Гиббсовские случайные величины и гиббсовско – пуассоновские предельные распределения

О.Ю. Воробьёв
Институт вычислительного моделирования СО РАН
Красноярский государственный университет
а/я 8699, Академгородок, Красноярск, 660036,
email: vorob@scn.ru

Под влиянием некоторых прикладных интерпретаций в статистической теории потребительского выбора [3] возникла идея гиббсовской случайной величины, которая строится из бернуллиевской случайной величины и показательного распределения. Есть надежда приблизить некоторые встречающиеся на практике распределения суммами гиббсовских случайных величин так же, как суммами бернуллиевских случайных величин, можно приблизить многие известные распределения.

Гиббсовские случайные величины

Гиббсовская случайная величина γ с параметрами (b, β, p) определяется на положительной полуоси функцией распределения

$$G(x) = \mathbf{P}(\gamma \le x) = \begin{cases} 1-p, & 0 \le x < b, \\ \\ 1 - p\ \exp\{-\beta(x-b)\}, & \text{иначе.} \end{cases}$$

Ключевое свойство гиббсовской случайной величины — она не принимает значений внутри отрезка $[0, b]$. Очевидно, что таким же свойством обладает и бернуллиевская случайная величина $(b = 1)$. Так что гиббсовскую случайную величину с параметрами $(1, \beta, p)$ можно рассматривать как обобщение бернуллиевской с параметром p — она приближается к последней при $\beta \to \infty$.

Именно на это ключевое свойство гиббсовской случайной величины — отделимость нуля — указали некоторые интерпретации статистической теории потребительского выбора [1], где стоимость случайной

Олег Воробьёв (ред.), Финансово-актуарная математика и смежные вопросы, Труды ФАМ'2002, Красноярск, ИВМ СО РАН

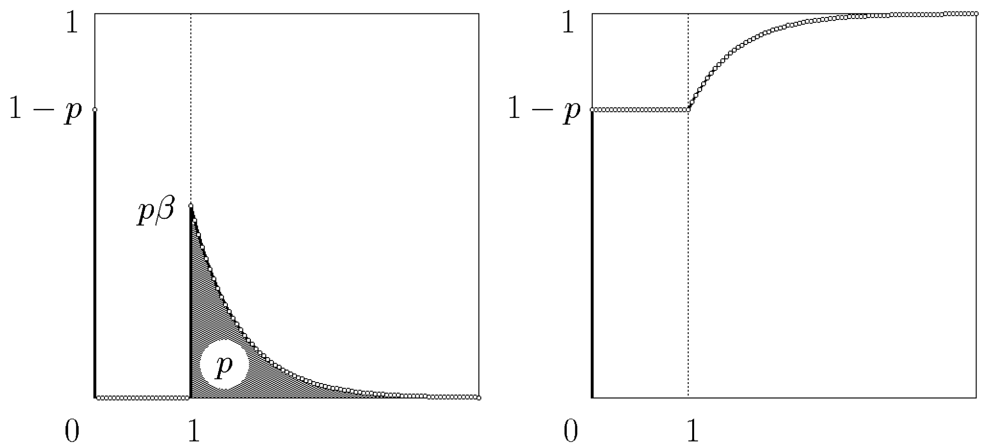

Рис. 1: Графики (слева) дискретно-непрерывной плотности (атом вероятности в нуле и плотность положительных значений) и (справа) функции распределения гиббсовской случайной величины с параметрами $(1, \beta, p)$.

покупки — это дискретно-непрерывная случайная величина, которая имеет распределение, существенно отделенное от нуля (для большинства товаров и услуг), в то время как нулевое значение стоимости покупки (соответствующие отсутствию покупки) имеет положительную вероятность.

Рассмотрим гиббсовские случайные величины

$$\gamma_1, \ldots, \gamma_n,$$

с одним и тем же распределением с параметрами $(1, \beta, p)$. Иначе говоря,

$$\mathbf{P}(\gamma_i > 0) = p, \quad \mathbf{P}(\gamma_i = 0) = 1 - p,$$

а условная плотность вероятности (при условии, что $\gamma > 0$) имеет показательный вид с параметром β:

$$f(x) = \begin{cases} 0, & x < 1 \\ \beta \exp\{-\beta(x-1)\}, & x \geq 1. \end{cases}$$

Вычислим моменты гиббсовской случайной величины. Рассмотрим случайную величину ξ имеющую показательное распределение. Ее плотность вероятности имеет вид:

$$f_\xi(x) = \beta \exp\{-\beta x\}$$

Известно, что моменты показательного распределения с параметром β равны

$$\mathbf{E}\xi^n = n!/\beta^n.$$

В частности, ожидание равно $\mathbf{E}\xi = 1/\beta$, а дисперсия имеет вид $\mathbf{D}\xi = \mathbf{E}\xi^2 - (\mathbf{E}\xi)^2 = 2/\beta^2 - 1/\beta^2 = 1/\beta^2$. Отсюда

$$\mathbf{E}\gamma^n = \mathbf{E}(\gamma^n|\gamma = 0) \cdot \mathbf{P}(\gamma = 0) + \mathbf{E}(\gamma^n|\gamma > 0) \cdot \mathbf{P}(\gamma > 0) =$$

$$= p \cdot \mathbf{E}(1+\xi)^n = p \cdot \sum_{i=0}^{n} C_n^{n-i}\mathbf{E}\xi^{n-i} =$$

$$= p \cdot \sum_{i=0}^{n} C_n^i (n-i)!(1/\beta)^{n-i} = p \cdot \frac{n!}{\beta^n}\sum_{i=0}^{n}\frac{\beta^i}{i!}.$$

$$\mathbf{E}\gamma = p\cdot(1+1/\beta), \quad \mathbf{E}\gamma^2 = p\cdot(1+2/\beta+2/\beta^2),$$

$$\mathbf{D}\gamma = \mathbf{E}\gamma^2 - (\mathbf{E}\gamma)^2 = p(1-p)\cdot(1+2/\beta+2/\beta^2) + p^2/\beta^2.$$

Предельная теорема для сумм гиббсовских случайных величин

Рассмотрим теперь сумму

$$S_\gamma^{(n)} = \gamma_1 + \ldots + \gamma_n$$

из n одинаково распределенных взаимно независимых гиббсовских случайных величин с параметрами $(1, \beta, p)$. Просто вычислить, что

$$\mathbf{E}S_\gamma^{(n)} = \sum_{i=1}^{n}\mathbf{E}\gamma_i = n\cdot p\cdot(1+1/\beta),$$

$$\mathbf{D}S_\gamma^{(n)} = \sum_{i=1}^{n}\mathbf{D}\gamma_i = n\cdot p(1-p)\cdot(1+2/\beta+2/\beta^2) + n\cdot p^2/\beta^2$$

Из центральной предельной теоремы следует, что после центрирования и нормирования функция распределения

$$F_n(x) = \mathbf{P}\left(\frac{S_\gamma^{(n)} - \mathbf{E}S_\gamma^{(n)}}{\sqrt{\mathbf{D}S_\gamma^{(n)}}} \le x\right)$$

всё менее отклоняется от функции нормального распределения Φ в том смысле, что

$$\sup_x |F_n(x) - \Phi(x)| \to 0$$

при $n \to \infty$. При этом

$$\sup_x |F_n(x) - \Phi(x)| = o\left(1 \Big/ \sqrt{\mathbf{D}S_\gamma^{(n)}}\right)$$

и порядок $\left(1 \Big/ \sqrt{\mathbf{D}S_\gamma^{(n)}}\right)$ не может быть улучшен. А поскольку

$$\mathbf{D}S_\gamma^{(n)} = n \cdot p(1-p) \cdot (1 + 2/\beta + 2/\beta^2) + n \cdot p^2/\beta^2,$$

то это означает, что аппроксимация F_n с помощью функции Φ при малых p может быть плохой. Возникает вопрос, а нельзя ли при малых p найти лучшую аппроксимацию, нежели так называемая нормальная, даваемая центральной предельной теоремой.

Гиббсовско-пуассоновское распределение

Рассмотрим конечное множество $\mathfrak{X}$ мощности $N = |\mathfrak{X}|$ и N-мерный гиббсовский случайный вектор

$$\boldsymbol{\gamma} = \{\gamma_x,\ x \in \mathfrak{X}\},$$

компоненты которого являются взаимно независимыми одинаково распределенными (b, β, p) – гиббсовскими случайными величинами. Как известно, эти случайные величины равны нулю с вероятностью $1-p$ и строго больше нуля с вероятностью p, когда они имеют β – показательное распределение, сдвинутое на b:

$$\gamma_x = \begin{cases} 0, & \text{с вероятностью } 1-p, \\ \xi_x, & \text{с вероятностью } p, \end{cases}$$

где ξ_x имеет β-показательную плотность, сдвинутую на b:

$$f_x(u) = \begin{cases} 0, & u < b, \\ \beta e^{-\beta(u-b)}, & \text{иначе.} \end{cases}$$

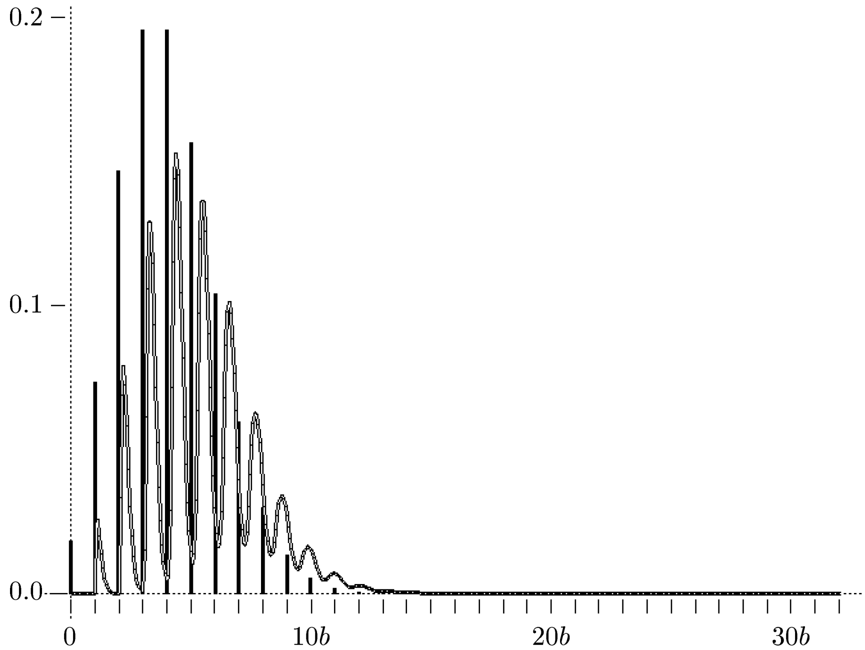

Рис. 2: График плотности (атома вероятности в нуле и плотности положительных значений) гиббсовско – пуассоновского распределения с параметрами $(b, \beta, \lambda) = (10, 1, 4)$ (светлая кривая). Темные прямоугольники — диаграмма вероятностей соответствующего λ-пуассоновского распределения: $\mathbf{P}(\xi_\lambda = bn) = e^{-\lambda}\lambda^n/n!$.

Рассмотрим сумму случайного подмножества $K_{\boldsymbol{\gamma}}$ ненулевых (b, β, p) – гиббсовских случайных величин

$$S(K_{\boldsymbol{\gamma}}) = \sum_{x \in K_\gamma} \xi_x,$$

где $K_{\boldsymbol{\gamma}}$ — случайное множество под $\mathfrak{X}$ с так называемым независимо – точечным распределением

$$p_{\boldsymbol{\gamma}}(X) = \mathbf{P}(K_{\boldsymbol{\gamma}} = X) = \prod_{x \in X} \mathbf{P}(x \in K_{\boldsymbol{\gamma}}) \prod_{x \in X^C} \mathbf{P}(x \notin K_{\boldsymbol{\gamma}}), \quad X \in 2^{\mathfrak{X}},$$

и с одинаковыми вероятностями покрытия

$$\mathbf{P}(x \in K_{\boldsymbol{\gamma}}) = \mathbf{P}(\gamma_x > 0) = p, \quad x \in \mathfrak{X},$$

так что

$$p_{\boldsymbol{\gamma}}(X) = p^{|X|}(1-p)^{|X^C|}, \quad X \in 2^{\mathfrak{X}}.$$

В силу определения случайного множества $K_{\boldsymbol{\gamma}}$ имеем

$$\sum_{x \in \mathfrak{X}} \gamma_x = \sum_{x \in K_\gamma} \xi_x = S(K_{\boldsymbol{\gamma}}).$$

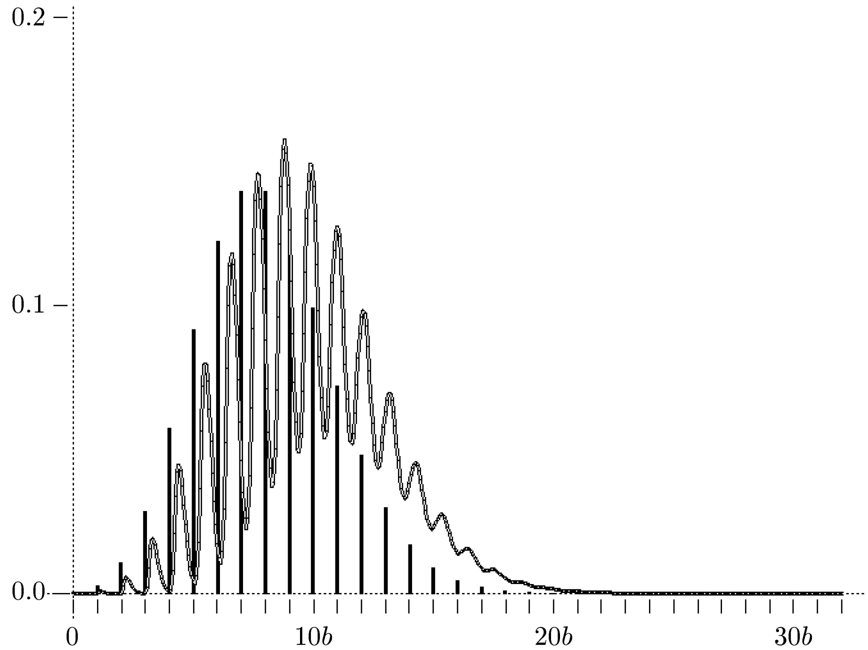

Рис. 3: График плотности (атома вероятности в нуле и плотности положительных значений) гиббсовско – пуассоновского распределения с параметрами $(b, \beta, \lambda) = (10, 1, 8)$ (светлая кривая). Темные прямоугольники — диаграмма вероятностей соответствующего λ-пуассоновского распределения: $\mathbf{P}(\xi_\lambda = bn) = e^{-\lambda}\lambda^n/n!$.

Оказывается, что при малых p, таких что $pN \approx \lambda$, хорошую аппроксимацию распределения суммы $S(K_{\boldsymbol{\gamma}})$ из N взаимно независимых гиббсовских случайных величин дает распределение

$$G_N(w) = p_{\boldsymbol{\gamma}}(\emptyset) + \int_0^w \sum_{X:0<b|X|\leq u} g_X^{(N)}(u - b|X|) p_{\boldsymbol{\gamma}}(X) du, \qquad (1)$$

где

$$g_X^{(N)}(u - b|X|) = \begin{cases} \frac{(\beta(u-b|X|))^{|X|}}{(|X|-1)!} \cdot \beta \cdot e^{-\beta(u-b|X|)}, & u \geq b|X|, \\ 0, & \text{иначе}, \end{cases} \qquad (2)$$

— условные гамма-плотности распределения суммы подмножества $X \subseteq \mathfrak{X}$ ненулевых β – показательных случайных величин (см. у Феллера [1], том 2, стр. 24), а

$$p_{\boldsymbol{\gamma}}(X) = \mathbf{P}(K_{\boldsymbol{\gamma}} = X), \quad X \in 2^{\mathfrak{X}},$$

— распределение вероятностей случайного множества $K_{\boldsymbol{\gamma}}$ ненулевых гиббсовских случайных величин, определяемое распределением исходного вектора гиббсовских случайных величин $\boldsymbol{\gamma}$.

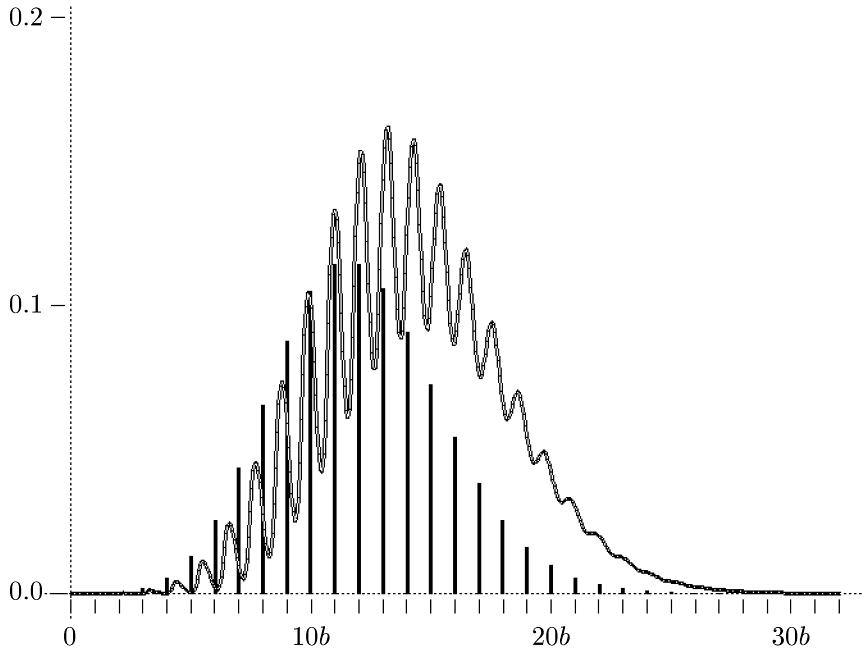

Рис. 4: График плотности (атома вероятности в нуле и плотности положительных значений) гиббсовско – пуассоновского распределения с параметрами $(b, \beta, \lambda) = (10, 1, 12)$ (светлая кривая). Темные прямоугольники — диаграмма вероятностей соответствующего λ-пуассоновского распределения: $\mathbf{P}(\xi_\lambda = bn) = e^{-\lambda}\lambda^n/n!$.

Иначе говоря, распределение (1)

$$G_N = \underbrace{G \star \ldots \star G}_{N} = G^{N\star}$$

— это свертка распределений N взаимно независимых одинаково распределенных гиббсовских случайных величин с параметрами (b, β, p). Как будет показано ниже G_N при $N \to \infty$ всё менее отклоняется от некоего предельного распределения $G_{b,\beta,\lambda}$, которое впервые вводится в данной работе, является частным случаем обобщенного распределения Пуассона, и которое называется *гиббсовско-пуассоновским распределением.*

Предельная теорема для редких гиббсовских случайных событий

Определение (гиббсовско – пуассоновское распределение). *Гиббсовско – пуассоновским распределением* с параметрами (b, β, λ) называется распределение на $[0, \infty)$, которое имеет атом в нуле:

$$\mathbf{P}(\gamma = 0) = e^{-\lambda}$$

и плотность положительных значений вида

$$g_{b,\beta,\lambda}(w) = \begin{cases} \beta \sum\limits_{0\ <\ n\ < w/b} \dfrac{(\lambda\beta(w-bn))^n}{n!(n-1)!} e^{-\lambda\beta(w-bn)}, & b < w, \\ 0, & 0 < w \leq b. \end{cases}$$

Функция $G_{b,\beta,\lambda}$ гиббсовско – пуассоновского распределения имеет вид

$$G_{b,\beta,\lambda}(w) = e^{-\lambda} \left\{ 1 + \sum_{0 < n < w/b} \frac{\lambda^n}{n!} \cdot n \cdot I_\beta(w-bn) \right\},$$

где

$$I_\beta(w-bn) = \frac{\beta^{n+1}}{n!} \int_0^{w-bn} u^n e^{-\beta u} du = 1 - e^{-\beta(w-bn)} \sum_{k=0}^{n} \frac{(\beta(w-bn))^k}{k!}$$

— известный интеграл.

Пусть G_N — функция распределения случайной величины

$$\sum_{n=1}^{N} \gamma_n$$

— суммы взаимно независимых одинаково распределенных (b, β, p) – гиббсовских случайных величин. Справедлива

Теорема. *Пусть $pN \to \lambda$ при $N \to \infty$. Тогда*

$$\sup_w |G_N(w) - G_{b,\beta,\lambda}(w)| \to 0$$

при $N \to \infty$.

Доказательство не требует никаких новых идей, кроме тех, которые обычно используются при доказательстве подобного рода предельных теорем.

Следствие. Случайная величина

$$\gamma = \sum_{n=1}^{\infty} \gamma_n$$

имеет гиббсовско – пуассоновскую функцию распределения $G_{b,\beta,\lambda}$ и плотность $g_{b,\beta,\lambda}$.

На рис. 2, 3 и 4 показаны примеры графиков плотности[1] $g_{b,\beta,\lambda}$ гиббсовско – пуассоновского распределения при различных значениях пуассоновского параметра $\lambda = 4, 8$ и 12. Слепой интервал $[0, b] = [0, 10]$ и параметр спуска $\beta = 1$ одинаковы для всех трех примеров.

Заключение

В работе введено новое предельное распределение, названное *гиббсовско – пуассоновским*, которое точно в той же степени является предельным распределением суммы гиббсовских случайных величин, в какой пуассоновское распределение — суммы бернуллиевских. Введенное распределение $G_{b,\beta,\lambda}$ — это всего лишь частный случай обобщенного пуассоновского распределения, обычно определяемого как λ-пуассоновская сумма

$$Q = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} F^{n\star}.$$

сверток

$$F^{n\star} = \underbrace{F \star \ldots \star F}_{n}$$

некоторого произвольного распределения F. Однако, распределение $G_{b,\beta,\lambda}$ явно заслуживает собственное имя — *гиббсовско – пуассоновское распределение* — по двум причинам.

Во-первых, оно часто встречается в наблюдениях за статистикой неотрицательных случайных величин, связанных, например, со стоимостями покупок. Во-вторых, ”дикобразный” вид его плотности не совсем обычен. Чаще всего при обработке статистических наблюдений пытаются тем или иным способом ”сгладить” оценку плотности, представляя себе в качестве идеала нечто, более или менее напоминающее нормальную плотность. Поэтому имеет прямой смысл ввести собственное наименование для ”ершистых” плотностей, чтобы выделить их

[1] Поскольку гиббсовско – пуассоновское распределение имеет атом в нуле, то изображение ”графика плотности” приходится понимать двояко (другого корректного способа изображения плотности нет): в нуле изображена вероятность нулевого значения, а на интервале $[b, \infty)$ плотность положительных значений.

из общей массы обобщенных пуассоновских плотностей, обратить внимание исследователей на их существование и довольно частую встречаемость, и тем самым предостеречь от излишнего стремления сглаживать "иголки" гиббсовско – пуассоновского "дикобраза", уродуя его в желанный нормальный "колокольчик".

Литература

[1] Феллер В. (1984) *Введение в теорию вероятностей и её приложения, т.1/т.2.* Москва: Мир, 528 с./752 с.

[2] Воробьёв О.Ю. (2002) *Теория случайных событий и её применение.* Красноярск: ИВМ СО РАН, 340 с.

[3] Голденок Е.Е. (2002) Статистическая модель потребительского выбора. *В этой книге.*