Statistik

Vorlesung 7 - Stetige Zufallsvariablen und Verteilungen

Prof. Dr. Sandra Eisenreich

Hochschule Landshut

Agenda

Stetige Zufallsvariable

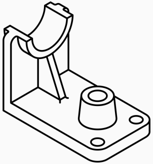
Motivation: stetige Zufallsvariable

Erinnerung: stetige Zufallsvariablen sind ZV mit kontinuierlichen reellen Zahlen als Ergebnissen (+Zusatzeigenschaft)

Beispiele: Die Dauer eines Vorgangs (z.B. zum Auskühlen eines Bauteils/Stoffs), die Temperatur eines Werksstoffs, die Länge eines Bauteils.... . . allgemein Zeit-/Längen- oder andere Messungen mit reellen Werten als Ergebnis!

Wahrscheinlichkeitsdichte - Herleitung

Beispiel: $X =$ Abweichung der Bauteillänge vom Soll.



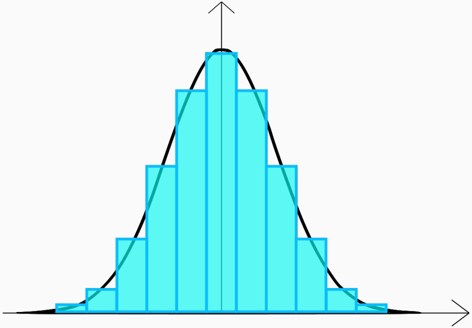
Was ist die Wahrscheinlichkeit, dass das Bauteil exakt um 0,1 mm vom Soll abweicht?

Antwort: 0! Wenn dieses Ereignis die Wahrscheinlichkeit $p > 0$ hätte, dann auch alle unendlich vielen Werte zwischen 0,1 mm und 0,1mm $+ 1\mu$m...

$\Rightarrow$ Im Fall von stetigen Zufallsvariablen kann man nur für Bereiche Wahrscheinlichkeiten angeben!

Wahrscheinlichkeitsdichte - Herleitung

Wir können die Abweichungen in 0.1mm-Abschnitte unterteilen (diskrete Werte wie bisher):

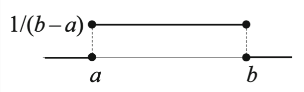


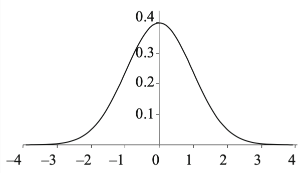
Die Wahrscheinlichkeit jedes Abschnitts = Fläche des Balken.

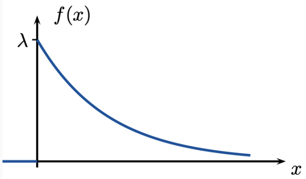
Wenn wir die Breite der Abschnitte immer kleiner werden lassen $\rightarrow$ stetige Kurve = (Wahrscheinlichkeits-)Dichte. Die Wahrscheinlichkeit für ein Intervall ist die Fläche unter der Kurve im Intervall.

Von diskret zu stetig

⇒ stetige Verteilungen bzw. ihre **Dichten** = "Grenzwerte" von diskreten Verteilungen.



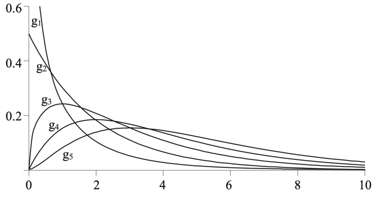


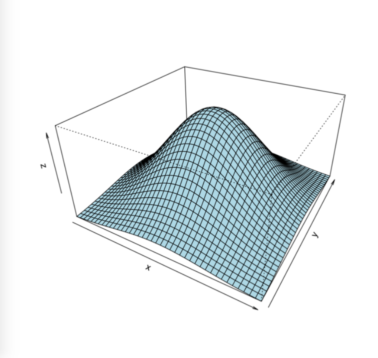


- diskrete Gleichverteilung → stetige Gleichverteilung
- standardisierte Binomialverteilung (Mittelwert 0 und Varianz 1) → "Gaußglocke" = Standardnormalverteilung
- geometrische Verteilung → Exponentialverteilung.

- Chi-Quadrat-Verteilung = Verteilung von $X_1^2 + \ldots + X_n^2$ für normalverteilte $X_1, \ldots, X_n$

- multivariate Normalverteilung = Gauß-Glocke für Zufallsvektoren (höherdimensionales Ergebnis).





Verteilungsfunktion

Merke: Bei stetigen Verteilungen hat ein einzelnes Ergebnis Wahrscheinlichkeit 0! Nur Bereiche haben Wahrscheinlichkeiten > 0 ($\rightarrow$ Verteilungsfunktion).

Beispiele:

- Wie groß ist die Wahrscheinlichkeit, dass der Zeiger zwischen 2 und 3 steht?
- Wie groß ist die Wahrscheinlichkeit, dass ein Bauteil keine Abweichung größer als $\pm$ 0.2mm aufweist?

Verteilungsfunktion

Wahrscheinlichkeiten können bei stetigen Verteilungen nur über die Verteilungsfunktion angegeben werden:

Erinnerung:

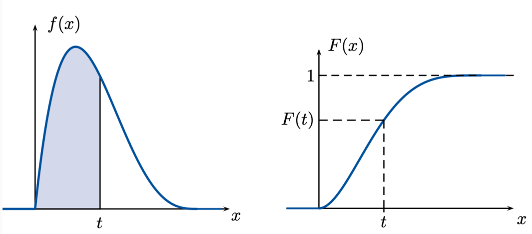
Ist X eine Zufallsvariable auf (Ω, P), so heißt die Funktion

$$F : \mathbb{R} \to [0, 1], \quad x \mapsto P(X \leq x)$$

die Verteilungsfunktion von X.

Dichte und Verteilungsfunktion - Herleitung

- Wahrscheinlichkeitsdichte: gibt also die Wahrscheinlichkeitsverteilung über alle möglichen Werte an, aber keine Wahrscheinlichkeiten!
- Verteilungsfunktion $P(X \leq x)$ = Fläche unterm Graphen bis $x = \int_{-\infty}^{x} f(t)dt$



Dichte und Verteilungsfunktion einer stetigen Zufallsvariablen

Definition (Stetige Zufallsvariable)

Eine Zufallsvariable X heißt stetig (verteilt), wenn es eine nichtnegative integrierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ mit der Eigenschaft

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

gibt, so dass die Verteilungsfunktion F von X die Darstellung

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \quad x \in \mathbb{R},$$

besitzt. In diesem Fall sagt man auch, X habe eine stetige Verteilung. Die Funktion f heißt Dichte von X, oder Dichte der Verteilungsfunktion von X.

Zusammenhang Dichte und Verteilung

Sei X eine stetige Zufallsvariable mit Verteilungsfunktion F und Dichte f.

- $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$.
- $x < y \Rightarrow F(x) \leq F(y)$
- $F'(x) = f(x)$ (wenn f bei x stetig ist)
- $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.

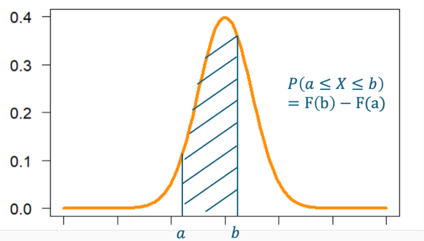
Bemerkung: f ist durch F nicht eindeutig festgelegt, man kann f an endlich vielen Stellen abändern ohne dass F sich ändert.

Wahrscheinlichkeiten von Intervallen

Ist F Verteilungsfunktion der stetigen Zufallsvariablen X mit der Dichte f, so gilt für alle reellen Zahlen $a < b$:

(a) $P(a < X \leq b) = F(b) - F(a) = \int_a^b f(t)dt$.

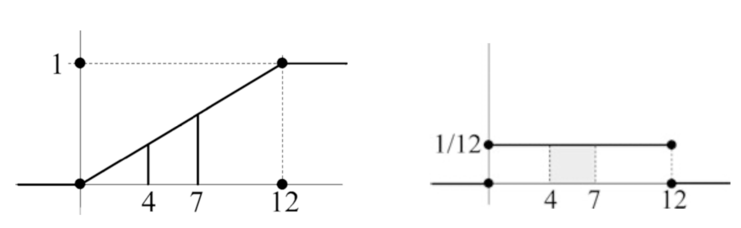
(b) $P(X > a) = 1 - F(a) = \int_a^\infty f(t)dt$.



Beispiele

Betrachten wir die Zufallsvariable $X : \Omega \to (0, 12]$ "Stellung des Studenzeigers der Uhr". Dann gilt:

- $P(X \leq 12) = 1$
- $P(X \leq 6) = \frac{1}{2}$
- $P(2 < X \leq 3) = \frac{1}{12}$



Verteilungsfunktion und Dichte der Zufallsvariablen X

Unabhängige stetige Zufallsvariablen

Definition

Zwei stetige Zufallsvariablen X und Y heißen unabhängig, wenn für alle $x, y \in \mathbb{R}$ gilt

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$$

Dies kann man für mehrere Zufallsvariablen verallgemeinern! Für diskrete und stetige unabhängige Zufallsvariablen gilt für alle $a, b, c, d \in \mathbb{R}$

$$P(a < X \leq b, c < Y \leq d) = P(a < X \leq b) \cdot P(c < Y \leq d)$$

Erwartungswert und Varianz von stetigen Zufallsvariablen

Motivation: Erwartungswert und Varianz einer stetigen Zufallsvariable

X: diskrete Zufallsvariable

$$E(X) = \sum_{x \in W(X)} P(X = x) \cdot x = \text{ Fläche unter der Funktion} P(X = x) \cdot x$$

X: stetige Zufallsvariable (ersetze $P(X = x)$durch Dichte $f(x)$):

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \text{ Fläche unter dem Graphen der Funktion} f(x) \cdot x$$

Für die Varianz gilt bei diskreten: $Var(X) = E(X^2) - E(X)^2$, diese Definition funktioniert hier ebenso.

Erwartungswert und Varianz einer stetigen Zufallsvariable

Sei X eine stetige Zufallsvariable mit Dichte f.

- Erwartungswert (falls existent):.

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

- Varianz:

$$\sigma^2 := Var(X) = E(X^2) - E(X)^2,$$

- Standardabweichung/Streuung:

$$\sigma = \sqrt{Var(X)}$$

Rechenregeln

Wie im Fall von disreten Zufallsvariablen gelten folgende Rechenregeln:

- für Zufallsvariablen X, Y, $a \in \mathbb{R}$, und $A \subset \Omega$ ein Ereignis gelten:
  (a) $E(X+Y) = E(X) + E(Y)$,
  (b) $E(aX) = aE(X)$,
  (c) $E(1_A) = P(A)$,
- für **unabhängige** Zufallsvariablen X, Y gilt:

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

$$Var(X+Y) = Var(X) + Var(Y).$$

Stetige Gleichverteilung

Die stetige Gleichverteilung

Die einfachste stetige Verteilung, genauso wie im diskreten Fall, ist die Gleichverteilung.

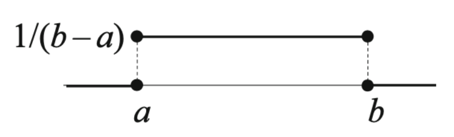


Abbildung 1: Die Dichte der Gleichverteilung

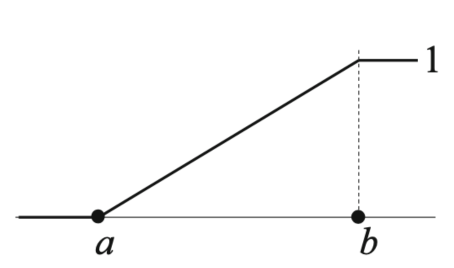


Abbildung 2: Die Verteilungsfunktion der Gleichverteilung

Achtung: hier können wir nur die Wahrscheinlichkeit für Teilbereiche von $[a, b]$ angeben!

Definition

Die Dichte der Gleichverteilung im Intervall $[a, b]$ hat die Form

$$w : \mathbb{R} \to \mathbb{R}, x \mapsto \begin{cases} 0 & x < a \text{ oder } x > b \\ \frac{1}{b-a} & a \leq x \leq b \end{cases}.$$

Die zugehörige Verteilungsfunktion ist damit

$$F : \mathbb{R} \to [0, 1], x \mapsto \begin{cases} 0 & x < a \\ 1 & x > b \\ \frac{x-a}{b-a} & a \leq x \leq b \end{cases}$$

(wobei $\frac{x-a}{b-a}$ die Stammfunktion von $\frac{1}{b-a}$ ist!)

Berechnung: Erwartungswert von Varianz

Erwartungswert:

$$
\begin{aligned}
E(X) &= \int_a^b xw(x)dx = \int_a^b \frac{x}{b-a}dx \\
&= \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)} \\
&= \frac{(a+b)(b-a)}{2(b-a)} \\
&= \frac{a+b}{2}
\end{aligned}
$$

Das ist genau der Mittelwert von a und b.

Varianz:

$$
\begin{aligned}
Var(X) &= E(X^2) - E(X)^2 \\
&= \int_a^b \frac{x^2}{b-a}dx - \left(\frac{a+b}{2}\right)^2 \\
&= \frac{(b-a)^2}{12}
\end{aligned}
$$

Beispiel: Zufallszahlengeneratoren

Die Ergebnisse eines Zufallszahlengenerators, der Zahlen zwischen 0 und 1 erzeugt, sind im Idealfall gleichverteilt.

Der Erwartungswert ist 0.5 und es ist zum Beispiel

$$P(0.1 < X < 0.2) = F(0.2) - F(0.1) = 0.2 - 0.1 = 0.1$$

Wie kann man einen Zufallszahlengenerator überprüfen?

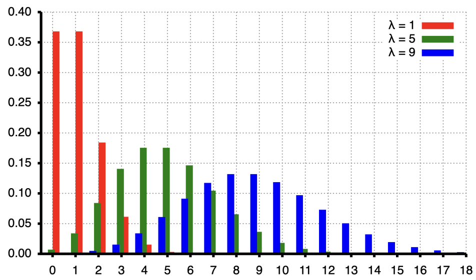
$\rightarrow$ beliebig viele Zahlen erzeugen und die entstandene Datenmenge analysieren.

Normalverteilung

Binomialverteilung bei großem n, kleinem p - Poissonverteilung

Für konstantes $\lambda = np$ konvergiert die Binomialverteilung für $n \to \infty$ (also für großes n und kleines p) gegen die Poisson-Verteilung:

$$b_{n,p}(k) \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

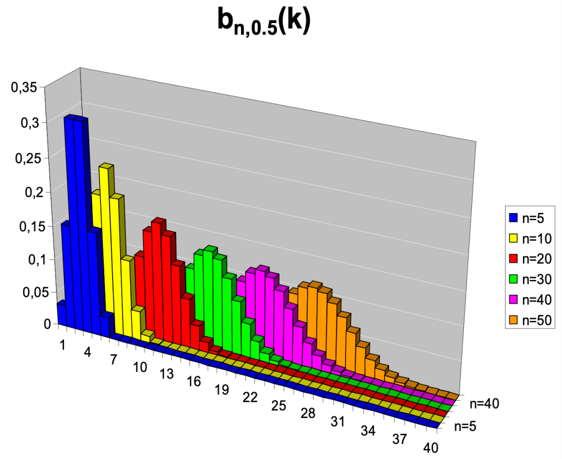


Frage: Was wenn n größer wird, aber p konstant bleibt?

Binomialverteilung bei großem n, konstantem p

Für größere n wird die Verteilung immer flacher und breiter, der Gipfel rückt nach rechts, denn für $n \to \infty$

$$E(X) = np \to \infty$$
$$Var(X) = np(1-p) \to \infty$$



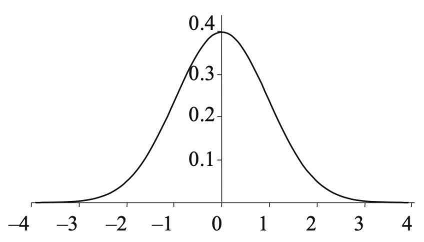
Es gibt also keine Grenzverteilung, an die sich die Binomialverteilungen annähern.

standardisierte Binomialverteilung bei großem n, konstantem p

Idee: $b_{n,p}$ für alle n "verschieben" auf $\mu = 0$ und "zusammenstauchen" auf $\sigma^2 = 1$.

- wir standardisieren: $X^* = \frac{X-\mu}{\sigma}$ $\Rightarrow$ Erwartungswert 0 und Varianz 1
- $X = k \Leftrightarrow X^* = k^* := \frac{k-\mu}{\sigma} = \frac{k-np}{\sqrt{np(1-p)}}$

Ausblick: für $n \to \infty$ bekommt man dann eine "Gauß-Glocke"



Beispiel: Binomialverteilung für $n = 16$, $p = \frac{1}{2}$

$$\mu = np = 8,\ \sigma = \sqrt{16 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 2$$

Histogramme für $b_{n,p}$ und $b^*_{n,p}$: die Säulenbreite für X^*:

$$(k+1)^* - k^* = \frac{(k+1) - \mu}{\sigma} - \frac{k - \mu}{\sigma} = \frac{1}{\sigma}$$

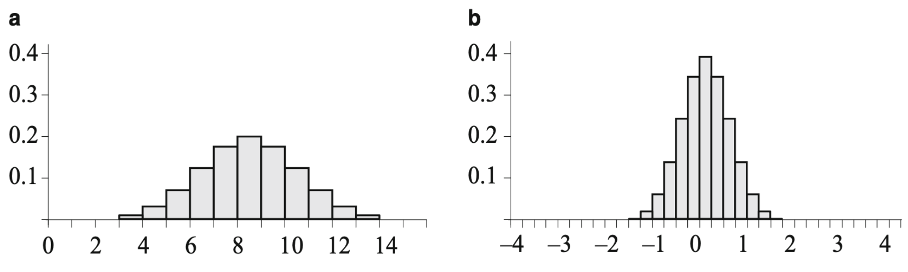


Abbildung 3: $b_{16,1/2}(k)$ und $b^*_{16,1/2}(k)$

Standardisierte Binomialverteilungen

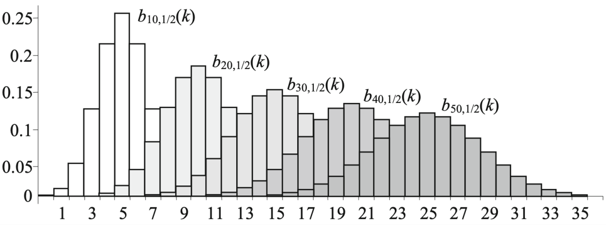


Abbildung 4: Die Binomialverteilung bei wachsendem n

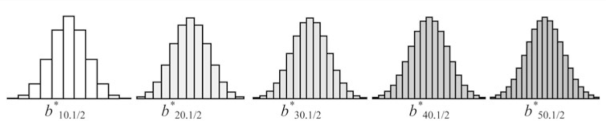
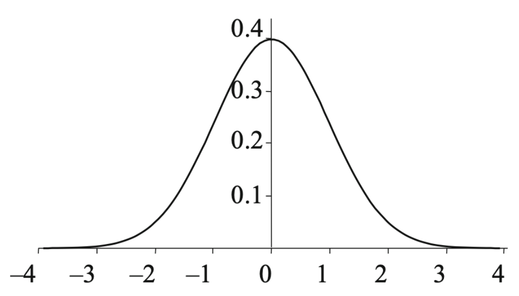


Abbildung 5: Standardisierte Binomialverteilungen

Konvergenz der Histogramme für $n \to \infty$

Der Grenzwertsatz von Moivre-Laplace: Es sei $\phi_n(x)$ die Höhe des Histogramms bei x. Für alle p zwischen 0 und 1 ist

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



Standardnormalverteilung

Eine Zufallsvariable N heißt standardnormalverteilt oder $\mathcal{N}(0,1)$-verteilt mit

- Dichte:

$$N(x|0,1) := \frac{1}{\sqrt{2\pi}} \exp(\frac{-x^2}{2})$$

- Verteilungsfunktion:

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

Sie hat Erwartungswert 0 und Varianz 1.

Rechenregel

Leider lässt sich das Integral $\int_{-\infty}^{x} \exp(\frac{-t^2}{2})dt$ nicht analytisch ausrechnen! Dies geht nur numerisch ($\rightarrow$ in Tabelle ablesen!).

Merke: Da $\phi(x)$ symmetrisch ist, gilt für $x > 0$

$$\Phi(-x) = 1 - \Phi(x)$$

und es müssen nur die positiven Werte tabelliert werden.

Verteilungsfunktion der Normalverteilung - Tabelle

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	0.46	0.6772	0.92	0.8212	1.38	0.9162	1.84	0.9671	2.30	0.9893	2.76	0.99711	3.22	0.99936
0.01	0.5040	0.47	0.6808	0.93	0.8238	1.39	0.9177	1.85	0.9678	2.31	0.9896	2.77	0.99720	3.23	0.99938
0.02	0.5080	0.48	0.6844	0.94	0.8264	1.40	0.9192	1.86	0.9686	2.32	0.9898	2.78	0.99728	3.24	0.99940
0.03	0.5120	0.49	0.6879	0.95	0.8289	1.41	0.9207	1.87	0.9693	2.33	0.9901	2.79	0.99736	3.25	0.99942
0.04	0.5160	0.50	0.6915	0.96	0.8315	1.42	0.9222	1.88	0.9700	2.34	0.9904	2.80	0.99744	3.26	0.99944
0.05	0.5199	0.51	0.6950	0.97	0.8340	1.43	0.9236	1.89	0.9706	2.35	0.9906	2.81	0.99752	3.27	0.99946
0.06	0.5239	0.52	0.6985	0.98	0.8365	1.44	0.9251	1.90	0.9713	2.36	0.9909	2.82	0.99760	3.28	0.99948
0.07	0.5279	0.53	0.7019	0.99	0.8389	1.45	0.9265	1.91	0.9719	2.37	0.9911	2.83	0.99767	3.29	0.99950
0.08	0.5319	0.54	0.7054	1.00	0.8413	1.46	0.9279	1.92	0.9726	2.38	0.9913	2.84	0.99774	3.30	0.99952
0.09	0.5359	0.55	0.7088	1.01	0.8438	1.47	0.9292	1.93	0.9732	2.39	0.9916	2.85	0.99781	3.31	0.99953
0.10	0.5398	0.56	0.7123	1.02	0.8461	1.48	0.9306	1.94	0.9738	2.40	0.9918	2.86	0.99788	3.32	0.99955
0.11	0.5438	0.57	0.7157	1.03	0.8485	1.49	0.9319	1.95	0.9744	2.41	0.9920	2.87	0.99795	3.33	0.99957
0.12	0.5478	0.58	0.7190	1.04	0.8508	1.50	0.9332	1.96	0.9750	2.42	0.9922	2.88	0.99801	3.34	0.99958
0.13	0.5517	0.59	0.7224	1.05	0.8531	1.51	0.9345	1.97	0.9756	2.43	0.9925	2.89	0.99807	3.35	0.99960

Abbildung 6: Die Verteilungsfunktion der Gaußschen Standardnormalverteilung - Auszug der Tablle

Wichtige Rechenregel zur Benutzung der Normalverteilung als Näherung für die Binomialverteilung: Die standardisierten Binomialverteilungen sind für große n (und alle möglichen p) näherungsweise standardnormalverteilt, daher gilt

$$P(k_1 \leq X \leq k_2) = P(k_1^* \leq X^* \leq k_2^*) \approx \Phi(k_2^*) - \Phi(k_1^*)$$

(für binomialverteile ZV X). Dabei ist

$$k_1^* = \frac{k_1 - \mu}{\sigma}, \quad k_2^* = \frac{k_2 - \mu}{\sigma}$$

Rechenregel: Merken!

Es geht noch etwas genauer:

Theorem

Sei X eine $b_{n,p}$-verteilte Zufallsvariable. Dann ist für $np > 5$ und $n(1-p) > 5$ die folgende Rechnung möglich:

$$P(k_1 \leq X \leq k_2) = \Phi\left(\frac{k_2 - np + 0.5}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - np - 0.5}{\sqrt{np(1-p)}}\right)$$

$$P(X \leq k) = \Phi\left(\frac{k - np + 0.5}{\sqrt{np(1-p)}}\right),$$

$$P(X < k) = \Phi\left(\frac{k - np - 0.5}{\sqrt{np(1-p)}}\right)$$

Beispiel: Flugbuchung

Problem: Eine Fluggesellschaft überbucht regelmäßig ihre Flüge zu 105%. Erfahrung sagt: 90% der Passagiere kommen zum Flug. Wie groß ist die Wahrscheinlichkeit, dass Passagier zurückgelassen werden muss?

Der Einfachheit halber: 100 Plätze, 105 Tickets verkauft.
Bernoulliexperiment: Passagier x kommt.

$$n = 105, p = 0.9, b_{n,p}(k) = P(\text{"} k \text{ Passagiere kommen"})$$

$$\left.\begin{array}{c} b_{n,p}(101) \\ b_{n,p}(102) \\ \vdots \\ b_{n,p}(105) \end{array}\right\} \text{Summe} \;=\; \text{W.keit für mehr als 100 Passagiere}$$

Diese Summe ist schwierig zu berechnen mit der Binomialverteilung.

Ergebnis: Flugbuchung

Approximation über die Standardnormalverteilung:

$np = 94.5$, $n(1-p) = 10.5$, also ist unsere Rechenregel anwendbar:

$$\begin{aligned} P(X > 100) &= 1 - P(X \leq 100) \\ &= 1 - \Phi\left(\frac{100 - 94.5 + 0.5}{\sqrt{94.5 \cdot 0.1}}\right) \\ &= 1 - \Phi(1.951) \\ &= 1 - 0.9744 \\ &= 0.0256 \end{aligned}$$

d.h. etwa 2.5% der Fälle.

Beispiel: Wahlumfrage

Problem: In einer Wahl haben 10% der Wähler die Partei A gewählt. In einer Wahlnachfrage werden 1000 Wähler befragt. Frage: Mit welcher Wahrscheinlichkeit weicht die Prognose mehr als $\pm 2\%$ vom richtigen Wahlergebnis ab?

Ergebnis: Wahlumfrage

Gesucht: $P(80 \leq \text{Anzahl Partei } A \leq 120)$

Die Befragung ist ein Bernoulliexperiment mit 1000 Stufen und $p = 0.1$, $X =$"Anzahl Wähler der Partei A" ist $b_{1000,0.1}$-verteilt

$$\begin{aligned} P(80 \leq X \leq 120) &= \Phi\left(\frac{120 - 100 + 0.5}{\sqrt{90}}\right) - \Phi\left(\frac{80 - 100 - 0.5}{\sqrt{90}}\right) \\ &= \Phi(2.16) - \Phi(-2.16) = 2\Phi(2.16) - 1 \\ &= 2 \cdot 0.9846 - 1 = 0.9692 \end{aligned}$$

Also ist die Wahrscheinlichkeit für Abweichung $> \pm 2\%$: 0.0308

Beispiel: Toast und Butter

Problem: Nehmen wir die Hypothese als wahr an, dass eine Toastscheibe mit 50% Wahrscheinlichkeit auf die gebutterte Scheibe fällt. Mache 100 Experimente. Wie groß ist die Wahrscheinlichkeit, dass Toast mehr als 52 mal auf die Butterseite fällt?

Ergebnis: Toast und Butter

X ist $b_{100,0.5}$-verteilt, die Rechenregel ist anwendbar

$$\begin{aligned} P(X > 52) &= 1 - P(X \leq 52) \\ &= 1 - \Phi\left(\frac{52 - 50 + 0.5}{\sqrt{25}}\right) \\ &= 1 - \Phi(0.5) \\ &= 1 - 0.6915 \\ &= 0.3085 \end{aligned}$$

Die allgemeine Normalverteilung(Wichtig!!!)

Die Standardnormalverteilung $N(0,1)$ kann man auch "verschieben" und "stauchen/auseinanderziehen".

- Ist $X \sim N(0,1)$ standardnormalverteilt, so hat die Zufallsvariable

  $$Y = \sigma X + \mu$$

  Erwartungswert μ und Varianz σ^2.
- Die zugehörige Verteilung heißt (Graußsche) Normalverteilung $N(\mu, \sigma^2)$.
- Eine ZV Y ist $N(\mu, \sigma^2)$-verteilt genau dann wenn die Standardisierte Y^* standardnormalverteilt ist.

Normalverteilung: Berechnung

Sei X eine $N(\mu, \sigma^2)$-verteilte Zufallsvariable. Dann gilt:

- Verteilungsfunktion F:

$$F(x) = P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

- Dichtefunktion:

$$N(x|\mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Rechenregel: Für $np > 5$ und $n(1-p) > 5$ ist eine $b_{n,p}$-verteilte Zufallsvariable annähernd $\mathcal{N}(np, np(1-p))$-verteilt.

Beispiel: Kugellagerkugeln

Der Durchmesser von Kugellagerkugeln aus einer Produktion sei $\mathcal{N}(45, 0.01^2)$-verteilt. Eine Kugel ist unbrauchbar, wenn sie um mehr als 0.03mm vom Sollwert 45mm abweicht. Wie groß ist Wahrscheinlichkeit für eine solche Abweichung?

Ergebnis: Kugellagerkugeln

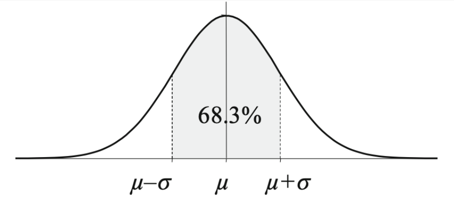
$$\begin{aligned} P(\text{Kugel brauchbar}) &= P(45 - 0.03 \leq X \leq 45 + 0.03) \\ &= \Phi\left(\frac{45.03 - 45}{0.01}\right) - \Phi\left(\frac{45 - 0.03 - 45}{0.01}\right) \\ &= \Phi(3) - \Phi(-3) \\ &= 2\Phi(3) - 1 \\ &= 2 \cdot 0.99865 - 1 \\ &= 0.9973 \end{aligned}$$

Also: 0.27% der Kugeln sind unbrauchbar.

Theorem

Ist X eine $N(\mu, \sigma^2)$-verteilte Zufallsvariable, so gilt für die Abweichungen vom Erwartungswert:

$$P(|X - \mu| \leq \sigma) \approx 0.6826$$
$$P(|X - \mu| \leq 2\sigma) \approx 0.9546$$
$$P(|X - \mu| \leq 3\sigma) \approx 0.9973$$



Theorem (Der zentrale Grenzwertsatz)

Eine Zufallsvariable, die sich aus vielen gleichverteilten Einflüssen additiv zusammensetzt, ist normalverteilt.

Konkret: Seien $X_1, X_2, \ldots, X_n$ unabhängige gleichverteilte Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 und

$$S_n = X_1 + X_2 + \ldots + X_n.$$

Dann ist S_n für große n (Regel: ab $n \geq 30$) annähernd $\mathcal{N}(n\mu, n\sigma^2)$-verteilt.

Beispiel: Würfeln mit Grenzwertsatz

Die Zufallsvariable "Würfeln mit einem Würfel" ist gleichverteilt mit $E(X) = 3.5$ und $Var(X) = 2.92$. Würfeln wir 1000 Mal und ist X_i das Ergebnis des i-ten Wurfes.

Wie groß ist die Wahrscheinlichkeit, dass die Augensumme um mehr als 100 vom Erwartungswert abweicht?

Ergebnis: Würfeln mit Grenzwertsatz

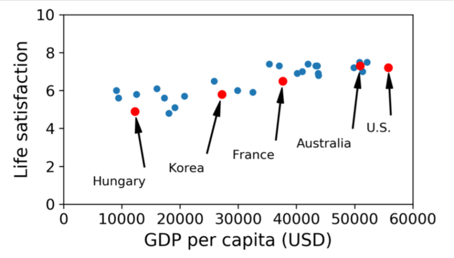
Die Augensumme $S_{1000} = X_1 + X_2 + \ldots + X_{1000}$ ist eine $\mathcal{N}(3500, 2920)$-verteilte Zufallsvariable.

$$
\begin{aligned}
P(3400 \leq S_{1000} \leq 3600) &= \Phi\left(\frac{3600 - 3500}{54}\right) - \Phi\left(\frac{3400 - 3500}{54}\right) \\
&= \Phi(1.85) - \Phi(-1.85) \\
&= 2\Phi(1.85) - 1 \\
&= 0.9356
\end{aligned}
$$

Also Wahrscheinlichkeit für größere Abweichung ca. 6.4%.

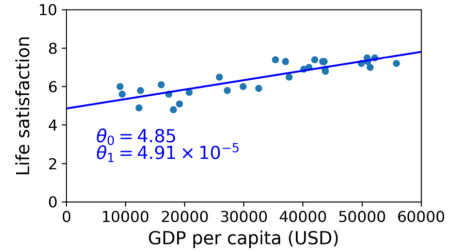
Anwendung der Normalverteilung in ML

Angenommen, wir haben Daten, die wie folgt aussehen:



Welche Vorhersage würden wir für einen Punkt auf der x-Achse, der nicht erfasst ist, machen? $\rightarrow$ lege "Durchschnittsgerade" $f(x)$ durch die Daten $=$ Lineare Regression

Lineare Regression: Statistische Betrachtung



In der Realität sind die Daten niemals alle auf einer Geraden - sie schwanken um diese Gerade herum $\rightarrow$ Gaußsche Normalverteilung mit Erwartungswert bei Input $x = f(y)$:

$$N(\mu = f(x), \sigma^2)$$

Exponentialverteilung

Exponentialverteilung: Motivation

Die Exponentialverteilung ist die stetige Variante der geometrischen Verteilung. Sie beschreibt die Wartezeit bis zum Eintreten eines Ereignisses.

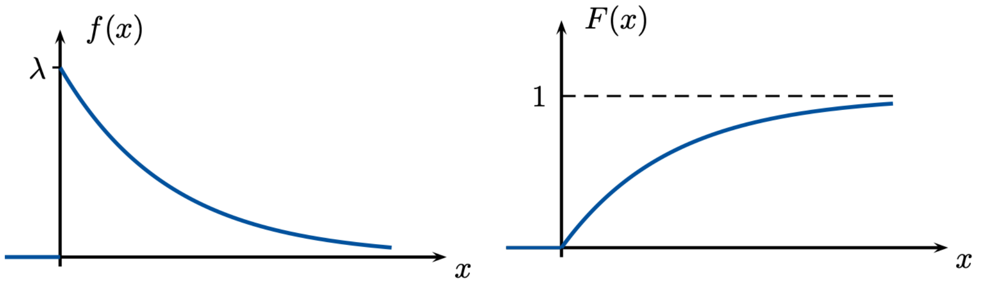


Abbildung 7: Dichte und Verteilungsfunktion der Exponentialverteilung

Exponentialverteilung

Eine ZV X ist exponentialverteilt mit Parameter λ, $X \sim \text{Exp}(\lambda)$ mit:

- Dichtefunktion:

$$f(x) := \begin{cases} \lambda e^{-\lambda x}, & \text{falls } x > 0 \\ 0 & \text{sonst} \end{cases}$$

- Verteilungsfunktion:

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{falls } x > 0, \\ 0 & \text{sonst} \end{cases}$$

Es gilt:

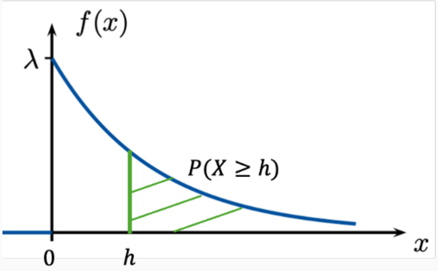
$$E(X) = \frac{1}{\lambda}, \quad Var(X) = \frac{1}{\lambda^2}$$

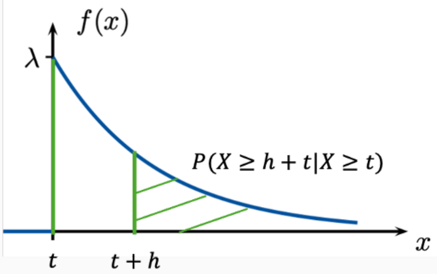
Gedächtnislosigkeit der Exponentialverteilung

Exponentialverteilung ist die einzige gedächtnislose stetige Verteilung:

$$P(X \geq t + h | X \geq t) = P(X \geq h).$$

Die Wahrscheinlichkeit, noch Zeitspanne h bis zum Eintreten warten zu müssen, wird nicht kleiner, wenn man bereits t lange gewartet hat!





Herleitung: Gedächtnislosigkeit

Es ist $P(X \leq t) = 1 - e^{-\lambda t}$. Dann ist $P(X \geq t) = 1 - P(X \leq t) = e^{-\lambda t}$. Also

$$\begin{aligned} P(X \geq t + h | X \geq t) &= \frac{P(\{X \geq t + h\} \cap \{X \geq t\})}{P(\{X \geq t\})} \\ &= \frac{P(X \geq t + h)}{P(X \geq t)} \\ &= \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} \\ &= e^{-\lambda h} = P(X \geq h). \end{aligned}$$

Beispiel: Festplattenausfall

Der Hersteller einer Festplatte gibt als mittlere Zeit bis zum Ausfall, die „Mean Time To Failure“ (MTTF), einen Wert von 70 Jahren an.

- Wie groß ist die Wahrscheinlichkeit, dass die Festplatte in einem Server im nächsten Jahr bzw. in den nächsten zwei Jahren ausfällt?
- Wie groß ist die Wahrscheinlichkeit, dass in 50 Rechnern, die im Dauerbetrieb laufen, alle Festplatten ein Jahr beziehungsweise zwei Jahre ohne Ausfall überstehen?

Ergebnis: Festplattenausfall

$\text{MTTF} = 70 = \text{Erwartungswert } 1/\lambda \Rightarrow \lambda = 1/70 = 0.014286$

- $P(X \leq 1) = 1 - e^{-0.014} = 1 - 0.9858 = 0.0142$ (ca. 1.4%)
  $P(X \leq 2) = 1 - e^{-2 \cdot 0.014} = 1 - 0.9718 = 0.0282$ (ca. 2.8%)
- $P(X > 1) = 0.9858, P(X > 2) = 0.9718$. Da die Festplattenausfälle voneinander unabhängig sind gilt:

$$P(\text{alle 50 Platten halten länger als 1 Jahr}) = 0.9858^{50} = 0.489$$

$$P(\text{alle 50 Platten halten länger als 2 Jahre}) = 0.9718^{50} = 0.239$$

Chi-Quadrat-Verteilung

Die χ^2-Verteilung spielt eine sehr wichtige Rolle in der Testtheorie (wenn man nachprüfen will, ob Daten wirklich Stichproben einer angenommenen Verteilung sind).

Definition (χ^2-Verteilung)

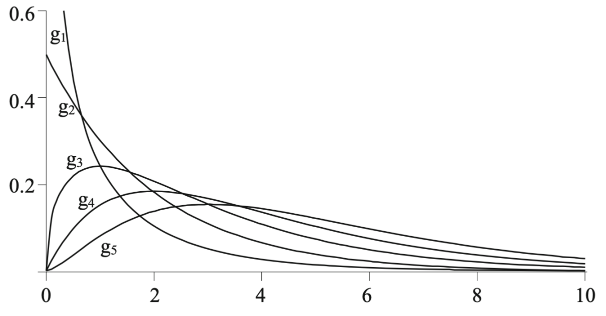
Sind die Zufallsvariablen $X_1, X_2, \ldots, X_n$ unabhängig und $\mathcal{N}(0,1)$-verteilt, so heißt die Verteilung der Zufallsvariablen

$$\chi_n^2 = X_1^2 + X_2^2 + \ldots + X_n^2$$

χ^2-Verteilung mit n Freiheitsgraden.

Dichtefunktion

Die χ^2-Verteilung ist Verteilung einer stetigen Zufallsvariablen. Die zum Grad n gehörige Dichtefunktion $g_n(x)$ ist von n abhängig und lässt sich aus der Dichte der Standardnormalverteilung mit Hilfe vollständiger Induktion berechnen.



$$g_1(x) = \frac{1}{\sqrt{2\pi x}} \cdot e^{-\frac{x}{2}}, g_2(x) = \frac{1}{2} \cdot e^{-\frac{x}{2}},$$
$$g_3(x) = \sqrt{\frac{x}{2\pi}} \cdot e^{-\frac{x}{2}}, g_4(x) = \frac{x}{4} \cdot e^{-\frac{x}{2}},$$
$$g_5(x) = \sqrt{\frac{x^3}{18\pi}} \cdot e^{-\frac{x}{2}}$$

Eigenschaften der χ^2-Verteilung

Theorem

Erwartungswert und Varianz der χ^2-Verteilung sind:

$$E(\chi_n^2) = n, \quad Var(\chi_n^2) = 2n$$

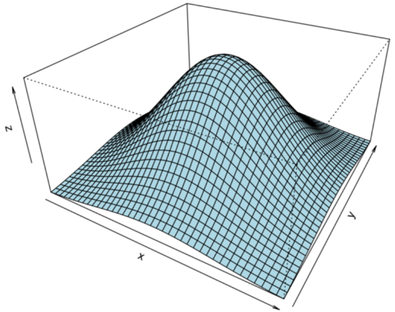
für große n ist

$$\chi_n^2 \sim \mathcal{N}(n, 2n)\text{-verteilt.}$$

Multivariate Verteilungen

Anschaulich: Die mehrdimensionale "Gauß-Glocke"

Angenommen, wir schauen uns nur den Zufallsvektor (X, Y) der Abweichungen in $x-$ und $y-$ Richtung an, dann hat dieser eine Wahrscheinlichkeitsdichte, die eventuell so aussieht:



Dies ist die multivariate Normalverteilung in zwei Dimensionen. Die Wahrscheinlichkeit für eine Abweichung in einer gewissen Fläche wäre das Volumen unter dem Graphen.

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y}$$

Multivariate Dichten und Verteilungen

Sind $X_1, \ldots, X_d$ stetige Zufallsvariablen auf (Ω, P), hat man

- multivariate Wahrscheinlichkeitsdichte: eine integrierbare Funktion $f : \mathbb{R}^d \to \mathbb{R}$
- multivariate Verteilungsfunktion $F(x_1, \ldots, x_d) \colon \mathbb{R}^d \to [0, 1]$:

$$P(X_1 \leq x_1, \ldots, X_d \leq x_d) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_d} f(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_d) d\tilde{x}_1 d\tilde{x}_2 \ldots d\tilde{x}_d$$

Im Kontrast dazu nennt man die bisher betrachteten Verteilungen univariat.

Multivariate Verteilung für unabhängige stetige Zufallsvariablen

Für unabhängige stetige Zufallsvariablen gilt:

$$f(x_1, \ldots, x_d) = f(x_1) \cdot \ldots f(x_d)$$

$$F(x_1, \ldots, x_d) = F(x_1) \cdot \ldots F(x_d)$$

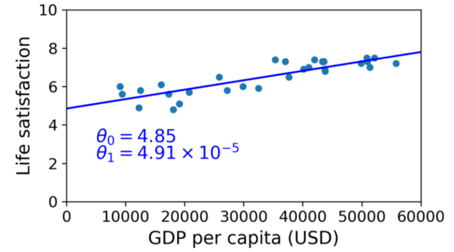
Nicht jedoch für abhängige!

Die multivariate (Gaußsche) Normalverteilung

Für Zufallsvektoren $\mathbf{X} = (X_1, \ldots, X_n)$ mit Erwartungswert $\mu = (\mu_1, \ldots, \mu_d)$ und Kovarianzmatrix Σ ist die multivariate (Gaußsche) Normalverteilung gegeben durch die Dichte:

$$N(\mathbf{x}|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right)$$

Anwendung im Machine Learning

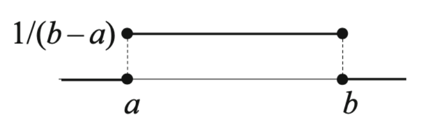


Wir erinnern uns an die Anwendung der Normalverteilung bei Linearer Regression: Die Daten sind um die Gerade herum normalverteilt.

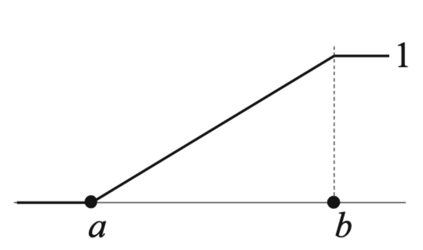
Wenn die Input-Daten mehrdimensional sind, braucht man die multivariate Normalverteilung um die Verteilung der Daten zu beschreiben.

$$N(\mu = \mathbf{f}(x), \Sigma)$$

Überblick stetiger Verteilungen: stetige Gleichverteilung

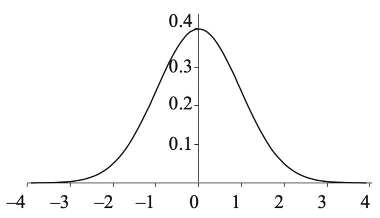


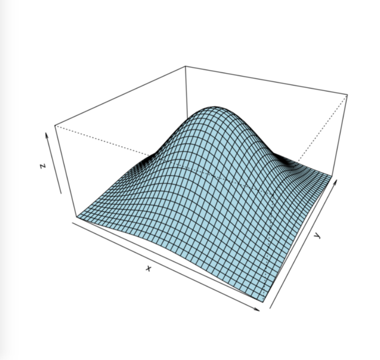
$$f : \mathbb{R} \to \mathbb{R}, x \mapsto \begin{cases} 0 & x < a \text{ oder } x > b \\ \frac{1}{b-a} & a \leq x \leq b \end{cases}.$$



$$F : \mathbb{R} \to [0,1], x \mapsto \begin{cases} 0 & x < a \\ 1 & x > b \\ \frac{x-a}{b-a} & a \leq x \leq b \end{cases}$$

$$E(X) = \frac{a+b}{2}, Var(X) = \frac{(b-a)^2}{12}$$





Univariate Normalverteilung: $N(x|\mu,\sigma^2)$ mit Erwartungswert μ, Varianz σ^2:

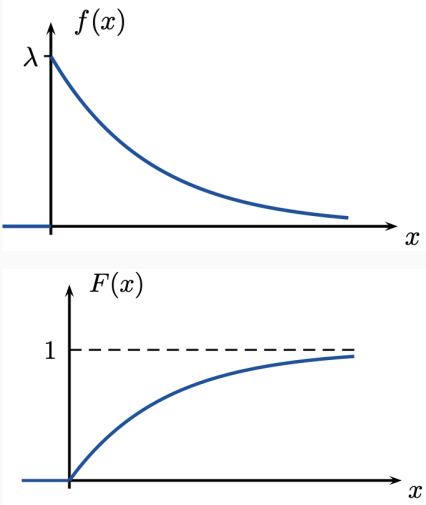
$$N(x|\mu,\sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$F(x) = P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Multivariate Normalverteilung:

$$N(\mathbf{x}|\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^d det(\Sigma)}} \cdot$$

$$\cdot \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)$$



$$f(x) := \begin{cases} \lambda e^{-\lambda x}, & \text{falls } x > 0 \\ 0 & \text{sonst} \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{falls } x > 0, \\ 0 & \text{sonst} \end{cases}$$

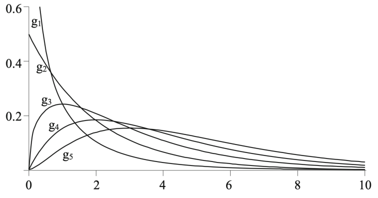
$$E(X) = \frac{1}{\lambda}, \quad Var(X) = \frac{1}{\lambda^2}$$

Überblick stetiger Verteilungen: Chi-Quadrat Verteilung

- Chi-Quadrat-Verteilung = Verteilung von $X_1^2 + \ldots + X_n^2$ für normalverteilte $X_1, \ldots, X_n$:

$$E(\chi_n^2) = n, \quad Var(\chi_n^2) = 2n$$

-



Verständnisfragen

1. Unter welchen Umständen kann die Binomialverteilung durch die Poisson-Verteilung angenähert werden?
2. Unter welchen Umständen kann die Binomialverteilung durch die Normalverteilung angenähert werden?
3. Warum tritt die Normalverteilung in vielen Anwendungsfällen auf?
4. Was sind die Gemeinsamkeiten von geometrischer Verteilung und Exponentialverteilung?
5. Die Wartezeit auf ein zufällig vorbeikommendes Taxi ist exponentialverteilt. Wie sieht es mit der Wartezeit auf einen Linienbus aus, der alle 10 Minuten fährt, wenn man nicht weiß, wann genau?
6. Wie hängen Binomialverteilung und hypergeometrische Verteilung zusammen?

Lösung

1. Für große n und kleine p so dass $np \leq 10$, $n > 1500p$.
2. Für $np > 5$ und $n(1-p) > 5$ ist eine $b_{n,p}$-verteilte Zufallsvariable annähernd $\mathcal{N}(np, np(1-p))$-verteilt.
3. Immer, wenn man die Ergebnisse vieler Zufallsvariablen aufsummiert, ist das Ergebnis normalverteilt.
4. Gedächtislosigkeit.
5. Wenn man einfach zufällig an die Bushaltestelle geht, ist die Wartezeit stetig gleichverteilt.
6. Binomialverteilung: ziehen mit Zurücklegen, hypergeometrische Verteilung: Ziehen ohne Zurücklegen, das heißt für große N nähern sich beide an.

Literatur

- Hartmann, Peter; Mathematik für Informatiker, Springer-Vieweg; 7. Auflage; 2019
- Henze, Norbert; Stochastik für Einsteiger; Springer; 10. Auflage; 2013