Интервальное оценивание параметров

Лекция №7

Статистические методы в ядерном эксперименте, ИЯФ, 2024 г.

1. Определения
2. Построение доверительных интервалов
3. Примеры

Определения

Пусть F – функция распределения непрерывной случайной величины. Число q_δ называется квантилью уровня δ, если $F(q_\delta) = \delta$.

Пусть имеется выборка $\{X_k\}_n$ и параметр $1 > \varepsilon > 0$. Интервал со случайными границами $(\theta_-(X_1, ..., X_n), \theta_+(X_1, ..., X_n))$ называется:

- доверительным интервалом для параметра θ уровня доверия $1 - \varepsilon$, если для любого $\theta \in \Theta$ имеет место:

$$P(\theta_-, \theta_+) \geq 1 - \varepsilon$$

- асимптотическим доверительным интервалом для параметра θ уровня доверия $1 - \varepsilon$, если для любого $\theta \in \Theta$ имеет место:

$$\lim_{n \to \infty} \inf P(\theta_-, \theta_+) \geq 1 - \varepsilon$$

Замечание 1: неравенство соответствует случаю дискретного распределения, когда нельзя получить точное равенство для произвольного ε.
Замечание 2: случайными являются границы интервала, поэтому читается эта запись как вероятность того, что интервал содержит истинное значение θ.

Построение доверительных интервалов (I)

Пусть $\{X_k\}_n$ – выборка объёма n из экспоненциального распределения $p(x|\lambda) = \frac{1}{\lambda} e^{-x/\lambda}$ $(\lambda > 0)$. Построим асимптотический доверительный интервал для параметра λ уровня доверия $1 - \varepsilon$.
Исходя из ЦПТ:

$$z = \frac{(\langle x \rangle_n - \mu)\sqrt{n}}{\sigma} = \frac{(\langle x \rangle_n - \lambda)\sqrt{n}}{\lambda} \underset{n \to \infty}{\sim} N(0,1)$$

Возьмём $q_{(1-\varepsilon/2)}$ – квантиль $N(0,1)$ уровня $1 - \varepsilon/2$, учитывая также, что $q_{\varepsilon/2} = -q_{(1-\varepsilon/2)}$, имеем:

$$P(-q_{(1-\varepsilon/2)} < (\langle x \rangle_n - \lambda)\sqrt{n}/\lambda < q_{(1-\varepsilon/2)}) \underset{n \to \infty}{\to} 1 - \varepsilon$$

Разрешив неравенство $-q_{(1-\varepsilon/2)} < (\langle x \rangle_n - \lambda)\sqrt{n}/\lambda < q_{(1-\varepsilon/2)}$ относительно λ, получим асимптотический доверительный интервал для параметра λ:

$$P\left(\frac{\langle x \rangle_n}{1 + \frac{q_{(1-\varepsilon/2)}}{\sqrt{n}}} < \lambda < \frac{\langle x \rangle_n}{1 - \frac{q_{(1-\varepsilon/2)}}{\sqrt{n}}} \right) \underset{n \to \infty}{\to} 1 - \varepsilon$$

Построение доверительных интервалов (II)

Схема построения доверительного интервала

1. Выбрать статистику $Y = Y(X, \theta)$ такую, что функция $Y(X, \theta)$ обратима относительно параметра θ и функция распределения Y известна или сходится к известному распределению F_Y, <u>причём F_Y не зависит от θ</u>.
2. По заданному уровню достоверности $1 - \varepsilon$ определить квантили q_1 и q_2 распределения F_Y такие, что:

$$P(q_1 < Y(X, \theta) < q_2) = 1 - \varepsilon$$

3. Разрешив неравенство $q_1 < Y(X, \theta) < q_2$ относительно θ, определить доверительный интервал.

<u>Замечание:</u> как правило, в качестве q_1 и q_2 берутся квантили уровней $\varepsilon/2$ и $1 - \varepsilon/2$, соответственно. Однако в общем случае при выборе квантилей необходимо руководствоваться соображением минимальности длины доверительного интервала.

Построение доверительных интервалов (III)

Гамма-распределение

$$\Gamma(x|\alpha,\lambda) = \begin{cases} \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}, & \text{если } x \geq 0 \\ 0, & \text{если } x < 0 \end{cases} = \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x} H(x)$$

ПФМ:

$$M'_x(t) = E(e^{xt}) = \frac{\alpha^\lambda}{\Gamma(\lambda)} \int\limits_0^\infty x^{\lambda-1} e^{-x(\alpha-t)} dx = \frac{\alpha^\lambda}{(\alpha-t)^\lambda} = \frac{1}{(1-\frac{t}{\alpha})^\lambda}$$

Пусть X_1, ..., X_n независимы и $X_i \in \Gamma(x|\alpha,\lambda_i)$, $i = 1, ..., n$. Тогда их сумма $S_n = X_1 + ... + X_n$ имеет гамма-распределение с параметрами α и $\lambda_1 + ... + \lambda_n$. ПФМ S_n:

$$M'_{S_n}(t) = \prod_{i=1}^n M'_{X_i}(t) = \frac{1}{(1-\frac{t}{\alpha})^{\sum\limits_{i=1}^n \lambda_i}} \Rightarrow \Gamma(x|\alpha, \sum_{i=1}^n \lambda_i)$$

Замечание 1: если $x \in N(0,1)$, то $x^2 \in \Gamma(x|1/2, 1/2)$.
Замечание 2: если $X_1, ..., X_k$ независимы и $X_i \in N(0,1)$, то с.в.:

$$\chi_k^2 = X_1^2 + ... + X_k^2$$

имеет гамма-распределение $\Gamma(x|1/2, k/2) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-x/2} H(x)$ и называется распределением хи-квадрат с k степенями свободы (для этого распределения будем использовать обозначение χ_k^2) .

Построение доверительных интервалов (IV)

Распределение Стьюдента

Если $X_1, ..., X_n$ независимы и $X_i \in N(\mu, \sigma)$ $(i = 1, .., n)$, рассмотрим с.в.:

$$t = \frac{(\langle x \rangle - \mu)\sqrt{n}}{s} = \frac{\langle x \rangle - \mu}{\sigma/\sqrt{n}} \Big/ \sqrt{\frac{s^2}{\sigma^2}} = \frac{g}{\sqrt{h}}, \quad g = \frac{\langle x \rangle - \mu}{\sigma/\sqrt{n}}, \quad h = \frac{s^2}{\sigma^2},$$

где $\langle x \rangle = \frac{1}{n}\sum_{k=1}^{n} X_k$ и $s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \langle x \rangle)^2$. Видно, что $g \in N(0,1)$, а $h \in \frac{\chi^2_{n-1}}{n-1}$, вычислим ф.п.в. для t. Совместная ф.п.в. для g и h:

$$p(g,h) = \frac{1}{\sqrt{2\pi}} e^{-g^2/2} \times \frac{(n-1)^{\frac{n-1}{2}}}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} h^{\frac{n-1}{2}-1} e^{-h(n-1)/2} H(h)$$

Сделаем замену переменных: $t = g/\sqrt{h}$, $u = h$, и $\frac{\partial(g,h)}{\partial(t,u)} = \sqrt{u}$:

$$T_{n-1}(t) = \int\limits_0^{\infty} p(t\sqrt{u}, u)\sqrt{u}du = \frac{\Gamma(\frac{n}{2})}{\sqrt{\pi(n-1)}\Gamma(\frac{n-1}{2})} \frac{1}{(1+\frac{t^2}{n-1})^{n/2}},$$

$$T_n(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})} \frac{1}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}}$$

Т.о. для с.в. t мы получили распределение Стьюдента с $(n-1)$ степенями свободы, $T_{n-1}(t)$.

Построение доверительных интервалов (V)

Обратим внимание, что $T_1(t) = \frac{1}{\pi(1+t^2)}$, т.е. распределение Коши. Распределение Стьюдента симметрично и при любом конечном n убывает медленнее нормального, имеет более длинные хвосты. Поэтому доверительный интервал, оцениваемый по t-распределению, больше доверительного интервала по нормальному распределению.
При $n \to \infty$ t-распределение стремится к нормальному:

$$\frac{1}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}} \underset{n\to\infty}{\to} \left(\left(1+\frac{t^2}{n}\right)^{-n}\right)^{1/2} \underset{n\to\infty}{\to} e^{-t^2/2}$$

$E(t) = 0$, а $D(t) = \frac{n}{n-2}$ (при $n > 2$).

Построение доверительных интервалов (VI)

Пусть $\{X_k\}_n$ – выборка объёма n из $N(\mu, \sigma^2)$, и
$\langle x \rangle = \frac{1}{n}\sum\limits_{k=1}^{n} X_k$, $s^2 = \frac{1}{n-1}\sum\limits_{k=1}^{n}(X_k - \langle x \rangle)^2$, $s'^2 = \frac{1}{n}\sum\limits_{k=1}^{n}(X_k - \mu)^2$ тогда:

- $\frac{\langle x \rangle - \mu}{\sigma/\sqrt{n}} \in N(0,1)$ (для оценки доверительного интервала для μ при известном σ^2)
- $\frac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i - \mu)^2 = \frac{ns'^2}{\sigma^2} \in \chi^2_n$ (для оценки доверительного интервала для σ^2 при известном μ)
- $\frac{\langle x \rangle - \mu}{\sqrt{s^2/n}} \in T_{n-1}$ (для оценки доверительного интервала для μ при неизвестном σ^2, вместо σ^2 используем выборочную дисперсию s^2)
- $\sum\limits_{i=1}^{n}\left(\frac{X_i - \langle x \rangle}{\sigma}\right)^2 = \frac{(n-1)s^2}{\sigma^2} \in \chi^2_{n-1}$ (для оценки доверительного интервала для σ^2 при неизвестном μ, вместо μ используем выборочное среднее $\langle x \rangle$)