Домашняя работа №1.

Задача №1

У Дианы есть выборка $X_1, \dots, X_n \sim iid\, N(0, \sigma^2)$. Она собирается использовать для параметра σ нестандартную оценку $\hat{\sigma} = \frac{1}{n} \sum_{i=1}^{n} |x_i|$. Вручную:

1. Найдите плотность распределения для случайной величины $|X|$.
2. Проверьте, является ли оценка $\hat{\sigma}$ смещенной, найдите её bias, если оценка смещена, то скорректируйте её
3. Найдите дисперсию скорректированной оценки.
4. Определите, является ли оценка состоятельной
5. Найдите для несмещенной оценки MSE

Решение:

(1) $F_{|x|}(t) = \mathbb{P}(|x| \le t) = \mathbb{P}(-t \le x \le t) = F_z(t) - F_z(-t) = F_z(t) - (1 - F_z(t)) =$

$= \begin{cases} 2F_z(t) - 1, & t \ge 0 \\ 0, & t < 0 \end{cases}$

$$f_{|x|}(t) = F'_{|x|}(t) = \begin{cases} 2f_x(t), & t \ge 0 \\ 0, & t < 0 \end{cases} = \begin{cases} 2 \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{\frac{-t^2}{2\sigma^2}}, & t \ge 0 \\ 0, & t < 0 \end{cases}$$

Ответ. $f_{|x|}(t) = \begin{cases} 2 \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-t^2}{2\sigma^2}}, & t \ge 0 \\ 0, & t < 0 \end{cases}$

(2) Несмещенность: $\mathbb{E}(\hat{\sigma}) = \sigma$

$$\mathbb{E}(\hat{\sigma}) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} |x_i|\right) = \frac{1}{n} \mathbb{E}\left(\sum_{i=1}^{n} |x_i|\right) = \frac{1}{n} \cdot n\, \mathbb{E}(|x_i|) = \mathbb{E}(|x_i|)$$

$$\mathbb{E}(|x_i|) = \int_{-\infty}^{\infty} f_{|x|}(x)\, |x|\, dx = \int_{-\infty}^{0} f_{|x|}(x) \cdot (-x)\, dx + \int_{0}^{+\infty} f_{|x|}(x) \cdot x\, dx =$$

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

$$= \left\{ \begin{aligned} & u = -\frac{x^2}{2\sigma^2} \\ & du = -\frac{2x}{2\sigma^2} dx = -\frac{x}{\sigma^2} dx \\ & -\sigma^2 du = x\, dx \end{aligned} \right\} = \frac{2}{\sqrt{2\pi}\sigma} \int_{0}^{-\infty} e^{u} \cdot (-\sigma^2)\, du = -\frac{2\sigma^2}{\sqrt{2\pi}\sigma} \int_{0}^{-\infty} e^{u}\, du =$$

также меняются пределы интегрирования

$$= \frac{-2\sigma}{\sqrt{2\pi}} \cdot e^{u} \Big|_0^{-\infty} = \frac{-2\sigma}{\sqrt{2\pi}} \cdot (e^{-\infty} - e^{0}) = -\frac{2\sigma}{\sqrt{2\pi}} (0 - 1) = \frac{2\sigma}{\sqrt{2\pi}}$$

Таким образом оценка смещена

$$\text{bias}(\hat{\sigma}) = \mathbb{E}(\hat{\sigma}) - \sigma = \frac{2\sigma}{\sqrt{2\pi}} - \sigma = \frac{(2-\sqrt{2\pi})\sigma}{\sqrt{2\pi}}$$

Корректировка: $\hat{\sigma}_{кор} = \frac{\sqrt{2\pi}}{2} \hat{\sigma}$

Ответ: оценка смещена, $\text{bias}(\hat{\sigma}) = \frac{2-\sqrt{2\pi}\,\sigma}{\sqrt{2\pi}}$, $\hat{\sigma}_{кор} = \frac{\sqrt{2\pi}}{2}\hat{\sigma}$

(3)

$$\operatorname{Var}(\hat{\sigma}_{кор}) = \operatorname{Var}\left(\frac{\sqrt{2\pi}}{2}\hat{\sigma}\right) = \operatorname{Var}\left(\frac{\sqrt{2\pi}}{2} \cdot \frac{1}{n} \sum_{i=1}^{n} |X_i|\right) =$$
$$= \left(\frac{\sqrt{2\pi}}{2n}\right)^2 \operatorname{Var}\left(\sum_{i=1}^{n} |X_i|\right) = \frac{2\pi}{4n^2} \cdot n \operatorname{Var}(|X_1|) = \frac{\pi}{2n} \operatorname{Var}(|X_1|)$$

$$\operatorname{Var}(|X_1|) = \mathbb{E}(|X_1|^2) - [\mathbb{E}(|X_1|)]^2$$

$$\mathbb{E}(|X_1|^2) = \int_{-\infty}^{\infty} f_{|X|}(x) \cdot |x|^2 dx = \int_{-\infty}^{0} -x^2 f_{|X|}(x) dx + \int_0^{+\infty} x^2 \cdot f_{|X|}(x) dx =$$
$$= \int_{-\infty}^{0} -x^2 \cdot 0 \, dx + \int_0^{+\infty} x^2 \cdot \frac{2}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{x^2}{2\sigma^2}} dx =$$
$$= 0 + \frac{2}{\sqrt{2\pi}\,\sigma} \int_0^{+\infty} x^2 e^{-\frac{x^2}{2\sigma^2}} dx = \left\{\begin{array}{ll} \int fg' = fg - \int f'g & \\ f = x & g' = x e^{-\frac{x^2}{2\sigma^2}} \\ f' = 1 & g = -\sigma^2 e^{-\frac{x^2}{2\sigma^2}} \end{array}\right\} =$$

$$= \frac{2}{\sqrt{2\pi}\sigma} \left(\underbrace{x \cdot (-\sigma)^2 e^{-\frac{x^2}{2\sigma^2}} \Big|_0^{+\infty}}_{0} - \int_0^{+\infty} -\sigma^2 e^{-\frac{x^2}{2\sigma^2}} dx \right) =$$

$$= \frac{2}{\sqrt{2\pi}\,\sigma} \sigma^2 \int_0^{+\infty} e^{-\frac{x^2}{2\sigma^2}} dx = \frac{2\sigma}{\sqrt{2\pi}} \int_0^{+\infty} e^{-\frac{x^2}{2\sigma^2}} dx = \left\{\begin{array}{l} u = \frac{x}{\sqrt{2}\sigma} \\ du = \frac{1}{\sqrt{2}\sigma} dx \\ \sqrt{2}\sigma \, du = dx \end{array}\right\} =$$

$$= \frac{2\sigma}{\sqrt{2\pi}} \int_0^{+\infty} \sqrt{2}\sigma e^{-u^2} du = \frac{2\sigma^2}{\sqrt{\pi}} \int_0^{+\infty} e^{-u^2} du = \underbrace{\frac{2}{\sqrt{\pi}} \cdot \int_0^{+\infty} e^{-u^2} du}_{\operatorname{erf}(+\infty) = 1} \cdot \sigma^2 = \sigma^2$$

Таким образом:

$$\mathbb{E}(|X_1|^2) = \sigma^2$$
$$\mathbb{E}(|X_1|)^2 = \left(\frac{2\sigma}{\sqrt{2\pi}}\right)^2 = \frac{4\sigma^2}{2\pi} = \frac{2\sigma^2}{\pi}$$

$$\operatorname{Var}(\hat{\sigma}_{кор}) = \frac{\pi}{2n} \operatorname{Var}(|X_1|) = \frac{\pi}{2n} \cdot \left(\sigma^2 - \frac{2\sigma^2}{\pi}\right) = \frac{\sigma^2(\pi - 2)}{2n}$$

Ответ: $\operatorname{Var}(\hat{\sigma}_{кор}) = \frac{\sigma^2(\pi-2)}{2n}$

④ состоятельность:

$\hat{\sigma} \xrightarrow{p} \sigma$, т.е $\operatorname{plim}_{n\to\infty} \hat{\sigma} = \sigma$

$$\operatorname{plim}_{n\to\infty} \hat{\sigma}_{\text{кор}} = \operatorname{plim}_{n\to\infty} \left(\frac{\sqrt{2\pi}}{2}\hat{\sigma}\right) = \operatorname{plim}_{n\to\infty} \left|\frac{\sqrt{2\pi}}{2} \cdot \underbrace{\frac{1}{n}\cdot\sum_{i=1}^{n}|x_i|}_{\xrightarrow{p} \mathbb{E}(|x_i|)}\right| = \operatorname{plim}_{n\to\infty} \left|\frac{\sqrt{2\pi}}{2}\cdot\frac{2\sigma}{\sqrt{2\pi}}\right| =$$

$= \sigma$ – оценка состоятельная

⑤ $MSE(\hat{\sigma}_{\text{кор}}) = MSE\left(\frac{\sqrt{2\pi}}{2}\hat{\sigma}\right) = Var\left(\frac{\sqrt{2\pi}}{2}\hat{\sigma}\right) + bias^2\left(\frac{\sqrt{2\pi}}{2}\hat{\sigma}\right) =$

$= \frac{\sigma^2(\pi-2)}{2n} + \left[\mathbb{E}\left(\frac{\sqrt{2\pi}}{2}\hat{\sigma}\right) - \sigma\right]^2 = \frac{\sigma^2(\pi-2)}{2n} + \left[\frac{\sqrt{2\pi}}{2}\mathbb{E}(\hat{\sigma}) - \sigma\right]^2 =$

$= \frac{\sigma^2(\pi-2)}{2n} + \left[\frac{\sqrt{2\pi}}{2}\cdot\frac{2\sigma}{\sqrt{2\pi}} - \sigma\right]^2 = \frac{\sigma^2(\pi-2)}{2n}$

Задача N2 Мы знаем, что для выборки

$X_1, \dots, X_n \sim iid\, N(\mu, \sigma^2)$ хорошими оценками параметров будут

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

У этих оценок есть ещё одна полезная особенность. Случайные величины $\hat{\mu}$ и $\hat{\sigma}^2$ независимы. Докажите это.

Hint если ковариация между двумя случайными величинами нулевая, они будут независимы только, если эти две случайные величины нормально распределены. Воспользуйтесь этим. Для начала найдите $cov(x_1, x_1 - \bar{x})$

Решение:

I этап Найдем $cov(x_1, x_1 - \bar{x})$

$cov(x_1, x_1 - \bar{x}) = cov(x_1, x_1) - cov(x_1, \bar{x}) = Var(x_1) -$

$cov(X_1, \frac{1}{n}(X_1+...+X_n)) = Var(X_1) - \frac{1}{n}cov(X_1,X_1) - \frac{1}{n}cov(X_1,X_2) -$
$- ... - \frac{1}{n}cov(X_1,X_n) = Var(X_1) - \frac{1}{n}Var(X_1) - \frac{1}{n}\cdot(n-1)cov(X_1,X_j) =$
$= \left\{ \begin{array}{l} \text{т.к } X_1 \text{ и } X_j \\ \text{независ, то} \\ cov(X_1,X_j)=0 \end{array} \right\} = Var(X_1) - \frac{1}{n}Var(X_1) = \frac{n-1}{n}Var(X_1)$

Найдем $cov(X_1; X_2 - \bar{X})$

$cov(X_1, X_2 - \bar{X}) = cov(X_1, X_2) - cov(X_1, \bar{X}) = cov(X_1, X_2) -$
$- cov(X_1, \frac{1}{n}(X_1+X_2+...+X_n)) = cov(X_1,X_2) - \frac{1}{n}cov(X_1,X_1) -$
$- \frac{1}{n}cov(X_1,X_2) - ... - \frac{1}{n}cov(X_1,X_n) = (1-\frac{1}{n})cov(X_1,X_2) -$
$- \frac{1}{n}Var(X_1) - \frac{1}{n}\cdot(n-2)\cdot cov(X_1,X_j) = \left\{ \begin{array}{l} 1.\ \text{т.к } X_1 \text{ и } X_2 \text{ iid} \\ \text{то } cov(X_1,X_2)=0 \\ 2.\ \text{т.к } X_1 \text{ и } X_j \text{ iid} \\ \text{то } cov(X_1,X_j)=0 \end{array} \right\}$
$= -\frac{1}{n}Var(X_1)$

II этап посчитаем следующее:

ⓐ $cov(\frac{1}{n}\sum_{i=1}^{n}X_i, \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})) = \frac{1}{n(n-1)}cov(\sum_{i=1}^{n}X_i; \sum_{i=1}^{n}(X_i - \bar{X})) =$
$= \frac{1}{n(n-1)}cov(X_1+...+X_n; (X_1-\bar{X})+...+(X_n-\bar{X})) =$
$= \frac{1}{n(n-1)}[\underbrace{cov(X_1, X_1-\bar{X})+...+cov(X_n; X_n-\bar{X})}_{n} +$
$+ \underbrace{cov(X_1, X_2-\bar{X})+...+cov(X_1, X_n-\bar{X})}_{n-1} + \underbrace{...}_{n \text{ таких}}] =$
$= \frac{1}{n(n-1)}[n\,cov(X_1, X_1-\bar{X}) + n(n-1)cov(X_1, X_2-\bar{X})] =$
$= \frac{1}{n(n-1)}[n\frac{n-1}{n}Var(X_1) + n(n-1)(-\frac{1}{n}Var(X_1))] =$
$= \frac{1}{n(n-1)}\cdot[(n-1)Var(X_1) - (n-1)Var(X_1)] = 0 \Rightarrow$
$\frac{1}{n}\sum_{i=1}^{n}X_i$ и $\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})$ — независимы

(б) $\operatorname{cov}(\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\bar{X}}, \frac{1}{n-1}(X_1-\bar{X})) = \frac{1}{n(n-1)}\operatorname{cov}(X_1+\ldots+X_n, X_1-\bar{X}) =$

$= \frac{1}{n(n-1)}[\operatorname{cov}(X_1, X_1-\bar{X}) + \underbrace{\operatorname{cov}(X_2, X_1-\bar{X}) + \ldots + \operatorname{cov}(X_n, X_1-\bar{X})}_{n-1}] =$

$= \frac{1}{n(n-1)}[\frac{n-1}{n}\operatorname{Var}(X_1) + (n-1)\cdot(-\frac{1}{n}\operatorname{Var}(X_i))] =$

т.к все случ. вел i.i.d $\operatorname{Var}(X_i) = \operatorname{Var}(X_1)$

$= \frac{1}{n(n-1)}[\frac{n-1}{n}\operatorname{Var}(X_1) - \frac{n-1}{n}\operatorname{Var}(X_1)] = 0$

$\Rightarrow$ $\bar{X}$ и $\frac{1}{n-1}(X_1-\bar{X})$ тоже независимы

(в) по аналогии со всеми другими X

IV этап

таким образом $\operatorname{cov}(\bar{X}; (X_i-\bar{X})) = 0$

$\operatorname{cov}(\bar{X}, \sum_{i=1}^{n}(X_i-\bar{X})) = \sum_{i=1}^{n}\operatorname{cov}(\bar{X}; (X_i-\bar{X})) = 0$ $\Rightarrow$ независимы

Согласно существующей теореме о том что функции от независимых переменных тоже независимые

Имеем для нашего случая

$\sigma^2 = f(X_i-\bar{X})$, f - возводит в квадрат

$\operatorname{cov}(\bar{X}, f(X_i-\bar{X})) = \operatorname{cov}(\bar{X}, (X_i-\bar{X})^2) = 0$

и $\operatorname{cov}(\bar{X}; \sum_{i=1}^{n} f(X_i-\bar{X})) = \operatorname{cov}(\bar{X}, \sum_{i=1}^{n}(X_i-\bar{X})^2) =$

$= \sum_{i=1}^{n}\operatorname{cov}(\bar{X}; (X_i-\bar{X})^2) = 0$ — независимы

Тогда $\hat{M}$ и $\hat{\sigma}^2$ тоже независимы что и требовалось доказать.

(Задача №3) Аналитеса Аня наряд поймала 400 покемонов-девочек и 100 покемонов-мальчиков. Среди девочек 250 оказались ядовитыми, среди мальчиков 60.

1. Методом моментов найдите точечную оценку для доли ядовитых покемонов среди мальчиков
2. Постройте 95% доверительный интервал для ядовитых покемонов-мальчиков. Какая у него получилась длина?
3. Постройте 95% доверительный интервал для ядовитых покемонов девочек. Какая у него получилась длина?
4. Постройте 95% доверительный интервал для разницы долей ядовитых покемонов среди девочек и мальчиков. Какой длины он оказался? Попадает ли ноль в этом интервал? Что это означает?
5. Нужно ли предположение о нормальности X_i и Y_i для решения предыдущих пунктов? А какие предположения нужны? Выпишите их.

(Решение)

Пусть $X_1, \dots, X_{400} \sim iid\ Bern(p_д)$ — где $X_i = 1$ — ядовитый покемон девочка, $X_i = 0$ — неядовитая

Пусть $Y_1, \dots, Y_{100} \sim iid\ Bern(p_м)$ — где $Y_i = 1$ — ядовитый покемон мальчик, $Y_i = 0$ — неядовитый.

$\{ \mathbb{E}(X_i) = p \quad Var(X_i) = p(1-p) \}$

1) $\bar{Y}_n \approx \mathbb{E}(Y)$, $\bar{Y}_n \approx p_м$

$\bar{Y}_n = \hat{p}_м^{MM} = \frac{60}{100} = 0,6$ — с помощью метода моментов

$\hat{p}_M^{MM} = 0{,}6 = \frac{1}{n}\sum_{i=1}^{n} y_i$

2) $\alpha = 0{,}05$ 95% Д.И для p_M

$\hat{p}_M = \overline{Y}_n \overset{asy}{\sim} N\left(p_M; \frac{\hat{p}_M^{MM}(1-\hat{p}_M^{MM})}{n}\right) = N\left(p_M; \frac{0{,}6\cdot(1-0{,}6)}{100}\right) \Rightarrow$

$$\frac{\hat{p}_M^{MM} - p_M}{\sqrt{\frac{\hat{p}_M^{MM}(1-\hat{p}_M^{MM})}{n}}} \overset{asy}{\sim} N(0,1)$$

Доверительный интервал: $1-\frac{\alpha}{2} = 1 - \frac{0{,}05}{2}\ 0{,}975$

$$\hat{p}_M^{MM} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}_M^{MM}(1-\hat{p}_M^{MM})}{n}}$$

$$0{,}6 \pm 1{,}96 \cdot \sqrt{\frac{0{,}6\cdot(1-0{,}6)}{100}}$$

$$0{,}6 - 1{,}96\cdot\frac{\sqrt{0{,}24}}{10} \le p_M \le 0{,}6 + 1{,}96\cdot\frac{\sqrt{0{,}24}}{10}$$

$$\underline{0{,}504 \le p_M \le 0{,}696}$$

длина дов. интервала:

$$2\cdot Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}_M^{MM}(1-\hat{p}_M^{MM})}{n}} = 2\cdot 1{,}96\cdot\frac{\sqrt{0{,}24}}{10} = \underline{\underline{0{,}192}}$$

3) $\alpha = 0{,}05$ 95% Д.И. для p_g.

С помощью метода найдем точечную оценку для доли ядовитых покемонов среди девочек

$\overline{X}_n \approx E(X)$

$\overline{X}_n = \hat{p}_g^{MM} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{250}{400} = 0{,}625$

$\hat{p}_g = \overline{X}_n \overset{asy}{\sim} N\left(p_g; \frac{\hat{p}_g^{MM}(1-\hat{p}_g^{MM})}{n}\right) = N\left(p_g; \frac{0{,}625\cdot(1-0{,}625)}{400}\right)$

$$\frac{\hat{p}_g^{MM} - p_g}{\sqrt{\frac{\hat{p}_g^{MM}(1-\hat{p}_g^{MM})}{n}}} \overset{asy}{\sim} N(0,1)$$

Доверительный интервал $1-\frac{\alpha}{2} = 1 - \frac{0{,}05}{2} = 0{,}975$

$$\hat{p}_g^{MM} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}_g^{MM}(1-\hat{p}_g^{MM})}{n}}$$

$$0,625 \pm 1,96 \cdot \frac{\sqrt{0,625\cdot(1-0,625)}}{20}$$

$$0,625 - 1,96 \cdot \frac{\sqrt{0,625\cdot 0,375}}{20} \le p_g \le 0,625 + 1,96 \cdot \frac{\sqrt{0,625\cdot 0,375}}{20}$$

$$0,578 \le p_g \le 0,672$$

длина д.и:

$$2\cdot 1,96 \cdot \frac{\sqrt{0,625\cdot 0,375}}{20} = 0,094$$

4) 95% д.и. для $p_g - p_M$

$$\hat{p}_g - \hat{p}_M \overset{asy}{\sim} N\left(p_g - p_M;\ \frac{\hat{p}_g(1-\hat{p}_g)}{n_g} + \frac{\hat{p}_M(1-\hat{p}_M)}{n_M}\right)$$

д.и: $(\hat{p}_g - \hat{p}_M) \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}_g(1-\hat{p}_g)}{n_g} + \frac{\hat{p}_M(1-\hat{p}_M)}{n_M}}$

$$0,625 - 0,6 \pm 1,96 \cdot \sqrt{\frac{0,625\cdot 0,375}{400} + \frac{0,6\cdot 0,4}{100}}$$

$$-0,082 \le p_g - p_M \le 0,132$$

длина ди:

$$2\cdot 1,96 \cdot \sqrt{\frac{0,625\cdot 0,375}{400} + \frac{0,6\cdot 0,4}{100}} = 0,214$$

Ноль попадает в д.и, значит разница долей статистически не значима, нельзя утверждать что доля ядовитых мальчиков отличается от девочек

5) Предположения о нормальности X_i, Y_i не нужны. Мы предполагаем, что слч. вел. независимы и распределены по закону Бернулли

Задача №4 Аналитесса Аня пытается выяснить какая доля жителей Москвы любит кофе. Она хочет построить оценку так чтобы ширина 95% доверительного интервала для неё оказалась не больше 0.1

1. Предположим что Аня ничего не знает о доле. Сколько наблюдений Ане необходимо собрать для достижения такой точности?
2. Аналитик Богдан проговорился Ане, что в прошлом году уже опрашивал так людей. У него оценка доли получилась в районе 0.7. Аня верит Богдану как себе. Сколько наблюдений ей надо собрать для достижения той же точности с учетом этой информации?

Решение:
Пусть $X_i \sim iid\ Bern(p)$
где $X_i = 1$, если человек любит кофе
$X_i = 0$, если человек не любит кофе

95% Д.И для p:
Согласно ЗБЧ и ММ хорошей оценкой доли будет среднее: $\hat{p}_{MM} = \bar{X}_n$, $\mathbb{E}(X_i) \approx \bar{X}_n$

по ЦПТ $\bar{X}_n = \hat{p}_{MM} \overset{asy}{\sim} N\left(p, \frac{\hat{p}(1-\hat{p})}{n}\right)$

$\hat{p} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, $\alpha = 0{,}05$, тогда $z_{1-\frac{\alpha}{2}} = z_{0,975} = 1{,}96$

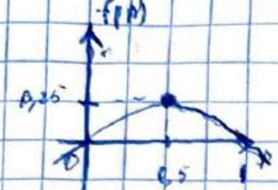
ширина - длина Д.И: $2 \cdot z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Согласно условию задачи: $2 \cdot z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0{,}1$

$\hat{p}$ - неизвестно, подберем максимальное значение которое $\hat{p}(1-\hat{p})$ может принять:

$f(p) = p(1-p) = p - p^2$ – парабола

$f'(p) = 1 - 2p = 0$, $p = \frac{1}{2}$ – вершина параболы



$f(\frac{1}{2}) = 0,5 - 0,25 = 0,25$

Имеем $2 \cdot Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{0,25}{n}} \le 0,1$

$2 \cdot 1,96 \cdot \sqrt{\frac{0,25}{n}} \le 0,1$

$n \geqslant 384,16$ – значит $\geqslant 385$ наблюдений

2) $\hat{p} = 0,7$

$2 \cdot Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}} \le 0,1$

$2 \cdot 1,96 \cdot \sqrt{\frac{0,7 \cdot 0,3}{n}} \le 0,1$

$n \geqslant 322,69$ – значит необходимо 323 наблюдения

Задача №5 Для идеальной картинки режиссёру рекламы грузовиков Volvo требуется 60 дублей. В каждом Жан Клод ван Дамм садится на шпагат. Пусть $U_1, U_2, \dots, U_{60}$ – независимо распределенные случайные величины, сколько сантиметров актёру не хватает до идеальной паралели, $U_i \sim U(0,1)$ и $X = U_1 + \dots + U_{60}$.

1. К какому распределению очень близко распределение случайной величины X? Укажите его параметры.
2. Найдите, чему приблизительно равна $(X > 20)$

Решение

В задаче не сказано про одинаково распределенные случ. вел, но подозреваю, что это учитывается. Т.е

$U_1, U_2, \dots U_{60} \sim iid\ U(0,1)$

(1) Согласно ЦПТ распределение случ. вел X будет очень близким к нормальному. Тогда его параметры будут такими $\mathbb{E}(U_i) = \frac{1}{2}$ $Var(U_i) = \frac{1}{12}$

$X \overset{asy}{\sim} N(60 \cdot \frac{1}{2}, 60 \cdot \frac{1}{12})$

$X \overset{asy}{\sim} N(30; 5) \Rightarrow Z = \frac{X-30}{\sqrt{5}} \overset{asy}{\sim} N(0;1)$

(2) $\mathbb{P}(X > 20) = 1 - \mathbb{P}(X \leq 20) = 1 - \mathbb{P}(\frac{X-30}{\sqrt{5}} \leq \frac{20-30}{\sqrt{5}}) =$
$= 1 - \mathbb{P}(Z \leq -\frac{10}{\sqrt{5}}) = 1 - (1 - \mathbb{P}(Z \leq \frac{10}{\sqrt{5}})) = P(Z \leq \frac{10}{\sqrt{5}}) =$
$= F_Z(\frac{10}{\sqrt{5}}) \approx 0{,}99$

Ответ: 1) Близко к нормальному согласно ЦПТ с параметрами $\mu = 30$ $\sigma^2 = 5$
2. $\mathbb{P}(X > 20) \approx 0{,}99$

Задача 6 Пусть $X_1, X_2, \dots$ независимо распределенные случайные величины, число дублей в каждом съемочном дне, которые не идеальны по вине водителя левого грузовика Volvo, $\mathbb{E}(X_i)=2$, $Y_1, Y_2, \dots$ независимо распределенные случайные величины, сколько дублей испортил водитель второго грузовика за каждый съемочный день $\mathbb{E}(Y_i)=4$. К какому числу сходится случайная величина $\frac{X_1+X_2+\dots+X_n}{Y_1+Y_2+\dots+Y_n}$ если режиссер перфекционист не ограничен во времени и деньгах т.е при $n \to \infty$

Решение

Как и 5 задаче про одинаковое распределение ничего не сказано, думаю оно подразумевается

Т.е $X_1, X_2, \dots \sim iid$ $Y_1, Y_2, \dots \sim iid$

Согласно ЗБЧ $\frac{X_1+X_2+\dots+X_n}{n} \xrightarrow{P} \mathbb{E}(X_i)$

$\frac{Y_1+Y_2+\dots+Y_n}{n} \xrightarrow{P} \mathbb{E}(Y_i)$

$\operatorname{plim}_{n\to\infty} \overline{X_n} = \mathbb{E}(X_i) = 2$ $\operatorname{plim} \overline{Y_n} = \mathbb{E}(Y_i) = 4$

Значит $\lim_{n\to\infty} \frac{X_1+X_2+\dots+X_n}{Y_1+Y_2+\dots+Y_n} = \lim_{n\to\infty} \frac{\frac{X_1+X_2+\dots+X_n}{n}}{\frac{Y_1+Y_2+\dots+Y_n}{n}} = \lim_{n\to\infty} \frac{\overline{X_n}}{\overline{Y_n}} =$

$= \frac{\operatorname{plim}_{n\to\infty} \overline{X_n}}{\operatorname{plim}_{n\to\infty} \overline{Y_n}} = \frac{2}{4} = \frac{1}{2}$ Ответ: $\frac{1}{2}$

Задача 7 В рекламе задействовано два грузовика. Расход топлива за i-ый день (в л/100км) у левого — случайного величина X_i, у правого — случайная величина Y_i, X_i и Y_i независимы $\forall i$. Известно, что $\mathbb{E}(X_i)=20$. $Var(X_i)=1$, $\mathbb{E}(Y_i)=18$, $Var(Y_i)=4$

За один съемочный день грузовики проезжают по 100 км. Поскольку сначала надо отрепетировать прок без Жан Клода Ван Дамма, а потом с ним грузовики будут нужны в 40 съемочных днях. Какова вероятность того, что за все время репетиций и съемок понадобится более 1550 литров топлива.

(Решение)

Обозначим $S_X = X_1 + \ldots + X_{40}$, $S_Y = Y_1 + \ldots + Y_{40}$

Согласно ЦПТ $S_X \overset{asy}{\sim} N(nE(X_i); n\,Var(X_i))$

$$S_X \overset{asy}{\sim} N(40 \cdot 20;\ 40 \cdot 1)$$

$$S_X \overset{asy}{\sim} N(800;\ 40)$$

Аналогично для S_Y: $S_Y \overset{asy}{\sim} (720; 160)$

Тогда $S_X + S_Y \overset{asy}{\sim} N(1520; 200)$

$$Z = \frac{(S_X + S_Y) - 1520}{\sqrt{200}} \overset{asy}{\sim} N(0,1)$$

$$\mathbb{P}(X_1 + \ldots + X_{40} + Y_1 + \ldots + Y_{40} \geq 1550) = \mathbb{P}(S_X + S_Y \geq 1550) =$$

$$= 1 - \mathbb{P}(S_X + S_Y < 1550) = 1 - \mathbb{P}\left(\frac{(S_X+S_Y) - 1520}{\sqrt{200}} < \frac{1550 - 1520}{\sqrt{200}}\right) =$$

$$= 1 - \mathbb{P}\left(Z < \frac{30}{10\sqrt{2}}\right) = 1 - \mathbb{P}\left(Z < \frac{3}{\sqrt{2}}\right) \approx 1 - \mathbb{P}(Z < 2{,}121) =$$

$$= 1 - F_Z(2{,}121) \approx 1 - 0{,}983 = 0{,}017$$

Ответ: $\mathbb{P}(X_1 + \ldots + X_{40} + Y_1 + \ldots + Y_{40} \geq 1550) \approx 0{,}017$

(Задача 8) После выхода рекламы компания Volvo провела опрос 10000 дальнобойщиков. У каждого спросили, считает ли он модель Volvo XC 40 более стильной, чем модель Volvo V90. Дальнобойщики отвечали „да" или „нет" равновероятно. Оцените вероятность того, что число положительных ответов отличалось от 5000 меньше, чем на 100

Решение

$X_1, \dots, X_{10000} \sim iid\ Bern(\frac{1}{2})$, где

X_i	0 (нет)	1 (да)
$\mathbb{P}$	$\frac{1}{2}$	$\frac{1}{2}$

$\mathbb{E}(X_i) = \frac{1}{2}$ $\quad Var(X_i) = \frac{1}{2} \cdot (1 - \frac{1}{2}) = \frac{1}{4}$

Нужно найти $\mathbb{P}(|X_1 + \dots + X_{10000} - 5000| < 100)$

Обозначим $S_X = X_1 + \dots + X_{10000}$

Согласно ЦПТ $S_X \overset{asy}{\sim} N(n\mathbb{E}(X_i); n\,Var(X_i))$

$S_X \overset{asy}{\sim} N(10000 \cdot \frac{1}{2}; 10000 \cdot \frac{1}{4})$

$S_X \overset{asy}{\sim} N(5000; 2500)$

$Z = \frac{S_X - 5000}{\sqrt{2500}} \overset{asy}{\sim} N(0;1)$

$\mathbb{P}(|(X_1 + \dots + X_{10000}) - 5000| < 100) = \mathbb{P}(|S_X - 5000| < 100) = \mathbb{P}(4900 < S_X < 5100) =$

$= \mathbb{P}\left(\frac{4900 - 5000}{\sqrt{2500}} < \frac{S_X - 5000}{\sqrt{2500}} < \frac{5100 - 5000}{\sqrt{2500}}\right) = \mathbb{P}(-2 < Z < 2) =$

$= F_Z(2) - F_Z(-2) = F_Z - (1 - F_Z(2)) = 2F_Z(2) - 1 \approx$

$\approx 2 \cdot 0{,}977 - 1 = 0{,}954$

Ответ: $\mathbb{P}(|(X_1 + \dots + X_{10000}) - 5000| < 100) \approx 0{,}954$