# О НОВЫХ СВОЙСТВАХ РАСПРЕДЕЛЕНИЙ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

**Е. А. Лебедев**

*Киевский национальный университет имени Тараса Шевченко*
*Киев, Украина*
*E-mail: leb@unicyb.kiev.ua*

Рассмотрены основные распределения математической статистики: Стьюдента, $\chi^2$ и Снедекора – Фишера. Изучены их свойства с целью построения для них эффективных вычислительных процедур рекуррентного типа.

*Ключевые слова*: распределение Гаусса, Стьюдента, $\chi^2$, Снедекора – Фишера, значение функции распределения, квантиль.

## 1. ВВЕДЕНИЕ

Распределения Стьюдента, $\chi^2$ и Снедекора – Фишера широко используются в математической статистике и ее применениях (см., например, [1]). Проблема вычисления значений функции распределения (ф. р.) и квантилей для этих распределений возникает при создании практически любого программного продукта прикладной статистики. В настоящей работе предложен новый подход для решения этой проблемы. Его конечным результатом являются простые вычислительные алгоритмы рекуррентного типа для указанных распределений.

Пусть $S_n(x)$, $\chi_n^2(x)$, $F_{n_1,n_2}(x)$ – ф. р. Стьюдента, $\chi^2$ и Снедекора – Фишера соответственно, $n$, $n_1$, $n_2$ – число степеней свободы для рассматриваемых распределений. Через $\tau_n$, $\chi_n^2$, $\eta_{n_1,n_2}$ будем обозначать случайные величины, которые отвечают ф. р. $S_n(x)$, $\chi_n^2(x)$ и $F_{n_1,n_2}(x)$. Известно, что

$$\tau_n \stackrel{d}{=} \frac{\xi_0}{\sqrt{(\xi_1^2+\ldots+\xi_n^2)/n}}, \tag{1}$$

$$\chi_n^2 \stackrel{d}{=} \xi_1^2+\ldots+\xi_n^2, \tag{2}$$

$$\eta_{n_1,n_2} \stackrel{d}{=} \frac{(\xi_1^2+\ldots+\xi_{n_1}^2)/n_1}{(\xi_{n_1+1}^2+\ldots+\xi_{n_1+n_2}^2)/n_2}. \tag{3}$$

Основываясь на (1) – (3), можно предположить, что для $S_n(x)$, $\chi^2_n(x)$, $F_{n_1,n_2}(x)$ существуют явные формулы, которыми эти ф. р. представляются через $\Phi(x)$ и элементарные функции. В разделах 2, 3 настоящей работы эта гипотеза полностью подтверждается. Более того, следствием проведенного анализа распределений Стьюдента, $\chi^2$ и Снедекора – Фишера являются эффективные процедуры для их вычисления.

## 2. ТОЧНОЕ И АППРОКСИМАТИВНОЕ ПРЕДСТАВЛЕНИЕ РАСПРЕДЕЛЕНИЯ СТЬЮДЕНТА

Английский статистик В. Госсет при анализе случайных отклонений выборочного среднего от истинного среднего значения пришел к закону распределения, который имеет ф. р.

$$S_n(x) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \int_{-\infty}^{x} (1+\frac{z^2}{n})^{-\frac{n+1}{2}} dz,$$

где $\Gamma(\alpha)$, $\alpha > 0$ – Эйлеров интеграл второго рода.

Поскольку свои работы В. Госсет публиковал под псевдонимом Student, то в математической статистике распределение $S_n(x)$ известно как распределение Стьюдента. Натуральный параметр «$n$» – число степеней свободы распределения.

Для $x > 0$ ф. р. Стьюдента можно представить следующим образом:

$$S_n(x) = \frac{1}{2} + k_n I_n(x),$$

$$k_n = \frac{1}{\sqrt{\pi n}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}, \quad I_n(x) = \int_0^x (1+\frac{z^2}{n})^{-\frac{n+1}{2}} dz.$$

Величины $k_n$, $I_n(x)$ для любого натурального $n$ и действительного $x > 0$ можно находить следующим образом.

**Лемма 1.** Последовательности $k_n$ и $I_n(x)$, $n = 1, 2, \ldots$, $x > 0$, удовлетворяют рекуррентным соотношениям:

$$k_{n+2} = \frac{k_n}{\sqrt{1-\frac{1}{(n+1)^2}}}, \quad k_1 = 1/\pi, \quad k_2 = \sqrt{2}/4, \tag{4}$$

$$I_{n+2}(\sqrt{n+2}\,x) = \sqrt{1-\frac{1}{(n+1)^2}}\, I_n(\sqrt{n}\,x) + \frac{1}{n+1}\sqrt{n+2}\,x\,(1+x^2)^{-\frac{n+1}{2}}, \tag{5}$$

$$I_1(x) = arctg\, x, \quad I_2(\sqrt{2}\,x) = \sqrt{2}\,x\,(1+x^2)^{-\frac{1}{2}}. \tag{6}$$

Соотношения (4) – (6) можно решить и получить явный вид ф. р. Стьюдента. Очевидно, этот вид будет зависеть от четности числа степеней свободы.

**Следствие 1.** Если $n = 2m+1$, то для $x > 0$

$$S_{2m+1}(x) = \frac{1}{2} + \frac{1}{\pi} arctg \frac{x}{\sqrt{2m+1}} + \tag{7}$$

$$+ \frac{1}{\pi} \frac{x}{\sqrt{2m+1}} \sum_{p=1}^{m} \frac{\sqrt{2p+1}}{2p} \left(1 + \frac{x^2}{2m+1}\right)^{-p} \left( \prod_{s=1}^{p} \sqrt{1 - \frac{1}{4}\frac{1}{s^2}} \right)^{-1}.$$

Если $n = 2m$, то для $x > 0$

$$S_{2m}(x) = \frac{1}{2} + \frac{2x}{\sqrt{m}} \sum_{p=1}^{m} \frac{\sqrt{2p}}{2p-1} \left(1 + \frac{x^2}{2m}\right)^{-p+1/2} \left( \prod_{s=2}^{p} \sqrt{1 - \frac{1}{(2s-1)^2}} \right)^{-1}. \tag{8}$$

Рекуррентные соотношения (4) – (6) представляют собой эффективный алгоритм, который позволяет находить значение $S_n(x)$ за $[n/2]$ шагов, $[\cdot]$ – целая часть числа. При больших $n$ можно использовать аппроксимацию распределения Стьюдента гауссовским законом. То, что при $n \to \infty$ $S_n(x)$ слабо сходится к $\Phi(x)$, следует из (1). Для того чтобы для больших $n$ вместо $S_n(x)$ использовать $\Phi(x)$, необходимо оценить скорость сходимости. Приведем один из вариантов решения этой задачи.

**Теорема 1.** Пусть $\rho(S_n, \Phi) = \sup\limits_{-\infty < x < \infty} |S_n(x) - \Phi(x)|$. Тогда

$$\rho(S_n, \Phi) \le \frac{1}{\pi e} \frac{\sqrt{2}}{2} \left(1 - \frac{\sqrt{e}}{2}\right) \frac{1}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right).$$

Доказательство теоремы основано на следующей оценке.

**Лемма 2.** Для любого $x \ge 0$ и $0 < \alpha < 1/\sqrt{e}$

$$|S_n(x) - \Phi(x)| \le \frac{1}{4\pi} \sqrt{\frac{2}{e}} \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{\pi n}} (\alpha\sqrt{e})^n + \tag{9}$$

$$+ \frac{1}{2\pi} \sqrt{\frac{2}{e}} (\alpha^{-1} - 1)\left[1 - (\alpha e^{(1-\alpha^2)/2})^n\right] \frac{1}{\sqrt{n}}.$$

Оценка (9) позволяет по точности вычислений определить номер $n$, начиная с которого можно вместо ф. р. $S_n(x)$ использовать $\Phi(x)$.

Отметим, что предложенный в этом разделе рекуррентный алгоритм легко программно реализуем. Его можно использовать и для нахождения квантилей любого порядка $0 < \alpha < 1$. Для этого решение уравнения $S_n(x) = \alpha$ находим методом деления отрезка пополам.

## 3. ФОРМУЛЫ ДЛЯ РАСПРЕДЕЛЕНИЙ $\chi^2$ И СНЕДЕКОРА – ФИШЕРА

Впервые ф. р. $\chi^2_n(x)$ получил астроном Ф. Хельмерт, который имел дело с гауссовской теорией ошибок. Позднее К. Пирсон назвал эту ф. р. «хи квадрат».

Последовательность $\chi^2_n(x)$ удовлетворяет рекуррентному соотношению.

**Лемма 3.** Для любого $n = 1, 2, \ldots$.

$$\chi^2_{n+2}(x) = \chi^2_n(x) - \delta_n \frac{x^{n/2}}{n!!} e^{-x/2}, \tag{10}$$

$$\chi^2_1(x) = 2\,\Phi(\sqrt{x}) - 1,\ \chi^2_2(x) = 1 - e^{-x/2},$$

где

$$\delta_n = \begin{cases} 1, & n = 2k, \\ \sqrt{\dfrac{2}{\pi}}, & n = 2k+1. \end{cases}$$

Как и в случае распределения Стьюдента, рекуррентное соотношение (10) можно решить и получить явные формулы для $\chi^2_n(x)$, которые будут содержать элементарные функции и функцию $\Phi(x)$.

**Следствие 2.** Если $n = 2m$, $m = 1, 2, \ldots$, то

$$\chi^2_{2m}(x) = 1 - e^{-x/2} \sum_{p=0}^{m-1} \frac{(x/2)^p}{p!}. \tag{11}$$

Если $n = 2m+1$, $m = 0, 1, \ldots$, то

$$\chi^2_{2m+1}(x) = 2\,\Phi(\sqrt{x}) - 1 - \sqrt{\frac{2}{\pi}}\, e^{-x/2} \sum_{p=1}^{m} \frac{x^{p-1/2}}{(2p-1)!!}. \tag{12}$$

Рекуррентное соотношение (10) дает значение $\chi^2_n(x)$ за $[n/2]$ итераций. Использование этой процедуры и метода деления отрезка пополам позволяет находить квантили.

Перейдем теперь к ф. р. Снедекора – Фишера $F_{n_1,n_2}(x)$. Нам будут необходимы следующие ее свойства ([1]):

а) симметрия

$$F_{n_1,n_2}(x) = 1 - F_{n_2,n_1}(1/x);$$

в) связь с бетта-распределением

$$F_{n_1,n_2}(x) = 1 - B_{n_2/2,n_1/2}(y), \text{ где } y = \frac{n_2}{n_2 + n_1 x};$$

с) рекуррентное соотношение для бетта-распределения

$$B_{a,b}(x) = \frac{1}{B(a,b)} \frac{x^a (1-x)^{b-1}}{a} + B_{a+1,b-1}(x),$$

где $B(a,b) = \int_0^1 x^{a-1} (1-x)^{b-1}\, dx$ – интеграл Эйлера первого рода.

Свойства а) – с) позволяют найти $F_{n_1,n_2}(x)$ в явном виде.

**Лемма 4.** Ф. р. Снедекора – Фишера представима в виде:

$$F_{n_1,n_2}(x) = 1 - B_{n_2/2,n_1/2}(y),\ y = \frac{n_2}{n_2 + n_1 x}, \tag{13}$$

причем

1) если $n_1 = 2k$, $n_2 = 2p$, то

$$B_{n_2/2,\,n_1/2}(y) = B_{p,k}(y) = \tag{14}$$

$$= \sum_{i=0}^{k-1} y^{p+i} (1-y)^{k-1-i} \frac{(p+k-1)!}{(p+i)!(k-1-i)!};$$

2) если $n_1 = 2k$, $n_2 = 2p+1$, то

$$B_{n_2/2,\,n_1/2}(y) = B_{p+1/2,\,k}(y) = \tag{15}$$

$$= \sum_{i=0}^{k-1} y^{p+1/2+i} (1-y)^{k-1-i} \frac{\prod_{\nu=1}^{p+k} (\nu - 1/2)}{\prod_{\mu=1}^{p+i+1} (\mu - 1/2)} \frac{1}{(k-1-i)!};$$

3) если $n_1 = 2k+1$, $n_2 = 2p+1$, то

$$B_{n_2/2,\,n_1/2}(y) = B_{p+1/2,\,k+1/2}(y) = \tag{16}$$

$$= \frac{1}{\pi} \sum_{i=0}^{k-1} y^{p+1/2+i} (1-y)^{k-1/2-i} \frac{(k+p)!}{\prod_{\nu=1}^{k-i} (\nu - 1/2) \prod_{\mu=1}^{p+i+1} (\mu - 1/2)} +$$

$$+ 1 - \frac{2}{\pi} arctg\, z - \frac{2}{\pi} \sum_{i=1}^{p+k} \frac{\frac{1}{2i} \frac{z}{(1+z^2)^i}}{\prod_{j=1}^{i} (1 - \frac{1}{2j})},$$

где $z = (y^{-1} - 1)^{1/2}$.

Формулы (13) – (16) получены с помощью процедуры рекуррентного типа, которая вычисляет значения ф. р. Снедекора – Фишера.

## Алгоритм вычисления $F_{n_1,n_2}(x)$

1. Идентификация параметров распределения Снедекора – Фишера.
Устанавливаем, какой из четырех случаев имеет место:

1) $n_1 = 2k$, $n_2 = 2p$;

2) $n_1 = 2k$, $n_2 = 2p+1$;

3) $n_1 = 2k+1$, $n_2 = 2p$;

4) $n_1 = 2k+1$, $n_2 = 2p+1$.

2. Вычисление ф. р.

Подсчет $F_{n_1,n_2}(x)$ будем проводить для четырех комбинаций предыдущего шага:

$$1)\quad y=\frac{n_2}{n_2+n_1x};$$

$$A_0=p\ln y+(k-1)\ln(1-y)+\sum_{\nu=1}^{k-1}\ln(1+p/\nu),$$

$$A_{i+1}=A_i+\ln[\frac{y}{1-y}\frac{k-1-i}{p+1+i}],\ i=0,1,...,k-2;$$

$$B_{p,k}(y)=\sum_{i=0}^{k-1}\exp(A_i),\ F_{n_1,n_2}(x)=1-B_{p,k}(y).$$

$$2)\quad y=\frac{n_2}{n_2+n_1x};$$

$$B_0=(p+\frac{1}{2})\ln y+(k-1)\ln(1-y)+\sum_{\nu=1}^{k-1}\ln(1+\frac{p+1/2}{\nu}),$$

$$B_{i+1}=B_i+\ln[\frac{y}{1-y}\frac{k-1-i}{p+3/2+i}],\ i=0,1,...,k-2;$$

$$B_{p+1/2,k}(y)=\sum_{i=0}^{k-1}\exp(B_i),\ F_{n_1,n_2}(x)=1-B_{p+1/2,k}(y).$$

3) Вычисления проводим по схеме пункта 2) с параметрами

$$n_1=2p,\ n_2=2k+1,\ 1/x,$$

и в завершение $F_{n_1,n_2}(x)=1-F_{2p,2k+1}(1/x)$.

$$4)\quad y=\frac{n_2}{n_2+n_1x},\ z=(y^{-1}-1)^{1/2},$$

$$D_0=-\ln\pi+(p+1/2)\ln y+(k-1/2)\ln(1-y)+$$

$$+\sum_{\nu=1}^{p+1}\ln(\frac{\nu}{\nu-1/2})+\sum_{\mu=1}^{k-1}\ln(\frac{p+1+\mu}{\mu-1/2})-\ln(k-1/2),$$

$$D_{i+1}=D_i+\ln[\frac{y}{1-y}\frac{k-i-1/2}{p+i+3/2}],\ i=0,1,...,k-2;$$

$$C_1=\frac{z}{1+z^2},\ C_{i+1}=\frac{2i}{2i+1}\frac{1}{1+z^2}C_i,\ i=1,2,...,p+k-1;$$

$$B_{p+1/2,k+1/2}(y)=\sum_{i=0}^{k-1}\exp(D_i)+1-\frac{2}{\pi}arctgz-\frac{2}{\pi}\sum_{i=1}^{p+k}C_i,$$

$$F_{n_1,n_2}(x)=1-B_{p+1/2,k+1/2}(y).$$

## 4. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

В контексте явных формул (7), (8) для ф. р. Стьюдента и (13)–(16) для ф. р. Снедекора – Фишера отметим следующий интересный факт. Согласно (1), (3) эти распределения определяются через стандартное гауссовское распределение, причем, как известно, его ф. р. $\Phi(x)$ не выражается через элементарные функции. Несмотря на это, $S_n(x)$ и $F_{n_1,n_2}(x)$ имеют такое представление.

В случае ф. р. $\chi^2_n(x)$ ситуация несколько усложняется: выражение для $\chi^2_n(x)$ содержит $\Phi(x)$ для нечетного числа степеней свободы. Учитывая последний факт (а также аппроксимацию $S_n(x)$ для больших $n$ функцией $\Phi(x)$), кратко остановимся на способах вычисления гауссовской ф. р.

Для $x \geq 0$ функцию $\Phi(x)$ можно представить в виде суммы знакопеременного ряда

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-u^2/2}\, du = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} (-1)^k\, a_k(t), \qquad (17)$$

где $a_k(t) = \dfrac{t^k \sqrt{2t}}{k!(2k+1)}$, $t = x^2/2$.

Анализ $a_k(t)$, $k = 0, 1, \ldots$ показывает, что при $x > \sqrt{6}$ члены этой последовательности возрастают на начальном отрезке номеров и достигают максимального значения при $k^\bullet = k^\bullet(x) = [(x^2 - 5 + \sqrt{x^4 - 6x^2 + 1})/4] + 1$, $[\cdot]$ – целая часть числа, причем

$$a_{k^\bullet}(x^2/2) = \frac{1}{\sqrt{\pi}} \exp(x^2/2 + o(x^2))\,(1 + o(x^2)). \qquad (18)$$

Так, при $x = 4$, $a_{k^\bullet} \cong 112$. Следовательно, с увеличением аргумента $x \geq 0$ значение $1/2 \leq \Phi(x) < 1$ получается в результате взаимного погашения больших чисел. Это, конечно, недостаток формулы (17) и при ее использовании необходимо ограничивать интервал для $x$ (например, рассматривать $x \in [0, 3]$). За пределами этого интервала для вычисления гауссовской ф. р. можно использовать разложение $[1 - \Phi(x)]/\Phi'(x)$ в цепную дробь (см. [2], с. 110).

## ЛИТЕРАТУРА

1. *Айвазян С. А., Евнюков И. С., Мешалкин Л. Д.* Прикладная статистика: Основы моделирования и первичная обработка данных. М.: Финансы и статистика, 1983. С. 471.

2. *Королюк В. С., Портенко Н. И., Скороход А. В., Турбин А. Ф.* Справочник по теории вероятностей и математической статистике. Киев: Наукова думка, 1978. С. 584.