Random Variables

COS 341 Fall 2002, lecture 21

Informally, a random variable is the value of a measurement associated with an experiment, e.g. the number of heads in n tosses of a coin. More formally, a random variable is defined as follows:

Definition 1 *A* random variable *over a sample space is a function that maps every sample point (i.e. outcome) to a real number.*

An *indicator random variable* is a special kind of random variable associated with the occurence of an event. The indicator random variable I_A associated with event A has value 1 if event A occurs and has value 0 otherwise. In other words, I_A maps all outcomes in the set A to 1 and all outcomes outside A to 0.

Random variables can be used to define events. In particular, any predicate involving random variables defines the event consisting of all outcomes for which the predicate is true. e.g. for random variables R_1, R_2, $R_1 = 1$ is an event, $R_2 \leq 2$ is an event, $R_1 = 1 \wedge R_2 \leq 2$ is an event.

Events derived from random variables can be used in expressions involving conditional probability as well. e.g.

$$\mathbf{Pr}(R_1 = 1 | R_2 \leq 2) = \frac{\mathbf{Pr}(R_1 = 1 \wedge R_2 \leq 2)}{\mathbf{Pr}(R_2 \leq 2)}$$

Independence of Random Variables

Definition 2 *Two random variables R_1 and R_2 are* independent*, if for all $x_1, x_2 \in \mathbb{R}$, we have:*

$$\mathbf{Pr}(R_1 = x_1 \wedge R_2 = x_2) = \mathbf{Pr}(R_1 = x_1) \cdot \mathbf{Pr}(R_2 = x_2)$$

An alternate definition is as follows:

Definition 3 *Two random variables R_1 and R_2 are* independent*, if for all $x_1, x_2 \in \mathbb{R}$, such that $\mathbf{Pr}(R_2 = x_2) \neq 0$, we have:*

$$\mathbf{Pr}(R_1 = x_1 | R_2 = x_2) = \mathbf{Pr}(R_1 = x_1) \cdot \mathbf{Pr}(R_2 = x_2)$$

In order to prove that two random variables are *not* independent, we need to exhibit a pair of values x_1, X_2 for which the condition in the definition is violated. On the other hand, proving independence requires an argument that the condition in the definition holds for all pairs of values x_1, x_2.

Mutual Independence

Definition 4 *Random variables $R_1, R_2, \ldots, R_t$ are mutually independent if, for all $x_1, x_2, \ldots, x_t \in \mathbb{R}$,*

$$\mathbf{Pr}\left(\bigcap_{i=1}^{t} R_i = x_i\right) = \prod_{i=1}^{t} \mathbf{Pr}(R_i = x_i).$$

Definition 5 *A collection of random variables is said to be k-wise independent if all subsets of k variables are mutually independent.*

Consider a sample space consisting of bit sequences of length 2, where all 4 possible two bit sequences are equally likely. Random variable B_1 is the value of the first bit, B_2 is the value of the second bit and B_3 is $B_1 \oplus B_2$. Here the variables B_1, B_2, B_3 are 2-wise independent, but they are not mutually independent.

Pairwise independence is another name for 2-wise independence, i.e. when we say that a collection of variables is pairwise independent, we mean that they are 2-wise independent.

Probability Density Functions

Probability density functions are used to describe the distribution of a random variable, i.e. the set of values a random variable takes and the probabilities associated with those values. This description of a random variable is independent of any experiment.

Definition 6 *The* probability density function *(pdf) for a random variable X is the function $f_X : (R) \to [0,1]$ defined by:*

$$f_X(t) = \mathbf{Pr}(X = t).$$

For a value t not in the range of X, $f_X(t) = 0$. Note that $\sum_{t \in \mathbb{R}} f_X(t) = 1$,

Definition 7 *The* cumulative distribution function *(cdf) for a random variable X is the function $F_X : \mathbb{R} \to [0,1]$ defined by:*

$$F_X(t) = \mathbf{Pr}(X \le t) = \sum_{s \le t} f_X(s).$$

Two common distributions enountered are the *uniform distribution* and the *binomial distribution.*

Uniform Distribution

Let U be a random variable that takes values in the range $\{1, \ldots, N\}$, such that each value is equally likely. Such a variable is said to be uniformly distributed. The pdf and cdf for this distribution are:

$$f_U(t) = \frac{1}{N}, \qquad F_U(t) = \frac{t}{N}, \text{ for} 1 \le k \le N.$$

Binomial Distribution

Let H be the number of heads in n inpdendent tosses of a biased coin. Each toss of the coin has probability p of being heads and probability $1-p$ of being tails. Such a variable is said to have a *binomial* distribution. The pdf of this distribution is given by

$$f_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

As a sanity check, we can verify that

$$\sum_{k=0}^{n} f_{n,p}(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1$$

Expected Value

Definition 8 *The* expectation $\mathbf{E}[X]$ *of a random variable* X *on a sample space* S *is defined as:*

$$\mathbf{E}[X] = \sum s \in SX(s) \cdot \mathbf{Pr}(\{s\}).$$

An equivalent definition is:

Definition 9 *The* expectation *of a random variable* X *is*

$$\mathbf{E}[X] = \sum_{t \in range(X)} t \cdot \mathbf{Pr}(X = t).$$

If the range of a random variable is non-negative integers, there is an another way to compute the expectation.

Theorem 1 *If* X *is a random variable which takes values in the non-negative integers, then*

$$\mathbf{E}[X] = \sum_{t=0}^{\infty} \mathbf{Pr}(X > i).$$

Proof: Note that

$$\mathbf{Pr}(X > t) = \mathbf{Pr}(X = t+1) + \mathbf{Pr}(X = t+2) + \mathbf{Pr}(X = t+3) + \cdots$$

Thus,

$$\begin{aligned}
\sum_{t=0}^{\infty} \mathbf{Pr}(X > t) &= \mathbf{Pr}(X > 0) + \mathbf{Pr}(X > 1) + \mathbf{Pr}(X > 2) + \cdots \\
&= \mathbf{Pr}(X = 1) + \mathbf{Pr}(X = 2) + \mathbf{Pr}(X = 3) + \cdots \\
&\quad \mathbf{Pr}(X = 2) + \mathbf{Pr}(X = 3) + \mathbf{Pr}(X = 4) + \cdots \\
&\quad \mathbf{Pr}(X = 3) + \mathbf{Pr}(X = 4) + \mathbf{Pr}(X = 5) + \cdots \\
&= 1 \cdot \mathbf{Pr}(X = 1) + 2 \cdot \mathbf{Pr}(X = 2) + 3 \cdot \mathbf{Pr}(X = 3) + \cdots \\
&= \sum_{t=0}^{\infty} t \cdot \mathbf{Pr}(X = t) \\
&= \mathbf{E}[X].
\end{aligned}$$

■

Linearity of Expectation

Theorem 2 (Linearity of Expectation) *For any random variables X_1 and X_2, and constants $c_1, c_2 \in \mathbb{R}$,*

$$\mathbf{E}[c_1 X_1 + c_2 X_2] = c_1 \, \mathbf{E}[X_1] + c_2 \, \mathbf{E}[X_2]$$

Note that the above theorem holds irrespective of the dependence between X_1 and X_2.

Corollary 1 *For any random variables $X_1, \ldots, X_k$, and constants $c_1, \ldots, c_k \in \mathbb{R}$,*

$$\mathbf{E}\left[\sum_{i=1}^{k} c_i X_i\right] = \sum_{i=1}^{k} c_i \, \mathbf{E}[X_i].$$

Conditional Expectation

Definition 10 *We define* conditional expectation*, $\mathbf{E}[X|A]$, of a random variable, given event A, to be*

$$\mathbf{E}[X|A] = \sum_k k \cdot \mathbf{Pr}(X = k|A).$$

The rules for expectation also apply to conditional expectation:

Theorem 3

$$\mathbf{E}[c_1 X_1 + c_2 X_2 | A] = c_1 \, \mathbf{E}[X_1|A] + c_2 \, \mathbf{E}[X_2|A].$$

The following theorem shows how conditional expectation allows us to compute the expectation by case analysis.

Theorem 4 (Law of Total Expectation) *If the sample space is the disjoint union of events $A_1, A_2, \ldots$, then*

$$\mathbf{E}[X] = \sum_i \mathbf{E}[X|A_i] \, \mathbf{Pr}(A_i).$$

Expected value of a product

In general, the expected value of the product of two random variables need not be equal to the product of their expectations. However, this holds when the random variables are independent:

Theorem 5 *For any two* independent *random variables, X_1 and X_2,*

$$\mathbf{E}[X_1 \cdot X_2] = \mathbf{E}[X_1] \cdot \mathbf{E}[X_2].$$

Corollary 2 *If random variables $X_1, X_2, \ldots, X_k$ are mutually independent, then*

$$\mathbf{E}\left[\prod_{i=1}^{k} X_i\right] = \prod_{i=1}^{k} \mathbf{E}[X_i].$$

Note that in general,

$$\mathbf{E}\left[\frac{1}{T}\right] \neq \frac{1}{\mathbf{E}[T]}.$$

Linearity of expectation also holds for infinite sums, provided the summations considered are absolutely convergent:

Theorem 6 (Infinite Linearity of Expectation) *Let $X_1, X_2, \ldots$ be random variables such that $\sum_{i=1}^{\infty} \mathbf{E}[|X_i|]$ converges. Then*

$$\mathbf{E}\left[\sum_{i=1}^{\infty} X_i\right] = \sum_{i-0}^{\infty} \mathbf{E}[X_i].$$

Deviation from the Mean

Theorem 7 (Markov's Theorem) *If X is a* nonnegative *random variable, then for all $t > 0$,*

$$\mathbf{Pr}[X \geq t] \leq \frac{\mathbf{E}[R]}{t}.$$

Proof: We will show that $\mathbf{E}[X] \geq t \cdot \mathbf{Pr}(X > t)$.

$$\begin{aligned}
\mathbf{E}[X] &= \sum_k k \cdot \mathbf{Pr}(X = k) \\
&\leq \sum_{k \geq t} k \cdot \mathbf{Pr}(X = k) \\
&\leq \sum_{k \geq t} t \cdot \mathbf{Pr}(X = k) \\
&= t \cdot \sum_{k \geq t} \mathbf{Pr}(X = k) \\
&= t \cdot \mathbf{Pr}(X \geq t)
\end{aligned}$$

■

Note that in order to apply Markov's theorem to random variable X, X must be nonnegative. Markov's theorem need not hold if X can take negative values.

An alternate to express Markov's theorem is as follows:

Corollary 3 *If X is a nonnegative random variable, then for any $c > 0$,*

$$\mathbf{Pr}(X \geq c \cdot \mathbf{E}[X]) \leq \frac{1}{c}.$$

Definition 11 *The* variance, *$\mathbf{Var}[X]$, of a random variable, X, is:*

$$\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2].$$

Theorem 8 (Chebyshev's Theorem) *Let X be a random variable, then for any $k > 0$,*

$$\mathbf{Pr}(|X - \mathbf{E}[X]| \geq k) \leq \frac{\mathbf{Var}[X]}{k^2}.$$

Proof: Note that

$$\mathbf{Pr}(|X - \mathbf{E}[X]| \geq k) = \mathbf{Pr}((X - \mathbf{E}[X])^2 \geq k^2.$$

Now we apply Markov's theorem to the random variable $(X - \mathbf{E}[X])^2$. This gives us the desired result. ■

Definition 12 *The* standard deviation *of a random variable X is denoted σ_X and is defined to be the square root of the variance:*

$$\sigma_X = \sqrt{\mathbf{Var}\, X} = \sqrt{E[(X - \mathbf{E}[X])^2]}.$$

Chebyshev's theorem can be restated in terms of standard deviation as follows:

Corollary 4 *If X is a random variable, then for any $c > 0$,*

$$\mathbf{Pr}(|X - \mathbf{E}[X]| \geq c \cdot \sigma_X) \leq \frac{1}{c^2}.$$

The variance can also be computed in a alternate, somewhat more convenient way:

Theorem 9

$$\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2. \tag{1}$$

Properties of Variance

Theorem 10 *Let X be a random variable, and let a and b be constants. Then,*

$$\mathbf{Var}[aX + b] = a^2\, \mathbf{Var}[X].$$

Proof: The proof is left as an exercise. Start with an expression for the variance using (1). ■

Theorem 11 *If $X_1, X_2, \ldots, X_n$ are pairwise independent random variables, then*

$$\mathbf{Var}[X_1 + X_2 + \ldots + X_n] = \sum_{i=1}^{n} \mathbf{Var}[X_i].$$

Proof: The proof is left as an exercise. Start with an expression for the variance using (1), and expand the expressions that you obtain. You will need to use Theorem 5 about the expectation of the product of two independent random variables. ■

$$
\begin{aligned}
\mathsf{E}(XY) &= \sum_{\omega\in\Omega} X(\omega)Y(\omega)\mathsf{Pr}(\omega) \\
&= \sum_{x}\sum_{y} xy\cdot \mathsf{Pr}(X=x \text{ and } Y=y) \\
&= \sum_{x}\sum_{y} xy\cdot \mathsf{Pr}(X=x)\mathsf{Pr}(Y=y) \\
&= \left(\sum_{x} x\cdot \mathsf{Pr}(X=x)\right)\left(\sum_{y} y\cdot \mathsf{Pr}(Y=y)\right) \\
&= \mathsf{E}(X)\mathsf{E}(Y),
\end{aligned}
$$

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

Consequently, $\mathsf{Var}(X) \leq \mathsf{E}(X^2)$.

Proof. Using the linearity of expectation and the fact that the expectation of a constant is itself, we have

$$\begin{aligned}
\mathsf{Var}(X) &= \mathsf{E}(X - \mathsf{E}(X))^2 \\
&= \mathsf{E}(X^2 - 2X\mathsf{E}(X) + (\mathsf{E}(X))^2) \\
&= \mathsf{E}(X^2) - 2\mathsf{E}(X)\mathsf{E}(X) + (\mathsf{E}(X))^2 \\
&= \mathsf{E}(X^2) - (\mathsf{E}(X))^2
\end{aligned}$$

Proposition 6.11. *Given a discrete probability space* (Ω, Pr), *for any random variable X and Y, if X and Y are independent, then*

$$\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y).$$

Proof. Recall from Proposition 6.9 that if X and Y are independent, then $\mathsf{E}(XY) = \mathsf{E}(X)\mathsf{E}(Y)$. Then, we have

$$\begin{aligned}
\mathsf{E}((X+Y)^2) &= \mathsf{E}(X^2 + 2XY + Y^2) \\
&= \mathsf{E}(X^2) + 2\mathsf{E}(XY) + \mathsf{E}(Y^2) \\
&= \mathsf{E}(X^2) + 2\mathsf{E}(X)\mathsf{E}(Y) + \mathsf{E}(Y^2).
\end{aligned}$$

Using this, we get

$$\begin{aligned}
\mathsf{Var}(X+Y) &= \mathsf{E}((X+Y)^2) - (\mathsf{E}(X+Y))^2 \\
&= \mathsf{E}(X^2) + 2\mathsf{E}(X)\mathsf{E}(Y) + \mathsf{E}(Y^2) - ((\mathsf{E}(X))^2 + 2\mathsf{E}(X)\mathsf{E}(Y) + (\mathsf{E}(Y))^2) \\
&= \mathsf{E}(X^2) - (\mathsf{E}(X))^2 + \mathsf{E}(Y^2) - (\mathsf{E}(Y))^2 \\
&= \mathsf{Var}(X) + \mathsf{Var}(Y),
\end{aligned}$$