Block 4:
Markov-Chain-Monte-Carlo-Methoden

Wintersemester 2018/19

Überblick

- Rückblick: Simulation von Zufallszahlen
- Ergodische Markovketten
- Der Metropolis-Hastings-Algorithmus
- Anwendungsbeispiel

Literatur:

- Christian Robert, George Casella: *Introducing Monte Carlo Methods with R, 2010*
- Christian Robert, George Casella: *Monte Carlo Statistical Methods, 2004*

Rückblick: Simulation von Zufallszahlen

- Simulation/Samplen von Zufallsvariable X mit Dichte f
  Anwendungen: Approximation von Erwartungswerten
  $\mathbb{E}[g(X)] \approx \frac{1}{n} \sum_{k=1}^{n} g(X_k)$
- Methoden aus Block 1:
  - Inversionsmethode (Inversion der Verteilungsfunktion)
  - spezielle Methoden für Normalverteilung, diskrete Verteilungen
  - Verwerfungsmethode
- Gemeinsamkeiten dieser Methoden:
  - erzeugen i.i.d.-Samples
  - basieren auf mehr oder weniger starken Annahmen (z.B. Existenz einer "guten" Kandidatendichte bei der Verwerfungsmethode)
- Problem: Annahmen sind häufig nicht erfüllt bzw. Kandidatendichten sind schwierig zu finden
  Beispiel: (Bayes) A-Priori-Verteilung θ, "Daten"
  $X = (X_1, \ldots, X_n)$ mit (bedingter) Dichte $f^{X|\theta}$
  $\Rightarrow$ A-Posteriori-Dichte: $f^{\theta|X} = \frac{f^{\theta,X}}{f^X} = \frac{f^{X|\theta} f^\theta}{f^X} \propto f^{X|\theta} f^\theta$

Ergodische Markovketten

- fundamentale Idee:
  - $(X_n)_{n\geq 0}$ sei eine Markovkette mit invarianter Verteilung μ (d.h. $P_\mu(X_1 \in A) = \mu(A)$ für alle Borelmengen A)
  - μ habe Lebesgue-Dichte f
  - *Ergodensatz* (Birkhoff): Wenn $(X_n)_{n\geq 0}$ *ergodisch* ist, dann gilt für alle $g \in L^1(\mu)$, dass

$$\frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \to \int g(x) f(x)\, dx = \mathbb{E}[g(X)], \quad n \to \infty.$$

  - für "große" n erzeugt also die Markovkette annähernd Samples bezüglich f
  - diese Samples sind nicht i.i.d.!
- *wie konstruiert man eine solche Markovkette?*

Der Metropolis-Hastings-Algorithmus

- Algorithmus:
  - wähle X_0 beliebig/zufällig, so dass $f(X_0) > 0$
  - gegeben X_n, erzeuge $Y_n \sim q(\cdot|X_n)$ (q Vorschlagsdichte)
  - Akzeptanzwahrscheinlichkeit:
    $r(X_n, Y_n) := \min\left\{1, \frac{f(Y_n)q(X_n|Y_n)}{f(X_n)q(Y_n|X_n)}\right\}$
  - setze

$$X_{n+1} = \begin{cases} Y_n & \text{mit Wahrscheinlichkeit } r(X_n, Y_n), \\ X_n & \text{mit Wahrscheinlichkeit } 1 - r(X_n, Y_n), \end{cases}$$

- das Ergebnis ist wirklich eine Markovkette!
- Spezialfall: $q(y|x) = q(x|y)$ ist symmetrisch (z.B. Gaussdichte)
  $\Rightarrow r(X_n, Y_n) = \min\{1, f(Y_n)/f(X_n)\}$
  Interpretation:
  - akzeptiere Y_n immer, wenn es in einem Bereich größerer Dichte liegt
  - wenn Y_n in einem Bereich kleinerer Dichte liegt, dann vertrauen wir dem Sample weniger und "werfen eine Münze"

Eigenschaften von $(X_n)_{n\geq 0}$ (1)

- Übergangskern der Markovkette ist

  $$K(x,A) = P(X_{n+1} \in A | X_n = x)$$
  $$= \int_A q(y|x)\, r(x,y)\, dy + \delta_x(A)\left(1 - \int q(y|x)\, r(x,y)\, dy\right),$$

  für Borelmenge A, d.h.
  $K(x,y) = q(y|x)\, r(x,y) + \delta_x(y)\left(1 - \int q(y|x)\, r(x,y)\, dy\right)$
- Übergangskern erfüllt *detailed balance* bezüglich f, d.h. es gilt
  $K(y,x)\, f(y) = K(x,y)\, f(x)$
  $\Rightarrow$ das Maß μ mit Dichte f ist invariante Verteilung von $(X_n)_{n\geq 0}$, da

  $$P_\mu(X_1 \in A) = \int K(x,A)\, f(x)\, dx = \int K(x,y)\, f(x)\, \mathbb{1}_A(y)\, d(x,y)$$
  $$= \int K(y,x)\, f(y)\, \mathbb{1}_A(y)\, d(x,y) = \int f(y)\, \mathbb{1}_A(y)\, d(y)$$
  $$= \mu(A)$$

Eigenschaften von $(X_n)_{n\geq 0}$ (2)

- wenn $q(y|x) > 0$ für alle $(x, y) \in E \times E$, wobei E=$supp(f)$, dann ist $(X_n)_{n\geq 0}$ ist *irreduzibel*, d.h. jeder Punkt im Support E von f kann in einem Schritt erreicht werden (denn $K(x, y) > 0$)
- man kann zeigen, dass $(X_n)_{n\geq 0}$ auch *aperiodisch* ist, wenn $P(X_{n+1} = X_n) > 0$
- eine *aperiodische irreduzible* Markovkette ist *ergodisch*, d.h. Ergodensatz ist anwendbar für $(X_n)_{n\geq 0}$
- f bzw. $q(\cdot|x)$ müssen nur bis auf konstante Faktoren bekannt sein

Eigenschaften von $(X_n)_{n\geq 0}$ (3)

- es sollte leicht sein von $q(\cdot|x)$ zu sampeln
- erzeugte Samples hängen stark von Konvergenzgeschwindigkeit gegen die stationäre Verteilung ab
- Samples sind *nicht* unabängig, Approximation erst nach Burn-in-Phase gut
- Vergleich mit der Verwerfungsmethode:
  Accept-Reject (Verwerfungsmethode):
  - erzeuge $U \overset{d}{\sim} \mathcal{U}(0,1)$, $Y \overset{d}{\sim} g$ (g = Kandidatendichte)
  - wenn $U \leq f(Y)/Mg(Y)$, dann *akzeptiere* $X := Y$
  - andernfalls *lehne* Y *ab* und kehre zum Anfang zurück

Welches q?

- das beste q wäre $q = f \Rightarrow r(X_n, Y_n) = 1$

1. Methode: der unabhängige Metropolis-Hastings-Algorithmus
   - $q(\cdot|x) \equiv q$
   - $r\left(X_n, Y_n\right) = \min\left\{1, \frac{f(Y_n)q(X_n)}{f(X_n)q(Y_n)}\right\}$
   - Ähnlich zur Verwerfungsmethode, Bedingung $f/q \leq M$ muss nicht erfüllt sein
2. Methode: der random-walk Metropolis-Hastings-Algorithmus
   - $Y_n = X_n + \epsilon_n$, ϵ_n i.i.d. mit Varianz σ^2 und symmetrisch mit Dichte q, so dass Y_n Dichte $q(\cdot - X_n)$ hat
   - $q(-x) = q(x)$
   - $r\left(X_n, Y_n\right) = \min\left\{1, \frac{f(Y_n)}{f(X_n)}\right\}$
   - Verhalten hängt stark von σ^2 ab

Anwendungsbeispiel

- 28. Januar 1986: Explosion der Raumfähre Challenger aufgrund von Materialermüdungserscheinungen an Dichtungsringen
- Wahrscheinlicher Grund: ungewöhnlich niedrige Außentemperatur von 31 Grad Fahrenheit (ca. 0 Grad Celsius)

Materialprobleme	1	1	1	1	0	0	0	0	0	0	0	0
Temperatur (F)	53	57	58	63	66	67	67	67	68	69	70	70
Materialprobleme	1	1	0	0	0	1	0	0	0	0	0	
Temperatur (F)	70	70	72	73	75	75	76	76	78	79	81	

Anwendungsbeispiel

- modelliere Materialprobleme mit *logistischer Regression*:
  - Annahme: wir beobachten $Y_i \overset{iid}{\sim} \text{Ber}(p(x_i))$
  - x_i=Temperatur, Y_i=Materialproblem Ja/Nein
  - $p(x_i) = P(Y_i = 1 | X_i = x_i) = \frac{\exp(\alpha + x_i\beta)}{1+\exp(\alpha+x_i\beta)}$, für Parameter $\alpha, \beta \in \mathbb{R}$
  - das ist ein *verallgemeinertes lineares Modell*
- Ziel: Bestimme α, β anhand der Daten und mache Vorhersagen für ungesehene Temperaturen & bestimme die Verteilung dieser Vorhersagen mittels wiederholter Simulationen.