§9. Моментные характеристики многомерных распределений. Мультиномиальное и многомерное нормальное распределения

Для описания положения в пространстве, рассеяния и формы многомерных распределений обычно используются *смешанные* центральные моменты, вычисляемые как математические ожидания от произведения различных степеней центрированных средними значениями компонент случайного вектора:

$$\mathbf{E}\left[(X_1 - \mu_1)^{k_1} \cdots (X_n - \mu)^{k_n}\right],$$

где $\mu_i = \mathbf{E}X_i$ – вектор средних значений компонент случайного вектора $X^{(n)}$. Мы будем иметь дело только с моментами второго порядка $\lambda_{ij} = \mathbf{E}(X_i - \mu_i)(X_j - \mu_j),\ i, j = 1, \ldots, n$. Матрица $\Lambda = \|\lambda_{ij}\|$ моментов второго порядка называется *ковариационной матрицей* или *матрицей ковариаций* $\operatorname{cov}(X_i, X_j) = \lambda_{ij}$. Естественно, диагональ ковариационной матрицы Λ составляют дисперсии $\sigma_i^2 = \lambda_{ii} = \operatorname{cov}(X_i, X_i)$ соответствующих компонент $X_i,\ i = 1, \ldots, n$ случайного вектора $X^{(n)}$, в то время как смешанные моменты λ_{ij} при $i \neq j$ характеризуют степень *линейной связности* компонент X_i и X_j. Этот термин требует специального обсуждения, ввиду его исключительной распространенности в приложениях многомерного статистического анализа.

Всё вертится около следующего лебеговского варианта неравенства Коши–Буняковского.

Неравенство Шварца. *Пусть X и Y – случайные величины, а $g(X)$ и $h(Y)$ – измеримые функции от соответствующих величин, обладающие конечными вторыми моментами. Тогда*

$$|\mathbf{E}g(X)h(Y)| \leq \left[\mathbf{E}g^2(X)\mathbf{E}h(Y)^2\right]^{1/2}$$

с равенством тогда и только тогда, когда функции g и h линейно связаны: существуют такие постоянные a и b, что $P\left(ag(X) + bh(Y) = 0\right) = 1$.

Применим это неравенство к функциям $g(X) = X - \mu_X$ и $h(Y) = Y - \mu_Y$, где $\mu_X = \mathbf{E}X,\ \mu_Y = \mathbf{E}Y$. Если случайные величины X и Y независимы, то, в силу предложения 8.3б $\operatorname{cov}(X, Y) = \mathbf{E}(X - \mu_X)(Y - \mu_Y) = \mathbf{E}(X - \mu_X)\mathbf{E}(Y - \mu_y) = 0$, то есть независимые случайные величины имеют нулевую ковариацию. Если же X и Y линейно связаны:

$Y-\mu_Y = a(X-\mu_X)$, то в неравенстве Шварца достигается знак равенства, так что $\mathrm{cov}(X,\,Y) = \lambda_{XY} = \mathbf{E}(X-\mu_X)(Y-\mu_Y) = \pm\sqrt{\mathbf{D}X\cdot\mathbf{D}Y} = \pm\sigma_X\sigma_Y > 0$ (естественно, мы предполагаем, что X и Y принимают по крайней мере два различных значения с ненулевой вероятностью). Эти два крайних значения в неравенстве Шварца оправдывают введение следующей меры линейной связности пары случайных величин.

Определение 9.1. Пусть X и Y – две случайные величины с конечными дисперсиями. Моментная характеристика

$$\rho = \rho_{XY} = \frac{\lambda_{XY}}{\sigma_X\sigma_Y}$$

называется *коэффициентом корреляции* между случайными величинами X и Y.

Итак, если X и Y независимы, то $\rho = 0$, если же $Y = a + bX$ при некоторых постоянных a и b, то $|\rho| = 1$, причем $\rho = -1$, если $a < 0$, и $\rho = +1$, если $a > 0$. Однако, **равенство $\rho = 0$ не означает, что случайные величины X и Y независимы!**

Пример 9.1 (*зависимых случайных величин с нулевым коэффициентом корреляции*). Покажем, что случайные величины X и Y, равномерно распределенные в круге радиуса r, зависимы, но $\rho_{XY} = 0$.

Действительно, совместная функция плотности $f(x,\,y)$ случайных величин X и Y (см. пример 8,1) отлична от нуля только в круге $x^2+y^2 \le r^2$ и принимает постоянное значение, равное $1/\pi r^2$, внутри этого круга. Маргинальные плотности

$$f^X(x) = \frac{2}{\pi r^2}\sqrt{r^2-x^2},\ \ |x| \le r; \qquad f^Y(y) = \frac{2}{\pi r^2}\sqrt{r^2-y^2},\ \ |y| \le r,$$

и $f^X(x) = f^Y(y) = 0$ вне квадрата $|x| \le r$, $|y| \le r$.

Имеем: $f^X(x)f^Y(y) = 4\pi^{-2}r^{-4}\left[(r^2-x^2)(r^2-y^2)\right]^{1/2}$, что, очевидно, не совпадает с $f(x,\,y) = 1/\pi r^2$ в области $x^2+y^2 \le r^2$. Таким образом, в силу предложения 8.2, случайные величины X и Y зависимы.

Покажем, что, тем не менее, $\rho_{XY} = 0$. Функция $f(x,\,y)$ центрально симметрична, и поэтому $\mu_X = \mu_Y = 0$. Далее,

$$\lambda_{XY} = \frac{1}{\pi r^2}\int\limits_{-r}^{r} x\,dx \int\limits_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} y\,dy = 0.$$

Но если $\lambda_{XY} = 0$, то и $\rho = 0$.

Для ковариации пары случайных величин справедливы формулы, аналогичные тем, что были получены для дисперсии в предложениях 6.1 и 8.3.

Предложение 9.1. *Для любой пары случайных величин $(X,\, Y)$ и независимых двумерных векторов $(X_1,\, Y_1), \ldots, (X_n,\, Y_n)$, обладающих конечными вторыми моментами, справедливы равенства*

$$(1) \quad \lambda_{XY} = \mathbf{E}XY - \mathbf{E}X\mathbf{E}Y,$$

$$(2) \quad \lambda_{S_X S_Y} = \sum_1^n \lambda_{X_i Y_i},$$

где $S_X = \sum_1^n X_i, \quad S_Y = \sum_1^n Y_i$.

Доказательство. (1) Имеем $\lambda_{XY} = \mathbf{E}\,[(X - \mathbf{E}X)(Y - \mathbf{E}Y)] = \mathbf{E}\,(XY - Y\mathbf{E}X - X\mathbf{E}Y + \mathbf{E}X\mathbf{E}Y) = \mathbf{E}XY - \mathbf{E}X\mathbf{E}Y$.

(2) Делаем столь же тривиальные выкладки, что и выше, и при этом не забывем, что среднее от произведения независимых случайных величин равно произведению средних:

$$\lambda_{S_X S_Y} = \mathbf{E}\left[\sum_1^n (X_i - \mathbf{E}X_i) \sum_1^n (Y_i - \mathbf{E}Y_i)\right] =$$

$$\mathbf{E}\left[\sum_1^n (X_i - \mathbf{E}X_i)(Y_i - \mathbf{E}Y_i) + \sum_{i \neq j} (X_i - \mathbf{E}X_i)(Y_j - \mathbf{E}Y_j)\right] =$$

$$\sum_1^n \mathbf{E}\,[(X_i - \mathbf{E}X_i)(Y_i - \mathbf{E}Y_i)] + \sum_{i \neq j} \mathbf{E}(X_i - \mathbf{E}X_i) \cdot \mathbf{E}(Y_j - \mathbf{E}Y_j).$$

Последнее слагаемое в правой части равно нулю, ибо $\mathbf{E}(X_i - \mathbf{E}X_i) = \mathbf{E}X_i - \mathbf{E}X_i = 0$, а первое слагаемое есть сумма ковариаций каждого вектора.

Изучим две наиболее распространенные многомерные вероятностные модели.

<u>*Лекция 15*</u>

Мультиномиальное распределение $\mathcal{M}(m,\, n, \mathbf{p})$. Рассматривается схема независимых испытаний, в каждом из которых может произойти одно из $m \geq$ событий $A_1, \ldots, A_m$ с вероятностями $p_1, \ldots, p_n$,

$\sum_1^m p_j = 1$. Типичный пример таких испытаний – наблюдения энтомолога по оценке численности видов насекомых, населяющих некоторый, достаточно изолированный район нашей планеты. Всего проводится n независимых испытаний, и регистрируются значения $x_1, \ldots, x_m$ компонент случайного вектора $X^{(m)} = (X_1, \ldots, X_m), \sum_1^m X_j = n$, где x_j – количество испытаний, в которых произошло событие A_j, $j = 1, \ldots, m$.

Легко видеть, что мы имеем дело с многомерным аналогом схемы Бернулли, и для вывода распределения $X^{(m)}$ естественно воспользоваться техникой стохастических представлений наблюдаемого случайного элемента в виде суммы индикаторов, то есть поступить по аналогии с примером 8.2. Свяжем с каждым i-ым испытанием случайный вектор $Y_i = X_{1i}, \ldots, X_{mi}$, каждая компонента X_{ji} которого принимает значение 1, если в i-ом испытании произошло событие A_j, и $X_{ji} = 0$ в противном случае. Таким образом, все компоненты Y_i равны нулю за исключением одной компоненты, равной единице, и номер этой компоненты совпадает с номером исхода (события A_j), которым завершилось i-ое испытание, $i = 1, \ldots, n$, $j = 1, \ldots, m$. Постулируется, что случайные векторы $Y_1, \ldots, Y_n$ независимы в совокупности (следствие независимости проведения испытаний).

При таком соглашении каждая компонента X_j наблюдаемого вектора $X^{(m)}$ имеет стохастическое представление

$$X_j = \sum_{i=1}^{n} X_{ji}, \tag{1}$$

в котором $X_{j1}, \ldots, X_{jn}$ независимы и одинаково $\mathrm{B}(1, \mathrm{p_j})$ распределены: принимают значение 1 с вероятностью p_j и значение 0 с вероятностью $1 - p_j$, $\quad j = 1, \ldots, m$. Из представления (1) следует, что вероятность любого события в n мультиномиальных испытаниях (значений, которые принимают векторы $Y_1, \ldots, Y_n$) определяется только от количествами $x_1, \ldots, x_m$ испытаний, которые завершились соответствующими исходами $A_1, \ldots, A_m$. Легко видеть, что эта вероятность равна $p_1^{x_1} \cdots p_m^{x_m}$, где $\sum_1^m x_j = n$. Теперь для того, чтобы вывести функцию плотности $f(x_1, \ldots, x_m) = P(X_1 = x_1, \ldots, X_m = x_m)$, достаточно решить комбинаторную задачу, которую мы умеем решать в случае $m = 2$: сколькими способами можно получить x_1 исходов A_1, x_2 исходов A_2, $\ldots$ x_m исходов A_m в $n = \sum_1^m x_i$ испытаниях? Решение задачи дают *мультиномиальные коэффициенты*

$$C_n^{x_1\ldots x_m} = \frac{n!}{x_1!\cdots x_m!}$$

(сравните с биномиальными коэффициентами C_n^x). Итак, функция плотности *мультиномиального распределения* $\mathcal{M}(\mathrm{m}, \mathrm{n}\,\mathbf{p})$ по m-кратному произведению считающих мер равна

$$f(x_1,\ldots,x_m) = \frac{n!}{x_1!\cdots x_m!}p_1^{x_1}\cdots p_m^{x_m},$$

в области $\sum_1^m x_j = n$ и $f(x_1,\ldots,x_m) = 0$ в случае целых $x_1,\ldots,x_m$, не удовлетворяющих последнему равенству, а также в случае дробных x_j, $j = 1,\ldots,m$.

Вычислим моментные характеристики мультиномиального распределения, используя стохастическое представление (1). Вектор средних

$$\mathbf{E}X_j = \sum_{i=1}^n \mathbf{E}X_{ji} = n(1.p_j + 0\cdot(1-p_j)) = np_j, \quad j = 1,\ldots,m;$$

вектор дисперсий

$$\sigma_j^2 = \mathbf{E}\left(\sum_{i=1}^n X_{ji}\right)^2 - (\mathbf{E}X_j)^2 = \sum_{i=1}^n \mathbf{E}X_{ji}^2 + \sum_{i\neq k}\mathbf{E}X_{ji}\mathbf{E}X_{jk} - n^2p_j^2 =$$

$$np_j + n(n-1)p_j^2 - n^2p_j^2 = np_j(1-p_j), \quad j = 1,\ldots,m.$$

Вычислим ковариации X_j с X_l при $j \neq l$:

$$\lambda_{jl} = \sum_{i=1}^n \mathbf{E}(X_{ji}X_{li}) + \sum_{i\neq k}\mathbf{E}X_{ji}\mathbf{E}X_{lk} - \mathbf{E}X_j\mathbf{E}X_l.$$

Поскольку при i-ом испытании только одна из компонент $X_{1i},\ldots,X_{mi}$ равна единице, а остальные равны нулю, то $X_{ji}X_{li} \equiv 0$, первая сумма в правой части последнего равенства равна нулю и $\lambda_{jl} = n(n-1)p_jp_l - n^2p_jp_l = -np_jp_l$. Коэффициенты корреляции

$$\rho_{jl} = -\frac{np_jp_l}{n\left(p_j(1-p_j)p_l(1-p_l)\right)^{1/2}} = -\sqrt{\frac{p_jp_l}{(1-p_j)(1-p_l)}}, \quad j \neq l.$$

Отрицательные значения коэффициентов корреляции есть следствие связей между компонентами наблюдаемого вектора: $\sum_1^m X_j = n$.

Многомерное нормальное распределение $\mathcal{N}_m(\mu,\, \Lambda)$**.** Мы трактовали мультиномиальную схему испытаний как многомерный аналог схемы независимых испытаний Бернулли. В таком случае естественно рассмотреть многомерный аналог предельной теоремы Муавра–Лапласа. Применяя формулу Стирлинга к мультиномиальным коэффициентам, нетрудно убедиться, что имеет место

Теорема 9.1. (Интегральная предельная теорема для мультиномиального распределения). *Для любых постоянных* $(a_1,\, b_1)$, $\ldots,(a_m,\, b_m)$ *и* $(m+1)$*-мерного случайного вектора* $X^{(m+1)} = (X_1,\ldots,$ $X_m,\, X_{m+1}) \sim \mathcal{M}(m+1,\, n,\, \mathbf{p})$ *справедливо асимптотическое представление*

$$\lim_{n\to\infty} P\left(a_1 \leq \frac{X_1 - np_1}{\sqrt{np_1(1-p_1}} < b_1,\ldots,a_m \leq \frac{X_m - np_m}{\sqrt{np_m(1-p_m}} < b_m\right) =$$

$$\frac{1}{(2\pi)^{m/2}\sqrt{|\mathbf{P}|}}\int\limits_{a_1}^{b_1}\cdots\int\limits_{a_m}^{b_m}\exp\left\{-\frac{1}{2}\mathbf{x}'\mathbf{P}^{-1}\mathbf{x}\right\}dx_1\cdots dx_m, \qquad (2)$$

где $\mathbf{x}' = (x_1,\ldots x_m)$, $\mathbf{x}$ *–аналогичный вектор столбец;* $\mathbf{P} = \|\rho_{ij}\|$ *– корреляционная матрица, в которой* $\rho_{ii} = 1$ *и*

$$\rho_{ij} = -\sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}},$$

если $i \neq j, \quad i,j = 1,\ldots,m;$ $\mathbf{P}^{-1}$ *– матрица, обратная к* $\mathbf{P}$, *наконец,* $|\mathbf{P}|$ *– определитель матрицы* $\mathbf{P}$.

Интеграл (2) определяет непрерывное распределение вероятностей на прямоугольниках (следовательно, и на борелевском поле) пространства R^m, причем функция плотности этого распределения

$$f_m(\mathbf{x}' \,|\, \mathbf{P}) = \frac{1}{(2\pi)^{m/2}\sqrt{|\mathbf{P}|}}\exp\left\{-\frac{1}{2}\mathbf{x}'\mathbf{P}^{-1}\mathbf{x}\right\}, \quad \mathbf{x}' \in \mathrm{R}^m.$$

Если теперь вместо корреляционной матрицы $\mathbf{P}$ мультиномиального распределения рассмотреть произвольную, положительно определенную корреляционную матрицу $\mathbf{P} = \|\rho_{ij}\|$, $\rho_{ii} = 1$, $i = 1,\ldots,m$, то $f_m(\mathbf{x}' \,|\, \mathbf{P})$ будет функцией плотности случайного вектора $X^{(m)} =$

$(X_1, \ldots, X_m)$, имеющего *стандартное m-мерное нормальное распределение* $\mathcal{N}_m(0,\, \mathbf{P})$. Далее, если ввести вектор средних $\mu = (\mu_1, \ldots, \mu_m)$ и вектор дисперсий $\sigma^2 = (\sigma_1^2, \ldots, \sigma_m^2)$, то случайный вектор $\sigma X_1 + \mu_1, \ldots, \sigma_m X_m + \mu_m$ будет иметь *многомерное нормальное распределение* $\mathcal{N}_m(\mu,\, \Lambda)$ с функцией плотности

$$\varphi_m(\mathbf{x}' \mid \mu,\, \Lambda) = \frac{1}{(2\pi)^{m/2}\sqrt{|\,\Lambda\,|}} \exp\left\{-\frac{1}{2}(x-\mu)'\Lambda^{-1}(x-\mu)\right\} =$$

$$\frac{1}{(2\pi)^{m/2}\sigma_1\cdots\sigma_m\sqrt{|\,\mathbf{P}\,|}} \exp\left\{-\frac{1}{2}\sum_{i,j}\frac{\mathbf{P}_{ij}}{|\,\mathbf{P}\,|}\cdot\frac{(x_i-\mu_i)(x_j-\mu_j)}{\sigma_i\sigma_j}\right\}, \quad \mathbf{x}' \in \mathrm{R}^m,$$

где ковариационная матрица $\Lambda = \sigma_1\cdots\sigma_m \cdot \mathbf{P}$, а $\mathbf{P}_{ij}/|\mathbf{P}|$ – элементы матрицы $\mathbf{P}^{-1}$, обратной к $\mathbf{P}$.

Нетрудно видеть, что если коэффициенты корреляции $\rho_{ij} = 0$, когда $i \neq j$, то есть $\mathbf{P}$ есть единичная матрица, а в матрице ковариаций Λ отличны от нуля только диагональные элементы $\sigma_1^2, \ldots, \sigma_m^2$, то нормальная функция плотности распадается в произведение маргинальных нормальных $\mathcal{N}_1(\mu_i,\, \sigma_i^2)$, $i = 1, \ldots, m$ функций плотности. Таким образом, справедливо

Предложение 9.2. *В случае нормального распределения случайного вектора некоррелированность его компонент влечет их независимость.*

Следует обратить особое внимание на требование *положительной определенности* корреляционной матрицы $\mathbf{P}$ или, что то же, ковариационной матрицы Λ. Если эти матрицы положительно полуопределены, то есть имеют ранг $r < m$, то мы получим *несобственное m-*мерное нормальное распределение, вся вероятностная масса которого будет сосредоточена на гиперплоскости R^r, а между компонентами случайного вектора $X^{(m)}$ будет существовать линейная зависимость.

Указанные свойства многомерного нормального распределения наиболее наглядно прослеживаются в случае $m = 2$ – нормального распределения на плоскости R^2. Функция плотности распределения случайного вектора $(X, Y) \sim \mathcal{N}_2(\mu,\, \Lambda)$ при $\rho = \rho_{XY} \neq \pm 1$ равна

$$\varphi_2(x,\, y \,|\, \mu,\, \Lambda) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot$$
$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right\}.$$

Как отмечалось выше, $\mu_1 = \mathbf{E}X$, $\mu_2 = \mathbf{E}Y$ есть вектор средних значений; $\sigma_1^2 = \mathbf{D}X$, $\sigma_2^2 = \mathbf{D}Y$ – дисперсии соответствующих случайных величин; $\rho = \mathrm{cov}(X,\, Y)/\sigma_1\sigma_2$, – коэффициент корреляции между X и Y. В том, что это действительно так, можно убедиться и непосредственным вычислением интегралов, определяющих соответствующие моментные характеристики. Маргинальные функции плотности $f^X(x)$ и $f^Y(y)$ находятся также непосредственным интегрированием совместной функции плотности $\varphi_2(x,\, y \,|\, \mu, \Lambda)$ по соответствующим переменным y и x :

$$f^X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\},\; f^Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\},$$

то есть $X \sim \mathcal{N}(\mu_1,\, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2,\, \sigma_2^2)$. Особо отметим, что *существуют многомерные распределения, отличные от нормального, но имеющие нормальные маргинальные распределения.*

Подобные эллипсы

$$\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right) = c^2$$

играют роль *кривых равных вероятностей*: нетрудно вычислить, что вероятность попадания $(X,\, Y)$ в область, ограниченную этим эллипсом, равна $1 - \exp\{-c^2\}$.

Форма эллипса равных вероятностей дает хорошее представление о виде поверхности $z = \varphi_2(x,\, y \,|\, \mu,\, \Lambda)$ нормальной плотности. При $\rho = 0$, $\sigma_1 = \sigma_2$ эллипсы превращаются в окружности. Когда ρ приближается к $+1$ или -1, эллипсы становятся более тонкими и вытянутыми, что является показателем стремления вероятностной массы сосредотачиваться около общей большой оси этих эллипсов.

Особый интерес представляет эллипс с $c = 2$, который называется *эллипсом рассеяния.* Он обладает достаточно высокой вероятностью $1 - e^{-4} \approx 0.98$ попадания случайной точки $(X,\, Y)$ внутрь эллипса и еще одним замечательным свойством: равномерное распределение по области, ограниченной эллипсом рассеяния, имеет те же моменты первого (μ_1, μ_2) и второго (σ_1^2, σ_2^2, $\rho\sigma_1\sigma_2$) порядков, что и нормальное распределение.

В заключение отметим, что двумерное нормальное распределение играет важную роль в *теории стрельб*: распределение координат точек попадания при стрельбе из закрепленного ствола хорошо согласуется с нормальным законом.