

исследования по анализу распределений, или ориентироваться на оценки сверху (2), (6) и (8).

Литература

1. **Руководство** по выражению неопределенности измерения / Пер. с англ., под ред. В. А. Слаева. СПб.: ВНИИМ, 1999.
2. **РМГ 43—2001**. ГСИ. Применение Руководства ИСО по выражению неопределенности измерений.

3. **Рубичев Н. А., Рябцев Г. Г.** Типовые ошибки применения статистических методов обработки измерительной информации и способы их устранения // Метрология. 2012. № 6. С. 3—16.

4. **Вентцель Е. С.** Теория вероятностей. М.: ГИФМЛ, 1958.
5. **РМГ 91—2009**. ГСИ. Совместное использование понятий «погрешность измерения» и «неопределенность измерения». Общие принципы.

Дата принятия 24.04.2013 г.

519.24

Оптимальный выбор количества интервалов дискретизации области изменения одномерной случайной величины при оценивании плотности вероятности

А. В. ЛАПКО, В. А. ЛАПКО

Институт вычислительного моделирования СО РАН, Красноярск, Россия,
e-mail: lapko@icm.krasn.ru

Исследованы асимптотические свойства регрессионной оценки плотности вероятности, синтез которой предполагает декомпозицию статистических данных. На этой основе определена зависимость количества интервалов дискретизации от объема исходной информации.

Ключевые слова: плотность вероятности, декомпозиция исходных данных, непараметрическая статистика, регрессионная оценка, асимптотические свойства, количество интервалов дискретизации.

The asymptotic properties of the regression estimator of the probability density the synthesis of which assumes the decomposition of statistical data are examined. On this basis the dependence of number of sampling intervals from the volume of original data is determined.

Key words: probability density, initial data decomposition, nonparametric statistic, regression estimator, asymptotic properties, number of sampling intervals.

Выбор количества интервалов дискретизации области изменения значений случайной величины является одной из задач математической статистики. Рассмотрим ее решение с позиций оптимизации структуры регрессионной оценки плотности вероятности.

Пусть $x^i, i = \overline{1, n}$ — выборка из n независимых наблюдений одномерной случайной величины x с неизвестной плотностью вероятности $p(x)$, которая допускает разложение в ряд Тейлора.

Для условий выборок большого объема n в [1, 2] предложена методика синтеза регрессионной оценки плотности вероятности, которая основана на декомпозиции исходных статистических данных и последующем анализе характеристик получаемых множеств случайных величин. Развитие результатов этих исследований приведено в [3].

Разобьем область определения $p(x)$ на N непересекающихся интервалов длиной 2β и сформируем множества случайных величин $X^j, j = \overline{1, N}$. В качестве характеристик X^j при-

ем частоту P^j попадания случайной величины x в j -й интервал и его центр z^j . На основе полученной информации составим статистическую выборку $V_1 = (z^j, p^j = P^j / (2\beta), j = \overline{1, N})$, где центры z введенных интервалов имеют равномерный закон распределения $p_0(z) = (2\beta N)^{-1}$, а объем полученной выборки N может быть значительно меньше объема n исходных статистических данных.

Приближением по эмпирическим данным искомой плотности вероятности $p(x)$ возьмем непараметрическую оценку условного математического ожидания [3]:

$$\tilde{p}(x) = c^{-1} \sum_{j=1}^N P^j \Phi\left((x - z^j)/c\right), \quad (1)$$

где $\Phi(u)$ — ядерная функция, удовлетворяющая условиям [4, 5]:

$$\Phi(u) = \Phi(-u), \quad 0 \leq \Phi(u) < \infty;$$

$$\int \Phi(u) du = 1, \quad \int u^2 \Phi(u) du = 1;$$

$$\int u^m \Phi(u) du < \infty \quad \text{при } 0 \leq m < \infty.$$

Для упрощения записи бесконечные пределы интегрирования здесь и далее опускаем.

Нетрудно убедиться, что регрессионная оценка плотности $\bar{p}(x)$ является нормированной функцией, т. е. удовлетворяет основному свойству плотности вероятности.

Коэффициенты размытости $c = c(N)$ ядерных функций в статистике (1) убывают с ростом количества N интервалов дискретизации области определения плотности вероятности. Их оптимальное значение, минимизирующее асимптотическое выражение среднего квадратического отклонения (СКО) погрешности аппроксимации

$$(2\Delta / (Nc)) \int p^2(x) dx \int \Phi^2(u) du + (c^4 / 4) \int (p^{(2)}(x))^2 dx$$

определяется формулой [3]:

$$\bar{c} = \left[\frac{2\Delta \|\Phi(u)\|^2 \|p(x)\|^2}{N \|p^2(x)\|^2} \right]^{1/5}, \quad (2)$$

где $\|\Phi(u)\|^2 = \int \Phi^2(u) du$; $\|p(x)\|^2 = \int p^2(x) dx$; $\|p^{(2)}(x)\|^2 = \int (p^{(2)}(x))^2 dx$; $p^{(2)}(x)$ — вторая производная плотности вероятности $p(x)$ от x ; Δ — длина интервала изменения значений случайной величины x .

Ниже на основе результатов исследования асимптотических свойств регрессионной оценки плотности вероятности определим аналитическую зависимость количества N интервалов дискретизации от объема исходных статистических данных и характеристик анализируемой плотности вероятности.

Асимптотические свойства регрессионной оценки плотности вероятности. Представим модификацию регрессионной оценки плотности вероятности (1) в виде

$$\bar{p}(x) = (nc)^{-1} \sum_{i=1}^N \sum_{j=1}^n h \left[\frac{(x^j - z^i)}{\beta} \right] \Phi \left[\frac{(x - z^i)}{c} \right], \quad (3)$$

где индикаторная функция

$$h \left[\frac{(x^j - z^i)}{\beta} \right] = \begin{cases} 1, & \text{если } |x^j - z^i| \leq \beta; \\ 0, & \text{если } |x^j - z^i| > \beta, \end{cases}$$

определяет принадлежность элементов выборки

$$V = (x^j, j=1, \bar{n}) \text{ интервалам } (z^i \pm \beta), i=1, \bar{N}.$$

В отличие от (1) в ее модификации (3) в явном виде присутствуют параметры процедуры дискретизации β , N и объем n исходных статистических данных. Поэтому в резуль-

тате анализа (3) можно получить соответствующее ей асимптотическое СКО $W_2(c, N)$, зависящее от коэффициента размытости ядерных функций регрессионной оценки плотности и количества интервалов N дискретизации области изменения случайной величины. Подставив в $W_2(c, N)$ оптимальное значение $\bar{c}$ из (2), получим выражение $W_2(N)$, минимизация которого по N даст аналитическую зависимость количества интервалов дискретизации от объема n исходных статистических данных.

Следуя предложенной схеме, вычисляем

$$\begin{aligned} M(\bar{p}(x)) &= (nc)^{-1} \sum_{i=1}^N \sum_{j=1}^n \int \int h \left[\frac{(x^j - z^i)}{\beta} \right] \times \\ &\times p(x^j) dx^j \Phi \left[\frac{(x - z^i)}{c} \right] p_0(z^i) dz^i = \\ &= (Nc^{-1}) \int \int h \left[\frac{(v - t)}{\beta} \right] p(v) dv \Phi \left[\frac{(x - t)}{c} \right] p_0(t) dt, \end{aligned} \quad (4)$$

где M — математическое ожидание; $p_0(t) = (2\beta N)^{-1}$ — плотность вероятности распределения центров интервалов дискретизации.

Тогда (4) запишем в виде

$$M(\bar{p}(x)) = (2\beta c)^{-1} \int \int h \left[\frac{(v - t)}{\beta} \right] p(v) dv \Phi \left[\frac{(x - t)}{c} \right] dt. \quad (5)$$

Проведем в (5) замену переменных $v = \beta u + t$, $t = x - cw$, и, разложив функцию $p(x - cw + \beta u)$ в ряд Тейлора в точке x , получим

$$\begin{aligned} M(\bar{p}(x)) &\approx 0,5 \int \int h(u) \Phi(w) \times \\ &\times \left[p(x) + (\beta u - cw) p^{(1)}(x) + \frac{(\beta u - cw)^2}{2} p^{(2)}(x) \right] du dw. \end{aligned}$$

Отсюда с учетом свойств ядерной $\Phi(w)$ и индикаторной $h(u) = 1$ для $u \in [-1; 1]$ функций запишем асимптотическое выражение

$$M(\bar{p}(x)) \sim p(x) + \left[\frac{(\beta^2 / 3 + c^2)}{2} \right] p^{(2)}(x). \quad (6)$$

Определим асимптотические свойства СКО:

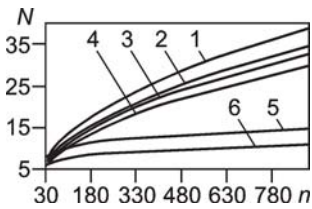
$$M(\bar{p}(x) - p(x))^2 = M(\bar{p}^2(x)) - 2p(x)M(\bar{p}(x)) + p^2(x). \quad (7)$$

Следуя ранее использованной технологии преобразований, сформируем асимптотическое выражение

$$\begin{aligned} M(\bar{p}^2(x)) &\approx (2\beta / c) p^2(x) \|\Phi(u)\|^2 + p^2(x) + \\ &+ p(x) p^{(2)}(x) (\beta^2 / 3 + c^2) + (p^{(2)}(x))^2 c^4 / 4 + \\ &+ (N / (nc)) p(x) \|\Phi(u)\|^2. \end{aligned} \quad (8)$$

С учетом (6) при достаточно больших объемах статистических данных второе слагаемое выражения (7) запишем как

$$\begin{aligned} -2p_1(x)M(\bar{p}(x)) &\approx -2p^2(x) - p(x) p^{(2)}(x) (\beta^2 / 3 + c^2) + \\ &+ 0(\beta^4) + 0(c^2 \beta^2). \end{aligned} \quad (9)$$



Зависимости количества N интервалов дискретизации области изменения значений случайной величины от объема n исходных статистических данных. Кривые 1—4 соответствуют результату (12), полученному для плотностей вероятности $p_j(x)$; $j = 1, 4$; кривые 4—6 рассчитаны по формулам из [6]

учетом ее асимптотической несмещенности (6) является состоятельной.

Анализ аппроксимационных свойств регрессионной оценки плотности вероятности. На основе результатов исследования асимптотических свойств $\bar{p}(x)$ (3) определим оптимальное количество интервалов дискретизации области изменения значений одномерной случайной величины x .

Подставим оптимальное значение коэффициента размытости (2) ядерных функций регрессионной оценки плотности вероятности (1) в критерий

$$W_2(c, N) = (\Delta / (Nc)) \|\rho(x)\|^2 \Phi(u)^2 + (N / (nc)) \|\Phi(u)\|^2 p(x) + (c^4 / 4) \|\rho^{(2)}(x)\|^2,$$

который получен путем интегрирования (10) и замены значения β на $\Delta / (2N)$.

В результате найдем

$$W_2(N) = \frac{3}{2^{6/5} N^{4/5}} \left[\left(\Delta \|\Phi(u)\|^2 \|\rho(x)\|^2 \right)^4 \|\rho^{(2)}(x)\|^2 \right]^{1/5} + \frac{N^{6/5}}{n} \left(\frac{\|\Phi(u)\|^4 \|\rho^{(2)}(x)\|^2}{2\Delta \|\rho(x)\|^2} \right)^{1/5}. \quad (11)$$

Для определения оптимального количества $\bar{N}$ интервалов дискретизации области изменения значений случайной величины, минимизирующего асимптотическую (при $n \rightarrow \infty$) погрешность аппроксимации, продифференцируем правую часть критерия (11) по N и приравняем к нулю полученное выражение

$$(6/5) \left(\left(\|\Phi(u)\|^2 \right)^4 \|\rho^{(2)}(x)\|^2 / \left(2N^9 \Delta \|\rho(x)\|^2 \right) \right)^{1/5} \times \times \left(N^2 / n - \Delta \|\rho(x)\|^2 \right) = 0.$$

Тогда, подставив результаты (8), (9) в (7), получим асимптотическое выражение для СКО

$$M(\bar{p}(x) - p(x))^2 \sim \sim (2\beta p^2(x) / c) \|\Phi(u)\|^2 + + (N / (nc)) \|\Phi(u)\|^2 p(x) + + (p^{(2)}(x))^2 c^4 / 4. \quad (10)$$

Из анализа (10) следует, что при выполнении условий $\beta/c \rightarrow 0$, $\beta \rightarrow 0$, $c \rightarrow 0$ и $N/(nc) \rightarrow 0$ (либо $nc\beta \rightarrow \infty$) при $n \rightarrow \infty$ оценка плотности вероятности (3) обладает свойством сходимости в СКО, а с

Отсюда оптимальное значение

$$\bar{N} = \sqrt{\Delta \|\rho(x)\|^2 n} \quad (12)$$

определяется видом восстанавливаемой плотности вероятности, областью ее определения и объемом исходных статистических данных. Полученная закономерность является объективной, так как не зависит от вида ядерных функций оценки плотности вероятности (3). Заметим, что выражение (12) не содержит информации о производных $p(x)$, поэтому появляется возможность использовать его при выборе количества интервалов дискретизации для широкого класса плотностей вероятностей.

Зависимость (12) по виду близка формуле Гаеде

$$N = \sqrt{n}, \quad (13)$$

причем для равномерной плотности вероятности $p(x)$ значения количества интервалов дискретизации, вычисленные по (12), (13) совпадают, так как в этих условиях произведение $\Delta \|\rho(x)\|^2 = 1$.

Проанализируем зависимость $\bar{N}$ от объема исходных статистических данных, соответствующую результату (12), полученному для ряда плотностей вероятности:

$$p_1(x) = (2\pi)^{-1/2} \exp(-x^2/2);$$

$$p_2(x) = 2x \quad \forall x \in [0; 1];$$

$$p_3(x) = \begin{cases} 3(1 - x^2/5) / (4\sqrt{5}) & \forall |x| < \sqrt{5}; \\ 0, & \forall |x| \geq \sqrt{5}; \end{cases}$$

$$p_4(x) = 1 \quad \forall x \in [0; 1].$$

При малых значениях n количество N интервалов дискретизации, полученных в соответствии с (12) и по известным формулам Старджесса, Брукса и Каррузера, Гаеде [6], сопоставимы (см. рисунок).

С ростом объема n исходных статистических данных наблюдается более значительное увеличение количества интервалов дискретизации $\bar{N}$ (12) по сравнению с результатами, получаемыми по указанным выше формулам. Очевидно, что увеличение $\bar{N}$ сопряжено с ростом объема статистической выборки V_1 , используемой при синтезе оценки плотности вероятности (1). При этом повышаются ее аппроксимационные свойства, что согласуется с результатами исследований [3, 7].

Влияние вида плотности вероятности на количество интервалов дискретизации в соответствии с (12) не столь велико.

Коэффициент $\sqrt{\Delta \|\rho(x)\|^2}$ в (12) для плотностей $p_j(x)$,

$j = 1, 4$ принимает значения: 1,3; 1,15; 1,1; 1.

Таким образом, регрессионная оценка плотности вероятности является эффективным средством обработки статистических данных большого объема. Ее синтез осуществляется путем декомпозиции исходной информации и последующего анализа количественных характеристик получаемых множеств случайных величин на основе кривой регрессии.

Исследование асимптотических свойств предложенной оценки плотности вероятности позволяет сформировать критерий для выбора оптимальных параметров ее структуры. Определена аналитическая зависимость количества интервалов декомпозиции области изменения значений случайной величины от объема статистических данных и характеристик восстанавливаемой плотности вероятности. Полученная аналитическая зависимость по виду близка к формуле Гаеде и совпадает с ней при оценивании плотности вероятности случайных величин с равномерным законом распределения.

Результаты исследований имеют важное значение в задачах оценивания плотности вероятности и проверки гипотез о распределениях случайных величин с использованием критерия Пирсона.

Л и т е р а т у р а

1. Лапко А. В., Лапко В. А. Непараметрические методики анализа множеств случайных величин // Автометрия. 2003. Т. 39. № 1. С. 54—61.

2. Lapko A. V., Lapko V. A. Nonparametric Methods For Random Variable Set Analysis // Proc. Optoelectron., Instrum. 2003. V. 39. № 1. P. 44—50.

3. Лапко А. В., Лапко В. А. Регрессионная оценка плотности вероятности и ее свойства // Системы управления и информационные технологии, 2012. Т. 49. № 3. С. 152—156.

4. Parzen E. On estimation of a probability density function and mode // Ann. math. statistic. 1962. V. 33. № 3. P. 1065—1076.

5. Епанечников В. А. Непараметрическая оценка многомерной плотности вероятности // Теория вероятностей и ее применения. 1969. Т. 14. Вып. 1. С. 156—161.

6. Лапко А. В., Лапко В. А. Сравнение непараметрических критериев проверки гипотез о распределениях случайных величин // Вестник СибГАУ. 2011. Т. 37. № 4. С. 48—52.

7. Лапко А. В., Лапко В. А. Анализ асимптотических свойств многомерной непараметрической регрессии // Вестник СибГАУ. 2012. Т. 42. № 2. С. 41—44.

Дата принятия 04.04.2013 г.

621.391

Модели надежности на основе модифицированных распределений Вейбулла

С. Я. ГРОДЗЕНСКИЙ

Московский государственный технический университет радиотехники, электроники и автоматики, Москва, Россия, e-mail: grodzensky44@mail.ru

Анализируется возможность применения модифицированных распределений Вейбулла для оценки и статистико-физического анализа надежности изделий. Предложено использовать смесь распределений — экспоненциального и Вейбулла.

Ключевые слова: надежность, ресурс, распределения Вейбулла и экспоненциальное, смесь распределений.

The possibility of application of modified Weibull distributions for evaluation and statistical-physical analysis of the reliability of the products is examined. The use of mixture of exponential and Weibull distributions is suggested.

Key words: reliability, resource, Weibull and exponential distributions, mixture of distributions.

Двухпараметрическая функция распределения вероятностей. В 1939 г. В. Вейбулл предложил функцию распределения, удобную для описания долговечности материалов [1]. При рассмотрении распределения отказов в форме $F(X) = 1 - \exp[-\varphi(X)]$ в [2] было отмечено, что единственным практическим путем достижения успеха является выбор простой функции, эмпирическая проверка и затем окончательный выбор, если нет ничего лучшего. В качестве простой служила двухпараметрическая функция распределения вероятностей

$$F(t) = 1 - \exp[-(t/T)^s], \quad (1)$$

где T , s — параметры масштаба и формы.

С середины 50-х годов прошлого века возрастает интерес к распределению Вейбулла, поскольку оно оказывается хорошей моделью для описания надежности сложных уст-

ройств [3]. Частными случаями (1) являются распределения экспоненциальное при параметре формы $s = 1$ и Рэлея при $s = 2$.

Основным показателем надежности многих видов промышленной продукции (в частности, изделий радиоэлектроники) является интенсивность отказов, или условная плотность вероятности возникновения отказа объекта, определяемая при условии, что до рассматриваемого момента времени отказ не возник. Интенсивность отказов — отношение их частоты к вероятности безотказной работы — выражается формулой

$$\lambda(t) = f(t)/P(t) = f(t) [1 - F(t)], \quad (2)$$

где $\lambda(t)$ — интенсивность отказов; $f(t)$ — частота отказов (плотность распределения случайной величины — времени t , физически определяет «скорость» падения надежности); $P(t)$ — вероятность безотказной работы в течение времени t .