WSTĘP DO MATEMATYKI FINANSOWEJ

BARTOSZ KOŁODZIEJEK

WYDZIAŁ MATEMATYKI I NAUK INFORMACYJNYCH

Laboratoria 7

Projekt „NERW 2 PW. Nauka - Edukacja - Rozwój - Współpraca" współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.

Zadanie 10 pn. „Modyfikacja programów studiów na kierunkach prowadzonych przez Wydział Matematyki i Nauk Informacyjnych", realizowane w ramach projektu „NERW 2 PW. Nauka - Edukacja - Rozwój - Współpraca", współfinansowanego jest ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.

6 czerwca 2023

0.1. **L7 - Metody MCMC.** Ćwiczymy metody MCMC (Markov Chain Monte Carlo). Będziemy całkowali metodami Monte Carlo z wykorzystaniem łańcuchów Markowa. Metody te opierają się o twierdzenia ergodyczne, a nie o prawa wielkich liczb jak tradycyjne metody Monte Carlo.

(1) Przypomnienie o Łańcuchach Markowa (ŁM):
Niech $\mathcal{S}$ będzie przeliczalnym zbiorem - będziemy go nazywali przestrzenią stanów.

- Ciąg zmiennych losowych $(X_n)_{n=0}^{\infty}$ o wartościach w $\mathcal{S}$ nazywamy łańcuchem Markowa, jeśli dla każdego $n \in \mathbb{N}$ i każdego ciągu $s_0, s_1, \ldots, s_n \in \mathcal{S}$ zachodzi
$$\mathbb{P}(X_n = s_n | X_{n-1} = s_{n-1}, \ldots, X_1 = s_1, X_0 = s_0) = \mathbb{P}(X_n = s_n | X_{n-1} = s_{n-1})$$
o ile powyższe prawdopodobieństwa mają sens (czyli zdarzenia w warunkowaniach mają dodatnie prawdopobieństwa).
- Łańcuch Markowa nazywamy jednorodnym (JŁM), gdy
$$\mathbb{P}(X_n = s_n | X_{n-1} = s_{n-1}) = \mathbb{P}(X_1 = s_n | X_0 = s_{n-1}) \qquad \text{dla każdego } n.$$
Dalej będziemy rozważali tylko jednorodne ŁM.
- Oznaczmy $p_{ij} = \mathbb{P}(X_1 = s_j | X_0 = s_i)$. Będziemy utożsamiali stan s_i z jego indeksem i. Wtedy, macierz $P = (p_{ij})_{i,j\in\mathcal{S}}$ nazywamy macierzą przejścią (w jednym kroku).
Macierz P jest macierza stochastyczną, tzn. jej wyrazy są nieujemne, a jej wiersze sumują się do 1.
Rozkład JŁM $(X_n)_{n=0}^{\infty}$ jest jednoznacznie określony przez macierz P oraz rozkład początkowy (rozkład X_0).
- Definiujemy również macierz przejścia w n krokach poprzez
$$P(n) = (p_{ij}(n))_{i,j\in\mathcal{S}} := P^n.$$
- Mówimy, że
  - stan s_k jest osiągalny ze stanu s_j (ozn.: $s_j \to s_k$), jeśli $p_{jk}(n) > 0$ dla pewnego n,
  - stany s_k i s_j komunikują się ($s_k \leftrightarrow s_j$), jeśli $s_j \to s_k$ oraz $s_k \to s_j$.
- JŁM nazywamy nieprzywiedlnym, jeśli wszystkie jego stany wzajemnie się komunikują.
- Stan s_j nazywamy stanem powracającym, jeśli JŁM startujący ze stanu s_j ($X_0 = s_j$), powróci do niego z prawdopodobieństwem 1 w skończonym (aczkolwiek dowolnie długim) czasie:
$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} \{X_n = s_j\} | X_0 = s_j\right) = 1.$$
W przeciwnym przypadku (gdy powyżej jest < 1), stan taki nazywamy stanem chwilowym.
⚠ Twierdzenie Lévy'ego: Błądzenie losowe w $\mathbb{Z}^d$ jest powracające wtedy i tylko wtedy, gdy $d \in \{1, 2\}$.
- Okresem stanu j nazywamy liczbę $o(j) = \mathrm{NWD}\{n \colon p_{jj}(n) > 0\}$.
- **TW.** W nieprzywiedlnym JŁM wszystkie stany są tego samego typu: jeżeli jeden jest powracający (odp. chwilowy, okresowy o okresie d), to wszystkie jego stany są powracające (chwilowe, okresowe o okresie d).
👁 Z uwagi na powyższy wynik, dla nieprzywiedlnych JŁM, jest sens mówić o łańcuchach powracających, chwilowych i okresowych. Mówimy, że JŁM jest okresowy jeśli wszystkie jego stany sa okresowe o okresie $d > 1$ (w przeciwnym przypadku, JŁM nazwiemy nieokresowym).
- Mówimy, że rozkład prawdopodobieństwa $\pi = (\pi_j)_{j\in\mathcal{S}}$ jest rozkładem stacjonarnym dla JŁM o macierzy przejścia P, jeśli (wektor π tutaj jest poziomy!)
$$\pi = \pi \cdot P,$$
czyli $\sum_{j\in\mathcal{S}} \pi_j = 1$, $\pi_j \geq 0$ oraz $\pi_j = \sum_{i\in\mathcal{S}} \pi_i p_{ij}$.
⚠ Jeśli X_0 ma rozkład stacjonarny π, to dla każdego n
$$X_n \stackrel{d}{=} X_0 \sim \pi.$$
- **TW.** (Twierdzenie ergodyczne). Niech $(X_n)_{n\geq 0}$ będzie nieprzywiedlnym i nieokresowym łańcuchem Markowa na skończonej przestrzeni stanów $\mathcal{S}$ z macierzą przejścia P. Wtedy:
  (a) istnieje dokładnie jeden rozkład stacjonarny π dla macierzy P,
  (b) $\pi_i > 0$ dla $i \in \mathcal{S}$,

(c) dla każdego $i \in \mathcal{S}$,

$$p_{ij}(n) \stackrel{n\to\infty}{\longrightarrow} \pi_j, \qquad j \in \mathcal{S}$$

oraz dla dowolnego rozkładu X_0 zachodzi

$$\mathbb{P}(X_n = j) \stackrel{n\to\infty}{\longrightarrow} \pi_j, \qquad j \in \mathcal{S}$$

(d) dla dowolnego X_0 zachodzi

$$\boxed{\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_k=j\}} \stackrel{1}{\longrightarrow} \pi_j, \qquad j \in \mathcal{S}.}$$

⚠ Zauważmy, że warunek (d) implikuje dla dowolnej funkcji f na $\mathcal{S}$,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) = \frac{1}{n}\sum_{k=1}^{n} f(X_k)\sum_{i\in\mathcal{S}} \mathbb{1}_{\{X_k=i\}} = \sum_{i\in\mathcal{S}} f(i)\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_k=i\}} \stackrel{1}{\longrightarrow} \sum_{i\in\mathcal{S}} f(i)\pi_i = \mathbb{E}_\pi[f(X)].$$

⚠ Warunek (d) przypomina Mocne Prawo Wielkich Liczb, ale faktycznie jest to niezależne twierdzenie. Zmienne $(X_n)_n$ są bardzo zależne!

- Szybkość zbieżności możemy próbować oceniać za pomocą CTG w wersji dla łancuchów Markowa (pamiętajmy jednak, że w zastosowaniach rzadko X_0 ma rozkład stacjonarny, a jedynie wiemy, że rozkład X_n dąży do stacjonarnego - bywa że bardzo szybko).

  TW. (CTG dla ŁM) Jeśli rozkład X_0 jest stacjonarny, to

$$\frac{1}{\sqrt{n}}\left(\sum_{k=1}^{n}(f(X_k) - \mathbb{E}_\pi[f(X)])\right) \stackrel{d}{\longrightarrow} Z \sim \mathrm{N}(0, \sigma_f^2),$$

  gdzie $\sigma_f^2 = \mathrm{Var}[f(X_0)] + 2\sum_{k=1}^{\infty} \mathrm{Cov}[f(X_0), f(X_k)]$.

- 📘 JŁM nazywamy odwracalnym (*reversible*), jeśli istnieje taki rozkład prawdopodobieństwa π na $\mathcal{S}$, że zachodzi

$$\pi_i p_{ij} = \pi_j p_{ji}, \qquad \forall\, i, j \in \mathcal{S}.$$

  ⚠ Jeśli JŁM jest odwracalny, to π z powyższego warunku jest jego rozkładem stacjonarnym:

$$\pi_j = \pi_j \cdot 1 = \pi_j \cdot \sum_{i\in\mathcal{S}} p_{ji} = \sum_{i\in\mathcal{S}} \pi_i p_{ij}.$$

  ⚠ Odwracalne ŁM to szczególnie ładne ŁM, ponieważ łatwo dla nich stwierdzić, że coś jest rozkładem stacjonarnym.

- ⚠ Powyżej rozważaliśmy ŁM o dyskretnej przestrzeni stanów, ale wszystko można przenieść na ogólniejsze przestrzenie $\mathcal{S}$ (w szczególności, gdy $\mathcal{S}$ jest zbiorem nieprzeliczalnym). Wtedy wiele obiektów trzeba przedefiniować, poniżej rozważymy (tylko) miary absolutnie ciągłe względem miary Lebesgue'a oraz $\mathcal{S} = \mathbb{R}^d$.

  Po pierwsze, ŁM nazywamy proces stochastyczny, którego gęstości warunkowe spełniają warunek

$$f_{X_n|X_{n-1}=s_{n-1},\dots,X_1=s_1,X_0=s_0}(s_n) = f_{X_n|X_{n-1}=s_{n-1}}(s_n), \qquad s_i \in \mathcal{S},\ i = 0,\dots,n.$$

  Taki proces nazywamy jednorodnym, jeśli

$$f_{X_n|X_{n-1}=s_{n-1}}(s_n) = f_{X_1|X_0=s_{n-1}}(s_n) \qquad \text{dla każdego } n.$$

  Nie istnieje coś takiego jak macierz przejścia $(p_{ij})_{ij}$ w jednym kroku, ponieważ stany są "ciągłe". Zamiast tego definiuje się gęstość prawdopodobieństwa przejścia w jednym kroku, czyli gęstość warunkową X_1 pod warunkiem X_0,

$$p(y \mid x) := f_{X_1|X_0=x}(y).$$

  Zatem $y \mapsto p(y \mid x)$ jest gęstością dla każdego $x \in \mathcal{S}$. Wtedy rozkładem stacjonarnym π (a właściwie gęstością tego rozkładu) ŁM o prawdopodobieństwie przejścia zadanym jak powyżej to funkcja spełniająca dla $y \in \mathcal{S}$

$$\pi(y) = \int_{\mathcal{S}} p(y \mid x)\pi(x)dx.$$

  Jeśli ŁM jest odwracalny, tzn. dla pewnej gęstości π zachodzi

$$\pi(x)p(y \mid x) = \pi(y)p(x \mid y), \qquad \forall x, y \in \mathcal{S},$$

  to π z warunku powyżej jest gęstością rozkładu stacjonarnego.

Istnieje wersja Twierdzenia ergodycznego dla takich ŁM. Warunek (d) z powyższego twierdzenia jest zastępowany przez

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{1} \int_{\mathcal{S}} f(x)\pi(x)dx = \mathbb{E}_\pi[f(X)]$$

dla "ładnych" funkcji f (np. ciągłych i ograniczonych, ale wystarczy po prostu całkowalnych względem π).

(2) Problem: załóżmy, że π jest pewnych rozkładem prawdopodobieństwa na $\mathcal{S}$. Celem jest estymacja wartości oczekiwanej z pewnej funkcji $h\colon \mathcal{S} \to \mathbb{R}$, czyli

$$\mathbb{E}_\pi[h(X)] = \sum_{s\in\mathcal{S}} h(s)\pi_s \qquad \left(\mathbb{E}_\pi[h(X)] = \int_{\mathcal{S}} h(s)\pi(s)ds\right).$$

Jeśli mamy doczynienia z sytuacją, gdy łatwo jest generować z rozkładu π, to możemy korzystać ze zwykłej metody Monte Carlo.

Metody MCMC pomagają w sytuacji, gdy

- rozkład π jest znany tylko z dokładnością do stałej normującej, tzn. znamy funkcję $f\colon \mathcal{S} \to \mathbb{R}_+$ taką, że

  $$\pi_i = c\, f(i) \qquad (\pi(x) = c\, f(x)),$$

  ale stała c nie jest znana.

Oczywiście, $c = 1/(\sum_{i\in\mathcal{S}} f(i))$ ($c = 1/\int_{\mathcal{S}} f(x)dx$), ale w zastosowaniach zbiór $\mathcal{S}$ bywa tak duży (przy dużej kosztochłonności obliczania f), że znalezienie tej sumy może nie być możliwe w interesującym nas horyzoncie czasowym.

(3) Niech $f\colon \mathcal{S} \to \mathbb{R}_+$ będzie funkcją jak powyżej. Żeby zrealizować nasz cel, czyli estymować $\mathbb{E}_\pi[h(X)]$, gdzie

$$\pi := (c\, f(i)\colon i \in \mathcal{S}),$$

skonstruujemy JŁM o rozkładzie stacjonarnym danym przez π.

⚠ Z konstrukcji tego ŁM będzie wynikało, że wystarczy, nam wiedza o f, a znajomość c oraz π nie jest konieczna. Taki typ problemów spotykany jest bardzo naturalnie podczas uprawiania statystyki Bayesowskiej, o czym później.

Niech $Q = (q_{ij})_{i,j\in\mathcal{S}}$ będzie (na tym etapie dowolną, ale myślmy już że nieprzywiedlną) macierzą przejścia na przestrzeni stanów $\mathcal{S}$. Za pomocą macierzy Q skonstruujemy macierz P dla której π będzie rozkładem stacjonarnym.

👁 Uwaga: w literaturze zamiast q_{ij} można spotkać oznaczenie $q(j|i)$.

Zdefiniujmy

$$\rho_{ij} = \frac{\pi_j\, q_{ji}}{\pi_i\, q_{ij}} = \frac{f(j)\, q_{ji}}{f(i)\, q_{ij}} \qquad \text{oraz} \qquad \alpha_{ij} = \min\{\rho_{ij}, 1\}$$

oraz rozważmy macierz $P = (p_{ij})_{i,j\in\mathcal{S}}$ zadaną przez

$$p_{ij} = \begin{cases} q_{ij}\alpha_{ij}, & \text{jeśli } i \neq j, \\ 1 - \sum_{j\neq i} q_{ij}\alpha_{ij}, & \text{jeśli } i = j. \end{cases}.$$

Macierz P jest oczywiście macierzą stochastyczną. Ponadto mamy dla $i \neq j$,

$$\pi_i p_{ij} = \begin{cases} \pi_j\, q_{ji}, & \rho_{ij} < 1, \\ \pi_i\, q_{ij}, & \rho_{ij} \geq 1. \end{cases}$$

oraz

$$\pi_j p_{ji} = \begin{cases} \pi_i\, q_{ij}, & \rho_{ji} < 1, \\ \pi_j\, q_{ji}, & \rho_{ji} \geq 1. \end{cases}$$

Żeby pokazać odwracalność ŁM o macierzy przejścia P (czyli $\pi_i p_{ij} = \pi_j p_{ji}$) wystarczy zauważyć, że $\rho_{ij} = 1/\rho_{ji}$. To oznacza z kolei, że π jest rozkładem stacjonarnym.

Powstaje ważne pytanie, w jaki sposób generować ŁM o macierzy przejścia P? Okazuje się to być bardzo łatwe, ale o tym następnym razem.

0.2. **L7 - Metody MCMC - ciąg dalszy.**

(1) Przypomnijmy, że celem jest estymacja wielkości

$$\mathbb{E}_\pi[h(X)] = \begin{cases} \sum_{s\in\mathcal{S}} h(s)\pi_s, \\ \int_{\mathcal{S}} h(x)\pi(s)ds, \end{cases}$$

gdzie π jest pewnym rozkładem prawdopodobieństwa.

(2) Pomysł: skonstruujemy JŁM $(X_n)_n$ o rozkładzie stacjonarnym π i skorzystamy z twierdzenia ergodycznego:

$$\frac{1}{n}\sum_{k=1}^{n} h(X_k) \xrightarrow{1} \mathbb{E}_\pi[h(X)].$$

(3) Z konstrukcji tego JŁM wynika, że wystarczy nam znajomość π z dokładnością do stałej, czyli znamy funkcję f, taką że $\pi = c \cdot f$, ale nie znamy stałej c. Tym samym, dla zadanej funkcji f (całkowalnej, nieujemnej) będziemy estymować

$$\mathbb{E}_\pi[h(X)] = \begin{cases} \frac{\sum_{s\in\mathcal{S}} h(s)f(s)}{\sum_{s\in\mathcal{S}} f(s)}, \\ \frac{\int_{\mathcal{S}} h(x)f(s)ds}{\int_{\mathcal{S}} f(s)ds}. \end{cases}$$

(4) Rozważmy najpierw przypadek dyskretny. Niech $Q = (q_{ij})_{i,j\in\mathcal{S}}$ będzie nieprzywiedlną macierzą przejścia na przestrzeni stanów $\mathcal{S}$. Pokazaliśmy już wcześniej, że macierz $P = (p_{ij})_{i,j\in\mathcal{S}}$ zdefiniowana poniżej jest
- macierzą stochastyczną,
- JŁM o tej macierzy jest odwracalny, tzn. $\pi_i p_{ij} = \pi_j p_{ji}$,
- rozkładem stacjonarnym tego łańcucha jest π.

Niech

$$\rho_{ij} = \frac{\pi_j\, q_{ji}}{\pi_i\, q_{ij}} = \frac{f(j)\, q_{ji}}{f(i)\, q_{ij}} \qquad \text{oraz} \qquad \alpha_{ij} = \min\{\rho_{ij}, 1\}$$

oraz $P = (p_{ij})_{i,j\in\mathcal{S}}$, gdzie

$$p_{ij} = \begin{cases} q_{ij}\alpha_{ij}, & \text{jeśli } i \neq j, \\ 1 - \sum_{j\neq i} q_{ij}\alpha_{ij}, & \text{jeśli } i = j. \end{cases}.$$

(5) Powstaje ważne pytanie, w jaki sposób generować ŁM o macierzy przejścia P? Okazuje się to być bardzo łatwe: odpowiednio modyfikujemy JŁM o macierzy Q. Poniższy algorytm nazywa się *algorytmem Metropolisa-Hastingsa*:

Algorytm 0.1 (Algorytm Metropolisa-Hastingsa)
Zainicjalizuj $X_0 \in \mathcal{S}$. [Dowolnie jak, byleby przyszłość była niezależna od tego wyboru.]
Dla chwil $t = 1, 2, \ldots$ wykonuj następujące kroki:
(i) w chwili t wygeneruj stan Y zgodnie z rozkładem $(q_{X_{t-1},k} \colon k \in \mathcal{S})$, [Innymi słowy, będąc w stanie X_{t-1}, zrób krok ŁM o macierzy przejścia Q, nowy stan nazwij Y],
(ii) Zaakceptuj stan Y z prawdopodobieństwem $\alpha_{X_{t-1},Y}$ i wtedy połóż $X_t = Y$. W przypadku braku akceptacji nie zmieniaj stanu, tzn. połóż $X_t = X_{t-1}$. Innymi słowy,

$$X_t = \begin{cases} Y, & \textit{z prawdopodobieństwem } \alpha_{X_{t-1},Y}, \\ X_{t-1}, & \textit{w p.p.} \end{cases}$$

O ile dla tak skonstruowanego ŁM $(X_n)_n$ zachodzi Twierdzenie ergodycznego , to wiemy, że dla dowolnej funkcji h na $\mathcal{S}$ mamy

$$\frac{1}{n}\sum_{k=1}^{n} h(X_k) \xrightarrow{1} \sum_{i\in\mathcal{S}} h(i)\pi_i.$$

Zauważmy, że jeśli JŁM generowany przez macierz Q był nieprzywiedlny, to skonstruowany ŁM $(X_n)_n$ będzie nieprzywiedlny i nieokresowy.

Macierz przejścia Q (a właściwie jej wiersze) nazywana jest *proposal distribution.*
Komentarze:
- rozważmy symetryczny rozkład proposal distribution, tzn. $q_{ij} = q_{ji}$. Powiedzmy, że nasz ŁM jest w stanie i oraz wylosowaliśmy stan j zgodnie z Q. Stan ten jest akceptowany z prawdopodobieństwem $\alpha_{ij} = \min\{\frac{f(j)}{f(i)}, 1\}$. W szczególności, jeśli $f(j) \geq f(i)$, to stan j zostanie automatycznie akceptowany.

Skonstruowany ŁM chętnie przebywa dłużej w stanach, gdzie funkcja f jest duża. W pewnym sensie, jest tutaj podobieństwo do importance sampling, przy czym tutaj kolejne wyrazy ciągu są zależne.

- Jak dobrać proposal distribution? Mamy tu pewną dowolność.
  W idealnym świecie byśmy wybrali $Q_{ij} = \pi_j$, dzięki czemu nowe stany zawsze by były akceptowane ($\alpha_{ij} = 1$), a ciąg $(X_n)_n$ byłby i.i.d. Żeby zobaczyć niezależność rozważmy

$$\mathbb{P}(X_n = j, X_{n-1} = i) = \mathbb{P}(X_n = j \mid X_{n-1} = i)\mathbb{P}(X_{n-1} = i) = Q_{ij}\mathbb{P}(X_{n-1} = i) = \pi_j\mathbb{P}(X_{n-1} = i).$$

  Ponieważ nie znamy π, nie możemy jednak dokonać takiego wyboru.
  W doborze Q powinniśmy zwrócić uwagę na to by skonstruowany ŁM "możliwie szybko" (ale nie za szybko) eksplorował przestrzeń stanów oraz żeby "dość często" (ale nie za często) akceptował nowe stany.
- W wersji ciągłej, algorytm M-H wygląda bardzo podobnie jak w wersji dyskretnej. Zamiast macierzy przejścia Q musimy zadać rodzinę gęstości (proposal distribution) $y \mapsto q(y \mid x)$ dla $x \in \mathcal{S}$.

  Algorytm 0.2 (Algorytm Metropolisa-Hastingsa, wersja absolutnie ciągła)
  Zainicjalizuj $X_0 \in \mathcal{S}$.
  Dla chwil $t = 1, 2, \ldots$ *wykonuj następujące kroki:*
  (i) w chwili t wygeneruj stan Y zgodnie z rozkładem o gęstośći $y \mapsto q(y \mid X_{t-1})$,
  (ii) Zaakceptuj stan Y z prawdopodobieństwem $\min\{\frac{f(Y)q(X_{t-1}|Y)}{f(X_{t-1})q(Y|X_{t-1})}, 1\}$ i wtedy połóż $X_t = Y$. W przypadku braku akceptacji połóż $X_t = X_{t-1}$.

(6) Bardzo prosty ciągły przykład. Zastosujemy algorytm M-H do generowania zmiennych z rozkładu $\mathrm{N}(0,1)$. Mamy $\mathcal{S} = \mathbb{R}$, a jako proposal distribution wybierzmy (symetryczny) rozkład

$$q(y \mid x) = q(y - x),$$

gdzie $q(x) = \frac{1}{2\delta}\mathbb{1}_{[-\delta,\delta]}(x)$ dla pewnej $\delta > 0$. Ponieważ mamy $q(x \mid y) = q(y \mid x)$, to $\alpha(x, y) = \min\{\frac{e^{-y^2/2}}{e^{-x^2/2}}, 1\}$ (nie musimy znać stałej normującej gęstość $\mathrm{N}(0,1)$).

Algorytm M-H przysposobiony do tego przykładu wygląda następująco:
Zainicjalizuj $X_0 = 0$ (na przykład).
Dla chwil $t = 1, 2, \ldots$ wykonuj następujące kroki:
(i) w chwili t wygeneruj stan Y zgodnie z rozkładem o gęstości

$$y \mapsto q(y \mid X_{t-1}) = q(y - X_{t-1}) = \frac{1}{2\delta}\mathbb{1}_{[-\delta,\delta]}(y - X_{t-1}) = \frac{1}{2\delta}\mathbb{1}_{[X_{t-1}-\delta, X_{t-1}+\delta]}(y).$$

Innymi słowy, $Y = X_{t-1} + U$, gdzie $U \sim \mathrm{U}([-\delta,\delta])$ jest niezależne od X_{t-1}.
(ii) Wylosuj liczbę $V \sim \mathrm{U}([0,1])$ oraz oblicz

$$X_t = \begin{cases} Y, & \text{jeśli } V \le \alpha(X_{t-1}, Y), \\ X_{t-1}, & \text{w p.p.} \end{cases}$$

Co wiemy o tak skonstruowanym ciągu $(X_n)_{n\ge 0}$?

- Twierdzenie mówi, że ma on asymptotycznie rozkład normalny.
- Gdyby $X_0 \sim \mathrm{N}(0,1)$, to wszystkie elementy ciągu miałyby również rozkład $\mathrm{N}(0,1)$.
- Elementy ciągu są bardzo skorelowane, tym bardziej im δ jest mniejsza. Szybkość zbieżności rozkładu empirycznego do rozkładu $\mathrm{N}(0,1)$ zależy od δ.

(7) Typowym zastosowaniem metod MCMC jest aproksymacja rozkładu a posteriori w statystyce Bayesowskiej. Czym jest statystyka Bayesowska?

- Niech $(f_\theta : \theta)$ będzie rodziną gęstości indeksowanych parametrem θ. (np. $f_\theta(x) = e^{-(x-\theta)^2/2}/\sqrt{2\pi}$).
- Dysponujemy obserwacją X z rozkładu o gęstości f_θ, przy czym nie znamy wartości θ. Na podstawie tej obserwacji chcemy dowiedzieć się czegoś o wykorzystanym θ.
  ⚠ Obserwację X powyżej należy rozumieć ogólnie, może być wielowymiarowa. W szczególności jedna obserwacja z rozkładu wielowymiarowego może być rozumiana jako ciąg obserwacji z rozkładów jednowymiarowych.
- Załóżmy, że gęstości f_θ są w rzeczywistości gęstościami warunkowymi pod warunkiem pewnej zmiennej losowej Θ, tzn.

$$f_\theta(x) = f_{X|\Theta=\theta}(x).$$

- Załóżmy, że zmienna losowa Θ ma gęstość f_Θ.
  ⚠ Ten i poprzedni krok są kluczowe w statystyce Bayesowskiej. Opierają się na filozoficznym (nieweryfikowalnym) założeniu, że nieznany parametr θ jest realizacją pewnej zmiennej losowej. Istnieją dwie

zwalczające się grupy: team częstościowcy (frequentists) oraz team Bayesowscy.

Rozkład zmiennej losowej Θ nazywamy rozkładem a priori; reprezentuje on wiedzę z jaką badacz rozpoczyna badanie. Istnieje wiele kryteriów wyboru rozkładu a priori. Wykorzystuje się np. tzw. rozkłady nieinformacyjne, np. rozkład jednostajny (często dziwny: na $[0, \infty)$ lub $\mathbb{R}$) lub rozkłady, dla których można podać jawną postać rozkładu a posteriori, o którym poniżej.

- Rozkładem a posteriori nazywamy rozkład o gęstości (gęstość parametru pod warunkiem danych)
$$f_{\Theta|X=x}(\theta) = \frac{f_{(\Theta,X)}(\theta, x)}{f_X(x)} = \frac{f_{X|\Theta=\theta}(x) f_\Theta(\theta)}{\int f_{X|\Theta=t}(x) f_\Theta(t) dt}.$$
Wersja dyskretna tego powyższego wzoru wygląda następująco
$$\mathbb{P}(\Theta = \theta_k \mid X = x_i) = \frac{\mathbb{P}(\Theta = \theta_k, X = x_i)}{\mathbb{P}(X = x_i)} = \frac{\mathbb{P}(X = x_i \mid \Theta = \theta_k)\mathbb{P}(\Theta = \theta_k)}{\sum_j \mathbb{P}(X = x_i \mid \Theta = \theta_j)\mathbb{P}(\Theta = \theta_j)}.$$
- Standardowym podejściem jest znalezienie parametru θ dla którego gęstość a posteriori jest największa. Gdybyśmy umieli symulować z rozkładu $\Theta \mid X = x$, to byśmy mogli wyznaczyć ten parametr na podstawie otrzymanego histogramu.
- W konkretnych zastosowaniach obliczenie stałej normującej gęstość rozkładu a posteriori (czyli mianownik) może nie być wykonalne. W takiej sytuacji znamy rozkład a posteriori z dokładnością do stałej normującej, więc zastosowanie metod MCMC jest tutaj bardzo naturalne.