Лекция 8. Условное матожидание. Равенство Вальда

1 апреля 2022 г.

1 Условное матожидание как с.в.

Мы до этого рассматривали функции от с.в. и говорили, что они также являются с.в. Например, если есть функция $g(x) : \mathbb{R} \to \mathbb{R}$, то мы можем рассмотреть композицию этой функции и с.в. X, а именно $g(X) : \Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}$.

Вспомним также, что такое условное матожидание. $E[X \mid Y = y]$ – это число, равное матожиданию с.в. X при условии, что с.в. $Y = y$ (*NB:* это условное матожидание может и не существовать или быть бесконечностью). То есть мы можем рассматривать это как функцию от y:

$$g(y) = E[X \mid Y = y]$$

А следовательно, мы можем рассмотреть и композицию этой функции с с.в. Y:

$$g(Y) = E[X \mid Y],$$

что является с.в.

А раз это с.в., то можно говорить о ее матожидании, дисперсии и прочих характеристиках.

2 Матожидание условного матожидания

Мы рассматриваем с.в. $Z = g(Y) = E[X \mid Y]$. Посчитаем ее матожидание в дискретном случае Y.

$$E[E[X \mid Y]] = E[g(Y)] = \sum_y p_Y(y)g(y) = \sum_y p_Y(y)E[X \mid Y = y] = E[X],$$

где последнее равенство — по формуле полной вероятности. То есть мы получили:

$$E[E[X \mid Y]] = E[X]$$

В англоязычной литературе это называется Law of iterated expectations, в русской — просто вариацией теоремы о полном матожидании (что есть правда). Аналогичное равенство можно получить и для непрерывной Y, заменив сумму интеграла, а функцию вероятностей — плотностью.

Вспомним пример про ломание палки из 6-ой лекции (непрерывные с.в, часть 3). Напомним: сначала ломаем палку длины ℓ в случайном месте $Y \sim U(0, \ell)$. Потом берем кусок $[0, Y]$ и опять ломаем его в случайном месте $X \sim U(0, Y)$. И найдем матожидание X следующим образом:

$$E[X] = E[E[X \mid Y]] = E\left[\frac{Y}{2}\right] = \frac{1}{2}E[Y] = \frac{\ell}{4}$$

Рассмотрим еще пример, чтобы лучше понять значение этого утверждения. например, вы сейчас хотите сделать прогноз о том, сколько вы заработаете на акциях в течение мая. Пусть это число будет равно X, и вас интересует $E[X]$. Но к началу мая сумарная стоимость ваших акций Y будет не такой, как сейчас, то есть она тоже будет с.в. И ваш новый прогноз на прибыл будет $E[X \mid Y = y]$. Однако если вернуться в настоящий момент, то для вас прогноз, сделанный в мае также будет являться случайной величиной. При этом $E[X] = E[E[X \mid Y]]$ можно читать так:

> *Матожидание прогноза, сделанного через месяц, равно матожиданию текущего прогноза.*

3 Матожидание условной дисперсии

Для начала ведем понятие условной дисперсии.

$$\mathrm{Var}(X \mid Y = y) = E\left[(X - E[X \mid Y = y])^2 \mid Y = y\right]$$

То есть это вариация, которая посчитана в том мире, где у нас уже произошло событие $Y = y$ и у нас соответсвтующим образом изменилась вероятностная мера для с.в. X. Но это опять функция от Y, поэтому мы можем снова определить с.в. $Z = \mathrm{Var}(X \mid Y)$ как с.в., которая принимает значение $\mathrm{Var}(X \mid Y = y)$, если Y принимает значение y.

Пример с ломанием палки, где $X \sim U(0, Y)$: $\mathrm{Var}(X \mid Y = y) = \frac{y^2}{12}$, при этом $\mathrm{Var}(X \mid Y) = \frac{Y^2}{2}$ — с.в. Но вот полная дисперсия в отличие от полного матожидания устроена сложнее:

$$\mathrm{Var}(X) = E[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(E[X \mid Y])$$

Давайте докажем это. Посчитаем сначала условную дисперсию при известном значении Y:

$$\mathrm{Var}(X \mid Y = y) = E[X^2 \mid Y = y] - (E[X \mid Y = y])^2$$

Так как это равенство выполнено для всех возможных значений y, то можем перейти к равенству с.в. (то есть при каждом элементарном исходе они принимают одинаковые значения).

$$\mathrm{Var}(X \mid Y) = E[X^2 \mid Y] - (E[X \mid Y])^2$$

А раз равны с.в., то равны и их матожидания

$$E[\mathrm{Var}(X \mid Y)] = E[X^2] - E[(E[X \mid Y])^2]$$

Но также мы можем посчитать вариацию с.в. $E[X \mid Y]$:

$$\mathrm{Var}(E[X \mid Y]) = E[(E[X \mid Y])^2] - (E[X])^2$$

Остается сложить последние два равенства, тогда справа будет дисперсия X, а слева то, что указано в формуле полной дисперсии.

4 Примеры, если требуются

Первый пример: (X, Y) такие, что:

- с вероятностью $\frac{1}{2}$ с.в. $Y = 1$, а $X \sim U(0, 1)$
- с вероятностью $\frac{1}{2}$ с.в. $Y = 2$, а $X \sim U(1, 3)$

Можем посчитать:

- $E[X \mid Y]$
- $E[X] = E[E[X \mid Y]]$
- $\mathrm{Var}(X \mid Y)$
- $E[\mathrm{Var}(X \mid Y)]$
- $\mathrm{Var}(E[X \mid Y])$

Более глобальный пример: Y — номер группы, известны размеры групп, X — оценка каждого элемента в группе. Известны матожидания и дисперсии X внутри каждой группы.

Можем посчитать:

- $E[X \mid Y]$

- $E[X] = E[E[X \mid Y]]$
- $\operatorname{Var}(E[X \mid Y])$
- $\operatorname{Var}(X \mid Y)$
- $E[\operatorname{Var}(X \mid Y)]$

Можно интерпретировать формулу полной дисперсии как средняя дисперсия внутри групп плюс дисперсия между группами.

5 Матожидание и дисперсия суммы случайного числа с.в.

Рассмотрим ситуацию: вы заходите в N магазинов, где N — некоторая с.в. В i-ом магазине вы тратите X_i рублей, причем

- Все X_i независимы друг от друга и от N
- Все X_i имеют одинаковое распределение, такое же, как какая-то с.в. X.

Сколько вы ожидаемо потратили? Пусть $Y = \sum_{i=1}^{N} X_i$, тогда

$$E[Y \mid N = n] = nE[X]$$

Это функция, то есть мы можем задать с.в. $Z = E[Y \mid N] = NE[X]$. А тогда мы знаем, что

$$E[Y] = E[E[Y \mid N]] = E[NE[X]] = E[N]E[X].$$

Заметим, что это верно только если матожидания N и X конечны. Это упрощенное равенство Вальда, которое мы рассмотрим подробнее позже.

$$\boxed{E\left[\sum_{i=1}^{N} X_i\right] = E[N]E[X]}$$

Посчитаем также дисперсию.

$$\operatorname{Var}(E[Y \mid N]) = \operatorname{Var}(NE[X]) = (E[X])^2 \operatorname{Var}(N),$$

так как $E[X]$ не зависит от N и считается константой.

$$\operatorname{Var}(Y \mid N = n) = \operatorname{Var}\left(\sum_{i=0}^{N} X_i \mid N = n\right) = \operatorname{Var}\left(\sum_{i=0}^{n} X_i\right) = n \operatorname{Var}(X) \quad (1)$$

$$\operatorname{Var}(Y \mid N) = N \operatorname{Var}(X) \quad (2)$$

$$E[\operatorname{Var}(Y \mid N)] = E[N \operatorname{Var}(X)] = E[N] \operatorname{Var}(X) \quad (3)$$

Теперь сложим $\operatorname{Var}(E[Y \mid N])$ и $E[\operatorname{Var}(Y \mid N)]$, получим дисперсию Y:

$$\operatorname{Var}\left(\sum_{i=1}^{N} X_i\right) = (E[X])^2 \operatorname{Var}(N) + E[N] \operatorname{Var}(X)$$

То есть дисперсия Y есть среднее число слагаемых умножить на дисперсию каждого, плюс она увеличивается дисперсией числа слагаемых, умноженных на квадрат среднего значения каждого слагаемого.

6 Равенство Вальда

У нас уже была упрощенная версия этой теоремы, однако давайте рассмотрим обобщенный случай.

Теорема 1 (Равенство Вальда) *Пусть $\{X_n\}_{n\in\mathbb{N}}$ — бесконечная последовательность вещественных с.в. Пусть также N — какая-то целочисленная неотрицательная с.в. (то есть N всегда принимает значения из $\mathbb{N}_0$). Допустим также, что выполнены все условия:*

1. *Все X_n имеют конечное матожидание.*
2. *Для любого $n \in \mathbb{N}$ верно, что $E[X_n \mathbb{1}_{\{N\geq n\}}] = E[X_n] \Pr(N \geq n)$ (можно трактовать как событие $N \geq n$ не очень зависит от с.в. X_n).*
3. *Следующий ряд сходится (трактовка: просто техническое требование для абсолютной сходимости $E[S_N]$ и $E[T_N]$, но также часто встречающееся на практике, особенно когда X_n неотрицательные).*

$$\sum_{n=1}^{+\infty} E[|X_n| \mathbb{1}_{\{N\geq n\}}] < +\infty$$

Тогда, если мы обозначим

$$S_N = \sum_{n=1}^{N} X_n,$$
$$T_N = \sum_{n=1}^{N} E[X_n],$$

то верно, что $E[S_N] = E[T_N]$

Если также N имеет конечное матожидание и все X_n имеют одинаковое конечное матожидание, то

$$E[S_N] = E[X_1]E[N].$$

Именно последнее равенство часто и имеют в виду, когда говорят о равенстве Вальда. Заметим, что по сравнению с тем, что у нас было раньше, мы не требуем полной независимости X_n от N, а также допускаем, что у X_n могут быть разные распределения.

Заметим, что в общем случае все три условия необходимы. Если не выполняется первое, то суммы S_N и T_N просто неопределены.

Необходимость второго условия демонстрируется следующим примером. Возьмем последовательность X_n, которые следуют распределению Бернулли с $p = 0.5$. И возьмем довольно зависимый от них $N = 1 - X_1$. В таком случае S_N всегда равно нуль (либо это сумма из нуля слагаемых, либо это одно слагаемое, равное нулю), однако $E[X_1]E[N] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. В данном примере есть явная зависимость события $N \geq 1$ от X_1, причем

$$E[X_1 \mathbb{1}_{\{N \geq 1\}}] = \frac{1}{2} \cdot 1 \cdot 0 + \frac{1}{2} \cdot 0 \cdot 1 = 0 \neq \frac{1}{2} \cdot \frac{1}{2} = E[X_1] \Pr[N \geq 1].$$

Для необходимости третьего (более технического) условия рассмотрим последовательность $X_n = \pm 2^n$, причем знак выбирается равновероятно. И возьмем $N = \min_n\{X_n = +2^n\}$, то есть первый раз, когда мы взяли положительную степень двойки. В такос случае выполняется первое условие, $E[X_n] = 0$, выполняется и второе, так как событие $N \geq n$ зависит только от $X_1, \ldots, X_{n-1}$, но не от X_n. При этом третье условие не выполнено:

$$E[|X_n| \mathbb{1}_{\{N \geq n\}}] = 2^n \cdot \left(\frac{1}{2}\right)^{n-1} = \frac{1}{2}.$$

В таком случае $S_N = 2$, так как независимо от N эта сумма равна $2^N - \sum_{n=1}^{N-1} 2^n = 2$. Но так как $E[X_n] = 0$, то справа в равенстве ноль.

Также пара примеров, когда все равботает. Очевидный пример — задача про сумму очков, которая выпадает на кости до первой выпавшей единицы. Мы раньше считали матожидание этой суммы так:

$$\sum_{i=1}^{+\infty} \frac{1}{6} \left(\frac{5}{6}\right)^{i-1} (4(i-1)+1) = 21$$

По равенству Вальда получается то же самое (только надо проверить удовлетворение всех условий)

$$E[S_N] = E[N]E[X_1] = 6 \cdot 3.5 = 21$$

Также неравенство Вальда работает при зависимых друг от друга X_n. Пусть есть какая-то с.в. Z с нулевым матожиданием, и мы определяем $X_n = (-1)^n Z$. По равенству Вальда матожидание их суммы равно нулю. Если рассуждать по-другому, то если N четное, то сумма равна нулю, а если нечетное, то $E[S_N] = E[-Z] = 0$. Но опять же надо проверить все условия.

Доказательство равенства Вальда

Сначала надо доказать абсолютную сходимость $E[S_N]$ и $E[T_N]$.

$$\begin{aligned} E[|S_N|] &= E\left[\sum_{i=1}^{+\infty} |S_i| \mathbb{1}_{N=i}\right] \leq E\left[\sum_{i=1}^{+\infty}\sum_{j=1}^{i} |X_j| \mathbb{1}_{N=i}\right] \\ &= E\left[\sum_{j=1}^{+\infty} |X_j| \sum_{i=j}^{+\infty} \mathbb{1}_{N=i}\right] = E\left[\sum_{j=1}^{+\infty} |X_j| \mathbb{1}_{N\geq j}\right] \\ &= \sum_{j=1}^{+\infty} E\left[|X_j| \mathbb{1}_{N\geq j}\right] < \infty, \end{aligned}$$

где переход со второй на третью строчку легален из-за конечности финального ряда, а последнее неравенство следует из условий теоремы.

$$\begin{aligned} E[|T_N|] &= \sum_{i=1}^{+\infty} |T_i| \Pr[N=i] \leq \sum_{i=1}^{+\infty} \Pr[N=i] \sum_{j=1}^{i} |E[X_j]| \\ &= \sum_{j=1}^{+\infty} |E[X_j]| \Pr[N \geq j] = \sum_{j=1}^{+\infty} |E[X_j \mathbb{1}_{N\geq j}]| \\ &\leq \sum_{j=1}^{+\infty} E[|X_j| \mathbb{1}_{N\geq j}] < \infty, \end{aligned}$$

где в середине второй строчки мы воспользовались вторым условием равенства, а в конце – третьим. Также мы дважды воспользовались неравенством треугольника (в обоих нестрогих неравенствах). Теперь легально доказать само равенство следующим образом:

$$\begin{aligned} E[S_N] &= E\left[\sum_{i=1}^{+\infty} X_i \mathbb{1}_{N\geq i}\right] = \sum_{i=1}^{+\infty} E\left[X_i \mathbb{1}_{N\geq i}\right] \\ &= \sum_{i=1}^{+\infty} E[X_i] \Pr[N \geq i] = \sum_{i=1}^{+\infty} E[X_i] \sum_{j=i}^{+\infty} \Pr[N=j] \\ &= \sum_{j=1}^{+\infty} \Pr[N=j] \sum_{i=1}^{j} E[X_i] = \sum_{j=1}^{+\infty} \Pr[N=j] T_j = E[T_N]. \end{aligned}$$

Мы везде могли менять местами порядок суммирования, так как все у нас сходится абсолютно, плюс воспользовались вторым условием при переходе на вторую строчку.

Наконец, если у всех X_n одинаковое матожидание, и у N конечное матожидание, то

$$E[S_N] = E[NE[X_1]] = E[N]E[X_1],$$

Так как $E[X_1]$ в матожидании расценивается как константа.