Вопросы по предыдущему

19.116. Пусть $x_1, \ldots, x_n$ — выборка из нормально распределенной генеральной совокупности $N(m, \sigma)$. Найти информацию Фишера $I_n(\sigma^2)$.

19.117 (продолжение). В условиях предыдущей задачи при известном математическом ожидании m оценивается дисперсия σ^2. Показать, что статистика

$$s_0^2 = \frac{1}{n}\sum (x_i - m)^2$$

является эффективной оценкой σ^2.

19.118*. Пусть $x_1, \ldots, x_n$ — выборка из генеральной совокупности, имеющей равномерное распределение $R(0, 1)$. Показать, что статистика $\tilde{m} = \frac{1}{2}(x^{(1)} + x^{(n)})$ является более эффективной оценкой математического ожидания, чем выборочное среднее.

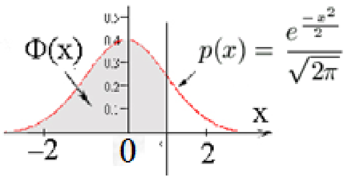
117

$$Ds_0^2 = D\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 = \frac{1}{n^2}nD(x_i - m)^2 = \frac{1}{n}\left(M(x_i - m)^4 - \left[M(x_i - m)^2\right]^2\right) =$$

$$= \frac{1}{n}\left(3\sigma^4 - (\sigma^2)^2\right) = \frac{2\sigma^4}{n} \qquad \xi_0 = (x_i - m) \sim N(0, \sigma^2) \quad M\xi_0^4 = 3\sigma^4$$

Инфо. Фишера: $\quad I_n(\sigma^2) = \dfrac{n}{2\sigma^4}$

О моментах стандартного нормального закона $N(0,1)$



$$\mu_n = \int_{-\infty}^{\infty} x^n \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx$$

$$\mu_{2k+1} = 0$$

$N(0,1), \quad \mu_{n+2} = (n+1)\mu_n$, $\mu_2 = 1,\ \mu_4 = (2+1)\mu_2 = 3,\ \ \mu_6 = (4+1)\mu_4 = 5\cdot 3 = 15, \ldots$

$N(0,\sigma^2),\ \times\sigma^{n+2}, \qquad \mu_2 = \sigma^2, \qquad \mu_4 = 3\sigma^4, \quad \mu_6 = 15\sigma^6$

118

$$x_1, x_2, \ldots, x_n \sim R[0,1] \qquad \hat{m} = \frac{1}{n}\sum_{i=1}^{n}\xi_i, \quad M\hat{m} = \frac{1}{2}, \quad D\frac{1}{n}\sum_{i=1}^{n}\xi_i = \frac{1}{12n},$$

$$\tilde{m} = \frac{1}{2}\left(x_{(1)} + x_{(n)}\right) \qquad M\tilde{m} = M\frac{1}{2}\left(x_{(1)} + x_{(n)}\right) = \frac{1}{2}, \quad D\frac{1}{2}\left(x_{(1)} + x_{(n)}\right) = ?$$

Решение.

1. Определить з.р. $x_{(n)}$, затем M $x_{(n)}$ и D $x_{(n)}$

Ф-ция распределения:

$$F_{x(n)}(x) = P\left(\max_{1\le i\le n}\xi_i < x\right) = P\left(\xi_1 < x, \xi_2 < x, ..., \xi_n < x\right) = F_\xi^n(x) = \begin{cases} 0^n, x < 0 \\ x^n, \quad 0 \le x \le 1, \\ 1^n, x > 1 \end{cases}$$

Плотность:

$$p_{x(n)}(x) = \begin{cases} nx^{n-1}, x \in [0,1], \\ 0,, x \notin [0,1]. \end{cases}$$

ясно, что $p_{x(1)}(x) = \begin{cases} n\left(1-x\right)^{n-1}, x \in [0,1], \\ 0, x \notin [0,1]. \end{cases}$?

$$1 - F_{x(1)}(x) = P\left(\min_{1\le i\le n}\xi_i \ge x\right) = P\left(\xi_1 \ge x, \xi_2 \ge x, ..., \xi_n \ge x\right) = \left(1 - F_\xi(x)\right)^n = \begin{cases} 1^n, x < 0 \\ \left(1-x\right)^n, 0 \le x \le 1, \\ 0^n, x > 1 \end{cases}$$

Дифференцируем слева и справа, и получаем плотность $p_{x(1)}(x)$

Мат. ожидание:

$$Mx_{(n)} = \int x p_{x(n)}(x)dx = \int_0^1 x\left(nx^{n-1}\right)dx = \frac{n}{n+1}$$

$$Mx^2_{(n)} = \int x^2 p_{x(n)}(x)dx = \int_0^1 x^2\left(nx^{n-1}\right)dx = \frac{n}{n+2}$$

$$Dx_{(n)} = Mx^2_{(n)} - \left(Mx_{(n)}\right)^2 = \frac{n\left(n+1\right)^2 - n^2\left(n+2\right)}{\left(n+2\right)\left(n+1\right)^2} = \frac{n}{\left(n+2\right)\left(n+1\right)^2} \sim \frac{1}{\left(n+1\right)^2}.$$

2. M.o. M $x_{(1)}$ = 1- M $x_{(n)}$, D $x_{(1)}$ = D $x_{(n)}$

$$M\left(\frac{1}{2}\left(x_{(1)} + x_{(n)}\right)\right) = \frac{1}{2}\left(1 - Mx_{(m)} + Mx_{(n)}\right) = \frac{1}{2}$$

3. Дисперсия $D\left(\frac{1}{2}\left(x_{(1)} + x_{(n)}\right)\right)$ через дисперсии и коэф. корр. и учесть $\frac{1+r}{2} \le 1$

$$D\left(\frac{1}{2}\left(x_{(1)} + x_{(n)}\right)\right) = \frac{1}{4}\left(Dx_{(1)} + Dx_{(n)} + 2r\sqrt{Dx_{(1)}Dx_{(n)}}\right) =$$

$$= \frac{1}{4}\left(Dx_{(n)} + Dx_{(n)} + 2rDx_{(n)}\right) = \frac{Dx_{(n)}}{4}\left(2+2r\right) = Dx_{(n)}\frac{1+r}{2} \le Dx_{(n)} = \frac{n}{\left(n+2\right)\left(n+1\right)^2} \sim \frac{1}{n^2}$$

$$D\frac{1}{n}\sum_{i=1}^n \xi_i = \frac{1}{12n}, \qquad n \ge 3$$

Тема 3. Достаточные статистики

вопрос о сжатии информации

3.1. Определение

Пусть имеется совокупность наблюдений $x \equiv (x_1, x_2, ..., x_n)$ случайного характера, т.е. одна из возможных реализаций многомерной СВ $\xi \equiv (\xi_1, \xi_2, ..., \xi_n)$ и $p_\xi(x;a)$ — закон распределения ξ (плотность или вероятность) зависит от **неизвестног о**параметра $a \equiv (a_1, a_2, ..., a_l)$.

Вводим некоторую статистику $\tau = T(\xi)$ — функцию (вообще говоря, многомерную) исходных наблюдений; $p_\tau(t;a)$ — соответствующийзакон распределения.

Имеем пару (ξ, τ) случайных величин.

Пусть мы знаем значение статистики $\tau = T(\xi)$. Можно ли для выводов относительно неизвестного *а* оставить только СВ $\tau = T(\xi)$, а ξ выбросить? **Надо посмотреть условное распределение** ξ **при условии известного значения** τ **.**

А) Пусть случайные величины дискретны

$$p_\xi \ \ x \,|\, \tau = T(\xi) = T(x);\, a \ \ = P \ \ \xi = x \,|\, \tau = T(\xi) = T(x);\, a \ \ =$$

$$= \frac{P\{\xi = x, \tau = T(\xi) = T(x);\, a\}}{P\{\tau = t = T(x);\, a\}} = \frac{P\{\xi = x;\, a\}}{P\{\tau = t = T(x);\, a\}} = \frac{p_\xi(x;\, a)}{p_\tau \ \ t = T(x);\, a} \quad \textbf{.(3.1)}$$

Здесь учтено, что из события

$A = \{\xi = x\}$ следует событие $B = \ \tau = T(\xi) = T(x)$:

$$P \ \ AB \ \ = P \ \ AB \ \ .$$

Пусть условное распределение **не зависит** от параметра a,т.е.

$$p_\xi \ \ x \,|\, \tau = T(x);\, a \ \ = \mathrm{const}(a).$$

Нужны ли нам тогда наблюдения ξ? Ведь если известно значение $\tau = T(x)$, то закон распределения ξ (условное распределение) от параметра не зависит, т.е. новой информации о параметре *а* в них нет. Следовательно, если $\tau = T(x)$ **известно**, ξ дополнительной информации не несёт. Можно ξ выбросить, оставив $\tau = T(x)$.

Определение. Статистика $\tau = T(x)$ называется **достаточной** для a, если **условное** распределения ξ, при условии **известного** τ , **не зависит** от a.

Б) Для непрерывных СВ формула (3.1) тоже верна

$$p_\xi \ \ x \,|\, \tau = T(\xi) = T(x);\, a \ \ = \frac{p_\xi(x;\, a)}{p_\tau \ \ t = T(x);\, a}$$

Итак, любые статистические выводы о параметре a можно делать на основании достаточной статистики $\tau = T(\xi)$.

3.2.Критерий факторизации

Теорема 3.1 (критерий факторизации). Статистика $T(\xi)$ является **достаточной** для параметра a **тогда и только тогда, когда** справедливо представление:

$$p_\xi(x;a) = g\left(T(x), a\right) h(x).$$

Существенным в этой записи является то, что ***множитель, зависящий от параметра** a**, от** x **зависит через функцию** $T(x)$.

Пример 3.1. Проведено **две** серии **независимых** испытаний события A.

В каждой серии вероятность «успеха» **неизвестна** и **равна** $P(A) = p \equiv a$.

Обозначим через ξ_1 и ξ_2 СВ, равные, соответственно, числу «успехов» в первой и во второй сериях. Очевидно, эти СВ распределены по биномиальному закону:

$$\xi_1 \sim Bi(n_1, p \equiv a) \text{ и } \xi_2 \sim Bi(n_2, p \equiv a).$$

В первой серии из n_1 испытаний имеем $\xi_1 = x_1$ успехов, а во второй серии из n_2 испытаний имеем $\xi_2 = x_2$ успехов.

Доказать **двумя** способами, что СВ, равная сумме $\tau = \xi_1 + \xi_2$, является **достаточной** статистикой для параметра $p \equiv a$.

1-й способ (с помощью условного распределения (3.1)).

Поскольку случайные величины $\xi_1 \sim Bi(n_1, p \equiv a)$ и $\xi_2 \sim Bi(n_2, p \equiv a)$ **независимы**, их сумма

$$\tau = \xi_1 + \xi_2$$

также имеет **биномиальное** распределение: $\tau = \xi_1 + \xi_2 \sim Bi(n_1 + n_2, p \equiv a)$. Выпишем закон её распределения:

$$p_\tau(t;a) = P\left(\tau = t; a\right) = C_{n_1+n_2}^t a^t (1-a)^{n_1+n_2-t}, \text{ где } t = 0, 1, 2, ..., (n_1 + n_2).$$

Условное распределение СВ $\xi = (\xi_1, \xi_2)$ при условии, что

$$\tau = \xi_1 + \xi_2 = x_1 + x_2 = t,$$

согласно формуле (3.1), имеет вид: $\xi = (\xi_1, \xi_2)$

$$p_\xi\left(x_1, x_2 \,|\, \tau = x_1 + x_2 = t; a\right) = \frac{P\{\xi_1 = x_1, \xi_2 = x_2; a\}}{P\{\tau = \xi_1 + \xi_2 = x_1 + x_2 = t; a\}} =$$

$$= \left.\frac{C_{n_1}^{x_1} a^{x_1} (1-a)^{n_1-x_1} \cdot C_{n_2}^{x_2} a^{x_2} (1-a)^{n_2-x_2}}{C_{n_1+n_2}^t a^t (1-a)^{n_1+n_2-t}}\right|_{t=x_1+x_2} = \frac{C_{n_1}^{x_1} \cdot C_{n_2}^{x_2}}{C_{n_1+n_2}^{x_1+x_2}} = \frac{C_{n_1}^{x_1} \cdot C_{N-n_1}^{t-x_1}}{C_N^t}. \quad (3.2)$$

Это условное распределение не зависит от параметра a, поэтому статистика $\tau = \xi_1 + \xi_2$ является **достаточной**. Получили гипергеометрическое распределение: в ящике находится $N = n_1 + n_2$ шаров, из которых n_1 — белые и n_2 — чёрные. Вынимаем $t = x_1 + x_2$ шаров; какова вероятность того, что среди вынутых шаров будет x_1 белых и $x_2 = t - x_1$ чёрных? Ф-ла выделена желтым.

2-й способ (**с помощью критерия факторизации**).

Записываем закон распределения для $\xi = (\xi_1, \xi_2)$:

$$p_\xi(x;a) = p_{\xi_1}(x_1;a)\cdot p_{\xi_2}(x_2;a) = C_{n_1}^{x_1} a^{x_1}(1-a)^{n_1-x_1}\cdot C_{n_2}^{x_2} a^{x_2}(1-a)^{n_2-x_2} =$$

$$= \left(\prod_{i=1}^{m=2} C_{n_i}^{x_i}\right)\cdot a^{\sum x_i}(1-a)^{\sum n_i - \sum x_i} = \left(\prod_{i=1}^{m=2} C_{n_i}^{x_i}\right)\cdot \boxed{\left(\frac{a}{1-a}\right)^{\sum x_i}(1-a)^{\sum n_i}}$$

и выделяем множитель, зависящий от параметра *а* (выделено желтым).

Он зависит от $x = (x_1, x_2)$ через $\sum x_i$, следовательно СВ $\tau = \sum_i \xi_i$ — **достаточная** статистика для параметра a. Заметим, что

$$M\tau = M(\xi_1 + \xi_2) = n_1 a + n_2 a = a(n_1 + n_2),$$

т.е., с точностью до множителя, $M\tau$ есть параметр. Если на множитель разделить, получим оценку для параметра:

$$\hat{a} = \frac{\tau}{n_1 + n_2} = \frac{\xi_1 + \xi_2}{n_1 + n_2}.$$

Пример 3.2. Пусть $\xi \equiv (\xi_1, \xi_2, ..., \xi_n)$ — выборка из показательного распределения, $\xi_i \sim E(a), i = \overline{1,n}$. Закон распределения для одного $(i\text{-го})$ наблюдения имеет вид:

$$p_{\xi_i}(x_i;a) = ae^{-ax_i}, \text{если } x_i \geq 0,$$

поэтому для выборки $\xi \equiv (\xi_1, \xi_2, ..., \xi_n)$ получаем:

$$p_\xi(x;a) = p_\xi(x_1, x_2, ..., x_n;a) = \prod_i p_{\xi_i}(x_i;a) = a^n e^{-a\sum x_i}, \text{если} \min x_i \geq 0.$$

По критерию факторизации делаем вывод, что СВ $\tau = \sum_i \xi_i$ — **достаточная** статистика для параметра a. Отметим, что можно использовать запись $\tau = \sum_i x_i$.

Замечание. Выпишем первые два момента достаточной статистики τ из примера 3.2:

$$M\tau = \frac{n}{a},\ D\tau = \frac{n}{a^2}.$$

$\frac{\tau}{n}$ является несмещенной оценкой для $\frac{1}{a}$. Проверим $M\frac{\tau}{n} = \frac{1}{n}\frac{n}{a} = \frac{1}{a}$,

Дисперсия оценки: $D\frac{\tau}{n} = \frac{1}{n^2}\frac{n}{a^2} = \frac{1}{na^2}$,

Можно по $\tau = \sum_i x_i$. строить доверительный интервал для параметра.

Пример 3.3. Пусть $\xi \equiv (\xi_1, \xi_2, ..., \xi_n)$ — выборка, $\xi_i \sim R[0,a]$. Плотность для одного наблюдения задаётся формулой: ***(пример был в лк, МП-оценка)***

$$p_{\xi_i}(x_i;a) = \begin{cases} 1/a, & x_i \in [0,a], \\ 0, & x_i \notin [0,a]. \end{cases}$$

Поэтому плотность для всех наблюдений ξ имеет вид:

$$p_\xi(x;a) = p_\xi(x_1, x_2, ..., x_n;a) = \prod_{i=1}^{n} p_{\xi_i}(x_i;a) =$$

$$= \begin{cases} (1/a)^n, \text{если } x_1 \in [0,a], ..., x_n \in [0,a], \\ 0 \quad \text{в остальных случаях}, \end{cases} = \begin{cases} (1/a)^n, \text{если } \max x_i \le a \text{ и } \min x_i \ge 0, \\ 0 \quad \text{в остальных случаях}. \end{cases}$$

Введем функцию $pos(u) = \begin{cases} 1, & u \ge 0, \\ 0, & u < 0. \end{cases}$ Тогда

$$p_\xi(x;a) = (1/a)^n\, pos\left(a - \max_i x_i\right) \cdot pos\left(\min_i x_i\right),$$

где множитель $(1/a)^n pos\left(a - \max_i x_i\right)$, зависящий от параметра a, от $x \equiv (x_1, x_2, ..., x_n)$ зависит через $\max_i x_i$. По критерию факторизации статистика $\tau = \max_i x_i$ является **достаточной** для параметра a.

Найдём **математическое ожидание** и **дисперсию** статистики $\tau = \max_i \xi_i$.

Более общая задача:

Пусть СВ ξ_i, $i = \overline{1,n}$, имеют функцию распределения $F(y)$. Тогда для функции распределения статистики τ справедлива формула:

$$F_\tau(y) = P\left(\tau = \max_i \xi_i < y\right) = P\{\xi_1 < y, ..., \xi_n < y\} = P^n\{\xi_i < y\} = F^n(y).$$

Зная функцию распределения $F_\tau(y)$, найдём плотность распределения:

$$p_\tau(y) = F_\tau'(y) = nF^{n-1}(y)\,p_\xi(y).$$

В нашем случае $\xi_i \sim R[0,a]$, $i = \overline{1,n}$, поэтому ф.р.

$$F_\tau(y)=\begin{cases} 0, & \text{если } y<0,\\ (y/a)^n, & \text{если } y\in[0,a],\\ 1, & \text{если } y>a. \end{cases}$$ плотность:

$$p_\tau(y)=\begin{cases} n\left(\dfrac{y}{a}\right)^{n-1}\cdot\dfrac{1}{a}, & \text{если } y\in[0,a],\\ 0, & \text{если } y\notin[0,a]. \end{cases}$$

Следовательно,

$$M\tau=\int_0^a y p_\tau(y)dy=\int_0^a yn\left(\frac{y}{a}\right)^{n-1}\cdot\frac{1}{a}dy=an\int_0^1 z^n dz=a\frac{n}{n+1}.$$

$$M\tau^2=\int_0^a y^2 p_\tau(y)dy=\int_0^a y^2 n\left(\frac{y}{a}\right)^{n-1}\cdot\frac{1}{a}dy=na^2\int_0^1 z^{n+1}dz=a^2\frac{n}{n+2}.$$

$$D\tau=M\tau^2-(M\tau)^2=a^2\frac{n}{n+2}-\left(a^2\frac{n}{n+1}\right)^2=\frac{a^2 n}{(n+2)(n+1)^2}.$$

Из формулы для $M\tau$ следует, что τ с точностью до множителя является несмещённой оценкой для a:

$$\hat{a}=\frac{n+1}{n}\tau=\frac{n+1}{n}\max_i \xi_i.$$

Характеристики этой оценки:

$$M\hat{a}=a,\ D\hat{a}=D(c\tau)=c^2 D\tau=\left(\frac{n+1}{n}\right)^2\frac{a^2 n}{(n+2)(n+1)^2}=\frac{a^2}{n(n+2)}.$$

Это пример **сверхэффективной оценки**, поскольку дисперсия убывает быстрее, чем $1/n$.

Пример 3.4. Рассмотрим выборку с наблюдениями, распределенными по закону Коши. Плотность распределения выборки имеет вид:

$$p_\xi(x;a)=p_\xi(x_1,x_2,\dots,x_n;a)=\prod_{i=1}^n \frac{2}{\pi}\frac{1}{1+(x_i-a)^2}$$

Это пример, когда **нет достаточной статистики** для параметра a, кроме тривиальной $(\xi_1,\xi_2,\dots,\xi_n)$.

3.3. Теорема Блекуэлла (это вопрос лекционный, перенесен на практ. занятие)

С помощью этой теоремы можно улучшать оценки

Теорема 3.2. Пусть $\varphi(\xi)$ — **несмещённая** оценка параметра a: $M\varphi(\xi)=a$

и $\tau=T(\xi)$ — **достаточная** для *а* статистика. Определим условное математическое ожидание:

$$\varphi^* = M(\varphi|\tau) = \varphi^*(\tau).$$

Проверим, что $M(\varphi|\tau)$ является функцией τ и от параметра *a* не зависит:

$$M(\varphi|\tau) = \int \varphi(x) p_\xi(x|\tau)\,dx.$$

Здесь $p_\xi(x|\tau)$— **условное** распределение СВξ при условии известного τ, которое (в силу достаточности статистики τ) не зависит от параметра a. Интеграл зависит от τ.

Тогда:

1) $\varphi^*(\tau)$ является **несмещённой** оценкой параметра $a: M\varphi^*(\tau) = a.$

2) $D\varphi^* \le D\varphi.$

Действительно, полное математическое ожидание $M\varphi(\xi)$ можно определить по условному м.о. при условии известного τ:

$$a = M\varphi = MM\ \varphi|\tau\ = M\ \varphi^*\ .$$

Аналогично, полная дисперсия:

$$D\varphi = MD\ \varphi|\tau\ + DM\ \varphi|\tau\ = MD\ \varphi|\tau\ + D\ \varphi^*\ \ge D\ \varphi^*\ .$$

Следствие. Оценка φ^0 с минимальной дисперсией ($МД$-оценка), если она существует, является **функцией достаточной статистики**.

Действительно, пусть $\varphi^0(\xi)$— MD-оценка, и $\tau = T(\xi)$— достаточная для *a* статистика.

Для оценки $\varphi^*(\tau) = M(\varphi^0|\tau)$ справедливо неравенство $D\varphi^* \le D\varphi^0$.

Но $D\varphi^0 \le D\varphi^*$, т.к. φ^0 — MD-оценка. Следовательно, $D\varphi^0 = D\varphi^*$, и $\varphi^*(\tau)$ — MD-оценка, но она — функция достаточной статистики.

Пример 3.5. Пусть $\xi \equiv (\xi_1, \xi_2, ..., \xi_n)$ — выборка, $\xi_i \sim R[0,a], i = \overline{1,n}.$ Рассмотрим следующую оценку, полученную методом моментов для параметра a: $m = \dfrac{a}{2}, a = 2m,$

$$\hat{a} = 2\hat{m} = 2\frac{1}{n}\sum_{i=1}^{n}\xi_i,$$

Она является несмещённой, поскольку

$$M\hat{a} = 2\cdot\frac{n}{n}\cdot\frac{a}{2} = a.$$

В примере 3.3 показано, что СВ $\tau = \max\limits_{1\le i\le n}\xi_i$ является **достаточной** статистикой для параметра a. Построим новую оценку

$$a^* = M\left\{\frac{2}{n}\sum_{j=1}^{n}\xi_j \,\middle|\, \max_{1\le i\le n}\xi_i = y\right\} = \frac{2}{n}\sum_{j=1}^{n} M\ \xi_j \,\Big|\, \max_{1\le i\le n}\xi_i = y\ =$$

$$= \frac{2}{n}\left[(n-1)\frac{y}{2}+y\right] = \frac{n+1}{n} y = \frac{n+1}{n}\max \xi_i .$$

Мы воспользовались тем, что если СВ ξ распределена по закону *R*[0,*a*], поэтому условное распределение ξ при условии, что $\xi < y,\ y \in [0, a]$, является равномерным *R*[0,*y*]. Действительно, функция условного распределения

$$F(x \mid \xi < y) = P\{\xi < x \mid \xi < y\} = \frac{P\{\xi < x, \xi < y\}}{P\{\xi < y\}} =$$

$$= \begin{cases} \dfrac{P\{\xi < x\}}{P\{\xi < y\}}, & \text{если } x \le y, \\ 1, & \text{если } x > y, \end{cases} = \begin{cases} \dfrac{x/a}{y/a}, & \text{если } x \le y, \\ 1, & \text{если } x > y, \end{cases} = \begin{cases} \dfrac{x}{y}, & \text{если } x \le y, \\ 1, & \text{если } x > y. \end{cases}$$

И потому $M\left(\xi_j \,\middle|\, \max\limits_{1 \le i \le n} \xi_i = y\right) = \begin{cases} y/2, & \text{если } \xi_j \ne \max\limits_{1 \le i \le n} \xi_i, \\ y, & \text{если } \xi_j = \max\limits_{1 \le i \le n} \xi_i. \end{cases}$

Таким образом, построили оценку

$$a^* = \frac{n+1}{n} \max_i \xi_i .$$

Можно проверить, что она является **несмещённой** оценкой (уже рассматривали эту оценку).

Сравним дисперсии двух оценок: дисперсия исходной оценки:

$$D\hat{a} = \frac{4n}{n^2} \cdot \frac{a^2}{12} = \frac{a^2}{3n}.$$

Дисперсия новой оценки:

$$Da^* = \frac{a^2}{n(n+2)} \sim \frac{a^2}{n^2}$$ (см. пример 3.3 в разделе 3.2).

Итак, $Da^* \le D\hat{a}$. Заметим также, что это пример **сверхэффекивной** оценки, дисперсия которой убывает быстрее, чем $1/n$.

Домашнее задание

3.1. Пусть $x_1, x_2, ..., x_n$ — выборка из генеральной совокупности, распределённой по **показательному закону** $E(a)$. Найти достаточную статистику для параметра a.

3.2. Пусть $x_1, x_2, ..., x_n$ — выборка из генеральной совокупности, имеющей **равномерное** распределение $R[a, b]$ с **неизвестными** параметрами (a, b). Найти достаточную статистику для этих параметров.

Ответ: $\tau = \left(\min_i \xi_i ,\ \max_i \xi_i \right)$.

3.3. Случайная величина X задана плотностью распределения:

$$q(x; a, b) = \begin{cases} \frac{1}{b} e^{-(x-a)/b}, & \text{если } x \geq a, \\ 0, & \text{если } x < a, \end{cases}$$

с **неизвестными** параметрами (a, b) (это **смещённое на *a* показательное** распределение). По выборке $x_1, x_2, ..., x_n$ найти достаточную статистику и оценить параметры.

Ответ: $\tau = \left(\sum_{i=1}^{n} \xi_i ,\ \min_i \xi_i \right)$. $\hat{b} = \frac{n}{n-1}\left(\overline{\xi} - \min_i \xi_i \right)$, $\hat{a} = \frac{n}{n-1}\left(\min_i \xi_i - \frac{\overline{\xi}}{n} \right)$