For convenience, the characteristic properties of the following six discrete random variable models are summarized in this section: *Equilikely*$(a, b)$, *Bernoulli*$(p)$, *Geometric*$(p)$, *Pascal*$(n, p)$, *Binomial*$(n, p)$, and *Poisson*$(\mu)$. For more details about these models, see Sections 6.1 and 6.2; for supporting software, see the random variable models library `rvms` in Appendix D and the random variate generators library `rvgs` in Appendix E.

### 6.4.1 EQUILIKELY

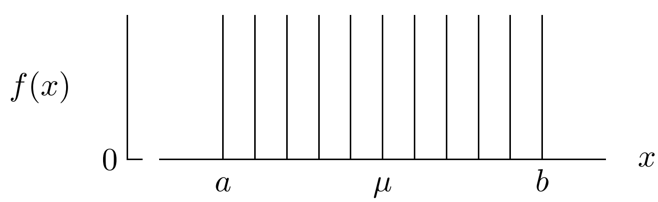
**Definition 6.4.1** The discrete random variable $X$ is *Equilikely*$(a, b)$ if and only if

- the parameters $a$, $b$ are integers with $a < b$
- the possible values of $X$ are $\mathcal{X} = \{a, a+1, \ldots, b\}$
- the pdf of $X$ is

$$f(x) = \frac{1}{b-a+1} \qquad\qquad x = a, a+1, \ldots, b$$

as illustrated in Figure 6.4.1

**Figure 6.4.1.**
*Equilikely*$(a, b)$ **pdf.**




- the cdf of $X$ is

$$F(x) = \frac{x-a+1}{b-a+1} \qquad\qquad x = a, a+1, \ldots, b$$

- the idf of $X$ is

$$F^*(u) = a + \lfloor (b-a+1)u \rfloor \qquad\qquad 0 < u < 1$$

- the mean of $X$ is

$$\mu = \frac{a+b}{2}$$

- the standard deviation of $X$ is

$$\sigma = \sqrt{\frac{(b-a+1)^2 - 1}{12}}.$$

An *Equilikely*$(a, b)$ random variable is used to model those situations where a discrete random variable is restricted to the integers between $a$ and $b$ inclusive and all values in this range are equally likely. A typical application will involve a model derived from a statement like "... an element is selected *at random* from a finite set ... " An *Equilikely* random variable is also known as a *discrete uniform*, *DU*, or *rectangular* random variable — terminology that is not used in this book.

### 6.4.2 BERNOULLI

**Definition 6.4.2** The discrete random variable $X$ is *Bernoulli*$(p)$ if and only if

- the real-valued parameter $p$ satisfies $0 < p < 1$
- the possible values of $X$ are $\mathcal{X} = \{0, 1\}$
- the pdf of $X$ is
$$f(x) = p^x(1-p)^{1-x} \qquad x = 0, 1$$
as illustrated in Figure 6.4.2 for $p = 0.6$

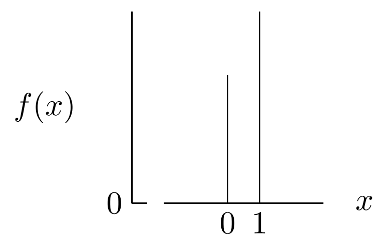



**Figure 6.4.2.** *Bernoulli*(0.6) **pdf.**

- the cdf of $X$ is
$$F(x) = (1-p)^{1-x} \qquad x = 0, 1$$
- the idf of $X$ is
$$F^*(u) = \begin{cases} 0 & 0 < u < 1-p \\ 1 & 1-p \le u < 1 \end{cases}$$
- the mean of $X$ is
$$\mu = p$$
- the standard deviation of $X$ is
$$\sigma = \sqrt{p(1-p)}.$$

A *Bernoulli*$(p)$ random variable is used to model the Boolean situation where only two outcomes are possible — success or failure, true or false, 1 or 0, etc. The parameter $p$ determines the probability of the two possible outcomes with the convention that

$$p = \Pr(\text{success}) = \Pr(X = 1)$$

and

$$1 - p = \Pr(\text{failure}) = \Pr(X = 0).$$

The random variables $X_1$, $X_2$, $X_3$, ..., define an *iid* sequence of so-called *Bernoulli*$(p)$ *trials* if and only if each $X_i$ is *Bernoulli*$(p)$ and each is statistically independent of all the others. The repeated tossing of a coin (biased or not) is the classic example of an *iid* sequence — a head is equally likely on each toss ($p$ does not change) and the coin has no memory of the previous outcomes (independence).

### 6.4.3 GEOMETRIC

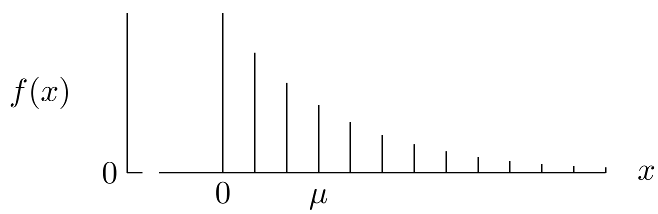
**Definition 6.4.3** The discrete random variable $X$ is *Geometric*$(p)$ if and only if

- the real-valued parameter $p$ satisfies $0 < p < 1$
- the possible values of $X$ are $\mathcal{X} = \{0, 1, 2, \ldots\}$
- the pdf of $X$ is

$$f(x) = p^x(1-p) \qquad\qquad x = 0, 1, 2, \ldots$$

  as illustrated in Figure 6.4.3 for $p = 0.75$

**Figure 6.4.3.**
*Geometric*(0.75) **pdf.**




- the cdf of $X$ is

$$F(x) = 1 - p^{x+1} \qquad\qquad x = 0, 1, 2, \ldots$$

- the idf of $X$ is

$$F^*(u) = \left\lfloor \frac{\ln(1-u)}{\ln(p)} \right\rfloor \qquad\qquad 0 < u < 1$$

- the mean of $X$ is

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

- the standard deviation of $X$ is

$$\sigma = \frac{\sqrt{p}}{1-p}.$$

A *Geometric*$(p)$ random variable is conventionally used to model the number of "successes" (1's) before the first "failure" (0) in a sequence of independent *Bernoulli*$(p)$ trials. For example, if a coin has $p$ as the probability of a head (success), then $X$ counts the number of heads before the first tail (failure). Equivalently, a *Geometric*$(p)$ random variable can be interpreted as a model for random sampling from a urn containing balls, a fraction $p$ of which are 1's with the remainder 0's. A *Geometric*$(p)$ random variable counts the number of 1's that are drawn, with replacement, before the first 0. If $X$ is *Geometric*$(p)$ then $X$ has the *memoryless* property — for any non-negative integer $x'$

$$\Pr(X \geq x + x' \mid X \geq x') = \Pr(X \geq x) = p^x \qquad\qquad x = 0, 1, 2, \ldots$$

*independent* of $x'$. An intuitive interpretation of this property is that a string of $x' - 1$ consecutive 1's is followed by a sequence of 0's and 1's that is probabilistically the same as a brand new sequence.

### 6.4.4 PASCAL

**Definition 6.4.4** The discrete random variable $X$ is *Pascal*$(n, p)$ if and only if

- the parameter $n$ is a positive integer
- the real-valued parameter $p$ satisfies $0 < p < 1$
- the possible values of $X$ are $\mathcal{X} = \{0, 1, 2, \ldots\}$
- the pdf of $X$ is

$$f(x) = \binom{n+x-1}{x} p^x (1-p)^n \qquad\qquad x = 0, 1, 2, \ldots$$

as illustrated in Figure 6.4.4 for $(n, p) = (5,\ 2/7)$

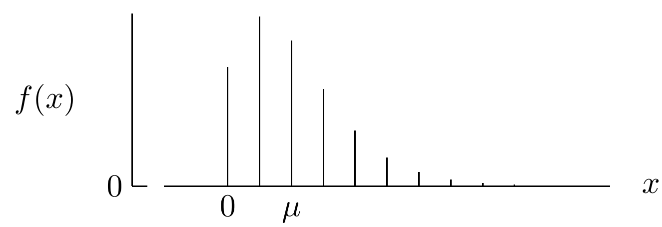



**Figure 6.4.4.** *Pascal*$(5, 2/7)$ **pdf.**

- the cdf of $X$ contains an incomplete beta function (see Appendix D)

$$F(x) = 1 - I(x+1,\ n,\ p) \qquad\qquad x = 0, 1, 2, \ldots$$

- except for special cases, the idf of $X$ must be determined by numerical inversion
- the mean of $X$ is

$$\mu = \frac{np}{1-p}$$

- the standard deviation of $X$ is

$$\sigma = \frac{\sqrt{np}}{1-p}.$$

A *Pascal*$(n, p)$ random variable is the number of 1's before the $n^{\text{th}}$ 0 in an *iid* sequence of independent *Bernoulli*$(p)$ trials. Therefore, a *Pascal*$(1, p)$ random variable and a *Geometric*$(p)$ random variable are equivalent. Moreover, $X$ is a *Pascal*$(n, p)$ random variable if and only if

$$X = X_1 + X_2 + \cdots + X_n,$$

where $X_1,\ X_2,\ \ldots,\ X_n$ is an *iid* sequence of *Geometric*$(p)$ random variables. In terms of the urn model presented in Section 6.4.3, a *Pascal*$(n, p)$ random variable counts the number of 1's that are drawn, with replacement, before the $n^{\text{th}}$ 0. A *Pascal*$(n, p)$ random variable is also called a *negative binomial* — terminology that is not used in this book.

### 6.4.5 BINOMIAL

**Definition 6.4.5** The discrete random variable $X$ is *Binomial*$(n, p)$ if and only if

- the parameter $n$ is a positive integer
- the real-valued parameter $p$ satisfies $0 < p < 1$
- the possible values of $X$ are $\mathcal{X} = \{0, 1, 2, \ldots, n\}$
- the pdf of $X$ is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad\qquad x = 0, 1, \ldots, n$$

as illustrated in Figure 6.4.5 for $(n, p) = (10, 0.3)$

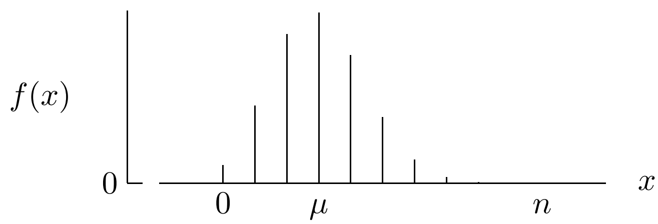



**Figure 6.4.5.** *Binomial*$(10, 0.3)$ **pdf.**

- the cdf of $X$ contains an incomplete beta function (see Appendix D)

$$F(x) = \begin{cases} 1 - I(x+1,\, n-x,\, p) & x = 0, 1, \ldots, n-1 \\ 1 & x = n \end{cases}$$

- except for special cases, the idf of $X$ must be determined by numerical inversion
- the mean of $X$ is

$$\mu = np$$

- the standard deviation of $X$ is

$$\sigma = \sqrt{np(1-p)}.$$

A *Binomial*$(n, p)$ random variable is the number of 1's in a sequence of $n$ independent *Bernoulli*$(p)$ trials. Therefore a *Binomial*$(1, p)$ random variable and a *Bernoulli*$(p)$ random variable are the same. Equivalently, $X$ is a *Binomial*$(n, p)$ random variable if and only if

$$X = X_1 + X_2 + \cdots + X_n,$$

where $X_1, X_2, \ldots, X_n$ is an *iid* sequence of *Bernoulli*$(p)$ random variables. For example, if a coin has $p$ as the probability of a head (success), then $X$ counts the number of heads in a sequence of $n$ tosses. In terms of the urn model, a *Binomial*$(n, p)$ random variable counts the number of 1's that will be drawn, with replacement, if exactly $n$ balls are drawn. Because drawing $x$ 1's out of $n$ is equivalent to drawing $n - x$ 0's it follows that $X$ is *Binomial*$(n, p)$ if and only if $n - X$ is a *Binomial*$(n, 1-p)$ random variable.

### 6.4.6 POISSON

**Definition 6.4.6** The discrete random variable $X$ is *Poisson*$(\mu)$ if and only if

- the real-valued parameter $\mu$ satisfies $\mu > 0$
- the possible values of $X$ are $\mathcal{X} = \{0, 1, 2, \ldots\}$
- the pdf of $X$ is

$$f(x) = \frac{\mu^x \exp(-\mu)}{x!} \qquad\qquad x = 0, 1, 2 \ldots$$

as illustrated in Figure 6.4.6 for $\mu = 5$

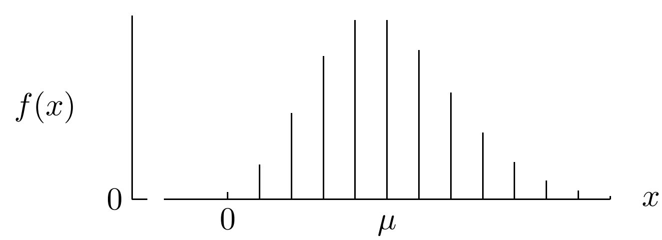



**Figure 6.4.6.** *Poisson*(5) **pdf.**

- the cdf of $X$ contains an incomplete gamma function (see Appendix D)

$$F(x) = 1 - P(x + 1,\, \mu) \qquad\qquad x = 0, 1, 2 \ldots$$

- except for special cases, the idf of $X$ must be determined by numerical inversion
- the mean of $X$ is $\mu$
- the standard deviation of $X$ is

$$\sigma = \sqrt{\mu}.$$

A *Poisson*$(\mu)$ random variable is a limiting case of a *Binomial* random variable. That is, let $X$ be a *Binomial*$(n, \mu/n)$ random variable ($p = \mu/n$). Fix the values of $\mu$ and $x$ and consider what happens in the limit as $n \to \infty$. The pdf of $X$ is

$$\begin{aligned} f(x) &= \frac{n!}{x!(n-x)!}\left(\frac{\mu}{n}\right)^x \left(1 - \frac{\mu}{n}\right)^{n-x} \\ &= \frac{\mu^x}{x!}\left(\frac{n!n^x}{(n-x)!(n-\mu)^x}\right)\left(1 - \frac{\mu}{n}\right)^n. \end{aligned}$$

It can be shown that

$$\lim_{n\to\infty}\left(\frac{n!n^x}{(n-x)!(n-\mu)^x}\right) = 1 \qquad\qquad \text{and} \qquad\qquad \lim_{n\to\infty}\left(1 - \frac{\mu}{n}\right)^n = \exp(-\mu).$$

Therefore

$$\lim_{n\to\infty} f(x) = \frac{\mu^x \exp(-\mu)}{x!}$$

which shows that for large values of $n$, *Binomial*$(n, \mu/n)$ and *Poisson*$(\mu)$ random variables are virtually identical, particularly if $\mu$ is small.

### 6.4.7 SUMMARY

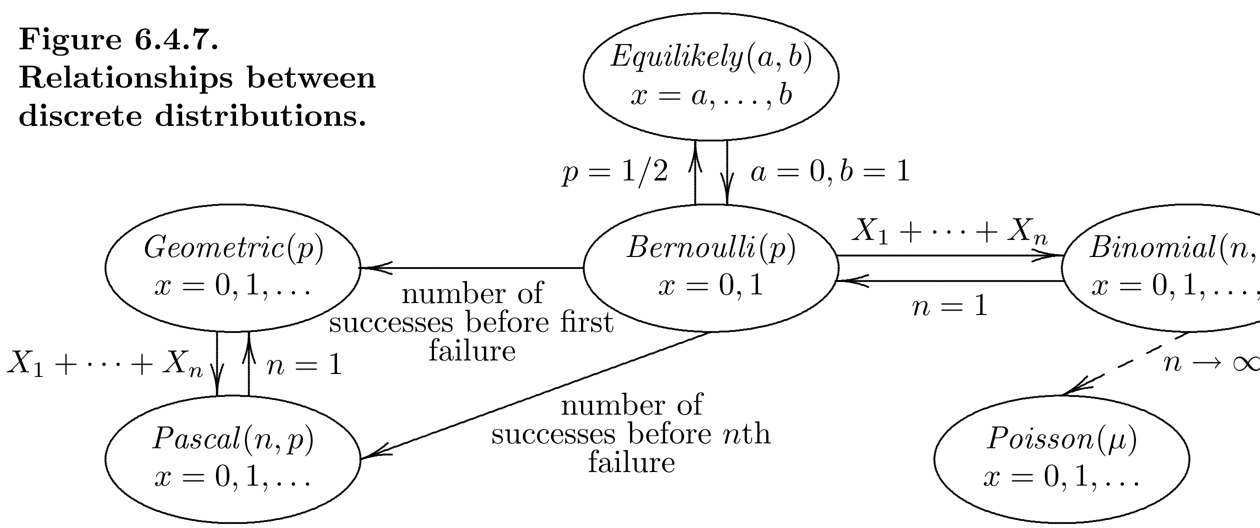
Effective discrete-event simulation modeling requires that the modeler be familiar with several parametric distributions which can be used to mimic the stochastic elements of the model. In order to choose the proper distribution, it is important to know

- how these distributions arise;
- their support $\mathcal{X}$;
- their mean $\mu$;
- their variance $\sigma^2$;
- the shape of their pdf.

Thus when a modeling situation arises, the modeler has a wide array of options for selecting a stochastic model.

It is also important to know how these distributions relate to one another. Figure 6.4.7 summarizes relationships between the six distributions considered in this text. Listed in each oval are the name, parameter(s), and support of each distribution. The solid arrows connecting the ovals denote special cases [e.g., the *Bernoulli*$(p)$ distribution is a special case of the *Binomial*$(n, p)$ distribution when $n = 1$] and transformations [e.g., the sum (convolution) of $n$ independent and identically distributed *Bernoulli*$(p)$ random variables has a *Binomial*$(n, p)$ distribution]. The dashed line between the *Binomial*$(n, p)$ and *Poisson*$(\mu)$ distributions indicates that the limiting distribution of a binomial random variable as $n \to \infty$ has a Poisson distribution.

**Figure 6.4.7. Relationships between discrete distributions.**




There are internal characteristics associated with these distributions that are not shown in Figure 6.4.7. One such example is that the sum of independent Poisson random variables also has a Poisson distribution. Another is that the sum of independent binomial random variables with identical parameters $p$ also has a binomial distribution.

The mean and variance of a random variable are special cases of what are called *moments*. The table below summarizes the first four moments of the distribution of the six discrete random variables surveyed in this chapter. The first moment, the mean $\mu = E[X]$, and the second moment about the mean, the variance $\sigma^2 = E[(X-\mu)^2]$, have been defined earlier. The skewness and kurtosis are the third and fourth standardized centralized moments about the mean, defined by

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \qquad \text{and} \qquad E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right].$$

The skewness is a measure of the symmetry of a distribution. A symmetric pdf has a skewness of zero. A positive skewness typically indicates that the distribution is "leaning" to the left, and a negative skewness typically indicates that the distribution is "leaning" to the right. The geometric distribution, for example, has a positive skewness for all values of its parameter $p$. The kurtosis is a measure of the peakedness and tail behavior of a distribution. In addition to being measures associated with particular distributions, the higher-order moments are occasionally used to differentiate between the parametric models.

| Distribution | Mean | Variance | Skewness | Kurtosis |
|---|---|---|---|---|
| *Equilikely*$(a, b)$ | $\frac{a+b}{2}$ | $\frac{(b-a+1)^2-1}{12}$ | $0$ | $\frac{3}{5}\left[3-\frac{4}{(b-a)(b-a+2)}\right]$ |
| *Bernoulli*$(p)$ | $p$ | $p(1-p)$ | $\frac{1-2p}{\sqrt{p(1-p)}}$ | $\frac{1}{p(1-p)}-3$ |
| *Geometric*$(p)$ | $\frac{p}{1-p}$ | $\frac{p}{(1-p)^2}$ | $\frac{1+p}{\sqrt{1-p}}$ | $\frac{p^2+7p+1}{p}$ |
| *Pascal*$(n, p)$ | $\frac{np}{1-p}$ | $\frac{np}{(1-p)^2}$ | $\frac{1+p}{\sqrt{n(1-p)}}$ | $3+\frac{6}{n}+\frac{(1-p)^2}{np}$ |
| *Binomial*$(n, p)$ | $np$ | $np(1-p)$ | $\frac{1-2p}{\sqrt{np(1-p)}}$ | $3-\frac{6}{n}+\frac{1}{np(1-p)}$ |
| *Poisson*$(\mu)$ | $\mu$ | $\mu$ | $\frac{1}{\sqrt{\mu}}$ | $3+\frac{1}{\mu}$ |

Although discussion is limited here to just six discrete distributions, there are many other parametric distributions capable of modeling discrete distributions. For a complete list of common discrete and continuous parametric distributions including their pdf, cdf, idf, and moments, we recommend the compact work of Evans, Hastings, and Peacock (2000) or the encyclopedic works of Johnson, Kotz, and Kemp (1993) and Johnson, Kotz, and Balakrishnan (1994, 1995, 1997).

We conclude this section with a brief discussion of the techniques and software associated with the evaluation of the pdf, cdf, and idf of the six discrete distributions surveyed in this chapter.

### 6.4.8 PDF, CDF AND IDF EVALUATION

**Library rvms**

Pdf's, cdf's, and idf's for all six of the discrete random variable models in this section can be evaluated by using the functions in the library `rvms`, listed in Appendix D. For example, if $X$ is *Binomial*$(n,p)$ then for any $x = 0, 1, 2, \ldots, n$

```
pdf = pdfBinomial(n, p, x);                    /* f(x) */
cdf = cdfBinomial(n, p, x);                    /* F(x) */
```

and for any $0.0 < u < 1.0$

```
idf = idfBinomial(n, p, u);                    /* F*(u) */
```

This library also has functions to evaluate pdf's, cdf's, and idf's for all the continuous random variables cataloged in Chapter 7.

**Alternative, Recursive Approaches for Calculating pdf Values**

Although this approach is not used in the library `rvms` to evaluate pdf's and cdf's, if $X$ is a discrete random variable then pdf values (and cdf values, by summation) can usually be easily generated recursively. In particular:

- if $X$ is *Geometric*$(p)$ then

$$\begin{aligned} f(0) &= 1 - p \\ f(x) &= p f(x-1) \qquad\qquad x = 1, 2, 3, \ldots \end{aligned}$$

- if $X$ is *Pascal*$(n,p)$ then

$$\begin{aligned} f(0) &= (1-p)^n \\ f(x) &= \frac{(n+x-1)p}{x} f(x-1) \qquad\qquad x = 1, 2, 3, \ldots \end{aligned}$$

- if $X$ is *Binomial*$(n,p)$ then

$$\begin{aligned} f(0) &= (1-p)^n \\ f(x) &= \frac{(n-x+1)p}{x(1-p)} f(x-1) \qquad\qquad x = 1, 2, 3, \ldots, n \end{aligned}$$

- if $X$ is *Poisson*$(\mu)$ then

$$\begin{aligned} f(0) &= \exp(-\mu) \\ f(x) &= \frac{\mu}{x} f(x-1) \qquad\qquad x = 1, 2, 3, \ldots \end{aligned}$$

### 6.4.9 EXERCISES

**Exercise 6.4.1** ($a$) Simulate the tossing of a fair coin 10 times and record the number of heads. ($b$) Repeat this experiment 1000 times and generate a discrete-data histogram of the results. ($c$) Verify numerically that the relative frequency of the number of heads is approximately equal to the pdf of a *Binomial*(10, 0.5) random variable.

**Exercise 6.4.2** Prove that a *Geometric*($p$) random variable has the memoryless property.

**Exercise 6.4.3** Derive the *Geometric*($p$), *Pascal*($n, p$), *Binomial*($n, p$), and *Poisson*($\mu$) recursive pdf equations.

**Exercise 6.4.4** Use the *Binomial*($n, p$) recursive pdf equations to implement the functions `pdfBinomial(n,p,x)` and `cdfBinomial(n,p,x)`. Compare your implementations, in terms of both accuracy and speed, with the corresponding functions in the library `rvms`. Use $n = 10, 100, 1000$ with $\mu = 5$ and $p = \mu/n$. The comparison can be restricted to possible values $x$ within the range $\mu \pm 3\sigma$.

**Exercise 6.4.5** Verify numerically that the pdf of a *Binomial*(25, 0.04) random variable is virtually identical to the pdf of a *Poisson*($\mu$) random variable for an appropriate value of $\mu$. Evaluate these pdf's in two ways: by using the appropriate pdf functions in the library `rvms` and by using the *Binomial*($n$ , $p$) recursive pdf equations.

**Exercise 6.4.6**[a] Prove or disprove: if $X$ is a *Pascal*($n, p$) random variable then the cdf of $X$ is $F(x) = I(n,\, x + 1,\, 1 - p)$.

**Exercise 6.4.7**[a] Prove or disprove: if $X$ is a *Binomial*($n, p$) random variable then the cdf of $X$ is $F(x) = I(n - x,\, x + 1,\, 1 - p)$.

**Exercise 6.4.8** Prove that if $X$ is *Binomial*($n, p$) then

$$\Pr(X > 0) = \sum_{x=0}^{n-1} p(1 - p)^x.$$

What is the probability interpretation of this equation?

**Exercise 6.4.9** ($a$) If you play Pick-3 (as in Example 6.1.14) once a day for 365 consecutive days, what is the probability that you will be ahead at the end of this period? ($b$) What is your expected yield (winnings) at the end of this period?

**Exercise 6.4.10** Let $X$ be a *Geometric*($p$) random variable. As an alternative to the definition used in this book, some authors define the random variable $Y = X + 1$ to be *Geometric*($p$). ($a$) What is the pdf, cdf, idf, mean, and standard deviation of $Y$? ($b$) Does this random variable also have the memoryless property?