Conditional expectation

If X and Y are random variables then

$$E(X \mid Y=y) = \sum_x x\, P(X=x \mid Y=y) \quad \text{if } X \text{ is a discrete r.v.}$$

$$= \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x \mid y)\, dx \quad \text{if } X \text{ is a continuous r.v.}$$

here, the r.v. Y can be either discrete or continuous.

Ex. 1 Suppose X, Y are jointly discrete with joint pmf as follows:

	$x=1$	$x=2$	$x=3$	row sum	marginal of Y:
$y=2$	.2	.1	0	.3	$\leftarrow P(Y=2)$
$y=4$	.1	.1	.2	.4	$\leftarrow P(Y=4)$
$y=6$	0	.2	.1	.3	$\leftarrow P(Y=6)$
column sum	.3	.4	.3		
Marginal of X:	$P(X=1)$	$P(X=2)$	$P(X=3)$		

Compute $E(X \mid Y=2)$, $E(X \mid Y=4)$ and $E(X \mid Y=6)$

Solution: $E(X \mid Y=2) = 1 \cdot P(X=1 \mid Y=2) + 2P(X=2 \mid Y=2) + 3P(X=3 \mid Y=2)$

$$= 1 \cdot \frac{P(X=1, Y=2)}{P(Y=2)} + 2\,\frac{P(X=2, Y=2)}{P(Y=2)} + 3\,\frac{P(X=3, Y=2)}{P(Y=2)}$$

$$= 1 \cdot \frac{.2}{.3} + 2 \cdot \frac{.1}{.3} + 3 \cdot \frac{0}{.3} = \frac{2}{3} + \frac{2}{3} = \boxed{\frac{4}{3}}.$$

Similarly,

$$E(X \mid Y=4) = 1\,P(X=1 \mid Y=4) + 2\,P(X=2 \mid Y=4) + 3\,P(X=3 \mid Y=4)$$

$$= 1\cdot\frac{1}{4} + 2\cdot\frac{1}{4} + 3\cdot\frac{2}{4}$$

$$= \frac{9}{4}.$$

and

$$E(X \mid Y=6) = 1\cdot P(X=1 \mid Y=6) + 2\,P(X=2 \mid Y=6) + 3\,P(X=3 \mid Y=6)$$

$$= 1\cdot\frac{0}{3} + 2\cdot\frac{2}{3} + 3\cdot\frac{1}{3}$$

$$= \frac{7}{3}.$$

<u>Ex.2</u>. Suppose X, Y joint continuous w/ joint pdf

$$f(x,y) = x\,e^{-x(1+y)} \quad \text{for } x>0,\ y>0.$$

(So that $f_Y(y) = \int_0^\infty x\,e^{-x(1+y)}\,dx = (1+y)^{-2}$ for $y>0$.)

Compute $E(X \mid Y=y)$.

$$E(X \mid Y=y) = \int_{-\infty}^{\infty} x\cdot f_{X|Y}(x|y)\,dx = \int_0^\infty x\cdot\frac{x\,e^{-x(1+y)}}{(1+y)^{-2}}\,dx$$

$$= \int_0^\infty (1+y)^2\cdot x^2\,e^{-x(1+y)}\,dy = \frac{2!}{(1+y)^3}\cdot(1+y)^2 = \frac{2}{1+y}.$$

Note: $X \mid Y=y \sim \text{Gamma}\left(2, \frac{1}{1+y}\right)$ so it is natural that $E(X \mid Y=y)$

be the product of α and β, in this case $2\cdot\frac{1}{1+y} = \frac{2}{1+y}$.

the random variable $E(X|Y)$.

When $Y=y$, we know (in principle) how to compute $E(X|Y=y)$. We can, therefore, think of $E(X|Y)$ to be the random variable that takes the value $E(X|Y=y)$ when $Y=y$.

For example, in Example 1 we found

$$E(X|Y=2) = \frac{4}{3} \quad , \quad E(X|Y=4) = \frac{9}{4} \text{ and } E(X|Y=6) = \frac{7}{3}.$$

So

$$E(X|Y) = \begin{cases} 4/3 & \text{if } Y=2 \\ 9/4 & \text{if } Y=4 \\ 7/3 & \text{if } Y=6. \end{cases}$$

But! We know $P(Y=2) = .3$, $P(Y=4) = .4$, $P(Y=6) = .3$ so we can even go further and say $E(X|Y)$ is the r.v. that takes the values

$$\frac{4}{3}, \ \frac{9}{4} \text{ and } \frac{7}{3}$$

with respective probabilities

$$.3, \ .4 \text{ and } .3 .$$

By the way let's compute the expected value of this r.v:

$$E(E(X|Y)) = \frac{4}{3}\cdot(.3) + \frac{9}{4}(.4) + \frac{7}{3}(.3) = 2.$$

Notice that $E(X) = 1(.3) + 2(.4) + 3(.3) = 2 = E(E(X|Y))\ldots$

this is NO ACCIDENT....

The Law of total expectation

$$E(X) = E\left(E(X|Y)\right).$$

Proof. Suppose X, Y are jointly continuous with joint pdf $f(x,y)$ and conditional pdf $f_{X|Y}(x|y)$. Then

$$E\left(E(X|Y)\right) = \int_{-\infty}^{\infty} E(X|Y=y) f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx \right\} f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \underbrace{f_{X|Y}(x|y) f_Y(y)}_{= f(x,y).}\, dx\, dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y)\, dx\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y)\, dy\, dx$$

$$= \int_{-\infty}^{\infty} x \left\{ \int_{-\infty}^{\infty} f(x,y)\, dy \right\} dx = \int_{-\infty}^{\infty} x f_X(x)\, dx = E(X).$$

$\boxtimes$

<u>Example</u> (a random sum of i.i.d. random variables)

Suppose N is a discrete r.v. with

$$P(N=n) = p_n \quad \text{for } n=0,1,2,3,\ldots$$

and suppose $X_1, X_2, X_3, \ldots$ are independent, identically distributed r.v.s. which are all independent of N.

Let μ_N, σ_N^2 and μ_X, σ_X^2 represent the means and variances of N and X, respectively.

Compute $E(S)$ where $S = \sum_{i=1}^{N} X_i = X_1 + X_2 + \cdots + X_N$.

First conditioning on $N=n$, we see

$$E(S|N=n) = E\left(\sum_{i=1}^{N} X_i \middle| N=n\right) = E\left(\sum_{i=1}^{n} X_i \middle| N=n\right) \xrightarrow{X\text{'s are indep of } N.} = E\left(\sum_{i=1}^{n} X_i\right)$$

$$= \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \mu_X = n\mu_X.$$

Now

$$E(E(S|N)) = \sum_{n=0}^{\infty} E(S|N=n)\,P(N=n) = \sum_{n=0}^{\infty} (n\mu_X)\,P(N=n)$$

$$= \mu_X \cdot \sum_{n=0}^{\infty} n\,p_n = \mu_X \cdot \mu_N.$$

By the law of total expectation

$$E(S) = E(E(S|N)) = \mu_X \mu_N.$$

$\blacksquare$

Continuing with the last example, let's compute $Var(S)$.

$$Var(S) = E(S^2) - \{E(S)\}^2 = E(S^2) - (\mu_X \mu_N)^2.$$

and we just need to compute the 2^{nd} moment of S.

$$S^2 = \left(\sum_{i=1}^{N} X_i\right)\left(\sum_{j=1}^{N} X_j\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} X_i X_j$$

$$= \sum_{i=1}^{N} X_i^2 + \sum_{\substack{i=1 \\ i \neq j}}^{N}\sum_{j=1}^{N} X_i X_j$$

$$E(S^2) = E\left(E(S^2 | N)\right)$$ so we start by computing the inner conditional expectation

$$E(S^2 | N=n)$$

$$= E\left(\sum_{i=1}^{N} X_i^2 + \sum_{i \neq j}^{N}\sum^{N} X_i X_j \,\Big|\, N=n\right) = E\left(\sum_{i=1}^{n} X_i + \sum_{i \neq j}^{n}\sum^{n} X_i X_j \,\Big|\, N=n\right)$$

$$= E\left(\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j}^{n}\sum^{n} X_i X_j\right)$$ since X_i's are independent of N.

$$= n E(X_1^2) + n(n-1)E(X_1 X_2)$$ by linearity of expectation and all X_i's have same distribution.

$$= n(\sigma_X^2 + \mu_X^2) + n(n-1)E(X_1)E(X_2)$$ since X_1, X_2 are independent

$$= n\sigma_X^2 + n\mu_X^2 + n(n-1)\mu_X^2 = n\sigma_X^2 \quad n^2\mu_X^2$$

Thus,

$$E(S^2) = E\left(E(S^2 | N)\right) = \sum_{n=0}^{\infty}(n\sigma_X^2 + n^2\mu_X^2)p_N(n) = \sigma_X^2\mu_N + \mu_X^2 E(N^2)$$

So

$$Var(S) = \sigma_X^2\mu_N + \mu_X^2 E(N^2) - \mu_X^2\mu_N^2 = \sigma_X^2\mu_N + \mu_X^2\sigma_N^2.$$

Example

Another useful property of the Conditional Expectation is the following:

Let X, Y be given and let h be any function.

Then
$$\boxed{E(X h(Y) \mid Y) = h(Y) E(X \mid Y).}$$

equivalently,

$$E(X h(Y) \mid Y=y) = h(y) E(X \mid Y=y).$$

Proof (is easy.) Use the definition of Conditional expectation (assume X is continuous r.v, for example).

$$E(X h(Y) \mid Y=y) = \int_{-\infty}^{\infty} x\, h(y)\, f_{X|Y}(x|y)\, dx = h(y) \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx$$
$$= h(y) E(X \mid Y=y).$$

Remark. The above property says that (loosely speaking) once Y is given, any function of Y can be treated as a constant. and therefore can be taken out of the (conditional) expectation.

Combining the above with the Law of total expectation we have another strategy for computing $E(X h(Y))$ (or, in fact, $E(g(X) h(Y))$)

$$E(g(X) h(Y)) = E\big(E(g(X) h(Y) \mid Y)\big) = E\big(h(Y) \underbrace{E(g(X) \mid Y)}\big)$$

and it may happen that $E(g(X)|Y)$ is easy to compute

Example

The law of total expectation can be used to give (yet) another way of computing $E(X)$ when

$X \sim \text{geometric}(p)$.

Let $X \sim \text{geometric}(p)$. Let Y be the outcome on the 1^{st} toss, so that $Y=1$ (is a success) with probability p and $Y=0$ (a failure) with probability $1-p$.

Let $\mu = E(X)$.

Now, $E(X \mid Y=1) = 1$ since a success happens on trial # 1 and therefore if $Y=1$, we have $X=1$. (easy.)

What about

$$E(X \mid Y=0)?$$

$$E(X \mid Y=0) = 1 + \mu$$

a failure happens on 1^{st} trial.

sample space

$$\Omega = \{S, FS, FFS, FFFS, FFFFS, FFFFFS, \ldots\}$$

These outcomes start with a failure

another copy of Ω

So

$$\mu = E(X) = 1 \cdot p + (1+\mu)\cdot(1-p) = p + (1-p) + \mu\cdot(1-p)$$

$$\mu = 1 + \mu(1-p) \Rightarrow p\cdot\mu = 1 \Rightarrow \mu = 1/p.$$

How about computing the Variance of a geometric?

$$E(X^2) = E\left(E(X^2 \mid Y)\right)$$

$$= E(X^2 \mid Y=1) \cdot p + E(X^2 \mid Y=0) \cdot (1-p)$$

$$E(X^2) = p + E\left((1+X)^2\right)(1-p)$$

$$= p + E\left(1 + 2X + X^2\right) \cdot (1-p)$$

$$= p + \left(1 + 2\overset{1/p}{\mu} + E(X^2)\right)(1-p)$$

$$= 1 + 2 \cdot \frac{1-p}{p} + (1-p)E(X^2)$$

$$pE(X^2) = 1 + \frac{2(1-p)}{p} \implies E(X^2) = \frac{1}{p} + \frac{2(1-p)}{p^2}.$$

So

$$\operatorname{var}(X) = E(X^2) - \{E(X)\}^2$$

$$= \frac{1}{p} + \frac{2(1-p)}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}.$$

Just as there is a notion of conditional expectation there is also a

Conditional Variance.

Here's the definition:

$$Var(X|Y) = E\left([X - E(X|Y)]^2 \,|\, Y \right).$$

or, equivalently

$$Var(X|Y=y) = E\left([X - E(X|Y=y)]^2 \,|\, Y=y \right).$$

Remark Jointly distributed random variables X and Y have a conditional distribution $X|Y=y$. In this way the

Conditional expectation and conditional variance are simply the mean and variance of this conditional distribution.

Let's verify this with example 2 from earlier.
In that example X, Y had joint pdf $f(x,y) = x e^{-x(1+y)}$ for $x>0, y>0$.

$f_Y(y) = \frac{1}{(1+y)^2}$ for $y>0$ and therefore, for $y>0$,

$$f_{X|Y}(x|y) = (1+y)^2 x e^{-x(1+y)} \quad \text{for } x>0.$$

which we note is the pdf of a Gamma$(2, \frac{1}{1+y})$ distribution.

That is, $X \mid Y=y \sim \text{Gamma}(\alpha, \beta)$ with $\alpha = 2$, $\beta = \frac{1}{1+y}$.

So the mean is then $\alpha\beta = 2 \cdot \frac{1}{1+y} = \frac{2}{1+y}$

and the variance is $\alpha\beta^2 = 2 \cdot \frac{1}{(1+y)^2} = \frac{2}{(1+y)^2}$.

We already saw (in example 2) that $E(X \mid Y=y) = \frac{2}{1+y}$

let's show the Conditional Variance of X given $Y=y$

is $\frac{2}{(1+y)^2}$ using the definition on the previous page ...

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

$$= E\left(\left[X - \frac{2}{1+y}\right]^2 \mid Y=y \right)$$

$$= E\left(X^2 - \frac{4}{1+y} \cdot X + \frac{4}{(1+y)^2} \mid Y=y \right)$$

$$= \int_0^\infty x^2 f_{X|Y}(x|y)\,dx - \frac{4}{1+y} \int_0^\infty x f_{X|Y}(x|y)\,dx + \frac{4}{(1+y)^2}$$

$$= (1+y)^2 \int_0^\infty x^2 \cdot x e^{-x(1+y)}\,dx - \frac{4}{1+y} \cdot \frac{2}{1+y} + \frac{4}{(1+y)^2}$$

$$= \frac{(1+y)^2 \cdot 3!}{(1+y)^4} - \frac{8}{(1+y)^2} + \frac{4}{(1+y)^2} = \frac{2}{(1+y)^2}.$$

exactly what we should expect. from the remark above.

Example (the Bivariate normal) Recall if Z_1, Z_2 are independent standard normals, then

$$X = \mu_X + \sigma_X Z_1 \quad \text{and}$$

$$Y = \mu_Y + \sigma_Y \cdot \rho Z_1 + \sigma_Y \sqrt{1-\rho^2} \cdot Z_2$$

has a bivariate Normal distribution.

Notice that

$$Z_1 = \frac{X - \mu_X}{\sigma_X}.$$

So that

$$Y = \mu_Y + \sigma_Y \rho \left(\frac{X - \mu_X}{\sigma_X} \right) + \sigma_Y \sqrt{1-\rho^2}\, Z_2$$

from which it follows that

$$Y|X \sim \text{Normal}\left(\mu_Y + \frac{\sigma_Y \rho}{\sigma_X}(X - \mu_X),\ \sigma_Y^2 (1-\rho^2) \right)$$

and immediately we have

$$E(Y|X) = \mu_Y + \frac{\sigma_Y \rho (X - \mu_X)}{\sigma_X} \quad \text{and} \quad Var(Y|X) = \sigma_Y^2 (1-\rho^2).$$

Notice, in particular, that

$$E(Y|X) = \mu_Y - \frac{\sigma_Y \mu_X \rho}{\sigma_X} + \frac{\sigma_Y \rho}{\sigma_X} X \quad \text{is just a linear transformation of } X$$

and

$$Var(E(Y|X)) = \frac{\sigma_Y^2 \rho^2}{\cancel{\sigma_X^2}} \cancel{Var(X)} = \sigma_Y^2 \rho^2 \implies \boxed{\rho^2 = \frac{Var(E(Y|X))}{Var(Y)}}.$$

The Moment-generating function (mgf)

For a r.v. X, we define the mgf of X as the function

$$M_X(s) = E(e^{sX})$$

when this function is finite (exists) in an open neighboorhood of $s=0$.

Let's assume for a r.v. X this function exists, then

since $e^y = 1 + y + \frac{y^2}{2!} + \frac{y^3}{3!} + \frac{y^4}{4!} + \cdots$

we have

$$e^{sX} = 1 + sX + \frac{s^2}{2!}X^2 + \frac{s^3}{3!}X^3 + \cdots$$

and, finally, by linearity of expectation

$$(*)\begin{cases} E(e^{sX}) = 1 + E(X)s + E(X^2)\frac{s^2}{2!} + E(X^3)\frac{s^3}{3!} + \cdots \\ \quad = \sum_{n=0}^{\infty} E(X^n)\frac{s^n}{n!} =: M_X(s). \end{cases}$$

The function $M_X(s)$ has all the moments $E(X^n)$ "encoded" into it. and if one wanted to find $E(X^n)$ for a particular integer $n \geq 1$ this function can be used to find it.

Ex. Suppose $X \sim \text{uniform}(0,1)$.

Compute the mgf $M_X(s)$.

Then use the representation on the previous page to find $E(X^n)$ for an arbitrary integer $n \geq 1$.

Solution. Use the Law of the Unconscious Statistician

$$M_X(s) = E(e^{sX})$$

$$= \int_0^1 e^{sx} \cdot 1 \, dx = \frac{e^{sx}}{s}\Big]_{x=0}^{x=1} = \frac{e^s - 1}{s}.$$

So the mgf of a Uniform$(0,1)$ is $M_X(s) = \frac{e^s - 1}{s}$.

Now,

$$e^s = 1 + s + \frac{s^2}{2!} + \frac{s^3}{3!} + \frac{s^4}{4!} + \dots + \frac{s^n}{n!} + \frac{s^{n+1}}{(n+1)!} + \dots$$

and

$$e^s - 1 = s + \frac{s^2}{2!} + \frac{s^3}{3!} + \frac{s^4}{4!} + \dots + \frac{s^n}{n!} + \frac{s^{n+1}}{(n+1)!}$$

and

$$\frac{e^s - 1}{s} = 1 + \frac{s}{2!} + \frac{s^2}{3!} + \frac{s^3}{4!} + \dots + \frac{s^{n-1}}{n!} + \frac{s^n}{(n+1)!} + \dots$$

$$= 1 + \left(\frac{1}{2}\right)s + \left(\frac{1}{3}\right)\frac{s^2}{2!} + \left(\frac{1}{4}\right)\frac{s^3}{3!} + \dots + \left(\frac{1}{n}\right)\frac{s^{n-1}}{(n-1)!} + \left(\frac{1}{n+1}\right)\frac{s^n}{n!} + \dots$$

When comparing to (*) on the previous page we see $E(X^n)$ is the coefficient of $\frac{s^n}{n!}$ in this expansion, so

$$E(X^n) = \frac{1}{n+1}.$$

KR

Since differentiation is a linear operation, it seems straightforward that

$$M_X'(s) = \frac{d}{ds}\left(M_X(s)\right) = \frac{d}{ds}E\left(e^{sX}\right) = E\left(\frac{d}{ds}\left(e^{sX}\right)\right)$$

$$= E\left(X e^{sX}\right).$$

From which it follows (plugging in $s=0$)

$$M_X'(0) = E\left(X e^{0\cdot X}\right) = E(X). \leftarrow \text{the } 1^{st} \text{ moment.}$$

Continuing in this way

$$M_X''(s) = \frac{d}{ds}\left(M_X'(s)\right) = \frac{d}{ds}E\left(Xe^{sX}\right) = E\left(\frac{d}{ds}\left(Xe^{sX}\right)\right)$$

$$= E\left(X^2 e^{sX}\right) \quad \text{and} \quad M_X''(0) = E(X^2) \text{ the second moment}$$

So that in principle we can also recover the moments of X by taking a suitable number of derivatives with respect to s of the mgf and then evaluate this at $s=0$:

$$\frac{d^n}{ds^n}M_X(s) = E\left(X^n e^{sX}\right) \Rightarrow \frac{d^n}{ds^n}\left(M_X(0)\right) = E\left(X^n\right).$$