Задание 1. Байесовские рассуждения

Курс: Байесовские методы в машинном обучении 2020

Вероятностные модели посещаемости курса

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 $\mathrm{Bin}(a,p_1)$ и $\mathrm{Bin}(b,p_2)$ 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 $\mathrm{Bin}(c,p_3)$. 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:

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

Рассмотрим несколько упрощённую версию модели 1. Известно, что биномиальное распределение $\mathrm{Bin}(n,p)$ при большом количестве испытаний и маленькой вероятности успеха может быть с высокой точностью приближено пуассоновским распределением $\mathrm{Poiss}(\lambda)$ с $\lambda = np$. Известно также, что сумма двух пуассоновских распределений с параметрами λ_1 и λ_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:

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

Рассмотрим теперь модель посещений нескольких лекций курса. Будем считать, что посещения отдельных лекций являются независимыми. Тогда:

$$
\begin{aligned}
&p(a,b,c_1,\dots,c_N,d_1,\dots,d_N) = p(a)p(b)\prod_{n=1}^{N} p(d_n|c_n)p(c_n|a,b),\\
&d_n|c_n \sim c_n + \mathrm{Bin}(c_n,p_3),\\
&c_n|a,b \sim \mathrm{Bin}(a,p_1) + \mathrm{Bin}(b,p_2),\\
&a \sim \mathrm{Unif}[a_{min},a_{max}],\\
&b \sim \mathrm{Unif}[b_{min},b_{max}].
\end{aligned}
\tag{3}
$$

Вариант 1

Рассматриваются модели 1 и 2 с параметрами $a_{min} = 75$, $a_{max} = 90$, $b_{min} = 500$, $b_{max} = 600$, $p_1 = 0.1$, $p_2 = 0.01$, $p_3 = 0.3$. Провести следующие исследования для обеих моделей:

1. Вывести формулы для всех необходимых далее распределений аналитически.

   По определению равномерного распеределния и из условия задачи:
   $p(a) = \frac{1}{16}[75 \le a \le 90]$
   $p(b) = \frac{1}{101}[500 \le b \le 600]$

   По определению биномиального распределения, из условий задач и так как плотность распределения суммы случайных велчин есть свертка плотностей распределений:
   $p(c|a,b) = \sum_i^c \binom{a}{i} p_1^i (1-p_1)^{a-i} \binom{b}{c-i} p_2^{c-i} (1-p_2)^{b-c+i}$. Но для ускорения расчетов будем считать, что $p(c|a,b) = \sum_{x=0}^c p(\text{Bin}(a,p_1) = x) p(\text{Bin}(b,p_2) = c-x)$
   $p(d|c) = p(c + \text{Bin}(c,p_3) = d) = p(\text{Bin}(c,p_3) = d-c)$

   По описанному выше и формуле полной вероятности:
   $p(c|a) = \sum_b p(c|a,b)p(b) = \frac{1}{101}\sum_{b=500}^{600} p(c|a,b)$
   $p(c|b) = \sum_a p(c|a,b)p(a) = \frac{1}{16}\sum_{a=75}^{90} p(c|a,b)$
   $p(c) = \sum_{a,b} p(c|a,b)p(a)p(b) = \frac{1}{1616}\sum_{a=75}^{90}\sum_{b=500}^{600}\sum_{i=0}^{c} \binom{a}{i}\binom{b}{c-i} p_1^i p_2^{c-i}(1-p_1)^{a-i}(1-p_2)^{b-c+i}$
   $p(d) = \sum_c p(d|c)p(c)$

   По формулам полной вероятности:
   $p(c|d) = \frac{p(d|c)p(c)}{p(d)}$
   $p(c|a,b,d) = \frac{p(a,b,c,d)}{p(a,b,d)} = \frac{p(d|c)p(c|a,b)p(a)p(b)}{\sum_c p(d|c)p(c|a,b)p(a)p(b)}$

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

   Воспользуемся определением математического ожидания и дисперсии.

$$\mathbb{E}x = \sum_x xp(x); \qquad \mathbb{D}x = \sum_x x^2 p(x) - (\mathbb{E}x)^2$$

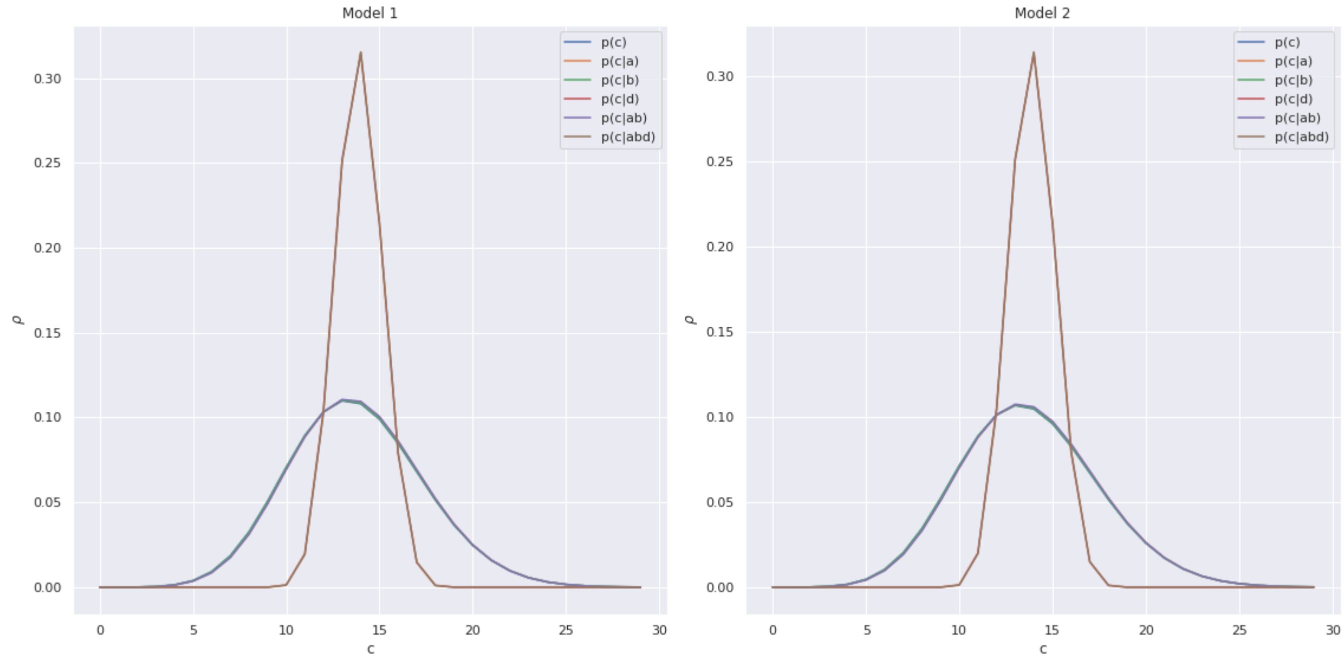
$$\mathbb{E}a = \frac{a_{\min} + a_{\max}}{2}; \qquad \mathbb{E}b = \frac{b_{\min} + b_{\max}}{2}$$

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

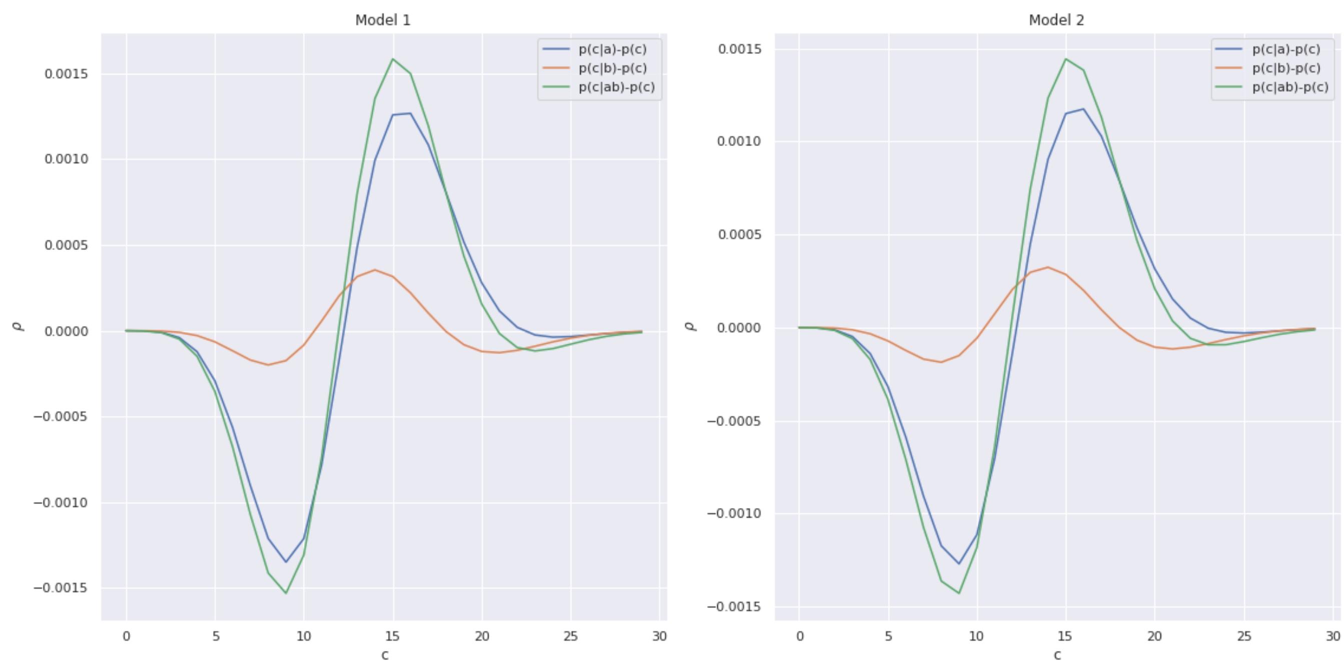
Модель		$p(a)$	$p(b)$	$p(c)$	$p(d)$
1	$\mathbb{E}$	82.50	550.00	13.75	17.88
	$\mathbb{D}$	21.25	850.00	13.17	25.14
2	$\mathbb{E}$	82.50	550.00	13.75	17.87
	$\mathbb{D}$	21.25	850.00	14.05	26.63

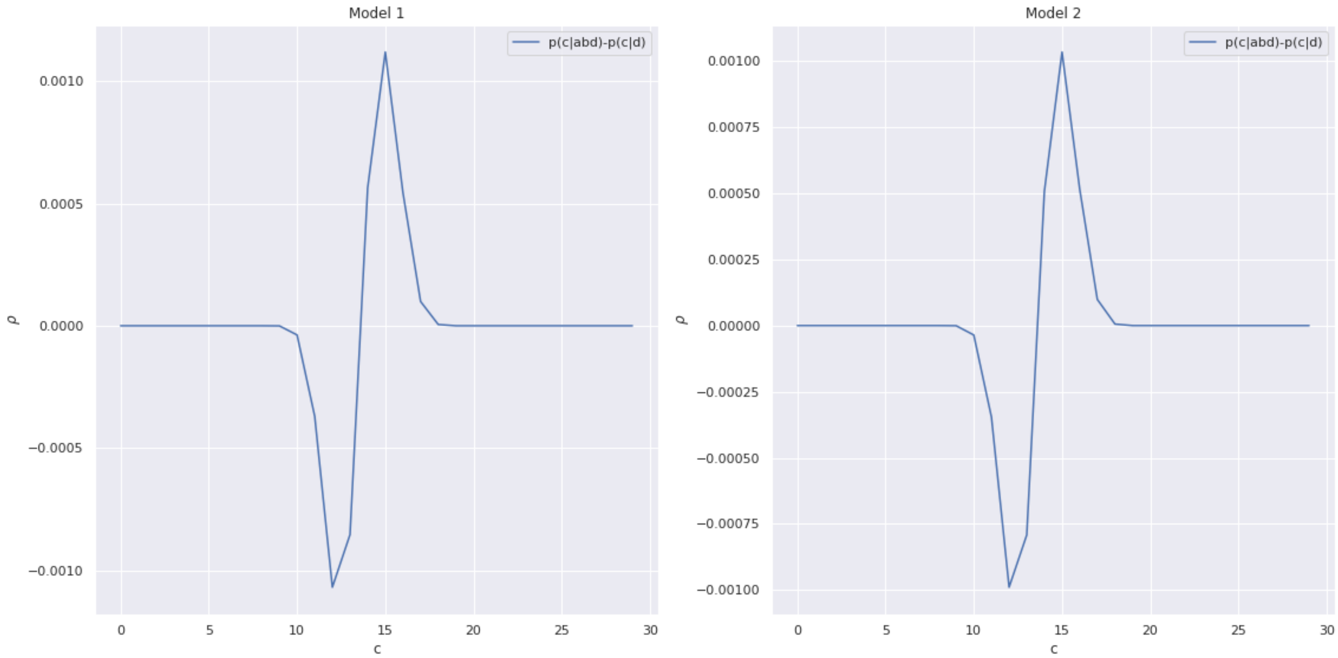
3. Пронаблюдать, как происходит уточнение прогноза для величины c по мере прихода новой косвенной информации. Для этого построить графики и найти мат.ожидание и дисперсию для распределений $p(c)$, $p(c|a)$, $p(c|b)$, $p(c|d)$, $p(c|a,b)$, $p(c|a,b,d)$ при параметрах a, b, d, равных мат.ожиданиям своих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

   Построим графики плотностей для первой и второй моделей



Заметим, что по графикам плотностей ничего толком не понятно – так малы некоторые уточнения. Понятно, что из-за d график плотности сильно меняется, однако ничего тольком разглядеть невозможно. Поэтому построим графики разностей “похожих“ плотностей распределений.



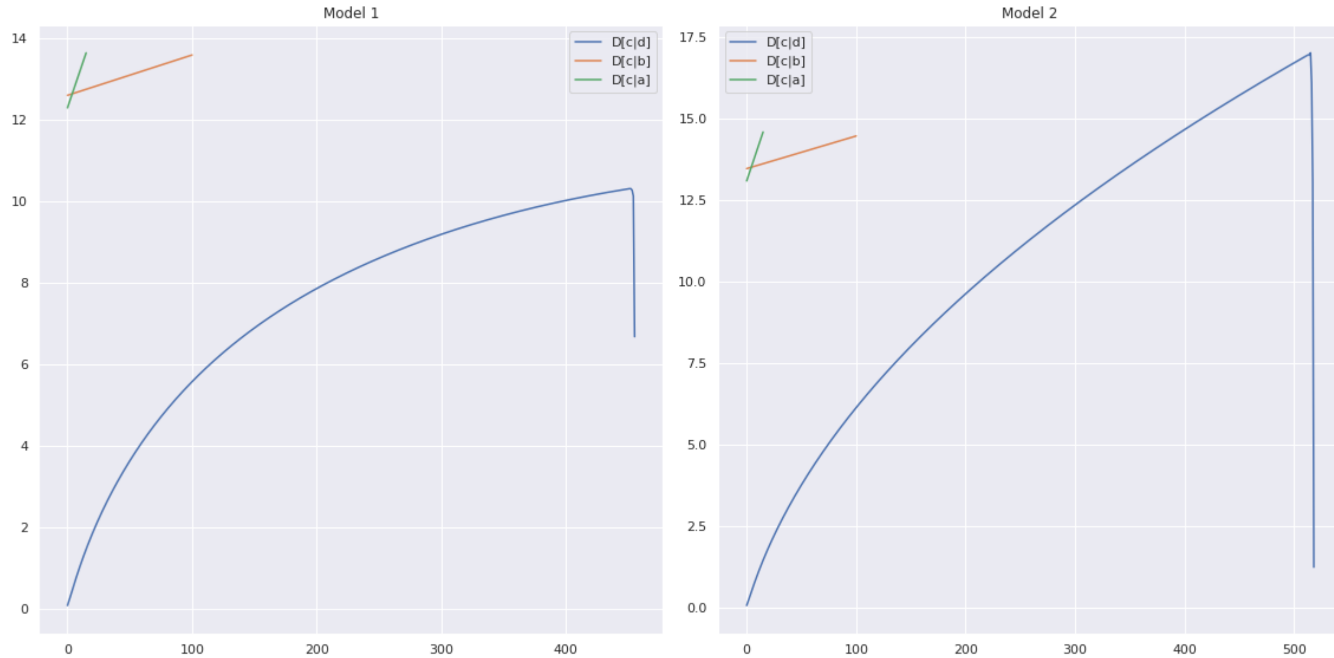


Посмотрев на них мы можем заключить, что сдвиг действительно происходит. И например заметить, что информаци и об a, и об b, меняет прогноз наибольшим образом (что интуитивно ясно из условия задачи). В целом это совпадает с интуитивным пониманием баесовского вывода. И таблицой из пункта 2.

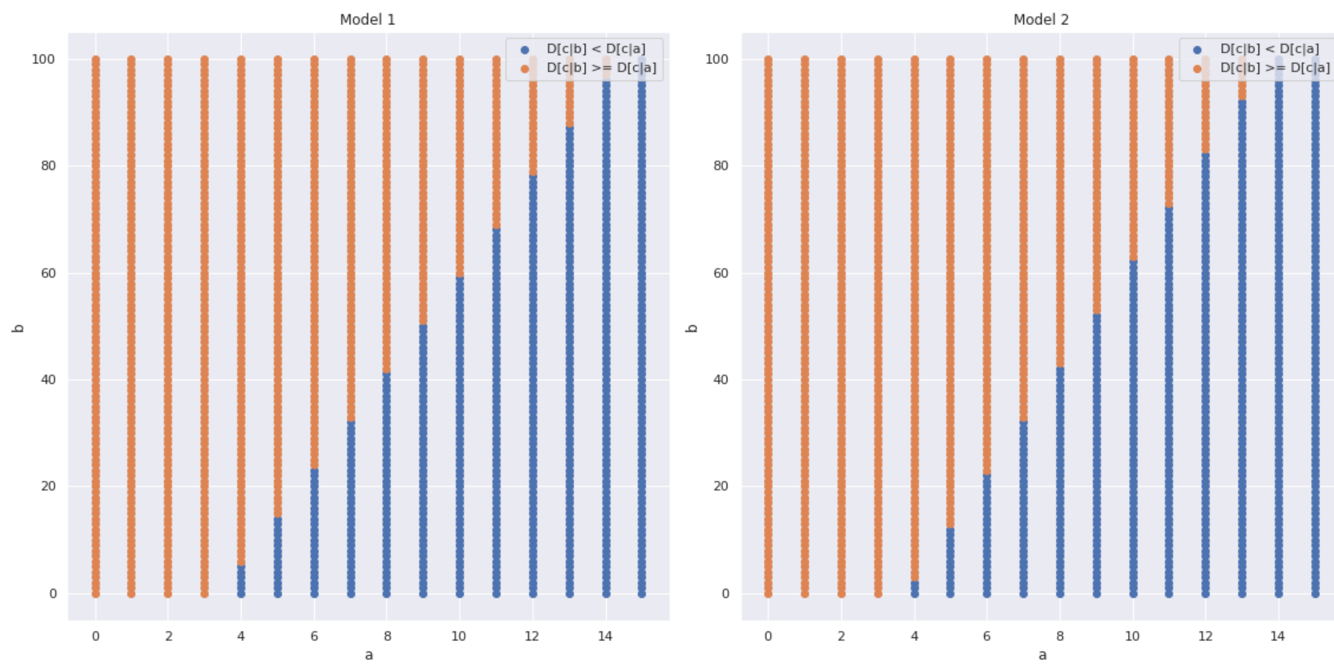
Модель		$p(c)$	$p(c\|a)$	$p(c\|b)$	$p(c\|d)$	$p(c\|a,b)$	$p(c\|a,b,d)$
1	$\mathbb{E}$	13.75	13.80	13.75	13.90	13.80	13.90
	$\mathbb{D}$	13.17	13.00	13.08	1.53	12.92	1.53
2	$\mathbb{E}$	13.75	13.80	13.75	13.89	13.80	13.90
	$\mathbb{D}$	14.05	13.88	13.96	1.54	13.80	1.54

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!

   Построим графики дисперсий. Заметим, что для первой модели указанные соотношения выполняются всегда, в отличии от второй модели.



Для проверки линейной разделимости построим требуемые множества. Кажется, что они линейноразделимы.



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

   Привожу таблицу временных замеров. Значения указаны в милисекундах.

Модель	$p(c)$	$p(c\|a)$	$p(c\|b)$	$p(c\|d)$	$p(c\|a,b)$	$p(c\|a,b,d)$	$p(d)$
1	77.10 ± 6.38	72 ± 5.15	73 ± 3.95	146 ± 4.72	72.4 ± 6.85	7600 ± 41.70	142 ± 3.74
2	83.60 ± 5.79	81.80 ± 3.04	84.9 ± 2.17	157 ± 5.58	82.40 ± 4.70	7700 ± 48.50	145 ± 3.50

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.

1) Можно назвать первую модель более точной, так как дисперсии получаемых величин меньше. 2) Для первой модели $\mathbb{D}[c|d] < \mathbb{D}[c|b]$ и $\mathbb{D}[c|d] < \mathbb{D}[c|a]$ выполняетс всегда, в отличии от второй. Хотелось бы мне связать эти два факта и сказать, что первая модель лучше поэтому. Но я к сожалению не до конца понимаю в каком смысле можно сравнить эти модели.