Script generated by TTT

Title: Meixner: Zue_DWT (26.06.2014)

Date: Thu Jun 26 14:12:31 CEST 2014

Duration: 52:32 min

Pages: 24

SS 2014

Zentralübung zur Vorlesung Diskrete Wahrscheinlichkeitstheorie

Dr. Werner Meixner

Fakultät für Informatik
TU München

http://www14.in.tum.de/lehre/2014SS/dwt/uebung/

26. Juni 2014

ZÜ VII

Übersicht:

1. Übungsbetrieb — Fragen, Probleme? Klausur!
2. Thema — Wiederholung TA 2 von Blatt 8
   Aufgabenstellung
   Vorüberlegung
   VA 2
3. Vorbereitung — VA Blatt 10

1. Übungsbetrieb

1.1 Fragen, Probleme?

Aktuelle Fragen?

2. Thema: Wiederholung TA 2 von Blatt 8

Die Aufgabe VA 2 wird analog wie Tutoraufgabe 2 von Blatt 8 gelöst und soll als Wiederholung dienen bzw. der Vorbereitung stochastischer Prozesse.

2.1 Aufgabenstellung

An der dänischen Grenze werden Grenzkontrollen durchgeführt. Im Schnitt treffen alle 30 Sekunden an der Grenzstation Personen ein, die zu kontrollieren sind.

Die Zeit zwischen zwei Kontrollen sei exponentialverteilt mit Parameter $\frac{1}{30}$. Wenn 2 Minuten lang kein Kontrollfall eingetroffen ist, dann machen die Grenzbeamten Ruhepause.

Seien $T_1, T_2, \ldots$ die Zeitspannen zwischen dem Eintreffen von zu kontrollierenden Personen und W die Wartezeit bis zur nächsten Ruhepause.

1. Berechnen Sie $\mathbb{E}[T_1 \mid T_1 \geq 120]$.
2. Berechnen Sie $\mathbb{E}[W]$.

2.2 Vorüberlegung

a) Es wird offenbar eine Folge von Zeitpunkten z_i betrachtet, zu denen Personen an der Grenzstation eintreffen.

b) Die Zeitdifferenzen $T_i = z_i - z_{i-1}$ werden als exponentialverteilt angenommen. Dies bedeutet, dass diese T_i nicht davon abhängen, wie lange noch keine Person eingetroffen ist (Gedächtnislosigkeit).

c) Der Parameter λ bedeutet „Anzahl der Personen pro Zeiteinheit" im Durchschnitt, hier also $\frac{1}{30}$ Personen pro Sekunde als Erwartungswert.

d) Alle T_i sind unabhängig, d. h. die Menge der T_i ist unabhängig.

e) Falls $T_i > 120$, dann gibt es eine Ruhepause.

Die Frage ist, wie lange man durchschnittlich warten muss, bis erstmalig $T_N > 120$ festgestellt wird.

2.3 VA 2

(1) Berechnen Sie $\mathbb{E}[T_1 \mid T_1 \geq 120]$.

Lösung

Sei $\lambda = \frac{1}{30}$.

Dann gilt

$$\mathbb{E}[T_1] = 30\,.$$

Es gilt

$$\mathbb{E}[T_1 \mid T_1 \geq 120] = 120 + \mathbb{E}[T_1] = 150\,.$$

Beide obigen Aussagen folgen aus der Tatsache, dass T_1 exponentialverteilt ist.

2.3 VA 2

(1) Berechnen Sie $\mathbb{E}[T_1 \mid T_1 \geq 120]$.

Lösung

Sei $\lambda = \frac{1}{30}$.

Dann gilt

$$\mathbb{E}[T_1] \;=\; 30\,.$$

Es gilt

$$\mathbb{E}[T_1 \mid T_1 \geq 120] \;=\; 120 + \mathbb{E}[T_1] \;=\; 150\,.$$

Beide obigen Aussagen folgen aus der Tatsache, dass T_1 exponentialverteilt ist.

(2) Berechnen Sie $\mathbb{E}[W]$.

Lösung

Wir wählen die Bezeichnung W' wie in TA 2 von Blatt 8, d.h.,

$$W = 120 + \sum_{j=1}^{N-1} T_j$$

oder

$$W = 120 + W'$$

mit

$$W' = \sum_{j=1}^{N-1} T_j\,.$$

N ist eine geometrisch verteilte Zufallsvariable mit Erfolgswahrscheinlichkeit $p = \Pr[T > 120]$.

Es gilt

$$\begin{aligned} p &= \Pr[T > 120] \\ &= 1 - \Pr[T \leq 120] \\ &= 1 - (1 - e^{-\lambda \cdot 120}) \\ &= e^{-4}\,. \end{aligned}$$

Berechnung von $\mathbb{E}[W']$:

Wir setzen $T = T_1$.

$$\begin{aligned} \mathbb{E}[W'] &= \sum_{n=1}^{\infty} \mathbb{E}[W' | N = n] \cdot p(1-p)^{n-1} \\ &= \sum_{n=1}^{\infty} \mathbb{E}[T | T \leq 120] \cdot (n-1) \cdot p(1-p)^{n-1} \\ &= \mathbb{E}[T | T \leq 120] \cdot \sum_{n=1}^{\infty} (n-1) \cdot p(1-p)^{n-1} \\ &= \mathbb{E}[T | T \leq 120] \cdot \mathbb{E}[N-1]. \end{aligned}$$

Diese Berechnung kann man verkürzen, wenn man den Satz über zufällige Summen von Zufallsvariablen anwendet.

Es gilt
$$\mathbb{E}[N-1] = e^4 - 1\,.$$

$\mathbb{E}[T|T \le 120]$ erhalten wir aus den Gleichungen

$$\begin{aligned}
\mathbb{E}[T] &= \mathbb{E}[T|T \le 120] \cdot \Pr[T \le 120] \;+\; \mathbb{E}[T|T \ge 120] \cdot \Pr[T \ge 120] \\
&= \mathbb{E}[T|T \le 120] \cdot (1 - e^{-4}) + 150 \cdot e^{-4} \\
&= 30\,.
\end{aligned}$$

Es folgt

$$\mathbb{E}[T|T \le 120] = \frac{30 - 150 \cdot e^{-4}}{1 - e^{-4}} = \frac{30 \cdot e^4 - 150}{e^4 - 1}\,.$$

Ergebnis

$$\begin{aligned}
\mathbb{E}[W'] &= 30 \cdot e^4 - 150\,, \\
\mathbb{E}[W] &= 30 \cdot e^4 - 150 + 120 \\
&\approx 26,8 \text{ (Minuten)}\,.
\end{aligned}$$

3. Vorbereitung

3.1 VA 1

Auf zwei unabhängigen Servern stehe ein Web-Dienst zur Verfügung. Es soll festgestellt werden, welcher Server schnellere Antwortzeiten liefert.

Dazu werden $n = 1000$ Anfragen an die Server geschickt und es wird festgestellt, von welchem Server die Antwort zuerst eintrifft. Dabei gehen wir davon aus, dass Pakete nicht gleichzeitig empfangen werden können.

In 540 Fällen antwortet Server A vor Server B.

Aufgabe

Wir wählen als Nullhypothese H_0 die Aussage, dass Server B im Mittel schneller ist als Server A.

Kann man für einen entsprechenden statistischen Test auf einem Signifikanzniveau von $\alpha = 0,04$ die Nullhypothese annehmen?

Formulieren Sie hierzu den Test und weisen Sie Ihre Behauptung nach.

Lösung

Sei X eine Bernoulli-verteilte Zufallsvariable mit Wert 1 genau dann, wenn A schneller antwortet als B. Wir schreiben $p = \Pr[X = 1]$.

Seien X_i für $i \in \{1, 2, \ldots, n\}$ unabhängige Wiederholungen von X mit $n = 1000$.
Dann ist die Testgröße $T = \sum_{i=1}^{n} X_i$ binomialverteilt.

Wir formulieren einen approximativen Binomialtest.

Nullhypothese H_0: $p < \frac{1}{2}$, Alternative H_1: $p \geq \frac{1}{2}$.

Die Nullhypothese mit Signifikanzniveau $\alpha = 0,04$ anzunehmen, heißt aber, die
triviale Alternative mit Signifikanzniveau $\alpha = 0,04$ abzulehnen.

Um also das Testschema des approximativen Binomialtests nach Vorlesung anwenden zu können, bezeichnen wir die

Alternative H_1 als H_0' und H_0 als H_1'.

Nun testen wir, ob die Hypothese H_0' mit Signifikanzniveau $\alpha = 0,04$ abgelehnt werden kann.

Wir setzen also

Nullhypothese H_0': $p \geq \frac{1}{2}$, Alternative H_1': $p < \frac{1}{2}$,

und wenden den approximativen Binomialtest wie folgt an.

Mit $p_0 = \frac{1}{2}$, $n = 1000$ und $T = h$ sei

$$Z = \frac{h - np_0}{\sqrt{np_0(1-p_0)}} = \frac{540 - 500}{\sqrt{250}} \approx 2.5298\ldots\ .$$

Das Ablehnungskriterium für H_0' mit $\alpha = 0.04$ ist $Z < z_{0.04} \approx -0{,}9599$.

Dieses Kriterium ist wegen

$$Z = \frac{540 - 500}{\sqrt{250}} > z_{0.04}$$

nicht erfüllt, d. h. wir können H_0 nicht mit Signifikanz $\alpha = 0,04$ annehmen.

Nun könnten wir sofort mit gleichem Test prüfen, ob H_0 mit Signifikanzniveau, d. h., maximaler Irrtumswahrscheinlichkeit $\alpha = 0.04$ ablehnen können. Und dies ist in der Tat so.

Es gilt nämlich, dass H_0 mit Signifikanzniveau $\alpha = 0.04$ abgelehnt werden kann. Dies folgt aus $Z > z_{1-\alpha} \approx 1{,}6667$.

Aber das war aber nicht die Frage.

Gefragt war, ob wir H_0 mit Signifikanzniveau $\alpha = 0.04$ annehmen können.

Nun könnten wir sofort mit gleichem Test prüfen, ob H_0 mit Signifikanzniveau, d. h., maximaler Irrtumswahrscheinlichkeit $\alpha = 0.04$ ablehnen können. Und dies ist in der Tat so.

Es gilt nämlich, dass H_0 mit Signifikanzniveau $\alpha = 0.04$ abgelehnt werden kann. Dies folgt aus $Z > z_{1-\alpha} \approx 1{,}6667$.

Aber das war aber nicht die Frage.

Gefragt war, ob wir H_0 mit Signifikanzniveau $\alpha = 0.04$ annehmen können.

3.2 VA 3

Seien $T_1, T_2, \ldots$ unabhängige exponentialverteilte Zufallsvariablen mit gleichem Parameter λ. Seien $S_n = \sum_{i=1}^{n} T_i$ für alle $n \in \mathbb{N}_0$.

Man zeige für alle $n \in \mathbb{N}$ für die Verteilungsfunktion von S_n, dass für alle $t \geq 0$ gilt:

$$F_{S_{n+1}}(t) = -\frac{\lambda^n t^n}{n!} e^{-\lambda t} + F_{S_n}(t)\,.$$

Bemerkung

Man vergleiche das Thema Gammaverteilung in ZÜ 6.

Beweis

Zunächst beweisen wir

$$F_{S_n}(t) - F_{S_{n+1}}(t) = \frac{1}{\lambda} f_{S_{n+1}}(t)\,.$$

$$\begin{aligned}
F_{S_{n+1}}(t) &= \int_{-\infty}^{\infty} f_T(T) \cdot \left(\int_{-\infty}^{t-T} f_{S_n}(S)\,\mathrm{d}S \right) \mathrm{d}T \\
&= \int_{-\infty}^{\infty} f_T(T) \cdot F_{S_n}(t-T)\,\mathrm{d}T \\
&= \int_{0}^{t} f_T(T) \cdot F_{S_n}(t-T)\,\mathrm{d}T \\
&= [F_T(T) F_{S_n}(t-T)]_{T=0}^{T=t} + \int_{0}^{t} F_T(T) f_{S_n}(t-T)\,\mathrm{d}T \\
&= \int_{0}^{t} (1 - \frac{1}{\lambda} f_T(T)) f_{S_n}(t-T)\,\mathrm{d}T \\
&= \int_{0}^{t} f_{S_n}(t-T)\,\mathrm{d}T - \frac{1}{\lambda} \int_{0}^{t} f_T(T) \cdot f_{S_n}(t-T)\,\mathrm{d}T \\
&= F_{S_n}(t) - \frac{1}{\lambda} f_{S_{n+1}}((t)\,.
\end{aligned}$$

Beweis:

Für $n = 1$ gilt die Formel, da für $t \geq 0$ $f_T(t) = \lambda e^{-\lambda t}$ gilt.

Induktionsschritt von n auf $n+1$ für alle $n \geq 1$:

$$\begin{aligned}
f_{S_{n+1}}(t) &= \int_0^t f_T(t-x) f_{S_n}(x)\,\mathrm{d}x \\
&= \int_0^t \lambda e^{-\lambda(t-x)} \cdot \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t}\,\mathrm{d}x \\
&= \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda t} \int_0^t x^{n-1}\,\mathrm{d}x \\
&= \frac{\lambda^{n+1} t^n}{n!} e^{-\lambda t}.
\end{aligned}$$