Übungsblatt 9

Mit $*$ und $**$ gekennzeichnete Aufgaben können zum Sammeln von Bonuspunkten verwendet werden. Lösungen von $*$-Aufgaben sind schriftlich abzugeben im Zettelkasten Nr. 5 auf dem D1 bis Mittwoch, 13.6.07, 11:00 Uhr. Lösungen von $**$-Aufgaben sind per Web-Formular unter `http://www.math.upb.de/~walter` ($\longrightarrow$ Lehre SS 07 $\longrightarrow$ Übungen) abzuliefern bis spätestens Mittwoch, 13.6.07, $23^{\underline{59}}$ Uhr.

Aufgabe 49:** (Erwartungswerte. 20 Punkte)

Dies ist eine Online-Aufgabe, die bis zum 13.6.06, 23^{59} Uhr, abzuliefern ist.

Betrachte die Verteilungsfunktion

$$F_X(r) = \begin{cases} 0 & \text{für } r < 0, \\ r/2 & \text{für } 0 \leq r < 1, \\ 3/5 & \text{für } 1 \leq r < 2, \\ (3 \cdot r + 1)/10 & \text{für } 2 \leq r < 3, \\ 1 & \text{für } 3 \leq r. \end{cases}$$

einer Zufallsvariablen X (die Daten werden vom Aufgabenserver zufällig abgeändert). Berechne den Erwartungswert!

Anleitung: Da keiner der Fälle „X ist diskret“ bzw. „X ist kontinuierlich“ in reiner Form vorliegt, benutze man Satz 2.21.4) der Vorlesung, um das Stieltjes-Integral des Erwartungswertes auf ein „normales“ Riemann-Integral zurückzuführen.

Musterlösung:

Es ist das Stieltjes-Integral

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} r \, dF_X(r)$$

zu berechnen. Da keiner der einfachen Fälle „X ist diskret“ bzw. „X ist kontinuierlich“ vorliegt, benutzen wir Satz 2.21.4) der Vorlesung mit $f(r) = r$, um das Stieltjes-Integral auf ein „normales“ Riemann-Integral zurückzuführen:

$$\int_a^b r \, dF_X(r) = b \cdot F_X(b) - a \cdot F_X(a) - \int_a^b F_X(r) \, dr.$$

Hierbei können wir $a = 0$ und $b = 3$ wählen, da außerhalb des Intervalls $[0, 3]$ die Verteilungsfunktion konstant ist und damit die Bereiche $(-\infty, 0]$ und $(3, \infty)$ nicht zum Stieltjes-Integral beitragen. Also:

$$\mathrm{E}(X) = 3 \cdot F_X(3) - 0 \cdot F_X(0) - \int_0^3 F_X(r) \, dr = 3 - \int_0^1 F_X(r) \, dr - \int_1^2 F_X(r) \, dr - \int_2^3 F_X(r) \, dr$$

$$= 3 - \int_0^1 \frac{r}{2} \, dr - \int_1^2 \frac{3}{5} \, dr - \int_2^3 \frac{3 \cdot r + 1}{10} \, dr = 3 - \frac{1}{4} - \frac{3}{5} - \frac{17}{20} = \frac{13}{10}.$$

Aufgabe 50: (Erwartungswerte)

Ein Raumschiff bestehe aus 3 Teilen A, B und C, die jeweils völlig zerstört werden, wenn sie von (mindestens) einem Meteoriten getroffen werden. Das gesamte Raumschiff fällt aus, wenn A oder (B und C) ausfallen. Wieviele Meteoriten müssen das Raumschiff im Durchschnitt treffen, bis es ausfällt?

Musterlösung:
Das Modell: die Meteoriteneinschläge seien durch

$$\Omega = \{(m_1, m_2, m_3, \ldots); m_i \in \{A, B, C\}\}$$

modelliert, wobei $m_i = A$ bedeutet, dass der i-te Meteoriteneinschlag A trifft. Zur Abkürzung setzen wir

$$\begin{aligned} A &= \{(A, m_2, m_3, \ldots); m_2, m_3, \ldots \in \{A, B, C\}\}, \\ BC &= \{(B, C, m_3, \ldots); , m_3, m_4, \ldots \in \{A, B, C\}\}, \\ CB &= \{(C, B, m_3, \ldots); , m_3, m_4, \ldots \in \{A, B, C\}\} \end{aligned}$$

usw. (z.B., nach Einschlag in A ist egal, wo weitere Meteoriten einschlagen etc.). Seien $p_A = P(A), p_B = P(B), p_C = P(C)$ mit $p_A + p_B + p_C = 1$ vorgegeben. Dies sind die Wahrscheinlichkeiten, dass ein einschlagender Meteorit jeweils A,B oder C trifft. Diese Wahrscheinlichkeiten sind etwa durch die relativen Größen der Bauteile A, B und C bestimmt.
Mit der Unabhängigkeit der Einschläge gilt

$$P(BC) = p_B\, p_C \ , \ P(BBA) = p_B^2\, p_A \ , \ P(CCB) = p_C^2\, p_B \ , \text{ etc.}$$

Sei X die Anzahl der Meteoriteneinschläge bis zum Totalausfall. Diese Variable hat die folgende Verteilung:

$$\begin{aligned} P(X=1) &= P(A) = p_A, \\ P(X=2) &= P(BA) + P(BC) + P(CA) + P(CB) = p_B(p_A + p_C) + p_C(p_A + p_B) \ , \\ P(X=3) &= P(BBA) + P(BBC) + P(CCA) + P(CCB) \\ &= p_B^2(p_A + p_C) + p_C^2(p_A + p_B) \ , \\ P(X=4) &= P(BBBA) + P(BBBC) + P(CCCA) + P(CCCB) \\ &= p_B^3(p_A + p_C) + p_C^3(p_A + p_B) \ , \\ &\ \vdots \\ P(X=k) &= P(\underbrace{B \ldots B}_{k-1} A) + P(\underbrace{B \ldots B}_{k-1} C) + P(\underbrace{C \ldots C}_{k-1} A) + P(\underbrace{C \ldots C}_{k-1} B) \\ &= p_B^{k-1}(p_A + p_C) + p_C^{k-1}(p_A + p_B) \ . \end{aligned}$$

Der Erwartungswert ist damit

$$
\begin{aligned}
\mathrm{E}(X) &= 1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3) + \ldots \\
&= p_A + \sum_{k=2}^{\infty} k\, p_B^{k-1}(p_A + p_C) + \sum_{k=2}^{\infty} k\, p_C^{k-1}(p_A + p_B) \\
&= p_A + \sum_{k=1}^{\infty} k\, p_B^{k-1}(p_A + p_C) \ - (p_A + p_C) \\
&\quad + \sum_{k=1}^{\infty} k p_C^{k-1}(p_A + p_B) \ - (p_A + p_B) \\
&= \underbrace{-p_A - p_B - p_C}_{-1} \ + \ (p_A + p_C) \frac{d}{dp_B} \sum_{k=0}^{\infty} p_B^k \ + \ (p_A + p_B) \frac{d}{dp_C} \sum_{k=0}^{\infty} p_C^k \\
&= -1 \ + \ (p_A + p_C) \frac{d}{dp_B} \frac{1}{1 - p_B} \ + \ (p_A + p_B) \frac{d}{dp_C} \frac{1}{1 - p_C} \\
&= -1 \ + \ \frac{p_A + p_C}{(1 - p_B)^2} \ + \ \frac{p_A + p_B}{(1 - p_C)^2} \ = \ \frac{1}{p_A + p_C} + \frac{1}{p_A + p_B} - 1.
\end{aligned}
$$

Mit der Zusatzannahme $p_A = p_B = p_C = 1/3$ (etwa: die Teile A,B und C sind gleich groß) folgt

$$
\mathrm{E}(X) = \frac{1}{\frac{2}{3}} + \frac{1}{\frac{2}{3}} - 1 = 2 \ .
$$

Aufgabe 51*: (Erwartungswerte. 20 Punkte)

Mein Lotto„system“ besteht darin, dass ich beim nächsten Mal so viele (unabhängige) Lottotips abgebe, wie die erste gezogene Lottozahl der letzten Ziehung angab. Nach wievielen Wochen darf ich das erste Mal mit „6 Richtigen“ rechnen? (Beachte Aufgabe 47.a).)

Musterlösung:

Das oft wiederholte Bernoulli-Experiment besteht aus den Teilprozessen:

1) Wähle $k \in \{1, \ldots, 49\}$,
2) Gib k Tippreihen ab.

Der „Erfolg“ besteht darin, mindestens ein Mal „6 Richtige“ zu haben. Mit der Formel der totalen Wahrscheinlichkeit gilt:

$$
P(\text{„Erfolg“}) = \sum_{k=1}^{49} P(\text{„mindestens ein Mal 6 Richtige“} \mid k \text{ Tippreihen}) \cdot \underbrace{P(k)}_{1/49} \ .
$$

Sei

$$
p_0 = P(\text{„6 Richtige bei einer Tippreihe“}) = \frac{1}{\binom{49}{6}} \ .
$$

Mit

$$
\begin{aligned}
& P(\text{„mindestens ein Mal 6 Richtige“} \mid k \text{ Tippreihen}) \\
&= 1 - P(\text{„keine 6 Richtige“} \mid k \text{ Tippreihen}) \\
&= \boxed{1 - (1 - p_0)^k} \ = \ 1 - (1 - kp_0 + \tfrac{k(k-1)}{2}\, p_0^2 + \ldots) \\
&= k\, p_0 - \tfrac{k(k-1)}{2}\, p_0^2 + \ldots \ \approx \ \boxed{k\, p_0}
\end{aligned}
$$

folgt für das zusammengesetzte Bernoulli-Experiment:

$$P(\text{„Erfolg“}) \;=\; \sum_{k=1}^{49} \Big(1-(1-p_0)^k\Big)\cdot\frac{1}{49} \;\approx\; \sum_{k=1}^{49} \frac{k\,p_0}{49} \;=\; 25\cdot p_0 \ .$$

Mit Aufgabe 47.a) ist der Erwartungswert der Anzahl der Versuche bis zum ersten Erfolg gegeben durch

$$\frac{1}{P(\text{„Erfolg“})} \approx \frac{1}{25\cdot p_0} = \frac{1}{25}\cdot\binom{49}{6} = \frac{13\,983\,816}{25} \approx 559\,353 \ .$$

Aufgabe 52*: (Erwartungswerte. 20 Punkte)

Der Anteil q einer Personengruppe hat eine Krankheit, die durch eine (teure) Blutuntersuchung entdeckt werden kann. Der Anteil $p = 1-q$ hat keine Erreger im Blut. Es sollen alle kranken Personen identifiziert werden. Vergleiche zwei Untersuchungsmethoden:

(1) *Einzelprüfung:* Jeder wird einzeln untersucht. Man braucht einen Test pro Person.

(2) *Gruppenprüfung:* Das Blut von jeweils r Personen wird gemischt und untersucht. Wenn der Test positiv ist, werden alle Personen der Gruppe nochmal einzeln geprüft.

Vergleiche die mittleren Kosten der einzelnen Methoden in Abhängigkeit von p und r. Bestimme für $p = 0.99$ den kostengünstigsten Wert von r (z.B. per Wertetabelle der Kosten für diverse r). (Zur Kontrolle: die Ersparnis von (2) gegenüber (1) liegt bei etwa 80%.)

Musterlösung:

Die Personengruppe bestehe aus N Personen. Bei der Einzelprüfung sind N Bluttests durchzuführen. Zur Gruppenuntersuchung teile die Menge von N Personen in $n = N/r$ gleichgroße Gruppen mit jeweils r Personen auf (es sei N ein ganzzahliges Vielfaches von r). Die Untersuchung einer Gruppe entspricht einer r-fachen Wiederholung des Bernoulli-Experiments „wähle eine einzelne Person und stelle fest, ob sie gesund ist“, also

$$\tilde p = P(\text{„alle Personen der Gruppe sind gesund“}) = \binom{r}{r} p^r\, q^0 = p^r \ .$$

Die Kosten einer Einzeluntersuchung sei 1. Sei K die Zufallsvariable „Gesamtkosten“.

Bestimme $\mathrm{E}(K)$, einfache Lösung (braucht Satz 2.33.1 der Vorlesung):

Sei $K = K_1 + \cdots + K_n$, wobei die Zufallsvariablen K_i die Kosten der Untersuchung der i-ten Gruppe sind. Ist die Gruppe gesund, fällt nur ein Bluttest an. Ist die Gruppe krank (d.h., fällt der Test positiv aus, weil mindestens eine Person der Gruppe krank ist), sind r weitere Tests für alle Mitglieder der Gruppe fällig. Damit folgt für $r > 1$:

$$\begin{aligned}\mathrm{E}(K_i) \;&=\; 1 \cdot P(\text{„die } i\text{-te Gruppe ist gesund“})\\ &\quad+ (1+r)\cdot P(\text{„die } i\text{-te Gruppe ist krank“})\\ &= p^r + (1+r)\,(1-p^r) \;=\; 1 + r - r\,p^r.\end{aligned}$$

Mit $\mathrm{E}(K) = \mathrm{E}(K_1) + \cdots + \mathrm{E}(K_n)$ (nach Satz 2.33.1 der Vorlesung) folgt

$$\mathrm{E}(K) = n\,(1 + r - r\,p^r) = N\,\frac{1 + r - r\,p^r}{r} = N\,\left(\frac{1}{r} + 1 - p^r\right) \ .$$

Bestimme $\mathrm{E}(\mathbf{K})$**, komplexere Lösung:**
Sei X die Anzahl der Gruppen, die sich bei Untersuchung von insgesamt n Gruppen als gesund herausstellt. Betrachtet man die Untersuchung einer Gruppe als Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $\tilde{p} = p^r$ für „Erfolg“ = „die Gruppe ist gesund“ = „alle Personen der Gruppe sind gesund“, so beschreibt X die Anzahl der Erfolge bei n-facher Wiederholung. Per Binomial-Verteilung mit Erfolgswahrscheinlichkeit $\tilde{p}$ gilt

$$P(\text{„genau } k \text{ Gruppen sind gesund“}) = P(X = k) = \binom{n}{k} \tilde{p}^k \, (1 - \tilde{p})^{n-k}.$$

Die Kosten im Falle $X = k$ sind $n + (n - k)\,r$ (n Gruppen werden untersucht, bei $n - k$ kranken Gruppen sind jeweils r Personen durch einen weiteren Test einzeln zu prüfen). Damit folgt

$$\begin{aligned} \mathrm{E}(K) &= \sum_{k=0}^{n} (n + (n-k)\,r)\, P(X = r) = \sum_{k=0}^{n} (n + (n-k)\,r) \binom{n}{k} \tilde{p}^k \, (1 - \tilde{p})^{n-k} \\ &= n + n\,r - r\,\mathrm{E}(X) = n\,(1 + r - r\,\tilde{p}) = n\,(1 + r - r\,p^r) \\ &= N\,\frac{1 + r - r\,p^r}{r} = N\,\left(\frac{1}{r} + 1 - p^r\right)\,. \end{aligned}$$

Optimierung der Kosten: Bei gegebenem p und fixiertem N sollen die mittleren Kosten durch geeignete Wahl von r minimiert werden. Leider ist das Minimum nicht durch eine geschlossene Formel ausdrückbar. Um abzuschätzen, in welchem Bereich für r wir suchen sollten, betrachten wir zur Vorbereitung zunächst eine Approximation für den Fall einer seltenen Krankeit, d.h., p liegt dicht bei 1. Setze dazu $p = 1 - \epsilon$ und entwickle $\mathrm{E}(K)$ um $\epsilon = 0$. Mit $p^r = (1 - \epsilon)^r = 1 - r\,\epsilon + O(\epsilon^2)$ ergibt sich

$$\mathrm{E}(K) = N\,\left(\frac{1}{r} + r\,\epsilon\right) + O(\epsilon^2)\,.$$

Wir minimieren den führenden Term:

$$\frac{d}{dr}\left(\frac{1}{r} + r\,\epsilon\right) = 0 \quad \Rightarrow \quad r = \frac{1}{\sqrt{\epsilon}} = \frac{1}{\sqrt{1-p}}\,.$$

Für $p = 0.99$, also $\epsilon = 1/100$, ergibt sich approximativ als optimale Gruppengröße $r = 10$. Den genauen Wert ermitteln wir aus einer Wertetabelle:

r	4	6	8	9	10	11	12	14	16	18
$1 - p^r + 1/r$	0.289	0.225	0.202	0.197	0.19561	0.19557	0.1969	0.203	0.211	0.221

Das Kostenminimum für $p = 0.99$ stellt sich für $r = 11$ ein. Es beträg etwa 19.6% der Kosten der Einzelprüfungen.