4.1 Introduction

It is frequently the case when an experiment is performed that we are mainly interested in some function of the outcome as opposed to the actual outcome itself.

- In tossing dice, we are often interested in the sum of the two dice and are not really concerned about the separate values of each die.
- We may be interested in knowing that the sum is 7 and not be concerned over whether the actual outcome was $(1, 6)$ or $(2, 5)$ or $(3, 4)$ or $(4, 3)$ or $(5, 2)$ or $(6, 1)$.
- In coin flipping, we may be interested in the total number of heads that occur and not care at all about the actual head-tail sequence that results.
- These quantities of interest, or more formally, these real-valued functions defined on

the sample space, are known as **random variables**.

Example 4.1a. Suppose that our experiment consists of tossing 3 fair coins. If we let Y denote the number of heads appearing, then Y is a random variable taking on one of the values 0, 1, 2, 3 with respective probabilities

- $P(Y = 0) = P(T, T, T) = \frac{1}{8}$

$$\begin{aligned} P(Y = 1) &= P\{(T, T, H), (T, H, T), (H, T, T)\} \\ &= \frac{3}{8} \\ P(Y = 2) &= P\{(T, H, H), (H, T, H), (H, H, T)\} \\ &= \frac{3}{8} \end{aligned}$$

- $P(Y = 3) = P(H, H, H) = \frac{1}{8}$
- We must have

$$1 = P\left(\bigcup_{i=0}^{3} \{Y = i\}\right) = \sum_{i=0}^{3} P\{Y = i\}$$

Example 4.1b. Three balls are to be randomly selected without replacement from an urn containing 20 balls numbered 1 through 20. If we bet that at least one of the drawn balls has a number as large as or larger than 17, what is the probability that we win the bet?

- X: The largest number selected.
- $P\{X = i\} = \frac{\binom{i-1}{2}}{\binom{20}{3}} \quad i = 3, \ldots, 20$
- From above:

$$
\begin{aligned}
P\{X = 20\} &= \frac{\binom{19}{2}}{\binom{20}{3}} = \frac{3}{20} = .150 \\
P\{X = 19\} &= \frac{\binom{18}{2}}{\binom{20}{3}} = \frac{51}{380} \approx .134 \\
P\{X = 18\} &= \frac{\binom{17}{2}}{\binom{20}{3}} = \frac{34}{285} \approx .119 \\
P\{X = 17\} &= \frac{\binom{16}{2}}{\binom{20}{3}} = \frac{2}{19} \approx .105
\end{aligned}
$$

- $P(X \geq 17) \approx .105 + .119 + .134 + .150 = .508$

Example 4.1c. Independent trials, consisting of the flipping of a coin having probability p of coming up heads, are continually performed until either a head occurs or a total of n flips is made.

- X: The number of times the coin is flipped.

$$\begin{aligned}
P\{X=1\} &= P\{H\} = p \\
P\{X=2\} &= P\{(T,H)\} = (1-p)p \\
P\{X=3\} &= P\{(T,T,H)\} = (1-p)^2 p \\
&\vdots \\
P\{X=n-1\} &= P\{(\underbrace{T,T,\ldots,T}_{n-2},H)\} \\
&= (1-p)^{n-2}p \\
P\{X=n\} &= P\{(\underbrace{T,T,\ldots,T}_{n-1},T),(\underbrace{T,T,\ldots,T}_{n-1},H)\} \\
&= (1-p)^{n-1}
\end{aligned}$$

- As a check:

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n}\{X=i\}\right) &= \sum_{i=1}^{n} P\{X=i\} \\
&= \sum_{i=1}^{n-1} p(1-p)^{i-1} + (1-p)^{n-1} \\
&= p\left[\frac{1-(1-p)^{n-1}}{1-(1-p)}\right] + (1-p)^{n-1} \\
&= 1-(1-p)^{n-1} + (1-p)^{n-1} \\
&= 1
\end{aligned}$$

Example 4.1d. Three balls are randomly chosen from an urn containing 3 white, 3 red, and 5 black balls. Suppose that we win \$1 for each white ball selected and lose \$1 for each red selected.

- X: Total winnings from the experiment.
- $P\{X = 0\} = \frac{\binom{5}{3}+\binom{3}{1}\binom{3}{1}\binom{5}{1}}{\binom{11}{3}} = \frac{55}{165}$
- $P\{X = 1\} = P\{X = -1\} = \frac{\binom{3}{1}\binom{5}{2}+\binom{3}{2}\binom{3}{1}}{\binom{11}{3}} = \frac{39}{165}$
- $P\{X = 2\} = P\{X = -2\} = \frac{\binom{3}{2}\binom{5}{1}}{\binom{11}{3}} = \frac{15}{165}$
- $P\{X = 3\} = P\{X = -3\} = \frac{\binom{3}{3}}{\binom{11}{3}} = \frac{1}{165}$
- $\sum_{i=0}^{3} P\{X = i\} + \sum_{i=1}^{3} P\{X = -i\}$
  $= \frac{55 + 39 + 15 + 1 + 39 + 15 + 1}{165} = 1$

- The probability that we win money is

$$\sum_{i=1}^{3} P\{X = i\} = \frac{55}{165} = \frac{1}{3}$$

Example 4.1e. Suppose that there are N distinct types of coupons and each time one obtains a coupon it is, independent of prior selections, equally likely to be any one of the N types.

- T: The number of coupons that needs to be collected until one obtains a complete set of at least one of each type.
- A_j: The event that no type j coupon is contained among the first $n, j = 1, \ldots, N$.

$$\begin{aligned} P\{T > n\} &= P\left(\bigcup_{j=1}^{N} A_j\right) \\ &= \sum_j P(A_j) - \sum\sum_{j_1<j_2}(A_{j_1}A_{j_2}) + \cdots \\ &\quad +(-1)^{k+1}\sum\sum\sum_{j_1<j_2<\cdots<j_k} P(A_{j_1}A_{j_2}\cdots A_{j_k}) \cdots \\ &\quad +(-1)^{N+1}P(A_1A_2\cdots A_N) \end{aligned}$$

- $P(A_j) = \left(\frac{N-1}{N}\right)^n$
- $P(A_{j_1}A_{j_2}) = \left(\frac{N-2}{N}\right)^n$

- $P(A_{j_1}A_{j_2}\cdots A_{j_k}) = \left(\frac{N-k}{N}\right)^n$
- We see that for $n > 0$,

$$\begin{aligned} P\{T > n\} &= N\left(\frac{N-1}{N}\right)^n - \binom{N}{2}\left(\frac{N-2}{N}\right)^n + \binom{N}{3}\left(\frac{N-3}{N}\right)^n - \cdots \\ &\quad +(-1)^N\binom{N}{N-1}\left(\frac{1}{N}\right)^n \\ &= \sum_{i=1}^{N-1}\binom{N}{i}\left(\frac{N-i}{N}\right)^n(-1)^{i+1} \end{aligned}$$

- $P\{T = n\} = P\{T > n-1\} - P\{T > n\}$
- D_n: The number of distinct types of coupons that contained in the first n selections.
- A: each is one of these k types.
- B: each of these k types is represented.
- We see that

$$P(A) = \left(\frac{k}{N}\right)^n$$

$$P(B|A) = 1 - \sum_{i=1}^{k-1}\binom{k}{i}\left(\frac{k-i}{k}\right)^n(-1)^{i+1}$$

- There are $\binom{N}{k}$ possible choices for the set of k types.

$$\begin{aligned} P\{D_n = k\} &= \binom{N}{k} P(AB) \\ &= \binom{N}{k} \left(\frac{k}{N}\right)^n \left[1 - \sum_{i=1}^{k-1} \binom{k}{i} \left(\frac{k-i}{k}\right)^n (-1)^{i+1}\right] \end{aligned}$$

Remark.

- Since one must collect at least N coupons to obtain a compete set, it follows that $P\{T > n\} = 1$ if $n < N$.
- From Eq. (1.2):

$$\sum_{i=1}^{N-1} \binom{N}{i} \left(\frac{N-i}{N}\right)^n (-1)^{i+1} = 1$$

- $\sum_{i=0}^{N-1} \binom{N}{i} \left(\frac{N-i}{N}\right)^n (-1)^{i+1} = 0$
- Set $j = N - i$,

$$\sum_{j=1}^{N} \binom{N}{j} j^n (-1)^{j-1} = 0$$

4.2 Distribution functions

- The cumulative distribution function (c.d.f.) of the random variable X:

$$F(b) = P\{X \leq b\} \qquad -\infty < b < \infty$$

- Some properties of the c.d.f. F:

1. F is a nondecreasing function; that is, if $a < b$, then $F(a) \leq F(b)$.
2. $\lim_{b \to \infty} F(b) = 1$.
3. $\lim_{b \to -\infty} F(b) = 0$.
4. F is right continuous. That is, for any b and any decreasing sequence b_n, $n \geq 1$, that converges to b, $\lim_{n \to \infty} F(b_n) = F(b)$.

Example 4.2a. The distribution function of the random variable X is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \le x < 1 \\ \frac{2}{3} & 1 \le x < 2 \\ \frac{11}{12} & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

A graph of $F(x)$ is presented in Fig. 4.1.

(a)

$$\begin{aligned} P\{X < 3\} &= \lim_n P\left\{X \le 3 - \frac{1}{n}\right\} \\ &= \lim_n F\left(3 - \frac{1}{n}\right) = \frac{11}{12} \end{aligned}$$

(b)

$$\begin{aligned} P\{X = 1\} &= P\{X \le 1\} - P\{X < 1\} \\ &= F(1) - \lim_n F\left(1 - \frac{1}{n}\right) \\ &= \frac{2}{3} - \frac{1}{2} = \frac{1}{6} \end{aligned}$$

(c) $P\left\{X > \frac{1}{2}\right\} = 1 - P\left\{X \leq \frac{1}{2}\right\} = 1 - F\left(\frac{1}{2}\right) = \frac{3}{4}$

(d) $P\{2 < X \leq 4\} = F(4) - F(2) = \frac{1}{12}$

4.3 Discrete random variable

For a discrete random variable X, we define the **probability mass function** $p(a)$ of X by

$$p(a) = P\{X = a\}$$

- X must assume one of the values $x_1, x_2, \ldots$
- $p(x_i) \geq 0 \quad i = 1, 2, \ldots$
- $p(x) = 0$ all other values of x
- $\sum_{i=1}^{\infty} p(x_i) = 1$
- If the probability mass function of X is

$$p(0) = \frac{1}{4} \quad p(1) = \frac{1}{2} \quad p(2) = \frac{1}{4}$$

we can represent this graphically as shown in Fig. 4.2.

- A graph of the probability mass function of the random variable representing the sum when two dice are rolled looks like the one shown in Fig. 4.3.

Example 4.3a. The probability mass function of a random variable X is given by $p(i) = c\lambda^i/i!$, $i = 0, 1, 2, \ldots$, where λ is some positive value. Find (a) $P\{X = 0\}$ and (b) $P\{X > 2\}$.

(a) Since $\sum_{i=0}^{\infty} p(i) = 1$, we have that

$$c \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = ce^{\lambda} = 1$$

- $c = e^{-\lambda}$
- $P\{X = 0\} = e^{-\lambda}\lambda^0/0! = e^{-\lambda}$
- X has a Poisson(λ) distribution.

(b)

$$P\{X > 2\} = 1 - P\{X \leq 2\}$$

$$= 1 - P\{X = 0\} - P\{X = 1\}$$
$$-P\{X = 2\}$$
$$= 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2}$$

- The cumulative distribution function F:

$$F(a) = \sum_{x \leq a} p(x)$$

- If X is a discrete random variable whose possible values are $x_1, x_2, x_3, \ldots$, where $x_1 < x_2 < x_3 < \cdots$, then its distribution function is a step function.

- If the probability mass function of X is

$$p(1) = \frac{1}{4} \quad p(2) = \frac{1}{2} \quad p(3) = \frac{1}{8} \quad p(4) = \frac{1}{8}$$

then its cumulative distribution function is

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{4} & 1 \leq a < 2 \\ \frac{3}{4} & 2 \leq a < 3 \\ \frac{7}{8} & 3 \leq a < 4 \\ 1 & 4 \leq a \end{cases}$$

4.4 Expected value

Expected value:

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

$$E[X] = \sum_{i=1}^{n} x_i p(x_i)$$

- The expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes it.
- If $p(0) = p(1) = \frac{1}{2}$, then $E[X] = 0(\frac{1}{2}) + 1(\frac{1}{2}) = \frac{1}{2}$.
- If $p(0) = \frac{1}{3}, p(1) = \frac{2}{3}$, then $E[X] = 0(\frac{1}{3}) + 1(\frac{2}{3}) = \frac{2}{3}$.
- If an infinite sequence of independent replications of an experiment is performed, then for any event, the proportion of time that E occurs will be $P(E)$.
- Consider a random variable that must take

on one of the values $X - 1, x_2, \ldots, x_n$ with respective probabilities $p(x_1), p(x_2), \ldots, p(x_n)$; and think of X as representing our winnings in a single game of chance.

- Now by the frequency interpretation, it follows that if we continually play this game, then the proportion of time that we win x_i will be $p(x_i)$.
- The average winnings per game will be

$$\sum_{i=1}^{n} x_i p(x_i) = E[X]$$

Example 4.4a. Find $E[X]$ where X is the outcome when we roll a fair die.

- $p(i) = \frac{1}{6},\ i = 1, 2, \ldots, 6.$
- $E[X] = 1(\frac{1}{6}) + 2(\frac{1}{6}) + 3(\frac{1}{6}) + 4(\frac{1}{6}) + 5(\frac{1}{6}) + 6(\frac{1}{6}) = \frac{7}{2}.$

Example 4.4b. We say that I is an indicator variable for the event A if

$$I = \begin{cases} 1 \text{ if } A \text{ occurs} \\ 0 \text{ if } A^c \text{ occurs} \end{cases}$$

Find $E[I]$.

- $p(1) = P(A), p(0) = 1 - P(A)$.
- We have that $E[I] = P(A)$.

Example 4.4c. A contestant on a quiz show is presented with two questions, questions 1 and 2, which he is to attempt to answer in some order chosen by him. If he decides to try question i, then he will be allowed to go on to question j, $j \neq i$ only if his answer to i is correct. If his initial answer is incorrect, he is not allowed to answer the other question. The contestant is to receive V_i dollars if he answers question i correctly, $i = 1, 2$. Thus, for instance, he will receive $V_1 + V_2$ dollars if both questions are correctly answered. If the probability that he knows the answer to question i is P_i, $i = 1, 2$, which question should he attempt first so as to maximize his expected winnings? Assume that the events E_i, $i = 1, 2$, that he knows the answer to question i, are independent events.

- If he attempts question 1 first, then he will win

$$\begin{array}{ll} 0 & \text{with probability } 1 - P_1 \\ V_1 & \text{with probability } P_1(1 - P_2) \\ V_1 + V_2 & \text{with probability } P_1 P_2 \end{array}$$

- His expected winnings in this case will be

$$V_1 P_1 (1 - P_2) + (V_1 + V_2) P_1 P_2$$

- If he attempts question 2 first, his expected winnings will be

$$V_2 P_2 (1 - P_1) + (V_1 + V_2) P_1 P_2$$

- It is better to try question 1 first if

$$V_1 P_1 (1 - P_2) \geq V_2 P_2 (1 - P_1)$$

equivalently, if $\dfrac{V_1 P_1}{1 - P_1} \geq \dfrac{V_2 P_2}{1 - P_2}$.

- If he is 60 percent certain of answering question 1, worth \$200, correctly and he is 80 percent certain of answering question 2, worth \$100, correctly, then he should attempt question 2 first because
$$400 = \frac{(100)(.8)}{.2} > \frac{(200)(.6)}{.4} = 300$$

Example 4.4d. A school class of 120 students are driven in 3 buses to a symphonic performance. There are 36 students in one of the buses, 40 in another, and 44 in the third bus. When the buses arrive, one of the 120 students is randomly chosen. Let X denote the number of students on the bus of that randomly chosen student, and find $E[X]$.

- $P\{X = 36\} = \frac{36}{120}$
- $P\{X = 40\} = \frac{40}{120}$
- $P\{X = 44\} = \frac{44}{120}$
- $E[X] = 36(\frac{3}{10}) + 40(\frac{1}{3}) + 44(\frac{11}{30}) = \frac{1208}{30} = 40.2667$

- The average number of students on a bus is $120/3 = 40$.
- The more students there are on a bus, then more likely a randomly chosen student would have been on that bus.
- Buses with many students are given more weight than those with fewer students.

The concept of expectation is analogous to the physical concept of the center of gravity of a distribution of mass (Fig. 4.5).

4.5 Expectation of a function of a random variable

$$E[g(X)]$$

Example 4.5a. Let X denote a random variable that takes on any of the values $-1, 0, 1$ with respective probabilities

$$P\{X = -1\} = .2 \quad P\{X = 0\} = .5$$

$$P\{X = 1\} = .3$$

Compute $E[X^2]$.

- Letting $Y = X^2$.
- $P\{Y = 1\} = P\{X = -1\} + P\{X = 1\} = .5$
- $P\{Y = 0\} = P\{X = 0\} = .5$
- $E[X^2] = E[Y] = 1(.5) + 0(.5) = .5$

Proposition 5.1: If X is a discrete random variable that takes on one of the values x_i, $i \geq 1$, with respective probabilities $p(x_i)$, then for any real-valued function g

$$E[g(X)] = \sum_i g(x_i)p(x_i)$$

Example 4.5b. A product, sold seasonally, yields a net profit of b dollars for each unit sold and a net loss of ℓ dollars for each unit left unsold when the season ends. The number of units of the product that are ordered at a specific department store during any season is a random variable having probability mass function $p(i)$, $i \geq 0$. If the store must stock this product in advance, determine the number of units the store should stock so as to maximize its expected profit.

- X: The number of units ordered.
- If s units are stocked, then the profit

$$P(s) = \begin{cases} bX - (s - X)\ell & \text{if } X \leq s \\ sb & \text{if } X > s \end{cases}$$

- The expected profit equals

$$\begin{aligned} E[P(s)] &= \sum_{i=0}^{s} [bi - (s-i)\ell]p(i) + \sum_{i=s+1}^{\infty} sbp(i) \\ &= (b+\ell)\sum_{i=0}^{s} ip(i) - s\ell \sum_{i=0}^{s} p(i) + sb\left[1 - \sum_{i=0}^{s} p(i)\right] \\ &= (b+\ell)\sum_{i=0}^{s} ip(i) - (b+\ell)s\sum_{i=0}^{s} p(i) + sb \\ &= sb + (b+\ell)\sum_{i=0}^{s}(i-s)p(i) \end{aligned}$$

- To determine the optimal value of s

$$\begin{aligned} E[P(s+1)] &= b(s+1) + (b+\ell)\sum_{i=0}^{s+1}(i-s-1)p(i) \\ &= b(s+1) + (b+\ell)\sum_{i=0}^{s}(i-s-1)p(i) \end{aligned}$$

- $E[P(s+1)] - E[P(s)] = b - (b+\ell) \sum_{i=0}^{s} p(i)$
- Stocking $s+1$ units will be better than stocking s units whenever

$$\sum_{i=0}^{s} p(i) < \frac{b}{b+\ell}$$

- Stocking s^*+1 items will lead to a maximum expected profit where s^* is the largest value of s satisfying the above inequality.

$$E[P(0)] < \cdots < E[P(s^*)] < E[P(s^*+1)] > E[P(s^*+2)] > \cdots$$

Corollary 5.1: If a and b constants, then

$$E[aX+b] = aE[X] + b$$

Mean: The weighted average of the possible values of X.

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

nth moment:

$$E[X^n] = \sum_{x:p(x)>0} x^n p(x)$$

4.6 Variance

$$W = 0 \quad \text{with probability } 1$$

$$Y = \begin{cases} -1 \text{ with probability } \frac{1}{2} \\ +1 \text{ with probability } \frac{1}{2} \end{cases}$$

$$Z = \begin{cases} -100 \text{ with probability } \frac{1}{2} \\ +100 \text{ with probability } \frac{1}{2} \end{cases}$$

- All have the same expectation, 0.
- There is much greater spread in possible value of Y than in those of W and in the possible values of Z than in those of Y.
- A reasonable way of measuring the possible variation of X would be to look at how far apart X would be from its mean on the average.
- Candidate $E[|X-\mu|]$ is inconvenient to deal with.

Definition: If X is a random variable with mean μ, then the variance of X, denoted by $\text{Var}(X)$, is defined by

$$\text{Var}(X) = E[(X-\mu)^2] = E[X^2] - \mu^2$$

Example 4.6a. Calculate $\text{Var}(X)$ if X represents the outcome when a fair die is rolled.

- Shown in Example 4.4a that $E[X] = \frac{7}{2}$.
-

$$\begin{aligned} E[X^2] &= 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right) \\ &= \left(\frac{1}{6}\right)(91) \end{aligned}$$

- $\text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$

Proposition: For any constants a and b,

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

In the terminology of mechanics, the variance variance represents the **moment of inertia**.

Standard deviation

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$

4.7 The Bernoulli and Binomial random variables

Bernoulli random variable:

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

Binomial random variable:

$$p(i) = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \quad i = 0, 1, \ldots, n$$

Example 4.7a. Five fair coins are flipped. If the outcomes are assumed independent, find the probability mass function of the number of heads obtained.

- Let X equal the number of heads (successes) that appear, then X is a binomial random variable with parameters $(n = 5, p = \frac{1}{2})$.

- $$P\{X=0\} = \binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = \frac{1}{32}$$
$$P\{X=1\} = \binom{5}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4 = \frac{5}{32}$$
$$P\{X=2\} = \binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \frac{10}{32}$$
$$P\{X=3\} = \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{10}{32}$$
$$P\{X=4\} = \binom{5}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1 = \frac{5}{32}$$
$$P\{X=5\} = \binom{5}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0 = \frac{1}{32}$$

Example 4.7b. It is known that screws produced by a certain company will be defective with probability .01 independently of each other. The company sells the screws in packages of 10 and offers a money-back guarantee that at most 1 of the 10 screws is defective. What proportion of package sold must the company replace?

- X: The number of defective screws in a package
- Then X is a binomial random variable with parameters (10, .01).
- The probability that a package will have to be replaced is
  $1 - P\{X = 0\} - P\{X = 1\}$
  $= 1 - \binom{10}{0}(.01)^0(.99)^{10} - \binom{10}{1}(.01)^1(.99)^9$
  $\approx .004$

Example 4.7c. The following gambling game, known as the wheel of fortune (or chuck-a-luck), is quite popular at many carnivals and gambling casinos: A player bets on one of the numbers 1 through 6. Three dice are then rolled, and if the number bet by the player appears i times, $i = 1, 2, 3$, then the player wins i units; on the other hand, if the number bet by the player does not appear on any of the dice, then the player loses 1 unit. Is this game fair to the player? (Actually, the game is played by spinning a wheel that comes to rest on a slot labeled by three of the numbers 1 through 6, but it is mathematically equivalent to the dice version.)

- Assume that the dice are fair and act independently of each other, then the number of times that the number bet appears is a binomial random variable with parameters $(3, \frac{1}{6})$.
- X: The player's winnings in the game, we have

$$P\{X = -1\} = \binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3 = \frac{125}{216}$$

$$P\{X = 1\} = \binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2 = \frac{75}{216}$$

$$P\{X = 2\} = \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1 = \frac{15}{216}$$

$$P\{X = 3\} = \binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0 = \frac{1}{216}$$

- $E[X] = \frac{-125+75+30+3}{216} = \frac{-17}{216}$

In the next example we consider the simplest form of the theory of inheritance as developed by G. Mendel (1822-1884).

Example 4.7d. Suppose that a particular trait (such as eye color or left handedness) of a person is classified on the basis of one pair of genes and suppose that d represents a dominant gene and r a recessive gene. Thus a person with dd genes is pure dominant, one with rr is pure recessive, and one with rd is hybrid. The pure dominant and the hybrid are alike in appearance. Children receive 1 gene from each parent. If, with respect to a particular trait, 2 hybrid parents have a total of 4 children, what is the probability that 3 of the 4 children have the outward appearance of the dominant gene?

- Assume that each child is equally likely to inherit either of 2 genes from each parent, the probabilities that the child of 2 hybrid parents will have dd, rr, or rd pairs of genes are, respectively, $1/4, 1/4, 1/2$.

- An offspring will have the outward appearance of the dominant gene if its gene pair is either dd or rd.
- The number of such children is $B(4, 3/4)$.
- The desired probability is

$$\binom{4}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^1 = \frac{27}{64}$$

Example 4.7e. Consider a jury trial in which it takes 8 of the 12 jurors to convict; that is, in order for the defendant to be convicted, at least 8 of the jurors must vote him guilty. If we assume that jurors act independently and each makes the right decision with probability θ, what is the probability that the jury renders a correct decision?

- If he is guilty, the probability of a correct decision is

$$\sum_{i=8}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$$

- If the defendant is innocent, the probability of the jury's rendering a correct decision is

$$\sum_{i=5}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$$

- If α represents the probability that the defendant is guilty, then, by conditioning on whether or not he is guilty, we obtain that the probability that the jury renders a correct decision is

$$\alpha \sum_{i=8}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i} + (1-\alpha) \sum_{i=5}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$$

Example 4.7f. A communication system consists of n components, each of which will, independently, function with probability p. The total system will be able to operate effectively if at least one-half of its components function.

(a) For what values of p is a 5-component system more likely to operate effectively than a 3-component system?

(b) In general, when is a $(2k+1)$-component system better than a $(2k-1)$-component system?

(a) • As the number of functioning components is a binomial random variable with parameters (n, p).

• The probability that a 5-component system will be effective is

$$\binom{5}{3}p^3(1-p)^2 + \binom{5}{4}p^4(1-p) + p^5$$

• The corresponding probability for a 3-component system is

$$\binom{3}{2}p^2(1-p) + p^3$$

• The 5-component system is better if

$$10p^3(1-p)^2 + 5p^4(1-p) + p^5 > 3p^2(1-p) + p^3$$

which reduces to

$$3(p-1)^2(2p-1) > 0$$

or

$$p > \frac{1}{2}$$

(b)
- In general, a system with $2k+1$ components will be better than one with $2k-1$ components if and only if $p > 1/2$.
- X: The number of the first $2k-1$ that function.
- P_{2k+1}(effective)
$= P\{X \geq k+1\} + P\{X = k\}(1-(1-p)^2) + P\{X = k-1\}p^2$
which follows since the $(2k+1)$-component system will be effective if either
  - (i) $X \geq k+1$;
  - (ii) $X = k$ and at least one of the remaining 2 components function; or
  - (iii) $X = k-1$ and both of the next 2 functions.

$$\begin{aligned} P_{2k-1}(\text{effective}) &= P\{X \geq k\} \\ &= P\{X = k\} + P\{X \geq k+1 \end{aligned}$$

$P_{2k+1}(\text{effective}) - P_{2k-1}(\text{effective})$
$= P\{X = k-1\}p^2 - (1-p)^2 P\{X = k\}$
$= \binom{2k-1}{k-1} p^{k-1}(1-p)^k p^2 - (1-p)^2 \binom{2k-1}{k} p^k (1-p)^{k-1}$
$= \binom{2k-1}{k} p^k (1-p)^k [p - (1-p)]$ since $\binom{2k-1}{k-1} = \binom{2k-1}{k}$
$> 0 \Leftrightarrow p > \frac{1}{2}$

4.7.1 Properties of binomial random variable

-
$$\begin{aligned} E[X^k] &= \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=1}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i} \\ & \quad i\binom{n}{i} = n\binom{n-1}{i-1} \\ &= np \sum_{i=1}^{n} i^{k-1} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \\ &= np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \end{aligned}$$

$$= npE[(Y+1)^{k-1}]$$

where Y is a binomial random variable with parameters $(n-1, p)$.

- $k = 1$, $E[X] = np$
- $k = 2$,

$$\begin{aligned} E[X^2] &= npE[Y+1] \\ &= np[(n-1)p+1] \end{aligned}$$

- $$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= np[(n-1)p+1] - (np)^2 \\ &= np(1-p) \end{aligned}$$

Proposition: If X is a binomial random variable with parameters n and p, then

$$\begin{aligned} E[X] &= np \\ \text{Var}(X) &= np(1-p) \end{aligned}$$

Proposition 7.1: If X is a binomial random variable with parameters (n, p), where $0 < p < 1$, then as k goes from 0 to n, $P\{X = k\}$ first increases monotonically and then decreases monotonically, reaching its largest value when k is the largest integer less than or equal to $(n+1)p$.

Example 4.7g.

- In a U.S. presidential election the candidate who gains the maximum number of votes in a state is awarded the total number of electoral college votes allocated to that state.
- The number of electoral college votes of a given state is roughly proportional to the population of that state – that us, a state of population size n has roughly nc electoral votes.
- Let us determine the average power in a close presidential election of a citizen in a state of size n, where by *average power* in

a close election, we mean the following:

- A vote in a state of size $n = 2k + 1$ will be decisive if the other $n - 1$ voters split their votes split their votes evenly between the two candidates.
- P {voter in state of size $2k + 1$ makes a difference }
  $= \binom{2k}{k}(\frac{1}{2})^k(\frac{1}{2})^k = \frac{(2k)!}{k!k!2^{2k}}$
- Make use of Stirling's approximation, which says that for k large,
  $$k! \sim k^{k+1/2}e^{-k}\sqrt{2\pi}$$
  where we say that $a_k \sim b_k$ when the ratio a_k/b_k approaches 1 as k approaches ∞.
- P {voter in state of size $2k + 1$ makes a difference }
  $\sim \frac{(2k)^{2k+1/2}e^{-2k}\sqrt{2\pi}}{k^{2k+1}e^{-2k}(2\pi)2^{2k}} = \frac{1}{\sqrt{k\pi}}.$
- Average power $= ncP\{\text{makes a difference}\}$
  $$\sim \frac{nc}{\sqrt{n\pi/2}}$$
  $$= c\sqrt{2n/\pi}.$$

4.7.2 Computing the binomial distribution function

- Suppose that X is $B(n,p)$.
- The key to computing its distribution function

$$P\{X \leq i\} = \sum_{k=0}^{i} \binom{n}{k} p^k (1-p)^{n-k} \quad i = 0, 1,$$

$$P\{X = k+1\} = \frac{p}{1-p}\frac{n-k}{k+1} P\{X = k\}$$

Example 4.7h.

- $X \sim B(6, .4)$.

$$P\{X = 0\} = (.6)^6 \approx .0467$$
$$P\{X = 1\} = \frac{4}{6}\frac{6}{1} P\{X = 0\} \approx .1866$$
$$P\{X = 2\} = \frac{4}{6}\frac{5}{2} P\{X = 1\} \approx .3110$$
$$P\{X = 3\} = \frac{4}{6}\frac{4}{3} P\{X = 2\} \approx .2765$$
$$P\{X = 4\} = \frac{4}{6}\frac{3}{4} P\{X = 3\} \approx .1382$$

$$P\{X = 5\} = \frac{42}{65} P\{X = 4\} \approx .0369$$
$$P\{X = 6\} = \frac{41}{66} P\{X = 5\} \approx .0041$$

Example 4.7i. If X is a $B(100, .75)$, find $P\{X = 70\}$ and $P\{X \leq 70\}$.

- $P\{X = 70\} \approx .04575381$
- $P\{X \leq 70\} \approx .14954105$

4.8 The Poisson random variable

Poisson probability distribution:

$$p(i) = P\{X = i\} = e^{-\lambda} \frac{\lambda^i}{i!} \quad i = 0, 1, 2, \ldots$$

- The Poisson random variable has a tremendous range of applications in diverse areas because it may be used as an approximation for a $B(n, p)$ when n is large and p is small enough so that np is a moderate size.

- If X is $B(n,p)$ and let $\lambda = np$. Then

$$\begin{aligned} P\{X=i\} &= \frac{n!}{(n-i)!i!}p^i(1-p)^{n-i} \\ &= \frac{n!}{(n-i)!i!}(\frac{\lambda}{n})^i(1-\frac{\lambda}{n})^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)}{n^i}\frac{\lambda^i}{i!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^i} \end{aligned}$$

- For n large and λ moderate,

$$\begin{aligned} \left(1-\frac{\lambda}{n}\right)^n &\approx e^{-\lambda} \\ \frac{n(n-1)\cdots(n-i+1)}{n^i} &\approx 1 \\ \left(1-\frac{\lambda}{n}\right)^i &\approx 1 \\ P\{X=i\} &\approx e^{-\lambda}\frac{\lambda^i}{i!} \end{aligned}$$

Examples of Poisson random variable:

1. The number of misprints on a page (or a group of pages) of a book.
2. The number of people in a community living to 100 years of age.

3. The number of wrong telephone numbers that are dialed in a day.
4. The number of packages of dog biscuits sold in a particular store each day.
5. The number of customers entering a post office on a given day.
6. The number of vacancies occurring during a year in the Supreme Court.
7. The number of α-particles discharged in a fixed period of time from some radioactive material.

Example 4.8a. Suppose that the number of typographical errors on a single page of this book has a Poison distribution with parameter $\lambda = \frac{1}{2}$. Calculate the probability that there is at least one error on this page.

- X: Denote the number of errors on this page.

- $P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-1/2} \approx .393$

Example 4.8b. Suppose that the probability that an item produced by a certain machine will be defective is .1. Find the probability that a sample of 10 items will contain at most 1 defective item.

- The desired probability is

$$\binom{10}{0}(.1)^0(.9)^{1}0 + \binom{10}{1}(.1)^1(.9)^9 = .7361$$

- The Poisson approximation yields the value $e^{-1} + e^{-1} \approx .7358$.

Example 4.8c. Consider an experiment that consists of counting the number of α-particles given off in a 1-second interval by 1 gram of radioactive material. If we know from past experience that, on the average, 3.2 such α-particles are given off, what is a good approximation to the probability that no more than 2 α-particles will appear?

- $X \sim$ Poisson(3.2)
- The desired probability is

$$P\{X \leq 2\} = e^{-3.2} + 3.2e^{-3.2} + \frac{(3.2)^2}{2}e^{-3.2}$$
$$\approx .3799$$

Before computing the expected value and variance of the Poisson random variable with parameter λ, recall that this random variable approximates a $B(n, p)$ when n is large, p is small, and $\lambda = np$.

- $np = \lambda$
- $np(1 - p) \approx \lambda$

Recursive relation for moments:

$$E[X^k] = \lambda E[(X + 1)^{k-1}]$$

Mean:

$$
\begin{aligned}
E[X] &= \sum_{i=0}^{\infty} \frac{ie^{-\lambda}\lambda^i}{i!} \\
&= \lambda \sum_{i=1}^{\infty} \frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\
&= \lambda
\end{aligned}
$$

Second moment:

$$
\begin{aligned}
E[X^2] &= \sum_{i=0}^{\infty} \frac{i^2 e^{-\lambda}\lambda^i}{i!} \\
&= \lambda \sum_{i=1}^{\infty} \frac{ie^{-\lambda}\lambda^{i-1}}{(i-1)!} \\
&= \lambda \sum_{j=0}^{\infty} \frac{(j+1)e^{-\lambda}\lambda^j}{j!} \\
&= \lambda[\sum_{j=0}^{\infty} \frac{je^{-\lambda}\lambda^j}{j!} + \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!}] \\
&= \lambda(\lambda+1)
\end{aligned}
$$

Variance: $\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \lambda$

Proposition: The expected value and variance of a Poisson random variable are both equal to its parameter λ.

Another use of the Poisson probability distribution arises in situations where "events" occur at certain points in time.

A Poisson random variable is usually a good approximation for diverse phenomena:

1. The number of earthquakes during some fixed time span.
2. The number of people enters a particular establishment (bank, post office, gas station, and so on).
3. The number of wars per year.
4. The number of electrons emitted from a heated cathode during a fixed time period.
5. The number of deaths in a given period of

time of the policyholders of a life insurance company.

Assume that for some positive constant λ the following assumptions hold true:

1. The probability that exactly 1 event occurs in a given interval of length h is equal to $\lambda h + o(h)$, where $o(h)$ stands for any function $f(h)$ that is such that $\lim_{h \to 0} f(h)/h = 0$.
2. The probability that 2 or more events occur in an interval of length h is equal to $o(h)$.
3. For any integers $n, j_1, j_2, \ldots, j_n$, and any set of n nonoverlapping intervals, if we define E_i to be the event that exactly j_i of the events under consideration occur in the ith of these intervals, then events $E_1, E_2, \ldots, E_n$ are independent.

$N(t) \sim P(\lambda)$: The number of events occurs in $(0, t]$.

Example 4.8d. Suppose that earthquakes occur in the western portion of the United States in accordance with assumptions 1, 2, and 3 with $\lambda = 2$ and with 1 week as the unit of time. (That is, earthquakes occur in accordance with the three assumptions at a rate of 2 per week.)

(a) Find the probability that at least 3 earthquakes occur during the next 2 weeks.

(b) Find the probability distribution of the time, staring from now, until the next earthquake.

(a)

$$\begin{aligned} P\{N(2) \geq 3\} &= 1 - P\{N(2) = 0\} - P\{N(2) = 1\} \\ &\quad - P\{N(2) = 2\} \\ &= 1 - e^{-4} + 4e^{-4} - \frac{4^2}{2}e^{-4} \\ &= 1 - 13e^{-4} \end{aligned}$$

(b) – X: Denote the amount of time (in weeks) until the next earthquake.

– $P\{X > t\} = P\{N(t) = 0\} = e^{-\lambda t}$

– $F(t) = P\{X \leq t\} = 1 - P\{X > t\} = 1 - e^{-\lambda t} = 1 - e^{2t}$

4.8.1 Computing the Poisson distribution function

- X is Poisson with parameter λ,

$$\frac{P\{X=i+1\}}{P\{X=i\}} = \frac{e^{-\lambda}\lambda^{i+1}/(i+1)!}{e^{-\lambda}\lambda^{i}/i!} = \frac{\lambda}{i+1}$$

$$\begin{aligned} P\{X=0\} &= e^{-\lambda} \\ P\{X=1\} &= \lambda P\{X=0\} \\ P\{X=2\} &= \frac{\lambda}{2}P\{X=1\} \\ &\vdots \\ P\{X=i+1\} &= \frac{\lambda}{i+1}P\{X=i\} \end{aligned}$$

Example 4.8e.

(a) Determine $P\{X \le 100\}$ when X is Poisson with mean 90.

(b) Determine $P\{Y \le 1075\}$ when Y is Poisson with mean 1000.

- From the text diskette we obtain the solution

(a) $P\{X \leq 100\} \approx .1714$;

(b) $P\{Y \leq 1075\} \approx .9894$.

4.9 Other discrete probability distributions

4.9.1 The geometric random variable

Geometric distribution: $G(p)$

$$P\{X = n\} = (1-p)^{n-1}p \quad n = 1, 2, \ldots$$

- Suppose that independent trials, each having a probability p, $0 < p < 1$, of being a success, are performed until a success occurs.
- $X \sim G(p)$: Number of trials required.

Example 4.9a. An urn contains N white and M black balls. Balls are randomly selected, one at a time, until a block one is obtained. If we assume that each selected ball is replaced before the next one is drawn, what is the probability that

(a) exactly n draws are needed;

(b) at least k draws are needed?

- Let X denote the number of draws needed to select a black ball, $p = \frac{M}{M+N}$.

(a) $P\{X = n\} = \left(\frac{N}{M+N}\right)^{n-1} \frac{M}{M+N} = \frac{MN^{n-1}}{(M+N)^n}$

(b)

$$\begin{aligned}
P\{X \geq k\} &= \frac{M}{M+N} \sum_{n=k}^{\infty} \left(\frac{N}{M+N}\right)^{n-1} \\
&= \left(\frac{M}{M+N}\right) \left(\frac{N}{M+N}\right)^{k-1} / \left[1 - \frac{N}{M+N}\right] \\
&= \left(\frac{N}{M+N}\right)^{k-1} \\
&= (1-p)^{k-1}
\end{aligned}$$

Example 4.9b. Find the expected value of a geometric random variable.

- $q = 1 - p$,

$$
\begin{aligned}
E[X] &= \sum_{n=1}^{\infty} nq^{n-1}p \\
&= p \sum_{n=0}^{\infty} \frac{d}{dq}(q^n) \\
&= p\frac{d}{dq}\left(\sum_{n=0}^{\infty} q^n\right) \\
&= p\frac{d}{dq}\left(\frac{1}{1-q}\right) \\
&= \frac{p}{(1-q)^2} \\
&= \frac{1}{p}
\end{aligned}
$$

Example 4.9c. Find the variance of a geometric random variable.

-

$$
\begin{aligned}
E[X^2] &= \sum_{n=1}^{\infty} n^2 q^{n-1}p \\
&= p \sum_{n=1}^{\infty} \frac{d}{dq}(nq^n) \\
&= p\frac{d}{dq}\left(\sum_{n=1}^{\infty} nq^n\right)
\end{aligned}
$$

$$= p\frac{d}{dq}\left(\frac{q}{1-q}E[X]\right)$$
$$= p\frac{d}{dq}[q(1-q)^{-2}]$$
$$= p\left[\frac{1}{p^2} + \frac{2(1-p)}{p^3}\right]$$
$$= \frac{2}{p^2} - \frac{1}{p}$$

- Since $E[X] = 1/p$,
$$\text{Var}(X) = \frac{1-p}{p^2}$$

4.9.2 The negative binomial random variable

- Suppose that independent trials, each having probability p, $0 < p < 1$, of being a success are performed until a total of r successes is accumulated.
- X: Number of trials required, then
$$P\{X = n\} = \binom{n-1}{r-1}p^r(1-p)^{n-r} \quad n = r, r+1, \ldots$$

X is said to be a *negative binomial* random variable with parameter (r, p).

- Y_1: The number of trials required for the first success.
- Y_2: The number of additional trials after the first success until the second success.
- $X = Y_1 + Y_2 + \cdots + Y_r$ where Y_i's are independently and identically distributed as $G(p)$.

Example 4.9d. If independent trials, each resulting in a success with probability p, are performed, what is the probability of r successes occurring before m failures?

- The solution will be arrived at by noting that r successes will occur before m failures if and only if the rth successes occurs no later than the $r + m - 1$ trial.

- The desired probability is

$$\sum_{n=r}^{r+m-1} \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

Example 4.9e. The Banach match problem. A pipe-smoking mathematician carries, at all times, 2 matchboxes, 1 in his left-hand pocket and 1 in his right-hand pocket. Each time he needs a match he is equally likely to take it from either pocket. Consider the moment when the mathematician first discovers that one of this matchboxes is empty. If it is assumed that both matchboxes initially contained N matches, what is the probability that there are exactly k matches in the other box, $k = 0, 1, \ldots, N$?

- E: The event that the mathematician first discovers that the right-hand matchbox is empty and there are k matches in the left-hand box at the time.
- $P(E) = \binom{2N-k}{N} \left(\frac{1}{2}\right)^{2N-k+1}$

- The desired result is

$$2P(E) = \binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k}$$

Example 4.9f. Compute the expected value and the variance of a negative binomial random variable with parameters r and p.

-

$$\begin{aligned}
E[X^k] &= \sum_{n=r}^{\infty} n^k \binom{n-1}{r-1} p^r (1-p)^{n-r} \\
&= \frac{r}{p} \sum_{n=r}^{\infty} n^{k-1} \binom{n}{r} p^{r+1} (1-p)^{n-r} \\
&= \frac{r}{p} \sum_{m=r+1}^{\infty} (m-1)^{k-1} \binom{m-1}{r} p^{r+1} (1-p)^{m-(r+1)} \\
&= \frac{r}{p} E[(Y-1)^{k-1}]
\end{aligned}$$

where Y is a negative binomial random variable with parameters $r+1$, p.

- $k = 1$, $E[X] = \frac{r}{p}$
- $k = 2$,

$$\begin{aligned}
E[X^2] &= \frac{r}{p} E[Y-1] \\
&= \frac{r}{p}\left(\frac{r+1}{p} - 1\right)
\end{aligned}$$

- $$\begin{aligned}\text{Var}(X) &= \frac{r}{p}\left(\frac{r+1}{p} - 1\right) - \left(\frac{r}{p}\right)^2 \\ &= \frac{r(1-p)}{p^2}\end{aligned}$$
- If independent trials, each of which is a success with probability p, are performed, then the expected value and variance of the number of trials that it takes to amass r successes is r/p and $r(1-p)/p^2$.
- For $G(p)$, $r = 1$.

Example 4.9g. Find the expected value and the variance of the number of times one must throw a die until the outcome 1 has occurred 4 times.

- $X \sim NB(r, p)$
- $r = 4$ and $p = \frac{1}{6}$,

$$E[X] = 24$$

$$\text{Var}(X) = \frac{4(\frac{5}{6})}{(\frac{1}{6})^2} = 120$$

4.9.3 The hypergeometric random variable

- Suppose that a sample of size n is to be chosen randomly (without replacement) from an urn containing N balls, of which m are white and $N - m$ are block.
- Let X denote the number of white balls selected, then

$$P\{X = i\} = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \quad i = 0, 1, \ldots, n \quad (9.4)$$

- A random variable X, whose probability mass function is given by Eq. (9.4) or some values of n, N, m is said to be a *hypergeometric* random variable.

Example 4.9h. An unknown number, say N, of animals inhabit a certain region. To obtain some information about the population size, ecologists often perform the following experiment: They first catch a number, say m, of these animals, mark them in some manner, and release them. After allowing the marked animals time to disperse throughout the region, a new catch of size, say n, is made. Let X denote the number of marked animals in this second capture. If we assume that the population of animals in the region remained fixed between the time of the two catches and that each time an animal was caught it was equally likely to be any of the remaining uncaught animals, it follows that is a hypergeometric random variable such that

$$P\{X = i\} = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \equiv P_i(N)$$

- Suppose now that X is observed to equal i.
- $P_i(N)$ represents the probability of the observed event when there are actually N an-

imals present in the region, it would appear that a reasonable estimate of N would be the value of N that maximizes $P_i(N)$. Such an estimate is called a *maximum likelihood* estimate.

- The maximization of $P_i(N)$ can most simply be done by first nothing that

$$\frac{P_i(N)}{P_i(N-1)} = \frac{(N-m)(N-n)}{N(N-m-n+i)}$$

the above ratio is greater than 1 if and only if

$$(N-m)(N-n) \geq N(N-m-n+i)$$

or, equivalently, if and only if

$$N \leq \frac{mn}{i}$$

- $P_i(N)$ is first increasing, and then decreasing, and reaches its maximum value at the largest integral value not exceeding mn/i. This value is thus the maximum likelihood estimate of N.

- Suppose that the initial catch consists of $m = 50$ animals of which are marked and then released.
- If a subsequent catch consists of $n = 40$ animals of which $i = 4$ are marked, then we would estimate that there are some 500 animals in the region.

Example 4.9i. A purchaser of electrical components buys them in lots of size 10. It is his policy to inspect 3 components randomly from a lot and to accept the lot only if all 3 are nondefective. If 30 percent of the lots have 4 defective components and 70 percent have only 1, what proportion of lots does the purchaser reject?

- A: The event that the purchaser accepts a lot.
- $P(A) = P(A|\text{lot has 4 defectives})\frac{3}{10} + P(A|$ $\text{lot has 1 defective})\frac{7}{10}$

$$= \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}}\left(\frac{3}{10}\right) + \frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}}\left(\frac{7}{10}\right)$$
$$= \frac{54}{100}$$

If n balls are randomly chosen without replacement from a set of N balls, of which the fraction $p = m/N$ is white, then the number of white balls selected is hypergeometric.

It would seem that when m and N are large in relation to n, it shouldn't make much difference whether the selection is being done with or without replacement.

$$P\{X = i\} = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \approx \binom{n}{i} p^i (1-p)^{n-i}$$

when $p = m/N$ and m and N are large in relation to n and i.

Example 4.9j. Determine the expected value and the variance of X, a hypergeometric random variable with parameters n, N, m.

- $$E[X^k] = \sum_{i=0}^{n} i^k P\{X = i\} = \sum_{i=1}^{n} i^k \binom{m}{i}\binom{N-m}{n-i} \Big/ \binom{N}{n}$$
- $i\binom{m}{i} = m\binom{m-1}{i-1}$ and $n\binom{N}{n} = N\binom{N-1}{n-1}$
- $$\begin{aligned} E[X^k] &= \frac{nm}{N} \sum_{i=1}^{n} i^{k-1} \binom{m-1}{i-1}\binom{N-m}{n-i} \Big/ \binom{N-1}{n-1} \\ &= \frac{nm}{N} \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{m-1}{j}\binom{N-m}{n-1-j} \Big/ \binom{N-1}{n-1} \\ &= \frac{nm}{N} E[(Y+1)^{k-1}] \end{aligned}$$

  where Y is a hypergeometric random variable with parameters $n-1, N-1, m-1$.
- $k = 1$, $E[X] = \frac{nm}{N}$
- $k = 2$,

$$\begin{aligned} E[X^2] &= \frac{nm}{N} E[Y+1] \\ &= \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1\right] \end{aligned}$$

- As $E[X] = nm/N$ we can conclude that
$$\text{Var}(X) = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right]$$
- If $p = m/N$ denote the fraction of balls that are white, then
$$\text{Var}(X) = \frac{N-n}{N-1}np(1-p)$$

Remark

- We shown in Example 4.9j that if n balls are randomly selected without replacement from a set of N balls, of which the fraction p are white, then the expected number of white balls chosen is np.
- If N is large in relation to n, then $\text{Var}(X) \approx np(1-p)$.

4.9.4 The Zeta (or Zipf) distribution

- A random variable is said to have a zeta (sometimes called the Zipf) distribution if its probability mass function is given by

$$P\{X = k\} = \frac{C}{k^{\alpha+1}} \quad k = 1, 2, \ldots$$

for some value of $\alpha > 0$.

- $C = \left[\sum_{k=1}^{\infty} \left(\frac{1}{k} \right)^{\alpha+1} \right]^{-1}$
- The zeta distribution owes its name to the fact that the function

$$\zeta(s) = 1 + \left(\frac{1}{2}\right)^s + \left(\frac{1}{3}\right)^s + \cdots + \left(\frac{1}{k}\right)^s + \cdots$$

  is known in mathematical disciplines as the Riemann zeta function.
- The zeta distribution was used by the Italian economist Pareto to describe the distribution of family incomes in a given country.
- It was G. K. Zipf who applied these distributions in a wide variety of different areas and popularized their use.

Summary

- Random variable: A real-valued function defined on the outcome of a probability experiment.
- Distribution function:

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

  All probabilities concerning X can be stated in terms of F.
- Probability mass function: Discrete random variable

$$p(x) = P\{X = x\}$$

- Expected value:

$$E[X] = \Sigma\, xp(x)$$

- Variance:

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

- Standard deviation: $\sqrt{\text{Var}(X)}$
- $B(n, p)$: $p(i) = \binom{n}{i} p^i (1-p)^{n-i}$

$$E[X] = np \qquad \text{Var}(X) = np(1-p)$$

- $P(\lambda)$: $p(i) = \frac{e^{-\lambda}\lambda^i}{i!}$

$$E[X] = \lambda \qquad \mathrm{Var}(X) = \lambda$$

- $G(p)$: $p(i) = p(1-p)^{i-1}$

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

- $NB(r,p)$: $p(i) = \binom{i-1}{r-1} p^r (1-p)^{i-r}$

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

- $HG(n,N,m)$: $p(i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$

$$E[X] = np \qquad \mathrm{Var}(X) = \frac{N-n}{N-1} np(1-p)$$

with $p = m/N$.