Abgabe des Blattes während der Vorlesung am Mittwoch, den 22.01. Die folgenden Fragen sind diese Woche fakultativ: Aufgabe 2, iii), iv) und v). Diese Fragen werden zum Blatt 12 gehören.

Aufgabe 1: 9 Punkte

Warnung: In dieser Aufgabe werden nur Punkte für sorgfältige Beweise gegeben. In anderen Worten: jeder approximative Beweis ist 0 Punkte wert.

Sei X eine reelle Zufallsvariable mit Verteilungsfunktion $F(t) = \mathbb{P}(X \leqslant t)$.

i) Beweisen Sie die folgenden Eigenschaften: F ist monoton steigend, $\lim_{t\to-\infty} F(t) = 0$, $\lim_{t\to+\infty} F(t) = 1$ und F ist rechtsseitig stetig.

ii) Zeigen Sie: F ist genau dann stetig auf $\mathbb{R}$, wenn $\mathbb{P}(X = t) = 0$ für alle $t \in \mathbb{R}$ gilt. Geben Sie Beispiele von reellen Verteilungen mit stetiger bzw. unstetiger Verteilungsfunktion.

iii) Zeigen Sie: F ist genau dann streng monoton steigend, wenn $\mathbb{P}(X \in I) > 0$ für jedes offenen Interval $I \subseteq \mathbb{R}$. Geben Sie Beispiele von reellen Verteilungen mit streng monotoner bzw. nicht streng monotoner Verteilungsfunktion.

iv) Sei nun die verallgemeinerte (rechtsstetige) Inverse von F definiert:

$$\begin{aligned} F^{-1} : [0,1] &\to \mathbb{R} \cup \{\pm\infty\} \\ u &\mapsto \inf\{t \in \mathbb{R} : F(t) > u\}, \end{aligned}$$

mit der Konvention $\inf \varnothing = +\infty$. Bestimmen und zeichnen Sie F^{-1} für jeden der obigen Beispielen.

v) Zeigen Sie: Falls F streng monoton ist, gilt $F^{-1} \circ F(t) = t$ für jede $t \in \mathbb{R}$.

vi) Zeigen Sie: Falls F stetig ist, gilt $F \circ F^{-1}(u) = u$ für jede $u \in (0,1)$. Also falls F streng monoton und stetig ist, ist F^{-1} die klassische Inverse von F.

vii) Zeigen Sie (ohne Vorraussetzung an F): F^{-1} ist rechtsseitig stetig.

viii) Sei $t \in \mathbb{R}$ fest. Zeigen Sie (ohne Vorraussetzung an F):

$$\{u : u < F(t)\} \subseteq \{u : F^{-1}(u) \leqslant t\} \subseteq \{u : u \leqslant F(t)\}.$$

ix) Zeigen Sie (ohne Vorraussetzung an F): sei Z eine gleichmäßig auf $[0,1]$ verteilte Zufallsvariable. So ist $X := F^{-1}(Z)$ auf $\mathbb{R}$ gemäß F verteilt.

So werden mit dem Computer Zufallsvariablen von beliebigen Verteilungen generiert. Die meisten Softwaren, die sie kennen (Maple, Mathematica, Scilab, Excel, ...) verfügen über eine Funktion, oft *rand* genannt, die i.i.d. Zufallsvariablen $Z_1, \dots, Z_n$ mit Verteilung $U[0,1]$ generiert. Um i.i.d. Zufallsvariablen $X_1, \dots, X_n$ mit beliebiger Verteilungsfunktion F zu generieren, kann man dann die $X_i = F^{-1}(Z_i)$ anwenden.

Aufgabe 2: 10=2+2+2+2+2 Punkte

Fishers exakter Test auf Unabhängigkeit

Von $n = 1200$ Verkehrsunfällen waren 280 mit tödlichem Ausgang. Davon ereigneten sich 80 bei einer Geschwindigkeit von mehr als 150km/h. Insgesamt ereigneten sich 1100 Verkehrsunfälle bei einer niedrigeren Geschwindigkeit.

Die Kontingenztafel lautet

A \ B	mehr als 150 km/h	weniger als 150 km/h	Σ
tödlich	$h_n(11) := 80$	$h_n(12) := 200$	$h_n^A(1) := 280$
nicht tödlich	$h_n(21) := 20$	$h_n(22) := 900$	$h_n^A(2) := 920$
Σ	$h_n^B(1) := 100$	$h_n^B(2) := 1100$	$n = 1200$

Es gelten folgenden Definitionen: $A = B = \{1, 2\}$, $E = A \times B$. Wir definieren Θ als die Menge aller strikt positiven Zähldichten auf E:

$$\Theta := \Theta_E = \left\{ \nu = (\nu(ij))_{ij \in E} \in (0,1)^E : \sum_{ij \in E} \nu(ij) = 1 \right\}$$

Unser statistisches Modell ist das n-fache Produktmodell $(E^n, \mathcal{P}(E)^{\otimes n}, \nu^{\otimes n} : \nu \in \Theta)$. Für jede $i \in A$ bezeichnen wir $\nu^A(i) = \sum_{j \in B} \nu(ij)$ (ν^B wird genauso definiert). Wir betrachten dazu die Häufigkeiten $h_n(ij) = |\{1 \leqslant k \leqslant n : X_k = ij\}|$ für jede $ij \in E$.

Wir möchten testen, ob die beiden Merkmale A und B unabhängig sind. Die Nullhypothese hat daher die Gestalt $H_0 : \nu = \nu^A \otimes \nu^B$; entsprechend wählen wir $\Theta_0 = \{\alpha \otimes \beta : \alpha \in \Theta_A, \beta \in \Theta_B\}$ und $\Theta_1 = \Theta \backslash \Theta_0$.

i) Zeigen Sie, dass $\nu = \nu^A \otimes \nu^B$ genau dann gilt, wenn $\nu(11) = \nu^A(1)\nu^B(1)$.

ii) Zeigen Sie: für alle $p, q, r, s \in \mathbb{N}_0$ und $\nu = \nu^A \otimes \nu^B \in \Theta_0$ gilt

$$\begin{aligned}\mathbb{P}_\nu\left[h_n(11) = p, h_n(12) = q, h_n(21) = r, h_n(22) = s\right] \\ &= \frac{n!}{p!q!r!s!} \nu(11)^p \nu(12)^q \nu(21)^r \nu(22)^s \\ &= \frac{n!}{p!q!r!s!} \nu^A(1)^{p+q} \nu^A(2)^{r+s} \nu^B(1)^{p+r} \nu^B(2)^{q+s}.\end{aligned}$$

iii) Für alle $n \in \mathbb{N}$, $k, n_A, n_B \in \mathbb{N}_0$ und $\nu \in \Theta_0$ gilt:

$$\mathbb{P}_\nu[h_n(11) = k | h_n^A(1) = n_A, h_n^B(1) = n_B] = \mathcal{H}_{n_B; n_A, n-n_A}(\{k\}) = \mathcal{H}_{n_A; n_B, n-n_B}(\{k\}),$$

wobei die hypergeometrische Verteilung in Serie 1 definiert wurde.

iv) Sei $\alpha > 0$. Zu beliebigen Häufigkeiten $n_A, n_B \in \{0, \dots, n\}$ wählt man zwei Zahlen $c_+(n_A, n_B)$ und $c_-(n_A, n_B)$ mit $H_{n_B; n_A, n-n_A}(\{c_-, \dots, c_+\}) \geqslant 1 - \alpha$ und mit $c_- \leqslant \frac{n_A n_B}{n} \leqslant c_+$ ($\frac{n_A n_B}{n}$ ist der Erwartungswert von $H_{n_B; n_A, n-n_A}$).

Wie würden Sie nun vorgehen, um zum Niveau α einen Test der Nullhypothese $H_0 : \nu = \nu^A \otimes \nu^B$ gegen die Alternative $H_1 : \nu \neq \nu^A \otimes \nu^B$ zu entwickeln? Beschreiben Sie den Fehler erster bzw. zweiter Art in diesem Fall.

v) Wie würden Sie vorgehen, um zum Niveau α einen Test der Nullhypothese $H_0 : \nu = \nu^A \otimes \nu^B$ gegen die Alternative $H_1' : \nu(11) > \nu^A(1)\nu^B(1)$ ("die Unfällen mit einer Geschwindigkeit von mehr als 150 km/h sind tödlicher") zu entwickeln? Beschreiben Sie auch den Fehler erster bzw. zweiter Art in diesem Fall.