Глава 5

Байесовский подход и его применение в задачах страхования, гарантийного обслуживания и принятия решений с проведением экспериментов

Будущее нельзя предвидеть, но можно изобрести.

Денис Габор

5.1. Основные положения байесовского подхода

Во многих задачах принятия решений априорная вероятностная информация о состояниях природы может быть изменена после получения новых экспертных оценок или в результате наблюдения соответствующих событий, связанных с состояниями и подтверждающих или опровергающих априорную информацию.

Как отмечается в справочнике [31], многие статистические задачи независимо от методов их решения обладают общим свойством: до того как получен конкретный набор данных, в качестве потенциально приемлемых для изучаемой ситуации рассматривается несколько вероятностных моделей. После того как получены данные, возникает выраженное в некотором виде знание об относительной приемлемости этих моделей. Одним из способов "пересмотра" относительной приемлемости вероятностных моделей является *байесовский подход*, основой которого выступает известная *теорема Байеса*.

Несмотря на то что так называемая традиционная частотная школа статистического вывода, представленная работами таких ученых, как Фишер, Ньюмен, Пирсон, и многими другими, доминирует в статистике в настоящее время, байесовские методы показали чрезвычайно стремительное развитие в последние десятилетия. Причина этого заключается в том, что байесовский подход имеет ряд существенных преимуществ, которые делают его достаточно привлекательным для широкого применения [32].

Основное отличие байесовского подхода от других статистических подходов состоит в том, что до того, как будут получены данные, лицо, принимающее решение, или статистик рассматривает степени своего доверия к возможным моделям и представляет их в виде вероятностей. Как только данные получены, теорема Байеса позволяет рассчитать новое множество вероятностей, которые представляют пересмотренные степени доверия к возможным моделям, учитывающие новую информацию, поступившую благодаря данным.

Статистические данные зачастую отсутствуют в реальных задачах анализа риска и принятия решений, что делает использование многих традиционных частотных подходов неправомерным [1]. Имеющаяся в распоряжении информация может содержать только субъективные оценки в виде экспертных оценок и суждений. Более того, ситуация, в которой принимается решение, может быть вообще новой и никогда ранее не анализируемой. Эти особенности усложняют процесс принятия решений и могут поставить под сомнение какие-либо выводы и заключения. Поэтому в такой ситуации байесовский подход может оказаться весьма полезным и эффективным.

Основой байесовского подхода в статистике является теорема Байеса. Предположим, что мы наблюдаем некоторую случайную величину Y, которая имеет плотность вероятности $p(y|\theta)$ с параметрами θ, но хотим сделать вывод о другой случайной величине θ, имеющей некоторое распределение вероятностей $\pi(\theta)$. Пусть в результате наблюдений получены статистические данные y (значения случайной величины Y). С одной стороны, из определения условной вероятности следует, что

$$\Pr(\theta|y) = \Pr(y,\theta)/\Pr(y).$$

С другой стороны, из определения условной вероятности также следует, что

$$\Pr(y,\theta) = \Pr(y|\theta)\Pr(\theta).$$

Подставляя второе равенство в первое, получаем известную формулу Байеса:

$$\Pr(\theta|y) = \frac{\Pr(y|\theta)\Pr(\theta)}{\Pr(y)}.$$

Если имеется m возможных переменных $(\theta_1, ..., \theta_m)$, то

$$\Pr(\theta_j|y) = \frac{\Pr(y|\theta_j)\Pr(\theta_j)}{\Pr(y)} = \frac{\Pr(y|\theta_j)\Pr(\theta_j)}{\sum_{i=1}^{m}\Pr(\theta_i)\Pr(y|\theta_i)}.$$

Распределение $\Pr(\theta)$ называется *априорным* распределением вероятностей возможных значений θ (это распределение принимается прежде, чем получены статистические данные). Распределение $\Pr(\theta|y)$ называется *апостериорным* распределением значений θ при условии, что наблюдались данные y (это распределение вычисляется после получения статистических данных). При этом переменные θ_j называются *гипотезами*, а y – *свидетельствами*, поддерживающими гипотезы.

Пример 5.1. Предположим, что необходимо определить надежность некоторой фирмы. При этом все фирмы можно разделить на три основные группы в зависимости от надежности. Первая группа – фирмы средней надежности, вторая группа – фирмы высокой надежности, третья группа – фирмы низкой надежности. Таким образом при анализе конкретной фирмы имеются три гипотезы: θ_i – фирма принадлежит i-й группе, $i = 1, 2, 3$. При этом из общей статистики организаций, занимающихся аналогичной деятельностью, известно, что 50% фирм имеют среднюю надежность, 30% фирм имеют высокую надежность, 20% фирм имеют низкую надежность. Используя эти данные, можно определить априорные вероятности гипотез: $\Pr(\theta_1) = 0.5$, $\Pr(\theta_2) = 0.3$, $\Pr(\theta_3) = 0.2$. Одними из признаков

Таблица 5.1. Априорные и условные вероятности

i	1	2	3
$\Pr(\theta_i)$	0.5	0.3	0.2
$\Pr(y_1\|\theta_i)$	0.4	0.8	0.3
$\Pr(y_2\|\theta_i)$	0.7	0.9	0.0

(свидетельствами) надежности фирмы является наличие прибыли у фирмы (y_1) и своевременный расчет с бюджетом (y_2). В реальных задачах таких признаков имеется обычно намного больше. Однако мы ограничимся только двумя, чтобы показать использование байесовского подхода. Из анализа аналогичных предприятий известно, что прибыль имеют 40% фирм средней надежности, 80% фирм высокой надежности и 30% фирм низкой надежности. Отсюда можно записать условные вероятности $\Pr(y_1|\theta_1) = 0.4$, $\Pr(y_1|\theta_2) = 0.8$ и $\Pr(y_1|\theta_3) = 0.3$. Также известно, что выплачивают в бюджет 70% фирм средней надежности, 90% фирм высокой надежности и 0% фирм низкой надежности. Отсюда можно записать условные вероятности $\Pr(y_2|\theta_1) = 0.7$, $\Pr(y_2|\theta_2) = 0.9$ и $\Pr(y_2|\theta_3) = 0$. Все исходные данные приведены в табл. 5.1.

Необходимо отметить, что условные вероятности $\Pr(y^c|\theta_i)$ противоположных свидетельств определяются из очевидного условия $\Pr(y^c|\theta_i) = 1 - \Pr(y|\theta_i)$. Так, если фирма не имеет прибыли, то $\Pr(y_1^c|\theta_1) = 0.6$, $\Pr(y_1^c|\theta_2) = 0.2$, $\Pr(y_1^c|\theta_3) = 0.7$.

В процессе сбора фактов вероятности гипотез будут повышаться, если факты поддерживают их, или уменьшаться, если факты опровергают их. Предположим, что мы имеем только одно свидетельство y_1, т.е. с вероятностью единица конкретная фирма имеет прибыль. Наблюдая y_1, мы вычисляем апостериорные вероятности для гипотез согласно формуле Байеса для одного свидетельства:

$$\Pr(\theta_1|y_1) = \frac{\Pr(y_1|\theta_1)\Pr(\theta_1)}{\sum_{i=1}^{3}\Pr(y_1|\theta_i)\Pr(\theta_i)} = 0.4,$$

$$\Pr(\theta_2|y_1) = \frac{\Pr(y_1|\theta_2)\Pr(\theta_2)}{\sum_{i=1}^{3}\Pr(y_1|\theta_i)\Pr(\theta_i)} = 0.48,$$

$$\Pr(\theta_3|y_1) = \frac{\Pr(y_1|\theta_3)\Pr(\theta_3)}{\sum_{i=1}^{3}\Pr(y_1|\theta_i)\Pr(\theta_i)} = 0.12.$$

Из результатов расчетов видно, что после того как y_1 произошло (стало известно, что фирма имеет прибыль), доверие к гипотезам

θ_1 и θ_3 понизилось, в то время как доверие к θ_2 возросло. Таким образом, мы модифицировали априорные вероятности и получили апостериорные вероятности гипотез. Если продолжить модификацию или пересчет на основе новых свидетельств, то полученные выше апостериорные вероятности гипотез становятся априорными. Например, предположим, что анализируемая фирма платит в бюджет, т.е. наблюдается y_2. Тогда апостериорные вероятности вычисляются по тем же формулам с заменой вероятностей $\Pr(\theta_i)$ вероятностями $\Pr(\theta_i|y_1)$ и свидетельств y_1 свидетельствами y_2, вычисленными на предыдущем этапе. В результате расчетов получаем $\Pr(\theta_1|y_2) = 0.393$, $\Pr(\theta_2|y_2) = 0.607$, $\Pr(\theta_3|y_2) = 0$.

Необходимо отметить, что те же результаты можно получить за один шаг, предполагая, что одновременно получены два свидетельства и эти свидетельства независимы. В этом случае апостериорная вероятность имеет вид

$$\Pr(\theta_j|y_1, y_2) = \frac{\Pr(y_1, y_2|\theta_j)\Pr(\theta_j)}{\sum_{i=1}^{3}\Pr(y_1, y_2|\theta_i)\Pr(\theta_i)}.$$

Предположение независимости свидетельств позволяет записать

$$\Pr(y_1, y_2|\theta_j) = \Pr(y_1|\theta_j) \cdot \Pr(y_2|\theta_j).$$

Отсюда

$$\Pr(\theta_j|y_1, y_2) = \frac{\Pr(y_1|\theta_j) \cdot \Pr(y_2|\theta_j)\Pr(\theta_j)}{\sum_{i=1}^{3}\Pr(y_1|\theta_i) \cdot \Pr(y_2|\theta_i) \cdot \Pr(\theta_i)}.$$

Аналогичные результаты можно получить и для непрерывного случая. Пусть $\theta = (\theta_1, ..., \theta_m)$ – вектор m непрерывных случайных величин, определенный на множестве Ω^m. Тогда теорема Байеса принимает вид

$$p(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{p(y)} = \frac{p(y|\theta)\pi(\theta)}{\int_{\Omega^m} p(y, \theta)\mathrm{d}\theta}.$$

Здесь $\pi(\theta)$ – априорная плотность вектора параметров θ, $p(\theta|y)$ – апостериорная плотность, вычисляемая исходя из предположения априорной плотности $\pi(\theta)$ и наблюдаемых данных y.

Метод максимума функции правдоподобия, используемый в статистическом оценивании параметров распределения, и байесовский подход достаточно близки. Пусть $L(\theta|y)$ – функция правдоподобия. Метод максимума функции правдоподобия заключается в том, что параметры $\theta \in \Omega^m$ выбираются таким образом, чтобы максимизировать функцию правдоподобия как функцию от θ. Правомерность использования метода максимума функции правдоподобия основывается на свойствах выборки большого объема, а именно если размер выборочных данных достаточно большой, то можно предположить:

1) нормальность статистики среднего значения случайной величины;

2) статистика отношения правдоподобия имеет χ^2-распределение.

Однако эти замечательные свойства не выполняются, если размер выборки мал. Поэтому альтернативным подходом является попытка начать статистический вывод с некоторых исходных предположений (догадок) о распределении неизвестных параметров $\pi(\theta)$. Тогда, используя теорему Байеса, можно вычислить значения апостериорного распределения:

$$p(\theta|y) = \frac{1}{p(y)} p(y|\theta)\pi(\theta).$$

Заметим, что $p(y|\theta) = L(\theta|y)$ – функция правдоподобия, $1/p(y)$ – некоторая постоянная величина по отношению к параметрам θ. Поэтому апостериорное распределение пропорционально произведению функции правдоподобия на априорное распределение или

$$p(\theta|y) \propto L(\theta|y)\pi(\theta),$$

где символ $\propto$ означает пропорциональность.

Если в задаче о надежности фирм функция правдоподобия была известна в явном виде, т.е. в виде набора вероятностей $p(y|\theta)$ (условных к значению параметра), то в общем случае эта функция представляется в виде некоторой функции от параметров, что значительно усложняет вычисления.

Постоянная величина $p(y)$ вычисляется исходя из условия нормализации, т.е. сумма апостериорных вероятностей или интеграл апостериорной плотности по множеству значений параметров θ равен единице. Тогда, например, для непрерывного случая

$$p(y) = \int_{\Omega^m} p(y|\theta)\pi(\theta)\mathrm{d}\theta.$$

Зависимость апостериорных вероятностей от априорных показывает, как много информации о значениях неизвестного параметра содержится в статистических данных. Если апостериорные вероятности сильно зависят от априорных, то, скорее всего, данные содержат мало информации. Если апостериорные вероятности слабо зависят от выбора априорного распределения, то данные являются информативными.

Таким образом, при использовании байесовского подхода кроме распределения вероятностей рассматриваемой случайной величины Y предполагается использование некоторого априорного распределения параметров θ функции распределения величины Y. Опираясь на статистические данные, априорное распределение параметров θ модифицируется путем умножения на функцию правдоподобия и нормализации. Результатом модификации является апостериорное распределение параметров θ. Другими словами, параметры распределения случайной величины сами являются случайными величинами с некоторым распределением, т.е. мы имеем неопределенность второго порядка: “случайные параметры случайной величины” или “распределение параметров распределения”.

Как теперь найти окончательно оценки параметров θ? Для простоты будем считать, что вектор состоит из одного элемента θ. Наибольшее распространение получили три точечные оценки параметра θ.

1. *Мода.* Оценка $\widehat{\theta}$ параметра выбирается исходя из максимума апостериорной плотности, т.е.

$$p(\widehat{\theta}|y) = \max_{\theta} p(\theta|y).$$

2. *Медиана.* Оценка $\widehat{\theta}$ параметра выбирается исходя из равенства $\Pr(\theta > \widehat{\theta}|y) = \Pr(\theta < \widehat{\theta}|y) = 0.5$ или

$$\int_{\widehat{\theta}}^{+\infty} p(\theta|y)\mathrm{d}\theta = \int_{-\infty}^{\widehat{\theta}} p(\theta|y)\mathrm{d}\theta = 0.5.$$

3. *Среднее.* Оценка $\widehat{\theta}$ параметра вычисляется как математическое ожидание, т.е.

$$\widehat{\theta} = \mathbb{E}[\theta|y] = \int_{-\infty}^{+\infty} \theta p(\theta|y)\mathrm{d}\theta.$$

Наиболее важным и одновременно сложным является вопрос выбора априорного распределения параметров. Одним из факторов здесь является то, что при наличии информативных статистических данных даже “плохое” априорное распределение не повлияет существенно на апостериорное. Другим важным фактором является сложность вычислений, особенно если расчеты апостериорного распределения производятся последовательно по мере поступления статистической информации. Поэтому на выбор априорного распределения влияет его принадлежность к так называемому классу *согласованных* распределений, т.е. таких распределений, что априорное и апостериорное являются одним и тем же распределением, но с разными параметрами. При этом согласованность определяется не только видом априорного распределения, но и видом функции правдоподобия, т.е. тип распределения должен сохраняться при умножении априорного распределения на функцию правдоподобия с учетом нормализации. К таким распределениям относятся: гамма-распределение, если функция правдоподобия является пуассоновской; бета-распределение, если функция правдоподобия является биномиальной; распределение Дирихле, если функция правдоподобия является полиномиальной. Подробное описание целого ряда согласованных распределений можно найти в справочнике [31].

С другой стороны, до появления каких-либо статистических данных или наблюдений о параметрах зачастую ничего не известно. Поэтому если отбросить требование простоты вычислений, то предпочтительное или “хорошее” (подходящее) априорное распределение должно минимально влиять на вывод, т.е. на апостериорное распределение, а также учитывать отсутствие априорной информации о параметрах. Априорные распределения, моделирующие отсутствие априорной информации, называются *неинформативными*.

Постулат Байеса–Лапласа говорит от том, что, когда заранее ничего не известно о параметре θ, априорное распределение следует принимать равномерным, т.е. все возможные исходы случайной величины θ имеют равные вероятности. Основной проблемой использования равномерного распределения в качестве неинформативного априорного распределения является то, что равномерное распределение неинвариантно по отношению к функциям параметра. Если мы ничего не знаем о параметре θ, то мы также ничего не знаем и, например, о функции $1/\theta$. Однако если θ имеет равномерное распределение, то $1/\theta$ уже не имеет равномерного распределения, хотя согласно постулату Байеса–Лапласа, $1/\theta$ должно иметь равномерное распределение. Кроме того, равномерное распределение нельзя использовать в качестве априорного, если множество значений параметра бесконечно. Следует также отметить существенную зависимость равномерного распределения от множества исходов. Можно привести следующий пример, иллюстрирующий эту зависимость. Пусть в мешке имеются шары красного цвета и каких-то других цветов. При этом неизвестно, сколько шаров красного цвета и сколько всех остальных шаров в мешке. Какова вероятность того, что первый вытащенный из мешка шар будет красный? Предполагая равномерное распределение, получаем вероятность 1/2. Пусть теперь в мешке имеются шары красного цвета, зеленого цвета и каких-то других. Тогда вероятность красного шара равна 1/3. Ситуация во втором случае принципиально не изменилась, так как зеленые шары можно отнести к шарам “другого цвета” применительно к первой ситуации. Однако мы получили совершенно разные вероятности красного шара.

В литературе существует достаточно большое количество подходов для выбора того или иного неинформативного априорного распределения [31], имеющих свои достоинства и недостатки. Однако наиболее интересным является подход, полностью отличающийся от большинства традиционных. Суть этого подхода заключается в следующем. Определим не одно априорное распределение, а целый класс $\mathcal{M}$ распределений π, для которого можно найти нижнюю и верхнюю вероятности события A как

$$\underline{P}(A) = \inf\{P_\pi(A) : \pi \in \mathcal{M}\},$$
$$\overline{P}(A) = \sup\{P_\pi(A) : \pi \in \mathcal{M}\}. \quad (5.1)$$

При определенных условиях множество $\mathcal{M}$ полностью определяется нижней и верхней функциями распределения вероятностей. Следует отметить, что класс $\mathcal{M}$ следует рассматривать "не как класс подходящих априорных распределений, а как подходящий класс априорных распределений". Это значит, что каждое отдельное распределение из класса не является подходящим или "хорошим" априорным распределением, так как ни одно отдельное распределение не может удовлетворительно моделировать отсутствие информации. Но весь класс в целом, определяемый верхним и нижним распределениями вероятностей, является подходящей моделью отсутствия информации. Когда априорной информации почти нет, $\underline{P}(A)$ для этого класса должно быть близко к 0, а $\overline{P}(A)$ близко к 1. Это означает, что априори событие A может иметь любую вероятность. В большинстве случаев класс распределений $\mathcal{M}$ определяется параметрически. Поэтому минимум и максимум в (5.1) обычно вычисляются по некоторым параметрам.

5.2. Распределение Дирихле

Для описания распределения Дирихле рассмотрим стандартную полиномиальную модель. Пусть $\Omega = \{\omega_1, ..., \omega_m\}$ – множество возможных исходов и имеется совокупность N наблюдений, независимо выбранных из Ω с одинаковыми вероятностями каждого исхода $\Pr\{\omega_j\} = \theta_j$ для всех $j = 1, ..., m$, где

$\theta_j \geq 0$ и $\sum_{j=1}^{m} \theta_j = 1$. Вероятность того, что из N наблюдений исход ω_j будет наблюдаться N_j раз, определяется из известной формулы полиномиального распределения с параметрами $\theta_1, ..., \theta_m$. Однако параметры $\theta_1, ..., \theta_m$ могут быть сами случайными величинами и иметь некоторое распределение или плотность вероятности $\pi(\theta)$. Одним из наиболее интересных распределений параметров (вероятностей) $\theta_1, ..., \theta_m$ является *распределение Дирихле*, которое является согласованным с полиномиальным распределением в том смысле, что априорное и апостериорное распределения являются распределениями Дирихле.

Априорное распределение Дирихле $(s, \mathbf{t})$ для случайного вектора вероятностей $\theta = (\theta_1, ..., \theta_m)$, где $\mathbf{t} = (t_1, ..., t_m)$, имеет функцию плотности вероятности

$$\pi(\theta) = \Gamma(s) \left(\prod_{j=1}^{m} \Gamma(st_j) \right)^{-1} \cdot \prod_{j=1}^{m} \theta_j^{st_j - 1}.$$

Здесь параметр $t_i \in (0, 1)$ является средним значением (математическим ожиданием) вероятности θ_i; параметр $s > 0$ определяет влияние априорного распределения на апостериорные вероятности; вектор $\mathbf{t}$ принадлежит внутренней области единичного симплекса размерности m, который будем обозначать $S(1, m)$; $\Gamma(\cdot)$ – гамма-функция, удовлетворяющая условиям $\Gamma(x + 1) = x\Gamma(x)$ и $\Gamma(1) = 1$. Необходимо отметить, что переменными распределения Дирихле являются вероятности $\theta_1, ..., \theta_m$, удовлетворяющие условию $\theta \in S(1, m)$. Это означает, что не только сами события считаются случайными, но и их вероятности.

Пример 5.2. Вернемся к примеру с надежностью фирм (см. пример 5.1). В этом примере можно выделить три события ($m = 3$): фирма имеет среднюю надежность (первое событие – ω_1), высокую надежность (второе событие – ω_2), низкую надежность (третье событие – ω_3). Вероятности этих событий являются случайными величинами, и их математические ожидания равны t_1, t_2, t_3. При этом предположение о том, что априорные вероятности надежности фирмы равны 0.5, 0.3, 0.2 соответственно, означает, что математические ожидания вероятностей надежности фирмы для априорного распределения Дирихле равны $t_1 = 0.5$, $t_2 = 0.3$, $t_3 = 0.2$. Однако вероятности надежности фирмы могут быть в действительности любые с плотностью вероятности $\pi(\theta)$.

После получения вектора наблюдений $\mathbf{n} = (n_1, ..., n_m)$, где n_j – число наблюдений исхода ω_j, умножая априорную плотность на функцию правдоподобия $L(\mathbf{n}|\theta)$, получаем апостериорную плотность:

$$p(\theta|\mathbf{n}) \propto \pi(\theta)L(\mathbf{n}|\theta) = \prod_{j=1}^{m} \theta_j^{n_j + st_j - 1},$$

которая может рассматриваться как плотность распределения Дирихле $(N + s, \mathbf{t}^*)$, где

$$\mathbf{t}^* = (t_1^*, ..., t_m^*), \quad t_j^* = \frac{n_j + st_j}{N + s}.$$

Другими словами, распределение Дирихле относится к классу согласованных распределений и при пересчете априорные параметры $\mathbf{t}$ преобразуются в $\mathbf{t}^*$. Это очень важное свойство, благодаря которому распределение Дирихле получило широкое распространение в байесовском анализе.

Пример 5.3. Предположим, что в результате обследования 10 фирм, было установлено, что 6 фирм имеют среднюю надежность, 3 фирмы – высокую надежность и 1 фирма – низкую надежность. Тогда $n_1 = 6$, $n_2 = 3$, $n_3 = 1$, $N = 10$. Пусть параметр s равен 1. Тогда апостериорные математические ожидания вероятностей надежности фирм становятся равны

$$t_1^* = \frac{6 + 1 \cdot 0.5}{10 + 1} = 0.591,$$
$$t_2^* = \frac{3 + 1 \cdot 0.3}{10 + 1} = 0.3,$$
$$t_3^* = \frac{1 + 1 \cdot 0.2}{10 + 1} = 0.109.$$

Таким образом, после получения новой статистической информации вероятности надежности фирм остаются случайными, но с новыми (апостериорными) математическими ожиданиями, полученными выше. При этом новая информация подтвердила априорное представление о вероятности фирм высокой надежности и изменила представление о вероятностях фирм низкой и средней надежности. С одной стороны, если принять $s = 0$, то апостериорные данные будут полностью определяться только имеющейся дополнительной информации в виде анализа 10 фирм. Полученные данные пересчета в этом случае не будут зависеть от выбора априорных вероятностей. С другой стороны, если параметр s принимает большие значения, то получаемая статистическая информация практически перестает влиять на апостериорные вероятности и поиск дополнительной информации теряет смысл.

Одним из важных свойств распределения Дирихле является то, что его маргинальные распределения также являются распределениями Дирихле с другими параметрами. В частности, если $m = 2$, то распределение Дирихле есть не что иное, как бета-распределение. Другими словами, распределение Дирихле есть обобщение бета-распределения при $m > 2$.

5.3. Бета-биномиальное распределение

Большое количество прикладных задач связано с необходимостью определения вероятности того, что в течение некоторого периода наблюдений произойдет не менее или не более заданного количества определенных событий. Если известна вероятность θ наступления события, то вероятность $p(k|\theta)$ того, что произойдет k событий из n возможных, определяется на основе биномиального распределения

$$p(k|\theta) = C_n^k\ \theta^k(1-\theta)^{n-k}.$$

Если вероятность θ точно неизвестна и она сама является случайным числом с некоторой функцией плотности вероятности $\pi(\theta|\vartheta)$, то вероятность того, что произошло ровно k событий, определяется из выражения

$$P(k) = \int_{\Omega} p(k|\theta) \cdot \pi(\theta|\vartheta) \mathrm{d}\theta.$$

Здесь ϑ – вектор параметров распределения π; $\Omega = [0, 1]$ – множество значений параметра θ.

В рассматриваемом случае для согласования априорного и апостериорного распределений в качестве функции π используется бета-распределение. Априорное бета-распределение случайной величины θ, обозначаемое $\mathrm{Beta}(a, b)$, имеет функцию плотности вероятности

$$\pi(\theta|a, b) = \mathrm{Beta}(a, b) = \frac{1}{\mathrm{B}(a, b)} \theta^{a-1} (1-\theta)^{b-1},\ 0 \leq \theta \leq 1.$$

Здесь $a > 0$, $b > 0$ – параметры распределения, т.е. $\vartheta = (a, b)$; $\mathrm{B}(a, b)$ – бета-функция.

Необходимо отметить, что бета-распределение используется в качестве априорного в рассматриваемой ситуации для поиска вероятности того, что произойдет k событий. В отличие от этого рассмотренный ранее частный случай распределения Дирихле используется для поиска вероятности того, что произойдет ровно одно событие.

Пусть в результате наблюдений из суммарного числа наблюдений N произошло K интересуемых нас событий. Тогда апостериорное бета-распределение $\pi(\theta|\vartheta, K)$ при условии получения статистических данных имеет следующий вид:

$$\pi(\theta|\vartheta, K) = \mathrm{Beta}(a + K, b + N - K).$$

Отсюда вероятность $P(k)$ того, что произойдет ровно k определенных событий в будущем при общем числе наблюдений n, равна

$$\begin{aligned}P(k) &= \int_{\Omega} p(k|\theta)\cdot\pi(\theta|\vartheta, K)\mathrm{d}\theta = \\ &= \int_0^1 C_n^k\theta^k(1-\theta)^{n-k}\cdot\mathrm{Beta}(a+K, b+N-K)\mathrm{d}\theta = \\ &= C_n^k\,\frac{\mathrm{B}(a+k+K, b+n+N-k-K)}{\mathrm{B}(a+K, b+N-K)}.\end{aligned}$$

Это – *бета-биномиальное распределение*. Вероятность того, что число интересуемых событий не превысит M, равна

$$P_M = \sum_{k=0}^{M} C_n^k\,\frac{\mathrm{B}(a+k+K, b+n+N-k-K)}{\mathrm{B}(a+K, b+N-K)}.$$

Рассмотрим применение бета-биномиального распределения в модели *индивидуального страхования*. Теория страхования и актуарные расчеты приобретают все большее значение в условиях развития рынка страховых услуг и повышающейся конкуренции на этом рынке страховых компаний. Поэтому в настоящее время существует огромное количество публикаций, посвященных определению параметров страхования, таких как оптимальный размер страховых премий, страховых выплат, оптимальное количество договоров и т.д. Случайная природа страховых случаев не позволяет однозначно определить параметры страхования и всегда существует некоторый риск разорения страховой компании. Актуарные расчеты призваны определять параметры страхования так, чтобы минимизировать риск разорения или ограничить его некоторым предельным значением [11].

Рассмотрим коротко известную модель индивидуального страхования дискретного времени, подробное описание и анализ которой представлен Бойковым [5]. Предположим, что компания заключает n однотипных договоров с одинаковым сроком действия t и возможными выплатами по i-му договору X_i. Страховая премия за каждый договор равна c. Тогда собранные к моменту времени t премии составляют $\Pi(t) = cn$, а суммарные выплаты

$$R(t) = X_1 + ... + X_n.$$

Здесь X_i – случайная величина, определяемая наступлением страхового случая за время t. Если обозначить индикатор наступления i-го страхового случая I_i, а размер выплаты y_i, то $X_i = I_i \cdot y_i$ и существует некоторая вероятность наступления i-го страхового случая $\Pr\{I_i = 1\} = q$. Вероятность неразорения определяется в этом случае как

$$P_M = \Pr\{\Pi(t) \geq R(t)\} = \Pr\{cn \geq X_1 + ... + X_n\}.$$

Если случайные величины $X_1, ..., X_n$ независимы, значения параметров c и n фиксированы, $y_i = y$ для всех $i = 1, ..., n$, то задача определения вероятности неразорения достаточно просто решается введением случайной величины K числа страховых случаев, имеющей биномиальное распределение $p(k|\theta)$ с параметром θ. При этом вероятность неразорения равна

$$P_M = \sum_{k=0}^{M} p(k|\theta).$$

Здесь M – максимальное количество исков, которые компания может оплатить, определяемое как

$$M = \lfloor \Pi(t)/y \rfloor = \lfloor cn/y \rfloor,$$

символ $\lfloor \cdot \rfloor$ означает целую часть числа "снизу".

Как видно из описания модели, вероятность неразорения может быть вычислена с использованием байесовского подхода, а именно с применением бета-биномиального распределения.

Пример 5.4. Рассмотрим числовой пример применения модели. Пусть страховая компания заключила $n = 100$ договоров. При наступлении страхового случая выплачивается сумма $y = 100$ по одному договору. Премия по одному договору составляет $c = 12$. Максимальное количество исков, которые компания может оплатить, равно

$$M = \lfloor cn/y \rfloor = 12 \cdot 100/100 = 12.$$

Предположим, что в результате наблюдений из суммарного числа $N = 20$ договоров ранее произошел $K = 1$ страховой случай. Тогда вероятность неразорения при выборе параметров $a = 1$ и $b = 1$ неинформативного априорного распределения равны

$$P_{12} = \sum_{k=0}^{12} C_{100}^{k} \ \frac{\mathrm{B}(a + k + 1, b + 100 + 20 - k - 1)}{\mathrm{B}(a + 1, b + 20 - 1)} =$$

$$= \sum_{k=0}^{12} C_{100}^{k} \ \frac{\mathrm{B}(1 + k + 1, 1 + 100 + 20 - k - 1)}{\mathrm{B}(1 + 1, 1 + 20 - 1)} = 0.744.$$

Как видно из полученных результатов, оценка вероятности неразорения 0.744 слишком пессимистична при столь небольшой выборке наблюдаемых в прошлом событий.

Предположим теперь, что из 200 договоров в прошлом произошло 10 страховых случаев (в той же пропорции). Тогда $P_{12} = 0.986$. Это говорит о существенной зависимости вычисляемой вероятности от априорного распределения при небольшом объеме статистических данных.

5.4. Отрицательное биномиальное распределение

Если рассматривается число определенных событий за некоторый интервал времени t, то одним из наиболее распространенных законов распределения для описания соответствующих вероятностей является распределение Пуассона с параметром λ, характеризующим интенсивность этих событий. Вероятность k событий за время t, в соответствии с распределением Пуассона, имеет следующий вид:

$$p(k) = \frac{(\lambda t)^k \exp(-\lambda t)}{k!}.$$

Если интенсивность событий λ точно неизвестна, то предполагая, что она сама является случайным числом с некоторой функцией плотности вероятности $\pi(\lambda|\vartheta)$, можно использовать байесовский подход. При этом для согласования априорного и апостериорного распределений функция π должна соответствовать плотности гамма-распределения. Априорное гамма-распределение случайной величины λ, обозначаемое

$\mathrm{Gamma}(a,b)$, имеет функцию плотности вероятности

$$\pi(\lambda) = \mathrm{Gamma}(a,b) = \\ = \frac{1}{\Gamma(a)} b^a \lambda^{a-1} \exp(-b\lambda), \ \ \lambda > 0.$$

Здесь $a > 0$, $b > 0$ – параметры распределения, т.е. $\vartheta = (a,b)$; $\Gamma(a)$ – гамма-функция.

Пусть в течении n периодов наблюдений в прошлом произошло в сумме $K = \sum_{i=1}^{n} k_i$ событий. Здесь k_i – число событий на i-м периоде наблюдений. Тогда апостериорная функция плотности вероятности $\pi(\lambda|K)$ гамма-распределения при условии получения статистических данных в форме K событий за n периодов наблюдений имеет следующий вид:

$$\pi(\lambda|K) = \mathrm{Gamma}(a+K, b+n).$$

Отсюда вероятность $P(k)$ того, что произойдет ровно k событий в будущем при условии $t = 1$, имеет вид:

$$P(k) = \int_0^\infty p(k|\lambda) \cdot \pi(\lambda|\vartheta) \mathrm{d}\lambda = \\ = \int_0^\infty \frac{\lambda^k \exp(-\lambda)}{k!} \cdot \mathrm{Gamma}(a,b) \cdot \mathrm{d}\lambda = \\ = \frac{\Gamma(a+k)}{\Gamma(a)k!} \cdot \left(\frac{b}{b+1}\right)^a \left(\frac{1}{b+1}\right)^k.$$

Это – *отрицательное биномиальное распределение.* Отсюда вероятность того, что число интересуемых событий не превысит M, равна

$$P_M = \sum_{k=0}^{M} \frac{\Gamma(a+k)}{\Gamma(a)k!} \cdot \left(\frac{b}{b+1}\right)^a \left(\frac{1}{b+1}\right)^k.$$

Вероятности $P(k)$ также определяются при помощи рекуррентной формулы

$$P(k) = \begin{cases} \left(\dfrac{b}{b+1}\right)^a, & k = 0 \\ \dfrac{a+k-1}{k(b+1)} \cdot P(k-1), & k \geq 1 \end{cases}.$$

Рассмотрим случай, когда времена предыдущих n периодов наблюдений не являются одинаковыми и равны соответственно $t_1, ..., t_n$. При этом суммарное время наблюдений равно $T = t_1 + ... + t_n$. Количество событий, наблюдаемое в каждом периоде, равно $k_1, ..., k_n$, а их суммарное число за время T равно $K = k_1 + ... + k_n$. Это более общий и интересный случай, так как на практике трудно ожидать, что исходные статистические данные собирались в течение тех же периодов времени, что и период прогнозирования t.

Вероятность k событий за время t при условии, что за период наблюдения t_1 произошло k_1 событий, равна

$$P(k) = \int_0^\infty \frac{(\lambda t)^k \exp(-\lambda t)}{k!} \cdot \mathrm{Gamma}(a + k_1, b + t_1) \mathrm{d}\lambda =$$
$$= \frac{\Gamma(a + k_1 + k)}{\Gamma(a + k_1) k!} \cdot \left(\frac{b + t_1}{b + t_1 + t}\right)^{a + k_1} \left(\frac{t}{b + t_1 + t}\right)^k .$$

Аналогичное выражение для $P(k)$ можно получить для произвольного значения $n \geq 1$. При этом, как следует из результирующих выражений, полученная вероятность не зависит от распределения числа событий $k_1, ..., k_n$ и от распределения времени наблюдения $t_1, ..., t_n$, а определяется только суммарным числом событий K и суммарным временем наблюдений T, т.е.

$$P(k) = \frac{\Gamma(a + K + k)}{\Gamma(a + K) k!} \cdot \left(\frac{b + T}{b + T + t}\right)^{a + K} \left(\frac{t}{b + T + t}\right)^k .$$

В частном случае, когда все временные интервалы одинаковы и имеют единичную длительность, вероятность k событий за единичный интервал времени имеет вид

$$P(k) = \frac{\Gamma(a + K + k)}{\Gamma(a + K) k!} \cdot \left(\frac{b + n}{b + n + 1}\right)^{a + K} \left(\frac{1}{b + n + 1}\right)^k .$$

Вероятность P_M определяется как

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

Необходимо отметить, что приведенное выражение для вероятности P_M может быть переписано в другом виде

$$P_M = 1 - \frac{\mathrm{B}_q(M+1,r)}{\mathrm{B}(M+1,r)},$$

где

$$r = a + K, \; q = \frac{t}{b+T+t};$$

$\mathrm{B}_q(M+1,r)$ – неполная бета-функция, определяемая как

$$\mathrm{B}_q(M+1,r) = \int_0^q x^M (1-x)^{r-1} \mathrm{d}x.$$

Если $q = 1$, то неполная бета-функция совпадает с бета-функцией, т.е. $\mathrm{B}_1(M+1,r) = \mathrm{B}(M+1,r)$.

Отрицательное биномиальное распределение с параметрами $a = 1$, $b = 1$ является одним из неинформативных априорных распределений.

Рассмотрим применение отрицательного биномиального распределения в модели гарантийных обязательств. Производство, закупки и продажа товара обычно связаны с риском того, что часть изделий после покупки или в процессе эксплуатации выйдет из строя, откажет или не сможет выполнять часть своих функций. Одним из способов минимизации или компенсации этого риска являются гарантийные обязательства. В соответствии с гарантийными обязательствами производитель определяет время, в течение которого отказавшие изделия будут отремонтированы, заменены или покупателю возмещен ущерб. Время действия гарантийного обязательства, величина возмещения ущерба, стоимость товара, его количество и т.д. могут рассматриваться как параметры гарантийных обязательств. При этом значения параметров выбираются так, чтобы потери вследствие отказов изделий в определенном смысле не превышали доходов от производства и продажи этих изделий. Поэтому неправильно оцененные параметры гарантийных обязательств могут привести к существенным потерям для производителя и продавца.

Оценка и оптимизация параметров гарантийных обязательств – одна из важных задач теории принятия решений. Как в любой задаче принятия решений, оптимальность решения или риск ошибочного решения полностью определяется имеющейся в распоряжении информацией о состояниях природы, которая в свою очередь зависит от знания распределения вероятностей, определенного на состояниях природы. В терминах гарантийных обязательств необходимо знать вероятностные характеристики надежности изделий или параметры распределений вероятностей, например числа отказавших изделий за заданное время.

Рассмотрим поток изделий между производителем и потребителем. Следующая схема гарантийных обязательств является достаточно типовой. Пусть n – число изделий, которые покупатель хотел бы приобрести. Здесь предполагается, что покупателем или потребителем изделий является магазин как промежуточная ступень между производителем и конечным покупателем. Можно рассматривать также и конечного покупателя, но тогда $n = 1$, и это является частным случаем общей схемы. Предположим, что все эти изделия одинаковые. На каждое изделие устанавливается срок гарантии t (единиц времени). Покупатель готов заплатить x (денежных единиц) за одно изделие, а также готов не требовать возмещения убытков при отказе не более z изделий в интервале времени $[0, t]$. Однако за каждый отказ изделия свыше z покупатель требует компенсации y денежных единиц за одно отказавшее изделие. В частном случае параметр z равен 0.

Задача оптимизации параметров гарантийных обязательств с точки зрения производителя заключается в выборе оптимальных в некотором смысле значений z и y.

Предположим, что себестоимость производства одного изделия равна c единиц. Тогда при отказе z или менее изделий в течение времени t прибыль производителя равна $B = n(x - c)$. Однако если произошло $i > z$ отказов изделий за время t, то производитель обязан выплатить по гарантийным обязательствам $(i-z)y$ и прибыль равна

$$B = n(x-c) - (i-z)y.$$

Заметим, что наиболее важный параметр гарантийных обязательств – время t. Поэтому для описания потока отказов в течение этого времени и для вычисления вероятности $p(i|\lambda)$ появления i отказов за это время распределение Пуассона с параметром λ является наиболее предпочтительным. Необходимо отметить, что биномиальное распределение также может использоваться здесь для описания вероятностей отказов.

Ожидаемые выплаты, в соответствии с гарантийными обязательствами, можно рассматривать как математическое ожидание

$$\sum_{i=z+1}^{n} y(i-z)p(i|\lambda).$$

Отсюда ожидаемый доход производителя при заданном распределении $p(i|\lambda)$ определяется как разность между прибылью $n(x-c)$ от продажи n изделий с учетом их цены, себестоимости и ожидаемыми выплатами в соответствии с гарантийными обязательствами, т.е.

$$\mathbb{E}_p B = n(x-c) - \sum_{i=z+1}^{n} y(i-z)p(i|\lambda). \tag{5.2}$$

Таким образом, ожидаемый доход зависит от интенсивности отказов λ. Если ожидаемый доход больше 0, то гарантийные обязательства можно считать неубыточными и эффективными. В противном случае, их необходимо пересмотреть в пользу производителя.

Заметим, что схема гарантийных обязательств близка к схеме страхования, но в отличие от цели задачи страхования цель данной задачи заключается в вычислении математического ожидания $\mathbb{E}_p B$, а не функции распределения вероятностей. Тем не менее схема доказательства остается аналогичной. Используя отрицательное биномиальное распределение, несложно получить следующее выражение для ожидаемого дохода при известных параметрах (a, b), а также при условии, что за время T наблюдений произошло K отказов:

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

Пример 5.5. Производитель продает партию товаров в размере $n = 100$ штук по цене $x = 200\$$ за единицу товара при его себестоимости $c = 160\$$. Срок гарантийных обязательств – $t = 1$ год. При испытаниях вышли из строя $K = 2$ изделия за суммарный период наблюдений $T = 3$. При отказе свыше $z = 1$ изделия осуществляется возмещение убытков. Найдем оптимальное значение выплат для возмещения убытков y за каждое отказавшее изделие. Используем сначала точное априорное отрицательное биномиальное распределение с параметрами $a = 1$ и $b = 1$. Тогда ожидаемый доход равен

$$\mathbb{E}_p B = 100 \cdot 40 - y \sum_{k=2}^{100} \frac{(k-1)\Gamma(3+k)}{\Gamma(3)k!} \left(\frac{4}{5}\right)^{3} \left(\frac{1}{5}\right)^{k} =$$
$$= 4000 - y \cdot 0.262.$$

Отсюда условие положительного ожидаемого дохода имеет вид:

$$4000 - y \cdot 0.262 \geq 0$$

и $y \leq 15267$. Это значение указывает на то, что производитель готов заплатить 15267\$ за каждую отказавшую в течение года единицу товара, чтобы иметь положительный ожидаемый доход.

5.5. Гамма-экспоненциальное распределение

Во многих случаях для описания распределения вероятностей непрерывных случайных величин X, например времени безотказной работы элементов оборудования, используют экспоненциальное распределение, которое имеет следующий вид:

$$\Pr\{X < t\} = F(t|\lambda) = 1 - e^{-\lambda t}.$$

Здесь λ – параметр распределения или интенсивность рассматриваемых событий. Если интенсивность событий λ точно неизвестна, то предполагая, что она сама является случайным числом с некоторой функцией плотности вероятности $\pi(\lambda|\vartheta)$, можно снова использовать байесовский подход. При этом для согласования априорного и апостериорного распределений функция π должна быть гамма-распределением $\mathrm{Gamma}(a, b)$.

Используя преобразования, представленные выше, получим вероятность того, что за время t в будущем произойдет событие

$$\begin{aligned}\Pr\{X < t\} = P(t) = \\ &= \int_0^\infty (1 - e^{-\lambda t}) \cdot \mathrm{Gamma}(a, b) \mathrm{d}\lambda = \\ &= 1 - \left(\frac{b}{b+t}\right)^a.\end{aligned}$$

Это – гамма-экспоненциальное распределение.

Если за время T произошло K событий, то апостериорная вероятность того, что за время t в будущем произойдет одно событие, например отказ, равна

$$\Pr\{X < t\} = P(t) = 1 - \left(\frac{b+T}{b+t+T}\right)^{a+K}.$$

5.6. Принятие решений с проведением экспериментов

Если вернуться к примеру в разделе 2.7, в котором рассматривался вопрос реализации продукции при проведении анализа рынка и при его отсутствии, то можно заметить, что в нем были указаны точные вероятности трех возможных исходов реализации точной продукции (см. рис. 2.4 и раздел 2.7). Однако во многих ситуациях принятия решений эти вероятности могут быть известны в виде некоторых априорных вероятностей, которые характеризуют не конкретную ситуацию, а отражают наши знания об аналогичных ситуациях, имеющих место в прошлом. Более того, в ряде ситуаций эти вероятности могут вообще отсутствовать.

В задаче с реализацией продукции для определения или уточнения вероятностей возможных исходов использовался предварительный анализ рынка. При этом все анализируемые события из Ω (успешная реализация, средняя реализация, плохая реализация) совпадали с событиями на рынке реализации продукции, что существенно упрощает задачу принятия решений. Однако во многих ситуациях можно определить только вероятности некоторых косвенных признаков рассматриваемых событий, которые так или иначе связаны с этими событиями. Значения этих признаков составляют некоторое множество $\Theta = \{\theta_1, ..., \theta_l\}$. Следовательно, возникает задача определения вероятностей событий $\omega_1, ..., \omega_m$ по известным вероятностям событий $\theta_1, ..., \theta_l$. Более того, если каким-либо образом получены условные вероятности $p(\theta_j|\omega_i)$ наступления события θ_j при условии наступления события ω_i, $i = 1, ..., m$, $j = 1, ..., l$, то задача определения вероятностей искомых событий из Ω решается при помощи байесовского подхода, в соответствии с которым вероятность $p(\omega_i|\theta_j)$ определяется по формуле[1]

$$p(\omega_i|\theta_j) = \frac{\pi(\omega_i)p(\theta_j|\omega_i)}{\sum_{k=1}^{m} \pi(\omega_k)p(\theta_j|\omega_k)}.$$

Здесь $\pi(\omega_k)$ – априорные вероятности рассматриваемых событий или состояний природы.

Эта формула может быть также переписана в виде

$$p(\omega_i|\theta_j) = \frac{\Pr(\omega_i, \theta_j)}{\Pr(\theta_j)}.$$

Здесь $\Pr(\omega_i, \theta_j)$ – вероятность совместного наступления событий ω_i и θ_j.

Вернемся к задаче с реализацией продукции, предполагая, что реализация продукции непосредственно зависит от уровня дохода жителей региона и анализ реализуемости продукции

[1]Эта формула была приведена в разделе 5.1 с другими обозначениями.

сводится к определению доходности жителей. Другими словами, уровень дохода является косвенным признаком, определяющим частично реализуемость продукции. Все доходы для упрощения задачи разделим на две части: "низкие" (θ_1) и "высокие" (θ_2), т.е. $\Theta = \{\theta_1, \theta_2\}$. Новое дерево решений, учитывающее проведение анализа рынка для уточнения вероятностей реализации продукции, показано на рис. 5.1.

Из анализа аналогичных рынков известно, что низкий доход преобладал в 10% регионов с успешной реализацией, в 20% регионов со средней реализацией и 60% регионов с плохой реализацией. Отсюда можно записать условные вероятности $\Pr(\theta_1|\omega_1) = 0.1$, $\Pr(\theta_1|\omega_2) = 0.2$ и $\Pr(\theta_1|\omega_3) = 0.65$. Заметим, что для рассматриваемой ситуации выполняется равенство $\Pr(\theta_1|\omega_i) = 1 - \Pr(\theta_2|\omega_i)$. Отсюда $\Pr(\theta_2|\omega_1) = 0.9$, $\Pr(\theta_2|\omega_2) = 0.8$ и $\Pr(\theta_2|\omega_3) = 0.35$.

Предположим, что известны априорные вероятности рассматриваемых событий $\omega_1, \omega_2, \omega_3$, которые равны $\pi(\omega_1) = 0.4$, $\pi(\omega_2) = 0.4$, $\pi(\omega_3) = 0.2$. Следует отметить, что эти вероятности априори характеризуют реализуемость в любом регионе. Нас же интересует конкретный регион, для которого известно, что население с низким доходом преобладает. Тогда, используя байесовский подход, можно вычислить апостериорные вероятности событий $\omega_1, \omega_2, \omega_3$ для данного региона. Эти вероятности равны

$$p(\omega_1|\theta_1) = \frac{0.4 \cdot 0.1}{0.4 \cdot 0.1 + 0.4 \cdot 0.2 + 0.2 \cdot 0.65} = 0.16,$$

$$p(\omega_2|\theta_1) = \frac{0.4 \cdot 0.2}{0.4 \cdot 0.1 + 0.4 \cdot 0.2 + 0.2 \cdot 0.65} = 0.32,$$

$$p(\omega_3|\theta_1) = \frac{0.2 \cdot 0.65}{0.4 \cdot 0.1 + 0.4 \cdot 0.2 + 0.2 \cdot 0.65} = 0.52.$$

Если в данном регионе преобладает высокий доход, то

$$p(\omega_1|\theta_1) = \frac{0.4 \cdot 0.9}{0.4 \cdot 0.9 + 0.4 \cdot 0.8 + 0.2 \cdot 0.35} = 0.48,$$

$$p(\omega_2|\theta_1) = \frac{0.4 \cdot 0.8}{0.4 \cdot 0.9 + 0.4 \cdot 0.8 + 0.2 \cdot 0.35} = 0.427,$$

$$p(\omega_3|\theta_1) = \frac{0.2 \cdot 0.35}{0.4 \cdot 0.9 + 0.4 \cdot 0.8 + 0.2 \cdot 0.35} = 0.093.$$

Зная апостериорные вероятности, можно определить ожидаемые значения полезности для всех самых правых вершин шансов по формуле (2.1). Ожидаемая полезность при отказе от анализа равна

$$\mathbb{E}_\pi \mathbf{u}_1 = 6 \cdot 0.4 + 2 \cdot 0.4 - 3 \cdot 0.2 = 2.6.$$

Ожидаемая полезность при выполнении анализа, по результатам которого получено преобладание низких доходов, равна

$$\mathbb{E}_\pi \mathbf{u}_2 = 6 \cdot 0.16 + 2 \cdot 0.32 - 3 \cdot 0.52 = 0.04.$$

Ожидаемая полезность при выполнении анализа, по результатам которого получено преобладание высоких доходов, равна

$$\mathbb{E}_\pi \mathbf{u}_3 = 6 \cdot 0.48 + 2 \cdot 0.427 - 3 \cdot 0.093 = 3.455.$$

Приведенные ожидаемые полезности показаны на рис. 5.1 курсивом. Априорные вероятности низкого и высокого доходов равны

$$\Pr(\theta_j) = \pi(\omega_1)p(\theta_j|\omega_1) + \pi(\omega_2)p(\theta_j|\omega_2) + \pi(\omega_3)p(\theta_j|\omega_3), \ j = 1, 2.$$

Отсюда

$$\Pr(\theta_1) = 0.4 \cdot 0.1 + 0.4 \cdot 0.2 + 0.2 \cdot 0.65 = 0.25,$$

$$\Pr(\theta_2) = 0.4 \cdot 0.9 + 0.4 \cdot 0.8 + 0.2 \cdot 0.35 = 0.75.$$

Тогда можно вычислить ожидаемое значение полезности до проведения анализа

$$\begin{aligned}\mathbb{E}_\pi \mathbf{u}_0 &= \mathbb{E}_\pi \mathbf{u}_2 \cdot \Pr(\theta_1) + \mathbb{E}_\pi \mathbf{u}_3 \cdot \Pr(\theta_2) = \\ &= 0.04 \cdot 0.25 + 3.455 \cdot 0.75 = 2.601.\end{aligned}$$

Таким образом, в начальной позиции ожидаемая полезность или прибыль без проведения анализа равна 2.6 млн.

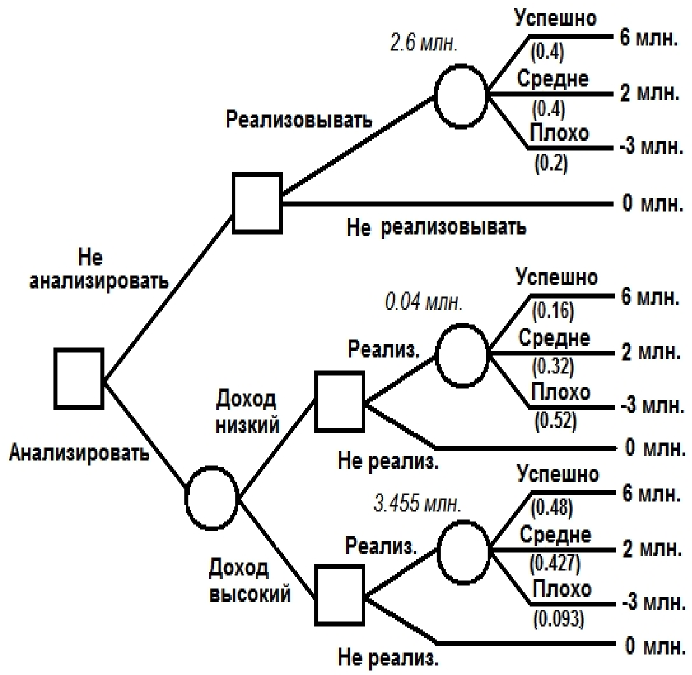

Рис. 5.1. Дерево решений с учетом апостериорных вероятностей

долларов, а ожидаемая прибыль при проведении анализа равна 2.601 млн. долларов. Следовательно, выполнение анализа имеет смысл, если его стоимость не превышает значения $2.601 - 2.6 = 0.001$ млн. долларов. Если это условие выполняется, то целесообразно проводить анализ и далее при любом исходе анализа осуществлять реализацию.

Завершая главу, можно отметить, что байесовские модели имеют целый ряд преимуществ по сравнению с частотными [33]. Одно из таких преимуществ заключается в том, что байесовские модели могут в принципе дать некоторые результаты, даже если совсем нет выборочных данных. Это обусловлено использованием априорного распределения вероятностей, которое при отсутствии статистических данных не изменяется, и пересчитанное по теореме Байеса апостериорное распределение совпадает с априорным.

В принятии решений байесовские модели позволяют проводить дополнительные эксперименты по уточнению состояний природы. На практике такое уточнение осуществляется в основном при помощи сбора дополнительной информации, а также с помощью проведения этих экспериментов. Это было наглядно продемонстрировано в примере 5.1, где дополнительная информация о прибыли и своевременном расчете с бюджетом позволяет уточнить априорные вероятности состояний природы (надежности фирмы).

В то же время, при планировании эксперимента необходимо определить, является ли он целесообразным с точки зрения затрат на его проведение. Для этого нужно сравнить ожидаемые дополнительные доходы или выгоды, которые можно получить при использовании дополнительной информации, приобретаемой в результате проведения эксперимента, и ожидаемые затраты на этот эксперимент. Так, для примера с анализом надежности фирмы получение информации о прибыли может потребовать проведения аудиторской проверки, на которую необходимо затратить определенные средства. Количество альтернативных действий в таких задачах увеличивается, так как появляется возможность не только выбора одной из имеющихся альтернатив, но и определения целесообразности проведения экспериментов или мероприятий для получения дополнительной информации. Примеры решения задач с проведением экспериментов можно также найти в книге [29].

5.7. Контрольные вопросы

1) Какое распределение вероятностей называется априорным?
2) Какое распределение вероятностей называется апостериорным?

3) Что является гипотезами в байесовском подходе?
4) Что является свидетельствами, поддерживающими гипотезы, в байесовском подходе?
5) Какие предположения, лежащие в основе метода максимума функции правдоподобия, используемого в статистическом оценивании параметров распределения, не всегда выполняются?
6) Какие распределения вероятностей называются неинформативными?
7) В чем заключается постулат Байеса–Лапласа?
8) Почему постулат Байеса–Лапласа не во всех случаях позволяет определить подходящее априорное распределение вероятностей?
9) Что означает согласованность распределений вероятностей в байесовском подходе?
10) Какой должна быть функция правдоподобия для согласования гамма-распределения?
11) Какой должна быть функция правдоподобия для согласования бета-распределение?
12) Каким должно быть априорное распределение для соблюдения согласованности, если функция правдоподобия является полиномиальной?
13) Какие особенности имеет модель индивидуального страхования дискретного времени?
14) Какие особенности имеет типовая схема гарантийных обязательств?
15) Может ли отрицательное биномиальное распределение применяться для моделирования индивидуального страхования дискретного времени?
16) Может ли бета-биномиальное распределение использоваться для определения параметров типовой схемы гарантийных обязательств?

17) Как оценить целесообразность проведения дополнительного эксперимента для уточнения вероятностей состояний природы?

5.8. Задачи

1) Определить целесообразность закупки партии изделии для продажи в розничной торговле в некотором регионе при наличии рекламы. Статистический анализ показал, что только в 40% регионов изделие пользуется спросом. Однако в 70% районах, где изделие пользуется спросом, есть его реклама. Реклама также имеется в 50% регионов, где изделие не пользуется спросом. а изделие пользуется спросом, а в 60% опрашиваемых хотели бы приобрести изделие, а 80% не нуждаются в нем.
2) Решить предыдущую задачу при условии, что в рассматриваемом регионе нет рекламы изделия.
3) Априори известно, что в 40% регионов некоторое изделие пользуется спросом. Однако в результате обследования 5 регионов, установлено, что в 3 регионах наблюдается высокий спрос, а в 2 – низкий. Используя распределение Дирихле с параметром $s = 1$, определить апостериорные математические ожидания вероятности высокого спроса.
4) Страховая компания заключила $n = 20$ договоров. При наступлении страхового случая выплачивается сумма $y = 40$ по одному договору. Премия по одному договору составляет $c = 6$. В результате наблюдений из суммарного числа $N = 10$ договоров ранее произошло $K = 2$ страховых случая. Используя бета-биномиальное распределение, определить вероятность разорения при выборе параметров $a = 1$ и $b = 1$ неинформативного априорного распределения.
5) За 100 дней продаж было куплено пять единиц определенного товара. Используя отрицательное биномиальное распределение определить вероятность того, что за 10 дней

будет куплено не более трех единиц товара. Параметры неинформативного априорного распределения – $a = 1$ и $b = 1$.

6) Производитель продает партию товаров в размере 50 штук по цене 80$ за единицу товара при его себестоимости 60$. Срок гарантийных обязательств равен двум годам. При испытаниях вышли из строя три изделия за суммарный период наблюдений четыре года. При любом отказе изделия осуществляется возмещение убытков. Используя отрицательное биномиальное распределение и параметры $a = 1$ и $b = 1$ неинформативного априорного распределения, определить максимальное значение компенсации за одно отказавшее изделие, чтобы обеспечить положительный ожидаемый доход.

7) За неделю было куплено две единицы товара. Используя гамма-экспоненциальное распределение и параметры $a = 1$ и $b = 1$ неинформативного априорного распределения, определить минимальное время, за которое с вероятностью не менее 0.95 будет продана одна единица товара.