Практическое задание 1. Байесовские рассуждения

Янаков Дмитрий Спартакович, 417 группа, 1 Вариант

3 октября, 2022

1 Вывод формул для распределений

$$p(a,b,c,d) = p(a)p(b)p(c|a,b)p(d|c)$$

Для обеих моделей:

$$p(a) = \frac{1}{a_{max} - a_{min} + 1}$$

$$p(b) = \frac{1}{b_{max} - b_{min} + 1}$$

Модель 1:

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

Модель 2:

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

Для обеих моделей:

$$p(c|a) = \sum_{b=b_{min}}^{b_{max}} p(c|a,b)p(b)$$

$$p(c|b) = \sum_{a=a_{min}}^{a_{max}} p(c|a,b)p(a)$$

$$p(c) = \sum_{a=a_{min}}^{a_{max}} \sum_{b=b_{min}}^{b_{max}} p(c|a,b)p(a)p(b)$$

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

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

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

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

То есть единственным отличием в данных моделях является подсчет распределения $p(c|a,b)$.

2 Математические ожидания и дисперсии априорных распределений

2.1 Математические ожидания

Для обеих моделей:

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

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

$$\mathbb{E}[c|a,b] = p_1 a + p_2 b$$
$$\mathbb{E}[d|c] = c + p_3 c$$

По свойству условного математического ожидания:

$$\mathbb{E}c = \mathbb{E}[\mathbb{E}[c|a,b]] = p_1\mathbb{E}a + p_2\mathbb{E}b$$

$$\mathbb{E}d = \mathbb{E}[\mathbb{E}[d|c]] = \mathbb{E}c + p_3\mathbb{E}c = (1+p_3)\mathbb{E}c$$

2.2 Дисперсии

Для обеих моделей:

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

Модель 1:

$$\mathbb{D}c = \mathbb{E}[\mathbb{D}[c|a,b]] + \mathbb{D}[\mathbb{E}[c|a,b]] = \mathbb{E}[p_1(1-p_1)a + p_2(1-p_2)b + \mathrm{Cov}(a,b)] + \mathbb{D}[p_1 a + p_2 b] =$$

$$= p_1(1-p_1)\mathbb{E}a + p_2(1-p_2)\mathbb{E}b + p_1^2\mathbb{D}a + p_2^2\mathbb{D}b$$

Модель 2:

$$\mathbb{D}c = \mathbb{E}[\mathbb{D}[c|a,b]] + \mathbb{D}[\mathbb{E}[c|a,b]] = \mathbb{E}[p_1 a + p_2 b + \mathrm{Cov}(a,b)] + \mathbb{D}[p_1 a + p_2 b] =$$

$$= p_1\mathbb{E}a + p_2\mathbb{E}b + p_1^2\mathbb{D}a + p_2^2\mathbb{D}b$$

$\mathrm{Cov}(a,b) = 0$, так как случайные величины a и b независимы.

Для обеих моделей:

$$\mathbb{D}d = \mathbb{E}[\mathbb{D}[d|c]] + \mathbb{D}[\mathbb{E}[d|c]] = \mathbb{E}[\mathbb{D}c + p_3(1-p_3)c + \mathrm{Cov}(c, Bin(c,p_3))] + \mathbb{D}[(1+p_3)c] =$$

$$= p_3(1-p_3)\mathbb{E}c + (1+p_3)^2\mathbb{D}c$$

$\mathbb{D}c = 0$, так как дисперсия берется по d, то есть в данном случае c – это константа, а дисперсия от константы равняется нулю. Аналогично для $\mathrm{Cov}(c, Bin(c,p_3))$, так как ковариация константы и случайной величины равняется нулю.

2.3 Числовые значения

Посчитаем числовые значения математических ожиданий и дисперсий априорных распределений при следующих параметрах: $a_{min} = 75$, $a_{max} = 90$, $b_{min} = 500$, $b_{max} = 600$, $p_1 = 0.1$, $p_2 = 0.01$, $p_3 = 0.3$.

Результаты приведены в таблице ниже.

	Модель 1	Модель 2
$\mathbb{E}a$	82.5	
$\mathbb{E}b$	550	
$\mathbb{E}c$	13.75	
$\mathbb{E}d$	17.875	
$\mathbb{D}a$	21.25	
$\mathbb{D}b$	850	
$\mathbb{D}c$	13.1675	14.0475
$\mathbb{D}d$	25.140575	26.627775

Таблица 1: Числовые значения мат.ожиданий и дисперсий априорных распределений

Видно, что первая модель лучше в смысле дисперсии для параметров c и d.

3 Уточнение прогноза

Выясним, как происходит уточнение прогноза для величины c по мере прихода новой косвенной информации. Для этого построим графики 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 (**Рис. 1**), а также найдем мат.ожидания и дисперсии соответствующих распределений (**Таблица 2**).

В графиках для распределений ось абсцисс урезана до значения $c = 30$, поскольку дальше идут значения близкие к нулю.

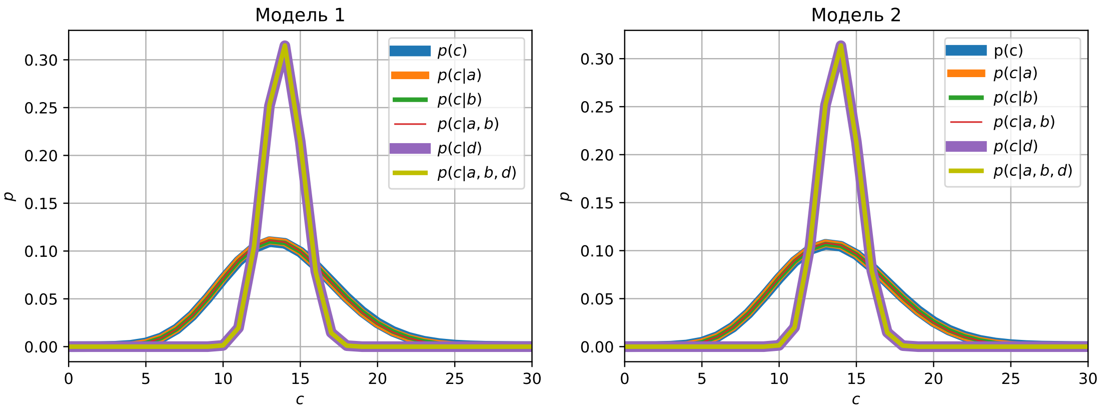

Рис. 1: Графики функций распределения

Можно заметить, что для обеих моделей кривые распределений делятся условно на две группы: те, у которых присутствует информация о d и те, у которых она отсутствует. Во втором случае кривые получаются более вытянутыми вверх и с меньшей дисперсией. Также видно, что добавление информации в виде мат.ожиданий a и b практически не повлияло на априорное распределение c.

<table>
<tr><th></th><th>Модель 1</th><th>Модель 2</th></tr>
<tr><td>$\mathbb{E}c$</td><td colspan="2">13.75</td></tr>
<tr><td>$\mathbb{E}[c|a]$</td><td colspan="2">13.7</td></tr>
<tr><td>$\mathbb{E}[c|b]$</td><td colspan="2">13.75</td></tr>
<tr><td>$\mathbb{E}[c|d]$</td><td>13.895971</td><td>13.893834</td></tr>
<tr><td>$\mathbb{E}[c|a,b]$</td><td colspan="2">13.7</td></tr>
<tr><td>$\mathbb{E}[c|a,b,d]$</td><td>13.890873</td><td>13.888971</td></tr>
<tr><td>$\mathbb{D}c$</td><td>13.1675</td><td>14.0475</td></tr>
<tr><td>$\mathbb{D}[c|a]$</td><td>12.91</td><td>13.785</td></tr>
<tr><td>$\mathbb{D}[c|b]$</td><td>13.0825</td><td>13.9625</td></tr>
<tr><td>$\mathbb{D}[c|d]$</td><td>1.533582</td><td>1.543943</td></tr>
<tr><td>$\mathbb{D}[c|a,b]$</td><td>12.825</td><td>13.7</td></tr>
<tr><td>$\mathbb{D}[c|a,b,d]$</td><td>1.529425</td><td>1.540229</td></tr>
</table>

Таблица 2: Числовые значения мат.ожиданий и дисперсий условных распределени

Мат.ожидания и дисперсии соответствующих распределений были посчитаны по определению (кроме $\mathbb{E}c$ и $\mathbb{D}c$, для которых были выведены формулы).

Как заметно из таблицы, мат.ожидания у обеих моделей примерно одинаковые, однако дисперсия меньше у первой модели.

4 Наибольший вклад в уточнение прогноза

Определим, какая из величин a, b, d вносит наибольший вклад в уточнение прогноза для величины c (в смысле дисперсии распределения).

Проверим, верно ли, что $\mathbb{D}[c|d] < \mathbb{D}[c|a]$ и $\mathbb{D}[c|d] < \mathbb{D}[c|b]$ для любых допустимых значений a, b, d. Для этого вычислим $\max\limits_{d} \mathbb{D}[c|d]$ и $\min\limits_{a} \mathbb{D}[c|a]$, $\min\limits_{b} \mathbb{D}[c|b]$.

Модель 1:

$$\max_d \mathbb{D}[c|d] = 10.29869, \quad \min_a \mathbb{D}[c|a] = 12.28, \quad \min_b \mathbb{D}[c|b] = 12.5875$$

Модель 2:

$$\max_d \mathbb{D}[c|d] = 17.0098, \quad \min_a \mathbb{D}[c|a] = 13.085, \quad \min_b \mathbb{D}[c|b] = 13.4625$$

Видно, что для первой модели соотношения на дисперсии выполняются, а для второй, вообще говоря, нет. Это означает, что для первой модели информация в виде d гарантированно даст больший вклад в уточнение прогноза, чем a или b.

Далее, найдём множество точек (a, b) таких, что $\mathbb{D}[c|b] < \mathbb{D}[c|a]$. Изобразим его графически (**Рис. 2**), а также покажем, что 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 разделимы.

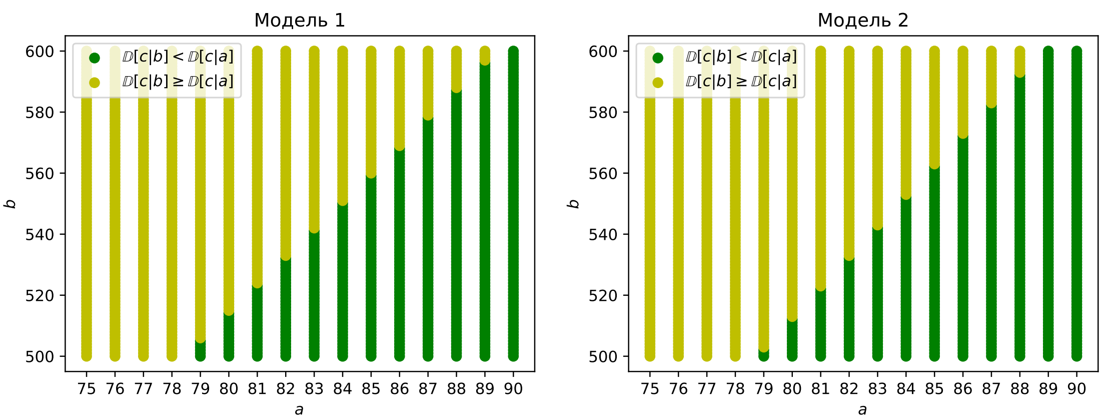

Рис. 2: Сравнение $\mathbb{D}[c|a]$ и $\mathbb{D}[c|b]$

Высоты зеленых столбцов равны:

- в модели 1: $[0, 0, 0, 0, 6, 15, 24, 33, 42, 51, 60, 69, 79, 88, 97, 101]$;
- в модели 2: $[0, 0, 0, 0, 3, 13, 23, 33, 43, 53, 63, 73, 83, 93, 101, 101]$.

То есть возрастание идет с постоянным шагом (от ненулевого значения до максимального, равного 101): в модели 1 – 9, в модели 2 – 10. Следовательно, множества линейно разделимы в обоих случаях, так как можно провести прямую линию, разделяющую их (правда, есть все-таки скачок 69 −> 79 в 1 модели, где шаг равняется 10, но это все равно не помешает провести прямую, так как сетка дискретная и прямая может пройти в зазоре).

Отсюда следует, что нельзя однозначно сказать, какая из величин a или b даёт больший вклад в уточнение прогноза для обеих моделей.

5 Временные замеры

Проведем замеры по времени для функций $p(c)$, $p(c|a)$, $p(c|b)$, $p(c|d)$, $p(c|a, b)$, $p(c|a, b, d)$, $p(d)$. Результаты автоматически усреднятся с помощью встроенной в *jupyter notebook* функции *timeit*, а в качестве условных параметров возьмём мат.ожидания соответствующих априорных распределений, округленных до ближайшего целого.

Результаты приведены в **Таблице 3**.

Видно, что вторая модель работает быстрее, чем первая. Это вполне ожидаемо, поскольку в первой модели распределение $p(c|a, b)$ имеет более сложную формулу для подсчета, чем во второй. И это распространяется на все остальные распределения, так как они высчитываются с помощью $p(c|a, b)$.

6 Сравнение двух моделей

Рассмотрев обе модели, можно сделать следующие выводы:

1. Первая модель работает дольше второй. Объясняется это тем, что для подсчета распределения $p(c|a, b)$ используется свертка.

2. Вторая модель имеет большую дисперсию, чем первая. Объясняется это тем, что Пуассоновское распределение аппроксимирует биномиальное распределение. Следовательно, из-за этого теряется точность.

3. Для первой модели добавление информации в виде d гарантированно улучшает результат (в смысле дисперсии), по сравнению с параметрами a и b. Для второй модели это неверно из-за аппроксимации.

	Модель 1	Модель 2
$p(c)$	0.07139	0.06767
$p(c\|a)$	0.00996	0.00365
$p(c\|b)$	0.00114	0.00075
$p(c\|d)$	0.26448	0.24502
$p(c\|a, b)$	0.00036	0.00014
$p(c\|a, b, d)$	0.04609	0.04365
$p(d)$	0.14360	0.13520

Таблица 3: Временные замеры распределений в секундах