Лекция 7 16.04.2012

Определение. Распределение случайной величины

$$F(m,n)=\frac{1/m\sum_{i=1}^{m}U_i^2}{1/n\sum_{j=1}^{n}Z_j^2}=\frac{1/m\,\chi^2(m)}{1/n\,\chi^2(n)}$$ называется F–распределением (распределением Фишера) с (m, n) степенями свободы.

Лемма Фишера

Пусть $X_1,...,X_n$ – выборка из нормальной генеральной совокупности, т.е. $X_i \sim N(m,\sigma^2)$, и $X_1,...,X_n$ независимы. Иными словами, $X=[X_1,...,X_n]' \sim N\left(\begin{bmatrix} m \\ \vdots \\ m \end{bmatrix}, \sigma^2 I_n\right)$. Обозначим $\iota_n=[\underbrace{1,...,1}_{n}]'$. Тогда $X \sim N\left(m\,\iota_n,\, \sigma^2 I_n\right)$. Обозначим также

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\ s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 .$$

Теорема (лемма Фишера). 1) Случайные величины $\bar{X}$ и s^2 независимы.
2) Случайная величина $\frac{n-1}{\sigma^2}s^2$ имеет распределение хи-квадрат с $n-1$ степенями свободы.
3) Случайная величина $t=\frac{(\bar{X}-m)\sqrt{n}}{s}$ имеет t–распределение с $n-1$ степенями свободы.

Лемма. 1) Пусть Z– стандартный нормальный n-вектор O – ортогональная $n\times n$-матрица. Тогда вектор $U=OZ$ является стандартным нормальным вектором.

2) Пусть D – идемпотетнтная матрица, $rank(D)=r$. Тогда $Z'DZ \sim \chi^2(r)$.

Доказательство. 1) Задача 3, домашнее задание 5.

2) Существует ортогональная матрица O, такая что $O'DO=\Lambda$, где Λ – диагональная матрица, на главной диагонали которой стоит r единиц и $n-r$ нулей. Тогда $Z'DZ=(OZ)'\Lambda(OZ)=U'\Lambda U=\sum_{i=1}^{r}U_i^2$, где U – стандартный нормальный вектор. Отсюда $Z'DZ \sim \chi^2(r)$.

Доказательство теоремы. 1) Введем $n\times n$ матрицы

$A=\frac{1}{n}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}=\frac{1}{n}\iota_n\iota_n',\ B=I_n-A$. Это идемпотентные матрицы рангов 1 и $n-1$ соответственно. Имеем $Ax=[\bar{x},...,\bar{x}]'=\bar{x}\cdot\iota_n,\ Bx=[x_1-\bar{x},...,x_n-\bar{x}]',\ x\in R^n$. Далее,

$Cov(AX, BX) = AV(X)B' = \sigma^2 AB = 0$. Поэтому в силу нормальности AX и BX независимы. Поэтому независимы и функции от них: $\bar{X}$ и $s^2 = \frac{1}{n-1}(BX)'(BX)$.

2) Ясно, что $Z_i = \frac{X_i - m}{\sigma} \sim N(0,1)$, вектор $Z = [Z_1, ..., Z_n]'$ является стандартным нормальным, и $\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sigma^2 \sum_{i=1}^{n}(Z_i - \bar{Z})^2 = \sigma^2 (BZ)'(BZ) = \sigma^2 Z' BZ = \sigma^2 \chi^2(n-1)$ в силу леммы.

3) Наконец, $t = \frac{(\bar{X} - m)\sqrt{n}}{s} = \frac{\bar{Z}\sqrt{n}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2}} = \frac{N(0,1)}{\sqrt{1/(n-1)\,\chi^2(n-1)}} = t(n-1)$, поскольку числитель и знаменатель независимы (в силу п. 1).

Предельные теоремы

Определение. Последовательность $\{\xi_n, n = 1,2,...\}$ случайных величин *сходится по вероятности* к сл.в. ξ, если $\forall \varepsilon > 0 \;\; P(|\xi_n - \xi| > \varepsilon) \to 0,\; n \to \infty$, обозначение $p\lim_{n\to\infty} \xi_n = \xi$.

Определение. Последовательность $\{\xi_n, n = 1,2,...\}$ случайных величин *сходится почти наверное* к сл.в. ξ, если $\lim\limits_{n\to\infty} \xi_n(\omega) = \xi(\omega)$ для почти всех $\omega \in \Omega$, обозначение $\lim_{n\to\infty} \xi_n = \xi$ п.н..

Определение. Последовательность $\{\xi_n, n = 1,2,...\}$ случайных величин *сходится в среднеквадратичном* к сл.в. ξ, если $\lim_{n\to\infty} E|\xi_n - \xi|^2 = 0$.

Определение. Последовательность $\{\xi_n, n = 1,2,...\}$ случайных величин *сходится по распределению* к сл.в. ξ, если $\lim_{n\to\infty} F_n(x) = F(x)$ в каждой точке непрерывности функции $F(x)$, где $F_n(x),\; F(x)$ – функции распределения случайных величин $\xi_n,\; \xi$, соответственно.

Можно показать, что

- из сходимости почти наверное вытекает сходимость по вероятности;
- если $p\lim_{n\to\infty} \xi_n = \xi$, то существует подпоследовательность $\{k_n\},\; n = 1, 2, ...$, такая что $\lim_{n\to\infty} \xi_{k_n} = \xi$ п.н.
- из сходимости в среднеквадратичном вытекает сходимость по вероятности
- из сходимости по вероятности вытекает сходимость по распределению
- сходимость по распределению эквивалентна следующим условиям
  - $E\left(f(\xi_n)\right) \to E\left(f(\xi)\right)$ для любой непрерывной ограниченной функции $f(\cdot)$
  - $\overline{\lim}\, P(\xi_n \in A) \le P(\xi \in A)$ для любого замкнутого множества A
  - $\underline{\lim}\, P(\xi_n \in B) \ge P(\xi \in B)$ для любого открытого множества B

Неравенство Чебышева. Пусть $\xi \ge 0$ – неотрицательная случайная величина. Тогда $P(\xi \ge a) \le \frac{E(\xi)}{a}$.

Доказательство. $E(\xi) = \int_0^\infty xf(x)dx \ge \int_a^\infty xf(x)dx \ge a\int_a^\infty f(x)dx = aP(\xi \ge a)$.

Следствие. $P(|X-m_X|>\varepsilon)\le\dfrac{V(X)}{\varepsilon^2}$.

Закон больших чисел

Теорема. Пусть $\{\xi_n, n=1,2,...\}$ – последовательность <u>независимых</u> сл. вел., $E(\xi_i)=m_i$, $V(\xi_i)=\sigma_i^2$ и пусть

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2=0. \qquad (*)$$

Тогда $p\lim_{n\to\infty}\left(\dfrac{\xi_1+...+\xi_n}{n}-\dfrac{m_1+...+m_n}{n}\right)=0$. В частности, если $\{\xi_n, n=1,2,...\}$ одинаково распределены, то $p\lim_{n\to\infty}\left(\dfrac{\xi_1+...+\xi_n}{n}\right)=m$.

Доказательство. Обозначим $X_n=\dfrac{\xi_1+...+\xi_n}{n}-\dfrac{m_1+...+m_n}{n}$. Тогда $E(X_n)=0,\ V(X_n)=\dfrac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2$. Из неравенства Чебышева следует, что $P(|X_n|>\varepsilon)\le\dfrac{V(X_n)}{\varepsilon^2}\to 0,\ n\to\infty$ в силу (*).

Пример. Пусть последовательно и независимо реализуется эксперимент со случайным исходом, и пусть m_n – число появлений события *A* в *n* опытах. Тогда m_n – это биномиальная случайная величина (n,p), где $p=\mathrm{P}(A)$. Тогда $\dfrac{m_n}{n}\to p,\ n\to\infty$, что служит оправданием устойчивости частот.

Центральная предельная теорема

Теорема (непрерывности, без доказательства). Пусть $\{F_n(x),\ x\in R,\ n=1,2,...\}$ – последовательность функций распределения и $\{\varphi_n(t),\ t\in R,\ n=1,2,...\}$ – соответствующая последовательность характеристических функций.

1) Если $F_n\xrightarrow{w}F$, где *F* – некоторая функция распределения, то $\varphi_n(t)\to\varphi(t),\ t\in R$, где $\varphi(t)$ – характеристическая функция *F*.

2) Если при каждом $t\in R$ существует $\lim\limits_{n\to\infty}\varphi_n(t)=\varphi(t)$ и функция $\varphi(t)$ непрерывная в точке $t=0$, то $\varphi(t)$ является характеристической функцией некоторого распределения *F*, и $F_n\xrightarrow{w}F$.

Пусть $\{\xi_n, n=1,2,...\}$ – последовательность <u>независимых одинаково распределенных</u> сл. вел., $E(\xi_i)=m$, $V(\xi_i)=\sigma^2,\ E|\xi_i^3|<+\infty$ Обозначим

$$T_n=\frac{S_n-mn}{\sigma\sqrt{n}},\ S_n=\sum_{i=1}^{n}\xi_i .$$

Центральная предельная теорема. Последовательность случайных величин $\{T_n\}$ сходится по распределению к стандартной нормальной величине. В частности, $P\left(T_n\le x\right)\to\Phi(x)$ для любого *x*.

Частный случай. Пусть $X_n \sim B(n,p)$ – биномиальная случайная величина с параметрами (n,p). Тогда $X_n = \sum_{i=1}^{n} \varepsilon_i$, где $\{\varepsilon_i,\, i=1,\dots,n\}$ – независимые бернуллиевские величины с параметром *p*, $E(\varepsilon_i) = p,\, V(\varepsilon_i) = p(1-p)$. Таким образом, к величинам X_n применима ЦПТ: $\dfrac{X_n - np}{\sqrt{np(1-p)}} \approx Z \sim N(0,1)$. Разделив числитель и знаменатель дроби на *n*, получим:

$\dfrac{\hat{p}_n - p}{\sqrt{\dfrac{p(1-p)}{n}}} \approx Z \sim N(0,1)$, где $\hat{p}_n = \dfrac{X_n}{n}$ – выборочная пропорция.

Пример. В 1980 около 50% американцев говорили, что инфляция является самой серьезной для страны проблемой. Каков шанс того, что выборочная пропорция будет отражать пропорцию генеральной совокупности с точностью 3 процента для случайного опроса 1500 американцев (размер, соответствующий опросам Гэллапа)?

Решение. Обозначим через *X* количество человек из 1500 опрошенных, которые считали, что инфляция является наиболее важной проблемой государства. Тогда *X* имеет биномиальное распределение *Bi*(1500;0.5). $\hat{p}$ – выборочная пропорция,

$\mathrm{E}\left(\hat{p}\right) = p = 0.5, \quad \mathrm{SE} = \sqrt{\dfrac{p(1-p)}{n}} = 0.013$. Используем нормальное приближение:

$$\mathrm{P}(0.47 \le \hat{p} \le 0.53) = \mathrm{P}\left(\frac{0.47-0.5}{0.013} \le \frac{\hat{p}-0.5}{0.013} \le \frac{0.53-0.5}{0.013}\right) = \mathrm{P}(-2.32 \le Z \le 2.32) = 0.980 .$$