Probability 2 - Notes 5

Conditional expectations $E(X|Y)$ as random variables

Conditional expectations were discussed in lectures (see also the second part of Notes 3). The goal of these notes is to provide a summary of what has been done so far. We start by reminding the main definitions and by listing several results which were proved in lectures (and Notes 3).

Let X and Y be two discrete r.v.'s with a joint p.m.f. $f_{X,Y}(x,y) = P(X = x, Y = y)$. Remember that the distributions (or the p.m.f.'s) $f_X(x) = P(X = x)$ of X and $f_Y(y) = P(Y = y)$ of Y are called the marginal distributions of the pare (X,Y) and that

$$f_X(x) = \sum_y f_{X,Y}(x,y) \text{ and } f_Y(y) = \sum_x f_{X,Y}(x,y).$$

If $f_Y(y) \neq 0$, the conditional p.m.f. of $X|Y = y$ is given by $f_{X|Y}(x|y) \stackrel{\text{def}}{=} \frac{f_{X,Y}(x,y)}{f_Y(y)}$ and the conditional expectation by

$$E(X|Y = y) \stackrel{\text{def}}{=} \sum_x x f_{X|Y}(x|y) \text{ and, more generally, } E(g(X)|Y = y) \stackrel{\text{def}}{=} \sum_x g(x) f_{X|Y}(x|y),$$

is defined for any real valued function $g(X)$. In particular, $E(X^2|Y = y)$ is obtained when $g(X) = X^2$ and

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

Remark. We always suppose that $\sum_x |g(x)| f_{X|Y}(x|y) \leq \infty$.

Definition. Denote $\varphi(y) = E(X|Y = y)$. Then $E(X|Y) \stackrel{\text{def}}{=} \varphi(Y)$. In words, $E(X|Y)$ is a random variable which is a function of Y taking value $E(X|Y = y)$ when $Y = y$.

The $E(g(X)|Y)$ is defined similarly. In particular $E(X^2|Y)$ is obtained when $g(X) = X^2$ and

$$Var(X|Y) = E(X^2|Y) - [E(X|Y)]^2.$$

Remark. Note that $E(X|Y)$ is a random variable whereas $E(X|Y = y)$ is a number (y is fixed).

Theorem 1. (i) $E[E(X|Y)] = E(X)$.
(ii) $Var(X) = Var[E(X|Y)] + E[Var(X|Y)]$.

Proof. See lecture or Notes 3.

Sums of random number of random variables (random sums).

Let $X_1, X_2, X_3, ...$ be a sequence of independent identically distributed random variables (i.i.d. random variables), each with the same distribution, each having common mean $a = E(X)$ and variance $\sigma^2 = Var(X)$. Here X is a r.v. having the same distribution as X_j. The sum $S = \sum_{j=1}^{N} X_j$ where the number in the sum, N is also a random variable and is independent of the X_j's. The following statement now follows from Theorem 1.

Theorem 2. (i) $E[S] = E(X) \times E(N) = aE(N)$.
(ii) $Var(S) = Var(X) \times E(N) + [E(X)]^2 \times Var(N) = \sigma^2 E(N) + a^2 Var(N)$.

Proof. (i) Since $E[S|N=n] = E[\sum_{j=1}^{n} X_j] = \sum_{j=1}^{n} E[X_j] = an$, we obtain that $E[S|N]] = aN$ (by the definition! of $E[S|N]]$). But then, from (i) of Theorem 1, we obtain that $E[S] = E[E[S|N]] = E[aN] = aE[N]$. □

(ii) Similarly $Var(S|N=n) = Var[\sum_{j=1}^{n} X_j] = n\sigma^2$ and hence $Var(S|N) = \sigma^2 N$. By Theorem 1, (ii) we have that

$$Var(S) = E[Var(S|N)] + Var(E[S|N]) = E[N\sigma^2] + Var(aN) = \sigma^2 E[N] + a^2 Var(N).$$

□

Example: finding $E(Y_n)$ and $Var(Y_n)$ for a branching process.

Remember that a BP $Y_n,\ n = 0, 1, 2, \ldots$, is defined by $Y_0 = 1$ and

$$Y_{n+1} = X_1^{(n)} + X_2^{(n)} + \ldots + X_{Y_n}^{(n)},$$

where r.v.'s $X_j^{(n)}$ are independent of each other and have the same distribution as a given integer-valued r.v. X.

Theorem 2 can be used in order to prove the following statements:

Suppose that $E(X) = \mu$, $Var(X) = \sigma^2$. Then:
(i) $E(Y_n) = \mu^n$
(ii)If $\mu \neq 1$, then $Var(Y_n) = \frac{\sigma^2 \mu^{n-1}(1-\mu^n)}{(1-\mu)}$. If $\mu = 1$ then $Var(Y_n) = n\sigma^2$.

Proof. Was given in lectures (and a different proof can be found in Notes 4).

Some additional properties of conditional expectations.

1. If X and Y are independent r.v.'s then $E(X|Y) = E(X)$.
Proof. As we know, X and Y are independent if and only if $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ or, equivalently, $f_{X|Y}(x|y) = f_X(x)$. But then $E(X|Y=y) = \sum_x x f_{X|Y}(x|y) = \sum_x x f_X(x) = E(X)$. □

2. $E[E(g(X)|Y)] = E(g(X))$
Proof. Set $Z = g(X)$. Statement (i) of Theorem 1 applies to any two r.v.'s. Hence, applying it to Z and Y we obtain $E[E(Z|Y)] = E(Z)$ which is the same as $E[E(g(X)|Y)] = E(g(X))$. □

This property may seem to be more general statement than (i) in Theorem 1. The proof above shows that in fact these are equivalent statements.

3. $E(XY|Y) = YE(X|Y)$.
Proof. $E(XY|Y=y) = E(yX|Y=y) = yE(X|Y=y)$ (because y is a constant). Hence, $E(XY|Y) = YE(X|Y)$ by the definition of the conditional expectation. □

Corollary. $E(XY) = E[YE(X|Y)]$. **Proof.** $E(XY) = E[E(XY|Y)] = E[YE(X|Y)]$. □

Exercise. Use the same method to prove that $E(Xh(Y)|Y) = h(Y)E(X|Y)$ for any real valued function $h(y)$.