Фамилия:...

Имя:...

Группа:...

Задача №1

Количество золота (в килограммах), добываемого гномом за день, не зависит от количества золота, добытого другими гномами и хорошо описывается распределением со следующей функцией плотности:

$$f_{X_1}(t) = \begin{cases} \theta e^{\theta(1-t)}, \text{ если } t \geq 1 \\ 0, \text{ в противном случае} \end{cases} \quad \text{, где } \theta > 0$$

Гном статистик посчитал, что за день 100 гномов, суммарно, добыли 200 килограмм золота.
Подсказка: $E(X_1) = \frac{\theta+1}{\theta}$.

1. Оцените параметр θ с помощью метода моментов и запишите реализацию полученной оценки. **(2 балла)**
2. Оцените параметр θ при помощи метода максимального правдоподобия и запишите реализацию полученной оценки (в данном пункте проверять условия второго порядка не нужно). **(2 балла)**
3. Проверьте соблюдение условий второго порядка. **(1 балл)**
4. Найдите реализацию оценки асимптотической дисперсии найденной в предыдущем пункте ММП оценки. **(2 балла)**
5. Постройте 90%-й асимптотический доверительный интервал для параметра θ. **(2 балла)**
6. На 10%-м уровне значимости протестируйте гипотезу о том, что $\theta = 1.1$, против альтернативы $\theta \neq 1.1$. При этом используйте оценку асимптотической дисперсии ММП оценки параметра θ. **(2 балла)**
7. Постройте 90%-й асимптотический доверительный интервал для $E(e^{-X_1})$ (воспользуйтесь дельта-методом). **(3 балла)**
8. На 5%-м уровне значимости протестируйте гипотезу о том, что $E(e^{-X_1}) = 0.2$, против альтернативы $E(e^{-X_1}) < 0.2$ (воспользуйтесь дельта-методом). **(3 балла)**

Подсказка: $\int e^{a+bt}dt = \frac{1}{b}e^{a+bt}$.

Решение:

1. Воспользуемся первым начальным моментом:

$$E(X_1) = \frac{\theta + 1}{\theta}$$

   Выразим параметр через первый начальный момент:

$$\theta E(X_1) = \theta + 1 \implies \theta(E(X_1) - 1) = 1 \implies \theta = \frac{1}{E(X_1) - 1}$$

   Отсюда получаем оценку:

$$\hat{\theta} = \frac{1}{\overline{X}_n - 1}$$

   Найдем ее реализацию:

$$\hat{\theta}(x) = \frac{1}{200/100 - 1} = 1$$

2. Запишем функцию правдоподобия:

$$L(\theta; x) = \prod_{i=1}^{n} \theta e^{\theta(1-x_i)}$$

   Логарифмируем функцию правдоподобия:

$$\ln L(\theta; x) = n\ln(\theta) + \theta \sum_{i=1}^{n}(1 - x_i)$$

   Рассмотрим условия первого порядка:

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

   Решая соответствующее равенство получаем оценку, которая, предположительно, является ММП оценкой:

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

   Найдем реализацию соответствующей оценки:

$$\hat{\theta}(x) = \frac{100}{200 - 100} = 1$$

3. Условия второго порядка соблюдены, поскольку вторая производная всегда отрицательна:

$$\frac{d^2 \ln L(\theta; x)}{d\theta^2} = -\frac{n}{\theta^2} < 0$$

4. Найдем информацию Фишера:

$$I(\theta) = -E\left(-\frac{n}{\theta^2}\right) = \frac{n}{\theta^2}$$

Отсюда получаем асимптотическую дисперсию ММП оценки:

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

Подставляя оценку вместо истинного значения параметра получаем оценку асимптотической дисперсии ММП оценки:

$$As.\hat{V}ar(\hat{\theta}) = \frac{\hat{\theta}^2}{n}$$

Найдем реализацию соответствующе оценки:

$$As.\hat{V}ar(\hat{\theta})(x) = \frac{1^2}{100} = 0.01$$

5. Построим 90%-й асимптотический доверительный интервал для θ:

$$\left(1 - 1.65\sqrt{0.01}, 1 + 1.65\sqrt{0.01}\right) = (0.835, 1.165)$$

6. Тестируется гипотеза $H_0 : \theta = 1.1$ против альтернативы $H_1 : \theta \neq 1.1$.

Посчитаем реализацию тестовой статистики.

$$T(x) = \frac{1 - 1.1}{\sqrt{0.01}} = -1$$

Вычислим p-value:

$$\text{p-value} = 2\min\left(1 - \Phi(-1), \Phi(-1)\right) \approx 0.317$$

Поскольку $0.317 > 0.1$, то нулевая гипотеза не отвергается на 10%-м уровне значимости.

В качестве альтернативного подхода к тестированию можно выписать критическую область:

$$\mathcal{T} = (-\infty, z_{0.05}) \cup (z_{0.95}, \infty) = (-\infty, -1.65) \cup (1.65, \infty)$$

.

Поскольку $-1 \notin (-\infty, -1.65)$ и $-1 \notin (1.65, \infty)$, то нулевая гипотеза не отвергается на 10%-м уровне значимости.

7. Найдем соответствующее математическое ожидание:

$$E(e^{X_1}) = \int_1^\infty e^{-t}\theta e^{\theta(1-t)}dt = \int_1^\infty \theta e^{\theta(1-t)-t}dt = \int_1^\infty \theta e^{\theta-(1+\theta)t}dt =$$
$$= -\frac{\theta}{1+\theta}e^{\theta-(1+\theta)t}|_1^\infty = \frac{\theta}{1+\theta}e^{-1}$$

По свойству инвариантности ММП оценок получаем ММП оценку соответствующего математического ожидания:

$$\hat{E}(e^{X_1}) = \frac{\hat{\theta}}{1+\hat{\theta}}e^{-1}$$

При помощи дельта метода найдем асимптотическую дисперсию:

$$As.Var(\hat{E}(e^{X_1})) = \left(\frac{d\frac{\theta}{1+\theta}e^{-1}}{d\theta}\right)^2 \times \frac{\theta^2}{n} = \frac{\theta^2}{n(1+\theta)^4}e^{-2}$$

Ее оценка будет иметь вид:

$$\widehat{As.Var}(\hat{E}(e^{X_1})) = \frac{\hat{\theta}^2}{n(1+\hat{\theta})^4}e^{-2}$$

Выпишем реализации найденных оценок:

$$\hat{E}(e^{X_1})(x) = \frac{1}{1+1}e^{-1} \approx 0.184$$
$$\widehat{As.Var}(\hat{E}(e^{X_1}))(x) = 0.0000846$$

Пользуясь полученными результатами запишем асимптотический доверительный интервал:

$$\left(0.184 - 1.65\sqrt{0.0000846}, 0.184 + 1.65\sqrt{0.0000846}\right) \approx (0.169, 0.199)$$

8. Тестируется гипотеза $H_0 : E(e^{-X_1}) = 0.2$ против альтернативы $H_0 : E(e^{-X_1}) < 0.2$.

   Запишем реализацию тестовой статистики:

$$T(x) = \frac{0.184 - 0.2}{\sqrt{0.0000846}} \approx -1.74$$

   Поскольку $-1.74 \in (-\infty, -1.65)$, то нулевая гипотеза отвергается на 5%-м уровне значимости.

   В качестве альтернативы можно рассчитать p-value:

$$\text{p-value} = \Phi(-1.74) \approx 0.04$$

   Поскольку $0.04 < 0.05$, то нулевая гипотеза отвергается на 5%-м уровне значимости.

Задача №2

На перекрестке автомобили, независимо друг от друга, поворачивают налево, поворачивают направо или едут вперед. По наблюдениям Лаврентия 35 машин проехали налево, 25 машин проехали направо и 40 машин проехали прямо. Все машины совершают поворот под углом 90 градусов (будем считать угол положительным независимо от того, совершается поворот направо или налево). Если же машина едет прямо, то угол ее поворота, очевидно, равняется нулю.

1. На 10%-м уровне значимости при помощи теста Хи-квадрат Пирсона протестируйте гипотезу о том, что каждая из трех возможных траекторий движения (прямо, налево или направо) автомобиля равновероятна. Выпишите нулевую и альтернативную гипотезы (в форме ограничений на параметры), критическую область теста и значение тестовой статистики. Сделайте вывод, сопоставив реализацию тестовой статистики с критической областью. **(3 балла)**

2. Повторите предыдущий пункт при помощи теста отношения правдоподобия (LR-тест). **(3 балла)**

3. Протестируйте на 10%-м уровне значимости гипотезу о том, что машины едут прямо в половине случаев, против альтернативы о том, что реже. Для осуществления вывода сопоставьте p-value теста с уровнем значимости. **(4 балла)**

4. На 10%-м уровне значимости протестируйте гипотезу о том, что математическое ожидание угла поворота проезжающего перекресток автомобиля равняется 50, против альтернативы о том, что меньше. **(5 баллов)**

Подсказка: Обозначим через $\chi^2_{q,d}$ квантиль уровня q Хи-квадра случайной величины с d степенями свободы. Известно, что $\chi^2_{0.9,2} \approx 4.6$, $\chi^2_{0.95,2} \approx 5.99$, $\chi^2_{0.975,2} \approx 7.38$, $\chi^2_{0.975,3} \approx 9.34$ и $\chi^2_{0.99,2} \approx 9.21$.
Значения логарифмов: $\ln(0.25) \approx -1.386$, $\ln(0.35) \approx -1.05$, $\ln(0.4) \approx -0.916$ и $\ln\left(\frac{1}{3}\right) \approx -1.099$

Решение:

1. Для удобства запишем информацию о реализации выборки в форме таблицы:

Траектория	налево	направо	прямо
Количество	35	25	40

Через p_1 и p_2 обозначим вероятности того, что машина поедет налево и направо соответственно. Всего в выборке содержится информация о $n = 35 + 25 + 40 = 100$ автомобилях.

Тестируется гипотеза $H_0 : p_1 = \frac{1}{3} \wedge p_2 = \frac{1}{3}$. Посчитаем реализацию тестовой статистики:

$$T(x) = \frac{\left(35 - \frac{100}{3}\right)^2}{\frac{100}{3}} + \frac{\left(25 - \frac{100}{3}\right)^2}{\frac{100}{3}} + \frac{\left(40 - \frac{100}{3}\right)^2}{\frac{100}{3}} = 3.5$$

Поскольку носитель распределения включает 3 элемента, то асимптотическое распределение тестовой статистики будет $\chi^2(3-1) = \chi^2(2)$. Следовательно, критическая область имеет вид $\mathcal{T}_{0.1} = (4.6, \infty)$. Так как $3.5 \notin (4.6, \infty)$, то нулевая гипотеза не отвергается на 10%-м уровне значимости.

2. Запишем логарифм функции правдоподобия:

$$\ln L(p_1, p_2) = 35\ln(p_1) + 25\ln(p_2) + 40\ln(1 - p_1 - p_2)$$

Максимизируя функцию правдоподобия получаем реализации ММП оценок:

$$\hat{p_1}(x) = 0.35, \qquad \hat{p_2}(x) = 0.25$$

Посчитаем логарифм правдоподобия в точке реализации ММП оценки

$$\ln L(0.35, 0.25) = 35\ln(0.35) + 25\ln(0.25) + 40\ln(1 - 0.25 - 0.35) \approx -108.053$$

Повторим расчет с учетом ограничений нулевой гипотезы:

$$\ln L\left(\frac{1}{3}, \frac{1}{3}\right) = 35\ln\left(\frac{1}{3}\right) + 25\ln\left(\frac{1}{3}\right) + 40\ln\left(\frac{1}{3}\right) \approx -109.861$$

Посчитаем реализацию тестовой статистики:

$$2 \times (-108.053 - (-109.861)) = 3.616$$

Поскольку нулевая гипотеза предполагает 2 ограничения, то при их соблюдении асимптотическое распределение тестовой статистики будет $\chi^2(2)$. Следовательно, критическая область имеет вид $\mathcal{T}_{0.1} = (4.6, \infty)$. Так как $3.616 \notin (4.6, \infty)$, то нулевая гипотеза не отвергается на 10%-м уровне значимости.

3. Обозначим через p вероятность того, что машина поедет прямо. Тестируется гипотеза $H_0 : p = 0.5$ против альтернативы $H_1 : p < 0.5$.

   Рассчитаем реализацию тестовой статистики:

$$T(x) = \frac{0.4 - 0.5}{\sqrt{\frac{0.5\times(1-0.5)}{100}}} = -2$$

   Посчитаем p-value с учетом того, что критическая область, в силу формы альтернативной гипотезы, является левосторонней:

$$\text{p-value} = \Phi(-2) \approx 0.023$$

   Поскольку $0.023 < 0.1$, то нулевая гипотеза отвергается на 10%-м уровне значимости.

4. У нас имеются $35+25 = 60$ наблюдений с 90 градусами и 40 наблюдений с 0 градусов. Посчитаем выборочное среднее:

$$\overline{x}_{100} = \frac{60 \times 90 + 40 \times 0}{100} = 54$$

   Найдем реализацию исправленной выборочной дисперсии:

$$\hat{\sigma}^2_{100} = \frac{60 \times (90-54)^2 + 40 \times (0-54)^2}{100-1} \approx 1963.636$$

   Посчитаем реализацию тестовой статистики:

$$T(x) = \frac{54 - 50}{\sqrt{\frac{1963.636}{100}}} \approx 0.9$$

   Посчитаем p-value с учетом того, что критическая область, в силу формы альтернативной гипотезы, является левосторонней:

$$\text{p-value} = \Phi(0.9) \approx 0.816$$

   Поскольку $0.816 > 0.1$, то нулевая гипотеза не отвергается на 10%-м уровне значимости.

Задача №3

Имеется выборка X_1, X_2 из нормального распределения, у которого математическое ожидание и дисперсия совпадают по модулю, то есть $Var(X_1) = |E(X_1)|$. Реализация выборки имеет вид $x = (2, -1)$. При помощи леммы Неймана-Пирсона тестируется гипотеза о том, что математическое ожидание наблюдения равняется 1 против альтернативы о том, что оно равняется -1.

1. Запишите тестовую статистику. **(1 балл)**
2. Найдите распределение тестовой статистики при условии верной нулевой гипотезы, а также при условии верной альтернативной гипотезы. **(5 баллов)**
3. Найдите критическую область теста для 10%-го уровня значимости. **(5 баллов)**
4. Сделайте вывод о том, отвергается ли нулевая гипотеза на 10%-м уровне значимости. **(1 балл)**
5. Посчитайте мощность выведенного теста при 10%-м уровне значимости. **(6 баллов)**

Решение:

1. Из условия известно, что $X_i \sim \mathcal{N}(\theta, |\theta|)$, где $i \in \{1, 2\}$.

   Запишем функцию плотности наблюдения:

   $$f_{X_i}(t) = \frac{1}{\sqrt{2\pi|\theta|}} e^{-\frac{(t-\theta)^2}{2|\theta|}}$$

   Тестируется гипотеза $H_0 : \theta = 1$ против альтернативы $H_1 : \theta = -1$. Запишем функцию правдоподобия для параметров, соответствующих этим гипотезам:

   $$L(1; X) = \frac{1}{2\pi} e^{-\frac{(X_1-1)^2-(X_2-1)^2}{2}}$$
   $$L(-1; X) = \frac{1}{2\pi} e^{-\frac{(X_1+1)^2-(X_2+1)^2}{2}}$$

   Выпишем тестовую статистику:

   $$T(X) = \frac{L(-1; X)}{L(1; X)} = e^{\frac{(X_1-1)^2+(X_2-1)^2}{2} - \frac{(X_1+1)^2+(X_2+1)^2}{2}} = e^{-2(X_1+X_2)}$$

2. Для удобства произведем монотонное преобразование над тестовой статистикой:

   $$T^*(X) = \ln(-0.5T(X)) = X_1 + X_2 \sim \mathcal{N}(2\theta, 2|\theta|)$$

   В результате получаем:

   $$T^*(X)|H_0 \sim \mathcal{N}(2, 2)$$
   $$T^*(X)|H_1 \sim \mathcal{N}(-2, 2)$$

3. Поскольку одно из преобразований было убывающим, то критическая область окажется левосторонней. Следовательно, необходимо найти квантиль уровня 0.1 распределения $T^*(X)|H_0$:

$$P\left(T^*(X) \leq q_{0.1}|H_0\right) = 0.1 \implies \Phi\left(\frac{q_{0.1} - 2}{\sqrt{2}}\right) = 0.1 \implies \frac{q_{0.1} - 2}{\sqrt{2}} \approx -1.28 \implies q_{0.1} \approx 0.19$$

   Отсюда получаем левостороннюю критическую область теста:

$$\mathcal{T}_{0.1} = (-\infty, 0.19)$$

4. Посчитаем реализацию тестовой статистики:

$$T(x) = 2 - 1 = 1$$

   Поскольку $1 \notin (-\infty, 0.19)$, то нулевая гипотеза не отвергается на 10%-м уровне значимости.

5. Найдем вероятность ошибки второго рода:

$$P(X_1 + X_2 > 0.19|H_1) = 1 - \Phi\left(\frac{0.19 + 2}{\sqrt{2}}\right) \approx 1 - 0.939 = 0.061$$

   Следовательно, мощность теста равняется $1 - 0.061 = 0.939$.