УДК 519.233.22

В. В. Гарбарук, В. Н. Фоменко

УЧЕТ АПРИОРНОЙ ИНФОРАЦИИ О ПАРАМЕТРЕ В МЕТОДЕ НАИБОЛЬШЕГО ПРАВДОПОДОБИЯ

Дата поступления: 15.05.2016
Решение о публикации: 16.06.2016

Цель: Создать вероятностную модель, в которой параметры распределения рассматриваются как компоненты случайного вектора, в рамках этой модели построить состоятельные статистические оценки параметров распределения генеральной совокупности, которые учитывали бы априорную информацию об исследуемых величинах. **Методы:** Применены такие теории вероятностей и математической статистики, как апостериорная оценка гипотез по методу Байеса, метод наибольшего правдоподобия, частные и условные распределения, теория статистического оценивания параметров распределения. Все исследование проводится в рамках стандартной системы аксиом теории вероятностей. **Результаты:** Даны точечные оценки параметров, обладающие большей эффективностью, чем стандартные оценки. Под эффективностью оценки в работе понимается отношение среднего квадрата отклонения предложенной статистики и стандартной эффективной статистики от оцениваемого параметра. Получены общие выражения, определяющие как локальную (при заданном значении параметра), так и интегральную (средневзвешенную по априорному распределению параметра) эффективность полученных оценок. На примере математического ожидания нормально распределенной генеральной совокупности вычисляется эффективность при априорном нормальном и гамма-распределении математического ожидания. Показано, что полученные оценки имеют более высокую точность, чем стандартные «эффективные» оценки. Результаты, относящиеся к эффективности оценок, проиллюстрированы графически. **Практическая значимость:** Повышается точность статистических методов исследования в случае, когда объем эмпирических данных невелик. В частности, предложенный подход может оказаться полезным при проведении дорогостоящих экспериментов, некоторых видов опросов внутри небольших групп населения.

Математическая статистика, параметр распределения, точечная оценка, нормальное распределение, гамма-распределение, байесовская оценка, априорная информация.

Viktor V. Garbaruk, Cand. Sci. (Eng.), professor, vmkaf@pgups.ru; ***Viktor N. Fomenko**, Dr. Sci. (Phys.-math.), professor, vfomenko1943@gmail.com (Petersburg State Transport University) TAKING INTO ACCOUNT A PRIORI INFORMATION ON PARAMETER IN MAXIMUM LIKELIHOOD METHOD

Objective: To create a probability model in which distribution parameters are considered as components of random vector, to build consistent statistical estimates for statistical population distribution parameters within the framework of this model capable of accounting for a priori information on parameters under study. **Methods:** Probability and mathematical statistics theories were applied, including a posteriori hypothesis estimate by Bayes' method, maximum-likelihood method, marginal and conditional distributions, distribution parameter estimation theory. The entire study is being carried out within the framework of standard axioms system of probability theory. **Results:** Point estimates of parameters were provided, possessing higher degree of efficiency than standard estimates. In this study, estimate efficiency is understood as mean-square ratio of deviation of proposed statistics and standard efficient statistic from the estimated parameter. Common expressions were obtained, determining both local (with set parameter value) and integral (weighted average by a priori distribution of a parameter) efficiency of estimates obtained. On the example of mathematical expectation of normally distributed statistical population, efficiency is calculated for a priori normal and gamma-distribution of mathematical expectation. The estimates obtained were demonstrated to be more accurate than standard "efficient" estimates. Results concerning estimate efficiency were graphically illustrated. **Practical importance:** Accuracy of statistical study methods is increased in case when the size of empirical data is small. The proposed approach can be useful when conducting expensive experiments or certain types of polls amongst small groups of population.

Mathematical statistics, distribution parameter, point estimate, normal distribution, gamma-distribution, Bayes estimate, a priori information.

При статистическом оценивании параметров априорная информация о них, как правило, не принимается во внимание [5, 7–9]. Это не очень существенно, если оценка основывается на выборке большого объема, которая содержит достаточную информацию о параметре, и сама может использоваться для создания или уточнения базы априорной информации. Однако при малом объеме опытных данных учет априорных знаний может существенно увеличить надежность и точность оценки. Важность данного аспекта подчеркивалась в нескольких работах в области математической статистики (см., например, [1, 2]). Изложенный подход основывается на байесовской оценке апостериорных вероятностей событий, причем в качестве последних рассматриваются события, предполагающие, что параметр распределения принимает некоторое значение. При этом вероятностное пространство никак не переопределяется. Очевидно, такая трактовка параметра распределения как случайной величины некорректна.

В данной работе мы вводим вероятностное пространство случайных векторов, одной из компонент которого является параметр распределения. Основной объект рассмотрения в нашем подходе – двумерный случайный вектор $\Xi = (\gamma, \xi)$. Плотность распределения вектора Ξ задается функцией $h(g,x)$, где g рассматривается как параметр распределения ξ. Функция $p(g) = \int h(g,x)dx$ устанавливает распределение параметра в отсутствие результатов измерений ξ (априорное распределения параметра).

Введем в рассмотрение плотность условного распределения [9, 14, 15]

$$f(x \mid g) = \frac{h(g,x)}{p(g)}.$$

Результат серии n измерений вектора Ξ представляется совокупностью из $n+1$ случайной величины:

$$(\gamma, \xi_1, \ldots, \xi_n),$$

где $\xi_1, \ldots, \xi_n$ – взаимно независимые случайные величины с той же плотностью распределения $f(x|\gamma)$, что и у генеральной совокупности ξ, γ играет роль параметра распределения случайной величины ξ.

Точечная оценка параметра

Если $\varphi(x_1, ..., x_n; g)$ – плотность распределения выборки [11–15], то

$$\varphi(x_1, \ldots, x_n; g) = p(g)\prod_{i=1}^{n} f(x_i \mid g);$$
$$\ln\varphi(x_1, \ldots, x_n; g) = \tag{1}$$
$$= \ln p(g) + \sum_{i=1}^{n} \ln f(x_i \mid g).$$

Определим условную плотность распределения выборки при заданном параметре

$$\omega(x_1, \ldots, x_n; g) =$$
$$= \frac{\varphi(x_1, \ldots, x_n; g)}{p(g)} = \prod_{i=1}^{n} f(x_i \mid g) \tag{2}$$

и условную плотность распределения параметра γ при заданной выборке

$$\psi(x_1,\dots,x_n;g)=\frac{\varphi(x_1,\dots,x_n;g)}{\int\varphi(x_1,\dots,x_n;g')dg'}.$$

Пусть g_n – стационарная точка плотности $\varphi(x_1, ..., x_n; g)$ или, что то же самое, величины $\ln\varphi(x_1, ..., x_n; g)$ относительно g, т. е. корень уравнения

$$\frac{\partial\ln\varphi(x_1,\dots,x_n;g_n)}{\partial g_n}=0. \qquad (3)$$

Можно показать, что статистика g_n – состоятельная (смещенная) оценка параметра g, что является ее необходимым и достаточным обоснованием. Однако доказательство этого факта выходит за рамки данной статьи и будет опубликовано в отдельной работе.

Для дальнейшего изложения уравнение (3) удобно записать следующим образом:

$$\frac{\partial\ln p(g_n)}{\partial g_n}+\sum_{i=1}^{n}\frac{\partial\ln f(x_i\,|\,g_n)}{\partial g_n}=0. \qquad (4)$$

Байесовский подход к точечной оценке

Отметим, что уравнение (3) может рассматриваться как байесовская оценка гипотезы о значении параметра [2–4]. Действительно, разобьем область изменения параметра g на K частей с точками деления g_k, $(g_{k+1}>g_k)$. Введем гипотезы

$$H_k=\{\omega\in\Omega\,|\,g(\omega)\in[g_k,g_{k+1})\}.$$

Тогда апостериорные вероятности гипотез

$$P(H_k\,|\,x_1,\dots,x_n)=$$
$$=\frac{P(H_k)P(x_1,\dots,x_n\,|\,H_k)}{\sum_{m=1}^{K}P(H_m)P(x_1,\dots,x_n\,|\,H_m)}. \qquad (5)$$

Для независимой выборки

$$P(x_1,\dots,x_n\,|\,H_k)=\prod_{i=1}^{n}\frac{\int_{g_k}^{g_{k+1}}f(x_i\,|\,g)p(g)dg}{\int_{g_k}^{g_{k+1}}p(g)dg}$$

и

$$P(H_k)=\int_{g_k}^{g_{k+1}}p(g)dg.$$

Переходя к пределу $\delta=\max_k(g_{k+1}-g_k)\to 0$, получаем из (5)

$$P(g\,|\,x_1,\dots,x_n)=$$
$$=\lim_{\delta\to 0}\frac{P(H_k)}{g_{k+1}-g_k}=$$
$$=\frac{p(g)\prod_{i=1}^{n}f(x_i\,|\,g)}{\int p(g')\prod_{i=1}^{n}f(x_i\,|\,g')dg'}=$$
$$=\psi(g\,|\,x_1,\dots,x_n).$$

Критерий Байеса состоит в выборе гипотезы с максимальной апостериорной вероятностью. Таким образом, оценка параметра принимается в виде

$$g_n=\arg\max_g\psi(g\,|\,x_1,\dots,x_n). \qquad (6)$$

В случае дифференцируемости плотности $\Psi\ (g|x_1, ..., x_n)$ выражение (6) эквивалентно уравнениям (3) и (4).

Эффективность точечной оценки параметра

Прежде всего, отметим, что в асимптотическом пределе [11, 12] $n\to\infty$ уравнение (3), определяющее оценку g_n, переходит в обычное уравнение метода наибольшего правдоподобия. Отсюда вытекает, что оценка g_n асимптотически эффективна и нормально распределена [13–15].

Минимально возможная (в регулярном случае) дисперсия несмещенной точечной оценки, согласно неравенству Рао – Крамера, равна $1 / (nJ(g))$, где $J(g) = \int(\partial \ln f(x|g) / \partial g)^2 \times \times f(x|g)dg$ – информация Фишера [6, 10, 14, 15]. Несмещенная оценка с такой дисперсией обычно называется эффективной и под эффективностью какой-либо оценки часто понимается отношение минимально возможной дисперсии к дисперсии исследуемой оценки. Однако в данном случае этот критерий неприменим, так как оценка g_n является смещенной.

В этой работе под эффективностью оценки мы понимаем квадрат отношения среднеквадратических ошибок для стандартной «эффективной» оценки $g_n^{(e)}$ и оценки g_n, полученной в данном подходе:

$$E_0 = \frac{\int dgp(g)\int\prod_{i=1}^n f(x_i \mid g)\left[g - g_n{}^{(e)}(x_1,\ldots,x_n)\right]^2 dx_i}{\int dgp(g)\int\prod_{i=1}^n f(x_i \mid g)[g - g_n(x_1,\ldots,x_n)]^2 dx_i}.$$

Величину E_0 мы называем интегральной эффективностью.

Так как

$$\int\prod_{i=1}^n f(x_i \mid g)(g - g_n{}^{(e)})^2 dx_i = \mathrm{D}\, g_n{}^{(e)} = (nJ(g))^{-1},$$

для эффективности получаем

$$E_0 = \frac{\int \frac{p(g)}{J(g)} dg}{n\int dgp(g)\int\prod_{i=1}^n f(x_i \mid g)[g - g_n(x_1,\ldots,x_n)]^2 dx_i}. \quad (7)$$

Эффективность оценки при фиксированном значении параметра g определяется выражением

$$E_0 = \frac{1}{nJ(g)\int\prod_{i=1}^n f(x_i \mid g)[g - g_n(x_1,\ldots,x_n)]^2 dx_i}. \quad (8)$$

Примеры точечных оценок для некоторых законов распределения

В этом разделе приведем точечные оценки математического ожидания в случае, когда исследуемая величина имеет нормальное распределение (m, σ) при заданной дисперсии σ^2. Предположим, что априорное распределение математического ожидания имеет либо нормальное, либо γ-распределение. Параметры априорного распределения снабжаются нижним индексом 0.

Априорная плотность распределения параметра и плотность распределения значений выборки для нормального закона равны

$$p(g) = p_N(g) = \\ = \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left[-\frac{(g - m_0)^2}{2\sigma_0{}^2}\right]; \\ f(x \mid g) = f_N(x) = \\ = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x - g)^2}{2\sigma^2}\right]. \quad (9)$$

Априорная плотность для γ-распределения имеет вид

$$p(g) = p_\gamma(g) = \\ = \frac{1}{\Gamma(r_0)}\left(\frac{r_0}{m_0}\right)^{r_0} g^{r_0 - 1}\exp\left(-\frac{r_0}{m_0}g\right). \quad (10)$$

Параметр формы r_0 связан с математическим ожиданием и дисперсией соотношением $r_0 = m_0^2/\sigma_0^2$. Ниже мы рассматриваем только случай $r_0 \geq 1$.

Уравнение (4) приводит к следующим выражениям для оценок:

$$NN: g_n = \frac{1}{\sigma^2 + n\sigma_0^2}\left(m_0\sigma^2 + \sigma_0^2\sum_{i=1}^{n} x_i\right); \quad (11а)$$

$$N\gamma: g_n = \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{r_0\sigma^2}{nm_0}\right) + \sqrt{\frac{1}{4}\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{r_0\sigma^2}{nm_0}\right)^2 + \frac{\sigma^2}{n}(r_0 - 1)}. \quad (11б)$$

Исследуем эффективность полученных оценок согласно критериям (7) и (8). Информация Фишера для нормального распределения в случае, когда в качестве параметра рассматривается математическое ожидание, равна

$$J(g) = -\mathrm{M}\frac{\partial^2 \ln f(\xi\,|\,g)}{\partial g^2} = \frac{1}{\sigma^2}.$$

Введем удобные обозначения – приведенные априорные математическое ожидание и дисперсию:

$$\tilde{m}_0 = \frac{m_0}{\sigma}; \tilde{\sigma}_0 = \frac{\sigma_0}{\sigma}.$$

Тогда формула для интегральной эффективности принимает вид

$$E_0 = 1 + \frac{1}{n\sigma_0^2}.$$

Локальная эффективность задается выражением

$$E(g) = 1 + \frac{2\tilde{\sigma}_0^2 - \left(\frac{g}{\sigma} - \tilde{m}_0\right)^2 + \frac{1}{n}}{n\left[1 + \frac{1}{n\tilde{\sigma}_0^2}\left(\frac{g}{\sigma} - \tilde{m}_0\right)^2\right]}.$$

Рис. 1 и 2 представляют интегральную и локальную эффективности для нормального априорного распределения. На рис. 2 $\tilde{m}_0 = \tilde{\sigma}_0 = 1$. Подъем при $\tilde{\sigma}_0 \to 0$ объясняется тем, что при малой априорной дисперсии априорная информация почти полностью определяет параметр.

Обратимся теперь к случаю априорного γ-распределения. Из (11б) и (10) получаем для интегральной эффективности оценки

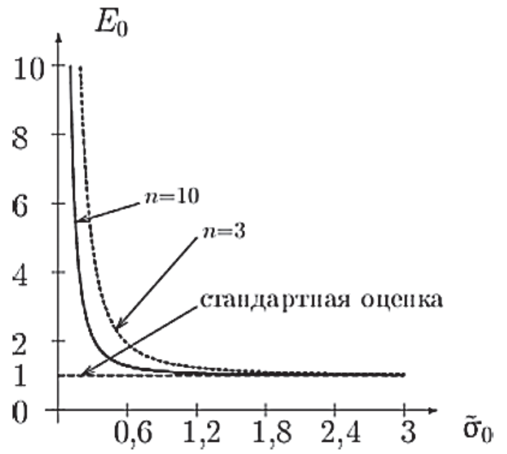

Рис. 1. Интегральная эффективность оценки математического ожидания при нормальном априорном распределении

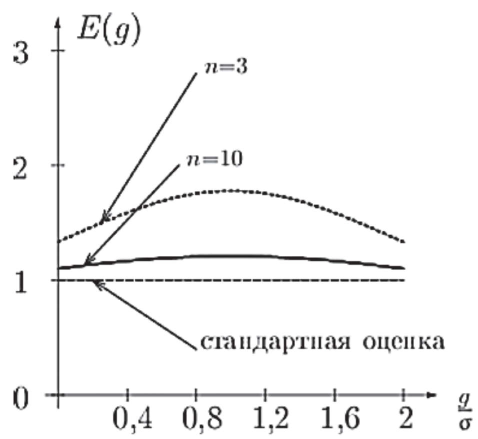

Рис. 2. Локальная эффективность оценки математического ожидания при нормальном априорном распределении при $\tilde{m}_0 = \tilde{\sigma}_0 = 1$

$$E_0 = \left[\begin{array}{c} \dfrac{n}{\Gamma(\dfrac{\tilde{m}_0{}^2}{\tilde{\sigma}_0{}^2})} \left(\dfrac{\tilde{m}_0}{\tilde{\sigma}_0{}^2} \right)^{\frac{\tilde{m}_0{}^2}{\tilde{\sigma}_0{}^2}} \int\limits_0^\infty \gamma^{\frac{\tilde{m}_0{}^2}{\tilde{\sigma}_0{}^2}-1} \times \\ \times \exp\left(-\dfrac{\tilde{m}_0}{\tilde{\sigma}_0{}^2}\gamma \right) \Phi_n(\gamma, \tilde{m}_0, \tilde{\sigma}_0) d\gamma \end{array} \right]^{-1}. \quad (12)$$

Эффективность при определенном значении параметра определяется формулой

$$E(g) = \\ = \frac{\int \prod_{i=1}^n f(x_i \mid g) \left[g - g_n^{(e)}(x_1, \ldots, x_n) \right]^2 dx_i}{\int \prod_{i=1}^n f(x_i \mid g) [g - g_n(x_1, \ldots, x_n)]^2 dx_i} = \\ = \frac{1}{\Phi_n(\frac{g}{\sigma}, \tilde{m}_0, \tilde{\sigma}_0)}. \quad (13)$$

В уравнениях (12) и (13) мы ввели обозначение

$$\Phi_n(\gamma, \tilde{m}_0, \tilde{\sigma}_0) = \sqrt{\frac{n}{2\pi}} \int\limits_{-\infty}^{\infty} \exp\left[-n(t-\gamma)^2 \right] \times \\ \times \left[\begin{array}{c} \dfrac{1}{2}\left(t - \dfrac{\tilde{m}_0}{\tilde{\sigma}_0{}^2} \right) + \\ + \sqrt{\dfrac{1}{4}\left(t - \dfrac{\tilde{m}_0}{\tilde{\sigma}_0{}^2} \right)^2 + \dfrac{1}{n}\left(\dfrac{\tilde{m}_0{}^2}{\tilde{\sigma}_0{}^2} - 1 \right)} - \gamma \end{array} \right]^2 dt.$$

На рис. 3 приведена интегральная эффективность E_0 (см. (12)) как функция $\tilde{\sigma}_0$ для случая $\tilde{\sigma}_0 = \tilde{m}_0$ (показательное распределение) при двух объемах выборки. Эффективность при учете априорной информации, по меньшей мере, в два раза превышает стандартную эффективную оценку. Рис. 4 показывает, как учет априорной информации влияет на локальную эффективность оценки. Оценка вычислена по формуле (13) при $\tilde{\sigma}_0 = \tilde{m}_0 = 1$.

Заключение

Исследования на примере математического ожидания нормально распределенной генеральной совокупности показывают, что использование априорной информации об изучаемом параметре может существенно (в несколько раз) уточнить его статистическую оценку. Это особенно полезно при выборке малого объема, когда точность стандартной оценки низка и эффект от априорной информации наибольший.

Следующим шагом в направлении использования априорных знаний могло бы стать построение оценок для нескольких параметров одновременно, принимая во внимание их корреляционные свойства, известные заранее.

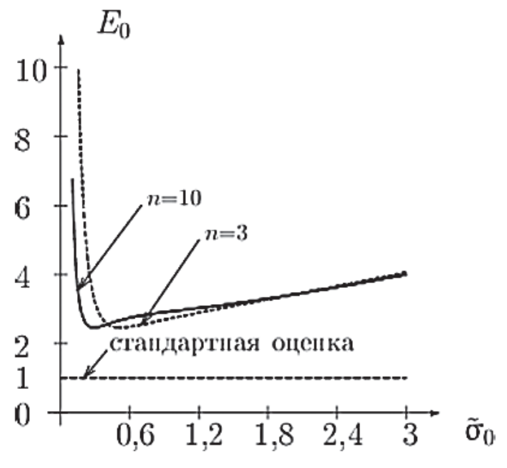

Рис. 3. Интегральная эффективность оценки математического ожидания при показательном априорном распределении

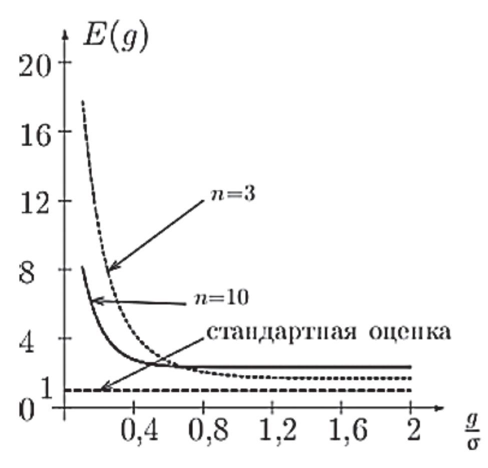

Рис. 4. Локальная эффективность оценки математического ожидания при показательном априорном распределении при $\tilde{m}_0 = 1$

Библиографический список

1. Айвазян С. А. Байесовский подход в эконометрическом анализе / С. А. Айвазян // Прикладная эконометрика. – 2008. – № 9. – С. 93–108.
2. Буре В. М. Байесовский подход / В. М. Буре, Л. В. Грауэр. – СПб. : ШАД, 2013. – 36 с.
3. Вентцель Е. С. Прикладные задачи теории вероятностей / Е. С. Вентцель, Л. А. Овчаров. – М. : Радио и связь, 1983. – 365 с.
4. Вентцель Е. С. Теория вероятностей / Е. С. Вентцель. – М. : Наука, 1969. – 564 с.
5. Ивченко Г. И. Математическая статистика / Г. И. Ивченко, Ю. И. Медведев. – М. : Высш. шк., 1984. – 352 с.
6. Кибзун А. И. Лекции по теории вероятностей / А. И. Кибзун, А. В. Наумов. – М. : МАИ, 2000. – 100 с.
7. Кожевников Ю. В. Введение в математическую статистику / Ю. В. Кожевников. – Казань : КГТУ, 1996. – 250 с.
8. Колемаев В. А. Теория вероятностей и математическая статистика / В. А. Колемаев, О. В. Староверов, В. Б. Турундаевский. – М. : Высш. шк., 1991. – 400 с.
9. Крамер Г. Математические методы статистики / Г. Крамер. – М. : Мир, 1975. – 648 с.
10. Лавренченко А. С. Лекции по математической статистике и теории случайных процессов / А. С. Лавренченко. – М. : МАИ, 1974. – 140 с.
11. Пугачев В. С. Введение в теорию вероятностей / В. С. Пугачев. – М. : Наука, 1968. – 368 с.
12. Пугачев В. С. Теория вероятностей и математическая статистика / В. С. Пугачев. – М. : Наука, 1979. – 496 с.
13. Розанов Ю. А. Лекции по теории вероятностей / Ю. А. Розанов. – М. : Наука, 1968. – 324 с.
14. Севастьянов Б. А. Курс теории вероятностей и математической статистики / Б. А. Севастьянов. – М. : Наука, 1982. – 255 с.
15. Чистяков В. П. Курс теории вероятностей / В. П. Чистяков. – М. : Наука, 1987. – 255 с.

References

1. Ayvazyan S. A. *Prikladnaya ekonometrika – Applied Econometrics*, 2008, no. 9, pp. 93-108.
2. Bure V. M. & Grauer L. V. Bayyesovskiy podkhod [Bayes' Approach]. St. Petersburg, ShAD, 2013. 36 p.
3. Ventsel Ye. S. & Ovcharov L. A. Prikladnyye zadachi teorii veroyatnostey [Applied Tasks for Probability Theory]. Moscow, Radio i svyaz, 1983. 365 p.
4. Ventsel Ye. S. Teoriya veroyatnostey [Probability Theory]. Moscow, Nauka, 1969. – 564 c.
5. Ivchenko G. I. & Medvedev Yu. I. Matematicheskaya statistika [Mathematical Statistics]. Moscow, Vysshaya shkola, 1984. 352 p.
6. Kibzun A. I. & Naumov A. V. Lektsii po teorii veroyatnostey [Lectures on Probability Theory]. Moscow, MAI, 2000. 100 p.
7. Kozhevnikov Yu. V. Vvedeniye v matematicheskuyu statistiku [Introduction to Mathematical Statistics]. Kazan, KGTU, 1996. 250 p.
8. Kolemayev V. A., Staroverov O. V. & Turundayevskiy V. B. Teoriya veroyatnostey i matematicheskaya statistika [Probability Theory and Mathematical Statistics]. Moscow, Vysshaya shkola, 1991. 400 p.
9. Kramer G. Matematicheskiye metody statistiki [Mathematical Mathods in Statistics]. Moscow, Mir, 1975. 648 p.
10. Lavrenchenko A. S. Lektsii po matematicheskoy statistike i teorii sluchaynykh protsessov [Lectures on Mathematical Statistics and Theory of Random Processes]. Moscow, MAI, 1974. 140 p.
11. Pugachev V. S. Vvedeniye v teoriyu veroyatnostey [Introduction to Probability Theory]. Moscow, Nauka, 1968. 368 p.
12. Pugachev V. S. Teoriya veroyatnostey i matematicheskaya statistika [Probability Theory and Mathematical Statistics]. Moscow, Nauka, 1979. 496 p.
13. Rozanov Yu. A. Lektsii po teorii veroyatnostey [Lectures in Probability Theory]. Moscow, Nauka, 1968. 324 p.
14. Sevastyanov B. A. Kurs teorii veroyatnostey i matematicheskoy statistiki [Probability Theory and Mathematical Statistics Course]. Moscow, Nauka, 1982. 255 p.
15. Chistyakov V. P. Kurs teorii veroyatnostey [Probability Theory Course]. Moscow, Nauka, 1987. 255 p.

ГАРБАРУК Виктор Владимирович – канд. техн. наук, доцент, профессор, vmkaf@pgups.ru; *ФОМЕНКО Виктор Николаевич – доктор физ.-мат. наук, профессор, профессор, vfomenko1943@gmail.com (Петербургский государственный университет путей сообщения Императора Александра I).