Einführung in die Wahrscheinlichkeitstheorie und Statistik
Wintersemester 2011/12

Lösungsvorschlag zur Probeklausur

Aufgabe 1.
Es seien X, Y unabhängige Zufallsvariablen mit Werten in $\mathbb{N} = \{1, 2, \ldots\}$ und

$$\mathbf{P}(X = n) = \mathbf{P}(Y = n) = \frac{1}{2^n} \quad (n = 1, 2, \ldots).$$

Berechnen Sie die folgenden Wahrscheinlichkeiten und vereinfachen Sie sie soweit wie möglich:

(a) $\mathbf{P}(\min(X, Y) \leq i)$,

(b) $\mathbf{P}(X = Y)$,

(c) $\mathbf{P}(X \text{ teilt } Y)$.

Lösung:

(a)

$$\begin{aligned}
\mathbf{P}(\min(X, Y) \leq i) &= 1 - \mathbf{P}(\min(X, Y) > i) \\
&= 1 - \mathbf{P}(X > i, Y > i) \\
&= 1 - \mathbf{P}(X > i)\mathbf{P}(Y > i) \\
&= 1 - \left(\sum_{n=i+1}^{\infty} \frac{1}{2^n} \right)^2 \\
&= 1 - \left(\frac{1}{2^i} \underbrace{\sum_{n=1}^{\infty} \frac{1}{2^n}}_{=1} \right)^2 \\
&= 1 - \left(\frac{1}{2^i} \right)^2 .
\end{aligned}$$

(b)

$$\begin{aligned}
\mathbf{P}(X=Y) &= \sum_{n=1}^{\infty} \mathbf{P}(X=k=Y) \\
&= \sum_{n=1}^{\infty} \mathbf{P}(X=k, Y=k) \\
&= \sum_{n=1}^{\infty} \mathbf{P}(X=k)\mathbf{P}(Y=k) \\
&= \sum_{n=1}^{\infty} \frac{1}{4^k} \\
&= \frac{1}{1-\frac{1}{4}} - 1 = \frac{1}{3}.
\end{aligned}$$

(c)

$$\begin{aligned}
\mathbf{P}(X \text{ teilt } Y) &= \sum_{n=1}^{\infty} \mathbf{P}(X=n, Y=k\cdot n \text{ für ein } k \in \mathbb{N}) \\
&= \sum_{n=1}^{\infty} \mathbf{P}(X=n) \sum_{k=1}^{\infty} \mathbf{P}(Y=k\cdot n) \\
&= \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \frac{1}{2^{n(k+1)}} \\
&= \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} \left(\frac{1}{2^{k+1}}\right)^n \\
&= \sum_{k=1}^{\infty} \frac{1}{2^{k+1}-1}.
\end{aligned}$$

Aufgabe 2.

Zu Beginn des 19. Jahrhunderts lag die Jahresmitteltemperatur in Karlsruhe bei 10.4 Grad Celsius. Die angehende Klimaforscherin Miri Swarm behauptet, dass sich die mittlere Jahrestemperatur in den letzten zwei Jahrhunderten erhöht hat. Dabei stützt sie sich auf die gemessenen Jahresmitteltemperaturen der Jahre 2000 bis 2008 (jeweils in Grad Celsius)

$$12.2 \quad 11.3 \quad 11.7 \quad 11.8 \quad 11.1 \quad 11.2 \quad 11.6 \quad 11.9 \quad 11.5.$$

(a) Stimmen Sie Miri Swarm zu? Formulieren Sie dieses Testproblem mathematisch und führen Sie dazu einen geeigneten Test durch, bei dem Sie annehmen, dass die Jahresmitteltemperatur näherungsweise normalverteilt ist mit bekannter Varianz $\sigma^2 = 1$. Dabei soll die Wahrscheinlichkeit, der Klimaforscherin zuzustimmen, obwohl gar keine Temperaturerhöhung vorliegt, maximal 5% betragen.

(b) Für welchen falschen Parameterbereich können Sie aus dieser Form von Test ein 95% Konfidenzintervall bestimmen? Geben Sie dieses Intervall explizit an!

HINWEISE: $\bar{X}_9 \approx 11.59$, Quantile der Standardnormalverteilung: $q_{0,05} = -1.64$, $q_{0,01} = -2,33$.

<u>Lösung:</u>

(a) <u>Mathematische Formulierung des Testproblems:</u>

$X_1, \ldots, X_9$ sind $\mathcal{N}(\mu, 1)$ verteilt mit unbekanntem μ. Es soll die Hypothese $H_0 : \mu \leq 10,4 =: \mu_0$ (man stimmt Miri Swarm nicht zu) gegen die Alternative $H_1 : \mu > 10,4$ (man stimmt Miri Swarm zu) getestet werden. Dabei soll der Fehler 1. Art kleiner als das Niveau $\alpha = 0,05$ sein.

Der gleichmäßig beste Test für dieses Testproblem ist aus der Vorlesung bekannt und hat folgende Form (vgl. Skript Satz 8.20):

$$\phi^*(X_1, \ldots, X_9) = \begin{cases} 1 & , \bar{X}_n > c^* \\ 0 & , \bar{X}_n \leq c^* \end{cases},$$

wobei $c^* = \mu_0 + \frac{\sigma}{\sqrt{n}} \underbrace{u_{1-\alpha}}_{-u_\alpha} = 10,4 + \frac{1}{3} \cdot 1,64 \leq 11$. Da $\bar{X}_9 \approx 11.59$ entscheidet man sich für die Alternative und stimmt der Klimaforscherin zu.

(b) Der Ablehnungsbereich des Tests aus (a) hat die Form $K(\mu_0) = (\mu_0, \infty)$. Demnach ist der falsche Parameterbereich (vgl. Skript Bemerkung 9.5) $\bar{K}(\mu) = \{\mu' | \mu > \mu'\} = (-\infty, \mu)$.

Das gesuchte Konfidenzintervall $S(X_1, \ldots, X_9)$ erhält man wie folgt:

$$\begin{aligned} S(X_1, \ldots, X_9) &= \left\{ \mu \in \mathbb{R} : \bar{X}_9 \leq \mu + \frac{1}{3} \cdot 1,64 \right\} \\ &= [\bar{X}_9 - \frac{1}{3} \cdot 1,64, \infty) \\ &= [11.04, \infty). \end{aligned}$$

Aufgabe 3.

Es seien $X_1, \ldots, X_n$ i.i.d. mit unbekanntem Erwartungswert μ und endlicher Varianz. Zur Schätzung von μ schlagen wir den folgenden Schätzer vor

$$\hat{\mu}_n = \frac{1}{n-2} \sum_{i=1}^{n} X_i.$$

(a) Berechnen Sie $\mathbf{E}\hat{\mu}_n$ sowie $\lim\limits_{n\to\infty} \mathbf{E}\hat{\mu}_n$!

(b) Berechnen Sie $\operatorname{var}(\hat{\mu}_n)$ sowie $\lim\limits_{n\to\infty} \operatorname{var}(\hat{\mu}_n)$!

(c) Ist $\hat{\mu}_n$ ein konsistenter Schätzer für μ?

<ins>Lösung:</ins>

(a) Es gilt $\mathbf{E}\hat{\mu}_n = \frac{1}{n-2} \sum\limits_{i=1}^{n} \mathbf{E}X_i = \frac{n}{n-2} \cdot \mu$ und $\lim\limits_{n\to\infty} \mathbf{E}\hat{\mu}_n = \mu$.

(b) Es gilt $\operatorname{var}(\hat{\mu}_n) = \frac{1}{(n-2)^2} \sum\limits_{i=1}^{n} \operatorname{var}(X_i) = \frac{n}{(n-2)^2} \cdot \operatorname{var}(X_1)$ und

$\lim\limits_{n\to\infty} \operatorname{var}(\hat{\mu}_n) = \lim\limits_{n\to\infty} \frac{n}{(n-2)^2} \cdot \operatorname{var}(X_1) = 0.$

(c) Schreibe $\hat{\mu}_n = \frac{n}{n-2} \cdot \frac{1}{n} \sum\limits_{i=1}^{n} X_i$. Es gilt $\frac{n}{n-2} \xrightarrow{\mathbf{P}} 1$ (Konvergenz gilt deterministisch) und $\frac{1}{n} \sum\limits_{i=1}^{n} X_i \xrightarrow{\mathbf{P}} \mu$ nach dem schwachen Gesetz der großen Zahlen. Damit konvergiert dann auch $\hat{\mu}_n = \frac{n}{n-2} \cdot \frac{1}{n} \sum\limits_{i=1}^{n} X_i$ in Wahrscheinlichkeit gegen μ (was man sich zum Beispiel anhand Proposition 14.9 überlegen kann) und die Konsistenz des Schätzers $\hat{\mu}_n$ ist gezeigt.

Aufgabe 4.

Seien X und Y gemeinsam normalverteilt mit $\mathbf{E}X = \mathbf{E}Y = 0$, $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$ und $\mathrm{Kov}(X,Y) = \frac{1}{2}$.

(a) Geben Sie die gemeinsame Dichte von X und Y an.

(b) Berechnen Sie $\mathbf{P}(X > \frac{1}{2}|Y = 1)$.

<u>Lösung:</u>

(a) Wir haben

$$f_{X,Y}(x,y) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \cdot \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})\right\}.$$

Erwartungswertvektor: $\underline{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

Kovarianzmatrix: $\Sigma = \begin{pmatrix} 1 & 0,5 \\ 0,5 & 1 \end{pmatrix} \Leftarrow \Sigma^{-1} = \frac{4}{3}\begin{pmatrix} 1 & -0,5 \\ -0,5 & 1 \end{pmatrix}$

Damit ergibt sich:

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{3}\pi} \cdot \exp\left\{-\frac{2}{3}(x^2 - xy + y^2)\right\}.$$

(b) Es ist

$$\begin{aligned}
\mathbf{P}(X > \frac{1}{2}|Y = 1) &= \int_{\frac{1}{2}}^{\infty} f_{X|Y=1}(x)dx \\
&= \frac{1}{f_Y(1)} \int_{\frac{1}{2}}^{\infty} f_{XY}(x,1)dx \\
&= \sqrt{2\pi} \cdot e^{0,5} \cdot \frac{1}{\sqrt{3}\pi} \int_{\frac{1}{2}}^{\infty} \exp\left\{-\frac{2}{3}(x^2 - x + 1)\right\} dx \\
&= \frac{\sqrt{2\pi}}{\sqrt{3}\pi} \cdot e^{0,5} \int_{\frac{1}{2}}^{\infty} \exp\left\{-\frac{2}{3}(x-0,5)^2 - 0,5\right\} dx \\
&= \frac{\sqrt{2\pi}}{\sqrt{3}\pi} \cdot \underbrace{\int_{\frac{1}{2}}^{\infty} \exp\left\{-\frac{2}{3}(x-0,5)^2\right\} dx}_{=\frac{\sqrt{6\pi}}{4}} \\
&= 0,5.
\end{aligned}$$

Aufgabe 5.

Es seien $X_1, \ldots, X_n \overset{iid}{\sim} NB(r,p)$ verteilt mit bekanntem positiven Parameter $r \in \mathbb{N}$ und unbekanntem Parameter $p \in (0,1)$. Bestimmen Sie den Maximum-Likelihood-Schätzer $\hat{p}_n$ für p (es ist auch zu zeigen, dass $\hat{p}_n$ die Likelihood <u>maximiert</u>) und berechnen Sie die Fisher-Information $I(p)$.

Hinweis: Die negative Binomialverteilung $NB(r,p)$ hat die Zähldichte

$$\mathbf{P}(X = k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r \text{ für alle } k \geq r.$$

<u>Lösung:</u>

Die Likelihood ist gegeben durch: $\mathbf{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \binom{x_i-1}{r-1}(1-p)^{x_i-r}p^r$.

Wir benutzen die Log-Likelihood für die Berechnung des MLE:

$$\begin{aligned} l(p) :&= \log\left(\prod_{i=1}^{n}\binom{x_i-1}{r-1}(1-p)^{x_i-r}p^r\right) \\ &= \sum_{i=1}^{n}\log\binom{x_i-1}{r-1} + \sum_{i=1}^{n}(x_i-r)\log(1-p) + \sum_{i=1}^{n} r\log(p). \end{aligned}$$

$$\frac{\partial l(p)}{\partial p} = -\frac{\sum_{i=1}^{n}(x_i-r)}{1-p} + \frac{\sum_{i=1}^{n} r}{p}$$

$$\begin{aligned} \frac{\partial l(p)}{\partial p} \overset{!}{=} 0 &\Leftrightarrow \frac{-\sum_{i=1}^{n} x_i p + nrp + nr - nrp}{p(1-p)} = 0 \\ &\Leftrightarrow \left(\sum_{i=1}^{n} x_i\right) p = nr \\ &\Leftrightarrow p = \frac{nr}{\sum_{i=1}^{n} x_i} =: \hat{p}_n. \end{aligned}$$

$$\frac{\partial^2 l(p)}{\partial^2 p} = -\frac{\sum_{i=1}^{n}(x_i-r)}{(1-p)^2} - \frac{\sum_{i=1}^{n} r}{p^2} =: h(p)$$

$h(\hat{p}_n) < 0$, also maximiert $\hat{p}_n$ die Likelihood.

Fisher-Information:

$$\begin{aligned}
\mathbf{E}\left(\frac{\partial l(p)}{\partial p}\right)^2 &= \mathbf{E}\left(\frac{nr - \sum\limits_{i=1}^{n} x_i p}{p(1-p)}\right)^2 \\
&= \frac{n^2r^2 - 2n^2rp\mathbf{E}X_1 + np^2\mathbf{E}X_1^2 + p^2n(n-1)(\mathbf{E}X_1)^2}{p^2(1-p)^2} \\
&= \ldots \\
&= \frac{nr(1-p)}{p^2(1-p)^2} \\
&= \frac{nr}{p^2(1-p)}.
\end{aligned}$$

[Beachte: Für eine $NB(r,p)$-verteilte Zufallsvariable X gilt: $\mathbf{E}X = \frac{r}{p}$ und $\mathbf{E}X^2 = \frac{r(r+1-p)}{p^2}$. In einer Klausur hätte man diese beiden Werte wohl angegeben, um die Aufgabe nicht zu lang werden zu lassen.]

Aufgabe 6.

Es sei $p \in (0,1)$ bekannt und $X \sim \mathcal{B}(N,p)$, wobei N *a priori* gemäß einer Poisson-Verteilung mit Parameter $\lambda > 0$ verteilt sei. Berechnen Sie die Posteriori-Verteilung von N und den zugehörigen Posteriori-Erwartungswert.

Lösung:

Posteriori-Verteilung:

$$\begin{aligned}
[N|X] &\propto [X|N] \cdot [N] \\
&= \binom{N}{X} p^X (1-p)^{N-X} \cdot \frac{\lambda^N}{N!} \\
&\propto \frac{((1-p)\lambda)^{N-X}}{(N-X)!}.
\end{aligned}$$

Dies entspricht dem Kern einer „verschobenen" Poissonverteilung zum Parameter $(1-p)\lambda$ mit Zähldichte:

$$p(N=k|X=x) = e^{-\lambda(1-p)} \cdot \frac{((1-p)\lambda)^{k-X}}{(k-X)!}, \quad k \in \{x,,x+1,\dots\}.$$

Posteriori-Erwartungswert:

$$\begin{aligned}
\mathbf{E}[N|X=x] &= \sum_{k=x}^{\infty} p(N=k|X=x) \cdot k \\
&= e^{-(1-p)\lambda} \sum_{k=x}^{\infty} \frac{((1-p)\lambda)^{k-x}}{(k-x)!} \cdot k \\
&= e^{-(1-p)\lambda} \sum_{k=0}^{\infty} \frac{((1-p)\lambda)^{k}}{(k)!} \cdot (k+x) \\
&= x + (1-p)\lambda.
\end{aligned}$$

[Beachte: Dieser Erwartungswert wurde auch schon auf Blatt 11, Aufgabe 2 berechnet.]

Sollte jemand Fehler bzw. Unklarheiten in diesem Lösungsvorschlag bemerken, so kann er mir gerne eine Mail zukommen lassen: martin.kroll@uni-heidelberg.de.