Probability Models

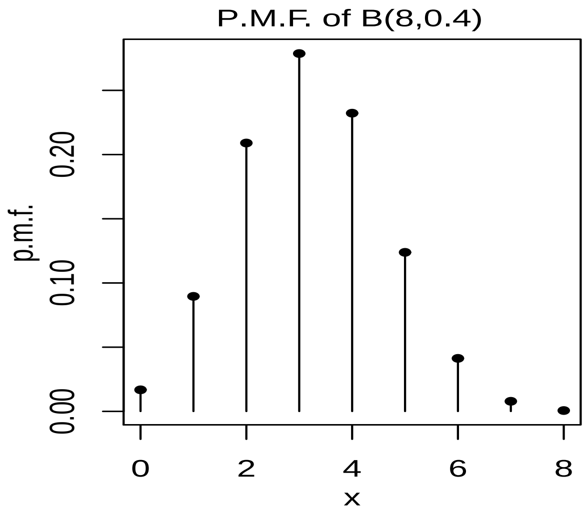
1 Discrete Models

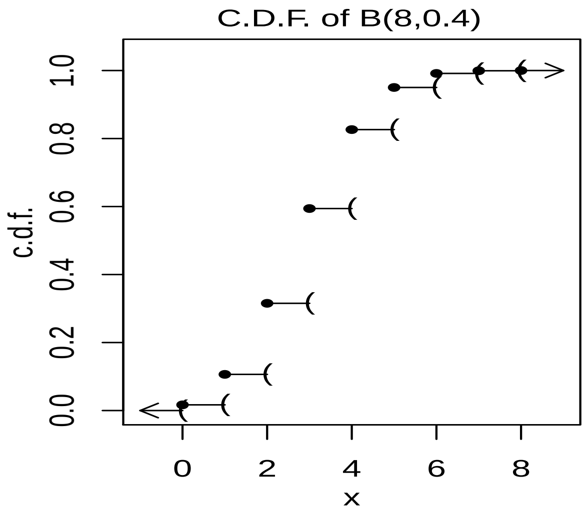
Here we discuss three important discrete distributions, which may be used to model different real life situations. In somewhat abstract terms, these real life situations will involve studying some aspect of a sequence of independent and identically distributed (i.i.d.) Bernoulli trials with constant probability of Success p.

1.1 Binomial Distribution

If $X \sim$B(n, p), it's probability mass function (p.m.f.) is given by $P(X = k) = \binom{n}{k} p^k(1-p)^{n-k}$ for $k = 0, 1, \ldots, n$ and its cumulative distribution function (c.d.f.) is given by $F(x) = P(X \le x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k(1-p)^{n-k} \binom{n}{k} p^k(1-p)^{n-k}$. Here are some plots of B(n, p) p.m.f.'s and c.d.f.'s and the `R` codes for generating them.

```
> x<-0:8
> p<-dbinom(x,8,0.4)
> c<-pbinom(x,8,0.4)
> pmfplot(x,p,"P.M.F. of B(8,0.4)")
> cdfplot(x,c,"C.D.F. of B(8,0.4)")
```





Here are some quantile calculations:

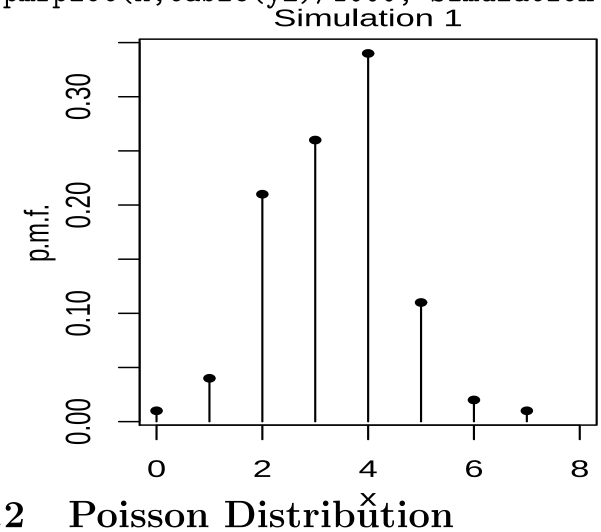
```
> qbinom(0.25,8,0.4)
[1] 2
> qbinom(0.75,8,0.4)
[1] 4
> qbinom(0.95,8,0.4)
[1] 5
> qbinom(0.999,8,0.4)
[1] 7
```

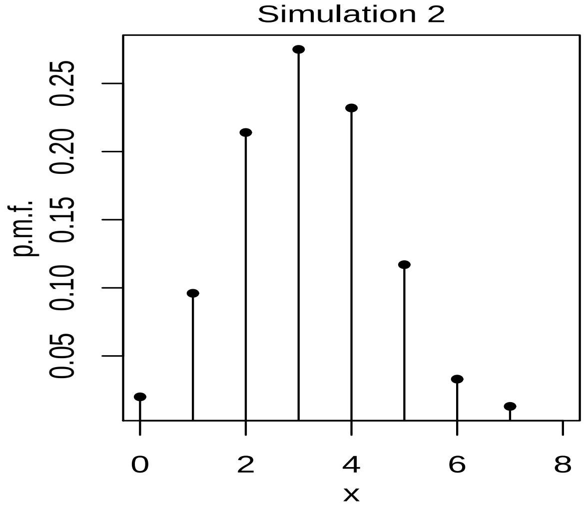
```
> qbinom(0.5,8,0.4)
[1] 3
> qbinom(0.85,8,0.4)
[1] 5
> qbinom(0.99,8,0.4)
[1] 6
> qbinom(0.9999,8,0.4)
[1] 8
```

Here are some simulations:

```
> y1<-rbinom(100,8,0.4)
> table(y1)

y1
 0  1  2  3  4  5  6  7
 1  4 21 26 34 11  2  1
> y2<-rbinom(1000,8,0.4)
> table(y2)
y2
  0   1   2   3   4   5   6   7
 20  96 214 275 232 117  33  13
> round(p,digits=4)
[1] 0.0168 0.0896 0.2090 0.2787 0.2322 0.1239 0.0413 0.0079 0.0007
> pmfplot(x,table(y1)/100,"Simulation 1")
> pmfplot(x,table(y2)/1000,"Simulation 2")
```





1.2 Poisson Distribution

For some $\lambda > 0$, if X ∼Poisson(λ) its p.m.f is given by $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$, and its c.d.f. is given by $F(x) = P(X \le x) = e^{-\lambda}\sum_{k=0}^{\lfloor x\rfloor}\frac{\lambda^k}{k!}$. Poisson distribution has been introduced earlier in Example 27 in Chapter 3 consisting of notes on Random Variables. It arises as a limiting distribution of B(n, p_n) for $n \to \infty$ such that $np_n \to \lambda$ as $n \to \infty$. This fact has actually been proved in Example 28 of Chapter 3 using characteristic function, where it is also mentioned that a direct proof will be provided in a later chapter. Here is the direct proof which is much more straight forward than the proof given in Example 28 of Chapter 3 using characteristic function.

If X ∼B(n, p_n), then for a fixed non-negative integer k

$$\lim_{n\to\infty} P(X = k) = \lim_{n\to\infty}\left(\begin{array}{c} n \\ k \end{array}\right)p_n^k(1-p_n)^{n-k}$$

$$= \lim_{n\to\infty}\frac{1}{k!}\times\frac{n(n-1)\cdots(n-k+1)(n-k)!}{(n-k)!}\times\frac{(np_n)^k}{n^k}\times\left(1-\frac{np_n}{n}\right)^{n-k}$$

$$= \frac{1}{k!}\times\left\{\lim_{n\to\infty}1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right\}\times\left(\lim_{n\to\infty}x_n\right)^k\times\lim_{n\to\infty}\left(1-\frac{x_n}{n}\right)^n\times\lim_{n\to\infty}(1-p_n)^{-k}$$

where $x_n = np_n$ is such that $\lim_{n\to\infty} x_n = \lambda$. Now the term in $\{\cdot\}$ equals 1, and it is well-known that $\lim_{n\to\infty}\left(1-\frac{x_n}{n}\right)^n = e^{-x}$ if $\lim_{n\to\infty} x_n = x$. Further $\lim_{n\to\infty} p_n = 0$ as np_n has a limit. Thus the above limit equals $e^{-\lambda}\frac{\lambda^k}{k!}$.

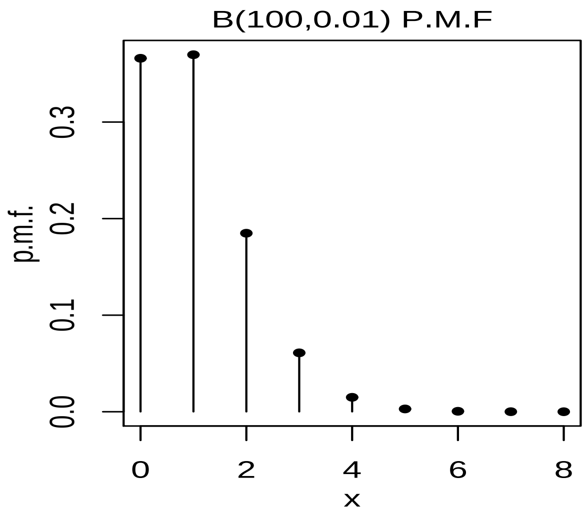
As in the case of Binomial distribution, the standard way to visualise a discrete distribution is to compare successive terms in its p.m.f.. In case of Poisson, for $k = 0, 1, 2, \ldots$ this yields

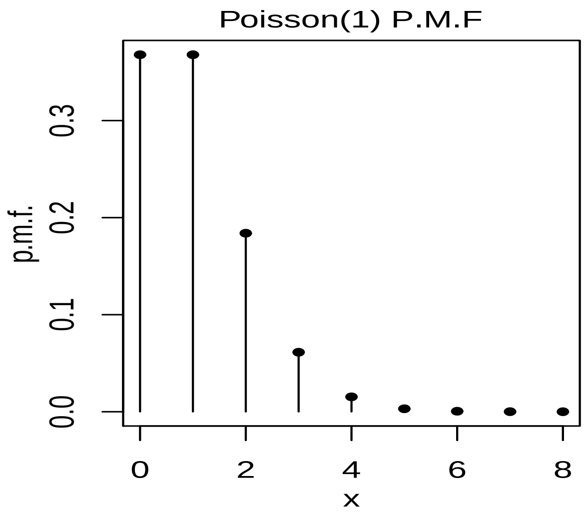
$$\frac{P(X=k+1)}{P(X=k)} \begin{matrix}>\\=\\<\end{matrix} 1 \iff \frac{\lambda}{k+1} \begin{matrix}>\\=\\<\end{matrix} 1 \iff k \begin{matrix}<\\=\\>\end{matrix} \lambda-1$$

This means that $P(X = k+1)$ will remain greater than $P(X = k)$ as long as $k < \lambda - 1$ and then the p.m.f. starts declining. Thus except for the rare occasion of λ being a positive integer, in all other cases, the maximal value of k for which $P(X = k+1) > P(X = k)$ is $\lfloor \lambda - 1 \rfloor$ yielding $\lfloor \lambda - 1 \rfloor + 1 = \lfloor \lambda \rfloor$ as the unique mode. In case λ is a positive integer, with $P(X = \lambda) = P(X = \lambda - 1)$ being the maximal value of the p.m.f., the distribution is bimodal with both $\lambda - 1$ and λ being its two modes.

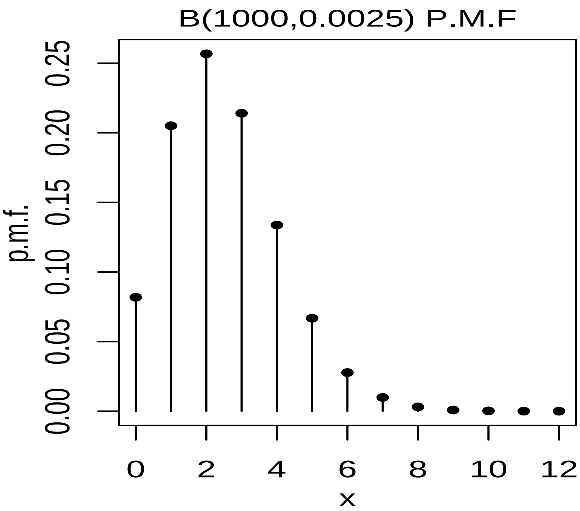
Here are some plots comparing the p.m.f.'s of the $\mathrm{B}(n, p)$ distributions with large n and small p and the corresponding Poisson($\lambda = np$) approximations. Also note how Poisson(1) is bimodal with 0 and 1 being its two modes while Poisson(2.5) is unimodal with its mode at 2.

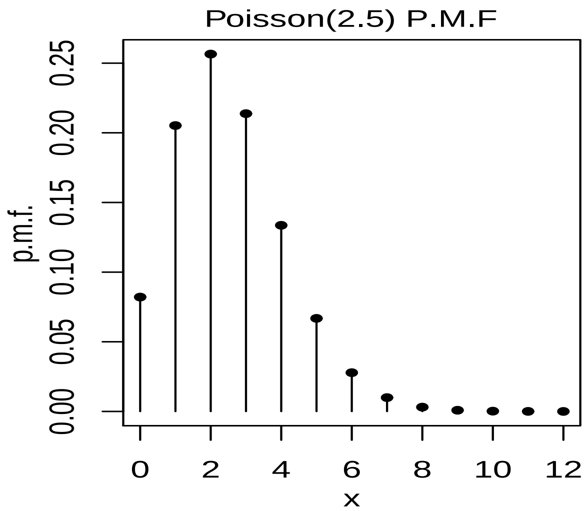
```
> p1<-dbinom(x,100,0.01)
> p2<-dpois(x,1)
> round(p1,digits=4)
[1] 0.3660 0.3697 0.1849 0.0610 0.0149 0.0029 0.0005 0.0001 0.0000
> round(p2,digits=4)
[1] 0.3679 0.3679 0.1839 0.0613 0.0153 0.0031 0.0005 0.0001 0.0000
> pmfplot(x,p1,"B(100,0.01) P.M.F")
> pmfplot(x,p2,"Poisson(1) P.M.F")
```





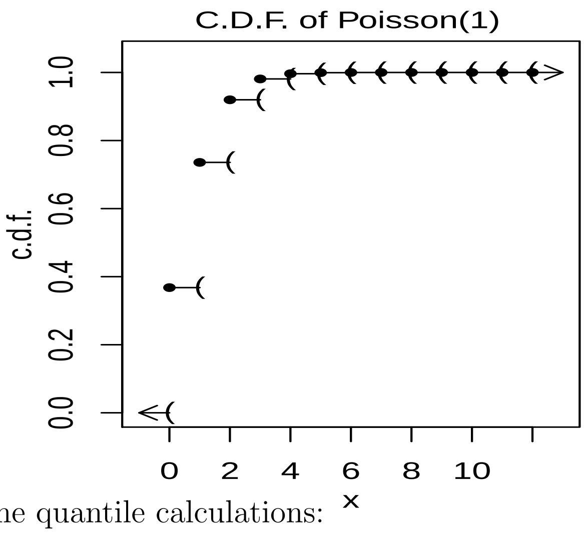

```
> x2<-0:12
> p3<-dbinom(x2,1000,0.0025)
> p4<-dpois(x2,2.5)
> round(p3,digits=4)
 [1] 0.0818 0.2051 0.2567 0.2141 0.1337 0.0668 0.0277 0.0099 0.0031 0.0008
[11] 0.0002 0.0000 0.0000
> round(p4,digits=4)
 [1] 0.0821 0.2052 0.2565 0.2138 0.1336 0.0668 0.0278 0.0099 0.0031 0.0009
[11] 0.0002 0.0000 0.0000
> pmfplot(x2,p3,"B(1000,0.0025) P.M.F")
> pmfplot(x2,p4,"Poisson(2.5) P.M.F")
```

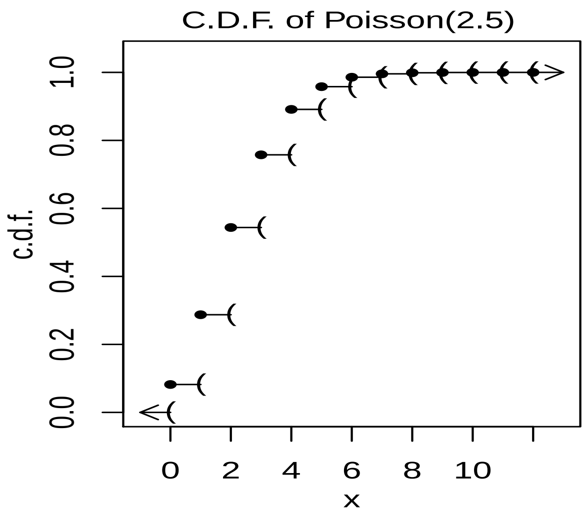




Here are the c.d.f.'s of these Poisson distributions.

```
> c1<-ppois(x,1)
> c2<-ppois(x2,2.5)
> cdfplot(x,c1,"C.D.F. of Poisson(1)")
> cdfplot(x2,c2,"C.D.F. of Poisson(2.5)")
```





Some quantile calculations:

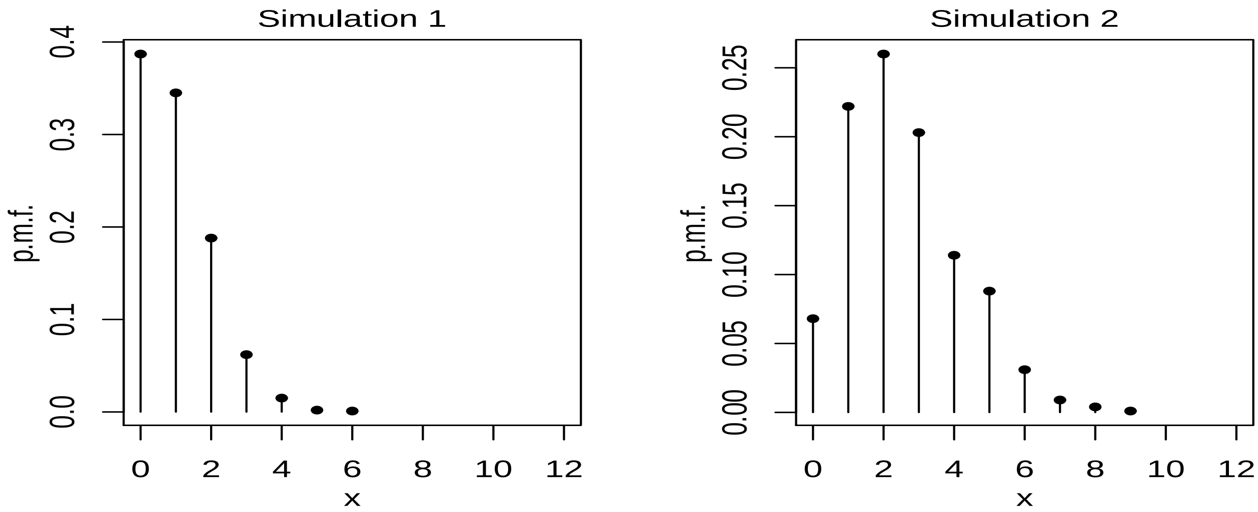
```
> qpois(0.25,1)
[1] 0
> qpois(0.5,1)
[1] 1
> qpois(0.75,1)
[1] 2
> qpois(0.9,1)
[1] 2
> qpois(0.99,1)
[1] 4
```

```
> qpois(0.25,2.5)
[1] 1
> qpois(0.5,2.5)
[1] 2
> qpois(0.75,2.5)
[1] 3
> qpois(0.9,2.5)
[1] 5
> qpois(0.99,2.5)
[1] 7
```

Some simulations:

```
> y1<-rpois(1000,1)
> table(y1)
y1
  0   1   2   3   4   5   6
387 345 188  62  15   2   1
> pmfplot(x,table(y1)/1000,"Simulation 1")
```

```
> y2<-rpois(1000,2.5)
> table(y2)
y2
 0   1   2   3   4   5   6  7  8  9
68 222 260 203 114  88  31  9  4  1
> pmfplot(x,table(y2)/1000,"Simulation 2")
```



In **Example 27** in page 69 of **Chapter 3** (the notes on Random Variables) it has been found that the moment generating function (m.g.f.) of the Poisson(λ) random variable is given by $M(t) = \exp\{\lambda(e^t - 1)\}$. Thus its mean is given by

$$M'(0) = \frac{d}{dt}M(t)\bigg|_{t=0} = \lambda e^t \exp\left\{\lambda\left(e^t - 1\right)\right\}\Big|_{t=0} = \lambda$$

and its variance is given by

$$M''(0) - (M'(0))^2 = \frac{d}{dt}\lambda e^t \exp\left\{\lambda\left(e^t - 1\right)\right\}\bigg|_{t=0} - \lambda^2 = \left(\lambda^2 e^{2t} + \lambda e^t\right)\exp\left\{\lambda\left(e^t - 1\right)\right\}\Big|_{t=0} - \lambda^2 = \lambda.$$

.

There is another very important way of viewing the Poisson distribution, in terms of what is called a Poisson Process. Consider keeping track of occurrence of some random events in space or time. It will be easier to explain it in terms of time. Starting at some time point labelled as 0, one is interested in counting the number of times a certain event of interest, such as arrival of customers/orders/phone-calls/failures etc. has occurred till some point of time t. Let $N(t)$ denote the number of times the event of interest has occurred in the time interval $[0, t]$. Since the events occur at random points of time, in order to get a handle on the distribution of $N(t)$, it is first necessary to postulate or characterise the nature of the randomness of the occurrence of these events as listed below. It is assumed that

i. $N(0) = 0$.

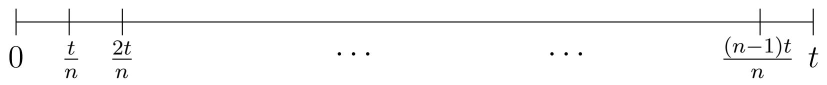
ii. If $(t_1, t_2] \cap (t_3, t_4] = \phi$, $N(t_2) - N(t_1)$, which denotes the number of events occurring in the time interval $(t_1, t_2]$, is independent of $N(t_4) - N(t_3)$, the number of events occurring in the time interval $(t_3, t_4]$. In words, number of events occurring in disjoint time intervals are independent of each other.

iii. $\lim_{h\to 0} \frac{1}{h} P\left[N(t+h) - N(t) = 1\right] = \lambda$ and $\lim_{h\to 0} \frac{1}{h} P\left[N(t+h) - N(t) \geq 2\right] = 0$. In words, probability of 2 or more events happening in a tiny interval $(t, t+h]$ of length h goes to 0 faster than the length of the interval itself, and the probability of exactly one event happening in that tiny interval $(t, t+h]$ of length h is proportional to the length of the interval with the proportionality constant λ, which in this light may be viewed as the *rate* of the occurrence of the event.

If events happen according to the above three postulates, then $N(t)$ is said to follow a Homogeneous Poisson Process (HPP), and it may be shown that $N(t)$ ~Poisson(λt). It is very important to note a couple of points. $N(t)$ is a *stochastic process* in which $\{N(t) : t \geq 0\}$ is considered in its totality and is called a HPP instead of considering it at only one given fixed time instance like say t_0 at which point $N(t_0)$, which is a random variable, happens to have a Poisson distribution. And the second point to note is that a HPP need not be restricted to counting occurrences of events only in time. The time intervals mentioned in the above postulates may easily be replaced by spatial intervals as well, and the HPP may again be used to model things like number of defects detected on the surface of a car of a certain area, or the number of chocolate chips found in a cookie of a certain volume.

For a rigorous proof that $N(t) \sim$Poisson(λt) one may look into the book *Stochastic Process*(1996), Second Edition, Wiley, by Sheldon M. Ross pp.60-63, but intuitively it is fairly straight forward to see why, as explained below.

Since we are interested in finding the distribution of the number of events occurring in the time interval $[0, t]$, let's divide the interval into n equal parts, each one of length $\frac{t}{n}$ as in the following diagram:



If n is large (actually the exact answer can be obtained only with $\lim_{n\to\infty}$), by postulate **iii**, in every interval of length $\frac{t}{n}$, there is a negligible probability of 2 or more events happening and the probability of one event happening is approximately same as $p_n = \lambda\frac{t}{n}$. That is (approximately speaking) in each one of those n intervals either the event of interest occurs (success) or it doesn't (failure). Thus a Bernoulli trial is taking place in each one of those n tiny intervals with the probability of success being p_n. Furthermore by postulate **ii**, since the n intervals are disjoint from each other, these n Bernoulli trials are mutually independent. Therefore $N(t)$, which is the total number of events happening in $[0, t]$, which in turn is same as the total number of successes in these n independent Bernoulli trials, may be approximated by a $\mathrm{B}(n, p_n)$ random variable with $\lim_{n\to\infty} np_n = \lambda t$. Thus as is shown in the beginning of this sub-section, as $n \to \infty$ the distribution of $N(t)$ is same as that of a Poisson(λt) random variable.

1.3 Negative Binomial Distribution

Recall that at the beginning of Chapter 4 on Discrete Distributions, it is stated that the fundamental building block of discrete distributions are Bernoulli trials (as we just saw even for the Poisson distribution above). Consider independent and identically distributed Bernoulli trials with probability of success p. Now consider the random variable X denoting the number of trials one has to wait till one gets n successes, for some fixed positive integer n. Such an X is said to have a Negative Binomial Distribution with parameters n an p and is denoted by $X \sim$NB(n, p). As examples think about counting the total number of tosses one has to conduct in order to get say 10 heads, or the total number of job interviews a candidate has to attend in order to end up with 2 job offers.

We begin by figuring out the p.m.f. of the NB(n, p) random variable. First note that the support of X, the set where it concentrates all its probabilities or the set of possible values that X can take, is given by $\mathcal{X} = \{n, n+1, n+2, \ldots\}$ which is countably infinite. In order to find the p.m.f. of X, fix a $k \in \{0, 1, 2, \ldots\}$ and consider the event $\{X = n+k\}$. To find $P(X = n+k)$ one needs to figure out the basic fundamental outcomes or ω's that belong to the event set $A = \{X = n + k\} = \{\omega \in \Omega : X(\omega) = n + k\}$. Note that as in the case of Binomial model, the ω's can be represented by a string of Successes (S's) and Failures (F's) such as $FSS \cdots SFS$ with the i-th symbol representing the outcome of the i-th Bernoulli trial for $i = 1, 2, \ldots$. Now for a ω to belong to A, it must have exactly n S's (as the experiment involves keeping on conducting the Bernoulli trials till one gets exactly n successes) and k F's (as the total number of trials is $n + k$ - that is exactly what $X = n + k$ means

- and there are n S's, the remaining k must be F's) with the last symbol necessarily being an S (as the experiment must stop as soon as the n-th success is observed). Since the trials are independent with $P(S) = p$, so that $P(F) = 1 - P(S) = 1 - p = q$ (say), for any $\omega \in A$, $P(\omega) = p^n q^k$. Thus in order to find $P(X = n + k)$, now all one has to do is figure out how many such distinct ω's could be there in the set A. Since such an ω is an $(n+k)$-long string of S and F's with n S's, k F's and an S in the end, it is same as choosing k positions from the $n + k - 1$ positions to be filled with F's and the remaining ones filled with S's, which can be done in $\binom{n+k-1}{k}$ ways. Thus there are $\binom{n+k-1}{k}$ many ω's in A with each ω having a probability of $p^n q^k$ yielding the Negative Binomial p.m.f. as $P(X = n + k) = \binom{n+k-1}{k} p^n q^k$ for $k = 0, 1, 2, \ldots$.

The proof that $\sum_{k=0}^{\infty} P(X = n + k) = 1$ and the name Negative Binomial come from the negative binomial coefficients and the negative binomial theorem introduced in page 61 of the Chapter 3: Random Variable notes. Recall that $\binom{n+k-1}{k} = (-1)^k \binom{-n}{k}$ and

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

with the last but one equality following from the negative binomial theorem.

As usual for visualisation of the distribution, like Binomial and Poisson, comparison of successive terms in the p.m.f. yields

$$\frac{P(X = n+k+1)}{P(X = n+k)} = \frac{(n+k)!k!(n-1)!q^{k+1}p^n}{(n+k-1)!(k+1)!(n-1)!q^k p^n} = \frac{n+k}{k+1} q \begin{array}{c} > \\ = \\ < \end{array} 1 \Longleftrightarrow k \begin{array}{c} < \\ = \\ > \end{array} \frac{n-1}{p} - n$$

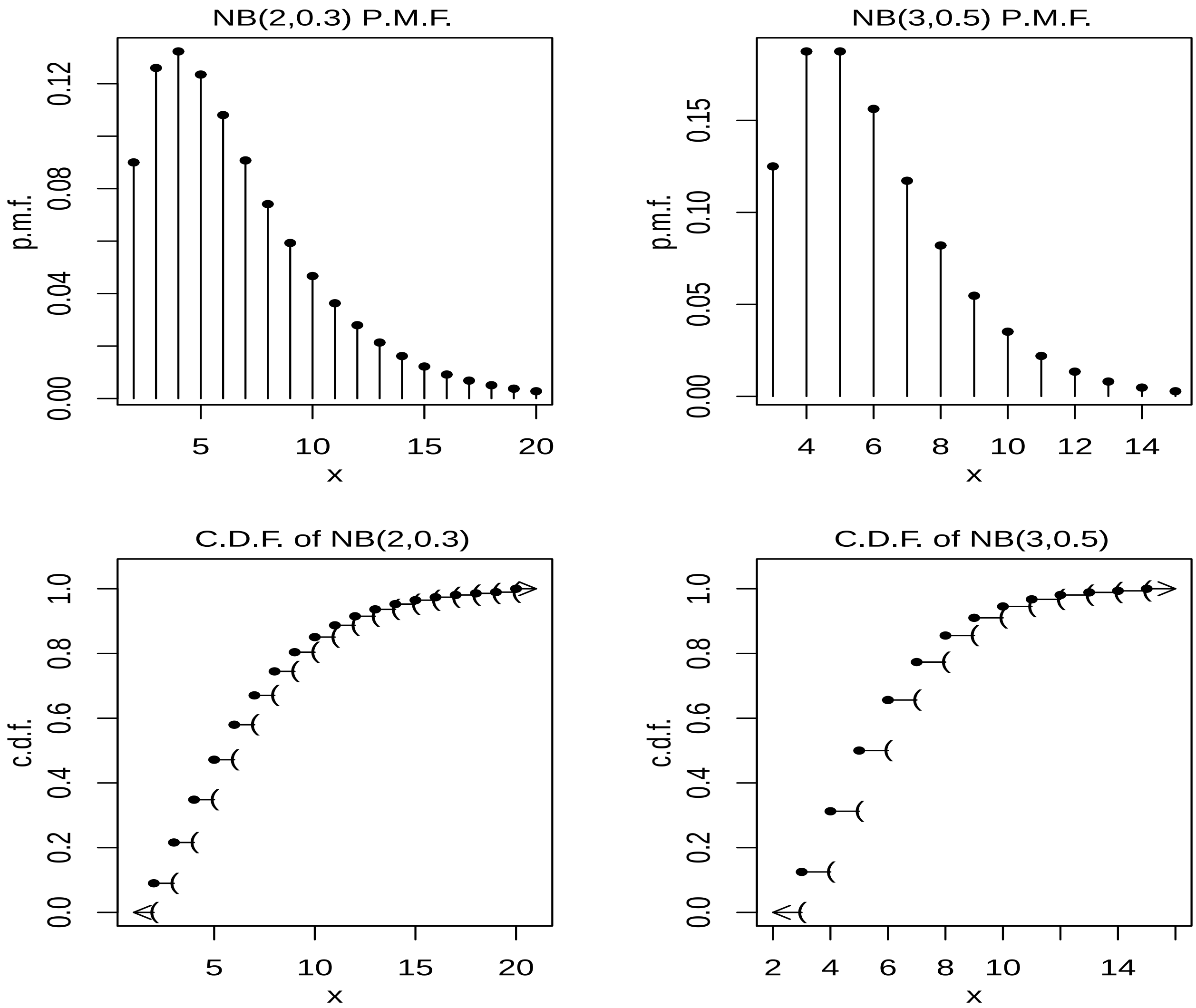
which goes on to show that starting from n, the p.m.f. increases only up to $k \leq \lfloor \frac{n-1}{p} - n \rfloor + 1 = \lfloor (n-1)\frac{q}{p} \rfloor$, after which starts declining. Thus if $(n-1)\frac{q}{p} \geq 0$ is not an integer, negative binomial distribution has a unique mode at $n + \lfloor (n-1)\frac{q}{p} \rfloor$, and in case $(n-1)\frac{q}{p}$ is an integer, like Binomial and Poisson, negative binomial also is bimodal with two side by side modes at $n + (n-1)\frac{q}{p}$ and $(n-1)\left(1 + \frac{q}{p}\right)$, with the exception of the case of $n = 1$, in which case it has a unique mode at 0.

Here are the plots of the p.m.f.'s of a couple of NB(n, p) distribution. Note the bimodal case for $n = 3$ and $p = 1/2$. Also note that the R's `xnbinom(...)` commands for `x=d,p,q` and `r` provide the desired quantities for an NB$(n, p) - n$ distribution with support $\{0, 1, 2, \ldots\}$, with the NB(n, p) distribution as defined here with support $\{n, n+1, n+2, \ldots\}$. Thus the slight modification with "$+n$" in the commands for the plots.

```
> x<-0:18
> pmfplot(x+2,dnbinom(x,2,0.3),"NB(2,0.3) P.M.F.")
> x2<-0:12
> pmfplot(x2+3,dnbinom(x2,3,0.5),"NB(3,0.5) P.M.F.")
```

The plots of the c.d.f.'s of the distributions also follow.

```
> cdfplot(x+2,pnbinom(x,2,0.3),"C.D.F. of NB(2,0.3)")
> cdfplot(x2+3,pnbinom(x2,3,0.5),"C.D.F. of NB(3,0.5)")
```



Some quantile calculations:

```
> qnbinom(0.25,2,0.3)
[1] 2
> qnbinom(0.5,2,0.3)
[1] 4
> qnbinom(0.75,2,0.3)
[1] 7
> qnbinom(0.9,2,0.3)
[1] 10
> qnbinom(0.99,2,0.3)
[1] 18
```

```
> qnbinom(0.25,3,0.5)
[1] 1
> qnbinom(0.5,3,0.5)
[1] 2
> qnbinom(0.75,3,0.5)
[1] 4
> qnbinom(0.9,3,0.5)
[1] 6
> qnbinom(0.99,3,0.5)
[1] 11
```

Now what about the moments? Though it is possible to obtain the moments by direct calculations with the NB(n,p) p.m.f. and the aid of the negative binomial theorem, deriving them using the m.g.f. route is the alternative adopted here.

Consider the random variable introduced in **Example 4** in Chapter 3, where one counts the number of Tails till the first Head appears in an independent sequence of tosses of a coin with $P(H) = p$. Here let's denote that random variable by Z_1, with Head and Tail replaced by Success and Failure respectively, and "independent tosses" replaced by i.i.d. sequence of Bernoulli trials. Now let $Y_1 = Z_1 + 1$ denote the number of independent Bernoulli trials required to obtain the first Success, and in general for $i \geq 2$ let Y_i denote the number of trials required between the $i-1$-st and the i-th Success.

Though it has not been formally said so earlier, for obvious reason Y_1 is said to have a **Geometric** distribution, or alternatively a **Geometric Random Variable** and is denoted by $Y_1 \sim$Geom(p). Now it is easy to see that all the $Y_1, Y_2, \ldots$ are i.i.d. Geom(p) and by the definition of the Negative Binomial random variable X given in the first paragraph of this sub-section, where it is defined as the number of trials required to obtain n Successes in an i.i.d. sequence of Bernoulli trials, $X = \sum_{i=1}^n Y_i$.

Now in **Example 4 (Continued)** in page 66 of §6.2 of **Chapter 3** it has been shown that the m.g.f. of Z_1 is given by $M_{Z_1}(t) = \frac{p}{1-qe^t}$. Thus the m.g.f. of $Y_1 = Z_1 + 1$ is given by

$$M_{Y_1}(t) = E\left[e^{tY_1}\right] = E\left[e^{t(Z_1+1)}\right] = e^t E\left[e^{tZ_1}\right] = e^t M_{Z_1}(t) = \frac{pe^t}{1-qe^t}$$

and by the result for the m.g.f. of i.i.d. sum proven in page 68 of **Chapter 3**, the m.g.f. of NB(n,p) random variable X is given by

$$M_X(t) = [M_{Y_1}(t)]^n = \left(\frac{pe^t}{1-qe^t}\right)^n.$$

Thus the mean of NB(n,p) is given by

$$E[X] = M_X'(0) = \frac{d}{dt}M_X(t)\bigg|_{t=0} = \frac{np^n e^{tn}}{(1-qe^t)^{n+1}}\bigg|_{t=0} = \frac{n}{p}$$

and

$$E\left[X^2\right] = M_X''(0) = \frac{d}{dt}M_X'(t)\bigg|_{t=0} = \frac{np^n e^{tn}\left(n+qe^t\right)}{(1-qe^t)^{n+2}}\bigg|_{t=0} = \frac{n(n+q)}{p^2}$$

yielding

$$V[X] = \frac{n(n+q)}{p^2} - \frac{n^2}{p^2} = n\frac{q}{p^2}.$$

There is a far more elegant way of deriving the mean and variance of the Negative Binomial distribution, which is as follows. Recall that the Negative Binomial $X = \sum_{i=1}^n Y_i$, where the Y_i's are i.i.d. Geom(p). Recall that in pp.10-11 of **Chapter 3** it has been proven that $E[Z_1] = \frac{q}{p}$ and $V[Z_1] = \frac{q}{p^2}$. Since Y_i's are i.i.d. and $Y_1 = Z_1 + 1$, $E\left[Y_i\right] = 1 + \frac{q}{p} = \frac{1}{p}$ and $V\left[Y_i\right] = V\left[Z_1\right] = \frac{q}{p^2}$. Now since $X = \sum_{i=1}^n Y_i$,

$$E[X] = \sum_{i=1}^n E\left[Y_i\right] = \sum_{i=1}^n \frac{1}{p} = \frac{n}{p} \quad \text{and} \quad V[X] = \sum_{i=1}^n V\left[Y_i\right] = \sum_{i=1}^n \frac{q}{p^2} = n\frac{q}{p^2}.$$

We will end this section by studying a very important property of the NB(1,p) or the Geom(p) random variable Y, which has p.m.f. $P[Y=n] = q^{n-1}p$ for $n = 1, 2, \ldots$. First note that it is one of those rare cases of a discrete model where the c.d.f. can be expressed in closed form (instead of just a finite sum) as follows:

$$\text{For } n = 1, 2, \ldots, P[Y \le n] = \sum_{k=1}^n q^{k-1}p = p\frac{1-q^n}{1-q} = 1 - q^n$$

so that $P[Y > n] = q^n$. Now for positive integers m and n,

$$P[Y = m+n | Y > m] = \frac{P[Y = m+n \;\&\; Y > m]}{P[Y > m]} = \frac{P[Y = m+n]}{P[Y>m]} = \frac{q^{m+n-1}p}{q^m} = q^{n-1}p = P[Y=n].$$

This is a remarkable property called the **lack of memory** property, that is enjoyed by a Geom(p) random variable. Now let us see why this property is called so. The lack of memory property, in this discrete positive integer valued random variable's case states that for any positive integers m and n, $P[Y = m+n|Y > m] = P[Y = n]$. To understand what it means, let's look at the concrete example of the coin tossing experiment, where one keeps on tossing a coin with $P(H) = p$, till the first Head appears and one observes the total number of tosses required to get there. The lack of memory property states that, given the fact that the number of tosses required to obtain the first Head is more than m, the probability that one has to toss it for another n times to get the first Head, is same as the probability of the required number of tosses equalling n in a new experiment where one starts counting afresh. The fact that one has already tossed m times and has not observed a Head yet, is of no consequence, and as if the experiment is starting anew all over again from the $m+1$-st toss. Or in other words, the counting (total number of tosses) process has *forgotten* that already the coin has been tossed m times without any success of obtaining a Head. Hence is the name of this property.

What is even more remarkable, is the fact that among all discrete random variables with the set of positive integers as its support, it is only the Geom(p) that has this lack of memory property. First note that the lack of memory property of the Geometric random variable Y may be alternatively viewed as follows:

$$P[Y > m+n|Y > m] = \frac{P[Y > m+n \;\&\; Y > m]}{P[Y > m]} = \frac{P[Y > m+n]}{P[Y > m]} = \frac{q^{m+n}}{q^m} = q^n = P[Y > n].$$

Now consider an arbitrary positive integer valued random variable Y, which has the lack of memory property $P[Y > m+n] = P[Y > m]P[Y > n]$. We are to show that such Y must be a Geometric random variable. For this purpose it will suffice to show that $P[Y > n] = q^n$ for some q. Define $q = P[Y > 1]$ satisfying $P[Y > n] = q^n$ for $n = 1$. Thus since the assertion is true for $n = 1$, the assertion will be proved by induction if one can show that $P[Y > n] = q^n \Rightarrow P[Y > n+1] = q^{n+1}$. By the lack of memory property,

$$P[Y > n+1] = P[Y > n]P[Y > 1] = q^n q = q^{n+1}$$

where the second equality follows from the induction hypothesis of $P[Y > n] = q^n$ and the definition of q. This proves that among all positive integer valued random variables, it is only the Geometric random variable that has the lack of memory property.

2 Continuous Models

Appendix

```
> pmfplot <- function(x,p,s)
+ {
+  n<-length(x)
+  plot(range(x),range(p),type="n",xlab="x",ylab="p.m.f.",main=s,font.main=1,
+       cex.main=1)
+  for(i in 1:n)
+  {
+   lines(c(x[i],x[i]),c(0,p[i]))
+   points(x[i],p[i],pch=20)
```

```
+  }
+ }

> cdfplot <- function(x,f,s)
+ {
+  n<-length(x)
+  plot(range(x)+c(-1,1),c(0,1.05),type="n",xlab="x",ylab="c.d.f.",main=s,font.main=1,ce
+  arrows(x[1],0,x[1]-1,0,length=0.1)
+  points(x[1],0,pch="(")
+  for(i in 1:(n-1))
+  {
+   lines(c(x[i],x[i+1]),c(f[i],f[i]))
+   points(x[i],f[i],pch=20)
+   points(x[i+1],f[i],pch="(")
+  }
+  arrows(x[n],1,x[n]+1,1,length=0.1)
+  points(x[n],1,pch=20)
+ }
```