

# Отчет по заданию №1

курс «Байесовские методы в машинном обучении»  
кафедра ММП ВМК МГУ

Михеев Борис, 417 группа  
Вариант 1

4 октября 2022 г.

## Постановка задачи

Рассмотрим модель посещаемости студентами ВУЗа одной лекции по курсу. Пусть аудитория данного курса состоит из студентов профильного факультета, а также студентов других факультетов. Обозначим через  $a$  количество студентов, поступивших на профильный факультет, а через  $b$  – количество студентов других факультетов. Пусть студенты профильного факультета посещают лекцию с некоторой вероятностью  $p_1$ , а студенты остальных факультетов – с вероятностью  $p_2$ . Обозначим через  $c$  количество студентов, посетивших данную лекцию. Тогда случайная величина  $c|a, b$  есть сумма двух случайных величин, распределённых по биномиальному закону  $Bin(a, p_1)$  и  $Bin(b, p_2)$  соответственно. Пусть далее на лекции по курсу ведётся запись студентов. При этом каждый студент записывается сам, а также, быть может, записывает своего товарища, которого на лекции на самом деле нет. Пусть студент записывает своего товарища с некоторой вероятностью  $p_3$ . Обозначим через  $d$  общее количество записавшихся на данной лекции. Тогда случайная величина  $d|c$  представляет собой сумму  $c$  и случайной величины, распределённой по биномиальному закону  $Bin(c, p_3)$ . Для завершения задания вероятностной модели осталось определить априорные вероятности для  $a$  и для  $b$ . Пусть обе эти величины распределены равномерно в своих интервалах  $[a_{min}, a_{max}]$  и  $[b_{min}, b_{max}]$  (дискретное равномерное распределение). Таким образом, мы определили следующую вероятностную модель (**модель 1**):

$$\begin{aligned} p(a, b, c, d) &= p(d|c)p(c|a, b)p(a)p(b), \\ d|c &\sim c + Bin(c, p_3), \\ c|a, b &\sim Bin(a, p_1) + Bin(b, p_2), \\ a &\sim Unif[a_{min}, a_{max}], \\ b &\sim Unif[b_{min}, b_{max}]. \end{aligned} \tag{1}$$

Рассмотрим несколько упрощённую версию модели 1. Известно, что биномиальное распределение  $Bin(n, p)$  при большом количестве испытаний и маленькой вероятности успеха может быть с высокой точностью приближено пуассоновским распределением  $Poiss(\lambda)$  с  $\lambda = np$ . Известно также, что сумма двух пуассоновских распределений с параметрами  $\lambda_1$  и  $\lambda_2$  есть пуассоновское распределение с параметром  $\lambda_1 + \lambda_2$  (для биномиальных распределений это неверно). Таким образом, мы можем сформулировать вероятностную модель, которая является приближённой версией модели 1 (**модель 2**):

$$\begin{aligned}
p(a, b, c, d) &= p(d|c)p(c|a, b)p(a)p(b), \\
d|c &\sim c + \text{Bin}(c, p_3), \\
c|a, b &\sim \text{Pois}(ap_1 + bp_2), \\
a &\sim \text{Unif}[a_{\min}, a_{\max}], \\
b &\sim \text{Unif}[b_{\min}, b_{\max}].
\end{aligned} \tag{2}$$

Рассматриваются модели 1 и 2 с параметрами  $a_{\min} = 75$ ,  $a_{\max} = 90$ ,  $b_{\min} = 500$ ,  $b_{\max} = 600$ ,  $p_1 = 0.1$ ,  $p_2 = 0.01$ ,  $p_3 = 0.3$ . Взять в качестве диапазона допустимых значений для величины  $c$  интервал  $[0, a_{\max} + b_{\max}]$ , а для величины  $d$  – интервал  $[0, 2(a_{\max} + b_{\max})]$ .

## Решение

### 1. Вывести формулы для всех необходимых далее распределений аналитически.

#### Модель 1

Т. к.  $a$  и  $b$  являются дискретными случайными величинами и распределены равномерно на соответствующих отрезках, применим для них известные формулы плотностей. Далее для удобства записи обозначим  $a_{\max} - a_{\min} + 1 = n_a$ ,  $b_{\max} - b_{\min} + 1 = n_b$ .

$$p(a) = \frac{[a_{\min} \leq a \leq a_{\max}]}{n_a}.$$

$$p(b) = \frac{[b_{\min} \leq b \leq b_{\max}]}{n_b}.$$

Т. к.  $c|a, b \sim \text{Bin}(a, p_1) + \text{Bin}(b, p_2)$ , то:

$$p(c|a, b) = p(c = \text{Bin}(a, p_1) + \text{Bin}(b, p_2)|a, b) = \sum_{i=0}^c C_a^i p_1^i (1 - p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1 - p_2)^{b-c+i}.$$

Заметим, что под знаком суммирования стоят произведения плотностей двух величин  $\text{Bin}(a, p_1)$  и  $\text{Bin}(b, p_2)$  в точках  $i$  и  $c - i$  соответственно. Используем это при вычислениях на компьютере с применением векторных операций, воспользовавшись [идеями со stackoverflow](#) и полученной формулой для  $p(c|a, b)$ : сгенерируем матрицы плотностей для каждого распределения при каждом  $c$  и  $a$  или  $b$  соответственно (с помощью `scipy`), а затем найдем искомое распределение по формуле путем перемножения соответствующих подматриц (в матрице для  $a$  возьмем первые  $c$  строк и транспонируем, в матрице для  $b$  возьмем последние  $c$  строк в обратном порядке).

Далее воспользуемся правилом суммирования вероятностей:

$$p(c|a) = \sum_{b=b_{\min}}^{b_{\max}} p(c|a, b)p(b) = \frac{1}{n_b} \sum_{b=b_{\min}}^{b_{\max}} \sum_{i=0}^c C_a^i p_1^i (1 - p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1 - p_2)^{b-c+i}.$$

$$p(c|b) = \sum_{a=a_{\min}}^{a_{\max}} p(c|a, b)p(a) = \frac{1}{n_a} \sum_{a=a_{\min}}^{a_{\max}} \sum_{i=0}^c C_a^i p_1^i (1 - p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1 - p_2)^{b-c+i}.$$

$$p(c) = \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} p(c|a, b)p(a)p(b) = \frac{1}{n_a n_b} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} \sum_{i=0}^c C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i}.$$

Т. к.  $d|c \sim c + Bin(c, p_3)$ , то:

$$p(d|c) = p(d = c + Bin(c, p_3)|c) = p(d - c = Bin(c, p_3)|c) = C_c^{d-c} p_3^{d-c} (1-p_3)^{2c-d}.$$

Используем правило суммирования вероятностей:

$$\begin{aligned} p(d) &= \sum_{c=0}^{a_{max}+b_{max}} p(d|c)p(c) = \sum_{c=0}^{a_{max}+b_{max}} C_c^{d-c} p_3^{d-c} (1-p_3)^{2c-d} p(c) = \\ &= \frac{1}{n_a n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} \sum_{i=0}^c C_c^{d-c} p_3^{d-c} (1-p_3)^{2c-d} C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i}. \end{aligned}$$

Применим далее теорему Байеса и правило суммирования вероятностей:

$$p(c|d) = \frac{p(d|c)p(c)}{p(d)}.$$

$$p(c|a, b, d) = \frac{p(a, b, c, d)}{p(a, b, d)} = \frac{p(d|c)p(c|a, b)p(a)p(b)}{\sum_{c=0}^{a_{max}+b_{max}} p(d|c)p(c|a, b)p(a)p(b)}.$$

## Модель 2

Формулы для данной модели во многом схожи с моделью 1. Распределения для  $a$  и  $b$  останутся теми же:

$$p(a) = \frac{[a_{min} \leq a \leq a_{max}]}{n_a}.$$

$$p(b) = \frac{[b_{min} \leq b \leq b_{max}]}{n_b}.$$

Главное отличие модели 2 от модели 1 состоит в распределении  $p(c|a, b)$ , т. к.  $c|a, b \sim \text{Poiss}(ap_1 + bp_2)$ :

$$p(c|a, b) = p(c = \text{Poiss}(ap_1) + \text{Poiss}(bp_2)|a, b) = \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)}.$$

Стоит отметить, что воспользовавшись свойствами бинома Ньютона, данное выражение можно привести к следующему виду:

$$\frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)} = \sum_{i=0}^c \frac{C_c^i (ap_1)^i (bp_2)^{c-i}}{c!} e^{-(ap_1 + bp_2)} = \sum_{i=0}^c \frac{(ap_1)^i (bp_2)^{c-i}}{i!(c-i)!} e^{-ap_1} e^{-bp_2}.$$

Это позволяет использовать при вычислениях ту же идею, что и в случае модели 1, но для пуассоновских распределений вместо биномиальных.

Далее, по аналогии с предыдущей моделью:

$$p(c|a) = \sum_{b=b_{min}}^{b_{max}} p(c|a, b)p(b) = \frac{1}{n_b} \sum_{b=b_{min}}^{b_{max}} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)}.$$

$$p(c|b) = \sum_{a=a_{min}}^{a_{max}} p(c|a, b)p(a) = \frac{1}{n_a} \sum_{a=a_{min}}^{a_{max}} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)}.$$

$$p(c) = \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} p(c|a, b)p(a)p(b) = \frac{1}{n_a n_b} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)}.$$

$$p(d|c) = p(d = c + \text{Bin}(c, p_3)|c) = p(d - c = \text{Bin}(c, p_3)|c) = C_c^{d-c} p_3^{d-c} (1 - p_3)^{2c-d}.$$

$$\begin{aligned} p(d) &= \sum_{c=0}^{a_{max}+b_{max}} p(d|c)p(c) = \sum_{c=0}^{a_{max}+b_{max}} C_c^{d-c} p_3^{d-c} (1 - p_3)^{2c-d} p(c) = \\ &= \frac{1}{n_a n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} C_c^{d-c} p_3^{d-c} (1 - p_3)^{2c-d} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)}. \end{aligned}$$

$$p(c|d) = \frac{p(d|c)p(c)}{p(d)}.$$

$$p(c|a, b, d) = \frac{p(a, b, c, d)}{p(a, b, d)} = \frac{p(d|c)p(c|a, b)p(a)p(b)}{\sum_{c=0}^{a_{max}+b_{max}} p(d|c)p(c|a, b)p(a)p(b)}.$$

**2. Найти математические ожидания и дисперсии априорных распределений  $p(a)$ ,  $p(b)$ ,  $p(c)$ ,  $p(d)$ .**

**Модель 1**

Для равномерно распределенных  $a$  и  $b$  используем известные формулы:

$$\mathbb{E}a = \sum_{a=a_{min}}^{a_{max}} p(a)a = \sum_{a=a_{min}}^{a_{max}} \frac{a}{n_a} = \frac{a_{min} + a_{max}}{2}.$$

$$\mathbb{D}a = \sum_{a=a_{min}}^{a_{max}} p(a)a^2 - (\mathbb{E}a)^2 = \sum_{a=a_{min}}^{a_{max}} \frac{a^2}{n_a} - \left( \frac{a_{min} + a_{max}}{2} \right)^2 = \frac{(a_{max} - a_{min} + 1)^2 - 1}{12}.$$

$$\mathbb{E}b = \sum_{b=b_{min}}^{b_{max}} p(b)b = \sum_{b=b_{min}}^{b_{max}} \frac{b}{n_b} = \frac{b_{min} + b_{max}}{2}.$$

$$\mathbb{D}b = \sum_{b=b_{min}}^{b_{max}} p(b)b^2 - (\mathbb{E}b)^2 = \sum_{b=b_{min}}^{b_{max}} \frac{b^2}{n_b} - \left( \frac{b_{min} + b_{max}}{2} \right)^2 = \frac{(b_{max} - b_{min} + 1)^2 - 1}{12}.$$

$$\mathbb{E}c = \sum_{c=0}^{a_{max}+b_{max}} p(c)c = \frac{1}{n_a n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} \sum_{i=0}^c C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i} c.$$

$$\begin{aligned} \mathbb{D}c &= \sum_{c=0}^{a_{max}+b_{max}} p(c)c^2 - (\mathbb{E}c)^2 = \\ &= \frac{1}{n_a n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} \sum_{i=0}^c C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i} c^2 - (\mathbb{E}c)^2. \end{aligned}$$

$$\begin{aligned} \mathbb{E}d &= \sum_{d=0}^{2a_{max}+2b_{max}} p(d)d = \sum_{d=0}^{2a_{max}+2b_{max}} \sum_{c=0}^{a_{max}+b_{max}} C_c^{d-c} p_3^{d-c} (1-p_3)^{2c-d} p(c)d = \\ &= \frac{1}{n_a n_b} \sum_{d=0}^{2a_{max}+2b_{max}} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} \sum_{i=0}^c C_c^{d-c} p_3^{d-c} (1-p_3)^{2c-d} C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i} d. \end{aligned}$$

$$\begin{aligned} \mathbb{D}d &= \sum_{d=0}^{2a_{max}+2b_{max}} p(d)d^2 - (\mathbb{E}d)^2 = \\ &= \frac{1}{n_a n_b} \sum_{d=0}^{2a_{max}+2b_{max}} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} \sum_{i=0}^c C_c^{d-c} p_3^{d-c} (1-p_3)^{2c-d} C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i} d^2 - \\ &\quad - (\mathbb{E}d)^2. \end{aligned}$$

Результаты вычислений данных величин приведены в табл. 1.

$\mathbb{E}a$	82.5	$\mathbb{D}a$	21.25
$\mathbb{E}b$	550	$\mathbb{D}b$	850
$\mathbb{E}c$	13.75	$\mathbb{D}c$	13.168
$\mathbb{E}d$	17.875	$\mathbb{D}d$	25.141

Таблица 1: Математические ожидания и дисперсии априорных распределений  $p(a)$ ,  $p(b)$ ,  $p(c)$ ,  $p(d)$ , модель 1.

## Модель 2

Изменяются только формулы для величин  $c$  и  $d$ :

$$\mathbb{E}a = \frac{a_{min} + a_{max}}{2}.$$

$$\mathbb{D}a = \frac{(a_{max} - a_{min} + 1)^2 - 1}{12}.$$

$$\mathbb{E}b = \frac{b_{min} + b_{max}}{2}.$$

$$\mathbb{D}b = \frac{(b_{max} - b_{min} + 1)^2 - 1}{12}.$$

$$\mathbb{E}c = \frac{1}{n_a n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1+bp_2)} c.$$

$$\mathbb{D}c = \frac{1}{n_a n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1+bp_2)} c^2 - (\mathbb{E}c)^2.$$

$$\mathbb{E}d = \frac{1}{n_a n_b} \sum_{d=0}^{2a_{max}+2b_{max}} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} C_c^{d-c} p_3^{d-c} (1-p_3)^{2c-d} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1+bp_2)} d.$$

$$\mathbb{D}d = \frac{1}{n_a n_b} \sum_{d=0}^{2a_{max}+2b_{max}} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} C_c^{d-c} p_3^{d-c} (1-p_3)^{2c-d} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1+bp_2)} d^2 - (\mathbb{E}d)^2.$$

Результаты вычислений приведены в табл. 2.

$\mathbb{E}a$	82.5	$\mathbb{D}a$	21.25
$\mathbb{E}b$	550	$\mathbb{D}b$	850
$\mathbb{E}c$	13.75	$\mathbb{D}c$	14.048
$\mathbb{E}d$	17.875	$\mathbb{D}d$	26.628

Таблица 2: Математические ожидания и дисперсии априорных распределений  $p(a)$ ,  $p(b)$ ,  $p(c)$ ,  $p(d)$ , модель 2.

Для обеих моделей математические ожидания априорных распределений, а также  $\mathbb{D}a$  и  $\mathbb{D}b$  совпадают,  $\mathbb{D}c$  и  $\mathbb{D}d$  больше для второй модели, вероятно в силу приближения биномиального распределения пуассоновским и погрешности этого приближения.

3. Пронаблюдать, как происходит уточнение прогноза для величины  $c$  по мере прихода новой косвенной информации. Для этого построить графики и найти мат.ожидание и дисперсию для распределений  $p(c)$ ,  $p(c|a)$ ,  $p(c|b)$ ,  $p(c|d)$ ,  $p(c|a,b)$ ,  $p(c|a,b,d)$  при параметрах  $a$ ,  $b$ ,  $d$ , равных мат.ожиданиям своих априорных распределений, округленных до ближайшего целого.

#### Модель 1

Построим графики плотностей распределений. Стоит отметить, что в данном случае исследования проводятся при фиксированных  $a$ ,  $b$ ,  $d$ . Для начала отложим по оси абсцисс весь диапазон значений  $c$ .

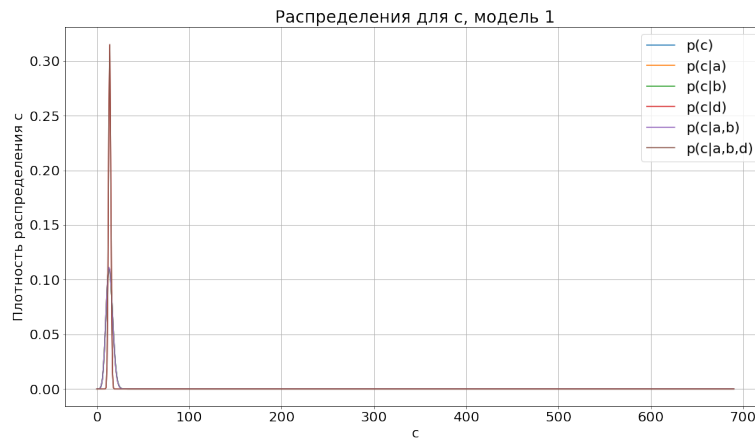


Рис. 1: Графики распределений для  $c$  на всем диапазоне значений  $c$ , модель 1.

По данному графику сложно о чем-то судить, распределения сосредоточены на интервале примерно от 0 до 30, в остальных точках плотности равны или близки к нулю. Рассмотрим детальнее отрезок сосредоточения.

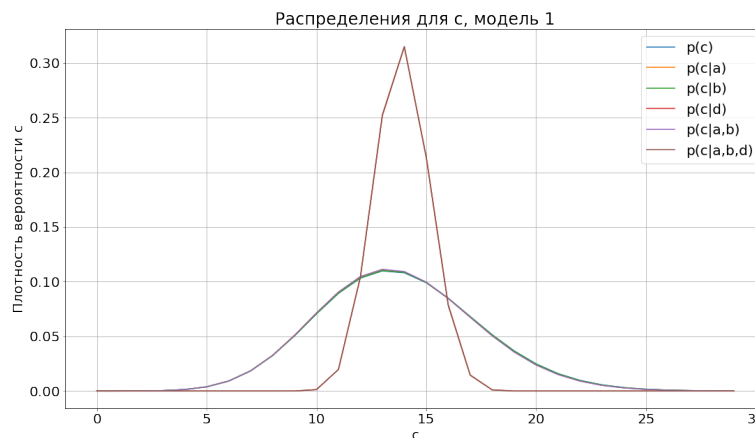


Рис. 2: Графики распределений для  $c$ , модель 1.

Графики очень близки друг к другу и наслаиваются, затруднительно заметить различия распределений. Судя по всему,  $p(c)$ ,  $p(c|a)$ ,  $p(c|b)$  и  $p(c|a, b)$  почти идентичны,  $p(c|d)$  и  $p(c|a, b, d)$  также очень близки друг к другу. Можно сказать, что информация о  $d$  заметно влияет на форму распределения, оно сильнее сосредотачивается, становится более «остроконечным», его максимум становится более выраженным и располагается близко к значению  $\mathbb{E}c$ , дисперсия уменьшается, что свидетельствует об уточнении прогноза для  $c$ .

Для более детального анализа построим графики данных распределений по отдельности и сравним их с априорным распределением  $p(c)$ .

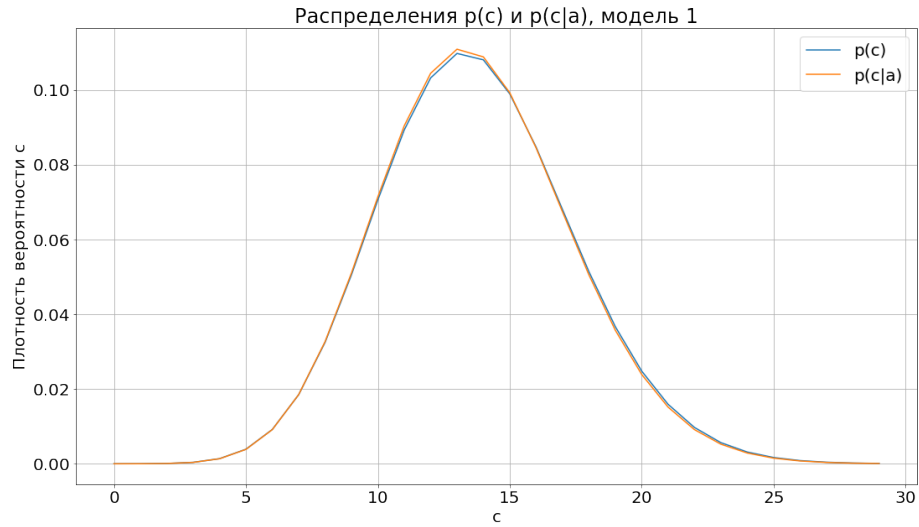


Рис. 3:  $p(c)$  и  $p(c|a)$ , модель 1.

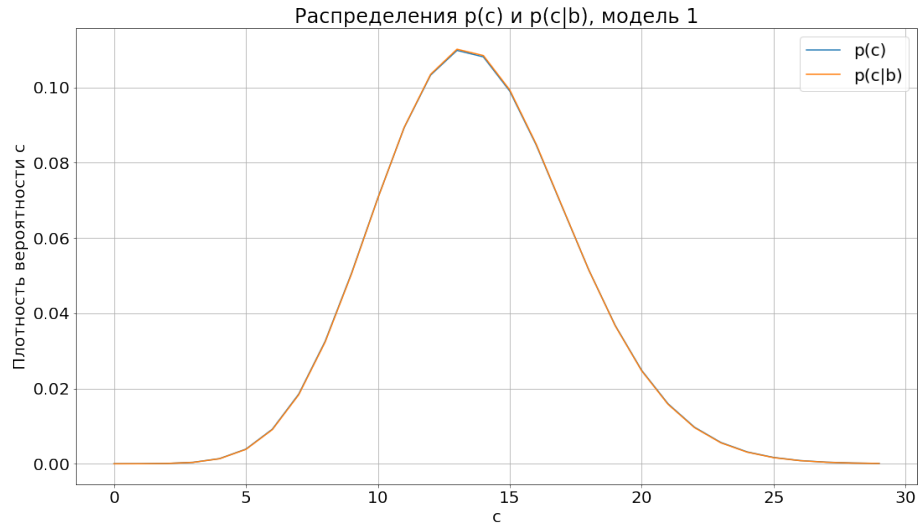


Рис. 4:  $p(c)$  и  $p(c|b)$ , модель 1.



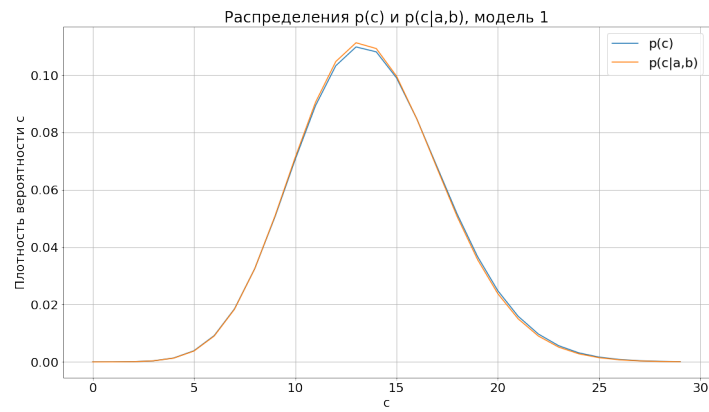


Рис. 5:  $p(c)$  и  $p(c|a,b)$ , модель 1.

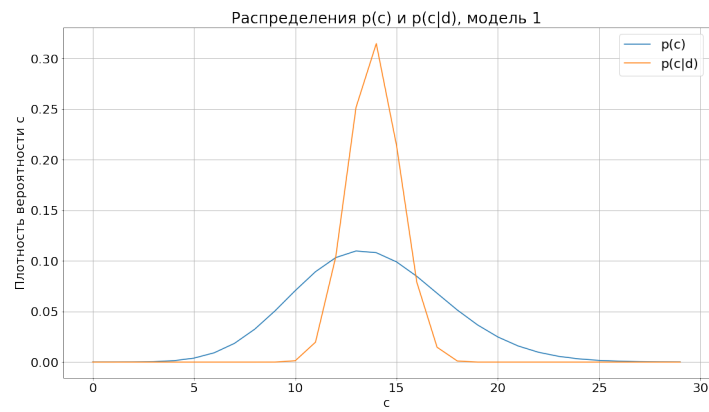


Рис. 6:  $p(c)$  и  $p(c|d)$ , модель 1.

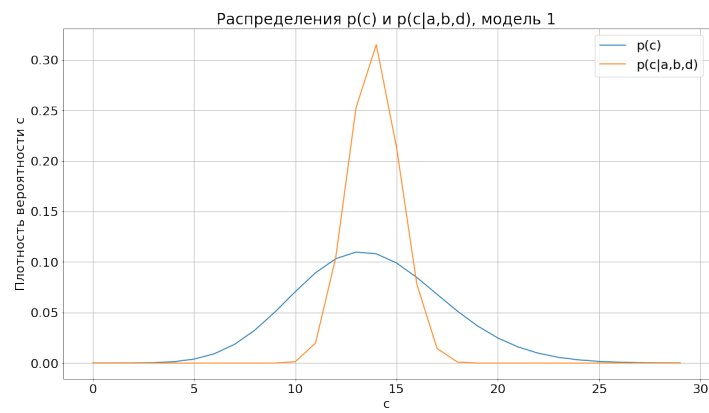


Рис. 7:  $p(c)$  и  $p(c|a,b,d)$ , модель 1.

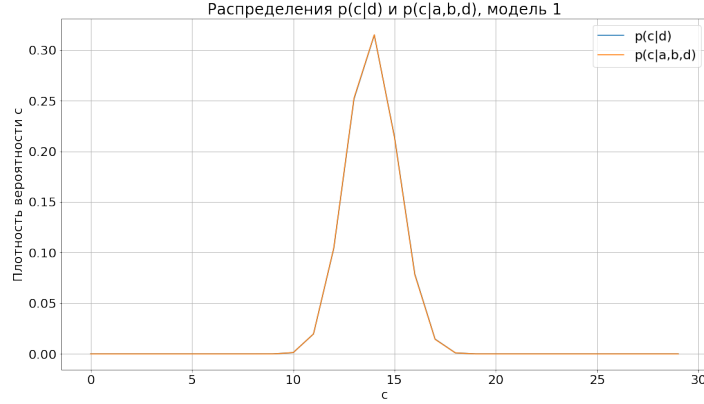


Рис. 8:  $p(c|d)$  и  $p(c|a, b, d)$ , модель 1.

По графикам можно сказать, что информация об  $a$  сильнее влияет на распределение и прогноз  $c$ , чем информация о  $b$  (она почти не оказывает влияния,  $p(c)$  и  $p(c|b)$  практически идентичны). Информация об  $a$  и  $b$  в совокупности оказывает влияние несколько большее, чем об  $a$  и  $b$  по отдельности, однако во всех перечисленных случаях распределения изменяются достаточно слабо и очень близки к априорному, т. е.  $a$  и  $b$  слабо влияют на уточнение прогноза  $c$ . Информация о  $d$  влияет на форму распределения значительно сильнее, значительно уменьшается дисперсия, отчетливее выделяется максимум, что свидетельствует об уточнении прогноза  $c$ . Стоит отметить, что распределения  $p(c|a, b, d)$  и  $p(c|d)$  почти полностью совпадают (рис. 8), т. е. при наличии информации о  $d$  информация об  $a$  и  $b$  почти не оказывает влияния.

Найдем матожидания и дисперсии для рассмотренных распределений. Приведем формулы для их нахождения:

$$\mathbb{E}c = \sum_{c=0}^{a_{max}+b_{max}} p(c)c = \frac{1}{n_a n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} \sum_{i=0}^c C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i} c.$$

$$\begin{aligned} \mathbb{D}c &= \sum_{c=0}^{a_{max}+b_{max}} p(c)c^2 - (\mathbb{E}c)^2 = \\ &= \frac{1}{n_a n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} \sum_{i=0}^c C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i} c^2 - (\mathbb{E}c)^2. \end{aligned}$$

$$\mathbb{E}(c|a) = \sum_{c=0}^{a_{max}+b_{max}} p(c|a)c = \frac{1}{n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{b=b_{min}}^{b_{max}} \sum_{i=0}^c C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i} c.$$

$$\begin{aligned} \mathbb{D}(c|a) &= \sum_{c=0}^{a_{max}+b_{max}} p(c|a)c^2 - (\mathbb{E}(c|a))^2 = \\ &= \frac{1}{n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{b=b_{min}}^{b_{max}} \sum_{i=0}^c C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i} c^2 - (\mathbb{E}(c|a))^2. \end{aligned}$$

$$\mathbb{E}(c|b) = \sum_{c=0}^{a_{max}+b_{max}} p(c|b)c = \frac{1}{n_a} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{i=0}^c C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i} c.$$

$$\begin{aligned}
\mathbb{D}(c|b) &= \sum_{c=0}^{a_{max}+b_{max}} p(c|b)c^2 - (\mathbb{E}(c|b))^2 = \\
&= \frac{1}{n_a} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{i=0}^c C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i} c^2 - (\mathbb{E}(c|b))^2. \\
\mathbb{E}(c|d) &= \sum_{c=0}^{a_{max}+b_{max}} p(c|d)c = \sum_{c=0}^{a_{max}+b_{max}} \frac{p(d|c)p(c)}{p(d)} c. \\
\mathbb{D}(c|d) &= \sum_{c=0}^{a_{max}+b_{max}} p(c|d)c^2 - (\mathbb{E}(c|d))^2 = \sum_{c=0}^{a_{max}+b_{max}} \frac{p(d|c)p(c)}{p(d)} c^2 - (\mathbb{E}(c|d))^2. \\
\mathbb{E}(c|a, b) &= \sum_{c=0}^{a_{max}+b_{max}} p(c|a, b)c = \sum_{c=0}^{a_{max}+b_{max}} \sum_{i=0}^c C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i} c. \\
\mathbb{D}(c|a, b) &= \sum_{c=0}^{a_{max}+b_{max}} p(c|a, b)c^2 - (\mathbb{E}(c|a, b))^2 = \\
&= \sum_{c=0}^{a_{max}+b_{max}} \sum_{i=0}^c C_a^i p_1^i (1-p_1)^{a-i} C_b^{c-i} p_2^{c-i} (1-p_2)^{b-c+i} c^2 - (\mathbb{E}(c|a, b))^2. \\
\mathbb{E}(c|a, b, d) &= \sum_{c=0}^{a_{max}+b_{max}} p(c|a, b, d)c = \frac{\sum_{c=0}^{a_{max}+b_{max}} p(d|c)p(c|a, b)p(a)p(b)c}{\sum_{c=0}^{a_{max}+b_{max}} p(d|c)p(c|a, b)p(a)p(b)}. \\
\mathbb{D}(c|a, b, d) &= \sum_{c=0}^{a_{max}+b_{max}} p(c|a, b, d)c^2 - (\mathbb{E}(c|a, b, d))^2 = \\
&= \frac{\sum_{c=0}^{a_{max}+b_{max}} p(d|c)p(c|a, b)p(a)p(b)c^2}{\sum_{c=0}^{a_{max}+b_{max}} p(d|c)p(c|a, b)p(a)p(b)} - (\mathbb{E}(c|a, b, d))^2.
\end{aligned}$$

Результаты вычислений приведены в табл. 3.

$\mathbb{E}c$	13.75	$\mathbb{D}c$	13.168
$\mathbb{E}(c a = \mathbb{E}a)$	13.7	$\mathbb{D}(c a = \mathbb{E}a)$	12.91
$\mathbb{E}(c b = \mathbb{E}b)$	13.75	$\mathbb{D}(c b = \mathbb{E}b)$	13.082
$\mathbb{E}(c d = \mathbb{E}d)$	13.896	$\mathbb{D}(c d = \mathbb{E}d)$	1.533
$\mathbb{E}(c a = \mathbb{E}a, b = \mathbb{E}b)$	13.7	$\mathbb{D}(c a = \mathbb{E}a, b = \mathbb{E}b)$	12.825
$\mathbb{E}(c a = \mathbb{E}a, b = \mathbb{E}b, d = \mathbb{E}d)$	13.891	$\mathbb{D}(c a = \mathbb{E}a, b = \mathbb{E}b, d = \mathbb{E}d)$	1.529

Таблица 3: Математические ожидания и дисперсии для распределений  $p(c)$ ,  $p(c|a)$ ,  $p(c|b)$ ,  $p(c|d)$ ,  $p(c|a, b)$ ,  $p(c|a, b, d)$ , модель 1.

Действительно, информация о  $d$  значительно уменьшает дисперсию распределения, и в то же время информация об  $a$  и  $b$  влияет на нее слабо. Наименьшего значения дисперсии удается достичь при знании  $a$ ,  $b$  и  $d$  одновременно, что вполне логично. Матожидания в целом изменяются слабо, но при наличии информации о  $d$  они становятся чуть больше. Стоит отметить, что данные получены при фиксированных  $a$ ,  $b$ ,  $d$ , равных матожиданиям своих априорных распределений.

## Модель 2

Построим графики рассматриваемых распределений. Во многом они очень близки к соответствующим распределениям в случае модели 1.

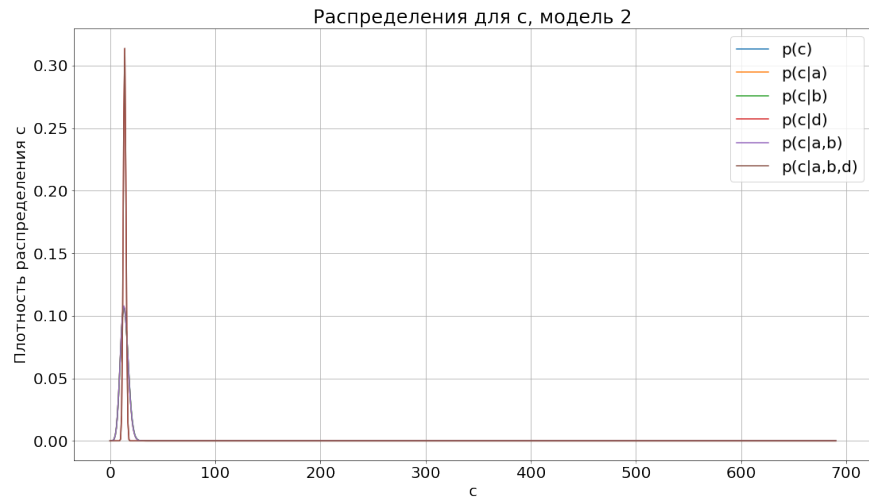


Рис. 9: Графики распределений для  $c$  на всем диапазоне значений  $c$ , модель 2.

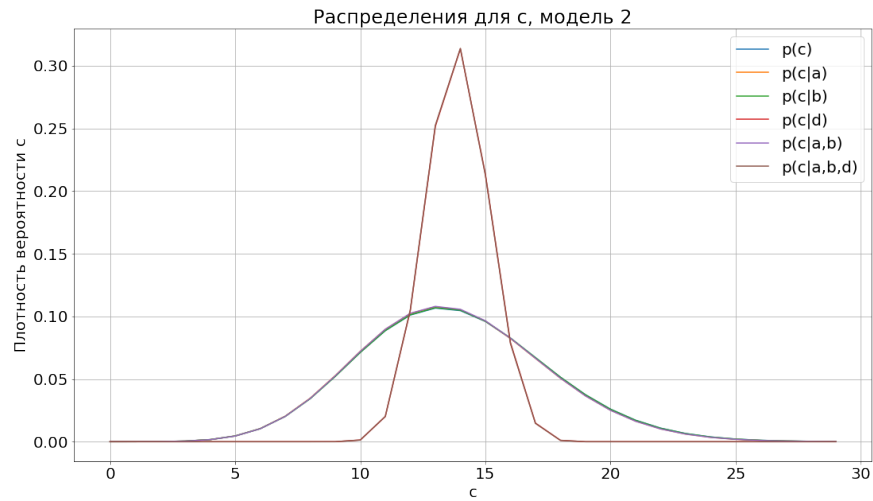


Рис. 10: Графики распределений для  $c$ , модель 2.

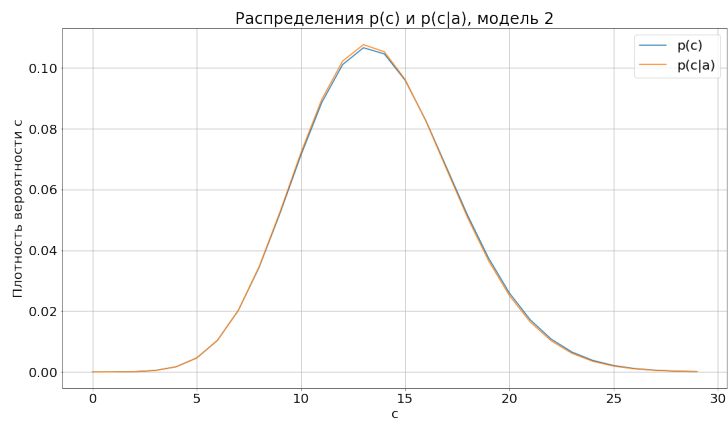


Рис. 11:  $p(c)$  и  $p(c|a)$ , модель 2.

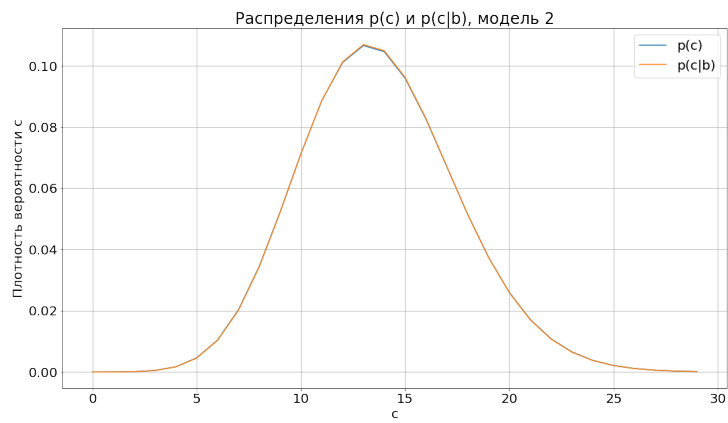


Рис. 12:  $p(c)$  и  $p(c|b)$ , модель 2.

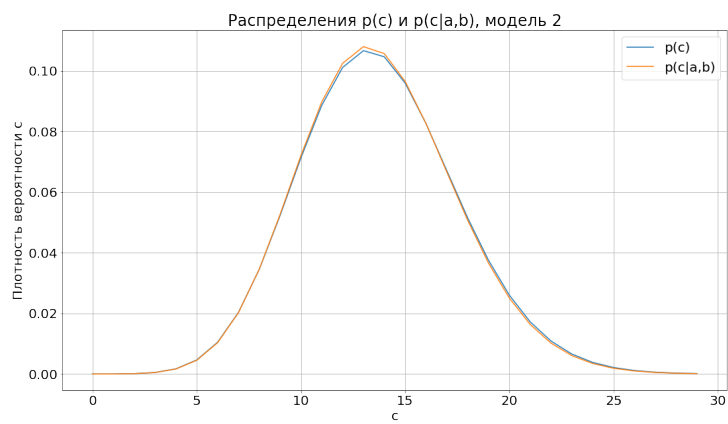


Рис. 13:  $p(c)$  и  $p(c|a,b)$ , модель 2.

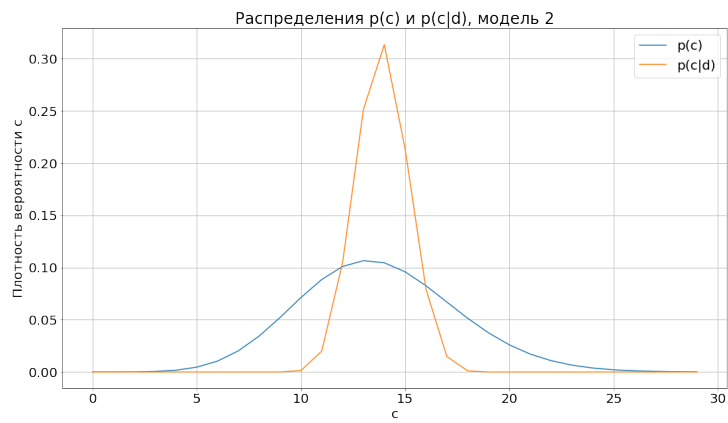


Рис. 14:  $p(c)$  и  $p(c|d)$ , модель 2.

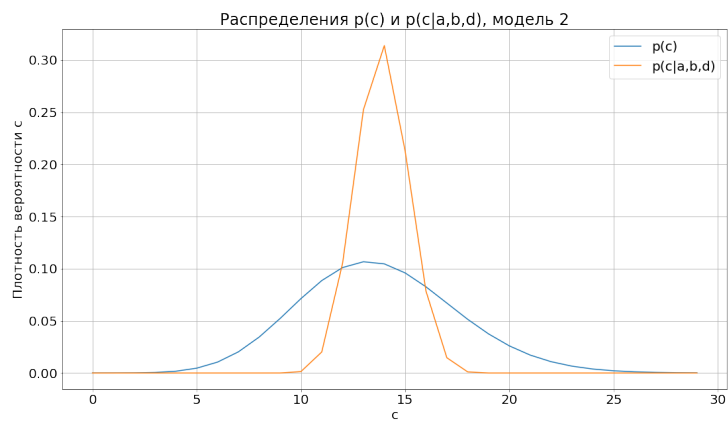


Рис. 15:  $p(c)$  и  $p(c|a, b, d)$ , модель 2.

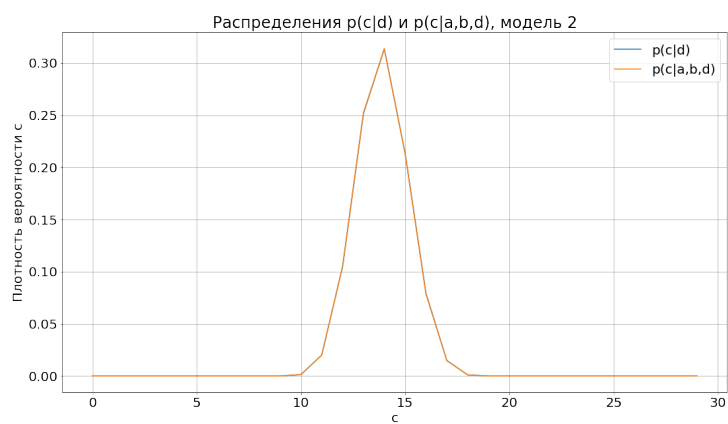


Рис. 16:  $p(c|d)$  и  $p(c|a, b, d)$ , модель 2.

В целом ситуация мало отличается от модели 1, и все наблюдения, выводы и замечания для нее верны и для модели 2.

Найдем матожидания и дисперсии для рассматриваемых распределений, приведем формулы для них.

$$\mathbb{E}c = \sum_{c=0}^{a_{max}+b_{max}} p(c)c = \frac{1}{n_a n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)} c.$$

$$\mathbb{D}c = \sum_{c=0}^{a_{max}+b_{max}} p(c)c^2 - (\mathbb{E}c)^2 = \frac{1}{n_a n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)} c^2 - (\mathbb{E}c)^2.$$

$$\mathbb{E}(c|a) = \sum_{c=0}^{a_{max}+b_{max}} p(c|a)c = \frac{1}{n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{b=b_{min}}^{b_{max}} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)} c.$$

$$\mathbb{D}(c|a) = \sum_{c=0}^{a_{max}+b_{max}} p(c|a)c^2 - (\mathbb{E}(c|a))^2 = \frac{1}{n_b} \sum_{c=0}^{a_{max}+b_{max}} \sum_{b=b_{min}}^{b_{max}} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)} c^2 - (\mathbb{E}(c|a))^2.$$

$$\mathbb{E}(c|b) = \sum_{c=0}^{a_{max}+b_{max}} p(c|b)c = \frac{1}{n_a} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)} c.$$

$$\mathbb{D}(c|b) = \sum_{c=0}^{a_{max}+b_{max}} p(c|b)c^2 - (\mathbb{E}(c|b))^2 = \frac{1}{n_a} \sum_{c=0}^{a_{max}+b_{max}} \sum_{a=a_{min}}^{a_{max}} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)} c^2 - (\mathbb{E}(c|b))^2.$$

$$\mathbb{E}(c|d) = \sum_{c=0}^{a_{max}+b_{max}} p(c|d)c = \sum_{c=0}^{a_{max}+b_{max}} \frac{p(d|c)p(c)}{p(d)} c.$$

$$\mathbb{D}(c|d) = \sum_{c=0}^{a_{max}+b_{max}} p(c|d)c^2 - (\mathbb{E}(c|d))^2 = \sum_{c=0}^{a_{max}+b_{max}} \frac{p(d|c)p(c)}{p(d)} c^2 - (\mathbb{E}(c|d))^2.$$

$$\mathbb{E}(c|a, b) = \sum_{c=0}^{a_{max}+b_{max}} p(c|a, b)c = \sum_{c=0}^{a_{max}+b_{max}} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)} c.$$

$$\mathbb{D}(c|a, b) = \sum_{c=0}^{a_{max}+b_{max}} p(c|a, b)c^2 - (\mathbb{E}(c|a, b))^2 = \sum_{c=0}^{a_{max}+b_{max}} \frac{(ap_1 + bp_2)^c}{c!} e^{-(ap_1 + bp_2)} c^2 - (\mathbb{E}(c|a, b))^2.$$

$$\mathbb{E}(c|a, b, d) = \sum_{c=0}^{a_{max}+b_{max}} p(c|a, b, d)c = \frac{\sum_{c=0}^{a_{max}+b_{max}} p(d|c)p(c|a, b)p(a)p(b)c}{\sum_{c=0}^{a_{max}+b_{max}} p(d|c)p(c|a, b)p(a)p(b)}.$$

$$\mathbb{D}(c|a, b, d) = \sum_{c=0}^{a_{max}+b_{max}} p(c|a, b, d)c^2 - (\mathbb{E}(c|a, b, d))^2 = \frac{\sum_{c=0}^{a_{max}+b_{max}} p(d|c)p(c|a, b)p(a)p(b)c^2}{\sum_{c=0}^{a_{max}+b_{max}} p(d|c)p(c|a, b)p(a)p(b)} - (\mathbb{E}(c|a, b, d))^2.$$

Результаты вычислений приведены в табл. 4.

$\mathbb{E}c$	13.75	$\mathbb{D}c$	14.048
$\mathbb{E}(c a = \mathbb{E}a)$	13.7	$\mathbb{D}(c a = \mathbb{E}a)$	13.785
$\mathbb{E}(c b = \mathbb{E}b)$	13.75	$\mathbb{D}(c b = \mathbb{E}b)$	13.962
$\mathbb{E}(c d = \mathbb{E}d)$	13.894	$\mathbb{D}(c d = \mathbb{E}d)$	1.544
$\mathbb{E}(c a = \mathbb{E}a, b = \mathbb{E}b)$	13.7	$\mathbb{D}(c a = \mathbb{E}a, b = \mathbb{E}b)$	13.7
$\mathbb{E}(c a = \mathbb{E}a, b = \mathbb{E}b, d = \mathbb{E}d)$	13.889	$\mathbb{D}(c a = \mathbb{E}a, b = \mathbb{E}b, d = \mathbb{E}d)$	1.54

Таблица 4: Математические ожидания и дисперсии для распределений  $p(c)$ ,  $p(c|a)$ ,  $p(c|b)$ ,  $p(c|d)$ ,  $p(c|a, b)$ ,  $p(c|a, b, d)$ , модель 2.

В общем ситуация очень близка к модели 1, наблюдаются все те же особенности и закономерности. Однако для всех распределений значения дисперсии стали немного больше, чем для модели 1, возможно, из-за приближения биномиального распределения пуассоновским, т. к. это приближение все же не является идеально точным. Информация о  $d$  все так же вносит наибольший вклад в уточнение прогноза для  $c$ .

В целом для обеих моделей информация о  $d$  сильнее всего влияет на распределение и прогноз  $c$ , информация об  $a$  и  $b$  влияет на распределение  $c$  незначительно, но информация об  $a$  изменяет распределение сильнее, чем информация о  $b$ . Можно сказать, что это вполне логично и согласуется с постановкой исследуемой вероятностной модели: число записавшихся на лекции ( $d$ ) действительно лучше позволяет оценить число реально присутствующих людей, чем знание количества людей на профильном ( $a$ ) и непрофильных ( $b$ ) факультетах, и знание количества людей с профильного факультета ( $a$ ) для оценки числа присутствующих полезнее, чем количество людей с непрофильных факультетов ( $b$ ), т. к. для них предмет лекции профильный, и они вероятнее всего на него придут, т. к. он ими более востребован и представляет для них интерес. Также вполне логична и естественна оценка параметров распределений математическими ожиданиями их априорных распределений, также согласуется с рассматриваемой моделью.

**4. Определить, какая из величин  $a$ ,  $b$ ,  $d$  вносит наибольший вклад в уточнение прогноза для величины  $c$  (в смысле дисперсии распределения). Для этого проверить верно ли, что  $\mathbb{D}(c|d) < \mathbb{D}(c|b)$  и  $\mathbb{D}(c|d) < \mathbb{D}(c|a)$  для любых допустимых значений  $a$ ,  $b$ ,  $d$ . Найти множество точек  $(a, b)$  таких, что  $\mathbb{D}(c|b) < \mathbb{D}(c|a)$ . Являются ли множества  $\{(a, b) | \mathbb{D}(c|b) < \mathbb{D}(c|a)\}$  и  $\{(a, b) | \mathbb{D}(c|b) \geq \mathbb{D}(c|a)\}$  линейно разделимыми? Ответ должен быть обоснован!**

Рассмотрим распределения  $p(c|a)$ ,  $p(c|b)$ ,  $p(c|d)$  при всевозможных допустимых значениях  $a$ ,  $b$  и  $d$ , найдем их дисперсии. Отметим, что в предыдущем пункте значения  $a$ ,  $b$  и  $d$  были конкретными и фиксированными. Также возникает проблема при вычислении  $p(c|d)$  для полного диапазона значений  $d$ , равного  $[0, 2(a_{\max} + b_{\max})]$ . Т. к. для расчетов в  $p(c|d)$  применяется теорема Байеса, а при достаточно больших  $d$  вероятность  $p(d)$  в знаменателе мала и близка к нулю (соответственно, и  $p(d|c)p(c)$  в числителе тоже близко к нулю), то из-за недостаточной точности вычислений на компьютере происходит деление 0 на 0. Для решения данной проблемы к нулевым значениям  $p(d)$  прибавлялся машинный эpsilon.  $p(c|d)$  при соответствующих  $d$  получалось равным 0, и  $\mathbb{D}(c|d)$  в таких случаях получалась нулевой, но вообще говоря в формуле для  $\mathbb{D}(c|d)$  будут произведения неопределенности  $\frac{0}{0}$  на достаточно большое  $d$ , и нулевое значение дисперсии вообще говоря некорректно. Однако данная проблема определяется ограничениями, связанными с машинной точностью. В целом лишь отчасти можно сказать, что присутствие нулевых дисперсий при больших  $d$  логично и согласуется с моделью, т. к. при



больших  $d$ , согласно формулам в моделях 1 и 2,  $c$  может принимать небольшое число достаточно больших значений, и дисперсия его прогноза может быть мала. В остальном же это не вполне корректно. Получаемый результат при вычислении  $\mathbb{D}(c|d)$  таким способом все же не является точным в силу недостаточной точности вычислений на компьютере (`scipy` порой с самого начала возвращает нулевые вероятности из-за недостаточной точности). Отметим, что использование логарифмов вероятностей в процессе подсчета распределений не позволило избавиться от данной проблемы, а также привело к увеличению времени работы функций из-за необходимости перехода от логарифмов обратно к вероятностям.

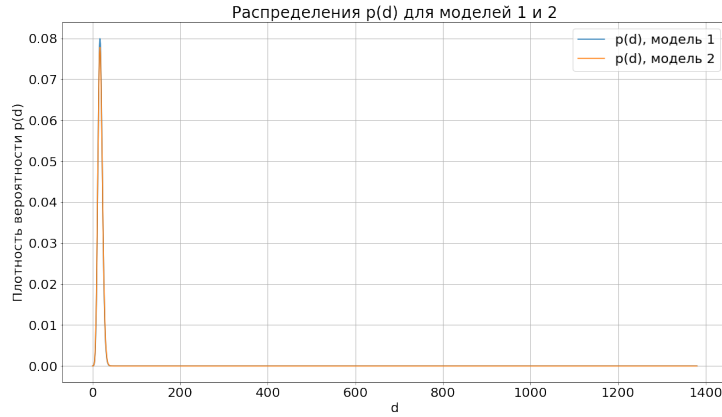


Рис. 17: Распределение  $p(d)$  для обеих моделей. При больших  $d$  значения  $p(d)$  практически равны 0, из-за чего при расчете  $p(c|d)$  по формуле Байеса возникает деление на 0.

## Модель 1

В результате экспериментов было получено, что:

$$\begin{aligned} \min\{\mathbb{D}(c|a)\} &= 12.28, \\ \max\{\mathbb{D}(c|a)\} &= 13.63, \\ \min\{\mathbb{D}(c|b)\} &= 12.587, \\ \max\{\mathbb{D}(c|b)\} &= 13.578, \\ \min\{\mathbb{D}(c|d)\} &= 0, \\ \max\{\mathbb{D}(c|d)\} &= 10.299. \end{aligned}$$

Таким образом, для модели 1  $\mathbb{D}(c|d) < \mathbb{D}(c|a)$  и  $\mathbb{D}(c|d) < \mathbb{D}(c|b)$  при любых значениях  $a$ ,  $b$  и  $d$ , т. е.  $d$  вносит наибольший вклад в уточнение прогноза для  $c$  в смысле дисперсии распределения.

Для большей наглядности можно изобразить дисперсии при разных значениях  $a$ ,  $b$  и  $d$  графически, однако рассматриваемые распределения зависят от разных параметров, количество их допустимых значений различно, и в итоге для каждого распределения будет рассматриваться различное число точек в различных координатах.

Для  $\mathbb{D}(c|a)$  и  $\mathbb{D}(c|b)$  наблюдается линейный рост с ростом значений параметров, для  $\mathbb{D}(c|d)$  зависимость уже нелинейная, и при больших  $d$   $\mathbb{D}(c|d)$  становится близким или равным 0, в основном в силу упомянутой недостаточной точности вычислений на компьютере.

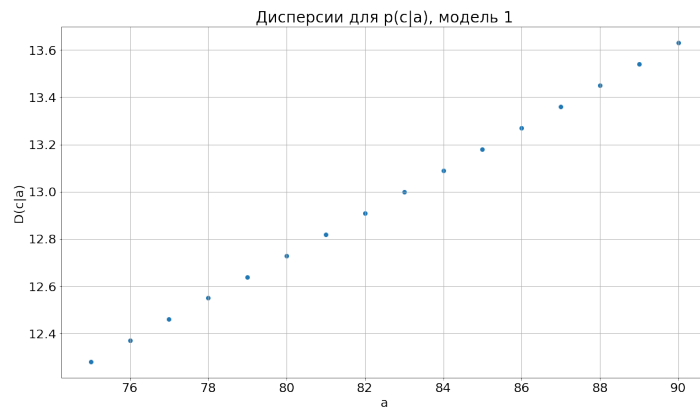


Рис. 18: Дисперсия  $\mathbb{D}(c|a)$  , модель 1.

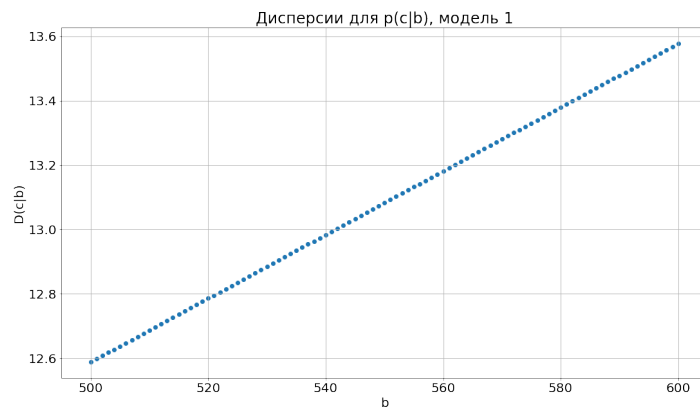


Рис. 19: Дисперсия  $\mathbb{D}(c|b)$  , модель 1.

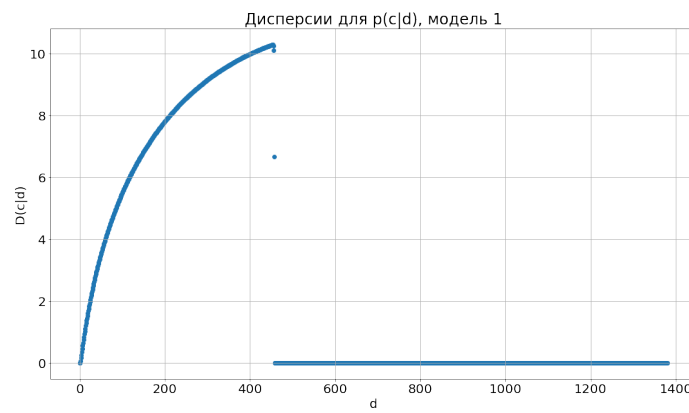


Рис. 20: Дисперсия  $\mathbb{D}(c|d)$  , модель 1.

## Модель 2

Для модели 2 было получено:

$$\min\{\mathbb{D}(c|a)\} = 13.085,$$

$$\max\{\mathbb{D}(c|a)\} = 14.585,$$

$$\min\{\mathbb{D}(c|b)\} = 13.462,$$

$$\max\{\mathbb{D}(c|b)\} = 14.462,$$

$$\min\{\mathbb{D}(c|d)\} = 0,$$

$$\max\{\mathbb{D}(c|d)\} = 17.01.$$

Для модели 2 условие  $\mathbb{D}(c|d) < \mathbb{D}(c|a)$ ,  $\mathbb{D}(c|d) < \mathbb{D}(c|b)$  уже не выполняется для всех значений  $a$ ,  $b$ ,  $d$ , даже без учета случаев возникновения нулевых значений  $\mathbb{D}(c|d)$  из-за машинной точности. Таким образом, при некоторых  $a$ ,  $b$  и  $d$  наибольший вклад в уточнение прогноза вносит  $a$  или  $b$ , а не  $d$ . По аналогии с предыдущей моделью можно рассмотреть графики дисперсий, во многом они схожи с рассмотренной первой моделью, и наблюдаемые в них зависимости те же.

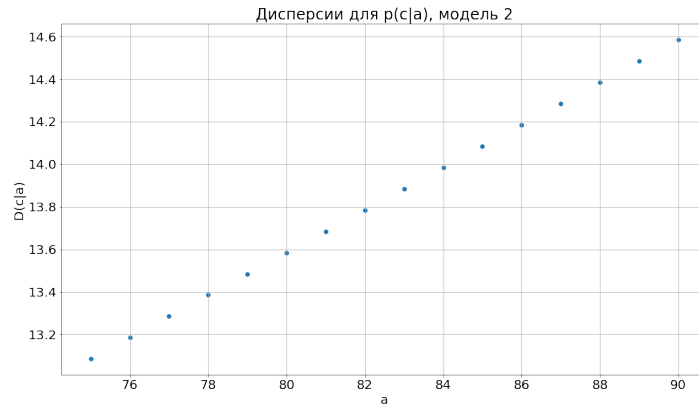


Рис. 21: Дисперсия  $\mathbb{D}(c|a)$ , модель 2.

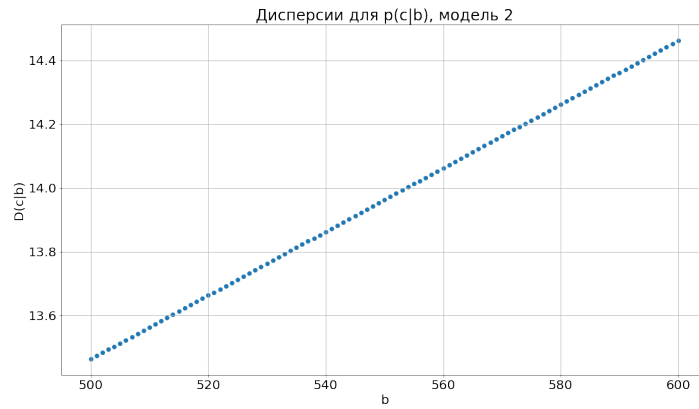


Рис. 22: Дисперсия  $\mathbb{D}(c|b)$ , модель 2.

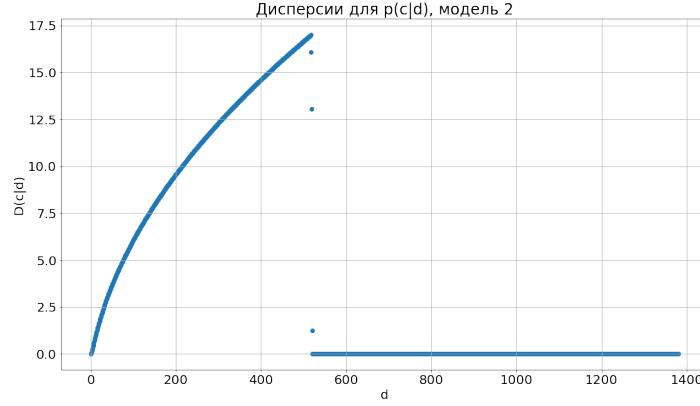


Рис. 23: Дисперсия  $\mathbb{D}(c|d)$ , модель 2.

Перейдем к рассмотрению множеств  $\{(a, b) | \mathbb{D}(c|b) < \mathbb{D}(c|a)\}$  и  $\{(a, b) | \mathbb{D}(c|b) \geq \mathbb{D}(c|a)\}$ . Изобразим соответствующие точки при всевозможных значениях  $a$  и  $b$ .

### Модель 1

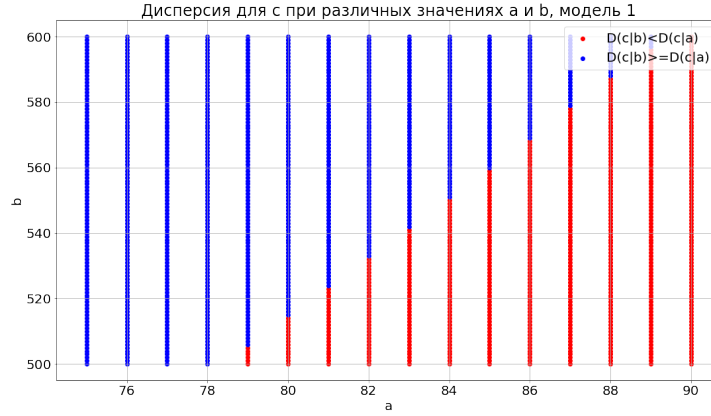


Рис. 24: Множества  $\{(a, b) | \mathbb{D}(c|b) < \mathbb{D}(c|a)\}$  и  $\{(a, b) | \mathbb{D}(c|b) \geq \mathbb{D}(c|a)\}$ , модель 1.

Действительно, судя по графикам, рассматриваемые множества являются линейно разделимыми в координатах  $(a, b)$ . В качестве более строгого доказательства можно применить линейный классификатор для разделения множества точек. В выборке точек в качестве признаков можно взять значения  $a$  и  $b$ , в качестве целевой переменной – индикатор того, что  $\mathbb{D}(c|b) < \mathbb{D}(c|a)$ . Если рассматриваемые множества линейно разделимы, то в ходе обучения классификатора удастся подобрать коэффициенты, определяющие разделяющую прямую в пространстве  $(a, b)$ , разбивающую точки с высокой точностью. Будем использовать алгоритм логистической регрессии, качество классификации будем оценивать по out-of-fold предсказаниям, т. е. разобьем выборку на некоторое число фолдов (в проводимых экспериментах на 5, как известная практическая рекомендация), и для каждого из них будем строить предсказания, обучая отдельный классификатор на всей остальной выборке. Метрикой качества выберем точность (ассигасу).

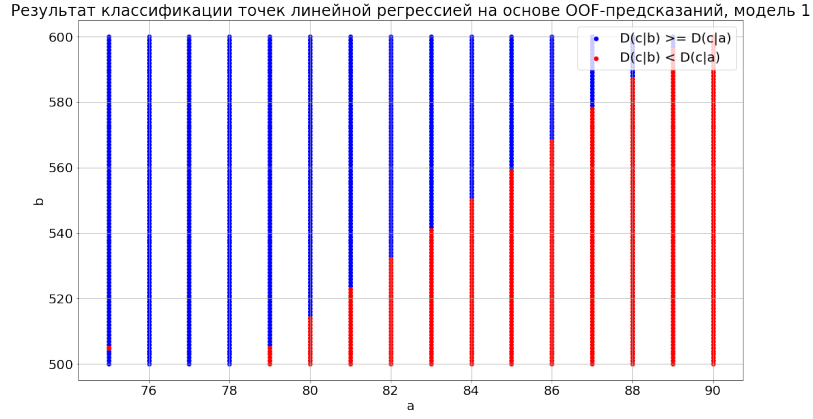


Рис. 25: Результат классификации точек  $(a, b)$  алгоритмом логистической регрессии, модель 1.

В результате действительно удалось качественно разделить множества точек прямой, предсказания алгоритма почти полностью соответствуют истинному разделению. Средняя точность классификации по фолдам составила 0.998. Таким образом, подтверждается гипотеза о линейной разделимости рассматриваемых множеств.

## Модель 2

Изобразим точки на плоскости  $(a, b)$ .

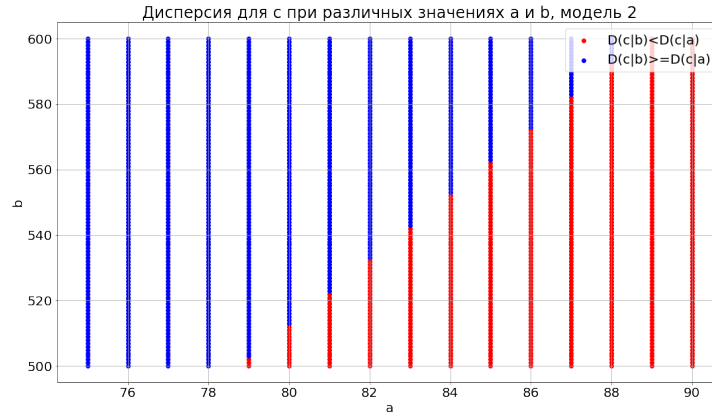


Рис. 26: Множества  $\{(a, b) | \mathbb{D}(c|b) < \mathbb{D}(c|a)\}$  и  $\{(a, b) | \mathbb{D}(c|b) \geq \mathbb{D}(c|a)\}$ , модель 2.

По графику также можно сказать, что соответствующие множества точек линейно разделимы. По аналогии с предыдущей моделью так же можем применить линейный классификатор и рассмотреть результат.

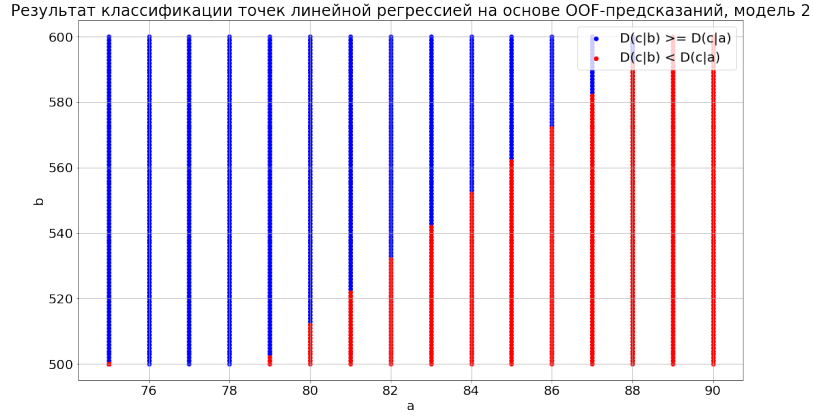


Рис. 27: Результат классификации точек  $(a, b)$  алгоритмом логистической регрессии, модель 2.

Аналогично удалось получить качественное разбиение плоскости прямой, близкое к истинному разбиению, средняя точность по фолдам составила 0.999, что свидетельствует о верности предположения о линейной разделимости исследуемых множеств.

## 5. Провести временные замеры по оценке всех необходимых распределений $p(c)$ , $p(c|a)$ , $p(c|b)$ , $p(c|d)$ , $p(c|a, b)$ , $p(c|a, b, d)$ , $p(d)$ .

Будем замерять время оценки всех распределений, указывая в качестве параметров  $a, b, d$  матожидания их априорных распределений и полный диапазон их допустимых значений. Время будем усреднять по 7 запускам соответствующих функций (используя `%timeit` в `Jupyter Notebook`). Результаты приведены в табл. 5.

	Модель 1	Модель 2
$p(c)$	107 мс	109 мс
$p(c a = \mathbb{E}a)$	74 мс	75 мс
$p(c a)$	112 мс	113 мс
$p(c b = \mathbb{E}b)$	7.02 мс	8.15 мс
$p(c b)$	110 мс	115 мс
$p(c d = \mathbb{E}d)$	177 мс	181 мс
$p(c d)$	179 мс	182 мс
$p(c a = \mathbb{E}a, b = \mathbb{E}b)$	3.78 мс	3.85 мс
$p(c a, b)$	110 мс	114 мс
$p(c a = \mathbb{E}a, b = \mathbb{E}b, d = \mathbb{E}d)$	71.7 мс	71.8 мс
$p(c a, b, d)$	10.9 с	11.1 с
$p(d)$	186 мс	192 мс

Таблица 5: Время оценки распределений  $p(c)$ ,  $p(c|a)$ ,  $p(c|b)$ ,  $p(c|d)$ ,  $p(c|a, b)$ ,  $p(c|a, b, d)$ ,  $p(d)$ , модели 1 и 2.

В целом время расчетов для обеих моделей почти не отличается, модель 2 работает немного дольше, вероятно, из-за вычисления экспоненты.

**6. Используя результаты всех предыдущих пунктов, сравнить две модели. Показать где максимально проявляется разница между ними (привести конкретный пример, не обязательно из экспериментов выше). Объяснить причины подобного результата.**

В целом исходя из определения моделей и результатов экспериментов, можно сказать, что модели 1 и 2 очень близки друг к другу. Однако модель 1 в определенном смысле является более точной, т. к. модель 2 по сути является ее приближением для случая больших  $a$  и  $b$  и малых  $p_1$  и  $p_2$  (приближение биномиального распределения пуассоновским), и т. к. в результате экспериментов выяснилось, что значения дисперсии для различных распределений для модели 1 меньше, чем для модели 2. Для обеих из них в большинстве случаев наибольший вклад в уточнение прогноза  $c$  вносит величина  $d$ , и для модели 1 это верно всегда, т. е.  $\mathbb{D}(c|d) < \mathbb{D}(c|a)$ ,  $\mathbb{D}(c|d) < \mathbb{D}(c|b)$  для любых  $a, b, d$ , но для модели 2 это выполняется не всегда. Также информация об  $a$  для каждой из моделей несколько важнее для прогноза  $c$ , чем информация о  $b$ . Таким образом, более предпочтительной в плане точности прогноза видится модель 1, в том числе и с точки зрения времени работы (вторая модель вычисляется несколько дольше первой).

Что касается проявления существенных различий между моделями, то на основании проведенных исследований возникает предположение, что заметные отличия могут начать проявляться, когда для второй модели перестает выполняться условие  $\mathbb{D}(c|d) < \mathbb{D}(c|a)$ ,  $\mathbb{D}(c|d) < \mathbb{D}(c|b)$  (что, вероятно, является большим различием в поведении, т. к. для модели 1 это условие выполняется для любых  $a, b, d$ ). Исходя из графиков, можно положить, что это происходит, грубо говоря, после  $d = 200$ . Рассмотрим распределения  $p(c|d)$  и  $p(c|a, b, d)$  при различных соответствующих значениях  $d$  и сравним их. В принципе можно сказать, что значения  $a$  и  $b$  по результатам проведенных экспериментов не так сильно влияют на прогноз  $c$ , поэтому возьмем их равными  $a = \mathbb{E}a$ ,  $b = \mathbb{E}b$  (округленными до ближайшего целого, по аналогии с предыдущими пунктами).

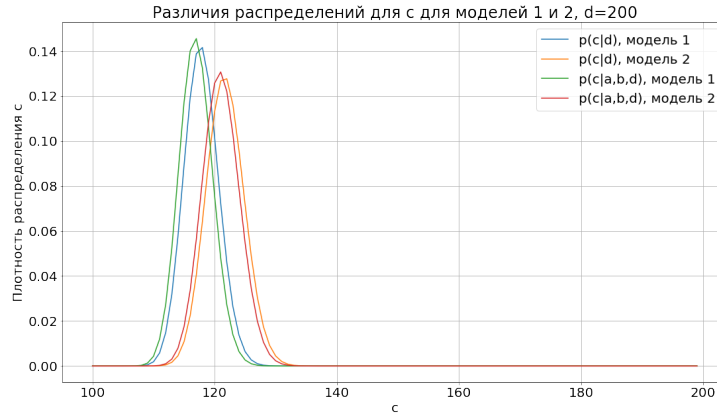


Рис. 28: Распределения  $p(c|d)$  и  $p(c|a, b, d)$  для моделей 1 и 2,  $d = 200$ ,  $a = \mathbb{E}a$ ,  $b = \mathbb{E}b$ .

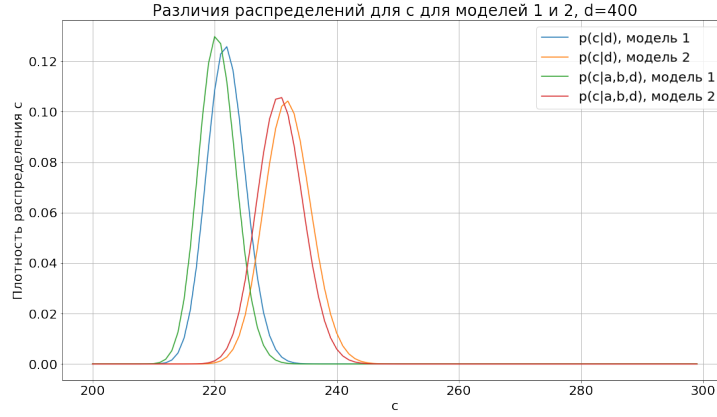


Рис. 29: Распределения  $p(c|d)$  и  $p(c|a, b, d)$  для моделей 1 и 2,  $d = 400$ ,  $a = \mathbb{E}a$ ,  $b = \mathbb{E}b$ .

Таким образом, распределения начинают сильнее различаться. С ростом  $d$  они начинают сильнее «разъезжаться», для модели 2 матожидание становится больше, чем для модели 1, и их разность становится больше. Также становится меньше значение максимума плотности, становится больше дисперсия. Но форма распределений для обеих моделей в целом остается одной и той же. Таким образом, можно сказать, что с ростом  $d$  в случае  $p(c|d)$  и  $p(c|a, b, d)$  модель 2 в отличие от модели 1 становится менее точной и более склонной к перепрогнозу  $c$ . В целом это может быть связано с тем, что вторая модель является по сути приближением первой. При этом стоит помнить о том, что для еще больших значений  $d$  (примерно после  $d = 450$ ) начинает проявляться проблема нулевых значений  $\mathbb{D}(c|d)$ , связанная с недостаточной точностью вычислений на компьютере.