Special Probability Distributions (Ghahramani, Ch. 5 & 7)

Certain classes of random experiment and random variables defined upon them turn up so often that we define and name standard functions as their distribution functions, pmfs or pdfs as appropriate.

We shall now discuss the most common examples of such random variables.

Discrete random variables

Some discrete distributions which arise frequently in modelling real world phenomena are:

- Bernoulli
- Binomial
- Geometric
- Negative Binomial
- Hypergeometric
- Poisson
- Uniform

Continuous random variables

Some continuous distributions which arise frequently in modelling real world phenomena are:

- Uniform
- Exponential
- Gamma
- Beta
- Pareto
- Weibull
- Normal
- Cauchy
- lognormal

Bernoulli Random Variables (Ghahramani 5.1)

If the random experiment has two possible outcomes (or two categories of outcomes)

- success and failure
- up and down
- defective and non-defective
- right and wrong
- true and false
- ...

A random experiment in which such a dichotomy is observed is called *a Bernoulli trial*.

Jacob Bernoulli [27/12/1654 – 16/08/1705]

We write $\Omega = \{S, F\}$ and let $X(S) = 1$ and $X(F) = 0$.

The random variable X is known as a *Bernoulli random variable*. If p is the probability of a success, then the probability mass function is

$$
\begin{aligned}
p(0) &= 1 - p \\
p(1) &= p
\end{aligned}
$$

The value $p \in [0, 1]$ is a *parameter*. By varying it, we get different members of the family of Bernoulli random variables.

We say X has a *Bernoulli distribution* with parameter p.

Bernoulli mean and variance

Applying our formulae for expectations we have

$$\begin{aligned}
\mathbb{E}(X) &= 0 \times (1-p) + 1 \times p = p. \\
\mathbb{E}(X^2) &= 0^2 \times (1-p) + 1^2 \times p = p. \\
V(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = p - p^2 = p(1-p).
\end{aligned}$$

Many random experiments take the form of sequences of Bernoulli trials. Moreover the physical processes that govern the experiments are such that it is reasonable to assume that the outcomes of the Bernoulli trials are mutually independent.

Thus, for example, the Bernoulli trials might be tosses of separate coins, or the same coin at different time points.

The Bernoulli, Binomial, Geometric and Negative Binomial random variables all arise in the context of a *sequence of independent Bernoulli trials*. They each summarise different aspects of the observed sequence of "successes" and "failures".

Binomial random variables (Ghahramani 5.1)

Consider a sequence of n independent Bernoulli trials with $p = \mathbb{P}(\text{success})$. The sample space Ω for such an experiment could be taken to be the set of all sequences of the form

$$\omega = \underbrace{S\ S\ F\ F\ S\ S\ F \ldots S}_{n \text{ letters}}$$

and the probability of any given sequence occurring is

$$\mathbb{P}(\{\omega\}) = p^{\text{no. of successes}} \times (1-p)^{\text{no. of failures}}. \qquad (*)$$

However we usually aren't interested in the precise sequence ω that comes up. More interesting is the total number of successes.

Define the random variable $N(\omega) = \text{No. of successes in } \omega$.

To find the *pmf* p_N of N, we need $\mathbb{P}(N = k)$, which is the probability of the event:

$$A_k = \{\omega : N(\omega) = k\} \quad k = 0, 1, \ldots, n.$$

From $(*)$ above we know that the probability of any given $\omega \in A_k$ is

$$p^k(1-p)^{n-k}.$$

Thus to get $\mathbb{P}(A_k)$ all we need to do is count how many ω there are in A_k.

This is known to be

$$\left(\begin{array}{c} n \\ \\ k \end{array}\right) = \frac{n!}{k!(n-k)!}.$$

So we have

$$p_N(k) = \mathbb{P}(N = k) = \mathbb{P}(A_k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Binomial distribution

If $X =$ no. of successes in n independent Bernoulli trials with $p = \mathbb{P}(\text{success})$ then

$$p_X(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n$$

and we say X has a *Binomial distribution* with parameters n and p and write $X \stackrel{d}{=} \mathrm{Bi}(n, p)$.

Note that we can then write $X \stackrel{d}{=} \mathrm{Bi}(1, p)$ for a Bernoulli distribution.

Using the Binomial Theorem we can verify that

$$\begin{aligned}\sum_{x=0}^{n} p_X(x) &= \sum_{x=0}^{n} \tbinom{n}{x} p^x (1-p)^{n-x} \\ &= (p+1-p)^n \\ &= 1^n = 1\end{aligned}$$

Note: The *Binomial Theorem* is very important in many areas of mathematics and is certainly something you should know. It states that for integer n

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

Binomial mean and variance

Applying our formulae for expectations we can deduce that

$$\begin{aligned} \mathbb{E}(X) &= np. \\ \mathbb{E}(X(X-1)) &= n(n-1)p^2. \\ V(X) &= \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2 \\ &= np(1-p). \end{aligned}$$

Example - Tay-Sachs disease

This is a hereditary metabolic disorder caused by a recessive genetic trait. When both parents are carriers their child has probability $\frac{1}{4}$ of being born with the disease (Mendel's first law).

If such parents have four children, what is the probability distribution for X = no. of children born with the disease? State assumptions.

Sampling with replacement

Suppose that a population consists of N objects, a proportion p of which are defective. A sample of n is obtained by selecting one object at random from the population, replacing it, selecting at random again, and so on.

We therefore have a sequence of independent Bernoulli trials with $\mathbb{P}(\text{success}) = p$.

Thus if X = number of defectives obtained in the sample, then $X \stackrel{d}{=} \text{Bi}(n, p)$.

Binomial distribution shape

We can show that the ratio of successive binomial probabilities $r(x)$ satisfies

$$r(x) = \frac{p_X(x)}{p_X(x-1)} = \frac{\frac{n+1}{x} - 1}{\frac{1}{p} - 1} \quad x = 1, 2, \ldots, n$$

which decreases as x increases.

The formula

$$\frac{p_X(x)}{p_X(x-1)} = \frac{\frac{n+1}{x} - 1}{\frac{1}{p} - 1} \quad x = 1, 2, \ldots, n$$

is useful for computing the binomial probabilities and is called the recursive formula.

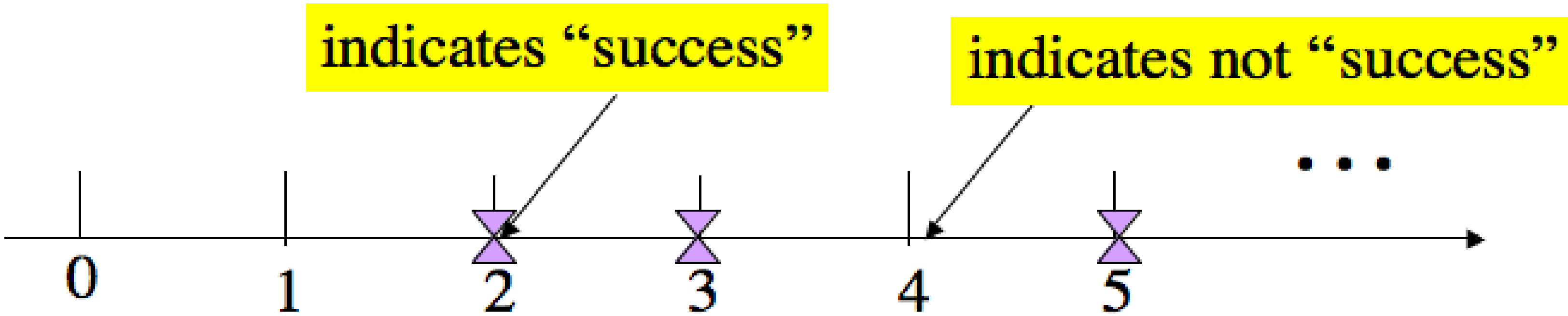
If $x < p(n+1)$ then $r(x) > 1$ and the pmf is increasing.
If $x > p(n+1)$ then $r(x) < 1$ and the pmf is decreasing.

So the Binomial distribution only has a single "peak".

Exercise: Find values of n and p so that two successive binomial probabilities are the same.

Geometric random variables (Ghahramani 5.3)

- Can we have infinitely many independent events all with the same probability (i.e. infinitely many Bernoulli trials)?



- Let N be the random variable which gives the number of "failures" before there is a "success".

- The values for N are $\{0, 1, 2, \dots\}$ and N is therefore a discrete random variable
- N is a time if the trials occur in time
- The pmf of N:

$$\begin{aligned}
\mathbb{P}(N=0) &= \mathbb{P}(\{\omega : N(\omega)=0\}) = \mathbb{P}(S) = p \\
\mathbb{P}(N=1) &= \mathbb{P}(\{\omega : N(\omega)=1\}) = \mathbb{P}(FS) = (1-p)p \\
\mathbb{P}(N=2) &= \mathbb{P}(\{\omega : N(\omega)=2\}) = \mathbb{P}(FFS) = (1-p)(1-p)p \\
& \dots \\
\mathbb{P}(N=n) &= (1-p)(1-p)\dots(1-p)p.
\end{aligned}$$

So the *pmf* for N is

$$p_N(n) = \mathbb{P}(N=n) = (1-p)^n p, \quad n = 0, 1, 2, \dots$$

- We say N has a *Geometric distribution* with parameter p and write $N \stackrel{d}{=} \mathrm{G}(p)$.

As a check look at

$$\begin{aligned}\sum_{n=0}^{\infty} \mathbb{P}(N=n) &= \sum_{n=0}^{\infty}(1-p)^n p \\ &= \frac{p}{1-(1-p)} \\ &= 1.\end{aligned}$$

Note: Remember that if $|x|<1$ then

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}.$$

This is the formula for the sum of the geometric series.

Important Note: The Geometric distribution is defined slightly differently in some texts, including Ghahramani, by counting the number of *trials* until the first success rather than the number of failures before it.

If we define $M =$ to be the number of trials until the first success then clearly $M = N + 1$ where $N \stackrel{d}{=} G(p)$ under our definition. So instead of taking the values $0, 1, 2, \ldots$ the random variable M takes values $1, 2, 3, \ldots$. The distribution of M is simply shifted in location by one unit to the right.

Example

A computer communication channel transmits bits correctly with probability 0.95 independently of all other bits. What is the probability that there are at least 2 bits transmitted correctly before there is one transmitted incorrectly?

Find the probability of at least 10 correct before the first incorrect transmission in a communication channel given that there are at least 8 correct.

Lack of memory property

A curious property of the geometric distribution is the so called "lack of memory" property. If $T \stackrel{d}{=} G(p)$ then, for $t = 0, 1, 2, \ldots$

$$\mathbb{P}(T \geq t) = p(1-p)^t + p(1-p)^{t+1} + \ldots = \frac{p(1-p)^t}{p} = (1-p)^t.$$

So for given $a, t = 0, 1, 2, \ldots$, we have:

$$\mathbb{P}(T - a \geq t \,|\, T \geq a) \quad = \quad (1-p)^t.$$

Hence given that the first a trials were all failures, the *"residual"* time $T - a$ till the first success will have the same $\mathrm{G}(p)$ distribution as the original T.

The information that there has been no successes in the past a trials has no effect on the future waiting time to a success: the process “forgets” — the past has no effect on the future.

Geometric mean and variance

Applying our formulae for expectations we can deduce that

$$\begin{aligned} \mathbb{E}(X) &= \frac{(1-p)}{p}, \\ V(X) &= \frac{(1-p)}{p^2}. \end{aligned}$$

Exercise

Use the formula in slide 150 to show

$$\mathbb{E}(X) = \frac{(1-p)}{p}.$$

Hint: If X is a discrete random variables taking values in the set of non-negative integers, the formula becomes

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} \mathbb{P}(X \geqslant n).$$

Negative Binomial random variables (Ghahramani 5.3)

Let us look at the following generalisation of geometric distributions.

Again we consider an infinite sequence of independent Bernoulli trials, each with success probability p. Let $r \geqslant 1$ be a given positive integer.

Question: Define Z to be the total number of failures before the r-th success. What is the distribution of Z?

We first observe that one way in which the event "$Z = z$" can occur is

$$\begin{array}{ccc|c} F\ F\ \ldots\ F & & S\ S\ \ldots\ S & S \\ \leftarrow\quad z\quad\rightarrow & & \leftarrow r-1 \rightarrow & \end{array}$$

The probability of this sequence is $(1-p)^z p^r$. If the first $z+r-1$ results are arranged amongst themselves, leaving the final rth S, the event $Z = z$ still occurs. This can be done in

$$\binom{z+r-1}{r-1}$$

ways; and for each arrangement, the probability is $(1-p)^z p^r$.

The pmf of Z is therefore given by

$$\begin{aligned}
p_Z(z) &= \mathbb{P}(Z = z) = \binom{z+r-1}{r-1} p^r (1-p)^z \\
&= \frac{r(r+1)\dots(r+z-1)}{z!} p^r (1-p)^z \\
&= \frac{(-1)^z(-r)(-r-1)\dots(-r-z+1)}{z!} p^r (1-p)^z \\
&= \binom{-r}{z} p^r (-(1-p))^z, \quad z = 0, 1, 2, \dots
\end{aligned}$$

where we define

$$\binom{x}{k} := \frac{x(x-1)\dots(x-k+1)}{k!}$$

for any real number x and nonnegative integer k.

Note: We define $\binom{x}{0} = 1$ for all real x.

If x is a positive integer and $k \leqslant x$, then $\binom{x}{k}$ is the usual binomial coefficient.

If r happens to be an integer we have shown that

$$\binom{-r}{z} = (-1)^z \binom{z+r-1}{r-1}.$$

Using

$$p_Z(z) = \binom{-r}{z} p^r (p-1)^z, \quad z = 0, 1, 2, \ldots$$

we can show that this is a well defined pmf for all real $r > 0$.

Extended Binomial Theorem

For any real r (as opposed to just integer r) we have

$$(1+b)^r = \sum_{k=0}^{\infty} \binom{r}{k} b^k$$

which converges provided $|b| < 1$.

Provided $p \neq 0$, it follows that

$$p^r \sum_{z=0}^{\infty} \binom{-r}{z} (p-1)^z = p^r (1 + (p-1))^{-r} = p^r p^{-r} = 1.$$

However to ensure that all the individual terms are non-negative we need to have $r > 0$.

Negative Binomial Distribution

If the random variable Z has pmf

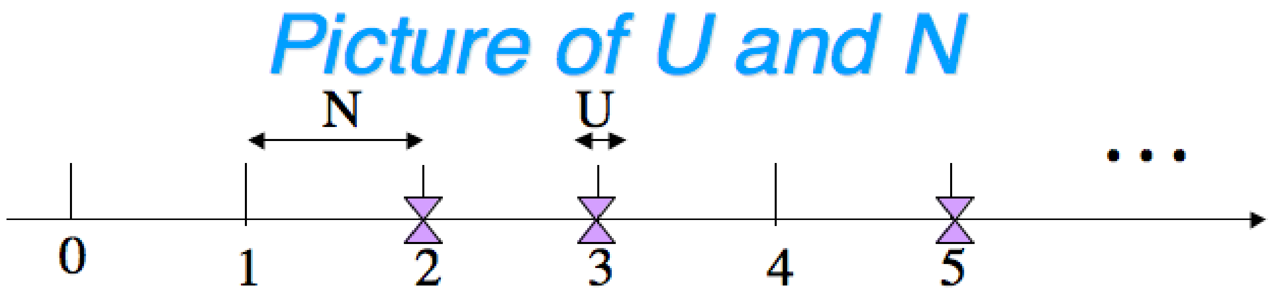
$$p_Z(z) = \binom{-r}{z} p^r (p-1)^z, \quad z = 0, 1, 2, \ldots$$

where $r > 0$ and $0 < p \leq 1$, then we say Z has a *Negative Binomial distribution* with parameters r and p and write $Z \stackrel{d}{=} \mathrm{Nb}(r, p)$.

In the special case where r is an integer then Z can be interpreted as the number of failures before the rth success in a sequence of independent Bernoulli trials with $p = \mathbb{P}(\text{success})$.

Connection with Geometric Distributions

- Let N be the random variable defined as the number of "failures" before the first "success". Then N follows $G(p)$.
- Let U be the random variable defined as the number of "failures" after the first "success" before the second "success". Also U follows $G(p)$.



- Note that N and U are different random variables, although they have the same distribution.

- Consider now the total number of failures, Z, till the rth success. Namely, $Z \stackrel{d}{=} \mathrm{NB}(r, p)$.
- For $r = 1$, Z is always N.
- For $r = 2$, $Z = N + U$.
- In the previous picture, for $r = 2$, we have $N = 1$, $U = 0$, $Z = 1$.
- For general r, the random variable Z is the sum of r independent $G(p)$ random variables:

$$Z = X_1 + \cdots + X_r$$

where X_i denotes the number of failures after the $(i-1)$-th "success" and before the i-th success.

Negative binomial mean and variance

The above decomposition

$$Z = X_1 + \cdots + X_r$$

allows us to deduce that

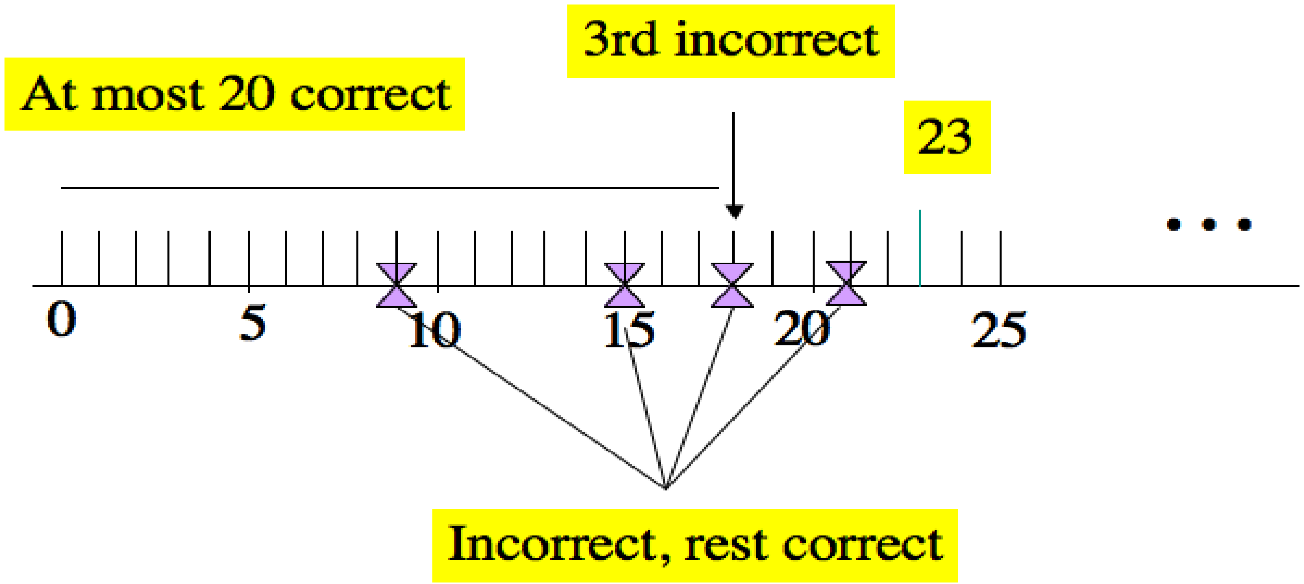
$$\begin{aligned} \mathbb{E}(X) &= \frac{r(1-p)}{p}, \\ V(X) &= \frac{r(1-p)}{p^2}. \end{aligned}$$

The precise reason why we can reach these formulas will be clear in the later stage of the course after we learn the concept of independent random variables.

Connection with Binomial Distributions

Consider the previous example of bits being transmitted correctly with probability 0.95. What is the probability that there are at most 20 correct transmissions till the third incorrect transmission?

- Here $r = 3$ and we require $\mathbb{P}(Z \leq 20)$
- Let X be the number of incorrect transmissions in the first 23 transmissions


At most 20 correct
3rd incorrect
23
• • •
0
5
10
15
20
25
Incorrect, rest correct

More generally, there is a relationship between the Negative Binomial and Binomial distributions. Indeed, the event $\{Z \leq n - r\}$ is the same as the event {at most $n - r$ failures before the rth success}, which is the same as the event {at most n trials to get r successes} which is the same as {number of successes in first n trials $\geq r$}.

To summarise, given $n \geqslant r$ and

$$Z \stackrel{d}{=} \mathrm{NB}(r, p),\ X \stackrel{d}{=} \mathrm{Bi}(n, p),$$

we have

$$\mathbb{P}(Z \leqslant n - r) = \mathbb{P}(X \geqslant r).$$

Example

To complete her degree a part time student needs to do three more subjects. Assuming she can only take one subject per semester, and that she passes a subject with probability 0.85 independently of her past results, find the probability that she will need more than 2 but not more than 3 years to graduate?

Important Note: Ghahramani defines the Negative binomial distribution as the sum of r independent Geometric random variables using its definition for the Geometric (which differs from ours). So in Ghahramani the Negative binomial takes values $(r, r+1, r+2, \ldots)$ and is simply shifted in location r units to the right.

Negative Binomial shape

We can show that the ratio of successive negative binomial probabilities $r(x)$ satisfies

$$r(z) = \frac{p_Z(z)}{p_Z(z-1)} = \left(\frac{r-1}{z} + 1\right)(1-p), \quad z = 1, 2, \ldots$$

which decreases as z increases.

The formula

$$\frac{p_Z(z)}{p_Z(z-1)} = \left(\frac{r-1}{z} + 1\right)(1-p), \quad z = 1, 2, \ldots$$

is useful for computing negative binomial probabilities and is called the recursive formula.

- If $z < \frac{1-p}{p}(r-1)$ then $r(z) > 1$ and the pmf is increasing.
- If $z > \frac{1-p}{p}(r-1)$ then $r(z) < 1$ and the pmf is decreasing.

So the Negative Binomial distribution only has a single "peak".

Hypergeometric random variables (Ghahramani 5.3)

When we looked at binomial random variables, we considered the experiment of sampling with replacement. In many, if not most, experiments of this type it is more natural to sample without replacement. If we perform such an experiment, then the number of successes is no longer binomially distributed.

We need a different distribution to describe the number of successes - the hypergeometric distribution.

- The population consists of N objects, a proportion p of which are defective.
- The number of defective items in the population is therefore $D = Np$.
- A sample of n is obtained by selecting n objects at random either all at once or sequentially without replacement. The two procedures are equivalent.
- Easiest to assume that the items are labelled: say, r_1, $\ldots$, r_D for the defetives and $b_1, \ldots, b_{N-D}$ for the non-defectives.

The pmf

Let X be number of defectives obtained in the sample of n.

Then the pmf of X is given by

$$p_X(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} \quad (x = 0, 1, 2, \ldots, n).$$

If $X = x$, then the sample contains x defectives and $n - x$ non-defectives. The defectives can be chosen in $\binom{D}{x}$ ways, and for each way of choosing the defectives, the non-defectives can be chosen in $\binom{N-D}{n-x}$ ways. So the number of ways of choosing a sample which contains x defectives is $\binom{D}{x}\binom{N-D}{n-x}$.

There are $\binom{N}{n}$ ways of choosing a sample of n from a population of N, and each is equally likely since the selection is made at random.

The expression for the pmf follows.

Note that $p_X(x) > 0$ only if $0 \leq x \leq D$ and $0 \leq n - x \leq N - D$, since otherwise one or other of $\binom{D}{x}$ or $\binom{N-D}{n-x}$ is zero.

Therefore $p_X(x) > 0$ only if $A \leq x \leq B$, where $A = \max(0, n + D - N)$, and $B = \min(n, D)$.

Nevertheless, we usually denote the set of possible values S_X as $\{0 \leq x \leq n\}$ allowing that some of these values may actually have zero probability.

Clearly, $p_X(x) \geq 0$. It can be shown that $\sum p_X(x) = 1$, by equating coefficients of s^n on both sides of the identity $(1 + s)^D(1 + s)^{N-D} = (1 + s)^N$.

Hypergeometric Distribution

If X has pmf $p_X(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}$, $(x = A, \ldots, B)$, where A and B are defined above, then we say that X has a *hypergeometric distribution* with parameters n, D and N and we write $X \stackrel{d}{=} \mathrm{Hg}(n, D, N)$.

Example

If a hand of five cards is dealt from a well-shuffled pack of fifty-two cards, the number of spades in the hand, $X \stackrel{d}{=} \mathrm{Hg}(n = 5, D = 13, N = 52)$.

Hypergeometric mean and variance

It can be shown that

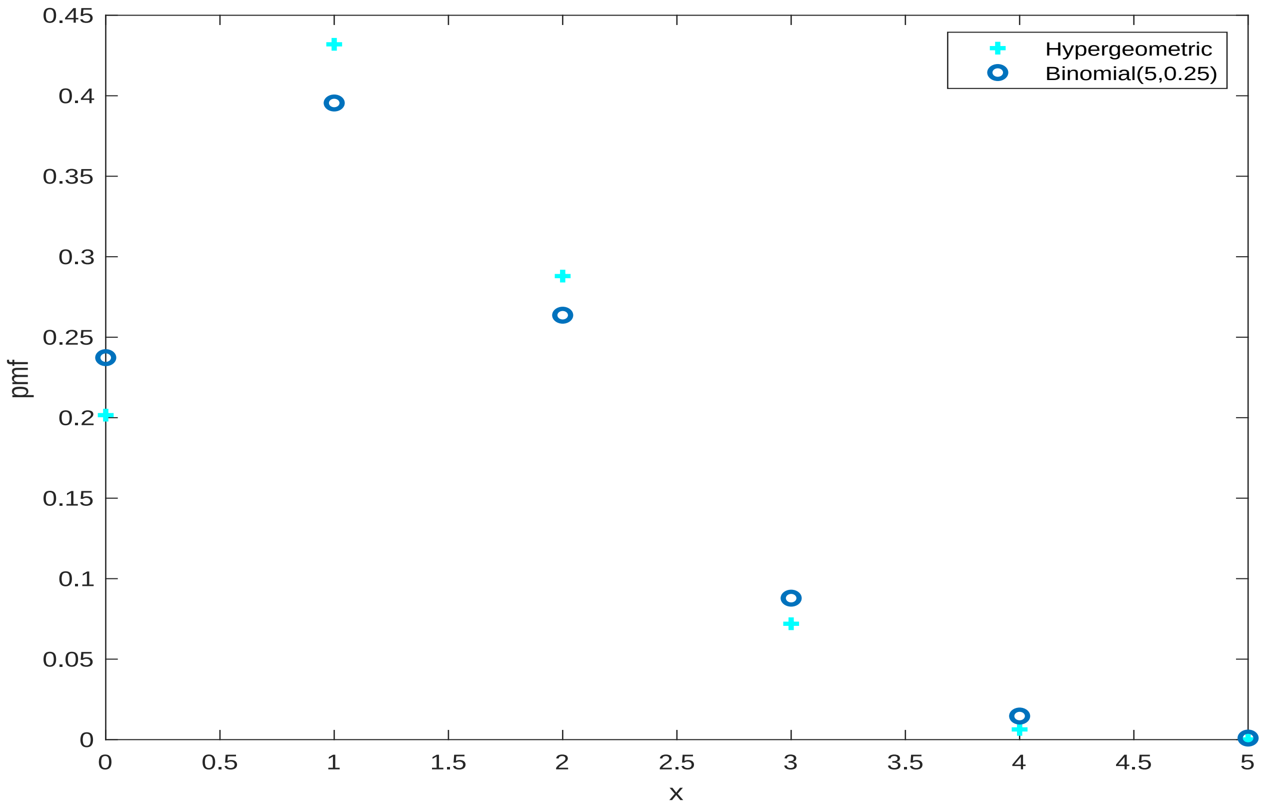
$$\begin{aligned}\mathbb{E}(X) &= \frac{nD}{N}.\\ V(X) &= \frac{nD(N-D)}{N^2} \times \left(1 - \frac{n-1}{N-1}\right).\end{aligned}$$

It is interesting to compare these formulae to those for the Binomial distribution with p replaced by the percentage of defectives $\frac{D}{N}$.

In particular, a hypergeometric distribution is approximated by the binomial distribution with parameters n and p if n is small compared to N, because there is not much difference between sampling with and without replacement

Example

Compute the hyper-geometric probabilities for $N = 24$, $n = 5$ and $p = 0.25$ and the approximating binomial probabilities and plot them



Poisson (1837) Random Variables (Ghahramani 5.2)

Recall that a Binomial random variable counts the total number of successes in a sequence of n independent Bernoulli trials. If we think of a "success" as an "event" then a Binomial random variable effectively counts events occurring in discrete time.

A Poisson random variable is an analogue of the Binomial random variable which effectively counts "events" occurring in continuous time. However both types count "events" so both are discrete random variables.

We derive the Poisson distribution via a limiting process involving sequences of Bernoulli trials as follows.

Assume that each Bernoulli trial takes a time $1/n$ to complete and that the probability of success in a Bernoulli trial is proportional to this time, say $\mathbb{P}(success) = \alpha/n$. Then, by time 1, we can complete n trials. Let N = number of "events" which occur by time 1. Then $N \stackrel{d}{=} \mathrm{Bi}(n, \frac{\alpha}{n})$ and hence

$$\mathbb{P}\big(N = k\big) = \tbinom{n}{k}\big(\tfrac{\alpha}{n}\big)^k\big(1 - \tfrac{\alpha}{n}\big)^{n-k},\ 0 \leq k \leq n.$$

Now we shrink the length of time for each trial and the success probability at the same rate, by letting $n \to \infty$.

It is a basic mathematical fact that

$$\lim_{n\to\infty}(1+\frac{x}{n})^n = e^x$$

and so

$$\lim_{n\to\infty}\left(1-\frac{\alpha}{n}\right)^n = e^{-\alpha}.$$

So we have for fixed $k \geq 0$,

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\big(N = k\big) &= \lim_{n\to\infty} \tbinom{n}{k}\left(\tfrac{\alpha}{n}\right)^k\left(1 - \tfrac{\alpha}{n}\right)^{n-k} \\
&= \lim_{n\to\infty} \frac{n!}{n^k(n-k)!}\,\frac{\alpha^k}{k!}\,\left(1 - \tfrac{\alpha}{n}\right)^n\left(1 - \tfrac{\alpha}{n}\right)^{-k} \\
&= 1 \times \frac{\alpha^k}{k!} \times e^{-\alpha} \times 1 \\
&= \frac{e^{-\alpha}\alpha^k}{k!}.
\end{aligned}
$$

If N has pmf $p_N(k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$ $(k = 0, 1, 2, \ldots)$, we say that N has a *Poisson distribution* with parameter λ, and we write $N \stackrel{d}{=} \mathrm{Pn}(\lambda)$.

So in this case $N \stackrel{d}{=} \mathrm{Pn}(\alpha)$.

Note that $p_N(k) \geq 0$, and that $\sum_{k=0}^{\infty} p_N(k) = 1$, since

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$$

This is the Taylor series expansion of e^{λ} and is another basic mathematical fact that is worth remembering.

Siméon Denis Poisson [21 June 1781 – 25 April 1840]

Use of Poisson Distributions

Poisson distributions might be a reasonable model for

- Radioactive emissions
- Arrival of calls at a call centre
- Times that people are diagnosed with a rare disease
- Number of earthquakes during a time period
- Number of crimes occurring in a specific area

Example

Assume cars pass an isolated petrol station on a country road at a constant mean rate of 5 per hour, or equivalently 2.5 per half hour. Let N denote the number of cars which pass the petrol station whilst it is temporarily closed for half an hour one Friday afternoon. What is the probability that the station missed out on three or more potential customers?

The average number of cars ('events') in half an hour is 2.5 so $N \stackrel{d}{=} \mathrm{Pn}(2.5)$.

$$
\begin{aligned}
\mathbb{P}(N \geq 3) &= 1 - \mathbb{P}(N \leq 2) \\
&= 1 - \left\{ \frac{e^{-2.5} 2.5^0}{0!} + \frac{e^{-2.5} 2.5^1}{1!} + \frac{e^{-2.5} 2.5^2}{2!} \right\} \\
&= 1 - \{0.0821 + 0.2052 + 0.2565\} \\
&= 0.4562
\end{aligned}
$$

Poisson mean and variance

Applying our formulae for expectations we can deduce that

$$\begin{aligned}
\mathbb{E}(X) &= \lambda, \\
\mathbb{E}(X(X-1)) &= \lambda^2, \\
V(X) &= \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2 \\
&= \lambda.
\end{aligned}$$

Note the interesting fact that the mean and variance of a Poisson random variable are equal.

Our original derivation of the Poisson distribution shows that it approximates the Binomial distribution under the right circumstances. As we shrunk the length of time for each trial and the success probability at the same rate, in the Binomial distribution for $N \stackrel{d}{=} \mathrm{Bi}(n, \frac{\alpha}{n})$ we had

$$\begin{aligned} \text{Number of trials:} &\quad n \to \infty \\ \text{Probability of success:} &\quad \tfrac{\alpha}{n} \to 0 \\ \text{Average number of ``events'':} &\quad (n)\tfrac{\alpha}{n} = \alpha. \end{aligned}$$

Poisson approximation to Binomial

If p is small, the Poisson distribution with $\lambda = np$ can be used as a convenient approximation to the Binomial distribution.

$\mathrm{Bi}(n, p) \approx \mathrm{Pn}(np)$ for p small.

A rough rule is that the approximation is satisfactory if $p \leq 0.05$.

Example

One in $10,000$ items from a production line is defective. The occurence of defects in successive items are independent. What is the probability that in a batch of $20,000$ items there will be at least 4 defective items?

Answer: 0.1429**.**

The Scope of Poisson Approximation

We can think of a binomial random variable $X \stackrel{d}{=} \text{Bi}(n, p)$ as the sum of n independent Bernoulli random variables with parameter p.

Poisson approximation can be used effectively in the following more general context of "sum of Bernoulli trials".

- The random variable X can be expressed as the sum of n Bernoulli random variables:

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

  where each $X_i \stackrel{d}{=} \text{Bernoulli}(p_i)$.

- n is very large, each p_i is very small, and the sum $p_1 + \cdots + p_n$ is of "normal scale".

- The dependence among these Bernoulli random variables $X_1, \cdots, X_n$ is weak: for each fixed i, the number of random variables that are dependent on X_i is much smaller than n.

If the above conditions are satisfied, then it is effective to use the Poisson approximation $X \overset{d}{\approx} \mathrm{Pn}(\lambda)$ with parameter $\lambda = p_1 + \cdots + p_n$.

Remark. The binomial case is a special situation: in this case all the X_i's are independent, $p_1 = \cdots p_n = p$, and $\lambda = np$.

Birthday Problem Revisited

Recall the Birthday Problem on Slide 38 to 41.

- There are n independent persons in a group.
- Each day of the year is equally likely to be the birthday of a person.

Question: What is the probability that "no two people can have the same birthday"?

In Slide 40, we approximate this probability as

$$\mathbb{P}(\text{no two people have same B-day}) \approx \exp\big(-\frac{n(n-1)}{720}\big).$$

We can recover this result using Poisson approximation. But we need to understand what are the underlying Bernoulli trials and success probabilities.

- Label the n people by $1, 2, \cdots, n$.
- We think of each "Bernoulli trial" as labelled by a pair of indices $i < j$ representing the pair of Person i and Person j.
- We define "success" for this trial (labelled by (i, j)) to be the event that Person i and Person j have the same birthday. The probability of this event (prob. of success) is $\frac{1}{365}$.
- There are $\binom{n}{2} = \frac{n(n-1)}{2}$ trials in total (number of pairs among the n persons.

The dependence among these trials is weak:

- For each given trial (i, j), it is independent of all those trials (k, l) where $\{k, l\}$ are different from $\{i, j\}$ (because the Persons k, l are independent of the Persons i, j in this case).
- There are $\binom{n-2}{2}$ trials that are independent of the trial (i, j).
- There are at most

$$\binom{n}{2} - \binom{n-2}{2} = 2n - 3$$

  trials that might have dependence on the trial (i, j). This number is much smaller than the total number $\binom{n}{2}$ of trials if n is large.

Let X be the total number of successes (namely, the number of pairs who have the same birthday).

We can use Poisson approximation $X \stackrel{d}{\approx} \mathrm{Pn}(\binom{n}{2} \cdot \frac{1}{365})$:

$$\mathbb{P}(\text{no two people have same birthday day})$$
$$= \mathbb{P}(X = 0) \approx \exp\big(-\binom{n}{2} \cdot \frac{1}{365}\big) = \exp\big(-\frac{n(n-1)}{720}\big).$$

The same idea can be used to approximate the probability that "no three people have the same birthday". Simple counting in this case becomes almost impossible. We use Poisson approximation.

- Each Bernoulli trial is labelled by a triple of indices $i < j < k$ (representing Persons i, j, k). "Success" means these three persons have the same birthday. Probability of success is $\frac{1}{365^2}$.
- There are a total of $\binom{n}{3}$ Bernoulli trials.
- Dependence among these trials is weak.

Let X be the number of successes (i.e. number of triples who have the same birthday). Then $X \overset{d}{\approx} \mathrm{Pn}\left(\binom{n}{3} \cdot \frac{1}{365^2}\right)$:

$$\mathbb{P}(\text{no three people have same B-day})$$
$$= \mathbb{P}(X = 0) \approx \exp\left(-\binom{n}{3} \cdot \frac{1}{365^2}\right).$$

When $n = 83$, this probability is approximately 0.5 – in a group of 83 people, with 50% chance there are at least three people having the same birthday.

Discrete uniform random variables

Consider the discrete random variable X having *pmf*

$$p_X(x) = \frac{1}{n-m+1} \qquad (x = m, m+1, \ldots, n)$$

where m and n are integers such that $m \leq n$. We say that X has a *discrete uniform distribution* on $m \leq x \leq n$, and we write $X \stackrel{d}{=} \mathrm{U}(m, n)$.

If X denotes the result of throwing a fair die, then $X \stackrel{d}{=} \mathrm{U}(1, 6)$.

If Y denotes the result of spinning a roulette wheel (with one zero), then $Y \stackrel{d}{=} \mathrm{U}(0, 36)$.

Discrete uniform mean and variance

Applying our formulae for expectations we can deduce that for $X \stackrel{d}{=} U(0, n)$

$$\begin{aligned}
\mathbb{E}(X) &= \frac{n}{2}, \\
\mathbb{E}(X^2) &= \frac{1}{6}n(2n+1), \\
V(X) &= \frac{1}{12}n(n+2).
\end{aligned}$$

Example

Consider a sequence of independent Bernoulli trials with probability of success p. We are given the additional information that in the first n trials there is exactly one success. Let X denote the number of the (random) trial at which this single success occurred. What is the pmf of X?