

Algorytmy Statystyki Praktycznej

Błażej Miasojedow

MIMUW

3 kwietnia 2019

Hierarchiczny model bayesowski

Mamy k grup (np powiatów), w j -tej grupie mamy N_j obserwacji

$$X_{1j}, \dots, X_{N_j, j} \sim N(\mu_j, 1/\tau_j)$$

Następnie

$$\mu_1, \dots, \mu_k \sim N(\mu, 1/\tau)$$

$$\tau_1, \dots, \tau_k \sim \text{Gamma}(a, b)$$

μ i τ odpowiadają wartości oczekiwanej i wariancji badanej cechy w całym kraju. Są to z reguły wartości nieznane i na nie również nakładamy rozkłady a’priori

$$\mu \sim N(m, v^2)$$

$$\tau \sim \text{Gamma}(\alpha, \lambda)$$

Rozkład a’posteriori

$$\pi(\mu_1, \dots, \mu_k, \tau_1, \dots, \tau_k, \mu, \tau | X, a, b, \alpha, \lambda, m, v^2) \propto \prod_{j=1}^k \left\{ \prod_{i=1}^{N_j} [p(X_{ij} | \mu_j, \tau_j)] \pi(\mu_j | \mu, \tau) \pi(\tau_j | a, b) \right\} \pi(\mu | m, v^2) \pi(\tau | \alpha, \lambda)$$

Problemy obliczeniowe

Potrzebujemy metod pozwalających:

- Aproksymować/ generować z rozkładu $\pi(\theta|X)$ znanego z dokładnością do stałej.
- Obliczać całki względem rozkładu jak powyżej.

Proste Monte Carlo

Chcemy obliczyć

$$I = \int f(x)\pi(x)dx,$$

gdzie π gęstość prawdopodobieństwa. Czyli z MPWL

$$I = Ef(X) \approx \frac{1}{n} \sum_{i=1}^n f(X_i) =: \hat{I}_n,$$

gdzie $X_1, \dots, X_n \sim \text{i.i.d.}, \pi$

Własności estymatora MC:

- $E\hat{I}_n = I$
- Jeśli $\text{Var}(f(X_i)) = \sigma^2$ to

$$E(I - \hat{I}_n)^2 = \frac{\sigma^2}{n}$$

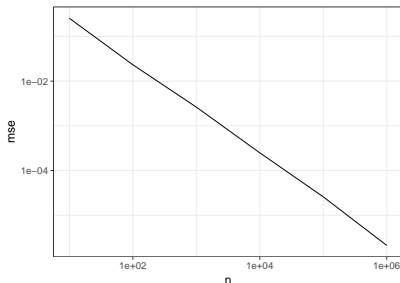
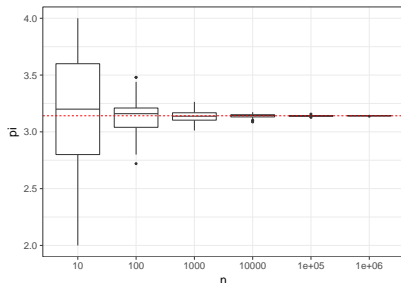
Przykład obliczenia liczby π

Jeśli $X \sim \text{Unif}([0, 1] \times [0, 1])$ to

$$\pi = 4E1(X_1^2 + X_2^2 \leq 1)$$

Jeśli $X^1, \dots, X^n \sim \text{Unif}([0, 1] \times [0, 1])$ to

$$\pi \approx \frac{1}{n} \sum_{i=1}^n 1((X_1^i)^2 + (X_2^i)^2 \leq 1)$$



Rejection Sampling I

Chcemy wylosować zmienną losową o gęstości f , ale potrafimy generować z g oraz wiemy, że

$$f(x) \leq M g(x)$$

Algorytm:

- 1 $Y \sim g$
- 2 $U \sim \text{Unif}([0, 1])$
- 3 Jeśli $U \leq \frac{f(Y)}{Mg(Y)}$ to zwracamy Y w.p.p wracamy do kroku 1

Poprawność: Niech $Z_i = (U_i, Y_i)$, $A = \{(y, u): u \leq \frac{f(y)}{Mg(y)}\}$,

$$N = \min\{n: Z_i \in A\}$$

Wówczas N i Z_N są niezależne oraz

$$P(Z_N \in B) = P(Z \in B | Z \in A)$$

Rejection Sampling II

Obliczmy najpierw

$$P(N = n) = P(Z_1 \notin A, \dots, Z_{n-1} \notin A, Z_n \in A) = q^{n-1}p,$$

gdzie $p = P(Z_1 \in A)$ oraz $q = 1 - p$ Teraz

$$\begin{aligned} P(N = n, Z_N \in B) &= P(Z_1 \notin A, \dots, Z_{n-1} \notin A, Z_n \in A \cap B) \\ &= P(Z_1 \notin A)^{n-1} P(Z_n \in B) \frac{P(Z_n \in A \cap B)}{P(Z_n \in B)} \end{aligned}$$

Teraz wystarczy obliczyć

$$P(Y \in B | Z \in A) = \frac{\int_B \int_0^{f(y)/Mg(y)} g(y) du dy}{\int \int_0^{f(y)/Mg(y)} g(y) du dy} = \frac{\int_B f(y) dy}{\int f(y) dy}$$

Oczekiwana liczba kroków algorytmu:

$$p = P(Z_1 \in A) = \int \frac{f(y)}{Mg(y)} g(y) dy = \frac{1}{M}$$

Stąd

$$EN = M$$

Przykład

Chcemy generować z X rozkładu gaussowskiego $N(0, 1)$ uciętego do odcinka $[3, 4]$ czyli

$$f(x) = \frac{\exp(-x^2/2)}{\int_3^4 \exp(-x^2/2)} 1(x \in [3, 4])$$

I korzystamy z dwóch gęstości pomocniczych g_1 gęstość $N(0, 1)$ i $g_2 \propto x \exp(-x^2/2) 1(x \in [3, 4])$ W pierwszym przypadku do wygenerowania próbki długości 100 potrzeba 73936 zmiennych z rozkładu g_1 natomiast dla g_2 wystarczyło 303 zmiennych.

Własności rejection sampling

- Jeśli potrafimy dobrze, przybliżyć **globalnie** gęstość to algorytm jest efektywny.
- Generuje próbkę i.i.d
- W przypadku użycia do obliczania całek “marnujemy” znaczną część wygenerowanych zmiennych losowych.

Importance sampling

Chcemy obliczyć

$$I = \int f(x)p(x)dx$$

p gęstość rozkładu. Dysponujemy próbką X_1, \dots, X_n i.i.d z rozkładu q . Zauważmy, że

$$I = \int f(x)p(x)dx = \int f(x)\frac{p(x)}{q(x)}q(x) = E_q f(X)w(X),$$

gdzie $w = \frac{p}{q}$.

$$I_{IS} = \frac{1}{n} \sum_{i=1}^n f(X_i)w(X_i)$$

Własności IS

- $E(I_{IS}) = I$
- $Var(I_{IS}) = \frac{Var_q(f(X)w(X))}{n}$
- Wymaga znajomości stałej normującej f

Samonormujący IS

Zauważmy, że

$$E_q(w(X)) = 1$$

Zastąpmy n przez estymator $\hat{n} = \sum_{i=1}^n w(X_i)$. Czyli

$$\bar{I}_{IS} = \frac{\sum_i f(X_i)w(X_i)}{\sum_{i=1}^n w(X_i)}$$

Teraz w występuje w liczniku i mianowniku.

- $E(\bar{I}_{IS}) \neq I$ ale $\lim_{n \rightarrow \infty} E\bar{I}_{IS} = I$
- $\text{Var}(\bar{I}_{IS}) = \mathcal{O}(1/n)$
- Można stosować do gęstości znanej z dokładnością do stałej normującej.

Efektywność IF

Proste MC

$$\text{MSE}(\hat{I}) = \frac{\sigma^2}{n}$$

Dla IS

$$\text{MSE}(\bar{I}_{\text{IS}}) = \frac{c}{n}$$

Effective Sample Size

Dla jakiej długości próbki i.i.d. dostaniemy ten sam błąd co nasz estymator. Dla IS ESS przybliża się następująco:

Traktujemy wagi $w_i = w(X_i)$ jako deterministyczne wówczas wariancja estymatora IS jest postaci

$$\frac{\sum w_i^2}{(\sum w_i)^2} \text{Var}_q(f(X_i))$$

I ESS definiujemy jako

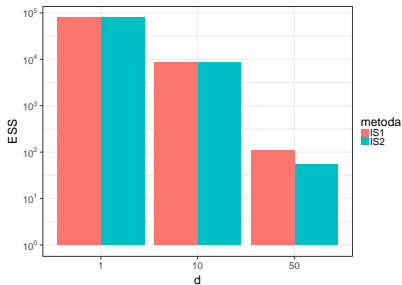
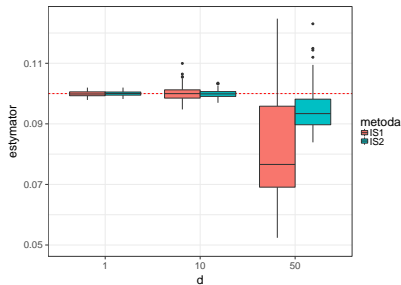
$$\text{ESS} = \frac{(\sum w_i)^2}{\sum w_i^2}$$

Przykład

$Y \sim N(0.1, 0.2^2 I_d)$ i chcemy obliczyć

$$I = \frac{1}{d} \sum_{i=1}^d EY_i = 0.1$$

Bedziemy korzystać z (samonormujący) IS z rozkładem propozycji $N(0, 0.2^2 I_d)$.



Własności IS

- Wykorzystuje całą próbkę do estymacji.
- Wrażliwa na dobór rozkładu pomocniczego. Trzeba dobrze przybliżyć rozkład globalnie.
- Nie wymaga znajomości stałej normującej

MCMC

Wadą standardowych MC jest to, że kolejne iteracje niezależą od poprzednich. Co powoduje, że każdy krok musi być dopasowany globalnie do rozkładu. Chcielibyśmy eksplorować rozkład lokalnie, np. w okolicy wyniku z poprzedniego kroku.

MCMC:

Chcemy wygenerować ergodyczny łańcuch Markowa X_1, \dots, X_n o rozkładzie stacjonarnym π . Wówczas na podstawie MPWL dla łańcuchów Markowa otrzymujemy

$$I = \int f(x)\pi(x)dx \approx \frac{1}{n} \sum_{i=1}^n f(X_i)$$

Z uwagi na to, że asymptotycznie mamy właściwy rozkład to stosuje się “burn-in” time

$$\hat{I}_{n_0, n} = \frac{1}{n} \sum_{i=n_0+1}^{n+n_0} f(X_i)$$

Rozkład stacjonarny i odwracalność I

Łańcuch Markowa z jądrem przejścia $P(x, dy)$ ma rozkład stacjonarny π wtedy i tylko wtedy gdy

$$\pi(A) = \int_X P(x, A)\pi(x)dx = \pi P.$$

Łańcuch Markowa z jądrem przejścia $P(x, dy)$ jest odwracalny względem π wtedy i tylko wtedy gdy

$$\int_{A \times B} P(x, dy)\pi(x)dx = \int_{A \times B} P(y, dx)\pi(y)dy$$

lub w skrócie

$$P(x, dy)\pi(dx) = P(y, dx)\pi(dy)$$

Rozkład stacjonarny i odwracalność II

Łańcuch Markowa jest odwracalny względem $\pi \implies \pi$ jest rozkładem stacjonarnym

$$\begin{aligned}\pi P(A) &= \int P(x, A) \pi(x) dx = \int \int_A \pi(x) P(x, dy) dx \\ &= \int_A \int P(y, dx) \pi(y) dy = \int_A \pi(y)\end{aligned}$$

Algorytm Metropolisa-Hastingsa

- $Q(x, dy)$ jądro przejścia o gęstości $q(x, y)$
- π rozkład docelowy

Algorytm MH

- 1 Generujemy $Y \sim Q(X_n, \cdot)$
- 2 Obliczamy prawdopodobieństwo akceptacji $\alpha(X_n, Y)$ gdzie

$$\alpha(x, y) = \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \wedge 1$$

3

$$X_{n+1} = \begin{cases} Y & \text{z prawdopodobieństwem } \alpha(X_n, Y) \\ X_n & \text{z prawdopodobieństwem } 1 - \alpha(X_n, Y) \end{cases}$$

Algorytm Metropolisa-Hastingsa

Algorytm MH generuje łańcuch Markowa odwracalny względem π Czyli

$$P_{MH}(x, y)\pi(x) = P_{MH}(y, x)\pi(y)$$

Jeśli $x \neq y$ to

$$P_{MH}(x, y) = \alpha(x, y)q(x, y)$$

Założmy, że $\alpha(x, y) < 1$ czyli $\alpha(x, y) = \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}$ to $\alpha(y, x) = 1$
(drugi przypadek symetryczny)

$$\begin{aligned} P_{MH}(x, y)\pi(x) &= \alpha(x, y)q(x, y)\pi(x) = \pi(y)q(y, x) \\ &= \pi(y)q(y, x)\alpha(y, x) = P_{MH}(y, x)\pi(y) \end{aligned}$$

Przykłady MH I

- Random Walk Metropolis $q(x, y) = q(|x - y|)$ np.
 $Q(x, dy) \sim N(x, \Sigma)$ wówczas

$$\alpha(x, y) = \frac{\pi(y)}{\pi(x)} \wedge 1$$

- Independent MH $q(x, y) = q(y)$

$$\alpha(x, y) = \frac{\pi(y)q(x)}{\pi(x)q(y)} \wedge 1$$

Przykłady MH II

- Metropolis Adjusted Langevin Algorithm

$$dX_t = \nabla \log(\pi(X_t))dt + \sqrt{2}dW_t$$

Dyskretyzacja Eulera- Maruyamy

$$X_{k+1} = X_k + \tau \nabla \log(\pi(X_k)) + \sqrt{2\tau}Z_k,$$

gdzie $Z_k \sim N(0, I)$.

MALA

$$Y = X_k + \tau \nabla \log(\pi(X_k)) + \sqrt{2\tau}Z_k$$

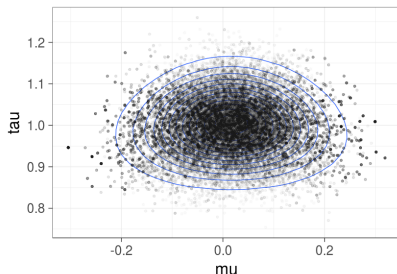
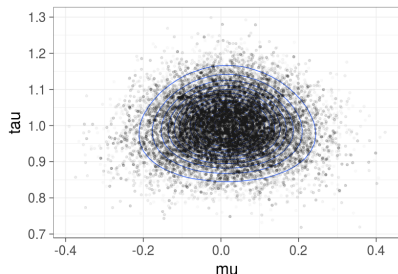
oraz

$$\alpha(x, y) = \frac{\pi(y)\phi(x; y + \tau \nabla \log(\pi(y)); 2\tau I)}{\pi(x)\phi(y; x + \tau \nabla \log(\pi(x)); 2\tau I)} \wedge 1$$

Przykład

X_1, \dots, X_n i.i.d $N(\mu, 1/\tau)$,

$$\pi(\mu) \sim N(0, 10), \quad \pi(\tau) \sim \text{Gamma}(1, 1)$$



	RWM	MALA
μ	-0.0013 (\pm 0.094)	0.0630 (\pm 0.150)
τ	1.0000 (\pm 0.070)	0.9700 (\pm 0.060)

Próbnik Gibbsa I

Mały krok próbnika Gibbsa:

$$X_i \sim \pi(X_i | X_{-i})$$

Oznaczmy przez $P_i(x, dy)$ jego jądro przejścia, jego gęstość

$$p_i(x, y) = \begin{cases} \pi(y_i | x_{-i}) & \text{jeśli } x_{-i} = y_{-i} \\ 0 & \text{w p.p.} \end{cases}$$

Systematic Scan

Wykonujemy kroki P_i w ustalonej kolejności czyli np.

$$X_1^n \sim \pi(\cdot | X_{-1}^{n-1})$$

$$X_2^n \sim \pi(\cdot | X_1^n, X_2^{n-1}, \dots, X_d^{n-1})$$

$$\vdots$$

$$X_k^n \sim \pi(\cdot | X_1^n, \dots, X_{k-1}^n, X_{k+1}^{n-1}, \dots, X_d^{n-1})$$

$$\vdots$$

Próbnik Gibbsa II

Random Scan

W każdym kroku losujemy współrzędną która ma być zmieniana

❶ Losujemy $j \sim \text{Unif}\{1, \dots, d\}$

❷

$$X_j^n \sim \pi(\cdot | X_{-j}^{n-1})$$

$$X_k^n = X_k^{n-1}, \quad \text{dla } k \neq j$$

Poprawność próbnika Gibbsa I

Mały krok próbnika Gibbsa jest odwracalny. Dla $x_{-i} = y_{-i}$

$$\begin{aligned} P_i(x, y)\pi(x) &= \pi(y_i|x_{-i})\pi(x_i, x_{-i}) \\ &= \frac{\pi(y_i, x_{-i})\pi(x_i|x_{-i})}{\pi(x_{-i})} \\ &= \frac{\pi(y_i, y_{-i})\pi(x_i|y_{-i})}{\pi(y_{-i})} \\ &= \pi(x_i|y_{-i})\pi(y_i, y_{-i}) = P_i(y, x)\pi(y) \end{aligned}$$

Poprawność próbnika Gibbsa II

Próbnik Gibbsa z losowym wyborem współrzędnych jest odwracalny.

$$P_{RS}(x, y) = \sum_{i=1}^d \frac{1}{n} P_i(x, y) \pi(x)$$

Każdy z P_i odwracalny to kombinacja wypukła jest odwracalna. Próbnik Gibbsa z deterministycznym wyborem współrzędnych jest nieodwracalny a zachowuje rozkład stacjonarny.

$$P_{SS}(x, y) =$$

$$P_1((x_1, \dots, x_d), (y_1, x_2, \dots, x_d)) \cdots$$

$$P_i((y_1, \dots, y_{i-1}, x_i, \dots, x_d), (y_1, \dots, y_i, x_{i+1}, \dots, x_d)) \cdots$$

$$P_d((y_d, \dots, y_{d-1}, x_d), (y_1, \dots, y_d))$$

Poprawność próbnika Gibbsa III

$$\begin{aligned}\pi P_{\text{SS}}(y) &= \int P_{\text{SS}}(x, y) \pi(x) dx \\ &= \int P_2 \cdots P_d \int P_1((x_1, \dots, x_d), (y_1, x_2, \dots, x_d)) \pi(x) dx_1 dx_{-1} \\ &= \int P_2((y_1, x_2, \dots, x_d), (y_1, y_2, \dots, x_d)) \cdots P_d((y_1, x_2, \dots, x_d)) dx_{-1} \\ &= \cdots = \pi(y)\end{aligned}$$

Przykład 1

$$\pi(\mu, \tau | X_1, \dots, X_n) \propto \tau^{a+n/2-1} \exp \left\{ -\tau \frac{\sum_{i=1}^n (X_i - \mu)^2}{2} - \frac{(\mu - m)^2}{2v^2} - \tau b \right\}$$

Pełne rozkłady warunkowe:

$$\pi(\mu | \tau, X_1, \dots, X_n) \sim N \left(\frac{n\bar{X}v^2 + m\tau^{-1}}{nv^2 + \tau^{-1}}, \frac{v^2\tau^{-1}}{nv^2 + \tau^{-1}} \right)$$

$$\pi(\tau | \mu, X_1, \dots, X_n) \sim \text{Gamma} \left(a + n/2, \frac{\sum_{i=1}^n (X_i - \mu)^2}{2} + b \right)$$

Próbnik Gibbsa dla modelu hierarchicznego

$$\pi(\mu_1, \dots, \mu_k, \tau_1, \dots, \tau_k, \mu, \tau | X, a, b, \alpha, \lambda, m, v^2) \propto \prod_{j=1}^k \left\{ \prod_{i=1}^{N_j} [p(X_{ij} | \mu_j, \tau_j)] \pi(\mu_j | \mu, \tau) \pi(\tau_j | a, b) \right\} \pi(\mu | m, v^2) \pi(\tau | \alpha, \lambda)$$

Pełne rozkłady warunkowe:

$$\pi(\mu_j | \tau_1, \dots, \tau_k, \tau, \mu, X_1, \dots, X_n) \sim N \left(\frac{N_j \bar{X}_j \tau^{-1} + \mu \tau_j^{-1}}{N_j \tau^{-1} + \tau_j^{-1}}, \frac{\tau^{-1} \tau_j^{-1}}{N_j \tau^{-1} + \tau_j^{-1}} \right)$$

$$\pi(\tau_j | \mu_1, \dots, \mu_k, \tau, \mu, X_1, \dots, X_n) \sim \text{Gamma} \left(a + N_j/2, \frac{\sum_{i=1}^{N_j} (X_{ij} - \mu_j)^2}{2} + b \right)$$

$$\pi(\mu | \tau_1, \dots, \tau_k, \mu_1, \dots, \mu_k, \tau, X_1, \dots, X_n) \sim N \left(\frac{k \bar{\mu} v^2 + m \tau^{-1}}{k v^2 + \tau^{-1}}, \frac{v^2 \tau^{-1}}{k v^2 + \tau^{-1}} \right)$$

$$\pi(\tau | \tau_1, \dots, \tau_k, \mu_1, \dots, \mu_k, \mu, X_1, \dots, X_n) \sim \text{Gamma} \left(\alpha + k/2, \frac{\sum_{i=1}^k (\mu_i - \mu)^2}{2} + \lambda \right)$$

Metropolis wewnątrz Gibbsa

Próbnik Gibbsa składa się z małych kroków postaci

$$P_i(\mathbf{x}, d\mathbf{y}) = \pi(d\mathbf{y}_i | \mathbf{x}_{-i}).$$

Może się zdarzyć, że generowanie z jednego (wielu) rozkładów warunkowych $\pi(\cdot | \mathbf{x}_{-i})$ jest trudne. Wówczas krok P_i można zastąpić krokiem Metropolisa Hastingsa o rozkładzie docelowym $\pi(\cdot | \mathbf{x}_{-i})$.

Potrzebujemy następujących własności łańcuchów Markowa

- Ergodyczność $P(X_n \in A | X_1 = x) \rightarrow \pi(A)$
- Prawo wielkich liczb $\hat{I}_n = \frac{1}{n} \sum_{i=1}^n f(X_i) \rightarrow \pi f$
- Centralne twierdzenie graniczne $\sqrt{n}(\hat{I}_n - \pi f) \rightarrow N(0, \sigma^2)$

Ergodyczność I

Definicja

Dla miar probabilistycznych μ, ν odległością pełnego wahania nazywamy

$$\|\mu - \nu\|_{\text{tv}} = \sup_A |\mu(A) - \nu(A)|$$

Jeśli $X \sim \mu$, $Y \sim \nu$ to

$$\|\mu - \nu\|_{\text{tv}} \leq \mathbb{P}(X \neq Y)$$

TWIERDZENIE

Jeśli łańcuch Markowa o rozkładzie stacjonarnym π jest π nieprzywiedlny i nieokresowy to

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{\text{tv}} \rightarrow 0$$

TWIERDZENIE

Jeśli πf istnieje oraz $\{X_i\}$ jest π -nieprzywiedlny to

$$\frac{\sum_{k=1}^n f(X_i)}{n} \rightarrow \pi f, \quad \text{p.n.}$$

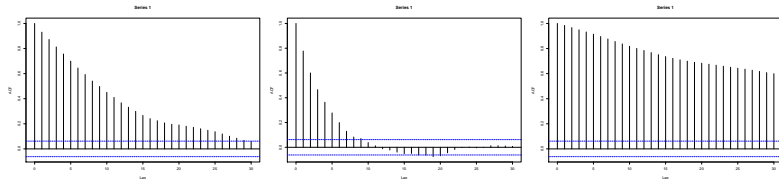
Centralne twierdzenie graniczne I

Przy pewnych założeniach o regularności

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n f(X_i) - \pi f \right) \rightarrow N(0, \sigma^2)$$

$$\sigma^2 = \text{Var}_{\pi}(f(X)) + \sum_{k=2}^{\infty} \text{cov}_{\pi}(X_1, X_k)$$

Stąd ważna w diagnostyce MCMC jest funkcja autokorelacji.



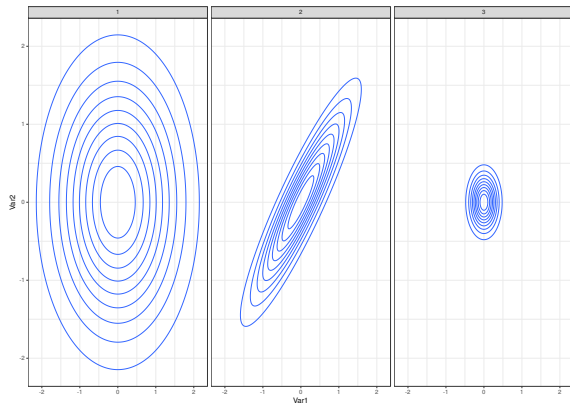
Problemy z algorytmem Metropolisa

RWM:

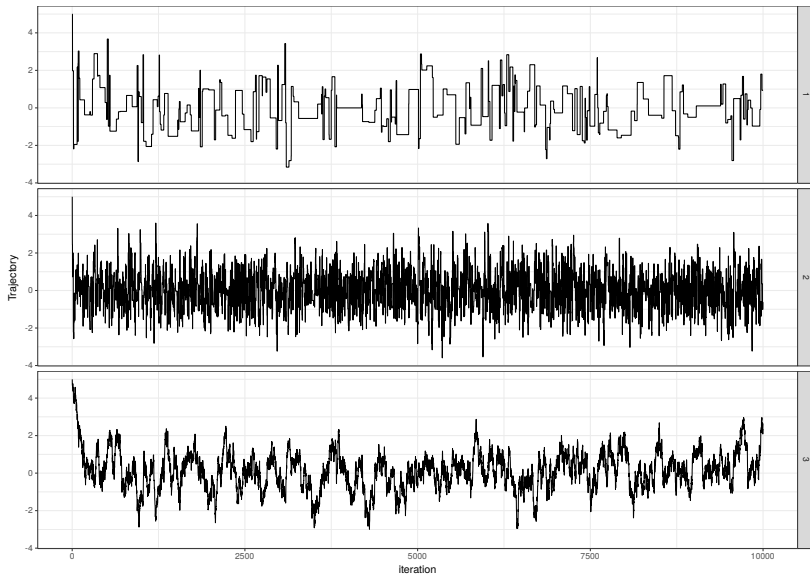
Propozycja $Y \sim N(X_n, \Sigma)$

Jak wybrać Σ ?

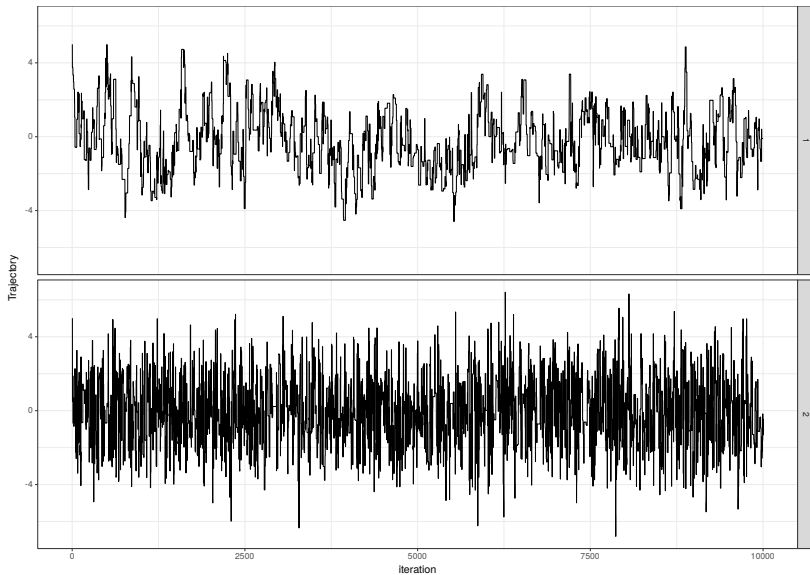
- Wielkość.
- Kształt.



Wielkość



Kształt



“Optymalny” wybór Σ

- Optymalana $\Sigma = \ell^2 \text{Cov}_\pi(X)$, gdzie ℓ^2 można zdefiniować na dwa sposoby.
- $\ell^2 \approx \frac{2.38^2}{d}$, gdzie d wymiar.
- ℓ^2 takie, że średnie prawdopodobieństwo akceptacji wynosi 0.234

Adaptacyjny RWM

Chcemy skonstruować algorytm, który automatycznie znajduje optymalną macierz Σ . W tym celu w każdym kroku algorytmu będziemy estymować macierz kowariancji oraz odpowiednio ją skalować. niech

$$\mu_n = \frac{n-1}{n}\mu_{n-1} + \frac{1}{n}X_n$$

$$\Gamma_n = \frac{n-1}{n}\Gamma_{n-1} + \frac{1}{n}(X_n - \mu_n)(X_n - \mu_n)^T$$

Wersja I:

$$\Sigma_n = \frac{2.38^2}{d}\Gamma_n + \epsilon I$$

Wersja II:

$$T_n = T_{n-1} + \frac{1}{n}(\alpha(X_n, Y_n) - 0.234)$$

$$\Sigma_n = \exp(T_n)\Gamma_n + \epsilon I$$

Uwagi o implementacji adaptacyjnego RWM I

Generowanie wielowymiarowego rozkładu gaussowskiego

Jeśli $X = (X_1, \dots, X_d)$ i.i.d $N(0,1)$ oraz $\Sigma = LL^T$ to

$$LX \sim N(0, \Sigma)$$

Typowo używa się rozkładu Choleskiego. Dla którego istnieje one-rank-update Jeśli

$$M_{n+1} = M_n + xx^T$$

To odpowiadające im rozkłady Choleskiego

$$L_{n+1} = \text{ORU}(L_n, x).$$

W adaptacyjnym RWM wystarczy pamiętać L_n rozkład Choleskiego Γ_n i

$$L_{n+1} = \text{ORU}\left(\frac{\sqrt{n}}{\sqrt{n+1}}L_n, \frac{1}{\sqrt{n+1}}(X_{n+1} - \mu_{n+1})\right)$$

Uwagi o implementacji adaptacyjnego RWM II

Generowanie propozycji:

$$Y_n = X_n + \exp(T_n/2)L_n Z_n + \sqrt{\epsilon}Z'_n$$

gdzie Z_n i Z'_n niezależne $N(0, I)$.

Układ Hamiltonowski

$$H(p, x) = K(p) + U(x)$$

K - energia kinetyczna, U- energia potencjalna Układ Hamiltonowski

$$\begin{aligned}\frac{dx_i}{dt} &= \frac{\partial H}{\partial p_i} \\ \frac{dp_i}{dt} &= -\frac{\partial H}{\partial x_i}\end{aligned}$$

Rozwiązanie zachowuje energię.

Związek z rozkładami prawdopodobieństwa

Rozkład Gibbsa

$$\pi(p, x) \propto \exp(-H(p, x))$$

Rozkład brzegowy dla dowolnego K

$$\pi(x) \propto \exp(-U(x))$$

Rozkład warunkowych

$$\pi(p|x) \propto \exp(-K(p)) = \pi(p)$$

Idealne Hamiltonowskie Monte Carlo

$$X_n \rightarrow X_{n+1}$$

- Wylosuj $P \sim \pi(p)$.
- Zgodnie z Hamiltonowską dynamiką

$$(P, X_n) = (x_0, p_0) \rightarrow (x_T, p_T) = (X_{n+1}, P')$$

Potrzeba K zadające rozkład łatwy do generowania, oraz
potrzeba umieć obliczyć rozwiązanie układu hamiltonowskiego
w czasie T .

Hamiltonian MCMC I

$$K(p) = \frac{1}{2} p^T p$$

Czyli $\pi(p) = N(0, I)$. Dodatkowo potrzeba numerycznie rozwiązywać układ hamiltonowski.

Leapfrog integrator:

Dla $1 \leq n \leq T/\epsilon$

$$p_{n+1/2} = p_n - \frac{\epsilon}{2} \frac{\partial U}{\partial x}(x_n)$$

$$x_{n+1} = x_n + \epsilon p_{n+1/2}$$

$$p_{n+1} = p_{n+1/2} - \frac{\epsilon}{2} \frac{\partial U}{\partial x}(x_{n+1})$$

Hamiltonian MCMC II

Jeżeli zamienimy znak $p_T \rightarrow -p_T$ to leapfrog integrator jest odwracalny. I żeby usunąć błąd numerycznego rozwiązania wystarczy dodać akceptację Metropolisa

$$\alpha((x_0, p_0), (x_T, -p_T)) = \exp(H(x_T, p_T) - H(x_0, p_0)) \wedge 1$$

Hamiltonian MCMC III

Algorytm

- 1 Generujemy $p \sim N(0, I)$
- 2 Przybliżamy rozwiązanie układu hamiltonowskiego (leapfrog integrator) $X_n = x_0, p = p_0 \rightarrow x_T, p_T$
- 3

$$X_{n+1} = \begin{cases} x_T & \text{z prawdopodobieństwem } \alpha((x_0, p_0), (x_T, -p_T)) \\ X_n & \text{z prawdopodobieństwem } 1 - \alpha((x_0, p_0), (x_T, -p_T)) \end{cases}$$