Информация о минимуме

1. Минимум включает базовые задачи по теории вероятностей и статистике, для решения которых достаточно применить некоторый типичный алгоритм.

2. Каждая контрольная работа и экзамен состоят из основной части и минимума. В минимум входят задачи, схожие с теми, что представлены далее. За полностью верное решение этих задач можно получить до 5 баллов из 10.

3. Задачи решаются в тестовой форме: необходимо выбрать единственный верный вариант из нескольких доступных.

4. За ответы, не содержащие расписанного решения, исходя из которого понятна логика выбора варианта в тесте, выставляется оценка 0 баллов.

5. Минимум включает как простые задачи, так и задачи повышенной сложности. За правильное решение всех простых задач выставляется 4 балла, а за решение сложной – 1 балл.

6. Задачи повышенной сложности отмечаются пометкой (**доп.**) в конце условия. Эти задачи в контрольной работе могут в заметно большей степени отличаться от своих аналогов в данном файле.

7. Оценка за минимум не нормируется.

Минимум 3

1. Имеется выборка $X_1, ..., X_5$ с реализацией $x = (2, 0, 0, -1, 3)$.

   а) Найдите реализацию вариационного ряда и его третьего члена.

   б) Посчитайте реализацию выборочного среднего.

   в) Вычислите реализацию исправленной выборочной дисперсии.

   г) Определите, чему равняется реализация третьего начального выборочного момента.

   д) Запишите реализацию выборочной функции распределения.

   е) Найдите реализацию несмещенной оценки вероятности $P(X_1 > 0.5)$, используя информацию о всех наблюдениях в выборке.

   ж) Вычислите реализацию несмещенной оценки вероятности $P(X_1 > 0.5|X_1 \leq 2)$, используя информацию о всех наблюдениях в выборке.

   Решение:

   а) Сортируя, в порядке возрастания, реализации наблюдений, получаем реализацию вариационного ряда:

   $$(-1, 0, 0, 2, 3)$$

   б) Рассчитаем реализацию выборочного среднего:

   $$\overline{x}_5 = \frac{-1 + 0 + 0 + 2 + 3}{5} = 0.8$$

   в) Посчитаем реализацию выборочной дисперсии:

   $$\hat{\sigma}_5^2 = \frac{(-1-0.8)^2 + (0-0.8)^2 + (0-0.8)^2 + (2-0.8)^2 + (3-0.8)^2}{5-1} = 2.7$$

   г) Вычислим значение реализации третьего начального выборочного момента:

   $$\overline{x^3} = \frac{-1^3 + 0^3 + 0^3 + 2^3 + 3^3}{5} = 6.8$$

   д) Запишем реализацию выборочной функции распределения:

   $$\hat{F}_n(t) = \begin{cases} 0, \text{ если } t < -1 \\ 0.2, \text{ если } -1 \leq t < 0 \\ 0.6, \text{ если } 0 \leq t < 2 \\ 0.8, \text{ если } 2 \leq t < 3 \\ 1, \text{ если } t \geq 3 \end{cases}$$

   е) Найдем реализацию соответствующей оценки:

   $$\hat{P}\left(X > 0.5\right)(x) = \frac{1}{5}\left(I(x_1 > 0.5) + I(x_2 > 0.5) + I(x_3 > 0.5) + I(x_4 > 0.5) + I(x_5 > 0.5)\right) =$$
   $$= \frac{1}{5}\left(I(2 > 0.5) + I(0 > 0.5) + I(0 > 0.5) + I(-1 > 0.5) + I(3 > 0.5)\right) =$$
   $$= \frac{1}{5}(1 + 0 + 0 + 0 + 1) = 0.4$$

ж) По аналогии с предыдущим пунктом получаем:

$$\hat{P}\left(X_1 > 0.5 | X_1 \leq 2\right)(x) = \frac{\hat{P}\left(X_1 > 0.5 \cap X_1 \leq 2\right)(x)}{\hat{P}\left(X_1 \leq 2\right)(x)} = \frac{\frac{1}{5}\left(1+0+0+0+0\right)}{\frac{1}{5}\left(1+1+1+1+0\right)} = 0.25$$

2. Имеется выборка $X_1, X_2, ..., X_n$ из распределения со следующей функцией плотности:

$$f_{X_i}(t) = \begin{cases} \frac{2(\theta-t)}{\theta^2}, \text{при } t \in [0,\theta] \\ 0, \text{в противном случае} \end{cases}, \text{где } \theta > 0$$

а) Определите, при каком α оценка $\hat{\theta}_3 = X_1 - 2X_2 + \alpha X_3$ окажется несмещенной.

б) Найдите $\alpha > 0$, при котором последовательность оценок $\hat{\theta}_n = \frac{X_1+...+X_n}{\alpha(n+1)}$ окажется состоятельной, либо покажите, что такого α не существует.

в) Реализация выборки имеет вид $x = (1.5, 3.5, 3)$. При помощи второго начального момента найдите реализацию оценки параметра θ.

Решение:

а) Сперва рассчитаем математическое ожидание наблюдения:

$$E(X_i) = \int_0^\theta \frac{x \times 2(\theta - x)}{\theta^2}dx = \frac{\theta}{3}$$

Пользуясь полученным результатам найдем математическое ожидание оценки:

$$E\left(\hat{\theta}_3\right) = E(X_1 - 2X_2 + \alpha X_3) = E(X_1) - 2E(X_2) + \alpha E(X_3) =$$
$$E(X_1) - 2E(X_1) + \alpha E(X_1) = \frac{\theta}{3} - \frac{2\theta}{3} + \frac{\alpha\theta}{3}$$

У несмещенной оценки математическое ожидание равняется истинному значению оцениваемого параметра, откуда:

$$\frac{\theta}{3} - \frac{2\theta}{3} + \frac{\alpha\theta}{3} = \theta$$

Решая соответствующее равенство для θ получаем, что $\alpha = 4$.

б) Сперва найдем, к чему стремится математическое ожидание оценки:

$$\lim_{n\to\infty} E\left(\hat{\theta}_n\right) = \lim_{n\to\infty} E\left(\frac{X_1 + ... + X_n}{\alpha(n+1)}\right) = \lim_{n\to\infty} \frac{E(X_1) + ... + E(X_n)}{\alpha(n+1)} =$$
$$= \lim_{n\to\infty} \frac{nE(X_1)}{\alpha(n+1)} = \lim_{n\to\infty} \frac{n \times \frac{\theta}{3}}{\alpha(n+1)} = \lim_{n\to\infty} \frac{n\theta}{3\alpha(n+1)} = \frac{\theta}{3\alpha} \lim_{n\to\infty} \frac{n}{n+1} = \frac{\theta}{3\alpha} \times 1 = \frac{\theta}{3\alpha}$$

Поскольку состоятельность требует выполнение асимптотическое несмещенности, то должно соблюдаться $\frac{\theta}{3\alpha} = \theta$, а значит $\alpha = \frac{1}{3}$.

Убедимся, что при соответствующем значении α дисперсия последовательности оценок стремится к нулю. В противном случае необходимого значения α не существует.

$$\lim_{n\to\infty} Var\left(\hat{\theta}_n\right) = \lim_{n\to\infty} Var\left(\frac{X_1 + ... + X_n}{\alpha(n+1)}\right) = \lim_{n\to\infty} \frac{Var(X_1) + ... + Var(X_n)}{(\alpha(n+1))^2} =$$
$$= \lim_{n\to\infty} \frac{nVar(X_1)}{\alpha^2(n+1)^2} = 0$$

Обратим внимание, что для нахождения предела нет необходимости в расчете выражения для $Var(X_i)$, поскольку дисперсия наблюдения является константой, не зависящей от объема выборки n.

В данном случае дисперсия последовательности оценок стремится к нулю при любом $\alpha > 0$, в том числе при $\alpha = \frac{1}{3}$, а значит при соответствующем значении последовательность оценок окажется состоятельной.

в) Найдем второй начальный момент наблюдения:

$$E(X_i^2) = \int_0^\theta \frac{x^2 \times 2(\theta - x)}{\theta^2} dx = \frac{\theta^2}{6}$$

Отсюда получаем, что:

$$\theta = \sqrt{6E(X_i^2)}$$

Следовательно, оценка метода моментов будет иметь вид:

$$\hat{\theta} = \sqrt{6\overline{X^2}_n} = \sqrt{2(X_1^2 + X_2^2 + X_3^2)}$$

В итоге получаем реализацию оценки:

$$\hat{\theta}(x) = \sqrt{2 \times (1.5^2 + 3.5^2 + 3^2)} \approx 6.86$$

3. Имеется выборка $X_1, X_2, ..., X_n$ из распределения со следующей функцией вероятности:

x	1	2	3
$P(X_1 = x)$	0.4	$0.3 + \theta$	$0.3 - \theta$

Реализация выборки имеет вид $x = (1, 2, 2, 1, 3)$

а) При помощи первого начального момента найдите оценку параметра θ методом моментов и посчитайте ее реализацию.

б) Проверьте, является ли найденная вами в предыдущем пункте оценка несмещенной.

Решение:

а) Найдем математическое ожидание наблюдения:

$$E(X_i) = 1 \times 0.4 + 2 \times (0.3 + \theta) + 3 \times (0.3 - \theta) = 1.9 - \theta$$

Отсюда получаем, что:

$$1.9 - \theta = E(X_i) \implies \theta = 1.9 - E(X_i)$$

Следовательно, оценка метода моментов имеет вид:

$$\hat{\theta} = 1.9 - \overline{X}$$

В результате получаем реализацию данной оценки:

$$\hat{\theta}(x) = 1.9 - \frac{1 + 2 + 2 + 1 + 3}{5} = 0.1$$

б) Оценка является несмещенной, поскольку:

$$E(\hat{\theta}) = E(1.9 - \overline{X}) = E(1.9) - E(\overline{X}) = 1.9 - (1.9 - \theta) = \theta$$

4. Имеется выборка X_1, X_2, X_3 из распределения со следующей функцией плотности:

$$f_{X_1}(t) = \begin{cases} 0.25t, \text{ если } t \in [1, 3] \\ 0, \text{ в противном случае} \end{cases}$$

Найдите вероятность того, что:

а) Наибольшее наблюдение в выборке окажется больше 2.

б) Наименьшее наблюдение в выборке не превысит 2. **(доп.)**

в) Медианное значение в выборке не превысит 2. **(доп.)**

Решение:

а) Сперва найдем функцию распределения наблюдения на носителе, то есть при $t \in [1, 3]$ получаем:

$$F_{X_1}(t) = \int\limits_1^t 0.25q dq = 0.125(t^2 - 1)$$

Применяя формулу для максимальной экстремальной статистики получаем:

$$P\left(X_{(3)} > 2\right) = 1 - P\left(X_{(3)} \leq 2\right) = 1 - F_{X_{(3)}}(2) = 1 - \left(F_{X_1}(2)\right)^3 =$$
$$= 1 - \left(0.125 \times (2^2 - 1)\right)^3 \approx 0.947$$

б) По аналогии с помощью формулы для минимальной экстремальной статистики получаем:

$$P\left(X_{(1)} \leq 2\right) = F_{X_{(1)}}(2) = 1 - \left(1 - F_{X_1}(2)\right)^3 =$$
$$= 1 - \left(1 - 0.125 \times (2^2 - 1)\right)^3 \approx 0.756$$

в) Можно вновь воспользоваться формулой:

$$P(X_{(2)} \leq 2) = F_{X_{(2)}}(2) = \sum_{k=2}^{3} C_3^k \left(1 - F_{X_1}(2)\right)^{3-k} \left(F_{X_1}(2)\right)^k =$$
$$= C_3^2 \left(1 - F_{X_1}(2)\right)^{3-2} \left(F_{X_1}(2)\right)^2 + C_3^3 \left(1 - F_{X_1}(2)\right)^{3-3} \left(F_{X_1}(2)\right)^3 =$$
$$= 3 \left(1 - F_{X_1}(2)\right) \left(F_{X_1}(2)\right)^2 + \left(F_{X_1}(2)\right)^3 =$$
$$= 3 \left(1 - 0.125 \times (2^2 - 1)\right) \left(0.125 \times (2^2 - 1)\right)^2 + \left(0.125 \times (2^2 - 1)\right)^3 \approx 0.316$$

В качестве альтернативы использованию готовой формулы нетрудно вывести ее самостоятельно. Для этого обратим внимание на то, что событие $X_{(2)} \leq 2$ происходит в случае, если хотя бы два наблюдения не больше 2. То есть если все наблюдения не больше 2 или ровно два наблюдения не меньше 2, а оставшееся – меньше 2.

$$\begin{gathered}
P(X_{(2)} \leq 2) = P\left(X_1 \leq 2 \cap X_2 \leq 2 \cap X_3 \leq 2\right) + \\
+P\left(X_1 \leq 2 \cap X_2 \leq 2 \cap X_3 > 2\right) + \\
+P\left(X_1 \leq 2 \cap X_2 > 2 \cap X_3 \leq 2\right) + \\
+P\left(X_1 > 2 \cap X_2 \leq 2 \cap X_3 \leq 2\right) = \\
= P\left(X_1 \leq 2 \cap X_2 \leq 2 \cap X_3 \leq 2\right) + 3P\left(X_1 \leq 2 \cap X_2 \leq 2 \cap X_3 > 2\right) = \\
= P(X_1 \leq 2)P(X_2 \leq 2)P(X_3 \leq 2) + 3P(X_1 \leq 2)P(X_2 \leq 2)\left(1 - P(X_3 \leq 2)\right) = \\
= F_{X_1}(2)^3 + 3F_{X_1}(2)^2\left(1 - F_{X_1}(2)\right) = \\
= \left(0.125(2^2-1)\right)^3 + 3\left(0.125(2^2-1)\right)^2\left(1 - (0.125(2^2-1)\right) \approx 0.316
\end{gathered}$$

5. Имеются независимые стандартные нормальные случайные величины $Z_1, \ldots, Z_5$. Найдите распределение случайной величины:

   а) $X = Z_1^2 + Z_3^2 + Z_5^2$.
   б) $X = \frac{2Z_1}{\sqrt{Z_2^2+Z_3^2+Z_4^2+Z_5^2}}$.
   в) $X = 1.5 \times \frac{Z_1^2+Z_2^2}{Z_3^2+Z_4^2+Z_5^2}$.
   г) $X = \frac{(Z_1+Z_2+Z_3+Z_4)^2}{4} + Z_5^2$. **(доп.)**

   Решение:

   а) Поскольку речь идет о распределении суммы квадратов трех независимы стандартных нормальных случайных величин, то $X \sim \chi^2(3)$.
   б) Поскольку $(Z_2^2 + Z_3^2 + Z_4^2 + Z_5^2) \sim \chi^2(4)$, то:

   $$X = \frac{Z_1}{\sqrt{\frac{1}{4}\left(Z_2^2 + Z_3^2 + Z_4^2 + Z_5^2\right)}} \sim t(4)$$

   в) Так как числитель и знаменатель независимы и имеют Хи-квадрат распределения, а $1.5 = \frac{3}{2}$, то очевидно, что: $X \sim F(2, 3)$.
   г) Перепишем выражение в виде:

   $$X = \left(\frac{Z_1 + Z_2 + Z_3 + Z_4}{2}\right)^2 + Z_5^2$$

   Пользуясь тем, что линейная комбинация независимых нормальных случайных величин имеет нормальное распределение, получаем:

   $$E\left(\frac{Z_1+Z_2+Z_3+Z_4}{2}\right) = \frac{E(Z_1)+E(Z_2)+E(Z_3)+E(Z_4)}{2} = \frac{0+0+0+0}{2} = 0$$
   $$Var\left(\frac{Z_1+Z_2+Z_3+Z_4}{2}\right) = \frac{Var(Z_1)+Var(Z_2)+Var(Z_3)+Var(Z_4)}{2^2} = \frac{1+1+1+1}{4} = 1$$
   $$\frac{Z_1+Z_2+Z_3+Z_4}{2} \sim \mathcal{N}\left(0, 1\right)$$

   В результате $X \sim \chi^2(2)$, поскольку представлен в виде суммы квадратов двух стандартных нормальных случайных величин.

6. Имеется выборка X_1, X_2, X_3, X_4, X_5 из распределения с функцией распределения:

$$F_{X_1}(t) = \begin{cases} 0, \text{ если } x < 2 \\ 0.5, \text{ если } 2 \leq x < 5 \\ 0.9 - \theta, \text{ если } 5 \leq x < 10 \\ 1, \text{ если } x \geq 10 \end{cases}$$

Где $\theta \in [0, 0.4)$. Реализация выборки имеет вид $x = (2, 5, 10, 5, 2)$

а) Запишите таблицу распределения для X_1.

б) Оцените параметр θ при помощи метода максимального правдоподобия и запишите реализацию его оценки.

Решение:

а) Найдем вероятности всех значений, входящих в носитель X_1:

$$P(X_1 = 2) = 0.5$$
$$P(X_1 = 5) = P(X_1 \leq 5) - P(X_1 < 5) = (0.9 - \theta) - 0.5 = 0.4 - \theta$$
$$P(X_1 = 10) = P(X_1 \leq 10) - P(X_1 < 10) = 1 - (0.9 - \theta) = 0.1 + \theta$$

Отсюда получаем таблицу распределения:

t	2	5	10
$P(X_1 = t)$	0.5	$0.4 - \theta$	$0.1 + \theta$

б) Запишем функцию правдоподобия:

$$L(\theta; x) = P(X_1 = 2) \times P(X_2 = 5) \times P(X_3 = 10) \times P(X_4 = 5) \times P(X_5 = 2) =$$
$$= 0.5 \times (0.4 - \theta) \times (0.1 + \theta) \times (0.4 - \theta) \times 0.5$$

Найдем логарифм функции правдоподобия:

$$\ln L(\theta; x) = \ln(0.5) + \ln(0.4 - \theta) + \ln(0.1 + \theta) + \ln(0.4 - \theta) + \ln(0.5)$$

Для того, чтобы найти максимум функции правдоподобия по оцениваемому параметру θ, рассмотрим условия первого порядка (FOC - first order conditions).

$$\frac{d \ln L(\theta; x)}{d\theta} = -\frac{1}{0.4 - \theta} + \frac{1}{0.1 + \theta} - \frac{1}{0.4 - \theta} = 0 \implies$$
$$\implies \frac{2}{0.4 - \theta} = \frac{1}{0.1 + \theta} \implies 0.2 + 2\theta = 0.4 - \theta$$

Решая соответствующее равенство получаем точку, подозреваемую на максимум $\theta^* = \frac{1}{15}$. Убедимся в том, что данная точка сооветствует максимуму, рассмотрев условия второго порядка:

$$\frac{d^2 \ln L(\theta; x)}{d\theta^2}|_{\theta = \theta^*} = -\frac{2}{(0.4 - \theta^*)^2} - \frac{1}{(0.1 + \theta^*)^2} = -\frac{2}{\left(0.4 - \frac{1}{15}\right)^2} - \frac{1}{\left(0.1 + \frac{1}{15}\right)^2} = -54 < 0$$

Таким образом, получаем реализацию ММП оценки $\hat{\theta} = \frac{1}{15}$.

7. Имеется выборка $X_1, ..., X_n$ из нормального распределения, причем $E(X_1) = 1$ и $Var(X_1) = \theta$. Реализация выборки имеет вид $x = (0, 2, -1)$.

   а) Оцените параметр θ при помощи метода максимального правдоподобия и запишите реализацию полученной оценки.

   б) Найдите реализацию оценки асимптотической дисперсии найденной в предыдущем пункте ММП оценки.

   в) Постройте 99%-й асимптотический доверительный интервал для параметра θ.

   г) На 1%-м уровне значимости протестируйте гипотезу о том, что $\theta = 1$, против альтернативы $\theta > 1$. Вычислите p-value. При этом используйте оценку асимптотической дисперсии ММП оценки параметра θ.

   д) Постройте 99%-й асимптотический доверительный интервал для $\ln(\theta)$ (воспользуйтесь дельта-методом). **(доп.)**

   е) На 1%-м уровне значимости протестируйте гипотезу о том, что $\ln(\theta) = 2$, против альтернативы $\ln(\theta) < 2$ (воспользуйтесь дельта-методом). **(доп.)**

Примечание: на практике при столь малом объеме выборки использовать асимптотические свойства ММП оценок, в частности, для построения доверительных интервалов и тестирования гипотез – не корректно. Однако, для простоты вычислений, в данном случае учитываются лишь несколько наблюдений

Решение:

а) Запишем функцию правдоподобия:

$$L(\theta; X) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} e^{\frac{-(X_i-1)^2}{2\theta}}$$

Найдем логарифм функции правдоподобия:

$$L(\theta; X) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\theta}} e^{\frac{-(X_i-1)^2}{2\theta}}\right) = -\frac{n}{2}\left(\ln(2\pi) + \ln(\theta)\right) - \sum_{i=1}^{n} \frac{(X_i-1)^2}{2\theta}$$

Для того, чтобы найти максимум функции правдоподобия по оцениваемому параметру θ, рассмотрим условия первого порядка (FOC - first order conditions).

$$\frac{d\ln L(\theta; x)}{d\theta} = -\frac{n}{2\theta} + \sum_{i=1}^{n} \frac{(X_i-1)^2}{2\theta^2} = 0 \implies$$
$$\implies \frac{n}{2\theta} = \sum_{i=1}^{n} \frac{(X_i-1)^2}{2\theta^2} \implies n = \sum_{i=1}^{n} \frac{(X_i-1)^2}{\theta}$$

Решая соответствующее равенство получаем точку, подозреваемую на максимум:

$$\theta^* = \sum_{i=1}^{n} \frac{(X_i-1)^2}{n}$$

Убедимся в том, что данная точка соответствует максимуму, рассмотрев условия второго порядка:

$$\frac{d^2 \ln L(\theta; x)}{d\theta^2}|_{\theta=\theta^*} = \frac{n}{2(\theta^*)^2} - \sum_{i=1}^{n} \frac{(X_i-1)^2}{(\theta^*)^3} \propto \frac{n}{2} - \sum_{i=1}^{n} \frac{(X_i-1)^2}{\theta^*} =$$

$$= \frac{n}{2} - \frac{\sum\limits_{i=1}^{n}(X_i-1)^2}{\left(\sum\limits_{i=1}^{n} \frac{(X_i-1)^2}{n}\right)} = \frac{n}{2} - n = -\frac{n}{2} < 0$$

Где знак $\propto$ обозначает пропорциональное равенство (с точностью до умножения на положительную константу, что не влияет на знак выражения). Таким образом, получаем ММП оценку:

$$\hat{\theta} = \sum_{i=1}^{n} \frac{(X_i-1)^2}{n}$$

Наконец, подставляя реализации наблюдений, получаем реализацию оценки:

$$\hat{\theta}(x) = \frac{(0-1)^2 + (2-1)^2 + (-1-1)^2}{3} = 2$$

б) Обратим внимание, что, поскольку $E(X_i) = 1$, то:

$$E((X_i-1)^2) = E((X_i - E(X_i))^2) = Var(X_i) = \theta$$

Пользуясь данным результатом найдем информацию Фишера:

$$I(\theta) = -E\left(\frac{n}{2\theta^2} - \sum_{i=1}^{n} \frac{(X_i-1)^2}{\theta^3}\right) = \sum_{i=1}^{n} \frac{E\left((X_i-1)^2\right)}{\theta^3} - \frac{n}{2\theta^2} =$$

$$= \sum_{i=1}^{n} \frac{Var(X_i)}{\theta^3} - \frac{n}{2\theta^2} = \frac{nVar(X_1)}{\theta^3} - \frac{n}{2\theta^2} = \frac{n}{\theta^2} - \frac{n}{2\theta^2} = \frac{n}{2\theta^2}$$

В качестве альтернативы, информацию Фишера, содержащуюся во всей выборке, можно было бы найди с помощью информации Фишера, содержащейся в одном наблюдении:

$$i(\theta) = -E\left(\frac{1}{2\theta^2} - \frac{(X_1-1)^2}{\theta^3}\right) = \frac{1}{2\theta^2} \implies I(\theta) = ni(\theta) = \frac{n}{2\theta^2}$$

Исходя из информации Фишера получаем выражение для асимптотической дисперсии ММП оценки:

$$As.Var(\hat{\theta}) = \frac{1}{I(\theta)} = \frac{2\theta^2}{n}$$

Найдем оценку асимптотической дисперсии за счет замены истинного значения параметра на его ММП оценку:

$$\widehat{As.Var}(\hat{\theta}) = \frac{2\hat{\theta}^2}{n}$$

Наконец, получаем реализацию данной оценки:

$$\widehat{As.Var}(\hat{\theta})(x) = \frac{2 \times 2^2}{3} = \frac{8}{3}$$

в) Формула для асимптотического доверительного интервала параметра, оцененного при помощи метода максимального правдоподобия, имеет вид:

$$\left(\hat{\theta} - z_{0.995}\sqrt{\widehat{As.Var}(\hat{\theta})}, \hat{\theta} + z_{0.995}\sqrt{\widehat{As.Var}(\hat{\theta})}\right)$$

Запишем реализацию соответствующего асимптотического доверительного интервала:

$$\left(2 - 2.58 \times \sqrt{\frac{8}{3}}, 2 + 2.58 \times \sqrt{\frac{8}{3}}\right) \approx (-2.2, 6.2)$$

г) Поскольку тестируется гипотеза о параметре, оценка которого была получена при помощи метода максимального правдоподобия, то можно воспользоваться следующей статистикой (при верной нулевой гипотезе в асимптотике она имеет стандартное нормальное распределение):

$$T(X) = \frac{\hat{\theta} - \theta_0}{\sqrt{\widehat{As.Var}(\hat{\theta})}}$$

Найдем реализацию соответствующей статистики:

$$T(x) = \frac{2-1}{\sqrt{\frac{8}{3}}} \approx 0.61$$

Поскольку тестируется правосторонняя гипотеза, то критическая область имеет вид $\mathcal{T}_{0.01} = (z_{1-0.01}, \infty) = (2.33, \infty)$. Поскольку $0.6 \notin (2.33, \infty)$, то нулевая гипотеза не отвергается. Также, рассчитаем p-value:

$$\text{p-value} \approx 1 - \Phi(0.61) \approx 0.27$$

Таким образом, нулевая гипотеза не отвергается на любом уровне значимости, меньше, чем 0.27, в том числе не отвергается на уровне значимости 0.01.

д) Пользуясь инвариантностью ММП оценок получаем ММП оценку $\ln(\theta)$:

$$\widehat{\ln(\theta)} = \ln(\hat{\theta})$$

В соответствии с дельта методом найдем ее асимптотическую дисперсию:

$$As.Var(\ln(\hat{\theta})) = \left(\frac{d\ln(\theta)}{d\theta}\right)^2 \times \frac{2\theta^2}{n} = \left(\frac{1}{\theta}\right)^2 \times \frac{2\theta^2}{n} = \frac{2}{3}$$

Оценка этой асимптотической дисперсии получается за счет замены параметра θ на его оценку $\hat{\theta}$. Однако, поскольку в данном случае θ не входит в соответствующее выражение, то автоматически получаем:

$$\widehat{As.Var}(\ln(\hat{\theta})) = \frac{2}{3}$$

Используя полученные результаты запишем реализацию искомого асимптотического доверительного интервала:

$$\left(\ln(2) - 2.58 \times \sqrt{\frac{2}{3}}, \ln(2) + 2.58 \times \sqrt{\frac{2}{3}}\right) \approx (-1.2, 2.6)$$

e) Найдем реализацию тестовой статистики:

$$T(x) = \frac{\ln(2) - 2}{\sqrt{\frac{2}{3}}} \approx -1.6$$

Поскольку тестируется левосторонняя гипотеза, то критическая область имеет вид $\mathcal{T}_{0.01} = (-\infty, -z_{1-0.01}) = (-\infty, -2.33)$. Поскольку $-1.6 \notin (-\infty, -2.33)$, то нулевая гипотеза не отвергается. Также, рассчитаем p-value:

$$\text{p-value} \approx \Phi(-1.6) \approx 0.054$$

Таким образом, нулевая гипотеза не отвергается на любом уровне значимости, меньше, чем 0.054, в том числе не отвергается на уровне значимости 0.01.

8. Имеются две независимые выборки $X_1, ..., X_{100}$ и $Y_1, ..., Y_{225}$ из распределений с конечными математическими ожиданиями и дисперсиями. По этим выборкам были рассчитаны выборочные средние и исправленные выборочные дисперсии $\overline{X}_{100} = 10$, $\overline{Y}_{225} = 20$, $\hat{\sigma}^2_X = 25$ и $\hat{\sigma}^2_Y = 100$. Найдите реализацию 90%-го асимптотического доверительного интервала для:

   а) $E(X_1)$
   б) $E(X_1 - Y_1)$
   в) $Var(X_1)$, предполагая, что X_1 имеет нормальное распределение.
   г) $Var(X_1)/Var(Y_1)$, предполагая, что X_1 и Y_1 имеют нормальное распределение.

Подсказки: квантили уровня 0.05 и 0.95 Хи-квадрат распределения с 99 степенями свободы равняются 77.05 и 123.22 соответственно. Квантили уровня 0.05 и 0.95 распределения Фишера $F(224, 99)$ равняются 0.76 и 1.34 соответственно. В качестве подсказки может быть дано большее число квантилей, среди которых необходимо будет выбрать подходящие.

Решение:

а) Найдем реализацию 90%-го асимптотического доверительного интервала для математического ожидания (среднего):

$$\left(10 - 1.65 \times \sqrt{\frac{25}{100}}, 10 + 1.65 \times \sqrt{\frac{25}{100}}\right) = (9.175, 10.825)$$

б) Запишем реализацию 90%-го асимптотического доверительного интервала для разницы математических ожиданий (средних) наблюдений из независимых выборок:

$$\left(10 - 20 - 1.65 \times \sqrt{\frac{25}{100} + \frac{100}{225}}, 10 - 20 + 1.65 \times \sqrt{\frac{25}{100} + \frac{100}{225}}\right) = (11.375, -8.625)$$

в) Выпишем реализацию 90%-го доверительного интервала для дисперсии наблюдения из нормального распределения при неизвестном математическом ожидании:

$$\left(\frac{(100-1) \times 25}{123.22}, \frac{(100-1) \times 25}{77.05}\right) \approx (20.09, 32.12)$$

г) Определим реализацию 90%-го доверительного интервала для отношения дисперсий наблюдений из нормального распределения при неизвестных математических ожиданиях:

$$\left(\frac{25}{100}\times 0.76, \frac{25}{100}\times 1.34\right) = (0.19, 0.335)$$

9. Имеются две независимые выборки $X_1, ..., X_{100}$ и $Y_1, ..., Y_{225}$ из распределений с конечными математическими ожиданиями и дисперсиями. По этим выборкам были рассчитаны выборочные средние и исправленные выборочные дисперсии $\overline{X}_{100} = 10$, $\overline{Y}_{225} = 20$, $\hat{\sigma}_X^2 = 25$ и $\hat{\sigma}_Y^2 = 100$. Рассчитайте p-value и сделайте вывод на 5%-м уровне значимости в тестах со следующими гипотезами:

а) $H_0 : E(X_1) = 11$ против $H_1 : E(X_1) < 11$.

б) $H_0 : E(X_1) = E(Y_1)$ против $H_1 : E(X_1) > E(Y_1)$

в) $H_0 : Var(X_1) = 22$ против $H_1 : Var(X_1) \neq 22$, предполагая, что X_1 имеет нормальное распределение. Подсказка: $F_{\chi^2(99)}(112.5) \approx 0.833$ и $F_{\chi^2(100)}(102.5) \approx 0.588$.

г) $H_0 : Var(X_1) = Var(Y_1)$ против $H_1 : Var(X_1) > Var(Y_1)$, предполагая, что X_1 и Y_1 имеют нормальное распределение. Подсказка: $F_{F_{100-1,225-1}}(0.25) \approx 0$ и $F_{F_{100-1,225-1}}(0.8) \approx 0.1$.

Решение:

а) Применим левосторонний тест о равенстве математического ожидания определенному значению для выборки из распределения с конечными математическим ожиданием и дисперсией:

$$T(x) = \frac{10-11}{\sqrt{\frac{25}{100}}} = -2$$
$$\text{p-value} = \Phi(-2) \approx 0.023$$

Поскольку p-value < 0.05, то нулевая гипотеза отвергается на 5%-м уровне значимости.

б) Воспользуемся правосторонним тест о равенстве математических ожиданий для независимых выборок из распределения с конечными математическими ожиданиями и дисперсиями:

$$T(x) = \frac{10-20}{\sqrt{\frac{25}{100}+\frac{100}{225}}} = -12$$
$$\text{p-value} = 1-\Phi(-12) \approx 1$$

Поскольку p-value > 0.05, то нулевая гипотеза не отвергается на 5%-м уровне значимости.

в) Будем использовать двухсторонний тест о равенстве дисперсии наблюдения выборки из нормального распределения определенному значению:

$$T(x) = \frac{(100-1)\times 25}{22} \approx 112.5$$
$$\text{p-value} = 2\min(1-F_{\chi^2(100-1)}(112.5), F_{\chi^2(100-1)}(112.5)) = 2\min(0.167, 0.833) = 0.334$$

Поскольку p-value > 0.05, то нулевая гипотеза не отвергается на 5%-м уровне значимости.

г) Применим правосторонний тест о равенстве дисперсий наблюдений выборок из нормального распределения:

$$T(x) = \frac{25}{100} = 0.25$$
$$\text{p-value} = 1 - F_{F_{100-1,225-1}}(0.25) \approx 1 - 0 = 0$$

Поскольку p-value > 0.05, то нулевая гипотеза не отвергается на 5%-м уровне значимости.

10. Имеются независимые выборки $X = (X_1, ..., X_{100})$ и $Y(Y_1, ..., Y_{64})$ из распределения Бернулли с параметрами p_X и p_Y соответственно. Пятый начальный выборочный момент по выборке X оказался равен 0.2, а сумма элементов выборки Y равна 16. Рассчитайте p-value и сделайте вывод на 10%-м уровне значимости в тестах со следующими гипотезами:

а) $H_0 : p_X = 0.15$ против $p_X \neq 0.15$.

б) $H_0 : p_X = p_Y$ против $p_X < p_Y$. **(доп.)**

Решение:

а) Применим двухсторонний тест о равенстве вероятности (доли) определенному значению, пользуясь тем, что все начальные выборочные моменты для выборки из распределения Бернулли совпадают:

$$\overline{X}_{100} = 0.2$$
$$T(x) = \frac{0.2 - 0.15}{\sqrt{\frac{0.15 \times (1-0.15)}{100}}} \approx 1.4$$
$$\text{p-value} = 2\min(\Phi(1.4), 1 - \Phi(1.4)) \approx 0.16$$

Поскольку p-value > 0.1, то нулевая гипотеза не отвергается на 10%-м уровне значимости.

б) Применим левосторонний тест о равенстве вероятностей (долей) определенному значению:

$$\overline{X}_{100} = 0.2 \qquad \overline{Y}_{64} = \frac{16}{64} = 0.25$$
$$\sum_{i=1}^{100} X_i = 0.2 \times 100 = 20 \qquad \sum_{i=1}^{64} Y_i = 16$$
$$Z_{100,64} = \frac{20 + 16}{100 + 64} = \frac{9}{41}$$
$$T(x) = \frac{0.2 - 0.25}{\sqrt{\frac{9}{41} \times \left(1 - \frac{9}{41}\right) \times (1/100 + 1/64)}} \approx -0.75$$
$$\text{p-value} = \Phi(-0.75) \approx 0.22$$

Поскольку p-value > 0.1, то нулевая гипотеза не отвергается на 10%-м уровне значимости.

11. Имеется выборка, состоящая из одного наблюдения X_1, имеющего равномерное распределение. Тестируется гипотеза $H_0 : X_1 \sim U(0, 1)$ против альтернативы $H_1 : X_1 \sim U(0, 2)$. Нулевая гипотеза отвергается, если $X_1 > 0.5$.

а) Запишите критическую область соответствующего теста.

б) Найдите уровень значимости данного теста.

в) Определите мощность этого теста.

Решение:

а) Критическая область теста $\mathcal{T}_\alpha = (0.5, \infty)$. Поскольку как при альтернативной, так и при нулевой гипотезах наблюдения не принимают значения больше 2, то критическую область можно также записать в виде $\mathcal{T}_\alpha = (0.5, 2)$.

б) Уровень значимости теста совпадает с вероятностью совершить ошибку первого рода. Поскольку $X_1|H_0 \sim U(0,1)$, то:

$$\alpha = P(X_1 \in (0.5, 2)|H_0) = P(X_1 > 0.5|H_0) = 1 - F_{X_1|H_0}(0.5) = \frac{0.5}{1} = 0.5$$

в) Сперва найдем вероятность ошибки второго рода, учитывая, что $X_1|H_1 \sim U(0,2)$:

$$\beta = P(X_1 \notin (0.5, 2)|H_1) = P(X_1 \in (0, 0.5)|H_1) = \frac{0.5}{2} = 0.25$$

В результате получаем, что мощность теста равняется:

$$1 - \beta = 1 - 0.25 = 0.75$$

12. Имеется выборка $X_1, ..., X_n$ из экспоненциального распределения с параметром λ. Пусть $n = 2$. Тестируется гипотеза $H_0 : \lambda = 1$ против альтернативы $H_1 : \lambda < 1$. Нулевая гипотеза отвергается, если $\max(X_1, ..., X_n) < k$.

а) Найдите значение параметра k, при котором уровень значимости теста окажется равен 0.25. **(доп.)**

б) Посчитайте p-value теста, если реализация выборки имеет вид $x = (1, 2)$. **(доп.)**

Решение:

а) Запишем функцию распределения тестовой статистики:

$$T(X) = \max(X_1, ..., X_n)$$
$$F_{T_X}(t) = (1 - e^{-\lambda t})^n = (1 - e^{-\lambda t})^2, \text{ где t > 0}$$

Используя найденное распределение рассчитаем уровень значимости теста:

$$\alpha = P(T(X) < k|H_0) = (1 - e^{-1\times k})^2 = 0.25 \implies 1 - e^{-k} = 0.5 \implies$$
$$\implies e^{-k} = 0.5 \implies -k = \ln(0.5) \implies k = -\ln(0.5) \approx 0.69$$

б) Пользуясь реализацией тестовой статистики $T(x) = \max(1, 2) = 2$ и тем, что критическая область является левосторонней, получаем:

$$\text{p-value} = F_{T(X)|H_0}(2) = (1 - e^{-1\times 2})^2 \approx 0.75$$