4. Spezielle Verteilungen

Lernziele:

- Sie unterscheiden die Binomial- und die hypergeometrische Verteilung.
- Sie unterscheiden die diskrete Poisson- und die stetige Exponentialverteilung, die über den Paramter λ zusammenhängen.
- Sie kennen die graphischen Darstellungen von Dichten der Gleich-, Exponential-, Normal-, Chiquadrat- und t-Verteilung.
- Sie kennen die Dichten der Gleich-, Exponential- und Normalverteilung und die Bedeutung ihrer Parameter.
- Sie wissen, für welche Anwendungsbereiche die oben genannten diskreten und stetigen Verteilungen geeignet sind, und transferieren diese Modelle auf andere Problemstellungen.
- Sie können Wahrscheinlichkeiten, Quantile und Zufallszahlen der oben genannten Verteilungen mit R berechnen bzw. generieren.

Literatur:

- Teschl Band 2, Kap. 28 + 29
- Arens et al., Kap 39
- Zucchini, Kap. 5 + 6

4.1 Diskrete Verteilungen

4.1.1 Bernoulliverteilung

- **Anwendungsmodell:**
  Indikatorvariable mit den Werten 1 bei Erfolg und 0 bei Misserfolg
- **Wahrscheinlichkeit:**
  $P(X = 1) = p$, $P(X = 0) = 1 - p$
- **Verteilung:**
  $X \sim B_{1,p}$
- **Erwartungswert:**
  $E[X] = p \quad = 1 \cdot p + 0 \cdot (1-p)$
- **Varianz:**
  $Var[X] = p(1-p) \quad = E[X^2] - (E[X])^2 = p - p^2$

  $E[X^2] = 1^2 \cdot p + 0^2 \cdot (1-p) = p$

4.1 Diskrete Verteilungen

4.1.2 Binomialverteilung

- **Anwendungsmodell:**
  Anzahl der Erfolge beim n-maligen Ziehen **mit** Zurücklegen

  Trefferwahrscheinlichkeit ändert sich nicht
- **Wahrscheinlichkeit:**
  $P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}, \quad k \in \{0, 1, \ldots, n\}$

  Anzahl der Pfade mit k-mal Treffer
- **Verteilung:**
  hat Verteilung

  $X \sim B_{n,p}$

  Parameter der Verteilung
- **Erwartungswert:**
  $E[X] = np$
- **Varianz:**
  $Var[X] = np(1-p)$
- **R-Funktionen:**

`dbinom(k,n,p)` $= P(X = k)$:	Wahrscheinlichkeits-/Dichtefunktion
`pbinom(k,n,p)` $= F(k)$:	Verteilungsfunktion
`qbinom(q,n,p)`:	q-Quantil
`rbinom(k,n,p)`:	k binomialverteilte Zufallszahlen

Beispiel 4.1.2: n-maliges Ziehen mit Zurücklegen

Ein Kommunikationsnetz mit n unabhängig voneinander arbeitenden Komponenten ist funktionsfähig, wenn mind. die Hälfte der Komponenten funktioniert. Die Wahrscheinlichkeit für die Funktionsfähigkeit einer Komponente ist 10%. $X \sim B_{n,0.1}$

(a) Mit welcher Wahrscheinlichkeit ist ein Netz mit 3 bzw. 5 Komponenten funktionsfähig?

(b) Für welche Werte von p ist die Wahrscheinlichkeit, dass ein System mit 5 Komponenten funktioniert größer als bei einem System mit 3 Komponenten?

(a) $n=3$: $P(X \geq 2) = P(X=2) + P(X=3) = \binom{3}{2} 0.1^2 \cdot 0.9 + \binom{3}{3} 0.1^3$
$= 1 - P(X \leq 1) = 1 - \text{pbinom}(1,3,0.1) \approx 2{,}8\%$

$n=5$: $P(X \geq 3) = P(X=3) + P(X=4) + P(X=5)$
$= 1 - P(X \leq 2) = 1 - \text{pbinom}(2,5,0.1) \approx 0{,}9\%$

(b) $\boxed{\begin{array}{l}\text{Für welches } p \text{ gilt:}\\ 1-\text{pbinom}(2,5,p) > 1-\text{pbinom}(1,3,p)\end{array}}$

Ansatz ist wichtig

n=5: $\underbrace{\binom{5}{3}p^3(1-p)^2}_{P(X=3)}+\underbrace{\binom{5}{4}p^4(1-p)}_{P(X=4)}+\underbrace{p^5}_{P(X=5)} > \binom{3}{2}p^2\cdot(1-p)+\binom{3}{3}p^3 \quad | : p^2$

$$10\,p\,(1-2p+p^2)+5p^2(1-p)+p^3 > 3(1-p)+p$$

$$\vdots$$

$$2p^3-5p^2+4p-1>0 \iff \underbrace{(2p^2-3p+1)}_{2(p-\frac{1}{2})(p-1)}(p-1)>0$$

$$2\left(p-\tfrac{1}{2}\right)\cdot\underbrace{(p-1)^2}_{>0}>0 \implies p>\tfrac{1}{2}$$

Für $p>\frac{1}{2}$ ist das Kommunikat.-netz mit 5 Komp. funktionsfähiger!

4.1 Diskrete Verteilungen

4.1.3 Hypergeometrische Verteilung

- **Anwendungsmodell:**
  Anzahl der Erfolge beim n-maligen Ziehen **ohne** Zurücklegen aus einer Menge mit M Elementen, die Erfolg bedeuten, und N Elementen, die Misserfolg bedeuten z.B. Lotto "6 aus 49"
- **Wahrscheinlichkeit:**
  $$P(X = k) = \frac{\binom{M}{k} \cdot \binom{N}{n-k}}{\binom{M+N}{n}}, \quad k \in \{0, 1, \ldots, \min\{n, M\}\}$$
- **Verteilung:**
  $X \sim H_{M,N,n}$ Parameter der Verteilung
- **Erwartungswert:**
  $E[X] = n \frac{M}{M+N}$ $\hat{=}$ Trefferwahrscheinlichkeit
- **Varianz:** $\to 1$, falls $n \ll M+N$
  $$Var[X] = \underbrace{n \frac{M}{M+N} \left(1 - \frac{M}{M+N}\right)}_{\hat{=}\text{ Binomialverteilung}} \frac{M+N-n}{M+N-1}$$
- **R-Funktionen:**
  dhyper(k,M,N,n) $= P(X = k)$
  phyper(k,M,N,n) $= F(k)$

Falls $20n \leq M + N$ und $M + N$ groß, dann ist der Unterschied zwischen Ziehen ohne bzw. mit Zurücklegen unwesentlich, so dass die Binomialverteilung mit $p = \frac{M}{M+N}$ als Approximation für die hypergeometrische Verteilung verwendet werden kann.

Beispiel 4.1.3: $M+N = 100$, $M = 10$, $N = 90$

Eine Charge mit 100 Glühbirnen wird angenommen, wenn sich in einer Stichprobe von $\underset{=n}{22}$ Glühbirnen maximal 2 defekte Glühbirnen befinden. Die Defektrate der Glühbirnen sei 10%. defekt $\hat{=}$ Erfolg

Wie groß ist die Wahrscheinlichkeit, dass die Charge angenommen wird?

Ges.: $P(X \leq 2) = \text{phyper}(2, 10, 90, 22) \approx 61.7\%$

$= P(X=0) + P(X=1) + P(X=2)$

4.1 Diskrete Verteilungen

4.1.4 Poisson-Verteilung

- **Anwendungsmodell:** "Verteilung der seltenen Ereignisse"
  Häufigkeit punktförmiger Ereignisse in einem Kontinuum
  Die durchschnittlich zu erwartende Anzahl der Erfolge λ pro Maßeinheit (i. a. Zeiteinheit) sei bekannt.
- **Wahrscheinlichkeit:**
  $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ mit $\sum_{k=0}^{\infty} P(X = k) = 1$, da $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$
- **Verteilung:**
  $X \sim P_\lambda$
- **Erwartungswert:**
  $E[X] = \lambda$, da $\sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} \lambda \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \lambda$
- **Varianz:**
  $Var[X] = \lambda$
- **R-Funktionen:**
  $\texttt{dpois(k,}\lambda\texttt{)} = P(X = k)$
  $\texttt{ppois(k,}\lambda\texttt{)} = F(k)$

Falls $n \geq 50, p \leq 0.1$ und $np \leq 10$, dann kann die Poisson-Verteilung mit $\lambda = np$ als Approximation für die Binomialverteilung verwendet werden.

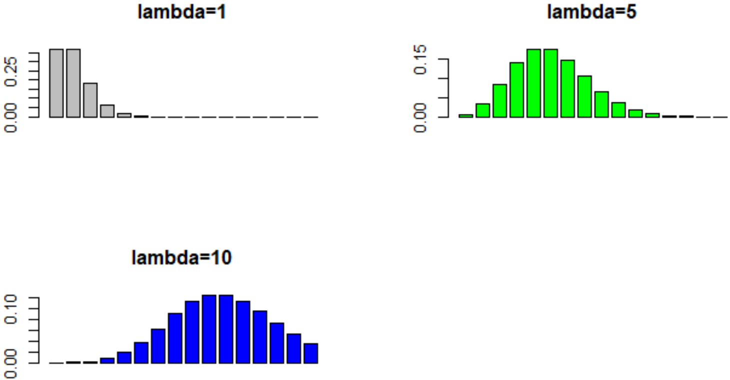

Abbildung: Poisson-Verteilungen für verschiedene Werte von λ

Beispiel 4.1.4:
Bei einer Hotline rufen im Durchschnitt $\overset{\lambda}{\overset{\shortparallel}{3}}$ Kunden $\overbrace{\text{pro Tag}}^{\text{Kontinuum}}$ an.

(a) Mit welcher Wahrscheinlichkeit ruft an einem bestimmten Tag genau ein Kunde an?

(b) Mit welcher Wahrscheinlichkeit rufen an einem bestimmten Tag mehr als 10 Kunden an?

a) Ges.: $P(X=1) = \mathrm{dpois}(1,3) \approx 14.9\,\%$

$$= \frac{3^1}{1!} e^{-3} = 3e^{-3}$$

b) Ges.: $P(X>10) = 1 - P(X \leq 10) = 1 - \mathrm{ppois}(10,3) \approx 0.03\%$

4.1 Diskrete Verteilungen

4.1.5 Gleichverteilung

- **Anwendungsmodell:**
  Alle Werte $\{x_1, \dots, x_n\}$ einer ZV X sind gleichwahrscheinlich.
- **Wahrscheinlichkeit:**
  $P(X = x_k) = \frac{1}{n}$
- **Verteilung:**
  $X \sim U_{\{x_1,\dots,x_n\}}$
- **Erwartungswert:**
  $E[X] = \frac{1}{n} \sum_{k=1}^{n} x_k = \bar{x}$
- **Varianz:**
  $Var[X] = \frac{1}{n} \sum_{k=1}^{n} x_k^2 - \bar{x}^2$
- **R-Funktion:**
  `sample(1:N,n)`: n Zufallszahlen zwischen 1 und N

4.2 Stetige Verteilungen

4.2.1 Stetige Gleichverteilung

- **Anwendungsmodell:**
  Zufallszahlen aus einem Intervall $[a, b]$
- **Dichte:**
  $f(x) = \frac{1}{b-a}$ für $x \in [a, b]$
- **Verteilung:**
  $X \sim U_{[a,b]}$
- **Erwartungswert:**
  $E[X] = \frac{a+b}{2}$
- **Varianz:**
  $Var[X] = \frac{(b-a)^2}{12}$
- **R-Funktionen:**
  `dunif(x,a,b)` $= f(x)$
  `punif(x,a,b)` $= F(x)$
  `runif(n)`: n Zufallszahlen zwischen 0 und 1

$$f(x) = \begin{cases} \frac{1}{30}, & \text{für } x \in [0,30] \\ 0, & \text{sonst} \end{cases}$$

Beispiel 4.2.1:

An einer Haltestelle fahren die Busse im 15 Minutentakt um 7:00, 7:15 usw. Ein Fahrgast kommt zu einem zufälligen Zeitpunkt zwischen 7:00 und 7:30 Uhr an die Haltestelle.

X: "Anzahl der Minuten, die Fahrgast nach 7:00 Uhr an Haltestelle ankommt."

(a) Wie groß ist die Wahrscheinlichkeit, dass er weniger als 5 Minuten warten muss?

(b) Wie groß ist die Wahrscheinlichkeit, dass er mindestens 12 Minuten warten muss?

punif(15,0,30) − punif(10,0,30)

(a) $P(\text{"Wartezeit} < 5\text{"}) = P(X \in]10;15[) + P(X \in]25;30[)$

$= \int_{10}^{15} f(x)\,dx + \int_{25}^{30} \frac{1}{30}\,dx = 2 \cdot \frac{5}{30} = \frac{1}{3}$

(b) $P(X \in]0;3[) + P(X \in]15;18[) = 2 \cdot \frac{3}{30} = \frac{1}{5}$

4.2 Stetige Verteilungen

4.2.2 Normalverteilung

- **Anwendungsmodell:**
  Beschreibt viele reale Situationen, ist insbesondere Grenzverteilung unabhängiger Summen
- **Dichte:**
  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$
- **Verteilung:**
  $X \sim N_{\mu,\sigma^2}$
- **Erwartungswert:**
  $E[X] = \mu$
- **Varianz:**
  $Var[X] = \sigma^2$
- **R-Funktionen:**
  `dnorm`$(\mathtt{x}, \mu, \sigma) = f(x)$
  `pnorm`$(\mathtt{x}, \mu, \sigma) = F(x)$
  `qnorm`$(\mathtt{q}, \mu, \sigma)$: q-Quantil

Eigenschaften:

- Maximalstelle von $f(x)$ bei $x = \mu$
- Wendestellen von $f(x)$ bei $x = \mu \pm \sigma$
- $X \sim N_{\mu,\sigma^2} \implies \overbrace{\underline{aX + b}}^{Y} \sim N_{\underbrace{a\mu+b}_{\mu_Y},\underbrace{a^2\sigma^2}_{\sigma_Y^2}}$ und $\frac{X-\mu}{\sigma} \sim N_{0,1}$
- $X_1 \sim N_{\mu_1,\sigma_1^2}$ und $X_2 \sim N_{\mu_2,\sigma_2^2} \implies X_1 + X_2 \sim N_{\mu_1+\mu_2,\sigma_1^2+\sigma_2^2}$

X_1, X_2 stochastisch unabh.

! Trans- formierte ZV sind wieder normalverteilt

Standardnormalverteilung $N_{0,1}$

- Dichte: $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$
- Verteilung: $\Phi(x) = \int\limits_{-\infty}^{x} \varphi(t)\,\mathrm{d}t$
- Quantile: Wegen der Achsensymmetrie von $\varphi(x)$ gilt:
  $\Phi(-x) = 1 - \Phi(x) \implies -x_p = x_{1-p}$

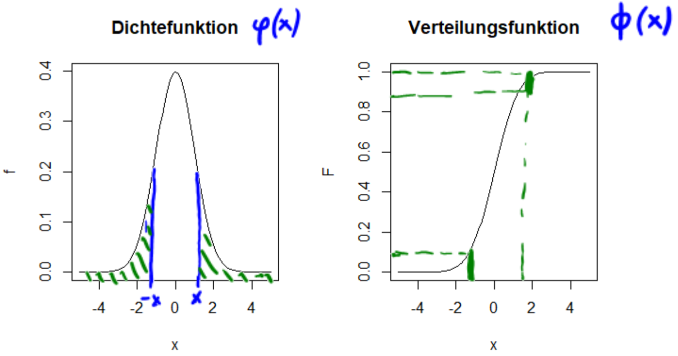

Abbildung: Dichte- und Verteilungsfkt. der Standardnormalverteilung

Beispiel 4.2.2:
Laut Klimadatenbank kann die Niederschlagsmenge M_x, die im Jahr x in Rosenheim fallen wird, durch eine normalverteilte ZV mit $\mu = 64$ und $\sigma^2 = 36$ modelliert werden.
Wie groß ist die Wahrscheinlichkeit, dass

$P(M_{2020} > 70) = 1 - P(M_{2020} \leq 70)$
$= 1 - \text{pnorm}(70, 64, 6)$

(a) die Niederschlagsmenge im Jahr 2020 > 70 ist?
(b) die Gesamtniederschlagsmenge in den beiden Jahren 2020 und 2021 > 140 ist?
(c) für die (unabhängigen) Niederschlagsmengen M_{2018} und M_{2019} gilt: $M_{2018} > M_{2019} + 16$?

(b) $P(M_{2020} + M_{2021} > 140)$ $\qquad M_{2020} + M_{2021} \sim N_{128, 72}$
$= 1 - \text{pnorm}(140, 128, \text{sqrt}(72)) \approx 7,9\,\%$

(c) $P(M_{2018} > M_{2019} + 16) = P(M_{2018} - M_{2019} > 16)$

$= P(M_{2018} + (-M_{2019}) > 16) = 1 - \text{pnorm}(16, 0, \text{sqrt}(72))$

$\approx 3\%$

$M_{2018} + (-M_{2019}) \sim N_{0,72}$

$E[-M_{2019}] = -E[M_{2019}] = -64$

$Var[-M_{2019}] = (-1)^2 \, Var[M_{2019}] = 36$

4.2 Stetige Verteilungen

4.2.3 Exponentialverteilung

- **Anwendungsmodell:** Modellierung von Lebensdauern, Wartezeiten
  Sei $Y_t \sim P_{\lambda t}$ im Intervall $[0, t]$ von t Zeiteinheiten, dann beschreibt die Exponentialverteilung die Wartezeit X bis zum Eintreten eines Ereignisses.
- **Dichte und Verteilungsfunktion:**
  $f(x) = \lambda e^{-\lambda x}\,(x \geq 0)$ und $F(x) = 1 - e^{-\lambda x}$
- **Verteilung:**
  $X \sim Exp_\lambda$
- **Erwartungswert:** (Berechnung mit partieller Integration)
  $E[X] = \frac{1}{\lambda}$
- **Varianz:** (Berechnung mit partieller Integration)
  $Var[X] = \frac{1}{\lambda^2}$
- **R-Funktionen:**
  $\texttt{dexp(x,}\lambda\texttt{)} = f(x)$
  $\texttt{pexp(x,}\lambda\texttt{)} = F(x)$

Eigenschaft:

- Eine exponentialverteilte ZV X ist gedächtnislos, d. h. $P(X > s + t | X > t) = P(X > s)$

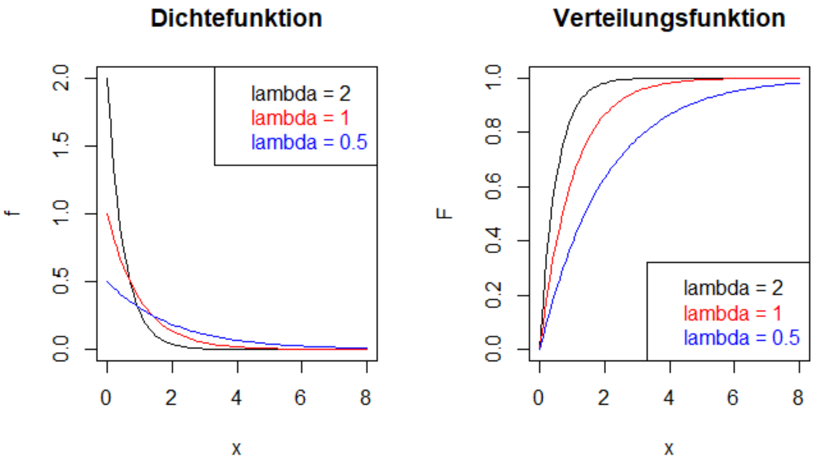

Abbildung: Dichte- und Verteilungsfkt. der Exponentialverteilung für $\lambda = 0.5$, 1 bzw. 2

Beispiel 4.2.3:

Die Lebensdauer einer Autobatterie entspreche einer Reichweite von 10000 km. Wie groß ist dann die Wahrscheinlichkeit, dass die Batterie bei einer 5000 km langen Fahrt nicht ausfällt?

4.2.4 Chiquadrat-Verteilung

$Z_1, \dots, Z_n$ seien unabhängige, standardnormalverteilte ZV $\Longrightarrow$
$X = Z_1^2 + \dots + Z_n^2$ hat Chiquadratverteilung mit n Freiheitsgraden

- **Anwendungsmodell:**
  Summen unabhängiger, standardnormalverteilter ZV
- **Verteilung:**
  $X \sim \chi_n^2$
- **Erwartungswert:**
  $E[X] = n$
- **Varianz:**
  $Var[X] = 2n$
- **R-Funktionen:**
  `dchisq(x,n)` $= f(x)$
  `pchisq(x,n)` $= F(x)$

Eigenschaft:

- $X_1 \sim \chi^2_{n_1}$ und $X_2 \sim \chi^2_{n_2} \implies X_1 + X_2 \sim \chi^2_{n_1+n_2}$

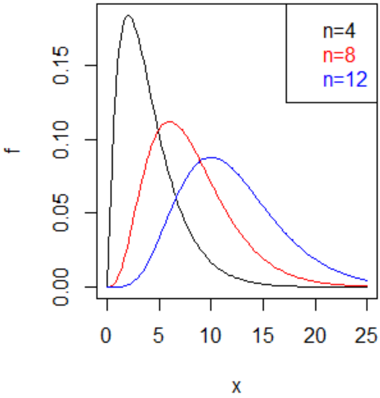

Abbildung: Dichtefunktionen der Chiquadratverteilung

Beispiel 4.2.4:
Es soll ein Zielpunkt in einem dreidimensionalen Raum getroffen werden. Der Messfehler X_i in jeder der 3 Koordinaten sei normalverteilt mit $\mu = 0$ und $\sigma = 2$ [m].
Wie groß ist die Wahrscheinlichkeit, dass der Abstand $D = \sqrt{X_1^2 + X_2^2 + X_3^2}$ zwischen Mess- und Zielpunkt größer als 3 [m] ist?

4.2.5 t-Verteilung

$Z \sim N_{0,1}$ und $X \sim \chi_n^2 \implies Y = \frac{Z}{\frac{X}{\sqrt{n}}}$ ist t-verteilt mit n Freiheitsgraden

- **Anwendungsmodell:**
  Schätz- und Testverfahren bei unbekannter Varianz
- **Verteilung:**
  $Y \sim t_n$
- **Erwartungswert:**
  $E[Y] = 0$ für $n > 1$
- **Varianz:**
  $Var[Y] = \frac{n}{n-2}$ für $n > 2$
- **R-Funktionen:**
  `dt(y,n)`
  `pt(y,n)`

Eigenschaften:

- Für $n \to \infty$: $t_n \to N_{0,1}$
- Achsensymmetrie der Dichtefunktion $\Longrightarrow -x_p = x_{1-p}$

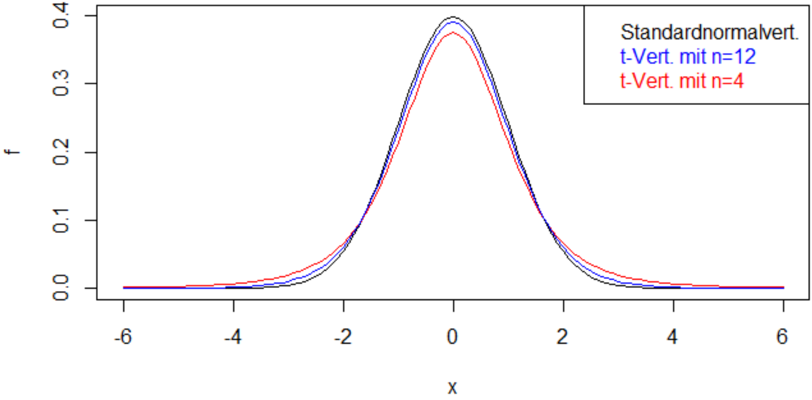

Abbildung: Dichtefunktionen der t-Verteilung