Technische Universität München
Fakultät für Informatik
Lehrstuhl für Effiziente Algorithmen
Prof. Dr. Ernst W. Mayr
Dr. Werner Meixner

Sommersemester 2010
Lösungsblatt 10
9. Juli 2010

Diskrete Wahrscheinlichkeitstheorie

Hausaufgabe 1 (5 Punkte)

Wir betrachten unabhängige stetige Zufallsvariablen X und Y, die beide auf dem Intervall $[0,1] \subseteq \mathbb{R}$ gleichverteilt sind. Sei $Z = \max\{X, Y\}$.

1. Berechnen Sie die Verteilungsfunktion F_Z.
2. Bestimmen Sie eine Funktion $u : [0,1] \to [0,1]$, so dass $u(X)$ die gleiche Verteilung wie Z besitzt.

Lösungsvorschlag

1. Die gemeinsame Dichte der Zufallsvariablen X, Y sei $f_{X,Y}(x,y)$. Aufgrund der Unabhängigkeit von X, Y gilt für $(x,y) \notin [0,1] \times [0,1]$ die gemeinsame Dichte 0 und für $(x,y) \in [0,1] \times [0,1]$

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) = 1 \cdot 1 = 1\,.$$

Wir berechnen die Verteilungsfunktion $F_Z(z)$. Offenbar gilt zunächst $F_Z(z) = 0$ bzw. $F_Z(z) = 1$ für $z \leq 0$ bzw. $1 \leq z$. Für $z \in [0,1] \subseteq \mathbb{R}$ gilt

$$\begin{aligned} F_Z(z) &= \Pr[\max\{X,Y\} \leq z] \\ &= \Pr[X \leq z, Y \leq z] \\ &= \int_{[0,z]\times[0,z]} f_{X,Y}(x,y)\,\mathrm{d}x\mathrm{d}y \\ &= \int_{[0,z]\times[0,z]} 1\,\mathrm{d}x\mathrm{d}y \\ &= z^2\,. \end{aligned}$$

2. Aus der Vorlesung ist bekannt, dass wir eine Simulation von F_Z aus der Inversen von F_Z erhalten können.

Wir rechnen direkt und setzen die Invertierbarkeit von u voraus. Sei $Y = u(X)$. Dann gilt

$$\begin{aligned} F_Y(y) &= \Pr[Y \leq y] \;=\; \Pr[u(X) \leq y] \\ &= \Pr[X \leq u^{-1}(y)] \\ &= F_X(u^{-1}(y)) \\ &= u^{-1}(y)\,. \end{aligned}$$

Aus der Gleichung $F_Z(y) = F_X(u^{-1}(y))$ folgt nun $y^2 = u^{-1}(y)$, mithin

$$u(x) = \sqrt{x}\,.$$

Hausaufgabe 2 (5 Punkte)

Sei $X \sim \text{Bin}(10^5, \frac{1}{2})$ eine binomialverteilte Zufallsvariable. Sei $k \in \{100, 1000\}$.

1. Berechnen Sie mit Hilfe der Ungleichung von Chebyshev eine untere Schranke für $\Pr[5 \cdot 10^4 - k \leq X \leq 5 \cdot 10^4 + k]$.

2. Approximieren Sie $\Pr[5 \cdot 10^4 - k \leq X \leq 5 \cdot 10^4 + k]$ mit Hilfe des Satzes von de Moivre.

Lösungsvorschlag

1. Es gilt $\mathbb{E}[X] = 5 \cdot 10^4$. Deshalb gilt $5 \cdot 10^4 - k \leq X \leq 5 \cdot 10^4 + k$ genau dann, wenn $|X - \mathbb{E}[X]| \leq k$ gilt. Mit $\text{Var}[X] = 5 \cdot 10^4 \cdot \frac{1}{2} = 25 \cdot 10^3$ erhalten wir

$$\begin{aligned}\Pr[5 \cdot 10^4 - k \leq X \leq 5 \cdot 10^4 + k] &= \Pr[|X - \mathbb{E}[X]| \leq k] \\ &= 1 - \Pr[|X - \mathbb{E}[X]| > k] \\ &\geq 1 - \frac{\text{Var}[X]}{(k+1)^2} \\ &= 1 - \frac{25 \cdot 10^3}{(k+1)^2}\,.\end{aligned}$$

   Für $k = 100$ bzw. $k = 1000$ erhalten wir die unteren Schranken $-1.45...$ bzw. $0.97505\ldots$. Die erste Schranke ist als negative Zahl offensichtlich wertlos.

2. Sei

$$X^* = \frac{X - np}{\sqrt{np(1-p)}} = \frac{X - 5 \cdot 10^4}{5 \cdot 10\sqrt{10}}\,.$$

   Wir approximieren $X^* \sim N(0, 1)$ und erhalten

$$\begin{aligned}\Pr[5 \cdot 10^4 - k \leq X \leq 5 \cdot 10^4 + k] &= \Pr\left[-\frac{k}{5 \cdot 10\sqrt{10}} \leq X^* \leq \frac{k}{5 \cdot 10\sqrt{10}}\right] \\ &= \Pr\left[-6.32456 \cdot 10^{-3}k \leq X^* \leq 6.32456 \cdot 10^{-3}k\right] \\ &= \Phi(6.32456 \cdot 10^{-3}k) - \Phi(-6.32456 \cdot 10^{-3}k) \\ &= 2 \cdot \Phi(6.32456 \cdot 10^{-3}k) - 1\,.\end{aligned}$$

   Für $k = 100$ bzw. $k = 1000$ erhalten wir ≈ 0.4714 bzw. ≈ 1.

Hausaufgabe 3 (5 Punkte)

Seien $X_1, X_2, \ldots$ identisch verteilt und unabhängig mit $E[X_i] = 0,6$ und $\sigma = 0,3$. Wir betrachten $Y_n = \sum_{i=1}^{n} X_i$. Berechnen Sie die folgenden Grenzwerte.

1. $\lim_{n\to\infty} \Pr[\frac{Y_n}{n} = 0,6]$.

2. $\lim_{n\to\infty} \Pr[0,5 < Y_n < 0,7]$.

3. $\lim_{n\to\infty} \Pr[0,59 < \frac{Y_n}{n} < 0,61]$.

Lösungsvorschlag

Bei Grenzwertbestimmungen der nachfolgenden Art wenden wir den Zentralen Grenzwertsatz an.

1. Für stetige Zufallsvariable X gilt $\Pr[X = r] = 0$ für jeden Wert $r \in \mathbb{R}$. Da wir Y_n durch eine Normalverteilung approximieren, erhält man $\Pr[Y_n = 0] = 0$. Trivialerweise ist der Grenzwert für $n \to \infty$ ebenfalls 0 und wir brauchen nichts zu berechnen.

2. Auch hier ist der Grenzwert Null, allerdings muss er nun wie folgt berechnet werden. Wie üblich transformieren wir die Variable Y_n in die Variable $Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}}$.

$$\begin{aligned}\Pr[0,5 < Y_n < 0,7] &= \Pr\left[\frac{0,5-0,6n}{0,3\sqrt{n}} < Z_n < \frac{0,7-0,6n}{0,3\sqrt{n}}\right] \\ &= \Phi\left(\frac{0,7-0,6n}{0,3\sqrt{n}}\right) - \Phi\left(\frac{0,5-0,6n}{0,3\sqrt{n}}\right).\end{aligned}$$

   Nun folgt

$$\begin{aligned}\lim_{n\to\infty} \Pr[0,5 < Y_n < 0,7] &= \lim_{n\to\infty}\left(\Phi\left(\frac{0,7-0,6n}{0,3\sqrt{n}}\right) - \Phi\left(\frac{0,5-0,6n}{0,3\sqrt{n}}\right)\right) \\ &= 0\,.\end{aligned}$$

3. Rechnung analog wie oben:

$$\begin{aligned}\Pr[0,59 < \tfrac{Y_n}{n} < 0,61] &= \Pr[0,59n < Y_n < 0,61n] \\ &= \Pr\left[\frac{0,59n-0,6n}{0,3\sqrt{n}} < Z_n < \frac{0,61n-0,6n}{0,3\sqrt{n}}\right] \\ &= \Phi\left(\frac{0,01n}{0,3\sqrt{n}}\right) - \Phi\left(\frac{-0,01n}{0,3\sqrt{n}}\right).\end{aligned}$$

   Nun folgt

$$\begin{aligned}\lim_{n\to\infty} \Pr[0,59 < \tfrac{Y_n}{n} < 0,61] &= \lim_{n\to\infty}\left(\Phi\left(\frac{0,01n}{0,3\sqrt{n}}\right) - \Phi\left(\frac{-0,01n}{0,3\sqrt{n}}\right)\right) \\ &= 1 - 0 = 1\,.\end{aligned}$$

Hausaufgabe 4 (5 Punkte)

In einem Unfallkrankenhaus treffen im Schnitt alle 20 Minuten Patienten zur Behandlung ein. Die Zeit zwischen zwei Behandlungsfällen sei exponentialverteilt mit Parameter $\frac{1}{20}$. Wenn 1 Stunde lang kein Patient eingetroffen ist, macht das Personal Ruhepause. Wir wollen wissen, welcher Zeitabstand zwischen zwei Ruhepausen zu erwarten ist.
Seien $T_1, T_2, \ldots$ die Zeitspannen zwischen dem Eintreffen zweier Behandlungsfälle und W die Wartezeit bis zur nächsten Ruhepause.

1. Geben Sie $\mathbb{E}[T_1 \mid T_1 \geq 60]$ an.

2. Geben Sie $\mathbb{E}[W \mid T_1 \geq 60]$ an.

3. Berechnen Sie $\mathbb{E}[W]$.

Lösungsvorschlag

1. Es gilt $\mathbb{E}[T_1] = 20$. Da die Exponentialverteilung gedächtnislos ist, ergibt sich
$$\mathbb{E}[T_1 \mid T_1 \geq 60] = 60 + \mathbb{E}[T_1] = 80\,.$$

2. Offenbar gilt
$$\mathbb{E}[W \mid T_1 \geq 60] = 60\,.$$

3. Es gilt
$$\mathbb{E}[W] \;=\; \mathbb{E}[W|T_1 \leq 60] \cdot \Pr[T_1 \leq 60] \;+\; \mathbb{E}[W|T_1 \geq 60] \cdot \Pr[T_1 \geq 60]\,.$$
Andererseits gilt
$$\mathbb{E}[W \mid T_1 \leq 60] \;=\; \mathbb{E}[T_1 \mid T_1 \leq 60] + \mathbb{E}[W]\,.$$
Einsetzen von $\mathbb{E}[W|T_1 \leq 60]$ und $\mathbb{E}[W|T_1 \geq 60] = 60$ liefert
$$\mathbb{E}[W] \;=\; (\mathbb{E}[T_1 \mid T_1 \leq 60] + \mathbb{E}[W]) \cdot \Pr[T_1 \leq 60] \;+\; 60 \cdot \Pr[T_1 \geq 60]\,.$$
Nun wird nach $\mathbb{E}[W]$ aufgelöst.
$$\mathbb{E}[W] \;=\; \frac{\mathbb{E}[T_1|T_1 \leq 60] \cdot \Pr[T_1 \leq 60] \;+\; 60 \cdot \Pr[T_1 \geq 60]}{\Pr[T_1 \geq 60]}\,.$$
Wir berechnen die benötigten Werte.
$\Pr[T_1 \leq 60] = 1 - e^{-\frac{1}{20}\cdot 60} = 1 - e^{-3}$ und daraus $\Pr[T_1 \geq 60] = e^{-3}$.
$\mathbb{E}[T_1|T_1 \leq 60]$ erhalten wir aus den Gleichungen
$$\begin{aligned}\mathbb{E}[T_1] &= \mathbb{E}[T_1|T_1 \leq 60] \cdot \Pr[T_1 \leq 60] \;+\; \mathbb{E}[T_1|T_1 \geq 60] \cdot \Pr[T_1 \geq 60]\\ &= \mathbb{E}[T_1|T_1 \leq 60] \cdot (1 - e^{-3}) + 80 \cdot e^{-3}\\ &= 20\end{aligned}$$
wie folgt:
$$\mathbb{E}[T_1|T_1 \leq 60] \;=\; \frac{20 - 80 \cdot e^{-3}}{1 - e^{-3}}\,.$$
Ergebnis durch Einsetzen:
$$\mathbb{E}[W] = 20(e^3 - 1) \approx 382 \text{ (Minuten)}\,.$$

Hinweis: *Auf den Übungsblättern in diesem Semester wird es grundsätzlich die drei Aufgabentypen Vorbereitungsaufgabe, Tutoraufgabe und Hausaufgabe geben. Die als Vorbereitung bezeichneten Aufgaben dienen der häuslichen Vorbereitung der Tutoraufgaben. Tutoraufgaben werden in den Übungsgruppen bearbeitet. Dabei wird die Lösung der Vorbereitungsaufgaben vorausgesetzt. Die Vorbereitungsaufgaben werden in der Zentralübung unterstützt.*

Vorbereitung 1

Eine Werbeagentur möchte am letzten Tag der Fußballweltmeisterschaft mit einer Blitzumfrage schätzen, welcher Anteil ϑ der per Bahn anreisenden Fußballfans einen Platz im Stadion hat. Jeder der 12 Mitarbeiter befragt so lange zufällig ausgewählte Fans, bis er einen Fan gefunden hat, der eine Karte für das Stadion besitzt. Die Anzahl der vom Mitarbeiter i befragten Fans sei X_i.

Wir nehmen an, dass alle X_i die gleiche geometrische Verteilung besitzen mit

$$\Pr_\vartheta[X_i = k] \;=\; (1-\vartheta)^{k-1} \cdot \vartheta\,,\ \forall k \in \mathbb{N}\,.$$

1. Man bestimme auf der Basis der ermittelten Stichprobenwerte

   $$3,\ 1,\ 2,\ 3,\ 2,\ 1,\ 2,\ 1,\ 5,\ 3,\ 2,\ 3$$

   einen Maximum-Likelihood-Schätzwert für ϑ.

2. Man gebe mit Hilfe des Zentralen Grenzwertsatzes ein konkretes 95%-Konfidenzintervall für ϑ an.

Lösungsvorschlag

1. Für die Likelihood-Funktion $L(\vec{x};\vartheta)$ der Stichprobe $\vec{x} = (x_1, \ldots, x_{12})$ gilt mit $n = 12$

   $$\begin{aligned} L(\vec{x};\vartheta) &= \prod_{i=1}^{n} \Pr_\vartheta[X_i = x_i] \\ &= \prod_{i=1}^{n} \left[(1-\vartheta)^{x_i-1} \cdot \vartheta\right]. \end{aligned}$$

   Wir schreiben $\sum_{i=1}^n x_i - n = n\bar{x} - n$ und erhalten

   $$L(\vec{x};\vartheta) \;=\; (1-\vartheta)^{n\bar{x}-n} \cdot \vartheta^n\,.$$

   Gesucht ist der Wert $\vartheta = \bar{\vartheta}$ mit $0 \le \vartheta \le 1$, für den $L(\vec{x};\vartheta)$ das Maximum in dem $[0,1]$-Intervall annimmt.

   Falls $\bar{x} = 1$, dann folgt sofort $\vartheta = 1$. Dies bedeutet, dass mit maximaler Wahrscheinlichkeit (die der Test erlaubt) jeder Fan eine Stadionkarte besitzt.

   Falls $\bar{x} \neq 1$, dann bestimmen wir das Maximum von ϑ zwischen 0 und 1 im ersten Schritt durch Bestimmung einer Nullstelle der Ableitung von L nach ϑ wie folgt.

$$\begin{aligned}
\frac{\mathrm{d}\, L(\vec{x};\vartheta)}{\mathrm{d}\,\vartheta} &= (n\bar{x}-n)(1-\vartheta)^{n\bar{x}-n-1}(-1)\vartheta^n + (1-\vartheta)^{n\bar{x}-n} n\vartheta^{n-1} \\
&= (1-\vartheta)^{n\bar{x}-n-1}\vartheta^{n-1}[(n\bar{x}-n)(-\vartheta)+n(1-\vartheta)] \\
&= (1-\vartheta)^{n\bar{x}-n-1}\vartheta^{n-1}[n-n\bar{x}\vartheta]
\end{aligned}$$

Als Nullstellen der Ableitung der Likelihood-Funktion erhalten wir $\vartheta = 0$, $\vartheta = 1$ und $\vartheta = \frac{1}{\bar{x}}$. Man beachte, dass $0 < \frac{1}{\bar{x}} < 1$ gilt. Da

$$L(\vec{x};0) = L(\vec{x};1) = 0 \quad \text{und} \quad L\left(\vec{x};\frac{1}{\bar{x}}\right) > 0$$

gilt, scheiden $\vartheta = 0$ und $\vartheta = 1$ als Maximumstellen aus. Es folgt

$$\bar{\vartheta} = \frac{1}{\bar{x}}\,.$$

Nun berechnen wir $\frac{1}{\bar{x}}$ aus der Stichprobe mit $\bar{x} = \frac{7}{3}$ und erhalten

$$\bar{\vartheta} = \frac{3}{7}\,.$$

Es haben also bei maximaler Testwahrscheinlichkeit $\frac{3}{7}$ der Fans Stadionkarten.

2. Die Stichprobe liefert den Wert $\bar{x} = \frac{7}{3}$ für die Zufallsvariable $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ mit $n = 12$. Für das Konfidenzintervall für ϑ zum Konfidenzniveau $1-\alpha = 0.95$ setzen wir an

$$\Pr\left[-c \le \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} \le c\right] \approx 0.95$$

mit dem Quantil $c = x_{1-\frac{\alpha}{2}} = x_{0.975} \approx 1.96$ und den Gleichungen $\mu_{\bar{X}} = \frac{1}{\vartheta}$ und $\sigma_{\bar{X}}^2 = \frac{1-\vartheta}{n\vartheta^2}$ für $n = 12$. Wir bestimmen nun die Menge aller ϑ, für die die Ungleichung $-c \le \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} \le c$ gilt.

$$\begin{aligned}
-c \le \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} \le c &\iff (\bar{X}-\mu_{\bar{X}})^2 \le c^2\sigma_{\bar{X}}^2 \\
&\iff (\bar{X}-\frac{1}{\vartheta})^2 \le c^2 \cdot \frac{1}{n} \cdot \frac{1-\vartheta}{\vartheta^2} \\
&\iff (\vartheta\bar{X}-1)^2 \le \frac{c^2}{n}\cdot(1-\vartheta) \\
&\iff \vartheta^2\bar{X}^2 + (\frac{c^2}{n}-2\bar{X})\vartheta + (1-\frac{c^2}{n}) \le 0
\end{aligned}$$

Nun setzen wir sämtliche Zahlenwerte ein und erhalten (wegen $c \approx 1.96$ näherungsweise)

$$-c \le \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} \le c \iff 49\vartheta^2 - 39.1188\vartheta + 6.1188 \le 0$$

Wenn wir nach bekannter Formel die Nullstellen der quadratischen Gleichung

$$49\vartheta^2 - 39.1188\vartheta + 6.1188 = 0$$

mit α_1, α_2 bezeichnen, wobei wir $\alpha_1 < \alpha_2$ annehmen können, dann ergibt sich als Lösung

$$\alpha_1 \leq \vartheta \leq \alpha_2$$

mit $\alpha_1 \approx 0.2135$ und $\alpha_2 \approx 0.5848$.

Tutoraufgabe 1

Ziel dieser Aufgabe ist es, einen ML-Schätzer für die Erfolgswahrscheinlichkeit $p \in (0,1)$ einer negativ binomialverteilten Zufallsvariablen zu bestimmen. Hierfür seien $X_1, \ldots, X_n$ unabhängige Zufallsvariable, wobei jedes X_i negativ binomialverteilt mit Erfolgswahrscheinlichkeit p bei m zu erzielenden Erfolgen sei, d. h., jedes X_i hat die Dichte

$$f_{X_i}(k) = \binom{k-1}{m-1} p^m (1-p)^{k-m} \quad (\text{mit } k \geq m)\,.$$

Der Parameter m sei bekannt. Zu schätzen ist p.

1. Es sei $\vec{k} = (k_1, \ldots, k_n)$ ein Stichprobenvector (mit $k \geq m$). Stellen Sie die Likelihood-Funktion $L(\vec{k}; p)$ auf.

2. Maximieren Sie $L(\vec{k}; p)$ und bestimmen Sie den entsprechenden ML-Schätzer für p.

3. Zeigen Sie, dass der hergeleitete ML-Schätzer i. A. nicht erwartungstreu ist.

   Hinweis: Verwenden Sie $\ln(1+x) = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{x^i}{i}$ für $x \in (-1, 1]$.

Lösungsvorschlag

1.

$$\begin{aligned}
L(\vec{k}; p) &= \prod_{i=1}^{n} f_{X_i}(k_i) \\
&= \prod_{i=1}^{n} \left(\binom{k_i - 1}{m-1} p^m (1-p)^{k_i - m} \right) \\
&= \left(\frac{p}{1-p} \right)^{mn} \cdot (1-p)^{\overbrace{\sum_{i=1}^{n} k_i}^{=:K}} \cdot \underbrace{\prod_{i=1}^{n} \binom{k_i - 1}{m-1}}_{=:C} \\
&= p^{mn} \cdot (1-p)^{K - mn} \cdot C\;.
\end{aligned}$$

2. Wir bestimmen das Maximum von L analog wie in der VA 1 durch Bildung der Ableitung und eine Begründung, dass die 2. Ableitung an der gefundenen Stelle negativ sein muss.

$$\begin{aligned}
\frac{\mathrm{d}\, L(\vec{k}; p)}{\mathrm{d}\, p} &= Cmn \cdot p^{mn-1} \cdot (1-p)^{K-mn} - Cp^{mn} \cdot (K - mn) \cdot (1-p)^{K-mn-1} \\
&= Cp^{mn-1} \cdot (1-p)^{K-mn-1} \cdot [mn(1-p) - p(K - mn)] \\
&= Cp^{mn-1} \cdot (1-p)^{K-mn-1} \cdot [mn - pK]\,.
\end{aligned}$$

Die Nullstellen sind $p = 0$, $p = 1$ und $p = \frac{mn}{K}$, wobei die mittlere Nullstelle das Maximum von L liefert. Der gesuchte Schätzer für die Erfolgswahrscheinlichkeit p ist also

$$U = \frac{mn}{\sum_{i=1}^{n} X_i}\,.$$

3. U ist erwartungstreu, falls $E[U] = p$ gilt. U ist aber nicht einmal für $m = n = 1$ erwartungstreu, wie die folgende Rechnung zeigt.

   Seien $m = n = 1$. Dann gilt $U = \frac{1}{X_1}$ und X_1 ist geometrisch verteilt:

$$\mathbb{E}[U] = \mathbb{E}\left[\frac{1}{X_1}\right] = \sum_{i=1}^{\infty} \frac{1}{i} \cdot p \cdot (1-p)^{i-1} = \frac{p}{1-p} \sum_{i=1}^{\infty} \frac{(1-p)^i}{i}\,.$$

   Wir berechnen die unendliche Reihe (siehe Hinweis):

$$-\ln p = -\ln(1 + (p-1)) = -\sum_{i=1}^{\infty} (-1)^{i+1} \frac{(p-1)^i}{i} = \sum_{i=1}^{\infty} \frac{(1-p)^i}{i}\,.$$

   Damit folgt

$$\mathbb{E}[U] = p \cdot \frac{-\ln p}{1-p} \neq p\,.$$

Tutoraufgabe 2

Beim Testen von Hypothesen bezeichnet man die zu überprüfende Hypothese (Nullhypothese) generell mit H_0 und die Alternative mit H_1. Ein Tierhändler erhält ein Paket mit 100 Frettchen. Er will testen, ob weniger als zehn (< 10) dieser Frettchen aggressiv und bissig sind. Dazu hält er zehn Frettchen seinen Finger hin und nimmt das Paket nur an, wenn ihn keines davon beißt (wir nehmen an, dass ein aggressives Frettchen sofort zubeißen würde). Wie lauten die Hypothesen des Händlers? Was ist das Signifikanzniveau des Tests?

Lösungsvorschlag

Die Zahl X der aggressiven Frettchen in der Stichprobe vom Umfang 10 ist hypergeometrisch verteilt mit $N = 100$ und S= Anzahl der aggressiven Frettchen im Paket. Der Händler testet die Hypothese

$$H_0 : S \leq 9$$

gegen

$$H_1 : S \geq 10$$

und verwirft die Nullhypothese dabei im Fall $X \geq 1$. Für $S = 9$ ist

$$\Pr_S[X > 0] = 1 - \Pr_S[X = 0] = 1 - \frac{\binom{9}{0}\binom{91}{10}}{\binom{100}{10}} = 62.9\%.$$

Das ist auch das effektive Niveau des Tests, denn für $S < 9$ ist die Verwerfungswahrscheinlichkeit kleiner.

Tutoraufgabe 3

Sei X eine Bernoulli-verteilte Zufallsvariable mit Erfolgswahrscheinlichkeit p. Mit einer Stichprobe von nur 2 Elementen soll die Hypothese $H_0 : p \geq \frac{2}{3}$ getestet werden. Die Testvariable sei $T = X_1 + X_2$, wobei X_1 und X_2 unabhängige Kopien von X sein sollen, d. h., X_1, X_2 sind ebenfalls Bernoulli-verteilt mit gleichem Parameter p. Der Ablehnungsbereich des Tests sei $K = \{0\}$.

1. Geben Sie die Verteilungsfunktion von T an. (Es ist nicht nach der Dichtefunktion von T gefragt!)

2. Berechnen Sie die Fehlerwahrscheinlichkeit 1. Art (Signifikanzniveau) α_1 .

3. Wir nehmen an, dass die Alternative $H_1 : p \leq \frac{1}{3}$ echt ist. Dies bedeutet, dass H_1 gelte, wenn H_0 nicht gilt.

   Berechnen Sie damit die Fehlerwahrscheinlichkeit 2. Art α_2 .

Lösungsvorschlag

1. T ist binomialverteilt auf den Werten aus $W_T = \{0, 1, 2\}$ mit Dichtefunktion

   $$f_T(t) = \binom{2}{t} p^t (1-p)^{2-t}$$

   Für die Verteilungsfunktion F_T ergibt sich

   $$\begin{aligned} F_T(0) &= (1-p)^2, \\ F_T(1) &= (1-p)^2 + \binom{2}{1} p^1 (1-p)^1, \\ &= 1 - p^2, \\ F_T(2) &= 1. \end{aligned}$$

   Im Übrigen gilt $F_T(t) = 1$ für $t \geq 2$.

2. $$\begin{aligned} \alpha_1 &= \max_{p \geq \frac{2}{3}} \Pr[T \in K] \\ &= \max_{p \geq \frac{2}{3}} F_T(0) \\ &= \max_{p \geq \frac{2}{3}} (1-p)^2 = \frac{1}{9} \,. \end{aligned}$$

3. $$\alpha_2 = \max_{p \leq \frac{1}{3}} \Pr[T \notin K] \,.$$

   $$\begin{aligned} \alpha_2 &= \max_{p \leq \frac{1}{3}} (1 - F_T(0)) \\ &= \max_{p \leq \frac{1}{3}} (1 - (1-p)^2) = \frac{5}{9} \,. \end{aligned}$$