

Verteilungen

Allgemein

| Begriffe | |
|---------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|
| Zufallsvariable X | Funktion $X : \Omega \rightarrow \mathbb{R}$, bzw. Zufallsexperiment mit Ausgang in \mathbb{R} . |
| Kum. Verteilungsfn. $F(x)$ | $F(x) = \mathbb{P}(X \leq x)$, $x \in \mathbb{R}$ |
| Erwartungswert $\mathbb{E}(X) = \mu_X$ | Maß für die mittlere Lage, "Durchschnitt von unendlich vielen Wiederholungen". |
| Varianz $\text{Var}(X) = \sigma_X^2$ ($= \mathbb{E}((X - \mu_X)^2)$) | Maß für die Streuung um die mittlere Lage. |
| Standardabweichung σ_X ($= \sqrt{\text{Var}(X)}$) | Maß für die Streuung um die mittlere Lage (<i>gleiche</i> Einheit wie die Zufallsvariable). |
| Variationskoeffizient σ_X / μ_X (für positive ZV) | Relative Streuung, <i>dimensionslos</i> . |
| Quantil q_α , $\alpha \in (0, 1)$ | Wert, der mit Wahrscheinlichkeit α unterschritten wird: $\mathbb{P}(X \leq q_\alpha) = \alpha$. |

Rechenregeln Erwartungswert und Varianz

$$\mathbb{E}(a + bX) = a + b \cdot \mathbb{E}(X), \quad a, b \in \mathbb{R}$$

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

$$\text{Var}(a + bX) = b^2 \text{Var}(X), \quad a, b \in \mathbb{R}$$

$$\text{Var}(a) = 0, \quad a \in \mathbb{R}$$

$$\mathbb{E}(a + bX + cY) = a + b \cdot \mathbb{E}(X) + c \cdot \mathbb{E}(Y), \quad a, b, c \in \mathbb{R}$$

$$\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y), \quad \text{falls } X, Y \text{ unabh. (unkorr.)}$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y), \quad \text{falls } X, Y \text{ unabh. (unkorr.)}$$

Diskrete Verteilungen

Grundbegriffe

$$\text{Wertebereich} \quad W = \{x_1, x_2, \dots\}$$

$$\text{Wahrscheinlichkeitsfunktion, Einzelwahrscheinlichkeit} \quad p(x_k) = \mathbb{P}(X = x_k) \in [0, 1]$$

$$\text{Es gilt: } \sum_{k \geq 1} p(x_k) = 1$$

$$\text{Erwartungswert} \quad \mathbb{E}(X) = \sum_{k \geq 1} x_k p(x_k)$$

$$\mathbb{E}(g(X)) = \sum_{k \geq 1} g(x_k) p(x_k)$$

Bernoulli-Verteilung [Bernoulli(p)]

X : Binäres Experiment: "Erfolg" (= 1) vs. "Misserfolg" (= 0) mit Erfolgswahrscheinlichkeit $p \in (0, 1)$. Erfolg und Misserfolg können selber definiert werden. Zusammenfassend

$$X = \begin{cases} 0 & \text{W'keit } 1-p \\ 1 & \text{W'keit } p \end{cases}$$

$$W = \{0, 1\}$$

$$\mathbb{P}(X = x) = p^x \cdot (1-p)^{1-x}, \quad x \in W$$

$$\mathbb{E}(X) = p$$

$$\text{Var}(X) = p \cdot (1-p)$$

Eigenschaften:

$$X_1, \dots, X_n \text{ i.i.d. } \sim \text{Bernoulli}(p): \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$$

(Summe von n unabh. Bernoulli(p)-ZV. ist eine Binomialvert.)

Binomialverteilung [Bin(n, p)]

X : Anzahl Erfolge bei n unabhängigen Bernoulli(p)-Experimenten.

$$W = \{0, 1, 2, \dots, n\}$$

$$\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in W$$

$$\mathbb{E}(X) = np$$

$$\text{Var}(X) = np(1-p)$$

Eigenschaften:

- $n = 1$: Bernoulli-Verteilung
- $X_1 \sim \text{Bin}(n_1, p), X_2 \sim \text{Bin}(n_2, p)$ unabh.:
 $X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$ (Achtung: Überall *gleiches* p !)

Approximationen:

- n groß, p klein: (Faustregel: $n \geq 50, p \leq 0.05$)
Poisson-Verteilung mit $\lambda = np$
- n groß: (Faustregel: $np(1-p) \geq 9$)
Normalverteilung mit $\mu = np, \sigma^2 = np(1-p)$

Geometrische Verteilung [Geom(p)]

X : Anzahl Wiederholungen von unabhängigen Bernoulli(p)-Experimenten bis zum **ersten Erfolg**.

$$W = \{1, 2, \dots\} \text{ (unbeschränkt!)}$$

$$\mathbb{P}(X = x) = p \cdot (1-p)^{x-1}, \quad x \in W$$

$$\mathbb{E}(X) = \frac{1}{p} \text{ (Wiederkehrperiode)}$$

$$\text{Var}(X) = \frac{1-p}{p^2}$$

Poisson-Verteilung [Pois(λ)]

X : Zähldaten

$$W = \{0, 1, 2, \dots\} \text{ (unbeschränkt!)}$$

$$\mathbb{P}(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x \in W$$

$$\mathbb{E}(X) = \lambda$$

$$\text{Var}(X) = \lambda$$

Eigenschaften:

$$X_1 \sim \text{Pois}(\lambda_1), X_2 \sim \text{Pois}(\lambda_2), \text{ unabh.: } X_1 + X_2 \sim \text{Pois}(\lambda_1 + \lambda_2)$$

Approximationen:

$$\lambda \text{ groß (Faustregel: } \lambda \geq 9): \text{ Normalverteilung mit } \mu = \lambda, \sigma^2 = \lambda$$

Stetige Verteilungen

Grundbegriffe

Wertebereich W enthält ein Intervall.

Dichte: Funktion $f(x) \geq 0$ mit

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Achtung: $f(x) > 1$ möglich!

Kumulative Verteilungsfunktion

$$F(x) = \int_{-\infty}^x f(u) du$$

$$f(x) = F'(x)$$

$$\mathbb{P}(X \in (a, b)) = \int_a^b f(x) dx = F(b) - F(a)$$

Erwartungswert

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x) dx, \quad \mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x) dx$$

Varianz

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx$$

Quantil

$$q_\alpha = F^{-1}(\alpha)$$

Achtung

$$\mathbb{P}(X = x) = 0.$$

Uniforme Verteilung [Uni(a, b)]

Stetige Version des Laplace-Modells, "keine Region bevorzugt".

$W = [a, b]$

Dichte

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b \text{ (außerhalb 0)}$$

Kumulative Verteilungsfunktion

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

$$\mathbb{E}(X) = \frac{a+b}{2}$$

$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

Eigenschaften:

- $X \sim \text{Uni}(a, b)$: $c + dX \sim \text{Uni}(c + da, c + db)$
- $X \sim \text{Uni}(0, 1)$: $Y = F_Y^{-1}(X)$ hat die Verteilungsfunktion F_Y (!)

Normalverteilung [$\mathcal{N}(\mu, \sigma^2)$]

Gauß'sche Glockenkurve.

Häufige Verteilung für Messfehler oder alles, was sich als Summe von "Einzeleffekten" interpretieren lässt (ZGWS).

$W = \mathbb{R}$

Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad x \in \mathbb{R}$$

Kumulative Verteilungsfunktion ist nicht geschlossen darstellbar,

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

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

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

$$\mathbb{P}(|X - \mu| \leq \sigma) = 0.68$$

$$\mathbb{P}(|X - \mu| \leq 2\sigma) = 0.95$$

$\mu = 0, \sigma = 1$: **Standardnormalverteilung** mit kum. Vert.fn $\Phi(\cdot)$

Eigenschaften für $X \sim \mathcal{N}(\mu, \sigma^2)$:

- $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$ (**Standardisierung**)
- $a + bX \sim \mathcal{N}(a + b\mu, b^2\sigma^2)$

Lognormalverteilung [Lognormal(μ, σ^2)]

Flexible Verteilungsfamilie für positive Messwerte. Form von fast symmetrisch bis deutlich rechtsschief. Analog zum zentralen Grenzwertsatz kann die Lognormalverteilung verwendet werden als Approximation für das Produkt vieler positiver i.i.d. Zufallsvariablen.

$W = \mathbb{R}_+$

Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2\right\}, \quad x \geq 0 \text{ (sonst 0)}$$

Kumulative Verteilungsfunktion

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \Phi\left(\frac{\log(x)-\mu}{\sigma}\right) & x > 0 \end{cases}$$

$$\mathbb{E}(X) = e^{\mu + \sigma^2/2}$$

$$\text{Var}(X) = e^{2\mu + 2\sigma^2}(e^{\sigma^2} - 1)$$

Eigenschaften für $X \sim \text{Lognormal}(\mu, \sigma^2)$:

- $\log(X) \sim \mathcal{N}(\mu, \sigma^2)$ bzw.
- $X = e^Y$ mit $Y \sim \mathcal{N}(\mu, \sigma^2)$

Exponentialverteilung [Exp(λ)]

Einfaches Modell für Warte- oder Ausfallzeiten, "gedächtnislos". Parameter λ wird als Ausfallrate interpretiert.

$W = \mathbb{R}_+$

Dichte

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0 \text{ (sonst 0)}$$

Kumulative Verteilungsfunktion

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

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

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

Eigenschaften für $X \sim \text{Exp}(\lambda)$:

$$cX \sim \text{Exp}(\lambda/c)$$

Wichtige Sätze

\sqrt{n} -Gesetz

Idee: Standardabweichung des Mittelwerts nimmt mit $1/\sqrt{n}$ ab.

X_1, \dots, X_n i.i.d. mit $\text{Var}(X) = \sigma_X^2$:

$$\text{Var}(\bar{X}_n) = \frac{\sigma_X^2}{n} \quad \sigma_{\bar{X}_n} = \frac{\sigma_X}{\sqrt{n}}$$

Also: Um die Standardabweichung eines **Mittelwerts** zu halbieren, braucht man *viermal* so viele Beobachtungen. Die Standardabweichung einer *Einzelmessung* bleibt natürlich *unverändert*.

Gesetz der großen Zahlen (GGZ)

Idee: Wenn man über viel mittelt, kommt man dem Erwartungswert immer näher.

Satz: X_1, \dots, X_n i.i.d. mit $\mathbb{E}(X) = \mu$:

$$\bar{X}_n \rightarrow \mu \quad (n \rightarrow \infty).$$

Zentraler Grenzwertsatz (ZGWS)

Idee: Summen oder Mittelwerte von i.i.d.-Zufallsvariablen sind approximativ normalverteilt.

Satz: X_1, \dots, X_n i.i.d. mit $\mathbb{E}(X) = \mu$ und $\text{Var}(X) = \sigma^2$:

$$S_n \approx \mathcal{N}(n\mu, n\sigma^2) \quad \bar{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$