Angewandte Bayes-Statistik

Volker Schmid

LMU, Institut für Statistik

Sommersemester 2017

Zielsetzung

Für komplexere Modelle ist die Posteriori of nicht mehr analytisch zugänglich. Insbesondere ist die Normalisierungskonstante, d.h., die marginale Likelihood

$$p(x) = \int f(x|\theta)p(\theta)d\theta$$

schwer zu berechnen. Auswege:

- Numerische Integration
- Approximation der Posteriori
- Simulationsverfahren

Simulationsbasierte Posteriori-Inferenz

Idee: Erzeuge Ziehungen aus der Posteriori-Verteilung und approximiere daraus Statistiken der Posteriori-Verteilung

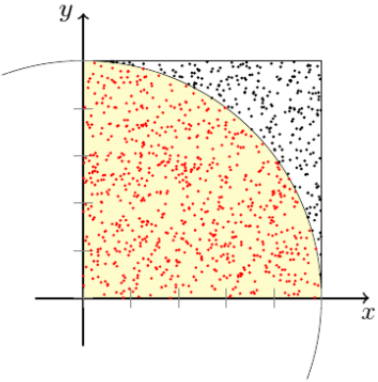
- Posteriori-Erwartungswert durch den Mittelwert
- Posteriori-Median über Median der Stichprobe
- Quantile der Posteriori-Verteilung über Quantile der Stichprobe
- HPD-Intervalle als kürzeste Intervalle, die $100(1-\alpha)\%$ der Stichprobe enthalten

Monte-Carlo-Integration

Definition 1 (Monte-Carlo-Integration).

Sei $f(x) > 0$ eine beliebige stetige Funktion mit bekanntem Wertebereich $[0, Y]$. $\int_a^b f(x)dx$ kann wie dann folgt approximiert werden.

- *Ziehe n gleichverteilte Zufallszahlen x aus $[a, b]$*
- *Ziehe unabhängig davon n gleichverteilte Zufallszahlen y aus $[0, Y]$*
- *Berechne den Anteil h der Punkte (x_i, y_i), die unterhalb der Funktion f liegen*
- $\int_a^b f(x)dx \approx h(b-a)Y$


y
x

Monte-Carlo-Schätzer

Ist $p(x)$ eine Dichte, so können Integrale der Form

$$E(g(x)) = \int g(x)p(x)dx$$

mit einer Stichprobe $x_1, \ldots, x_m$ aus $p(x)$ durch den Stichprobenmittelwert

$$\bar{g}_m = \frac{1}{m} \sum_{i=1}^{m} g\left(x_i\right)$$

approximiert werden. Aus dem starken Gesetz der grossen Zahlen folgt

$$\lim \frac{1}{m} \sum_{i=1}^{m} g(x_i) \to \int g(x)p(x)dx.$$

Monte-Carlo-Schätzer

Die Varianz des Monte-Carlo-Schätzers $\bar{g}_m$ ist gegeben durch

$$Var(\bar{g}_m) = \frac{1}{m} \int (g(X) - E(g(x)))^2 \, p(x) dx = \frac{1}{m} Var(g).$$

Der Approximationsfehler verringert sich mit steigendem m. Es folgt sogar aus dem zentralen Grenzwertsatz

$$\sqrt{m} \, (g_m - E(g(x))) \sim N(0, Var(g)).$$

Schätzer für $Var(\bar{g}_m)$ ist

$$\widehat{Var(\bar{g}_m)} = \frac{1}{m-1} \sum_{i=1}^{m} (g(x_i) - \bar{g}_m)^2$$

Inversionsmethode

Gegeben sei die Verteilungsfunktion $F(x)$ einer Zufallsvariablen X.
Sei $u \sim U[0,1]$. Dann ist

$$Y = F^{-1}(u) = \inf\{y : F(y) \geq u\} \sim X$$

Acception-Rejection-Methode

Ziel: Wir wollen aus einer Dichtefunktion $f(x)$ ziehen. Gegeben sei eine Dichtefunktion $g(x)$, nach der wir problemlos Zufallszahlen ziehen können. Es existiere ein c, so dass

$$cg(z) \geq f(z)$$

für alle z mit $f(z) > 0$. Dann können Zufallszahlen gemäß $f(x)$ wie folgt gezogen werden:

- ziehe Z gemäß $g(z)$
- akzeptiere Z mit Wahrscheinlichkeit $\frac{f(z)}{g(z)c}$

Squezed Acception-Rejection-Sampling

Gegeben sei eine untere Schranke $s(z) \leq f(z)$. Für u Ziehung aus $U[0,1]$ akzeptiere Z

- wenn $u \leq \frac{s(z)}{cg(z)}$
- wenn $u \leq \frac{f(z)}{cg(z)}$

Der zweite Schritt kann ausgelassen werden, wenn bereits im ersten Schritt akzeptiert wurde.

Markov Chain Monte Carlo

- Ziel: Ziehungen aus der Posterioriverteilung
- Simulationsbasierte Inferenz
- Funktioniert für komplexe und hochdimensionale Probleme
- Idee: Erzeuge eine Markovkette, deren stationäre Verteilung die Posterioriverteilung ist
- Ziehungen sind voneinander abhängig

Markov Chain Monte-Carlo-Methoden

Mit der Übergangsmatrix $\boldsymbol{P}$ einer irreduziblen, aperiodischen Markovkette, deren stationäre Verteilung $\boldsymbol{\pi}$ ist, können Zufallszahlen $Y \sim \pi$ wie folgt erzeugt werden:

- Wahl eines beliebigen Startwerts $y^{(0)}$
- Simulation der Realisierungen einer Markovkette der Länge m mit Übergangsmatrix $\boldsymbol{P}$: $(y^{(1)}, \ldots, y^{(m)})$

Ab einem gewissen Index b, dem *burn-in* geht der Einfluss der Startverteilung verloren und daher gilt approximativ

$$y^{(t)} \sim \pi, \text{ für } i = b, \ldots, m.$$

Die Ziehungen sind identisch verteilt, aber nicht unabhängig.

Gibbs-Sampling

Beim Gibbs-Sampling zieht man abwechselnd aus den Full Conditional Posteriors (bollständig bedingte Posterioriverteilung) der einzelnen Parameter(blöcke).

Beispiel 1 (Normalverteilung).

Modell:

$$x_i \sim N(\mu, \sigma^2)$$

Likelihood:

$$p(\boldsymbol{x}|\mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{\sigma^2}\sum(x_i - \mu)^2\right)$$

Semi-konjugierte Prioris:

$$\begin{aligned} \mu &\sim N(m_0, s_0) \\ \tau = \sigma^{-2} &\sim Ga(a, b) \\ \mu &\perp \tau \end{aligned}$$

Posteriori-Verteilung:

$$\begin{aligned} p(\mu,\tau|\boldsymbol{x}) &\propto \tau^{n/2}\exp\left(-\frac{\tau}{2}\sum(x_i-\mu)^2\right) \\ &\cdot \exp\left(-\frac{1}{2s_0}(\mu-m_0)^2\right)\tau^a\exp(-b\tau) \end{aligned}$$

Full conditional von μ:

$$p(\mu,|\boldsymbol{x},\tau) \propto \exp\left(-\frac{\tau}{2}\sum(x_i-\mu)^2 - \frac{1}{2s_0}(\mu-m_0)^2\right)$$

Es handelt sich um den Kern der $\mathrm{N}(s^{-1}m, s^{-1})$-Verteilung mit $m = \tau\sum x_i + m_0/s_0$ und $s = n\tau + s_0^{-1}$.
Posteriori-Verteilung:

$$\begin{aligned} p(\mu,\tau|\boldsymbol{x}) &\propto \tau^{n/2}\exp\left(-\frac{\tau}{2}\sum(x_i-\mu)^2\right) \\ &\cdot \exp\left(-\frac{1}{2s_0}(\mu-m_0)^2\right)\tau^a\exp(-b\tau) \end{aligned}$$

Full conditional von τ:

$$p(\tau|\boldsymbol{x},\mu) \propto \tau^{a+n/2} exp\left(-(b+0.5\sum(x_i-\mu)^2)\right)$$

Es handelt sich um den Kern der
$\text{Ga}(a+n/2, b+0.5\sum(x_i-\mu)^2)$-Verteilung.
Gibbs-Sampler:

1. Wähle Startwert τ_0
2. Ziehe $\mu \sim N(s^{-1}m, s^{-1})$
3. Ziehe $\tau \sim Ga(a+n/2, b+0.5\sum(x_i-\mu)^2)$
4. Iteriere 2 und 3 für $m = 1, \ldots, M$

Metropolis-Hastings-Algorithmus

- Ziehe θ^* aus einer Vorschlagsverteilung (Proposal) mit Dichte $q(\theta|\theta^{(k-1)})$
- Akzeptiere θ^* wird mit Wahrscheinlichkeit

$$\alpha = \min\left(1, \frac{p(\theta^*|x)q(\theta^{(k-1)}|\theta^*)}{p(\theta^{(k-1)}|x)q(\theta^*|\theta^{(k-1)})}\right)$$

- Andernfalls setze $\theta^{(k)} = \theta^{(k-1)}$.

Vorschlagsdichten

- Independence Proposal: Vorschlagsverteilung ist unabhängig vom aktuellen Wert
- Symetrisches Proposal: $q(\theta^*|\theta^{(k-1)}) = q(\theta^{(k-1)}|\theta^*)$, die Vorschlagsdichte kürzt sich aus der Akzeptanzwahrscheinlichkeit (Metropolis-Algorithmus):

$$\alpha = \frac{p(\theta^*|x)}{p(\theta^{k-1}|x)}$$

  Jeder Vorschlag mit $p(\theta^*|x) > p(\theta^{k-1}|x)$ wird angenommen!
- Random Walk Proposal: Vorschlagsverteilung ist ein Random Walk

$$\theta^* = \theta^{(k-1)} + \epsilon, \epsilon \sim f$$

  also $q(\theta^*|\theta^{(k-1)}) = f(\theta^* - \theta^{(k-1)})$.

Random Walk

Random Walk wird in der Regel mit Normalverteilung konstruiert:

$$\theta^* \sim N(\theta^{(k-1)}, C)$$

mit vorgegebener Kovarianzmatrix C.

- Eine zu kleine Varianz führt zu hohen Akzeptanzraten, aber ineffizienten, da stark autokorrelierten Ziehungen. Im Extremfall $C \to 0$ führt zu $\alpha = 1$, $\tau \to \infty$.
- Eine zu große Varianz führt zu zu großen Schritten, Vorschläge liegen in den Enden der Posterioriverteilung, sehr kleine Akzeptanzraten.
- Tuning der Kovarianzmatrix notwendig

Multivariate Verteilung

- Metropolis-Hastings-Algorithmus kann für θ-Vektoren durchgeführt werden
- Akzeptanzraten jedoch i.d.R. geringer mit höherer Dimension
- Alternative ist der komponentenweise Metropolis-Hastings: Jede Komponente des Parameters wird einzeln (skalar oder blockweise) aufdatiert. Sei $\theta = (\theta_1, \theta_2)$:

$$\alpha = \min\left(1, \frac{p(\theta_1^*|x, \theta_2^{(k-1)})q(\theta_1^{(k-1)}|\theta^*, \theta_2^{(k-1)})}{p(\theta_1^{(k-1)}|x, \theta_2^{(k-1)})q(\theta_1^*|\theta^{(k-1)}, \theta_2^{(k-1)})}\right)$$

- Updates können in fester oder zufälliger Weise erfolgen