Семинар №9. Доверительные интервалы.

Определение. Пара статистик $(T_1(X), T_2(X))$ (где X — наблюдение) называется доверительным интервалом уровня доверия γ для параметра θ, если для $\forall\theta \in \Theta \subset \mathbb{R}$ выполнено

$$P_\theta(T_1(X) < \theta < T_2(X)) \geq \gamma.$$

Доверительный интервал называется точным, если $\forall\theta \in \Theta$

$$P_\theta(T_1(X) < \theta < T_2(X)) = \gamma.$$

<u>Метод центральной статистики.</u>

Пусть случайная величина $G(X,\theta)$ такова, что её распределение известно и не зависит от θ. Пусть также $G(X,\theta)$ строго монотонна и непрерывна по θ.

Пусть $p_1 < p_2$, $p_i \in [0,1]$, таковы, что $p_2 - p_1 = \gamma$. Обозначим через G_{p_i} p_i-тую квантиль распределения $G(X,\theta)$ (т.е. $P(G(X,\theta) \leq G_{p_i}) = p_i$, если распределение $G(X,\theta)$ непрерывно). Пусть $T_i(X)$ — решения уравнений $G(X, T_i(X)) = G_{p_i}$. Тогда

$$P_\theta(T_1(X) < \theta < T_2(X)) = P_\theta(G_{p_1} < G(X,\theta) < G_{p_2}) = p_2 - p_1 = \gamma,$$

т.е. $(T_1(X), T_2(X))$ – доверительный интервал уровня доверия γ.

Замечание. От доверительного интервала требуется как можно меньшая длина, чтобы "локализовать" неизвестное значение параметра. С этой целью p_1 и p_2 выбираются, как правило, симметричными относительно $0,5$, ведь у распределения могут быть тяжёлые хвосты.

Это всё, безусловно, прекрасно, но без примера не особо понятно. Посему
Пример. Пусть $X_1, \ldots, X_n$ — выборка из распределения с плотностью $p_\theta(x) = e^{-(x-\theta)} I\{x \geq \theta\}$ (экспоненциальное распределение со сдвигом). Построить доверительный интервал уровня доверия γ.

<u>Решение.</u> Заметим, что $X_i - \theta \sim Exp(1)$ — не зависит от θ. Тогда возьмём $G(X,\theta) = \sum\limits_{i=1}^{n} X_i - n\theta \sim \Gamma(1,n)$, так как сумма независимых, одинаково экспоненциально распределенных случайных величин имеет гамма-распределение. Выберем $u_{\frac{1-\gamma}{2}}$ и $u_{\frac{1+\gamma}{2}}$ — квантили уровня $\frac{1-\gamma}{2}$ и $\frac{1+\gamma}{2}$ из распределения $\Gamma(1,n)$ соответственно. Тогда

$$\gamma = \frac{1+\gamma}{2} - \frac{1-\gamma}{2} = P_\theta\left(u_{\frac{1-\gamma}{2}} < \sum_{i=1}^{n} X_i - n\theta < u_{\frac{1+\gamma}{2}}\right) = P_\theta\left(\overline{X} - \frac{u_{\frac{1+\gamma}{2}}}{n} < \theta < \overline{X} - \frac{u_{\frac{1-\gamma}{2}}}{n}\right),$$

т.е. доверительный интервал для θ в нашей задаче такой: $\left(\overline{X} - \frac{u_{\frac{1+\gamma}{2}}}{n}; \overline{X} - \frac{u_{\frac{1-\gamma}{2}}}{n}\right)$. ■

<u>Асимптотические доверительные интервалы.</u>

Определение. $X = (X_1, \ldots, X_n)$ — выборка из $P \in \{P_\theta, \theta \in \Theta\}$. Последовательность случайных интервалов $\left(T_1^{(n)}(X_1, \ldots, X_n), T_2^{(n)}(X_1, \ldots, X_n)\right)$ называется асимптотическим доверительным интервалом уровня доверия γ для θ, если для $\forall\theta \in \Theta$

$$\liminf_{n\to\infty} P_\theta\left(T_1^{(n)}(X_1, \ldots, X_n) < \theta < T_2^{(n)}(X_1, \ldots, X_n)\right) \geq \gamma.$$

Он называется точным, если $\forall\theta \in \Theta$

$$\liminf_{n\to\infty} P_\theta\left(T_1^{(n)}(X_1,\ldots,X_n) < \theta < T_2^{(n)}(X_1,\ldots,X_n)\right) = \gamma.$$

Метод построения. Пусть $\widehat{\theta}_n(X)$ – асимптотически нормальная оценка θ, т.е., по определению,

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{d_\theta} N(0, \sigma^2(\theta)).$$

Тогда

$$\frac{\sqrt{n}\left(\widehat{\theta}_n - \theta\right)}{\sigma(\theta)} \xrightarrow{d_\theta} N(0,1).$$

У нас есть большое желание построить какой-нибудь доверительный интервал для θ при использовании квантилей нормального распределения, но нам мешает $\sigma(\theta)$ в знаменателе (имеется в виду, что не всегда из доверительного интервала для статистики $\frac{\sqrt{n}(\widehat{\theta}_n-\theta)}{\sigma(\theta)}$ мы можем получить доверительный интервал для параметра θ). Оказывается, если $\sigma(\theta)$ непрерывна по θ, то вместо θ в $\sigma(\theta)$ можно подставить $\widehat{\theta}_n$ (действительно, ведь тогда $\sigma(\widehat{\theta}_n) \xrightarrow{d_\theta} \sigma(\theta)$ по теореме о наследовании сходимости), причём сходимость к нормальному закону сохранится (это следует из леммы Слуцкого).
Если $u_{1-\frac{\alpha}{2}}$ и $u_{\frac{\alpha}{2}}$ – квантили $N(0,1)$, $\alpha = 1-\gamma$ (а $u_{1-\frac{\alpha}{2}} = -u_{\frac{\alpha}{2}}$ в силу симметрии распределения $N(0,1)$), то

$$P_\theta\left(-u_{1-\frac{\alpha}{2}} < \frac{\sqrt{n}\left(\widehat{\theta}_n - \theta\right)}{\sigma(\widehat{\theta}_n)} < u_{1-\frac{\alpha}{2}}\right) = P_\theta\left(\widehat{\theta}_n - \frac{u_{1-\frac{\alpha}{2}}\sigma(\widehat{\theta}_n)}{\sqrt{n}} < \theta < \widehat{\theta}_n + \frac{u_{1-\frac{\alpha}{2}}\sigma(\widehat{\theta}_n)}{\sqrt{n}}\right) \longrightarrow$$

$$\xrightarrow[n\to\infty]{} 1-\alpha = \gamma,$$

т.е. асимптотический доверительный интервал для θ при наличии асимптотически нормальной оценки $\widehat{\theta}_n$ выглядит так: $\left(\widehat{\theta}_n - \frac{u_{1-\frac{\alpha}{2}}\sigma(\widehat{\theta}_n)}{\sqrt{n}}; \widehat{\theta}_n + \frac{u_{1-\frac{\alpha}{2}}\sigma(\widehat{\theta}_n)}{\sqrt{n}}\right)$.

Пример. Пусть $X_1,\ldots,X_n$ – выборка из $Bern(\theta)$. Построить асимптотический доверительный интервал уровня доверия γ для параметра θ.
Решение. Из ЦПТ, $\sqrt{n}(\overline{X}-\theta) \xrightarrow{d_\theta} N(0,\theta(1-\theta))$, то есть асимптотически нормальную оценку для построения асимптотического доверительного интервала мы получили. Осталось просто подставить нашу оценку и значение асимптотической дисперсии в ту формулу, которую мы вывели выше:

$$\left(\overline{X} - \frac{u_{\frac{1+\gamma}{2}}}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})},\ \overline{X} + \frac{u_{\frac{1+\gamma}{2}}}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})}\right)$$

– искомый асимптотический доверительный интервал для параметра θ уровня доверия γ, где $u_{\frac{1+\gamma}{2}}$ – по-прежнему квантиль уровня $\frac{1+\gamma}{2}$ из распределения $N(0,1)$. ■