Elements of Probability and Statistics

Lecture 07: Random Variables, Discrete Probability Distributions

IAI, TCG-CREST August 29, 2023

4.1 (Review) Random Variables

Defn. Random Variables: For a random experiment with sample space S, a random variable is a function from S to the real numbers $\mathbb{R}$.

Thus a random variable X assigns a numerical value $X(s)$ to each possible outcome $s \in S$ of the experiment.

An interpretation of random variables is that they provide numerical summaries of experiments, which can be more convenient to work with compared to working directly with arbitrary subsets of the sample space. The well-developed fields of algebra and calculus can also be easily applied on probability functions defined from the real line to the $[0, 1]$ interval.

Ex.1. 2 coin tosses: We can define X to be the number of heads. Thus X assigns 2 to the outcome HH, assigns 1 to the outcomes HT and TH, and 0 to the outcome TT. If Y is the number of tails, we can similarly define Y, and also establish a relationship Y = 2 - X between the random variables. This means that $Y(s) = 2 - X(s)\ \forall s \in S$.

Ex.2. 1 dice roll: Let A be a random variable indicating an even outcome $\{2, 4, 6\}$. Thus $A(s) = 1$ if s is even, otherwise $A(s) = 0$. Similarly, let B be a random variable indicating an odd outcome $\{1, 3, 5\}$, i.e., $B(s) = 1$ if s is odd, otherwise $B(s) = 0$. Then we can establish the following relationship between them, $B = 1 - A$.

(Review) Defn. Discrete Random Variable: A random variable X is said to be a discrete random variable if it maps the outcomes of a random experiment to a finite list of values $a_1, ..., a_n$, or a countably infinite list of values $a_1, a_2, ...$

The probability of any event is then written as $P(X = x) \equiv P(\{s \in S : X(s) = x\})$.

If X is a discrete random variable, then the finite or countably infinite set of of values x such that $P(X = x) > 0$ are called the support of X.

The distribution of a random variable provides a way to specify the probabilities of all events associated with a random variable. There are several equivalent ways of expressing the distribution of a random variable.

(Review) Defn. The Probability Mass Function (PMF) p_X of a discrete random variable X is the function given by $p_X(x) = P(X = x)$.

In general $p_X(x) > 0$ if x is in the support of X, and $p_X(x) = 0$ otherwise.

For a function to be a valid PMF, it must satisfy the following two conditions:

- Non-negativity: $p(x) > 0$ if $x = x_i$ for some i, and $p(x) = 0$ otherwise;
- The PMF must sum to one: $\sum_{i=1}^{\infty} p(x_i) = 1$.

Ex.3. Find the PMF of X, which is the number of heads for two coin tosses. Similarly, find the PMF for Y, which is the number of tails for two coin tosses. What is the distribution of $X = Y$?

Note that X and Y here are two different random variables, but they have the same distribution. This demonstrates that two different random variables can have the same distribution.

There are some distributions that occur quite commonly across a wide range of problems. When working on an unknown problem, identifying a distribution from a similar problem can help in solving it. Hence some popular discrete distributions are discussed next.

(Review) Defn. Bernoulli Distribution: Let a discrete random variable X have only two possible values, described as 0 or 1, or success or failure, etc.. X is said to be distributed as a Bernoulli Distribution with parameter p, or $X \sim \text{Bern}(p)$, if $P(X = 1) = p$ and $P(X = 0) = 1 - p$, where $0 < p < 1$.

Any event A has associated with it a Bernoulli indicator random variable which indicates whether the event has occured or not.

Defn. Indicator Random Variable: The indicator random variable of an event A, which can be written as $I_A \sim \text{Bern}(p)$, equals 1 if A occurs, otherwise it equals 0. Hence $P(I_A = 1) = p = P(A)$, whereas $P(I_A = 0) = 1 - p = 1 - P(A)$.

Defn. Bernoulli Trial: A random experiment that can either result in a 'success' or in a 'failure' is called a Bernoulli trial.

Thus any Bernoulli random variable can be interpreted as an indicator of success for a particular Bernoulli trial; the random variable is one for success and zero for failure.

The extension of this interpretation of Bernoulli random variables naturally leads to describing a distribution for n number of Bernoulli trials.

(Review) Defn. Binomial Distribution: We assume the occurrence of n Bernoulli trials, where the success probability of each trial is p. Let the event X describe the number of successes that have occured. Then X is said to follow a Binomial distribution with parameters n and p, written as $X \sim \text{Bin}(n, p)$, for a positive integer n and $0 < p < 1$. The PMF of X is given by,

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Q. For two coin tosses, describe as Binomial distributions the distributions of random variable X which is the number of heads, and the random variable Y which is the number of tails.

Defn. Expectation of a discrete random variable: If a discrete random variable X has as supports $x_1, x_2, ...$, then the expected value of X is the weighted average of the supports, where the probability of each support is the weight associated with it:

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i) \ .$$

If X and Y are discrete random variables with the same distribution, then $E(X) = E(Y)$.

Ex. Expectation a Bernoulli random variable: Let $X \sim \text{Bern}(p)$, and $q = 1 - p$. Then,

$$E(X) = 1p + 0q = p \ .$$

Def. The probability of an event A is the expected value of its indicator variable I_A, i.e., $P(A) = E(I_A)$.

Theorem. Linearity of Expectation: For any random variables X and Y, and constant c,

$$E(X+Y) = E(X) + E(Y) \ ,$$
$$E(cX) = cE(X) \ .$$

Defn. Variance of a random variable:

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

The square root of the variance is called the standard deviation, i.e., $\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$.

Theorem. For any random variable X,

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

Ex. Variance of a Bernoulli random variable: Let $X \sim \mathrm{Bern}(p)$, and $q = 1-p$. Then,

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = p - p^2 = pq \ .$$

Theorem. The following are important properties of Variance:

- $\mathrm{Var}(X) \geq 0$, with equality iff $P(X = a) = 1$ for some constant a.
- $\mathrm{Var}(X + c) = \mathrm{Var}(X)$.
- $\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$.
- If X and Y are independent, then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(X)$.

A corollary of the above properties: Variance is not linear.

Ex. Binomial Distribution: Let $X \sim \mathrm{Bin}(n,p)$. Then prove,

$$E(X) = np,$$
$$\mathrm{Var}(X) = npq \ .$$

Ex. Geometric Distribution: Let $X \sim \mathrm{Geom}(n,p)$, and therefore $P(X = k) = (1-p)^{k-1}p$. Then prove,

$$E(X) = \frac{1}{p}$$
$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$

Ex. Hypergeometric Distribution: Let $X \sim \mathrm{HypGeom}(n,p)$, and therefore $P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$. Then prove,

$$E(X) = n\frac{w}{w+b}$$

Avisek Gupta
Postdoctoral Fellow
IAI, TCG CREST.