УДК 621.039.5

ЯДЕРНЫЕ ОЦЕНКИ ПРИ РАСЧЕТАХ ГРАНИЦ ТОЛЕРАНТНЫХ ИНТЕРВАЛОВ

Осечкина Т.А.

Санкт-Петербургский государственный лесотехнический университет им. С.М. Кирова, Санкт-Петербург, e-mail: mathschool_pstu@mail.ru

В статье описана методика непараметрического ядерного оценивания толерантных пределов показателей надежности деталей машин для их сертификации и стандартизации. Отмечено, что для некоторых законов распределения, в частности, нормального, логнормального, показательного, Гаусса, Вейбулла, формулы оценок толерантных пределов получены. Но в некоторых случаях вид закона распределения исследуемой величины неизвестен, а в условиях малых выборок статистическая проверка закона распределения оказывается невозможной. В этом случае используются непараметрические оценки характеристик надежности деталей. Основные результаты в данном направлении получены на основе теории порядковых статистик с использованием свойств бета-функции. Такие оценки обычно жестко связывают количество наблюдений и доверительную вероятность, либо количество наблюдений и необходимую долю генеральной совокупности. В предлагаемой статье получены формулы для определения толерантных интервалов характеристик надежности на основе ядерных оценок. Приводится алгоритм получения толерантных пределов. Приводятся сравнительные результаты применения алгоритма при нахождении толерантных пределов усталостных характеристик по различным методикам. Для расчетов оценок ядерным методом использовалась функция Епанечникова. Показано, что оценки, полученные по предлагаемому алгоритму, в некоторых случаях улучшают толерантные пределы, полученные с помощью порядковых статистик. В частности, если распределение генеральной совокупности согласуется с нормальным или логнормальным законами распределения.

Ключевые слова: толерантные пределы, толерантные интервалы, надежность, непараметрические оценки, ядерные оценки

NUCLEAR ESTIMATES FOR CALCULATIONS OF TOLERANT INTERVALS

Osechkina T.A.

St. Petersburg State Forestry University named after S.M. Kirov, St. Petersburg, e-mail: mathschool_pstu@mail.ru

The article describes the method of non-parametric nuclear assessment of tolerant limits of reliability indicators of machine parts for their certification and standardization. It is noted that for some distribution laws, in particular, normal, lognormal, exponential, Gauss, Weibull, formulas for estimating tolerant limits have been obtained. But in some cases, the form of the distribution law of the quantity being studied is unknown, and in the conditions of small samples the statistical verification of the distribution law is impossible. In this case, nonparametric estimates of the reliability characteristics of the parts are used. The main results in this direction were obtained on the basis of the theory of order statistics using the properties of the beta function. Such estimates are usually tightly related to the number of observations and the confidence probability, or the number of observations and the necessary share of the population. In the proposed article, formulas are obtained for determining tolerant intervals of reliability characteristics based on nuclear assessments. An algorithm for obtaining tolerance limits is given. Comparative results of the application of the algorithm when finding tolerant limits of fatigue characteristics by various methods are given. To calculate the estimates using the nuclear method, the Epechnikov function was used. It is shown that the estimates obtained by the proposed algorithm, in some cases, improves the tolerance limits obtained using ordinal statistics. In particular, if the distribution of the general population is consistent with the normal or lognormal distribution laws.

Keywords: tolerance limits, tolerance intervals, reliability, nonparametric evaluation, nuclear evaluation

Указание границ толерантных интервалов является одним из основных условий международной сертификации промышленных изделий.

Толерантный интервал – интервал, определяемый по выборке, относительно которого можно утверждать с уровнем доверия γ, что он содержит, по крайней мере, указанную долю p_0 совокупности.

Границы статистического толерантного интервала – толерантные границы.

Уровень доверия $\gamma = 1 - \alpha$ – это вероятность того, что толерантный интервал будет содержать не менее $p_0 \cdot 100\,\%$ совокупности.

Вероятность попадания СВ X с функцией распределения $F_X(x)$ в интервал $[L, U]$ равна $F_X(U) - F_X(L)$. То есть границы толерантного интервала – это решение уравнения

$$P\Big(\big(F(U) - F(L)\big) \geq p_0\Big) = \gamma$$

для двустороннего интервала $[L, U]$ и

$$P\big(F(U) \geq p_0\big) = \gamma,$$

$$P\big(1 - F(L) \geq p_0\big) = \gamma$$

для односторонних интервалов $(-\infty; U]$ и $[L; +\infty)$ соответственно.

Обычно двусторонние толерантные интервалы ищутся в форме $[M - k\sigma; M + k\sigma]$, односторонние – в форме $\left(-\infty; M + k\sigma\right]$ или $\left[M - k\sigma; \infty\right)$.

На данный момент достаточно полно изучен вопрос оценки толерантных пределов, как двусторонних, так и односторонних, для некоторых известных распределений.

Так, в работах [1] выводятся оценки толерантных пределов нормально распределенной генеральной совокупности, в работах [1, 2] рассматриваются оценки толерантных пределов показательно распределенных совокупностей. В ряде работ предлагаются оценки толерантных границ для распределений Вейбула, Гаусса, логнормального распределения [3] сведением их к нормальному распределению.

Но в некоторых случаях априори вид распределения неизвестен. В этом случае необходимо либо предварительно выдвинуть гипотезу о виде распределения и проверить ее, что в случае ограниченной информации может вызвать затруднения, либо использовать непараметрические критерии для требуемых оценок.

Значительная часть результатов математической статистики основана на предположении о том, что информации, имеющейся у потребителя, достаточно для представления участвующих в задаче распределений в виде некоторых функций с конечным числом параметров. Однако на практике это предположение часто оказывается невыполнимым. Потребности в создании статистических процедур, не предполагающих знание вида распределений, отвечает ветвь математической статистики, получившая название непараметрическая статистика. Непараметрическая статистика рассматривает только такие ситуации, в которых о функциональном виде распределений ничего не известно. Теория непараметрического оценивания плотности распределения широко представлена в работах Л. Девроя и Л. Дьёрфи, Парзена, Розенблатта, Ш. Закса, Хансена и Кумпенса. Обзор укакзанных методов приведен в [2]. Оказалось, что, несмотря на малый объем априорной информации, используемой при построении непараметрических процедур, они обладают высокой эффективностью. Потери эффективности при переходе от параметрических к непараметрическим процедурам (в случае истинности параметрической модели) незначительны и довольно часто составляют всего несколько процентов. Эффективность непараметрической процедуры по сравнению с фиксированной параметрической резко возрастает при отклонении истинных распределений от расчетных. В большинстве случаях непараметрические процедуры оказываются асимптотически оптимальными.

Материалы и методы исследования

Широко известна [2] непараметрическая оценка толерантных границ с использованием β-распределения. Для упорядоченной выборки $\left\{X_k\right\}, k = \overline{1;n},$ из произвольно распределенной генеральной совокупности решение неравенства $np^{n-1} - (n-1)p^n \le 1 - \gamma$ определяет долю p генеральной совокупности, попадающей в интервал $\left[X_1, X_n\right]$ с вероятностью γ. Но полученная таким образом оценка жестко связывает долю p, доверительную вероятность γ и объем выборки n. То есть при одном и том же объеме выборки увеличение доли p ведет к уменьшению доверительной вероятности γ и наоборот.

Теория непараметрических доверительных и толерантных интервалов развивалась в исследованиях Уилкса, Оуэна, Гаттмена, Роббинса, Фрэзера, Барлоу и Прошана, Хансона и Кумпенса и других. В этих работах строятся функции специального вида, мажорирующие вероятность попадания случайной величины в заданный интервал, представленной в виде упорядоченной статистики. В основном эти оценки связаны с β-функцией.

В работах Хансена и Купменса [2] предлагается искать верхнюю толерантную границу для выборки X_i в виде

$$X_{(n-k+j)} + b\left(X_{(n-k)} - X_{(n-k-j)}\right), 1 \le k + j \le n - 1, 0 \le k \le n - 2,$$

где коэффициент b является решением уравнения

$$I_p(2, N-1) + p^{1/b} \frac{\Gamma(N+1)\Gamma\left(\frac{2b-1)}{b}\right)}{\Gamma(2)\Gamma\left(N + \frac{b-1}{b}\right)} \cdot \left[1 - I_p\left(\frac{2b-1}{b}, N-1\right)\right] = \gamma.$$

В ряде отечественных работ [3–5] для непараметрических оценок используется ядерный метод получения функции распределения. Согласно указанной методике, функция плотности вероятности рассматриваемой случайной величины определяется формулой

$$\tilde{f}_\xi(t;\sigma) = \frac{1}{n\sigma}\sum_{i=1}^{n} V\left(\frac{t-\xi_i}{\sigma}\right), \tag{1}$$

где $\xi_1;\xi_2;...;\xi_n$ – выборка, полученная в результате наблюдений за исследуемым объектом; n – объем выборки; $V(\cdot)$ – ядерная функция, удовлетворяющая условиям регулярности

$$V(y) = V(-y), 0 \le V(y) < \infty, \int_{-\infty}^{\infty} V(y)dy = 1,$$

$\sigma > 0$ – параметр сглаживания, такой, что $\lim\limits_{n\to\infty} \sigma n = \infty$.

В качестве ядерной функции чаще всего выбираются следующие функции [6]:

$V(t) = \frac{1}{2} I(|t| \le 1)$ – прямоугольное ядро;

$V(t) = 0,75(1-t^2)\cdot I(|t| \le 1)$ – ядро Епанечникова;

$V(t) = \frac{15}{16}(1-t^2)^2 I(|t| \le 1)$ – квадратическая ядерная функция.

В рассматриваемых функциях $I(\vartheta) = \begin{cases} 1, & \text{если условие } \vartheta \text{ выполнено;} \\ 0, & \text{если условие } \vartheta \text{ не выполнено.} \end{cases}$

Точность оценки зависит как от выбора функции-ядра, так и выбора параметра сглаживания. В работе [7] предложена методика поиска оптимального значения параметра сглаживания.

В работе [6] получена оценка доверительного интервала для теоретической функции плотности вероятности при $n \ge 100$

$$\hat{f}_\xi(x) - u_\beta \cdot \sqrt{\frac{D_\eta(x)}{n}} < f_\xi(x) < \hat{f}_\xi(x) + u_\beta \cdot \sqrt{\frac{D_\eta(x)}{n}}$$

и при небольших значениях $n \le 100$:

$$\hat{f}_\xi(x) - t_{n-1;(1-\beta)/2} \cdot \sqrt{\frac{D_\eta(x)}{n}} < f_\xi(x) < \hat{f}_\xi(x) + t_{n-1;(1-\beta)/2} \cdot \sqrt{\frac{D_\eta(x)}{n}}.$$

Результаты исследования и их обсуждение

Рассуждая аналогично [6], проведем оценку одностороннего толерантного интервала. Обозначим $\eta_i = \frac{1}{\sigma} V\left(\frac{x-\xi_i}{\sigma}\right)$. Случайные величины η_i при каждом x независимы.

Согласно центральной предельной теореме, при больших значениях n, оценка функции плотности вероятности, как суммы независимых величин, является при каждом x нормально распределенной случайной величиной, т.е.

$$\hat{f}_\xi(x) = \frac{1}{n}\sum_{i=1}^{n} \eta_i(x) \approx N\left(\frac{1}{n}\sum_{i=1}^{n} M_{\eta_i}(x), \sqrt{S_n^2}\right),$$

где

$$M_\eta(x) = \frac{1}{n\sigma}\int_{-\infty}^{+\infty}\sum_{i=1}^{n} V_i\left(\frac{x-\tau}{\sigma}\right)\cdot \hat{f}_\xi(\tau)d\tau,$$

$$D_{\eta}(x)=\frac{1}{n^2\sigma^2}\int\limits_{-\infty}^{+\infty}\sum_{i=1}^{n}V_i^2\left(\frac{x-\tau}{\sigma}\right)\times$$

$$\times\hat{f}_{\xi}(\tau)d\tau-M_{\eta}^2(x). \qquad (2)$$

Соответственно, случайная величина $M_{\hat{f}}(x)-\hat{f}_{\xi}(x)$ имеет нормальное распределение с нулевым математическим ожиданием и той же дисперсией:

$$M_{\hat{f}}(x)-\hat{f}_{\xi}(x)\approx N\left(0,\sqrt{\frac{D_{\eta}(x)}{n}}\right).$$

Тогда

$$P\left(M_{\hat{f}}(x)-\hat{f}_{\xi}(x)<\delta\right)=\Phi\left(\frac{\delta}{\sqrt{D_{\eta}(x)/n}}\right)$$

и при доверительной вероятности β:

$$P\left(M_{\hat{f}}(x)-\hat{f}_{\xi}(x)<\delta\right)=\beta\Rightarrow\Phi\left(\frac{\delta}{\sqrt{D_{\eta}(x)/n}}\right)=\beta,$$

$$\frac{\delta}{\sqrt{D_{\eta}(x)/n}}=\Phi^{-1}(\beta)=u_{\beta},$$

$$\delta=u_{\beta}\sqrt{\frac{D_{\eta}(x)}{n}}$$

(u_{β} – β-квантиль стандартного нормального распределения).

Окончательно получаем, что с вероятностью β верна оценка

$$M_{\hat{f}}(x)-\hat{f}_{\xi}(x)<u_{\beta}\cdot\sqrt{\frac{D_{\eta}(x)}{n}}$$

или

$$f_{\xi}(x)<\hat{f}_{\xi}(x)+u_{\beta}\cdot\sqrt{\frac{D_{\eta}(x)}{n}}.$$

При небольших значениях n используется распределение Стьюдента с n – 1 степенью свободы, т.е.

$$\frac{M_{\eta}-\hat{f}_{\xi}(t)}{\sqrt{D_{\eta}/n}}\approx t_{n-1,\beta}$$

и

$$f_{\xi}(x)<\hat{f}_{\xi}(x)+t_{n-1;\beta}\cdot\sqrt{\frac{D_{\eta}(x)}{n}}. \qquad (3)$$

Пусть $\hat{f}$ – оценка функции плотности вероятности f для случайной величины X, представленной выборкой $\xi_1;\xi_2;\ldots;\xi_n$. Оценка L нижней толерантной границы для доли p_0 с доверительной вероятностью γ является решением уравнения

$$P\left(1-F(L)\geq p_0\right)=\gamma. \qquad (4)$$

Проинтегрировав неравенство (3) и выполнив необходимые оценки, получим

$$F_{\xi}(x)<\hat{F}_{\xi}(x)+\int\limits_{-\infty}^{x}t_{n-1;\beta}\cdot\sqrt{\frac{D_{\eta}(\tau)}{n}}d\tau,$$

$$1-\hat{F}_{\xi}(x)-\int\limits_{-\infty}^{x}t_{n-1;\beta}\cdot\sqrt{\frac{D_{\eta}(\tau)}{n}}d\tau<1-F_{\xi}(x),$$

где

$$M_{\eta}(x)=\frac{1}{n\sigma}\int\limits_{-\infty}^{+\infty}\sum_{i=1}^{n}V_i\left(\frac{x-\tau}{\sigma}\right)\cdot\hat{f}_{\xi}(\tau)d\tau,$$

$$D_{\eta}(x)=\frac{1}{n^2\sigma^2}\int\limits_{-\infty}^{+\infty}\sum_{i=1}^{n}V_i^2\left(\frac{x-\tau}{\sigma}\right)\cdot\hat{f}_{\xi}(\tau)d\tau-M_{\eta}^2(x).$$

Другими словами, неравенство

$$1-F_{\xi}(x)>1-\hat{F}_{\xi}(x)-\int\limits_{-\infty}^{x}t_{n-1;\beta}\cdot\sqrt{\frac{D_{\eta}(\tau)}{n}}d\tau$$

верно с вероятностью β. Тогда, если верно неравенство

$$1-\hat{F}_{\xi}(L)-\int\limits_{-\infty}^{L}t_{n-1;\beta}\cdot\sqrt{\frac{D_{\eta}(\tau)}{n}}d\tau\geq p_0,$$

то с вероятностью β будет верно неравенство

$$1-F_{\xi}(L)\geq p_0,$$

т.е. $P\left(1-F_{\xi}(L)\geq p_0\right)=\beta$.

Следовательно, уравнение (4), определяющее значение оценки нижней границы L одностороннего толерантного интервала, верно в случае, если β = γ. Таким образом, оценку границы левостороннего толерантного интервала для доверительной вероятности γ можно найти как решение неравенства

$$1-\hat{F}_{\xi}(L)-\int\limits_{-\infty}^{L}t_{n-1;\gamma}\cdot\sqrt{\frac{D_{\eta}(\tau)}{n}}d\tau\geq p_0 \qquad (5)$$

при $n\leq 100$ и неравенства

$$1-\hat{F}_{\xi}(L)+\int\limits_{-\infty}^{L}u_{\gamma}\cdot\sqrt{\frac{D_{\eta}(\tau)}{n}}d\tau\geq p_0 \qquad (6)$$

при $n > 100$.

Таким образом, получен алгоритм построения нижней толерантной границы с помощью ядерной оценки функции плотности.

1. Оценить выборочное среднее, выборочную дисперсию СВ, представленной выборкой $X_i, i = \overline{1;n}$. Выбрать оптимальное значение параметра сглаживания.

2. Найти промежутки ненулевых значений ядер: $\left[u_j; u_{j+1}\right], u_k \in \left\{X_i - \sigma; X_i + \sigma\right\}$.

3. Определить на каждом из промежутков $\left[u_j; u_{j+1}\right]$ функцию

$$f(\sigma,t) = \frac{1}{n\sigma}\sum_{i=1}^{n} V\left(\frac{t-\xi_i}{\sigma}\right) = \frac{1}{n\sigma}\sum_{i=1}^{n}\frac{3}{4}\left(1-\left(\frac{t-\xi_i}{\sigma}\right)^2\right) I\left(\left(\frac{t-\xi_i}{\sigma}\right)\le 1\right).$$

4. Определить на каждом из промежутков $\left[u_j; u_{j+1}\right]$ функцию $F(x)$:

$$F(x) = \int_{-\infty}^{+\infty}\frac{1}{n\sigma}\sum_{i=1}^{n}\frac{3}{4}\left(1-\left(\frac{t-\xi_i}{\sigma}\right)^2\right) I\left(\left(\frac{t-\xi_i}{\sigma}\right)\le 1\right)dt.$$

5. Определить значение статистики $t_{n-1,\gamma}$.

6. Определить дисперсию оценки по формуле (1).

7. Найти нижнюю толерантную границу, решив неравенство (5) ((6)).

Заключение

Для апробации была рассмотрена выборка, полученная в результате усталостных характеристик 78 экспериментальных образцов. Варианты выборки принадлежали отрезку [195, 230]. Были получены оценки нижней толерантной границы для $\gamma = 0,95$, $p = 0,9$ а) по формулам для случая нормально распределенной генеральной совокупности $L = 191,44$, б) по методу Хансена – Кумпенса $L = 144,15$, в) по предложенному алгоритму ядерной оценки $L = 188,1$.

Список литературы

1. ГОСТ Р ИСО16269-6-2005. Статистические методы. Статистическое представление данных. Определение статистических толерантных интервалов. Введ. с 01.09.2005. М.: Стандартинформ, 2005. 29 с.

2. Закс Ш. Теория статистических выводов / Пер. с англ. М.: Мир, 1975. 767 с.

3. Antonov A., Polyakov A., Rodionov A., Chepurko V. Application of generalised linear model for time-dependent trend assessment – a case study for the ageing psa network. Reliability Engineering and System Safety. 2009. June. Vol. 94. no. 6. P. 1021–1029.

4. Antonov A., Zulyaeva N., Belousov A. et al. The statistical analysis of operating reliability of electropump units cn 60-180 for reactors VVER-1000 by root estimation methods. VI International Conference MMR 2009 – Mathematical methods in reliability. Theory. Methods. Applications.: extended abstracts. Moscow, 2009. 22–29 June. P. 521–525.

5. Antonov A., Belova K., Chepurko V. On one method of reliability coefficients calculation for objects in non-homogeneous event flows. Mathematical and Statistical Models and Methods in Reliability. Applications to Medicine, Finance, and Quality Control / Ed. by V.V. Rykov, N. Balakrishnan, M.S. Nikulin. -Statistics for Industry and Technology. Springer, 2010. P. 51–67.

6. Антонов А.В., Зюляева Н.Г., Белоусов А.Я. и др. Статистический анализ эксплуатационной надежности электронасосных агрегатов ЦН 60–180 реакторов ВВЭР-1000 методами корневого оценивания // Автоматика и телемеханика. 2010. Т. 71. № 7. С. 160–172.

7. Антонов А.В., Соколов С.В., Чепурко В.А. Вероятностные методы оценки остаточной наработки восстанавливаемых элементов ЯЭУ в условиях ограниченности исходных данных/ Ядерная физика и инжиниринг. 2011. Т. 2. № 5. С. 421–424.