

## Метод наибольшего правдоподобия для дискретных случайных величин (МНП ДСВ)

Для ДСВ: есть реализация случайной выборки  $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ . Ищем значение параметра  $\theta$  (параметров), которое бы максимизировало вероятность одновременной реализации этих исходов.

$L(\theta) = P(X_1 = x_1)P(X_2 = x_2) \dots P(X_n = x_n) \rightarrow \max$  - функция правдоподобия.

$\ln(L(\theta))$  - логарифмическая функция правдоподобия.

Решаем задачу  $L(\theta) \rightarrow \max$  или  $\ln(L(\theta)) \rightarrow \max$ .

Нужные формулы с логарифмами:

(на области определения)  $\ln(a \cdot b) = \ln a + \ln b, \ln a^k = k \ln a, \ln e^k = k, (\ln x)' = \frac{1}{x}, (\ln f)' = \frac{1}{f} f'$

Для нахождения максимума/минимума функции надо взять производную, сравнить ее с нулем, сделать выводы.

Каждый раз следите за тем, чтобы переменные имели смысл – вероятность от 0 до 1, интенсивность от 0, число испытаний целое и т.д.

(основных задач 14, номера 11.15 – 11.18 необязательные, 11.19 и 11.20 дополнительные)

11.1	Количество покупателей в магазине подчиняется закону Пуассона с неизвестным параметром $\lambda$ . Первый продавец говорит, что в среднем в магазин приходит $\hat{\lambda}_1 = 6$ покупателей в час, второй – $\hat{\lambda}_2 = 7$ покупателей. Для того чтобы понять, чья оценка точнее и лучше оценивает неизвестный параметр из генеральной совокупности, продавцы решили взять выборку объема два, то есть посмотреть, сколько покупателей будет за два часа. За первый час их было 5, за второй – 8. а) Какая из оценок лучше – $\hat{\lambda}_1 = 6$ или $\hat{\lambda}_2 = 7$ ? б) С помощью МНП оценить $\lambda$ .
11.2	Студенты трех группы сдают экзамен, причем все работы они пишут независимо друг от друга, вероятность получения отличной оценки для всех одинакова и равна $p$ . В первой группе было получено 5 отличных оценок из 12, во второй 4 из 12, в третьей – 4 из 15. Методом наибольшего правдоподобия оценить $p$ . При решении задачи использовать логарифмическую функцию правдоподобия.
11.3	Студенты двух групп сдают экзамен, причем все работы они пишут независимо друг от друга. Вероятность получения "отлично" для всех студентов первой группы одинакова. Для второй группы вероятность тоже одинакова, но в два раза больше чем в первой. В первой группе было получено 2 "отлично" из 4, во второй 3 из 5. Найти оценку вероятности получения "отлично" для студента первой группы. Использовать логарифмическую функцию правдоподобия.
11.4	Случайные величины $X_1, X_2, X_3$ , распределенные по закону Пуассона с параметрами $\lambda, 2\lambda$ и $2\lambda$ соответственно, приняли значения $X_1 = 1, X_2 = 3, X_3 = 5$ . Оценить значение параметра $\lambda$ методом наибольшего правдоподобия. Использовать логарифмическую функцию правдоподобия.
11.5	Случайные величины $X_1, X_2, X_3$ , распределенные по закону Пуассона с параметрами $2\lambda, \lambda$ и $3\lambda$ соответственно, приняли значения $X_1 = 1, X_2 = 3, X_3 = 8$ . Оценить значение параметра $\lambda$ методом наибольшего правдоподобия. Использовать логарифмическую функцию правдоподобия.
11.6	В магазине работают три кассы. Количество покупателей, обслуживаемых каждым из кассиров, распределено по закону Пуассона. При этом первый из кассиров самый опытный – в среднем на обслуживание одного покупателя у него уходит в два раза меньше времени, чем у остальных продавцов. За время, которое мы наблюдали за кассирами, первый кассир обслужил 6 человек, второй – 3, третий – 4. Методом наибольшего правдоподобия оценить параметр $\lambda$ – интенсивность для первого кассира.
11.7	Студент А утверждает, что забрасывает мяч в корзину с вероятностью $p_1 = 0.3$ , студент Б с этим не согласен – по его мнению А попадает в корзину с $p_2 = 0.2$ . Для того чтобы оценить, чья оценка лучше, А проводит три серии бросков до первого попадания. В первой серии он попал с четвертого раза, во второй с пятого, в третьей со второго. Будем считать, что все броски независимы и вероятность постоянная. а) Чему равна вероятность реализации именно таких трех серий, если вероятность попасть в корзину при одном броске равна $p_1 = 0.3$ ? б) Чему равна вероятность реализации именно таких трех серий, если вероятность попасть в корзину при одном броске равна $p_2 = 0.2$ ? в) Какая из этих оценок лучше? г) Найти вероятность $p$ , при которой вероятность появления именно таких трех серий будет наибольшей.
11.8	Тренер и ученик стреляют в цель до первого попадания каждый. Известно, что тренер попадает в цель с вероятностью в два раза большей, чем ученик. Методом максимального правдоподобия оценить вероятность

	попадания учеником в цель при единичном выстреле, если известно, что тренер попал со второго раза, а ученик – с пятого. При решении задачи использовать логарифмическую функцию правдоподобия.
<b>МНП НСВ</b> Для НСВ с плотностью $f(\theta, x)$ , зависящей от параметра (параметров), функция правдоподобия равна $L(\theta) = f(\theta, x_1)f(\theta, x_2) \dots f(\theta, x_n)$ . Можно максимизировать ее, можно логарифмическую.	
<b>11.9</b>	Случайные величины $X_1, X_2, X_3$ , распределенные по показательному закону с параметрами $\lambda, 2\lambda$ и $3\lambda$ соответственно, приняли значения $X_1 = 4, X_2 = 2, X_3 = 3$ . Оценить значение параметра $\lambda$ методом наибольшего правдоподобия.
<b>11.10</b>	В магазине работают три кассы. Время обслуживания покупателей каждым из кассиров распределено по показательному закону. При этом первый из кассиров самый опытный – среднее время обслуживания покупателя у него в два раза меньше чем у оставшихся двух. Первый кассир обслужил очередного покупателя за минуту, второй – за две минуты, третий – за полторы. Методом наибольшего правдоподобия оценить параметр $\lambda$ – интенсивность для первого кассира.
<b>11.11</b>	Случайные величины $X_1 \sim N(m; 3^2)$ и $X_2 \sim N(m; 4^2)$ приняли значения 5 и 9 соответственно. Методом наибольшего правдоподобия оценить значение параметра $m$ .
<b>11.12</b>	Андрей и Борис независимо друг от друга играют в покер в интернете. Выигрыш каждого из них за день – это случайная величина, распределенная по нормальному закону, причем известно, что у них одинаковое матожидание выигрыша $m$ (тысяч рублей), но разные стандартные отклонения – у Андрея 1 а у Бориса 2 тысячи рублей. За последний день они выиграли по 2 и 3 тысячи рублей соответственно. Методом наибольшего правдоподобия оценить значение параметра $m$ . <i>(в условии сказано: случайные величины <math>X_1 \sim N(m; 1^2)</math> и <math>X_2 \sim N(m; 2^2)</math> приняли значения 2 и 3 соответственно)</i>
<b>11.13</b>	Случайные величины $X_1 \sim N(0; \sigma^2)$ и $X_2 \sim N(0; (3\sigma)^2)$ приняли значения 2 и 4 соответственно. Методом наибольшего правдоподобия оценить значение параметра $\sigma$ .
<b>11.14</b>	Случайные величины $X_1 \sim N(0; \sigma^2)$ и $X_2 \sim N(0; (2\sigma)^2)$ приняли значения 2 и 3 соответственно. Методом наибольшего правдоподобия оценить значение параметра $\sigma$ .
Задачи на сравнение качества точечных оценок ( <b>11.15-11.18</b> ) можно не решать, они более теоретические, имеют отношение к обоснованию того, что мы действительно используем «хорошие» формулы, и можно доказать, что они дают хороший (лучший из возможных) результат.	
<b>11.15</b>	Независимые случайные величины $X_1, X_2$ и $X_3$ имеют одинаковое матожидание $m$ но разные стандартные отклонения – $\sigma, 2\sigma$ и $3\sigma$ соответственно. В качестве оценки матожидания мы рассматриваем три варианта: $\hat{X}_1 = \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3, \hat{X}_2 = \frac{1}{6}X_1 + \frac{1}{5}X_2 + \frac{1}{4}X_3, \hat{X}_3 = \frac{1}{6}X_1 + \frac{1}{3}X_2 + \frac{1}{2}X_3$ . Какая из этих оценок лучше? <i>Указание – проверить несмещенность, у несмещенных сравнить дисперсии.</i>
<b>11.16</b>	Случайные величины $X_1$ и $X_2$ распределены по одному закону и независимы. Среди всех несмещенных оценок матожидания вида $c_1X_1 + c_2X_2$ найти оценку с наименьшей дисперсией.
<b>11.17</b>	Случайные величины $X_1$ и $X_2$ имеют одинаковое матожидание $m$ но разные стандартные отклонения – $\sigma$ и $2\sigma$ соответственно. Среди всех несмещенных оценок матожидания ГС вида $c_1X_1 + c_2X_2$ найти оценку с наименьшей дисперсией.
<b>11.18</b>	Независимые случайные величины $X_1$ и $X_2$ распределены по одному закону. Найти матожидание выражений а) $X_1^2 + X_2^2 - 2 \cdot \left(\frac{X_1+X_2}{2}\right)^2$ б) $\frac{X_1^2 + X_2^2 - 2 \cdot \left(\frac{X_1+X_2}{2}\right)^2}{2}$ в) какое из этих выражений является исправленной выборочной дисперсией выборки объема 2 и в чем мы убедились? (напомним, что $S^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$ )
<b>дополнительные задачи</b>	
<b>11.19</b>	Два приятеля договариваются о встрече. Первый из них изучал ТВиМС, второй постоянно опаздывает, причем время опоздания это случайная величина, распределенная равномерно на отрезке $[0; b]$ . Последние три раза опоздания составили 5 минут, 14 и 6 (этот приятель непредсказуем, время очередного опоздания не зависит от предыдущих опозданий). Помогите приятелю, изучавшему ТВиМС, оценить параметр $b$ методом наибольшего правдоподобия.
<b>11.20</b>	Рассмотрим две компании А и Б, состоящие двух офисов – центрального и регионального. Представители обеих компаний участвуют в отраслевой конференции и выступают в сессии, посвященной вопросам равноправия в отрасли. Представитель компании А сравнивает процент женщин в центральных офисах обеих компаний (в А он выше) и в региональных офисах (снова в А он выше), и из этих двух сравнений делается вывод о том, что в компании А женщины представлены лучше. Однако, представитель компании Б приводит данные не по отдельным офисам, а сразу по всей компании – своей и А, и оказывается, что в компании Б процент женщин выше. Может ли такое быть? Привести пример или обосновать невозможность. (это чисто математическая задача, без всяких фокусов)