1 домашнее задание по ПСМО 21/22

Антон Антонов

10 декабря 2021 г.

1.

1.a.

$$y = \begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{3} \end{bmatrix} \tag{1}$$

Данные предположения логичны, так как у нас есть 300 исходов разных событий, логично предположить, что метод максимального правдоподобия придет к результату, что лучшая оценка - такая же вероятность, как и в 300 реализовавшихся исходах.

1.b.

L = $\mathrm{p}_1^{150} p_2^{100}(1-p_1-p_2)^{50}$
$l = 150log(p_1) + 100log(p_2) + 50log(1-p_1-p_2)$
$l'_{p_1} = \frac{150}{p_1} - \frac{50}{1-p_1-p_2}$
$\frac{150}{\hat{p}_1} - \frac{50}{1-\hat{p}_1-\hat{p}_2} = 0$
$\hat{p}_1 = \frac{3-3\hat{p}_2}{4}$ (1)
$l'_{p_2} = \frac{100}{p_2} - \frac{50}{1-p_1-p_2}$
$\frac{100}{\hat{p}_2} - \frac{50}{1-\hat{p}_1-\hat{p}_2} = 0$
$\hat{p}_2 = \frac{2-2\hat{p}_1}{3}$ (2)
Подставим значения из первого уравнения во второе
(1) $\rightarrow$ (2) $\hat{p}_{1ML} = \frac{1}{2}$ $\hat{p}_{2ML} = \frac{1}{3}$

1.c.

$$
\begin{aligned}
&l = 150log(p_1) + 100log(p_2) + 50log(1 - p_1 - p_2)\\
&l_R = 150log(0.2) + 100log(p_2) + 50log(0.8 - p_2)\\
&l'_R = \frac{100}{p_2} - \frac{50}{0.8 - p_2}\\
&\frac{100}{\hat{p}_2} - \frac{50}{0.8 - \hat{p}_2} = 0\\
&\hat{p}_2 = \frac{80}{150}
\end{aligned}
$$

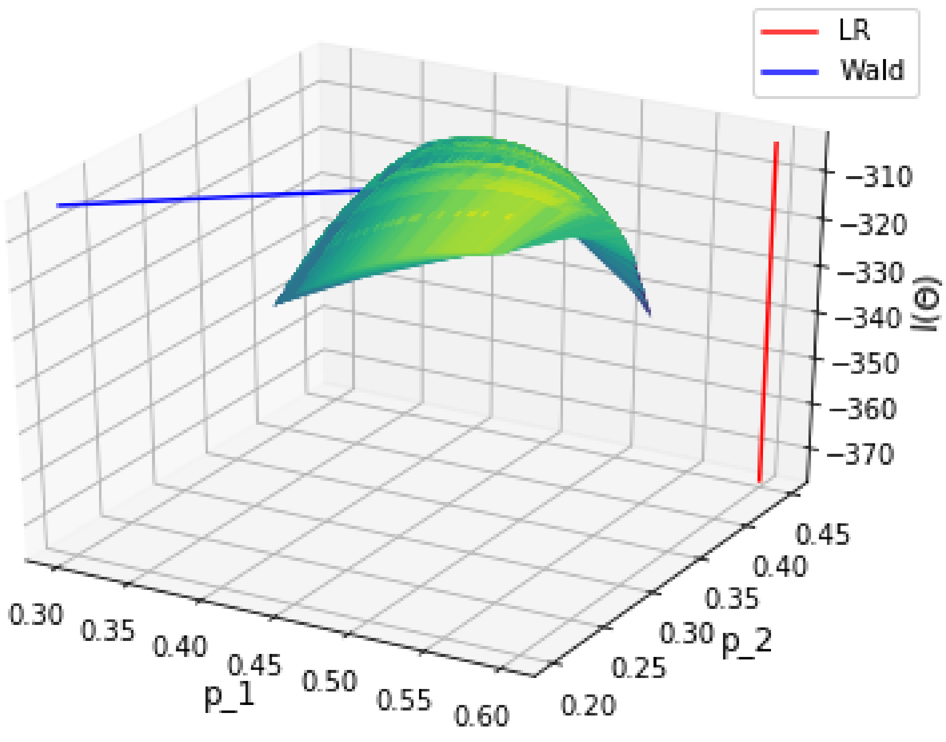
$$
\begin{aligned}
&LR : 2(l_{UR} - l_R) \approx 133.89 > 3.84 \sim \chi^2_{1 0.05}\\
&\text{не принимаем} H_0
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
&l''_{p_1} = -\frac{150}{p_1^2} - \frac{50}{(1 - p_1 - p_2)^2}\\
&I_f = (-H) = \frac{150}{p_1^2} + \frac{50}{(1 - p_1 - p_2)^2}\\
&\hat{I}_f = 4453.125\\
&LM : \frac{l'(\hat{\Theta}_R)^2}{V\hat{a}r(l'(\hat{\Theta}_R)} \approx \frac{316406.250}{4453.125} = 71.05 > 3.84 \sim \chi^2_{1 0.05}\\
&\text{Не принимаем} H_0
\end{aligned}
$$

1.d.

$$
\begin{gathered}
LR : 2(l(\hat{\Theta}_{UR}) - l(\hat{\Theta}_R)) \approx 145.55 > 5.99 \sim \chi^2_{20.05} \\
\text{не принимаем} H_0 \\
W : (\hat{\Theta}_{UR} - \Theta_R)^T [V\hat{a}r(\Theta_{UR})]^{-1} (\hat{\Theta}_{UR} - \Theta_R) \\
\hat{I}_f = -H(\hat{\Theta}) = - \begin{bmatrix} -2400 & -1800 \\ -1800 & -2700 \end{bmatrix} \\
(\hat{\Theta}_{UR} - \Theta_R)^T [V\hat{a}r(\Theta_{UR})]^{-1} (\hat{\Theta}_{UR} - \Theta_R) = 240.00 > 5.99 \sim \chi^2_{20.05} \\
\text{не принимаем} H_0
\end{gathered} \tag{3}
$$

1.e.



1.f.

Представим, что наши данные получены из Бернуллевского распределения, у нас выпадает либо зеленые ручки с вероятностью p_3, либо остальные - $1 - p_3$. Тогда

$$\frac{\hat{\Theta}_{ML} - \Theta}{\sqrt{\frac{Var(\hat{\Theta})}{n}}} \rightarrow \mathbb{N}(0,1)$$
$$\Theta \in [\hat{\Theta} - z_{\alpha/2}\sqrt{Var(\hat{\Theta})}\ ,\ \hat{\Theta} + z_{\alpha/2}\sqrt{Var(\hat{\Theta})}]$$
$$Var(\hat{\Theta}) = \hat{p}_3(1 - \hat{p}_3)$$
$$\Theta \in [(1 - \frac{1}{2} - \frac{1}{3}) - 1.96(\sqrt{\frac{1}{300}\frac{1}{6}\frac{5}{6}})\ ;\ (1 - \frac{1}{2} - \frac{1}{3}) + 1.96(\sqrt{\frac{1}{300}\frac{1}{6}\frac{5}{6}})]$$
$$\Theta \in [(0.124\ ;\ 0.209] \tag{4}$$

1.g.

Переобозначим наши переменные:

$$
\begin{aligned}
& a = p_1 + p_2 \\
& b = p_2 \\
& l = 150log(a-b) + 100log(b) + 50log(1-a) \\
& l'_a = \frac{150}{a-b} - \frac{50}{1-a} \\
& l'_a = 0 \\
& \hat{a} = \frac{3+\hat{b}}{4} \\
& l'_b = -\frac{150}{a-b} + \frac{100}{b} \\
& l'_b = 0 \\
& \hat{b} = \frac{2}{5}\hat{a} \\
& \hat{b} \to \hat{a} \to \hat{a} = \frac{15}{18}\ ;\ \hat{b} = \frac{1}{3} \\
& \hat{V}ar(a) = \hat{I}_{F_a}^{-1} \\
& \hat{I}_F = -H(\hat{\Theta}) = -\begin{bmatrix} -\frac{150}{(\hat{a}-\hat{b})^2} - \frac{50}{(1-\hat{a})^2} & \frac{150}{(\hat{a}-\hat{b})^2} \\ \frac{150}{(\hat{a}-\hat{b})^2} & -\frac{150}{(\hat{a}-\hat{b})^2} - \frac{100}{\hat{b}^2} \end{bmatrix} = \begin{bmatrix} 2400 & -600 \\ -600 & 1500 \end{bmatrix} \\
& \hat{V}ar(a) = \frac{1}{2160} \\
& a \in [\frac{15}{18} - 2.576\sqrt{\frac{1}{2160\cdot 300}}\ ;\ \frac{15}{18} - 2.576\sqrt{\frac{1}{2160\cdot 300}}] \\
& a \in [0.830\ ;\ 0.837]
\end{aligned}
\tag{5}
$$

1.h.

Чтобы построить доверительный интервал для нашей оценки, нагенерируем 1000 бутстреп выборок с помощью мультиномиального распределения с параметрами $[\frac{1}{2}\ ;\ \frac{1}{3}\ ;\ \frac{1}{6}]$ по 300 повторений. И используем percentile bootstrap.

```
import numpy as np

dots = np.random.multinomial(300,
                             [1.0 / 2, 1.0 / 3, 1.0 / 6]
                             1000)[:, :1]
sorted_dots = np.sort(dots.flatten())[50:950]
print(sorted_dots[0], sorted_dots[-1])
```

$\hat{p}_1 \in [128\ ;\ 139]$

1.i.

Выстрелила - пользуется значительно большим спросом чем суммарно ручки другого цвета. То есть ручку данного цвета покупают намного чаще чем все остальные ручки. Если представить покупку ручки с. в. с 3 элементарными исходами (покупка ручки разных цветов), то по плотности распределения "выстрелившая"ручка должна быть самой высокой даже больше чем сумма высоты остальных.

1.j.

$$\begin{aligned}&H_0 : p_1 > 1 - p_1 \rightarrow p_1 > 0.5\\&H_1 : p_1 \ngtr 0.5\\&LR : 2(l(\hat{\Theta}_{UR}) - l(\hat{\Theta}_R)) = 2 \cdot 0 = 0 < 2.706 \sim \chi^2_{1 0.05}\\&H_0 \text{ не отвергается}\end{aligned} \tag{6}$$

2.

2.a.

$$
\begin{aligned}
&n = 10^6 \\
&L = \prod_{i=1}^{n} \frac{2x_i}{b} exp(-\frac{x_i^2}{b}) = (\frac{2}{b})^n \prod_{i=1}^{n} x_i exp(-n\frac{\bar{x^2}}{b}) \\
&l = nln(\frac{2}{b}) + \sum_{i=1}^{n} ln(x_i) - n\frac{\bar{x^2}}{b} \\
&l' = \frac{nb}{2}(-\frac{2}{b^2}) + n\frac{\bar{x^2}}{b^2} \\
&l' = 0 \\
&\hat{b}_{ML} = \bar{x^2} = 20
\end{aligned}
\tag{7}
$$

2.b.

$$
\begin{aligned}
\text{LR}: 2(l(\hat{\Theta}_{UR} - l(\hat{\Theta}_R) = 2(10^6 ln(\frac{2}{20}) + \sum_{i=1}^{n} ln(x_i) - 10^6\frac{20}{20} - \\
- 10^6 ln(\frac{2}{3}) - \sum_{i=1}^{n} ln(x_i) + 10^6\frac{20}{3}) = 7.54 \cdot 10^6 > 3.84 \sim \chi^2_{10.05} \\
H_0 \text{ не принимается}
\end{aligned}
\tag{8}
$$

2.c.

$$
\begin{aligned}
\text{LM}: \frac{l'(\hat{\Theta}_R)^2}{V\hat{a}r(l'(\hat{\Theta}_R)} = \frac{l'(\hat{\Theta}_R)^2}{I_F(\Theta_R)} = \frac{(10^6(-1+\frac{20}{1^2}))^2}{10^6(40-1)} \approx 9.26 \cdot 10^6 > 3.84 \sim \chi^2_{10.05} \\
H_0 \text{ не принимается}
\end{aligned}
\tag{9}
$$

Значение p-value слишком мало, не считается с достаточной точностью функциями из stats, выводит 0.

2.d.

По теореме Крамера-Рао:

$$
\begin{aligned}
&Var(\hat{\Theta}) \geq \frac{1}{\hat{I}_F} \\
&\frac{\hat{\Theta}_{ML} - \Theta}{\sqrt{\frac{1}{\hat{I}_F}}} \rightarrow \mathbb{N}(0,1) \\
&\Theta \in [\hat{\Theta} - z_{\alpha/2} Var(\hat{\Theta})\ ,\ \hat{\Theta} + z_{\alpha/2} Var(\hat{\Theta})] \\
&\hat{I}_F = -H(\hat{\Theta}) = 10^6(\frac{40}{20^3} - \frac{1}{20^2}) = 2500 \\
&\Theta \in [20 - 1.96\sqrt{(\frac{1}{2500})}\ ;\ 20 + 1.96\sqrt{(\frac{1}{2500})}] \\
&1 \notin CI \rightarrow H_0 \text{ не принимается}
\end{aligned}
\tag{10}
$$

3.

3.a.

Давайте, изначально выведем общую формулу для $\hat{p}$.
В данном случае у нас используется биномиальное распределение, так что формулу для параметра будем выводить с помощью метода максимального правдоподобия.

$$\sum_{i=1}^{N} (\frac{c_i}{2}) p^2 (1-p)^{c_i - 2} \to max_p$$

, где c_i - количество бросков в i эксперименте до окончания.

$$\sum_{i=1}^{N} ln(\frac{c_i}{2}) + 2ln(p) + (c_i - 2)ln(1-p) \to max_p$$

$$\sum_{i=1}^{N} \frac{2}{\hat{p}} - \frac{c_i - 2}{1 - \hat{p}} = 0$$

$$\frac{2N}{\hat{p}} = \frac{N\bar{c} - 2N}{1 - \hat{p}}$$

$$2N - 2N\hat{p} = N\bar{c}\hat{p} - 2N\hat{p}$$

$$\hat{p} = \frac{2}{\bar{c}}$$

Рассчитаем $\bar{c} = \frac{3+3+2+4}{4} = 3$.

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

3.b.

$$a = p^2 + 3p^3 - 1$$

Чтобы получить $\hat{a}$ - подставим туда уже оцененные p, пользуясь свойством инвариантности оценок ММП и непрерывностью функции a.

$$\hat{a} = \hat{p}^2 + 3\hat{p}^3 - 1 = (\frac{2}{3})^2 + 3(\frac{2}{3})^3 - 1 = \frac{1}{3}$$

3.c.

Нам известно, что:

$$E(M) = \frac{2}{p}$$

, где M - количество бросков до конца игры (в первом подпункте - C).
Отсюда можно выразить настоящую p:

$$p = \frac{2}{E(M)}$$

Чтобы доказать состоятельность, нам нужно показать, используя выражения из первого подпункта, что:

$$\lim_{n \longrightarrow +\infty} P(|\frac{2}{\bar{c}_n} - p| \leq \epsilon) = 1$$

Подставим p, которую мы выразили ранее:

$$\lim_{n \longrightarrow +\infty} P(|\frac{2}{\bar{c}_n} - \frac{2}{E(M)}| \leq \epsilon) = 1$$

Можно внести предел под вероятность:

$$P(\lim_{n \longrightarrow +\infty} |\frac{2}{\bar{c}_n} - \frac{2}{E(M)}| \leq \epsilon) = 1$$

По закону больших чисел, при $n \rightarrow +\infty, \bar{c}_n \rightarrow E(M)$, тогда условие будет выполняться, наша оценка - состоятельна.

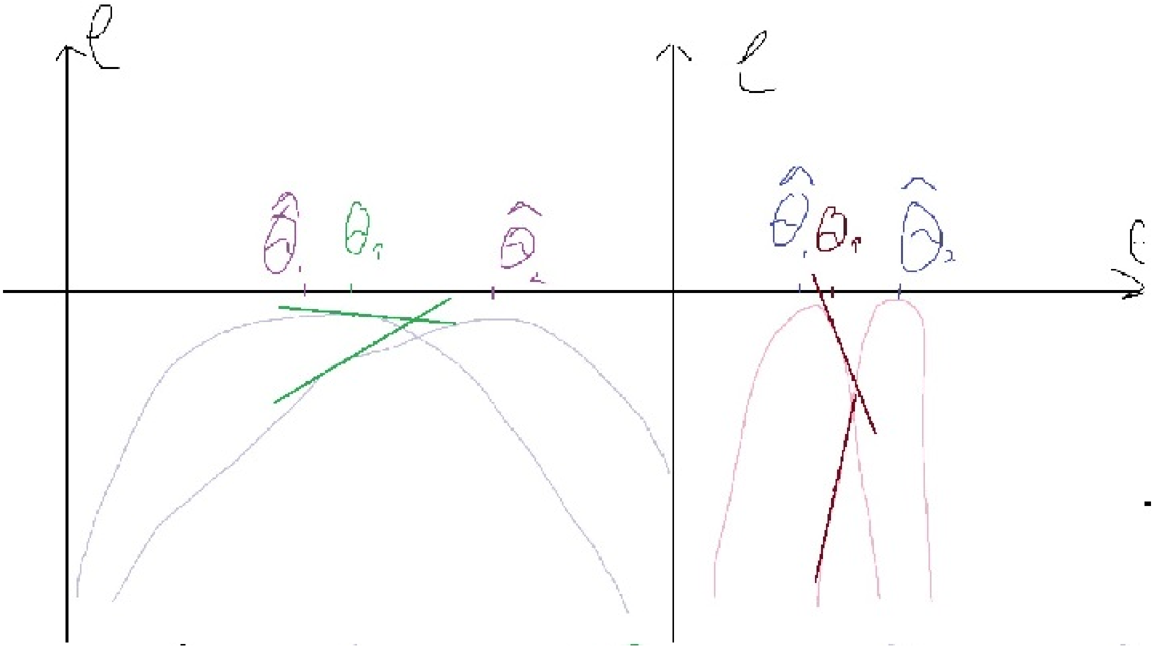
4.

4.a.

$$
\begin{aligned}
&KL(f(y|\Theta)||f(y|\phi)) = \int_{-\infty}^{+\infty} f(y|\Theta)log(f(y|\Theta))dy - \\
&-\int_{-\infty}^{+\infty} f(y|\Theta)log(f(y|\phi))dy = \\
&= |\text{Разложим логарифм второго члена в ряд Тейлора}| = \\
&= \int_{-\infty}^{+\infty} f(y|\Theta)log(f(y|\Theta))dy - \int_{-\infty}^{+\infty} f(y|\Theta)log(f(y|\Theta))dy - \\
&-\int_{-\infty}^{+\infty} f(y|\Theta)(log(f(y|\Theta))'(\phi-\Theta)dy^{*} - \\
&-\int_{-\infty}^{+\infty} f(y|\Theta)\frac{(log(f(y|\Theta))''}{2}(\phi-\Theta)^2dy + O((\phi-\Theta)^3) = \\
&= -\frac{1}{2}(\phi-\Theta)^2\int_{-\infty}^{+\infty} f(y|\Theta)\frac{\partial^2}{\partial\Theta^2}(log(f(y|\Theta))dy + O((\phi-\Theta)^3) = \\
&= -\frac{1}{2}E(H(y|\Theta))(\phi-\Theta)^2 + O((\phi-\Theta)^3) = \frac{1}{2}I_F(\phi-\Theta)^2 + O((\phi-\Theta)^3) \\
&* \ : \ \int_{-\infty}^{+\infty} f(y|\Theta)(log(f(y|\Theta))'(\phi-\Theta)dy = \\
&= (\phi-\Theta)\int_{-\infty}^{+\infty} f(y|\Theta)(\frac{\partial}{\partial\Theta}log(f(y|\Theta))dy = \\
&= (\phi-\Theta)\int_{-\infty}^{+\infty} f(y|\Theta)(\frac{1}{f(y|\Theta)})\frac{d}{d\Theta}f(y|\Theta)dy = \\
&= (\phi-\Theta)\int_{-\infty}^{+\infty} f(y|\Theta)(\frac{1}{f(y|\Theta)})\frac{d}{d\Theta}f(y|\Theta)dy = \\
&= (\phi-\Theta)\frac{d}{d\Theta}\int_{-\infty}^{+\infty} f(y|\Theta)dy = \\
&= 0
\end{aligned}
\tag{11}
$$

4.b.

Информация Фишера показывает дисперсию производной логарифма правдоподобия в точке настоящего значения параметра. Можно представить картинку с правдоподобием, это будет функция похожая на отрицательную параболу в осях (параметр, логарифм правдоподобия). Если это функция более пологая, то при некотором изменении выборки максимум логарифма правдоподобия изменится на значительную величину по оси параметра (в сравнении с пологой функцией), но в точке настоящего параметра производные будут не так сильно отличаться, как в таком же случае при крутой функции. При крутой функции при аналогичном изменении выборки максимумы не так сильно поменяются в сравнении, зато производные в настоящей точки будут достаточно сильно меняться. Из чего можно сделать вывод, если при построении разных логорифмов правдоподобия производная меняется сильно в точке настоящего параметра, то это нам дает намного больше информации об этом параметре из-за поведения функции, чем когда они меняются медленно, потому что это будет значит, что функции пологие $\rightarrow$ ближе к экстремуму. На рисунке можно увидеть иллюстрация примера данной логики, слева - пологая, справа - крутая.



5.

5.a.

Полумна неглупая, поэтому предполагает зависимость такого рода - чем полезнее зелье, тем оно менее популярно. Это связано с тем, что даже мир магов живет в капиталистическом строе, поэтому зельеварам крайне невыгодно продавать и рекламировать зелья высокой полезности, потому что, один раз использовав полезное зелье, волшебники не пойдут покупать его еще раз, так как проблема была решена. В связи с этим, они лоббируют менее полезные зелья, которые в итоге и становятся хитами продаж.
Полумна не успела доучиться в Хогвартсе, поэтому просто не знала, что на коэффициенты ее модели тоже наложен ряд ограничений, выведенных практиками. β_1 отвечает за максимальную популярность на рынке, всегда больше или равна 0, но не может быть больше 10 (максимальная условная единица популярности). β_2 - прямо пропорционален скорости убывания популярности при росте полезности.

$$\begin{aligned} L = \beta_1^{2n} exp(-\beta_2 \sum_{i=1}^{n} x_i) \prod_{i=1}^{n} u_i \\ l = 2nln(\beta_1) - \beta_2 \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} ln(u_i) \end{aligned} \tag{12}$$

Производные по нашим параметрам нам не дадут найти экстремальные точки, так что это надо сделать учитывая наши ограничения. Зная, что беты положительные, можно увидеть, что от первой беты функция логарифма правдоподобия зависит положительно, а от второй - отрицательно (если сумма $x_i > 0$, положительно, если сумма меньше) $\rightarrow$ максимум будет при максимальной $\hat{\beta}_{1ML} = 10$ и минимальной $\hat{\beta}_{2ML} = 0$.

5.b.

$$
\begin{aligned}
&W:(\hat{\Theta}_{UR}-\Theta_R)^T[V\hat{a}r(\Theta_{UR})]^{-1}(\hat{\Theta}_{UR}-\Theta_R)\\
&\hat{I}_F=-H(\hat{\Theta})=\begin{bmatrix}\frac{2n}{\hat{\beta}_1^2} & 0\\ 0 & 0\end{bmatrix}=\begin{bmatrix}\frac{1000}{100} & 0\\ 0 & 0\end{bmatrix}\\
&W:\begin{bmatrix}9\\ -2\end{bmatrix}^T\begin{bmatrix}\frac{1000}{100} & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}9\\ -2\end{bmatrix}=810>5.99\sim\chi^2_{20.05}\\
&H_0 \text{ не принимается}
\end{aligned}
\tag{13}
$$

6.

$n = 100, \overline{X} = 20, \overline{X^2} = 400$
$\hat{p}_{ML} = ?$
$p \in [0,1]$

6.a.

Первоначально стоит исследовать функцию, определить ее область определения, точки экстремума и второй знак производной.

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

$$Support(p) : p \in (0, 50) \bigcup (50, \inf)$$

$\sqrt{X_1 + ... + X_n} = \sqrt{\overline{X} * n} = 20\sqrt{5}$
$X_1^2 + ... + X_n^2 = \overline{X^2} * n = 40000$

$$l(p) = \frac{20\sqrt{5}}{50 - p} + \frac{ln(p)}{40000}$$

$$ML : \hat{p} = argmax_p l(p)$$

$$l'(p) = \frac{20\sqrt{5}}{(50 - p)^2} + \frac{1}{40000p}$$

$$\frac{20\sqrt{5}}{(50 - \hat{p})^2} + \frac{1}{40000\hat{p}} = 0$$

$$\hat{p}_1 = -\frac{50}{-1 + 8000\sqrt{5} + 40\sqrt{200000 - 10\sqrt{5}}} < 0$$

Данное решение не удовлетворяет нашей области определения, из-за ln(p) $\rightarrow$ p $>$ 0

$$\hat{p}_2 = -50 * (-1 + 8000\sqrt{5} + 40\sqrt{200000 - 10\sqrt{5}}) < 0$$

Данное решение так же не удовлетворяет нашей области определения, из-за ln(p) $\rightarrow$ p $>$ 0
Проверим знак производной на отрезке [0, 1].

$$l'(0.1) \approx 0.02 > 0 \rightarrow$$

функция растет на этом промежутке, то есть максимально значение будет в точке $\hat{p} = 1$

Ответ: $\hat{p} = 1$

6.b.

$$l(p) = \frac{(p^2 - ln(p))\overline{X^2}}{\overline{X}} = \frac{(p^2 - ln(p))400}{20} = 20(p^2 - ln(p))$$

$$Support(p) : p \in (0, \inf)$$

$$l'(p) = 2p - \frac{1}{p}$$

$$2\hat{p} - \frac{1}{\hat{p}} = 0$$

$$\hat{p} = \frac{\sqrt{2}}{2} \approx 0.7$$

Проверим знаки производной на $(0, \frac{\sqrt{2}}{2})(\frac{\sqrt{2}}{2}, 1]$.

$$l'(\frac{1}{e}) = \frac{2}{e} - e < 0$$

$$l'(1) = 2 - 1 > 0$$

Выходит, что $\frac{\sqrt{2}}{2}$ - точка минимума.
Есть смысл проверить значения функции в краевых точках, но так как 0 - не входит в область определения, то посчитаем предел при $\hat{p} \to 0$.

$$l(1) = 20$$

$$\lim_{\hat{p} \longrightarrow 0} 20((\hat{p})^2 - ln(\hat{p})) = +\infty$$

Ответ: нет решения.

7.

$p = \mathbb{N}(1, 2)$
$q = Exp(1)$
$r = Bin(3, 0.5)$

7.a.

$KL(p||q) = \infty$, так как $supp(p) \not\subset supp(q)$

7.b.

$$
\begin{aligned}
KL(q||p) = CE(q||p) - H(q) = \\
&= \int_{supp(q)} f_q(w) \log_2 \frac{1}{f_p(w)} dw - \int_{supp(q)} f_q(w) \log_2 \frac{1}{f_q(w)} dw = \\
&= \int_{supp(q)} exp(-w) \log_2 \frac{\sqrt{4pi}}{exp(-\frac{1}{4}(w-1)^2)} dw - \\
&- \int_{supp(q)} exp(-w) \log_2 \frac{1}{exp(-w)} dw = \\
&= -\frac{ln(\frac{exp(-\frac{1}{4})}{2\sqrt{\pi}})}{ln2} - \frac{1}{ln2} \approx 0.7437
\end{aligned} \tag{14}
$$

7.c.

$KL(p||r) = \infty$, так как $supp(p) \not\subset supp(r)$

7.d.

$KL(q||r) = \infty$, так как $supp(q) \not\subset supp(r)$

7.e.

Если p умножить/сложить на константу, то ответ на пункт b изменится, например, если вычесть единичку, то мат ожидание станет равно 0, полученное и экспоненциальное распределение станут более похожими на графике плотностей на саппорте q, по идее, KL(q||p) должно будет снизиться, но это неточно, надо проверять.

8.

8.a.

$$
\begin{aligned}
H(X|Y) = -\int_{x,y} p(x,y)log(\frac{p(x,y)}{p(y)})dxdy = \\
= \int_{x,y} p(x,y)log(\frac{1}{p(x,y)})dxdy - \int_{x,y} p(x,y)log(\frac{1}{p(y)})dxdy = \\
= \int_{x,y} p(x,y)log(\frac{1}{p(x,y)})dxdy - \int_{y} p(y)log(\frac{1}{p(y)})dy = \\
= H(X,Y) - H(Y)
\end{aligned}
\tag{15}
$$

Аналогично доказываем, что $H(Y|X) = H(X,Y) - H(X)$.

$$
I(X,Y) = I(Y,X)
$$
$$
H(X) - H(X|Y) = H(Y) - H(Y|X)
$$
$$
H(X) - H(X,Y) + H(Y) = H(Y) - H(X,Y) + H(X)
$$
$$
H(X,Y) \equiv H(X,Y) \rightarrow I(X,Y) \equiv I(Y,X)
$$

8.b.

Воспользуемся формулой, которую вывели в предпоследней строчки предыдущего подпункта для взаимной информации.

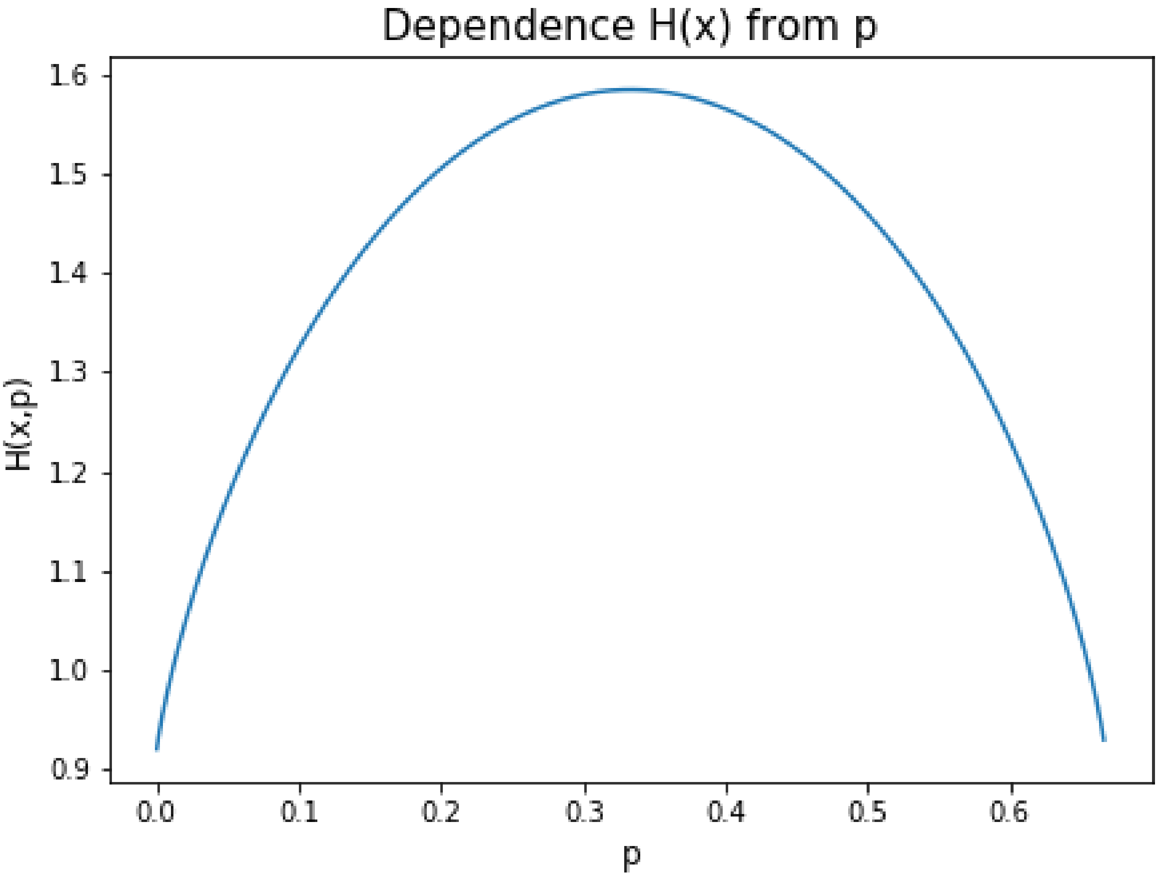
$$
\begin{aligned}
&I(X,Y) = H(X) + H(Y) - H(X,Y) = \\
&= \int_x p(x)log(\frac{1}{p(x)})dx + \int_y p(y)log(\frac{1}{p(y)})dy - \\
&\int_{x,y} p(x,y)log(\frac{1}{p(x,y)})dxdy = \\
&= \int_{x,y} p(x,y)log(\frac{1}{p(x)})dxdy + \int_{x,y} p(x,y)log(\frac{1}{p(y)})dxdy - \\
&\int_{x,y} p(x,y)log(\frac{1}{p(x,y)})dxdy = \\
&= \int_{x,y} p(x,y)log(\frac{p(x,y)}{p(x)p(y)})dxdy = KL(p(x,y)||p(x)p(y))
\end{aligned}
\tag{16}
$$

8.c.

По теории информации, то сколько лишней информации нам потребуется узнавать, если узнавать X, Y будем отдельно друг от друга, в отличие от того, сколько бы нам потребовалось, если бы узнавали вместе. Имеется ввиду, что можно было бы использовать информацию X при дальнейшем разборе Y. То есть, если они никак не пересекаются, то взаимная информация будет равна 0.

9.

9.a.



9.b.

При $p = \frac{1}{3}$ H(x,p) будет максимальна. Если все события будут равновероятными, то достигается максимальная неопределенность, потому что у нас, можно сказать, что практически нет никакой информации, так как для нас все исходы одинаково вероятны. Но, скажем, если бы $p = 0$, то это бы понизило Энтропию, потому что бы мы знали, что один исход невероятен, у нас осталось бы только 2 исхода.

10.

10.a.

11.

11.a.

$dim\, V = 1$(Один базисный вектор)
$dim\, W = 2$(Любая переменная выражается через 2 других)
$dim\, (V \cap W) = 0$(Если подставить значения 1 во 2, то не сойдется -> пе
$dim\, V^{\perp} = 3 - 1 = 2$(Всего 3-мерное пространство -> в ортогональном пр
$dim\, W^{\perp} = 3 - 2 = 1$

(17)

11.b.

$$\mathrm{u}_{projV} = X(X^TX)^{-1}X^Tu = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^T (\begin{bmatrix}1 & 1 & 1\end{bmatrix} \begin{bmatrix}1 & 1 & 1\end{bmatrix}^T)^{-1} \begin{bmatrix}1 & 1 & 1\end{bmatrix} u =$$

$= \frac{1}{3}\begin{bmatrix}u_1 + u_2 + u_3\\ u_1 + u_2 + u_3\\ u_1 + u_2 + u_3\end{bmatrix}$
$||u_{projV}||^2 = \frac{3}{9}(u_1 + u_2 + u_3)^2$
$u_{projW} = X(X^TX)^{-1}X^Tu = \begin{bmatrix}1 & 0\\ 0 & 1\\ 3 & 2\end{bmatrix} (\begin{bmatrix}1 & 0\\ 0 & 1\\ 3 & 2\end{bmatrix}^T \begin{bmatrix}1 & 0\\ 0 & 1\\ 3 & 2\end{bmatrix})^{-1} \begin{bmatrix}1 & 0\\ 0 & 1\\ 3 & 2\end{bmatrix}^T u =$
$= \begin{bmatrix}\frac{5}{14}u_1 - \frac{3}{7}u_2 + \frac{3}{14}u_3\\ -\frac{3}{7}u_1 + \frac{5}{7}u_2 + \frac{1}{7}u_3\\ \frac{3}{14}u_1 + \frac{1}{7}u_2 + \frac{13}{14}u_3\end{bmatrix}$, В матрице X столбцы базисных векторов
$||u_{projW}||^2 = (\frac{5}{14}u_1 - \frac{3}{7}u_2 + \frac{3}{14}u_3)^2 + (-\frac{3}{7}u_1 + \frac{5}{7}u_2 + \frac{1}{7}u_3)^2 +$
$+ (\frac{3}{14}u_1 + \frac{1}{7}u_2 + \frac{13}{14}u_3)^2$
$u_{projV \cap W} = \begin{bmatrix}0\\ 0\\ 0\end{bmatrix}$

$||u_{projV\cap W}||^2 = 0$
Надо найти 2 базисных вектора$V^{\perp}$
Они будут перпендикулярны базысному векторуV
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \\ -1 & -1 \end{bmatrix}$$
$$u_{projV^{\perp}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ -1 & -1 \end{bmatrix} (\begin{bmatrix} 0 & 1 \\ 1 & 0 \\ -1 & -1 \end{bmatrix}^T \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ -1 & -1 \end{bmatrix})^{-1} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ -1 & -1 \end{bmatrix}^T u =$$
$$= \frac{1}{3} \begin{bmatrix} 2u_1 - 1u_2 - 1u_3 \\ -1u_1 + 2u_2 - 1u_3 \\ -1u_1 - 1u_2 + 2u_3 \end{bmatrix}$$
$||u_{projV^{\perp}}||^2 = \frac{1}{9}((2u_1 - 1u_2 - 1u_3)^2 + (-1u_1 + 2u_2 - 1u_3)^2 +$
$+ (-1u_1 - 1u_2 + 2u_3)^2)$
Аналогично надо найти базисный вектор
$$\begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}$$
$$u_{projW^{\perp}} = \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix} (\begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}^T \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix})^{-1} \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}^T u = \frac{1}{14} \begin{bmatrix} 9u_1 + 6u_2 - 3u_3 \\ 6u_1 + 4u_2 - 2u_3 \\ -3u_1 - 2u_2 + 1u_3 \end{bmatrix}$$
$||u_{projW^{\perp}}||^2 = \frac{1}{196}((9u_1 + 6u_2 - 3u_3)^2 + (6u_1 + 4u_2 - 2u_3)^2 + (-3u_1 - 2u_2 + 1u_3)^2)$

11.c.

По теореме, если вектор имеет многомерное стандартное распределение и происходит проекция на подпространство размерности k, то квадрат длины проекции будет иметь χ^2_k распределение.

$$\begin{aligned}
&||u_{projV}||^2 \sim \chi^2_1 \\
&||u_{projW}||^2 \sim \chi^2_2 \\
&||u_{projV\cap W}||^2 \text{никак не будет распределен} \qquad (18)\\
&||u_{projV^\perp}||^2 \sim \chi^2_2 \\
&||u_{projW^\perp}||^2 \sim \chi^2_1
\end{aligned}$$

12.

12.a.

$$
\begin{aligned}
&RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \xrightarrow{\hat{\beta}_0,\, \hat{\beta}_1} min \\
&\frac{dRSS}{d\hat{\beta}_0} = -2(n\bar{y} - n\hat{\beta}_0 - n\hat{\beta}_1\bar{x}) \\
&(n\bar{y} - n\hat{\beta}_0 - n\hat{\beta}_1\bar{x}) = 0 \\
&\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \\
&\frac{dRSS}{d\hat{\beta}_1} = -2\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\
&\sum_{i=1}^{n}(x_i y_i - x_i\bar{y} + \hat{\beta}_1\bar{x}x_i - \hat{\beta}_1 x_i^2) = 0 \\
&\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sum_{i=1}^{n} x_i(x_i - \bar{x})} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})}
\end{aligned}
\tag{19}
$$

12.b.

$$
\begin{aligned}
&\sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \sum_{i=1}^{n}(y_i - \bar{y} + \hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i) = \\
&= n\bar{y} - n\bar{y} + \hat{\beta}_1(n\bar{x} - n\bar{x}) = 0
\end{aligned}
\tag{20}
$$

12.c.

$$
\bar{\hat{y}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_0 + \hat{\beta}_1 x_i = \frac{1}{n}\sum_{i=1}^{n}\bar{y} + \hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i = \bar{y} + \frac{\hat{\beta}_1}{n}(n\bar{x} - n\bar{x}) = \bar{y}
\tag{21}
$$

12.d.

$$\begin{aligned}
&\hat{y} = \bar{y} \\
&\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} = \bar{y}
\end{aligned} \tag{22}$$

12.e.

Это утверждение следует из второй половины подпункта a, где мы находим производную по $\hat{\beta}_1$, мы приравниваем нужное нам равенство 0, чтобы найти $\hat{\beta}_1$. Из этого следует, что подставив оценённые беты в равенство у нас так же выйдет 0.

13.

13.a.

$$\hat{\beta} = (X^T X)^{-1} X^T y = (\text{посчитано с помощью numpy}) = \begin{bmatrix} 2.67 \\ 0.04 \\ -0.17 \end{bmatrix} = \\ = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \tag{23}$$

13.b.

$$\hat{y} = X\hat{\beta} = \begin{bmatrix} 2.17 \\ 2.72 \\ 2.54 \\ 2.65 \\ 0.72 \end{bmatrix} \tag{24}$$

13.c.

$$\begin{aligned} TSS &= \sum_{i=1}^{n} (y_i - \bar{y})^2 = 5.97 \\ ESS &= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = 2.77 \\ RSS &= TSS - ESS = 3.20 \\ R^2 &= \frac{ESS}{TSS} = 0.46 \end{aligned} \tag{25}$$

13.d.

$$\hat{\sigma} = \sqrt{\frac{RSS}{n-k}} = \sqrt{\frac{3.20}{5-3}} = 1.26 \tag{26}$$

13.e.

$$Var(\hat{\beta}) = Var((X^TX)^{-1}X^Ty) = (X^TX)^{-1}X^TVar(y)X(X^TX)^{-1} =$$
$$= \sigma^2(X^TX)^{-1}$$
$$\hat{Var}(\hat{\beta}) = \hat{\sigma}^2(X^TX)^{-1} = \begin{bmatrix} 0.65 & -0.05 & -0.07 \\ -0.04 & 0.01 & 0.002 \\ -0.07 & 0.002 & 0.02 \end{bmatrix} \tag{27}$$

13.f.

$$\begin{aligned} &H_0 : \beta_1 = 1 \\ &H_1 : \beta_1 \neq 1 \\ &t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{V}ar(\hat{\beta}_1)}} = \frac{0.04 - 1}{\sqrt{0.01}} = -9.6 \sim t_{5-3} \\ &-t_{crit,2} = -4.3 \\ &H_0 \text{ отвергается, так как } t_{obs} < -t_{crit,2} \end{aligned} \tag{28}$$

13.g.

$$\begin{aligned} &H_0 : \beta_2 = 1 \\ &H_1 : \beta_2 < 1 \\ &t = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\hat{V}ar(\hat{\beta}_2)}} = \frac{-0.17 - 1}{\sqrt{0.02}} = -8.27 \sim t_{5-3} \\ &-t_{crit,2} = -1.89 \\ &H_0 \text{ отвергается, так как } t_{obs} < -t_{crit,2} \end{aligned} \tag{29}$$

13.h.

$$
\begin{aligned}
&H_0 : \beta_1 = \beta_2 = 0\\
&H_1 : \beta_1^2 + \beta_2^2 > 0\\
&UR : \text{the same}\\
&R : y_i = \beta_0 + u_i\ ,\ \hat{\beta}_0 = \bar{y} = 2.16\\
&RSS_R = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2 = TSS_{UR}\\
&\text{(Это выполняется только в нашем случае, а не в обещм)}\\
&F = \frac{(RSS_R - RSS_{UR})/(k_{UR} - k_R)}{RSS_{UR}/(n - k_{UR})} = \frac{(5.97 - 3.20)/(3 - 1)}{3.20/(5 - 3)} = 0.87 \sim F_{2,2}\\
&F_{crit,2,2} = 19(5\%\text{уровень значимости})\\
&H_0 \text{ не отвергается, так как } F_{obs} < F_{crit,2,2}
\end{aligned}
\tag{30}
$$

13.i.

$$
\begin{aligned}
&H_0 : \beta_1 = \beta_2\\
&H_1 : \beta_1 \neq \beta_2\\
&UR : \text{the same}\\
&R : y_i = \beta_0 + \beta_1(x_{i1} + x_{i2}) + u_i \ , \ x_1 + x_2 = \begin{bmatrix} 1. & 4.1\\ 1. & 14.2\\ 1. & -2.9\\ 1. & 2.5\\ 1. & 11.3 \end{bmatrix}\\
&R : \ \hat{\beta} = (X.TX)^{-1}X^Ty = \begin{bmatrix} 2.41\\ -0.04 \end{bmatrix}\\
&R : \hat{y} = \begin{bmatrix} 2.23445699\\ 1.80226413\\ 2.53399659\\ 2.30292318\\ 1.92635911 \end{bmatrix}\\
&RSS_R = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 5.62\\
&F = \frac{(RSS_R - RSS_{UR})/(k_{UR} - k_R)}{RSS_{UR}/(n - k_{UR})} = \frac{(5.62 - 3.20)/(3 - 2)}{3.20/(5 - 3)} = 1.51 \sim F_{1,2}\\
&F_{crit,1,2} = 18.51 (5\% \text{уровень значимости})\\
&H_0 \text{ не отвергается, так как } F_{obs} < F_{crit,1,2}
\end{aligned}
\tag{31}
$$

13.j.

$$
\begin{aligned}
&\beta_1 \in [\hat{\beta}_1 - t_{0.025,2}\sqrt{\hat{V}ar(\hat{\beta}_1)} \ ; \ \hat{\beta}_1 + t_{0.975,2}\sqrt{\hat{V}ar(\hat{\beta}_1)}]\\
&\beta_1 \in [0.04 - 4.3 \cdot 0.1 \ ; \ 0.04 + 4.3 \cdot 0.1]\\
&\beta_1 \in [-0.39 \ ; \ 0.47]
\end{aligned}
\tag{32}
$$

13.k.

$$\hat{y}_6 = \hat{\beta}_0 + \hat{\beta}_1 x_{1,6} + \hat{\beta}_2 x_{2,6} = 2.67 + 0.04 \cdot 10 - 0.17 \cdot 7 = 1.78 \quad (33)$$

13.l.

$$\begin{aligned}
&V\hat{a}r(\hat{y}_6) = V\hat{a}r(\hat{\beta}_0) + 100V\hat{a}r(\hat{\beta}_1) + 49V\hat{a}r(\hat{\beta}_2) + 10cov(\hat{\beta}_0, \beta_1) + \\
&+ 7cov(\hat{\beta}_0, \beta_2) + 70cov(\hat{\beta}_1, \beta_2) = \\
&= 0.65 + 1 + 0.98 - 0.5 - 0.49 + 0.14 = 1.78 \\
&E(y_6 | x_{1,6}, x_{2,6}) \in [1.78 \pm \sqrt{1.78} t_{0.025,2}] \\
&E(y_6 | x_{1,6}, x_{2,6}) \in [1.78 \pm 5.74]
\end{aligned} \quad (34)$$

14. X†

14.a.

$$EX = \frac{1}{N}\sum_{i=1}^{n} X_i \qquad (35)$$

$$\hat{E}X = \frac{1}{N}\sum_{i=1}^{n} X_i = \bar{X} = 6/3 = 2 \qquad (36)$$

14.b.

$EX = moment_1$
$EM \cdot p = \bar{x},$
Где EM ожидаемое число до окончания игры
$\bar{x} = 2$
учитывая, что игра кончается при 2 орлах, которые будем считать $= 1$
$\hat{p}_{MM} = \dfrac{\bar{x}}{EM} = \dfrac{2}{40} = 0.05$

(37)