IEOR 6711: Conditional expectation

Here we review some basic properties of conditional expectation that are useful for doing computations and give several examples to help the reader memorize these properties. (A more rigorous account can be found, for example, in Karlin and Taylor, Pages 5-9 in Ch. 1 and then Pages 302–305 in Ch. 6.)

Recall for two rvs X and Y that $E(X|Y)$ is itself a rv and is a function of Y, say $g(Y)$; so then for example $E(X|Y=i)=g(i)$. The idea is that besides the part of X determined by Y, the rest of X is averaged out with the expected value. By treating Y as if it was a constant, and then integrating (averaging) out the rest yields the conditional expectation. If X is determined by Y (for example $X=Y$ or some function of Y), then $E(X|Y)=X$; nothing has been averaged out.

Basic properties

1. $E(X)=E[E(X|Y)]$ for any rv Y. For example suppose $X=\sum_{n=1}^{N} U_n$, where the U_i are iid and independent of the rv N. Letting $Y=N$ yields $E(X|Y)=NE(U)$, and thus $E[E(X|Y)]=E(N)E(U)$. We then conclude that $E(X)=E(N)E(U)$. This illustrates how conditional expectations can serve as a useful tool towards computing an expected value; they allow one to compute $E(X)$ in two steps: first condition on some Y, and then take the expected value of that result. The point is that $E(X|Y)$ and X have the same expected value, so if your objective is to compute $E(X)$, then by choosing Y wisely, computing the expected value of $E(X|Y)$ can be easier to do than directly computing $E(X)$. Moreover, $Var[E(X|Y)]\leq Var[X]$ (with equality holding if and only if X is a deterministic function of Y.) Thus not only does $E(X|Y)$ have the same expected value as X but is also has *a smaller variance*. This fact can be taken advantage of in stochastic simulation since it yields a method of reducing the variance for estimation of some (difficult to compute exactly) $E(X)$: Generate n (large) iid copies of $E(X|Y)$ (and average) instead of generating n iid copies of X (and averaging). The strong law of large numbers and central limit theorem can be invoked on both scenarios, but with the conditional one having smaller variance, and hence smaller (more accurate) confidence intervals for estimating $E(X)$.

2. $E(X|Y)=E[E(X|Z,Y)|Y]$ for any rv Z. (This is a generalization of (1).) That is, conditioning on more information first, $\{Z,Y\}$, and then less information second, $\{Y\}$, is the same as only conditioning on the less information. More generally if $\mathcal{G}\subset\mathcal{F}$ are two $\sigma-$ fields of events, then $E(X|\mathcal{G})=E[E(X|\mathcal{F})|\mathcal{G}]$. For example, given a stochastic process $\{X_n\}$ and another one $\{U_n\}$ define $\mathcal{G}_n=\sigma\{X_0,\ldots,X_n\}$ and $\mathcal{F}_n=\sigma\{U_0,\ldots,U_n\}$, and suppose that $\mathcal{G}_n\subset\mathcal{F}_n,\ n\geq 0$. Then, for example,

$$E(X_{n+1}\mid\mathcal{G}_n)=E[E(X_{n+1}\mid\mathcal{F}_n)\mid\mathcal{G}_n],\ n\geq 0.$$

As in (1), this can help do computations, such as proving that $\{X_n\}$ is a martingale: If you can show that $E(X_{n+1} \mid \mathcal{F}_n) = X_n,\ n \geq 0$, then it follows that $E(X_{n+1} \mid \mathcal{G}_n) = X_n,\ n \geq 0$, and hence that $\{X_n\}$ is a martingale. Again, by wisely choosing some alternative $\{U_n\}$ for which $\mathcal{G}_n \subset \mathcal{F}_n,\ n \geq 0$, can greatly simplify computing the desired $E(X_{n+1} \mid \mathcal{G}_n)$ by allowing it to be done in two steps: first compute $E(X_{n+1} \mid \mathcal{F}_n)$ and then take its conditional expectation with respect to $\mathcal{G}_n$.

We consider some special cases

(a) $Z = X$ itself. Then $E(X|Z,Y) = E(X|X,Y) = X$, and hence $E[E(X|Z,Y)|Y] = E(X|Y)$.

(b) $X = Z + Y$. Then $E(X|Z,Y) = E(Z+Y|Z,Y) = Z+Y$, and hence $E[E(X|Z,Y)|Y] = E(Z+Y|Y) = E(Z|Y) + Y$. (Recall that $E(Y|Y) = Y$.)

(c) $X = \sum_{n=1}^{N(S)} U_n$, where $\{U_n\}$ are iid with mean $E(U)$, and independently $\{N(t)\}$ is a Poisson process at rate λ and independently $S > 0$ is a rv with mean $E(S)$. Let $Z = N(S)$ and $Y = S$. Recall that given S, $N(S)$ has a Poisson distribution with mean λS. Thus $E(X|Z,Y) = N(S)E(U)$, and thus $E[E(X|Z,Y)|Y] = E[N(S)E(U)|S] = E(U)E[N(S)|S] = E(U)\lambda S$. On the other hand $E(X|Y) = \lambda S E(U)$, the same thing.

3. If X_1 is determined by Y (e.g., is a function of Y; $X_1 = g(Y)$), then $E(X_1X_2|Y) = X_1E(X_2|Y)$. The idea here is that since we are treating Y as a constant, and since X_1 is determined by Y it too can be treated as a constant. A special case is $E(XY|Y) = YE(X|Y)$.