Übungsblatt 14

Es sind keinerlei Abgaben und Korrekturen für dieses Blatt mehr vorgesehen.

Aufgabe 76: (Moivre-Laplace beim Flohsprung)

Ein Floh startet auf dem Ursprung der Zahlengeraden und springt jeweils mit der Wahrscheinlichkeit $p = 0.6$ um zwei Einheiten nach rechts bzw. mit der Wahrscheinlichkeit $q = 0.4$ um eine Einheit nach links. Sei Y_n die Position nach n Sprüngen.

a) Bestimme Erwartungswert und Streuung von Y_n!

b) Mit welcher Wahrscheinlichkeit befindet sich der Floh nach 10 000 Sprüngen im Intervall $[7\,700, 8\,090]$?

Musterlösung:
Sei S_n die Anzahl der Erfolge (= Sprünge nach rechts) bei n-facher Wiederholung des Bernoulli-Experiments „Flohsprung“. Es gilt $Y_n = 2\,S_n - (n - S_n) = 3\,S_n - n$.

a) Die Bi(n, p)-verteilte Variable S_n hat Erwartunswert $\mathrm{E}(S_n) = n\,p$ und Streuung $\sigma(S_n) = \sqrt{n\,p\,q}$:

$$\mathrm{E}(Y_n) = 3\,\mathrm{E}(S_n) - n = 3\,n\,p - n = 0.8\,n\ .$$

Es gilt $\sigma(\alpha\,X + \beta) = |\alpha|\,\sigma(X)$:

$$\begin{aligned}\mathrm{Var}(\alpha\,X + \beta) &= \mathrm{E}((\alpha X + \beta)^2) - (\mathrm{E}(\alpha X + \beta))^2\\ &= \alpha^2\,\mathrm{E}(X^2) + 2\,\alpha\,\beta\,\mathrm{E}(X) + \beta^2 - \alpha^2\,(\mathrm{E}(X))^2 - 2\,\alpha\,\beta\,\mathrm{E}(X) - \beta^2\\ &= \alpha^2\,(\mathrm{E}(X^2) - (\mathrm{E}(X))^2) = \alpha^2\,\mathrm{Var}(X)\ .\end{aligned}$$

Damit folgt unmittelbar

$$\sigma(Y_n) \;=\; \sigma(3\,S_n - n) \;=\; 3\,\sigma(S_n) \;=\; 3\,\sqrt{n\,p\,q} \;\approx\; 1.47\,\sqrt{n}\ .$$

b) Mit $n = 10\,000$ gilt:

$$\begin{aligned} P(7700 \le Y_n \le 8090) \;&=\; P(7700 \le 3\,S_n - n \le 8100)\\ =\; P(17\,700 \le 3\,S_n \le 18\,090) \;&=\; P(5\,900 \le S_n \le 6\,030)\ .\end{aligned}$$

Nach Folgerung 4.9 der Vorlesung kann diese Wahrscheinlichkeit für die Bi(n, p)-verteilte Variable S_n mittels Moivre-Laplace berechnet werden:

$$\begin{aligned} P(\underbrace{5\,900}_{a'} \le S_n \le \underbrace{6\,030}_{b'}) \;&\approx\; \frac{1}{\sqrt{2\,\pi}} \int_{\frac{a'-n\,p}{\sqrt{n\,p\,q}}}^{\frac{b'-n\,p}{\sqrt{n\,p\,q}}} e^{-x^2/2}\,dx\\ &=\; \frac{1}{2}\,\Big(\Phi\Big(\frac{b' - n\,p}{\sqrt{n\,p\,q}}\Big) + \Phi\Big(\frac{|a' - n\,p|}{\sqrt{n\,p\,q}}\Big)\Big)\\ &=\; \frac{1}{2}\,\Big(\Phi\Big(\frac{30}{\sqrt{10^4 \cdot 0.6 \cdot 0.4}}\Big) + \Phi\Big(\frac{100}{\sqrt{10^4 \cdot 0.6 \cdot 0.4}}\Big)\Big)\\ &=\; \frac{1}{2}\,(\Phi(0.6124) + \Phi(2.0412)) \;=\; 0.709\ .\end{aligned}$$

Aufgabe 77: (Moivre-Laplace in Mario Puzos 'Fools Die')
Der Croupier am Tisch 1 eines Casinos in Las Vegars arbeitet mit Spielern in betrügerischer Weise zusammen, indem er das Roulette-Rad so präpariert hat, dass die 17 mit einer Wahrscheinlichkeit von 1.1/38 erscheint. (Normal wäre 1/38: Las Vegas hat beim Roulette eine Doppel-Null). Nach auffälligen Verlusten beginnt das Casino Aufzeichnungen zu führen. Wie lange müssen die Ergebnisse des ersten Tisches aufgezeichnet werden, um mit 90%-iger Sicherheit sagen zu können, dass das Rad manipuliert ist?

Musterlösung:
Das Bernoulli-Experiment „die 17 erscheint" wird n-mal wiederholt, die W'keit p soll durch die relative Häufigkeit $\bar{X}_n$ abgeschätzt werden. Nach vielen Aufzeichnungen wird das Casino eine relative Häufigkeit $\bar{X}_n \approx 1.1/38$ feststellen.
Wie genau sollte man p festlegen, um von „Manipulation" sprechen zu können (beachte, auch bei einem fairen Rad wird nicht exakt $p = 1/38$ gelten)? Wäre das Rad fair, so müsste $|p - 1/38| < \epsilon$ gelten mit unbekanntem, aber sehr kleinem $\epsilon \ll 0.1/38$. Wir wollen mit 90%-iger Sicherheit sagen können, dass

$$p \ \geq \ \frac{1}{38} + \epsilon$$

gilt, dann sollte man von 'Manipulation' reden.
Ein gemessener Mittelwert $\bar{x}_n \approx \frac{1.1}{38}$ für $\bar{X}_n$ liege vor. Wir fragen nach

$$P\Big(\bar{X}_n - p \leq \bar{x}_n - \frac{1}{38} - \epsilon\Big),$$

was wir gemäß Interpretation 2, 4.3 der Skriptes nach Ersetzen von $\bar{X}_n$ durch den Messwert $\bar{x}_n$ interpretieren als die Sicherheit

$$P\Big(\bar{x}_n - p \leq \bar{x}_n - \frac{1}{38} - \epsilon\Big) = P\Big(\frac{1}{38} + \epsilon \leq p\Big),$$

mit der die Aussage $p \ \geq \ \frac{1}{38} + \epsilon$ gilt. Hierfür ist der Sollwert $S = 0.9$ vorgegeben. Nach den Folgerung 4.10 des Skripts gilt in der Moivre–Laplace–Näherung:

$$S \ = \ 0.9 \ = \ P\Big(\bar{X}_n - p \leq \bar{x}_n - \frac{1}{38} - \epsilon\Big) \ \approx \ \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{(\bar{x}_n - \frac{1}{38} - \epsilon)\ \sqrt{\frac{n}{p\,q}}} e^{-x^2/2}\, dx$$

$$= \ \frac{1}{2}\ \Big(\Phi\Big(\ (\bar{x}_n - \frac{1}{38} - \epsilon)\ \sqrt{\frac{n}{p\,q}}\ \Big) + 1\Big)\ .$$

Es folgt

$$\Phi\Big((\bar{x}_n - \frac{1}{38} - \epsilon)\ \sqrt{\frac{n}{p\,q}}\ \Big) \ \approx \ 2 \cdot 0.9 - 1 \ = \ 0.8$$

$$\Rightarrow \quad \Big(\bar{x}_n - \frac{1}{38} - \epsilon\Big)\ \sqrt{\frac{n}{p\,q}} \ = \ \Phi^{-1}(0.8) \approx 1.282$$

$$\Rightarrow \quad n \ \approx \ \frac{p\,q \cdot 1.282^2}{\Big(\bar{x}_n - \frac{1}{38} - \epsilon\Big)^2}\ .$$

Die Gültigkeit dieser Gleichung kann erst im Laufe der Aufzeichnung überprüft werden, wenn Werte $\bar{x}_n$ für $\bar{X}_n$ vorliegen. Setzen wir für eine „Hochrechnung“ $\bar{x}_n \approx 1.1/38$ und vernachlässigen wir $\epsilon \ll 0.1/38$, so gilt mit $p \approx 1/38$, $q \approx 37/38$ (bzw. $p \approx 1.1/38$, $q \approx 36.9/38$, das macht praktisch keinen Unterschied):

$$n \;\approx\; \frac{\frac{1}{38} \cdot \frac{37}{38} \cdot 1.282^2}{\left(\frac{0.1}{38} - \epsilon\right)^2} \;\approx\; \frac{\frac{1}{38} \cdot \frac{37}{38} \cdot 1.282^2}{\left(\frac{0.1}{38}\right)^2} \;\approx\; 6\,081\,.\,04\ .$$

Aufgabe 78: (Moivre-Laplace vor der Wahl)

Wieviele Wähler muss man befragen, um den Stimmenanteil einer Partei mit einer Sicherheit von 95.4% auf eine absolute Genauigkeit von ±1% vorhersagen zu können? Beantworte diese Frage einmal über das Gesetz der großen Zahl und dann über die Moivre–Laplace–Näherung und vergleiche. Macht es einen Unterschied, ob man am Stimmenanteil der CDU oder der Grünen interessiert ist? Gib realistische Werte an!

Musterlösung:

Betrachte das Bernoulli-Experiment X: „frage einen Wähler, ob er die Partei 'XYZ' wählen wird, und werte die Antwort 'ja' als Erfolg“. Die Erfolgsw'keit $p = P(„\text{ja}“)$ entspricht dem Anteil der Bevölkerung, die diese Partei wählen wird, also $p =$ „Stimmenanteil von 'XYZ'“.

Durch n Befragungen soll p durch die relative Häufigkeit $\bar{X}_n = S_n/n$ geschätzt werden ($S_n =$ Anzahl der Erfolge). Das Gesetz der großen Zahl liefert

$$S \;=\; P(|\bar{X}_n - \mu| < \epsilon) \;\geq\; 1 - \frac{\sigma^2}{\epsilon^2\, n}\ .$$

Mit dem Erwartungswert $\mu = \mathrm{E}(X) = p$ und der Varianz $\sigma^2 = p\,q = p\,(1-p) \leq 1/4$ folgt

$$n \;\geq\; \frac{p\,(1-p)}{(1-S)\,\epsilon^2}$$

bei approximativ bekanntem p bzw.

$$n \;\geq\; \frac{1}{4\,(1-S)\,\epsilon^2}$$

bei völlig unbekanntem p. Mit $S = 0.954$ und $\epsilon = 0.01$ reichen damit

$$n \;\geq\; \frac{1}{4 \cdot 0.046 \cdot 0.01^2} \;\approx\; 54\,348 \text{ (Befragungen)}$$

um den Stimmenanteil aller Parteien mit einer Sicherheit von 95.4% auf ±1% festlegen zu können. Mit Approximationen von p bekommt man bessere Werte, z.B.:

$$\begin{array}{llllll} \text{CDU}: & p \;\approx\; 0.35 & \Rightarrow & n \;\geq\; 49\,457 & \text{Befragungen reichen,} \\ \text{Grüne}: & p \;\approx\; 0.07 & \Rightarrow & n \;\geq\; 14\,152 & \text{Befragungen reichen.} \end{array}$$

Nach Moive-Laplace gilt

$$S \;=\; P(|\bar{X}_n - \mu| < \epsilon) \;\approx\; \Phi\Big(\epsilon\,\sqrt{\frac{n}{p\,q}}\Big)\ ,$$

also

$$n \;\approx\; \frac{p\,(1-p)}{\epsilon^2}\,(\Phi^{-1}(S))^2 \;\leq\; \frac{1}{4\,\epsilon^2}\,(\Phi^{-1}(S))^2\ .$$

Mit $S = 0.954$, $\Phi^{-1}(0.954) \approx 2.0$, $\epsilon = 0.01$ folgt

$$n \;\approx\; p\,(1-p)\;40\,000 \;\leq\; 10\,000\;.$$

Damit reichen 10 000 Befragungen für alle Parteien. Mit Approximationen von p bekommt man bessere Werte, z.B.:

$$\begin{array}{llllll} \text{CDU}: & p \;\approx\; 0.35 & \Rightarrow & n \;\approx\; 9\,100 & \text{Befragungen reichen,} \\ \text{Grüne}: & p \;\approx\; 0.07 & \Rightarrow & n \;\approx\; 2\,604 & \text{Befragungen reichen.} \end{array}$$

Aufgabe 79: (Moivre-Laplace am Wahlabend)

Am Wahlabend wurden bereits 1 000 000 Stimmen ausgezählt. Wie genau lässt sich das Wahlergebnis voraussagen, wenn man eine 99%-ige Sicherheit der Hochrechnung fordert? Beantworte diese Frage einmal über das Gesetz der großen Zahl und dann über die Moivre–Laplace–Näherung und vergleiche. Macht es einen Unterschied, ob man am Stimmenanteil der CDU oder der Grünen interessiert ist? Gib realistische Werte an!

Musterlösung:

Betrachte das Bernoulli-Experiment X: „wähle einen Stimmzettel, sehe nach, ob die Partei 'XYZ' gewählt wurde, und werte das Ergebnis 'ja' als Erfolg“. Die Erfolgsw'keit $p = P(„\text{ja}“)$ entspricht dem Stimmenanteil von 'XYZ'. Betrachte die Auszählung von n Stimmzetteln als n-fache Wiederholung des Bernoulli-Experiments (was nur approximativ stimmt, denn die Auswahl eines auszuzählenden Stimmzettels aus der Menge alle Stimmzettel geschieht ja „ohne Zurücklegen“, d.h., das Experiment ändert sich mit jedem ausgezählten Stimmzettel geringfügig).

Bei n ausgezählten Stimmzetteln soll p durch die relative Häufigkeit $\bar{X}_n = S_n/n$ geschätzt werden (S_n = Anzahl der Erfolge). Das Gesetz der großen Zahl liefert

$$S \;=\; P(|\bar{X}_n - \mu| < \epsilon) \;\geq\; 1 - \frac{\sigma^2}{\epsilon^2\; n}\;.$$

Mit dem Erwartungswert $\mu = \mathrm{E}(X) = p$ und der Varianz $\sigma^2 = p\,q = p\,(1-p) \leq 1/4$ folgt

$$\epsilon \;\leq\; \sqrt{\frac{p\,(1-p)}{n\,(1-S)}} \;\leq\; \sqrt{\frac{1}{4\,n\,(1-S)}}\;.$$

Für $n = 10^6$ sind damit die Stimmenanteile aller Parteien mit der Sicherheit $S = 0.99$ auf die Genauigkeit

$$\epsilon \;\leq\; \sqrt{p\,(1-p)}\;0.01 \;\leq\; 0.005$$

(also 0.5%) festgelegt. Mit Approximationen von p bekommt man genauere Werte, z.B.:

$$\begin{array}{llllll} \text{CDU}: & p \;\approx\; 0.35 & \Rightarrow & \epsilon \;\leq\; 0.0048 & (0.48\%), \\ \text{Grüne}: & p \;\approx\; 0.07 & \Rightarrow & \epsilon \;\leq\; 0.0026 & (0.26\%). \end{array}$$

Nach Moive-Laplace gilt

$$S \;=\; P(|\bar{X}_n - \mu| < \epsilon) \;\approx\; \Phi\Big(\epsilon\,\sqrt{\frac{n}{p\,q}}\Big)\;,$$

also

$$\epsilon \;\approx\; \sqrt{\frac{p\,(1-p)}{n}}\;\Phi^{-1}(S) \;\leq\; \sqrt{\frac{1}{4\,n}}\;\Phi^{-1}(S)\;.$$

Mit $S = 0.99$, $\Phi^{-1}(0.99) \approx 2.58$, $n = 10^6$ folgt

$$\epsilon \;\approx\; \sqrt{p\,(1-p)}\;0.00258 \;\leq\; 0.00129\;.$$

Damit sind die Stimmenanteile aller Parteien auf mindestens 0.13% genau bestimmt. Mit Approximationen von p bekommt man genauere Werte, z.B.:

$$\begin{array}{lllll} \text{CDU}: & p \;\approx\; 0.35 & \Rightarrow & \epsilon \;\approx\; 0.0012 & (0.12\%), \\ \text{Grüne}: & p \;\approx\; 0.07 & \Rightarrow & \epsilon \;\approx\; 0.0007 & (0.07\%). \end{array}$$

Dies ist erstaunlich und wenig intuitiv: hat man von den etwa 50 Millionen Wählern nur 1 Millionen ausgezählt, ist das Wahlergebnis praktisch bekannt.
