Probability and Statistics I

STAT 3600 – Fall 2021

Le Chen
lzc0090@auburn.edu

Last updated on
July 4, 2021

Auburn University
Auburn AL

Chapter 2. Discrete Distributions

Definition 2.1-1 Given a random experiment with an outcome space S, a function X that assigns one and only one real number $X(s) = x$ to each element s in S is called a *random variable*.
The *space* of X is the set of real numbers $\{x : X(s) = x, s \in S\}$, where $s \in S$ means that the element s belongs to the set S.

Example 2.1-1 In the experiment of tossing a coin once, we might define the r.v. X as

$$X(H) = 1 \quad \text{and} \quad X(T) = 0.$$

Note that we could also define another r.v., say Y or Z, with

$$Y(H) = 0,\ Y(T) = 1 \qquad \text{or} \qquad Z(H) = 0,\ Z(T) = 0.$$

Example 2.1-2 Consider an experiment in which a person is selected at random from some population and her height in inches is measured. This height is a random variable.

Events defined by random variables

If X is a r.v. and x is a fixed real number, we can define the event $(X = x)$ as

$$(X = x) = \{s : X(s) = x\}.$$

Similarly, for fixed numbers x, x_1,and x_2 , we can define the following events:

$$(X \leq x) = \{s : X(s) \leq x\}$$
$$(X > x) = \{s : X(s) > x\}$$
$$(x_1 < X \leq x_2) = \{s : x_1 < X(s) \leq x_2\}$$

These events have probabilities that are denoted by

$$\begin{aligned}
\mathbb{P}(X = x) &= \mathbb{P}\{s : X(s) = x\} \\
\mathbb{P}(X \leq x) &= \mathbb{P}\{s : X(s) \leq x\} \\
\mathbb{P}(X > x) &= \mathbb{P}\{s : X(s) > x\} \\
\mathbb{P}(x_1 < X \leq x_2) &= \mathbb{P}\{s : x_1 < X(s) \leq x_2\} \\
\mathbb{P}(X \in C) &= \mathbb{P}(s : X(s) \in C).
\end{aligned}$$

Definition 2.1-2 Let X be a random variable. The *distribution* of X is the collection of all probabilities of the form $\mathbb{P}(X \in C)$ for all sets C of real numbers such that $\{X \in C\}$ is an event.

Example 2.1-3 In the experiment of tossing a fair coin three times, the sample space S, consists of eight equally likely sample points $S = \{HHH, ..., TTT\}$. If X is the r.v. giving the number of heads obtained, find the distribution of X.

Definition 2.1-3 A random variable X is called *discrete* if it takes a finite or countable number (sequence) of values:

$$X \in \{x_1, x_2, x_3, ...\} .$$

It is completely described by telling the probability of each outcome. Distribution defined by:

$$\mathbb{P}(X = x_k) = f(x_k), \quad k = 1, 2, \cdots$$

is called the *probability mass function (p.m.f.)* of the discrete r.v.

Definition 2.1-4 The closure of the set $\{x : f(x) > 0\}$ is called the *support of (the distribution of)* X.

Theorem 2.1-1 (Properties of probability mass function) Let X be a discrete r.v. It is p.m.f. $f(x) = \mathbb{P}(X = x)$ satisfies the following properties:

(a) $f(x) \geq 0$, for all $x \in S$;

(b) $\sum_{x \in S} f(x) = 1$;

(c) $\mathbb{P}(X \in A) = \sum_{x \in A} f(x)$ where $A \subset S$.

Definition 2.1-5 (Cumulative distribution function) We call the function defined by

$$F(x) = \mathbb{P}\left(X \le x\right), \qquad -\infty < x < \infty,$$

the *cumulative distribution function* and abbreviate it as *cdf*.

Definition 2.1-6 When a pmf is constant on the space or support, we say that the distribution is *uniform* over that space.

Example 2.1-4 Let X be a discrete r.v. over a finite number of values $\{1, 2, 3, ..., m\}$ with the pmf:

$$f(x) = \frac{1}{m}, \qquad x = 1, 2, 3, \cdots, m.$$

Then X is uniform over $\{1, 2, 3, ..., m\}$. Its cdf is given by

$$F(x) = \mathbb{P}(X \leq x) = \begin{cases} 0, & x < 1, \\ \dfrac{k}{m}, & k \leq x < k+1, \\ 1, & m \leq x. \end{cases}$$

This is a (right-continuous) step function. Draw it...

Example 2.1-5 Let

$$f(x) = \begin{cases} \frac{3}{4}\left(\frac{1}{4}\right)^x & \text{if } x = 0, 1, 2, \cdots \\ 0 & \text{otherwise.} \end{cases}$$

(a) Verify that the function $f(x)$ defined by is a pmf of a discrete r.v. X.
(b) Find (i) $f(2) = \mathbb{P}(X = 2)$; (ii) $\mathbb{P}(X \leq 2)$; (iii) $\mathbb{P}(X \geq 1)$.

Example 2.1-6 Let X be the number of accidents per week in a factory. Let the pmf of X be

$$f(x) = \frac{1}{(x+1)(x+2)},\ x = 0, 1, 2, \cdots .$$

Find the conditional probability of $X \geq 4$, given that $X \geq 1$.

Definition 2.2-1 If $f(x)$ is the p.m.f. of the random variable X of the discrete type with space S, and if the summation

$$\sum_{x \in S} u(x)f(x), \quad \text{which is sometimes written} \quad \sum_{S} u(x)f(x),$$

exists, then the sum is called the *mathematical expectation* or the *expected value* of the function $u(X)$, and it is denoted by $\mathbb{E}[u(X)]$. That is,

$$\mathbb{E}[u(X)] = \sum_{x \in S} u(x)f(x).$$

Remark 2.2-1 The usual definition of mathematical expectation of $u(X)$ requires that the sum *converge absolutely*, that is, that

$$\sum_{x \in S} |u(x)| f(x)$$

converge and be finite. The reason for the absolute convergence is that it allows one, in the advanced proof of

$$\sum_{x \in S_X} |u(x)| f(x) = \sum_{y \in S_Y} |y| g(y),$$

to rearrange the order of the terms in the $x-$summation.

Example 2.2-1 Let the random variable X have the p.m.f.

$$f(x) = \frac{1}{5}, \quad x \in S,$$

where $S = \{-2, -1, 0, 1, 2\}$. Then find (a) $\mathbb{E}[X]$; (b) $\mathbb{E}[X^2]$; (c) $\mathbb{E}[X^3]$.

Theorem 2.2-1 When it exists, the mathematical expectation $\mathbb{E}$ satisfies the following properties:

(a) If c is a constant, then $\mathbb{E}(c) = c$.

(b) If c is a constant and u is a function, the

$$\mathbb{E}[cu(X)] = c\mathbb{E}[u(X)].$$

(c) If c_1 and c_2 are constants and u_1 and u_2 are functions, then

$$\mathbb{E}[c_1 u_1(X) + c_2 u_2(X)] = c_1\mathbb{E}[u(X)] + c_2\mathbb{E}[u_2(X)].$$

Example 2.2-2 For the pmf given in Example 2.2-1, find (a) $\mathbb{E}[X(3 - 2X)]$; (b) $\mathbb{E}[3X^2 + 4X^3 - 5]$.

Example 2.2-3 Let X have a *hypergeometric distribution* in which n objects are selected from $N = N_1 + N_2$, that is, the pmf is given by

$$f(x) = \mathbb{P}(X = x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}, \quad x \in \{0, 1, \cdots, n\}.$$

Then find $\mathbb{E}[X]$.

Example 2.2-4 In the casino game called *high-low*, there are three possible bets. Assume that \$1 is the size of the bet. A pair of fair six-sided dice is rolled and their sum is calculated. If you bet *low*, you win \$1 if the sum of the dice is $\{2, 3, 4, 5, 6\}$. If you bet *high*, you win \$1 if the sum of the dice is $\{8, 9, 10, 11, 12\}$. If you bet on $\{7\}$, you win \$4 if the sum is 7 is rolled. Otherwise, you lose on each of the three bets. In all three cases, your original dollar is returned if you win. Find the expected value of the game to the bettor for each of these three bets.

Exercises from textbook: 2.2-1, 2.2-2, 2.2-4-2.2-7, 2.2-11, 2.2-12

Definition 2.3-1 The *mean (or expected value)* of a discrete r.v. X, denoted by μ or $\mathbb{E}(X)$, is defined by

$$\mu = \mathbb{E}(X) = \sum_{x \in S} x f(x).$$

Example 2.3-1 Say an experiment has probability of success p, where $0 < p < 1$, and probability of failure $q = 1 - p$. This experiment is repeated independently until the first success occurs; say this happen on the X trial. Clearly the space of X is $S_x = \{1, 2, 3, 4, \cdots\}$.

What is $\mathbb{P}(X = x)$, where $x \in S_x$?

We must observe $x - 1$ failures and then a success to have this happen. Thus, due the independence, the probability is

$$f(x) = \mathbb{P}(X = x) = q \cdot q \cdots q \cdot p = q^{x-1}p, \;\; x \in S_X.$$

Since p and q are positive, this is a pmf because

$$\sum_{x \in S_X} q^{x-1}p = \sum_{x=0}^{\infty} q^{x-1}p = \frac{p}{1-q} = \frac{p}{p} = 1.$$

This distribution is called *geometric distribution*.

The mean of this distribution is

$$\mu = \sum_{x=1}^{\infty} x f(x) = (1)p + (2)qp + (3)q^2 p + \cdots$$

and

$$q\mu = (q)p + (2)q^2 p + (3)q^3 p + \cdots .$$

If we subtract these second of these two equations from the first, we have

$$\begin{aligned} (1-q)\mu &= p + pq + pq^2 + pq^3 + \cdots \\ &= \sum_{x=0}^{\infty} pq^x = \frac{p}{1-q} = 1. \end{aligned}$$

That is,

$$\mu = \frac{1}{1-q} = \frac{1}{p}.$$

For illustration, if $p = 1/10$, we would expect $\mu = 10$ trials are needed on average to observe a success. This certainly agrees with out intuition.

Definition 2.3-2 The rth *moment about the origin* of a discrete r.v. X is defined by

$$\mathbb{E}(X^n) = \sum_{x \in S} x^n f(x).$$

Definition 2.3-3 The *variance* of a discrete r.v. X, denoted by σ^2 or $\mathrm{Var}(X)$, is defined by

$$\sigma^2 = \mathrm{Var}(X) = \mathbb{E}\{[X - \mathbb{E}(X)]^2\} = \sum_{x \in S} (x - \mu)^2 f(x).$$

Definition 2.3-4 The *standard deviation* of a r.v. X, denoted by σ, is the positive square root of $\mathrm{Var}(X)$, i.e.,

$$\sigma = \sqrt{\mathrm{Var}(X)}.$$

Example 2.3-2 Consider a discrete r.v. X whose p.m.f. is given by

$$f(x) = \begin{cases} \frac{1}{3} & \text{if } x = -1, 0, 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find the mean, variance, and standard deviation of X.

Example 2.3-3 Let a r.v. X denote the outcome of throwing a fair die. Find the mean and variance of X.

Example 2.3-4 Find the mean and variance of a r.v. X, which has a uniform distribution on the first m positive integers.
Hint: Use the formula:

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2} \qquad \text{and} \qquad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Definition 2.3-5 The *nth moment about the point b* is defined as

$$\mathbb{E}[(X-b)^n] = \sum_{x \in S} (x-b)^n f(x).$$

Remark 2.3-1 $\mathrm{Var}(X)$ is the second moment about the mean μ.

Definition 2.3-6 The *rth factorial moment* is defined as

$$\mathbb{E}[(X)_r] = \mathbb{E}[X(X-1)(X-2)\cdots(X-r+1)]$$

Remark 2.3-2

$$\begin{aligned}\sigma^2 = \mathrm{Var}(X) &= \mathbb{E}[X(X-1)] + \mathbb{E}(X) - [\mathbb{E}(X)]^2 \\ &= \mathbb{E}[(X)_2] + \mu - \mu^2.\end{aligned}$$

Example 2.3-5 Find the variance of of the hypergeometric distribution considered in Example 2.3-3.

Hint: $\mathbb{E}(X) = n\frac{N_1}{N}$ and $\mathbb{E}[X(X-1)] = \frac{n(n-1)(N_1)(N_1-1)}{N(N-1)}$.

Definition 2.3-7 Let X be a random variable of the discrete type with pmf $f(x)$ and space S. If there is a positive number h such that

$$\mathbb{E}(e^{tX}) = \sum_{x \in S} e^{tx} f(x)$$

exists and is finite for $-h < t < t$, then the function defined by

$$M(t) = \mathbb{E}(e^{tX})$$

is called the *moment-generating function of* X (or of the distribution of X). This function is often abbreviated as mgf.

Properties of moment generating function

1. $M(0) = 1$.

2. If the space of S is $\{b_1, b_2, b_3, \cdots\}$, then the moment generating function is given by the expansion

$$M(t) = e^{tb_1} f(b_1) + e^{tb_2} f(b_2) + e^{tb_3} f(b_3) + \cdots .$$

Thus, the coefficient of e^{tb_i} is the probability

$$f(b_i) = \mathbb{P}(X = b_i).$$

3. If the moment generating function exists, then

$$M'(0) = \mathbb{E}(X) = \mu$$

$$M''(0) = \mathbb{E}(X^2)$$

and, in general,

$$M^{(r)}(0) = \mathbb{E}(X^r)$$

Example 2.3-6 Suppose X has the geometric distribution of Example 2.3-1; that is, the pmf is

$$f(x) = q^{x-1}p, \quad x = 1, 2, 3, \cdots .$$

Find mgf of X.

Example 2.3-7 Define the p.m.f. and give the values of μ, σ^2, and σ when the moment generating function of X is given by (a) $M(t) = 1/3 + (2/3)e^t$; and (b) $M(t) = (0.25 + 0.75e^t)$.

Example 2.3-8 If the moment generating function of X is

$$M(t) = \frac{2}{5}e^t + \frac{1}{5}e^{2t} + \frac{2}{5}e^{3t}.$$

Find the pmf, mean, and variance.

Example 2.3-9 Suppose the mgf of X is

$$M(t) = \frac{e^t/2}{1 - e^t/2}, \qquad t < \ln(2).$$

Find the pmf, mean, and variance.

Hint: Use $(1-z)^{-1} = 1 + z + z^2 + z^3 + \cdots, \qquad -1 < z < 1.$

Exercises from textbook: Section 2.3: 1, 2, 3, 4, 5, 8, 9, 11, 13, 19

Definition 2.4-1 A r.v. X is called a *Bernoulli* r.v. with parameter p if its p.m.f. is given by

$$f(x) = \mathbb{P}(X = x) = p^x (1-p)^{1-x}, \qquad x = 0 \text{ or } 1,$$

where $0 \leq p \leq 1$.

A Bernoulli r.v. X is associated with some experiment where an outcome can be classified as either a "success" or a "failure," and the probability of a success is p and the probability of a failure is $q = 1 - p$. Such experiments are often called Bernoulli trials.

Theorem 2.4-1 The mean and variance of the Bernoulli r.v. X are

$$\mu = \mathbb{E}(X) = p,$$
$$\sigma^2 = \mathrm{Var}(X) = p(1-p).$$

Definition 2.4-2 A r.v. X is called a *binomial* r.v. with parameters (n, p) if its pmf is given by

$$f(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \cdots, n$$

where $0 \leq p \leq 1$ and

$$\binom{n}{x} = \frac{n!}{x!(n-x)!},$$

which is known as the binomial coefficient.

Remark 2.4-1 Recall that if n is positive integer, then

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} b^x a^{n-x}.$$

Thus, if we use binomial expansion, then sum of the binomial probabilities is

$$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [p+(1-p)]^n = 1,$$

a result that had to follow from the fact that $f(x)$ is pmf.

Remark 2.4-2 A binomial r.v. X is associated with some experiments in which n independent Bernoulli trials are performed and X represents the number of successes that occur in the n trials. Note that a Bernoulli r.v. is just a binomial r.v. with parameters $(1, p)$.

We now use the binomial expansion to find the mgf for a binomial random variable and then the mean and variance.

Theorem 2.4-2 Let $X \sim \text{binom}(n, p)$. Then

$$M_X(t) = [(1-p) + pe^t]^n, \quad -\infty < t < \infty.$$

Solution. The mgf is given by

$$\begin{aligned} M_X(t) &= E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= [(1-p) + pe^t]^n, \quad -\infty < t < \infty, \end{aligned}$$

from the expansion of $(a+b)^n$ with $a = 1-p$ and $b = pe^t$. $\square$

Theorem 2.4-3 Let $X \sim \text{binom}(n, p)$. Then

$$\mu = np \quad \text{and} \quad \sigma^2 = np(1-p).$$

Solution. By Theorem 2,

$$\mu = \mathbb{E}(X) = M'(0) = n[(1-p) + pe^t]^{n-1}(pe^t)\Big|_{t=0} = np,$$

$$\sigma^2 = \text{Var}(X) = \mathbb{E}[X^2] - (E[X])^2 = M''(0) - [M'(0)]^2 = np(1-p).$$

$\square$

Example 2.4-1 A binary source generates digits 1 and 0 randomly with probabilities 0.6 and 0.4, respectively.
(a) What is the probability that two 1s and three 0s will occur in a five-digit sequence?
(b) What is the probability that at least three 1s will occur in a five-digit sequence?

Example 2.4-2 A fair coin is flipped 10 times. Find the probability of the occurrence of 5 or 6 heads.

Example 2.4-3 For $0 \le p \le 1$, and $n = 2, 3, \cdots$, determine the value of

$$\sum_{x=2}^{n} x(x-1)\binom{n}{x} p^x (1-p)^{n-x}.$$

Exercises from textbook: Section 2.4: 1, 3, 4, 5, 7abc, 10, 17, 20.

Consider a collection of $N = N_1 + N_2$ similar objects, N_1 of them belonging to one of the two dichotomous classes (red chips, say) and N_2 of them belonging to the second class (blue chips, say).

A collection of n objects is selected from these N objects at random and without replacement.

Find the probability that exactly x of these n objects belong to the first class and $n - x$ belong to the second. Clearly, we need

$$0 \le x \le N_1 \quad \text{and} \quad 0 \le n - x \le N_2, \tag{1}$$

which are equivalent to

$$\max(n - N_2, 0) \le x \le \min(n, N_1).$$

We can select x objects from the fist class in any one of $\binom{N_1}{x}$ ways and $n-x$ objects from the second class in any one of $\binom{N_2}{n-x}$ ways.

By multiplication principle, the product $\binom{N_1}{x}\binom{N_2}{n-x}$ equals the number of ways the joint operation can be performed.

If we assume that each of the $\binom{N}{n}$ ways of selecting n objects from $N = N_1 + N_2$ objects has the same probability, it follows that the desired probability is

$$f(x) = \mathbb{P}(X = x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}, \qquad \max(n - N_2, 0) \le x \le \min(n, N_1).$$

Then we say the random variable X has a hypergeometric distribution with parameters N_1, N_2 and n, denoted as $HG(N_1, N_2, n)$.

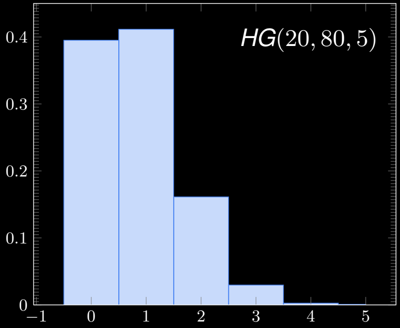
Example 2.5-1 A lot (collection) consisting of 100 fuses is inspected by the following procedure: Five fuses are chased at random and tested; if all five blow at the correct amperage, the lot is accepted. Suppose that the lot contains 20 defective fuses. If X is a random variable equal to the number of defective fuses in the sample of 5, the probability of accepting is

$$\mathbb{P}(X = 0) = \frac{\binom{20}{0}\binom{80}{5}}{\binom{100}{5}} = 0.3193.$$

More generally, the pmf of X is

$$f(x) = \mathbb{P}(X = x) = \frac{\binom{20}{x}\binom{80}{5-x}}{\binom{100}{5}}, \qquad x = 0, 1, 2, 3, 4, 5.$$

x	0	1	2	3	4	5
$f(x)$	$\frac{149380}{378131}$	$\frac{933625}{2268786}$	$\frac{182875}{1134393}$	$\frac{78375}{2646917}$	$\frac{2375}{934206}$	$\frac{38}{467103}$
approx.	0.3951	0.4115	0.1612	0.02961	0.002542	0.00008135



Theorem 2.5-1 Suppose that X follows $HG(N_1, N_2, n)$. Then

$$\mathbb{E}(X) = n\left(\frac{N_1}{N}\right) \qquad \text{and} \qquad \mathrm{Var}(X) = n\left(\frac{N_1}{N}\right)\left(\frac{N_2}{N}\right).$$

Remark 2.5-1 Check Examples 2.2-3 and 2.3-5.

Let the random variable X denote the number of trials needed to observe the rth success in a sequence of independent Bernoulli trials. That is, X is the trial number on which the rth success is observed.

By the multiplication rule of probabilities, the pmf of X-say, $g(x)$- equals the product of the probability

$$\binom{x-1}{r-1} p^{r-1}(1-p)^{x-r} = \binom{x-1}{r-1} p^{r-1} q^{x-r}$$

of obtaining exactly $r-1$ successes in the first $x-1$ trials and the probability p of success on the rth trial. Thus, the pmf of X is

$$g(x) = \binom{x-1}{r-1} p^{r}(1-p)^{x-r} = \binom{x-1}{r-1} p^{r} q^{x-r}, \quad x = r, r+1, \cdots .$$

We say that X has a negative binomial distribution with parameter (r, p).

Remark 2.6-1 The reason for calling this distribution the negative binomial distribution is as follows:

Consider $h(w) = (1-w)^{-r}$, the binomial $(1-w)$ with the negative exponent $-r$. Using Maclaurin's series expansion, we have

$$(1-w)^{-r} = \sum_{k=0}^{\infty} \frac{h^{(k)}(0)}{k!} w^k = \sum_{k=0}^{\infty} \binom{r+k-1}{r-1} w^k, \quad -1 < w < 1.$$

If we let $x = k + r$ in the summation, then $k = x - r$ and

$$(1-w)^{-r} = \sum_{x=r}^{\infty} \binom{r+x-r-1}{r-1} w^{x-r} = \sum_{x=r}^{\infty} \binom{x-1}{r-1} w^{x-r},$$

the summand of which is, expect for the factor p^r, the negative binomial probability when $w = q$. In particular, the sum of the probabilities for the negative binomial distribution is 1 because

$$\sum_{x=r}^{\infty} g(x) = \sum_{x=r}^{\infty} \binom{x-1}{r-1} p^r q^{x-r} = p^r (1-q)^{-r} = 1.$$

The case $r = 1$

If $r = 1$ in the negative binomial distribution, we note that X has a geometric distribution, since the pmf consists of the term of a geometric series, namely,

$$g(x) = p(1-p)^x, \quad x = 1, 2, 3, \cdots$$

Remark 2.6-2 Recall that for a geometric, the sum is given by

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r} \qquad \text{when } |r| < 1.$$

Thus, for the geometric distribution,

$$\sum_{x=1}^{\infty} g(x) = \sum_{x=1}^{\infty} (1-p)^{x-1} p = \frac{p}{1-(1-p)} = 1,$$

so that $g(x)$ does satisfy the properties of a pmf.

From the sum of a geometric series, we also note that when k is an integer,

$$\mathbb{P}(X > k) = \sum_{x=k+1}^{\infty} (1-p)^{x-1} p = \frac{(1-p)^k p}{1-(1-p)} = (1-p)^k = q^k.$$

Thus, the value of the cdf at a positive integer k is

$$\mathbb{P}(X \leq k) = \sum_{x=1}^{k} (1-p)^{x-1} p = 1 - P(X > k) = 1 - q^k.$$

General case $r \geq 1$

Theorem 2.6-1 Let X follow a negative binomial distribution with parameters (r, p). Then

$$\mathbb{E}(X) = \frac{r}{p} \quad \text{and} \quad \mathrm{Var}(X) = \frac{rq}{p}, \quad \text{where } q = 1 - p.$$

This theorem is proved by the following example.

Example 2.6-1 Show that the moment generating function of negative binomial random variable X is

$$M(t) = \frac{(pe^t)^r}{[1-(1-p)e^t]^r}, \qquad \text{where } t < -\ln(1-p).$$

Example 2.6-2 Suppose that a sequence of independent tosses are made with a coin for which the probability of obtaining a head on each given toss is $1/30$.
(a) What is the expected number of tosses that will be required in order to obtain five heads?
(b) What is the variance of the number of tosses that will be required in order to obtain five heads?

Remark 2.6-3 Recall that when the moment-generating function exists, derivatives of all orders exist at $t = 0$. Thus, it is possible to represent $M(t)$ as a Maclaurin's series, namely,

$$M(t) = M(0) + M^{'}(0)\left(\frac{t}{1!}\right) + M^{''}(0)\left(\frac{t^2}{2!}\right) + M^{'''}(0)\left(\frac{t^3}{3!}\right) + \cdots$$

Here, $M^{(k)}(0)$ gives the k-th moment.

On the other hand, in many cases, knowing all moments can help us determine the underlying r.v. or distribution.

Example 2.6-3 Let $\mathbb{E}(X^r) = 5^r$, $r = 1, 2, 3, \cdots$. Find the moment-generating function $M(t)$ of X and the pmf of X.

Example 2.6-4 Consider the experiment of throwing a fair dice.
(a) Find the probability that it will take less than six tosses to throw a 6.
(b) Find the probability that it will take more than six tosses to throw a 6.
(c) Find the average number of rolls required in order to obtain a 6.

Exercises form textbook: Section 2.6: 1, 2, 3, 4, 6, 7, 8.

Definition 2.7-1 Let the number of occurrences of some event in a given continuous interval be counted. Then we have an *approximate Poisson process* with parameter $\lambda > 0$ if the following conditions are satisfied:

(a) The numbers of occurrences in non overlapping subintervals are independent.

(b) The probability of exactly one occurrence in a sufficiently short subinterval of length h is approximately λh.

(c) The probability of two or more occurrences in a sufficiently short subinterval is essentially zero.

Definition 2.7-2 A r.v. X is called a *Poisson* r.v. with parameter $\lambda(>0)$ if its pmf is given by

$$f(x) = \mathbb{P}(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \qquad x = 0, 1, 2, 3, \cdots .$$

The corresponding cdf of X is

$$F(x) = e^{-\lambda}\sum_{k=0}^{n}\frac{\lambda^k}{k!} \qquad n \leq x < n+1.$$

The moment-generating function of Poisson r.v. X is

$$M(t) = \mathbb{E}(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)},$$

from which one obtain the mean and variance of the Poisson r.v. X

$$\mu = \mathbb{E}(X) = \lambda \quad \text{and} \quad \sigma^2 = \mathrm{Var}(X) = \lambda.$$

Remark 2.7-1 In the case of large n and small p, we have that

$$\binom{n}{k} p^k (1-p)^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!} \qquad np = \lambda$$

which indicates that the binomial distribution can be approximated by the Poisson distribution.

The Poisson r.v. has a tremendous range of applications in diverse areas because it may be used as an approximation for binomial r.v. with parameters (n, p) when n is large and p is small enough so that np is of a moderate size.

Some examples of Poisson r.v.'s include

1. The number of telephone calls arriving at a switching center during various intervals of time
2. The number of misprints on a page of a book
3. The number of customers entering a bank during various intervals of time.

Example 2.7-1 A noisy transmission channel has a per-digit error probability $p = 0.01$.

(a) Calculate the probability of more than one error in 10 received digits.

(b) Repeat (a), using the Poisson approximation.

Example 2.7-2 The number of telephone calls arriving at a switchboard during any $10-$minute period is known to be a Poisson r.v. X with $\lambda = 2$.
(a) Find the probability that more than three calls will arrive during any 10-minute period.
(b) Find the probability that no calls will arrive during any $10-$minute period.

Exercises from textbook: 2.7-1, 2.7-2,2.7-3, 2.7-5,2.7-9, 2.7-11.