Lecture #20: Random Variables and Expectation
Key Concepts

Some — but not all — of this material is review material.

Random Variables

Definition. Let Ω be a sample space. A ***random variable over Ω*** is a (total) function $X : \Omega \to \mathbb{R}$.

We will often be interested in random variable whose ranges are particular *subsets* V of $\mathbb{R}$ — so that these functions can also be viewed as functions $X : \Omega \to V$ (as well as functions $X : \Omega \to \mathbb{R}$).

Definition. An ***indicator random variable*** over a sample space Ω, is a random variable X such that $X(\sigma) \in \{0, 1\}$ for all $\sigma \in \Omega$.

Once again, let Ω be a sample space and let $X : \Omega \to \mathbb{R}$.

- We will write "$X = r$" as the name of the event

$$\{\sigma \in \Omega \mid X(\sigma) = r\} \subseteq \Omega.$$

- We will write "$X \geq r$" as the name of the event

$$\{\sigma \in \Omega \mid X(\sigma) \geq r\} \subseteq \Omega.$$

- "$X \leq r$", "$X > r$", $X < r$", and $X \neq r$" can be used as the names for (corresponding) events in the same way.

The Expected Value of a Random Variable

Definition. Let Ω be a sample space with probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$, and let $X : \Omega \to \mathbb{R}$ be a random variable over Ω.

Suppose that

$$\sum_{\sigma \in \Omega} \mathsf{P}(\sigma) \times |X(\sigma)|$$

is finite — that is, " less than $+\infty$".[1]

Then the ***expected value of X, with respect to probability distribution P***, is the value

$$\mathsf{E}[X] = \sum_{\sigma \in \Omega} \mathsf{P}(\sigma) \times X(\sigma).$$

Recall, from the lecture on "Conditional Probability", that if Ω is a sample space, $P : \Omega \to \mathbb{R}$ is a probability distribution, and $B \subseteq \Omega$ is an event such that $\mathsf{P}(B) > 0$, then a ***conditional probability distribution*** $P_B : \Omega \to \mathbb{R}$ can be defined by setting

$$\mathsf{P}_B(\sigma) = \begin{cases} \frac{\mathsf{P}(\sigma)}{\mathsf{P}(B)} & \text{if } \sigma \in B, \\ 0 & \text{if } \sigma \notin B \end{cases}$$

for every outcome $\sigma \in \Omega$.

Definition. If X is a random variable then the ***conditional expectation of X given B*** is the expected value of X with the respect to the conditional probability P_B:

$$\mathsf{E}[X \mid B] = \sum_{\sigma \in \Omega} \mathsf{P}_B(\sigma) \times X(\sigma).$$

Theorem 1. *Let Ω be a sample space with probability distribution P : $\Omega \to \mathbb{R}$, let $B \subseteq \Omega$ be an event such that $\mathsf{P}(B) > 0$ and $\mathsf{P}(B^C) > 0$, and let X be a random variable. Then*

$$\mathsf{E}[X] = \mathsf{E}[X \mid B] \times \mathsf{P}(B) + \mathsf{E}[X \mid B^C] \times \mathsf{P}(B^C).$$

[1] This is a "technical restriction" that you will not need to worry about whenever Ω is a finite set.

Linearity of Expectation

Let Ω be a sample space with a probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$, and let $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ be random variables, for some positive integer n. "$X_1 + X_2 + \cdots + X_n$" denotes a random variable such that

$$(X_1 + X_2 + \cdots + X_n)(\sigma) = X_1(\sigma) + X_2(\sigma) + \cdots + X_n(\sigma)$$

for each outcome $\sigma \in \Omega$. Similarly, if $X : \Omega \to \mathbb{R}$ and $a, b \in \mathbb{R}$ then "$a \cdot X + b$" denotes a random variable such that

$$(aX + b)(\sigma) = a \cdot X(\sigma) + b$$

for every outcome $\sigma \in \Omega$, as well.

Theorem 2 (Linearity of Expectation)**.** *Let Ω be a sample space with probability distribution $P : \Omega \to \mathbb{R}$.*

(a) If $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ are random variables over Ω, for a positive integer n, then

$$\mathsf{E}[X_1 + X_2 + \cdots + X_n] = \mathsf{E}[X_1] + \mathsf{E}[X_2] + \cdots + \mathsf{E}[X_n].$$

(b) If $X : \Omega \to \mathbb{R}$ is a random variable over Ω and $a, b \in \mathbb{R}$ then

$$\mathsf{E}[a \cdot X + b] = a \cdot \mathsf{E}[X] + b.$$

Independent Random Variables

Once again, suppose that Ω is a sample space with probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$. Consider random variables $X : \Omega \to V_X$ and $Y : \Omega \to V_Y$, where $V_X, V_Y \subseteq \mathbb{R}$. Recall that, for $a \in V_X$ and $b \in V_Y$,

$$\text{“}X = a\text{”} = \{\sigma \in \Omega \mid X(\sigma) = a\}$$

and

$$\text{“}Y = b\text{”} = \{\tau \in \Omega \mid Y(\tau) = b\},$$

so that

$$\text{“}X = a \wedge Y = b\text{”} = \{\mu \in \Omega \mid X(\mu) = a \text{ and } Y(\mu) = b\}.$$

Definition. The above random variables X and Y are ***independent*** if

$$\mathsf{P}(X = a \wedge Y = b) = \mathsf{P}(X = a) \times \mathsf{P}(Y = b)$$

for all values $a \in V_X$ and $b \in V_Y$.

Theorem 3. *Let Ω be a sample space with probability distribution $P : \Omega \to \mathbb{R}$, and let $X, Y : \Omega \to \mathbb{R}$. If X and Y are independent random variables then*

$$\mathsf{E}[X \times Y] = \mathsf{E}[X] \times \mathsf{E}[Y].$$

Pairwise Independent Random Variables

Once again, let Ω be a sample space with probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$. Let k be a positive integer and let

$$X_1 : \Omega \to V_1, X_2 : \Omega \to V_2, \ldots, X_k : \Omega \to V_k$$

be random variables over Ω (so that $V_1, V_2, \ldots, V_k \subseteq \mathbb{R}$).

Definition. The random variables $X_1, X_2, \ldots, X_k$ are ***pairwise independent*** if X_i and X_j are independent, for every pair of numbers i and j such that $1 \le i < j \le k$.

Mutually Independent Random Variables

Finally, let Ω be a sample space with probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$. Let k be a positive integer and let

$$X_1 : \Omega \to V_1, X_2 : \Omega \to V_2, \ldots, X_k : \Omega \to V_k$$

be random variables over Ω (so that $V_1, V_2, \ldots, V_k \subseteq \mathbb{R}$).

Definition. The random variables $X_1, X_2, \ldots, X_k$ are ***mutually independent*** if the following condition is satisfied: ***For every*** subset $S \subseteq \{1, 2, \ldots, k\}$ and ***for all*** combinations of $a_i \in V_i$, for $i \in S$,

$$\mathsf{P}\left(\bigwedge_{i \in S} (X_i = a_i)\right) = \prod_{i \in S} \mathsf{P}(X_i = a_i).$$