Пусть $\gamma \in (0,1)$, $T_1(X)$ и $T_2(X)$ статистики, причём $T_1(X) < T_2(X)$

Интервал (T_1, T_2) называется γ-доверительным, или доверительным уровня значимости γ для параметра θ, если $\mathbb{P}_\theta(\{T_1(X) < \theta < T_2(X)\}) \geq \gamma \ \forall\gamma$

Аналогично вводится понятие доверительного интервала для функций от параметра

$G(X,\theta)$ называется центральной статистикой, если

1) G монотонна на параметре

2) Распределение $G(X,\theta)$ не зависит от параметра θ

Метод построения доверительного интервала

Пусть g_1 и g_2 такие, что $g_1 < g_2$ и $g_2 - g_1 = \gamma$

И пусть F_G - функция распределения центральной статистики G такая, что $F_G(g_2) - F_G(g_1) = \gamma$

Тогда $\mathbb{P}_\theta(\{g_1 < G(X,\theta) < g_2\}) = \gamma \ \forall\theta$

Находим T_1 и T_2, решая уравнение $G(X,\theta) = g_1, g_2$ (решаем относительно θ)

$\mathbb{P}_\theta(\{T_1 < \theta < T_2\}) = \mathbb{P}_\theta(\{G(X,T_1) < G(X,\theta) < G(X,T_2)\}) = \mathbb{P}_\theta(\{g_1 < G(X,\theta) < g_2\}) = \gamma$ в силу монотонности

Примеры построения доверительных интервалов

1) $N(\theta,\sigma^2)$, строим интервал для параметра θ

Центральная статистика - $G(X,\theta) = \sqrt{n}\frac{\overline{X}-\theta}{\sigma}$ (статистика монотонна по параметру и имеет распределение $N(0,1)$ - не зависит от параметра, значит статистика центральная)

2) $N(\theta_1,\theta_2^2)$, строим для θ_1)

$G(X,\theta) = \sqrt{n-1}\frac{\overline{X}-\theta}{\sqrt{S^2}}$ (статистика центральная по аналогичным рассуждениям)

3) $N(a,\theta^2)$, надо построить не для параметра, а для функции от него θ^2

$\frac{X_i-a}{\theta} \sim N(0,1)$

$G(X,\theta) = \sum_{i=1}^n \frac{(X_i-a)^2}{\theta^2} \sim \chi^2_n$ (условия центральной статистики выполнены)

Нужно выбрать g_1 и g_2 так, чтобы $F_G(g_2) - F_G(g_1) \geq \gamma$

$F(g) = \int\limits_0^g K_n(x)dx$, где $K_n(x) = \frac{x^{n/2-1}}{\Gamma(n/2)2^{n/2}},\ x > 0$

В качестве g_1 и g_2 будем брать квантили

Пусть $F(g_1) = \alpha_1$, тогда $g_1 = \chi^2_{n,\alpha_1}$ и $F(g_2) = 1-\alpha_2$, тогда $g_2 = \chi^2_{n,1-\alpha_2}$, $1-\alpha_2-\alpha_1 \geq \gamma$

Решаем уравнение $G(X,\theta) = \frac{1}{\theta^2}\sum\limits_{i=1}^n (X_i-a)^2 = g_1 = \chi^2_{n,\alpha_1}$ относительно θ^2

$\theta^2 = \frac{1}{\chi^2_{n,\alpha_1}}\sum\limits_{i=1}^n (X_i-a)^2$ для g_1, $\theta^2 = \frac{1}{\chi^2_{n,1-\alpha_2}}\sum\limits_{i=1}^n (X_i-a)^2$ для g_2

Доверительный интервал имеет вид $\left(\frac{1}{\chi^2_{n,\alpha_1}}\sum\limits_{i=1}^n (X_i-a)^2;\ \frac{1}{\chi^2_{n,1-\alpha_2}}\sum\limits_{i=1}^n (X_i-a)^2\right)$, где $\alpha_1 + \alpha_2 = 1-\gamma$

Таких интервалов много (α_1 и α_2 можно выбрать многими способами), надо выбрать кратчайший интервал

Для построения кратчайшего доверительного интервала нужно минимизировать $\frac{1}{\chi^2_{n,\alpha_1}} - \frac{1}{\chi^2_{n,1-\alpha_2}}$

$\chi^2_{n,\alpha_1} = F_G^{-1}(\alpha_1)$

$(F_G^{-1}(\alpha_1))' = \frac{1}{F'_G(F_G^{-1}(\alpha_1))} = \frac{1}{K_n(g_1)}$

$(\frac{1}{\chi^2_{n,\alpha_1}} - \frac{1}{\chi^2_{n,1-\alpha_2}})' = \frac{1}{g_1^2}\cdot\frac{1}{K_n(g_1)} - \frac{1}{g_2^2}\cdot\frac{1}{K_n(g_2)} = 0$

Получили систему уравнений:

$$\begin{cases} g_1^2 K_n(g_1) = g_2^2 K_n(g_2) \\ \alpha_1 + \alpha_2 = 1-\gamma \end{cases}$$

$$\begin{cases} g_1^2 g_1^{n/2-1} e^{-g_1/2} = g_2^2 g_2^{n/2-1} e^{-g_2/2} \\ 1-\alpha_2 = \alpha_1 + \gamma \end{cases}$$

Из неё вытекает, что α_1 и $1-\alpha_2$ должны быть симметричны относительно $\frac{1}{2}$, $F(g_1) = \frac{1-\gamma}{2} = \alpha_1$, $F(g_2) = \frac{1+\gamma}{2} = 1-\alpha_2$

Мы получили симметричный/центральный интервал, $g_1 = \chi_{n,(1-\gamma)/2}$, $g_2 = \chi_{n,(1+\gamma)/2}$, вероятность попасть левее него равна вероятности попасть правее

4) Оптимальный интервал для параметра в модели $N(\theta, \sigma^2)$

$G(X,\theta) = \sqrt{n}\frac{\overline{X}-\theta}{\sigma}$

$\Phi(g)$ - интеграл вероятности, $\Phi(g_2) - \Phi(g_1) = \gamma$

$G(X,T_1) = g_1$, $G(X,T_2) = g_2$ или наоборот, чтобы было $T_1 < T_2$

Решаем уравнение $G(X,T) = g$

$\frac{\sqrt{n}(\overline{X}-T)}{\sigma} = g$

$\overline{X} - T = \frac{g\sigma}{\sqrt{n}}$

$T = \overline{X} - \frac{g\sigma}{\sqrt{n}}$

$T_1 = \overline{X} - \frac{g_2\sigma}{\sqrt{n}}$, $T_2 = \overline{X} - \frac{g_1\sigma}{\sqrt{n}}$, так как $-\frac{g_1\sigma}{\sqrt{n}} > -\frac{g_2\sigma}{\sqrt{n}}$

$l = T_2 - T_1 = (g_2 - g_1)\frac{\sigma}{\sqrt{n}}$, мы должны минимизировать l или же $g_2 - g_1$ при соблюдении условия $\Phi(g_2) - \Phi(g_1) = \gamma$

Решаем задачу методом множителя Лагранжа, вводим новую функцию с параметром

$\phi(g_1, g_2, \lambda) = (g_2 - g_1) + \lambda(\Phi(g_2) - \Phi(g_1) - \gamma)$ - ищем минимум получившейся функции

$grad\ \phi = (-1 - \lambda\Phi'(g_1),\ 1 + \lambda\Phi'(g_2),\ \Phi(g_2) - \Phi(g_1) - \gamma) = (0,\ 0,\ 0)$

$1 + \lambda\Phi'(g_1) = 1 + \lambda\Phi'(g_2) = 0 \Rightarrow \Phi'(g_1) = \Phi'(g_2)$

$\frac{1}{\sqrt{2\pi}}e^{-\frac{g_1^2}{2}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{g_2^2}{2}}$

$g_1^2 = g_2^2$

$g_1 = -g_2$

$\gamma = \Phi(g_2) - \Phi(-g_2) = \Phi(g_2) - (1 - \Phi(g_2))$

$\Phi(g_2) = \frac{1+\gamma}{2}$, $g_2 = c_\gamma$, $g_1 = -c_\gamma$ - центральный оптимальный доверительный интервал

5) Доверительный интервал для параметра равномерного распределения $R((0,\theta))$

На выбор 2 центральные статистики

Первая это $\frac{\overline{X}-\frac{\theta}{2}}{\theta} = \frac{\overline{X}}{\theta} - \frac{1}{2}$, при объёме выборки (то есть n) > 12 распределение $\sqrt{n}(\frac{\overline{X}}{\theta} - \frac{1}{2})$ будет приближенно-нормальным, и оценка γ тоже будет приближенной - поэтому мы выберем следующую центральную статистику

Вторая это $\frac{X_{(n)}}{\theta}$, в этом случае квантилями будут просто корни n-ой степени

Убедимся, что $(\frac{X_{(n)}}{\theta})^n \sim R((0,1))$

$(\frac{X_{(n)}}{\theta})^n < x \Leftrightarrow \frac{X_{(n)}}{\theta} < \sqrt[n]{x}$

$F_{(n)}(\sqrt[n]{x}) = (\sqrt[n]{x})^n = x$

Упростим себе задачу ещё больше - вместо $\frac{X_{(n)}}{\theta}$ в качестве центральной статистики возьмём $(\frac{X_{(n)}}{\theta})^n$, тогда $F_G(g) = g$

Нужно найти g_1 и g_2 ($g_1 < g_2$) чтобы $F(g_2) - F(g_1) = \gamma$ или же $g_2 - g_1 = \gamma$

$g_2 = g_1 + \gamma$ должно быть меньше одного, тогда достаточно взять любое $g_1 < 1 - \gamma$

Теперь ищем статистики T_1 и T_2

$(\frac{X_{(n)}}{T})^n = g$

$\frac{X_{(n)}}{T} = \sqrt[n]{g}$

$T = \frac{X_{(n)}}{\sqrt[n]{g}}$

$T_1 = \frac{X_{(n)}}{\sqrt[n]{g_2}}$, $T_2 = \frac{X_{(n)}}{\sqrt[n]{g_1}}$

Продолжение в следующей лекции