Functions of random variables

Suppose X is a random variable and g a function. Then we can form a new random variable $Y := g(X)$. What is $Eg(X)$? In order to use the definition of expectations we need to first compute the distribution of $g(X)$, and that can be time consuming. Then, we set

$$Eg(X) = \sum_x xP\{g(X) = x\}.$$

The following theorem shows a simpler way out.

Theorem 1 (Law of the unconscious statistician). *Provided that $g(X)$ has a well-defined expectation, we have*

$$Eg(X) = \sum_w g(w)P\{X = w\}.$$

Proof. We know that

$$\begin{aligned} Eg(X) &= \sum_z zP\{g(X) = z\} = \sum_z z \sum_{w:\, g(w)=z} P\{X = w\} \\ &= \sum_w \sum_{z:\, g(w)=z} zP\{X = w\}, \end{aligned}$$

after reordering the sums. The preceding quantity is clearly equal to $\sum_w \sum_{z:\, g(w)=z} g(w)P\{X = w\}$. But for every w there is only one z such that $g(w) = z$. Therefore, $\sum_{z:\, g(w)=z} 1 = 1$. The theorem follows. □

Example 1 (Moments). The kth *moment* of a random variable X is defined as $E(X^k)$, provided that the expectation is defined. Thanks to the law of the unconscious statistician,

$$E(X^k) = \sum_a a^k P\{X = a\}.$$

For instance, suppose X denotes the number of dots rolled on a roll of a fair die. Then,

$$E(X^k) = \left(\frac{1}{6} \times 1\right) + \left(\frac{1}{6} \times 2^k\right) + \cdots + \left(\frac{1}{6} \times 6^k\right).$$

Variance and standard deviation

When $E(X^2)$ and $E(X)$ are defined, we can define the *variance* of a random variable X as

$$\text{Var}(X) := E(X^2) - (EX)^2.$$

Proposition 1. *It is always the case that*

$$|EX| \leq \sqrt{E(X^2)} \qquad \text{(the Cauchy–Schwarz inequality).}$$

In particular, EX is well defined and finite if $E(X^2) < \infty$. Moreover, if $E(X^2) < \infty$, then

$$\text{Var}(X) = E\left(|X - EX|^2\right).$$

The latter proposition shows that $\text{Var}(X) \geq 0$. Therefore, it makes sense to consider its square root: The *standard deviation* of X is defined as

$$\text{SD}(X) := \sqrt{\text{Var}(X)}.$$

Example 2. Suppose X denotes the number of dots rolled on a roll of a fair die. Then,

$$E(X^2) = \frac{1^2 + 2^2 + \cdots + 6^2}{6} = \frac{91}{6}, \qquad E(X) = \frac{1 + \cdots + 6}{6} = \frac{7}{2}.$$

Therefore,

$$\text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

And $\text{SD}(X) = \sqrt{35/12} \approx 1.71$.

Expectation of functions of more than one variable

Similar ideas as before lead us to the following: If (X, Y) is a random vector then for every function g of two variables,

$$Eg(X, Y) = \sum_{(a,b)} g(a, b)P\{X = a, Y = b\},$$

provided that the sum is well defined.

Example 3 (Quick proof of the addition rule)**.**

$$\begin{aligned} E(\alpha X + \beta Y) &= \sum_{a,b} (\alpha a + \beta b)P\{X = a, Y = b\} \\ &= \alpha \sum_{(a,b)} P\{X = a, Y = b\} + \beta \sum_{(a,b)} P\{X = a, Y = b\}. \end{aligned}$$

Because $\sum_b P\{X = a, Y = b\} = P\{X = a\}$ and $\sum_a P\{X = a, Y = b\} = P\{Y = b\}$ [marginals!], we find that $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$, as before.

Example 4 (Expectation of a product)**.** If (X, Y) is a random vector, then

$$E(XY) = \sum_{(a,b)} abP\{X = a, Y = b\}.$$

If, in addition, X and Y are independent, then $P\{X = a, Y = b\} = P\{X = a\}P\{Y = b\}$, therefore as long as the following expectations are all defined and finite, then

$$E(XY) = \sum_a aP\{X = a\} \sum_b bP\{Y = b\} = E(X)E(Y)!$$

And more generally [still assuming independence],

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)],$$

provided that the expectations are all defined and finite.

Proposition 2. *Suppose (X, Y) is a random vector and $E(X^2)$ and $E(Y^2)$ are finite. Then,*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\{E(XY) - EX \cdot EY\}.$$

In particular, if $X_1, \ldots, X_n$ are independent, then

$$\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n).$$

Proof. Because $E(X^2)$ and $E(Y^2)$, the Cauchy–Schwarz inequality tells us that EX, EY, $\text{Var}(X)$, and $\text{Var}(Y)$ are all defined and finite. Therefore,

$$\begin{aligned}\text{Var}(X+Y) &= E\left[(X+Y)^2\right] - \left[E(X+Y)\right]^2 \\ &= \left[E(X^2) + E(Y^2) + 2E(XY)\right] - \left[(EX)^2 + (EY)^2 + 2EX \cdot EY\right],\end{aligned}$$

and this does the job since $E(X^2) - (EX)^2$ is the variance of X and $E(Y^2) - (EY)^2$ is the variance of Y. When X and Y are independent, $E(XY) = EX \cdot EY$, and the variance of $X + Y$ simplifies to the sum of the individual variances of X and Y. In particular, if X_1 and X_2 are independent then $\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2)$. And the remainder of the proposition follows from this and induction [on n]. □

Example 5 (Variance of binomials)**.** Let X have the binomial distribution with parameters n and p. We have seen that we can write X as

$$X = I_{A_1} + \cdots + I_{A_n},$$

where the A_j's are independent events with $P(A_j) = p$ each. And this is how we discovered the important identity,

$$E(X) = np.$$

Note that $I_{A_j}^2 = I_{A_j}$ [being a 0/1-valued object]. It follows that

$$\text{Var}(I_{A_j}) = E\left(I_{A_j}^2\right) - \left(EI_{A_j}\right)^2 = E(I_{A_j}) - p^2 = p - p^2 = p(1-p) = pq.$$

Consequently,

$$\text{Var}(X) = npq \qquad \text{and} \qquad \text{SD}(X) = \sqrt{npq}.$$

Direct computations of these are possible, but involved. For instance, in order to compute $\text{Var}(X)$ directly we first can attempt to find $E(X^2)$. But that means that we have to evaluate

$$E(X^2) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k},$$

which can be done but is tedious. On the other hand, the converse is easy [thanks to indicator functions]: $E(X^2) = \text{Var}(X) + (EX)^2 = npq + n^2p^2$.

Here is a final remark on standard deviations; it tells us how the standard deviation is changed under a linear change of variables:

Proposition 3. *If α and β are constants, then*

$$\text{SD}(\alpha X + \beta) = |\alpha|\text{SD}(X),$$

provided that $E(X^2) < \infty$.

Proof. Because $(\alpha X + \beta)^2 = \alpha^2 X^2 + \beta^2 + 2\alpha\beta X$,

$$E\left[(\alpha X + \beta)^2\right] = \alpha^2 E(X^2) + \beta^2 + 2\alpha\beta E(X).$$

And

$$\left[E(\alpha X + \beta)\right]^2 = \alpha^2 (EX)^2 + \beta^2 - 2\alpha\beta E(X).$$

Subtract the preceding 2 displays to find that

$$\mathrm{Var}(\alpha X + \beta) = \alpha^2 \mathrm{Var}(X).$$

The proposition follows upon applying a square root to both sides of the latter identity. □