Теория вероятностей и математическая статистика

Лекция 16

Некоторые распределения и интервальные оценки

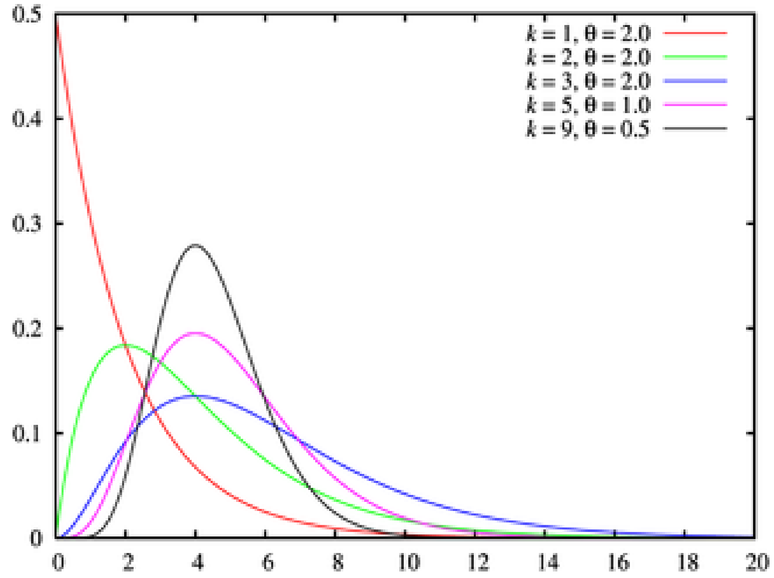
Гамма-распределение

Гамма-функцией (Эйлера) называется функция $\Gamma(p) = \int_0^{+\infty} t^{p-1} e^{-t}\, dt$.

Если $p \in \mathbb{N}$, то $\Gamma(p) = (p-1)!$

Гамма-распределением $\Gamma(\lambda, \alpha)$ с параметрами $\lambda > 0$ и $\alpha > 0$ называется распределение со следующей плотностью

$$p(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0, \\ 0, & x \le 0. \end{cases}$$



Если $X \sim N(0, \sigma^2)$, то $X^2 \sim \Gamma\left(\frac{1}{2\sigma^2}, \frac{1}{2}\right)$.

$\Gamma(\lambda, 1) = \mathrm{Exp}(\lambda)$.

Теорема: для независимых $X_i \sim \Gamma(\lambda, \alpha_i)$, $i \in \overline{1, n}$, имеем $\sum_{i=1}^n X_i \sim \Gamma(\lambda, \sum_{i=1}^n \alpha_i)$.

Теорема: если $X \sim \Gamma(\lambda, \alpha)$, то $kX \sim \Gamma\left(\frac{\lambda}{k}, \alpha\right)$.

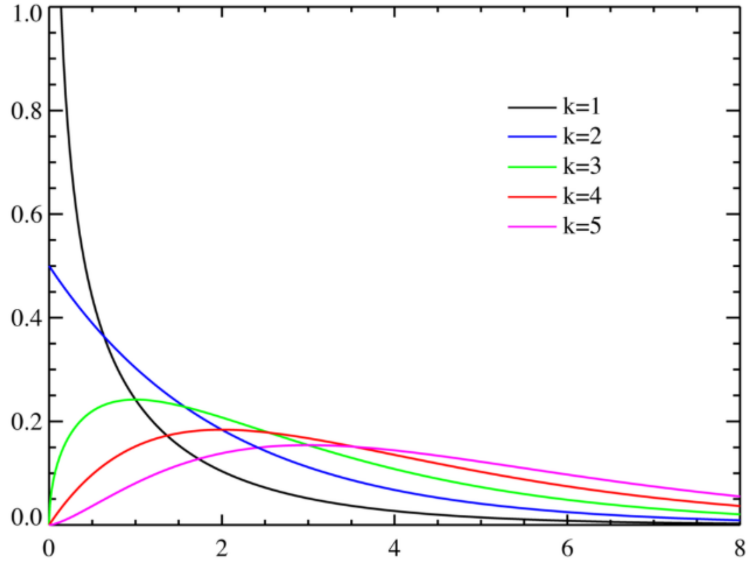
Распределение хи-квадрат

Распределением хи-квадрат $\chi^2(m)$ с m степенями свободы называется распределение суммы квадратов m независимых стандартных нормальных случайных величин.

Теорема: для независимых $X_i \sim N(0,1),\ i \in \overline{1,m},$ имеем $\sum_{i=1}^{m} X_i^2 \sim \Gamma\left(\frac{1}{2},\ \frac{m}{2}\right) = \chi^2(m)$.

Распределение $\chi^2(m)$ имеет следующую плотность:

$$p(x) = \begin{cases} \dfrac{1}{2^{m/2}\,\Gamma(m/2)}\, x^{\frac{m}{2}-1} e^{-\frac{x}{2}}, & x > 0, \\ 0, & x \le 0. \end{cases}$$



Теорема: для $\vec{X}_n \sim N(\mu, \sigma^2)$ имеем

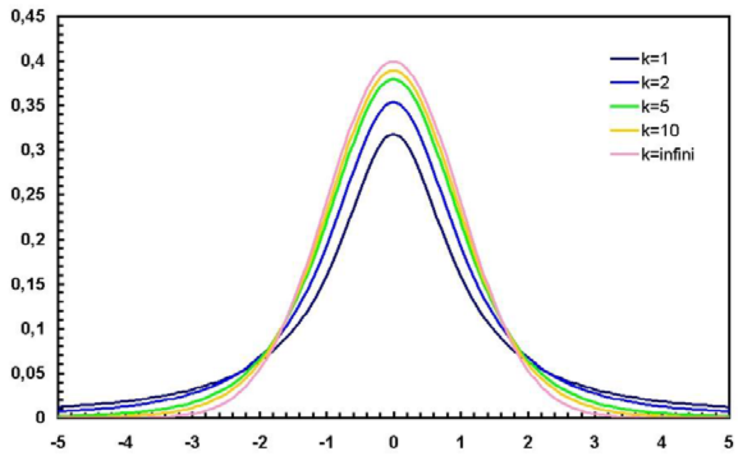
$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n), \qquad \frac{(n-1)\cdot S^2\left(\vec{X}_n\right)}{\sigma^2} = \frac{n\cdot \hat{\sigma}^2\left(\vec{X}_n\right)}{\sigma^2} \sim \chi^2(n-1).$$

Теорема: для независимых $X_i \sim \operatorname{Exp}(\lambda)$ имеем $2\lambda \sum_{i=1}^{n} X_i \sim \chi^2(2n)$.

Распределение Стью́дента (t-распределение)

Распределением Стьюдента (t-***распределением***) с m степенями свободы называется распределение со следующей плотностью

$$p(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{\pi m}\,\Gamma\left(\frac{m}{2}\right)}\left(1+\frac{x^2}{m}\right)^{-\frac{m+1}{2}}.$$



Теорема: для $\vec{X}_n \sim N(\mu, \sigma^2)$ имеем

$$\frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} \sim N(0,1), \qquad \frac{(\bar{X}-\mu)\sqrt{n}}{S(\vec{X}_n)} \sim T(n-1).$$

Свойство квантилей: $t_p(m) = -t_{1-p}(m)$.

Распределение Стьюдента используется для построения оценок и проверки гипотез для математического ожидания.

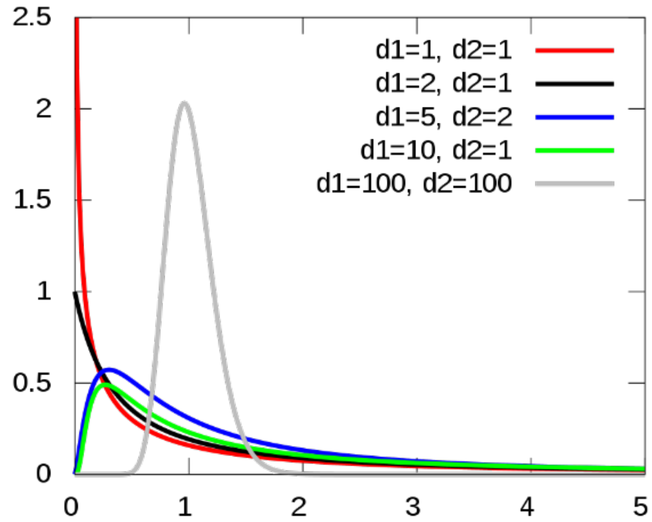
Распределение Фишера (Снедекора)

Бета-функцией называется функция $B(p,q) = \dfrac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)} = \int_0^1 u^{p-1}(1-u)^{q-1}\,du.$

Распределением Фишера (***Снедекора***) $F(n,m)$ с n и m степенями свободы называется распределение величины $\frac{X/n}{Y/m}$, где $X \sim \chi^2(n)$ и $Y \sim \chi^2(m)$.

Распределение Фишера имеет следующую плотность:

$$p(x) = \begin{cases} \dfrac{\left(\frac{n}{m}\right)^{n/2}}{B\left(\frac{n}{2}, \frac{m}{2}\right)} \dfrac{x^{\frac{n}{2}-1}}{\left(1+\frac{nx}{m}\right)^{\frac{n+m}{2}}}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$



Свойство квантилей: $f_p(n,m) = \dfrac{1}{f_{1-p}(m,n)}.$

Распределение Фишера используется для построения оценок для отношения дисперсий и проверке гипотезы о равенстве дисперсий.

Теорема о распределениях, связанных с двумерной генеральной совокупностью (X, Y)

Теорема: если $\vec{X}_n$ и $\vec{Y}_m$ – независимы, $\vec{X}_n \sim N(\mu_1, \sigma_1^2)$, $\vec{Y}_m \sim N(\mu_2, \sigma_2^2)$, то

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1),$$

$$\frac{\sum_{i=1}^{n}(X_i - \mu_1)^2 / n\sigma_1^2}{\sum_{i=1}^{m}(Y_i - \mu_2)^2 / m\sigma_2^2} \sim F(n, m), \qquad \frac{S^2\left(\vec{X}_n\right)/\sigma_1^2}{S^2\left(\vec{Y}_m\right)/\sigma_2^2} \sim F(n-1, m-1).$$

Если дополнительно $\sigma_1 = \sigma_2$, то

$$\sqrt{\frac{mn(m+n-2)}{m+n}} \cdot \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{(n-1)S^2\left(\vec{X}_n\right) + (m-1)S^2\left(\vec{Y}_m\right)}} \sim T(m+n-2).$$

Интервальная оценка и центральная статистика

Пусть СВ X имеет распределение $F(x,\theta)$ с параметром θ.

Интервальной оценкой параметра θ ***с коэффициентом доверия*** (***надёжностью***) γ называется интервал $\left(\underline{\theta}(\vec{X}_n), \overline{\theta}(\vec{X}_n)\right)$ такой, что $P\left(\underline{\theta}(\vec{X}_n) < \theta < \overline{\theta}(\vec{X}_n)\right) = \gamma$.

Точностью оценивания называется половина от длины интервала в интервальной оценке $\frac{\overline{\theta}(\vec{X}_n) - \underline{\theta}(\vec{X}_n)}{2}$.

Статистика $T\left(\vec{X}_n, \theta\right)$ называется ***центральной***, если её функция распределения не зависит от параметра θ.

Для построения интервальных оценок нужна центральная статистика, непрерывная и монотонная по θ:

- выбираем α и β так, чтобы $1-\alpha-\beta=\gamma$, тогда
$$P\left(h_\alpha < T\left(\vec{X}_n, \theta\right) < h_{1-\beta}\right) = \gamma,$$
где h_α и $h_{1-\beta}$ – квантили распределения статистики T;
- обычно $\beta = \alpha$ и границы доверительного интервала находятся из условий
$$T\left(\vec{x}_n, \underline{\theta}\right) = h_\alpha, \qquad T\left(\vec{x}_n, \overline{\theta}\right) = h_{1-\alpha}, \qquad \text{если } T\left(\vec{X}_n, \theta\right) \text{ – возрастает;}$$
$$T\left(\vec{x}_n, \underline{\theta}\right) = h_{1-\alpha}, \qquad T\left(\vec{x}_n, \overline{\theta}\right) = h_\alpha, \qquad \text{если } T\left(\vec{X}_n, \theta\right) \text{ – убывает.}$$

Доверительные оценки для математического ожидания нормального распределения с известной и неизвестной дисперсией

$$T(\vec{X}_n, \mu) = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \sim N(0,1), \qquad P\left(u_\alpha < \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} < u_{1-\alpha}\right) = 1 - 2\alpha,$$

$$u_\alpha = -u_{1-\alpha}, \qquad \mu \in \left(\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha},\ \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}\right).$$

В силу центральной предельной теоремы этой оценкой можно пользоваться и в случае неизвестного распределения с известной дисперсией, но тогда эта оценка будет приближённой.

$$\frac{(\bar{X} - \mu)\sqrt{n}}{S(\vec{X}_n)} \sim T(n-1), \qquad t_p(m) = -t_{1-p}(m),$$

$$\mu \in \left(\bar{X} - \frac{S(\vec{X}_n)}{\sqrt{n}} t_{1-\alpha}(n-1),\ \bar{X} + \frac{S(\vec{X}_n)}{\sqrt{n}} t_{1-\alpha}(n-1)\right).$$

Доверительные оценки для разности математических ожиданий нормальных распределений с известными и неизвестными (но равными) дисперсиями

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim N(0,1), \qquad \mu_1-\mu_2 \approx \bar{X}-\bar{Y} \pm \sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\, u_{1-\alpha}.$$

$$\sigma_1=\sigma_2, \ \sqrt{\frac{mn(m+n-2)}{m+n}} \cdot \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{(n-1)S^2\left(\vec{X}_n\right)+(m-1)S^2\left(\vec{Y}_m\right)}} \sim T(m+n-2),$$

$$\mu_1-\mu_2 \approx \bar{X}-\bar{Y} \pm t_{1-\alpha}(m+n-2) \cdot \sqrt{\frac{(m+n)\left((n-1)S^2\left(\vec{X}_n\right)+(m-1)S^2\left(\vec{Y}_m\right)\right)}{mn(m+n-2)}}\ .$$

Доверительные оценки для дисперсии и среднеквадратичного отклонения нормального распределения с известным и неизвестным математическим ожиданием

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n), \qquad \sigma^2 \in \left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha}(n)},\ \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha}(n)}\right).$$

$$\frac{(n-1)\cdot S^2\left(\vec{X}_n\right)}{\sigma^2} = \frac{n\cdot\hat{\sigma}^2\left(\vec{X}_n\right)}{\sigma^2} \sim \chi^2(n-1), \qquad \sigma^2 \in \left(\frac{(n-1)\,S^2\left(\vec{X}_n\right)}{\chi^2_{1-\alpha}(n-1)},\ \frac{(n-1)\,S^2\left(\vec{X}_n\right)}{\chi^2_{\alpha}(n-1)}\right).$$

Отсюда же интервальные оценки и для σ:

$$\sigma \in \left(\sqrt{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha}(n)}},\ \sqrt{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha}(n)}}\right),$$

$$\sigma \in \left(\sqrt{\frac{(n-1)\,S^2\left(\vec{X}_n\right)}{\chi^2_{1-\alpha}(n-1)}},\ \sqrt{\frac{(n-1)\,S^2\left(\vec{X}_n\right)}{\chi^2_{\alpha}(n-1)}}\right).$$

Доверительные оценки для отношения дисперсий нормальных распределений с известными и неизвестными математическими ожиданиями

$$\frac{\sum_{i=1}^{n}(X_i-\mu_1)^2/n\sigma_1^2}{\sum_{i=1}^{m}(Y_i-\mu_2)^2/m\sigma_2^2} \sim F(n,m), \qquad f_p(n,m)=\frac{1}{f_{1-p}(m,\ n)},$$

$$\frac{\sigma_1^2}{\sigma_2^2} \in \left(\frac{m\ \sum_{i=1}^{n}(X_i-\mu_1)^2}{n\ f_{1-\alpha}(n,m)\ \sum_{i=1}^{m}(Y_i-\mu_2)^2},\ \frac{m\ f_{1-\alpha}(m,n)\ \sum_{i=1}^{n}(X_i-\mu_1)^2}{n\ \sum_{i=1}^{m}(Y_i-\mu_2)^2}\right).$$

$$\frac{S^2(\vec{X}_n)/\sigma_1^2}{S^2(\vec{Y}_m)/\sigma_2^2} \sim F(n-1,m-1),$$

$$\frac{\sigma_1^2}{\sigma_2^2} \in \left(\frac{S^2(\vec{X}_n)}{S^2(\vec{Y}_m)\ f_{1-\alpha}(n-1,m-1)},\ \frac{S^2(\vec{X}_n)\ f_{1-\alpha}(m-1,n-1)}{S^2(\vec{Y}_m)}\right).$$

Доверительная оценка для параметра экспоненциального распределения

$$\vec{X}_n \sim \mathrm{Exp}(\lambda), \qquad T\left(\vec{X}_n, \lambda\right) = 2\lambda n\bar{X} \sim \chi^2(2n), \qquad \lambda \in \left(\frac{\chi^2_{\alpha}(2n)}{2n\bar{X}},\ \frac{\chi^2_{1-\alpha}(2n)}{2n\bar{X}}\right).$$

Приближённая доверительная оценка для параметра *p* биномиального распределения

$$\vec{X}_n \sim B(n,p), \qquad p = MX,$$

поэтому, как и выше для математического ожидания нормального распределения,

$$p \in \left(\bar{X} - \frac{\sigma}{\sqrt{n}}\, u_{1-\alpha},\ \bar{X} + \frac{\sigma}{\sqrt{n}}\, u_{1-\alpha}\right),$$

где $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{k}{n}$ – относительная частота успехов,

$$\sigma = \sqrt{pq} \approx \sqrt{\bar{X}(1-\bar{X})} = \sqrt{\frac{k}{n}\left(1 - \frac{k}{n}\right)},$$

$$p \in \left(\frac{k}{n} - \sqrt{\frac{k(n-k)}{n^3}}\, u_{1-\alpha},\ \frac{k}{n} + \sqrt{\frac{k(n-k)}{n^3}}\, u_{1-\alpha}\right).$$