Глава 5

Условное математическое ожидание

5.1 Базовая часть

В этом разделе мы не будем касаться математической статистики, а дополним ваши знания теории вероятностей очень важным понятием.

5.1.1 Определение и примеры

Определение УМО

Пусть задано вероятностное пространство $(\Omega, \mathcal{F}, \mathbf{P})$ и случайная величина X на нем, $\mathbf{E}|X| < \infty$. Кроме того, пусть задана сигма-алгебра $\mathcal{A} \subseteq \mathcal{F}$. Условным математическим ожиданием X относительно $\mathcal{A}$ называют случайную величину (!) $Y = \mathbf{E}(X|\mathcal{A})$, такую, что:

1. Y — $\mathcal{A}$-измерима,

2. для любого $B \in \mathcal{A}$ $\mathbf{E}XI_B = \mathbf{E}YI_B$.

Такая величина существует и единственна с точностью до множества меры 0 (т.е. любые две различных ее версии равны п.н.). Это напрямую вытекает из теоремы Радона-Никодима, поскольку заряд

$$Q(A) = \mathbf{E}XI_A$$

абсолютно непрерывен около меры, которую $\mathbf{P}_X$ индуцирует на $\mathcal{A}$. Если вас напугали последние слова – не переживайте, это никак не влияет на восприятие дальнейшего.

УМО в частных случаях

Посмотрим что это за объект для различных сигма-алгебр $\mathcal{A}$.

Пример 1. Если $\mathcal{A} = \{\Omega, \emptyset\}$, то мы ищем величину Y, измеримую около такой сигма-алгебры. Это означает, что прообразы борелевских множеств около Y есть либо Ω, либо $\emptyset$, т.е. $Y = const$. Так как $\mathbf{E}Y = \mathbf{E}X$, имеем $Y = \mathbf{E}X$ п.н. Поэтому в случае $\mathcal{A} = \{\Omega, \emptyset\}$ условное мат. ожидание будет просто мат. ожиданием.
Если $\mathcal{A} = \sigma(X)$, то мы ищем случайную величину Y, измеримую относительно $\sigma(X)$, такую что $\mathbf{E}XI_B = \mathbf{E}YI_B$ для всех $B \in \sigma(X)$. Очевидно, что подходящей величиной является X, в силу единственности, это и будет $\mathbf{E}(X|\mathcal{A})$. Таким образом, математическое ожидание случайной величины относительно порожденной ей сигма-алгебры есть она сама.

Рассмотрим сигма-алгебру $\mathcal{A}$, порожденную разбиением D_i, $i \leq \infty$. Тогда на каждом из D_i величина $Y = \mathbf{E}(X|\mathcal{A})$ должна быть постоянной. Действительно, если бы она принимала несколько значений на множестве D_i, то мы могли бы выбрать одно из них (x) и рассмотреть $B = \{\omega : Y(\omega) = x\}$. Оно должно лежать в $\mathcal{A}$, при этом оно не содержит D_i, но пересекается с ним, чего быть не может. Поэтому наша случайная величина будет принимать некоторое постоянное значение a_i на D_i. В силу второго свойства $\mathbf{E}(XI_{D_i}) = \mathbf{P}(D_i)a_i$, откуда

$$a_i = \mathbf{E}(X|D_i) = \mathbf{E}(XI_{D_i})/\mathbf{P}(D_i).$$

Пример 2. Пусть X_i - независимые бернуллиевские с параметром $1/2$, а $\mathcal{A}$ - сигма-алгебра, порожденная разбиением $D_k = \{\omega : X_1 + X_2 + X_3 = k\}$, $k \leq 3$. Тогда

$$\mathbf{E}(X_1|D_k) = \mathbf{E}(X_1|X_1 + X_2 + X_3 = k) = \sum_{i=0}^{1} i\mathbf{P}(X_1 = i|X_1 + X_2 + X_3 = k) =$$

$$\mathbf{P}(X_2 + X_3 = k-1)\mathbf{P}(X_1 = 1)/\mathbf{P}(X_1 + X_2 + X_3 = k) = C_2^{k-1}/C_3^k = k/3,$$

откуда $\mathbf{E}(X_1|\mathcal{A})$ — сл.в., принимающая значения $k/3$ там, где $X_1 + ... + X_3 = k$. Но это, очевидно, просто $\overline{X}$.

Получается, что имея некоторую функцию от омега, мы приближаем ее кусочно-постоянными функциями, постоянными на множествах разбиения. Условное математическое ожидание — это некоторая аппроксимация случайных величин, измеримых относительно более богатой сигма-алгебре, величинами, измеримыми около более бедной.

5.1.2 Условные распределения

Определение

Зачастую используют запись $\mathbf{E}(X|Z)$, где X, Z сл.в. Эта запись равносильна $\mathbf{E}(X|\sigma(Z))$, где $\sigma(Z)$ сигма-алгебра, порожденная Z, т.е. $\{Z^{-1}(B), B \in \mathcal{B}(R)\}$. Все отображения, измеримые около $\sigma(Z)$, представляют собой борелевские функции $g(Z)$ п.н., поэтому нахождение условного матожидания равносильно поиску функции $g(z)$. Мы будем обозначать эту функцию $\mathbf{E}(X|Z = z)$. При этом $\mathbf{E}(X|Z) = g(Z)$.

Получается, что для пары дискретных величин X, Z мы получаем

$$g(z) = \mathbf{E}(X|Z = z) = \sum_{k=0}^{\infty} x_k \mathbf{P}(X = x_k|Z = z).$$

Определение 1. Условным распределением $X|Z$ называют набор мер

$$\mathbf{P}(X \in A|Z = z) := \mathbf{E}(I_{X \in A}|Z = z)$$

при всех возможных z.

При каждом z — это некоторая мера, при этом

$$\mathbf{E}(X|Z = z) = \int_{\mathbb{R}} x\mathbf{P}(X \in dx|Z = z)$$

Примеры

Теперь изучим абсолютно-непрерывный случай. Если величины X, Z величины с совместной плотностью $f_{X,Z}$, то положим

$$f_{X|Z}(x|z) = \frac{f_{X,Z}(x,z)}{\int_R f_{X,Z}(u,z)du} = \frac{f_{X,Z}(x,z)}{f_Z(z)},$$

(с плотностью такого вида мы уже сталкивались, когда изучали байесовские оценки). Тогда

$$\mathbf{E}(X|Z=z) = \int\limits_{\mathbb{R}} x f_{X|Z}(x|z)dx = g(z),\ Y = E(X|Z) = g(Z).$$

Почему именно так? Первому условию наша величина заведомо удовлетворяет, поскольку она есть функция от Z, а значит $\sigma(Z)$ измерима. Кроме того, если $B \in \sigma(Z)$, то $B = Z^{-1}(C)$ для некоторого $C \in \mathcal{B}(R)$, откуда

$$\mathbf{E}(YI_B) = \int\limits_C g(z) f_Z(z)dz = \int\limits_C \frac{\int_{\mathbb{R}} x f_{X,Z}(x,z)dxdz}{f_Z(z)} f_Z(z)dz = \int\limits_{Z^{-1}(C)} X(\omega)P(X \in d\omega) = \mathbf{E}(XI_B).$$

Одно такое условное матожидание мы уже считали — байесовская оценка была условным математическом ожиданием θ при условии $X_1, ..., X_n$.

Пример 3. Пусть X, Y — независимые $\exp(1)$ сл.в. Найти $\mathbf{E}(X|X/Y)$. Запишем совместную плотность $X, X/Y$. Здесь нам понадобится формула, связывающая плотность $g(\vec{X})$ и $\vec{X}$, где $g : \mathbb{R}^n \to \mathbb{R}^n$ — гладкая.

$$f_{g(\vec{X})}(\vec{x}) = f_{\vec{X}}(g^{-1}(\vec{x}))|J|,$$

где J — Якобиан g^{-1}.
Если $X = u$, $X/Y = v$, то $X = u$, $Y = u/v$. Определитель матрицы Якоби такой замены будет $-u/v^2$, откуда

$$f_{X,X/Y}(u,v) = f_{X,Y}(u, u/v) \cdot \frac{u}{v^2} = \frac{ue^{-u}e^{-u/v}}{v^2}.$$

При этом

$$f_{X/Y}(v) = \int_{\mathbb{R}^+} \frac{u}{v^2} e^{-u(1+1/v)}du = (1+1/v)^{-2}\Gamma(2)/v^2 = 1/(1+v)^2.$$

Значит $f_{X|X/Y}(u|v) = u(1+1/v)^2 e^{-u(1+1/v)}$, откуда условное распределение X при условии $X/Y = v$ является гамма-распределением $\Gamma(2, v/(1+v))$.

Значит математическое ожидание этой величины $2v/(1+v)$, т.е.

$$\mathbf{E}(X|X/Y) = \frac{2X}{Y+X}.$$

5.1.3 Свойства УМО

Базовые свойства

С применениями у.м.о. мы познакомимся позднее. Сейчас изучим ряд свойств у.м.о.:

1. У.м.о. как и м.о. линейно, у.м.о. константы равно этой константе п.н., если $X \geq Y$ п.н., то $\mathbf{E}(X|\mathcal{A}) \geq \mathbf{E}(Y|\mathcal{A})$ п.н.
2. $\mathbf{E}(\mathbf{E}(X|\mathcal{A})) = \mathbf{E}X$. Более того, если $\mathcal{A}_1 \subseteq \mathcal{A}_2$, то $\mathbf{E}(\mathbf{E}(X|\mathcal{A}_1)|\mathcal{A}_2) = \mathbf{E}(\mathbf{E}(X|\mathcal{A}_2)|\mathcal{A}_1) = \mathbf{E}(X|\mathcal{A}_1)$ п.н.
3. $\mathbf{E}(XY|\mathcal{A}) = Y\mathbf{E}(X|\mathcal{A})$ п.н., если Y $\mathcal{A}$-измерима.
4. $\mathbf{E}(X|Y) = \mathbf{E}X$, если X не зависит от Y.

Первое свойство прямо следует из исходного определения, свойства 2, 3, 4 легко осознать с точностью до определения в терминах проекции.

Условной вероятностью события A при условии сигма-алгебры $\mathcal{A}$ называют $\mathbf{P}(A|\mathcal{A}) = \mathbf{E}(I_A|\mathcal{A})$. Для у.м.о. оказываются верными модификации неравенств Чебышева, Маркова, Иенсена, теоремы о монотонной и мажорируемой сходимости.

Пример 4. Пусть X_n – пуассоновские величины с параметром n, N - случайная величина с натуральными значениями, не зависящая от X_n, $\mathbf{E}N^2 < \infty$. Как найти $\mathbf{E}NX_N$? Воспользуемся свойством 2: $\mathbf{E}(NX_N) = \mathbf{E}(\mathbf{E}(NX_N|N)) = \mathbf{E}(N\mathbf{E}(X_N|N))$. Но $\mathbf{E}(X_N|N = k) = \mathbf{E}(X_k|N = k) = \mathbf{E}X_k = k$ (свойство 4). Значит $\mathbf{E}X_N = \mathbf{E}N^2$.

Формула Байеса

Заметим, что часто удобно считать условное распределение $Y|X$ и неудобно $X|Y$, например, так было в примере 4. В случае дискретных величин нам помогает формула Байеса:

$$\mathbf{P}(X = k|Y = l) = \frac{\mathbf{P}(Y = l|X = k)\mathbf{P}(X = k)}{\sum_i \mathbf{P}(Y = l|X = i)\mathbf{P}(X = i)}.$$

В непрерывным она также верна (и столь же очевидна):

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{\mathbb{R}} f_{Y|X}(y|x)f_X(x)dx}$$

Отсюда

$$\mathbf{E}(X|Y = y) = \frac{\int_{\mathbb{R}} xf_{Y|X}(y|x)f_X(x)dx}{\int_{\mathbb{R}} f_{Y|X}(y|x)f_X(x)dx}.$$

Мы уже видели такие рассуждения, когда получили формулу для байесовской оценки в случае квадратичного риска.

Пример 5. С помощью формулы Байеса решим задачу примера 4:
$f_{X/Y|X}(y|x) = f_{1/Y}(y/x)/x = f_Y(x/y)x/y^2$, откуда

$$\mathbf{E}(X|X/Y = y) = \frac{\int_{\mathbb{R}^+} xe^{-x/y}x/y^2e^{-x}dx}{\int_{\mathbb{R}^+} e^{-x/y}x/y^2e^{-x}dx} = \frac{2y}{y+1},$$

что совпадает с ответом, полученным ранее.

5.2 Факультатив

Поиск условного распределения из физических соображений

Итак, мы определили условное математическое ожидание и нашли способы его нахождения в случае дискретных и абсолютно-непрерывных величин. Большую роль для нас при этом имело условное распределение $\mathbf{P}(X \in A|Y = k)$ для дискретных Y или условная плотность $f_{X|Y}(x|y)$, по которым мы и считали наше м.о. В некоторых случаях условное распределение находится из естественных соображений об условном поведении:

Пример 6. Рассмотрим условное распределение величины XY при условии X, где X, Y — независимы и имеют плотности. Но при фиксированном X распределение XY — распределение Y, умноженной на "константу" X, т.е. $f_{XY|X}(y|x) = f_Y(y/x)/x$, где мы воспользовались формулой плотности величины, умноженной на константу. Более формально можно доказать эту формулу из определения УМО.

Пример 7. Пусть X имеет некоторое распределение. Какое распределение имеет $X|X$? Очевидно, что

$$\mathbf{E}(I_{X\in B}|X = x) = I_{x\in B}.$$

Проверим, что это соответствует определению. Мы должны проверить, что при всех $A \in \sigma(X)$ (т.е. $A = \{X \in C\}$ при некотором C):

$$\mathbf{E}(I_{X\in B}; A) = \mathbf{E}(I_{X\in B}; A)$$

Это верно, значит $I_{X\in B}$ действительно $\mathbf{E}(I_{X\in B}|X)$.

Пример 8. Аналогичным образом, если $X \sim \mathcal{N}(0,1)$, то $X|X^2$ равновероятно принимает значения $\pm\sqrt{X^2}$, откуда $\mathbf{E}(X|X^2) = 0$.
Как убедиться в этом с точки зрения определения?
Проверим, что для любого $B \in \mathcal{B}(\mathbb{R})$ и события $A \in \sigma(X^2)$ (т.е. для какого-то $C \in \mathcal{B}(\mathbb{R})$ $A = \{\omega : X^2(\omega) \in C\}$) выполнено соотношение $\mathbf{E}(I_{X\in B}; A) = \mathbf{E}(I_{Y\in B}; A)$, где $Y = |X|Z$, $Z = \pm 1$ равновероятно. Первая вероятность совпадает с $\mathbf{P}(X \in B; A)$, вторая с $(\mathbf{P}(|X| \in B; A) + \mathbf{P}(-|X| \in B; A))/2$. При этом

$$\mathbf{P}(|X| \in B; A) + \mathbf{P}(-|X| \in B; A) = \mathbf{P}(X \in B, A, X > 0) + \mathbf{P}(X \in -B, A, X < 0) + \mathbf{P}(X \in -B, A, X > 0) + \mathbf{P}(X \in B, A, X < 0).$$

В силу симметричности X и симметричности A относительно X, $\mathbf{P}(X \in -B, A, X > 0) = \mathbf{P}(X \in B, A, X < 0)$, откуда получается искомое утверждение.

5.2.1 L^2-интерпретация УМО

Разбирая случай дискретных величин, мы нашли интерпретацию условного мат.ожидания — аппроксимацию случайной величины X борелевскими функциями от Z. Более точно эту интерпретацию дает следующее утверждение.
Пусть L^2 — множество случайных величин с $\mathbf{E}X^2 < \infty$, $(X,Y)_{L^2} = \mathbf{E}XY$ — скалярное произведение на этом пространстве. Рассмотрим проекцию величины X на пространство величин H, измеримых относительно $\mathcal{A}$ (в случае $\mathbf{E}(X|Z)$ — измеримых функций от Z), то есть $Y : X - Y$ ортогонально всем $\tilde{Y} \in H$. Такое Y п.н. единственно и совпадает с у.м.о. X при условии $\mathcal{A}$ (при условии Z) п.н.
Чтобы лучше понять откуда берется такое определение, возьмем более общее определение $\mathbf{E}(X|Y)$ — это такая функция $f(Y)$, что $\mathbf{E}(f(Y)h(Y)) = \mathbf{E}(Xh(Y))$ для всех функций h, для которых математические ожидания существуют. Конечно в определении это требовалось только на индикаторных функциях $h(Y)$, но они порождают все измеримые функции.
В $L^2(P)$ указанное математическое ожидание — это скалярное произведение. Значит $f(Y) - X$ ортогонально $h(Y)$, что и утверждается.

Иногда это определение более удобно для подсчета условного математического ожидания. Рассмотрим классический пример:

Пример 9. Пусть (X,Z) — двумерный нормальный вектор с нулевым вектором средних и матрицей ковариации Σ.
Напомню, что двумерным нормальным вектором называют вектор $\vec{X}$ с х.ф.

$$\psi_{\vec{X}}(\vec{t}) = Ee^{i(\vec{t},\vec{X})} = e^{i(\vec{t},\vec{\mu})}e^{-\frac{1}{2}\vec{t}^{\intercal}\Sigma^2\vec{t}},$$

где $\vec{\mu}$, $\vec{\Sigma}^2$ — заданные параметры распределения (вектор средних и матрица ковариации).
Тогда $\mathbf{E}(X|Z)$ - это такая $g(Z)$, что $X - f(Z)$ ортогонально $h(Z)$ для любой борелевской g.
Заметим, что если $\mathbf{E}(X - f(Z)) = 0$, а $X - f(Z)$ не зависит от Z, то $(X - f(Z), g(Z))_{L^2} = \mathbf{E}(X - f(Z))\mathbf{E}g(Z) = 0$, так что для нахождения $\mathbf{E}(X|Z)$ достаточно найти $f(Z)$, такую что $X - f(Z)$ не зависит от Z и $\mathbf{E}X = \mathbf{E}f(Z)$. Поищем $f(Z)$ в виде $aZ + b$. Тогда $(X - f(Z), Z)$ — нормальный вектор (линейное преобразование нормального вектора) и независимость $X - f(Z)$ и Z равносильна их некореллированности. Отсюда имеем уравнения

$$a\mathbf{E}Z + b = \mathbf{E}X,\ \ \mathbf{E}XZ - a\mathbf{E}Z^2 - b\mathbf{E}Z = 0,$$

откуда $b = 0$, $a = cov(X, Z)/DZ$. Следовательно, $\mathbf{E}(X|Z) = Zcov(X, Z)/DZ$.