

Stochastik für Informatiker



Dr. rer. nat. Johannes Riesterer

Verteilung

Sei $(\mathbb{R}, \mathcal{B}, P)$ ein Wahrscheinlichkeitsraum über \mathbb{R} mit Borell'scher σ -Algebra \mathcal{B} . Die Abbildung

$$F_P(x) = P((-\infty, x])$$

wird Wahrscheinlichkeitsverteilung genannt.

Verteilung einer Zufallsvariablen

Ist (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine reelle Zufallsvariable, so erhalten wir die zugehörige Wahrscheinlichkeitsverteilung

$$F_X(x) = P_X(X \leq x) .$$

Ist f eine Dichte auf \mathbb{R} und $F_X = P_f$ so schreiben wir auch $X \sim P_f$

Verteilung

Ist $f : \mathbb{R} \rightarrow [0, 1]$ eine dichte und P_f das zugehörige Wahrscheinlichkeitsmaß, dann gilt

$$\frac{d}{dx} F_{P_f}(x) = f(x)$$

(Hauptsatz der Integral- und Differentialrechnung)

Gleichverteilung

Die Gleichverteilung $U(a, b)$ auf einem Intervall $(a, b) \subset \mathbb{R}$ ist definiert durch

$$\text{Dichte: } f(x) := \frac{1_{(a,b)}}{|a-b|}$$

$$\Rightarrow \text{Verteilung: } F(x) = P_f((-\infty, x)) = \int_{-\infty}^x \frac{1_{(a,b)}}{|b-a|} dt$$

$$= \begin{cases} 0 & \text{für } x \leq a \\ \frac{x-a}{|b-a|} & \text{für } a \leq x \leq b \\ 1 & \text{für } x \geq b \end{cases}$$

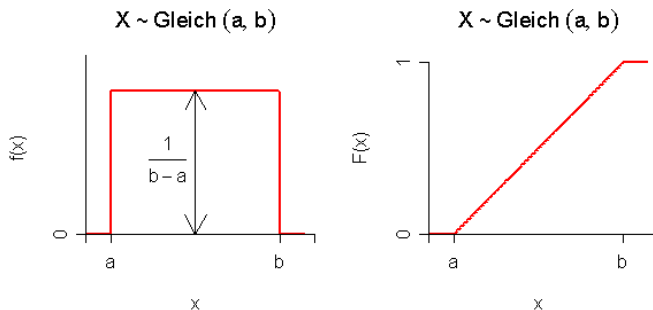


Figure: Quelle: Wikipedia

Gleichverteilung

Sei $X \sim U(a,b)$.

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{b-a} \int_a^b x \cdot 1 dx = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{a+b}{2}$$

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{1}{b-a} \int_a^b x^2 \cdot 1 dx - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{1}{3} \frac{b^3 - a^3}{b-a} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{1}{12} (4b^2 + 4ab + 4a^2 - 3a^2 - 6ab - 3b^2) = \frac{1}{12} (b-a)^2\end{aligned}$$

Normalverteilung

Die Normalverteilung $N(\mu, \sigma^2)$ auf \mathbb{R} ist definiert durch

Dichte: $f(x) := \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

\Rightarrow Verteilung: $F(x) = N(\mu, \sigma^2)(-\infty, x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$

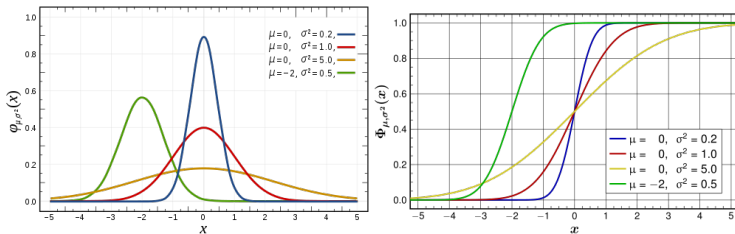


Figure: Quelle: Wikipedia

Normalverteilung

Sei $X \sim N(\mu, \sigma)$.

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

$$\mathbb{V}(X) = \sigma^2$$

Poissonverteilung

Die Poissonverteilung $Pois(\lambda)$ auf $\mathbb{N}_{\geq 0}$ ist definiert durch

$$P_{\lambda}(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$\Rightarrow F_{\lambda}(n) = \sum_{k=0}^n P_{\lambda}(k) = e^{-\lambda} \sum_{k=0}^n \frac{\lambda^k}{k!}$$

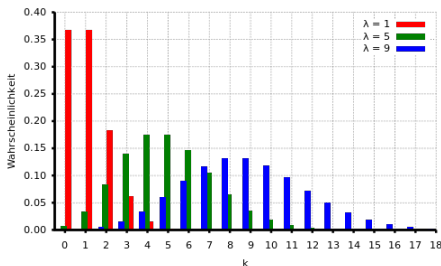


Figure: Quelle: Wikipedia

Poisson Verteilung

Die Poisson Verteilung beschreibt das Auftreten von seltenen Ereignissen und spielt bei Zählprozessen eine wichtige Rolle.

Poisson Verteilung

$$X \sim \text{Pois}(\lambda)$$

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

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

Bernoulliverteilung

Die Bernoulliverteilung Für $\Omega = \{0, 1\}$ und $p \in [0, 1]$ ist definiert durch

$$P(\omega) = p^\omega (1 - p)^{1-\omega}$$

- Werfen einer Münze: Kopf (Erfolg), $p = 1/2$, und Zahl (Misserfolg), $q = 1/2$.
- Werfen eines Würfels, wobei nur eine 6 als Erfolg gewertet wird: $p = 1/6$, $q = 5/6$.
- Qualitätsprüfung (einwandfrei, nicht einwandfrei).
- Betrachte sehr kleines Raum/Zeit-Intervall: Ereignis tritt ein ($p \gtrapprox 0$), tritt nicht ein ($q \lesssim 1$).

Binomialverteilung

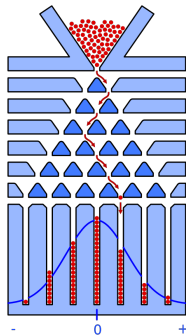


Figure: Quelle: Wikipedia

Erraten des Bereichs von Zufallsvariablen

Angenommen man findet einen Apparat (Fluxkompensator?), der zufällig Zahlen in einem Intervall $[0, \rho]$ ausgibt. Anhand von Beobachtungen der Zahlen möchte man ρ schätzen.



Figure: Quelle: forevergeek

Erraten des Bereichs von Zufallsvariablen - Modell

Wir machen die Annahme, dass alle Zahlen in dem Intervall gleich wahrscheinlich auftreten und nehmen n Stichproben X_1, \dots, X_n . Einen Schätzer für ρ bezeichnen wir mit T_n .

Erraten des Bereichs von Zufallsvariablen - Schätzer 1

Eine einfache und einleuchtende Idee ist es, ρ durch die größte beobachtete Zahl zu schätzen, also $T_n^{\max} := \max(X_1, \dots, X_n)$.

Erraten des Bereichs von Zufallsvariablen - Schätzer 2

Da das Auftreten der Zahlen gleich wahrscheinlich ist, ist der Erwartungswert des Zufallsexperiments $\rho/2$. Unter Berufung auf das schwache Gesetz der Großen Zahlen erscheint der Schätzer $T_n^E := 2 \cdot \left(\frac{1}{n} \sum_{i=1}^n X_i \right)$ sehr plausibel.

Erraten des Bereichs von Zufallsvariablen - Vergleich

Welcher Schätzer ist besser und in welchem Sinn?

Erraten des Bereichs von Zufallsvariablen - Vergleich Konvergenz

$$\begin{aligned} P(|T_n^{\max} - \rho| \geq \epsilon) &= P(T_n^{\max} \leq \rho - \epsilon) \\ &= P(X_1 \leq \rho - \epsilon, \dots, X_n \leq \rho - \epsilon) = \left(\frac{\rho - \epsilon}{\rho}\right)^n \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

Erraten des Bereichs von Zufallsvariablen - Vergleich Konvergenz

$$P(|T_n^E - \rho| \geq \epsilon) = P\left(\left|\frac{1}{n} \sum X_i - \frac{\rho}{2}\right| \geq \frac{\epsilon}{2}\right) \xrightarrow{n \rightarrow \infty} 0$$

(Gesetz der Großen Zahlen)

Erraten des Bereichs von Zufallsvariablen - Vergleich Erwartungswert

Da $P(T_n^{max} \leq c) = (\frac{c}{\rho})^n$ und $\frac{d}{dx}(\frac{c}{\rho})^n = \frac{n}{\rho^n} x^{n-1}$ und damit

$$\mathbb{E}(T_n^{max}) = \int_0^\rho x \frac{n}{\rho^n} x^{n-1} dx = \frac{n}{\rho^n} \int_0^\rho x^n dx = \frac{n}{n+1} \rho \xrightarrow{n \rightarrow \infty} \rho$$

Erraten des Bereichs von Zufallsvariablen - Vergleich Erwartungswert

$$\mathbb{E}(T_n^E) = \frac{2}{n} \sum \mathbb{E}(X_i) = \rho$$

Erraten des Bereichs von Zufallsvariablen - Vergleich Varianz

$$\mathbb{V}(T_n^{\max}) = \frac{4}{n} \mathbb{V}(X_1) = \frac{4}{n\rho} \int_0^\rho \left(x - \frac{\rho}{2}\right)^2 dx = \frac{\rho^2}{3n}$$

Erraten des Bereichs von Zufallsvariablen - Vergleich Varianz

$$\begin{aligned} \mathbb{V}(T_n^E) &= \mathbb{E}((T_n^E)^2) - (\mathbb{E}(T_n^E))^2 \\ &= \int_0^\rho x^2 \frac{n}{\rho^n} x^{n-1} dx - \left(\frac{n\rho}{n+1}\right)^2 = \frac{n\rho^2}{(n+1)^2(n+2)} \end{aligned}$$

Statistisches Modell

Ein statistisches Modell ist ein Tripel $(\mathcal{X}, \mathcal{F}, P_\rho : \rho \in \Theta)$ bestehend aus einer σ -Algebra \mathcal{F} über dem Grundraum \mathcal{X} und einer indizierten Menge (mit mindestens zwei Elementen) von Maßen $\{P_\rho\}_{\rho \in \Theta}$. Für $\Theta \subset \mathbb{R}^n$ bezeichnet man es auch als parametrisiertes statistisches Modell.

Statistik

Sei $(\mathcal{X}, \mathcal{F}, P_\rho : \rho \in \Theta)$ ein statistisches Modell und (Σ, \mathcal{G}) eine σ -Algebra über Σ . Eine Zufallsvariable

$$X : \mathcal{X} \rightarrow \Sigma$$

wird Statistik für $(\mathcal{X}, \mathcal{F}, P_\rho : \rho \in \Theta)$ genannt.

Schätzer

Sei $(\mathcal{X}, \mathcal{F}, P_\rho : \rho \in \Theta)$ ein statistisches Modell, (Σ, \mathcal{G}) eine σ -Algebra über Σ und

$$\tau : \Theta \rightarrow \Sigma$$

$$\rho \mapsto \tau(\rho)$$

eine Abbildung. Eine Statistik $T : \mathcal{X} \rightarrow \Sigma$ wird Schätzer für τ genannt.

Konsistenz

Eine Schätzfolge $T_n : \mathcal{X} \rightarrow \Sigma$ heißt konsistent, falls

$$P(|T_n - \tau(\rho)| \geq \epsilon) \xrightarrow{n \rightarrow \infty} 0 \text{ für alle } \epsilon > 0 \text{ und alle } \tau \in \theta$$

also $T_n \rightarrow \tau(\rho)$ für $n \rightarrow \infty$ (stochastisch).

Konsistenz

Ein Schätzer $T : \mathcal{X} \rightarrow \Sigma$ heißt Erwartungstreu (unbiased), falls

$$\mathbb{E}(T) = \tau(\rho) \text{ für alle } \rho \in \theta.$$

Andernfalls heißt $\mathbb{E}(T) - \tau(\rho)$ der Bias oder der systematische Fehler.

Stichprobenmittel und Stichprobenvarianz

Für $n \geq 2$ sei $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_\rho^n := \prod (P_\rho)_i)$ das Produktmodell, $X : \mathbb{R} \rightarrow \mathbb{R}$ eine Zufallsvariable, $X_i(x_1, \cdot, x_n) = x_i$ die Projektion auf die i -te Koordinate und $m(\rho) = \mathbb{E}(X)$ so wie $v(\rho) = \mathbb{V}(X)$. Des weiteren seien X so wie X_i unabhängig und identisch verteilt. Dann sind das Stichprobenmittel $M = \frac{1}{n} \sum_{i=1}^n X_i$ und die Stichprobenvarianz $V = \frac{1}{n-1} \sum_{i=1}^n (X_i - M)^2$ erwartungstreue Schätzer für m bzw. v .

Beweis

Sei $\rho \in \Theta$ fest. Wegen linearität des EW und u.i.v. ist $\mathbb{E}(V) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = m(\rho)$.

Beweis

Aus Linearität des EW und $\mathbb{E}(X_i - M) = 0$, u.i.v, stochastische Unabhängigkeit folgt

$$\begin{aligned}(n-1)\mathbb{E}(V) &= \sum_{i=1}^n \mathbb{V}(X_i - M) \\ &= n\mathbb{V}\left(\frac{n-1}{n}X_1 - \frac{1}{n}\sum_{j=2}^n X_j\right) \\ &= n\left(\left(\frac{n-1}{n}\right)^2 + (n-1)\frac{1}{n^2}\right)v(\rho) = (n-1)v(\rho)\end{aligned}$$

Durch Teilen beider Seiten durch $n-1$ folgt die Behauptung.

Likelihood Funktion

Für eine parametrisierte Dichte $p_\rho : \mathcal{X} \rightarrow [0, 1]$ heißt die Funktion

$$p : \mathcal{X} \times \Theta \rightarrow [0, 1]$$

$$p(x, \rho) := p_\rho(x)$$

die zugehörige Likelihood Funktion und

$$p_x : \Theta \rightarrow [0, 1]$$

$$p_x(\rho) := p(x, \rho)$$

die Likelihood Funktion zum Beobachtungswert $x \in \mathcal{X}$.

Im diskreten Fall setzen wir $p(x, \rho) := P_\rho(\{x\})$

Maximum Likelihood Schätzer

Ein Schätzer $T : \mathcal{X} \rightarrow \Sigma$ heißt Maximum-Likelihood-Schätzer, falls

$$p(x, T(x)) = \max_{\rho \in \Theta} p(x, \rho) \text{ für alle } x \in \mathcal{X}$$

also der Schätzwert $T(x)$ eine Maximalstelle der Funktion p_x auf Θ ist.

Erraten des Bereichs von Zufallsvariablen - Maximum Likelihood-Schätzer

Die Likelihood Funktion ist hier gegeben durch

$$p_x(\rho) = \begin{cases} \frac{1}{\rho^n} & \text{falls } x_1, \dots, x_n \leq \rho \\ 0 & \text{sonst} \end{cases}$$

Somit ist der Schätzer $T_n^{\max}(x) := \max(x_1, \dots, x_n)$ der Maximum-Likelihood-Schätzer.

Physikalische Messung - Maximum Likelyhood-Schätzer

Gegeben ist ein Sensor mit einer unbekannten Messgenauigkeit. Wir nehmen deshalb an, dass die Messung des Sensors einer Normalverteilung folgt, wobei der Erwartungswert (Messwert) und die Varianz (Streuung um Messwert) unbekannt sind. Wir machen n unabhängige Messungen und erhalten damit das Normalverteilte Produktmodell $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \prod_{i=1}^n N(m, v) : m \in \mathbb{R}, v > 0)$.

Physikalische Messung - Maximum Likelihood-Schätzer

Die Likelihood-Funktion ist gegeben durch

$$p_x(\rho) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi v}} e^{-\frac{(x_i - m)^2}{2v}} = \frac{1}{\sqrt{2\pi v}} e^{-\sum_{i=1}^n \left(\frac{(x_i - m)^2}{2v}\right)}$$

mit $x = (x_1, \dots, x_n)$ und $\rho = (m, v)$.

Physikalische Messung - Maximum Likelihood-Schätzer

Um diesen Ausdruck zu maximieren, muss m so gewählt werden, dass die quadratische Fehlersumme $\sum_{i=1}^n (x_i - m)^2$ minimal wird. Das ist der Fall für $m = M(x) = \frac{1}{n} \sum_{i=1}^n x_i$.

Physikalische Messung - Maximum Likelihood-Schätzer

Des weiteren muss v so gewählt werden, dass

$\frac{1}{\sqrt{2\pi v}} e^{-\sum_{i=1}^n \left(\frac{(x_i - M(x))^2}{2v} \right)}$ maximal wird. Differenziert man den Logarithmus dieses Ausdrucks nach v , erhalten wir

$$\begin{aligned} & -\frac{d}{dv} \left(\frac{n}{2} \log(v) + \frac{1}{2v} \sum_{i=1}^n (x_i - M(x))^2 \right) \\ &= -\frac{n}{2v} + \frac{1}{2v^2} \sum_{i=1}^n (x_i - M(x))^2 \end{aligned}$$

Dieser Term ist maximal, falls der letzte Term verschwindet. Dies ist er Fall für $v = V(x) = \frac{1}{n} \sum_{i=1}^n (x_i - M(x))^2$