Laplace-Experiment

$$P(E) = \frac{|E|}{|\Omega|} = \frac{\text{Anzahl günstige Ergebnisse}}{\text{Anzahl aller Ergebnisse}} \rightarrow \text{Gleichverteilung}$$

Produkte

$$|A_1 \times A_2 \times \dots \times A_k| = |A_1| * |A_2| * \dots * |A_k|$$

Summen

$$|A_1 \cup A_2 \cup \dots \cup A_k| = |A_1| + |A_2| + \dots + |A_k|$$

Fakultät

$$n! := n * (n-1) * \dots * 2 * 1 \sim \sqrt{2 * \pi * n} * (\frac{n}{e})^n$$

Binominalkoeffizient

$$\binom{n}{k} := \frac{n!}{(n-k)! * k!}$$

Urnenmodel

	zurücklegen	nicht zurücklegen
geordnet	n^k	$n!$ oder $\frac{n!}{(n-k)!}$
ungeordnet	$\binom{k+n-1}{k}$	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

algemeine Wahrscheinlichkeiten

$$P(E_1 \cup E_2 \cup E_3 \cup \dots) = P(E_1) + P(E_2) + P(E_3) + \dots$$
$$\forall E_1, E_2 \subseteq \Omega : P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$
$$E = \{e_1, e_2, \dots\} \rightarrow P(E) = P(\{e_1\}) + P(\{e_2\}) + \dots$$

Z-Dichte

Die Funktion $f_P : \Omega \rightarrow [0,1]$ mit $f_P(w) = P(\{w\})$ heisst Zähldichte von P.

$$P(E) = \sum_{e \in E} f_P(e)$$
$$P(E) = \sum_{e \in \Omega} f_P(e) = 1$$

bedingte Wahrscheinlichkeit

Es sei $B \subseteq \Omega$ mit $P(B) > 0$. Dann heisst $P(A|B)$ (elementare) bedingte Wahrscheibnlichkeit von A unter B.

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

Formel von Bayes

$$P(A|B) = \frac{P(A)}{P(B)} * P(B|A)$$

totele Wahrscheinlichkeit

Es sei $B_i (i \in I)$ eine Zerlegung von Ω (d.h. die B_i sind paarweise disjunkt und $\Omega = U_{i \in I} B_i$) mit $P(B_i) > 0$.

$$P(A) = \sum_{i \in I} P(A|B_i) * P(B_i)$$

positive prädiktive Wert

Für $0 < P(A) < 1$ gilt mit $\Omega = A \cup A^c$ insbesondere:

$$P(A|B) = \frac{P(A) * P(A|B)}{P(B|A) * P(A) + P(B|A^c) * P(A^c)}$$

stochastische Unabhängigkeit

Zwei Ereignisse $A, B \subseteq \Omega$ heissen stochastisch unabhängig, falls

$$P(A \cap B) = P(A) * P(B) \equiv \underbrace{P(A|B)}_{\frac{P(A \cap B)}{P(B)}} = P(A) \leftarrow (B) \neq 0]$$

Mehrstufige Zufallsexperimente

Gegeben ist eine Urne mit 4 weissen und 2 schwarzen Kugeln. Wir ziehen dreimal ohne Zurücklegen: Was ist die Warhscheinlichkeit A: ¨Dritte Kugel weiss¨?

$\Omega = \{w, s\} \times \{w, s\} \times \{w, s\}$ und $A = \{(w, w, w), (w, s, w), (s, w, w), (s, s, w)\}$

$$P(A) = f(w, w, w) + f(w, s, w) + f(s, w, w) + f(s, s, w)$$

Verteilungsfunktion

Es sei $X : \Omega \rightarrow X$ eine Zufallsvariable, wobei $X \in \mathbb{R}$ eine endliche oder abzählbare Menge ist. Zudem sei f die Zähldichte von X.

$$F(x) = P(X \leq x) = \sum_{t \in X : t \leq x} f(t)$$

Erwartungswert, Varianz & Standardabweichung

$$E(X) = \sum_{x \in X} x * f(x) = \sum_{x \in X} x * P(X = x)$$
$$V(X) = \sum_{x \in X} (x - E)^2 * f(x) = \sum_{x \in X} (x - E)^2 * P(X = x) = E(X^2) - E(X)^2$$
$$\sigma(X) = \sqrt{V(X)}$$

Bernoulli-Verteilung

kann zwei Werte annehmen, 0 (Treffer) oder 1 (kein Treffer)

$$X \sim B(p) : P(0) = 1 - p$$
$$P(1) = p$$
$$E(X) = p$$
$$V(X) = p * (1 - p)$$

Binomial-Verteilung

Anzahl Treffer (Wahrscheinlichkeit p) in n unabhängigen Versuchen

$$X \sim Bin(n,p) : P(X = k) = \binom{n}{k} * p^k * (1-p)^{n-k}, k = 0, 1, \ldots, n$$
$$E(X) = n * p$$
$$V(X) = n * p * (1 - p)$$

Dichte: binopdf(k, n, p) | **Verteilungsfunktion**: binocdf(k, n, p)

geometirsche-Verteilung

Versuche bis erster Treffer (Wahrscheinlichkeit p) in n unabhängigen Veruschen

$$X \sim Geo(n,p) : P(X = k) = (1-p)^{k-1} * p, k = 0, 1, \ldots, n$$
$$E(X) = \frac{1}{p}$$
$$V(X) = \frac{1-p}{p^2}$$

Dichte: geopdf($k-1$, p) | **Verteilungsfunktion**: geocdf($k-1$, p)

Poisson-Verteilung

Verteilung für seltene Ereignisse mit im Schnitt λ Ereignisse pro Zeit/Ort

$$X \sim Poi(\lambda) : P(X = k) = \frac{\lambda^k}{k!} * e^{-\lambda}, k = 0, 1, \ldots, n$$
$$E(X) = \lambda$$
$$V(X) = \lambda$$

Dichte: poisspdf(k, λ) | **Verteilungsfunktion**: poisscdf(k, λ)

stetige Zufallsvariable

$\rightarrow$ kontinuierlicher Wertebereich
$\rightarrow$ beschreibt Wahrscheinlichkeit das ZV Wert in Bereich annimmt.
$\rightarrow$ Wahrscheinlichkeit als Dichte ausgedrückt.

$$Dichtefunktion : P(a \leq X \leq b) = \int_a^b f(x)dx$$
$$Erwartungswert : E(X) = \int_{-\infty}^{\infty} x * f(x)dx$$
$$Varianz : V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 * f(x)dx = E(X^2) - E(X)^2$$
$$Standartabweichung : \sigma(X) = \sqrt{V(X)}$$
$$Verteilungsfunktion : F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

stetige Gleichverteilung-Verteilung

die konstante Dichte für $s \leq x \leq t$ auf dem Intervall$[s, t]$.

$$X \sim U[s,t] : f(x) = \frac{1}{t-s}$$
$$E(x) = \frac{s+t}{2}$$
$$V(x) = \frac{1}{12} * (t-s)^2$$

Standardormalverteilung

$$X \sim N(0,1) : \varphi(x) = \frac{1}{\sqrt{2\pi}} * e^{-\frac{x^2}{2}}$$
$$E(x) = 0$$
$$V(x) = 1$$

Dichte: normpdf(x) | **Verteilungsfunktion**: normcdf(x)

Normalverteilung

Die Normalverteilung mit Parametern μ und σ.

$$X \sim N(\mu,\sigma) : \varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
$$E(x) = \mu$$
$$V(x) = \sigma$$

Dichte: normpdf(x, μ, σ) | **Verteilungsfunktion**: normcdf(x, μ, σ)

Standardisierung: $X \sim M(\mu, \sigma) \rightarrow X - \mu \sim N(0, \sigma) \rightarrow Z \sim N(0, 1)$ mit $Z = \frac{x-\mu}{\sigma}$
Quantil: Gegeben $\alpha \in (0, 1)$. Für welche z_α gilt P$(x \leq z_\alpha = \alpha)$
Matlab: $x =$ norminv(p)

Exponentialverteilung

Die Wahrscheinlichkeit dass X einen Wert grösser als t annimt, sinkt exponentiell.

$$X \sim Poi(\lambda) : P(X = t) = \frac{\lambda^k}{k!} * e^{-\lambda}, \text{ k=0,1, ..., n}$$
$$E(x) = \frac{1}{\lambda}$$
$$V(x) = \frac{1}{\lambda^2}$$

Dichte: exppdf(k, $\frac{1}{\lambda}$) | **Verteilungsfunktion**: expcdf(k, $\frac{1}{\lambda}$)

Eigenschaften

$$E(X+Y) = E(X) + E(Y)$$
$$E(a * X) = a * E(X)$$
$$E(X + c) = E(X) + c$$
$$V(X + c) = V(X)$$
$$V(a * X) = a^2 * V(X)$$
$$E(g(X)) = \sum_x g(x) * P(X = x) \leftarrow [\mathbb{R} \rightarrow \mathbb{R}]$$
$$E(g(X)) \neq g(E(X))$$

Unabhängigkeit

$$P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) = P(X_1 \leq x_1) * P(X_2 \leq x_2), \ldots, P(X_n \leq x_n)$$
$$P(g(X,Y) = z) = \sum_{x \in X} \sum_{y \in \Upsilon : g(x,y) = z} f_X(x) * f_Y(y)$$
$$X \sim Poi(\lambda_1), Y \sim Poi(\lambda_2) \Rightarrow X + Y \sim Poi(\lambda_1 + \lambda_2)$$
$$X \sim Bin(n_1, p), Y \sim Bin(n_2, p) \Rightarrow X + Y \sim Bin(n_1 + n_2, p)$$

Additionstheorem der Normalverteilung

$$a_1 X_1 + a_2 X_2 + \cdots + a_n X_n \sim N(a_1 \mu_1 + a_2 \mu_2 + \cdots + a_n \mu_n, \sqrt{a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2})$$

Eigenschaft von Erwartungswert und Varianz

Es seien X und Y Zufallsvariablen und $a, c \in \mathbb{R}$. Dan gilt:

$$E(X+Y) = E(X) + E(Y)$$
$$E(aX) = aE(X)$$
$$E(X + c) = E(X) + c$$
$$V(X + c) = V(X)$$
$$V(aX) = a^2 V(X)$$

Falls X und Y unabhängig sind: $V(X+Y) = V(X) + V(Y)$

Wenn X diskret ist und $f_X$die Zählichte von X: $E(g(X)) = \sum_x g(x) * f_X(x)$

Wenn X stetig ist und f_X ide Dichte von X: $E(g(X)) = \int_{-\infty}^{\infty} g(x) * f_X(x) dx$

Ungleichung von Tschabyscheff

$$P(\mid X - \mu \mid \geq k) \leq \frac{\sigma^2}{k^2}$$

Diese Abschätzung gilt für alle möglichen Verteilungen von X. Sie ist deshalb in manchen Fällen recht grob.

Grentzwertsätze: Gesetz der Grossen Zahlen

$$P(\mid \frac{X}{n} - p \mid \geq \epsilon) \rightarrow 0 (n \rightarrow \infty)$$

Zentraler Grenzwertsatz

Es sei $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsvariablen eines Wahrscheinlichkeitsraumes, welche alle dieselbe Verteilung mit Erwartungswert μ und Varianz σ^2 haben. Dan gilt für grosse n:

Die Summe $S_n = X_1 + \cdots + X_n$ besitzt näherungsweise die Normalverteilung $N(\mu_n, \sigma_n)$ mit $\mu_n = n * \mu$ und $\sigma_n = \sqrt{n} * \sigma$. Es gilt also näherungsweise $\frac{S_n - \mu_n}{\sqrt{n * \mu}} \sim N(0,1)$.

Präzise gilt für alle $z \in \mathbb{R}$:

$$P(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq z) \rightarrow \Phi(z)(n \rightarrow \infty)$$

Grenzwertsatz von de Moivre und Laplace

Für grosse n ist

$$P(a \leq X \leq b) \approx \Phi(\frac{b - np}{\sqrt{np(1-p)}}) - \Phi(\frac{a - np}{\sqrt{np(1-p)}})$$

Diese Approximation ist für $n > \frac{9}{p(1-p)}$ hinfreichend genau.

Etwas genauer wird es mit der sogenannten **Stetigkeitskorrektur**:

$$P(a \leq X \leq b) \approx \Phi(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}) - \Phi(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}})$$

Statistische Tests

1. Man stellt die Nullhypothese H_0 und die Alternativhypothese H_1 auf. Das bestmögliche Ergebnis ist die Widerlegung von H_0
2. Man wählt ein Testverfahren zb Experimente oder Stichproben
3. Aufgrund des gewählten Tests bestimmt man die so genannte Testgrösse
4. Man wählt ein Signifikanzniveau α und bestimmt aufgrund dessen den Verwerfungsbereich. Dieser wird so festgelegt, dass bei Zutreffen von H_0 die Testgrösse mit Wahrscheinlichkeit $\leq \alpha$ dort drin liegt.
5. Liegt der gemäss (3) berechnete Wert der Testgrösse im Verwerfungsbereich, so lehnen wir H_0 ab. Man sagt in diesem Fall, auch, das Ergebniss des Tests sei signifikant auf dem α-niveau

Konfidenzintervall

Wenn die Zufallsvariable X die Anzahl an die Personen zählt, die mit ja antworten, dann ist $X \sim Bin(n,p)$, mit$n = 500$ in unserem Beispiel.
Bekanntlich gilt approximativ: $\frac{X-np}{\sqrt{np(1-p)}} \sim N(0,1)$.
Nun sei $\alpha = \frac{1+Q}{2}$ und z_a das α-Quantil von $N(0,1)$.

Die Wahrscheinlichkeit, dass das Intervall den Wert p enthält, also gerade Q.

$$\left[\frac{1}{n+z_\alpha^2}\left(X+\frac{z_\alpha^2}{2}-z_\alpha\sqrt{\frac{X(n-X)}{n}+\frac{z_\alpha^2}{4}}\right), \frac{1}{n+z_\alpha^2}\left(X+\frac{z_\alpha^2}{2}+z_\alpha\sqrt{\frac{X(n-X)}{n}+\frac{z_\alpha^2}{4}}\right)\right]$$

Approximativ:

$$\left[\frac{k}{n}-\frac{z_\alpha}{n}\sqrt{\frac{k(n-k)}{n}}, \frac{k}{n}+\frac{z_\alpha}{n}\sqrt{\frac{k(n-k)}{n}}\right]$$

---

by Jan Fässler