Prof. Dr. M. Schweizer | ETH Zürich | Sommer 2011

Wahrscheinlichkeit und Statistik
BSc D-INFK

1. **a)** (iii) **b)** (ii) **c)** (i) **d)** (ii) **e)** (ii) **f)** (iii) **g)** (ii) **h)** (i) **i)** (ii) **j)** (i)

2. Für ein heruntergeladenes Dokument benützen wir folgende Notation:

$$\begin{aligned} A &= \{\text{die Antivirensoftware } A \text{ löst für das Dokument Alarm aus}\}, \\ B &= \{\text{die Antivirensoftware } B \text{ löst für das Dokument Alarm aus}\}, \\ V &= \{\text{das Dokument enthält Viren der Art } V\}, \\ W &= \{\text{das Dokument enthält Viren der Art } W\}. \end{aligned}$$

a) (1 Punkt) Gesucht ist $P[B]$. Mit der Formel für die totale Wahrscheinlichkeit erhalten wir

$$\begin{aligned} P[B] &= P[B|V]P[V] + P[B|W]P[W] + P[B|(V \cup W)^c]P[(V \cup W)^c] \\ &= 0.9 * 0.01 * 0.9 + 0.8 * 0.01 * 0.1 + 0.01 * 0.99 = 0.0188. \end{aligned}$$

b) (2 Punkte) Gesucht ist die Wahrscheinlichkeit $P[V \cup W|A^c]$. Mit dem Satz von Bayes erhalten wir

$$\begin{aligned} P[V \cup W|A^c] = P[V|A^c] + P[W|A^c] &= \frac{P[A^c|V]P[V] + P[A^c|W]P[W]}{1 - P[A]} \\ &= \frac{0.2 * 0.01 * 0.9 + 0.3 * 0.01 * 0.1}{1 - 0.00889} = \frac{210}{99111}, \end{aligned}$$

wobei

$$\begin{aligned} P[A] &= P[A|V]P[V] + P[A|W]P[W] + P[A|(V \cup W)^c]P[(V \cup W)^c] \\ &= 0.8 * 0.01 * 0.9 + 0.7 * 0.01 * 0.1 + 0.001 * 0.99 = 0.00889. \end{aligned}$$

Bitte wenden!

c) (i) (2 Punkte) Die erwarteten Kosten pro heruntergeladenem Dokument für das Produkt A sind

$$
\begin{aligned}
&100 * P[W \cap A^c] + 10 * P[V \cap A^c] + 1 * P[(V \cup W)^c \cap A] \\
&\quad = 100 * P[A^c|W]P[W] + 10 * P[A^c|V]P[V] + 1 * P[A|(V \cup W)^c]P[(V \cup W)^c] \\
&\quad = 100 * 0.3 * 0.01 * 0.1 + 10 * 0.2 * 0.01 * 0.9 + 1 * 0.001 * 0.99 \\
&\quad = 0.04899.
\end{aligned}
$$

Die erwarteten Kosten ohne Antivirensoftware betragen

$$100 * P[W] + 10 * P[V] = 100 * 0.01 * 0.1 + 10 * 0.01 * 0.9 = 0.19.$$

(ii) (3 Punkte) Die erwarteten Kosten pro heruntergeladenem Dokument für das Produkt B betragen

$$
\begin{aligned}
&100 * P[W \cap B^c] + 10 * P[V \cap B^c] + 1 * P[(V \cup W)^c \cap B] \\
&\quad = 100 * P[B^c|W]P[W] + 10 * P[B^c|V]P[V] + 1 * P[B|(V \cup W)^c]P[(V \cup W)^c] \\
&\quad = 100 * 0.2 * 0.01 * 0.1 + 10 * 0.1 * 0.01 * 0.9 + 1 * 0.01 * 0.99 \\
&\quad = 0.0389.
\end{aligned}
$$

Aus $n * 0.0389 + 500 \leq n * 0.04899$ folgt $n \geq 49554.01388$. Also ist es ab 49555 Dokumenten sinnvoll, B statt A zu wählen.

Falls mit dem Wert 0.05 statt 0.04899 gerechnet wurde, dann gilt $n \geq 45045.045$. Also ist es ab 45045 Dokumenten sinnvoll, B statt A zu wählen.

d) (2 Punkte) Die Anzahl der Warnungen X ist binomial verteilt mit $n = 1000$ und $p = P[A]$. Deshalb gilt

$$
\begin{aligned}
P[X < 10] &= \sum_{k=0}^{9} \binom{1000}{k} p^k (1-p)^{(1000-k)} \\
&= \sum_{k=0}^{9} \binom{1000}{k} 0.00889^k (0.99111)^{(1000-k)}.
\end{aligned}
$$

3. **a)** (1 Punkt) Die Verteilungsfunktion ist das Integral der Dichtefunktion, also

$$F_Y(t) = \int_0^t \alpha s^{\alpha-1} e^{-s^\alpha} ds = e^{-s^\alpha}\Big|_t^0 = 1 - e^{-t^\alpha}.$$

Siehe nächstes Blatt!

b) (1 Punkte) Wenn **a**) gelöst wurde: Da $F_Y(t) = 1 - e^{-t^\alpha}$, folgt $1 - F_Y(t) = e^{-t^\alpha}$. Sonst,

$$1 - F_Y(t) = P[Y > t] = \int_t^\infty f_Y(s)ds = \int_t^\infty \alpha s^{\alpha-1} e^{-s^\alpha} ds = e^{-s^\alpha}\Big|_\infty^t = e^{-t^\alpha}.$$

Wir rechnen

$$\lambda(t) = \alpha t^{\alpha-1} e^{-t^\alpha} \frac{1}{e^{-t^\alpha}} = \alpha t^{\alpha-1}.$$

Daraus folgt, dass $\lambda(t) = 2t$ ist genau dann, wenn $\alpha = 2$ ist.

c) (1 Punkt) Wenn **a**) gelöst wurde,

$$P[X > 5] = P[Y + 3 > 5] = P[Y > 2] = 1 - P[Y \leq 2] = e^{-4}.$$

Sonst,

$$P[X > 5] = P[Y > 2] = \int_2^\infty f_Y(s)ds = \int_2^\infty 2se^{-s^2} ds = e^{-s^2}\Big|_\infty^2 = e^{-4}.$$

d) (2 Punkte) Nur 10% aller Rücksendungen werden nicht angenommen, wenn 90% aller Rücksendungen akzeptiert werden. Es gilt also

$$P[X \leq t] = 0.9.$$

Wir haben $P[X \leq t] = 0.9$ genau dann, wenn $P[Y \leq t - 3] = 0.9$. Weiter berechnen wir die Inverse von F_Y

$$F_Y(t-3) = 0.9 \quad \Leftrightarrow \quad t = 3 + F_Y^{-1}(0.9).$$

Aber $F_Y^{-1}(t) = \sqrt{-\log(1-t)}$ und wir erhalten

$$t = 3 + (-\log(0.1))^{\frac{1}{2}} = 4.52.$$

e) (2 Punkte) Wir benutzen die Definition der bedingten Wahrscheinlichkeit

$$\begin{aligned} P[Y \leq t | Y < 7] &= 1 - P[Y > t | Y < 7] = 1 - \frac{P[t < Y < 7]}{P[Y < 7]} \\ &= 1 - \frac{F_Y(7) - F_Y(t)}{F_Y(7)} = \frac{F_Y(t)}{F_Y(7)}. \end{aligned}$$

Da $F_Y(7) = 1 - e^{-49}$, erhalten wir

$$P[Y \leq t] = \begin{cases} \frac{1-e^{-t^2}}{1-e^{-49}} & t \in [0, 7], \\ 0 & t < 0, \\ 1 & t > 7. \end{cases}$$

Bitte wenden!

f) i) (1 Punkt) Die Randdichte entsteht aus der gemeinsamen Dichtefunktion durch Wegintegrieren der anderen Variable. Also

$$f_Z(z) = \int_2^{12} f_{Z,U}(z,u)du = \int_2^{12} \frac{z}{10\sigma^2} e^{-\frac{z^2}{2\sigma^2}} du = \frac{z}{\sigma^2} e^{-\frac{z^2}{2\sigma^2}}, \quad z \geq 0.$$

ii) (1 Punkt) Die Dichte $f_Z(z)$ hat die Form (2), wenn $\alpha = 2$, und $\beta = \sigma\sqrt{2}$.

g) (1 Punkt) Wie in **f)**

$$f_U(u) = \int_0^{\infty} \frac{z}{10\sigma^2} e^{-\frac{z^2}{2\sigma^2}} dz = \frac{1}{10} \int_0^{\infty} \frac{z}{\sigma^2} e^{-\frac{z^2}{2\sigma^2}} dz = \frac{1}{10} e^{-\frac{z^2}{2\sigma^2}} \Big|_{\infty}^{0} = \frac{1}{10}, \quad u \in [2,12].$$

Also ist U auf dem Interval $[2,12]$ gleichverteilt.
Anderer Lösungsweg: $f_{Z,U}$ hängt auf $[0,\infty) \times [2,12]$ nicht von u ab; also muss (ohne Rechnung!) die Randdichte von U konstant sein, also identisch $\frac{1}{10}$ (wegen Integral 1). **Hier muss man also beim Korrigieren sehr aufpassen!!**

4. a) (2 Punkte) Es handelt sich um eine ungepaarte Stichprobe. Die Begründung ist, dass es keine natürliche Möglichkeit gibt, Beobachtungen der ersten und Beobachtungen der zweiten Versuchsreihe zu Paaren zusammenzufügen.

b) (2 Punkte) Da der Administrator feststellen möchte, ob das neue System leistungsstärker ist als das alte, sollte er einen einseitigen Test durchführen, der prüft, ob die Laufzeit auf den neuen Systemen niedriger ist als auf den alten. Die Nullhypothese besagt daher $H_0\colon \mu_X = \mu_Y$, die Alternativhypothese
$H_A\colon \mu_X > \mu_Y$.

c) (3 Punkte) Wir führen einen ungepaarten Zweistichproben-Test bei Normalverteilung und bekannter Varianz durch.

i) Die Teststatistik lautet

$$T := \frac{\bar{X}_{n_X} - \bar{Y}_{n_Y} - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}.$$

T ist unter H_0 standardnormalverteilt, $T \sim \mathcal{N}(0,1)$.

ii) Eine starke Reduktion der Laufzeit entspricht grossen Werten für T. Daher wollen wir dann verwerfen, wenn T gross ist. Es folgt

$$0.05 = P_{H_0}[T > c_>] = 1 - P_{H_0}[T \leq c_>],$$

und nach dem ersten Punkt ergibt sich $c_> = \Phi^{-1}(0.95) \approx 1.64$. Wir setzen in die Teststatistik ein und erhalten

$$T(\omega) = \frac{114 - 104}{\sqrt{262.5(\frac{1}{21} + \frac{1}{21})}} = \frac{10}{5} = 2 > 1.64.$$

Siehe nächstes Blatt!

Die Nullhypothese wird verworfen, der Administrator sollte die Systeme kaufen.

d) (3 Punkte) Wir führen den t-Test durch.

i) Die Teststatistik des t-Tests lautet

$$T := \frac{\bar{X}_{n_X} - \bar{Y}_{n_Y} - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}, \quad \text{wobei}$$
$$S := \frac{1}{n_X + n_Y - 2}((n_X - 1)S_X^2 + (n_Y - 1)S_Y^2).$$

Der Parameter der t-Verteilung ist $n_X + n_Y - 2 = 40$.

ii) Wieder wollen wir dann verwerfen, wenn T gross ist. Es folgt

$$0.05 = P_{H_0}[T > c_>] = 1 - P_{H_0}[T \leq c_>].$$

Wegen $T \sim t_{40}$ unter H_0 folgt $c_> \approx 1.68$. Es gilt $S = \frac{1}{40}(20 \cdot 280 + 20 \cdot 245) = 262.5$. Somit ergibt sich als Wert der Teststatistik

$$T(\omega) = \frac{114 - 104}{\sqrt{262.5(\frac{1}{21} + \frac{1}{21})}} = \frac{10}{5} = 2 > 1.68.$$

Wieder wird die Nullhypothese verworfen, der Administrator sollte also die Systeme kaufen.