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

Sebastian Holtz

Projektaufgaben Block 4

1. Betrachte die Situation aus Aufgabe 2 in Block 1 und erzeuge mit dem Metropolis-Hastings-Algorithmus Samples bezüglich der Standardnormalverteilung. Diskutiere die Ergebnisse anhand der folgenden Punkte:

   - Histogramme, QQ-Norm-Plots und Kolmogorov-Smirnov-Tests,
   - das Verhalten für Werte außerhalb von $[-2, 2]$,
   - Anzahl der abgelehnten Samples,
   - Konvergenzeigenschaften der Simulation bezüglich Stationarität,
   - Konvergenz von empirischen Momenten $\frac{1}{n}\sum_{k=1}^{n} X_k^m$, $m \geq 1$,
   - Autokorrelationen (mit der Funktion `acf`).

   Verwende die beiden folgenden konkreten Methoden aus der Vorlesung:

   (a) Der unabhängige Metropolis-Hastings-Algorithmus mit der Dichte der Laplace-Verteilung zum Parameter $\alpha > 0$,

   $$g_\alpha(x) = \frac{\alpha}{2} \exp\left(-\alpha \left|x\right|\right), \quad x \in \mathbb{R},$$

   als Vorschlagsdichte q. Diskutiere darüberhinaus die Wahl von α und vergleiche mit dem Accept-Reject-Algorithmus aus Aufgabe 4.3 in Block 1.

   (b) Der random walk Metropolis-Hastings-Algorithmus, wobei $\varepsilon_n \sim \mathcal{U}(-\delta, \delta)$ für $\delta > 0$, d.h. die Vorschlagsdichte ist $q(y|x) = \frac{1}{2\delta}\mathbb{1}_{[x-\delta, x+\delta]}(y)$.

2. Betrachte die Beobachtungen des Challenger-Datensatzes aus der Vorlesung, welche mittels der logistischen Regression

$$Y_i \sim \mathrm{Ber}(p(x_i)), \quad p(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}, \quad i = 1, \ldots, 23, \tag{0.1}$$

modelliert werden sollen.

(a) Fitte und analysiere mittels `glm` ein verallgemeinertes lineares Modell. Zeichne die zugehörige logistische Funktion und Beobachtungen in einen gemeinsamen Plot. Recherchiere und beschreibe (kurz) den Hosmer-Lemeshow-Test und überprüfe mit diesem, ob das Modell den Daten gerecht wird. Welche Schätzung für $P(Y_i = 1 | X_i = 31)$ liefert das Modell?

Recherchiere und erkläre das Konzept der *Bayes-Analyse* und implementiere dazu den Prior von $\vartheta = (\alpha, \beta)$ mittels Verteilung

$$\pi(\alpha, \beta) = \pi_\alpha(\alpha | b)\pi_\beta(\beta) = \frac{1}{b} e^\alpha e^{-e^\alpha / b}, \ \alpha \in \mathbb{R},$$

wobei b einen sogenannten Hyperparameter darstellt.

(b) Zeige, dass für $A \sim \pi_\alpha$ die Identität

$$\mathbb{E}[A] = \log(b) - \gamma,$$

gilt und gib eine Näherung von γ (Recherchieren!). Wähle nun b, so dass

$$\mathbb{E}[A] = \hat{\alpha} \tag{0.2}$$

erfüllt ist, wobei $\hat{\alpha}$ der MLE von α (gemäß Aufgabenteil (a)) ist. Beschreibe kurz deine Intuition zu dieser Wahl von b.

(c) Bestimme den Likelihood des Modells (0.1) und gib die Posterior-Verteilung (bis auf Konstanten) von ϑ an. Wähle ferner die Kandidatendichte

$$q(\alpha, \beta) = \pi_\alpha(\alpha | \hat{b})\varphi(\beta),$$

wobei $\hat{b}$ gemäß (0.2) gewählt wird und $\varphi(\beta)$ Normalverteilung mit (datenabhängiger) Erwartung $\hat{\beta}$ und Varianz $\sigma_{\hat{\beta}}^2$ ist ($\hat{\beta}$ bezeichnet den MLE von β, s. Aufgabenteil (a)). Ist q eine unabhängige Vorschlagsdichte? Gib eine Formel für die Akzeptanzwahrscheinlichkeit an.

(d) Implementiere nun den Metropolis-Hastings-Algorithmus. Erzeuge und analysiere Grafiken zu folgenden Fragestellungen:
- Verteilung von gesampelten α und β,
- Konvergenzverhalten der Erwartungen von α und β,
- durchschnittliche logistische Funktion und deren Variation,
- Verteilung der vorhergesagten Fehlerwahrscheinlichkeiten bei 45 und 65 Grad Fahrenheit.

Lösungen in einer zip-Datei per Mail bis zum 15. Februar, 23:59 Uhr, abgeben.