Section 4.1 Random Variables

Oftentimes, we are not interested in the specific outcome of an experiment. Instead, we are interested in a *function of the outcome*.

EXAMPLE: Consider rolling a fair die twice.

$$S = \{(i, j) : i, j \in \{1, \dots, 6\}\}$$

Suppose we are interested in computing the sum, i.e. we have placed a bet at a craps table. Let X be the sum. Then $X \in \{2, 3, \dots, 12\}$ is random as it depends upon the outcome of the experiment. It is a *random variable*. We can compute probabilities associated with X.

$$P(X = 2) = P\{(1, 1)\} = 1/36$$

$$P(X = 3) = P\{(1, 2), (2, 1)\} = 2/36$$

$$P(X = 4) = P\{(1, 3), (2, 2), (1, 3)\} = 3/36$$

Can write succinctly

Sum, i	2	3	4	5	6	...	12
$P(X = i)$	1/36	2/36	3/36	4/36	5/36	...	1/36

DEFINITION: Let S be a sample space. Then a function $X : S \to \mathbb{R}$ is a <u>*random variable*</u>.

EXAMPLE: Consider a bin with 5 white and 4 red chips. Let S be the sample space associated with selecting three chips, *with replacement*, from the bin. Let X be the number of white chips chosen. The sample space is

$$S = \{(a, b, c) : a, b, c \in \{W, R\}\}.$$

We have

$$X(R, R, R) = 0$$

$$X(W, R, R) = X(R, W, R) = X(R, R, W) = 1$$

$$X(W, W, R) = X(W, R, W) = X(R, W, W) = 2$$

$$X(W, W, W) = 3$$

We now find probabilities:

$$P(X = 0) = P(R, R, R) = \left(\frac{4}{9}\right)^3$$

$$P(X = 1) = P((W, R, R) \cup (R, W, R) \cup (R, R, W))$$

$$= P(W, R, R) + P(R, W, R) + P(R, R, W) = 3 \cdot \frac{5}{9} \cdot \left(\frac{4}{9}\right)^2$$

$$P(X = 2) = 3 \cdot \left(\frac{5}{9}\right)^2 \cdot \frac{4}{9}$$

$$P(X = 3) = \left(\frac{5}{9}\right)^3$$

Note that

$$\left(\frac{4}{9}\right)^3 + 3 \cdot \frac{5}{9} \cdot \left(\frac{4}{9}\right)^2 + 3 \cdot \left(\frac{5}{9}\right)^2 \cdot \frac{4}{9} + \left(\frac{5}{9}\right)^3 = 1$$

What if *no replacement*? Then,

$$P(X = 0) = \frac{\binom{4}{3}}{\binom{9}{3}}, \qquad P(X = 1) = \frac{\binom{5}{1}\binom{4}{2}}{\binom{9}{3}}$$

$$P(X = 2) = \frac{\binom{5}{2}\binom{4}{1}}{\binom{9}{3}}, \qquad P(X = 3) = \frac{\binom{5}{3}}{\binom{9}{3}}$$

Again, we can show that this sums to one.

REMARK: Note that if X and Y are random variables on the sample space S, then so are $X + Y$, $X - Y$, $aX + bY$ (for a, $b \in \mathbb{R}$), XY, X/Y (so long as $Y \neq 0$). Also, if $f : \mathbb{R} \to \mathbb{R}$, then $f(X)$ is also a random variable. So, X^2, X^3, $\sin(X)$, etc. are all random variables. Also have mix of two ideas: $X^2 + Y^2$, $X^2 + Y - 3Z^4$, etc.

EXAMPLE: Consider choosing a cube with side-length chosen randomly from the interval $(0, 1)$. Let X be the side-length of the cube chosen. Let $Y = X^2$ be the side surface area of the cube chosen, and let $Z = X^3$ be the volume of the cube chosen.

By assumption, $P(X > 1/2) = 1/2$. What is $P(Y > 1/4)$ and $P(Z > 1/8)$? We have

$$P(Y > 1/4) = P(X^2 > 1/4) = P(X > 1/2) = 1/2$$

$$P(Z > 1/8) = P(X^3 > 1/8) = P(X > 1/2) = 1/2$$

Also,

$$P(Z < 1/2) = P(X^3 < 1/2) = P(X < 0.7937) = 0.7937$$

Typically we want the probabilities associated with a random variable, which can be captured with the distribution function of X.

DEFINITION: If X is a random variable, then the function F_X, or F, defined on $(-\infty, \infty)$ by $F_X(t) = F(t) = P(X \le t)$ is called the <u>*distribution function*</u> of X. Sometimes called <u>*cumulative distribution function*</u>.

Note that if $a \le b$, then $\{X \le a\} \subset \{X \le b\}$. Therefore, for $a \le b$ we have

$$F(a) = P\{X \le a\} \le P\{X \le b\} = F(b)$$

and we see that the distribution function of a random variable is non-decreasing. We will see other properties of F in Section 4.10.

EXAMPLES:

1. Consider rolling a die. Let $S = \{1, 2, 3, 4, 5, 6\}$. Let $X : S \to \mathbb{R}$ be given by $X(s) = s^2$. Then, $P(X = i) = 1/6$ for each $i \in \{1, 4, 9, 16, 25, 36\}$. Note, however, that F is defined for all $t \in \mathbb{R}$. So function is constant between numbers and

$$F(1) = P(X \leq 1) = P(X = 1) = 1/6$$

$$F(4) = P(X \leq 4) = P(\ X = 1 \ or\ X = 4) \ = P(X = 1) + P(X = 4) = 2/6, \text{ etc.}$$

2. Reconsider cube example. We have $P(X \leq t) = t$ for all $t \in (0, 1)$, so

$$F_X(t) = \begin{cases} 0, & t \leq 0 \\ t, & 0 < t \leq 1 \\ 1, & 1 < t \end{cases}$$

For the side surface area, $Y = X^2$ and so $P(Y \leq t) = P(X^2 \leq t)$. So,

$$F_Y(t) = \begin{cases} 0, & t \leq 0 \\ \sqrt{t}, & 0 < t \leq 1 \\ 1, & 1 < t \end{cases}$$

Similarly for volume.

Section 4.2 Discrete Random Variables

Consider a random variable $X : S \to \mathbb{R}$. Let $\mathcal{R}(X)$ be the <u>*range*</u> of X. If $\mathcal{R}(X)$ is finite or countably infinite, then X is said to be <u>*discrete*</u>.

For a discrete random variable X, we define the <u>*probability mass function*</u> $p(a)$ or $p_X(a)$ of X by

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

The probability mass function $p(a)$ is positive for at most a countable number of values of a. That is, if $\mathcal{R}(X) = \{x_1, x_2, \ldots\}$ is the range of X, then

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

$$p(x) = 0 \text{ for all other values of } x$$

Because X must take one of those values we also see that

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

Often helpful to present probability mass functions graphically.

EXAMPLES:

1. Consider rolling a fair die. Let X be a value rolled. Then $\mathcal{R}(X) = \{1, 2, 3, 4, 5, 6\}$, and for each $x_i \in \mathcal{R}(X)$, $p(x_i) = P\{X = x_i\} = 1/6$. Plot as vertical bars.

2. Suppose we are gambling on a game in which we win each round with a probability of 0.25. We will play until we win one round. Let X be the number of rounds we play. What is the probability mass function for X?

Solution: We see that

$$p(1) = P\{X = 1\} = P\{\text{win first}\} = 0.25, \quad p(2) = P\{\text{first lose, then win}\} = 0.75 \cdot 0.25$$

In general, for any integer $k \geq 1$, we have

$$p(k) = P\{k - 1 \text{ losses, then 1 win}\} = 0.75^{k-1} \cdot 0.25$$

Note that (by geometric series)

$$\sum_{k=1}^{\infty} p(k) = \sum_{k=1}^{\infty} 0.75^{k-1} \cdot 0.25 = 0.25 \sum_{k=0}^{\infty} 0.75^{k} = 0.25 \cdot \frac{1}{1 - 0.75} = 1$$

This is an example of a *geometric* random variable.

3. Let X be a discrete random variable with probability mass function

$$p(i) = \begin{cases} \dfrac{c\lambda^i}{i!}, & \text{if } i \in \{0, 1, 2, 3, \dots\} \\ 0, & \text{else} \end{cases}$$

This is an example of a *Poisson* random variable with parameter λ. For this to be a probability mass function, what must c be?

We must have $\sum_{i=0}^{\infty} p(i) = 1$. Thus,

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

Thus, $c = e^{-\lambda}$. Thus, for example

(i) $P\{X = 0\} = e^{-\lambda}$

(ii) $P\{X \geq 2\} = 1 - P\{X \leq 1\} = 1 - P\{X = 0\} - P\{X = 1\} = 1 - e^{-\lambda} - \lambda e^{-\lambda}$

Note connection with distribution functions. Discrete random variable if and only if distribution function has only jumps. So it is a step function. Note that

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

4. Consider a random variable X with probability mass function

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

Its distribution function is

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

Section 4.3 Expected Value

Consider a random variable X with range $\{x_1, x_2, \ldots\}$. Imagine each x_i as a possible winnings (in dollars) from the occurrence of a game of chance (X represents the outcome of one such game). Suppose that the probability mass function for X is given by

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

Suppose you play this game many times. Intuitively, how much do you expect to have won? What, then, is your average winnings per turn? Said another way, how much do you *expect* to win in a single occurrence of the game?

Most reasonable people would probably argue as follows: If n is big, after n plays I will win x_i dollars approximately $p(x_i) \cdot n$ times. (This will be shown to be true, in some sense, later in the class. It's called the law of large numbers.) Therefore, my total winnings are approximately

$$\sum_i x_i p(x_i) n$$

Dividing by n gives us the average winnings per game to be

$$\sum_i x_i p(x_i)$$

DEFINITION: Let X be a discrete random variable with probability mass function $p(x)$. Then the *expectation* or the *expected value* of X, denoted by $E[X]$, is

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

EXAMPLES:

1. Let X be the outcome of a roll of a fair die. Find $E[X]$.

Solution: We know that $p(1) = p(2) = \ldots = p(6) = 1/6$, thus

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}$$

This result can be generalized in the following natural way:

2. Consider a random variable taking values in $\{1, \ldots, n\}$ with $P\{X = i\} = 1/n$ for each $i \in \{1, \ldots, n\}$. We say that X is distributed uniformly over $\{1, \ldots, n\}$. What is the expectation?

Solution: We have

$$E[X] = \sum_{i=1}^{n} iP\{X = i\} = \sum_{i=1}^{n} i\frac{1}{n} = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$$

3. Suppose that X takes values on $\{0, 1, 2, \ldots\}$ with probability mass function

$$P\{X = x\} = e^{-\lambda}\frac{\lambda^x}{x!}$$

Then X has a *Poisson distribution* with parameter λ. The expected value is

$$E[X] = \sum_{x=0}^{\infty} xe^{-\lambda}\frac{\lambda^x}{x!} = \lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = \lambda e^{-\lambda}e^{\lambda} = \lambda$$

Note that if $X \geq 0$, i.e. the range only takes nonnegative values, then $E[X] \geq 0$.

4. Consider a Bernoulli random variable X that takes a value of 0 or 1 and the p.m.f. is $p(0) = 1 - p$ and $p(1) = p$ for some $0 < p < 1$. In this case

$$E[X] = 0 \cdot (1 - p) + 1 \cdot p = p$$

5. We say that $I = I_A = 1_A$ (all these notations are used) is an indicator function for the event A if

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

(example: rolling a die and letting $A = \{1, 2, 5\}$, then I_A gives the value 1 to the outcomes are 1, 2, or 5 and zero otherwise). Let's find $E[I_A]$. We know that the range of I_A is simply 1 and 0. We also know that $p(1) = P(A)$ and $p(0) = P(A^c)$. Therefore,

$$E[I_A] = 1 \cdot p(1) + 0 \cdot p(0) = P(A)$$

6. A school class of 120 students is driven somewhere in 3 buses. There are 36, 40, and 44 students on the bus 1, 2, and 3, respectively. Upon arrival, one of the 120 students is chosen randomly and we let X be the number of students on the chosen students' bus. What is $E[X]$?

6. A school class of 120 students is driven somewhere in 3 buses. There are 36, 40, and 44 students on the bus 1, 2, and 3, respectively. Upon arrival, one of the 120 students is chosen randomly and we let X be the number of students on the chosen students' bus. What is $E[X]$?

Solution: We know that all 120 students are equally likely to be chosen. We also know that the range of X is $\{36, 40, 44\}$. Finally, the probability mass function is given by

$$P\{X = 36\} = \frac{36}{120}, \quad P\{X = 40\} = \frac{40}{120}, \quad P\{X = 44\} = \frac{44}{120}$$

Thus,

$$E[X] = 36 \cdot \frac{36}{120} + 40 \cdot \frac{40}{120} + 44 \cdot \frac{44}{120} = 40.2667$$

Obviously, the average number of students per bus is $120/3 = 40$, showing the expected number of students on a randomly chosen students' bus is larger than the average. This is a general property. Basically, buses with many students are weighted heavily. This is easily seen if there were 118 students on one bus and 1 in each of the others. The expected number of students on a randomly chosen students' bus is then easily seen to be around 118.

Section 4.4 Expectation of a Function of a Random Variable

Suppose that we have a discrete random variable, X, and a function g, and suppose that we want to understand the random variable $g(X)$. More to the point, we want to be able to compute the expected value of X. Note that $g(X)$ is itself a random variable, so it has a probability mass function. Once we know it, we can compute the expected value.

EXAMPLE: Let $Y = X^2$, where X is a random variable taking values $-1, 0, 1$ with probabilities

$$P\{X = -1\} = 0.2, \quad P\{X = 0\} = 0.5, \quad P\{X = 1\} = 0.3$$

The range of Y is $\{0, 1\}$. Also,

$$P\{Y = 0\} = P\{X = 0\} = 0.5$$

$$P\{Y = 1\} = P\{X = 1\} + P\{X = -1\} = 0.5$$

Therefore, the expected value can be computed as normal

$$E[X^2] = E[Y] = 0 \cdot P\{Y = 0\} + 1 \cdot P\{Y = 1\} = 1 \cdot \frac{1}{2} = 0.5$$

Note that

$$0.5 = E[X^2] \neq (E[X])^2 = (-1 \cdot 0.2 + 0.3)^2 = 0.01$$

There is a nice way to make this computation.

THEOREM: If X is a discrete random variable with range x_i for $i \geq 1$, with respective probabilities $p(x_i)$, then for any real-valued function g we have

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

Proof: The proof proceeds by grouping together terms of $\sum_i g(x_i)p(x_i)$ that have the same value $g(x_i)$. This is just how we computed the last example! So, let y_j, for $j \geq 1$, represent all possible values of $g(x_i)$. Note that $\{y_j\}$ is just the range of $g(X)$. We have

$$\begin{aligned}
\sum_i g(x_i)p(x_i) &= \sum_j \sum_{i:g(x_i)=y_j} g(x_i)p(x_i) \\
&= \sum_j \sum_{i:g(x_i)=y_j} y_j p(x_i) \\
&= \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i) = \sum_j y_j P\{g(X) = y_j\} \\
&= E[g(X)] \ \blacksquare
\end{aligned}$$

Properties of Expectations:

1. For any $\alpha_1, \alpha_2 \in \mathbb{R}$ we have

$$\boxed{E[\alpha_1 g_1(X) + \alpha_2 g_2(X)] = \alpha_1 E[g_1(X)] + \alpha_2 E[g_2(X)]}$$

In fact,

$$\begin{aligned}
E[\alpha_1 g_1(X) + \alpha_2 g_2(X)] &= \sum_i (\alpha_1 g_1(x_i) + \alpha_2 g_2(x_i))p(x_i) \\
&= \alpha_1 \sum_i g_1(x_i)p(x_i) + \alpha_2 \sum_i g_2(x_i)p(x_i) \\
&= \alpha_1 E[g_1(X)] + \alpha_2 E[g_2(X)]
\end{aligned}$$

In particular, for any $a,\ b \in \mathbb{R}$ we have

$$\boxed{E[aX + b] = aE[X] + b}$$

2. If $X \geq 0$, then $E[X] \geq 0$.

3. If $f \leq g$, then $E[f(X)] \leq E[g(X)]$.

EXAMPLE: Suppose X has $\mathcal{R}(X) = \{-2, 4, 6\}$ and $p(-2) = 1/7,\ p(4) = 2/7$, and $p(6) = 4/7$. What is $E[X(X-2)]$?

Solution: We have

$$E[X(X-2)] = (-2) \cdot (-4) \cdot p(-2) + 4 \cdot 2 \cdot p(4) + 6 \cdot 4 \cdot p(6) = 8 \cdot \frac{1}{7} + 8 \cdot \frac{2}{7} + 24 \cdot \frac{4}{7} = \frac{56}{7} = \frac{120}{7}$$

or

$$\begin{aligned}
E[X(X-2)] &= E[X^2 - 2X] = E[X^2] - 2E[X] \\
&= 4 \cdot \frac{1}{7} + 16 \cdot \frac{2}{7} + 36 \cdot \frac{4}{7} - 2\left(-2 \cdot \frac{1}{7} + 4 \cdot \frac{2}{7} + 6 \cdot \frac{4}{7}\right) = \frac{120}{7}
\end{aligned}$$

Section 4.5 Variance

Variance gives a measure on the "spread" of a random variable around the mean, whereas expected value gave the mean.

Let $\mu = E[X]$ denote the mean of a random variable X. We define the *variance* and *standard deviation* of X to be

$$\boxed{\begin{aligned}&\mathrm{Var}(X) = E[(X-\mu)^2]\\ &\sigma_X = \sqrt{\mathrm{Var}(X)}\end{aligned}}$$

Sometimes more convenient alternative arises from linearity of expectations:

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

Thus,

$$E[X^2] = \mathrm{Var}(X) + \mu^2 \geq \mu^2 = E[X]^2$$

where $E[X^2]$ is called *second moment*. In general, $E[X^n]$ is *nth moment* of the random variable.

EXAMPLE: Consider rolling a fair die. What is the variance of X, the value rolled?

Solution: We know from before that $E[X] = 7/2$. Now we need $E[X^2]$. Again, $\mathcal{R}(X) = \{1,2,3,4,5,6\}$ and $p(i) = 1/6$ for all $i = 1,2,\ldots,6$. Thus,

$$E[X^2] = 1\cdot p(1) + 4\cdot p(2) + \ldots + 36\cdot p(6) = \frac{1}{6}(1+4+9+16+25+36) = \frac{91}{6}$$

Thus,

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$$

THEOREM: For any constants a and b we have

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

Proof: Let $\mu = E[X]$. We have

$$\mathrm{Var}(aX+b) = E[(aX+b-a\mu-b)^2] = E[a^2(X-\mu)^2] = a^2\mathrm{Var}(X) \blacksquare$$

EXAMPLE: Consider again a Poisson random variable with parameter $\lambda > 0$. That is $\mathcal{R}(X) = \{0,1,2,\ldots\}$ and for $k \in \{0,1,2,\ldots\}$,

$$p(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

What is the variance? Recall, we know that $E[X] = \lambda$. We need $E[X^2]$. Therefore,

$$E[X^2] = \sum_{k=0}^{\infty} k^2 e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} k^2\frac{\lambda^k}{k!} = e^{-\lambda}\lambda\sum_{k=1}^{\infty} k\frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\lambda\sum_{k=0}^{\infty}(k+1)\frac{\lambda^k}{k!}$$
$$= \lambda\left(\sum_{k=0}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} + e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\right) = \lambda(E[X]+1) = \lambda^2+\lambda$$

So, Var(X) $= E[X^2] - E[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda = E[X]$.

Section 4.6 The Bernoulli and Binomial Random Variables

Bernoulli random variable: Suppose we have an experiment and care about 2 events: success or failure (s or f). A random variable X is a <u>*Bernoulli*</u> random variable if $X(s) = 1$ for $s \in success$ and $X(f) = 0$ for $f \in failure$. Thus, for some $0 < p < 1$,

$$\boxed{p(1) = P\{X = 1\} = p \quad \text{and} \quad p(0) = P\{X = 0\} = 1 - p = q \quad (p(x) = 0 \text{ otherwise})}$$

Properties:

$$E[X] = 1 \cdot p(1) + 0 \cdot p(0) = p$$

$$E[X^2] = 1^2 \cdot p(1) + 0^2 \cdot p(0) = p$$

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

$$\sigma_X = \sqrt{p(1-p)}$$

EXAMPLE: There are 100 tagged and numbered (1 - 100) rabbits. You set a trap at night that is guaranteed to catch one. You need one that is numbered 37 - 50. Let X be 1 if you catch a desired rabbit and zero otherwise. Then X is Bernoulli with

$$P\{X = 1\} = 14/100, \quad P\{X = 0\} = 86/100$$

Binomial random variable: Now consider n independent repeated trials of a Bernoulli random variable. Let X be the number of successes in the n trials. Clearly,

$$R(X) = \{0, 1, \dots, n\}$$

What is $P\{X = k\}$ for $k \le n$? Each specific sequence of trials that has k successes and $n - k$ failures has a probability of happening of $p^k(1-p)^{n-k}$. How many ways can we get k successes and $n - k$ failures? It's the same question as how many ways can we choose k slots from n. Thus, the answer is $\binom{n}{k}$. Therefore,

$$p(k) = P\{X = k\} = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } k \in \{0, 1, 2, \dots, n\} \\ 0 & \text{otherwise} \end{cases}$$

DEFINITION: Any random variable with probability mass function given above is a *binomial* random variable with parameters $(n,\ p)$.

Note that by binomial theorem (TYPO IN THE BOOK IN THIS EQUATION ON PAGE 135)

$$\sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = (p + (1-p))^n = 1$$

EXAMPLE: There is a product that is defective with a probability of 0.01. They are sold in packages of 10. The company offers money back if two or more are defective. What is the probability that a given package will be returned for cash?

Solution: Let X denote the number of defective products in a given package. We want to know $P\{X \geq 2\}$. Clearly, X is a binomial$(10, 0.01)$ random variable. It is obviously easier to compute $P\{X \geq 2\} = 1 - P\{X \leq 1\}$. Thus, we need

$$P\{X = 0\} = \binom{10}{0} \cdot 0.01^0 \cdot 0.99^{10} = 0.90438, \quad P\{X = 1\} = \binom{10}{1} \cdot 0.01^1 \cdot 0.99^9 = 0.09135$$

Therefore, the desired probability is

$$P\{X \geq 2\} = 1 - P\{X = 0\} - P\{X = 1\} \approx 0.0043$$

Thus, 0.4 percent of the packages are eligible for recall.

Expectations and Variance of Binomial random variable. Let X be binomial$(n,\ p)$. We will compute the moments of X, $E[X^k]$. We have

$$E[X^k] = \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i}$$

Note that the following identity holds

$$i\binom{n}{i} = i \cdot \frac{n!}{i!(n-i)!} = n \cdot \frac{(n-1)!}{(i-1)!(n-i)!} = n\binom{n-1}{i-1}$$

Therefore,

$$E[X^k] = np\sum_{i=1}^{n} i^{k-1}\binom{n-1}{i-1} p^{i-1}(1-p)^{n-i} = np\sum_{j=0}^{n-1}(j+1)^{k-1}\binom{n-1}{j} p^j (1-p)^{n-1-j} = npE[(Y+1)^{k-1}]$$

where Y is a binomial$(n-1, p)$ random variable. Letting $k = 1$, we see that

$$E[X] = np$$

Setting $k = 2$ and using the preceding formula yields

$$E[X^2] = npE[Y+1] = np[(n-1)p + 1] = n^2p^2 - np^2 + np$$

Therefore,

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = n^2p^2 - np^2 + np - n^2p^2 = np - np^2 = np(1-p)$$

Interesting: Note that this is n times the variance of the Bernoulli random variable.

PROPOSITION: If X is a binomial random variable with parameters (n, p), 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 $(n + 1)p$.

Proof: Consider the ratio $P\{X = k\}/P\{X = k - 1\}$ and ask when is this greater than or less than 1. We have

$$\frac{P\{X = k\}}{P\{X = k - 1\}} = \frac{\dfrac{n!}{(n-k)!k!} \cdot p^k (1-p)^{n-k}}{\dfrac{n!}{(n-k+1)!(k-1)!} \cdot p^{k-1}(1-p)^{n-k+1}} = \frac{(n-k+1)p}{k(1-p)}$$

Thus, $P\{X = k\} \geq P\{X = k - 1\}$ if and only if

$$(n-k+1)p \geq k(1-p) \quad \Longrightarrow \quad np - pk + p \geq k - kp \quad \Longrightarrow \quad (n+1)p \geq k$$

Note that we have the following corollary which shows how to compute the probabilities of a binomial(n, p) iteratively:

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

Section 4.7 The Poisson Random Variable

Recall that a random variable X is said to be a *Poisson* random variable with parameter $\lambda > 0$ if for $k \in \{0, 1, 2, \ldots\}$,

$$\boxed{P\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!}}$$

One of the reasons that Poisson random variables are so useful is that they are an approximation to the binomial random variable in the following sense. Consider a binomial(n, p) random variable Y with big n and small p, but $np = \lambda$ moderate. Thus, there are many independent experiments, each with a small probability of occurring, but expected number of occurrences is moderate. Then the probability of k successes is approximately

$$\begin{aligned}
P\{Y = k\} &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{(n-k)!k!}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n(n-1)\ldots(n-(k-1))}{n^k} \cdot \frac{\lambda^k}{k!} \cdot \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^k} \\
&= [\text{take limit as } n \to \infty] \approx \frac{n^k}{n^k} \cdot \frac{\lambda^k}{k!} \cdot \frac{e^{-\lambda}}{1} = e^{-\lambda}\frac{\lambda^k}{k!}
\end{aligned}$$

EXAMPLE: A community of 100,000 people is struck with a fatal disease in which the chance of death is 0.0001. What is the probability that at most 8 people will die?

EXAMPLE: A community of 100,000 people is struck with a fatal disease in which the chance of death is 0.0001. What is the probability that at most 8 people will die?

Solution: This is binomial with parameters $n = 10^5$ and $p = 10^{-4}$. So, if X is the random variable giving number of deaths, then

$$P\{X \leq 8\} = \sum_{k=0}^{8} \binom{10^5}{k} (10^{-4})^k (1 - 10^{-4})^{10^5 - k} = \sum_{k=0}^{8} \frac{10^5!}{k!(10^5 - k)!} (10^{-4})^k (1 - 10^{-4})^{10^5 - k}$$

Computationally this formula is too complicated. Another option is to use the Poisson approximation with $\lambda = np = 10$:

$$P\{X \leq 8\} \approx \sum_{k=0}^{8} e^{-10} \frac{10^k}{k!} = 0.3328197$$

The "Actual answer" computed with Maple is 0.332808. But the Poisson approach was (relatively) instant, whereas the Binomial computation took 43.5 seconds.

Recall, we've already seen that for $k \in \{0, 1, 2, \ldots\}$, $p(k) = e^{-\lambda} \dfrac{\lambda^k}{k!}$ is a probability mass function. We also know that

$$\boxed{E[X] = \lambda, \quad \mathrm{Var}(X) = \lambda}$$

EXAMPLES:

1. Number of misprints on a page.

2. The number of people in a group who survive to age 100.

3. Number of certain natural disasters (earthquake/volcano) happening in a given year.

Another benefit: you only need the average occurrence rate of event, not n or p.

EXAMPLES:

1. Disney World averages 5 injuries per day. What is the probability that there will be more than 6 injuries tomorrow?

1. Disney World averages 5 injuries per day. What is the probability that there will be more than 6 injuries tomorrow?

Solution: It seems like there is not enough information to be able to answer this question. However, we know that $E[X] = np = 5$. We also know that n is big, p is small and $np = \lambda = 5$. Thus, X is approximately Poisson with parameter 5. Therefore,

$$P\{X \geq 7\} = 1 - P\{X < 7\} \approx 1 - \sum_{k=0}^{6} e^{-5}\frac{5^k}{k!} = 0.2378$$

Note that it would be impossible to solve this using the binomial view because

$$P\{X \geq 7\} = 1 - P\{X < 7\} = 1 - \sum_{k=0}^{6} \binom{n}{k} p^k (1-p)^k$$

but we don't know p or n.

2. Suppose that there are, on average, 1.2 typos on a page of a given book. What is the probability that there are more than two errors on a given page?

Solution: We model the number of errors on a given page, X, as a Poisson random variable with parameter 1.2. Then the desired probability is

$$P\{X \geq 3\} = 1 - P\{X \leq 2\} = 1 - e^{-1.2}\sum_{k=0}^{2} \frac{1.2^k}{k!} = 0.120512$$

Not even the assumptions laid our above are strictly necessary. We have the following

Poisson Paradigm: Consider n events, with p_i being the probability that the ith event occurs. If all the p_i are small and the trials are independent or "weakly dependent", then the number of these events that occur approximately has a Poisson distribution with mean $\sum_{i=1}^{n} p_i$.

Poisson Processes

Suppose that starting at time 0 we start counting events (earthquakes, arrivals at a post-office, number of meteors). For each t, we obtain a number $N(t)$ giving the number of events up to time t. $N(t)$ is a random variable. So we have an infinite (uncountable) number of random variables. Let $N(t)$ be a discrete random variable with range $\{0, 1, 2, 3, \ldots\}$. We make the following assumptions:

1. The probability of exactly one event occurring in a given time interval of length h is equal to $\lambda h + o(h)$, where $o(h)$ stands for any function g that satisfies $g(h)/h \to 0$ as $h \to 0$. So, $g(h) = h^2$ is $o(h)$, but $g(h) = h$ is not.

2. The probability that 2 or more events occur in an interval of length h is $o(h)$.

3. For any integers n, $j_1, \ldots, j_n$ 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 the events $E_1, E_2, \ldots, E_n$ are independent.

We now show that under the above assumptions the number of events happening in any length of time t has a Poisson distribution with parameter λt. That is, we will show that

$$P\{N(t) = k\} = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

We begin by breaking the interval $[0,\ t]$ up in to n subintervals of length t/n (n is large). We have

$P\{N(t) = k\} = P\{k$ of the subintervals contain exactly 1 event and other $n-k$ contain no events$\}$ $+P\{N(t) = k$ and at least one of the subintervals contains 2 or more events$\}$

Let A and B be the two events on the right hand side. We start by showing that $P(B) \to 0$ as $n \to \infty$ (so events happen one at a time). We will use Boole's inequality $P(\cup A_i) \leq \sum P(A_i)$:

$$\begin{aligned}
P(B) &\leq P\{\text{at least one subinterval has 2 or more events}\} \\
&= P\left\{\bigcup_{i=1}^{n}\{i\text{th subinterval contains 2 or more}\}\right\} \\
&\leq \sum_{i=1}^{n} P\{i\text{th subinterval contains 2 or more}\} \\
&= \sum_{i=1}^{n} o(t/n) = n \cdot o(t/n) = t\left[\frac{o(t/n)}{t/n}\right]
\end{aligned}$$

Thus, as $n \to 0,\ P(B) \to 0$. Now we just need to handle the set A. We see from assumptions 1 and 2 that

$$P\{\text{ 0 events occur in any interval of length } h\} = 1 - [\lambda h + o(h) + o(h)] = 1 - \lambda h - o(h)$$

Now using assumption 3 (independence of intervals), we see that

$$P(A) = \binom{n}{k}\left[\frac{\lambda t}{n} + o(t/n)\right]^k \left[1 - \left(\frac{\lambda t}{n} + o(t/n)\right)\right]^{n-k}$$

Using the same argument as for the original Poisson approximation, as $n \to \infty$, we find that

$$P(A) \to e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

Therefore, we see that

$$P\{N(t) = k\} = e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \quad k = 0, 1, 2, \ldots$$

So, $N(t)$ is a Poisson random variable with parameter λt. We say that $N(t)$ is a *Poisson process* with rate (or intensity) λ. Moreover, $E[N(t)] = \lambda t$ and so $E[N(1)] = \lambda$. So λ can be reconstructed through observation (wait one unit of time, record, then another, record, etc. using stationarity).

EXAMPLE: Suppose that you are watching people arriving at a doctors office and they are arriving at an average rate of 3 per hour. What is the probability that no one will arrive in the next hour? What is the probability that from 12pm – 5pm there will be 2 full hours (i.e. 2pm – 4pm) that are not necessarily contiguous in which no one will arrive?

Solution: Let $N(t)$ be the number of people to arrive in the next t hours. Then $N(t)$ is Poisson with rate 3. Thus,

$$P\{N(1)=0\}=e^{-3\cdot 1}\frac{(3\cdot 1)^0}{0!}=e^{-3}=0.04978$$

For the second problem, by stationarity and independence of increments, the probability of any hour long block having n visits is the same for each block. Therefore, if X is the number of hour long blocks in which nobody shows up, X is binomial with parameters 5 and $p=e^{-3}$. Therefore,

$$P\{X=2\}=\binom{5}{2}p^2(1-p)^{5-2}=\binom{5}{2}e^{-3\cdot 2}(1-e^{-3})^3=0.02889$$

Computing the Poisson random variable. Note that if X is Poisson with parameter $\lambda>0$, then

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

Therefore,

$$\boxed{P\{X=i+1\}=\frac{\lambda}{i+1}P\{X=i\}}$$

Section 4.8 Other Discrete Random Variables

The geometric random variable:

Suppose that independent trials, each having a probability of p, $0<p<1$, of success are performed until a success occurs. Let X equal the number of trials required. Then the range of X is $\{1,2,\ldots\}$ and the probability mass function is given by

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

Note that by Geometric series we have

$$\sum_{n=1}^{\infty}P(X=n)=\sum_{n=1}^{\infty}(1-p)^{n-1}p=p\sum_{n=1}^{\infty}(1-p)^{n-1}=p\sum_{n=0}^{\infty}(1-p)^{n}=p\cdot\frac{1}{1-(1-p)}=1$$

A random variable with this probability mass function is called a <u>*geometric*</u> random variable.

EXAMPLE: What is the minimum number of draws required to get an Ace with probability 0.95?

EXAMPLE: What is the minimum number of draws required to get an Ace with probability 0.95?

Solution: We want the smallest n such that $P(X \leq n) \geq 0.95$. Therefore, we want

$$1 - P(X > n) \geq 0.95 \quad \text{or} \quad P(X > n) \leq 0.05$$

We have

$$P(X > n) = \sum_{k=n+1}^{\infty} \frac{1}{13}\left(\frac{12}{13}\right)^{k-1} = \frac{1}{13}\left(\frac{12}{13}\right)^{n} \sum_{i=0}^{\infty}\left(\frac{12}{13}\right)^{i} = \left(\frac{12}{13}\right)^{n}$$

So we need $(12/13)^n \leq 0.05$ or $n \ln(12/13) \leq \ln(0.05)$ or $n \geq \ln(0.05)/\ln(12/13) = 37.42$. Thus, n is 38.

Expected Value and Variance: We can compute the expected value of a geometric random variable. One proof using summations is in the book. Here we show another way. Set $q = 1-p$. Then,

$$E(X) = \sum_{i=1}^{\infty} iq^{i-1}p = p\sum_{i=1}^{\infty} iq^{i-1}$$

The radius of convergence of $\sum_{i=0}^{\infty} q^i = 1/(1-q)$ is $|q| < 1$. Thus, you can differentiate termwise to see that

$$\frac{1}{(1-q)^2} = \sum_{i=1}^{\infty} iq^{i-1}$$

Therefore

$$E(X) = p\sum_{i=1}^{\infty} iq^{i-1} = \frac{p}{p^2} = \frac{1}{p} \quad \Longrightarrow \quad \boxed{E(X) = \frac{1}{p}}$$

We can also use this method to compute the variance. Taking a second derivative yields

$$\begin{aligned} \frac{2}{(1-q)^3} = \sum_{i=2}^{\infty} i(i-1)q^{i-2} = \sum_{i=1}^{\infty}(i+1)iq^{i-1} = \sum_{i=1}^{\infty} i^2q^{i-1} + \sum_{i=1}^{\infty} iq^{i-1} &= \frac{1}{p}E[X^2] + \frac{1}{p}E[X] \\ &= \frac{1}{p}E[X^2] + \frac{1}{p^2} \end{aligned}$$

So,

$$\frac{2}{p^3} = \frac{1}{p}E[X^2] + \frac{1}{p^2} \quad \Longrightarrow \quad E[X^2] = \frac{2}{p^2} - \frac{1}{p} \quad \Longrightarrow \quad \text{Var}(X) = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \boxed{\frac{1-p}{p^2}}$$

Memoryless Property: Let X be Geometric. Then for all $n,\ m \geq 1$ we have

$$\begin{aligned} P(X > n+m \mid X > m) &= \frac{P(X > n+m)}{P(X > m)} = \frac{n+m \text{ failures to start}}{m \text{ failures to start}} \\ &= \frac{(1-p)^{n+m}}{(1-p)^m} = (1-p)^n = P(X > n) \end{aligned}$$

So probability that next n will be failures given first m were is same as probability that first n are failures. This is intuitive by independence. However, Geometric is the only discrete random variable with this property.

Negative Binomial:

Generalized geometric.

- Consider a sequence of Bernoulli trials with probability of success $= p$.
- Let X be the number of trials until the rth success occurs.

Then X is *Negative Binomial* with parameters $r, \ p$.

- Note that NegBin$(1, p)$=Geom(p).

The range is $R(X) = \{r, r+1, r+2, \ldots\}$ and for $n \in R(X)$ (using independence)

$$\begin{aligned}
P(X = n) &= P(r-1 \text{ successes in } n-1 \text{ trials and then success on } n\text{th}) \\
&= P(r-1 \text{ successes in } n-1 \text{ trials}) P(\text{success}) \\
&= P(\text{binom}(n-1, p) = r-1)p \\
&= \binom{n-1}{r-1} p^{r-1}(1-p)^{n-1-(r-1)}p \\
&= \binom{n-1}{r-1} p^{r}(1-p)^{n-r}
\end{aligned}$$

We have

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

EXAMPLE: The Red Sox and Brewers are playing for the World Series. The probability that the Red Sox win each game is 0.55. What is the probability that the Brewers win in 6?

EXAMPLE: The Red Sox and Brewers are playing for the World Series. The probability that the Red Sox win each game is 0.55. What is the probability that the Brewers win in 6?

Solution: Let X be the number of games needed until Brewers win 4th. This is NegBin$(r = 4, p = 0.45)$. Therefore,

$$P(X = 6) = \binom{5}{3} \cdot 0.45^4 \cdot 0.55^{6-4} = 0.12404$$

Note also, that

$$P(X = 4) = \binom{3}{3} \cdot 0.45^4 \cdot 0.55^0 = 0.041006$$

$$P(X = 5) = \binom{4}{3} \cdot 0.45^4 \cdot 0.55^1 = 0.090214$$

$$P(X = 7) = \binom{6}{3} \cdot 0.45^4 \cdot 0.55^3 = 0.136448$$

Therefore, the probability that the Brewers win the series is

$$\begin{aligned} P(\text{Brewers Win}) &= P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) \\ &= 0.041006 + 0.090214 + 0.12404 + 0.136448 \approx 0.3917 \end{aligned}$$

One can check the following

p	P(Brewers Win)
0.4	0.28979
0.35	0.199845
0.3	0.126

Hypergeometric random variable:

- Suppose we have an urn with N balls of which m are white and $N - m$ are black.
- Now we choose a sample size of n from the N balls.
- Let X denote the number of white balls selected.

Then,

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

Note that we must have

$$i \leq \min\{m,\ n\}$$

else the probability is zero (which still holds for our formula). Also, as there are a total of $N - m$ black balls, we must also have

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

However, our formula is still valid for if we didn't have that, the second term is zero.

Any random variable with this probability mass function for some m, N, n is said to be a *hypergeometric* random variable. Letting $p = m/N$, we have

$$E[X] = \frac{nm}{N} \left(= \frac{m}{N} n = pn \right)$$
$$\text{Var}(X) = \frac{nm(N-m)}{N^2} \left(1 - \frac{n-1}{N-1}\right) \left(= np(1-p) \left(1 - \frac{n-1}{N-1}\right)\right)$$

EXAMPLE (Maximum likelihood estimate): Suppose we want to estimate the number of rabbits in a region. How can we do so?

1. Catch m animals and mark them. Then release.

2. At a later date, catch n animals. Now let X denote the number of marked animals in this second capture.

3. Assume the population remained fixed between two captures and each animal is equally likely to be caught.

4. Clearly, X is hypergeometric with parameters N, m, n. Thus,

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

5. Now we assume that that X is observed to be i. What is our best guess at N?

6. The idea is to find the value of N that maximizes $P_i(N)$. Note that

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

Therefore, the ratio is greater than one if and only if (after reducing)

$$N \le \frac{nm}{i}$$

Thus, $P_i(N)$ first increasing and then decreasing, and reaches its maximum value at the largest integer less than or equal to nm/i. This value is the maximum likelihood estimate of N.

Does this make sense? Think of this way. The proportion of marked animals is m/N. Assuming that i/n is of the same proportion and then solving yields

$$\frac{i}{n} = \frac{m}{N} \quad \Longrightarrow \quad N \approx \frac{mn}{i}$$

Section 4.9 Expected Value of Sums of Random Variables

In this section, we will show, in discrete case, that the expected value of a sum of random variables is the sum of the expected values. (this is true in the continuous case also, we'll show that later.)

Consider a random variable X with sample space S that is finite or countably infinite. We will write the expected value in a different way:

1. For $s \in S$, denote by $X(s)$ the value that X takes at s.

2. For $s \in S$, denote $p(s) = P(\{s\})$.

3. Note that for two random variables X, Y we can let $Z = X + Y$ by defining $Z(s) = X(s) + Y(s)$.

For example, if an experiment consists of flipping a coin five times and X gave the number of heads in the first three flips and Y gives the number of heads in the final four, then the outcome $s = (h,\ t,\ h,\ t,\ h)$ has

$$\begin{aligned} X(s) &= 2 \\ Y(s) &= 2 \\ Z(s) &= X(s) + Y(s) = 4 \end{aligned}$$

THEOREM: We have

$$\boxed{E[X] = \sum_{s \in S} X(s)p(s)}$$

Proof: Let the range of X be denoted by $\{x_i\}$ for $i \geq 1$. Also, for each such x_i let

$$S_i = \{s \in S \mid X(s) = x_i\}$$

denote the event that X equals x_i (note that these partition S). We have

$$\begin{aligned} E[X] &= \sum_{i=1}^{\infty} x_i P\{X = x_i\} \\ &= \sum_i x_i P(S_i) \\ &= \sum_i x_i \sum_{s \in S_i} p(s) \\ &= \sum_i \sum_{s \in S_i} x_i p(s) \\ &= \sum_i \sum_{s \in S_i} X(s)p(s) \\ &= \sum_{s \in S} X(s)p(s) \ \blacksquare \end{aligned}$$

COROLLARY: For random variables $X_1, X_2, \ldots, X_n$ we have

$$\boxed{E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]}$$

Proof: Let $Z = \sum_{i=1}^{n} X_i$. Then,

$$\begin{aligned} E[Z] = \sum_{s\in S} Z(s)p(s) &= \sum_{s\in S}(X_1(s) + \ldots + X_n(s))p(s) \\ &= \sum_{s\in S} X_1(s)p(s) + \ldots + \sum_{s\in S} X_n(s)p(s) \\ &= E[X_1] + \ldots + E[X_n] \blacksquare \end{aligned}$$

EXAMPLES:

1. Consider n trials where the ith has probability p_i of success. What is the expected number of successes $E[X]$?

Solution: Let

$$X_i = \begin{cases} 1, & \text{if trial } i \text{ is a success} \\ 0, & \text{if trial } i \text{ is a failure} \end{cases}$$

Then $X = \sum_i X_i$ and so

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p_i$$

Note that independence was not required here. As a special case, $p_i \equiv p$, we get the mean of the binomial. However, we also get the mean of the hypergeometric. Note that n balls are chosen and the probability that the ith is white is always m/N. That is, if Z_i is 1 when a white ball is chosen on ith choice and zero otherwise, and if X is a total number of white balls chosen, then

$$X = \sum_{i=1}^{n} Z_i$$

and

$$E[X] = \sum_{i=1}^{n} E[Z_i] = \sum_{i=1}^{n} P\{Z_i = 1\} = np = n(m/N)$$

Even though hypergeometric trials are dependent, the result holds.

2. 20 people working in an office have a holiday gift exchange. Suppose that there is no organization of the exchange and each person buys a gift for one of the other 19 selected at random without knowledge of who the others select. What is the expected number of people who receive no gift?

2. 20 people working in an office have a holiday gift exchange. Suppose that there is no organization of the exchange and each person buys a gift for one of the other 19 selected at random without knowledge of who the others select. What is the expected number of people who receive no gift?

Solution: For person i, let Z_i be one if this person does not receive a gift. Then, if we let N be the number of people who receive no gift, $N = \sum_{i=1}^{20} Z_i$. Therefore,

$$E[N] = \sum_{i=1}^{20} P\{Z_i = 1\}$$

Person i does not receive a gift if each of the other 19 people independently choose someone else. Thus, this happens with probability $(18/19)^{19}$ (binomial($n = 19,\ p = 18/19$) and asking for 19 successes). Thus,

$$E[N] = \sum_{i=1}^{20} P\{Z_i = 1\} = 20\left(\frac{18}{19}\right)^{19} \approx 7.16$$

4.10 Properties of Distribution Functions

Recall that for a random variable X we define

$$F(t) = F_X(t) = P\{X \leq t\}$$

The following hold:

1. F is non-decreasing: $a < b \implies F(a) \leq F(b)$.

2. $\lim_{b\to\infty} F(b) = 1$.

3. $\lim_{b\to-\infty} F(b) = 0$.

4. F is right continuous: Let $b_n > 0$ be such that $b_n \to b$, then $\lim_{n\to\infty} F(b_n) = F(b)$.

The final three can be proven by the following fact. A sequence of events $\{E_n\}$ is said to be increasing if

$$E_1 \subset E_2 \subset \ldots \subset E_n \subset \ldots$$

and it is said to be decreasing if

$$E_1 \supset E_2 \supset \ldots \supset E_n \supset \ldots$$

For an increasing sequence we define

$$\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$$

and for a decreasing sequence we define

$$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$$

See Section 2.6 for a proof of the following:

THEOREM: If $\{E_n\}$ is either increasing or decreasing, then

$$\boxed{\lim_{n\to\infty} P(E_n) = P\left(\lim_{n\to\infty} E_n\right)}$$

Thus, to prove property 2 we let b_n denote an increasing sequence that goes to ∞, then $\{X \le b_n\}$ is an increasing sequence whose union is $\{X < \infty\}$. So,

$$\lim_{n\to\infty} P\{X \le b_n\} = P\{X < \infty\} = 1$$

All probability questions about X can be answered in terms of the c.d.f., F. For example,

$$P\{a < X \le b\} = F(b) - F(a) \quad \text{for all} \quad a < b \qquad (*)$$

This equation can best be seen to hold if we write the event $\{X \le b\}$ as the union of the mutually exclusive events $\{X \le a\}$ and $\{a < X \le b\}$. That is,

$$\{X \le b\} = \{X \le a\} \cup \{a < X \le b\}$$

so

$$P\{X \le b\} = P\{X \le a\} + P\{a < X \le b\}$$

which establishes equation $(*)$.

Also, say we want $P\{X < b\}$. Then,

$$P\{X < b\} = P\left(\bigcap_{n=1}^{\infty}\{X \le b - 1/n\}\right) = \lim_{n\to\infty} P\{X \le b - 1/n\} = \lim_{n\to\infty} F(b - 1/n) = F(b^-)$$

i.e. the limit from the left (which is not necessarily the value $F(b)$).