Статистичні методи теорії ризику

© Ямненко Р.Є

КНУ імені Тараса Шевченка
механіко-математичний факультет
магістратура "Актуарна та фінансова математика"

I семестр 2012

Зміст

Баєсівське оцінювання параметрів

Задача оцінювання параметра – це задача прийняття рішень, у якій рішенням статистика є оцінка значень параметра $\mathbf{W} = (W_1, \ldots, W_k)$, що належить множині $\Theta \subset \mathbb{R}^k$, $k \geq 1$. Оскільки статистик оцінює значення $\mathbf{W}$, то множина рішень D співпадає з множиною значень параметра Θ. Припустимо для простоти, що $\Theta = D = \mathbb{R}^k$, маючи на увазі, що ймовірність того, що $\mathbf{W}$ лежить у деяких областях $\mathbb{R}^k$, може бути рівною 0.

Рішення статистика $\mathbf{d} = (d_1, \ldots, d_k) \in \mathbb{R}^k$ – це його оцінка значення $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)$ параметра $\mathbf{W}$, а його втрати $L(\boldsymbol{\nu}, \mathbf{d})$ відображають розбіжність між значенням $\boldsymbol{\nu}$ та оцінкою $\mathbf{d}$.

Тому в задачах оцінювання вважають, що функція втрат $L(\boldsymbol{\nu}, \mathbf{d})$ має вигляд

$$L(\boldsymbol{\nu}, \mathbf{d}) = \gamma(\boldsymbol{\nu})\Lambda(\boldsymbol{\nu} - \mathbf{d}), \tag{1}$$

де Λ – це невід'ємна функція від вектора похибок $(\boldsymbol{\nu} - \mathbf{d})$ така, що $\Lambda(0) = 0$, $\gamma(\boldsymbol{\nu})$ – невід'ємна вагова функція, яка визначає відносну значущість заданого вектора похибок для різних значень параметра $\mathbf{W}$. Якщо втрати $L(\boldsymbol{\nu}, \mathbf{d})$ залежать лише від вектора похибок $(\boldsymbol{\nu} - \mathbf{d})$, то можна вважати, що функція $\gamma(\boldsymbol{\nu})$ стала на всьому просторі $\mathbb{R}^k$.

Розглянемо задачу оцінювання, в якій функція втрат $L(\boldsymbol{\nu}, \mathbf{d})$ має вигляд (1), і припустимо, що апостеріорна щільність розподілу параметра $\mathbf{W}$ дорівнює $u_{\mathbf{x}}(\boldsymbol{\nu})$. Баєсівське рішення $\mathbf{d}^*$ або, у нашому випадку, баєсівську оцінку $\mathbf{d}^*$ визначають, як точку $\mathbf{d} \in \mathbb{R}^k$, в якій досягається мінімум ризику $R(\mathbf{d}, u_{\mathbf{x}})$:

$$R(\mathbf{d}, u_{\mathbf{x}}) = \int\limits_{\mathbb{R}^k} \gamma(\boldsymbol{\nu})\, \Lambda(\boldsymbol{\nu} - \mathbf{d}) u_{\mathbf{x}}(\boldsymbol{\nu}) d\boldsymbol{\nu}. \tag{2}$$

Зауважимо, що такий же інтеграл мінімізують і в задачі оцінювання, у якій функція втрат $\tilde{L}(\boldsymbol{\nu}, \mathbf{d})$ має вигляд $\tilde{L}(\boldsymbol{\nu}, \mathbf{d}) = \Lambda(\boldsymbol{\nu} - \mathbf{d})$, а апостеріорна щільність розподілу $\tilde{u}_{\mathbf{x}}$ параметра W задовольняє умові пропорційності $\tilde{u}_{\mathbf{x}}(\boldsymbol{\nu}) \propto \gamma(\boldsymbol{\nu}) u_{\mathbf{x}}(\boldsymbol{\nu})$.

Іншими словами, одні й ті самі баєсівські рішення отримуємо незалежно від того, чи є невід'ємна функція $\gamma(\boldsymbol{\nu})$ множником при функції втрат, чи вона є множником при щільності розподілу параметра $\mathbf{W}$. Цей результат є наслідком формули (2). Тому під час обговорення задач теорії оцінювання можна вважати, що функція $\gamma(\boldsymbol{\nu})$ у виразі (1) є сталою.

Якщо параметр W одновимірний, тобто його значення лежать в $\mathbb{R}^1$, то функцію втрат у задачах оцінювання часто вибирають такою, що має вигляд

$$L(\nu, d) = a|\nu - d|^b, \tag{3}$$

де $a > 0$, $b > 0$. Розглянемо більш детально такі функції втрат для значень $b = 1$ та $b = 2$.

Квадратична функція втрат

Найкраще вивченою функцією втрат у задачах оцінювання дійсного параметра W є квадратична функція. Вона задається рівністю

$$L(\nu, d) = (\nu - d)^2. \tag{4}$$

Функція втрат (4) зручна для математичних викладок. Крім того, наступні не зовсім строгі, але корисні міркування показують, чому така функція втрат часто використовується.

Припустимо, що втрати $L(\nu, d)$ залежать лише від різниці $\nu - d$. Нехай $L(\nu, d) = \Lambda(\nu - d)$, де Λ – невід'ємна двічі диференційовна функція така, що $\Lambda(0) = 0$. Якщо розкласти функцію Λ в ряд Тейлора з точністю до членів другого порядку, то отримаємо

$$L(\nu, d) = \Lambda(\nu - d) \approx a_0 + a_1(\nu - d) + a_2(\omega - d)^2. \tag{5}$$

Якщо статистик має достатню кількість інформації про значення W, щоб вибрати оцінку d, яка буде близькою до ν, то члени більш високого порядку в (5) відносно малі та ними можна знехтувати.

Той факт, що $\Lambda(0) = 0$, означає, що $a_0 = 0$, а з невід'ємності функції Λ випливає, що $a_1 = 0$ та $a_2 \geq 0$. Отже, за рахунок вибору системи одиниць співвідношення (5) зводиться до (4).

Теорема

Баєсівська оцінка d^* для квадратичної функції втрат (4) та довільного апріорного розподілу W дорівнює його математичному сподіванню:

$$d^* = EW,$$

а баєсівський ризик дорівнює дисперсії W:

$$R^* = DW.$$

Доведення

Якщо функція втрат задається рівністю (4), то баєсівське рішення d^* для довільного апріорного розподілу W є таким, що мінімізує значення ризику

$$R(d) = \mathrm{E}(W - d)^2 = \mathrm{E}W^2 - 2d\,\mathrm{E}W + d^2.$$

Цей квадратичний поліном буде мінімальним, коли $d = \mathsf{E}W$. За такого вибору d мінімальне значення ризику дорівнює

$$R(d^*) = \mathsf{E}(W - \mathsf{E}W)^2 = \mathsf{D}W.$$

Припустимо, що $\mathbf{X}$ – спостереження з (узагальненою) щільністю розподілу $f_\nu(\mathbf{x})$ за умови $W = \nu$. Нехай $u_{\mathbf{x}}(\nu)$ позначає апостеріорну щільність розподілу W за умови $\mathbf{X} = \mathbf{x}$. Тоді легко знайти баєсівську оцінку $\delta^*(\mathbf{x})$ та відповідний баєсівський ризик $R^* = R^*(u)$ для квадратичної функції втрат (4).

Теорема

Баєсівська оцінка $\delta^*(x)$ для квадратичної функції втрат (4) та довільного апостеріорного розподілу W має вигляд

$$\delta^*(x) = E(W|\mathbf{X} = \mathbf{x}) = \int_{\mathbb{R}} \nu u_{\mathbf{x}}(\nu) d\nu, \tag{6}$$

а баєсівський ризик дорівнює

$$R^* = E(D(W|\mathbf{X})).$$

Доведення

Зі сказаного в параграфі ?? та доведення попередньої теореми випливає, що для довільного спостереженого значення $\mathbf{X} = \mathbf{x}$ баєсівське рішення має вигляд (6), тобто дорівнює середньому значенню апостеріорного розподілу W. Далі, після спостереження значення $\mathbf{x}$ та вибору оцінки $\delta^*(\mathbf{x})$ ризик дорівнює умовній дисперсії $\mathrm{D}(W|\mathbf{X} = \mathbf{x})$ за умови апостеріорного розподілу W. Тому баєсівський ризик має вигляд $R(\delta^*) = \mathrm{E}(\mathrm{D}(W|\mathbf{X}))$.

Заувага

Для обчислення баєсівського ризику $\mathrm{E}(\mathrm{D}(W|\mathbf{X}))$ можна використовувати формулу повної дисперсії (??):

$$\mathrm{D}W = \mathrm{E}(\mathrm{D}\,(W|\mathbf{X})] + \mathrm{D}\,[\mathrm{E}(W|\mathbf{X})]. \tag{7}$$

Приклад

Як приклад, в якому кожне середнє існує та легко підраховується, розглянемо повторну вибірку $\mathbf{X} = (X_1, \ldots, X_n)$ із розподілу Пуассона з невідомим значенням параметра W. Припустимо, що апріорним розподілом W є гамма-розподіл з параметрами α та β. За теоремою 1.6 апостеріорний розподіл W при $X_i = x_i, i = 1, \ldots, n$ – це гамма-розподіл з параметрами $\alpha + \sum_{i=1}^{n} x_i$ та $\beta + n$.

Нехай функція втрат має вигляд (4) з $a = 1$. З виразу для середнього значення гамма-розподілу видно, що баєсівська оцінка δ^* визначається рівністю

$$\delta^*(X_1, \ldots, X_n) = \frac{\alpha + \sum_{i=1}^{n} X_i}{\beta + n}. \tag{8}$$

Для довільних значень $X_1, \ldots, X_n$ дисперсія апостеріорного розподілу дорівнює

$$\mathsf{D}(W|X_1, \ldots, X_n) = \frac{\alpha + \sum\limits_{i=1}^{n} X_i}{(\beta + n)^2}. \tag{9}$$

Оскільки $\mathsf{E}(X_i\,|W) = W$, $i = 1, \ldots, n$, то

$$\mathsf{E}(X_i) = \mathsf{E}[\mathsf{E}(X_i\,|W)] = \mathsf{E}(W) = \frac{\alpha}{\beta}. \tag{10}$$

Тому зі співвідношень (??) (9) і (10) виводимо наступну формулу для баєсівського ризику:

$$\rho^*(p) = \frac{\alpha}{\beta(\beta + n)}. \tag{11}$$

Припустимо тепер, що ціна одного спостереження у вибірці дорівнює c ($c > 0$) і що статистик може вибирати об'єм вибірки. З формули (11) видно, що для вибірки об'ємом n спостережень загальний ризик дорівнює

$$\frac{\alpha}{\beta(\beta + n)} + cn. \tag{12}$$

Цей загальний ризик буде мінімальним при

$$n = \left(\frac{\alpha}{c\beta}\right)^{1/2} - \beta. \tag{13}$$

Нехай функція втрат L замість співвідношення (4) задається:

$$L(\nu, d) = \frac{(\nu - d)^2}{\nu}. \tag{14}$$

Знову припустимо, що апріорний розподіл W є гамма-розподіл з параметрами α і β. Тоді для довільних значень спостережень $X_1, \dots, X_n$ таких, що $\alpha + \sum_{i=1}^{n} X_i > 1$, баєсівська оцінка $\delta^*(X_1, \dots, X_n)$ має вигляд

$$\delta^*(X_1, \dots, X_n) = \frac{\alpha + \sum_{i=1}^{n} X_i - 1}{\beta + n}. \tag{15}$$

Баєсівський ризик $\rho^*(\xi)$ дорівнює

$$\rho^*(\xi) = \frac{1}{(\beta + n)}. \tag{16}$$

Загальний ризик буде мінімальним при

$$n = \left(\frac{1}{c}\right)^{1/2} - \beta. \tag{17}$$

Втрати, пропорційні абсолютній величині похибки

Припустимо, що збиток пропорційний абсолютній величині похибки. Нехай для $\nu \in \mathbb{R}$ та $d \in \mathbb{R}$ збиток $L(\nu, d)$ має вигляд

$$L(\nu, d) = |\nu - d|. \tag{18}$$

Кажуть, що число m – це *медіана розподілу* W, якщо $\mathrm{P}(W \geq m) \geq \frac{1}{2}$ та $\mathrm{P}(W \leq m) \geq \frac{1}{2}$. Кожний розподіл має принаймні одну медіану, але не обов'язково єдину.

Наступна теорема показує, що медіана розподілу W буде баєсівською оцінкою параметра для функції втрат (18).

Теорема

Баєсівська оцінка $\delta^*(x)$ для функції втрат (4), пропорційних абсолютній величині похибки, та довільної функції щільності $u_{\mathbf{x}}(\nu)$ розподілу W за умови спостереження значення $\mathbf{x}$ в.в. $\mathbf{X}$ є медіаною апостеріорного розподілу W.

Доведення

Баєсівське рішення $d^* = \delta^*(\mathbf{x}) \in D$ для кожного спостереженого значення $\mathbf{X} = \mathbf{x}$ можна отримати, мінімізуючи вираз

$$G(d) = \mathsf{E}(L(W, d)|\mathbf{X} = \mathbf{x}) = \int_{\mathbb{R}} |\nu - d| u_{\mathbf{x}} \mathrm{d}\nu$$

$$= \int_{-\infty}^{d} (d - \nu) u_{\mathbf{x}} \mathrm{d}\nu + \int_{d}^{-\infty} (\nu - d) u_{\mathbf{x}} \mathrm{d}\nu.$$

Звідси

$$\frac{\mathrm{d}G(d)}{\mathrm{d}d} = \int_{-\infty}^{d} u_{\mathbf{x}} \mathrm{d}\nu - \int_{d}^{-\infty} u_{\mathbf{x}} \mathrm{d}\nu = 0,$$

а це означає, що d є розв'язком рівняння

$$\mathsf{P}(W \leq d|\mathbf{X} = \mathbf{x}) = \mathsf{P}(W \geq d|\mathbf{X} = \mathbf{x}),$$

т.б. баєсівське рішення – медіана апостеріорного р-лу W.

Втрати “все або нічого”

Припустимо, що функцію втрат визначено таким чином:

$$L(\nu, d) = \begin{cases} 0, & \text{коли } d = \nu; \\ 1, & \text{коли } d \neq \nu. \end{cases} \tag{19}$$

Таку функцію втрат називають втратами 0/1 чи “все або нічого”.

Нагадаємо, що *модою* абсолютно неперервного розподілу називають точку максимуму щільності розподілу ймовірностей.

Теорема

Баєсівська оцінка $\delta^*(x)$ для функції втрат "все або нічого" (20) та довільної функції щільності $u_{\mathbf{x}}(\nu)$ розподілу W за умови спостереження значення $\mathbf{x}$ в.в. $\mathbf{X}$ є модою апостеріорного розподілу W.

Доведення

Оскільки у цьому випадку диференціювання виразу $G(d) = \mathrm{E}(L(W, d)|\mathbf{X} = \mathbf{x})$ не можливе, застосуємо перехід до границі. Нехай для $\varepsilon > 0$

$$\tilde{L}(\nu, d) = \begin{cases} 0, & \text{коли } \nu - \varepsilon < d < \nu + \varepsilon; \\ 1, & \text{інакше.} \end{cases} \tag{20}$$

Тоді $\tilde{L}(\nu, d)$ прямує до функції втрат (20), коли $\varepsilon \to 0$. Для $\tilde{L}$ очікувані втрати відносно апостеріорного розподілу $u_{\mathbf{x}}(\nu)$ дорівнюють

$$\tilde{G}(d) = \mathsf{E}(\tilde{L}(W, d)|\mathbf{X} = \mathbf{x}) = 1 - \int_{d-\varepsilon}^{d+\varepsilon} u_{\mathbf{x}}(\nu) d\nu = 1 - 2\varepsilon u_{\mathbf{x}}(d)$$

при достатньо малому ε. Як бачимо, мінімум $\tilde{G}(d)$ досягаємо, поклавши d рівним моді апостеріорного розподілу параметра W.

Оцінювання векторного параметра

Розглянемо задачу оцінювання вектора

$$\mathbf{W} = (W_1, \ldots, W_k), \quad k \geq 2.$$

Стандартною функцією втрат для такої задачі є квадратична функція втрат L, яка визначена для всіх $\mathbf{w} \in \mathbb{R}^k$ та $\mathbf{d} \in \mathbb{R}^k$ за формулою

$$L(\mathbf{w}, \mathbf{d}) = (\mathbf{w} - \mathbf{d})' A (\mathbf{w} - \mathbf{d}). \tag{21}$$

Тут A – симетрична невід'ємно визначена $k \times k$ матриця. Якщо A – додатно визначена, то кожному $\neq 0$ вектору похибок $\mathbf{w} - \mathbf{d}$ відповідає додатний збиток. Якщо матриця A не є додатно визначеною, то існують ненульові вектори похибок з нульовими втратами – наприклад, коли статистик цікавиться оцінкою лише кількох компонент $\mathbf{W}$, а значення інших для нього несуттєві.

Припустимо, що для $\mathbf{W}$ існує вектор середніх $\mathsf{E}(\mathbf{W}) = \mu$ та коваріаційна матриця $\operatorname{cov}(\mathbf{W}) = \Sigma$. Баєсівська оцінка при заданому розподілі $\mathbf{W}$ – це точка $\mathbf{d} \in \mathbb{R}^k$, яка мінімізує математичне сподівання

$$\mathsf{E}[(\mathbf{W} - \mathbf{d})' A\,(\mathbf{W} - \mathbf{d})]$$

$$= \mathsf{E}\{[(\mathbf{W} - \mu)' + (\mu - \mathbf{d})']\,A\,[(\mathbf{W} - \mu) + (\mu - \mathbf{d})]\}$$
$$= \mathsf{E}[(\mathbf{W} - \mu)'\,A\,(\mathbf{W} - \mu)] + (\mu - \mathbf{d})' A (\mu - \mathbf{d}).$$

У правій частині рішення $\mathbf{d}$ не входить під знак мат. сподівання. Оскільки A – невід'ємно визначена, то другий доданок правої частини невід'ємний для всіх значень $\mathbf{d}$. Отже, $\mathbf{d} \in \mathbb{R}^k$ – баєсівська оцінка тоді і лише тоді, коли

$$(\mu - \mathbf{d})' A (\mu - \mathbf{d}) = 0. \tag{22}$$

Звідси випливає, що значення $\mathbf{d} = \mu$ є баєсівською оцінкою для $\mathbf{W}$. Далі, якщо A – додатно визначена матриця, то це значення $\mathbf{d}$ буде єдиною баєсівською оцінкою. Якщо матриця A не є додатно визначеною, то знайдуться інші значення $\mathbf{d}$, які задовольняють співвідношення (22). Підкреслимо, що вектор середніх $\mu = \mathrm{E}(\mathbf{W})$ завжди є баєсівською оцінкою для довільної симетричної невід'ємно визначеної матриці A.

Можна показати, що

$$\mathrm{E}[(\mathbf{W} - \mu)' A (\mathbf{W} - \mu)] = \mathrm{tr}(A\Sigma), \tag{23}$$

де $\mathrm{tr}(Z)$ – слід матриці Z, тобто сума її діагональних елементів. Отже, математичне сподівання збитку для довільної баєсівської оцінки $\mathbf{d}$ дорівнює $\mathrm{tr}(A\Sigma)$.

Нехай $\mathbf{X}$ – спостереження з умовною щільністю розподілу $f_{\boldsymbol{\nu}}(\mathbf{x})$ за умови $\mathbf{W} = \boldsymbol{\nu}$. З викладеного вище зрозуміло, що баєсівську вирішувальну функцію δ^* визначають з рівності $\delta^*(\mathbf{X}) = \mathsf{E}(\mathbf{W}|\mathbf{X})$, а баєсівський ризик R^* при заданій апріорній щільності розподілу p дорівнює

$$R^* = \mathrm{tr}\{A\,\mathsf{E}[\mathrm{cov}(\mathbf{W}|\mathbf{X})]\}. \tag{24}$$

Тут $\mathrm{cov}(\mathbf{W}|\mathbf{X})$ – коваріаційна матриця апостеріорного розподілу $\mathbf{W}$ при заданому значенні $\mathbf{X} = \mathbf{x}$.

Спряжені сімейства розподілів

Припустимо, що $\mathbf{X} = (X_1, \ldots, X_n)$ – повторна випадкова вибірка з генеральної сукупності із заданою функцією розподілу $F_\nu(\mathbf{x})$ чи щільністю $f_\nu(\mathbf{x})$, $\nu \in \Theta \subseteq \mathbb{R}^k$, $k \geq 1$, де значення ν параметра W частково чи повністю невідоме.

Нас цікавить значення певного функціоналу $g(\nu)$, що залежить від параметра ν. Для цього будемо шукати таку функцію $T(\mathbf{X})$, що "якнайкраще" оцінює $g(\nu)$ і залежить лише від вектора спостережень $\mathbf{X}$. Функцію $T(\mathbf{X})$ називають оцінкою $g(\nu)$.

З математичної точки зору вираз "якнайкраще" представляє функція втрат $L(\nu, T(\mathbf{x}))$, яка дорівнює збиткам від вибору оцінки $T(\mathbf{x})$ за спостережним значенням $\mathbf{x}$ для справжнього значення ν. Тоді функція ризику оцінки T має вигляд

$$R_\nu(T) = \mathsf{E}(L(\nu, T(\mathbf{x}))|W = \nu)$$

$$= \int_{\mathbb{R}^n} L(\nu, T(\mathbf{x}))\mathrm{d}F_\nu(\mathbf{x}) = \int_{\mathbb{R}^n} L(\nu, T(\mathbf{x}))f_\nu(\mathbf{x})\mathrm{d}\mathbf{x}. \qquad (25)$$

Оскільки $\mathbf{X} = (X_1, \dots, X_n)$ – повторна вибірка, тобто спостереження X_i, $i = 1, \dots, n$, – н.о.р. в.в. із функцією розподілу $F_\nu(x)$ чи щільністю $f_\nu(x)$, то сумісну функцію розподілу вибірки $\mathbf{X}$ можна записати, як

$$F_\nu(\mathbf{x}) = \prod_{i=1}^{n} F_\nu(x_i), \quad \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n,$$

Сумісна щільність

$$f_\nu(\mathbf{x}) = \prod_{i=1}^{n} f_\nu(x_i) \tag{26}$$

є нічим іншим, як функцією вірогідності.

Нехай параметр W має апріорну щільність $u(\nu)$, $\nu \in \Theta$. Його апостеріорну щільність $u_{\mathbf{x}}(\nu)$, тобто щільність W за умови, що вибіркові значення **x** відомі, визначають за допомогою базової формули (**??**):

$$u_{\mathbf{x}}(\nu) = \frac{f_{\nu}(\mathbf{x})u(\nu)}{f_{\mathbf{X}}(\mathbf{x})}, \tag{27}$$

де

$$f_{\mathbf{X}}(\mathbf{x}) = \int\limits_{\Theta} f_{\mathbf{X},W}(\mathbf{x},\nu)\mathrm{d}\nu = \int\limits_{\Theta} f_{\mathbf{X},W=\nu}(\mathbf{x}|\nu)u(\nu)\mathrm{d}\nu = \int\limits_{\Theta} f_{\nu}(\mathbf{x})u(\nu)\mathrm{d}\nu$$

– безумовна щільність випадкового вектора **X**.

Часто (27) зручніше виражати не в термінах вибіркових значень, а за допомогою значення деякої (достатньої) вибіркової статистики $T(\mathbf{X})$, зокрема вибіркового середнього $\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$. Наприклад, часто має сенс розглядати

$$u_{\bar{x}}(\nu) = \frac{f_{\nu}(\bar{x})u(\nu)}{f_{\mathbf{X}}(\bar{x})}, \tag{28}$$

де $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$. На практиці ці два підходи еквівалентні.

Зручний спосіб знаходження апостеріорної щільності полягає у використання пропорційності. Оскільки функція $f_{\mathbf{X}}(\mathbf{x})$ не залежить від ν, є обмеженою та інтегровною, то з (??) випливає, що

$$u_{\mathbf{x}}(\nu) \propto f_{\nu}(\mathbf{x})u(\nu), \quad \nu \in \Theta, \tag{29}$$

тобто апостеріорна щільність параметра W пропорційна добутку апріорної щільності та функції вірогідності.

Відповідно до вищесказаного, планування та аналізування експерименту виявляються досить нескладними, якщо існує стандартне сімейство розподілів параметра W, що має таку властивість:

- якщо апріорний розподіл W належить цьому сімейству, то для будь-якого розміру вибірки n і будь-яких значень спостережень у вибірці апостеріорний розподіл W також належить цьому сімейству.

Сімейство розподілів, які мають таку властивість, називають *замкнутим відносно процесу вибору*. Його називають також *спряженим сімейством розподілів* через особливий зв'язок, який існує між розподілами параметра і розподілами спостережень.

Вибірка з розподілу Бернуллі

Припустимо, наприклад, що $\mathbf{X} = (X_1, \ldots, X_n)$ – повторна вибірка з розподілу Бернуллі з невідомим значенням параметра W. Для заданого значення $W = \nu$ функція вірогідності $f_\nu(\mathbf{x})$ дорівнює

$$f_\nu(\mathbf{x}) = \prod_{i=1}^{n} f_\nu(x_i) = \nu^y (1-\nu)^{n-y}, \quad \text{де} \quad y = \sum_{i=1}^{n} x_i. \tag{30}$$

Припустимо також, що апріорний розподіл W є бета-розподілом із заданими значеннями параметрів $\alpha > 0$ та $\beta > 0$. Тоді апріорна щільність розподілу ймовірностей параметру W має вигляд

$$u(\nu) \propto \nu^{\alpha-1} (1-\nu)^{\beta-1}, \quad 0 < \nu < 1. \tag{31}$$

Тому зі співвідношень (29), (31) апостеріорна щільність розподілу ймовірностей $u_{\mathbf{x}}(\nu)$ параметра W після спостереження значень $X_i = x_i$, $i = 1, 2, \ldots, n$, є

$$u_{\mathbf{x}}(\nu) \propto \nu^{\alpha+y-1}(1-\nu)^{n-y+\beta-1}. \tag{32}$$

Зі співвідношення (32) видно, що апостеріорним розподілом W є бета-розподіл з параметрами $\alpha + y$ і $\beta + n - y$. Ми довели наступну теорему.

Теорема

Нехай $X_1, \ldots, X_n$ *– повторна вибірка з розподілу Бернуллі з невідомим значенням параметра* W. *Припустимо, що апріорний розподіл* W *є бета-розподіл з параметрами* α *і* β, $\alpha > 0, \beta > 0$. *Тоді апостеріорний розподіл* W *за умови, що випадкові величини* X_i, $i = 1, 2, \ldots, n$, *набувають значення* $x_i, i = 1, 2, \ldots, n$, *є бета-розподіл з параметрами* $\alpha + y$ *і* $\beta + n - y$, *де* $y = \sum_{i=1}^{n} x_i$. *Іншими словами, сімейство бета-розподілів спряжене з сімейством розподілів Бернуллі.*

Вибірки з пуассонівського, від'ємного біноміального та експоненційного розподілів

Теорема

Нехай $X_1, \ldots, X_n$ *– повторна вибірка з розподілу Пуассона з невідомим значенням середнього значення параметра* W*. Припустимо, що апріорним розподілом* W *є гамма-розподіл з параметрами* $\alpha > 0$ *та* $\beta > 0$*. Тоді апостеріорним розподілом* W *за умови, що в.в.* X_i *набувають значення* x_i*,* $i = 1, 2, \ldots, n$*, є гамма-розподіл з параметрами* $\alpha + \sum_{i=1}^{n} x_i$ *та* $\beta + n$*.*

Доведення

Позначимо $y = \sum_{i=1}^{n} x_i$, тоді з умов теореми випливає, що для $\nu > 0$

$$f_\nu(\mathbf{x}) \propto \nu^y e^{-n\nu}, \quad \mathbf{x} = (x_1, \ldots, x_n), \quad \text{та} \quad u(\nu) \propto \nu^{\alpha-1} e^{-\beta\nu}. \tag{33}$$

Отже, зі співвідношень (29) та (33) випливає, що

$$u_{\mathbf{x}}(\nu) \propto \nu^{\alpha+y-1} e^{-(\beta+n)\nu}.$$

Таким чином, апостеріорна щільність розподілу W є щільністю гамма-розподілу з параметрами $\alpha + y$ та $\beta + n$.

В якості міри розсіювання розподілу додатної в.в. X часто вживають *коефіцієнт варіації*, який визначається співвідношенням

$$\mathrm{CV}(X) = \frac{\sqrt{\mathrm{D}X}}{\mathrm{E}X} = \frac{\sigma_X}{m_1}. \tag{34}$$

Нагадаємо, що коли в.в. X має гамма-розподіл з параметрами α та β, то $\mathrm{E}X = \frac{\alpha}{\beta}$, $\mathrm{D}X = \frac{\alpha}{\beta^2}$. З цих співвідношень та теореми 1.6 випливає, що коефіцієнт варіації апостеріорного розподілу W дорівнює

$$\mathrm{CV}_{\mathbf{x}}(W) = \mathrm{CV}(W|\mathbf{X} = \mathbf{x}) = \left(\alpha + \sum_{i=1}^{n} x_i\right)^{-\frac{1}{2}}.$$

Нехай $\varepsilon > 0$ – фіксоване число. Припустимо, що спостереження розподілу Пуассона проводять доти, поки коефіцієнт варіації апостеріорного розподілу W не стане меншим за ε. Тоді спостереження потрібно продовжувати до моменту першого виконання нерівності

$$\alpha + \sum_{i=1}^{n} x_i \geq \frac{1}{\varepsilon^2}.$$

Наступні теореми описують спряжені сімейства розподілів для вибірок з від'ємного біноміального та експоненційного (показникового) розподілів. Доведення теорем аналогічні доведенню попередньої теореми.

Теорема

Нехай $X_1, \ldots, X_n$ *– повторна вибірка з від'ємного біноміального розподілу з параметрами* r *і* W, *де* $r > 0$ *– фіксоване, а значення* W *невідоме. Нехай апріорний розподіл* W *є бета-розподілом із параметрами* $\alpha > 0$, $\beta > 0$. *Тоді апостеріорний розподіл* W *за умови, що* $X_i = x_i$, $i = 1, 2, \ldots, n$, *є бета-розподілом із параметрами* $\alpha + rn$ *та* $\beta + \sum\limits_{i=1}^{n} x_i$.

Теорема

Нехай $X_1, X_2, \ldots, X_n$ *– вибірка з експоненційного розподілу з невідомим значенням параметра* W*. Припустимо, що апріорним розподілом* W *є гамма-розподіл з параметрами* $\alpha > 0$ *та* $\beta > 0$*. Тоді апостеріорний розподіл* W *за умови, що* $X_i = x_i$*,* $i = 1, 2, \ldots, n$*, є гамма-розподілом із параметрами* $\alpha + n$ *і* $\beta + \sum\limits_{i=1}^{n} x_i$.

Спряжені сімейства для вибірок з нормального розподілу

Ми почнемо з розгляду нормальних розподілів із відомою мірою точності, або, що те саме, з відомою дисперсією.

Теорема

Нехай $X_1, X_2, \ldots, X_n$ *– повторна вибірка з нормального розподілу з невідомим значенням середнього* W *та заданою мірою точності* $r > 0$. *Припустимо, що апріорний розподіл* W *– це нормальний розподіл із середнім* μ *й мірою точності* $\tau > 0$. *Тоді апостеріорним розподілом* W *за умови спостереження значень* $X_i = x_i$, $i = 1, 2, \ldots, n$, *є нормальний із середнім* μ' *та мірою точності* $\tau' = \tau + nr$, *де*

$$\mu' = \frac{\tau\mu + nr\bar{x}}{\tau + nr}. \tag{35}$$

Доведення

Для довільного значення $\nu \in \mathbb{R}$ умовна щільність розподілу в.в. $X_1, X_2, \ldots, X_n$ задовольняє умові

$$f_\nu(\mathbf{x}) \propto \exp\left\{-\frac{r}{2}\sum_{i=1}^{n}(x_i-\nu)^2\right\}, \quad \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n. \quad (36)$$

Але

$$\sum_{i=1}^{n}(x_i-\nu)^2 = n(\nu-\bar{x})^2 + \sum_{i=1}^{n}(x_i-\bar{x})^2. \quad (37)$$

Оскільки останній доданок у рівності (37) не містить ν, то ми можемо переписати співвідношення (36) так:

$$f_\nu(\mathbf{x}) \propto \exp\left\{-\frac{nr}{2}(\nu-\bar{x})^2\right\}. \quad (38)$$

Апріорна щільність розподілу $u(\nu)$ параметра W задовольняє співвідношення

$$u(\nu) \propto \exp\left\{-\frac{\tau}{2}(\nu-\mu)^2\right\}. \tag{39}$$

Апостеріорна щільність розподілу $u_{\mathbf{x}}(\nu)$ параметра W пропорційна добутку функцій, що фігурують у (38) й (39). Можна показати, що

$$\tau(\nu-\mu)^2 + nr(\nu-\bar{x})^2 = \tau + nr\left(\nu-\mu'\right)^2 + \frac{\tau nr(\bar{x}-\mu)^2}{\tau+nr}. \tag{40}$$

Оскільки останній доданок у (40) не містить ν, то він може бути включений у нормуючий множник, і ми одержуємо таке співвідношення:

$$u_{\mathbf{x}} \propto \exp\left\{-\frac{\tau + nr}{2}\left(\nu - \mu'\right)^2\right\}, \qquad (41)$$

де μ' визначено у (35). Зі співвідношення (41) випливає, що апостеріорним розподілом W є нормальний із середнім μ' та мірою точності $\tau + nr$.

Ця теорема показує, чому краще виражати результати спостережень через міру точності, а не через дисперсію. Середнє μ' апостеріорного розподілу W можна записати у вигляді

$$\mu' = \frac{nr}{\tau + nr}\bar{x} + \frac{\tau}{\tau + nr}\mu. \tag{42}$$

Ми бачимо, що μ' – це зважене середнє $\bar{x}$ та μ, де $\bar{x}$ – значення вибіркового середнього, а μ – середнє апріорного розподілу W. Тому можна розглядати середнє апостеріорного розподілу μ' як зважене середнє оцінки $\bar{x}$ параметра W, побудованої за вибіркою, та оцінки параметра W, одержаної, виходячи з апріорного розподілу.

Ваги оцінок, $\bar{x}$ та μ, у цьому усередненні пропорційні nr та τ, де nr – міра точності умовного розподілу вибіркового середнього при будь-якому фіксованому значенні W, а τ – міра точності апріорного розподілу W. Чим більшим є розмір вибірки n та чим вища точність r кожного спостереження, тим більше вага, що дається $\bar{x}$.

Вигляд міри точності $\tau' = \tau + nr$ апостеріорного розподілу W досить простий. Точність зростає на r одиниць при кожному спостереженні незалежно від одержаних значень спостережень. Тому під час збільшення кількості спостережень розподіл W все більше концентруватиметься навколо свого середнього, значення якого залежать від спостережень.

У наступній теоремі розглянемо нормальний розподіл, для якого значення середнього задане, а значення міри точності невідоме.

Теорема

Нехай $X_1, X_2, \ldots, X_n$ – повторна вибірка з нормального розподілу із заданим значенням середнього μ та невідомим значенням міри точності W, і нехай апріорний розподіл W є гамма-розподілом із параметрами $\alpha > 0$ та $\beta > 0$. Тоді апостеріорним розподілом W при $X_i = x_i$, $i = 1, 2, \ldots, n$, є гамма-розподіл з параметрами $\alpha + \frac{n}{2}$ та β',

$$\beta' = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2. \tag{43}$$

Для гамма-розподілу з параметрами α і β коефіцієнт варіації дорівнює $\alpha^{-\frac{1}{2}}$. З теореми випливає, що коефіцієнт варіації апостеріорного розподілу W спадає, коли розмір вибірки зростає.

Розглянемо тепер вибірку з нормального розподілу, у якого і середнє, і міра точності невідомі. Спряженим сімейством у цій задачі має бути деяке сімейство двовимірних розподілів.

Теорема

Нехай $X_1, X_2, \ldots, X_n$ – повторна вибірка з нормального розподілу із невідомими значеннями середнього M та міри точності R. Припустимо, що апріорний сумісний розподіл M та R такий: умовним розподілом M за даного $R = r > 0$ є нормальний із середнім $\mu \in \mathbb{R}$ та мірою точності τr, $\tau > 0$, а розподілом R є гамма-розподіл із параметрами $\alpha > 0$ та $\beta > 0$. Тоді апостеріорний сумісний розподіл M та R за умови спостереження $X_i = x_i$, $i = 1, 2, \ldots, n$, має наступний вигляд: умовний розподіл M за даного $R = r$ є нормальним із середнім μ' та мірою точності $(\tau + n)\, r$, де

$$\mu' = \frac{\tau\mu + n\bar{x}}{\tau + n}, \tag{44}$$

Доведення

Позначимо для даних значень $M = m \in \mathbb{R}$ та $R = r > 0$ умовну сумісну щільність розподілу випадкової вибірки $\mathbf{X} = (X_1, \ldots, X_n)$ через $f_{m,r}(\mathbf{x})$, $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, а сумісну щільність розподілу ймовірностей M та R через $u(m, r)$. Тоді

$$f_{m,r}(\mathbf{x}) \propto r^{\frac{n}{2}} \exp\left\{-\frac{r}{2}\sum_{i=1}^{n}(x_i - m)^2\right\} \tag{46}$$

та

$$u(m, r) \propto r^{\frac{1}{2}} \exp\left\{-\frac{\tau r}{2}(m-\mu)^2\right\} r^{\alpha-1} e^{-r\beta}. \tag{47}$$

Апостеріорна сумісна щільність розподілу $u_{\mathbf{x}}(m, r)$ параметрів M та R пропорційна добутку правих частин співвідношень (46) та (47). З рівностей (37) та (40) випливає, що ця щільність розподілу може бути задана співвідношенням

$$u_{\mathbf{x}}(m, r) \propto \left(r^{\frac{1}{2}} \exp\left\{ -\frac{(\tau + n)\, r}{2} \left(m - \mu' \right)^2 \right\} \right) \left(r^{\alpha + \frac{n}{2} - 1} e^{-r\beta'} \right). \tag{48}$$

Тут μ' визначається рівністю (44), β' – рівністю (45).

Функція у першій парі круглих дужок у співвідношенні (48), як функція від m, повинна бути пропорційною умовній щільність розподілу M за відомого значення R, тому що змінна m не входить у вираз у другій парі дужок. Але для кожного фіксованого значення r функція у лівій парі дужок пропорційна щільності нормального розподілу, середнє значення й міра точності якого задані в умовах теореми. Звідси випливає, що функція у правій парі дужок повинна бути пропорційною щільності розподілу R. Отже, розподілом R є гамма-розподіл, параметри якого зазначені у теоремі.

Якщо сумісна щільність розподілу ξ параметрів M та R – це щільність нормального–гамма розподілу, яка визначена у співвідношенні (47), то умовний розподіл M для будь-якого заданого значення $R = r$ буде нормальним, але розподіл M таким не буде. Маргінальна щільність $u_M(m)$ розподілу параметра M дорівнює

$$u_M\left(m\right) = \int\limits_0^{\infty} u\left(m, r\right) \mathrm{d}r. \tag{49}$$

Якщо скористатись символом пропорційності та пропустити всі множники, що не містять m, то з (47) одержимо, що $u_M(m)$ має наступний вигляд:

$$u_M(m) \propto \left(\beta + \frac{\tau}{2}(m-\mu)^2\right)^{-\frac{2\alpha+1}{2}} \propto \left(1 + \frac{1}{2\alpha}\frac{\alpha\tau(m-\mu)^2}{\beta}\right)^{-\frac{2\alpha+1}{2}}. \tag{50}$$

Порівнюючи функцію, визначену у співвідношенні (50), зі щільністю t-розподілу, бачимо, що розподіл M є t-розподілом з 2α степенями вільності, параметром зміщення μ та мірою точності $\frac{\alpha\tau}{\beta}$. Апостеріорний розподіл M знаходимо заміною μ, τ, α, β, їхніми апостеріорними значеннями, які вказані в теоремі. Отже, кількість степеней вільності $2\alpha + n$ апостеріорного розподілу не залежить від спостережених значень $x_1, \ldots, x_n$, проте параметр зміщення та точність апостеріорного розподілу від них залежать.

Вибірка з рівномірного розподілу

Теорема

Нехай $X_1, \ldots, X_n$ *– повторна вибірка з рівномірного розподілу на інтервалі* $(0, W)$*, де значення параметра* W *невідоме. Припустимо, що апріорний розподіл* W *– це розподіл Парето з параметрами* $\nu_0 > 0$ *та* $\alpha > 0$*, тобто*

$$u(\nu) = \frac{\alpha \nu_0^{\alpha}}{\nu^{\alpha+1}}, \quad \nu > \nu_0. \tag{51}$$

Тоді апостеріорний розподіл W *за умови* $X_i = x_i$*,* $i = 1, 2, \ldots, n$ *є розподілом Парето з параметрами* ν_0' *та* $\alpha + n$*, де*

$$\nu_0' = \max\{\nu_0, x_1, \ldots, x_n\}. \tag{52}$$

Доведення

З (51) випливає, що для $\nu > \nu_0$ апріорна щільність розподілу $u(\nu)$ параметра W має вигляд

$$u(\nu) \propto \frac{1}{\nu^{\alpha+1}}. \tag{53}$$

Щільність розподілу $f_\nu(\mathbf{x})$ задається рівностями (??) та (26) та дорівнює

$$f_\nu(\mathbf{x}) = \frac{1}{\nu^n}, \quad 0 < \max\{x_1, x_2, \ldots, x_n\} < \nu. \tag{54}$$

Звідси випливає, що апостеріорна щільність розподілу $u_{\mathbf{x}}(\nu)$ параметра W буде додатною лише для таких значень ν, для яких $\nu > \nu_0$ та $\nu > \max\{x_1, \ldots, x_n\}$. Тому $u_{\mathbf{x}}(\nu) > 0$ лише тоді, коли $\nu > \nu_0'$, де ν_0' визначено у (52). Отже, для $\nu > \nu_0'$

$$u_{\mathbf{x}}(\nu) \propto \frac{1}{\nu^{\alpha+n+1}}.$$

А значить, апостеріорний розподіл W має бути розподілом Парето зі значеннями параметрів, визначеними у теоремі.

Теорема

Нехай $X_1, \ldots, X_n$ *– повторна вибірка з рівномірного розподілу на інтервалі* (W_1, W_2)*, де значення* W_1 *та* W_2 *невідомі. Припустимо, що апріорним сумісним розподілом* W_1 *та* W_2 *є двосторонній двовимірний розподіл Парето з параметрами* r_1*,* r_2 *та* α*, де* $r_1 < r_2$*,* $\alpha > 0$*, тобто*

$$u(\nu_1, \nu_2) = \begin{cases} \frac{\alpha(\alpha+1)(r_2-r_1)^\alpha}{(\nu_2-\nu_1)^{\alpha+2}} & \text{при } \nu_1 < r_1,\ \nu_2 > r_2, \\ 0 & \text{в інших випадках,} \end{cases} \tag{55}$$

зокрема

$$EW_1 = \frac{\alpha r_1 - r_2}{\alpha - 1}, \quad E(W_2) = \frac{\alpha r_2 - r_1}{\alpha - 1}, \tag{56}$$

$$DW_1 = DW_2 = \frac{\alpha(r_2 - r_1)^2}{(\alpha - 1)^2(\alpha - 2)}. \tag{57}$$

Тоді апостеріорний сумісний розподіл W_1 та W_2 за умови $X_i = x_i$, $i = 1, 2, \ldots, n$, є двостороннім двовимірним розподілом Парето з параметрами r_1', r_2' та $\alpha + n$, де

$$r_1' = \min\{r_1, x_1, \ldots, x_n\}, \quad r_2' = \max\{r_2, x_1, \ldots, x_n\}. \tag{58}$$

Приклад

Розглянемо повторну вибірку $X_1, \ldots, X_n$ з нормального розподілу з невідомим значенням середнього W і заданою мірою точності r. Припустимо, що потрібно оцінити значення W при функції втрат, пропорційної абс. величині похибки (18). З теореми 1.9 випливає, що якщо апріорний розподіл W нормальний з середнім μ та мірою точності τ, то для довільних значень $X_1, \ldots, X_n$ апостеріорний розподіл W нормальний з середнім $\left(\tau\mu + nr\bar{X}\right)/(\tau + nr)$ та мірою точності $\tau + nr$. Оскільки єдина медіана нормального розподілу співпадає з його середнім, то баєсівська оцінка δ^*:

$$\delta^*(X_1, \ldots, X_n) = \frac{\tau\mu + nr\bar{X}}{\tau + nr}. \tag{59}$$

Баєсівський ризик $\rho^*(p)$ дорівнює

$$\rho^*(p) = \mathsf{E}|W - \delta^*(X_1, \ldots, X_n)|. \tag{60}$$

Апостеріорний розподіл випадкової величини $[W - \delta^*(X_1, \ldots, X_n)]$ при $X_i = x_i, i = 1, \ldots, n$, буде нормальним з середнім 0 та мірою точності $\tau + nr$, незалежно від значень $x_1, \ldots, x_n$. Далі, якщо Y – нормальна випадкова величина з середнім 0 та мірою точності τ, то

$$\mathsf{E}|Y| = \left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}. \tag{61}$$

Отже, для довільних значень $X_1, \ldots, X_n$ середнє значення у правій частині (60) дорівнює $\left(\frac{2}{\pi(\tau+nr)}\right)^{\frac{1}{2}}$.

Із співвідношення (60) випливає, що

$$\rho^*(p) = \left[\frac{2}{\pi(\tau + nr)}\right]^{\frac{1}{2}}. \tag{62}$$

Припустимо тепер, що ціна одного спостереження дорівнює c ($c > 0$) і статистик може вибирати об'єм вибірки. З рівності (62) видно, що для вибірки об'ємом n загальний ризик дорівнює

$$\left[\frac{2}{\pi(\tau + nr)}\right]^{\frac{1}{2}} + cn \tag{63}$$

і він мінімізується при

$$n = \frac{1}{(2\pi rc^2)^{\frac{1}{3}}} - \frac{\tau}{r}. \tag{64}$$

Вправа

Розподіл величини позову є нормальним із невідомим середнім μ та відомим стандартним відхилом $\$50$. На основі попереднього досвіду відповідним апріорним розподілом параметра μ є нормальний з середнім $\$300$ та стандартним відхилом $\$20$.

i) Обчисліть апріорну ймовірність того, що середнє значення μ розподілу величини позову менше $\$270$.

ii) Випадкова вибірка з 10 поточних спостережень має середнє значення $\$270$.

   a) Знайдіть апостеріорний розподіл μ.

   b) Обчисліть апостеріорну ймовірність того, що μ менше $\$270$ та прокоментуйте отриману відповідь.

Вправа

Ризик складається з п'яти полісів. За кожним полісом щомісяця може виникнути один позов з імовірністю θ, ймовірність того, що протягом місяця позовів буде два або більше є незначною. Апріорний розподіл θ є рівномірним на $(0, 1)$. Всього за цим ризиком протягом періоду 12 місяців було 10 позовів.

i) Знайдіть апостеріорний розподіл θ.
ii) Визначте баєсівську оцінку θ за:
    (a) квадратичної функції втрат;
    (b) втрат "все або нічого".

Вправа

Страхова компанія забезпечує гарантію на певний електричний ґаджет. На початку 2006 року під гарантією перебувало $4\,500$ пристроїв, кожен з яких має ймовірність q повністю вийти з ладу протягом року (незалежно один від одного). Апріорний розподіл q – це бета-розподіл із середнім $0,015$ та стандартним відхилом $0,005$. Знаючи, що у 2006 році 58 ґаджетів повністю вийшли з ладу, визначте апостеріорний розподіл q.

Вправа

У несиметричної монети ймовірність випадіння герба є невідомою сталою p. Відомо, що p має бути або $0,4$, або $0,75$. Апріорні уявлення щодо p можна описати за допомогою розподілу: $\mathrm{P}(p = 0,4) = 0,6$, $\mathrm{P}(p = 0,75) = 0,4$. Монету підкинули 6 разів, з яких чотири вона випала гербом. Знайти апостеріорний розподіл p.

Офісний працівник щодня отримує випадкову кількість листів електронною поштою. Кількість листів за день має пуассонівський розподіл із невідомим параметром μ. Апріорні уявлення щодо μ відповідають гамма-розподілу з середнім 50 та стандартним відхилом 15. За 10-денний період працівник загалом отримав 630 листів.
Обчисліть баєсівську оцінку μ за умови втрат “все або нічого”.

Вправа

Нехай $y_1, \ldots, y_n$ – випадкова вибірка з інтервалу $[0, \theta]$, де $\theta > 0$ – невідома стала. Апріорні уявлення щодо θ задані розподілом зі щільністю

$$u(\theta) = \begin{cases} \alpha\beta^{\alpha}\theta^{-(1+\alpha)}, & \text{коли } \theta > \beta; \\ 0, & \text{інакше}, \end{cases}$$

де α та β – додатні сталі.

i) Покажіть, що апостеріорний розподіл θ для заданого y_1 має таку саму форму, як і апріорний, та визначте відповідні параметри.

ii) Запишіть апостеріорний розподіл θ для заданих $y_1, \ldots, y_n$.

Вправа

Ризик складається з п'яти полісів. За кожним полісом щомісяця може виникнути один позов з імовірністю θ, ймовірність того, що протягом місяця позовів буде два або більше є незначною. Апріорний розподіл θ є рівномірним на $(0, 1)$. Всього за цим ризиком протягом періоду 12 місяців було 10 позовів.

i) Знайдіть апостеріорний розподіл θ.
ii) Визначте баєсівську оцінку θ за:
   (a) квадратичної функції втрат;
   (b) втрат "все або нічого".