Agenda 12.1.24:

- Frohes neues Jahr! ü
- Kommende Übungstermine:
  - 26.1. Besprechung EA 6 & Fragen allgemein (⇝ gern per Mail)
  - 9.2. Besprechung EA 7 & Fragen

  Weitere Fragen können auch im Forum gestellt werden (alternativ: Sprechstunde)
- Alte Klausur ist im Forum zu finden. Achtung: Vorlesungsinhalte + Einsendeaufgaben haben sich in diesem Semester geändert, daher liebe Beispiele + Aufgaben aus diesem Semester zur Orientierung nehmen!
- Heute: EA 5 [1,2,3a & 4] + Tipps EA 6

---

1. Grundbegriffe (10 Punkte). Ein Call-Center möchte die minimale und maximale Anrufdauer schätzen. Wir nehmen dabei an, dass die Anrufzeiten unabhängig voneinander sind und dass Dauer jedes Anrufs gleichverteilt ist mit einer „natürlichen" Mindestdauer $a \geq 0$ und maximalen Dauer $b > a$ (zum Beispiel wenn es keine Frage gibt, die in unter a Minuten gelöst werden kann und keine Kund:innen, die länger als b Minuten Zeit haben). Das Call-Center beobachtet nun 1000 Gespräche, von denen das kürzeste 5 Minuten und das längste 137 Minuten gedauert hat. Daraus schließt das Call-Center, dass 5 und 137 Minuten die Mindest- und die Höchstdauer sind.

a) (4 Punkte) Geben Sie ein statistisches Modell (gemäß der Definition) an, das zur obigen Situation passt.

b) (2 Punkte) Handelt es sich um ein stetiges oder diskretes Modell?

c) (2 Punkte) Schreiben Sie in der Notation der Vorlesung auf, welcher Schätzer hier verwendet wurde.

d) (2 Punkte) Geben Sie eine Abbildung τ an, die jedem $\vartheta \in \Theta$ in ihrem statistischen Modell die Kenngröße „Erwartungswert der Anrufdauer" zuordnet. Geben Sie einen Schätzer für τ an.

Aus dem Skript:

Wir abstrahieren nun von diesem Beispiel. Sei dafür zuerst eine Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ gegeben, auf dem im Folgenden alle Zufallsvariablen definiert seien.

- Gegeben sei eine Zufallsvariable X mit Werten in einer Menge $\mathcal{X}$, die wir auch *Zustandsraum* von X nennen werden (im obigen Beispiel war $\mathcal{X} = \{0,1\}^n$).
- Die Zufallsvariable X besitze eine Verteilung, die von einem Parameter ϑ abhängt (im obigen Beispiel $X = (X_1, \ldots, X_n)$, X_i unabhängig mit $X_i \sim \mathrm{Ber}(\vartheta)$), wobei ϑ aus einer zulässigen Menge Θ von Parametern gewählt werden darf, also $\vartheta \in \Theta$ ($\vartheta \in [0,1]$).
- Ist X bezüglich ϑ verteilt, so bezeichnen wir diese Verteilung mit P_ϑ, d.h. $X \sim P_\vartheta = \mathbb{P}_X$. Für jedes $\vartheta \in \Theta$ ist P_ϑ also eine Wahrscheinlichkeitsverteilung auf $\mathcal{X}$.

Definition 5.2.2. Sei $\mathcal{X}$ eine Menge und $\mathcal{F}$ ein Ereignisfeld definiert auf $\mathcal{X}$. Sei $\{P_\vartheta\}_{\vartheta \in \Theta}$ eine Familie von Wahrscheinlichkeitsmaßen definiert auf dem messbaren Raum $(\mathcal{X}, \mathcal{F})$. Das Tripel $(\mathcal{X}, \mathcal{F}, \{P_\vartheta\}_{\vartheta \in \Theta})$ heißt *statistisches Modell* oder auch *statistischer Raum*. $\mathcal{X}$ heißt auch *Stichprobenraum*, Θ nennt man *Parameterraum*.

Wir werden im Folgenden auf die Erwähnung des Ereignisfeldes $\mathcal{F}$ verzichten und auch das Tupel $(\mathcal{X}, \{P_\vartheta\}_{\vartheta \in \Theta})$ als statistisches Modell bezeichnen.

↳ σ-Alg. siehe Lösungsskizze

A1) a)
- X_i = Anrufdauer des i-ten Gesprächs $i \in \{1, \ldots, 1000 = n\}$

  $X = (X_1, \ldots, X_{1000})$, $\mathcal{X} = [0, \infty)^{1000}$
- $\vartheta = (a,b)$ ist der (unbekannte) Parameter und zweidimensional

  $\Theta = \{(a,b) \mid 0 \leq a \leq b < \infty\}$

- Die X_i sind unter $\mathbb{P}_\vartheta = \mathbb{P}_{(a,b)}$ iid. gleichverteilt über $[a,b]$

  d.h. X_i hat Dichte

$$f_i(x_i) = \frac{1}{b-a} \mathbb{1}_{(a,b)}(x_i)$$

$X = (X_1, \ldots X_n)$ hat Dichte

$$f(x_1, \ldots x_n) \underset{\text{iid.}}{=} \prod_{i=1}^{n} f_i(x_i) = \prod_{i=1}^{n} \frac{1}{b-a} \mathbb{1}_{(a,b)}(x_i)$$

$$= \frac{1}{(b-a)^n} \underbrace{\prod_{i=1}^{n} \mathbb{1}_{(a,b)}(x_i)}_{= \mathbb{1}_{(a,b)^n}(x_1, \ldots x_n)} = \frac{1}{(b-a)^n} \mathbb{1}_{(a,b)^n}(x_1, \ldots x_n)$$

b) stetig oder diskret? beides ok! Hier haben wir ein stetiges Modell angenommen.

Definition 5.2.4. Sei $(\mathcal{X}, \{P_\vartheta\}_{\vartheta \in \Theta})$ ein statistisches Modell. Ist $\mathcal{X}$ eine höchstens abzählbare Teilmenge des $\mathbb{R}^d$, so sprechen wir von einem *diskreten statistischen Modell*. Ist $\mathcal{X}$ eine Teilmenge des $\mathbb{R}^d$ und besitzt jede Verteilung P_ϑ eine Dichtefunktion $\rho_\vartheta \colon \mathcal{X} \to [0, \infty)$, so nennen wir das Tripel ein *stetiges statistisches Modell*.

c)

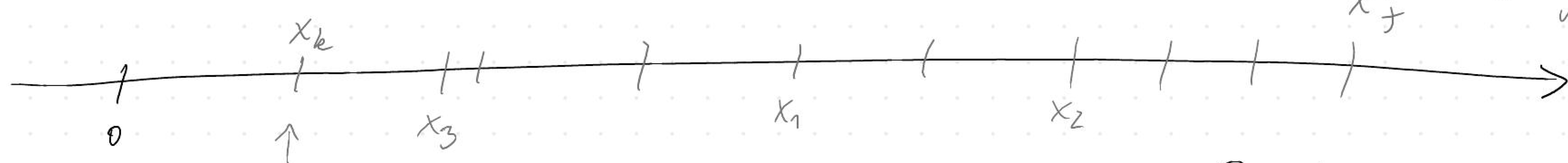
Definition 5.3.1. Eine Abbildung $T\colon \mathcal{X} \to \Theta$ heißt *Schätzer*, *Punktschätzer* oder auch *Parameterschätzer*.

Hier $\quad T: [0,\infty)^n \to \{(a,b) : 0 \le a \le b < \infty\}$

d.h. $\quad \underbrace{T(x_1, \ldots x_n)}_{\text{Anrufdauern}} = (\hat{a}, \tilde{b})$ für bestimmte $\hat{a}, \tilde{b}$.

Hier $\quad T(x_1, \ldots x_n) = \left(\min\limits_{i=1..n} x_i \,,\ \max\limits_{i=1,..n} x_i \right)$

beobachtete max Anrufdauer
= geschätzte mögliche max. Anrufdauer



beobachtete min Anrufdauer
= geschätzte mögliche min Anrufdauer

Ein anderer Schätzer könnte die Abbildung
$\tilde{T}(x_1, \ldots x_n) = (3,\ 5 \cdot \max\limits_{i=1,..n} x_i)$
sein (erfüllt aber i.A. keine Gütekriterien).

---

d)

Definition 5.3.4. Sei $(\mathcal{X}, \{P_\vartheta\}_{\vartheta\in\Theta})$ ein statistisches Modell. Sei $\tau\colon \Theta \to \mathbb{R}$ eine Abbildung, die jedem $\vartheta \in \Theta$ eine Kenngröße $\tau(\vartheta) \in \mathbb{R}$ zuordnet. Eine Abbildung $T\colon \mathcal{X} \to \mathbb{R}$ heißt dann *Schätzer für* τ.

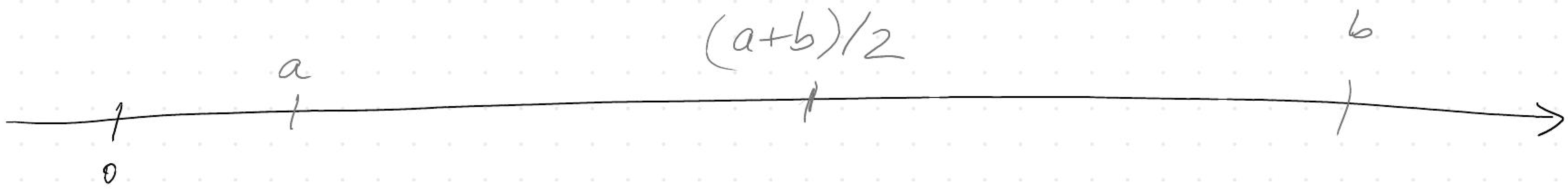
$\tau: \{(a,b) : 0 \le a \le b < \infty\} \to \mathbb{R}$

d.h. $\quad \tau(a,b) = \mathbb{E}_{a,b}[X_1]$ "Erwartungswert der Anrufdauer" als Funktion von (a,b).

Jedem Paar (a,b) wird der EW der Anrufdauer $\tau(a,b)$ zugeordnet.

Wir berechnen die rechte Seite:

$$\mathbb{E}_{a,b}[X_1] = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{2}\,\frac{b^2 - a^2}{b-a} = \frac{b+a}{2}. \Rightarrow \tau(a,b) = \frac{b+a}{2}$$



Eine Abbildung $U: \mathcal{X} \to \mathbb{R}$ heißt Schätzer für τ. $\to$ wieder viele Möglichkeiten

z.B. empirischer Mittelwert

$$U(x_1, \ldots, x_n) = \frac{1}{n} \sum_{k=1}^{n} x_k$$

oder $\quad U(x_1 \ldots x_n) = \tau(T(x_1, \ldots x_n))$ mit

$$U(x_1 \ldots x_n) = \frac{\min\limits_{i=1..n} x_i + \max\limits_{i=1..n} x_i}{2}$$

Gaußsches Produktmod.:
$(X_1 \ldots X_{1000})$ iid.
und jedes X_i ist normalverteilt
d.h. $F(x) = \int \frac{1}{\sqrt{\ }} \ldots$

Hier: Produktmodell
$(X_1 \ldots X_{1000})$ iid.
und jedes X_i ist gleichverteilt
d.h. $F(x) = \int_a^x \frac{1}{b-a}\,dy$

2. Erwartungstreue in Poisson-Modellen (10 Punkte). Wir betrachten als statistisches Modell ein Poisson-Modell $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0), \mathbb{P}_\theta)$ mit $\mathbb{P}_\theta(\{n\}) = \frac{\theta^n}{n!e^\theta}$ für alle $n \in \mathbb{N}_0$, also die Poisson-Verteilung.

a) (3 Punkte) Zeigen Sie, dass $T(x) = x$ ein erwartungstreuer Schätzer für θ ist.

b) (3 Punkte) Zeigen Sie, dass $T(x) = x^2 - x$ ein erwartungstreuer Schätzer für $\tau(\theta) = \theta^2$ ist.

Wir definieren $f : (0,\infty) \times \mathbb{N} \to (0,\infty)$ durch

$$f(x,n) = \frac{x^n}{n!(e^x - 1)}.$$

c) (2 Punkte) Es sei X eine Poisson-verteilte Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega', \mathcal{F}', \mathbb{P}')$ mit Parameter $\lambda > 0$. Zeigen Sie, dass $\mathbb{P}'[X = n \mid X > 0] = f(\lambda, n)$ für alle $n \in \mathbb{N}$ gilt.

Insbesondere ist für jedes $\lambda > 0$ durch $f(\lambda, n)$ eine Wahrscheinlichkeitsverteilung auf $\mathbb{N}$ (ohne Null) definiert. Wir betrachten nun als statistisches Modell ein bedingtes Poisson-Modell $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \tilde{\mathbb{P}}_\theta)$ mit $\tilde{\mathbb{P}}_\theta(\{n\}) = f(\theta, n)$.

d) (2 Punkte) Zeigen Sie, dass $T(x) = 1 + (-1)^x$ ein erwartungstreuer Schätzer für $\tau(\theta) = 1 - e^{-\theta}$ ist. Ist dies eine sinnvolle Wahl?

Tipp: $\sum_{j=0}^{\infty} \frac{\theta^j}{j!} = e^\theta$ $\quad k! = k\cdot(k-1)(k-2)\cdot\ldots\cdot 1 \quad 0! := 1$

Definition 5.3.12. Sei $(\mathcal{X}, \{P_\vartheta\}_{\vartheta\in\Theta})$ ein statistisches Modell. Ein Schätzer $T : \mathcal{X} \to \Theta$ heißt *erwartungstreu* oder *unverzerrt* (englisch: *unbiased*), wenn für jede P_ϑ-verteilte Zufallsvariable X gilt:

$$\mathbb{E}_\vartheta(T) = \vartheta.$$

Ist $T : \mathcal{X} \to \mathbb{R}$ ein Schätzer für eine Kenngröße $\tau : \Theta \to \mathbb{R}$, so heißt dieser erwartungstreu, wenn

$$\mathbb{E}_\vartheta(T) = \tau(\vartheta)$$

ist.

Man beachte, dass hier $\mathbb{E}_\vartheta(T) = \mathbb{E}(T(X))$ mit $X \sim P_\vartheta$ ist gemäß Definition 5.2.9.

a) Wir müssen $\mathbb{E}_\theta[T] = \mathbb{E}[T(X)] = \theta$ für $X \sim \mathbb{P}_\theta$ zeigen.

$k! = k\cdot(k-1)\cdot(k-2)\cdot\ldots\cdot 1$

$$\mathbb{E}_\theta[T] \overset{\text{Def}}{=} \mathbb{E}[T(X)] \overset{\text{EW}}{=} \sum_{k=0}^{\infty} k \cdot \frac{\theta^k}{k!\,e^\theta} = 0 + \sum_{k=1}^{\infty} k \frac{\theta^k}{k!\,e^\theta} = \sum_{k=1}^{\infty} \cancel{k} \frac{\theta\cdot\theta^{k-1}}{\cancel{k}\cdot(k-1)!\,e^\theta}$$

$$= \frac{\theta}{e^\theta} \sum_{k=1}^{\infty} \frac{\theta^{k-1}}{(k-1)!} \overset{\text{versch.}}{=} \frac{\theta}{e^\theta} \sum_{j=0}^{\infty} \frac{\theta^j}{j!} \overset{\text{Tipp}}{=} \frac{\theta}{\cancel{e^\theta}} \cancel{e^\theta} = \theta.$$

b) Wir müssen zeigen: $\mathbb{E}_\theta[T] = \theta^2$.

$$\mathbb{E}_\theta[T] = \sum_{k=0}^{\infty} T(k) \frac{\theta^k}{k!\,e^\theta} = \sum_{k=0}^{\infty} \underbrace{(k^2 - k)}_{= k(k-1)} \frac{\theta^k}{k!\,e^\theta}$$

$$= \underset{k=0}{0} + \underset{k=1}{0} + \sum_{k=2}^{\infty} \cancel{k}\cdot(\cancel{k-1}) \frac{\theta^2\cdot\theta^{k-2}}{\cancel{k}\cdot(\cancel{k-1})\cdot(k-2)!\,e^\theta}$$

$$= \frac{\theta^2}{e^\theta} \underbrace{\sum_{k=2}^{\infty} \frac{\theta^{k-2}}{(k-2)!}}_{= \left(\underset{k=2}{\frac{\theta^0}{0!}} + \underset{k=3}{\frac{\theta^1}{1!}} + \ldots\right)} \overset{\text{versch.}}{=} \frac{\theta^2}{e^\theta} \underbrace{\sum_{j=0}^{\infty} \frac{\theta^j}{j!}}_{= \left(\underset{j=0}{\frac{\theta^0}{0!}} + \underset{j=1}{\frac{\theta^1}{1!}} + \ldots\right)} = \frac{\theta^2}{e^\theta}\cdot e^\theta = \theta^2$$

Genauso für $\theta^3, \theta^4, \ldots$ usw!

c) $X \sim \mathrm{Poi}(\lambda) \implies \mathbb{P}'[X = n] = \frac{\lambda^n}{n!\,e^\lambda}$, $n \in \mathbb{N}_0$. Sei $n \in \mathbb{N}$ (ohne 0).

$$\mathbb{P}'[X = n \mid X > 0] \overset{\text{Def.}}{=} \frac{\mathbb{P}[X = n, X > 0]}{\mathbb{P}[X > 0]} \overset{n>0}{=} \frac{\mathbb{P}[X = n]}{\mathbb{P}[X > 0]}$$

$$= \frac{\mathbb{P}[X = n]}{1 - \mathbb{P}[X = 0]} = \frac{\frac{\lambda^n}{n!\,e^\lambda}}{1 - \frac{\lambda^0}{0!\,e^\lambda}} = \frac{\frac{\lambda^n}{n!\,e^\lambda}}{1 - e^{-\lambda}} = \frac{\lambda^n/n!}{e^\lambda(1 - e^{-\lambda})}$$

$$= \frac{\lambda^n/n!}{e^\lambda - 1} = f(\lambda, n).$$

d) Stat. Modell mit $\tilde{\mathbb{P}}_\theta[X = n] = f(\lambda, n)$. $T(x) = 1 + (-1)^x$ ist erw. tr. Schätzer für $\tau(\theta) = 1 - e^{-\theta}$:

$$\tilde{\mathbb{E}}_\theta[T(X)] = \sum_{k=1}^{\infty} T(k) f(\theta, k) = \sum_{k=1}^{\infty} (1 + (-1)^k)\cdot\frac{\theta^k/k!}{e^\theta - 1} = \text{rechnen} = 1 - e^{-\theta}$$

siehe Lösungsskizze

$$= \sum_{k=1}^{\infty} \frac{\theta^k/k!}{e^\theta - 1} + \sum_{k=1}^{\infty} \frac{(-\theta)^k/k!}{e^\theta - 1}$$

0 ergänzen

$$= \underbrace{\frac{\theta^0/0!}{e^\theta-1} + \sum_{k=1}^{\infty} \frac{\theta^k/k!}{e^\theta-1}}_{\sum_{k=0}} + \underbrace{\frac{(-\theta)^0/0!}{e^\theta-1} + \sum_{k=1}^{\infty} \frac{(-\theta)^k/k!}{e^\theta-1}}_{\sum_{k=0}} - 2\cdot\frac{1}{e^\theta-1}$$

$$= \frac{1}{e^\theta-1} \underbrace{\sum_{k=0}^{\infty} \frac{\theta^k}{k!}}_{e^\theta} + \frac{1}{e^\theta-1} \underbrace{\sum_{k=0}^{\infty} \frac{(-\theta)^k}{k!}}_{e^{-\theta}} - \frac{2}{e^\theta-1}$$

$$= \frac{e^\theta}{e^\theta-1} + \frac{e^{-\theta}}{e^\theta-1} - \frac{2}{e^\theta-1} = \frac{e^\theta - 1 + e^{-\theta} - 1}{e^\theta-1} = 1 + \underbrace{\frac{e^{-\theta}-1}{e^\theta-1}}_{= e^{-\theta}\frac{1-e^\theta}{e^\theta-1}}$$

$$= 1 - e^{-\theta}.$$

3. Maximum-Likelihood-Schätzer und Konfidenzintervall (10 Punkte). Wir betrachten Daten $X_1, \ldots, X_n$, die unabhängig und jeweils binomialverteilt mit Parametern (m, p) sind. Der Parameter $m \in \mathbb{N}$ sei bekannt und $p \in (0,1)$ sei unbekannt.

a) (5 Punkte) Zeigen Sie, dass $T(x) = \frac{1}{nm}\sum_{\ell=1}^{n} x_\ell$ der Maximum-Likelihood-Schätzer für p ist.

b) (2 Punkte) Sei $\alpha \in (0,1)$ und bezeichne Φ die Verteilungsfunktion der Standard-Normalverteilung. Zeigen Sie mit dem Zentralen Grenzwertsatz, dass für $\epsilon : (0,1) \to (0, \infty)$ mit

$$\epsilon(p) = \frac{\Phi^{-1}(1-\frac{\alpha}{2})\sqrt{p(1-p)}}{\sqrt{nm}} = \sqrt{\beta}\sqrt{p(1-p)}, \qquad \beta = \frac{[\Phi^{-1}(1-\frac{\alpha}{2})]^2}{nm},$$

approximativ gilt:

$$\mathbb{P}_p\left(\left|\frac{\sum_{l=1}^{n} X_l}{nm} - p\right| < \epsilon(p)\right) \geq 1 - \alpha$$

c) (3 Punkte) Zeigen Sie, dass $|T(X) - p| < \sqrt{p(1-p)}\sqrt{\beta} \Leftrightarrow p \in (p_-(X), p_+(X))$ für

$$p_\pm(X) = \frac{(2T(X) + \beta)}{2(1+\beta)} \pm \frac{\sqrt{\beta^2 + 4T(X)[1-T(X)]\beta}}{2(1+\beta)}$$

gilt und konstruieren Sie ein (approximatives) Konfidenzintervall zum Niveau α.

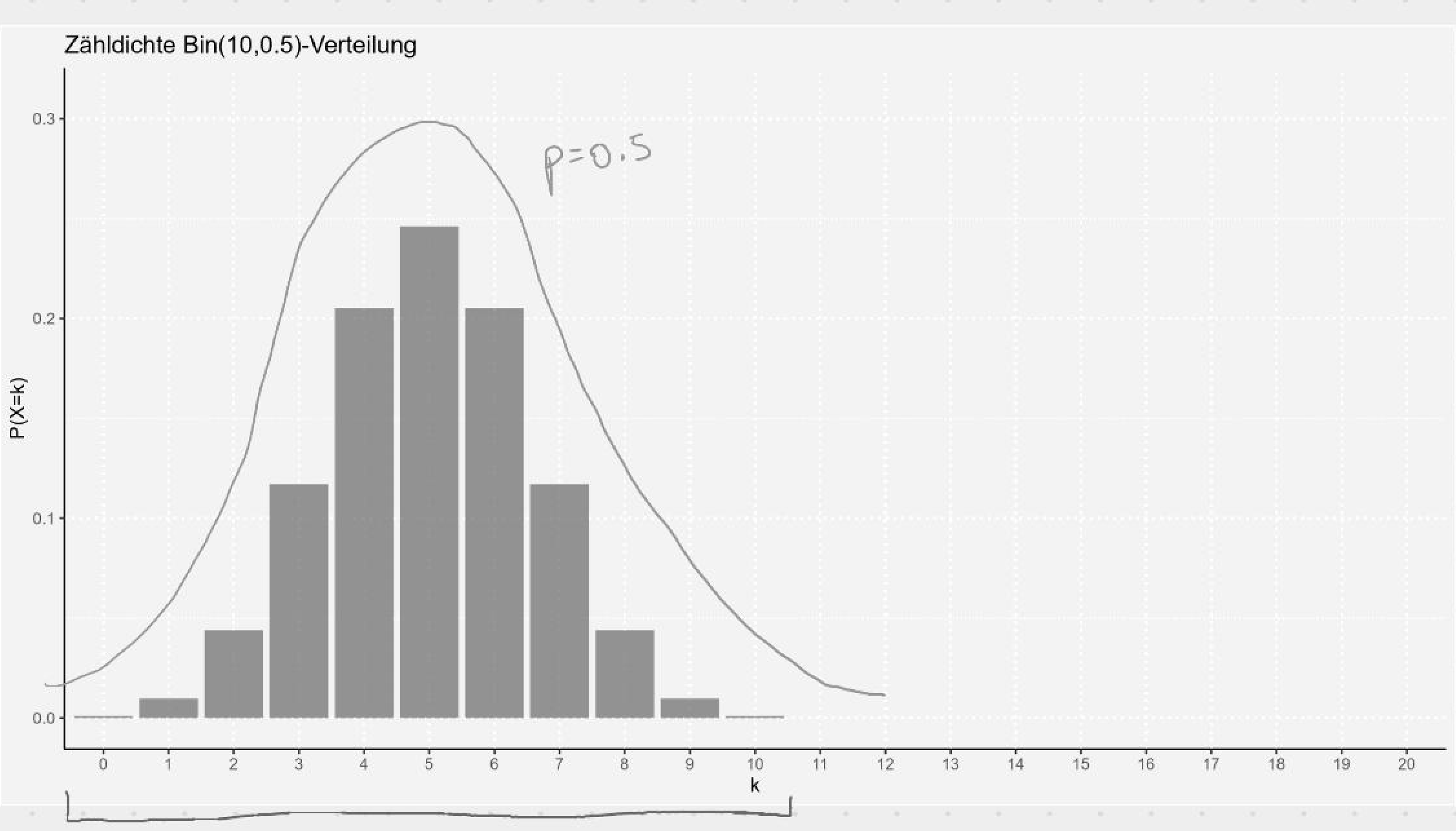
Heute wichtig: ML-Schätzer
(Rest siehe Lösungsskizze + Beispiele der Vorlesung, ggf. im Forum nachfragen)

A3 a)

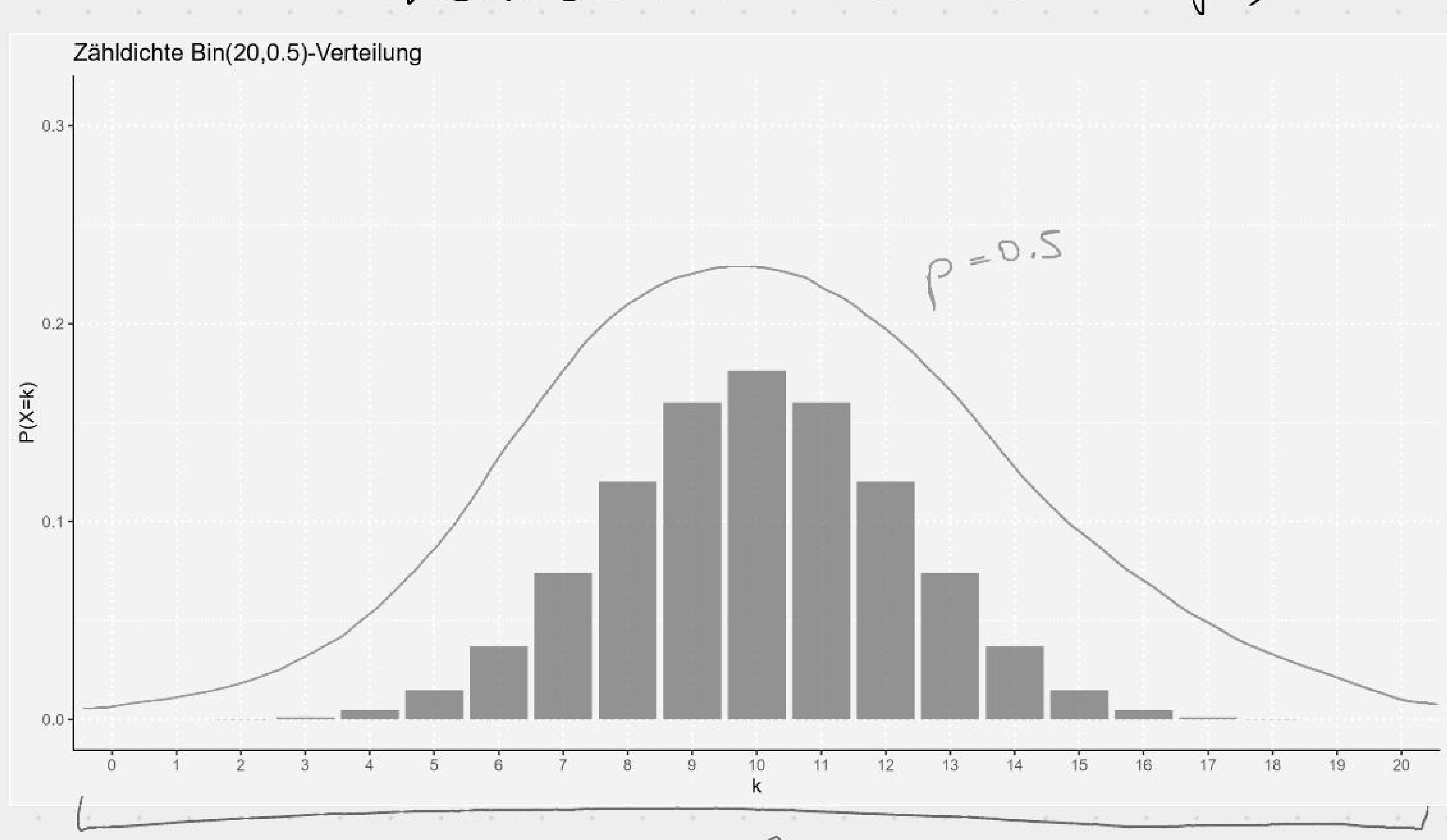
$X_i \sim Bin(m, p)$ $\quad$ m bekannt, p unbekannt

$$\mathbb{P}_p[X_i = k] = \binom{m}{k} p^k (1-p)^{m-k} \quad \text{für } k = 0, 1, \ldots, m$$

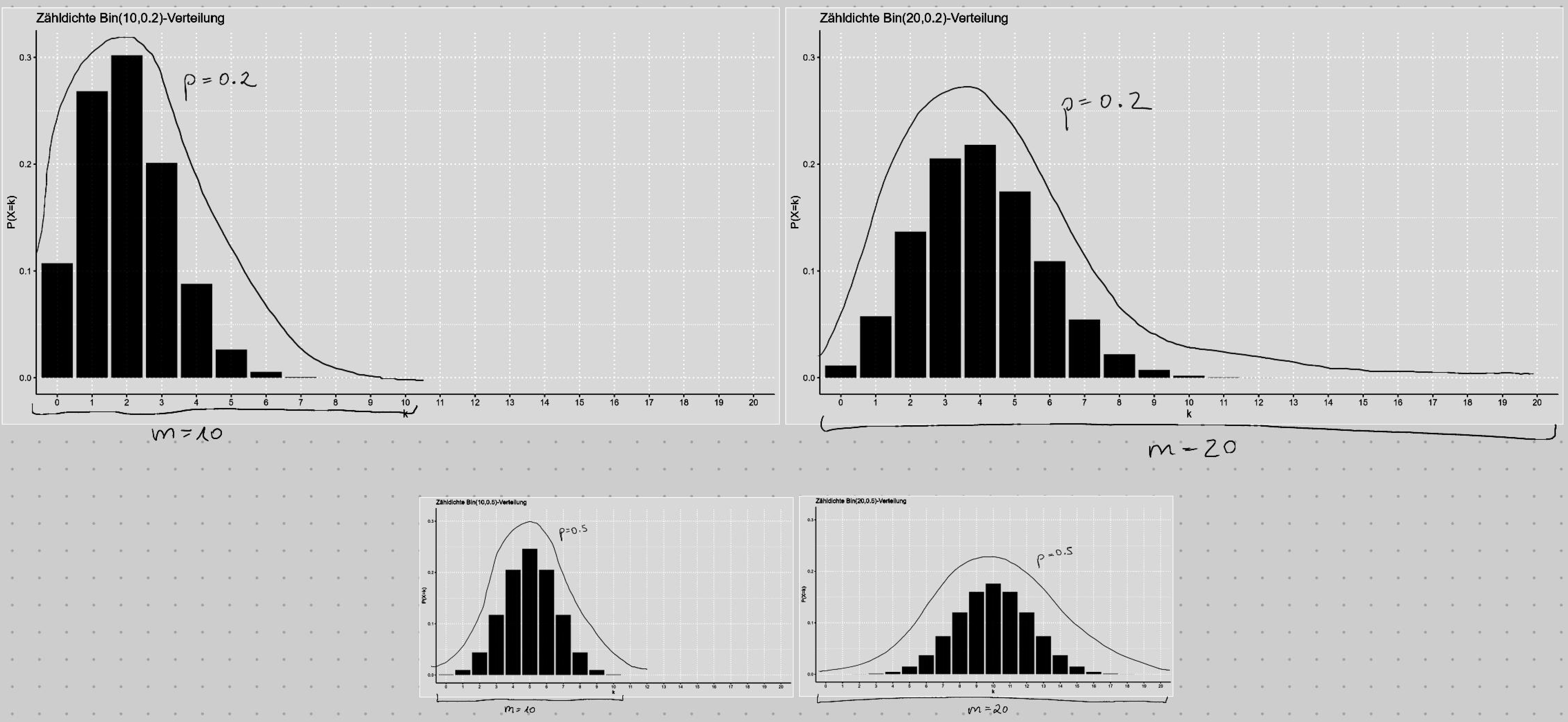
Wie beeinflussen m und p die Verteilung? m beeinflusst "Breite" der Werte, p beeinflusst "Tendenz" zu größeren/kleineren Werten (Schiefe)

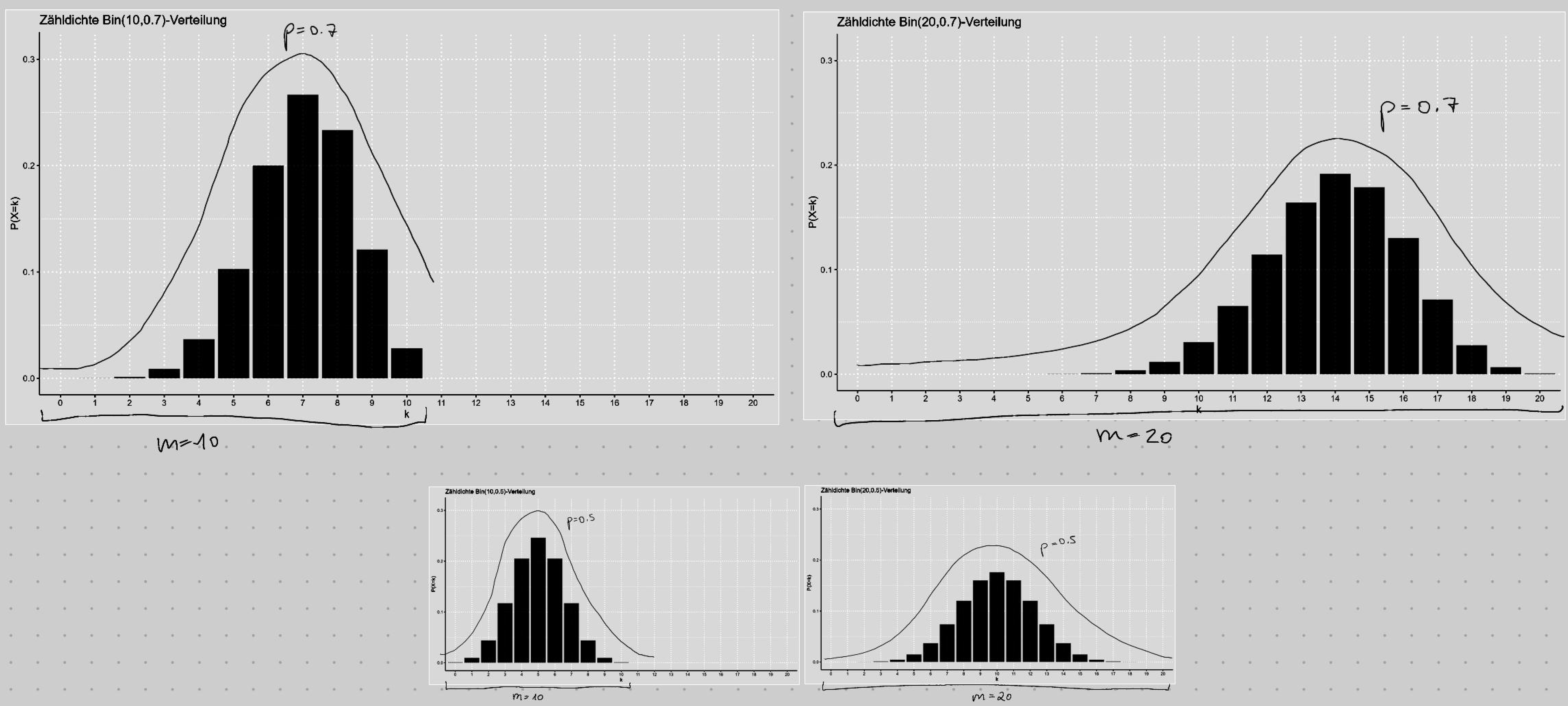


$m = 10$



$m = 20$





D.h. wir wollen in a) aus beobachteten Daten, z.B.

i	1	2	3	4	5	6	7	8	9	10
x_i	3	1	1	4	3	1	0	1	3	2

$m = 10$, $n = 10$

ein p schätzen (m ist bekannt). Hier: Wir vermuten, dass p klein ist.

Idee ML: Wenn die Daten bekannt sind, für welches p würden wir sie "am wahrscheinlichsten" beobachten?

[→ z.B. $p = 0.8$ ist hier (intuitiv) eher unwahrscheinlich, $p \approx 0$ aber auch!]

1. Wir betrachten die Likelihoodfunktion:

$$\varrho(x_1,\dots,x_n,p) = \prod_{i=1}^{n} \mathbb{P}[X_i = x_i] = \prod_{i=1}^{n} \left[\binom{m}{x_i} p^{x_i} (1-p)^{m-x_i}\right] \quad (*)$$

2. Der ML Schätzer T soll die Funktion maximieren, d.h.

$$\varrho(x_1, \dots, x_n, T(x_1 \dots x_n)) = \max_{p \in (0,1)} \varrho(x_1, \dots x_n, p) \quad \text{für alle } x_1, \dots x_n.$$

3. Die Funktion in (*) ist etwas unhandlich. Wir betrachten daher die log-Likelihood-Funktion (wie in Bsp 5.3.23):

$$\log(\varrho(x_1, \dots x_n, p)) = \sum_{i=1}^{n} \log\left[\binom{m}{x_i} p^{x_i} (1-p)^{m-x_i}\right] \quad \Big| \; \log(a \cdot b) = \log(a) + \log(b)$$

$$= \sum_{i=1}^{n} \log\left[\binom{m}{x_i}\right] + x_i \log[p] + (m - x_i) \log[(1-p)]$$

$$= \sum_{i=1}^{n} \log\left[\binom{m}{x_i}\right] + \left(\sum_{i=1}^{n} x_i\right) \cdot \log(p) + n \cdot m \log(1-p)$$

$$- \left(\sum_{i=1}^{n} x_i\right) \log(1-p) \quad \longrightarrow \max \text{ bzgl. } p!$$

4. Um die log-Likelihood zu maximieren, leiten wir nach p ab:

$$\frac{d}{dp} \log(\varrho(x_1, \dots x_n, p)) = \left(\sum_{i=1}^{n} x_i\right) \frac{1}{p} - n \cdot m \frac{1}{1-p} + \left(\sum_{i=1}^{n} x_i\right) \frac{1}{1-p} \stackrel{!}{=} 0$$

$$\Rightarrow \quad nm \cdot p = \left(\sum_{i=1}^{n} x_i\right) (1 - p + p)$$

$$\Rightarrow \quad p = \frac{1}{n \cdot m} \sum_{i=1}^{n} x_i .$$

(dies ist wirklich ein Maximum $\rightsquigarrow$ 2. Abl.)

5. D.h. $\quad T(x_1, \dots x_n) = \frac{1}{nm} \sum_{i=1}^{n} x_i \quad$ ist der ML-Schätzer!

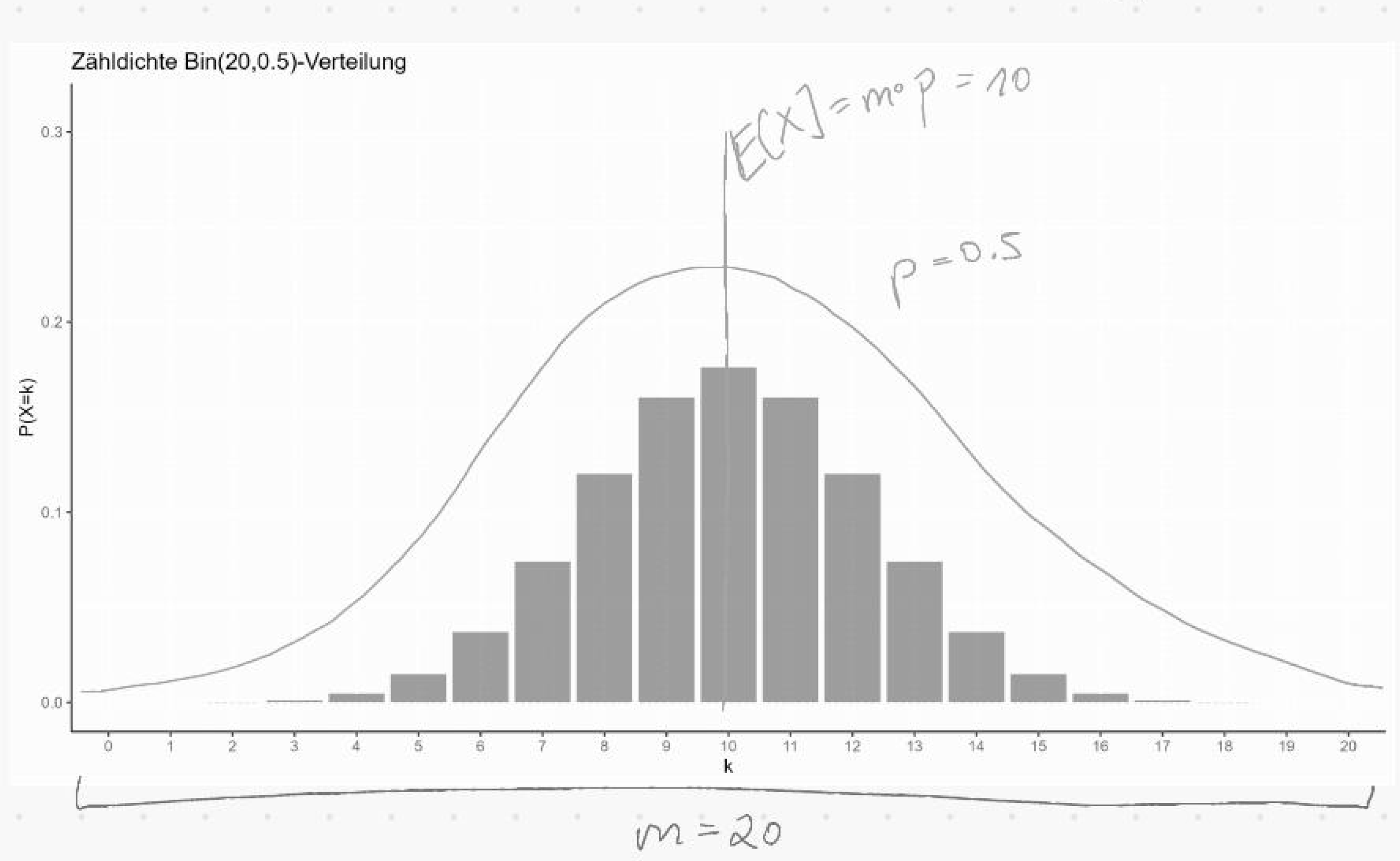
Im Beispiel:

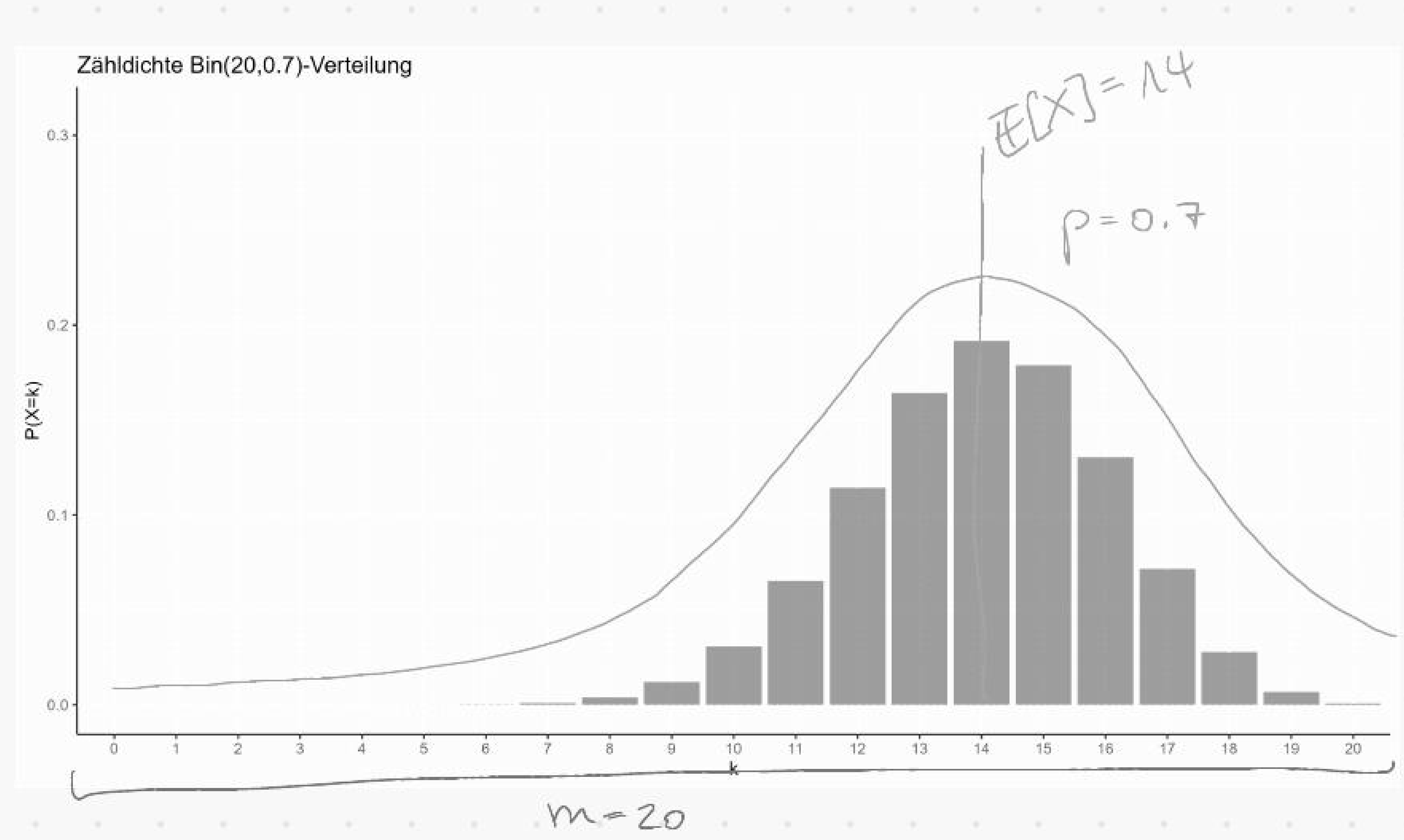
i	1	2	3	4	5	6	7	8	9	10
x_i	3	1	1	4	3	1	0	1	3	2

$$T(3,1,1,\dots) = \frac{1}{10 \cdot 10} (3 + 1 + 1 + \dots) = \frac{19}{100} = 0.19 = \hat{p}$$

(Daten wurden erzeugt mit $p = 0.2$).

Der ML-Schätzer ist intuitiv* auch sinnvoll hier: Es gilt $\mathbb{E}_p[X_i] = mp$.





und wir wissen ja, dass $\frac{1}{n} \sum x_i$ ein „guter" Schätzer für den EW ist (tatsächlich auch ein ML-Schätzer). Es ist also nicht überraschend, dass nun $\frac{1}{m} \frac{1}{n} \sum x_i \approx \frac{1}{m} \cdot m \cdot p$ ein guter Schätzer für p ist.

* Achtung, etwas unsaubere Argumentation

4. Generalised Linear Models & ML-Schätzer (10 Punkte). Wir betrachten folgende Situation. Wir wollen ein Modell erarbeiten, welches die Wahrscheinlichkeit vorhersagen kann, dass ein Gebäude bei einem Starkregenereignis stark beschädigt wird. Wir vermuten, dass diese Wahrscheinlichkeit vom Alter A und der Stärke des Unwetters S abhängt. Wir modellieren dies, indem wir annehmen, dass $Y_i \in \{0,1\}$ beschreibt, ob Gebäude i beschädigt wird (dabei steht 0 für 'kein Schaden' und 1 für 'Schaden'). Weiter nehmen wir an, dass $A_i \in \{0,1,2,3,4,5\}$ die Alterskategorie des Hauses und $S_i \in \{1,2,3,4,5\}$ die Stärke des jeweiligen Unwetters bezeichnet. Wir nehmen an, die Y_i sind unabhängig voneinander mit $Y_i \sim B(1,p_i)$, wobei

$$p_i = \frac{\mathrm{e}^{\beta_0+\beta_1 A_i+\beta_2 S_i}}{1+\mathrm{e}^{\beta_0+\beta_1 A_i+\beta_2 S_i}},$$

bzw. $\ln(p_i/(1-p_i)) = \beta_0+\beta_1 A_i+\beta_2 S_i$. Hier ist $\beta = (\beta_0,\beta_1,\beta_2) \in \mathbb{R}^3$ ein Vektor, der nicht von i abhängt. Wir nehmen an, wir haben n Datensätze bzw. Stichproben Y_i, A_i, S_i, $i = 1,\dots,n$, zur Verfügung.

a) (4 Punkte) Beschreiben Sie, wie die Idee der Maximum-Likelihood-Methode genutzt werden kann, um β zu bestimmen und zeigen Sie, dass

$$\max_{\beta\in\mathbb{R}^3}\Big(\sum_{i:y_i=1}(\beta_0+\beta_1 a_i+\beta_2 s_i) - \log(1+\mathrm{e}^{\beta_0+\beta_1 a_i+\beta_2 s_i}) - \sum_{i:y_i=0}\log(1+\mathrm{e}^{\beta_0+\beta_1 a_i+\beta_2 s_i})\Big)$$

ein sinnvoller Ansatz für das zugehörige Optimierungsproblem ist.

Wir nehmen nun an, wir haben folgende Daten beobachtet:

A_i	0	0	2	4	5
S_i	1	5	1	5	1
Y_i	0	1	1	1	0

b) (3 Punkte) Die Daten wurden mit den Werten $\beta_0 = -1$, $\beta_1 = 0.08$ und $\beta_2 = 0.022$ simuliert. Berechnen Sie die p_i für dieses Beispiel. Welche Interpretation haben die Parameter $\beta_0, \beta_1, \beta_2$? Was bedeuten positive/negative Werte?

c) (3 Punkte) Angenommen $\beta_0 = -1$ ist bekannt und β_1, β_2 sollen mit der Maximum-Likelihood-Methode geschätzt werden. Berechnen Sie die Werte für β_1 und β_2 (mit dem Computer) oder lesen Sie einen approximativen maximalen Wert in folgendem Bild ab, welches die Likelihood-Funktion farblich mit β_1 auf der x-Achse und β_2 auf der y-Achse zeigt.

A4) $\mathbb{P}[\text{Gebäude } i \text{ wird beschädigt}]$

$= \mathbb{P}[y_i = 1] = p_i(A_i, S_i)$

Alter A_i Stärke S_i

insbesondere

$$p_i = \frac{e^{\beta_0+\beta_1\cdot A_i+\beta_2 S_i}}{1+e^{\beta_0+\beta_1\cdot A_i+\beta_2 S_i}}$$

wobei $\beta_0, \beta_1, \beta_2$ unbekannte Parameter sind, die wir schätzen wollen.

Erinnerung: bei der linearen Regr. machen wir etwas Ähnliches "$Z_i = \beta_0 + \beta_1 W_i$" ⇝ kleinste Quadrate

Hier brauchen wir ein komplizierteres Verfahren, um die $\beta_0, \beta_1, \beta_2$ zu schätzen. Wir haben aber Datensätze mit y_i, A_i, S_i zur Verfügung!

Idee: Für welche $\beta_0, \beta_1, \beta_2$ sind die beobachteten Daten am wahrscheinlichsten? $\Rightarrow$ Max. Likelihood!

a) Hier Likelihood-Fkt. für Y, gegeben A & S. Der Parameterraum ist $\Theta = \mathbb{R}^3 = \{(\beta_0,\beta_1,\beta_2) : \beta_i \in \mathbb{R}\}$.

$$f: \underset{\text{Werte } A}{\{1,2,3,4,5\}} \times \underset{\text{Werte } S}{\{1,2,3\}} \times \underset{\text{Werte } y}{\{0,1\}} \times \underset{\text{Werte } \beta_0,\beta_1,\beta_2}{\mathbb{R}^3} \to [0,\infty)$$

$$(a,s,y,\beta) \mapsto f_\beta(a,s,y)$$

$$= \prod_{i=1}^{n} \mathbb{P}[y_i = y_i \mid A = a, S = s] = \prod_{i:\, y_i=1} p_i \cdot \prod_{j:\, y_j=0} (1-p_j)$$

wobei $p_i = \dfrac{e^{\beta_0+\beta_1 a_i+\beta_2 s_i}}{1+e^{\beta_0+\beta_1 a_i+\beta_2 s_i}}$. Insgesamt

$$f(a,s,y,\beta) = \left(\prod_{i:\, y_i=1} \frac{e^{\beta_0+\beta_1 a_i+\beta_2 s_i}}{1+e^{\beta_0+\beta_1 a_i+\beta_2 s_i}}\right)\left(\prod_{j:\, y_j=0} \frac{1}{1+e^{\beta_0+\beta_1 a_j+\beta_2 s_j}}\right).$$

Maximieren wir dies bzgl. β, finden wir die Parameter, die den Zusammenhang im Sinne der Max. Likelihood am besten beschreiben.
In der Aufgabenstellung steht die log-Likelihood-Funktion, die wir alternativ maximieren können.

b)

A_i	0	0	2	4	5
S_i	1	5	1	5	1
Y_i	0	1	1	1	0
p_i	0.2732888	0.2911098	0.3061884	0.3612368	0.3593929

Wir setzen $\beta_0 = -1$ $\beta_1 = 0.08$ $\beta_2 = 0.022$ ein

$$p_1 = \frac{e^{-1+0.08\cdot 0+0.022\cdot 1}}{1+e^{-1+0+0.022\cdot 1}} \approx 0.2733$$

Einfluss von $\beta_0, \beta_1, \beta_2$? Wir schreiben um:

$$p_i = \frac{e^{\cdots}}{1+e^{\cdots}} = \frac{1}{1+e^{-\beta_0-\beta_1 A_1-\beta_2 A_2}} \quad \text{Schadenw'keit}$$

$\rightsquigarrow \frac{1}{1+e^{-\beta_0}}$ wachsend in β_0 & gleich $1/2$ für $\beta_0 = 0$ ($< 1/2$ für $\beta_0 < 0$)

gibt "Grundw'keit" an, dass ein Schaden passiert (vor anderen Effekten

$\rightsquigarrow$ Wir wissen, dass $A_i > 0$ & dass die Werte größer werden, wenn das Haus älter ist. Wir vermuten, dass dann auch die Schadenw'keit steigt, daher erwarten wir $\beta_1 > 0$ ($\beta_1 < 0$ würde den umgekehrten Zusammenhang modellieren).

$\rightsquigarrow$ ähnlich für β_2.

Achtung: Wenn wir die Tabelle ansehen, ist das größte p_i in der letzten Spalte. Allerdings gilt hier $y = 0$ (kein Schaden).
Man kann sich überlegen, dass die ML-Methode (ohne die β_i zu kennen) dadurch gestört werden kann – besonders bei so einer kleinen Stichprobe.

c) Tatsächlich bekommen wir für die Daten (mit dem PC)

$$\beta_1 = -0.23 \quad \text{und} \quad \beta_2 = 1.1$$

heraus. Wegen des negativen Zsh. mit dem Alter (ältere Häuser sind in diesem Modell weniger anfällig) sind diese Werte unplausibel!
$\to$ weitere Beispiele in Lösungsskizze.

Tipps:

1. Grundbegriffe (10 Punkte). In einer Sendung von 10 Geräten befindet sich eine unbekannte Anzahl fehlerhafter Geräte, wobei der Fehler jeweils nur durch eine sehr kostspielige Qualitätskontrolle festgestellt werden kann. Ein Abnehmer, der an einer völlig einwandfreien Lieferung interessiert ist, führt folgende Eingangskontrolle durch: Er prüft 5 Geräte. Sind diese alle einwandfrei, so nimmt er die Sendung an, sonst lässt er sie zurückgehen.

a) (3 Punkte) Beschreiben Sie das Vorgehen testtheoretisch.

b) (2 Punkte) Was sind Fehler erster und zweiter Art in diesem Modell?

c) (3 Punkte) Ermitteln Sie das effektive Niveau des Testverfahrens.

d) (2 Punkte) Wieviele Geräte müssen überprüft werden, wenn die Wahrscheinlichkeit für eine irrtümliche Annahme der Sendung kleiner gleich 0.1 sein soll?

Tipps:

1. Für a): Formulieren Sie die Hypothese, den kritischen Bereich und berechnen sie die W'keit, dass die Testgröße X im kritischen Bereich liegt.
2. Die Aufgabe hat mit Kombinatorik zu tun. Angenommen wir betrachten eine Urne mit ϑ roten und $s = N - \vartheta$ schwarzen Kugeln, aus welcher n-mal ohne Zurücklegen gezogen wird. Dann besitzt die Anzahl X der gezogenen roten Kugeln eine so genannte *hypergeometrische* Verteilung:

$$\mathbb{P}_\vartheta[X = k] = \frac{\binom{\vartheta}{k}\binom{N-\vartheta}{n-k}}{\binom{N}{n}} \mathbb{1}_{\{k=0,1,\ldots,\vartheta\wedge n\}}$$

3. Zu d): Wir suchen nach n, sodass $\mathbb{P}_\vartheta[\phi(X) = 0] \leq 0.1$ für alle $\vartheta \in \Theta_1$ gilt. Man darf ohne Beweis verwenden, dass $\frac{\vartheta\binom{N-\vartheta}{n}}{\binom{N}{n}}$ fallend in ϑ ist.

2. Verspätete Züge (10 Punkte). Wir betrachten ein Beförderungsunternehmen. Für ein Machine-Learning-Modell, welches voraussagen soll, ob ein Zug mehr als 16 Minuten Verspätung hat, wollen wir herausfinden ob der Wochentag ein relevantes Merkmal ist. Wir haben die folgenden Daten beobachtet:

	Mo	Di	Mi	Do	Fr	Sa	So
Anzahl verspätete Züge	31	31	14	29	40	16	14
Anzahl pünktl. Züge	119	69	86	91	110	84	66

Testen Sie mit einem χ^2-Test zum Niveau $\alpha = 0.05$, ob der Wochentag in das Modell miteinbezogen werden sollte.

Tipps:

1. Eine Tabelle für die absoluten und relativen Häufigkeiten machen. Die h^A, h^B und L^A, L^B ergeben sich dann leicht als Summen der Zeilen bzw. Spalten.
2. Die Musterlösung ist so lang, weil noch einmal detailliert die Notation der Vorlesung betrachtet wird. Wenn dies auch machen möchte, ist es eine gute Übung. Man kann die Lösung aber auch knapper aufschreiben.

3. Bester Test (10 Punkte). Für eine Stichprobe vom Umfang n wird angenommen, dass die Daten normalverteilt sind mit Mittelwert θ und Varianz $\sigma^2 = 1.5$.

a) (4 Punkte) Konstruieren Sie einen optimalen Test zum Niveau $\alpha = 0.05$ für die Hypothese $\theta_0 = 25$ gegen die Alternative $\theta_1 = 25.5$.

b) (4 Punkte) Berechnen Sie Wahrscheinlichkeit für einen Fehler 2. Art.

c) (2 Punkte) Wie groß muss n gewählt werden, damit der Fehler 2. Art kleiner als 0.05 wird?

Tipps:

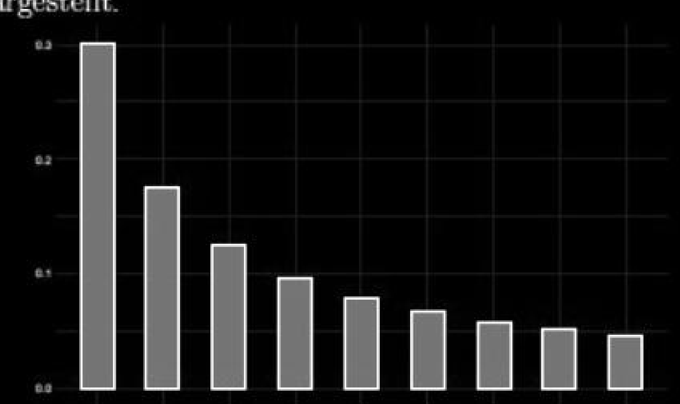
1. „bester Test“ und „optimal“ sind Stichworte, nach denen wir in den Vorlesungsnotizen suchen sollten (wir müssen nicht selber optimieren – es gibt ein Resultat dazu).
2. Man braucht in manchen Teilaufgaben die Normalverteilungstabelle in Anhang D der Vorlesung. Dabei kann es hilfreich sein, dass wenn $X_k \sim \mathcal{N}(\mu, \sigma^2)$, dann $\frac{n^{-1}\sum_{k=1}^n X_k - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ (Warum?).
3. Für a): bei Bedarf kann man sich an folgenden Zwischenschritten orientieren:
   - Likelihood-Funktion und $\rho_0(x) = \rho(\theta_0, x)$, $\rho_1(x) = \rho(\theta_1, x)$ berechnen
   - $R(x_1, \ldots, x_n)$ nach Def. 6.2.10 aufschreiben und vereinfachen. Herauskommen könnte
   $$R(x_1, \ldots, x_n) = \begin{cases} \exp\left(-\frac{101}{12}n + \frac{1}{3}\sum_{k=1}^n x_k\right), & \rho_0 > 0\,, \\ \infty\,, & \text{sonst}\,. \end{cases}$$
   - Man sucht c, sodass $G_\phi(\theta_0) = \mathbb{P}_{\theta_0}[R(X) > c] = \alpha$. Herauskommen könnte
   $$c = \exp\Big(\frac{1.65\sqrt{n}}{\sqrt{6}} - \frac{n}{12}\Big).$$
4. Für b): Wir müssen $\mathbb{P}_{\theta_1}[R(X) < c]$ betrachten (Warum?). Herauskommen soll $\Phi\left(1.65 - \frac{\sqrt{n}}{\sqrt{6}}\right)$.
5. Für c): Zwischenergebnis aus b) verwenden. Man beachte $1 - \Phi(-x) = \Phi(x)$, wenn man die Tabelle ansieht.

4. Benford's Law (10 Punkte). 1881 beobachtete der amerikanische Mathematiker Simon Newcomb, dass die vorderen Seiten seiner Logarithmentafeln deutlich stärker abgegriffen waren, als die hinteren. Dies bedeutete, dass also häufiger Zahlen mit führender Ziffer 1 nachgeschlagen wurden, als mit höheren führenden Ziffern. Diese Beobachtung führte zur Entdeckung von Benford's Law, welches (intuitiv) besagt, dass in vielen realen Datensätzen (zum Beispiel Populationszahlen, Sterberaten, Sportstatistiken) anteilig etwa 30% der Zahlen mit einer 1 beginnen, 17.5% mit einer 2 und diese Anteile weiter fallen bis etwa 5%, was der Anteil der mit einer 9 beginnenden Zahlen ist. Die zugehörige Verteilung kann wie folgt beschrieben werden. Eine positive Zufallsvariable X heißt *Benfordsch*, falls

$$\mathbb{P}[X \in E_i] = \log_{10}\Big(1 + \frac{1}{i}\Big), \quad \text{wobei } E_i = \{x > 0 : \text{führende Ziffer von } x \text{ ist } i\}, \qquad i = 1, \ldots, 9\,.$$

Approximationen für die jeweiligen Wahrscheinlichkeiten sind in der folgenden Tabelle aufgelistet und im Balkendiagramm dargestellt.

1	0.3010
2	0.1761
3	0.1249
4	0.0969
5	0.0792
6	0.0669
7	0.0580
8	0.0512
9	0.0458



In vielen Szenarien passt diese Verteilung erstaunlich gut zu „echten“ Daten. Darum wird diese Verteilung auch verwendet, um gefälschte bzw. manipulierte Datensätze zu erkennen, von denen man normalerweise erwarten würde, dass sie Benford's Law erfüllen (zum Beispiel bei Wahlergebnissen und Steuerdaten, aber auch bei Bilddateien).

erste Ziffer	Beobachtungen Deutschland	Beobachtungen China
1	47	149
2	24	84
3	20	79
4	15	42
5	10	30
6	11	46
7	6	25
8	12	23
9	5	21

Wir betrachten nun das Handelsvolumen von Aktien, die für die jeweils gelisteten Unternehmen an den betrachteten Märkten (Deutschland und China) am 5.12.23 bzw. 6.12.23 gehandelt wurden. Wir nutzen die Auswertung der Website `http://shiny.calpoly.sh/BenfordData/`, um in der Tabelle oben direkt die Häufigkeiten der beobachteten ersten Ziffern der Volumina in der jeweiligen Geldwährung (EUR, CNY) aufzulisten.

a) (4 Punkte) Nutzen Sie einen geeigneten statistischen Test zum Niveau $\alpha = 0.1$, um zu überprüfen, ob die an der deutschen Börse gehandelten Volumina Benford's law genügen.

b) (4 Punkte) Nutzen Sie einen geeigneten statistischen Test zum Niveau $\alpha = 0.1$, um zu überprüfen, ob die an der chinesischen Börse gehandelten Volumina Benford's law genügen.

c) (2 Punkte) Als 'p-Wert' bezeichnen wir die Wahrscheinlichkeit, dass die Teststatistik (unter der Nullhypothese) einen mindestens so extremen Wert ergibt wie in den betrachteten Daten. Bestimmen Sie (approximativ) den p-Wert für b) mithilfe der folgenden Tabelle und beschreiben Sie, was dieser hier aussagt.

$\chi^2_{n,\tilde\alpha}$	$\tilde\alpha = 0.904$	0.903	0.902	0.901	0.9	0.1
$n = 1$	2.7707695	2.7541889	2.7377935	2.7215796	2.7055435	0.0157908
2	4.6868142	4.6660886	4.6455756	4.6252709	4.6051702	0.2107210
3	6.3445226	6.3208948	6.2974996	6.2743324	6.2513886	0.5843744
4	7.8819369	7.8559436	7.8301997	7.8047003	7.7794403	1.0636232
5	9.3469601	9.3189184	9.2911408	9.2636220	9.2363569	1.6103080
6	10.7624970	10.7326224	10.7030252	10.6736997	10.6446407	2.2041307
7	12.1415171	12.1099685	12.0787093	12.0477339	12.0170366	2.8331069
8	13.4921822	13.4590829	13.4262844	13.3937808	13.3615661	3.4895391
9	14.8200146	14.7854640	14.7512250	14.7172911	14.6836566	4.1681590
10	16.1289529	16.0930335	16.0574357	16.0221530	15.9871792	4.8651821

Tipps:

1. Die Situation zuerst wieder in die Notation der Vorlesung 'übersetzen'. Dafür kann man sich an Abschnitt 6.2.5 orientieren.
2. In der Musterlösung wurde für a) und b) die Tabelle aus Anhang D der Vorlesungsnotizen verwendet. Man könnte aber auch die genaueren Werte aus der Tabelle in c) nutzen. Eigentlich wird die Tabelle aber nur für den p-Wert benötigt.