Probability and Statistics (ENM 503)

Michael A. Carchidi

November 10, 2015

Chapter 5 - Discrete Random Variables

The following notes are based on the textbook entitled: *A First Course in Probability* by Sheldon Ross (9th edition) and these notes can be viewed at

https://canvas.upenn.edu/

after you log in using your PennKey user name and Password.

1. Range Sets and Probability Mass Functions

In this chapter we shall discuss some probability terminology and concepts for discrete distributions. The nature of a random variable X (Discrete, Continuous, or Mixed) is based on the range of possible values for X. A *discrete* random variable is a variable X which is capable of being assigned a discrete set of numerical values $x_1, x_2, x_3, ..., x_n$, where n could be infinite. We call the set of possible value for X the *range set* of X and is denoted by

$$R_X = \{x_1, x_2, x_3, ..., x_n\}, \tag{1a}$$

where n could be infinite. A *probability function* defined on R_X, denoted by

$$P : R_X \to [0, 1],$$

is a rule that assigns a probability to each possible value of the random variable X,

$$\{p_1, p_2, p_3, ..., p_n\} \to \{x_1, x_2, x_3, ..., x_n\},$$

respectively, so that $p_k = P(X = x_k)$, which we shall also denote by $p(x_k)$. Since $P : R_X \to [0, 1]$ must be a probability function, it satisfy the probability axioms so that

$$0 \leq p(x) \leq 1 \tag{1b}$$

for each $x \in R_X$ and

$$\sum_{x \in R_X} p(x) = 1. \tag{1c}$$

Note that since we may assign $p(x) = 0$ for all $x \notin R_X$, we may assume that $p(x)$ is defined for all $-\infty < x < +\infty$ and by defining $p(x) = 0$ when $x \notin R_X$, we may write Equation (1c) as

$$\sum_{x=-\infty}^{+\infty} p(x) = \sum_{x} p(x) = 1. \tag{1d}$$

We call the collection of pairs $(x_k, p(x_k))$ for $k = 1, 2, 3, \ldots, n$, the *probability distribution* of the random variable X and $p(x_k)$ for $k = 1, 2, 3, ..., n$, is called the probability mass function (pmf) for X.

Example #1: A Discrete Random Variable

The number of customers arriving in a bank on a given day is a discrete random variable having the possible range of values given by $R_X = \{0, 1, 2, \ldots\}$ with $|R_X| = \infty$. ■

Example #2: A Discrete Random Variable

Suppose that a fair coin is flipped n times and the number of heads is recorded. If this number is denoted by X, then X is a discrete random variable with $R_X = \{0, 1, 2, ..., n\}$ with $|R_X| = n + 1$. ■

Example #3: A Discrete Random Variable

Suppose that n, m-sided die (with sides numbered 1 through m) are rolled and the sum is recorded. If this recorded number is denoted by X, then X is a discrete random variable with

$$R_X = \{n, n + 1, n + 2, ..., nm - 1, nm\}$$

with $|R_X| = nm - n + 1$. Specifically, if $n = 4$ and $m = 6$, then $R_X = \{4, 5, 6, ..., 24\}$, with $|R_X| = 24 - 4 + 1 = 21$. ■

Example #4: A Discrete Random Variable

Suppose that a fair coin is flipped until a head appears and the number of flips is recorded. If this number is denoted by X, then X is a discrete random variable with $R_X = \{1, 2, 3, ...\}$ with $|R_X| = \infty$. ■

Example #5: A Discrete Random Variable

Suppose that a box contains g "good" items and b "bad" items and suppose that n of these items is drawn from the box without replacement and the number of good items drawn is recorded. If this number is denoted by X, then X is a discrete random variable with

$$R_X = \{x_{\min}, x_{\min} + 1, x_{\min} + 2, ..., x_{\max}\},$$

where

$$x_{\min} = \max(n - b, 0) \qquad \text{and} \qquad x_{\max} = \min(g, n).$$

and $|R_X| = x_{\max} - x_{\min} + 1$. ■

Computing Probabilities

Let A be a subset of R_X and let x be a value of X. Then the probability that x is in A is given by

$$P(x \in A) = \sum_{x \in A} p(x) \tag{2}$$

where the sum is only over those x that are in A.

Example #6: A Discrete Random Variable

A total of m balls are to be randomly selected, without replacement, from an urn that contains $n \geq m$ balls numbered 1 through n. If X is the *largest numbered* ball selected, then X is a random variable with range set

$$R_X = \{m, m + 1, m + 2, ..., n\}.$$

Because each of the

$$\binom{n}{m}$$

possible selections of m of the n balls is equally likely, the event that $X = k$ occurs only if the one of the m balls selected is numbered k (which can occur only one way) and the remaining $m-1$ balls selected is numbered $k-1$ or lower and this can occur as many as

$$\binom{1}{1} \times \binom{k-1}{m-1} = \binom{k-1}{m-1}$$

ways. Therefore, the probability that $X = k$ is

$$p(k) = \frac{\binom{k-1}{m-1}}{\binom{n}{m}} \tag{3a}$$

for $k = m, m+1, m+2, ..., n$. ■

Continuous Random Variables

If the range set R_X of a random variable X is a continuous interval such as $a < x < b$, or a union of continuous intervals, then X is called a continuous random variable. For example, the lifetime of a person is a continuous random variable X, with $R_X = \{x|x > 0\}$. Note that it is not wrong to not have a upper bound on the value of X since we could always assign a probability of zero (or least very close to zero) to an age past a certain value. We shall leave the discussion of continuous random variables to the next chapter.

Mixed Random Variables

If the range set R_X of a random variable X is the union of discrete (D) and continuous (C) sets, then X is called a mixed random variable. For example

$$R_X = R_X^{\mathrm{D}} \cup R_X^{\mathrm{C}} = \{1, 2, 3\} \cup [4, 5]$$

is the set consisting of the integers 1, 2, and 3, as well as all the real numbers between 4 and 5, inclusive, and we shall leave the discussion of mixed random variables also to the next chapter.

2. Cumulative Distribution Functions

The cumulative distribution function (cdf) of a discrete random variable X, denoted by $F(x)$, equals the probability that the random variable X assumes a value less than or equal to x, *i.e.*,

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y). \tag{4a}$$

Note that $F(x)$ must then satisfy the following three properties: (i) F is a non-decreasing function of x, so that $F(a) \leq F(b)$ whenever $a \leq b$, (ii)

$$\lim_{x \to -\infty} F(x) = 0 \qquad \text{and} \qquad \lim_{x \to +\infty} F(x) = 1 \tag{4b}$$

and (iii)

$$P(a < X \leq b) = F(b) - F(a) \tag{4c}$$

for all $a < b$, so that

$$P(X = a) = F(a) - F(a-1). \tag{4d}$$

Note that a typical plot of $F(x)$ for a discrete random variable is "staircase" in structure with the height of each step at x equaling the pmf function value at x.

Example #7: A Staircase Graph

Suppose that X is a random variables with distribution function

$$\{(-1, 1/6), (2, 3/6), (4, 2/6)\}$$

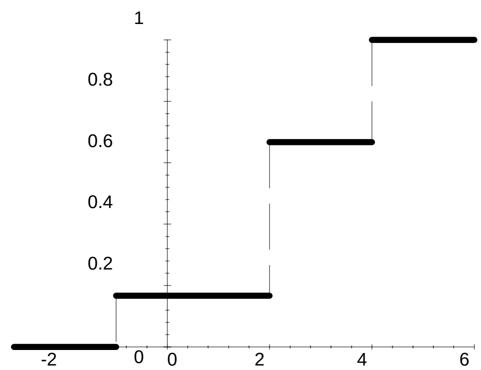
so that

$$p(x) = \begin{cases} 1/6, & \text{for} \quad x = -1 \\ 3/6, & \text{for} \quad x = +2 \\ 2/6, & \text{for} \quad x = +4 \end{cases}$$

and then

$$F(x) = \begin{cases} 0, & \text{for} \quad x < -1 \\ 1/6, & \text{for} \quad -1 \leq x < +2 \\ 4/6, & \text{for} \quad +2 \leq x < +4 \\ 1, & \text{for} \quad +4 \leq x \end{cases}$$

and a plot of F versus x is shown below.



Plot of F versus x for all x

Note that in this graph, we have $F(-1) = 1/6$, $F(+2) = 4/6$, and $F(+4) = 1$ showing that the left part of each step is closed and the right part is opened and the height of each step is $1/6$, $3/6$ and $2/6$, at $x = -1$, $x = +2$ and $x = +4$, respectively, and these are $p(-1)$, $p(+2)$ and $p(+4)$, respectively. ■

Example #8: Example #6 Revisited

A total of m balls are to be randomly selected, without replacement, from an urn that contains $n \geq m$ balls numbered 1 through n. If X is the *largest numbered* ball selected, then X is a random variable with range set

$$R_X = \{m, m+1, m+2, ..., n\}$$

and we saw in Example #6 that the probability that $X = k$ is

$$p(k) = \frac{\binom{k-1}{m-1}}{\binom{n}{m}}$$

for $k = m, m+1, m+2, ..., n$. If we now want to compute $F(x) = P(X \leq x)$, then

$$F(x) = P(X \leq x) = \sum_{j=m}^{x} \frac{\binom{j-1}{m-1}}{\binom{n}{m}} = \frac{1}{\binom{n}{m}} \sum_{j=m}^{x} \binom{j-1}{m-1}.$$

Using the binomial coefficient identity

$$\sum_{j=a}^{b} \binom{j-1}{a-1} = \binom{b}{a}, \tag{5}$$

the expression for $F(x)$ reduces to

$$F(x) = \frac{\binom{x}{m}}{\binom{n}{m}} \tag{3b}$$

for $x = m, m+1, m+2, ..., n$. $\blacksquare$

3. Expectation Value (Mean) of a Discrete Random Variable

If X is a discrete random variable with range set R_X, then the expectation value of X (if it exist) is defined by

$$E(X) = \sum_{x \in R_X} xp(x) \tag{6a}$$

and since $p(x) = 0$ for $x \notin R_X$, we may write this as

$$E(X) = \sum_{x=-\infty}^{+\infty} xp(x) = \sum_{x} xp(x), \tag{6b}$$

and it is easy to see that

$$E(\alpha X + \beta) = \sum_{x} (\alpha x + \beta)p(x) = \alpha \sum_{x} xp(x) + \beta \sum_{x} p(x)$$

which, via Equation (1d) and Equation (6a), reduces to

$$E(\alpha X + \beta) = \alpha E(X) + \beta \tag{6c}$$

for constants α and β. In fact, if X_p for $p = 1, 2, 3, ..., q$ are a set of discrete random variables (all with the same pmf p) and if α_p, for $p = 1, 2, 3, ..., q$ and β are constants, then it should be clear that

$$E(\alpha_1 X_1 + \cdots + \alpha_q X_q + \beta) = \alpha_1 E(X_1) + \cdots + \alpha_q E(X_q) + \beta. \tag{6d}$$

Example #9: $E(X)$ Need Not Exist for Every Discrete Distribution

Note that $E(X)$ need not exist for every discrete distribution. For example, consider the discrete random variable X having distribution function

$$(x, p(x)) = \left(x, \frac{6}{(\pi x)^2} \right)$$

for the range set $R_X = \{1, 2, 3, \ldots\}$. This is a valid distribution function since

$$0 \leq p(x) = \frac{6}{(\pi x)^2} \leq 1$$

for each $x \in \{1, 2, 3, \ldots\}$, and

$$\sum_{x \in R_X} p(x) = \sum_{x=1}^{\infty} \frac{6}{(\pi x)^2} = \frac{6}{\pi^2} \sum_{x=1}^{\infty} \frac{1}{x^2} = \frac{6}{\pi^2} \times \frac{\pi^2}{6} = 1.$$

If we try to compute $E(X)$, we find that

$$E(X) = \sum_{x=1}^{\infty} x p(x) = \sum_{x=1}^{\infty} x \frac{6}{(\pi x)^2} = \frac{6}{\pi^2} \sum_{x=1}^{\infty} \frac{1}{x} \to \infty$$

and hence does not converge. ■

The Mean of a Random Variable X

Note that when $E(X)$ does exist, then this is call the *mean* (often denoted by μ) of the random variable X.

Example #10: The Mean of X Need Not Be in R_X

Note that the mean of a random variable X need not be a possible value of X. For example, in the roll of a fair die, we have $R_X = \{1, 2, 3, 4, 5, 6\}$ with $p(x) = 1/6$ for $x = 1, 2, 3, 4, 5, 6$, and $p(x) = 0$ for all other values of x. Then

$$\mu = E(X) = \sum_{x} x p(x)$$

leads to

$$\mu = (1)\left(\frac{1}{6}\right) + (2)\left(\frac{1}{6}\right) + (3)\left(\frac{1}{6}\right) + (4)\left(\frac{1}{6}\right) + (5)\left(\frac{1}{6}\right) + (6)\left(\frac{1}{6}\right) = 3.5$$

and $3.5 \notin R_X$. ■

Example #11: Center of Mass From Physics

Suppose that n particles having masses m_1, m_2, ..., m_n are distributed on the x axis at positions x_1, x_2, ..., x_n, respectively. If $m = m_1 + m_2 + \cdots + m_n$ is the total mass of the n-particle system, its center-of-mass $x_{\rm cm}$ is defined as

$$x_{\rm cm} = \frac{m_1x_1 + m_2x_2 + \cdots + m_nx_n}{m_1 + m_2 + \cdots + m_n}$$

which we may write as

$$x_{\rm cm} = \left(\frac{m_1}{m}\right)x_1 + \left(\frac{m_2}{m}\right)x_2 + \cdots + \left(\frac{m_n}{m}\right)x_n.$$

By assigning probabilities m_j/m to x_j, we note that these probabilities are valid since $0 \leq m_j \leq 1$ for $j = 1, 2, 3, ..., n$ and

$$\sum_{j=1}^{n}\left(\frac{m_j}{m}\right) = 1.$$

Then $x_{\rm cm}$ can be viewed as the expected value of a random variable having possible values x_1, x_2, ..., x_n with probabilities m_1/m, m_2/m, ..., m_n/m, respectively. ■

Moments of X and the Variance of X

If n is a *nonnegative* integer, the quantity

$$E((X - c)^n) = \sum_{x \in R_X}(x - c)^n p(x) = \sum_{x}(x - c)^n p(x) \tag{7a}$$

is called the nth moment of X about the point c. Note that the zeroth moment is always equal to one and if $c = 0$, then $E(X^n)$ is called the nth moment of X (about zero implied) and if $c = \mu = E(X)$, $E((X - \mu)^n)$ is called the nth moment of X about its mean μ. The second moment of X about its mean μ is called the *variance* of X, and is given by

$$V(X) = E((X-\mu)^2) \tag{7b}$$

and this is always *non-negative*. Using Equation (6d), this can be reduced to

$$V(X) = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2$$

or

$$V(X) = E(X^2) - (E(X))^2. \tag{7c}$$

It should be noted that if α and β are constants, then

$$\begin{aligned} V(\alpha X + \beta) &= E((\alpha X + \beta)^2) - (E(\alpha X + \beta))^2 \\ &= E(\alpha^2 X^2 + 2\alpha\beta X + \beta^2) - (\alpha E(X) + \beta)^2 \\ &= \alpha^2 E(X^2) + 2\alpha\beta E(X) + \beta^2 - \alpha^2 (E(X))^2 - 2\alpha\beta E(X) - \beta^2 \end{aligned}$$

which reduces to

$$V(\alpha X + \beta) = \alpha^2 E(X^2) - \alpha^2 (E(X))^2 = \alpha^2 (E(X^2) - (E(X))^2)$$

and so

$$V(\alpha X + \beta) = \alpha^2 V(X). \tag{7d}$$

The *standard deviation* of X is given by

$$\sigma(X) = \sqrt{V(X)} \qquad \text{and so} \qquad \sigma(\alpha X + \beta) = |\alpha| \sigma(X) \tag{7e}$$

for constants α and β. Note that the mean of X is a measure of the *central tendency* of X and the variance of X (if it exists) is a measure of the *spread* or variation of possible values of X around this mean. Note that μ and σ are measured in the same physical units as X. A unitless measure of the spread of a distribution is called the *coefficient of variation* of X and is defined as

$$\text{cv} = \frac{\sigma}{\mu} = \frac{\sqrt{V(X)}}{E(X)} \tag{7f}$$

when $E(X) \neq 0$.

Modes

A mode of a discrete random variable X is a value (or values) of X that occurs (occur) most frequently. Note that a mode need not be unique and all modes are in R_X. In the simple example of tossing a fair die, all values of X in $R_X = \{1, 2, 3, 4, 5, 6\}$ are modes since they all occur with the same (maximum) probability of $1/6$.

4. Expectation Value of a Function of a Random Variable

Suppose that we are given a discrete random variable X along with its pmf and that we want to compute the expected value of some function of X, say $Y = H(X)$. One way to compute this is to note that since X is a random variable, so is $Y = H(X)$ and we could compute its pmf, call it $q(y)$ and then compute

$$E(Y) = \sum_{y} yq(y). \tag{8a}$$

Example #12: The Function of a Discrete Random Variable

Suppose that X is a random variable having known pmf $p(x)$ and suppose that Y is a random variable related to X through some known function $H(X)$, so that $Y = H(X)$. To determine the pmf (call it $q(y)$) of the random variable $Y = H(X)$, we simply note that if the range set of X is $R_X = \{x_1, x_2, x_3, ..., x_n\}$, then the range set of Y is

$$R_Y = \{H(x_1), H(x_2), H(x_3), ..., H(x_n)\}$$

which we write as

$$R_Y = \{y_1, y_2, y_3, ..., y_m\}$$

where $m \leq n$ with

$$q(y) = \sum_{x|y=H(x)} p(x). \tag{8b}$$

For a specific example, suppose that X is the sum of two fair dice with distribution $\{(x, p(x))\}$ given by

$$\left\{\left(2, \frac{1}{36}\right), \left(3, \frac{2}{36}\right), \left(4, \frac{3}{36}\right), \left(5, \frac{4}{36}\right), \left(6, \frac{5}{36}\right), \left(7, \frac{6}{36}\right),\right.$$

$$\left(8, \frac{5}{36}\right), \left(9, \frac{4}{36}\right), \left(10, \frac{3}{36}\right), \left(11, \frac{2}{36}\right), \left(12, \frac{1}{36}\right)\right\}$$

and suppose that $Y = (X-7)^2$. Then

$$\begin{aligned} R_Y &= \{(2-7)^2, (3-7)^2, (4-7)^2, (5-7)^2, (6-7)^2, \\ &\quad (7-7)^2, (8-7)^2, (9-7)^2, (10-7)^2, (11-7)^2, (12-7)^2\} \\ &= \{25, 16, 9, 4, 1, 0, 1, 4, 9, 16, 25\} \end{aligned}$$

or

$$R_Y = \{0, 1, 4, 9, 16, 25\}$$

with pmf $q(y)$ such that

$$q(0) = p(7) = \frac{6}{36} \quad , \quad q(1) = p(6) + p(8) = \frac{5}{36} + \frac{5}{36} = \frac{10}{36}$$

and

$$q(4) = p(5) + p(9) = \frac{4}{36} + \frac{4}{36} = \frac{8}{36}$$

and

$$q(9) = p(4) + p(10) = \frac{3}{36} + \frac{3}{36} = \frac{6}{36}$$

and

$$q(16) = p(3) + p(11) = \frac{2}{36} + \frac{2}{36} = \frac{4}{36}$$

and

$$q(25) = p(2) + p(12) = \frac{1}{36} + \frac{1}{36} = \frac{2}{36}.$$

Thus we find that $\{(y, q(y))\}$ is

$$\left\{\left(0, \frac{6}{36}\right), \left(1, \frac{10}{36}\right), \left(4, \frac{8}{36}\right), \left(9, \frac{6}{36}\right), \left(16, \frac{4}{36}\right), \left(25, \frac{2}{36}\right)\right\}.$$

The expected value of Y is then

$$E(Y) = 0\left(\frac{6}{36}\right) + 1\left(\frac{10}{36}\right) + 4\left(\frac{8}{36}\right) + 9\left(\frac{6}{36}\right) + 16\left(\frac{4}{36}\right) + 25\left(\frac{2}{36}\right)$$

or $E(Y) = 35/6 \simeq 5.83$. ■

We may also compute the expected value of $Y = H(X)$ using the fact that

$$E(Y) = \sum_y yq(y) = \sum_y y \sum_{x|y=H(x)} p(x) = \sum_{x|y=H(x)} \sum_y yp(x)$$

which reduces to just

$$E(Y) = \sum_x H(x)p(x). \tag{8c}$$

Example #13: The Mean of a Function of a Discrete Random Variable

Going back to Example #12, we have

$$E(Y) = \sum_x H(x)p(x) = \sum_x (x-7)^2 p(x)$$

which becomes

$$\begin{aligned} E(Y) &= (2-7)^2\left(\frac{1}{36}\right) + (3-7)^2\left(\frac{2}{36}\right) + (4-7)^2\left(\frac{3}{36}\right) + (5-7)^2\left(\frac{4}{36}\right) \\ &\quad + (6-7)^2\left(\frac{5}{36}\right) + (7-7)^2\left(\frac{6}{36}\right) + (8-7)^2\left(\frac{5}{36}\right) + (9-7)^2\left(\frac{4}{36}\right) \\ &\quad + (10-7)^2\left(\frac{3}{36}\right) + (11-7)^2\left(\frac{2}{36}\right) + (12-7)^2\left(\frac{1}{36}\right) \end{aligned}$$

and reduces to $E(Y) = 35/6$, as before. ■

Example #14: Maximizing Expected Profit

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

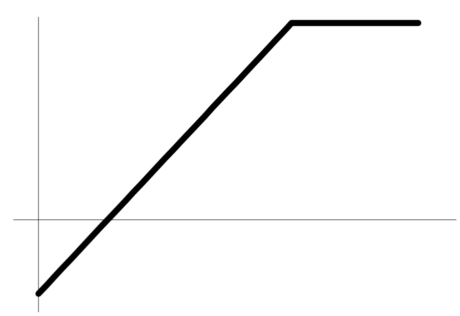
To solve this, let X denote the number of units ordered (*i.e.*, demanded by customers) and let s be the number of units that are stocked by the store. Then the profit, call it $H(s)$, can be expressed as

$$H(s) = \begin{cases} \alpha X - (s-X)\beta, & \text{when} \quad X \leq s \\ \\ s\alpha & \text{when} \quad X \geq s \end{cases}$$

or

$$H(s) = \begin{cases} (\alpha+\beta)X - s\beta, & \text{when} \quad X \leq s \\ \\ s\alpha & \text{when} \quad X \geq s \end{cases}.$$

A typical plot of $H(s)$ versus X is shown in the following figure.



Typical plot of $H(s)$ versus X

The expected profit can be computed as

$$\begin{aligned} E(H(s)) &= \sum_{x\geq 0} H(s)p(x) = \sum_{x=0}^{s} H(s)p(x) + \sum_{x\geq s+1} H(s)p(x) \\ &= \sum_{x=0}^{s}((\alpha+\beta)x - s\beta)p(x) + \sum_{x\geq s+1} s\alpha p(x) \\ &= (\alpha+\beta)\sum_{x=0}^{s} xp(x) - s\beta\sum_{x=0}^{s} p(x) + s\alpha \sum_{x\geq s+1} p(x) \\ &= (\alpha+\beta)\sum_{x=0}^{s} xp(x) - s\beta\sum_{x=0}^{s} p(x) + s\alpha \left(1 - \sum_{x=0}^{s} p(x)\right) \end{aligned}$$

or

$$E(H(s)) = s\alpha + (\alpha+\beta)\sum_{x=0}^{s} xp(x) - s(\alpha+\beta)\sum_{x=0}^{s} p(x)$$

or

$$E(H(s)) = s\alpha + (\alpha+\beta)\sum_{x=0}^{s}(x-s)p(x).$$

To determine the *optimum* value of s, let us investigate what happens to the profit when we increase the value of s by one unit and then look at

$$\Delta E(H(s)) = E(H(s+1)) - E(H(s)).$$

This leads to

$$E(H(s+1)) = (s+1)\alpha + (\alpha+\beta)\sum_{x=0}^{s+1}(x-s-1)p(x)$$

or

$$E(H(s+1)) = (s+1)\alpha + (\alpha+\beta)\sum_{x=0}^{s}(x-s-1)p(x)$$

since the last term in this sum is zero, and so, setting

$$\Delta E(H(s)) = E(H(s+1)) - E(H(s),$$

we have

$$\begin{aligned}\Delta E(H(s)) &= (s+1)\alpha + (\alpha+\beta)\sum_{x=0}^{s}(x-s-1)p(x)\\ &\quad -s\alpha - (\alpha+\beta)\sum_{x=0}^{s}(x-s)p(x)\end{aligned}$$

which reduces to

$$\Delta E(H(s)) = \alpha + (\alpha+\beta)\sum_{x=0}^{s}(x-s-1-x+s)p(x)$$

or

$$\Delta E(H(s)) = \alpha - (\alpha+\beta)\sum_{x=0}^{s} p(x) = \alpha - (\alpha+\beta)F(s)$$

where $F(s)$ is the cdf of X evaluated at $X = s$. Now when $\Delta E(H(s)) > 0$, then increasing the value of s to $s+1$ will increase profits while when $\Delta E(H(s)) < 0$, increasing the value of s to $s+1$ will decrease profits. Therefore, profits will increase as long as

$$\Delta E(H(s)) = \alpha - (\alpha + \beta)F(s) \geq 0.$$

Since $F(s)$ is an increasing function of s, this says that we should determine the *largest* value of s for which

$$F(s) \leq \frac{\alpha}{\alpha + \beta}$$

and this will result in a maximum expected profit. ■

5. A Survey of Important Discrete Distributions

The following six discrete distributions stand out in applications and so each will be discussed in some detail. These are the: (i) Bernoulli, (ii) Binomial, (iii) Geometric, (iv) Pascal, (v) Hypergeometric, (vi) Uniform, and (vii) Poisson distributions. We begin with the Bernoulli distribution since the binomial, geometric and Pascal distributions are generated by this one. Even the hypergeometric come from a Bernoulli-type idea.

5.1 The Bernoulli Distribution With Parameter p

Consider an experiment of a single trial, which can be marked as either a "success" or as a "failure". Let the random variable $X = s$ if the experiment resulted in a success while $X = f$ if the experiment resulted in a failure and the words "success" and "failure" are not meant to imply that one is good (success) and the other is bad (failure). For example, the experiment might be the simply flip of a coin and we could arbitrarily call heads a success and tail a failure, or the other way around.

An experiment in which there are only two outcomes is called a *Bernoulli trial* and the resulting distribution is called a *Bernoulli distribution*. It has a range set consisting of just the two point $X = s$ (which is a numerical value associated with a *success*) and $X = f$ (which is a numerical value associated with a *failure*) so that $R_X = \{s, f\}$. The pmf of this random variable is defined by

$$P(X = x) = \begin{cases} 1 - p, & \text{for} \quad x = f \\ \\ p, & \text{for} \quad x = s \end{cases} \tag{9a}$$

for $0 \leq p \leq 1$, where p is called the probability of a success and $q = 1 - p$ is the probability of a failure. The first and second moments are

$$E(X) = sP(X = s) + fP(X = f)$$

or

$$E(X) = sp + f(1 - p) \tag{9b}$$

and

$$E(X^2) = sP(X = s) + f^2 P(X = f) = s^2 p + f^2(1 - p)$$

and so the variance is given by

$$\sigma^2 \equiv V(X) = E(X^2) - \mu^2 = s^2 p + f^2(1 - p) - (sp + f(1 - p))^2,$$

which reduces to

$$V(X) = p(1 - p)(s - f)^2. \tag{9c}$$

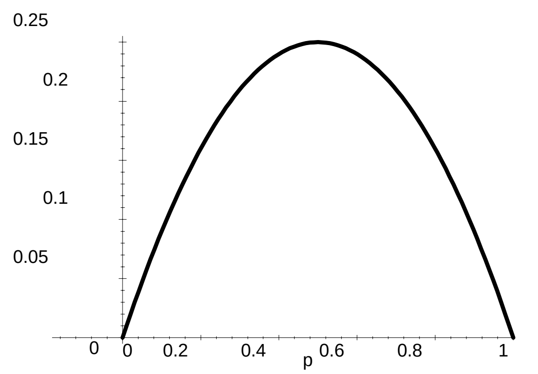
The choice for the value of s and f is completely arbitrary and most people like to choose $f = 0$ (for failure) and $s = 1$ (for success). In this case, we call this the "standard" Bernoulli distribution and the above equations reduce to

$$P(X = x) = \begin{cases} 1 - p, & \text{for} \quad x = 0 \\ \\ p, & \text{for} \quad x = 1 \end{cases} \tag{10a}$$

and

$$E(X) = p \qquad \text{and} \qquad V(X) = p(1 - p). \tag{10b}$$

A plot of this variance versus p below shows that the *maximum* variance in the Bernoulli distribution is given by $1/4$ and occurs when $p = 1/2$.



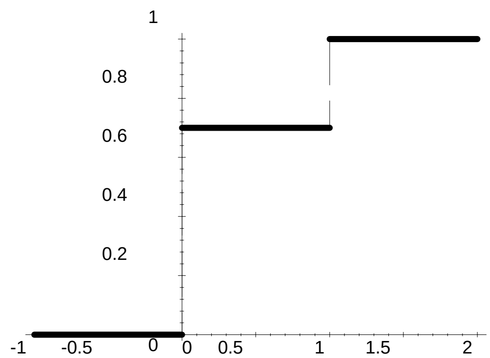
Plot of σ^2 versus p (Bernoulli)

This is reasonable since p close to zero would result in $X = 0$ most of the time, thereby showing little variation in the value of X, while p close to one would result in $X = 1$ most of the time, also showing little variation in the value of X. When $p = 1/2$, then $X = 0$ and $X = 1$ are equally likely to occur which should result in the largest variation in X.

The cumulative distribution function for the standard Bernoulli distribution is given by

$$F(x) = P(X \leq x) = \begin{cases} 0, & \text{for } \; x < 0 \\ 1 - p, & \text{for } \; 0 \leq x < 1 \\ 1, & \text{for } \; 1 \leq x \end{cases} \tag{11}$$

which looks like two stairs of heights $1-p$ and p, respectively, as illustrated below for $p = 0.3$.



Plot of F versus x for $p = 0.3$

Note that $F(-\infty) = 0$, $F(+\infty) = 1$ and $F(x_1) \leq F(x_2)$ for $x_1 \leq x_2$. Even through the pmf $P(x)$ is non-zero for only the two values $x = 0$ and $x = 1$, it is still defined for all x and so if its cdf, $F(x)$.

The Bernoulli distribution is quite simple and itself does not have many applications. The real value of the Bernoulli distribution is that it can be used to *generate other discrete distributions* that do have many applications. We illustrate the first one as the binomial distribution with parameters p and n.

5.2 The Binomial Distribution With Parameters n and p

The random variable X that denotes the number of successes in n *independent and identical* Bernoulli trials has a binomial distribution with parameter p, where $0 \leq p \leq 1$. To determine the pmf of this distribution, we note that there are

$$\binom{n}{x}$$

different ways to choose x successes (and hence $n - x$ failures) out of n trials since order does not matter here, and because of the independence of these trials, each one has a probability of $p^x(1-p)^{n-x}$. Therefore, the pmf of this distribution is given by

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \tag{12a}$$

for $0 \leq p \leq 1$ and for $x = 0, 1, 2, \ldots, n$.

Since we may write

$$X = X_1 + X_2 + X_3 + \cdots + X_n$$

with each of $X_1, X_2, X_3, \ldots, X_n$ being a Bernoulli distribution and all independent, we must have

$$E(X) = E(X_1 + X_2 + X_3 + \cdots + X_n) = \sum_{k=1}^{n} E(X_k) = \sum_{k=1}^{n} p$$

or

$$E(X) = np. \tag{12b}$$

In general, we shall show that

$$V(X_1 + X_2 + X_3 + \cdots + X_n) \neq V(X_1) + V(X_2) + V(X_3) + \cdots + V(X_n)$$

unless all of the X_k's are independent. For now, let us just assume that

$$V(X_1 + X_2 + X_3 + \cdots + X_n) = V(X_1) + V(X_2) + V(X_3) + \cdots + V(X_n)$$

as long as all the X_k's are independent. Applying this to the Binomial distribution, we find that

$$V(X) = V(X_1 + X_2 + X_3 + \cdots + X_n) = \sum_{k=1}^{n} V(X_k) = \sum_{k=1}^{n} p(1-p)$$

or

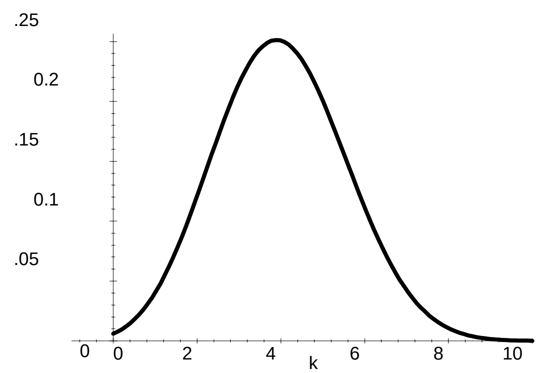
$$V(X) = np(1-p). \tag{12c}$$

A plot of V versus p shows a maximum or $V_{\max} = n/4$ at $p = 1/2$ and zero variance at the endpoint $p = 0$ (which gives all n failures) and at $p = 1$ (which gives all n successes), and these results make sense.

The cumulative distribution function for the binomial distribution is given by

$$F(x) = P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k} \tag{12d}$$

for $x = 0, 1, 2, 3, \ldots, n$. This *cannot be simplified* in any closed form and so quite often, these results are tabulated for different values of p and n or they must be computed using a computer or calculator A typical plot of a Binomial distribution (drawn as a continuous curve) using $n = 10$ and $p = 0.4$ is shown in the following figure.



Plot of the Binomial Distribution
($n = 10$ and $p = 0.4$)

The mean of the Binomial distribution (np), if it is an integer, is also the mode, which is the point in which the pmf is the largest. More generally, we have

$$x_{\text{mode}} = \lfloor np \rfloor.$$

Example #14 - A Binomial Distribution

A production process manufactures computer chips which are, on average, 2% nonconforming. Every day a random sample of size 50 is taken from the production process. If the sample contains more than two nonconforming chips, the production process is stopped. Determine the probability that the production process is stopped by this sampling scheme. Considering the sampling production process as $n = 50$ Bernoulli trials, each with $p = 0.02$, the total number of nonconforming chips in the sample X, would have a binomial distribution

$$P(X = x) = \binom{50}{x} (0.02)^x (0.98)^{50-x}$$

for $x = 0, 1, 2, \ldots, 50$, and zero otherwise. Then

$$\begin{aligned} P(X &> 2) = 1 - P(X \leq 2) \\ &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \end{aligned}$$

$$
\begin{aligned}
&= 1-\binom{50}{0}(0.02)^0(0.98)^{50-0}-\binom{50}{1}(0.02)^1(0.98)^{50-1} \\
&\quad -\binom{50}{2}(0.02)^2(0.98)^{50-2} \\
&= 1-0.364-0.372-0.186 \\
&= 0.078
\end{aligned}
$$

and so the probability that the production process is stopped by this sampling scheme is approximately 7.8%. Note that the mean number of nonconforming chips is given by

$$E(X) = np = (50)(0.02) = 1$$

and the variance is $V(X) = np(1-p) = (50)(0.02)(0.98) = 0.98$. ■

Example #15 - Is This a Fair Game

The game chuck-a-luck is quite popular at many carnivals and gambling casinos. A player bets on one of the numbers 1, 2, 3, 4, 5 and 6. Three fair dice are then rolled, and if the number bet by the player appears k times ($k = 1, 2, 3$), then the player wins k dollars and if the number bet by the player does not appear on any of the dice, then the player loses 1 dollar (which is the cost for playing the game). Determine if this game is fair and if not, what should be the charge for playing the game to make it fair?

To solve this, we assume that the dice are fair and act independently of one another. Then the number of times that the number bet appears is a binomial random variable with parameters $n = 3$ and $p = 1/6$. Hence, letting X denote the *player's winnings* in the game, we find that

$$R_X = \{-1, 1, 2, 3\}$$

with

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

and

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

and

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

and

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

The expected winnings, which is the player's winnings after playing the game many times, is then

$$E(X) = (-1)\left(\frac{125}{216}\right) + 1\left(\frac{75}{216}\right) + 2\left(\frac{15}{216}\right) + 3\left(\frac{1}{216}\right) = -\frac{17}{216}$$

or $E(X) \simeq -0.08 < 0$, showing that the game is slightly unfair to the player.

A fair game is one in which $E(X) = 0$. Letting c be the cost for playing this game, it will be fair when

$$E(X) = (-c)\left(\frac{125}{216}\right) + 1\left(\frac{75}{216}\right) + 2\left(\frac{15}{216}\right) + 3\left(\frac{1}{216}\right) = 0$$

which leads to

$$E(X) = \frac{1}{2} - c\left(\frac{125}{216}\right) = 0 \qquad \text{or} \qquad c = \frac{108}{125} = 0.864$$

which says that the game is fair when it cost 86.4¢ to play. Note that the answer here is not just $1.00 - 0.08 = 92$¢.■

5.3 The Geometric Distribution With Parameter p

The geometric distribution is related to a sequence of independent and identical Bernoulli trials; the random variable of interest, X is defined to be the number of trials *to achieve the first success*. This distribution with parameter $0 \leq p \leq 1$ is then given by

$$P(X = x) = p(1-p)^{x-1} \tag{13a}$$

for $x = 1, 2, 3, \ldots$, and zero otherwise. This is because

$$P(X = x) = P(F \cap F \cap F \cap \cdots F \cap F \cap S)$$

with $(x-1)$ failures (F) and one success (S). The first moment is given by

$$E(X) = \sum_{x=1}^{\infty} xp(1-p)^{x-1} = p\sum_{x=1}^{\infty} x(1-p)^{x-1}$$

To simplify this, let use look at some special sums.

Some Special Sums

Using the fact that

$$\sum_{x=1}^{\infty} r^x = \frac{r}{1-r} \tag{14a}$$

for $|r| < 1$, we find that

$$\frac{d}{dr}\sum_{x=1}^{\infty} r^x = \frac{d}{dr}\left(\frac{r}{1-r}\right) \qquad \text{or} \qquad \sum_{x=1}^{\infty} xr^{x-1} = \frac{1}{(1-r)^2}$$

or

$$\sum_{x=1}^{\infty} xr^x = \frac{r}{(1-r)^2} \tag{14b}$$

for $|r| < 1$, and

$$\frac{d}{dr}\sum_{x=1}^{\infty} xr^x = \frac{d}{dr}\left(\frac{r}{(1-r)^2}\right) \qquad \text{or} \qquad \sum_{x=1}^{\infty} x^2r^{x-1} = \frac{1+r}{(1-r)^3}$$

or

$$\sum_{x=1}^{\infty} x^2r^x = \frac{(1+r)r}{(1-r)^3} \tag{14c}$$

for $|r| < 1$.

Using these sums, we now find that

$$E(X) = p\sum_{x=1}^{\infty} x(1-p)^{x-1} = p\left(\frac{1}{(1-(1-p))^2}\right)$$

or

$$E(X) = \frac{1}{p} \tag{13b}$$

and then the second moment is

$$E(X^2) = \sum_{x=1}^{\infty} x^2 p(1-p)^{x-1} = p \sum_{x=1}^{\infty} x^2 (1-p)^{x-1} = p \frac{(1+(1-p))}{(1-(1-p))^3}$$

or simply

$$E(X^2) = \frac{2-p}{p^2}.$$

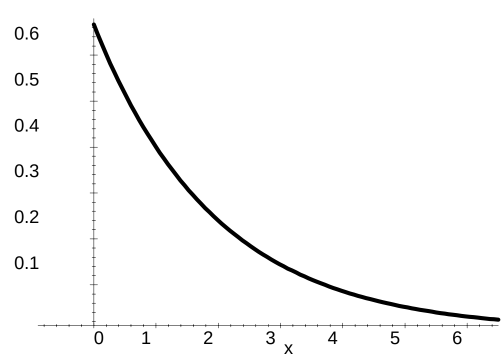
The variance is then given by

$$V(X) = E(X^2) - (E(X))^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

or

$$V(X) = \frac{1-p}{p^2}. \tag{13c}$$

Note that $V = 0$ when $p = 1$, which is reasonable since for $p = 1$, the first success will always occur on the first trial. However $V \to \infty$ as $p \to 0$, which is also reasonable since the number of trials needed for the first success is then "all over the place". A typical plot of a Geometric distribution using $p = 0.4$ is shown in the following figure.



Plot of the Geometric Distribution
($p = 0.4$)

Example #16 - The Geometric Distribution

A production process manufactures computer chips which are, on average, 2% nonconforming. Every day a random sample of size 50 is taken from the production process. If the sample contains more than two nonconforming chips, the production process is stopped. Let us determine the probability that the first nonconforming chip is the fifth one inspected. This is given by

$$P(X = 5) = (0.02)(1 - 0.02)^4 = 0.018.$$

or $P = 1.8\%$. ■

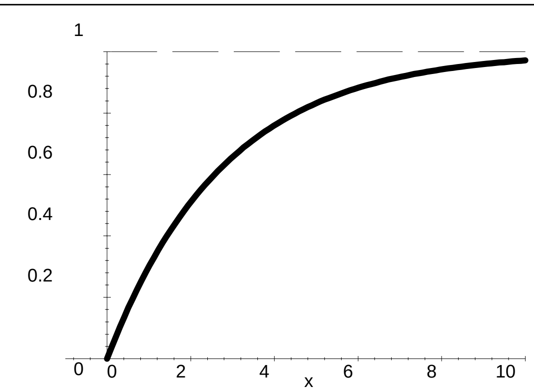
The cumulative distribution function for the geometric distribution is given by

$$F(x) = P(X \leq x) = \sum_{k=1}^{x} p(1-p)^{k-1} = \sum_{k=0}^{x-1} p(1-p)^k = p\left\{\frac{1-(1-p)^x}{1-(1-p)}\right\}$$

which reduces to

$$F(x) = 1 - (1-p)^x \tag{13d}$$

for $x = 1, 2, 3, \ldots$. A plot of this (shown as a continuous curve and for $p = 0.3$) is below.



Plot of F versus x for $p = 0.3$

It shows $F(x_1) \leq F(x_2)$ when $x_1 \leq x_2$, as it must.

Example #17: Shoot-Look-Shoot

One method used to model the killing of a target in warfare is to used the shoot-look-shoot policy where a salvo of weapons is fired at a target, the target is examined (looked at) to see if it is dead. If the target is not dead, then another salvo is fired and so on until the target is dead. Suppose the probability that a target is killed in a single salvo is p, then the probability that x salvos are needed to kill the target is given by

$$P(X = x) = p(1-p)^{x-1}$$

for $x = 1, 2, 3, \ldots$, assuming that killing the target is a success and not killing the target is a failure. This, of course assumes that the target is not weaken by previous shots, which is not completely realistic. ■

The Memoryless Property of the Geometric Distribution

The geometric distribution has the *memoryless* property which states that

$$P(X > s+t \mid X > s) = P(X > t) \tag{14}$$

for all $0 \leq s$ and $0 \leq t$. The proof of this follows by writing

$$P(X > s+t \mid X > s) = \frac{P(X > s+t \ \cap \ X > s)}{P(X > s)} = \frac{P(X > s+t\ )}{P(X > s)}$$

or

$$P(X > s+t \mid X > s) = \frac{1 - P(X \leq s+t\ )}{1 - P(X \leq s)} = \frac{1 - F(s+t)}{1 - F(s)}.$$

But for the geometric distribution, we have

$$F(x) = 1 - (1-p)^x \qquad \text{so that} \qquad 1 - F(x) = (1-p)^x,$$

and so

$$P(X > s+t \mid X > s) = \frac{(1-p)^{s+t}}{(1-p)^s} = (1-p)^t = 1 - F(t)$$

or simply

$$P(X > s+t \mid X > s) = P(X > t)$$

and so this part of the proof is complete.

Suppose now that a discrete distribution satisfies the memoryless property so that

$$P(X > s + t \mid X > s) = P(X > t)$$

for all $0 \leq s$ and $0 \leq t$. Then

$$\frac{P(X > s + t \ \cap \ X > s)}{P(X > s)} = P(X > t) \qquad \text{or} \qquad \frac{P(X > s + t\ )}{P(X > s)} = P(X > t)$$

or

$$P(X > s + t\ ) = P(X > t)P(X > s).$$

This says that

$$1 - F(s + t) = (1 - F(t))(1 - F(s)) = 1 - F(t) - F(s) + F(t)F(s)$$

or

$$F(t) + F(s) = F(s + t) + F(t)F(s).$$

Replacing t by $t - 1 \geq 0$, we have

$$F(t - 1) + F(s) = F(s + t - 1) + F(t - 1)F(s)$$

and subtracting these we get

$$F(t) - F(t - 1) = F(s + t) - F(s + t - 1) + (F(t) - F(t - 1))F(s)$$

or

$$P(t) = P(s + t) + P(t)F(s) \qquad \text{or} \qquad P(s + t) = P(t)(1 - F(s)).$$

Setting $s = 1$ and $t = x$, we have

$$P(x + 1) = P(x)(1 - F(1)) = (1 - p)P(x) \qquad \text{for} \qquad p \equiv F(1)$$

and for $x = 1, 2, 3, ...$ This says that

$$P(2) = (1 - p)P(1) \quad , \quad P(3) = (1 - p)P(2) = (1 - p)^2 P(1)$$

and so on to yield

$$P(x) = (1 - p)^x P(1)$$

for $x = 1, 2, 3, ...$ Since

$$\sum_{x=1}^{\infty} P(x) = 1 \quad \text{we have} \quad \sum_{x=1}^{\infty}(1-p)^x P(1) = P(1)\sum_{x=1}^{\infty}(1-p)^x = \frac{(1-p)P(1)}{p} = 1$$

yielding $P(1) = p/(1-p)$, and hence

$$P(x) = (1-p)^{x-1}p$$

for $x = 1, 2, 3, ...,$ which is a geometric distribution. Thus we see that *the geometric distribution is the only discrete distribution which has the memoryless property*

$$P(X > s + t \mid X > s) = P(X > t)$$

for all $0 \leq s$ and $0 \leq t$.

5.4 The Pascal Distribution With Parameters r and p

As seen earlier, the geometric distribution is related to a sequence of independent and identical Bernoulli trials; the random variable of interest, X, is defined to be the number of trials to achieve a success for the *first* time. An obvious generalization to a geometric distribution arises if we ask for the number of trials to achieve a success for the rth time with $r = 1, 2, 3, \ldots$. Such a distribution is called the Pascal distribution.

The Pascal distribution is related to a sequence of Bernoulli trials; the random variable of interest, X, is defined to be the number of trials to achieve a success for the rth time with $r = 1, 2, 3, \ldots$.. Now $X = x$ if and only if a success occurs on the rth trial and $r-1$ successes occurred in the previous $x-1$ trials. The probability that $r-1$ successes occurred in the previous $x-1$ trials is binomial and has probability

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

and multiplying this by the probability of the rth success (p), we find that

$$P(X = x) = \binom{x-1}{r-1}p^r(1-p)^{x-r} \tag{15a}$$

gives the probability that it takes x trials to achieve a success for the rth time with $r = 1, 2, 3, \ldots$ and $x = r, r+1, r+2, \ldots.$ This is called the Pascal distribution with parameters $r = 1, 2, 3, \ldots$, and $0 \leq p \leq 1$ is given by for $x = r, r+1, r+2, r+3, \ldots$, and zero otherwise. For $r = 1$, we (of course) get the geometric distribution.

$$P(X = x) = \binom{x-1}{1-1} p^1 (1-p)^{x-1} = p(1-p)^{x-1}$$

for $x = 1, 2, 3, \ldots.$

The Negative Binomial Distribution - Optional

The Pascal distribution is also called the negative binomial distribution for the following reason. Using the binomial coefficient definition

$$\binom{a}{b} = \frac{a(a-1)(a-2)\cdots(a-(b-1))}{b!}, \tag{16a}$$

we find that

$$\begin{aligned}
\binom{-r}{k} &= \frac{(-r)(-r-1)(-r-2)\cdots(-r-(k-1))}{k!} \\
&= (-1)^k \frac{(r)(r+1)(r+2)\cdots(r+(k-1))}{k!} \\
&= (-1)^k \frac{(r+k-1)!}{k!(r-1)!} = (-1)^k \frac{(k+r-1)!}{k!(r-1)!}
\end{aligned}$$

which says that

$$\binom{k+r-1}{r-1} = (-1)^k \binom{-r}{k}. \tag{16b}$$

Then

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} = \binom{(x-r)+r-1}{r-1} p^r (1-p)^{x-r}$$

or simply

$$P(X = x) = \binom{-r}{x-r} (-1)^{x-r} p^r (1-p)^{x-r} \tag{15b}$$

for $x = r, r+1, r+2, r+3, \ldots$, and zero otherwise. Using this expression we note that

$$\sum_{x=r}^{\infty} P(X=x) = \sum_{x=r}^{\infty} \binom{-r}{x-r} p^r (p-1)^{x-r} = p^r \sum_{x=0}^{\infty} \binom{-r}{x} (p-1)^x$$

which reduces

$$\sum_{x=r}^{\infty} P(X=x) = p^r (1+p-1)^{-r} = 1$$

as it should. In addition, the first and second moments can now be computed using

$$\begin{aligned} E(X-r) &= \sum_{x=r}^{\infty} (x-r) P(X=x) = \sum_{x=r}^{\infty} (x-r) \binom{-r}{x-r} p^r (p-1)^{x-r} \\ &= p^r \sum_{x=0}^{\infty} x \binom{-r}{x} (p-1)^x = p^r \sum_{x=1}^{\infty} (-r) \binom{-r-1}{x-1} (p-1)^x \\ &= -rp^r (p-1) \sum_{x=1}^{\infty} \binom{-r-1}{x-1} (p-1)^{x-1} \end{aligned}$$

resulting in

$$E(X) - E(r) = -rp^r (p-1)(1+p-1)^{-r-1} = r\left(\frac{1-p}{p}\right)$$

or simply

$$E(X) = r + r\left(\frac{1-p}{p}\right) = \frac{r}{p}. \tag{15b}$$

In addition we can look at $E((X-r)(X-r-1))$ and the student should show that

$$E(X^2) = r\left(\frac{1-p+r}{p^2}\right) \tag{15c}$$

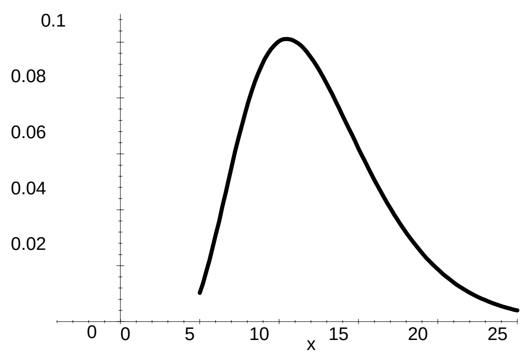
resulting in

$$V(X) = E(X^2) - (E(X))^2 = r\left(\frac{1-p+r}{p^2}\right) - \frac{r^2}{p^2}$$

or

$$V(X) = \frac{r(1-p)}{p^2}. \tag{15d}$$

A typical plot of a Pascal distribution using $r = 5$ and $p = 0.4$ is shown in the following figure.



Plot of the Pascal Distribution
($r = 5$ and $p = 0.4$)

showing that it has a mode. To estimate this node we look at

$$\begin{aligned}
P(x) - P(x-1) &= \binom{x-1}{r-1} p^r (1-p)^{x-r} - \binom{x-1-1}{r-1} p^r (1-p)^{x-1-r} \\
&= \frac{(x-1)!}{(r-1)!(x-r)!} p^r (1-p)^{x-r} - \frac{(x-2)!}{(r-1)!(x-1-r)!} p^r (1-p)^{x-1-r} \\
&= \left(\frac{(x-1)(1-p)}{(x-r)} - 1 \right) \frac{p^r (1-p)^{x-1-r} (x-2)!}{(r-1)!(x-1-r)!}
\end{aligned}$$

which is increasing as long as

$$\frac{(x-1)(1-p)}{(x-r)} - 1 > 0$$

which says that

$$x < 1 + \frac{r-1}{p}.$$

The mode of the distribution is then

$$x_{\text{mode}} = \left\lfloor 1 + \frac{r-1}{p} \right\rfloor.$$

Example #18: Achieving r Successes Before m Failures

If independent trials, each resulting in a success with probability p, are performed, let us compute the probability that r successes occur before m failures.

To answer this, we note that r successes will occur before m failures *if and only if* the rth success occurs no later than the $r+m-1$ trial. This follows because if the rth success occurs before or at the $r+m-1$ trial, then it must have occurred before the mth failure, and conversely. Hence the desired probability is

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

which cannot be simplified any further. ■

Example #19: The Banach Matchbox Problem

At all times, a pipe-smoking mathematician carries 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 his matchboxes is empty. If it is assumed that both matchboxes initially contain N matches, let us compute the probability that there are exactly k matches (for $k = 0, 1, 2, ..., N$) left in the other box, when it is discovered that one of the boxes is empty.

To answer this, let E denote 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 this time. Now this event will occur *if and only if* the $(N+1)$st choice of the right-hand matchbox is made at the $(N+1+N-k)$th trial since this will result in N matches being removed from the right-hand pocket and $N-k$ matches being removed from the left-hand pocket. This is Pascal with $p = 1/2$, $r = N+1$ and $x = 2N-k+1$ and hence occurs with probability

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

Since there is an equal probability that it is the left-hand box that is first discovered to be empty and there are k matches in the right-hand pocket at that time, the desired probability must be twice this result, which leads to

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

as the desired probability for $k = 0, 1, 2, ..., N$. ■

A Relation Between The Pascal and Binomial Distributions

It is interesting to note that if X has a binomial distribution with parameters n and p so that X is the number of successes in n Bernoulli trials with the probability of a success being p, and if Y has a Pascal distribution with parameters r and p so that Y is the number of Bernoulli trials required to obtain r successes with the probability of a success being p, then the event $Y > n$ is *equivalent* to the event $X < r$. *This follows because if the number of Bernoulli trials required to obtain r successes is larger than n, then the number of successes in n Bernoulli trials must have been less than r, and conversely.* Therefore these events must occur with the same probability so that

$$P(Y > n) = P(X < r) \qquad \text{and hence also} \qquad P(Y \leq n) = P(X \geq r)$$

since

$$P(Y \leq n) = 1 - P(Y > n) \qquad \text{and} \qquad P(X \geq r) = 1 - P(X < r).$$

This allows one to used binomial-distribution probabilities to evaluate probabilities associated with the Pascal distribution.

Example #20

If we want to evaluate the probability that more than 10 repetitions are required to obtain the 3rd success with $p = 0.2$, then

$$P(Y > 10) = P(X < 3) = \sum_{k=0}^{2} \binom{10}{k} (0.2)^k (0.8)^{10-k} = 0.678$$

or $P(Y > 10) = 67.8\%$. ■

5.5 The Hypergeometric Distribution With Parameters g, b and n

Suppose we have a lot of $g + b$ items, g of which are good and b of which are bad, and suppose we choose at random $n \leq g + b$ items from this lot, *without replacement and with order unimportant.* If we let X be the number of good items chosen, then $X = x$ if and only if we obtain precisely x good items from the g good items and precisely $n - x$ bad items from the b bad items. Since order is unimportant, this can occur in

$$\binom{g}{x} \times \binom{b}{n-x}$$

possible ways and since there are

$$\binom{g+b}{n}$$

total ways to choose the n items from the $g + b$ items, (all events being equally likely) we must find that

$$P(X = x) = \frac{\binom{g}{x}\binom{b}{n-x}}{\binom{g+b}{n}} \tag{17a}$$

for $x = x_{\min}, x_{\min} + 1, x_{\min} + 2, \ldots, x_{\max}$, and zero otherwise, where

$$x_{\min} = \max(n - b, 0) \qquad \text{and} \qquad x_{\max} = \min(g, n).$$

Using the binomial coefficient identity

$$\sum_{x=x_{\min}}^{x_{\max}} \binom{g}{x}\binom{b}{n-x} = \binom{g+b}{n}$$

it is easy to show that

$$\mu \equiv E(X) = n\left(\frac{g}{g+b}\right) \tag{17b}$$

and

$$\sigma^2 \equiv V(X) = n\left(\frac{g}{g+b}\right)\left(\frac{b}{g+b}\right)\left(\frac{g+b-n}{g+b-1}\right) \tag{17c}$$

give the mean and variance of the distribution. Note that V is at a maximum at

$$n = \frac{g+b}{2}$$

when $g+b$ is even.

The Mode of the Hypergeometric Distribution

Note that the mode of the hypergeometric distribution can be estimated using the fact that

$$\frac{P(X=x)}{P(X=x-1)} = \frac{\binom{g}{x}\binom{b}{n-x}/\binom{g+b}{n}}{\binom{g}{x-1}\binom{b}{n-(x-1)}/\binom{g+b}{n}} = \frac{\binom{g}{x}\binom{b}{n-x}}{\binom{g}{x-1}\binom{b}{n-x+1}}$$

which reduces to

$$\frac{P(X=x)}{P(X=x-1)} = \frac{g!b!(x-1)!(g-x+1)!(n-x+1)!(b-n+x-1)!}{x!(g-x)!(n-x)!(b-n+x)!g!b!}$$

or

$$\frac{P(X=x)}{P(X=x-1)} = \frac{(g-x+1)(n-x+1)}{x(b-n+x)}$$

and this is greater than 1 (showing that $P(X=x)$ is increasing) as long as

$$\frac{(g-x+1)(n-x+1)}{x(b-n+x)} > 1$$

which reduces to

$$x < \frac{(n+1)(g+1)}{g+b+2}.$$

This says that x_{mode} is the largest integer less than or equal to this value of x, which is the floor of this function and so

$$x_{\text{mode}} = \left\lfloor \frac{(n+1)(g+1)}{g+b+2} \right\rfloor . \tag{17d}$$

Example #21: A Maximum Likelihood Estimate

An unknown number, say N, of animals inhabit a certain region. To obtain some information about the size of the population (the value of N), ecologists often perform the following experiment. They first catch a number, say g, of these animals, mark them in some manner, and then release them. After allowing the marked animals time to disperse throughout the region, a new catch of size n

is made. Letting X denote the number of marked animals in this second capture, we assume that the population of animals in the region remains fixed from the time between the two catches and that each time an animal was caught, it was equally likely to be any of the remaining un-caught animals, if follows that X is hypergeometric with g marked animals and b unmarked animals with $N = g + b$. Assuming that the new catch of size x is the value of X having the highest probability of occuring (maximum likelihood) so that $x_{\text{mode}} = x$, we then have

$$x_{\text{mode}} = \left\lfloor \frac{(n+1)(g+1)}{g+b+2} \right\rfloor = \left\lfloor \frac{(n+1)(g+1)}{N+2} \right\rfloor$$

which says that

$$x_{\text{mode}} \leq \frac{(n+1)(g+1)}{N+2} < x_{\text{mode}} + 1$$

or

$$\frac{(n+1)(g+1)}{x_{\text{mode}}+1} - 2 < N \leq \frac{(n+1)(g+1)}{x_{\text{mode}}} - 2$$

and then the average of these, namely

$$N = \frac{(n+1)(g+1)}{2}\left(\frac{1}{x_{\text{mode}}} + \frac{1}{x_{\text{mode}}+1}\right) - 2$$

can be used to estimate the value of N. For example, suppose that the initial catch of animals is $g = 50$, which are marked and then released. If a subsequent catch consisting of $n = 40$ animals has $x = 4$ previously marked, and assuming that what we caught the second time around occurs with the highest probability and is therefore most likely (maximum likelihood) then $x_{\text{mode}} = 4$, then

$$N = \frac{(40+1)(50+1)}{2}\left(\frac{1}{4} + \frac{1}{4+1}\right) - 2 \simeq 468.475$$

showing that there are *most likely* around 468 animals in the region. ■

Non-Independent Bernoulli Trials

Note that the hypergeometric distribution can be viewed as non-independent, non-indentical Bernoulli trials resulting in the value of p *not being constant*. For example, thinking of good as a success and bad as a failure, the first draw has

$$p_1 = \frac{g}{g+b}$$

as the probability of success. Then the second draw has

$$p_2 = \frac{g}{g+b-1}$$

as the probability of success, if the first draw was a bad item (failure) and the second draw has

$$p_2 = \frac{g-1}{g+b-1}$$

as the probability of success, if the first draw was a good item (success). The third draw has

$$p_3 = \frac{g}{g+b-2}$$

as the probability of success, if both of the first two draws were bad items (failures), it has

$$p_3 = \frac{g-1}{g+b-2}$$

as the probability of success, if one of the first two draws was a good item (success) and the other was a bad item (failure), and it has

$$p_3 = \frac{g-2}{g+b-2}$$

as the probability of success, if both of the first two draws were good items (successes). This can be continued until either you run out of good items or you run out of bad items.

Example #22

Small electric motors are shipped in lots of $g+b = 50$. Before such a shipment is accepted, an inspector chooses $n = 5$ of these motors and inspects them. If none of these motors tested are found to be defective, the lot is accepted. If one or more are found to be defected, the entire shipment is inspected. Suppose that there are, in fact, three defective motors in the lot. Determine the probability that 100 percent inspection is required.

To answer this we first note that a motor is either defective ($g = 3$) or non-defected ($b = 47$). Letting X be the number of defective motors found, 100%

inspection requires that $X \geq 1$. This leads to

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{3}{0}\binom{47}{5-0}}{\binom{50}{5}} = 0.28,$$

or $P(X \geq 1) = 28\%$. ■

5.6 The Uniform Distribution on n Values

The Uniform Distribution on n values has a range set given by

$$R_X = \{a+1, a+2, a+3, \ldots, a+n\} \quad \text{with pmf} \quad P(X = x) = \frac{1}{n} \tag{18a}$$

for each $x = a+1, a+2, a+3, \ldots, a+n$ and $n \geq 1$. The cdf is given by

$$F(x) = \sum_{k=a+1}^{x} \frac{1}{n} = \frac{x-a}{n}. \tag{18b}$$

for $x = a+1, a+2, a+3, \ldots, a+n$ and $n \geq 1$. In most application, $a = 0$ and then the cdf of this distribution is given by

$$F(x) = \sum_{k=1}^{x} \frac{1}{n} = \frac{x}{n} \tag{18c}$$

for $x = 1, 2, 3, \ldots, n$ and $n \geq 1$. The mean of this distribution is given by

$$E(X) = \sum_{k=1}^{n} k\frac{1}{n} = \frac{1}{n}\sum_{k=1}^{n} k = \frac{1}{n}\frac{n(n+1)}{2}$$

or

$$E(X) = \frac{n+1}{2} \tag{18d}$$

and the second moment of this distribution is

$$E(X^2) = \sum_{k=1}^{n} k^2\frac{1}{n} = \frac{1}{n}\sum_{k=1}^{n} k^2 = \frac{1}{n}\frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$$

resulting in a variance of

$$V(X) = E(X^2) - (E(X))^2 = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2,$$

which reduces to

$$V(X) = \frac{n^2 - 1}{12}, \tag{18e}$$

for $n = 1, 2, 3, ...$

5.7 The Poisson Distribution With Parameter α

The *Poisson* probability mass function with parameter $\alpha > 0$ is given by

$$P(X = x) = \frac{e^{-\alpha}\alpha^x}{x!} \tag{19a}$$

for $x = 0, 1, 2, 3, \ldots$, and zero otherwise. It should be clear that

$$0 \leq P(X = x) = \frac{e^{-\alpha}\alpha^x}{x!} \leq 1$$

for all $x = 0, 1, 2, ...$, and

$$\sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} \frac{e^{-\alpha}\alpha^x}{x!} = e^{-\alpha} \sum_{x=0}^{\infty} \frac{\alpha^x}{x!} = e^{-\alpha}e^{\alpha} = 1$$

as it must. The first and second moments are given by

$$E(X) = \sum_{x=0}^{\infty} x\frac{e^{-\alpha}\alpha^x}{x!} = \sum_{x=1}^{\infty} \frac{e^{-\alpha}\alpha^x}{(x-1)!} = e^{-\alpha} \sum_{x=0}^{\infty} \frac{\alpha^{x+1}}{x!} = e^{-\alpha}\alpha e^{\alpha},$$

or

$$E(X) = \alpha, \tag{19b}$$

and

$$\begin{aligned}
E(X^2) &= \sum_{x=0}^{\infty} x^2\frac{e^{-\alpha}\alpha^x}{x!} = \sum_{x=1}^{\infty} x\frac{e^{-\alpha}\alpha^x}{(x-1)!} \\
&= e^{-\alpha}\sum_{x=0}^{\infty}(x+1)\frac{\alpha^{x+1}}{x!} = e^{-\alpha}\sum_{x=0}^{\infty} x\frac{\alpha^{x+1}}{x!} + e^{-\alpha}\sum_{x=0}^{\infty}\frac{\alpha^{x+1}}{x!} \\
&= e^{-\alpha}\sum_{x=1}^{\infty}\frac{\alpha^{x+1}}{(x-1)!} + e^{-\alpha}\alpha e^{\alpha} = e^{-\alpha}\sum_{x=0}^{\infty}\frac{\alpha^{x+2}}{x!} + e^{-\alpha}\alpha e^{\alpha} \\
&= e^{-\alpha}\alpha^2 e^{\alpha} + \alpha = \alpha^2 + \alpha,
\end{aligned}$$

respectively, and so the variance is given by

$$V(X) = E(X^2) - (E(X))^2 = \alpha^2 + \alpha - \alpha^2 = \alpha. \tag{19c}$$

The cumulative distribution function for the Poisson distribution is given by

$$F(x) = P(X \leq x) = \sum_{k=0}^{x} \frac{e^{-\alpha}\alpha^k}{k!}$$

for $x = 0, 1, 2, 3, \ldots$, and since this cannot be reduced to closed form, it must be tabulated or computed using a calculator or computer. *Note that the Poisson distribution has the property that its mean equals it variance, but it is not the only discrete distribution that has this property.*

Example #23

A computer-terminal repair person is "beeped" each time there is a call for service. The number of beeps per hour is known to occur in accordance with a Poisson distribution with a mean of $\alpha = 2$ per hour. The probability of three beeps in the next hour is given by

$$P(X = 3) = \frac{e^{-2}2^3}{3!} = 0.18.$$

The probability of two or more beeps in a 1-hour time period is given by

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - \frac{e^{-2}2^0}{0!} - \frac{e^{-2}2^1}{1!} = 0.594$$

or $P(X \geq 2) = 59.4\%$. ■

Example #24: Lead-Time Demand

The lead time in an inventory problem is defined as the time between placing of an order for restocking the inventory of a certain item and the receipt of that order. During this lead time, additional orders for this item may occur, resulting in the item's inventory going to zero (known as a stockout) or the item's inventory going below zero (known as a back order). The lead-time demand in an inventory system is this accumulation of demand for an item from the point at which an order is placed until the order is received, *i.e.*,

$$L = \sum_{i=1}^{T} D_i$$

where L is the total lead-time demand, D_i is the lead-time demand during the ith time period in which a reorder was placed and T is the number of time periods in which a reorder was placed during the lead time L. Note that in general the D_is and T may be random variables resulting in L being a random variable.

An inventory manager desires that the probability of a stockout not exceed a certain value during the lead time. For example, it may be required that the probability of a stockout during the lead time is not to exceed 5%. If the lead-time demand is Poisson distributed, when reorder should occur can be determined so that the lead-time demand does not exceed a specified value with a specified *protection probabaility.*

For example, assuming that the lead-time demand is Poisson with mean $\alpha = 10$ units demanded and that a 95% protection from a stockout is desired. Then we would like to determine the smallest value of x such that the probability that the lead-time demand does not exceed x is greater that or equal to 0.95, *i.e.*,

$$P(X \leq x) = F(x) = \sum_{k=0}^{x} \frac{e^{-10}10^k}{k!} \geq 0.95.$$

Using a table of value for $F(x)$ or just trial-and-error leads to $F(15) = 0.951 \geq 0.95$ so that $x = 15$, which means that reorders should be made when the inventory reaches 15 units. ■

The Sum of Independent Discrete Variables - Convolution Sums

Suppose that X and Y are independent discrete random variables with pmfs $p(x)$ and $q(y)$, respectively, and suppose that $Z = X + Y$, then to determine the pmf of z, we use the Theorem of total probability and write

$$P(Z = z) = \sum_{x} P(Z = z|X = x)P(X = x) = \sum_{x} P(Y = z - x|X = x)P(X = x)$$

which says that

$$r(z) \equiv P(Z = z) = \sum_x P(X = x)P(Y = z - x)$$

since X and Y are assumed independent and this says that

$$r(z) = \sum_x p(x)q(z - x). \tag{20a}$$

The student should show that this can also be written as

$$r(z) = \sum_y p(z - y)q(y). \tag{20b}$$

The sums in these two expressions are called *convolution* sums.

Example #25: Sum of Two Independent Uniform Distributions

Suppose that X is discrete uniform with $R_X = \{1, 2, 3, ..., n\}$ and Y is also discrete uniform with $R_Y = \{1, 2, 3, ..., n\}$ and suppose that $Z = X + Y$ where X and Y are independent. To determine the pmf of Z, we start with $p(x) = 1/n$ for $x = 1, 2, 3, ..., n$ and $q(y) = 1/n$ for $y = 1, 2, 3, ..., n$. Then from Equation (20a), we have

$$r_n(z) = \sum_x p(x)q(z - x).$$

Since $p(x) \neq 0$ requires that $1 \leq x \leq n$ and $q(z-x) \neq 0$ requires that $1 \leq z - x \leq n$, or

$$z - n \leq x \leq z - 1.$$

This says that

$$\max(1, z - n) \leq x \leq \min(n, z - 1)$$

so that

$$r_n(z) = \sum_{x=\max(1,z-n)}^{\min(n,z-1)} p(x)q(z - x) = \sum_{x=\max(1,z-n)}^{\min(n,z-1)} \frac{1}{n}\frac{1}{n}$$

or

$$r_n(z) = \frac{\min(n, z - 1) - \max(1, z - n) + 1}{n^2}$$

for $z = 2, 3, 4, ..., 2n$. Note that for $1 \leq z \leq n$, we have $\min(n, z-1) = z-1$ and $\max(1, z-n) = 1$ and

$$r_n(z) = \frac{z-1-1+1}{n^2} = \frac{z-1}{n^2}$$

and for $n < z \leq 2n$, we have $\min(n, z-1) = n$ and $\max(1, z-n) = z-n$, resulting in

$$r_n(z) = \frac{n-(z-n)+1}{n^2} = \frac{2n-z+1}{n^2}.$$

Thus we have

$$r_n(z) = \frac{1}{n^2} \times \begin{cases} z-1, & \text{for} \quad z = 2, 3, 4, ..., n \\ \\ 2n-z+1, & \text{for} \quad z = n+1, n+2, n+3, ..., 2n \end{cases}$$

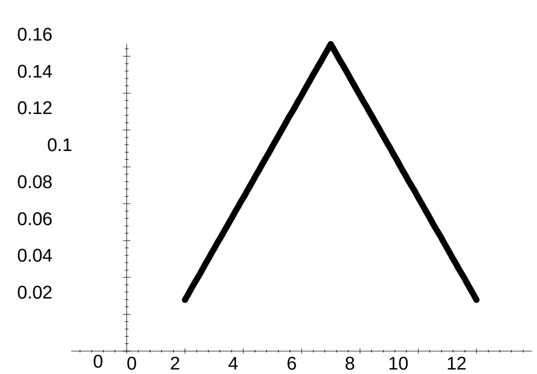
and note that $z-1 = 2n-z+1$ when $z = n+1$, and so we may also write this as

$$r_n(z) = \frac{1}{n^2} \times \begin{cases} z-1, & \text{for} \quad z = 2, 3, 4, ..., n, n+1 \\ \\ 2n-z+1, & \text{for} \quad z = n+1, n+2, n+3, ..., 2n \end{cases}.$$

In the case when $n = 6$, which is recording of sum of the roll of two 6-sided dice, we have

$$r_6(z) = \frac{1}{36} \times \begin{cases} z-1, & \text{for} \quad z = 2, 3, 4, 5, 6, 7 \\ \\ 13-z, & \text{for} \quad z = 7, 8, 9, 10, 11, 12 \end{cases}$$

which is plotted in the figure below



Plot of $r_6(z)$ versus z

and it is also the expected result. ■

The Sum of Independent Poisson Random Variables

Suppose that X_1 and X_2 are two Poisson random variables with parameters α_1 and α_2, respectively, and suppose that $X = X_1 + X_2$, then the pmf of X can be computed using the convolution sums as

$$\begin{aligned} P(X &= x) = \sum_{x_1=0}^{x} P(X = x \,|\, X_1 = x_1) P(\, X_1 = x_1) \\ &= \sum_{x_1=0}^{x} P(X_2 = x - x_1) P(\, X_1 = x_1) = \sum_{x_1=0}^{x} \frac{e^{-\alpha_2}\alpha_2^{x-x_1}}{(x-x_1)!}\frac{e^{-\alpha_1}\alpha_1^{x_1}}{x_1!} \\ &= \sum_{x_1=0}^{x} \frac{e^{-\alpha_2}\alpha_2^{x-x_1}}{(x-x_1)!}\frac{e^{-\alpha_1}\alpha_1^{x_1}}{x_1!} = \sum_{x_1=0}^{x} \frac{e^{-(\alpha_1+\alpha_2)}\alpha_1^{x_1}\alpha_2^{x-x_1}}{x_1!(x-x_1)!} \\ &= \frac{e^{-(\alpha_1+\alpha_2)}}{x!}\sum_{x_1=0}^{x} x!\frac{e^{-(\alpha_1+\alpha_2)}\alpha_1^{x_1}\alpha_2^{x-x_1}}{x_1!(x-x_1)!} = \frac{e^{-(\alpha_1+\alpha_2)}}{x!}\sum_{x_1=0}^{x}\binom{x}{x_1}\alpha_1^{x_1}\alpha_2^{x-x_1} \end{aligned}$$

and so we have

$$P(X = x) = \frac{e^{-(\alpha_1+\alpha_2)}(\alpha_1+\alpha_2)^x}{x!}$$

which shows that $X = X_1 + X_2$ is also Poisson with parameter $\alpha_1 + \alpha_2$. In general is X_1, X_2, X_3, …, X_n are all Poisson with parameters α_1, α_2, α_3, …, α_n, then

$$X = X_1 + X_2 + X_3 + \cdots + X_n$$

is Poisson with parameter

$$\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_n.$$

A practical and extremely important of the Poisson distribution is now discussed.

6. A Stationary Poisson Process Having Constant Rate λ

Consider a sequence of random events such as the arrival of units at a shop, or customers arriving at a bank, or web-site hits on the internet. These events may be described by a counting function $N(t)$ (defined for all $0 \leq t$), which equals the *number of events* that occur in the closed time interval $[0, t]$. We assume that $t = 0$ is the point at which the observations begin, whether or not an arrival occurs at that instant. Note that $N(t)$ is a *random variable* and the possible values of $N(t)$ (*i.e.*, its range set) are the non-negative integers: $0, 1, 2, 3, \ldots$, for t a non-negative parameter.

A counting process $N(t)$ is called a *Poisson process* with *mean rate* (per unit time) λ if the following assumptions are fulfilled.

A1: *Arrivals occur one at a time*: This implies that the probability of 2 or more arrivals in a very small (*i.e.*, infinitesimal) time interval Δt is zero *compared to* the probability of 1 or less arrivals occurring in the same time interval Δt.

A2: *$N(t)$ has stationary increments*: The distribution of the numbers of arrivals between t and $t + \Delta t$ depends only on the length of the interval Δt and not on the starting point t. Thus, arrivals are completely random without rush or slack periods. In addition, the probability that a *single arrival* occurs in a small time interval Δt is proportional to Δt and given by $\lambda \Delta t$ where λ is the mean arrival rate (per unit time).

A3: *$N(t)$ has independent increments*: The numbers of arrivals during non-overlapping time intervals are independent random variables. Thus, a large or small number of arrivals in one time interval has no effect on the number of arrivals in subsequent time intervals. Future arrivals occur completely at random, independent of the number of arrivals in past time intervals.

Given that arrivals occur according to a Poisson process, (*i.e.*, meeting the three assumptions A1, A2, and A3), let us derive an expression for the probability that n arrivals ($n = 0, 1, 2, 3, \ldots,$) occur in the time interval $[0, t]$. We shall denote this probability by $P_n(t)$, so that

$$P_n(t) = P(N(t) = n) \tag{21a}$$

for $n = 0, 1, 2, 3, \ldots$, and, of course

$$\sum_{n=0}^{\infty} P_n(t) = 1 \tag{21b}$$

for all time t.

Computing $P_0(t)$

First let us consider computing $P_0(t)$ which is the probability that no arrivals occur in the time interval $[0, t]$. From the above equation we may write this as

$$P_0(t) = 1 - P_1(t) - \sum_{n=2}^{\infty} P_n(t),$$

and for a small time interval $[0, \Delta t]$, this becomes

$$P_0(\Delta t) = 1 - P_1(\Delta t) - \sum_{n=2}^{\infty} P_n(\Delta t).$$

From assumption A2, we may say that $P_1(\Delta t) \simeq \lambda \Delta t$ for some arrival rate λ, and from assumption A1, we have $P_n(\Delta t) \simeq 0$ for $n \geq 2$. This leads to $P_0(\Delta t) \simeq 1 - \lambda \Delta t$, for small Δt. Now no arrivals by time $t + \Delta t$ means no arrivals by time t and no arrivals between t and $t + \Delta \dot{t}$. Therefore, we may write that

$$P_0(t + \Delta t) = P(N(t + \Delta t) = 0) = P((N(t) = 0) \cap (N(t + \Delta t) - N(t) = 0))$$

which, because of independence (assumption A3), can be written as

$$P_0(t + \Delta t) = P(N(t) = 0) P(N(t + \Delta t) - N(t) = 0)$$

since $[0, t]$ and $[t, t + \Delta t]$ are non-overlapping time intervals. Therefore we have

$$P_0(t + \Delta t) \simeq P_0(t) P_0(\Delta t) = P_0(t)(1 - \lambda \Delta t)$$

which yields

$$\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} \simeq -\lambda P_0(t).$$

In the limit as $\Delta t \to 0$, this becomes

$$\lim_{\Delta t \to 0} \frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} \equiv \frac{dP_0(t)}{dt} = -\lambda P_0(t). \qquad (22)$$

Solving the ODE $P_0'(t) = -\lambda P_0(t)$

Solving this differential equation for $P_0(t)$, we write

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) \qquad \text{or} \qquad \frac{dP_0(t)}{P_0(t)} = -\lambda dt$$

which leads to

$$\int \frac{dP_0(t)}{P_0(t)} = -\lambda \int dt \qquad \text{or} \qquad \ln(P_0(t)) = -\lambda t + C_1$$

or simply

$$P_0(t) = e^{C_1 - \lambda t} = e^{C_1} e^{-\lambda t} = Ae^{-\lambda t}$$

with $A = e^{C_1}$. Now $P_n(0) = 0$ for $n \geq 1$, since no arrivals can occur if no time has elapsed and hence

$$P_0(0) = 1 - \sum_{n=1}^{\infty} P_n(0) = 1 - 0 = 1.$$

Therefore we may say that $P_0(0) = Ae^{-\lambda 0} = A = 1$, and so we finally find that

$$P_0(t) = e^{-\lambda t} = P(N(t) = 0) \tag{23}$$

for all time $t \geq 0$.

Computing $P_n(t)$ for $n = 1, 2, 3, ...$

Next we consider $P_n(t)$ for $n = 1, 2, 3, ...$, and note that

$$P_n(t + \Delta t) = P(N(t + \Delta t) = n)$$

which can occur if and only if one of the following $n + 1$ events has occurred:

- $N(t) = 0$ and $N(t + \Delta t) - N(t) = n$, or
- $N(t) = 1$ and $N(t + \Delta t) - N(t) = n - 1$, or
- $N(t) = 2$ and $N(t + \Delta t) - N(t) = n - 2$, and so on until,
- $N(t) = n - 1$ and $N(t + \Delta t) - N(t) = 1$, or

- $N(t) = n$ and $N(t + \Delta t) - N(t) = 0$.

This leads to

$$\begin{aligned}
P_n(t + \Delta t) &= P(N(t + \Delta t) = n) \\
&= \sum_{x=0}^{n} P((N(t) = x) \cap (N(t + \Delta t) - N(t) = n - x)) \\
&= \sum_{x=0}^{n} P(N(t) = x) P(N(t + \Delta t) - N(t) = n - x) \\
&\simeq \sum_{x=0}^{n} P_x(t) P_{n-x}(\Delta t) \\
&= \sum_{x=0}^{n-2} P_x(t) P_{n-x}(\Delta t) + P_{n-1}(t) P_1(t) + P_n(t) P_0(\Delta t)
\end{aligned}$$

Using assumption A1, we have $P_{n-x}(\Delta t) \simeq 0$ for small Δt and $x \leq n-2$, $P_1(\Delta t) \simeq \lambda \Delta t$, and $P_0(\Delta t) \simeq 1 - \lambda \Delta t$. Therefore we see that

$$\begin{aligned}
P_n(t + \Delta t) &= \sum_{x=0}^{n-2} P_x(t) P_{n-x}(\Delta t) + P_{n-1}(t) P_1(t) + P_n(t) P_0(\Delta t) \\
&\simeq \sum_{x=0}^{n-2} P_x(t)(0) + P_{n-1}(t)(\lambda \Delta t) + P_n(t)(1 - \lambda \Delta t)
\end{aligned}$$

or

$$P_n(t + \Delta t) \simeq P_{n-1}(t)(\lambda \Delta t) + P_n(t)(1 - \lambda \Delta t)$$

which may be re-written as

$$\frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} \simeq \lambda P_{n-1}(t) - \lambda P_n(t).$$

In the limit as $\Delta t \to 0$, this yields

$$\lim_{\Delta t \to 0} \frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} \equiv \frac{dP_n(t)}{dt} = \lambda P_{n-1}(t) - \lambda P_n(t)$$

for $n \geq 1$ and $t \geq 0$.

Solving the ODE $P_n'(t) = \lambda P_{n-1}(t) - \lambda P_n(t)$

If we write this as

$$\frac{dP_n(t)}{dt} + \lambda P_n(t) = \lambda P_{n-1}(t)$$

and multiply both sides by $e^{\lambda t}$, we get

$$e^{\lambda t}\frac{dP_n(t)}{dt} + \lambda e^{\lambda t} P_n(t) = \lambda e^{\lambda t} P_{n-1}(t)$$

or simply

$$\frac{d}{dt}(e^{\lambda t} P_n(t)) = \lambda e^{\lambda t} P_{n-1}(t),$$

which (after integrating) leads to

$$e^{\lambda t} P_n(t) = C_2 + \int_0^t \lambda e^{\lambda z} P_{n-1}(z) dz$$

or simply

$$e^{\lambda t} P_n(t) = \lambda \int_0^t e^{\lambda z} P_{n-1}(z) dz$$

since $P_n(0) = 0$ for $n \geq 1$ results in $C_2 = 0$. This leads to

$$P_n(t) = \lambda e^{-\lambda t} \int_0^t e^{\lambda z} P_{n-1}(z) dz \tag{24}$$

for $n = 1, 2, 3, \ldots$. Consequently we have

$$P_1(t) = \lambda e^{-\lambda t} \int_0^t e^{\lambda z} P_0(z) dz = \lambda e^{-\lambda t} \int_0^t e^{\lambda z} e^{-\lambda z} dz = \lambda e^{-\lambda t} \int_0^t dz$$

or $P_1(t) = (\lambda t) e^{-\lambda t}$. For $n = 2$, we have

$$P_2(t) = \lambda e^{-\lambda t} \int_0^t e^{\lambda z} P_1(z) dz = \lambda e^{-\lambda t} \int_0^t e^{\lambda z} (\lambda z) e^{-\lambda z} dz = \lambda^2 e^{-\lambda t} \int_0^t z dz$$

or simply

$$P_2(t) = \frac{e^{-\lambda t} (\lambda t)^2}{2}.$$

It is easy to see that in general,

$$P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \tag{25}$$

for $n \geq 0$ and $t \geq 0$.

Therefore we see that if arrivals occur according to a Poisson process, meeting the three assumptions A1, A2, and A3, the probability that $N(t)$ is equal to n, *i.e.*, the probability that n arrivals occur in the time interval $[0, t]$, is given by

$$P(N(t) = n) = P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \tag{26a}$$

for $t \geq 0$ and $n = 0, 1, 2, 3, \ldots$. We see then that this is a *Poisson distribution* with parameter $\alpha = \lambda t$, which has mean and variance both given by

$$E(N(t)) = V(N(t)) = \alpha = \lambda t. \tag{26b}$$

Note that for any times t and s with $s < t$, assumption A2 implies that the random variable $N(t) - N(s)$, representing the number of arrivals in the interval $[s, t]$, is also Poisson with parameter $\lambda(t - s)$ so that

$$P(N(t) - N(s) = n) = \frac{e^{-\lambda(t-s)}(\lambda(t - s))^n}{n!} \tag{27}$$

for $n = 0, 1, 2, 3, \ldots$, and $t \geq s \geq 0$.

Some Properties of a Poisson Process

There are two important properties of the Poisson process known as Random Splitting and Random Pooling which we now discuss.

Random Splitting

Consider a Poisson process $N(t)$ having rate λ. Suppose that each time an event occurs it is classified as either: type 1 (with probability p_1), type 2 (with probability p_2), …, type k (with probability p_k), where of course

$$p_1 + p_2 + p_3 + \cdots + p_k = 1, \tag{28a}$$

then if $N_j(t)$ for $j = 1, 2, 3, \ldots, k$, represents the random variable of the number of type j events occurring in $[0, t]$, it can be shown that

$$N(t) = N_1(t) + N_2(t) + \cdots + N_k(t) \tag{28b}$$

and $N_j(t)$ for $j = 1, 2, 3, \ldots, k$, are all Poisson processes having rates λp_j, respectively. Furthermore each of these is independent of each other.

Example #25

If the arrival of customers in a bank is a Poisson process with parameter λ. We may break these customers up into disjoint classes (*i.e.*, male and female, or customers younger than 30, between 31 and 50, and older than 51), and the separate classes will all form a Poisson process with rates given by λp, where p is the probability that a particular class exists. For example, $p = 1/2$ in the case of male and female classes. ■

Random Pooling

Consider k different independent Poisson processes $N_j(t)$ for $j = 1, 2, 3, \ldots, k$, having rates λ_j, respectively. If

$$N(t) = N_1(t) + N_2(t) + \cdots + N_k(t) \tag{29a}$$

then, using the reproductive property of the Poisson distribution, shown earlier in this chapter, we must have $N(t)$ being a Poisson process with rate

$$\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_k. \tag{29b}$$

7. A Nonstationary Poisson Process Having Rate Function $\lambda_{\text{NSPP}}(t)$

If assumption A1 and A3 above are maintained, but assumption A2 is relaxed, then we have a *Non-Stationary Poisson Process* (NSPP), which is characterized by a rate that is not constant, but rather is a function of time t, so that $\lambda = \lambda_{\text{NSPP}}(t)$. This is known as an *arrival-rate function*, and it gives *the arrival rate AT time t*. This is useful for situations in which the arrival rate varies during the period of interest, including, meal times for restaurants, phone calls during business hours, and orders for pizza delivery around 6:00 PM.

The key to working with a NSPP is the *expected number of arrivals BY time t* (*i.e.*, in the time interval from 0 to t) defined as

$$\Lambda_{\text{NSPP}}(t) = \int_0^t \lambda_{\text{NSPP}}(s)ds. \tag{30}$$

Note that when $\lambda_{\text{NSPP}}(s) = \lambda$ (a constant), this gives $\Lambda_{\text{NSPP}}(t) = \lambda t$, which is what we had earlier for a *Stationary Poisson Process* (SPP). To be useful as a arrival-rate function, $\lambda_{\text{NSPP}}(t)$ must be nonnegative and integrable. Note also that

$$\bar{\lambda}_{\text{NSPP}}(t) = \frac{1}{t}\int_0^t \lambda_{\text{NSPP}}(s)ds = \frac{1}{t}\Lambda_{\text{NSPP}}(t) \tag{31}$$

gives an *average arrival rate function* for the NSPP over the first t units of time. Dropping the subscript NSPP, we see that

$$\Lambda(t) = \int_0^t \lambda(s)ds = \bar{\lambda}t. \tag{32}$$

Now let $N(t)$ be the arrival function for SPP and let $\mathcal{N}(t)$ be the arrival function for NSPP. The fundamental assumption for working with NSPPs is that

$$P(\mathcal{N}(t) = n|\Lambda(t)) = P(N(t) = n|\bar{\lambda}) \qquad \text{with} \qquad \bar{\lambda} = \frac{1}{t}\int_0^t \lambda(s)ds,$$

and since

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

we see that

$$P(\mathcal{N}(t) = n|\Lambda(t)) = \frac{e^{-\bar{\lambda}t}(\bar{\lambda}t)^n}{n!} \qquad \text{with} \qquad \bar{\lambda} = \frac{1}{t}\int_0^t \lambda(s)ds$$

or just

$$P(\mathcal{N}(t) = n|\Lambda(t)) = \frac{e^{-\Lambda(t)}(\Lambda(t))^n}{n!} \quad \text{with} \quad \Lambda(t) = \int_0^t \lambda(s)ds. \tag{33}$$

If the time interval is $[a, b]$ instead of $[0, t]$, then we have

$$P(\mathcal{N}(b) - \mathcal{N}(a) = n|\Lambda(t)) = P(N(b) - N(a) = n|\bar{\lambda})$$

with

$$\bar{\lambda} = \frac{1}{b-a}\int_a^b \lambda(s)ds,$$

and since

$$P(N(b) - N(a) = n|\bar{\lambda}) = \frac{e^{-\bar{\lambda}(b-a)}(\bar{\lambda}(b-a))^n}{n!}$$

we have

$$P(\mathcal{N}(b) - \mathcal{N}(a) = n|\Lambda) = \frac{e^{-\Lambda}\Lambda^n}{n!} \qquad \text{with} \qquad \Lambda = \int_a^b \lambda(s)ds$$

Let us illustrate this idea with an example.

Example #26 - A Non-Stationary Poisson Process

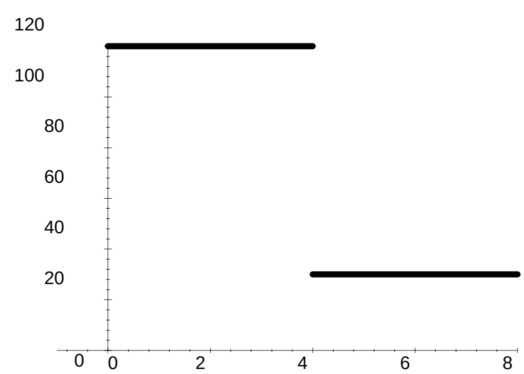
Suppose that arrivals to a post office occur at a rate of 2 per minute from 8:00 AM until 12:00 PM, then drop to 1 every 2 minutes until the post office closes at 4:00 PM. Determine the probability distribution on the number of arrivals between 11:00 AM and 2:00 PM. To solve this we let $t = 0$ correspond to 8:00 AM. Then the situation could be modeled as a NSPP with rate function

$$\lambda(t) = \begin{cases} 2 \text{ customers/minute}, & \text{for} \quad 0 \text{ hours} \le t < 4 \text{ hours} \\ \\ 1/2 \text{ customers/minute}, & \text{for} \quad 4 \text{ hours} \le t \le 8 \text{ hours} \end{cases}$$

where time t is in units of *hours*. The first thing we must do is get the units to mesh properly. Therefore, let us change $\lambda(t)$ to be on a per hour rate and so we write

$$\lambda(t) = \begin{cases} 120 \text{ customers/hour}, & \text{for} \quad 0 \text{ hours} \leq t < 4 \text{ hours} \\ \\ 30 \text{ customers/hour}, & \text{for} \quad 4 \text{ hours} \leq t \leq 8 \text{ hours} \end{cases}$$

A plot of this is shown in the figure below.



Plot of $\lambda(t)$ versus time t (in hours)

This says that the expected number of arrivals by time t (in hours) is

$$\Lambda(t) = \int_0^t \lambda(s)ds = \int_0^t 120ds = 120t.$$

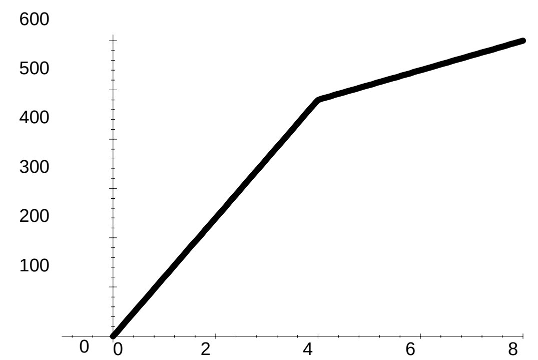
for 0 hours $\leq t \leq 4$ hours, and

$$\begin{aligned} \Lambda(t) &= \int_0^t \lambda(s)ds = \int_0^4 \lambda(s)ds + \int_4^t \lambda(s)ds \\ &= \int_0^4 120ds + \int_4^t 30ds = 360 + 30t \end{aligned}$$

for 4 hours $\leq t \leq 8$ hours, and so

$$\Lambda(t) = \begin{cases} 120t, & \text{for} \quad 0 \text{ hours} \leq t < 4 \text{ hours} \\ \\ 360 + 30t, & \text{for} \quad 4 \text{ hours} \leq t \leq 8 \text{ hours} \end{cases}.$$

A plot of this is shown in the figure below.

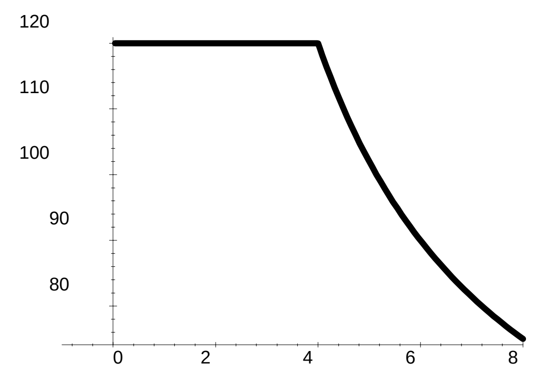


Plot of $\Lambda(t)$ versus time t (in hours)

We also note that

$$\bar{\lambda}(t) = \frac{1}{t}\Lambda(t) = \begin{cases} 120, & \text{for} \quad 0 \text{ hours} \leq t < 4 \text{ hours} \\ 30 + 360/t, & \text{for} \quad 4 \text{ hours} \leq t \leq 8 \text{ hours} \end{cases}$$

and a plot of this is shown in the figure below.



Plot of $\bar{\lambda}(t)$ versus time t (in hours)

Now since 2:00 PM and 11:00 AM correspond to $t = 6$ hours and $t = 3$ hours, respectively, we have

$$\Lambda = \int_3^6 \lambda(s)ds = \int_3^4 \lambda(s)ds + \int_4^6 \lambda(s)ds$$

or

$$\Lambda = \int_3^4 120 ds + \int_4^6 30 ds = 180 \text{ customers},$$

so that

$$P(\mathcal{N}(6) - \mathcal{N}(3) = n) = \frac{e^{-180}(180)^n}{n!}$$

for $n = 0, 1, 2, 3, \ldots$, which is a Poisson Distribution with $\alpha = 180$. ■