# Stochastik 1 Hausaufgaben Blatt 13

Jun Wei Tan*

*Julius-Maximilians-Universität Würzburg*

(Dated: January 29, 2025)

**Problem 1.** In einer Meinungsumfrage soll die Zustimmung oder Ablehnung eines generellen Tempolimits in der Bevölkerung geschätzt werden. Dazu werden $n$ zufällig ausgewählte Personen befragt. Dabei wird $S_n/n$, die relative Anzahl der Befürworter unter den befragten Personen, als Schätzung für die Zustimmungsrate $p$ verwendet.

(a) Begründen Sie, weshalb $S_n$ näherungsweise als binomialverteilt, $S_n \sim \mathrm{Bin}(n,p)$, angenommen werden kann.

(b) Verwenden Sie den Satz von de Moivre-Laplace, um folgende Approximation herzuleiten:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \approx 2\left(1 - \Phi\left(\frac{\sqrt{n}\varepsilon}{\sqrt{p(1-p)}}\right)\right),$$

wobei $\Phi$ die Verteilungsfunktion der Standard-Normalverteilung bezeichnet.

(c) Wie viele Personen sollte ein Meinungsforschungsinstitut befragen, um sicherzustellen, dass mit einer Wahrscheinlichkeit von mindestens $\alpha \in (0,1)$ die relative Anzahl der Befürworter nicht mehr als 5% von der wahren Zustimmungsrate $p$ abweicht?

*Proof.* (a) Die Anzahl der zustimmenden Menschen ist $Np$, wobei $N$ die gesamte Anzahl von Personen ist. Wenn man eine Person zufällig wählt, ist die Wahrscheinlichkeit dafür, dass sie zustimmt, $p$. Für $N$ groß genug bleibt die Wahrscheinlichkeit $p$ bei weitere zufällige gewählte Personen. Damit ist die Verteilung eine Bernoulli-Verteilung.

(b) Nach Moivre-Laplace ist

$$\mathbb{P}\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \approx \Phi(b) - \Phi(a)$$

$$\mathbb{P}\left(a\sqrt{np(1-p)} \leq S_n - np \leq b\sqrt{np(1-p)}\right) \approx \Phi(b) - \Phi(a)$$

* jun-wei.tan@stud-mail.uni-wuerzburg.de

$$\mathbb{P}\left(a\sqrt{\frac{p(1-p)}{n}} \le \frac{S_n}{n} - p \le b\sqrt{\frac{p(1-p)}{n}}\right) \approx \Phi(b) - \Phi(a)$$

$$\mathbb{P}\left(a \le \frac{S_n}{n} - p \le b\right) \approx \Phi\left(\frac{b\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(\frac{a\sqrt{n}}{\sqrt{p(1-p)}}\right)$$

Wegen der Symmetrie ist

$$\begin{aligned}\mathbb{P}\left(a \le \left|\frac{S_n}{n} - p\right| \le b\right) &= 2\mathbb{P}\left(a \le \frac{S_n}{n} - p \le b\right) \\ &\approx 2\left(\Phi\left(\frac{b\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(\frac{a\sqrt{n}}{\sqrt{p(1-p)}}\right)\right)\end{aligned}$$

Aufgrund der Stetigkeit ist

$$\mathbb{P}\left(a \le \left|\frac{S_n}{n} - p\right|\right) = \lim_{b\to\infty} \mathbb{P}\left(a \le \left|\frac{S_n}{n} - p\right| \le b\right).$$

Weil $\lim_{b\to\infty} \Phi(b) = 1$, ist

$$\mathbb{P}\left(a \le \left|\frac{S_n}{n} - p\right| \le b\right) \approx 2\left(1 - \Phi\left(\frac{a\sqrt{n}}{\sqrt{p(1-p)}}\right)\right).$$

(c) Wir wollen: $\epsilon = 5\% = 0.05$ und

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) \le 1 - \alpha.$$

Nach (b) ist

$$\begin{aligned}1 - \Phi\left(\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}\right) &\lesssim \frac{1}{2} - \frac{\alpha}{2} \\ \Phi\left(\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}\right) &\gtrsim \frac{1+\alpha}{2} \\ n &\gtrsim \frac{p(1-p)}{\epsilon^2}\Phi^{-1}\left(\frac{1+\alpha}{2}\right).\end{aligned}$$

□

**Problem 2.** (a) Seien $X_1, \ldots, X_n$ identisch verteilte, reellwertige Zufallsvariablen (nicht notwendigerweise unabhängig) mit $\mathbb{E}[X_1^2] < \infty$. Angenommen, es existiert ein festes $h \ge 1$, so dass $\mathrm{Cov}(X_j, X_k) = 0$ für $|j-k| \ge h$. Zeigen Sie unter dieser Annahme für $S_n = \sum_{i=1}^n X_i$ die Abschätzung

$$\mathrm{Var}(S_n) \le 2nh\,\mathrm{Var}(X_1).$$

(b) Folgern Sie, unter Verwendung von (a), dass auch unter diesen Annahmen ein schwaches Gesetz der großen Zahlen erfüllt ist.

*Proof.* (a) Es gilt

$$\begin{aligned}
\operatorname{Var}(S_n) &= \sum_{i,j=1}^{n} \operatorname{Cov}(X_i, X_j) && \text{Linearität} \\
&= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ |j-i|<h}}^{n} \operatorname{Cov}(X_i, X_j) \\
&\le \sum_{i=1}^{n} \sum_{\substack{j=1 \\ |j-i|<h}}^{n} \sqrt{\operatorname{Var}(X_i)}\sqrt{\operatorname{Var}(X_j)} && \text{Cauchy-Schwarz} \\
&= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ |j-i|<h}}^{n} \operatorname{Var}(X_1) && \text{identisch verteilt} \\
&\le \sum_{i=1}^{n} 2h\operatorname{Var}(X_1) \\
&= 2hn\operatorname{Var}(X_1)
\end{aligned}$$

(b) Wegen Linearität des Erwartungwerts ist noch $\mathbb{E}(S_n) = n\mathbb{E}(X_1)$. Nach (a) ist $\operatorname{Var}(S_n) < \infty$. Damit gilt nach Tschebyschev:

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1)\right| \ge \epsilon\right) &= \mathbb{P}\left(\left|\frac{S_n - n\mathbb{E}(X_1)}{n}\right| \ge \epsilon\right) \\
&= \mathbb{P}\left(\left|\frac{S_n - \mathbb{E}(S_n)}{n}\right| \ge \epsilon\right) \\
&= \mathbb{P}\left(|S_n - \mathbb{E}(S_n)| \ge n\epsilon\right) \\
&\frac{1}{(n\epsilon)^2}\operatorname{Var}(S_n) \\
&\le \frac{1}{(n\epsilon)^2}(2hn\operatorname{Var}(X_1)) \\
&\le \frac{2h\operatorname{Var}(X_1)}{n\epsilon^2}
\end{aligned}$$

also

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1)\right| \ge \epsilon\right) = 0. \qquad \square$$