МГУ им.Ломоносова
Факультет ВМК кафедра ММП

Задание 1. Байесовские рассуждения

Курс: Байесовские методы в машинном обучении 2021

Батшева Анастасия 417 группа

Москва
2021

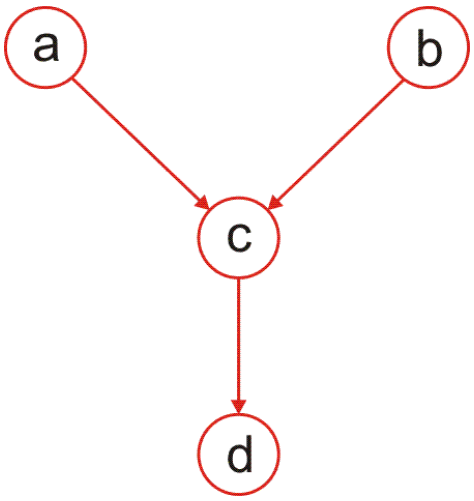
1 Формулировка задания

Приведена формулировка задания для варианта **1**:

Вероятностные модели посещаемости курса

Рассмотрим модель посещаемости студентами ВУЗа одной лекции по курсу:
a - количество студентов на профильном факультете
$\mathbf{p_1}$ - вероятность посещения лекции студентом профильного факультета
b - количество студентов других факультетах
$\mathbf{p_2}$ - вероятность посещения лекции студентом непрофильного факультета
c - количество студентов, действительно посетивших данную лекцию
d - количество студентов, *теоретически* посетивших данную лекцию
$\mathbf{p_3}$ - вероятность, с которой студент отмечает своего друга

Вероятностная модель 1:
$p(a, b, c, d) = p(d|c)p(c|a, b)p(a)p(b)$
$d|c \sim c + \text{Bin}(c, p_3)$
$c|a, b \sim \text{Bin}(a, p_1) + \text{Bin}(b, p_2)$
$a \sim \text{Unif}[a_{min}, a_{max}]$
$b \sim \text{Unif}[b_{min}, b_{max}]$



Вероятностная модель 2:
$p(a, b, c, d) = p(d|c)p(c|a, b)p(a)p(b)$
$d|c \sim c + \text{Bin}(c, p_3)$
$c|a, b \sim \text{Poiss}(ap_1 + bp_2)$
$a \sim \text{Unif}[a_{min}, a_{max}]$
$b \sim \text{Unif}[b_{min}, b_{max}]$

Провести следующие исследования для обеих моделей:
1. Вывести формулы для всех необходимых далее распределений аналитически.
2. Найти математические ожидания и дисперсии априорных распределений $p(a)$, $p(b)$, $p(c)$, $p(d)$.
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.
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 $\{\,(a, b) \mid \mathbb{D}[c|b] < \mathbb{D}[c|a]\,\}$ и $\{\,(a, b) \mid \mathbb{D}[c|b] \geq \mathbb{D}[c|a]\,\}$ линейно разделимыми? Ответ должен быть обоснован!
5. Провести временные замеры по оценке всех необходимых распределений $p(c)$, $p(c|a)$, $p(c|b)$, $p(c|d)$, $p(c|a, b)$, $p(c|a, b, d)$, $p(d)$.
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.

Взять в качестве диапазона допустимых значений для величины c интервал $[0, a_{max} + b_{max}]$, а для величины d – интервал $[0, 2(a_{max} + b_{max})]$.
Исследование должно быть выполнено на компьютере, однако за дополнительные аналитические выкладки в пунктах 2-4 будут ставиться дополнительные баллы. При оценке выполнения задания будет учитываться эффективность программного кода - любая из функций должна работать быстрее секунды на скалярных входах (для этого код должен реализовываться векторно). По всем пунктам задания должен быть проведен анализ результатов и сделаны выводы.

2 Входные данные

$a_{min} = 75, \quad a_{max} = 90, \quad b_{min} = 500, \quad b_{max} = 600, \quad p_1 = 0.1, \quad p_2 = 0.01, \quad p_3 = 0.3$

3 Теоретическая часть

Поскольку в нашем случае все случайные величины дискретны, то актуальны следующие формулы:

1. Если ξ - дискретная случайная величина $\sim p(\xi)$, тогда:
   (a) $\mathbb{E}\xi = \sum\limits_{\xi} p(\xi)\xi$
   (b) $\mathbb{D}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2$
2. Если $\xi_0, \xi_1, \ldots, \xi_n$ - $n+1$ дискретных с.в., то:
   $p(\xi_0) = \sum\limits_{\xi_1,\ldots,\xi_n} p(\xi_0|\xi_1, \ldots, \xi_n)p(\xi_1, \ldots, \xi_n)$ - **правило суммирования**
3. Если ψ - дискретная случайная величина, зависящая от ξ: $\sim p(\psi|\xi)$, тогда:
   (a) $\mathbb{E}\psi|\xi = \sum\limits_{\psi} p(\psi|\xi)\psi$
   (b) $\mathbb{D}\psi|\xi = \mathbb{E}\psi^2|\xi - (\mathbb{E}\psi|\xi)^2$
   (c) $p(\psi) =$ правило суммирования $= \sum\limits_{\xi} p(\psi|\xi)p(\xi) \Rightarrow$
   (d) $\mathbb{E}\psi = \sum\limits_{\psi} p(\psi)\psi = \sum\limits_{\psi}[\sum\limits_{\xi} p(\psi|\xi)p(\xi)]\psi = \sum\limits_{\xi}[\sum\limits_{\psi} p(\psi|\xi)\psi]p(\xi) = \mathbb{E}_\xi[\mathbb{E}\psi|\xi]$
   (e) $\mathbb{D}\psi = \mathbb{E}\psi^2 - (\mathbb{E}\psi)^2 = \mathbb{E}_\xi[\mathbb{E}\psi^2|\xi] - (\mathbb{E}\psi)^2 = \mathbb{E}_\xi[\mathbb{D}\psi|\xi] + \mathbb{E}_\xi[(\mathbb{E}\psi|\xi)^2] - (\mathbb{E}\psi)^2 = \mathbb{E}_\xi[\mathbb{D}\psi|\xi] + \mathbb{D}_\xi[\mathbb{E}\psi|\xi] + (\mathbb{E}_\xi[\mathbb{E}\psi|\xi])^2 - (\mathbb{E}\psi)^2 = \mathbb{E}_\xi[\mathbb{D}\psi|\xi] + \mathbb{D}_\xi[\mathbb{E}\psi|\xi]$

Более того, далее будут использованы формулы для мат.ожидания и дисперсии стандартных распределений, взятые из доступных источников.

1. Если $\xi \sim \text{Unif}[\xi_{min}, \xi_{max}]$, то $\mathbb{E}\xi = \frac{\xi_{max}+\xi_{min}}{2}$, $\mathbb{D}\xi = \frac{(\xi_{max}-\xi_{min}+1)^2-1}{12}$
2. Если $\xi \sim \text{Bin}(p)$, то $\mathbb{E}\xi = np$, $\mathbb{D}\xi = np(p-1)$
3. Если $\xi \sim \text{Poiss}(p)$, то $\mathbb{E}\xi = p$, $\mathbb{D}\xi = p$

3.1

Вывести формулы для всех необходимых далее распределений аналитически.

Применим формулы выше для определения распределений:

1. $\boxed{p(a) = \text{Unif}[a_{min}, a_{max}]}$
2. $\boxed{p(b) = \text{Unif}[b_{min}, b_{max}]}$
3. $\boxed{p(c|a,b) = \sum\limits_{a_i+b_i=c} p(a_i|a)p(b_i|b)}$

   $c = a_i + b_i$, где a_i, b_i: $\begin{cases} p(a_i|a) = \text{Bin}(a, p_1) \text{ or } \text{Poiss}(ap_1) \\ p(b_i|b) = \text{Bin}(b, p_2) \text{ or } \text{Poiss}(bp_2) \end{cases} \Rightarrow$

   $\mathbb{P}(c = c^*|a,b) = \sum\limits_{a_i^*+b_i^*=c^*} \mathbb{P}(a_i = a_i^*, b_i = b_i^*|a,b) =$ поскольку a и b независимы $=$

   $\sum\limits_{a_i^*+b_i^*=c^*} \mathbb{P}(a_i = a_i^*|a)\mathbb{P}(b_i = b_i^*|b) \Rightarrow p(c|a,b) = \sum\limits_{a_i+b_i=c} p(a_i|a)p(b_i|b)$
4. $\boxed{p(c) = \sum\limits_{a_i+b_i=c} p(a_i)p(b_i)}$

   $p(c) = \sum\limits_{a,b} p(c,a,b) = \sum\limits_{a,b} p(c|a,b)p(a)p(b) = \sum\limits_{a,b} \sum\limits_{a_i+b_i=c} p(a_i|a)p(b_i|b)p(a)p(b) =$

   Поменяем порядок суммирования $= \sum\limits_{a_i+b_i=c} \sum\limits_{a,b} p(a_i|a)p(b_i|b)p(a)p(b) =$

   $\sum\limits_{a_i+b_i=c} \sum\limits_{a} p(a_i, a) \sum\limits_{b} p(b_i, b) = \sum\limits_{a_i+b_i=c} \sum\limits_{a} p(a_i, a)p(b_i) = \sum\limits_{a_i+b_i=c} p(a_i)p(b_i)$
5. $\boxed{p(c|a) = \sum\limits_{b} p(c|a,b)p(b) = \sum\limits_{a_i+b_i=c} p(a_i|a)p(b_i)}$

   $p(c|a) = \frac{p(c,a)}{p(a)} = \frac{\sum_b p(c,a,b)}{p(a)} = \frac{\sum_b p(c|a,b)p(a)p(b)}{p(a)} = \sum\limits_{b} p(c|a,b)p(b) = \sum\limits_{a_i+b_i=c} \sum\limits_{b} p(a_i|a)p(b_i|b)p(b) =$

   $\sum\limits_{a_i+b_i=c} p(a_i|a)p(b_i)$
6. $\boxed{p(c|b) = \sum\limits_{a} p(c|a,b)p(a) = \sum\limits_{a_i+b_i=c} p(b_i|b)p(a_i)}$
7. $\boxed{p(d|c) = \text{Bin}(c + c_{const}, p3)}$

   $c_{const} = c$ - сдвиг по оси значений

   $d|c = c + c_i|c$, где $c_i|c \sim \text{Bin}(c, p3)$
8. $\boxed{p(d) = \sum\limits_{c} p(d|c)p(c)}$
9. $\boxed{p(c|d) = \frac{p(d|c)p(c)}{p(d)}}$
10. $\boxed{p(c|a,b,d) = \frac{p(d|c)p(c|a,b)}{p(d)}}$

3.2

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

1. $\boxed{\mathbb{E}a = 82.5,\ \mathbb{D}a = 21.25}$ (1),(2)

   $\mathbb{E}a = \frac{a_{max}+a_{min}}{2} = \frac{90+75}{2} = 82.5$
   $\mathbb{D}a = \frac{(a_{max}-a_{min}+1)^2-1}{12} = \frac{(90-75+1)^2-1}{12} = \frac{255}{12} = 21.25$

2. $\boxed{\mathbb{E}b = 550,\ \mathbb{D}a = 850}$ (1),(2)

   $\mathbb{E}b = \frac{b_{max}+b_{min}}{2} = \frac{600+500}{2} = 550$
   $\mathbb{D}b = \frac{(b_{max}-b_{min}+1)^2-1}{12} = \frac{(600-500+1)^2-1}{12} = \frac{10200}{12} = 850$

3. $\boxed{\mathbb{E}c = 13.75,\ \mathbb{D}d = 13.17}$ (1), $\boxed{\mathbb{E}c = 13.75,\ \mathbb{D}d = 14.05}$ (2)

   Так как $c = a_i + b_i$, а a_i, b_i независимы и зависят от a, b соответственно, то:

   (a) Поскольку у $\mathrm{Bin}(x, p)$ и у $\mathrm{Poiss}(xp)$ мат.ожидания одинаковы и равны xp, следующее верно для обеих моделей
   $\mathbb{E}c = \mathbb{E}a_i + \mathbb{E}b_i$
   $\begin{cases} \mathbb{E}a_i = \mathbb{E}_a[\mathbb{E}a_i|a] = \mathbb{E}_a[ap_1] = p_1\mathbb{E}a = 0.1 \cdot 82.5 = 8.25 \\ \mathbb{E}b_i = +\mathbb{E}_b[\mathbb{E}b_i|b] = \mathbb{E}_b[bp_2] = p_2\mathbb{E}b = 0.01 \cdot 550 = 5.5 \end{cases} \Rightarrow \mathbb{E}c = 13.75$

   (b) $\mathbb{D}c = \mathbb{D}a_i + \mathbb{D}b_i$
   $\begin{cases} \mathbb{D}a_i = \mathbb{E}_a[\mathbb{D}a_i|a] + \mathbb{D}_a[\mathbb{E}a_i|a] \\ \mathbb{D}b_i = \mathbb{E}_b[\mathbb{D}b_i|b] + \mathbb{D}_b[\mathbb{E}b_i|b] \end{cases}$
   Поскольку дисперсии (1) $\mathrm{Bin}(x, p) = xp(1-p)$ и (2) $\mathrm{Poiss}(xp) = xp$, то
   (1) $\begin{cases} \mathbb{D}a_i = \mathbb{E}_a[ap_1(1-p_1)] + \mathbb{D}_a[ap_1] \\ \mathbb{D}b_i = \mathbb{E}_b[bp_2(1-p_2)] + \mathbb{D}_b[bp_2] \end{cases} \Rightarrow \begin{cases} \mathbb{D}a_i = p_1(1-p_1)\mathbb{E}a + p_1^2\mathbb{D}a \\ \mathbb{D}b_i = p_2(1-p_2)\mathbb{E}b + p_2^2\mathbb{D}b \end{cases} \Rightarrow \begin{cases} \mathbb{D}a_i = 7.64 \\ \mathbb{D}b_i = 5.53 \end{cases}$
   (2) $\begin{cases} \mathbb{D}a_i = \mathbb{E}_a[ap_1] + \mathbb{D}_a[ap_1] \\ \mathbb{D}b_i = \mathbb{E}_b[bp_2] + \mathbb{D}_b[bp_2] \end{cases} \Rightarrow \begin{cases} \mathbb{D}a_i = p_1\mathbb{E}a + p_1^2\mathbb{D}a \\ \mathbb{D}b_i = p_2\mathbb{E}b + p_2^2\mathbb{D}b \end{cases} \Rightarrow \begin{cases} \mathbb{D}a_i = 8.46 \\ \mathbb{D}b_i = 5.58 \end{cases}$

   $\Rightarrow \mathbb{D}c = 13.17(1)$ и $\mathbb{D}c = 14.05(2)$

4. $\boxed{\mathbb{E}d = 17.88,\ \mathbb{D}d = 25.14}$ (1) и $\boxed{\mathbb{E}d = 17.88,\ \mathbb{D}d = 26.63}$ (2)

   (a) $\mathbb{E}d = \mathbb{E}_c[\mathbb{E}d|c] = \mathbb{E}_c[c + \mathbb{E}[c_i|c]] = \mathbb{E}c + \mathbb{E}_c[\mathbb{E}c_i|c] = \mathbb{E}c + p_3\mathbb{E}c = (1+p_3)\mathbb{E}c = 1.3 \cdot 13.75 = 17.875$

   (b) $\mathbb{D}d = \mathbb{E}_c[\mathbb{D}d|c] + \mathbb{D}_c[\mathbb{E}d|c] = \mathbb{E}_c[\mathbb{D}[c_i|c]] + \mathbb{D}_c[c + \mathbb{E}[c_i|c]] = \mathbb{E}_c[cp_3(1-p_3)] + \mathbb{D}_c[c + cp_3] = p_3(1-p_3)\mathbb{E}c + (1+p_3)^2\mathbb{D}c = 2.89 + 1.69\mathbb{D}c \Rightarrow$
   $\mathbb{D}c = 25.14(1)$ и $\mathbb{D}c = 26.63(2)$

Сравним с результатами эксперимента:

```
Prior distributions, model 1:
p(a)  E: 82.50   D: 21.25   time: 0.0001
p(b)  E: 550.00  D: 850.00  time: 0.0001
p(c)  E: 13.75   D: 13.17   time: 0.0277
p(d)  E: 17.87   D: 25.14   time: 0.0923

Prior distributions, model 2:
p(a)  E: 82.50   D: 21.25   time: 0.0000
p(b)  E: 550.00  D: 850.00  time: 0.0000
p(c)  E: 13.75   D: 14.05   time: 0.0109
p(d)  E: 17.88   D: 26.63   time: 0.0731
```

Как видно, значения совпадают. Экспериментальные мат.ожидания и дисперсии считались по единым формулам: $\mathbb{E}x = \sum\limits_x p(x)x$, $\mathbb{D}x = \mathbb{E}x^2 - (\mathbb{E}x)^2$

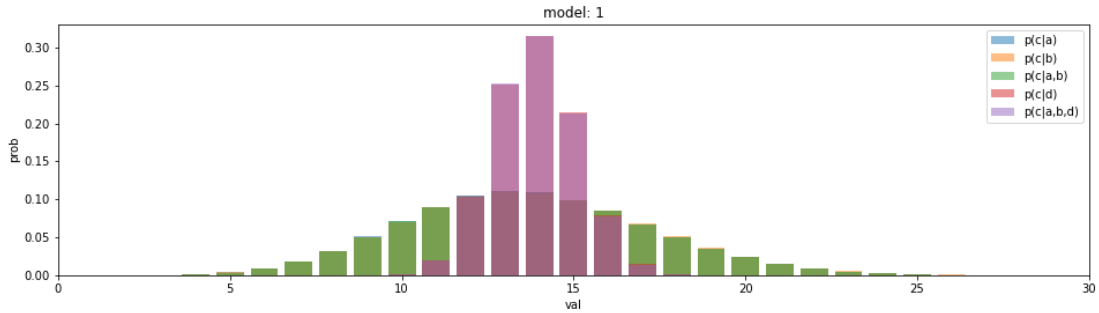
3.3

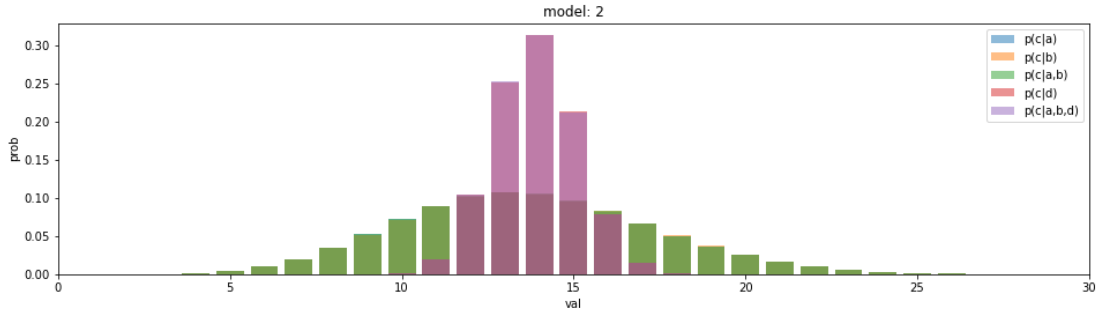
Пронаблюдать, как происходит уточнение прогноза для величины c по мере прихода новой косвенной информации. Для этого построить графики и найти мат.ожидание и дисперсию для распределений $p(c)$, $p(c|a)$, $p(c|b)$, $p(c|d)$, $p(c|a,b)$, $p(c|a,b,d)$ при параметрах a, b, d, равных мат.ожиданиям своих априорных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

```
Posterior distributions, model 1:
a = 82
b = 550
d = 18
p(c|a)     E: 13.70   D: 12.91   time: 0.0174
p(c|b)     E: 13.75   D: 13.08   time: 0.0042
p(c|a,b)   E: 13.70   D: 12.82   time: 0.0046
p(c|d)     E: 13.90   D: 1.53    time: 0.1173
p(c|a,b,d) E: 13.89   D: 1.53    time: 0.0763

Posterior distributions, model 2:
a = 82
b = 550
d = 18
p(c|a)     E: 13.70   D: 13.79   time: 0.0118
p(c|b)     E: 13.75   D: 13.96   time: 0.0056
p(c|a,b)   E: 13.70   D: 13.70   time: 0.0046
p(c|d)     E: 13.89   D: 1.54    time: 0.0937
p(c|a,b,d) E: 13.89   D: 1.54    time: 0.0734
```

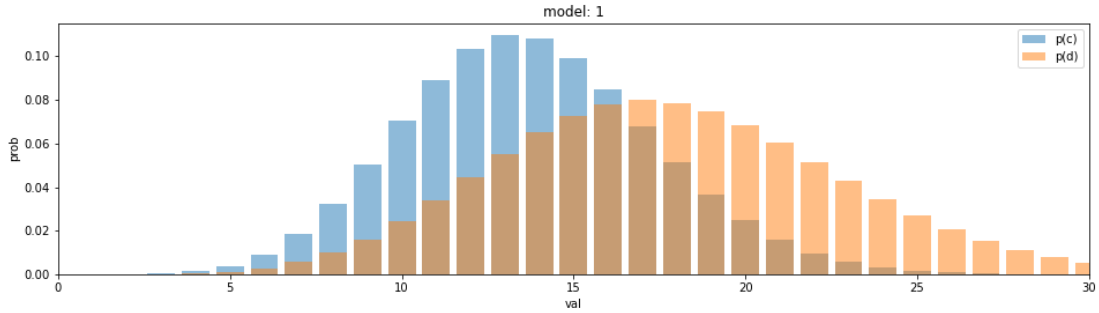
Судя по дисперсиям апостериорных распределений, лучше всех остальных прогнозирует c величина d, потому как дисперсии у распределений $p(c|d), p(c|a,b,d)$ существенно (почти в 10 раз) меньше, чем у других.

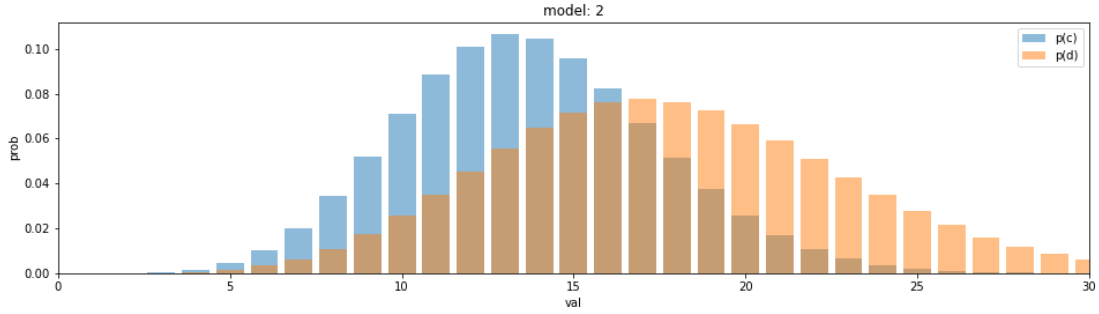




Видно, что $p(c|d), p(c|a,b,d)$ ближе к $\sigma(c)$, чем остальные, то есть именно они лучше всего уточняют величину c. При этом на самом деле дисперсия у $p(c|a,b,d)$ чуть меньше, чем у $p(c|d)$ потому как присутствует дополнительная информация о a,b, однако эта разница меньше одной сотой (первые два знака после запятой одинаковы).

Так же в задании необходимо построить график распределения $p(c)$. Ради интереса построим так же график для $p(d)$.





Получается довольно интересно: наглость студентов записывать своих друзей не катастрофически велика :) Именно поэтому величина d лучше всего предсказывает значение c. Инспектор, имея на руках список теоретически присутствующих, наверняка знает, что число практически присутствующих не может превышать d (студент, честно пришедший на лекцию, гарантированно отмечается), и при этом не меньше половины d, округленной в большую сторону, так как каждый студент не отмечает больше одного друга. Однако зная лишь общее число студентов на факультетах, число пришедших можно оценить от 0 (в худшем случае вроде субботы перед новогодними праздниками) до a + b (в лучшем случае, когда за посещение лекции ставят автоматы). Таким образом, знание числа отметившихся существенно сужает диапазон возможных значений величины c.

3.4

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 $\{\,(a,b) \mid \mathbb{D}[c|b] < \mathbb{D}[c|a]\,\}$ и $\{\,(a,b) \mid \mathbb{D}[c|b] \geq \mathbb{D}[c|a]\,\}$ линейно разделимыми? Ответ должен быть обоснован!

1. Вычислим дисперсии $\mathbb{D}[c|a], \mathbb{D}[c|b], \mathbb{D}[c|d]$ при всех возможных значениях параметров a, b, d для обеих моделей

2. Для каждой модели обозначим на плоскости a, b с помощью цвета множества:
   $(a,b)|\mathbb{D}[c|b] < \mathbb{D}[c|a]$ - точки (a,b) фиолетового цвета
   $(a,b)|\mathbb{D}[c|b] \geq \mathbb{D}[c|a]$ - точки (a,b) желтого цвета

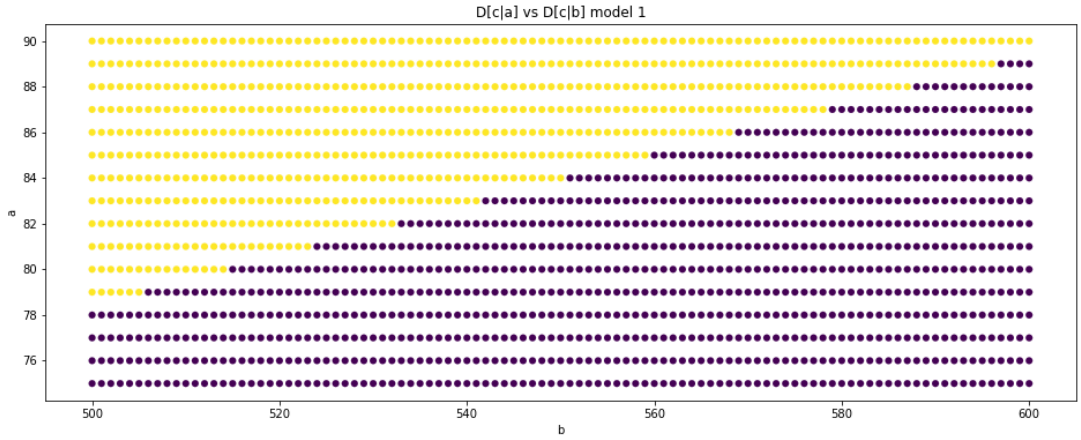
3. Проверим, верно ли, что $\mathbb{D}[c|d] < \mathbb{D}[c|b]$ и $\mathbb{D}[c|d] < \mathbb{D}[c|a]$ для любых допустимых значений a, b, d. Для этого сравним максимум $\mathbb{D}[c|d]$ среди всех значений d с минимумом $\mathbb{D}[c|b]$ и $\mathbb{D}[c|a]$ среди всех значений b и a соответственно.

Теоретически, множества разделимы, так как
$c = a_i + b_i$
$\mathbb{D}c|a = \mathbb{D}a_i|a + \mathbb{D}b_i = ap_1(1-p_1) + \mathbb{D}b_i$
$\mathbb{D}c|b = \mathbb{D}a_i + \mathbb{D}b_i|b = \mathbb{D}a_i + bp_2(1-p_2)$

$\mathbb{D}[c|b] < \mathbb{D}[c|a] \iff \mathbb{D}a_i + bp_2(1-p_2) < ap_1(1-p_1) + \mathbb{D}b_i \iff b < a\frac{p_1(1-p_1)}{p_2(1-p_2)} + \frac{(\mathbb{D}a_i - \mathbb{D}b_i)}{p_2(1-p_2)} = aw_1 + w_0$
Поэтому проверку на разделимость будем осуществлять с помощью этой прямой $b = aw_1 + w_0$

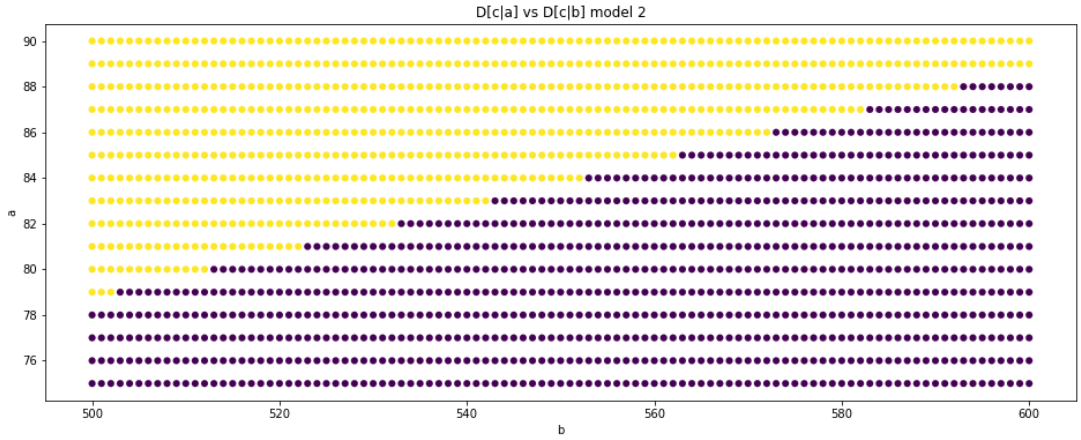
```
model 1:
max D[c|d] = 10.2987 < 12.2800 = min D[c|a]: True
max D[c|d] = 10.2987 < 12.5875 = min D[c|b]: True

line b = 9.09a + -212.88 is True separation line
```




```
model 2:
max D[c|d] = 12.8942 < 13.0850 = min D[c|a]: True
max D[c|d] = 12.8942 < 13.4625 = min D[c|b]: True

line b = 10.00a + -287.75 is True separation line
```



Отсюда видно, что:
1. $\mathbb{D}[c|d] < \mathbb{D}[c|b]$ и $\mathbb{D}[c|d] < \mathbb{D}[c|a]$ для любых допустимых значений a, b, d, что подтверждает результаты пункта 3.
2. Множества $(a, b)|\mathbb{D}[c|b] < \mathbb{D}[c|a]$ и $(a, b)|\mathbb{D}[c|b]\mathbb{D}[c|a]$ являются линейно разделимыми

3.5

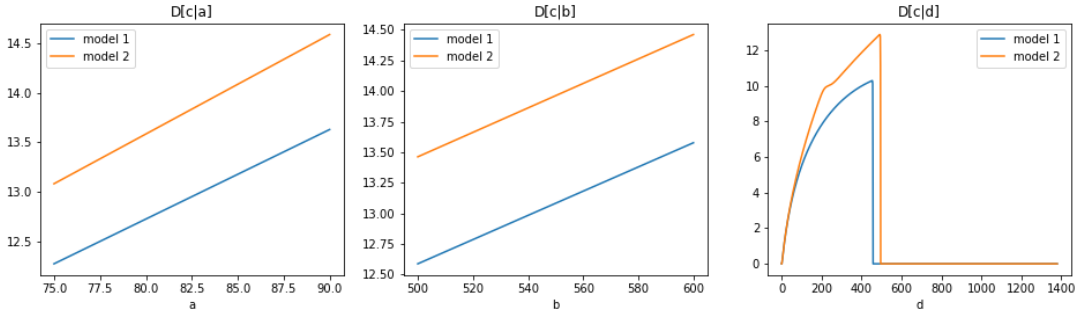
Провести временные замеры по оценке всех необходимых распределений $p(c), p(c|a), p(c|b)$**,** $p(c|d), p(c|a, b), p(c|a, b, d), p(d)$**.**

Все замеры времени были проведены в пункте 1.

3.6

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.

Изначально (в пункте 1) можно было видеть отличие дисперсий у распределений $p(c|a), p(c|b), p(c|d)$.



Однако основное отличие моделей - разные распределений $c|a, b$:
$c|a, b \sim \text{Bin}(a, p_1) + \text{Bin}(b, p_2)$ - Первая вероятностная модель
$c|a, b \sim \text{Poiss}(ap_1 + bp_2)$ - Вторая вероятностная модель
Посмотрим, насколько они отличаются визуально.

Так же известно, что сумма пуассоновских распределений дает пуассоновское, а вот сумма биномиальных не дает биномиального.

1. Доказательство первого факта довольно тривиальное (через функцию вероятности и свертку в биномиальные коэффициенты, было на семинаре).

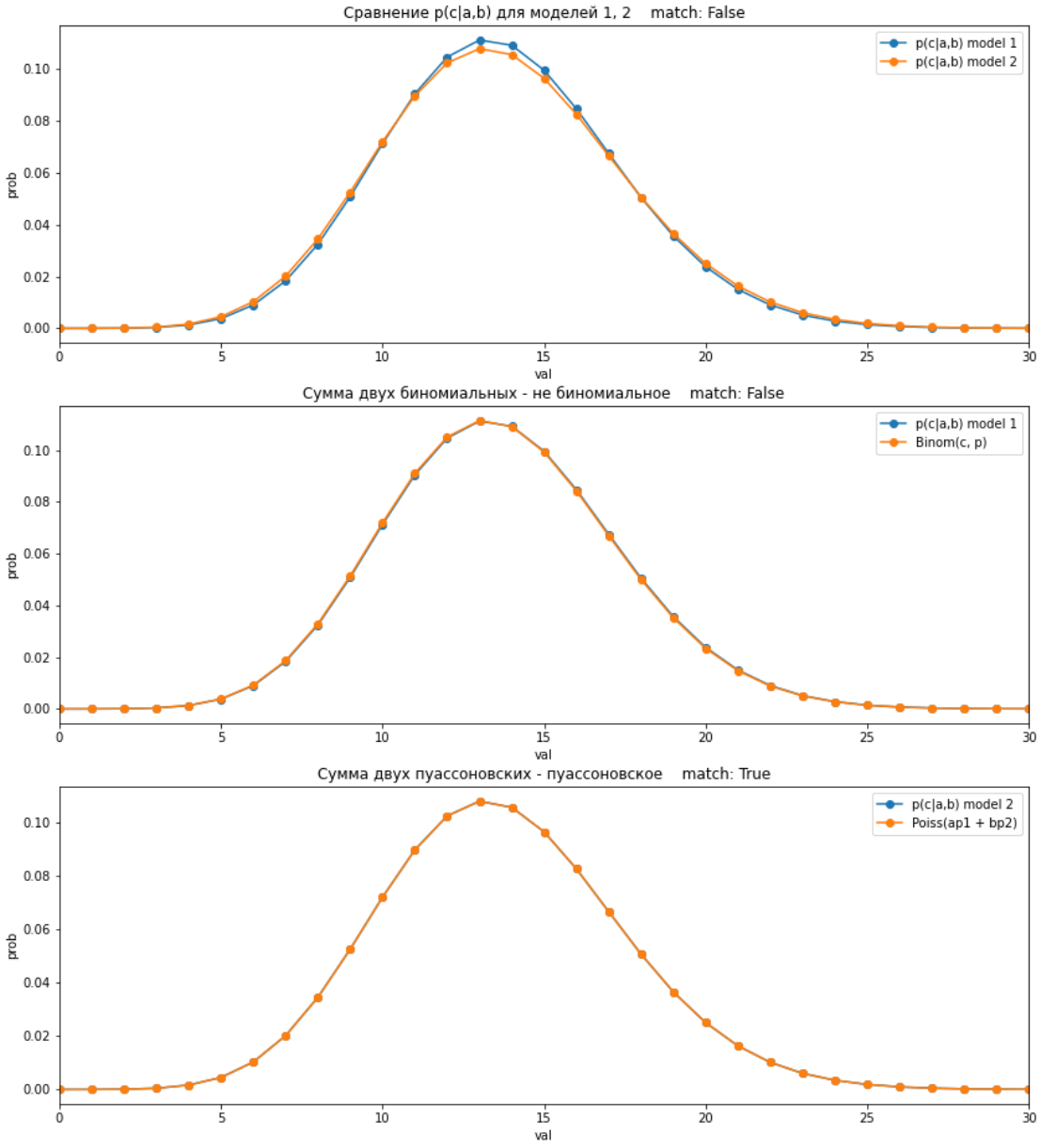
2. Допустим, что второй факт неверен, и можно найти биномиальное распределение $B(c, p)$, совпадающее с суммой данных биномиальных $B_1(a, p1)$, $B_2(b, p2)$ Тогда совпадают моменты:
$$\begin{cases} \mathbb{E}[B] = \mathbb{E}[B_1 + B_2] \\ \mathbb{D}[B] = \mathbb{D}[B_1 + B_2] \end{cases} \Rightarrow \begin{cases} cp = ap_1 + bp_2 \\ cp(1-p) = ap_1(1-p_1) + bp_2(1-p_2) \end{cases} \Rightarrow$$
$$\begin{cases} c = \frac{ap_1+bp_2}{p} \\ p = 1 - \frac{ap_1(1-p_1)+bp_2(1-p_2)}{ap_1+bp_2} \end{cases} \Rightarrow \begin{cases} c = \frac{ap_1+bp_2}{p} \\ p = \frac{ap_1^2+bp_2^2}{ap_1+bp_2} \end{cases} \Rightarrow \begin{cases} c = \frac{(ap_1+bp_2)^2}{ap_1^2+bp_2^2} \\ p = \frac{ap_1^2+bp_2^2}{ap_1+bp_2} \end{cases}$$

Проверим с помощью эксперимента, совпадают ли распределения:

1. $p(c|a, b)$ для модели 1 и $p(c|a, b)$ для модели 2

2. $\text{Poiss}(a, p_1) + \text{Poiss}(b, p_2)$ и $\text{Poiss}(ap_1 + bp_2)$

3. $\text{Bin}(a, p_1) + \text{Bin}(b, p_2)$ и $\text{Bin}(, p)$



Как видно, первые два распределения не совпадают (match = False), а последние - совпадают. Результаты эксперимента подтверждают теорию. А так же 2 модель "сохраняет"распределение относительно суммы, а 1 - нет.