

Интервальные доверительные оценки для показателей качества бинарных классификаторов — ROC-кривых, AUC для случая малых выборок

© С.Ю. Гуськов, В.В. Лёвин

АО «Банк ЗЕНИТ», Москва, 127566, Россия

На основе представления полиномиального распределения как условного совместного распределения независимых пуассоновских случайных величин получены доверительные интервалы для эмпирических функций распределения, построенных по группированным данным. Оценки доверительных границ далее применяются при построении доверительных интервалов для ROC-кривых. Указанные оценки могут быть использованы в автоматизированных процедурах контроля качества продукции при обнаружении и идентификации несплошностей, аномалий в структуре конструкционных материалов и их элементов, для повышения надежности и эффективности этих процедур при малых объемах выборок.

Ключевые слова: *доверительные интервалы, эмпирические функции распределения, полиномиальное распределение, пуассоновское распределение, ROC-кривая, бинарный классификатор.*

Введение. В работе представлен подход к построению доверительных границ для показателей качества бинарных классификаторов — ROC-кривых и связанных с ними характеристик (AUC, AR, CAP [1, 2]) для случая малых выборок. Предложены оценки доверительных границ для эмпирических функций распределения на основе группированных данных с использованием точных доверительных интервалов для пуассоновского распределения [3–6]. На основе этих оценок построены доверительные границы для ROC-кривых и связанных с ними характеристик [7–9].

Оценки показателей качества классификации могут быть использованы в автоматизированных процедурах выявления дефектов, основанных на обработке визуальной информации, которая поступает в ходе неразрушающего контроля (например, в автомобильной промышленности — снимки структуры отливок металла для выпуска автодеталей, полученные с применением компьютерной томографии, и т. п. [10–13]). Основой автоматизированных процедур контроля качества являются алгоритмы, реализующие бинарные классификаторы. По этим алгоритмам определяют наличие или отсутствие дефектов на основе анализа снимков структуры образцов продукции. Использование предлагаемых в статье доверительных оценок для показателей качества (ROC-кривых и связанных с ними характери-

стик) позволяет повысить эффективность работы автоматизированных процедур при малых объемах выборки.

ROC-кривая (рис. 1) представляет собой графическую характеристику качества бинарного классификатора — зависимость доли верных положительных классификаций от доли ложных положительных классификаций при варьировании уровня отсечения решающего правила.

Множество классов $Y = \{-1, +1\}$. Классификацию «+1» получают объекты, которые по своим характеристикам могут быть отнесены к положительным, а классификацию «-1» — объекты с противоположными характеристиками. В зависимости от содержательной интерпретации задачи к *положительным* относят объекты, выявление которых необходимо обеспечить в первую очередь. При этом в качестве ошибки I рода рассматривается неверная классификация положительных объектов.

Положительные объекты в медицинской диагностике — это пациенты с симптомами заболевания, которых необходимо выявить на ранней стадии болезни с использованием диагностического теста; в машиностроении при приемочном контроле качества продукции — это дефектные изделия, которые должны быть отсеяны с помощью процедур приемки качества; в банковском деле — клиенты, у которых высок риск невозврата долга и предоставление кредита которым нежелательно.

Далее для определенности положительные объекты будем обозначать как «плохие» (Bad), а отрицательные — как «хорошие» (Good).

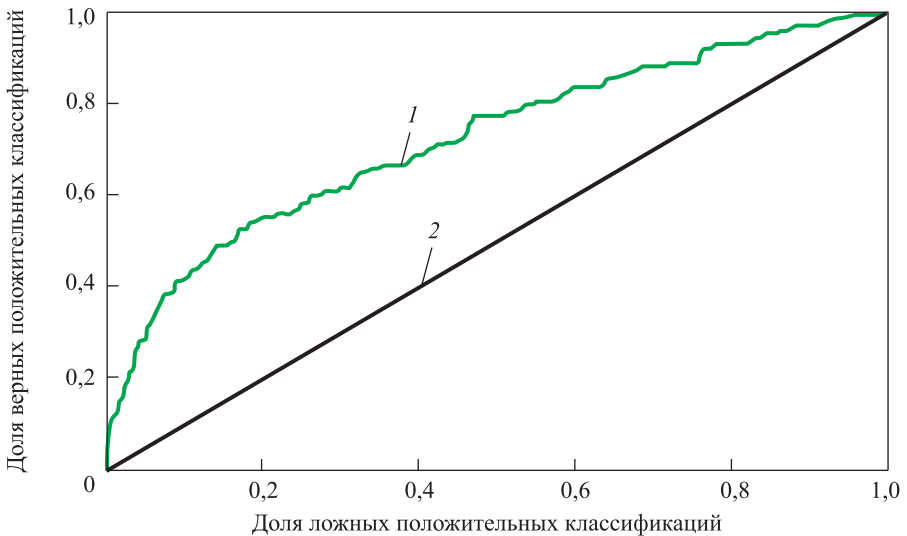


Рис. 1. Примеры различных ROC-кривых при наличии (кривая 1) и отсутствии (кривая 2) различающей способности у классификатора

Правило классификации задается в виде функции

$$\alpha(x) = \text{sign}(f(x, \bar{w}) - w_0),$$

где x — классифицируемый объект; $f(x, \bar{w})$ — дискриминантная функция; \bar{w} — вектор параметров, определяемый по обучающей выборке; w_0 — уровень отсечения, $0 < w_0 < 1$.

Для оценки качества классификаторов рассматривают следующие абсолютные и относительные показатели [1].

Абсолютные показатели качества различения (определяются для заданного уровня отсечения w_0):

- *TP (True Positives)* — верно классифицированные положительные объекты (так называемые *истинно положительные случаи*);
- *TN (True Negatives)* — верно классифицированные отрицательные объекты (*истинно отрицательные случаи*);
- *FN (False Negatives)* — положительные объекты, классифицированные как отрицательные (ошибка I рода). Это «ложный пропуск» — событие ошибочно не обнаруживается (*ложно отрицательные случаи*);
- *FP (False Positives)* — отрицательные объекты, классифицированные как положительные (ошибка II рода). Это «ложное обнаружение», так как при отсутствии события ошибочно выносятся решение о его присутствии (*ложно положительные случаи*).

В таблице приведено разбиение результатов классификации в зависимости от качества объекта (положительный или отрицательный объект) и от принимаемого решения (положительная или отрицательная классификация).

Таблица сопряженности (ошибок)

Модель	Фактически	
	Положительно	Отрицательно
Положительно	<i>TP</i>	<i>FP</i>
Отрицательно	<i>FN</i>	<i>TN</i>

Относительные показатели качества различения (определяются для заданного уровня отсечения w_0):

- *TPR (True Positives Rate)* — доля истинно положительных случаев,

$$TPR = \frac{TP}{TP + FN};$$

- *FPR (False Positives Rate)* — доля ложно положительных случаев,

$$FPR = \frac{FP}{TN + FP}.$$

В ряде приложений (медицинская диагностика, различение сигналов в радиотехнике) используют следующие относительные показатели:

- Se (*Sensitivity*) — чувствительность — доля истинно положительных случаев,

$$Se = \frac{TP}{TP + FN};$$

- Sp (*Specificity*) — специфичность — доля истинно отрицательных случаев, которые были правильно идентифицированы моделью,

$$Sp = \frac{TN}{TN + FP}.$$

Между показателями TPR , FPR и Se , Sp существует следующая связь:

$$FPR = 1 - Sp; \quad Se = TPR.$$

Все возможные ROC-кривые могут располагаться между ROC-кривой, соответствующей отсутствию различающей способности у классификатора (кривая совпадает с графиком функции $Y = X$), и ROC-кривой, соответствующей идеальному классификатору, безошибочно отличающему «плохих» от «хороших» (совпадает с графиком зависимости $Y = 1$) (рис. 2).

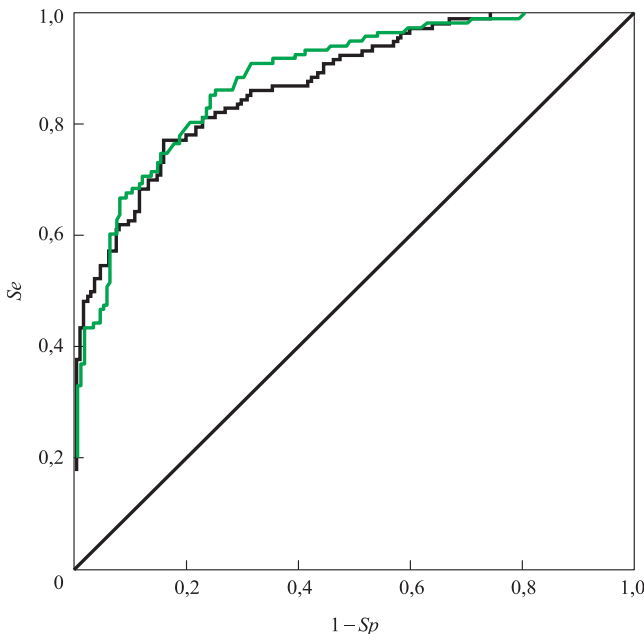


Рис. 2. Взаимное расположение ROC-кривых

Количественной характеристикой классификатора является показатель AUC (*Area Under Curve*), равный площади под ROC-кривой. Возможные значения AUC (в %) варьируются в интервале $[0; 100]$. Приемлемыми считаются классификаторы с $AUC \geq 70\%$.

Обозначим через $F_B(z)$ и $F_G(z)$ функции распределения соответственно «плохих» (*Bad*) и «хороших» (*Good*) объектов, непрерывные для $0 \leq z \leq 1$. Тогда для ROC-кривой справедливо представление

$$ROC(z) = F_B(F_G^{-1}(z)).$$

Соответственно для показателя AUC справедливо представление

$$AUC = \int_0^1 TPR(FPR) d(FPR).$$

Поскольку ROC-кривая и показатель AUC определяются по результатам наблюдений, для практического использования получаемых показателей необходимо иметь интервальные доверительные оценки для ROC-кривой и AUC, особенно для случая малых выборок (число наблюдений результатов классификации $n = 50 \dots 100$), когда асимптотические оценки имеют большую погрешность.

Интервальные оценки. Для построения интервальных доверительных оценок функций распределения F_G, F_B используется группировка наблюдений.

Интервал значений $[0; 1]$ разбивается на N подынтервалов $A_{jN} = \left[\frac{j-1}{N}; \frac{j}{N} \right]$, $j = 1, 2, \dots, N$, $\bigcup_{j=1}^N A_{jN} = A = [0; 1]$.

Функции $F_{G(B)N}(z)$ (оценки F_B, F_G по группированным данным) определяются как

$$F_{G(B)N} = \begin{cases} 0, & z \leq 0, \\ \frac{1}{n} \sum_{j=1}^{\lfloor Nz \rfloor + 1} v_{jN}^{G(B)}, & 0 < z \leq 1, \\ 1, & z > 1, \end{cases}$$

где $F_{G(B)N}$ — функции распределения для «хороших» (*Good*) и «плохих» (*Bad*) наблюдений соответственно, построенные по группированным данным; n — количество наблюдаемых данных; N — количество интервалов разбиения отрезка $[0; 1]$.

Здесь $\bar{v}_N = (v_{1N}^{G(B)}, \dots, v_{NN}^{G(B)})$ — случайный вектор частот попаданий «хороших» («плохих») наблюдений в соответствующие интервалы. Он имеет полиномиальное распределение $M(n; p_{1N}, \dots, p_{NN})$:

$$P(v_{1N} = k_1, \dots, v_{NN} = k_N) = \begin{cases} \frac{n!}{k_1! \dots k_N!} p_{1N}^{k_1} \dots p_{NN}^{k_N}, & \sum_{j=1}^N k_j = n, \\ 0, & \sum_{j=1}^N k_j \neq n. \end{cases}$$

Известно, что полиномиальное распределение является совместным распределением независимых пуассоновских случайных величин при условии, что их сумма фиксирована. Таким образом, если $\xi_{1N}, \dots, \xi_{NN}$ — независимые пуассоновские случайные величины, т.

е. $p(\xi_{jN} = k) = \frac{(np_{jN})^k}{k!} e^{-np_{jN}}, j = 1, 2, \dots, N$, то

$$P(v_{1N} = k_1, \dots, v_{NN} = k_N) = P\left(\xi_{1N} = k_1, \dots, \xi_{NN} = k_N \left| \sum_{j=1}^N \xi_{jN} = n \right.\right), \quad (1)$$

где λ_j — параметр пуассоновского распределения.

Для среднего значения np_{jN} отдельной пуассоновской случайной величины ξ_{jN} можно указать точный доверительный интервал с уровнем доверия $\gamma = 1 - \varepsilon$, где ε — заданный уровень значимости. Если k — наблюдавшееся значение случайной величины ξ_{jN} , имеющей пуассоновское распределение $\Pi(np_{jN})$, то для среднего значения np_{jN} точный доверительный интервал имеет вид

$$P\left\{\frac{1}{2}\chi_{2k_j, \frac{\varepsilon}{2}}^2 \leq np_{jN} \leq \frac{1}{2}\chi_{2k_j+2, 1-\frac{\varepsilon}{2}}^2\right\} = \gamma = 1 - \varepsilon.$$

Здесь $\chi_{m, \varepsilon}^2$ — квантиль распределения хи-квадрат с m степенями свободы и уровнем значимости ε .

Таким образом, для неизвестного параметра p_{jN} имеем доверительный интервал:

$$P\left\{\frac{1}{2n}\chi_{2k_j, \frac{\varepsilon}{2}}^2 \leq p_{jN} \leq \frac{1}{2n}\chi_{2k_j+2, 1-\frac{\varepsilon}{2}}^2\right\} = 1 - \varepsilon.$$

Для независимых $\xi_{1N}, \xi_{2N}, \dots, \xi_{NN}$, где $\xi_{jN} \sim \Pi(np_{jN}), j = 1, 2, \dots, N$, получаем доверительный параллелепипед

$$D_N(p_{1N}, \dots, p_{NN}) = \left\{ (p_{1N}, \dots, p_{NN}) \left(\frac{1}{2n} \chi_{2k_j, \frac{\varepsilon}{2}}^2 \leq p_{jN} \leq \frac{1}{2n} \chi_{2k_j+2, 1-\frac{\varepsilon}{2}}^2 \right), \right. \\ \left. j = 1, 2, \dots, N \right\}$$

с уровнем доверия γ_N , равным

$$P \left\{ \frac{1}{2n} \chi_{2k_j, \frac{\varepsilon}{2}}^2 \leq p_{jN} \leq \frac{1}{2n} \chi_{2k_j+2, 1-\frac{\varepsilon}{2}}^2, j = 1, 2, \dots, N \right\} = \\ = \prod_{j=1}^N P \left\{ \frac{1}{2n} \chi_{2k_j, \frac{\varepsilon}{2}}^2 \leq p_{jN} \leq \frac{1}{2n} \chi_{2k_j+2, 1-\frac{\varepsilon}{2}}^2 \right\} = (1-\varepsilon)^N. \quad (2)$$

Здесь k_1, k_2, \dots, k_N — наблюдавшиеся значения случайных величин $\xi_{1N}, \dots, \xi_{NN}$.

Используя связь между полиномиальным и пуассоновским распределением, получаем следующее утверждение.

Утверждение 1. Для полиномиального случайного вектора $\bar{v}_N = (v_{1N}, \dots, v_{NN}) \sim M(n; p_{1N}, \dots, p_{NN})$, распределение которого является совместным распределением независимых пуассоновских случайных величин при условии, что их сумма фиксирована, с неизвестными p_{1N}, \dots, p_{NN} , множество $D_N(p_{1N}, p_{2N}, \dots, p_{NN})$ является точным доверительным множеством с уровнем доверия $\gamma_1 = 1 - \varepsilon_1 = (1 - \varepsilon)^{N-1} n^{-n} n! e^n$, где ε_1 — уровень значимости для множества $D_N(p_{1N}, p_{2N}, \dots, p_{NN})$.

Доказательство. В силу соотношений (1) и (2)

$$B = P \left\{ \frac{1}{2n} \chi_{2v_{jN}, \frac{\varepsilon}{2}}^2 \leq p_{jN} \leq \frac{1}{2n} \chi_{2v_{jN}+2, 1-\frac{\varepsilon}{2}}^2, j = 1, 2, \dots, N \right\} = \\ = \frac{P \left\{ \frac{1}{2n} \chi_{2\xi_{jN}, \frac{\varepsilon}{2}}^2 \leq p_{jN} \leq \frac{1}{2n} \chi_{2\xi_{jN}+2, 1-\frac{\varepsilon}{2}}^2, j = 1, 2, \dots, N-1, \sum_{j=1}^N \xi_{jN} = n \right\}}{P \left\{ \sum_{j=1}^N \xi_{jN} = n \right\}} = B^*, \\ P \left\{ \sum_{j=1}^N \xi_{jN} = n \right\} = \frac{n^n e^{-n}}{n!},$$

так как сумма $\sum_{j=1}^N \xi_{jN}$ независимых пуассоновских случайных величин $\xi_{jN} \sim \Pi(np_{jN})$ также имеет распределение Пуассона с параметром $\sum_{j=1}^N np_{jN} = n$.

В то же время

$$B^* = \frac{(1-\varepsilon)^{N-1}}{P\left\{\sum_{j=1}^N \xi_{jN} = n\right\}} = (1-\varepsilon)^{N-1} n^{-n} n! e^n = \gamma_1,$$

что и требовалось доказать.

Примечание 1. Необходимо отметить, что случайный вектор $\frac{\bar{v}_N}{n} = \left(\frac{\bar{v}_{1N}}{n}, \dots, \frac{\bar{v}_{MN}}{n}\right)$ представляет собой гистограмму, построенную по группированным наблюдениям. Таким образом, множество $D_N(p_{1N}, p_{2N}, \dots, p_{NN})$ есть точное доверительное множество для гистограммы, которая является оценкой плотности распределения.

Примечание 2. Пусть n — число наблюдений, а N — число интервалов группировки. При заданных $n, \varepsilon, \varepsilon_1$ подходящее N можно выбрать по формуле

$$N = \left\lceil \frac{\ln(1-\varepsilon_1) + n \ln n - n - \ln(n!)}{\ln(1-\varepsilon)} \right\rceil + 1.$$

Примечание 3. В формуле (2) для каждого p_{jN} можно задать свой уровень значимости ε_j и получить соответствующий доверительный интервал для p_{jN} с уровнем доверия $\gamma_j = 1 - \varepsilon_j$, $j = 1, 2, \dots, N$.

Утверждение 2. Оценки $F_{GN}(z)$, $F_{BN}(z)$ при $N \rightarrow \infty$ и $0 < c_1 \leq \frac{n}{N} \leq c_2 < \infty$ при некоторых константах c_1, c_2 являются асимптотически несмещенными и асимптотически состоятельными оценками соответственно для $F_G(z)$, $F_B(z)$.

Доказательство. Математическое ожидание оценок

$$EF_{G(B)N}(z) = \frac{1}{n} \sum_{j=1}^{[Nz]+1} Ev_{jN}^{G(B)} = \frac{1}{n} \sum_{j=1}^{[Nz]+1} np_{jN}^{G(B)} = F_{G(B)}\left(\frac{[Nz]+1}{N}\right),$$

и в силу непрерывности функций $F_G(z)$ и $F_B(z)$

$$F_{G(B)}\left(\frac{[Nz]+1}{N}\right) - F_{G(B)}(z) \rightarrow 0.$$

Обозначим $R_{iN}^{G(B)} = \sum_{j=1}^i v_{jN}$, тогда

$$\text{Var } R_{iN}^{G(B)} = \sum_{j=1}^i np_{jN}^{G(B)}(1-p_{jN}^{G(B)}) - 2 \sum_{1 \leq j_1 < j_2 \leq i} np_{j_1N}^{G(B)}p_{j_2N}^{G(B)},$$

$$\begin{aligned} \text{Var } F_{G(B)N}(z) &= \sum_{j=1}^{[Nz]+1} \frac{p_{jN}^{G(B)}(1-p_{jN}^{G(B)})}{n} - \\ &- 2 \sum_{1 \leq j_1 < j_2 \leq [Nz]+1} \frac{p_{j_1N}^{G(B)}p_{j_2N}^{G(B)}}{n}, \end{aligned}$$

где $p_{jN}^{G(B)} = F_{G(B)}\left(\frac{j}{N}\right) - F_{G(B)}\left(\frac{j-1}{N}\right)$.

Для оценки $p_{jN}^{G(B)}$ используют полиномы Бернштейна:

$$B_M(F_{G(B)}; z) = \sum_{k=0}^M F_{G(B)}\left(\frac{k}{M}\right) C_M^k z^k (1-z)^{M-k}, \quad z \in [0; 1],$$

где M — произвольное натуральное число; C_M^k — число сочетаний из M по k .

Известно, что при $M \rightarrow \infty$ полиномы $B_M(F_{G(B)}; z)$ равномерно сходятся по z к $F_{G(B)}(z)$, т. е. справедлива теорема Вейерштрасса о равномерном приближении непрерывной функции полиномами.

Выберем $M = M(N)$ так, чтобы

$$\max_{z \in [0; 1]} |F_{G(B)}(z) - B_M(F_{G(B)}; z)| < \frac{1}{N}.$$

Поскольку $B_M(F_{G(B)}; z)$ — полином, он имеет ограниченную производную по $z \in [0; 1]$. Следовательно, оценка

$$\begin{aligned} p_{jN}^{G(B)} &= F_{G(B)}\left(\frac{j}{N}\right) - F_{G(B)}\left(\frac{j-1}{N}\right) \leq \\ &\leq \left| F_{G(B)}\left(\frac{j}{N}\right) - B_M\left(F_{G(B)}; \frac{j}{N}\right) + B_M\left(F_{G(B)}; \frac{j}{N}\right) - \right. \\ &\left. - B_M\left(F_{G(B)}; \frac{j-1}{N}\right) + B_M\left(F_{G(B)}; \frac{j-1}{N}\right) - F_{G(B)}\left(\frac{j-1}{N}\right) \right| \leq \end{aligned}$$

$$\begin{aligned} &\leq \left| F_{G(B)}\left(\frac{j}{N}\right) - B_M\left(F_{G(B)}; \frac{j}{N}\right) \right| + \left| B_M\left(F_{G(B)}; \frac{j}{N}\right) - B_M\left(F_{G(B)}; \frac{j-1}{N}\right) \right| + \\ &\quad + \left| F_{G(B)}\left(\frac{j}{N}\right) - B_M\left(F_{G(B)}; \frac{j}{N}\right) \right| \leq \frac{c}{N}, \end{aligned}$$

где $c = \text{const}$.

Отсюда следует оценка

$$\text{Var} F_{G(B)N}(z) \leq \sum_{j=1}^{[Nz]+1} \frac{p_{jN}^{G(B)}}{n} \leq \frac{c}{n} \sum_{j=1}^{[Nz]+1} \frac{1}{N} = \frac{c}{N} \frac{[Nz]+1}{N} \xrightarrow{N \rightarrow \infty} 0,$$

что и требовалось доказать.

Определим верхнюю $F_{G(B)N}^V(z)$ и нижнюю $F_{G(B)N}^L(z)$ граничные функции:

$$F_{G(B)N}^V(z) = \begin{cases} 0, & z \leq 0, \\ \frac{1}{2n} \sum_{j=1}^{[Nz]+1} \chi_{2\nu_{jN}G(B)+2, 1-\frac{\varepsilon}{2}}^2, & 0 < z \leq 1, \\ 1, & z > 1; \end{cases}$$

$$F_{G(B)N}^L(z) = \begin{cases} 0, & z \leq 0, \\ \frac{1}{2n} \sum_{j=1}^{[Nz]+1} \chi_{2\nu_{jN}G(B), \frac{\varepsilon}{2}}^2, & 0 < z \leq 1, \\ 1, & z > 1. \end{cases}$$

Утверждение 3. Функции $F_{G(B)N}^V(z)$, $F_{G(B)N}^L(z)$ являются точными доверительными границами для $F_{G(B)N}(z)$ с уровнем доверия γ_1 .

Доказательство. Для заданных k_1, k_2, \dots, k_N случайное событие $\{\xi_{1N} = k_1, \xi_{2N} = k_2, \dots, \xi_{NN} = k_N\}$ эквивалентно событию $\{\xi_{1N} = k_1, \xi_{1N} + \xi_{2N} = k_1 + k_2, \dots, \xi_{1N} + \dots + \xi_{NN} = k_1 + \dots + k_N\}$ и, следовательно, их вероятности совпадают, что и требовалось доказать.

Примечание 4. Функции $F_{G(B)N}^{L(V)}$ составляют доверительную полосу для всей функции $F_{G(B)N}(z)$ с заданным уровнем доверия γ_1 . При этом, как следует из утверждения 2 об асимптотической состоятельности (при $N \rightarrow \infty$) $F_{G(B)N}(z)$ для оценки истинной функции распределения $F_{G(B)}(z)$, максимальное расхождение между $F_{G(B)N}(z)$

и $F_{G(B)}(z)$ имеет порядок $1/N$. Поэтому для оценки $F_{G(B)}(z)$ доверительным интервалом с уровнем доверия γ_1 является

$$\left(F_{G(B)N}^V(z) + \frac{1}{N}, F_{G(B)N}^L(z) - \frac{1}{N} \right).$$

С помощью $F_{G(B)N}(z)$ строится оценка

$$ROC_N(z) = F_{BN} \left(F_{GN}^{-1}(z) \right),$$

а с помощью $F_{G(B)N}^{L(V)}$ — верхняя и нижняя доверительные оценки

$$ROC_N^L(z) = F_{GN}^L \left((F_{BN}^L)^{-1}(z) \right); ROC_N^V(z) = F_{GN}^V \left((F_{BN}^V)^{-1}(z) \right),$$

которые составляют доверительную полосу для функции $ROC(z)$ с заданным уровнем доверия γ_1 .

Аналогично по оценке $ROC_N(z)$ определяется оценка AUC_N , а по ROC_N^V , ROC_N^L строятся AUC_N^V , AUC_N^L — верхняя и нижняя границы доверительного интервала для AUC:

$$AUC_N^L(z) = \int_0^1 F_{GN}^L \left((F_{BN}^L)^{-1}(z) \right) dz;$$

$$AUC_N^V(z) = \int_0^1 F_{GN}^V \left((F_{BN}^V)^{-1}(z) \right) dz.$$

Заключение. В работе для случая малых выборок построены доверительные интервалы для ROC-кривых и показателя AUC с использованием связи полиномиального распределения с условным распределением независимых пуассоновских случайных величин. Указанные доверительные интервалы могут применяться для доверительного оценивания характеристик бинарных классификаторов в автоматизированных процедурах контроля качества продукции в случае малого числа наблюдений, когда асимптотические доверительные интервалы имеют большую погрешность.

Автоматизированные процедуры выявления дефектов и контроля качества продукции повышают уровень качества выпускаемой продукции в связи с сокращением времени обработки, снижением трудозатрат при анализе образцов и уменьшением количества субъективных ошибок по сравнению с «ручным» контролем качества операторами. Предлагаемые в статье доверительные интервальные оценки для показателей качества бинарных классификаторов (ROC, AUC и др.) позволяют повысить надежность и эффективность этих автоматизированных процедур.

ЛИТЕРАТУРА

- [1] Engelmann B., Hayden E., Tasche D. Testing rating accuracy. *RISK*, 2003, vol. 16, pp. 82–86.
- [2] Stein R.M. Benchmarking default prediction models pitfalls and remedies in model validation. *J. of Risk Model Validation*, 2007, vol. 1, no. 1, pp. 77–113.
- [3] Большев Л.Н. Сравнение интенсивностей простейших потоков. *Теория вероятностей и ее применение*, 1962, т. 7, с. 353–355.
- [4] Большев Л.Н. О построении доверительных пределов. *Теория вероятностей и ее применение*, 1965, т. 7, с. 356–358.
- [5] Garwood F. Fiducial limits for Poisson distribution. *Biometrika*, 1936, vol. 28, pp. 437–442.
- [6] Stevens W.L. Fiducial limits of the parameter of discontinuous distribution. *Biometrika*, 1950, vol. 37, pp. 117–129.
- [7] Sofus A. Macskassy and Foster Provost, Confidence Bands for ROC Curves: Methods and an Empirical Study. *CeDER Working Paper 02–04*. Stern School of Business, New York University, Jan. 2004, 15 p.
- [8] Jokiel-Rokita A., Pulit M. Nonparametric estimation of the ROC curve based on smoothed empirical distribution functions. *Statistical Computing*, 2013, vol. 23, pp. 703–712.
- [9] Baklizi A. A Simple Method for Finding Empirical Likelihood Type Intervals for the ROC Curve. *J. of Modern Applied Statistical Methods*, 2007, vol. 6, no. 2, pp. 589–595.
- [10] Le Meur Y., Vignolle J.-M., Chanussot J. Practical use of receiver operating characteristic analysis to assess the performances of defect detection algorithms. *J. of Electronic Imaging, Society of Photo-optical Instrumentation Engineers (SPIE)*, 2008, vol. 17, no. 3, p. 10.1117.
- [11] Dobrzański L.A., Krupiński M., Sokolowski J.H. Methodology of automatic quality control of aluminium castings. *J. of Achievements in Materials and Manufacturing Engineering*, 2007, vol. 20, no. 1–2, pp. 69–78.
- [12] Щеголев С.С., Мотков А.В. Комплексный автоматизированный метод выявления дефектов колец подшипников. *Интернет-журнал «НАУКОВЕДЕНИЕ»*, 2014, вып. 4 (23). URL: <http://naukovedenie.ru/PDF/110TVN414.pdf> (дата обращения 25.11.2014).
- [13] Иванов В.И., Коновалов Н.Н., Дергачев А.Н. Использование вероятностных методов для оценки эффективности неразрушающего контроля. *Технологии техносферной безопасности*, 2014, вып. 6 (58). URL: <http://agps-2006.narod.ru/ttb/2014-6/15-06-14.ttb.pdf> (дата обращения 25.11.2014).

Статья поступила в редакцию 10.12.2014

Ссылку на эту статью просим оформлять следующим образом:

Гуськов С.Ю., Лёвин В.В. Интервальные доверительные оценки для показателей качества бинарных классификаторов — ROC-кривых, AUC для случая малых выборок. *Инженерный журнал: наука и инновации*, 2015, вып. 3. URL: <http://engjournal.ru/catalog/mesc/idme/1376.html>



Гуськов Сергей Юрьевич родился в 1992 г., окончил НИУ ВШЭ в 2014 г. Экономист АО «Банк ЗЕНИТ». Область научных интересов: применение математических методов в решении экономических задач. e-mail: ser0432@yandex.ru



Лёвин Владимир Владимирович родился в 1951 г., окончил МИЭМ в 1969 г. Канд. физ.-мат. наук, доцент, начальник отдела скоринга АО «Банк ЗЕНИТ». Автор более 50 публикаций. Область научных интересов: предельные теоремы теории вероятностей и их применение в построении оптимальных статистических процедур. e-mail: vladimir.levin.51@mail.ru

Confidence interval estimation for quality factors of binary classifiers – ROC curves, AUC for small samples

© S.Yu. Gus'kov, V.V. Lyovin

JSC “Bank ZENITH”, Moscow, 127566, Russia

Polynomial distribution being presented as conditional joint distribution of independent Poisson random variables we build confidence intervals for sum polygons based on grouped data. We then use these estimates to build confidence intervals for ROC curves. These estimations then could be used in automatic defect detection and quality control procedures to find and to identify inhomogeneities and anomalies in structure of constructional materials and their elements for the end to improve robustness and efficiency of these procedures for small samples.

Keywords: *confidence intervals, sum polygons, connection between polynomial distribution and Poisson distribution, ROC curves, binary classifiers.*

REFERENCES

- [1] Engelmann B., Hayden E., Tasche D. Testing rating accuracy. *RISK*, 2003, vol. 16, pp. 82–86.
- [2] Stein R.M. Benchmarking default prediction models pitfalls and remedies in model validation. *J. of Risk Model Validation*, 2007, vol.1, no. 1, pp. 77–113.
- [3] Bol'shev L.N. *Teoriya veroyatnostey i ee primeneniye — Theory of Probability and its Applications*, 1962, vol. 7, pp. 353–355.
- [4] Bol'shev L.N. *Teoriya veroyatnostey i ee primeneniye — Theory of Probability and its Applications*, 1965, vol. 7, pp. 356–358.
- [5] Garwood F. Fiducial limits for Poisson distribution. *Biometrika*, 1936, vol. 28, pp. 437–442.
- [6] Stevens W.L. Fiducial limits of the parameter of discontinuous distribution. *Biometrika*, 1950, vol. 37, pp. 117–129.
- [7] Sofus A. Maeskassy and Foster Provost, Confidence Bands for ROC Curves: Methods and an Empirical Study. *CeDER Working Paper 02–04*. Stern School of Business, New York University, Jan. 2004, 15 p.
- [8] Jokiel-Rokita A., Pulit M. Nonparametric estimation of the ROC curve based on smoothed empirical distribution functions. *Statistical Computing*, 2013, vol. 23, pp. 703–712.
- [9] Baklizi A. A Simple Method for Finding Empirical Likelihood Type Intervals for the ROC Curve. *J. of Modern Applied Statistical Methods*, 2007, vol. 6, no. 2, pp. 589–595.
- [10] Le Meur Y., Vignolle J.-M., Chanussot J. Practical use of receiver operating characteristic analysis to assess the performances of defect detection algorithms. *J. of Electronic Imaging, Society of Photo-optical Instrumentation Engineers (SPIE)*, 2008, vol. 17, no. 3, pp.10.1117.
- [11] Dobrzański L.A., Krupiński M., Sokolowski J.H. Methodology of automatic quality control of aluminium castings. *J. of Achievements in Materials and Manufacturing Engineering*, 2007, vol. 20, no. 1–2, pp. 69–78.

- [12] Shchegolev S.S., Motkov A.V. The complex automated method of detection of defects of rings of bearings. On-line journal “*Naukovedenie*”, July–August 2014, issue 4 (23). Available at: <http://naukovedenie.ru/PDF/110TVN414.pdf> (accessed 25 November 2014).
- [13] Ivanov V.I., Konovalov N.N., Dergachev A.N. *Tekhnologii tekhnosfernoi bezopasnosti — Technospheric safety technology*, 2014, issue 6 (58). Available at: <http://agps-2006.narod.ru/ttb/2014-6/15-06-14.ttb.pdf> (accessed 25 November 2014).

Gus'kov S.Yu. (b. 1992), economist at JSC “Bank ZENITH”. Scientific interests: applying mathematical methods for solving economic problems.
e-mail: ser0432@yandex.ru

Lyovin V.V. (b. 1951), Ph.D. (Phys. & Math.), associate professor, head of the Scoring Department of JSC “Bank ZENITH”. Scientific interests: limiting theorems of probability theory and application of these for developing optimal statistical procedures.
e-mail: vladimir.levin.51@mail.ru; v.levin@zenit.ru