*Projektpraktikum II (Stochastik)*
Wintersemester term 2018/19
Humboldt-Universität zu Berlin

Sebastian Holtz

# Projektaufgaben Block 3

1. In einer Studie wurde anhand von Autopsiedaten ermittelt wie sich Lungenkrebs im Körper ausbreitet (Beispiel 2.12 im Buch von Dobrow). Für insgesamt 50 verschiedene Positionen im Körper wurden die Übergangswahrscheinlichkeiten einer Markovkette berechnet, die stets in Zustand 23 (=Lunge) startet. Lade die Übergangsmatrix vom Web über die Datei *cancer.csv.* Berechne für drei verschiedene Positionen $x \in \{1, \dots, 50\}$ im Körper deiner Wahl die mittlere Anzahl von Schritten die der Krebs braucht, um dorthin zu gelangen (mittlere Eintrittszeit $\mathbb{E}[\tau_x]$, siehe Vorlesung). Verwende die beiden folgenden Methoden, um $\mathbb{E}[\tau_x]$ zu approximieren:

   (a) Es gilt laut Vorlesung $\mathbb{E}[\tau_x] = \sum_{k=0}^{\infty} kP(\tau_x = k)$. Weiterhin gilt (muss nicht gezeigt werden), dass $P(\tau_x = k) = \sum_{i,j \neq x} \tilde{P}_{ij}^{k-1} P_{jx} \mu(i)$, wobei $\mu$ die Anfangsverteilung und $P$ die Übergangsmatrix ist. Außerdem ist $(\tilde{P}_{ij})$ die Übergangsmatrix $P$, wenn man die $x$-te Spalte und die $x$-te Zeile von $P$ durch Nullen ersetzt, und $\tilde{P}^{k-1} = \tilde{P}^{k-2}\tilde{P}$.

   (b) Simuliere 10.000 Mal die Markovkette und bestimme $\mathbb{E}[\tau_x]$ durch Mittelwertbildung. Wähle dazu einen maximalen Zeithorizont und verwende nur Pfade bei denen der Zielzustand auch erreicht wird.

2. Lade die Tabelle *traffic.csv* vom Web in R. Sie enthält Verkehrsdaten von vorbeifahrenden Autos an einem Autobahnabschnitt in Abständen von 15 Minuten. Bestimme eine Intensitätsfunktion $\tilde{\lambda}(t)$ für $0 \leq t < 24$ als mittlere Anzahl an beobachteten Autos pro Minute, indem du $\tilde{\lambda}(t)$ stückweise konstant definierst auf den 96 möglichen 15-Minuten-Intervallen. Wir wollen einen Stresstest für den Autobahnabschnitt betrachten anhand der beobachteten Daten. Simuliere dazu einen Poisson-Prozess auf $[0, 24)$ mit gesteigerter Intensität $\lambda(t) = 1.2\tilde{\lambda}(t)$. Wiederhole die Simulation 10.000 Mal (d.h. für 10.000 Tage) und bearbeite die Fragen unten. Diskutiere die Ergebnisse kurz.

   (a) Wieviele Autos kommen insgesamt in 24 Stunden im Schnitt an dem Autobahnabschnitt vorbei?

   (b) In welcher Stunde (0 - 24) fährt im Schnitt das 1000. Auto vorbei? Verwende 24 in der Berechnung, wenn es insgesamt weniger als 1000 Autos gab an einem Tag.

   (c) Wieviele Autos fahren durchschnittlich zwischen 1:30 und 1:45 vorbei?

(d) Bestimme für jede der 10.000 Simulationen die einstündige Periode $[t, t+1)$, wobei $t \in \{0, 0.25, 0.5, 0.75, \ldots, 23.0\}$, in der die meisten Autos vorbeifahren. Erzeuge ein Histogramm für diese Perioden.

(e) Erzeuge ein Histogramm für die Anzahl der vorbeifahrenden Autos in dieser Periode gemäß den 10.000 Simulationen.

3. Sei $B = (B_t)_{t \geq 0}$ eine $d$-dimensionale Brownsche Bewegung, $d \geq 1$, d.h. $B_t = (B_t^{(1)}, \ldots, B_t^{(d)})$, wobei die $B_t^{(i)}$ unabhängige eindimensionale Brownsche Bewegungen sind, $i = 1, \ldots, d$. Sei $T_{R,x} = \inf\{t > 0 : \|B_t + x\| = R\}$, für $R > 0$ und $x \in \mathbb{R}^d$, der erste Zeitpunkt zu dem $B_t + x$ den Rand des Balls mit Radius $R$ im $\mathbb{R}^d$ trifft. Man kann zeigen, dass folgender Satz gilt:

   **Satz (Polyá):** Sei $x \in \mathbb{R}^d$ mit $\|x\| \geq R$. Dann ist $P(T_{R,x} < \infty) = \left(\frac{R}{\|x\|}\right)^{d-2}$.

   Illustriere mit Hilfe dieses Satzes und anhand von Simulationen die folgende Aussage von Polyá:

   > Ein betrunkener Mensch findet stets zurück nach Hause, ein betrunkener Vogel aber vielleicht nicht.

   *Hinweis:* Man kann zeigen, dass $P(T_{R,x} < \infty) = \lim_{r \to \infty} P(T_{R,x} < r)$. Wähle $r$ ausreichend groß und simuliere $B_t$ für $0 \leq t \leq r$ mit einem hinreichend kleinen Gitter.

4. Sei $X = (X_t)_{t \geq 0}$ ein stochastischer Prozess, der die stochastische Differentialgleichung $dX_t = (\mu - X_t)dt + \sigma dB_t$ erfüllt für eine Zufallsvariable $X_0$, $\mu \in \mathbb{R}$, $\sigma > 0$. $X$ heißt dann *Ornstein-Uhlenbeck-Prozess*. Man kann zeigen, dass die exakte Lösung durch $X_t = X_0 e^{-t} + \mu(1 - e^{-t}) + \int_0^t \sigma e^{s-t} dB_s$ gegeben ist, wobei $X_0$ der Anfangswert ist.

   (a) Bestimme Erwartungswert und Varianz von $X_t$ über die exakte Lösung. Du darfst verwenden, dass $\int_0^t f(s) dB_s \overset{d}{\sim} N(0, \int_0^t f^2(s) ds)$ für $f \in L^2(\mathbb{R})$ gilt. Stelle eine Vermutung für eine Grenzverteilung von $X_t$ auf für $t \to \infty$.

   (b) Simuliere und plotte Pfade bezüglich $X_t$ für verschiedene deterministische Startwerte $X_0$ und $\mu = 1$, $\sigma = 1$. Beschreibe das Ergebnis. Interpretiere die Rolle von $\mu$ und $\sigma$.

---

Lösungen in einer zip-Datei per Mail bis zum 20. Januar, 23:59 Uhr, abgeben.