Chapter 2 – Random variables

Def'n: A random variable is obtained by assigning a numerical value to each outcome of a random experiment.

A discrete random variable is an r.v. that assumes a finite (or countably infinite) range.

→ # of people who drop the course

A continuous random variable is an r.v. with an interval (either finite or infinite) of real numbers for its range.

→ average alcohol intake by a student, average alcohol outtake by a student

Notation: X = random variable; x = particular value;
$P(X = x)$ denotes probability that X equals the value x.

Ex2.1) Toss a coin 3 times. Let X be the number of heads obtained from the tosses.

Table 2X1

x	$P(X = x)$

Discrete Random Variables

The last example involved a discrete random variable. In fact, Table 2X1 represented a probability distribution for X, or a description of the probabilities associated with the possible values of X.

For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, a probability mass function (*pmf*) is a function such that

1. $f(x_i) = P(X = x_i)$
2. $0 \le f(x_i) \le 1$
3. $\sum_{i=1}^{n} f(x_i) = 1$

Ex2.2) Verify that the following is a *pmf*: $f(x) = (3/4)(1/4)^x$, $x = 0, 1, 2, \ldots$

(Proving properties 1 & 2 are trivial.) For Property 3, recognize

$$\sum_{i=0}^{\infty} \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^{x_i} = \sum_{i=1}^{\infty} ar^{i-1}$$

or, in other words, it's a geometric series with $|r| < 1$, so it's convergent. Thus,

$$\sum_{i=0}^{\infty} \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^{x_i} = \frac{3/4}{1-(1/4)} = \frac{3/4}{3/4} = 1$$

Hence, the above function is a *pmf*.

Ex2.3) Using the *pmf* from Ex2.2,

(a) $P(X = 2) =$

(b) $P(X \leq 2) =$

(c) $P(X \geq 2) =$

(d) $P(X \geq 1) =$

2.1.3 Cumulative Distribution Functions

Def'n: The cumulative distribution function of a discrete r.v. X, denoted as $F(x)$, is

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

$F(x)$ satisfies the following properties:

1. $0 \leq F(x) \leq 1$
2. If $x \leq y$, then $F(x) \leq F(y)$

Ex2.4) Consider the *cdf* (corresponding plot drawn in class)

$$F(x) = \begin{cases} 0 & x < 1 \\ 0.7 & 1 \leq x < 4 \\ 0.9 & 4 \leq x < 7 \\ 1 & 7 \leq x \end{cases}$$

Verification of properties should be easy. Determining the *pmf* requires noting that the only points that receive nonzero probability are 1, 4, and 7. The *pmf* at each point is the change in *cdf* at the point. Thus,

a) $P(X \leq 4) =$ b) $P(X > 7) =$ c) $P(X \leq 5) =$

d) $P(X > 4) =$ e) $P(X \leq 2) =$

2.2 Continuous Random Variables

Def'n: For a continuous r.v. X with some interval, a probability density function (*pdf*) is a function such that

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$ (Limits are "generic" here)
3. $P(a \leq X \leq b) = \int_a^b f(x)dx$ (figure drawn in class)
4. $P(X = x) = 0 \rightarrow P(a \leq X \leq b) = P(a < X < b)$

Ex2.5) Suppose the *pdf* of X is

$$f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & elsewhere \end{cases}$$

Determine $P(X < 2)$, $P(2 \leq X < 4)$, and $P(X \geq 4)$.

$P(X < 2) = \int_0^2 f(x)dx =$

$P(2 \leq X < 4) = \int_2^4 f(x)dx =$

$P(X \geq 4) = \int_4^\infty f(x)dx =$

Note: The three probabilities cover the entire range of X. Thus, their sum equals 1.

2.2.3 Cumulative Distribution Functions

Def'n: The cumulative distribution function of a continuous r.v. X is

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u)du \qquad \text{for } -\infty < x < \infty$$

$F(x)$ satisfies the following properties:

1. $0 \leq F(x) \leq 1$
2. If $x \leq y$, then $F(x) \leq F(y)$
3. $P(a \leq X \leq b) = F(b) - F(a)$

Note that $f(x) = \frac{dF(x)}{dx}$

Ex2.6) Using the *pdf* from Ex2.5, find the *cdf*.

$F(x) = P(X \leq x) = \int_0^x f(u)du =$

Therefore,

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x}, & x \geq 0 \end{cases}$$

Note that $P(X < 2) = F(2)$, $P(2 \leq X < 4) = F(4) - F(2)$, and $P(X \geq 4) = 1 - F(4)$. These will give the same answers as above in Ex2.5.

2.3 Expectations of a Random Variable

Def'n: The mean (or, expected value) of the discrete r.v. X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \sum_{i=1} x_i f(x_i) = \sum_{i=1} x_i p_i$$

If X is a discrete random variable with *pmf* of $f(x)$,

$$E[h(X)] = \sum_{i=1} h(x_i) f(x_i)$$

Ex2.7) Using Ex2.1,

$\mu = \sum x_i f(x_i) = \sum x_i P(X = x_i) =$

Ex2.8) Consider $h(X) = X^2$. Then,

$\sum x_i^2 f(x_i) = \sum x_i^2 P(X = x_i) =$

Thus, $E[X^2] = \sum x_i^2 f(x_i) =$

Def'n: The mean (or, expected value) of a continuous random variable X is

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

The median for a continuous random variable X is the value x for which $F(x) = 0.5$.
If X is a continuous r.v. with *pdf* of $f(x)$,

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

Ex2.9) Using Ex2.5, what are E(X) and $E(X^2)$?

Ex2.10) Using Ex2.5, what is the median of X?

2.4 Variance of a Random Variable

The variance of any random variable X, denoted as σ^2 or $V(X)$, is

$$\sigma^2 = Var(X) = V(X) = E[(X-\mu)^2] = E[(X-E(X))^2] \Rightarrow E(X^2) - [E(X)]^2$$

The standard deviation of X is $\sigma = \sqrt{\sigma^2}$

Ex2.11) Using Ex2.7 and Ex2.8,

$\sigma^2 = V(X) = E(X^2) - [E(X)]^2 =$

$\sigma = \sqrt{\sigma^2} =$

Ex2.12) Using Ex2.9,

$\sigma^2 = V(X) = E(X^2) - [E(X)]^2 =$

$\sigma = \sqrt{\sigma^2} =$

2.6 Combinations and Functions of Random Variables

For any constants a and b,

Means:	*Variances*:
1. $E(a) = a$	1. $V(a) = 0$
2. $E(aX) = aE(X)$	2. $V(aX) = a^2V(X)$
3. $E(aX + b) = aE(X) + b$	3. $V(aX + b) = a^2V(X)$
4. $E(aX \pm bY) = aE(X) \pm bE(Y)$	4. $V(aX \pm bY) = a^2V(X) + b^2V(Y) \pm 2ab\text{cov}(X, Y)$

Def'n: Given r.v.'s $X_1, X_2, \ldots, X_n$ and constants $a_1, a_2, \ldots, a_n, b$,

$$Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n + b$$

is a linear combination of $X_1, X_2, \ldots, X_n$.

The fourth rule of each can be extended for any linear combination such that

$$E(Y) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n) + b$$

$$V(Y) = a_1^2V(X_1) + a_2^2V(X_2) + \ldots + a_n^2V(X_n) + 2\sum_{i<j}\sum a_ia_j \text{cov}(X_i, X_j)$$

If $X_1, X_2, \ldots, X_n$ are independent,

$$V(Y) = a_1^2V(X_1) + a_2^2V(X_2) + \ldots + a_n^2V(X_n)$$

Ex2.14) If X_1, X_2, and Y are independent random variables such that

$E(X) = E(X_1) = E(X_2) = 4$ $\quad$ $E(Y) = -3$

$V(X) = V(X_1) = V(X_2) = 3$ $\quad$ $V(Y) = 1$

a) Find the mean and standard deviation of $W = X_1 + X_2$. (Note that $W \neq 2X$)

b) Find the mean and standard deviation of $T = 4X_1 - 3Y - \pi$.

If $X_1, X_2, \ldots, X_n$ are independent, random variables with $E(X_i) = \mu$ and $V(X_i) = \sigma^2$, then

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

is a random variable with

$$E(\overline{X}) = \mu \qquad \text{and} \qquad V(\overline{X}) = \frac{\sigma^2}{n}$$