Занятие 3. Доверительные интервалы и границы.

Домашнее задание:

Глава 19, задачи: 157, 158, 173, 174, 183, 184, 187, 188, 192.

Задача 3.1.

Задана выборка $(\xi_1,..., \xi_n)$ из нормального распределения $N(m,\sigma^2)$, в котором m – неизвестный параметр и σ^2 – известное числовое значение. Построить доверительный интервал для m с уровнем доверия $P_д$.

Решение:

Рассмотрим выборочное среднее $\hat{m}_1(\xi_1,..., \xi_n)$:

$$\hat{m}_1(\xi_1,..., \xi_n) = \frac{1}{n}\sum_{i=1}^{n} \xi_i .$$

Поскольку $(\xi_1,..., \xi_n)$ является выборкой, то все величины ξ_i совместно независимы и имеют нормальное распределение $N(m,\sigma^2)$, отсюда следует, что сумма $\sum_{i=1}^{n} \xi_i$ также имеет нормальное распределение, тогда и выборочное среднее $\hat{m}_1(\xi_1,..., \xi_n)$ имеет нормальное распределение:

$$\hat{m}_1(\xi_1,..., \xi_n) \sim N\left(m, \frac{\sigma^2}{n}\right).$$

Функция распределения выборочного среднего $F_{\hat{m}_1}(x)$ может быть выражена через функцию Лапласа $\Phi(x)$ (функцию распределения стандартного нормального распределения $N(0,1)$):

$$F_{\hat{m}_1}(x) = \Phi\left(\frac{x-m}{\frac{\sigma}{\sqrt{n}}}\right).$$

Образуем новую величину $\varphi(\xi_1,..., \xi_n; m)$:

$$\varphi(\xi_1,..., \xi_n; m) = F_{\hat{m}_1}(\hat{m}_1(\xi_1,..., \xi_n)) .$$

Поскольку выборочное среднее $\hat{m}_1(\xi_1,..., \xi_n)$ подставляется в свою функцию распределения $F_{\hat{m}_1}(x)$, то, как известно из курса теории вероятностей, полученная подстановкой случайная величина $F_{\hat{m}_1}(\hat{m}_1(\xi_1,..., \xi_n))$ будет иметь равномерное распределение $R[0,1]$, тогда такое же распределение будет иметь и величина $\varphi(\xi_1,..., \xi_n; m)$:

$$\varphi(\xi_1,..., \xi_n; m) = F_{\hat{m}_1}(\hat{m}_1(\xi_1,..., \xi_n)) = \Phi\left(\frac{\hat{m}_1(\xi_1,..., \xi_n) - m}{\frac{\sigma}{\sqrt{n}}}\right) \sim R[0,1] .$$

Легко видеть, что величина $\varphi(\xi_1,..., \xi_n; m)$ является центральной статистикой для m, действительно, распределение $\varphi(\xi_1,..., \xi_n; m)$ полностью известно (и не зависит от неизвестного m), величина $\varphi(\xi_1,..., \xi_n; m)$ как функция m является непрерывной (поскольку непрерывной является функция $\Phi(x)$) и при всех реализациях выборки величина $\varphi(\xi_1,..., \xi_n; m)$ является убывающей по m функцией (поскольку функция $\Phi(x)$ является возрастающей).

Таким образом, величина $\varphi(\xi_1,..., \xi_n; m)$ может быть использована при построении доверительного интервала для m. Выберем число $\alpha = \frac{1 - P_д}{2}$, тогда:

$$P\{\alpha < \varphi(\xi_1,..., \xi_n; m) < 1 - \alpha\} = 1 - \alpha - \alpha = 1 - 2\alpha = P_д,$$

поскольку величина $\varphi(\xi_1,..., \xi_n; m)$ имеет равномерное распределение $R[0,1]$. Преобразовывая полученное равенство получим следующую цепочку равенств:

$$P\{\alpha < \varphi(\xi_1,..., \xi_n; m) < 1 - \alpha\} = P_д,$$

$$P\left\{\alpha < \Phi\left(\frac{\hat{m}_1(\xi_1,..., \xi_n) - m}{\frac{\sigma}{\sqrt{n}}}\right) < 1 - \alpha\right\} = P_д,$$

$$P\left\{\Phi^{-1}(\alpha) < \frac{\hat{m}_1(\xi_1,..., \xi_n) - m}{\frac{\sigma}{\sqrt{n}}} < \Phi^{-1}(1 - \alpha)\right\} = P_д,$$

$$P\left\{\hat{m}_1(\xi_1,..., \xi_n) - \Phi^{-1}(1 - \alpha)\frac{\sigma}{\sqrt{n}} < m < \hat{m}_1(\xi_1,..., \xi_n) - \Phi^{-1}(\alpha)\frac{\sigma}{\sqrt{n}}\right\} = P_д,$$

$$P\left\{\frac{1}{n}\sum_{i=1}^{n}\xi_i - \Phi^{-1}(1 - \alpha)\frac{\sigma}{\sqrt{n}} < m < \frac{1}{n}\sum_{i=1}^{n}\xi_i - \Phi^{-1}(\alpha)\frac{\sigma}{\sqrt{n}}\right\} = P_д.$$

Поскольку последнее равенство справедливо при любом значении неизвестного параметра m, тогда и точная нижняя грань:

$$\inf_m P\left\{\frac{1}{n}\sum_{i=1}^{n}\xi_i - \Phi^{-1}(1 - \alpha)\frac{\sigma}{\sqrt{n}} < m < \frac{1}{n}\sum_{i=1}^{n}\xi_i - \Phi^{-1}(\alpha)\frac{\sigma}{\sqrt{n}}\right\} = P_д.$$

Отсюда следует, что интервал:

$$\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i - \Phi^{-1}(1 - \alpha)\frac{\sigma}{\sqrt{n}}; \frac{1}{n}\sum_{i=1}^{n}\xi_i - \Phi^{-1}(\alpha)\frac{\sigma}{\sqrt{n}}\right)$$

является доверительным интервалом для m с уровнем доверия $P_д$.

Заметим, что полученный доверительный интервал совпадает с известным наикратчайшим доверительным интервалом для m, действительно, величина

$$\Phi^{-1}(1 - \alpha) = \Phi^{-1}\left(1 - \frac{1 - P_д}{2}\right) = \Phi^{-1}\left(\frac{1 + P_д}{2}\right)$$

является квантилью стандартного нормального распределения уровня $\frac{1 + P_д}{2}$, с другой стороны, в силу известного свойства $\Phi(x) = 1 - \Phi(-x)$:

$$\Phi(\Phi^{-1}(\alpha)) = 1 - \Phi(-\Phi^{-1}(\alpha)),$$

$$\alpha = 1 - \Phi(-\Phi^{-1}(\alpha)),$$

$$\Phi(-\Phi^{-1}(\alpha)) = 1 - \alpha,$$

$$-\Phi^{-1}(\alpha) = \Phi^{-1}(1 - \alpha),$$

то есть $-\Phi^{-1}(\alpha)$ является квантилью уровня $1 - \alpha = \frac{1 + P_д}{2}$.

Ответ:

$$\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i - \Phi^{-1}(1-\alpha)\frac{\sigma}{\sqrt{n}};\frac{1}{n}\sum_{i=1}^{n}\xi_i - \Phi^{-1}(\alpha)\frac{\sigma}{\sqrt{n}}\right).$$

Задача 3.2.

Задана выборка $(\xi_1,..., \xi_n)$ из равномерного распределения $R[0,a]$, где $a>0$ неизвестный параметр. Построить нижнюю и верхнюю доверительные границы для a с уровнем доверия $P_д$.

Решение:

Рассмотрим порядковую статистику $\xi_{(n)} = \max\limits_{1\le i\le n} \xi_{(i)}$, функция распределения $F_{(n)}(x)$ которой, как известно, имеет вид:

$$F_{(n)}(x) = F_{\xi}^{n}(x),$$

где $F_{\xi}(x)$ – функция распределения случайных величин выборки ξ_i. Поскольку все величины ξ_i имеют равномерно распределение $R[0,a]$, то:

$$F_{\xi}(x) = \begin{cases} 0 & ,x<0 \\ \dfrac{x}{a} & ,0\le x\le a \\ 1 & ,a<x \end{cases}.$$

Тогда для функции $F_{(n)}(x)$ получим выражение:

$$F_{(n)}(x) = \begin{cases} 0 & ,x<0 \\ \left(\dfrac{x}{a}\right)^n & ,0\le x\le a \\ 1 & ,a<x \end{cases}.$$

Определим величину $\varphi(\xi_1,..., \xi_n;a)$ следующим образом:

$$\varphi(\xi_1,..., \xi_n;a) = F_{(n)}(\xi_{(n)}).$$

Поскольку все величины выборки ξ_i имеют равномерно распределение $R[0,a]$, то с вероятностью 1 величины $\xi_i \in [0,a]$, тогда и $\xi_{(n)} = \max\limits_{1\le i\le n} \xi_{(i)} \in [0,a]$ и следовательно:

$$\varphi(\xi_1,..., \xi_n;a) = F_{(n)}(\xi_{(n)}) = \left(\frac{\xi_{(n)}}{a}\right)^n.$$

Поскольку случайная величина $\xi_{(n)}$ подставляется в свою функцию распределения $F_{(n)}(x)$, то получаемая в результате величина $F_{(n)}(\xi_{(n)})$, как известно, имеет равномерное распределение $R[0,1]$. Таким образом:

$$\varphi(\xi_1,..., \xi_n;a) = \left(\frac{\xi_{(n)}}{a}\right)^n \sim R[0,1].$$

Легко видеть, что величина $\varphi(\xi_1,..., \xi_n;a)$ является центральной статистикой для a, действительно, распределение величины $\varphi(\xi_1,..., \xi_n;a)$ полностью известно и не зависит от a, при всех $a>0$ величина $\varphi(\xi_1,..., \xi_n;a)$ является непрерывной по a функцией, и при всех реализациях выборки $(\xi_1,..., \xi_n)$ величина $\varphi(\xi_1,..., \xi_n;a)$ является убывающей по a.

Поскольку $\varphi(\xi_1,..., \xi_n;a)$ имеет равномерное распределение $R[0,1]$, то:

$$P\{\varphi(\xi_1,..., \xi_n;a) < P_д\} = P_д,$$

тогда для любого $a>0$:

$$P\left\{\left(\frac{\xi_{(n)}}{a}\right)^n < P_{\text{д}}\right\} = P_{\text{д}},$$

$$P\left\{\frac{\xi_{(n)}}{\sqrt[m]{P_{\text{д}}}} < a\right\} = P_{\text{д}}.$$

Поскольку последнее равенство справедливо при всех $a > 0$, то:

$$\inf_{a>0} P\left\{\frac{\xi_{(n)}}{\sqrt[m]{P_{\text{д}}}} < a\right\} = P_{\text{д}}.$$

Откуда следует, что статистика $\frac{\xi_{(n)}}{\sqrt[m]{P_{\text{д}}}}$ является нижней доверительной границей для a с уровнем доверия $P_{\text{д}}$.

Аналогично,

$$P\{1 - P_{\text{д}} < \varphi(\xi_1,..., \xi_n; a)\} = 1 - (1 - P_{\text{д}}) = P_{\text{д}},$$

тогда для любого $a > 0$:

$$P\left\{1 - P_{\text{д}} < \left(\frac{\xi_{(n)}}{a}\right)^n\right\} = P_{\text{д}},$$

$$P\left\{a < \frac{\xi_{(n)}}{\sqrt[n]{1 - P_{\text{д}}}}\right\} = P_{\text{д}}.$$

Откуда,

$$\inf_{a>0} P\left\{a < \frac{\xi_{(n)}}{\sqrt[n]{1 - P_{\text{д}}}}\right\} = P_{\text{д}},$$

И по определению статистика $\frac{\xi_{(n)}}{\sqrt[n]{1 - P_{\text{д}}}}$ является верхней доверительной границей для a с уровнем доверия $P_{\text{д}}$.

Ответ:

$\frac{\max\limits_{1\le i\le n} \xi_{(i)}}{\sqrt[m]{P_{\text{д}}}}$ – нижняя доверительная граница;

$\frac{\max\limits_{1\le i\le n} \xi_{(i)}}{\sqrt[n]{1 - P_{\text{д}}}}$ – верхняя доверительная граница.

Задача 3.3.

Задана выборка $(\xi_1,..., \xi_n)$ из нормального распределения $N(m, \sigma^2)$, где m и σ – неизвестные параметры. Построить доверительные интервалы для σ^2 с уровнем доверия $P_{\text{д}}$ при:

1) $n = 10$, $P_{\text{д}} = 0.98$;

2) $n = 10$, $P_{\text{д}} = 0.9$;

3) $n = 5$, $P_{\text{д}} = 0.98$;

Решение:

Пусть статистика $\tilde{\mu}_2(\xi_1,..., \xi_n)$ обозначает исправленную выборочную дисперсию:

$$\tilde{\mu}_2(\xi_1,\dots,\xi_n) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\xi_i - \frac{1}{n}\sum_{i=1}^{n}\xi_i\right)^2 .$$

Согласно теореме Фишера статистика $\frac{n-1}{\sigma^2}\tilde{\mu}_2(\xi_1,\dots,\xi_n)$ имеет распределение хи-квадрат с числом степеней свободы $n-1$:

$$\frac{n-1}{\sigma^2}\tilde{\mu}_2(\xi_1,\dots,\xi_n) \sim \chi^2(n-1) .$$

Пусть y_1 и y_2 являются квантилями распределения $\chi^2(n-1)$ уровней $\frac{1-P_д}{2}$ и $\frac{1+P_д}{2}$ соответственно, тогда:

$$P\left\{y_1 < \frac{n-1}{\sigma^2}\tilde{\mu}_2(\xi_1,\dots,\xi_n) < y_2\right\} = P_д ,$$

$$P\left\{\frac{(n-1)}{y_2}\tilde{\mu}_2(\xi_1,\dots,\xi_n) < \sigma^2 < \frac{(n-1)}{y_1}\tilde{\mu}_2(\xi_1,\dots,\xi_n)\right\} = P_д .$$

Поскольку полученное равенство справедливо при всех $\sigma > 0$, то интервал:

$$\left(\frac{(n-1)}{y_2}\tilde{\mu}_2(\xi_1,\dots,\xi_n); \frac{(n-1)}{y_1}\tilde{\mu}_2(\xi_1,\dots,\xi_n)\right)$$

является доверительным интервалом для σ^2 с уровнем доверия $P_д$.

В следующей далее таблице приведены интервалы для случаев 1)-3).

n	$P_д$	$\frac{1-P_д}{2}$	$\frac{1+P_д}{2}$	y_1	y_2	$\left(\frac{(n-1)}{y_2}\tilde{\mu}_2; \frac{(n-1)}{y_1}\tilde{\mu}_2\right)$
10	0.98	0.01	0.99	2.09	21.66	$(0.42\,\tilde{\mu}_2; 4.31\,\tilde{\mu}_2)$
10	0.9	0.05	0.95	3.33	16.92	$(0.55\,\tilde{\mu}_2; 2.7\,\tilde{\mu}_2)$
5	0.98	0.01	0.99	0.55	15.09	$(0.27\,\tilde{\mu}_2; 7.27\,\tilde{\mu}_2)$

Ответ:

1) $(0.42\,\tilde{\mu}_2; 4.31\,\tilde{\mu}_2)$,

2) $(0.55\,\tilde{\mu}_2; 2.7\,\tilde{\mu}_2)$,

3) $(0.27\,\tilde{\mu}_2; 7.27\,\tilde{\mu}_2)$,

где $\tilde{\mu}_2 = \tilde{\mu}_2(\xi_1,\dots,\xi_n) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\xi_i - \frac{1}{n}\sum_{i=1}^{n}\xi_i\right)^2$.

Задача 3.4.

Задана выборка $(\xi_1,\dots,\xi_n)$ из показательного распределения $E(a)$, где $a > 0$ неизвестный параметр. Построить:

1) доверительный интервал для a с уровнем доверия $P_д$,

2) нижнюю доверительную границу для a с уровнем доверия $P_д$,

3) верхнюю доверительную границу для a с уровнем доверия $P_д$.

Решение:

1) Рассмотрим величину $\varphi(\xi_1,\dots,\xi_n;a)$:

$$\varphi(\xi_1,\dots,\xi_n;a) = \sum_{i=1}^{n} 2a\xi_i$$

и определим распределение этой величины. Поскольку $\varphi(\xi_1,..., \xi_n; a)$ является суммой независимых случайных величин, то наиболее просто установить распределение с помощью характеристической функции $\varphi(\xi_1,..., \xi_n; a)$.

Поскольку случайные величины выборки ξ_i независимы и имеют показательное распределение $E(a)$, то совместная плотность вероятности $p_\xi(x_1,..., x_n; a)$ случайных величин выборки имеет вид:

$$p_\xi(x_1,..., x_n; a) = \prod_{i=1}^{n} ae^{-ax_i}.$$

Пусть случайные величины $\eta_i = 2a\xi_i$ или иначе $\eta_i = f_i(\xi_i)$, где $f_i(x) = 2ax$:

$$\begin{cases} \eta_1 = f_1(\xi_1) = 2a\xi_1 \\ \eta_2 = f_2(\xi_2) = 2a\xi_2 \\ \quad\cdots \\ \eta_n = f_n(\xi_n) = 2a\xi_n \end{cases} \leftrightarrow \begin{cases} \xi_1 = f_1^{-1}(\eta_1) = \dfrac{\xi_1}{2a} \\ \xi_2 = f_2^{-1}(\eta_2) = \dfrac{\xi_2}{2a} \\ \quad\cdots \\ \xi_n = f_n^{-1}(\eta_n) = \dfrac{\xi_n}{2a} \end{cases},$$

где $f_i^{-1}(y) = \dfrac{y}{2a}$.

Совместная плотность вероятности $p_\eta(y_1,..., y_n; a)$ величин η_i в соответствии с правилом преобразования плотностей вероятностей выражается через плотность вероятности $p_\xi(x_1,..., x_n; a)$ случайных величин выборки ξ_i следующим образом:

$$p_\eta(y_1,..., y_n; a) = p_\xi(f_1^{-1}(y_1),..., f_n^{-1}(y_n); a)\frac{\partial(f_1^{-1}, f_2^{-1},..., f_n^{-1})}{\partial(y_1, y_2,.., y_n)},$$

где якобиан преобразования:

$$\frac{\partial(f_1^{-1}, f_2^{-1},..., f_n^{-1})}{\partial(y_1, y_2,.., y_n)} = \begin{vmatrix} \dfrac{\partial f_1^{-1}}{\partial y_1} & \dfrac{\partial f_2^{-1}}{\partial y_2} & \cdots & \dfrac{\partial f_n^{-1}}{\partial y_n} \\ \dfrac{\partial f_2^{-1}}{\partial y_1} & \dfrac{\partial f_2^{-1}}{\partial y_2} & \cdots & \dfrac{\partial f_2^{-1}}{\partial y_n} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial f_n^{-1}}{\partial y_1} & \dfrac{\partial f_n^{-1}}{\partial y_2} & \cdots & \dfrac{\partial f_n^{-1}}{\partial y_n} \end{vmatrix} = \begin{vmatrix} \dfrac{1}{2a} & 0 & \cdots & 0 \\ 0 & \dfrac{1}{2a} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \dfrac{1}{2a} \end{vmatrix} = \left(\frac{1}{2a}\right)^n.$$

Таким образом,

$$p_\eta(y_1,..., y_n; a) = \prod_{i=1}^{n} ae^{-a\frac{y_i}{2a}}\left(\frac{1}{2a}\right)^n = \left(\frac{1}{2a}\right)^n a^n \prod_{i=1}^{n} e^{-\frac{y_i}{2}} = \left(\frac{1}{2}\right)^n \prod_{i=1}^{n} e^{-\frac{y_i}{2}} = \prod_{i=1}^{n}\frac{1}{2}e^{-\frac{y_i}{2}}.$$

Исходя из вида плотности вероятности $p_\eta(y_1,..., y_n; a)$, легко видеть, что случайные величины η_i независимы в совокупности и имеют показательное распределение $E\left(\dfrac{1}{2}\right)$. Вернемся к рассмотрению величины $\varphi(\xi_1,..., \xi_n; a)$:

$$\varphi(\xi_1,..., \xi_n; a) = \sum_{i=1}^{n} 2a\xi_i = \sum_{i=1}^{n}\eta_i.$$

Поскольку случайные величины η_i независимы в совокупности, то характеристическая функция $\chi_\varphi(t)$ суммы $\sum_{i=1}^{n}\eta_i$ равна произведению характеристических функций $\chi_{\eta_i}(t)$ случайных величин η_i:

$$\chi_\varphi(t) = \prod_{i=1}^{n}\chi_{\eta_i}(t).$$

Все случайные величины η_i имеют одинаковое распределение $E\left(\frac{1}{2}\right)$, характеристическая функция которого известна:

$$\chi_{\eta_i}(t) = \frac{1}{(1-2it)}.$$

Таким образом,

$$\chi_\varphi(t) = \prod_{i=1}^{n}\frac{1}{(1-2it)} = (1-2it)^{-n} = (1-2it)^{-\frac{2n}{2}}.$$

Известно что распределение хи-квадрат с k степенями свободы имеет характеристическую функцию $(1-2it)^{-\frac{k}{2}}$, откуда следует, что величина $\varphi(\xi_1,..., \xi_n;a)$ имеет распределение хи-квадрат с $2n$ степенями свободы $\chi^2(2n)$:

$$\varphi(\xi_1,...,\xi_n;a) = \sum_{i=1}^{n}2a\xi_i = \sum_{i=1}^{n}\eta_i \sim \chi^2(2n)$$

Легко видеть, что величина $\varphi(\xi_1,..., \xi_n;a)$ является центральной статистикой для a: распределение величины $\varphi(\xi_1,..., \xi_n;a)$ полностью известно и не зависит от a, при всех $a>0$ величина $\varphi(\xi_1,..., \xi_n;a)$ является непрерывной по a функцией, и при всех реализациях выборки $(\xi_1,..., \xi_n)$ величина $\varphi(\xi_1,..., \xi_n;a)$ является возрастающей по a.

2) Пусть теперь z_1 и z_2 являются квантилями распределения хи-квадрат с $2n$ степенями свободы $\chi^2(2n)$ уровней $\frac{1-P_д}{2}$ и $\frac{1+P_д}{2}$ соответственно, тогда очевидно:

$$P\{z_1 < \varphi(\xi_1,..., \xi_n;a) < z_2\} = F_{x^2_{2n}}(z_2) - F_{x^2_{2n}}(z_1) = \frac{1+P_д}{2} - \frac{1-P_д}{2} = P_д,$$

где $F_{x^2_{2n}}(z)$ – функция распределения $\chi^2(2n)$. Преобразовывая неравенство слева, получим следующее равенство вероятностей:

$$P\left\{z_1 < \sum_{i=1}^{n}2a\xi_i < z_2\right\} = P_д,$$

$$P\left\{\frac{z_1}{2\sum_{i=1}^{n}\xi_i} < a < \frac{z_2}{2\sum_{i=1}^{n}\xi_i}\right\} = P_д.$$

Поскольку последнее равенство справедливо для любого произвольного $a>0$, то по определению интервал:

$$\left(\frac{z_1}{2\sum_{i=1}^{n}\xi_i}; \frac{z_2}{2\sum_{i=1}^{n}\xi_i}\right)$$

является доверительным интервалом для a с уровнем доверия $P_д$.

3) Поскольку центральная статистика $\varphi(\xi_1,..., \xi_n; a)$:

$$\varphi(\xi_1,..., \xi_n; a) = 2a\sum_{i=1}^{n} \xi_i$$

является функцией возрастающей по a, то для построения нижней доверительной границы необходимо ограничить центральную статистику снизу:

$$P\{z_l < \varphi(\xi_1,..., \xi_n; a)\} = P_д .$$

Откуда:

$$1 - P\{\varphi(\xi_1,..., \xi_n; a) \le z_l\} = P_д ,$$

$$P\{\varphi(\xi_1,..., \xi_n; a) \le z_l\} = 1 - P_д ,$$

и величина z_l должна быть равна квантили уровня $1 - P_д$ распределения центральной статистики $\chi^2(2n)$.

Для получения нижней доверительной границы остается лишь преобразовать неравенство:

$$P\{z_l < \varphi(\xi_1,..., \xi_n; a)\} = P_д ,$$

$$P\left\{z_l < 2a\sum_{i=1}^{n} \xi_i\right\} = P_д ,$$

$$P\left\{\frac{z_l}{2\sum_{i=1}^{n} \xi_i} < a\right\} = P_д .$$

Таким образом, статистика $\dfrac{z_l}{2\sum_{i=1}^{n} \xi_i}$ является нижней доверительной границей для a с уровнем доверия $P_д$.

4) Аналогично для построения верхней доверительной границы необходимо ограничить центральную статистику сверху:

$$P\{\varphi(\xi_1,..., \xi_n; a) < z_u\} = P_д ,$$

откуда непосредственно следует, что z_u является квантилью уровня $P_д$ распределения центральной статистики $\chi^2(2n)$.

Далее преобразование неравенства, стоящего под знаком вероятности приводит к верхней доверительной границе:

$$P\{\varphi(\xi_1,..., \xi_n; a) < z_u\} = P_д ,$$

$$P\left\{2a\sum_{i=1}^{n} \xi_i < z_u\right\} = P_д ,$$

$$P\left\{a < \frac{z_u}{2\sum_{i=1}^{n} \xi_i}\right\} = P_д .$$

Таким образом, статистика $\dfrac{z_u}{2\sum\limits_{i=1}^{n}\xi_i}$ является верхней доверительной границей для a с уровнем доверия $P_д$.

Ответ:

1) доверительный интервал:

$$\left(\frac{z_1}{2\sum\limits_{i=1}^{n}\xi_i};\frac{z_2}{2\sum\limits_{i=1}^{n}\xi_i}\right),$$

где z_1 и z_2 являются квантилями распределения хи-квадрат с $2n$ степенями свободы уровней $\dfrac{1-P_д}{2}$ и $\dfrac{1+P_д}{2}$ соответственно;

2) нижняя доверительная граница:

$$\frac{z_l}{2\sum\limits_{i=1}^{n}\xi_i},$$

где z_l – квантиль уровня $1-P_д$ распределения $\chi^2(2n)$;

3) верхняя доверительная граница:

$$\frac{z_u}{2\sum\limits_{i=1}^{n}\xi_i},$$

z_u – квантиль уровня $P_д$ распределения $\chi^2(2n)$.

Задача 3.5.

Задана выборка $(\xi_1,..., \xi_n)$ из распределения Коши $Co\,(a,1)$ с неизвестным параметром a. Построить приближенный доверительный интервал для a с уровнем доверия $P_д$.

Решение:

Построим оценку параметра a по методу порядковых статистик. Известно, что распределение Коши $Co\,(a,b)$ имеет плотность вероятности:

$$p_\xi(x;a)=\frac{1}{\pi}\frac{b}{(x-a)^2+b^2}.$$

Легко видеть, что функция плотности вероятности симметрична относительно прямой $x=a$, откуда следует, что медиана (квантиль уровня 0.5) $x_{0.5}$ совпадает с параметром a:

$$a=x_{0.5}.$$

Взяв в качестве оценки медианы $x_{0.5}$ порядковую статистику с номером $\left[\dfrac{n}{2}\right]+1$, получим оценку параметра a:

$$\hat{a}(\xi_1,..., \xi_n)=\xi_{\left(\left[\frac{n}{2}\right]+1\right)}.$$

В соответствии с теоремой Крамера распределение порядковой статистики $\xi_{\left(\left[\frac{n}{2}\right]+1\right)}$ с ростом объема выборки n стремится к нормальному распределению $N\left(x_{0.5}, \frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{n\frac{1}{\pi^2}}\right)$ или иначе $N\left(a, \frac{\pi^2}{4n}\right)$, поскольку $a = x_{0.5}$. Откуда следует, что распределение оценки $\hat{a}(\xi_1,..., \xi_n)$ при больших n можно считать приближенно нормальным:

$$\hat{a}(\xi_1,..., \xi_n) \sim N\left(a, \frac{\pi^2}{4n}\right).$$

Рассмотрим величину $\varphi(\xi_1,..., \xi_n; a)$, которая представляет собой нормированную оценку $\hat{a}(\xi_1,..., \xi_n)$:

$$\varphi(\xi_1,..., \xi_n; a) = \frac{\hat{a}(\xi_1,..., \xi_n) - a}{\sqrt{\frac{\pi^2}{4n}}} = \frac{\xi_{\left(\left[\frac{n}{2}\right]+1\right)} - a}{\sqrt{\frac{\pi^2}{4n}}}.$$

Легко видеть, что при больших n величина $\varphi(\xi_1,..., \xi_n; a)$ имеет приближенно стандартное нормальное распределение $N(0,1)$:

$$\varphi(\xi_1,..., \xi_n; a) = \frac{\xi_{\left(\left[\frac{n}{2}\right]+1\right)} - a}{\sqrt{\frac{\pi^2}{4n}}} \sim N(0,1)$$

Откуда следует, что величина $\varphi(\xi_1,..., \xi_n; a)$ является центральной статистикой для параметра a, поскольку распределение $\varphi(\xi_1,..., \xi_n; a)$ полностью известно, функция $\varphi(\xi_1,..., \xi_n; a)$ как функция параметра a является непрерывной и одновременно при всех реализациях выборки величина $\varphi(\xi_1,..., \xi_n; a)$ является убывающей по a функцией.

Пусть y является квантилью стандартного нормального распределения $N(0,1)$ уровня $\frac{1+P_д}{2}$, тогда:

$$P\{-y < \varphi(\xi_1,..., \xi_n; a) < y\} \approx P_д,$$

$$P\left\{-y < \frac{\xi_{\left(\left[\frac{n}{2}\right]+1\right)} - a}{\sqrt{\frac{\pi^2}{4n}}} < y\right\} \approx P_д,$$

$$P\left\{\xi_{\left(\left[\frac{n}{2}\right]+1\right)} - y\sqrt{\frac{\pi^2}{4n}} < a < \xi_{\left(\left[\frac{n}{2}\right]+1\right)} + y\sqrt{\frac{\pi^2}{4n}}\right\} \approx P_д.$$

Поскольку последнее приближенное равенство справедливо при всех возможных конечных значениях a, то интервал:

$$\left(\xi_{\left(\left[\frac{n}{2}\right]+1\right)} - y\sqrt{\frac{\pi^2}{4n}};\xi_{\left(\left[\frac{n}{2}\right]+1\right)} + y\sqrt{\frac{\pi^2}{4n}}\right)$$

можно считать приближенным доверительным для параметра a с уровнем доверия $P_д$.

Ответ:

$\left(\xi_{\left(\left[\frac{n}{2}\right]+1\right)} - y\sqrt{\frac{\pi^2}{4n}};\xi_{\left(\left[\frac{n}{2}\right]+1\right)} + y\sqrt{\frac{\pi^2}{4n}}\right)$, где y – квантиль стандартного нормального распределения $N(0,1)$ уровня $\frac{1+P_д}{2}$.

Задача 3.6.

Задана выборка $(\xi_1,..., \xi_n)$ из распределения Пуассона $Po(\lambda)$ с неизвестным параметром λ. Построить приближенный доверительный интервал для λ с уровнем доверия $P_д$.

Решение:

Построим по заданной выборке оценку $\hat{\lambda}(\xi_1,..., \xi_n)$ параметра λ (например, методом моментов):

$$\hat{\lambda}(\xi_1,..., \xi_n) = \frac{1}{n}\sum_{i=1}^{n}\xi_i .$$

К величинам выборки ξ_i применима центральная предельная теорема, поэтому с ростом n распределение суммы (и оценки) будет стремиться к нормальному распределению $N\left(\lambda, \frac{\lambda}{n}\right)$. Нормируя оценку $\hat{\lambda}(\xi_1,..., \xi_n)$, получим величину $\varphi(\xi_1,..., \xi_n;\lambda)$:

$$\varphi(\xi_1,..., \xi_n;\lambda) = \frac{\hat{\lambda}(\xi_1,..., \xi_n) - \lambda}{\sqrt{\frac{\lambda}{n}}} \sim N(0,1),$$

распределение которой при больших n приближенно совпадает со стандартным нормальным распределением.

Заметим, что величина $\varphi(\xi_1,..., \xi_n;\lambda)$ является центральной статистикой для параметра λ, поскольку распределение величины $\varphi(\xi_1,..., \xi_n;\lambda)$ известно (приближенно), и сама величина $\varphi(\xi_1,..., \xi_n;\lambda)$ как функция λ является непрерывной и убывающей одновременно при всех реализациях выборки $(\xi_1,..., \xi_n)$.

Пусть y является квантилью стандартного нормального распределения $N(0,1)$ уровня $\frac{1+P_д}{2}$, тогда выполняется приближенное равенство:

$$P\{-y < \varphi(\xi_1,..., \xi_n;\lambda) < y\} \approx P_д,$$

$$P\left\{-y < \frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\lambda}{n}}} < y\right\} \approx P_д,$$

$$P\left\{\left|\frac{\hat{\lambda}-\lambda}{\sqrt{\frac{\lambda}{n}}}\right|<y\right\}\approx P_{д},$$

$$P\left\{(\hat{\lambda}-\lambda)^2<y^2\frac{\lambda}{n}\right\}\approx P_{д},$$

$$P\left\{\lambda^2-\left(2\hat{\lambda}+\frac{y^2}{n}\right)\lambda+\hat{\lambda}^2<0\right\}\approx P_{д},$$

$$P\left\{\frac{2\hat{\lambda}+\frac{y^2}{n}-\sqrt{\left(2\hat{\lambda}+\frac{y^2}{n}\right)^2-4\hat{\lambda}^2}}{2}<\lambda<\frac{2\hat{\lambda}+\frac{y^2}{n}+\sqrt{\left(2\hat{\lambda}+\frac{y^2}{n}\right)^2-4\hat{\lambda}^2}}{2}\right\}\approx P_{д},$$

$$P\left\{\frac{2\hat{\lambda}+\frac{y^2}{n}-\sqrt{4\hat{\lambda}\frac{y^2}{n}+\frac{y^4}{n^2}}}{2}<\lambda<\frac{2\hat{\lambda}+\frac{y^2}{n}+\sqrt{4\hat{\lambda}\frac{y^2}{n}+\frac{y^4}{n^2}}}{2}\right\}\approx P_{д}$$

Пренебрегая слагаемыми порядка $\frac{1}{n}$ вне корня и слагаемыми порядка $\frac{1}{n^2}$ под корнем, получим приближенное равенство:

$$P\left\{\frac{2\hat{\lambda}-\sqrt{4\hat{\lambda}\frac{y^2}{n}}}{2}<\lambda<\frac{2\hat{\lambda}+\sqrt{4\hat{\lambda}\frac{y^2}{n}}}{2}\right\}\approx P_{д}$$

$$P\left\{\hat{\lambda}-\sqrt{\hat{\lambda}\frac{y^2}{n}}<\lambda<\hat{\lambda}+\sqrt{\hat{\lambda}\frac{y^2}{n}}\right\}\approx P_{д}.$$

Поскольку последнее приближенное равенство справедливо при всех допустимых значениях параметра λ, то интервал:

$$\left(\hat{\lambda}-\sqrt{\hat{\lambda}\frac{y^2}{n}};\hat{\lambda}-\sqrt{\hat{\lambda}\frac{y^2}{n}}\right),$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i-\sqrt{\frac{1}{n}\sum_{i=1}^{n}\xi_i\frac{y^2}{n}};\frac{1}{n}\sum_{i=1}^{n}\xi_i-\sqrt{\frac{1}{n}\sum_{i=1}^{n}\xi_i\frac{y^2}{n}}\right)$$

можно считать приближенным доверительным интервалом для параметра λ с уровнем доверия $P_{д}$.

Ответ:

$\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i-\sqrt{\frac{1}{n}\sum_{i=1}^{n}\xi_i\frac{y^2}{n}};\frac{1}{n}\sum_{i=1}^{n}\xi_i-\sqrt{\frac{1}{n}\sum_{i=1}^{n}\xi_i\frac{y^2}{n}}\right)$, где y – квантиль стандартного нормального распределения $N(0{,}1)$ уровня $\frac{1+P_{д}}{2}$.

Задача 3.7.

Задана выборка $(\xi_1,..., \xi_n)$ из распределения с плотность вероятности $p_\xi(x;q)$:

$$p_\xi(x;q) = q\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} + (1-q)\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}},$$

где q – неизвестный параметр ($0 < q < 1$). Построить приближенный доверительный интервал для q с уровнем доверия P_∂.

Решение:

Построим оценку параметра q по методу моментов. Заметим, что если величина ξ имеет плотность вероятности $p_\xi(x;q)$, тогда математическое ожидание $M_q[\xi]$:

$$M_q[\xi] = \int_{-\infty}^{\infty} xp_\xi(x;q)dx = \int_{-\infty}^{\infty} x\left(q\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} + (1-q)\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}}\right)dx =$$

$$= q\int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx + (1-q)\int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}}dx = q\cdot 0 + (1-q)\cdot 1 = 1-q.$$

Откуда,

$$q = 1 - M_q[\xi]$$

и оценка по методу моментов $\hat{q}(\xi_1,..., \xi_n)$ для параметра q:

$$\hat{q}(\xi_1,..., \xi_n) = 1 - \frac{1}{n}\sum_{i=1}^{n}\xi_i.$$

К величинам выборки ξ_i применима центральная предельная теорема, поэтому с ростом n распределение суммы $\sum_{i=1}^{n}\xi_i$ будет стремиться к нормальному распределению, поэтому распределение оценки $\hat{q}(\xi_1,..., \xi_n)$ будет также стремиться к нормальному распределению с параметрами $N\left(M_q[\hat{q}], D_q[\hat{q}]\right)$:

$$M_q[\hat{q}] = 1 - M_q\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i\right] = 1 - \frac{1}{n}\sum_{i=1}^{n}M_q[\xi_i] = 1 - \frac{1}{n}n(1-q) = q,$$

$$D_q[\hat{q}] = D_q\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i\right] = \frac{1}{n^2}D_q\left[\sum_{i=1}^{n}\xi_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}D_q[\xi_i].$$

Дисперсию $D_q[\xi_i]$ удобно вычислить с помощью второго начального момента $M_q[\xi_i^2]$:

$$M_q[\xi_i^2] = \int_{-\infty}^{\infty} x^2 p_\xi(x;q)dx = \int_{-\infty}^{\infty} x^2\left(q\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} + (1-q)\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}}\right)dx =$$

$$= q\int_{-\infty}^{\infty} x^2\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx + (1-q)\int_{-\infty}^{\infty} x^2\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}}dx = q\cdot(1+0^2) + (1-q)\cdot(1+1^1) = q + (1-q)\cdot 2.$$

Тогда,

$$D_q[\xi_i] = M_q[\xi_i^2] - \left(M_q[\xi_i]\right)^2 = q + (1-q)\cdot 2 - (1-q)^2 = 1 + q - q^2,$$

$$D_q[\hat{q}] = \frac{1}{n^2}n(1+q-q^2) = \frac{1+q-q^2}{n}$$

Таким образом, при больших n распределение оценки $\hat{q}(\xi_1,..., \xi_n)$ приближенно совпадает с нормальным распределением:

$$\hat{q}(\xi_1,..., \xi_n) \sim N\left(q, \frac{1+q-q^2}{n}\right),$$

при $n \to \infty$.

Нормируя оценку $\hat{q}(\xi_1,..., \xi_n)$, получим величину $\varphi(\xi_1,..., \xi_n; q)$:

$$\varphi(\xi_1,..., \xi_n; q) = \frac{\hat{q}(\xi_1,..., \xi_n) - q}{\sqrt{\frac{1+q-q^2}{n}}} \sim N(0,1),$$

при $n \to \infty$.

Заметим, что величина $\varphi(\xi_1,..., \xi_n; \lambda)$ является центральной статистикой для параметра q , поскольку распределение величины $\varphi(\xi_1,..., \xi_n; q)$ известно (приближенно), и сама величина $\varphi(\xi_1,..., \xi_n; q)$ как функция q является непрерывной и убывающей одновременно при всех реализациях выборки $(\xi_1,..., \xi_n)$.

Пусть y является квантилью стандартного нормального распределения $N(0,1)$ уровня $\frac{1+P_д}{2}$, тогда выполняется приближенное равенство:

$$P\{-y < \varphi(\xi_1,..., \xi_n; q) < y\} \approx P_д ,$$

$$P\left\{-y < \frac{\hat{q}-q}{\sqrt{\frac{1+q-q^2}{n}}} < y\right\} \approx P_д ,$$

$$P\left\{\left|\frac{\hat{q}-q}{\sqrt{\frac{1+q-q^2}{n}}}\right| < y\right\} \approx P_д ,$$

$$P\left\{\left|\hat{q}-q\right| < y\sqrt{\frac{1+q-q^2}{n}}\right\} \approx P_д$$

Заметим, что дальнейшее преобразование неравенства потребует возведения в квадрат и вызовет необходимость решения громоздкого квадратного уравнения. Этого можно избежать, если исключить параметр q в выражении, стоящем справа от знака больше, например, заменив всю правую часть оценкой сверху. Прежде всего, заметим что:

$$\sqrt{\frac{1+q-q^2}{n}} \le \max_{0<q<1} \sqrt{\frac{1+q-q^2}{n}},$$

и если справедливо неравенство:

$$\left|\hat{q}-q\right| < y\sqrt{\frac{1+q-q^2}{n}},$$

тогда выполняется и неравенство:

$$\left|\hat{q}-q\right| < \max_{0<q<1} \sqrt{\frac{1+q-q^2}{n}},$$

то есть из выполнения первого неравенства следует выполнение второго неравенства:

$$\left|\hat{q}-q\right| < y\sqrt{\frac{1+q-q^2}{n}} \Rightarrow \left|\hat{q}-q\right| < \max_{0<q<1} \sqrt{\frac{1+q-q^2}{n}} .$$

Отсюда следует, что и вероятность выполнения второго неравенства не меньше вероятности выполнения второго неравенства:

$$P\left\{\left|\hat{q}-q\right| < y\max_{0<q<1}\sqrt{\frac{1+q-q^2}{n}}\right\} \geq P\left\{\left|\hat{q}-q\right| < y\sqrt{\frac{1+q-q^2}{n}}\right\} \approx P_д .$$

Поскольку $\max\limits_{0<q<1}\sqrt{\frac{1+q-q^2}{n}} = \frac{5}{4}$, то:

$$P\left\{\left|\hat{q}-q\right| < y\frac{5}{4}\right\} \geq P_д ,$$

$$P\left\{\hat{q} - y\frac{5}{4} < q < \hat{q} + y\frac{5}{4}\right\} \geq P_д .$$

Поскольку полученное неравенство справедливо при всех допустимых значениях параметра q, то интервал:

$$\left(\hat{q} - y\frac{5}{4}; \hat{q} - y\frac{5}{4}\right),$$

$$\left(1 - \frac{1}{n}\sum_{i=1}^{n}\xi_i - y\frac{5}{4}; 1 - \frac{1}{n}\sum_{i=1}^{n}\xi_i + y\frac{5}{4}\right),$$

является приближенным доверительным интервалом для параметра q с уровнем доверия не меньше $P_д$.

Ответ:

$\left(1 - \frac{1}{n}\sum_{i=1}^{n}\xi_i - y\frac{5}{4}; 1 - \frac{1}{n}\sum_{i=1}^{n}\xi_i + y\frac{5}{4}\right)$, где y – квантиль стандартного нормального распределения $N(0,1)$ уровня $\frac{1+P_д}{2}$.

Задача 3.8.

Задана случайная величина ξ, имеющая нормальное распределение $N(m,\sigma^2)$, где m и σ^2 – неизвестные параметры. Построить приближенную верхнюю границу для σ с уровнем доверия не меньше $P_д$.

Решение:

Рассмотрим величину $\frac{|\xi|}{\sigma}$ и неравенство $y < \frac{|\xi|}{\sigma}$: если бы величина y была выбрана таким образом, что вероятность:

$$P\left\{y < \frac{|\xi|}{\sigma}\right\} \geq P_д ,$$

то отсюда следовало бы что:

$$P\left\{\sigma < \frac{|\xi|}{y}\right\} \geq P_д ,$$

то есть статистика $\frac{|\xi|}{y}$ являлась бы верхней доверительной границей с уровнем доверия не меньше $P_д$. Остается лишь определить значение величины y.

Для определения величины y рассмотрим исходную вероятность:

$$P\left\{y < \frac{|\xi|}{\sigma}\right\} = 1 - P\left\{\frac{|\xi|}{\sigma} < y\right\} = 1 - P\left\{-y < \frac{\xi}{\sigma} < y\right\}.$$

По условию величина ξ имеет нормальное распределение $N(m,\sigma^2)$, поэтому величина $\frac{\xi}{\sigma}$ имеет нормальное распределение $N\left(\frac{m}{\sigma},1\right)$, тогда:

$$P\left\{y<\frac{|\xi|}{\sigma}\right\}=1-P\left\{-y<\frac{\xi}{\sigma}<y\right\}=1-\int_{-y}^{y}\frac{1}{\sqrt{2\pi}}e^{-\frac{\left(x-\frac{m}{\sigma}\right)^2}{2}}dx\geq 1-\max_{m,\sigma}\int_{-y}^{y}\frac{1}{\sqrt{2\pi}}e^{-\frac{\left(x-\frac{m}{\sigma}\right)^2}{2}}dx\ .$$

Легко видеть, что наибольшего значения величина интеграла принимает при $\frac{m}{\sigma}=0$, тогда:

$$P\left\{y<\frac{|\xi|}{\sigma}\right\}\geq 1-\int_{-y}^{y}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=1-(\Phi(y)-\Phi(-y))=1-\Phi(y)+\Phi(-y)=$$

$$=\Phi(-y)+\Phi(-y)=2\Phi(-y),$$

где $\Phi(x)$ – функция распределения стандартного нормального распределения $N(0,1)$.

Пусть $\Phi(-y)=\frac{P_д}{2}$, тогда:

$$P\left\{y<\frac{|\xi|}{\sigma}\right\}\geq 2\Phi(-y)=2\frac{P_д}{2}=P_д\,,$$

что и требовалось для построения приближенной верхней границы.

Ответ:

$\frac{|\xi|}{y}$, где y – квантиль уровня $\frac{P_д}{2}$ стандартного нормального распределения $N(0,1)$.