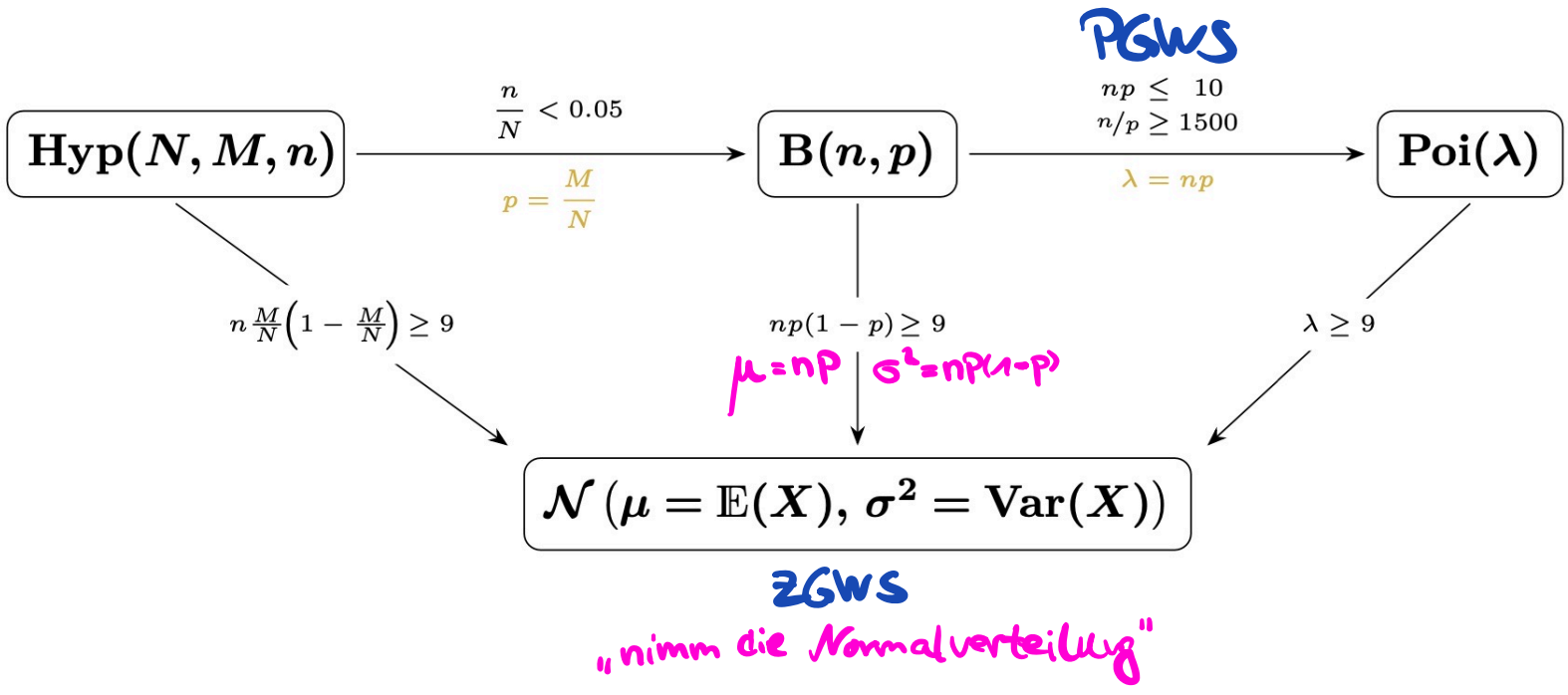
Approximation von Verteilungen

Ausgewählte Approximationen & Bedingungen



Aufgabe: Vor der mündlichen Prüfung gibt ein Mathematik Professor einen Fragenkatalog raus. Der Prüfling bereitet sich auf 90 dieser Fragen vor und lässt nach dem Prinzip „Mut zur Lücke“ die restlichen 10 Fragen aus. In der Prüfung werden zufällig drei Fragen aus dem Katalog gewählt. Mit welcher Wahrscheinlichkeit wählt der Professor mindestens eine der nicht vorbereiteten Aufgaben aus? Berechnen Sie die Wahrscheinlichkeit exakt und mit einer geeigneten Näherung.

$X = \{$ Anzahl nicht vorbereiteten Aufgaben bei $n = 3$ Fragen $\}$

$X \sim \mathrm{Hyp}(N=100, M=10, n=3)$

$$\mathbb{P}(X=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X=0) = 1 - \frac{\binom{10}{0}\cdot\binom{90}{3}}{\binom{100}{3}}$$

$$= \frac{67}{245} \approx 0{,}273469\ldots$$

Näherung: $\frac{n}{N} = \frac{3}{100} = 0{,}03 < 0{,}05$ ✓

$$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$\Rightarrow \mathrm{Hyp}(N=100, M=10, n=3) \overset{*}{\approx} B(n=3, p=\frac{M}{N}=0{,}1)$

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X=0) = 1 - \binom{3}{0}\cdot 0{,}1^0 \cdot 0{,}9^3 = 0{,}271$$

$Cov(X,Y) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = Var(X)$

Rechenregeln für Erwartungswert & Varianz

X, Y Zufallsvariablen, $a \in \mathbb{R}$ const.

$\mathbb{E}(X \cdot Y) - \mathbb{E}(X)\,\mathbb{E}(Y)$

(1) $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

(2) $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}(X)$

(3) $\mathbb{E}(a) = a$

(1) $Var(X+Y) = Var(X) + Var(Y) + 2 \cdot Cov(X,Y)$

(2) $Var(a \cdot X) = a^2 \cdot Var(X)$

Bsp: $Var(-X) = Var(X)$

(3) $Var(a) = 0$

Bsp: $X_1, \dots, X_n$ i.i.d. (independent, identically distributed) mit $\mu := \mathbb{E}(X_i)$, $\sigma^2 := Var(X_i)$

$$S_n := \sum_{i=1}^{n} X_i$$

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) \overset{(1)}{=} \sum_{i=1}^{n} \underbrace{\mathbb{E}(X_i)}_{=\mu} = n\mu$$

$$Var(S_n) = Var\left(\sum_{i=1}^{n} X_i\right) \overset{(1)}{=} \sum_{i=1}^{n} \underbrace{Var(X_i)}_{\sigma^2} = n\sigma^2$$

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) \overset{(2)}{=} \frac{1}{n} \cdot \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) \overset{(1)}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} \underbrace{\mathbb{E}(X_i)}_{\mu} = \frac{1}{n} \cdot n\mu = \mu$$

$$Var(\overline{X}_n) = Var\left(\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right) \overset{(2)}{=} \frac{1}{n^2} \cdot Var\left(\sum_{i=1}^{n} X_i\right) \overset{(1)}{=} \frac{1}{n^2} \cdot \sum_{i=1}^{n} \underbrace{Var(X_i)}_{=\sigma^2} = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

Poissonscher Grenzwertsatz (PGWS)

"große n & kleine p"

- $n \in \mathbb{N}_{\geq 1}$: $X_n \sim B(n, p_n)$ mit $\lim_{n\to\infty} n \cdot p_n = \lambda > 0$
- $Y \sim \text{Poi}(\lambda)$

$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$ = Binomial $\qquad$ $\frac{\lambda^k}{k!} e^{-\lambda}$ = Poisson

$$\Rightarrow \forall k \in \mathbb{N}: \lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(Y = k)$$

Beweis:

$$\mathbb{P}(X_n = k) = \binom{n}{k} \cdot p_n^k \cdot (1-p_n)^{n-k} = \frac{n!}{k! \cdot (n-k)!} \cdot \left(\frac{np_n}{n}\right)^k \cdot \left(1 - \frac{np_n}{n}\right)^{n-k}$$

($n! = n \cdot (n-1) \cdot \ldots \cdot (n-k+1) \cdot (n-k)!$)

$$= \underbrace{\frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{n^k}}_{= \frac{n}{n} \cdot \frac{n-1}{n} \cdot \ldots \cdot \frac{n-k+1}{n}} \cdot \underbrace{\frac{(np_n)^k}{k!}}_{\to \frac{\lambda^k}{k!}} \cdot \underbrace{\left(1 - \frac{np_n}{n}\right)^n}_{\to e^{-\lambda}} \cdot \underbrace{\left(1 - \frac{np_n}{n}\right)^k}_{\to 1}$$

($\frac{n}{n} = 1$, $\frac{n-1}{n} \to 1$, $\frac{n-k+1}{n} \to 1$; $np_n \to \lambda$)

$$\xrightarrow{n\to\infty} \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \mathbb{P}(Y = k) \quad \forall k \in \mathbb{N}$$

Zentraler Grenzwertsatz (ZGWS)

- $X_1, X_2, X_3, \ldots$ i.i.d. ($n \geq 30$ gilt als geeignet bei allgemeinen Verteilungen)
- $\forall i \in \mathbb{N}: \mu := \mathbb{E}(X_i) < \infty, \quad \sigma^2 := \mathrm{Var}(X) < \infty$

$$\Rightarrow \quad Z_n := \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} \xrightarrow{n \to \infty} Z \sim \mathcal{N}(0,1)$$

$$Z_n := \frac{\overline{X_n} - \mathbb{E}(\overline{X_n})}{\sqrt{\mathrm{Var}(\overline{X_n})}} \xrightarrow{n \to \infty} Z \sim \mathcal{N}(0,1)$$

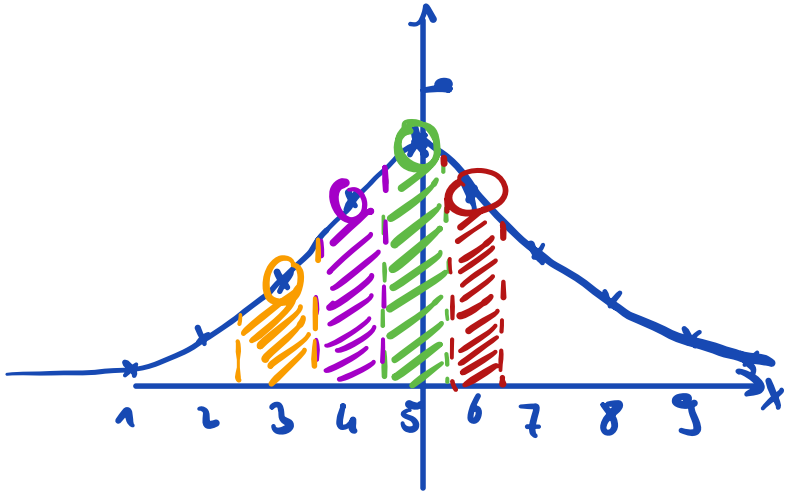
Diskret → Stetig (ZGWS)

$$X \sim \text{diskret} \overset{\text{ZGWS}}{\approx} N(\mu = E(X), \sigma^2 = Var(X))$$

(1) Intervall erzeugen

(2) Stetigkeitskorrektur

- obere X-Grenze „+0,5"
- untere X-Grenze „−0,5"



Aufgabe: Bei einer Briefkastenwerbeaktion geht eine Firma davon aus, dass es in 1 von 1000 Fällen zu einem Kaufvertrag kommt. Bei einer solchen Aktion wurden 10.000 Werbebriefe an zufällig ausgewählte Adressen verschickt. Wie groß ist die Wahrscheinlichkeit, dass mehr als zehn Kaufverträge zustande kommen? Nutzen Sie zur Berechnung die exakte Verteilung, den Poissonschen Grenzwertsatz und den Zentralen Grenzwertsatz. Welche der Näherungen sind geeignet?

$X = \{\text{Anzahl Kaufverträge bei } n = 10.000 \text{ Werbeaktionen}\}$

$X \sim B(n=10.000,\ p = \frac{1}{1000})$

$$P(X > 10) = 1 - P(X \leq 10) = 1 - \sum_{k=0}^{10} P(X=k)$$ (Binomial)

$$= 1 - \sum_{k=0}^{10} \binom{10.000}{k} \left(\frac{1}{1000}\right)^k \cdot \left(\frac{999}{1000}\right)^{10.000-k} = 0,41696023$$

PGWS: $B(n=10.000, p=\frac{1}{1000}) \approx Poi(\lambda = np = 10)$

$$P(X > 10) = 1 - P(X \leq 10) = 1 - \sum_{k=0}^{10} P(X=k)$$ (Poisson)

$$= 1 - \sum_{k=0}^{10} \frac{10^k}{k!} e^{-10} = 1 - e^{-10} \cdot \sum_{k=0}^{10} \frac{10^k}{k!} = 0,41696024$$

ZGWS: $P(X > 10) = P(\underbrace{11}_{a} \leq X \leq \underbrace{10000}_{b}) \overset{\text{ZGWS}}{\approx} \Phi\left(\frac{10.000 - 10 + 0{,}5}{\sqrt{9{,}99}}\right) - \Phi\left(\frac{11 - 10 - 0{,}5}{\sqrt{9{,}99}}\right)$

(Stetigkeitskorrektur)

$B(n=10.000, p=\frac{1}{1000}) \approx N(\mu = np = 10;\ \sigma^2 = np(1-p) = 9{,}99)$

$= \underbrace{\Phi(3160{,}85)}_{\approx 1} - \underbrace{\Phi(0{,}16)}_{\approx 0{,}563559}$

$= 0{,}436441$

Bei Normalverteilung

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

$P(X \geq a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$

(*) $P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$

① Intervall erzeugt

② $\mu = np$ und $\sigma^2 = np(1-p)$

③ wie bei Normalverteilung (*) + Stetigkeitskorrektur

Aufgabe: Bei jungen Wasserschildkröten ist die Überlebenswahrscheinlichkeit sehr gering. Nur jede Fünftausendste Wasserschildkröte eines Jahrgangs wird erwachsen. Angenommen es gibt in diesem Jahr 7500 Nachkommen der Wasserschildkröten. Berechnen Sie mit der exakten Verteilung und mit einer geeigneten Näherung die Wahrscheinlichkeit dafür, dass mehr als eine junge Wasserschildkröte dieses Jahrgangs erwachsen wird.

$X = \{$ Anzahl erwachsener Schildkröten bei $n = 7500$ Schildkröten $\}$

$X \sim B(n = 7500, p = \frac{1}{5000})$

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$P(X > 1) = 1 - P(X \leq 1) = 1 - (P(X=0) + P(X=1))$$

$$= 1 - \left(\binom{7500}{0} \cdot \left(\frac{1}{5000}\right)^0 \cdot \left(\frac{4999}{5000}\right)^{7500} + \underbrace{\binom{7500}{1}}_{7500} \cdot \left(\frac{1}{5000}\right)^1 \cdot \left(\frac{4999}{5000}\right)^{7499} \right) \approx \underline{0{,}44219}$$

PGWS geeignet?: $np = 1{,}5 \leq 10$ ✓

$\frac{n}{p} = \frac{7500}{\frac{1}{5000}} = 37\,500\,000 \geq 1500$ ✓

$B(n = 7500; p = \frac{1}{5000}) \approx Poi(\lambda = np = 1{,}5)$

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$P(X > 1) = 1 - P(X \leq 1) = 1 - (P(X=0) + P(X=1))$$

$$= 1 - \left(\frac{1{,}5^0}{0!} e^{-1{,}5} + \frac{1{,}5^1}{1!} e^{-1{,}5} \right) = 1 - 2{,}5 \cdot e^{-1{,}5} = \underline{0{,}44217}$$

Aufgabe: Wie groß ist die Wahrscheinlichkeit beim 1000-maligen Werfen einer Münze genau 500 mal das Ergebnis „Kopf“ zu erhalten? Nutzen Sie eine geeignete Näherung.

$X = \{$ Anzahl „Kopf“ bei $n = 1000$ Münzwürfen $\}$

$X \sim B(n=1000,\ p=0{,}5)$

$$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

(exakt: $\mathbb{P}(X=500) = \binom{1000}{500} \underbrace{0{,}5^{500} \cdot 0{,}5^{500}}_{0{,}5^{1000}} = 0{,}025225\ldots$)

ZGWS: $np(1-p) = 1000 \cdot 0{,}5 \cdot 0{,}5 = 250 \geq 9$ ✓ ist geeignet

$B(n=1000, p=0{,}5)$

$\approx$

$N(\mu = np = 500;\ \sigma^2 = np(1-p) = 250)$

$$\mathbb{P}(X=500) = \mathbb{P}(500 \leq X \leq 500) = \Phi\left(\frac{500-500+0{,}5}{\sqrt{250}}\right) - \Phi\left(\frac{500-500-0{,}5}{\sqrt{250}}\right)$$

$$= \Phi(0{,}03) - \underbrace{\Phi(-0{,}03)}_{=1-\Phi(0{,}03)} = 2 \cdot \underbrace{\Phi(0{,}03)}_{0{,}511967} - 1 = 0{,}023934$$

Summen normalverteilter Zufallsvariablen

- $\forall i \in \{1, \dots, n\}: \; X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ unkorreliert → linear unabhängig

$$\Rightarrow \; S_n := \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\frac{1}{n} \sum_{i=1}^{n} \mu_i, \frac{1}{n^2} \cdot \sum_{i=1}^{n} \sigma_i^2\right)$$

Anmerkung: X, Y unabhängig $\Rightarrow$ X, Y unkorreliert → nicht linear abhängig

X, Y korreliert $\not\Rightarrow$ X, Y abhängig

↙ linear abhängig

Aufgabe: Die drei Mitarbeiter eines Kleinunternehmens bekommen für die gefahrenen Kilometer zur Arbeitsstelle einen Fahrtkostenzuschuss von ihrem Arbeitgeber. Für eine Gesamtkostenanalyse des Unternehmens werden die als unabhängig und normalverteilt angesehenen gefahrenen Kilometer aus den vergangenen Jahren geschätzt. Es ergeben sich auf einen Monat gerechnet die Erwartungswerte von 20 km, 30 km und 50 km mit Standardabweichungen von 3 km, 4 km und 5 km. Wie groß ist die Wahrscheinlichkeit, dass alle Mitarbeiter zusammen in einem Monat mehr als 110 km gegenüber ihrem Arbeitgeber geltend machen?

$X_i = \{$ Gefahrene km von Mitarbeiter $i\}$

$X_1 \sim N(\mu_1 = 20; \sigma_1^2 = 3^2)$

$X_2 \sim N(\mu_2 = 30; \sigma_2^2 = 4^2)$

$X_3 \sim N(\mu_3 = 50; \sigma_3^2 = 5^2)$

$$S_3 = \sum_{i=1}^{3} X_i \sim N\left(\mu = \sum_{i=1}^{3} \mu_i = 100; \sigma^2 = \sum_{i=1}^{3} \sigma_i^2 = 50\right)$$

$$P(S_3 > 110) = 1 - \Phi\left(\frac{110 - E(S_3)}{\sqrt{Var(S_3)}}\right) = 1 - \Phi\left(\frac{110-100}{\sqrt{50}}\right)$$

$$= 1 - \Phi\left(\frac{10}{\sqrt{25 \cdot 2}}\right) = 1 - \Phi(\sqrt{2}) = 1 - \Phi(1{,}41) = 0{,}07927$$

0,920730