§ 68: WICHTIGE KONTINUIERLICHE VERTEILUNGEN

68.1. Motivation

Zufallsvariablen sind nicht immer diskret, sie können oft auch jede beliebige reelle Zahl in einem Intervall $[c, d]$ annehmen. Beispiele für solche „kontinuierlichen" Zufallsvariablen sind Größe, Gewicht oder Zeit.

In diesen Fällen macht es wenig Sinn, die Ws. anzugeben, dass die Zufallsvariable einen bestimmten Wert annimmt (diese Ws. ist 0). Wir müssen Ws. für Intervalle betrachten. Hierzu sind Begriffe wie Dichten notwendig.

68.2. Def.: Sei X eine kontinuierliche Zufallsvariable. Existiert eine Funktion $f: \mathbb{R} \to \mathbb{R}$ mit

a) $f(x) \geq 0 \quad \forall x \in \mathbb{R}$

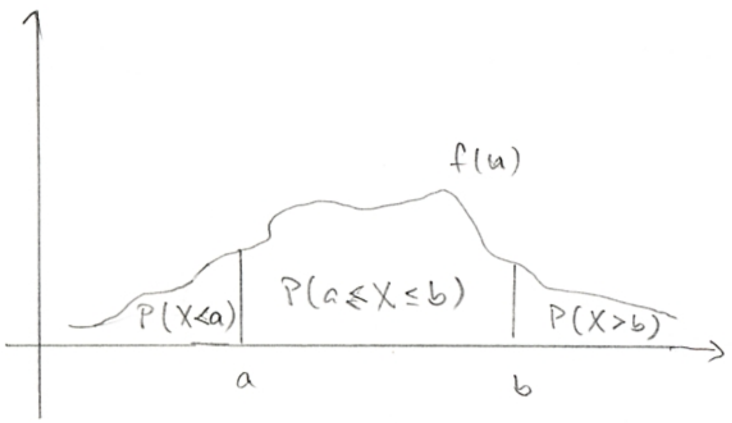
b) $\int_{-\infty}^{\infty} f(x)\, dx = 1$

und

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx,$$

so nennt man f die <u>Dichte</u> von X.

68.3. Veranschaulichung



Erwartungswert und Varianz von X sind definiert durch

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx, \qquad \sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

68.4. Def.: Sei X eine kontinuierliche Zufallsvariable mit Dichte f. Dann nennt man ihre Stammfunktion

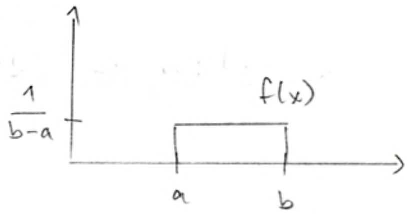
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\, du$$

die <u>Verteilungsfunktion</u> von X.

68.5. <u>Beispiel: Kontinuierliche Gleichverteilung</u>

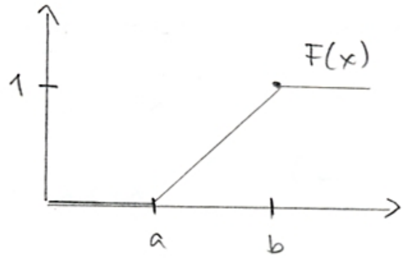
Eine kontinuierliche Zufallsvariable, die auf $[a,b]$ gleich verteilt ist, hat die Dichte

$$f(x) = \begin{cases} \frac{1}{b-a} & (a \leq x \leq b) \\ 0 & (\text{sonst}) \end{cases}$$



und die Verteilungsfunktion

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



Wir kommen nun zur wichtigsten kontinuierlichen Verteilung:

68.6. <u>Die Standardnormalverteilung</u>

Ist X eine kontinuierliche Zufallsvariable mit der Dichte

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

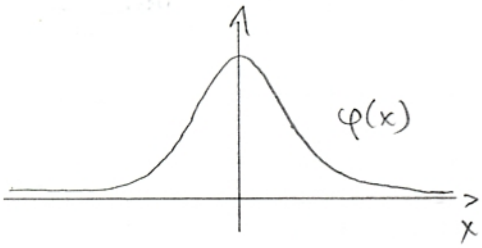
So kann man $P(a < X \leq b)$ mit der Stammfunktion

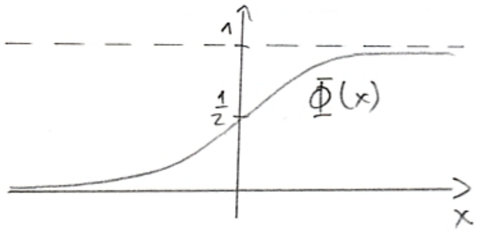
$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$$

berechnen:

$$P(a < X \leq b) = \Phi(b) - \Phi(a)$$

φ nennt man <u>normale Dichte</u>, und Φ ist die <u>Standardnormalverteilung</u> (<u>$N(0,1)$-Verteilung</u>).



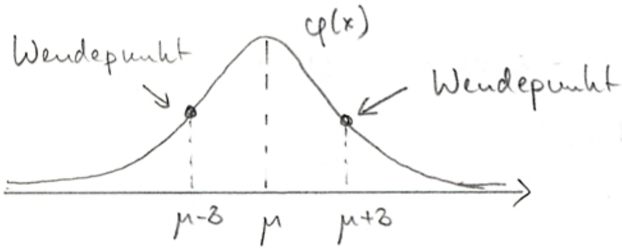


$\Phi(x)$ ist nicht analytisch auswertbar, liegt aber tabelliert vor. Eine standartnormalverteilte Zufallsvariable hat Erwartungswert 0 und Varianz 1. Daher heißt sie $N(0,1)$-verteilt.

68.7. Die allgemeine Normalverteilung

Eine kontinuierliche Zufallsvariable X genügt einer allgemeinen Normalverteilung ($N(\mu,\sigma^2)$-Verteilung, Gauß-Verteilung) mit EW μ und Varianz σ^2, wenn ihre Dichte gegeben ist durch

$$\boxed{\varphi(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$



Es gilt:

$$P(\mu-\sigma \le X \le \mu+\sigma) \approx 68\,\%$$
$$P(\mu-2\sigma \le X \le \mu+2\sigma) \approx 95{,}5\,\%$$
$$P(\mu-3\sigma \le X \le \mu+3\sigma) \approx 99{,}7\,\%$$

Ist X $N(\mu,\sigma^2)$-verteilt, so ist $Y := \frac{X-\mu}{\sigma}$ $N(0,1)$-verteilt.
Somit ist die Tabelle der Standardnormalverteilung ausreichend.

68.8. Approximation der Binomialverteilung durch die Gaußverteilung

Eine Binomialverteilung mit n Einzelexperimenten mit Ws. p kann man durch eine allgem. Normalverteilung mit Erwartungswert np und Standardabweichung $\sigma = \sqrt{np(1-p)}$ approximieren:

$$b_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \underbrace{\Phi\left(\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)}_{\text{Fläche über dem Intervall } [k-\frac{1}{2}, k+\frac{1}{2}]}$$

Diese Approximation ist gut für $np > 5$ und $n(1-p) > 5$, d.h. insbesondere für große n oder $p \approx 0{,}5$.

68.9. Beispiel

Wie groß ist die Ws., dass in 6000 Würfen eines fairen Würfels die Sechs mindestens 1100 Mal auftritt?

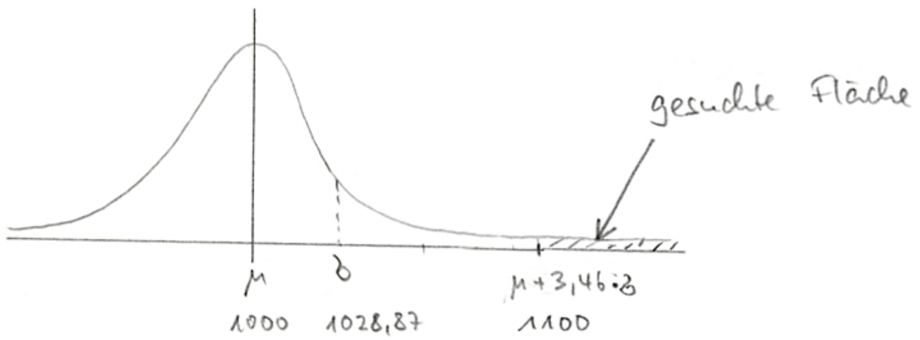
$n = 6000$, $p = \frac{1}{6}$

Wegen $np = 1000 > 5$ und $n(1-p) = 5000 > 5$ ist die Approximation durch die Gaußverteilung sinnvoll:

$\mu = n \cdot p = 1000$

$\sigma = \sqrt{np(1-p)} = \sqrt{6000 \cdot \frac{1}{6} \cdot \frac{5}{6}} \approx 28{,}87$

$1100 = \underbrace{1000}_{\mu} + 3{,}46 \cdot \underbrace{28{,}87}_{\sigma}$



Die Ws., mehr als 1100 Sechsen zu würfeln, beträgt

$$1 - \underbrace{\Phi(3{,}46)}_{\text{in Tabelle nachschlagen}} \approx 0{,}00028 .$$

Der wichtigste Grund für die Bedeutung der Gaußverteilung ist der folgende Satz:

68.10. Satz (Zentraler Grenzwertsatz)

Seien $X_1, X_2, \ldots$ unabhängige Zufallsvariablen, die alle die gleiche Verteilung und somit auch den selben Erwartungswert μ und die selbe Varianz σ^2 besitzen. Ferner sei $Y_n := X_1 + \ldots + X_n$ und

$$Z_n := \frac{Y_n - n\mu}{\sigma\sqrt{n}}$$

bezeichne die Standardisierte von Y_n (vgl. 65.15). Dann konvergiert die Verteilungsfunktion $F_n(x)$ von Z_n für $n \to \infty$ gegen die Standardnormalverteilung $\Phi(x)$.

68.11. Bemerkungen

b) Beachte, dass die einzelnen Zufallsvariablen nicht normalverteilt sein müssen. Ihre Verteilung kann beliebig sein!

a) Der Beweis von Satz 68.10 ist aufwändig. Siehe z.B. R. Nelson: Probability, Stochastic Processes and Queueing Theory, Springer, New York, 1995. Abschnitt 5.5.6.

c) Die Normalverteilung ist also eine sinnvolle Approximation in allen Fällen, in denen sich eine Zufallsvariable aus vielen gleichartigen Einzeleinflüssen zusammensetzt.
Beispiel: Eine Messung wird oft wiederholt. Dann approximieren die Ergebnisse eine Gaußverteilung.