Core course for *App.AI*, *Bioinfo.*, *Data Sci.&Eng.*, *Dec.Analytics*, *Q.Fin.*, *Risk Mgmt.* and *Stat.* Majors:

STAT2601A Probability and Statistics I (2023-2024 First Semester)

Chapter 4: Discrete Distributions

4.1 Random Variables

Definition 4.1.
A *random variable* $X : \Omega \to \mathbb{R}$ is a numerical valued function defined on a sample space. In other words, a number $X(\omega)$, providing a measure of the characteristic of interest, is assigned to each outcome ω in the sample space. ◀

Remarks

Always keep in mind that X is a function rather than a number. The value of X depends on the outcome. We write $X = x$ to represent the event $\{\omega \in \Omega | X(\omega) = x\}$ and $X \leq x$ to represent the event $\{\omega \in \Omega | X(\omega) \leq x\}$.

Example 4.1.

Let X be the number of aces in a hand of three cards drawn randomly from a deck of 52 cards. Denote A as an ace card and N as a non-ace card. Then, the sample space is

$$\Omega = \{AAA, AAN, ANA, ANN, NAA, NAN, NNA, NNN\}.$$

The support of X is $X(\Omega) = \{0, 1, 2, 3\}$ which is a countable set. Hence, X is discrete.

$X : \Omega \to \{0, 1, 2, 3\}$ such that

$$\begin{aligned} X(AAA) &= 3, \\ X(AAN) &= X(ANA) = X(NAA) = 2, \\ X(ANN) &= X(NAN) = X(NNA) = 1, \\ X(NNN) &= 0. \end{aligned}$$

Using simple probability calculations, we have

$$\begin{aligned} \Pr(X = 0) &= \Pr(\{NNN\}) = \frac{\binom{48}{3}}{\binom{52}{3}} = 0.78262, \\ \Pr(X = 1) &= \Pr(\{ANN, NAN, NNA\}) = \frac{\binom{4}{1}\binom{48}{2}}{\binom{52}{3}} = 0.20416, \\ \Pr(X = 2) &= \Pr(\{AAN, ANA, NAA\}) = \frac{\binom{4}{2}\binom{48}{1}}{\binom{52}{3}} = 0.01303, \\ \Pr(X = 3) &= \Pr(\{AAA\}) = \frac{\binom{4}{3}}{\binom{52}{3}} = 0.00018. \end{aligned}$$

★

Example 4.2.

The annual income ω of a randomly selected citizen has a sample space $\Omega = [0, \infty)$. Suppose the annual income is taxable if it exceeds c. Let X be the taxable income. Then the space of X is also $[0, \infty)$ and $X : \Omega \to [0, \infty)$ such that

$$X(\omega) = \begin{cases} 0, & \text{if } \omega \leq c; \\ \omega - c, & \text{if } \omega > c. \end{cases}$$

★

Note

Conventionally, we use capital letters $X, Y, \ldots$ to denote the random variables and small letters $x, y, \ldots$ to denote the possible numerical values (or *realizations*) of these variables.

4.2 Distribution of the Discrete Type

Definition 4.2.
A random variable X defined on the sample space Ω is called a *discrete random variable* if $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$ is countable (e.g., $X : \Omega \to \{0, 1, 2, \ldots\}$).

◀

4.2.1 Probability Mass Function and Distribution Function

Definition 4.3.
The *probability mass function* (*pmf*) of a discrete random variable X is defined as

$$p(x) = \Pr(X = x), \qquad x \in X(\Omega),$$

where $X(\Omega)$ is the countable set of possible values of X.

◀

Example 4.3.

For the previous example of card drawing, the pmf of X is

$$p(0) = 0.78262, \qquad p(1) = 0.20416, \qquad p(2) = 0.01303, \qquad p(3) = 0.00018.$$

★

Conditions for a pmf

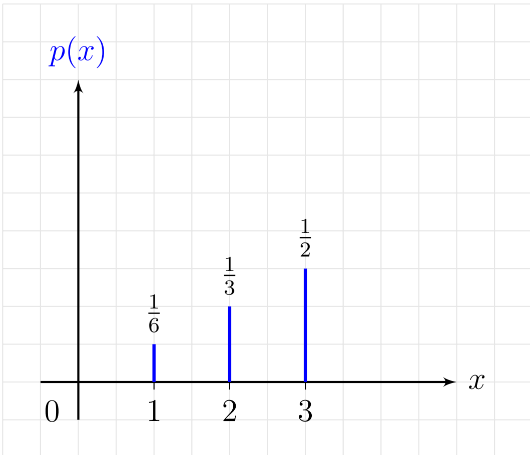
Since pmf is defined through probability, we have the following conditions for p to be a valid pmf:

1. $p(x) \geq 0$ for all $x \in X(\Omega)$;
2. $\displaystyle\sum_{x \in X(\Omega)} p(x) = 1$;
3. $\Pr(X \in A) = \displaystyle\sum_{x \in A} p(x)$ where $A \subset X(\Omega)$.

Example 4.4.

Is $p(x) = \dfrac{x}{6}$, $x = 1, 2, 3$, a valid pmf?

Solution:



$X(\Omega) = \{1, 2, 3\}$

1. $p(x) = \dfrac{x}{6} > 0$ for all $x = 1, 2, 3$;
2. $\displaystyle\sum_{i=1}^{3} p(x) = \frac{1}{6} + \frac{1}{3} + \frac{1}{2} = 1$;
3. $\Pr(X \leq 2) = p(1) + p(2) = \dfrac{1}{2}$.

Definition 4.4.

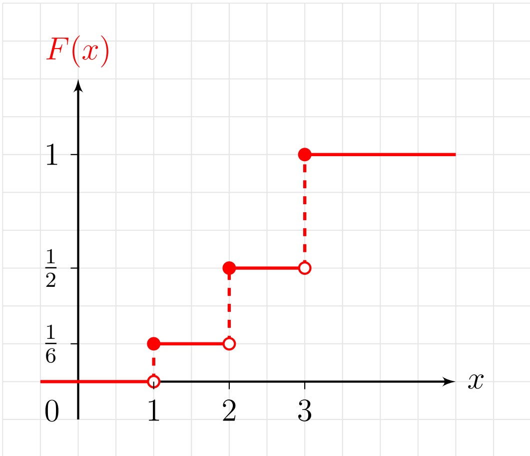
The *(cumulative) distribution function (cdf)* of a discrete random variable X is defined as

$$F(x) = \Pr(X \leq x) = \sum_{t \leq x} p(t), \qquad -\infty < x < \infty.$$

Example 4.5.

Continue with **Example 4.4.**, $p(x) = \dfrac{x}{6}$, $x = 1, 2, 3$, we have

$$\begin{aligned}
F(1) &= \Pr(X \le 1) = \Pr(X = 1) = \frac{1}{6} \\
F(1.5) &= \Pr(X \le 1.5) = \Pr(X = 1) = \frac{1}{6} = F(1.2601) = F(1.999999) = \cdots \\
F(2) &= \Pr(X \le 2) = p(1) + p(2) = \frac{1}{2} \\
F(3) &= \Pr(X \le 3) = p(1) + p(2) + p(3) = 1
\end{aligned}$$



As can be seen, the cdf of a discrete random variable would be a *step-function* with $p(x)$ as the size of the jumps at the possible value x.

Properties of a cdf

1. $F(x)$ is non-decreasing, i.e., if $a \le b$, then $F(a) \le F(b)$.
2. $F(-\infty) = \lim_{b \to -\infty} F(b) = 0$.
3. $F(\infty) = \lim_{b \to \infty} F(b) = 1$.
4. $F(x)$ is right-continuous. That is, for any b and any decreasing sequence $\{b_n, n \ge 1\}$ that converges to b, $\lim_{n \to \infty} F(b_n) = F(b)$.
5. $F(x)$ is a step-function if X is a discrete random variable. The step size at the point $x \in X(\Omega)$ is $\Pr(X = x)$.
6. The cdfs are useful in describing probability models. The probabilities attributed to events concerning a random variable X can be reconstructed from the cdf of X, i.e., the cdf of X completely specifies the random behaviour of X.

4.2.2 Mathematical Expectation

Example 4.6.

Consider the following two games. In each game, three fair dice will be rolled.

Game 1: If all the dice face up with same number, then you win \$24 otherwise you lose \$1.

Game 2: You win \$1, \$2, or \$3, according to one die face up as six, two dice face up as six or three dice face up as six, respectively. If no dice face up as six, then you lose \$1.

Which game is a better choice?

Solution:

To make a better decision, one may consider the amount one will win (or lose) in the long run. First we need to evaluate the probabilities of win or lose in each game. Let X, Y be the amounts of money you will win in one single trial of game 1 and game 2 respectively. A negative value means you lose money.

For game 1,

$$\begin{aligned}
\Pr(X=24) &= \Pr(\text{same number on three dice}) = 6 \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}, \\
\Pr(X=-1) &= 1 - \frac{1}{36} = \frac{35}{36}.
\end{aligned}$$

For game 2,

$$\begin{aligned}
\Pr(Y=1) &= 3 \times \frac{1}{6} \times \frac{5}{6} \times \frac{5}{6} = \frac{25}{72}, \\
\Pr(Y=2) &= 3 \times \frac{5}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{5}{72}, \\
\Pr(Y=3) &= \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216}, \\
\Pr(Y=-1) &= 1 - \frac{25}{72} - \frac{5}{72} - \frac{1}{216} = \frac{125}{216}.
\end{aligned}$$

Suppose we play game 1 for 36000 times. Since the relative frequency is a good estimate of the probability when number of trials is large, approximately in 1000 times we will win \$24 and 35000 times we will lose \$1. So in these 36000 trials of game 1, we win

$$24 \times 1000 + (-1) \times 35000 = -11000.$$

Approximately we will lose \$11000 in 36000 trials of game 1. The average amount we win in each trial is

$$\frac{-11000}{36000} = -\frac{11}{36}.$$

This is the long term average of gain if we play game 1 repeatedly. Indeed it can be calculated as

$$\begin{aligned}
-\frac{11}{36} &= \frac{24 \times 1000 + (-1) \times 35000}{36000} \\
&= 24 \times \frac{1}{36} + (-1) \times \frac{35}{36} \\
&= 24 \times \Pr(X=24) + (-1) \times \Pr(X=-1).
\end{aligned}$$

Similarly, the long term average of gain if we play game 2 repeatedly is

$$\begin{aligned} & 1 \times \Pr(Y = 1) + 2 \times \Pr(Y = 2) + 3 \times \Pr(Y = 3) + (-1) \times \Pr(Y = -1) \\ = \; & 1 \times \frac{25}{72} + 2 \times \frac{5}{72} + 3 \times \frac{1}{216} + (-1) \times \frac{125}{216} = -\frac{17}{216} > -\frac{11}{36}. \end{aligned}$$

Therefore game 2 is better than game 1 in terms of long term average gain.

However, since in the long run you will lose money in both games, the best strategy is do not gamble at all.

Definition 4.5.
Let X be a discrete random variable with pmf $p(x)$. The *mathematical expectation* (*expected value*) of X is defined by

$$\mathrm{E}(X) = \sum_{x \in X(\Omega)} x p(x),$$

provided that the summation exists. In general, for any function g, the expected value of $g(X)$ is

$$\mathrm{E}[g(X)] = \sum_{x \in X(\Omega)} g(x) p(x).$$

For example, $\mathrm{E}(X^2) = \sum_{x \in X(\Omega)} x^2 p(x)$, $\mathrm{E}(\ln X) = \sum_{x \in X(\Omega)} (\ln x) p(x)$, ..., etc.

Properties

1. If c is a constant, then $\mathrm{E}(c) = c$.

$$\mathrm{E}(c) = \sum_{x \in X(\Omega)} c p(x) = c \sum_{x \in X(\Omega)} p(x) = c.$$

2. If c is a constant, then $\mathrm{E}[cg(X)] = c\mathrm{E}[g(X)]$.

$$\mathrm{E}[cg(X)] = \sum_{x \in X(\Omega)} c g(x) p(x) = c \sum_{x \in X(\Omega)} g(x) p(x) = c\mathrm{E}[(g(X)].$$

3. If $c_1, c_2, \ldots, c_n$ are constants, then $\mathrm{E}\left[\sum_{i=1}^{n} c_i g_i(X)\right] = \sum_{i=1}^{n} c_i \mathrm{E}\left[g_i(X)\right]$.

   For example, $\mathrm{E}(5X + 3X^2) = 5\mathrm{E}(X) + 3\mathrm{E}(X^2)$.

   However, $\mathrm{E}(X^2) \neq [\mathrm{E}(X)]^2$, $\mathrm{E}(\ln X) \neq \ln \mathrm{E}(X)$.

4. $X(\omega) \geq Y(\omega)$ for all $\omega \in \Omega \Longrightarrow \mathrm{E}(X) \geq \mathrm{E}(Y)$.

5. $\mathrm{E}(|X|) \geq |\mathrm{E}(X)|$.

Example 4.7.

Continue with the previous gambling example (**Example 4.6.**), for game 1,

$$\begin{aligned}
\mathrm{E}(X^2) &= (24)^2\times\frac{1}{36}+(-1)^2\times\frac{35}{36}=\frac{611}{36}=16.9722\\
\mathrm{E}\left\{\left[X-\left(-\frac{11}{36}\right)\right]^2\right\} &= \left(24+\frac{11}{36}\right)^2\times\frac{1}{36}+\left(-1+\frac{11}{36}\right)^2\times\frac{35}{36}=16.8788
\end{aligned}$$

Alternatively,

$$\begin{aligned}
\mathrm{E}\left\{\left[X-\left(-\frac{11}{36}\right)\right]^2\right\} &= \mathrm{E}(X^2)+\left(\frac{11}{36}\right)^2+2\left(\frac{11}{36}\right)\mathrm{E}(X)\\
&= 16.9722+0.09336+2(0.3056)(-0.3056)=16.8788
\end{aligned}$$

The value $\mathrm{E}\left\{[X-\mathrm{E}(X)]^2\right\}$ can tell us the variation of our gains among long term trials of game 1.

★

4.2.3 Mean and Variance

Definition 4.6.

If X is a discrete random variable with pmf $p(x)$ and space $X(\Omega)$, then $\mathrm{E}(X)$ is called the (population) *mean* of X (of the distribution) and is usually denoted by μ. It is a measure of central location of the random variable X.

Furthermore, $\mathrm{E}\left[(X-\mu)^2\right]=\sum_{x\in X(\Omega)}(x-\mu)^2p(x)$ is called the (population) *variance* of X (of the distribution) and is usually denoted by σ^2 or $\mathrm{Var}(X)$.

The positive square root of the variance, $\sigma=\sqrt{\sigma^2}=\sqrt{\mathrm{Var}(X)}$, is called the (population) *standard deviation* of X (of the distribution). Both σ and σ^2 are measures of spread.

◀

Remarks

1. The variance, and hence the standard deviation, provide a measure of the variation of the probability distribution of the random variable X. Together with the mean μ, which is the measure of the central location, they give an informative summary on the probabilistic behavior of the random variable. Usually, a general random variable X would often take a value within two standard deviations from the mean, i.e., from $\mu-2\sigma$ to $\mu+2\sigma$. However, there are always exceptions.

2. A useful computational formula for the calculation of the variance is

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

Example 4.8.

Consider the probability distribution for the returns on stock A and B provided below:

State	Probability	Return on Stock A (X)	Return on Stock B (Y)
1	0.2	1%	10%
2	0.3	2%	6%
3	0.3	3%	2%
4	0.2	4%	−2%

Expected return on stock A:

$$\mu_{\mathrm{A}} = \mathrm{E}(X) = 0.01 \times 0.2 + 0.02 \times 0.3 + 0.03 \times 0.3 + 0.04 \times 0.2 = 2.5\%$$

Expected return on stock B:

$$\mu_{\mathrm{B}} = \mathrm{E}(Y) = 0.1 \times 0.2 + 0.06 \times 0.3 + 0.02 \times 0.3 + (-0.02) \times 0.2 = 4\%$$

Variance and standard deviation of return on stock A:

$$\begin{aligned}
\mathrm{E}(X^2) &= 0.01^2 \times 0.2 + 0.02^2 \times 0.3 + 0.03^2 \times 0.3 + 0.04^2 \times 0.2 = 0.00073 \\
\sigma_{\mathrm{A}}^2 &= \mathrm{Var}(X) = \mathrm{E}(X^2) - \mu_{\mathrm{A}}^2 = 0.00073 - (0.025)^2 = 0.000105 \\
\sigma_{\mathrm{A}} &= \sqrt{0.000105} = 1.02\%
\end{aligned}$$

Variance and standard deviation of return on stock B:

$$\begin{aligned}
\mathrm{E}(Y^2) &= 0.1^2 \times 0.2 + 0.06^2 \times 0.3 + 0.02^2 \times 0.3 + (-0.02)^2 \times 0.2 = 0.00328 \\
\sigma_{\mathrm{B}}^2 &= \mathrm{Var}(Y) = \mathrm{E}(Y^2) - \mu_{\mathrm{B}}^2 = 0.00328 - (0.04)^2 = 0.00168 \\
\sigma_{\mathrm{B}} &= \sqrt{0.00168} = 4.10\%
\end{aligned}$$

Although stock B offers a higher expected return than stock A, it is also riskier since its variance and standard deviation are greater than stock A's.

This, however, is only part of the picture because most investors choose to hold securities as part of a diversified portfolio.

★

Properties

Let X be a random variable and a, c be two constants.

1. $\mathrm{E}(aX+c) = a\mathrm{E}(X)+c$.

   Proof.

$$\begin{aligned}\mathrm{E}(aX+c) &= \sum(ax+c)p(x)\\ &= a\sum xp(x)+c\sum p(x)\\ &= a\mathrm{E}(X)+c\end{aligned}$$

□

2. $\mathrm{Var}(aX+c) = a^2\mathrm{Var}(X)$.

   Proof.

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

□

Example 4.9.

Consider the stock returns in **Example 4.8.**. Suppose a simple portfolio allocates 70% of the fund to invest in stock B and 30% to a term deposit with 2% fixed interest rate. Then the overall return of the portfolio can be expressed as

$$R = 0.7\times Y + 0.3\times 0.02 = 0.7Y + 0.006.$$

Therefore the expected return is

$$\mu_R = \mathrm{E}(R) = 0.7\mathrm{E}(Y) + 0.006 = 0.7\times 0.04 + 0.006 = 3.4\%,$$

with the risk evaluated as

$$\begin{aligned}\sigma_R^2 &= \mathrm{Var}(R) = (0.7)^2\times\mathrm{Var}(Y) = (0.7)^2\times 0.00168 = 0.0008232,\\ \sigma_R &= \sqrt{0.0008232} = 2.87\%.\end{aligned}$$

This simple example demonstrates the risk-return trade off. Comparing to a total assets of stock B, this portfolio can reduce the risk substantially with only little sacrifice in the expected return.

★

4.2.4 Moment and Moment Generating Function

Definition 4.7.
Let r be a positive integer. $\mathrm{E}(X^r)$ is called the r-th *moment* of X. $\mathrm{E}\left[(X-b)^r\right]$ is called the r-th moment of X about b if it exists. It is also called the r-th *central moment* if $b=\mu$. For example, $\mu=\mathrm{E}(X)$ is the 1st moment of X and $\sigma^2=\mathrm{Var}(X)$ is the 2nd central moment of X.

◀

Definition 4.8.
Let X be a discrete random variable with pmf $p(x)$ and space $X(\Omega)$, then

$$M_X(t)=\mathrm{E}(e^{tX})=\sum_{x\in X(\Omega)} e^{tx}p(x)$$

is called the *moment generating function* (*mgf*) of X if it exists. The domain of $M_X(t)$ is the set of all real numbers t such that e^{tX} has a finite expected value.

◀

Example 4.10.

Suppose X is a random variable with pmf $p(x)=2\left(\frac{1}{3}\right)^x$, $x=1,2,3,\ldots$

Then the moment generating function of X is

$$M_X(t)=\mathrm{E}(e^{tX})=\sum_{x=1}^{\infty} e^{tx}2\left(\frac{1}{3}\right)^x=\sum_{x=1}^{\infty} 2\left(\frac{e^t}{3}\right)^x.$$

For the closed-form expression for $M_X(t)$ to exist, the series must be converging, i.e., $\frac{e^t}{3}<1$. Therefore, for $t<\ln 3$,

$$M_X(t)=2\times\frac{\left(\frac{e^t}{3}\right)}{1-\left(\frac{e^t}{3}\right)}=\frac{2e^t}{3-e^t}.$$

$M_X(t)$ is undefined if $t\geq\ln 3$.

★

Properties of moment generating function

It can be used for "generating" moments.

$$\boxed{M_X^{(r)}(0) = \frac{\mathrm{d}^r}{\mathrm{d}t^r} M_X(t)\bigg|_{t=0} = \mathrm{E}(X^r)}$$

Proof.

$$\begin{aligned}
M_X(t) &= \mathrm{E}(e^{tX}) = \sum_{x \in X(\Omega)} e^{tx} p(x) && \Longrightarrow \quad M_X(0) = 1 \\
M_X'(t) &= \sum_{x \in X(\Omega)} x e^{tx} p(x) && \Longrightarrow \quad M_X'(0) = \sum_{x \in X(\Omega)} x p(x) = \mathrm{E}(X) \\
M_X''(t) &= \sum_{x \in X(\Omega)} x^2 e^{tx} p(x) && \Longrightarrow \quad M_X''(0) = \sum_{x \in X(\Omega)} x^2 p(x) = \mathrm{E}(X^2) \\
&\vdots
\end{aligned}$$

□

Example 4.11.

For **Example 4.10.**,

$$M_X'(t) = \frac{(3 - e^t)(2e^t) - (2e^t)(-e^t)}{(3 - e^t)^2} \qquad \Longrightarrow \mu = M_X'(0) = \frac{4 + 2}{4} = \frac{3}{2}.$$

$\sigma^2 = M_X''(0) - [M_X'(0)]^2$ may be difficult to find.

Consider the *cumulant generating function* (*cgf*) $R_X(t) = \ln M_X(t)$,

$$\begin{aligned}
R_X'(t) &= \frac{M_X'(t)}{M_X(t)} && \Longrightarrow \quad R_X'(0) = \frac{M_X'(0)}{M_X(0)} = \mu \\
R_X''(t) &= \frac{M_X(t)M_X''(t) - M_X'(t)M_X'(t)}{[M_X(t)]^2} && \Longrightarrow \quad R_X''(0) = \frac{M_X(0)M_X''(0) - [M_X'(0)]^2}{[M_X(0)]^2} = \sigma^2
\end{aligned}$$

Therefore,

$$M_X(t) = \frac{2e^t}{3 - e^t} \qquad \Longrightarrow R_X(t) = \ln 2 + t - \ln(3 - e^t),$$

$$R_X'(t) = 1 + \frac{e^t}{3 - e^t} \qquad \text{and} \qquad R_X''(t) = \frac{(3 - e^t)e^t - e^t(-e^t)}{(3 - e^t)^2}.$$

Hence,

$$\mu = R_X'(0) = 1 + \frac{1}{3 - 1} = \frac{3}{2} \qquad \text{and} \qquad \sigma^2 = R_X''(0) = \frac{2 + 1}{4} = \frac{3}{4}.$$

★

Remarks

Moment generating function uniquely characterizes the distribution. That is, if X and Y have the same moment generating function, then they must have the same distribution.

4.3 Common Discrete Distributions

4.3.1 Bernoulli Trials and Binomial Distribution

Bernoulli experiment

Possible outcome	X	Probability
Success	1	p
Failure	0	$1-p$

Some example

Trial	Success	Failure
Tossing a coin	Head	Tail
Birth of a child	Boy	Girl
Pure guess in multiple choice	Correct	Wrong
Randomly choose a voter	Support	Not support
Randomly select a product	Non-defective	Defective

Definition 4.9. (Bernoulli Distribution)
pmf of X:

$$p(x) = p^x(1-p)^{1-x}, \qquad x = 0, 1.$$

We call the distribution of X as a *Bernoulli distribution* with success probability p, denoted by

$$X \sim \mathrm{Ber}(p).$$

The outcome of each experiment is called a Bernoulli trial.

$$\mu = p, \qquad \mathrm{E}(X^2) = p, \qquad \sigma^2 = p - p^2 = p(1-p).$$

◀

Definition 4.10. (Binomial Distribution)
Let X be the random variable denoting the number of successes in n Bernoulli trials. If these n Bernoulli trials are:

(i) having the **same success probability** p, and

(ii) **independent**, i.e., the success probability of any trial is not affected by the outcome of other trials;

then X is said to have a *binomial distribution* with n trials and success probability p. It is denoted as

$$X \sim \mathrm{B}(n,p) \qquad \text{or} \qquad X \sim \mathrm{Bin}(n,p).$$

◀

Example 4.12.

Let X be the number of boys in a family with four children.

Denote Success (S): boy, Failure (F): girl

Value of X	0	1	2	3	4
Outcome	FFFF	SFFF FSFF FFSF FFFS	SSFF SFSF SFFS FSSF FSFS FFSS	SSSF SSFS SFSS FSSS	SSSS
Probability	$(1-p)^4$	$p(1-p)^3$	$p^2(1-p)^2$	$p^3(1-p)$	p^4
No. of permutations	$\binom{4}{0}$	$\binom{4}{1}$	$\binom{4}{2}$	$\binom{4}{3}$	$\binom{4}{4}$

Therefore, the pmf of X is given by

$$p(x) = \Pr(X = x) = \binom{4}{x} p^x (1-p)^{4-x}, \qquad x = 0, 1, 2, 3, 4.$$

★

In general, the pmf of $X \sim \mathrm{B}(n, p)$ is given by

$$p(x) = \Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n.$$

Recall the *binomial theorem*, for positive integer n,

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r}.$$

Hence, $p(x)$ is the $(x+1)$-th term in the expansion of $[p + (1-p)]^n$.

Distribution function:

$$F(x) = \Pr(X \le x) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i}, \qquad x = 0, 1, 2, \ldots, n.$$

Moment generating function:

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} = (pe^t + 1 - p)^n \qquad \text{for all } t.$$

From this mgf, one can easily derive

$$\boxed{\mu = np \qquad \text{and} \qquad \sigma^2 = np(1-p).}$$

Example 4.13.

An examination paper consists of 50 multiple choice questions, with 5 choices for each question. A student goes into the examination without knowing a thing, and tries to answer all the questions by pure guessing. Let X be the number of questions this student can answer correctly. Then obviously $X \sim \mathrm{B}(50, 0.2)$.

On average, he will get $\mathrm{E}(X) = 50 \times 0.2 = 10$ correct answers by pure guessing, and the corresponding variance is $\mathrm{Var}(X) = 50 \times 0.2 \times 0.8 = 8$.

The probability of getting 15 correct answers by pure guessing is

$$\Pr(X = 15) = p(15) = \binom{50}{15}(0.2)^{15}(0.8)^{35} = 0.02992.$$

Suppose that two marks will be given for each correct answer, while half mark will be deducted for each incorrect answer. Let Y be the total score this student can get. Then,

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

On average, he will get

$$\mathrm{E}(Y) = \mathrm{E}(2.5X - 25) = 2.5\mathrm{E}(X) - 25 = 2.5 \times 10 - 25 = 0$$

mark and the corresponding variance is

$$\mathrm{Var}(Y) = \mathrm{Var}(2.5X - 25) = (2.5)^2\mathrm{Var}(X) = (2.5)^2 \times 8 = 50.$$

If the passing mark is set to 40, the probability that he will pass the examination is

$$\begin{aligned} \Pr(Y \geq 40) &= \Pr(2.5X - 25 \geq 40) \\ &= \Pr(X \geq 26) \\ &= \sum_{x=26}^{50} \binom{50}{x}(0.2)^x(0.8)^{50-x} \\ &= 0.000000492. \end{aligned}$$

Example 4.14.

Eggs are sold in boxes of six. Each egg has independently of the others a probability 0.2 being cracked. A shopper requires three boxes of eggs and regards as satisfactory if a box contains not more than two cracked eggs.

Let X be the number of cracked eggs in a particular box. Then $X \sim \mathrm{B}(6, 0.2)$.

$$\begin{aligned}
\Pr(\text{a box is satisfactory}) &= \Pr(X \leq 2) \\
&= \Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2) \\
&= (0.8)^6 + \binom{6}{1}(0.2)(0.8)^5 + \binom{6}{2}(0.2)^2(0.8)^4 \\
&= 0.90112.
\end{aligned}$$

Let Y be the number of satisfactory boxes in five boxes. Then $Y \sim \mathrm{B}(5, 0.90112)$.

$$\begin{aligned}
& \Pr(\text{at least 3 boxes are satisfactory}) \\
= {} & \Pr(Y \geq 3) \\
= {} & \Pr(Y = 3) + \Pr(Y = 4) + \Pr(Y = 5) \\
= {} & \binom{5}{3}(0.90112)^3(0.09888)^2 + \binom{5}{4}(0.90112)^4(0.09888) + (0.90112)^5 \\
= {} & 0.99171.
\end{aligned}$$

★

Properties of the Binomial Distribution

1. When n is equal to 1, the binomial distribution $\mathrm{B}(1, p)$ is just the Bernoulli distribution $\mathrm{Ber}(p)$.

2. $X \sim \mathrm{B}(n, p)$ can be viewed as a sum of n independent Bernoulli random variables,

$$X = \sum_{i=1}^{n} Y_i, \qquad \text{where } Y_i \overset{\text{i.i.d.}}{\sim} \mathrm{Ber}(p).$$

3. If $X \sim \mathrm{B}(n, p)$, then $\Pr(X = k)$ first increases monotonically and then decreases monotonically, it attains its largest value when k is the largest integer less than or equal to $(n + 1)p$.

4. If $X \sim \mathrm{B}(n, p)$, then

$$\Pr(X = k + 1) = \frac{p}{1 - p} \frac{n - k}{k + 1} \Pr(X = k).$$

4.3.2 Geometric and Negative Binomial Distribution

Definition 4.11. (Geometric Distribution)
Suppose we perform a sequence of independent Bernoulli trials with success probability p. Let X be the number of trials performed until the first success is obtained. Then X is said to have a *geometric distribution*. It is denoted by

$$X \sim \text{Geo}(p).$$

The pmf of X is given by

$$p(x) = \Pr(X = x) = (1-p)^{x-1}p, \qquad x = 1, 2, \ldots.$$

◀

Distribution function:

$$\begin{aligned} F(x) &= \Pr(X \le x) = \sum_{i=1}^{x} p(i) \\ &= p\sum_{i=1}^{x}(1-p)^{i-1} \\ &= p\frac{1-(1-p)^x}{1-(1-p)} \\ &= 1-(1-p)^x, \qquad x = 1, 2, \ldots. \end{aligned}$$

Moment generating function:

$$\begin{aligned} M_X(t) &= \sum_{x=1}^{\infty} e^{tx}(1-p)^{x-1}p \\ &= \sum_{y=0}^{\infty} e^{t(y+1)}(1-p)^y p \\ &= pe^t\sum_{y=0}^{\infty}[(1-p)e^t]^y \\ &= \frac{pe^t}{1-(1-p)e^t}, \qquad t < -\ln(1-p). \end{aligned}$$

From this mgf, one can easily derive

$$\boxed{\mu = \frac{1}{p} \qquad \text{and} \qquad \sigma^2 = \frac{1-p}{p^2}.}$$

Example 4.15.

In a casino game, *roulette*, suppose you bet on the number 00 for every trial. Let X be the number of games played until you win once, then $X \sim \text{Geo}(1/38)$.

On average, you will need to play for $\text{E}(X) = 38$ games in order to win once. The probability of no more than 4 games played until your first win is

$$\Pr(X \leq 4) = F(4) = 1 - \left(1 - \frac{1}{38}\right)^4 = 0.1012.$$

★

Example 4.16.

Often packages that you buy in a store include a card with a picture, or other items of a set, and you try to collect all of the N possible cards. We would be interested in the expected number of trials one should make in order to collect a set of the cards. Define

$$\begin{aligned} X_1 &= 1, \quad \text{the number of trials until we see the first new card;} \\ X_i &= \text{the number of trials after the } (i-1)\text{-th new card until the } i\text{-th new card,} \end{aligned}$$

and assume that the packages are independent with equal chances to contain the N possible cards. Then for $i > 1$, the distribution of X_i is geometric with success probability $\dfrac{N-i+1}{N}$ and therefore $\text{E}(X_i) = \dfrac{N}{N-i+1}$. Let W be the number of trials needed for collecting the whole set of N different cards. Then $W = \sum_{i=1}^{N} X_i$ and therefore

$$\text{E}(W) = \sum_{i=1}^{N} \text{E}(X_i) = \sum_{i=1}^{N} \frac{N}{N-i+1} = N \sum_{i=1}^{N} \frac{1}{i}.$$

In particular, if $N = 9$, then

$$\text{E}(W) = 9 \times \left(1 + \frac{1}{2} + \cdots + \frac{1}{9}\right) = 25.4607.$$

★

Remarks

1. There is another definition of the geometric distribution. Let Y be the number of **failures** before the first success. Then obviously $Y = X - 1$.

$$\begin{aligned}
p(y) &= (1-p)^y p, \qquad y = 0, 1, 2, \ldots \\
F(y) &= 1-(1-p)^{y+1}, \qquad y = 0, 1, 2, \ldots \\
M_Y(t) &= \frac{p}{1-(1-p)e^t}, \qquad t < -\ln(1-p) \\
\mu &= \frac{1-p}{p} \\
\sigma^2 &= \frac{1-p}{p^2}
\end{aligned}$$

2. Consider for non-negative integers a and b,

$$\begin{aligned}
\Pr(X > a+b) &= 1-[1-(1-p)^{a+b}] \\
&= (1-p)^a(1-p)^b \\
&= \Pr(X > a)\Pr(X > b).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Pr(X > a+b | X > a) &= \frac{\Pr(X > a+b, X > a)}{\Pr(X > a)} \\
&= \frac{\Pr(X > a+b)}{\Pr(X > a)} \\
&= \Pr(X > b).
\end{aligned}$$

Hence, conditional on no success before a-th trial, the probability of no success in the next b trials is equal to the unconditional probability of no success before the first b-th trial. This property is called the *memoryless property*. Among all discrete distributions, geometric distribution is the only one that has the memoryless property.

Example 4.17.

Consider the roulette example. Since the geometric distribution is memoryless, although you may have already lost 100 games in a row, the probability of waiting more than 5 games till you win is just the same as the scenario that you did not lose the 100 games at all. This dispels the so called *Gambler's fallacy* that "cold hands" are more (less) likely to come up after observing a series of "hot hands".

★

Definition 4.12. (Negative Binomial Distribution)
Suppose we perform a sequence of independent Bernoulli trials with success probability p. Let X be the number of trials until a total of r successes are accumulated. (Totally there are X trials to produce r successes, and the X-th trial is a success.) Then X is said to have a *negative binomial distribution*. It is denoted by

$$X \sim \mathrm{NB}(r, p).$$

The pmf of X is given by

$$p(x) = \Pr(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \qquad x = r, r+1, r+2, \ldots.$$

◀

Recall the *negative binomial series*,

$$\begin{aligned} \frac{1}{(1-a)^r} &= 1 + ra + \frac{1}{2!}r(r+1)a^2 + \cdots + \frac{1}{k!}r(r+1)\cdots(r+k-1)a^k + \cdots \\ &= \sum_{k=0}^{\infty} \binom{r+k-1}{k} a^k = \sum_{k=0}^{\infty} \binom{r+k-1}{r-1} a^k = \sum_{y=r}^{\infty} \binom{y-1}{r-1} a^{y-r}. \end{aligned}$$

Hence, $p(x)$ is the $(x-r+1)$-th term in the expansion of $p^r[1-(1-p)]^{-r}$.

Distribution function:

$$F(x) = \sum_{i=r}^{x} \binom{i-1}{r-1} p^r (1-p)^{i-r}, \qquad x = r, r+1, r+2, \ldots.$$

Moment generating function:

$$\begin{aligned} M_X(t) &= \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} p^r (1-p)^{x-r} \\ &= p^r e^{rt} \sum_{x=r}^{\infty} \binom{x-1}{r-1} [(1-p)e^t]^{x-r} \\ &= \left[\frac{pe^t}{1-(1-p)e^t}\right]^r, \qquad t < -\ln(1-p). \end{aligned}$$

From this mgf, one can easily derive

$$\boxed{\mu = \frac{r}{p} \qquad \text{and} \qquad \sigma^2 = \frac{r(1-p)}{p^2}.}$$

Example 4.18.

Fermat and Pascal are sitting in a cafe in Paris and decide to play the simplest of all games, flipping a coin. If the coin comes up head, Fermat gets a point. If the coin comes up tail, Pascal gets a point. The first to get 10 points wins the total pot worth 100 Francs. But then a strange thing happens. Fermat is winning 7 points to 6, when he receives an urgent message that a friend is sick, and he must rush to his home town of Toulouse immediately. Of course Pascal understands, but later, in correspondence, the problem arises: how should the 100 Francs be divided?

Solution:

Let X be the number of additional games they need to play so that Fermat can get 3 more points. Then X is the number of trials until 3 heads (successes) are obtained. Therefore X is a negative binomial random variable with $r = 3$ and $p = 0.5$. The probability mass function is given by

$$p(x) = \binom{x-1}{3-1}(0.5)^3(0.5)^{x-3} = \binom{x-1}{2}(0.5)^x, \qquad x = 3, 4, 5, \ldots.$$

For Fermat to win the game, Pascal should get less than 4 points before he gets 3 points, i.e., X must be less than 7. Therefore,

$$\begin{aligned}
\Pr(\text{Fermat wins}) &= \Pr(X < 7) \\
&= \binom{3-1}{2}(0.5)^3 + \binom{4-1}{2}(0.5)^4 + \binom{5-1}{2}(0.5)^5 + \binom{6-1}{2}(0.5)^6 \\
&= 0.65625.
\end{aligned}$$

Hence Fermat should receive 65.625 Francs, while Pascal should receive 34.375 Francs.

Consider a general *problem of points* in which a player needs N points to win. Suppose the intermediate scores are m to n when the game is interrupted. Let $r = N - m$, $s = N - n$ respectively be the additional points the two sides need in order to win the game. Then the probability that the side with m points would win the game is

$$\sum_{i=r}^{r+s-1} \binom{i-1}{r-1} p^r (1-p)^{i-r}.$$

★

Remarks

1. If r is equal to 1, then the negative binomial distribution $\text{NB}(1,p)$ becomes the geometric distribution $\text{Geo}(p)$, i.e., $\text{NB}(1,p) \equiv \text{Geo}(p)$.

2. There is another definition of the negative binomial distribution. Let Y be the number of **failures** before the r-th success. Then obviously, $Y = X - r$.

$$\begin{aligned} p(y) &= \binom{y+r-1}{r-1} p^r (1-p)^y, \qquad y = 0, 1, 2, \ldots \\ M_Y(t) &= \left[\frac{p}{1-(1-p)e^t}\right]^r, \qquad t < -\ln(1-p) \\ \mu &= \frac{r(1-p)}{p} \\ \sigma^2 &= \frac{r(1-p)}{p^2} \end{aligned}$$

3. A technique known as inverse binomial sampling is useful in sampling biological populations. If the proportion of individuals possessing a certain characteristic is p and we sample until we see r such individuals, then the number of individuals sampled is a negative binomial random variable.

4.3.3 Hypergeometric Distribution

Definition 4.13. (Hypergeometric Distribution)
Suppose we have N objects with m objects as type I and $(N-m)$ objects as type II. A sample of n objects is randomly drawn without replacement from the N objects. Let X be the number of type I objects in the sample. Then X is said to have a *hypergeometric distribution*. It is denoted by

$$X \sim \text{Hyp}(N, m, n).$$

The pmf of X is given by

$$p(x) = \Pr(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}, \qquad \max\left[n-(N-m), 0\right] \le x \le \min(n, m).$$

The mean and variance are

$$\mu = np \qquad \text{and} \qquad \sigma^2 = \left(\frac{N-n}{N-1}\right) np(1-p),$$

where $p = \dfrac{m}{N}$ is the proportion of type I objects in the population of N objects. ◀

Remarks

The hypergeometric distribution can be regarded as a finite population counterpart of the binomial distribution.

$$\begin{array}{lcl} \text{Sampling with replacement} & \Longrightarrow & \text{Binomial distribution} \\ \text{Sampling without replacement} & \Longrightarrow & \text{Hypergeometric distribution} \end{array}$$

The variance of a hypergeometric distribution and that of a binomial distribution differ by the multiplier $\dfrac{N-n}{N-1}$ and this expression is therefore called the *finite population correction factor*.

Example 4.19.

Let X be the number of 2's in a hand of 13 cards drawn randomly from a deck of 52 cards. Then X has a hypergeometric distribution with $N = 52$, $m = 4$, $n = 13$.

$$\begin{aligned} \mathrm{E}(X) &= 13 \times \frac{4}{52} = 1 \\ \mathrm{Var}(X) &= \left(\frac{52-13}{52-1}\right)(13)\left(\frac{4}{52}\right)\left(1-\frac{4}{52}\right) = 0.7059 \\ \Pr(X=3) &= \frac{\binom{4}{3}\binom{48}{10}}{\binom{52}{13}} = 0.04120 \end{aligned}$$

Example 4.20. (Capture and Re-capture Experiment)

To estimate the population size of a specific kind of animal in a certain region, e.g., number of fish in a lake, ecologists often perform the following procedures.

1. Catch m fish from the lake.
2. Tag these m fish with a certain marker and release them back to the lake.
3. After a certain period of time, catch n fish from the lake.
4. Count the number of tagged fish found in this new sample, denote as X.

It follows that X is a hypergeometric random variable such that

$$P_i(N) := \Pr(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}.$$

Consider

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

This ratio is greater than 1 if and only if

$$(N-m)(N-n) \geq N(N-m-n+i) \quad \Longleftrightarrow \quad N \leq \frac{mn}{i}.$$

Hence for fixed m, n, i, the value $P_i(N)$ is first increasing, and then decreasing, and it attains its maximum value at the largest integral value not exceeding $\frac{mn}{i}$. Therefore a reasonable estimation of the population size N is

$$\left\lfloor \frac{mn}{i} \right\rfloor,$$

where i is the number of tagged fish we found in the new sample.

This kind of estimation is known as the *maximum likelihood estimation* (*MLE*).

4.3.4 Poisson Distribution

Definition 4.14. (Poisson Distribution)
A random variable X, taking on one of the values $0, 1, 2, \ldots$, is said to have the *Poisson distribution* with parameter λ ($\lambda > 0$) if

$$p(x) = \Pr(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots.$$

It is denoted as

$$X \sim \text{Po}(\lambda) \qquad \text{or} \qquad X \sim \text{Poi}(\lambda).$$

Distribution function:

$$F(x) = \sum_{i=0}^{x} \frac{e^{-\lambda}\lambda^i}{i!}, \qquad x = 0, 1, 2, \ldots.$$

Moment generating function:

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} \\ &= e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} \\ &= e^{-\lambda}e^{\lambda e^t} \\ &= e^{\lambda(e^t-1)}, \qquad \text{for all } t. \end{aligned}$$

From this mgf, one can easily derive

$$\boxed{\mu = \sigma^2 = \lambda.}$$

If $X \sim \text{Po}(\lambda)$, one can note a computational formula as

$$\Pr(X = k+1) = \frac{\lambda}{k+1} \Pr(X = k).$$

Definition 4.15.
Let an event E with occurrence in time obeying the following postulates:

1. *Independence* – The number of times E occurs in non-overlapping time intervals are independent.

2. *Lack of clustering* – The probability of two or more occurrences in a sufficiently short interval is essentially zero.

3. *Rate* – The probability of exactly one occurrence in a sufficient short time interval of length h is approximately λh, i.e., directly proportional to h.

Denote $N(t)$ as the number of occurrences of E within the time interval $[0, t]$. Then $\{N(t), t \geq 0\}$ is said to be a *Poisson process* and the probabilistic behaviour of $N(t)$ can be modelled by the Poisson distribution with parameter λt.

◀

The formal derivation of the distribution of $N(t)$ requires the knowledge of differential equations and is omitted here. The following provides an informal justification of the result.

First we may partition the time interval into n subintervals each with length $h = t/n$.

For n sufficiently large, i.e., h sufficiently short,

$$\begin{aligned}
\Pr(\text{one occurrence in a subinterval}) &= \lambda h = \frac{\lambda t}{n} && \text{(by postulate 3)}, \\
\Pr(\text{more than one occurrence in a subinterval}) &= 0 && \text{(by postulate 2)}.
\end{aligned}$$

Hence each subinterval can be regarded as a Bernoulli trial with success probability $\lambda t/n$. From postulate 1, the occurrences in the subintervals are independent. Therefore we have n independent Bernoulli trials each with success probability $\lambda t/n$. Hence,

$$N(t) \sim \text{B}\left(n, \frac{\lambda t}{n}\right),$$

$$p(x) = \Pr(N(t) = x) = \binom{n}{x} \left(\frac{\lambda t}{n}\right)^x \left(1 - \frac{\lambda t}{n}\right)^{n-x}, \qquad x = 0, 1, 2, \ldots, n.$$

Since we need the subintervals to be sufficiently small, we should consider the limit of the above expression when $n \to \infty$,

$$\begin{aligned}
p(x) &= \lim_{n\to\infty} \binom{n}{x} \left(\frac{\lambda t}{n}\right)^x \left(1 - \frac{\lambda t}{n}\right)^{n-x} \\
&= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{x!} \frac{(\lambda t)^x}{n^x} \left(1 - \frac{\lambda t}{n}\right)^{n-x} \\
&= \frac{(\lambda t)^x}{x!} \lim_{n\to\infty} \frac{n}{n} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right) \left(1 - \frac{\lambda t}{n}\right)^{-x} \left(1 - \frac{\lambda t}{n}\right)^{n} \\
&= \frac{e^{-\lambda t}(\lambda t)^x}{x!}.
\end{aligned}$$

Thus, the pmf of $N(t)$ is given by

$$p(x) = \Pr(N(t) = x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \qquad x = 0, 1, 2, \ldots.$$

Therefore, $N(t)$ has a Poisson distribution with parameter λt, i.e., $N(t) \sim \mathrm{Po}(\lambda t)$.

Remarks

1. Note that
$$\mathrm{E}\left[N(t)\right] = \lambda t \qquad \Longrightarrow \mathrm{E}\left[\frac{N(t)}{t}\right] = \lambda.$$
Therefore λ can be interpreted as the average number of occurrence per unit time interval. The value of λ depends on the time unit used.

2. According to above derivation, the Poisson distribution can be used to approximate a binomial distribution when n is large and p is small.

Theorem 4.1. (Poisson Approximation to Binomial Distribution)
When n is large and p is small such that np is bounded, then the binomial distribution $\mathrm{B}(n,p)$ can be approximated by $\mathrm{Po}(np)$, i.e.,

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \approx \frac{e^{-np}(np)^x}{x!}.$$

This approximation should be successful if $n \geq 100$ and $np \leq 10$.

▲

Example 4.21.

Customers arrive to a departmental store in accordance with a Poisson distribution, with a rate of 30 per hour. What is the probability that there will be less than 10 customers arriving in the next half hour? What is the probability that there will be more than 6 customers arriving in ten minutes?

Solution:

$$\text{Time unit} = \text{hour} \qquad \Longrightarrow \lambda = 30$$

For the number of customers arriving in half an hour, the mean is $\lambda t = 30 \times 0.5 = 15$, i.e., $X = N(0.5) \sim \text{Po}(15)$. The probability that there will be less than 10 customers arriving is

$$\begin{aligned}\Pr(X < 10) &= \Pr(X = 0) + \Pr(X = 1) + \cdots + \Pr(X = 9) \\ &= e^{-15}\left(1 + 15 + \frac{15^2}{2!} + \frac{15^3}{3!} + \cdots + \frac{15^9}{9!}\right) \\ &= 0.0699.\end{aligned}$$

For the number of customers arriving in ten minutes, the mean is $\lambda t = 30 \times \frac{1}{6} = 5$, i.e., $Y = N(1/6) \sim \text{Po}(5)$. The probability that there will be more than 6 customers arriving is

$$\begin{aligned}\Pr(Y > 6) &= 1 - \Pr(Y \leq 6) \\ &= 1 - \Pr(Y = 0) - \Pr(Y = 1) - \cdots - \Pr(Y = 6) \\ &= 1 - e^{-5}\left(1 + 5 + \frac{5^2}{2!} + \frac{5^3}{3!} + \cdots + \frac{5^6}{6!}\right) \\ &= 1 - 0.7622 \\ &= 0.2378.\end{aligned}$$

★

Example 4.22.

In a certain manufacturing process in which glass items are being produced, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that on the average 1 in every 1000 of these items produced has one or more bubbles. What is the probability that a random sample of 8000 will yield fewer than 8 items possessing bubbles?

Solution:

Let X be the number of items possessing bubbles.

Then $X \sim \text{B}\left(8000, \frac{1}{1000}\right)$ can be approximated by Po(8).

The required probability is

$$\begin{aligned}\Pr(X < 8) &= \Pr(X \leq 7) \\ &= \sum_{x=0}^{7} \binom{8000}{x} (0.001)^x (0.999)^{8000-x} \\ &\approx \sum_{x=0}^{7} \frac{e^{-8} 8^x}{x!} \\ &= 0.452960809.\end{aligned}$$

★

Example 4.23.

Suppose that the total number of goals in a soccer match of the English Premier League follows the Poisson distribution with $\lambda = 2.8$, i.e., on average there are 2.8 goals per match. Determine, in a soccer match,

(a) the probability that there will be more than 2 goals;

(b) the probability that there will be even number of goals (zero is counted as even).

Solution:

Let X be the number of goals in a particular match. Then X follows Po(2.8), i.e.,

$$\Pr(X = x) = \frac{e^{-2.8}(2.8)^x}{x!}, \qquad x = 0, 1, 2, \ldots .$$

(a)

$$\begin{aligned}
\Pr(X > 2) &= 1 - \Pr(X \le 2) \\
&= 1 - [\Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2)] \\
&= 1 - e^{-2.8}\left[1 + 2.8 + \frac{(2.8)^2}{2!}\right] \\
&= 0.5305
\end{aligned}$$

(b)

$$\begin{aligned}
\Pr(X \text{ is even}) &= \Pr(X = 0) + \Pr(X = 2) + \Pr(X = 4) + \cdots \\
&= e^{-2.8}\left[1 + \frac{(2.8)^2}{2!} + \frac{(2.8)^4}{4!} + \cdots\right]
\end{aligned}$$

Recall the Taylor series expansion of the exponential function,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots ,$$

we have

$$\begin{aligned}
e^{2.8} &= 1 + 2.8 + \frac{(2.8)^2}{2!} + \frac{(2.8)^3}{3!} + \frac{(2.8)^4}{4!} + \cdots , \\
e^{-2.8} &= 1 - 2.8 + \frac{(2.8)^2}{2!} - \frac{(2.8)^3}{3!} + \frac{(2.8)^4}{4!} - \cdots .
\end{aligned}$$

Therefore,

$$e^{2.8} + e^{-2.8} = 2\left[1 + \frac{(2.8)^2}{2!} + \frac{(2.8)^4}{4!} + \cdots\right].$$

Hence,

$$\begin{aligned}
\Pr(X \text{ is even}) &= e^{-2.8} \times \frac{e^{2.8} + e^{-2.8}}{2} \\
&= \frac{1 + e^{-5.6}}{2} \\
&= 0.5018.
\end{aligned}$$

★

Example 4.24.

Random variables that are often modelled by the Poisson distribution include the following:

1. Number of misprints on a page of a book.
2. Number of people in a community living to 100 years of age.
3. Number of wrong telephone numbers that are dialled in a day.
4. Number of packages of dog biscuits sold in a particular store each day.
5. Number of customers entering a post office on a given day.
6. Number of vacancies occurring during a year in the Supreme Court.
7. Number of α-particles discharged in a fixed period of time from some radioactive material.
8. Number of earthquakes occurring during some fixed time span.
9. Number of wars per year.
10. Number of electrons emitted from a heated cathode during a fixed time period.
11. Number of deaths in a given period of time of the policyholders of a life insurance company.
12. Number of flaws in a certain type of drapery material.

★

Note that binomial distribution, negative binomial distribution (the version with non-negative integer support), and Poisson distribution are often used to model counting variables.

Distribution	**Symbol**	**Support**	**Mean**		**Variance**
Binomial					
Negative Binomial *					
Poisson					

* For the version counting the number of failure(s) before the r-th success

∼ End of Chapter 4 ∼