Review:- Probability Distributions
STT 465

In statistics, we have to deal with 'events' and their probabilities very frequently. In order to evaluate the probabilities of the 'events' more efficiently, we often use random variables. A **sample space** is set of all possible 'events'. Let us denote the sample space by S
A **random variable** X is a real valued, measurable function from the sample space S onto the real line $\mathbb{R}$. $X : S \to \mathbb{R}$. Random variables can broadly be classified as either discrete or continuous.
Specific random variables are characterized by their probability distributions. A random variable can be uniquely identified by it's **cumulative distribution function** (*cdf*), usually denoted by F. In other words, $F(a) = P(X \leq a)$

Discrete Random Variable

A random variable that can take on at most a countable number of possible values is said to be **Discrete**. For a discrete random variable X, we define the **probability mass function** (*pmf*) by $p(a) = P(X = a)$ for any $a \in \mathbb{R}$. The *pmf* is positive for at most a countable number of values of a. That is, if X must assume one of the values $x_1, x_2, \ldots$ then one must have the following:-

$$P(x_i) \geq 0 \text{ for all i = 1,2, } \ldots$$

$$\sum_{i=0}^{\infty} P(x_i) = 1$$

We are also interested in moments of the random variable. for a discrete random variable X, define it's **expected value** as

$$E(X) = \mu = \sum_{i=0}^{\infty} x_i p(x_i)$$

The **variance** is defined as

$$Var(X) = E(X - E(X))^2$$

Exercise 1 Let X be a random variable. If $a, b \in \mathbb{R}$ and $b \neq 0$. Prove the following:-
a) $Var(X) = E(X^2) - E^2(X)$
b) $E(a + bX) = a + bE(X)$
c) $Var(a + bX) = b^2 Var(X)$
note that the above results are true only when the corresponding moments $E(X)$, $E(X^2) < \infty$

Bernoulli and Binomial distribution

Suppose that a trial, or an experiment, whose outcome can be classified as either a success or a failure is performed. If we let $X = 1$ when the outcome is a success and $X = 0$ when it is a failure, then the *pmf* of X is given by

$$p(1) = P(X = 1) = p$$

$$p(0) = P(X = 0) = 1 - p$$

where $0 \leq p \leq 1$ is the probability that the trial is a success. A random variable X with the above *pmf* is said to be a **Bernoulli** random variable.

Exercise 2 For Bernoulli random variable X, show that $E(X) = p$, $Var(X) = p(1 - p)$

Suppose now that n independent trials, each of which results in a success with probability p and in a failure with probability $1 - p$, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a **binomial** random variable with parameters (n, p). Thus, a Bernoulli random variable is just a binomial random variable with parameters $(1, p)$.
The *pmf* of a binomial random variable having parameters (n, p) is given by

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \text{ for } k = 0, 1, \ldots, n$$

Exercise 3 Prove the following where $X \sim Bin(n,p)$ with *pmf* $p(k) = P(X=k)$
a)$\sum_{k=0}^{n} p(k) = 1$
b)If $X \sim Bin(n,p)$, then $Y = n - X \sim Bin(n, 1-p)$

Exercise 4 $X_1, X_2, \ldots X_n$ are n independent and identically distributed Bernoulli random variables with success probability p. Define $Y = \sum_{i=1}^{n} X_i$.
a)Show that $Y \sim Bin(n,p)$
b)Use the above representation of the binomial random variable to show that $E(Y) = np$ and $Var(Y) = np(1-p)$.

Continuous Random Variable

Let X be a random variable whose set of possible values is uncountable. We say that X is a **continuous** random variable if there exists a nonnegative function $f(x)$, defined for all $x \in \mathbb{R}$ such that, for any measurable set B of real numbers

$$P(X \in B) = \int_B f(x)dx$$

The function f is called the **probability density function** (*pdf*) of the random variable X.
One must have the following for any *pdf* $f(x)$

$$f(x) \geq 0 \text{ for all } x \in \mathbb{R}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

The **cumulative distribution function** (*cdf*) is defined as

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(X)dx \text{ for all } a \in \mathbb{R}$$

The **expectation** and **variance** of a continuous random variable is defined as following:-

$$E(X) = \mu = \int_{-\infty}^{\infty} xf(x)dx$$

$$Var(X) = E(X - E(X))^2 = E[X - \mu]^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

Exercise 5 Do Exercise 1 part a)-c) for a continuous random variable X with *pdf* $f(x)$.

Normal Distribution

We say that X is a normal random variable with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ if the density of X is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp^{-\frac{1}{2\sigma^2}(x-\mu)^2} \text{ for all } -\infty < x < \infty$$

Properties
$X \sim N(\mu, \sigma^2)$ Define $Z = \frac{X-\mu}{\sigma}$
a) $Z \sim N(0,1)$.
b) $E(Z) = 0$, $Var(Z) = 1$.
c) This density function of Z is a bell-shaped curve that is symmetric about 0. Thus $E(Z) = \text{median}(Z) = mode(Z) = 0$.
d) $E(X) = \mu$ and $Var(X) = \sigma^2$. The density of X is symmetric around μ.
e) If $X_1 \sim N(\mu_1, \sigma_2^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, and they are independent then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. In fact any linear combination of normal random variables will also be normally distributed
f) If $Z_1, Z_2, \ldots, Z_n$ are independent and identically distributed $N(0,1)$ random variables, then $S = \frac{\sum_{i=1}^{n} Z_i}{\sqrt{n}} \sim N(0,1)$. In other words, sum of independent normal random variables is also normal.

Beta Distribution

A random variable is said to have a beta distribution if its density is given by the following

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \qquad \text{for } a,b > 0 \text{ and } 0 \leq x \leq 1$$

where

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}$$

Properties
Let $X \sim Beta(a,b)$ with the above *pdf*.
a) $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ where $\Gamma(k) = \int_0^\infty e^x x^{k-1} dx$ is the gamma function for $k > 0$
b)The gamma function defined above satisfies the recursion $\Gamma(k) = (k-1)\Gamma(k-1)$ as long as $k > 1$
c)Using b) we can show that for any positive integer k, $\Gamma(k) = (k-1)!$
d) $E(X) = \frac{a}{a+b}$, $Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$

Exercise 6 If $X \sim Beta(a,b)$, define $Y = 1 - X$. Then $Y \sim Beta(b,a)$