

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	1
2	Оценки MLE (Maximum Likelihood Estimation) и MAP (Maximum a posteriori)	3
2.1	MLE	3
2.2	MAP	3
2.3	СВЯЗЬ MLE и MAP	4
3	Сопряжённые распределения	4
4	Заключение	7
5	Список литературы	8
6	Приложения	8

1 Введение

В данной работе рассматривается возможное применение байесовского подхода к задачам страхования.

Зарождение известной теоремы Байеса можно проследить в его эссе, посвящённом решению одной из задач "Доктрины Случайности" ("The Doctrine of Chances") [1] - первой книге по теории вероятностей впервые опубликованной в 1718 году Муавром. Сегодня под байесовским подходом 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 1. (Условная вероятность)

Пусть A - событие, причём $P(A) > 0$. Условной вероятностью события B при условии A называют следующую величину:

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

Теорема 1. Пусть (Ω, \mathcal{F}, P) - вероятностное пространство.

Пусть $B_i, 1 \leq i \leq n$ - непересекающиеся события, объединение которых даёт всё Ω . Тогда верно следующее:

1. (Формула полной вероятности) $P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$
2. (Формула Байеса) $P(B_l|A) = \frac{P(B_l)P(A|B_l)}{P(A)} = \frac{P(B_l)P(A|B_l)}{\sum_{i=1}^n P(A|B_i)P(B_i)}$

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

Подобное соотношение распространяется 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. Подробнее про различие подходов можно прочитать в работе Alexandra Hotti [2]. Отмечу, что основное 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.

Пример 1. Предположим, что у нас есть некоторый посёлок, в котором уже расположено несколько домов. Новый застройщик решил застраховать свой объект от пожара, который в этой местности не редок. Предлагая ему контракт, мы заранее обговариваем условия 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 об их распределении. Так, $\theta_1 \sim \text{Bern}(p = 0.2)$, $\theta_2 \sim \text{Bern}(0.3)$, $\theta_3 \sim \text{Bern}(0.5)$.

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 переменные, но и непрерывные (скажем, площадь дома) и категориальные (с конечным множеством значений, скажем, количество этажей).

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, окутанных 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 противоположного 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 расчёту 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.

$$\begin{aligned} P(\theta_1 = 1|y_1 = 1) &= \frac{P(y_1=1|\theta_1=1)P(\theta_1)}{P(y_1=1)} = \frac{P(y_1=1|\theta_1=1)P(\theta_1=1)}{\sum_{i=1}^3 P(y_1=1|\theta_i=1)P(\theta_i=1)} = 0.041, \\ P(\theta_2 = 1|y_1 = 1) &= \frac{P(y_1=1|\theta_2=1)P(\theta_2)}{P(y_1=1)} = \frac{P(y_1=1|\theta_2=1)P(\theta_2=1)}{\sum_{i=1}^3 P(y_1=1|\theta_i=1)P(\theta_i=1)} = 0.186, \\ P(\theta_3 = 1|y_1 = 1) &= \frac{P(y_1=1|\theta_3=1)P(\theta_3)}{P(y_1=1)} = \frac{P(y_1=1|\theta_3=1)P(\theta_3=1)}{\sum_{i=1}^3 P(y_1=1|\theta_i=1)P(\theta_i=1)} = 0.773 \\ P(\theta_1 = 1|y_2 = 1) &= \frac{P(y_2=1|\theta_1=1)P(\theta_1)}{P(y_2=1)} = \frac{P(y_2=1|\theta_1=1)P(\theta_1=1)}{\sum_{i=1}^3 P(y_2=1|\theta_i=1)P(\theta_i=1)} = 0.222, \\ P(\theta_2 = 1|y_2 = 1) &= \frac{P(y_2=1|\theta_2=1)P(\theta_2)}{P(y_2=1)} = \frac{P(y_2=1|\theta_2=1)P(\theta_2=1)}{\sum_{i=1}^3 P(y_2=1|\theta_i=1)P(\theta_i=1)} = 0.222, \\ P(\theta_3 = 1|y_2 = 1) &= \frac{P(y_2=1|\theta_3=1)P(\theta_3)}{P(y_2=1)} = \frac{P(y_2=1|\theta_3=1)P(\theta_3=1)}{\sum_{i=1}^3 P(y_2=1|\theta_i=1)P(\theta_i=1)} = 0.556 \end{aligned}$$

Отсюда 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 ожидаемые 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.

$$E(p) = \sum_{i=1}^3 p_i P(\theta_i = 1|y_j = 1)$$

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

2 Оценки MLE (Maximum Likelihood Estimation) и MAP (Maximum a posteriori)

2.1 MLE

Вспомним содержание метода максимума правдоподобия (Maximum Likelihood Estimation). Если нам известна плотность $p(y|\theta)$ некоторого признака, зависящего от неизвестного параметра θ , то для его оценки необходимо максимизировать правдоподобие $l(\theta|y) := p(y|\theta) = \prod_{i=1}^n p(y_i|\theta)$. Подробнее с MLE можно познакомиться в книге Pattern Recognition and Machine Learning [3].

Пример 2. Немного видоизменим наш первый пример и скажем, что пожар может либо случиться, либо нет. Это событие есть не что иное, как бернуллиевская случайная величина с неизвестным нам параметром $\theta \in \Theta = [0, 1]$. Оценим же вероятность возникновения пожара:

$$\begin{aligned} P(y_i = 1|\theta) &= \theta^{y_i}(1 - \theta)^{1-y_i}, \\ L(\theta|y) = \log(l(\theta|y)) &= \sum_{i=1}^n \log(P(y_i|\theta)) = \sum_{i=1}^n y_i \log \theta + (1 - y_i) \log(1 - \theta) \rightarrow \max_{\theta \in \Theta} \\ \hat{\theta}_{MLE} &= \frac{1}{n} \sum_{i=1}^n y_i. \end{aligned}$$

Такая оценка получена после логарифмирования (именно это монотонное преобразование упрощает поиск экстремума) и дифференцирования выражения по параметру θ , а затем 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 ранее исходов. Однако, если наших наблюдений было мало, то скорее всего будет перевес в какую-то сторону, скажем, если в 9 случаях из 10 был пожар, то такая оценка выдаст значение $\frac{9}{10}$, что может быть далеко от истины. Словом, как и для любой состоятельной оценки нам необходимо достаточно много наблюдений.

2.2 MAP

В байесовском же подходе дела обстоят несколько иначе. Предположим, что мы не обладаем никаким знанием об 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 $p(\theta|y) = 2\theta$

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

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int_{\Theta} p(y|\theta)p(\theta)d\theta} = 6\theta(1 - \theta)$$

Ну и для большей удовлетворённости перейдём к третьему дому. Вновь примем в качестве априорного распределения апостериорное, полученное на прошлом шаге, $p(\theta) = 6\theta(1 - \theta)$, пожара в нём не было, т.е. $p(y|\theta) = 1 - \theta$ и несложными вычислениями получаем

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

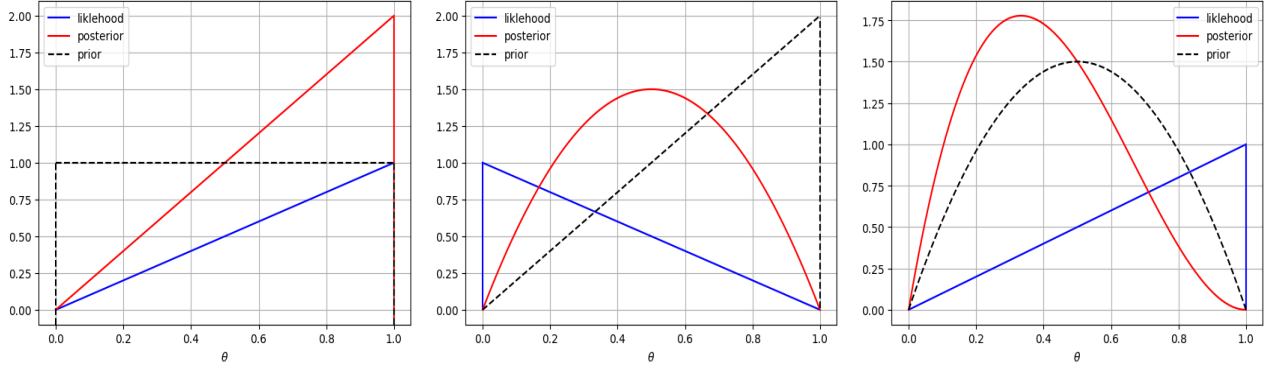


Рис. 1: Распределение неизвестного параметра после 1, 2 и 3-го наблюдения

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, воспользуемся максимизацией апостериорного распределения (Maximum a posteriori = MAP). Подобно MLE, прологарифмируем плотность и найдём критическую точку:

$$\begin{aligned} \log p(\theta|y) &= \log p(y|\theta) + \log p(\theta) - \log(y) \rightarrow \max_{\theta \in \Theta} \quad (*) \\ &\sim \\ \log p(y|\theta) + \log p(\theta) &\rightarrow \max_{\theta \in \Theta}. \end{aligned}$$

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 зависело от θ (например, так, как это было в первой итерации), то и задача (*) была бы эквивалентна задаче 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:

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

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx.$$

Вот примеры некоторых бета-распределений:

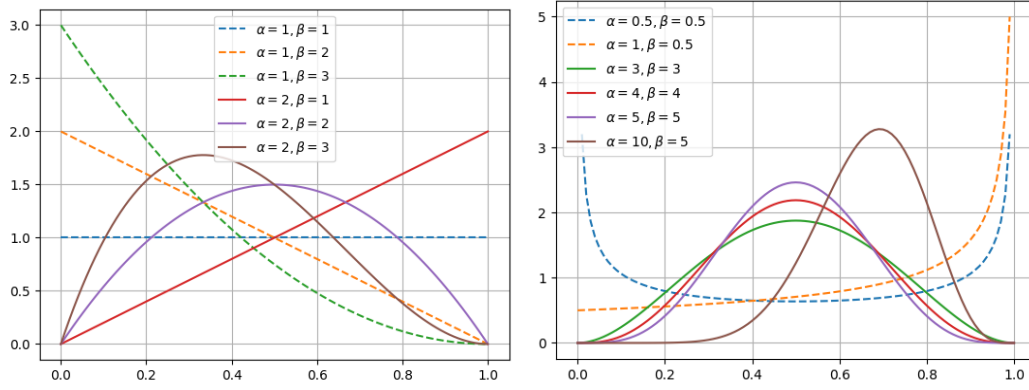


Рис. 2: Плотности бета-распределения

В случае бета-распределения получим следующую оценку:

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

И кажется, что на каждом шаге вычислений, нам не составит труда посчитать числитель в формуле Байеса и найти оценку неизвестного параметра, несмотря 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 2. 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 математическим ожиданием 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.

Пример 3. Обратимся к пуассоновскому распределению. Напомню, что удобно использовать именно это распределение для описания вероятностей событий, происходящих за промежутки времени t , и работать с пуассоновским процессом $N(t)$. Вероятностная функция масс такого распределения с параметром (интенсивностью) $\lambda \in (0, \infty)$ имеет вид:

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

а процесс в момент времени t будет распределён по закону Пуассона с параметром λt .

Если же нам не известна интенсивность, то мы можем сделать предположение о её распределении $\pi(\lambda|\nu)$, причём выбирать стоит из класса гамма-распределений, дабы работать с сопряжёнными семействами распределений. Напомню, что если случайная величина распределена по гамма-закону с параметрами $\nu = (a, b)$, $a, b > 0$, то её функция плотности имеет вид:

$$f(x) = \begin{cases} x^{a-1} b^a e^{-bx} \frac{1}{\Gamma(a)} & x > 0 \\ 0 & x \leq 0 \end{cases}.$$

Пусть в течении n промежутков времени ($T = t_1 + \dots + t_n$) в прошлом произошло $K = k_1 + \dots + k_n$ событий. Так, положим тогда $\pi(\lambda|\nu) = \text{Gamma}(a + K, b + T)$. Тогда вероятность k событий, в момент времени t равна:

$$\begin{aligned} P(k) &= \int_0^\infty p(k|\lambda) \pi(\lambda|\nu) d\lambda = \int_0^\infty \frac{(\lambda t)^k e^{-\lambda t}}{k!} \lambda^{a+K-1} (b+T)^{a+K} e^{-(b+T)\lambda} \frac{1}{\Gamma(a+K)} d\lambda = \\ &= \frac{(b+T)^{a+K}}{k! \Gamma(a+K)} \int_0^\infty (\lambda t)^k e^{-\lambda t} \lambda^{a+K-1} e^{-(b+T)\lambda} d\lambda = \frac{t^k (b+T)^{a+K}}{k! \Gamma(a+K)} \int_0^\infty \lambda^{k+a+K-1} e^{-\lambda(t+b+T)} d\lambda = \\ &= \left[x = \frac{\lambda}{t+b+T} \right] = \frac{t^k (b+T)^{a+K}}{k! \Gamma(a+K) (t+b+T)^{k+a+K}} \int_0^\infty x^{k+a+K-1} e^{-x} dx = \\ &= \frac{\Gamma(a+k+K)}{k! \Gamma(a+K)} \left(\frac{t}{t+b+T} \right)^k \left(\frac{b+T}{t+b+T} \right)^{a+K}. \end{aligned}$$

Тогда вероятность того, что число событий не превысит Q равна:

$$\sum_{k=0}^Q \frac{\Gamma(a+k+K)}{k! \Gamma(a+K)} \left(\frac{t}{t+b+T} \right)^k \left(\frac{b+T}{t+b+T} \right)^{a+K}$$

Исследуем теперь следующую модель: пусть есть производитель, выпускающий детали по стоимости x с себестоимостью $c < x$, есть покупатель, готовый купить n таких деталей. При этом на каждую деталь устанавливается гарантийный срок t и производитель готов выплатить компенсацию y за каждое бракованное устройство свыше количества z (за бракованные устройства меньшего количества компенсации не предполагается).

В таком случае ожидаемая прибыль производителя $n(x - c)$. Ожидаемые убытки $(i - z)y$, но лишь в случае $i > z$. Теперь для моделирования выплат перейдём к плотности количества брака. Пусть $i \sim p(i|\lambda)$. В таком случае математическое ожидание выплат $\sum_{i=z+1}^\infty y(i - z)p(i|\lambda)$, отсюда математическое ожидание прибыли производителя $n(x - c) - \sum_{i=z+1}^\infty y(i - z)p(i|\lambda)$. Используя результат для априорного распределения, полученный выше, имеем математическое ожидание прибыли:

$$n(x - c) - y \sum_{k=z+1}^n \frac{(k-1)\Gamma(a+k+K)}{k!\Gamma(a+K)} \left(\frac{t}{t+b+T}\right)^k \left(\frac{b+T}{t+b+T}\right)^{a+K}$$

Пример 4. Производитель продаёт $n = 100$ деталей по цене $x = 220\$$ и себестоимостью $c = 180\$$ за единицу товара. Гарантийные обязательства действуют $t = 1$ год. Известно, что в прошлой партии за $T = 3$ года бракованными оказались $K = 2$ детали. Производитель готов возместить убытки y за каждую бракованную продукцию свыше $i = 1$ детали. Требуется найти оптимальное значение y .

Будем использовать априорное распределение, полученное выше, с параметрами $a = 1, b = 1$. В таком случае ожидаемая прибыль $100 \cdot 40 - \sum_{k=2}^{100} \frac{(k-1)\Gamma(3+k)}{\Gamma(3)k!} \left(\frac{4}{5}\right)^3 \left(\frac{1}{5}\right)^k \approx 4000 - y \cdot 0.262$. Отсюда получаем $y \leq 15267\$$.

4 Заключение

В работе были рассмотрены некоторые методы применения байесовского подхода к задачам страхования. Такой подход используется для оценки вероятности наступления страхового случая на основе имеющихся данных. Вот несколько примеров применения байесовского подхода в страховании:

- **Оценка вероятности наступления страхового случая на основе исторических данных.** Страховая компания может использовать данные о прошлых страховых случаях, чтобы оценить вероятность наступления подобного случая в будущем. Байесовский подход позволяет учитывать новые данные и корректировать оценки вероятности на основе этой информации.
- **Прогнозирование потерь от страховых случаев.** Например, страховая компания может использовать данные о страховых случаях, произошедших в определенном регионе и в определенный период времени, чтобы оценить вероятность наступления страхового случая в будущем и определить размер страховой премии.
- **Определение оптимальных условий страхования.** Например, страховая компания может использовать данные о клиенте (возраст, пол, место жительства и т. д.), чтобы оценить вероятность наступления страхового случая и предложить наиболее подходящие условия страхования.

В целом, байесовский подход в страховании позволяет компаниям качественнее оценивать риски и принимать более обоснованные решения при определении условий страхования и расчёте страховых премий, за счёт построения более точных моделей прогнозирования[4].

5 Список литературы

- [1] A. De Moivre. The Doctrine of Chances. 1718.
- [2] Alexandra Hotti. Bayesian insurance pricing using informative prior estimation techniques. 2020.
- [3] Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006.
- [4] Uraivan Jaroengeratikun, Winai Bodhisuwan, Ampai Thongteeraparp. A Bayesian Inference of Non-Life Insurance Based on Claim Counting Process with Periodic Claim Intensity. 2012.

6 Приложения

- [1] Код с графиками распределений