ETH Zürich HS 2015 D-INFK
Prof. Dr. P. Embrechts

Wahrscheinlichkeit und Statistik

Serie 9

Übung 9-1. Seien $X_1, X_2, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit absolut stetiger Dichtefunktion f und Verteilungsfunktion F. Definiere die zwei Zufallsvariablen

$$X_{(1)} = \min\{X_1, X_2, \ldots, X_n\} \quad \text{und} \quad X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}.$$

(a) Bestimme die Dichtefunktionen $f_{(1)}$ von $X_{(1)}$ und $f_{(n)}$ von $X_{(n)}$.

(b) Nehme an, dass X_k exponential verteilt ist mit Parameter $\lambda > 0$. Definiere die Folge

$$a_n := \frac{\log n}{\lambda} \quad \text{und} \quad b_n := \frac{1}{\lambda} \quad \text{für } n \in \mathbb{N}.$$

Zeige, dass

$$\lim_{n\to\infty} P\left[\frac{X_{(n)} - a_n}{b_n} \leq t\right] = e^{-e^{-t}} \quad \text{für } t > 0.$$

Bemerkung. Vergleiche mit dem zentralen Grenzwertsatz im Skript (Satz 5.5. S. 114). Die Grenzverteilungsfunktion $e^{-e^{-t}}\mathbb{1}_{\{t>0\}}$ heisst *Gumbel Verteilung* und spielt eine fundamentale Rolle in der Extremwerttheorie.

Übung 9-2. Ein grosses Kaufhaus möchte herausfinden, welcher Anteil seiner Kunden schon mindestens ein Mal Ladendiebstahl begangen hat. Dazu werden n Kunden anonym befragt. Da anzunehmen ist, dass trotz Anonymität viele Menschen dies nicht gerne zugeben, wird der Test folgendermassen modifiziert. Jeder der Befragten wirft (unabhängig von den anderen Befragten) geheim eine (faire) Münze und anwortet nur dann wahrheitsgetreu, falls die Münze Kopf zeigt. Wenn die Münze Zahl zeigt, antwortet der Befragte in jedem Fall mit "Nein". Sei p die Wahrscheinlichkeit, dass der Befragte mindestens einen Ladendiebstahl begangen hat. Ferner sei S_n die Anzahl der Nein-Antworten unter den n befragten Personen.

(a) Wie gross ist die Wahrscheinlichkeit, dass eine unter den Befragten beliebig ausgewählte Person mit "Nein" antwortet?

(b) Wie gross ist die mittlere Anzahl der Nein-Antworten bei $n = 1000$ Befragten?

(c) Wir nehmen an, dass $p = 20\%$. Finde nun für $\delta = 1\%$ eine obere Schranke für die Wahrscheinlichkeit, dass bei 1000 befragten Kunden mindestens 909 mit "Nein" antworten.

Bemerkung. Benutze ein Resultat aus der Vorlesung über Chernoff-Schranken.

Übung 9-3. Seien X und Y zwei unabhängig und identisch exponential verteilte Zufallsvariablen mit Parameter λ. Definiere die Zufallsvariablen

$$U := \frac{X}{X+Y} \quad \text{und} \quad V := X + Y.$$

Wir bezeichnen die dazugehörigen Dichtefunktionen mit f_U und f_V , sowie die gemeinsame Dichtefunktion $f_{(U,V)}$.

(a) Berechne f_U und f_V.

(b) Berechne $f_{(U,V)}$. Sind U und V unabhängig?

Siehe nächste Seite!

Challenge Serie 9. Diese Aufgabe ist die Fortsetzung von Aufgabe 1. Seien $X_1, X_2, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit absolut stetiger Dichtefunktion f und Verteilungsfunktion F. Wir ordnen (X_k) ihrer Grösse entsprechend an und bezeichnen sie mit $(X_{(k)})$, d.h., mit Wahrscheinlichkeit 1 gilt

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}. \tag{1}$$

Die Folge $(X_{(k)})$ heisst *Ordnungsstatistik* von (X_k). Mit dieser Notation gilt insberondere

$$X_{(1)} = \min\{X_1, X_2, \ldots, X_n\} \quad \text{und} \quad X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}.$$

(a) Bestimme die gemeinsame Dichtefunktion $f_{(1),\ldots,(n)}$ von $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$.

(b) Bestimme mit einem heuristischen Argument die Dichtefunktion $f_{(k)}$ von $X_{(k)}$, $k = 1, \ldots, n$.
Tipp. Verwende die Notation aus dem Skript auf S. 86:

$$P[x < X_{(k)} < x + \mathrm{d}x] \approx f_{(k)}(x)\,\mathrm{d}x.$$

(c) Nehme an, dass X_k uniform auf $[0,1]$ verteilt ist. Berechne $E[X_{(k)}]$ und $\mathrm{Var}(X_{(k)})$.
Tipp. Verwende die Formel

$$\int_0^1 x^{m-1}(1-x)^{n-1}\,\mathrm{d}x = \frac{(m-1)!(n-1)!}{(m+n-1)!} \quad \text{für } m, n \in \mathbb{N}.$$

(d) Sei $r = \frac{p}{q}$ eine rationale Zahl in $[0,1]$ und seinen $X_1, X_2, \ldots, X_{qn}$ unabhängig und identisch uniform auf $[0,1]$ verteilte Zufallsvariablen, wobei $n \in \mathbb{N}$. Zeige, dass zu jedem $\varepsilon > 0$

$$\lim_{n\to\infty} P[\left|X_{(pn)} - r\right| > \varepsilon] = 0,$$

wobei $X_{(pn)}$ die $(p\,n)$–te Ordungsstatistik von (X_k) bezeichnet.
Bemerkung. Vergleiche mit dem schwachen Gesetz der grossen Zahl im Skript.