However, if F_X is constant on some interval, then F_X^{-1} is not well defined by (2). The problem is avoided by defining $F_X^{-1}(y)$ for $0 < y < 1$ by

$$F_X^{-1}(y) = \inf\{x : F_X(x) \geq y\}. \tag{3}$$

At the end point of the range of y, $F_X^{-1}(1) = \infty$ if $F_X(x) < 1$ for all x and, for any F_X, $F_X^{-1}(0) = -\infty$.

Theorem 1.4 *(Probability integral transformation) Let X have continuous cdf $F_X(x)$ and define the random variable Y as $Y = F_X(X)$. Then Y is uniformly distributed on $(0,1)$, that is, $P(Y \leq y) = y$, $0 < y < 1$.*

Proof: For $Y = F_X(X)$ we have, for $0 < y < 1$,

$$\begin{aligned} P(Y \leq y) &= P(F_X(X) \leq y) \\ &= P(F_X^{-1}[F_X(X)] \leq F_X^{-1}(y)) \\ &= P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) = y. \end{aligned}$$

At the endpoints we have $P(Y \leq y) = 1$ for $y \geq 1$ and $P(Y \leq y) = 0$ for $y \leq 0$, showing that Y has a uniform distribution.

The reasoning behind the equality

$$P(F_X^{-1}[F_X(X)] \leq F_X^{-1}(y)) = P(X \leq F_X^{-1}(y))$$

is somewhat subtle and deserves additional attention. If F_X is strictly increasing, then it is true that $F_X^{-1}(F_X(x)) = x$. However, if F_X is flat, it may be that $F_X^{-1}(F_X(x)) \neq x$. Then $F_X^{-1}(F_X(x)) = x_1$, since $P(X \leq x) = P(X \leq x_1)$ for any $x \in [x_1, x_2]$. The flat cdf denotes a region of 0 probability $P(x_1 < X \leq x) = F_X(x) - F_X(x_1) = 0$. □

2 Expected values

Definition 2.1 *The expected value or mean of a random variable $g(X)$, denoted by $Eg(X)$, is*

$$Eg(X) = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } X \text{ is continuous} \\ \sum_{x \in \mathcal{X}} g(x) f_X(x) = \sum_{x \in \mathcal{X}} g(x) P(X = x) & \text{if } X \text{ is discrete,} \end{cases}$$

provided that the integral or sum exists. If $E|g(X)| = \infty$, we say that $Eg(X)$ does not exist.

Example 2.1 *(Exponential mean) Suppose X has an exponential (λ) distribution, that is, it has pdf given by*

$$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda}, \quad 0 \le x < \infty, \lambda > 0.$$

Then EX is given by

$$EX = \int_0^\infty x \frac{1}{\lambda} e^{-x/\lambda} dx = \lambda.$$

Example 2.2 *(Binomial mean) If X has a binomial distribution, its pmf is given by*

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n,$$

where n is a positive integer $0 \le p \le 1$, and for every fixed pair n and p the pmf sums to 1.

$$\begin{aligned}
EX &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n-1} n \binom{n-1}{x-1} p^x (1-p)^{n-x} \quad \left(x\binom{n}{x} = n\binom{n-1}{x-1}\right) \\
&= \sum_{y=0}^{n-1} n \binom{n-1}{y} p^{y+1} (1-p)^{n-(y+1)} \quad (\textit{substitute } y = x-1) \\
&= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} \\
&= np.
\end{aligned}$$

Example 2.3 *(Cauchy mean) A classic example of a random variable whose expected value does not exist is a Cauchy random variable, that is, one with pdf*

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad -\infty < x < \infty.$$

It is straightforward to check that $\int_{-\infty}^{\infty} f_X(x) dx = 1$, but $E|X| = \infty$. Write

$$E|X| = \int_{-\infty}^{\infty} \frac{|x|}{\pi} \frac{1}{1+x^2} dx = \frac{2}{\pi} \int_0^\infty \frac{x}{1+x^2} dx.$$

For any positive number M,

$$\int_0^M \frac{x}{1+x^2} dx = \frac{1}{2} \log(1+x^2)|_0^M = \frac{1}{2} \log(1+M^2).$$

Thus,

$$E|X| = \frac{1}{\pi} \lim_{M \to \infty} \log(1+M^2) = \infty$$

and EX does not exist.

Theorem 2.1 *Let X be a random variable and let a, b, and c be constants. Then for any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,*

a. $E(ag_1(X) + bg_2(X) + c) = aEg_1(X) + bEg_2(X) + c.$

b. *If* $g_1(x) \geq 0$ *for all* x, *then* $Eg_1(X) \geq 0$.

c. *If* $g_1(x) \geq g_2(x)$ *for all* x, *then* $Eg_1(X) \geq Eg_2(X)$.

d. *If* $a \leq g_1(x) \leq b$ *for all* x, *then* $a \leq Eg_1(X) \leq b$.

Example 2.4 *(Minimizing distance) Find the value of b which minimizes the distance $E(X-b)^2$.*

$$\begin{aligned} E(X-b)^2 &= E(X - EX + EX - b)^2 \\ &= E(X-EX)^2 + (EX-b)^2 + 2E((X-EX)(EX-b)) \\ &= E(X-EX)^2 + (EX-b)^2. \end{aligned}$$

Hence $E(X-b)^2$ is minimized by choosing $b = EX$.

When evaluating expectations of nonlinear functions of X, we can proceed in one of two ways. From the definition of $Eg(X)$, we could directly calculate

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

But we could also find the pdf $f_Y(y)$ of $Y = g(X)$ and we would have

$$Eg(X) = EY = \int_{-\infty}^{\infty} y f_Y(y) dy.$$

3 Moments and moment generating functions

Definition 3.1 *For each integer n, the n^{th} moment of X (or $F_X(x)$), μ'_n, is*

$$\mu'_n = EX^n.$$

The n^{th} central moment of X, μ_n, is

$$\mu_n = E(X-\mu)^n,$$

where $\mu = \mu'_1 = EX$.

Theorem 3.1 *The variance of a random variable X is its second central moment, $VarX = E(X-EX)^2$. The positive square root of $VarX$ is the standard deviation of X.*