Binomialverteilung

Konstruktionsprinzip:

- Ein Zufallsexperiment wird n mal unabhängig durchgeführt.
- Wir interessieren uns jeweils nur, ob ein bestimmtes Ereignis A eintritt oder nicht.
- X = „Häufigkeit, mit der Ereignis A bei n unabhängigen Versuchen eintritt“.
- Träger von X: $\mathcal{X} = \{0, 1, 2, \ldots, n\}$.

Wahrscheinlichkeitsfunktion

- Bezeichne $\pi = P(A)$ die Wahrscheinlichkeit für A in *einem* Experiment.
- Das Ereignis $X = x$ tritt z.B. auf, wenn in den ersten x Versuchen A eintritt und anschließend nicht mehr. Die Wahrscheinlichkeit dafür ist

$$\begin{aligned} P(A_1 \cap \ldots \cap A_x \cap \bar{A}_{x+1} \cap \ldots \cap \bar{A}_n) &= \underbrace{\pi \cdot \ldots \cdot \pi}_{x \text{ mal}} \underbrace{(1-\pi) \cdot \ldots \cdot (1-\pi)}_{n-x \text{ mal}} \\ &= \pi^x (1-\pi)^{n-x}. \end{aligned}$$

- Insgesamt gibt es $\binom{n}{x}$ Möglichkeiten für die Verteilung der x Erfolge (Auftreten von A) auf n Plätze. Damit gilt:

$$P(X = x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}.$$

Binomialverteilung

Definition

Eine Zufallsvariable heißt binomialverteilt mit den Parametern n und π, kurz $X \sim B(n, \pi)$, wenn sie die Wahrscheinlichkeitsfunktion

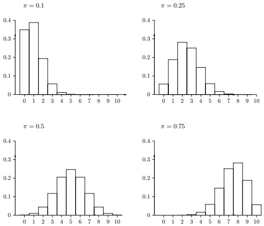
$$f(x) = \begin{cases} \binom{n}{x} \pi^x (1-\pi)^{n-x}, & x = 0, 1, \ldots, n \\ 0, & \textit{sonst} \end{cases}$$

besitzt.

Die $B(1, \pi)$-Verteilung heißt auch Bernoulliverteilung.

Wahrscheinlichkeitshistogramme

Binomialverteilungen mit $n = 10$



Erwartungswert und Varianz I

- Zur Berechnung von Erwartungswert und Varianz der Binomialverteilung ist folgende Darstellung hilfreich:

$$X = X_1 + \ldots + X_n$$

mit den binären Variablen

$$X_i = \begin{cases} 1 & \text{falls } A \text{ beim } i\text{-ten Versuch eintritt,} \\ 0 & \text{sonst.} \end{cases}$$

- Die X_i sind stochastisch unabhängig mit

$$\begin{aligned} \mathbb{E}(X_i) &= 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) = \pi \\ \mathrm{Var}(X_i) &= \mathbb{E}(X_i^2) - (\mathbb{E}(X_i))^2 = 1 \cdot P(X_i = 1) - \pi^2 = \pi(1 - \pi). \end{aligned}$$

Erwartungswert und Varianz I

- Erwartungswert der Binomialverteilung:

$$\mathbb{E}(X) = \mathbb{E}(X_1 + \ldots + X_n) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n) = n\pi$$

Die direkte Berechnung über

$$\mathbb{E}(X) = \sum_{i=1}^{n} i \binom{n}{i} \pi^i (1 - \pi)^{n-i} = \ldots = n\pi$$

ist deutlich komplizierter!

- Varianz der Binomialverteilung:

$$\mathrm{Var}(X) = \mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n) = n\pi(1 - \pi)$$

Beispiel: Wahlprognose

- 100 Wahlberechtigte werden befragt.
- 30% aller Wahlberechtigten wählen Partei S.

→ Wie groß ist die Wahrscheinlichkeit, dass von den 100 Befragten mehr als 50 die Partei S wählen?

$X \sim B(100, 0.3)$

$$\begin{aligned} P(X \geq 50) &= P(X = 50) + P(X = 51) + \ldots + P(X = 100) \\ &= \binom{100}{50} \cdot 0.3^{50} \cdot 0.7^{50} + \ldots \\ &= 0.0000206 \end{aligned}$$

Eigenschaften der Binomialverteilung

- Symmetrieeigenschaft (vertausche Rolle von A und $\bar{A}$):
  Sei $X \sim B(n, \pi)$ und $Y = n - X$. Dann gilt $Y \sim B(n, 1 - \pi)$.
- Summeneigenschaft: Seien $X \sim B(n, \pi)$ und $Y \sim B(m, \pi)$. Sind X und Y unabhängig, so gilt

$$X + Y \sim B(n + m, \pi)$$

Wichtige Voraussetzung: Gleiches π

Poisson Verteilung

(vgl. z.B. Fahrmeir et. al))
Eine weitere wichtige diskrete Verteilung ist die Poisson-Verteilung. Sie modelliert die Anzahl (eher seltener) Ereignisse in einem Zeitintervall (Unfälle, Todesfälle; Sozialkontakte, deviante Verhaltensmuster, etc.).

Definition

Eine Zufallsvariable X mit der Wahrscheinlichkeitsfunktion

$$f(x) = P(X = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda}, & x \in \{0, 1, \ldots\} \\ 0, & sonst \end{cases}$$

heißt Poisson-verteilt mit Parameter (oder Rate) $\lambda > 0$, kurz $X \sim Po(\lambda)$. Es gilt

$$\mathbb{E}(X) = \lambda, \quad \mathrm{Var}(X) = \lambda$$

Poisson Modell als Verteilung seltener Ereignisse

Die Poisson-Verteilung kann auch als Näherungsmodell für eine Binomialverteilung gesehen werden, wenn die Anzahl der Versuchswiederholungen n groß und die „Trefferwahrscheinlichkeit“ π sehr klein ist (seltene Ereignisse!).
Der Erwartungswert λ ist dann gleich $n \cdot \pi$.
Es gilt also abgekürzt geschrieben

$$X \sim B(n, \pi) \underset{\substack{n \text{ groß} \\ \pi \text{ klein}}}{\Longrightarrow} X \approxeq Po(n \cdot \pi)$$

Hat man mehrere unabhängige „Poisson-Prozesse“, also dynamische Simulationen, bei denen die Ereignisanzahl Poisson-verteilt ist, also z.B. verschiedene deviante Verhaltensmuster, so ist die Gesamtanzahl der einzelnen Ereignisanzahlen wieder Poisson-verteilt: genauer gilt

Addition von Poisson-verteilten ZV

Sind $X \sim Po(\lambda_X), Y \sim Po(\lambda_Y)$ voneinander unabhängig, so gilt

$$X + Y \sim Po(\lambda_X + \lambda_Y).$$

Beachte, die Unabhängigkeit (genauer die Unkorreliertheit, siehe später) ist wesentlich. Hat man als Extremfall, z.B. zwei Ereignisse bei denen das eine das andere voraussetzt (Scheidungen, Scheidungen mit Streit um das Sorgerecht für Kinder), so ist die Gesamtzahl nicht mehr Poisson-verteilt. Es muss gelten, wenn $X + Y$ Poisson-verteilt wäre:

$$\mathrm{Var}(X + Y) = \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = \mathrm{Var}(X) + \mathrm{Var}(Y),$$

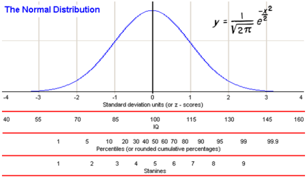
was aber bei abhängigen (korrelierten) X und Y verletzt ist.

Normalverteilung I

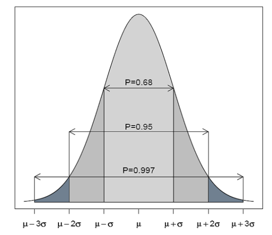
Die Normalverteilung ist wohl das wichtigste Verteilungsmodell der Statistik, denn

- viele Zufallsvariablen sind (nach Transformation) (ungefähr) normalverteilt.
- beim Zusammenwirken vieler zufälliger Einflüsse ist der geeignet aggregierte Gesamteffekt oft approximativ normalverteilt (Zentraler Grenzwertsatz, Kap. 1.7).
- die asymptotische Grenzverteilung, also die Verteilung bei unendlich großem Stichprobenumfang, typischer statistischer Größen ist die Normalverteilung.

Normalverteilung II



Normalverteilung III



Definition Normalverteilung

Definition

Eine stetige Zufallsvariable X heißt normalverteilt mit den Parametern μ und σ^2, in Zeichen $X \sim N(\mu, \sigma^2)$, wenn für ihre Dichte gilt:

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

und standardnormalverteilt, in Zeichen $X \sim N(0;1)$, falls $\mu = 0$ und $\sigma^2 = 1$ gilt (π ist hier die Kreiszahl $\pi = 3.14\ldots$).

Grundlegende Eigenschaften

a) Die *Dichte der Standardnormalverteilung* wird oft mit $\varphi(x)$ bezeichnet, also

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

und die zugehörige Verteilungsfunktion mit

$$\Phi(x) = \int_{-\infty}^{x} \varphi(u)du.$$

b) $\Phi(x)$ lässt sich nicht in geschlossener Form durch bekannte Funktionen beschreiben $\Longrightarrow$ numerische Berechnung, Tabellierung.

c) μ und σ^2 sind genau der Erwartungswert und die Varianz, also, wenn $X \sim N\mu, \sigma^2)$, dann

$$\mathbb{E}(X) = \mu \text{ und } \mathrm{Var}(X) = \sigma^2.$$

d) Die Dichte ist symmetrisch um μ, d.h. $f(\mu - x) = f(\mu + x)$.

Rechnen mit der Normalverteilung I

- Es gilt:
$$\Phi(-x) = 1 - \Phi(x)$$
(folgt aus der Symmetrie der Dichte).
- Tabelliert sind die Werte der Verteilungsfunktion $\Phi(z) = P(Z \leq z)$ für $z \geq 0$.
Ablesebeispiel: $\Phi(1.75) = 0.9599$
- Funktionswerte für negative Argumente: $\Phi(-z) = 1 - \Phi(z)$
- Die z-Quantile ergeben sich über die Umkehrfunktion.
Beispielsweise ist $z_{0.9599} = 1.75$ und $z_{0.9750} = 1.96$.

Rechnen mit der Normalverteilung II

Zentrale Idee: X zu standardnormalverteilter Zufallsvariable umformen, d.h. standardisieren. Dabei muss die rechte Seite analog mit transformiert werden:

$$\begin{aligned} X \leq a \quad &\Leftrightarrow \quad X - \mu \leq a - \mu \\ &\Leftrightarrow \quad \frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma} \end{aligned}$$

das heißt

$$P(X \leq a) = P(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}).$$

Wegen

$$\frac{X-\mu}{\sigma} \sim \mathrm{N}(0;1)$$

gilt dann

$$P(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}) = \Phi(\frac{a-\mu}{\sigma}),$$

so dass sich der folgende Zusammenhang ergibt:

$$P(X \leq a) = \Phi(\frac{a-\mu}{\sigma}).$$

Beispiel: IQ

Der IQ st so konstruiert, dass er in der Bevölkerung normalverteilt ist mit Erwartungswert 100 und Standardabweichung 15.

- $P(IQ \leq 110) = ?$
- $P(IQ \leq 70) =$
- Wie groß ist q, damit gilt $P(IQ \geq q) = 0.01$?

Abgeschlossenheit gegenüber Linearkombinationen

Seien X_1 und X_2 unabhängig und $X_i \sim N(\mu_i, \sigma_i^2),\ i = 1, 2$. Ferner seien b, a_1, a_2 feste reelle Zahlen. Dann gilt

$$\begin{aligned} Y_1 &:= a_1 X_1 + b \sim N(a_1\mu_1 + b; a_1^2\sigma_1^2) \\ Y_2 &:= a_1 X_1 + a_2 X_2 \sim N(a_1\mu_1 + a_2\mu_2; a_1^2\sigma_1^2 + a_2^2\sigma_2^2). \end{aligned}$$

Das Ergebnis lässt sich auf mehrere Summanden verallgemeinern.