Домашнє завдання №5

2024

Зміст

Задача 2. MCMC algorithms **2**

Завдання (a): MCMC algorithm . . . 2

Завдання (b): Gibbs Sampler . . . 3

Задача 3. Empirical Bayes for exponential families **5**

Постановка задачі . . . 5

Визначення математичного сподівання . . . 6

Визначення баєсової оцінки вектора θ . . . 7

Задача 4. Gamma-Poisson empirical Bayes **11**

Постановка задачі . . . 11

Допоміжні теоретичні викладки . . . 11

Визначення MLE для параметра $\widehat{\sigma}$ . . . 13

Визначення баєсової оцінки параметра θ_i . . . 15

Задача 5. Gibbs Sampler for Gamma-Poisson model **17**

Завдання (a): Gibbs Sampler . . . 17

Задача 2. MCMC algorithms

Завдання (a): MCMC algorithm

Нехай $\lambda(\theta)$ — деякий апостеріорний розподіл невідомого параметра $\theta \in \mathbb{R}$. Алгоритм MCMC має на меті згенерувати такий ланцюг Маркова $\left\{\theta^{(t)}\right\}_{t\geqslant 0}$ на множині станів $\mathbb{R}$, для якого $\lambda(\theta)$ є інваріантним розподілом. Нижче наведено один з варіантів MCMC, а саме — алгоритм Метрополіса-Гастінгса:

1. Задати початкове значення $\theta^{(0)} \in \mathbb{R}$ та розподіл пропозицій $q(y\,|\,x)$:

$$\forall t \geqslant 1 \text{ та } \forall x, y \in \mathbb{R} : q(y\,|\,x) = P\left(\theta^{(t)} = y\,|\,\theta^{(t-1)} = x\right) \tag{2.1}$$

2. Для $t = 1, \ldots, M$ повторювати кроки:

   (а) Згенерувати кандидата $v \sim q(v\,|\,u)$, де $u = \theta^{(t-1)}$;

   (б) Обчислити імовірність прийняття $\alpha(v\,|\,u)$ кандидата v :

$$\alpha(v\,|\,u) = \min\left\{1,\ \frac{\lambda(v)\,q(u\,|\,v)}{\lambda(u)\,q(v\,|\,u)}\right\} \tag{2.2}$$

   (в) З імовірністю $\alpha(v\,|\,u)$ покласти $\theta^{(t)} = v$, з імовірністю $1 - \alpha(v\,|\,u)$ покласти $\theta^{(t)} = u$.

Завдання: показати, що $\lambda(\theta)$ — стаціонарний розподіл для ланцюга $\left\{\theta^{(t)}\right\}_{t\geqslant 0}$.

Розв'язок

Означення 1. Розподіл π ланцюга Маркова з розподілом перехідних імовірностей $Q(y\,|\,x)$ називають стаціонарним, якщо

$$\forall y \in \mathbb{R} : \pi(y) = \int\limits_{\mathbb{R}} Q(y\,|\,x)\,\pi(x)\,dx \tag{2.3}$$

Теорема 1. *Якщо для деякого розподілу π ланцюга Маркова і деякого розподілу станів $Q(y\,|\,x)$ виконується рівняння балансу*

$$\forall x, y \in \mathbb{R} : \pi(x)\,Q(y\,|\,x) = \pi(y)\,Q(x\,|\,y), \tag{2.4}$$

то π є стаціонарним розподілом для $Q(y\,|\,x)$.

Доведення. *Якщо $\forall x, y \in \mathbb{R}$ справедливо*

$$\pi(x)\,Q(y\,|\,x) = \pi(y)\,Q(x\,|\,y), \tag{2.5}$$

тоді проінтегрувавши обидві частини рівняння, матимемо

$$\int\limits_{\mathbb{R}} \pi(x)\, Q(y\,|\,x)\, dx = \int\limits_{\mathbb{R}} \pi(y)\, Q(x\,|\,y)\, dx, \tag{2.6}$$

а отже, згідно властивостей розподілу станів отримуємо

$$\int\limits_{\mathbb{R}} \pi(x)\, Q(y\,|\,x)\, dx = \int\limits_{\mathbb{R}} \pi(y)\, Q(x\,|\,y)\, dx = \pi(y) \tag{2.7}$$

Таким чином, в силу Озн. 1 розподіл π є стаціонарним. □

Алгоритм MCMC будує ланцюг Маркова з розподілом станів такого вигляду:

$$\forall x, y \in \mathbb{R} : Q(y\,|\,x) = q(y\,|\,x)\, \alpha(y\,|\,x) \tag{2.8}$$

Відтак, перевіримо справедливість рівняння балансу для апостеріорного розподілу $\lambda(\theta)$, згенерованого за допомогою MCMC:

$$\begin{aligned}
\forall x, y \in \mathbb{R} : \lambda(x)\, Q(y\,|\,x) &= \lambda(x)\, q(y\,|\,x)\, \alpha(y\,|\,x) = \\
&= \lambda(x)\, q(y\,|\,x) \cdot \min\left\{1,\, \frac{\lambda(y)\, q(x\,|\,y)}{\lambda(x)\, q(y\,|\,x)}\right\} = \\
&= \min\left\{\lambda(x)\, q(y\,|\,x),\, \lambda(x)\, q(y\,|\,x)\, \frac{\lambda(y)\, q(x\,|\,y)}{\lambda(x)\, q(y\,|\,x)}\right\} = \\
&= \min\left\{\lambda(x)\, q(y\,|\,x),\, \lambda(y)\, q(x\,|\,y)\right\} = \\
&= \lambda(y)\, q(x\,|\,y) \cdot \min\left\{\frac{\lambda(x)\, q(y\,|\,x)}{\lambda(y)\, q(x\,|\,y)},\, 1\right\} = \\
&= \lambda(y)\, q(x\,|\,y)\, \alpha(x\,|\,y) = \\
&= \lambda(y)\, Q(x\,|\,y)
\end{aligned} \tag{2.9}$$

Отже, за Теоремою 1 розподіл $\lambda(\theta)$ є стаціонарним для ланцюга Маркова, згенерованого згідно з алгоритмом MCMC.

Завдання (b): Gibbs Sampler

Нехай $\lambda(\theta)$ — апостеріорний розподіл вектора параметрів $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$. Вибірка Гіббса полягає у тому, щоб симулювати вибірку із заданого розподілу $\lambda(\theta)$, генеруючи значення чергового параметра при фіксованих (ініціалізованих на першій ітерації) значеннях решти параметрів.

Завдання: показати, що вибірка Гіббса є частковим випадком алгоритму Метрополіса-Гастінгса.

Розв'язок

Алгоритм Метрополіса-Гастінгса генерує ланцюг Маркова, крок за кроком приймаючи стан-пропозицію, згенеровану із деякого розподілу пропозицій, з певною імовірністю α (яка обчислюється з огляду на заданий апостеріорний розподіл). При цьому, фактично, апостеріорний роподіл може включати й інші параметри, які в рамках MCMC покладені фіксованим значенням.

Вибірка Гіббса, своєю чергою, на кожному кроці завжди (тобто з імовірністю $\alpha = 1$) приймає стан-пропозицію, згенеровану безпосередньо із апостеріорного розподілу. Іншими словами, вибірка Гіббса — частковий випадок алгоритму Метрополіса-Гастінгса з апостеріорним розподілом пропозицій та імовірністю прийняття пропозиції $\alpha = 1$. Формальний опис розв'язку наведено за *посиланням*.

Задача 3. Empirical Bayes for exponential families

Постановка задачі

Нехай задано випадковий вектор $\overrightarrow{X} = (X_1, \ldots, X_n)$, який коротко позначимо як X, з s-параметризованої експоненціальної сім'ї розподілів з вектором параметрів $\overrightarrow{\theta} = (\theta_1, \ldots, \theta_s)$, у короткому записі θ. Тоді щільність розподілу вектора X матиме вигляд:

$$p_\theta(x) \equiv p(X = x \mid \theta) = e^{A(\theta)T(x) - B(\theta)} h(x) \tag{3.1}$$

Згідно з умовою задачі функція $A(\theta) = \theta$. Іншими словами, вектор X належить експоненціальній сім'ї розподілів у так званій канонічній формі:

$$p_\theta(x) = e^{\theta T(x) - B(\theta)} h(x) \tag{3.2}$$

Розписуючи скалярний добуток векторів θ та $T(x)$, матимемо такий еквівалентний запис:

$$p_\theta(x) = e^{\sum\limits_{j=1}^{s} \theta_j T_j(x) - B(\theta)} h(x) \tag{3.3}$$

Крім того, в рамках задачі вектор θ є випадковим вектором з деякого (апріорного) розподілу $\lambda_\gamma(\theta)$, який своєю чергою параметризований фіксованим значенням параметра γ.

Позначаючи апостеріорний розоділ вектора параметрів θ як $\lambda_\gamma(\theta \mid x)$, а безумовний розподіл випадкового вектора X як $q_\gamma(x)$, у завданні слід показати, що для $i = \overline{1, n}$:

$$\mathbb{E}_\gamma \left[\sum_{j=1}^{s} \theta_j \frac{\partial T_j(x)}{\partial x_i} \mid X = x \right] = \frac{\partial}{\partial x_i} \ln q_\gamma(x) - \frac{\partial}{\partial x_i} \ln h(x) \tag{3.4}$$

А у випадку $s = n$ та $T(x) = x$, а також покладаючи значення параметра γ як оцінку максимальної правдоподібності

$$\widehat{\gamma} = \underset{\gamma}{\operatorname{argmax}}\, L(X \mid \gamma) = \underset{\gamma}{\operatorname{argmax}}\, q_\gamma(x), \tag{3.5}$$

необхідно продемонструвати, що баєсова оцінка вектора θ (тобто математичне сподівання апостеріорного розподілу) дорівнює, відповідно

$$\mathbb{E}_{\widehat{\gamma}} \left[\theta \mid X = x\right] = \nabla_x \big(\ln q_{\widehat{\gamma}}(x) - \ln h(x)\big), \tag{3.6}$$

де оператор набла визначений так:

$$\nabla_x = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right) \tag{3.7}$$

Визначення математичного сподівання

Розглянемо отримання математичного сподівання у виразі (3.4). Наведемо задану баєсівську модель:

$$(X_1, \ldots, X_n) \,|\, (\theta_1, \ldots, \theta_s) \sim p_\theta(x) \tag{3.8}$$

$$(\theta_1, \ldots, \theta_s) \sim \lambda_\gamma(\theta) \tag{3.9}$$

Запишемо згідно із формулою Баєса вираз для визначення апостеріорного розподілу $\lambda_\gamma(\theta \,|\, x)$ вектора θ при заданому апріорному розподілі $\lambda_\gamma(\theta)$, безумовному розподілі даних $q_\gamma(x)$ та функції правдоподібності $p_\theta(x)$ (3.3):

$$\lambda_\gamma(\theta \,|\, x) = \frac{p_\theta(x)\lambda_\gamma(\theta)}{q_\gamma(x)} = \frac{h(x)}{q_\gamma(x)}\, \lambda_\gamma(\theta)\, e^{\sum\limits_{j=1}^{s} \theta_j T_j(x) - B(\theta)} \tag{3.10}$$

Прологарифмуємо отриманий вираз:

$$\ln \lambda_\gamma(\theta \,|\, x) = \ln h(x) - \ln q_\gamma(x) + \ln \lambda_\gamma(\theta) + \sum_{j=1}^{s} \theta_j T_j(x) - B(\theta) \tag{3.11}$$

Продиференціюємо (3.11) за координатою x_i, враховуючи відсутню залежність від координати у функціях $\lambda_\gamma(\theta)$ та $B(\theta)$:

$$\frac{\partial}{\partial x_i} \ln \lambda_\gamma(\theta \,|\, x) = \frac{\partial}{\partial x_i} \ln h(x) - \frac{\partial}{\partial x_i} \ln q_\gamma(x) + \sum_{j=1}^{s} \theta_j \frac{\partial T_j(x)}{\partial x_i} \tag{3.12}$$

Таким чином, отримуємо співвідношення для шуканої суми:

$$\sum_{j=1}^{s} \theta_j \frac{\partial T_j(x)}{\partial x_i} = \frac{\partial}{\partial x_i} \ln \lambda_\gamma(\theta \,|\, x) + \frac{\partial}{\partial x_i} \ln q_\gamma(x) - \frac{\partial}{\partial x_i} \ln h(x) \tag{3.13}$$

Обчислимо математичне сподівання виразу (3.13), використовуючи властивість лінійності математичного сподівання та враховуючи відсутню залежність від вектора θ у частини доданків:

$$\mathbb{E}_\gamma \left[\sum_{j=1}^{s} \theta_j \frac{\partial T_j(x)}{\partial x_i} \,|\, X = x \right] = \mathbb{E}_\gamma \left[\frac{\partial}{\partial x_i} \ln \lambda_\gamma(\theta \,|\, x) \,|\, X = x \right] + \frac{\partial}{\partial x_i} \ln q_\gamma(x) - \frac{\partial}{\partial x_i} \ln h(x) \tag{3.14}$$

Детальніше розглянемо перший доданок у правій частині рівняння (3.14). За означенням умовного математичного сподівання для неперервної випадкової величини $g(\xi) \,|\, \eta$:

$$\mathbb{E}\left[g(\xi) \,|\, \eta\right] \overset{\text{def}}{=} \int\limits_{\mathbb{R}} g(y)\, f_{\xi \,|\, \eta}(y \,|\, \eta)\, dy \tag{3.15}$$

Таким чином:

$$\mathbb{E}_\gamma\left[\frac{\partial}{\partial x_i}\ln\lambda_\gamma(\theta\,|\,x)\,|\,X=x\right]=\int\limits_{\mathbb{R}^s}\frac{\partial}{\partial x_i}\ln\lambda_\gamma(\theta\,|\,x)\,\lambda_\gamma(\theta\,|\,x)\,d\theta \tag{3.16}$$

У викладках нижче послідовно продиференціюємо функцію логарифма (1), скористаємося інтегральним правилом Лейбніца для винесення оператора диференціювання за межі знаку інтегрування (2) та використаємо властивість, що інтеграл довільної щільності розподілу на всій області визначення дорівнює одиниці (3):

$$\begin{aligned}\mathbb{E}_\gamma\left[\frac{\partial}{\partial x_i}\ln\lambda_\gamma(\theta\,|\,x)\,|\,X=x\right]&=\int\limits_{\mathbb{R}^s}\frac{\partial}{\partial x_i}\ln\lambda_\gamma(\theta\,|\,x)\,\lambda_\gamma(\theta\,|\,x)\,d\theta=\\&\overset{(1)}{=}\int\limits_{\mathbb{R}^s}\frac{\frac{\partial}{\partial x_i}\lambda_\gamma(\theta\,|\,x)}{\lambda_\gamma(\theta\,|\,x)}\lambda_\gamma(\theta\,|\,x)\,d\theta=\\&=\int\limits_{\mathbb{R}^s}\frac{\partial}{\partial x_i}\lambda_\gamma(\theta\,|\,x)\,d\theta=\\&\overset{(2)}{=}\frac{\partial}{\partial x_i}\int\limits_{\mathbb{R}^s}\lambda_\gamma(\theta\,|\,x)\,d\theta=\\&\overset{(3)}{=}0\end{aligned}\tag{3.17}$$

Отже, математичне сподівання (3.14) отримає шуканий вигляд:

$$\mathbb{E}_\gamma\left[\sum_{j=1}^{s}\theta_j\frac{\partial T_j(x)}{\partial x_i}\,|\,X=x\right]=\frac{\partial}{\partial x_i}\ln q_\gamma(x)-\frac{\partial}{\partial x_i}\ln h(x)\tag{3.18}$$

Визначення баєсової оцінки вектора θ

Нехай тепер задано баєсівську модель

$$(X_1,\ldots,X_n)\,|\,(\theta_1,\ldots,\theta_s)\sim p_\theta(x)\tag{3.19}$$

$$(\theta_1,\ldots,\theta_n)\sim\lambda_{\widehat{\gamma}}(\theta)\tag{3.20}$$

над n-параметризованою експоненціальною сім'єю розподілів

$$p_\theta(x)=e^{\theta T(x)-B(\theta)}h(x)\tag{3.21}$$

при фіксованому значенні параметра $\widehat{\gamma}$ як оцінки максимальної правдоподібності (3.5). Крім того, у розподілі (3.21) покладено $T(x)=x$, що буде використано у викладках нижче.

Процедура пошуку математичного сподівання (3.18) у попередньому підрозділі вказує шлях пошуку баєсової оцінки вектора θ. Тож першим кроком аналогічним чином запишемо згідно із формулою Баєса вираз для визначення апостеріорного розподілу $\lambda_{\widehat{\gamma}}(\theta \,|\, x)$ вектора θ при заданому апріорному розподілі $\lambda_{\widehat{\gamma}}(\theta)$, безумовному розподілі даних $q_{\widehat{\gamma}}(x)$ та функції правдоподібності $p_\theta(x)$ (3.21):

$$\lambda_{\widehat{\gamma}}(\theta \,|\, x) = \frac{p_\theta(x)\lambda_{\widehat{\gamma}}(\theta)}{q_{\widehat{\gamma}}(x)} = \frac{h(x)}{q_{\widehat{\gamma}}(x)}\, \lambda_{\widehat{\gamma}}(\theta)\, e^{\theta T(x) - B(\theta)}, \tag{3.22}$$

де $\theta \cdot T(x)$ є нічим іншим як скалярним добутком двох векторів. Прологарифмуємо отриманий вираз:

$$\ln \lambda_{\widehat{\gamma}}(\theta \,|\, x) = \ln h(x) - \ln q_{\widehat{\gamma}}(x) + \ln \lambda_{\widehat{\gamma}}(\theta) + \theta T(x) - B(\theta) \tag{3.23}$$

Оскільки розмірність вектора θ збігається із розмірністю вектора X, застосуємо до (3.23) оператор набла ∇_x (3.7):

$$\nabla_x \ln \lambda_{\widehat{\gamma}}(\theta \,|\, x) = \nabla_x \ln h(x) - \nabla_x \ln q_{\widehat{\gamma}}(x) + \nabla_x \ln \lambda_{\widehat{\gamma}}(\theta) + \nabla_x\big(\theta T(x)\big), \tag{3.24}$$

де на відміну від попереднього підрозділу у рівності (3.12) доданок з апріорним розподілом $\nabla_x \ln \lambda_{\widehat{\gamma}}(\theta)$ не дорівнює нулю, адже значення гіперапараметра $\widehat{\gamma}$ як оцінки максимальної правдоподідності (3.5) містить функціональну залежність від даних x.

Наступні кроки спрямуємо на визначення дії оператора набла на скалярний добуток двох векторів:

$$\nabla_x\big(\theta T(x)\big) = \nabla_x(\theta) T(x) + \theta\, \nabla_x\big(T(x)\big) = \theta\, \nabla_x\big(T(x)\big), \tag{3.25}$$

а оскільки в рамках завдання покладено $T(x) = x$, то як наслідок

$$\nabla_x\big(\theta T(x)\big) = \theta \tag{3.26}$$

На додачу до властивості адитивності диференціального оператора набла, в результаті вираз (3.24) отримує вид:

$$\theta = \nabla_x \ln \lambda_{\widehat{\gamma}}(\theta \,|\, x) - \nabla_x \ln \lambda_{\widehat{\gamma}}(\theta) + \nabla_x\big(\ln q_{\widehat{\gamma}}(x) - \ln h(x)\big) \tag{3.27}$$

Аналогічно до (3.14) обчислимо умовне математичне сподівання:

$$\begin{split} \mathbb{E}_{\widehat{\gamma}}\left[\theta \,|\, X = x\right] = \mathbb{E}_{\widehat{\gamma}}\left[\nabla_x \ln \lambda_{\widehat{\gamma}}(\theta \,|\, x) \,|\, X = x\right] - \mathbb{E}_{\widehat{\gamma}}\left[\nabla_x \ln \lambda_{\widehat{\gamma}}(\theta) \,|\, X = x\right] + \\ + \nabla_x\big(\ln q_{\widehat{\gamma}}(x) - \ln h(x)\big) \end{split} \tag{3.28}$$

Згідно з означенням (3.15) та відповідно до викладок (3.17):

$$\begin{aligned}
\mathbb{E}_{\widehat{\gamma}}\left[\nabla_x \ln \lambda_{\widehat{\gamma}}(\theta \,|\, x) \,|\, X = x\right] &= \int\limits_{\mathbb{R}^n} \nabla_x \ln \lambda_{\widehat{\gamma}}(\theta \,|\, x)\, \lambda_{\widehat{\gamma}}(\theta \,|\, x)\, d\theta = \\
&= \int\limits_{\mathbb{R}^n} \frac{\nabla_x \lambda_{\widehat{\gamma}}(\theta \,|\, x)}{\lambda_{\widehat{\gamma}}(\theta \,|\, x)}\, \lambda_{\widehat{\gamma}}(\theta \,|\, x)\, d\theta = \\
&= \int\limits_{\mathbb{R}^n} \nabla_x \lambda_{\widehat{\gamma}}(\theta \,|\, x)\, d\theta = \\
&= \nabla_x \int\limits_{\mathbb{R}^n} \lambda_{\widehat{\gamma}}(\theta \,|\, x)\, d\theta = \\
&= 0
\end{aligned} \tag{3.29}$$

Розпишемо другий доданок математичного сподівання (3.28):

$$\mathbb{E}_{\widehat{\gamma}}\left[\nabla_x \ln \lambda_{\widehat{\gamma}}(\theta) \,|\, X = x\right] = \int\limits_{\mathbb{R}^n} \nabla_x \ln \lambda_{\widehat{\gamma}}(\theta)\, \lambda_{\widehat{\gamma}}(\theta \,|\, x)\, d\theta = \int\limits_{\mathbb{R}^n} \frac{\nabla_x \lambda_{\widehat{\gamma}}(\theta)}{\lambda_{\widehat{\gamma}}(\theta)}\, \lambda_{\widehat{\gamma}}(\theta \,|\, x)\, d\theta \tag{3.30}$$

За правилом диференціювання складеної функції:

$$\nabla_x \lambda_{\widehat{\gamma}}(\theta) = \frac{\partial \lambda_{\widehat{\gamma}}(\theta)}{\partial x} = \frac{\partial \lambda_{\widehat{\gamma}}(\theta)}{\partial \gamma} \cdot \frac{\partial \gamma}{\partial x} \tag{3.31}$$

Відтак, враховуючи інтегральне правило Лейбніца, матимемо такий еквівалентний запис:

$$\mathbb{E}_{\widehat{\gamma}}\left[\nabla_x \ln \lambda_{\widehat{\gamma}}(\theta) \,|\, X = x\right] = \frac{\partial \gamma}{\partial x} \int\limits_{\mathbb{R}^n} \frac{\partial \lambda_{\widehat{\gamma}}(\theta)}{\partial \gamma} \cdot \frac{\lambda_{\widehat{\gamma}}(\theta \,|\, x)}{\lambda_{\widehat{\gamma}}(\theta)}\, d\theta \tag{3.32}$$

З іншого боку, гіперапарметр $\widehat{\gamma}$ є оцінкою максимальної правдоподібності (3.5) безумовного розподілу даних $q_\gamma(x)$, а отже, похідна від функції $q_{\widehat{\gamma}}(x)$ в точці $\widehat{\gamma}$ дорівнює нулю:

$$\widehat{\gamma} = \operatorname*{argmax}_{\gamma} q_\gamma(x) \iff \frac{\partial}{\partial \gamma} q_{\widehat{\gamma}}(x) = 0 \tag{3.33}$$

За формулою повної ймовірності (4.11) у застосуванні до повної групи подій, яка утворена вектором θ з неперервних випадкових величин:

$$q_{\widehat{\gamma}}(x) = \int\limits_{\mathbb{R}^n} p_\theta(x)\, \lambda_{\widehat{\gamma}}(\theta)\, d\theta \tag{3.34}$$

Як наслідок, матимемо:

$$
\begin{aligned}
\frac{\partial}{\partial\gamma}\, q_{\widehat{\gamma}}(x) = \frac{\partial}{\partial\gamma} \int\limits_{\mathbb{R}^n} p_\theta(x)\, \lambda_{\widehat{\gamma}}(\theta)\, d\theta = \\
&= \int\limits_{\mathbb{R}^n} \frac{\partial}{\partial\gamma} \Big(p_\theta(x) \cdot \lambda_{\widehat{\gamma}}(\theta) \Big)\, d\theta = \\
&= \int\limits_{\mathbb{R}^n} \frac{\partial p_\theta(x)}{\partial\gamma}\, \lambda_{\widehat{\gamma}}(\theta)\, d\theta + \int\limits_{\mathbb{R}^n} p_\theta(x)\, \frac{\partial \lambda_{\widehat{\gamma}}(\theta)}{\partial\gamma}\, d\theta = \\
&= \int\limits_{\mathbb{R}^n} p_\theta(x)\, \frac{\partial \lambda_{\widehat{\gamma}}(\theta)}{\partial\gamma}\, d\theta
\end{aligned}
\tag{3.35}
$$

Наостанок, за формулою Баєса (3.22):

$$
\lambda_{\widehat{\gamma}}(\theta \,|\, x) = \frac{p_\theta(x)\lambda_{\widehat{\gamma}}(\theta)}{q_{\widehat{\gamma}}(x)} \iff p_\theta(x) = \frac{\lambda_{\widehat{\gamma}}(\theta \,|\, x)}{\lambda_{\widehat{\gamma}}(\theta)}\, q_{\widehat{\gamma}}(x) \tag{3.36}
$$

Тоді в силу (3.33) вираз (3.35) отримає вид:

$$
\frac{\partial}{\partial\gamma}\, q_{\widehat{\gamma}}(x) = 0 \iff q_{\widehat{\gamma}}(x) \int\limits_{\mathbb{R}^n} \frac{\lambda_{\widehat{\gamma}}(\theta \,|\, x)}{\lambda_{\widehat{\gamma}}(\theta)} \cdot \frac{\partial \lambda_{\widehat{\gamma}}(\theta)}{\partial\gamma}\, d\theta = 0 \tag{3.37}
$$

Відповідно, шукане математичне сподівання (3.32) буде дорівнювати нулю:

$$
\mathbb{E}_{\widehat{\gamma}} \left[\nabla_x \ln \lambda_{\widehat{\gamma}}(\theta) \,|\, X = x \right] = 0 \tag{3.38}
$$

Отже, остаточний вигляд баєсової оцінки вектора θ (3.28) записуватиметься таким формульним співвідношенням:

$$
\mathbb{E}_{\widehat{\gamma}} \left[\theta \,|\, X = x \right] = \nabla_x \Big(\ln q_{\widehat{\gamma}}(x) - \ln h(x) \Big) \tag{3.39}
$$

Коментар. У підсумку можна зауважити, що на вигляд баєсової оцінки для експоненціальної сім'ї розподілів у канонічній формі впливають лише няавні дані та емпірична оцінка значення гіперпараметра апріорного розподілу (а не його безпосередній вигляд).

Задача 4. Gamma-Poisson empirical Bayes

Постановка задачі

Розглянемо модель такого вигляду:

$$X_{ij}\,|\,\theta_i \overset{\text{ind}}{\sim} \text{Poiss}(\theta_i),\ i=\overline{1,n},\ j=\overline{1,m}, \tag{4.1}$$

$$\theta_i \overset{\text{i.i.d.}}{\sim} \text{Gamma}(k,\widehat{\sigma}),\ i=\overline{1,n}, \tag{4.2}$$

при цьому k є відомим, а $\widehat{\sigma}$ визначене як оцінка максимальної правдоподібності. Завдання полягає у тому, щоб віднайти баєсову оцінку параметра θ_i, тобто математичне сподівання апостеріорного розподілу параметра θ_i. Також зазначимо, що відповідно до (4.1) та (4.2) дані організовані таким чином:

$$\begin{gathered} \theta_1 : X_{11}, X_{12}, \ldots, X_{1m} \\ \theta_2 : X_{21}, X_{22}, \ldots, X_{2m} \\ \ldots \\ \theta_n : X_{n1}, X_{n2}, \ldots, X_{nm} \end{gathered} \tag{4.3}$$

Іншими словами, для кожної згенерованої випадкової величини θ_i з Гамма-розподілу (4.2) генерується набір випадкових величин $X_{i1}, X_{i2}, \ldots, X_{im}$ з розподілу Пуасона з параметром θ_i.

Допоміжні теоретичні викладки

Перш за все, наведемо імовірнісні розподіли випадкових величин, про які ітиме мова. Нехай неперервна випадкова величина ξ має Гамма-розподіл з параметрами $\alpha > 0$ та $\beta > 0$:

$$\xi \sim Gamma(\alpha, \beta) \tag{4.4}$$

На рівні позначень існує дві еквівалентні параметризації α та β :

- «shape parameter» $\alpha = k$ та «scale parameter» $\beta = \sigma$;
- «shape parameter» $\alpha = k$ та «inverse scale parameter» $\beta = 1/\sigma$, який ще називають «rate parameter».

Відповідно до кожної з параметризацій щільність розподілу $f_\xi(x)$ має відповідний вигляд. У випадку «scale parameter» $\beta = \sigma$:

$$\xi \sim \text{Gamma}(k,\sigma) \iff f_\xi(x) = \frac{x^{k-1}}{\sigma^k \Gamma(k)}\, e^{-x/\sigma}\, \mathbb{1}(x>0), \tag{4.5}$$

де, відповідно, математичне сподівання та дисперсія складають

$$\mathbb{E}\xi = k\sigma,\ \mathrm{Var}\xi = k\sigma^2 \tag{4.6}$$

У випадку ж «rate parameter» $\beta = 1/\sigma$:

$$\xi \sim \mathrm{Gamma}(k, 1/\sigma) \iff f_\xi(x) = \frac{x^{k-1}}{\Gamma(k)}\,\sigma^k e^{-x\sigma}\,\mathbb{1}(x > 0), \tag{4.7}$$

де математичне сподівання та дисперсія

$$\mathbb{E}\xi = \frac{1}{\sigma},\ \mathrm{Var}\xi = \frac{1}{\sigma^2} \tag{4.8}$$

Коментар. Згідно з умовою завдання, у роботі розглядатиметься Гамма-розподіл саме зі «scale parameter» $\beta = \sigma$. Спершу я не звернув уваги на цю приписку в умові завдання (адже раніше не зустрічав різну параметризацію у Гамма-розподілі), що спричинило деяку плутанину в процесі розписування викладок. Пізніше я відловив неточність і вирішив задля систематизації та ясності вказати означення, наведені вище.

Дискретна випадкова величина η має розподіл Пуасона з параметром λ, якщо:

$$\eta \sim Poiss(\lambda) \iff P(\eta = x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{4.9}$$

Кажуть, що дискретна випадкова величина η має від'ємний біноміальний розподіл з параметром кількості успіхів r та ймовірністю успіху p, якщо:

$$\eta \sim NB(r, p) \iff P(\eta = x) = \tbinom{x+r-1}{x}(1-p)^x p^r \tag{4.10}$$

Наостанок, наведемо формулу повної ймовірності для деякої події A та повної групи подій $\{H_i : i = 1, 2, 3, \ldots\}$, визначеної та тому ж імовірнісному просторі, що і подія A :

$$P(A) = \sum_i P(A \,|\, H_i)P(H_i) \tag{4.11}$$

Формула (4.11) застосовна і до пошуку ймовірності події A за умови деякої іншої події C :

$$\begin{aligned} P(A \,|\, C) &= \frac{P(A, C)}{P(C)} = \\ &= \frac{\sum\limits_i P(A, H_i, C)}{P(C)} = \\ &= \frac{\sum\limits_i P(A \,|\, H_i, C)P(H_i \,|\, C)P(C)}{P(C)} = \\ &= \sum_i P(A \,|\, H_i, C)P(H_i \,|\, C) \end{aligned} \tag{4.12}$$

Визначення MLE для параметра $\widehat{\sigma}$

Визначимо оцінку максимальної правдоподібності параметра $\widehat{\sigma}$ у Гамма-розподілі (4.2):

$$\widehat{\sigma} = \underset{\sigma}{\operatorname{argmax}} \, L(X_{11}, X_{12}, \dots, X_{1m}; \dots; X_{n1}, X_{n2}, \dots, X_{nm} \,|\, \sigma) \tag{4.13}$$

З огляду на незалежність наборів $X_{i1}, X_{i2}, \dots, X_{im}$, функція правдоподібності матиме вид:

$$\begin{aligned} L(X_{11}, X_{12}, \dots, X_{1m}; \dots; X_{n1}, X_{n2}, \dots, X_{nm} \,|\, \sigma) &\overset{\text{ind}}{=} \prod_{i=1}^{n} f(X_{i1}, X_{i2}, \dots, X_{im} \,|\, \sigma) = \\ &\overset{\text{ind}}{=} \prod_{i=1}^{n} \prod_{j=1}^{m} f(X_{ij} \,|\, \sigma) \end{aligned} \tag{4.14}$$

Використаємо формулу повної ймовірності (4.11) та формулу (4.12) у застосуванні до повної групи подій, яка утворена неперервною випадковою величиною θ_i. Тоді розподіл $f(X_{ij} \,|\, \sigma)$ у виразі (4.14) отримає вид:

$$f(X_{ij} \,|\, \sigma) = \int\limits_{\mathbb{R}} f(X_{ij} \,|\, \theta_i, \sigma) f(\theta_i \,|\, \sigma) \, d\theta_i \tag{4.15}$$

Враховуючи розподіл Пуасона для X_{ij} (4.1) та Гамма-розподілені величини θ_i (4.2), матимемо:

$$\begin{aligned} f(X_{ij} \,|\, \sigma) &= \int\limits_{\mathbb{R}} f(X_{ij} \,|\, \theta_i, \sigma) f(\theta_i \,|\, \sigma) \, d\theta_i = \\ &= \int\limits_{\mathbb{R}} \frac{\theta_i^{X_{ij}} e^{-\theta_i}}{X_{ij}!} \cdot \frac{\theta_i^{k-1}}{\sigma^k \Gamma(k)} \, e^{-\theta_i/\sigma} \, \mathbb{1}(\theta_i > 0) \, d\theta_i = \\ &= C \, \sigma^{-k} \int\limits_{0}^{\infty} \theta_i^{X_{ij}+k-1} \, e^{-\theta_i (1+\frac{1}{\sigma})} \, d\theta_i \end{aligned} \tag{4.16}$$

За означенням, Гамма-функція в деякій точці a при умові $\operatorname{Re}(a) > 0$ записується таким чином:

$$\Gamma(a) = \int\limits_{0}^{\infty} y^{a-1} \, e^{-y} \, dy \tag{4.17}$$

Використаємо заміну $y = bx$ й запишемо Гамма-функцію у точці $a + 1:$

$$\Gamma(a+1) = \int\limits_{0}^{\infty} y^{a} \, e^{-y} \, dy = \int\limits_{0}^{\infty} (bx)^{a} \, e^{-bx} \, b \, dx = b^{a+1} \int\limits_{0}^{\infty} x^{a} \, e^{-bx} \, dx \tag{4.18}$$

Як наслідок, матимемо:

$$\int_0^\infty x^a\, e^{-bx}\, dx = b^{-a-1}\, \Gamma(a+1) \tag{4.19}$$

Застосуємо (4.19) до розподілу (4.16), в результаті отримаємо:

$$\begin{aligned} f(X_{ij} \,|\, \sigma) &= C\, \sigma^{-k} \int_0^\infty \theta_i^{X_{ij}+k-1}\, e^{-\theta_i\left(1+\frac{1}{\sigma}\right)}\, d\theta_i = \\ &= C\, \sigma^{-k} \left(1+\frac{1}{\sigma}\right)^{-X_{ij}-k} \Gamma(X_{ij}+k) = \\ &= C\, \sigma^{-k} \left(1+\frac{1}{\sigma}\right)^{-X_{ij}-k} \end{aligned} \tag{4.20}$$

Тоді функція правдоподібності (4.14) матиме такий остаточний вигляд:

$$L = \prod_{i=1}^{n}\prod_{j=1}^{m} f(X_{ij} \,|\, \sigma) = \prod_{i=1}^{n}\prod_{j=1}^{m} C\, \sigma^{-k} \left(1+\frac{1}{\sigma}\right)^{-X_{ij}-k} = C\, \sigma^{-nmk} \left(1+\frac{1}{\sigma}\right)^{-nm\overline{X}-nmk}, \tag{4.21}$$

де позначено

$$\overline{X} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij} \tag{4.22}$$

Оскільки довільна диференційовна функція та логарифм від неї досягають екстремумів в однакових точках, то пошук розв'язку рівняння

$$\frac{d}{d\sigma}\ln L = 0 \tag{4.23}$$

призведе до оцінки максимальної правдоподібності парамета σ :

$$\begin{aligned} \frac{d}{d\sigma}\ln L = \frac{d}{d\sigma}\left(\ln C - nmk\ln\sigma - (nm\overline{X}+nmk)\ln\left(1+\frac{1}{\sigma}\right)\right) = \\ = -\frac{nmk}{\sigma} + \frac{nm\overline{X}+nmk}{1+\frac{1}{\sigma}}\,\sigma^{-2} = \\ = -\frac{nmk}{\sigma} + \frac{nm\overline{X}+nmk}{\sigma(\sigma+1)}, \end{aligned} \tag{4.24}$$

а відтак матимемо

$$\frac{d}{d\sigma}\ln L = 0 \iff \widehat{\sigma} = \frac{\overline{X}}{k} \tag{4.25}$$

Визначення баєсової оцінки параметра θ_i

Відштовхуючись від $h(\theta_i)$ — апріорного Гамма-розподілу (4.2) з параметрами k й $\widehat{\sigma}$ — та наявних даних $X_{i1}, X_{i2}, \ldots, X_{im}$ (4.1), застосуємо формулу Баєса для пошуку апостеріорного розподілу параметра θ_i :

$$h(\theta_i \,|\, X_{i1}, X_{i2}, \ldots, X_{im}) \propto h(\theta_i) L(X_{i1}, X_{i2}, \ldots, X_{im} \,|\, \theta_i) \tag{4.26}$$

Отже, апріорний розподіл має вигляд:

$$h(\theta_i) = \frac{\theta_i^{k-1}}{\widehat{\sigma}^k \Gamma(k)} \, e^{-\theta_i/\widehat{\sigma}} \, \mathbb{1}(\theta_i > 0) \tag{4.27}$$

Функція правдоподібності, своєю чергою, розписуватиметься таким чином:

$$L(X_{i1}, X_{i2}, \ldots, X_{im} \,|\, \theta_i) \overset{\text{ind}}{=} \prod_{j=1}^{m} P\left(X_{ij} \,|\, \theta_i\right) = \prod_{j=1}^{m} \frac{\theta_i^{X_{ij}} e^{-\theta_i}}{X_{ij}!} = \frac{\theta_i^{\sum\limits_{j=1}^{m} X_{ij}} e^{-m\theta_i}}{X_{i1}! X_{i2}! \cdots X_{im}!} \tag{4.28}$$

Отже, матимемо:

$$L(X_{i1}, X_{i2}, \ldots, X_{im} \,|\, \theta_i) = C \, \theta_i^{m\overline{X}_i} \, e^{-m\theta_i} \tag{4.29}$$

Відтак апостеріорний розподіл матиме вигляд:

$$\begin{aligned} h(\theta_i \,|\, X_{i1}, X_{i2}, \ldots, X_{im}) &\propto h(\theta_i) L(X_{i1}, X_{i2}, \ldots, X_{im} \,|\, \theta_i) = \\ &= \theta_i^{k-1} \, e^{-\theta_i/\widehat{\sigma}} \, \theta_i^{m\overline{X}_i} \, e^{-m\theta_i} \, \mathbb{1}(\theta_i > 0) = \\ &= \theta_i^{k+m\overline{X}_i-1} \, e^{-\theta_i(\widehat{\sigma}^{-1}+m)} \, \mathbb{1}(\theta_i > 0) \end{aligned} \tag{4.30}$$

Окремо розпишемо множник експоненти:

$$e^{-\theta_i(\widehat{\sigma}^{-1}+m)} = e^{-\theta_i\left(\frac{1}{\widehat{\sigma}}+m\right)} = e^{-\theta_i\left(\frac{1+\widehat{\sigma}m}{\widehat{\sigma}}\right)} = e^{-\theta_i/\left(\frac{\widehat{\sigma}}{1+\widehat{\sigma}m}\right)} \tag{4.31}$$

Таким чином, впізнаємо у виразі (4.30) Гамма розподіл (4.5) із відповідними параметрами:

$$\theta_i \,|\, X_{i1}, X_{i2}, \ldots, X_{im} \sim \text{Gamma}\left(k + m\overline{X}_i, \frac{\widehat{\sigma}}{1+\widehat{\sigma}m}\right) \tag{4.32}$$

А тоді відповідно до (4.6), математичне сподівання апостеріорного розподілу (тобто баєсова оцінка параметра θ_i) буде такою:

$$\mathbb{E}(\theta_i \,|\, X_{i1}, X_{i2}, \ldots, X_{im}) = \left(k + m\overline{X}_i\right)\left(\frac{\widehat{\sigma}}{1+\widehat{\sigma}m}\right) \tag{4.33}$$

Враховуючи оцінку максимальної правдоподібності $\widehat{\sigma} = \overline{X}/k$ (4.25), матимемо:

$$\mathbb{E}(\theta_i \mid X_{i1}, X_{i2}, \ldots, X_{im}) = m\left(k/m + \overline{X}_i\right)\left(\frac{\overline{X}}{k + \overline{X}m}\right) = \left(k/m + \overline{X}_i\right)\left(\frac{\overline{X}}{k/m + \overline{X}}\right), \tag{4.34}$$

де позначено

$$\overline{X}_i = \frac{1}{m}\sum_{j=1}^{m} X_{ij} \text{ та } \overline{X} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij} \tag{4.35}$$

Коментар. Під час виведення MLE (4.25) для параметра σ, інтеграл у формулі (4.16) був обчислений без використання інформації про від'ємно біноміально розподілені дані. Насправді, я досі не до кінця усвідомив, яким чином слід було б використати цей факт у зв'язці із параметром σ без заданих параметрів розподілу r та p. Можливо, я недогледів певне перетворення між розподілами.

Задача 5. Gibbs Sampler for Gamma-Poisson model

Нехай задано Баєсову модель такого вигляду:

$$X_{ij} \,|\, \theta_i \overset{\text{ind}}{\sim} \text{Poiss}(\theta_i),\ i = \overline{1,n},\ j = \overline{1,m}, \tag{5.1}$$

$$\theta_i \,|\, \sigma \overset{\text{i.i.d.}}{\sim} \text{Gamma}(k,\sigma),\ i = \overline{1,n}, \tag{5.2}$$

$$\sigma^{-1} \sim \text{Exp}(1), \tag{5.3}$$

при цьому k є відомим. Іншими словами, для кожної згенерованої випадкової величини σ з експоненціального розподілу генерується значення θ_i з Гамма-розподілу, і надалі — набір випадкових величин $X_{i1},\dots,X_{im}$ з розподілу Пуасона з параметром θ_i.

Завдання (a): Gibbs Sampler

Завдання полягає у тому, щоб навести повні викладки одного кроку вибірки Гіббса для генерування апостеріорних розподілів параметрів θ_i та σ. Однак, перш ніж переходити до алгоритму, наведемо явний вигляд розподілів, які розглядатимуться у подальших міркуваннях.

Дискретна випадкова величина ξ має розподіл Пуасона з параметром λ, якщо:

$$\xi \sim Poiss(\lambda) \iff P(\xi = x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{5.4}$$

Неперервна випадкова величина η має Гамма-розподіл з «shape parameter» $k > 0$ та «scale parameter» $\sigma > 0$, якщо:

$$\eta \sim \text{Gamma}(k,\sigma) \iff f_\eta(x) = \frac{x^{k-1}}{\sigma^k \Gamma(k)}\, e^{-x/\sigma}\, \mathbb{1}(x>0) \tag{5.5}$$

Кажуть, що неперервна випадкова величина η^{-1} має обернений Гамма-розподіл з параметрами $k > 0$ та $\sigma > 0$, якщо:

$$\eta^{-1} \sim \text{IG}(k,\sigma) \iff f_{\eta^{-1}}(x) = \frac{x^{-k-1}}{\Gamma(k)}\, \sigma^k e^{-\sigma/x}\, \mathbb{1}(x>0) \tag{5.6}$$

Наостанок наведемо вигляд щільности експоненціального розподілу з параметром λ для неперервної випадкової величини ζ :

$$\zeta \sim \text{Exp}(\lambda) \iff f_\zeta(x) = \lambda e^{-\lambda x}\, \mathbb{1}(x>0), \tag{5.7}$$

при цьому випадкова величина ζ^{-1} матиме щільність виду

$$\zeta^{-1} \sim \text{Exp}(\lambda) \iff f_{\zeta^{-1}}(x) = -\lambda x^{-2} e^{-\lambda/x}\, \mathbb{1}(x>0) \tag{5.8}$$

Тож нехай ініційовано пару значень $\left(\sigma^{(0)}; \theta_1^{(0)}, \ldots, \theta_n^{(0)}\right)$. Тоді один крок вибірки Гіббса складатиметься із таких пунктів:

1. Використовуючи значення $\sigma^{(0)}$, згенерувати значення $\theta_i^{(1)}$ з так званого «full conditional distribution for parameter θ_i» — $f(\theta_i \,|\, \sigma^{(0)}; X_{i1}, \ldots, X_{im})$;

2. Маючи значення $\theta_i^{(1)}$, згенерувати значення $\sigma^{(1)}$ з «full conditional distribution for parameter σ» — $f(\sigma \,|\, \theta_1^{(1)}, \ldots, \theta_n^{(1)}; X_{11}, \ldots, X_{1m}; \ldots; X_{n1}, \ldots, X_{nm})$;

Зауважимо, що повний умовний розподіл одного параметра пропорційний сумісному розроділу, в якому інший параметр вважається фіксованим. Продемонструємо це, застосувавши ланцюгове правило:

$$\underbrace{f(\theta_i, \sigma \,|\, X_{i1}, \ldots, X_{im})}_{\text{сумісний розподіл}} = \underbrace{f(\theta_i \,|\, \sigma; X_{i1}, \ldots, X_{im})}_{\text{повний умовний розподіл}} \underbrace{f(\sigma \,|\, X_{i1}, \ldots, X_{im})}_{\text{не залежить від } \theta_i} \tag{5.9}$$

Таким чином переконуємося, що повний умовний розподіл як фунція від θ_i пропорційний сумісному розподілу, в якому значення σ покладено фіксованим. Аналогічним чином для параметра σ:

$$\underbrace{f(\theta_i, \sigma \,|\, X_{i1}, \ldots, X_{im})}_{\text{сумісний розподіл}} = \underbrace{f(\sigma \,|\, \theta_i; X_{i1}, \ldots, X_{im})}_{\text{повний умовний розподіл}} \underbrace{f(\theta_i \,|\, X_{i1}, \ldots, X_{im})}_{\text{не залежить від } \sigma} \tag{5.10}$$

Отже, опишемо один крок вибірки Гіббса:

1. Маючи $\sigma^{(0)}$, згенеруємо значення $\theta_i^{(1)}$ із відповідного умовного розподілу. Послідовно скористаємося такими викладками: щойно продемонтрованою властивістю (5.9) у переході (1), формулою Баєса у переході (2), ланцюговим правилом у переході (3) та незалежністю значень вибірки даних у кроці (4):

$$\begin{aligned} f(\theta_i \,|\, \sigma^{(0)}; X_{i1}, \ldots, X_{im}) &\overset{(1)}{\propto} f(\theta_i, \sigma^{(0)} \,|\, X_{i1}, \ldots, X_{im}) \propto \\ &\overset{(2)}{\propto} f(X_{i1}, \ldots, X_{im} \,|\, \theta_i, \sigma^{(0)}) f(\theta_i, \sigma^{(0)}) = \\ &\overset{(3)}{=} f(X_{i1}, \ldots, X_{im} \,|\, \theta_i, \sigma^{(0)}) f(\theta_i \,|\, \sigma^{(0)}) f(\sigma^{(0)}) = \\ &\overset{(4)}{=} \prod_{j=1}^{m} f(X_{ij} \,|\, \theta_i, \sigma^{(0)}) f(\theta_i \,|\, \sigma^{(0)}) f(\sigma^{(0)}) \end{aligned} \tag{5.11}$$

З огляду на вигляд функцій щільностей заданої Баєсової моделі (5.1) – (5.3), матимемо:

$$\begin{aligned} f(\theta_i \,|\, \sigma^{(0)}; X_{i1}, \ldots, X_{im}) \propto \prod_{j=1}^{m} \frac{\theta_i^{X_{ij}} e^{-\theta_i}}{X_{ij}!} \times \frac{\theta_i^{k-1}}{\left[\sigma^{(0)}\right]^k \Gamma(k)} e^{-\theta_i/\sigma^{(0)}} \mathbb{1}(\theta_i > 0) \times \\ \times \left[\sigma^{(0)}\right]^{-2} e^{-1/\sigma^{(0)}} \mathbb{1}(\sigma^{(0)} > 0) \end{aligned} \tag{5.12}$$

Відкидаючи множники, які не мають фунцкіональної залежності від θ_i, отримуємо такий вираз:

$$f(\theta_i \,|\, \sigma^{(0)}; X_{i1}, \ldots, X_{im}) \propto \theta_i^{m\overline{X}_i+k-1}\, e^{-\theta_i(m+1/\sigma^{(0)})}\, \mathbb{1}(\theta_i > 0) \tag{5.13}$$

А відтак, аналогічно до перетворень у формулі (4.30) з'ясовуємо вигляд апостеріорного розподілу параметра θ_i :

$$\theta_i \,|\, \sigma^{(0)}; X_{i1}, \ldots, X_{im} \sim \mathrm{Gamma}\left(k + m\overline{X}_i, \frac{\sigma^{(0)}}{1+\sigma^{(0)}m}\right) \tag{5.14}$$

Отже, використовуючи $\sigma^{(0)}$, генеруємо значення $\theta_i^{(1)}$ із розподілу (5.14).

2. Маючи $\theta_i^{(1)}$, згенеруємо значення $\sigma^{(1)}$ із відповідного повного умовного розподілу. В силу аналогічних кроків, як це показано для виведення формули (5.11), матимемо:

$$\begin{aligned} &f(\sigma \,|\, \theta_1^{(1)}, \ldots, \theta_n^{(1)}; X_{11}, \ldots, X_{1m}; \ldots; X_{n1}, \ldots, X_{nm}) \propto \\ &\quad \propto f(X_{11}, \ldots, X_{1m}; \ldots; X_{n1}, \ldots, X_{nm} \,|\, \theta_1^{(1)}, \ldots, \theta_n^{(1)}; \sigma) f(\theta_1^{(1)}, \ldots, \theta_n^{(1)} \,|\, \sigma) f(\sigma), \end{aligned} \tag{5.15}$$

а отже, зважаючи на незалежність випадкових величин $\theta_1^{(1)}, \ldots, \theta_n^{(1)}$, отримаємо

$$\begin{aligned} &f(\sigma \,|\, \theta_1^{(1)}, \ldots, \theta_n^{(1)}; X_{11}, \ldots, X_{1m}; \ldots; X_{n1}, \ldots, X_{nm}) \propto \\ &\qquad \propto \prod_{i=1}^{n} f(X_{i1}, \ldots, X_{im} \,|\, \theta_i^{(1)}, \sigma) \prod_{i=1}^{n} f(\theta_i^{(1)} \,|\, \sigma) f(\sigma), \end{aligned} \tag{5.16}$$

і як наслідок властивості вибірки даних:

$$\begin{aligned} &f(\sigma \,|\, \theta_1^{(1)}, \ldots, \theta_n^{(1)}; X_{11}, \ldots, X_{1m}; \ldots; X_{n1}, \ldots, X_{nm}) \propto \\ &\qquad \propto \prod_{i=1}^{n}\prod_{j=1}^{m} f(X_{ij} \,|\, \theta_i^{(1)}, \sigma) \prod_{i=1}^{n} f(\theta_i^{(1)} \,|\, \sigma) f(\sigma) \end{aligned} \tag{5.17}$$

Відтак, з огляду на вигляд функцій щільностей заданої Баєсової моделі, вираз розписуватиметься як:

$$\begin{aligned} &f(\sigma \,|\, \theta_1^{(1)}, \ldots, \theta_n^{(1)}; X_{11}, \ldots, X_{1m}; \ldots; X_{n1}, \ldots, X_{nm}) \propto \\ &\propto \prod_{i=1}^{n}\prod_{j=1}^{m} \frac{\left[\theta_i^{(1)}\right]^{X_{ij}} e^{-\theta_i^{(1)}}}{X_{ij}!} \times \prod_{i=1}^{n} \frac{\left[\theta_i^{(1)}\right]^{k-1}}{\sigma^k \Gamma(k)}\, e^{-\theta_i^{(1)}/\sigma} \mathbb{1}(\theta_i^{(1)} > 0) \times \\ &\qquad \times \sigma^{-2} e^{-1/\sigma} \mathbb{1}(\sigma > 0) \end{aligned} \tag{5.18}$$

Відкидаючи множники, які не мають фунцкіональної залежності від σ, отримуємо такий вираз:

$$f(\sigma \,|\, \theta_1^{(1)}, \ldots, \theta_n^{(1)}; X_{11}, \ldots, X_{1m}; \ldots; X_{n1}, \ldots, X_{nm}) \propto \\ \propto \sigma^{-nk-2}\, e^{-(n\overline{\theta_i^{(1)}}+1)/\sigma}\, \mathbb{1}(\sigma > 0) \quad (5.19)$$

У виразі, наведеному вище, впізнаємо обернений Гамма-розподіл (5.6):

$$\sigma \,|\, \theta_i^{(1)}; X_{11}, \ldots, X_{nm} \sim \mathrm{IG}\left(nk + 1, n\overline{\theta_i^{(1)}} + 1\right) \quad (5.20)$$

Отже, використовуючи набір $\theta_i^{(1)}$, генеруємо значення $\sigma^{(1)}$ із розподілу (5.20).