

Условные математические ожидания относительно σ -алгебр и отображений

В.В. Некруткин

кафедра статистического моделирования <http://statmod.ru>, матмех СПбГУ
Материалы к специальному курсу, 2018 г.

Содержание

1	Элементарные условные распределения	3
1.1	Основные формулы	3
1.2	Упражнения	7
2	Условные математические ожидания относительно σ-алгебр	12
2.1	Определение и простейшие свойства	12
2.2	УМО относительно отображений и функция регрессии	13
2.3	Предельный переход под знаком УМО	17
2.4	Дальнейшие свойства УМО	18
2.5	УМО как оператор проектирования. Условная дисперсия	21
2.6	Регулярные варианты условных распределений	25
2.7	Теорема о монотонном классе. Применение к УМО	28
3	Гауссовская регрессия и условные гауссовские распределения	31
3.1	Гауссовская регрессия	31
3.2	Условные гауссовские распределения	35
3.3	Рекуррентное моделирование гауссовских векторов	36
	Список литературы	37

Понятия условных математических ожиданий и условных распределений, несомненно, входят в круг минимально развитого вероятностного образования. С другой стороны, при первом знакомстве эти понятия могут показаться слишком абстрактными. Чтобы несколько смягчить эту неприятность, мы начнем с более простых и интуитивно очевидных вещей.

1 Элементарные условные распределения

На практике часто встречается ситуация, когда нам хочется узнать какую-то информацию о распределении случайной величины ξ_1 , в то время как доступна для непосредственного наблюдения (измерения) другая случайная величина ξ_2 .

Если ξ_2 имеет дискретное распределение и принимает значения y_i с положительными вероятностями p_i , то легко определить понятие *условного распределения ξ_1 при условии $\xi_2 = y_i$* равенством¹

$$P(\xi_1 \in A | \xi_2 = y_i) = \frac{P(\xi_1 \in A, \xi_2 = y_i)}{P(\xi_2 = y_i)}, \quad A \in \mathcal{B}_R.$$

Обозначим левую часть этого равенства $\mathcal{Q}(A, y_i)$. Из свойств условных вероятностей сразу же следует, что при любом y_i условное распределение $\mathcal{Q}(\cdot, y_i)$ является вероятностной мерой. Поэтому можно говорить об *условном математическом ожидании случайной величины $f(\xi_1)$ при условии $\xi_2 = y_i$* :

$$E(f(\xi_1) | \xi_2 = y_i) = \int_{\mathbb{R}} f(x) \mathcal{Q}(dx, y_i) = \frac{E(f(\xi_1), \xi_2 = y_i)}{P(\xi_2 = y_i)},$$

если только этот интеграл существует. Другие понятия, связанные с распределениями случайных величин и векторов, также легко переносятся на условные распределения, если случайная величина ξ_2 дискретна.

Возникает вопрос — как ввести понятие условного распределения ξ_1 при условии $\xi_2 = y$ в том случае (например), когда $P(\xi_2 = y) = 0$ для любого $y \in \mathbb{R}$? Строгий и полный ответ формулируется в терминах так называемых условных математических ожиданий относительно σ -алгебр и отображений. Это особая теория, и мы ее не будем пока трогать.

Если же у случайных величин ξ_1, ξ_2 есть совместная плотность распределения, ситуация сильно упрощается. Ею мы и будем заниматься.

1.1 Основные формулы

Дальнейшие построения приведены для случая, когда обе случайные величины ξ_1, ξ_2 принимают значения в $\mathbb{R}$. Это ограничение введено только для удобства записи, все формулы сохранятся (с очевидными изменениями), если ξ_1 и ξ_2 — случайные вектора, принимающие значения в евклидовых пространствах произвольных (не обязательно равных) размерностей. Единственное, что требуется — это существование совместной плотности распределения у ξ_1 и ξ_2 .

Определение 1.1. Пусть случайные величины ξ_1, ξ_2 обладают совместной плотностью распределения $p_{\xi_1 \xi_2}(x, y) = p(x, y)$. Обозначим $p_{\xi_2}(x) = \int p(x, y) dy$ — плотность распределения случайной величины ξ_2 . Тогда функция

$$p(x|y) = p_{\xi_1|\xi_2}(x|y) = \begin{cases} \frac{p(x, y)}{p_{\xi_2}(y)} & \text{при } p_{\xi_2}(y) \neq 0, \\ p_{\xi_1}(x) & \text{при } p_{\xi_2}(y) = 0 \end{cases}$$

называется *условной плотностью распределений ξ_1 при условии $\xi_2 = y$* .

¹Конечно, здесь существенно предположение о том, что $P(\xi_2 = y_i) > 0$ при всех y_i .

Замечание 1. Функция $p(\cdot | y)$ является плотностью некоторого распределения при любом $y \in \mathbb{R}$. Действительно, при тех y , при которых $p_{\xi_2}(y) = 0$, это очевидно, а при остальных y

$$\int_{\mathbb{R}} \frac{p(x, y)}{p_{\xi_2}(y)} dx = \frac{p_{\xi_2}(y)}{p_{\xi_2}(y)} = 1.$$

Определение 1.2. Вероятностная мера

$$\mathcal{Q}(A, y) = P(\xi_1 \in A | \xi_2 = y) \stackrel{\text{def}}{=} \int_A p(x|y) dx, \quad A \in \mathcal{B}_R, \quad (1.1.1)$$

называется *условным распределением ξ_1 при условии $\xi_2 = y$* , а интеграл

$$E(f(\xi_1) | \xi_2 = y) \stackrel{\text{def}}{=} \int_{\mathbb{R}} f(x) \mathcal{Q}(dx, y) = \int_{\mathbb{R}} f(x) p(x|y) dx \quad (1.1.2)$$

(в предположении, что правая часть последнего равенства существует) — *условным математическим ожиданием случайной величины $f(\xi_1)$ при условии $\xi_2 = y$* .

Соответствующим образом условная дисперсия $D(f(\xi_1) | \xi_2 = y)$ определяется как²

$$\begin{aligned} D(f(\xi_1) | \xi_2 = y) &\stackrel{\text{def}}{=} \int_{\mathbb{R}} \left(f(x) - (E(f(\xi_1) | \xi_2 = y)) \right)^2 \mathcal{Q}(dx, y) = E(f^2(\xi_1) | \xi_2 = y) - (E(f(\xi_1) | \xi_2 = y))^2 = \\ &= \int_{\mathbb{R}} f^2(x) p(x|y) dx - \left(\int_{\mathbb{R}} f(x) p(x|y) dx \right)^2. \end{aligned} \quad (1.1.3)$$

Замечание 2. Поясним определение (1.1.1). Если для любого сколь угодно малого $h > 0$ вероятность $P(\xi_2 \in (y - h, y + h))$ положительна, то мы по определению условной вероятности получим, что

$$P(\xi_1 \in A | \xi_2 \in (y - h, y + h)) = \frac{P(\xi_1 \in A, \xi_2 \in (y - h, y + h))}{P(\xi_2 \in (y - h, y + h))}.$$

Естественно ожидать, что

$$P(\xi_1 \in A | \xi_2 = y) = \lim_{h \rightarrow 0} P(\xi_1 \in A | \xi_2 \in (y - h, y + h)),$$

если только предел в правой части существует. Покажем, что в «хорошем» случае это действительно так. Действительно,

$$P(\xi_1 \in A | \xi_2 \in (y - h, y + h)) = \frac{\int_A dx \int_{y-h}^{y+h} dt p(x, t)}{\int_{y-h}^{y+h} p_{\xi}(t) dt} = \frac{\int_A dx \frac{1}{2h} \int_{y-h}^{y+h} dt p(x, t)}{\frac{1}{2h} \int_{y-h}^{y+h} p_{\xi_2}(t) dt}. \quad (1.1.4)$$

Если плотность $p_{\xi_2}(t)$ непрерывна и положительна в окрестности точки y , то при $h \rightarrow 0$

$$\frac{1}{2h} \int_{y-h}^{y+h} p_{\xi_2}(t) dt \rightarrow p_{\xi_2}(y)$$

при $h \rightarrow 0$. Аналогично, легко привести условия на совместную плотность $p(x, t)$, при которых

$$\frac{1}{2h} \int_{y-h}^{y+h} dt p(x, t) \rightarrow p(x, y), \quad h \rightarrow 0,$$

²Проверьте второе равенство в этой цепочке.

причем этот предельный переход можно совершать под знаком интеграла, стоящего в числителе правой части (1.1.4).

Таким образом, «при благоприятных условиях»

$$\lim_{h \rightarrow 0} P(\xi_1 \in A \mid \xi_2 \in (y - h, y + h)) = \frac{\int_A dx p(x, y)}{p_{\xi_2}(y)} = \int_A p(x|y) dx,$$

что и соответствует формуле (1.1.1).

Свойства условных распределений.

1. Если ξ_1 и ξ_2 независимы, то $p_{\xi_1|\xi_2}(x|y) = p_{\xi_1}(x)$. Это равенство непосредственно следует из Определения 1.1 и из того факта, что $p(x, y) = p_{\xi_1}(x)p_{\xi_2}(y)$ для независимых ξ_1 и ξ_2 .
Отсюда сразу же вытекает, что $P(\xi_1 \in A \mid \xi_2 = y) = P(\xi_1 \in A)$ и $E(f(\xi_1) \mid \xi_2 = y) = Ef(\xi_1)$.
2. Имеет место «формула полной вероятности» для распределения случайной величины ξ_1 :

$$P(\xi_1 \in A) = \int_{\mathbb{R}} P(\xi_1 \in A \mid \xi_2 = y) p_{\xi_2}(y) dy. \quad (1.1.5)$$

Действительно, согласно (1.1.1)

$$\begin{aligned} \int_{\mathbb{R}} P(\xi_1 \in A \mid \xi_2 = y) p_{\xi_2}(y) dy &= \int_{\mathbb{R}} p_{\xi_2}(y) dy \int_A p(x|y) dx = \int_{y: p_{\xi_2}(y) \neq 0} p_{\xi_2}(y) dy \int_A p(x|y) dx = \\ &= \int_{y: p_{\xi_2}(y) \neq 0} p_{\xi_2}(y) dy \int_A \frac{p(x, y)}{p_{\xi_2}(y)} dx = \int_A dx \int_{y: p_{\xi_2}(y) \neq 0} dy p(x, y) = \\ &= \int_A dx \int_{\mathbb{R}} dy p(x, y) = \int_A p_{\xi_1}(x) dx \end{aligned} \quad (1.1.6)$$

Из (1.1.5) сразу же вытекает аналогичная формула для математических ожиданий:

$$Ef(\xi_1) = \int_{\mathbb{R}} E(f(\xi_1) \mid \xi_2 = y) p_{\xi_2}(y) dy. \quad (1.1.7)$$

3. Приведем несколько более общую формулу для условных математических ожиданий:³

$$E(g(\xi_1, \xi_2) \mid \xi_2 = y) = E(g(\xi_1, y) \mid \xi_2 = y). \quad (1.1.8)$$

Ограничимся «показательством» этого утверждения в стиле Замечания 2 (знак $\stackrel{\bullet}{=}$ означает здесь «ожидаемое равенство»). При фиксированном y

$$\begin{aligned} E(g(\xi_1, \xi_2) \mid \xi_2 = y) &\stackrel{\bullet}{=} \lim_{h \rightarrow 0} E(g(\xi_1, \xi_2) \mid \xi_2 \in [y - h, y + h]) = \\ &= \lim_{h \rightarrow 0} \frac{E(g(\xi_1, \xi_2), \xi_2 \in [y - h, y + h])}{P(\xi_2 \in [y - h, y + h])} = \lim_{h \rightarrow 0} \frac{\int_{\mathbb{R}} dx \frac{1}{2h} \int_{y-h}^{y+h} g(x, t) p(x, t) dt}{\frac{1}{2h} \int_{y-h}^{y+h} p_{\xi_2}(t) dt} \stackrel{\bullet}{=} \\ &\stackrel{\bullet}{=} \frac{\int_{\mathbb{R}} dx g(x, y) p(x, y)}{p_{\xi_2}(y)} = \int_{\mathbb{R}} g(x, y) p(x|y) dx = E(g(\xi_1, y) \mid \xi_2 = y), \end{aligned}$$

причем при последнем переходе используется равенство (1.1.3) с заменой $f(x)$ на $f_y(x) = g(x, y)$ при фиксированном y .

³Интуитивно формула (1.1.8) совершенно очевидна: действительно, если мы знаем, что $\xi_2 = y$, то что нам может помешать заменить ξ_2 на y в выражении, стоящем перед значком условия?

4. Полагая $g(x, y) = \mathbf{1}_B(x, y)$ при $B \in \mathcal{B}_R$, получим из (1.1.8), что

$$P((\xi_1, \xi_2) \in B | \xi_2 = y) = P((\xi_1, y) \in B | \xi_2 = y), \quad (1.1.9)$$

что является обобщением (1.1.1).

5. Если случайные величины ξ_1, ξ_2 независимы, то равенства (1.1.8) и (1.1.9) переходят соответственно в

$$E(g(\xi_1, \xi_2) | \xi_2 = y) = Eg(\xi_1, y) \quad \text{и} \quad P((\xi_1, \xi_2) \in B | \xi_2 = y) = P((\xi_1, y) \in B).$$

6. Аналогично формулам (1.1.5) и (1.1.7),

$$P((\xi_1, \xi_2) \in B) = \int_{\mathbb{R}} P((\xi_1, y) \in B | \xi_2 = y) p_{\xi_2}(y) dy \quad (1.1.10)$$

и

$$Eg(\xi_1, \xi_2) = \int_{\mathbb{R}} E(g(\xi_1, y) | \xi_2 = y) p_{\xi_2}(y) dy. \quad (1.1.11)$$

7. Если случайные величины ξ_1, ξ_2 независимы, то равенства (1.1.10) и (1.1.11) переходят соответственно в

$$P((\xi_1, \xi_2) \in B) = \int_{\mathbb{R}} P((\xi_1, y) \in B) p_{\xi_2}(y) dy \quad \text{и} \quad Eg(\xi_1, \xi_2) = \int_{\mathbb{R}} E(g(\xi_1, y)) p_{\xi_2}(y) dy.$$

1.2 Упражнения

Некоторые из предложенных задач могут быть решены различными способами. Здесь предполагается, что они должны решаться с помощью понятий, связанных с условными плотностями.

1. Найти распределение суммы двух независимых случайных величин, если известны их плотности. (Получить формулу свертки с помощью условных распределений.)
2. На отрезок $[0, 1]$ бросают точку, а потом на левый из образовавшихся отрезков снова бросают точку. Найти распределение ее координаты.
3. Вектор (ξ_1, ξ_2) равномерно распределен в единичном круге с центром в нуле. Найти условное распределение ξ_1 при условии, что $\xi_2 = y$.
4. Вектор (ξ, η) равномерно распределен в области $D \subset \mathbb{R}^2$. Найти условное распределение ξ при условии $\eta = y$.
5. Случайные величины $\xi_1, \dots, \xi_n \in U(0, 1)$ и независимы. Доказать, что случайная величина $\eta_n = \prod_{i=1}^n \xi_i$ имеет плотность $p_n(x) = (-\ln x)^{n-1}/(n-1)!$
6. Случайные величины ξ_1 и ξ_2 независимы и показательно распределены с параметром λ . Доказать, что $\xi_1 + \xi_2$ и ξ_1/ξ_2 независимы.
7. На отрезок $[0, 1]$ бросают точку (α — ее координата). После этого проводят n испытаний Бернулли с вероятностью успеха α . Найти распределение числа успехов.
8. Семейство случайных величин $\{\xi_\lambda, \lambda > 0\}$ обладает тем свойством, что ξ_λ показательно распределено с параметром λ . Случайная величина η не зависит от семейства и показательно распределена с параметром 1. Найти распределение случайной величины ξ_η .
9. Случайные величины $\xi_1, \xi_2 \in U(0, 1)$ и независимы. Обозначим $\xi_{[1]} = \min(\xi_1, \xi_2)$ и $\xi_{[2]} = \max(\xi_1, \xi_2)$. Найти а) условное распределение $\xi_{[1]}$ при условии $\xi_{[2]} = y$ и б) условное распределение ξ_1 при условии $\xi_{[2]} = y$.
10. На отрезок $[0, 1]$ бросают точку, а потом на правый из образовавшихся отрезков снова бросают точку. Найти распределение ее координаты.
11. На отрезок $[0, 1]$ бросают точку, а потом на меньший из образовавшихся отрезков снова бросают точку. Найти распределение ее координаты.
12. На отрезок $[0, 1]$ бросают точку, а потом и на левый и на правый из образовавшихся отрезков снова бросают по точке. Найти совместное распределение их координат.
13. Случайная величина ξ показательно распределена с параметром 1. На отрезок $[0, 1/\xi]$ бросают точку. Найти распределение ее координаты.
14. Пусть $\xi_\lambda \in U(0, 1)$ при $\lambda > 0$, а случайная величина $\eta \in \text{EXP}(1)$ и не зависит от $\{\xi_\lambda\}$. Найти распределение с.в. ξ_η .
15. Пусть $\xi \in U(0, 1)$. На отрезок $[0, 1/\xi]$ бросают точку. Найти распределение ее координаты.
16. Пусть $\xi_\lambda \in \Pi(\lambda)$ при любом $\lambda > 0$. Случайная величина η не зависит от всех ξ_λ и имеет показательное распределение с параметром 1. Найти распределение с.в. ξ_η .
17. Пусть $\xi_\lambda \in \Pi(\lambda)$ при любом $\lambda > 0$. Случайная величина η не зависит от всех ξ_λ и имеет гамма-распределение с параметрами (k, μ) , k — целое. Найти (и идентифицировать) распределение с.в. ξ_η .

18. В единичный круг с центром в нуле бросают точку. Проводят окружность с центром в нуле через эту точку. Потом в образовавшееся кольцо снова бросают точку. Найти совместное распределение ее полярных координат.
19. В квадрат $[0, 1]^2$ бросают точку $\xi = (\xi_1, \xi_2)$. После этого в прямоугольник $[0, \xi_1] \times [0, \xi_2]$ бросают точку $\bar{\eta}$. Найти распределение ее координат. Являются ли они независимыми?
20. Пусть ξ_1, ξ_2 независимы и имеют распределение $N(0, 1)$. Доказать, что $\xi_1 + \xi_2$ и $\xi_1 - \xi_2$ независимы.
21. Найти распределение $\xi_1/(\xi_1 + \xi_2)$, если ξ_1 и ξ_2 независимы, одинаково распределены и а) имеют показательное распределение с λ , б) равномерно распределены на $[0, 1]$, в) имеют $N(0, 1)$ -распределение.
22. Найти условное распределение $\mathcal{L}(\xi_1 | \xi_1 + \xi_2 = y)$, если ξ_1 и ξ_2 независимы, одинаково распределены и а) имеют показательное распределение с λ , б) равномерно распределены на $[0, 1]$, в) имеют $N(0, 1)$ -распределение.
23. Пусть $\xi_1, \dots, \xi_n \in U(0, 1)$ и независимы. Положим $\eta_1 = \min(\xi_1, \dots, \xi_n)$, $\eta_2 = \max(\xi_1, \dots, \xi_n)$. Найти условное распределение η_1 при условии, что $\eta_2 = y$.
24. На отрезок $[0, 1]$ бросают 2 точки, разбивающие его на 3 отрезка. После этого на центральный отрезок бросают еще точку (ее координата — η). Найти (именно в таком порядке) а) $E\eta$, $D\eta$; б) характеристическую функцию η ; в) распределение η .
25. На отрезок $[0, 1]$ бросают 2 точки, разбивающие его на 3 отрезка. После этого на правый отрезок бросают еще точку (ее координата — η). Найти распределение η .
26. Случайная величина ξ показательна распределена с $\lambda = 1$. Условное распределение η при условии $\xi = x$ есть $U(-1/x, 1/x)$. Найти распределение η .
27. Пусть $\xi_1, \xi_2 \in N(0, 1)$ и независимы. Найти условное распределение $\eta = 2\xi_1 - \xi_2$ при условии $\xi_1 + \xi_2 = x$.
28. Пусть $\xi_1, \xi_2 \in U(0, 1)$ и независимы. Найти условное распределение $\eta = 2\xi_1 - \xi_2$ при условии $\xi_1 + \xi_2 = x$.
29. Пусть $\xi_1, \xi_2 \in N(0, 1)$ и независимы. Найти распределение $\eta = \xi_1 + \alpha\xi_2$, где $\alpha \in U(0, 1)$ и не зависит от ξ_1, ξ_2 .
30. Пусть $\xi_1, \xi_2, \xi_3 \in N(0, 1)$ и независимы. Найти совместное распределение $\xi_1 + \xi_2$ и $\xi_2 + \xi_3$.
31. Случайный вектор (ξ_1, ξ_2) равномерно распределен в единичном круге с центром в нуле. Если $\sqrt{\xi_1^2 + \xi_2^2} = x$, то $\eta \in \text{EXP}(x)$. Найти распределение η .
32. Пусть $\beta_1, \beta_2 \in U(0, 1)$ и независимы, $\eta = \max(\beta_1, \beta_2)$. Имея реализацию $\eta = x$, моделируем $\xi \in \text{EXP}(x)$. Найти распределение ξ .
33. Пусть $\beta_1, \beta_2 \in \text{EXP}(1)$ и независимы, $\eta = \min(\beta_1, \beta_2)$. Если $\eta = a$, то $\xi \in U(0, a)$. Найти $D\xi$.
34. Пусть $\xi \in \text{EXP}(1)$. Если $\xi = a$, то $\eta \in N(a, 1)$. Найти распределение η .
35. Пусть $\xi \in U(0, \sqrt{2})$. В квадрат $(-\xi, \xi) \times (-\xi, \xi)$ бросают точку. Найти совместное распределение ее координат.

36. Вектор $\bar{\xi} = (\xi_1, \xi_2, \xi_3)^T$ равномерно распределен в единичном шаре с центром в нуле. В трехмерный шар радиуса $\sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}$ бросают точку. Найти распределение расстояния от этой точки до нуля.
37. Пусть $\beta_1, \beta_2 \in U(0, 1)$ и независимы. Найти распределение $\beta_1 + \beta_2$ при условии, что $\beta_1 - \beta_2 = x$.
38. Пусть $\beta_1, \beta_2, \beta_3 \in U(0, 1)$ и независимы. Найти распределение $\beta_1 + \beta_2$ при условии, что $\max(\xi_2, \xi_3) = x$.
39. Пусть $\beta_1, \beta_2 \in N(0, 1)$ и независимы. Найти совместное распределение (β_1, β_2) при условии, что $\xi_1^2 + \xi_2^2 = x$.
40. Сначала бросают точку на отрезок $(0, 1)$. Если результат бросания имеет координату x , то моделируют случайную величину $\xi \in \text{EXP}(x)$. Найти распределение ξ .
41. Пусть $\alpha \in U(0, 1)$. На отрезок $(0, 1/\alpha)$ бросают точку ξ . Найти распределение ξ .
42. Случайные величины ξ_s, η_t независимы при любых $s, t > 0$, причем распределения ξ_s и η_s совпадают и имеют плотность

$$p_s(x) = \begin{cases} \exp(-x + s) & \text{при } x \geq s, \\ 0 & \text{иначе.} \end{cases}$$

Случайные величины $\alpha_1, \alpha_2 \in U(0, 1)$ и независимы как между собой, так и от ξ_s, η_t . Найти распределение случайной величины $\xi_{\alpha_1} + \xi_{\alpha_2}$.

43. Пусть $\beta_1, \beta_2 \in U(0, 1)$ и независимы. Найти распределение случайной величины

$$\xi = \frac{\beta_1}{a\beta_1 + b\beta_2}, \quad a, b > 0.$$

44. Пусть $\xi_\lambda \in \text{EXP}(\lambda)$, $\lambda > 0$. Случайная величина η не зависит от $\{\xi_\lambda, \lambda > 0\}$ и имеет плотность распределения $2x$, $0 < x < 1$. Найти распределение случайной величины ξ_η .
45. Вектор $(\alpha_1, \alpha_2)^T$ равномерно распределен в квадрате $[0, 1]^2$. Найти распределение случайного вектора $(\eta_1, \eta_2)^T$, равномерно распределенного в квадрате $[0, \max(\alpha_1, \alpha_2)]^2$.
46. При любом фиксированном $x > 0$ случайные величины $\xi_x, \eta_x \in \text{EXP}(x)$ и независимы. Случайная величина $\beta \in \text{EXP}(1)$ и не зависит от $\{\xi_x, \eta_x\}$, $x > 0$. Найти совместное распределение $(\xi_\beta, \xi_\beta + \eta_\beta)$.
47. Пусть $\beta_1, \beta_2 \in U(0, 1)$ и независимы. Обозначим

$$\xi_1 = \beta_1 + \beta_2, \quad \xi_2 = 2\beta_1 - 2\beta_2.$$

Найти условное распределение ξ_1 при условии $\xi_1 + \xi_2 = t$.

48. Пусть $\beta \in U(0, 1)$. Если $\beta = x$, то в круг радиуса x с центром в нуле бросают точку. Найти совместное распределение ее координат.
49. Точку β_1, β_2 бросают в единичный квадрат. После этого в прямоугольник со сторонами $\beta_1, \beta_1 + \beta_2$ бросают точку. Найти совместное распределение ее координат.
50. Пусть $\beta_1, \beta_2 \in U(0, 1)$ и независимы. Найти условное распределение β_1 при условии а) $\beta_1 + \beta_2 = x$, б) $\beta_1 - \beta_2 = x$.

51. Случайные величины ξ_1, ξ_2, α независимы, причем $\xi_1, \xi_2 \in N(0, 1)$, а $\alpha \in U(0, 1)$. Найти распределение случайной величины

$$\eta = \frac{\xi_1 + \alpha \xi_2}{\sqrt{1 + \alpha^2}}.$$

52. Пусть $\xi \in U(-1, 1)$. Условное распределение η при условии $\xi = x$ есть $N(x, 1)$. Найти $P(\eta < 0)$, $E\eta$ и $D\eta$.
53. Случайные величины $\xi_p \in \text{Geom}(p)$, $0 < p < 1$. С.в. $\alpha \in U(0, 1)$ и не зависит от $\{\xi_p\}$. Найти распределение $\eta = \xi_\alpha$.
54. Пусть $\alpha \in U(0, 1)$. Производят испытания Бернулли с вероятностью успеха α до 2-го успеха. Найти распределение числа испытаний.
55. Пусть $\alpha \in U(0, 1)$. Производят испытания Бернулли с вероятностью успеха α^2 до 1-го успеха. Найти распределение числа испытаний.
56. Пусть $\xi \in \text{Geom}(p)$, а η имеет показательное распределение с параметром $\xi + 1$. Найти распределение η .
57. Пусть $\beta_1, \beta_2 \in U(0, 1)$ и независимы. Найти совместное распределение (β_1, β_2) при условии, что $\max(\beta_1, \beta_2) = x$.
58. Пусть $\beta_1, \beta_2 \in \text{EXP}(1)$ и независимы. Найти условное распределение $\max(\beta_1, \beta_2)$ при условии, что $\min(\beta_1, \beta_2) = x$.
59. Пусть $\beta_1, \beta_2 \in U(0, 1)$ и независимы. Найти распределение β_1 при условии, что $\max(\beta_1, \beta_2) = x$.
60. Случайный вектор $\bar{\xi}_1$ р.р. в квадрате с вершинами $(-1, 0), (-1, 1), (0, 0), (0, 1)$. Случайный вектор $\bar{\xi}_2$ р.р. в квадрате с вершинами $(0, 0), (0, 1), (1, 1), (1, 0)$. Случайный вектор $\bar{\xi}_3$ р.р. в квадрате с вершинами $(-1, -1), (-1, 0), (0, 0), (0, -1)$. Случайный вектор $\bar{\xi}_4$ р.р. в квадрате с вершинами $(0, -1), (0, 0), (1, 0), (1, -1)$.
Пусть $\alpha \in U(0, 1)$ и не зависит от $(\bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3, \bar{\xi}_4)$. Найти распределение вектора $\bar{\xi}_\eta$, где $\eta = \lfloor 4\alpha^2 + 1 \rfloor$.
61. Пусть $\beta_1, \beta_2 \in N(0, 1)$ и независимы. Найти распределение

$$\xi = \frac{\beta_1 \beta_2}{\beta_1^2 + \beta_2^2}.$$

62. Пусть $\xi \in N(0, 1)$. При $a \in \mathbb{R}$ случайные величины $\eta_a \in N(a, 1)$ и не зависят от ξ . Найти распределение η_ξ .
63. Случайные вектора $\bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3$ независимы и р.р. на единичной окружности с центром в нуле. Найти вероятность того, что хорды с концами $((0, 1)^T, \bar{\xi}_1)$ и $(\bar{\xi}_2, \bar{\xi}_3)$ пересекаются.
64. Пусть $\beta_1, \beta_2 \in N(0, 1)$ и независимы. В круг радиуса $1/\sqrt{\beta_1^2 + \beta_2^2}$ с центром в нуле бросают точку. Найти распределение ее расстояния до центра.
65. Пусть $\xi \in U(0, 1)$. В квадрат $(0, 1/\xi) \times (0, 1/\xi)$ бросают точку. Найти совместное распределение ее координат.
66. В круг радиуса 1 с центром в нуле бросают точку $\bar{\xi} = (\xi_1, \xi_2)$. После этого в круг радиуса $\sqrt{\xi_1^2 + \xi_2^2}$ с центром в нуле бросают точку $\bar{\eta}$. Найти вероятность того, что $\bar{\eta}$ отстоит от центра меньше, чем на 0.5 и находится в 1-й четверти.

67. Пусть $\xi_k \in \text{EXP}(k)$, $k = 1, 2, \dots$. $\eta \in \text{GEOM}(p)$ и не зависит от ξ_k , $k \geq 1$. Найти распределение случайной величины $\xi_{\eta+1}$.
68. Пусть $\xi_1, \xi_2 \in N(0, 1)$ и независимы. Кроме того, при $\lambda > 0$ случайные величины $\beta_\lambda, \eta_\lambda \in \text{EXP}(\lambda)$ независимы между собой и от ξ_1, ξ_2 . Найти распределение $\beta_{|\xi_1|} + \eta_{|\xi_2|}$.
69. Случайный вектор $\bar{\xi} = (\xi_1, \xi_2)^T$ гауссовский с нулевым средним, единичными дисперсиями и коэффициентом корреляции ρ . Найти условное распределение ξ_1 при условии $\xi_1 + \xi_2 = x$.

2 Условные математические ожидания относительно σ -алгебр

2.1 Определение и простейшие свойства

Определение 2.1. Пусть $(\Omega, \mathcal{F}, P)$ — некоторое вероятностное пространство и $\mathcal{F}_0 \subset \mathcal{F}$ — σ -алгебра, содержащаяся в $\mathcal{F}$. Пусть, кроме того, $\xi : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B}_R)$ — случайная величина с конечным математическим ожиданием.

Случайная величина ξ' называется *условным математическим ожиданием*⁴ ξ относительно⁵ σ -алгебры $\mathcal{F}_0$, если, во-первых, ξ' измерима относительно $\mathcal{F}_0$ и, во-вторых, для любого $A \in \mathcal{F}_0$

$$\int_A \xi' dP = \int_A \xi dP. \quad (2.1.1)$$

Стандартным обозначением для УМО относительно σ -алгебры $\mathcal{F}_0$ является $\xi' = E(\xi | \mathcal{F}_0)$, иногда вместо этого обозначения используют более короткое $E_{\mathcal{F}_0}\xi$.

В случае, когда $\xi = \mathbb{I}_A$ с $A \in \mathcal{F}$, вместо $E(\xi | \mathcal{F}_0)$ пишут $P(A | \mathcal{F}_0)$ и говорят об *условной вероятности события A при условии $\mathcal{F}_0$* .

Докажем теорему существования и единственности условных математических ожиданий.

Теорема 1. Условное математическое ожидание $E(\xi | \mathcal{F}_0)$ существует и единственно P-п.в.

Доказательство. Нужное нам утверждение сразу же следует из теоремы Радона-Никодима. Действительно, рассмотрим вероятностное пространство⁶ $(\Omega, \mathcal{F}_0, P)$ и положим $\nu(A) = \int_A \xi dP$ при $A \in \mathcal{F}_0$.

Ясно, что ν является конечным зарядом и что $\nu \prec P$. Поэтому, согласно теореме Радона-Никодима,⁷ существует такая $\mathcal{F}_0$ -измеримая функция ξ' , что $\nu(A) = \int_A \xi' dP$. Из этой же теоремы следует, что ξ' единственна с точностью до $\mathcal{F}_0$ -измеримых множеств P-меры ноль. $\square$

Замечание 3. Отметим, что, в отличие от обычного математического ожидания, условное математическое ожидание является случайной величиной, то есть функцией $\omega \in \Omega$ (не обязательно постоянной). Более того, эта функция не является единственной, и поэтому можно говорить об УМО как о классе эквивалентных случайных величин (например, как об элементе $\mathbb{L}_{\mathcal{F}_0}^1 = \mathbb{L}^1(\mathcal{F}_0, P)$), а не как об одном каком-то представителе этого класса.

Приведем несколько простейших примеров УМО относительно различных σ -алгебр $\mathcal{F}_0$.

1. (минимальная σ -алгебра $\mathcal{F}_0$). Если $\mathcal{F}_0 = (\Omega, \emptyset)$, то $E(\xi | \mathcal{F}_0) = E\xi$. Действительно, любая постоянная измерима относительно $\mathcal{F}_0$, а при $A = \Omega$ или $A = \emptyset$, очевидно,

$$\int_A E\xi dP = \int_A \xi dP.$$

2. (максимальная σ -алгебра $\mathcal{F}_0$). Если $\mathcal{F}_0 = \mathcal{F}$, то P-п.в. $E(\xi | \mathcal{F}_0) = \xi$ для любой суммируемой случайной величины ξ .

Действительно, здесь ξ по определению измерима относительно $\mathcal{F}_0$, а условие (2.1.1) при подстановке $\xi' = \xi$ превращается в тождество.

⁴Сокращенно — УМО.

⁵или — «при условии»

⁶Мы сохраняем здесь обозначение P, хотя на самом деле речь идет о сужении распределения P с σ -алгебры $\mathcal{F}$ на σ -алгебру $\mathcal{F}_0$.

⁷для конечных зарядов

3. (дискретная σ -алгебра $\mathcal{F}_0$). Возьмем $\mathcal{F}_0 = (\Omega, A, A^C, \emptyset)$, где $A \in \mathcal{F}$ и $P(A), P(A^C) > 0$. Тогда⁸

$$E(\xi | \mathcal{F}_0)(\omega) = \begin{cases} \frac{1}{P(A)} \int_A \xi dP & \text{при } \omega \in A, \\ \frac{1}{P(A^C)} \int_{A^C} \xi dP & \text{при } \omega \in A^C. \end{cases}$$

Приведем несколько простейших свойств УМО. Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство, $\mathcal{F}_0 \subset \mathcal{F}$ — σ -алгебра и ξ, η — случайные величины с конечными математическими ожиданиями.

Предложение 1. (Простейшие свойства УМО).

1. Если $\xi = c = \text{const}$, то $E(\xi | \mathcal{F}_0) = c$ почти всюду по мере P .
2. Для любых постоянных a, b

$$E(a\xi + b\eta | \mathcal{F}_0) = aE(\xi | \mathcal{F}_0) + bE(\eta | \mathcal{F}_0) \quad P\text{-п.в.} \quad (2.1.2)$$

3. Если $\xi \geq 0$, то $E(\xi | \mathcal{F}_0) \geq 0$ с вероятностью 1.
4. Если $\xi \geq \eta$, то $E(\xi | \mathcal{F}_0) \geq E(\eta | \mathcal{F}_0)$ с вероятностью 1.
5. $E(|\xi| | \mathcal{F}_0) \geq |E(\xi | \mathcal{F}_0)|$ с вероятностью 1.
6. $E(E(\xi | \mathcal{F}_0)) = E\xi$.

Доказательство. 1. Постоянная c , конечно же, измерима относительно $\mathcal{F}_0$. Постановка $\xi' = \xi = c$ в (2.1.1) завершает доказательство.

2. Обозначим $\zeta = a\xi + b\eta$ и $\zeta' = aE(\xi | \mathcal{F}_0) + bE(\eta | \mathcal{F}_0)$. Тогда при $A \in \mathcal{F}_0$

$$\int_A \zeta' dP = a \int_A E(\xi | \mathcal{F}_0) dP + b \int_A E(\eta | \mathcal{F}_0) dP = a \int_A \xi dP + b \int_A \eta dP = \int_A \zeta dP.$$

Осталось заметить, что ζ' измерима относительно $\mathcal{F}_0$.

3. Обозначим $\xi' = E(\xi | \mathcal{F}_0)$ и положим $A = \{\omega : \xi' < 0\}$. Пусть $P(A) > 0$. Тогда

$$0 > \int_A \xi' dP = \int_A \xi dP,$$

что противоречит неотрицательности ξ .

4. Доказательство сводится к рассмотрению случайной величины $\beta = \xi - \eta$ и последовательному применению пунктов 3 и 2 настоящего Предложения.

5. Следует из того, что $|\xi| \geq \xi$ и $|\xi| \geq -\xi$ с дальнейшим применением п. 4 Предложения.

6. Непосредственно сводится к (2.1.1) выбором $A = \Omega$. $\square$

Замечание 4. Отметим, что пункты 1 – 5 Предложения 1 аналогичны соответствующим свойствам математических ожиданий, то есть интегралов. В дальнейшем мы увидим, что это не является случайным совпадением. Шестой пункт, конечно, имеет другую природу.

2.2 УМО относительно отображений и функция регрессии

Вернемся к определению условных математических ожиданий. Рассмотрим вероятностное пространство $(\Omega, \mathcal{F}, P)$ и суммируемую случайную величину ξ , определенную на этом пространстве. Кроме того, введем отображение $T : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{A})$, где $(D, \mathcal{A})$ — некоторое измеримое пространство.

Отображение T порождает σ -алгебру $\sigma(T) \subset \mathcal{F}$, которая определяется равенством $\sigma(T) = \{T^{-1}B, B \in \mathcal{A}\}$. Выбрав $\mathcal{F}_0 = \sigma(T)$, можно рассматривать УМО $E(\xi | \sigma(T))$ относительно σ -алгебры $\sigma(T)$, которое обозначается $E(\xi | T)$ и называется *условным математическим ожиданием ξ относительно отображения T* .

⁸Проверьте! А что изменится, если, например, $P(A) = 0$?

Замечание 5. Заметим, что любая σ -алгебра $\mathcal{F}_0 \subset \mathcal{F}$ порождается измеримым отображением $T : (\Omega, \mathcal{F}) \mapsto (\Omega, \mathcal{F}_0)$ таким, что $T(\omega) = \omega$. Поэтому понятия УМО относительно σ -алгебры и УМО относительно отображения эквивалентны.

Обозначение $E(\xi | T)$ обретает содержательный смысл благодаря следующему утверждению.

Лемма 1. Пусть $(D, \mathcal{A})$ — некоторое измеримое пространство, $T : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{A})$ и σ -алгебра $\mathcal{F}_0 = \sigma(T)$ порождена отображением T . Обозначим $\mathcal{P}_T$ распределение случайной величины T .

1. Для того, чтобы случайная величина $\xi : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B}_R)$ была измерима относительно $\sigma(T)$, необходимо и достаточно, чтобы существовала такая функция $g : (D, \mathcal{A}) \mapsto (\mathbb{R}, \mathcal{B}_R)$, что $\xi = g \circ T$.

2. Пусть $\xi = g \circ T$, $\xi_1 = g_1 \circ T$ и $\xi_1 = \xi$ п.в. по мере P . Тогда $g_1 = g$ п.в. по мере $\mathcal{P}_T$.

Доказательство. 1. В части достаточности утверждение очевидно. Для доказательства необходимости предположим сначала, что ξ является простой $\sigma(T)$ -измеримой функцией, то есть что

$$\xi = \sum_{i=1}^n c_i \mathbb{I}_{A_i},$$

где $A_i \in \sigma(T)$ и $A_i \cap A_j = \emptyset$ при $i \neq j$. Тогда $A_i = T^{-1}B_i$, где $B_i \in \mathcal{A}$, причем множества B_i можно выбрать попарно непересекающимися.⁹

Положим $B = \cup_i B_i$ и определим функцию g равенством

$$g(t) = \begin{cases} c_i & \text{при } t \in B_i, \\ 0 & \text{при } t \in D \setminus B. \end{cases}$$

Тогда

$$\xi(\omega) = \sum_{i=1}^n c_i \mathbb{I}_{A_i}(\omega) = \sum_{i=1}^n c_i \mathbb{I}_{T^{-1}B_i}(\omega) = \sum_{i=1}^n c_i \mathbb{I}_{B_i}(T\omega) = g(T\omega).$$

Перейдем к общему случаю. Представим ξ в виде поточечного предела простых $\sigma(T)$ -измеримых случайных величин ξ_n , каждая из которых согласно доказанному имеет вид $\xi_n = g_n \circ T$.

Положим

$$D_0 = \{t \in D : \text{существует конечный предел } \lim_n g_n(t)\}$$

и определим функцию $g : (D, \mathcal{A}) \mapsto (\mathbb{R}, \mathcal{B}_R)$ равенством¹⁰

$$g(t) = \begin{cases} \lim_n g_n(t) & \text{при } t \in D_0, \\ 0 & \text{иначе.} \end{cases}$$

Так как при любом ω последовательность $g_n(T\omega)$ имеет предел, равный $\xi(\omega)$, то $T\omega \in D_0$ для любого $\omega \in \Omega$. Поэтому $\xi(\omega) = \lim_n g_n(T\omega) = g(T\omega)$.

2. Второе утверждение леммы вытекает из цепочки равенств

$$\int_B g d\mathcal{P}_T = \int_{T^{-1}B} \xi dP = \int_{T^{-1}B} \xi_1 dP = \int_B g_1 d\mathcal{P}_T,$$

выполняющейся для любого $B \in \mathcal{A}$. □

⁹Проверьте!

¹⁰Функция g является измеримой, поскольку предел измеримых функция измерим.

Пример 1. Пусть $(\Omega, \mathcal{F}) = (D, \mathcal{A}) = (\mathbb{R}, \mathcal{B}_R)$. Положим $T(\omega) = |\omega|$. Ответьте на следующие вопросы: а) Как устроена σ -алгебра $\sigma(T)$? б) Какие функции являются измеримыми относительно этой σ -алгебры? в) Очевидно ли Вам в этом случае первое утверждение Леммы 1?

Согласно Лемме 1 условное математическое ожидание случайной величины ξ относительно отображения $T : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{A})$ имеет вид

$$E(\xi | T)(\omega) = g(T\omega),$$

где функция $g : (D, \mathcal{A}) \mapsto (\mathbb{R}, \mathcal{B}_R)$ называется *функцией регрессии* ξ на T и обычно обозначается как $g(t) = E(\xi | T = t)$.

Замечание 6. Это обозначение имеет прозрачную интерпретацию. Предположим, что нас интересует совместное поведение результатов двух измерений величин ξ и T в некотором случайном эксперименте, причем T доступно непосредственному наблюдению, а ξ — нет.

Тогда величину $E(\xi | T = t)$ можно понимать как среднее значение ξ при условии, что T приняло значение t (см. ниже Пример 2). Такая интерпретация часто делает естественными более сложные свойства УМО.

Определение 2.1 и Теорема 1 для случая $\mathcal{F}_0 = \sigma(T)$ легко переформулируются на языке функции регрессии. Действительно, так как при $B \in \mathcal{A}$

$$\int_{T^{-1}B} g(T) dP = \int_B g dP_T,$$

то (2.1.1) превращается в равенство

$$\int_{T^{-1}B} \xi dP = \int_B g dP_T, \quad (2.2.1)$$

а Теорема 1 (вместе со вторым пунктом Леммы 1) утверждает, что функция g , для которой (2.2.1) выполняется для любого $B \in \mathcal{A}$, существует и единственна $\mathcal{P}_T$ -почти всюду.

Таким образом, чтобы найти условное математическое ожидание $\xi' = E(\xi | T)$ относительно отображения T , достаточно найти такую функцию $g : (D, \mathcal{A}) \mapsto (\mathbb{R}, \mathcal{B}_R)$, для которой выполняются равенства (2.2.1), а затем положить $\xi' = g \circ T$.

Опишем две ситуации, когда можно явно выписать вид функции регрессии.

Пример 2. (Примеры вычисления функций регрессии)

1. (*Случай дискретного T*). Пусть $D = \{t_1, \dots, t_n, \dots\}$ является конечным или счетным множеством, а σ -алгебра $\mathcal{A}$ состоит из всех подмножеств множества D . Обозначим $A_i = \{\omega : T\omega = t_i\}$ и предположим дополнительно, что $P(A_i) > 0$. Тогда

$$g(t_i) = E(\xi | T = t_i) = \frac{1}{P(A_i)} \int_{A_i} \xi dP. \quad (2.2.2)$$

Для доказательства достаточно проверить равенство (2.2.1) для $B = \{t_i\}$. Тогда

$$\int_B g dP_T = g(t_i)P(A_i) = \int_{A_i} \xi dP,$$

что и требовалось доказать.

2. (*Случай абсолютно непрерывного (ξ, T)*). Пусть совместное распределение ξ, T имеет производную Радона-Никодима (иначе говоря, совместную плотность) $p(x, t)$ относительно меры $\mu(dx) \otimes \nu(dt)$, определенной на σ -алгебре $\mathcal{B}_R \times \mathcal{A}$ подмножеств множества $\mathbb{R} \times D$. Обозначим

$$p_T(t) = \int_{\mathbb{R}} p(x, t) \mu(dx).$$

Тогда

$$E(\xi | T = t) = \begin{cases} \frac{\int_{\mathbb{R}} xp(x, t)\mu(dx)}{p_T(t)} & \text{если } p_T(t) \neq 0, \\ 0 & \text{если } p_T(t) = 0. \end{cases} \quad (2.2.3)$$

Докажем равенство (2.2.3). Обозначим $D_0 = \{t : p_T(t) \neq 0\}$. Поскольку, как нетрудно видеть, $p_T = dP_T/d\nu$, то $P_T(D_0) = P(T \in D_0) = 1$.

Поэтому для любого $B \in \mathcal{A}$ интеграл по множеству B от правой части (2.2.3) имеет вид

$$\int_{B \cap D_0} \nu(dt) \int_{\mathbb{R}} xp(x, t)\mu(dx) = E(\xi, T \in B \cap D_0) = E(\xi, T \in B) = \int_{T^{-1}B} \xi dP.$$

Тем самым (2.2.3) доказано.¹¹

Утверждения Предложения 1 легко переформулируются в терминах функции регрессии. Например, вместо (2.1.2) мы будем иметь равенство.

$$E(a\xi + b\eta | T = t) = aE(\xi | T = t) + bE(\eta | T = t) \quad \mathcal{P}_{T\text{-п.в.}}$$

Аналогично переписываются неравенства пп. 3 – 5 Предложения. Что касается тождества шестого пункта, то оно приобретет вид¹²

$$E\xi = \int_D E(\xi | T = t) P_T(dt). \quad (2.2.4)$$

Приведем еще одно полезное утверждение о свойствах УМО, сформулировав его в терминах отображений. Пусть $\xi : (\Omega, \mathcal{F}) \mapsto (D_1, \mathcal{D}_1)$, $\beta : (\Omega, \mathcal{F}) \mapsto (D_2, \mathcal{D}_2)$, $D = D_1 \times D_2$ и $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2$.

Лемма 2. Если почти всюду $E(\eta | \xi, \beta) = \varphi(\xi)$, где φ — некоторая измеримая функция, то $\varphi(\xi) = E(\eta | \xi)$.

Доказательство. Возьмем произвольное множество $B \in \sigma(\xi)$. Ясно, что $B \in \sigma(\xi, \beta)$. Поэтому

$$\int_B E(\eta | \xi, \beta) dP = \int_B \eta dP.$$

С другой стороны,

$$\int_B E(\eta | \xi, \beta) dP = \int_B \varphi(\xi) dP.$$

Следовательно,

$$\int_B \eta dP = \int_B \varphi(\xi) dP,$$

и, поскольку $\varphi(\xi)$ измерима относительно $\sigma(\xi)$, осталось сослаться на определение УМО. $\square$

При изучении дальнейших свойств условных математических ожиданий мы будем использовать оба варианта записи УМО — через σ -алгебры и через функции регрессии.

¹¹В примерах опущены доказательства того, что правые части (2.2.2) и (2.2.3) измеримы относительно $\sigma(T)$. Убедитесь в этом сами.

¹²Убедитесь в этом.

2.3 Пределный переход под знаком УМО

Существуют различные варианты условий, при которых можно переходить к пределу под знаком УМО (см., например, [1, теор. 2 §7 гл.2]). Приведем один из них, аналогичный теореме Лебега о мажорируемой сходимости.

Предложение 2. Пусть $(\Omega, \mathcal{F}, P)$ — некоторое вероятностное пространство и $\mathcal{F}_0$ — σ -алгебра, содержащаяся в $\mathcal{F}$. Рассмотрим последовательность случайных величин ξ_n и случайную величину ξ такие, что $\xi_n \rightarrow \xi$ п.в. и $|\xi_n| \leq \eta$ с $E\eta < \infty$.

Тогда $E(\xi_n | \mathcal{F}_0) \rightarrow E(\xi | \mathcal{F}_0)$ и $E(|\xi_n - \xi| | \mathcal{F}_0) \rightarrow 0$ с вероятностью 1.

Доказательство. Заметим, что почти всюду $|\xi| \leq \eta$,¹³ Обозначим $\beta_n = \sup_{k \geq n} |\xi_k - \xi|$. Тогда с вероятностью 1 $\beta_n \downarrow 0$ и $\beta_n \leq 2\eta$ и, следовательно, $E\beta_n \rightarrow 0$ по «обычной» теореме Лебега о мажорируемой сходимости. Поэтому (здесь мы пользуемся¹⁴ Предложением 1)

$$|E(\xi_n | \mathcal{F}_0) - E(\xi | \mathcal{F}_0)| \leq E(|\xi_n - \xi| | \mathcal{F}_0) \leq E(\beta_n | \mathcal{F}_0) \downarrow \beta,$$

где β — некоторая P -п.в. неотрицательная случайная величина. При этом

$$0 \leq E\beta \leq E(E(\beta_n | \mathcal{F}_0)) = E\beta_n \rightarrow 0.$$

Следовательно, $P(\beta = 0) = 1$ и утверждение доказано. $\square$

Следствие 1. 1. Пусть $\eta_n \geq 0$, $\sum_{i=1}^n \eta_i \uparrow \xi$ с вероятностью 1 и $E\xi < \infty$. Тогда

$$E(\xi | \mathcal{F}_0) = E\left(\sum_{i=1}^{\infty} \eta_i | \mathcal{F}_0\right) = \sum_{i=1}^{\infty} E(\eta_i | \mathcal{F}_0) \quad P\text{-п.в.} \quad (2.3.1)$$

2. Пусть $\{A_i\}_{i \geq 1}$ — конечное или счетное семейство попарно непересекающихся событий и $A = \cup_{i \geq 1} A_i$. Тогда

$$P(A | \mathcal{F}_0) = P\left(\cup_{i \geq 1} A_i | \mathcal{F}_0\right) = \sum_{i \geq 1} P(A_i | \mathcal{F}_0) \quad P\text{-п.в.} \quad (2.3.2)$$

3. Пусть $\xi : (\Omega, \mathcal{F}) \mapsto (G, \mathcal{G})$, $T : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{A})$ и $\{C_i\}_{i \geq 1}$ — последовательность попарно дизъюнктивных элементов $\mathcal{G}$. Тогда

$$P(\xi \in \cup_{i \geq 1} C_i | T = t) = \sum_{i \geq 1} P(\xi \in C_i | T = t) \quad \mathcal{P}_T\text{-п.в.} \quad (2.3.3)$$

Доказательство. Если обозначить $\xi_n = \sum_{1 \leq i \leq n} \eta_i$ и положить $\eta = \xi$, то становится ясно, что первое утверждение непосредственно следует из Предложения 2.

Равенство (2.3.2) сводится к (2.3.1) выбором $\eta_i = I_{A_i}$, а (2.3.3) — к (2.3.2) выбором $A_i = \xi^{-1}C_i$ и записью соответствующего равенства в регрессионной форме. $\square$

Замечание 7. Функцию $P(\xi \in \cdot | T = t)$, определенную на σ -алгебре $\mathcal{G}$, хочется назвать условным распределением ξ при условии $T = t$. Действительно, из Предложения 1 следует, что $\mathcal{P}_T$ -почти всюду

$$P(\xi \in G | T = t) = 1 \quad \text{и} \quad P(\xi \in C | T = t) = 1 - P(\xi \notin C | T = t)$$

для любого $C \in \mathcal{G}$. В то же время (2.3.3) имеет, казалось бы, смысл счетной аддитивности этой функции.

¹³Поэтому $E|\xi| < \infty$.

¹⁴А как именно?

Здесь, однако, не все так просто. Дело в том, что в (2.3.1) исключительное множество (то есть то подмножество G , для которого (2.3.1) не выполняется) зависит, вообще говоря, от набора попарно дизъюнктивных множеств C_i , и для разных наборов C_i эти исключительные множества совпадать (и даже иметь непустое пересечение) не обязаны.

Поскольку интересующих нас наборов множеств C_i может быть очень много (это зависит от σ -алгебры $\mathcal{G}$), то, хотя каждое исключительное множество имеет нулевую меру, в принципе может оказаться, что не существует такого t , при котором равенство (2.3.1) выполнено для всех наборов C_i . Это, конечно, совсем не похоже на счетную аддитивность.

Решение этой проблемы связано с понятием регулярного варианта условного распределения, которое мы будем рассматривать в разделе 2.6.

2.4 Дальнейшие свойства УМО

Как и раньше, рассмотрим вероятностное пространство $(\Omega, \mathcal{F}, P)$ и σ -алгебру $\mathcal{F}_0 \subset \mathcal{F}$.

Предложение 3. 1. Пусть ξ и η — случайные величины такие, что ξ является $\mathcal{F}_0$ -измеримой, $E|\eta| < \infty$ и $E|\xi\eta| < \infty$. Тогда

$$E(\xi\eta | \mathcal{F}_0) = \xi E(\eta | \mathcal{F}_0) \quad P\text{-п.в.} \quad (2.4.1)$$

В частности, если ξ суммируема и $\mathcal{F}_0$ -измерима, то $E(\xi | \mathcal{F}_0) = \xi$ с вероятностью 1.

2. Если суммируемая случайная величина ξ и σ -алгебра $\mathcal{F}_0$ независимы,¹⁵ то

$$E(\xi | \mathcal{F}_0) = E\xi \quad P\text{-п.в.} \quad (2.4.2)$$

3. Рассмотрим σ -алгебры $\mathcal{F}_1$ и $\mathcal{F}_2$ такие, что $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$. Если $E|\xi| < \infty$, то P -п.в.

$$E(E(\xi | \mathcal{F}_1) | \mathcal{F}_2) = E(E(\xi | \mathcal{F}_2) | \mathcal{F}_1) = E(\xi | \mathcal{F}_1). \quad (2.4.3)$$

Доказательство. 1. Пусть $\xi = I_B$, где $B \in \mathcal{F}_0$. Обозначим $\beta = \xi E(\eta | \mathcal{F}_0)$. Тогда при $A \in \mathcal{F}_0$ по определению УМО

$$\int_A \beta dP = \int_{A \cap B} E(\eta | \mathcal{F}_0) dP = \int_{A \cap B} \eta dP = \int_A \xi \eta dP.$$

Так как β является $\mathcal{F}_0$ -измеримой, то равенство (2.4.1) доказано для случая, когда ξ представляет собой индикатор $\mathcal{F}_0$ -измеримого множества. Используя Предложение 1, сразу же получаем, что (2.4.1) имеет место для простых $\mathcal{F}_0$ -измеримых¹⁶ функций ξ .

Перейдем теперь к общему случаю. Прежде всего заметим, что равенство (2.4.1) достаточно¹⁷ доказывать для неотрицательных ξ, η . Возьмем последовательность неотрицательных простых $\mathcal{F}_0$ -измеримых функций ξ_n таких, что поточечно $\xi_n \uparrow \xi$. Тогда $\xi_n \eta \uparrow \xi \eta$.

Как уже доказано,

$$E(\xi_n \eta | \mathcal{F}_0) = \xi_n E(\eta | \mathcal{F}_0) \quad P\text{-п.в.} \quad (2.4.4)$$

Правая часть (2.4.4) стремится к $\xi E(\eta | \mathcal{F}_0)$, а левая, согласно Предложению 2, с вероятностью 1 к $E(\xi \eta | \mathcal{F}_0)$. Тем самым первое утверждение доказано.

2. Так как $E I_A \xi = P(A) E\xi$ при $A \in \mathcal{F}_0$, то утверждение следует из цепочки равенств

$$\int_A E\xi dP = P(A) E\xi = \int_A \xi dP.$$

¹⁵По определению это означает, что для любых $B \in \mathcal{B}_R$ и $A \in \mathcal{F}_0$ события A и $\{\omega : \xi(\omega) \in B\}$ независимы.

¹⁶то есть имеющих вид $\sum_{i=1}^m c_i I_{B_i}$, где $B_i \in \mathcal{F}_0$

¹⁷Почему?

3. Поскольку $\xi_1 \stackrel{\text{def}}{=} E(\xi | \mathcal{F}_1)$ является $\mathcal{F}_2$ -измеримой случайной величиной, то равенство

$$E(\xi_1 | \mathcal{F}_2) = \xi_1 \quad \text{Р-п.в.}$$

является частным случаем (2.4.1).

Обозначим $\xi_2 = E(E(\xi | \mathcal{F}_2) | \mathcal{F}_1)$ и возьмем $A \in \mathcal{F}_1$. Поскольку $A \in \mathcal{F}_2$, то

$$\int_A E(\xi | \mathcal{F}_2) dP = \int_A \xi dP.$$

Поэтому

$$\int_A \xi_2 dP = \int_A E(\xi | \mathcal{F}_2) dP = \int_A \xi dP = \int_A E(\xi | \mathcal{F}_1) dP.$$

Следовательно, $\xi_2 = E(\xi | \mathcal{F}_1)$ с вероятностью 1. Доказательство закончено. $\square$

Замечание 8. Перепишем равенства (2.4.1) и (2.4.2) в регрессионных терминах и обсудим получившиеся результаты.¹⁸

Если ввести отображение $T : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{A})$ и положить $\mathcal{F}_0 = \sigma(T)$, то (2.4.1) автоматически перепишется как

$$E(\varphi(T)\eta | T) = \varphi(T) E(\eta | T) \quad \text{Р-п.в.,} \quad (2.4.5)$$

где $\varphi(T)$ заменяет собой ξ . В регрессионных терминах (2.4.5) приобретет вид

$$E(\varphi(T)\eta | T = t) = \varphi(t) E(\eta | T = t) = E(\varphi(t)\eta | T = t) \quad \mathcal{P}_T\text{-п.в.,} \quad (2.4.6)$$

что представляется совершенно естественным в рамках интерпретации функции регрессии, приведенной в Замечании 6. Действительно, если мы знаем, что $T = t$, то произведение $\varphi(T)\eta$ должно превратиться в $\varphi(t)\eta$, после чего постоянная¹⁹ $\varphi(t)$ просто выносится за знак условного математического ожидания $E(\varphi(t)\eta | T = t)$. В частности, «совершенно очевидно», что

$$E(\varphi(T) | T = t) = \varphi(t),$$

иначе просто быть не может! А условие « $\mathcal{P}_T$ -п.в.» — это всего-навсего результат формализации интуитивно очевидного факта.

Из (2.4.6) сразу же следует (см. (2.2.4)), что

$$E(\varphi(T)\eta) = \int_D \varphi(t) E(\eta | T = t) \mathcal{P}_T(dt) \quad (2.4.7)$$

Это тождество еще будет обсуждаться в дальнейшем.

Что касается равенства (2.4.2), то при выборе $\mathcal{F}_0 = \sigma(T)$ независимость ξ и $\mathcal{F}_0$ означает всего лишь независимость ξ и T . Поэтому, переписав (2.4.2) в регрессионной форме

$$E(\xi | T = t) = E\xi \quad \mathcal{P}_T\text{-п.в.,} \quad (2.4.8)$$

мы снова приходим к интуитивной очевидности этого равенства: конечно, если ξ и T независимы, то знание значения T никак не влияет на нашу информацию о распределении (и, следовательно, о среднем значении) случайной величины ξ .

¹⁸Сделайте то же самое для равенства (2.4.3), на которое чуть позже мы посмотрим с другой точки зрения.

¹⁹то есть не зависящая от ω

Теорема 2. (Неравенство Иенсена для УМО).

Рассмотрим, кроме вероятностного пространства $(\Omega, \mathcal{F}, P)$, σ -алгебры $\mathcal{F}_0 \subset \mathcal{F}$ и случайной величины ξ с конечным математическим ожиданием, еще и измеримую функцию $\varphi : (\mathbb{R}, \mathcal{B}_R) \mapsto (\mathbb{R}, \mathcal{B}_R)$ такую, что $E|\varphi(\xi)| < \infty$. Если функция φ выпукла вниз, то

$$E(\varphi(\xi) | \mathcal{F}_0) \geq \varphi(E(\xi | \mathcal{F}_0)) \quad P\text{-п.в.} \quad (2.4.9)$$

Доказательство. Опишем выпуклость функции φ следующим образом: для любой точки $x_0 \in \mathbb{R}$ через точку $(x_0, \varphi(x_0))$ можно провести такую прямую, что весь график функции $y = \varphi(x)$ будет находиться не ниже этой прямой.²⁰ С формальной точки зрения это означает, что для любого x_0 существует такое $c = c(x_0)$, что

$$\varphi(x) \geq \varphi(x_0) + c(x_0)(x - x_0). \quad (2.4.10)$$

Равенства (2.4.10) уже достаточно для доказательства обычного неравенства Иенсена,²¹ доказывая (2.4.9), нам еще придется потрудиться.

Прежде всего, заметим,²² что функцию $z \mapsto c(z)$ можно выбрать $\mathcal{B}_R$ измеримой. Затем, подставив в (2.4.10) вместо x случайную величину ξ , а вместо $x_0 = E(\xi | \mathcal{F}_0)$, получим неравенство

$$\varphi(\xi) \geq \varphi(E(\xi | \mathcal{F}_0)) + c(E(\xi | \mathcal{F}_0)) (\xi - E(\xi | \mathcal{F}_0)). \quad (2.4.11)$$

Если бы мы точно знали, что у правой части неравенства (2.4.11) есть конечное математическое ожидание, то взяв УМО относительно σ -алгебры $\mathcal{F}_0$ от обеих частей (2.4.11), мы получили бы, используя Предложения 1 и 3, что P п.в.

$$E(\varphi(\xi) | \mathcal{F}_0) \geq \varphi(E(\xi | \mathcal{F}_0)) + c(E(\xi | \mathcal{F}_0)) E((\xi - E(\xi | \mathcal{F}_0)) | \mathcal{F}_0) = \varphi(E(\xi | \mathcal{F}_0)),$$

то есть необходимое нам утверждение. Однако, поскольку поведение (кроме постулированной измеримости) функции $c(x)$ нам, вообще говоря, неизвестно, такой путь не является корректным.

Введем обозначения

$$c_n(z) = \begin{cases} c(z) & \text{если } |c(z)| \leq n, \\ 0 & \text{если } |c(z)| > n, \end{cases}$$

$\Delta = c(E(\xi | \mathcal{F}_0))$, $\Delta_n = c_n(E(\xi | \mathcal{F}_0))$ и $B_n = \{\omega : |\Delta| \leq n\}$. Тогда, очевидно, Δ_n является $\mathcal{F}_0$ -измеримой случайной величиной, $B_n \in \mathcal{F}_0$ и $B_n \uparrow \Omega$ при $n \rightarrow \infty$. Кроме того, так как Δ_n ограничена, то, обозначая

$$\theta_n = \Delta_n E(\xi - E(\xi | \mathcal{F}_0)) = \mathbb{I}_{B_n} c(E(\xi | \mathcal{F}_0)) (\xi - E(\xi | \mathcal{F}_0)),$$

получим, что P -п.в.

$$E(\theta_n | \mathcal{F}_0) = \Delta_n E(\xi - E(\xi | \mathcal{F}_0)) = 0,$$

и, следовательно, для любого $A \in \mathcal{F}_0$

$$\int_A \theta_n dP = \int_A E(\theta_n | \mathcal{F}_0) dP = 0.$$

²⁰В гладком случае, конечно, эта прямая является касательной к графику функции φ , в общем случае — одной из опорных прямых.

²¹Для этого нужно подставить в (2.4.10) вместо x случайную величину ξ , вместо x_0 — ее среднее $E\xi$, и взять математическое ожидание от обеих частей получившегося неравенства. В результате получим, что $E\varphi(\xi) \geq \varphi(E\xi)$.

²²Точнее — примем этот факт, легко проверяемый в конкретных ситуациях, без доказательства.

Следовательно, возвращаясь к (2.4.11), видим, что для любого $A \in \mathcal{F}_0$

$$\int_{A \cap B_n} \varphi(\xi) dP \geq \int_{A \cap B_n} \varphi(E(\xi | \mathcal{F}_0)) dP + \int_{A \cap B_n} \theta_n dP = \int_{A \cap B_n} \varphi(E(\xi | \mathcal{F}_0)) dP.$$

Устремляя n к бесконечности, приходим к неравенству

$$\int_A \varphi(\xi) dP \geq \int_A \varphi(E(\xi | \mathcal{F}_0)) dP,$$

которое, ввиду произвольности $A \in \mathcal{F}_0$, эквивалентно (2.4.9). $\square$

Замечание 9. Конечно, в регрессионной форме неравенство Иенсена (2.4.9) имеет вид

$$E(\varphi(\xi) | T = t) \geq \varphi(E(\xi | T = t)) \quad \mathcal{P}_T\text{-п.в.} \quad (2.4.12)$$

2.5 УМО как оператор проектирования. Условная дисперсия

В этом разделе мы будем интерпретировать случайные величины (в том числе условные математические ожидания относительно σ -алгебр и отображений) как классы совпадающих почти всюду между собой функций, переходя тем самым на язык функционального анализа. Кроме того, мы будем предполагать, что $E\xi^2 < \infty$.

УМО как проектор. Как и раньше, зафиксируем вероятностное пространство $(\Omega, \mathcal{F}, P)$, σ -алгебру $\mathcal{F}_0 \subset \mathcal{F}$ и случайную величину ξ с конечным математическим ожиданием.

Пусть дополнительно $E\xi^2 < \infty$. Тогда мы можем рассматривать ξ как элемент Гильбертова пространства $\mathbb{L}^2(\mathcal{F}) = \mathbb{L}^2(\mathcal{F}, P)$ функций, измеримых относительно σ -алгебры $\mathcal{F}$ и имеющих конечные вторые моменты по мере P . Как всегда, это пространство оснащено скалярным произведением $(\eta_1, \eta_2) = \int_{\Omega} \eta_1 \eta_2 dP$.

Поскольку функция $\varphi(x) = x^2$ выпукла вниз, то для этой функции выполняется неравенство Иенсена, то есть

$$(E(\xi | \mathcal{F}_0))^2 \leq E(\xi^2 | \mathcal{F}_0) \quad P\text{-п.в.}$$

Беря математическое ожидание от обеих частей этого равенства, получим, что

$$E(E(\xi | \mathcal{F}_0))^2 \leq EE(\xi^2 | \mathcal{F}_0) = E\xi^2.$$

Тем самым мы видим, что из условия $\xi \in \mathbb{L}^2(\mathcal{F})$ следует, что $\xi' \stackrel{\text{def}}{=} E(\xi | \mathcal{F}_0) \in \mathbb{L}^2(\mathcal{F})$. При этом ξ' измерима относительно $\mathcal{F}_0$ и $E\xi' = E\xi$, так что $D\xi' \leq D\xi$.

Эти свойства имеют простое объяснение. Рассмотрим множество $\mathbb{L}^2(\mathcal{F}_0)$ функций, измеримых относительно σ -алгебры $\mathcal{F}_0$ и имеющих конечные вторые моменты. Очевидно, $\mathbb{L}^2(\mathcal{F}_0) \subset \mathbb{L}^2(\mathcal{F})$, причем $\mathbb{L}^2(\mathcal{F}_0)$ является замкнутым линейным подпространством $\mathbb{L}^2(\mathcal{F})$.

Предложение 4. Если $E\xi^2 < \infty$, то условное математическое ожидание $E(\xi | \mathcal{F}_0)$ является результатом ортогонального проектирования в $\mathbb{L}^2(\mathcal{F})$ функции ξ на $\mathbb{L}^2(\mathcal{F}_0)$.

Доказательство. Случайная величина $\xi' \in \mathbb{L}^2(\mathcal{F}_0)$ является ортогональной проекцией ξ на подпространство $\mathbb{L}^2(\mathcal{F}_0)$ тогда и только тогда, когда $\xi - \xi'$ ортогональна η для любой $\eta \in \mathbb{L}^2(\mathcal{F}_0)$. Проверим это свойство, выбрав $\xi' = E(\xi | \mathcal{F}_0)$. Действительно,

$$E(\xi - E(\xi | \mathcal{F}_0))\eta = E\xi\eta - E(\eta E(\xi | \mathcal{F}_0)) = E\xi\eta - E(E(\xi\eta | \mathcal{F}_0)) = E\xi\eta - E\xi\eta = 0.$$

Доказательство закончено. $\square$

Замечание 10. 1. Обозначим $\xi' = E(\xi | \mathcal{F}_0)$. Поскольку для любой $\eta \in \mathbb{L}^2(\mathcal{F}_0)$

$$\begin{aligned} E(\xi - \eta)^2 &= E((\xi - \xi') + (\xi' - \eta))^2 = \\ &= E(\xi - \xi')^2 + E(\xi' - \eta)^2 + 2E(\xi - \xi')(\xi' - \eta) = E(\xi - \xi')^2 + E(\xi' - \eta)^2 \geq E(\xi - \xi')^2, \end{aligned}$$

то, кроме «геометрической» интерпретации УМО суммируемой с квадратом случайной величины ξ относительно σ -алгебры $\mathcal{F}_0$, мы получаем и «оптимизационную» интерпретацию:

$$E(\xi - E(\xi | \mathcal{F}_0))^2 = \min_{\eta \in \mathbb{L}^2(\mathcal{F}_0)} E(\xi - \eta)^2. \quad (2.5.1)$$

2. Наиболее выпукло последнее свойство проявляется в регрессионной постановке. Пусть $\xi \in \mathbb{L}^2(\mathcal{F})$ и $T : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{A})$. Как нам выбрать функцию $\varphi : (D, \mathcal{A}) \mapsto (\mathbb{R}, \mathcal{B}_R)$, чтобы случайная величина $\varphi(T)$ наилучшим образом аппроксимировала ξ в среднеквадратическом?

Ответ содержится в (2.5.1): искомая функция φ совпадает с функцией регрессии $E(\xi | T = t)$, а сама случайная величина $\varphi(T)$ — с условным математическим ожиданием $E(\xi | T)$.

3. В случае, когда $E\xi^2 < \infty$, равенство (2.4.3) Предложения 3 приобретает понятный геометрический смысл. Действительно, поскольку $\mathbb{L}^2(\mathcal{F}_1) \subset \mathbb{L}^2(\mathcal{F}_2) \subset \mathbb{L}^2(\mathcal{F})$, то не важно, в каком порядке проектировать ξ на $\mathbb{L}^2(\mathcal{F}_1)$ и $\mathbb{L}^2(\mathcal{F}_2)$ — результатом все равно будет проекция ξ на меньшее из этих подпространств.

4. По определению, оператор $\Gamma : X \mapsto X$ называется оператором проектирования, если $\Gamma^2 = \Gamma$. Поскольку при $E|\xi| < \infty$

$$E(E(\xi | \mathcal{F}_0) | \mathcal{F}_0) = E(\xi | \mathcal{F}_0),$$

то оператор взятия условного математического ожидания всегда является оператором проектирования в $\mathbb{L}^1 = \mathbb{L}^1(\mathcal{F}, P)$. Если же мы ограничиваемся случайными величинами с конечными вторыми моментами, то этот оператор оказывается ортогональным проектором $\mathbb{L}^2(\mathcal{F}) \mapsto \mathbb{L}^2(\mathcal{F}_0)$.

Условная дисперсия. Основное дисперсионное тождество. Требование конечности второго момента у случайной величины ξ позволяет ввести понятия не только условного математического ожидания, но и условной дисперсии относительно σ -алгебры. Но сначала дадим следующее определение.

Определение 2.2. Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство. События $A_1, \dots, A_n$ называются *условно независимыми относительно σ -алгебры $\mathcal{F}_0 \subset \mathcal{F}$* , если

$$P(A_1 \cap \dots \cap A_n | \mathcal{F}_0) = \prod_{i=1}^n P(A_i | \mathcal{F}_0) \quad P\text{-п.в.} \quad (2.5.2)$$

Случайные величины $\xi_1, \dots, \xi_n$ называются *условно независимыми относительно σ -алгебры $\mathcal{F}_0 \subset \mathcal{F}$* , если для любых борелевских множеств $B_1, \dots, B_n$ события $\{\omega : \xi_i(\omega) \in B_i\}$ являются условно независимыми.

Замечание 11. Регрессионный вариант равенств, определяющих условную независимость случайных величин, очевидно, имеет вид

$$P(\xi_1 \in B_1, \dots, \xi_n \in B_n | T = t) = \prod_{i=1}^n P(\xi_i \in B_i | T = t) \quad P_{T\text{-п.в.}},$$

где $B_i \in \mathcal{B}_R$. Этот вариант делает определение интуитивно очевидным. Пусть, например, α_1, α_2 и T — независимые случайные величины, равномерно распределенные на $(0, 1]$. Положим $\xi_1 = \alpha_1 T$ и $\xi_2 = \alpha_2 T$. Случайные величины ξ_1 и ξ_2 , конечно, зависимы, но при условии $T = t$ они независимы и равномерно распределены на $(0, t]$.

Лемма 3. Пусть ξ и η условно независимы при условии $\mathcal{F}_0$. Если $E|\xi| < \infty$, $E|\eta| < \infty$ и $E|\xi\eta| < \infty$, то Р-п.в.

$$E(\xi\eta | \mathcal{F}_0) = E(\xi | \mathcal{F}_0) E(\eta | \mathcal{F}_0). \quad (2.5.3)$$

Доказательство. Приведем лишь краткую схему доказательства, которое во многом аналогично доказательству первого пункта Предложения 3.

Прежде всего, если ξ и η представляют собой индикаторы $\mathcal{F}$ -измеримых множеств, то (2.5.3) следует из (2.5.2). Отсюда сразу же следует равенство (2.5.3) для простых условно-независимых случайных величин.

Переход к общему случаю тоже делается стандартным образом. Заметив, что (2.5.3) достаточно доказывать для неотрицательных ξ и η , нужно аппроксимировать их снизу простыми случайными величинами, а затем применить Предложение 2 о предельном переходе под знаком УМО. $\square$

Определение 2.3. Пусть $E\xi^2 < \infty$. Условной дисперсией случайной величины ξ относительно σ -алгебры $\mathcal{F}_0$ называется случайная величина $D(\xi | \mathcal{F}_0) \stackrel{\text{def}}{=} E((\xi - E(\xi | \mathcal{F}_0))^2 | \mathcal{F}_0)$.

Условная дисперсия обладает следующими свойствами, выполняющимися Р-п.в.

Предложение 5. Пусть ξ и η обладают вторыми конечными моментами. Тогда

1. $D(\xi | \mathcal{F}_0) = E(\xi^2 | \mathcal{F}_0) - (E(\xi | \mathcal{F}_0))^2$;²³
2. условная дисперсия $D(\xi | \mathcal{F}_0)$ неотрицательна и равна нулю тогда и только тогда, когда ξ является $\mathcal{F}_0$ -измеримой;
3. для любых $\mathcal{F}_0$ -измеримых a и b таких, что $E|b| < \infty$ и $Ea^2\xi^2 < \infty$

$$D(a\xi + b | \mathcal{F}_0) = a^2 D(\xi | \mathcal{F}_0) \quad \text{Р-п.в.};$$

4. если ξ и η условно независимы при условии $\mathcal{F}_0$, то $D(\xi + \eta | \mathcal{F}_0) = D(\xi | \mathcal{F}_0) + D(\eta | \mathcal{F}_0)$ Р-п.в.

Доказательство. Доказательство этого утверждения повторяет (с учетом свойств условных математических ожиданий) доказательство аналогичных свойств обычной дисперсии и поэтому опускается.²⁴ $\square$

Предложение 6. Если $E\xi^2 < \infty$, то имеет место равенство

$$D\xi = E(D(\xi | \mathcal{F}_0)) + D(E(\xi | \mathcal{F}_0)). \quad (2.5.4)$$

Доказательство. Действительно,

$$E(D(\xi | \mathcal{F}_0)) = E(E(\xi^2 | \mathcal{F}_0)) - E(E(\xi | \mathcal{F}_0))^2 = E\xi^2 - E(E(\xi | \mathcal{F}_0))^2 = E(\xi - E(\xi | \mathcal{F}_0))^2$$

и

$$D(E(\xi | \mathcal{F}_0)) = E(E(\xi | \mathcal{F}_0))^2 - \left(E(E(\xi | \mathcal{F}_0))\right)^2 = E(E(\xi | \mathcal{F}_0))^2 - (E\xi)^2.$$

Складывая эти два равенства, получаем (2.5.4). $\square$

Равенство (2.5.4) называется *основным дисперсионным тождеством*. Для краткости его иногда записывают в виде $D\xi = ED_{\mathcal{F}_0}(\xi) + DE_{\mathcal{F}_0}(\xi)$.

²³Сравните это равенство с (1.1.3).

²⁴Проведите это доказательство. А как изменится п. 3 этого предложения, если в нем a и b не постоянные, а $\mathcal{F}_0$ -измеримые случайные величины?

Замечание 12. Равенство (2.5.4) имеет простой геометрический смысл. Во-первых, числа $D\xi$, $ED_{\mathcal{F}_0}(\xi)$ и $DE_{\mathcal{F}_0}(\xi)$ не изменяются,²⁵ если ξ заменить на $\xi - E\xi$. Следовательно, можно считать для простоты, что $E\xi = 0$.

Далее, при этом условии $D\xi$ — это квадрат нормы элемента ξ пространства $\mathbb{L}_{\mathcal{F}}^2$. Точно также $DE_{\mathcal{F}_0}(\xi)$ является квадратом нормы проекции $\xi' = E(\xi | \mathcal{F}_0)$ случайной величины ξ на подпространство $\mathbb{L}_{\mathcal{F}_0}^2$. Поскольку $\xi - \xi'$ ортогонально ξ' , а

$$E\xi\xi' = E\left(E(\xi\xi' | \mathcal{F}_0)\right) = E\left(\xi'E(\xi | \mathcal{F}_0)\right) = E(\xi')^2$$

и поэтому

$$E(\xi - \xi')^2 = E\xi^2 - 2E\xi\xi' + E(\xi')^2 = E\xi^2 - E(\xi')^2 = ED_{\mathcal{F}_0}(\xi),$$

то (2.5.4) представляет собой просто-напросто вариант теоремы Пифагора в $\mathbb{L}_{\mathcal{F}}^2$:

$$\|\xi\|_2^2 = \|\xi - \xi'\|_2^2 + \|\xi'\|_2^2.$$

Корреляционное отношение. Пусть σ -алгебра $\mathcal{F}_0 = \sigma(T)$ порождается отображением $T : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{A})$. Тогда с условной дисперсией связано понятие *корреляционного отношения*, определяемого как

$$r^2(\xi | T) = 1 - \frac{ED_T(\xi)}{D\xi} = \frac{DE_T(\xi)}{D\xi}, \quad (2.5.5)$$

где обозначения $DE_T(\xi)$ и $ED_T(\xi)$ используются вместо $ED_{\mathcal{F}_0}(\xi)$ и $DE_{\mathcal{F}_0}(\xi)$. Корреляционное отношение является мерой функциональной зависимости ξ от T : оно неотрицательно, не превосходит 1, равно 0 в случае, когда ξ и T независимы и принимает значение 1, если ξ является $\sigma(T)$ -измеримой случайной величиной.

Замечание 13. Полезно сравнить корреляционное отношение с квадратом коэффициента корреляции. Действительно, пусть $(D, \mathcal{A}) = (\mathbb{R}, \mathcal{B})$ и $ET^2 < \infty$. Рассмотрим вместо величины

$$\min_{\varphi} E(\xi - \varphi(T))^2 = E(\xi - E(\xi | T))^2 = ED_T(\xi)$$

(минимум берется по всем φ таким, что $E\varphi^2(T) < \infty$) величину²⁶ $\min_{a,b} E(\xi - (aT + b))^2$. Нетрудно посчитать,²⁷ что

$$\min_{a,b} E(\xi - (aT + b))^2 = D\xi (1 - \rho^2),$$

где ρ — коэффициент корреляции между ξ и T . Поэтому вместо формулы (2.5.5), которая может быть записана в виде

$$r^2(\xi | T) = 1 - \frac{\min_{\varphi} E(\xi - \varphi(T))^2}{D\xi},$$

²⁵Проверьте!

²⁶То есть здесь мы ортогонально проектируем ξ на множество линейных функций от T . Иначе говоря, мы ищем линейную регрессию ξ на T и изучаем ее остаток. Для коэффициента корреляции условие $T \in \mathbb{R}$ является необходимым, для корреляционного отношения оно совершенно необязательно.

²⁷Это следует из явного вида оптимальных значений a_{opt}, b_{opt} :

$$a_{opt} = \rho \sqrt{D\xi/D_T}, \quad b_{opt} = -a_{opt}ET + E\xi.$$

у нас получается

$$\rho^2(\xi, T) = 1 - \frac{\min_{a,b} E(\xi - (aT + b))^2}{D\xi}.$$

Очевидно, что $r^2(\xi | T) \geq \rho^2(\xi, T)$. При этом полезно иметь в виду, что $\rho^2(\xi, T) = \rho^2(T, \xi)$, в то время как, вообще говоря, $r^2(\xi | T) \neq r^2(T | \xi)$.

Пример 3. Пусть $T \in U(-1, 1)$ и $\xi = T^2$. Так как ξ есть функция от T , то $r^2(\xi | T) = 1$. Сосчитаем коэффициент корреляции между ξ и T . Так как $ET = 0$, то $\text{Cov}(\xi, T) = E\xi T = ET^3 = 0$, то есть ξ и T некоррелированы.

Найдем теперь величину $r^2(T | \xi)$. Нетрудно видеть, что для любого $x \in (0, 1)$ имеет место равенство $E(T | \xi = x) = 0$. Действительно, условное распределение $\mathcal{L}(T | \xi = x)$ автоматически сосредоточено в точках $\pm\sqrt{x}$, причем веса этого распределения обязаны быть одинаковыми ввиду четности функции T^2 .

Следовательно, $E(T | \xi) = 0$ почти всюду и поэтому

$$r^2(T | \xi) = \frac{DE_\xi(T)}{D(T)} = \frac{D(E(T | \xi))}{D(T)} = 0.$$

Таким образом, ξ функционально зависит от T , но эта зависимость абсолютно ($\rho(\xi, T) = 0$) не является линейной. В то же время никаких следов функциональной зависимости T от ξ нет, так как $r^2(T | \xi) = 0$.

Конечно, такой же результат будет верен для любой суммируемой с квадратом случайной величины вида $\xi = \varphi(T)$, если только φ — четная функция, не равная некоторой постоянной.²⁸

А теперь — 3 задачи.

Задача 1. 1. Найдите $r^2(\xi | T)$ и $r^2(T | \xi)$ а случае, когда $\xi, T \in \mathbb{R}$, $E\xi = ET = 0$, $0 < E\xi^2 ET^2 < \infty$ и $\xi = aT + b$ с $a \neq 0$. Сравните результат с $\rho(\xi, T)$.

2. Решите предыдущую задачу, если $\xi = aT + b + \varepsilon$, где ε и T независимы, причем $E\xi = 0$ и $E\xi^2 < \infty$.

2. Пусть $\xi \in U(0, 1)$, $T = \lfloor n \rfloor$, где $n > 1$ целое. Найдите $r^2(\xi | T)$, $r^2(T | \xi)$ и $\rho(\xi, T)$. Как ведут себя эти характеристики при $n \rightarrow \infty$?

2.6 Регулярные варианты условных распределений

В разделе 2.3 (см. Замечание 7) обсуждались сложности, связанные с формализацией (интуитивно очевидного) понятия «условного распределения относительно отображения». В этом разделе будет указан способ разрешения этих сложностей.

Определение 2.4. Рассмотрим вероятностное пространство $(\Omega, \mathcal{F}, P)$ и два измеримых пространства $(D, \mathcal{A})$ и $(E, \mathcal{E})$. Пусть $\xi : (\Omega, \mathcal{F}) \mapsto (E, \mathcal{E})$ и $T : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{A})$.

Функция $Q : \mathcal{E} \times D \mapsto \mathbb{R}$ называется *регулярным вариантом условного распределения ξ относительно T* , если

1. для любого $t \in D$ функция $Q(\cdot; t)$ является вероятностной мерой;
2. для любого $A \in \mathcal{E}$ функция $Q(A; \cdot)$ является $\mathcal{A}$ -измеримой;
3. для любого $A \in \mathcal{E}$

$$Q(A; t) = P(\xi \in A | T = t) \quad \mathcal{P}_T\text{-п.в.} \quad (2.6.1)$$

Замечание 14. 1. Если $(E, \mathcal{E}) = (\Omega, \mathcal{F})$ и $\xi(\omega) = \omega$, то функция Q называется *регулярным вариантом условной вероятности относительно отображения T* .

2. Если ξ и T независимы, то $Q(A; t) = P(\xi \in A)$ при любом t . Это сразу же следует из равенства (2.4.8) (см. Замечание 8 к Предложению 3).

²⁸ А что будет, если $\xi = |T + 0.5|$?

Полезность введения понятия регулярного варианта условного распределения сразу следует из следующего утверждения.

Предложение 7. Рассмотрим в условиях Определения 2.4 функцию $g : (E, \mathcal{E}) \mapsto (\mathbb{R}, \mathcal{B}_R)$ такую, что $E|g(\xi)| < \infty$. Тогда

$$E(g(\xi) | T = t) = \int_E g(x) Q(dx; t) \quad \mathcal{P}_T\text{-п.в.} \quad (2.6.2)$$

и

$$Eg(\xi) = \int_D \mathcal{P}_T(dt) \int_E g(x) Q(dx; t). \quad (2.6.3)$$

Доказательство. При $A \in \mathcal{E}$ и $g = \mathbb{I}_A$ равенство (2.6.2) превращается в (2.6.1). Поэтому (2.6.2) имеет место для любой простой функции g . Дальше все стандартно: рассматриваем неотрицательные функции g , берем последовательность простых функций g_n таких, что поточечно $g_n \uparrow g$ и переходим к пределу под знаком как обычного интеграла (в правой части (2.6.2)) так и под знаком условной вероятности (в левой части (2.6.2)).

Что касается (2.6.3), то

$$Eg(\xi) = \int_D \mathcal{P}_T(dt) E(g(\xi) | T = t)$$

и остается воспользоваться (2.6.2). □

Замечание 15. 1. Число $Q(A; t)$, конечно же, имеет смысл вероятности события $\xi \in A$ при условии, что отображение T приняло значение t .

2. Формула (2.6.3) представляет собой обобщение элементарной формулы полной вероятности²⁹: имеется двухэтапный эксперимент, на первом этапе которого измеряется значение T , а на втором — ξ , причем распределение ξ зависит от уже известного значения T . Если эта зависимость выражается регулярным вариантом Q условного распределения, то вся информация о распределении ξ записывается в виде (2.6.3).

3. Из (2.6.3), в частности, следует, что распределения случайных величин ξ и η совпадают, если совпадают их регулярные варианты распределений относительно одного и того же отображения T .³⁰

Приведем еще один пример использования понятия регулярного варианта условного распределения, связанный с основным дисперсионным тождеством.

Предложение 8. Рассмотрим вероятностное пространство $(\Omega, \mathcal{F}, P)$ и измеримое отображение $T : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{A})$. Пусть $\xi_1, \dots, \xi_n$ — случайные величины, обладающие следующими свойствами:

1. $E\xi_i^2 < \infty$ при $1 \leq i \leq n$;
2. $\xi_1, \dots, \xi_n$ условно независимы при условии T ;
3. Существуют совпадающие регулярные варианты $Q_i(dx; t) = Q(dx; t)$ условных распределений ξ_i относительно T .

Запишем основное дисперсионное тождество для случайной величины $\xi = \xi_1$ в виде $D\xi = ED_T(\xi) + DE_T(\xi)$. Тогда

$$D\left(\frac{\xi_1 + \dots + \xi_n}{n}\right) = \frac{ED_T(\xi)}{n} + DE_T(\xi).$$

²⁹Вернее, «полного математического ожидания».

³⁰Почему?

Доказательство. Обозначим $\eta_n = (\xi_1 + \dots + \xi_n)/n$. Тогда $D\eta_n = ED_T(\eta_n) + DE_T(\eta_n)$. Поскольку согласно (2.6.2)

$$E(\xi_i | T = t) = \int_{\mathbb{R}} xQ_i(dx; t) = \int_{\mathbb{R}} xQ(dx; t) = E(\xi | T = t) \quad \mathcal{P}_T\text{-п.в.},$$

то $E(\eta_n | T) = E(\xi | T)$ с вероятностью 1 и $DE_T(\eta_n) = DE_T(\xi)$.

С другой стороны, так как случайные величины ξ_i условно независимы при условии T , то, согласно Предложению 5, с вероятностью 1

$$D(\eta_n | T) = n^{-2} \sum_{i=1}^n D(\xi_i | T) = n^{-1} D(\xi | T),$$

причем последний переход обосновывается точно так же, как равенство $E(\eta_n | T) = E(\xi | T)$. Утверждение доказано. $\square$

Перейдем теперь к вопросам существования регулярных вариантов условных распределений. Начнем с двух частных случаев.

Предложение 9. 1. Пусть в условиях Определения 2.4 $D = \{t_1, \dots, t_n, \dots\}$ является конечным или счетным множеством, а σ -алгебра $\mathcal{A}$ состоит из всех подмножеств множества D . Предположим дополнительно, что $P(T = t_i) > 0$ при всех i . Тогда существует регулярный вариант $Q(dx; t)$ условного распределения ξ относительно T , который имеет вид

$$Q(A; t_i) = \frac{P(\xi \in A, T = t_i)}{P(T = t_i)}.$$

2. Пусть совместное распределение пары отображений (ξ, T) имеет плотность $p(x, t)$ относительно меры $\mu(dx) \otimes \nu(dt)$, определенной на σ -алгебре $\mathcal{E} \times \mathcal{A}$ подмножеств множества $E \times D$. Обозначим

$$p_T(t) = \int_{\mathbb{R}} p(x, t) \mu(dx).$$

Тогда существует регулярный вариант $Q(dx; t)$ условного распределения ξ относительно T , имеющий вид

$$Q(A; t) = \begin{cases} \frac{\int_A p(x, t) \mu(dx)}{p_T(t)} & \text{при } p_T(t) \neq 0, \\ P(\xi \in A) & \text{при } p_T(t) = 0. \end{cases} \quad (2.6.4)$$

Доказательство. При доказательстве по существу повторяются рассуждения Примера 2 (раздел 2.2). Кратко покажем это для второго утверждения. Прежде всего, правая часть (2.6.4) удовлетворяет первым двум условиям Определения 2.4. Осталось проверить третье условие.

Согласно (2.2.1) это означает, что для любого $B \in \mathcal{A}$ должно выполняться равенство

$$\int_{T^{-1}B} \mathbb{I}_A(\xi) dP = \int_B Q(A; t) \mathcal{P}_T(dt). \quad (2.6.5)$$

Левая часть (2.6.5), очевидно, равна $P(\xi \in A, T \in B)$. Что касается правой части, обозначим $D_0 = \{t : p_T(t) \neq 0\}$ и заметим, что $P(T \in D_0) = 1$. Поскольку $\mathcal{P}_T(dt) = p_T(t) \nu(dt)$, то имеет место цепочка равенств

$$\begin{aligned} & \int_{B \cap D_0} \left(\int_A \frac{p(x, t) \mu(dx)}{p_T(t)} \right) \mathcal{P}_T(dt) + \int_{B \cap D_0^c} P(\xi \in A) \mathcal{P}_T(dt) = \\ & = \int_{B \cap D_0} \left(\int_A p(x, t) \mu(dx) \right) \nu(dt) = P(\xi \in A, T \in B \cap D_0) = P(\xi \in A, T \in B). \end{aligned}$$

Утверждение доказано. $\square$

Замечание 16. Конечно, в правой части (2.6.4) для случая $p_T(t) = 0$ может стоять любое распределение, определенное на σ -алгебре $\mathcal{A}$.

Хотя Предложение 9 охватывает широкий класс ситуаций, в которых существуют регулярные варианты условных распределений (и дает принципиальную возможность их выписывать), хочется иметь для этого общие условия, не связанные с дискретностью или абсолютной непрерывностью распределений. Сформулируем без доказательства соответствующее утверждение.³¹

Теорема 3. Пусть $(E, \rho, \mathcal{E})$ — полное метрическое сепарабельное пространство с борелевской σ -алгеброй. Если ξ — случайная величина со значениями в $(E, \mathcal{E})$, то для любого измеримого пространства $(D, \mathcal{A})$ и любого отображения $T : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{A})$ существует регулярный вариант условного распределения ξ относительно T .

Из Теоремы 3, в частности, следует, что существует регулярный вариант условного распределения любого случайного вектора со значениями в евклидовом пространстве любой размерности.

2.7 Теорема о монотонном классе. Применение к УМО

Для дальнейшего изучения свойств условных распределений нам понадобится следующее определение.

Определение 2.5. Система подмножеств $\mathcal{M}$ некоторого множества E называется *монотонным классом*, если она замкнута относительно счетных монотонных объединений и пересечений. Иначе говоря, $\mathcal{M}$ должна обладать следующими свойствами:

1. если $A_i \in \mathcal{M}$ и $A_i \subset A_{i+1}$, то $\bigcup_i A_i \in \mathcal{M}$;
2. если $A_i \in \mathcal{M}$ и $A_i \supset A_{i+1}$, то $\bigcap_i A_i \in \mathcal{M}$.

Очевидно, любая σ -алгебра является монотонным классом. Обратное, конечно, неверно.³²

Лемма 4. 1. Пусть $\mathcal{E}$ — некоторая система подмножеств множества E . Рассмотрим всевозможные монотонные классы, содержащие $\mathcal{E}$ и состоящие из подмножеств множества E . Тогда среди этих классов существует минимальный.

2. Пусть $\mathcal{E}$ — некоторая алгебра подмножеств множества E . Для того, чтобы $\mathcal{E}$ была σ -алгеброй, необходимо и достаточно, чтобы $\mathcal{E}$ было монотонным классом.

Доказательство. 1. Прежде всего, семейство монотонных классов, содержащих $\mathcal{E}$, не пусто, так как туда входит множество всех подмножеств E , являющееся σ -алгеброй. Рассмотрим пересечение всех этих монотонных классов (оно снова не пусто, так как содержит $\mathcal{E}$) и обозначим это пересечение $\mathcal{M}$. Это $\mathcal{M}$ и будет искомым минимальным монотонным классом.³³

2. Конечно, необходимость очевидна. Так как $\mathcal{E}$ уже является алгеброй, достаточно показать, что $\mathcal{E}$ замкнуто относительно счетных объединений. Пусть $A_i \in \mathcal{E}$, $i \geq 1$ Тогда

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_1 \cup A_2) \cup (A_1 \cup A_2 \cup A_3) \cup \dots,$$

откуда немедленно следует требуемое. □

Теорема 4. (теорема о монотонном классе.)³⁴

Пусть $\mathcal{E}$ — алгебра подмножеств множества E и $\sigma(\mathcal{E})$ — минимальная σ -алгебра, содержащая $\mathcal{E}$. Если монотонный класс $\mathcal{M}$ удовлетворяет условию $\mathcal{E} \subset \mathcal{M} \subset \sigma(\mathcal{E})$, то $\mathcal{M} = \sigma(\mathcal{E})$.

³¹ Доказательство (чуть-чуть в другой формулировке) можно найти в [1, гл. II §7] или в [2, гл. IV, §46, утв. 46.4 и 46.5].

³² Почему?

³³ Проверьте это.

³⁴ Другие варианты этой теоремы вместе с обсуждением см. в [1, гл. II §2].

Доказательство. Прежде всего, можно ограничиться случаем, когда $\mathcal{M}$ — минимальный монотонный класс, содержащий $\mathcal{E}$. Кроме того, согласно Лемме 4 нам достаточно показать, что $\mathcal{M}$ является алгеброй.

Действительно, $E \in \mathcal{M}$ так как $\mathcal{M}$ содержит алгебру $\mathcal{E}$. Докажем, что $\mathcal{M}$ замкнут относительно дополнений. Обозначим

$$\mathcal{M}^* = \{A \subset E : A \in \mathcal{M} \text{ и } A^c \in \mathcal{M}\}$$

и докажем, что $\mathcal{M}^* = \mathcal{M}$. По определению $\mathcal{M}^* \subset \mathcal{M}$. Поэтому (так как $\mathcal{M}$ минимальный монотонный класс) нам достаточно доказать, что $\mathcal{M}^*$ является монотонным классом.

Пусть $A_i \in \mathcal{M}^*$ и $A_i \uparrow A$. Тогда $A \in \mathcal{M}$ (так как $A_i \in \mathcal{M}$ и $\mathcal{M}$ — монотонный класс). Кроме того, $A_i^c \downarrow A^c \in \mathcal{M}$, так как $A_i^c \in \mathcal{M}$ и $A_i^c \supset A_{i+1}^c$. Тем самым $A \in \mathcal{M}^*$ и $\mathcal{M}^*$ — монотонный класс.

Осталось доказать, что $\mathcal{M}$ замкнут относительно конечных объединений. Доказательство будет проходить в два аналогичных этапа. Сначала возьмем $A \in \mathcal{E}$, обозначим

$$\mathcal{M}_A = \{B \in \mathcal{M} : A \cup B \in \mathcal{M}\}$$

и покажем, что $\mathcal{M}_A = \mathcal{M}$.

Очевидно, $\mathcal{M}_A \subset \mathcal{M}$ (это по определению) и $\mathcal{M}_A \supset \mathcal{E}$ (так как $\mathcal{E}$ — алгебра). Кроме того, $\mathcal{M}_A$ — монотонный класс. Действительно, пусть $B_n \in \mathcal{M}_A$ и $B_n \uparrow B$. Так как $A \cup B_n \uparrow A \cup B$ и $A \cup B_n \in \mathcal{M}$, то $A \cup B \in \mathcal{M}$ (для убывающих $B_n \in \mathcal{M}_A$ все аналогично).

Следовательно, так как $\mathcal{M}$ — минимальный монотонный класс, $\mathcal{M}_A = \mathcal{M}$ и $A \cup B \in \mathcal{M}$ при любых $A \in \mathcal{E}$ и $B \in \mathcal{M}$.

Пусть теперь $A \in \mathcal{M}$, и $\mathcal{M}^A = \{B \in \mathcal{M} : A \cup B \in \mathcal{M}\}$. Для завершения доказательства теоремы достаточно показать, что $\mathcal{M}^A = \mathcal{M}$. Это происходит повторением предыдущих рассуждений: $\mathcal{M}^A \subset \mathcal{M}$, $\mathcal{M}^A \supset \mathcal{E}$ (для доказательства этого факта и нужен первый этап) и $\mathcal{M}^A$ — монотонный класс. $\square$

Следующее утверждение, важное для работы с УМО, иллюстрирует стиль применения теоремы о монотонном классе.³⁵

Предложение 10. Пусть $(\Omega, \mathcal{F}, P)$ — некоторое вероятностное пространство, $\xi : (\Omega, \mathcal{F}) \mapsto (D_1, \mathcal{D}_1)$, $T : (\Omega, \mathcal{F}) \mapsto (D_2, \mathcal{D}_2)$ и $f : (D_1 \times D_2, \mathcal{D}_1 \times \mathcal{D}_2) \mapsto (\mathbb{R}, \mathcal{B}_R)$. Если $E|f(\xi, T)| < \infty$, то

$$E(f(\xi, T) | T = t) = E(f(\xi, t) | T = t) \quad \mathcal{P}_T\text{-п.в.} \quad (2.7.1)$$

Доказательство. Согласно (2.4.6), равенство (2.7.1) верно в случае, когда $f(x, t) = f_1(x)f_2(t)$ (и выполняются естественные условия суммируемости, необходимые для (2.4.6)). В частности, это имеет место при $f(x, t) = \mathbb{I}_{A \times B}(x, t)$, где $A \in \mathcal{D}_1$ и $B \in \mathcal{D}_2$. Отсюда сразу же следует,³⁶ что (2.7.1) выполняется для $f(x, t) = \mathbb{I}_{C_0}(x, t)$, где C_0 — любой элемент алгебры $\mathcal{E}$, порожденной «ячейками» $A \times B$ с $A \in \mathcal{D}_1$ и $B \in \mathcal{D}_2$.

Докажем, что (2.7.1) имеет место для $f(x, t) = \mathbb{I}_C(x, t)$, где $C \in \mathcal{D}_1 \times \mathcal{D}_2$. Иначе говоря, докажем равенство

$$P((\xi, T) \in C | T = t) = P((\xi, t) \in C | T = t) \quad \mathcal{P}_T\text{-п.в.,}$$

которое эквивалентно тому, что

$$P((\xi, T) \in C, T \in B) = \int_B P((\xi, t) \in C | T = t) \mathcal{P}_T(dt) \quad (2.7.2)$$

³⁵То есть: если некоторое свойство выполняется для элементов алгебры (например, для алгебры ячеек в $\mathbb{R}^d$ — это обычно несложно проверяется) и обладает свойствами монотонности, то оно выполняется и для всех элементов соответствующей σ -алгебры (например, для борелевской σ -алгебры в $\mathbb{R}^d$).

³⁶Проверьте!

для любого $B \in \mathcal{D}_2$. Мы уже знаем, что (2.7.2) верно при $C \in \mathcal{E}$. С другой стороны, по определению $\mathcal{D}_1 \times \mathcal{D}_2 = \sigma(\mathcal{E})$. Поэтому, согласно Теореме 4 достаточно показать, что семейство $\mathcal{M}$ множеств C , для которых выполнено (2.7.2), является монотонным классом.

Поскольку, как нетрудно видеть,³⁷ $\mathcal{M}$ замкнуто относительно дополнений, нужно разобраться лишь с монотонностью «по возрастанию». Здесь тоже нет проблем: левая часть (2.7.2) замкнута относительно предельного перехода при $C_n \uparrow C$ по свойству меры, а правая — согласно Предложению 2 о предельном переходе под знаком УМО.

Дальнейшее стандартно (см. доказательство первого утверждения Предложения 3): в равенстве (2.7.1) мы переходим от индикаторов (для них оно доказано) к простым функциям f , затем (снова предельный переход под знаком УМО) — к неотрицательным суммируемым функциям и, наконец, избавляемся от условия неотрицательности. $\square$

Следствие 2. Если в условиях Предложения 10 существует регулярный вариант $Q(dx; t)$ условного распределения ξ при условии T , то

$$E(f(\xi, T) | T = t) = \int_{D_1} f(x, t) Q(dx; t) \quad \mathcal{P}_T\text{-п.в.} \quad (2.7.3)$$

и

$$Ef(\xi, T) = \int_{D_2} \mathcal{P}_T(dt) \int_{D_1} f(x, t) Q(dx; t). \quad (2.7.4)$$

Доказательство. Для доказательства (2.7.3) используем (2.7.1). Правая часть последнего равенства имеет вид $E(g(\xi) | T = t)$, где (при фиксированном t) $g(\xi) = f(\xi, t)$. Применяя (2.6.2), получаем (2.7.3). Что касается (2.7.4), то это равенство следует из (2.6.3). $\square$

³⁷ Действительно?

3 Гауссовская регрессия и условные гауссовские распределения

3.1 Гауссовская регрессия

Рассмотрим вероятностное пространство $(\Omega, \mathcal{F}, P)$ и гауссовский вектор

$$\bar{\zeta} = \begin{pmatrix} \bar{\xi} \\ \eta \end{pmatrix},$$

где $\eta \in \mathbb{R}$, а $\bar{\xi} = (\xi_1, \dots, \xi_d)^T \in \mathbb{R}^d$. Обозначим $\bar{m}_\xi = E\bar{\xi} = (m_1, \dots, m_d)^T$ математическое ожидание вектора $\bar{\xi}$ и m_η — математическое ожидание случайной величины η .

Теорема 5. Существуют такие постоянные $a_1, \dots, a_d$ и b , что почти наверное

$$E(\eta | \xi_1, \dots, \xi_d) = a_1 \xi_1 + \dots + a_d \xi_d + b. \quad (3.1.1)$$

Доказательство. Предположим сначала, что $m_1 = \dots = m_d = m_\eta = 0$.

Обозначим $\mathbb{L}^2 = \mathbb{L}^2_{\mathcal{F}}(P)$ гильбертово пространство случайных величин, суммируемых с квадратом по мере P , $\mathbb{L}^2_{\bar{\xi}} \subset \mathbb{L}^2$ — замкнутое линейное подпространство этого пространства, состоящее из функций, зависящих только от $\xi_1, \dots, \xi_d$ и, наконец, $\mathcal{L}^2_{\bar{\xi}} — замкнутое линейное подпространство $\mathbb{L}^2_{\bar{\xi}}$ с образующими $\mathbf{1}, \xi_1, \dots, \xi_d$.³⁸ Конечно, скалярное произведение элементов $\gamma_1, \gamma_2 \in \mathbb{L}^2$ определяется как $E\gamma_1\gamma_2$.$

Обозначим $\pi_{\bar{\xi}}$ оператор ортогонального проектирования из $\mathbb{L}^2$ на $\mathcal{L}^2_{\bar{\xi}}$. Тогда существуют такие постоянные $a_1, \dots, a_d$ и b , что $\pi_{\bar{\xi}}(\eta) = a_1 \xi_1 + \dots + a_d \xi_d + b$, причем $\eta - \pi_{\bar{\xi}}(\eta)$ ортогонально $\mathbf{1}, \xi_1, \dots, \xi_d$. Из ортогональности $\phi \stackrel{\text{def}}{=} \eta - \pi_{\bar{\xi}}(\eta)$ единице сразу же следует, что $E\phi = 0$ и

$$0 = E\eta = a_1 E\xi_1 + \dots + a_d E\xi_d + b = b,$$

то есть $b = 0$.

Далее, случайный вектор $(\phi, \xi_1, \dots, \xi_d)^T$ является гауссовским как линейное преобразование гауссовского вектора $\bar{\zeta}$. При этом случайная величина ϕ ортогональна всем ξ_i и (так как $E\eta = E\xi_i = 0$), некоррелирована с ξ_i при всех i .

Это означает, что ϕ и $\xi_1, \dots, \xi_d$ независимы. Следовательно, ϕ независима с любой случайной величиной вида $\varphi(\xi_1, \dots, \xi_d)$. Если дополнительно потребовать, чтобы $E\varphi^2(\xi_1, \dots, \xi_d) < \infty$, то отсюда будет следовать, что ϕ ортогональна $\varphi(\xi_1, \dots, \xi_d)$. Действительно, в этом случае

$$E\phi\varphi(\xi_1, \dots, \xi_d) = E\phi E\varphi(\xi_1, \dots, \xi_d) = 0,$$

так как $E\phi = 0$. Поэтому случайная величина $\pi_{\bar{\xi}}(\eta) = a_1 \xi_1 + \dots + a_d \xi_d$ является ортогональной проекцией η на линейное пространство $\mathbb{L}^2_{\bar{\xi}}$. Согласно Предложению 4 и Замечанию 10, такая ортогональная проекция и является условным математическим ожиданием $E(\eta | \xi_1, \dots, \xi_d)$.

Теперь нам осталось избавиться от предположения о том, что $E\xi_1 = \dots = E\xi_d = E\eta = 0$. Действительно, так как σ -алгебра, порожденная случайными величинами $\xi_1, \dots, \xi_d$ совпадает³⁹ с σ -алгеброй, порожденной случайными величинами $\xi_1 - m_1, \dots, \xi_d - m_d$, то

$$\begin{aligned} E(\eta | \xi_1, \dots, \xi_d) &= E(\eta - m_\eta | \xi_1 - m_1, \dots, \xi_d - m_d) + m_\eta = \\ &= a_1(\xi_1 - m_1) + \dots + a_d(\xi_d - m_d) + m_\eta = a_1 \xi_1 + \dots + a_d \xi_d + b, \end{aligned}$$

где $b = m_\eta - (a_1 m_1 + \dots + a_d m_d)$. □

³⁸Здесь мы для простоты изложения отождествляем случайные величины с их классами эквивалентности, являющимися элементами $\mathbb{L}^2$. Заметим, что это вполне согласуется с определением условного математического ожидания, которое единственно P -почти всюду. Под $\mathbf{1}$ имеется ввиду функция, тождественно равная единице.

³⁹Почему?

Замечание 17. Ортогональную проекцию η на линейное пространство $\mathcal{L}_\xi^2$ обычно называют *линейной регрессией η на $\xi_1, \dots, \xi_d$* . Поэтому Теорему 5 можно кратко сформулировать следующим образом: *гауссовская регрессия является линейной*.

Замечание 18. Тот факт, что регрессия совпадает с линейной регрессией, отнюдь не характеризует гауссовские распределения. Например, рассмотрим произвольные суммируемые независимые случайные величины $\xi_1, \dots, \xi_n$ с нулевыми математическими ожиданиями и положим $S_i = \xi_1 + \dots + \xi_i$. Тогда⁴⁰

$$E(S_n | S_1, \dots, S_{n-1}) = S_{n-1}$$

и, поскольку $\sigma(S_1, \dots, S_{n-1}) = \sigma(\xi_1, \dots, \xi_{n-1})$,⁴¹ то

$$E(S_n | \xi_1, \dots, \xi_{n-1}) = \xi_1 + \dots + \xi_{n-1}.$$

Если потребовать дополнительно существование у случайных величин ξ_i конечность вторых моментов, то как раз и получится⁴² нужный нам пример.

Хотя этот пример и выглядит несколько экзотично, он отнюдь не единственный в своем роде.⁴³

Остановимся теперь на явном виде коэффициентов $a_1, \dots, a_d$ (как уже показано, свободный член b выражается через a_i и m_i). Для этого снова достаточно⁴⁴ рассмотреть случай $m_1 = \dots = m_d = m_\eta = 0$. Обозначим Σ_ζ ковариационную матрицу вектора $\bar{\zeta}$, Σ_ξ ковариационную матрицу вектора $\bar{\xi} = (\xi_1, \dots, \xi_d)^T$, и $\bar{\beta} = (\beta_1, \dots, \beta_d)^T$ — вектор из ковариаций $\beta_i = E\eta \xi_i$. Ясно, что

$$\Sigma_\zeta = \begin{pmatrix} \Sigma_\xi & \bar{\beta} \\ \bar{\beta}^T & \sigma_\eta^2 \end{pmatrix} \quad (3.1.2)$$

где $\sigma_\eta^2 = D\eta$.

Из того, что $\eta - (a_1\xi_1 + \dots + a_d\xi_d)$ ортогональна $\xi_1, \dots, \xi_d$, сразу же следует, что

$$\beta_i = E\eta \xi_i = \sum_{k=1}^d a_k E\xi_k \xi_i, \quad i = 1, \dots, d. \quad (3.1.3)$$

Обозначая $\bar{a} = (a_1, \dots, a_d)^T$ и переписывая систему (3.1.3) в матричном виде, получим, что коэффициенты a_i удовлетворяют равенству

$$\bar{\beta} = \Sigma_\xi \bar{a}. \quad (3.1.4)$$

Дальнейшее зависит от ковариационной матрицы Σ_ξ . Если она невырождена, то единственное решение системы (3.1.3) имеет вид $\bar{a} = \Sigma_\xi^{-1} \bar{\beta}$, а

$$E(\eta | \xi_1, \dots, \xi_d) = (\bar{a}, \bar{\xi}) = \bar{a}^T \bar{\xi} = \bar{\beta}^T \Sigma_\xi^{-1} \bar{\xi}. \quad (3.1.5)$$

Если же $\det \Sigma_\xi = 0$, то ситуация оказывается немного более хитрой. Дело в том, что в этом случае случайные величины $\xi_1, \dots, \xi_d$ оказываются линейно зависимыми⁴⁵ и, хотя проекция $\pi_\xi(\eta)$ по-прежнему определена однозначно, ее выражение через образующие⁴⁶ $\xi_1, \dots, \xi_d$ пространства $\mathcal{L}_\xi^2$ однозначным не является.

⁴⁰ А почему?

⁴¹ А это очевидно?

⁴² А почему?

⁴³ А не связано ли это с понятием мартингалов?

⁴⁴ Почему?

⁴⁵ А почему?

⁴⁶ Но не базис!

Если нужно найти явное выражение для $\bar{a}$, то общий путь для этого может выглядеть следующим образом. В пространстве $\mathcal{L}_\xi^2$ ищется базис (обозначим его $\xi'_1, \dots, \xi'_p$ с $p < d$) и далее мы переходим к вычислению регрессии $E(\eta | \xi'_1, \dots, \xi'_p)$ с невырожденной ковариационной матрицей Σ'_ξ вектора $(\xi'_1, \dots, \xi'_p)^T$. Выбор подходящего базиса, конечно, не является однозначным и зависит от решаемой (практической) задачи. Например, в некоторых⁴⁷ ситуациях он может быть обусловлен интерпретационными соображениями.

Существует, однако, и стандартный способ выбора этого базиса. Он связан с так называемыми «псевдообратными» матрицами. Рассмотрим собственные числа⁴⁸ $\lambda_1, \dots, \lambda_d$ матрицы Σ_ξ . Обозначим $U_1, \dots, U_d$ соответствующую ортонормированную систему собственных векторов. Тогда⁴⁹ $\Sigma_\xi = \sum_{i=1}^d \lambda_i U_i U_i^T$. Матрица

$$\Sigma_\xi^- = \sum_{i: \lambda_i \neq 0} \frac{1}{\lambda_i} U_i U_i^T$$

называется⁵⁰ *псевдообратной* к матрице Σ_ξ .

Нетрудно видеть, что в невырожденном случае $\Sigma_\xi^- = \Sigma_\xi^{-1}$. Действительно, тогда Σ_ξ^- является невырожденной⁵¹ и, поскольку все λ_i не равны нулю, то

$$\Sigma_\xi^- \Sigma_\xi = \left(\sum_{i=1}^d \frac{1}{\lambda_i} U_i U_i^T \right) \left(\sum_{j=1}^d \lambda_j U_j U_j^T \right) = \sum_{i,j=1}^d \frac{\lambda_j}{\lambda_i} U_i U_i^T U_j U_j^T = \sum_{i=1}^d U_i (U_i^T U_i) U_i^T = \sum_{i=1}^d U_i U_i^T = \mathbf{I}_d,$$

где $\mathbf{I}_d$ — единичная $d \times d$ матрица.⁵²

Предложение 11. Вектор $\bar{a}_0 \stackrel{\text{def}}{=} \Sigma_\xi^- \bar{\beta}$ удовлетворяет равенству (3.1.4). При этом

$$\|\bar{a}_0\| = \min \{ \|\bar{a}\| : \Sigma_\xi \bar{a} = \bar{\beta} \}. \quad (3.1.6)$$

Доказательство. Прежде всего, для любого вектора $\bar{c} = (c_1, \dots, c_d)^T$

$$\Sigma_\xi \bar{c} = \sum_{i: \lambda_i \neq 0} \lambda_i U_i U_i^T \bar{c} = \sum_{i: \lambda_i \neq 0} \alpha_i U_i, \quad (3.1.7)$$

где $\alpha_i = \lambda_i (U_i, \bar{c})$. Следовательно, вектор $\bar{\beta}$ в правой части (3.1.4) имеет вид $\bar{\beta} = \sum_{i: \lambda_i \neq 0} \gamma_i U_i$ с некоторыми $\gamma_j \in \mathbb{R}$, причем эти γ_j не зависят от выбора $\bar{a}$ в правой части (3.1.4).⁵³ Поэтому

$$\sum_{j: \lambda_j \neq 0} U_j U_j^T \bar{\beta} = \sum_{j: \lambda_j \neq 0} U_j U_j^T \left(\sum_{i: \lambda_i \neq 0} \gamma_i U_i \right) = \sum_{i,j: \lambda_i, \lambda_j \neq 0} \gamma_i U_j U_j^T U_i = \sum_{i: \lambda_i \neq 0} \gamma_i U_i = \bar{\beta}.$$

Наконец,

$$\Sigma_\xi \Sigma_\xi^- \bar{\beta} = \left(\sum_{i=1}^d \lambda_i U_i U_i^T \right) \left(\sum_{j: \lambda_j \neq 0} \frac{1}{\lambda_j} U_j U_j^T \right) \bar{\beta} = \sum_{i,j: \lambda_j \neq 0} \frac{\lambda_i}{\lambda_j} U_i U_i^T U_j U_j^T \bar{\beta} = \sum_{j: \lambda_j \neq 0} U_j U_j^T \bar{\beta} = \bar{\beta}, \quad (3.1.8)$$

то есть $a = \Sigma_\xi^- \bar{\beta}$ удовлетворяет равенству (3.1.4).

Далее, любое решение $\bar{a}$ системы (3.1.4) имеет вид $\bar{a}_0 + \bar{a}_*$, где $\bar{a}_*$ удовлетворяет условию $\Sigma_\xi \bar{a}_* = \mathbf{0}$. Из этого равенства и (3.1.7) сразу же выводится,⁵⁴ что $\bar{a}_*$ представляет собой линейную комбинацию векторов U_j , соответствующих $\lambda_j = 0$. Поскольку $\bar{a}_0$ является линейной комбинацией тех U_j , для которых $\lambda_j \neq 0$, то $\bar{a}_0$ и $\bar{a}_*$ ортогональны. Поэтому $\|\bar{a}\|^2 = \|\bar{a}_0\|^2 + \|\bar{a}_*\|^2$ и утверждение доказано. $\square$

⁴⁷Часто — статистических

⁴⁸с учетом их кратности

⁴⁹Проверьте!

⁵⁰Общее определение и свойства псевдообратных матриц (к любой прямоугольной матрице) можно найти, например, в [3, §3.4].

⁵¹Почему?

⁵²А почему из этого равенства следует, что $\Sigma_\xi^- = \Sigma_\xi^{-1}$?

⁵³Почему?

⁵⁴Как?

Замечание 19. Таким образом, в вырожденном случае вектор $\bar{a}_0 = \Sigma_{\xi}^{-} \bar{\beta}$ не только задает одно из представлений проекции η на $\mathbb{L}_{\xi}^2$, но и имеет наименьшую длину из всех пригодных для этого векторов $\bar{a}$. Это свойство, конечно, может быть очень важно для приложений.

Использование псевдообратной матрицы приводит к равенству

$$E(\eta | \xi_1, \dots, \xi_d) = \bar{a}_0^T \bar{\xi} = \bar{\beta}^T \Sigma_{\xi}^{-} \bar{\xi} \quad (3.1.9)$$

вместо (3.1.5).⁵⁵

Следствие 3. Рассмотрим теперь случай, когда гауссовский вектор $\bar{\zeta} \in \mathbb{R}^{d+k}$ имеет вид

$$\bar{\zeta} = \begin{pmatrix} \bar{\xi} \\ \bar{\eta} \end{pmatrix},$$

где $\bar{\xi} = (\xi_1, \dots, \xi_d)^T \in \mathbb{R}^d$, а $\bar{\eta} = (\eta_1, \dots, \eta_k)^T \in \mathbb{R}^k$. Нас будет интересовать условное математическое ожидание

$$E(\bar{\eta} | \bar{\xi}) \stackrel{\text{def}}{=} (E(\eta_1 | \bar{\xi}), \dots, E(\eta_k | \bar{\xi}))^T. \quad (3.1.10)$$

Обозначим $\Sigma_{\xi\xi}$ ковариационную матрицу вектора $\bar{\xi}$, $\Sigma_{\eta,\eta}$ — ковариационную матрицу вектора $\bar{\eta}$ и $\Sigma_{\xi\eta}$ — $d \times k$ -матрицу, состоящую из ковариаций $\text{Cov}(\xi_i, \eta_j)$ ($1 \leq i \leq d, 1 \leq j \leq k$). Аналогичным образом пусть $\Sigma_{\eta\xi} = \Sigma_{\xi\eta}^T$ — это $k \times d$ -матрица, состоящая из ковариаций $\text{Cov}(\eta_j, \xi_i)$.

Как и раньше, будем считать, что случайные вектора $\bar{\xi}$ и $\bar{\eta}$ имеют нулевые средние. Кроме того, для определенности будем использовать формулу (3.1.9), не предполагая заранее что матрица $\Sigma_{\xi\xi}$ является невырожденной.

Поскольку условное математическое ожидание (3.1.10) определяется покомпонентно, то, применяя равенство (3.1.9) к каждой компоненте, мы получим формулу⁵⁶

$$E(\bar{\eta} | \bar{\xi}) = \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-} \bar{\xi}. \quad (3.1.11)$$

Пример 4. Приведем один пример, формально не относящийся ни к гауссовским распределениям, ни к вероятности вообще.

Рассмотрим измеримое пространство $(D, \mathcal{A})$ с конечной мерой μ и пространство $\mathbb{L}_{\mu}^2$ функций, определенных на D , измеримых относительно σ -алгебры $\mathcal{A}$ и суммируемых с квадратом по мере μ .

Пусть $f, \varphi_1, \dots, \varphi_d$ — элементы $\mathbb{L}_{\mu}^2$, при этом дополнительно предположим, что функции φ_i линейно независимы.⁵⁷ Поставим задачу наилучшей аппроксимации в $\mathbb{L}_{\mu}^2$ функции f с помощью линейных комбинаций постоянной функции $\varphi_0 = \mathbf{1}$ и функций φ_i .

Иначе говоря, нам нужно найти такие числа $a_1, \dots, a_d$ и b , чтобы достигался минимум выражения

$$\int_D (f - (a_1 \varphi_1 + a_2 \varphi_2 + \dots + a_d \varphi_d + b))^2 d\mu.$$

Такая постановка задачи характерна для так называемого Метода Наименьших Квадратов (МНК), который в общем виде сводится к нахождению параметров $\theta_1, \dots, \theta_{d+1}$ минимизирующих норму разности

$$\|f(x) - \varphi(x; \theta_1, \dots, \theta_{d+1})\|$$

в некотором нормированном пространстве, причем функция f считается известной, функция φ — известной с точностью до параметров $\theta_1, \dots, \theta_{d+1}$, а сам вектор $(\theta_1, \dots, \theta_{d+1})$ принадлежит параметрическому множеству Θ .

⁵⁵А какому базису в $\mathcal{L}_{\xi}^2$ соответствует выбор $\bar{a} = \Sigma_{\xi}^{-} \bar{\beta}$?

⁵⁶Проверьте! А как будет выглядеть эта формула без предположения о нулевых математических ожиданиях?

⁵⁷Точнее, линейно независимы почти всюду по мере μ .

3.2 Условные гауссовские распределения

Все явные формулы предыдущего раздела для вычисления гауссовской регрессии основаны на том, что ортогональная проекция η на линейное пространство случайных величин, измеримых относительно $\xi_1, \dots, \xi_d$ и суммируемых с квадратом совпадает с ортогональной проекцией η на линейное конечномерное пространство с образующими $1, \xi_1, \dots, \xi_d$.

Иными словами, в результате мы решали задачу о минимизации выражения $E(\eta - \sum_{i=1}^d a_i \xi_i - b)^2$ по всевозможным наборам $(a_1, \dots, a_d, b)$. Конечно, решение такой задачи не зависит от того, имеет ли вектор $\bar{\zeta}$ гауссовское распределение или нет, и определяется только первыми двумя моментами этого вектора. Как только мы переходим к условным гауссовским распределениям, ситуация в корне меняется.⁵⁸

Предложение 12. Регулярный вариант условного распределения $\mathcal{L}(\bar{\eta} | \bar{\xi} = \bar{x})$ в условиях Следствия 3 имеет вид

$$\mathcal{L}(\bar{\eta} | \bar{\xi} = \bar{x}) = N_k \left(E(\bar{\eta} | \bar{\xi} = \bar{x}), \Sigma_{\eta\eta} - \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \Sigma_{\xi\eta} \right). \quad (3.2.1)$$

Доказательство. Рассмотрим случайный вектор $\bar{\mathcal{E}} = \bar{\eta} - E(\bar{\eta} | \bar{\xi})$. Этот вектор является гауссовским (как линейное преобразование вектора $\bar{\zeta}$), имеет нулевое среднее и не зависит от $\bar{\xi}$ (так как каждая его компонента ортогональна $\xi_1, \dots, \xi_d$).

Поэтому для любого борелевского k -мерного множества A

$$\begin{aligned} P(\bar{\eta} \in A | \bar{\xi} = \bar{x}) &= P(\bar{\eta} - E(\bar{\eta} | \bar{\xi}) + E(\bar{\eta} | \bar{\xi}) \in A | \bar{\xi} = \bar{x}) = P(\bar{\eta} - E(\bar{\eta} | \bar{\xi}) + E(\bar{\eta} | \bar{\xi} = \bar{x}) \in A | \bar{\xi} = \bar{x}) = \\ &= P(\bar{\mathcal{E}} + E(\bar{\eta} | \bar{\xi} = \bar{x}) \in A | \bar{\xi} = \bar{x}) = P(\bar{\mathcal{E}} + E(\bar{\eta} | \bar{\xi} = \bar{x}) \in A). \end{aligned}$$

Тем самым мы доказали, что условное распределение $\mathcal{L}(\bar{\eta} | \bar{\xi} = \bar{x})$ является гауссовским и имеет среднее $E(\bar{\eta} | \bar{\xi} = \bar{x})$. Все, что осталось, это сосчитать ковариационную матрицу $\Sigma_{\mathcal{E}}$ вектора $\bar{\mathcal{E}}$.

Поскольку (см. (3.1.11)) $E(\bar{\eta} | \bar{\xi}) = \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \bar{\xi}$, то

$$\begin{aligned} \Sigma_{\mathcal{E}} &= E\bar{\mathcal{E}}\bar{\mathcal{E}}^T = E \left(\bar{\eta} - \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \bar{\xi} \right) \left(\bar{\eta} - \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \bar{\xi} \right)^T = \\ &= E\bar{\eta}\bar{\eta}^T - E\bar{\eta} \left(\Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \bar{\xi} \right)^T - E \left(\Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \bar{\xi} \right) \bar{\eta}^T + E \left(\Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \bar{\xi} \right) \left(\Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \bar{\xi} \right)^T = \\ &= \Sigma_{\eta\eta} - \left(E\bar{\eta} \bar{\xi}^T \right) \Sigma_{\xi\xi}^{-1} \Sigma_{\xi\eta} - \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \left(E\bar{\xi} \bar{\eta}^T \right) + \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} E \left(\bar{\xi} \bar{\xi}^T \right) \Sigma_{\xi\xi}^{-1} \Sigma_{\xi\eta} = \\ &= \Sigma_{\eta\eta} - \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \Sigma_{\xi\eta} - \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \Sigma_{\xi\eta} + \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \Sigma_{\xi\xi} \Sigma_{\xi\xi}^{-1} \Sigma_{\xi\eta} = \Sigma_{\eta\eta} - \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \Sigma_{\xi\eta}, \end{aligned}$$

причем в последнем переходе используется равенство $\Sigma_{\xi\xi} \Sigma_{\xi\xi}^{-1} \Sigma_{\xi\eta} = \Sigma_{\xi\eta}$, которое следует из (3.1.8), так как каждый столбец матрицы $\Sigma_{\xi\eta}$ является линейной комбинацией тех векторов U_i , для которых $\lambda_i \neq 0$. Утверждение доказано. $\square$

Замечание 20. При доказательстве Предложения 12 использовался тот факт, что средние векторов $\bar{\xi}$ и $\bar{\eta}$ равны нулю. На самом деле, как нетрудно понять, формула (3.2.1) верна и без этого предположения.⁵⁹

⁵⁸Это обстоятельство объясняет, с одной стороны, частое применение линейных методов в многомерной статистике (все выражения определяются только первыми и вторыми моментами, которые, в принципе, хорошо оцениваются) и, с другой стороны, популярность предположения о гауссовском (генеральном) распределении многомерных выборок в теоретических исследованиях — все формулы будут такими же, как и без этого предположения, зато появляется возможность строить доверительные области, проверять статистические гипотезы и пр.

⁵⁹Проверьте!

Пример 5. Рассмотрим частный случай $d = k = 1$ и найдем вид функции регрессии $E(\eta | \xi = x)$ и условного распределения $\mathcal{L}(\eta | \xi = x)$ для двумерного случайного гауссовского вектора $\bar{\zeta} = (\xi, \eta)^T$ со средним $E\bar{\zeta} = (m_\xi, m_\eta)^T$ и ковариационной матрицей

$$\Sigma_{\bar{\zeta}} = \begin{pmatrix} \sigma_\xi^2 & \rho\sigma_\xi\sigma_\eta \\ \rho\sigma_\xi\sigma_\eta & \sigma_\eta^2 \end{pmatrix},$$

так что $\sigma_\xi^2 = D\xi$, $\sigma_\eta^2 = D\eta$ и $\rho = \rho(\xi, \eta)$. В общих обозначениях ковариационной матрицы (3.1.2) это означает, что ковариационная матрица Σ_ξ является одномерной и равной σ_ξ^2 , а вектор $\bar{\beta}$ тоже одномерен и равен $\rho\sigma_\xi\sigma_\eta$.

Предположение о том, что матрица Σ_ξ невырождена, тем самым сводится к условию $\sigma_\xi^2 > 0$. Будем дальше это предполагать.⁶⁰ Конечно, $\Sigma_\xi^{-1} = 1/\sigma_\xi^2$.

Теперь, используя формулу (3.1.5) в ее очевидном варианте с ненулевыми средними, получим, что

$$E(\eta | \xi = x) = \bar{\beta}^T \Sigma_\xi^{-1} (x - m_\xi) + m_\eta = \rho\sigma_\xi\sigma_\eta \frac{1}{\sigma_\xi^2} (x - m_\xi) + m_\eta = \rho \frac{\sigma_\eta}{\sigma_\xi} (x - m_\xi) + m_\eta.$$

Желающие могут узнать в этой формуле традиционную для статистики формулу (одномерной) линейной регрессии.

Для нахождения условного распределения $\mathcal{L}(\eta | \xi = x)$, конечно же, нужно использовать формулу (3.2.1), причем, так как вид функции регрессии уже найден, то осталось найти выражение для (одномерной) ковариационной матрицы $\Sigma_{\eta\eta} - \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \Sigma_{\xi\eta}$. Это совсем легко: $\Sigma_{\eta\eta} = \sigma_\eta^2$, $\Sigma_{\eta\xi} = \Sigma_{\xi\eta} = \rho\sigma_\xi\sigma_\eta$ и $\Sigma_{\xi\xi}^{-1} = 1/\sigma_\xi^2$, так что

$$\Sigma_{\eta\eta} - \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \Sigma_{\xi\eta} = \sigma_\eta^2 - (\rho\sigma_\xi\sigma_\eta)^2 / \sigma_\xi^2 = \sigma_\eta^2 (1 - \rho^2),$$

и окончательно

$$\mathcal{L}(\eta | \xi = x) = N(E(\eta | \xi = x), \sigma_\eta^2 (1 - \rho^2)).$$

Ясно, что при $|\rho| = 1$ это распределение вырождается в меру Дирака, сосредоточенную в точке $E(\eta | \xi = x)$.⁶¹

3.3 Рекуррентное моделирование гауссовских векторов

Рассмотрим теперь задачу моделирования гауссовского вектора $\bar{\xi} = (\xi_1, \dots, \xi_d)^T$ с заданными вектором средних $\bar{m}_\xi = E\bar{\xi}$ и ковариационной матрицей Σ_d . Обозначим Σ_n ковариационную матрицу вектора $\bar{\xi}_n = (\xi_1, \dots, \xi_n)^T$ и положим $\sigma_n^2 = D\xi_n$, $n = 1, \dots, d$.

Здесь мы будем заниматься только одним — рекуррентным — способом моделирования вектора $\bar{\xi}$. Это означает, что мы будем переходить от размерности n к размерности $n + 1$, и моделировать случайную величину ξ_{n+1} исходя из того, что случайный вектор $\bar{\xi}_n$ уже построен. Кроме того, «источником случайности» мы будем считать независимые случайные величины $\varepsilon_1, \dots, \varepsilon_d$, имеющие стандартное нормальное распределение. Наконец, как и в предыдущем разделе, мы будем использовать матрицы, псевдообратные к Σ_n , не предполагая невырожденности последних.

Прежде всего заметим, что достаточно рассмотреть случай $E\bar{\xi} = \mathbf{0}$, переход к общей ситуации очевиден: если $\bar{\xi}' \in N_d(\mathbf{0}, \Sigma)$ и $\bar{m} \in \mathbb{R}^d$, то $\bar{\xi}' + \bar{m} \in N_d(\bar{m}, \Sigma)$.

Итак, пусть случайный вектор $\bar{\xi}_n$ уже промоделирован. Тогда, согласно (3.1.9), $E(\xi_{n+1} | \bar{\xi}_n) = \bar{\beta}_n^T \Sigma_n^{-1} \bar{\xi}_n$, где $\bar{\beta}_n = E\xi_{n+1} \bar{\xi}_n$. Кроме того (см. Предложение 12), случайная величина $\xi_{n+1} - \bar{\beta}_n^T \Sigma_n^{-1} \bar{\xi}_n$

⁶⁰ А как будет выглядеть гауссовские регрессия и условное распределение, если $\sigma_\xi^2 = 0$?

⁶¹ А как «на пальцах» объяснить этот факт?

не зависит от $\bar{\xi}_n$ и, имеет нормальное распределение со нулевым средним и дисперсией $\Delta_{n+1}^2 = \sigma_{n+1}^2 - \bar{\beta}_n^T \Sigma_n^- \bar{\beta}_n$.

Тогда моделирующая формула для ξ_{n+1} будет выглядеть как

$$\xi_{n+1} = \bar{\beta}_n^T \Sigma_n^- \bar{\xi}_n + \Delta_{n+1} \varepsilon_{n+1}, \quad n = 1, \dots, d-1. \quad (3.3.1)$$

Если сюда добавить инициализацию $\xi_1 = \sigma_1 \varepsilon_1$, то вся процедура рекуррентного моделирования будет описана. Именно такая процедура приведена в [4, §5.2, Лемма 4] в предположении невырожденности матриц Σ_n .

Замечание 21. Конечно, такая процедура является в общем случае очень трудоемкой, так как требует на каждом шаге (псевдо)-обращений матриц со все растущими размерами.⁶² Однако, если правую часть (3.3.1) выразить через ε_i , то мы окажется, что⁶³

$$\xi_{n+1} = a_{n+1,1} \varepsilon_1 + \dots + a_{n+1,n+1} \varepsilon_{n+1}, \quad n = 1, \dots, d-1,$$

где a_{ij} — некоторые постоянные коэффициенты. Вычисление коэффициентов a_{ij} производится рекуррентно, оно связано с разложением Холецкого (см. [4, §5.2, формула 5.2.5]).

Список литературы

- [1] А.Н. Ширяев (2004), Вероятность-1, М., Изд-во МЦНМО.
- [2] К. Парасарати (1980), Введение в теорию вероятностей и теорию меры, М., Мир.
- [3] Г. Стренг (1980), Линейная алгебра и ее применения, М., Мир.
- [4] В. В. Некруткин (2017), Моделирование распределений, <http://statmod.ru/wiki/books:vv>.

⁶²Есть и исключения.

⁶³Проверьте!