Домашнее задание 1

Дородный Дмитрий СКБ172

1 ноября 2019 г.

1 Распределение Бореля-Таннера

1.1 Основные характеристики

Если принять $\alpha > 0$, а r - натуральное число, то $p(x, r, \alpha)$ - распределение Бореля-Таннера:
$p(x, r, \alpha) = \frac{r}{(x-r)!} * x^{x-r-1} * e^{-\alpha x} * \alpha^{x-r}$, где $x = r, r+1...$
В 1942 Борель вывел распределение для случая $r = 1$, а затем Таннер в 1952 показал, что на самом деле оно выполняется для любого целого положительного r.
В отличии от большинства дискретных распределений, определенных на каком-то конечном подмножестве целых чисел, распределение Бореля-Таннера работает для лоюбой целочисленной положительной начальной точки.
Посчитаем математиеское ожидание как момент:
$\sum xp(x) = xp(x, r, \alpha) = xAe^{-\alpha x} * \alpha^{x-r}$
Пусть $\beta = \alpha e^{-\alpha}$, тогда $\beta^x = \alpha^x e^{-\alpha x}$, тогда момент можно записать как: $\sum xA\beta^x\alpha^r$, где α^r не зависит от x, следовательно, можно переписать как: $\alpha^r\mu = \sum xA\beta^x$.
По свойству нормировки сумма вероятностей равна единице:
$\sum \frac{r}{(x-r)!} * x^{x-r-1} * e^{-\alpha x} * \alpha^{x-r} = \sum Ax^{x-r-1}e^{-\alpha x} = \sum A\beta^{x-r}e^{-\alpha r}$, откуда подстановкой в уравнение момента получаем $\alpha^r = \sum A\beta^x$ (1)
Возведем β в степень и домножим на экспоненту, чтобы получить требуемые исходные коэффициенты:
$\beta^{x-r}e^{-\alpha r} = e^{-\alpha x}\alpha^{x-r}$.
Тогда производная β по α:
$\frac{d\alpha}{d\beta} = \frac{\alpha}{\beta(1-\alpha)}$
и вторая:
$\frac{d^2\alpha}{d\beta^2} = \frac{(2-\alpha)\alpha^2}{(1-\alpha)^3\beta^2}$,
также продифференцируем (1) по α:
$\beta r\alpha^{r-1}\frac{d\alpha}{d\beta} = \sum Ax\beta^x$.
Подставив выражение для производной и просуммировав по x, для $x = r, r+1, ...$ получаем выражение для математического ожидания:
$\mu = \frac{r}{1-\alpha}$.
Чтобы вывести дисперсию, продифференцируем (1) два раза, и воспользо-

вавшись формулой для второй производной получаем:
$\mu_2 = \sigma^2 + \mu^2 = \sum x^2 p(x, r, \alpha)$, откуда дисперсия $\sigma^2 = \frac{\alpha r}{(1-\alpha)^3}$
Рассмотрим так же и другие характеристики этого распределения:

Функция распределения

$$F(x) = \begin{cases} 0 & x \leq r \\ \text{re}^{-\alpha r} \sum_{i=0}^{k=r} \frac{\alpha^t}{i!} e^{-\alpha i} (r+i)^{i-1} & k < x \leq k+1, k = r, r+1, ... \end{cases}$$

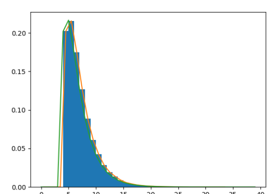
Производящая функция

$\phi_x(t) = rt^r e^{-\alpha r} \sum_{k=0}^{\infty} \frac{(\alpha t)^k}{k!} e^{-\alpha k} (r+k)^{k-1}$

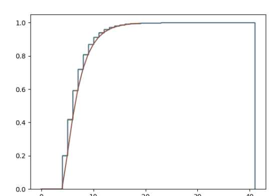
Характеристическая функция

$\phi_x(t) = re^{(it-\alpha)r} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} e^{(it-\alpha)k} (r+k)^{k-1}$

На картинке: синим - гистограмма частот с параметрами $\alpha = 0.4$, $r = 4$



Эмпирическая и теоритические графики функций распределения с теми же параметрами



1.1.1 Интерпретации

Рассмотрим типичную интерпретацию: пусть есть система с постоянным временем обслуживания и одним входом, на который поступает очередь, в виде стационарного пуассоновского потока (т.е. параметр $\lambda = const$, следовательно константная вероятность что за заданный период времени произойдет ровно m обслуживаний), со скоростью в среднем $\alpha, 0 < \alpha < 1$ заявок за время обработки одной заявки. В начальный момент времени очередь имеет длину $r, r \geq 1$. X - чисило обслуживаний за время, требующееся для полного обслуживания очереди, в которой было r заявок в начальный момент времени.

Распределение бореля-таннера также встречается в процессах с ветвлениями, например для определения размера вспышек болезней. Пусть распространение инфекции происходит как гомогенное дерево Галтона-Ватсона ($X_{n+1} = \sum_{j=1}^{X_n} \epsilon_j^{(n)}$, X_n - по смыслу количество потомков в n поколении). Предположим, что инфекция распространяется так, что каждый больной инфицирует Z здоровых людей, которые играют роль потомства в этом случае. Получается распределение Z повинуется степенной ф-ции. $P(Z = r) = \frac{\theta^r}{r! e^\theta}$. Пусть X - размер вспышки инфекции, получившейся из $S = s$ начальных больных, тогда $P(X = x, x) = Const(x, s) * \frac{\theta^{x-s}}{A(\theta)^x}$). Тогда если распределение "потомков"(заражения) происходило по пуассоновским законам, то размер вспышки эпидемии будет подчиняться распределению Бореля-Таннера: $P(X = x, s) = \frac{s * x^{x-s-1} \lambda^{x-s} e^{-x\lambda}}{(x-s)!}, x = s, s+1...$

1.1.2 Моделирование

Для моделирования использовался рекурсивный алгоритм:
$p(x) = \frac{r}{(x-r)!} x^{x-r-1} e^{-\alpha x} \alpha^{x-r}$
$p(x-1) = \frac{r}{(x-r-1)!} (x-1)^{x-r-2} e^{-\alpha(x-1)} \alpha^{x-r-1}$
$= \frac{r}{(x-r)!} x^{x-r-1} e^{-\alpha x} \alpha^{x-r} * (x-r)(x-1)^{x-r-2} e^{\alpha} \alpha^{-1} \frac{1}{x^{x-r-1}}$
$= p(x) * \frac{x-r}{x} (\frac{x-1}{x})^{x-r-2} * e^{\alpha} * \frac{1}{\alpha}$
Получаем
$p(x) = p(x-1) * \frac{x}{x-r} (\frac{x}{x-1})^{x-r-2} e^{-\alpha} \alpha, x = r+1, r+2....$
Первый член:
$p(r) = \frac{r}{(r-r)!} r^{r-r-1} e^{-\alpha r} \alpha^{r-r} = e^{-\alpha r}$
Начальное состояние $\alpha, r, x = r, p = e^{-\alpha r}$, ρ - случайная равномерно распределенная от 0 до 1 величина
В цикле обновляем $\rho = \rho - p$ (по смыслу - вероятность обработки заявки на данном шаге). Если $r < 0$, то x - сгенерированная величина. Если нет, то увеличиваем x на 1 и вычисляем следующий шаг p умножая текущее p на $\frac{x}{x-r} (\frac{x}{x-1})^{x-r-2} e^{-\alpha} \alpha$
Код доступен по ссылке

2 Гамма распределение

2.1 Основные характеристики

Пусть случаная величина X имеет гамма-распределение, тогда её плотность вероятности будет выглядеть так:

$f_X(x) = \begin{cases} x^{k-1}\frac{e^{-\frac{x}{\theta}}}{\theta^k\Gamma(k)}, & \text{если } x \geq 0 \\ 0, & \text{если } x < 0 \end{cases}$

Для начала найдем характеристическую функцию случайной величины X:

$\int\limits_0^{+\infty} e^{itx}x^{k-1}\frac{e^{-\frac{x}{\theta}}}{\theta^k\Gamma(k)}dx = \frac{1}{\theta^k\Gamma(k)}\int\limits_0^{+\infty} x^{k-1}e^{-x(\frac{1}{\theta}-it)}dx = |z = x(\frac{1}{\theta} - it) \Rightarrow dz = (\frac{1}{\theta} - it)dx| = \frac{1}{\theta^k\Gamma(k)(\frac{1}{\theta}-it)^k}\int\limits_0^{+\infty} z^{k-1}e^{-z}dz = \frac{1}{(1-it\theta)^k}$

Теперь найдем математическое ожидание и дисперсию:

$iEX = \frac{d(1-it\theta)^{-k}}{dt}|_{t=0} \Rightarrow EX = k\theta$

$iEX^2 = \frac{d^2(1-it\theta)^{-k}}{dt^2}|_{t=0} \Rightarrow EX^2 = k(k+1)\theta^2$

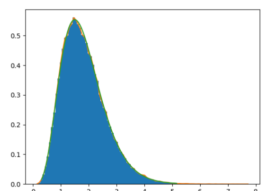
Так как $DX = EX^2 - (EX)^2$, получаем:

$DX = k\theta^2$

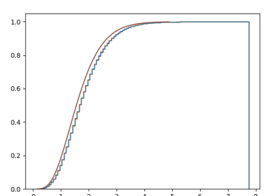
Функция распределения:

$\int\limits_0^x t^{k-1}\frac{e^{-\frac{t}{\theta}}}{\theta^k\Gamma(k)}dt = |\frac{t}{\theta} = z| = \frac{1}{\Gamma(k)}\int\limits_0^{\frac{x}{\theta}} z^{k-1}e^{-z}dz$

На картинке: синим - гистограмма частот с параметрами $\lambda = 5$, $\alpha = 3$



Эмпирическая и теоритические графики функций распределения с теми же параметрами



2.1.1 Интерпретации

Гамма-распределение присутствует в процессах, где важно время ожидания, по сути своей определяя время ожидания между пуассоновскими событиями. Пуассоновская функция вероятности: $P(k) = \frac{\lambda^k * e^- \lambda}{k!}$, характеризует вероятность, что на интервале произойдет x событий, λ - среднее количество событий на интервале в общем случае. Если вместо среднего количества использовать размер интервала t и частоту событий r ($\lambda = rt$) выражение можно преобразовать как: $P(N(t) = k) = \frac{(rt)^k e^{-rt}}{k!}$, где $N(t)$ - количество событий, произошедших между настоящим временем и временем t. Очевидно, это является гамма распределением с параметрами $\alpha = k+1$ и $\theta = \frac{1}{\lambda}$: $G(k, \frac{1}{\lambda}) = \frac{x^{k-1}e^{-x\lambda}}{\Gamma(k)\lambda^{-k}} = \frac{\lambda(x\lambda)^{k-1}e^{-x\lambda}}{(k-1)!}$. Пример: допустим, рыбак рассчитывает ловить одну рыбу за полчаса. Рассчитаем вероятность того, что ему придется рыбачить от 2 до 4 часов чтобы поймать 4 рыбы: $P(2 \leq X \leq 4) = \sum_{x=2}^{4} \frac{x^{4-1}e^{\frac{-x}{2}}}{\Gamma(4)2^4} = 0.123$.

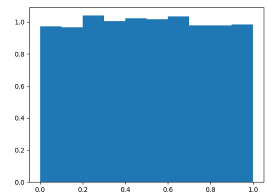
Нетипичная интерпретация - использование для оценки надежности технических систем. Пусть устройство отказывает тогда, когда отказывают не менее k его элементов. Если отказ элемента - пуассоновское событие $P_n(t) = \frac{(\lambda\tau)^n}{n!}$, то отказы элементов подчинены экспоненциальному распределению $P(x) = e^{-\lambda x}$. Получаем, что плотность вероятности отказа утройства, отказ которого вызывается отказом k элементов будет $f(t) = \frac{\lambda_0^k t^{k-1}}{(k-1)!}e^{-\lambda_0 t}$, где λ_0 - исходная интенсивность отказов элементов, что является гамма распределением с параметрами k, t.

2.1.2 Моделирование

Используется теорема Джонка-Ньюмэна-Оделла-Виттейкера: пусть a, b - заданные константы, а U, V - равномерно от 0 до 1 распределенные случайные величины. Тогда, если $U^{\frac{1}{a}} + V^{\frac{1}{b}} \leq 1$, случайная величина $\frac{U^{\frac{1}{a}}}{U^{\frac{1}{a}} + V^{\frac{1}{b}}}$ распределена по бета распределению. Начальное состояние - $\theta = \frac{1}{\lambda}, \alpha$. Возьмем три независимые равномерно распределенные от 0 до 1 случайные величины $r_1, r_2 < r_3$. Пусть $S_1 = r_1^{\frac{1}{\alpha}}, S_2 = r_2^{\frac{1}{1-\alpha}}$. Тогда когда выполняется условие теоремы: $S_1 + S_2 \leq 1$, выводим сгенерированную величину как $E\frac{S_1}{S_1+S_2} = -\frac{ln(r_3)}{\lambda}\frac{S_1}{S_1+S_2}$ - Алгоритм работает для $\alpha < 1$. Для остальных случаев разделяем параметр α на целую и дробную части (по свойству о композиции законов распределения сумма по разным параметрам формы с одним и тем же параметром масштаба будет давать исходное гамма распределение), и целую часть моделируем как сумма $[\alpha]$ показательно сгенерированных величин. (Т.к. $\Gamma([\alpha], \lambda) = \sum \Gamma(1, \lambda)$). Для $\alpha > 1$ получаем метод генерации: $y - \sum \frac{ln(r_i)}{\lambda} = y - \frac{ln(r_1 * \ldots * r_{[\alpha]})}{\lambda}$, где y - величина сгенерированная по алгоритму, представленному выше.
Код доступен по ссылке

3 Генератор случайных равномернораспределенных величин

использовалась библиотечная ф-ция, имеющая равномерное распределение по документации. При построении гистограммы действительно видно, что



она распределена равномерно от 0 до 1: