Список використаних джерел

Список основної літератури

1. Боровков А. А. Математическая статистика / А. А. Боровков. – Москва: Наука, 1984. – 472 с.
2. Ивченко Г. И. Математическая статистика / Г. И. Ивченко, Ю. И. Медведев. – Москва: Высшая школа, 1984. – 248 с.
3. Карташов М. В. Теорія ймовірностей та математична статистика / М. В. Карташов. – К.: ТВіМС, 2004. – 307 с.
4. Слюсарчук П. В. Теорія ймовірностей і математична статистика / П. В. Слюсарчук. – Ужгород: Вид-во «Карпати» , 2005. – 184 с.

Список додаткової літератури

5. Гихман И. И. Теория вероятностей и математическая статистика / И. И. Гихман, А. В. Скороход, М. Й. Ядренко. – Киев: Вища школа, 1988. – 320 с.
6. Севастьянов Б.А. Курс теории вероятностей и математической статистики / Б.А. Севастьянов. – Москва: Наука,1982. – 256 с.

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД

«УЖГОРОДСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ»

КАФЕДРА ТЕОРІЇ ЙМОВІРНОСТЕЙ І МАТЕМАТИЧНОГО АНАЛІЗУ

О. О. СИНЯВСЬКА

ДОДАТКОВІ РОЗДІЛИ СТАТИСТИКИ

Ужгород – 2017

УДК 519.22 (075.8)
ББК 22.172
С-38

О. О. Синявська. **Додаткові розділи статистики.** Конспект лекцій. – Ужгород, 2017. – 124 с.

Навчальний посібник містить повний курс лекцій з дисципліни «Додаткові розділи статистики» та рекомендований для студентів спеціальностей «математика», «прикладна математика» ,«статистика».

Рекомендовано до друку вченою радою ДВНЗ «Ужгородський національний університет», протокол протокол № 3 від 24 лютого 2017 р.

Рецензенти: **Слюсарчук П.В.,** кандидат фізико-математичних наук, завідувач кафедри теорії ймовірностей і математичного аналізу (ДВНЗ «Ужгородський національний університет»);
Мулеса О.Ю., кандидат технічних наук, доцент кафедри кібернетики і прикладної математики (ДВНЗ «Ужгородський національний університет»).

Означення 13.3. Оцінка $\hat{\theta}_1$ називається *асимптотично байєсівською* або або *асимптотично мінімаксною*, якщо для будь-якої іншої оцінки θ^* виконується відповідно

$$\lim_{n\to\infty} \sup \left[M_n\left(\hat{\theta}_1-\theta\right)^2 - M_n\left(\hat{\theta}-\theta\right)^2\right] \le 0,$$

$$\lim_{n\to\infty} sup\left[\sup_{t\in\Gamma} M_t n\left(\hat{\theta}_1-\theta\right)^2 - \sup_{t\in\Gamma} M_t n\left(\hat{\theta}-\theta\right)^2\right] \le 0.$$

Відшукання асимптотично байєсівських і асимптотично мінімаксних оцінок можливе, як ми побачимо, при вельми широких припущеннях.

В багатовимірному випадку (коли $\theta \in R^k$ вектор) оцінка $\hat{\theta}_Q = M(\theta/X)$ буде мінімізувати

$$v(\hat{\theta}) = M\left(\hat{\theta}-\theta\right)V\left(\hat{\theta}-\theta\right)^T = MM_\theta\left(\hat{\theta}-\theta\right)V\left(\hat{\theta}-\theta\right)^T =$$

$$= \int M_t\left(\hat{\theta}-t\right)V\left(\hat{\theta}-t\right)^T q(t)\lambda(dt)$$

для будь-якої невід'ємно визначеної матриці V або, що те саме, мінімізувати усереднене (з вагою $q(t)$) середньоквадратичне відхилення $\hat{\theta}-\theta$ по будь-якому напрямку $a \in R^k$.

Означення 13.4. Оцінка $\hat{\theta}_Q$ називається *байєсівською*, якщо для будь-якої іншої оцінки $\hat{\theta}$ і для будь-якої невід'ємно визначеної матриці V

$$v\left(\hat{\theta}_Q\right) \le v\left(\hat{\theta}\right).$$

Оцінка θ_1^* називається асимптотично байєсівською, якщо

$$\lim_{n\to\infty} sup\left[nv\left(\hat{\theta}_1\right) - nv(\hat{\theta}_Q)\right] \le 0.$$

Означення 13.5. Оцінка $\bar{\theta}^*$ називається *мінімаксною*, якщо для будь-якої оцінки θ^* і для будь-якої невід'ємно визначеної матриці V

$$\sup_{t\in\Gamma} M_t\left(\hat{\theta}_Q-t\right)V\left(\hat{\theta}_Q-t\right)^T - \sup_{t\in\Gamma} M_t\left(\hat{\theta}-t\right)V\left(\hat{\theta}-t\right)^T \le 0.$$

Оцінка θ_1^* називається *асимптотично мінімаксною*, якщо

$$\lim_{n\to\infty} sup\left[\sup_{t\in\Gamma} M_t n\left(\hat{\theta}_1-t\right)^2 V\left(\hat{\theta}_1-t\right)^T - \sup_{t\in\Gamma} M_t n(\bar{\theta}^*-t)^2 V(\bar{\theta}^*-t)^T\right] \le 0.$$

При $N+1=\sqrt{n}/2$ ця оцінка буде співпадати з оцінкою p^*, визначеною у (13.8), і в силу теореми 2 буде мінімаксною. Розподіл Q буде найгіршим. Він концептується зі збільшенням n навколо «найгіршого» значення параметра $p=1/2$, при якому дисперсія оцінки $\bar{x}$, рівна $p\,(1-p)/n=1/(4n)$, буде максимальна. Сама оцінка $\bar{x}$ не мінімаксна, так як

$$\sup_p \frac{p(1-p)}{n}=\frac{1}{4n}>\frac{1}{4(1+\sqrt{n})^2}.$$

В той же час зрозуміло, що для всіх значень p, які лежать поза вузьким околом точки $p=1/2$, оцінка $\bar{x}$ все ж буде кращою, чим p_Q^* - це буде мати місце для всіх p, для яких

$$p(1-p)<\frac{1}{4(1+1/\sqrt{n})^2}.$$

В загальному випадку пошук точних виразів (явних функцій від X) для байєсівських і мінімаксних оцінок не завжди можливий. Тому природно використовувати також асимптотичний підхід.

Перед тим як вводити відповідне визначення, нагадаємо, що байєсівські і мінімаксні оцінки $\hat{\theta}_Q$ і $\bar{\theta}^*$ визначались нерівностями

$$M(\hat{\theta}_Q-\theta)^2-M(\hat{\theta}-\theta)^2\le 0,$$

$$\sup_{t\in\Gamma} M_t(\bar{\theta}^*-t)^2-\sup_{t\in\Gamma} M_t\left(\hat{\theta}-t\right)^2\le 0 \qquad (13.9)$$

для будь-якої оцінки $\hat{\theta}$. Було б нерозумно визначати асимптотичну байєсовість і мінімаксність оцінок, просто добавляючи до лівих частин знак граничного переходу $\lim_{n\to\infty}$, так як звичайно для а. н. оцінок $M_\theta(\hat{\theta}-\theta)^2\sim\sigma^2(\theta)/n$ ліві частини в (13.9) будуть прямувати до нуля. Тому природно розглядати відношення складених в (13.9). Враховуючи, що в подальшому ми будемо мати справу головним чином з оцінками, для яких $M_\theta(\hat{\theta}-\theta)^2$ має порядок малості $1/n$, можна еквівалентним чином використати наступне визначення.

ЗМІСТ

Приклад 13.2. Нехай $X \in B_p$, $x_j, j = 1, \dots, n$, приймають значення 1 і 0 відповідно з ймовірностями p і $1-p, p \in \Theta = [0,1]$. Як ми знаємо, в цьому випадку для оцінки $p^* = \bar{x}$ справедливо

$$M_p(\bar{x}-p)^2 = p\,(1-p)/n,$$

так що критерій теореми 2 не виконаний. Розглянемо оцінку

$$p^* = \frac{\bar{x}+\frac{1}{2\sqrt{n}}}{1+\frac{1}{\sqrt{n}}}. \qquad (13.8)$$

Для неї похибка

$$M_p(p^*-p)^2 = \left(1+\frac{1}{\sqrt{n}}\right)^{-2} M_p\left(\bar{x}-p+\frac{1}{2\sqrt{n}}-\frac{p}{\sqrt{n}}\right)^2 =$$

$$\frac{n}{\left(1+\sqrt{n}\right)^2}\left(\frac{p(1-p)}{n}+\frac{(1-2p)^2}{4n}\right) = \frac{1}{4\left(1+\sqrt{n}\right)^2}$$

не залежить від p. Якщо ми переконаємося тепер, що оцінка (13.8) є байєсівською, то ми встановимо тим самим її мінімаксність. Розглянемо апріорний розподіл $Q \simeq B(N+1, N+1)$, де $B(\lambda_1, \lambda_2)$ – це бета-розподіл зі щільністю

$$\frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} t^{\lambda_1-1}(1-t)^{\lambda_2-1}.$$

Тоді, оскільки

$$f_t(X) = t^{\bar{x}n}(1-t)^{n(1-\bar{x})}, \qquad q(t) = \frac{\Gamma(2N+2)}{\Gamma^2(N+1)} t^N (1-t)^N,$$

то апостеріорний розподіл буде мати щільність $q(t/X)$, яка як функція від t буде пропорційна $f_t(X)q(t)$ або, що теж, пропорційна

$$t^{N+\bar{x}n}(1-t)^{N+(1-\bar{x})n}.$$

Це означає, що апостеріорний розподіл співпадає з $B(N+\bar{x}n+1, N+n(1-\bar{x})+1)$. Так як середнє значення розподілу $B(\lambda_1, \lambda_2)$ рівне $\lambda_1/(\lambda_1+\lambda_2)$, то байєсівська оцінка p_Q^*, відповідна Q, буде рівна

$$p_Q^* = \frac{N+\bar{x}n+1}{2N+n+2} = \frac{\bar{x}+(N+1)/n}{1+2\,(N+1)/n}.$$

Приклад 13.1. Нехай $X \simeq N(\alpha, 1)$. З'ясуймо, що собою являє байєсівська оцінка $\alpha^*_{Q^{(k)}}$ параметра α із апріорним нормальним розподілом $Q^{(k)} \simeq N(0, k)$. В цьому випадку ми повинні покласти $\lambda(dt) = dt$,

$$q^{(k)}\,(\mathrm{t}) = \frac{1}{\sqrt{2\pi k}} \mathrm{e}^{-\frac{\mathrm{t}^2}{2\mathrm{k}}}.$$

Апостеріальний розподіл $Q_X^{(k)}$ буде мати щільність $q^{(k)}(\mathrm{t}/X)$, пропорційну (як функція від t) $q^{(k)}\,(\mathrm{t}), \mathrm{f_t}(\mathrm{X})$ або, що також , пропорційну

$$exp\left\{-\frac{t^2}{2k} - \frac{1}{2}\sum(x_i - t)^2\right\}.$$

Із рівності

$$-\frac{t^2}{2}\left(\frac{1}{k} + n\right) + \bar{x}nt = -\frac{1}{2}\left(\frac{1}{k} + n\right)\left(t - \frac{\bar{x}n}{\frac{1}{k} + n}\right)^2 + \frac{(\bar{x}n)^2}{2\left(\frac{1}{k} + n\right)}$$

випливає, що

$$Q_X^{(k)} \simeq N(\frac{\bar{x}nk}{1 + nk}, \frac{k}{1 + nk}).$$

Так як байєсівська оцінка $\alpha^*_{Q^{(k)}}$ параметра α рівна математичному сподіванню апостеріорного розподілу, то звідси отримаємо

$$\alpha^*_{Q^{(k)}} = \frac{\bar{x}nk}{1 + nk} = \frac{\bar{x}}{1 + \frac{1}{nk}}.$$

Дисперсія апостеріорного розподілу $\sigma^2_{Q_X^{(k)}} = \frac{k}{1+nk}$ від X не залежить. Отже, в силу (13.4) середньоквадратична похибка байєсівської оцінки рівна $\frac{k}{1+nk} \to \frac{1}{n}$ при $k \to \infty$. Тому для оцінки $\alpha^* = \bar{x}$ маємо

$$M_t(\bar{x} - t)^2 = \frac{1}{n} = \lim_{k\to\infty} \int M_t\left(\alpha^*_{Q^{(k)}} - t\right)^2 q^{(k)}(t)dt,$$

і, отже, оцінка $\alpha^* = \bar{x}$ мінімаксна. «Найгіршим» розподілом тут був би рівномірний розподіл на всій прямій, якщо б такий існував.

ВСТУП

Математична статистика – це розділ математики, що базується на поняттях та методах теорії ймовірностей та призначений для одержання обґрунтованих висновків за результатами спостережень над відповідним випадковим експериментом.

На сьогодні, у зв'язку з стрімким розвитком багатьох природничих та технічних наук, актуальним є широке застосування імовірнісних та статистичних методів у багатьох галузях науки і техніки. Метою проведення статистичних досліджень є пошук та дослідження співвідношень між статистичними даними та їх використання для вивчення, прогнозування і прийняття рішень.

Серед основних задач математичної статистики виділяють:

1. *Оцінка невідомих параметрів розподілу ймовірностей значень випадкової змінної на основі отриманої вибірки.*

2. *Побудова довірчих інтервалів для параметрів розподілу ймовірностей генеральної сукупності.*

3. *Перевірка гіпотез про вигляд невідомого розподілу або про величину параметрів розподілу, вигляд якої відомий.*

4. *Задача виявлення тенденцій, закономірностей, залежностей між випадковими змінними на основі даних вибірок з цих змінних.*

Навчальний посібник містить курс лекцій з дисципліни «Додаткові розділи статистики» та базується на дисципліні «Теорії ймовірностей і математична статистика». Особливістю посібника є його теоретична спрямованість.

Курс складається з 13 розділів та спрямований на викладення сучасних основ математичної статистики. У першому розділі описано основні закони розподілу випадкових величин та їх основні числові характеристики. У другому розділі наведено основні твердження про розподіл лінійних і квадратичних форм від гауссівських випадкових величин. Третій розділ присвячений короткому викладу основ моделювання випадкових величин. У четвертому розділі наведено поняття

Таким чином, мінімаксна оцінка – це байєсівська оцінка, яка «вирівнює» похибки $M_t(\bar{\theta}^* - t)^2$ при різних t. Це означає, що апріорний розподіл $\bar{Q}$, відповідне цій оцінці, змушує бути однаково уважним до всіх можливих значень θ, а не орієнтуватися, як це роблять байєсівські оцінки $\hat{\theta}_Q$, відповідні іншим апріорним розподілам $Q \neq \bar{Q}$, на якісь виділені (вірогідніші) значення θ. Оскільки в останньому випадку ми використовуємо додаткову інформацію про θ, то звичайно, що при $Q \neq \bar{Q}$ оцінки $\hat{\theta}_Q$ будуть володіти меншими значеннями безумовних середньоквадратичних відхилень

$$\int M_t\left(\hat{\theta}_Q - t\right)^2 Q(dt) \leq \int M_t\left(\hat{\theta}_Q - t\right)^2 \bar{Q}(dt).$$

У зв'язку з цим розподіл $\bar{Q}$ в теоремі 13.2, відповідно мінімаксній оцінці $\bar{\theta}^*$, часто називають *найгіршим*.

Так як такий найгірший розподіл $\bar{Q}$ не завжди існує (це буває, у випадках, коли Θ – необмежена множина), розглянемо модифікований критерій для відшукання мінімаксної оцінки.

Теорема 13.3. Якщо існують оцінка $\hat{\theta}_1$ і послідовність розподілу $Q^{(k)}$ зі щільностями $q^{(k)}$ такі, що при всіх t

$$M_t\left(\hat{\theta}_1 - t\right)^2 \leq \lim_{k\to\infty} sup \int M_t\left(\theta^*_{Q^{(k)}} - t\right)^2 q^{(k)}(\mathrm{t})\,\lambda(dt),$$

то оцінка $\hat{\theta}_1$- мінімаксна.

Доведення цієї теореми настільки ж просте. Для будь-якої оцінки θ^* справедливо

$$\sup_t M_t\left(\hat{\theta} - t\right)^2 \geq \int M_t\left(\hat{\theta} - t\right)^2 q^{(k)}(\mathrm{t})\lambda(dt) \geq$$
$$\int M_t\left(\theta^*_{Q^{(k)}} - t\right)^2 q^{(k)}(\mathrm{t})\,\lambda(dt).$$

Звідси випливає, що

$$\sup_t M_t\left(\hat{\theta} - t\right)^2 \geq \lim_{k\to\infty} sup \int M_t\left(\theta^*_{Q^{(k)}} - t\right)^2 q^{(k)}(\mathrm{t})\,\lambda(dt) \geq$$
$$\geq M_t\left(\hat{\theta}_1 - t\right)^2. \blacksquare$$

$$inf_{\theta^*} \sup_{t\in\Gamma} M_t\big(\hat{\theta}-t\big)^2 = \sup_{t\in\Gamma} M_t(\bar{\theta}^*-t)^2. \qquad (13.5)$$

Встановимо деякі корисні зв'язки між байєсівськими і мінімаксними оцінками.

Теорема 13.1. Позначимо через $\hat{\theta}_Q$ байєсівську оцінку для апріорного розподілу Q зі щільністю q. Якщо існують оцінка $\hat{\theta}_1$ і розподіл Q, такі що при всіх t

$$M_t\big(\hat{\theta}_1-t\big)^2 \le \int M_u\left(\hat{\theta}_Q-u\right)^2 q(u)\lambda(du), \qquad (13.6)$$

то оцінка $\hat{\theta}_1$ мінімаксна.

Доведення. Нехай $\hat{\theta}$ - будь-яка інша оцінка. Тоді

$$\sup_t M_t\big(\hat{\theta}-t\big)^2 \ge \int M_t\big(\hat{\theta}-t\big)^2 q(t)\lambda(dt) \ge$$

$$\ge \int M_t\big(\hat{\theta}_Q-t\big)^2 q(t)\lambda(dt) \ge M_t\big(\hat{\theta}_1-t\big)^2. \blacksquare$$

Зауважимо, що для майже всіх t, які належать носію $N_Q=\{t: q(t)>0\}$ розподілу Q, в нерівності (6) з необхідністю повинна виконуватись рівність, так як в протилежному випадку ми дістали б

$$\int M_t\big(\hat{\theta}_1-t\big)^2 q(t)\lambda(dt) < \int M_t\big(\hat{\theta}_Q-t\big)^2 q(t)\lambda(dt),$$

що суперечить визначенню байєсівської оцінки.

Це зауваження дозволяє нам сформулювати наступний критерій мінімаксності оцінки, еквівалентний теоремі 13.1.

Теорема 13.2. Якщо оцінка $\hat{\theta}$

1) байєсівська для для деякого розподілу Q,

2) $M_t\big(\hat{\theta}-t\big)^2 = c = const$ при $t \in N_Q$,

3) $M_t\big(\hat{\theta}-t\big)^2 \le c$ для інших t,

то $\hat{\theta}$ - мінімаксна оцінка.

Якщо $\hat{\theta}=\hat{\theta}_Q=\bar{\theta}^*$ задовільняє цьому критерію, то, очевидно,

$$\sup_t M_t(\bar{\theta}^*-t)^2 = \int M_t(\bar{\theta}^*-t)^2\, q(t)\lambda(dt). \qquad (13.7)$$

вибірки, емпіричного розподілу та доведено теорему Глівенко-Кантеллі. Також у цьому розділі розглянуто поняття статистики, оцінки та властивості оцінок.

У п'ятому розділі викладено основні методи побудови точкових оцінок невідомих параметрів розподілу. Значна увага приділяється опису методу моментів. Шостий та сьомий розділи посібника присвячені відповідно побудові оптимальних оцінок на основі нерівності Крамера-Рао та побудові оптимальних оцінок за допомогою достатності і незміщенності. У восьмому розділі обґрунтовано основні принципи оцінювання параметрів розподілу методом максимальної вірогідності. У дев'ятому розділі описано процес побудови інтервальних оцінок невідомих параметрів розподілу, зокрема для параметрів нормальних спостережень.

Десятий розділ присвячений перевірці статистичних гіпотез. У ньому наведено основні терміни та поняття з теорії перевірки гіпотез, непараметричні критерії для функції розподілу, рангові критерії перевірки гіпотез, головні положення щодо критерію хі-квадрат Пірсона, а також розглянено перевірку гіпотез про значення параметрів нормальної вибірки. У одинадцятому розділі розглянуто критерій відношення вірогідностей і лема Неймана-Пірсона та деякі поняття послідовного аналізу. Дванадцятий розділ містить метод найменших квадратів для лінійної регресії, а тринадцятий – байєсівський та мінімаксний підходи до оцінювання параметрів.

В списку літератури перераховані підручники та посібники, які можна використовувати для підготовки та самостійного опрацювання студентами до даної дисципліни. Підрозділи, формули, приклади, твердження і т. п. мають наскрізну нумерацію у кожному розділі. При посиланні на об'єкт з якогось розділу наводиться номер формули, твердження тощо. Закінчення доведень позначається ■.

РОЗДІЛ 1. Закони розподілу, що використовуються при реалізації техніки статистичних обчислень

Зі стохастичним експериментом можна пов'язати певне висловлювання про його результат. Оскільки для кожної елементарної події можна встановити, справедливе дане висловлювання чи ні, то довільному висловлюванню відповідає певна множина елементарних подій, а саме: така підмножина простору елементарних подій, для елементів якої справджується дане висловлювання. *Випадковими подіями* називаються підмножини простору елементарних подій, для яких справджуються певні висловлювання про результат стохастичного експерименту.

Поняття випадкової події дозволяє вивчати якісні (дихотомічні) наслідки стохастичних експериментів. На практиці важливо вивчати також кількісні результати (наприклад: кількість аверсів при 2 підкиданнях монети). Для визначення такої кількості необхідно кожній елементарній події поставити у відповідність певне число – кількісний результат експерименту, тобто визначити функцію на просторі елементарних подій.

1.1. Розподіли дискретних випадкових величин

Означення 1.1. *Дискретною випадковою величиною* називається відображення *ξ*: Ω→R, що має скінченну або зліченну множину значень $\xi(\Omega) = \{\xi(\omega), \omega \in \Omega\} = \{x_1, \dots, x_n\}$, причому для кожного можливого значення x_n прообраз $\{\omega: \xi(\omega) = x_n\}$ є випадковою подією.

Означення 1.2. *Розподілом* дискретної випадкової величини *ξ(ω)* зі значеннями $\{x_n\}$ називається послідовність ймовірностей

$$p_n \equiv P\{\xi = x_n\}, n \geq 1.$$

$$\hat{\theta}_Q = M(\theta/X) = \int tq(t/x)\ \lambda(dt) = \int tQx(dt). \qquad (13.2)$$

Означення 13.1. Оцінка $\hat{\theta}_Q$, визначена формулами (13.2), (13.1), називається *байєсівською*, відповідною апріорному розподілу Q з щільністю $q(t)$.

Відмітимо ще раз, що для байєсівської оцінки безумовне середньоквадратичне відхилення

$$M(\hat{\theta}-\theta)^2 = MM\left(\left(\hat{\theta}-\theta\right)^2/\theta\right) = MM_\theta(\hat{\theta}-\theta)^2 =$$

$$= \int M_t\left(\hat{\theta}-t\right)^2 q(t)\,\lambda(dt) \qquad (13.3)$$

приймає найменше можливе значення. Співвідношення (13.3) показує, що байєсівська оцінка мінімізує середнє значення (з даною ваговою функцією $q(t)\lambda(dt)$) величини $M_t\left(\hat{\theta}-t\right)^2$.

Іншими словами, якщо θ вибирається випадково, з щільністю $q(t)$, то байєсівська оцінка є найкращою в розумінні середньоквадратичного підходу. Середньоквадратичне відхилення (13.3) байєсівської оцінки можна зобразити у вигляді:

$$M(\hat{\theta}_Q-\theta)^2 = \int M_t\left(\hat{\theta}_Q-\theta\right)^2 q(t)\lambda(dt) =$$

$$\iint \left(t-\hat{\theta}_Q\right)^2 f_t(x)q(t)\lambda(dt)\mu^n(dx) = \int \sigma^2_{Q_x} f(x)\mu^n(dx) = M\sigma^2_{Q_X},$$

де $\sigma^2_{Q_X}$- дисперсія апостеріорного розподілу Q_X:

$$\sigma^2_{Q_X} = \int \left(t-\hat{\theta}_Q\right)^2 q(t/X)\lambda(dt) = \int \left(t-M(\theta/X)\right)^2 Q_X(dt). \qquad (13.4)$$

Інший підхід до порівняння оцінок заснований на порівнянні $sup_{t\in\Gamma} M_t(\hat{\theta}-t)^2$, де $\Gamma \subset \Theta$ – задана підмножина Θ (Γ або співпадає з Θ, або рівне тій його частині, відносно якої вдалось встановити, що $\theta \in \Gamma$).

Означення 13.2. Оцінка $\bar{\theta}^*$ називається *мінімаксною*, якщо будь-якої іншої оцінки $\hat{\theta}$

$$\sup_{t\in\Gamma} M_t(\bar{\theta}^*-t)^2 \leq \sup_{t\in\Gamma} M_t\left(\hat{\theta}-t\right)^2.$$

Іншими словами, для мінімаксної оцінки досягається

РОЗДІЛ 13. Байєсівський та мінімаксний підходи до оцінювання параметрів

Суть байєсівського підходу полягає в тому, що невідомий параметр θ розглядається як випадкова величина з деякою (відомою чи невідомою) щільністю розподілу $q(t), t \in \Theta$, відносно міри λ, яка як і міра μ в умові (A_μ), найчастіше складатиме або міру Лебега, або зліченну міру. Щільність $q(t)$ називається *апріорною*, тобто даною до експерименту. Байєсівський підхід припускає, що невідомий параметр θ вибраний випадковим чином з розподілу зі щільністю $q(t)$.

Нехай, далі, $f_t(x), t \in \Theta, x \in æ^n$ – функція правдоподібності, яка при кожному t є щільністю розподілу в $æ^n$. Тому функція $f(x,t) = f_t(x)q(t)$ є щільність деякого розподілу в $æ^n \times \Theta$ відносно міри $\mu^n \times \lambda$, яку можна інтерпретувати як щільність спільного розподілу X і θ. При такому підході в силу теореми функція $f_t(x), x \in æ^n$є умовна щільність X при умові $\theta = t$:

$$f_t(x) = f(x/t), \qquad M_\theta g(X) = M(g(X)/\theta).$$

В цих розглядах формальна сторона справи потребує, щоб $f_t(x)$ була вимірною по t і x функцією. Надалі, будемо припускати, що ця властивість має місце.

Поряд з $f(x/t)$ ми можемо виписати умовну щільність $q(t/x)$ величини θ при умові $X = x$:

$$q(t/x) = \frac{f_t(x)q(t)}{f(x)},\ f(x) = \int f_t(x)q(t)\lambda(dt). \qquad (13.1)$$

Ця щільність визначає так званий апостеріорний (тобто після експеримента) розподіл θ, який ми будемо позначати Q_x. Рівність (13.1) називається *формулою Байєса* для щільності апостеріорного розподілу.

Властивість умовного математичного сподівання стосовно байєсівського випадку означає наступне: серед всіх функцій $\hat{\theta} = \varphi(x)$ найкращою оцінкою для θ (в розумінні мінімізації $M(\theta - \varphi(X))^2$) є функція

Ця послідовність є невід'ємною і в сумі становить 1, тобто є *дискретним розподілом імовірностей.*

Значна кількість дискретних розподілів пов'язана із такою схемою випробувань.

Означення 1.3. *Схемою випробувань Бернуллі* називається стохастичний експеримент, що зводиться до послідовності:

(а) незалежних у сукупності випробувань;

(б) кожне з яких є дихотомічним – закінчується одним із двох результатів: успіх або неуспіх, причому

(в) ймовірність успіху не залежить від номера випробування.

Для побудови ймовірнісного простору, який відповідає схемі Бернуллі, позначимо через n кількість випробувань, p – ймовірність успіху в одному випробуванні, $q = 1 - p$ – ймовірність неуспіху. Якщо ототожнити значення 1 з успіхом та 0 із неуспіхом, то елементарна подія матиме вигляд

$$\omega = \{\omega_1, \dots, \omega_n\}, \omega_k \in \{0,1\}\ .$$

Для відшукання ймовірностей зафіксуємо точку $\varepsilon = \{\varepsilon_1, \dots, \varepsilon_n\} \in \{0,1\}^n$ та розглянемо одноточкову подію

$$\{\varepsilon\} = \{\omega : \omega = \varepsilon\} = \bigcap_{k=1}^n \{\omega : \omega_k = \varepsilon_k\}.$$

Звідси

$$P\{\omega : \omega = \varepsilon\} = \prod_{k=1}^n P\{\omega : \omega_k = \varepsilon_k\} =$$

$$= \prod_{k=1}^n p^{\varepsilon_k} q^{1-\varepsilon_k} = p^{\nu_n(\varepsilon)} q^{n-\nu_n(\varepsilon)},$$

де функція

$$\nu_n(\varepsilon) = \sum_{k=1}^n \omega_k = |k : \omega_k = 1|,$$

збігається з загальною кількістю успіхів в елементарній події $\omega = (\omega_1, \dots, \omega_n)$. Як в усякому дискретному ймовірнісному просторі, ймовірність будь-якої події дорівнює сумі ймовірностей сприятливих елементарних подій.

Означення 1.4. Дискретна випадкова величина ξ має *біноміальний розподіл* із параметрами n та p, позначення $\xi \simeq B(n,p)$, якщо вона збігається з загальною кількістю успіхів у схемі випробувань Бернуллі з n випробуваннями та ймовірністю успіху p в одному випробуванні.

Для обчислення розподілу ξ зауважимо, що за означенням $\xi(\omega) = \nu_n(\omega)$, де функцію ν_n визначено вище при описанні схеми Бернуллі. Тому

$$P\{\xi = k\} = \sum_{\omega:\,\xi(\omega)=k} p^{\nu_n(\omega)} q^{n-\nu_n(\omega)} = |\{\omega: \xi(\omega) = k\}| p^k q^{n-k}.$$

Оскільки кількість елементарних подій у події $\{\xi = k\}$ дорівнює кількості сполучень (k-елементних підмножин множини з n елементів) C_n^k, то остаточно маємо

$$P\{\xi = k\} = p_n(k) = C_n^k p^k q^{n-k}, k = \overline{0,n}.$$

Послідовність $(p_n(k),\ \ k = \overline{0,n})$ називається *біноміальним розподілом* імовірностей. Випадкова величина $\xi \simeq B(n,p)$ має наступні числові характеристики: $M\xi = p, D\xi = npq$.

Означення 1.5. Дискретна випадкова величина ξ має *геометричний розподіл* із параметром p, позначення $\xi \simeq G(p)$, якщо вона збігається із кількістю випробувань до першого успіху в нескінченній послідовності випробувань Бернуллі з імовірністю успіху p в одному випробуванні.

Геометричний розподіл імовірностей має вигляд

$$P\{\xi = k\} = p(1-p)^{k-1}, k = 1,2, \ldots.$$

Випадкова величина $\xi \simeq G(p)$ має наступні числові характеристики: $M\xi = \frac{1}{p}, D\xi = \frac{1-p}{p^2}$.

Означення 1.6. Дискретна випадкова величина ξ має *негативний геометричний розподіл* із параметрами p, r, позначення $\xi \simeq G(p,r)$, якщо вона збігається з кількістю випробувань до r-го успіху в нескінченній послідовності випробувань Бернуллі з імовірністю успіху p в одному випробуванні. Розподіл має вигляд

$$P\{\xi = k\} = C_{k-1}^{r-1} p^r (1-p)^{k-r}, k = 1,2, \ldots.$$

Дійсно, кожна сприятлива елементарна подія довжини k містить $k- r$ неуспіхів та r успіхів, причому останнім має бути успіх. Тому ймовірність такої події

Можна довести, що визначник цієї матриці дуже близький до нуля навіть при великих k. Отже, обчислення оцінки МНК стикатиметься з великими труднощами.

Тому для покращення якості оцінювання в поліноміальній схемі доцільно розглянути її в еквівалентному вигляді

$$\xi_j = \sum_{i=0}^{k} \theta_i \varphi_i(x_j) + \varepsilon_j\,, \quad j = \overline{1,n},$$

де $\varphi_i(x)$ – деякі поліноми ступеня i від x. Матриця V у такій схемі набуває вигляду

$$V = \left(\sum_{j=1}^{n} \varphi_i(x_j) \varphi_l(x_j)\,, 0 \le i,l \le k \right).$$

Для покращення якостей оцінки МНК поліноми φ_i обирають так, щоб матриця V була діагональною. З огляду на її визначення розглянемо таку білінійну форму (скалярний добуток) від функцій f,g: $(f,g) = \sum_{j=1}^{n} f(x_j) g(x_j)$.

Визначимо рекурентно послідовність *поліномів Чебишева*

$$\varphi_0(x) \equiv 1, \varphi_k(x) = x^k - \sum_{i=0}^{k-1} \varphi_i(x) \frac{(x^k,\varphi_i)}{(\varphi_i,\varphi_i)}, \qquad k \ge 1.$$

За індукцією з цього означення виводимо, що для всіх $i < k$

$$(\varphi_k,\varphi_i) = (x^k,\varphi_i) - (\varphi_i,\varphi_i)(x^k,\varphi_i)/(\varphi_i,\varphi_i) = 0.$$

Отже, поліноми Чебишева попарно ортогональні, матриця $V = \left((\varphi_i,\varphi_i)\delta_{ik}\right)$ є діагональною і оцінка МНК дорівнює

$$\hat{\theta} = V^{-1}TX = \left(\sum_{j=1}^{n} \xi_j \varphi_i(x_j) / (\varphi_i,\varphi_i), \qquad i = \overline{0,k} \right).$$

Оскільки коваріаційна матриця оцінки $\hat{\theta}$ за теоремою про незміщеність і розподіл оцінки МНК дорівнює діагональній матриці $\sigma^2 V^{-1}$, то координати $\hat{\theta}_i$ некорельовані, а у випадку нормальних похибок – незалежні.

$$\hat{a} = \frac{1}{n\bar{\sigma}^2}\sum_{j=1}^{n}\xi_j\left(\overline{x^2} - x_j\bar{x}\right) = -\bar{x}\hat{b} + \frac{1}{n\bar{\sigma}^2}\sum_{j=1}^{n}\xi_j\left(\overline{x^2} - (\bar{x})^2\right) =$$

$$= \frac{1}{n}\sum_{j=1}^{n}\xi_j - \bar{x}\hat{b} = \hat{\mu}_n - \bar{x}\hat{b}.$$

Оцінка $\hat{b}$ називається *вибірковим коефіцієнтом кореляції* векторів (ξ_j), (x_j). За теоремою про незміщеність і розподіл оцінки МНК $D(\hat{b}) == \sigma^2 (V^{-1})_{22} = \sigma^2 / n\bar{\sigma}^2$. Якщо спостереження ξ_j отримано з активного експерименту, тобто статистик має можливість обирати точки спостережень (x_j), то для мінімізації дисперсії оцінки $\hat{b}$ слід максимізувати величину $\bar{\sigma}^2$.

12.3. Поліноміальна регресія

Попередню задачу лінійної апроксимації спостережень можна узагальнити, якщо розглянути схему поліноміальної регресії:

$$\xi_j = \sum_{i=0}^{k}\theta_i x_j^i + \varepsilon_j\,, \quad j = \overline{1,n},$$

де (x_j) – задані точки спостережень. Незважаючи на нелінійний характер залежності від x, дана модель залишається лінійною за невідомим параметром $\theta = (\theta_1, \dots, \theta_k)$. Тому для аналізу поліноміальної регресії можна застосувати всі результати розділу про лінійну регресію, якщо визначити матрицю перетворень $T = \left(x_j^i, \quad i = \overline{0,k}, \quad j = \overline{1,n}\right)$.

Припустимо, що точки спостережень (x_j) "рівномірно розподілені" на відрізку $(0{,}1)$, тобто $\sum_{j=1}^{n} g(x_j) \approx \int_0^1 g(x)dx$ для $g \in C(0{,}1)$. Тоді матриця $V = TT'$ дорівнює

$$V = \left(\sum_j x_j^i x_j^l\,, 0 \le i,l \le k\right) \approx (1/(i+l+1), 0 \le i,l \le k).$$

дорівнює $p^r(1-p)^{k-r}$, а загальна кількість їх збігається з кількістю розміщень $r-1$ успіху на перших $n-1$ місці.

Випадкова величина $\xi \simeq G(p,r)$ має наступні числові характеристики: $M\xi = \frac{r}{p}, D\xi = \frac{r(1-p)}{p^2}$.

Гіпергеометричний розподіл виникає, наприклад, у випробуваннях, коли з комплекту, що складається з N предметів, n з яких мають певну властивість (нестандартні, пофарбовані тощо), відбирається навмання m предметів (одноразово, або послідовно без повернення до комплекту), а випадкова величина ξ – кількість предметів із зазначеною властивістю серед відібраних, позначення $\xi \simeq H(N,n,m)$.

Для обчислення розподілу величини ξ приймемо класичне означення ймовірностей та зауважимо, що кількість всіх елементарних подій дорівнює C_N^m – кількості способів обрати m предметів із загального числа N. Кількість елементарних подій, що сприяють події $\{\xi = k\}$, очевидно, дорівнює за основним правилом комбінаторики $C_n^k C_{N-n}^{m-k}$– добуткові кількості способів обрати k предметів із загальної кількості n та $m-k$ предметів із кількості $N-n$. Отже, гіпергеометричний розподіл має вигляд

$$P\{\xi = k\} = \frac{C_n^k C_{N-n}^{m-k}}{C_N^m}.$$

Випадкова величина $\xi \simeq H(N,n,m)$ має наступні числові характеристики: $M\xi = n\frac{m}{N}, D\xi = n\frac{m}{N}\left(1-\frac{m}{N}\right)\left(1-\frac{n-1}{N-1}\right)$.

Зауваження 1.1. Якщо проінтерпретувати N як загальну кількість виробів, що виготовлені на підприємстві, n – як кількість бракованих серед них, m – як об'єм вибіркової партії, що була перевірена відділом контролю якості, то випадкова кількість виявлених бракованих виробів матиме гіпергеометричний розподіл. Цей факт використовують, зокрема, при статистичному контролі якості продукції.

Означення 1.7. Дискретна випадкова величина має розподіл Пуассона з параметром $\lambda > 0$, позначення $\xi \simeq \Pi(\lambda)$, якщо

$$P\{\xi = n\} = \frac{\lambda^n}{n!}e^{-\lambda}, n = 0{,}1{,}2 \,\dots.$$

Випадкова величина $\xi \simeq \Pi(\lambda)$ має наступні числові характеристики: $M\xi = D\xi = \lambda$.

Нормованість пуассонівських імовірностей випливає з формули розкладу в ряд Тейлора функції exp*(λ)*. Розподіл Пуассона широко використовується в теорії ймовірностей та статистиці для моделювання випадкових потоків подій, таких як кількість телефонних викликів за певний час, кількість захворювань за період, кількість випадково обраних точок у певному об'ємі тощо. Одночасно він є граничним для біноміального розподілу.

1.2. Розподіли неперервних випадкових величин

Сукупність ймовірностей подій вигляду *{* $\xi = x$*}* не завжди повністю описує величину ξ. Наприклад, для випадкової точки на відрізку всі ці ймовірності дорівнюють нулю. Тому в загальному випадку на відміну від дискретного для визначення розподілу випадкової величини використовують прообрази інтервалів.

Означення 1.8. *Функцією розподілу* випадкової величини ξ називається функція $F_\xi(x)$ дійсного аргументу x, яка задається рівністю

$$F_\xi(x) = P\{\xi < x\}, x \in R.$$

Дискретна функція розподілу – функція розподілу дискретної випадкової величини – є сталою в кожному околі своєї точки неперервності та збігається із сумою своїх стрибків на інтервалі $(-\infty, x)$. Якщо випадкова величина ξ набуває значень $\{x_n, n \geq 1\}$ із ймовірностями $\{p_n, n \geq 1\}$, то за означенням:

$$F_\xi(x) = P\left(\bigcup_{n:x_n<x}\{\xi = x_n\}\right) = \sum_{n:x_n<x} p_n.$$

Означення 1.9. Випадкова величина ξ та її функція розподілу F_ξ називаються *абсолютно неперервними*, якщо існує невід'ємна функція $f_\xi(x)$ така, що

$$F_\xi(x) = \int_{-\infty}^{x} f_\xi(y)dy, \forall x \in R,$$

де інтеграл слід розуміти

(a) як інтеграл Рімана – Стілтьєса для кусково-неперервної функції $f_\xi(x)$,

$$= \sigma^2(ZV^{-1}T + C - ZV^{-1}T)(ZV^{-1}T + C - ZV^{-1}T)' =$$
$$= \sigma^2 ZV^{-1}Z' + \sigma^2(C - ZV^{-1}T)(C - ZV^{-1}T)' \geq$$
$$\geq \sigma^2 ZV^{-1}Z' = Cov_\theta\big(Z\hat{\theta}\big) = Cov_\theta(\hat{z}),$$

де нерівність матриць означає одночасну нерівність всіх відповідних діагональних елементів і використано ортогональність матриць $ZV^{-1}T$ І $C - ZV^{-1}T$. Остання випливає з наведеної вище тотожності $Z = CT'$ та рівностей

$$ZV^{-1}T(C - ZV^{-1}T)' = ZV^{-1}(TC' - TT'V^{-1}Z') =$$
$$= ZV^{-1}(TC' - Z') = ZV^{-1}(TC' - (CT')') = 0. \blacksquare$$

12.2. Проста лінійна регресія

Розглянемо для прикладу випадок $k = 2$. Нехай

$$\theta = (a, b), \qquad t_{1j} = 1, \qquad t_{2j} = x_j, \qquad j = \overline{1,n}.$$

Модель регресії матиме вигляд

$$\xi_j = a + bx_j + \varepsilon_j, \qquad j = \overline{1,n}.$$

Матриці T, V задаються рівностями

$$T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix},$$

$$V = TT'n\begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{pmatrix}, \qquad \bar{x} = \sum_{j=1}^{n} x_j, \quad \overline{x^2} = \sum_{j=1}^{n} x_j^2.$$

Звідси обчислюємо обернену матрицю V^{-1} та оцінки МНК

$$V^{-1}T = \frac{1}{n\bar{\sigma}^2}\begin{pmatrix} \overline{x^2} - x_1\bar{x} & \overline{x^2} - x_2\bar{x} & & \overline{x^2}x_n\bar{x} \\ x_1 - \bar{x} & x_2 - \bar{x} & \cdots & x_n - \bar{x} \end{pmatrix},$$

$$\bar{\sigma}^2 = \overline{x^2} - (\bar{x})^2 = \frac{1}{n}\sum_{j=1}^{n}\big(x_j - \bar{x}\big)^2,$$

$$\hat{b} = \frac{1}{n\bar{\sigma}^2}\sum_{j=1}^{n}\xi_j\big(x_j - \bar{x}\big),$$

$$\widehat{\theta} = V^{-1}TX = V^{-1}T(T'\theta + \varepsilon) = \theta + V^{-1}T\varepsilon.$$

Тому внаслідок теореми про розклад суми квадратів відхилень

$$SS_v + SS_r = (X - T'\theta)^2 = \left(X - T'\widehat{\theta}\right)^2 + \left(T'\left(\widehat{\theta} - \theta\right)\right)^2 =$$

$$= \left(T'\theta + \varepsilon - T'(\theta + V^{-1}T\varepsilon)\right)^2 + (T'(\theta + V^{-1}T\varepsilon) - T'\theta)^2 =$$

$$= \left((I - T'V^{-1}T)\varepsilon\right)^2 + (T'V^{-1}T\varepsilon)^2 = \left((I - \Pi)\varepsilon\right)^2 + (\Pi\varepsilon)^2 =$$

$$= \varepsilon'(I - \Pi)'(I - \Pi)\varepsilon + \varepsilon'\Pi'\Pi\varepsilon = \varepsilon'(I - \Pi)\varepsilon + \varepsilon'\Pi\varepsilon.$$

Звідси

$$SS_v/\,\sigma^2 + SS_r/\,\sigma^2 = \varepsilon'(I - \Pi)\varepsilon/\,\sigma^2 + \varepsilon'\Pi\varepsilon/\,\sigma^2 =$$

$$= \zeta'U(I - \Pi)U'\zeta + \zeta'U\Pi U'\zeta =$$

$$= \zeta'\left(I - I_n(k)\right)\zeta + \zeta' I_n(k)\zeta = \sum_{i=k+1}^{n} \zeta_i^2 + \sum_{i=1}^{k} \zeta_i^2 = \chi_{n-k}^2 + \chi_k^2,$$

де величини в правій частині незалежні за теоремою про векторні перетворення незалежних величин ζ_i і мають за означенням відповідні хі-квадрат розподіли. ■

Теорема 12.4 (теорема Гаусса – Маркова). Нехай $rang\, T = k, \widehat{\theta}$ – оцінка методу найменших квадратів параметра θ, а вектор

$$z = Z\theta \in R^m, \qquad m \le k,$$

породжений лінійним перетворенням $Z: \in R^k \to \in R^m$. Тоді оцінка $z = Z\widehat{\theta}$ є лінійною незміщеною оцінкою для z, і має найменшу дисперсію (тобто найменші діагональні елементи коваріаційної матриці) у класі всіх лінійних за X незміщених оцінок вектора z.

Доведення. Нехай $\tilde{z} = CX$ є довільною лінійною незміщеною оцінкою для z. Тоді

$$M_\theta \tilde{z} = CM_\theta X = CT'\theta = z = Z\theta, \qquad \forall\theta \in R^k,$$

звідки виводимо тотожність

$$Z = CT'.$$

За теоремою про коваріаційну матрицю лінійного перетворення

$$Cov_\theta(\tilde{z}) = CCov_\theta(X)C' = CCov_\theta(\varepsilon)C' = \sigma^2 CC' =$$

(б) як інтеграл Лебега – Стілтьєса для вимірної функції $f_\xi(x)$.

Функція $f_\xi(x)$ називається *щільністю розподілу* випадкової величини ξ та функції розподілу F_ξ.

Означення 1.10. Випадкова величина ξ *має* рівномірний розподіл *на відрізку* $[a,b]$, *позначення* $\xi \simeq \mathrm{U}(a,b)$, якщо її щільність є сталою всередині цього відрізку, та дорівнює нулю поза ним.

Це означає, що ймовірність попадання величини в якусь множину всередині відрізку як інтеграл від щільності пропорційна довжині цієї множини і не залежить від її положення – тобто виконується умова рівноймовірності значень. З умови нормованості випливає, що щільність розподілу рівномірної на $[a,b]$ величини дорівнює

$$f_\xi(x) = \begin{cases} \dfrac{1}{b-a}, x \in [a,b], \\ 0, x \notin [a,b], \end{cases}$$

а функція розподілу має вигляд

$$F_\xi(x) = \begin{cases} 0, x < a \\ \dfrac{x-a}{b-a}, x \in [a,b], \\ 1, x > b. \end{cases}$$

Випадкова величина $\xi \simeq \mathrm{U}(a,b)$ має наступні числові характеристики: $M\xi = \frac{a+b}{2}, D\xi = \frac{(b-a)^2}{12}$.

Означення 1.11. Випадкова величина ξ має *стандартний нормальний розподіл*, позначення $\xi \simeq \mathrm{N}(0,1)$, якщо її щільність дорівнює

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \forall x \in R,$$

тобто функція розподілу має вигляд

$$F_\xi(x) = \Phi(x) \equiv \int_{-\infty}^{x} f_\xi(y)dy.$$

Випадкова величина $\xi \simeq \mathrm{N}(0,1)$ має наступні числові характеристики: $M\xi = 0, D\xi = 1$.

Означення 1.12. Випадкова величина ξ має *нормальний* розподіл із параметрами μ і $\sigma > 0$ (що позначається як $\xi \simeq \mathrm{N}(\mu, \sigma^2)$), якщо її можна зобразити у вигляді лінійного перетворення $\xi = \mu + \sigma\zeta$, де ζ – стандартна нормальна величина – тобто має стандартний нормальний розподіл. Нормальна щільність розподілу величини ξ має вигляд

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \forall x \in R.$$

Це рівняння можна взяти за інше означення величини ξ.

Випадкова величина $\xi \simeq \mathrm{N}(\mu, \sigma^2)$ має наступні числові характеристики: $M\xi = \mu, D\xi = \sigma^2$.

Означення 1.13. Випадкова величина ξ має *показниковий* розподіл *(або експоненційний* розподіл*)* із параметром λ>0, позначення $\xi \simeq \mathrm{Exp}(\lambda)$, якщо вона абсолютно неперервна, невід'ємна, а її щільність і функція розподілу

$$f(x) = \lambda\exp(-\lambda x),\; F(x) = 1 - \exp(-\lambda x), \forall x \geq 0.$$

Випадкова величина $\xi \simeq \mathrm{Exp}(\lambda)$ має наступні числові характеристики: $M\xi = \frac{1}{\lambda}, D\xi = \frac{1}{\lambda^2}$.

Означення 1.14. Випадкова величина ξ має *розподіл Парето*, якщо для деяких $\lambda > 0, \alpha > 0$ її функція розподілу дорівнює

$$P\{\xi < x\} = \begin{cases} 1 - (\lambda x)^{-\alpha}, x \geq \dfrac{1}{\lambda} \\ 0, x < \dfrac{1}{\lambda}. \end{cases}$$

Означення 1.15. Випадкова величина ξ має *розподіл Коші*, якщо її можна зобразити у вигляді $\xi = tg\,\varphi$, де випадковий кут φ є рівномірно розподіленим на відрізку $[-\pi/2, \pi/2]$.

Оскільки $\phi \in [-\pi/2, \pi/2]$, то

$$P\{\xi < x\} = P\{\varphi < arctgx\} = \frac{1}{2} + \frac{1}{\pi}arctgx,$$

і щільність ξ має вигляд

$$f_\xi(x) = \frac{1}{\pi}\frac{1}{1+x^2}.$$

Теорема 12.4 (про незміщеність і розподіл оцінки МНК). Нехай $rang\,T = k$ і має місце нормальна модель*:* $\varepsilon \simeq N_n(0, \sigma^2 I)$. Тоді*:*

(а) оцінка МНК $\widehat{\theta}$ є незміщеною для θ, причому $\widehat{\theta} \simeq N_n(0, \sigma^2 V^{-1})$,

(б) статистика $\widehat{\sigma}_{n-k}{}^2 = SS_v/(n-k)$ є незміщеною оцінкою для σ^2,

(в) статистики $SS_v/\,\sigma^2$ *та* $SS_r/\,\sigma^2$ незалежні й мають хі-квадрат розподіли з $n-k$ і k ступенями свободи відповідно.

Зауваження 12.1. Відношення нормованих статистик із пункту *(в)* має розподіл Фішера після нормування на кількості ступенів свободи і використовується для перевірки гіпотези про суттєвість залежності теоретичного середнього спостережень від параметра, що оцінюється. **Доведення.**

За теоремою про лінійні перетворення нормальних векторів статистика $\widehat{\theta} = V^{-1}TX$ має нормальний розподіл, тому *(а)* випливає з рівнянь

$$M_\theta\widehat{\theta} = MV^{-1}T(T'\theta + \varepsilon) = \theta + V^{-1}TM\varepsilon = 0,$$

$$Cov_\theta(\widehat{\theta}) = V^{-1}TCov(X)(V^{-1}T)' = \sigma^2 V^{-1}TT'V^{-1} = \sigma^2 V^{-1},$$

де використано теорему про коваріаційну матрицю лінійного перетворення.

Твердження *(б)* є наслідком *(в)*, оскільки

$$M_\theta\widehat{\sigma}_{n-k}{}^2 = M_\theta SS_v/(n-k) = \sigma^2 M_\theta(SS_v/\,\sigma^2)/(n-k) =$$

$$= \sigma^2 M_\theta(\chi^2_{n-k})/(n-k) = \sigma^2(n-k)/(n-k) = \sigma^2.$$

Для доведення *(в)* зауважимо, що $n \times n$ –матриця $\Pi \equiv T'V^{-1}T$ симетрична, ідемпотентна:

$$\Pi^2 = T'V^{-1}TT'V^{-1}T = T'V^{-1}VV^{-1}T = T'V^{-1}T = \Pi$$

та має ранг k. Тому існує ортонормована матриця U, що приводить її до діагональної матриці вигляду

$$U\Pi U' = I_n(k) = \left(\delta_{ij}1_{i\leq k}, i, j = \overline{1,n}\right).$$

Позначимо $\zeta = U\varepsilon/\sigma$. Тоді $\zeta \simeq N_n(0, I)$ – стандартний нормальний вектор за теоремою про лінійні перетворення нормальних векторів, оскільки $\varepsilon/\sigma \simeq N_n(0, I)$ – стандартний нормальний вектор за умовою. Далі, за теоремою про рівняння методу найменших квадратів

$$\left(X - T'(\theta + h)\right)^2 - (X - T'\theta)^2 = 2h'T(X - T'\theta) + o(h), \qquad h \to 0.$$

Якщо θ є точкою екстремуму, то ліва частина не змінює знак при зміні знака h, тому диференціал Гато є нульовим

$$d[(X - T'\theta)^2, h] = 2h'T(X - T'\theta) = 2h'(TX - V\theta) = 0, \qquad \forall h \in \mathrm{K}^n,$$

що і доводить (а).

(б) З неперервності та невід'ємності квадратичної функції випливає існування хоча б однієї точки локального мінімуму. Тому рівняння МНК завжди має принаймні один розв'язок. З (а) робимо висновок, що єдина оцінка МНК дорівнює $V^{-1}TX$.

Достатність є наслідком наступної теореми.

Теорема 12.3 (про розклад суми квадратів відхилень). Нехай θ – істинне значення параметра, а $\widehat{\theta}$ – довільний розв'язок рівняння МНК.

Тоді повна сума квадратів відхилень дорівнює сумі двох сум квадратів:

$$(X - T'\theta)^2 = \left(X - T'\widehat{\theta}\right)^2 + \left(T'\left(\widehat{\theta} - \theta\right)\right)^2,$$

і набуває найменшого значення в точці $\theta = \widehat{\theta}$.

Означення 12.2. Суми в зображенні повної суми квадратів відхилень

$$SS_v = \left(X - T'\widehat{\theta}\right)^2, \ SS_r = \left(T'\left(\widehat{\theta} - \theta\right)\right)^2$$

називаються відповідно *сумою квадратів відхилень від регресії* та *залишковою сумою квадратів.*

Доведення. Обчислимо

$$(X - T'\theta)^2 = \left(X - T'\theta + T'\left(\widehat{\theta} - \theta\right)\right)^2 =$$

$$= \left(X - T'\widehat{\theta}\right)^2 + \left(T'\left(\widehat{\theta} - \theta\right)\right)^2 + 2\left(T'\left(\widehat{\theta} - \theta\right)\right)\left(X - T'\widehat{\theta}\right) =$$

$$= \left(X - T'\widehat{\theta}\right)^2 + \left(T'\left(\widehat{\theta} - \theta\right)\right)^2 + 2\left(\widehat{\theta} - \theta\right)'\left(TX - V\widehat{\theta}\right) =$$

$$= \left(X - T'\widehat{\theta}\right)^2 + \left(T'\left(\widehat{\theta} - \theta\right)\right)^2,$$

оскільки θ є розв'язком рівняння МНК. ■

Теорема 1.1 (про розподіл Ерланга). Нехай величини $(\xi_n, n \ \geq \ 1)$ незалежні в сукупності, однаково розподілені і мають показниковий розподіл із параметром λ. Тоді сума $\zeta_n = \ \xi_1 + \ldots + \xi_n$має *розподіл Ерланга* порядку n зі щільністю:

$$f_n(x) = \lambda(\lambda x)^{n-1} \exp\left(-\lambda x\right)/(\mathrm{n}-1)!\,, x \geq 0.$$

Означення 1.16. Невід'ємна випадкова величина ζ має *гама-розподіл* із параметрами $\lambda, \alpha \ > \ 0$, позначення $\xi \simeq \Gamma(\lambda, \alpha)$, якщо її щільність дорівнює

$$f(x) = \lambda(\lambda x)^{\alpha-1} \exp\left(-\lambda x\right)/\Gamma(\alpha)\,, x \geq 0$$

Нехай $X_1, \ldots, X_n$ – незалежні в сукупності стандартні нормальні випадкові величини: $X_k \simeq N(0,1), k = 1, \ldots, n$.

Означення 1.17. Випадкова величина

$$\chi^2(n) = X_1^2 + \cdots + X_n^2$$

має *розподіл Пірсона або* χ^2*-розподіл* ("хі" - квадрат) з n ступенями свободи.

Означення 1.18. Випадкова величина $T(n)$ має $t-$розподіл, або *розподіл Стьюдента*, із n ступенями свободи, якщо її можна подати у вигляді

$$T(n) = \frac{\zeta}{\sqrt{\frac{1}{n}\chi^2(n)}},$$

де $\zeta \simeq N(0,1)$ – стандартна нормальна величина, а $\chi^2(n)$ – незалежна від неї величина із χ^2-розподілом та n ступенями свободи.

Означення 1.19. Випадкова величина F має *розподіл Фішера*, або ж *Снедекора – Фішера*, із (n, m) ступенями свободи, якщо її можна подати у вигляді

$$F = \frac{\frac{1}{n}\chi^2(n)}{\frac{1}{m}\chi^2(m)},$$

де $\chi^2(n)$, $\chi^2(m)$ – незалежні величини із χ^2 –розподілом та n, m ступенями свободи відповідно.

Означення 1.20. Випадковий вектор $\zeta = (\zeta_1, \ldots, \zeta_n)^T$ називається *n-вимірним стандартним нормальним вектором*, позначення $\xi \simeq N_n(0, I)$, якщо випадкові величини $\zeta_1, \ldots, \zeta_n$ є незалежними в сукупності стандартними нормальними величинами.

Зауваження 1.1. З теореми про критерій незалежності абсолютно неперервних величин та з означення стандартної нормальної величини випливає, що вектор ζ є стандартним нормальним тоді й тільки тоді, коли його сумісна щільність має вигляд

$$f_\zeta(x) = \prod_{k=1}^{n}(2\pi)^{-1/2}e^{-x_k^2/2} = (2\pi)^{-n/2}exp\left(-\sum_{k=1}^{n}\frac{x_k^2}{2}\right)=$$

$$= (2\pi)^{-n/2}\exp\left(-x^T x/2\right).$$

Означення 1.21. Випадковий вектор $\xi = (\xi_1, \dots, \xi_n)^T$ називається *n-вимірним нормальним вектором*, якщо його можна зобразити у вигляді лінійного невиродженого перетворення стандартного нормального вектора ζ, тобто якщо для деяких вектора $m \in R^n$ та невиродженої матриці A виконується рівність $\xi = m + A\zeta$.

Теорема 1.2 (про сумісну щільність нормального вектора). Випадковий вектор ξ є *n-вимірним нормальним вектором* тоді й тільки тоді, коли його сумісна щільність має вигляд

$$f_\zeta(x) = (2\pi)^{-n/2}|\mathrm{detV}|^{-1/2}\exp\left(-\frac{1}{2}(x-m)^T V^{-1}(x-m)\right), x \in R^n,$$

де $m \in R^n$, а матриця V є симетричною додатно визначеною. Якщо $\xi = m + A\zeta$, де $\zeta \simeq N_n(0, I)$, то $m = m, V = AA^T$.

Зауваження 1.2. Останнє позначення для випадку, коли $m = 0$ і $V = I$ – одинична матриця, відповідає за означенням стандартному нормальному вектору.

Теорема 1.3 (про інтерпретацію параметрів нормального розподілу). Нехай нормальний вектор $\zeta \simeq N_n(m, V)$. Тоді $M\xi = m, Cov(\xi) = V$.

Означення 12.1. Оцінкою методу найменших квадратів векторного параметра θ називається статистика

$$\widehat{\theta} = \arg\min_{\theta \in R^k}(X - T'\theta)^2.$$

Очевидно, що функція $\widehat{\theta}$ обчислюється за X і T, тобто є оцінкою. Існування оцінки МНК очевидне, оскільки квадратична функція неперервна та обмежена знизу.

Теорема 12.1 (про співвідношення між оцінкою МНК та ОМВ). Припустимо, що похибки ε_i незалежні у сукупності і однаково нормально розподілені:

$$\varepsilon \simeq N_n(0,\ \sigma^2 I).$$

Тоді оцінка МНК збігається з оцінкою максимальної вірогідності.

Доведення. За умовою вибірка

$$X \simeq T'\theta + N_n(0,\ \sigma^2 I) = N_n(T'\theta,\ \sigma^2 I)$$

є нормальним вектором, тому логарифмічна функція вірогідності дорівнює

$$\ln L(X, \theta) = -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}(X - T'\theta)^2.$$

Очевидно, що точка максимуму $L(X, \theta)$ за θ збігається з оцінкою МНК. ■

Теорема 12.2 (про рівняння методу найменших квадратів).

(а) Вектор $\widehat{\theta} = \widehat{\theta}(X) \in R^k$ збігається з оцінкою методу найменших квадратів тоді й тільки тоді, коли він є розв'язком рівняння МНК

$$V\widehat{\theta} = TX,$$

де

$$V \equiv TT'$$

– симетрична матриця розміру $k \times k$.

(б) Якщо $rang\ T = k$, то $det\ V = 0$ і єдина оцінка МНК має вигляд

$$\widehat{\theta} = V^{-1}TX.$$

Доведення. *Необхідність.*

(а) Обчислимо головну частину приросту квадратичної функції в околі точки θ (диференціал Гато):

РОЗДІЛ 12. Метод найменших квадратів для лінійної регресії

У регресійному аналізі вивчається проблема кількісного впливу одних змінних (наприклад, відсоткового складу різних домішок у сплаві) на числові значення інших змінних (наприклад, міцності сплаву). Одночасно враховується стохастичний характер величин, що спостерігаються.

12.1. Модель лінійної регресії

В теорії лінійної регресії розглядається модель лінійної залежності

$$\xi_j = \sum_{i=1}^{k} \theta_i \, t_{ij} + \varepsilon_j, \;\; j = \overline{1,n},$$

де $X = \left(\xi_j, \; j = \overline{1,n} \right)$ – вектор спостережень (вибірка), $\theta = \left(\theta_i, i = \overline{1,k} \right) \in R^k$ – невідомий вектор параметрів, що підлягає оцінці, його розмірність $k < n$, матриця $T = \left(t_{ij}, \; i = \overline{1,k}, \; j = \overline{1,n} \right)$ вважається відомою, вектор $\varepsilon = \left(\varepsilon_j, \; j = \overline{1,n} \right)$ містить випадкові похибки вимірювань.

Природне припущення полягає в тому, що випадкові похибки ε_i є незалежними у сукупності, однаково розподіленими, мають нульове середнє (відсутня систематична похибка) та дисперсію σ^2.

Усі вектори будемо розглядати як вектори-стовпчики. У векторній формі модель лінійної регресії має вигляд

$$X = T'\theta + \varepsilon,$$

де символ $'$ визначає транспонування. Вектор похибок задовольняє умови

$$M_\theta \varepsilon = 0, \;\; Cov_\theta \varepsilon = \sigma^2 I$$

з одиничною матрицею I.

Для оцінки невідомого параметра θ К. Гаусс запропонував і використав *метод найменших квадратів* (МНК).

РОЗДІЛ 2. Розподіл лінійних і квадратичних форм від нормально розподілених випадкових величин

Нехай $\xi_k, k = 1, \dots, n$ – незалежні стандартно нормально розподілені випадкові величини такі, що $\xi_k \simeq N(0,1)$ та $\xi = (\xi_1, \dots, \xi_n)^T$

Лема 2.1 (Фішера). Нехай A – ортогональне перетворення в R^n. Тоді компоненти вектора $\tilde{\xi} = A\xi$ – незалежні і стандартно нормально розподілені випадкові величини.

Доведення. Розглянемо випадковий вектор $\tilde{\xi} = A\xi$ такий, що

$$\begin{pmatrix} \tilde{\xi}_1 \\ \vdots \\ \tilde{\xi}_n \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix}.$$

Звідси $\tilde{\xi}_i = a_{i1}\xi_1 + \dots + a_{in}\xi_n, i = 1, \dots, n$. Очевидно, що $\xi \simeq N_n(0, E)$.

Із властивостей нормально розподілених векторів випливає, що $\tilde{\xi}$ – нормальний вектор, оскільки $M\tilde{\xi} = 0$ так як $\xi_i \simeq N(0,1) \Rightarrow a_{ki}\xi_i \simeq N(0, a_{ki}^2)$ та $\tilde{\xi}_i \simeq N(0, a_{k1}^2 + \dots + a_{kn}^2)$, а кореляційна матриця рівна

$$M\tilde{\xi}\tilde{\xi}^T = M(A\xi\xi^T A^T) = A(M\xi\xi^T)A^T = E, M\tilde{\xi}_i\tilde{\xi}_j = \begin{cases} 0, i \neq j; \\ 1, i = j. \end{cases}$$

Отже, $\tilde{\xi} \simeq N_n(0, E)$. ■

Лема 2.2. Нехай $Q_j = Q_j(x_1, \dots, x_n), j = \overline{1,k}$ – квадратичні форми рангу $n_j \geq 1$, $n_1 + \dots + n_k = n$ і $\sum_{j=1}^{k} Q_j = \sum_{j=1}^{n} x_j^2$. Тоді випадкові величини $Q_j(\xi) = Q_j(\xi_1, \dots, \xi_n)$ мають χ^2 –розподіл із n_j-ступенями вільностей і є взаємно незалежними. При цьому існує ортогональне перетворення B, що одночасно зводить всі квадратичні форми Q_j до суми квадратів, і якщо $y = Bx$, то $Q_j(x) = y_{N_{j-1}+1}^2 + \dots + y_{N_j}^2$, $N_j = n_1 + \dots + n_j, N_0 = 0$.

Доведення. Оскільки квадратичні форми Q_j мають ранг n_j, то їх можна зобразити у вигляді суми квадратів лінійних форм

$$Q_j(x) = \sum_{i=N_{j-1}+1}^{N} \delta_i \left(\sum_{s=1}^{n} b_{i_s} x_s\right)^2, \text{ де } \delta_i = \pm 1.$$

Враховуючи, що $n_1 + \dots + n_k = n$, одержимо

$$\sum_{j=1}^{k} Q_j = \sum_{j=1}^{n} x_j^2 = \sum_{i=1}^{n} \delta_i \left(\sum_{s=1}^{n} b_{i_s} x_s\right)^2 = x^T B^T I B x,$$

де I – діагональна матриця з елементами δ_i. Із рівності двох симетричних квадратичних форм випливає рівність їх коефіцієнтів. Отже, $B^T I B = E \Rightarrow (B^T)^{-1} B^{-1} = I$.

Матриця B – невироджена, тому I – невід'ємно визначена матриця і $I = E$, $\delta_i = B^T B = E$, а тому перетворення B є ортогональним. На основі леми 3.1 вектор $\eta = B\xi$ – незалежний стандартно нормально розподілений, тобто $\eta \simeq N_n(0, E)$. Так як, $Q_j(x) = y_{N_{j-1}+1}^2 + \dots + y_{N_j}^2$, то $Q_j(\xi) = \eta_{N_{j-1}+1}^2 + \dots + \eta_{N_j}^2$ і ортогональне перетворення B одночасно приводить квадратичні форми Q_j до суми квадратів, і величини $Q_j(\xi)$ – незалежні і мають розподіл χ^2 з n_j-ступенями вільностей. ■

Теорема 2.1. Нехай та $\xi_1, \dots, \xi_n$ – незалежні стандартно нормально розподілені випадкові величини, $\xi_k \simeq N(0,1), k = \overline{1,n}$. Позначимо $\bar{\xi} = \frac{1}{n}\sum_{k=1}^{n} \xi_k$, $s^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(\xi_k - \bar{\xi}\right)^2$. Тоді

1) випадкова величина $\bar{\xi}$ має стандартний нормальний розподіл;

2) $(n-1)s^2$ має χ^2 –розподіл із $(n-1)$-ступенями вільностей;

3) величини $\bar{\xi}$ і s^2 – незалежні.

Доведення. Розглянемо вираз

$$\frac{(n-1)s^2}{n} = \frac{1}{n}\sum_{k=1}^{n}\left(\xi_k^2 - \bar{\xi}^2\right)^2,$$

$$(n-1)s^2 + n\bar{\xi}^2 = \sum_{k=1}^{n} \xi_k^2. \qquad (2.1)$$

Із умов теореми випливає, що $\bar{\xi} \simeq N\left(0, \frac{1}{n}\right)$, оскільки $\xi_k \simeq N(0,1), k = \overline{1,n}$ і $M\bar{\xi} = 0$, $D\bar{\xi} = \frac{1}{n}$.

Далі, $Q_1 = n\bar{\xi}^2$ – квадратична форма рангу 1, розподіленою за χ^2 –розподілом із 1-ступенем вільностей. Оскільки, $Q_2 = (n-1)s^2 = \sum_{k=1}^{n}\left(\xi_k - \bar{\xi}\right)^2$ є квадратичною формою рангу від $\xi_1, \dots, \xi_n$, а за формулою (2.1) $Q_1 + Q_2$ є сумою квадратів

$$a + b \le \alpha + \beta.$$

Отже, для будь-яких заданих $\alpha, \beta \in (0,1)$ існує послідовний критерій (з граничними значеннями $c_j = \ln C_j$), що має не більшу за $\alpha + \beta$ суму ймовірностей похибок першого та другого роду.

Для наближеного обчислення середньої кількості необхідних спостережень у послідовному критерії використовується *тотожність Вальда*:

$$M_\theta S_\tau = m_\theta M_{\theta\tau},\ m_\theta = M_\theta S_1 = M_\theta h(\xi_1).$$

Остання величина не перевищує $C_0 = \exp(c_0) < 1$, на події A_{0k}, і не менша за $C_1 = \exp(c_1) > 1$ на події A_{1k}. Позначимо B_{ik} борелєві підмножини $\mathbb{R}^k$, що отримуються з A_{ik} заміною величин ξ_j на j-ті координати y_j вектора $x_k = (y_1, \dots y_k) \in \mathbb{R}^k$. Тоді $A_{ik} = \{X_k \in B_{ik}\}$. Із включення $x_k \in B_{1k}$ випливає, що $L_k(X_k, \theta_1)/L_k(X_k, \theta_0) = \exp\left(\sum_{j=1}^k h(y_j)\right) = \exp(c_1) = C_1$. Аналогічно, при $x_k \in B_{0k}$ маємо $L_k(X_k, \theta_1)/L_k(X_k, \theta_0) = \exp\left(\sum_{j=1}^k h(y_j)\right) = \exp(c_0) = C_0$. Звідси

$$P(A_{1k}|H_0) = \int_{B_{1k}} L_k(X_k, \theta_0)\mu_k(dx_k) \le \int_{B_{1k}} \frac{L_k(X_k, \theta_1)}{C_1}\mu_k(dx_k) = \frac{P(A_{1k}|H_1)}{C_1},$$

$$P(A_{0k}|H_1) = \int_{B_{0k}} L_k(X_k, \theta_1)\mu_k(dx_k) \le \int_{B_{0k}} C_0 L_k(X_k, \theta_0)\mu_k(dx_k) = C_0\, P(A_{1k}|H_0).$$

Нехай α, β – ймовірності похибок першого та другого роду для побудованого вище послідовного критерію. Тоді

$$\alpha = P\left(\bigcup_{k=1}^{n} A_{1k}|H_0\right) = \sum_{k=1}^{n} P(A_{1k}|H_0) \le \sum_{k=1}^{n} P(A_{1k}|H_1)/C_1 =$$

$$= \left(1 - \sum_{k=1}^{n} P(A_{0k}|H_1) - P(A_n|H_1)\right)/C_1 = (1-\beta)\;/\;C_1,$$

$$\beta = \sum_{k=1}^{n} P(A_{0k}|H_1) - P(A_n|H_1)) \le$$

$$\le \sum_{k=1}^{n} C_0 P(A_{1k}|H_0) - C_0 P(A_n|H_0)) = (1-\alpha)\;/\;C_0$$

Отже, критичні значення $c_i = \ln C_i$ та похибки в послідовному критерії відношення вірогідностей пов'язані системою нерівностей

$$\alpha \le (1-\beta)/C_1\,, \beta \le (1-\alpha)\;/\;C_0$$

Розглянемо послідовний критерій з граничними значеннями

$$C_1' = (1-\beta)/\alpha\,,\;\; C_0' = \beta/(1-\alpha)$$

для яких попередні нерівності перетворюються на рівності. Нехай *a; b* – ймовірності похибок для цього критерію. Тоді з нерівностей для ймовірностей похибок отримуємо

$$\frac{a}{(1-b)} \le \frac{1}{C_1'} = \frac{\alpha}{(1-\beta)}, \frac{b}{(1-a)} \le \frac{1}{C_0'} = \beta/(1-\alpha),$$

звідки множенням на знаменники та додаванням виводимо нерівність

компонент $\xi_1, \dots, \xi_n$, то внаслідок леми 2.2 Q_2 буде квадратичною формою рангу $(n-1)$, тому величина $(n-1)s^2$ буде мати розподіл χ^2 із $(n-1)$-ступенями вільностей. Квадратичні форми Q_j – незалежні, форми Q_1 і Q_2 – незалежні, а тому величини $\bar{\xi}$ і s^2 – незалежні. ■

Нехай $X_1, \dots, X_n$ – вибірка із нормально розподілених випадкових величин, $X_k \simeq N(a, \sigma^2), k = \overline{1,n}$. Позначимо $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$, $s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$. Тоді очевидно, що $M\bar{X} = a, D\bar{X} = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$. Тоді випадкова величина $\eta = \sqrt{n}\frac{\bar{X}-a}{\sigma} \simeq N(0,1)$. Величина

$$\eta = \frac{1}{\sigma}\left(\frac{1}{n}\sum_{k=1}^{n} X_k - a\right)\sqrt{n} = \frac{1}{\sigma n}\sum_{k=1}^{n}(X_k - a)\sqrt{n} =$$

$$= \frac{1}{n}\sum_{k=1}^{n}\frac{(X_k - a)}{\sigma}\sqrt{n} \simeq N(0,1).$$

Далі,

$$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{k=1}^{n}(X_k - \bar{X})^2}{\sigma^2} = \sum_{k=1}^{n}\left(\frac{X_k - \bar{X}}{\sigma}\right)^2 =$$

$$= \sum_{k=1}^{n}\left(\frac{X_k - a}{\sigma} - \frac{\bar{X} - a}{\sigma}\right)^2 = \sum_{k=1}^{n}\left(\xi_k - \bar{\xi}\right)^2.$$

Таким чином, за теоремою 2.1 маємо:

1) $\bar{X}$, s^2 – незалежні;

2) величина $\frac{(n-1)s^2}{\sigma^2}$ має χ^2 –розподіл із $(n-1)$-ступенями вільностей;

3) η і $\frac{(n-1)s^2}{\sigma^2}$ – незалежні випадкові величини.

Із попередніх міркувань випливає наступна теорема.

Теорема 2.2. Якщо $\xi_1, \dots, \xi_n$ – незалежні і нормально розподілені випадкові величини, $\xi_k \simeq N(a, \sigma^2), k = \overline{1,n}$, то величина

$$\tau = \frac{\sqrt{n}\left(\bar{\xi} - a\right)}{s} = \frac{\sqrt{n}\left(\bar{\xi} - a\right)}{\sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(\xi_k - \bar{\xi}\right)^2}}$$

має розподіл Стьюдента з $(n-1)$-ступенями вільностей.

(0) якщо $\{S_1 \epsilon (c_{01}, c_{11}), \dots, S_{k-1} \epsilon (c_{0k-1}, c_{1k-1}, S_k \leq c_{0k})\}$ для деякого $k \leq n$ то спостереження припиняються в такий момент *k;* і приймається нульова гіпотеза H_0;

(1) якщо $\{S_1 \epsilon (c_{01}, c_{11}), \dots, S_{k-1} \epsilon (c_{0k-1}, c_{1k-1}, S_k \leq c_{0k})\}$ для для деякого $k \leq n$ то спостереження припиняються в такий момент *k;* і приймається альтернативна гіпотеза H_1;

(n) інакше приймається рішення на користь H_0.

Вибір знаків сталих обгрунтовується критерієм Колмогорова посиленого закону великих чисел, внаслідок якого $S_k \overset{P_1}{\to} -\infty, k \to \infty$за нульової гіпотези та $S_k \overset{P_1}{\to} \infty, k \to \infty$ за альтернативи, відповідно до знаку математичного сподівання одного доданку. Звідси випливає також, що ймовірність події в (n) прямує до нуля при $n \to \infty$.

На відміну від критерію відношення вірогідностей послідовний критерій спирається на випадкову кількість спостережень

$$\tau = \min\Big(n, \inf\big(k \geq 1 \colon S_k \notin (c_{0k}; c_{1k})\big)\Big).$$

Якщо обрати $c_{0k} = -\infty, k \leq n$, $c_{1k} = \infty, k < n$ i $c_{1n} = \ln l_\alpha$, то отримаємо критерій відношення вірогідностей як частковий випадок узагальненого критерію. Тому мінімальна середня кількість випробувань у всьо-му класі послідовних критеріїв із заданими ймовірностями похибок α, β першого та другого роду не більша за відповідну кількість для простого критерію відношення вірогідностей: $\inf M\tau \leq n$. Отже, використання послідовного критерію може призвести до зменшення кількості спостережень при збереженні якості критерію.

Для уточнення властивостей послідовного критерію припустимо, що сталі $c_{0k} = c_0$, $c_{1k} = c_1$, не залежать від *k*. Позначимо A_{0k} подію в умові (0) критерію, A_{1k} – в умові (1) та A_n– у (n). За означенням всі ці події попарно несумісні та утворюють повну групу подій. Нехай випадковий вектор $X_k = (\xi_1, \dots \xi_k)$ містить перші *k* спостережень. Його функція вірогідності дорівнює $L_k(X_k, \theta) = \prod_{j=1}^{k} f\big(\xi_j, \theta\big)$, а відношення вірогідностей має вигляд

$$L_k(X_k, \theta_1)/L_k(X_k, \theta_0) = \exp\left(\sum_{j=1}^{k} h(\xi_j)\right) = \exp(S_k).$$

Обчислимо його потужність

$$P_{\theta_1}(X \epsilon W) = F_1(W) = F_1(W \setminus W^*) + F_1(W \cap W^*) = F_1(W^*) + F_1(W \setminus W^*) -$$

$$F_1(W^* \setminus W) = \int_{W \setminus W^*} l_{01}(x) dF_0(x) - \int_{W^* \setminus W} l_{01}(x) dF_0(x) \le$$

$$F_1(W^*) + l_\alpha F_0(W \setminus W^*) - l_\alpha F_0(W^* \setminus W) \le F_1(W^*) + l_\alpha F_0(W) - l_\alpha F_0(W^*) \le$$

$$\le F_1(W^*) + l_\alpha - l_\alpha = F_1(W^*),$$

де справедливість передостанньої нерівності є наслідком означення області W^*; згідно з яким $l_{01}(x) \le l_\alpha$ пр$x \notin W^*$; та $l_{01}(x) \ge l_\alpha$ при $x \in W^*$; а остання нерівність випливає з вибору l_α; оскільки вірогідний рівень W^* дорівнює $F_0(W^*) = \alpha$. ■

11.2. Поняття про послідовний аналіз

Для ілюстрації розширення правил формування статистичних висновків, що дозволило б приймати певні рішення після отримання кожного чергового спостереження, розглянемо задачу перевірки простих гіпотез $\text{H}_i: \theta = \theta_i, i = 0{,}1$ для кратної вибірки $X = \xi_1, \dots \xi_n$. Дану ідею запропонував та розвинув А. Вальд у своїх працях, що створили основу теорії послідовного статистичного аналізу. Позначимо через $f(y, \theta)$ щільність одного спостереження. За теоремою про функцію вірогідності кратної вибірки логарифм відношення вірогідностей дорівнює

$$\ln l_{01}(x) = \ln \frac{L_n(X,\theta_1)}{L_n(X,\theta_0)} = \sum_{j=1}^{n} h(\xi_j) = S_n, h(y) = \ln \frac{f(y,\theta_1)}{f(y,\theta_0)},$$

де доданки $h(\xi_j)$незалежні в сукупності та однаково розподілені.

У теоремі про властивості інформації за Кульбаком доведено, що за умов консистентності ОМВ $M_{\theta_0} h(\xi_1) < 0$ *:* Аналогічно можна довести, що

$$M_{\theta_0} h(\xi_1) > 0.$$

Найбільш потужний критерій відношення вірогідностей має критичну область вигляду $\{l_{01}(x) \ge l_\alpha\} = \{Sn \ge \ln l_\alpha\}$.

Нехай для кожного $k = \overline{1,n}$ задано сталі $c_{0k} < 0 < c_{1k}$. Позначимо $S_k = \sum_{j=1}^{k} h(\xi_j)$. Розглянемо *узагальнений послідовний критерій відношення вірогідностей*:

РОЗДІЛ 3. Моделювання випадкових величин

Моделювання – це один із найбільш расповсюджених способів вивчення процесів і явищ. Під моделюванням розуміється процес побудови, вивчення і застосування моделей.

При моделюванні випадкової величини виникає питання про *базис* (тобто сукупність випадкових величин із заданими характеритиками) для одержання випадкових величин на будь-якій наперед заданій множині можливих значень і при будь-якому наперед заданому законі розподілу. В якості базису візьмемо рівномірно розподілену випадкову величину на проміжку *(0,1).*

При моделюванні випадкових величин неможливо відтворити експеримент з випадковими величинами. Тому замість них використовують *псевдовипадкові числа* – такі числа, які за своїми властивостями до випадкових величин, але можуть бути отримані за допомогою деякого обчислювального алгоритму.

3.1. Моделювання дискретних випадкових величин

Розглянемо дискретну випадкову величину ξ з розподілом, наведеним в таблиці 1.1 .

Таблиця 3.1 – Розподіл випадкової величини ξ

x_1	x_2	...	x_n
p_1	p_2	...	p_n

де $p_i = P\{\xi = x_i\}$. Для того, щоб обчислити значення цієї величини розділимо інтервал $0 \le y < 1$ на інтервали Δ_i, такі, що довжина Δ_i рівна p_i.

Теорема 3.1. Випадкова величина ξ, яка визначається формулою

$$\xi = x_i, \text{ коли } \gamma \in \Delta_i \tag{3.1}$$

має розподіл імовірностей, наведений в таблиці 3.1.

Теорему 3.1 можна узагальнити на випадкову величину, яка може приймати нескінченну послідовність значень $x_1, x_2, \dots, x_n, \dots$ і має розподіл, наведений в таблиці 3.2.

Таблиця 3.2 – Розподіл випадкової величини, яка приймає нескінченну послідовність значень

x_1	x_2	$\cdots$	x_n	$\cdots$
p_1	p_2	$\cdots$	p_n	$\cdots$

В цьому випадку числа x_n і p_n задається формулами, і обчислення їх при кожному розрахунку випадкової величини виду (3.1) може виявитися досить трудомістким. Тоді потрібно вибрати число n_0 так, щоб сума імовірностей $p_1 + \cdots + p_{n_0}$ була достатньо близькою до 1, і значення $x_1, \dots, x_{n_0}$ і $p_1 + \cdots + p_{n_0}$ вибрати завчасно. Обчислювати x_i і p_i за формулами прийдеться тільки при $i > n_0$, а це буде достатньо рідко.

3.2. Моделювання випадкових подій

Моделювання випадкових подій зводиться до моделювання дискретних випадкових величин. Розглянемо чотири задачі, в кожній з яких потрібно змоделювати послідовність однакових незалежних випровувань.

М1. В кожному із випробувань може наступити або не наступити деяка подія A, імовірність настання якої $P(A) = p$ задана.

Розглянемо випадкову величину ξ, яка називається *індикатором* події A, яка рівна 1 при настанні A і 0 при настанні протилежної події $\bar{A}$. Розподіл ξ задається таблицею 3.3:

За умови абсолютної неперервності $F_1 \ll F_0$існує вимірна щільність міри $l_{01}(x)$ така, що

$$F_1(B) = \int_B l_{01}(x) dF_0(x)$$

для всіх вимірних множин $B \in \Sigma$ вибіркового простору S: За означенням функції вірогідності міри $F_i(B)$мають щільності $L(x, \theta_i)$ відносно фіксованої міри λ на Σ: Тому за теоремою про заміну змінної статистика $l_{01}(x)$ збігається з емпіричним відношенням вірогідностей

$$l_{01}(x) = \frac{L(X,\theta_1)}{L(X,\theta_0)}.$$

Теорема 11.1 (лема Неймана – Пірсона). Нехай нульова гіпотеза та альтернатива є простими гіпотезами і $F_1 \ll F_0$. Якщо для даного вірогідного рівня $\alpha \epsilon (0,1)$ існує стала $l_\alpha > 0$ така, що

$$P_{\theta_0}(l_{01}(X) \geq l_\alpha) = \alpha$$

то критерій відношення вірогідностей $(l_{01}(X), [l_\alpha, \infty))$ з критичною областю

$$W^* = \{x \in S: l_{01}(x) \geq l_\alpha\}$$

є найбільш потужним рівня α.

Зауваження 11.2. Функція

$$P_{\theta_0}(l_{01}(X) \geq l) = F_0(\{x \epsilon S: l_{01}(x \geq l)\})$$

не зростає за l і набуває значень із відрізка [0,1] Тому умова теореми щодо існування l_α виконується, якщо випадкова величина $l_{01}(X)$ є абсолютно неперервною. Якщо ж для даного рівня αумова існування точного критичного рівня не виконана, то можна скористатись рандомізованим критерієм, як це описано вище в розділі про рангові критерії.

Доведення. Розглянемо довільний критерій із критичною областю W рівня α; тобто

$$P_{\theta_0}(X \epsilon W) = F_0(W) \leq \alpha.$$

РОЗДІЛ 11. Найбільш потужні критерії, лема Неймана-Пірсона

Розглянемо задачу перевірки статистичної гіпотези $H_0: \theta \epsilon \Theta_0$ проти альтернативи $H_1: \theta \epsilon \Theta_1$ на підставі вибірки X зі значеннями у вибірковому просторі $(S, \Sigma, \mathcal{L})$ Як вже відомо, кожен критерій перевірки гіпотези однозначно задається вибірковою критичною областю $W \epsilon \Sigma$: гіпотеза H_0відкидається, якщо $X \epsilon W$ і не відкидається (приймається), якщо $X \notin W$. Вибіркова критична область визначається через статистичний критерій як така підмножина вибіркового простору:

$$W = \{x \in S: \hat{\kappa}(x) \epsilon D_1\}$$

де $\hat{\kappa}(x) = \hat{\kappa}$ – статистика критерію, а D_1 – її критична область. Нагадаємо, що вірогідний рівень критерію визначається найбільшою з імовір-ностей похибок першого роду $P_\theta(X \epsilon W), \theta \epsilon W$.

Означення 11.1. Критерій із критичною областю W^* є *найбільш потужним* рівня α , якщо його вірогідний рівень дорівнює α*;* причому довільний критерій із критичною областю W рівня α має не більшу потужність, ніж W^*

$$P_\theta(X \epsilon W^*) \geq P_\theta(X \epsilon W)\ \forall \theta \epsilon \Theta_1.$$

11.1. Критерій відношення вірогідностей

Задача відшукання найбільш потужного критерію не завжди має розв'зок, оскільки часто неможливо максимізувати значення потужності одночасно при декількох значеннях параметра. Однак у випадку простих гіпотез та альтернатив найбільш потужний критерій існує.

Припустимо, що нульова гіпотеза та її альтернатива є простими гіпотезами:

$$H_i: \theta = \theta_i, i = 0, 1.$$

Позначимо через $F_i(B) = P_{\theta i}(X \epsilon B), B \epsilon \Sigma$ відповідні розподіли вибірок. Ці розподіли відомі повністю за означенням простої гіпотези.

Таблиця 3.3 – Індикатор події A

0	1
p	$p-1$

Згідно теореми 3.1 для здійснення кожного випробування потрібно знайти випадкове число γ і перевірити нерівність $\gamma < p$. Якщо вона виконується, то подія A в цьому випадку відбулася, а якщо $\gamma \leq p$, то не виконується.

М2. З випробуванням пов'язана повна група попарно незалежних подій $A_1, \dots, A_n$ і задані імовірності $P(A_i) = p_i$.

Для моделювання таких випробувань розглянемо випадкову величину ξ – номер події, яка настала. Очевидно, розподіл ξ виражається таблицею 3.4

Таблиця 3.4 – Розподіл ξ

1	2	...	n
p_1	p_2	...	p_n

Для здійснення кожного випробування вибирають випадкове число γ і за теоремою 3.1 знаходять значення ξ. Якщо $\xi = i$, то відбулася подія A_i.

М3. З випробуванням пов'язані два незалежні спільні події A і B, імовірності яких задані $P(A) = p_A,\ P(B) = p_B$.

В зв'язку з незалежністю подій A і B можна послідовно моделювати їх настання, в кожному випробуванні: спочатку за числом γ_1 методом М1 визначити, чи наступила подія A, а також точно так само за числом γ_2, визначити чи наступила подія B. Однак часто більш зручний другий спосіб. Розглянемо повну групу попарно незалежних подій, яка складається з чотирьох подій:

$$A_1 = AB,\ A_2 = A\bar{B}, A_3 = \bar{A}B, A_4 = \bar{A}\bar{B} \qquad (3.2)$$

імовірність цих подій можна обчислити :

$$p_1 = p_A p_B, \qquad p_2 = p_A(1 - p_B),$$
$$p_3 = p_B(1 - p_A),\; p_4 = (1 - p_A(1 - p_B).$$

Відповідно, метод М2 дозволяє, дозволяє використовуючи одне випадкове число γ, визначити, яка з цих чотирьох подій наступить в моделюючому випробуванні.

М4. З випробуванням пов'язані дві незалежні події A і B, і задані імовірності $P(A) = p_A$, $P(B) = p_B$, $P(AB) = p_{AB}$.

В цьому випадку потрібно розглянути повну групу подій (3.2), тільки імовірності цих подій обчислюються інакше:

$$p_1 = p_{AB}, \qquad p_2 = p_A - p_B,$$
$$p_3 = p_B - p_A,\; p_4 = 1 - p_A - p_B + p_{AB}.$$

В цьому випадку можна здійснити послідовне моделювання подій A і B, використовуючи два випадкові числа γ_1 і γ_2. Спочатку за числом γ_1 (методом М1) визначаємо, чи настала подія A. Якщо, A настала, то знаючи умовну імовірність $P(B/A) = p_{AB} \,/\, p_A$, можна за допомогою числа γ_2 визначити, чи настала подія B: умовою настання події B служить виконання нерівності $\gamma_2 < P(B/A)$. Якщо подія A не настала, то ймовірність настання B знаходять за допомогою умовної імовірності $P(B \,/\, \bar{A})$, яка рівна

$$P(B/\bar{A}) = (p_B - p_{AB})/(1 - p_A).$$

3.3. Моделювання неперервних випадкових величин

Нехай випадкова величина ξ визначена в інтервалі $a < x < b$ і має щільність $p(x) > 0$ при $a < x < b$. Позначимо через $F(x)$ функцію розподілу ξ, яка при $a < x < b$ рівна

$$F(x) = \int_a^x p(u)du.$$

Випадок $a = -\infty$ і $b = \infty$ не виключається.

Теорема 3.2. Випадкова величина ξ, яка задовольняє рівності

$$F(\xi) = \gamma \tag{3.3}$$

Гіпотеза про відношення дисперсій при відомих середніх. Одночасно спостерігаються незалежні кратні нормальні вибірки $X = (\xi_1, \ldots, \xi_n)$ та $Y = (\eta_1, \ldots, \eta_m$ з розподілами $\xi_1 \simeq N(\mu_1, \sigma_1^2), \eta_1 \simeq N(\mu_2, \sigma_2^2)$ з відомими середніми μ_k та невідомими дисперсіями σ_k^2 . Нульова гіпотеза формулюється як $H_0 : \sigma_2^2 = \rho\sigma_1^2$, де відношення ρ відоме (зокрема ρ = 1). Тоді за гіпотези H_0 статистика

$$\hat{\phi}_{n,m} = \frac{\hat{\sigma}_n^2/\sigma_1^2}{\hat{\sigma}_m^2/\sigma_2^2} = \rho\frac{\hat{\sigma}_n^2}{\hat{\sigma}_m^2}$$

має розподіл Фішера з *n, m* ступенями свободи. Відповідно критерій $(\hat{\phi}_{n,m}, [0, w_{\alpha/2}/\rho] \cup [w_{1-\alpha/2/\rho}, \infty))$ має рівень *α*, де границі w_α критичної області обрані з умови $P(\hat{\phi}_{n,m} < w_\alpha) = \alpha$.

Гіпотеза про відношення дисперсій при невідомих середніх. При невідомих середніх слід обрати нормовані вибіркові дисперсії. У цьому випадку статистика

$$\hat{\phi}_{n-1,m-1} = \frac{\hat{s}_n^2/\sigma_1^2}{\hat{s}_m^2/\sigma_2^2} = \rho\,\frac{\hat{s}_n^2}{\hat{s}_m^2}$$

має розподіл Фішера з *n − 1, m − 1* ступенями свободи. Тому критерій вигляду $(\hat{\phi}_{n-1,m-1}, [0, w_{\alpha/2}/\rho] \cup [w_{1-\alpha/2/\rho}, \infty))$ має рівень α. Іноді на практиці для перевірки гіпотези про рівність середніх одночасно використовують критерій про відношення дисперсій (для обґрунтування припущення про рівність дисперсій), а потім вже критерій про різницю середніх, що оснований на такому припущенні.

має за нульової гіпотези стандартний нормальний розподіл, а його критична область відповідає великим її значенням. Критерій ($|\hat{\zeta}_{nm}|, [x_\alpha, \infty)$) для перевірки гіпотези про різницю середніх має вірогідний рівень α та є консистентним.

Гіпотеза про різницю середніх при невідомій дисперсії. Нехай, спостерігаються незалежні кратні нормальні вибірки $X = (\xi_1, \ldots, \xi_n)$ та $Y = (\eta_1, \ldots, \eta_m)$ з розподілами $\xi_1 \simeq N(\mu_1, \sigma_1^2), \eta_1 \simeq N(\mu_2, \sigma_2^2)$ з невідомими середніми μ_kта однаковими невідомими дисперсіями σ^2 .

Нульова гіпотеза має вигляд $H_0 : \mu_2 - \mu_1 = \Delta$, де величина Δвідома, зокрема, можливо, $\Delta = 0$.

Нехай $\hat{\mu}_m, \hat{\mu}_n$ – відповідні вибіркові середні для X та Y , а, $\hat{s}_n^2, \hat{s}_m^2$ – нормовані вибіркові дисперсії. Тоді з незалежності вибірок X та Yі з теорем про векторні перетворення незалежних величин та про вибіркові середнє та дисперсію нормальної вибірки випливає, що величини $\hat{\mu}_n, \hat{\mu}_m, \hat{s}_n^2, \hat{s}_m^2$ незалежні у сукупності. Тому випадкова величина

$$\sqrt{\frac{nm}{n+m}}\ (\hat{\mu}_m - \hat{\mu}_n - \Delta)/\sigma,$$

Що має нульове середнє та одиничну дисперсію, є нормально розподіленою і до того ж не залежить від суми

$$((n-1)\hat{s}_n^2 + (m-1)\hat{s}_m^2)/\sigma^2,$$

яка має *хі*-квадрат розподіл із (*n*+*m*−*2*)-ступенями свободи за означенням. Отже, за нульової гіпотези *H₀* внаслідок означення розподілу Стьюдента статистика, що утворена відношенням:

$$\hat{\tau}_{n+m-2} = \sqrt{\frac{nm}{n+m}} \frac{\hat{\mu}_m - \hat{\mu}_n - \Delta}{\sqrt{((n-1)\hat{s}_n^2 + (m-1)\hat{s}_m^2)/(n+m-2)}}$$

має розподіл Стьюдента з ($n + m - 2$)-ступенями свободи. Критерій Стьюдента ($|\hat{\tau}_{n+m-2}|, [y_{n+m-2,\alpha}, \infty)$) для перевірки гіпотези про різницю середніх має вірогідний рівень α. Критерій є консистентним.

і має щільність розподілу $p(x)$.

Доведення*:* Оскільки функція $F(x)$ строго зростає на інтервалі (a, b) від $F(a) = 0$ до $F(b) = 1$, то рівняння (3.3) має єдиний корінь при кожному γ. При цьому рівні імовірності

$$P(x < \xi < x + dx) = P(F(x) < \gamma < F(x + dx)).$$

Оскільки, випадкова величина γ рівномірно розподілена в інтервалі (a, b), тоді

$$P(x < \xi < x + dx) = F(x + dx) - F(x) = p(x)dx,$$

що і потрібно було довести. ■

В тих випадках, коли рівняння (3.3) аналітично розв'язується відносно ξ, отримується явна формула $\xi = G(\gamma)$ для вибору випадкової величини ξ, де $G(\gamma)$ – обернена функція по відношенню до $y = F(x)$. В інших випадках можна рівняння (3.3) розв'язувати чисельно. Якщо обєм накопичувача дозволяє, то зручно скласти таблицю функції $G(y)$, $0 < y < 1$, і за нею знаходити значення ξ. Іноді зручно використовувати таблицю функції $F(x)$, $a < x < b$, і знаходити значення ξ оберненої інтерполяції.

3.4. Метод обернених функцій

Теореми 3.1 і 3.2 представлють частинні випадки загального методу, який називають *методом обернених функцій*.

Розглянемо довільну випадкову величину ξ з функцією розподілу

$$F(x) = P\{\xi < x\}.$$

Обернену по відношенню до $F(x)$ функцію $G(y)$ визначають наступним чином. Спочатку, доповнимо графік функції $y = F(x)$ вертикальними відрізками в точках розриву до неперервної лінії $y = F_0(x)$; функція $y == F_0(x)$, взагалі кажучи, неоднозначна. Цю лінію можна записати рівнянням вигляду $x = G_0(y)$, де функція $G_0(y)$ знову не обов'язково повинна бути однозначною: інтервалам сталості $F_0(x)$ відповідають вертикальні відрізки $G_0(y)$ і навпаки. Покладемо $G(y) = G_0(y)$ в точках неперервності і

$$G(y) = G(y + 0)$$

в точках розриву.

Побудована таким чином однозначна функція $G(y)$ не спадає при $0 < y < 1$ і неперервна справа в усіх точках. Функції $F(x)$ і $G(y)$ зв'язані наступною властивістю: $G(y) < x$ тоді і тільки тоді, коли $y < F(x)$ означає, що точка $(G(y), y)$ розташована на лінії $y = F_0(x)$ одночасно і ліві і нижні точки $(x, F(x))$.

Теорема 3.3. Нехай функція розподілу F має обернену функцію F^{-1}. Тоді функцією розподілу випадкової величини $\eta = F^{-1}(\alpha)$ є F.

Доведення: Знайдемо функцію розподілу η:

$$F_\eta(x) = P(\eta < x) = P(F^{-1}(\alpha) < x) = P\big(\alpha < F(x)\big) = F(x). \blacksquare$$

Зауваження 3.1. 1) Метод оберненої функції ефективний, якщо вдається отримати просту формулу для F^{-1}. А це не завжди можливо. 2) Умова оборотності функції F обмежена. Не кожна функція розподілу має обернену, але від цієї умови можна відмовитись.

Нехай F функція розподілу довільної випадкової величини ξ. При $0 < y < 1$ задамо функцію G наступним чином:

$$G(y) = \inf\{\, t : F(t) > y\}.$$

Теорема 3.4. Випадкова величина

$$\eta = G(\alpha)$$

має функцію розподілу $F(x)$.

Доведення*:* Враховуючи рівність $P\big(\alpha < F(x)\big) = F(x)$, доведемо, що

$$P(G(\alpha) < x) = P\big(\alpha < F(x)\big).$$

Для цього достатньо показати, що умова $G(y) < x$ і $y < F(x)$ рівносильні.

1. Нехай $G(y) < x$. Тоді $\inf\{\, t : F(t) > y\} < x$, і, значить $F(x) > y$.

2. Нехай $y < F(x)$. Позначимо $z = G(y) = \inf\{\, t : F(t) > y\}$. Тоді $F(z - \varepsilon) \le y$ для всіх $\varepsilon > 0$. За неперервністю функції F зліва, $F(z) \le y$. Таким чином , $F(z) \le y < F(x)$, і за рахунок монотонності F, $G(y) = z < x$. $\blacksquare$

$$\hat{\tau}_{n-1}(\mu) = \sqrt{n}\,\frac{\hat{\mu}_n - \mu}{\hat{s}_n}$$

має розподіл Стьюдента з $(n-1)$-ступенем свободи. Тому критерій вигляду $(|\hat{\tau}_{n-1}(\mu_0)|, [y_{n-1,\alpha}, \infty))$ для перевірки гіпотези $H_0 : \mu = \mu_0$ має вірогідний рівень α. Критерій є консистентним.

Гіпотеза про дисперсію при невідомому середньому. За теоремою про вибіркові середнє та дисперсію нормальної вибірки величина

$$\hat{\chi}^2_{n-1}(\sigma) = (n - 1)\hat{s}^2_n / \sigma^2$$

Має *хі*-квадрат розподіл з $(n-1)$-ступенем свободи. Відповідно критерій вигляду $\hat{\chi}^2_{n-1}(\sigma_0), [z_{n-1,\alpha}, \infty)$ має вірогідний рівень α.

Гіпотеза про різницю середніх при відомих дисперсіях. Нехай спостерігаються дві незалежні кратні нормальні вибірки $X = (\xi_1, \dots, \xi_n)$ та $Y = (\eta_1, \dots, \eta_m)$, що мають розподіли

$$\xi_1 \simeq N(\mu_1, \sigma_1^2), \eta_1 \simeq N(\mu_2, \sigma_2^2)$$

з невідомими середніми μ_k та відомими дисперсіями σ_k^2.

Нульова гіпотеза формулюється як

$$H_0 : \mu_2 - \mu_1 = \Delta,$$

де величина Δ відома. Часто розглядають випадок, коли Δ = 0.

Нехай $\hat{\mu}_n, \hat{\mu}_m$ – відповідні вибіркові середні для X і Y . Величини $\hat{\mu}_n, \hat{\mu}_m$ незалежні за теоремою про векторні перетворення незалежних величин, причому $\hat{\mu}_n - \mu_1 \simeq N(0, \sigma_1^2/n), \hat{\mu}_m - \mu_2 \simeq N(0, \sigma_2^2/m)$ за теоремою про вибіркові середнє та дисперсію нормальної вибірки. Тому за нульової гіпотези

$$\hat{\mu}_m - \hat{\mu}_n - \Delta \simeq N(0, \sigma_1^2/n + \sigma_2^2/m),$$

за теоремою про нормальність суми незалежних нормальних векторів.

Отже, статистика критерію

$$\hat{\zeta}_{nm} = (\hat{\mu}_m - \hat{\mu}_n - \Delta) / \sigma_1^2/n + \sigma_2^2/m$$

Тодi $(|\hat{\varkappa}(X, \theta_0)|, [x_\alpha, \infty))$ є критерієм рівня α для перевірки гіпотези $H_0 : \theta = \theta_0$.

Нехай $X = (\xi_1, \ldots, \xi_n)$ – кратна вибірка з нормальним розподілом спостережень $N(\mu, \sigma^2)$. Наведені критерії містять статистики: $\hat{\mu}_n$ – вибіркове середнє, $\hat{\sigma}_n^2$ – вибіркова дисперсія, та $\hat{s}_n^2$ – нормована вибіркова дисперсія. x_α означає двобічний квантиль вірогідного рівня $1 - \alpha$ для стандартного нормального розподілу: $P(|\zeta| < x_\alpha) = 1 - \alpha$. Аналогічний зміст мають $y_{n\alpha}$ для розподілу Стьюдента з n ступенями свободи, та $z_{n\alpha}$ – для xi-квадрат розподілу з n ступенями свободи.

Гіпотеза про середнє при відомій дисперсії. Нехай дисперсія σ^2 відома. Випадкова величина

$$\hat{\varkappa}_n(\mu, \sigma) = \sqrt{n}\,(\hat{\mu}_n - \mu)/\sigma$$

має стандартний нормальний розподіл за теоремою про вибіркові середнє та дисперсію нормальної вибірки. Тому двобічний критерій вигляду $(|\hat{\varkappa}_n(\mu, \sigma))|, [x_\alpha, \infty))$ для перевірки гіпотези $H_0 : \mu = \mu_0$ має вірогідний рівень α. Критерій є консистентним, оскільки при $\mu_0 \neq \mu$ статистика

$$\hat{\varkappa}_n(\mu_0, \sigma) = \sqrt{n}(\hat{\mu}_n - \mu + \mu - \mu_0)/\sigma \to^P \pm\infty, n \to \infty.$$

Гіпотеза про дисперсію при відомому середньому. При відомому середньому μ вибіркова дисперсія $\hat{\sigma}_n^2$ є статистикою, а величина

$$\hat{\chi}_n(\sigma) = n\hat{\sigma}_n^2 / \sigma^2 = \sum_{k=1}^{n} (\xi_k - \mu)^2 / \sigma^2$$

має xi-квадрат розподіл із n ступенями свободи за теоремою про вибіркові середнє та дисперсію нормальної вибірки. Тому критерій $(\hat{\chi}_n^2(\sigma_0), [z_{n\alpha}, \infty))$ для перевірки гіпотези $H_0 : \sigma = \sigma_0$ має вірогідний рівень α.

Гіпотеза про середнє при невідомій дисперсії. За теоремою про статистику Стьюдента від нормальної вибірки величина

Теорема 3.5. Нехай $F(x)$ - функція розподілу, і F_0 - звуження функції F на відрізок $[a, b]$. Нехай функція F_0 має обернену F_0^{-1}, і , крім того $F(a) = 0$ і $F(b) = 1$, тоді $G = F_0^{-1}$.

Доведення*:* Нехай $0 < y < 1$. Тоді

$$G(y) = \inf\{t \in \mathbb{R} : F(t) > y\} =$$
$$= \inf\{t \in [a, b] : F(t) > y\} = \inf\{t \in [a, b] : F_0(t) > y\} =$$
$$= \inf\{t \in [a, b] : t > F_0^{-1}(y)\} = F_0^{-1}(y). \blacksquare$$

РОЗДІЛ 4. Статистичний простір, вибірка. Статистики та оцінки. Емпірична фінкція розподілу

4.1. Статистична вибірка

Означення 4.1. *Статистичною вибіркою* називається довільна вимірна функція $X : \Omega \to S$ зі значеннями у вимірному вибірковому просторі (S, Σ, λ), де:

S – деяка множина (вибірковий простір),

Σ – сигма-алгебра підмножин S,

λ – деяка сигма-скінченна міра на Σ.

Вважається, що значення $X(\omega) = x$ є відомим для статистика (тобто спостерігається) і може використовуватись для отримання статистичних висновків.

Надалі вибірковим простором буде обиратися переважно евклідів простір $S = R^n$ із борелєвою сигма-алгеброю $\Sigma = B(R)$, тому під вибіркоюслід розуміти звичайний випадковий вектор, що спостерігається в стохастичному експерименті. У більшості випадків мірою λ слугує або точкова міра – відносно неї кожна одноточкова множина з певного класу має одиничне значення міри (у випадку дискретної вибірки X), або ж міра Лебега, якщо $S \subset R^n$ і вибірка X має сумісну щільність. У випадку, коли вектор X містить всю наявну інформацію про стохастичний експеримент, для спрощення часто вважають, що простір елементарних подій збігається з вибірковим простором: $(\Omega, \zeta) = (S, \Sigma)$, а окремі спостереження є елементарними подіями: $X(\omega) = \omega$.

Означення 4.2. Міра μ на деякому вимірному просторі (S, Σ) *абсолютно неперервна* відносно міри λ, (позначення $\mu << \lambda$), якщо для довільної множини $B \in \Sigma$ із $\lambda(B) = 0$ випливає $\mu(B) = 0$.

За теоремою Радона – Нікодима ця властивість еквівалентна існуванню вимірної інтегровної за мірою λ функції $f(x), x \in S$, такої, що

Якщо оцінка $\widehat{\theta}_n$є оцінкою максимальної вірогідності за емпіричними частотами $\hat{\nu}_n = (\hat{\nu}_{n1}, \dots, \hat{\nu}_{nk})$, тобто

$$\hat{\theta}_n = argmax_{\theta \in \Theta} L(\hat{\nu}_n, \theta) = argmax_{\theta \in \Theta} \prod_{i=1}^{k} p_i^{\hat{\nu}_{ni}}(\theta),$$

то *модифікована статистика хі-квадрат*

$$\hat{\chi}^2(\hat{\theta}_n) \equiv \sum_{i=1}^{k} \frac{\left(\hat{\nu}_{ni} - np_i(\hat{\theta}_n)\right)^2}{np_i(\hat{\theta}_n)}$$

слабко збігається до величини з *хі*-квадрат розподілом і кількістю ступенів свободи, що скоригована на число оцінених параметрів:

$$\hat{\chi}^2(\hat{\theta}_n) \to^W \chi^2_{\mathrm{k}-1-\mathrm{d}}, n \to \infty.$$

Критичними для гіпотези H_0 є великі значення статистики.

Модифікований критерій узгодженості хі-квадрат задається парою $(\chi^2_{k-1-d}\ (n), [z_\alpha, \infty))$, де критичний рівень z_α обчислюється з рівняння $P(\chi^2_{k-1-d} < z_\alpha) = 1 - \alpha$. З теореми 4.2 про асимптотику модифікованої статистики *хі*-квадрат випливає, що вірогідний рівень критерію наближено (при великих об'ємах вибірки) дорівнює α.

10.6. Перевірка гіпотез про параметри нормальних спостережень

Існує прямий зв'язок між перевіркою гіпотез про параметри та побудовою довірчих (надійних) інтервалів для них. А саме, довірчий інтервал для невідомого параметра можна розглядати для множину прийнятних значень для його гіпотетичного значення, а доповнення цього інтервалу – як відповідну критичну область для нульової гіпотези. Наприклад, більшість довірчих інтервалів для нормальних спостережень вигляду $[\hat{\theta}_1, \hat{\theta}_1]$ побудовані так, що

$$\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = \{|\hat{\varkappa}(X, \theta)| < x_\alpha\}$$

для деякої опорної величини$\hat{\varkappa}(X, \theta)$. Тут критичне значення x_αобране з умови

$$P\left(b\hat{\theta}_1 < \theta < \hat{\theta}_2\right) = 1 - \alpha.$$

а статистика *xi*-квадрат

$$\hat{\chi}^2_{k-1}(n) = \sum_{i=1}^{k} (\hat{v}_{ni} - np_i)^2 / np_i$$

слабко збігається за припущенням. Тому $\tilde{\chi}^2_{k-1}(n) \to^P \infty, n \to \infty$,і для кожного критичного рівня x_αпотужність $P_\theta(\tilde{\chi}^2_{k-1}(n) \geq x_\alpha) \to 1, n \to \infty$. Отже, критерій є консистентним.

Критерій *xi*-квадрат для складних гіпотез. Розглянемо поліноміальну схему випробувань, в якій розподіл спостереження залежить від значення векторного параметра $\theta \in \Theta \subset R^d$:

$$P_\theta(\zeta_1 = x_i) = p_i(\theta), i = \overline{1,k}.$$

Функції$p_i(\theta)$ вважаються повністю відомими і утворюють при кожному θ дискретний розподіл: $p_i(\theta) > 0, \sum_{i=1}^{k} p_i(\theta) = 1$. Як і вище, результати спостережень представлені вектором емпіричних частот $\hat{v}_n = (\hat{v}_{n1}, \dots, \hat{v}_{nk})$.

Розглянемо випадок, коли параметр θ не тільки невідомий, але й відсутні будь-які припущення про його значення. Тоді спочатку доведеться висунути якесь припущення щодо значення θ на підставі спостережень (тобто фактично оцінити це значення), а потім перевірити статистичну гіпотезу про відповідність істинного параметра його оцінці. Очевидно, що використання наведеного вище критерію *xi*-квадрат стає некоректним, оскільки підстановка замість значення θ його оцінки змінює розподіл статистики *x*і-квадрат $\hat{\chi}^2_{k-1}(n)$ і, зокрема, деформує рівень критерію.

Теорема 10.4 (про асимптотику модифікованої статистики *xi*- квадрат). Припустимо, що у поліноміальній схемі випробувань з k значеннями у одному випробуванні виконуються умови:

(1) $inf_{\theta,i} p_i(\theta) > 0$,

(2) $p_i(\theta) \in C^2(\Theta), \forall i = \overline{1,k}$,

(3) $rang\big(\partial p_i(\theta)/\partial\theta, i = \overline{1,k}\big) = d \equiv dim\Theta < k$.

$$\mu(B) = \int_B f(x)\,\lambda(dx)$$

Означення 4.3. Функція $f = d\mu/d\lambda$ називається *щільністю* міри μ відносно λ.

У деяких розділах статистики використовується умова підпорядкованості. Вона полягає в тому, що розподіл вибірки є абсолютно неперервним (має щільність) відносно фіксованої міри λ на вибірковому просторі :

$$P_\theta(X \epsilon .) << \lambda(\cdot), \forall\, \theta \,\epsilon\, \Theta.$$

У будь-якому випадку для зліченного параметричного простору Θ така міра завжди існує. Дійсно, довільна міра зі зліченної множини ймовірнісних мір ($\mu_\theta, \theta \in \Theta$) абсолютно неперервна відносно міри

$$\lambda = \sum_{\theta \,\epsilon\, \Theta} 2^{-n(\theta)}\, \mu_\theta$$

де $n(\theta)$ – номер елемента θ у послідовності Θ.

4.2. Функція вірогідності

Означення 4.4. Нехай $X : \Omega \to S$ – вибірка зі значеннями у вимірному просторі (S, Σ, λ), яка задовольняє умову підпорядкованості. *Функцією вірогідності* (або *функцією правдоподібності*) вибірки називається сумісна щільність розподілу вибірки відносно міри λ у вибірковому просторі:

$$L(x,\theta) \equiv \frac{dP_\theta(X \epsilon .)}{d\lambda(.)}(x),$$

тобто така вимірна за *x* функція, що для всіх $B \in \Sigma$ і всіх $\theta \in \Theta$

$$P_\theta(X \epsilon B) = \int_B L(x,\theta)\lambda(dx).$$

Для абсолютно неперервної вибірки X міра λ є мірою Лебега, функція вірогідності як функція *x* є сумісною щільністю розподілу:

$$P_\theta(X \epsilon B) = \int_B L(x,\theta)dx,$$

а для дискретної вибірки λ – точкова міра, функція вірогідності є дискретним розподілом імовірностей:

$$P_\theta(X \epsilon B) = \sum_{x \,\epsilon\, B \cap S} L(x,\theta).$$

Означення 4.5. *Вибірковою функцією вірогідності* (*емпіричною функцією вірогідності)* називається випадкова величина, що отримується в результаті підстановки у функцію вірогідності замість аргумента $x \in S$ значення вибірки як випадкового вектора

$$L(X,\theta) \equiv L(x,\theta)|_{\,x=X}.$$

4.3. Кратні вибірки

Часто у статистиці використовується схема багатократних спостережень.

Означення 4.6. Вимірний простір $(S, \Sigma, \lambda) = (R, \mathfrak{B}, \upsilon)^n$ є *n-кратним прямим добутком* вимірного простору (R, $\mathfrak{B}$, υ), якщо

$$S = R^n \equiv \{x = (x_1, \dots, x_n), x_k \epsilon R\},$$

$$\sum = \mathfrak{B} \otimes \dots \otimes \mathfrak{B} \equiv \sigma[B_1 \times B_2 \times \dots \times B_n, B_k \,\epsilon\, \mathfrak{B}, k = \overline{1,n}],$$

а міра $\lambda = \upsilon \times \dots \times \upsilon$ визначається на прямокутниках як добуток

$$\lambda(B_1 \times B_2 \times \dots \times B_n) \equiv \upsilon(B_1) \dots \upsilon(B_1).$$

Наприклад, $(R^n, B(R^n), L_n) = (R, B(R), L_1)^n$, де L_n – n-вимірна міра Лебега (довжина, площа, об'єм ...).

Означення 4.7. Нехай вибірковий простір (S, Σ, λ) є *n-кратним прямим добутком* (R,$\mathfrak{B}$,υ)^n. Випадковий вектор $X : \Omega \to S$ називається n-кратною вибіркою, якщо $X = (\xi_1, \dots, \xi_n)$ складається з незалежних у сукупності однаково розподілених випадкових величин $\xi_k, k = \overline{1,n}$, зі значеннями у просторі (R, $\mathfrak{B}$, υ), тобто

розподіл збігається з наперед заданим розподілом: $H_0 : \theta = (p_1, \dots, p_k)$. Як статистику критерію оберемо статистику хі-квадрат

$$\chi^2_{k-1}(n) = \sum_{i=1}^{k} \frac{(\hat{\upsilon}_{ni} - np_i)^2}{np_i} = n \sum_{i=1}^{k} \frac{(\hat{\upsilon}_{ni}/n - p_i)^2}{p_i}$$

що визначена в теоремі4.1 (Пірсона про асимптотику статистики *хі* -квадрат). Зауважимо, що за теоремою Бореля про асимптотику частоти успіху має місце збіжність

$$\hat{\upsilon}_{ni}/n \to^{P1} P(\varsigma_1 = x_i), n \to \infty.$$

Якщо гіпотеза H_0 не виконується, тобто остання гранична ймовірність не дорівнює p_i при певному i; то має місце збіжність

$$\chi^2_{k-1}(n) \to^{P1} \infty, n \to \infty.$$

Отже, критичними для гіпотези H_0 є великі значення статистики (4.2).

Критерій узгодженості хі-квадрат для перевірки гіпотези H_0 має вигляд $(\chi^2_{k-1}(n), [x_\alpha, \infty))$, де критичний рівень x_α знаходиться з умови

$$P(\chi^2_{k-1} < x_\alpha) = 1 - \alpha.$$

З теореми Пірсона випливає, що рівень критерію при великих об'ємах вибірки наближено дорівнює α.

Критерій *хі*-квадрат є консистентним для кожної простої альтернативи. Дійсно, якщо гіпотетичне значення $\theta = \left(\tilde{p}_i, i = \overline{1,k}\right)$ не дорівнює істинному $\left(p_i, i = \overline{1,k}\right)$, то

$$\chi^2_{k-1}(n) = \sum_{i=1}^{k} (\hat{\upsilon}_{ni} - n\tilde{p}_i)^2 / n\tilde{p}_i = \sum_{i=1}^{k} (np_i - n\tilde{p}_i + \hat{\upsilon}_{ni} - np_i)^2 / n\tilde{p}_i \geq$$

$$\sum_{i=1}^{k} (np_i - n\tilde{p}_i)^2/2 - (\hat{\upsilon}_{ni} - np_i)^2 / n\tilde{p}_i \geq n\Delta - c\chi^2_{k-1}(n),$$

де

$$\Delta = \sum_{i=1}^{k} (p_i - \tilde{p}_i)^2 / 2\,\tilde{p}_i > 0, c = max_i(p_i/\tilde{p}_i),$$

Оскільки вектори γ_j незалежні і однаково розподілені, то за класичною центральною граничною теоремою для випадкових векторів має місце слабка збіжність при $n \to \infty$:

$$\eta(n) = \sum_{j=1}^{n} \gamma_j/\sqrt{n} \to^{W} \ \xi \simeq N_k(0, I - q \cdot q').$$

З неперервності квадратичної функції та з означення слабкої збіжності векторів звідси випливає слабка збіжність випадкових величин

$$\hat{\chi}^2_{k-1}(n) = \eta^2(n) \to^{W} \xi^2 \, , n \ \to \infty.$$

За означенням вектор $q = \left(\sqrt{p_i}, i = \overline{1,k}\right)^T$ має одиничну норму. Нехай $(q_1, \ldots, q_{k-1})$ – доповнення q до ортонормованого базису в $\mathbb{R}^k$, а ортонормована матриця $U \equiv (q_1, \ldots, q_{k-1}, q)^T$. Тоді $Uq = e = (0, \ldots, 0, 1)^T$.

Розглянемо випадковий вектор $\beta = U\xi$.За теоремою про лінійні перетворення нормальних векторів вектор $\beta = (\beta_1, \ldots, \beta_k)^T$ є нормальним. Оскільки за теоремою про коваріаційну матрицю лінійного перетворення

$$M_\theta \beta = 0, \qquad Cov(\beta) = UCov(\xi)U^T = U(I - q \cdot q^T)U^T =$$
$$= UU^T - Uq \cdot (Uq)^T = I - e \cdot e^T = \left(\delta_{ij} 1\mathrm{I}_{i<k,} i, j = \overline{1,k}, \right)$$

то випадкові величини $(\beta_1, \ldots, \beta_{k-1})$ є незалежними у сукупності стандартними нормальними величинами, а $\beta_k = \ 0$, оскільки

$$D\beta_k = Cov(\beta)_{kk} = 0.$$

Тоді за означенням *хі*-квадрат розподілу $\xi^2 \simeq \chi^2_{k-1}$, оскільки:

$$\xi^2 = \xi'\xi = (U'\beta)'U'\beta = \beta' UU'\beta = \beta^2 = \sum_{i=1}^{k-1} \beta_i^2 \simeq \chi^2_{k-1}.$$

Отже, $\hat{\chi}^2_{k-1}(n) \to^{W} \xi^2 \simeq \chi^2_{k-1}$. ■

Критерій *хі*-квадрат для простих гіпотез. Припустимо, що статистичний простір відповідає наведеній вище поліноміальній схемі випробувань, причому невідомим параметром є розподіл результату одного випробування: $\theta \ = \ (p_1, \ldots, p_k)$. Розглянемо задачу перевірки простої гіпотези H_0, яка полягає в тому, що цей

$$P_\theta(X \epsilon B_1 \times B_2 \times \ldots \times B_n) = \prod_{k=1}^{n} P_\theta(\xi_1 \epsilon B_k) \ \forall \ B_k \in \mathfrak{B}, \theta \in \Theta.$$

Означення 4.8. Число *n* називається об’ємом вибірки X.

Якщо X – кратна вибірка, її вибіркова функція вірогідності позначається через $L_n(\mathrm{X}, \theta)$, де n – об’єм вибірки.

Означення 4.9. *Функцією вірогідності спостереження* для кратної вибірки називається щільність розподілу величини ξ_1 відносно міри υ :

$$f(y, \theta) = \frac{dP_\theta(\xi_1 \epsilon \, .)}{dv(.)}(y),$$

тобто така вимірна функція *f*, що

$$P_\theta(\xi_k \ \epsilon \ B) = \int_B f(y, \theta) v(dy), \forall \ \mathrm{B} \ \epsilon \, \mathfrak{B}, \forall \ \theta \in \Theta.$$

Теорема 4.1. (про функцію вірогідності кратної вибірки). Нехай кратна вибірка $X = (\xi_1, \ldots, \xi_n)$ містить спостереження ξ_k зі значеннями в просторі $(R, \mathfrak{B}, \upsilon)$. Припустимо, що ξ_k мають функцію вірогідності спостережень $f(y, \theta)$. Тоді функція вірогідності всієї вибірки X дорівнює добуткові

$$L_n(y, \theta) = \prod_{k=1}^{n} f(x_k, \theta) \, , x = (x_1, \ldots, x_n).$$

Доведення є наслідком теореми про критерій незалежності абсолютно неперервних величин, оскільки вибірковий вектор $X = (\xi_1, \ldots, \xi_n)$ утворений саме незалежними однаково розподіленими величинами.

4.4. Статистики та оцінки

Означення 4.10. *Статистикою* називається довільна вимірна функція від вибірки: $T = T(X)$, яка не містить значень невідомого параметра θ.

Множина значень статистики є довільним вимірним простором (C, $\mathcal{C}$). Найчастіше це евклідів простір: $(C, \mathcal{C})$ = (R^m, $\mathfrak{B}$ (R^m,)).

Означення 4.11. *Оцінкою* невідомого параметра $\theta \in \Theta$ називається будь-яка статистика зі значеннями у параметричному просторі Θ.

Щоб підкреслити спеціальний характер оцінки, часто її зображають у вигляді $\hat{\theta}$. Очевидно, оцінка є засобом для прогнозування, передбачення, оцінювання значення невідомого параметра θ на підставі спостережень *X*.

Зауваження 4.1. Якщо при кожному n спостерігається кратна вибірка X об'єму *n*, а спосіб, яким утворена оцінка, один і той самий (не залежить від об'єму вибірки n), то поняття "оцінка" використовують також у широкому розумінні як "послідовність оцінок", що утворені за одним правилом при різних значеннях об'єму вибірки. Послідовності оцінок позначаються через $\widehat{\theta}_n$.

Приклад 4.1. Вибіркове середнє є оцінкою, що дорівнює середньому арифметичному спостережень, які утворюють вибірку. Однак це є послідовність оцінок, оскільки при кожному значенні об'єму вибірки є окрема статистика.

4.5. Властивості оцінок

Надалі для двох випадкових величин ξ, η запис $\xi \simeq \eta$ означає, що ці величини мають однакові функції розподілу, отже, і однакові породжені міри Лебега – Стілтьєса. Символом $N(\mu, \sigma^2)$ позначатимемо випадкову величину з нормальним розподілом та середнім μ і дисперсією σ^2. Якість тієї чи іншої оцінки потребує порівняльного аналізу. Для порівняння оцінок чи їх послідовностей будемо використовувати такі поняття.

Означення 4.12. Оцінка $\hat{\theta}$ називається *незміщеною*, якщо її математичне сподівання збігається з точним значенням θ:

$$M_\theta\ \widehat{\theta} = \theta, \forall \theta \in \Theta.$$

Означення 4.13. Оцінка $\widehat{\theta}_n$ називається *асимптотично незміщеною*, якщо має місце асимптотична збіжність середніх

$$M_\theta\ \widehat{\theta}_n \to \theta, n \to \infty, \forall \theta \in \Theta.$$

Теорема 10.3 (теорема Пірсона про асимптотику статистики *xi* – квадрат). Нехай $\hat{\upsilon}_n = (\hat{\upsilon}_{n1}, \dots, \hat{\upsilon}_{nk})$ – вектор емпіричних частот у схемі з n поліноміальними випробуваннями $X = (\varsigma_1, \dots, \varsigma_n)$ та розподілом імовірностей $\theta = (p_1, \dots, p_k) > 0$ окремих спостережень. Тоді статистика хі-квадрат

$$\hat{\chi}^2_{k-1}(n) \equiv \sum_{i=1}^{k} \frac{(\hat{\upsilon}_{n1} - np_i)^2}{np_i}$$

має граничний *xi*-квадрат розподіл із $k-1$ ступенями свободи, що незалежить від параметра θ :

$$\hat{\chi}^2_{k-1}(n) \to^W \hat{\chi}^2_{k-1}, n \to \infty.$$

Доведення. Розглянемо *k*-вимірний випадковий вектор

$$\eta(n) = \left((\hat{\upsilon}_{ni} - np_i)/\sqrt{np_i}, i = \overline{1,k}\right)$$

та для кожного $j = \overline{1,n}$ незалежні однаково розподілені вектори

$$\gamma_j = \left(\left(1\mathrm{I}_{\{\xi_j = x_i\}} - p_i\right)/\sqrt{np_i}, i = \overline{1,k}\right).$$

За означенням емпіричних частот $\hat{\upsilon}_{ni}$ та вектора $\eta(n)$

$$\eta(n) = \left(\sum_{j=1}^{n}\left(1\mathrm{I}_{\{\xi_j = x_i\}} - p_i\right)/\sqrt{np_i}, i = \overline{1,k}\right) = \sum_{j=1}^{n} \gamma_j/\sqrt{n},$$

$$\hat{\chi}^2_{k-1}(n) = \eta^2(n) = \eta'(n)\eta(n).$$

Обчислимо середнє та коваріацію одного доданку в сумі для $\eta(n)$:

$$E_{\theta\gamma_1} = \left((P_\theta(\xi_1 = x_i) - p_i)/\sqrt{p_i}\,, i = \overline{1,k}\right) = 0,$$

$$Cov(\gamma_1) = \left(E_\theta\left(1\mathrm{I}_{\{\xi_j = x_i\}} - p_i\right)\left(1\mathrm{I}_{\{\xi_j = x_l\}} - p_l\right)/\sqrt{p_i p_l}, \qquad i, l = \overline{1,k}\right) =$$

$$= \left((p_i\delta_{il} - p_i p_l)/\sqrt{p_i p_l}\ i, l = \overline{1,k}\right) = \left((\delta_{il} - p_i p_l), i, l = \overline{1,k}\right) = I - q \cdot q^T,$$

де I – одинична матриця, а матриця $q \cdot q^T = \sqrt{p_i p_l}, i, l = \overline{1,k}$розміру$k \times k$ утворена декартовим добутком вектора-стовпчика $q \equiv \left(\sqrt{p_i}, i = \overline{1,k}\right)^T$ на свій транспонований вектор-рядок.

Є кількістю тих випробувань із числа n, що привели до значення x_i.

Оскільки вектор параметрів θ кратної вибірки X містить розподіл спостережень, то її функція вірогідності має вигляд

$$L(X,\theta)=\prod_{j=1}^{n}\left(\sum_{i=1}^{k}p_i 1\mathrm{I}_{\{\varsigma_1=x_i\}}\right)=\prod_{j=1}^{n}\left(\prod_{i=1}^{k}{p_i}^{1\mathrm{I}\{\varsigma_1=x_i\}}\right)=\prod_{i=1}^{k}{p_i}^{\hat{v}_{ni}},$$

а вектор відносних емпіричних частот $\hat{v}_n/n$ збігається з оцінкою максимальної вірогідності параметра θ вигляду:

$$arg\ max_{\theta\in\Theta}L(X,\theta)=arg\ max_{\theta\in\Theta}\sum_{i=1}^{k}\hat{v}_{ni}\ln p_i=$$

$$=\left(\hat{v}_n/n, i=\overline{1,k}\right)=\hat{v}_n/n.$$

Якщо інтерпретувати при кожному фіксованому i подію $\{\varsigma_1=x_i\}$ як j-й успіх у послідовності з n випробувань Бернуллі, а решту значень ς_j – як не успіх, прийдемо до висновку, що величина $\hat{v}_{ni}$ є відповідною кількістю успіхів у n випробуваннях Бернуллі. Тому вона має біноміальний розподіл із параметрами n і p_i, зокрема, $M_\theta\hat{v}_{ni}=np_{\mathrm{i}}, D_\theta\hat{v}_{ni}=np_{\mathrm{i}}(1-p_{\mathrm{i}})$, а відносна частота успіхів $\hat{v}_{ni}/n$ є строго консистентною асимптотичнонормальною оцінкою ймовірності p_i за теоремою про властивості відносної частоти. Оскільки переріз подій одиничної ймовірності має ймовірність 1, то вектор $\hat{v}_n/n$ є строго консистентною оцінкою для дискретного розподілу θ :

$$P_\theta(lim_{n\to\infty}\hat{v}_n/n=\theta)=P_\theta\left(\bigcap_{i=1}^{k}\{lim_{n\to\infty}\hat{v}_{ni}/n=p_{\mathrm{i}}\}\right)=1.$$

Емпіричні частоти залежні, причому випадковий вектор $\hat{v}_n\in R^k$ лежить у $(k-1)$-вимірному підпросторі: $\sum_{i=1}^{k}\hat{v}_{ni}=n$.

Статистика *хі*-квадрат

Означення 4.14. Оцінка $\hat{\theta}_n$ називається *консистентною* (або *слушною*, *спроможною*), якщо вона збігається за ймовірністю до істинного значення θ:

$$\hat{\theta}_n\overset{P_\theta}{\to}\theta, n\to\infty, \forall\theta\in\Theta,$$

тобто

$$P_\theta\left(\left|\hat{\theta}_n-\theta\right|\geq\varepsilon\right)\to 0, n\to\infty, \forall\,\varepsilon>0, \forall\theta\in\Theta.$$

Означення 4.15. Оцінка $\hat{\theta}_n$ називається *строго консистентною,* якщо вона збігається з імовірністю 1 до істинного значення θ :

$$\hat{\theta}_n\overset{\text{м.н.}}{\longrightarrow}\theta, n\to\infty, \forall\theta\in\Theta,$$

тобто

$$P_\theta(\exists\lim_{n\to\infty}\hat{\theta}_n=\theta)=1, \forall\theta\in\Theta.$$

Зауваження 4.2. Наведені властивості стосуються оцінок $\hat{\theta}$ для значення невідомого параметру θ. У випадку, коли цей параметр – векторний, доцільно розглядати також оцінки $\hat{\tau}$ для значень деякої функції $\tau=\tau(\theta)$ від параметра θ. Сформульовані вище означення поширюються також і на дану схему, якщо замінити θ на $\tau(\theta)$, а $\hat{\theta}$ на $\hat{\tau}$.

Означення 4.17. Оцінка $\hat{\theta}_n$ параметра θ називається *асимптотично нормальною*, якщо знайдеться числова нормуюча послідовність $c_n=c_n(\theta)$ така, що має місце слабка збіжність

$$c_n\left(\hat{\theta}_n-\theta\right)\overset{W_\theta}{\longrightarrow}\zeta\simeq N(0,1), n\to\infty, \forall\theta\in\Theta.$$

За теоремою про еквівалентність слабкої збіжності та в основному асимптотична нормальність еквівалентна збіжності

$$P_\theta(c_n(\hat{\theta}_n-\theta)<x)\to\Phi(x), n\to\infty, \forall x\in\mathbb{R}, \forall\theta\in\Theta,$$

де Φ – стандартна нормальна функція розподілу.

Зауваження 4.3 (про побудову інтервальної оцінки). Якщо оцінка $\hat{\theta}_n$ є асимптотично нормальною, а $c_n\sim\sqrt{n}/\sigma$, то за теоремою про добуток слабко збіжної послідовності зі збіжною

$$\sqrt{n}(\widehat{\theta_n}-\theta)\overset{W}{\to}\sigma\zeta\equiv\eta\simeq N(0,\sigma^2), n\to\infty.$$

У цьому разі величина σ^2 називається асимптотичною дисперсією оцінки $\hat{\theta}_n$. Істинне її значення $\sigma^2 = \sigma^2(\theta)$ є відомою функцією параметра θ. З наведеної слабкої збіжності випливає, що розподіл нормованої величини $\sqrt{n}(\hat{\theta}_n - \theta)/\sigma(\theta)$ наближається до розподілу стандартної нормальної величини ζ. Оберемо для заданого рівня $p \in (0,1)$ значення x_p так, щоб

$$P(|\zeta| \le x_p) = p.$$

Для знаходження x_p досить чисельно розв'язати рівняння

$$\Phi(x_p) = \frac{1+p}{2}.$$

Наприклад, $x_{0.997} \approx 3$ за правилом трьох сигма.

Тоді при великих n наближено

$$P_\theta \sqrt{n}|\hat{\theta}_n - \theta|/\sigma(\theta) \le x_p \backsimeq P(|\zeta| \le x_p) = p.$$

Припустимо, що функція $\sigma(\theta)$ неперервна, а оцінка $\hat{\theta}_n$ – консистентна. З цих припущень випливає збіжність за ймовірністю $\sigma(\hat{\theta}_n) \overset{P}{\to} \sigma(\theta)$.Тому можна наближено замінити $\sigma(\theta)$ на $\sigma(\hat{\theta}_n)$ під знаком імовірності і стверджувати (внаслідок теореми про добуток слабко збіжної послідовності зі збіжною), що при великих n подія $\{\sqrt{n}|\hat{\theta}_n - \theta|/\sigma(\hat{\theta}_n) \le x_p\}$ наближено теж має ймовірність p. Отже, властивості асимптотичної нормальності та консистентності дають можливість наближеної побудови інтервальних асимптотично незміщених оцінок надійності р для невідомого параметра, що мають вигляд

$$P_\theta\left(\hat{\theta}_n - x_p\sigma(\hat{\theta}_n)/\sqrt{n} \le \theta \le \hat{\theta}_n + x_p\sigma(\hat{\theta}_n)/\sqrt{n}\right) \approx p, \forall\theta \in \Theta.$$

Означення 4.17. Оцінка $\hat{\theta}_n$ називається *локально незміщеною*, локально консистентною тощо, якщо відповідна властивість виконується лише для значень параметра $\theta \in \Theta$ із деякого околу істинного значення.

$$D\hat{\rho}_n = \hat{\rho}_n^2 = 144c_n^2 \sum_{i,j=1}^{n} Cov(v_k, v_j)Cov(\tau_i, \tau_j) = \frac{1}{n-1}.$$

Критичними для гіпотези незалежності є великізамодулем значення $\hat{\rho}_n$. Тому критична область критерію про незалежність має вигляд $D_1 = \{\rho: |\rho| \ge r_\alpha\}$. Для забезпечення рівня α за нульової гіпотези при малих об'ємах n використовують табульовані значення квантилей розподілу статистики Спірмена, а при великих n – асимптотичнунормальність

$$\sqrt{n-1}\hat{\rho}_n \to^W \varsigma \simeq N(0,1),$$

звідки наближено $r_\alpha = x_{\alpha/2}/\sqrt{n-1}$, де $x_{\alpha/2}$– квантиль рівня $1-\alpha/2$ стандартного нормального розподілу.

10.5. Критерій *хі*-квадрат Пірсона у поліноміальній схемі Бернуллі

Критерій *хі*-квадрат – це найбільш універсальний з відомих критеріїв іможе бути застосований для великої кількості статистичних моделей.

Користуючись групуванням спостережень, або ж підрахунком емпіричних частот для певної групи подій, часто вибірковий вектор можна звестидо вигляду $X = (\varsigma_1, \dots, \varsigma_n)$, де окремі спостереження незалежні у сукупності, однаково розподілені та набувають k різних значень $\{x_1, \dots, x_k\}$:

$$P_\theta(\varsigma_1 = x_i) = p_i > 0, i = \overline{1,k}, \sum_{i=1}^{k} p_i = 1.$$

Така схема випробувань є узагальненням біноміальної схеми випробувань Бернуллі, де *k=2*; і називається *поліноміальною схемою випробувань*. Параметром у даній схемі виступає невідомий дискретний розподіл $\theta = \left(p_i, i = \overline{1,k}\right)$.

Аналогом відносної частоти успіхів для поліноміальної схеми є векторемпіричних частот: $\hat{v}_n = (\hat{v}_{n1}, \dots, \hat{v}_{nk})$, де величина

$$\hat{v}_{ni} = \sum_{j=1}^{n} 1\mathrm{I}_{\{\varsigma_1 = x_i\}} = |j: \varsigma_1 = x_i|, i = \overline{1,k},$$

H_0 про незалежність X і Y та неперервність спільної функції розподілу вектори v і τ незалежні в сукупності та рівномірно розподілені на множині $\prod_n$ перестановок порядку n. Зокрема, для математичних сподівань, дисперсій та коваріацій рангів справедливі наведені у попередньому розділі формули з заміноюоб’єму вибірки N на n :

$$Mv_k = M\tau_k = \frac{n+1}{2}, \qquad Dv_k = (n^2-1)/12,$$

$$Cov\left(v_k, v_j\right) = -(n+1)/12, j \neq k.$$

Використаємо для перевірки H_0 статистику, що визначає *вибірковий коефіцієнт кореляції* між векторами рангів:

$$\hat{\rho}_n = \sum_{k=1}^{n} (v_k - Mv_k)\,(\tau_k - M\tau_k)/\hat{\sigma}_v \hat{\sigma}_\tau,$$

$$\hat{\sigma}_v^2 = \sum_{k=1}^{n} (v_k - Mv_k)^2, \hat{\sigma}_\tau^2 = \sum_{k=1}^{n} (\tau_k - M\tau_k)^2.$$

За нерівністю Коші статистика $\hat{\rho}_n$ набуває значень із відрізку $[-1;\ 1]$. При повній тотожності векторів рангів $v = \tau$ що свідчить про повну позитивну залежність між X i Y , маємо $\hat{\rho}_n = 1$, а при повній протилежності $\hat{\rho}_n = -1$. Отже, великі за абсолютною величиною значення статистики $\hat{\rho}_n$ вказують на залежність векторів рангів та відповідних вибірок.

З використанням теореми про розподіл вектора рангів можна підрахувати, що

$$\hat{\sigma}_v^2 = \hat{\sigma}_\tau^2 = n(n^2-1)/12.$$

Звідси знаходимо еквівалентне зображення для *статистики Спірмена*

$$\hat{\rho}_n = 12c_n \sum_{k=1}^{n} (v_k - (n+1)/2)(\tau_k - (n+1)/2) =$$

$$= 1 - 6c_n \sum_{k=1}^{n} (v_k - \tau_k)^2.$$

де $c_n = 1/n(n^2-1)$. За нульової гіпотези

$$M\hat{\rho}_n = 12c_n \sum_{k=1}^{n} M(v_k - (n+1)/2)(\tau_k - (n+1)/2) = 0,$$

4.6. Вибірка

Означення 4.18. Множина значень $x_1, \dots, x_n$ величини ξ, що одержані в результаті n експериментів, називається *вибіркою* обсягу n.

Означення 4.19. Множину всіх значень величини ξ ми будемо називати *генеральною сукупністю.*

При одночасному дослідженні двох (або більшої кількості) ознак вибірка буде складатись із упорядкованих пар (або упорядкованих наборів) чисел.

Для розв’язування багатьох задач важливо знати розподіл досліджуваної випадкової величини, бо на його основі приймають ті чи інші рішення.

Нехай задано вибірку (статистичні дані) $x_1, \dots, x_n$ для величини ξ. Безпосередньо із вибірки важко зробити якісь висновки про властивості досліджуваної величини, тому необхідно провести первинну обробку результатів спостережень.

Означення 4.20. Статистичні дані $x^{(1)}, \dots, x^{(n)}$, що записані у порядку зростання $\left(x^{(i)} \leq x^{(i+1)}\right)$, називають варіаційним рядом, а його члени $x^{(i)}$ – варіантами.

У варіаційному ряді значення можуть повторюватись. Якщо ми будемо записувати тільки різні значення $x_1, \dots, x_k$ $k \leq n$ і, якщо значення x_i зустрічається у вибірці n_i раз, то число n_i називають частотою, а $\frac{n_i}{n}$– відносною частотою значення $x_{(i)}$.

Означення 4.21. Перелік різних варіант і відповідних їм частот (відносних частот) називають емпіричним або статистичним розподілом (статистичним рядом). Його можна подати у вигляді такої таблиці (надалі дужки в індексах будемо опускати):

Таблиця 4.1.

Варіанти x_i	x_1	x_2	…	x_k
Частоти n_i	n_1	n_2	…	n_k

Відзначимо, що $n_1 + n_2 + \cdots + n_k = n$ – обсяг вибірки. Такий розподіл називають дискретним.

Для графічного зображення дискретного розподілу використовують полігон частот (або відносних частот) – це ламана із вершинами в точках (x_i, n_i) (або $(x_i, n_i/n)$).

Якщо досліджується дискретна випадкова величина, кількість можливих значень якої досить велике, або досліджується неперервна випадкова величина, то будують інтервальний статистичний ряд або інтервальний розподіл. Для цього множину всіх значень $x_1, \dots, x_n$, від найменшого (x_{min}) до найбільшого (x_{max}), розбивають на певну кількість інтервалів і для кожного із інтервалів вказують число статистичних даних, які в нього попадають. Інтервали можуть мати однакову або різну довжину. При розбитті на нерівні інтервали їх необхідно підбирати так, щоб розподіл ξ в кожному із них був приблизно рівномірним. Кількість інтервалів залежить від мети дослідження. Інтервальний розподіл можна подати у вигляді такої таблиці:

Таблиця 4.2.

Інтервали	(a_0, a_1)	(a_1, a_2)	…	(a_{s-1}, a_s)
Частоти n_i	n_1	n_2	…	n_s

Тут n_i кількість статистичних даних, що належать i-му інтервалу, а $n_1 + n_2 + \cdots + n_s = n$. Якщо замість n_i поставити $\frac{n_i}{n}$, то будемо мати емпіричний розподіл відносних частот. Для подальшої обробки результатів спостережень від інтервального розподілу переходять до дискретного, замінюючи кожен інтервал його серединою.

Графічним зображенням інтервального розподілу є гістограма – сукупність прямокутників, основами яких є інтервали групування, а висоти дорівнюють

зі спільною неперервною функцією розподілу $G(y) = P(\eta_1 < y)$. Для перевірки нульової *гіпотези однорідності*

$$H_0 : G = F;\ X \text{ та } Y \text{ незалежні, кратні вибірки,}$$

використаємо об'єднану вибірку

$$Z = (\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_m),$$

та позначимо через v_j ранг j-го спостереження у вибірці *Z*. За нульової гіпотези вибірка *Z* є кратною вибіркою об'єма $N = n + m$, тому вектор v рівномірно розподілений на множині перестановок $\prod_n$ порядку *N*.

Визначимо *статистику Вілкоксона*

$$S_{nm} = \sum_{j=1}^{n} v_j,$$

яка задає сумарний ранг вибірки *X* в об'єднаній вибірці *Z*.

За нульової гіпотези H_0 розподіл статистики S_{nm} однозначно визначається лише значеннями *n; m*. Це дає можливість знайти критичний статистики для забезпечення заданого вірогідного рівня α. Вигляд критичної області визначається альтернативою.

Нехай альтернативою є від'ємний зсув вибірки *Y* відносно *X* на $\Delta > 0$, тобто

$$H_1 : G(y) = F(y + \Delta);\ \forall y \in R;\ X \text{ та } Y \text{ незалежні, кратні.}$$

У цьому разі $\eta_1 \doteq \xi_1 - \Delta$,отже, значення статистики S_{nm} будуть переважно більшими при виконанні альтернативи порівняно з нульовою гіпотезою H_0, оскільки елементи вибірки *X* отримуватимуть переважно більші ранги. Тому критичними для нульової гіпотези слід вважати великі значення статистики. Отже, критична область повинна мати вигляд $[x_{nm}(\alpha); \infty)$;де критичний рівень $x_{nm}(\alpha)$ визначається умовою $P\left(S_{mn} \geq x_{nm}(\alpha)\right) \approx \alpha$.

Критерій незалежності Спірмена. Розглянемо дві кратні вибірки $X = (\xi_1, \dots, \xi_n)$ та $Y = (\eta_1, \dots, \eta_n)$ однакового об'єму *n*. Позначимо через $v = (v_1, \dots, v_n)$ та $\tau = (\tau_1, \dots, \tau_n)$– вектори рангів цих вибірок. При виконанні гіпотези

$$P_\theta(b_{nm}\Delta - \hat{\varkappa}_n - \hat{\varkappa}_m \geq x_\alpha), \text{де } \Delta = sup_{x\epsilon R}|G(x) - F(x)| > 0,$$
$$\hat{\varkappa}_n = \sqrt{n} sup_{x\epsilon R}\left|\hat{F}_n(x) - F(x)\right| \to^W \varkappa',$$
$$\hat{\varkappa}_m = \sqrt{m} sup_{x\epsilon R}\left|\hat{G}_m(x) - G(x)\right| \to^W \varkappa'',$$

за теоремою Колмогорова про відхилення емпіричної функції розподілу. Тому $b_{nm}\Delta - \tilde{\varkappa}_n - \tilde{\varkappa}_m \to^P \infty$; отже, потужність критерію Смірнова прямує до одиниці при кожній простій альтернативі. Критерій є консистентним.

10.4. Деякі рангові критерії

Означення 10.8. *Ранговими статистиками* вибірки $X = (\xi_1, \ldots, \xi_n)$ називаються координати вектора рангів $v = (v_1, \ldots, v_n)$, де значення v_k задаєномер k-го спостереження ξ_k у варіаційному ряді $\left(\xi_{(k)}, k = \overline{1, n}\right)$:

$$\xi_k = \xi_{(v_k)}, (\xi_1, \ldots, \xi_n) = \left(\xi_{v_1}, \ldots, \xi_{v_n}\right).$$

Для кратної вибірки з неперервною функцією розподілу вектор рангів v рівномірно розподілений на множині всіх перестановок $\prod_n$ порядку n.

*Зауваження 3.1.*Увипадку, коли функція розподілу спостережень (ξ_k) не є неперервною (наприклад, коли ці спостереження є цілозначними), не можна виключити, що серед них знайдуться однакові. У цьому разі ранги визначають так, щоб вони були однаковими для однакових спостережень,однак щоб сума таких рангів не змінилася. Тому при виконанні подій

$$\xi_{n_1} = \cdots = \xi_{n_k} = x, \text{та } \xi_{(r-1)} < \xi_{(r)} = \cdots = \xi_{(r+k-1)} = x < \xi_{(r+k)}$$

обирають

$$v_{n_1} = v_{n_2} = \cdots = v_{n_k} = \frac{\left(r + (r+1) + .. + (r + k - 1)\right)}{k} = r + \frac{k-1}{2}.$$

Критерій однорідності Вілкоксона. Розглянемо кратну вибірку $X = (\xi_1, \ldots, \xi_n)$, в якій окремі спостереження мають невідому неперервну функцію розподілу $F(x) = P(\xi_1 < x)$ і одночасно незалежну кратну вибірку $Y = (\eta_1, \ldots, \eta_m)$

відношенню частоти інтервалу до його довжини $\frac{n_i}{h_i}$. Якщо висоти рівні $\frac{n_i}{nh_i}$, то таку сукупність прямокутників називають гістограмою відносних частот. Площа i-го прямокутника дорівнює відносній частоті $\frac{n_i}{n}$, яка при $n \to \infty$ збігається за ймовірністю до ймовірності попадання значень випадкової величини у відповідний інтервал. Якщо довжина h інтервалу мала, то ця ймовірність приблизно рівна $p(x)h$, де $p(x)$ – щільність. Отже, верхню межу гістограми можна розглядати як статистичний аналог щільності розподілу досліджуваної випадкової величини.

4.7. Вибіркові характеристики

Означення 4.22. Будь-які характеристики, які знаходяться на основі статистичний даних, називаються *емпіричними* або *статистичними характеристиками*, а характеристики, які знаходяться на основі розподілу досліджуваної величини, називаються *теоретичними характеристиками*.

Використання всіх статистичних даних для аналізу ознаки не завжди доцільно. Для аналізу властивостей досліджуваної ознаки на основі статистичних даних використовують числові характеристики вибірки (статистичні характеристики). До основних характеристик вибірки відносять величини, які характеризують середнє значення та розсіювання можливих значень досліджуваної величини.

Означення 4.23. Однією із основних характеристик середнього значення є *вибіркова середня* $\overline{\mathrm{x}}$ – середнє арифметичне результатів спостережень:

$$\bar{x} = \frac{1}{n}\sum_i x_i. \tag{4.1}$$

Попередню уяву про розсіювання статистичних даних дає розмах варіювання $R = x_{max} - x_{min}$, але ця величина є досить грубою характеристикою розсіювання.

Означення 4.24. До основних характеристик розсіювання статистичних даних відносять *вибіркову дисперсію*: (середнє арифметичне квадратів відхилень результатів спостережень від вибіркової середньої)

$$D_B = \frac{1}{n}\sum_i (x_i - \bar{x})^2 \qquad (4.2)$$

і вибіркове середнє квадратичне відхилення $\sigma_B = \sqrt{D_B}$

Для незгрупованих даних вибіркова середня і вибіркова дисперсія знаходяться за формулами (4.1) і (4.2). Якщо ж статистичні дані згруповані, то

$$\bar{x} = \frac{1}{n}\sum_i x_i n_i,\ D_B = \frac{1}{n}\sum_i (x_i - \bar{x})^2\, n_i \qquad (4.3)$$

де x_i – різні результати спостережень випадкової величини ξ, а n_i – відповідні частоти.

Для характеристики середнього значення розглядають також середнє геометричне, середнє гармонійне, моду, медіану, квантилі тощо. Розглянемо емпіричну функцію розподілу $F_n(x)$. Число $\widehat{x_p} = \max\{x\colon F_n(x) \le p\}$ називається емпіричною p - квантиллю.

Означення 4.25. Можна записати таку формулу:

$$\hat{x}_p = \begin{cases} x_{np}, & np - \text{ціле}, \\ x_{[np]+1}, & np - \text{не ціле}, \end{cases}$$

де $x_{[np]}$ – член варіаційного ряду із номером $[np]$. *Медіана* – це квантиль $\hat{x}_{1/2}$.

Означення 4.26. *Мода* – це член варіаційного ряду із найбільшою частотою.

Означення 4.27. *Емпіричним моментом* r - го порядку (r>0) називають величину

$$\hat{\alpha}_r = \frac{1}{n}\sum_i n_i\, x_i{}^r,$$

а центральним емпіричним моментом r - го порядку –

$$\hat{\mu}_r = \frac{1}{n}\sum_i n_i\,(x_i - \bar{x})^r.$$

Означення 4.28. *Асиметрія* випадкової величини – відношення центрального моменту 3-го порядку до куба середнього квадратичного відхилення цієї величини.

Критерій однорідності Смірнова. Одночасно з вибіркою $X = (\xi_1, \dots, \xi_n)$, розглянемо кратну вибірку $Y = (\eta_1, \dots, \eta_m)$ об'ємом m; що не залежить від X, із неперервною функцією розподілу окремих спостережень $G(y) = \mathbb{P}(\eta_1 < y)$. Параметризуючи розподіл повного вектора спостережень *(X; Y)* як пару функцій розподілу $\theta = (F, G)$, розглянемо складну нульову *гіпотезу однорідності*

$H_0\,:\, F = G,\ F$ – неперервна; *X і Y*– незалежні, кратні

проти альтернативи

$H_1\,:\, F \ne G,\ F, G$– неперервні; *X і Y* – незалежні, кратні.

Позначимо емпіричну функцію розподілу другої вибірки

$$\hat{G}_m(y) = \tfrac{1}{m}\textstyle\sum_{k=1}^m 1\mathrm{I}_{\{\eta_k < y\}}.$$

Теорема 10.2(теорема Смірнова про відхилення емпіричних функцій розподілу). За гіпотези H_0 розподіл статистики Смірнова:

$$\hat{\varkappa}_{nm} = \sqrt{\frac{mn}{m+n}}\, sup_{x\in\mathbb{R}}\left|\hat{F}_n(x) - \hat{G}_m(x)\right|$$

не залежить від вигляду невідомої функції розподілу $F = G$, і для всіх $x \in \mathbb{R}$ має місце збіжність

$$P(\hat{\varkappa}_{nm} < x) \to P(\varkappa < x) = K(x);\ n, m \to 1,$$

де функція визначена в (10.1).

Критерій Смірнова задається парою $(\hat{\varkappa}_{nm};\ [x_\alpha;\ \infty))$, де критичний рівень x_α визначається за вірогідним рівнем α умовою $K(x_\alpha) = 1 - \alpha$. Тут вигляд критичної області обумовлений характером статистики $\hat{\varkappa}_{nm}$. Рівень критерію наближено дорівнює

$$P(\hat{\varkappa}_n \ge x_\alpha) \approx P(\varkappa \ge x_\alpha) = 1 - K(x_\alpha) = \alpha.$$

Для оцінки потужності припустимо, що виконується альтернатива H_1

та позначимо $b_{nm} = \sqrt{\frac{mn}{m+n}}$.

Потужність критерію дорівнює

$$P_\theta\left(b_{nm} sup_{x\epsilon R}\left|\hat{F}_n(x) - \hat{G}_m(x)\right| \ge x_\alpha\right) =$$

$$= P_\theta\left(b_{nm} sup_{x\epsilon R}\left|\hat{F}_n(x) - F(x) + G(x) - \hat{G}_m(x) + F(x) - G(x)\right| \ge x_\alpha\right) \ge$$

(б) і має місце слабка збіжність $\hat{\varkappa}_n \to^W \varkappa, n \to \infty$, причому гранична величина має функцію розподілу Колмогорова:

$$P(\varkappa < x) = K(x) \equiv \sum_{k=-\infty}^{+\infty}(-1)^k exp(-2k^2x^2) \ , x > 0. \qquad (10.1)$$

За теоремою Колмогорова про відхилення емпіричної функції розподілу розподіл статистики $\hat{\varkappa}_n(X)$ за нульової гіпотези не залежить від невідомої функції розподілу F та має місце збіжність в основному

$$P(\hat{\varkappa}_n < x) \to P(\varkappa < x) = K(x), n \to \infty,$$

дефункція розподілу Колмогорова $K(x)$ повністю відома і визначена рівністю (10.1).

Критерій Колмогорова задається парою $\left(\hat{\varkappa}_n, [x_\alpha; \infty)\right)$, де критичний рівень x_α знаходитьсяза вірогідним рівнем α з умови $K(x_\alpha) = 1 - \alpha$. Нульова гіпотеза про узгодженість відхиляється, якщо $\hat{\varkappa}_n \geq x_\alpha$. Вірогідний рівень критерію наближено (і тим точніше, чим більший обсяг вибірки) дорівнює

$$P(\hat{\varkappa}_n \geq x_\alpha) \approx P(\varkappa \geq x_\alpha) = 1 - K(x_\alpha) = \alpha.$$

Отже, для скінченних n критерій Колмогорова є наближеним. Його можна зробити точним, якщо для побудови критичного рівня замість $K(x)$ використати точне значення функції розподілу статистики $\hat{\varkappa}_n$.

Для дослідження потужності критерію припустимо, що розглядається нульова гіпотеза $H_0 : \theta = F$, де F – гіпотетична функція розподілу спостережень, а істинна функція розподілу відповідає альтернативі $H_1 : \theta = \theta_1 \equiv G$. Тоді потужність критерію дорівнює

$$P_{\theta_1}\left(\sqrt{n} sup_{x\epsilon R}\left|\hat{F}_n(x) - F(x)\right| \geq x_\alpha\right) =$$

$$P_{\theta_1}\left(\sqrt{n} sup_{x\epsilon R}\left|\hat{F}_n(x) - G(x) + G(x) - F(x)\right| \geq x_\alpha\right) \geq$$

$$P_{\theta_1}\left(\sqrt{n}\Delta - \tilde{\varkappa}_n \geq x_\alpha\right), \text{де } \Delta = sup_{x\epsilon R}|F(x) - G(x)| > 0,$$

$$\tilde{\varkappa}_n = \sqrt{n} sup_{x\epsilon R}\left|\hat{F}_n(x) - G(x)\right| \to^W \varkappa, n \to \infty,$$

за теоремою Колмогорова у припущенні H_1, та $P_{\theta_1}\left(\sqrt{n}\Delta - \tilde{\varkappa}_n \geq x_\alpha\right) \to 1, n \to \infty$. Отже, потужність критерію Колмогорова прямує до одиниці при кожній простій альтернативі. Критерій є консистентним.

Означення 4.29. *Ексцес* випадкової величини – відношення центрального моменту 4-го порядку до четвертого степеня середнього квадратичного відхилення цієї величини, зменшене на число 3.

4.8. Емпірична функція розподілу

Розглянемо статистичний простір, в якому спостерігається кратна вибірка $X = (\xi_1, \dots, \xi_n)$ зі спостереженнями $\xi_k, k = \overline{1,n}$, що мають невідому теоретичну функцію розподілу $F(x) = P(\xi_k < x)$. У цьому випадку параметром θ статистичного простору є функція розподілу F.

Означення 4.32. Емпіричною функцією розподілу називається така параметрична сім'я статистик:

$$\hat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n}\prod{}_{\{\xi_k<x\}} = , x \in R.$$

Функція $\hat{F}_n(x)$ будується лише за значеннями вибірки та заданим аргументом x, є кусково-сталою та має прирости величини $1/n$ у точках $\xi_k, k = \overline{1,n}$. При кожній фіксованій елементарній події ω вона є дискретною функцією розподілу як функція аргументу x.

Теорема 4.2. (про властивості емпіричної функції розподілу). Для кожного x ∈ R значення $\widehat{F_n(x)}$ емпіричної функції розподілу:

1) має біноміальний розподіл:

$$P\left(\hat{F}_n(x) = \frac{k}{n}\right) = C_n^k F^k(x)\left(1 - F(x)\right)^{n-k}, k = \overrightarrow{0,n},$$

2) є незміщеною оцінкою значення теоретичної функції розподілу:

$$M\hat{F}_n(x) = F(x),$$

3) є строго консистентною оцінкою для цього значення:

$$\hat{F}_n(x) \xrightarrow{\text{м.н.}} F(x), n \to \infty,$$

4) є асимптотично нормальною оцінкою з асимптотичною дисперсією $F(x)(1 - F(x))$:

$$\sqrt{n}\left(\hat{F}_n(x) - F(x)\right) \xrightarrow{W} \eta \backsimeq N\left(0, F(x)\left(1 - F(x)\right)\right), n \to \infty.$$

Доведення. Зафіксуємо x. Розглянемо послідовність із n випробувань, в яких k-м успіхом називається подія $\{\xi_k < x\}$. Оскільки величини ξ_k незалежні в сукупності та однаково розподілені, то дана послідовність є схемою випробувань Бернуллі, причому ймовірність успіху не залежить від номера випробування та дорівнюватиме $\theta = P(\xi_k < x) = F(x)$. За означенням величина $\hat{F}_n(x)$ збігатиметься з відносною частотою успіхів. Тому всі вказані властивості емпіричної функції розподілу випливають із наведеної вище теореми про властивості відносної частоти – частотної оцінки ймовірності успіху у випробуваннях Бернуллі. ■

Емпірична функція розподілу $\hat{F}_n(x)$ є випадковим процесом: вона одночасно є функцією елементарної події $\omega \in \Omega$ та аргумента $x \in$ R.

Теорема 4.3 (теорема Глівенко – Кантеллі). Емпірична функція розподілу є рівномірно строго консистентною оцінкою функції розподілу F:

$$\sup_{x \in R}\left|\hat{F}_n(x) - F(x)\right| \xrightarrow{\text{м.н.}} 0, n \to \infty.$$

Доведення. Зафіксуємо $m \in$ N і визначимо при $0 \le k < m$ величини

$$x_{mk} = sup(x \in R : F(x) < k/m),$$

де $sup\,\emptyset \equiv -\infty$, та визначимо $x_m = \infty$. Ця послідовність не спадає за k та у припущенні невиродженості F містить хоча б один скінченний елемент. У випадку, коли $F(x) = \mathbf{1}_{\{c<x\}}$, твердження теореми очевидне.

Зауважимо, що $F(x_{mk}) = F(x_{mk} - 0) \le k/m$ та $F(x_{mk} + 0) \ge k/m$ за умови $x_{mk} > -\infty$, оскільки з $F(x_{mk} + 0) < k/m$ випливало б існування $x > x_{mk}$ таких, що $F(x) < k/m$.

Визначимо випадкову величину

$$d_{mn} = \max_{0 \le k \le m}\left|\hat{F}_n(x_{mk}) - F(x_{mk})\right|.$$

10.3. Непараметричні критерії для функції розподілу

Розглянемо кратну вибірку $X = (\xi_1, \dots, \xi_n)$, в якій окремі спостереження мають невідому неперервну функцію розподілу $F(x)=P(\xi_1<x)$. З формального погляду дана схема відповідає параметричному просторові Θ; що містить усі функції розподілу, а нульова гіпотеза щодо значенняпараметра $\theta = F$ для заданої функції розподілу має вигляд

$$H_0 : P(\xi_1<x) = F(x);\ \forall x \in \mathbb{R};$$

X – кратна вибірка вигляду.

Критерій узгодженості Колмогорова. Перевіряється гіпотеза узгодженості кратної вибірки з заданою функцією розподілу F, тобто припущення H_0 про те, що функція розподілуокремих спостережень збігається з наперед заданою неперервною функцією розподілу $F(x)$. Як альтернативу будемо розглядати клас усіх неперервних функцій розподілу, відмінних від F:

$$H_1 : P(\xi_1<x) = G(x);\ \forall x \in \mathbb{R};\ G \neq F;\ G \in \mathbb{C}(R);$$

X – кратна вибірка.

Критерій можна вважати непараметричним, адже невідомою є вся функція розподілу, а не окремі параметри.

Побудова критерію ґрунтується на *статистиці Колмогорова* :

$$\hat{\varkappa}_n(X) = \sqrt{n} sup_{x \in R}\left|\hat{F}_n(x) - F(x)\right|,$$

де $\hat{F}_n(x)$ – емпірична функція розподілу:

$$\hat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n} 1\mathrm{I}_{\{\xi_k<x\}}.$$

Теорема 10.1 (Колмогорова про відхилення емпіричної функції розподілу). Якщо теоретична функція розподілу $F(x)$ неперервна, то:

(а) розподіл статистики Колмогорова:

$$\hat{\varkappa}_n \equiv \sqrt{n} sup_{x \in R}\left|\hat{F}_n(x) - F(x)\right|$$

не залежить від вигляду невідомої функції F,

Зауваження 10.2. Якщо нульова гіпотеза є простою гіпотезою, то вірогідний рівень задається одним числом: P_{θ_0} *(X∈W)*, оскільки при $\theta = \theta_0$ розподіл вибірки відомий. Це число називають *P-значенням критерію*.

Імовірність $P_\theta(X\in W)$ попадання вибірки у критичну область як функція від усіх значень параметра $\theta\in\Theta_0\cup\Theta_1$ називається *оперативною характеристикою критерію*.

Маючи на меті одночасну мінімізацію ймовірностей помилок першого та другого роду, для створення якісного критерію треба відшукати таку статистику $\hat{\varkappa}(X)$, яка була б ”чутливою” до істинного значення параметра:

– при виконанні нульової гіпотези набувала переважно значень із множини D_0 (наприклад, ”досить помірних” значень) – що зменшує ймовірність похибки першого роду,

– при виконанні альтернативи набувала переважно значень у доповненні $D_1=D\backslash D_0$ (наприклад, ”надмірно великих значень”) – що зменшує ймовірність помилки другого роду.

Означення 10.5. Критерій $(\hat{\varkappa}(X), D_1)$ має *вірогідний рівень* α, якщо ймовірності помилок першого роду не перевищують α :

$$P_\theta(\hat{\varkappa}(X) \in D_1) = P_\theta(X\in W) \le \alpha,\ \forall\theta\in\Theta_0,$$

іхочабодназцихнерівностейєрівністю(хоча б для одного θ).

Зауваження 10.3. Якщо нульова гіпотеза є простою гіпотезою, то в означенні рівня критерію обов’язково має місце рівність. Дійсно, в такому випадку маємо єдине значення параметру і одну нерівність, яка за означенням перетворюється на рівність.

Означення 10.6. Критерій $(\hat{\varkappa}(X), D_1)$ є *незміщеним* критерієм, якщо його потужність не менша за його рівень α :

$$P_\theta(\hat{\varkappa}(X) \in D_1) = P_\theta(X\in W) \ge \alpha,\ \forall\theta\in\Theta_1.$$

Означення 10.7. Критерій $(\widehat{\varkappa_n},\ D_1)$ вірогідного рівня α є консистентним критерієм, якщо його потужність прямує до одиниці:

$$P_\theta(\widehat{\varkappa_n} \in D_1) \to 1,\ n\to\infty,\ \forall\theta\in\Theta_1.$$

За теоремою про властивості емпіричної функції розподілу має місце збіжність $\hat{F}_n(x_{mk}) - F(x_{mk}) \xrightarrow{\text{м.н.}} 0$ при $n\to\infty$ для кожного k. Тому $d_{mn} \xrightarrow{\text{м.н.}} 0$ при $n \to \infty$ для кожного фіксованого m.

Для кожного x існує k таке, що $x \in (x_{mk}, x_{m,k+1}]$. Якщо $x_{mk} > -\infty$, то

$$\hat{F}_n(x) - F(x) \le \hat{F}_n(x_{m,k+1}) - F(x_{mk} + 0) =$$
$$= \hat{F}_n(x_{m,k+1}) - \hat{F}_n(x_{m,k+1}) + \hat{F}_n(x_{m,k+1}) - \hat{F}_n(x_{mk} + 0) \le$$
$$\le d_{mn} + (k+1)/m - k/m = d_{mn} + 1/m.$$

У випадку $x_{mk} = -\infty$ за означенням $F(x) \ge k/m$ для всіх x, звідки отримуємо заміною $x_{mk} + 0$ на x таку ж саму оцінку.

Аналогічно доводимо відповідну оцінку знизу. Зважаючи на довільність x, з отриманих нерівностей виводимо, що

$$\sup_{x\in R}\left|\hat{F}_n(x) - F(x)\right| \le d_{mn} + 1/m.$$

Отже, для довільного $m \ge 1$ з імовірністю 1 виконується нерівність

$$\overline{\lim_{n\to\infty}} \sup_{x\in R}\left|\hat{F}_n(x) - F(x)\right| \le \overline{\lim_{n\to\infty}}\, d_{mn} + 1/m = 1/m,$$

звідки отримуємо при $m\to\infty$ твердження теореми. ■

Теорема 4.4 (теорема Колмогорова про відхилення емпіричної функції розподілу). Якщо теоретична функція розподілу F неперервна,

1) то розподіл статистики Колмогорова:

$$\hat{\aleph}_n \equiv \sqrt{n}\sup_{x\in R}\left|\hat{F}_n(x) - F(x)\right|$$

не залежить від вигляду невідомої функції F,

2) і має місце слабка збіжність

$$\hat{\aleph}_n \xrightarrow{W} \aleph, n \to \infty,$$

причому гранична величина має функцію розподілу Колмогорова:

$$P(\aleph < x) = K(x) \equiv \sum_{k=-\infty}^{+\infty}(-1)^k e^{-2k^2x^2}, x > 0.$$

Доведення. 1) Для спрощення припустимо, що функція F строго монотонна. Тоді коректно визначена обернена до F функція

$$F^{(-1)}(u) = sup(x : F(x) < u), 0 < u < 1,$$

причому $F(F^{(-1)}(u)) = u$.

Зробимо заміну змінної $x = F^{(-1)}(u), 0 < u < 1$, в означенні емпіричної функції розподілу

$$\widehat{F_n}\left(F^{(-1)}(u)\right) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{\xi_k < F^{(-1)}(u)\}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{F(\xi_k) < u\}}$$

де за монотонністю та означенням оберненої функції використано тотожність $\left\{\xi_k < F^{(-1)}(u)\right\} = \{F(\xi_k) < u\}$, а функція $\mathbf{1}_A: X \to \{0,1\}$, де X – довільна множина, визначена наступним чином

$$\mathbf{1}_A(x) = \begin{cases} 1, x \in A; \\ 0, x \notin A \end{cases}$$

є індикатором множини $A \subseteq X$.

Величини $\alpha_k \equiv F(\xi_k)$ незалежні за теоремою про перетворення незалежних величин, та рівномірно розподілені на [0,1]:

$$P(F(\xi_k) < u) = P(\xi_k < F^{(-1)}(u)) =$$
$$= F(F^{(-1)}(u)) = u, \forall u \in [0,1].$$

З означення отримуємо $\{F^{(-1)}(u), 0 < u < 1\} = \{x : 0 < F(x) < 1\}$, а при $F(x) \in \{0,1\}$ виводимо рівність $\widehat{F_n}(x) = F(x)$ м.н. Тому заміна $x = F^{(-1)}(u)$ в означенні статистики Колмогорова веде до рівності м.н.

$$\widehat{\aleph}_n = \sqrt{n} \sup_{u \in (0,1)} \left|\frac{1}{n}\widehat{F_n}\left(F^{(-1)}(u)\right) - F\left(F^{(-1)}(u)\right)\right| =$$

$$= \sqrt{n} \sup_{u \in (0,1)} \left|\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{\alpha_k < u\}} - u\right|.$$

Отже, за теоремою про обчислення ймовірностей і математичного сподівання функції від випадкового вектора розподіл величини $\widehat{\aleph_n}$ як функції від вектора $(\alpha_1, \dots, \alpha_n)$ з незалежних рівномірно розподілених величин визначається однозначно і не залежить від F.

(2.1) якщо воно справджується, то нульова гіпотеза H_0 на підставі спостережень X відкидається і приймається альтернатива H_1,

(2.0) якщо ж це включення не справджується, то нульова гіпотеза H_0 на підставі спостережень X не може бути відкинута, отже, приймається, а альтернатива H_1 відкидається.

Зауваження 10.1. Перевірка гіпотези ні в якому разі не є доведенням справедливості чи несправедливості припущення гіпотези. Невдача у відхиленні H_0 означає лише, ще немає досить вагомих свідчень для відхилення H_0 – це і є зміст висновку про те, що ця гіпотеза приймається. Про справедливість припущення можна казати лише при наявності багатосторонніх свідчень на його користь, як статистичних, так і інших.

10.2. Рівень та потужність критерію

Означення 10.3. *Помилкою(похибкою) першого роду* статистичного критерію називається відхилення нульової гіпотези за умови, що вона справджується. *Помилкою(похибкою)* другого роду називається прийняття нульової гіпотези за умови, коли справджується альтернатива.

Вказані помилкові рішення пов'язані з відповідними подіями. Їх імовірності називаються *ймовірностями помилок першого та другого роду*.

Означення 10.4. Нехай нульова гіпотеза має вигляд $H_0: \theta \in \Theta_0$, а альтернатива – $H_1: \theta \in \Theta_1$. *Вірогідним рівнем* (або *критичним рівнем*) статистичного критерія (1.3) називається функція від θ, що задає ймовірності помилок першого роду:

$$P_\theta(\hat{\varkappa}(X) \in D_1) = P_\theta(X \in W),\ \theta \in \Theta_0.$$

Потужністю критерію називається ймовірність відсутності помилки другого роду (тобто ймовірність правильного – альтернативного – висновку при альтернативі), щозадається функцією:

$$1 - P_\theta(\hat{\varkappa}(X) \in D_0) = P_\theta(\hat{\varkappa}(X) \in D_1) = P_\theta(X \in W),\ \theta \in \Theta_1.$$

робиться на підставі розгляду значення певної функції від вибірки – *статистики критерію*. Ця функція $\hat{\varkappa}(X): S \rightarrow D$ є довільною статистикою (вимірною функцією від вибірки) зі значеннями в деякому вимірному просторі *D*.

Для визначення результату статистичного висновку щодо якісної властивості розподілу спостережень на основі значення статистики $\hat{\varkappa}$досить розбити множину *D* на дві частини: $D = D_0 \cup D_1$ й у випадку

(0) включення $\hat{\varkappa} \in D_0$– приймати нульову гіпотезу,

(1) при $\hat{\varkappa} \in D_1$ – приймати альтернативу і відхиляти нульову гіпотезу.

Множина D_1, на якій відхиляється нульова гіпотеза, називається *критичною областю* статистики критерію. Очевидно, що довільний алгоритм дихотомічного вибору на підставі значення вибірки X можна подати у наведеному вигляді, обираючи, наприклад, $D = \{0, 1\}$, $D_0 = \{0\}$, $D_1 = \{1\}$ та відповідно конструюючи *D*-значну статистику критерію.

На практиці частіше обирають $D = R$, $D_0 = (-\infty, x0)$, $D_1 = [x0, \infty)$. У цьому випадку статистика критерію є числовою величиною, а x_0 визначає *критичний рівень* статистики критерію. Нульова гіпотеза відкидається за умови перевищення статистикою критерію критичного рівня: $\hat{\varkappa} \geq x_0$.

У загальному випадку з критичною областю статистики D_1 можна пов’язати *критичну область вибірки*

$$W = \{x \in S : \hat{\varkappa}(x) \in D_1\} = \{x \in S : \delta(x) = H_1\}.$$

При потраплянні вибіркового вектора X у критичну область нульова гіпотеза відкидається, у іншому випадку відкидається альтернатива.

Означення 10.2. *Статистичним критерієм* (статистичним тестом) називається пара $(\hat{\varkappa}(X), D_1)$, що утворена *D*-значною статистикою критерію $\hat{\varkappa}(X)$ та її критичною областю $D_1 \subset D$.

Алгоритм перевірки статистичної гіпотези H_0 проти альтернативи H_1 за допомогою критерію виконується в два етапи:

(1) обчислюють значення $\hat{\varkappa} = \hat{\varkappa}(X)$,

(2) перевіряють включення $\hat{\varkappa} \in D_1$, що еквівалентне $X \in W$,

2) Обчислення граничного розподілу статистики Колмогорова спирається на граничні теореми теорії випадкових процесів і не наводиться. ■

Теорема Колмогорова дає можливість обчислити інтервальну асимптотично незміщену оцінку надійності р для функції розподілу

$$P(|\widehat{F_n}(x) - x_p/\sqrt{n} \leq F(x) \leq P(|\widehat{F_n}(x) + x_p/\sqrt{n}, \forall x \in R|) \approx p,$$

де x_p – розв’язок рівняння $K(x_p) = p$.

Для більш точного оцінювання замість K слід використати формули для функцій розподілу статистик $\widehat{\aleph}_n$, що були знайдені В. С. Королюком.

4.8. Два типи статистик

Нехай дано вимірну функцію S від n аргументів. Вибіркова характеристика $S(X) = S(x_1, \dots, x_n)$ часто називається також статистикою. Будь яка статистика є випадковою величиною. Її розподіл повністю визначається розподілом $P(B) = P(x_1 \in B)$ (нагадаємо, що $S(X)$ можна розглядати як випадкову величину, задану на $(\mathfrak{X}^n, \mathfrak{A}_{\mathfrak{X}}^n, P)$, де P n-кратний прямий добуток одновимірних розподілів x_1.

Виділимо два класи статистик, які часто будуть зустрічатися в подальшому. Вони будуть побудовані за допомогою наступних двох видів функціоналів $G(F)$ від функцій розподілу F:

I. Функціонали виду

$$G(F) = h\left(\int g(x)dF(x)\right),$$

де g - задана борелева функція, h - функція, неперервна в точці

$$a = \int g(x)dF_0(x),$$

де F_0 така, що$X \in F_0$.

II. Функціонали $G(F)$, неперервні в «точці» F_0 в рівномірній метриці: $G(F^{(n)}) \rightarrow G(F_0)$, якщо $\sup_x\left|F^{(n)}(x) - F_0(x)\right| \rightarrow 0$, носії *) розподілів $F^{(n)}$ належать носію F_0. Тут F_0 знову є функція, для якої $X \in F_0$.

Відповідні класи статистик ми визначимо за допомогою рівності

$$S(X) = G(F_n^*),$$

де F_n^* - емпірична функція розподілу. Тоді ми отримаємо

I. Клас статистик I типу, які представлені у вигляді

$$S(X) = h\left(\int g(x)dF_n^*(x)\right) = h\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right).$$

Очевидно, всі вибіркові моменти мають вигляд адитивних статистик $\frac{1}{n}\sum_{i=1}^{n} g(x_i)$ і відносяться до статистик I типу.

II. Клас статистик, які ми будемо називати статистиками II типу або статистиками, неперервними в точці F_0.

Зрозуміло, що, наприклад, вибіркова медіана буде неперервною статистикою в точці F, якщо існує медіана ζ, $F(\zeta) = 1/2$, і F неперервна і строго зростає в точці ζ.

Належність функціоналів одному з названих класів, звичайно, не є альтернативною. Функціонал $G(F)$ може не належати до жодного з цих класів або належати обом відразу. Наприклад, якщо G є функціоналом I типу, носій F зосереджений на відрізку $[a, b]$ $(F(a) = 0, F(b) = 1)$, а функція g має обмежену варіацію на $[a, b]$, то G буде одночасно функціоналом II типу, так як в цьому випадку функціонал

$$\int g(x)dF(x) = g(b) - \int_a^b F(x)dg(x)$$

неперервний відносно F у рівномірній метриці. Сказане означає, що статистики I типу $\bar{x}$ і S^2 будуть також статистиками II типу, якщо $X \in P$ і P зосереджено на кінцевому інтервалі.

РОЗДІЛ 10. Перевірка статистичних гіпотез

Статистичною гіпотезою називається довільне припущення про розподіл вибірки, яка спостерігається у стохастичному експерименті.

Часто статистичні гіпотези формулюють у вигляді $H_o: \theta \epsilon \Theta_0$, де $\Theta_0 \subset \Theta$ – певна параметрична підмножина.

Статистичною *альтернативою* називають таке припущення про розподіл вибірки, яке вважається виконаним у випадку, коли не справджується основна статистична гіпотеза. Статистичні альтернативи мають вигляд, $H_1: \theta \epsilon \Theta_1$, де $\Theta_1 \subset \Theta$. Якщо $\Theta_1 = \Theta \setminus \Theta_0$, то альтернатива може не формулюватися явно. Якщо ж має місце строге включення $\Theta_0 \cup \Theta_1 \subset \Theta$, то основна гіпотеза H_0 на відміну від альтернативної гіпотези H_1, називається *нульовою гіпотезою*.

Означення 10.1. Статистична гіпотеза $H_0: \theta \epsilon \Theta_0$ називається *простою гіпотезою*, якщо множина $\Theta_0 = \{\theta_0\}$ є одноелементною, причому розподіл P_{θ_0} $(X \in \cdot)$ відомий повністю.

Приклад 10.1.

а) Нормальна вибірка з відомою дисперсією. Невідомим параметром нормального розподілу є середнє: $\theta = \mu$, дисперсія σ^2 вважається відомою. При фіксації μ щільність спостережень відома, тому гіпотеза $H_0: \theta = \mu_0$– проста.

б) Нормальна вибірка з невідомою дисперсією. Якщо параметр $\theta = (\mu, \sigma^2)$, то при виконанні гіпотези $H_0: \theta = \mu_0$ вибіркові щільності можуть мати різні дисперсії, тому ця гіпотеза не є простою.

10.1. Статистика критерію, критична область

Абстрактно критерій для перевірки H_0 проти H_1 на підставі вибірки X можна визначити як функцію $\delta(X) : S \to \{H_0, H_1\}$, що для кожного вектора X приймає висновок на користь однієї з гіпотез. Однак на практиці статистичний висновок

$$(1) \qquad P\left(n\,\hat{\sigma}_n^2/z_{n\alpha} \le \sigma^2 \le n\,\hat{\sigma}_n^2/z_{n,2-\alpha}\right) =$$
$$= P\left(z_{n,2-\alpha} \le \chi_n^2 \le z_{n\alpha}\right) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha.$$

Оцінка середнього при невідомій дисперсії. За теоремою про статистику Стьюдента від нормальної вибірки статистика $\tau_n = \sqrt{n}(\widehat{\mu_n} - \mu)\,/\,\hat{s}_n$ має розподіл Стьюдента з $n-1$ ступенем свободи. Тому
$$P\left(\widehat{\mu_n} - \frac{\hat{s}_n}{\sqrt{n}} y_{n-1,\alpha} \le \mu \le \widehat{\mu_n} + \frac{\hat{s}_n}{\sqrt{n}} y_{n-1,\alpha}\right) = P\left(|\tau_n| \le y_{n-1,\alpha}\right) = 1 - \alpha.$$

Оцінка дисперсії при невідомому середньому. Оскільки $(n-1)\,\hat{s}_n^2/\sigma^2 \simeq \chi_{n-1}^2$, то
$$P\left((n-1)\,\hat{s}_n^2/z_{n-1,\alpha} \le \sigma^2 \le (n-1)\,\hat{s}_n^2/z_{n-1,2-\alpha}\right) =$$
$$P\left(z_{n-1,2-\alpha} \le \chi_{n-1}^2 \le z_{n-1,\alpha}\right) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha.$$

РОЗДІЛ 5. Точкове оцінювання невідомих параметрів. Метод моментів. Незміщенні оптимальні оцінки

5.1. Метод підстановки

Оцінка - це те ж саме, що і статистика, тобто будь яка вимірна функція $\widehat{\theta}$ від вибірки. Неформально, зміст, вкладений в цей новий термін, полягає в тому, що оцінками θ^* ми називаємо лише статистики, призначені для використання замість невідомого параметра θ. Іншими словами, $\widehat{\theta}$ - деяке наближення для θ, що базується на вибірці. Величину $\widehat{\theta}$ називають також точковою оцінкою для θ, на відміну від інтервальних оцінок, які будуть розглянуті пізніше.

Задання оцінки зазвичай передбачає задання функцій (від вибірок X_n), визначених при всіх можливих значеннях n. Тому надалі, термін «оцінка» буде означати сім'ю статистик $\widehat{\theta} = \hat{\theta}_n\,(X_n)$, визначених для всіх $n = 1,2,\ldots$, де $\widehat{\theta}$ - функція на $\mathfrak{X}^n$ або, що те ж саме, одну функцію $\widehat{\theta} = \widehat{\theta}\,(n, X_\infty)$, визначену на добутку множини цілих чисел і $\mathfrak{X}^\infty$.

Будемо вважати, що в постановці задачі про оцінку визначено множину Θ можливих значень параметра θ і сім'ю $\mathcal{P}$ можливих розподілів P вибірки X (це можуть бути, скажімо, лише нормальні розподіли, або розподіли Пуассона, для яких необхідно оцінити невідомі параметри α, λ). Якщо які-небудь обмеження на θ (на P) відсутні, то можна вважати, що Θ ($\mathcal{P}$) співпадає з евклідовим простором відповідної розмірності (з множиною всіх розподілів).

Якщо для позначення параметра, замість θ, використовується яка-небудь інша буква, наприклад λ, то оцінки цього параметра будуть позначатися тим самим способом: додаванням до λ верхнього індексу «зірочка». Наприклад, для параметра α нормального закону природно розглядати оцінку
$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Вибіркові моменти, що використовуються для оцінки

$$\mathrm{M}x_1 = \int x\mathrm{P}(dx) \ \text{ і } \ Dx_1 = \int (x - \mathrm{M}x_1)^2\mathrm{P}(dx),$$

мають свої спеціальні традиційні позначення

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \ \text{ і } \ S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3.$$

Як було сказано раніше, для даного параметра можна вказати як завгодно багато різних оцінок, і перш ніж обговорювати, яким чином в кожній конкретній ситуації слід порівнювати їхні переваги, ми зупинимося на деяких загальних «регулярних» методах їх побудови.

Ці методи поєднують у собі найбільш розумні підходи до проблеми оцінки і дозволять нам отримувати надалі, найкращі, в тому чи іншому сенсі, оцінки.

В основі майже всіх прийомів оцінювання лежить наступний основний метод, який можна назвати *методом підстановки емпіричного розподілу* (або, коротко, *методом підстановки*).

Нехай $X_n \in P$, і невідомий параметр θ представимо у вигляді деякого функціоналу G від розподілу P:

$$\theta = G(P).$$

Нехай, надалі, P_n^*, як і раніше, означає емпіричний розподіл. Тоді відповідно до методу підстановки, в якості оцінки $\widehat{\theta}$ необхідно взяти функцію

$$\widehat{\theta} = G(\hat{P}_n).$$

Такі оцінки ми будемо називати *оцінками за методом підстановки* або, коротко, просто *оцінками підстановки*.

Функціонал G іноді буває заданий в неявному вигляді як розв'язок деякого рівняння $H(\theta, P) = 0$, відносно θ. У цьому випадку, оцінками підстановки, відповідно до основного визначення ми будемо називати будь який розв'язок рівняння $H(\theta, \hat{P}_n) = 0$.

Якщо відомо, що множина можливих значень параметра $\theta \in R^k$ обмежена областю Θ із R^k, яка не співпадає з R^k, то цю інформацію можна врахувати при

має розподіл Стьюдента *із* $n-1$ *ступенем свободи.*

Доведення. За теоремою про вибіркові середнє та дисперсію нормальної вибірки після ділення на σ чисельник $\sqrt{n}(\widehat{\mu}_n - \mu) / \sigma$ має стандартний нормальний розподіл, і не залежить від знаменника

$$\widehat{s}_n/\sigma = \sqrt{(n-1)\widehat{s}_n{}^2/\sigma^2(n-1)} \simeq \sqrt{\chi^2_{n-1}/(n-1)}.$$

Тому весь дріб τ_{n-1} має розподіл Стьюдента. ■

9.2. Побудова довірчих інтревалів для параметрів нормального розподілу

Розглянемо двосторонні інтервальні оцінки для параметрів кратної нормальної вибірки $X = \xi_1 .. \xi_n, \xi_k \simeq N(\mu, \sigma^2)$.

Нехай x_α означатиме двосторонній квантиль вірогідного рівня $1-\alpha$ для стандартного нормального розподілу:

$$P(|\varsigma| \le x_\alpha) = 1 - \alpha \text{ , } \Phi(x_\alpha) = 1 - {}^{\alpha}/_{2}.$$

Аналогічний зміст мають квантилі $y_{n\alpha}$ для розподілу Стьюдента, $z_{n\alpha}$ – для хі-квадрат розподілу:

$$P(|\tau_n| \le y_{n\alpha}) = 1 - \alpha \text{ , } P(x_n^2 \le z_{n\alpha}) = 1 - \alpha/2.$$

Доведення наведених нижче оцінок є очевидним наслідком теореми 9.2 та означення відповідних квантилей.

Оцінка середнього при відомій дисперсії. Оскільки при відомій дисперсії σ^2 статистика $\varsigma = \sqrt{n}(\widehat{\mu_n} - \mu) / \sigma \simeq N(0,1)$ є стандартною нормальною, то

$$P\left(\widehat{\mu_n} - \frac{\sigma}{\sqrt{n}}x_\alpha \le \mu \le \widehat{\mu_n} + \frac{\sigma}{\sqrt{n}}x_\alpha\right) = P(|\varsigma| \le x_\alpha) = 1 - \alpha.$$

Оцінка дисперсії при відомому середньому. У випадку відомого середнього σ для відповідно модифікованої вибіркової дисперсії $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^{n}(\xi_k - \mu)^2$ зі співвідношення $n\,\widehat{\sigma}_n^2/\sigma^2 \simeq \chi_n^2$ виводимо, що

$$\sqrt{n}\,\frac{\widehat{\mu}_n-\mu}{\sigma} \simeq N(0,1), \qquad \frac{(n-1)\widehat{s}_n^{\,2}}{\sigma^2} \simeq \chi^2_{n-1},$$

$$n\,\frac{\widehat{\sigma}_n^{\,2}}{\sigma^2} = \frac{1}{\sigma^2}\sum_{k=1}^{n}(\xi_k-\mu)^2\,\chi^2_n.$$

Доведення. Розглянемо лінійне перетворення

$$\zeta = (\,\zeta_k = (\xi_k - \mu)/\sigma, \qquad k=\overline{1,n}).$$

Вектор ζ є стандартним нормальним вектором, тому що за означенням величини $(\xi_k-\mu)/\sigma$ є незалежними стандартними нормальними. Оскільки перші дві статистики в твердженні теореми збігаються відповідно з нормованими вибірковим середнім та вибірковою дисперсією для вибірки ζ, то перше твердження теореми є наслідком теореми про вибіркові моменти стандартної нормальної вибірки. Третя статистика збігається із сумою $\sum_{k=1}^{n}\zeta_k^2$ і має хі-квадрат розподіл за означенням. ■

Означення 9.2. Випадкова величина τ_n має $t-$*розподіл, або розподіл Стьюдента,* із n ступенями свободи, якщо її можна подати у вигляді

$$\tau_n = \frac{\zeta}{\sqrt{\chi_n^2/n}},$$

де $\zeta \simeq N(0,1)$ – стандартна нормальна величина, а χ_n^2 – незалежна від неї величина із хі-квадрат розподілом та n ступенями свободи.

Означення 9.3. Випадкова величина $\phi_{n,m}$ має *розподіл Фішера*, або ж *Снедекора – Фішера*, із n, m ступенями свободи, якщо її можна подати у вигляді

$$\phi_{n,m} = \frac{\chi_n^2\,/\,n}{\chi_m^2\,/\,m},$$

де χ_n^2, χ_m^2 – незалежні величини із хі-квадрат розподілом та n, m ступенями свободи відповідно.

Теорема 9.3 (про статистику Стьюдента від нормальної вибірки). Нехай $X=(\xi_1,\dots,\xi_n)$– кратна вибірка з нормальним розподілом $\xi_k \simeq N(\mu,\sigma^2)$. Тоді статистика

$$\tau_{n-1} = \sqrt{n}\,\frac{\widehat{\mu}_n-\mu}{\widehat{s}_n}$$

побудові оцінок підстановки. Припустимо, що область Θ -замкнена, і нехай $\mathcal{P}$ - множина можливих розподілів вибірки X, $\Theta = \{G(P)\}_{P\in\mathcal{P}}$. Визначимо функціонал $G_1(P)$ для довільного P як значення
$t \in \Theta$, для якого досягається

$$\min_{t\in\Theta}|t-G(P)| = |G_1(P)-G(P)|; \tag{5.1}$$

тому $G_1(P)$ – точка з Θ, найближча до $G(P)$. Оскільки, $G_1(P)=G(P)=\theta$, якщо $P\in\mathcal{P}$, то оцінка

$$\widehat{\theta} = G_1(\hat{P}_n) \tag{5.2}$$

разом з $G(\hat{P}_n)$, буде оцінкою підстановки, при цьому, множина можливих значень $\widehat{\theta}$ буде належати Θ.

Про оцінки (5.1), (5.2) будемо говорити, що вони отримані звуженням методу підстановки.

Наприклад, нехай оцінюється параметр α нормального розподілу $\Phi_{\alpha,1}$, і нам наперед відомо, що $\alpha\in[0,1]$. Тоді може виявитися, що оцінка $\hat{\alpha}=\bar{x}\notin[0,1]$ (очевидно, що $\bar{x}=\int t\,d\hat{F}_n(t)$ - оцінка підстановки). Звуження методу підстановки рекомендує в якості оцінки взяти точку з $[0,1]$, найближчу до $\bar{x}$.

Відзначимо тепер, що в сформульованому вигляді метод підстановки не завжди має сенс. Справа в тому, що вихідний функціонал G може виявитися не визначений на множині емпіричних розподілів. Нехай, наприклад, наперед відомо, що розподіл P належить класу $\mathcal{P}$ абсолютно безперервних за мірою Лебега розподілів, тому кожне $P\in\mathcal{P}$ має щільність f. Нас цікавить значення

$$\theta = G(P) = \int f^2(x)dx = \int\left(\frac{dP}{dx}\right)^2 dx.$$

Зрозуміло, що в цьому випадку $G(\hat{P}_n)$ не має сенсу, оскільки $\hat{P}_n$ - дискретний розподіл. У таких випадках метод підстановки завжди можна природним чином модифікувати так, щоб він зберіг свою суть. У наведеному прикладі, де $G(P)$ є функціоналом від щільності f, слід в якості $\widehat{\theta}$, відповідно до методу підстановки, розглянути значення $G(\hat{P}_n)$, де $\hat{P}_n$ - згладжений емпіричний розподіл.

Може виявитися також, що в деяких випадках $G(P_n^*)$ існуватиме не для всіх X_n, а лише для $X_n \in A_n$, де $P(X_n \in A_n) \to 1$ при $n \to \infty$. Ця обставина для подальшого викладу ролі грати не буде, і для визначеності можна взяти $G(\hat{P}_n) = 0$ для $X_n \notin A_n$.

У цьому параграфі ми для простоти будемо припускати, що $G(\hat{P}_n)$ буде існувати при всіх $X_n \in \mathfrak{X}^n$ і що $\widehat{\theta}$ - випадкова величина, тобто функція $G(\hat{P}_n)$ здійснює вимірне зображення $\mathfrak{X}^n$ в R^k, де k - розмірність θ.

Принцип підстановки є досить природнім підходом до задачі, оскільки, як ми вже знаємо, розподіл $\hat{P}_n$ необмежено наближається до P із зростанням n.

Нехай $X_n = [X_\infty]_n$ та F, як зазвичай, функція розподілу, яка відповідає P.

Теорема 5.1. *Нехай $\theta = G(P)$ і функціонал G належить одному з двох класів: або він представляється у вигляді*

$$G(P) = h\left(\int g(x)\, dF(x)\right), \tag{I}$$

де h - функція, неперервна в точці $a = \int g(x)\, dF_0(x)$ (функціонал I типу), або у вигляді

$$G(P) = G_1(F), \tag{II}$$

де функціонал G_1 неперервний у точці F_0 у рівномірній метриці (функціонал II типу). Тоді якщо$X \in F_0$, то $\widehat{\theta} = G(\hat{P}_n)$ є строго консистентна оцінка:

$$\widehat{\theta} \underset{\text{м.н.}}{\longrightarrow} \theta.$$

5.2. Вибіркові моменти, метод моментів

Велику групу статистик утворюють вибіркові моменти, які за законом великих чисел є природними оцінками для теоретичних моментів - математичних сподівань степеневих функцій від випадкової величини.

Означення 5.1. Нехай ξ - випадкова величина. Її (нецентральним) *теоретичним моментом* порядку $k \in \mathbb{N}$ називається число

$$\mu_k \equiv M\xi^k,$$

де χ^2_{n-1} – випадкова величина з хі-квадрат розподілом, що має $n-1$ ступінь свободи.

Доведення. Визначимо вектор $q = \left(\frac{1}{\sqrt{n}},\ k = \overline{1,n}\right)$, що має одиничну евклідову норму. Тоді $\widehat{\mu}_n = \zeta^T q/\sqrt{n}$ і

$$(n-1)\widehat{s}_n{}^2 = \sum_{k=1}^{n} (\zeta_k - \widehat{\mu}_n)^2 = \sum_{k=1}^{n} \zeta_k^2 - \sqrt{n}\,\widehat{\mu}_n = \zeta^T\zeta - (\zeta^T q)(q^T\zeta).$$

Нехай $(q_1, \dots, q_{n-1})$ – доповнення вектора q до ортонормованого базису в $\mathbb{R}^n$, що існує за теоремою Грама – Шмідта. Позначимо $U = (q_1, \dots, q_{n-1}, q)^T$ ортонормовану матрицю з рядками $q_1^T, \dots, q_{n-1}^T, q^T$, що породжена цим базисом. Зауважимо, що за ортонормованістю значення $U_q = (0, \dots, 0, 1) \equiv e$ є одиничним вектором. Крім того, за теоремою про лінійні перетворення нормальних векторів вектор $\eta = U\zeta$ знову є стандартним нормальним вектором. Тому

$$(n-1)\widehat{s}_n{}^2 = \zeta^T\zeta - (\zeta^T q)(q^T\zeta) = \zeta^T(I - q\cdot q^T)\zeta =$$

$$= \eta^T U(I - q\cdot q^T)U^T\eta = \eta^T(I - e\cdot e^T)\eta = \sum_{k=1}^{n-1} \eta_k^2 \simeq \chi^2_{n-1}$$

за означенням хі-квадрат розподілу, оскільки $\eta_k,\ k = \overline{1, n-1}$, є незалежними у сукупності величинами із стандартним нормальним розподілом.

За теоремою про векторні перетворення незалежних величин статистика χ^2_{n-1} не залежить від $\eta_n = q^T\zeta = \sqrt{n}\,\widehat{\mu}_n$, оскільки величини $(\eta_1, \dots, \eta_{n-1})$ та η_n незалежні внаслідок означення стандартного нормального вектора $\eta = (\eta_1, \dots, \eta_n)$.

Нарешті, середнє $\widehat{\mu}_n$ як лінійне перетворення за теоремою про лінійні перетворення нормальних векторів має нормальний розподіл, із середнім 0 і дисперсією $D\,\widehat{\mu}_n = \frac{n}{n^2} = \frac{1}{n}$. ■

Теорема 9.2 (про вибіркові середнє та дисперсію нормальної вибірки). Нехай $X = (\xi_1, \dots, \xi_n)$ – кратна вибірка з нормальним розподілом спостережень $\xi_k \simeq N(\mu, \sigma^2)$. Тоді статистики $\widehat{\mu}_n$ *і* $\widehat{s}_n{}^2$ незалежні, причому

РОЗДІЛ 9. Довірчі інтервали. Інтервальні оцінки параметрів нормальних спостережень

Означення 9.1. *Інтервальною оцінкою* (або *довірчим інтервалом*) невідомого параметра θ називається випадковий інтервал $(\hat{\theta}_1, \hat{\theta}_2)$, який покриває невідоме точне значення θ із заданою ймовірністю (надійністю) γ, тобто

$$P\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = \gamma.$$

Випадкові величини $\hat{\theta}_1$ та $\hat{\theta}_2$, які є функціями вибіркового вектора $X = (x_1, \dots, x_n)$, називаються відповідно *нижньою* та *верхньою* межами довірчого інтервалу; ймовірність γ ще називають *довірчою ймовірністю*.

Точність оцінювання параметра θ характеризується довжиною довірчого інтервалу $\hat{\theta}_2 - \hat{\theta}_1$ і залежить від обсягу вибірки n та надійності γ.

На практиці надійність оцінки γ беруть такою, щоб подію з ймовірністю γ можна було вважати практично вірогідною, наприклад, γ = 0,95; 0,99; 0,999. Часто величину β записують у вигляді $\gamma = \alpha - 1 \in (0,1)$. Мале число α називається *рівнем довіри* оцінки.

9.1. Статистики від нормальних вибірок

Теорема 9.1 (про вибіркові моменти стандартної нормальної вибірки). Нехай вибірка $\zeta = (\zeta_1, \dots, \zeta_n)$ є стандартним нормальним вектором. Тоді вибіркові моменти

$$\hat{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} \zeta_k, \qquad \hat{s}_n^{\,2} = \frac{1}{n-1}\sum_{k=1}^{n} (\zeta_k - \hat{\mu}_n)^2$$

незалежні, причому

$$\sqrt{n}\,\hat{\mu}_n \simeq N(0,1), \qquad (n-1)\hat{s}_n^{\,2} \simeq \chi^2_{n-1},$$

за умови інтегрованості величини ξ^k.

Центральним теоретичним моментом порядку k називається число

$$\mu_k^0 \equiv M(\xi - \mu)^k,$$

де $\mu \equiv \mu_1$ - математичне сподівання величини ξ. Зокрема, $\mu_2^0 = \sigma^2$ - дисперсія ξ.

Зауваження 5.1. Значення центральних моментів використовуються в теорії розподілів для означення таких спеціальних характеристик:

$k_a = \mu_3^0/\sigma^3$ - *коефіцієнт асиметрії* (нульовий для симетричних розподілів),

$k_e = \mu_4^0/\sigma^4 - 3$ - *коефіцієнт ексцесу* (нульовий для нормальних спостережень).

Означення 5.2. Нехай $X = (\xi_1, \dots, \xi_n)$ - вибірка. (Нецентральним) *вибірковим моментом* порядку k називається статистика

$$\hat{\mu}_{kn} \equiv \frac{1}{n}\sum_{i=1}^{n} \xi_i^k.$$

Центральним вибірковим моментом порядку k називається статистика

$$\hat{\mu}_{kn}^0 \equiv \frac{1}{n}\sum_{i=1}^{n} (\xi_i - \hat{\mu}_n)^k,$$

де $\hat{\mu}_n \equiv \hat{\mu}_{1n}$ - *вибіркове середнє*. Зокрема, $\hat{\mu}_{2n}^0 \equiv \hat{\sigma}_n^2$ - *вибіркова дисперсія*.

Зауваження 5.2. Важливою властивістю центральних моментів є *інваріантність відносно зсувів*: вони не змінюються при одночасному зсуві всіх спостережень на сталу:

$$\hat{\mu}_{kn}^0 = \frac{1}{n}\sum_{i=1}^{n}\left(\xi_i - \frac{1}{n}\sum_{j=1}^{n}(\xi_j)\right)^k = \frac{1}{n}\sum_{i=1}^{n}\left(\xi_i - c - \frac{1}{n}\sum_{j=1}^{n}(\xi_j - c)\right)^k.$$

Зауваження 5.3. Як і у випадку теоретичних моментів, справедлива тотожність

$$\hat{\sigma}_n^2 = \hat{\mu}_{2,n} - (\hat{\mu}_n)^2.$$

Дійсно,

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \hat{\mu}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}\xi_i^2 - 2(\hat{\mu}_n)^2 + (\hat{\mu}_n)^2.$$

Теорема 5.2 (про моменти вибіркових моментів). Нехай $X = (\xi_1, \dots, \xi_n)$ - кратна вибірка. Тоді

$$M\hat{\mu}_{kn} = \mu_k, \qquad D\hat{\mu}_{kn} = \frac{1}{n}(\mu_{2k} - \mu_k^2).$$

Зокрема, $M\hat{\mu}_n = \mu$, $D\hat{\mu}_n = \sigma^2/n$. Крім того,

$$M\hat{\sigma}_n^2 = \frac{n-1}{n}\sigma^2, D\hat{\sigma}_n^2 = \frac{n-1}{n^3}\left((n-1)\mu_4^0 - (n-3)\sigma^4\right).$$

Доведення. Незміщеність нецентральних моментів є очевидним наслідком лінійності математичного сподівання. Вираз для їх дисперсій випливає з незалежності в сукупності і однакової розподіленості степеневих функцій $\left(\xi_i^k, i = \overline{1,n}\right.$ та з теореми про дисперсію суми незалежних величин:

$$D\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i^k\right) = \frac{1}{n^2}\sum_{i=1}^{n}D\left(\xi_i^k\right) = \frac{1}{n^2}nD(\xi_1^k) = \frac{1}{n}\left(M\xi_1^{2k} - \left(M\xi_1^k\right)^2\right).$$

Середнє для вибіркової дисперсії обчислюється з урахуванням останнього зауваження шляхом піднесення до квадрату під знаком суми:

$$M\hat{\sigma}_n^2 = M(\hat{\mu}_{n2} - \hat{\mu}_n^2) = M\hat{\mu}_{n2} - M\hat{\mu}_n^2 = \mu_2 - \frac{\left(M\sum_{i=1}^{n}\xi_i^2 + M\sum_{i\neq j}\xi_i\xi_j\right)}{n^2} =$$

$$\mu_2 - \mu_2/n - \mu^2(n-1)/n = (\mu_2 - \mu^2)(n-1)/n = \sigma^2\,(n-1)/n.$$

При обчисленні дисперсії $\hat{\sigma}_n^2$ можна вважати, $\mu = 0$, оскільки оцінка $\hat{\sigma}_n^2$ інваріантна відносно зсувів, зокрема, не змінюється при відніманні від спостережень їх спільного середньогоμ. Тому:

$$n^4 M(\hat{\sigma}_n^2)^2 = M\left(n\sum\xi_k^2 - \left(\sum\xi_k\right)^2\right) = M\left((n-1)\sum\xi_k^2 - 2\sum_{i<j}\xi_i\xi_j\right)^2 =$$

$$(n-1)^2 M\left(\sum\xi_k^2\right)^2 + 4M\left(\sum_{i<j}\xi_i\xi_j\right)^2 - 4(n-1)M\left(\sum\xi_k^2\sum_{i<j}\xi_i\xi_j\right) =$$

З опуклості функції $ln\,x$ виводимо елементарну нерівність $ln\,x\ \leq\ x-1$ при $x \in \mathbb{R}$, звідки за монотонністю математичного сподівання

$$I(\theta,\theta_0) = -M_{\theta_0}\ln\zeta \geq -M_{\theta_0}(\zeta-1) = 0.$$

Оскільки нерівність $ln\,x\ \leq\ x-1$ є строгою при $x \neq 1$, то за властивістю позитивності математичного сподівання від невід’ємної та нену льової м.н. випадкової величини $\zeta - 1 - \ln\zeta$ інформація за Кульбаком дорівнює нулю тоді й тільки тоді, коли $\zeta = 1$ м.н., тобто коли $f(y,\theta) = f(y,\theta_0)$ майже для всіх y на просторі значень спостережень. За умовою консистентності остання властивість еквівалентна $\theta = \theta_0$.

Неперервність $I(\theta,\theta_0)$ випливає з неперервності щільності спостережень та з теореми Лебега про мажоровану збіжність:

$$\lim_{n\to\infty} I(\theta_n,\theta_0) = -\lim_{n\to\infty} M_{\theta_0}\ln\frac{f(\xi_1,\theta_n)}{f(\xi_1,\theta_0)} = -M_{\theta_0}\lim_{n\to\infty}\ln\frac{f(\xi_1,\theta_n)}{f(\xi_1,\theta_0)} = I(\theta,\theta_0),$$

оскільки величини під знаком математичного сподівання за умов консистентності не перевищують інтегровну величину 2η. ■

(c1) Для всіх $\theta \neq \theta_0$ відношення $f(y,\theta)/f(y,\theta_0)$ не дорівнює майже скрізь одиниці на просторі значень спостережень $y \in R$ – тобто функції $f(y,\theta)$, $f(y,\theta_0)$ відрізняються на множині додатної міри.

(c2) Параметрична множина Θ є компактною підмножиною евклідового простору $\mathbb{R}^d$.

(c3) При кожному y функція $f(y,\cdot)$ є строго додатною та неперервно диференційовною за θ, причому величини $\ln f(\xi_1,\theta)$ та $\partial \ln f(\xi_1,\theta)/\partial\theta$ мажоруються інтегровною випадковою величиною η :

$$|\ln f(\xi_1,\theta)\ | \le \eta, \forall\theta \in \Theta,$$

$$|\partial \ln f(\xi_1,\theta)/\partial\theta| \le \eta, M\eta < \infty, \forall\theta \in \Theta.$$

8.3. Інформація за Кульбаком

Означення 8.2. *Функцією інформації за Кульбаком* називається числова функція

$$I(\theta,\theta_0) = -M_{\theta_0} \ln\frac{f(\xi_1,\theta)}{f(\xi_1,\theta_0)}.$$

Теорема 8.2 (про властивості інформації за Кульбаком). За умов консистентності ОМВ:

(а) $I(\theta,\theta_0) \ge 0$,

(б) $I(\theta,\theta_0) = 0$ *тоді й тільки тоді, коли* $\theta = \theta_0$,

(в) $I(\theta,\theta_0)$ *неперервна за* θ.

Доведення. Розглянемо випадкову величину

$$\zeta = \frac{f(\xi_1,\theta)}{f(\xi_1,\theta_0)}.$$

За формулою з теореми про обчислення математичного сподівання функції від випадкової величини та з означення щільності $f(y,\theta)$ обчислимо

$$M_{\theta_0}\zeta = \int_R \frac{f(y,\theta)}{f(y,\theta_0)} f(y,\theta_0)\nu(dy) = \int_R f(y,\theta)\nu(dy) = 1.$$

$$(n-1)^2(n\mu_4^0 + n(n-1)\sigma^4) + 2n(n-1)\sigma^4 - 0 =$$
$$(n-1)[(n-1)\mu_4^0 + (n^2-2n+3)\sigma^4],$$
$$D\hat{\sigma}_n^2 = M(\hat{\sigma}_n^2)^2 - (M\hat{\sigma}_n^2)^2 = M(\hat{\sigma}_n^2)^2 - \sigma^4(n-1)^2/n^2 =$$
$$(n-1)\big((n-1)\mu_4^0 - (n-3)\sigma^4\big)/n^3,$$

де використане припущення $M\xi_i = 0$. ■

З теореми про моменти вибіркових моментів випливає, що вибіркова дисперсія $\hat{\sigma}_n^2$ є зміщеною оцінкою для теоретичної дисперсії σ^2. Тому часто використовують її скоригований незміщений варіант.

Означення 5.3. *Нормованою вибірковою дисперсією* називається статистика

$$\hat{s}_n^2 \equiv \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \hat{\mu}_n)^2 = \frac{n}{n-1}\hat{\sigma}_n^2,$$

яка, очевидно, є незміщеною оцінкою дисперсії:

$$M\hat{s}_n^2 = \frac{n}{n-1}M\hat{\sigma}_n^2 = \sigma^2 \ .$$

Теорема 5.3 (про властивості вибіркових моментів). Нехай $X = (\xi_1,\dots,\xi_n)$ - кратна вибірка.

(а) Якщо$M|\xi_1|^k < \infty$, то нецентральний вибірковий момент $\hat{\mu}_{kn}$ є незміщеною та строго консистентною оцінкою відповідного теоретичного моменту μ_k.

(б) Якщо $M\xi_1^{2k} < \infty$, то оцінки $\hat{\mu}_{kn}$ є асимптотично нормальними:

$$\sqrt{n}(\hat{\mu}_{kn} - \mu_k) \xrightarrow{W} \eta \simeq N(0,\mu_{2k} - \mu_k^2).$$

Доведення. Випадкові величини $\zeta_i \equiv \xi_i^k$ є незалежними за теоремою про перетворення незалежних величин та однаково розподіленими за теоремою про збереження однакової розподіленості, оскільки отримані застосуванням однієї (степеневої порядку k) функції до величин з однаковими розподілами. Умова (а) гарантує справедливість критерію Колмогорова посиленого закону великих чисел, що і доводить (а).

Для доведення (б) застосуємо класичну центральну граничну теорему. Скінченність математичних сподівань та дисперсій, а також формула для дисперсії одного доданка наведені в теоремі про моменти вибіркових моментів. Тому асимптотична нормальність є наслідком класичної центральної граничної теореми. ■

Вибіркові моменти можна розглядати одночасно у вигляді випадкового *вектора вибіркових моментів* наперед заданої розмірності d:

$$\hat{\mu}_n^{(d)} \equiv \left(\hat{\mu}_{kn}, k = \overline{1,d}\right), \quad M\hat{\mu}_n^{(d)} = \mu^{(d)} \equiv \left(\mu_k, k = \overline{1,d}\right).$$

Теорема 5.4 (про асимптотику вектора вибіркових моментів). Нехай $X = (\xi_1, \dots, \xi_n)$ - кратна вибірка.

(а) Якщо $M|\xi_1|^d < \infty$, то векторна статистика $\hat{\mu}_n^{(d)}$ є строго консистентною оцінкою для вектора $\mu^{(d)}$.

(б) Якщо $M\xi_1^{2d} < \infty$, то вектор вибіркових моментів є асимптотично нормальним:

$$\sqrt{n}\left(\hat{\mu}_n^{(d)} - \mu^{(d)}\right) \xrightarrow{W} \eta \simeq N_d\left(0, V^{(d)}\right), n \to \infty,$$

з коваріаційною матрицею $V^{(d)} = (\mu_{k+l} - \mu_k\mu_l, \ \ k, l = \overline{1,d})$.

Доведення.

(а). З теореми про властивості вибіркових моментів виводимо, що $\hat{\mu}_{kn} \xrightarrow{P_\theta 1} \mu_k$ для всіх $k = \overline{1,d}, \ \theta \in \Theta$. Тому

$$P_\theta\left(\lim \hat{\mu}_n^{(d)} = \mu^{(d)}(\theta)\right) = P_\theta\left(\bigcap_{k=1}^{d}\left\{\lim \hat{\mu}_{kn} = \mu_k\right\}\right) = 1.$$

(б). Розглянемо послідовність випадкових векторів

$$\gamma_j \equiv \xi_j^k - \mu_k, k = \overline{1,d}, \qquad j \geq 1.$$

Вони є незалежними за теоремою про перетворення незалежних величин, та однаково розподіленими, оскільки отримані з однаково розподілених величин застосуванням однієї функції. Дана послідовність задовольняє умови класичної центральної граничної теореми для випадкових векторів із параметрами $m = 0, \ V = V^{(d)}$, тому що

$$M\gamma_j = 0, \operatorname{Cov}(\gamma_j) = \left(M\left(\xi_j^k - \mu_k\right)\left(\xi_j^l - \mu_l\right), k, l = \overline{1,d}\right) =$$

$$M_\theta\,\widehat{\theta}_n = \int_0^\infty \frac{n}{x}\theta\frac{(\theta x)^{n-1}}{(n-1)!}exp(\theta x)dx = \theta\frac{n}{n-1} \neq \theta \ \ \forall n \geq 1 \ \forall\theta > 0.$$

Отже, у загальному випадку не можна розраховувати на незміщеність ОМВ.

4. *Нормальна вибірка*. Як показано вище в розділі про оцінювання середнього для нормальних спостережень,

$$\ln L(X,\theta) - \frac{n}{2}\ln 2\pi = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}\left(\widehat{\mu}_{2n} - 2\,\widehat{\mu}_n\mu + \mu^2\right) =$$
$$= -\frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}\left(\widehat{\mu}_{2n} - (\widehat{\mu}_n)^2 + (\widehat{\mu}_n - \mu)^2\right) =$$
$$= -\frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}\left(\widehat{\sigma}_n{}^2 + (\widehat{\mu}_n - \mu)^2\right) \leq = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}\,\widehat{\sigma}_n{}^2 =$$
$$= -\frac{n}{2}\left(\ln\sigma^2 + \frac{1}{\sigma^2}\,\widehat{\sigma}_n{}^2\right) \leq -\frac{n}{2}\left(\ln\,\widehat{\sigma}_n{}^2 + \frac{1}{\widehat{\sigma}_n{}^2}\,\widehat{\sigma}_n{}^2\right) =$$
$$= -\frac{n}{2}\left(\ln\,\widehat{\sigma}_n{}^2 + 1\right),$$

оскільки функція $\ln s + \frac{a}{s}$ набуває найменшого значення при $s = a$. Дані нерівності є рівностями при $\mu = \widehat{\mu}_n, \ \sigma^2 = \widehat{\sigma}_n{}^2$. Тому оцінка максимальної вірогідності параметра $\theta = (\mu, \sigma^2)$ дорівнює

$$\widehat{\theta}_n = \arg\max \ln L(X,\theta) = \left(\widehat{\mu}_n, \widehat{\sigma}_n{}^2\right).$$

Зокрема, за теоремою про інваріантність оцінки максимальної вірогідності для границь вірогідного інтервалу з правила трьох сигма має вигляд

$$\widehat{\mu \pm 3\sigma} = \widehat{\mu}_n \pm 3\sqrt{\widehat{\sigma}_n{}^2}.$$

8.2. Умови консистентності ОМВ

Нехай X $=(\xi_1, \dots, \xi_n)$– кратна вибірка зі щільністю $f(y,\theta)$ одного спостереження.

У даному розділі символом θ_0 будемо позначати істинне значення параметра θ.

Розглянемо такі умови консистентності ОМВ:

Зауваження 8.1. Для практичного обчислення ОМВ як кореня рівняння $U(X, \widehat{\theta}) = 0$ часто використовують метод Ньютона – Рафсона, який складається з двох етапів.

(1) Обирають якусь консистентну оцінку $\widehat{\theta}_0 = \widehat{\theta}_0(X)$.

(2) Рекурентно обчислюють

$$\widehat{\theta}_{m+1} = \widehat{\theta}_m + \frac{U(X, \widehat{\theta}_m)}{I(\widehat{\theta}_m)},$$

де $I(\theta)$ – функція інформації за Фішером.

Доведено, що за певних умов границя $\lim \widehat{\theta}_m$ існує та збігається з ОМВ.

8.1. Приклади обчислення оцінок максимальної вірогідності

1. *Схема Бернуллі:*

$$\arg\max \ln L(X, \theta) = \arg\max \ln \theta^{\upsilon_n(X)} (1-\theta)^{n-\upsilon_n(X)} =$$

$$= \arg\max\big(\theta \ln \upsilon_n(X) + (1-\theta)\ln\big(n - \upsilon_n(X)\big)\big) = \frac{\upsilon_n(X)}{n} = \widehat{\theta}_n.$$

2. *Пуассонівська вибірка:*

$$\ln L(X, \theta) = \ln \prod_{k=1}^{n} \frac{\theta^{\xi_k}}{\xi_k!} e^{-\theta} = \ln \theta \sum_{k=1}^{n} \xi_k - n\theta + \ln h(X),$$

$$\widehat{\theta}_n = \arg\max \ln L(X, \theta) = \frac{1}{n} \sum_{k=1}^{n} \xi_k = \widehat{\mu}_n.$$

3. *Показникова вибірка* (зміщеність ОМВ):

$$\ln L(X, \theta) = \ln \prod_{k=1}^{n} \theta exp(-\theta \xi_k) = n \ln \theta - \theta \sum_{k=1}^{n} \xi_k,$$

$$\widehat{\theta}_n = \arg\max \ln L(X, \theta) = \frac{n}{\sum_{k=1}^{n} \xi_k} = \frac{1}{\widehat{\mu}_n}.$$

Зауважимо, що сума $\sum_{k=1}^{n} \xi_k$ має розподіл Ерланга з параметрами n, θ, тому

$$\left((\mu_{k+l} - \mu_k \mu_l),\ \ k, l = \overline{1, d}\right) = V^{(d)}.$$

Тому за вказаною теоремою

$$\sqrt{n}\left(\hat{\mu}_n^{(d)} - \mu^{(d)}\right) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \gamma_j = \eta_n \xrightarrow{W} \eta \simeq N_d\left(0, V^{(d)}\right)\ . \blacksquare$$

5.3. Незміщені оптимальні оцінки, оптимальність

У межах задачі оцінювання невідомого параметра θ важливу роль відіграє умова незміщеності, оскільки вона сприяє задовільній якості оцінки навіть при невеликих об’ємах вибірки.

Означення 5.4. Статистика $T = T(X)$ називається *незміщеною оцінкою* значення функції $\tau = \tau(\theta)$ від невідомого параметра $\theta \in \Theta$, якщо $dim\, T = dim\, \tau$ і

$$M_\theta T = \tau(\theta),\ \ \forall \theta \in \Theta.$$

Приклад 5.1. Для кратної вибірки $X = (\xi_1, \dots, \xi_n)$:

(а) вибіркова дисперсія $\hat{\sigma}_n^2$ є незміщеною оцінкою функції $\frac{n-1}{n}\sigma^2$ від невідомої дисперсії σ^2, а нормована вибіркова дисперсія $\hat{s}_n^2$ є незміщеною оцінкою σ^2,

(б) статистики ξ_1 та $(\xi_1 - \xi_2)^2/2$ є незміщеними оцінками середнього та дисперсії – хоча не зовсім ефективними.

Якщо існують незміщені оцінки, то їх можна порівнювати між собою за величиною середньоквадратичного відхилення від значення, яке оцінюється, та одночасно збігається із середнім цих оцінок.

Позначимо через

$$\Gamma_\tau = \{T = T(X) : M_\theta T = \tau(\theta),\qquad \forall \theta \in \Theta\}$$

клас усіх незміщених оцінок дійсної параметричної функції $\tau(\theta) : \Theta \to R$.

Означення 5.5. Статистика $T^* \in \Gamma_\tau$ називається *оптимальною незміщеною оцінкою* значення $\tau(\theta)$, або ж незміщеною оцінкою з рівномірно найменшою дисперсією, якщо вона має найменшу дисперсію у класі всіх незміщених оцінок:

$$D_\theta T^* \le D_\theta T,\qquad \forall T \in \Gamma_\tau,\qquad \forall \theta \in \Theta.$$

Теорема 5.5 (про єдність оптимальної оцінки). Якщо $\mathrm{T}_1, \mathrm{T}_2$ – дві оптимальні оцінки для значення $\tau = \tau(\theta)$, то $\mathrm{T}_1 = \mathrm{T}_2$ майже напевне.

Доведення. Позначимо $\sigma^2(\theta) = D_\theta T_1 = D_\theta T_2$. Розглянемо оцінку $T = (T_1 + T_2)/2$. Оскільки $M_\theta T = (M_\theta T_1 + M_\theta T_2)/2 = \tau$, то $\mathrm{T} \in \Gamma_\tau$, і за означенням $\sigma^2(\theta)$

$$\sigma^2(\theta) \le D_\theta T = M_\theta(T_1 - \tau + T_2 - \tau)^2/4 =$$
$$(D_\theta T_1 + D_\theta T_2 + 2Cov(\mathrm{T}_1, \mathrm{T}_2))/4 =$$
$$(2\sigma^2(\theta) + 2M(T_1 - \tau)(T_2 - \tau))/4 \le \sigma^2(\theta),$$

де остання нерівність випливає з нерівності Коші. Отже, за означенням оптимальної оцінки $\sigma^2(\theta) = D_\theta T$, і нерівність Коші має перетворитися на рівність. За зауваженням про рівність у нерівності Коші існує невипадкова стала b така, що $T_2 - \tau = b(T_1 - \tau)$ м.н. Тому

$$\sigma^2(\theta) = Cov(\mathrm{T}_1, \mathrm{T}_2) = bCov(\mathrm{T}_1, \mathrm{T}_1) = bD_\theta T_1 = b\sigma^2(\theta).$$

Звідси $b = 1$ та $\mathrm{T}_1 = \mathrm{T}_2$ м.н. ■

Приклад 5.2.

1. Оцінювання ймовірності успіху за одним спостереженням з геометричним розподілом. Нехай вибірка $X = \xi$ містить одну випадкову величи-ну, що має геометричний розподіл $G(1-\theta)$ із невідомою ймовірністю неуспіху $\theta \in (0{,}1)$. Для довільної статистики $T = T(X)$ умова незміщеності при оцінюванні функції $\tau = \tau(\theta)$ має вигляд

$$M_\theta T(X) = \sum\nolimits_{n \ge 1} (1-\theta)\theta^{n-1} T(n) = \tau(\theta), \forall \theta \in (0{,}1).$$

Нехай $\tau(\theta) = \theta$. Тоді з єдиності розкладу функції $\theta/(1-\theta)$ у ряд Тейлора виводимо, що існує єдина незміщена оцінка для θ, до того ж не дуже змістовна: $T(X) = 0$ при $X = 1$ та $T(X) = 1$ для інших X.

Якщо ж $\tau(\theta) = 1/\theta$, то незміщених оцінок взагалі не існує, оскільки функція $1/\theta(1-\theta)$ не припускає розклад у збіжний ряд Тейлора на відрізку $(0{,}1)$.

2. Оцінювання дисперсії у загальній нормальній моделі. Вибірка

РОЗДІЛ 8. Оцінки максимальної вірогідності

Будемо припускати, що вибірка X задовольняє умову підпорядкованості її розподілу деякій мірі у вибірковому просторі. За такої умови повністю визначена вибіркова функція вірогідності $L(X,\theta)$. Значення цієї функції і дають критерій "найбільшої ймовірності".

Означення 8.1. *Оцінкою максимальної вірогідності* (ОМВ, *EML*) параметру θ за вибіркою X називається статистика

$$\widehat{\theta} = \widehat{\theta}(X) \equiv \arg\max_{\theta \in \Theta} L(X,\theta),$$

де $L(X,\theta)$ – вибіркова функція вірогідності, тобто така статистика $\widehat{\theta} = \widehat{\theta}(X)$, що

$$L(X,\theta) \le L(X, \widehat{\theta}), \forall \theta \in \Theta.$$

У випадку кратної вибірки ОМВ позначається як θ_n, де n – об'єм вибірки.

Необхідною умовою існування ОМВ є припущення існування точки максимуму. ОМВ визначена однозначно за умови єдиності цієї точки.

Теорема 8.1 (про рівняння максимальної вірогідності). Якщо $\Theta \subset \mathbb{R}^d$, максимум в означенні оцінки максимальної вірогідності досягається всередині параметричної множини, а функція вірогідності диференційовна, то ОМВ задовольняє рівняння максимальної вірогідності.

$$\frac{\partial}{\partial\theta} \ln L(X,\theta)|_{\theta = \widehat{\theta}} = 0,$$

тобто ОМВ є коренем функції впливу $U(X,\theta)$:

$$U(X, \widehat{\theta}) = 0.$$

У випадку векторного параметра рівняння максимуму вірогідності є векторним, і перетворюється на систему рівнянь максимальної вірогідності для кожної координати вектора- градієнта.

Отже, $D_\theta g(T) \leq D_\theta T_0,\ \forall\theta \in \Theta,\ \forall\, T_0 \in \Gamma_\tau$, тобто $g(T)$ - оптимальна незміщена оцінка для $\tau(\theta)$. ■

Приклад 7.1. Рівномірна на $[0, \theta]$ вибірка (існування суперефективної оптимальної оцінки в нерегулярній моделі).

Нехай X – n-кратна вибірка з рівномірним розподілом $\xi_k \simeq U(0, \theta), k = \overline{1, n}, \theta \in \Theta = \mathbb{R}_+$. Як зроблено вище в прикладі достатніх статистик, перевіряємо, що в даному випадку статистика $T(X) = \max\ \xi_k$ є достатньою. Ця статистика є повною, оскільки має функцію розподілу

$$P_\theta(T < x) = (x/\theta)^n, 0 \leq x \leq \theta,$$

та з рівностей

$$M_\theta \varphi(T) = \theta^{-n} \int_0^\theta \varphi(x) n x^{n-1} dx = 0,\ \forall\theta > 0,$$

випливає, що $\varphi(x) = 0$ майже для всіх $x \geq 0$ та $\varphi(T) = 0$ м.н.

Отже, статистика $T(X) = \max\ \xi_k$ є оптимальною незміщеною оцінкою свого математичного сподівання $M_\theta T(X) = \tau(\theta) = \frac{n\theta}{n+1}$. Слід зазначити, що дисперсія цієї оцінки дорівнює

$$D_\theta T = \theta^2 n/(n+2)\ (n+1)^2) = O(1/n^2), n \to \infty,$$

і має суттєво вищий порядок малості, ніж гарантує нерівність Крамера - Рао. Цей факт пов'язаний з тим, що в даній схемі не виконується умова регулярності, тому критерій Крамера-Рао не може бути застосований.

$X = (\xi_1, \dots, \xi_n)$ складається з незалежних у сукупності величин із нормальним розподілом $\xi_k \simeq N(\mu, \sigma^2)$. Як було з'ясовано в теоремі про моменти вибіркових моментів, нормована вибіркова дисперсія

$$\hat{s}_n^2 = \frac{1}{n-1} \sum\nolimits_{i=1}^{n} (\xi_i - \hat{\mu}_n)^2 = \frac{n}{n-1} \hat{\sigma}_n^2$$

є незміщеною оцінкою невідомої дисперсії σ^2. З тієї ж теореми виводимо, що відповідне середньоквадратична похибка дорівнює

$$D_\theta \hat{s}_n^2 = \left(\frac{n}{n-1}\right)^2 D_\theta \hat{\sigma}_n^2 = \frac{1}{n(n-1)} \left((n-1)\mu_4^0 - (n-3)\sigma^4\right) =$$

$$\frac{1}{n(n-1)} \left((n-1)3\sigma^4 - (n-3)\sigma^4\right) = \frac{2\sigma^4}{n-1},$$

де враховане співвідношення $\mu_4^0 = 3\sigma^4$ для моментів нормального розподілу.

Одночасно розглянемо зміщену оцінку для σ^2 вигляду $\hat{\tau}_n^2 = t\hat{s}_n^2$, де t - деяка стала. Обчислимо середньоквадратичну похибку

$$M_\theta(\hat{\tau}_n^2 - \sigma^2)^2 = M_\theta(t(\hat{s}_n^2 - \sigma^2) - (1-t)\sigma^2)^2 =$$

$$t^2 D_\theta \hat{s}_n^2 + (1-t)^2\sigma^4 = (2t^2/(n-1) + (1-t)^2)\sigma^4.$$

Вираз у правій частині набуває найменшого значення при $t = \frac{n-1}{n+1} \neq 1$. Отже, незміщена оцінка $\hat{s}_n^2$ з погляду середньоквадратичного відхилення від оцінюваного значення σ^2 гірша за зміщену оцінку $\hat{\tau}_n^2$. Це і не дивно – адже мінімум у класі всіх оцінок часто є меншим від мінімуму в певному підкласі.

РОЗДІЛ 6. Нерівність Крамера-Рао, ефективність

Розглянемо кратну вибірку $X = (\xi_1, \dots, \xi_n)$ з функцією вірогідності

$$L(x,\theta) = \prod_{k=1}^{n} f(x_k,\theta), \qquad x = (x_1, \dots, x_n) \in S, \theta \in \Theta,$$

де $f(x,\theta)$ – щільність одного спостереження.

Припускатимемо, що параметричний простір евклідів: $\Theta \subset \mathbb{R}^d$, а функція вірогідності $L(x,\theta)$ – диференційовна за параметром θ.

6.1. Функція впливу

Означення 6.1. *Функцією впливу*, або *функцією внеску* вибірки X називається частинна похідна за параметром θ від логарифма вибіркової функції вірогідності:

$$U(X,\theta) \equiv \frac{\partial}{\partial\theta} lnL(X,\theta).$$

У випадку кратної вибірки функцією впливу спостереження ξ називається похідна за θ від логарифма функції вірогідності спостереження:

$$u(\xi,\theta) \equiv \frac{\partial}{\partial\theta} lnf(\xi,\theta).$$

Зауваження 6.1. Якщо параметр $\theta \in \Theta \subset \mathbb{R}$ скалярний, то функція впливу є дійсною функцією від випадкової величини, тобто випадковою величиною. Якщо ж $\theta \in \Theta \subset \mathbb{R}^d$, то похідну слід розуміти як вектор, що складений із частинних похідних (градієнт), а функція впливу буде d-вимірним випадко-вим вектором:

$$U(X,\theta) \equiv \left(U_k(X,\theta), k = \overline{1,d}\right), \qquad U_k(X,\theta) \equiv \frac{\partial}{\partial\theta_k} lnL(X,\theta).$$

Важливе значення та назва "функція впливу" пояснюється таким міркуванням. Якщо припустити, що розподіл вибірки взагалі не залежить від невідомого параметра θ (параметр не впливає на розподіл), то відповідна функція впливу буде тотожнім нулем. Одночасно зауважимо, що якраз у цьому випадку марно

Отже, $D_\theta T_0 \ge D_\theta T_1$ для довільної незміщеної оцінки T_0, що й доводить теорему. ■

Теорема 7.3 (теорема Рао-Блекуела-Колмогорова). Якщо існує оптимальна оцінка $T^* \in \Gamma_\tau$, а $T = T(x)$ - деяка достатня статистика, то $T^* = H^*(T)$ м.н. є вимірною функцією від цієї статистики.

Доведення. Оберемо в теоремі про дисперсію функції від достатньої статистики незміщену оцінку $T_0 = T^*$, звідки $D_\theta T^* \ge D_\theta H(T)$ для вимірної функції H, що побудована в цій теоремі, причому $H(T) \in \Gamma_\tau$. Однак за означенням оптимальної оцінки $D_\theta H(T) \ge D_\theta T^*$. Отже, обидві нерівності є рівностями для всіх θ і кожна з оцінок $T^*, H(T)$ є оптимальною. З теореми про єдиність оптимальної оцінки виводимо, що $T^* = H^*(T)$ м.н. для функції $H^* \equiv H$. ■

Означення 7.2. Достатня статистика $T = T(X)$ називається *повною* статистикою, якщо для кожної вимірної функції $\varphi(\cdot)$ із тотожностей

$$M_\theta \varphi(T) = 0, \qquad \forall\theta \in \Theta,$$

випливає, що $\varphi(T) = 0$ майже напевне.

Теорема 7.4 (про оптимальність повної достатньої статистики). Якщо існує повна достатня статистика, то довільна вимірна функція від неї є оптимальною оцінкою свого математичного сподівання.

Доведення. Нехай $T = T(X)$ - повна достатня статистика, а g- вимірна функція. Визначимо функцію $\tau(\theta) = M_\theta g\big(T(X)\big)$. Якщо $T_0 \in \Gamma_\tau$ - довільна незміщена оцінка для значення $\tau = \tau(\theta)$, то внаслідок теореми про дисперсію функції від достатньої статистики знайдеться вимірна функція H така, що статистика $T_1 \equiv H(T)$ знову є незміщеною, причому $D_\theta T_1 \le D_\theta T_0$, $\forall\theta \in \Theta$. За умовою незміщеності $M_\theta T_1 = M_\theta H(T) = \tau(\theta)$, де за означенням $\tau(\theta) = M_\theta g(T)$. Звідси $M_\theta\big(H(T) - g(T)\big) = 0, \forall\theta \in \Theta$. Тому з повноти T випливає $H(T) - g(T) = 0$ м.н., тобто $T_1 \equiv H(T) = g(T)$ м.н., причому функція g вже не залежить від вибору T_0.

$$H(t) = M_\theta(T_0(X)|T(X)=t) \equiv \frac{M_\theta\big(T_0(X)1_{\{T(X)=t\}}\big)}{P_\theta(T(X)=t)} =$$

$$\sum_{x\in S} T_0(x)\, P_\theta(X=x|T(X)=t) = \sum_{x\in S} T_0(x)\, l(x,t),$$

яка за теоремою про властивість достатньої статистики не залежить від параметра θ.

Визначимо статистику

$$T_1(X) = H\big(T(X)\big).$$

Вона є незміщеною, оскільки за формулою повної ймовірності та за означенням функцій $H(t)$ і $l(x,t)$

$$M_\theta T_1(X) = M_\theta H\big(T(X)\big) = \sum_t H(t) P_\theta(T(X)=t) =$$

$$\sum_t \left(\sum_{x\in S} T_0(x)\, l(x,t)\right) P_\theta(T(X)=t) =$$

$$\sum_{x\in S} T_0(x) \sum_t P_\theta(X=x|T(X)=t) P_\theta(T(X)=t) =$$

$$\sum_{x\in S} T_0(x) P_\theta(X=x) = M_\theta T_0(X) = \tau(\theta), \forall\theta\in\Theta.$$

Далі, обчислимо дисперсію

$$D_\theta T_0 = M_\theta\big(T_0 - H(T) + H(T) - \tau(\theta)\big)^2 =$$

$$M_\theta\big(T_0 - H(T)\big)^2 + D_\theta H(T) \ge D_\theta H(T) = D_\theta T_1,$$

оскільки за означенням $H(T)$ математичне сподівання добутку дорівнює нулю:

$$M_\theta\big(T_0 - H(T)\big)\big(H(T)-\tau(\theta)\big) = \sum_t M_\theta\big(T_0 - H(T)\big)\big(H(T)-\tau(\theta)\big)1_{\{T=t\}} =$$

$$\sum_t M_\theta\big(T_0 - H(t)\big)\big(H(t)-\tau(\theta)\big)1_{\{T=t\}} =$$

$$\sum_t \Big(M_\theta T_0 1_{\{T=t\}} - H(t)P_\theta(T=t)\Big)\Big(H(t)-\tau(\theta)\Big) =$$

$$\sum_t \big(M_\theta(T_0/T=t) - H(t)\big)P_\theta(T=t)\big(H(t)-\tau(\theta)\big) = 0.$$

намагатись оцінити параметр на основі спостережень, чий розподіл не залежить від значення такого параметра. Тому відносно "малі" значення функції впливу вказують на брак інформації щодо параметра в наявних спостереженнях.

Теорема 6.1 (про функцію впливу кратної вибірки). Якщо $X = (\xi_1, \dots, \xi_n)$ - кратна вибірка, то вибіркова функція впливу дорівнює сумі функцій впливу спостережень, які її утворюють:

$$U(X,\theta) = \sum\nolimits_{k=1}^{n} u(\xi_k,\theta).$$

Доведення випливає з теореми про функцію вірогідності кратної вибірки та лінійності частинної похідної:

$$\frac{\partial}{\partial\theta} lnL(X,\theta) = \frac{\partial}{\partial\theta} ln \prod\nolimits_{k=1}^{n} f(\xi_k,\theta) \; =$$

$$\tfrac{\partial}{\partial\theta}\textstyle\sum_{k=1}^{n} lnf(\xi_k,\theta) = \sum_{k=1}^{n} \tfrac{\partial}{\partial\theta} lnf(\xi_k,\theta) = \sum_{k=1}^{n} u(\xi_k,\theta). \blacksquare$$

Приклади

1. *Пуассонівська вибірка*. Для кратної вибірки з розподілом Пуассона $P(\lambda)$ та невідомим параметром $\theta = \lambda$

$$u(y,\theta) = \frac{\partial}{\partial\theta} \ln(\theta^y \exp(-\theta/y!)) = y/\theta - 1, y \in \mathbb{Z}_+,$$

$$U(X,\theta) = \sum\nolimits_{k=1}^{n} (\xi_k/\theta - 1) = n(\hat\mu_n/\theta - 1).$$

2. *Вибірка з гама-розподілу*. Для кратної вибірки з гама-розподілом $\Gamma(\lambda,\alpha)$ та невідомим параметром $\theta = (\lambda,\alpha)$

$$lnf(y,\theta) = \alpha ln\lambda + (\alpha-1)lny - \lambda y - ln\Gamma(\alpha),$$

$$u_1(y,\theta) = \alpha/\lambda - y,$$

$$u_2(y,\theta) = ln\lambda + lny - \psi(\alpha), \qquad \psi(\alpha) = \frac{\partial}{\partial\alpha}\, ln\Gamma(\alpha),$$

$$U_1(X,\theta) = n(\alpha/\lambda - \hat\mu_n),$$

$$U_2(X,\theta) = n\; ln\lambda + ln\prod\nolimits_{k=1}^{n}\xi_k - n\psi(\alpha).$$

У даному розділі будемо припускати, що виконані певні ***умови регулярності*** на функцію вірогідності вибірки. Ці умови не будемо точно деталізувати, однак позначимо їх.

1. Множина тих значень вибірки X, для яких функція вірогідності $L(X,\theta)$ строго додатна, не залежить від θ.
2. Функція вірогідності $L(X,\theta)$ двічі неперервно диференційована за параметром θ.
3. Функція впливу $U(X,\theta)$ інтегровна у квадраті: $M_\theta U^2(X,\theta) < \infty\ \forall\theta$.
4. Знак похідної за параметром θ можна внести під знак інтегралів $M_\theta L(X,\theta)$ за аргументом $x \in S$ від функції вірогідності $L(x,\theta)$, а також від її похідних за θ.

6.2. Властивості функції впливу, інформація за Фішером

Теорема 6.2 (про центрованість функції впливу). За умов регулярності функція впливу центрована:

$$M_\theta U(X,\theta) = 0, \qquad \forall\theta \in \Theta$$

Доведення. За означенням функції вірогідності

$$P_\theta(X \in S) = 1 = \int_S L(x,\theta)\lambda(dx),$$

оскільки вибірка X набуває значень у вибірковому просторі S, а $L(x,\theta)$ – її щільність розподілу відносно міри λ.

Позначимо $S_0 = \{x \in S: L(x,\theta) > 0\}$ вимірну підмножину вибіркового простору $(S,\sum)$, яка не залежить від θ за умовами регулярності. Використо-вуючи ці умови та властивості інтегралу Лебега від нульової функції, з наведеної вище тотожності дістанемо

$$0 = \frac{\partial}{\partial\theta}\int_S L(x,\theta)\lambda(dx) = \frac{\partial}{\partial\theta}\int_{S_0} L(x,\theta)\lambda(dx) =$$

Пара $(\hat\mu_n, \hat\mu_{2n})$ утворює достатню статистику, так само як і $(\hat\mu_n, \hat\sigma_n^2)$ - оскільки між цими парами статистик є ізоморфізм.

7.2. Умовний розподіл вибірки

Теорема 7.1 (про властивість достатньої статистики). Якщо $T(X)$ - достатня статистика, то умовний розподіл

$$P_\theta(X = x|T(X) = t) = l(x,t)$$

не залежить від θ.

Доведення проведемо в припущенні, що вибірка є дискретним випадковим вектором. Зауважимо, що дана умовна ймовірність дорівнює нулю (і не залежить від θ), якщо $T(X) \neq t$. Отже, можна припустити, що $T(X) = t$. За означенням умовної ймовірності

$$P_\theta(X = x|T(X) = t) = \frac{P_\theta(X = x, T(X) = t)}{P_\theta(T(X) = t)} =$$

$$\frac{P_\theta(X = x, T(x) = t)}{P_\theta(T(X) = t)} = L(X,\theta)\Big/\sum_{y:T(y)=t} L(y,\theta) =$$

$$g(T(x),\theta)\,h(x)\,j(\theta)\Big/\sum_{y:T(y)=t} g(T(y),\theta)\,h(y)\,j(\theta) =$$

$$g(t,\theta)\,h(x)\Big/\sum_{y:T(y)=t} g(t,\theta)\,h(y) = h(x)\Big/\sum_{y:T(y)=t} h(y) = l(x,t). \blacksquare$$

7.3. Достатність та оптимальність

Теорема 7.2 (про дисперсію функції від достатньої статистики). Нехай $T = T(X)$ - достатня статистика, а $T_0 \in \Gamma_\tau$ - довільна незміщена оцінка функції $\tau(\theta)$. Тоді знайдеться функція $T_1(X) = H\big(T(X)\big)$ від достатньої статистики $T(X)$, яка також є незміщеною та має не більшу дисперсію, ніж T_0, при кожному $\theta \in \Theta$.

Доведення. Нехай $T_0 \in \Gamma_\tau$ - довільна незміщена оцінка. Розглянемо функцію

РОЗДІЛ 7. Достатні статистики та оптимальність

Означення 7.1. Статистика $T = T(X)$ називається *достатньою статистикою* для вибірки X, якщо вибіркова функція вірогідності має вигляд

$$L(X,\theta) = g(T(X),\theta)\ h(X)\ j(\theta)$$

для деяких вимірних невід'ємних функцій g, h, j.

Зауважимо, що на розмірність достатньої статистики обмеження не накладаються, однак при виборі цієї статистики доцільно мінімізувати її розмірність. В усякому випадку, достатні статистики існують: наприклад, статистика $T(X) = X$ завжди є достатньою, якщо обрати $g = L, h \equiv 1, j \equiv 1$.

7.1. Приклади достатніх статистик

1. *Пуассонівська вибірка*. Якщо X - n-кратна вибірка з розподілом Пуассона, $\xi_1 \simeq \Pi(\lambda)$, то

$$\ln L(X,\theta) = \ln \prod_{k=1}^{n} \frac{\lambda^{\xi_k}}{\xi_k!} \exp(-\lambda) =$$

$$\ln(\lambda) \sum_{k=1}^{n} \xi_k - \sum_{k=1}^{n} \ln(\xi_k!) - n\lambda,$$

отже існує скалярна достатня статистика $T(X) = \hat{\mu}_n \equiv \frac{1}{n}\sum_{k=1}^{n} \xi_k$.

2. *Рівномірна вибірка*. Якщо X - n -кратна вибірка з $\xi_1 \simeq U(a,b)$, то

$$L(X,\theta) = (b-a)^{-n} 1_{\left\{a<\min \xi_k, \max \xi_k < b\right\}}.$$

Отже, достатньою є двовимірна статистика

$$T(X) = \left(\min_{1\le k\le n} \xi_k\ , \max_{1\le k\le n} \xi_k\right) = \left(\xi_{(1)}, \xi_{(n)}\right)$$

3. *Нормальна вибірка*. Якщо X - n -кратна вибірка з $\xi_1 \simeq N(\mu,\sigma^2)$, то

$$\ln L(X,\theta) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{n}{2\sigma^2}(\hat{\mu}_{2n} - 2\hat{\mu}_n\mu + \mu^2).$$

$$\int_{S_0} \frac{\partial}{\partial\theta} L(x,\theta)\lambda(dx) = \int_{S_0} \left(\frac{\partial}{\partial\theta} lnL(x,\theta)\right) L(x,\theta)\lambda(dx) =$$

$$\int_{S} \left(\frac{\partial}{\partial\theta} lnL(x,\theta)\right) L(x,\theta)\lambda(dx) = M_\theta U(X,\theta),$$

де остання рівність випливає з теореми про обчислення математичного сподівання функції від неперервної величини, і використовується очевидна тотожність

$$\frac{\partial}{\partial\theta} L(x,\theta) = \left(\frac{\partial}{\partial\theta} lnL(x,\theta)\right) L(x,\theta)$$

на множині S_0, де $L(x,\theta) > 0$.■

Означення 6.2. Нехай параметр $\theta \in \Theta \subset \mathbb{R}$ - скалярний. *Інформацією за Фішером* у вибірці X називається функція

$$I(\theta) \equiv D_\theta U(X,\theta) = M_\theta U^2(X,\theta).$$

Друга рівність в означенні випливає з теореми про центрованість функції впливу та з властивості дисперсії.

Теорема 6.3 (про обчислення інформації за Фішером). За умов регулярності справедлива тотожність

$$I(\theta) = -M_\theta \frac{\partial^2}{\partial\theta^2} lnL(x,\theta) = -M_\theta \frac{\partial}{\partial\theta} U(X,\theta), \qquad \forall\theta \in \Theta.$$

Доведення. Продиференціюємо тотожність із теореми про центрованість функції впливу та використаємо умови регулярності:

$$0 = \frac{\partial}{\partial\theta} M_\theta U(X,\theta) = \frac{\partial}{\partial\theta} \int_S U(x,\theta) L(x,\theta)\lambda(dx) =$$

$$\frac{\partial}{\partial\theta} \int_{S_0} U(x,\theta) L(x,\theta)\lambda(dx) = \int_{S_0} \frac{\partial}{\partial\theta} (U(x,\theta) L(x,\theta))\lambda(dx) =$$

$$\int_{S_0} \left(\frac{\partial}{\partial\theta} U(x,\theta)\right) L(x,\theta)\lambda(dx) + \int_{S_0} U(x,\theta) \frac{\partial}{\partial\theta} L(x,\theta)\lambda(dx) =$$

$$\int_{S} \left(\frac{\partial}{\partial\theta} U(x,\theta)\right) L(x,\theta)\lambda(dx) +$$

$$\int_{S_0} U(x,\theta)\left(\frac{\partial}{\partial\theta} lnL(x,\theta)\right) L(x,\theta)\lambda(dx) =$$

$$M_\theta \frac{\partial}{\partial\theta} U(X,\theta) + M_\theta U^2(X,\theta) = M_\theta \frac{\partial^2}{\partial\theta^2} lnL(x,\theta) + I(\theta),$$

де множина $S_0 = \{x \in S: L(x,\theta) > 0\}$. ■

Теорема 6.4 (про адитивність інформації за Фішером). Якщо $X = (\xi_1, \dots, \xi_n)$ – кратна вибірка, то

$$I(\theta) = ni(\theta),$$

де

$$i(\theta) = D_\theta u(\xi_1, \theta)$$

інформація за Фішером в одному спостереженні.

Доведення. За теоремою про функцію впливу кратної вибірки та за теоремою про дисперсію суми незалежних величин

$$I(\theta) = D_\theta U(X,\theta) = D_\theta \sum\nolimits_{k=1}^{n} u(\xi_k,\theta) = \sum\nolimits_{k=1}^{n} D_\theta u(\xi_k,\theta) = ni(\theta),$$

де незалежність величин $u(\xi_k,\theta)$ випливає з теореми про перетворення незалежних величин і використано однакову розподіленість цих величин. ■

Приклади.

1) *Схема Бернуллі*. Нехай вибірка $X = (\chi_1,\dots,\chi_n)$ містить результати випробувань Бернуллі з невідомою ймовірністю успіху θ, де χ_k - індикатор успіху в k-му випробуванні. Функція вірогідності одного спостереження є щільністю відносно точкової міри і має вигляд

$$f(y,\theta) = \theta 1_{y=1} + (1-\theta)1_{y=0} = \theta^y(1-\theta)^{1-y}, \qquad y \in \{0,1\},$$

теоретична функція впливу дорівнює

$$u(y,\theta) = \frac{\partial}{\partial\theta} ln\, f(y,\theta) = \frac{y}{\theta} - \frac{1-y}{1-\theta} = \frac{y-\theta}{\theta(1-\theta)},$$

функції вірогідності та впливу всієї вибірки

$$L(X,\theta) = \prod\nolimits_{k=1}^{n} \theta^{\chi_k}(1-\theta)^{1-\chi_k} = \theta^{\nu_n}(1-\theta)^{n-\nu_n}, \qquad \nu_n = \sum\nolimits_{k=1}^{n} \chi_k,$$

$$U_1(X,\theta) = \frac{n}{\sigma^2}(\hat{\mu}_n - \mu),$$

$$U_2(X,\theta) = -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}(\hat{\mu}_{2n} - 2\hat{\mu}_n\mu + \mu^2).$$

Оберемо векторну сталу $c = (c_1, c_2) = (\sigma^2/n, 0)$. Тоді у критерії (б) нерівності Крамера-Рао для векторного параметра

$$T^* - \mu = c_1 U_1(X,\theta) + c_2 U_2(X,\theta) = \hat{\mu}_n - \mu$$

має місце рівність для $T^* = \hat{\mu}_n$, тобто вибіркове середнє є оптимальною оцінкою і при невідомій дисперсії.

$$lnL(X,\theta) = -\frac{n}{2}ln2\pi - \frac{n}{2}ln\sigma^2 - \frac{1}{2\sigma^2}\sum\nolimits_{k=1}^{n}(\xi_k-\mu)^2 =$$

$$= -\frac{n}{2}ln2\pi - \frac{n}{2}ln\sigma^2 - \frac{n}{2\sigma^2}(\hat{\mu}_{2n} - 2\hat{\mu}_n\mu + \mu^2).$$

де $\hat{\mu}_n$ – вибіркове середнє, $\hat{\mu}_{2n}$ – другий нецентральний вибірковий момент. Звідси знаходимо функцію впливу

$$U(X,\theta) = \frac{n}{\sigma^2}(\hat{\mu}_n - \mu).$$

За критерієм ефективності Крамера-Рао оптимальна оцінка μ повинна задо-вольняти рівняння

$$T^* - \mu = c\frac{n}{\sigma^2}(\hat{\mu}_n - \mu) = \hat{\mu}_n - \mu,$$

що має місце при виборі $c = \sigma^2/n$. Отже, вибіркове середнє $T^* = \hat{\mu}_n$ є оптимальною оцінкою середнього при відомій дисперсії.

Оцінювання дисперсії при відомому середньому

Розглянемо таку ж нормальну кратну вибірку $X = (\xi_1, \dots, \xi_n)$, $\xi_k \simeq N(\mu, \sigma^2)$, де відоме середнє μ, а дисперсія $\sigma^2 = \theta$ оцінюється. У цьому випадку функція впливу має вигляд

$$U(X,\theta) = -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}(\hat{\mu}_{2n} - 2\hat{\mu}_n\mu + \mu^2).$$

Вибором множника $c = 2\sigma^4/n$ забезпечується рівність $T - \sigma^2 = cU$ у критерії ефективності Крамера-Рао, а оптимальна оцінка дорівнює вибірко-вій дисперсії з урахуванням відомого середнього

$$T^* = \tilde{\sigma}_n^2 = \hat{\mu}_{2n} - 2\hat{\mu}_n\mu + \mu^2 = \frac{1}{n}\sum\nolimits_{k=1}^{n}(\xi_k-\mu)^2.$$

Оцінювання середнього при невідомій дисперсії

Розглянемо нормальну кратну вибірку $X = (\xi_1, \dots, \xi_n)$, $\xi_k \simeq N(\mu, \sigma^2)$, де невідомі як μ, так і σ^2, параметр $\theta = (\mu, \sigma^2)$, а оцінюється середнє $\mu = \tau(\theta)$. У цьому випадку координати векторної функції впливу мають вигляд

$$U(X,\theta) = \sum\nolimits_{k=1}^{n} u(\chi_k,\theta) = \frac{n}{\theta(1-\theta)}\left(\hat{\theta}_n - \theta\right), \qquad \hat{\theta}_n = \frac{\nu_n}{n},$$

функції інформації за Фішером мають вигляд

$$i(\theta) = M_\theta\left(\frac{\chi_1-\theta}{\theta(1-\theta)}\right)^2 = \frac{D_\theta\chi_1}{\theta^2(1-\theta)^2} = \frac{1}{\theta(1-\theta)}, \qquad I(\theta) = \frac{n}{\theta(1-\theta)}.$$

2) *Показниковий розподіл*. Розглянемо кратну вибірку $X = (\xi_1, \dots, \xi_n)$ із показниковим розподілом спостережень $\xi_k \simeq Exp(\theta)$. Функції впливу мають вигляд

$$u(y,\theta) = \frac{\partial}{\partial\theta}(ln\theta - \theta y) = 1/\theta - y,$$

$$U(X,\theta) = \sum\nolimits_{k=1}^{n} u(\xi_k,\theta) = n/\theta - n\hat{\mu}_n,$$

а функції інформації за Фішером у спостереженні та у вибірці дорівнюють

$$i(\theta) = M_\theta(1/\theta - \xi_1)^2 = D_\theta\xi_1 = 1/\theta^2, \qquad I(\theta) = n/\theta^2 \ .$$

6.3. Нерівність Крамера-Рао

Розглянемо задачу оцінювання значення дійсної параметричної функції $\tau(\theta)$ у класі Γ_τ незміщених її оцінок.

Теорема 6.5 (про нерівність та критерій Крамера-Рао). Нехай $\Theta \subset \mathbb{R}$ і виконані умови регулярності.

(а) Для довільної незміщеної оцінки $T = T(X) \in \Gamma_\tau$ при всіх $\theta \in \Theta$ справедлива нерівність Крамера-Рао

$$M_\theta(T-\tau)^2 \equiv D_\theta T \geq \frac{\tau_\theta^2(\theta)}{I(\theta)},$$

де $\tau_\theta(\theta) = \frac{d}{d\theta}\tau(\theta)$, $I(\theta)$ – інформація за Фішером у вибірці X.

(б) Рівність у нерівності (а) має місце тоді й тільки тоді, коли оцінка T є лінійною функцією від функції впливу вибірки:

$$T(X) - \tau(\theta) = c(\theta)U(X,\theta) \text{ м.н.}, \forall\theta \in \Theta,$$

для деякої дійсної $c(\theta)$, причому у випадку рівності

$$c(\theta) = \tau_\theta(\theta)/I(\theta).$$

Означення 6.4. Оцінка $T \in \Gamma_\tau$ називається *ефективною оцінкою* параметричної функції τ_θ, якщо нерівність Крамера-Рао для неї є рівністю, тобто у випадку, коли ця оцінка має найменше можливе значення дисперсії у класі Γ_τ всіх незміщених оцінок.

Твердження (б) дає критерій ефективності Крамера-Рао.

Зауваження 6.2. З означення та єдності оптимальної оцінки випливає, що ефективна оцінка є оптимальною оцінкою свого математичного сподівання, тобто критерій Крамера-Рао є достатньою умовою оптимальності для регулярних спостережень.

Зауваження 6.3. Нерівність Крамера-Рао для дисперсії $D_\theta T$ має місце також і для зміщених оцінок T, якщо в правій частині τ_θ^2 замінити на $(\tau_\theta - b_\theta)^2$, де $b_\theta = \frac{d}{d\theta} b_\theta, b_\theta = M_\theta T - \tau_\theta$. Дійсно, у даному випадку оцінка T є незміщеною для значення $\tau + b$.

Доведення теореми.

За означенням незміщеності та за теоремою про обчислення математичного сподівання функції від випадкової величини

$$M_\theta T(X) = \int_S T(x)L(x,\theta)\mu(dx) = \tau(\theta), \forall\theta \in \Theta.$$

За умови регулярності з урахуванням означення функції впливу та теореми про обчислення математичного сподівання функції від випадкової величини знайдемо

$$\tau_\theta(\theta) = \frac{\partial}{\partial\theta}\int_S T(x)L(x,\theta)\mu(dx) = \int_S T(x)\frac{\partial}{\partial\theta}L(x,\theta)\mu(dx) =$$

$$\int_S T(x)U(x,\theta)L(x,\theta)\mu(dx) =$$

$$M_\theta T(X)U(X,\theta) = M_\theta\big(T(X) - \tau(\theta)\big)U(X,\theta),$$

де в останній рівності використана теорема про центрованість функції впливу.

Згідно з нерівністю Коші з отриманої тотожності виводимо

$$\tau_\theta^2(\theta) = (M_\theta\big(T(X) - \tau(\theta)\big)U(X,\theta))^2 \le$$

$$M_\theta\big(T(X) - \tau(\theta)\big)^2 M_\theta U^2(X,\theta) = D_\theta T \cdot I(\theta),$$

за означенням $I(\theta)$. Це доводить твердження (а).

Для того, щоб нерівність Коші перетворювалася на рівність, необхідно й достатньо, щоб множники під знаком математичного сподівання були лінійно пов'язані майже напевне при кожному θ, тобто $T(X) - \tau(\theta) = c(\theta)U(X,\theta)$ м.н.. Звідси виводимо твердження (б) теореми. Вираз для значення відповідної сталої знаходимо після підстановки тотожності критерію рівності в отриману вище тотожність:

$$\tau_\theta(\theta) = M_\theta\big(T(X) - \tau(\theta)\big)U(X,\theta) = M_\theta c(\theta)U^2(X,\theta) = c(\theta)I(\theta). \blacksquare$$

Зауваження 6.4. У випадку, коли $X = (\xi_1, \dots, \xi_n)$ – кратна вибірка, інформація за Фішером у знаменнику дорівнює $I(\theta) = ni(\theta)$, отже нерівність Крамера-Рао набуває вигляду

$$D_\theta\left(\sqrt{n}(T - \tau)\right) = nD_\theta T \ge \frac{\tau_\theta^2(\theta)}{i(\theta)} \equiv \tilde{\sigma}_{opt}^2.$$

Означення 6.5. Якщо послідовність оцінок $T_n \in \Gamma_\tau$ є асимптотично нормальною з асимптотичною дисперсією σ_a^2, тобто

$$\sqrt{n}(T_n - \tau) \overset{w}{\to} \eta \simeq N(0, \sigma_a^2),$$

причому $\sigma_a^2 = \tilde{\sigma}_{opt}^2$ збігається з найменшим можливим за нерівністю Крамера-Рао значенням, то оцінка (T_n) називається *асимптотично ефективною*.

6.4. Приклад оцінювання параметрів нормальних спостережень

Оцінювання середнього при відомій дисперсії

Нехай $X = (\xi_1, \dots, \xi_n)$ – кратна вибірка з нормальним розподілом спостережень: $\xi_k \simeq N(\mu, \sigma^2)$, де дисперсія σ^2 відома, а середнє $\mu = \theta$ треба оцінити. Логарифмічна функція вірогідності має вигляд