

§68: WICHTIGE KONTINUIERLICHE VERTEILUNGEN

68.1 Motivation

Zufallsvariablen sind nicht immer diskret, sie können oft auch jede beliebige reelle Zahl in einem Intervall annehmen. Beispiele für solche „kontinuierlichen“ Zufallsvariablen sind Größe, Gewicht und Zeit. In diesen Fällen macht es wenig Sinn, die Wahrscheinlichkeit anzugeben, dass eine Zufallsvariable einen bestimmten Wert annimmt (diese Wahrscheinlichkeit ist 0). Wir müssen Wahrscheinlichkeiten für Intervalle betrachten. Hierzu sind Begriffe wie Dichten notwendig.

68.2 Def.:

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

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

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

und

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

so nennt man $f(x)$ die Dichte von X .

68.3 Veranschaulichung

(siehe Vorlesung im Internet)

68.4 Def.:

Sei X eine kontinuierliche Zufallsvariable mit Dichte f . Dann nennt man die Stammfunktion

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

die Verteilungsfunktion von X .

68.5 Beispiel: Kontinuierliche Gleichverteilung

Eine kontinuierliche Zufallsvariable mit Gleichverteilung auf $[a, b]$ 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 Die Standardnormalverteilung

Ist X eine kontinuierliche Zufallsvariable mit der Dichte

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

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

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

berechnen:

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

φ nennt man normale Dichte, und Φ ist die Standardnormalverteilung.

Φ ist nicht analytisch auswertbar, liegt aber tabelliert vor. Eine standardnormalverteilte Zufallsvariable hat Erwartungswert 0 und Varianz 1. Daher heißt sie auch $N(0, 1)$ -verteilt.

68.7 Die allgemeine Normalverteilung

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

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

Es gilt:

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &\approx 68\% \\ P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &\approx 95.5\% \\ P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &\approx 99.7\% \end{aligned}$$

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

zu 68.2: Erwartungswert und Varianz von X sind definiert durch

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

68.8 Approximation der Binomialverteilung durch die Gaußverteilung

Eine Binomialverteilung mit n Einzelexperimenten mit Wahrscheinlichkeit p kann

man durch eine Gaußverteilung mit Erwartungswert np und Standardabweichung $\sqrt{np(1-p)}$ 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 Wahrscheinlichkeit, dass in 6000 Würfeln eines fairen Würfels die Sechs mindestens 1100 Mal auftritt?

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

Wegen $n \cdot p = 1000 > 5$ und $n(1-p) = 5000 > 5$ ist die Gaußapproximation sinnvoll.

$$\mu = np = 6000 \cdot \frac{1}{6} = 1000, \quad \sigma = \sqrt{np(1-p)} = \sqrt{6000 \cdot \frac{1}{6} \cdot \frac{5}{6}} \approx 28.87$$

$$1100 = \underbrace{1000}_{\mu} + 3.46 \cdot \underbrace{28.87}_{\sigma}$$

Die Wahrscheinlichkeit, mehr als 1100 Sechsen zu würfeln, ist

$$1 - \underbrace{\Phi(3.46)}_{\text{in Tabelle nachschlagen}} \approx 0.00028$$

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

68.10 Satz: (Zentraler Grenzwertsatz)

Seien $X_1, X_2, \dots$ unabhängige Zufallsvariablen, die alle die gleiche Verteilung und somit alle den gleichen Erwartungswert μ und die gleiche Varianz σ^2 besitzen.

Ferner sei $Y_n := X_1 + \dots + X_n$ und $Z_n := \frac{Y_n - n\mu}{\sqrt{n}\sigma}$ bezeichne die Standardisierte von Y_n (vgl. 65.15).

Dann konvergiert die Verteilungsfunktion $F_n(x)$ von Z_n für $n \rightarrow \infty$ gegen die Standardnormalverteilung $\Phi(x)$.

68.11 Bemerkungen

- (a) Der Beweis von Satz 68.10 ist relativ aufwendig. Siehe z.B. R. Nelson: Probability, Stochastic Processes and Queueing Theory. Springer, New York, 1995
- (b) Beachte, dass die einzelnen Zufallsvariablen X_i nicht normalverteilt sein müssen. Ihre Verteilung kann beliebig sein!

- (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.