

Байесовский подход в статистике

Лекция №11

Статистические методы в ядерном эксперименте, ИЯФ, 2024 г.

- 1 Условная вероятность, формула Байеса
- 2 Байесовский подход
- 3 Дискретный случай
- 4 Непрерывный случай
- 5 Примеры

Условная вероятность

Пусть событию E_1 соответствуют k_1 элементарных исходов, событию E_2 — k_2 исходов, событию $E_1 E_2$ — k_{12} исходов. Пусть полное число элементарных исходов равно n . Тогда вероятность $P(E_1 E_2)$ события $E_1 E_2$ равна:

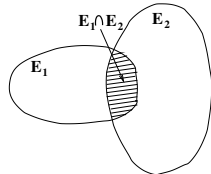
$$P(E_1 E_2) = \frac{k_{12}}{n}, \quad P(E_1) = \frac{k_1}{n}, \quad P(E_2) = \frac{k_2}{n}$$

$$P(E_1 E_2) = \frac{k_{12}}{n} = \frac{k_1}{n} \cdot \frac{k_{12}}{k_1} = P(E_1) \cdot \frac{k_{12}}{k_1} = P(E_1) \cdot P(E_2|E_1)$$

$$P(E_1 E_2) = P(E_2) \cdot P(E_1|E_2)$$

$P(E_2|E_1) = \frac{k_{12}}{k_1}$ - вероятность события E_2 при условии, что событие E_1 уже произошло;

$P(E_2|E_1)$ называется условной вероятностью. В общем случае $P(E_2|E_1) \neq P(E_1|E_2)$.



События E_1 и E_2 называют независимыми если $P(E_1 E_2) = P(E_1)P(E_2)$, или: $P(E_2|E_1) = P(E_2)$. Т.е. если событие E_1 произошло, и это не влияет на вероятность события E_2 , то события E_1 и E_2 независимы.

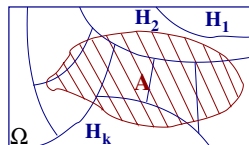
Несовместность и независимость событий - разные свойства !

Формула полной вероятности (Байеса)

Набор попарно несовместных событий H_1, H_2, \dots ($H_i H_j = \emptyset, \forall i \neq j$) таких, что $P(H_k) > 0, \forall k$ и $\cup H_k = \Omega$, называется полной группой событий или разбиением пространства Ω .

Тогда вероятность любого события A может быть записана как:

$P(A) = \sum_k P(H_k)P(A|H_k)$ - формула полной вероятности



Действительно:

$$P(A) = P(A \cdot \Omega) = P(A \cdot (\cup H_k)) = P(\cup (A H_k)) = \sum_k P(A H_k) = \sum_k P(H_k)P(A|H_k)$$

Из $P(A H_k) = P(H_k)P(A|H_k) = P(A)P(H_k|A)$, получаем:

Формула Байеса

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{P(A)} = \frac{P(H_k)P(A|H_k)}{\sum_i P(H_i)P(A|H_i)}$$

Байесовский подход

До того, как будут получены данные, физик рассматривает степени своего доверия к возможным моделям (гипотезам) и представляет их в виде вероятностей (априорных вероятностей). Как только данные получены, теорема Байеса позволяет физику рассчитать новое множество вероятностей (апостериорные вероятности), которые представляют собой пересмотренные степени доверия к возможным моделям, учитывающие новую информацию из экспериментальных данных.



Байесовский подход, дискретный случай (I)

- Рассматриваем список моделей $\{H_1, H_2, \dots, H_k\}$, составляющий исчерпывающее множество возможных взаимоисключающих вероятностных моделей, пригодных для описания изучаемой ситуации. Присваиваем этим моделям значения априорных вероятностей:

$$\{P(H_1), P(H_2), \dots, P(H_k)\}, 0 \leq P(H_i) \leq 1, i = 1, \dots, k, \sum_{i=1}^k P(H_i) = 1. \text{ Каждая}$$

вероятностная модель H_i определяет распределение вероятностей множества возможных данных, которые можно было бы получить.

- Если D – данные, полученные в эксперименте, то вероятности данных, определяемых каждой из альтернативных моделей H_i будут задаваться с помощью условных вероятностей $\{P(D|H_1), P(D|H_2), \dots, P(D|H_k)\}$. $P(D|H_i)$ называют правдоподобием H_i при заданных данных D .
- Затем физик пересматривает априорные вероятности $P(H_i)$ в свете новой информации, содержащейся в данных D . Это значит, что рассчитываются апостериорные вероятности $\{P(H_1|D), P(H_2|D), \dots, P(H_k|D)\}$, зависящие теперь от данных D . Теорема Байеса позволяет вычислить апостериорные вероятности $P(H_i|D)$, если известны априорные вероятности $P(H_i)$ и правдоподобия $P(D|H_i)$:

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{P(D)}, i = 1, 2, \dots, k$$

$$P(D) = P(D|H_1)P(H_1) + P(D|H_2)P(H_2) + \dots + P(D|H_k)P(H_k) = \sum_{i=1}^k P(D|H_i)P(H_i).$$

Обратим внимание, что $\sum_{i=1}^k P(H_i|D) = 1$ также как и $\sum_{i=1}^k P(H_i) = 1$.

Байесовский подход, дискретный случай (II)

Замечание: Если знания извлекаются из нескольких наборов данных ($D_1 \cup D_2$), то:

$$P(H_i|D_1 \cup D_2) = \frac{P(D_1 \cup D_2|H_i)P(H_i)}{P(D_1 \cup D_2)}, \quad i = 1, 2, \dots, k,$$

причём конечные апостериорные вероятности можно получать поэтапно:

$$P(H_i) \rightarrow P(H_i|D_1) \rightarrow P(H_i|D_1 \cup D_2),$$

или

$$P(H_i) \rightarrow P(H_i|D_2) \rightarrow P(H_i|D_1 \cup D_2).$$

Пример 1: В некотором царстве животных генотипам ВВ и Вв соответствуют животные с чёрной окраской, генотипам bb – с коричневой, а при всех спариваниях неизменно получают помёт из семи детёнышей. Чёрное животное, о котором известно, что оно получилось в результате спаривания Вв×Вв, само спаривается с коричневым животным, и оказывается, что все семь детёнышей – чёрные. Каковы вероятности того, что чёрный родитель имеет генотип ВВ или Вв соответственно ?

Пример 2: А если нам известно только то, что одно из спариваний Вв×Вв или ВВ×Вв привело к появлению чёрного родителя ?

Если Θ – множество возможных значений параметра θ с априорной плотностью вероятности $\pi(\theta)$ и $F(x^*|\theta)$ – правдоподобие θ при заданных наблюдениях x^* , то апостериорная плотность вероятности $p(\theta|x^*)$ для параметра θ рассчитывается по формуле Байеса:

$$p(\theta|x^*) = \frac{F(x^*|\theta)\pi(\theta)}{p(x^*)}, \quad p(x^*) = \int_{\Theta} F(x^*|\theta)\pi(\theta)d\theta.$$

Напомним, что $F(x^*|\theta)$ для единичного измерения случайной величины x есть просто ф.п.в. X , взятая в точке $x = x^*$, т.е. $F(x^*|\theta) = p(x^*|\theta)$. Причём при использовании Байесовского подхода ф.п.в., зависящая от параметра θ рассматривается как условная ф.п.в. x : $p(x|\theta)$.

Если найдена ф.п.в. для параметра θ , $p(\theta|x^*)$, то можно используя теорию вероятности вычислить любую характеристику.

Т.е. байесовский подход сводит задачу статистики к задаче теории вероятностей.

Особенности байесовского подхода (I)

- В современной теории вероятностей существуют два подхода к тому, что называть случайностью.
- В частотном подходе предполагается, что случайность есть объективная неопределенность.
- В байесовском подходе предполагается, что случайность есть мера нашего незнания. Например, случайность при бросании кости связана с незнанием динамических характеристик кубика, сукна, руки кидающего, сопротивления воздуха итп.
- При интерпретации случайности как объективной неопределенности единственным возможным средством анализа является проведение серии испытаний. При этом вероятность события интерпретируется как предел частоты наступления этого события в n испытаниях при $n \rightarrow \infty$. Т.о. все величины четко делятся на случайные и детерминированные.
- Идея байесовского подхода заключается в переходе от априорных знаний (или точнее незнаний) к апостериорным с учетом наблюдаемых явлений. Все величины и параметры считаются случайными. Байесовские методы работают даже при объеме выборки $n = 0$. В этом случае апостериорное распределение равно априорному. В качестве оценок неизвестных параметров выступают апостериорные распределения, т.е. решить задачу оценивания некоторой величины, значит найти её апостериорное распределение.
- Начиная с 1930 г. байесовские методы подвергались резкой критике и практически не использовались. В настоящее время (с начала 1990 г.) наблюдается возрождение байесовских методов, которые оказались в состоянии решить многие проблемы статистики и машинного обучения.
- Если при получении оценки параметра (его доверительного интервала) при анализе экспериментальных данных в ФЭЧ использовались байесовские методы, то в конце анализа часто проводятся серии псевдоэкспериментов (Toy Monte Carlo) для уточнения частотных свойств найденных доверительных интервалов.

- Результаты расчётов в байесовском подходе зависят от неизвестной функции априорной вероятности $\pi(\theta)$, в неё часто закладываются субъективные ожидания.
- Постулат Байеса (1763 г.): Если распределение априорной вероятности для θ неизвестно, то, предполагая, что все значения θ априори равновероятны ($\pi(\theta) = \text{const}$), мы тем самым сможем охарактеризовать нашу полную неосведомлённость относительно θ . Автор сам был настолько неуверен в справедливости данного постулата, что так его и не опубликовал при жизни.
- Параметр θ не является случайным параметром в классическом частотном понятии теории вероятностей. Параметр θ имеет вполне определённое значение, просто мы его не знаем.

Байесовский подход, непрерывный случай (II)

Если для параметра θ получена апостериорная плотность вероятности $p(\theta|x^*)$, то часто используемые характеристики:

- Мода θ^* (или наиболее вероятное значение θ). Если $\pi(\theta) = \text{const}$, то θ^* часто совпадает с оценкой методом максимального правдоподобия.
- Среднее $\bar{\theta}$ и дисперсия σ_{θ}^2 :

$$\bar{\theta} = \int_{-\infty}^{+\infty} \theta p(\theta|x^*) d\theta, \quad \sigma_{\theta}^2 = \int_{-\infty}^{+\infty} (\theta - \bar{\theta})^2 p(\theta|x^*) d\theta.$$

- Медиана θ_m определяется как:

$$\int_{-\infty}^{\theta_m} p(\theta|x^*) d\theta = \int_{\theta_m}^{+\infty} p(\theta|x^*) d\theta.$$

- Доверительный интервал для θ уровня значимости $1 - \epsilon$:

$$\int_a^b p(\theta|x^*) d\theta = 1 - \epsilon,$$

причём (a, b) есть интервал наивысшей апостериорной плотности вероятности уровня $100\%(1 - \epsilon)$. Т.е. $\forall \theta \in (a, b)$ и $\theta' \notin (a, b)$, $p(\theta|x^*) \geq p(\theta'|x^*)$.

Пример 3: Из N событий n событий прошло критерии отбора. Какова эффективность p (вероятность) этих критериев отбора ?

Пример 4: Пусть из N событий все прошли критерии отбора. Какова эффективность p (вероятность) этих критериев отбора ?

Пример 5: Пусть в одном экспериментальном заходе из N_1 событий n_1 событий прошло критерии отбора, а в другом заходе – из N_2 событий было отобрано n_2 событий. Какова средняя эффективность p этих критериев отбора ?

Пример 6: Иногда в модели существуют скрытые параметры \mathbf{s} – параметр, который невозможно определить из эксперимента. Например, роль таких параметров играют источники систематических ошибок. Байесовский вывод позволяет легко учесть такие параметры в анализе: их следует учитывать так же, как и остальные параметры, и в конце по ним надо проинтегрировать:

$$p(\theta|x^*) = \frac{\int p(x^*|\theta, \mathbf{s})\pi(\theta, \mathbf{s})d\mathbf{s}}{\int p(x^*|\theta, \mathbf{s})\pi(\theta, \mathbf{s})d\theta d\mathbf{s}}$$

Пусть в эксперименте измерили некую величину x с точностью σ_x . При этом измерительная шкала сдвинута на неизвестную величину с нулевым ожиданием и дисперсией σ_y . Требуется оценить истинное значение μ и ошибку измерения σ истинного значения величины.

Пример 7: В эксперименте измерили массу нейтрино и получили значение $x = -5.41$ эВ. В результате моделирования было получено, что разрешение установки $\sigma_x = 3.3$ эВ. Что можно сказать о массе нейтрино m ?

Частотный подход (Неймана) (III)

| s | 1σ | 2σ | 3σ | 4σ | 5σ | 6σ |
|---------------|-----------|-----------------------|----------------------|----------------------|----------------------|----------------------|
| ε | 0.1587 | $2.275 \cdot 10^{-2}$ | $1.35 \cdot 10^{-3}$ | $3.15 \cdot 10^{-5}$ | $2.87 \cdot 10^{-7}$ | $1.00 \cdot 10^{-9}$ |

- случай верхнего предела: $\mu \leq \mu_{\text{up}} = x_{\text{obs}} + \sigma s(\varepsilon)$,
где уровню достоверности $CL=95\%$ ($\varepsilon = 0.05$) соответствует $s = 1.64$.
- случай симметричного доверительного интервала:
 $x_{\text{obs}} - \sigma s(\varepsilon) \leq \mu \leq x_{\text{obs}} + \sigma s(\varepsilon)$

Верхний предел в Байесовском подходе

