5. Discrete Random Variables and Probability Distributions

5.1 Random Variables

So far, considered sample spaces

$$\Omega \equiv \{\omega_1, \omega_2, \ldots\}$$

and events

$$E \equiv \{\omega_i \in E\} \subseteq \Omega.$$

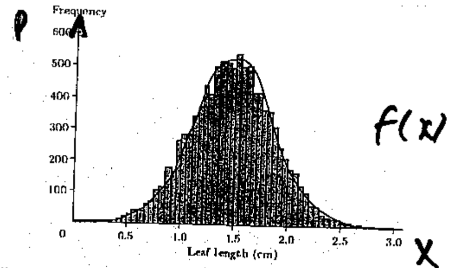


Figure 2.11 A smooth curve fitted to a histogram

$f(x) = -x$

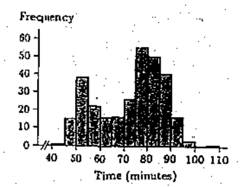
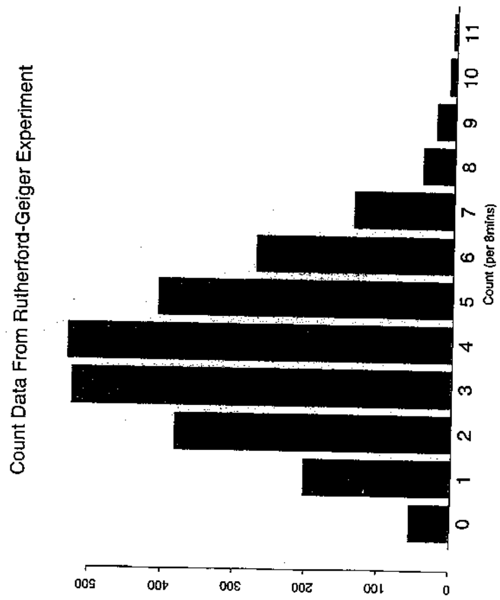


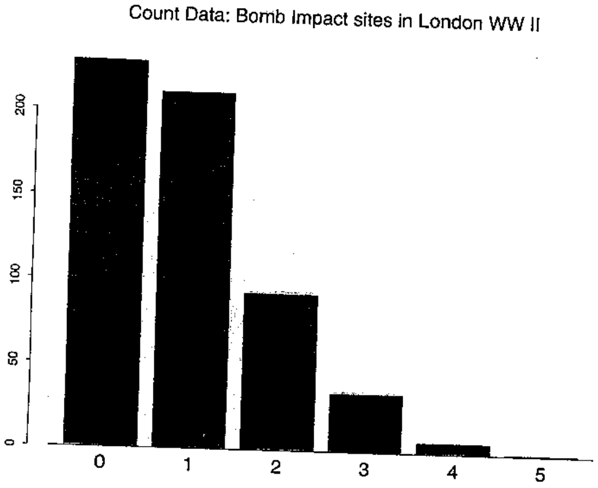
Figure 3.18 Waiting times between eruptions, Old Faithful geyser

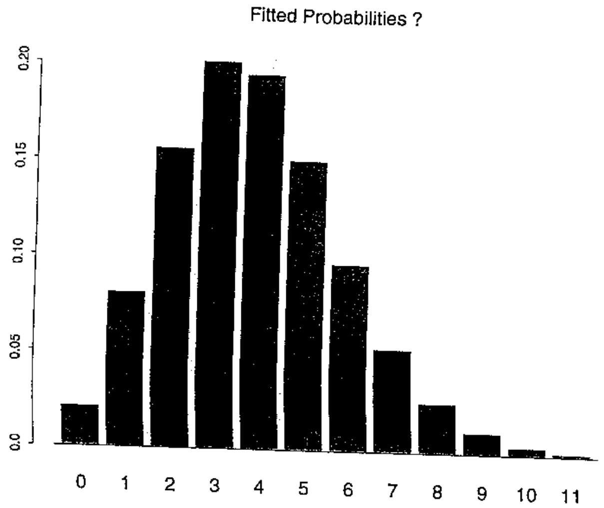
Rutherford, E. and Geiger, H. (1910) The probability variation in the distribution of alpha particles. *Philosophical Magazine* Sixth Series, **20**, 698–704.

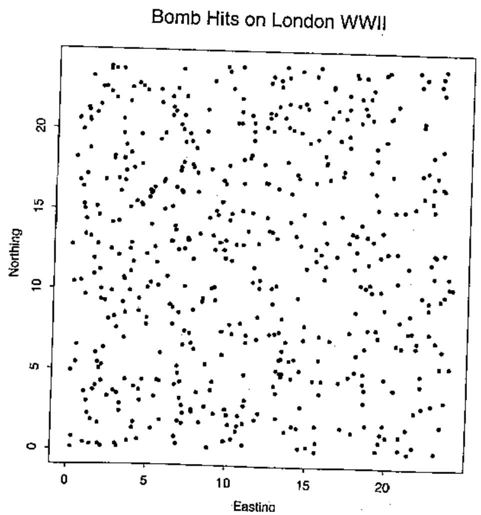
Table 4.4 Emissions of alpha particles

Count	Frequency	Fit
0	57	54
1	203	210
2	383	407
3	525	525
4	532	509
5	408	395
6	273	255
7	139	141
8	49	68
9	27	30
10	10	11
11	4	4
12	2	1
> 12	0	1









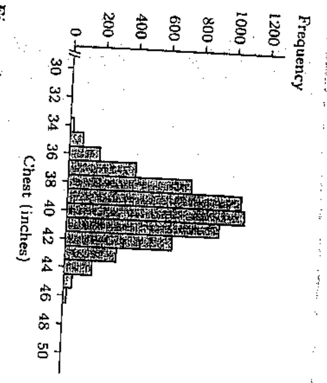
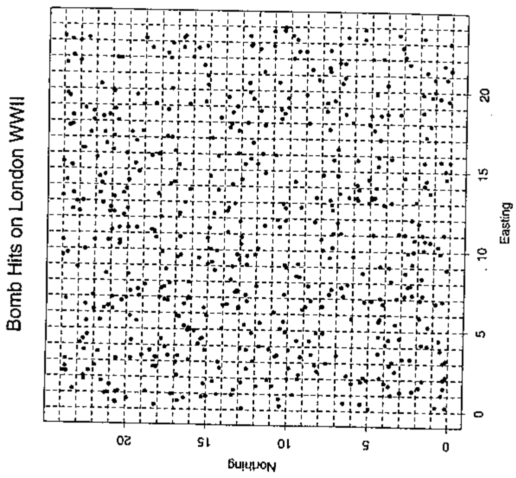
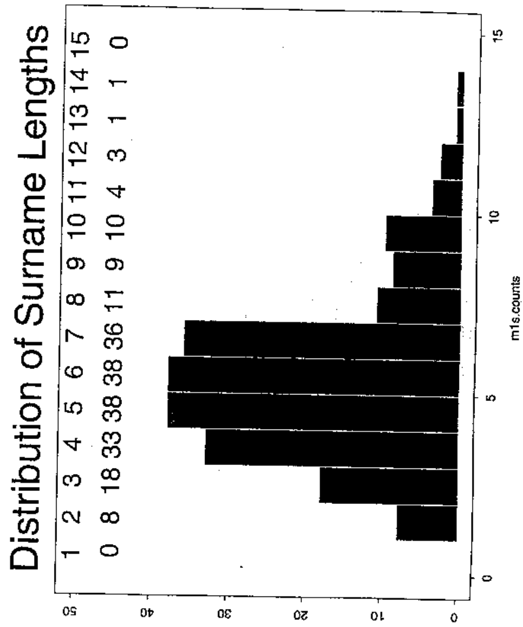


Figure 1.18 Chest measurements (inches)





COURSEWORK HINT

Hypergeometric Formula

Given N, R, n

formula calculates probability for different values of r

$$\frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}$$

That is: if "Pop^N contains R" is T_R and "sample contains r" is S_r

Formula gives

$$P(S_r \mid T_R)$$

but, if you observe S_r, you may wish to calculate

$$P(T_R \mid S_r)$$

However, the events

$T_0, T_1, \dots T_N$

Partition Ω

(and hence imply a partition of S_r for each r ....)

DEADLINE EXTENSION:

NOON, MONDAY 19th

(in the lecture ...)

i.e. X assigns some real number x to sample outcome $\omega \in \Omega$.

EXAMPLE Fair coin tossed three times

SAMPLE OUTCOMES

HHH	ω_1
HHT	ω_2
HTH	ω_3
$\vdots$	
TTT	ω_8

Could consider events

$E \equiv$ "3H in total"

$F \equiv$ "2H in total"

To simplify notation, we now seek a way of referring to general experiments via a "universal" (common) sample space

(NOT SPECIFIC TO ANY SINGLE EXPERIMENT)

We introduce a mapping from Ω to $\mathbb{R}$

i.e define

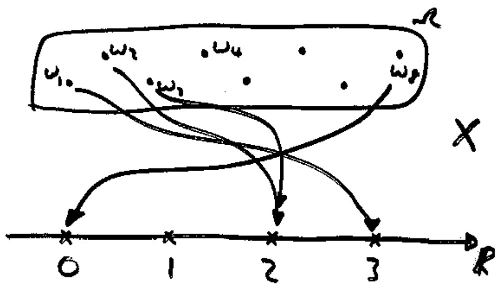
$$X : \Omega \longrightarrow \mathbb{R}$$
$$\omega \longmapsto x$$

so that $X(\omega) = x$

Define mapping $X : \Omega \longrightarrow \mathbb{R}$ by

$$X(\omega) = \# H \ \omega \text{ contains}$$

i.e
$$X(\omega_1) = 3$$
$$X(\omega_2) = 2$$
$$\vdots$$
$$X(\omega_8) = 0$$



Or could define mapping $X : \Omega \longrightarrow \mathbb{R}$
where

$$X(\omega) = \#\ \text{"runs" in } \omega$$

("run" - consecutive outcomes of the same type)

then

$$X(\omega_1) = 1$$
$$X(\omega_2) = 2$$
$$\vdots \qquad \vdots$$
$$X(\omega_8) = 1$$

X is termed a "random variable" (or "random quantity")

because there is UNCERTAINTY as to which value X will take before the experiment is carried out.

(but we can work out a list of possible values for X).

MAJOR ADVANTAGE

As $X(\omega) \in \mathbb{R}$, we can consider the "sample space" for X, which we know by construction must be a subset of $\mathbb{R}$; this will be true for all experiments.

Typically random variables will be associated with numerical summaries

e.g "# Heads is 3" $\equiv$ "$X = 3$"

"Count is 117" $\equiv$ "$X = 117$"

etc

Also interested in collections of sample outcomes

"# H is ≤ 3" $\equiv$ "# H is 0" $\cup \ldots \cup$ "# H is 3"

$\equiv$ "$X \leq 3$"

Note on notation

X (upper case) - Random variable (i.e. map)

x (lower case) - any real value.

- Do not confuse the two!

If $\mathbb{X}$ is countable, the random variable X is termed a

Discrete Random Variable

For event $E \subseteq \Omega$, the image is

$$\mathbb{X}_E \equiv \{x : X(\omega) = x \text{ for some } \omega \in E\}$$

and, by construction we must have

$$P(E) \equiv P[X \in \mathbb{X}_E]$$

event in Ω

event in $\mathbb{R}$

Definition

For random variable X defined on sample space Ω, the range of X, denoted $\mathbb{X}$, is the set of real numbers

$$\{x : X(\omega) = x, \text{ some } \omega \in \Omega\}$$

(the "image" of Ω under X ?)

Typically, $\Omega \equiv \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$

$\therefore \mathbb{X} \equiv \{x_1, x_2, \ldots, x_n, \ldots\}$

i.e. both sets are Countable lists

$\therefore$ We can consider probabilities

$$P[X \in \mathbb{X}_E]$$

directly. Specifically, we consider only probabilities such as

$$P[X = x]$$

or

$$P[X \leq x]$$

for real values x

To complete the probability specification, we seek the formulae which specify these kinds of probabilities – that is, we seek a function f that gives

$$f(x) = P[X = x]$$

for values of $x \in \mathbb{X}$

5.2 THE PROBABILITY MASS FUNCTION

DEFINITION

For discrete random variable X with range $\mathbb{X} = \{x_1, x_2, \ldots\}$, the PROBABILITY MASS FUNCTION of X is denoted f_X, and is defined by

$$f_X(x) = P[X = x]$$

for all $x \in \mathbb{R}$, but where

$$f_X(x) \equiv 0$$

if $x \notin \mathbb{X}$

EXAMPLE Fair coin tossed n times

$\Omega \equiv \{$"All sequences of results that are possible"$\}$.

$|\Omega| = 2^n$ (MULTIPLICATION PRINCIPLE)

$\mathcal{E} \Rightarrow$ "ALL SEQUENCES EQUALLY LIKELY"

(H/T EQUALLY LIKELY ON ANY TOSS + MUTUAL INDEPENDENCE)

$\therefore$ For any sequence ω_i

$$P(\{\omega_i\}) = \frac{1}{2^n}$$

Can represent the sequences as

HTHHH $\cdots$ TH

10111 $\cdots$ 01

$\therefore$ a binary sequence representation may be useful in computations.

Define discrete random variable (r.v.) X by

$$X(\omega) = \#\,H_s \ \omega \text{ contains.}$$

Now X is a "many-to-one" map

(many sequences map to the same real number)

e.g $n = 4$

ω_1	HTHH	$X(\omega_1) = 3$
ω_2	HHHT	$X(\omega_2) = 3$

How many sequences are there like this

$$\binom{n}{x}$$

(i.e. select x positions from n without replacement, unordered)

$$\therefore \quad P[X = x] = \binom{n}{x}\left(\frac{1}{2}\right)^n$$

(NOTE: Let E_x be the event in Ω containing all sequences with x H. Then

$E_0, E_1, \ldots E_n$ PARTITION Ω

The range of X is

$$\mathbb{X} \equiv \{0, 1, 2, \ldots, n\}$$

and we now seek

$$f_X(x) = P[X = x]$$

Now, the event "$X = x$" is equivalent to the collection sequences in Ω that contain x H out of n.

Also E_x contains $\binom{n}{x}$ equally likely ω_i)

$\therefore$ A full specification in this experiment is

$$f_X(x) = \binom{n}{x}\left(\frac{1}{2}\right)^n \quad x \in \mathbb{X}$$

and

$$f_X(x) = 0 \qquad x \notin \mathbb{X}$$

<u>NOTE</u> Could define f_X <u>pointwise</u> (rather than using a formula)

e.g.

$$f_X(x) = \begin{cases} \left(\frac{1}{2}\right)^n & x = 0 \\ n\left(\frac{1}{2}\right)^n & x = 1 \\ \frac{n(n-1)}{2}\left(\frac{1}{2}\right)^n & x = 2 \\ \vdots & \end{cases}$$

- define $f_X(x)$ for all points at which it is non-zero ?

<u>NOTE</u>

$E_0, E_1, \ldots, E_n$ partition Ω

$\Rightarrow$ "$X=0$", "$X=1$", ..., "$X=n$" are events that <u>partition</u> $\mathbb{X}$

and $\sum_{x=0}^{n} P(E_x) = 1$

$$\Longrightarrow \sum_{x=0}^{n} P[X=x] = 1$$

The "duality" between the partition of Ω and the partition of $\mathbb{X}$ implies the following <u>important conditions</u>.

We <u>must</u> have that the probability mass function (p.m.f.) satisfies

(i) $f_X(x) \geq 0$

(ii) $\sum_{x \in \mathbb{X}} f_X(x) = 1$

(so that, in fact, $0 \leq f_X(x) \leq 1$)

The p.m.f. f_X describes the <u>PROBABILITY DISTRIBUTION</u> of X

QUANTITATIVELY — NUMERICAL

QUALITATIVELY — "SHAPE" OF DISTRIBUTION

The precise form of f_X depends on experimental context.

⌈But, always remember, the p.m.f. is merely a real-function of a single variable

just like functions

$$f(x) = \cos x$$
$$f(x) = x^2 + 3x - 2$$
$$f(x) = e^x$$

that happens to obey some specific conditions/rules. ⌋

If $\mathbb{X}$ is finite, we just need to ensure

$$0 \leq f(x)$$

for $x \in \mathbb{X}$, because

$$\sum_{x \in \mathbb{X}} f(x) = c > 0$$

where c is a finite constant

$\therefore$ can define $f_X(x) = \frac{1}{c} f(x)$

CONSTRUCTING f_X.

Given that we require

(i) $f_X(x) \geqslant 0 \quad x \in \mathbb{X}$

(ii) $\displaystyle\sum_{x \in \mathbb{X}} f_X(x) = 1$

What general guidelines do we have for constructing a pmf ?

Consider a function $f(\cdot)$ defined on $\mathbb{X}$.....

as then clearly

(i) $f_X(x) \geqslant 0$

(ii) $\displaystyle\sum_{x \in \mathbb{X}} f_X(x) = 1$

by construction.

This method is a good general way of constructing pmfs

as f_X has the same shape as f, but is re-scaled to sum to 1.

If $\mathbb{X}$ is countable but <u>infinite</u>

not quite so straightforward . . . .

Need that

$$\text{(i)} \quad f(x) \geqslant 0$$

but <u>also</u> that

$$\text{(ii)} \quad \sum_{x \in \mathbb{X}} f(x) = c$$

where c is some <u>finite</u> constant

i.e. We require that the sequence $\{p_x\}$ defined by $p_x = f(x)$ are terms in a <u>convergent</u> series.

Then we may again define

$$f_X(x) = \frac{1}{c} f(x)$$

$\therefore$ We will derive pmf's from "famous" convergent series.

<u>EXAMPLE</u> For $\mathbb{X} \equiv \{1, 2, 3, \ldots\}$

consider

$$f(x) = \theta^x$$

for some constant θ, $0 \leq \theta \leq 1$

- i.e. a GEOMETRIC PROGRESSION

$$\sum_{x \in \mathbb{X}} f(x) = \theta + \theta^2 + \theta^3 + \cdots$$

$$= \frac{\theta}{1-\theta} = c$$

$\therefore$ Define $f_X(x) = \frac{1}{c} f(x) = (1-\theta)\theta^{x-1}$

<u>EXAMPLE</u> For $\mathbb{X} \equiv \{1, 2, 3, \ldots\}$

$$f(x) = \frac{1}{x^2}$$

- a famously convergent series . . .

$$\sum_{x \in \mathbb{X}} f(x) = 1 + \frac{1}{4} + \frac{1}{9} + \cdots$$

$$= \frac{\pi^2}{6} = c \text{, say.}$$

$\therefore$ Define

$$f_X(x) = \frac{1}{c} f(x) = \frac{6}{\pi^2 x^2}$$

EXAMPLE For $\mathbb{X} = \{0, 1, 2, \ldots\}$

$$f(x) = \frac{\lambda^x}{x!}$$

for some $\lambda > 0$.

$$\sum_{x \in \mathbb{X}} f(x) = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots$$

$$= e^{\lambda} \quad = c \text{ say}$$

$$\therefore f_X(x) = \frac{1}{c} f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

EXAMPLE Ex 5 1.

(a) X - # pupils with birthday on New Years Day.

$$\mathbb{X} = \{0, 1, 2, 3, \ldots, 500\}$$

Recall the occupancy problem

"allocate r items to n boxes; what is the probability that box 1 contains precisely k items?"

Here $r = 500$

$n = 365$

Probability was

$$\binom{r}{k}\left(\frac{1}{n}\right)^k\left(1 - \frac{1}{n}\right)^{r-k} \quad 0 \le k \le r.$$

$\therefore$ $P[X = x]$ is given by

$$f_X(x) = \binom{500}{x}\left(\frac{1}{365}\right)^x\left(1 - \frac{1}{365}\right)^{500-x}$$

$$= \frac{500!}{x!(500-x)!}\left(\frac{1}{365}\right)^x\left(1 - \frac{1}{365}\right)^{500-x}$$

$$= \frac{1}{x!}\frac{500!}{(500-x)!}\left(\frac{1/365}{1 - 1/365}\right)^x \times$$

$$\left(1 - \frac{500}{365 \times 500}\right)^{500}$$

Now $\dfrac{500!}{(500-x)!} \approx (500)^x$

$$\left(\frac{1/365}{1 - 1/365}\right)^x \approx \left(\frac{1}{365}\right)^x$$

$$\left(1 - \frac{500}{365 \times 500}\right)^{500} \approx e^{-500/365}$$

$\therefore$

$$f_X(x) \simeq \frac{\lambda^x e^{-\lambda}}{x!} \qquad \lambda = \frac{500}{365}$$

- see solutions for full details
- we will formalize this proof later .....

Numerical values for exact and approximate probabilities do equate.

$\therefore P[X = x]$ is given by

$$\underbrace{\binom{x-1}{2}\left(\tfrac{1}{2}\right)^2\left(\tfrac{1}{2}\right)^{x-3}}_{\substack{x-1 \text{ tosses containing} \\ \text{precisely } 2\ H.}} \times \underbrace{\tfrac{1}{2}}_{H.}$$

$\therefore$

$$f_X(x) = \binom{x-1}{2}\left(\tfrac{1}{2}\right)^3\left(\tfrac{1}{2}\right)^{x-3}$$

$$x = 3, 4, 5, \ldots.$$

EXAMPLE Ex 5, 6.

X - # tosses required to obtain 3 Heads in a sequence of independent tosses.

$\mathbb{X} = \{3, 4, 5, 6, \ldots\}$

Now, "$X = x$" corresponds to

A sequence of $x-1$ tosses containing two H somewhere in the sequence

then

a H to finish the sequence.

The pmf f_X for a discrete r.v. X describes the probability distribution; we now consider two useful "functionals" related to f_X

PROBABILITY GENERATING FUNCTION

EXPECTATION.

DEFINITION

For sequence of real numbers $\{a_i \,;\, i \geq 0\}$, the function

$$g(t) = \sum_{i=0}^{\infty} a_i t^i$$

is the **GENERATING FUNCTION** of $\{a_i\}$.

(whenever the sum is convergent)

e.g. if $a_i = \binom{n}{i}$ $i=0,1,2,\ldots n$

and zero otherwise

$$\begin{aligned} g(t) &= \sum_{i=0}^{\infty} a_i t^i \\ &= \sum_{i=0}^{n} \binom{n}{i} t^i \\ &= (1+t)^n \\ &= 1 + nt + \frac{n(n-1)}{2} t^2 \ldots \end{aligned}$$

DEFINITION

For discrete random variable X with range $\mathbb{X} = \{x_1, x_2, \ldots\}$ and p.m.f. f_X, the **PROBABILITY GENERATING FUNCTION** (PGF), G_X, is defined by

$$G_X(t) = \sum_{x_i \in \mathbb{X}} f_X(x_i)\, t^{x_i}$$

(i.e. the generating function of the probabilities $p_i = P[X = x_i]$)

Typically, $\mathbb{X} = \{0, 1, 2, \ldots\}$ so then

$$G_X(t) = \sum_{x=0}^{\infty} f_X(x)\, t^x$$

- definition holds whenever the sum is convergent.

USEFUL PROPERTY : UNIQUENESS

- there is a 1-1 correspondence between G_X and f_X.

$\therefore$ knowledge of $G_X \iff$ knowledge of f_X

PROBABILITY MASS FUNCTION CALCULATIONS USING MAPLE

```
> n:=20;
  for x from 0 by 1 while x < n+1 do
     f_X[x] := binomial(n,x)*(1/2)^n;
  od;
```

$\binom{n}{x}\left(\frac{1}{2}\right)^n$

Numerical values of Probabilities

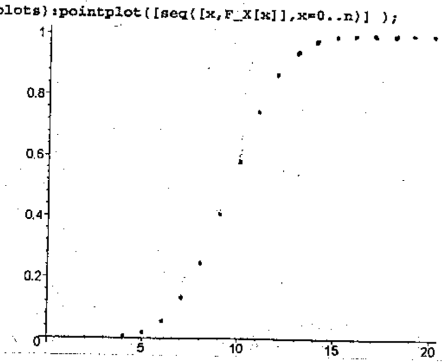
```
> seq(binomial(n,x)*(1/2)^n,x=0..20);
```

$$\frac{1}{1048576}, \frac{5}{262144}, \frac{95}{524288}, \frac{285}{262144}, \frac{4845}{1048576}, \frac{969}{65536}, \frac{4845}{131072}, \frac{4845}{65536}, \frac{62985}{524288}, \frac{20995}{131072}, \frac{46189}{262144},$$
$$\frac{20995}{131072}, \frac{62985}{524288}, \frac{4845}{65536}, \frac{4845}{131072}, \frac{969}{65536}, \frac{4845}{1048576}, \frac{285}{262144}, \frac{95}{524288}, \frac{5}{262144}, \frac{1}{1048576}$$

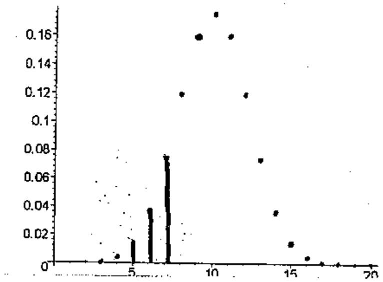
CUMULATIVE PROBABILITY

```
> F_X[0]:=f_X[0];
> for x from 1 by 1 while x < n+1 do
     F_X[x] := F_X[x-1] + f_X[x];
  od;
```

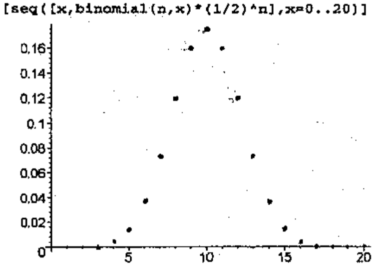
```
> with(plots):pointplot([seq([x,F_X[x]],x=0..n)] );
```

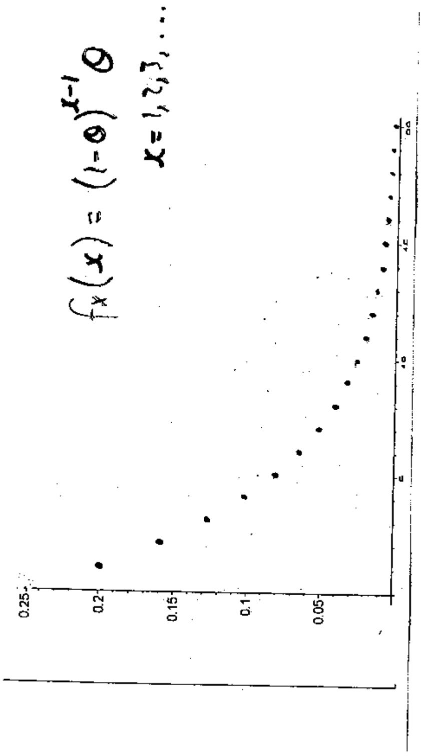


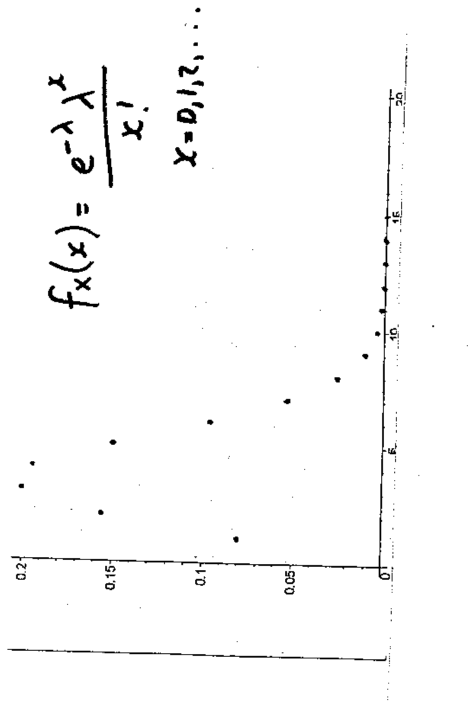
```
> with(plots):pointplot([seq([x,f_X[x]],x=0..n)] );
```

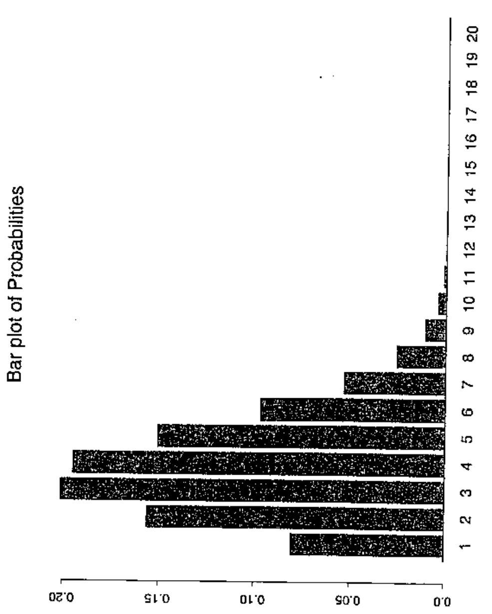


```
> pointplot([seq([x,binomial(n,x)*(1/2)^n],x=0..20)] );
```









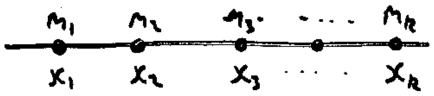
DEFINITION

For discrete r.v. X with range $\mathbb{X} = \{x_1, x_2, \ldots\}$ and pmf f_X, the **EXPECTATION** of X is defined by

$$E_{f_X}[X] = \sum_{x_i \in \mathbb{X}} x_i f_X(x_i)$$

Analogy:

Imagine masses $m_1, m_2, \ldots, m_k$
at locations $x_1, x_2, \ldots, x_k$



then the "centre of mass" of the system is at

$$\frac{\sum_{i=1}^{k} x_i m_i}{\sum_{i=1}^{k} m_i}$$

For probability interpretation, let

$$m_i = P[X = x_i] = f_X(x_i)$$

$\therefore$ "centre of probability mass" is at

$$\sum_{i=1}^{k} x_i f_X(x_i) \qquad \left(\sum_{i=1}^{k} m_i = 1\right)$$

which is the **EXPECTATION** for a finite range $\mathbb{X}$..

The expectation of a probability distribution/random variable is a measure of "location"

i.e. $E_{f_X}[X]$ gives an indication of where (on the real line) the bulk of the probability mass is located.

NOTE $E_{f_X}[X]$ is always a real constant, and not a function of x

EXAMPLE $\mathbb{X} \equiv \{0, 1, 2, \ldots\}$

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

$$E_{f_X}[X] = \sum_{x=0}^{\infty} x f_X(x)$$

$$= \sum_{x=0}^{\infty} x . \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \lambda e^{-\lambda} . e^{\lambda} = \underline{\underline{\lambda}}$$

IMPORTANT NOTE

$E_{f_X}[X]$ need not be finite!

e.g. $f_X(x) = \dfrac{6}{\pi^2 x^2}$ $\quad x = 1, 2, 3, \ldots$

Then

$$E_{f_X}[X] = \sum_{x=1}^{\infty} x \cdot f_X(x)$$

$$= \sum_{x=1}^{\infty} x \cdot \frac{6}{\pi^2 x^2} = \frac{6}{\pi^2} \sum_{x=1}^{\infty} \frac{1}{x}$$

$$= \infty$$

We will return to, study, and extend these definitions of

PGF

and

EXPECTATION

later in the course.

5.3 THE CUMULATIVE DISTRIBUTION FUNCTION

DEFINITION

For discrete r.v. X with range $\mathbb{X} \equiv \{x_1, x_2, \ldots\}$ so that

$$x_1 < x_2 < \ldots$$

the CUMULATIVE DISTRIBUTION FUNCTION (cdf) of X is denoted F_X and defined by

$$F_X(x) = P[X \leq x]$$

for $x \in \mathbb{R}$

i.e. $F_X(x_1) = P[X = x_1] = f_X(x_1)$

$$F_X(x_2) = P[X = x_1] + P[X = x_2]$$

$$= f_X(x_1) + f_X(x_2)$$

etc, so that

$$F_X(x_i) = \sum_{j=1}^{i} P[X = x_j]$$

$$= \sum_{j=1}^{i} f_X(x_j)$$

For $x \notin \mathbb{X}$

$$P[X \leq x] \equiv P[X \leq x_k]$$

say, where x_k is the largest element of $\mathbb{X}$ such that

$$x_k \leq x$$

$\therefore$ can define F_X for all real values of x.

F_X is the "cumulative" probability function which sums successive terms in the pmf f_X

F_X is completely specified by f_X (and vice-versa).

$$f_X(x_i) = F_X(x_i) - F_X(x_{i-1})$$

How does F_X behave?

PROPERTIES OF F_X

(i) F_X is a STEP FUNCTION with "steps" of magnitude $f_X(x_1), f_X(x_2), \ldots$ at positions $x_1, x_2, \ldots$

(ii) F_X is NON-DECREASING

– as we are successively adding on probabilities, if $a < b$ we must have

$$F_X(a) \leq F_X(b)$$

but between "step" locations there is no increase in F_X

i.e between x_i and x_{i+1} there is zero probability content

$\therefore$

$$F_X(x_{i+1} - \delta) = F_X(x_i)$$

for small $\delta > 0$.

(iii) If $x < x_1$

$$F_X(x) = P[X \leq x] = 0$$

nd in the limit as $x \longrightarrow \infty$

$$\lim_{x \to \infty} F_X(x) = F_X(\infty) = 1$$

$$\text{" } P[X \leq \infty] = 1. \text{ "}$$

In summary: F_X must have the following properties

(i) F_X is NON-DECREASING

(ii) F_X is CONTINUOUS FROM THE RIGHT

(iii) $\lim_{x \to -\infty} F_X(x) = 0$

(iv) $\lim_{x \to \infty} F_X(x) = 1$

It is not usually possible to compute a "closed-form" formula for F_X

- usually we will concentrate on mass function f_X, and only compute F_X numerically (using MAPLE etc).

EXAMPLE $\mathbb{X} = \{1, 2, 3, \ldots\}$

$$f_X(x) = (1-\theta)^{x-1}\theta \quad x = 1, 2, 3, \ldots$$

For $x \in \mathbb{X}$, we have

$$\begin{aligned} F_X(x) &= P[X \leq x] \\ &= \sum_{t=1}^{x} P[X = t] \\ &= P[X=1] + P[X=2] + \ldots \\ &\quad + P[X = x] \end{aligned}$$

$$= \theta + (1-\theta)\theta + (1-\theta)^2\theta + \dots$$
$$\dots + (1-\theta)^{x-1}\theta$$

$$= \theta\left[\frac{1-(1-\theta)^x}{1-(1-\theta)}\right] \quad \text{G.P.}$$

$$= 1-(1-\theta)^x \qquad x = 1,2,3,\dots$$

$$\therefore \; F_X(x) = 1-(1-\theta)^x \qquad x = 1,2,3,\dots$$

NOTE

eg $F_X(2.2) = P[X \le 2.2]$

$$= P[X=1] + P[X=2] = \underline{\underline{F_X(2)}}$$

We have now studied general properties of the two methods of specifying discrete probability distributions:

Key elements
(i) X
(ii) $\mathbb{X}$
(iii) f_X and F_X

We will now study specific "<u>named</u>" distributions corresponding to particular experimental conditions.

5.4 <u>THE BERNOULLI DISTRIBUTION</u>

Suppose $\Omega \equiv \{\omega_1, \omega_2\}$

+ define
$$X(\omega_1) = 0, \quad X(\omega_2) = 1 \Rightarrow \mathbb{X} \equiv \{0,1\}$$

Suppose
$$P[X=0] = P(\{\omega_1\}) = 1-\theta$$
$$P[X=1] = P(\{\omega_2\}) = \theta$$

for some $0 \le \theta \le 1$

Then

$$f_X(x) = \begin{cases} 1-\theta & x=0 \\ \theta & x=1 \end{cases}$$

$$= \theta^x(1-\theta)^{1-x} \qquad x \in \{0,1\}$$

and zero otherwise.

$$F_X(x) = \begin{cases} 0 & x<0 \\ 1-\theta & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

This is the <u>BERNOULLI DISTRIBUTION</u>

$$X \sim \text{Bernoulli}(\theta).$$

NOTE

PGF (PROBABILITY GENERATING FUNC.)

$$G_X(t) = \sum_{x \in \mathbb{X}} f_X(x)\, t^x$$

$$= \sum_{x=0}^{1} \theta^x (1-\theta)^{1-x}\, t^x$$

$$= (1-\theta) + \theta t$$

EXPECTATION

$$E_{f_X}[X] = \sum_{x=0}^{1} x . f_X(x) = \theta.$$

This simple expermental situation and probability specification is the basis for many of the more complicated "named" distributions.

We consider mutually independent repeats of a Bernoulli experiment, and vary the numerical summary / experimental conditions to define a variety of distributions.

5.5 THE BINOMIAL DISTRIBUTION

Consider a sequence of n Bernoulli experiments that are independent and identical in implementation.

Define r.v. X by

$$X = \text{"\# times } \omega_2 \text{ occurs"}$$

- total number of HEADS
  1s
  SUCCESSES.

Then

$$\mathbb{X} \equiv \{0, 1, 2, \ldots n\}$$

and if $P(\{\omega_2\}) = \theta$ for each repeat then

$$f_X(x) = P[X = x] = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

for $0 \leq x \leq n$, and zero otherwise.

as there are $\binom{n}{x}$ sequences containing precisely x ω_2s, and each has probability $\theta^x (1-\theta)^{n-x}$

[USING THE PARTITION, AND CHAIN RULE]

This is the pmf of the

BINOMIAL DISTRIBUTION

NOTE

For the binomial distribution

$$F_X(x) = \sum_{t=0}^{\text{floor}(x)} f_X(t)$$

is not available in "closed form"

Alternative interpretation

Sampling from a POPULATION comprising

Proportion	θ	TYPE I
	$1-\theta$	TYPE II

either FINITE POP^N WITH REPLACEMENT

or INFINITE POP^N WITHOUT REPLACEMEN

Alternative construction

Let X_i refer to the i^{th} trial

then $X_i \sim \text{Bernoulli}(\theta)$

i.e. $X_i = \begin{cases} 0 & \text{w.p } 1-\theta \\ 1 & \text{w.p } \theta. \end{cases} \quad \forall i=1,\dots n$

Then, by construction, we have

$$X = \sum_{i=1}^{n} X_i$$

where $X_1, X_2, \dots X_n$ are independent and identically distributed **i.i.d**

$$X \sim \text{Binomial}(n, \theta)$$

PGF

$$G_X(t) = \sum_{x=0}^{n} f_X(x)\, t^x$$

$$= \sum_{x=0}^{n} \binom{n}{x} \theta^x (1-\theta)^{n-x} t^x$$

$$= (1-\theta+\theta t)^n$$

- combine θ^x and t^x as $(\theta t)^x$
- note binomial expansion form.

EXPECTATION

$$E_{f_X}[X] = \sum_{x=0}^{n} x\, f_X(x)$$

$$= \sum_{x=1}^{n} x . \frac{n!}{x!(n-x)!} \theta^x (1-\theta)^{n-x}$$

$$= n\theta \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} \theta^{x-1}(1-\theta)^{n-x}$$

$$= n\theta \sum_{y=0}^{n-1} \binom{n-1}{y} \theta^y (1-\theta)^{n-y-1} \qquad y = x-1$$

$$= n\theta (1-\theta+\theta)^{n-1}$$

$$= \underline{\underline{n\theta}}$$

Special Case n large, θ small

00000001 000000001 000001 0

Consider $n \longrightarrow \infty$

$\theta \longrightarrow 0$

but where $n\theta = \lambda$ is constant

$\therefore \theta = \frac{\lambda}{n}$, and consider λ to be the rate at which successes occur per n trials.

Now

$$G_X(t) = (1-\theta+\theta t)^n$$

$$= \left(1 - \frac{\lambda}{n}(1-t)\right)^n$$

$\therefore$ as $n \longrightarrow \infty$ with λ, t constant

$$G_X(t) \longrightarrow e^{-\lambda(1-t)}$$

(by definition

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

$\therefore$ In the limit as $n \longrightarrow \infty$

$$G_X(t) = \exp\{\lambda(t-1)\}$$

Can we deduce the form of f_X in this limiting case ?

$$G_X(t) = \exp\{\lambda(t-1)\}$$

$$= \sum_{i=0}^{\infty} \frac{\{\lambda(t-1)\}^i}{i!}$$

$$= \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \underbrace{\left\{ \sum_{j=0}^{i} \binom{i}{j} t^j (-1)^{i-j} \right.}_{\text{BINOMIAL EXPANSION}}$$

$$= \sum_{i=0}^{\infty} \sum_{j=0}^{i} \frac{\lambda^i}{i!} \frac{i!}{j!(i-j)!} t^j (-1)^{i-j}$$

$$= \sum_{j=0}^{\infty} \left\{ \sum_{i=j}^{\infty} \frac{\lambda^i}{j!(i-j)!} (-1)^{i-j} \right] t^j$$

$$= \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \left\{ \sum_{i=j}^{\infty} \frac{(-\lambda)^{i-j}}{(i-j)!} \right] t^j$$

$$= \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \left\{ \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} \right] t^j$$

$$= \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} t^j$$

NOTE

Can show

$$P[X = x] = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

$$= \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$\longrightarrow \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{as } n \to \infty.$$

see Ex 5 Qu 1 SOLUTIONS.

$$\therefore \quad G_X(t) = \sum_{j=0}^{\infty} \frac{\lambda^j e^{-\lambda}}{j!} t^j$$

BUT we know that G_X is the generating function for the probabilities $P[X = x]$

$$\therefore \quad f_X(x) = P[X = x] = \frac{\lambda^x e^{-\lambda}}{x!}$$

in this limiting case!

NOTE When considering $n \longrightarrow \infty$, we deduce

$$\mathbb{X} \equiv \{0, 1, 2, \dots\}$$

5.6 THE POISSON DISTRIBUTION

Discrete r.v. X

Range $\mathbb{X} \equiv \{0, 1, 2, \dots\}$

$$f_X(x) = \frac{e^{-\lambda} \lambda^x}{x!} \qquad x = 0, 1, 2, \dots$$

This is the Poisson distribution with parameter λ.

$$X \sim \text{Poisson}(\lambda)$$

Interpretation + Context

(i) Limiting case of Binomial

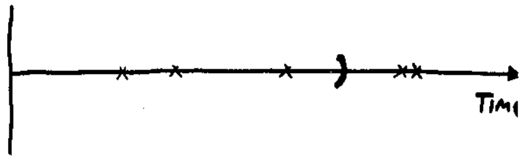
$n \longrightarrow \infty$ $n\theta = \lambda$ constant.
$\theta \longrightarrow 0$

i.e "SUCCESSES RARE, OCCUR AT CONSTANT RATE λ"

(ii) Used as a general model for count data

e.g "COUNT EVENTS OCCURRING IN FIXED TIME PERIOD"

THE POISSON PROCESS MODEL



x - Event occurs

- events occur
  - (i) at random
  - (ii) at constant rate λ per unit time
  - (iii) independently of other events

Consider interval $[0,t)$.

Define discrete r.v. X_t by

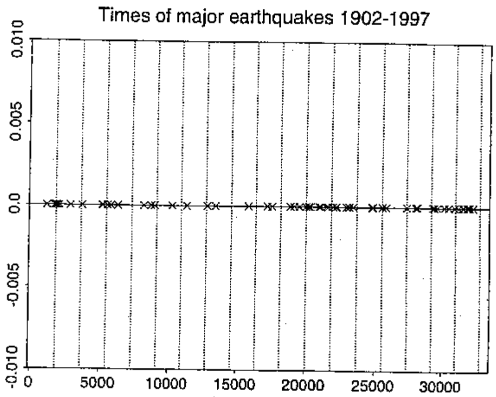
X_t = " # events that occur in $[0,t)$"

$\Rightarrow \mathbb{X}_t \equiv \{0,1,2,\ldots\}$

Can prove that

$$P[X_t = n] = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \quad n=0,1,\ldots$$

i.e $X_t \sim \text{Poisson}(\lambda t)$



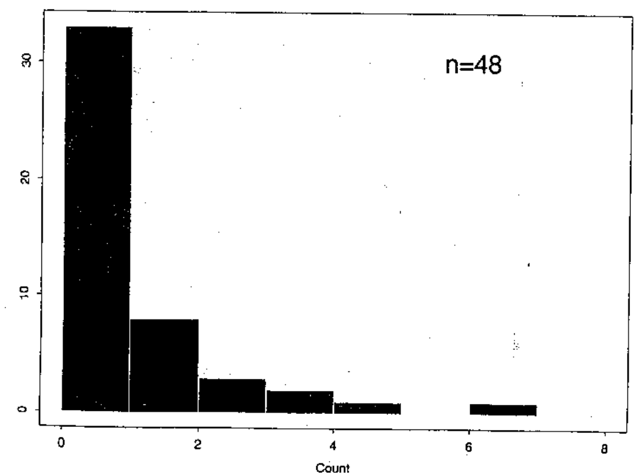
Break down 95 years into 19x5 year blocks + count the number in each block

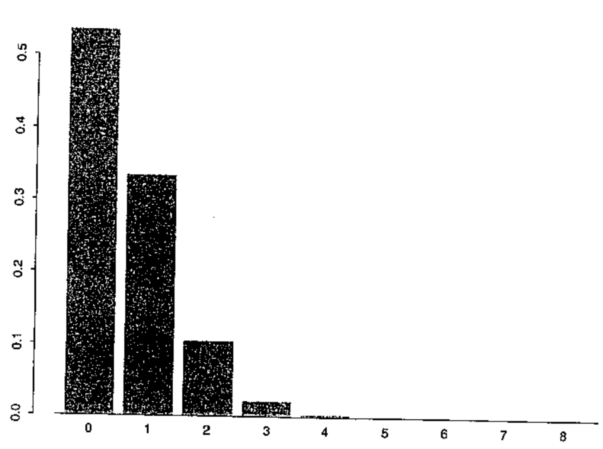
TABLE 4.2.3 OBSERVED HORSE-KICK FATALITIES

x = Number of Deaths	Observed Number of Corps-Years in Which x Fatalities Occurred
0	109
1	65
2	22
3	3
4	1
	200

Altogether there were 122 fatalities [109(0) + 65(1) + 22(2) + 3(3) + 1(4)], meaning that the observed fatality *rate* was 122/200, or 0.61 fatalities per corps-year. For reasons that will be discussed in Chapter 5, Bortkiewicz set λ equal to 0.61 and proposed as a model for X,

$$P(X = x) = p(x, 0.61) = \frac{e^{-0.61}(0.61)^x}{x!}$$





Uses :

- any counting experiment where $\mathcal{X} = \{0, 1, 2, 3, \ldots\}$ and events occur at constant rate.
- # earthquakes in one year
- # particles released in 8 mins from radioactive source
- # deaths per year due to horse-kick in the Prussian cavalry

PGF $G_X(t) = \exp\{\lambda(t-1)\}$

$$= e^{-\lambda} \exp\{\lambda t\} \quad \text{etc}$$

proved previously

EXPECTATION

$$E_{f_X}[X] = \lambda$$

proved previously

λ - controls "shape" of pmf
- controls "location" or "centre of probability mass"

$$f_X(x) = P[X=x]$$

$$= (1-\theta)^{x-1}\theta \qquad x=1,2,\dots$$

[need $x-1$ FAILURES, then a SUCCESS]

This is the GEOMETRIC pmf

$$X \sim Geometric(\theta)$$

CDF For $x \in \mathbb{X}$

$$F_X(x) = 1-(1-\theta)^x$$

proved previously.

5.7 THE GEOMETRIC DISTRIBUTION

Experiment : Repeated, independent Bernoulli(θ) trials.

Stop after 1st success.

e.g.
0001
001
0000001

possible outcome sequences.

X - # trials required to obtain first success

$$\Rightarrow \mathbb{X} \equiv \{1,2,3,\dots\}$$

PGF $G_X(t) = \sum_{x=1}^{\infty} f_X(x)t^x$

$$= \sum_{x=1}^{\infty} (1-\theta)^{x-1}\theta t^x$$

$$= \frac{\theta t}{1-(1-\theta)t} \qquad \text{G.P.}$$

EXPECTATION

$$E_{f_X}[X] = \sum_{x=1}^{\infty} x\, f_X(x)$$

$$= \sum_{x=1}^{\infty} x\,(1-\theta)^{x-1}\theta$$

$$= \frac{1}{\theta} \qquad \text{(MAPLE)}$$

5·8 THE NEGATIVE BINOMIAL DISTRIBUTION

Experiment : Repeated, independent Bernoulli (θ) trials.

Stop after $\underline{n^{th}}$ success $(n=1,2,\dots)$

eg $n=3$

001011	6
001000101	9
1011	4

Define discrete r.v. X - # trials require

$\Rightarrow \mathbb{X} = \{n, n+1, n+2, \dots\}$

$$f_X(x) = P[X=x]$$

$$= \underline{\binom{x-1}{n-1}\theta^{n-1}(1-\theta)^{x-n}} \times \theta$$

[need $n-1$ successes in $x-1$ trials, then a final success on the x^{th} trial — first part follows BINOMIAL dist^n.

$$\therefore f_X(x) = \binom{x-1}{n-1}\theta^n(1-\theta)^{x-n}$$

$x = n, n+1, \dots$

This is the NEGATIVE BINOMIAL pmf

$$X \sim NegBinomial(n, \theta)$$

NOTE

$$\sum_{x\in\mathbb{X}} f_X(x) = 1$$

$$\Longleftrightarrow \sum_{x=n}^{\infty}\binom{x-1}{n-1}\theta^n(1-\theta)^{x-n} = 1$$

$$\Longleftrightarrow \sum_{y=0}^{\infty}\binom{y+n-1}{n-1}\theta^n(1-\theta)^y = 1$$

$y = x-n$.

[RECALL THE NEGATIVE BINOMIAL EXPANSION

$$\frac{1}{(1-t)^n} = \sum_{y=0}^{\infty}\binom{y+n-1}{y}t^y$$
]

NOTE

If $X_1, X_2, \dots, X_n$ are i.i.d discrete r.v.s with

$$X_i \sim Geometric(\theta) \quad i=1,2,\dots,n$$

then if $X = \sum_{i=1}^{n} X_i$, we have

$$X \sim NegBinomial(n, \theta)$$

e.g $n=4$

$$\overbrace{\underbrace{001}_{X_1}\,\underbrace{00001}_{X_2}\,\underbrace{01}_{X_3}\,\underbrace{001}_{X_4}}^{X}$$

PGF $G_X(t) = \left\{ \dfrac{\theta t}{1-(1-\theta)t} \right\}^n$

Proof Exercise ....

(uses negative binomial expansion)

NOTE Consider discrete r.v. Y

$Y = X - n$ $\boxed{\text{Ex 6, Q 2.}}$

$\mathbb{Y} \equiv \{0, 1, 2, \dots\}$

$f_Y(y) = P[Y=y] = P[X-n=y]$

$= P[X = n+y] = f_X(n+y)$

– defines "alternative" representation of NegBinomial (as in MAPLE...)

5.9 THE HYPERGEOMETRIC DISTRIBUTION

Experiment Sampling without replacement from finite population.

Population size : N

TYPE I items : R
TYPE II items : $N-R$

Sample of size n.

Define discrete r.v. X by

X = # TYPE I items in the sample.

$\mathbb{X} \equiv \{ \text{Max}\{0, n-(N-R)\}, \dots, \text{Min}\{n, R\} \}$

and

$$f_X(x) = \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}}$$

$$= \frac{\binom{n}{x}\binom{N-n}{R-x}}{\binom{N}{R}}$$

5.10 PROBABILITY GENERATING FUNCTION

Recall for a discrete r.v X with range $\mathbb{X}$, pmf f_X.

PGF

$$G_X(t) = \sum_{x \in \mathbb{X}} f_X(x)\, t^x$$

(when this sum is convergent...)

Each f_X has a unique corresponding G_X.

Uses

(i) to identify distributions "UNIQUENESS"

(ii) to calculate probabilities in "complex" problems.

The next theorem is a KEY result in the use of pgfs

SUMS OF RANDOM VARIABLES: A KEY PGF RESULT

Suppose that X_1 and X_2 are discrete random variables with ranges $\mathbb{X}_1$ and $\mathbb{X}_2$, probability mass functions f_{X_1} and f_{X_2} and probability generating functions G_{X_1} and G_{X_2} respectively. Suppose also that X_1 and X_2 are independent. Define discrete random variable Y by

$$Y = X_1 + X_2$$

Using the Theorem of Total Probability, by partitioning on the different possible values of X_1, for y in an appropriate range $\mathbb{Y}$

$$\begin{aligned} f_Y(y) = \mathrm{P}[\,Y = y\,] &= \sum_x \mathrm{P}[\,(X_1 = x) \cap (X_2 = y - x)\,] \\ &= \sum_x \mathrm{P}[\,X_1 = x\,]\,\mathrm{P}[\,X_2 = y - x\,] \\ &= \sum_x f_{X_1}(x) f_{X_2}(y - x) \end{aligned}$$

where the second line follows from the independence of X_1 and X_2, and all summations are over $x \in \mathbb{X}_1$. Hence we now have an expression for the pmf for the new variable Y.

Now,

$$\begin{aligned} G_Y(t) &= \sum_y f_Y(y) t^y = \sum_y \left\{ \sum_x f_{X_1}(x) f_{X_2}(y - x) \right\} t^y \\ &= \sum_{x_1} \sum_{x_2} f_{X_1}(x_1) f_{X_2}(x_2) t^{x_1 + x_2} \qquad \text{changing variables to } x_1 = x \quad x_2 = y - x \\ &= \left\{ \sum_{x_1} f_{X_1}(x_1) t^{x_1} \right\} \left\{ \sum_{x_2} f_{X_2}(x_2) t^{x_2} \right\} \\ &= G_{X_1}(t)\, G_{X_2}(t) \end{aligned}$$

so therefore

$$G_Y(t) = G_{X_1}(t)\, G_{X_2}(t)$$

EXTENSION

If $X_1, X_2, \ldots, X_n$ are independent discrete random variables with pgfs $G_{X_1}, G_{X_2}, \ldots, G_{X_n}$ respectively, and discrete random variable Y is defined by

$$Y = \sum_{i=1}^{n} X_i$$

then by induction on n using the above result we have that

$$G_Y(t) = \prod_{i=1}^{n} G_{X_i}(t)$$

so that if $X_1, X_2, \ldots, X_n$ are also identically distributed with pgf G_X then

$$G_Y(t) = \{G_X(t)\}^n$$

(see, for example, the Bernoulli/Binomial pgfs or the Geometric/Negative Binomial pgfs)

THEOREM

Suppose discrete r.vs X_1 and X_2 are independent with pgfs G_{X_1} and G_{X_2}. Then if discrete r.v. Y is defined by

$$Y = X_1 + X_2$$

we have

$$G_Y(t) = G_{X_1}(t)\, G_{X_2}(t)$$

for the pgf of Y.

PROOF See Ex 6 Qu 9/10.
+ HANDOUT.

$$P[Y = y] = \sum_x P[(Y = y) \cap (X_1 = x)] \quad \text{(TH}^{\text{M}}\text{ TOTAL PROB}$$

$$= \sum_x P[(X_1 = x) \cap (X_2 = y - x)]$$

$$= \sum_x P[X_1 = x]\, P[X_2 = y - x] \quad \text{(INDEPENDENCE}$$

$$\therefore \quad f_Y(y) = \sum_x f_{X_1}(x)\, f_{X_2}(y - x)$$

where summations are over $x \in \mathbb{X}_1$, for appropriate $y \in \mathbb{Y}$

$$\therefore G_Y(t) = \sum_y f_Y(y)\, t^y$$

$$= \sum_y \left\{ \sum_x f_{X_1}(x) f_{X_2}(y-x) \right\} t^y$$

$$= \left\{ \sum_x f_{X_1}(x)\, t^x \right\} \left\{ \sum_z f_{X_2}(z)\, t^z \right\} \quad z = y - x$$

$$= G_{X_1}(t)\, G_{X_2}(t)$$

EXTENSION

By induction on n, if $X_1, \ldots, X_n$ are independent and

$$Y = \sum_{i=1}^{n} X_i$$

then

$$G_Y(t) = \prod_{i=1}^{n} G_{X_i}(t)$$

$\therefore$ can routinely calculate the pgf for sums of independent discrete rvs.