УДК 681.513.7

ПРИНЦИП СОГЛАСОВАННОСТИ И БЕЙЕСОВСКОЕ ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ

С. И. Гуров

Московский Государственный университет им. М.В. Ломоносова, ф-т ВМиК
Россия, г. Москва,
e-mail: *sgur@cs.msu.su, gurov@ccas.ru.*

Работа выполнена при финансовой поддержке РФФИ (код проекта 01-01-00885)

Abstract

In the context of Bayes approach to statistical tasks the new method of concretization of a prior distribution based on principle named «concordant principle» is putting forward. The application of suggested method to the task of interval estimation of parameters of binomial, Poisson and normal distributions is shown.

Введение

Основная проблема применения бейесовского подхода для получения выводов на основе статистического эксперимента состоит в необходимости конкретизации априорного распределения.

При наличии у исследователя результатов аналогичных экспериментов, проводимых ранее, возможно использование того или иного метода восстановления априорного распределения в рамках эмпирического бейесовского подхода. При отсутствии указанных данных такой возможности нет. В этих случаях обычно прибегают к постулату Бейеса[1], который устанавливает, что если ничего не известно в параметре θ, и он изменяется на конечном интервале, то в качестве априорного распределения принимают равномерное. В 1948г. Г. Джефферс расширил постулат Бейеса, предложив т.н. неинформативное априорное распределение для неизвестного параметра θ, лишь в случае полубесконечной области изменения θ [16], [12].

Известны и другие способы конкраприорного распределения, основанные на использовании принципв наименьшей информативности. понятий субъективной вероятности, функции полезности, доверия и т.д. [4]. Однако применение их весьма ограничено. В [3] рассматриваются методы параметрического оценивания плотности вероятностей; ими можно воспользоваться для восстановленияаприорного распределения. Сделанными замечениями и ограничивается *анализ последних достижений*

[1]его называют также принципом равновероятности или законом недостаточного основания Лапласа

и публикаций, посвящённых рассматриваемой проблеме.

Целью работы является формулировка нового принципа конкретизации априорного распределения неизвестной величины θ, определяющей распределение $P_\theta(\xi)$ случайной величины ξ, значения которой наблюдаются в статистическом эксперименте и обладающего свойством естественности, согласованности с данной задачей. предлагаемый принцип состоит в указании (1) некоторого естественного для данной статистической модели класса $\mathfrak{G}_c$ возможных априорных распределений, параметризованного конечномерным параметром γ и (2) метода нахождения значения γ по результатам статистического эксперимента.

Данный принцип мы будем называть *принципом согласованности*, поскольку он полагает, с одной стороны, однотипность априорных плотностей и функции правдоподобия переметра θ, и с другой — равенство его частотной (классической) и бейесовской точечных оценок. Требование согласованности двух указанных видов представляется вполне естественным. Заметим, что приравнивание частотной и бейесовской оценок применяется в математической статистике при рассмотрении минимаксных точечных оценок и нахождения соответствующего т.н. наименее благоприятного априорного распределения. При определении класса априорных распределений $\mathfrak{G}_c$ используются понятия фидуциальной теории Р. Фишера. таким образом. предлагаемый принцип согласованности объединяет классический частотный, бейесовский и идуциальный подходы математической статистики.

Полученное на основе указанного принципа априорное распределение определяет апостериорное, которое и используется для получения статистических решений, в нашей работе — для доверительного оценивания неизвестного параметра θ.

Данный подход в начале был предложен без обоснования в [5], [6] и развит затем в [7] и [8].

1. Точечное и интервальное оценивание параметров распределений случайных величин

Данный раздел носит вспомагательно-справочный характер. Здесь также будет приведено важное для дальнейшего определение сопряжённого относительно некоторой статистической модели семейства априорных распределений. все неопределённые понятия и ссылки могут быть найдены в [1], [2], [12], [13] и [15].

Рассмотрим модель простого выбора, связанную с однократным наблюдением над случайным элементом ξ, принимающим значения в $\mathcal{X}$ и связанную с ней параметризованную статистическую структуру $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathcal{P})$. Здесь $\mathcal{B}(\mathcal{X})$ — σ-алгебра событий на $\mathcal{X}$, а $\mathcal{P} = \{P_\theta(\xi), \theta \in \Theta\}$ – параметрическое семейство распределений на $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, где Θ — конечномерное пространство изменения параметра θ.

При проведении n элементарных (наблюдений) образуется структура $(\mathcal{X}^n, \mathcal{B}(\mathcal{X})^n, \mathcal{P}^n)$. $\mathcal{X}^n$, как известно, называют выборочным пространством. Это множество всевозможных значений статистических данных, полученных в ходе выполнения n независимо проведённых единичных наблюдений над случайной величины ξ. Статистическую структуру мы будем называть по типу распределения ξ (биномиальная, пуассоновская, нормальная и т.д.).

Далее мы полагаем, что $\Theta \subseteq \mathbb{R}$ и $\mathcal{X} \subseteq \mathbb{R}$, как в подавляющем большинстве практических задач. Будем также считать, что и $P_\theta(\xi)$, и всеостальные рассматриваемые ниже распределения имеют плотности (дифференцируемы по соответствующей вероятностмере). Распределения мы будем обозначать прописными, а соответствующие плотности — строчными буквами: так $p_\theta(\xi)$ есть плотность распределения $P_\theta(\xi)$.

В рамках частотного подхода найдём точесную оценку неизвестного истинного значения θ^*. Известно, что при не очень жёстких ограничениях многими «хорошими» статистическими свойствами обладает оценка $\hat{\theta}_L$ максимального правдоподобия (МП–оценка), определяемая как

$$\hat{\theta}_L = \arg\max_{\theta \in \bar{\Theta}} L(\theta, x) = \hat{\theta}_L(x),$$

где $\bar{\Theta}$ — замыкание множества Θ, а $L(\theta, x) \propto p_\theta(x)$ — функция правдоподобия параметра θ для данных $x = (x_1, \ldots, x_n) \in \mathcal{X}^n \subseteq \mathbb{R}^n$ и

$$L(\theta, x) = \prod_{i=1}^{n} p_\theta(x_i).$$

Будем далее предполагать существование доверительной области для θ в виде интервала $J(\eta) = (\theta^-(x, \eta), \theta^+(x, \eta)) \in \Theta$, где η — коэффициент доверия. Указание на величину η и функциональную зависимость границ интервала будем иногда опускать и использовать обозначение $|J(\eta)| = (\theta^+ - \theta^-)$ для длины интервала $J(\eta)$.

Рассмотрим кратко методику потроения неймановских (наиболее селективных), фидуциальных и бейесовких доверительных интервалов. Построение кратчайших доверительных интервалов мы рассматривать не будем, поскольку в большинстве практически важных случаев они совпадают с неймановскими.

В случае, когда $P_\theta(\xi)$ — непрерывное распределение, построение неймановских доверительных интервалов основано на следующем утверждении.

Теорема 1. *Предположим, что $(x_1, \ldots, x_n)$ — выборка из распределения с непрерывной функцией распределения $F(x|\theta)$. Пусть функция $q(x_1, \ldots, x_n|\ \theta)$*

1. *определена в каждой точке θ интервала (θ_1, θ_2), содержащего θ^*, и в каждой точке выборочного пространства $\mathcal{X}^n = \mathbb{R}^n$, за исключением, возможно, лишь множество вероятностной меры 0,*
2. *непрерывна и монотонно убывает или возрастает по θ и*
3. *имеет функцию распределения, не зависящую от θ.*

Если (q_1, q_2) — интервал, для которого $P\{q_1 < q < q_2 | \ \theta\} = \eta$, то при истинном значении параметра θ, равном θ^*, решения θ^- и θ^+(где $\theta^- < \theta^+$) уравнений

$$q(x_1, \ldots, x_n | \ \theta) = \begin{cases} q_1 \\ q_2 \end{cases}$$

существуют, и (θ^-, θ^+) — 100η–процентный доверительный интервал для θ^*.

Здесь ясно, что если функция $q(x_1, \ldots, x_n | \ \theta)$ строго монотонна, то интервал (q_1, q_2) определяется однозначно.

Пусть функция распределения некоторой точесной оценки $\hat{\theta}$ есть $F(t| \ \theta)$. тогда в качестве $q(x_1, \ldots, x_n | \ \theta)$ можно взять $F(\hat{\theta}| \ \theta)$. Эту функцию называют неймановским доверительным распределением статистики $\hat{\theta}$. В качестве $\hat{\theta}$ можно взять МП–оценку $\hat{\theta}_L$. Для её функции распределения $F(\hat{\theta}_L| \ \theta)$ условия регулярности 1) – 3) в вышеприведённом утверждении выполняются почти во всех интересных для практики случаях. Далее, обычно рассматривают центральные интервалы[2], приняв вместь q_1 и q_2 величины P и $1-P$, где $0.5 \leqslant P < 1$. Тогда η будет равняться $2P-1$ и границы θ^-, θ^+ неймановских интервалов $J_N(2P-1) = (\theta^-, \theta^+)$ определяется как решения уравнений

$$F(\hat{\theta}_L| \ \theta) = \begin{cases} P, \\ 1-P. \end{cases} \tag{1}$$

Если $P_\theta(\xi)$ — дискретное распределение, то при определении доверительных интервалов возникают неопределённости, связанные с нестрогими неравенствами, появляющимися в (1) вместо равенств. Для снятия этих неопределённостей эффективен приём рандомизации, заключающийся в добавлении к дискретной случайной величине равномерно распределённой непрерывной так, чтобы полученная сумма имела бы непрерывное распределение. В результате рандомизации доверительные интервалы даже укорачиваются[3], но появляются два распределения, по одному из которых определяют верхнюю, а по другому нижнюю границу интервала $J(\eta)$.

Далее мы рассмотрим понятия фидуциальной теории Р. Фишера, и фидуциальные доверительные интервалы будем также называть фишеровскими. Если в статистической структуре $(\mathcal{X}, \mathcal{B}(\mathcal{X}, \mathcal{P}))$ и параметрическом семействе $\{P_\theta(\xi), \theta \in \Theta\}$, где $\mathcal{X} \subseteq \mathcal{R}$ и $\Theta \subseteq \mathcal{R}$ функция распределения $F(t| \ \theta)$ статистики s убывает с ростом θ так, что

[2]Нецентральные интервалы используют в случае, когда есть основания обеспечить разные вероятности нахождения доверительного интервала слева или права от истинного значения определяемой величины.

[3]Неопределённость, связанная с введением дополнительной случайной величины оказывается меньше устранённой неопределённости, связанной с неравенствами

$F^*(t|\ \theta) = 1 - F(t|\ \theta)$, рассматриваемая как функция от θ обладает свойствами функции распределения, F^*, называют *фидуциальным распределением*[4]. Когда в качестве s используется достаточная статистика, указанные условия обычно выпоняются.

Таким образом, с точностью до мультипликативной константы фидуциальное распределение получается, когда функция правдоподобия $L(\theta, x)$ рассматривается как функция от θ (и здесь θ выступает как случайная величина). Поэтому фидуциальные доверительные интервалы называют также интервальными оценками максимального правдоподобия.

В случае, когда $P_\theta(\xi)$ — непрерывное распределение, границы θ^-, θ^+ фидуциального интервала определяют решения системы:

$$\begin{cases} F(\theta^+|\ x, \gamma) - F(\theta^-|\ x, \gamma) = \eta, \\ f(\theta^+|\ x, \gamma) = f(\theta^-|\ x, \gamma), \end{cases}$$

где $h(\theta|\ x, \gamma)$ — непрерывная унимодулярная плотность апостериорного распределения. Указанные выше условия обычно записывают компактнее: границы θ^-, θ^+ интервала определяются как решения уравнений (1), где в левой части стоит фидуциальная функция распределения $F(\theta|\ x, \gamma)$.

Если $P_\theta(\xi)$ — дискретное распределение, то возникают те же неопределённости, что и в частотном случае. Здесь также появляются два распределения для определения верхнеи нижней границ каждого интервала.

Перейдём теперь к рассмотрению бейесовского подхода к построению интервальных оценок. Предположим, что задан класс $\{G(\theta|\ \gamma), \gamma \in \Gamma\}$ априорных распределений параметра θ с функциями распределений $G(\theta|\ \gamma)$, где γ — конечномерный параметр из некоторой области Γ. Тогда апостериорное распределение $H(\theta|\ x, \gamma)$ параметра θ находят из соотношения

$$dH(\theta|\ x, \gamma) \propto L(\theta, x) \cdot G(\theta|\gamma). \tag{2}$$

Отсюда, в частности, видно, что априорная мера может и не быть вероятностной.

Совокупность всевозможных апостериорных распределений данной задачи образует множество $\mathfrak{H} = \{H(\theta|\ x, \gamma),\ x \in \mathcal{X}^n,\ \gamma \in \Gamma\}$. Мы считаем функции из $\mathfrak{H}$ строго вогнутыми по θ, что имеет место почти для всех практически важных случаев.

Удобно выбирать априорное распределение из специально подобранного для данной статистической структуры семейства распределений $\mathfrak{G} = \{G(\theta|\ \gamma), \gamma \in \Gamma\}$, а именно из т.н. сопряжённых семейств распределений. Семейство распределений $\mathfrak{G}$ называется *сопряжённым семейством априорных распределений* относительно статистической модели $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathcal{P})$ получения статистических данных $x \in \mathcal{X}$, если при

[4]«Фидуциальное распределение не является распределением вероятности в смысле частотной теории. Это новое понятие,выражающее интенсивность нашей веры в различные возможные значения параметра» [12].

выборе в качестве априорного распределения $G = G(\theta|\ \gamma_0) \in \mathfrak{G}$ апостериорное распределение H также принадлежит $\mathfrak{G}$, т.е. $H = G(\theta|\ \gamma(x, \gamma_0))$. Введённое таким образом понятие сопряжённости совпадает с согласованностью в смысле [13] семейств распределений ξ и θ, и класс $\mathfrak{G}$ будет содержать фидуциальные распределения θ. Сопряжёнными будут, например, семейства бета–распределений относительно биномиальной и отрицательно–биномиальной статистической моделей, гамма–распределений относительно пуассоновской и экспоненциальной моделей, семейства односторонних и двусторонних распределений Паретто относительно равномерных на $(0, \theta)$ и (θ_1, θ_2) моделей соответственно и т.д.

Бейесовской точечной оценкой $\hat{\theta}_B$ величины θ^* при квадратичной функции потерь будет математическое ожидание $\mu_H = \mu_H(x, \gamma)$ апостериорного распределения. Мы ограничимся рассмотрением именно таких, используемых в подавляющем большинстве практических случаев функций потерь считать, что

$$\hat{\theta}_B = \mu_H = \hat{\theta}_B(x, \gamma).$$

Бейесовские доверительные интервалы получают также, как в доверительно–интервальной и фидуциальной теориях. т.е. используя системы вида 1, где в левой части стоит функция апостериорного распределения $H(\theta|\ x, \gamma)$. Также остаётся справедливым всё сказанное относительно случая. когда $P_\theta(\xi)$ — дискретное распределение.

Если в качестве априорного принять равномерное (пропорциональное $d\theta$) распределение, апостериорное распределение будет совпадать с фидуциальным, а бейесовские интервалы (их в эслучае будем обозначать $J(x, \eta, \gamma)$ или $J_B(\eta)$) — с неймановскими и фишеровскими.

2. Принцип согласованности

Принцип согласованности включает в себя два момента. Первый (ПС1) связан с определением класса $\mathfrak{G}_c = \{G(\theta|\ \gamma), \gamma \in \Gamma_c\}$, из которого выбираются априорные распределения.

Для данной статистической модели $\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathcal{P}$ рассмотрим сопряжённое семейство априорных распределений $\mathfrak{G} = \{G(\theta|\ \gamma), \gamma \in \Gamma\}$, у которых $\Gamma \subseteq \mathbb{R}^k$.

Отметим, что обычно $\mathfrak{G}$ включаетв себя класс некоторых специальных непрерывных распределений $\{G(\theta|\ \gamma) \in \Gamma_0\}$ (нормальное, бета-,гамма- и т.д.), встречающихся в математической статистике, и замыкание Γ_0 есть Γ. Например, равномерное распределение данного параметра, очевидно, входит в семейство сопряжённых распределений дял любой статистической модели. семейство всех гамма–распределений $\{G_\theta(a, b)\}$ (с параметрами $a > 0,\ b > 0$) является сопряжённым для статистической модели наблюдения случайной величины, распределённой по закону Пуассона. Равномерное распределение величины θ на $(0, \infty)$ не является элементом $\{G_\theta(a, b)\}$, однако получается асимптотически при $a = 1, b \to 0$.

С другой стороны, класс $\mathfrak{G}$, очевидно, содержит все фидуциальные распрделения $\{F(\theta|\ \gamma), \gamma \in \Gamma_f \subseteq \Gamma\}$ данной статистической модели и область Γ_f соответствует всем таким распределениям.

Среди компонент вектора $\gamma = (\gamma_1, \ldots, \gamma_k)$ из Γ_f имеются принимающие дискретные значения. Всегда, нарпимер. имеется компонента, соответствующая числу испытаний n. Можем считать, что это γ_1. Формально полагая $n = 0$ в $\mathfrak{G}$ обычно получают равномерное распределение. Поэтому считаем, что $\gamma_1 \in \{0, 1, \ldots\}$.

Положим теперь все дискретные координаты вектора $\gamma \in\ \Gamma_f$ непрерывными в соответствующих областях ограничений. Нпример, в приведённых выше обозначениях полагаем $\gamma_1 \in \mathbb{R}_{\geqslant 0}$. В результате такого расширения получаем однозначно оперделённую область Γ_c такую, что $\Gamma_f \subset \Gamma_c \subseteq \Gamma$.

Определяемые областью Γ_c распределения образуют в $\mathfrak{G}$ подкласс, который мы назовём *согласованным семейством априорных распределений* относительно данной статистической модели.

ПС1:: *Класс* $\mathfrak{G}_c$ *является согласованным семейством априорных распределений* $\{G(\theta|\ \gamma), \gamma \in \Gamma_c\}$ *относительно статистической модели* $\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathcal{P}$.

Полученное по формуле Бейеса апостериорное распределение H будем записывать в виде $H(\theta|\ x, \gamma_1)$, где $\gamma_1 = \gamma(x, \gamma_0)$. Понятно, что если $\mathfrak{G}_c$ — класс всевозможных апостериорных распределений, получаемых в рамках данной статистической модели, то $\mathfrak{H}_c \subseteq \mathfrak{G}_c$.

Ясно также, что класс $\mathfrak{G}_c$ содержит все распределения, плотности которых пропорциональны функциям правдоподобия параметра θ при всевозможных вариантах исходов произвольного числа элементарных экспериментов:

$$dG(\theta|\ \gamma) \propto L(\theta, x),\ x \in \mathcal{X}^n,\ n = 1, 2, \ldots.$$

Представляется, что такой выбор класса $\mathfrak{G}_c$ адекватно отражает специфику статистической модели.

Второй момент (ПС2) принципа согласованности указывает способ нахождения параметра γ, конкретизирующего априорное распределение для данной задачи в условиях наблюденных результатов статистического эксперимента. Принцип согласованности определяет, что параметр γ находится как решение (при полученных значениях x и заданном коэффициенте доверия η) оптимизационной задачи

ПС2:: $\qquad |J(x, \eta, \gamma)| \to \max, \quad \hat{\theta}_L(x) = \hat{\theta}_B(x, \gamma), \quad \gamma \in \Gamma_c.$

В случае отсутствия решения, считаем, что в данных условиях принцип согласованности неприменим. Область $\mathcal{X}_c \subseteq \mathcal{X}^n$ в которой решение задачи (ПС2) существует, назовём *областью применимости* принципа согласованности. При сделанных предположениях относительно свойств апостериорных распределений решение, очевидно, будет единственным.

Пусть $x \in \mathcal{X}_c$ и найдено решение γ^* задачи (ПС2). Оно определит плотности априорного $g(\theta|\gamma^*)$, и апостериорного $h(\theta|x,\gamma^*)$ распределений величины θ, обеспечивающие наибольшую величину доверительной области при условия совпадения точечных оценок, полученных в рамках частотного и бейесовского подходов. Это позволяет утверждать, что в сделанных предположениях достоверность (значение коэффициента доверия) справедливости включения $\theta^* \in J(x,\eta,\gamma^*)$ будет не менее η при любых γ и x. Полученные доверительные пределы θ^-, θ^+ и сам интервал $J_c = J(x,\eta,\gamma^*)$ будем называть *согласованными.* Ясно, что величина согласованного доверительного интервала будет не больше соответствующего классического бейесовского.

Против применения предлагаемого принципа могут быть выдвинуто возражение, связанное с тем, что часто оцениваемые параметры являются неизвестными, но фиксированными (неслучайными), и поэтому при их нахождении применимы лишь классические методы. Однако, если придерживаться т.н. «субъектного» подхода в статистических задачах оценивания, (считать, что априорное распределение является мерой нашего незнания), то применение бейесовского подхода является оправданным [11].

Отметим, что в некоторых случаях согласованные доверительные пределы совпадают с полученными традиционно. С другой стороны, в некоторых случаях применение принципа согласованности приводит к сокращению доверительных интервалов. Ниже рассмотрены соответствующие примеры.

3. Согласованное интервальное оценивнаие параметра биномиального распределения

Рассмотрим применение предложенного подхода на примере задачи интервального оценивания неизвестой вероятности числа успехов m в n испытаниях Бернулли:

$$Bi_m(n,p) = \begin{pmatrix} n \\ m \end{pmatrix} p^m (1-p)^{n-m}.$$

Пусть величина ξ, принимающая значения 0 или 1, имеет распределение

$$P\{\xi = t|p^*\} = Bi_t(1,p^*) = (p^*)^t (1-p^*)^{1-t};\ p^* \in (0,1),\ t \in \{0,1\},$$

где p^* — неизвестная, но фиксированная величина, для которой требуется построить интервальную оценку. Как обычно, мы интерпретируем 1 и 0 соответственно как появление или отсутствие некоторого случайного события X в данном эксперименте. Пусть в результате проведения n таких элементарных экспериментов получена выборка $x = (x_1, \ldots, x_n)$. В рассматриваемом случае
$\mathcal{X} = \{0,1\}$, $\mathcal{BX}^n$ — булева алгебра n-мерных двоичных наборов,

$$\theta = p^* = p, \quad \Theta = (0,1) \subset \mathbb{R}.$$

Обозначим через $\bar{x}$ выборочное среднее $(1/n)\sum\limits_{i=1}^{n} x_i$. Функция распределения $\bar{x}$ есть

$$P\{\bar{x} \leqslant m/n|\; n, p\} = \sum_{i=0}^{m} \binom{n}{i} p^i (1-p)^{n-i} =$$
$$= I_{1-p}(n-m, m+1) = 1 - I_p(m+1, n-m);$$
$$0 \leqslant m \leqslant n, \quad n = 1, 2, \ldots,$$

где $I_p(a,b)$ - неполная B-функция [5], которая определена для положительных параметров a и b. Заметим, что в нашем случае функция распределения $\overline{x}$ является непрерывной и строго монотонно убывающей по $p = \theta$.

Пусть при проведении n элементарных испытаний событие X наблюдалось m раз. Тогда функция правдоподобия $L(p, x)$ величины ξ есть

$$L(p, x) = p^m (1-p)^{n-m},$$

а МП-оценкой величины p будет являться наблюденное значение величины $\overline{x}$

$$\hat{p}_L = \overline{x} = \frac{m}{n}.$$

Известно, что эта оценка является несмещенной, эффективной и состоятельной, а несмещенная функция оценки $\hat{D}_{\{\hat{p}_L\}}$ её дисперсии есть:

$$\hat{D}_{\{\hat{p}_L\}} = \frac{m(n-m)}{n^2(n-1)}. \tag{3}$$

Поскольку распределение ξ дискретно, в рамках классического частотного подхода доверительные интервалы $J = (p^-, p^+)$ с коэффициентом доверия $\eta = 2P - 1$ накрывающего значение p^* определяются [2] как решения следующих уравнений:

$$\begin{cases} I_{p^+}(m, n-m+1) = 1 - P, \\ I_{p^-}(m+1, n-m) = P. \end{cases} \tag{4}$$

Сопряженным семейством априорных распределений будет множество B-распределений $\Im = \{I_p(a,b) | \gamma \in \Gamma\}$, где $\gamma = (a,b)$ и $\Gamma = R_{>0} \times R_{>0}$ с математическими ожиданиями и дисперсиями

$$\mu_I = \frac{a}{a+b}, \sigma_I^2 = \frac{ab}{(a+b)^2(a+b+1)}.$$

Случай $a = b = 1$ соответствует здесь равномерному распределению на (0,1).

В соответствии с (ПС1) класс $\Im_c$ для данной задачи есть $\{I_p(a,b) | \gamma \in \Gamma_c\}$, где $\gamma = (a,b)$ и $\Gamma_c = R_{\geq 1} \times R_{\geq 1}$. Таким образом, плотность некоторого данного априорного

[5] бета–функция

распределения p представляет собой B-распределение

$$g(p|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} = Be_p(a,b); \tag{5}$$

$$p \in (0,1), \gamma = (a,b) \in \Gamma_c = R_{\geq 1} \times R_{\geq 1}.$$

Плотность вероятности апостериорного распределения будет равняться

$$h(p|x,a,b) = Be_p(m+a, n-m+b), \tag{6}$$

и бейесовская точечная оценка $\hat{p_B}$ параметра p есть

$$\hat{p}_B = \mu_H = \frac{m+a}{n+a+b}.$$

По плотности априорного распределения (6) найдем согласованный доверительный интервал $J_c(\eta) = (p^-, p^+)$ с коэффициентом доверия $\eta = (P+1)/2(0.5 \leq P < 1)$: его границы суть решения уравнений

$$\begin{cases} I_{p^+}(m+a-1, n-m+b) = 1-P, \\ I_{p^-}(m+a, n-m+b-1) = P. \end{cases} \tag{7}$$

Таким образом, задача оптимизации (ПС2) записывается в виде

$$(p^+ - p^-) \to max, \frac{m}{n} = \frac{m+a}{n+a+b} = \widehat{p},\ 1 \leq a, 1 \leq b, \tag{8}$$

где p^+ и p^- - решения уравнений (7).

Первое ограничение (равенство частотной и бейесовской оценок) определяет область применимости принципа согласованности: χ_c в данной задаче есть $1 \leq m \leq n-1$. Соответственно, мы исключаем из рассмотрения 0-события (заметим, что в практике доверительного оценивания эти случае принято рассматривать отдельно [10]. В указанных границах имеем $a = bm/(n-m)$.

Нетрудно видеть, что условие $(p^+ - p^-) \to max$ равносильно $t \to min$, где t - такое минимальное натуральное, что t-я производная от $h(p|x,a,b)$ в точке $p=0$ или $p=1$ не равна 0. Отсюда получаем, что решение оптимизационной задачи (8) есть

$$\begin{cases} a = 1, b = \frac{n-m}{m}, \text{если } 1 \leq m \leq \frac{n}{2}, \\ a = \frac{m}{n-m}, b = 1, \text{если } \frac{n}{2} < m \leq n-1. \end{cases} \tag{9}$$

Полученное решение полностью определяет плотности определяет плотности априорного (5) апостериорного (6) распределений. Точечная оценка $\hat{p}$ истинного значения p^* будет выражаться через определенные параметры a и b как $\hat{p} = 1/(b+1)$ или $\hat{p} = a/(a+1)$ соответственно в первом или во втором случае (9) и для дисперсии этой оценки верна формула (3).

Границы доверительного интервала $J_c(\eta)$ определяются уравнениями (7). Ясно, что $|J_c(\eta)| \leq |J_B(\eta)|$, причем равенство достигается лишь когда $m = n/2$. Например, при $n = 10$, $m = 1$, $\eta = 0.9(P = 0.95)$ по таблицам [2] имеем:

$$J_B = (0.005, 0.394); \; J_c = (0.003, 0.238),$$

т.е. длина доверительного интервала сократилась почти на 40%. Очевидно, что симметричность плотностей апостериорных распределений при $\hat{p} = 1/2 - r$ и $\hat{p} = 1/2 + r, \; 0 < r < 1/2$ повлечет равенство длин доверительных интервалов для этих случаев [6].

Интересно сравнить полученное значение b с максимально правдоподобным в предположении малого $p = m/n (m \neq 0)$, т.е. когда можно считать, что $a = 1$. Найдем МП-оценку параметра b распределения $Be_p(1, b)$. Функция правдоподобия в этом случае будет иметь вид

$$L(b, x) = b \left(1 - \frac{m}{n}\right)^{b-1}.$$

Нетрудно найти, что её максимум достигается при

$$\hat{b} = \frac{1}{\ln \frac{n}{n-m}}.$$

Поскольку $b = (n - m)/m$, имеем

$$\frac{1}{\hat{b}} = \ln(1 + \frac{1}{b}) = \frac{1}{b} - \frac{1}{2b^2} + \frac{1}{3b^3} - \cdots ,$$

далее, обращая ряд (см., например, [9]) получаем, что

$$\hat{b} = b + \frac{1}{2} - \frac{1}{12b} - \frac{5}{24b^2} - o\left(\frac{1}{b^3}\right),$$

т.е. $\hat{b} \approx b + 0.5$ хорошей с точностью: относительная ошибка не превосходит полупроцента при $b \geq 4$ (или$\hat{p} \leq 1/5$).

Таким образом, согласованная оценка для b в условиях сделанных предположений относительно мало отличается от её МП-оценки. Это отличие будет ещё меньше для длин соответствующих апостериорных интервальных оценок параметра p.

Понятно, что аналогичные утверждения справедливы и для a при $b = 1$ и $\hat{p} \approx 1$.

---

[6] Следует сказать, что для представления точечной оценки вероятностей редких событий в [14] было предложено в качестве плотности априорного распределения использовать именно $B_{C_p}(1, b)$ с достаточно большим b, однако не было приведено ни обоснования данному выбору, ни каких – либо указаний на возможный способ определения параметра b. Эта идея Э. Лемана и послужила толчком к появлению данной работы.

4. Согласованное интервальное оценивание параметра распределения Пуассона

Пусть случайная величина ξ распределена по закону Пуассона, т.е.

$$P\{\xi = t|\lambda^*\} = P_{o_t}(\lambda^*) = \frac{(\lambda^*)^t e^{-\lambda^*}}{t!}, t = 0, 1, ...,$$

где λ^* - неизвестная, но фиксированная величина, для которой требуется построить интервальную оценку. Заметим, что функция распределения $P_{o_t}(\lambda)$ является непрерывной и строго монотонно убывающей по $\lambda = 0$.

Функция распределения ξ при $\lambda^* = \lambda$ есть

$$P\{\xi \leq t|\lambda\} = \sum_{t=0}^{t} \frac{\lambda^i e^{-\lambda}}{i!} = X^2(2\lambda, 2t+2), \ t = 0, 1, ...,$$

где $X^2(x, n)$ интервал вероятностей χ^2-распределения с n степенями свободы.

Таким образом [2],

$$P_{o_t}(\lambda) = X^2(2\lambda, 2t+2) - X^2(2\lambda, 2t).$$

Пусть в результате проведения n элементарных экспериментов получена выборка $x = (x_1, \ldots, x_n)$ значений ξ. Обозначим $m = \sum\limits_{i=1}^{n} x_i$. Величина m является достаточной статистикой с функцией распределения

$$P\{m \leq t|n, \lambda\} = \sum_{i=0}^{t} \frac{(n\lambda)^i}{i!} e^{-n\lambda}, \ t = 0, 1, \ldots.$$

Обозначим $d = max\{x_1, \ldots, x_n\}$ и пусть $r_0, \ldots, r_d$- (абсолютные) частоты появления чисел $0, 1, \ldots, d$ в выборке x. Тогда функция правдоподобия $L(\lambda, x)$ величины ξ запишется как

$$L(\lambda, x) = \prod_{i=0}^{d} \left(\frac{\lambda^i}{i!} e^{-\lambda} \right)^{r_i} \propto \lambda^m \exp(-\lambda n) = \lambda^{(m+1)-1} \exp(-\lambda n). \tag{10}$$

Легко показать, что максимум $L(\lambda, x)$ по λ достигается на значении $\bar{x} = m/n$, т.е. МП-оценка для λ есть

$$\hat{\lambda}_L = \bar{x} = \frac{m}{n}.$$

Данная оценка является несмещенной и состоятельной. Кроме того, величина $\sqrt{n/\lambda}(\bar{x} - \lambda)$ асимптотически имеет нормальное стандартное распределение. [7]

[7] поэтому при больших n приблеженные границы доверительного интервала являются корнями некоторого квадратного уравнения

В силу дискретности распределения ξ,в рамках частотного подхода неймановские доверительные интервалы $J = (\lambda^-, \lambda^+)$ с коэффициентом доверия $\eta = 2P - 1$ накрывающие значение λ^* определяются [2] как решения следующих уравнений:

$$\begin{cases} X^2(2n\lambda^-, 2m) = P, \\ X^2(2n\lambda^+, 2m+2) = 1 - P. \end{cases} \tag{11}$$

Перейдем к построению согласованного доверительного интервала для λ^*. Из (10) следует, что $\{g(\lambda|\gamma), \gamma \in \Gamma_c\}$ есть подкласс Г-распределений $G_\lambda(a,b)$ с плотностями

$$g(\lambda|a,b) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}\exp(-\lambda b) = g_\lambda(a,b),\ \lambda > 0,\ \gamma = (a,b) \in \Gamma_c = R_{\geq 1} \times R_{\geq 0}.$$

В общем случае Г-распределение определено при произвольных неотрицательных параметрах формы a и масштаба b, его математическое ожидание μ_G и дисперсия σ_G^2 суть

$$\mu_G = \frac{a}{b}, \quad \sigma_G^2 = \frac{a}{b^2}.$$

Ранее было отмечено, что равомерное распределение не есть какое-либо Г-распределение, но получается как предельное при $a = 1$ и $b \to 0$.

В нашей задаче плотность вероятности апостериорного распределения будет

$$h(\lambda|x,a,b) = g_\lambda(m+a, n+b), \tag{12}$$

и бейесовская точечная оценка $\hat\lambda_B$ параметра λ есть

$$\hat\lambda_B = \mu_H = \frac{m+a}{n+b}.$$

Из (10), (11) (12) следует, что границы согласованного доверительного интервала $J_c(\eta) = (\lambda^-, \lambda^+)$ с коэффициентом доверия $\eta = (P+1)/2 (0.5 \leq P < 1)$ находятся как решение уравнений

$$\begin{cases} X^2(2(n+b)\lambda^-, 2(m+a-1)) = P, \\ X^2(2(n+b)\lambda^+, 2(m+a)) = 1 - P, \end{cases} \tag{13}$$

Таким образом, задача оптимизации (ПС2) записывается в виде

$$(\lambda^+ - \lambda^-) \to \max, \frac{m}{n} = \frac{m+a}{n+b} = \hat p, 1 \leq a, 0 \leq b, \tag{14}$$

где λ^+иλ^- -решения уравнений (13)).

Равенство частотной и бейесовской оценок сразу дает $a = bm/n$. Условие $a \geq 1$ влечет $m \neq 0$, что определяет область применимости χ_c метода согласованного оценивания. Далее, нетрудно видеть, что условие $(\lambda^+ - \lambda^-) \to \max$ равносильно максимизации дисперсии апостериорного распределения. Отсюда легко получаем, что решение оптимизационной задачи (14) есть

$$a = 1,\ b = \frac{n}{m},$$

и границы согласованного доверительного интервала с коэффициентом доверия $\eta = 2P - 1$ будут определяться из уравнений

$$\begin{cases} X^2(2(n+n/m)\lambda^-, 2m) = P, \\ X^2(2(n+n/m)\lambda^+, 2m+2) = 1-P. \end{cases} \tag{15}$$

При этом очевидно, что длины согласованных интервалов будут на $100\%/(m+1)$ меньше соответственных бейесовских.

Заметим, что в данном случае интервалы полученные с использованием постулата Джеффрейса (в полубесконечной области изменения θ априорное распределение следует принять пропорциональным $d\theta/\theta$) совпадают с классическим [12].

Пусть получено $m \neq 0$. Сравним полученное нами b с максимально правдоподобным в предположении $a = 1$. Найдем МП-оценку параметра b распределения $G_\lambda(1, b)$ при полученном значении $\lambda = m/n$.Функция правдоподобия в этом случае будет иметь вид

$$L(b, x) = b \exp\left(-\frac{m}{n}\right) b.$$

Нетрудно видеть, что её максимум достигается при

$$\hat{b} = \frac{n}{m} = b,$$

т.е. согласованная оценка для b совпадает с её МП- оценкой в условиях сделанного предположения.

5. Интервальное оценивание параметров нормального распределения

Рассмотрим задачу доверительного оценивания математичекого ожидания $\mu = \theta$ нормального распределения $N(\mu, \sigma)$ при известной дисперсии σ^2.

Зададимся коэффициентом доверия η и будем считать, что результатом n наблюдений за величиной $\xi \sim N(\mu, \sigma)$ является выборка $x = (x_1, \ldots, x_n)$. Выборочное среднее $\bar{x}$ есть $(1/n) \sum\limits_{i=1}^{n} x_i$ и, как известно,$\hat{\mu_L} = \bar{x}$.

Классические границы доверительного интервала находят, основываясь на том, что распределение $\bar{x}$ есть $N(\mu, \sigma/\sqrt{n})$,и, следовательно, случайная величина $(\bar{x} - \mu)\sqrt{n}/\sigma$ имеет стандартное нормальное распределение $N(0, 1)$, свободное (независимое) от параметра μ. В силу этого, искомый доверительный интервал для μ есть

$$(\bar{x} - \frac{\sigma}{\sqrt{n}} t_\eta, \bar{x} + \frac{\sigma}{\sqrt{n}} t_\eta),$$

где t_η — квантиль уровня $(\eta+1)/2$ стандартного нормального распределения. Таким образом, при данном η длина доверительного интервала полностью определяется значением среднего квадратичного отклонения $\sigma/\sqrt{n}$ нормального распределения выборочного среднего.

Найдем теперь аналогичный согласованный доверительный интервал. Функцией правдоподобия, соответствующей статистическим данным x будет

$$L(\theta, x) = \left(\frac{1}{\sqrt{2\Pi}\sigma}\right)^n \exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2] \propto \exp[-\frac{n}{2\sigma^2}(\theta-\bar{x})^2].$$

В силу этого

$$g(\theta|a,b) \propto \exp\left[-\frac{a}{2\sigma^2}(\theta-b)^2\right],\ \gamma=(a,b)\in\Gamma_c = R_{\geq 0}\times R$$

(случай $a=0$ соответствует здесь равномерному распределению).

По (2) имеем

$$h(\theta|x,a,b) \propto \exp\left[-\frac{n+a}{2\sigma^2}(\theta-m)^2\right], \text{где}\, m=\frac{n\bar{x}+ab}{n+a}$$

Таким образом, $h = N(m, \sigma/\sqrt{n+a})$.

В соответствие с (ПС2) должно выполняться равенство

$$\hat{\mu}_B = m = \frac{n\bar{x}+ab}{n+a} = \bar{x}$$

В то же время понятно, что требование $|J(x,\eta,\gamma)| \to max$ означает максимализацию дисперсии плотности h, или $a \to min$. Отсюда $a=0$, вышепреведенное равенство выполняется, плотность апостериорного распределения h совпадает с плотностью распределения выборочного среднего, а согласованный интервал — с классическим.

С другой стороны,

$$L(a,b|x) = \exp\left[-\frac{a}{2\sigma^2}(\bar{x}-b^2)\right],\ a\geq 0,$$

и $\hat{a}=0$ является МП-оценкой параметра a.

Аналогично показывается, что и в двух других классических задачах интервального оценивания параметров нормального распределения (σ при известном μ и неизвестных σ и μ) согласованные интервалы совпадают с полученными традиционно. Видимо в силу этих почти очевидных результатов перспективность предлагаемого принципа согласованности не была замечена ранее.

Заключение

Рассмотренные примеры демонстрируют применение предложенного принципа согласованноси к задаче интервального оценивания. Принцип согласованности и основанный на нем метод определения интервальных оценок неизвестных величин очерчивают новый эффективный подход к решению статистических задач. Это являетя *основным выводом* данной статьи.

Интересно отметить, что если в трех расмотренных нами случаях биномиальной, пуассоновской и нормальной статистических моделях в качестве априорной невероятностной меры взять, как в подходе Г. Джефрейса, меру, пропорциональную количеству информации $I(\theta)$, то математическое ожидание полученного апостериорного распределения будет совпадать с МП-оценкой параметра θ. Кроме того, было показано, что (возможно, при принятии некоторых дополнительных естественных предложений) значения оценок параметров априорного распределения, полученные по предлагаемому принципу согласованности, совпадают или почти совпадают с максимально правдоподобными.

Перспективным дальнейшим исследованием является сравнительный анализ всех известных на сегодняшний день методов интервального доверительного оценивания.

Автор признателен Ю.И. Журавлеву за неизменную поддержку и В.Е. Бенингу за ценные консультаци.

Список литературы

1. Беляев Ю.К., Носко В.П. Основные понятия и задачи мтематической статстики: Учеб. пособие.— М.:Изд–во МГУ, ЧеРо, 1998. — 192 с.
2. Большев Л.Н., Смирнов Н.В. Таблицыматематической статистики. — М.: Наука, 1983. —416 с.
3. Бостанжиян В.А. Определение плотности вероятности. Необходимый объем выборки. — М.: Наука, 1971. — 160с.
4. Володин И.Н. Бейесовский подход // Вероятность и математическая статистика: Энциклопедия. — М.:Большая Российская энциклопедия, 1999.— С.38.
5. Гуров С.И. Оценка надежности классифицируюих алгоритмов. — М.:Издательский отдел ф–та ВМиК МГУ, 2002. — 44с.
6. Гуров С.И. Как оценить надежность алгоритмов класификаций. II. Интервальные оценки // Таврический Вестник информатики и математики, Вып. 2, 2003.— Симферополь:КНЦ НАН Украина. — С.4–15.
7. Гуров С.И.,Уткина О.Ф. Определение вероятности ошибки распознования с восстановлением его априорного распределения // "Математические методы распознования образов"(ММРО–11). Доклады XI Всероссийской конференции.—М.:ВЦ РАН, 2003(в печати).
8. Гуров С.И. Принцип согласованности для конкретизации априорного распределения и его применение к задаче доверительного оценивания // Spectral and Evolution Problems: Proceeding of the 14th Crimean Autumn Mathematical School — Symposium. Vol.14. — Simferopol: Crimean Scientific Center of Ukrainean National Academy of Scainces.2004(в печати).
9. Двайт Г.Б. Таблицы интегралов и другие математические формулы. — М.: Наука, 1973. — 228с.
10. Закс Л. Статистическое оценивание. — М.: Статистика, 1976. — 560 с.
11. Кендал М., Стюарт А. Теория распределений. — М.: Наука, 1966.
12. Кендал М., Стюарт А. Статистические выводы и связи. — М.: Наука, 1973.
13. Климов Г.П. Теория вероятности и математическая статистика. — М.: Изд. Моск. ун–та, 1983. — 318 с.
14. Леман Э. Теория точечного оценивания. — М.: Наука, 1991. — 448 с.
15. Уилкс С. Математическая статистика. — М.: Наука, 1967. — 632 с.
16. Jeffreys H. (1961) Theory of Probability, Oxford University Press.