

# Сюжеты с кружка по эконометрике

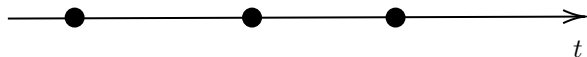
2 июля 2019 г.

## 1 Идеи распределений

### 1.1 Экспоненциальное распределение и распределение Пуассона через АПП

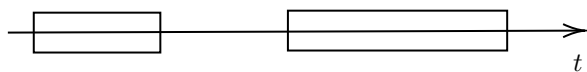
Аксиомы Пуассоновского потока (АПП):

1. Время непрерывно.
2. На оси времени происходят точечные происшествия.



«Точечные» означает, что происшествие не длится, скажем, 2 секунды, а случается «в точке» на оси времени.

3. Количества происшествий на непересекающихся интервалах независимы.



4. Закон распределения количества происшествий стабилен во времени. Это означает, что на двух временных интервалах одинаковой длины количества происшествий распределены одинаково.

5. На малом интервале времени  $\Delta t$ :

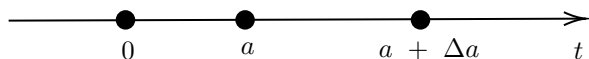
- вероятность двух и более происшествий мала по сравнению с  $\Delta t$ :

$$\mathbb{P}(2 \text{ и более происшествий}) = o(\Delta t).$$

- вероятность ровно одного происшествия пропорциональна  $\Delta t$  с точностью до  $o$ -малых:

$$\mathbb{P}(\text{ровно 1 происшествие}) = \lambda \Delta t + o(\Delta t).$$

Посмотрим, как из АПП можно вывести экспоненциальное распределение и распределение Пуассона. Рассмотрим следующий участок на оси времени:



Попробуем найти вероятность того, что за промежуток времени  $[0; a]$  произойдёт ровно 0 происшествий:

$$\mathbb{P}(0 \text{ происшествий за } [0; a]).$$

Для этого рассмотрим вероятность того, что ровно 0 происшествий произойдёт за промежуток времени  $[0; a + \Delta a]$ :

$$\mathbb{P}(0 \text{ происшествий за } [0; a + \Delta a]).$$

Так как  $[0; a]$  и  $[a; a + \Delta a]$  – непересекающиеся интервалы, то по аксиоме 3 количество происшествий на них независимы, а значит совместная вероятность раскладывается в произведение:

$$\mathbb{P}(0 \text{ происшествий за } [0; a + \Delta a]) = \mathbb{P}(0 \text{ происшествий за } [0; a]) \times \mathbb{P}(0 \text{ происшествий за } [a; a + \Delta a]).$$

А  $\mathbb{P}(0 \text{ происшествий за } [a; a + \Delta a])$  можно рассчитать по аксиоме 5. На этом интервале может произойти ноль, одно или два и более событий. Так как нас интересует первый вариант, вычтем из единицы вероятности второго и третьего вариантов:

$$\begin{aligned} \mathbb{P}(0 \text{ происшествий за } [a; a + \Delta a]) &= 1 - \mathbb{P}(1 \text{ происшествие за } [a; a + \Delta a]) - \\ &\quad - \mathbb{P}(2 \text{ и более происшествий за } [a; a + \Delta a]). \end{aligned}$$

А эти вероятности – ровно то, что стоит в аксиоме 5, только в нашем случае интервал  $\Delta t$  – это  $\Delta a$ . Таким образом, получаем:

$$\begin{aligned} \mathbb{P}(0 \text{ происшествий за } [a; a + \Delta a]) &= 1 - \mathbb{P}(1 \text{ происшествие за } [a; a + \Delta a]) - \\ &\quad - \mathbb{P}(2 \text{ и более происшествий за } [a; a + \Delta a]) = \\ &= 1 - \lambda \Delta a - o(\Delta a) - o(\Delta a) = 1 - \lambda \Delta a - o(\Delta a) \text{ (по свойствам } o\text{-малых)}. \end{aligned}$$

Вернёмся к начальному выражению:

$$\mathbb{P}(0 \text{ происшествий за } [0; a + \Delta a]) = \mathbb{P}(0 \text{ происшествий за } [0; a]) \times (1 - \lambda \Delta a - o(\Delta a)).$$

Раскроем скобки:

$$\mathbb{P}(0 \text{ происшествий за } [0; a + \Delta a]) - \mathbb{P}(0 \text{ происшествий за } [0; a]) = \mathbb{P}(0 \text{ происшествий за } [0; a])(-\lambda \Delta a + o(\Delta a)),$$

так как  $\mathbb{P}(0 \text{ происшествий за } [0; a]) \times o(\Delta a) = o(\Delta a)$  (вероятность лежит в пределах от нуля до единицы, то есть мала).

Поделим обе части на  $\Delta a$ :

$$\frac{\mathbb{P}(0 \text{ происшествий за } [0; a + \Delta a]) - \mathbb{P}(0 \text{ происшествий за } [0; a])}{\Delta a} = \mathbb{P}(0 \text{ происшествий за } [0; a])(-\lambda) + \frac{o(\Delta a)}{\Delta a}.$$

А теперь возьмём предел правой и левой части при  $\Delta a \rightarrow 0$ . Заметим, что слева стоит ни что иное как производная  $\mathbb{P}(0 \text{ происшествий за } [0; a])$  (по определению производной), если мы рассматриваем эту вероятность как функцию. Также по определению:

$$\lim_{\Delta a \rightarrow 0} \frac{o(\Delta a)}{\Delta a} = 0.$$

Обозначим вероятность  $\mathbb{P}(0 \text{ происшествий за } [0; a])$  как  $z(a)$ , то есть явно скажем, что это некоторая функция:

$$\mathbb{P}(0 \text{ происшествий за } [0; a]) \equiv z(a).$$

Тогда выражение выше можно записать как:

$$z'_a = -\lambda z(a) + 0.$$

Общее решение этого дифференциального уравнения:

$$z(a) = C e^{-\lambda a}.$$

Заметим, что  $z(a)$  уже похожа на функции распределения экспоненциального распределения и распределения Пуассона. Восстановим константу  $C$ . Логично предположить, что  $z(0) = 1$ , то есть вероятность того, что за промежуток времени  $[0; 0]$  произойдёт 0 происшествий, равна единице. Отсюда:

$$1 = C e^0 \Rightarrow C = 1.$$

Таким образом:

$$z(a) = e^{-\lambda a}.$$

Что такое  $z(a)$ ? По нашей записи это вероятность того, что в промежуток времени  $[0; a]$  произойдёт 0 происшествий.

Распределение Пуассона как раз моделирует случайную величину, которая показывает число событий, произошедших за фиксированный промежуток времени. Предположим, что  $a = 1$ , то есть  $z(a) = e^{-\lambda}$  (промежуток времени фиксирован от 0 до 1). Таким образом,  $z(a)$  – это вероятность того, что случайная величина, имеющая Пуассоновское распределение, примет значение 0 на промежутке времени  $[0; 1]$ . Покажем это:

$$\begin{aligned} X &\sim \text{pois}(\lambda), \\ \mathbb{P}(X = 0) &= \frac{e^{-\lambda} \times \lambda^0}{0!} = e^{-\lambda}. \end{aligned}$$

Заметим, что  $z(a)$  можно интерпретировать по-другому: если это вероятность того, что за промежуток времени  $[0; a]$  произойдёт 0 происшествий, то это же вероятность того, что первое происшествие произойдёт после точки  $a$  на временной оси. Пусть случайная величина  $Y$  показывает время до первого происшествия. Тогда:

$$z(a) = \mathbb{P}(Y \geq a).$$

Перепишем данное выражение в терминах функции распределения:

$$P(Y \leq a) = 1 - z(a) = 1 - e^{-\lambda a}.$$

Получили функцию распределения экспоненциального распределения! Экспоненциальное распределение моделирует время между двумя происшествиями. У нас получилось, что  $Y \sim \exp(\lambda)$  и  $Y$  – время до первого происшествия, то есть экспоненциальное распределение также показывает время от начала отсчёта до первого происшествия. Ввиду свойства отсутствия памяти, два этих определения эквивалентны (так как после наступления нового происшествия, точка отсчёта смещается в точку этого происшествия).

Покажем, что распределение Пуассона можно вывести для любого числа происшествий. Например, проверим, что наши рассуждения верны и для  $\mathbb{P}(1 \text{ происшествие за } [0; a])$ . Будем рассматривать фиксированный промежуток времени  $[0; 1]$  (то есть при  $a = 1$ ). Ожидаемый результат:

$$\begin{aligned} X &\sim \text{pois}(\lambda). \\ \mathbb{P}(X = 1) &= e^{-\lambda} \lambda. \end{aligned}$$

Для краткости обозначим искомую вероятность за  $u(a)$ :

$$\mathbb{P}(1 \text{ происшествие за } [0; a]) \equiv u(a).$$

Применим ту же схему. Вероятность того, что произошло ровно 1 происшествие за промежуток времени  $[0; a + \Delta a]$  раскладывается в сумму вероятностей, что ровно 1 происшествие произошло либо за  $[0; a]$ , либо за  $[a; a + \Delta a]$ . Первая вероятность – это произведение вероятностей, что за  $[0; a]$  произошло ровно одно происшествие, а за  $[a; a + \Delta a]$  произошло ноль происшествий. Вторая вероятность – это произведение вероятностей, что за  $[0; a]$  произошло ноль происшествий, а за  $[a; a + \Delta a]$  произошло ровно одно происшествие. Запишем эти рассуждения в наших обозначениях:

$$u(a + \Delta a) = u(a)(1 - \lambda \Delta a - o(\Delta a)) + z(a)(\lambda \Delta a + o(\Delta a)).$$

Снова раскроем скобки и поделим обе части на  $\Delta a$ :

$$\frac{u(a + \Delta a) - u(a)}{\Delta a} = \lambda z(a) - \lambda u(a) + \frac{o(\Delta a)}{\Delta a}.$$

В пределе при  $\Delta a \rightarrow 0$  получаем:

$$u'_a = \lambda e^{-\lambda a} - \lambda u(a).$$

Условие  $u(0) = 0$  опять же выполняется. Если решить данное дифференциальное уравнение, получим:

$$u(a) = \frac{e^{-\lambda a} \lambda a}{1!}.$$

При  $a = 1$  получаем:

$$u(a) = e^{-\lambda} \lambda.$$

Получили, что ожидали. Можно проверить результаты и для большего числа происшествий.

Важный факт, который получаем из двоякой интерпретации  $z(a)$ : если предполагаем, что время между двумя происшествиями распределено экспоненциально, то их количество распределено по Пуассону. И наоборот, если считаем, что количество происшествий распределено по Пуассону, то время между двумя происшествиями распределено экспоненциально.

## 1.2 Гамма-распределение: $\gamma(k, \lambda)$

Гамма распределение будем рассматривать на примере ловли червячков птичкой. По смыслу, случайная величина  $s \sim \gamma(k, \lambda)$  показывает суммарное время на ловлю  $k$  червячков, если матожидание количества червячков за единицу времени равно  $\lambda$ . Раз это суммарное время, то гамма-распределение – это ещё и сумма случайных величин, распределённых экспоненциально с параметром  $\lambda$ .

Выведем общую формулу гамма-распределения по индукции, начав со случая  $\gamma(3, \lambda)$ . Пусть  $Y_1, Y_2, Y_3 \sim \exp(\lambda)$ . Тогда:

$$f(y_1, y_2, y_3) dy_1 \wedge dy_2 \wedge dy_3 = \lambda e^{-\lambda y_1} \lambda e^{-\lambda y_2} \lambda e^{-\lambda y_3} dy_1 \wedge dy_2 \wedge dy_3.$$

Обозначим:  $S_3 = Y_1 + Y_2 + Y_3$ ,  $S_2 = Y_1 + Y_2$ ,  $S_1 = Y_1$ ;  $R_2 = S_1/S_2$ ,  $R_3 = S_2/S_3$ ,  $R_4 = S_3/S_4$ . По смыслу,  $R_i$  – какая доля времени потрачена на поимку  $(i-1)$  червячка, если известно время на поимку  $i$  червячков. Перейдём от  $Y_1, Y_2, Y_3$  к  $R_2, R_3, S_3$ .

$$Y_1 = S_1 = R_2 \times R_3 \times S_3.$$

$$Y_2 = S_2 - S_1 = (1 - R_2) \times R_3 \times S_3.$$

$$Y_3 = S_3 - S_2 = (1 - R_3) \times S_3.$$

Далее нам нужно подставить эти выражения в дифференциальную форму выше. Прежде чем сделать это, выведем часть «с птичками».

$$\begin{aligned} dy_1 \wedge dy_2 &= d(r_2 r_3 s_3) \wedge d((1 - r_2) r_3 s_3) = (d[r_2] r_3 s_3 + r_2 d[r_3] s_3) \wedge (-d[r_2] r_3 s_3 + (1 - r_2) d[r_3] s_3) = \\ &= (1 - r_2) r_3 s_3 d[r_2] \wedge d[r_3] s_3 + r_2 r_3 s_3 d[r_2] \wedge d[r_3] s_3 = r_3 s_3 d[r_2] \wedge d[r_3] s_3. \end{aligned}$$

$$\begin{aligned} dy_1 \wedge dy_2 \wedge dy_3 &= \\ &= r_3 s_3 d[r_2] \wedge d[r_3] s_3 \wedge (-d[r_3] s_3 + (1 - r_3) ds_3) = \\ &= r_3 s_3 d[r_2] \wedge (d[r_3] s_3 + r_3 d[s_3]) \wedge (-d[r_3] s_3 + (1 - r_3) d[s_3]) = \\ &= r_3 s_3 d[r_2] \wedge (s_3(1 - r_3) d[r_3] d[s_3] + s_3 r_3 d[r_3] d[s_3]) = \\ &= r_3 s_3^2 dr_2 dr_3 ds_3. \end{aligned}$$

Теперь подставим всё, что нашли, в дифференциальную форму:

$$f(y_1, y_2, y_3) dy_1 \wedge dy_2 \wedge dy_3 = \lambda^3 e^{-\lambda s_3} r_3 s_3^2 dr_2 dr_3 ds_3.$$

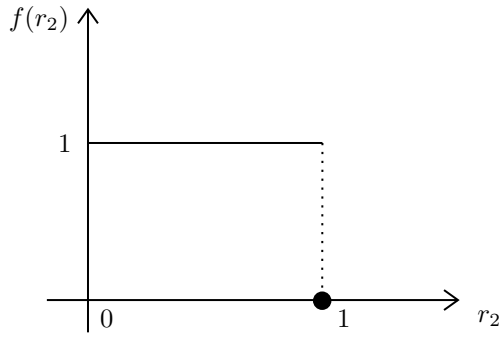
Отсюда:

$$f(r_2, r_3, s_3) = \lambda^3 e^{-\lambda s_3} r_3 s_3^2.$$

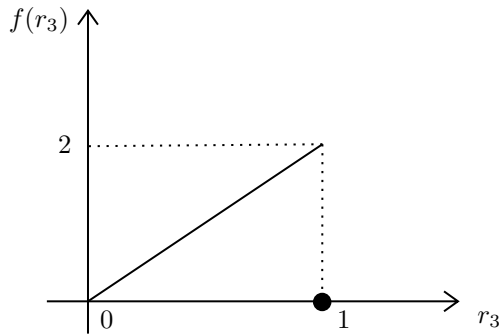
Замечаем, что совместная функция плотности раскладывается на произведение индивидуальных функций плотности:

$$f(r_2, r_3, s_3) = f(r_2) f(r_3) f(s_3) = [C_1] [C_2 r_3] [C_3 e^{-\lambda s_3} s_3^2].$$

Восстановим константы. Понятно, что  $r_2 \sim U[0, 1]$ , (от 0 до 1, так как это доля времени) – а значит,  $C_1 = 1$ :



Далее понимаем, что  $r_3$  тоже распределено от 0 до 1, так как это доля времени. Посмотрим на функцию плотности  $r_3$ :



Чтобы интеграл под функцией плотности равнялся 1,  $C_2$  должна быть равна 2. Выходит, что  $C_3 = \frac{\lambda^3}{3}$ .

На самом деле, константы нам не так важны. Обобщим, что получили про гамма-распределение. Если  $S_k$  – суммарное время на ловлю  $k$  червячков, если ожидаемое количество пойманных червячков за единицу времени равно  $\lambda$ , то  $S_k \sim \gamma(k, \lambda)$  и  $f(s) = Ce^{-\lambda s} s^{k-1}$ .

Покажем, что это верно и для следующего шага индукции. Перейдём от  $r_2, r_3, s_3$  к  $r_2, r_3, r_4, s_4$ . Тогда  $s_3 = s_4 \times r_4$ . Вернёмся к дифференциальной форме:

$$1 \times 2r_3 \times \frac{\lambda^3}{2} e^{-\lambda s_3} s_3^2 dr_2 dr_3 ds_3.$$

При добавлении  $y_4$  форма домножится на  $\wedge \lambda e^{-\lambda y_4} dy_4$ . Понятно, что  $y_4 = (1 - r_4)s_4$ .

Рассчитаем новую часть «с птичками».

$$ds_3 \wedge dy_4 = d(s_4 r_4) \wedge d((1 - r_4)s_4) = (d[s_4]r_4 + s_4 d[r_4]) \wedge (-d[r_4]s_4 + (1 - r_4)d[s_4]) = (r_4 s_4 + s_4(1 - r_4))d[r_4] \wedge d[s_4] = s_4 dr_4 ds_4.$$

Понятно, что  $s_3^2 = (s_4 r_4)^2$ . Тогда новая дифференциальная форма имеет вид (подставляем всё, что получили):

$$1 \times 2r_3 \times 3r_4^2 \times \frac{\lambda^4}{3!} s_4^3 e^{-\lambda s_4} dr_2 dr_3 dr_4 ds_4.$$

Совпадает с нашей формулой выше с точностью до константы. Аналогично можно сделать следующий шаг индукции для  $ds_4 \wedge dy_5$ .

Так как гамма-распределение – это сумма экспоненциальных, легко вывести матожидание и дисперсию гамма-распределения:

$$\mathbb{E}(\gamma) = k \times \frac{1}{\lambda}.$$

$$Var(\gamma) = Var(Y_1 + \dots + Y_k) = kVar(Y_1) = k \times \frac{1}{\lambda^2}.$$

Понятно, что при  $k = 1$  гамма-распределение – это экспоненциальное распределение.

Полное общее определение:  $S_k \sim \gamma(k, \lambda)$ , где  $S_k$  – суммарное время на ловлю  $k$  червячков,  $\lambda$  – ожидаемое количество пойманных червячков за единицу времени. Тогда:

$$\begin{aligned} f(s_k) &= \frac{\lambda^k}{(k-1)!} e^{-\lambda s_k} s_k^{k-1}. \\ \mathbb{E}(S_k) &= \frac{k}{\lambda}. \\ \text{Var}(S_k) &= \frac{k}{\lambda^2}. \end{aligned}$$

### 1.3 Бета-распределение: $\beta(a, b)$

По смыслу: ловим  $(a + b)$  червячков,  $a$  отдаём,  $b$  оставляем себе. Тогда, если  $R$  – доля времени на поимку  $a$  червячков, то  $R \sim \beta(a, b)$ .

Пусть  $Z_1$  – время на поимку  $a$  червячков, а  $Z_2$  – доля на поимку  $b$  червячков. Тогда  $Z_1 \sim \gamma(a, \lambda)$ ,  $Z_2 \sim \gamma(b, \lambda)$ . Дифференциальная форма:

$$f(z_1, z_2) dz_1 \wedge dz_2 = \frac{\lambda^a}{(a-1)!} e^{-\lambda z_1} z_1^{a-1} \times \frac{\lambda^b}{(b-1)!} e^{-\lambda z_2} z_2^{b-1} dz_1 \wedge dz_2.$$

Перейдём к нашим обозначениям  $r$  и  $s$ . Понятно, что  $Z_1 = Y_1 + \dots + Y_a$ ,  $Z_2 = Y_{a+1} + \dots$ . Тогда обозначим:

$$\begin{aligned} s &= Z_1 + Z_2. \\ r &= \frac{Z_1}{Z_1 + Z_2}. \end{aligned}$$

Тогда  $Z_1 = rs$ ,  $Z_2 = (1 - r)s$ . Часть «с птичками»:

$$d(rs) \wedge d((1 - r)s) = sdr \wedge ds \text{ (считали выше).}$$

Подставляем всё в дифференциальную форму:

$$\frac{\lambda^{a+b}}{(a-1)!(b-1)!} e^{-\lambda s} r^{a-1} (1-r)^{b-1} s^{(a-1)+(b-1)+1} dr \wedge ds.$$

Перегруппируем:

$$\frac{\lambda^{a+b}}{(a+b-1)!} s^{a+b-1} e^{-\lambda s} \times \frac{(a+b-1)!}{(a-1)!(b-1)!} r^{a-1} (1-r)^{b-1} dr \wedge ds.$$

Замечаем, что до знака  $\times$  стоит функция плотности гамма-распределения  $f(s)$  с параметрами  $k = a+b$  и  $\lambda$ . А после этого знака –  $f(r)$ , и по нашему обозначению  $r$ , это и есть функция плотности бета-распределения.

Выпишем отдельно: если  $R$  – доля времени на поимку  $a$  червячков, а всего ловим  $(a + b)$  червячков, то  $R \sim \beta(a, b)$  и:

$$\begin{aligned} f(r) &= \frac{(a+b-1)!}{(a-1)!(b-1)!} r^{a-1} (1-r)^{b-1}. \\ \mathbb{E}(R) &= \frac{a}{a+b}. \end{aligned}$$

### 1.4 Нормальное распределение (через Гаусса)

Предпосылки:

1. Есть истинная величина  $\mu$ .
2.  $y_i = \mu + u_i$ ,  $u_i$  – независимы и симметричны около 0 (то есть свидетели равновероятно завышают и занижают показания).

3.  $\bar{y} = \hat{\mu}_{ML} \forall y_1 \dots y_{n+1}$ .

4?  $f(u)$  – дифференцируемая.

Тогда  $u_i \sim N(0, \sigma^2)$ .

Выпишем правдоподобие:

$$L = f(y_1 - \mu)f(y_2 - \mu) \dots f(y_{n+1} - \mu).$$

$$\ell = \sum_{i=1}^{n+1} \ln(f(y_i - \mu)).$$

$$\ell'_\mu = - \sum_{i=1}^{n+1} \frac{f'(y_i - \mu)}{f(y_i - \mu)}.$$

По предпосылке 3:

$$\sum_{i=1}^{n+1} \frac{f'(y_i - \bar{y})}{f(y_i - \bar{y})} = 0 \forall y_1 \dots y_{n+1}.$$

Так как выполняется для любых  $y_i$ , возьмём конкретные показания. Для них должно выполняться:

$$\begin{aligned} y_1 - \bar{y} &= a, \\ y_2 - \bar{y} &= a, \\ &\dots \\ y_n - \bar{y} &= a, \\ y_{n+1} - \bar{y} &= -na. \end{aligned}$$

Получаем:

$$n \frac{f'(a)}{f(a)} + \frac{f'(-na)}{f(-na)} = 0 \forall n, a.$$

Так как  $f$  – симметричная дифференцируемая:

$$\frac{f'(na)}{f(na)} = n \frac{f'(a)}{f(a)}.$$

Обозначим:  $h(x) = \frac{f'(x)}{f(x)}$ . Получаем, что  $h(na) = nh(a) \forall n, a$ . Это означает, что  $h(x) = kx$ . Получаем дифференциальное уравнение:

$$\frac{f'(x)}{f(x)} = kx.$$

Решение:

$$f(x) = C_1 e^{\frac{kx^2}{2}}.$$

## 2 Лайфхаки

1. Гамма-функция:

