Глава 10

Основы теории–Методы Байеса

§1 Введение

Байесовская теория предлагает совершенно другой подход к статистике. Байесовская версия оценки, рассмотренная здесь, касается основных ситуаций оценки параметра распределения случайной выборки. Классическая оценка заключается в методе нахождения максимума функции правдоподобия.

Фундаментальное различие в Баейсовского подхода от классического заключается в том, что параметр θ рассматривается как случайная величина.

В классической статистике θ фиксирована, но неизвестная величина. Данное предположение приводит к трудностям в интерпретации, требующейся для классических доверительных интервалов, которые являются случайными. После взятия данных по наблюдениям и расчета числового интервала, отпадает предположение о вероятности. Т.е. выражение $P(10.45 < \theta < 13.26) = 0.95$ не имеет смысла, т.к. θ не является случайной величиной.

В Байесовской статистике таких трудностей не возникает, и вероятностные утверждения могут быть применены относительно значений параметра θ.

§2 Теорема Байеса

Пусть $B_1, ..., B_k$ некоторые попарно несовместимые события пространства элементарных событий S, такие что $P(B_i) \neq 0$ для $i = 1, ..., k$, тогда для некоторого события $A \in S$ и $P(A) \neq 0$ верно:

$$P(B_r|A) = \frac{P(A|B_r)P(B_r)}{P(A)}, \text{где } P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

для $r = 1, ..., k$

Пример 10.10
Три изготовителя поставляют одежду розничному продавцу. 60 % одежды прибывают от изготовителя 1, 30 % от изготовителя 2 и 10 % от изготовителя 3. 10 % одежды от изготовителя 1, 5% от изготовителя 2 и 15% от изготовителя 3 являются дефектными. Определить вероятность, поступления дефектной одежды от изготовителя 3.

Решение Пусть A - событие, что прибыла дефектная одежда.

Пусть B_i - событие, прибыла одежда от производителя i. Согласно Байесовской теореме получаем:

$$P(B_3|A) = \frac{(0.15)(0.1)}{(0.1)(0.6) + (0.05)(0.3) + (0.15)(0.1)} = \frac{0.015}{0.09} = 0.167$$

Хотя производитель 3 поставляет только 10% одежды, его доля в поставке дефектной одежды составляет около 17%.

§3 Априорные и апостериорные распределения

Предположим, что $\underline{X} = (X_1, ..., X_n)$ случайная выборка с плотностью или функцией вероятности $f(x;\theta)$ и требуется оценить параметр θ.

Т.к параметр θ - случайная величина, она будет иметь распределение. Это позволяет использовать любые доступные предположения о возможных значениях для θ до получения каких-либо значений. Эти предположения выражаются через априорное распределение .

Затем после получения подходящих значений, определяется апостериорное распределение θ, и формируются основные выводи касающиеся θ.

3.1 Замечания

Т.к θ - случайная величина, она должна обозначаться Θ и ее априорная плотность должна записываться как $f_\Theta(\theta)$. Однако, для простоты нет мы не различаем Θ и θ, плотность параметра обозначается как $f(\theta)$. Заметим, что предполагается, что θ непрерывная случайная величина. Это предположение остается верным, даже в случае если X дискретная случайная величина, например имеющая Биномиальное или Пуассоновское распределение с параметрами (p или λ), принимающими непрерывное значение ((0,1) и (0, ∞) соответственно).

3.2 Определение апостериорной плотности

Предположим $\underline{X}$ случайная выборка с $f(x|\theta)$ и θ имеет априорную плотность $f(\theta)$.

Апостериорная плотность $\theta|\underline{X}$ определяется, используя определение условной плотности:

$$f(\theta|\underline{X}) = \frac{f(\theta, \underline{X})}{f(\underline{X})} = \frac{f(\underline{X}|\theta)f(\theta)}{f(\underline{X})}$$

Заметим, что $f(\underline{X}) = \int f(\underline{X}|\theta)f(\theta)d\theta$. Этот результат аналогичен непрерывной версии Байесовской теоремы.

Часто удобно записать полученный результат в терминах статистики, обозначая случайную выборку $\overline{X}$:

$$f(\theta|\overline{X}) = \frac{f(\overline{X}|\theta)f(\theta)}{f(\overline{X})}$$

На практике эти обозначения эквивалентны.

Удобным способом обозначения апостериорной плотности - использование пропорциональности. $f(\underline{X})$ не определяет θ единственным образом. Для достижения единственности требуется определение константы. Следовательно:

$$f(\theta|\underline{X}) \propto f(\underline{X}|\theta)f(\theta)$$

Также заметим, что $f(\underline{X})$ - совместная плотность распределения элементов выборки не является чем-то отличным от функции правдоподобия. Таким образом апостериорное распределение пропорционально функции правдоподобия.

Для данной функции правдоподобия, если априорное распределение приводит к апостериорному распределению, принадлежащему тому же семейству, что и априорное распределение, тогда априорное распределение называют сопряженным априорным для этой функции правдоподобия.

§4 Функция ущерба

Чтобы получить оценку параметра θ, необходимо для начала определить функцию ущерба. Это мера понесенных ущерба, при использовании $g(\underline{X})$ в качестве оценки параметра θ. Функция ущерба выбрана так, что она принимает значение ноль, когда оценка точна, что значит $g(\underline{X}) = \theta$, и принимает положительное значение и не возрастает, когда $g(\underline{X})$ далека от θ. Существует одна наиболее часто используемая функция ущерба, называемая квадратичная ошибка ущерба. Две остальные также применяются на практике.

Таким образом, Байесовская оценка $g(\underline{X})$ минимизирует ожидаемые потери при выводе апостериорного распределения.

Главная функция ущерба - квадратичные потери определена формулой:

$$L(g(\underline{x}), \theta) = [g(x) - \theta]^2$$

аналогом ее в классической статистике является среднеквадратичная ошибка.

Вторая функция ущерба - абсолютная ошибка ущерба определяется формулой:

$$L(g(\underline{x}), \theta) = |g(x) - \theta|$$

Третья функция ущерба - "0/1" или "все-ничего" потеря определяется формулой:

$$L(g(\underline{x}), \theta) = \begin{cases} 0, \text{ при } g(x) = \theta \\ 1, \text{ при } g(x) \neq \theta \end{cases}$$

Байесовской оценкой, которая достигается минимизацией ожидаемых потерь, для каждой из этих трех функций ущерба является математическое ожидание, медиана и мода апостериорного распределения соответственно. Ожидаемые апостериорные потери равны:

$$EPL = E[L(g(\underline{x}), \theta)] = \int L(g(\underline{x}), \theta) f(\theta|\underline{x}) d\theta$$

4.1 Квадратичная ошибка ущерба

Для простоты будем писать g вместо $g(\underline{x})$.

$$EPL = \int (g-\theta)^2 f(\theta|\underline{x})d\theta$$

$$\frac{d}{dg}EPL = 2\int (g-\theta) f(\theta|\underline{x})d\theta$$

Приравнивание к нулю $\Rightarrow g\int f(\theta|\underline{x})d\theta = \int \theta f(\theta|\underline{x})d\theta$

но $\int f(\theta|\underline{x})d\theta = 1$ более того $g = \int \theta f(\theta|\underline{x})d\theta = E[\theta|\underline{x}]$.

Ясно, что это минимизирует EPL/

Получили, что Байесовская оценка с использованием квадратичной потери является математическое ожиданием апостериорного распределения.

4.2 Абсолютная ошибка ущерба

Снова для простоты будем писать g вместо $g(\underline{x})$.

$$EPL = \int |g-\theta| f(\theta|\underline{x})d\theta$$

Предполагая, что θ принимает значения из интервала $(-\infty, \infty)$, получаем:

$$EPL = \int\limits_{-\infty}^{g} (g-\theta) f(\theta|\underline{x})d\theta + \int\limits_{g}^{\infty} (\theta-g) f(\theta|\underline{x})d\theta$$

Более того:

$$\frac{d}{dg}EPL = \int\limits_{-\infty}^{g} f(\theta|\underline{x})d\theta - \int\limits_{g}^{\infty} f(\theta|\underline{x})d\theta$$

[Используется формула $\frac{d}{dy}\int\limits_{a(y)}^{b(y)} f(x,y)dx = \int\limits_{a(y)}^{b(y)} \frac{\partial}{\partial y} f(x,y)dx + b'(y)f(b(y),y) - a'f(a(y),y)$]

приравнивая к нулю получаем:

$$\int\limits_{-\infty}^{g} f(\theta|\underline{x})d\theta = \int\limits_{g}^{\infty} f(\theta|\underline{x})d\theta$$

Следовательно $P(\theta \leqslant g) = P(\theta \geqslant g)$

откуда следует медиана апостериорного распределения.

4.3 Бинарная ошибка ущерба

Вместо подхода дифференцирования будет использован прямой подход с ограничениями.

Рассмотрим

$$L(g(\underline{x}), \theta) = \begin{cases} 0, \text{ при } g - \epsilon < \theta < g + \epsilon \\ 1, \text{ иначе} \end{cases}$$

Тогда при $\epsilon \to 0$ получаем требуемую функцию ущерба.

$$EPL = 1 - \int\limits_{g-\epsilon}^{g+\epsilon} f(\theta|\underline{x}) d\theta = 1 - 2\epsilon f(\theta|\underline{x})$$

для маленьких значений ϵ.

Приравняв g к моде $f(\theta|\underline{x})$ получим минимум этого выражения.

Пример 10.11 Для оценки биномиальной вероятности θ единичного наблюдения X с априорным распределением θ - $Beta(\alpha, \beta)$ исследуйте форму апостериорного распределения θ и определите Байесовскую оценку θ, используя квадратичную ошибку ущерба.

Решение Используя отношение пропорциональности, можем опустить любые константы. Таким образом, получаем:

Априорное распределение: $f(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$,опуская константу $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$

Функция правдоподобия: $f(\theta|X) \propto \theta^{x}(1-\theta)^{n-x}$, пропуская константу $\binom{n}{k}$.

Следовательно $f(\theta|X) \propto \theta^{x}(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1} = \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$

можно заметить, что не принимая во внимание пропущенные константы, получаем плотность Beta - распределения случайной величины. Более того получаем апостериорное распределение θ с параметрами $(x + \alpha)$ и $(n - x + \beta)$.

Также можно заметить, что апостериорная плотность и априорная плотность принадлежат одному семейству распределений. Следовательно, сопряженное априорное распределение для биномиального распределения - Beta-распределение. Байесовская оценка с использованием квадратичной ошибки ущерба - математическое ожидание этого распределения: $\frac{x+\alpha}{(x+\alpha)(n-x+\beta)} = \frac{x+\alpha}{n+\alpha+\beta}$

§5 Вопросы студентам

В1 Используется ли метод Байеса?

О1 Метод достаточно часто используется, например, в общем страховании в теории доверия, которую мы рассмотрим в главе 6.

В2 Не является ли метод Байеса произвольным, т.к. вы можете выбрать любую априорную и апостериорную функцию?

О2 Во многих ситуациях априорное распределение выбирается согласно фактически наблюдаемому распределению, как показано в наших примерах. Функция правдоподобия должна основываться на серьезности недооценки истинного значения, например, снижение прибыли компании в случае планирования на основе ложных значений. Таким образом, предположения не настолько произвольны, как может показаться на первый взгляд.

Однако, существует некоторая степень произвольности в любых методах статистической оценки. Например, метод моментов произвольно присваивает ожидаемую степень, вместо некоторого другого набора функций. В то же время, метод максимального правдоподобия подразумевает использование неинформативной априорной функции и функции бинарной ошибки ущерба, получаемые через предположения. Байесовский подход пытается охватить всю доступную информацию, делая очевидные предположения.

В3 Функция бинарная ошибки ущерба, кажется, не является непрерывным распределением. Т.к. вы не можете точно

оценить неизвестный параметр, вы уверены, что не ошибаетесь каждый раз?

О3 Вы правы. Для непрерывных распределений функция бинарная ошибки ущерба является математической идеализацией. Она похожа на обобщенную функцию Дирака дельта $\delta(x-a)$, которая используется в физике. Эта функция принимает значение 0 везде, кроме точки $x = a$, а в этой точке она неопределена. Однако, она имеет полезное свойство: $\int\limits_{-\infty}^{\infty} \delta(x-a)f(x)dx = f(a)$, что может быть использовано для упрощения вычислений.

5.1 Подсказки для ответов на вопросы

1. Основные вопросы по данной главе заключаются в нахождении апостериорного распределения и дальнейшего его использования для решения задачи.

2. При нахождении апостериорного распределения следует учитывать следующее:

   - в априорной и апостериорной функции распределения неизвестный параметр рассматривается как случайная величина, следовательно этот параметр занимает место x в формуле функции распределения.

   - существует множество одинаковых по значению переменных, например λ и λ' или μ и m.

   Поэтому важно понимать значение переменных и помнить:

   - в априорном и апостериорном распределении вы принимаете параметр распределения (например θ) за случайную переменную

   - и функция правдоподобия - функция от неизвестного параметра, а наблюдаемая случайная величина принимается за константу.

3. Функция квадратичной ошибки ущерба - наиболее часто используемая функция, поэтому убедитесь, что вы правильно поняли ее применение.

4. Заметим, что вам нет необходимости запоминать таблицу сопряженных функций, но на экзамене вас могут спросить их вывод.

§6 Ответы на вопросы для самоподготовки

Решение 5.1

Число людей, делающих вычисления без ошибок равно:

	Могут	Не могут	Всего
Актуарии	60	40	100
Бухгалтера	80	120	200
Всего	140	160	300

Итак вероятность, что представитель, кто думает, что $2+3=4$ является актуарием равна $40/160 = 1/4$

Решение 5.2

Определим следующие события (где h мала)

A_S : индивидуальный иск типа S
A_M : индивидуальный иск типа M
A_L : индивидуальный иск типа L
B : размер индивидуального иска лежит между \$5000 и \$$(5000+h)$
Для данного θ вероятность события B равна:

$$P(B|\Theta=\theta) = \int\limits_{5000}^{5000+h} \frac{2\theta^2}{x^3}dx$$

Мы можем аппроксимировать интеграл, используя типичное значение функции, умноженное на длину интервала

$$P(B|\Theta=\theta) \doteq \frac{2\theta^2}{5000^3}h$$

Отсюда получаем следующие вероятности:

$$P(B|\Theta=\theta_S) = \frac{2*100^2}{5000^3}h = 0.00000016h$$

$$P(B|\Theta=\theta_M) = \frac{2*1000^2}{5000^3}h = 0.000016h$$

$$P(B|\Theta=\theta_L)=\frac{2*2500^2}{5000^3}h=0.0001h$$

Известно, что
$P(\Theta=\theta_S)=0.80$
$P(\Theta=\theta_M)=0.15$
$P(\Theta=\theta_L)=0.05$

Используя формулу Байеса, получаем:

$$\begin{aligned}&P(\Theta=\theta_S|B)\\&\quad=\frac{P(B|\Theta=\theta_S)P(\Theta=\theta_S)}{P(B|\Theta=\theta_S)P(\Theta=\theta_S)+P(B|\Theta=\theta_M)P(\Theta=\theta_M)+P(B|\Theta=\theta_L)P(\Theta=\theta_L)}\\&\quad=\frac{0.00000016h*0.80}{0.00000016h*0.80+0.000016h*0.15+0.0001h*0.05}\\&\quad=0.0170\end{aligned}$$

Аналогично:
$P(\Theta=\theta_M|B)=0.3188$
$P(\Theta=\theta_L|B)=0.6642$

Итак, апостериорная вероятность типа S,L и M равна 1.70%, 31.88% и 66.42%.

(Далее мы рассмотрим более легкий подход к решению подобной проблемы через пропорции.)

<u>Решение 5.3</u>

Вероятность выбрать 1 красную и 2 черных карты из колоды, состоящей из R красных карт равна:

$$3*\frac{R(52-R)(51-R)}{52*51*50}$$

т.к. каждая из ситуаций : (RBB),(BRB),(BBR) - имеет одинаковую вероятность.

Функция правдоподобия равна:
$P(1\text{кр.}+2\text{чер.}|R=0)=0$
$P(1\text{кр.}+2\text{чер.}|R=13)=3(13)(39)(38)/(52)(51)(50)=114k$
$P(1\text{кр.}+2\text{чер.}|R=26)=3(26)(26)(25)/(52)(51)(50)=100k$
$P(1\text{кр.}+2\text{чер.}|R=39)=3(39)(13)(12)/(52)(51)(50)=36k$
$P(1\text{кр.}+2\text{чер.}|R=52)=0$
где $k=\frac{3*13^2}{52*51*50}$

Функция априорного распределения является константой (1/5 в каждом случае), т.к. выбирается один случайным образом.

Тогда из формулы "Апостериорное распределение$\propto$ Априорное распределение $*$ Функция правдоподобия апостериорное распределение того, что $R = 0, 13, 26, 39, 52$ следующее:
0, 45.6%, 40%, 14.4%, 0

Решение 5.4

Априорное распределение λ $Exp(\lambda')$ плотностью
$e^{\lambda'\lambda}$, $\lambda > 0$

X_i имеют $Exp(\lambda)$ распределение, поэтому функция правдоподобия пропорциональна
$\lambda^n e^{-(\sum x_i + \lambda')\lambda}$ $\lambda > 0$

Сравнивая полученную формулу с распределениями в таблице, видим, что она имеет ту же форму и значения, что и Gamma-распределение (помните, что λ - случайная величина) с параметрами $n+1$ и $\sum x_i + \lambda'$

<u>Решение 5.5</u>

Нам необходимо выбрать такое $\overline{\theta}$, чтобы минимизировать ожидаемые потери $E = E[I(\overline{\theta} \neq \theta)]$.
Из определения соотношения и индикатора следует, что выражение равно $1 * P(\overline{\theta} \neq \theta)$.
Итак нам необходимо выбрать такое $\overline{\theta}$, чтобы минимизировать вероятность того, что она не равна θ или максимизировать вероятность того, что она равна θ.
По определению это достигается выбором моды апостериорного распределения θ.

<u>Решение 5.6</u>

Нам необходимо выбрать такое $\overline{\theta}$, чтобы минимизировать ожидаемые потери $E = E[|\overline{\theta} - \theta|]$.

Если $f(\theta)$ апостериорное распределение, мы можем разделить соотношение на две части:
$E = \int\limits_{-\infty}^{\overline{\theta}} (\overline{\theta} - \theta) f(\theta) d\theta - \int\limits_{\overline{\theta}}^{\infty} (\overline{\theta} - \theta) f(\theta) d\theta$
Дифференцируя по $\overline{\theta}$, получаем:
$\frac{E}{d\overline{\theta}} = \int\limits_{-\infty}^{\overline{\theta}} f(\theta) d\theta - \int\limits_{\overline{\theta}}^{\infty} f(\theta) d\theta$
Приравнивая к нулю, получаем:
$\int\limits_{-\infty}^{\overline{\theta}} f(\theta) d\theta = \int\limits_{\overline{\theta}}^{\infty} f(\theta) d\theta$
Итак,$\overline{\theta}$ - медиана апостериорного распределения.

<u>Решение 5.7</u>

Апостериорное распределения для Пуассоновской функции правдоподобия и априорное распределение даны в таблице, как $Gamma(\sum x + 1, n + \lambda')$ (Проверьте, что вы можете получить эти результаты). Имеем $n = 10, \sum x = 31, \lambda' = 0.2$.
Следовательно апостериорное распределение для λ - $Gamma(32, 10.2)$.
Используя квадратичную функцию ущерба, получаем, что Байесовская оценка - математическое ожидание апостериорного распределения, т.е. $\overline{\lambda} = \frac{32}{10.2} = 3.14$

Решение 5.8

Апостериорное распределение для λ - $Gamma(32, 10.2)$.
Используя то, что если X имеет $Gamma(\alpha, \delta)$, тогда $2\delta X$ - $\chi^2_{2\alpha}$, имеем 20.4λ имеет χ^2_{64} распределение.
Интерполируя данные таблицы между χ^2_{60} и χ^2_{70} получаем
$0.95 \doteq P(43.79 < 20.4\lambda < 87.99) = P(2.15 < \lambda < 4.31)$

Таким образом 95% Байесовский доверительный интервал для λ равен (2.15,4.31).

Решение 5.9

Из таблиц результатов, представленных раньше в главе, получаем, что апостериорная вероятность для μ равна $N\left(\frac{49.32}{10}, \frac{8}{10}\right) = N(4.932, 0.8)$.
Итак 95% доверительный интервал равен:
$4.932 \pm 1.96 * \sqrt{0.8} = (3.18, 6.69)$

Часть VI