Random Walk Algorithms: Lecture 3

David A. Meyer

In which are introduced the notions of a random variable, its probability distribution function, its expectation value and variance, and selected properties of the latter, to be used subsequently.

Random Variables

It is often useful to calculate some numerical quantity associated with a random outcome. For example, if we flip a coin multiple times, we might want to understand the probability of seeing k heads. To formalize this idea, we define random variables:

DEFINITION. A real-valued random variable (to which we will usually refer just as a random variable^{*}) X is a map $X : \Omega \to \mathbb{R}$. If $S \subseteq \mathbb{R}$,

$$\Pr(X \in S) = \sum_{\omega \in \Omega | X(\omega) \in S} \Pr(\omega).$$

In particular,

$$\Pr(X = x) = \Pr(X \in \{x\}) = \sum_{\omega \in \Omega | X(\omega) = x} \Pr(\omega).$$

The map

$$f_X : \mathbb{R} \to \mathbb{R}$$
$$x \mapsto \Pr(X = x)$$

is the probability distribution function for X.

EXAMPLE. Let H(head) = 1 and H(tail) = 0. If Pr(head) = p, then Pr(H = 1) = p and Pr(H = 0) = 1 - p.

Expectation value

DEFINITION. The expectation value of a random variable X is

$$\mathsf{E}[X] = \sum_{\omega \in \Omega} X(\omega) \operatorname{Pr}(\omega) = \sum_{x \in \mathbb{R}} x f_X(x).$$

We can think of the expectation value as the average (or the mean) of the values of the random variable (weighted by their probabilities of occurring).

^{*} Although if we study random walks on \mathbb{Z}^d for d > 1 we must consider \mathbb{R}^d -valued random variables.

EXAMPLE. $\mathsf{E}[H] = p \cdot 1 + (1-p) \cdot 0 = p.$

Now, for two coin flips we can define two random variables: $H_i : \Omega \to \mathbb{R}$, with $H_i(\omega) = 1$ if flip $i \in \{1, 2\}$ is head up. Consider the sum $H_1 + H_2 : \Omega \to \mathbb{R}$. This is also a random variable, with $(H_1 + H_2)$ (head, head) = 2, *etc.*, so it counts the number of heads up in two coin flips. We can compute its expectation value using the definition:

$$\begin{split} \mathsf{E}[H_1 + H_2] &= \sum_{\omega} (H_1 + H_2)(\omega) \operatorname{Pr}(\omega) \\ &= \sum_{\omega} H_1(\omega) \operatorname{Pr}(\omega) + \sum_{\omega} H_2(\omega) \operatorname{Pr}(\omega) \\ &= \mathsf{E}[H_1] + \mathsf{E}[H_2] = 2p. \end{split}$$

This result is an instance of a general theorem, the proof of which is essentially the same calculation:

THEOREM. Let X, Y be two real-valued random variables on a countable sample space Ω . Then $\mathsf{E}[X + Y] = \mathsf{E}[X] + \mathsf{E}[Y]$, and $\mathsf{E}[\alpha X] = \alpha \mathsf{E}[X]$, for all $\alpha \in \mathbb{R}$.

We often refer to this result as "expectation value is linear". Notice that the dependence or independence of the events $\{H_1 = 1\}$ and $\{H_2 = 1\}$, for example, was irrelevant to the calculation. Nevertheless, we need a definition for independence of random variables:

DEFINITION. Two random variables X, Y are independent if for all $x, y \in \mathbb{R}$, the events $\{X = x\}$ and $\{Y = y\}$ are independent.

Variance

In the theorem stating that expectation value is linear, we implicitly used the following idea: Let $g : \mathbb{R} \to \mathbb{R}$. Then $g \circ X : \Omega \to \mathbb{R}$ is a random variable if X is a random variable. Thus,

$$\mathsf{E}[g(X)] = \sum_{\omega} g(X(\omega)) \operatorname{Pr}(\omega).$$

EXAMPLES. When $g : \mathbb{R} \to \mathbb{R}$ is defined by $g(x) = \alpha x$, we have the second statement in the theorem above. Even more simply, consider $\mathsf{E}[c]$, for a constant $c \in \mathbb{R}$. Let $g : \mathbb{R} \to \mathbb{R}$ be the constant function g(x) = c. Then

$$\mathsf{E}[c] = \sum_{\omega} c \operatorname{Pr}(\omega) = c \sum_{\omega} \operatorname{Pr}(\omega) = c.$$

Finally, for the specific random variable H defined above, H^2 is a random variable (using $g(x) = x^2$). But since $1^2 = 1$ and $0^2 = 0$, $H^2 = H$.

We saw above that the expectation value of the sum of two random variables is the sum of their expectation values. What about the expectation value of a product? To analyze this situation, we let $X(\Omega) = \{x_1, \ldots, x_r\}$ and $Y(\Omega) = \{y_1, \ldots, y_s\}$. (Notice that r and s need not be the same, since X and Y might not be 1-to-1 maps, and might have codomains of different cardinalities.) Then $(XY)(\Omega) \subseteq \{x_i y_j \mid 1 \le i \le r, 1 \le j \le s\}$, so

$$E[XY] = \sum_{i,j} x_i y_j \Pr(X = x_i \land Y = y_j)$$

= $\sum_{i,j} x_i y_j \Pr(X = x_i) \Pr(Y = y_j)$ if X, Y independent
= $\sum_i x_i \Pr(X = x_i) \sum_j y_j \Pr(Y = y_j)$
= $E[X]E[Y].$

This proves:

THEOREM. Let X, Y be two independent real-valued random variables on a countable sample space Ω . Then $\mathsf{E}[XY] = \mathsf{E}[X]\mathsf{E}[Y]$.

Notice that this theorem definitely does *not* say that for an arbitrary random variable X, $\mathsf{E}[X^2] = \mathsf{E}[XX] = \mathsf{E}[X]\mathsf{E}[X] = \mathsf{E}[X]^2$, because X is not independent of itself. Nevertheless, we now have enough machinery to define:

DEFINITION. The variance of a random variable X is

$$\mathsf{Var}[X] = \mathsf{E}\big[(X - \mathsf{E}[X])^2\big],$$

which we can think of as the average of the squared deviations from the mean. (We might like it to be the average of the deviations from the mean, but that would be 0. We might also have defined it to be the average of the absolute values of the deviations from the mean, but absolute values are more difficult to handle analytically than squares. Furthermore, it is the variance that occurs in the normal distribution, as Fisher pointed out when he introduced the word a century ago [1].)

PROPOSITION. Let X be a random variable and $c \in \mathbb{R}$. Then Var[X + c] = Var[X].

Proof. This follows immediately from the definition:

$$\mathsf{Var}[X+c] = \mathsf{E}\big[(X+c-\mathsf{E}[X+c])^2\big] = \mathsf{E}\big[(X-\mathsf{E}[X])^2\big] = \mathsf{Var}[X],$$

using the linearity of expectation value.

We can also use the linearity of expectation value to expand the definition of variance:

$$\begin{aligned} \mathsf{Var}[X] &= \mathsf{E}\big[X^2 - 2\mathsf{E}[X]X + \mathsf{E}[X]^2\big] \\ &= \mathsf{E}[X^2] - 2\mathsf{E}[X]\mathsf{E}[X] + \mathsf{E}[X]^2 \\ &= \mathsf{E}[X^2] - \mathsf{E}[X]^2. \end{aligned}$$

EXAMPLE. Using this formula we can easily compute the variance for the random variable H defined earlier, $Var[H] = E[H^2] - E[H]^2 = E[H] - E[H]^2 = p - p^2 = p(1-p)$.

We can also use this formula to prove the following:

PROPOSITION. Let X be a random variable and $\alpha \in \mathbb{R}$. Then $\mathsf{Var}[\alpha X] = \alpha^2 \mathsf{Var}[X]$.

So we cannot expect variance to be additive, like expectation value, since $Var[X + X] = Var[2X] = 4Var[X] \neq Var[X] + Var[X]$ (unless Var[X] = 0). Nevertheless, it is sometimes:

PROPOSITION. Let X, Y be random variables. If E[XY] = E[X]E[Y], e.g., when X and Y are independent, then Var[X + Y] = Var[X] + Var[Y].

Proof. Again we use the formula we derived from the definition, together with the linearity of expectation value:

$$\begin{split} \mathsf{Var}[X+Y] &= \mathsf{E}[(X+Y)^2] - \mathsf{E}[X+Y]^2 \\ &= \mathsf{E}[X^2 + 2XY + Y^2] - \left(\mathsf{E}[X] + \mathsf{E}[Y]\right)^2 \\ &= \mathsf{E}[X^2] - \mathsf{E}[X]^2 + \mathsf{E}[Y^2] - \mathsf{E}[Y]^2 + 2\left(\mathsf{E}[XY] - \mathsf{E}[X]\mathsf{E}[Y]\right) \\ &= \mathsf{Var}[X] + \mathsf{Var}[Y], \end{split}$$

using the hypothesis that $\mathsf{E}[XY] = \mathsf{E}[X]\mathsf{E}[Y]$.

We will use these properties of expectation value and variance when we begin to analyze random walks in the next lecture.

Reference

[1] R. A. Fisher, "The correlation between relatives on the supposition of Mendelian inheritance", Transactions of the Royal Society of Edinburgh **52** (1918) 399-433.