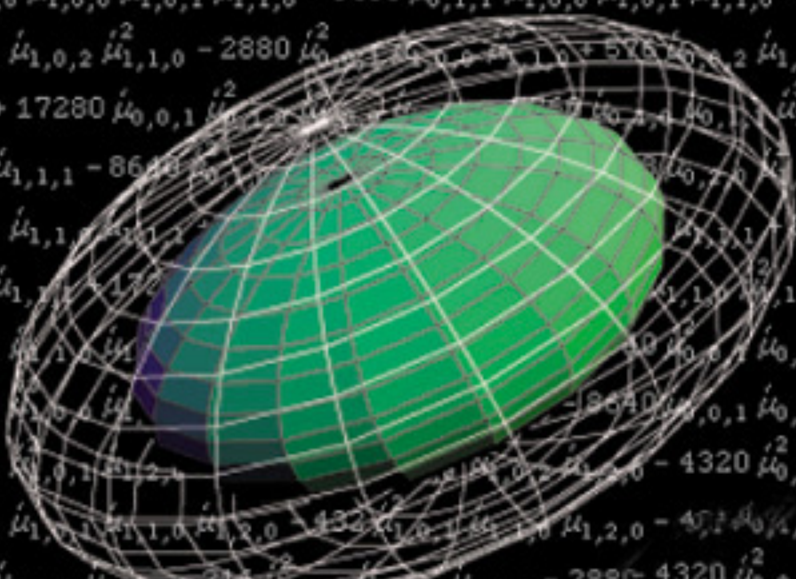
SPRINGER TEXTS IN STATISTICS

MATHEMATICAL STATISTICS *with* *Mathematica*®



COLIN ROSE
MURRAY D. SMITH

Mathematical Statistics with *Mathematica*

Chapter 3 – Discrete Random Variables

Please reference this 2002 edition as:

Rose, C. and Smith, M.D. (2002)
Mathematical Statistics with Mathematica, Springer-Verlag, New York.

Latest edition

For the latest up-to-date edition, please visit: www.mathStatica.com

Chapter 3

Discrete Random Variables

3.1 Introduction

In this chapter, attention turns to random variables induced from experiments defined on countable, or enumerable, sample spaces. Such random variables are termed *discrete*; their values can be mapped in one-to-one correspondence with the set of integers. For example, the experiment of tossing a coin has two possible outcomes, a head and a tail, which can be mapped to 0 and 1, respectively. Accordingly, random variable X, taking values $x \in \{0, 1\}$, represents the experiment and is discrete.

The distinction between discrete and continuous random variables is made clearer by considering the *cumulative distribution function* (cdf). For a random variable X (discrete or continuous), its cdf $F(x)$, as a function of x, is defined as

$$F(x) = P(X \leq x), \quad \text{for all } x \in \mathbb{R}. \tag{3.1}$$

Now inspect the following cdf plots given in Fig. 1.

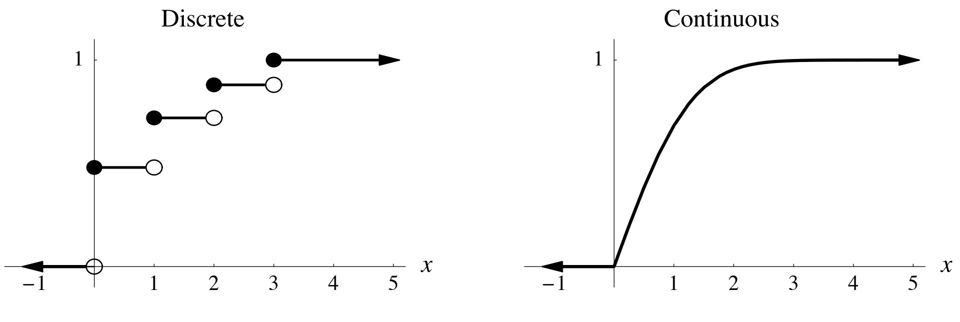


Fig. 1: Discrete and Continuous cumulative distribution functions

The left-hand panel depicts the cdf of a discrete random variable. It appears in the form of a step function. By contrast, the right-hand panel shows the cdf of a continuous random variable. Its cdf is everywhere continuous.

○ ***List Form and Function Form***

The discrete random variable X depicted in Fig. 1 takes values 0, 1, 2, 3, with probability 0.48, 0.24, 0.16, 0.12, respectively. We can represent these details about X in two ways, namely *List Form* and *Function Form*. Table 1 gives List Form.

$P(X=x)$:	0.48	0.24	0.16	0.12
x:	0	1	2	3

Table 1: List Form

We enter List Form as:

```
f1 = {0.48, 0.24, 0.16, 0.12};

domain[f1] = {x, {0, 1, 2, 3}} && {Discrete};
```

Table 2 gives Function Form.

$$P(X=x) \;=\; \frac{12}{25\,(x+1)}; \quad x \in \{0, 1, 2, 3\}$$

Table 2: Function Form

We enter Function Form as:

```
f2 = 12 / (25 (x + 1));

domain[f2] = {x, 0, 3} && {Discrete};
```

Both List Form (`f1`) and Function Form (`f2`) express the same facts about X, and both are termed the *probability mass function* (pmf) of X. Notice especially the condition `{Discrete}` added to the `domain` statements. This is the device used to tell **mathStatica** that X is discrete. Importantly, appending the discreteness flag is *not* optional; if it is omitted, **mathStatica** will interpret the random variable as continuous.

The suite of **mathStatica** functions can operate on a pmf whether it is in List Form or Function Form — as we shall see repeatedly throughout this chapter. Here, for example, is a plot of the pmf of X from the List Form:

```
PlotDensity[f₁];
```

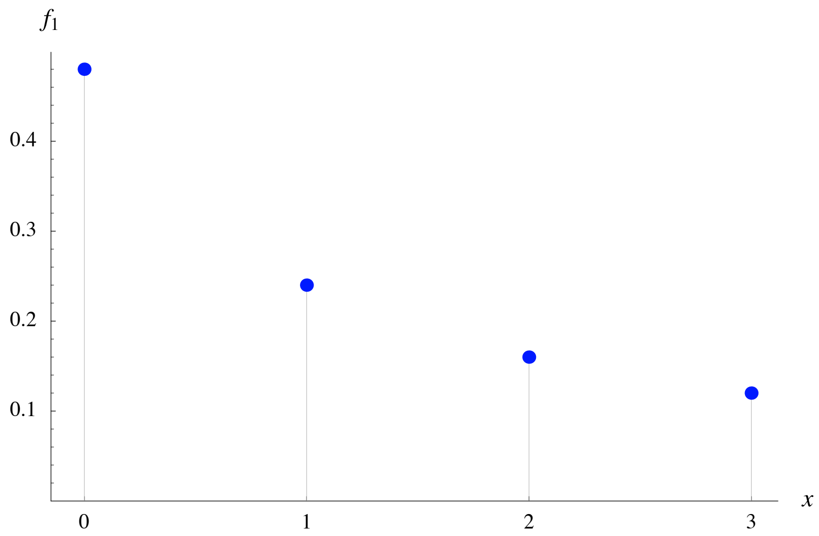


Fig. 2: The pmf of X

As a further illustration, here is the mean of X, using f_2:

```
Expect[x, f₂]
```

$$\frac{23}{25}$$

In general, the *expectation* of a function $u(X)$, where X is a discrete random variable, is defined as

$$E[u(X)] = \sum_x u(x)\, P(X = x) \tag{3.2}$$

where summation is over all values x of X. For example, here is $E[\cos(X)]$ using f_1:

```
Expect[Cos[x], f₁]
```

0.42429

§3.2 examines aspects of probability through the experiment of 'throwing' a die. §3.3 details the more popular discrete distributions encountered in practice (see the range provided in **mathStatica**'s *Discrete* palette). Mixture distributions are examined in §3.4, for they provide a means to generate many further distributions. Finally, §3.5 discusses pseudo-random number generation of discrete random variables.

There exist many fine references on discrete random variables. In particular, Fraser (1958) and Hogg and Craig (1995) provide introductory material, while Feller (1968, 1971) and Johnson *et al*. (1993) provide advanced treatments.

3.2 Probability: 'Throwing' a Die

The study of probability is often motivated by experiments such as coin tossing, drawing balls from an urn, and throwing dice. Many fascinating problems have been posed using these simple, easily replicable physical devices. For example, Ellsberg (1961) used urn drawings to show that there are (at least) two different types of uncertainty: one can be described by (the usual concept of) probability, the other cannot (ambiguity). More recently, Walley (1991, 1996) illustrated his controversial notion of imprecise probability by using drawings from a bag of marbles. Probability also attracts widespread popular interest: it can be used to analyse games of chance, and it can help us analyse many intriguing paradoxes. For further discussion of many popular problems in probability, see Mosteller (1987). For discussion of probability theory, see, for example, Billingsley (1995) and Feller (1968, 1971).

In this, and the next two sections, we examine discrete random variables whose domain of support is the set (or subset) of the non-negative integers. For discrete random variables of this type, there exist generating functions that can be useful for analysing a variable's properties. For a discrete random variable X taking non-negative integer values, the *probability generating function* (pgf) is defined as

$$\Pi(t) = E\left[t^X\right] = \sum_{x=0}^{\infty} t^x P(X = x) \tag{3.3}$$

which is a function of dummy variable t; it exists for any choice of $t \leq 1$. The pgf is similar to the moment generating function (mgf); indeed, subject to existence conditions (see §2.4 B), the mgf $M(t) = E[\exp(t X)]$ is equivalent to $\Pi(\exp(t))$. Likewise, $\Pi(\exp(i\, t))$ yields the characteristic function (cf). The pgf generates probabilities via the relation,

$$P(X = x) = \frac{1}{x!} \left. \frac{d^x \Pi(t)}{d t^x} \right|_{t=0} \qquad \text{for } x \in \{0, 1, 2, \ldots\}. \tag{3.4}$$

⊕ ***Example 1:*** Throwing a Die

Consider the standard six-sided die with faces labelled 1 through 6. If X denotes the upmost face resulting from throwing the die onto a flat surface, such as a table-top, then X may be thought of as a discrete random variable with pmf given in Table 3.

$P(X = x)$	:	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
x	:	1	2	3	4	5	6

Table 3: The pmf of X

This pmf presupposes that the die is fair. The pmf of X may be entered in either List Form:

```
f = Table[1/6, {6}];
domain[f] = {x, Range[6]} && {Discrete};
```

… or Function Form:

```
g = 1/6;
domain[g] = {x, 1, 6} && {Discrete};
```

The pgf of X may be derived from either representation of the pmf; for example, for the List Form representation:

```
pgf = Expect[t^x, f]
```

$$\frac{1}{6}\, t\,(1 + t + t^2 + t^3 + t^4 + t^5)$$

The probabilities can be recovered from the pgf using (3.4):

```
Table[1/x! D[pgf, {t, x}], {x, 1, 6}] /. t → 0
```

$$\left\{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right\}$$

⊕ ***Example 2:*** The Sum of Two Die Rolls

Experiments involving more than one die have often been contemplated. For example, in 1693, Samuel Pepys wrote to Isaac Newton seeking an answer to a die roll experiment, apparently posed by a Mr Smith. Smith's question concerned the chances of throwing a minimum number of sixes with multiple dice: at least 1 six from a throw of a box containing 6 dice, at least 2 sixes from another box containing 12 dice, and 3 or more sixes from a third box filled with 18 dice. We leave this problem as an exercise for the reader to solve (see §3.3 B for some clues). The correspondence between Newton and Pepys, including Newton's solution, is given in Schell (1960).

The experiment we shall pose is the sum S obtained from tossing two fair dice, X_1 and X_2. The outcomes of X_1 and X_2 are independent, and their distribution is identical to that of X given in *Example 1*. We wish to derive the pmf of $S = X_1 + X_2$. In order to do so, consider its pgf:

$$\Pi_S(t) = E\left[t^S\right] = E\left[t^{X_1 + X_2}\right].$$

By independence $E[t^{X_1 + X_2}] = E[t^{X_1}]\, E[t^{X_2}]$ and by identicality $E[t^{X_1}] = E[t^{X_2}] = E[t^X]$, so $\Pi_S(t) = E[t^X]^2$. In *Mathematica*, the pgf of S is simply:

```
pgfS = pgf^2
```

$$\frac{1}{36}\, t^2\,(1 + t + t^2 + t^3 + t^4 + t^5)^2$$

Now the domain of support of S is the integers from 2 to 12, so by (3.4) the pmf of S, say $h(s)$, in List Form, is:

```
h = Table[1/s! D[pgfS, {t, s}], {s, 2, 12}] /. t → 0
```

$$\left\{\frac{1}{36}, \frac{1}{18}, \frac{1}{12}, \frac{1}{9}, \frac{5}{36}, \frac{1}{6}, \frac{5}{36}, \frac{1}{9}, \frac{1}{12}, \frac{1}{18}, \frac{1}{36}\right\}$$

```
domain[h] = {s, Range[2, 12]} && {Discrete}
```

{s, {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}} && {Discrete}

⊕ ***Example 3:*** The Sum of Two Unfair Dice

Up until now, any die we have 'thrown' has been fair or, rather, assumed to be fair. It can be fun to consider the impact on an experiment when an unfair die is used! With an unfair die, the algebra of the experiment can rapidly become messy, but it is in such situations that *Mathematica* typically excels. There are two well-known methods of die corruption: loading it (attaching a weight to the inside of a face) and shaving it (slicing a thin layer from a face). In this example, we contemplate a shaved die. A shaved die is no longer a cube, and its total surface area is less than that of a fair die. Shaving upsets the relative equality in surface area of the faces. The shaved face, along with its opposing face, will have relatively more surface area than all the other faces. Consider, for instance, a die whose 1-face has been shaved. Then both the 1-face and the 6-face (opposing faces of a die sum to 7) experience no change in surface area, whereas the surface area of all the other faces is reduced.[1] Let us denote the increase in the probability of a 1 or 6 by δ. Then the probability of each of 2, 3, 4 and 5 must decrease by $\delta/2$ $(0 \le \delta < 1/3)$. The List Form pmf of X, a 1-face shaved die, is thus:

```
f = {1/6 + δ, 1/6 - δ/2, 1/6 - δ/2, 1/6 - δ/2, 1/6 - δ/2, 1/6 + δ};
domain[f] = {x, Range[6]} && {Discrete};
```

We now repeat the experiment given in *Example 2*, only this time we use dice which are 1-face shaved. We may derive the List Form pmf of S exactly as before:

```
pgf = Expect[t^x, f];        pgfS = pgf^2;
h = Table[1/s! D[pgfS, {t, s}], {s, 2, 12}] /. t → 0 //
  Simplify
```

$$\left\{\frac{1}{36}(1+6\delta)^2, \frac{1}{18}+\frac{\delta}{6}-\delta^2, \frac{1}{12}-\frac{3\delta^2}{4}, \frac{1}{18}(2-3\delta-9\delta^2),\right.$$
$$\frac{1}{36}(5-12\delta-9\delta^2), \frac{1}{6}+3\delta^2, \frac{1}{36}(5-12\delta-9\delta^2),$$
$$\left.\frac{1}{18}(2-3\delta-9\delta^2), \frac{1}{12}-\frac{3\delta^2}{4}, \frac{1}{18}+\frac{\delta}{6}-\delta^2, \frac{1}{36}(1+6\delta)^2\right\}$$

```
domain[h] = {s, Range[2, 12]} && {Discrete};
```

Figure 3 depicts the pmf of S using both fair and unfair dice. Both distributions are symmetric about their mean, 7, with a greater probability with shaved 1-face dice of sums of 2, 7 and 12. Moreover, as the distribution now appears fatter-tailed under shaved 1-face dice, we would expect the variability of its distribution to increase — a fact that can be verified by executing `Var[s, h /. δ → {0, 0.1}]`.

```
PlotDensity[h /. δ → {0, 0.1}];
```

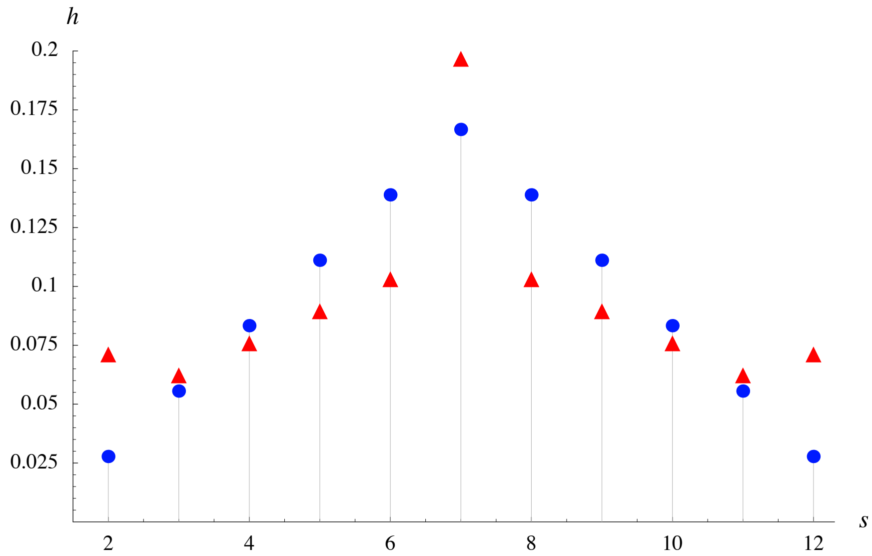


Fig. 3: The pmf of S for fair dice (●) and 1-face shaved dice (▲)

⊕ ***Example 4:*** The Game of Craps

The game of craps is a popular casino game. It involves a player throwing two dice, one or more times, until either a win or a loss occurs. The player wins on the first throw if the dice sum is 7 or 11. The player loses on the first throw if a sum of either 2, 3 or 12 occurs. If on the first throw the player neither wins nor loses, then the sum of the dice is referred to as the *point*. The game proceeds with the player throwing the dice until either the point occurs, in which case the player wins, or a sum of 7 occurs, in which case the player loses. When the dice are fair, it can be shown that the probability of the player winning is $244/495 \simeq 0.49293$.

It is an interesting task to verify the probability of winning the game with *Mathematica*. However, for the purposes of this example, we use simulation methods to estimate the probability of winning. The following inputs combine to simulate the outcome of one game — returning 1 for a win, and 0 for a loss. First, here is a simple function that simulates the roll of a fair die:

```
TT := Random[Integer, {1, 6}]
```

Next is a function that simulates the first throw of the game, deciding whether to stop or simulate further throws:

```
Game := (s = TT + TT;
        Which[s == 7 || s == 11,          1,
   Ω          s == 2 || s == 3 || s == 12,  0,
              True,                        MoreThrows[s]])
```

Finally, if more than one throw is needed to complete the game:

```
MoreThrows[p_] := (s = TT + TT;
                   Which[s == p,   1,
                         s == 7,   0,
                         True,   MoreThrows[p]])
```

Notice that the `MoreThrows` function calls itself if a win or loss does not occur. In practice this will not result in an infinite recurrence because the probability that the game continues forever is zero. Let our estimator be the proportion of wins across a large number of games. Here is a simulated estimate of the probability of winning a game:

```
SampleMean[ Table[Game, {100000}] ] // N
```

```
0.49162
```

As a further illustration of simulation, suppose that a gambler starting with an initial fortune of $5 repeatedly wagers $1 against an infinitely rich opponent — the House — until his fortune is lost. Assuming that a win pays 1 to 1, the progress of his fortune from one game to the next can be represented by the function:

```
fortune[x_] := x - 1 + 2 Game
```

For example, here is one particular sequence of 10 games:

```
NestList[fortune, 5, 10]
```

```
{5, 4, 5, 4, 3, 4, 5, 6, 5, 4, 3}
```

After these games, his fortune has dropped to $3, but as he is not yet ruined, he can still carry on gaming! Now suppose we wish to determine how many games the player can expect to play until ruin. To solve this, we take as our estimator the average length of a large number of matches. Here we simulate just 100 matches, and measure the length of each match:

```
matchLength = Table[
 NestWhileList[fortune, 5, Positive] // Length, {100}] - 1
```

```
{4059, 7, 37, 3193, 5, 5, 171, 45, 35, 15, 61, 573, 15, 125, 39, 67, 33,
 13, 73, 11, 287, 27, 89, 49, 13, 3419, 2213, 4081, 11, 89, 697, 127,
 179, 125, 33, 31, 9, 59, 973, 51, 5, 53, 613, 13, 13, 19, 19, 105, 53,
 29, 163, 561, 107, 11, 25, 5, 435, 35, 7, 21, 27, 33, 19, 147, 61, 339,
 101, 53, 239, 51, 23, 23, 403, 439, 6327, 7, 85, 5, 35, 107, 125, 49,
 83, 33, 17, 439, 29, 15, 49, 9, 103, 13, 35, 43, 107, 145, 9, 45, 27, 81}
```

Then our estimate is:

```
SampleMean[matchLength] // N
```

```
334.16
```

In fact, the simulator estimator has performed reasonably well in this small trial, for the exact solution to the expected number of games until ruin can be shown to equal:

$$5\Big/\left(\frac{251}{495}-\frac{244}{495}\right)\ //\ N$$

```
353.571
```

For details on the gambler's ruin problem see, for example, Feller (1968, Chapter 14). ■

3.3 Common Discrete Distributions

This section presents a series of discrete distributions frequently applied in statistical practice: the Bernoulli, Binomial, Poisson, Geometric, Negative Binomial, and Hypergeometric distributions. Each distribution can be input into *Mathematica* with the **mathStatica** *Discrete* palette. The domain of support for all of these distributions is the set (or subset) of non-negative integers.

3.3 A The Bernoulli Distribution

The Bernoulli distribution (named for Jacques Bernoulli, 1654–1705) is a fundamental building block in statistics. A Bernoulli distributed random variable has a two-point support, 0 and 1, with probability p that it takes the value 1, and probability $1-p$ that it is zero-valued. Experiments with binary outcomes induce a Bernoulli distributed random variable; for example, the ubiquitous coin toss can be coded 0 = tail and 1 = head, with probability one-half ($p=\frac{1}{2}$) assigned to each outcome if the coin is fair.

If X is a Bernoulli distributed random variable, its pmf is given by $P(X=x)=p^x(1-p)^{1-x}$, where $x\in\{0, 1\}$, and parameter p is such that $0<p<1$; p is often termed the success probability. From **mathStatica**'s *Discrete* palette:

$$\begin{aligned} &f\ =\ p^x\ (1-p)^{1-x}; \\ &\text{domain}[f] = \{x,\ 0,\ 1\}\ \&\&\ \{0<p<1\}\ \&\&\ \{\text{Discrete}\}; \end{aligned}$$

For example, the mean of X is:

```
Expect[x, f]
```

```
p
```

Although simple in structure, the Bernoulli distribution forms the backbone of many important statistical models encountered in practice.

⊕ ***Example 5:*** A Logit Model for the Bernoulli Response Probability

Suppose that sick patients are given differing amounts of a curative drug, and they respond to treatment after a fixed period of time as either 1 = cured or 0 = sick. Assume response $X \sim \text{Bernoulli}(p)$. Let y denote the amount of the drug given to a patient. Presumably the probability p that a patient is cured depends on y, all other factors held fixed. This probability is assumed to be governed, or modelled, by the logit relation:

$$p = \frac{1}{1 + e^{-(\alpha + \beta y)}};$$

Here, $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ are unknown parameters whose values we wish to deduce or estimate. This will enable us to answer, for example, the following type of question: "If a patient receives a dose y^*, what is his chance of cure?". To illustrate, here is a set of artificial data on $n = 20$ patients:

$x = 0$:	7	17	14	3	15	19	11	6	20	12
$x = 1$:	46	33	19	32	43	34	51	16	35	30

Table 4: Dosage given (artificial data)

At the end of the treatment, 10 patients responded 0 = sick (the top row), while the remaining 10 patients were cured (the bottom row). The dosage y that each patient received appears in the body of the table. We may enter this data as follows:

```
dose = { 7, 17, 14,  3, 15, 19, 11,  6, 20, 12,
        46, 33, 19, 32, 43, 34, 51, 16, 35, 30};

response = { 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
             1, 1, 1, 1, 1, 1, 1, 1, 1, 1};
```

The observed log-likelihood function is (see Chapter 11 and Chapter 12 for further details):

```
obslogL = Log[Times @@ (f /. {y → dose, x → response})];
```

We use `FindMaximum` to find the maximum of the log-likelihood with respect to values for the unknown parameters:

```
sol = FindMaximum[obslogL, {α, 0}, {β, 0}][[2]]

{α → -7.47042, β → 0.372755}
```

Given the data, the parameters of the model (α and β) have been estimated at the indicated values. The fitted model for the probability p of a cure as a function of dosage level y is given by:

```
p /. sol
```

$$\frac{1}{1+e^{7.47042-0.372755\,y}}$$

The fitted p (the smooth curve) along with the data (the circled points) are plotted in Fig. 4.

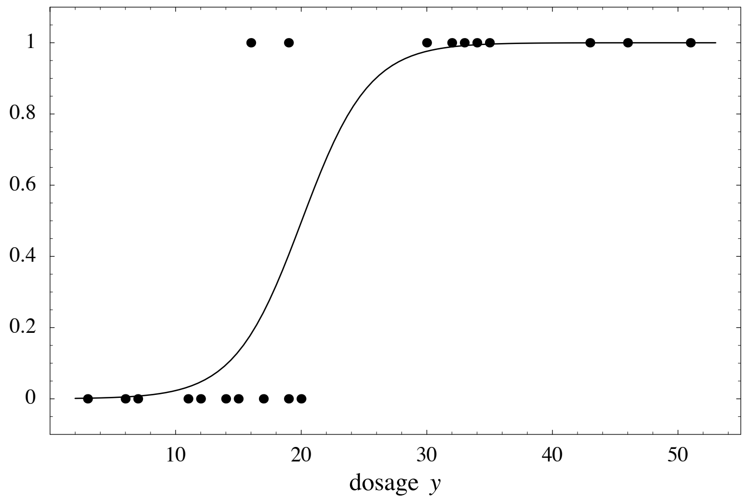


Fig. 4: Data and fitted p

Evidently, the fitted curve shows that if a patient receives a dosage of 20 units of the drug, then they have almost a 50% chance of cure (execute `Solve[(p/.sol)==0.5,y]`). Of course, room for error must exist when making a statement such as this, for we do not know the true values of the parameters α and β, nor whether the logistic formulation is the correct functional form.

```
Clear[p]
```

3.3 B The Binomial Distribution

Let $X_1, X_2, \ldots, X_n$ be n mutually independent and identically distributed Bernoulli(p) random variables. The discrete random variable formed as the sum $X = \sum_{i=1}^{n} X_i$ is distributed as a Binomial random variable with index n and success probability p, written $X \sim$ Binomial(n, p); the domain of support of X is the integers (0, 1, 2, …, n). The pmf and its support may be entered directly from **mathStatica**'s *Discrete* palette:

```
        f  =  Binomial[n, x] p^x (1 - p)^(n-x);
domain[f] =  {x, 0, n} &&
             {0 < p < 1, n > 0, n ∈ Integers} && {Discrete};
```

The Binomial derives its name from the expansion of $(p+q)^n$, where $q = 1 - p$. Here is the graph of the pmf, with p fixed at 0.4 and the index n taking values 10 (circles) and 20 (triangles):

```
PlotDensity[f /. {p → 0.4, n → {10, 20}}];
```

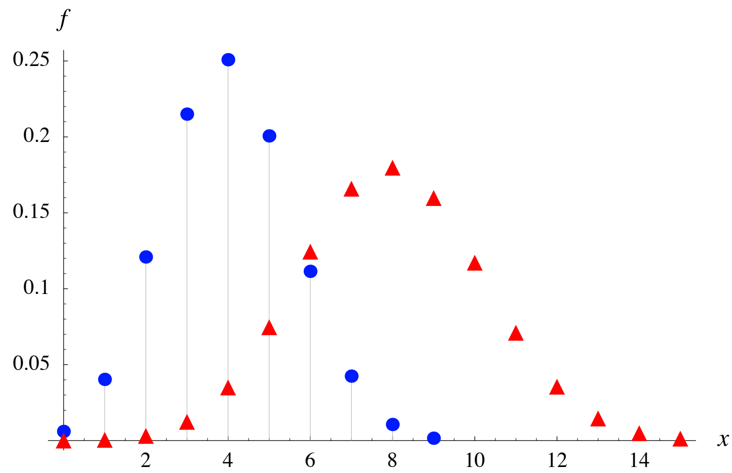


Fig. 5: Probability mass functions of X: $n = 10$, $p = 0.4$ (●); $n = 20$, $p = 0.4$ (▲)

The Binomial cdf, $P(X \leq x)$ for $x \in \mathbb{R}$, appears complicated:

```
Prob[x, f]
```

```
1 -
 ((1 - p)^(-1+n-Floor[x]) p^(1+Floor[x]) Γ[1 + n] Hypergeometric2F1[1,
     1 - n + Floor[x], 2 + Floor[x], p/(-1 + p)]) /
  (Γ[n - Floor[x]] Γ[2 + Floor[x]])
```

Figure 6 plots the cdf—it has the required step function appearance.

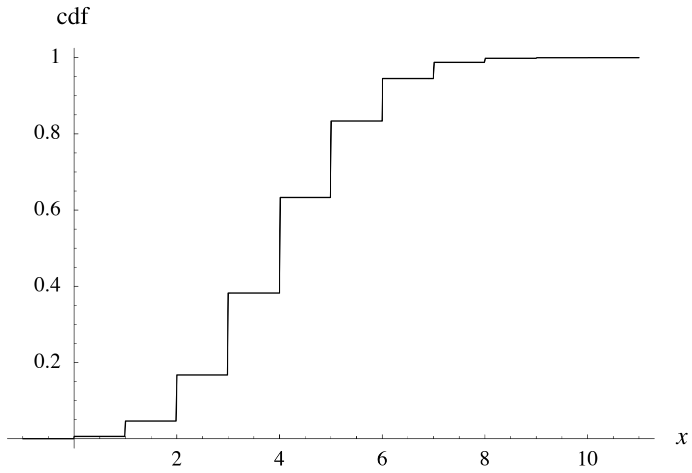


Fig. 6: The cdf of X: $n = 10$, $p = 0.4$

The mean, variance and other higher order moments of a Binomial random variable may be computed directly using `Expect`. For example, the mean $E[X]$ is:

```
μ = Expect[x, f]
```

```
n p
```

The variance of X is given by:

```
v = Var[x, f]
```

```
-n (-1 + p) p
```

Although the expression for the variance has a minus sign at the front, the variance is strictly positive because of the restriction on p.

Moments may also be obtained via a generating function method. Here, for example, is the central moment generating function $E[\exp(t(X-\mu))] = e^{-t\mu} E[\exp(tX)]$. In **mathStatica**:

```
mgfc = ⅇ^(-t μ) Expect[ⅇ^(t x), f]
```

```
ⅇ^(-n p t) (1 + (-1 + ⅇ^t) p)^n
```

Using `mgfc`, the i^{th} central moment $\mu_i = E[(X-\mu)^i]$ is obtained by differentiation with respect to t (i times), and then setting t to zero. To illustrate, when computing Pearson's measure of kurtosis $\beta_2 = \mu_4 / \mu_2^2$:

```
D[mgfc, {t, 4}] /. t → 0
------------------------  // FullSimplify
           v^2
```

```
-1 + 3 (-2 + n) (-1 + p) p
--------------------------
       n (-1 + p) p
```

```
ClearAll[μ, v]
```

The Binomial distribution has a number of linkages to other statistical distributions. For example, if $X \sim \text{Binomial}(n, p)$ with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$, then the standardised discrete random variable

$$Y = \frac{X - np}{\sqrt{np(1-p)}}$$

has a limiting $N(0, 1)$ distribution, as n becomes large. In some settings, the Binomial distribution is itself a limiting distribution; *cf.* the Ehrenfest Urn. The Binomial distribution is also linked to another common discrete distribution — the Poisson distribution — which is discussed in §3.3 C.

⊕ ***Example 6:*** The Ehrenfest Urn

In physics, the Ehrenfest model describes the exchange of heat between two isolated bodies. In probabilistic terms, the model can be formulated according to urn drawings. Suppose there are two urns, labelled A and B, containing, in total, m balls. Starting at time $t = 0$ with some initial distribution of balls, the experiment proceeds at each $t \in \{1, 2, \ldots\}$ by randomly drawing a ball (from the entire collection of m balls) and then moving it from its present urn into the other. This means that if urn A contains $k \in \{0, 1, 2, \ldots, m\}$ balls (so B contains $m - k$ balls), and if the chosen ball is in urn A, then there are now $k - 1$ balls in A and $m - k + 1$ in B. On the other hand, if the chosen ball was in B, then there are now $k + 1$ balls in A and one fewer in B. Let X_t denote the number of balls in urn A at time t. Then, X_{t+1} depends only on X_t, its value being either one more or one less. Because each variable in the sequence $\{X_t\} = (X_1, X_2, X_3, \ldots)$ depends only on its immediate past, $\{X_t\}$ is said to form a Markov chain. The *conditional* pmf of X_{t+1}, given that $X_t = k$, appears Bernoulli-like, with support points $k + 1$ and $k - 1$.

When the chosen ball comes from urn B, we have

$$P\big(X_{t+1} = k + 1 \,\big|\, X_t = k\big) \;=\; \frac{m - k}{m} \;=\; 1 - \frac{k}{m} \tag{3.5}$$

while if the chosen ball comes from urn A, we have

$$P\big(X_{t+1} = k - 1 \,\big|\, X_t = k\big) \;=\; \frac{k}{m}\,. \tag{3.6}$$

The so-called limiting distribution of the sequence $\{X_t\}$ is often of interest; it is sometimes termed the long-run *unconditional* pmf of X_t.[2] It is given by the list of probabilities $p_0, p_1, p_2, \ldots, p_m$, and may be found by solving the simultaneous equation system,

$$p_k \;=\; \sum_{j=0}^{m} p_j\, P\big(X_{t+1} = k \mid X_t = j\big), \qquad k \in \{0, 1, 2, \ldots, m\} \tag{3.7}$$

along with the adding-up condition,

$$p_0 + p_1 + p_2 + \cdots + p_m = 1. \tag{3.8}$$

Substituting (3.5) and (3.6) into equations (3.7) yields, with some work, the equation system written as a function of m:

```
Ehrenfest[m_] := Join[
  Table[p_k == (1 - (k - 1)/m) p_(k-1) + (k + 1)/m p_(k+1), {k, 1, m - 1}],
  {p_0 == p_1/m, p_m == p_(m-1)/m, Sum[p_i, {i, 0, m}] == 1}]
```

To illustrate, let $m = 5$ be the total number of balls distributed between the two urns. The long-run pmf is obtained as follows:

```
Solve[Ehrenfest[5], Table[p_i, {i, 0, 5}]]
```

$$\left\{\left\{p_0 \to \frac{1}{32},\ p_1 \to \frac{5}{32},\ p_2 \to \frac{5}{16},\ p_3 \to \frac{5}{16},\ p_4 \to \frac{5}{32},\ p_5 \to \frac{1}{32}\right\}\right\}$$

Now consider the Binomial(m, $\frac{1}{2}$) distribution, whose pmf is given by:

$$f = \text{Binomial}[m, x] \left(\frac{1}{2}\right)^m;$$

```
domain[f] = {x, 0, m} && {Discrete};
```

Computing all probabilities finds:

```
Table[f /. m → 5, {x, 0, 5}]
```

$$\left\{\frac{1}{32}, \frac{5}{32}, \frac{5}{16}, \frac{5}{16}, \frac{5}{32}, \frac{1}{32}\right\}$$

which is equivalent to the limiting distribution of the Ehrenfest Urn when $m = 5$. In fact, for arbitrary m, the limiting distribution of the Ehrenfest Urn is Binomial(m, $\frac{1}{2}$). ■

3.3 C The Poisson Distribution

The Poisson distribution is an important discrete distribution, with vast numbers of applications in statistical practice. It is particularly relevant when the event of interest has a small chance of occurrence amongst a large population; for example, the daily number of automobile accidents in Los Angeles, where there are few accidents relative to the total number of trips undertaken. In fact, a link between the Binomial distribution and the Poisson can be made by allowing the Binomial index n to become large and the success probability p to become small, but simultaneously maintaining finiteness of the mean (see *Example 2* of Chapter 8). The Poisson often serves as an approximation to the Binomial distribution. For detailed material on the Poisson distribution see, amongst others, Haight (1967) and Johnson *et al*. (1993, Chapter 4).

A discrete random variable X with pmf:

$$f = \frac{e^{-\lambda} \lambda^x}{x!};$$

```
domain[f] = {x, 0, ∞} && {λ > 0} && {Discrete};
```

is said to be Poisson distributed with parameter $\lambda > 0$; in short, $X \sim$ Poisson(λ). Figure 7 plots the pmf when $\lambda = 5$ and $\lambda = 10$.

```
PlotDensity[f /. λ → {5, 10}];
```

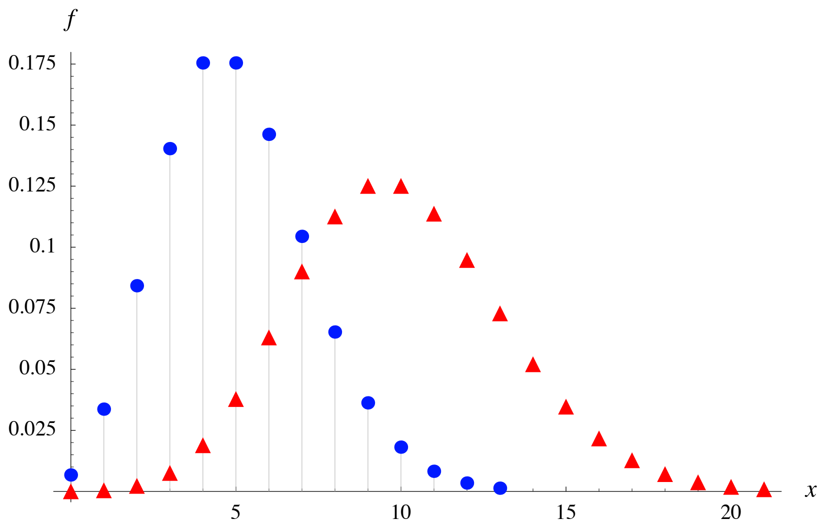


Fig. 7: Poisson pmf: $\lambda = 5$ (●), $\lambda = 10$ (▲)

From the graph, we see that the mass of the distribution shifts to the right as parameter λ increases. The Poisson mean is given by:

```
Expect[x, f]
```

λ

Curiously, the Poisson variance:

```
Var[x, f]
```

λ

… is identical to the mean. This feature alone serves to distinguish the Poisson from many other discrete distributions. Moreover, all cumulants of X are equal to λ, as can be shown by using the cumulant generating function:

```
cgf = Log[Expect[e^(t x), f]]
```

$\text{Log}[e^{(-1+e^t)\lambda}]$

To illustrate, here are the first 10 cumulants of X:

```
Table[D[cgf, {t, i}], {i, 10}] /. t → 0
```

{λ, λ, λ, λ, λ, λ, λ, λ, λ, λ}

⊕ ***Example 7:*** Probability Calculations

Let $X \sim$ Poisson(4) denote the number of ships arriving at a port each day. Determine:
(i) the probability that four or more ships arrive on a given day, and
(ii) repeat part (i) knowing that at least one ship arrives.

Solution: Begin by entering in X's details:

$$f = \frac{e^{-\lambda}\,\lambda^x}{x\,!} \;/.\; \lambda \to 4; \qquad \text{domain}[f] = \{x,\ 0,\ \infty\}\ \&\&\ \{\text{Discrete}\};$$

(i) The required probability simplifies to $P(X \geq 4) = 1 - P(X \leq 3)$. Thus:

```
pp = 1 - Prob[3, f] // N
```

```
0.56653
```

(ii) We require the conditional probability $P(X \geq 4 \mid X \geq 1)$. For two events A and B, the conditional probability of A given B is defined as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad \text{provided } P(B) > 0.$$

In our case, $A = \{X \geq 4\}$ and $B = \{X \geq 1\}$, so $A \cap B = \{X \geq 4\}$. We already have $P(X \geq 4)$, and $P(X \geq 1)$ may be found in the same manner. The conditional probability is thus:

$$\frac{\text{pp}}{1 - \text{Prob}[0,\ f]}$$

```
0.5771
```

⊕ ***Example 8:*** A Conditional Expectation

Suppose $X \sim$ Poisson(λ). Determine the conditional mean of X, given that X is odd-valued.

Solution: Enter in the details of X:

$$f = \frac{e^{-\lambda}\,\lambda^x}{x\,!}; \quad \text{domain}[f] = \{x,\ 0,\ \infty\}\ \&\&\ \{\lambda > 0\}\ \&\&\ \{\text{Discrete}\};$$

We require $E[X \mid X \in \{1, 3, 5, \ldots\}]$. This requires the pmf of $X \mid X \in \{1, 3, 5, \ldots\}$; namely, the distribution of X given that X is odd-valued, which is given by:

$$f1 = \frac{f}{\text{Sum}[\text{Evaluate}[f],\ \{x,\ 1,\ \infty,\ 2\}]};$$

```
domain[f1] = {x, 1, ∞, 2} && {λ > 0} && {Discrete};
```

Then, the required expectation is:

```
Expect[x, f1]
```

```
λ Coth[λ]
```

3.3 D The Geometric and Negative Binomial Distributions

○ ***The Geometric Distribution***

A Geometric experiment has similar properties to a Binomial experiment, *except* that the experiment is stopped when the first success is observed. Let p denote the probability of success in repeated independent Bernoulli trials. We are now interested in the probability that the *first* success occurs on the x^{th} trial. Then X is said to be a Geometric random variable with pmf:

$$P(X = x) = p(1-p)^{x-1}, \quad x \in \{1, 2, 3, \ldots\}, \quad 0 < p < 1. \tag{3.9}$$

This can be entered with **mathStatica**'s *Discrete* palette:

```
f = p (1 - p)^(x-1);
domain[f] = {x, 1, ∞} && {0 < p < 1} && {Discrete};
```

Here, for example, is a plot of the pmf when $p = 0.6$:

```
PlotDensity[f /. p → .6, AxesOrigin → {3 / 4, 0}];
```

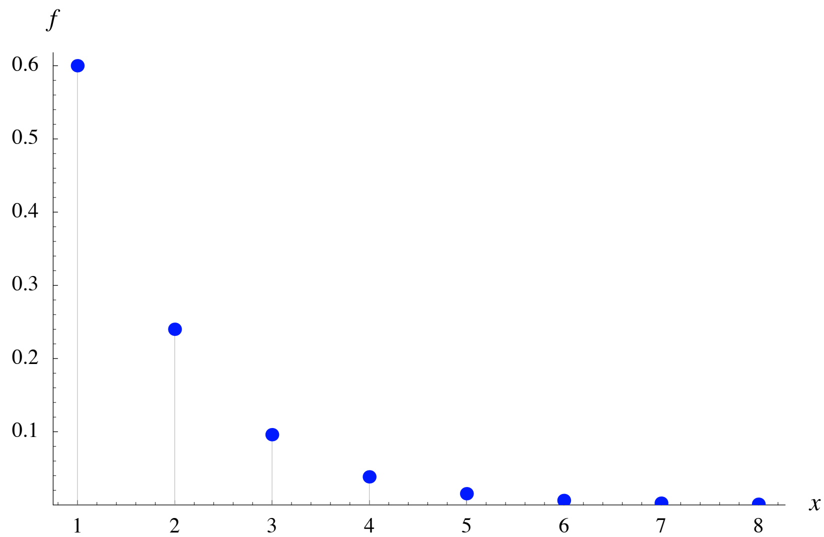


Fig. 8: The pmf of the Geometric distribution ($p = 0.6$)

○ ***The Waiting-Time Negative Binomial Distribution***

A Waiting-Time Negative Binomial experiment has similar properties to the Geometric experiment, *except* that the experiment is now stopped when a *fixed* number of successes occur. As before, let p denote the probability of success in repeated independent Bernoulli trials. Of interest is the probability that the r^{th} success occurs on the y^{th} trial. Then Y is a Waiting-Time Negative Binomial random variable with pmf,

$$P(Y = y) \;=\; \binom{y-1}{r-1} p^r (1-p)^{y-r} \tag{3.10}$$

for $y \in \{r, r+1, r+2, \ldots\}$ and $0 < p < 1$. We enter this as:

```
         h  = Binomial[y - 1, r - 1] p^r (1 - p)^(y-r);
domain[h] = {y, r, ∞} && {0 < p < 1, r > 0} && {Discrete};
```

The mean $E[Y]$ and variance are, respectively:

```
Expect[y, h]
```

$\frac{r}{p}$

```
Var[y, h]
```

$\frac{r - p\,r}{p^2}$

○ ***The Negative Binomial Distribution***

As its name would suggest, the Waiting-Time Negative Binomial distribution (3.10) is closely related to the Negative Binomial distribution. In fact, the latter may be obtained from the former by transforming $Y \to X$, such that $X = Y - r$:

```
f = Transform[x == y - r, h]
```

$(1-p)^x\, p^r$ Binomial[-1 + r + x, -1 + r]

with domain:

```
domain[f] = TransformExtremum[x == y - r, h]
```

{x, 0, ∞} && {0 < p < 1, r > 0} && {Discrete}

as given in the *Discrete* palette. When r is an integer, the distribution is sometimes known as the Pascal distribution. Here is its pgf:

```
Expect[t^x, f]
```

$p^r\, (1 + (-1 + p)\, t)^{-r}$

3.3 E The Hypergeometric Distribution

```
ClearAll[T, n, r, x]
```

Urn models in which balls are repeatedly drawn *without replacement* lead to the Hypergeometric distribution. This contrasts to sampling *with replacement* which leads to the Binomial distribution. To illustrate the former, suppose that an urn contains a total of T balls, r of which are red ($1 \le r < T$). The experiment proceeds by drawing one-by-one a sample of n balls from the urn without replacement ($1 \le n < T$).[3] Interest lies in determining the pmf of X, where X is the number of red balls drawn.

The domain of support of X is $x \in \{0, 1, \ldots, \min(n, r)\}$, where $\min(n, r)$ denotes the smaller of n and r. Next, consider the probability of a particular sequence of n draws, namely x red balls followed by $n - x$ other colours:

$$\left(\frac{r}{T} \times \frac{r-1}{T-1} \times \cdots \times \frac{r-x+1}{T-x+1}\right)\left(\frac{T-r}{T-x} \times \frac{T-r-1}{T-x-1} \times \cdots \times \frac{T-r-(n-x-1)}{T-x-(n-x-1)}\right)$$

$$= \frac{r!}{(r-x)!} \; \frac{(T-r)!}{(T-r-n+x)!} \; \frac{(T-n)!}{T!}$$

$$= \binom{T-n}{r-x} \Big/ \binom{T}{r}.$$

In total, there are $\binom{n}{x}$ arrangements of x red balls amongst the n drawn, each having the above probability. Hence, the pmf of X is

$$f(x) = \binom{n}{x}\binom{T-n}{r-x} \Big/ \binom{T}{r}$$

where $x \in \{0, 1, \ldots, \min(n, r)\}$. We may enter the pmf of X as:

```
f = Binomial[n, x] Binomial[T - n, r - x] / Binomial[T, r];

domain[f] = {x, 0, n} && {Discrete};
```

We have set `domain[f]={x,0,n}`, rather than `{x,0,Min[n,r]}`, because **mathStatica** does not support the latter. This alteration does not affect the pmf.[4]

The Hypergeometric distribution gets its name from the appearance of the Gaussian hypergeometric function in its pgf:

```
pgf = Expect[t^x, f]

(Γ[1 - n + T] Hypergeometric2F1Regularized[-n,
    -r, 1 - n - r + T, t]) / (Binomial[T, r] Γ[1 + r])
```

Here are the mean and variance of X:

```
Expect[x, f]
```

$$\frac{n\,r}{T}$$

```
Var[x, f]
```

$$\frac{n\,r\,(n-T)\,(r-T)}{(-1+T)\,T^2}$$

⊕ ***Example 9:*** The Number of Ace Cards

Obtain the pmf of the distribution of the number of ace cards in a poker hand. Then plot it.

Solution: In this example, the 'urn' is the deck of $T = 52$ playing cards, and the 'red balls' are the ace cards, so $r = 4$. There are $n = 5$ cards in a hand. Therefore, the pmf of the number of ace cards in a poker hand is given by:

```
Table[f /. {T → 52, n → 5, r → 4}, {x, 0, 4}]
```

$$\left\{\frac{35673}{54145}, \frac{3243}{10829}, \frac{2162}{54145}, \frac{94}{54145}, \frac{1}{54145}\right\}$$

which we may plot as:

```
PlotDensity[f /. {T → 52, n → 5, r → 4}, {x, 0, 4},
  AxesOrigin → {-0.25, 0}];
```

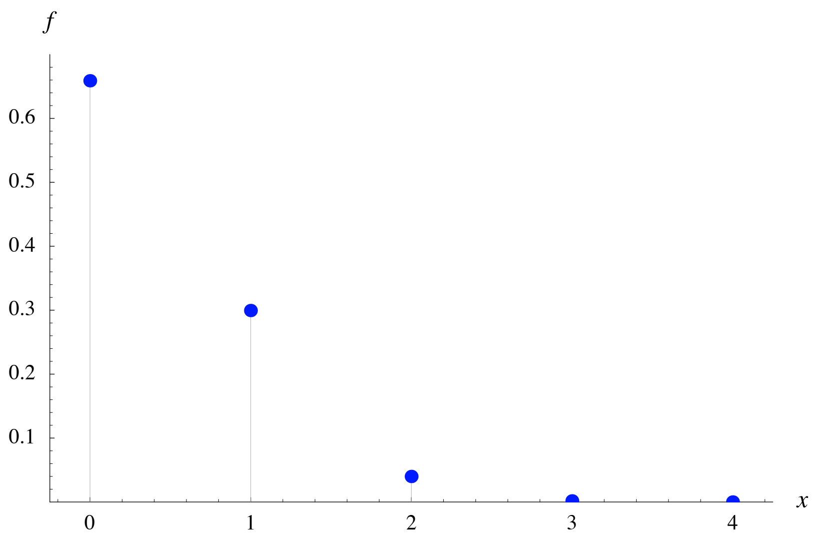


Fig. 9: The pmf of X, the number of ace cards in a poker hand

3.4 Mixing Distributions

At times, it may be necessary to use distributions with unusual characteristics, such as long-tailed behaviour or multimodality. Unfortunately, it can be difficult to write down from scratch the pdf/pmf of a distribution with the desired characteristic. Fortunately, progress can usually be made with the method of *mixing distributions*. Two prominent approaches to mixing are presented here: component-mixing (§3.4 A) and parameter-mixing (§3.4 B). The first type, component-mixing, forms distributions from linear combinations of other distributions. It is a method well-suited for generating random variables with multimodal distributions. The second type, parameter-mixing, relaxes the assumption of fixed parameters, allowing them to vary according to some specified distribution.

3.4 A Component-Mix Distributions

Component-mix distributions are formed from linear combinations of distributions. To fix notation, let the pmf of a discrete random variable X_i be $f_i(x) = P(X_i = x)$ for $i = 1, \ldots, n$, and let ω_i be a constant such that $0 < \omega_i < 1$ and $\sum_{i=1}^{n} \omega_i = 1$. The linear combination of the component random variables defines the n-component-mix random variable,

$$X \sim \omega_1 X_1 + \omega_2 X_2 + \cdots + \omega_n X_n \tag{3.11}$$

and its pmf is given by the weighted average

$$f(x) = \sum_{i=1}^{n} \omega_i f_i(x). \tag{3.12}$$

Importantly, the domain of support of X is taken to be all points x contained in the union of support points of the component distributions.[5] Titterington *et al.* (1985) deals extensively with distributions formed from component-mixes.

⊕ ***Example 10:*** A Poisson Two-Component-Mix

Let $X_1 \sim \text{Poisson}(2)$ and $X_2 \sim \text{Poisson}(10)$ be independent, and set $\omega_1 = \omega_2 = \frac{1}{2}$. Plot the pmf of the two-component-mix $X \sim \omega_1 X_1 + \omega_2 X_2$.

Solution: The general form of the pmf of X can be entered directly from (3.12):

$$f_1 = \frac{e^{-\theta}\,\theta^x}{x!}; \qquad f_2 = \frac{e^{-\lambda}\,\lambda^x}{x!}; \qquad f = \omega_1 f_1 + \omega_2 f_2;$$

As both X_1 and X_2 are supported on the set of non-negative integers, then this is also the domain of support of X. As the parameter restrictions are unimportant in this instance, the domain of support of X may be entered into *Mathematica* simply as:

```
domain[f] = {x, 0, ∞} && {Discrete};
```

The plot we require is:

```
PlotDensity[f /. {θ → 2, λ → 10, ω₁ → 1/2, ω₂ → 1/2}];
```

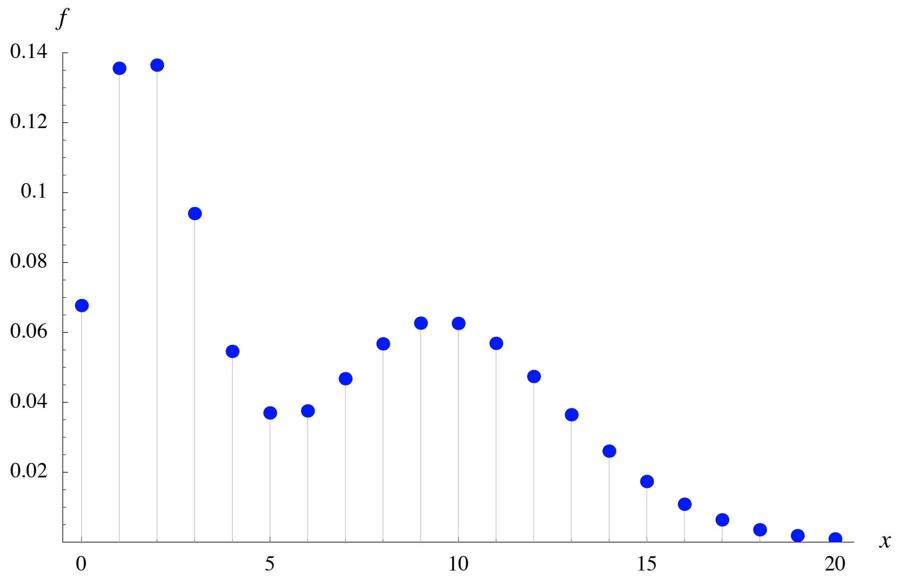


Fig. 10: The pmf of a Poisson–Poisson component-mix

For our chosen mixing weights, the pmf of X is bimodal, a feature not shared by either of the components. We return to the Poisson two-component-mix distribution in §12.6, where maximum likelihood estimation of its parameters is considered. ■

○ *Zero-Inflated Distributions*

A survey of individual consumption patterns can often return an excessively large number of zero observations on consumption of items such as cigarettes. *Zero-Inflated distributions* can be used to model such variables. They are just a special case of (3.11), and are formed from the two-component-mix,

$$X \sim \omega_1 X_1 + \omega_2 X_2 = (1-\omega) X_1 + \omega X_2 \tag{3.13}$$

where, because $\omega_1 + \omega_2 = 1$, we can express the mix with a single weight ω. In this component-mix, zero-inflated distributions correspond to nominating X_1 as a degenerate distribution with all its mass at the origin; that is, $P(X_1 = 0) = 1$. The distribution of X is therefore a modification of the distribution of X_2. If the domain of support of X_2 does not include zero, then this device serves to add zero to the domain of support of X. On the other hand, if X_2 can take value zero, then $P(X = 0) > P(X_2 = 0)$ because ω is such that $0 < \omega < 1$. In both scenarios, the probability of obtaining a zero is boosted.

⊕ ***Example 11:*** The Zero-Inflated Poisson Distribution

Consider the two-component-mix (3.13) with $P(X_1 = 0) = 1$, and $X_2 \sim \text{Poisson}(\lambda)$. In this case, X has the so-called Zero-Inflated Poisson distribution, or ZIP for short. The pmf of X is

$$P(X = x) = \begin{cases} 1 - \omega + \omega e^{-\lambda} & \text{if } x = 0 \\ \omega \dfrac{e^{-\lambda} \lambda^x}{x!} & \text{if } x \in \{1, 2, \ldots\} \end{cases}$$

where $0 < \omega < 1$ and $\lambda > 0$. To obtain the pgf, we only require the pgf of X_2, denoted $\Pi_2(t)$, since

$$\begin{aligned} \Pi(t) &= \sum_{x=0}^{\infty} t^x P(X = x) \\ &= (1 - \omega) + \omega \sum_{x=0}^{\infty} t^x P(X_2 = x) \\ &= (1 - \omega) + \omega \Pi_2(t). \end{aligned} \tag{3.14}$$

For our example, the pgf of $X_2 \sim \text{Poisson}(\lambda)$ is:

```
f₂ = e^-λ λ^x / x! ;
domain[f₂] = {x, 0, ∞} && {λ > 0} && {Discrete};

pgf₂ = Expect[t^x, f₂]

e^((-1+t) λ)
```

Then, by (3.14), the pgf of X is:

```
pgf = (1 - ω) + ω pgf₂

1 - ω + e^((-1+t) λ) ω
```

Taking, for example, $\omega = 0.5$ and $\lambda = 5$, $P(X = 0)$ is quite substantial:

```
pgf /. {ω → 0.5, λ → 5, t → 0}

0.503369
```

… when compared to the same chance for its Poisson component alone:

```
pgf₂ /. {λ → 5, t → 0} // N

0.00673795
```

3.4 B Parameter-Mix Distributions

When the distribution of a random variable X depends upon a parameter θ, the (unknown) true value of θ is usually assumed fixed in the population. In some instances, however, an argument can be made for relaxing parameter fixity, which yields our second type of mixing distribution: parameter-mix distributions.

Two key components are required to form a parameter-mix distribution, namely the conditional distribution of the random variable given the parameter, and the distribution of the parameter itself. Let $f(x \mid \Theta = \theta)$ denote the density of $X \mid (\Theta = \theta)$, and let $g(\theta)$ denote the density of Θ. With this notation, the so-called '$g(\theta)$ parameter-mix of $f(x \mid \Theta = \theta)$' is written as

$$f(x \mid \Theta = \theta) \underset{\Theta}{\bigwedge} g(\theta) \tag{3.15}$$

and is equal to

$$E_\Theta\big[f(x \mid \Theta = \theta)\big] \tag{3.16}$$

where $E_\Theta[\,]$ is the usual expectation operator, with its subscript indicating that the expectation is taken with respect to the distribution of Θ. The solution to (3.16) is the unconditional distribution of X, which is the statistical model of interest. For instance,

$$\text{Binomial}(N,\, p) \underset{N}{\bigwedge} \text{Poisson}(\lambda)$$

denotes a Binomial(N, p) distribution in which parameter N (instead of being fixed) has a Poisson(λ) distribution. In this fashion, many distributions can be created using a parameter-mix approach; indeed the parameter-mix approach is often used as a device in its own right for developing new distributions. Table 5 lists five parameter-mix distributions (only the first three are discrete distributions).

Negative Binomial $(r,\, p)$	=	$\text{Poisson}(L) \underset{L}{\bigwedge} \text{Gamma}\big(r,\, \frac{1-p}{p}\big)$
Holla $(\mu,\, \lambda)$	=	$\text{Poisson}(L) \underset{L}{\bigwedge} \text{InverseGaussian}(\mu,\, \lambda)$
Pólya–Aeppli $(b,\, \lambda)$	=	$\text{Poisson}(\Theta) \underset{\Theta}{\bigwedge} \text{Gamma}(A,\, b) \underset{A}{\bigwedge} \text{Poisson}(\lambda)$
Student's $t(n)$	=	$\text{Normal}(0,\, S^2) \underset{S^2}{\bigwedge} \text{InverseGamma}(\frac{n}{2},\, \frac{2}{n})$
Noncentral Chi-squared $(n,\, \lambda)$	=	$\text{Chi-squared}(n + 2K) \underset{K}{\bigwedge} \text{Poisson}(\frac{\lambda}{2})$

Table 5: Parameter-mix distributions

For extensive details on parameter-mixing, see Johnson *et al.* (1993, Chapter 8). The following examples show how to construct parameter-mix distributions with **mathStatica**.

⊕ ***Example 12:*** A Binomial–Poisson Mixture

Find the distribution of X, when it is formed as Binomial(N, p) $\underset{N}{\bigwedge}$ Poisson(λ).

Solution: Of the two Binomial parameters, index N is permitted to vary according to a Poisson(λ) distribution, while the success probability p remains fixed. Begin by entering the key components. The first distribution, say $f(x)$, is the conditional distribution $X \mid (N = n) \sim$ Binomial(n, p):

```
        f  = Binomial[n, x] p^x (1 - p)^(n-x);
domain[f] = {x, 0, n} &&
            {0 < p < 1, n > 0, n ∈ Integers} && {Discrete};
```

The second is the parameter distribution $N \sim$ Poisson(λ):

```
        g  = ⅇ^-λ λ^n / n!;
domain[g] = {n, 0, ∞} && {λ > 0} && {Discrete};
```

From (3.16), we require the expectation E_N[Binomial(N, p)]. The pmf of the parameter-mix distribution is then found by entering:

```
Expect[f, g]
```

$$\frac{e^{-p\lambda}\,(p\,\lambda)^x}{x!}$$

The mixing distribution is discrete and has a Poisson form: $X \sim$ Poisson($p\,\lambda$). ■

⊕ ***Example 13:*** A Binomial–Beta Mixture: The Beta–Binomial Distribution

Consider a Beta parameter-mix of the success probability of a Binomial distribution:

$$\text{Binomial}(n, P) \underset{P}{\bigwedge} \text{Beta}(a, b).$$

The conditional distribution $X \mid (P = p) \sim$ Binomial(n, p) is:

```
        f  = Binomial[n, x] p^x (1 - p)^(n-x);
domain[f] = {x, 0, n} &&
            {0 < p < 1, n > 0, n ∈ Integers} && {Discrete};
```

The distribution of the parameter $P \sim$ Beta(a, b) is:

```
g = p^(a-1) (1 - p)^(b-1) / Beta[a, b];    domain[g] = {p, 0, 1} && {a > 0, b > 0};
```

We obtain the parameter-mix distribution by evaluating $E_P[f(x \mid P = p)]$ as per (3.16):

```
Expect[f, g]
```

$$\frac{\text{Binomial}[n, x]\, \Gamma[b+n-x]\, \Gamma[a+x]}{\text{Beta}[a, b]\, \Gamma[a+b+n]}$$

This is a Beta–Binomial distribution, with domain of support on the set of integers $\{0, 1, 2, \ldots, n\}$. The distribution is listed in the *Discrete* palette. ■

⊕ ***Example 14:*** A Geometric–Exponential Mixture: The Yule Distribution

Consider an Exponential parameter-mix of a Geometric distribution:

$$\text{Geometric}(e^{-W}) \underset{W}{\bigwedge} \text{Exponential}\left(\tfrac{1}{\lambda}\right).$$

For a fixed value w of W, the conditional distribution is Geometric with the set of positive integers as the domain of support. The Geometric's success probability parameter p is coded as $p = e^{-w}$, which will lie between 0 and 1 provided $w > 0$. Here is the conditional distribution $X \mid (W = w) \sim \text{Geometric}(e^{-w})$:

```
         f  = p (1 - p)^(x-1) /. p → e^-w;
domain[f] = {x, 1, ∞} && {w > 0} && {Discrete};
```

Parameter W is such that $W \sim \text{Exponential}\left(\tfrac{1}{\lambda}\right)$:

```
         g  = λ e^(-λ w) ;
domain[g] = {w, 0, ∞} && {λ > 0};
```

The parameter-mix distribution is found by evaluating:

```
Expect[f, g]
```

$$\frac{\lambda\, \Gamma[x]\, \Gamma[1+\lambda]}{\Gamma[1+x+\lambda]}$$

This is a Yule distribution, with domain of support on the set of integers $\{1, 2, 3, \ldots\}$, with parameter $\lambda > 0$. The Yule distribution is also given in **mathStatica**'s *Discrete* palette. The Yule distribution has been applied to problems in linguistics. Another distribution with similar areas of application is the Riemann Zeta distribution. It too may be entered from **mathStatica**'s *Discrete* palette. The Riemann Zeta distribution has also been termed the Zipf distribution, and it may be viewed as the discrete analogue of the continuous Pareto(a, 1) distribution; see Johnson *et al.* (1993, Chapter 11) for further details. ■

⊕ ***Example 15:*** Modelling the Change in the Price of a Security (Stocks, Options, etc.)

Let the continuous random variable Y denote the change in the price of a security (measured in natural logarithms) using daily data. In economics and finance, it is common practice to assume that $Y \sim N(0, \sigma^2)$. Alas, empirically, the Normal model tends to under-predict both large and small price changes. That is, many empirical densities of price changes appear to be both more peaked and have fatter tails than a Normal pdf with the same variance; see, for instance, Merton (1990, p. 59). In light of this, we need to replace the Normal model with another model that exhibits the desired behaviour. Let there be t transactions in any given day, and let $Y_i \sim N(0, \omega^2)$, $i \in \{1, \dots, t\}$, represent the change in price on the i^{th} transaction.[6] Thus, the daily change in price is obtained as $Y = Y_1 + Y_2 + \cdots + Y_t$, a sum of t random variables. For Y_i independent of Y_j $(i \neq j)$, we now have $Y \sim N(0, t\,\omega^2)$, with pdf $f(y)$:

$$\mathbf{f} = \frac{1}{\sqrt{t}\,\omega\,\sqrt{2\pi}}\,\mathrm{Exp}\left[-\frac{y^2}{2\,t\,\omega^2}\right];$$

```
domain[f] = {y, -∞, ∞} && {ω > 0, t > 0, t ∈ Integers};
```

Parameter-mixing provides a resolution to the deficiency of the Normal model. Instead of treating t as a fixed parameter, we are now going to treat it as a discrete random variable $T = t$. Then, Y is a random-length sum of T random variables, and is in fact a member of the Stopped-Sum class of distributions; see Johnson *et al.* (1993, Chapter 9). In parameter-mix terms, f is the conditional model $Y \mid (T = t)$. For the purposes of this example, let the parameter distribution $T \sim$ Geometric(p), with density $g(t)$:

$$\mathbf{g} = p\,(1-p)^{t-1};$$

```
domain[g] = {t, 1, ∞} && {0 < p < 1} && {Discrete};
```

The desired mixture is

$$N(0, T\,\omega^2) \underset{T}{\bigwedge} \text{Geometric}(p) \;=\; E_T\left[f(y \mid T = t)\right]$$

which we can attempt to solve as:

```
Expect[f, g]
```

$$\sum_{t=1}^{\infty} \frac{e^{-\frac{y^2}{2\,t\,\omega^2}}\,(1-p)^{-1+t}\,p}{\sqrt{2\pi}\,\sqrt{t}\,\omega}$$

This does not evaluate further, in this case, as there is no closed form solution to the sum. However, we can proceed by using numerical methods.[7] Figure 11 illustrates. In the left panel, we see that the parameter-mix pdf (the solid line) is more peaked in the neighbourhood of the origin than a Normal pdf (the dashed line). In the right panel (which zooms-in on the distribution's right tail), it is apparent that the tails of the pdf are fatter than a Normal pdf. The parameter-mix distribution exhibits the attributes observed in empirical practice.

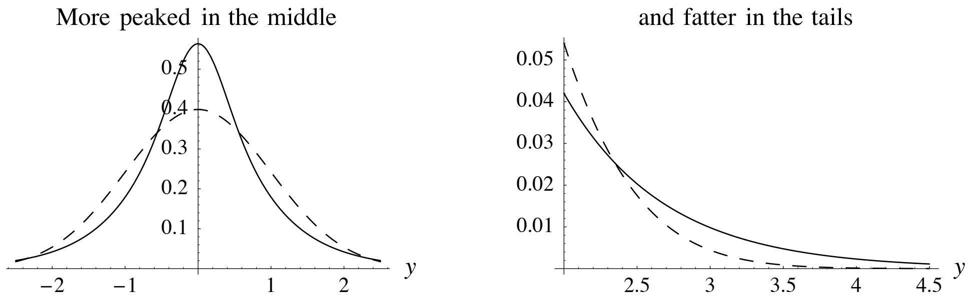


Fig. 11: Parameter-mix pdf (——) versus Normal pdf (– – –)

If a closed form solution is desired, we could simply select a different model for $g(t)$; for example, a Gamma or a Rayleigh distribution works nicely. While the latter are continuous densities, they yield the same qualitative results. For other models of changes in security prices see, for example, Fama (1965) and Clark (1973). ■

3.5 Pseudo-Random Number Generation

3.5 A Introducing `DiscreteRNG`

Let a discrete random variable X have domain of support $\Omega = \{x : x_0, x_1, \ldots\}$, with cdf $F(x) = P(X \leq x)$ and pmf $f(x) = P(X = x)$ such that $\sum_{x \in \Omega} f(x) = 1$. This section tackles the problem of generating pseudo-random copies of X. One well-known approach is the inverse method: if u is a pseudo-random drawing from the Uniform(0, 1), the (continuous) uniform distribution defined on the unit interval, then $x = F^{-1}(u)$ is a pseudo-random copy of X. Of course, this method is only desirable if the inverse function of F is computationally tractable, and this, unfortunately, rarely occurs. In this section, we present a discrete pseudo-random number generator entitled `DiscreteRNG` that is virtuous in two respects. First, it is universal — it applies in principle to any discrete univariate distribution without alteration. This is achieved by constructing $F^{-1}(u)$ as a lookup table, instead of trying to do so symbolically. Second, this approach is surprisingly efficient. Given that pluralism and efficiency are usually mutually incompatible, the attainment of both goals is particularly pleasing. Detailed discussion of the function appears in Rose and Smith (1997).

The **mathStatica** function `DiscreteRNG[`n, f`]` generates n pseudo-random copies of a discrete random variable X, with pmf f. It allows f to take either Function Form or List Form. We illustrate its use with both input types by example.

⊕ ***Example 16:*** The Poisson Distribution

Suppose that $X \sim$ Poisson(6). Then, in Function Form, its pmf $f(x)$ is:

$$f = \frac{e^{-\lambda} \lambda^x}{x!} \;/.\; \lambda \to 6;$$

```
domain[f] = {x, 0, ∞} && {Discrete};
```

Here, then, are 10 pseudo-random copies of X:

```
DiscreteRNG[10, f]
```

```
{5, 6, 9, 3, 5, 9, 2, 8, 5, 7}
```

and here are a few more:

```
data = DiscreteRNG[50000, f]; // Timing
```

```
{0.38 Second, Null}
```

Notice that it took `DiscreteRNG` a fraction of a second to produce 50000 Poisson(6) pseudo-random numbers!

In order to check how effective `DiscreteRNG` is in replicating the true distribution, we contrast the relative empirical distribution of the generated data with the true distribution of X using the **mathStatica** function `FrequencyPlotDiscrete`. The two distributions are overlaid as follows:

```
FrequencyPlotDiscrete[data, f];
```

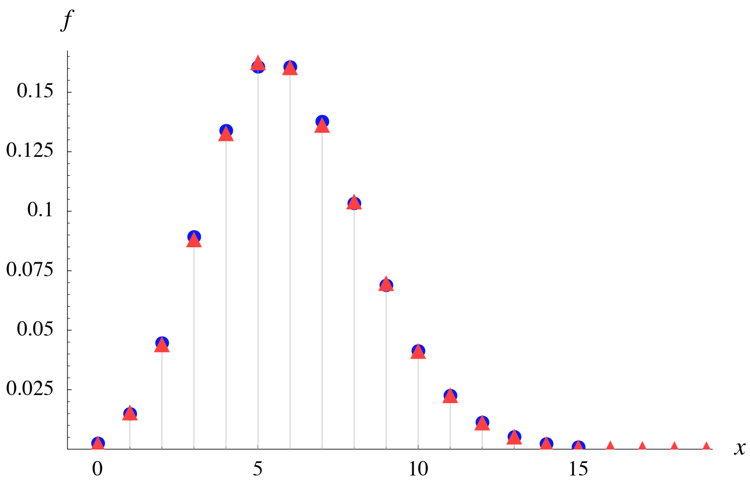


Fig. 12: Comparison of the empirical pmf (▲) to the Poisson(6) pmf (●)

The triangles give the generated empirical pmf, while the circles represent the true Poisson(6) pmf. The fit is superb. ■

⊕ ***Example 17:*** A Discrete Distribution in List Form

The previous example dealt with Function Form input. `DiscreteRNG` can also be used for List Form input. Suppose that random variable X is distributed as follows:

$P(X = x)$:	0.1	0.4	0.3	0.2
x :	-1	$3/2$	π	4.4

Table 6: The pmf of X

X's details in List Form are:

```
        f  = {0.1, 0.4, 0.3, 0.2};
domain[f] = {x, {-1, 3/2, π, 4.4}} && {Discrete};
```

Here are eight pseudo-random numbers from the distribution:

```
DiscreteRNG[8, f]
```

```
{1.5, 3.14159, 1.5, 4.4, 3.14159, 4.4, 4.4, 3.14159}
```

And here are a few more:

```
data = DiscreteRNG[50000, f]; // Timing
```

```
{0.39 Second, Null}
```

The empirical pmf overlaid with the true pmf is given by:

```
FrequencyPlotDiscrete[data, f];
```

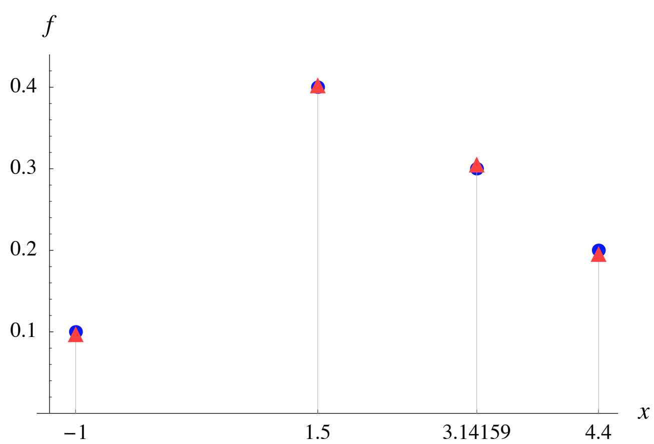


Fig. 13: Comparison of empirical pmf (▲) to true pmf (●)

Once again, the fit is superb. ■

⊕ ***Example 18:*** Holla's Distribution

Because `DiscreteRNG` is a general solution, it can generate random numbers, in principle, from any discrete distribution, not just the limited number of distributions that have been pre-programmed into *Mathematica*'s Statistics package. Consider, for example, Holla's distribution (see Table 5 for its parameter-mix derivation):

$$\mathbf{f} = \frac{1}{x!}\left(e^{\lambda/\mu}\sqrt{\frac{2}{\pi}}\sqrt{\lambda}\left(\frac{2}{\lambda}+\frac{1}{\mu^2}\right)^{\frac{1}{4}(1-2x)}\text{BesselK}\left[\frac{1}{2}-x,\ \sqrt{\lambda\left(2+\frac{\lambda}{\mu^2}\right)}\right]\right);$$

```
domain[f] = {x, 0, ∞} && {μ > 0, λ > 0} && {Discrete};
```

It would be a substantial undertaking to attempt to generate pseudo-random numbers from Holla's distribution using the inverse method. However, for given values of μ and λ, `DiscreteRNG` has no trouble in performing the task. Here is the code to produce 50000 pseudo-random copies:

```
data = DiscreteRNG[50000, f /. {μ → 1, λ → 4}]; // Timing
```

```
{0.39 Second, Null}
```

We again compare the empirical distribution to the true distribution:

```
FrequencyPlotDiscrete[data, f /. {μ → 1, λ → 4}];
```

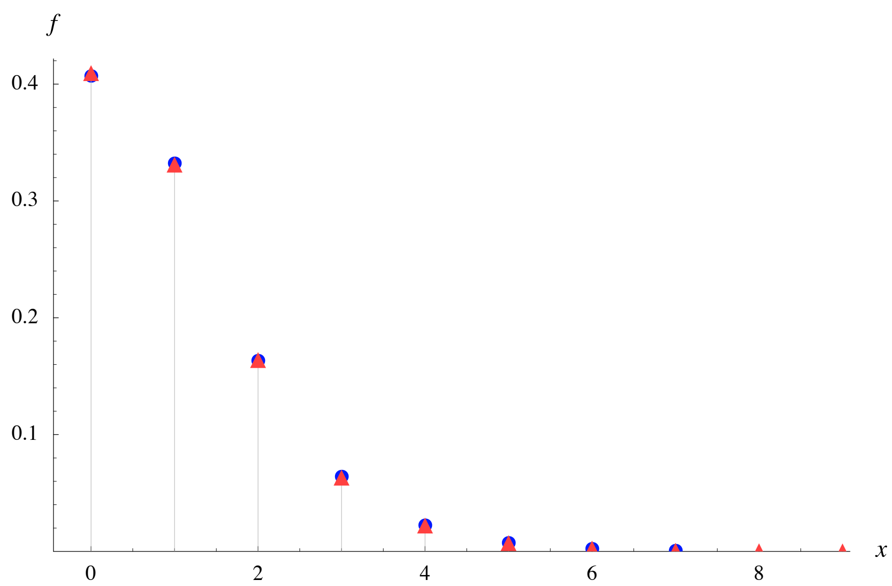


Fig. 14: Comparison of empirical pmf (▲) to Holla's distribution (●)

○ ***Computationl Efficiency***

Mathematica's `Statistics`DiscreteDistributions`` package includes the Poisson distribution used in *Example 16*, so we can compare the computational efficiency of *Mathematica*'s generator to **mathStatica**'s generator. After loading the Statistics add-on:

```
<<Statistics`
```

generate 50000 copies from Poisson(6) using the customised routine contained in this package:

```
dist = PoissonDistribution[6];
RandomArray[dist, {50000}]; // Timing
```

```
{77.67 Second, Null}
```

By contrast, **mathStatica** takes just 0.38 seconds (see *Example 16*) to generate the same number of copies. Thus, for this example, `DiscreteRNG` is around 200 times faster than *Mathematica*'s Statistics package, even though `DiscreteRNG` is a general solution that has not been specially optimised for the Poisson. In further comparative experiments against the small range of discrete distributions included in the *Mathematica* Statistics package, Rose and Smith (1997) report complete efficiency dominance for `DiscreteRNG`.

3.5 B Implementation Notes

`DiscreteRNG` works by internally constructing a numerical lookup table of the specified discrete random variable's cdf.[8] When generating many pseudo-random numbers from a particular discrete distribution, it therefore makes sense to ask for all the desired pseudo-random numbers in one go, rather than repeatedly constructing the lookup table. The contrast in performance is demonstrated by the following timings for a Riemann Zeta distribution:

$$f = \frac{x^{-(\rho+1)}}{\text{Zeta}[1+\rho]} \;/.\; \rho \to 3;$$

```
domain[f] = {x, 1, ∞} && {Discrete};
```

The first input below calls `DiscreteRNG` 1000 times, whereas the second generates all 1000 in just one call and is clearly far more efficient:

```
Table[DiscreteRNG[1, f], {1000}]; // Timing
```

```
{12.31 Second, Null}
```

```
DiscreteRNG[1000, f]; // Timing
```

```
{0. Second, Null}
```

A numerical lookup table is naturally limited to a finite number of elements. Thus, if the discrete random variable has an infinite number of points of support (as in the Riemann Zeta case), then the tail(s) of the distribution must be censored. The default in `DiscreteRNG` is to automatically censor the left and right tails of the distribution in such a way that less than $\varepsilon = 10^{-6}$ of the density mass is disturbed at either tail. The related **mathStatica** function `RNGBounds` identifies the censoring points and calculates the probability mass that is theoretically affected by censoring. For example, for $X \sim$ Riemann Zeta(3):

```
RNGBounds[f]
```

```
— The density was not censored below.
— Censored above at x = 68 . This can affect
   9.58084×10^-7 of the density mass.
```

The printed output tells us that when `DiscreteRNG` is used at its default settings, it can generate copies of X from the set $\Omega^* = \{1, \ldots, 68\}$. By censoring at 68, outcomes $\Omega_* = \{69, 70, 71, \ldots\}$ are reported as 68. Thus, the censored mass is not lost; it is merely shifted to the censoring point. The density mass shifted in this way corresponds to $P(X \in \Omega_*)$ which is equal to:

```
1 - Prob[68, f] // N
```

```
9.58084 × 10^-7
```

as reported by `RNGBounds` above.

If censoring at $x = 68$ is not desirable, tighter (or weaker) tolerance levels can be set. `RNGBounds[`f, $\underline{\varepsilon}$, $\overline{\varepsilon}$`]` can be used to inspect the effect of arbitrary tolerance settings, while `DiscreteRNG[`n, f, $\underline{\varepsilon}$, $\overline{\varepsilon}$`]` imposes those settings on the generator; $\underline{\varepsilon}$ is the tolerance setting for the left tail, and $\overline{\varepsilon}$ is the setting for the right tail. For example:

```
RNGBounds[f, 10^-8, 10^-8]
```

```
— The density was not censored below.
— Censored above at x = 313 . This can affect
   9.99556×10^-9 of the density mass.
```

Thus, `DiscreteRNG[`n, f, 10^{-8}, 10^{-8}`]` will generate n copies of $X \sim$ Riemann Zeta(3), with outcomes restricted to the integers in $\Omega^* = \{1, 2, \ldots, 313\}$; censoring occurs on the right at 313 which results in just under 10^{-8} of the density mass being shifted to that point. The reason the censoring point has 'blown out' to 313 is because the Riemann Zeta distribution is long-tailed.

`DiscreteRNG` and `RNGBounds` are defined for tolerance settings $\varepsilon \geq 10^{-15}$. Setting ε outside this interval is not meaningful and may cause problems. (It is also assumed that $\varepsilon < 0.25$, although this constraint should never be binding.) In List Form examples, the distribution is never censored, so `RNGBounds` does not apply and, by design, will not evaluate. For Function Form examples, we recommend that whenever `DiscreteRNG` is applied, the printed output from `RNGBounds` should also be inspected.

Finally, by constructing a lookup table, `DiscreteRNG` trades off a small fixed cost in return for a lower marginal cost. This trade-off will be particularly beneficial if a large number of pseudo-random numbers are required. If only a few are needed, it may not be worthwhile. The fixed cost is itself proportional to the size of the lookup table. For instance, a Discrete Uniform such as $f = 10^{-6}$ defined on $\Omega = \{1, \ldots, 10^6\}$ will require a huge lookup table. Here, a technique such as `Random[Integer,{1,10^6}]` is clearly more appropriate.

3.6 Exercises

1. Let random variable X take values 1, 2, 3, 4, 5, with probability $\frac{1}{55}$, $\frac{4}{55}$, $\frac{9}{55}$, $\frac{16}{55}$, $\frac{25}{55}$, respectively.

(i) Enter the pmf of X in List Form, plot the pmf, and then evaluate $E[X]$.

(ii) Enter the pmf of X in Function Form, and evaluate $E[X]$.

(iii) Repeat (i) and (ii) when X takes values 1, 3, 5, with probability $\frac{1}{35}$, $\frac{9}{35}$, $\frac{25}{35}$, respectively.

2. Enter the Binomial(n, p) pmf from **mathStatica**'s *Discrete* palette. Express the pmf in List Form when $n = 10$ and $p = 0.4$.

3. Derive the mean, variance, cdf, mgf and pgf for the following distributions whose pmf may be entered from **mathStatica**'s *Discrete* palette: (i) Geometric, (ii) Hypergeometric, (iii) Logarithmic, and (iv) Yule.

4. Using the shaved 1-face dice described in *Example 3*, plot the probability of winning Craps against δ.

5. A gambler aims to increase his initial capital of \$5 to \$10 by playing Craps, betting \$1 per game. Using simulation, estimate the probability that the gambler can, before ruin (*i.e.* his balance is depleted to \$0), achieve his goal.

6. In a large population of n individuals, each person must submit a blood sample for test. Let p denote the probability that an individual returns a positive test, and assume that p is small. The test designer suggests pooling samples of blood from m individuals, testing the pooled sample with a single test. If a negative test is returned, then this one test indicates that all m individuals are negative. However, if a positive test is returned, then the test is carried out on each individual in the pool. For this sampling design, determine μ (the expected number of tests), and the optimal value of m when $p = 0.01$. Assume all individuals in the population are mutually independent, and that p is the same across all individuals.

7. What are the chances of throwing: (i) at least 1 six from a throw of a box containing 6 dice, (ii) at least 2 sixes from another box containing 12 dice, and (iii) 3 or more sixes from a third box filled with 18 dice?

8. An urn contains 20 balls, 4 of which are coloured red. A sample of 5 balls is drawn one-by-one from the urn. What is the probability that one of the balls drawn is red:

(i) if each ball that is drawn is returned to the urn?

(ii) if each ball that is drawn is set aside?

9. Experience indicates that a firm will, on average, fire 3 workers per year. Assuming that the number of employees fired per year is Poisson distributed, what is the probability that in the coming year the firm will: (i) not fire any workers, and (ii) fire at least 4 workers?

10. Let a random variable $X \sim \text{Poisson}(\lambda)$. Determine the smallest value of λ such that $P(X \leq 1) \leq 0.05$.

11. Determine the pmf of the following parameter-mixes, and plot it at the indicated values of the parameters:

(i) Binomial(N, p) $\underset{N}{\bigwedge}$ Binomial(m, q). Plot for $p = \frac{3}{4}$, $q = \frac{1}{2}$, $m = 10$.

(ii) Negative Binomial(R, p) $\underset{R}{\bigwedge}$ Geometric(q). Plot for $p = \frac{1}{4}$, $q = \frac{2}{3}$.

(iii) Poisson(Θ) $\underset{\Theta}{\bigwedge}$ Lindley(δ). Plot for $\delta = 1$.

12. (i) Use `DiscreteRNG` to generate 20000 pseudo-random drawings from the Geometric(0.1) distribution. Then use `FrequencyPlotDiscrete` to plot the empirical distribution, with the true distribution superimposed on top.

(ii) Repeat (i), this time using *Mathematica*'s Statistics package pseudo-random number generator:

```
RandomArray[GeometricDistribution[0.1], 20000] + 1
```

(the "+1" is required because the Geometric distribution hardwired in the Statistics package includes 0 in its domain of support).

(iii) Report on any discrepancies you observe between the empirical and true distributions.

13. (i) Generate 20000 pseudo-random numbers from a Zero-Inflated Poisson distribution (parameters ω and λ; see *Example 11*) when $\omega = 0.6$ and $\lambda = 4$. Compare the empirical distribution to the theoretical distribution.

(ii) Generate 20000 pseudo-random numbers from a Poisson two-component-mix distribution (parameters ω, λ and θ; see *Example 10*) when $\omega = 0.6$, $\lambda = 9$ and $\theta = 3$. Compare the empirical distribution to the theoretical distribution.