4.3 Poisson - Verteilung

Situation: Die Poisson - Verteilung stellt eine gute Näherung für die Binomialverteilung mit vielen Wiederholungen (d.h. n groß) und kleiner Erfolgswahrscheinlichkeit p dar.
Meist ist dann n und p nicht bekannt, sondern nur die durchschnittlich zu erwartende Anzahl der Erfolge $\lambda = \underbrace{n \cdot p}_{\text{EW der Binomialvert.}}$.

"Verteilung der seltenen Ereignisse"

Beispiele für Poisson - verteilte ZVen:

(i) Anzahl der Druckfehler auf einer Buchseite

λ: durchschnittl. Anzahl der Druckfehler auf einer Seite, z.B. 2

n: Anzahl der Zeichen auf einer Seite, z.B. $n = 4000$

p: Wahrscheinlichkeit für Druckfehler, z.B. $p = \frac{1}{2000}$

(ii) Anzahl der Transistoren, die am ersten Tag ausfallen

Verteilung: $X \sim P_\lambda$ mit λ: durchschnittl. zu erwartende Anzahl der Treffer

Wahrscheinlichkeitsfkt.:

$$p(x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda} \quad \text{für } x \in \{0, 1, 2, \dots\}$$

Reihendarstellung für $e^\lambda = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!}$

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda} \cdot \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} \cdot e^{\lambda} = 1$$

Erwartungswert und Varianz:

$$E[X] = \lambda$$
$$Var[X] = \lambda$$

R-Funktionen:

dpois(x, λ) $= p(x)$

ppois(x, λ) $= F(x)$

Approximation der Binomialverteilung durch Poissonverteilung:

Erwartungswert der Binomialverteilung: $n \cdot p$

Für $X_n \sim B_{n,\frac{\lambda}{n}}$ ($\frac{\lambda}{n} = p$) (n groß $\Rightarrow p = \frac{\lambda}{n}$ klein) gilt:

$$\lim_{n\to\infty} P(X_n = x) = \lim_{n\to\infty} \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

$$= \lim_{n\to\infty} \underbrace{\binom{n}{x}\left(\frac{\lambda}{n}\right)^x}_{\to \frac{\lambda^x}{x!}} \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}} \cdot \overbrace{\left(1-\frac{\lambda}{n}\right)^{-x}}^{\to 1}$$

$\left(\lim_{n\to\infty} \left(1+\frac{\lambda}{n}\right)^n = e^{\lambda}\right)$

$$= \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

Vorteil: Für große n erspart man sich die Berechnung der Binomialkoeff. $\binom{n}{x}$.

Faustregel für die "Zulässigkeit" dieser Approximation:

$n \geq 50$, $p \leq 0.1$ und $\lambda = n \cdot p \leq 10$

Beispiel: Bei einer Hotline rufen im Durchschnitt 3 ($\lambda = 3$) Kunden pro Tag an.

(a) Wie groß ist die Wahrscheinlichkeit, dass heute genau 1 Kunde anruft?

$P(X=1) = \text{dpois}(1,3) \approx 14.9\,\%$

(b) Wie groß ist die Wahrscheinlichkeit, dass mehr als 10 Kunden anrufen?

$P(X>10) = 1 - P(X \leq 10) = 1 - F(10) = 1 - \text{ppois}(10,3)$

$\approx 0.03\,\%$

4.4 Gleichverteilung

4.4.1 Diskrete Gleichverteilung

Situation: Alle Werte $\{x_1, x_2, \dots, x_n\}$ einer diskreten ZV X sind gleichwahrscheinlich.

Beispiele: (i) Werfen eines Würfels

(ii) Ziehen eines Loses aus einer Urne

Verteilung: $X \sim U_{\{x_1, \dots, x_n\}}$

Wahrscheinlichkeitsfkt.: $p(x_i) = \frac{1}{n}$

Erwartungswert und Varianz:

$$E[X] = \sum_{i=1}^{n} \frac{1}{n} x_i = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} \quad (\text{Mittelwert})$$

$$Var[X] \underset{\text{Verschiebungssatz}}{=} \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

R-Funktion: sample(...) erzeugt gleichverteilte Zufallszahlen

4.4.2 Stetige Gleichverteilung

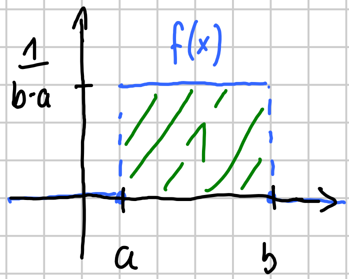
Situationen: (i) Zufallszahlen aus einem Intervall $[a,b]$

(ii) Ankunftszeit bei einer Party oder Haltestelle in einem Zeitintervall $[t_1, t_2]$

Verteilung: $X \sim U_{[a,b]}$

Dichtefunktion:

$$f(x) = \begin{cases} \frac{1}{b-a} & , \; x \in]a,b[\\ 0 & , \; \text{sonst} \end{cases}$$



Bed. an Dichte: $\int_{-\infty}^{\infty} f(x)\, dx = 1 = \int_a^b \frac{1}{b-a}\, dx$

Verteilungsfkt.:
ist stetig

$$F(x) = \begin{cases} 0 & , \; \text{für } x < a \\ \int_a^x f(t)\, dt = \frac{x-a}{b-a} & , \; \text{für } a \leq x < b \\ 1 & , \; \text{für } x \geq b \end{cases}$$

Erwartungswert und Varianz:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \frac{1}{b-a} \int_a^b x\, dx = \frac{\overset{=(b+a)(b-a)}{b^2 - a^2}}{2(b-a)} = \frac{a+b}{2}$$

Mittelpunkt des Intervalls $[a,b]$

$$Var[X] = \frac{(b-a)^2}{12} \quad \text{(s. 7. Übung)}$$

R-Funktionen: dunif(x, a, b)
punif(x, a, b)

Beispiel: An einer Haltestelle fahren die Busse im 15 Minutentakt um 7:00, 7:15 usw.
Ein Fahrgast kommt zu einem zufälligen Zeitpunkt zw. 7:00 und 7:30 Uhr an die Haltestelle.
Wie groß ist die Wahrscheinlichkeit, dass er
(a) weniger als 5 Minuten
(b) mindestens 12 Minuten warten muss?

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

$$X \in \left[\underset{a}{\underset{\parallel}{0}}, \underset{b}{\underset{\parallel}{30}}\right]$$

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

$$= \underbrace{\int_{10}^{15} \frac{1}{30}\, dx}_{\text{punif}(5,0,30)} + \int_{25}^{30} \frac{1}{30}\, dx = 2 \cdot \frac{1}{6} = \frac{1}{3}$$

b) $$P(X \in \,]0;3[\,) + P(X \in \,]15;18[\,) = 2 \cdot \frac{1}{10} = \frac{1}{5}$$

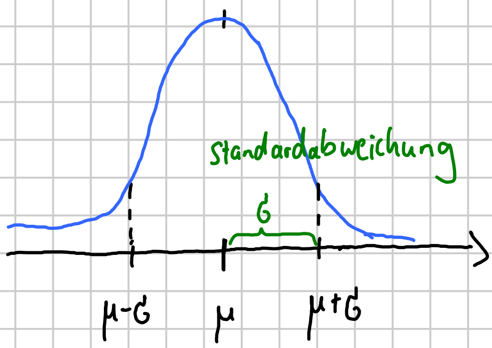
4.5 Normalverteilung

In der Praxis am häufigsten auftretende Verteilung (symmetrisch um den EW und rasch abfallend).

Eine ZV X hat eine Normalverteilung mit dem Lageparameter μ und dem Streuungsparameter $\sigma > 0$, wenn die Dichtefkt. lautet

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

"Gaußsche Glockenkurve"



Verteilung: $X \sim N_{\mu,\sigma^2}$ mit $\mu = E[X]$, $\sigma^2 = Var[X]$

Wichtige Eigenschaften:

(1) Maximalwert der Dichte bei $x = \mu$

(2) Wendepunkte der Dichtefkt. bei $x = \mu \pm \sigma$

(3) $X \sim N_{\mu,\sigma^2} \Longrightarrow aX + b \sim N_{a\mu+b,\, a^2\sigma^2}$

Falls X normalverteilt ist, dann ist auch $aX+b$ normalverteilt.

$$\frac{1}{\sigma}X - \frac{\mu}{\sigma} = \frac{X-\mu}{\sigma} \sim N_{0,1}$$ Standardnormalverteilung

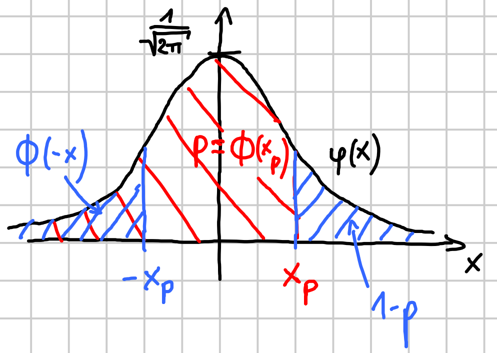
(4) $X_1 \sim N_{\mu_1,\sigma_1^2}$, $X_2 \sim N_{\mu_2,\sigma_2^2}$, X_1, X_2 unabhängig $\Longrightarrow X_1 + X_2 \sim N_{\mu_1+\mu_2,\, \sigma_1^2+\sigma_2^2}$

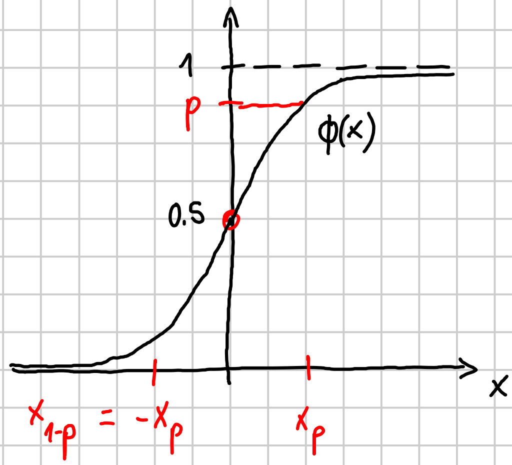
Die Summe von normalverteilten ZVen ist wieder normalverteilt

Bemerkung: Jede normalverteilte ZV X kann durch die Standardisierung $\frac{X-\mu}{\sigma}$ auf Standardnormalvert. mit EW 0 und Varianz 1 gebracht werden:

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

Verteilung $\phi(x) = \int_{-\infty}^{x} \varphi(t)\, dt$ kann nicht analytisch angegeben werden, sondern nur numerisch berechnet werden





Wegen der Symmetrie von $\varphi(x)$ gilt:

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

$\Rightarrow$ Für die p-Quantile: $-x_p = x_{1-p}$ bzw. $x_p = -x_{1-p}$

R-Funktionen:

Beachte: hier Standardabweichung

dnorm (x, μ, σ)

pnorm (x, μ, σ)

Defaultwerte: 0,1

Beispiel: Laut Klimadatenbank kann die Niederschlagsmenge M_x, die im Jahr x in Rosenheim fallen wird, durch eine normalverteilte ZV mit $\mu = 64$ [cm] und $\sigma^2 = 36$ [cm²] modelliert werden.

Wie groß ist die Wahrscheinlichkeit, dass

(a) die Niederschlagsmenge im nächsten Jahr mehr als 70 [cm] ist,

(b) die Gesamtniederschlagsmenge in den beiden nächsten Jahren größer als 140 [cm] ist,

(c) die Niederschlagsmenge M_{2017} größer ist als $M_{2018} + 16$. (M_{2017} und M_{2018} seien unabh.)

(a) $$P(M_{2019} > 70) = 1 - P(M_{2019} \leq 70) = 1 - \mathrm{pnorm}(70, 64, 6) \approx 15.9\,\%$$

$M_{2019} \sim N_{64,36}$

(b) $$P(M_{2019} + M_{2020} > 140) = 1 - \mathrm{pnorm}(140, 128, \mathrm{sqrt}(72)) \approx 7.9\,\%$$

Summe von normalverteilten ZV ist wieder normalverteilt mit

$$E[M_{2019} + M_{2020}] = E[M_{2019}] + E[M_{2020}] = 128$$

M_{2019}, M_{2020} unabh. $\Rightarrow$ $\mathrm{Var}[\ldots] = \mathrm{Var}[M_{2019}] + \mathrm{Var}[M_{2020}] = 72$

$\sigma = \sqrt{72}$

(c) $$P(M_{2017} > M_{2018} + 16) = P(M_{2017} - M_{2018} > 16) =$$

$$E[M_{2017} + (-M_{2018})] = 64 - 64 = 0$$

$$\mathrm{Var}[M_{2017} + (-M_{2018})] = \mathrm{Var}[M_{2017}] + (-1)^2\,\mathrm{Var}[M_{2018}] = 72$$

$$= 1 - \mathrm{pnorm}(16, 0, \mathrm{sqrt}(72)) \approx 3\,\%$$