Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia, 2018, Iss. 75, pp. 40–47

УДК 519.2:681.2

Поліноміальні оцінки параметрів для даних з експоненційним степеневим розподілом

Заболотній С. В., Чепинога А. В., Бондаренко Ю. Ю., Рудь М. П.

Черкаський державний технологічний університет

E-mail: *maxxium23@gmail.com*

В роботі запропонований оригінальний підхід до знаходження оцінок результатів багаторазових вимірювань при випадкових похибках, що описується моделлю експоненційного степеневого (узагальненого гаусового) розподілу. В основі даного підходу лежить метод максимізації поліному (ММПл), який базується на математичному апараті стохастичних поліномів Кунченка та описі випадкових величин статистиками вищих порядків (моментами або кумулянтами). Приведено вирази для знаходження поліноміальних оцінок з використанням аналітичних (методом Кардано) і чисельних (метод Ньютона-Рафсона) розв'язків. Показано, що при степені стохастичного полінома $r \leqslant 2$ поліноміальні оцінки вироджуються в лінійні оцінки середнього арифметичного. При використані поліномів ступеня $r = 3$ відносна точність поліноміальної оцінки збільшується. Коефіцієнт зменшення дисперсії оцінок залежить від величини кумулянтних коефіцієнтів 4-го і 6-го порядку, які характеризують ступінь відмінності від гаусової моделі. Шляхом багаторазових статистичних випробувань (методом Монте-Карло) досліджені властивості нормалізації поліноміальних оцінок і проведено порівняльний аналіз їх точності з відомими оцінками (середнім, медіаною і серединою розмаху). Побудовано області ефективності для кожного із методів в залежності від параметра форми експоненційного степеневого розподілу і обсягу вибірки.

Ключові слова: експоненційний степеневий розподіл; стохастичні поліноми; статистики вищих порядків; оцінка параметра

DOI: 10.20535/RADAP.2018.75.40-47

Вступ

Методи статистичного оцінювання є одним із базових математичних інструментів, що застосовується як в загальній метрологічній теорії, так і в прикладних задачах, пов'язаних, зокрема, із вимірюванням параметрів радіосигналів та характеристик радіотехнічних пристроїв і систем. Необхідність їх використання породжена дією різноманітних шумів та завад, що спотворюють результати вимірювань. В таких ситуаціях типовим метрологічним підходом є представлення вимірювальної моделі у вигляді адитивної взаємодії інформативного параметра і випадкової похибки. Компенсація впливу таких похибок заснована на проведенні багаторазових вимірювань. Зазвичай при обробці результатів вимірювань приймається, що розподіл експериментальних даних є симетричним відносно свого центру. При цьому істинне значення інформативного параметра може бути визначене як координата цього центру [1, 2].

Найпростішим і часто застосовуваним на практиці є звичайне усереднення результатів багаторазових вимірювань. Проте з точки зору математичної статистики використання лінійної оцінки параметрів у вигляді середнього арифметичного є оптимальним (за критерієм мінімізації дисперсії оцінок) лише для достатньо вузького класу випадкових похибок, зокрема тих, що адекватно описуються гаусовим (нормальним) законом розподілу ймовірностей. Широке поширення цієї моделі обґрунтовується відомим наслідком центральної граничної теореми. Проте такий розподіл не відображає всього різноманіття реальних похибок. Це призводить до необхідності використання інших, більш загальних не гаусових моделей [2–4].

Одна із відомих альтернатив базується на понятті узагальненого розподілу помилок, обґрунтованого в роботі Суботіна [5]. На цій основі було формалізовано модель так званого експоненціального степеневого розподілу (ЕСР), що описується функцією виду:

$$w(x) = \frac{1}{2\sigma p^{1/p}\,\Gamma(1+1/p)} \exp\left(-\frac{|x-\theta|^p}{p\sigma^p}\right), \quad (1)$$

де θ — центр розподілу, σ — параметр масштабу, p — параметр форми.

Очевидно, що при значенні параметр форми $p = 2$ модель виду (1) відповідає гаусовому закону, тому її також часто називають узагальненим гаусовим розподілом ймовірностей [6–8]. Зазначимо, що для значень $p < 2$ функція (1) має гостроверхий характер з пологими спадами (при $p = 1$ відповідає розподілу Лапласа). Із зростанням значень $p > 2$ розподіл стає більш плосковершинним і в граничному випадку (при $p \to \infty$) трансформується в рівномірний закон (див. рис. 1). Таким чином, універсальність моделі (1) пояснюється тим, що її імовірнісні властивості фактично залежать лише від одного параметра форми p, варіювання яким дозволяє суттєво їх змінювати.

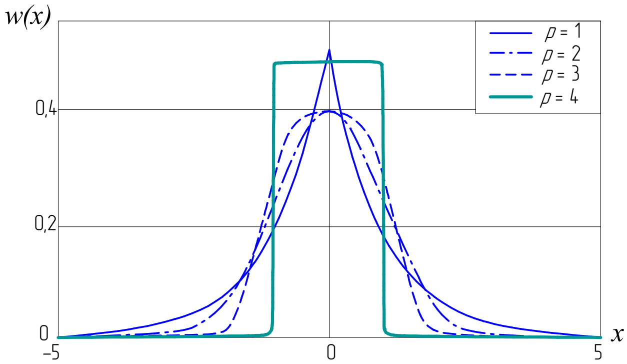

Рис. 1. Симетричні розподіли на основі експоненціальної моделі при різних значеннях параметра форми p

Відомо, що практичне використання будь-яких розподілів вимагає розвинених обчислювальних процедур статистичного оцінювання їх параметрів. Існує два основних підходи до знаходження оцінок параметрів моделі (1), що базуються відповідно на методі моментів та методі максимальної правдоподібності. Перший характеризується простотою алгоритмічної реалізації, але має відносно невисоку точність. Другий призводить до складних чисельних процедур і є лише асимптотично ефективним, вимагаючи достатньо великих обсягів вибірки [7–9].

Необхідно також відзначити, що за останні десятиліття сформоване ціле сімейство експоненціальних степеневих розподілів за рахунок поширення на двовимірний [6] і багатовимірний [10] випадки, а також розробки мультимодальних [11] та асиметричних модифікацій [12–14].

1 Мета дослідження

Відомо, що знаходження оцінки координати центру симетрично-розподілених випадкових даних може бути отриманий на основі різних статистик: медіани, середнього, середини розмаху, знакових та рангових оцінок. Аналіз монографій [2, 3], в яких наведені результати досліджень щодо використання ЕСР як моделі випадкових похибок, показує, що ефективність тієї чи іншої статистики суттєво залежить від величини параметру форми p. Зокрема, як вже зазначалося, для гаусового розподілу (при $p = 2$) найбільш ефективною є оцінка середнього, для розподілу Лапласа (при $p = 1$) — медіанна оцінка, а при рівномірному розподілі (для $p \to \infty$) — середина розмаху. Деякі дослідники рекомендують застосовувати зважені суми різних статистик, використовуючи в якості критерію вибору величину коефіцієнту ексцесу, оцінку якої достатньо просто отримати [15, 16].

В даній роботі пропонується використати оригінальний підхід до статистичного оцінювання параметрів, що базується на методі максимізації полінома (ММПл) [17]. Цей відносно новий метод статистичного оцінювання використовує опис випадкових величин у вигляді скінченої кількості усереднених статистик, наприклад, моментів або кумулянтів. У роботі [18] здійснено порівняльний аналіз ефективності ММПл-оцінок і оцінок середнього для різних типів симетричних розподілів (арксінусного, рівномірного, трапецієподібного, трикутного), а також досліджені властивості емпіричних розподілів цих оцінок. У роботі [19] на прикладі моделі трапецієподібного та роботі [20] для бімодальних розподілів більш детально досліджені області ефективності ММПл-оцінок порівняно з класичними оцінками.

Дане дослідження є безпосереднім продовженням робіт [18–20]. Його основною метою є порівняльний аналіз точності ММПл-оцінок при використанні в якості ймовірнісної моделі помилок експоненціального степеневого розподілу. Методологія досліджень передбачає отримання виразів, що аналітично описують дисперсії ММПл-оцінок, та побудову (шляхом статистичного моделювання методом Монте-Карло) областей ефективності оцінок, що отримуються різними методами в залежності від значення параметру форми ЕСР і обсягу вибіркових даних.

2 Математична постановка задачі

Нехай θ — інформативний параметр, значення якого необхідно оцінити на основі аналізу вектору $\vec{x} = \{x_1, x_2, ...x_n\}$. Цей вектор містить незалежні і однаково розподілені вибіркові значення, отримані в результаті багаторазових вимірювань. Припускається, що математична модель статистичних даних адекватно описується експоненціальним степеневим розподілом виду (1). При цьому значення параметрів масштабу σ і форми p є апріорно невідомими.

3 Знаходження оцінок методом максимізації полінома

У роботі [17] показано, що знаходження оцінок деякого скалярного параметру θ методом максимізації поліномів із вибірки однаково-розподілених випадкових величин $\vec{x} = \{x_1, x_2, ...x_n\}$ може бути зведено до пошуку розв'язку рівняння загального виду:

$$\sum_{i=1}^{r} h_i\ (\theta) \left[\frac{1}{n} \sum_{v=1}^{n} f_i\,(x_v) - \Psi_i\,(\theta) \right] \Bigg|_{\theta=\hat{\theta}} = 0, \qquad (2)$$

де r — степінь полінома, $f_i\,(x)$, $i = \overrightarrow{1,r}$ — набір певним чином упорядкованих базисних функцій, а $\Psi_i\,(\theta) = E\,\{f_i\,(x)\}$ — їх математичні сподівання, що залежать від параметра, який оцінюється. Вагові коефіцієнти $h_i\ (\theta)$ в (2) знаходяться за умови забезпечення мінімуму дисперсії оцінок параметра θ при застосуванні полінома степені r.

Якщо використати в якості базисних функцій степеневі перетворення виду $f_i\,(x) = x^i$, то їх математичні сподівання є початковими моментами $\alpha_i\,(\theta) = E\left\{x^i\right\}$. У такому випадку оптимальні вагові коефіцієнти $h_i\ (\theta)$, $i = \overrightarrow{1,r}$ можуть бути знайдені із вирішення системи лінійних алгебраїчних рівнянь виду:

$$\sum_{i=1}^{r} h_i\,(\theta)\ \ F_{i,j}\,(\theta) = \frac{d}{d\theta}\alpha_j\,(\theta)\,, \quad j = \overrightarrow{1,r}, \qquad (3)$$

де $F_{i,j}\,(\theta) = \alpha_{i+j}\,(\theta) - \alpha_i\,(\theta)\,\alpha_j\,(\theta)$, $i,j = \overrightarrow{1,r}$. Таким чином, при використанні полінома степені r для формування рівняння (2) необхідно мати частковий ймовірнісний опис у вигляді послідовності моментів до $2r$-го порядку включно.

Показано [17, 18], що при застосуванні степеневих базисних функцій ММПл-оцінки, що отримуються при $r = 1$, є еквівалентними оцінкам середнього для довільного закону розподілу випадкових величин. Крім того, за умови симетрії розподілу ММПл-оцінки для степені $r = 2$, також вироджуються в лінійні оцінки.

Для застосування ММПл при степені $r = 3$ наведемо співвідношення для перших 6-и початкових моментів (тут і в подальшому залежність в позначеннях від параметру θ для більшої компактності виразів будемо опускати):

$$\begin{aligned} &\alpha_1 = \theta; \\ &\alpha_2 = \theta^2 + \mu_2, \\ &\alpha_3 = \theta^3 + 3\theta\mu_2, \\ &\alpha_4 = \theta^4 + 6\theta^2\mu_2 + 3\mu_2^2 + \mu_4, \\ &\alpha_5 = \theta^5 + 10\theta^3\mu_2 + 15\theta\mu_2^2 + 5\theta\mu_4, \\ &\alpha_6 = \theta^6 + 15\theta^4\mu_2 + 45\theta^2\mu_2^2 + 15\mu_2^3 + \\ &\quad + 15\theta^2\mu_4 + 15\mu_2\mu_4 + \mu_6, \end{aligned} \qquad (4)$$

де μ_i — центральні моменти розподілу (1), які залежать лише від параметрів масштабу σ і форми p [9]:

$$\mu_i = \Gamma\left(\frac{i+1}{p}\right)\Gamma\left(\frac{1}{p}\right)^{\frac{i-2}{2}}\Gamma\left(\frac{3}{p}\right)^{-\frac{i}{2}}\sigma^i. \qquad (5)$$

Якщо використати поліном степені $r = 3$, то рівняння для знаходження ММПл-оцінок $\hat{\theta}$ для випадку симетрично-розподілених даних може бути записано у вигляді:

$$h_1 \sum_{v=1}^{n} (x_v - \theta) + h_2 \sum_{v=1}^{n} \left[x_v^2 - \left(\theta^2 + \mu_2\right)\right] + \\ + h_3 \sum_{v=1}^{n} \left[x_v^3 - \left(\theta^3 + 3\theta\mu_2\right)\right] \Bigg|_{\theta=\hat{\theta}} = 0, \qquad (6)$$

де h_1-h_3 — оптимальні коефіцієнти, які знаходяться шляхом аналітичного розв'язку системи рівнянь (3) методом Крамера (з урахуванням виразів (4)), мають вигляд:

$$\begin{aligned} h_1 &= \frac{3\theta^2\left(\mu_4 - 3\mu_2^2\right) + 3\mu_4\mu_2 - \mu_6}{\mu_2^{-2}\left(\mu_4^2 - \mu_2\mu_6\right)}, \\ h_2 &= \frac{-3\theta\left(\mu_4 - 3\mu_2\right)}{\mu_2^{-2}\left(\mu_4^2 - \mu_2\mu_6\right)}, \\ h_3 &= \frac{\mu_4 - 3\mu_2}{\mu_2^{-2}\left(\mu_4^2 - \mu_2\mu_6\right)}. \end{aligned} \qquad (7)$$

Підставивши коефіцієнти (7) в (6), після перетворень отримаємо кубічне рівняння відносно параметра, що оцінюється:

$$A\theta^3 + B\theta^2 + C\theta + D\,\big|_{\theta=\hat{\theta}} = 0, \qquad (8)$$

де $A = 1$, $B = -3\hat{\alpha}_1$, $C = 3\hat{\alpha}_2 - \frac{\mu_6 - 3\mu_4\mu_2}{\mu_4 - 3\mu_2^2}$, $D = \hat{\alpha}_1\frac{\mu_6 - 3\mu_4\mu_2}{\mu_4 - 3\mu_2^2} - \hat{\alpha}_3$.

Необхідно відзначити, що в рівнянні (8) фігурують вибіркові статистичні початкові моменти $\hat{\alpha}_i = \frac{1}{n}\sum_{v=1}^{n} x_v^i$, $i = \overrightarrow{1,3}$.

В роботі [18] показано, що розв'язок кубічного рівняння виду (8) може бути отриманий в аналітичному вигляді на основі формул Кардано. Проте їх використання для отримання ММПл-оцінок ускладнюється наявністю нелінійних перетворень (обчислення квадратних і кубічних коренів), а також потенційною можливістю отримання декількох дійсних рішень, що потребує додаткової перевірки за критерієм кількості добутої інформації [17, 18].

З обчислювальної точки зору більш ефективним може бути використання чисельних ітераційних процедур розв'язку нелінійних рівнянь. Як початкове наближення для пошуку кореня рівняння (8) логічно використати лінійну оцінку шуканого параметра у вигляді середнього $\hat{\alpha}_1$, а як умову зупинки ітерацій прив'язати до величини середньоквадратичного відхилення лінійних оцінок $\delta = \sqrt{\left(\hat{\alpha}_2 - \alpha_1^2\right)/n}$, яку можна масштабувати на деякий множник Δ, що буде визначати необхідний ступінь точності (наприклад, $\Delta = 10^{-3}$). Зазначимо, що відповідні статистики $\hat{\alpha}_1$ та $\hat{\alpha}_2$ обчислюються у разі формування коефіцієнтів кубічного рівняння (8).

Якщо для чисельного розв'язку кубічного рівняння (8) використати метод Ньютона-Рафсона [21], то алгоритм обчислення ММПл-оцінок можна представити у вигляді наступного псевдокоду:

$$k \leftarrow 1$$
$$\hat{\theta}_k \leftarrow \hat{\alpha}_1$$
$$\Delta \leftarrow 10^{-3}$$
$$\delta \leftarrow \Delta\sqrt{(\hat{\alpha}_2 - \alpha_1^2)/n}$$
$$\text{while } \left|\hat{\theta}_k - \hat{\theta}_{k-1}\right| > \delta$$
$$k \leftarrow k+1$$
$$\hat{\theta}_k \leftarrow \hat{\theta}_{k-1} - \frac{A\theta_{k-1}^3 + B\theta_{k-1}^2 + C\theta_{k-1} + D}{3A\theta_{k-1}^2 + 2B\theta_{k-1} + C}$$

Рис. 2. Алгоритм обчислення ММПл-оцінок при степені поліному $r = 3$ на основі методу Ньютона-Рафсона

4 Точність оцінок методу максимізації полінома

Доведено [17], що ММПл-оцінки є слушними і асимптотично незміщеними. Аналітичне обчислення дисперсії таких оцінок базується на понятті кількості добутої інформації, величина якої в загальному випадку може бути знайдена за формулою:

$$J_{rn(\theta)} = n\sum_{i=1}^{r} h_i(\theta)\frac{d}{d\theta}\alpha_i(\theta). \tag{9}$$

Ця кількісна характеристика за своїм статистичним змістом подібна до інформації за Фішером [22], оскільки при $n \to \infty$ її обернена величина асимптотично прямує до дисперсії параметра, що оцінюється, тобто:

$$\sigma^2_{(\theta)r} = \lim_{n\to\infty} J^{-1}_{rn(\theta)}. \tag{10}$$

Для порівняльного аналізу точності оцінок використаємо поняття коефіцієнта зменшення дисперсії [17–20]:

$$g_{(\theta)\,r} = \frac{\sigma^2_{(\theta)\,PMM3}}{\sigma^2_{(\theta)\,PMM1}}. \tag{11}$$

Цей коефіцієнт є відношенням дисперсії ММПл-оцінок параметра θ, які знаходяться у разі застосування поліному r-го порядку до дисперсії оцінок ММПл-оцінок при $r = 1$. Оскільки оцінки середнього еквівалентні ММПл-оцінкам при степені $r = 1$, то співпадають і їх дисперсії, тобто

$$\sigma^2_{(\theta)PMM1} = \sigma^2_{(\theta)\,mean}.$$

Відомо з [22], що дисперсія $\sigma^2_{(\theta)\,mean}$ оцінок середнього не залежить від величини параметра θ, а визначається як відношення центрального моменту 2-го порядку μ_2 до обсягу вибірки n. А оскільки оцінки середнього еквівалентні ММПл-оцінкам при $r = 1$, то співпадають і їх дисперсії

$$\sigma^2_{(\theta)PMM1} = \sigma^2_{(\theta)\,mean} = \frac{\mu_2}{n}. \tag{12}$$

Використовуючи співвідношення (9) і (10), можна отримати аналітичний вираз, що для асимптотичного випадку (при $n \to \infty$) дозволяє теоретично визначати дисперсію ММПл-оцінок при степені поліному $r = 3$:

$$\sigma^2_{(\theta)PMM3} = \frac{\mu_2}{n}\left[\frac{\mu_2\mu_6 - \mu_4^2}{9\mu_2^3 - 6\mu_2\mu_4 + \mu_6}\right]. \tag{13}$$

З урахуванням виразу (5), величину коефіцієнту зменшення дисперсії можна представить у вигляді:

$$g_{(\theta)\,3} = \frac{\Gamma^2\left(\frac{1}{p}\right)}{\Gamma\left(\frac{3}{p}\right)} \cdot \frac{\Gamma\left(\frac{3}{p}\right)\Gamma\left(\frac{7}{p}\right) - \Gamma^2\left(\frac{5}{p}\right)}{9\Gamma^3\left(\frac{3}{p}\right) - 6\Gamma\left(\frac{5}{p}\right)\Gamma\left(\frac{3}{p}\right)\Gamma\left(\frac{1}{p}\right) + \Gamma\left(\frac{7}{p}\right)\Gamma^2\left(\frac{1}{p}\right)}. \tag{14}$$

Очевидно, що теоретичне значення коефіцієнта зменшення дисперсії ММПл-оцінок залежить виключно від параметра форми p експоненціального степеневого розподілу. Графік цієї залежності представлено на рис. 3

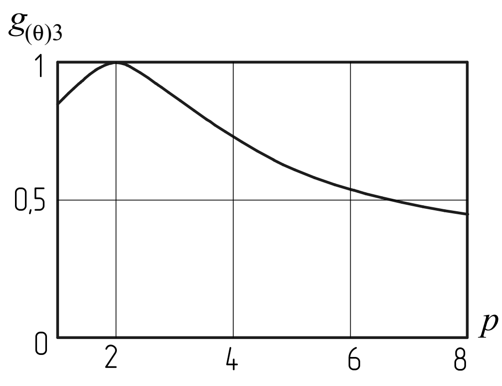

Рис. 3. Залежність коефіцієнту зменшення дисперсії $g_{(\theta)\,3}$ від параметра форми p ЕСР

Аналіз залежності, представленої на рис. 3, підтверджує вже відзначений раніше факт про однакову точність ММПл-оцінок і оцінок середнього в ситуації гаусового закону розподілу помилок (що відповідає значенню параметра форми $p = 2$). При всіх інших значеннях p потенційна точність ММПл є вищою. Особливо суттєвим (більше ніж у два рази) такий виграш може бути для плосковершинних розподілів з великими значеннями параметра p.

5 Статистичне моделювання

Для порівняльного аналізу ефективності ММПл-оцінок з іншими відомими методами оцінювання координати центру симетричних розподілів в середовищі Mathematica було розроблено програмний комплекс, що реалізує статистичне моделювання методом Монте-Карло. Він дозволяє на основі багаторазових експериментів з однаковими ймовірнісними властивостями вхідних даних проводити зіставлення точності оцінок, що отримуються на основі різних статистик: середнього арифметичного, медіани, центру розмаху та ММПл-оцінок при $r = 3$. По аналогії до (11), як кількісний критерій, що характеризує ефективність нового методу, використаємо емпіричні значення коефіцієнту відношення дисперсії:

$$\hat{g}_{(\theta)3} = \frac{\hat{\sigma}^2_{(\theta)PMM3}}{\hat{\sigma}^2_{(\theta)mean}},\quad \hat{q}_{(\theta)\,3} = \frac{\hat{\sigma}^2_{(\theta)PMM\,3}}{\hat{\sigma}^2_{(\theta)\,median}},\quad \hat{r}_{(\theta)\,3} = \frac{\hat{\sigma}^2_{(\theta)PMM\,3}}{\hat{\sigma}^2_{(\theta)\,mid-range}}, \tag{15}$$

де $\hat{\sigma}^2_{(\theta)\,mean}$, $\hat{\sigma}^2_{(\theta)\,median}$, $\hat{\sigma}^2_{(\theta)\,mid-range}$, $\hat{\sigma}^2_{(\theta)PMM\,3}$ — усереднені на основі m експериментів значення дисперсії оцінок, що отримуються із застосуванням середнього, медіани, центру розмаху та ММПл при $r = 3$ відповідно.

Очевидно, що на достовірність результатів моделювання в певній мірі впливають як величина обсягу вибірки n, так і кількість проведених експериментів m при фіксованих значеннях параметрів експоненціальних степеневих розподілів.

Необхідно також відзначити, що алгоритм обчислення ММПл-оцінок передбачає наявність інформації про параметри моделі помилок. Проте для більшості реальних ситуацій така інформація апріорно відсутня. У цьому випадку можна використати адаптивний підхід [19, 20], який полягає у обчисленні апостеріорних оцінок центральних моментів (необхідних для знаходження коефіцієнтів кубічного рівняння (8)), на основі співвідношення:

$$\hat{\mu}_i = \frac{1}{n}\sum_{v=1}^{n}\left(x_v - \bar{x}\right)^{\,i}, \tag{16}$$

де $\bar{x} = \hat{\alpha}_1 = \frac{1}{n}\sum\limits_{v=1}^{n} x_v$.

Сукупність результатів статистичного моделювання для різних значень параметра форми ЕСР ($p = 1 \div 10$) та обсягів вибіркових значень ($n = 20 \div 200$), що отримані при $m = 10^4$ багаторазових експериментів, представлено в табл. 1.

Аналіз теоретичних та експериментальних значень коефіцієнтів відношення дисперсії $g_{(\theta)3}$ та $\hat{g}_{(\theta)3}$ відповідно показує, те що існує певна кореляція аналітичних розрахунків та результатів, отриманих шляхом статистичного моделювання. Як можна помітити, зі збільшенням обсягу вибірки n розбіжність між теоретичними та експериментальними даними зменшується. Так при $n = 20$ розбіжність складає до 20 %, а вже при $n = 200$ зменшується до 2-3 %. Оскільки експериментальні результати прямують до теоретично розрахованих значень, це підтверджує асимптотичні властивості величини кількості добутої інформації (9), яка використовується при розрахунку дисперсії ММПл-оцінок.

На рис. 4 наведені границі, що розділяють області найбільшої ефективності (на основі критерію мінімуму дисперсії) при застосуванні різних методів оцінювання. Ці області отримані шляхом статистичного моделювання методом Монте-Карло (для $m = 10^4$ експериментів) при різних значеннях параметра форми ЕСР p та обсягів вибірки n.

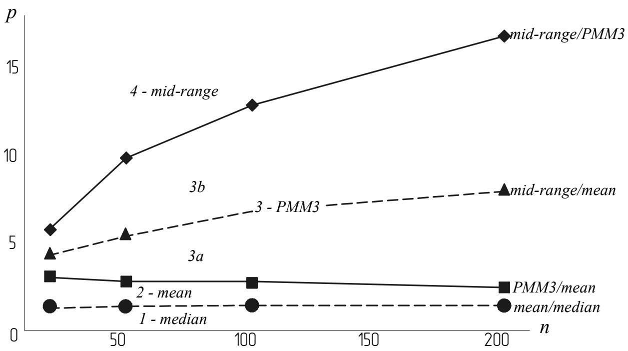

Рис. 4. Області ефективності методів знаходження оцінок координати центру сімейства симетричних експоненціальних розподілів

На основі наведених результатів можна зробити наступні висновки. Для значень $p < 1.3$ (область 1 — гостроверхінні розподіли) найбільш ефективною є медіанна оцінка. Для діапазону значень в межах $1.3 < p < 3$ (область 2) більш точною є статистика середнього. Для значень $p > 3$ більш ефективними стають ММПл-оцінки. Межа, що розділяє області 3 найбільш ефективних ММПл-оцінок і область 4 ефективних оцінок у вигляді середини розмаху суттєво залежить від обсягу вибіркових значень і носить параболічний характер. Наприклад, для $n = 20$ вона знаходиться біля значення $p \approx 6$, для $n = 50$ — $p \approx 10$, для $n = 100$ — $p \approx 13$, а для $n = 200$ в районі $p \approx 17$.

Більш наочну візуалізацію відносного ступеня ефективності ММПл-оцінок дає 3D-графік функції $Eff\,(p, n)$, представлений на рис. 5. Дану поверхню умовно можна поділити на декілька частин (залежно від границь ефективності різних методів, наведених на рис. 4. Зокрема, область 3 відповідає значенням параметрів p, n при яких ефективність ММПл-оцінок є кращою порівняно із іншими мето-

Табл. 1 Коефіцієнти відношення дисперсії оцінок

p	$g_{(\theta)3}$	Результати статистичного моделювання											
		$\hat{g}_{(\theta)3}$				$\hat{q}_{(\theta)\,3}$				$\hat{r}_{(\theta)\,3}$			
		n											
		20	50	100	200	20	50	100	200	20	50	100	200
10	0.4	0.63	0.51	0.44	0.43	0.27	0.19	0.16	0.15	1.38	1.12	0.81	0.6
5	0.61	0.8	0.71	0.62	0.61	0.36	0.29	0.25	0.23	0.9	0.6	0.39	0.24
3	0.89	0.99	0.97	0.9	0.89	0.52	0.5	0.42	0.4	0.61	0.37	0.21	0.12
2	1	1.13	1.08	1.04	1.01	0.77	0.7	0.67	0.66	0.39	0.2	0.1	0.06
1	0.85	0.97	0.85	0.83	0.81	1.4	1.36	1.36	1.51	0.12	0.05	0.02	0.01

дами, тобто виконується умова:

$$\min\left\{\hat{\sigma}^2_{(\theta)PMM3}, \hat{\sigma}^2_{(\theta)mean}, \hat{\sigma}^2_{(\theta)median}, \hat{\sigma}^2_{(\theta)mid-range}\right\} = \hat{\sigma}^2_{(\theta)PMM3}.$$

Сама поверхня відносної ефективності $Eff\,(p,n)$ для області 3 представляє собою інтерполяцію множини результатів статистичного моделювання і складається із двох частин: для 3a — це множина обернених значень коефіцієнтів ефективності $Eff\,(p,n) = \hat{g}^{-1}_{(\theta)3}$ (за умови $\min\left\{\hat{\sigma}^2_{(\theta)mean}, \hat{\sigma}^2_{(\theta)median}, \hat{\sigma}^2_{(\theta)mid-range}\right\} = \hat{\sigma}^2_{(\theta)mean}$), а для 3b $Eff\,(p,n) = \hat{r}^{-1}_{(\theta)\,3}$ (за умови $\min\left\{\hat{\sigma}^2_{(\theta)mean}, \hat{\sigma}^2_{(\theta)median}, \hat{\sigma}^2_{(\theta)mid-range}\right\} = \hat{\sigma}^2_{(\theta)mid-range}$).

Для тих випадків, коли ефективність ММПл-оцінок є нижчою відносно хоча б одного із інших методів (області 1, 2 і 4), для кращого візуального сприйняття величину функції ефективності штучно визначено як $Eff\,(p,n) = 1$.

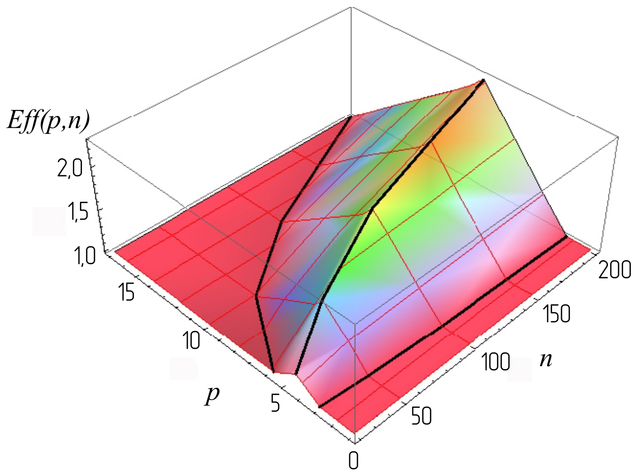

Рис. 5. Відносна ефективність ММПл-оцінок

Аналіз рис. 5 показує, що відносне зменшення дисперсії ММПл-оцінок може бути достатньо суттєвим (більше ніж у 2 рази).

Висновки

Дослідження, проведені в даній роботі, дозволяють зробити загальний висновок про потенційно високу ефективність застосування методу максимізації поліному для знаходження оцінок координати центру випадкових даних, які адекватно можуть бути описані моделлю у вигляді експоненціального степеневого розподілу. Важливим фактором є те, що застосування запропонованого підходу не потребує апріорної інформації про значення параметрів такої моделі. Показано, що алгоритм знаходження ММПл-оцінок інформативного параметру при степені полінома $r = 3$ зводиться до вирішення кубічного рівняння. Коефіцієнти такого рівняння формуються із використанням апостеріорних оцінок центральних моментів до 6-го порядку стохастичної складової (випадкової похибки). Розглянуто можливість використанням процедури Ньютона-Рафсона для визначення коренів.

На основі поняття кількості добутої інформації отримано аналітичний вираз, який описує залежність коефіцієнту зменшення дисперсії ММПл-оцінок (порівняно із оцінками середнього) в залежності від параметру форми експоненціального степеневого розподілу.

Шляхом статистичного моделювання проведено порівняльний аналіз ефективності (за критерієм мінімуму дисперсії) ММПл-оцінок із відомими непараметричними оцінками, на основі статистик середнього, медіани і середини розмаху. Побудовані області ефективності для кожного із методів в залежності від величини параметра форми експоненціальних розподілів та об'єму вибірки. Показано, що точність запропонованого підходу може суттєво (більше ніж у 2 рази) перевищувати класичні непараметричні оцінки.

Одним із подальших напрямків досліджень ефективності застосування методу максимізації поліному може бути порівняльний аналіз точності і складності обчислювальних процедур знаходження оцінок координати центру експоненційних відносно параметричного підходу на основі максимальної правдоподібності.

References

[1] *The International Vocabulary of Metrology, Basic and General Concepts and Associated Terms (VIM)*, JCGM 200:2012 [ISO/IEC Guide 99].

[2] Novitskii P. V. and Zograf I. A. (1991) *Otsenka pogreshnostei rezul'tatov izmerenii* [Estimation of errors of measurement results], Moskow, Energoatomizdat Publ., 304 p.

[3] Box G.E. and Tiao G.C. (1992) *Bayesian Inference in Statistical Analysis*. DOI: 10.1002/9781118033197

[4] Maugey T., Gauthier J., Pesquet-Popescu B. and Guillemot C. (2010) Using an exponential power model forwyner ziv video coding. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. DOI: 10.1109/icassp.2010.5496065

[5] Subbotin M.T. (1923) On the law of frequency of error, *Mat. Sb.*, Vil. 31, No 2, pp. 296-301.

[6] Taguchi T. (1978) On a generalization of Gaussian distribution. *Annals of the Institute of Statistical Mathematics*, Vol. 30, Iss. 1, pp. 211-242. DOI: 10.1007/bf02480215

[7] Varanasi M.K. and Aazhang B. (1989) Parametric generalized Gaussian density estimation. *The Journal of the Acoustical Society of America*, Vol. 86, Iss. 4, pp. 1404-1415. DOI: 10.1121/1.398700

[8] Nadarajah S. (2005) A generalized normal distribution. *Journal of Applied Statistics*, Vol. 32, Iss. 7, pp. 685-694. DOI: 10.1080/02664760500079464

[9] Crowder G.E. and Moore A.H. (1983) Adaptive Robust Estimation Based on a Family of Generalized Exponential Power Distributions. *IEEE Transactions on Reliability*, Vol. R-32, Iss. 5, pp. 488-495. DOI: 10.1109/tr.1983.5221739

[10] Lindsey J.K. (1999) Multivariate Elliptically Contoured Distributions for Repeated Measurements. *Biometrics*, Vol. 55, Iss. 4, pp. 1277-1280. DOI: 10.1111/j.0006-341x.1999.01277.x

[11] Hassan, M. Y. and Hijazi, R. H. (2010) A bimodal exponential power distribution, *Pak. J. Statist*, Vol. 26, No 2, pp. 379–396.

[12] Fernandez C., Osiewalski J. and Steel M.F.J. (1995) Modeling and Inference with ν-Spherical Distributions. *Journal of the American Statistical Association*, Vol. 90, Iss. 432, pp. 1331-1340. DOI: 10.1080/01621459.1995.10476637

[13] Komunjer I. (2007) Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics*, Vol. 22, Iss. 5, pp. 891-921. DOI: 10.1002/jae.961

[14] Zhu D. and Zinde-Walsh V. (2009) Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, Vol. 148, Iss. 1, pp. 86-99. DOI: 10.1016/j.jeconom.2008.09.038

[15] Zakharov I. P. and Shtefan N. V. (2002) Definition of effective distribution center value at statistical processing of measurement observations, *Radioelektronika ta informatyka*, No. 3 (20), pp. 97-99

[16] Warsza, Z. L., Galovska, M. (2009) About the best measurand estimators of trapezoidal probability distributions. *Przegląd Elektrotechniczny*, Vol. 85, No. 5, pp.86–91.

[17] Kunchenko Yu. P. (2002) *Polynomial parameter estimations of close to Gaussian random variables*, Aachen: Shaker Verlag. DOI:10.1007/978-3-319-77174-3

[18] Warsza Z.L. and Zabolotnii S.W. (2017) A Polynomial Estimation of Measurand Parameters for Samples of Non-Gaussian Symmetrically Distributed Data. *Advances in Intelligent Systems and Computing*, pp. 468-480. DOI: 10.1007/978-3-319-54042-9_45

[19] Warsza Z. and Zabolotnii S. (2017) Evaluation of the Uncertainty of Trapeze Distributed Measurements by the Polynomial Maximization Method. *Pomiary Automatyka Robotyka*, Vol. 21, Iss. 4, pp. 59-65. DOI: 10.14313/par_226/59

[20] Zabolotnii S. V., Kucheruk V. Yu., Warsza Z. L. and Khassenov A. K. (2018) Polynomial Estimates of Measurand Parameters for Data from Bimodal Mixtures of Exponential Distributions, *Vestnik Karagandinskogo universiteta*, No. 2 (90), pp. 71-80.

[21] Mathews J. H. and Fink K. D. (2004) *Numerical methods using MATLAB*, London, Pearson.

[22] Cramér H. (1946) *Mathematical Methods of Statistics (PMS-9)*. DOI: 10.1515/9781400883868

Полиномиальные оценки параметров для данных с экспоненциальным степенным распределением

Заболотний С. В., Чепинога А. В., Бондаренко Ю. Ю., Рудь М. П.

В работе предложен оригинальный подход к нахождению оценок результатов многократных измерений при случайных погрешностях, описываемых моделью экспоненциального степенного (обобщенного гауссова) распределения. В основе данного подхода лежит метод максимизации полинома (ММПл), основанный на математическом аппарате стохастических полиномов Кунченко и частичном описании случайных величин статистиками высших порядков (моментами или кумулянтами). Приведены теоретические основы ММПл для нахождения оценок информативного параметра из выборки одинаково распределенных случайных величин. Получены аналитические выражения для нахождения полиномиальных оценок. Показано, что при степени стохастического полинома $r \leqslant 2$ полиномиальные оценки вырождаются в линейные оценки среднего арифметического. При использовании полиномов степени $r = 3$ относительная точность полиномиальных оценок увеличивается. Рассмотрены особенности использования численных процедур (метод Ньютона-Рафсона) для нахождения решений стохастических уравнений. Для асимптотического случая (при $n \to \infty$) получены аналитические выражения, описывающие дисперсию ММПл-оценок. Показано, что теоретическое значение коэффициента уменьшения дисперсии ММПл-оценок (по сравнению с линейными оценками среднего арифметического) зависит от величины кумулянтних коэффициентов 4-го и 6-го порядка случайной погрешности. Путем многократных статистических испытаний (метод Монте-Карло) осуществлен сравнительный анализ точности полиномиальных оценок с известными непараметрическими оценкам (средним, медианой и серединой размаха). Показано, что с увеличением объема выборки n расхождение между теоретическими и экспериментальными данными уменьшается. Построены области эффективности для каждого из методов в зависимости от параметра формы экспоненциального степенного распределения и объема выборки. Показано, что точность предложенного подхода может существенно (более чем в 2 раза) превышать точность классических непараметрических оценок.

Ключевые слова: экспоненциальное степенное распределение; стохастические полиномы; статистики высших порядков; оценка параметра

Polynomial parameter estimation of exponential power distribution data

Zabolotnii S. V., Chepynoha A. V., Bondarenko Yu. Yu., Rud M. P.

The paper proposes an original approach to obtain the results of multiple measurements at random errors, which are described by the exponential power (generalized Gaussian) distribution model. The approach is based on the polynomial maximization method (PMM), which is based on Kunchchenko's mathematical apparatus using stochastic polynomials and a partial description of random variables of high-order statistics (moments or cumulants). The theoretical foundations of PMM are presented in relation to finding the estimates of the informative parameter from an equally distributed random variables sample. There are analytical expressions for finding polynomial estimations. It is shown that $r \leqslant 2$, then polynomial estimates degenerate in linear arithmetic mean estimates . If the polynomial degree $r = 3$ then the relative accuracy of polynomial estimations increases. The features of numerical procedures (Newton-Raphson method) for finding the stochastic equation roots are considered. Obtained analytical that describe the dispersion of the PMM estimates for an asymptotic case (for $n \to \infty$). It is shown that the theoretical value of reduction coefficient variance of PMM estimates (in comparison with the linear mean estimates) depends on the magnitude of the random error cumulative coefficients of the 4th and 6th order. Through multiple statistical tests (Monte Carlo method) carried out a comparative accuracy analysis of polynomial estimates with known nonparametric estimates (median, mid-range and mean). It is shown that with increasing size of sample the difference between theoretical and experimental data decreases. The efficiency areas for each method are constructed, depending on the exponential power distribution parameter and sample size. It is shown that the accuracy of the proposed approach can significantly (more than twofold) exceed the classical nonparametric estimation.

Key words: exponential power distribution; stochastic polynomials; high-order statistics; parameter estimation