

# Домашнє завдання №5

# Зміст

<b>Задача 2. MCMC algorithms</b>	<b>2</b>
Завдання (a) . . . . .	2
Завдання (b) . . . . .	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

## Задача 2. MCMC algorithms

### Завдання (а)

Нехай  $\lambda(\theta)$  — деякий апостеріорний розподіл невідомого параметра  $\theta \in \mathbb{R}$ . Алгоритм МСМС 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 МСМС, 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)

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 апіорному 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:

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

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

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 Пуасона 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:

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

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.