МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.ЛОМОНОСОВА
Механико-математический факультет
Кафедра теории вероятностей

Байесовские методы оценивания в страховании

Bayesian approach in insurance

Курсовая работа
Студента 4 курса 409 группы
Королькова Даниила Александровича

Научный руководитель
доктор физико-математических наук, профессор
Булинская Екатерина Вадимовна

Москва, 2023г.

Содержание

1 Введение

Зарождение известной теоремы Байеса можно проследить в его эссе, посвящённому решению одной из задач "Доктрины Случайности"("The Doctrine of Chances") - первой книге по теории вероятностей впервые опубликованной в 1718 году Муавром. Сегодня под теоремой Байеса мы понимаем следующее выражение

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$

несложно получаемое за два шага из определения условной вероятности.

Определение(Условная вероятность)

Пусть (Ω, F, P) - вероятностное пространство. Пусть $B \in \Omega$ и $P(B) > 0$. Тогда условной вероятностью А при условии В назовём следующее выражение

$$P(A|B) = \frac{P(AB)}{P(B)}.$$

Здесь $P(A)$ - априорная (prior) вероятность, $P(A|B)$ - апостериорная (posterior) вероятность. К слову, $P(B|A)$ может быть интерпретирована, как правдоподобие, а $c = P(B)$ - нормирующий множитель.

Так же отметим, что если множество Ω имеет разбиение (возможно, счётное), то есть $\Omega = \bigsqcup\limits_{i=0}^{n} B_i$, то $P(A) = \sum\limits_{i=0}^{n} P(A|B_i)P(B_i)$. Последнее принято называть формулой полной вероятноти.

Подобное соотношение распространияется и на плотности, а именно:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}.$$

Сегодня школа статистики поделилась на два класса: частотная и байесовская школа. Отличие байесовского подхода заключается в том, что мы заранее делаем предположение о распределении случайной величины. Причём разумно делать такое предположение опираясь на некоторые прошлые наблюдения. Дабы всё стало понятнее, перейдём, наконец, к примеру.

Пример 1

Предположим, что у нас есть некоторый посёлок, в котором уже расположено несколько домов. Новый застройщик решил застроховать свой объект от пожара, который в этом местности не редок. Предлагая ему контракт, мы заранее обговариваем условаия возмещений и рассчитываем риски, учитывая три варианта исхода: пожар может быть сильным (θ_1), слабым (θ_2) или не быть вовсе (θ_3). При этом по опыту прошлых лет, нам известно, что сильные пожары бывали в 20% домов, средние - в 30% и в остальных 50% домов пожаров не было вовсе. Иными словами, о случайных величинах $\theta_1, \theta_2, \theta_3$ мы обладаем неким априорным знанием, поэтому на этом этапе мы совершаем предположение о их распределении. Так, $\theta_1 \sim Bern(p = 0.2), \theta_2 \sim Bern(0.3), \theta_3 \sim Bern(0.5)$.

Так же для нашей модели рассмотрим некоторые характеристики (переменные, features), которыми обладают наблюдения (то есть дома). Для упрощения, скажем, что y_1 - переменная, отвечающая за наличие у дома металлического перекрытия, а y_2 - наличие у дома деревянных окон. Безусловно, на практике рассматриваются не только бинарные переменны, но и непрерывные (скажем, площадь дома) и категориальные (с конечным множеством значений, скажем, количество этажей), тогда вместо вероятностей всюду далее удобнее рассматривать плотности.

Рассмотрим первую характеристику домов. Известно, что металлическое перекрытие было у 10% домов, подвергшихся сильному пожару, 30% домов со средним пожаром и среди тех, кого пожар не затронул, находилось 75% домов с характеристикой $y_1 = 1$.

Рассмотрим вторую характеристику. Известно, что деревянные окна были у 45% домов, окутаннх сильным пожаром, 30% домов со средним и среди тех, кого пожар не затронул, было 30% домов характеристикой $y_2 = 1$.

Из полученных наблюдений можно вывести условные вероятности, а именно $P(y_1 = 1|\theta_1 = 1) = 0.1, P(y_1 = 1|\theta_2 = 1) = 0.3, P(y_1 = 1|\theta_3 = 1) = 0.75$ и $P(y_2 = 1|\theta_1 = 1) = 0.45, P(y_2 = 1|\theta_2 = 1) = 0.3, P(y_2 = 1|\theta_3 = 1) = 0.3$. Ясно так же, что вероятность противоположноего события $P(y_i = 0|\theta_j) = 1 - P(y_i = 1|\theta_j), i = 1, 2, j = 1, 2, 3$.

Теперь, когда все необходимые наблюдения проведены, можно подходить к рассчёту вероятности возгорания нового дома застройщика. Пусть нам известны его характеристики $y_1 = 1$ и $y_2 = 1$ и по ним, лишь по двум переменным, мы и будем предсказывать вероятность возгорания.

$$P(\theta_1 = 1|y_1 = 1) = \frac{P(y_1=1|\theta_1=1)P(\theta_1)}{P(y_1=1)} = \frac{P(y_1=1|\theta_1=1)P(\theta_1=1)}{\sum\limits_{i=1}^{3} P(y_1=1|\theta_i=1)P(\theta_i=1)} = 0.041,$$
$$P(\theta_2 = 1|y_1 = 1) = \frac{P(y_1=1|\theta_2=1)P(\theta_2)}{P(y_1=1)} = \frac{P(y_1=1|\theta_2=1)P(\theta_2=1)}{\sum\limits_{i=1}^{3} P(y_1=1|\theta_i=1)P(\theta_i=1)} = 0.186,$$
$$P(\theta_3 = 1|y_1 = 1) = \frac{P(y_1=1|\theta_3=1)P(\theta_3)}{P(y_1=1)} = \frac{P(y_1=1|\theta_3=1)P(\theta_3=1)}{\sum\limits_{i=1}^{3} P(y_1=1|\theta_i=1)P(\theta_i=1)} = 0.773$$

$$P(\theta_1 = 1|y_2 = 1) = \frac{P(y_2=1|\theta_1=1)P(\theta_1)}{P(y_2=1)} = \frac{P(y_2=1|\theta_1=1)P(\theta_1=1)}{\sum\limits_{i=1}^{3} P(y_2=1|\theta_i=1)P(\theta_i=1)} = 0.222,$$
$$P(\theta_2 = 1|y_2 = 1) = \frac{P(y_2=1|\theta_2=1)P(\theta_2)}{P(y_2=1)} = \frac{P(y_2=1|\theta_2=1)P(\theta_2=1)}{\sum\limits_{i=1}^{3} P(y_2=1|\theta_i=1)P(\theta_i=1)} = 0.222,$$
$$P(\theta_3 = 1|y_2 = 1) = \frac{P(y_2=1|\theta_3=1)P(\theta_3)}{P(y_2=1)} = \frac{P(y_2=1|\theta_3=1)P(\theta_3=1)}{\sum\limits_{i=1}^{3} P(y_2=1|\theta_i=1)P(\theta_i=1)} = 0.556$$

Остюда видим, что наиболее вероятный исход для дома с характеристиками $y_1 = 1$ и $y_2 = 1$ - это третий исход - отсутствие пожара. Более того, это апостериорное наблюдение скорректировало наши априорные знания.

Так, если бы мы рассчитывали наши ожидаемы убытки лишь от этого контракта лишь по одной характеристике j, то необходимо было бы просуммировать стоимости возмещений (p_1, p_2, p_3 по каждому из страховых случаев соответственно) с весами вероятностями, т.е. $E(p) = \sum\limits_{i=1}^{3} p_i P(\theta_i = 1|y_j = 1)$

Так же, инетересным представляется случай, когда переменные $y_i i = 1, ..., n$ независимы. В таком случае $P(\theta_j = 1|y_1 = 1, ..., y_n = 1) = \frac{P(y_1=1|\theta_j=1)...P(y_n=1|\theta_j=1)P(\theta_j=1)}{\sum\limits_{i=1}^{3} P(y_1=1|\theta_i=1)...P(y_n=1|\theta_j=1)P(\theta_i=1)}$.

2 Оценки MLE и MAP

2.1 MLE

Вспомним содержание метода максимума правдомподобия (Maximum Likelihood Estimation). Если нам известна плотность $p(y|\theta)$ некоторого признака, зависящего от неизвестного параметра θ, то для его оценки необходимо максимизировать $L(\theta|y) := p(y|\theta) = \prod\limits_{i=1}^{n} p(y_i|\theta)$.

Пример 2

Немного видоизменим наш первый пример и скажем, что пожар может либо случиться, либо нет. Это событие есть ни что иное, как бернуллиевская случайная величина с неизвестным нам параметром $\theta \in \Theta = [0, 1]$. Оценим же вероятность возникновения пожара:

$$P(y_i = 1|\theta) = \theta^{y_i}(1-\theta)^{1-y_i},$$

$$l(\theta|y) = log(L(\theta|y)) = \sum_{i=1}^{n} log(P(y_i|\theta)) = \sum_{i=1}^{n} y_i log\theta + (1-y_i)log(1-\theta) \to \max_{\theta\in\Theta}$$

$$\hat{\theta}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Такая оценка получена после логарифмирования (именно это монотонное преобразование упрощает поиск экстремума) и дифференцирования экстремального выражения по параметру θ и приравнивания результата к нулю. Оценка полученная таким путём достаточно проста и очевидна: если хотим оценить вероятность пожара, то достаточно просто взять среднее арифметическое всех наблюдаемых исходов ранее. Однако, если наших наблюдей было мало, то скорее всего будет перевес в какую-то сторону, скажем, если в 9 случаев из 10 был пожар, то такая оценка выдаст значение $\frac{9}{10}$, что может быть далеко от истины. Словом, как и для любой состоятельньой оценки нам необходимо достаточно много наблюдений.

2.2 MAP

В Байсовском же подходе дела обстоят несколько иначе. Предположим, что мы не обладаем никаким знанием о истинном распределении параметра θ и можем лишь предположить, что априорное распределение равномерное и $p(\theta) = 1$ на $[0, 1]$. Пускай мы пронаблюдали лишь за одним домом в нашем посёлке и в нём пожар произошёл, то есть $p(y|\theta) = \theta$.

Теперь же несложно найти и апостериорную плотность, зная, что $p(y) = \frac{1}{2}$. Так по формуле Байеса получим $p(\theta|y) = 2\theta$

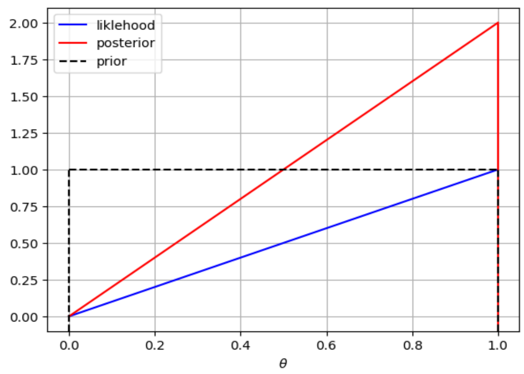

Рис. 1: Распределение неизвестного параметра

Теперь перейдём ко второму дому в нашем посёлке и теперь, вместо того, чтобы доверять первоначальному априорному распределению $p(\theta) = 1$, возьмём в качестве нового априорного распределения только что полученное апостериорное. Пусть во втором доме пожара-таки не было, то есть $p(y|\theta) = 1 - \theta$. Тогда

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int\limits_\theta p(y|\theta)p(\theta)d\theta} = 6\theta * (1 - \theta)$$

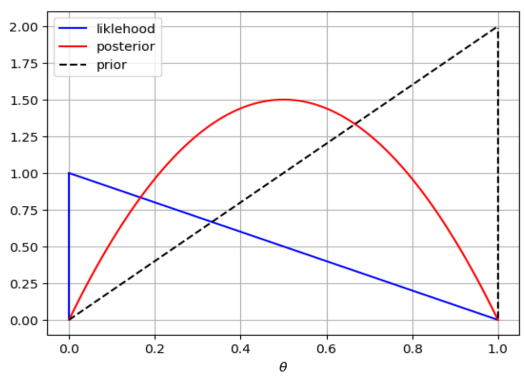

Рис. 2: После второго наблюдения

Ну и для большей удовлетворённости перейдём к третьему дому. Откалибруем вновь априорное распределение $p(\theta) = 6\theta * (1 - \theta)$, пожара в нём не было, т.е. $p(y|\theta) = 1 - \theta$ и несложными вычислениями получаем

$$p(\theta|y) = 12\theta(1 - \theta)^2$$

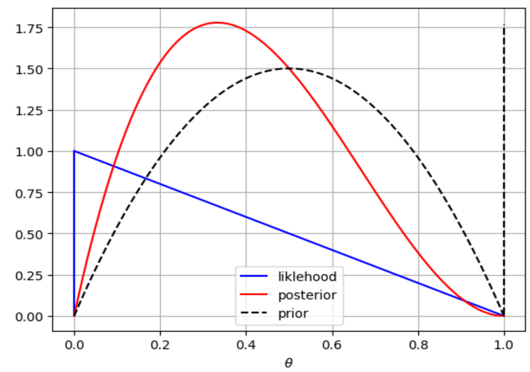

Рис. 3: После третьего наблюдения

Так, проделав несколько итераций, мы уже можем заметить отличие от MLE - сейчас мы работали с распределениями и наши оценки - распределения, в то время как метод максимального правдоподобия - точечное оценивание и его оценка - некий скаляр. И чтобы сейчас ответить на вопрос, какое θ выбрать, востользуемся максимизациец постериорного распределения (Maximum a posteriori = MAP). Подобно MLE, прологарифмируем плотность и найдём критическую точку:

$$logp(\theta|y) = logp(y|\theta) + logp(\theta) - log(y) \to \max_{\theta \in \Theta} \qquad (*)$$
$$\sim$$
$$logp(y|\theta) + logp(\theta) \to \max_{\theta \in \Theta}$$

То есть та же задача максимизация правдоподобия плюс некий добавок в виде априорного распределения.

2.3 Связь MLE и MAP

Ясно, что MLE является частным случаем MAP. Действительно, если бы априорное распределение не зависило от θ (например, так, как это было в первой итерации), то и задача (*) была бы эквивалентна задачи максимизации правдоподобия

3 Сопряжённые распределения

В примере выше на каждом новом шаге мы получали новое распределение. На самом деле, мы получали Бета-распределение. Напомню, что это распределение случайной величины X с параметрами $\alpha, \beta > 0$ и плотностью вида:

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)},$$
$$B(\alpha,\beta) = \int\limits_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$$

Вот примеры некотороых Бета-распределений:

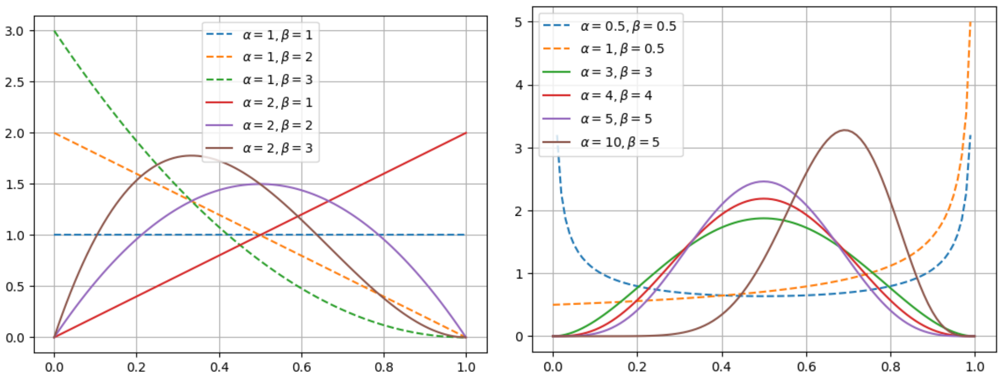

Рис. 4: Плотности Бета-распределения

В случае Бета-распределения получим следующую оценку:

$$\hat{\theta}_{MAP} = \frac{\sum\limits_{i=1}^{n} y_i + (\alpha - 1)}{m + (\alpha - 1) + (\beta - 1)}$$

И кажется, что на каждом шаге вычислений, нам не составит труда посчитать числитель в формуле Байеса и найти оценку неизвестного параметра, не смотря на то, что в знаменателе может стоять выражение, подсчёт которого может быть сильно затруднён. Так вот как раз, чтобы таких проблем не возникало и результат подсчёта апостериорного распределения мог бы использоваться в следующей итерации, как априорное, стоит обратить внимание на так называемые сопряжённые распределения - такое семейство пар, которые не выводят нас их класса априорного распределения (как в этом случае Бернулли и Бета-распределения).

Определение

Пусть $x \sim \mathcal{O}(\theta), \theta \sim \mathcal{B}(\beta_0)$. Тогда семейства $\mathcal{O}$ и $\mathcal{B}$ называются сопряжёнными, если $p(\theta|x) \sim \mathcal{B}(\beta_1)$

Вот некоторые из распространённых пар сопряжённых распределений:

- $p(y|\theta)$ - распределение Бернулли с параметром θ, $p(\theta)$ - Бета-распределение.
- $p(y|\mu)$ - нормальное с некоторым матожиданием μ и фиксированной дисперсией σ^2, и $p(\theta)$ - нормальное.
- $p(y|\lambda)$ - показательное с параметром λ, $p(\lambda)$ - Гамма-распределение
- $p(y|\lambda)$ - пуассоновское с параметром λ, $p(\lambda)$ - Гамма-распределение
- $p(y|\theta)$ - равномерное на отрезке $[0, \theta]$, $p(\theta)$ - Парето

Зачастую легче работать с экспоненциальным классом распределений, для него всегда удобно подбирать сопряжённое $p(\theta)$. К слову, большинство перечисленных выше распределений как раз-таки экспоненциальные.

4 MCMC

5 Список литературы

[1] Учебник по машинному обучению. Глава: 4.6. Байесовский подход к оцениванию.
[2] А.Г.Дьяконов. "Анализ малых данных", глава: Байесовский подход.
[3] Christopher M. Bishop. "Pattern Recognition and Machine Learning"
[4] Alexandra Hotti. "Bayesian insurance pricing using informative prior estimation techniques".