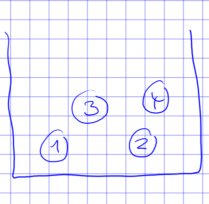
Unabhängigkeit von Zufallsvariablen:

X und Y heißen <u>unabhängig</u> falls für $\boxed{\text{alle}}$ $x \in X(\Omega)$, $y \in Y(\Omega)$ gilt

$$\boxed{\underbrace{\mathbb{P}(\{X=x\} \cap \{Y=y\})}_{= \mathbb{P}(X=x,\, Y=y)} = \mathbb{P}(X=x) \cdot \mathbb{P}(Y=y)}$$

<u>Beispiel</u>: Urnenmodell



Urne mit 4 nummerierten Kugeln

2x ziehen mit Zurücklegen

$$\Omega = \{ \underbrace{(w_1, w_2)}_{w} : w_1, w_2 \in \{1, \dots, 4\} \}$$

Laplace-Raum mit $\mathbb{P}(w) = \frac{1}{16}$

$X(w) = w_1 + w_2$ Summe der beiden Zahlen

$Y(w) = \min(w_1, w_2)$ Minimum der beiden Zahlen

$X(\Omega) = \{2, 3, \dots, 8\}$

$Y(\Omega) = \{1, \dots, 4\}$

<u>Frage</u>: Sind X und Y <u>unabhängig?</u>

Ereignisse der Form $\{X=x\}$, $x \in X(\Omega)$, $\{Y=y\}$, $y \in Y(\Omega)$ betrachten

z.B. $\{X=2,\ Y=2\}$

$\{X=2\}$, $\{Y=2\}$

$$\mathbb{P}(X=2,\ Y=2) = 0$$

$$\mathbb{P}(X=2) = \frac{1}{16} > 0$$

$$\mathbb{P}(Y=2) > 0$$

$$\mathbb{P}(X=2) \cdot \mathbb{P}(Y=2) \neq 0$$

$$\mathbb{P}(X=2,\ Y=2)$$

Damit sind X und Y nicht unabhängig, denn ich habe ein Gegenbeispiel gefunden.

$X = \sum_{k=1}^{n} Y_k$, $Y_k \sim Ber(p)$, die Y_k $\boxed{\text{unabhängige}}$ Zufallsvariablen

dann $X \sim Bin(n, p)$

<u>Letztes Mal</u>: $X \sim Ber(p)$

$\Rightarrow \mathbb{V}(X) = p(1-p)$

<u>jetzt</u>: Binomialverteilung

Sei $X \sim B(n, p)$.

Trick: $X = \sum_{k=1}^{n} Y_k$,

wobei die Y_k <u>unabhängige</u> Bernoulli-verteilte ZV mit Parameter p sind

$$\underline{\mathbb{V}(X)} = \mathbb{V}\left(\sum_{k=1}^{n} Y_k\right) \overset{\text{unabhängig}}{=} \sum_{k=1}^{n} \mathbb{V}(Y_k)$$

$$= \sum_{k=1}^{n} p(1-p) = \underline{\underline{n \cdot p(1-p)}}$$

<u>Bsp</u>: Varianz im Glücksspiel vom letzten Mal:

Glücksspiel faire Münze, Einsatz 1 €, 5 Wiederholungen

$$X_i = \begin{cases} 1 & \text{falls } i\text{-tes Spiel gewonnen} \\ -1 & \text{falls } i\text{-tes Spiel verloren} \end{cases}$$

Gewinn nach 5 Spielen: $X_1 + \dots + X_5$

die X_i sind <u>unabhängig</u>, $\mathbb{P}(X_i = 1) = \frac{1}{2}$

$$\mathbb{E}[X] = 0$$

$$\mathbb{V}(X) = \mathbb{V}(X_1) + \dots + \mathbb{V}(X_5)$$

$$\mathbb{V}(X_1) = \underbrace{\mathbb{E}[X_1^2]}_{=1} - \underbrace{\mathbb{E}[X_1]^2}_{=0} = \underline{\underline{1}}$$

$$\underline{\underline{\mathbb{V}(X) = 5}}$$

Einsatz auf <u>1000 €</u> erhöhen:

$$\tilde{X} = 1000 X_1 + \dots + 1000 X_5$$

$$\mathbb{E}[\tilde{X}] = 0$$

$$\mathbb{V}(\tilde{X}) = \mathbb{V}(1000 \cdot (X_1 + \dots + X_5))$$

$$= \mathbb{V}(1000 \cdot X)$$

Vorfaktor wird quadratisch

$$= \boxed{(1000)^2} \underbrace{\mathbb{V}(X)}_{=5}$$

$$= \underline{5'000'000}$$

also eine viel höhere Streuung als vorher.

<u>Beweis der Chebyshev-Ungleichung</u>

X diskret

$$\underline{\mathbb{V}(X)} = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

$$= \sum_{x \in X(\Omega)} \underbrace{(x - \mathbb{E}[X])^2}_{\geq 0} \underbrace{\mathbb{P}(X=x)}_{\geq 0}$$

Definition Erwartungswert

$$\geq \sum_{\substack{x \in X(\Omega): \\ |x - \mathbb{E}[X]| > a}} \underbrace{(x - \mathbb{E}[X])^2}_{\geq a^2} \mathbb{P}(X=x)$$

$$\geq a^2 \sum_{x : |x - \mathbb{E}[X]| > a} \mathbb{P}(X = x)$$

$$= a^2 \, \mathbb{P}(|X - \mathbb{E}[X]| > a)$$

umformen

$$\Longleftrightarrow \boxed{\mathbb{P}(|X - \mathbb{E}[X]| > a) \leq \frac{\mathbb{V}(X)}{a^2}} \quad \square$$

<u>Anwendung</u>: Bildchensammelproblem (coupon collector probl.)

n verschiedene Bilder, erhält zufällig Päckchen mit jeweils 1 Bild

X = Anzahl Päckchen, die man sammeln muss, um jedes der n Bilder mindestens 1x zu haben

Letztes Mal: $\mathbb{E}[X] \approx n \cdot \ln(n)$

<u>jetzt</u>: Suche eine Abschätzung für $\mathbb{P}(X \geq 2\,\mathbb{E}[X])$

<u>Idee</u>: Verwende die Chebyshev-Ungleichung

<u>Denn</u>: $\mathbb{P}(X \geq 2\mathbb{E}[X])$

$$= \mathbb{P}(X - \mathbb{E}[X] \geq \mathbb{E}[X])$$

$$\leq \mathbb{P}(|X - \mathbb{E}[X]| \geq \mathbb{E}[X])$$

$$= \mathbb{P}(|X - \mathbb{E}[X]| \geq n \ln(n))$$

Um Chebyshev anzuwenden, benötige ich die Varianz.

Trick vom letzten mal:

$$\boxed{X = \sum_{k=1}^{n} Y_k}, \quad Y_k \sim Geo\Big(\overbrace{\frac{n-k+1}{n}}^{=: p_k}\Big)$$

die Y_k sind unabhängig.

$$\mathbb{V}(Y_k) = \frac{1 - p_k}{p_k^2}$$

Varianz der geometrischen Verteilung.

$$\mathbb{V}(X) = \sum_{k=1}^{n} \mathbb{V}(Y_k) = \sum_{k=1}^{n} \frac{1 - p_k}{p_k^2}$$

$$= \sum_{k=1}^{n} \frac{k+1}{n \left(\frac{n-k+1}{n}\right)^2}$$

$$= \sum_{k=1}^{n} \frac{n^2 (k+1)}{n (n-k+1)^2}$$

$$= n^2 \underbrace{\sum_{k=1}^{n} \frac{\overbrace{k+1}^{\leq 1}}{n (n-k+1)^2}}_{\leq C < \infty}$$

$$\leq C \cdot n^2$$

$$\underline{\mathbb{P}(|X - \mathbb{E}[X]| \geq n \log n)} \underset{\text{Chebyshev}}{\leq} \frac{\mathbb{V}(X)}{(n \log n)^2}$$

$$\leq \frac{C \cdot n^2}{n^2 (\log n)^2}$$

$$\leq \underline{\frac{C}{(\log n)^2} \xrightarrow[n \to \infty]{} 0}$$

aus der Analysis:

$$\sum_{\ell=1}^{\infty} \frac{1}{\ell^2} \leq C < \infty$$

$$\sum_{k=1}^{n} \frac{k+1}{n (n-k+1)^2} \leq \sum_{k=1}^{n} \frac{1}{(\underbrace{n-k+1}_{=\ell})^2}$$

$$\leq \sum_{\ell=1}^{n} \frac{1}{\ell^2} < C$$