Some facts about expectation

October 22, 2010

1 Expectation identities

There are certain useful identities concerning the expectation operator that I neglected to mention early on in the course. Now is as good a time as any to talk about them. Here they are:

- Suppose that X is a continuous random variable having pdf $f(x)$, and suppose that $\alpha(x)$ is some function. Then,

  $$\mathbb{E}(\alpha(X)) \; = \; \int_{-\infty}^{\infty} \alpha(x) f(x) dx.$$

  There is also a discrete analogue of this identity.

- Suppose that $X_1, ..., X_k$ are random variables and that $\lambda_1, ..., \lambda_k$ are real numbers. Then,

  $$\mathbb{E}(\lambda_1 X_1 + \cdots + \lambda_k X_k) \; = \; \lambda_1 \mathbb{E}(X_1) + \cdots + \lambda_k \mathbb{E}(X_k).$$

  This property is called "Linearity of Expectation".

- Suppose that $X_1, ..., X_k$ are independent random variables. Then, the expectation of a product is the product of expectations; that is

  $$\mathbb{E}(X_1 \cdots X_k) \; = \; \mathbb{E}(X_1)\mathbb{E}(X_2) \cdots \mathbb{E}(X_k).$$

  Note: You really do need independence here. There are examples of r.v.'s that are not independent, for which this identity fails to hold.

- Suppose that X and Y are random varaibles. Then,

  $$\mathbb{E}_y \mathbb{E}(X | Y = y) \; = \; \mathbb{E}(X).$$

  This identity is called the "Tower Property of Expectation".

2 Proofs of the identities

2.1 Proof of the first identity

The idea of the proof is just to make a change of variable: Let $Y = \alpha(X)$, and let $g(y)$ denote the pdf for Y. For the purposes of this proof (to make it simpler to describe), let us assume that α is an increasing function. Then, we have that

$$\mathbb{P}(Y \leq y) \ = \ \int_{-\infty}^{\alpha^{-1}(t)} f(t)dt;$$

and so,

$$g(y) \ = \ f(\alpha^{-1}(y))\frac{d\ \alpha^{-1}(y)}{dy}.$$

We then get that

$$\mathbb{E}(Y) \ = \ \int_{-\infty}^{\infty} yg(y)dy \ = \ \int_{-\infty}^{\infty} yf(\alpha^{-1}(y))\frac{d\alpha^{-1}(y)}{dy}dy.$$

Making the substitution $x = \alpha^{-1}(y)$ we find that

$$\mathbb{E}(Y) \ = \ \int_{-\alpha^{-1}(-\infty)}^{\alpha^{-1}(\infty)} \alpha(x)f(x)dx.$$

The proof in the discrete case is even easier. We will not bother to give it here.

2.2 Proof of linearity of expectation

Suppose $f(x_1, ..., x_k)$ is the joint pdf for $(X_1, ..., X_k)$. Then, by definition,

$$\begin{aligned}
\mathbb{E}(\lambda_1 X_1 + \cdots + \lambda_k X_k) \ &= \ \int_{\mathbb{R}^k} (\lambda_1 x_1 + \cdots + \lambda_k x_k) f(x_1, ..., x_k) dx_1 \cdots dx_k \\
&= \ \sum_{i=1}^{k} \int_{\mathbb{R}} \lambda_i x_i \left(\int_{\mathbb{R}^{k-1}} f(x_1, ..., x_k) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_k \right) dx_i \\
&= \ \sum_{i=1}^{k} \lambda_i \int_{\mathbb{R}} x_i f_i(x_i) dx_i \\
&= \ \sum_{i=1}^{k} \lambda_i \mathbb{E}(X_i),
\end{aligned}$$

where f_i denotes the marginal pdf for X_i.

The discrete case is proved similarly.

2.3 Expectation of a product of independent random variables

Suppose $X_1, ..., X_k$ are independent random variables, having joint pdf $f(x_1, ..., x_k)$. Independence here implies that

$$f(x_1, ..., x_k) \;=\; f_1(x_1) \cdots f_k(x_k),$$

where the $f_i(x_i)$ are marginal pdfs. We have then that

$$\begin{aligned}
\mathbb{E}(X_1 \cdots X_k) &= \int_{\mathbb{R}^k} x_1 \cdots x_k f(x_1, ..., x_k) dx_1 \cdots dx_k \\
&= \int_{\mathbb{R}^k} x_1 \cdots x_k f_1(x_1) \cdots f_k(x_k) dx_1 \cdots dx_k \\
&= \left(\int_{\mathbb{R}} x_1 f_1(x_1) dx_1 \right) \cdots \left(\int_{\mathbb{R}} x_k f_k(x_k) dx_k \right) \\
&= \mathbb{E}(X_1) \cdots \mathbb{E}(X_k).
\end{aligned}$$

2.4 Proof of the tower property of expectation

Before we prove this particular identity we need to discuss what the conditional expectation notation even means: Given a joint pdf $f(x, y)$, we define something called the *conditional pdf* given by

$$f(x|y) \;=\; \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx} \;=\; \frac{f(x, y)}{h(y)},$$

where $h(y)$ is the marginal pdf for Y.

Basically, in the discrete case this would be $\mathbb{P}(X = x|Y = y)$; in the continuous case we have that for a region $A \subseteq \mathbb{R}^2$,

$$\mathbb{P}(X \in A \mid Y = y) \;=\; \int_A f(x|y) dx dy.$$

Once we have this notion, we can define conditional expectation as follows:

$$\mathbb{E}(X|Y = y) \;=\; \int_{\mathbb{R}} x f(x|y) dx.$$

And then we have that

$$\begin{aligned}
\mathbb{E}_y \mathbb{E}(X|Y=y) &= \int_{-\infty}^{\infty} \mathbb{E}(X|Y=y) g(y) dy \\
&= \int_{-\infty}^{\infty} g(y) \left(\int_{\mathbb{R}} x f(x|y) dx \right) dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x g(y) \frac{f(x,y)}{h(y)} dx dy \\
&= \int_{-\infty}^{\infty} x g(x) dx \\
&= \mathbb{E}(X),
\end{aligned}$$

where $g(x)$ is the marginal pdf for X, given by

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