

# Домашнє завдання №1

# Зміст

<b>Задача 2. MCMC algorithms</b>	<b>2</b>
Завдання (a): MCMC algorithm . . . . .	2
Завдання (b): Gibbs Sampler . . . . .	3
<b>Задача 3. Empirical Bayes for exponential families</b>	<b>5</b>
Постановка задачі . . . . .	5
Визначення математичного сподівання . . . . .	6
Визначення баєсової оцінки вектора $\theta$ . . . . .	7
<b>Задача 4. Gamma-Poisson empirical Bayes</b>	<b>11</b>
Постановка задачі . . . . .	11
Допоміжні теоретичні викладки . . . . .	11
Визначення MLE для параметра $\hat{\sigma}$ . . . . .	13
Визначення баєсової оцінки параметра $\theta_i$ . . . . .	15
<b>Задача 5. Gibbs Sampler for Gamma-Poisson model</b>	<b>17</b>
Завдання (a): Gibbs Sampler . . . . .	17
Завдання (b): R-base Gibbs Sampler . . . . .	20
Завдання (c): R-jags Gibbs Sampler . . . . .	25

## Задача 2. MCMC algorithms

### Завдання (а): MCMC algorithm

Нехай  $\lambda(\theta)$  — деякий апостеріорний розподіл невідомого параметра  $\theta \in \mathbb{R}$ . Алгоритм MCMC має на меті згенерувати такий ланцюг Маркова  $\{\theta^{(t)}\}_{t \geq 0}$  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  $q(y | x)$  :

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

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

- (а) Згенерувати кандидата  $v \sim q(v | u)$ , де  $u = \theta^{(t-1)}$ ;
- (б) Обчислити імовірність прийняття  $\alpha(v | u)$  кандидата  $v$  :

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

- (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  $\{\theta^{(t)}\}_{t \geq 0}$ .

### Розв'язок

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

$$\forall y \in \mathbb{R} : \pi(y) = \int_{\mathbb{R}} Q(y | x) \pi(x) dx \quad (2.3)$$

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

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

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

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

$$\pi(x) Q(y | x) = \pi(y) Q(x | y), \quad (2.5)$$

тоді проінтегрувавши обидві частини рівняння, матимемо

$$\int_{\mathbb{R}} \pi(x) Q(y | x) dx = \int_{\mathbb{R}} \pi(y) Q(x | y) dx, \quad (2.6)$$

а отже, згідно властивостей розподілу станів отримуємо

$$\int_{\mathbb{R}} \pi(x) Q(y | x) dx = \int_{\mathbb{R}} \pi(y) Q(x | y) dx = \pi(y) \quad (2.7)$$

Таким чином, в силу Озн. 1 розподіл  $\pi$  є стаціонарним.  $\square$

Алгоритм МСМС будує ланцюг Маркова з розподілом станів такого вигляду:

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

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

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

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

## Завдання (b): Gibbs Sampler

Нехай  $\lambda(\theta)$  — апостеріорний розподіл вектора параметрів  $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ . 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  $\alpha$  (яка обчислюється з огляду на заданий апостеріорний розподіл). При цьому, фактично, апостеріорний розподіл може включати й інші параметри, які в рамках МСМС 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  $\vec{X} = (X_1, \dots, X_n)$ , який коротко позначимо як  $X$ , з  $s$ -параметризованої експоненціальної сім'ї розподілів з вектором параметрів  $\vec{\theta} = (\theta_1, \dots, \theta_s)$ , у короткому записі  $\theta$ . Тоді щільність розподілу вектора  $X$  матиме вигляд:

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

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

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

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

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

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

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

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

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

$$\hat{\gamma} = \operatorname{argmax}_{\gamma} L(X | \gamma) = \operatorname{argmax}_{\gamma} q_\gamma(x), \quad (3.5)$$

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

$$\mathbb{E}_{\hat{\gamma}} [\theta | X = x] = \nabla_x (\ln q_{\hat{\gamma}}(x) - \ln h(x)), \quad (3.6)$$

де оператор набла визначений так:

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

## Визначення математичного сподівання

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

$$(X_1, \dots, X_n) | (\theta_1, \dots, \theta_s) \sim p_\theta(x) \quad (3.8)$$

$$(\theta_1, \dots, \theta_s) \sim \lambda_\gamma(\theta) \quad (3.9)$$

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

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

Прологарифмуємо отриманий вираз:

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

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

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

Таким чином, отримуємо співвідношення для шуканої суми:

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

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

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

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

$$\mathbb{E} [g(\xi) | \eta] \stackrel{\text{def}}{=} \int_{\mathbb{R}} g(y) f_{\xi | \eta}(y | \eta) dy \quad (3.15)$$

Таким чином:

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

У викладах 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):

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

Отже, математичне сподівання (3.14) отримає шуканий вигляд:

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

## Визначення баєсової оцінки вектора $\theta$

Нехай тепер задано баєсівську модель

$$(X_1, \dots, X_n) | (\theta_1, \dots, \theta_s) \sim p_\theta(x) \quad (3.19)$$

$$(\theta_1, \dots, \theta_n) \sim \lambda_{\hat{\gamma}}(\theta) \quad (3.20)$$

над  $n$ -параметризованою експоненціальною сім'єю розподілів

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

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



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

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

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

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

Оскільки розмірність вектора  $\theta$  збігається із розмірністю вектора  $X$ , застосуємо до (3.23) оператор набла  $\nabla_x$  (3.7):

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

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

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

$$\nabla_x (\theta T(x)) = \nabla_x (\theta) T(x) + \theta \nabla_x (T(x)) = \theta \nabla_x (T(x)), \quad (3.25)$$

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

$$\nabla_x (\theta T(x)) = \theta \quad (3.26)$$

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

$$\theta = \nabla_x \ln \lambda_{\hat{\gamma}}(\theta | x) - \nabla_x \ln \lambda_{\hat{\gamma}}(\theta) + \nabla_x (\ln q_{\hat{\gamma}}(x) - \ln h(x)) \quad (3.27)$$

Аналогічно до (3.14) обчислимо умовне математичне сподівання:

$$\begin{aligned} \mathbb{E}_{\hat{\gamma}} [\theta | X = x] &= \mathbb{E}_{\hat{\gamma}} [\nabla_x \ln \lambda_{\hat{\gamma}}(\theta | x) | X = x] - \mathbb{E}_{\hat{\gamma}} [\nabla_x \ln \lambda_{\hat{\gamma}}(\theta) | X = x] + \\ &\quad + \nabla_x (\ln q_{\hat{\gamma}}(x) - \ln h(x)) \end{aligned} \quad (3.28)$$

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

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

Розпишемо другий доданок математичного сподівання (3.28):

$$\mathbb{E}_{\hat{\gamma}} [\nabla_x \ln \lambda_{\hat{\gamma}}(\theta) | X = x] = \int_{\mathbb{R}^n} \nabla_x \ln \lambda_{\hat{\gamma}}(\theta) \lambda_{\hat{\gamma}}(\theta | x) d\theta = \int_{\mathbb{R}^n} \frac{\nabla_x \lambda_{\hat{\gamma}}(\theta)}{\lambda_{\hat{\gamma}}(\theta)} \lambda_{\hat{\gamma}}(\theta | x) d\theta \tag{3.30}$$

За правилом диференціювання складеної функції:

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

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

$$\mathbb{E}_{\hat{\gamma}} [\nabla_x \ln \lambda_{\hat{\gamma}}(\theta) | X = x] = \frac{\partial \gamma}{\partial x} \int_{\mathbb{R}^n} \frac{\partial \lambda_{\hat{\gamma}}(\theta)}{\partial \gamma} \cdot \frac{\lambda_{\hat{\gamma}}(\theta | x)}{\lambda_{\hat{\gamma}}(\theta)} d\theta \tag{3.32}$$

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

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

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

$$q_{\hat{\gamma}}(x) = \int_{\mathbb{R}^n} p_{\theta}(x) \lambda_{\hat{\gamma}}(\theta) d\theta \tag{3.34}$$

Як наслідок, матимемо:

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

Наостанок, за формулою Баєса (3.22):

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

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

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

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

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

Отже, остаточний вигляд баєсової оцінки вектора  $\theta$  (3.28) записуватиметься таким формульним співвідношенням:

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

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

## Задача 4. Gamma-Poisson empirical Bayes

### Постановка задачі

Розглянемо модель такого вигляду:

$$X_{ij} | \theta_i \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\theta_i), \quad i = \overline{1, n}, \quad j = \overline{1, m}, \quad (4.1)$$

$$\theta_i \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(k, \hat{\sigma}), \quad i = \overline{1, n}, \quad (4.2)$$

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

$$\begin{aligned} \theta_1 &: X_{11}, X_{12}, \dots, X_{1m} \\ \theta_2 &: X_{21}, X_{22}, \dots, X_{2m} \\ &\dots \\ \theta_n &: X_{n1}, X_{n2}, \dots, X_{nm} \end{aligned} \quad (4.3)$$

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

### Допоміжні теоретичні викладки

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

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

На рівні позначень існує дві еквівалентні параметризації  $\alpha$  та  $\beta$ :

- «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»  $\beta = \sigma$ :

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

де, відповідно, математичне сподівання та дисперсія складають

$$\mathbb{E}\xi = k\sigma, \quad \text{Var}\xi = k\sigma^2 \quad (4.6)$$

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

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

де математичне сподівання та дисперсія

$$\mathbb{E}\xi = \frac{1}{\sigma}, \quad \text{Var}\xi = \frac{1}{\sigma^2} \quad (4.8)$$

**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  $\eta$  має розподіл Пуасона з параметром  $\lambda$ , якщо:

$$\eta \sim \text{Pois}(\lambda) \iff P(\eta = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (4.9)$$

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

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

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

$$P(A) = \sum_i P(A | H_i) P(H_i) \quad (4.11)$$

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

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

## Визначення MLE для параметра $\hat{\sigma}$

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

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

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

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

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

$$f(X_{ij} | \sigma) = \int_{\mathbb{R}} f(X_{ij} | \theta_i, \sigma) f(\theta_i | \sigma) d\theta_i \quad (4.15)$$

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

$$\begin{aligned} f(X_{ij} | \sigma) &= \int_{\mathbb{R}} f(X_{ij} | \theta_i, \sigma) f(\theta_i | \sigma) d\theta_i = \\ &= \int_{\mathbb{R}} \frac{\theta_i^{X_{ij}} e^{-\theta_i}}{X_{ij}!} \cdot \frac{\theta_i^{k-1}}{\sigma^k \Gamma(k)} e^{-\theta_i/\sigma} \mathbb{1}(\theta_i > 0) d\theta_i = \\ &= C \sigma^{-k} \int_0^{\infty} \theta_i^{X_{ij}+k-1} e^{-\theta_i(1+\frac{1}{\sigma})} d\theta_i \end{aligned} \quad (4.16)$$

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

$$\Gamma(a) = \int_0^{\infty} y^{a-1} e^{-y} dy \quad (4.17)$$

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

$$\Gamma(a + 1) = \int_0^{\infty} y^a e^{-y} dy = \int_0^{\infty} (bx)^a e^{-bx} b dx = b^{a+1} \int_0^{\infty} x^a e^{-bx} dx \quad (4.18)$$

Як наслідок, матимемо:

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

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

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

Тоді функція правдоподібності (4.14) матиме такий остаточний вигляд:

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

де позначено

$$\bar{X} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m X_{ij} \quad (4.22)$$

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

$$\frac{d}{d\sigma} \ln L = 0 \quad (4.23)$$

призведе до оцінки максимальної правдоподібності параметра  $\sigma$  :

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

а відтак матимемо

$$\frac{d}{d\sigma} \ln L = 0 \iff \hat{\sigma} = \frac{\bar{X}}{k} \quad (4.25)$$

## Визначення баєсової оцінки параметра $\theta_i$

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

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

Отже, апіорний розподіл має вигляд:

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

Функція правдоподібності, своєю чергою, розписуватиметься таким чином:

$$L(X_{i1}, X_{i2}, \dots, X_{im} | \theta_i) \stackrel{\text{ind}}{=} \prod_{j=1}^m P(X_{ij} | \theta_i) = \prod_{j=1}^m \frac{\theta_i^{X_{ij}} e^{-\theta_i}}{X_{ij}!} = \frac{\theta_i^{\sum_{j=1}^m X_{ij}} e^{-m\theta_i}}{X_{i1}! X_{i2}! \dots X_{im}!} \quad (4.28)$$

Отже, матимемо:

$$L(X_{i1}, X_{i2}, \dots, X_{im} | \theta_i) = C \theta_i^{m\bar{X}_i} e^{-m\theta_i} \quad (4.29)$$

Відтак апостеріорний розподіл матиме вигляд:

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

Окремо розпишемо множник експоненти:

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

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

$$\theta_i | X_{i1}, X_{i2}, \dots, X_{im} \sim \text{Gamma} \left( k + m\bar{X}_i, \frac{\hat{\sigma}}{1 + \hat{\sigma}m} \right) \quad (4.32)$$

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

$$\mathbb{E}(\theta_i | X_{i1}, X_{i2}, \dots, X_{im}) = (k + m\bar{X}_i) \left( \frac{\hat{\sigma}}{1 + \hat{\sigma}m} \right) \quad (4.33)$$



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

$$\mathbb{E}(\theta_i | X_{i1}, X_{i2}, \dots, X_{im}) = m (k/m + \bar{X}_i) \left( \frac{\bar{X}}{k + \bar{X}m} \right) = (k/m + \bar{X}_i) \left( \frac{\bar{X}}{k/m + \bar{X}} \right), \quad (4.34)$$

де позначено

$$\bar{X}_i = \frac{1}{m} \sum_{j=1}^m X_{ij} \text{ та } \bar{X} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m X_{ij} \quad (4.35)$$

**Коментар.** Під час виведення MLE (4.25) для параметра  $\sigma$ , 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  $\sigma$  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:

$$X_{ij} | \theta_i \stackrel{\text{i.i.d.}}{\sim} \text{Pois}(\theta_i), \quad i = \overline{1, n}, \quad j = \overline{1, m}, \quad (5.1)$$

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

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

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

### Завдання (а): Gibbs Sampler

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

Дискретна випадкова величина  $\xi$  має розподіл Пуасона з параметром  $\lambda$ , якщо:

$$\xi \sim \text{Pois}(\lambda) \iff P(\xi = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (5.4)$$

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

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

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

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

Наостанок наведемо вигляд щільності експоненціального розподілу з параметром  $\lambda$  для неперервної випадкової величини  $\zeta^{-1}$ :

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

при цьому випадкова величина  $\zeta$  матиме щільність виду

$$f_{\zeta}(x) = \lambda x^{-2} e^{-\lambda/x} \mathbb{1}(x > 0) \quad (5.8)$$

Тож нехай ініційовано пару значень  $(\sigma^{(0)}; \theta_1^{(0)}, \dots, \theta_n^{(0)})$ . 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  $\theta_i$ » —  $f(\theta_i | \sigma^{(0)}; X_{i1}, \dots, X_{im})$ ;
2. Маючи значення  $\theta_i^{(1)}$ , згенерувати значення  $\sigma^{(1)}$  з «full conditional distribution for parameter  $\sigma$ » —  $f(\sigma | \theta_1^{(1)}, \dots, \theta_n^{(1)}; X_{11}, \dots, X_{1m}; \dots; X_{n1}, \dots, X_{nm})$ ;

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

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

Таким чином переконуємося, що повний умовний розподіл як функція від  $\theta_i$  пропорційний сумісному розподілу, в якому значення  $\sigma$  покладено фіксованим. Аналогічним чином для параметра  $\sigma$ :

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

Отже, опишемо один крок вибірки Гіббса:

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

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

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

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

Відкидаючи множники, які не мають функціональної залежності від  $\theta_i$ , отримуємо такий вираз:

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

А відтак, аналогічно до перетворень у формулі (4.30) з'ясуємо вигляд апостеріорного розподілу параметра  $\theta_i$ :

$$\theta_i | \sigma^{(0)}; X_{i1}, \dots, X_{im} \sim \text{Gamma} \left( k + m\bar{X}_i, \frac{\sigma^{(0)}}{1 + \sigma^{(0)}m} \right) \quad (5.14)$$

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:

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

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

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

і як наслідок властивості вибірки даних:

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

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

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

Відкидаючи множники, які не мають функціональної залежності від  $\sigma$ , отримуємо такий вираз:

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

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

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

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

## Завдання (b): R-base Gibbs Sampler

Виконаємо засобами мови R вибірку Гіббса згідно із викладками, наведеними у попередньому підрозділі. Вибірку даних (5.1) задамо з файлу `hw5.csv`. Покладемо  $n = k = 3$  та  $m = 100$ , тобто матимемо Баєсову модель вигляду

$$X_{ij} | \theta_i \stackrel{\text{ind}}{\sim} \text{Poiss}(\theta_i), \quad i = \overline{1, 3}, \quad j = \overline{1, 100}, \quad (5.21)$$

$$\theta_i | \sigma \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(3, \sigma), \quad i = \overline{1, 3}, \quad (5.22)$$

$$\sigma^{-1} \sim \text{Exp}(1) \quad (5.23)$$

Тож параметрами вибірки слугуватимуть значення  $\sigma$  та вектор  $(\theta_1, \theta_2, \theta_3)$ . Поклавши початкове значення  $\sigma^{(0)} = 1$ , проведемо 100 кроків для «розгону» алгоритму і фіксуватимемо наступні  $N = 10\,000$  кроків для визначення апостеріорного розподілу невідомих параметрів.

Наведемо нижче основні програмні блоки. До прикладу, у Лістингу 1 продемонстровано функцію для оновлення параметрів  $(\theta_1, \theta_2, \theta_3)$ .

Лістинг 1: Крок оновлення вектора параметрів  $(\theta_1, \theta_2, \theta_3)$

```

1 update_theta <- function(x, sigma, k, n, m) {
2   theta_values <- array(dim = n)
3
4   for (i in 1:n) {
5     shape_theta <- k + m * mean(x[, i])
6     scale_theta <- sigma / (1.0 + sigma * m)
7
8     theta_values[i] <- rgamma(1, shape = shape_theta, scale = scale_theta)
9   }
10
11   return(theta_values)
12 }
```

Нижче у Лістингу 2 продемонстровано програмну реалізацію оновлення значення  $\sigma$ .

### Лістинг 2: Крок оновлення параметра $\sigma$

```
1 update_sigma <- function(x, theta, k, n, m) {  
2   shape_sigma <- n * k + 1  
3   scale_sigma <- n * mean(theta) + 1  
4  
5   sigma_value <- 1 / rgamma(1, shape = shape_sigma, scale = 1 / scale_sigma)  
6  
7   return(sigma_value)  
8 }
```

Запуск основного алгоритму вибірки Гіббса (Лістинг 3) показано нижче у Лістингу 4.

### Лістинг 3: Імплементация алгоритму вибірки Гіббса

```
1 gibbs <- function(data, n_iter, init, prior, n, m) {  
2   # Initialize arrays to store variables  
3   sigma_out <- array(dim = c(n_iter, 1))  
4   theta_out <- array(dim = c(n_iter, 3))  
5  
6   sigma_now <- init$sigma  
7  
8   # Gibbs sampler  
9   for (i in 1:n_iter) {  
10    theta_now <- update_theta(  
11      x = data, sigma = sigma_now, k = prior$k, n = n, m = m  
12    )  
13    sigma_now <- update_sigma(  
14      x = data, theta = theta_now, k = prior$k, n = n, m = m  
15    )  
16  
17    theta_out[i, ] <- theta_now  
18    sigma_out[i, ] <- sigma_now  
19  }  
20  
21  cbind(sigma = sigma_out[, 1], theta1 = theta_out[, 1], theta2 = theta_out[, 2],  
22        theta3 = theta_out[, 3])  
23 }
```

### Лістинг 4: Запуск вибірки Гіббса

```
1 # Load the dataset  
2 x <- read.csv("hw5.csv")  
3  
4 # Set main parameters  
5 n <- ncol(x)  
6 m <- nrow(x)  
7  
8 prior <- list()  
9 prior$k <- 3  
10  
11 init <- list()
```

```

12 init$sigma <- 1.0
13
14 set.seed(53)
15 posterior <- gibbs(
16   data = x,
17   n_iter = 10e3 + 100,
18   init = init,
19   prior = prior,
20   n = n,
21   m = m
22 )
23
24 # Exclude first 100 iterations (burn-in period)
25 posterior <- tail(posterior, -100)

```

На Рис. 1 продемонстровано результати: «trace plot» отриманого ланцюга Маркова та апостеріорний розподіл параметрів  $\sigma$  та  $\theta_1$ . Перш ніж перейти до подальших висновків, проведемо огляд автокореляції утвореного ланцюга. Автокореляція як така вказує на міру лінійної залежності (в межах  $[-1, 1]$ ) поточного значення відносно попередніх значень (lags). З Рис. 2 бачимо, що автокореляція значень практично відсутня.

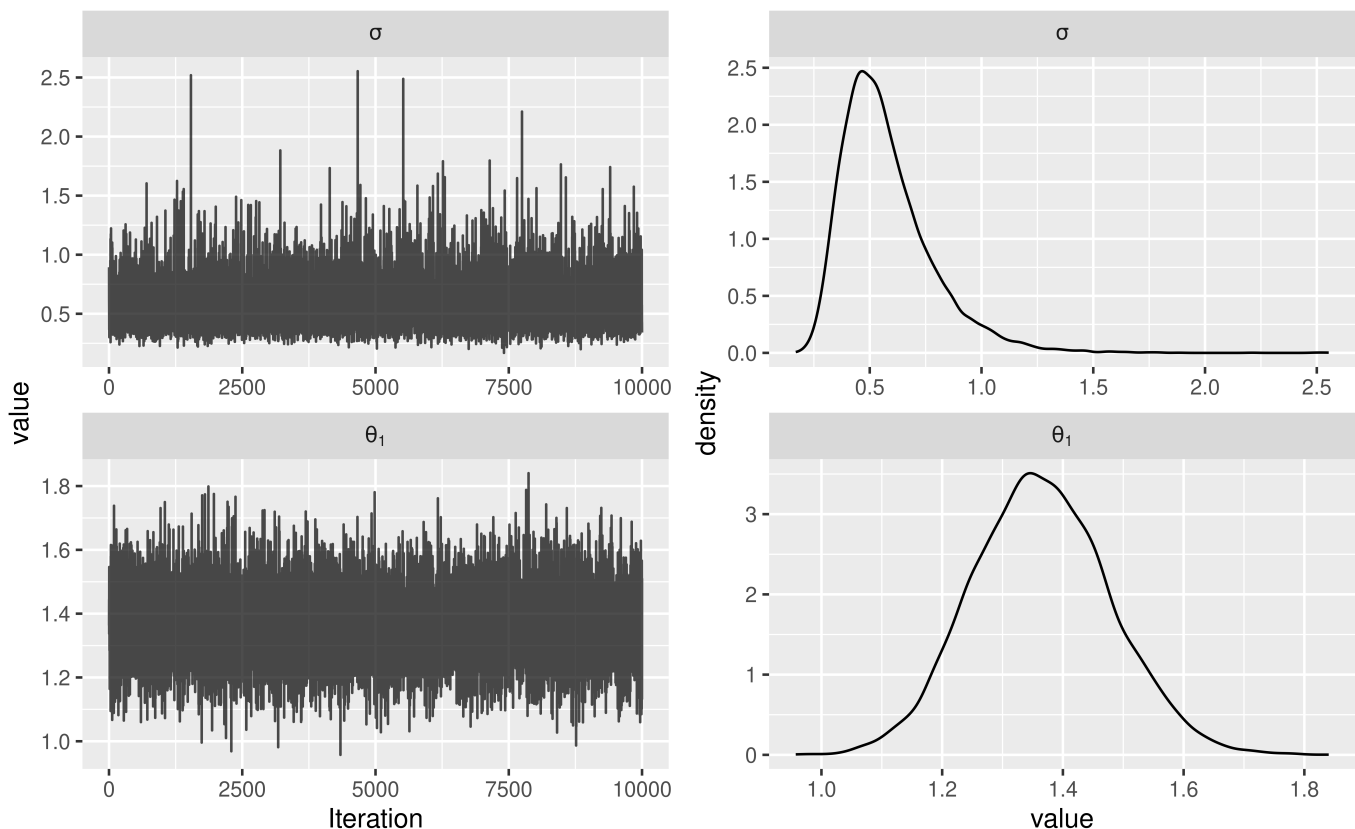


Рис. 1: Апостеріорний розподіл параметрів  $\sigma$  та  $\theta_1$  (засоби R-base)

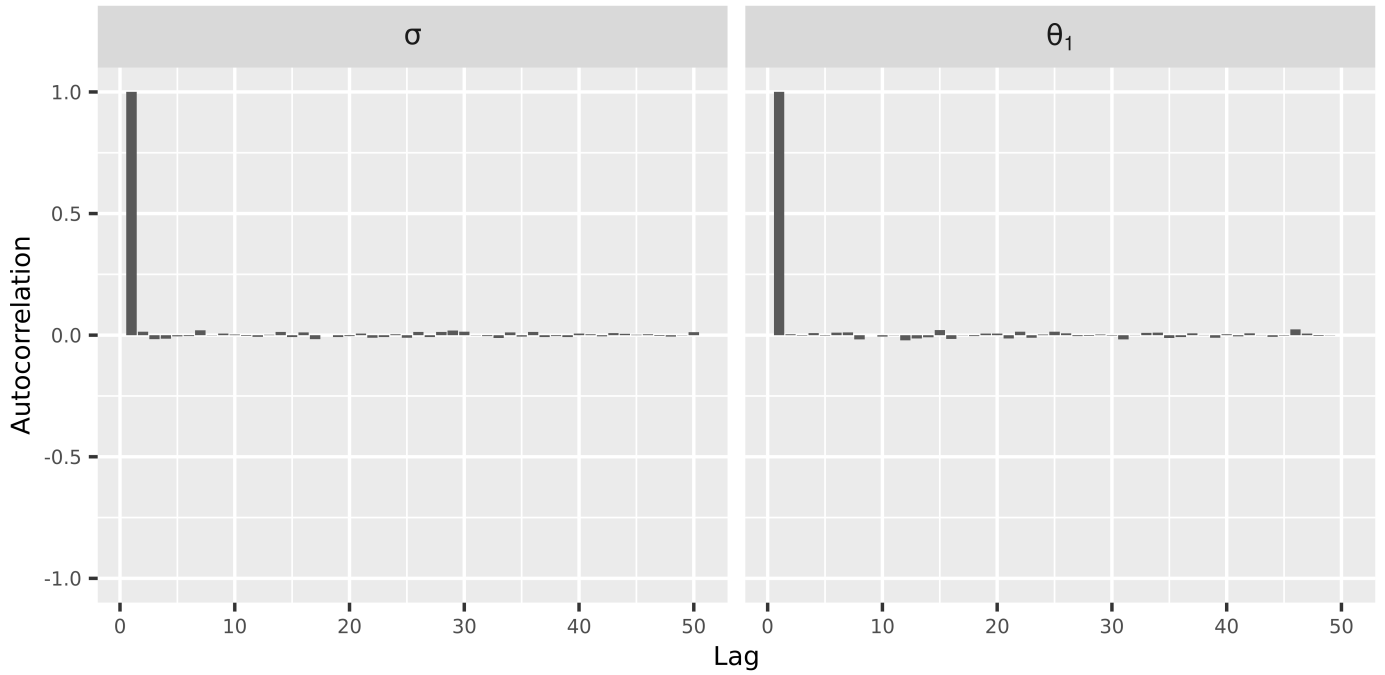


Рис. 2: Автокореляційний графік параметрів  $\sigma$  та  $\theta_1$  (засоби R-base)

Аналогічні висновки про слабку коеляцію можна зробити із табличного вигляду автокореляційного графіка (Табл. 1).

	$\sigma$	$\theta_1$
Lag 0	1.0000	1.0000
Lag 1	0.0141	0.0038
Lag 5	-0.0036	0.0099
Lag 10	-0.0027	-0.0009
Lag 50	-0.0018	0.0036

Таблиця 1: Автокореляційна таблиця параметрів  $\sigma$  та  $\theta_1$

У випадку ж сильної автокореляції значень у пригоді може стати виявлення так званого ефективного розміру вибірки (effective size), який вказуватиме, скільки значень у ланцюзі безпосередньо містять інформацію про стаціонарний розподіл. До прикладу, якщо згенеровано ланцюг довжиною  $N = 10\,000$ , а ефективний розмір вибірки складає  $N_{e.s.} = 1000$ , то це означає, що при обчисленні статистичних оцінок про істинний стаціонарний розподіл кожне  $N/N_{e.s.} = 100$ -те значення містить цінну інформацію, а інші значення є високо корельованими.



Для ланцюга, отриманого на Рис. 1,  $N_{e.s.}^{\sigma} = N_{e.s.}^{\theta_1} = 10\,000$ , що при довжині ланцюга  $N = 10\,000$  зайвий раз підкреслює відсутність автокореляції, адже кожне значення містить корисну інформацію про стаціонарний розподіл.

Таким чином, наостанок наведемо деякі статистичні характеристики отриманих апостеріорних розподілів з точністю до чотирьох значущих знаків (Табл. 2), зокрема вкажемо значення «класичної» Баєсової оцінки як умовного математичного сподівання.

Parameter	Empirical mean	Standard deviation	Standard error	2.5% - quantile	97.5% - quantile
$\sigma$	0.5768	0.2054	0.002054	0.3002	1.0800
$\theta_1$	1.3640	0.1139	0.001139	1.1478	1.5930

Таблиця 2: Результати імплементації вибірки Гіббса засобами R-base

У задачах статистичного аналізу для оцінки невідомого параметра імовірнісної моделі використовують не лише точкові, але й інтервальні оцінки. При цьому, у так званому частотному підході такою оцінкою слугує довірчий інтервал (confidence interval), а у Баєсівському підході — імовірний інтервал (credible interval). У декількох наступних абзацах коротко наведемо відмінності між цими поняттями.

Нехай задано вибірку  $X_1, \dots, X_n$  з деякого розподілу  $f(x; \mu)$ , де невідомим параметром слугує  $\mu$ . У частотному підході значення  $\mu$  вважається невідомим та фіксованим. У ролі точкової оцінки може бути використана довільна статистика (скажімо, вибіркове середнє), і окремим питанням стоятиме визначення властивостей такої оцінки (змістовність, незміщеність тощо). Водночас, судження щодо значень параметра  $\mu$  можна зробити на основі довірчого інтервалу зі статистик  $T_1(\vec{X})$  та  $T_2(\vec{X})$ :

$$P(T_1(\vec{X}) < \mu < T_2(\vec{X})) \geq \gamma \quad (5.24)$$

Інтервал (5.24) називають довірчим інтервалом рівня довіри  $\gamma$ . Отриманий інтервал є випадковим проміжком, який слід трактувати так: серед (нескінченних) спроб побудувати довірчий інтервал  $(T_1(\vec{x}), T_2(\vec{x}))$  для щоразу іншої реалізації вибірки  $\vec{x}$ , щонайменше  $\gamma\%$  з побудованих інтервалів міститимуть істинне значення невідомого параметра  $\mu$ . Тобто для конкретного наявного набору даних  $\vec{x}$  побудований інтервал або містить, або не містить істинний параметр. Іншими словами, значення  $\gamma$  має не імовірнісну інтерпретацію наявності/відсутності істинного значення параметра у побудованому довірчому інтервалі, а радше «рівнем впевненості» безпосередньо у статистичній процедурі побудови інтервалу.

На противагу частотному підходу, Баєсівська модель вважає параметр  $\mu$  невідомою випадковою величиною. Відтак, шляхом переоцінки деякого апіорного розподілу  $\mu \sim g(x)$  із залученням конкретного набору даних  $\vec{x}$ , судження про значення параметра виконують з отриманого переоціненого апостеріорного розподілу  $g^*(x)$ . Подальшим кроком може бути побудова точкової оцінки чи визначення імовірного інтервалу

$$P(q_1 < \mu < q_2) \stackrel{\text{def}}{=} \int_{q_1}^{q_2} g^*(x) dx, \quad (5.25)$$

який означатиме, власне, імовірність параметра  $\mu$  належати проміжку  $(q_1, q_2)$  при наявних даних  $\vec{x}$ .

Отже, повертаючись до Баєсової моделі (5.1) – (5.3), з огляду на результати побудованих апостеріорних розподілів (Табл. 2), 95%-імовірний інтервал для значень параметра  $\theta_1$  матиме вигляд:

$$P(1.1478 < \theta_1 < 1.5930) = 0.95 \quad (5.26)$$

Крім того, згідно вимог завдання порівняємо наведене значення зі значенням оцінки (4.34), отриманої в рамках Завдання 4:

$$\hat{\theta}_1 = (k/m + \bar{X}_1) \left( \frac{\bar{X}}{k/m + \bar{X}} \right) = 1.3608 \quad (5.27)$$

### Завдання (с): R-jags Gibbs Sampler

Тепер скористаємося засобами мови R, пакетом `rjags`, для реалізації вибірки Гіббса. Вибірку даних (5.1) задамо з файлу `hw5.csv`. Покладемо  $n = k = 3$  та  $m = 100$ , тобто матимемо Баєсову модель вигляду

$$X_{ij} | \theta_i \stackrel{\text{ind}}{\sim} \text{Pois}(\theta_i), \quad i = \overline{1, 3}, \quad j = \overline{1, 100}, \quad (5.28)$$

$$\theta_i | \sigma \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(3, \sigma), \quad i = \overline{1, 3}, \quad (5.29)$$

$$\sigma^{-1} \sim \text{Exp}(1) \quad (5.30)$$

Перш ніж рухатися далі, наголосимо, що параметр  $\sigma$  у Гамма-розподілі є «scale parameter», в той час як засаби `rjags` надають можливість задати модель виключно через «rate parameter». У такому разі, еквівалентний запис моделі матиме вид:

$$X_{ij} | \theta_i \stackrel{\text{ind}}{\sim} \text{Pois}(\theta_i), \quad i = \overline{1, 3}, \quad j = \overline{1, 100}, \quad (5.31)$$

$$\theta_i | \sigma \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(3, \sigma^{-1}), \quad i = \overline{1, 3}, \quad (5.32)$$

$$\sigma^{-1} \sim \text{Exp}(1) \quad (5.33)$$

Відтак, запис моделі через синтаксис `rjags` наведено на Лістингу 5, а безпосередній запуск алгоритму вказано у Лістингу 6.

#### Лістинг 5: Задана модель засобами `rjags`

```
1 jags_model_string <- "model {
2   for (i in 1:(n*m)) {
3     x[i] ~ dpois(theta[group[i]])
4   }
5
6   for (j in 1:n) {
7     theta[j] ~ dgamma(k, invsigma)
8   }
9
10  invsigma ~ dexp(1.0)
11 }"
```

#### Лістинг 6: Запуск вибірки Гіббса засобами `rjags`

```
1 # Convert three flows of data (x1,x2,x3) into one list and label each datapoint
   accordingly (group 1.0, 2.0 or 3.0)
2 x_jags <- array(dim = c(m * n, 2))
3 colnames(x_jags) <- c("x", "group")
4
5 x_jags[1:100, 1] <- x[, 1]
6 x_jags[1:100, 2] <- 1.0
7
8 x_jags[101:200, 1] <- x[, 2]
9 x_jags[101:200, 2] <- 2.0
10
11 x_jags[201:300, 1] <- x[, 3]
12 x_jags[201:300, 2] <- 3.0
13
14 data_jags <- list(
15   x = x_jags[, 1],
16   group = x_jags[, 2],
17   n = n,
18   m = m,
19   k = prior$k
20 )
21
22 model <- jags.model(
23   textConnection(jags_model_string),
24   data = data_jags,
25   n.chains = 3,
26 )
27
28 update(model, 100) # burn-in period
29
30 params <- c("theta", "invsigma")
31
32 model_simulation <- coda.samples(
33   model = model,
34   variable.names = params,
35   n.iter = 10e3
36 )
```

В рамках додаткового дослідження збіжності МСМС було запущено три ланцюги з різних початкових точок відліку. У першому наближенні з Рис. 3 можна зробити висновок, що бажана збіжність досягнута.

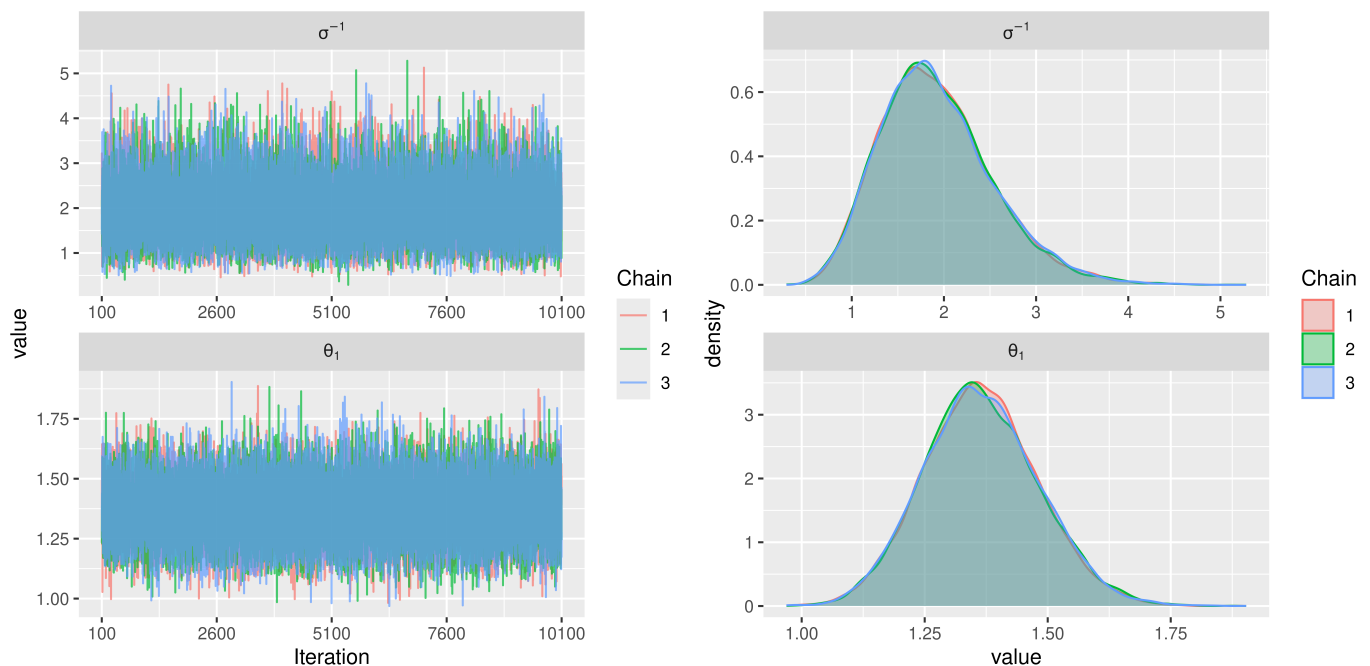


Рис. 3: Апостеріорний розподіл параметрів  $\sigma^{-1}$  та  $\theta_1$  (засоби R-jags)

Тим не менш, додатковим кроком проведемо діагностику Гелмана-Рубіна (Gelman-Rubin diagnostic), яка полягає у визначенні своєрідної міри розрізненості між декількома ланцюгами. Діагностика обчислює варіабельність (розкид, дисперсію) всередині ланцюгів, порівнюючи її з дисперсією між ланцюгами. Якщо всі ланцюги наблизилися до стаціонарного розподілу, то варіабельність між ланцюгами має бути відносно невеликою, а так званий «коефіцієнт потенційного зменшення масштабу» (potential scale reduction factor), який є результатом діагностики, має бути близьким до одиниці. Якщо значення значно перевищує одиницю, то можна зробити висновок, що ланцюги ще не збіглися за задану кількість ітерацій.

Бачимо (Рис. 4), що значення коефіцієнта близьке до одиниці як для параметра  $\sigma$ , так і для параметра  $\theta_1$ . Більш того, спостерігаємо, що бажана збіжність досягнута вже на початкових ітераціях, тому «період розгону» ланцюга (burn-in period) було обачно покласти невеликим: 100 перших ітерацій (Лістинг 6).

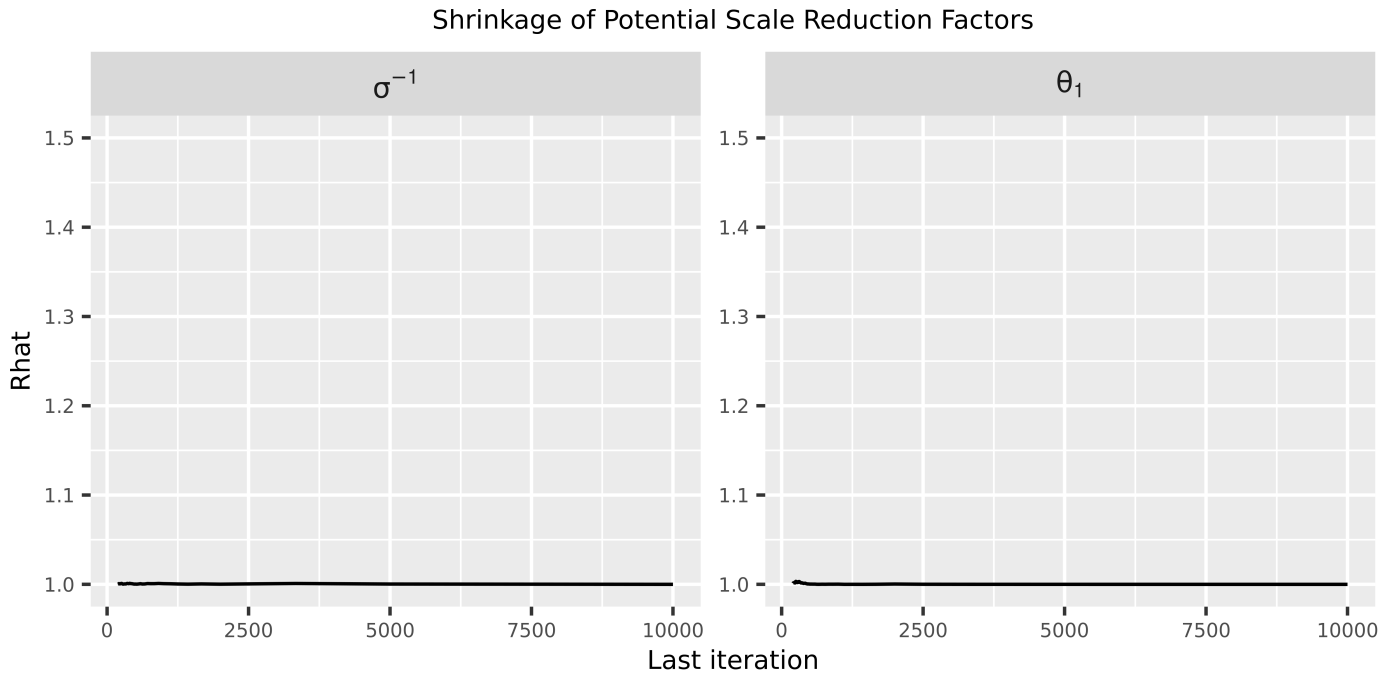


Рис. 4: Діагностика Гелмана-Рубіна для параметрів  $\sigma^{-1}$  та  $\theta_1$

Таким чином, у підсумку розглянемо перший з трьох ланцюгів, для якого дослідимо ознаки автокореляції (Рис. 5) та побудуємо Баєсівські оцінки (Табл. 3), які можна вважати змістовними та надійними, адже ефективний розмір вибірки (effective size) при довжині ланцюга  $N = 10\,000$  є значним:  $N_{e.s.}^{\sigma^{-1}} = N_{e.s.}^{\theta_1} = 10\,000$ .

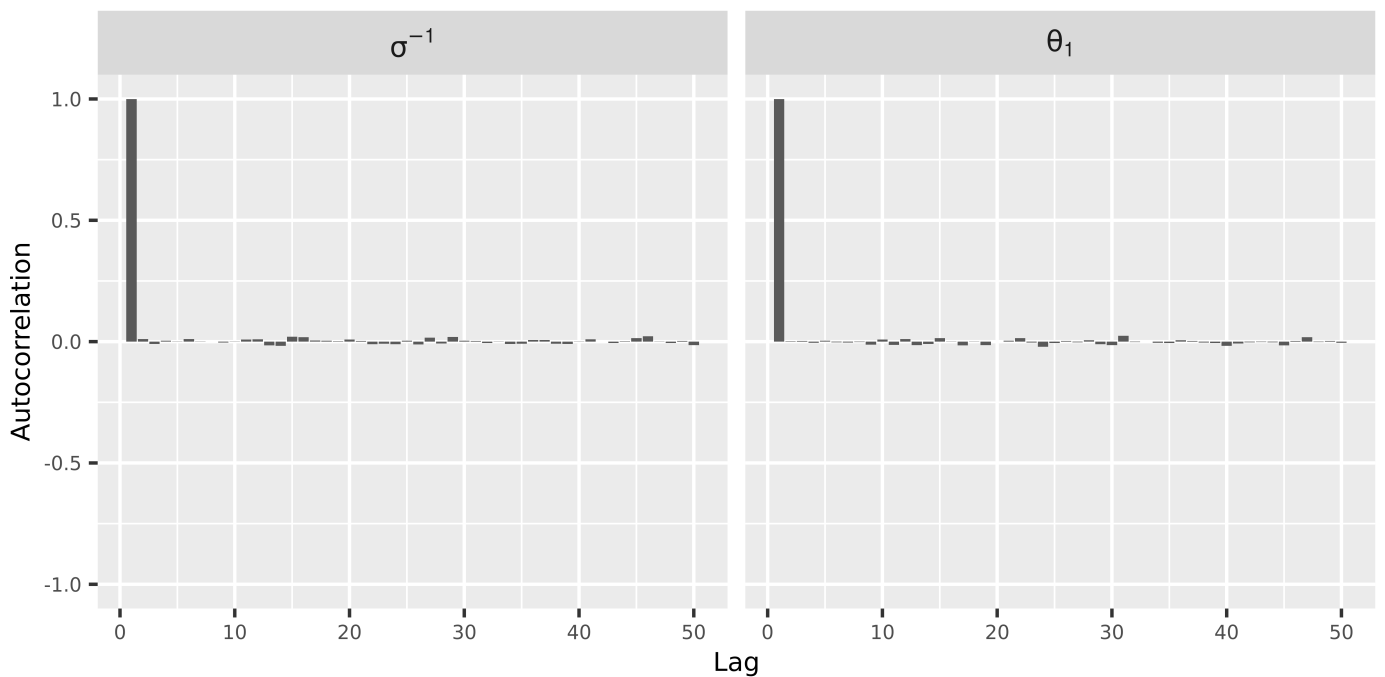


Рис. 5: Автокореляційний графік параметрів  $\sigma^{-1}$  та  $\theta_1$  (засоби R-jags)

Parameter	Empirical mean	Standard deviation	Standard error	2.5% - quantile	97.5% - quantile
$\sigma^{-1}$	0.9210	0.6130	0.006130	0.9131	1.2990
$\theta_1$	1.3640	0.1153	0.001153	1.1458	1.5970

Таблиця 3: Результати імплементації вибірки Гіббса засобами R-jags

Наостанок, 95%-імовірний інтервал для значень параметра  $\theta_1$  матиме вигляд:

$$P(1.1458 < \theta_1 < 1.5970) = 0.95, \quad (5.34)$$

що практично повторює отримані результати (5.26). Водночас, значення оцінки (4.34), отриманої в рамках Завдання 4, є таким:

$$\hat{\theta}_1 = (k/m + \bar{X}_1) \left( \frac{\bar{X}}{k/m + \bar{X}} \right) = 1.3608 \quad (5.35)$$