Теория вероятности

Дима Трушин

Семинар 5

Случайные векторы (продолжение)

Условные вероятности и условные математические ожидания Пусть даны две дискретные случайные величины ξ и η

$$\xi \sim \begin{cases} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{cases} \quad \text{и} \quad \eta \sim \begin{cases} b_1 & b_2 & \dots & b_n & \dots \\ q_1 & q_2 & \dots & q_n & \dots \end{cases}$$

Тогда определены условные вероятности[1]

$$P(\xi = a \mid \eta = b) = \frac{P(\xi = a, \eta = b)}{P(\eta = b)}$$

В случае, если $b \neq b_i$ ни для какого i (то есть при условии $P(\eta = b) = 0$), то мы полагаем эту вероятность равной нулю. В этом случае можно определить условное математическое ожидание

$$\mathbb{E}(\xi \mid \eta = b) = \sum_{i=1}^{\infty} a_i P(\xi = a_i \mid \eta = b)$$

Пусть теперь ξ и η – непрерывные случайные величины с плотностями $p_\xi(x)$ и $p_\eta(y)$ и совместной плотностью $p_{\xi,\eta}(x, y)$. Тогда можно определить условную плотность

$$p_{\xi|\eta=y}(x) = p(\xi = x \mid \eta = y) = \frac{p_{\xi,\eta}(x, y)}{p_\eta(y)}$$

В этом случае можно определить условную вероятность следующим образом

$$P(\xi \in A \mid \eta = y) = \int\limits_A p_{\xi|\eta=y}(x)\, dx = \int\limits_A \frac{p_{\xi,\eta}(x, y)}{p_\eta(y)}\, dx$$

В этом случае условное математическое ожидание определяется так

$$\mathbb{E}(\xi \mid \eta = y) = \int\limits_{\mathbb{R}} x p_{\xi|\eta=y}(x)\, dx = \int\limits_{\mathbb{R}} x \frac{p_{\xi,\eta}(x, y)}{p_\eta(y)}\, dx$$

Смысл условного распределения $P(\xi \in A \mid \eta = y)$ и в дискретном и в непрерывном случае одинаковый. Оно показывает, какова вероятность, что ξ покажет результат из множества A, при условии, что η показало значение y. Важно, что для дискретных случайных величин условное распределение тоже будет дискретным, а для непрерывных – непрерывным. Про условное математическое ожидание $\mathbb{E}(\xi \mid \eta = y)$ надо думать так. Оно показывает, какое в среднем значение будет показывать ξ среди тех попыток, когда η показывает y. Обратите внимание, что $g(y) = \mathbb{E}(\xi \mid \eta = y)$ можно рассматривать как случайную величину на $\mathbb{R}$ с распределением η. Тогда ее математическое ожидание будет[2]

$$\mathbb{E}_y(g(y)) = \mathbb{E}(\mathbb{E}(\xi \mid \eta = y)) = \mathbb{E}(\xi)$$

[1] Чтобы в этой формуле посчитать числитель, не достаточно знать распределение каждой из случайных величин. Надо полностью знать их совместное распределение. Однако, я не стал его вводить явно.

[2] Индекс y в математическом ожидании указывает по какой из случайных величин мы берем математическое ожидание. Это очень часто встречаемое обозначение особенно полезно во втором равенстве, чтобы понять, какого черта вообще происходит.

Если бы случайные величины ξ и η были бы независимы, то $\mathbb{E}(\xi \mid \eta = y) = \mathbb{E}(\xi)$. То есть мы варьируем значение η с помощью y и при этом среднее для ξ остается одним и тем же. Однако, если случайные величины зависимы, то при разных значениях η случайная величина ξ будет иметь разные средние $\mathbb{E}(\xi \mid \eta = y)$. Меняя y мы будем менять среднее для ξ при условии, что $\eta = y$. Но если мы возьмем и усредним все эти средние по η (с учетом меры для η) то мы в итоге получим обычное среднее для ξ. Именно эта последняя мысль и заключается в формуле для математического ожидания от условного ожидания. Еще на эту формулу можно смотреть как на интегрирование по частям.

Условное математическое ожидание относительно индикатора Если $\xi\colon \Omega \to \mathbb{R}$ – произвольная случайная величина и $A \subseteq \Omega$ – некоторое событие, то можно рассмотреть условное математическое ожидание ξ относительно χ_A, а именно функцию $\mathbb{E}(\xi \mid \chi_A = y)$. Математическое ожидание $\mathbb{E}(\xi \mid \chi_A = 1)$ совпадает с математическим ожиданием ξ, но по вероятностной мере $P_A(*) = P(* \mid A)$, то есть $\mathbb{E}(\xi \mid \chi_A = 1) = \int_\Omega \xi \, dP_A$. Таким образом это будет усреднение значений ξ по всем исходам, которые встречаются в A. Кроме того, $\mathbb{E}(\xi \mid \chi_A = 0)$ совпадает с $\mathbb{E}(\xi \mid \chi_{\bar{A}} = 1)$, а значит является усреднением по всем значениям ξ на элементарных исходах не из A.

Замечание про общий подход Скажу по секрету, что есть общий подход к условному математическому ожиданию. Тогда можно определить случайную величину $\mathbb{E}(\xi \mid \eta)$, которая является усреднением ξ на всех событиях, выражаемых в терминах η. Потому сохраняется более общая формула

$$\mathbb{E}(\mathbb{E}(\xi \mid \eta)\chi_A) = \mathbb{E}(\xi\chi_A), \quad \text{для любого } \; A = \eta^{-1}(B), \quad \text{где } \; B \subseteq \mathbb{R}$$

Распределение суммы независимых случайных величин

Пусть на некотором вероятностном пространстве (Ω, P) даны две независимые случайные величины ξ и η. Предположим, что мы знаем распределение каждой из них, то есть знаем меры P_ξ и P_η на прямой $\mathbb{R}$. Наша задача найти меру для суммы $P_{\xi+\eta}$.

Дискретный случай. Если обе случайные величины ξ и η дискретны, то и сумма их будет дискретна. Пусть исходные случайные величины заданы следующим образом

$$\xi \sim \begin{cases} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{cases} \quad \text{и} \quad \eta \sim \begin{cases} b_1 & b_2 & \dots & b_n & \dots \\ q_1 & q_2 & \dots & q_n & \dots \end{cases}$$

Тогда все возможные значения для суммы $\xi + \eta$ имеют вид $a_i + b_j$. А соответствующая мера считается так

$$P(\xi + \eta = x) = \sum_{a_i + b_j = x} p_i q_j$$

Непрерывный случай. Свертка. Теперь предположим, что ξ непрерывна и задана плотностью $p(x)$, а η имеет плотность $q(y)$. Утверждается, что случайная величина $\xi + \eta$ так же будет непрерывной с плотностью

$$p_{\xi+\eta}(x) = \int\limits_{\mathbb{R}} p(t)q(x-t)\,dt = \int\limits_{\mathbb{R}} p(x-t)q(t)\,dt$$

Действительно, давайте для начала вычислим функцию распределения

$$F_{\xi+\eta}(x) = P(\xi + \eta \leqslant x) = \int\limits_{u+v \leqslant x} p(u)q(v)\,dudv$$

Теперь можно сделать две замены переменных (одну для получения первого выражения, другую для – второго).

$$\begin{cases} s = u + v \\ t = u \end{cases} \quad \text{или} \quad \begin{cases} s = u + v \\ t = v \end{cases}$$

В матричном виде эти замены выглядят

$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v \\ u \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Обратная замена

$$\begin{pmatrix} v \\ u \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix}$$

Теперь пользуемся формулой

$$dudv = |\det(C)| dsdt \quad \text{где} \quad \begin{pmatrix} u \\ v \end{pmatrix} = C \begin{pmatrix} s \\ t \end{pmatrix}$$

Получаем

$$F_{\xi+\eta}(x) = \int\limits_{s \leqslant x} p(t)q(s-t)\, dsdt = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{\infty} p(t)q(s-t)\, dtds = \int\limits_{-\infty}^{x} \left(\int\limits_{-\infty}^{\infty} p(t)q(s-t)\, dt \right) ds$$

Аналогично для второй замены

$$F_{\xi+\eta}(x) = \int\limits_{s \leqslant x} p(s-t)q(t)\, dsdt = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{\infty} p(s-t)q(t)\, dtds = \int\limits_{-\infty}^{x} \left(\int\limits_{-\infty}^{\infty} p(s-t)q(t)\, dt \right) ds$$

Так как плотность – это производная от функции распределения, то нужный ответ получается дифференцированием по x.

Интеграл вида

$$\int\limits_{\mathbb{R}} p(t)q(x-t)\, dt = \int\limits_{\mathbb{R}} p(x-t)q(t)\, dt$$

Называется сверткой функций p и q. То есть мы проверили, что для независимых непрерывных величин, плотность суммы есть свертка плотностей слагаемых.

Характеристики случайных векторов

Если на вероятностном пространстве (Ω, P) определен случайный вектор $\bar{\xi}\colon \Omega \to \mathbb{R}^n$, то для него определены аналоги математического ожидания и дисперсии. Математическое ожидание будет некоторым вектором, который в среднем приближает наш случайный, а вот дисперсия будет матрицей квадратичной формы, отвечающей за случайные отклонения от этого среднего значения.

Математическое ожидание Математическое ожидание опять же определяется непонятной формулой

$$\mathbb{E}(\bar{\xi}) = \int\limits_{\Omega} \bar{\xi}\, dP = \int\limits_{\mathbb{R}^n} \bar{x}\, dP_{\bar{\xi}}$$

Здесь $\bar{x} = (x_1, \ldots, x_n)$ – вектор из переменных. Давайте придадим ей смысл в случае дискретного и непрерывного распределений.

Пусть случайный вектор $\bar{\xi}$ имеет дискретное распределение

$$\xi \sim \begin{cases} a_1 & \ldots & a_n & \ldots & \in \mathbb{R}^n \\ p_1 & \ldots & p_n & \ldots & \in [0,1] \end{cases}$$

Тогда

$$\mathbb{E}(\bar{x}) = \sum_{i=1}^{\infty} a_i p_i = \sum_{i=1}^{\infty} a_i P(\bar{\xi} = a_i)$$

Пусть теперь $\bar{\xi}$ имеет непрерывное распределение с плотностью $p(x_1, \ldots, x_n)$. Тогда

$$\mathbb{E}(\bar{\xi}) = \int\limits_{\mathbb{R}^n} \bar{x} p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

Обратите внимание, что тут интегрируется векторная функция $\bar{x}p(x_1, \ldots, x_n)$, где $\bar{x} = (x_1, \ldots, x_n)$.

В обоих случаях, для вектора $\bar{\xi} = (\xi_1, \ldots, \xi_n)$, верна формула

$$\mathbb{E}(\bar{\xi}) = (\mathbb{E}(\xi_1), \ldots, \mathbb{E}(\xi_n))$$

То есть математическое ожидание можно вычислять покоординатно. Для начала надо вычислить распределения каждой из координаты, а потом посчитать матожидание для каждой координаты.

Матрица ковариации В случае нескольких случайных величин мы будем определять не только среднее отклонение от среднего, но и корелированность каждой пары случайных величин. По определению ковариацией случайных величин ξ и η называется следующее выражение

$$\operatorname{cov}(\xi, \eta) = \mathbb{E}\left((\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta)\right)$$

Обратите внимание, что $\mathbb{D}(\xi) = \operatorname{cov}(\xi, \xi)$.

Если $\bar{\xi} = (\xi_1, \ldots, \xi_n)$ – случайный вектор на некотором вероятностном пространстве, то матрицей ковариации этого вектора называется

$$\Sigma(\bar{\xi}) = \begin{pmatrix} \operatorname{cov}(\xi_1, \xi_1) & \ldots & \operatorname{cov}(\xi_1, \xi_n) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(\xi_n, \xi_1) & \ldots & \operatorname{cov}(\xi_n, \xi_n) \end{pmatrix}$$

Важное замечание. Эта матрица всегда является симметричной и неотрицательно определенной. А значит, в случае, если она не вырождена, то она задает некоторое скалярное произведение на $\mathbb{R}^n$, а именно $\beta(x, y) = x^t \Sigma(\bar{\xi})^{-1} y$.[3] Мы знаем, что можно представить матрицу $\Sigma(\bar{\xi})$ в следующем виде

$$\Sigma(\bar{\xi}) = UDU^t, \text{ где } U = (u_1 | \ldots | u_n) \text{ – ортогональная матрица и } D = \operatorname{diag}(d_1, \ldots, d_n)$$

В этом случае давайте посмотрим на поверхность $x^t \Sigma(\bar{\xi})^{-1} x = 1$ – это поверхность, показывающая границу среднего отклонения вектора $\bar{\xi}$ от его математического ожидания. Эта граница образует эллипс. Векторы $u_1, \ldots, u_n$ – будут осями этого эллипса, а числа $d_1, \ldots, d_n$ характеризуют на сколько этот эллипс вытянут вдоль каждой из осей $u_1, \ldots, u_n$ (корни из d_i показывают максимальное значение x_i по модулю).

Если для двух случайных величин ξ и η мы имеем $\operatorname{cov}(\xi, \eta) = 0$, то говорят, что случайные величины не коррелированы. Отметим, что если случайные величины независимы, то они и не коррелированы, однако, обратное вообще говоря не верно.

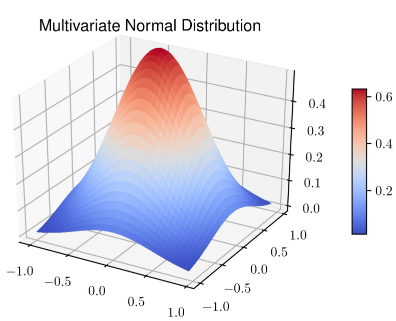
Многомерное гауссово распределение Я хочу задать аналог гауссова распределения в пространстве $\mathbb{R}^n$. Для этого зафиксируем некоторый вектор $a \in \mathbb{R}^n$ и матрицу $\Sigma \in \mathrm{M}_n(\mathbb{R})$ такую, что Σ симметрична и положительна определена. Тогда можно задать непрерывное распределение со следующей плотностью

$$p_{\bar{\xi}}(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-a)^t \Sigma^{-1}(x-a)}$$

В этом случае получаем

1. $\mathbb{E}(\bar{\xi}) = a$.
2. $\operatorname{cov}(\xi_i, \xi_j) = \Sigma_{i,j}$.

Ниже на картинке нарисована плотность нормального распределения со средним 0 и матрицей ковариации $\frac{1}{2}E$:



Важно, что многомерное гауссово распределение полностью определяется своими матожиданием и матрицей ковариации на столько, что его координаты независимы тогда и только тогда, когда они некоррелированы.

[3] Обратите внимание на обратную степень!

Геометрия гауссова распределения Пусть у нас задан случайный гауссов вектор ξ с плотностью

$$p_\xi(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-a)^t \Sigma^{-1}(x-a)}$$

Тогда выражение под экспонентой $(x-a)^t\Sigma^{-1}(x-a)$ является квадратичной формой относительно вектора $x-a$. Так как матрица Σ является положительно определенной, то уравнение $(x-a)^t\Sigma^{-1}(x-a) = 1$ задает эллипсоид с центром в точке a. Точка a является математическим ожиданием нашей случайной величины ξ, потому мы будем ожидать, что ответы ξ будут копиться близко к этой точке. А указанный эллипсоид показывает границу среднего отклонения значения ξ от математического ожидания.

Мы знаем, что существует ортонормированный базис $v_1, \ldots, v_n$ такой, что матрица Σ будет диагональна $D = \operatorname{diag}(d_1, \ldots, d_n)$. Пусть $C = (v_1|\ldots|v_n)$ – матрица перехода от стандартного базиса к базисе $v_1, \ldots, v_n$. Тогда матрица C ортогональна, то есть $C^tC = E$. В этом случае $D = C^t\Sigma C$.

Давайте сдвинем начало координат в a и выберем координатные оси вдоль векторов $v_1, \ldots, v_n$. Тогда в новом базисе $v_1, \ldots, v_n$ и новым началом координат случайный вектор ξ имеет вид $\eta = C^t(\xi - a)$. Вектор η будет нормально распределенным вектором с центром в точке 0 и матрицей ковариации $D = \operatorname{diag}(d_1, \ldots, d_n)$. То есть

$$p_\eta(x) = \frac{1}{\sqrt{(2\pi)^n d_1 \ldots d_n}} e^{-\frac{1}{2}\left(\frac{x_1^2}{d_1} + \ldots + \frac{x_n^2}{d_n}\right)} = \frac{1}{\sqrt{2\pi d_1}} e^{-\frac{x_1^2}{2d_1}} \ldots \frac{1}{\sqrt{2\pi d_n}} e^{-\frac{x_n^2}{2d_n}}$$

Значит d_1 будет дисперсией первой координаты для η, d_2 для второй и т.д. Кроме того, координаты вектора η оказались независимыми и нормально распределены. Потому мы знаем, что i-я координата в среднем меняется от $-\sqrt{d_i}$ до $\sqrt{d_i}$, то есть в среднем значения вектора будут заключены в прямоугольник. Однако, этот результат можно уточнить, значения будут реже бывать в углах этого прямоугольника и в итоге будут сосредоточены в среднем внутри упомянутого выше эллипсоида.