Universität Bremen
Fachbereich Mathematik

Stochastik
Sommersemester 2007

Lösungsskizzen zur 10. Übung

Aufgabe 3

Bei einer Bernoulli-Kette gibt es n Versuche, und bei jedem Versuch haben wir die Wahrscheinlichkeit p für einen Treffer und die Wahrscheinlichkeit $(1-p)$ für keinen Treffer. Es kann also nach n Versuchen $0, 1, 2, \ldots, n$ Treffer geben. Bei einer Bernoulli-Kette haben wir eine Binomialverteilung. In diesem Fall ist also:

$$\begin{aligned} \mu &= n \cdot p \\ \mathrm{Var(X)} &= n \cdot p \cdot (1-p) \\ \sigma &= \sqrt{\mathrm{Var(X)}}, \end{aligned}$$

wobei μ der Erwartungswert $E(X)$ ist, $\mathrm{Var(X)}$ die Varianz und σ die Streuung (oder auch Standardabweichung).

a) Für $n = 24$ und $p = 0,4$ ergeben sich so $\mu = 9,6$ und $\sigma = 2,4$. Das zu betrachtende Intervall $[\mu - \sigma; \mu + \sigma]$ ist daher

$$[7,2\,;\,12].$$

Nun ist die Frage, welche Treffer innerhalb dieses Intervalls liegen. Alle möglichen Treffer sind: $0, 1, 2, 3, \ldots, 24$, also insbesondere nur ganze Zahlen. Wenn wir uns das Intervall auf dem Zahlenstrahl vorstellen, wird deutlich, dass nur die Treffer

$$8, 9, 10, 11, 12$$

in dem Intervall liegen. Die Wahrscheilichkeit dafür, dass die Trefferzahl im Intervall liegt, ist daher

$$\sum_{k=8}^{12} P(k; 24; 0,4).$$

Die Binomialverteilung $P(k; n; p)$ ist dabei

$$P(k; n; p) := \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}.$$

Die Wahrscheinlichkeitswerte können mit dem Tabellenkalkulationsprogramm der Veranstaltungshomepage abgelesen werden. Es ergibt sich so:

$$\sum_{k=8}^{12} P(k; 24; 0,4) = \sum_{k=1}^{12} P(k; 24; 0,4) - \sum_{k=1}^{7} P(k; 24; 0,4) \approx 0,694.$$

Achtung: Man darf das Intervall $[7,2\,;\,12]$ auf keinen Fall runden! Warum auch?! Es ist die Frage, welche Treffer genau innerhalb dieses Intervalls liegen. Es gibt keinen Grund, das Intervall zu vergrößern oder zu verkleinern und dann zu gucken, welche Treffer drinnen liegen.

b) Für $n = 48$ und $p = 0,4$ ergeben sich so $\mu = 19,2$ und $\sigma \approx 3,39$. Die Wahrscheinlichkeit dafür, dass die Trefferzahl im Intervall $[\mu - \sigma; \mu + \sigma]$ liegt, ist gerundet $0,697$.

c) Für $n = 96$ und $p = 0,4$ ergeben sich so $\mu = 38,4$ und $\sigma \approx 4,8$. Die Wahrscheinlichkeit dafür, dass die Trefferzahl im Intervall $[\mu - \sigma; \mu + \sigma]$ liegt, ist gerundet $0,7023$.

Aufgabe 4

Bei A errechnen wir den Erwartungswert des Verlustes durch

$$E(X) = 1\,000\,000 \cdot 0,1 = 100\,000.$$

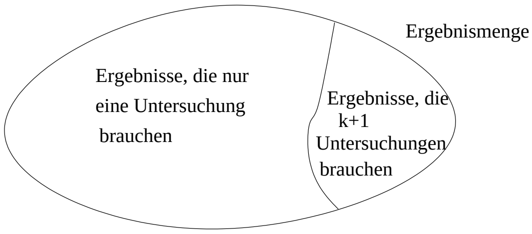
Bei B rechnen wir

$$\begin{aligned} E(X) &= 200\,000 \cdot 0,1 + 200\,000 \cdot 0,1 + 200\,000 \cdot 0,1 + 200\,000 \cdot 0,1 + 200\,000 \cdot 0,1 \\ &= 0,1 \cdot (5 \cdot 200\,000) = 0,1 \cdot 1\,000\,000 = 100\,000. \end{aligned}$$

Am Erwartungswert können wir also nicht ablesen, welche Strategie günstiger ist.

Aufgabe 5

Eine einzelne Probe ist mit einer Wahrscheinlichkeit von $0,01$ krank. Es werden jedoch nicht einzelne Proben gestestet, sondern Mischungen von k ($k \geq 1$) Proben. Sehen wir uns als erstes die Ergebnismenge Ω aller Mischungen an:



Da wir einen *Erwartungswert für die zu erwartenden Untersuchungen* erhalten wollen, wählen wir die Zufallsvariable $X : \Omega \to \{x_1, x_2\}$, wobei

$$x_1 = 1 \text{ und } x_2 = k + 1.$$

Wir erhalten für den Erwartungswert nun allgemein

$$E(X) = \sum_{i=1}^{2} x_i \cdot P(X = x_i) \quad = \quad 1 \cdot P(X = 1) + (k + 1) \cdot P(X = k + 1).$$

(Wir erinnern uns daran, dass $P(X = x_i)$ nur eine Abkürzung für

$$P(\{\omega \in \Omega \mid X(\omega) = x_i\})$$

ist, also für die Wahrscheinlichkeit, dass diejenigen Ergebnisse auftreten, die von der Zufallsvariable X auf den Wert x_i abgebildet werden. Zum Beispiel ist $P(X = k+1)$ die Wahrscheinlichkeit dafür, dass die Ergebnisse auftreten, die von der Zufallsvariable auf $k+1$ abbildet werden. Anders ausgedrückt, ist das die Wahrscheinlichkeit dafür, dass eine Mischung von k Proben $(k+1)$-mal getestet werden muss, also die Wahrscheinlichkeit dafür, dass eine Mischung als krank getestet wird.)

k=5: Im Fall $k = 5$ müssen wir uns nun überlegen, was $P(X = 1)$ und $P(X = 6)$ sind. Die Wahrscheinlichkeit dafür, dass eine Mischung, die aus fünf Proben besteht, gesund getestet wird, ist

$$0{,}99^5,$$

da für jede einzelne Probe die Wahrscheinlichkeit $0{,}99$ ist. Also ist $P(X = 1) = 0{,}99^5$. Umgekehrt ist $P(X = 6)$ gerade die Gegenwahrscheinlichkeit, also

$$P(X = 6) = 1 - 0{,}99^5.$$

Darum ist

$$E(X) = 1 \cdot P(X = 1) + 6 \cdot P(X = 6) = 1 \cdot 0{,}99^5 + 6 \cdot (1 - 0{,}99^5) \approx 1{,}245.$$

k=10: Im Fall $k = 10$ ist mit den gleichen Begründungen

$$E(X) = 1 \cdot P(X = 1) + 11 \cdot P(X = 11) = 1 \cdot 0{,}99^{10} + 11 \cdot (1 - 0{,}99^{10}) \approx 1{,}956.$$

k=20: Im Fall $k = 10$ ist

$$E(X) = 1 \cdot P(X = 1) + 21 \cdot P(X = 21) = 1 \cdot 0{,}99^{20} + 21 \cdot (1 - 0{,}99^{20}) \approx 4{,}642.$$

Um herauszubekommen, welche Anzahl von Proben in der Mischung optimal ist, sehen wir uns das Verhältnis des Erwartungswerts zur Anzahl der Proben an, also

$$\frac{E(X)}{k}.$$

Durch Differenzialrechnung (den Bruch nach k ableiten, Null setzen, usw.) oder durch Ausprobieren kommt man dazu, dass eine Mischung von 11 Proben optimal ist.

Aufgabe 6

Die Zufallsvariable ist gegeben durch $X : \Omega \to \{x_1, x_2, x_3, x_4\}$. Wir wissen aus dem Aufgabentext, dass die Werte x_1, x_2, x_3, x_4 mit der gleichen Wahrscheinlichkeite auftreten. Da die Summe aller Wahrscheinlichkeiten gleich Eins sein muss, ist daher für alle $i \in \{1, 2, 3, 4\}$

$$P(X = x_i) = \tfrac{1}{4}.$$

Die Varianz ist definiert als

$$v(m) := \sum_{i=1}^{4}((x_i - m)^2 \cdot P(X = x_i)).$$

Also ist

$$v(m) = \sum_{i=1}^{4}((x_i - m)^2 \cdot \frac{1}{4}).$$

Um zu erfahren, wann die Varianz minimal wird, rechnen wir die erste und zweite Ableitung von $v(m)$ aus. Mit der Kettenregel bekommen wir

$$v'(m) = \sum_{i=1}^{4}(-\frac{1}{2}(x_i - m)),$$

denn für eine Funktion $f(m) = (x_1 - m)^2 \cdot \frac{1}{4}$ ist die erste Ableitung

$$f'(m) = 2 \cdot (x_1 - m) \cdot (-1) \cdot \frac{1}{4} = -\frac{1}{2}(x_1 - m).$$

Achtung: Man leitet immer nach *der* Variable ab, die „in Klammern" steht, in unserem Fall m. Alle anderen Variablen, also auch $x_1, \ldots, x_4$ sind beim Ableiten nichts anderes als Konstanten. Wenn also $g(m) = 2 \cdot m + x$, dann ist $g'(m) = 2$. Die Variable x ist beim Ableiten nach m nur eine Konstante, die wegfällt.

Nun setzen wir $v'(m)$ gleich Null, um den Extremalpunkt herauszufinden. Wir erhalten:

$$\begin{aligned}
& \sum_{i=1}^{4}(-\frac{1}{2}(x_i - m)) = 0 \\
\Leftrightarrow \quad & -\frac{1}{2} \cdot (x_1 + x_2 + x_3 + x_4) + 4 \cdot \frac{1}{2} \cdot m = 0 \\
\Leftrightarrow \quad & 4m = x_1 + x_2 + x_3 + x_4 \\
\Leftrightarrow \quad & m = \frac{x_1 + x_2 + x_3 + x_4}{4} = \frac{\sum_{i=1}^{4} x_i}{4}.
\end{aligned}$$

Für

$$m^* := \frac{\sum_{i=1}^{4} x_i}{4}$$

gibt es also einen Extremalpunkt. Wir prüfen nun, ob er ein Minimum oder ein Maximum ist. Dafür betrachten wir die zweite Ableitung $v''(m)$, also

$$v'(m) = \sum_{i=1}^{4}(-\frac{1}{2}) \cdot (-1) = 2.$$

Es ist also auch $v'(m^*) = 2$ und daher haben wir an der Stelle m^* tatsächlich ein Minimum.