{(b-a)} = \tfrac{(b-a)^2}{12}
\end{aligned}$$

Beispiel 4.19 (Exponentialverteilung) Partielle Integration ergibt:

$$\begin{aligned}
\mathrm{E}(X) &= \int_0^\infty \lambda x \mathrm{e}^{-\lambda x}\,\mathrm{d}x = \lambda\Big[-\tfrac{x}{\lambda}\,\mathrm{e}^{-\lambda x}\Big|_0^\infty + \int_0^\infty \tfrac{1}{\lambda}\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x\Big] = -\tfrac{1}{\lambda}\,\mathrm{e}^{-\lambda x}\Big|_0^\infty = \tfrac{1}{\lambda} \\
\mathrm{V}(X) &= \int_0^\infty \lambda\big(x - \tfrac{1}{\lambda}\big)^2\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \tfrac{1}{\lambda^2}.
\end{aligned}$$

Der Erwartungswert einer exponentialverteilten Größe ist der Kehrwert des Parameters λ.

Beispiel 4.20 (Dreiecksverteilung) Mit den Bezeichnungen von Beispiel **??** gilt:

$$\mathrm{E}(X) = \frac{1}{3}(a+m+b), \quad \mathrm{V}(X) = \frac{1}{18}(a^2+m^2+b^2-ma-mb-ab).$$

Beispiel 4.21 (Betaverteilung) Hier gilt

$$\mathrm{E}(X) = \frac{sa+rb}{r+s}, \qquad \mathrm{V}(X) = \frac{rs}{(r+s)^2(r+s+1)}\,(b-a)^2.$$

Beispiel 4.22 (Gumbelverteilung) Durch Rückführung auf Standardform und Integration erhält man:

$$\mathrm{E}(X) = x_0 + \gamma/\lambda, \qquad \mathrm{V}(X) = \pi^2/(6\lambda^2).$$

Dabei ist

$$\gamma \approx 0.577216\ldots = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k} - \log n$$

die so genannte *Euler-Mascheroni'sche Konstante*.

4.6 Die Verteilungsfunktion einer Zufallsgröße

Die *kumulative Verteilungsfunktion* oder kurz **Verteilung** einer Zufallsgröße X ist die reelle Funktion

$$F_X(x) = P(X \leq x) = P\big(\{\omega \in \Omega : X(\omega) \leq x\}\big).$$

Für die Verteilungsfunktion einer diskreten Zufallsgröße gilt

$$F_X(x) = P(X \leq x) = \sum_{x_j \leq x} p_j\,;$$

diese ist stückweise konstant und springt jeweils beim Wert x_j um p_j. Die Abbildung 4.1 zeigt die Einzelwahrscheinlichkeiten und die Verteilungsfunktion einer Alternativverteilung.

Abbildung 4.1: Verteilungsfunktion einer diskreten Zufallsgröße.

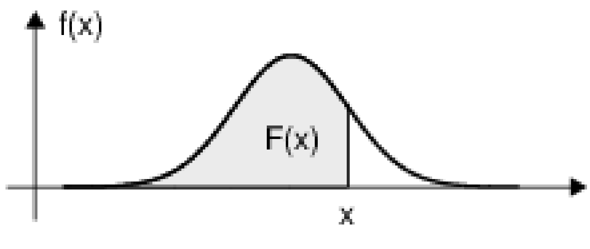
An der Verteilungsfunktion lassen sich die Einzelwahrscheinlichkeiten bei den Sprungstellen ablesen:

$$\begin{aligned} P(X = x_j) &= \text{Sprung von F(x) bei x} = \text{x}_\text{j} \\ &= F(x_j+) - F(x_j-). \end{aligned}$$

Kenntnis der Verteilungsfunktion ist also gleichbedeutend mit Kenntnis der Einzelwahrscheinlichkeiten.

Für die Verteilungsfunktion einer kontinuierlichen Zufallsgröße gilt:

$F_X(x) = \displaystyle\int_{-\infty}^{x} f_X(t)\,\mathrm{d}t$, im Bild:



Aus dem Hauptsatz der Differential- und Integralrechnung folgt die Beziehung

$$\frac{\mathrm{d}}{\mathrm{d}x} F_X(x) = f_X(x);$$

die Dichtefunktion kann durch Ableiten der Verteilungsfunktion gewonnen werden.

Die Abbildung 4.2 zeigt die Dichte und Verteilungsfunktion der Standardnormalverteilung.

Abbildung 4.2: Verteilungsfunktion einer kontinuierlichen Zufallsgröße.

Zusammenfassend halten wir fest, dass eine Zufallsgröße durch Angabe ihrer Verteilungsfunktion vollständig charakterisiert ist.

4.7 Funktionen von Zufallsgrößen

Die Aufgabe, Funktionen von Zufallsgrößen auszuwerten, tritt häufig auf, zum Beispiel bei der Ermittlung der Verteilung der Antwort eines mechanischen Systems, wenn die Eingabewerte Zufallsgrößen sind. Man spricht auch von Transformationen von Zufallsgrößen. Wir beginnen mit dem *diskreten Fall.*

Sei X eine Zufallsgröße mit Werten $x_1, \cdots, x_k$ und $u = U(x)$ eine bijektive Funktion (eineindeutige Zuordnung). Die Werte von $U = U(X)$ sind dann

$$u_1 = U(x_1), \ldots, u_k = U(x_k).$$

Im einfachen Beispiel der Quadratfunktion ist $u = U(x) = x^2$; die funktionale Zuordnung ist im Bereich $x \geq 0$ eineindeutig, vergleiche Abbildung 4.3.

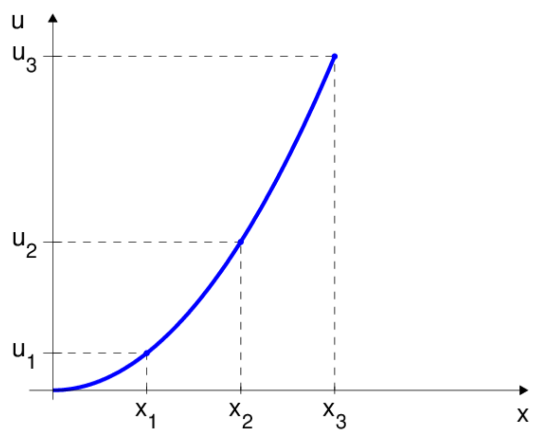


Abbildung 4.3: Transformation einer Zufallsgröße.

Wir suchen die Einzelwahrscheinlichkeiten $P(U = u_j)$. Diese lassen sich leicht mit Hilfe der Umkehrfunktion ausdrücken:

$$P(U = u_j) = P\big(U(X) = u_j\big) = P\big(X = U^{-1}(u_j)\big) = P(X = x_j).$$

Der Erwartungswert der transformierten Zufallsgröße U ergibt sich zu

$$\mathrm{E}(U) = \sum_{j=1}^{k} u_j P(U = u_j) = \sum_{j=1}^{k} U(x_j) P(X = x_j).$$

Im Beispiel der Quadratfunktion $u = x^2$ erhält man das zweite Moment

$$\mathrm{E}(X^2) = \sum_{j=1}^{k} x_j^2 P(X = x_j)$$

Auf diese Weise erhält man auch eine neue Interpretation der Varianz als Erwartungswert der transformierten Größe $U = (X - \mu)^2$, wobei $\mathrm{E}(X) = \mu$ den Erwartungswert der ursprunglichen Größe bedeutet:

$$\mathrm{V}(X) = \mathrm{E}\big((X - \mu)^2\big) = \sum_{j=1}^{k} (x_j - \mu)^2 P(X = x_j).$$

Im *kontinuierlicher Fall* erfolgt die Berechnung der Verteilung von $U = U(X)$ am besten mit Hilfe der Verteilungsfunktionen:

$$F_U(u) = P(U \le u) = P\big(U(X) \le u\big) = P\big(X \le U^{-1}(u)\big) = \int_{-\infty}^{U^{-1}(u)} f_X(x)\,\mathrm{d}x = F_X\big(U^{-1}(u)\big).$$

Die Dichte $F_U(u)$ erhält man durch Differenzieren:

$$f_U(u) = \frac{\mathrm{d}}{\mathrm{d}u} F_U(u) = \frac{\mathrm{d}}{\mathrm{d}x} F_X\big(U^{-1}(u)\big) \cdot \frac{\mathrm{d}}{\mathrm{d}u} U^{-1}(u) = f_X\big(U^{-1}(u)\big) \frac{\mathrm{d}}{\mathrm{d}u} U^{-1}(u)$$

Mit Hilfe der Substitution

$$x = U^{-1}(u), \quad \mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}u} U^{-1}(u)\,\mathrm{d}u, \quad u = U(x)$$

lässt sich der Erwartungswert berechnen:

$$\mathrm{E}(U) = \int_{-\infty}^{\infty} u f_U(u)\,\mathrm{d}u = \int_{-\infty}^{\infty} u f_X\big(U^{-1}(u)\big) \frac{\mathrm{d}}{\mathrm{d}u} U^{-1}(u)\,\mathrm{d}u = \int_{-\infty}^{\infty} U(x) f_X(x)\,\mathrm{d}x.$$

Ähnlich wie im diskreten Fall erhält man die Varianz als Erwartungswert der transformierten Größe $U = (X - \mu)^2$ mit $\mu = \mathrm{E}(X)$:

$$\mathrm{V}(X) = \mathrm{E}(U) = \mathrm{E}\big((X - \mu)^2\big) = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x)\,\mathrm{d}x.$$

Anwendung 4.23 (**Zentrierung der Normalverteilung**) Sei X nach $\mathcal{N}(\mu, \sigma^2)$ verteilt. Dann ist

$$U = \frac{X - \mu}{\sigma}$$

nach $\mathcal{N}(0,1)$ verteilt. Zur Herleitung dieser wichtigen Tatsache betrachten wir die Verteilungsfunktion

$$\begin{aligned} P(U \leq u) &= P\Big(\frac{X-\mu}{\sigma} \leq u\Big) = P(X \leq \mu + \sigma u) \\ &= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\mu+\sigma u} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} \mathrm{e}^{-y^2/2}\,\mathrm{d}y, \end{aligned}$$

wobei im letzten Schritt die Substitution $y = \frac{x-\mu}{\sigma}$ verwendet wurde. Das Ergebnis ist aber gerade die Verteilungsfunktion von $\mathcal{N}(0,1)$.

Die übliche Bezeichnungsweise für die Standardnormalverteilung ist

$$\varphi(u) = \tfrac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-u^2/2}, \quad \Phi(u) = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} \mathrm{e}^{-y^2/2}\mathrm{d}y.$$

In Tabelle 4.1 findet man ihre Werte für $0 \leq u \leq 3.69$.

Beispiel 4.24 (**Zentrierung und Tabellenrechnung**) Sei X nach $\mathcal{N}(\mu, \sigma^2)$ verteilt. Wir wollen die Wahrscheinlichkeit berechnen, dass ein Wert in einfacher Streudistanz vom Erwartungswert μ auftritt, das ist die Wahrscheinlichkeit

$$P(\mu - \sigma \leq X \leq \mu + \sigma).$$

Umformen zeigt, dass diese Wahrscheinlichkeit gleich

$$P(-\sigma \leq X - \mu \leq \sigma) = P\Big(-1 \leq \frac{X-\mu}{\sigma} \leq 1\Big) = P(-1 \leq U \leq 1)$$

ist, wobei die Größe U nach $\mathcal{N}(0,1)$ verteilt ist. In Tabelle 4.1 sind nur einseitige Wahrscheinlichkeiten angegeben. Wir machen uns die Symmetrie der Normalverteilung zunutze:

Abbildung 4.4: Symmetrischer einfacher Streubereich der Standardnormalverteilung.

Die Abbildung zeigt, dass gilt

$$\begin{aligned} P(-1 \leq U \leq 1) &= P(-\infty < U \leq 1) - P(-\infty < U < -1) \\ &= P(-\infty < U \leq 1) - P(U > 1) \\ &= P(U \leq 1) - \big(1 - P(U \leq 1)\big) \\ &= 2P(U \leq 1) - 1 = 2 \cdot 0.8413 - 1 = 0.683 \approx 68\%. \end{aligned}$$

wobei die Zahlenwerte der Tabelle zu entnehmen sind. In ähnlicher Weise berechnet man die Prozentanteile in zwei- und dreifacher Streudistanz (Abbildung **??**).

Als konkretes Beispiel nehmen wir die mittlere Tagestemperatur im Jänner 2001 in Innsbruck. In Abschnitt **??** haben wir eine Normalverteilung mit Parametern $\mu = 0.76$, $\sigma^2 = 9.1085$, $\sigma = 3.018$ angepasst. Unter der Annahme, dass dies ein zutreffendes Modell für die Jännertemperaturen ist, können wir die Wahrscheinlichkeit ausrechnen, dass die mittlere Tagestemperatur größer als 5° ist:

$$\begin{aligned} P(X \geq 5) &= 1 - P(X \leq 5) = 1 - P(X - 0.76 \leq 4.24) = 1 - P\big(\tfrac{X-0.76}{3.018} \leq \tfrac{4.24}{3.018}\big) \\ &= 1 - P\big(U \leq \tfrac{4.24}{3.018}\big) = 1 - P(U \leq 1.4049) \approx 1 - \Phi(1.405). \end{aligned}$$

Wir entnehmen der Tabelle: $\Phi(1.40) = 0.9192$, $\Phi(1.41) = 0.9207$ und mittels Interpolation $\Phi(1.405) \approx 0.92$. Insgesamt also

$$P(X \geq 5) = 1 - \Phi(1.405) = 0.08 \approx 8\%,$$

wie vor Abbildung **??** behauptet.

Anwendung 4.25 (**Die Lognormalverteilung**) Diese wird häufig zur Beschreibung rechtsschiefer unimodaler Verteilungen positiver Größen verwendet. Eine Zufallsgröße U ist lognormalverteilt, wenn ihr (natürlicher) Logarithmus normalverteilt ist. Man verwendet dann auch die Parameter der assoziierten Normalverteilung zur Spezifikation:

$$U \sim \mathcal{L}og\mathcal{N}(\mu, \sigma^2) \qquad \Longleftrightarrow \qquad X = \log U \sim \mathcal{N}(\mu, \sigma^2).$$

Umgekehrt gilt natürlich ebenso: ist X normalverteilt nach $\mathcal{N}(\mu, \sigma^2)$, so ist $U = \mathrm{e}^X$ lognormalverteilt nach $\mathcal{L}og\mathcal{N}(\mu, \sigma^2)$. Ist $f_X(x)$ die Dichtefunktion der Normalverteilung, so erhält man die Dichtefunktion der Lognormalverteilung zu

$$\begin{aligned} f_U(u) &= f_X\big(U^{-1}(u)\big)\frac{\mathrm{d}}{\mathrm{d}u}U^{-1}(u) = f_X(\log u)\frac{1}{u} \\ &= \begin{cases} \frac{1}{\sqrt{2\pi}\sigma u}\,\mathrm{e}^{-(\log u - \mu)^2/2\sigma^2}, & u > 0, \\ 0\,, & u \leq 0\,. \end{cases} \end{aligned}$$

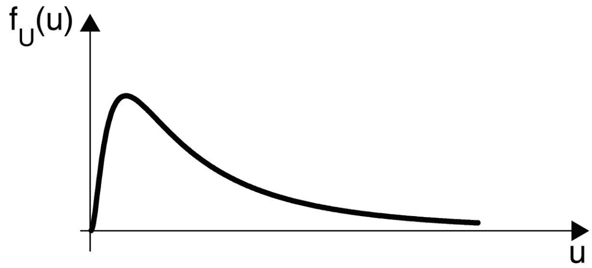


Abbildung 4.5: Dichte der Standardlognormalverteilung.

Erwartungswert und Varianz der Lognormalverteilung berechnen sich zu

$$\mathrm{E}(U) = \mathrm{e}^{\mu + \sigma^{2/2}}, \;\; \mathrm{V}(U) = \mathrm{e}^{2\mu + \sigma^2}(\mathrm{e}^{\sigma^2} - 1).$$

Anwendung 4.26 (Eine nützliche Formel) Die Varianz einer Zufallsgröße X lässt sich wie folgt berechnen:

$$\mathrm{V}(X) = \mathrm{E}(X^2) - \big(E(X)\big)^2.$$

Um das zu sehen, fassen wir die Varianz – wie oben dargelegt – als Funktion der Zufallsgröße X auf und können weiter umrechnen, mit $\mu = \mathrm{E}(X)$:

$$\begin{aligned} \mathrm{V}(X) &= \mathrm{E}\big((X-\mu)^2\big) = \mathrm{E}\big(X^2 - 2\mu X + \mu^2\big) \\ &= \mathrm{E}(X^2) - 2\mu\mathrm{E}(X) + \mu^2 = \mathrm{E}(X^2) - \mu^2 = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2. \end{aligned}$$

Wir haben dabei verwendet, dass das Bilden des Erwartungswertes eine lineare Operation ist, also mit Addition und Skalarmultiplikation vertauschbar ist, wie die Formeln am Beginn des Abschnitts zeigen.

4.8 Anhang: Verteilungen und ihre Anwendungen

In diesem Anhang werden einige weitere Verteilungstypen vorgestellt und typische Anwendungen der besprochenen Verteilungen zusammengestellt.

Die wichtigsten Verteilungen sind auch in MATLAB implementiert. Welche Verteilungen in MATLAB implementiert sind, findet man in der MATLAB-Dokumentation

```
doc stats
```

oder im Internet unter

`https://de.mathworks.com/help/stats/supported-distributions.html`

Es gibt in MATLAB auch die Möglichkeit Objekt-orientiert mit den sogenannten `probability distribution object` - Objekten zu arbeiten. Dazu werden wir hier aber nicht eingehen. Wie die Verteilungen in Matlab abgefragt werden können, folgt unten.

Diskrete Verteilungen:

Diskrete Gleichverteilung: Zufallsgrößen, die endlich viele Werte mit gleicher Wahrscheinlichkeit annehmen. Modell des ungezinkten Würfels. Anwendung bei statistischen Tests, zum Beispiel Test auf Trend einer Zeitreihe.

Alternativverteilung: X nimmt die Werte 1 (Erfolg) mit Wahrscheinlichkeit p und 0 (Misserfolg) mit Wahrscheinlichkeit $1-p$ an. Anwendung: Ja-Nein-Entscheidungen, Modellierung des Eintretens oder Nichteintretens eines gegebenen Ereignisses.

Binomialverteilung: Bernoulli-Experiment: n-malige unabhängige Wiederholung eines 0-1-Versuches mit Erfolgswahrscheinlichkeit p. Eine binomialverteilte Größe X nimmt die Werte $0, 1, 2, \dots, n$ an und beschreibt die Anzahl der Erfolge bei n 0-1-Versuchen.

Geometrische Verteilung: Beschreibt die Anzahl der Misserfolge bis zum ersten Erfolg bei Wiederholung eines 0-1-Versuches.

Negative Binomialverteilung: Beschreibt die Anzahl der 0-1-Versuche, die erforderlich sind, um $r \geq 1$ Erfolge zu erzielen. Für $r = 1$ reduziert sie sich auf die geometrische Verteilung. Anwendung in der Versicherungsmathematik.

Hypergeometrische Verteilung: Anwendung in der Qualitätskontrolle. Gegeben: N Werkstücke, davon sind M defekt. Es wird eine Probe von K Stück entnommen. Die hypergeometrische Verteilung beschreibt die Anzahl der defekten Stücke unter den K entnommenen.

Poissonverteilung: Beschreibt die Anzahl von Ereignissen, die in einem gegebenen Zeitintervall eintreten. Die Wartezeit zwischen den Ereignissen ist exponentialverteilt. Anwendungen: Anzahl eintreffender Anrufe in einer Telefonzentrale, Anzahl eintreffender Jobs auf einem EDV-Server, Anzahl von Sturmfluten oder Regenfronten in einem Jahr, Anzahl von Unfällen eines Versicherungsnehmers im Jahr.

Kontinuierliche Verteilungen

Kontinuierliche Gleichverteilung: Grundverteilung bei der Erzeugung von Zufallszahlen. Koninuierliche Verteilung mit geringstem Informationsgehalt (Verwendung in der Bayes-Statistik als Priorverteilung).

Dreiecksverteilung: Wahrscheinlichkeitstheoretisches Modell einer Dreipunktschätzung (minimaler, häufigster, maximaler Wert).

Betaverteilung: Vierparametrige Verteilung, die unimodale Zufallsgrößen mit Werten in einem Intervall $[a, b]$ modelliert. Die beiden anderen Parameter erlauben es, verschiedene Glattheits- und Schiefegrade zu erreichen.

Normalverteilung: Eine weit anwendbare symmetrische Verteilung. Summen unabhängiger Zufallsgrößen hinreichend hoher Summandenzahl sind näherungsweise normalverteilt. Stichprobenmittel sind näherungsweise normalverteilt. Summen und Vielfache normalverteilter Größen sind normalverteilt.

Lognormalverteilung: Eine Zufallsgröße ist lognormalverteilt, wenn ihr natürlicher Logarithmus normalverteilt ist. Modellierung nicht negativer, rechtsschiefer Verteilungen mit *fat tails*, also langsamem Abfall im Unendlichen.

Exponentialverteilung: Kombinatorisches Modell für die Lebensdauer eines radioaktiven Partikels. Wird allgemein zur Modellierung von Lebensdauern (z. B. von Geräten) und Wartezeiten verwendet. Modelliert die Wartezeit zwischen zwei aufeinanderfolgenden Ereignissen eines Poissonprozesses.

Erlangverteilung: Eine Summe von exponentialverteilten Größen ist Erlang-verteilt.

Gammaverteilung: Allgemeine Verteilung, die als Spezialfälle die Exponentialverteilung, die Chi-Quadrat-Verteilung und die Erlangverteilung enthält. Für ganzzahlige Potenzen p modelliert sie die Wartezeit bis zum p-ten Ereignis in einem Poissonprozess.

Laplaceverteilung: Besteht aus zwei um Null gespiegelten, aneinandergefügten Exponentialverteilungen.

Gumbelverteilung: Verteilung der Extremwerte (Maxima und Minima) einer Größe, deren Verteilung exponentiell im Unendlichen abklingt. Anwendung: Hochwasserstatistik.

Weibullverteilung: Verallgemeinerung der Exponentialverteilung. Anwendungen in der Extremwertstatistik und zur Beschreibung von Lebensdauern.

Paretoverteilung: Beschreibt nichtnegative Zufallsgrößen, deren kleinen Werte mit sehr hoher und deren große Werte mit kleiner Wahrscheinlichkeit auftreten. Anwendungsbeispiel: Vermögensverteilung in einer Bevölkerung, Größenverteilung nach Einwohnerzahl von Städten.

Cauchyverteilung: Verteilung einer Zufallsgröße, deren extreme Werte eine sehr hohe Wahrscheinlichkeit besitzen. Der Erwartungswert des Betrags $|X|$ einer Cauchy-verteilten Größe X ist ∞. Anwendung in der Simulation.

Maxwellverteilung: Eine Wahrscheinlichkeitsverteilung der Statistischen Physik. Beschreibt die Verteilung der Teilchengeschwindigkeiten in einem idealen Gas.

Chi-Quadrat-Verteilung: Wichtige Testverteilung. Die Summe der Quadrate von n unabhängigen, standardnormalverteilten Größen ist χ^2-verteilt mit n Freiheitsgraden. Lineare Abhängigkeiten unter den Größen führen zur Reduktion der Freiheitsgrade. Die Stichprobenvarianz ist χ^2-verteilt mit $n-1$ Freiheitsgraden.

t-Verteilung: Wichtige Testverteilung. Ist X standardnormalverteilt und unabhängig von der mit n Freiheitsgraden χ^2-verteilten Größe Y, so besitzt $X/\sqrt{Y/n}$ eine t-Verteilung mit n Freiheitsgraden. Wird auch als Student-Verteilung bezeichnet.

F-Verteilung: Wichtige Testverteilung. Sind V_m und V_n unabhängig und χ^2-verteilt mit m bzw. n Freiheitsgraden, so ist nV_m/mV_n F-verteilt mit (m,n) Freiheitsgraden. Wird auch als Fisher-Verteilung bezeichnet.

Anwendung in MATLAB

Dichtefunktionen

Die Dichtefunktionen der wichtigsten Verteilungen können in MATLAB mit Hilfe der Funktion `pdf` generiert werden.

```
y = pdf(name,x,A,B,C,...)
```

Der Wert `y` gibt den Wert der Dichtefunktion der jeweiligen Verteilung (`name`) an der Stelle `x` zurück. Die Anzahl der Parameter `A,B,C,...` hängen von der jeweiligen Verteilung ab und werden für die Charakterisierung der Verteilung benötigt.

z.B. Normalverteilung mit Erwartungswert $\mu = 1$ und Streuung $\sigma = 5$ für die Werte $x =$ -2, -1, 0, 1, 2.

```
x = [-2 -1 0 1 2];

mu = 1;
sigma = 5;
y = pdf('Normal',x,mu,sigma)
```

Verteilungsfunktionen

Die Verteilungsfunktion der wichtigsten Verteilungen können in MATLAB mit Hilfe der Funktion `cdf` erzeugt werden.

```
y = cdf(name,x,A,B,C,...)
```

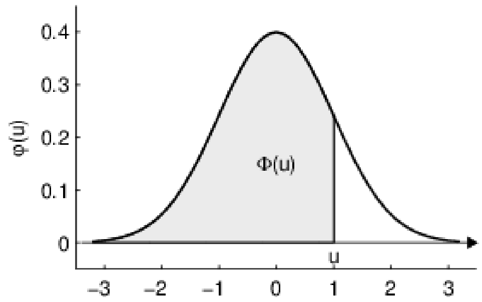
Der Wert `y` gibt den Wert der Verteilungsfunktion der jeweiligen Verteilung (`name`) an der Stelle `x` zurück. Die Anzahl der Parameter `A,B,C,...` hängen von der jeweiligen Verteilung `name` ab und werden für die Charakterisierung der Verteilung (`name`) benötigt.

z.B. Verteilungsfunktion für die Normalverteilung mit Erwartungswert $\mu = 1$ und Streuung $\sigma = 5$ für die Werte $x =$ -2, -1, 0, 1, 2.

```
x = [-2 -1 0 1 2];

mu = 1;
sigma = 5;
y = cdf('Normal',x,mu,sigma)
```

Tabelle der Standardnormalverteilung $\mathbf{N(0,1)}$



u	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09	φ(u)
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	0.3989
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	0.3970
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	0.3910
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	0.3814
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	0.3683
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	0.3521
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	0.3332
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	0.3123
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	0.2897
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	0.2661
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	0.2420
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	0.2179
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	0.1942
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	0.1714
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	0.1497
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	0.1295
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	0.1109
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	0.0940
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	0.0790
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	0.0656
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	0.0540
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	0.0440
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	0.0355
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	0.0283
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	0.0224
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	0.0175
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	0.0136
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	0.0104
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	0.0079
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	0.0060
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990	0.0044
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993	0.0033
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995	0.0024
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997	0.0017
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998	0.0012
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.0009
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.0006

Tabelle 4.1: Werte der kumulativen Standardnormalverteilung.