

Wahrscheinlichkeitsrechnung und Statistik für Studierende der Informatik



Vorlesung 11

Stetige Zufallsvariablen

Definition 3.1. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Eine Funktion $X : \Omega \rightarrow \mathbb{R}$ heißt *stetige Zufallsvariable*, falls es eine integrierbare, nicht negative reelle Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

gibt mit der Eigenschaft

$$P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

$$\implies \int_{-\infty}^{\infty} f(t) dt = 1$$

Die Funktion

$$\begin{aligned} F : \mathbb{R} &\rightarrow [0, 1] \\ x &\mapsto P(X \leq x) = \int_{-\infty}^x f(t) dt \end{aligned}$$

heißt *Verteilungsfunktion* von X , die Funktion f heißt *Dichte* der Zufallsvariablen X .

Falls f im Punkt x stetig ist, gilt $F'(x) = f(x)$.

Stetige Zufallsvariablen

Definition 3.6. (a) Ist X eine stetige Zufallsvariable mit der Dichtefunktion f , so heißt

$$E(X) := \int_{-\infty}^{\infty} x f(x) dx$$

der *Erwartungswert* von X .

(b) Ist X eine stetige Zufallsvariable mit der Dichtefunktion f derart, dass $E(X^2)$ existiert, so definiert man die *Varianz* durch

$$\text{Var}(X) := \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx.$$

$\sigma = \sigma_X = \sqrt{\text{Var}(X)}$ heißt *Standardabweichung* von X .

**Alles was Sie über $E(X)$ und $\text{Var}(X)$ wissen wenn X diskret ist,
gilt auch wenn X stetig ist!**

$$E(X_1 + X_2) = E(X_1) + E(X_2), \quad \text{Var}(X) = E(X^2) - E(X)^2, \quad \text{Tschebyscheff, ...}$$

Normal-Verteilung

Die Funktion $f: \mathbb{R} \rightarrow \mathbb{R}^+$ mit

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

ist eine Dichtefunktion.

Es gilt tatsächlich $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}x^2\right) dx = 1.$

Standardnormalverteilung

Definition 3.8. Eine stetige Zufallsvariable heißt *normal-verteilt* mit den Parametern μ und σ (kurz: $X \sim N(\mu, \sigma^2)$), wenn die Dichtefunktion folgende Gestalt hat:

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

Wir sagen kurz: X ist $N(\mu, \sigma^2)$ -verteilt. Die Dichtefunktion aus Beispiel 3.4 ergibt sich mit $\sigma = 1$ und $\mu = 0$; die $N(0, 1)$ -Verteilung heißt *Standardnormalverteilung*.

Normal-Verteilung

Definition 3.8. Eine stetige Zufallsvariable heißt *normal-verteilt* mit den Parametern μ und σ (kurz: $X \sim N(\mu, \sigma^2)$), wenn die Dichtefunktion folgende Gestalt hat:

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

Wir sagen kurz: X ist $N(\mu, \sigma^2)$ -verteilt. Die Dichtefunktion aus Beispiel 3.4 ergibt sich mit $\sigma = 1$ und $\mu = 0$; die $N(0, 1)$ -Verteilung heißt *Standardnormalverteilung*.

Es gilt tatsächlich $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right) dx = 1.$

$f(x)$ ist symmetrisch um μ . D.h.: $f(\mu - a) = f(\mu + a)$ für alle $a \in \mathbb{R}$.

$f(x)$ nimmt das Maximum an, wenn $x = \mu$.

Normal-Verteilung

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx = 1$$

Satz 3.9. Ist X $N(\mu, \sigma^2)$ -verteilt, so gilt

$$E(X) = \mu, \quad \text{Var}(X) = \sigma^2.$$

Bemerkungen 3.10. Ist X $N(0, 1)$ -verteilt ($X \sim N(0, 1)$), so bezeichnet man üblicherweise die zugehörige Verteilungsfunktion mit Φ , d.h.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{1}{2}t^2\right) dt.$$

Nicht elementar! D.h.: es gibt keine „einfachere“ Beschreibung!

Substitution:

$$\int_{-\infty}^{t_0} \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx = \int_{-\infty}^{\frac{t_0-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}t^2\right) dt = \Phi\left(\frac{t_0-\mu}{\sigma}\right)$$

Also gilt für $X \sim N(\mu, \sigma)$: $F(t_0) = P(X \leq t_0) = \Phi\left(\frac{t_0-\mu}{\sigma}\right).$

Normal-Verteilung

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}t^2\right) dt$$

Tabelle mit den Werten der Verteilungsfunktion ϕ der Standardnormalverteilung

x	$\phi(x)$												
0,00	0,5000	0,53	0,7019	1,06	0,8554	1,59	0,9441	2,12	0,9830	2,65	0,9960	3,18	0,99926
0,01	0,5040	0,54	0,7054	1,07	0,8577	1,60	0,9452	2,13	0,9834	2,66	0,9961	3,19	0,99929
0,02	0,5080	0,55	0,7088	1,08	0,8599	1,61	0,9463	2,14	0,9838	2,67	0,9962	3,20	0,99931
0,03	0,5120	0,56	0,7123	1,09	0,8621	1,62	0,9474	2,15	0,9842	2,68	0,9963	3,21	0,99934
0,04	0,5160	0,57	0,7157	1,10	0,8643	1,63	0,9485	2,16	0,9846	2,69	0,9964	3,22	0,99936
0,05	0,5199	0,58	0,7190	1,11	0,8665	1,64	0,9495	2,17	0,9850	2,70	0,9965	3,23	0,99938
0,06	0,5239	0,59	0,7224	1,12	0,8686	1,65	0,9505	2,18	0,9854	2,71	0,9966	3,24	0,99940
0,07	0,5279	0,60	0,7258	1,13	0,8708	1,66	0,9515	2,19	0,9857	2,72	0,9967	3,25	0,99942
0,08	0,5319	0,61	0,7291	1,14	0,8729	1,67	0,9525	2,20	0,9861	2,73	0,9968	3,26	0,99944
0,09	0,5359	0,62	0,7324	1,15	0,8749	1,68	0,9535	2,21	0,9865	2,74	0,9969	3,27	0,99946
0,10	0,5398	0,63	0,7357	1,16	0,8770	1,69	0,9545	2,22	0,9868	2,75	0,9970	3,28	0,99948
0,11	0,5438	0,64	0,7389	1,17	0,8790	1,70	0,9554	2,23	0,9871	2,76	0,99711	3,29	0,99950
0,12	0,5478	0,65	0,7422	1,18	0,8810	1,71	0,9564	2,24	0,9875	2,77	0,99720	3,30	0,99952
0,13	0,5517	0,66	0,7454	1,19	0,8830	1,72	0,9573	2,25	0,9878	2,78	0,99728	3,31	0,99953
0,14	0,5557	0,67	0,7486	1,20	0,8849	1,73	0,9582	2,26	0,9881	2,79	0,99736	3,32	0,99955
0,15	0,5596	0,68	0,7518	1,21	0,8869	1,74	0,9591	2,27	0,9884	2,80	0,99744	3,33	0,99957
0,16	0,5636	0,69	0,7549	1,22	0,8888	1,75	0,9599	2,28	0,9887	2,81	0,99752	3,34	0,99958
0,17	0,5675	0,70	0,7580	1,23	0,8907	1,76	0,9608	2,29	0,9890	2,82	0,99760	3,35	0,99960
0,18	0,5714	0,71	0,7612	1,24	0,8925	1,77	0,9616	2,30	0,9893	2,83	0,99767	3,36	0,99961
0,19	0,5754	0,72	0,7642	1,25	0,8944	1,78	0,9625	2,31	0,9896	2,84	0,99774	3,37	0,99962

$$\Phi(-x) + \Phi(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}t^2\right) dt = 1 \quad \implies \Phi(-x) = 1 - \Phi(x)$$

Normal-Verteilung

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}t^2\right) dt$$

Beispiel NV: Ein Werkstück soll eine Bohrung erhalten mit einem Durchmesser von 50mm. Die Toleranzgrenzen sind $t_u = 49,97\text{mm}$ und $t_o = 50,04\text{mm}$. Es sei bekannt, dass die von den Bohrautomaten erstellten Bohrungen $N(\mu, \sigma^2)$ verteilt sind, wobei $\mu = 50\text{mm}$ und $\sigma = 0.02\text{mm}$ gelten soll. Ein Werkstück ist Ausschuss, wenn der Durchmesser größer als t_o ausfällt. Ist der Durchmesser kleiner als t_u , so muss eine Nachbohrung durchgeführt werden.

(a) Wie groß ist die Wahrscheinlichkeit, dass ein Werkstück in den Toleranzgrenzen liegt?

Sei X die zugehörige Zufallsvariable. Dann ist folgender Wert gesucht

$$\begin{aligned} P(t_u \leq X \leq t_o) &= P(X \leq t_o) - P(X \leq t_u). \\ &= \Phi\left(\frac{t_o - \mu}{\sigma}\right) - \Phi\left(\frac{t_u - \mu}{\sigma}\right) \\ &= \Phi(2) - \Phi(-1,5) = \Phi(2) - (1 - \Phi(1,5)) \\ &\approx 0.9104 \end{aligned}$$

Normal-Verteilung

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}t^2\right) dt$$

Beispiel NV: Ein Werkstück soll eine Bohrung erhalten mit einem Durchmesser von 50mm. Die Toleranzgrenzen sind $t_u = 49,97\text{mm}$ und $t_o = 50,04\text{mm}$. Es sei bekannt, dass die von den Bohrautomaten erstellten Bohrungen $N(\mu, \sigma^2)$ verteilt sind, wobei $\mu = 50\text{mm}$ und $\sigma = 0.02\text{mm}$ gelten soll. Ein Werkstück ist Ausschuss, wenn der Durchmesser größer als t_o ausfällt. Ist der Durchmesser kleiner als t_u , so muss eine Nachbohrung durchgeführt werden.

(b) Wie groß ist die Wahrscheinlichkeit, dass ein Werkstück nachgebessert werden muss?

Sei X die zugehörige Zufallsvariable. Dann ist folgender Wert gesucht

$$\begin{aligned} P(X \leq t_u) \\ &= \Phi\left(\frac{t_u - \mu}{\sigma}\right) = \Phi(-1,5) = 1 - \Phi(1,5) \\ &\approx 0.0668 \end{aligned}$$

Normal-Verteilung

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}t^2\right) dt$$

Beispiel NV: Ein Werkstück soll eine Bohrung erhalten mit einem Durchmesser von 50mm. Die Toleranzgrenzen sind $t_u = 49,97\text{mm}$ und $t_o = 50,04\text{mm}$. Es sei bekannt, dass die von den Bohrautomaten erstellten Bohrungen $N(\mu, \sigma^2)$ verteilt sind, wobei $\mu = 50\text{mm}$ und $\sigma = 0.02\text{mm}$ gelten soll. Ein Werkstück ist Ausschuss, wenn der Durchmesser größer als t_o ausfällt. Ist der Durchmesser kleiner als t_u , so muss eine Nachbohrung durchgeführt werden.

(c) Wie groß ist die Wahrscheinlichkeit, dass ein Werkstück Ausschuss ist?

Sei X die zugehörige Zufallsvariable. Dann ist folgender Wert gesucht

$$\begin{aligned} P(X \geq t_o) &= 1 - P(X \leq t_o) \\ &= 1 - \Phi\left(\frac{t_o - \mu}{\sigma}\right) = 1 - \Phi(2) \\ &\approx 0.0228 \end{aligned}$$

Alternativ: Es muss gelten $P(X \leq t_u) + P(t_u \leq X \leq t_o) + P(X \geq t_o) = 1$.

Normal-Verteilung

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}t^2\right) dt$$

Bemerkung 3.13. Häufig ist man an der Wahrscheinlichkeit interessiert, dass die $N(\mu, \sigma^2)$ -verteilte Zufallsvariable Werte in einem zu μ symmetrischen Intervall $[\mu - k\sigma, \mu + k\sigma]$ mit $k \in \mathbb{N}$ annimmt. Es ist üblich, die Abweichung von μ in Einheiten von σ anzugeben. Deshalb spricht man vom $k\sigma$ -Intervall. Wir erhalten

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1 .$$

Speziell für $k = 1, 2, 3$ ergeben sich folgende Werte

$$P(\mu - 1 \cdot \sigma \leq X \leq \mu + 1 \cdot \sigma) = 2\Phi(1) - 1 \approx 0.6826 ,$$

$$P(\mu - 2 \cdot \sigma \leq X \leq \mu + 2 \cdot \sigma) = 2\Phi(2) - 1 \approx 0.9544 ,$$

$$P(\mu - 3 \cdot \sigma \leq X \leq \mu + 3 \cdot \sigma) = 2\Phi(3) - 1 \approx 0.9974 .$$

Also liegen ca. 68 % der beobachteten Werte bei einer $N(\mu, \sigma^2)$ -verteilten Zufallsvariablen zwischen $\mu - \sigma$ und $\mu + \sigma$, ca. 95 % liegen zwischen $\mu - 2\sigma$ und $\mu + 2\sigma$ und ca. 99.7 % liegen zwischen $\mu - 3\sigma$ und $\mu + 3\sigma$.

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}t^2\right) dt$$

Definition 4.1. a) Ist X eine Zufallsvariable mit dem Erwartungswert $\mu = E(X)$ und der Varianz $\sigma^2 = \text{Var}(X) > 0$, so gilt für die Zufallsvariable

$$X^* := \frac{X - \mu}{\sigma}$$

nach den Sätzen 2.12 und 2.16

$$E(X^*) = 0 \quad \text{und} \quad \text{Var}(X^*) = 1.$$

X^* heißt die zu X gehörende *standardisierte Variable*.

b) Zwei Zufallsvariable X und Y auf demselben Wahrscheinlichkeitsraum (Ω, \mathcal{A}, p) heißen *gleichverteilt* oder *identisch verteilt*, wenn ihre Verteilungsfunktionen übereinstimmen.

$$a, b \in \mathbb{R}: \quad E(aX + b) = aE(X) + b \quad \text{und} \quad \text{Var}(aX + b) = a^2 \text{Var}(X)$$

$$a \leq X^* \leq b \iff \sigma a \leq X - \mu \leq \sigma b \iff \sigma a + \mu \leq X \leq \sigma b + \mu$$

Grenzwertsätze

Binomialverteilung reloaded:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}t^2\right) dt$$

Ist X_n binomialverteilt mit n und p , so ist $P(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = B_{n,p}(k)$

und die Werte von X_n sind genau die Elemente aus $\{0, \dots, n\}$.

$$\text{Dann ist } X_n^* = \frac{X_n - np}{\sqrt{np(1-p)}}$$

und die Werte von X_n^* sind die Elemente aus $\left\{ \frac{k - np}{\sqrt{np(1-p)}} \mid k \in \{0, \dots, n\} \right\}$.

Wir wollen die Wahrscheinlichkeiten von X_n^* als Flächeninhalt einer Säule darstellen.

Die Breite einer Säule ist $\frac{1}{\sqrt{np(1-p)}}$. Damit der Flächeninhalt gleich $B_{n,p}(k)$ ist, muss

die Höhe einer Säule gleich $\sqrt{np(1-p)} \cdot B_{n,p}(k)$ sein.

Die Treppenfunktion, die über alle diese Säulen läuft, ist

$$f_n(x) = \begin{cases} \sqrt{np(1-p)} \cdot B_{n,p}(k) & \text{falls } \frac{k - np - \frac{1}{2}}{\sqrt{np(1-p)}} \leq x < \frac{k - np + \frac{1}{2}}{\sqrt{np(1-p)}} \\ 0 & \text{sonst} \end{cases}$$

Grenzwertsätze

Binomialverteilung reloaded:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}t^2\right) dt$$

Ist X_n binomialverteilt mit n und p , so ist

$$P(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = B_{n,p}(k)$$

und die Werte von X_n sind genau die Elemente aus $\{0, \dots, n\}$.

$$\text{Dann ist } X_n^* = \frac{X_n - np}{\sqrt{np(1-p)}}$$

und die Werte von X_n^* sind die Elemente aus

$$\left\{ \frac{k - np}{\sqrt{np(1-p)}} \mid k \in \{0, \dots, n\} \right\}.$$

Wir wollen die Wahrscheinlichkeiten von X_n^* als Flächeninhalt einer Säule darstellen.

Die Breite einer Säule ist $\frac{1}{\sqrt{np(1-p)}}$. Damit der Flächeninhalt gleich $B_{n,p}(k)$ ist, muss die Höhe einer Säule

gleich $\sqrt{np(1-p)} \cdot B_{n,p}(k)$ sein.

Die Treppenfunktion, die über alle diese Säulen läuft, ist

$$f_n(x) = \begin{cases} \sqrt{np(1-p)} \cdot B_{n,p}(k) & \text{falls } \frac{k - np - \frac{1}{2}}{\sqrt{np(1-p)}} \leq x < \frac{k - np + \frac{1}{2}}{\sqrt{np(1-p)}} \\ 0 & \text{sonst} \end{cases}$$

Es ist $\int_{-\infty}^{\infty} f_n dx = \sum_{k=0}^n B_{n,p}(k) = 1 \implies f_n$ ist eine Dichtefunktion!

$$\lim_{n \rightarrow \infty} f_n(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}t^2\right) dt$$

Satz 4.3 (Satz von de Moivre-Laplace). *Es sei $0 < p < 1$ und X_n $B(n, p)$ -verteilt sowie X_n^* die zu X_n gehörende standardisierte Zufallsvariable. Dann gilt für alle $a < b$:*

$$\lim_{n \rightarrow \infty} P(a \leq X_n^* \leq b) = \phi(b) - \phi(a),$$

wobei ϕ (wie üblich) die Verteilungsfunktion der Standardnormalverteilung bezeichnet.

Es folgt:

Ist X eine $B(n, p)$ -verteilte Zufallsvariable, dann gilt

$$P(X \leq x) \approx \Phi\left(\frac{x - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right)$$

Faustregel: gute Approximation, wenn $np(1-p) > 9$.

Grenzwertsätze

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}t^2\right) dt$$

$$X \text{ ist } B(n, p)\text{-verteilt: } P(X \leq x) \approx \Phi\left(\frac{x - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right)$$

In einem Kurs sind 220 Studierende angemeldet. In dem Hörsaal sind 160 Plätze. Studierende erscheinen mit einer Wahrscheinlichkeit von 0,75 zu einer Vorlesung. Wie groß ist die Wahrscheinlichkeit, dass der Platz ausreicht?

Sei X die Zufallsvariable, die die Anzahl von Studierenden in einer Vorlesung misst. Dann ist X binomialverteilt, mit $n = 220$ und $p = 0,75$.

Gesucht: $P(X \leq 160)$

Faustregel erfüllt: $np(1-p) = 220 \cdot 0,75 \cdot 0,25 = 41,25 > 9$

$$\text{Es folgt: } P(X \leq 160) \approx \Phi\left(\frac{160 - 165 + 0,5}{\sqrt{41,25}}\right) = \Phi(-0,7006\dots)$$

$$= 1 - \Phi(0,7006\dots) \approx 1 - 0,758 = 0,242$$

Genaueres Ergebnis mit Binomialverteilung: 0,2397...

Grenzwertsätze

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}t^2\right) dt$$

$$X \text{ ist } B(n, p)\text{-verteilt: } P(X \leq x) \approx \Phi\left(\frac{x - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right)$$

In einem Hörsaal sind 160 Plätze. Studierende erscheinen mit einer Wahrscheinlichkeit von 0,75 zu einer Vorlesung. Wie viele Studierende dürfen angenommen werden, damit der Platz mit einer Wahrscheinlichkeit von 0,95 ausreicht?

Sei X die Zufallsvariable, die die Anzahl von Studierenden in einer Vorlesung misst. Dann ist X binomialverteilt, mit unbekanntem n und $p = 0,75$.

Gesucht: größtes $n \in \mathbb{N}$, mit $P(X \leq 160) \geq 0,95$

$$P(X \leq 160) \approx \Phi\left(\frac{160 - n \cdot 0,75 + 0,5}{\sqrt{n \cdot 0,75 \cdot 0,25}}\right) \text{ und } \Phi(1,64) < 0,95 \text{ und } \Phi(1,65) > 0,95$$

Löse: $\frac{160 - n \cdot 0,75 + 0,5}{\sqrt{n \cdot 0,75 \cdot 0,25}} = 1,65$ **(Quadratische Gleichung in \sqrt{n})** $n = 200,51\dots$

Fazit: Es dürfen höchstens 200 Studierende angenommen werden.