# Chapter 3 Discrete Random Variables

## Definition 3.1

A random variable that can take on at most a countable number of possible values is said to be discrete. For a discrete random variable *X*, we define the **probability mass function** (pmf) $p_X(a)$ **of** *X* by

$$p_X(a) = P(X = a).$$

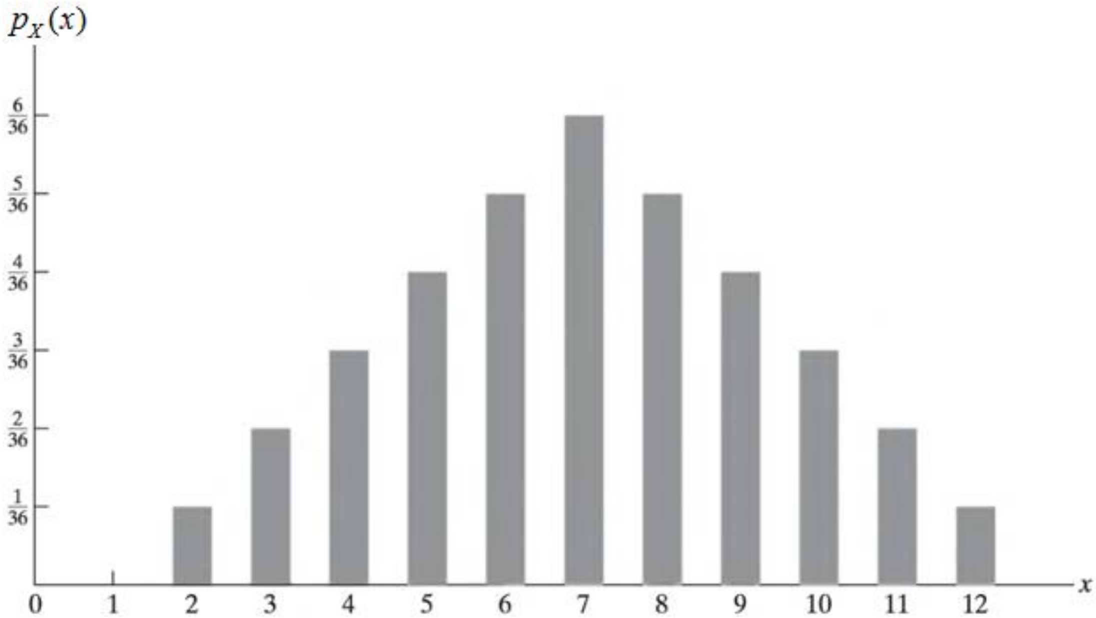
The probability mass function $p_X(a)$ is positive for at most a countable number of values of $a$. That is, if *X* must assume one of the values $x_1, x_2, \cdots,$ then

$$p_X(x_i) \geq 0 \quad \text{for } i = 1, 2, \cdots$$
$$p_X(x) = 0 \quad \text{for any other values of } x.$$

Since *X* must take on one of the values $x_i,$ we have

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

A graph of the probability mass function of the random variable represents the sum of two dice rolled.




Recall that the **distribution function** $F_X(x)$, **for a random variable** $X$ is the probability that $X$ is less than or equal to a given number. That is, $F_X(x) = \Pr(X \le x)$.

Proposition 3.2

The **distribution function** $F_X$ **of a discrete random variable** $X$ is given by

$$F_X(x) = \sum_{a \le x} p_X(a).$$

If $X$ is a discrete random variable whose possible values are $x_1, x_2, x_3, \cdots,$ where $x_1 < x_2 < x_3 < \cdots,$ then the distribution function $F_X$ of $X$ is a step function. That is, the value of $F_X$ is constant in the intervals $(x_{i-1}, x_i)$ and then takes a step (or jump) of size $p_X(x_i)$ at $x_i$.

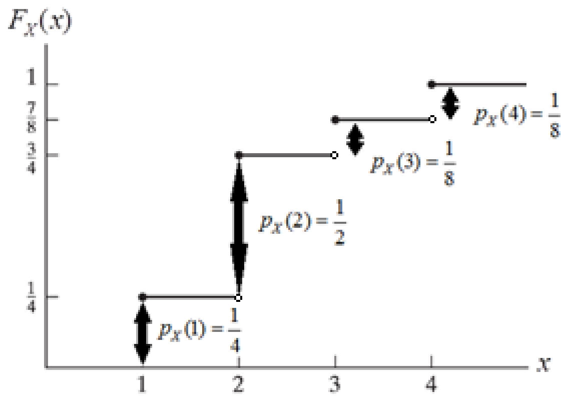
Example 3.3

If $X$ has a probability mass function given by

$$p_X(x) = \begin{cases} \frac{1}{4} & x = 1 \\ \frac{1}{2} & x = 2 \\ \frac{1}{8} & x = 3,4 \\ 0 & \text{otherwise,} \end{cases}$$

then its distribution function is

$$F_X(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{4} & 1 \le x < 2 \\ \frac{1}{4} + \frac{1}{2} = \frac{3}{4} & 2 \le x < 3 \\ \frac{1}{4} + \frac{1}{2} + \frac{1}{8} = \frac{7}{8} & 3 \le x < 4 \\ \frac{1}{4} + \frac{1}{2} + \frac{1}{8} + \frac{1}{8} = 1 & 4 \le x \end{cases}$$

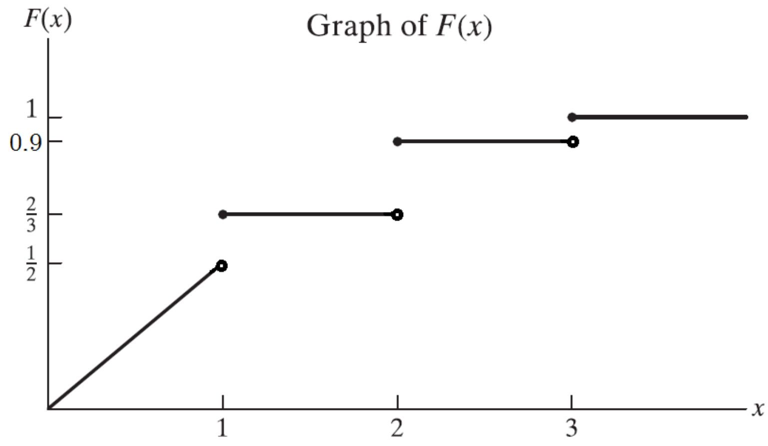



Example 3.4

The distribution function of the random variable $X$ is given by

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x/2 & 0 \le x < 1 \\ 2/3 & 1 \le x < 2 \\ 0.9 & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

A graph of $F_X(x)$ is presented in the below figure.




$$P(2) = F(2) - \lim_{x \to 2^-} F(x) = 0.9 - \frac{2}{3} = \frac{7}{30}$$
$$P(0.3) = 0$$
$$P\big((0.5,1)\big) = \lim_{x \to 1^-} F(x) - F(0.5) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$
$$P\big((0.5,1]\big) = F(1) - F(0.5) = \frac{2}{3} - \frac{1}{4} = \frac{5}{12}$$

Indeed, $X$ is neither discrete nor continuous. Its probability mass/density function is

$$f_X(x) = \begin{cases} 1/2 & 0 \le x < 1 \\ 1/6 & x = 1 \\ 7/30 & x = 2 \\ 0.1 & x = 3 \\ 0 & \text{otherwise} \end{cases}$$

Definition 3.5

If $X$ is a discrete random variable having a probability mass function $p_X(x)$, then the **expectation**, or the **expected value**, **of** $X$, denoted by $E[X]$, is defined by

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

Example 3.6

We say that $I_A$ is an indicator for the event *A* if

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise.} \end{cases}$$

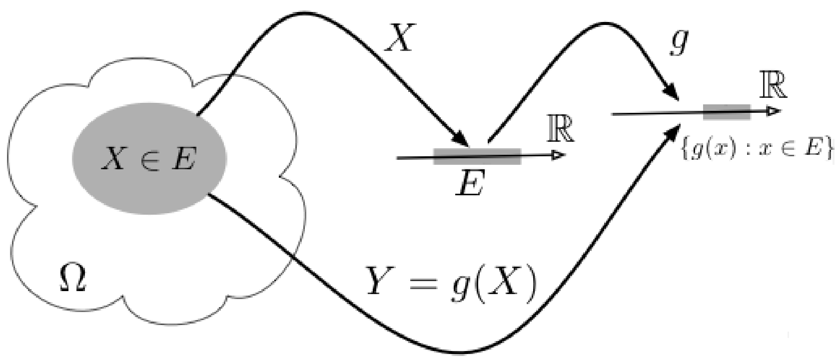
Find $E[I_A]$.

Solution

Since $p(1) = P(A),\ p(0) = 1 - P(A),$ we have

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

Since *g*(*X*) will equal *g*(*x*) whenever *X* is equal to *x*, it seems reasonable that $E[g(X)]$ should just be a weighted average of the values *g*(*x*), with *g*(*x*) being weighted by the probability that *X* is equal to *x*. That is, the following definition is quite intuitive.




Definition 3.7

If *X* is a discrete random variable, then for any real-valued function *g*,

$$E[g(X)] = \sum_{x:p_X(x)>0} g(x)\,p_X(x).$$

Theorem 3.8

If *a* and *b* are constants, then

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

Proof

$$E[aX+b]=\sum_x (ax+b)p(x)=a\sum_x xp(x)+b\sum_x p(x)=aE[X]+b.$$

Remark 3.9

In general, $E[g(X)]$ and $g\left(E[X]\right)$ are not equal. If $g''(x)>0$ (or $g''(x)<0$ ) for all $x$, then $g\left(E[X]\right)\leq E[g(X)]$ (or $g\left(E[X]\right)\geq E[g(X)]$).

Example 3.10

Let $X$ denote a random variable probability mass function

$$p_X(x)=\begin{cases} 0.2 & x=-1 \\ 0.5 & x=0 \\ 0.3 & x=1 \\ 0 & \text{otherwise} \end{cases}$$

Compute $E[X^2]$.

Solution

Let $Y=X^2$. Then the probability mass function of $Y$ is given by

$$p_Y(y)=P(Y=y)=P(X^2=y)=\begin{cases} 0.2+0.3=0.5 & y=1 \\ 0.5 & y=0 \\ 0 & \text{otherwise} \end{cases}$$

Hence,

$$E[X^2]=E[Y]=1(0.5)+0(0.5)=0.5.$$

Clearly, $E[X]=-1(0.2)+0(0.5)+1(0.3)=0.1.$ Note that

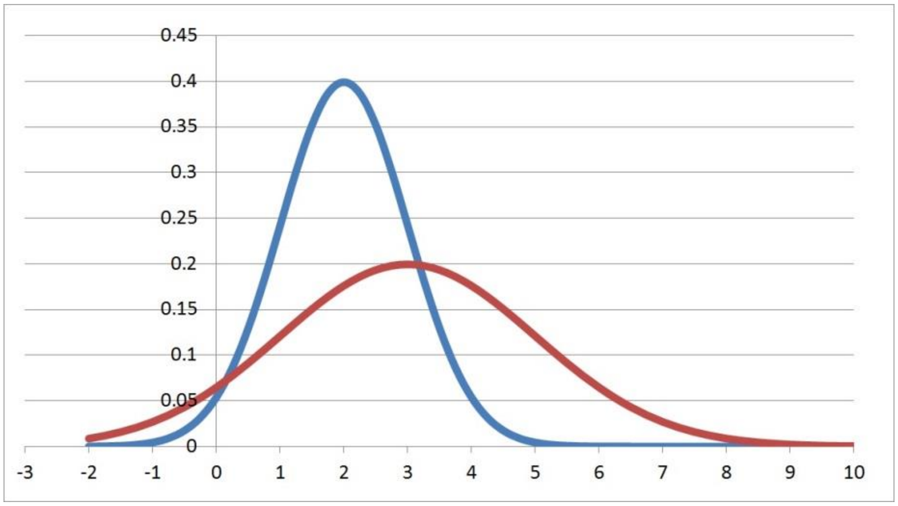
$$E[X^2]=0.5\neq 0.01=E[X]^2.$$

Definition 3.11

If $X$ is a random variable with mean $\mu=E[X]$, then the **variance of** $X$, denoted by $Var(X)$, is defined by

$$Var(X)=E\left[(X-\mu)^2\right].$$

By definition, the variance of $X$ is the average value of $(X-\mu)^2$. Since $(X-\mu)^2 \geq 0$, the variance is always larger than or equal to zero. A large value of the variance means that $(X-\mu)^2$ is often large, so $X$ often takes values far from its mean. This means that the distribution (red curve) is spread out. On the other hand, a low variance means that the distribution (blue curve) is concentrated around its average. $Var(X) = 0$ if and only if $X$ is a constant.




An alternative formula for $Var(X)$ is derived as follows:

$$\begin{aligned}
Var(X) &= E\left[(X-\mu)^2\right] \\
&= \sum_x (x-\mu)^2 p(x) \\
&= \sum_x \left(x^2 - 2\mu x + \mu^2\right) p(x) \\
&= \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) \\
&= E[X^2] - 2\mu^2 + \mu^2 \\
&= E[X^2] - \mu^2 \\
&= E[X^2] - E[X]^2.
\end{aligned}$$

Theorem 3.12

For any constants $a$ and $b$,

$$Var(aX+b) = a^2 Var(X).$$

Proof

Note that $E[aX+b] = aE[X]+b = a\mu + b.$

$$\begin{aligned} Var(aX+b) &= E\left[(aX+b-a\mu-b)^2\right] \\ &= E\left[a^2(X-\mu)^2\right] \\ &= a^2E\left[(X-\mu)^2\right] \\ &= a^2Var(X). \end{aligned}$$

Definition 3.13 (Bernoulli($p$))

Suppose that a trial, or an experiment, whose outcome can be classified as either a success or a failure is performed. Let $X$ = 1 when the outcome is a success and $X$ = 0 when it is a failure. A random variable $X$ is said to be a **Bernoulli random variable with parameter** $p$ if the probability mass function of $X$ is given by

$$\begin{aligned} p(0) &= P(X=0) = 1-p \\ p(1) &= P(X=1) = p \end{aligned}$$

where $p$, $0 < p < 1$, is the probability that the trial is a success.

Theorem 3.14

Let $X$ be a Bernoulli random variable with parameter $p$. Then

$$E[X] = p \quad \text{and} \quad Var[X] = p(1-p).$$

Suppose now that $n$ independent trials, each of which results in a success with probability $p$ or in a failure with probability $1-p,$ are to be performed. If $X$ represents the number of successes that occur in the $n$ trials, then $X$ is said to be a binomial random variable with parameters ($n$, $p$). Note that there are $\binom{n}{i}$ different sequences of the $n$ outcomes leading to $i$ successes and $n-i$ failures. The probability of any particular sequence of $n$ outcomes containing $i$ successes and $n-i$ failures is, by the assumed independence of trials, $p^i(1-p)^{n-i}$.

Definition 3..15 (Binomial($n$, $p$))

The probability mass function of a **binomial random variable with parameters** (*n*, *p*) is given by

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

Thus, $X = Y_1 + \cdots + Y_n$ is a binomial random variable with parameters (*n*, *p*) if $Y_1, \cdots, Y_n \sim \text{Bernoulli}(p)$ are independent. In particular, Bernoulli(*p*) = Binomial(1, *p*)

Theorem 3.16

Let *X* be a binomial random variable with parameter (*n*, *p*). Then

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

Example 3.17

In an examination paper with 50 multiple choice questions (5 choices for each), there is only one correct answer for each question. A student answers all the questions by pure guess. Suppose that two marks will be given for each correct answer and half mark will be deducted for each incorrect answer. Then what are the expectation and variance of the total score he can obtain?

Solution

Let *X* be the number of correct answers he gets by pure guess. Then $X \sim Bin(n = 50, p = 0.2)$. Number of incorrect answers is $50 - X$. Let *Y* be the total score that he can obtain. Then

$$Y = 2X - 0.5(50 - X) = 2.5X - 25.$$

Hence $E[Y] = 2.5E[X] - 25 = 2.5np - 25 = 0$ and $Var[Y] = 2.5^2 Var[X] = 2.5^2 np(1-p) = 50.$

Definition 3.18

For a nonnegative integer valued random variable *X*, the **probability generating function** (pgf) is $P_X(z) = E\left(z^X\right) = p(0) + p(1)z + p(2)z^2 + \cdots$ for all *z* for which the expectation exists.

Clearly, $p(n) = \frac{1}{n!}\frac{d^n}{dz^n}P_X(z)\bigg|_{z=0}$ and $P_X(1) = p(0) + p(1) + p(2) + \cdots = 1.$

Theorem 3.19

Let $P_X(z)$ be probability generating function of a nonnegative integer valued random variable *X*. Then

$$E[X] = P_X'(1) \quad \text{and} \quad Var[X] = P_X''(1) + P_X'(1) - P_X'(1)^2.$$

Proof

$$P_X'(z) = p(1) + 2p(2)z + 3p(3)z^2 + \cdots + ip(i)z^{i-1} + \cdots$$
$$P_X''(z) = 2p(2) + 3\cdot 2\,p(3)z + \cdots + i(i-1)\,p(i)z^{i-2} + \cdots$$
$$P_X'(1) = p(1) + 2p(2) + 3p(3) + \cdots + ip(i) + \cdots = E[X]$$
$$P_X''(1) = 2p(2) + 3\cdot 2\,p(3) + \cdots + i(i-1)\,p(i) + \cdots = E[X(X-1)] = E[X^2] - E[X]$$
$$Var[X] = E[X^2] - E[X]^2 = P_X''(1) + P_X'(1) - P_X'(1)^2$$

Definition 3.20 (Poisson$(\lambda)$)

A random variable *X* that takes on one of the values $0,1,2,\cdots$ is said to be a **Poisson random variable with parameter** $\lambda$ if, for some $\lambda > 0$,

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

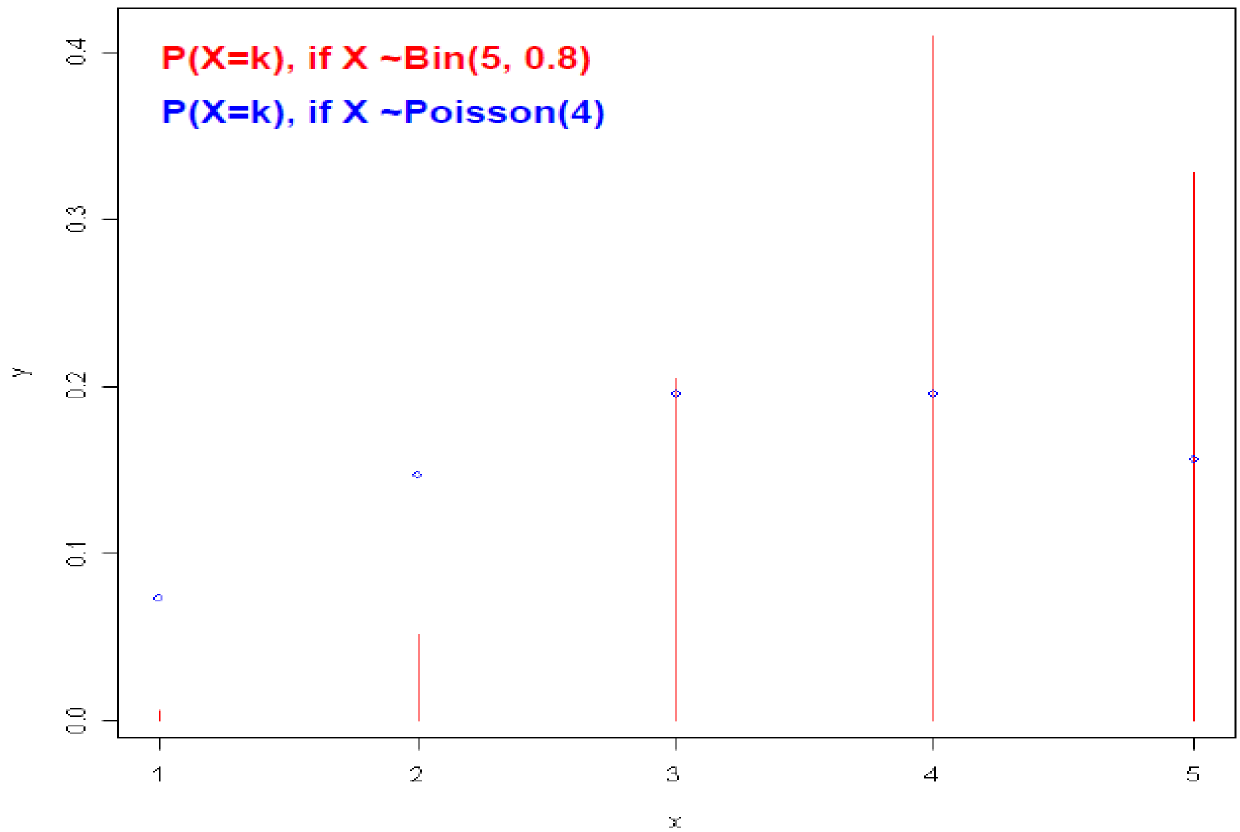
This equation defines a probability mass function, since

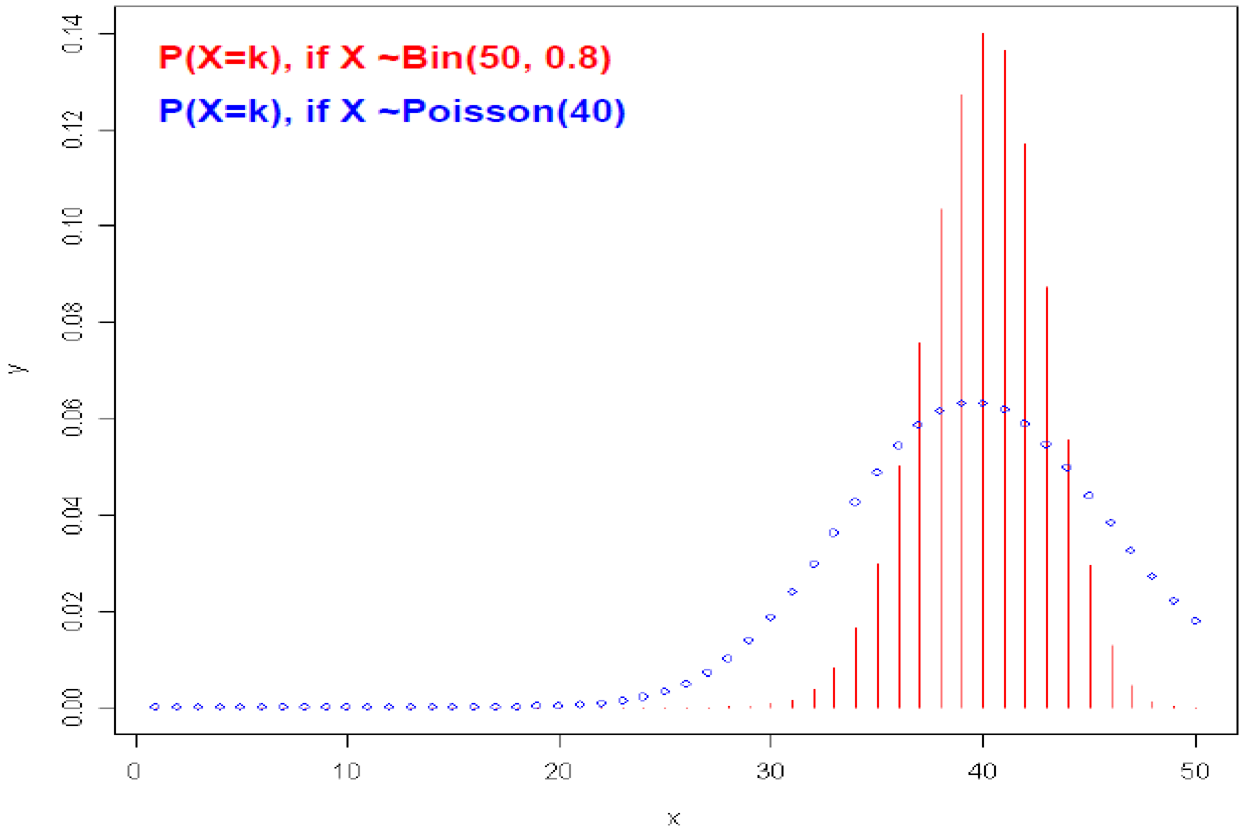
$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1.$$

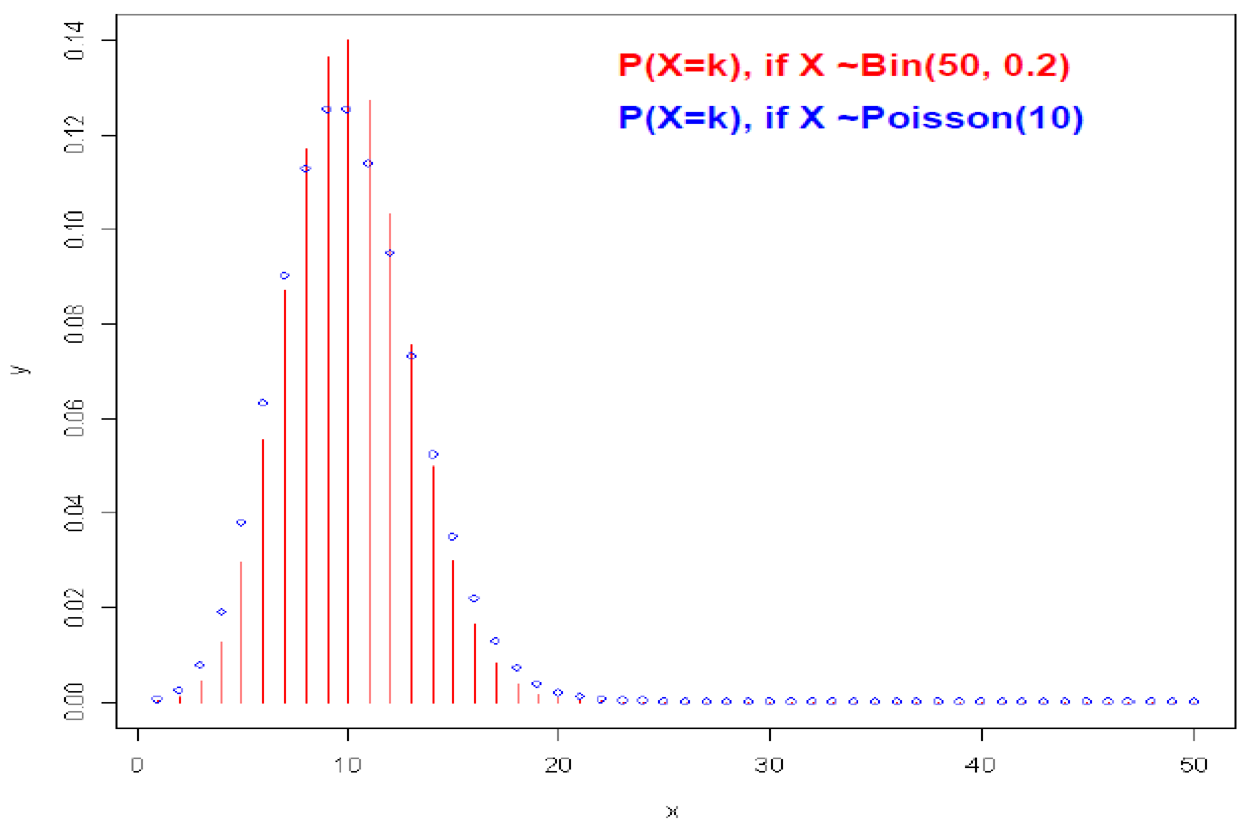
Poisson random variable can be used as an approximation for a binomial random variable with parameters (*n*, *p*) when *n* is large and *p* is small enough so that *np* is a constant. To see this, suppose that *X* is a binomial random variable with parameters (*n*, *p*), and let $\lambda = np.$ Then

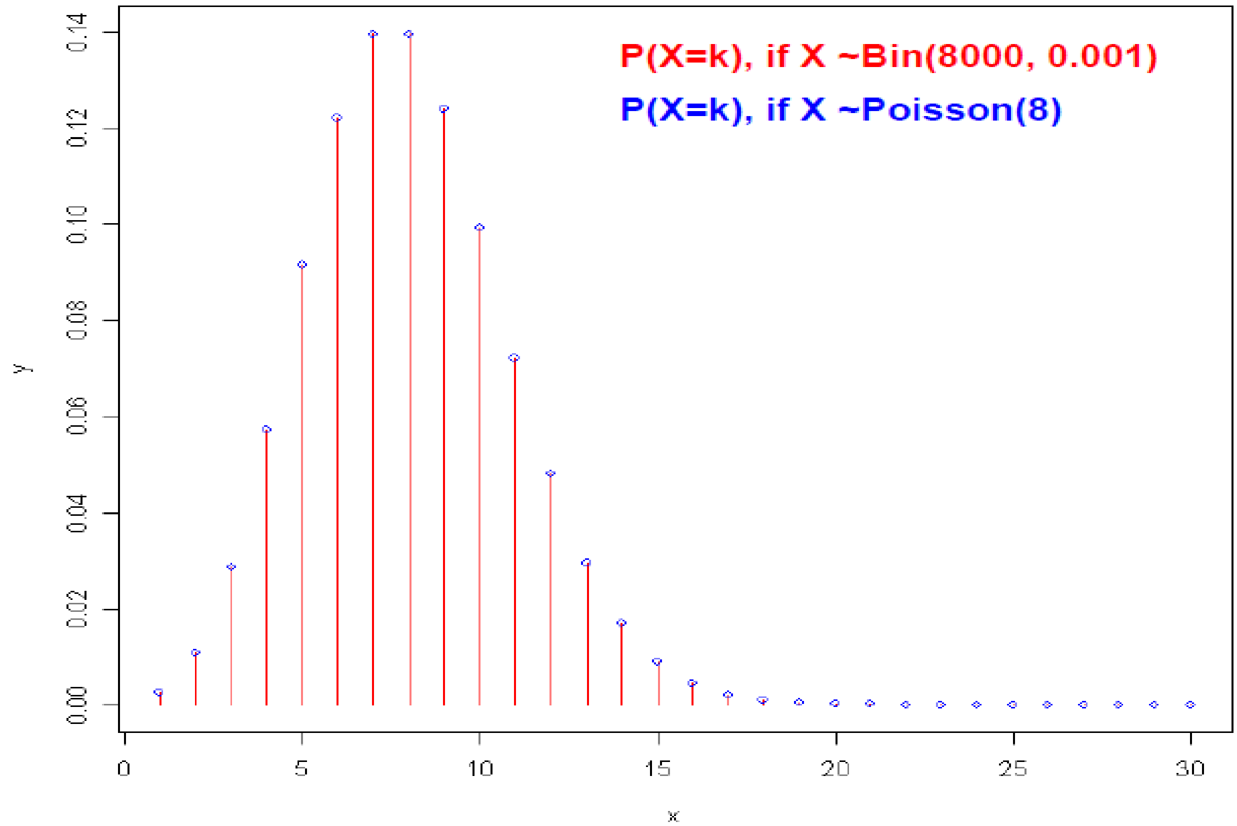
$$
\begin{aligned}
P(X=i) &= \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\
&= \frac{n(n-1)\cdots(n-i+1)}{i!} \frac{\lambda^i}{n^i} \left(1-\frac{\lambda}{n}\right)^{n-i} \\
&= \frac{\lambda^i}{i!} \cdot \frac{n(n-1)\cdots(n-i+1)}{n^i} \left(1-\frac{\lambda}{n}\right)^{n} \Big/ \left(1-\frac{\lambda}{n}\right)^{i} \\
&\approx e^{-\lambda} \frac{\lambda^i}{i!} \qquad \left[\text{as } n \to \infty, p \to 0\right]
\end{aligned}
$$

$$
\binom{100}{15}(0.1)^{15}(0.9)^{85} = 0.032682438154, \quad e^{-10}\frac{10^{15}}{15!} = 0.0347180696306841
$$




P(X=k), if X ~Bin(50, 0.8)
P(X=k), if X ~Poisson(40)
y
x
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0
10
20
30
40
50


P(X=k), if X ~Bin(50, 0.2)
P(X=k), if X ~Poisson(10)
y
x
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0
10
20
30
40
50




Since a binomial random variable has expected value $np = \lambda$ and variance $np(1-p) \approx \lambda$ (since $p$ is small), it would seem that both the expected value and the variance of a Poisson random variable would equal its parameter $\lambda$.

Theorem 3.21

Let $X$ be a Poisson random variable with parameter $\lambda$. Then

$$E[X] = Var[X] = \lambda.$$

Example 3.22

Company XYZ provides a warranty on a product that it produces. Each year, the number of warranty claims follows a Poisson distribution with mean $\lambda$. The probability that no warranty claims are received in any given year is 0.60.

Company XYZ purchases an insurance policy that will reduce its overall warranty claim payment costs. The insurance policy will pay nothing for the first warranty claim received and 5000 for each claim thereafter until the end of the year.

Calculate the expected amount of annual insurance policy payments to Company XYZ.

Solution

Let $N$ denote the number of warranty claims received. Then,

$$0.6 = P(N=0) = e^{-\lambda} \quad \Rightarrow \quad \lambda = -\ln(0.6) = 0.5108.$$

The expected yearly insurance payments are:

$$\begin{aligned}
&5000\left[P(N=2)+2P(N=3)+\cdots\right] \\
&=5000\left[P(N=1)+2P(N=2)+3P(N=3)+\cdots\right]-5000\left[P(N=1)+P(N=2)+P(N=3)+\cdots\right] \\
&=5000E[N]-5000\left[1-P(N=0)\right] \\
&=5000(0.5108)-5000(1-0.6) \\
&=554
\end{aligned}$$

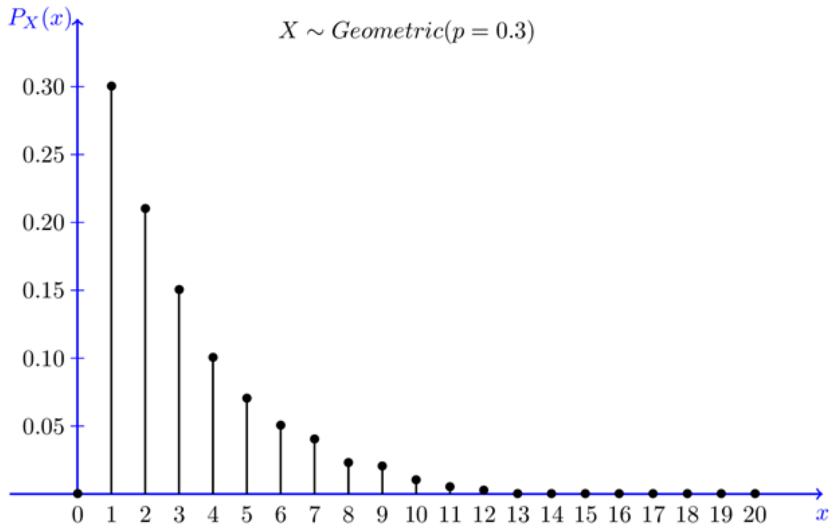
Instead of having a fixed number $n$ of trials to be chosen in advance, suppose we keep performing trials until we have achieved a given number of success.

Definition 3.23 (Geometric($p$))

Let $X$ denote the number of trial until the first success. $X$ has a **Geometric distribution with parameter** $p$:

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

where $p > 0$ is the probability of success.




Theorem 3.24

Let $X$ be a geometric random variable with parameter $p$. Then

$$P\left(X > n\right) = (1-p)^n, \quad E[X] = \frac{1}{p} \quad \text{and} \quad Var[X] = \frac{1-p}{p^2}.$$

In other words, if independent trials having a common probability $p$ of being successful are performed until the first success occurs, then the expected number of required trials equals $\frac{1}{p}$. For instance, the expected number of rolls of a fair die that it takes to obtain the value 1 is 6.

Example 3.25

Let $Y$ denote the number of failures before the first success. Then $Y = X - 1$ and

$$P(Y = k) = (1-p)^k p, \quad k = 0, 1, \cdots$$

Then $E[Y] = \frac{1}{p} - 1$ and $Var[Y] = \frac{1-p}{p^2}$.

Example 3.26

Automobile policies are separated into two groups: low-risk and high-risk. Rahul examines low-risk policies, continuing until a policy with a claim is found and then stopping. Toby follows the same procedure with high-risk policies. Each low-risk policy has a 10% probability of having a claim. Each high-risk policy has a 20% probability of having a claim. The claim statuses of polices are mutually independent.

Calculate the probability that Rahul examines fewer policies than Toby.

Solution

$P(\text{Rahul examines exactly } n \text{ policies}) = 0.1(0.9)^{n-1}.$

$P(\text{Toby examines more than } n \text{ policies}) = 0.8^n.$

$$P(\text{Rahul examines fewer policies than Toby}) = \sum_{n=1}^{\infty} 0.1(0.9)^{n-1} 0.8^n = 0.08 \sum_{n=1}^{\infty} 0.72^{n-1} = \frac{0.08}{1-0.72} = \frac{2}{7}$$

Theorem 3.27 (Memoryless property)

Let $X$ be a geometric random variable with parameter $p$. Then

$$P(X > m+n \mid X > n) = P(X > m) \quad \text{for all } m, n \geq 0.$$

Note that the geometric distribution satisfies the important property of being **memoryless**, meaning that if a success has not yet occurred at some given point, the probability distribution of the number of additional failures does not depend on the number of failures already observed. For instance, suppose a die is being rolled until a 1 is observed. If the additional information were provided that the die had already been rolled three times without a 1 being observed, the probability distribution of the number of further rolls is the same as it would be without the additional information.

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. In order for the $r$-th success to occur at the $n$-th trial, there must be $r-1$ successes in the first $n-1$ trials and the $n$-th trial must be a success. Let $X$ equal the number of trials to obtain $r$ successes. We have the following definition.

Definition 3.28 (Negative Binomial ($r$, $p$))

Any random variable $X$ whose probability mass function is given by

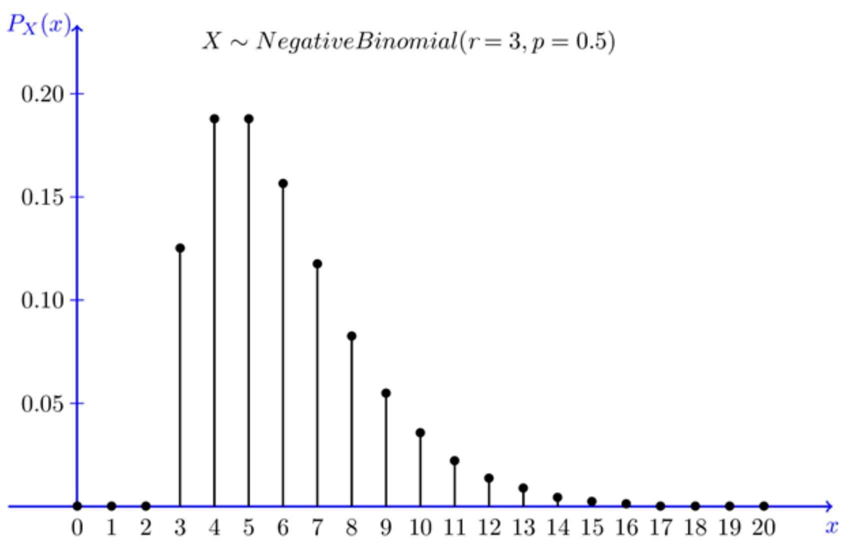
$$P(X=n)=\binom{n-1}{r-1}p^r(1-p)^{n-r}, \quad n=r,r+1,\cdots$$

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

In Binomial($n$, $p$), number of trials $n$ is fixed but number of successes is varied. In Negative Binomial ($r$, $p$), number of successes $r$ is fixed but number of trials is varied.

The number of trials required to obtain $r$ successes can be expressed as $X=Y_1+Y_2+\cdots+Y_r$, where $Y_1$ equals the number of trials required for the first success, $Y_2$ the number of additional trials after the first success until the second success occurs, $Y_3$ the number of additional trials until the third success, and so on. Because the trials are independent and all have the same probability of success, it follows that $Y_1,Y_2,\cdots,Y_r$ are all geometric random variables.

Note that Geometric($p$) = Negative Binomial (1, $p$).




Theorem 3.29

Let $X$ be a negative binomial random variable with parameters ($r$, $p$). Then

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

Let $Y$ be the number of failures before the $r$-th success. Then $Y = X - r$. The probability mass function of $Y$ is

$$P(Y = j) = \binom{j+r-1}{r-1} p^r (1-p)^j \qquad j = 0, 1, 2, \cdots$$

Its mean and variance are

$$E[Y] = \frac{r}{p} - r \quad \text{and} \quad Var[Y] = \frac{r(1-p)}{p^2}.$$