Theorie Übung 7

Stichprobenverteilung:

- Wir kennen die Populationsparameter (meist μ, σ^2) nicht mehr und müssen sie schätzen

$\Longrightarrow$ aus Gründen der Erwartungstreue und statistischer Effizienz verwenden wir den folgenden Schätzer:

$$\bar{X} = \left(\frac{X_1 + \ldots + X_N}{N} \right)$$

, wobei $\{X_j\}_{j=1}^{N}$ unabhängige Züge aus einer beliebigen Populationsverteilung mit $\mathbb{E}(X) = \mu_X$ und $\mathrm{Var}(X) = \sigma_X^2$ sind. Wir haben dann:

$$\mathbb{E}(\bar{X}) = \mu_X \quad \text{(erwartungstreu)}$$

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{N} \quad \text{("effizient")}$$

Zentraler Grenzwertsatz ($N \geq 50$)

Seien $X_1, X_2, \dots, X_N$ unabhängige, identisch verteilte Zufallsvariablen mit $\mathbb{E}(X_k) = \mu$ und $Var(X_k) = \sigma^2 \in (0, \infty)$.
Sei $\bar{X} = \frac{X_1 + \dots + X_N}{N} = N^{-1} \sum_{j=1}^{N} X_j$. Dann gilt:

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{N}} \xrightarrow[N \to \infty]{d} \mathcal{N}(0,1) \quad \left(\sim \mathcal{N}(0,1) \right)$$

Oder nicht standardisiert:

$$\bar{X} \sim \mathcal{N}\left(\mu_X, \frac{\sigma^2}{N}\right)$$

Somit können wir Inferenz betreiben, da wir wissen wie $\bar{X}$ verteilt ist.

↳ Binomialverteilung ist nur eine „Anwendung" davon, das werden wir in Aufgabe 2.) noch genauer anschauen.

Statistik
Übungsaufgaben 7

Stichprobenverteilung - Zentraler Grenzwertsatz - Unverzerrtheit eines Schätzers

B.4.12.3 A particular experiment generates a random variable X that has only two outcomes: $X = 1$ (success) with probability $p = 0.6$ and $X = 0$ (failure) with probability $(1 - p) = 0.4$. Consider a random sample consisting of $N = 3$ independent replications of this experiment. Find the exact sampling distribution of the sample mean.

1. Beschreiben Sie den Zentralen Grenzwertsatz (ZGS) mit seinen Annahmen und Implikationen. Finden Sie ein Beispiel in dem eine oder mehrere dieser Annahmen nicht erfüllt sind.

B.4.12.17 The population mean for a random variable X is $\mu = 40$. The population variance is $\sigma^2 = 81$. For a (large) random sample of size N drawn from this population, find the following:

a) The expected value and the variance of the sample mean $\bar{X}$ when $N = 36$.

b) The probability that $P(\bar{X} \geqslant 41)$ in the above case.

c) The probability $P(38.5 \leqslant \bar{X} \leqslant 40.5)$ when $N = 64$.

2. Sei X eine Zufallsvariable welche einer Binomial Verteilung folgt mit $N = 100$ und $p = 0.2$. Berechne folgende Wahrscheinlichkeiten mittels der Normal-Approximation mit Stetigkeitskorrektur.

a) $P(X \leqslant 25)$

b) $P(X > 30)$

c) $P(15 < X < 22)$

3. Gegeben sei eine Zufalls-Stichprobe, $X_1, \ldots, X_N$, aus einer Verteilung mit Erwartungswert μ und Varianz σ^2. Welche der folgenden Schätzer für μ sind erwartungstreu

a) $\hat{\mu}_1 = X_1$

b) $\hat{\mu}_2 = \frac{X_1 + X_N}{2}$

c) $\hat{\mu}_3 = \bar{X} + 1/N$

d) Wie beurteilen Sie die Varianz der 3 Schätzer im Vergleich zu der Varianz von $\bar{X}$?

4. Sei $X_1, \ldots, X_N$ eine Zufalls-Stichprobe von einer Verteilung mit Erwartungswert μ und Varianz σ^2.

a) Zeigen Sie, dass

$$E\left[\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2\right] = \frac{N-1}{N}\sigma^2 \ .$$

Hinweis: starten Sie mit

$$E\left[\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2\right] = E\left[\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu + \mu - \bar{X})^2\right] \ .$$

b) Zeigen Sie, dass daher die Stichprobenvarianz

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

ein unverzerrter Schätzer für σ^2 ist.

B.4.12.3 A particular experiment generates a random variable X that has only two outcomes: $X = 1$ (success) with probability $p = 0.6$ and $X = 0$ (failure) with probability $(1-p) = 0.4$. Consider a random sample consisting of $N = 3$ independent replications of this experiment. Find the exact sampling distribution of the sample mean.

$X_j \sim \text{Bernoulli}(p)$, $j = 1,2,3$

$X = \sum_{j=1}^{3} X_j \sim \text{Binomial}(N,p)$, $N=3$

- $\bar{X} = (\overset{0/1}{X_1} + \overset{0/1}{X_2} + \overset{0/1}{X_3})/3$, wobei $|\{X_1 + X_2 + X_3\}| = 2^3$

$\mathbb{P}(\bar{X} = \bar{x})$, $\bar{x} \in \{0, 1/3, 2/3, 1\}$

- $\{0,0,0\}$ mit $\mathbb{P}(\bar{X} = \underbrace{\bar{x}}_{=0}) = \underbrace{\binom{3}{0}}_{=1} \cdot \underbrace{0.6^0}_{=1} \cdot 0.4^3$ (kein Erfolg)

- $\left.\begin{array}{l}\{0,0,1\}\\ \{1,0,0\}\\ \{0,1,0\}\end{array}\right\}$ ein Erfolg mit $\mathbb{P}(\bar{X} = \bar{x} = 1/3) = \underbrace{\binom{3}{1}}_{=3} \cdot 0.6 \cdot 0.4^2$

- $\left.\begin{array}{l}\{1,1,0\}\\ \{1,0,1\}\\ \{0,1,1\}\end{array}\right\}$ zwei Erfolge mit $\mathbb{P}(\bar{X} = \bar{x} = 2/3) = \underbrace{\binom{3}{2}}_{=3} \cdot 0.6^2 \cdot 0.4$

- $\{1,1,1\}$ } drei Erfolge mit $P(\bar{X}=\bar{x}=1) = \underbrace{\binom{3}{3}}_{=1} \cdot 0.6^3$

$\bar{X}=\bar{x}$	0	1/3	2/3	1
$P(\bar{X}=x)$	0.064	0.288	0.43	0.216

1. Beschreiben Sie den Zentralen Grenzwertsatz (ZGS) mit seinen Annahmen und Implikationen. Finden Sie ein Beispiel in dem eine oder mehrere dieser Annahmen nicht erfüllt sind.

Seien $X_1, X_2, \ldots, X_n$ unabhängige, identisch verteilte Zufallsvariablen mit $\mathbb{E}(X_k) = \mu$ und $Var(X_k) = \sigma^2 \in (0, \infty)$.

Sei $\bar{X} = \frac{X_1 + \ldots + X_n}{N} = N^{-1} \sum_{j=1}^{N} X_j$ Dann gilt:

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{N}} \xrightarrow[N \to \infty]{d} \mathcal{N}(0,1)$$

· Gegenbeispiel : $X_1, \ldots, X_6$ und

$$X_2 = -X_1$$

$$X_4 = -X_3$$

$$X_6 = -X_5$$

$$\Rightarrow \bar{X} = \frac{X_1 + \ldots + X_6}{6} = \frac{X_1 + (-X_1) + X_3 + (-X_3) + X_5 + (-X_5)}{6} = 0$$

($\bar{X}$ ist konstant und nicht $\bar{X}$ in $\mathcal{N}(\mu_X, \sigma_X^2 / N)$)

B.4.12.17 The population mean for a random variable X is $\mu = 40$. The population variance is $\sigma^2 = 81$. For a (large) random sample of size N drawn from this population, find the following:

a) The expected value and the variance of the sample mean $\bar{X}$ when $N = 36$.

b) The probability that $P(\bar{X} \geqslant 41)$ in the above case.

c) The probability $P(38.5 \leqslant \bar{X} \leqslant 40.5)$ when $N = 64$.

Theorie: $\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{N}\right)$ und somit

$$\bar{Z} = \frac{\bar{X} - \mu_X}{\frac{\sigma_X}{\sqrt{N}}} \sim N(0,1)$$

(a) $$\mathbb{E}(\bar{X}) = \mathbb{E}\left(N^{-1} \sum_{i=1}^{N} X_i\right) = N^{-1}\left(\mathbb{E}(X_1) + \dots + \mathbb{E}(X_N)\right)$$

$$= N^{-1} \underbrace{(\mu + \dots + \mu)}_{N \cdot \mu} = \mu = 40$$

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\Big(\underbrace{N^{-1}}_{:=a} \underbrace{\sum_{i=1}^{N} X_i}_{:=Y}\Big) = N^{-2} \sum_{i=1}^{N} \underbrace{\mathrm{Var}(X_i)}_{} = N^{-2} \underbrace{\sum_{i=1}^{N} \sigma^2}_{N\sigma^2}$$

$(\mathrm{Var}(aY) = a^2 \mathrm{Var}(Y))$

$$= N^{-2} N \sigma^2 = \frac{\sigma^2}{N} = \frac{81}{36} = 2.25$$

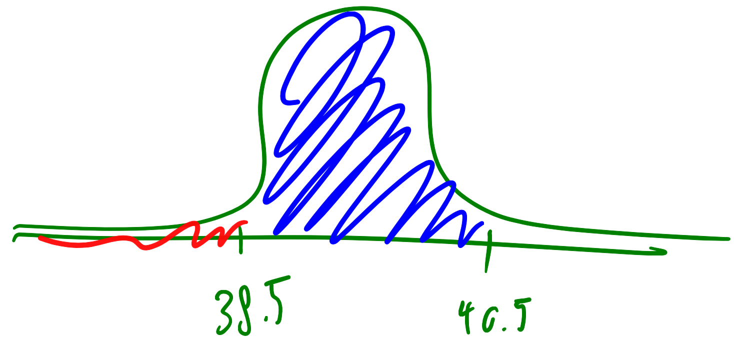
(b) $$P(\bar{X} \geqslant 41) = P\Big(\underbrace{\frac{\bar{X} - \mu_X}{6/\sqrt{N}}}_{\bar{Z} \sim N(0,1)} \geqslant \frac{40 - 41}{\sqrt{\frac{81 = 6^2}{36}}}\Big) = P(\bar{Z} \geqslant 0.6) =$$

$$1 - P(\bar{Z} \leq 0.6)$$

$$= 0.25$$

$\bar{X} \sim N\left(\mu_x, \frac{\sigma_x^2}{N}\right) \quad Z = \frac{\bar{X} - \mu_x}{\sigma/\sqrt{N}} \sim N(0,1)$

(c) $P(38.5 \le \bar{X} \le 40.5)$



$= P(\bar{X} \le 40.5) - P(\bar{X} \le 38.5)$

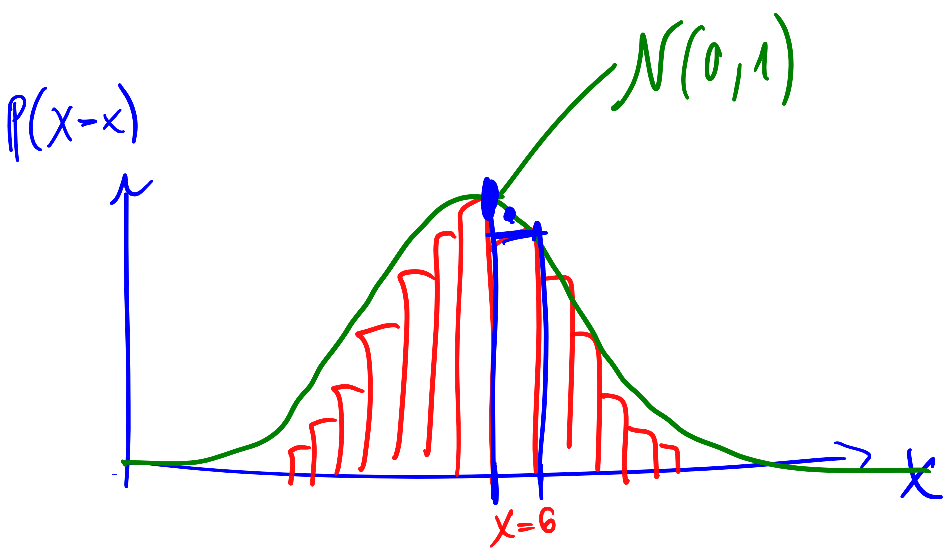
$P(\bar{X} \ge 38.5) = 1 - P(X \le 38.5)$

$= P\left(Z \le \frac{40.5 - 40}{\sqrt{\frac{81}{64 = N}}}\right) - P\left(Z \le \frac{38.5 - 40}{\sqrt{\frac{81}{64}}}\right)$

$= P(Z \le 0.4) - P(Z \le -1.3)$

$= P(Z \le 0.4) - (1 - P(Z \le 1.3))$

$= 0.65 - (1 - 0.9) = 0.55$

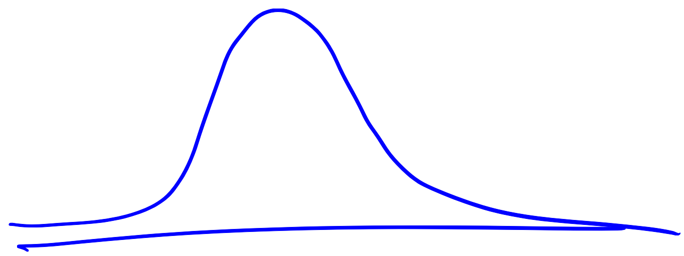


2. Sei X eine Zufallsvariable welche einer Binomial Verteilung folgt mit $N = 100$ und $p = 0.2$. Berechne folgende Wahrscheinlichkeiten mittels der Normal-Approximation mit Stetigkeitskorrektur.

a) $P(X \leqslant 25)$

b) $P(X > 30)$

c) $P(15 < X < 22)$



Theorie: $X_i \sim \text{Bernoulli}(p)$, sodass $\mathbb{E}(X_i) = p$ und $\text{Var}(X_i) = p(1-p)$

Wir haben $\hat{p} = \dfrac{(X_1 + \dots + X_N)}{N}$ und $N\hat{p} = X_1 + \dots + X_n$.

$E(N\hat{p})$ und $\text{Var}(N\hat{p})$ sind somit

$$\mathbb{E}(N\hat{p}) = \mathbb{E}(X_1 + \dots + X_n) = \underbrace{\underbrace{\mathbb{E}(X_1)}_{=p} + \dots + E(X_n)}_{N \text{ mal}} = Np$$

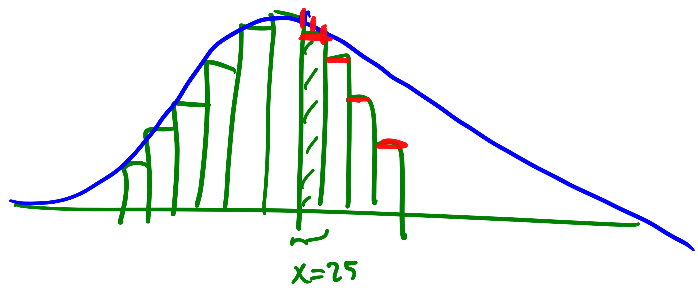
$$\begin{aligned} \text{Var}(N\hat{p}) &= \text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) \\ &= p(1-p) + \dots + p(1-p) \\ &= Np(1-p) \end{aligned}$$

$\Rightarrow$ da $N\hat{p}$ ein Mittelwert ist und auch die anderen Annahmen des ZGS erfüllt sind, haben wir

$$N\hat{p} \sim \mathcal{N}(Np, Np(1-p))$$

Github: enussl

(a) $P(X \le 25) = P(X < 26)$



↳ Stetigkeitskorrektur und rechnen

$$P(X \le 25.5) = P\left(Z \le \frac{25.5 - \overbrace{Np}^{\mu_X}}{\sqrt{\underbrace{Np(1-p)}_{\sigma_X^2}}}\right) = P(Z \le 1.375)$$

Tabelle
$= 0.91$

$X \in \mathbb{N}$

(b) $P(X > 30) = P(X \geq 30)$

↳ Stetigkeitskorrektur gibt

$$P(X \geq 30.5) = 1 - P(X \le 30.5)$$

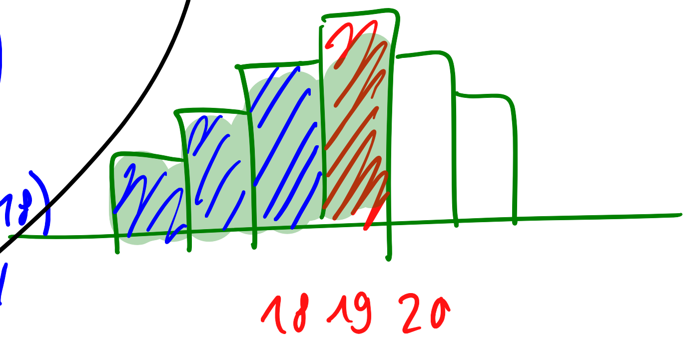
$$= 1 - P\left(Z \le \frac{30.5 - 20}{\sqrt{100 \cdot 0.2(1-0.2)}}\right) = 1 - P(Z \le 2.625)$$

$$= 0.00435$$

(c) $\mathbb{P}(15 < X < 22) = \underbrace{\mathbb{P}(X < 22)}_{\mathbb{P}(X \le 21)} - \underbrace{\mathbb{P}(X \le 15)}_{\mathbb{P}(X < 16)}$

$\mathbb{P}(18 < X < 20)$

$= \mathbb{P}(X < 20) - \mathbb{P}(X \le 18)$



Stetigkeitskorrektur:

$$\mathbb{P}(X \le 21.5) - \mathbb{P}(X \le 15.5)$$

$$= \mathbb{P}(Z \le 0.375) - \mathbb{P}(Z \le -1.125)$$

$$= \mathbb{P}(Z \le 0.375) - \left(1 - \mathbb{P}(Z \le 1.125)\right)$$

$$\overset{\text{Tabelle}}{=} 0.64 - (1 - 0.86) \approx 0.51$$

3. Gegeben sei eine Zufalls-Stichprobe, $X_1, \ldots, X_N$, aus einer Verteilung mit Erwartungswert μ und Varianz σ^2. Welche der folgenden Schätzer für μ sind erwartungstreu

a) $\hat{\mu}_1 = X_1$

b) $\hat{\mu}_2 = \frac{X_1 + X_N}{2}$

c) $\hat{\mu}_3 = \bar{X} + 1/N$

d) Wie beurteilen Sie die Varianz der 3 Schätzer im Vergleich zu der Varianz von $\bar{X}$?

Theorie:

- Erwartungstreu : $\mathbb{E}(\hat{\theta}) = \theta$

4. Sei $X_1, \ldots, X_N$ eine Zufalls-Stichprobe von einer Verteilung mit Erwartungswert μ und Varianz σ^2.

a) Zeigen Sie, dass

$$E\left[\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2\right] = \frac{N-1}{N}\sigma^2 \ .$$

Hinweis: starten Sie mit

$$E\left[\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2\right] = E\left[\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu + \mu - \bar{X})^2\right] \ .$$

b) Zeigen Sie, dass daher die Stichprobenvarianz

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

ein unverzerrter Schätzer für σ^2 ist.