Лекция 4

Робастные оценки

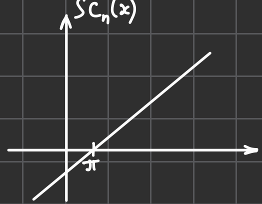
Опр: пусть оценка $\hat{\theta}_n$ построена по выборке $X_1,\dots,X_n$. Затем добавлено ещё одно наблюдение x и оценка $\hat{\theta}_{n+1}$ построена по $X_1,\dots,X_n$ и x. Тогда кривой чувствительности, измеряющей влияние наблюдения x на оценку $\hat{\theta}_n$ называется функция:

$$SC_n(x) = \frac{\hat{\theta}_{n+1} - \hat{\theta}_n}{\frac{1}{n+1}} = (n+1)\left(\hat{\theta}_{n+1} - \hat{\theta}_n\right)$$

Опр: оценка $\hat{\theta}$ называется B-робастной, если $SC_n(x)$ ограничена

Пример: пусть $\hat{\theta} = \bar{x}$

$$SC_n(x) = (n+1)\left(\frac{1}{n+1}\left(\sum_{i=1}^{n} x_i + x\right) - \frac{1}{n}\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} x_i + x - \left(\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} x_i\right) = x - \bar{x}$$



Выборочная медиана

эл-т вариационного ряда

$$\hat{\mu} = \begin{cases} X_{(k+1)}, & \text{если } n = 2k+1 \\ \dfrac{X_{(k)} + X_{(k+1)}}{2}, & \text{если } n = 2k \end{cases}$$

Опр: пороговой точкой (BP) ε_n^* оценки $\hat{\theta}$, построенной по выборке $x_1,\dots,x_n$ называется

$$\varepsilon_n^* = \frac{1}{n}\max\left\{m : \max_{i_1,\dots,i_m}\ \sup_{y_1,\dots,y_m} \left|\hat{\theta}(z_1,\dots,z_n)\right| < \infty\right\},$$ где выборка $z_1,\dots,z_n$

получена заменой значений $x_{i_1},\dots,x_{i_m}$ на произвольные значения $y_1,\dots,y_m$

Доверительные интервалы

$X_1,\dots,X_n \sim F(x,\theta)$, $\theta \subset \Theta \in \mathbb{R}^1$

Опр: пусть построены статистики $T_1(X_1,\dots,X_n)$ и $T_2(X_1,\dots,X_n)$, т.ч. $T_1(x) < T_2(x)$ и $P(T_1(x) < \theta < T_2(x)) = 1-\alpha$

$0 < \alpha < 1$

Тогда интервал $(T_1(x), T_2(x))$ называется доверительным

Опр: случайная функция $G(X_1,\dots,X_n,\theta) = G(X,\theta)$ называется центральной (опорной) статистикой, если:

1) $G(x,\theta)$ непрерывна и монотонна по θ
2) $F_G(x)$ не зависит от θ

$P(G(x,\theta) < Z_{1-\alpha}) = 1-\alpha$

$P(Z_{\alpha} < G(x,\theta)) = 1-\alpha$

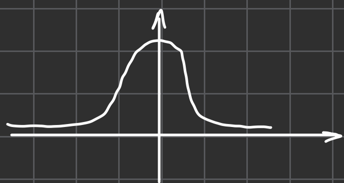
$P(Z_{\frac{\alpha}{2}} < G(x,\theta) < Z_{1-\frac{\alpha}{2}}) = 1-\alpha$

Распределение хи-квадрат (χ^2)

Опр: пусть СВ $\xi_1,\dots,\xi_m \sim N(0,1)$ и независимы, тогда СВ $\eta = \sum_{i=1}^{m} \xi_i^2 \sim \chi^2(m)$

Распределение Стьюдента

Опр: пусть $\xi_0, \xi_1, \dots, \xi_m \sim N(0,1)$ и независимы, тогда СВ $\xi = \dfrac{\xi_0}{\sqrt{\frac{1}{m}\sum_{i=1}^{m} \xi_i^2}} \sim t(m)$



$m = 1$ – распределение Коши

Распределение Фишера

Опр: пусть СВ $\xi_1 \sim \chi^2(m)$, $\xi_2 \sim \chi^2(n)$ и ξ_1 и ξ_2 – независимые. Тогда СВ $F = \dfrac{\frac{1}{m}\xi_1}{\frac{1}{n}\xi_2} \sim F(m,n)$

Теорема Фишера: пусть $x_1,\dots,x_n$ порождена СВ $x \sim N(m, \sigma^2)$, тогда:

1) $\dfrac{nS^2}{\sigma^2} = \sum_{i=1}^{n}\left(\dfrac{x_i - \bar{x}}{\sigma}\right)^2 \sim \chi^2(n-1)$

2) $\bar{x}$ и S^2 – независимые СВ

Примеры:

1) $X_1,\dots,X_n \sim N(\theta, \sigma^2)$, σ^2 – известно, ДИ для θ

ДИ – доверительный интервал

$\hat{\theta} = \bar{x} \sim N(\theta, \frac{\sigma^2}{n})$

$\dfrac{\sqrt{n}(\bar{x} - \theta)}{\sigma} \sim N(0;1)$

$P\left(Z_{\frac{\alpha}{2}} < \dfrac{\sqrt{n}(\bar{x}-\theta)}{\sigma} < Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$

$P\left(\dfrac{-Z_{1-\alpha/2}\,\sigma}{\sqrt{n}} - \bar{x} < -\theta < \dfrac{Z_{1-\alpha/2}\,\sigma}{\sqrt{n}} - \bar{x}\right) = 1-\alpha$

$P\left(\bar{x} - \dfrac{Z_{1-\alpha/2}\cdot\sigma}{\sqrt{n}} < \theta < \bar{x} + \dfrac{Z_{1-\alpha/2}\cdot\sigma}{\sqrt{n}}\right) = 1-\alpha$

$\left(\bar{x} - \dfrac{Z_{1-\frac{\alpha}{2}}\cdot\sigma}{\sqrt{n}}; \bar{x} + \dfrac{Z_{1-\frac{\alpha}{2}}\cdot\sigma}{\sqrt{n}}\right)$

2) $X_1,\dots,X_n \sim N(m, \theta_2^2)$, m – известно, ДИ для θ_2^2

$\sum_{i=1}^{m}\left(\dfrac{X_i - m}{\theta_2}\right)^2 \sim \chi^2(n)$

$$P\left(\chi^2_{n,\frac{\alpha}{2}} < \frac{\sum_{i=1}^{n}(x_i - m)^2}{\theta_2^2} < \chi^2_{n,1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(\frac{\sum_{i=1}^{n}(x_i - m)^2}{\chi^2_{n,1-\frac{\alpha}{2}}} < \theta_2^2 < \frac{\sum_{i=1}^{n}(x_i - m)^2}{\chi^2_{n,\frac{\alpha}{2}}}\right) = 1-\alpha$$

③ $X_1, \dots, X_n \sim N(\theta_1, \theta_2^2)$, θ_1 и θ_2 - неизв., ДИ для θ_2^2

$$\boxed{\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\theta_2}\right)^2} \sim \chi^2(n-1)$$

$$= G(X, \theta^2)$$

$$P\left(\chi^2_{n-1,\frac{\alpha}{2}} < \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\theta_2^2} < \chi^2_{n-1,1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{\chi^2_{n-1,1-\frac{\alpha}{2}}} < \theta_2^2 < \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{\chi^2_{n-1,\frac{\alpha}{2}}}\right) = 1-\alpha$$

④ $X_1, \dots, X_n \sim N(\theta_1, \theta_2^2)$, θ_1 и θ_2 - неизвестны

ДИ для θ_1 $\qquad t_{n,\alpha} = -t_{n,1-\alpha}$

$$\frac{\sqrt{n}\left(\frac{\bar{x}-\theta_1}{\sigma}\right)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma}\right)^2}} = \frac{\sqrt{n}(\bar{x}-\theta_1)}{\tilde{S}} \sim t(n-1)$$

$$P\left(t_{n-1,\frac{\alpha}{2}} < \frac{\sqrt{n}(\bar{x}-\theta_1)}{S} < t_{n-1,1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(\bar{x} - \frac{t_{n-1,1-\frac{\alpha}{2}}\,\tilde{S}}{\sqrt{n}} < \theta_1 < \bar{x} + \frac{t_{n-1,1-\frac{\alpha}{2}}\cdot\tilde{S}}{\sqrt{n}}\right) = 1-\alpha$$