6.5. Распределение «хи-квадрат»

Пусть X_i (i=1,…,n) – система независимых нормированных нормально распределенных СВ с МО, равным 0, и единичной дисперсий. Тогда СВ χ^2 , представляющая собой сумму квадратов этих величин,

$$\chi^2 = \sum_{i=1}^{n} x_i^2$$

распределена по закону “хи-квадрат” с k=n степенями свободы.

Если на величины X_i (i=1,…,n) наложено r связей, то число степеней свободы $k=n-r$.

Плотность этого распределения определяется: $0 \le \chi^2 < \infty$,

$$p(\chi^2) = \frac{1}{2^{k/2}\Gamma(k/2)}(\chi^2)^{(k/2)-1}\exp(-\chi^2/2)$$

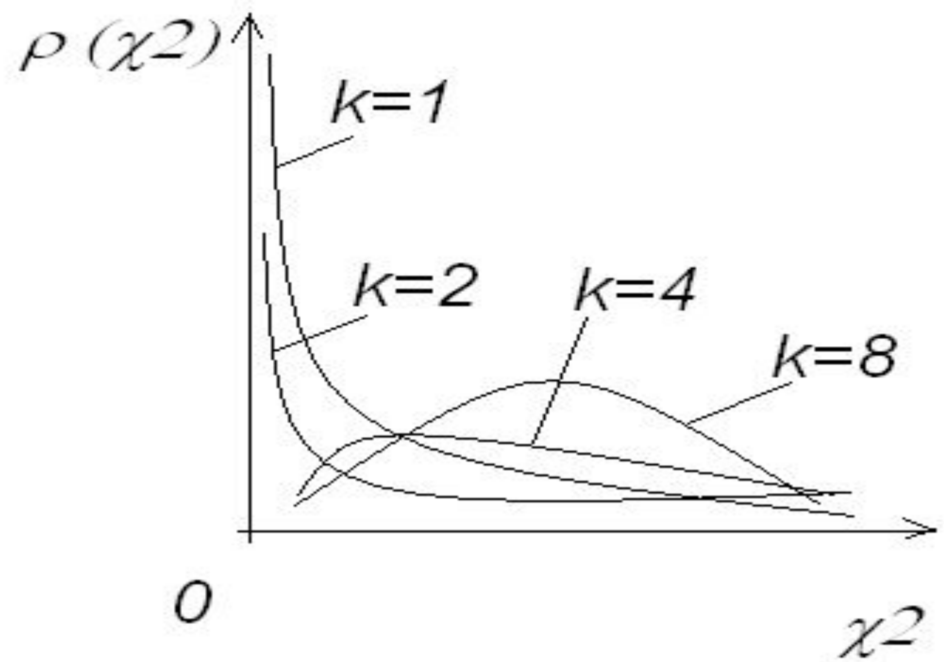
где

$$\Gamma(x) = \int_0^\infty t^{x-1} exp(-t)dt$$

гамма-функция (интеграл Эйлера второго рода). В частности $\Gamma(n+1)=n!$.

Из определения плотности вероятности распределения χ^2 следует, что распределение “хи-квадрат” определяется одним параметром – числом степеней свободы k. С увеличением числа степеней свободы распределение “хи-квадрат” медленно приближается к нормальному.

При $k=n>30$ χ^2 – распределение достаточно хорошо представляется нормальным законом с $M[\chi^2]=n$ и $D[\chi^2]=n$. На рисунке показано, как изменяется характер распределения χ^2 при увеличении числа степеней свободы k.



6.6. Распределение Стьюдента

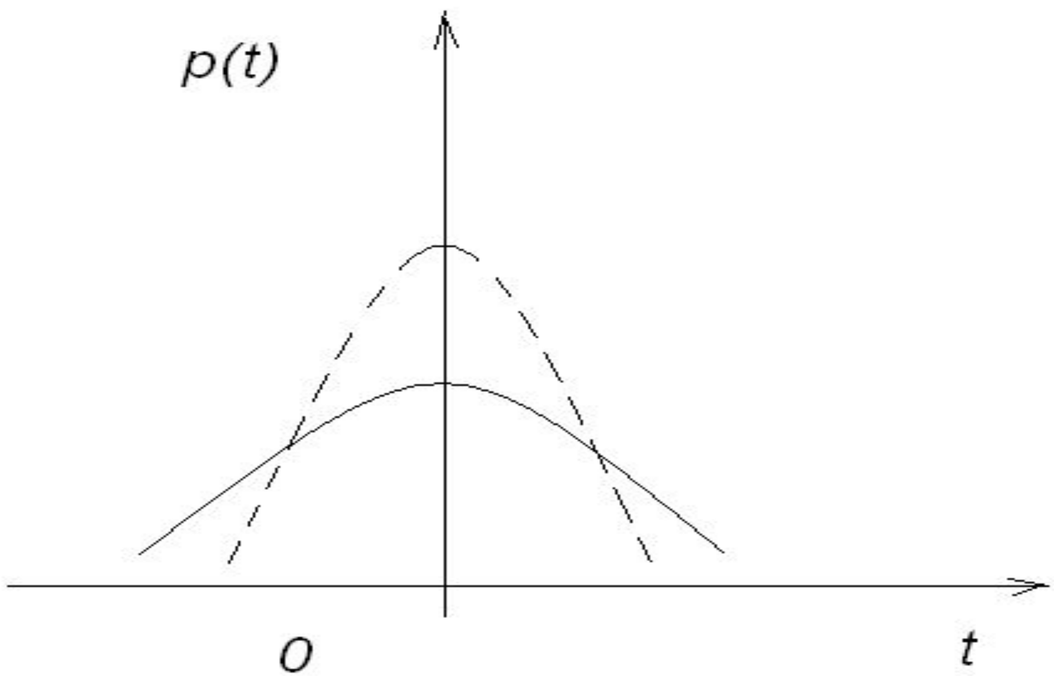
Пусть случайные величины Z, X_1, X_2, …, X_n подчинены нормальному закону распределения с нулевым средним и произвольной дисперсией. Пусть далее величина Z не зависит от X_i, $i = 1, \ldots, n$, и среди X_i имеется ровно k линейно независимых величин.

Тогда случайная величина

$$t = \frac{Z}{\sqrt{\frac{1}{k}\sum_{i=1}^{n} X_i^2}}$$

имеет распределение Стьюдента (*t*-распределение), с плотностью распределения

$$p(t) = \frac{1(\Gamma(k+1)/2)}{\sqrt{k\pi}\,\Gamma(k/2)}\left(1+\frac{t^2}{k}\right)^{-(k+1)/2}$$



Заметим, что t-распределение не зависит от σ^2. Величина t, определенная для нормированных случайных величин с нулевым средним и единичной дисперсией, также распределена по закону Стьюдента.

Распределение Стьюдента симметрично относительно начала координат. С возрастанием числа степеней свободы быстро приближается к нормальному закону распределения.

Для нормированных СВ распределения Стьюдента приближается к нормальному закону с характеристиками $M[t] = 0$ и $D[t] = k / (k - 2)$.

6.7. F-распределение Фишера

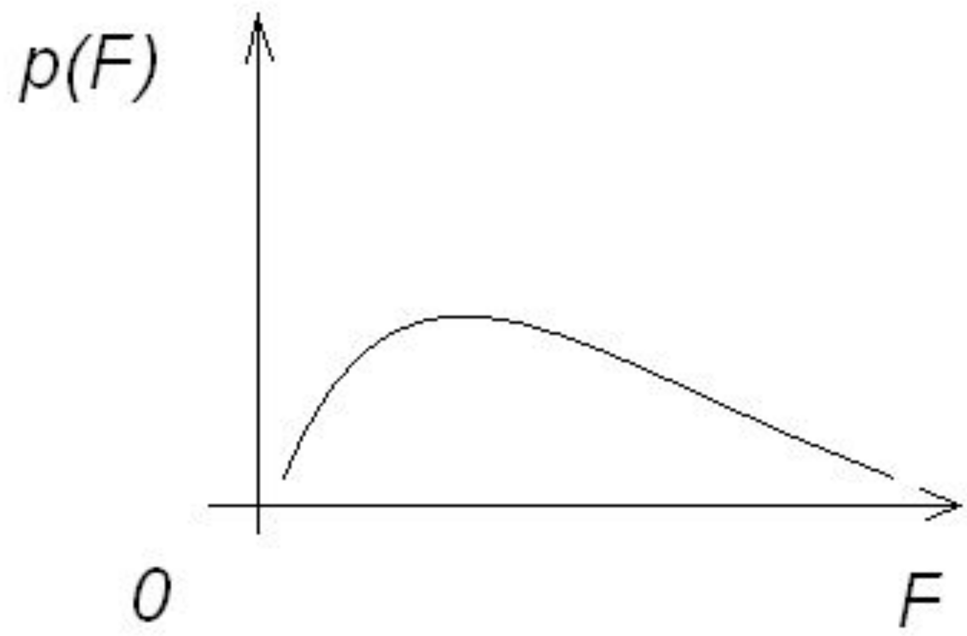
Если X и Y – независимые случайные величины, распределенные по закону χ^2 со степенями свободы k_1 и k_2 , то величина

$$F = \frac{X / k_1}{Y / k_2}$$

имеет F-распределение Фишера со степенями свободы k_1 и k_2.

Плотность этого распределения определяется выражением

$$p(F)=\frac{k_1^{k_1/2}k_2^{k_2/2}\Gamma((k_1+k_2)/2)}{\Gamma(k_1/2)\Gamma(k_2/2)}F^{(k_1-2)/2}(k_2+k_1F)^{-(k_1+k_2)/2}$$



F-распределение Фишера характеризуется 2 параметрами - числами степенями свободы k_1 и k_2.

6.8. Первичная обработка результатов измерений

Первичная обработка результатов измерений состоит из последовательного выполнения следующих шагов.

1. Построение случайной выборки измерений и простого статистического ряда.
2. Построение вариационного ряда
3. Грубые ошибки измерений. Исключение грубых ошибок.
4. Оценка математического ожидания случайной величины.
5. Оценка дисперсии случайной величины.
6. Оценка вероятности случайного события.
7. Оценка функции и плотности распределения случайной величины.

Рассмотрим более детально вопросы исключения грубых ошибок и оценки вероятности случайного события.

Получив выборку наблюдений случайной величины X с функцией распределения $F(x)$ следует убедиться, что она действительно соответствует этой функции распределения.

Так как в процессе измерений предполагаемая статистическая обстановка может нарушиться и среди реализаций x_i могут появляться ошибочные, т.е. не соответствующие $F(x)$ значения.

Обычно в качестве ошибочных подразумевают x_{min} и x_{max} и их называют грубыми ошибками, если установлено их несоответствие закону F(x).

Если $F(x)$ известно, то вопрос об ошибочности x_{max} может быть решен следующим образом. Зная $F(x)$, можно найти

$F_{(n)}(x)$ – функцию распределения

$$X_{(n)} = X_{\max}.$$

Тогда задаваясь вероятностью $\beta \cong 1$ практически достоверного события, из уравнения

$$P(X_{\max} < t_\beta) = F_{(n)}(t_\beta) = \beta$$

можно найти границу t_{β}, правее которой появление реализации x_{max} в соответствии с принципом практической уверенности невозможно.

Отсюда следует решающее правило: если $x_{max} \geq t_{\beta}$, то x_{max} считают грубой ошибкой, в противном случае x_{max} считают согласующейся с законом распределения $F(x)$.

В случае независимых измерений

$$F(t_{\beta}) = \sqrt[n]{\beta}$$

Аналогично решается вопрос об ошибочности x_{min} . Здесь определяется граница t_α из условия:

$$P(X_{min} < t_\alpha) = F_{(1)}(t_\alpha) = \alpha$$

где α=1-β - вероятность практически невозможного события.

Затем применяют решающее правило принципа практической уверенности:

x_{min} – грубая ошибка, если $x_{min} < t_\alpha$; x_{min} не противоречит $F(x)$ – в противном случае.

При независимых измерениях t_α находится из уравнения:

$$F(t_\alpha) = 1 - \sqrt[n]{1 - \alpha}$$

Чаще $F(x)$ бывает неизвестной. Тогда для решения поставленной задачи применяют частные приемы.

Например, если $F(x)$ нормального закона распределения с неизвестными параметрами $m = M[X]$ и $\sigma^2 = D[X]$, то строят вспомогательную случайную величину

$$T = \frac{X_{max} - \bar{X}}{S}$$

где $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ – оценка среднеквадратического отклонения

Затем устанавливают ее функцию распределения $F_T(t) = P(T < t)$ далее находят верхнюю границу t_β допустимых значений T из уравнения $F_T(t_\beta) = \beta = 1 - \alpha$.

Верхней границей допустимых значений x_{max} становится величина

$$\bar{x} + st_\beta$$

В итоге получаем следующее частное решающее правило: если

$$x_{max} < \bar{x} + st_{\beta}$$

то она считается соответствующей нормальному распределению;

в противном случае величина x_{max} считается грубой ошибкой.

Анализ ошибочности $x_{\min}$ выполняется аналогично по решающему правилу:

$$\mathrm{x}_{\min} \geq \bar{x} - st_{\beta}$$

то $x_{\min}$ считается соответствующей нормальному закону; в противном случае величину $x_{\min}$ считают грубой ошибкой. Для определения границ t_{β} составлены специальные таблицы, входом которых служат n и α=1-β.

Оценим вероятность $P(A)=p$ появления события A в n опытах.

В качестве оценки рассмотрим ч*астоту событий*:

$$p^* = m^*/n,$$

где m^* - число опытов, в которых наблюдалось событие A,

n – общее число опытов.

Из т.Бернулли следует, что оценка вероятности события р* является состоятельной, является оценкой сходящейся по вероятности к оцениваемому параметру.

Определим математическое ожидание и дисперсию оценки p^*. Т.к. m^* - случайная величина, распределенная по биномиальному закону с математическим ожиданием

$M[m^*] = np$ и дисперсией $D[m^*] = npq$, то

$$M(p^*) = M\left(\frac{m^*}{n}\right) = \frac{1}{n} M(m^*) = \frac{np}{n} = p$$

$$D(p^*) = D\left(\frac{m^*}{n}\right) = \frac{1}{n^2} D(m^*) = \frac{npq}{n^2} = \frac{pq}{n}$$

Т.о., оценка вероятности случайного события p^* является также несмещенной, т.е. оценкой, математическое ожидание которой равно оцениваемому параметру, и асимптотически эффективной, т.е. состоятельной оценкой, дисперсия которой с увеличением объема выборки стремится к нулю.