CSE 107
Probability and Statistics for Engineers
Expected Value Rules

The simplest version of the Expected Value Rule occurs when we compose a discrete random variable with a function of one variable.

Theorem 1
Let $X: \Omega \to \mathbb{R}$ be a discrete random variable, and $g: \mathbb{R} \to \mathbb{R}$ be a function. Then

$$E[g(X)] = \sum_{x \in \text{range}(X)} g(x) p_X(x)$$

where $p_X(x)$ denotes the PMF of X.

Proof:
Let $Y = g \circ X = g(X)$ and observe that for any $y \in \mathbb{R}$,

$$\begin{aligned}
\{\, Y = y \,\} &= \{\omega \in \Omega \mid Y(\omega) = y\} \\
&= \{\omega \in \Omega \mid g\big(X(\omega)\big) = y\} \\
&= \bigcup_{\{x \mid g(x) = y\}} \{\omega \in \Omega \mid X(\omega) = x\} \\
&= \bigcup_{\{x \mid g(x) = y\}} \{\, X = x \,\}.
\end{aligned}$$

The last expression is a disjoint union, so by axiom (iii),

$$\begin{aligned}
p_Y(y) &= P(Y = y) \\
&= \sum_{\{x \mid g(x) = y\}} P(X = x) \\
&= \sum_{\{x \mid g(x) = y\}} p_X(x).
\end{aligned}$$

Therefore

$$\begin{aligned}
E[g(X)] &= E[Y] \\
&= \sum_{y \in \text{range}(Y)} y \cdot p_Y(y)
\end{aligned}$$

$$= \sum_{y} y \cdot \left(\sum_{\{x | g(x) = y\}} p_X(x) \right)$$

$$= \sum_{y} \sum_{\{x | g(x) = y\}} y \cdot p_X(x)$$

$$= \sum_{y} \sum_{\{x | g(x) = y\}} g(x) \cdot p_X(x)$$

$$= \sum_{x} g(x) \cdot p_X(x).$$

■

To establish the two-variable version, we use the Total Expectation Theorem for discrete random variables. This says that if events $A_1, A_2, \dots, A_n$ form a partition of the sample space Ω, and Z is a discrete random variable, then

$$E[Z] = \sum_{i=1}^{n} P(A_i) E[Z | A_i]$$

This formula remains valid even if the sequence of events forming the partion is infinite, making the above expression an infinite series.

Theorem 2
Let X and Y be discrete random variables, $g: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a function of two variables, and $Z = g(X, Y)$. Then

$$E[Z] = E[g(X, Y)] = \sum_{x} \sum_{y} g(x, y) p_{X,Y}(x, y)$$

where the sum is over all $x \in \text{range}(X)$ and $y \in \text{range}(Y)$.

Proof:
Observe that the events $\{Y = y\}$ for all $y \in \text{range}(Y)$, form a partition of Ω. The total expectation theorem gives

$$E[Z] = \sum_{y \in \text{range}(Y)} P(Y = y) E[Z | Y = y]$$

$$= \sum_{y} p_Y(y) E[g(X, Y) | Y = y]$$

$$= \sum_y p_Y(y) E[g(X,y)|Y=y]$$

$$= \sum_y p_Y(y) \sum_x g(x,y)\, p_{X|Y}(x|y)$$

$$= \sum_y \sum_x g(x,y) p_Y(y) p_{X|Y}(x|y)$$

$$= \sum_x \sum_y g(x,y) p_{X,Y}(x,y)$$

as required. Line 4 above follows since for each fixed $y \in \text{range}(Y)$, $g(x,y)$ can be considered a function of the single variable x, and from the 1-variable, conditional expected value rule.

■

Theorem 3
Let $X: \Omega \to \mathbb{R}$ be a continuous random variable. Then

$$E[X] = \int_0^\infty P(X > x) dx - \int_0^\infty P(X < -x) dx$$

Proof:
Consider the first term.

$$\int_0^\infty P(X > x)\, dx = \int_0^\infty \left(\int_x^\infty f_X(t) dt \right) dx$$

$$= \int_0^\infty \left(\int_0^t f_X(t) dx \right) dt$$

$$= \int_0^\infty f_X(t) \left(\int_0^t dx \right) dt$$

$$= \int_0^\infty t f_X(t) dt$$

The second line follows by integrating over the infinite triangle $\{(x,t) |\ 0 \le x \le t\}$ in two ways. Similarly, the second term becomes

$$\int_0^\infty P(X < -x)dx = \int_0^\infty \left(\int_{-\infty}^{-x} f_X(t)dt \right) dx$$

$$= \int_{-\infty}^{0} \left(\int_0^{-t} f_X(t)dx \right) dt$$

$$= \int_{-\infty}^{0} f_X(t) \left(\int_0^{-t} dx \right) dt$$

$$= \int_{-\infty}^{0} (-t) f_X(t)dt$$

Again, the second line follows by integrating over the infinite triangle $\{(x,t) | \, 0 \le x \le -t\}$ in two ways. Therefore

$$E[X] = \int_{-\infty}^{\infty} t f_X(t)dt$$

$$= \int_0^\infty t f_X(t)dt + \int_{-\infty}^{0} t f_X(t)dt$$

$$= \int_0^\infty t f_X(t)dt - \int_{-\infty}^{0} (-t) f_X(t)dt$$

$$= \int_0^\infty P(X > x)dx - \int_0^\infty P(X < -x)dx$$

as claimed. ■

Theorem 4
Let $X: \Omega \to \mathbb{R}$ be a continuous random variable, and $g: \mathbb{R} \to \mathbb{R}$ be a function. Then

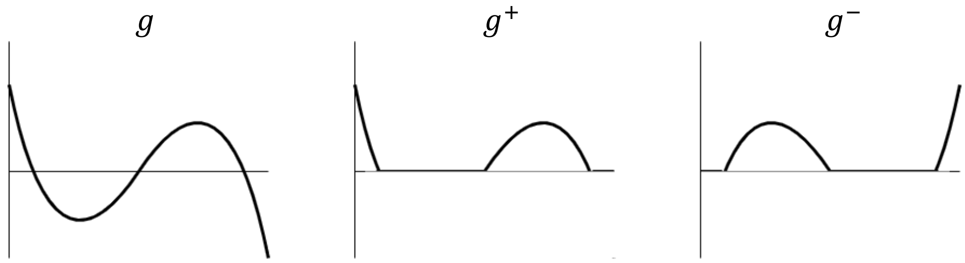
$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)dx$$

where $f_X(x)$ denotes the PDF of X.

Proof:
First define the two non-negative functions $g^+(x) = \max(g(x), 0)$ and $g^-(x) = \max(-g(x), 0)$ called the *positive part* and the *negative part* of $g(x)$, respectively. Then

$$g(x) = g^+(x) - g^-(x)$$

For instance



Now let $U = g(X)$. Then Theorem 3 says

$$E[U] = \int_0^\infty P(U > u)du - \int_0^\infty P(U < -u)du$$

i.e.

$$E[g(X)] = \int_0^\infty P(g(X) > u)du - \int_0^\infty P(g(X) < -u)du$$

The first term is

$$\int_0^\infty P(g(X) > u)du = \int_0^\infty \int_{\{x \mid g(x)>u\}} f_X(x)dx\,du$$

$$= \int_{-\infty}^\infty \int_{\{u \mid 0\le u<g(x)\}} f_X(x)du\,dx$$

$$= \int_{-\infty}^\infty \left(\int_0^{g(x)} du\right) f_X(x)dx$$

$$= \int_{-\infty}^\infty g^+(x) f_X(x)dx$$

The second term is

$$\int_0^\infty P(g(X) < -u)du = \int_0^\infty \int_{\{x \mid g(x)<-u\}} f_X(x)dx\, du$$

$$= \int_{-\infty}^\infty \int_{\{u \mid 0\le u<-g(x)\}} f_X(x)du\, dx$$

$$= \int_{-\infty}^\infty \left(\int_0^{-g(x)} du \right) f_X(x)dx$$

$$= \int_{-\infty}^\infty g^-(x)f_X(x)$$

Thus

$$E[g(X)] = \int_{-\infty}^\infty g^+(x)f_X(x)dx - \int_{-\infty}^\infty g^-(x)f_X(x)$$

$$= \int_{-\infty}^\infty \big(g^+(x) - g^-(x)\big)f_X(x)dx$$

$$= \int_{-\infty}^\infty g(x)f_X(x)dx$$

as claimed. ∎