Conditional Expectation

We have previously seen the conditional probability mass function and expectation. We now consider taking expectation with respect to a conditional probability mass function. Two important reasons for considering this are:

- Conditioning arguments can facilitate the computation of expectations.
- The conditional expectation can be viewed as a prediction of a random variable given certain available information.

There are two notions of conditional expectation. The first we will consider is the conditional expectation given an event.

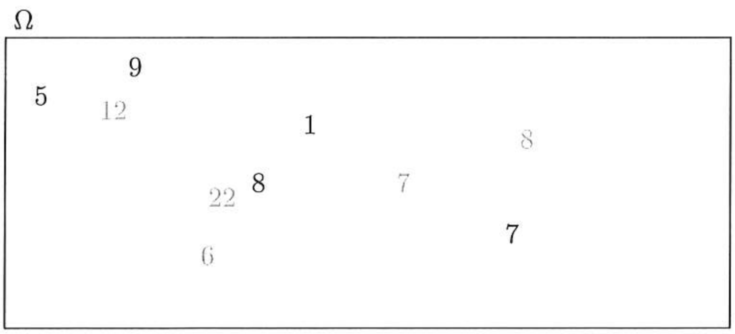
Definition. (Conditional expectation given an event) Let A be an event that occurs with positive probability. The conditional expectation of Y given A is

$$\mathbb{E}[Y \mid A] = \sum_y y\,\mathbb{P}(Y = y \mid A).$$

In particular, if $A = \{X = x\}$, then

$$\mathbb{E}[Y \mid X = x] = \sum_y y\,\mathbb{P}(Y = y \mid X = x) = \sum_y y\, f_{Y \mid X}(y \mid x).$$

Example. The sample space depicted below consists of black and grey numbers.



Assume that we will choose a colour number pair $\omega \in \Omega$ from this collection uniformly at random. Define the random variable $Y(\omega)$ as the number associated with the selection ω and the random variable $X(\omega)$ as

$$X(\omega) = \begin{cases} 1, & \text{if } \omega \text{ black,} \\ 0, & \text{if } \omega \text{ grey.} \end{cases}$$

$\mathbb{P}(Y=y \mid X=1) = 1/5$

$\mathbb{P}(Y=y, X=1)/\mathbb{P}(X=1)$

Question. What is the expected value of Y if we know that X is equal to 1? $y \in \{5, 9, 1, 8, 7\}$

$$\boxed{\mathbb{E}[Y \mid X = 1] = 1 \times \tfrac{1}{5} + 5 \times \tfrac{1}{5} + 7 \times \tfrac{1}{5} + 8 \times \tfrac{1}{5} + 9 \times \tfrac{1}{5}}$$

$= \frac{30}{5} = 6.$

Question. What is the expected value of our selection if we know that X is equal to 0?

$$\mathbb{E}[Y \mid X = 0] = \frac{1}{5}\times 6 + \frac{1}{5}\times 7 + \frac{1}{5}\times 8 + \frac{1}{5}\times 12 + \frac{1}{5}\times 22 = \frac{55}{5} = 11.$$

We can relate the conditional expectation $\mathbb{E}[Y \mid X = x]$ to the unconditional expectation $\mathbb{E}[Y]$ as follows: For any random variables X and Y defined in the same random experiement

$$\begin{aligned}
\sum_x \mathbb{E}[Y \mid X = x]\,\mathbb{P}(X = x) &= \sum_x \left\{\sum_y y\,\mathbb{P}(Y=y \mid X=x)\right\} \mathbb{P}(X=x) \\
&= \sum_x \sum_y y\,\mathbb{P}(Y=y \mid X=x)\,\mathbb{P}(X=x) \\
&= \\
&= \sum_y \cancel{x}\, y \left[\sum_x \mathbb{P}(Y=y, X=x)\right] \\
&= \sum_y y\,\mathbb{P}(Y=y) = \mathbb{E}(Y)
\end{aligned}$$

Example. Compute $\mathbb{E}[Y]$ from $\mathbb{E}[Y \mid X = 1]$ and $\mathbb{E}[Y \mid X = 0]$.

$$\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}[Y \mid X=1]\,\mathbb{P}(X=1) + \mathbb{E}[Y \mid X=0]\,\mathbb{P}(X=0) \\
&= 6 \times \tfrac{1}{2} + 11 \times \tfrac{1}{2} = 8\tfrac{1}{2}
\end{aligned}$$

The other notion of conditional expectation is the conditional expectation given a random variable.

Think of $\mathbb{E}[Y \mid X = x]$ as a function of x

Definition. (Conditional expectation given a random variable) The expression $\mathbb{E}[Y \mid X]$ is a random variable $g(X)$ that takes the value $\mathbb{E}[Y \mid X = x]$ when $X = x$.

Question. What is the support of the random variable $\mathbb{E}[Y \mid X]$? (previous example) $\{6, 11\}$

Since $\mathbb{E}[Y \mid X]$ is a random variable, we may take its expectation $\mathbb{E}[\mathbb{E}[Y \mid X]]$.

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[Y \mid X]] &:= \sum_x \mathbb{E}[Y \mid X = x]\,\mathbb{P}(X = x) \\
&= \mathbb{E}[Y].
\end{aligned}$$

This is very useful when we wish to know the expectation of a random sum of random variables.

Example. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with common mean μ, and let

$$S_n = \sum_{i=1}^{n} X_i .$$

$\mathbb{E}X$ = Expectation of X
= mean of X

n fixed

$$\boxed{\mathbb{E}[S_n] = \mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]} = n\mu .$$

Now let N be a random variable which takes values in the non-negative integers. Then,

$$\begin{aligned} \mathbb{E}[S_N] &= \mathbb{E}[\mathbb{E}[S_N \mid N]] \quad ; \quad \mathbb{E}[S_N \mid N] = N\mu \\ &= \mathbb{E}[N\mu] = \mu\,\mathbb{E}[N] \\ &= \end{aligned}$$

Two final properties of conditional expectation that we may have occassion to use are:

Property. Let $Y_1, Y_2, \ldots, Y_n$ and X be random variables defined in the same random experiment. Then

$$\mathbb{E}\left[\sum_{i=1}^{n} Y_i \middle| X\right] = \sum_{i=1}^{n} \mathbb{E}[Y_i \mid X].$$

(like 'linearity of expectations')

Property. If X and Y are independent, then $\mathbb{E}[Y \mid X]$ is constant and equal to $\mathbb{E}[Y]$.

$$\mathbb{E}[Y \mid X = x] = \sum_y y\,\mathbb{P}(Y=y \mid X=x)$$

$$\text{(by independence)} = \sum_y y\,\mathbb{P}(Y=y) = \mathbb{E}[Y]$$

Moment Generating Functions

Let X be a *non-negative* and *integer-valued* random variable.

The **moment generating function** (MGF) of X is the function $M : \mathbb{R} \to (0, \infty)$ defined by

$$M_X(s) := \mathbb{E}\,e^{sX} = \sum_{n=0}^{\infty} e^{sn}\mathbb{P}(X = n),$$

defined for all $s \in \mathbb{R}$ for which $\mathbb{E}\,e^{sX}$ exists (is finite). We will require that $M_X(s)$ exist for all s in some open interval containing the origin.

Question. What does $M_X(0)$ equal?

$$M_X(0) = \sum_{n=0}^{\infty} e^{n \times 0}\,\mathbb{P}(X=n) = \sum_{n=0}^{\infty} \mathbb{P}(X=n) = 1 .$$

$$\sum_{n=1}^{\infty} x^n = x \sum_{n=0}^{\infty} x^n$$

$$|x|<1 \quad \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

Example. Find the MGF of $X \sim$ Geometric(p).

$$M_X(s) = \sum_{n=1}^{\infty} e^{sn} p(1-p)^{n-1}$$

$$= \frac{p}{1-p} \sum_{n=1}^{\infty} e^{sn} (1-p)^n = \frac{p}{1-p} \sum_{n=1}^{\infty} \left[e^s(1-p)\right]^n$$

$$= \frac{p}{(1-p)} \cdot \frac{(1-p)e^s}{1-(1-p)e^s}$$

$$= pe^s/(1-(1-p)e^s)$$

for. $e^s(1-p) < 1$

$s + \log(1-p) < 0$

$s < -\log(1-p)$

Whenever $M_X(s)$ is defined, it can be determined from the distribution of X. Is the converse true? Given a moment generating function $M_X(s)$, can the distribution of X be determined? Fortunately, the answer is YES; moment generating functions have the **uniqueness property** – *two pmf's are the same if and only if their MGF's are the same.*

Recall from the properties of expectation that when X and Y are independent

$$\mathbb{E}\left[g(X)\, h(Y)\right] = \mathbb{E}\, g(X) \mathbb{E}\, h(Y)\,.$$

For the MGF we just defined this implies, for $Z = X + Y$, with X and Y independent (non-neg., integer-valued) random variables:

$$M_Z(s) = M_{X+Y}(s) = M_X(s) M_Y(s)$$

Exercise: Prove the formula above.

$$M_Z(s) = \mathbb{E}\left[e^{s(X+Y)}\right]$$
$$= \mathbb{E}\left[e^{sX} \cdot e^{sY}\right]$$
$$=$$
$$= \mathbb{E}\left[e^{sX}\right]\mathbb{E}\left[e^{sY}\right] \text{(by independence)}$$
$$= M_X(s) M_Y(s)$$

Using this fact and the uniqueness property, to find the distribution of a sum of a collection of random variables we simply need to multiply the MGF's of the random variables.

Example. Find the MGF of $X \sim$ Binomial(n,p). Recall that if $X_1, \ldots, X_n$ be independent Bernoulli(p) random variables. Then $\sum_{i=1}^n X_i$ has a Binomial(n,p) distribution. We can therefore compute the MGF of X by computing the MGF of $\sum_{i=1}^n X_i$.

Recall that if $X_i \sim \mathsf{Bernoulli}(p)$, then $\mathbb{P}(X_i = x) = p^x(1-p)^{1-x}$ for $x = 0, 1$.

$$M_{X_i}(s) = \mathbb{E}(e^{sX_i})$$
$$= (1-p)e^{s\times 0} + pe^{s\times 1} = 1-p+pe^s$$

Therefore,

$X \sim \text{Binomial}(n,p)$

$$M_X(s) = \prod_{i=1}^{n} M_{X_i}(s) \left\{= M_{X_1}(s) \times \cdots \times M_{X_n}(s)\right\}$$
$$= (1-p+pe^s)^n$$

Example. In the summary of formulas at the start of the workbook, the ~~PGF~~ MGF of $X \sim \mathsf{Poisson}(\lambda)$ is given as:

$$M_X(s) =$$

Hence, for $X \sim \mathsf{Poisson}(\lambda)$ and $Y \sim \mathsf{Poisson}(\mu)$ (indep.), the MGF of $Z = X + Y$ is:

$$M_Z(s) =$$

Which shows that $\boxed{Z \sim \quad\quad}$.

A useful property of the MGF is that we can obtain the *moments* of X by *differentiating* M and evaluating it at $s = 0$.

Differentiating $M(s)$ with respect to s gives

$$M'(s) = \frac{\mathrm{d}\mathbb{E}\,\mathrm{e}^{sX}}{\mathrm{d}s} = \mathbb{E}\,X\mathrm{e}^{sX}$$
$$M''(s) = \frac{\mathrm{d}\mathbb{E}\,X\mathrm{e}^{sX}}{\mathrm{d}s} = \mathbb{E}\,X^2\mathrm{e}^{sX}$$
$$M'''(s) = \mathbb{E}\,X^3\mathrm{e}^{sX}$$
$$\vdots$$
$$M^{(k)}(0) = \mathbb{E}\left[X^k\right]$$

In particular

$$\mathbb{E}[X] =$$

and

further discussion after lecture on MGF of geometric
Pg 60.

$$\frac{p}{1-p}\sum_{n=1}^{\infty}[e^s(1-p)]^n = \frac{p}{1-p}\sum_{n=1}^{\infty}[e^s(1-p)]^{n-1}\cdot e^s(1-p)$$

$$= \frac{p}{1-p}\, e^s(1-p)\sum_{n=1}^{\infty}[e^s(1-p)]^{n-1}$$

$$= \frac{p}{1-p}\, e^s(1-p)\sum_{k=0}^{\infty}[e^s(1-p)]^{k}$$

$$\sum_{n=1}^{\infty} x^n = \sum_{n=1}^{\infty} x^{n-1}\cdot x = x\sum_{n=1}^{\infty} x^{n-1} = x\sum_{k=0}^{\infty} x^k$$