Discrete probability distributions I

Deependra Dhakal

https://rookie.rbind.io

Academic year 2019-2020

Probability distributions

Sample Spaces and Events[1]

An **experiment** is any action or process whose outcome is subject to uncertainty. The **sample space** of an experiment, denoted $\mathcal{S}$, is the set of all possible outcomes of that experiment. An **event** is any subset of $\mathcal{S}$.

---

[1]https://jasonbintz.rbind.io/courses/math333/notes/probability-distributions/

Axioms of Probability

Let $\mathcal{S}$ be a sample space, E an event and $\{E_1, E_2, ...\}$ a countable collection of pairwise disjoint events. Then,

1. $P(E) \geq 0$
2. $P(\mathcal{S}) = 1$
3. $P\left(\bigcup\limits_{i=1}^{\infty} E_i\right) = \sum\limits_{i=1}^{\infty} P(E_i)$

Random Variable

Let $\mathcal{S}$ be a sample space and let $X : \mathcal{S} \to \mathbb{R}$. Then X is a **random variable**.

Problem 1

For a coin toss we have $\mathcal{S} = \{T, H\}$. Let $X : \mathcal{S} \to \mathbb{R}$ with $X(T) = 0$ and $X(H) = 1$.

A random variable is a **discrete random variable** if its set of values is countable. A random variable is **continuous** if its set of values is an interval. (That's not quite accurate, but it's good enough.)

Probability Distributions

When probabilities are assigned to the outcomes in $\mathcal{S}$, these in turn determine probabilities associated with the values of any random variable X defined on $\mathcal{S}$. The **probability distribution of** X describes how the total probability of 1 is distributed among the values of X.

Probability Mass Function (pmf)

Let $X : \mathcal{S} \to \mathbb{R}$ be a discrete random variable. The **probability mass function of X** (pmf) is the function $p : \mathbb{R} \to [0, 1]$ such that for all $x \in \mathbb{R}$, $p(x) = P(X = x)$.

Problem 2

The pmf for the random variable in the previous example where the probability of heads is θ is given by

$$p(x) = \begin{cases} 1 - \theta & \text{if } x = 0 \\ \theta & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

Probability mass functions can be visualized with a line graph. For $\theta = 0.2, 0.5,$ and 0.8:
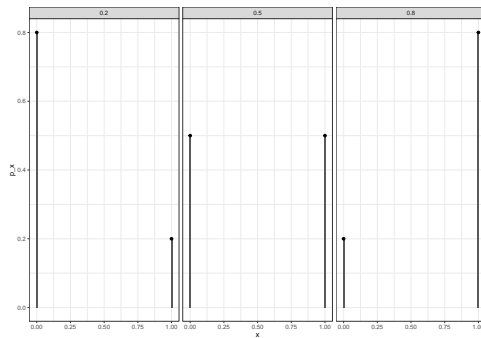
It's worth noting a couple of properties of a pmf:

1. $p(x) \geq 0, x \in \mathbb{R}$
2. $p(x) = 0, x \notin X(\mathcal{S})$
3. $\sum_{x \in X(\mathcal{S})} p(x) = 1$

Problem 3

Consider the experiment of tossing a coin twice and counting the number of heads. Give the sample space, define a sensible random variable and draw a line graph for the pmf.

Solution: $\mathcal{S} = \{TT, TH, HT, HH\}$. Define $X : \mathcal{S} \to \mathbb{R}$ by $X(TT) = 0$, $X(TH) = X(HT) = 1$ and $X(HH) = 2$. Then the pmf is given by

$$p(x) = \begin{cases} 0.25 & \text{if } x = 0 \\ .5 & \text{if } x = 1 \\ 0.25 & \text{if } x = 2 \\ 0 & \text{otherwise} \end{cases}$$
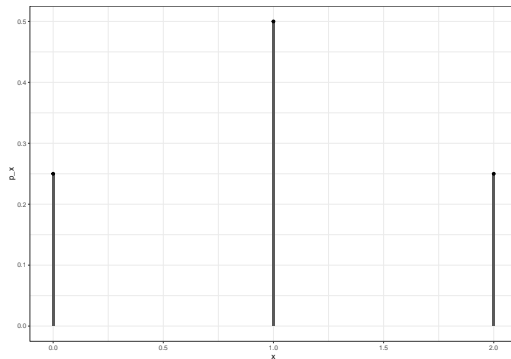
Problem 4

Consider the experiment of tossing a biased coin (where the probability of heads is θ) until it lands on heads. Give the sample space, define a sensible random variable and draw a line graph for the pmf.

Solution: $\mathcal{S} = \{H, TH, TTH, TTTH, ...\}$. Define $X : \mathcal{S} \to \mathbb{R}$ by the number of flips required. Then the pmf is given by

$$p(x) = \begin{cases} (1-\theta)^{x-1}\theta, & x = 1, 2, 3, ... \\ 0 & \text{otherwise} \end{cases}$$

Probability Density Function (pdf)

Let X be a continuous random variable. The **probability density function of** X (pdf) is a function $f(x)$ such that for $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx.$$

A couple of properties:

1. $f(x) \geq 0, x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$
3. $P(X = c) = \int_c^c f(x)\,dx = 0$

Example: A **uniform random variable** on $[a, b]$ is a model for "pick a random number between a and b." The pdf is given by

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}.$$

Example: Let X be **normal random variable**. Then X has the pdf given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where μ and σ^2 are the mean and variance of X. These are properties of X that we now discuss.

Mean and Variance

Let X be a random variable. The **mean** or **expected value** of X is defined by

$$\mathbb{E}(X) = \begin{cases} \displaystyle\sum_{x \in X(\mathcal{S})} x \cdot p(x), & X \text{ discrete} \\ \displaystyle\int_{-\infty}^{\infty} x \cdot f(x)\, dx, & X \text{ continuous} \end{cases}.$$

The **variance** of X is given by

$$\mathbb{V}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2 \right).$$

The expected value is often denoted μ and the variance, σ^2. The square root of the variance, σ, is called the **standard deviation.**

Highest density interval

The mean gives us one way to quantify the central tendency of a distribution. Variance is one way to quantify the spread of a distribution. Another way to quantify the spread of a distribution is to specify the highest density interval (HDI). The 95% HDI for a distribution, for example, is an interval I such that for $x \in I$, $f(x) > W$ for some fixed W and

$$\int_I f(x)\, dx = 0.95.$$

Using R to plot a normal random variable.

A continuous probability distribution function f, as a function of random variable x (Hence, $f(x)$), is created. The graph shows the pdf of a normal rv with mean 0 and standard deviation 0.2.
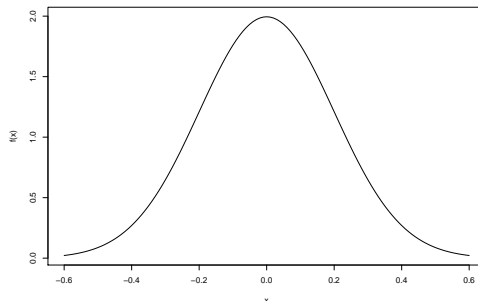
We can estimate the area under this curve. What should it be close to? Why?

[1] 0.997294665511095

How much of the density is between $x = -0.2$ and 0.2?

[1] 0.682588662598608

Distribution versus density function[2]

In general, a **distribution** function gives cumulative probability, $P(X \leq x)$, hence it is sometimes also referred to as cumulative probability function. A **density** function, however gives the exact probability of occurance of an event over given number of trials, where each trial has a certain probability of occurance associated with it.

In R, variants of density functions are prefixed with "d", such as **d**binom, **d**geom, **d**pois and those of distribution functions are prefixed with "p", such as **p**binom, **p**geom, **p**pois.

---

[2]Teetor (2011)

Survival function

The complement of the cumulative probability is called the **survival function**, $P(X > x)$. All of the distribution functions defined in R allow finding this right-tail probability simply by specifying `lower.tail = FALSE`.

For example, in

```
## [1] 0.0546875
```

We see that the probability of observing $X > 7$ is about 0.055.

Interval probability

The interval probability, $P(x_1 < X \leq x_2)$, is the probability of observing X between the limits x_1 and x_2. It is simply calculated as the difference between two cumulative probabilities: $P(X \leq x_2) - P(X \leq x_1)$. Here is $P(3 < X \leq 7)$ for our binomial variable:

```
## [1] 0.7734375
```

R lets you specify multiple values of x for these functions and will return a vector of the corresponding probabilities. Here we calculate two cumulative probabilities, $P(X \leq 3)$ and $P(X \leq 7)$, in one call to pbinom:

```
## [1] 0.1718750 0.9453125
```

This leads to a one-liner for calculating interval probabilities. The `diff` function calculates the difference between successive elements of a vector. We apply it to the output of `pbinom` to obtain the difference in cumulative probabilities – in other words, the interval probability:

```
## [1] 0.7734375
```

Discrete uniform distribution

Definition

A r.v. X is said to have a discrete uniform distribution over the range $[1, n]$ if its p.m.f is expressed as follows:

$$P(X = x) = \begin{cases} \frac{1}{n} & \text{for } x = 1, 2, .., n \\ 0, & \text{otherwise} \end{cases}$$

Bernoulli distribution

Definition

A r.v. X is said to have a Bernoulli distribution with parameter p if its p.m.f is given by:

$$P(X = x) = \begin{cases} p^x(1 - p)^{1-x} & \text{for } x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

Definition simplified[3]

A Bernouli Disribution is the probability distribution of a random variable which takes the value 1 with probability p and value 0 with probability $1 - p$, i.e.

$$\begin{cases} 1 - p & \text{for } k = 0 \\ p & \text{for } k = 1 \end{cases}$$

---

[3]http://benalexkeen.com/discrete-probability-distributions-bernoulli-binomial-poisson/

Problem 1

In a population, approximately 10% of the people are left-handed ($p = 0.1$). We want to know, out of a random sample of 10 people, what is the probability of 3 these 10 people being left handed ?

We assign a 1 to each person if they are left handed and 0 otherwise:

$$P(X = 1) = 0.1$$
$$P(X = 0) = 0.9$$

A Binomial distribution is derived from the Bernoulli distribution. Let's start with a simpler problem.

What is the probability of the first 3 people we pick being left-handed, followed by 7 people being right-handed ?

This is given by: $0.1^3 \times 0.9^7$. Which is:

```
## [1] 0.0005
```

What if we wanted the last 3 people to be left-handed ? This is given by: $0.9^7 \times 0.1^3$. This is same as previous.

The fact is it does not matter how we arrange the 3 people, we always have the same proability.

So we have to add up all the ways we can arrange the 3 people being picked.

There are 10! ways to arrange 10 people and there are 3! ways to arrange the 3 people that are picked and 7! ways to arrange the 7 that aren't picked.

Thus ways in which 3 people being picked are picked is given by: $\frac{10!}{3!7!}$, which is:

[1] 120

This can be generalized to say that, "10 choose 3". The "n choose k" notation is written as:

$$\binom{N}{k} = \frac{n!}{k!(n-1)!}$$

We can now calculate the probability that there are 3 left-handed people in a random selection of 10 people as:

$$P(X = 3) = \binom{10}{3}(0.1)^3(0.9)^7$$

This equals 0.057395628.

This expression of "n choose k" is implemented in r function `choose`.

```
## [1] 120
```

This can be further generalized as:

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

This probability can be calculated using r's native `dbinom` function. This is also known as binomial density.

```
## [1] 0.057395628
```

For obtaining different 0, 1, 2, 3 and 4 successful outcomes (left handed people) (assuming each have same frequency of occurance (10%) in the population), binomial density function approximation could be used as follows:

```
## # A tibble: 5 x 2
##   selection_of probability
##          <int>       <dbl>
## 1            0       0.349
## 2            1       0.387
## 3            2       0.194
## 4            3      0.0574
## 5            4      0.0112
```

We could plot our probabilities for each values upto all 10 people being left-handed:
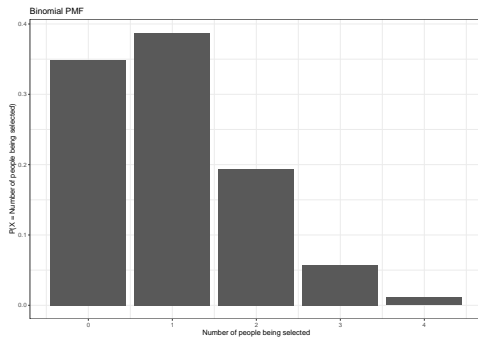
We can see there is almost negligible chance of getting more than 6 left-handed people in a random group of 10 people.

The probability of obtaining either 0, 1, 2, 3 **or** 4 successes (left handed people) (it can be restated as: the probability of having less than 4 successes) in a random selection of 10 people, can be obtained by summing over the binomial density vector.

[1] 0.9983650626

This is same as calculating cumulative probability density using pbinom function.

[1] 0.9983650626

The quantile is defined as the smallest value x such that $F(x) \geq p$, where F is the distribution function.

Problem 2

On an American roulette wheel there are 38 squares:

- 18 black
- 18 red
- 2 green

We bet on black 5 times in a row, what are the chances of winning more than half of these ?

The problem can be formulated as:

$$P(X > 5) = \sum_{i=6}^{10} \binom{10}{i} \left(\frac{18}{38}\right)^{i} \left(1 - \frac{18}{38}\right)^{n-i}$$

The probability value can be obtained by following r expression:

[1] 0.314125043967762

Simulating bernoulli trial

We can generate a sequence of 20 Bernoulli trials – random successes or failures. We use TRUE to signify a success and FALSE otherwise.

R could implement the simulation via `sample` function. By default, sample will choose equally among the set elements and so the probability of selecting either TRUE or FALSE is 0.5. With a Bernoulli trial, the probability p of success is not necessarily 0.5. You can bias the sample by using the prob argument of sample; this argument is a vector of probabilities, one for each set element. Suppose we want to generate 20 Bernoulli trials with a probability of success $p = 0.8$. We set the probability of FALSE to be 0.2 and the probability of TRUE to 0.8.

```
## [1]    TRUE   TRUE   TRUE   TRUE  FALSE   TRUE  FALSE   TRUE   TRUE   TRUE   TRUE
## [12]   TRUE  FALSE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE
```

The resulting sequence is clearly biased towards TRUE. This special case of a binary-valued sequence can be simulated using another built-in function rbinom, the random generator for binomial variates.

```
## [1]  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE
## [12]  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE
```

Exact binomial test

Note that exact probability of success in Bernoulli experiment is different from probability values for acceptance of hypothesis from exact test of a simple null. Like all other hypothesis testing, it requires sides of alternative be defined and (defaults to "two.sided") and confidence level be specified.

Under (the assumption of) simple Mendelian inheritance, a cross between plants of two particular genotypes produces progeny $\frac{1}{4}$ of which are "dwarf" and $\frac{3}{4}$ of which are "giant", respectively (Conover and Conover (1980); p. 97f.)

In an experiment to determine if this assumption is reasonable, a cross results in progeny having 243 dwarf and 682 giant plants.

If "giant" is taken as success, the null hypothesis is that $p = \frac{3}{4}$ and the alternative that $p! = \frac{3}{4}$.

```
## 
##  Exact binomial test
## 
## data:  c(682, 243)
## number of successes = 682, number of trials = 925, p-value =
## 0.382491559575
## alternative hypothesis: true probability of success is not equal to 0.75
## 95 percent confidence interval:
##  0.707668264079039 0.765406558241525
## sample estimates:
## probability of success
##      0.737297297297297
```

Assuming data of P(probability values) and V(observation values) for a discrete probability distribution. Calculation of mean and variances of discrete probability distribution can be done as follows:

```
## # A tibble: 1 x 3
##   probability_sum probability_variance probability_sd
##             <dbl>                <dbl>          <dbl>
## 1            3.44                 8.49           2.91
```

Let's make use of dataset about sales of car on a saturday by a car-salesman.

```
## [1] 0.89

## [1] 1.7779

## # A tibble: 6 x 2
##   numsold probability
##     <int>       <dbl>
## 1       0      0.0312
## 2       1      0.156
## 3       2      0.312
## 4       3      0.312
## 5       4      0.156
## 6       5      0.0312
```

Probability of exactly 8 successes out of 10 times

[1] 0.28849861690833

Probability of successes 6 times or less.

Question: why is the probability of success 6 times or less is smaller than individual probabilities? $\longrightarrow$ Because individual probability of success is high i.e., 0.76)

[1] 0.201248110790396

Generate 100 sample of successes in times out of 10 trials with given probability of success

Having confidence limits below, we can say with 95% confidence that out of 10 trials between "lower confidence limit" successes and "uper confidence limit" successes will occur.

```
## 2.5%
##    5

## 97.5%
##    10
```

Binomial distribution

Definition

A r.v. X is said to follow binomial distribution if it assumes only non-negative values and its p.m.f. is given by:

$$P(X = x) = p(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & x = 0, 1, 2..., n; q = 1 - p \\ 0, & \text{otherwise} \end{cases}$$

The assignment of probabilities in the definition of binomial distribution is permissible because,

$$\sum_{x=0}^{n} p(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (q + p)^n = 1$$

Let us suppose that n trials constitute an experiment. Then, if this experiment is repeated N times, the frequency function of the binomial distribution is given by:

$$f(x) = Np(x) = N \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x}; x = 0, 1, 2, ..., n$$

and the expected frequencies of 0, 1, 2,..., n successes are the successive terms of a binomial expansion, $N(q + p)^n$, $q + p = 1$.

Binomial theorem

We know the fact that,

$$k \binom{n}{k} = \frac{kn!}{k!(n-k)!} = \frac{n!}{(k-1)!(n-k)!} = \frac{n(n-1)!}{(k-1)!(n-k)!} = n \binom{n-1}{k-1} \; :$$

So,

$$\sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} n\binom{n-1}{k-1} p^k (1-p)^{n-k}$$

$$= n \sum_{k=0}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k}$$

$$= n \sum_{k=0}^{n-1} \binom{n-1}{k} p^{k+1} (1-p)^{n-k-1}$$

$$= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-k-1}$$

$$= np \Big(p + (1-p) \Big)^{n-1} \qquad \text{binomial theorem}$$

$$= np \ .$$

Physical/Experimental conditions for Binomial distribution

1. Each trial results in two exhaustive and mutually disjoint outcomes, termed as success and failure.
2. The number of trials 'n' is finite.
3. The trials are independent of each other.
4. The probability of success 'p' is constant in each trial.

Note: The trials satisfying the conditions 1, 3 and 4 are also called **Bernoulli trials**.

Problem 1

A coffee connoisseur claims that he can distinguish between a cup of instant coffee and a cup of percolator coffee 75% of the time. It is agreed that his claim will be accepted if he correctly identifies at least 5 of the 6 cups. Find his chances of having the claim (i) accepted, (ii) rejected, when he does have the ability he claims.

If p denotes the probability of a correct distinction between a cup of instant coffee and a cup of percolator coffee, then we are given:

$$p = \frac{75}{100} = \frac{3}{4} \implies q = 1 - p = \frac{1}{4}, \text{ and } n = 6$$

If the random variable X denotes the number of correct distinctions, then by the Binomial probability law, the probability of correct identification out of 6 cups is:

$$P(X = x) = p(x) = \binom{6}{x} \left(\frac{3}{4}\right)^x \left(\frac{1}{4}\right)^{6-x} ; x = 0, 1, 2, ..., 6$$

1. The probability of the claim being accepted is:

$$P(X \geq 5) = p(5) + p(6) = \binom{6}{5}\left(\frac{3}{4}\right)^5\left(\frac{1}{4}\right)^{6-5} + \binom{6}{6}\left(\frac{3}{4}\right)^6\left(\frac{1}{4}\right)^{6-6} = 0.534$$

2. The probability of claim being rejected is:

$$P(X \leq 4) = 1 - P(X \geq 5) = 1 - 0.534 = 0.466$$

Problem 2

In a binomial distribution consisting of 5 independent trials, probabilities of 1 and 2 successes are 0.4096 and 0.2048 respectively. Find the parameter 'p' of the distribution.

Let $X \sim B(n, p)$. In usual notations, we are given: $n = 5$, $p(1) = 0.4096$ and $p(2) = 0.2048$.

According to Binomial probability law:

$$P(X = x) = p(x) = \binom{5}{x} p^x (1 - p)^{5-x}; x = 1, 2, ..., 5$$

Now,

$$p(1) = \binom{5}{1} p(1-p)^4 \quad = 0.4096... \text{ (expression 1)}$$

$$p(2) = \binom{5}{2} p^2(1-p)^3 \quad = 0.2048... \text{(expression 2)}.$$

Dividing (expression 1) by (expression 2), we get:

$$\frac{\binom{5}{1} p(1-p)^4}{\binom{5}{2} p^2(1-p)^3} = \frac{0.4096}{0.2048} \implies \frac{5(1-p)}{10p} = 2 \implies p = \frac{1}{5} = 0.2$$

For a detailed discourse on:

1. Moments of Binomial distribution
2. Examples relating to moments of Binomial distribution
3. Recurrence relation for the moments of Binomial distribution (Renovsky formula)
4. Factorial momments of Binomial distribution
5. Mean deviation about mean of Binomial distribution
6. Mode of binomial distribution
7. Moment generating function of Binomial distribution

8 Additive property of Binomial distribution
9 Characteristic function of Binomial distribution
10 Cumulants of the Binomial distribution
11 Recurrance relation of cumulants of Binomial distribution
12 Probability generating function of Binomial distribution
13 Recurrence relation for the probabilities of Binomial distribution

Seven coins are tossed and number of heads noted. The experiment is repeated 128 times and the following distribution is obtained:

```
## # A tibble: 8 x 2
##    `Number of heads` Frequencies
##                <dbl>       <dbl>
## 1                  0           7
## 2                  1           6
## 3                  2          19
## 4                  3          35
## 5                  4          30
## 6                  5          23
## 7                  6           7
## 8                  7           1
```

Fit a binomial distribution distribution assuming, (i) The coin is unbiased (ii) The nature of the coin is not known. (iii) Probability of a head for four coins is 0.5 and for the remaining three coins is 0.45.

In fitting binomial distribution, first of all the mean and variance of the data are equated to np and npq respectively. Then the expected frequencies are calculated from these values of n and p. Here $n = 7$ and $N = 128$.

Case I. When the coin is unbiased:

$$p = q \hspace{4cm} = \frac{1}{2} \implies \frac{p}{q} = 1$$

$$p(0) = q^n = \left(\frac{1}{2}\right)^7 = 128 \text{ so that } f(0) = Nq^n = 128\left(\frac{1}{2}\right)^7 = 1$$

Using the recurrence formula:

$$p(x+1) = \left\{ \frac{n-x}{x+1} \cdot \frac{p}{q} \right\} p(x)$$

various probabilities, viz, $p(1), p(2), \dots$ can be easily calculated as shown below:

x	f	p	q	fx	$\frac{n-x}{x+1}$	$\frac{n-x}{x+1} \cdot \frac{p}{q}$	Expected frequency $f(x) = Np(x)$
0	7	0.5	0.5	0	7.00	7.00	1
1	6	0.5	0.5	6	3.00	3.00	7
2	19	0.5	0.5	38	1.67	1.67	21
3	35	0.5	0.5	105	1.00	1.00	35
4	30	0.5	0.5	120	0.60	0.60	35
5	23	0.5	0.5	115	0.33	0.33	21
6	7	0.5	0.5	42	0.14	0.14	7
7	1	0.5	0.5	7	0.00	0.00	1

Case II. When the nature of coin is not known, then

$$Mean = np = \bar{x} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i = \frac{433}{128} = 3.38; n = 7$$

$$\therefore p = \frac{3.3828}{7} = 0.4826 \text{ and } q = 1 - p = 0.51675, \implies \frac{p}{q} = 0.93521$$

$$f(0) = Nq^7 = 128(0.5167)^7 = 1.2593 \text{ (using logarithms)}$$

x	f	p	q	fx	$\frac{n-x}{x+1}$	$\frac{n-x}{x+1} \cdot \frac{p}{q}$	Expected frequency $f(x) = Np(x)$
0	7	0.48	0.52	0	7.00	6.55	1.26
1	6	0.48	0.52	6	3.00	2.81	8.24
2	19	0.48	0.52	38	1.67	1.56	23.12
3	35	0.48	0.52	105	1.00	0.94	36.10
4	30	0.48	0.52	120	0.60	0.56	33.76
5	23	0.48	0.52	115	0.33	0.31	18.95
6	7	0.48	0.52	42	0.14	0.13	5.85
7	1	0.48	0.52	7	0.00	0.00	0.77

The probability for generating function (p.g.f), say $P_X(s)$ for the 4 coins and $P_Y(s)$ for the remaining 3 coins are given by:

$$P_X(s) = (0.5 + 0.5s)^4$$
$$P_Y(s) = (0.55 + 0.45s)^3$$

Since all the throws are independent, the p.g.f $P_{X+Y}(s)$ for the whole experiment is given by:

$$
\begin{aligned}
P_{X+Y}(s) &= P_X(s)P_Y(s) \\
&= (0.50 + 0.50s)^4 \times (0.55 + 0.45)^3 \\
&= (0.0625 + 0.25s + 0.375s^2 + 0.25s^3 + 0.0625s^4) \times (0.166 + 0.408s + 0.334s^2 + (
\end{aligned}
$$

$$f(x) = N \times \text{Coefficient of } s^x \text{ in } P_{X+Y}(s)$$
$$\therefore f(0) = 128 \times 0.0625 \times 0.16637 = 1.1331$$
$$f(1) = 128(0.25 + 0.166 + 0.4084 \times 0.0625) = 8.5910$$
$$f(2) = 128(0.284) = 36.347$$
$$f(3) = 128(0.184) = 23.567$$
$$f(4) = 128(0.261) = 33.353$$
$$f(5) = 128(0.146) = 18.693$$
$$f(6) = 128(0.044) = 5.589$$
$$f(7) = 128(0.006) = 0.729$$

These frequencies, rounded to the nearest integer, keeping in minid that total frequency
is N $= 128$, are given below:

x	f
0	1
1	9
2	36
3	23
4	33
5	19
6	6
7	1

Bibliography

Further study

Also see: Gupta (2002)

References

Conover, William Jay, and William Jay Conover. 1980. "Practical Nonparametric Statistics." Wiley New York.

Gupta, VK, SC; Kapoor. 2002. *Fundamentals of Mathematical Statistics: A Modern Approach*. Eleventh. Sultan Chand; Sons.

Teetor, Paul. 2011. *R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics*. " O'Reilly Media, Inc.".