

Math-Net.Ru

Общероссийский математический портал

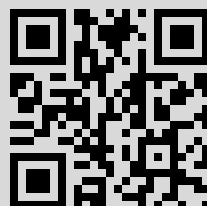
А. Н. Колмогоров, Метод медианы в теории ошибок, *Матем. сб.*, 1931, том 38, номер 3-4, 47–50

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 193.232.24.201

5 июня 2018 г., 17:33:58



Метод медианы в теории ошибок

А. Н. Колмогоров (Москва)

При допущении нормального закона распределения ошибок метод средней арифметической, как хорошо известно, является наилучшим для вычисления истинного значения наблюдаемой величины. Метод медианы в этом случае хуже, хотя, как показал Нааг, и незначительно. Однако, если гипотеза нормального распределения не удовлетворяет фактам, встает вопрос об отыскании наилучшего метода, соответствующего данному закону распределения. В частности во многих случаях, в которых считается необходимым исключение «анормальных наблюдений», было бы методологически более правильно исследовать общий закон распределения и найти наиболее соответствующий ему метод вычисления истинного значения.

В этой статье я хочу указать, как, зная закон распределения ошибок, определить степень точности метода медианы и сравнить ее с точностью метода средней арифметической. Какой из двух методов окажется предпочтительней, зависит от характера принятого закона распределения; однако, в силу теоремы II, *при совершенно неизвестном и могущем сильно уклоняться от нормального законе распределения надежнее употреблять метод медианы.*

Метод исследования в существенных чертах принадлежит Нааг'у¹ применившему его, однако, только к случаю нормального закона распределения.

Пусть вероятность ошибке отдельного наблюдения лежать между x и $x + dx$ равна $f(x)$. Мы будем предполагать, что $f(x)$ непрерывна и $f(m) \neq 0$ для медианы m :

$$\int_{-\infty}^m f(x) dx = \frac{1}{2}.$$

Пусть, далее

$$a = \int_{-\infty}^{+\infty} x f(x) dx$$

— математическое ожидание ошибки. Обозначим через $x_1, x_2, \dots, x_n$ результаты n последовательных наблюдений, через a_n — их среднее арифметическое и через m_n — их медиану. Мы ограничимся случаем нечетного $n = 2k + 1$, когда $m_n = x_{k+1}$,

¹ Comptes Rendus, 179, p. 1388 и Applications au tir, note II, в Traité du calcul des probabilités, Borel.

предполагая, что x_i занумерованы в порядке их возрастания. Разности $(a_n - a)$ и $(m_n - m)$ будут (как это следует из дальнейшего) порядка $\frac{1}{\sqrt{n}}$, поэтому естественно положить:

$$\begin{aligned} \alpha_n &= (a_n - a) \sqrt{n}, \\ \mu_n &= (m_n - m) \sqrt{n}, \end{aligned}$$

и исследовать законы распределения α_n и μ_n . Для α_n , как известно, имеем непрерывный закон распределения: вероятность α_n лежать между α и $\alpha + d\alpha$ равняется $\varphi_n(\alpha) d\alpha$, где $\varphi_n(\alpha)$ вычисляется по хорошо известным формулам. Если дисперсия

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - a)^2 f(x) dx$$

существует, то $\varphi_n(\alpha)$ сходится к нормальному закону распределения

$$\varphi(\alpha) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\alpha^2}{2\sigma^2}}$$

с дисперсией σ .

Вероятность m_n лежать между t и $t + dt$ равняется:

$$u_n(t) dt = \frac{(2k+1)!}{(k!)^2} \left(\frac{1 - \Theta^2(t)}{4} \right)^k f(t) dt,$$

где

$$\Theta(t) = 2 \int_m^t f(s) ds.$$

Вероятность же μ_n лежать между μ и $\mu + d\mu$ равна:

$$\begin{aligned} \psi_n(\mu) &= u_n(t) \frac{d\mu}{\sqrt{2k+1}}, \\ \mu &= (t - m) \sqrt{2k+1}, \\ t &= m + \frac{\mu}{\sqrt{2k+1}}. \end{aligned}$$

Теорема 1. $\psi_n(\mu)$ с увеличением n стремится к нормальному закону распределения

$$\psi(\mu) = \frac{1}{\sqrt{2\pi}\sigma_m} e^{-\frac{\mu^2}{2\sigma_m^2}},$$

с дисперсией

$$\sigma_m^2 = \frac{1}{2f(m)}.$$

В самом деле:

$$\psi_n(\mu) = \frac{\sqrt{2k+1} (2k)!}{(k!)^2 4^k} \{1 - \Theta^2(t)\}^k f(t).$$

По формуле Стирлинга

$$\frac{\sqrt{2k+1} (2k)!}{(k!)^2 4^k} \rightarrow \sqrt{\frac{2}{\pi}}.$$

² См. вывод аналогичной формулы (4) у Нагъа.

Далее ясно, что

$$f(t) \rightarrow f(m).$$

Остается рассмотреть множитель $\{1 - \Theta^2(t)\}^k$. Во-первых, имеем

$$\Theta(t) = m + \frac{\mu}{\sqrt{2k+1}} = \frac{\mu f(m)}{\sqrt{2k+1}} (1 + \varepsilon),$$

$$\int_m^{\infty} f(s) ds$$

где ε стремится к нулю с возрастанием n . Следовательно,

$$\Theta^2(t) = \frac{\mu^2 f(m)}{2k} (1 + \varepsilon'),$$

$$\{1 - \Theta^2(t)\}^k = e^{-\frac{\mu^2 f^2(m)}{2} (1 + \varepsilon'')},$$

где ε' и ε'' тоже стремятся к нулю. Сопоставляя полученные выражения, получаем

$$\psi_n(\mu) \rightarrow \sqrt{\frac{2}{\pi}} f(m) e^{-\frac{\mu^2 f^2(m)}{2}},$$

что и доказывает нашу теорему.

Мы видим таким образом, что при неограниченном возрастании n разности $m_n - m$ и $a_n - a$ имеют бесконечно малые дисперсии. Следовательно, систематической ошибкой в методе медианы будет m , а в методе средней арифметической a . Что же касается сравнительной точности обоих методов, то все решается отношением

$$\lambda = \frac{\sigma_m}{\sigma} = \frac{1}{2f(m)\tau}.$$

В случае нормального распределения $f(x)$ отношение λ вычислено еще у Наад'а и равно

$$\lambda_g = \sqrt{\frac{\pi}{2}} \sim \frac{5}{4}.$$

Легко построить примеры законов распределения, для которых λ принимает любые значения между 0 и ∞ . Однако имеет место следующая теорема.

Теорема II. Для законов распределения с одним максимумом отношение λ может принимать любые значения в промежутке

$$0 < \lambda < \sqrt{3},$$

но не может быть больше $\sqrt{3}$.

Максимальное значение $\sqrt{3}$ достигается, если

$$f(x) = a, \quad x \leq \frac{1}{2a},$$

$$f(x) = 0, \quad x > \frac{1}{2a}.$$

но этот закон распределения уже не удовлетворяет условию непрерывности: $\lambda = 0$, если $\sigma = +\infty$.

La méthode de la médiane dans la théorie des erreurs

Par A. Kolmogoroff (Moscou)
(Résumé)

Soit $f(x)dx$ la probabilité de l'erreur comprise entre x et $x+dx$, m la médiane de cette distribution et m_n la médiane des n ($n=2k+1$) observations. La loi de distribution pour la quantité

$$\mu_n = (m_n - m) \sqrt{n}$$

tend, quand $n \rightarrow \infty$, vers la loi de Gauss avec la dispersion

$$\sigma_m = \frac{1}{2f(m)},$$

si seulement $f(x)$ est continue et $f(m) \neq 0$.
