Probability and Random Process

COSE382

Wonzoo Chung

Conditional Expectation given Event

Definition 9.1.1 (Conditional expectation given an event). Let A be an event with $P(A) > 0$. The *conditional expectation of Y given A* is

$$E(Y|A) = \sum_y yP(Y = y|A),$$

if Y is a discrete r.v. If If Y is continuous

$$E(Y|A) = \int_{-\infty}^{\infty} yf(y|A)dy,$$

where the conditional PDF $f(y|A)$ is defined as

$$f(y|A) = \frac{d}{dy}P(Y \leq y|A) = \frac{P(A|Y = y)f(y)}{P(A)}.$$

Example 9.1.4 (Life expectancy). Fred is 30 years old, and he hears that the average life expectancy in his country is 80 years. Should he conclude that, on average, he has 50 years of life left? No, there is a crucial piece of information that he must condition on: the fact that he has lived to age 30 already. Letting T be Fred's lifespan, we have the cheerful news that

$$E(T) < E(T|T \geq 30).$$

The left-hand side is Fred's life expectancy at birth (it implicitly conditions on the fact that he is born), and the right-hand side is Fred's life expectancy given that he reaches age 30.

Theorem 9.1.5 (Law of total expectation). Let $A_1, \ldots, A_n$ be a partition of a sample space ($\dot{\bigcup} A_i = S$), with $P(A_i) > 0$ for all i, and let Y be a random variable on this sample space. Then

$$E(Y) = \sum_{i=1}^{n} E(Y|A_i)P(A_i).$$

Proof. Will be given in 9.3.7, Adam's law .

Example 9.1.8 (Geometric expectation redux). Let $X \sim$Geom(p). X is the number of failures (T) before the first success (H). To get $E(X)$, we condition on the first trial: if H, then X is 0; if T, then one trial is added and we start over again, by memoryless property. Therefore,

$$\begin{aligned} E(X) &= E(X|\text{first toss } H) \cdot p + E(X|\text{first toss } T) \cdot (1-p) \\ &= 0 \cdot p + (1 + E(X)) \cdot (1-p), \end{aligned}$$

which gives $E(X) = \dfrac{(1-p)}{p}$.

Example 9.1.9 (Time until HH vs. HT). What is the expected number of tosses of a fair coin until HT appears for the first time? What about the expected number of tosses until HH ?

Solution: Let W_{HT} be the number of tosses until HT and W_{HH} be the number of tosses until HH.

$$
\begin{aligned}
E(W_{HT}) &= E(\text{number of trials for the first } H) \\
&\quad + E(\text{number of trials for the first } T \text{ after the first } H) \\
&= E(FS(1/2)) + E(FS(1/2)) = 4 \\
E(W_{HH}) &= E(W_{HH}|\text{first toss } T)\frac{1}{2} + E(W_{HH}|\text{first toss } H)\frac{1}{2}.
\end{aligned}
$$

By memoryless property,

$$
\begin{aligned}
E(W_{HH}|\text{first toss } T) &= 1 + E(W_{HH}) \\
E(W_{HH}|\text{first toss } H) &= 2 \cdot \frac{1}{2} + (2 + E(W_{HH}))\frac{1}{2}.
\end{aligned}
$$

Therefore,

$$
E(W_{HH}) = \left(2 \cdot \frac{1}{2} + (2 + E(W_{HH})) \cdot \frac{1}{2}\right)\frac{1}{2} + (1 + E(W_{HH}))\frac{1}{2}.
$$

Solving for $E(W_{HH})$, we get $E(W_{HH}) = 6$.

Conditional Expectation given r.v.

Definition 9.2.1 (Conditional expectation given an r.v.)
The *conditional expectation of* Y *given* X, denoted by,

$$E(Y|X),$$

is defined as a random variable $g(X)$, a function of random variable X,

$$g(x) = E(Y|X = x) = \int y f(y|X = x) dy$$

- $E(Y|X)$ is not a constant but a random variable as a function of X.
- We can consider $E(E(Y|X))$ and $\mathrm{Var}E(Y|X)$

$$\begin{aligned} E(E(Y|X)) &= \int E(Y|X = x) f(x) dx \\ \mathrm{Var}(E(Y|X)) &= \int \left[E(Y|X = x) - E(E(Y|X = x))\right]^2 f(x) dx \end{aligned}$$

Example 9.2.4. Let $X \sim \text{Unif}(0, 1)$ and $Y \sim \text{Unif}(0, x)$ given $X = x$. Find $E(Y|X)$, and its mean and variance.

Solution:

$(Y|X = x)$ ~Unif(0, x). Then $E(Y|X = x) = x/2$, so by plugging in X for x, we have

$$E(Y|X) = X/2.$$

The expected value of $E(Y|X)$ is

$$E(E(Y|X)) = E(X/2) = 1/4.$$

The variance of $E(Y|X)$ is

$$\text{Var}(E(Y|X)) = \text{Var}(X/2) = 1/48.$$

Theorem 9.3.1 (Dropping what's independent). If X and Y are independent, then

$$E(Y|X) = E(Y)$$

Proof: $E(Y|X = x) = E(Y)$ for all x, hence $E(Y|X) = E(Y)$

Theorem 9.3.2 (Taking out what's known). For any function h,

$$E(h(X)Y|X) = h(X)E(Y|X).$$

Proof: $E(h(x)Y|X = x) = h(x)E(Y|X = x)$ for all x, hence $E(h(X)Y|X) = h(X)E(Y|X)$.

Note that $E(X|X) = X = XE(1|X) = X$

Theorem 9.3.4 (Linearity). $E(Y_1 + Y_2|X) = E(Y_1|X) + E(Y_2|X)$.

Note that "$E(Y|X_1 + X_2) \neq E(Y|X_1) + E(Y|X_2)$"

Example 9.3.3. Let $Z \sim N(0, 1)$ and $Y = Z^2$. Find $E(Y|Z)$ and $E(Z|Y)$.

Solution:

1)

$$E(Y|Z) = E(Z^2|Z) = Z^2 E(1|Z) = Z^2$$

2) Given y,

$$E(Z|Y = y) = \frac{1}{2}E(Z|Z = \sqrt{y}) + \frac{1}{2}E(Z|Z = -\sqrt{y}) = \frac{1}{2}\sqrt{y} - \frac{1}{2}\sqrt{y} = 0$$

Hence, $E(Z|Y) = 0$

Example 9.3.6 Let $X_1 \ldots, X_n$ be i.i.d., and $S_n = X_1 + \cdots + X_n$. Find $E(X_1|S_n)$.

Solution: By symmetry,

$$E(X_1|S_n) = E(X_2|S_n) = \cdots = E(X_n|S_n),$$

and by linearity,

$$E(X_1|S_n) + \cdots + E(X_n|S_n) = E(S_n|S_n) = S_n.$$

Therefore,

$$E(X_1|S_n) = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Theorem 9.3.7 (Adam's law). For any r.v.s X and Y,

$$E(E(Y|X)) = E(Y).$$

Proof. We present the proof in the case where X and Y are both discrete (the proofs for other cases are analogous). Let $E(Y|X) = g(X)$. We proceed by applying LOTUS, expanding the definition of $g(x)$ to get a double sum, and then swapping the order of summation:

$$\begin{aligned}
E(E(Y|X)) &= \sum_x E(Y|X=x)P(X=x) \\
&= \sum_x \left(\sum_y yP(Y=y|X=x) \right) P(X=x) \\
&= \sum_x \sum_y yP(X=x)P(Y=y|X=x) \\
&= \sum_y y \sum_x P(X=x, Y=y) \\
&= \sum_y yP(Y=y) = E(Y).
\end{aligned}$$

Theorem 9.3.8 (Adam's law with extra conditioning). For any r.v.s X, Y, Z, we have

$$E(E(Y|X,Z)|Z) = E(Y|Z).$$

Proof. We interpret $E(Y|X,Z)$ as a function of two r.v.s $g(X,Z)$. For a fixed $Z = z$, $E(Y|X, Z = z)$ is $g(X,z)$, a function of X,

$$E(E(Y|X,Z)|Z = z) \quad = \quad E(g(X,z)) = E(E(Y|X, Z = z)) = E(Y|Z = z)$$

for all z. Hence, $E(E(Y|X,Z)|Z) = E(Y|Z)$.

Theorem 9.3.9 (Projection interpretation). For any function $h : \mathbb{R} \to \mathbb{R}$, the r.v. $Y - E(Y|X)$ is uncorrelated with $h(X)$,

$$E((Y - E(Y|X))h(X)) = E(Y - E(Y|X))E(h(X))$$

Equivalently,

$$E((Y - E(Y|X))h(X)) = 0.$$

Proof. We have

$$\begin{aligned} E((Y - E(Y|X))h(X)) &= E(h(X)Y) - E(h(X)E(Y|X)) \\ &= E(h(X)Y) - E(E(h(X)Y|X)) \\ &= E(h(X)Y) - E(h(X)Y) \\ &= 0 \end{aligned}$$

Example 9.3.10 (Linear regression). *Linear regression*, in its most basic form, assumes that the conditional expectation of Y, the response, is *linear* in X, the data:

$$E(Y|X) = a + bX.$$

(a) Show that an equivalent way to express this is to write $Y = a + bX + \epsilon$ where ϵ is an r.v. (called the *error*) with $E(\epsilon|X) = 0$.

(b) Solve for the constants a and b in terms of $E(X)$, $E(Y)$, $\mathrm{Cov}(X, Y)$, and $\mathrm{Var}(X)$.

Solution:

(a) Let define the error ϵ as $\epsilon := Y - (a + bX)$. Then

$$E(\epsilon|X) = E(Y|X) - E(a + bX|X) = E(Y|X) - (a + bX).$$

If $E(\epsilon|X) = 0$ then $E(Y|X) = a + bX$ and vice versa.

(b) Taking the covariance of X and $Y = a + bX + \epsilon$, we have

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(X, a) + b\mathrm{Cov}(X, X) + \mathrm{Cov}(X, \epsilon) = b\mathrm{Var}(X).$$

$\mathrm{Cov}(X, \epsilon) = 0$ due to $\epsilon = Y - E(Y|X)$ and by Adam's law, we have $E(Y) = a + bE(X)$. Thus,

$$b = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}, \quad a = E(Y) - bE(X) = E(Y) - \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} \cdot E(X).$$

Conditional Variance given r.v.

Definition 9.5.1 (Conditional variance).

The *conditional variance of* Y *given* X is

$$\mathrm{Var}(Y|X) = E((Y - E(Y|X))^2|X).$$

This is equivalent to

$$\mathrm{Var}(Y|X) = E(Y^2|X) - (E(Y|X))^2.$$

Like $E(Y|X)$, $\mathrm{Var}(Y|X)$ is a random variable, and it is a function of X.

Theorem 9.5.4 (Eve's law). For any r.v.s X and Y,

$$\mathrm{Var}(Y) = E(\mathrm{Var}(Y|X)) + \mathrm{Var}(E(Y|X)).$$

The ordering of E's and Var's on the right-hand side spells EWE, whence the name Eve's law. Eve's law is also known as the law of total variance or the variance decomposition formula.

Proof. Let $g(X) = E(Y|X)$. By Adam's law, $E(g(X)) = E(Y)$. Then

$$\begin{aligned} E(\mathrm{Var}(Y|X)) &= E(E(Y^2|X) - g(X)^2) = E(Y^2) - E(g(X)^2), \\ \mathrm{Var}(E(Y|X)) &= E(g(X)^2) - (Eg(X))^2 = E(g(X)^2) - (EY)^2. \end{aligned}$$

Adding these equations, we have Eve's law.

Example 9.6.1 (Random sum). N customers in a day, where N is an r.v. with finite $E(N)$ and $\text{Var}(N)$. Let X_j be the amount spent by the jth customer with $E(X_j) = \mu$ and $\text{Var}(X_j) = \sigma^2$. N and all the X_j are independent. Find the mean and variance of the random sum $X = \sum_{j=1}^{N} X_j$.

Solution: Note that $E(X) \neq N\mu$ but,

$$E(X|N) = E(\sum_{j=1}^{N} X_j|N) = \sum_{j=1}^{N} E(X_j|N) = \sum_{j=1}^{N} E(X_j) = N\mu.$$

By Adam's law,

$$E(X) = E(E(X|N)) = E(N\mu) = \mu E(N).$$

Furthermore,

$$\text{Var}(X|N) = \text{Var}(\sum_{j=1}^{N} X_j|N) = \sum_{j=1}^{N} \text{Var}(X_j|N) = \sum_{j=1}^{N} \text{Var}(X_j) = N\sigma^2.$$

By Eve's law,

$$\begin{aligned} \text{Var}(X) &= E(\text{Var}(X|N)) + \text{Var}(E(X|N)) = E(N\sigma^2) + \text{Var}(N\mu) \\ &= \sigma^2 E(N) + \mu^2 \text{Var}(N). \end{aligned}$$