Прикладная статистика в машинном обучении

Домашнее задание #1

Часть 1

Дедлайн: 19 ноября, 23:59 МСК

Правила игры

1. Домашнее задание состоит из двух частей. Часть 1 содержит 13 обязательных и две бонусных задачи и предполагает решение «от руки». Часть 2 содержит 3 обязательных задачи и предполагает программное решение.

2. Домашнее задание оценивается в 80 баллов. При этом часть 1 оценивается в 65 баллов, а часть 2 – в 15 баллов. По умолчанию за каждый пункт каждой задачи можно получить 1 балл. Однако за некоторые пункты некоторых задач можно получить другое количество баллов, которое явно указано в скобках рядом с меткой пункта.

3. Каждый пункт оценивается с промежутком 0.5. Например, если за пункт можно получить максимум 1 балл, то за полностью корректное решение ставится 1 балл, за решение с небольшими ошибками ставится 0.5 балла, за решение с серьёзными ошибками или неправильное решение ставится 0 баллов. Для пунктов, за которые можно получить максимум 2 балла, в зависимости от решения можно получить 2, 1.5, 1 и т.д. баллов. При этом пункты проверяются независимо друг от друга: если пункт $t+1$ зависит от численных результатов пункта t, и в пункте t допускается ошибка, из-за которой в пункт $t+1$ приходят неверные входные данные, то при корректном решении пункт $t+1$ оценивается в максимальное количество баллов, которое можно за него получить.

4. Бонусные задачи X и Y приведены в конце части 1 и обозначены значком †. Эти задачи необязательны к решению и учитываются сверх установленных 80 баллов. Баллы за корректно решённые бонусные задачи прибавляются к набранным баллам, даже если в сумме получается больше 80 баллов (оценка за домашнюю работу в этом случае будет больше 10, и так и будет внесена в таблицу с оценками).

5. Весь код должен быть написан на Python, R, C или C++.

6. Решения принимаются до **12 ноября 2021 года, 23:59 МСК** включительно. Работы, отправленные после дедлайна, проверяются, но **не оцениваются**.

7. Все решения нужно загрузить в личный репозиторий на GitHub Classroom.

8. Репозиторий должен содержать: PDF-файл с решениями задач части 1 и файл с кодом с решениями задач части 2. Решение задач части 1 можно набрать в любом электронном редакторе или написать от руки, а затем сделать качественный скан. Все решения должны быть расположены в правильном порядке в одном файле. Файлы должны быть названы по типу «name_surname_group_hw1_part1.pdf» и «name_surname_group_hw1_part2.ext», где вместо ext может быть .py, .ipynb, .R, .c, .cpp. Если решение части 2 разбивается на несколько файлов кода, то в репозиторий нужно загрузить все файлы, а в README.md подробно указать, что содержит каждый файл.

9. Разрешается использовать без доказательства любые результаты, встречавшиеся на лекциях или семинарах по курсу, если получение этих результатов не является вопросом задания. Разрешается использовать любые свободные источники с указанием ссылки на них.

10. Плагиат не допускается. При обнаружении случаев списывания, 0 за работу выставляется всем участникам нарушения, даже если можно установить, кто у кого списал.

Note: в заголовках задачи указана субъективная оценка выполнения, суммарно эта часть решена примерно на **40/65**.

Задача 1. Просто компания - решено на 10/10

Компания «Напиши-ка» производит три вида ручек: синие, красные и зелёные. Глава аналитического отдела компании Данил хочет понять, какая из ручек скорее всего «выстрелит», а какая не будет пользоваться успехом у покупателей. Для этого он анализирует выборку в 300 проданных ручек. Оказалось, что из них 150 синих, 100 красных и 50 зелёных ручек. Данил уверен, что ручки продаются независимо друг от друга, и вероятность того, что будет продана синяя ручка, равна p_1, а что красная p_2.

Ручка	С	К	З
N	150	100	50
P	p_1	p_2	$1 - p_1 - p_2$

[а] Обозначим $p = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$. Найдите $\hat{p}_{ML}$ интуитивно, не выписывая правдоподобие, и поясните, как вы это сделали.

В силу независимости продажи ручек кажется очевидным предположить, что вероятности купить ручку сообразны долям уже проданных ручек исходя из имею выборки

[б] Выпишите функцию правдоподобия и найдите $\hat{p}_{ML}$ как точку её глобального максимума.

$L(X|p) = p_1^{150} \cdot p_2^{100} \cdot (1 - p_1 - p_2)^{50}$

$l = lnL = 150lnp_1 + 100lnp_2 + 50ln(1 - p_1 - p_2) \to max_{p1,p2}$

$l'_{p1} = \frac{150}{p1} - \frac{50}{1-p_1-p_2}$

$\hat{p}_2 = \frac{2-2p_1}{3}$

$l'_{p2} = \frac{100}{p2} - \frac{50}{1-p_1-p_2}$

$4 \cdot \hat{p}_1 = 3 - 3\frac{2-2p_1}{3}$

$\hat{p}_1 = \frac{1}{2}, \hat{p}_2 = \frac{2-1}{3} = \frac{1}{3}$

Проверим, что найденные оценки действительно максимизируют функцию правдоподобия:

$l''_{p_1p_1} = \frac{-150}{p_1^2} - \frac{50}{(1-p_{1p}2)^2}$

$l''_{p_2p_2} = \frac{-100}{p_2^2} - \frac{50}{(1-p_{1p}2)^2}$

$l''_{p_1p_2} = -\frac{50}{(1-p_{1p}2)^2}$

Получаем такой гессиан:

$H(\hat{p}_{ML}) = \begin{pmatrix} -2400 & -1800 \\ -1800 & -2700 \end{pmatrix}$

$\Delta_1 = -2400 \ \Delta_2 = 2400 \cdot 2700 - 1800^2 > 0$

В соответствии с критерием Сильвестра имеет отрицательно определённую матрицу, что свидетельствует об успешном нахождении максимума функции правдоподобии, который полностью совпадает с интуитивной оценкой, данной в предыдущем пункте

[в] Проверьте гипотезу

$$\begin{cases} H_0 : p_1 = 0.2, \\ H_A : p_1 \neq 0.2 \end{cases}$$

на уровне значимости 5% при помощи тестов LR и LM.

$LR = 2(max_{p_1,p_2} l(p_1, p_2) - max_{p_2} l(p_1, p_2))$

$max_{p_1,p_2} l(p_1, p_2) = l(\hat{p}_1, \hat{p}_2) = 150ln0.5 + 100ln\frac{1}{3} + 50ln\frac{1}{6} = -303.42$

$max_{p_2} l(p_1 = 0.2, p_2) = l(p_1 = 0.2, \hat{p}_2) = 150ln0.2 + 100ln\frac{1}{3} + 50ln(0.8 - \frac{1}{3}) = -389.384$

$LR = 171.928$

Даже без подбора критического значения для LR-теста очевидно, что p-value нулевой гипотезы примерно равно 0, следовательно она отвергается на любом разумном уровне значимости

$LM = \hat{s}(p_1 = 0.2)^T \hat{I}_F(p_1 = 0.2)^{-1} \cdot \hat{s}(p_1 = 0.2)$

$\hat{s}(p_1 = 0.2) = \begin{pmatrix} \frac{150}{p1} - \frac{50}{1-p_1-p_2} \\ \frac{100}{p2} - \frac{50}{1-p_1-p_2} \end{pmatrix} = \begin{pmatrix} 642.857 \\ 192.857 \end{pmatrix}$

$\hat{I}_F(p_1 = 0.2) = E(-H) = \begin{pmatrix} 3979.36 & 229.36 \\ 229.358 & 1129.36 \end{pmatrix}$

$\hat{I}_F(p_1 = 0.2)^{-1} = \frac{1}{3979.36 \cdot 1129.36 - 229.36^2} \begin{pmatrix} 1129.36 & -229.36 \\ -229.358 & 3979.36 \end{pmatrix}$

Агрегируя получившиеся значение, можем получить наблюдаемое значение LM-статистики - $LM = 125.601$.

В случае LM, так же, как и в случае с LR-теста, нулевая гипотеза $p_1 = 0.2$ отвергается на любом разумном уровне значимости.

[г] Проверьте гипотезу

$$\begin{cases} H_0 : \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0.2 \end{pmatrix}, \\ \\ H_A : \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \neq \begin{pmatrix} 0.3 \\ 0.2 \end{pmatrix} \end{cases}$$

на уровне значимости 5% при помощи тестов LR и W.

Я не понимаю, куда у меня пропал листочек с этим пунктом...

[д] Постройте график логарифма правдоподобия в трёхмерной плоскости. Покажите на графике $\hat{p}_{ML}$ визуальную интерпретацию тестов LR и W для гипотезы из предыдущего пункта.

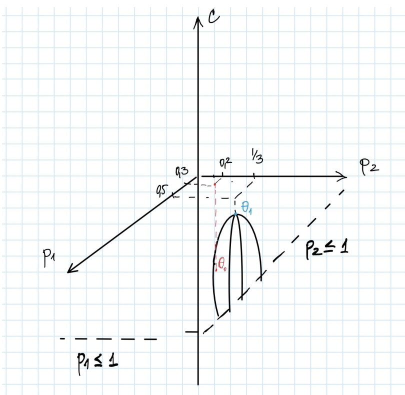


Рис. 1: — График логарифма правдоподобия в трёхмерной плоскости

[е] Постройте 95%-ый доверительный интервал для p_3.

[ж] Постройте 99%-ый доверительный интервал для $p_1 + p_2$.

[з] Постройте 90%-ый доверительный интервал для $\hat{p}_1$.

Подсказка: помните, что мы работаем в рамках частотного подхода.

[и] Приведите разумное интерпретируемое определение того, что ручка «выстрелила».

[к] Пользуясь определением из предыдущего пункта, сформулируйте гипотезу о том, что «выстрелит» ручка синего цвета и проверьте её при помощи любого из тестов LR, LM или W на уровне значимости 5%.

Решение оставшихся пунктов приведено на картинке ниже:

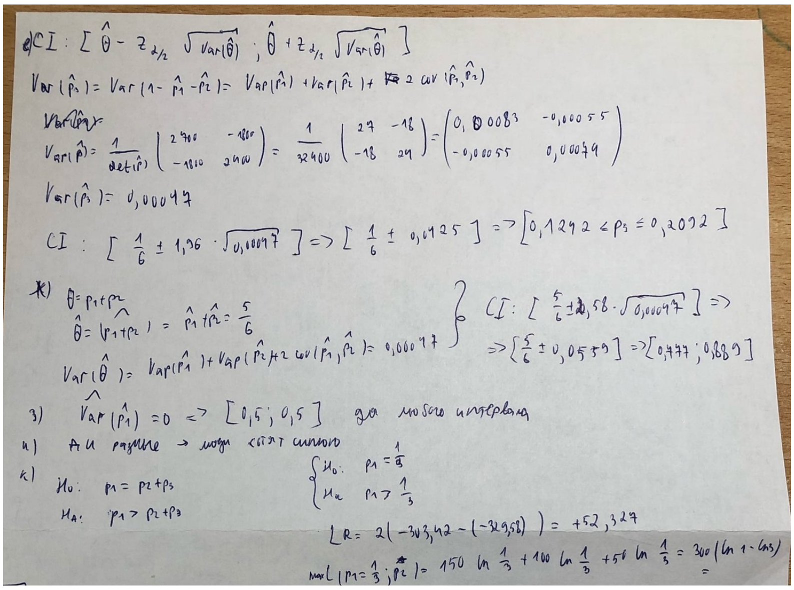


Рис. 2: — Решение второй части задачи 1

Задача 2. Анекдоточная - решено на 4/4

Станислав знает, что хороший анекдот должен быть не очень коротким, но и не слишком длинным. Время, за которое Станислав произносит один анекдот, – это непрерывная случайная величина с плотностью

$$f(x|b) = \begin{cases} \frac{2x}{b} e^{-\frac{x^2}{b}}, & \text{если } x > 0, \\ 0, & \text{иначе,} \end{cases}$$

где b – некоторый параметр. Станислав собрал случайную выборку по продолжительности рассказанных им анекдотов: $X_1, X_2, \ldots, X_n$, где $n = 10^6$. Оказалось, что $\sum X_i^2/n = 20$, $\sum X_i/n = 2$.

[а] Найдите $\hat{b}_{ML}$.

$\hat{b}_{ML} = \arg\max L(X|\theta)$

[б] Проверьте гипотезу

$$\begin{cases} H_0 : b = 3, \\ H_A : b \neq 3 \end{cases}$$

на уровне значимости 5% при помощи теста LR.

[в] Рассчитайте LM-статистику для проверки гипотезы

$$\begin{cases} H_0 : b = 1, \\ H_A : b \neq 1 \end{cases}$$

Чему приблизительно равно соответствующее p-value?

[г] Проверьте гипотезу из предыдущего пункта, построив соответствующий доверительный интервал для b.

Решение на листочке:

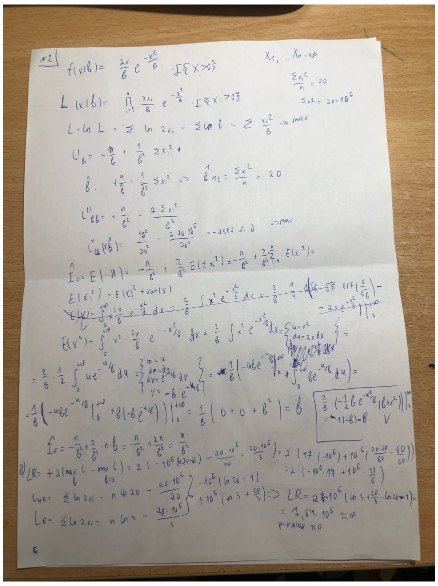


Рис. 3: — Решение первой части задачи 2

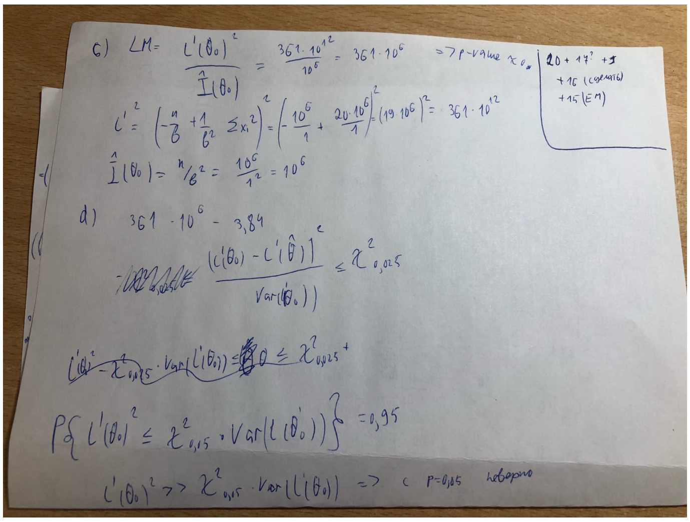


Рис. 4: — Решение второй части задачи 2

Задача 3. «Я не дерево. Я энт». - решено на 4/5

Исследователь Матвей подбрасывает монетку с вероятностью орла p до тех пор, пока не выпадет два орла (всего, не обязательно подряд). Он сыграл четыре игры, и оказалось, что первая завершилась за 3 хода, вторая – за 3 хода, третья – за 2 хода, четвёртая – за 4 хода. Будем считать, что подбрасывания в течение одной игры независимы. Также предположим, что игры происходили независимо друг от друга.

X	0, 1	2	3	4	-\\-	k
P	0	p^2	$C_1^2p^2(1-p)$	$C_1^3p^2(1-p)^2$	-\\-	$C_1^{k+1}p^2(1-p)^k-2$

Перед нами модифицированное геометрическое распределение, которое описывает число испытаний до k-ого успеха, где k=2 в нашем случае.

[a] **(2 балла)** Найдите $\hat{p}_{ML}$.

$L(X_1, X_2, X_3, X_4|p) = \prod((1-p)p^2)^2 \cdot (p^2)\ cdot((1-p)^2p^2) = (1-p)^4 \cdot p^8 \to max_p$

$l = lnL(X|p) = 4ln(1-p) + 8ln(p) \to max_p$

X	0, 1	2	3	4	-\\-	k
P	0	p^2	$2 \cdot p^2(1-p)$	$3 \cdot p^2(1-p)^2$	-\\-	$(k-1) \cdot p^2(1-p)^k - 2$

$l_p^{'} = -\frac{4}{1-p} + \frac{8}{p} = 0$

$\hat{p}_{ML} = \frac{2}{3}$

Проверим, что действительно нашли максимум:

$l_p^{''}p|_{p=\frac{2}{3}} = -\frac{4}{(1-p)^2} - \frac{8}{p^2}|_{p=\frac{2}{3}} = -4 \cdot 3^2 - \frac{8 \cdot 3^2}{4} = -36 - 18 = -54 > 0 \Rightarrow$ действительно максимум

[б] Найдите $\hat{a}_{ML}$ для нового параметра $a = (p^2 + 3p^3 - 1)$.

Вспомним свойство оценок максимального правдоподобия: $\widehat{g(\theta)}_{ML} = g(\widehat{\theta}_{ML})$. Таким образом, мы можем переписать исходное выражение так: $\hat{a}_{ML} = \hat{p}_{ML}^2 + 3\hat{p}_M L^3 - 1 = \frac{2^2}{3^2} + 3 \cdot \frac{2^3}{3^3} - 1 = \frac{4}{9} + \frac{8}{9} - 1 = \frac{1}{3}$

[в] **(2 балла)** Покажите, что $\hat{p}$ является состоятельной оценкой p.

Оценка параметра является состоятельной тогда и только тогда, когда сама оценка является как минимум асимптотически несмещённой, а дисперсия параметра стремится к 0 при увеличении выборки. Проверим выполнение данных свойств, предварительно выписав оценку максимального правдоподобия в общем виде:

$l = lnL(X|p) = \sum(x_i - 2)ln(1-p) + 2n \cdot ln(p) \to max_p$

$l_p^{'} = -\frac{\sum(x_i-2)}{1-p} + \frac{2n}{p} = 0$

$p = \frac{2n}{\sum x_i}$

$E(\hat{p}) = E\left(\frac{2 \cdot n}{\sum x_i}\right) = \frac{2n}{\sum E(x_i)} = \frac{2n \cdot p}{2n} = p \Rightarrow$ оценка ММП является несмещённой

$Var(\hat{p}) = Var\left(\frac{2 \cdot n}{\sum x_i}\right) = \frac{4n^2}{\sum Var(x_i)} = \frac{4n}{Var(x)}$

Непонятно, зачем нужна эта подсказка, если дисперсию всё равно надо искать, причём искать через начальные моменты:

$\mathrm{Var}(X) = E(X^2) - E(X)^2$

$E(X^2) = \sum_{i=2}^{\infty} p_i x_i^2 = \sum_{i=2}^{\infty}(i-1) \cdot p^2(1-p)^{i-2} \cdot i^2 = \frac{2(3-p)}{p^4}$

$\mathrm{Var}(X) = \frac{2(3-p)}{p^4} - \frac{4}{p^2} = \frac{2(3-p-2p^2)}{p^4}$

$\widehat{Var}(X) = \frac{81 \cdot 2(3 - \frac{2}{3} - 2\frac{4}{9}}{16} = 10.125 \cdot 1.444 = 14.625$

Так, получается, что оценка несостоятельна, что очевидным образом указывают на ошибку в расчёте, поскольку оценки ММП всегда являются состоятельными. Попробуем расписать дисперсию параметра по другому:

$Var(\hat{p}) = \frac{Var(2 \cdot n)}{Var(\sum x_i)} = \frac{2n}{\sum Var(x_i)} = \frac{2}{Var(x)}$ - так тоже не получается показать состоятельность оценки, поскольку дисперсия единичной реализации случайной величины не может зависеть от размера выборки. Не понимаю, как доказать это здесь.

Подсказка: для решения **Задачи X** потребуется доказать, что если M – число ходов, за которое завершится игра, то $\mathbb{E}(M) = \frac{2}{p}$. В этой задаче можно пользоваться этим утверждением без доказательства.

Задача 4. Полезное утверждение - не решено

Гарри никак не может понять, почему при большой информации Фишера оценки максимального правдоподобия лежат к истинному параметру ближе, чем при малой информации Фишера. Гермиона решает продемонстрировать аналитическую интуицию, стоящую за этим утверждением:

«Если взять выборку независимых одинаково распределённых случайных величин $Y_1, \ldots, Y_N$, каждая из которых имеет функцию плотности или функцию вероятности $f(y|\theta)$, и предположить, что выполнены все необходимые условия регулярности, то при $\phi \to \theta$:

$$D_{KL}\left[f(y|\theta) \| f(y|\phi)\right] = \frac{1}{2} I_f(\theta)(\phi - \theta)^2 + O((\phi - \theta)^3)».$$

[а] **(2 балла)** Докажите утверждение Гермионы либо для случая функций плотности, либо для случая функций вероятности.

[б] **(2 балла)** Поясните Гарри, почему при большей информации Фишера ML-оценки лежат ближе к истинному параметру.

Подсказка: $H(f) = -\mathbb{E}(\ln f)$, аналогично для кросс-энтропии.

(По мотивам: Williams, Weighing the Odds)

Задача 5. Модель для зелий - решено частично

Полумна хочет построить предсказательную модель, которая бы описывала зависимость популярности зелья y_i от силы его положительного влияния x_i. Обе величины являются количественными непрерывными переменными на $\mathbb{R}$. Предположим, что Полумна знает, как измерить популярность и силу влияния и верит, что искомая зависимость имеет следующий вид:

$$y_i = (\beta_1)^2 e^{-\beta_2 x_i} u_i,$$

где β_1 и β_2 – неизвестные положительные коэффициенты, u_i – случайная ошибка, причём $\ln u_i \sim \mathcal{N}(0, 2)$.

[а] **(2 балла)** Введите любые разумные ограничения на переменные. Найдите $\hat{\beta}_1$ и $\hat{\beta}_2$ методом максимального правдоподобия.

[б] **(2 балла)** Полумна собрала выборку, для которой оказалось, что

$$\sum_{i=1}^{n} \ln y_i = 100, \ \sum_{i=1}^{n} x_i = 50, \ \sum_{i=1}^{n} x_i \ln y_i = 200,$$

$$\sum_{i=1}^{n} x_i^2 = 2500, \ \sum_{i=1}^{n} (\ln y_i)^2 = 10000, \ \sum_{i=1}^{n} e^{-\hat{\beta}_2 x_i} = 1,$$

$$n = 500.$$

Проверьте гипотезу

$$\begin{cases} H_0 : \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \\ \\ H_A : \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \neq \begin{pmatrix} 1 \\ 2 \end{pmatrix} \end{cases}$$

на уровне значимости 5% при помощи теста W.

Подсказка: $\ln Y \sim N(\mu, \sigma^2) \Rightarrow \mathbb{E}(Y) = e^{\mu + \frac{\sigma^2}{2}}$.

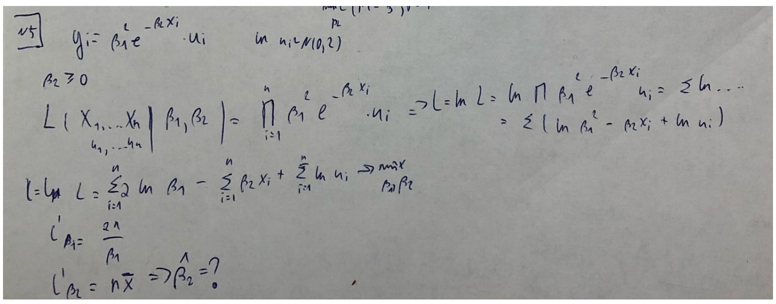


Рис. 5: — Попытка решения задачи 5

Задача 6. Функции правдоподобия - решено на 2.5/4

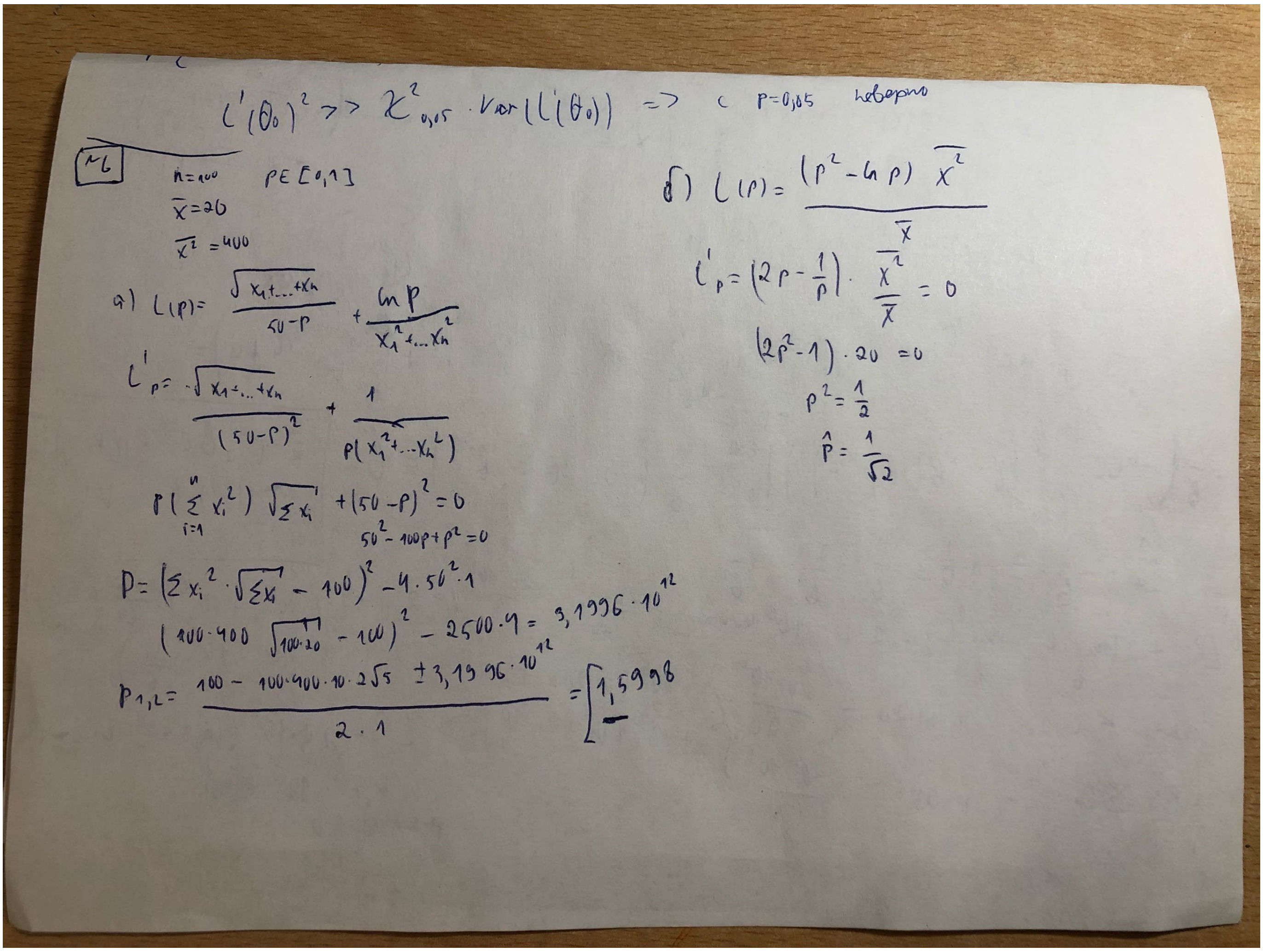
Пусть $X_1, \ldots, X_n$ – выборка независимых одинаково распределённых величин из распределения с параметром $p \in [0, 1]$. Известно, что $n = 100$, $\bar{X} = 20$, $(\bar{X^2}) = 400$. Найдите $\hat{p}_{ML}$ для следующих функций (можно либо вывести в явном виде, либо использовать математический анализ):

[а]

$$\ell(p) = \frac{\sqrt{X_1 + \ldots + X_n}}{50 - p} + \frac{\ln p}{X_1^2 + \ldots + X_n^2}.$$

[б]

$$\ell(p) = \frac{(p^2 - \ln p)\bar{X^2}}{\bar{X}}.$$



$L'(\theta_0)^2 >> \chi^2_{0,05} \cdot Var(L'(\theta_0)) \Rightarrow$ с $p=0,05$ неверно

№6

$n=100$ $p \in [0,1]$

$\bar{x}=20$

$\bar{x^2}=400$

а) $L(p) = \frac{\sqrt{x_1+\ldots+x_n}}{50-p} + \frac{\ln p}{x_1^2+\ldots x_n^2}$

$L'_p = \frac{\sqrt{x_1+\ldots+x_n}}{(50-p)^2} + \frac{1}{p(x_1^2+\ldots x_n^2)}$

$p(\sum_{i=1}^n x_i^2)\sqrt{\sum x_i} + (50-p)^2 = 0$

$50^2 - 100p + p^2 = 0$

$D = (\sum x_i^2 \cdot \sqrt{\sum x_i} - 100)^2 - 4 \cdot 50^2 \cdot 1$

$(400 \cdot 400 \sqrt{100 \cdot 20} - 100)^2 - 2500 \cdot 4 = 3,1996 \cdot 10^{12}$

$p_{1,2} = \frac{100 - 100 \cdot 400 \cdot 10 \cdot 2\sqrt{5} \pm 3,1996 \cdot 10^{12}}{2 \cdot 1} = \left[\begin{array}{l} 1,5998 \\ - \end{array}\right.$

б) $L(p) = \frac{(p^2 - \ln p)\bar{x^2}}{\bar{x}}$

$L'_p = (2p - \frac{1}{p}) \cdot \frac{\bar{x^2}}{\bar{x}} = 0$

$(2p^2 - 1) \cdot 20 = 0$

$p^2 = \frac{1}{2}$

$\hat{p} = \frac{1}{\sqrt{2}}$

Рис. 6: — Решение задачи 6

Задача 7. Дивергент - решено на 3.5/5

Рассмотрим распределения $p = \mathcal{N}(1, 2)$, $q = \text{Exp}(1)$, $r = \text{Bin}(3, 0.5)$. Для каждого пункта приведите математическое обоснование ответа.

[а] Найдите $D_{KL}(p \| q)$.

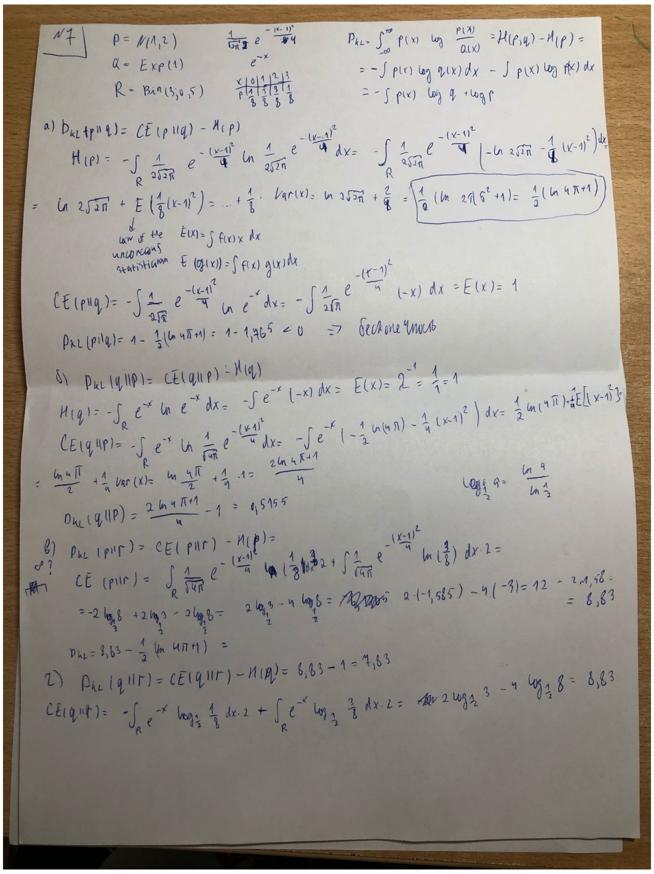
[б] Найдите $D_{KL}(q \| p)$.

[в] Найдите $D_{KL}(p \| r)$.

[г] Найдите $D_{KL}(q \| r)$.

[д] Возможно ли применить линейное преобразование к p, q или r так, чтобы ответ на хотя бы один из пунктов выше изменился?

Есть предположение, что дивергенция между P Q изменится, если прологарифмировать P и тем самым получить логнормальное распределение. Энтропия нормального и логнормального разная, но есть сомнение насчёт компенсации энтропии через кросс-энтропию с таким же знаком.



№7

$P = N(1,2)$ $\quad \frac{1}{\sqrt{2\pi}2} e^{-\frac{(x-1)^2}{4}}$

$Q = Exp(1)$ $\quad e^{-x}$

$R = Bin(3;0,5)$

x	0	1	2	3
p	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$D_{KL} = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{q(x)} = H(p,q) - H(p) =$
$= -\int p(x) \log q(x) dx - \int p(x) \log p(x) dx$
$= -\int p(x) \log q + \log p$

а) $D_{KL}(p\|q) = CE(p\|q) - H(p)$

$H(p) = -\int_R \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-1)^2}{4}} \ln \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-1)^2}{4}} dx = -\int_R \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-1)^2}{4}} \left(-\ln 2\sqrt{2\pi} - \frac{1}{8}(x-1)^2\right) dx =$

$= \ln 2\sqrt{2\pi} + E\left(\frac{1}{8}(x-1)^2\right) = \ldots + \frac{1}{8} \cdot Var(x) = \ln 2\sqrt{2\pi} + \frac{2}{8} = \boxed{\frac{1}{2}(\ln 2\pi\sigma^2 + 1) = \frac{1}{2}(\ln 4\pi + 1)}$

law of the unconscious statistician $\quad E(x) = \int f(x) x\, dx$ $\quad E(g(x)) = \int f(x) g(x) dx$

$CE(p\|q) = -\int \frac{1}{2\sqrt{\pi}} e^{-\frac{(x-1)^2}{4}} \ln e^{-x} dx = -\int \frac{1}{2\sqrt{\pi}} e^{-\frac{(x-1)^2}{4}} (-x) dx = E(x) = 1$

$D_{KL}(p\|q) = 1 - \frac{1}{2}(\ln 4\pi + 1) = 1 - 1{,}765 < 0 \Rightarrow$ бесконечность

б) $D_{KL}(q\|p) = CE(q\|p) - H(q)$

$H(q) = -\int_R e^{-x} \ln e^{-x} dx = -\int e^{-x}(-x) dx = E(x) = 2^{-1} = \frac{1}{1} = 1$

$CE(q\|p) = -\int_R e^{-x} \ln \frac{1}{\sqrt{4\pi}} e^{-\frac{(x-1)^2}{4}} dx = -\int e^{-x}\left(-\frac{1}{2}\ln(4\pi) - \frac{1}{4}(x-1)^2\right) dx = \frac{1}{2}\ln(4\pi) + \frac{1}{4}E[(x-1)^2] =$

$= \frac{\ln 4\pi}{2} + \frac{1}{4} Var(x) = \ln \frac{4\pi}{2} + \frac{1}{4} \cdot 1 = \frac{2\ln 4\pi + 1}{4}$ $\qquad \log_{\frac{1}{2}} 9 = \frac{\ln 9}{\ln \frac{1}{2}}$

$D_{KL}(q\|p) = \frac{2\ln 4\pi + 1}{4} - 1 = 0{,}5155$

в) $D_{KL}(p\|r) = CE(p\|r) - H(p) =$

$CE(p\|r) = \int_R \frac{1}{\sqrt{4\pi}} e^{-\frac{(x-1)^2}{4}} \ln\left(\frac{1}{8}\right) dx \cdot 2 + \int \frac{1}{\sqrt{4\pi}} e^{-\frac{(x-1)^2}{4}} \ln\left(\frac{3}{8}\right) dx \cdot 2 =$

$= -2\log_{\frac{1}{2}} 8 + 2\log_{\frac{1}{2}} 3 - 2\log_{\frac{1}{2}} 8 = 2\log_{\frac{1}{2}} 3 - 4\log_{\frac{1}{2}} 8 = 2 \cdot (-1{,}585) - 4 \cdot (-3) = 12 - 2 \cdot 1{,}58 = 8{,}83$

$D_{KL} = 8{,}83 - \frac{1}{2}(\ln 4\pi + 1) =$

г) $D_{KL}(q\|r) = CE(q\|r) - H(q) = 8{,}83 - 1 = 7{,}83$

$CE(q\|r) = -\int_R e^{-x} \log_{\frac{1}{2}} \frac{1}{8} dx \cdot 2 + \int_R e^{-x} \log_{\frac{1}{2}} \frac{3}{8} dx \cdot 2 = 2\log_{\frac{1}{2}} 3 - 4\log_{\frac{1}{2}} 8 = 8{,}83$

Рис. 7: — Решение задачи 7

Задача 8. Между молотом и наковальней - решено на 4/4

Одной из симметричных альтернатив KL-дивергенции является *взаимная информация*: для случайных величин X и Y она определяется как

$$I(X,Y) = H(X) - H(X|Y),$$

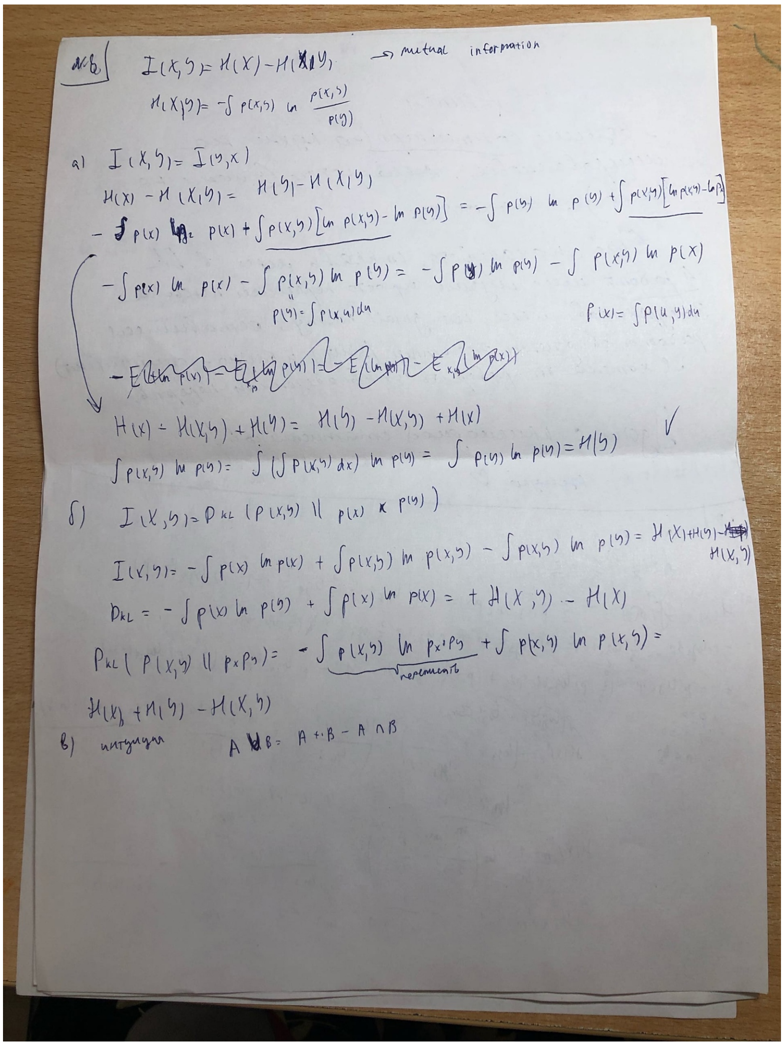
где $H(X|Y) = -\int p(x,y)\ln\dfrac{p(x,y)}{p(y)}$.

[а] Покажите, что $I(X,Y) = I(Y,X)$.

[б] **(2 балла)** Покажите, что $I(X,Y) = D_{KL}(p(x,y)\|p(x)\times p(y))$.

[в] Поясните интуитивную интерпретацию $I(X,Y)$.

Взаимная информация является отражением того количества информации, которое мы могли бы узнать о распределении Y, если бы изучали только распределение X. Оно помогает найти взаимозависимости в данных. Очевидно, если X и Y абсолютно независимы, то количество взаимной информации будет равняться 0.



№8 $I(X,Y) = H(X) - H(X|Y)$ → mutual information

$H(X|Y) = -\int p(x,y) \ln \frac{p(x,y)}{p(y)}$

а) $I(X,Y) = I(Y,X)$

$H(X) - H(X|Y) = H(Y) - H(Y|X)$

$-\int p(x) \ln p(x) + \int p(x,y)[\ln p(x,y) - \ln p(y)] = -\int p(y) \ln p(y) + \int p(x,y)[\ln p(x,y) - \ln p(x)]$

$-\int p(x) \ln p(x) - \int p(x,y) \ln p(y) = -\int p(y) \ln p(y) - \int p(x,y) \ln p(x)$

$p(y) = \int p(x,y)dx$ $p(x) = \int p(x,y)dy$

$H(x) - H(x,y) + H(y) = H(y) - H(x,y) + H(x)$ ✓

$\int p(x,y) \ln p(y) = \int (\int p(x,y)dx) \ln p(y) = \int p(y) \ln p(y) = H(y)$

б) $I(X,Y) = D_{KL}(p(x,y) \| p(x) \times p(y))$

$I(X,Y) = -\int p(x) \ln p(x) + \int p(x,y) \ln p(x,y) - \int p(x,y) \ln p(y) = H(X) + H(Y) - H(X,Y)$

$D_{KL} = -\int p(x) \ln p(y) + \int p(x) \ln p(x) = + H(X,Y) - H(X)$

$D_{KL}(p(x,y) \| p_x p_y) = -\underbrace{\int p(x,y) \ln p_x p_y}_{\text{перемножить}} + \int p(x,y) \ln p(x,y) =$

$H(x) + H(y) - H(x,y)$

в) интуиция $A \cup B = A + B - A \cap B$

Рис. 8: — Решение задачи 8

Задача 9. Хорошая задача на экзамен - решено на 2/2

Случайная величина X принимает значение 0 с вероятностью p, значение 1 с вероятностью $1/3$ и значение 2 с вероятностью $2/3 - p$.

[а] Постройте график зависимости $H(X)$ как функцию от p

$$H(X) = -\sum_{i=1}^{3}[p(x)\cdot\log_3(p(x))] = -p\log_3(p) - \frac{1}{3}\log_3\frac{1}{3} - \left(\frac{2}{3}-p\right)\log_3\left(\frac{2}{3}-p\right) =$$

$$= -p\log_3(p) - \left(\frac{2}{3}-p\right)\log_3\left(\frac{2}{3}-p\right) + \left(p+\frac{1}{3}+\frac{2}{3}-p\right) = 1 - p\log_3(p) - \left(\frac{2}{3}-p\right)\log_3\left(\frac{2}{3}-p\right)$$

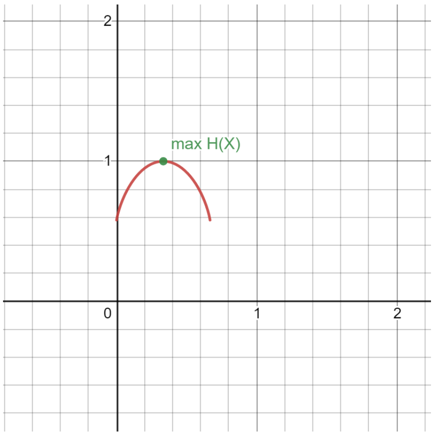


Рис. 9: — График зависимости энтропии от значений параметра p, построенный в Desmos

[б] При каком p энтропия будет максимальна? Поясните полученный результат.

Интуитивно очевидно, что при $p = \frac{1}{3}$ энтропия будет максимально, поскольку тогда при реализации выборки с каждым новым наблюдением мы будем получать наименьшее количество информации. Проверим:

$$H(X)'_p = -\log_3(p) - \tfrac{p}{p\ln 3} + \tfrac{\ln(2-3p)+1}{\ln 3} = \tfrac{\ln(2-3p)-\ln(3p)}{\ln 3}$$

$$\ln(2-3p) - \ln(3p) = 0$$

$$\hat{p} = \frac{1}{3}$$

Проверим, что действительно нашли максимум:

$$H(X)''_{pp}|_{p=1/3} = \frac{1}{\ln 3}\cdot\left(\frac{-3}{2-3p} - \frac{1}{p}\right)|_{p=1/3} = -\frac{2}{\ln 3} < 0$$

Итак, предположение, подтвердилось — $\arg\max_p H(X) = \frac{1}{3}$

Задача 10. Порисуем!

Рассмотрим модель множественной регрессии $y = X\beta + u$, которая оценивается при помощи МНК. Число наблюдений равно $n = 400$, число регрессоров равно $k = 10$, включая константный. Все регрессоры ортогональны друг другу.

[а] Долорес Амбридж строит регрессию по константному и следующим за ним четырём регрессорам. Корнелиус Фадж строит регрессию по константному и оставшимся пяти регрессорам. Покажите на *единой* картинке МНК $\hat{y}$, TSS, ESS, RSS и R^2 в их регрессиях.

[б] Альбус Дамблдор строит регрессию по всем 10 регрессорам. Покажите на той же картинке МНК $\hat{y}$, TSS, ESS, RSS и R^2 в его регрессии.

[в] **(2 балла)** Гарри Поттер хочет сравнить регрессии Амбридж и Дамблдора при помощи F-теста. Напомним, что

$$F = \frac{(RSS_R - RSS_{UR})/(k_{UR} - k_R)}{RSS_{UR}/(n - k_{UR})}.$$

Покажите на картинке МНК RSS_R, RSS_{UR} и угол, квадрату тангенса которого пропорциональна F-статистика.

[г] Приведите геометрическую интерпретацию F-теста.

Задача 11. Подпространства - почти не решено

Рассмотрим пространство $\mathbb{R}^3$ и два подпространства в нём

$$W = \{(x_1, x_2, x_3) | 3x_1 + 2x_2 - x_3 = 0\}$$

и

$$V = \mathrm{Lin}[(1, 1, 1)^T].$$

[а] Найдите $\dim V$, $\dim W$, $\dim(V \cap W)$, $\dim V^{\perp}$, $\dim W^{\perp}$. $dimV = 3$, так как это линейное подпространство $dimW = 2$, поскольку есть линейная зависимость между значениями вектора (x_1, x_2, x_3) $\dim(V \cap W) = 2$ $\dim V^{\perp} = 2$ $\dim W^{\perp} = 3$

[б] Найдите проекцию произвольного вектора u на V, W, $V \cap W$, $V^{\perp}$, $W^{\perp}$. Найдите квадрат длины каждой проекции.

[в] Как распределён квадрат длины проекции в каждом случае, если дополнительно известно, что вектор u имеет многомерное стандартное нормальное распределение?

Задача 12. Парная регрессия - решено на 5/5

Исследователь Борис работает с обычной парной регрессией

$$y_i = \beta_0 + \beta_1 X_i + u_i,$$

которую он оценивает при помощи МНК:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i.$$

[а] Просто для удобства выпишите RSS в этой регрессии и условия первого порядка в задаче минимизации.

$RSS = \sum(y_i - \hat{y}_i) = \sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \to min_{\hat{\beta}_0, \hat{\beta}_1}$

$RSS'_{\hat{\beta}_1} = -2\sum\left(x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)\right) \Rightarrow \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 = \sum x_i y_i$

$RSS'_{\hat{\beta}_0} = -2\sum\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) \Rightarrow n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i = \sum y_i$

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum(x_i \cdot y_i) - \frac{1}{n}\sum x_i \cdot \frac{1}{n}\sum y_i}{\frac{1}{n}\sum(x_i - \frac{1}{n}\sum x_i)^2}$$

$$\hat{\beta}_0 = \frac{1}{n}\sum y_i - \hat{\beta}_1 \frac{1}{n}\sum x_i$$

[б] Докажите, что $\sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$.

$\sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 x_i) = n \cdot (\frac{1}{n}\sum y_i - \hat{\beta}_1 \frac{1}{n}\sum x_i) + \hat{\beta}_1 \sum x_i = \sum y_i \Rightarrow \sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$

[в] Докажите, что $\bar{y} = \bar{\hat{y}}$.

По аналогии с предыдущим пунктом:
$\sum_{i=1}^{n}(y_i - \hat{y}_i) = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \bar{y} - \bar{\hat{y}} = 0$
(от добавления константы в обе части уравнения равенство между суммой истинных и оценённых значений целевой переменной не перестаёт быть верным)

[г] Докажите, что точка $(\bar{x}, \bar{y})$ лежит на линии оценённой регрессии.

Очевидно, чтобы доказать данное суждение, необходимо доказать справедливость следующего равенства:

$$\bar{\hat{y}} = \hat{\beta}_0 + \hat{\beta}_1\bar{x}$$

Отметим, что оценка среднего очевидно равняется среднему от оценок, поскольку оценка константы в МНК (как и везде, наверное) равна самой константе. Перепишем средние по определению, а также заменим $\hat{\beta}_0$ полным выражением из условия первого порядка:

$$\frac{1}{n}\sum\hat{y}_i = \frac{1}{n}\sum y_i - \hat{\beta}_1\frac{1}{n}\sum x_i + \hat{\beta}_1\frac{1}{n}\sum x_i$$

Теперь очевидно, что доказательство данного выражения сводится к предыдущему пункту:

$$\frac{1}{n}\sum\hat{y}_i = \frac{1}{n}\sum y_i$$

Справедливость данного утверждения уже была показана ранее.

[д] Докажите, что $\sum_{i=1}^{n} x_i(y_i - \hat{y}_i) = 0$.

$$\sum_{i=1}^{n} x_i(y_i - \hat{y}_i) = \sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \sum_{i=1}^{n} x_i(y_i - \frac{1}{n}\sum y_i - \hat{\beta}_1\frac{1}{n}\sum x_i - \hat{\beta}_1 x_i) =$$
$$= \sum_{i=1}^{n} x_i(y_i - \overline{y} - \hat{\beta}_1(x_i - \overline{x})) = \sum_{i=1}^{n} x_i\left((y_i - \overline{y}) - \frac{\frac{1}{n}\sum(x_i\cdot y_i) - \frac{1}{n}\sum x_i\cdot\frac{1}{n}\sum y_i}{\frac{1}{n}\sum(x_i - \frac{1}{n}\sum x_i)^2}\cdot(x_i - \overline{x})\right)$$

Поработаем с выражением рядом с $\hat{\beta}_1$ отдельно, применив свойства эмпирической оценки дисперсии для знаменателя оценки $\hat{\beta}_1$:

$$\sum_{i=1}^{n} x_i\cdot(x_i - \overline{x})\frac{\frac{1}{n}\sum(x_i\cdot y_i) - \frac{1}{n}\sum x_i\cdot\frac{1}{n}\sum y_i}{\frac{1}{n}\sum(x_i - \frac{1}{n}\sum x_i)^2} = n(\frac{1}{n}\sum(x_i\cdot y_i) - \frac{1}{n}\sum x_i\cdot\frac{1}{n}\sum y_i)\cdot\frac{\sum_{i=1}^{n} x_i\cdot(x_i - \overline{x})}{\sum(x_i - \frac{1}{n}\sum x_i)^2} =$$
$$= n(\frac{1}{n}\sum(x_i\cdot y_i) - \frac{1}{n}\sum x_i\cdot\frac{1}{n}\sum y_i)\cdot\frac{\sum x_i^2 - \overline{x}\sum x_i}{\sum x_i^2 - \overline{x}\sum x_i} = n(\overline{xy} - \overline{x}\cdot\overline{y})$$

В итоге получаем:

$$\sum_{i=1}^{n} x_i(y_i - \overline{y} - n(\overline{xy} - \overline{x}\cdot\overline{y}) = \sum(x_i\cdot y_i) - \frac{1}{n}\sum x_i\cdot\overline{y} - n(\frac{1}{n}\sum(x_i\cdot y_i) - \frac{1}{n}\sum x_i\cdot\frac{1}{n}\sum y_i) = 0$$

Задача 13. Гипотезы в линейной регрессии - решено на 12/12

Линейная регрессионная модель задаётся в следующем виде:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i.$$

Предположим, что $u \sim \mathcal{N}(0, \sigma^2 I)$. Известно, что

$$X = \begin{pmatrix} 1 & 1 & 3.1 \\ 1 & 12 & 2.2 \\ 1 & -3 & 0.1 \\ 1 & 2 & 0.5 \\ 1 & 0 & 11.3 \end{pmatrix}, y = \begin{pmatrix} 1.1 \\ 2.5 \\ 2.2 \\ 4 \\ 1 \end{pmatrix}$$

В процессе решения используйте калькулятор, все числа округляйте до сотых.

[а] Найдите $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$.

$\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2] = (X^T X)^{-1} X^T Y = [2.67, 0.04, -0.17]$

[б] Найдите $\hat{y}$.

$\hat{y} = X\beta = [2.17, 2.72, 2.54, 2.65, 0.72]$

[в] Найдите TSS, ESS, RSS и R^2.

$TSS = \sum(y_i - \overline{y})^2 = 5.97$

$ESS = \sum(\hat{y}_i - \overline{\hat{y}_i})^2 = 2.77$

$RSS = \sum(y_i - \hat{y}_i)^2 = 3.20$

$R^2 = \frac{ESS}{TSS} = 0.46$

[г] Найдите $\hat{\sigma}$.

$\hat{\sigma} = \frac{RSS}{n-k} = \frac{3.20}{5-3} = 1.60$

[д] Найдите $\widehat{\mathrm{Var}}(\hat{\beta})$.

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \hat{\sigma}^2 \cdot (X^T X)^{-1} = \begin{pmatrix} 0.65 & -0.04 & -0.07 \\ -0.04 & 0.01 & 0.00 \\ -0.07 & 0.00 & 0.02 \end{pmatrix}$$

[е] На уровне значимости 5% проверьте гипотезу

$$\begin{cases} H_0 : \beta_1 = 1, \\ H_1 : \beta_1 \neq 1. \end{cases}$$

$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1)}} = \frac{0.04-1}{0.01} = -76.92$

$t_{cr} = St(\nu = 5-3 = 2, \alpha = 0.05) = 4.30$

$|t| > t_{cr} \Rightarrow$ гипотеза отвергается, с вероятностью ошибки первого рода $\alpha = 0.05$ $\beta_1 ! = 1$

[ж] На уровне значимости 10% проверьте гипотезу

$$\begin{cases} H_0 : \beta_2 = 1, \\ H_1 : \beta_2 < 1. \end{cases}$$

$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1)}} = \frac{-0.17-1}{0.02} = -60.26$

$t_{cr} = St(\nu = 5-3 = 2, \alpha = 0.1) = 1.88$

$t < t_{cr} \Rightarrow$ гипотеза не отвергается, с вероятностью ошибки первого рода $\alpha = 0.1$ $\beta_2 < 1$. Это связано с тем, что у нас левосторонняя критическая область, в которой как раз лежит наблюдаемое значение оценки коэффициента.

[з] Проверьте регрессию на значимость в целом.

$X_r = [\overline{1}], \beta_r = \overline{Y} = 2.16, \hat{Y}_r = [\overline{2.16}], RSS_r = 5.97 = TSS$

$F_{obs} = \frac{RSS_r - RSS_{ur}}{k_{UR} - k_R} / \frac{RSS_{UR}}{n - k_{UR}} = \frac{ESS_{UR}}{k_{UR} - k_R} / \frac{RSS_{UR}}{n - k_{UR}} = \frac{2.77}{3-1} / \frac{3.20}{5-3} = 0.87$

$F_{cr} = F(\alpha = 0.05, \nu_1 = 2, \nu_2 = 2) = 18.99$

$p - value_{F_{obs}} = F^{-1}(x = 0.87, \nu_1 = 2, \nu_2 = 2) = 0.54$

Гипотеза о незначимости регрессионной модели не отвергается на любом разумном уровне значимости, т.е. модель не значима

[и] На уровне значимости 5% проверьте гипотезу

$$\begin{cases} H_0 : \beta_1 = \beta_2, \\ H_1 : \beta_1 \neq \beta_2. \end{cases}$$

$X_r = [\bar{1}, X_1 + X_2]$, $\beta_r = [2.41, -0.04]$, $\hat{Y}_r = [2.23, 1.80, 2.53, 2.30, 1.93]$, $RSS_r = 5.62$

$F_{obs} = \frac{RSS_r - RSS_{ur}}{k_{UR} - k_R} / \frac{RSS_{UR}}{n - k_{UR}} = \frac{5.62 - 3.2}{3 - 2} / \frac{3.20}{5 - 3} = 0.76$

$F_{cr} = F(\alpha = 0.05, \nu_1 = 1, \nu_2 = 2) = 18.51$

$p - value_{F_{obs}} = F^{-1}(x = 0.87, \nu_1 = 2, \nu_2 = 2) = 0.48$

Гипотеза о равенстве коэффициентов при X_1, X_2 не отвергается на любом разумном уровне значимости, значит равенство коэффициентов вполне вероятно.

[к] Постройте 95%-ый доверительный интервал для β_1.

$\beta_1 \in [\hat{\beta}_1 - t \cdot \sqrt{\widehat{\text{Var}}(\hat{\beta}_1)}; \hat{\beta}_1 + t \cdot \sqrt{\widehat{\text{Var}}(\hat{\beta}_1)}]$

$\beta_1 \in [0.04 - 4.3 \cdot 0.02; 0.04 + 4.3 \cdot 0.02]$

$\beta_1 \in [-0.05; 0.13]$

[л] Пусть $x_{1,6} = 10$, $x_{2,6} = 7$. Найдите $\hat{y}_6$.

$\hat{y}_6 = \hat{\beta}_0 + \hat{\beta}_1 x_{1,6} + \hat{\beta}_2 x_{2,6} = 2.67 + 0.04 \cdot 10 - 0.17 \cdot 7 = 1.88$

[м] Постройте 95%-ый доверительный интервал для $\mathbb{E}(y_6 | x_{1,6}, x_{2,6})$.

$\mathbb{E}(y_6 | x_{1,6}, x_{2,6}) \in [\hat{y}_{new} - se(\hat{y}_{new} \cdot t_{cr}; \hat{y}_{new} - se(\hat{y}_{new} \cdot t_{cr}]$

$$\begin{aligned} \widehat{\text{Var}}(\hat{y}_{new}) = \\ = \widehat{\text{Var}}(\hat{\beta}_0 + 10\hat{\beta}_1 + 7\hat{\beta}_2) = \widehat{\text{Var}}(\hat{\beta}_0) + 100\widehat{\text{Var}}(\hat{\beta}_1) + 49\widehat{\text{Var}}(\hat{\beta}_2) + \\ +2 \cdot 10 \cdot \widehat{cov}(\hat{\beta}_0, \hat{\beta}_1) + 2 \cdot 7 \cdot \widehat{cov}(\hat{\beta}_0, \hat{\beta}_2) + 2 \cdot 7 \cdot 10 \cdot \widehat{cov}(\hat{\beta}_1, \hat{\beta}_2) = \\ = 0.65^{0.5} + 100 \cdot 0.01^{0.5} + 49 \cdot 0.02^{0.5} + 20 \cdot (-0.04) + 14 \cdot (-0.07) + 140 \cdot (0.00) = \\ = 17.74 - 1.54 = 17.20 \end{aligned}$$

Задача X†. Методы моментов и первого шага - решено на 1/7

Альтернативой методу максимального правдоподобия является *метод моментов*, суть которого заключается в том, чтобы приравнять теоретические моменты как функции от оцениваемых параметров к их выборочным аналогам, и из полученной системы найти оценки.

[а] Пункт для тренировки. Рассмотрим выборку X_1, X_2, $X_3 \sim i.i.d.\ \mathcal{N}(\mu, 1)$. Оказалось, что $X_1 = 1$, $X_2 = 2$, $X_3 = 3$. Найдите $\hat{\mathbb{E}}(X_1)_{MM}$.

Формализуем суть метода моментов для первого момента:

$E(X) = \overline{X}$

Подставим необходимое:

$E(X) = \mu, \overline{x} = \frac{1+2+3}{3} = 2 \Rightarrow \hat{\mu}_{MM} = 2$

[б] **(6 баллов)** Исследователь Матвей подбрасывает монетку с вероятностью орла p до тех пор, пока не выпадет два орла (всего, не обязательно подряд). Оказалось, что среднее число ходов, за которое завершится игра, равно 40. Найдите $\hat{p}_{MM}$.

Подсказка: докажите, что если M – число ходов, за которое завершится игра, то $\mathbb{E}(M) = \frac{2}{p}$.

Задача Y†. Известное неравенство - очень интересно, но не решено

(6 баллов)

Рассмотрим линейную модель

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + u_i,$$

где $u_i \sim \mathcal{N}(0, 1)$, а все предпосылки ТГМ выполнены. Исследователь Вадим тестирует гипотезу вида

$$\begin{cases} H_0 : \beta_1 = C \\ H_A : \beta_1 \neq C, \end{cases}$$

где C – некоторая константа, при помощи тестов LR, LM и W. Докажите, что в такой постановке всегда верно, что $LM \leqslant LR \leqslant W$.