Expectations and Variances

Guy Lebanon

January 6, 2006

A RV $X : \Omega \to \mathbb{R}$ is described by probabilities assigned to quantities such as $P(X \in A)$. Alternatively, it is described by a pmf (for discrete) or pdf (for continuous) function. It is desirable to obtain a more succinct summary of the RV in terms of several numbers (rather than a function). This leads us to the concept of expectation and variances of RV.

Definition 1. *The expectation of a RV X is the real number*

$$\mathsf{E}(X) = \begin{cases} \sum_{x\in\mathbb{R}} x p_X(x) & X \textit{ is a discrete RV} \\ \int_{-\infty}^{\infty} x f_X(x)dx & X \textit{ is a continuous RV} \end{cases}.$$

In the definition above we assume that the integral or sum converges and even the stronger condition $\int |x| f_X(x)dx < \infty$. If the integral or sum does not converge to a real number we say that the expectation does not exist. Note that the sum in the definition for a discrete RV contains only finite or countable terms.

For example, the expectation of a uniform RV (continuous) on (a, b) is

$$\mathsf{E}(X) = \int \frac{1}{b-a} x dx = \frac{1}{b-a} \frac{x^2}{2}\Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

The expectation of a Bernoulli RV (discrete) is

$$\mathsf{E}(X) = 1p_X(1) + 0p_X(0) = 1 \cdot \theta = \theta.$$

The expectation of a RV expressing the result of throwing a fair die is

$$\mathsf{E}(X) = 1p_X(1) + \cdots 6p_X(6) = \frac{1}{6}(1 + \cdots + 6) = 21/6 = 3.5.$$

The expectation describes the mean or average of a RV. To see this consider averaging the result of a die over k throws (we denote by n_k the result of the k throw)

$$\begin{aligned} \text{average} =& \frac{n_1 + \ldots + n_k}{k} = \frac{1(\text{number of times we get 1}) + \cdots + 6(\text{number of times we get 6})}{k} \\ =& \, 1 \cdot \text{frequency of getting } 1 + \cdots + 6 \cdot \text{frequency of getting } 6. \end{aligned}$$

The above equals $\sum_x x p_X(x)$ under the frequency interpretation of probability. The expectation is then just a formal definition of average - but one that applies to either classical, frequency or subjective interpretation and applies to both discrete and continuous RVs.

The expectation of the RV $g(X)$ is given by $\sum_y y p_{g(X)}(y)$ or $\int y f_{g(X)}(y)dy$. It is sometimes easier to use the following formulas that compute $\mathsf{E}(g(X))$ in terms of the pdf or pmf of X:

$$\mathsf{E}(g(X)) = \begin{cases} \sum_{x\in\mathbb{R}} g(x) p_X(x) & \text{X is a discrete RV} \\ \int_{-\infty}^{\infty} g(x) f_X(x)dx & \text{X is a continuous RV} \end{cases}. \tag{1}$$

An important property of expectation is the linearity property

Theorem 1. *For any constants $a, b \in \mathbb{R}$, $Y = aX + b$ is a RV whose expectation is $\mathsf{E}(Y) = a\mathsf{E}(X) + b$*

Proof. Using Equation (1) we have

$$\mathsf{E}(aX+b) = \sum_x (ax+b)p_X(x) = a\sum_x p_X(x) + b\sum_x p_X(x) = a\mathsf{E}(x) + b\cdot 1$$

for discrete RV. The proof for continuous RV is the same except that the sum is replaced with the integral. □

Note that if $a=0$ we get $\mathsf{E}(X+b) = \mathsf{E}(X)$ and if $b=0$ we get $\mathsf{E}(aX) = a\mathsf{E}(X)$.

The expectation of a RV tells us the average or the mean. It is also important to know how much the RV deviates from the mean. This way we can summarize the spread of the RV X around its mean $\mathsf{E}(X)$. This is formalized by the concept of variance.

Definition 2. *The variance of a RV X is the expectation of the RV $Y = (X - \mathsf{E}(X))^2$:*

$$\mathsf{Var}(X) = \mathsf{E}((X - \mathsf{E}(X))^2).$$

The standard deviation of a RV X is $\sqrt{\mathsf{Var}(X)}$.

Since the variance is the mean of the squared deviation of the RV from its mean, it is small for RVs with peaked or narrow pdf or pmf and large for RV with diffuse or wide pdf or pmf.

As an example, the variance of the uniform RV on (a,b) is

$$\begin{aligned}
\mathsf{Var}(X) = \mathsf{E}((X - \mathsf{E}X)^2) &= \int_a^b \frac{1}{(b-a)}(x-(a+b)/2)^2 dx = \frac{1}{b-a}\int_a^b (x^2 - x(a+b) + (a+b)^2/4)dx \\
&= \frac{1}{b-a}\left(\frac{b^3-a^3}{3} - (a+b)\frac{b^2-a^2}{2} + \frac{(a+b)^2}{4}(b-a)\right) = \frac{b^3-a^3}{3(b-a)} - (a+b)\frac{b+a}{2} + \frac{(a+b)^2}{4} \\
&= \frac{(a-b)(a^2+ab+b^2)}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{a^2+b^2-2ab}{12} = \frac{(a-b)^2}{12}.
\end{aligned}$$

Note the dependence of $\mathsf{Var}(X)$ on the length of the interval $b-a$: the larger it is, the less peaked the pdf is, and the larger the variance is.

An alternative expression for the variance that is often simpler to compute is

$$\mathsf{Var}(X) = \mathsf{E}((X - \mathsf{E}(X))^2) = \mathsf{E}(X^2 - 2X\mathsf{E}X + (\mathsf{E}x)^2) = \mathsf{E}(X^2) - 2\mathsf{E}X\mathsf{E}X + (\mathsf{E}X)^2 = \mathsf{E}(X^2) - (\mathsf{E}X)^2$$

where we used above the linearity property of the expectation.

In contrast to the linearity of expectations, for variances we have

$$\mathsf{Var}(aX) = \mathsf{E}(a^2X^2) - (\mathsf{E}aX)^2 = a^2\mathsf{E}(X^2) - (a\mathsf{E}X)^2 = a^2(\mathsf{E}(X^2) - (\mathsf{E}X)^2) = a^2\mathsf{Var}(X)$$

and

$$\mathsf{Var}(X+b) = \mathsf{E}((X+b-\mathsf{E}(X+b))^2) = \mathsf{E}((X+b-\mathsf{E}(X)+b))^2) = \mathsf{Var}(X)$$

which put together form:

$$\mathsf{Var}(aX+b) = a^2\mathsf{Var}(X)$$

It is worth mentioning that a constant b can be considered as a random variable of the constant function $b:\Omega \to \mathbb{R}$, $b(\omega) = b$. As such it is a "deterministic" RV that does not have any real randomness. Its expectation is $\mathsf{E}(b) = \sum_x bp_X(x) = b$ and its variance is $\mathsf{Var}(b) = \mathsf{E}(b - \mathsf{E}(b)))^2 = E(0) = 0$.