ST2054 ~~9/22/08~~

Difference between Stats & Prob.

Prob is concerned with models for uncertainty
↳ will usually involve parameters

Stats is concerned wit using data to help choose an appropriate model + use data to estimate the parameters of dat model.

Terms

1) Random experiment : A process of doing / observing something where the outcome is uncertain

Eg Coin toss - outcomes H, T.
Dice throw -
Choosing card from deck
A measurement, A Game, Weather

2) Outcome : Is one of the experiment can terminate

eg. coin toss - outcome H, T
Dice throw - 1, 2, 3, 4, 5, 6

H,T 2 cards. counts
Continuous

3) Sample Space / outcome space :
The set of all outcomes for a particular experiment

denote by S or Ω

4) Event : A subset of the sample / out come space
Denote by capital letters . . .

e.g Die Throw → E - getting a 6

To specify an event we can either specify a property that describes the event
or
List the outcomes (elementry outcomes) that make up the event

$$\{\text{Even outcome}\} \equiv \{2, 4, 6\}$$

Note: We say an event has happened if Any of the outcomes in that event have accured

5) Null Event: ϕ
This is the impossible event

6) Unions and Intersection of Events
The union of a collection of events is the set of outcomes such that each outcome belongs to at least of the events of the collection

The intersection of a collection of events is the set of outcomes such that each outcome belongs to every one of the events of the collection.

Properties (from set theorey)

$A \cup B \equiv B \cup A$
$A \cap B = B \cap A$ } Comutitive law

$(A \cup B) \cup C = A \cup (B \cup C)$
$(A \cap B) \cap C = A \cap (B \cap C)$ } Associative

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ DISTRIBUTIVE

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

7) Complement of event A (A^c)

Is defined as the event that A does not occur.

" Difference of events ($A \setminus B$)

defined $A \cap B^c$

Properties of events.

1) Mutually exclusive: Events such that $A \cap B = \phi$

- extends to any number of events

$A, B, C, \ldots$ are mutually exclusive if every pair are mutally exclusive

eg. A, B, C are M.E

$A \cap B = \phi$

$A \cap C = \phi$

$B \cap C = \phi$

$A \cap B \cap C = \phi$

2) Collectively Exhaustive events:

collection of events $\{E_i\}$ are said to be coll. exhasive if

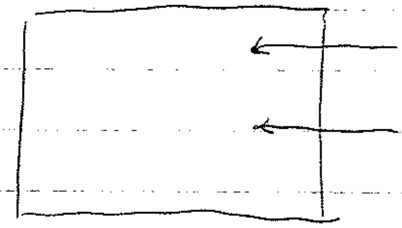
The union of the outcomes of the events make up S

eg. E_1: Rolling an even number

E_2: Rolling a 1, 3, 5

Use of Venn Diagrams (to Display Events)

Consider a random experiment: Throw a dart at a rectangle board (bound to hit it)

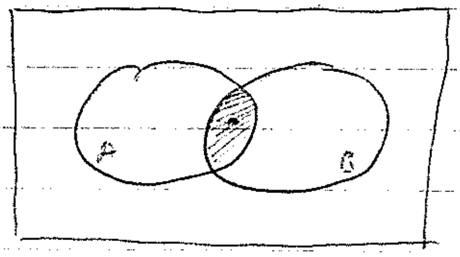


one outcome (pt of contact

Another outcome

so the union of all the outcomes is the S or the rectangle.

[S made up of all pts in rectangle]

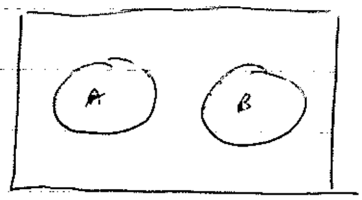


Event A: occurs if dart hits any pt in A

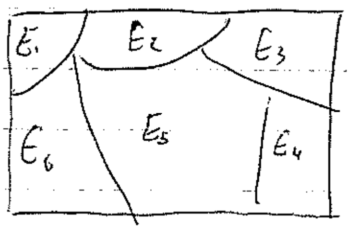
Event B: occurs if dart hits any pt in B

If dart hits red spot $A \cap B$ has occured.

A^c:



E_A, E_B are mutually exclusive

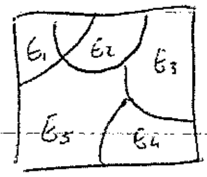


Note $\{E_i\}$ are mutually exclusive (ie $E_i \cap E_j = \phi$) no intersection

Also $\{E_i\}$ are collectively exhasitive

Union of all possible outcomes $= \Omega$

☆ Note: Possible to be C.E but not M.E



Not m.E cuz $E_1 \cap E_2 \neq \emptyset$
but coll. Ex. cuz union of all Events $= \Omega$.

ex: Die : E_1 : Even outcome
E_2 : 4 or 5 or 6
E_3 : 1 or 3

} C.E Not M.E

Partition of a sample Space
Suppose we have a collection $\{E_i\}$ which is m. Ex & Coll Ex.
we refer to this collection as a partition of the sample space Ω

Probability

17th Century Pascal | Fermat .. Bernoulli De Moivre Bayes.

We will think of prob as a measure of the likely hood that some event will occur

Russian dude proposed we sud specify some minimal assumptions for da probability set function $P(E)$

These assumptions are known as the Axioms of Probability

→ we have a collection of events (for any Ran exp) $\{E_i\}$ and these are m. ex.
(the collection is countably infinite)

Then, we associate a ~~volume~~ value $(P(E_i))$ wit each event

1) $0 \leq P(E_i) \leq 1$

2) $P(\Omega) = 1$

$E_1, \ldots E$ ME

3) $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$ [Axiom of countable additivity]

Deductions from these axioms:

i) We show that $P(\phi) = 0$

Proof: Consider $\{E_i\}$ such that

E_1 is Ω

$E_2, E_3, E_4, \ldots$ are each ϕ

☆ $\{E_i\}$ are M. Ex

Apply Axiom 3: $P(\bigcup_{i=1}^{\infty} E_i) = \sum P(E_i)$

$\bigcup_{i=1}^{\infty} E_i = \Omega$

$$P(\Omega) = P(E_1) + P(E_2) + P(E_3) + \ldots$$

$$= P(\Omega) + P(\phi) + P(\phi)$$

$$1 = 1 + P(\phi) + P(\phi) + \ldots$$

and since all $P(A_i) \geq 0$

$$\Rightarrow P(\phi) = 0$$

ii) Finite collection of events $E_1, E_2, E_3, \ldots, E_n$ all M.E
Then $P(\bigcup_{i=1}^{n} E_i) = \sum_{i}^{n} P(E_i)$

Proof: Introduce $\{A_i\}$
such that $A_i = E_i$ for $i = 1, 2, \ldots, n$
$A_i = \phi$, for $i > n$

the collection $\{A_i\}$ consists of M.E events so...
xiom 3 applies: Note that $\bigcup_{i=1}^{\infty} A_i = (\bigcup_{i=1}^{n} E_i) \cup A_{i+1} \cup A_{i+2} + \ldots$

But $\bigcup_{i=1}^{\infty} A_i$ is just $\bigcup_{i=1}^{n} E_i$

so dat $P(\bigcup_{i=1}^{\infty} A_i) = P(\bigcup_{i=1}^{n} E_i)$

But $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i}^{\infty} P(A_i)$

$= \sum_{i}^{n} P(E_i) + \sum_{i=n+1}^{\infty} P(\phi)$

$\Rightarrow P(\bigcup_{i=1}^{n} E_i) = \sum_{i}^{n} P(E_i)$

iii) For any event A, $P(A^c) = 1 - P(A)$

Proof: Note dat Ω / Sample space $= A \cup A^c$
 $\hookrightarrow$ is covers everytin

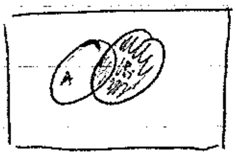
Apply $P(\cdot)$ [Probability set function]
to both sides
means by which you find the prob of a certain event occuring

$\Rightarrow P(\Omega) = P(A \cup A^c)$
$= P(A \cup A^c) = P(A) + P(A^c)$
$1 = P(A) + P(A^c)$
$\Rightarrow P(A^c) = 1 - P(A)$

iv) 2 Events A, B for a random expt.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Proof :



Consider

$$* \; B = (B \cap A) \cup (B \cap A^c)$$

= d portion of

Also $A \cup B = A \cup (B \cap A^c)$ [M.E]

Apply $P(\cdot)$ to both sides

$P(A \cup B) = P[A \cup (B \cap A^c)]$ cuz M.E

(by deduction 2)

$= P(A) + P(B \cap A^c)$ (**)

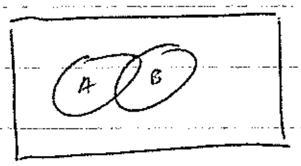
Similarly in case of *

$$P(B) = P(B \cap A) + P(B \cap A^c)$$

Substitute for $P(B \cap A^c)$ into (**)

$$P(A \cup B) = P(A) + [P(B) - P(B \cap A)]$$
$$P(A \cup B) = P(A) + P(B) - P(B \cap A)$$

Intuitive discussion :



Claim that

$P(A) = \frac{\square A}{\square}$ - Area of region A / Area of full rectangle

$$P(B) = \frac{\square B}{\square}$$

Consider next $P(A \cup B) = \dfrac{\square(A \cup B)}{\square}$

but $\square(A \cup B) = \square A + \square B - \square(A \cap B)$

Divide across by $\square$

$$\frac{\square(A \cup B)}{\square} = \frac{\square(A)}{\square} + \frac{\square(B)}{\square} - \frac{\square(A \cap B)}{\square}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

V) LAW OF TOTAL PROBABILITY

we have a collection $\{E_i\}$ of M. Ex events

$\{E_i\}$ is a partition - M.E (finite or infinite)

and C. Exhaustive

and A is some other event

Then

$$P(A) = \sum_{i=1}^{\infty} P(A \cap E_i)$$

Proof: Since $\{E_i\}$ is a partition then $\Omega = \bigcup_{i=1}^{\infty} E_i$

and $A \cap \Omega = A \cap (\bigcup_{i=1}^{\infty} E_i)$ compact way of $A \cap (E_1 \cup E_2)$

$A \cap \Omega = A$ $\quad$ $A = \bigcup_{i=1}^{\infty} (A \cap E_i)$ $\quad$ $(A \cap E_1) \cup (A \cap E_2) \Rightarrow \bigcup_{i=1}^{2} (A \cap E_i)$

Note that the events $\{A \cap E_i\}$ are M. Excl.

Axiom 3 $\Rightarrow$ $P(A) = P[\bigcup_{i=1}^{\infty} (A \cap E_i)]$

$$= \sum_{i=1}^{\infty} P(A \cap E_i)$$

Axiom 3.

Vi) we had the addition Law

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Can show $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C)$
$- P(B \cap C)$
$+ P(A \cap B \cap C)$

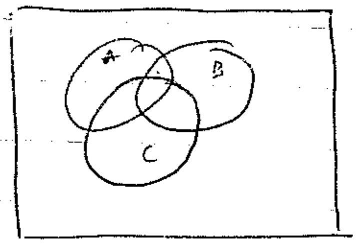
Proof: with $P(A \cup B \cup C) = P(A \cup D)$ ie $B \cup C = D$
$= P(A) + P(D) - P(A \cap D)$

$= P(A) + P(B \cup C) - P(A \cap (B \cup C))$
$= P(A) + P(B) + P(C) - P(B \cap C) - P[(A \cap B) \cup (A \cap C)]$

$= P(A) + P(B) + P(C) - P(B \cap C) - \underbrace{\{P(A \cap B) + P(A \cap C) - P[(A \cap B) \cap (A \cap C)]\}}_{\text{Addition law.}}$

$\cup C) = P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) + P[A \cap B \cap C]$

* $P[(A \cap B) \cap (A \cap C)]$ is simplfied to $P[A \cap B \cap C]$



$A \cup B \cup C$
Area $(A \cup B \cup C)$
$= \square(A) + \square(B) + \square(C) - \square(A \cap B) - \square(A \cap C)$
$- \square(B \cap C) + \square A \cap B \cap C.$

$P(A$ or B or $C)$ when m.E $= P(A) + P(B) + P(C)$

$P(A$ or B or $C)$ when not m.E $= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C)$
$- P(B \cap C) + P(A \cap B \cap C)$

In general:
$P(E_1$ or E_2 or E_3 ...... or $E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2}^{n} P(E_{i_1} \cap E_{i_2})$

$+ \sum_{i_1 < i_2 < i_3}^{n} P(E_{i_1} \cap E_{i_2} \cap E_{i_3}) - \ldots$

$+ (-1)^{n+1} P(E_1 \cap E_2 \cap \ldots \cap E_n)$

ST2054

TO SHOW

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) - \sum_{i<j} P(E_i \cap E_j) + \sum_{i<j<k} P(E_i \cap E_j \cap E_k)$$
$$+ \cdots (-1)^n \sum_{i_1<i_2<\cdots<i_{n-1}}^{n} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_{n-1}}) + (-1)^{n+1} P(E_1 \cap E_2 \cap \cdots E_n)$$

PROOF (INDUCTION)

For $(n=2)$, we have shown that $P(E_1 \cup E_2) = \sum_{i=1}^{2} P(E_i) - P(E_1 \cap E_2)$

Assume true for m, try to deduce for $(m+1)$.

$$P\left(\bigcup_{i=1}^{m+1} E_i\right) = P\left(\bigcup_{i=1}^{m} E_i \cup E_{m+1}\right) = P\left[\bigcup_{i=1}^{m} E_i\right] + P\left[E_{m+1}\right] - P\left[\left(\bigcup_{i=1}^{m} E_i\right) \cap E_{m+1}\right]$$

$$= \left[\sum_{i=1}^{m} P(E_i) + P(E_{m+1}) - \sum_{i<j} P(E_i \cap E_j) + \cdots\cdots\right] - \underbrace{P\left[\bigcup_{i=1}^{m} (E_i \cap E_{m+1})\right]}$$

$$= \sum_{i=1}^{m+1} P(E_i) + \Big(- \sum_{i<j}^{m} P(E_i \cap E_j) \qquad - \sum_{i=1}^{m} P(E_i \cap E_{m+1})$$
$$+ \sum_{i<j<k}^{m} P(E_i \cap E_j \cap E_k) \qquad - \Big\{ - \sum_{i<j}^{m} P\big(\underbrace{(E_i \cap E_{m+1}) \cap (E_j \cap E_{m+1})}_{E_i \cap E_j \cap E_{m+1}}\big)$$
$$- - - - - - - - - -$$
$$+ (-1)^m \sum_{i_1<i_2<\cdots i_{m-1}}^{m} P(E_{i_1} \cap E_{i_2} \cap \cdots\cdots \cap E_{i_{m-1}}) \qquad - \Big\{ (-1)^m \sum_{i_1<i_2 \cdots < i_{m-1}}^{m} P(E_{i_1} \cap E_{i_2} \cap \cdots E_{i_{m-1}} \cap E_{m+1}$$
$$+ (-1)^{m+1} P(E_1 \cap E_2 \cdots\cdots \cap E_m) \Big) \qquad - (-1)^{m+1} P(E_1 \cap E_2 \cap \cdots\cdots E_m \cap E_{m+1}$$

$$= \sum_{i=1}^{m+1} P(E_i) - \sum_{j<j}^{m+1} P(E_i \cap E_j) + \sum_{i<j<k}^{m+1} P(E_i \cap E_j \cap E_k)$$
$$+ \cdots\cdots\cdots (-1)^{m+1} \sum_{i_1<i_2<\cdots<i_m}^{m+1} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m})$$
$$+ (-1)^{m+2} P(E_1 \cap E_2 \cap \cdots \cap E_m \cap E_{m+1})$$

which $\Rightarrow$ Proposition is true for $(m+1)$. (Completes proof by induction)

NOTE ON 2nd LAST TERM ON RHS (UNDERLINED)

$$- (-1)^m \sum_{i_1<i_2 \cdots i_{m-1}}^{m} P(E_{i_1} \cap E_{i_2} \cap \cdots E_{i_{m-1}} \cap E_{m+1})$$

$$= (-1)^{m+1} \Big[P(E_1 \cap E_2 \cap \cdots\cdots \cap E_{m-1} \cap E_{m+1}$$
$$+ P[E_2 \cap E_3 \cap \cdots\cdots \cap E_m \cap E_{m+1}$$
$$+ P[E_1 \cap E_3 \cap \cdots\cdots \cap E_m \cap E_{m+1}$$
$$+ P[E_1 \cap E_2 \cap E_4 \cap \cdots\cdots \cap E_m \cap E_{m+1}$$
$$- - - - - - - - - - -$$
$$+ P[E_1 \cap E_2 \cap E_3 \cap \cdots E_{m-3} \cap E_{m-2} \cap E_{m-1} \cap E_{m+1} \Big]$$

WHEN WE ADD $(-1)^{m+1} P[E_1 \cap E_2 \cap \cdots E_m$ TO THIS

WE GET:

$$(-1)^{m+1} \sum_{i_1<i_2<\cdots<i_m}^{m+1} P(E_1 \cap E_2 \cap \cdots \cap E_{i_m})$$

How do we Find Numerical Values

There are 3 Types /approaches

i) Classical Prob. (A Priori Prob.)

ii) Empirical Prob. (Statistical Prob.)
e.g: Life tables. Try to repeat exp. N times.

iii) Subjective Probability Assignment
Idea is to use intuition to find value of P(A)
e.g Man Utd vs Liverpool
ii) cant be repeated numerous times under identical cond.
i) Utd + Liverpool don't have d exact same prob of victory - at old Trafford Utd prob will win

ii) The event A occurs, denote by n_A
Then we take as

$$P(A) \rightarrow \frac{n_A}{N}$$

But, there are many cases where the R. exp can't be repeated.

i) Classical Probability
This approach is based on noting that the random experiment has symmetry present so we can claim dat all the possible outcomes are equally LikeLy

Coin Toss (2 outcomes => P(H) = P(T))
Dice - Throw (6 outcomes => P(ω) = 1/6)

☆ If there are M outcomes and all equally likely the prob of any specified outcome is $\frac{1}{M}$.

$= \frac{1}{16}$

$P(\text{Queen}) = P(Q_H \cap Q_S \cap Q_C \cap Q_D)$ — m.e.

$\Rightarrow P(Q_H) + P(Q_S) + P(Q_C) + P(Q_D)$

$= \frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \frac{1}{52}$

$= 4/52$

Examples: i) 2 coins, thrown together

$P(\text{getting 1 Head}) = \frac{1}{2}$

4 coins, throw together

$P(\text{getting 2 Heads}) = \; ?$

H H T T

i) $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16}$

ii) H T T H $= 1/16$

iii) T T H H $= 1/16$

iv) T H H T $= 1/16$ — m.e.

v) T H T H vi) H T H T $= 1/16$

so $P(i \cap ii \cap iii \cap iv) = P(i) + P(ii) + P(iii) + P(iv) + P(v) + P(vi)$

or $\boxed{2}\ \boxed{2}\ \boxed{2}\ \boxed{2} = 16$ total

$\boxed{H}\ \boxed{H}\ \boxed{T}\ \boxed{T}$ so like 1ser

$= \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16}$

its $\frac{4!}{2! \times 2!} = 6$

$= \frac{6}{16}$.

- In dis small example it was possible to list all the out comes (Innumeration) & choose the one we wanted

Ex: 2 dice thrown

FOR CLASSIC: $\frac{\text{NO OF FAVOURABLE}}{\text{Total No. of outcomes}}$ — DRAW BOXES

WGB_G08 — ST2054 5-6 Thursday

$E \subseteq F \Rightarrow E^c \subseteq F$

Computation of Probabilities. — Classical / Empirical / Subjective

(1) Classical.

- Look for elementary events and symmetries and ~~compute~~ build up probabilities from them.

Examples. (i)
Prob of Queen

$$P[Q] = P[Q_H \cup Q_S \cup Q_C \cup Q_D] = \frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \frac{1}{52} = \frac{1}{13}$$

(ii)
Toss 2 coins : Prob of 1 Head.

HH HT TH TT

$$P[1 \text{ head}] = P[HT \cup TH] = \cancel{\frac{1}{4} + \frac{1}{4}}$$

$$= P[HT] + P[TH]$$

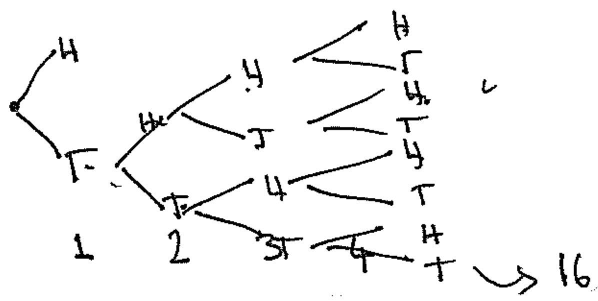
$$= P(H)P(T) + P(T)P(H)$$

$$= \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{1}{2}$$

Toss 4 coins : Prob 2 Heads

2222 = 16 possible outcomes

ME { HHTT, . . .
HTTH
TTHH
THHT
THTH
HTHT



121

$$\frac{\#\text{FAVORABLE}}{\#\text{POSSIBLE}}$$

~~Example~~ D

(III) THROW 2 DICE Prob sum = 3

$$= \frac{2}{36}$$

1 — 2
2 — 1
3
4
5
6

FIRST SECOND

36 POSSIBILITIES

(IV) Galileo. 3 DICE : PROB SUM = 9?
10?

1
2
3
4
5
6

$$P[9] = \frac{25}{216}$$

$$P(10) = \frac{27}{216}$$

	$=9$	$=10$	
6	1+2+6	1 3 6	6
6	1+3+5	1 4 5	6
3	1+4+4	2 2 6	3
3	2 2 5	2 3 5	6
6	2 3 4	2 4 4	3
1	3 3 3	3 3 4	3

- For bigger Tasks:

Example: gamblers wud throw 3 dice $\quad \underline{\Sigma = 9 \text{ vs } \Sigma = 10}$

reasoned that $P(9)$ should be $= P(10)$

But deir experiece was suspect

Galilea listed the sample space ($6 \times 6 \times 6$ outcomes)

$$P(9) = \frac{25}{216} \qquad P(10) = \frac{27}{216}$$

If no. of outcomes (N) is very large then innumeration / Listing of outcomes is not practical

So we need help in counting (N) and in counting ^{N}r $^{N}C_r$

nos of favourable outcomes. BIRTHDAY PROBLEM

COUNTING METHOD'S ... See sheet (3)

Problem: (from Sheet 3 stuff).

Estimation of 'animal abundance'

eg (lake: how many fish r there)

Step 1: Capture N_1 fish

mark them all

release them

Step 2: later, capture n fish

Check how many are marked $- x$

Step 3: How to estimate N

Get sample prob. = HYPER

BIRTHDAY PROBLEM

n INDIVIDUALS : ~~WHAT IS THE~~ PROB THAT AT least 2 have same Birthday

365 DAYS TO CHOOSE FROM.

OF POSSIBLE CONFIGS FOR n BIRTHDAYS

$$= 365^n$$

# of WAYS ALL n BIRTHDAYS ARE DISTINCT.

$$= 365 \cdot 364 \cdots (365-n+1)$$

PROB ALL BIRTHDAYS DISTINCT $= \dfrac{365 \ldots (365-n+1)}{365^n}$

PROB AT LEAST 2 THE SAME $= 1 - \dfrac{365 \ldots (365-n+1)}{365^n}$

n	P(A)
4	.016
23	.507
32	.753

Let's check.

SAMPLING

$(C_1, \ldots C_N)$

CHOOSE n items with replacement

\# of CONFIGURATIONS $= N^n$

OUTCOMES

$\{C_{i_1}\ C_{i_2}\ \ C_{i_n}\}$

2. CHOOSE n items without replacement.

\# of configurations $= N^{(n)}$

$= N(N-1)\ldots[N-n+1]$

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

3. LABELED Items

$\{C_1 \ldots C_N\}$

N_1 OBJECTS OF TYPE 1

N_2 - - - 2

N_k - - - k

$N_1 + N_2 + \ldots N_k = N$

COUNTS ON CARDS WITHOUT REPLACEMENT

$(C_1, \ldots, C_N)$ cards.

N_1 of Type 1
N_2 - - - 2

N_K of Type K

DRAW A SET of $n \leq N$ cards WITHOUT REPLACEMENT

\# of POSSIBLE sets of cards (sezen)

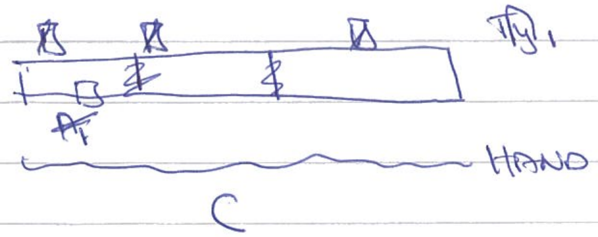
$$= N \ldots (N-n+1) = N^{(n)}$$

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

HAND n_1 of Type 1
n_2 - - 2
n_K K

$$\sum n_k = n$$

C = How many hands with of type $(n_1, n_2, \ldots n_K)$
Here the order does not matter



$\text{ARR} = n_1! \, n_2! \ldots n_K! \; n!$ $\quad C * \text{ARRANGEMENTS} = N_1^{(n_1)} \ldots N_K^{(n_K)}$

$$C = \frac{N_1^{(n_1)} \cdots N_K^{(n_K)}}{n_1! \ldots n_K! \; n!}$$

$$C \neq \frac{N_1!}{n_1!(N_1-n_1)! \; n!} \cdots$$

$$\text{Prob of Hand} = \frac{\displaystyle\prod_{k=1}^{K} \frac{N_k!}{n_k!\,(N_k-n_k)!}}{\dfrac{N!}{n!\,(N-n)!}}$$

$$= \frac{\displaystyle\prod_{k=1}^{K} {}^{N_k}C_{n_k}}{{}^{N}C_{n}}.$$

~~Sample of~~

Hypergeometric Dist.

$\binom{N}{n}$

ARRANGEMENTS OF $x_1, \ldots x_J$ $\sum x_i = n$ cards.

$n! = x_1! \ldots x_J!\, A$

$A = \dfrac{n!}{x_1! \ldots x_J!}$

$N \begin{cases} N_1 = S \\ N_2 = N - N_1 = F \end{cases}$

Sample n items without replacement

What is the prob that there will be $\underline{x \text{ of types } S}$

$N^{(n)}$ = # of samples

$N_1^{(x)} N_2^{(n-x)}$ = # of Samples in specific order.

$\binom{n}{x}$ orderings.

# of ways to get result $= \dfrac{\binom{n}{x} N_1^{(x)} N_2^{(n-x)}}{N^{(n)}}$

$$\text{Prob.} = \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}}$$

$$\frac{n!}{x!\,(n-x)!} \cdot \frac{N_1!}{(N_1-x)!} \cdot \frac{N_2!}{(N_2-(n-x))!}$$

$$\frac{N!}{(N-n)!}$$

# of ways to choose x S's and $(n-x)$ F' in a sequence of length n.

$n!$ # of sequences

$$\cancel{\binom{n}{x} \cdot x! \, (n-x)!}$$

GIVEN A Sequence with x S and $(n-x)$ F

$x!$ ways to re-order S

$(n-x)!$ - - - F

$n!$ = # of ordered sequences.

$$n! = \binom{n}{x} \cdot x! \; n!$$

$$\binom{n}{x} = \frac{n!}{x! \, (n-x)!}$$

(①②③④⑤) □□□ 1 2 3   3 1 2
2 3 1
1 3 2
2 1 3
3 2 1

3 2 1

$$\frac{x}{n} = \frac{N_1}{N}$$

If the denominator is □□□ not ( ) then must count different patterns.

FRIDAY 9–10 WGB33

Example : Deck of 52 : select (choose) $n = 5$ cards
what is Prob [2 Hearts , 3 non-Hearts]
Here (in a hand of cards) order doesn't matter ... a full house is a full house !!

$$\frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}} = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}} = .2743$$

Example : Deck of 52 : select $n = 5$ cards
one by one with replacement
what is Prob [2 hearts , 3 non-Hearts]
here each selection is from 52

☆ Couldn't answer this using Combinations
must use ARRANGEMENTS

denominator $= 52^5$: $\boxed{52}\ \boxed{52}\ \boxed{52}\ \boxed{52}\ \boxed{52}$

Now suppose the 2 Hearts r in Positions 1 & 2

$\boxed{13}\ \boxed{13}\ \boxed{39}\ \boxed{39}\ \boxed{39}$

However this count ignores other ~~ARRANGEMENT~~ patterns such as

$\boxed{H}\ \boxed{Non}\ \boxed{Non}\ \boxed{Non}\ \boxed{H}$

The number of such patterns for Hearts & Non-Hearts is $\binom{5}{2}$ or $\frac{{}^5P_5}{3! \times 2!}$ ~~coz~~ not 120 --- think of MISSISSIPPI

For each such pattern we have $13^2 \cdot 39^3$ Arrangements

$\Rightarrow$ Prob (2 H, 3 Non-Hearts) $= \frac{\binom{5}{2}\, 13^2 \cdot 39^3}{52^5}$

$= .2638$

If denominator is $\square\square\square$ not ( ) den must alc for diff patters ~~Bt~~ $\frac{5!}{2! \times 3!}$.

Example of Binomial prob : $\binom{5}{2}\left(\frac{1}{4}\right)^2($

see sheet 4

Another use of hypergeometric Prob :

LOTTO : For one panel of choosing 6 from 42

What is Prob (4 match)

$$\frac{\binom{6}{4}\binom{36}{2}}{\binom{42}{6}} \approx \frac{1}{555.1}$$

Prob 3 : $\dfrac{\binom{6}{3}\binom{36}{3}}{\binom{42}{6}} \approx \dfrac{1}{36.7}$

Industry : statistical quality control → Acceptance Sampling.

Supplizes arrive in Batches of size N
Instead of testing all N
- select a sample of n (from N)
test all n.

let X denote number of defective items

If $X \le c$ accept the batch (all N)
if $X > c$ reject the batch

The specification of values of n and c is an issue to be decided.

Example : $N = 20$ select $5 = n$ $c = 0$
($c=0$ samples r considered poor)
reject if $X \geq 1$

Find i) P(Acceptance)

let N_1 denote the number of defectives in batch

$N_1 = 10$

$P(X=0) : \dfrac{\binom{10}{5}\binom{10}{0}}{\binom{20}{5}} = .016$ ✓

cuz we wanted this to be small cuz it was a very bad batch

(.2)(.4)(.3)
(.2)(.6)(.3)
(.5)(.4)(.3)
$\div$ (.2)(.4)(.3) = .024

$N_1 = 2$

$$\frac{\cancel{\binom{10}{2}}\binom{10}{3}}{\binom{20}{5}} = \cancel{.5526}$$

$$\frac{\binom{2}{0}\binom{18}{5}}{\binom{20}{5}} = .5526$$

CONDITIONAL PROB, BAYES Theorem & Independance

:01 Consider a random expt associated events 2 of which are A & B

Suppose were told that event B has occured

Given this, what is the prob that event A has occured.

This is an example of conditional probability, and we denote this by $P(A/B)$

clearly we need to consider the event $A \cap B$ (since for A to occur the event $A \cap B$ must occ

Because B has occured, the collection of possible outco is restricted to those outcomes in B.

We view the outcome space as having contracted to those outcomes in B.

Seems reasonable to evaluate $P(A|B)$ by expressing $P(A \cap B)$ as a fraction of $P(B)$

Arising from this discussion we now define the conditional prob of A given B as follows

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{provided } P(B) > 0$$

know dis off

FORMAL: $P(\cdot|B)$ is a Pr. $P(B)>0$: (i) (ii) $P[A^c|B] = \frac{P(A^c \cap B)}{P(B)}$ (iii)

Examples: 1) Pair of dice : score (T)

Suppose we're told that the score is odd

event B

$$P(T \leq 7 \mid T \text{ is odd})$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(3 \text{ or } 5 \text{ or } 7)}{P(3 \text{ or } 5 \text{ or } 7 \text{ or } 9 \text{ or } 11)}$$

2	3	4	5	6	7
3	4	5	6	7	
4	5	6	7		
5	6	7			
6	7				
7					

$$\frac{12/36}{\frac{2+4+6+4+2}{36}}$$

$$= \frac{2}{3} = \frac{12}{18}$$

2) Newspaper A ← 60% of Adults
Newspaper B ← 20% of Adults
7% --- Both.

An adult is randomly chosen we ask if he reads B.

P(read A | read B) : P(A | B)

	B	B^c	
A	.07	.53	.6
A^c	.13	.27	.4
	.2	.8	1

$$\frac{P(A \cap B)}{P(B)} = \frac{.07}{.2} = .35$$

- multiplication LAW of prob

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\Rightarrow \quad P(A \cap B) = P(B)\,P(A|B)$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

$$\Rightarrow \quad P(B \cap A) = P(A) \cdot P(B|A)$$

$$P(B \cap A) = P(A \cap B)$$

$$P(A)\,P(B|A) = P(B) \cdot P(A|B) = P(B \cap A) = P(A \cap B)$$

from example 2).

$$P(\text{Person reads Both } A, B) = P(A \cap B) = P(A|B) \cdot P(B)$$

$$\qquad .07 \qquad\qquad .2 \qquad .35$$

3) Select 2 cards (one after another) from 52

$$P(K_1 \cap Q_2) = P(K_1 | Q_2) \cdot P(Q_2)$$

$$= P(Q_2 | K_1) \cdot P(K_1)$$

- see here we don't have equal se. 2nd one is a useful one - usually 1 useful 1 one not

$$= \left(\frac{4}{51}\right)\left(\frac{4}{52}\right)$$

P(K, Q any order).

$\boxed{K}\ \boxed{Q}$ or $\boxed{Q}\ \boxed{K}$

$$\left(\frac{4}{52}\right)\left(\frac{4}{51}\right) + \left(\frac{4}{52}\right)\left(\frac{4}{51}\right).$$

or

$$\frac{\binom{4}{1}\binom{4}{1}\binom{44}{0}}{\binom{52}{2}}$$

4) 2 child families

$\frac{1}{4}$ BB $\boxed{2}\ \boxed{2}$

$\frac{1}{4}$ BG

$\frac{1}{4}$ GB

$\frac{1}{4}$ GG

Event B: There is a boy in the family

$$P(BB \mid B) = \frac{P(BB \cap B)}{P(B)} \quad \frac{1/4}{3/4} = \frac{1}{3}$$

"If" $\Rightarrow$ $P(A \mid B)$.

5) Disease.

Disease D has a .5% incident rate

Sympton S is associated with D as follows:

- 96% of those with D show S
- 5% of those without D show S

False Positive: $S \cap D^c$ - symptoms but no Disease

good ✓ - positive

False Negative: $S^c \cap D$ disease but no Sym

bad - negative - den just add False

P (person shows S)

$P(S) = P[\text{Symptom \& disease}] \; \cancel{\text{or}} + \; P[\text{Symptom \& no disease}]$

$= P[S \cap D] + P[S \cap D^c]$

True Positive   False Positive

$= P(S|D).P(D) + P(S|D^c).P(D^c)$

$= (.96)(.005) + (.05)(.995)$

– cud one $P(D|S).P(S)$ but this is no use to us (don't know $P(S)$) so don't use dis one. —

An important question is :

$P(D|S)$

opposite

$$P(D|S) = \frac{P(D \cap S)}{P(S)} = \frac{P(S|D).P(D)}{P(S)}$$

-- use dis one cuz other term no use.

$$= \frac{(.96)(.005)}{(.96)(.005) + (.05)(.995)} = .088$$

so this test is not actually dat accurate cuz only 8.8% accurate. [redacted]

opposite

$P(D^c|S) = 1 - .088$

or

$$= \frac{P(D^c \cap S)}{P(S)} = \frac{P(S|D^c).P(D^c)}{P(S)} = \frac{(.05)(.995)}{.05455}$$

opposite

$$P(D^c|S^c) = \frac{P(D^c \cap S^c)}{P(S^c)} = \frac{P(S^c|D^c).P(D^c)}{P(S^c)}$$

$$= \frac{(.95)(.995)}{1 - .05455} = .999788$$

ESTIMATING THE NUMBER OF FISH IN A LAKE

Suppose there are N fish (of some variety) in the lake.
We (somehow) catch N_1 fish, tag them, and then release them.
Later, we catch n fish, and suppose we find that x of them are tagged.

QUESTION: How can we use the values N_1, n, x to estimate N?

The idea here is to choose as the value of N whatever value produces the largest probability for the observed data.
(This approach to estimation is termed the Maximum Likelihood approach).

$$\text{The probability of finding } x \text{ marked fish} = \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}}$$

and we denote this by $L(N)$, where L stands for likelihood.
Our task is to find the value of N for which $L(N)$ is maximum

Consider $\frac{L(N)}{L(N-1)}$. As long as this is >1, the likelihood is increasing.

$$\text{Now } \frac{L(N)}{L(N-1)} = \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}} \Bigg/ \frac{\binom{N_1}{x}\binom{N-1-N_1}{n-x}}{\binom{N-1}{n}}$$

$$= \frac{(N-N_1)}{(N-N_1)-(n-x)} \frac{(N-n)}{N}$$

This is >1 as long as

$$(N-N_1)(N-n) > (N-N_1-n+x)(N)$$

I.E. as long as

$$N_1 n > N x$$

I.E. as long as

$$N < \frac{N_1 n}{x}$$

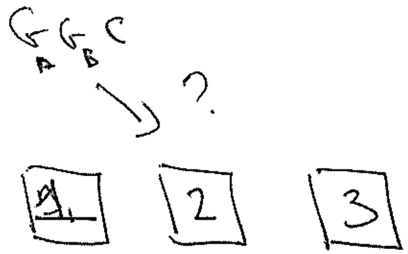
And as soon as $N > \frac{N_1 n}{x}$, then $\frac{L(N)}{L(N-1)} < 1$

This suggests that we should estimate N using

$$\hat{N} = \left[\frac{N_1 n}{x}\right] \leftarrow \text{denotes the integer part.}$$

NOTE: If $\left(\frac{N_1 n}{x}\right)$ is an integer, then $L(\hat{N}) = L(\hat{N}-1)$.

MONTE HALL: ADDITIONAL INFORMATION CAN CHANGE ODDS.

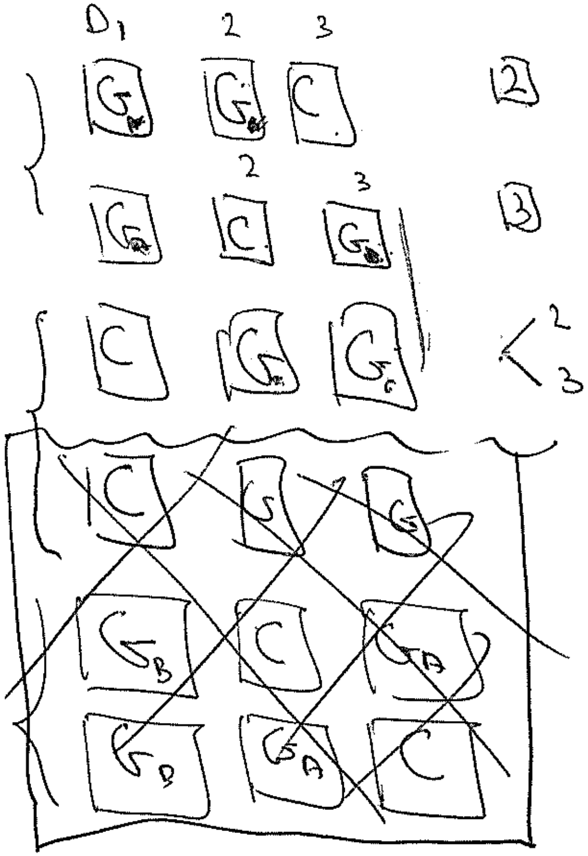


STEP 1 : PICK A DOOR

STEP 2 : OF DOORS NOT PICKED AT LEAST ONE DOES NOT HAVE A CAR. ONE IS REVEALED

- SHOULD PICK BE ALTERED BASED ON STEP2 INFO?

OUTCOMES (DOOR 1 PICKED)



REVEALED $\frac{1}{2}$ .50

$$P[\text{WIN NO CHANGE}] = \frac{2}{6} = \frac{1}{3}$$

$$P[\text{WIN W/CHANGE}] = \frac{4}{6} = \frac{2}{3}$$

~~P[WIN CHANGE AND WITH 2 REVEALED]~~ =

DOUBLE CHANCES OF WINNING.!

~~THIS IS NOT~~ $P[W \text{ with change} \mid D_1]\, P[D_1] + \ldots = 0\tfrac{1}{3} + 1\tfrac{1}{3} + 1\tfrac{1}{3}$

$1\tfrac{1}{3} + 0\tfrac{1}{2} + 0\tfrac{1}{3}$

D_j = CAR BEHIND DOOR j $\quad P[\text{WIN W/OUT} \mid D_1] = \tfrac{1}{3} \quad P[\text{WIN W/OUT} \mid D_2] = 0$

Note: $P(D)$ may be thought of as a prior probability (prior to the test result)

$P(D|S)$ is termed a Posterior probability (if after the test result is known)

We have $P(D) = .005$ || $P(D^c) = .995$

$P(D|S) = .088$ || $P(D^c/S^c) = .999788$

BAYES THEOREM

Here we were using Bayes' Theorem

Based on a simple derivation from conditional Prob.

Theorem: Consider event A and a partition (of Sample Space) $\{H_j\}_{j=1,2,\dots,n}$

Then
$$P(H_k|A) = \frac{P(A|H_k).P(H_k)}{P(A)}$$

$$= \frac{P(A|H_k).P(H_k)}{\sum_{i=1}^{n} P(A|H_i)\,P(H_i)}$$

like $P(S)$ in last example

Proof: i)
$$P(H_k|A) = \frac{P(H_k \cap A)}{P(A)}$$

$$= \frac{P(A|H_k).P(H_k)}{P(A)}$$

ii)
$$P(A) = P(A \cap \Omega)$$

$$= P[A \cap (\bigcup_{i=1}^{n} H_i)]$$

$$= P[\bigcup_{i=1}^{n} (A \cap H_i)]$$

$= \sum_{i}^{n} P(A \cap H_i)$

$P(A) = \sum_{i=1}^{n} P(A|H_i) \cdot P(H_i)$ - Another version of LAW OF TOTAL PROBABILITY

$= \dfrac{P(A|H_k)\ P(H_k)}{\sum_{i=1}^{n} P(A|H_i)\ P(H_i)}$

Example: 10 dice, Take That 4 are crooked
($\to$ 6 with prob .3)

Suppose a dice is picked out (from 10)
tossed, gives a 6
crooked 6 obtained.
Find $P(C/6)$

$P(C/6) = \dfrac{P(C \cap 6)}{P(6)} = \dfrac{P(6/C)\ P(C)}{P(6)}$

* 6 out of 10

Need $P(6) = P(6/C) \cdot P(C) + P(6/\text{Straight})\ P(\text{Straight})$
$= (.3)(.4) + (1/6)(.6)^*$
$.22$
$\Rightarrow P(C/6) = \dfrac{.12}{.22}$

$= .545445$

NOTE:

II $P(A|\underline{B}) = 1 - P(\bar{A}|\underline{B})$

eg. Q 5 Pros set 2.
P(gives correct Ans / Knows) = 1 - P(gives wrong Ans / Knows)
1 - 0 = 1 - (0)
B part must be same.

STATISTICAL INDEPENDANCE

We say event A is independand (not influenced by) of B if

$$P(A|B) = P(A)^*$$

Note:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

=> event A is independant of event B if

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

=> if $P(A \cap B) = P(A).P(B)^*$

Note: This condition is equivilant to

$$\frac{P(A \cap B)}{P(A)} = P(B) \quad [P(A) > 0]$$

$$P(B|A) = P(B)$$

(Thus if A is independent of B, B is independant of A)

Quite often, independance of events A, B is defined as

Definition: Events A, B are independant iff

$$P(A \cap B) = P(A).P(B).$$

Compare this with

$$P(A \cap B) = P(A/B).P(B)$$
$$= P(B/A).P(A)$$

Examples : of independance

Group of 100 students

Arrange into Freq table.

	F	m	
S	24	11	35
NS	26	39	65
	50	50	

Pick at radom a person

Consider $P(S \& F) = P(S \cap F) = \frac{24}{100}$

$P(S) = \frac{35}{100}$

$P(F) = \frac{50}{100}$

We note $P(S \cap F) \neq P(S)\,P(F)$

so smoking is dependant on weather your female

$$P(S|F) = \frac{P(S \cap F)}{P(F)} = \frac{P(F|S)\,.\,P(S)}{P(F)} = \frac{24}{50} = .48$$

$P(S|F) \neq P(S)$ so

smoking and female are dependant events

Suppose :

	F	m
S	21	14
NS	39	26
	60	40

3 Things : i) $P(S|F) = P(S)$ ii) $P(F|S) = P(F)$ iii) $P(S \& F) = P(S)\,P(F)$

Note $P(S \cap F) = \frac{21}{100} = .21$

$P(S) = .35$; $P(F) = .6$

$P(S \cap F) = P(S).P(F)$

so

$P(S/F) = \frac{21}{60} = .35$

$P(S/F) = P(S)$

These demonstrate that events S, F are independant

Note: Usually, we use the concept of independance by reasoning that independance of events is present.
- Then we can use

$P(A \cap B) = P(A).P(B)$
$P(A/B) = P(A)$

Example: Consider 2 horse races
Race 1 : Bet on Horse A (suppose $P(A \text{ wins}) = 1/5$)
Race 2 : Bet on Horse B (suppose $P(B \text{ wins}) = 1/10$)

Accumulator bet: Collect winnings if both win
event $A \cap B$

Want $P(A \cap B) = ? = P(A).P(B)$
If there is independance

$= \frac{1}{5} \cdot \frac{1}{10} = \frac{1}{50}$

Example : Reliability of complex systems.

1) A system with 10 components
reliability of component = P (Stays working in specified T)
= .95

$P(\text{System Fails}) = 1 - P[\text{System works}]$
$1 - P[\text{all 10 components don't fail}]$
$= 1 - P[C_1 \cap C_2 \cap \dots \cap C_{10}]$
If indep. assumed
$= 1 - [P(C_1) . P(C_2) \dots . P(C_{10})]$
$= 1 - (.95)^{10}$
$= .4$

2) A system with 1000 components.
reliability of components = .999

$P(\text{Fails}) = 1 - (.999)^{1000}$
$= .6323$

3) 1000 comp
.9999
$P(\text{Fails}) = 1 - (.9999)^{1000}$
$= .095$

Pairwise Independance

For 2 events, we know that

$$P(A \cap B) = P(A).P(B)$$

is our condition for independance

lets now consider events A, B, C

Suppose we have

$$\left.\begin{aligned} P(A \cap B) &= P(A).P(B) \\ P(A \cap C) &= P(A).P(C) \\ P(C \cap B) &= P(C).P(B) \end{aligned}\right\}$$ If all true events A, B, C are Pairwise independant

Can we say events A, B, C r <u>fully</u> independant

NO!

Suppose we have n events $E_1, E_2, \ldots, E_n$
we define mutual independance for these n events

If $P[E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_m}] = P[E_{i_1}].P[E_{i_2}]. \ldots .P[E_{i_m}]$

for all subsets of size m from the n events
for $m = 2$ or 3 or .... or n

for $n = 3$, we need 1 more condition along with 3 already given:

$$P(A \cap B \cap C) = P(A).P(B).P(C)$$

If all 4 are true, we have full indepence

Pairwise

Example: Consider expt with 4 outcomes

$\Omega \leftrightarrow \{O_1, O_2, O_3, O_4\}$ (equally likely)

lets define some events $E_1 : \{O_1 \overset{or}{,} O_2\}$

$E_2 : \{O_1 \overset{or}{,} O_3\}$

$E_3 : \{O_1 \overset{or}{,} O_4\}$

$P(E_1) = P(E_2) = P(E_3) = \frac{1}{2}$

Also : $E_1 \cap E_2 = O_1$

$E_1 \cap E_3 = O_1$

$E_2 \cap E_3 = O_1$

cob

$E_1 \cap E_2 \cap E_3 = O_1$

$P(E_1 \cap E_2) = P(O_1) = 1/4 = P(E_1).P(E_2)$

$P(E_1 \cap E_3) = 1/4 = P(E_1).P(E_3)$

$P(E_2 \cap E_3) = 1/4 = P(E_2).P(E_3)$

Thus the events E_1, E_2, E_3 pairwise independant

But $P(E_1 \cap E_2 \cap E_3) = P(O_1) = \frac{1}{4}$

$\neq P(E_1).P(E_2).P(E_3)$

So events are not mutually independant.

The Binomial Distribution

Consider a r. expt with 2 outcomes S, F
Suppose this is repeated n times
denote $P(S) = p$
$P(F) = Q = (1-p)$

We assume the n repetitions are mutually independent & that $P(S)$ remains same for all n

We're interested in X, denoting the number of successes in the n trials
(Terminology: Each repitition is called a Bernoulli trial
=> an exp with just 2 outcomes)

Note: X is our first example of a random variable (more ltr)

Possible values for X: $0, 1, 2, \dots, n$

We next try to find $P(X=k)$ [k successes]

- The notation $X=k$ means getting k successes
getting $n-k$ failures.

There are quite a few orderings that produce $(X=k)$
eg. $\underbrace{SSS\dots S}_{k}\ \underbrace{FFF\dots F}_{n-k}$ ← Arrangements !

$$P(X=k) = P(A_1 \cup A_2 \cup \dots \cup A_n \dots)$$
$$= \sum_{i=1} P(A_i)$$

Lets find $P(A_1)$

$$P(A_1) = P(SSS\dots SFFF\dots F)$$
$$= \underbrace{P(S)\,P(S)\dots P(S)}_{k}\ \underbrace{P(F)\,P(F)\dots P(F)}_{n-k} \quad \dots \text{ since independance}$$

$$= [P^k(S)]^k\,[P(F)]^{n-k}$$

$$= P^k (Q)^{n-k} = P^k [1-P]^{n-k}$$

Similarly:

$P(A_2) = P^k Q^{n-k}$ cuz S occurs k times
F ocurs $n-k$ times

For all other A:

$$P(x=k) = \sum P^k Q^{n-k}$$

How many Arrangements? $\binom{n}{k}$

:ob

=>

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

$\sum_k P_k = 1$

Connection wit binomial in algebra

$$(Q+P)^n = \sum_{k=0}^{n} \binom{n}{k} P^k Q^{n-k}$$

$\sum_k CP$

$\sum_k \binom{n}{k} p^k (1-p)^{n-k} \frac{1}{1-1} = 1$

Probability Distribution: This specification of probability values is called Prob. dist for a random variable X.

Example: Manufacturing process
Sample of n items each hour
test all n, count # failure (x)

Rule: Say $n = 20$
If $x \leq 3$ carry on
If $x > 3$ stop

We suppose various values for P (Prob of failure)

$P = .05$ From tables $P(X \geq 4) \Rightarrow P(X > 3) = .0159$
$P = .1$ $P(X \geq 4) = .1330$

Example of binomial:
SPC suppose of size 20
Test all 20 (some acceptable
some not

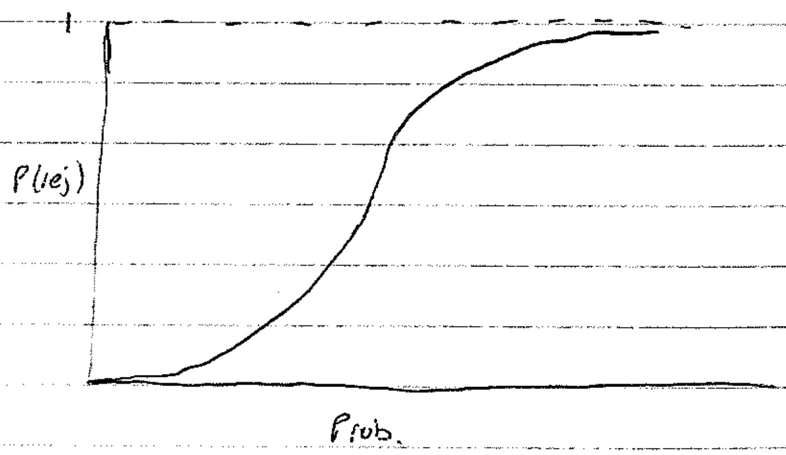
X = # of ~~good~~ bad. items

Model: 20 Bernoulli trials (each with P(success) = P ← defective items
Assuming independence and that P remains const.
Then X is Binomial $(b(n,p))$

Suppose our decision rule: OK if $X \leq 3$

Draw the operating characteristic
↓
Plot of P(rejection) against value of P.

P	Prob $(X \geq 4)$
.05	.0159
.1	.133
.15	.3523
.2	.5886
.25	.7748



Random Variables

Design matter. Find n, to give a "good shape" to curve.

A random variable is a variable whose value is determined (to some extent) by chance

or

A random variable is a real-valued function defined on a sample space

ob Example : Game of chance :

Pay €1 to enter.

Die is throw, get €1 back if 1

get €5 back if 6

2, 3, 4, 5 , no pay

Sample space : 1 2 3 4 5 6

Determine the net gain from 1 play.

1 → net gain = 0

2, 3, 4, 5 → net gain = −1

6 → net gain = 4

G	-1	0	4
Prob	$4/6$	$1/6$	$1/6$

– An example of a discrete probability distribution

Discrete → value of $G \in \mathbb{Z}$

Note : A probability distribution is a specification of how probabilities are distributed over all the possible values of a random variable.

We start with discrete random variables

(and the associated distributions are called discrete distributions.

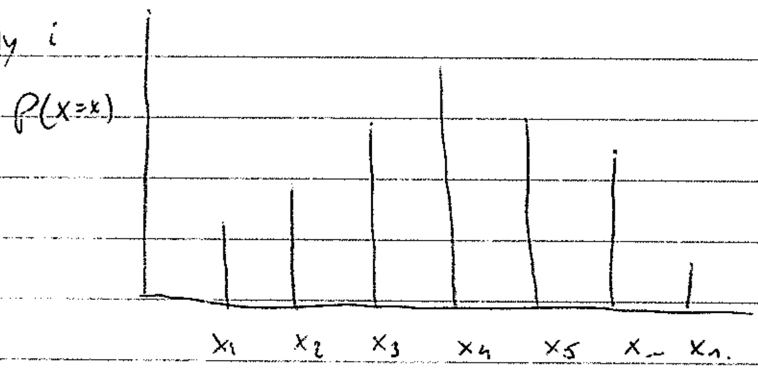
Terminology:

1) Probability functions: This is the particular specification of a discrete prob distribution in which we provide $P(X = x_i)$ for all possible values of the random variable X

X	x_1	x_2	x_3	---	x_n
Prob	p_1	p_2	p_3	---	p_n

← The probability function for x

Graphically:



eg Binomial:

X	0	1	----	r	
Prob	$(1-p)^n$	$np(1-p)^{n-1}$		$\binom{n}{r} p^r (1-p)^{n-r}$	

NOTE:

1) $p_i \geq 0$

2) $\sum p_i = 1$

Next we look at a series of independant random variables :

1) <u>Bernoulli random variable</u> (X)

Associated with a bernoulli Trial $X = 1$, $X = 0$

X	0	1
Prob	$1-p$	p

2) Binomial Random Variable (X)

we have n bernoulli trials

a) 2 outcomes S, F : Denote $P(S)$ by p

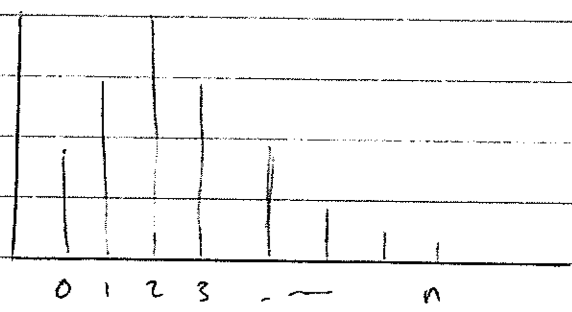
b) Trials are independant

c) p remains constant

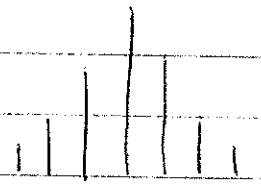
$X = \#$ Successes

X is $b(n, p)$

Shape depends on p



If $P = \frac{1}{2}$



3) Geometric Random Variable

Sequence of Bernoulli Trials

let x = # trials up to 1^{st} success

values for $x = 1, 2, 3, \ldots, n, \ldots, \infty$

$$P(x=k) = P[\underbrace{FFF\ldots F}_{k-1}\ S]$$

$$= (1-P)^{k-1} \cdot P$$

Easy to check $\sum_{k=1}^{\infty} P(x=k) = 1$

Prob set 3 Q.2. using binomial

$P(x=1) \rightarrow n(.5)^n$

: Prob of getting one head in n trials

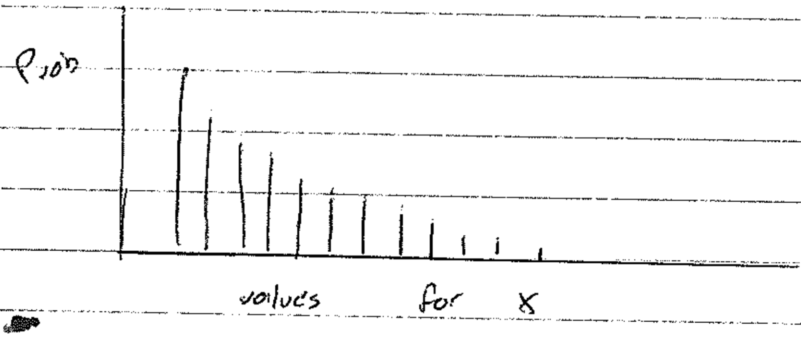
could get it on the first try $2^{nd}, \ldots, n^{th}$.

using geometric:

$P(x=n) = (.5)^n$

: Prob of getting one head in n trials

but getting it on n^{th} try only.



4) Negative Binomial Random Variable (Pascal)

Bernoulli sequence of trials

Y = # trials up to r^{th} success

ie like how many trials till 4^{th} success.

Values : $r, r+1, r+2, \ldots, \infty$

if $r=1$ Neg Binomial $\rightarrow$ Geometric.

$P(Y=k)$ = Prob [$(r-1)$ successes in 1^{st} $(k-1)$ trials $\underline{\cap}$ success on k^{th} trial]

cuz independent :-

$= P[(r-1) \text{ success in } (k-1) \text{ trials}] \cdot P[\text{success in } k^{th} \text{ trial}]$

$$\binom{k-1}{r-1} p^{r-1} \cdot (1-p)^{(k-1)-(r-1)} \cdot p$$

$$\binom{k-1}{r-1} p^{r} (1-p)^{k-r}$$

Note: compare Ans with $X \sim b(n,p)$
for what $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

If $r=1$ we are back to geometric.

Negative Binomial arises from:

Note: $$\binom{-r}{j} = \frac{(-r)(-r-1)\dots(-r-j+1)}{j!} = \frac{(-1)^j r(r+1)\dots(r+j-1)}{j!}$$

$$= \frac{(-1)^j (r+j-1)(r+j-2)\dots(r+1)(r)}{j!}$$

we need it cuz: $(1+x)^{-r} = \sum_{j=0}^{\infty} \binom{-r}{j} x^j$

$$(**) \quad = (-1)^j \binom{r+j-1}{j} = (-1)^j \binom{r+j-1}{r-1}$$

Name comes from:

$$\sum_{k=r}^{\infty} P(Y=k) = \sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r (1-p)^{k-r} = 1$$

$$p^r \sum_{k=r}^{\infty} \binom{k-1}{r-1} (1-p)^{k-r}$$

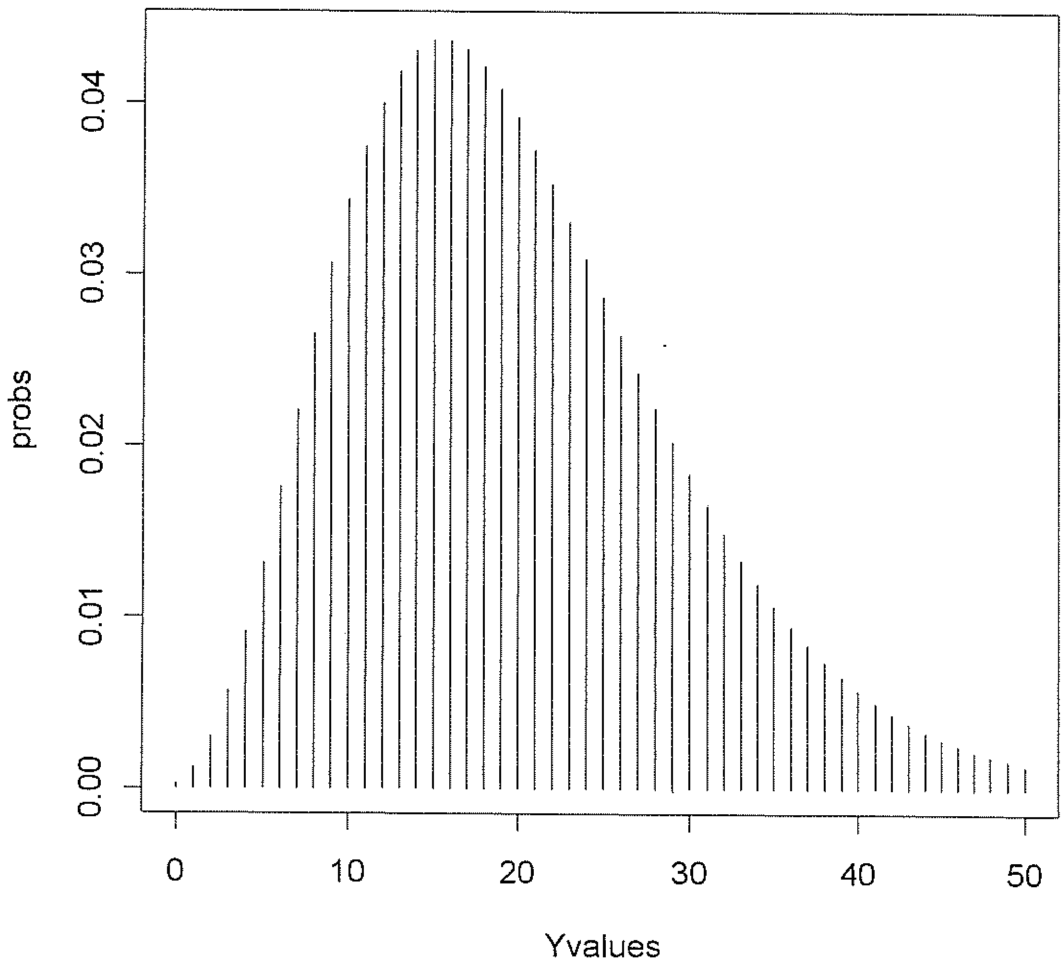
change index : let $j = k-r$ [as k runs from $r \to \infty$, j runs from $0 \to \infty$

PLOTTING THE SHAPE OF THE NEGATIVE BINOMIAL DISTRIB.


```
R : Copyright 2004, The R Foundation for Statistical Computing
Version 2.0.1  (2004-11-15), ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]
```

START

```
> Yvalues=0:50
> probs=dnbinom(Yvalues,5,0.2)
> probs
 [1] 0.000320000 0.001280000 0.003072000 0.005734400 0.009175040 0.013212058
 [7] 0.017616077 0.022145925 0.026575110 0.030709016 0.034394098 0.037520834
[13] 0.040022223 0.041869403 0.043065671 0.043639880 0.043639880 0.043126470
[19] 0.042168104 0.040836480 0.039203021 0.037336210 0.035299689 0.033151013
[25] 0.030940945 0.028713197 0.026504490 0.024344865 0.022258162 0.020262603
[31] 0.018371426 0.016593546 0.014934192 0.013395517 0.011977169 0.010676790
[37] 0.009490480 0.008413182 0.007439024 0.006561601 0.005774209 0.005070037
[43] 0.004442318 0.003884446 0.003390062 0.002953120 0.002567931 0.002229182
[49] 0.001931958 0.001671735 0.001444379
>
> sum(probs)
[1] 0.9913492
> plot(Yvalues,probs,"h")
```



ST2054

R Help on 'dnbinom'

help(dnbinom)

```
NegBinomial              package:stats              R Documentation

The Negative Binomial Distribution

Description:

     Density, distribution function, quantile function and random
     generation for the negative binomial distribution with parameters
     'size' and 'prob'.

Usage:

     dnbinom(x, size, prob, mu, log = FALSE)
     pnbinom(q, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
     qnbinom(p, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
     rnbinom(n, size, prob, mu)

Arguments:

       x: vector of (non-negative integer) quantiles.

       q: vector of quantiles.

       p: vector of probabilities.

       n: number of observations. If 'length(n) > 1', the length is
          taken to be the number required.

    size: target for number of successful trials, or dispersion
          parameter (the shape parameter of the gamma mixing
          distribution).

    prob: probability of success in each trial.

      mu: alternative parametrization via mean: see Details

log, log.p: logical; if TRUE, probabilities p are given as log(p).

lower.tail: logical; if TRUE (default), probabilities are P[X <= x],
          otherwise, P[X > x].

Details:
```

I HAVE OMITTED SOME (CURRENTLY LESS RELEVANT) DETAILS

```
     If an element of 'x' is not integer, the result of 'dnbinom' is
     zero, with a warning.

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

Value:

     'dnbinom' gives the density, 'pnbinom' gives the distribution
     function, 'qnbinom' gives the quantile function, and 'rnbinom'
     generates random deviates.

See Also:

     'dbinom' for the binomial, 'dpois' for the Poisson and 'dgeom' for
     the geometric distribution, which is a special case of the
     negative binomial.
```

$$= p^r \sum_{k=r}^{\infty} = p^r \sum_{j=0}^{\infty} \binom{k-1}{j} (1-p)^j \qquad (*)$$

Since $j = k-r$, then $k = j+r$

$$(*) \Rightarrow p^r \sum_{j=0}^{\infty} \binom{j+r-1}{j} (1-p)^j$$

$$= p^r \sum_{j=0}^{\infty} (-1)^j \binom{-r}{j} (1-p)^j \qquad \text{using } (**)$$

$$p^r [1-(1-p)]^{-r}$$

$$= p^r [p]^{-r} = 1 \quad \checkmark$$

what does it look like?
see sheet.

CONNECTION BETWEEN NEG Binomial & (Ord) Binomial

Sequence of Bernoulli Trials

- Think of 1st n trials

let X_n = # successes in 1st n trials

Y_r = # Trials needed to get to r^{th} success

Consider event $(X < r)$ it must be that $(Y > n)$

$\therefore$ $(X < r) \Rightarrow Y > n$

Consider event $(Y > n)$. If $Y > n$ occurs then $(X < r)$

CONNECTION WITH BINOMIAL

$X_n =$ # of successes in n trials

$Y_r =$ # of trials to get r successes.

EVENTS

$E = [X_n \leq r]$ = r or fewer successes in n trials.

$F = [Y_r \geq n]$ = r or fewer successes in n trials.

$E = F$

$$P(E) = \sum_{k=0}^{r} \binom{n}{k} p^r q^{n-r}$$

~~P(F)~~

$$P(F) = \sum_{k=n}^{\infty} \binom{k-1}{r-1} p^r q^{k-r}$$

E easier to calculate

Example

Thus the event $(X < r)$ and $(Y > n)$ are essentially the same.

Then $P(X < r) = P(Y > n)$

Alternatively: $P(X \geq r) = P(Y \leq n)$
↓
In binomial tables

We could use these to determine the neg binomials

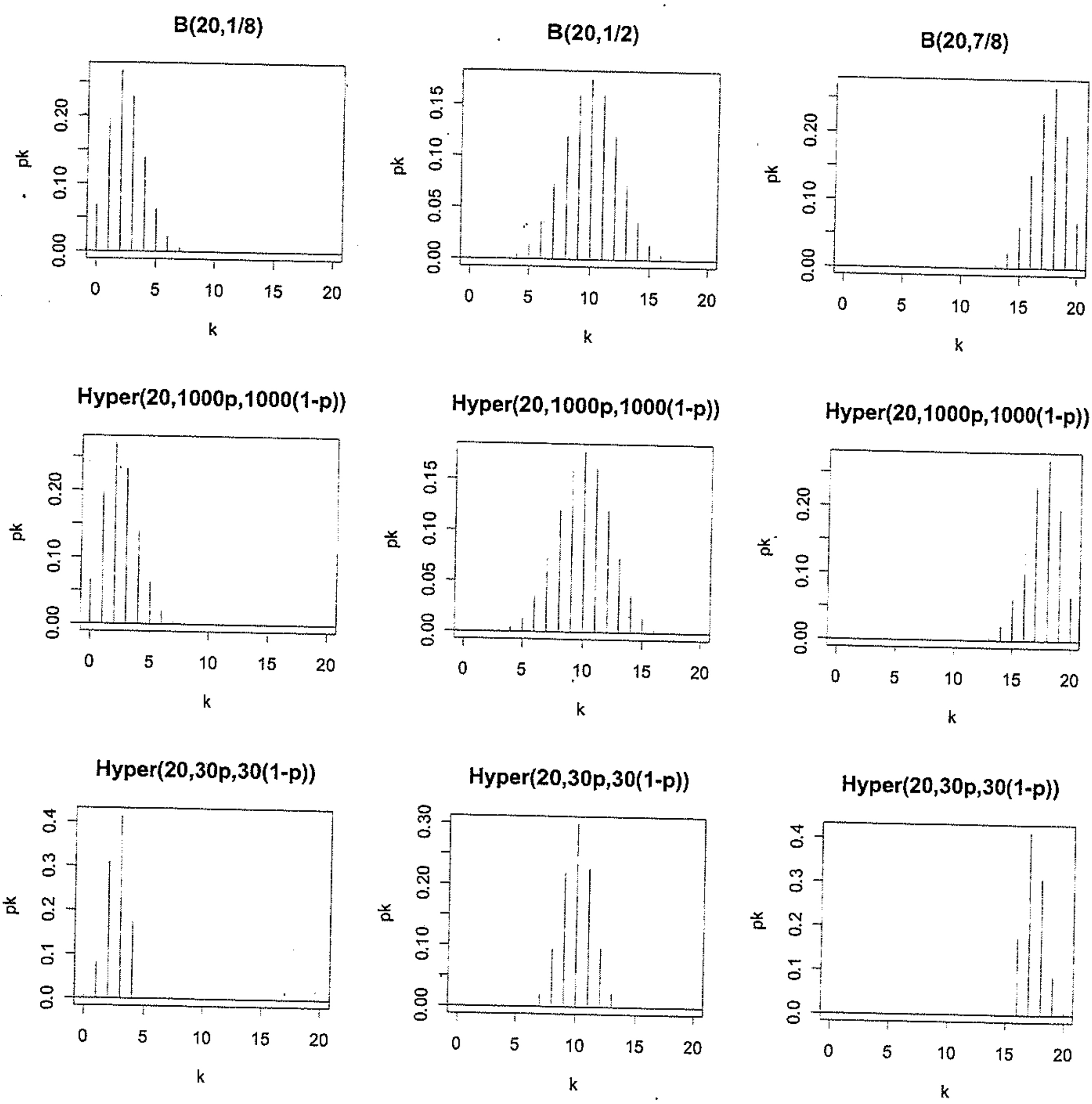
Example: Seq of bernoulli Trials with $p = .2$
Find $P[Y > 20]$ for $r = 5$.

$$P(Y_5 > 20) = 1 - P(Y_5 \leq 20) = 1 - P(X_{20} \geq 5)$$

$\text{bi}(20, .2)$

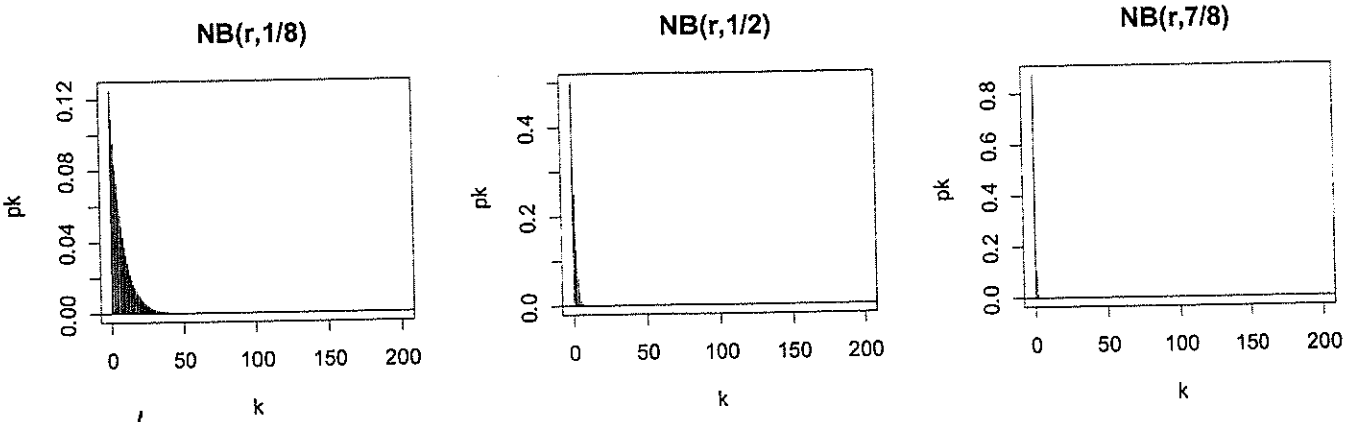
$$1 - .3704$$
$$.6296$$

Binomial and Hypergeometric.

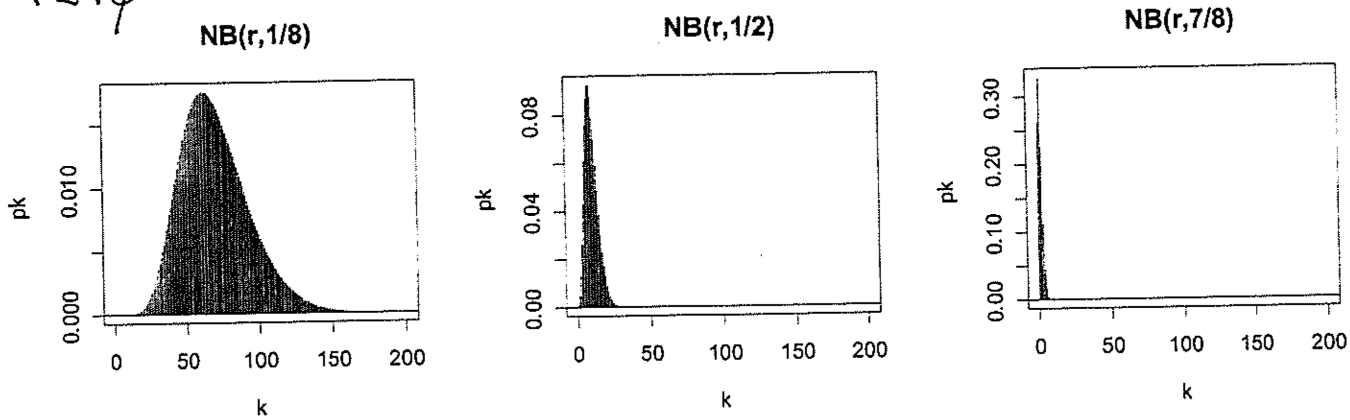


NEGATIVE BINOMIALS

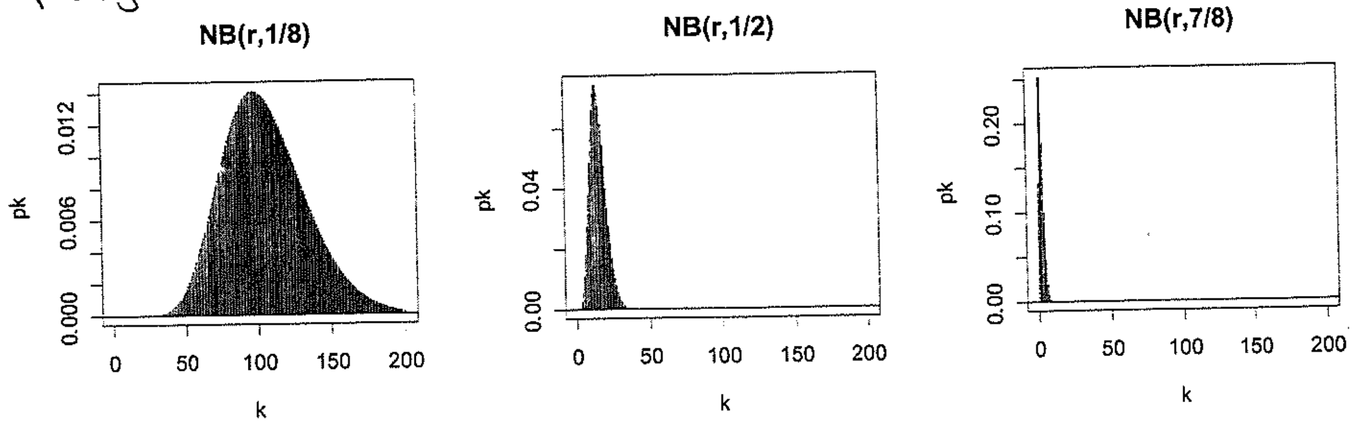
r = 1



r = 10



r = 15



~~Binomials~~

```
# ST2054 Binomial/Geometric/Negative Binomials

# 1. Compare Binomial and Hypergeometric

par(mfrow=c(3,3))
n=20 ; k=c(0:n)
p=.125;pk = dbinom(k, n, p);plot(k,pk,type="h",main="B(20,1/8)");abline(h=0)
p=.5;pk = dbinom(k, n, p);plot(k,pk,type="h",main="B(20,1/2)");abline(h=0)
p=.875;pk = dbinom(k, n, p);plot(k,pk,type="h",main="B(20,7/8)");abline(h=0)

N=1000
p=.125;N1=round(N*p);pk = dhyper(k,
N1,N-N1,n);plot(k,pk,type="h",main="Hyper(20,1000p,1000(1-p))");abline(h=0)
p=.5;N1=round(N*p);pk = dhyper(k,
N1,N-N1,n);plot(k,pk,type="h",main="Hyper(20,1000p,1000(1-p))");abline(h=0)
p=.875;N1=round(N*p);pk = dhyper(k,
N1,N-N1,n);plot(k,pk,type="h",main="Hyper(20,1000p,1000(1-p))");abline(h=0)

N=30
p=.125;N1=round(N*p);pk = dhyper(k,
N1,N-N1,n);plot(k,pk,type="h",main="Hyper(20,30p,30(1-p))");abline(h=0)
p=.5;N1=round(N*p);pk = dhyper(k,
N1,N-N1,n);plot(k,pk,type="h",main="Hyper(20,30p,30(1-p))");abline(h=0)
p=.875;N1=round(N*p);pk = dhyper(k,
N1,N-N1,n);plot(k,pk,type="h",main="Hyper(20,30p,30(1-p))");abline(h=0)


# 2. Negative Binomials

r=1    # Geomertric
n=200 ; k=c(0:n)
p=.125;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,1/8)");abline(h=0)
p=.5;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,1/2)");abline(h=0)
p=.875;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,7/8)");abline(h=0)

r=10
n=200 ; k=c(0:n)
p=.125;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,1/8)");abline(h=0)
p=.5;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,1/2)");abline(h=0)
p=.875;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,7/8)");abline(h=0)


r=15
n=200 ; k=c(0:n)
p=.125;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,1/8)");abline(h=0)
p=.5;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,1/2)");abline(h=0)
p=.875;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,7/8)");abline(h=0)
```

NegBinomial package:base R Documentation

The Negative Binomial Distribution

Description:

Density, distribution function, quantile function and random generation for the negative binomial distribution with parameters 'size' and 'prob'.

Usage:

```
dnbinom(x, size, prob, mu, log = FALSE)
pnbinom(q, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
qnbinom(p, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
rnbinom(n, size, prob, mu)
```

Arguments:

x: vector of (non-negative integer) quantiles.

q: vector of quantiles.

p: vector of probabilities.

n: number of observations. If 'length(n) > 1', the length is taken to be the number required.

size: target for number of successful trials, or dispersion parameter (the shape parameter of the gamma mixing distribution).

prob: probability of success in each trial.

mu: alternative parametrization via mean: see Details

log, log.p: logical; if TRUE, probabilities p are given as log(p).

lower.tail: logical; if TRUE (default), probabilities are P[X <= x], otherwise, P[X > x].

Details:

The negative binomial distribution with 'size' = n and 'prob' = p has density

```
p(x) = Gamma(x+n)/(Gamma(n) x!) p^n (1-p)^x
```

for x = 0, 1, 2, ...

This represents the number of failures which occur in a sequence of Bernoulli trials before a target number of successes is reached.

A negative binomial distribution can arise as a mixture of Poisson distributions with mean distributed as a gamma ('pgamma') distribution with scale parameter '(1 - prob)/prob' and shape parameter 'size'. (This definition allows non-integer values of 'size'.) In this model 'prob' = 'scale/(1+scale)', and the mean is 'size * (1 - prob)/prob'

The alternative parametrization (often used in ecology) is by the _mean_ 'mu', and 'size', the _dispersion parameter_, where 'prob' = 'size/(size+mu)'. In this parametrization the variance is 'mu + mu^2/size'.

If an element of 'x' is not integer, the result of 'dnbinom' is zero, with a warning.

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

$n = 100$
$p = .0278$

k	Binomial Probability	Poisson Approximation ← $m = 2.78$
0	.0596	.0620
1	.1705	.1725
2	.2414	.2397
3	.2255	.2221
4	.1564	.1544
5	.0858	.0858
6	.0389	.0398
7	.0149	.0158
8	.0050	.0055
9	.0015	.0017
10	.0004	.0005
11	.0001	.0001

The approximation is quite good. ■

The Poisson frequency function can be used to approximate binomial probabilities for large n and small p. This suggests how Poisson distributions can arise in practice. Suppose that X is a random variable that equals the number of times some event occurs in a given interval of time. Heuristically, let us think of dividing the interval into a very large number of small subintervals of equal length, and let us assume that the subintervals are so small that the probability of more than one event in a subinterval is negligible relative to the probability of one event, which is itself very small. Let us also assume that the probability of an event is the same in each subinterval and that whether an event occurs in one subinterval is independent of what happens in the other subintervals. The random variable X is thus nearly a binomial random variable, with the subintervals consitituting the trials, and, from the limiting result above, X has nearly a Poisson distribution.

The preceding argument is not formal, of course, but merely suggestive. But, in fact, it can be made rigorous. The important assumptions underlying it are (1) what happens in one subinterval is independent of what happens in any other subinterval, (2) the probability of an event is the same in each subinterval, and (3) events do not happen simultaneously. The same kind of argument can be made if we are concerned with an area or a volume of space rather than with an interval on the real line.

The Poisson distribution is of fundamental theoretical and practical importance. It has been used in many areas, including the following:

- The Poisson distribution has been used in the analysis of telephone systems. The number of calls coming into an exchange during a unit of time might be modeled as a Poisson variable if the exchange services a large number of customers who act more or less independently.
- One of the earliest uses of the Poisson distribution was to model the number of alpha particles emitted from a radioactive source during a given period of time.
- The Poisson distribution has been used as a model by insurance companies. For example. the number of freak acidents, such as falls in the shower. for a large

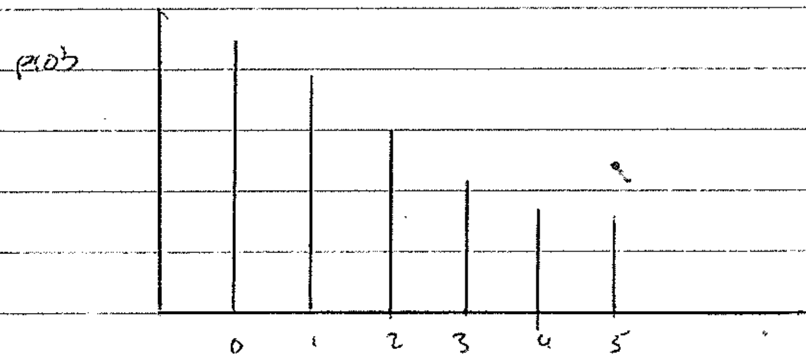
TUESDAY 1-2 Emma Preece 40

Poisson Probability Distribution

X = # occasions of a rare event

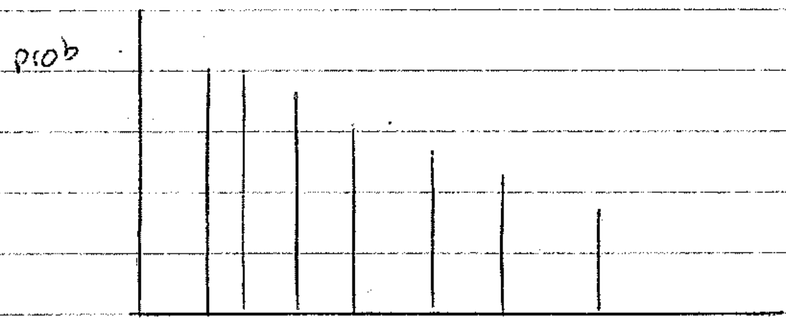
We'll see that $P(X=k) = \frac{m^k e^{-m}}{k!}$ $\left(\frac{\lambda^k e^{-\lambda}}{k}\right)$

(m is the parameter).

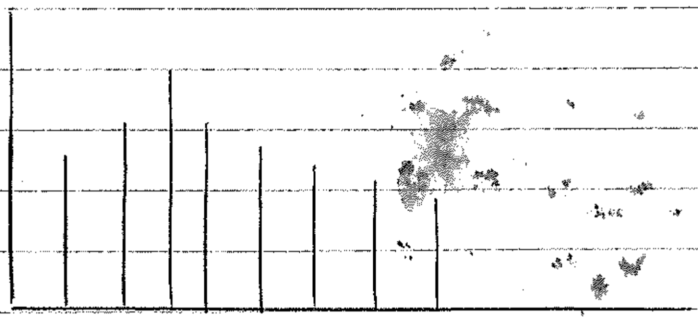


depends on m

$m < 1$



$m = 1$



$m > 1$

4/

λ has an interpretation as Average # of occurances per unit time.

$\lambda t = m$

$$P(x=k) = \frac{m^k e^{-m}}{k!}$$

Example: Find $P(x \geq 5)$ with $m = 5$.

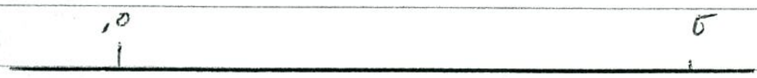
$= .559$

$\cdot 1088 = 1 - P(X<5) = 1 - (P[X=0] + P(X=1) + \ldots P(X=4))$

<u>Poission Distribution is an approximation of Binomial dist.</u>

In certain circumstances we can approximate the Binomial dist. using the poisson Distribution.

consider poission again.



Divide into n short time invertals each t/n.

Assume we have a Poisson Distribution $\frac{\lambda t}{n}$

For each short time interval, we can get 1 occurance will probability $\sim \lambda h \to \frac{\lambda t}{n}$

and prob of >1 occurance is negligable.

Thus we have n short intervals for each of which we get 1 occurance with prob $\sim \lambda \left(\frac{t}{n}\right)$

or 0 occurances with prob $\sim 1 - \lambda \left(\frac{t}{n}\right)$

Cur dey r non-overlapping intervals the # of occurances are all indep.

Each short time interval corresponds to a bernoulli trial

X_t = # occurances in the entire interval $(0, t)$

We get $(X_t = k)$ when k of the n short intervals results in success (ie an occurance)

Thus:

$$P(X_t = k) = P[\,k \text{ successes over the } n \text{ short intervals}\,]$$

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

Now let n get larger $(\to \infty)$ and the approximate of the poisson Process (by this collection of n short intervals) improves:

We want to find

$$\lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k}$$

$$\lim_{n\to\infty} \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} \frac{(\lambda t)^k}{n^k} \frac{(1 - \lambda t/n)^n}{(1 - \lambda t/n)^k}$$

switch

$$= \lim_{n\to\infty} \left[\frac{n(n-1)\ldots(n-k+1)}{\underbrace{(n)(n)\ldots(n)}_{k \text{ times}}} \right] \frac{(\lambda t)^k}{k!} \frac{\left(1 - \frac{\lambda t}{n}\right)^n}{\left(1 - \frac{\lambda t}{n}\right)^k}$$

$$\frac{(\lambda t)^k}{k!} \lim_{n\to\infty} \left[\quad \right] \lim_{n\to\infty} \frac{\left(1 - \frac{\lambda t}{n}\right)^n}{\left(1 - \frac{\lambda t}{n}\right)^k}$$

$$\frac{(\lambda t)^k}{k!} \; [1] \; \frac{\lim_{n\to\infty} \left(1 - \frac{\lambda t}{n}\right)^n}{\lim_{n\to\infty} \left(1-\frac{\lambda t}{n}\right)^k \to 1} \quad --- \quad \text{note: } \lim_{n\to\infty} \left(1+\frac{x}{n}\right)^n = e^x$$

$$= P(X_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

so:

$$\lim_{n\to\infty} \binom{n}{k} \left(1 - \frac{\lambda t}{n}\right)^{n-k} \left(\frac{\lambda t}{n}\right)^k = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad (*)$$

$$\frac{\lambda t}{n} = p \Rightarrow \lambda t = np.$$

Since λt is fixed then as $n \to \infty$ must have $p \to 0$ (in such a way that np remains fixed)

Thus $(*)$ equation means that

$$\lim_{\substack{n\to\infty \\ (np \text{ fixed})}} \binom{n}{k} p^k (1-p)^{n-k} = \frac{(np)^k e^{-np}}{k!}$$

Consequence of this:

When n is large

p is small

we can use a poisson prob function to approximate Binomial prob function

rule: $n \geq 100$ $\quad p \leq .01$ $\quad np \leq 20$

population of people in a given time period might be modeled as a Poisson distribution, since the accidents would presumably be rare and independent (provided there was only one person in the shower.)

- The Poisson distribution has been used by traffic engineers as a model for light traffic. The number of vehicles that pass a marker on a roadway during a unit of time can be counted. If traffic is light, the individual vehicles act independently of each other. In heavy traffic, however, one vehicle's movement may influence another's, so the approximation might not be good.

EXAMPLE B This amusing classical example is from von Bortkiewicz (1898). The number of fatalities that resulted from being kicked by a horse was recorded for 10 corps of Prussian cavalry over a period of 20 years, giving 200 corps-years worth of data. These data and the probabilities from a Poisson model with $\lambda = .61$ are displayed in the following table. The first column of the table gives the number of deaths per year, ranging from 0 to 4. The second column lists how many times that number of deaths was observed. Thus, for example, in 65 of the 200 corps-years, there was one death. In the third column of the table, the observed numbers are converted to relative frequencies by dividing them by 200. The fourth column of the table gives Poisson probabilities with the parameter $\lambda = .61$. In chapters 8 and 9 we discuss how to choose a parameter value to fit a theoretical probability model to observed frequencies and methods for testing goodness of fit. For now we will just remark that the value $\lambda = .61$ was chosen to match the average number of deaths per year.

hardly

Number of Deaths per Year	*Observed*	*Relative Frequency*	*Poisson Probability*
0	109	.545	.543
1	65	.325	.331
2	22	.110	.101
3	3	.015	.021
4	4	.005	.003

■

The Poisson distribution often arises from a model called a **Poisson process** for the distribution of random events in a set S, which is typically one-, two-, or three-dimensional, corresponding to time, a plane, or a volume of space. Basically, this model states that if $S_1, S_2, \ldots, S_n$ are disjoint subsets of S, then the numbers of events in these subsets, $N_1, N_2, \ldots, N_n$, are independent random variables that follow Poisson distributions with parameters $\lambda|S_1|, \lambda|S_2|, \ldots, \lambda|S_n|$, where $|S_i|$ denotes the measure of S_i (length, area, or volume, for example). The crucial assumptions here are that events in disjoint subsets are independent of each other and that the Poisson parameter for a subset is proportional to the subset's size. Later, we will see that this latter assumption implies that the average number of events in a subset is proportional to its size.

EXAMPLE C Suppose that an office receives telephone calls as a Poisson process with $\lambda = .5$ per min. The number of calls in a 5-min. interval follows a Poisson distribution with

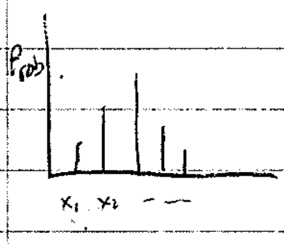
(Cumulative) Distribution Function

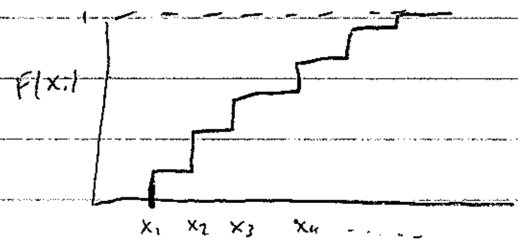
This is $P[X \leq x]$

For x discrete $(\in \mathbb{Z}^+)$, CDF look

Prob function: we specify $P(X = x_i)$ for each possible x_i

Distribution function: we specify $P(X \leq x_i)$ for each possible x_i "$F(x)$"





Note: Tables of binomial & poisson give $1 - F(x_{i-1})$ for each x_i

Continuous Random variable.

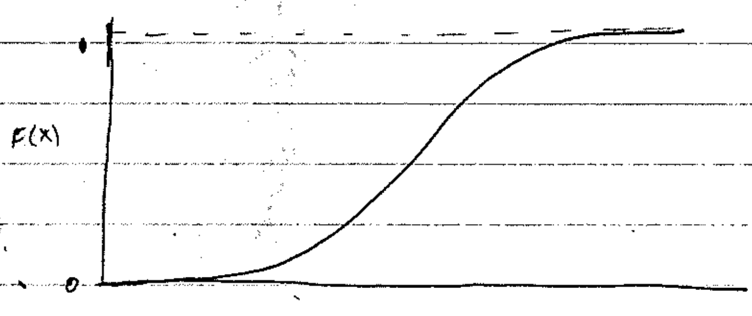
Here the values of X are in some continuous range

eg. 0 to 1

or $0 \to \infty$ Continuous → can take any value

or $-\infty \to +\infty$ Discrete → only values $\in \mathbb{N}$

Shape of the $F(x)$ is quite different



$F(x)$ is continuous.

Our interest is in D.F's (dist Functions) which are differentiable everywhere

Prob. Desity Functions

(except for a finite # of points in any finite interval)

Denote the derivative of $F(x)$ by $f(x)$

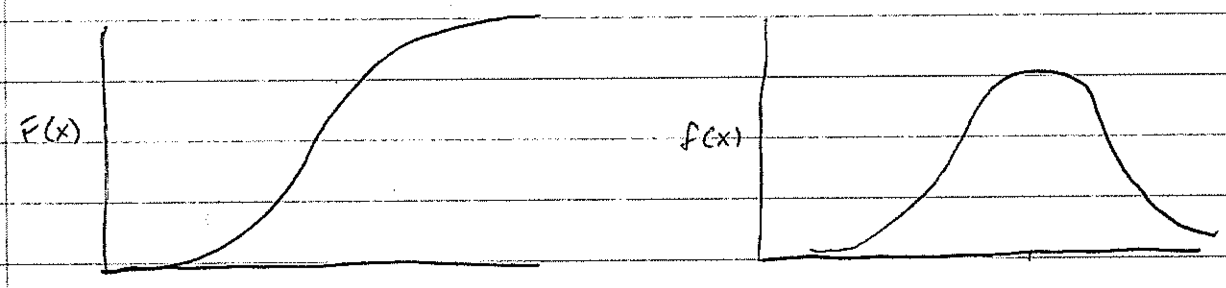
$f(x)$ is called

Probability Desity Function (pdf)

of random variable x or "distribution"

$$f(x) = F'(x)$$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

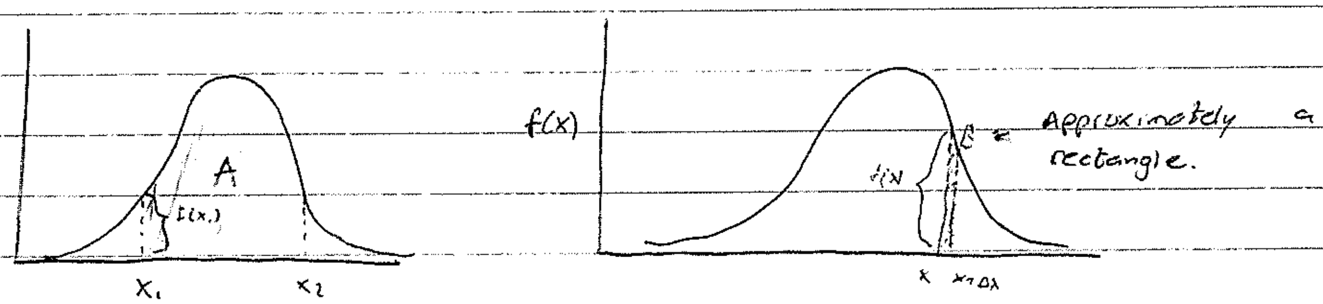


Why "density" in density function?

Consider $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$

$$= \int_{x_1}^{x_2} f(t)\,dt \quad - \underline{\text{Area } A}.$$

$$\left.\begin{array}{l} x_1 \to x \\ x_2 \to x + \Delta x \end{array}\right\} \quad P(x < X \leq x + \Delta x) = \int_{x}^{x+\Delta x} f(t)\,dt \quad - \underline{\text{Area } B}$$



$F(x_1)^*$ $\qquad$ $f(x)^*$

$$P[x < X \leq x + \Delta x] \simeq f(x)\,\Delta x$$

Approximation improves for smaller Δx

$\frac{dF(x)}{dx} = \text{P.D.F}$

$\int_{-\infty}^{y} f(y)\,dy = P(Y \le y)$
$= F(Y)$

pdf $f(x)$

(cdf)(DF) $F(x) = \int_{-\infty}^{x} f(u)\,du$

$P(X \le x)$

AREA $= F(x)$

519.2 Mull - Concepts of prob.

Welcome to business class

519.2 Mill - Problems in prob.

519 Sves - Problems in prob theory

519.2 Lars.

$I_{[Y \leq 4]}$

$P[Z \leq z \mid Y \leq 4]$

$\min(Y, 4)$

$P[Z \leq z]$

$= P[\min(Y,4) < z]$ $\quad z \geq 4$

1

$=$

Differentiate $F(t)$ to get PDF for L

$\Rightarrow f(t) = \lambda e^{-\lambda t}$

(we recognize as an Negative Exp pdf)

☆ This pdf has the unusual property referred to as "<u>No memory</u>" (Think of coin (Heads / tails)) if throw is heads. P(tails) still = 1/2

$$P\left(L > t+s \mid L > s\right) = P\left[L > t\right]$$

ie (magnitude of S has no effects on the prob)

$$\frac{P\left[(L > t+s) \cap (L > s)\right]}{P(L > s)}$$

$$= \frac{P\left(L > t+s\right]}{P(L > s)}$$

$$= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} \rightarrow P(L > t)$$

3) $X \sim \text{Gamma}(\alpha, \beta)$

$$f(x) = \frac{1}{\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} \frac{1}{\beta}$$

Shape: if $\alpha = 1 \rightarrow Ne(1/\beta)$ $\quad f(x) = \frac{1}{\beta} e^{-x/\beta}$

what about $\Gamma(\alpha) \overset{?}{=} (\alpha-1)!$

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$$

To show $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

$$\Gamma\left(\tfrac{1}{2}\right) = \int_0^\infty x^{-\frac{1}{2}} e^{-x}\, dx$$

$$= 2\int_0^\infty e^{-u^2}\, du \qquad u^2 = x, \quad 2u\,du = dx$$

$$\Gamma^2\left(\tfrac{1}{2}\right) = 4\left(\int_0^\infty e^{-u^2} du\right)\left(\int_0^\infty e^{-v^2} dv\right)$$

$$= 4\int_0^\infty \int_0^\infty e^{-(u^2+v^2)}\, du\, dv \qquad u = r\cos\theta, \quad v = r\sin\theta$$

$$= 4\int_0^\infty \int_0^{\frac{\pi}{2}} e^{-r^2}\, r\, dr\, d\theta \qquad u^2+v^2 = r^2, \quad du\,dv \longrightarrow r\,dr\,d\theta$$

$$= 4\left(\int_0^\infty e^{-r^2} r\, dr\right)\left(\int_0^{\frac{\pi}{2}} d\theta\right)$$

$$= 4\left(\tfrac{1}{2}\right)\left(\tfrac{\pi}{2}\right)$$

$$\Gamma^2\left(\tfrac{1}{2}\right) = \pi$$

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}$$

Notice $\Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right) = \frac{3}{4}\sqrt{\pi}$

DISTRIB.

Figure 2-11 shows several gamma densities. Gamma densities provide a fairly flexible class for modeling nonnegative random variables.

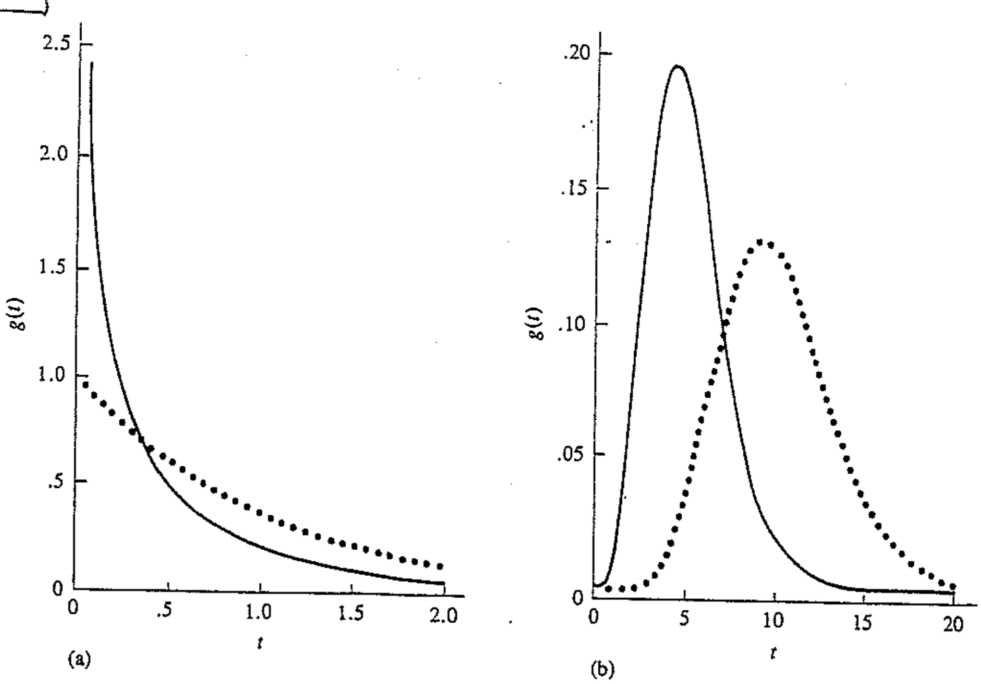


Figure 2-11. Gamma densities, (a) $\alpha = .5$ (solid) and $\alpha = 1$ (dotted) and (b) $\alpha = 5$ (solid) and $\alpha = 10$ (dotted); $\lambda = 1$ in all cases.

EXAMPLE A. The patterns of occurrence of earthquakes in terms of time, space, and magnitude are very erratic, but attempts are sometimes made to construct probabilistic models for these events. The models may be used in a purely descriptive manner or, more ambitiously, for purposes of predicting future occurrences and consequent damage.

Figure 2-12 shows the fit of a gamma density and an exponential density to the observed times separating a sequence of small earthquakes (Udias and Rice, 1975). The gamma density clearly gives a better fit ($\alpha = .509$ and $\lambda = .00115$). Note that an exponential model for interoccurrence times would be memoryless; that is, knowing that an earthquake had not occurred in the last t time units would tell us nothing about the probability of occurrence during the next s time units. The gamma model does not have this property. In fact, although we will not show this, the gamma model with these parameter values has the character that there is a large likelihood that the next earthquake will immediately follow any given one and this likelihood decreases monotonically with time. □

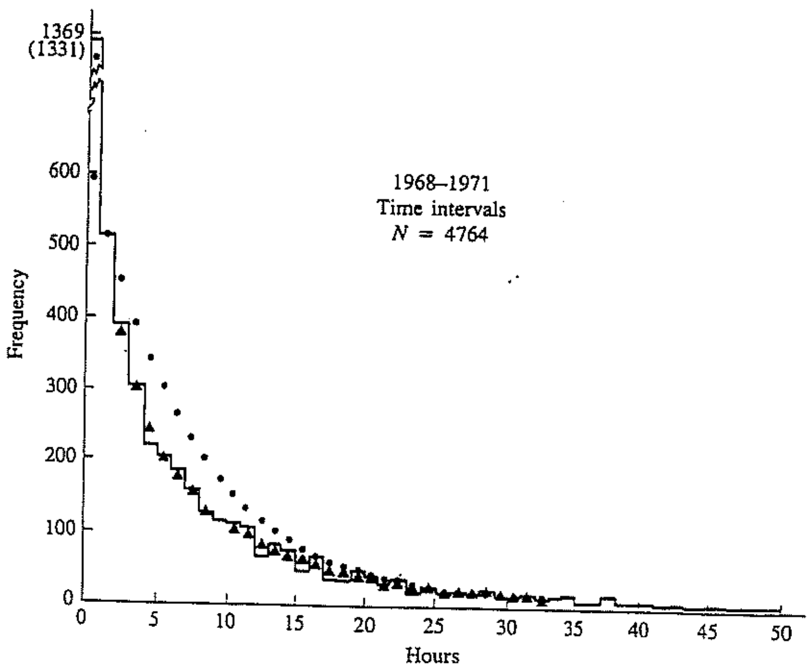


Figure 2-12. Fit of gamma distribution (triangles) and of exponential distribution (circles) to times between microearthquakes.

We can show $\Gamma(\alpha) = (\alpha - 1)\,\Gamma(\alpha - 1)$

case 1) If $\alpha = n$ $\quad \Gamma(\alpha) = (n-1)!$

case 2) $\alpha = \frac{n}{2}$ $\quad \Gamma\left(\frac{n}{2}\right) = \left(\frac{n}{2} - 1\right)\Gamma\left(\frac{n}{2} - 1\right)$

$$= \left(\frac{n}{2} - 1\right)\left(\frac{n}{2} - 2\right) \cdots \Gamma\left(\frac{1}{2}\right)$$

Can show $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

Connection with Poission Process (Gamma dist)

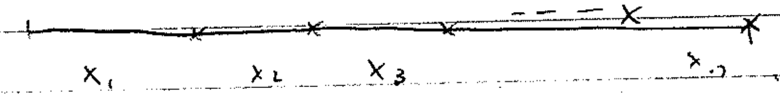
Although the gamma dist has many uses that are unconnected with the Poisson Process

But there is an incresting connection with Poisson.

* see H 10

Note: L_n = (time to 1st occ) + (time from 1st to 2nd) + ... + (time from $(n-1)^{st}$ to n^{th})

$$L_n = X_1 + X_2 + \dots + X_n$$



Each x_i is neg exp (λ).

For $\alpha = n$ (integer) then

L_n is $= \sum_{i=1}^{n} X_i$

when X_i = Time up to next occurence (for a poisson process)

15 dist funcs

L_n is sum of negative exponential random variables and has a gamma Distribution

gamma is prob density function.

4) Normal Distribution

P.D.F

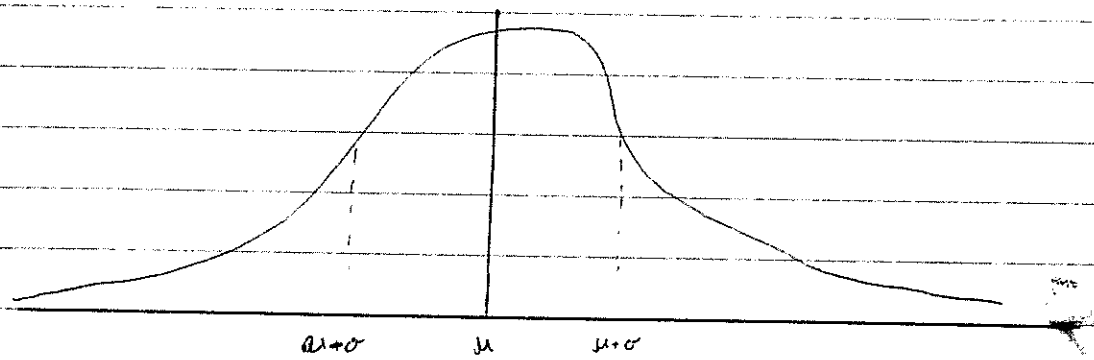
$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

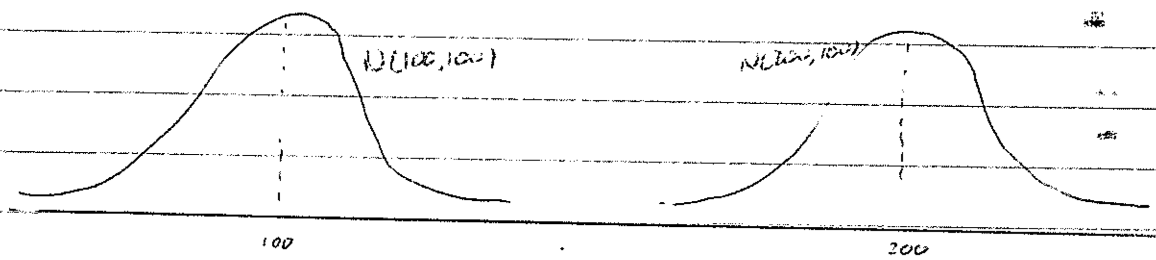
. μ = mean

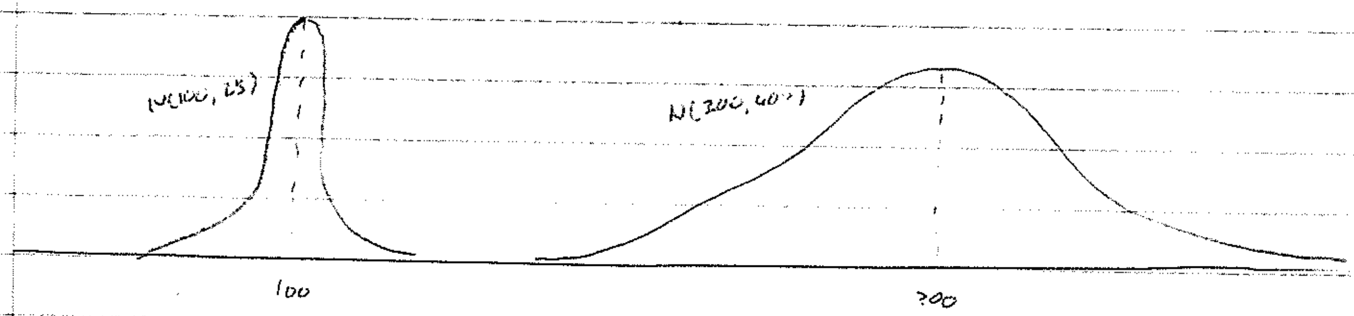
σ = Standard diviation

Shape : depends on μ and σ

$N(\mu, \sigma^2)$







Why important?

1) In nature (and man made phenomena) the normal curve can closely describe the shapes of many frequency distributions.

(eg distribution of heights (of plants, people etc)

– distributions of IQ measurements.

2) The distribution of averages ($\bar{x}$) follows a Normal Curve (Central limit theorem)

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$\sigma > 0$

$\mu \in \mathbb{R}$

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

let $z = \dfrac{x-\mu}{\sigma}$ $\qquad dz = \dfrac{1}{\sigma}\,dx$

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

This is normal PDF $\mu = 0$ $\sigma = 1$

Known as the standard normal PDF

let $t = \dfrac{z^2}{2}$ $\qquad dt = z\,dz \Rightarrow dz = \dfrac{1}{\sqrt{2}} t^{-\frac{1}{2}} dt$

$$I = 2\int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad \text{– Symmetry.}$$

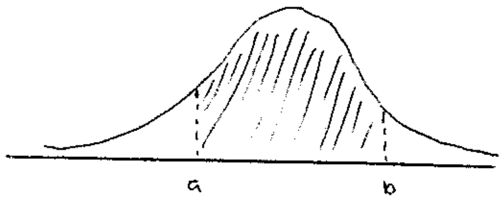
$$I = 2 \int_0^\infty \frac{1}{\sqrt{2}\sqrt{\pi}} e^{-t} \cdot \frac{1}{\sqrt{2}} t^{-\frac{1}{2}} \, dt$$

$$= \frac{1}{\sqrt{\pi}} \underbrace{\int_0^\infty t^{-\frac{1}{2}} e^{-t} \, dt}_{\Gamma(\frac{1}{2})} \qquad \Big| \; \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \, dt$$

$$I = \frac{1}{\sqrt{\pi}} \cdot \sqrt{\pi} = 1$$

$X = N(\mu, \sigma^2)$

$P[a < X \leq b]$



$= P(X \leq b) - P(X \leq a)$

$= F(b) - F(a)$

See sheet from 1st year.

$$P(X > x) = P\left(\frac{X - \mu}{\sigma} > \frac{x - \mu}{\sigma} \right)$$

$\uparrow$ $N(\mu, \sigma^2)$

$\uparrow$ $N(0,1)$

Example: $X \sim N(100, 100)$

Find $P(X > 120)$.

$$= P\left(\frac{X - 100}{10} > \frac{120 - 100}{10} \right).$$

$P(Z > 2) = .02275.$

Beta Distribution:

Random variable (continuous) on $(0, 1)$

The Prob Density Function (PDF):

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha - 1}(1 - x)^{\beta - 1} \quad \text{for } 0 < x < 1$$

where: $B(\alpha, \beta)$ is such that $\int_0^1 f(x)\,dx = 1$ and can be shown that $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$

Note: α, β are the parameters of the Beta distribution. They control the shape of the P.D.F. see H10 on shape.

Weibull Distribution:

The CDF (Cumulative Dist Function)

$$F(x) = 1 - e^{-(x/\theta)^\beta}$$

Differentiate w.r.t x, we'll get P.D.F

$$f(x) = F'(x) = \beta \left(\frac{x}{\theta}\right)^{\beta - 1} \frac{1}{\theta} e^{-(x/\theta)^\beta}$$

$$= \beta x^{\beta - 1} \cdot \frac{1}{\theta^\beta} e^{-(x/\theta)^\beta}$$

2 parameters β, θ: β = shape parameter

θ = scale parameter

SHAPES OF THE BETA DENSITY ST2054

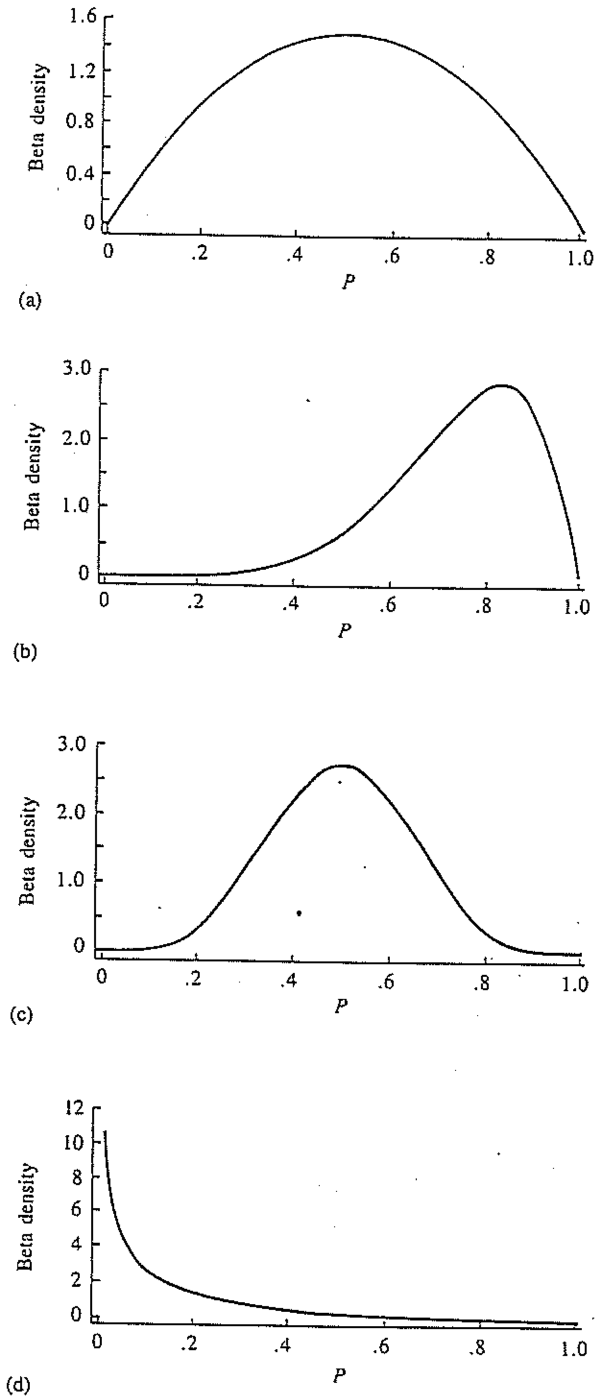
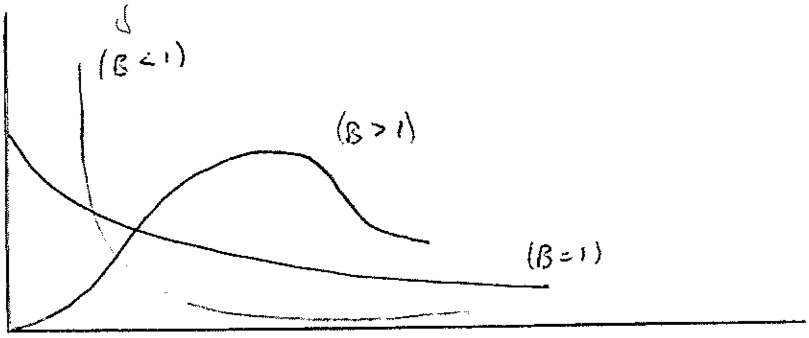


Figure 15-6. Beta density functions for various values of a and b: (a) $a = 2, b = 2$; (b) $a = 6, b = 2$; (c) $a = .5, b = 4$; and (d) $a = 6, b = 6$.

Shapes :



Log - Normal Distribution

Suppose Z is $N(\mu, \sigma^2)$

$P[Y \le y]$

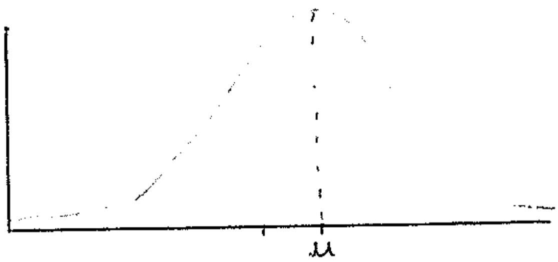
$= P[\log Y \le \log y]$

$= \Phi\left(\frac{\log y - \mu}{\sigma}\right)$.

Then $Y = e^Z$ will have a log-normal distribution.

Range of values for Y is $0 \to \infty$

Z is $N(\mu, \sigma^2)$



$Y \sim LN(-2, 4)$

$P[Y \le .32]$

$= P[\log Y \le \log(.32)]$

$= \Phi\left[\frac{\log(.32) + (2)}{2}\right]$

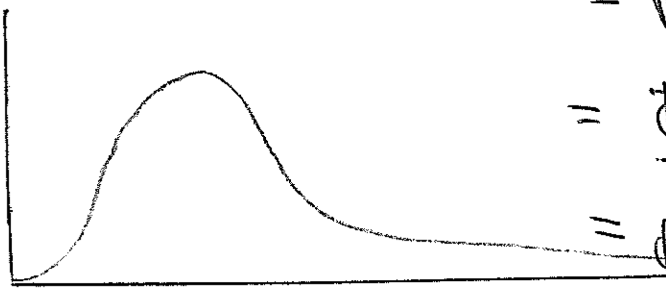
$= \Phi\left[\frac{-1.14 + 2}{2}\right]$

$= \Phi(.43) = .667$

$Y = \log N(\mu, \sigma^2)$

$f(y)$

used for claim size distribution.



<u>Location parameters</u> & <u>Scale Parameters</u>

Consider a random variable with Dist Funct $F(x)$

if we can write $F(x)$ as $F(x - \gamma)$

eg : For $X \sim N(\mu, \sigma^2)$

μ = location Parameter

$$F(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Note: An equivalent condition is that the P.D.F can be written as $f(x-\gamma)$ ... γ = location parameter

Scale Parameter:

If the P.D.F can be written in form

$$f(x) = \frac{1}{\theta} g\left(\frac{x}{\theta}\right) \qquad \theta = \text{scale parameter}$$

Examples: Normal PDF $= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$\Rightarrow \sigma$ = scale Parameter.

Gamma PDF $= \frac{1}{\Gamma(\gamma)} \left(\frac{x}{\beta}\right)^{\gamma-1} e^{-(x/\beta)} \frac{1}{\beta}$

$\Rightarrow \beta$ = scale parameter.

Weibull Dist Funct = $F(x) = 1 - e^{-(x/\theta)^{\beta}}$

PDF $\Rightarrow f(x) = \beta \left(x/\theta\right)^{\beta-1} 1/\theta \; e^{-(x/\theta)^{\beta}}$

$\Rightarrow \theta$ = scale parameter.

Whatever it's "over" = scale Parameter.

EXPECTED VALUES for Random Variables.

Vague def: Exp. value is a long term average value for the random variable (if one cud observe it many times)

Discrete case first: x_i has Prob $f(x_i)$

$$\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \\ f(x_1) & f(x_2) & & f(x_n) \end{array}$$

Example: Game of chance: charge €1 for entry

Die is thrown → Prizes won

G = Net Gain from 1 play.

G	-1	0	4
Prob	$4/6$	$1/6$	$1/6$

Play over & over again. Ask what is long term outlook.

Average values for G is ~ 0

The reason for this is cuz Exp. value of $G = 0$

$$E(G) = (4/6)(-1) + 0(1/6) + 4(1/6) = 0$$

<u>Definition of expected value for Random Variables.</u>

X is discrete with prob function $f(x_i)$. Then Exp. value of X is given by

$$E(X) = \sum_{\text{all } x_i} x_i \, P(x_i)$$

Provided the sum converges absolutely

ie $\sum_{\text{all } x_i} |x_i| \, |P(x_i)| < \infty$

If we don't have absolute convergence then we say $E(X)$ does not exist

$$F_x(x_i) = P[X \le x_i]$$

Lets also define $E[g(X)]$ where $g(X)$ is a function of random var X

Definition

$$E[g(X)] = \sum_{\text{all } x_i} g(x_i)\, P[X = x_i] \quad (*)$$

exists ~~prod~~ provided $\sum_{\text{all } x_i} |g(x_i)|\, P[X = x_i] < \infty$

If X cont. $E[g(x)] = \int g(x) f(x)\,dx$ prov. $\int |g| f(x)\,dx < \infty$

Note: lets denote $g(x)$ by y. y is also a random variable.

$E[a g_1(x) + b g_2(x)] = a E[g_1(x)] + b E[g_2(x)]$

- so we don't need a definition for $E(Y)$

what we could do is to get the Prob Distribution for Y and then apply the definition of $E(Y)$ with that prob dist.

However: its often quite difficult to find the Prob Distribution for $Y = g(x)$

- and because of this its easier to get $E[g(x)]$ using $(*)$

ii) later we look at methods for finding the prob dist. of Y (a function of X)

Important Example of $g(x) = [X - E(x)]^2$

$\mu = E(x)$

we will need to find $E[(X-\mu)^2]$

variance of a random variable

Denote $E[(x-\mu)^2] = V(x) = \sigma^2$

Related quantity is Standard Div of X

$SD(X) = \sqrt{V(x)} = \sigma$

— Example: Find the variance of $X \sim b(n,p)$

Soln.: In evaluating $V(x)$ its useful to know

$V(x) = E[(x-\mu)^2] = E(x^2) - [E(x)]^2$

Proof: $V(x) = E[(x-\mu)^2] = E[x^2 - 2x\mu + \mu^2]$

$= \sum (x_i^2 - 2x_i\mu + \mu^2)\, P(X=x_i)$

$= \sum_{\text{all } x_i} x_i^2 P(X=x_i) - 2\mu \sum x_i P(X=x_i) + \mu^2 \sum P(X=x_i)$

$= E(x^2) - 2\mu E(x) + \mu^2 (1)$

$= E(x^2) - 2\mu \cdot \mu + \mu^2$

$= E(x^2) - \mu^2$

$V(x) = E(x^2) - [E(x)]^2$

Lets find $E(x)$ for $x \sim b(n,p)$

$E(x) = np$

$$E(x) = \sum_{k=0}^{n} k\, P(x=k)$$

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

First term = 0

$$= \sum_{\underline{k=1}}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \qquad \Big|\ Q = 1-p$$

$$= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k Q^{n-k} \quad \text{.. by cancellation.}$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!\,[(n-1)-(k-1)]!} p^{k-1} Q^{(n-1)-(k-1)}$$

$$\Big|\ k-1 = j$$

$$= np \sum_{j=0}^{n-1} \frac{(n-1)!}{j!\,[(n-1)-j]!} p^j Q^{(n-1)-j}$$

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

$$= np\,[p+Q]^{n-1} \quad \text{.. but } p+Q = 1$$

$$\Rightarrow \underline{\underline{np. = E(x)}}$$

NOTE! Expected values for discrete var

$= \sum(x_i)\,.\,P(x_i)$ .. always.

cont: 1 var $E(g(x)) = \int g(x)\, f(x)\,dx$.

2 var: $E(g(x,y)) = \iint g(x,y)\,[\text{particular pdf}]\,dy\,dx$

eg $E(xy) = \iint xy\, f(x,y)\,dy\,dx$.

lets get $V(x)$ = $X \sim b(n,p)$.

$E(x) = np.$

$E(x^2) \Big| = E[x(x-1)] = \sum_{k=0}^{n} k(k-1)\, P(x=k)$

$= \sum_{k=2}^{n} k(k-1) \left[\frac{n!}{k!(n-k)!} \right. p^k (1-p)^{n-k}$

Just manipulation

$= n(n-1) \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!\,[(n-2)-(k-2)]!} p^2 . p^{k-2} (1-p)^{(n-2)-(k-2)}$

$J = k-2.$

$n(n-1)\, p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j}$

$= n(n-1)\, p^2 \,[\, p+1-p \,]^{n-2}$

$E[x(x-1)] = n(n-1)\, p^2$ *

$E[x^2 - x] = Ex^2 - Ex$

$\Rightarrow \quad E(x^2) = E[(x(x-1))]^* + Ex$

Thus $E(x^2) = n(n-1)\,p^{2*} + Ex$

$= n(n-1)\,p^2 + np.$

Then :

$V(x) = E(x^2) - [E(x)]^2$

$= n(n-1)\,p^2 + np - (E(x))^2$

$$= n(n-1)p^2 + np - (np)^2$$

$$= -np^2 + np$$

$$= V(x) = np(1-p)$$

$$E(x) = np.$$

Learn these.

---

like $E(x) = x . P(x = x_i)$

Example: $X \sim \text{Poisson}(m)$: $\dfrac{m^k e^{-m}}{k!}$

$$E(x) = \sum_{\text{all } k} k \, P(x=k)$$

$$E(x) = \sum_{k=0}^{\infty} k \frac{m^k e^{-m}}{k!}$$

$$= \sum_{k=1}^{\infty} \frac{m^k e^{-m}}{(k-1)!}$$

$$= m \sum_{k=1}^{\infty} \frac{m^{k-1} e^{-m}}{(k-1)!} \qquad k-1 = j$$

$$= m \sum_{j=0}^{\infty} \underbrace{\frac{m^j e^{-m}}{j!}}_{\text{Prob}(x=j)} = m(1)$$

$$= m.$$

$$E(x) = m. \text{ for poisson.}$$

Thus the parameter (m) of the poisson Dist is the mean value.

Note: Remember the poisson process.

we found $P(X_t = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$

Thus $E(X_t) = \lambda t$

(λt) is the average # occurances in time t
and λ must be the average # occurances per unit time.

λ often called Rate of poisson process.

Next: $V(X)$ for $X \sim \text{Poisson}(m)$

useful to consider $E(X(X-1))$

$$\sum_{k=0}^{\infty} k(k-1) \frac{m^k e^{-m}}{k!}$$

First 2 terms = 0

$$= \sum_{k=2}^{\infty} \frac{m^k e^{-m}}{(k-2)!} \qquad j = k-2$$

$$m^2 \sum_{j=0}^{\infty} \frac{m^j e^{-m}}{j} \quad . \text{ think abt it } \sum_{j=0}^{\infty} P(X=j) = 1$$

$$= m^2 (1)$$

$$E[X(X-1)] = m^2 =$$

$$E(X^2) = E(X^2 - X) + E(X)$$
$$= m^2 + E(X)$$

$= m^2 + m.$

$$V(x) = E(x^2) - [E(x)]^2$$

$$= m^2 + m - (m)^2$$

$$= m$$

$$\left.\begin{array}{l} V(x) = m. \\ E(x) = m. \end{array}\right\} \text{ For poisson.}$$

Next : $X \sim$ Geometric.

$X =$ # trials up to 1^{st} success.

$$P(X = k) = \underbrace{(1-p)^{k-1}}_{k-1 \text{ fails}} \cdot \underbrace{p}_{1 \text{ success}}$$

$$E(x) = 1/p$$

Why?

$$E(x) = \sum_{k=1}^{\infty} k \cdot P(X = k)$$

$$= \sum_{k=1}^{\infty} k (1-p)^{k-1} p.$$

$$E(x) = p + 2(1-p)p + 3(1-p)^2 p + 4(1-p)^3 p + \dots$$

Multiply across by $(1-p)$.

$$\text{Variance} = E(x^2) - [E(x)]^2$$
$$= E[x - E(x)]^2$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx. \quad f(x).$$

$$= \int_{-\infty}^{\infty} x^2 \cdot \underbrace{\frac{1}{\Gamma(x)} \left(\frac{x}{\beta}\right)^{\gamma-1} e^{-x/\beta} \cdot \frac{1}{\beta}}_{f(x)}\, dx$$

$$= \frac{1}{\Gamma(x)} \int_{0}^{\infty} \beta^2 \left(\frac{x}{\beta}\right)^{\gamma+1} e^{-x/\beta}\, \frac{1}{\beta}\, dx$$

$$T = x/\beta.$$

$$\frac{\beta^2}{\Gamma(x)} \underbrace{\int_{0}^{\infty} t^{\gamma+1} e^{-t}\, dt}_{\Gamma(\gamma+2)}$$

$$\Gamma(x+2) = \gamma+1\, \Gamma(x+1) = (\gamma+1)(\gamma)\,\Gamma(\gamma)$$

N.B

$$\Gamma(\gamma) = \int_0^\infty t^{\gamma-1} e^{-t}\, dt$$

$$\Gamma(\gamma+1) = \gamma\, \Gamma(\gamma).$$

$$\frac{\beta^2}{\Gamma(\gamma)} (\gamma+1)(\gamma)\,\Gamma(\gamma) = \gamma(\gamma+1)\,\beta^2.$$

Thus: $V(x) = E(x^2) - [E(x)]^2$

$$= \gamma(\gamma+1)\beta^2 - (\gamma\beta)^2$$

$$V(x) = \gamma\beta^2.$$

$$\left.\begin{aligned} E(x) &= \gamma\beta. \\ V(x) &= \gamma\beta^2 \end{aligned}\right\} \text{gamma.}$$

$E(x) = \mu$ } Normal distribution,

$E(x) = \sigma^2$

MOMENTS FOR A RANDOM VARIABLE

These are more descriptive measures for a random var (and its distribution).

Consider $\underbrace{E[x-a]^k}_{k^{th} \text{ moment about the constant } a.}$ $k \in \mathbb{Z}^+$, $a \in \mathbb{R}$

condition: $\int |x-a|^k f(x)\,dx < \infty$

Can make choices for a and k

$A = 0$: we have $E[x]^k$

These are called moments about the origin

Denoted by μ_k'

but μ_1' is denoted by μ

$A = \mu$: we have $E(x-\mu)^k$

These are known as moments about the mean (μ) or central moments

Notation: $E(x-\mu)^k$ or μ_k

Most important of these is μ_2 which we have already seen

$V(x) = \mu_2 = Var(x) = \sigma^2$

Higher moments: when $k > 2$.

only $k=3$ and (to a lesser extent) $k=4$ are of interest.

$E(x-\mu)^3$ is supposed to tell us about the skewness of the dis of x.

$\frac{E(x-\mu)^3}{\sigma^3}$: coefficient of skewness.

Prob distributions for which $E(x)$ doesn't exist.

Discrete example.

1) $x = 1, 2, 3, 4, \ldots, n, \ldots \infty$

with $Prob[x=i] = k\frac{1}{i^2}$ for $i = 1, 2, \ldots$

Find k to ensure that $\sum_{i=1}^{\infty} Prob[x=i] = 1$

$\Rightarrow k \sum_{i=1}^{\infty} 1/i^2 = 1$

$k\left[\frac{\pi^2}{6}\right] = 1$

$k = \frac{6}{\pi^2}$

$E(x) = \sum x_i \, Prob[x = x_i]$

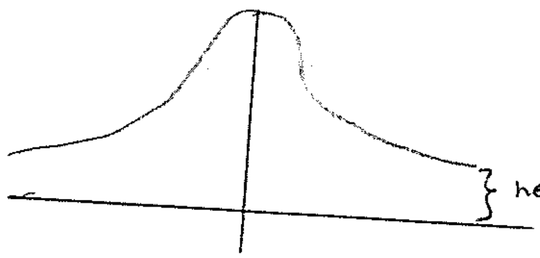
$E(x) = \sum_{i=1}^{\infty} i\left(\frac{6}{\pi^2}\right)\frac{1}{i^2}$

$= \frac{6}{\pi^2}\sum_{i=1}^{\infty}\frac{1}{i} = \infty \Rightarrow E(x)$ doesn't exist.

2 nice thats the prob.

2) Continuous: X having cauchy distribution

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad \text{for} \quad -\infty < x < \infty$$



heavy tailed dist. - more area under tails than $N(0,1)$.

Check area under curve = 1.

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \qquad = \int_{-\infty}^{\infty} \frac{1}{\pi} \cdot \frac{1}{1+x^2}\,dx$$

let $x = \tan\theta \qquad d\theta = \sec^2\theta\, d\theta = 1 + \tan^2\theta$

$$\int_{-\pi/2}^{\pi/2} \frac{1}{\pi} \frac{1}{1+\tan^2\theta} \sec^2\theta\, d\theta = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} d\theta = 1$$

$$E(x) = \int_{-\infty}^{\infty} |x|\, f(x)\, dx$$

$$= \int_{-\infty}^{\infty} |x| \frac{1}{\pi(1+x^2)}\,dx$$

cuz of symmetry

$$= 2\int_{0}^{\infty} \frac{x}{\pi(1+x^2)}\,dx$$

$$= 2 \cdot \frac{1}{2} \int_{0}^{\infty} \frac{d(x^2)}{\pi(1+x^2)} = \frac{1}{\pi}\int_{0}^{\infty} \frac{d(x^2)}{(1+x^2)}$$

$$= \frac{1}{\pi} \ln(1+x^2) \Big]_0^\infty \rightarrow \infty$$

$\Rightarrow E(x)$ does not exist.

What does $E(x)$ not existing imply?

$E(x)$ is long term average value if exp. is observed a large number of time.

In the cauchy dist. large + or large − values occur quite often (This is cuz of large area under tail).

---

note need lecture notes on this.

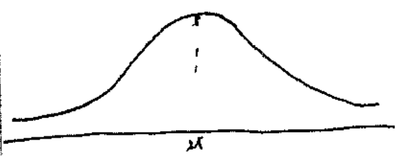
Linear Function of a Random Variable

Random var. X

$y = a + bx$ (a, b constants)

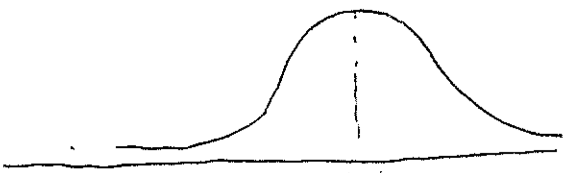
We showed $E(ax+b) = aE(x) + b$

$$V(ax+b) = a^2 V(x)$$



P. Den. F.

$Y = (1)X + b$.



Same shape but shifted.

APPLICATION : n Bernoulli trials

X = # of successes

$X = b(n, p)$

$\hat{p}$ = Sample proportion

Consider $\hat{p} = \frac{X}{n}$ ← x is random. : $aX + 0$ where $a = \frac{1}{n}$.

random too

Possible values of $\hat{p}$: $0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}$.

we want $E(\hat{p})$, $V(\hat{p})$

$E(\hat{p}) = \frac{1}{n} E(x)$

$\Downarrow$

$\frac{1}{n}(np) = p$

So the expected value of our estimate proportion is the actual proportion unbiased

$V(\hat{p}) = \left(\frac{1}{n}\right)^2 V(X)$

$\Downarrow$

$\frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$.

Variance is $\div n$. so if we have more n, variance decreases x is more accurate we control the accuracy.

- If E(estimate) = actual thing we're estimating :
"unbiased" estimator.

- If we want to estimate p, we could think of using $\hat{p} = \frac{x}{n}$ as our estimator for p

$\therefore$ coz $E(\hat{p}) = p$ we see that $\hat{p}$ is an unbiased Estimator for p

- and also that the reliability of $\hat{p}$ as an estimator will improve as n is made large.

CONVERGENCE of a sequence of Random var.

Consider $\hat{p}$ $\left(= \frac{x}{n}\right)$

This depends on n, so lets change notation
$\hat{p}$ now $= \hat{p}_n$

How does $\hat{p}_n$ behave as n gets larger.

This leads us to define:

$$P|X| > \varepsilon| \leq \frac{E(|X|)}{a}$$

convergence in probability:
Consider a sequence of random variables $\{X_n\}$
We say that the sequence converges in probability to the value c (a constant) if for every $\varepsilon > 0$

$$\lim_{n\to\infty} Prob\left[\,|X_n - c| < \varepsilon\,\right] = 1$$

For large enuf n, the difference between X_n and c can be as small as we want

With regard to the sequence $\{\hat{p}_n\}$, it turns out that $\hat{p}_n$ converges in probability to p

To show this we need a few more results.

In caught:

i) Markov Inequality:

Consider some random variable X and some function $g(x) \geq 0$ and constant $c > 0$

g operating on rand var.

rand variable.

$$\text{Prob}\left[g(X) \geq c \right] \leq \frac{E[g(X)]}{c}$$

rember: $E(Y) = \int_{\text{all } y} |y| f(y)\,dy$

PROOF: $E[g(X)] = \int_{\text{all } x} g(x) f(x)\,dx$

$$= \int_{\{\text{all } x \text{ such that } g(x) \geq c\}} g(x) f(x)\,dx + \int_{\{\text{all } x \text{ such that } g(x) < c\}} g(x) f(x)\,dx$$

~~both~~ 2^{nd} term is ≥ 0 since $g(x) \geq 0$

$$\Rightarrow E[g(X)] \geq \int_{\{\text{all } x \text{ such that } g(x) \geq c\}} g(x) f(x)\,dx$$

$$\geq \int_{\{\text{all } x \text{ s.t } g(x) \geq c\}} c f(x)\,dx = c \int_{\text{all } x : g(x) \geq c} f(x)\,dx \quad *$$

But when we evaluate $\int_{\text{all } x \text{ s.t } g(x) \geq c} f(x)\,dx$ we are actually getting

$$P[g(X) \geq c]$$

Ex: To get $P(x > a)^{**} = \int_a^{\infty} f(x)\,dx$.

You integrate over interval that ** is true. $a \to \infty$

$$E[g(x)] \geq c P[g(x) \geq c]$$

$$P[g(x) \geq c] \leq \frac{1}{c} E[g(x)]$$

APPLICATION: Suppose X is possion $m = 5$

choose $g(x) = x$.

Note that (for x - poisson) $g(x) \geq 0$

lets choose $c = 10$

Then from Markov inequality

$$P[x \geq 10] \leq \frac{E[x]}{10}$$

from tables

$.0318 \leq .5$ / correct but not very sharp.

CHEBYSHEV INEQUALITY.

Suppose we have some Random variable X for which $V(x) = \sigma^2$ exists.

Then $$P[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$

where $\mu = E(x)$

$k \in Z^+$ usually $k \geq 1$

"The prob that the rand. var differs from its mean by at least $k\sigma$ is at most $\frac{1}{k^2}$

For many distributions the actual value of $P[|X-\mu| \geq k\sigma]$ is considerably less than the $\frac{1}{k^2}$ given by the inequality.

But There are examples where the chebyshev bound is sharp:

x	80	100	120
Prob	$1/8$	$3/4$	$1/8$

$\underline{E(x) = 100} = \mu$ (by chance) $= 100$

$10 + 75 + 15 = 100$

$V(x) = E[x-\mu]^2$

$= \frac{1}{8}(80-100)^2 + \frac{3}{4}(100-100)^2 + \frac{1}{8}(120-100)^2$

$= \frac{400}{8} + 0 + \frac{400}{8}$

$\underline{V(x) = 100}$

$\underline{\sigma = 10}$ $\quad \underline{k=2}$

lets get: $P[|X-\mu| \geq 2\sigma]$ for the dist

$= P[|X-100| \geq 20]$

$= P[(x=80) \text{ or } (x=120)] = \frac{1}{8} + \frac{1}{8} = 1/4$

Bound $= 1/k^2 = 1/4$

What important about the chebyshev Inequality is that it applies to all probability distributions (as long as $E(x)$ & $V(x)$ exist).

Our principle application is to prove

Weak LAW of LARGE numbers.

Consider a seq of Bernoulli trials with Prob (succ) - p

Consider n trials, let $\hat{p} = \frac{x}{n}$ (x = # successes in n trials)

$\hat{p}$ estimate for p (prob of success) Think of heads / Tails

We will show that $\hat{p}_n$ converges in probability to p as $n \to \infty$

Remark: convergance in Prob.

we say that $\hat{p}_n$ converges in probability to p

if for any positive $\varepsilon > 0$

$$\lim_{n \to \infty} P[\, |\hat{p}_n - p| < \varepsilon \,] = 1$$

ie $$\lim_{n \to \infty} \text{Prob}[\, p - \varepsilon < \hat{p}_n < p + \varepsilon \,] = 1$$

Proof: Our rand. var is $\hat{p}_n$ & weve already shown

$E(\hat{p}_n) = p = \mu$

$V(\hat{p}_n) = \frac{p(1-p)}{n} = \sigma^2$

Use chebyshev : $x = \hat{p}_n$

so.

$$P\left(|\hat{p}_n - p| \geq k \sqrt{\frac{p(1-p)}{n}} \right] \leq \frac{1}{k^2}$$

get complemotary event

$P(A) \leq 1/k^2$

$P(A^c) = 1 - P(A)$

$1 - P(A) \geq -\frac{1}{k^2}$

$1 - P(A) \geq 1 - 1/k^2$

$$P\left[\,|\hat{P}_n - P| < k\sqrt{\frac{P(1-P)}{n}}\,\right] \geq 1 - 1/k^2$$

$$\varepsilon = k\sqrt{\frac{P(1-P)}{n}}$$

$$\varepsilon^2 = k^2\left(\frac{P(1-P)}{n}\right) \qquad \geq 1 - \frac{P(1-P)}{n\varepsilon^2}$$

$$1/k^2 = \frac{P(1-P)}{n\varepsilon^2}$$

as $n \to \infty$

$$\lim_{n\to\infty} P\left[\,|\hat{P}_n - P| < \varepsilon\,\right] \geq 1$$ can't be greater than 1

$$\lim_{n\to\infty} P\left[\,|\hat{P}_n - P| < \varepsilon\,\right] = 1$$

convergence in Prob.

Weak LAW of large numbers. See M10.

μ = Population mean.

$\bar{X}_n$: mean of sample of size n.

$$\bar{X}_n \xrightarrow{P} \mu$$

Bivariate & Multivariate Distributions

We consider 2 (or more) random variables, similtaniously.

We want to know about their joint Probability Distribution

One can descibe this using a bivariate distribution function.

$$F(x_1, x_2) = \text{Prob}\left[X_1 \leq x_1 \cap X_2 \leq x_2\right]$$

Proof: use markov Inequality.

- choose $g(x) = (x-\mu)^2$

  note $g(x) \geq 0$ as required by markov.

- Choose $c = k^2\sigma^2$

Note that $E[g(X)] = E[(X-\mu)^2] = \sigma^2$

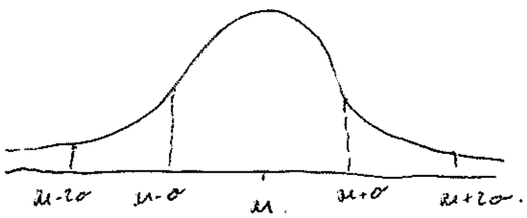
Consider markov for these choices:

$$P[(x-\mu)^2 \geq k^2\sigma^2] \leq \frac{E[g(x)]}{c} = \frac{\sigma^2}{k^2\sigma^2}$$

$$P[|x-\mu| \geq k\sigma] \leq \frac{1}{k^2}$$ - remember

Example: Suppose x is $N(\mu, \sigma^2)$ & $k=2$.

Thus $P[|x-\mu| \geq 2\sigma] \leq \frac{1}{4}$ ← chebyshev bound.



$\mu-\sigma$, $\mu+\sigma$ at pts of inflection.

$|x-\mu| \geq 2\sigma$ corresponds to x values below $(\mu-2\sigma)$ and above $(\mu+2\sigma)$

$N(0,1)$ Tables Tell us its actually .0456

WEAK LAW OF LARGE NUMBERS

(A RESULT REQUIRED IN THE COURSE [illegible])

FIRST, WE LOOK AT THE CONCEPT OF CONVERGENCE IN PROBABILITY:

DEFN LET $Z_1, Z_2, \ldots$ BE A SEQUENCE OF RANDOM VARIABLES. THE SEQUENCE $\{Z_n\}$ CONVERGES IN PROBABILITY TO b IF FOR ANY GIVEN NUMBER $\varepsilon > 0$

$$\lim_{n\to\infty} P(|Z_n - b| < \varepsilon) = 1$$

(ROUGHLY SPEAKING, THIS SAYS THAT $\{Z_n\}$ CONVERGES IN PROBABILITY TO b IF THE PROB. DISTRIB OF Z_n BECOMES MORE AND MORE CONCENTRATED AROUND b AS $n \to \infty$.)

NOTATION: $\operatorname{plim}_{n\to\infty} Z_n = b$ $\boxed{\text{OR}}$ $Z_n \xrightarrow{P} b$

WEAK LAW OF LARGE NUMBERS (KHINTCHINE'S THEOREM)

IF THE COMMON DISTRIBUTION OF THE INDEPENDENT, IDENTICALLY DISTRIBUTED VARIABLES $X_1, X_2, \ldots$ HAS A FINITE FIRST MOMENT μ, THEN FOR THE SEQUENCE OF AVERAGES $\overline{X}_n = \sum_1^n X_i$, WE HAVE

$$\operatorname{plim}_{n\to\infty} \overline{X}_n = \mu$$

PROOF: WE SHALL ASSUME THAT THE DISTRIBUTION OF THE X_i HAS A FINITE VARIANCE σ^2. (THEOREM ALSO HOLDS WHEN VARIANCE DOES NOT EXIST, I.E. INFINITE VARIANCE).

WE USE THE CHEBYSHEV INEQUALITY:

$$P[|X-\mu| \geq k\sigma] \leq \frac{1}{k^2}$$

$$\Rightarrow P[|X-\mu| < k\sigma] \geq 1 - \frac{1}{k^2}$$

{R. VAR X, WITH MEAN μ VAR σ^2

NOW FOR THE R. VAR $\overline{X}_n$, $E\overline{X}_n = \mu$

AND $V(\overline{X}_n) = \sigma^2/n$ {SINCE X_i ARE iid

FOR ANY $\varepsilon > 0$, CHOOSE k SO THAT $\frac{k\sigma}{\sqrt{n}} = \varepsilon$ [I.E. $k = \frac{\varepsilon\sqrt{n}}{\sigma}$

$$\Rightarrow P[|\overline{X}_n - \mu| < \varepsilon] \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$$

$$\Rightarrow \lim_{n\to\infty} P[|\overline{X}_n - \mu| < \varepsilon] = 1$$

(I.E. $\operatorname{plim}_{n\to\infty} \overline{X}_n = \mu$)

TO SHOW

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) - \sum_{i<j} P(E_i \cap E_j) + \sum_{i<j<k} P(E_i \cap E_j \cap E_k)$$

$$+ \cdots (-1)^n \sum_{i_1<i_2<\cdots<i_{n-1}}^{n} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_{n-1}}) + (-1)^{n+1} P(E_1 \cap E_2 \cap \cdots E_n)$$

PROOF (INDUCTION)

For $(n=2)$, we have shown that $P(E_1 \cup E_2) = \sum_{i=1}^{2} P(E_i) - P(E_1 \cap E_2)$

Assume true for m, try to deduce for $(m+1)$.

$$P\left(\bigcup_{i=1}^{m+1} E_i\right) = P\left(\bigcup_{i=1}^{m} E_i \cup E_{m+1}\right) = P\left[\bigcup_{i=1}^{m} E_i\right] + P\left[E_{m+1}\right] - P\left[\left(\bigcup_{i=1}^{m} E_i\right) \cap E_{m+1}\right]$$

$$= \left[\sum_{i=1}^{m} P(E_i) + P(E_{m+1}) - \sum_{i<j} P(E_i \cap E_j) + \cdots\right] - P\left[\bigcup_{i=1}^{m} (E_i \cap E_{m+1})\right]$$

$$= \sum_{i=1}^{m+1} P(E_i) + \Big(- \sum_{i<j}^{m} P(E_i \cap E_j) - \sum_{i=1}^{m} P(E_i \cap E_{m+1})$$

$$+ \sum_{i<j<k}^{m} P(E_i \cap E_j \cap E_k) - \Big\{ - \sum_{i<j}^{m} P\big(\underbrace{(E_i \cap E_{m+1}) \cap (E_j \cap E_{m+1})}_{E_i \cap E_j \cap E_{m+1}}\big)$$

$$- - - - - -$$

$$+ (-1)^m \sum_{i_1<i_2<\cdots i_{m-1}}^{m} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_{m-1}}) - \Big\{ (-1)^m \sum_{i_1<i_2<\cdots<i_{m-1}}^{m} P(E_{i_1} \cap E_{i_2} \cap \cdots E_{i_{m-1}} \cap E_{m+1})$$

$$+ (-1)^{m+1} P(E_1 \cap E_2 \cdots \cap E_m) \Big) - (-1)^{m+1} P(E_1 \cap E_2 \cap \cdots E_m \cap E_{m+1})$$

$$= \sum_{i=1}^{m+1} P(E_i) - \sum_{j<j}^{m+1} P(E_i \cap E_j) + \sum_{i<j<k}^{m+1} P(E_i \cap E_j \cap E_k)$$

$$+ \cdots\cdots (-1)^{m+1} \sum_{i_1<i_2<\cdots<i_m}^{m+1} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m})$$

$$+ (-1)^{m+2} P(E_1 \cap E_2 \cap \cdots \cap E_m \cap E_{m+1})$$

which $\Rightarrow$ Proposition is true for $(m+1)$. (Completes proof by induction)

NOTE ON 2nd LAST TERM ON RHS (UNDERLINED)

$$-(-1)^m \sum_{i_1<i_2\cdots i_{m-1}}^{m} P(E_{i_1} \cap E_{i_2} \cap \cdots E_{i_{m-1}} \cap E_{m+1})$$

$$= (-1)^{m+1} \Big[P(E_1 \cap E_2 \cap \cdots \cap E_{m-1} \cap E_{m+1})$$
$$+ P[E_2 \cap E_3 \cap \cdots \cap E_m \cap E_{m+1}$$
$$+ P[E_1 \cap E_3 \cap \cdots \cap E_m \cap E_{m+1}$$
$$+ P[E_1 \cap E_2 \cap E_4 \cap \cdots \cap E_m \cap E_{m+1}$$
$$- - - - + P[E_1 \cap E_2 \cap E_3 \cap \cdots E_{m-3} \cap E_{m-2} \cap E_{m-1} \cap E_{m+1} \Big]$$

WHEN WE ADD $(-1)^{m+1} P[E_1 \cap E_2 \cap \cdots E_m$ TO THIS

WE GET:

$$(-1)^{m+1} \sum_{i_1<i_2<\cdots<i_m}^{m+1} P(E_1 \cap E_2 \cap \cdots \cap E_{i_m})$$

How do we Find Numerical Values

There are 3 Types /approaches

i) Classical Prob. (A Priori Prob.)

ii) Empirical Prob (statistical Prob)
e.g: Life tables. Try to repeat exp. N times.

iii) Subjective Probability Assignment
Idea is to use intuition to find value of P(A)
e.g Man Utd vs Liverpool
ii) cant be repeated numerous times under identical cond.
i) Utd + Liverpool don't have d exact same prob of victory - at Old Trafford Utd prob will win

ii) The event A occurs, denote by n_A
Then we take as

$$P(A) \rightarrow \frac{n_A}{N}.$$

But, there are many cases where the R. exp can't be repeated.

i) Classical Probability
This approach is based on noting that the random experiment has symmetry present so we can claim dat all the possible outcomes are Equally LikeLY

Coin Toss (2 outcomes => P(H) = P(T))
Dice - Throw (6 outcomes => P(ω) = 1/6)

★ If there are M outcomes and all equally likely the prob of any specified outcome is $\frac{1}{M}$.

$\frac{1}{16}$

$P(\text{Queen}) = P(Q_H \cap Q_S \cap Q_C \cap Q_D)$ – M.E.

$\Rightarrow P(Q_H) + P(Q_S) + P(Q_C) + P(Q_D)$

$= \frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \frac{1}{52}$

$= 4/52$

Examples: i) 2 coins, thrown together

$P(\text{getting 1 Head}) = \frac{1}{2}$

4 coins, throw together

$P(\text{getting 2 Heads}) = \frac{1}{4}$?

H H T T

i) $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16}$

ii) H T T H $= 1/16$

iii) T T H H $= 1/16$

iv) T H H T $= 1/16$ — M.E.

v) T H T H vi) H T H T $= 1/16$

so $P(i \cap ii \cap iii \cap iv) = P(i) + P(ii) + P(iii) + P(iv) + P(v) + P(vi)$

or $\boxed{2}\ \boxed{2}\ \boxed{2}\ \boxed{2} = 16$ total $= \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16}$

$\boxed{H}\ \boxed{H}\ \boxed{T}\ \boxed{T}$ so like 1ser

its $\frac{4!}{2! \times 2!} = 6$ $= \cancel{\frac{1}{4}}\ \frac{6}{16}$.

– In dis small example it was possible to list all the out comes (innumeration) & choose the one we wanted

Ex: 2 dice thrown

FOR CLASSIC: $\frac{\text{NO OF FAVOURABLE}}{\text{Total No. of outcomes}}$ – DRAW BOXES

WGB_G08 — ST2054 5-6 Thursday

$E \in \mathcal{F} \Rightarrow E^c \in \mathcal{F}$

Computation of Probabilities.

- Classical
- Empirical
- Subjective

(1) Classical.

- Look for elementary events and symmetries and ~~compute~~ build up probabilities from them.

Examples. (i)

Prob of Queen

$$P[Q] = P[Q_H \cup Q_S \cup Q_C \cup Q_D]$$
$$= \frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \frac{1}{52} = \frac{1}{13}$$

(ii)

Toss 2 coins : Prob of 1 Head.

HH HT TH TT

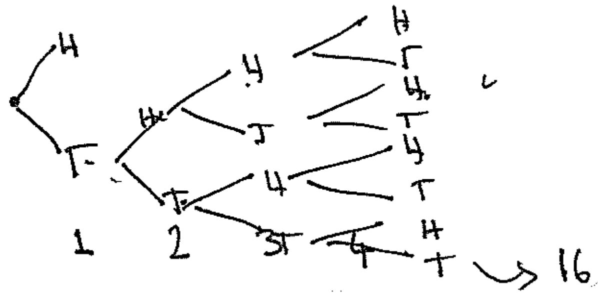
$$P[1 \text{ head}] = P[HT \cup TH] = \cancel{\frac{1}{4} + \frac{1}{4}}$$
$$= P[HT] + P[TH]$$
$$= P(H)P(T) + P(T)P(H)$$
$$= \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{1}{2}$$

Toss 4 coins : Prob 2 Heads

2222 = 16 possible outcomes

ME
HHTT, . .
HTTH
TTHH
THHT
THTH
HTHT

.121



$$\frac{\#\text{FAVORABLE}}{\#\text{POSSIBLE}}$$

~~Example~~ D

(III) Throw 2 DICE   Prob sum = 3

$= \frac{2}{36}$

1	2
2	1
3	
4	
5	
6	
FIRST	SECOND

36 POSSIBILITIES

(IV) Galileo. 3 DICE : PROB SUM = 9 ?
10 ?

1
2
3
4
5
6

$P[9] = \frac{25}{216}$

$P(10) = \frac{27}{216}$

	$=9$	$=10$	
6	1+2+6	1 3 6	6
6	1+3+5	1 4 5	6
3	1+4+4	2 2 6	3
3	2 2 5	2 3 5	6
6	2 3 4	2 4 4	3
1	3 3 3	3 3 4	3

BIRTHDAY PROBLEM

n INDIVIDUALS : ~~WHAT IS THE~~ PROB THAT
AT least 2 have same Birthday

365 DAYS TO CHOOSE FROM.

OF POSSIBLE CONFIGS FOR n BIRTHDAYS

$$= 365^n$$

# of WAYS ALL n BIRTHDAYS ARE DISTINCT.

$$= 365 \cdot 364 \cdots (365-n+1)$$

PROB ALL BIRTHDAYS DISTINCT $= \dfrac{365 \ldots (365-n+1)}{365^n}$

PROB AT LEAST 2 THE SAME $= 1 - \dfrac{365 \ldots (365-n+1)}{365^n}$

n	$P(A)$
4	.016
23	.507
32	~~.507~~ .753

Let's check.

- For bigger Tasks:

Example: gamblers wud throw 3 dice $\Sigma = 9$ vs $\Sigma = 10$

reasoned that $P(9)$ should be $= P(10)$

But deir experiece was suspect

Galilea listed the sample space ($6 \times 6 \times 6$ outcomes)

$P(9) = \frac{25}{216}$ $P(10) = \frac{27}{216}$

If no. of outcomes (N) is very large then innumeration / Listing of outcomes is not practical

So we need help in counting (N) and in counting nov of favourable outcomes. BIRTHDAY PROBLEM N^r ^NC_r

COUNTING METHOD'S ... See sheet (3)

Problem: (from Sheet 3 stuff).

Estimation of 'animal abundance'

eg (lake: how many fish r there)

Step 1: Capture N_1 fish

mark them all

release them

Step 2: later, capture n fish

Check how many are marked - x

Step 3: How to estimate N

Test sample prob. = HYPGV

SAMPLING

$(C_1, \; \ldots \; C_N)$

CHOOSE n items with replacement

$$\# \text{ of CONFIGURATIONS} = N^n$$

OUTCOMES

$\{C_{i_1}\ C_{i_2}\ \ C_{i_n}\}$

2. CHOOSE n items without replacement.

$$\# \text{ of configurations} = N^{(n)}$$

$$= N(N-1)\ldots[N-n+1]$$

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

3. LABELED Items

$\{C_1 \ \ldots \ C_N\}$

N_1 OBJECTS OF TYPE 1

N_2 . . . 2

$N_{\cancel{5}}$ . . . $\cancel{k}$

$N_1 + N_2 \ldots N_{\cancel{k}} = N$

COUNTS ON CARDS WITHOUT REPLACEMENT

$(c_1, \dots, c_N)$ cards.

N_1 of Type 1
N_2 . . . 2

N_K of Type K

DRAW A SET OF $n \le N$ cards WITHOUT REPLACEMENT

# of possible sets of size n cards

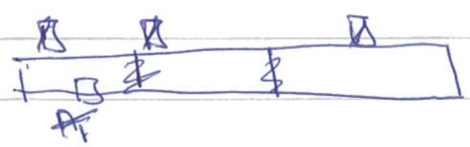
$$= N \dots (N-n+1) = N^{(n)}$$

$$= \frac{N!}{\cancel{n!}(N-n)!}$$

HAND n_1 of Type 1
n_2 . . . 2
n_K K

$$\sum n_k = n$$

C = How many hands with of type $(n_1, n_2, \dots n_K)$
Here the order does not matter



Ty 1

HAND

C

nRR = $n_1!\,n_2! \dots n_K!\; n!$ $C \times$ ARRANGEMENTS = $N_1^{(n_1)} \dots N_K^{(n_K)}$

$$C = \frac{N_1^{(n_1)} \cdots N_K^{(n_K)}}{n_1! \cdots n_K! \; n!}$$

$$C \neq \frac{N_1!}{n_1!(N_1-n_1)! \; n!} \cdots$$

$$\text{Prob of Hand} = \frac{\prod_{k=1}^{K} \frac{N_k!}{n_k!(N_k-n_k)!}}{\frac{N!}{n!(N-n)!}}$$

$$= \frac{\prod_{k=1}^{K} {}^{N_k}C_{n_k}}{\cancel{\prod} \; {}^{N}C_{n}.}$$

~~Sample of~~

Hypergeometric Dist.

$\cancel{\binom{N}{n}}$

ARRANGEMENTS OF $x_1, \ldots x_J$ $\sum x_i = n$ cards.

$n! = x_1! \ldots x_J! \, A$

$A = \dfrac{n!}{x_1! \ldots x_J!}$

$N \to N_1 = S$, $N_2 = N - N_1 = F$

Sample n items without replacement
What is the prob that there will be x of type S

$N^{(n)}$ = # of samples

$N_1^{(x)} N_2^{(n-x)}$ = # of samples in specific order.

$\binom{n}{x}$ orderings.

# of ways to get result $= \dfrac{\binom{n}{x} N_1^{(x)} N_2^{(n-x)}}{\cancel{N^{(n)}}}$

$$\text{Prob.} = \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}}$$

$$\frac{\dfrac{n!}{x!\,(n-x)!} \cdot \dfrac{N_1!}{(N_1-x)!} \cdot \dfrac{N_2!}{(N_2-(n-x))!}}{\dfrac{N!}{(N-n)!}}$$

# of ways to choose x S's and
$(n-x)$ F' in a sequence of length n.

$n!$ # of sequences

$\cancel{\binom{n}{x} \cdot x! \, (n-x)!}$

GIVEN A Sequence with x S and $(n-x)$ F

$x!$ ways to re-order S
$(n-x)!$ - - - F

$n!$ = # of ordered sequences.

$$n! = \binom{n}{x} \cdot x! \, n!$$

$$\binom{n}{x} = \frac{n!}{x! \, (n-x)!}.$$

ESTIMATING THE NUMBER OF FISH IN A LAKE

Suppose there are N fish (of some variety) in the lake.
We (somehow) catch N_1 fish, tag them, and then release them.
Later, we catch n fish, and suppose we find that x of them are tagged.

QUESTION: How can we use the values N_1, n, x to estimate N?

The idea here is to choose as the value of N whatever value produces the largest probability for the observed data. (This approach to estimation is termed the Maximum Likelihood approach).

$$\text{The probability of finding } x \text{ marked fish} = \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}}$$

and we denote this by $L(N)$, where L stands for likelihood.
Our task is to find the value of N for which $L(N)$ is maximum

Consider $\frac{L(N)}{L(N-1)}$. As long as this is >1, the likelihood is increasing.

$$\text{Now } \frac{L(N)}{L(N-1)} = \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}} \Big/ \frac{\binom{N_1}{x}\binom{N-1-N_1}{n-x}}{\binom{N-1}{n}}$$

$$= \frac{(N-N_1)}{(N-N_1)-(n-x)} \frac{(N-n)}{N}$$

This is >1 as long as

$$(N-N_1)(N-n) > (N-N_1-n+x)(N)$$

i.e. as long as

$$N_1 n > N x$$

i.e. as long as

$$N < \frac{N_1 n}{x}$$

And as soon as $N > \frac{N_1 n}{x}$, then $\frac{L(N)}{L(N-1)} < 1$

This suggests that we should estimate N using

$$\hat{N} = \left[\frac{N_1 n}{x}\right] \longleftarrow \text{denotes the integer part.}$$

NOTE: If $\left(\frac{N_1 n}{x}\right)$ is an integer, then $L(\hat{N}) = L(\hat{N}-1)$.

(☐☐☐☐☐) ☐☐☐ 1 2 3 · 3 1 2 · 2 3 1 · 1 3 2 · 2 1 3 · 3 2 1 · 3 2 1

$$\frac{x}{n} = \frac{N_1}{N}$$

If the denominator is ☐☐☐ not ( ) then must count different patterns.

FRIDAY 9-10 WGB33

Example: Deck of 52: select (choose) $n=5$ cards
what is Prob [2 Hearts, 3 non-Hearts]
Here (in a hand of cards) order doesn't matter ... a full house is a full house!!

$$\frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}} = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}} = .2743$$

Example: Deck of 52: select $n=5$ cards
one by one with replacement
what is Prob [2 hearts, 3 non-Hearts]
here each selection is from 52

☆ Couldn't answer this using Combinations
must use ARRANGEMENTS

denominator $= 52^5$: $\boxed{52}\ \boxed{52}\ \boxed{52}\ \boxed{52}\ \boxed{52}$

Now suppose the 2 Hearts r in Positions 1 & 2

$\boxed{13}\ \boxed{13}\ \boxed{39}\ \boxed{39}\ \boxed{39}$

However this count ignores other ~~ARRANGEMENT~~ patterns such as

$\boxed{H}\ \boxed{Non}\ \boxed{Non}\ \boxed{Non}\ \boxed{H}$

The number of such patterns for Hearts & Non-Hearts is $\binom{5}{2}$ or $\frac{{}^5P_5}{3! \times 2!}$ ~~coz~~ not 120 --- think of MISSISSIPPI

For each such pattern we have $13^2 \cdot 39^3$ Arrangements

$\Rightarrow$ Prob (2 H, 3 Non-Hearts) $= \dfrac{\binom{5}{2}\, 13^2 \cdot 39^3}{52^5}$

$= .2638$

If denominator is $\square\square\square$ not $(\ )$ den must a/c for diff patterns $\frac{5!}{2! \times 3!}$.

Example of Binomial prob: $\binom{5}{2}\left(\frac{1}{4}\right)^2($

see sheet 4

Another use of hypergeometric Prob:

LOTTO: For one panel of choosing 6 from 42

What is Prob (4 match)

$$\frac{\binom{6}{4}\binom{36}{2}}{\binom{42}{6}} \approx \frac{1}{555.1}$$

Prob 3 : $\frac{\binom{6}{3}\binom{36}{3}}{\binom{42}{6}} \approx \frac{1}{36.7}$

Industry : statistical quality control → Acceptance Sampling.

Supplizes arrive in Batches of size N
Instead of testing all N
- select a sample of n (from N)
test all n.

let X denote number of defective items

If $x \leq c$ accept the batch (all N)
if $x > c$ reject the batch

The specification of values of n and c is an issue to be decided.

Example : $N = 20$ select $5 = n$ $c = 0$
($c=0$ samples r considered poor)
reject if $x \geq 1$

Find i) P(Acceptance)

let N_1 denote the number of defectives in batch

$N_1 = 10$

$P(X=0) : \frac{\binom{10}{5}\binom{10}{0}}{\binom{20}{5}} = .016$ ✓

cuz we wanted this to be small cuz it was a very bad batch

(.2)(.4)(.3)
(.2)(.6)(.3)
(.8)(.4)(.3)
+ (.2)(.4)(.3) = .024

$N_1 = 2$

$$\cancel{\frac{\binom{10}{2}\binom{10}{3}}{\binom{20}{5}}} = \cancel{.5526}$$

$$\frac{\binom{2}{0}\binom{18}{5}}{\binom{20}{5}} = .5526$$

CONDITIONAL PROB, BAYES THEOREM & Independance

Consider a random expt associated events 2 of which are A & B

Suppose were told that event B has occured

Given this, what is the prob that event A has occured.

This is an example of conditional probability and we denote this by $P(A|B)$

clearly we need to consider the event $A \cap B$ (since for A to occur the event $A \cap B$ must occur

Because B has occured, the collection of possible outcomes is restricted to those outcomes in B.

We view the outcome space as having contracted to those outcomes in B.

Seems reasonable to evaluate $P(A|B)$ by expressing $P(A \cap B)$ as a fraction of $P(B)$

Arising from this discussion we now <u>define</u> the conditional prob of A given B as follows

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{provided } P(B) > 0$$

know dis off

FORMAL: $P(\cdot|B)$ is a Pr. $P(B)>0$: (i) (ii) $P[A^c|B] = \frac{P(A^c \cap B)}{P(B)}$ (iii)

Examples : 1) Pair of dice : score (T)

Suppose we're told that the score is odd
event B

$$P(T \leq 7 \mid T \text{ is odd})$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(3 \text{ or } 5 \text{ or } 7)}{P(3 \text{ or } 5 \text{ or } 7 \text{ or } 9 \text{ or } 11)}$$

2	3	4	5	6	7
3	4	5	6	7	
4	5	6	7		
5	6	7			
6	7				
7					

$$\frac{12/36}{\frac{2+4+6+4+2}{36}}$$

$$= \frac{2}{3} = \frac{12}{18}$$

2) Newspaper A ← 60% of Adults
Newspaper B ← 20% of Adults
7% --- Both.

An adult is randomly chosen we ask if he reads B.

$$P(\text{read } A \mid \text{read } B) : P(A|B)$$

	B	B^c	
A	.07	.53	.6
A^c	.13	.27	.4
	.2	.8	1

$$\frac{P(A \cap B)}{P(B)} = \frac{.07}{.2} = .35$$

- multiplication LAW of prob

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\Rightarrow \quad P(A \cap B) = P(B)\,P(A|B)$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

$$\Rightarrow \quad P(B \cap A) = P(A)\,.\,P(B|A)$$

$$P(B \cap A) = P(A \cap B)$$

$$P(A)\,P(B|A) = P(B)\,.\,P(A|B) = P(B \cap A) = P(A \cap B)$$

from example 2).

$$P(\text{Person reads Both } A, B) = \underset{.07}{P(A \cap B)} = \underset{.2}{P(A|B)}\,.\,\underset{.35}{P(B)}$$

3) Select 2 cards (one after another) from 52

$$P(K_1 \cap Q_2) = P(K_1 | Q_2)\,.\,P(Q_2)$$
$$= P(Q_2 | K_1)\,.\,P(K_1)$$

- see here we don't [illegible] [illegible] se. 2nd one is [illegible] useful one usually useful ! one not

$$= \left(\frac{4}{51}\right)\left(\frac{4}{52}\right)$$

	B	B^c	
A	.07	.53	.6
A^c	.13	.27	.4
	.2	.8	1

$$\frac{P(A \cap B)}{P(B)} = \frac{.07}{.2} = .35$$

- multiplication LAW of prob

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\Rightarrow \quad P(A \cap B) = P(B)P(A|B)$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

$$\Rightarrow \quad P(B \cap A) = P(A).P(B|A)$$

$$P(B \cap A) = P(A \cap B)$$

$$P(A).P(B|A) = P(B).P(A|B) = P(B \cap A) = P(A \cap B)$$

from example 2).

$$P(\text{Person reads Both } A, B) = \underset{.07}{P(A \cap B)} = \underset{.2}{P(A|B)}.\underset{.35}{P(B)}$$

3) Select 2 cards (one after another) from 52

$$P(K_1 \cap Q_2) = P(K_1|Q_2).P(Q_2)$$
$$= P(Q_2|K_1).P(K_1)$$

- see here we don't care [illegible] se. 2nd one is a useful one . usually 1 useful 1 one not

$$= \left(\frac{4}{51}\right)\left(\frac{4}{52}\right)$$

$P(K, Q \text{ any order})$.

$\boxed{K}\ \boxed{Q}$ or $\boxed{Q}\ \boxed{K}$

$$\left(\frac{4}{52}\right)\left(\frac{4}{51}\right) + \left(\frac{4}{52}\right)\left(\frac{4}{51}\right)$$

or

$$\frac{\binom{4}{1}\binom{4}{1}\binom{44}{0}}{\binom{52}{2}}$$

4) 2 child families

$\frac{1}{4}$	BB
$\frac{1}{4}$	BG
$\frac{1}{4}$	GB
$\frac{1}{4}$	GG

$\boxed{2}\ \boxed{2}$

Event B: There is a boy in the family

$$P(BB \mid B) = \frac{P(BB \cap B)}{P(B)} \quad \frac{1/4}{3/4} = \frac{1}{3}$$

'If' $\Rightarrow P(A \mid B)$.

5) Disease.

Disease D has a .5% incident rate

Sympton S is associated with D as follows:

- 96% of those with D show S
- 5% of those without D show S

False Positive : $S \cap D^c$ - symptoms but no Disease

good ✓ - positive

False Negative : $S^c \cap D$ disease but no Sym

bad - negative - den just add false

P (person shows S)

$P(S) = P[(\text{Symptom \& disease}] \cancel{\text{or}} + P[\text{Symptom \& no disease}]$

$= P[S \cap D] + P[S \cap D^c]$

True Positive   False Positive

$= P(S|D).P(D) + P(S|D^c).P(D^c)$

$= (.96)(.005) + (.05)(.995)$

– cud ave $P(D|S).P(S)$ but this is no use to us (don't know $P(S)$) so don't use dis one. —

An important question is :

opposite

$P(D|S)$

$P(D|S) = \dfrac{P(D \cap S)}{P(S)} = \dfrac{P(S|D).P(D)}{P(S)}$ -- use dis one cuz other term no use.

$= \dfrac{(.96)(.005)}{(.96)(.005) + (.05)(.995)} = .088$

so this test is not actually dat accurate cuz only 8.8% accurate. [redacted]

opposite

$P(D^c|S) = 1 - .088$

or

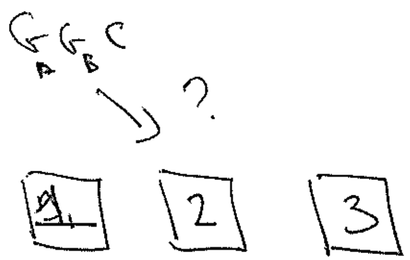
$= \dfrac{P(D^c \cap S)}{P(S)} = \dfrac{P(S|D^c).P(D^c)}{P(S)} = \dfrac{(.05)(.995)}{.05455}$

opposite

$P(D^c|S^c) = \dfrac{P(D^c \cap S^c)}{P(S^c)} = \dfrac{P(S^c|D^c).P(D^c)}{P(S^c)}$

$= \dfrac{(.95)(.995)}{1 - .05455} = .999788$

MONTE HALL : ADDITIONAL INFORMATION CAN CHANGE ODDS.

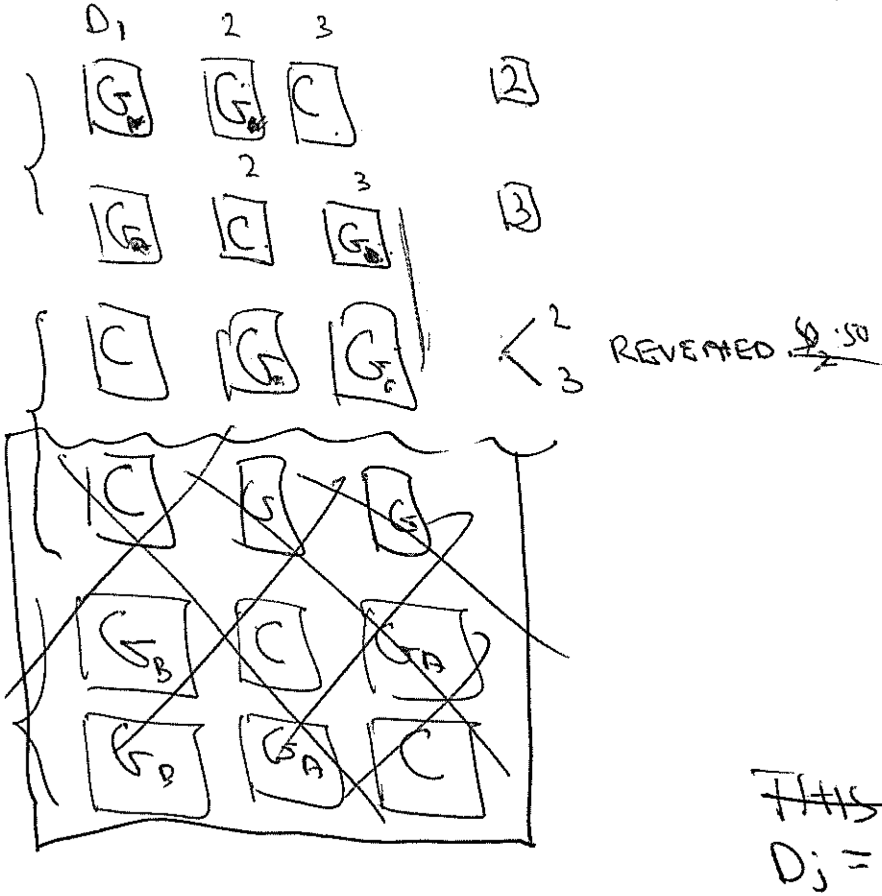


STEP 1 : PICK A DOOR

STEP 2 : OF DOORS NOT PICKED AT LEAST ONE DOES NOT HAVE A CAR. ONE IS REVEALED

- SHOULD PICK BE ALTERED BASED ON STEP 2 INFO?

OUTCOMES (DOOR 1 PICKED)



$$P[\text{WIN NO CHANGE}] = \frac{2}{6} = \frac{1}{3}$$

$$P[\text{WIN W/ CHANGE}] = \frac{4}{6} = \frac{2}{3}$$

~~P[WIN CHANGE AND WITH 2 REVEALED]~~ =

DOUBLE CHANCES OF WINNING. !

~~THIS IS NOT~~ $P[W \text{ with switch} \mid D_1]\, P[D_1] + \ldots = 0\tfrac{1}{3} + 1\tfrac{1}{3} + 1\tfrac{1}{3}$

$1\tfrac{1}{3} + 0\tfrac{1}{2} + 0\tfrac{1}{3}$

D_j = CAR BEHIND DOOR j $\quad P[W \text{ without} \mid D_1] = 1 \quad P[W \text{ without} \mid D_2] = 0$

Note: $P(D)$ may be thought of as a prior probability (prior to the test result)

$P(D|S)$ is termer a Posterior probability (if after the test result is known)

We have $P(D) = .005$ || $P(D^c) = .995$

$P(D|S) = .088$ || $P(D^c/S^c) = .999788$

BAYES THEOREM

Here we were using Bayes' Theorem

Based on a simple derivation from conditions Prob.

Theorem: Consider event A and a partition (of Sample Space) $\{H_j\}_{j=1,2,\dots,n}$

Then $P(H_k|A) = \frac{P(A|H_k).P(H_k)}{P(A)}$

$$= \frac{P(A|H_k).P(H_k)}{\sum_{i=1}^{n} P(A|H_i)\,P(H_i)}$$

like $P(S)$ in last example

Proof: i) $P(H_k|A) = \frac{P(H_k \cap A)}{P(A)}$

$$= \frac{P(A|H_k).P(H_k)}{P(A)}$$

ii) $P(A) = P(A \cap \Omega)$

$$= P[A \cap (\bigcup_{i=1}^{n} H_i)]$$

$$= P[\bigcup_{i=1}^{n} (A \cap H_i)]$$

$$= \sum^{n} P(A \cap H_i)$$

$$P(A) = \sum_{i=1}^{n} P(A|H_i) \cdot P(H_i)$$ - Another version of LAW OF TOTAL PROBABILITY

$$= \frac{P(A|H_k)\ P(H_k)}{\sum_{i=1}^{n} P(A|H_i)\ P(H_i)}$$

Example: 10 dice, Take that 4 are crooked ($\to$ 6 with prob .3)

Suppose a dice is picked out (from 10) tossed, gives a 6

crooked 6 obtained.

Find $P(C/6)$

$$P(C/6) = \frac{P(C \cap 6)}{P(6)} = \frac{P(6/C)\ P(C)}{P(6)}$$

* 6 out of 10

Need $P(6) = P(6/C) \cdot P(C) + P(6/\text{straight})\ P(\text{straight})$

$= (.3)(.4) + (1/6)(.6)^*$

$.22$

$\Rightarrow P(C/6) = \frac{.12}{.22}$

$= .545445$

NOTE:

IV $P(A|B) = 1 - P(\bar{A}|B)$

eg. VO 5 Pros set 2

P(gives correct Ans / Knows) = ~~?~~ 1 - P(gives wrong Ans / Knows)

1 - 0 = 1 - (0)

B part must be same.

STATISTICAL INDEPENDANCE

We say event A is independand (not influenced by) of B if

$$P(A|B) = P(A)^{*}$$

Note:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$\Rightarrow$ event A is independant of event B if

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

$$\Rightarrow \text{ if } \quad P(A \cap B) = P(A).P(B)^{*}$$

Note: This condition is equivilant to

$$\frac{P(A \cap B)}{P(A)} = P(B) \qquad [P(A) > 0]$$

$$P(B|A) = P(B)$$

(Thus if A is independent of B, B is independant of A)

Quite often, independance of events A, B is defined as

Definition: Events A, B are independant iff

$$P(A \cap B) = P(A).P(B).$$

Compare this with

$$P(A \cap B) = P(A/B).P(B)$$
$$= P(B/A).P(A)$$

Examples : of independance

Group of 100 students

Arrange into Freq table.

	F	m	
S	24	11	35
NS	26	39	65
	50	50	

Pick at radom a person

Consider $P(S \& F) = P(S \cap F) = \frac{24}{100}$

$P(S) = \frac{35}{100}$

$P(F) = \frac{50}{100}$

We note $P(S \cap F) \neq P(S)\,P(F)$

so smoking is dependent on weather your female

$P(S|F) = \frac{P(S \cap F)}{P(F)} = \frac{P(F|S)\,.\,P(S)}{P(F)} = \frac{24}{50} = .48$

$P(S|F) \neq P(S)$ so

smoking and female are dependant events

Suppose :

	F	m
S	21	14
NS	39	26
	60	40

3 Things : i) $P(S|F) = P(S)$ ii) $P(F|S) = P(F)$ iii) $P(S \& F) = P(S)\,P(F)$

Note $P(S \cap F) = \frac{21}{100} = .21$

$P(S) = .35$; $P(F) = .6$

$P(S \cap F) = P(S) . P(F)$

so

$P(S/F) = \frac{21}{60} = .35$

$P(S/F) = P(S)$

These demonstrate that events S, F are independant

Note: Usually, we use the concept of independance by reasoning that independance of events is present.
- Then we can use

$$P(A \cap B) = P(A) . P(B)$$
$$P(A/B) = P(A)$$

Example: Consider 2 horse races

Race 1 : Bet on Horse A (suppose $P(A \text{ wins}) = 1/5$)

Race 2 : Bet on Horse B (suppose $P(B \text{ wins}) = 1/10$)

Accumulator bet: Collect winnings if both win

event $A \cap B$

Want $P(A \cap B) = ? = P(A) . P(B)$

If there is independance

$= \frac{1}{5} . \frac{1}{10} = \frac{1}{50}$

Example : Reliability of complex systems.

1) A system with 10 components
reliability of component = P (Stays workin in specified T)
= .95

$$P(\text{System Fails}) = 1 - P[\text{System works}]$$
$$1 - P[\text{all 10 components don't fail}]$$
If indep. assumed
$$= 1 - P[C_1 \cap C_2 \cap \dots \cap C_{10}]$$
$$= 1 - [P(C_1) . P(C_2) \dots . P(C_{10})]$$
$$= 1 - (.95)^{10}$$
$$= .4$$

2) A system with 1000 components.
reliability of components = .999

$$P(\text{Fails}) = 1 - (.999)^{1000}$$
$$= .6323$$

3) 1000 comp
.9999
$$P(\text{Fails}) = 1 - (.9999)^{1000}$$
$$= .095$$

Pairwise Independance

For 2 events, we know that

$$P(A \cap B) = P(A).P(B)$$

is our condition for independance

lets now consider events A, B, C

Suppose we have

$$\left.\begin{aligned} P(A \cap B) &= P(A).P(B) \\ P(A \cap C) &= P(A).P(C) \\ P(C \cap B) &= P(C).P(B) \end{aligned}\right\}$$ If all true events A, B, C are Pairwise independant

Can we say events A, B, C r <u>fully</u> independant

NO!

Suppose we have n events $E_1, E_2, \ldots, E_n$
we define mutual independance for these n events

If $P[E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_m}] = P[E_{i_1}].P[E_{i_2}]. \ldots .P[E_{i_m}]$

for all subsets of size m from the n events
for $m = 2$ or 3 or .... or n

for $n = 3$, we need 1 more condition along with 3 already given:

$$P(A \cap B \cap C) = P(A).P(B).P(C)$$

If all 4 are true, we have full indepence

Pairwise

Example: Consider expt with 4 outcomes

$\Omega \leftrightarrow \{O_1, O_2, O_3, O_4\}$ (equally likely)

lets define some events $E_1 : \{O_1 \overset{or}{,} O_2\}$

$E_2 : \{O_1 \overset{or}{,} O_3\}$

$E_3 : \{O_1 \overset{or}{,} O_4\}$

$P(E_1) = P(E_2) = P(E_3) = \frac{1}{2}$

Also : $E_1 \cap E_2 = O_1$

$E_1 \cap E_3 = O_1$

$E_2 \cap E_3 = O_1$

$E_1 \cap E_2 \cap E_3 = O_1$

$P(E_1 \cap E_2) = P(O_1) = 1/4 = P(E_1).P(E_2)$

$P(E_1 \cap E_3) = 1/4 = P(E_1).P(E_3)$

$P(E_2 \cap E_3) = 1/4 = P(E_2).P(E_3)$

Thus the events E_1, E_2, E_3 pairwise independant

But $P(E_1 \cap E_2 \cap E_3) = P(O_1) = 1/4$

$\neq P(E_1).P(E_2).P(E_3)$

So events are not mutually independant.

tion

The Binomial Distribution

Consider a r. expt with 2 outcomes S, F
Suppose this is repeated n times
denote $P(S) = p$
$P(F) = Q = (1-p)$

We assume the n repetitions are mutually independent & that P(S) remains same for all n

We're interested in X, denoting the number of successes in the n trials
(Terminology: Each repitition is called a Bernoulli trial
=> an exp with just 2 outcomes)

Note: X is our first example of a random variable (more later)

Possible values for X: 0, 1, 2, ..., n

We next try to find $P(X=k)$ [k successes]

- The notation $X=k$ means getting k successes
getting n-k failures.

There are quite a few orderings that produce $(X=k)$
eg. $\underbrace{SSS\ldots S}_{k}\ \underbrace{FFF\ldots F}_{n-k}$ ← Arrangements

$$P(X=k) = P(A_1 \cup A_2 \cup \ldots \cup A_n\ldots)$$
$$= \sum_{i=1} P(A_i)$$

Lets find $P(A_1)$

$$P(A_1) = P(SSS\ldots SFFF\ldots F)$$
$$= \underbrace{P(S)\,P(S)\ldots P(S)}_{k}\,\underbrace{P(F)\,P(F)\ldots P(F)}_{n-k} \quad \ldots \text{since independence}$$

<u>BIRTHDAY PROBLEM</u>

n INDIVIDUALS : ~~WHAT IS THE~~ PROB THAT
AT least 2 have same birthday

365 DAYS TO CHOOSE FROM.

OF POSSIBLE CONFIGS FOR n BIRTHDAYS

$$= 365^n$$

# of WAYS ALL n BIRTHDAYS ARE DISTINCT.

$$= 365 \cdot 364 \cdots (365-n+1)$$

PROB ALL BIRTHDAYS DISTINCT $= \dfrac{365 \ldots (365-n+1)}{365^n}$

PROB AT LEAST 2 THE SAME $= 1 - \dfrac{365 \ldots (365-n+1)}{365^n}$

n	$P(A)$
4	.016
23	~~.484~~ .507
32	~~.507~~ .753

<u>Let's Check.</u>

$$= \left[P^k(S)\right]^k \left[P(F)\right]^{n-k}$$

$$= P^k (Q)^{n-k} \quad = P^k \left[1-P\right]^{n-k}$$

Similarly :

$P(A_2) = P^k Q^{n-k}$ cuz S occurs k times

F occurs n-k times

For all other A:

$$P(x=k) = \sum P^k Q^{n-k}$$

How many Arrangements ? $\binom{n}{k}$

=>

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

$\sum_k P_k = 1$

Connection wit binomial in algebra

$$(Q+P)^n = \sum_{k=0}^{n} \binom{n}{k} P^k Q^{n-k}$$

$\sum_k C P$

$\sum_k \binom{n}{k} p^k (1-p)^{n-k} \frac{1}{1-1} = 1$

Probability Distribution : This specification of probability values is called Prob. dist for a random variable X.

Example: Manufacturing process
Sample of n items each hour
test all n, count # failure (x)

Rule : Say n = 20
If $x \leq 3$ carry on
If $x > 3$ stop

We suppose various values for P (Prob of failure)

$P = .05$ From tables $P(X \geq 4) \Rightarrow P(X > 3) = .0159$

$P = .1$ $P(X \geq 4) = .1330$

Example of binomial:

SPC suppose of size 20

Test all 20 (some acceptable

some not

$X =$ # of ~~good~~ bad. items

defective items

Model: 20 Bernoulli trials (each with P(success) = P

Assuming independence and that P remains const.

Then X is Binomial $(b(n, p))$

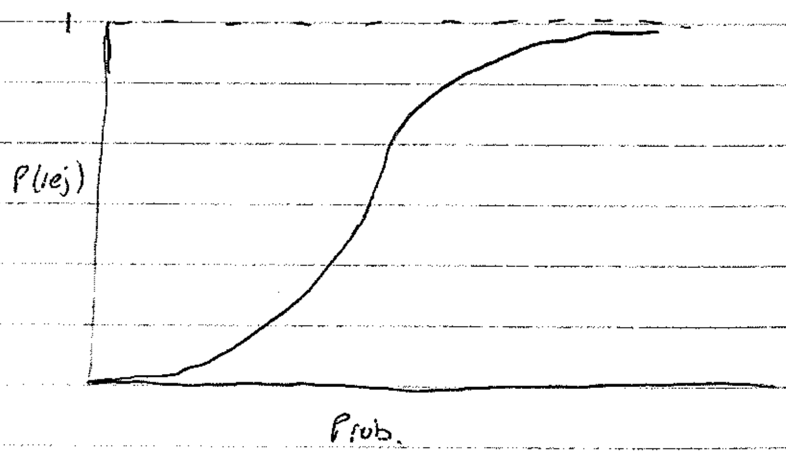
Suppose our decision rule: OK if $X \leq 3$

Draw the operating characteristic

↓

Plot of P(rejection) against value of P.

P	Prob $(X \geq 4)$
.05	.0159
.1	.133
.15	.3523
.2	.5886
.25	.7748



Random Variables

Design matter. Find n, to give a "good shape" to curve.

A random variable is a variable whose value is determined (to some extent) by chance

or

A random variable is a real-valued function defined on a sample space

ob Example : Game of chance :

Pay €1 to enter.

Die is throw, get €1 back if 1
get €5 back if 6
2, 3, 4, 5, no pay

Sample space : 1 2 3 4 5 6

Determine the net gain from 1 play.

1 → net gain = 0
2, 3, 4, 5 → net gain = −1
6 → net gain = 4

G	-1	0	4
Prob	$\frac{4}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

– An example of a discrete probability distribution

Discrete → value of $G \in \mathbb{Z}$

Note : A probability distribution is a specification of how probabilities are distributed over all the possible values of a random variable.

We start with discrete random variables (and the associated distributions are called discrete distributions.

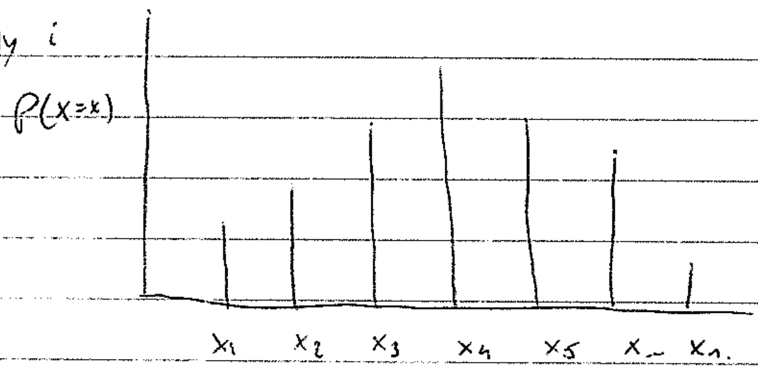
Terminology:

1) Probability functions : This is the particular specification of a discrete prob distribution in which we provide $P(X = x_i)$ for all possible values of the random variable X

X	x_1	x_2	x_3	---	x_n
Prob	p_1	p_2	p_3	---	p_n

← The probability function for X

Graphically :



eg Binomial :

X	0	1	----	r	
Prob	$(1-p)^n$	$np(1-p)^{n-1}$		$\binom{n}{r} p^r (1-p)^{n-r}$	

NOTE :

1) $p_i \geq 0$

2) $\sum p_i = 1$

Next we look at a series of independant random variables :

1) <u>Bernoulli random variable</u> (X)

Associated with a bernoulli Trial $X=1$,
$X=0$

X	0	1
Prob	$1-p$	p

2) Binomial Random Variable (X)

we have n bernoulli trials

a) 2 outcomes S, F : Denote $P(S)$ by p

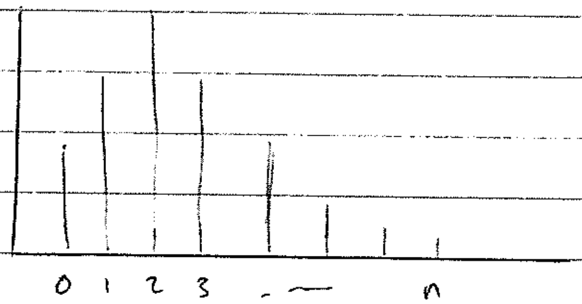
b) Trials are independant

c) p remains constant

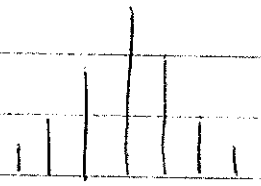
$X = \#$ Successes

X is $b(n, p)$

Shape depends on p



If $P = \frac{1}{2}$



3) Geometric Random Variable

Sequence of Bernoulli Trials

let X = # trials up to 1^{st} success

Values for $X = 1, 2, 3, \ldots, n, \ldots, \infty$

$$P(X=k) = P[\underbrace{FFF\ldots F}_{k-1} S]$$

$$= (1-P)^{k-1} P.$$

Easy to check $\sum_{k=1}^{\infty} P(X=k) = 1$

Prob set 3 Q.2. using binomial

$P(X=1) \to n(.5)^n$

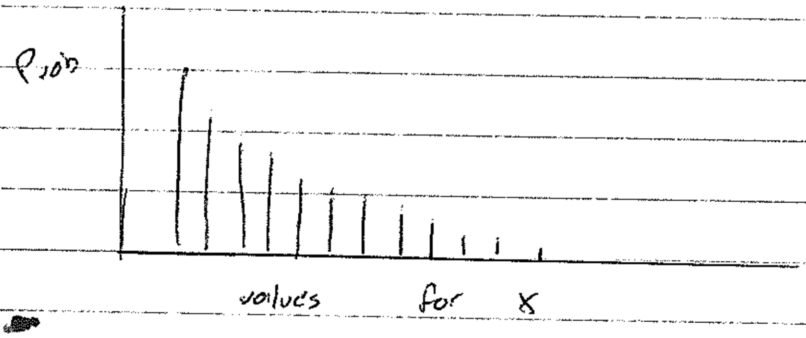
= Prob of getting one head in n trials

could get it on the first try $2^{nd}, \ldots, n^{th}$.

using geometric:

$P(X=n) = (.5)^n$

= Prob of getting one head in n trials but getting it on n^{th} try only.



4) Negative Binomial Random Variable (Pascal)

Bernoulli sequence of trials

Y = # trials up to r^{th} success

ie like how many trials till 4^{th} success.

Values : $r, r+1, r+2, \ldots, \infty$

if $r=1$ Neg Binomial $\to$ Geometric.

$P(Y=k)$ = Prob [$(r-1)$ successes in 1^{st} $(k-1)$ trials $\cap$ success on k^{th} trial]

cuz independant :-

$= P[(r-1) \text{ success in } (k-1) \text{ trials}] \cdot P[\text{success in } k^{th} \text{ trial}]$

$$\binom{k-1}{r-1} p^{r-1} (1-p)^{(k-1)-(r-1)} \cdot p$$

$$\binom{k-1}{r-1} p^{r} (1-p)^{k-r}$$

Note: compare Ans wit $X \sim b(n,p)$
for what $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

If $r=1$ we are back to geometric.

Negative Binomial arises from:

Note: $\binom{-r}{j} = \dfrac{(-r)(-r-1)\dots(-r-j+1)}{j!} = \dfrac{(-1)^j r(r+1)\dots(r+j-1)}{j!}$

$$= \frac{(-1)^j (r+j-1)(r+j-2)(\dots(r+1)(r)}{j!}$$

$$= (-1)^j \binom{r+j-1}{j} \; (**) = (-1)^j \binom{r+j-1}{r-1}$$

we need it cuz: $(1+x)^{-r} = \sum_{j=0}^{\infty} \binom{-r}{j} x^j$

Name comes from:

$$\sum_{k=r}^{\infty} P(Y=k) = \sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r (1-p)^{k-r} = 1$$

$$p^r \sum_{k=r}^{\infty} \binom{k-1}{r-1} (1-p)^{k-r}$$

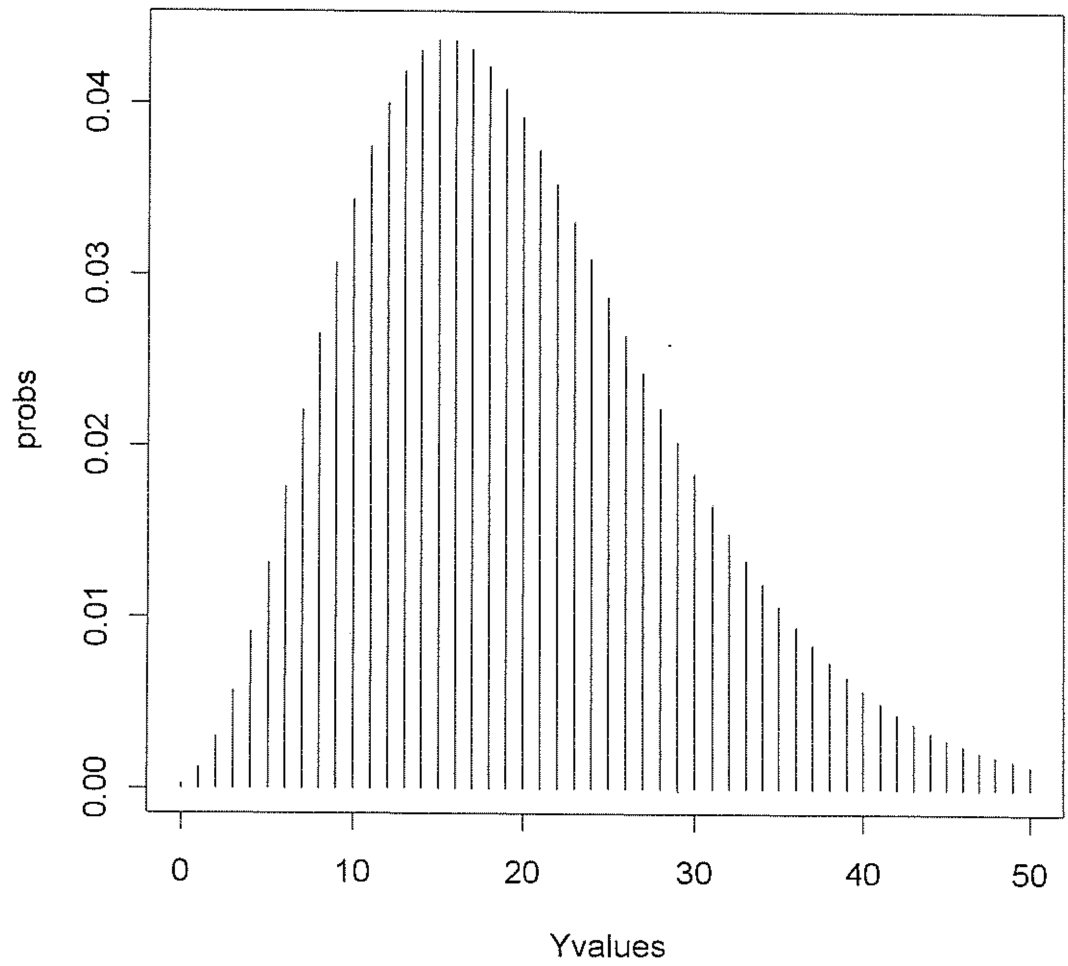
change index : let $j = k-r$ [as k runs from $r \to \infty$
j runs from $0 \to \infty$

PLOTTING THE SHAPE OF THE NEGATIVE BINOMIAL DISTRIB.

```
R : Copyright 2004, The R Foundation for Statistical Computing
Version 2.0.1  (2004-11-15), ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]
```

START

```
> Yvalues=0:50
> probs=dnbinom(Yvalues,5,0.2)
> probs
 [1] 0.000320000 0.001280000 0.003072000 0.005734400 0.009175040 0.013212058
 [7] 0.017616077 0.022145925 0.026575110 0.030709016 0.034394098 0.037520834
[13] 0.040022223 0.041869403 0.043065671 0.043639880 0.043639880 0.043126470
[19] 0.042168104 0.040836480 0.039203021 0.037336210 0.035299689 0.033151013
[25] 0.030940945 0.028713197 0.026504490 0.024344865 0.022258162 0.020262603
[31] 0.018371426 0.016593546 0.014934192 0.013395517 0.011977169 0.010676790
[37] 0.009490480 0.008413182 0.007439024 0.006561601 0.005774209 0.005070037
[43] 0.004442318 0.003884446 0.003390062 0.002953120 0.002567931 0.002229182
[49] 0.001931958 0.001671735 0.001444379
>
> sum(probs)
[1] 0.9913492
> plot(Yvalues,probs,"h")
```



$$= p^r \sum_{k=r}^{\infty} = p^r \sum_{j=0}^{\infty} \binom{k-1}{j} (1-p)^j \quad (*)$$

Since $j = k-r$, then $k = j + r$

$$(*) \Rightarrow p^r \sum_{j=0}^{\infty} \binom{j+r-1}{j} (1-p)^j$$

$$= p^r \sum_{j=0}^{\infty} (-1)^j \binom{-r}{j} (1-p)^j \quad \text{using } (**)$$

$$p^r [1-(1-p)]^{-r}$$

$$= p^r [p]^{-r} = 1 \quad \checkmark$$

what does it look like?
see sheet.

CONNECTION BETWEEN NEG Binomial & (Ord) Binomial

Sequence of Bernoulli Trials

- Think of 1st n trials
  let X_n = # successes in 1st n trials
  Y_r = # Trials needed to get to r^{th} success

Consider event $(X < r)$ it must be that $(Y > n)$

i.e. $(X < r) \Rightarrow Y > n$

Consider event $(Y > n)$. If $Y > n$ occurs then $(X < r)$

CONNECTION WITH BINOMIAL

$X_n =$ # of successes in n trials

$Y_r =$ # of trials to get r successes.

EVENTS

$E = [X_n \leq r]$ = r OR Fewer successes in n trials.

$F = [Y_r \geq n]$ = r OR Fewer successes in n Trials.

$E = F$

$$P(E) = \sum_{k=0}^{r} \binom{n}{k} p^r q^{n-r}$$

$$P(F) = \sum_{k=n}^{\infty} \binom{k-1}{r-1} p^r q^{k-r}$$

E easier to calculate

Example

Thus the event $(X < r)$ and $(Y > n)$
are essentially the same.

Then $P(X < r) = P(Y > n)$

Alternatively: $P(X \geq r) = P(Y \leq n)$

$\downarrow$

In binomial
tables.

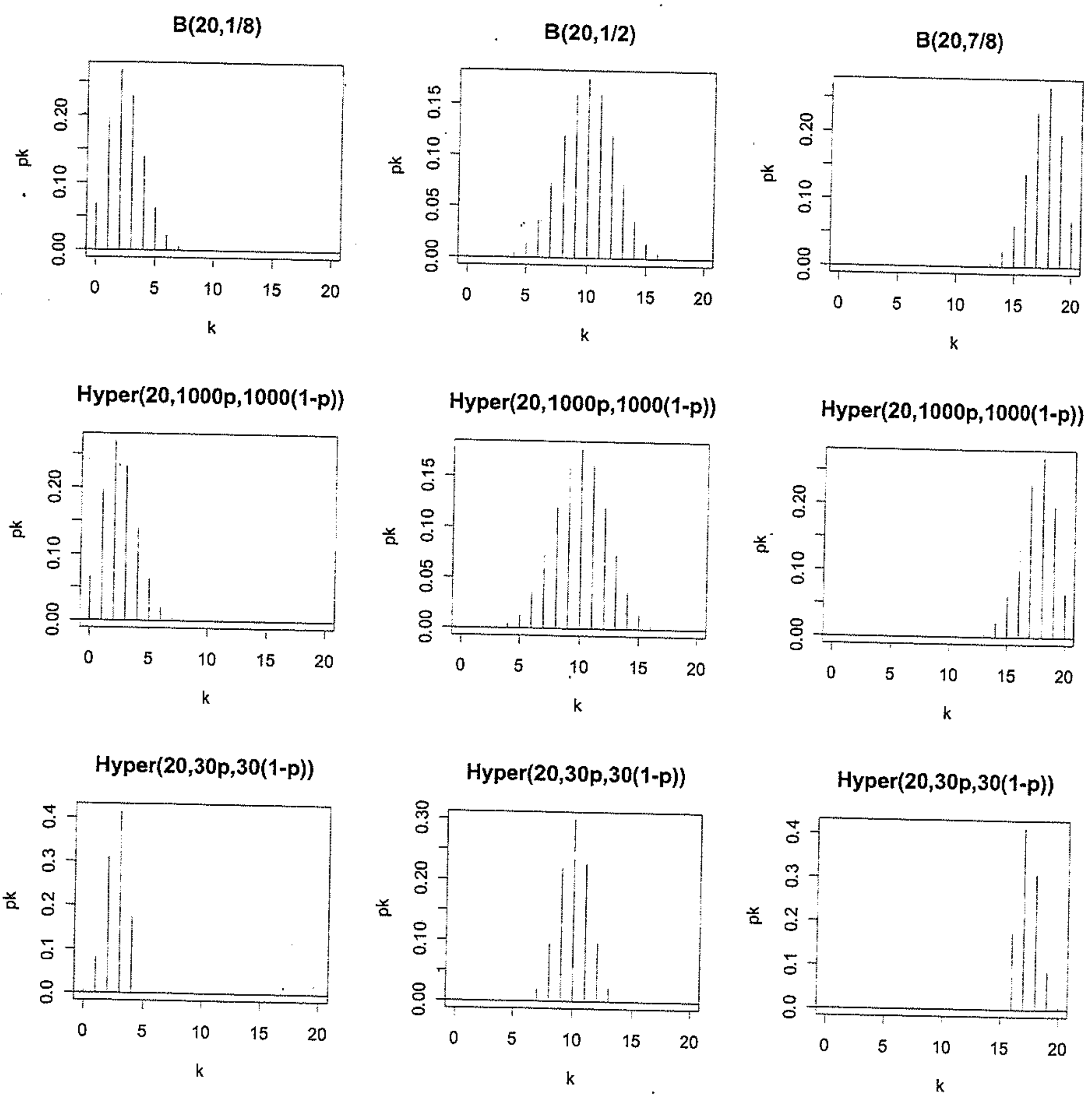
We could use these to determine the neg binomials

Example: Seq of bernoulli Trials with $p = .2$
Find $P[Y > 20]$ for $r = 5$.

$$P(Y_5 > 20) = 1 - P(Y_5 \leq 20) = 1 - P(X_{20} \geq 5)$$
$$bi(20, .2)$$

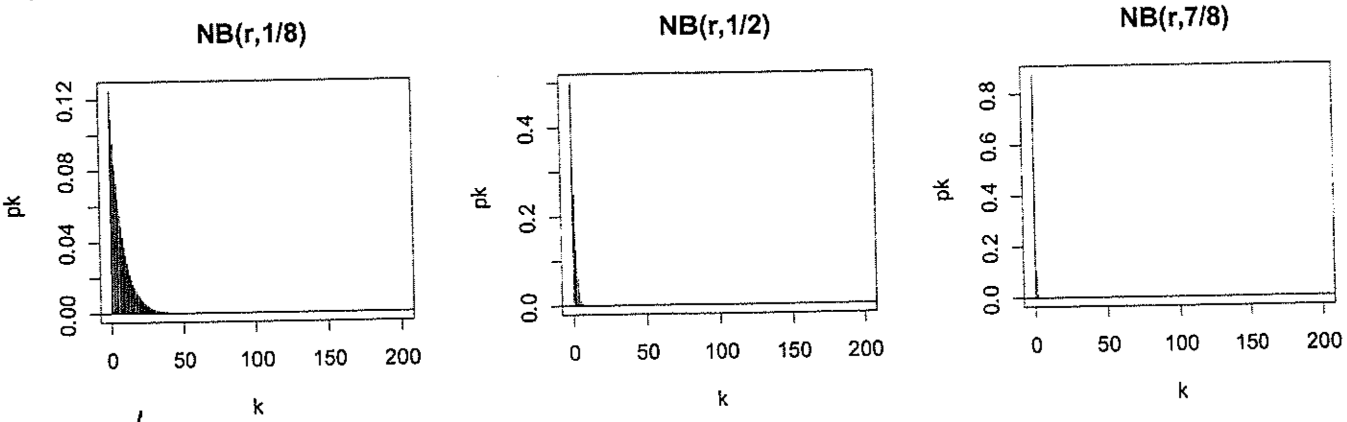
$$1 - .3704$$
$$.6296$$

Binomial and Hypergeometric.

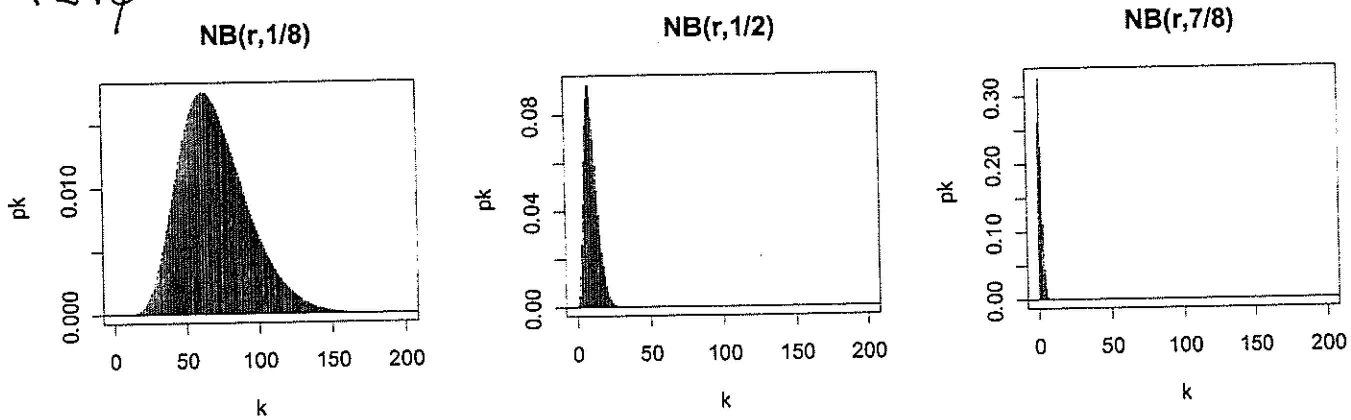


NEGATIVE BINOMIALS

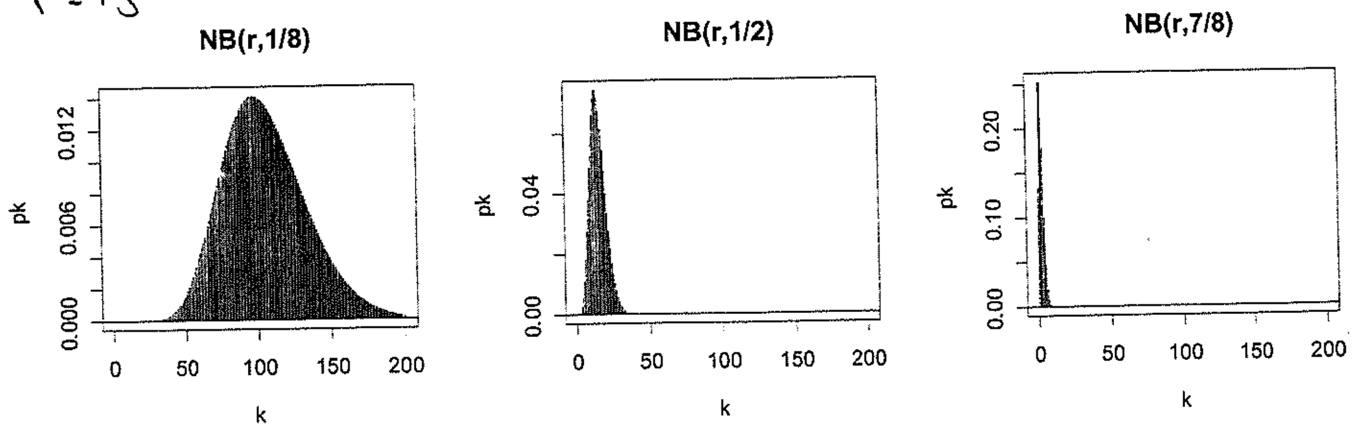
r = 1



r = 10



r = 15



~~Binomials~~

```
# ST2054 Binomial/Geometric/Negative Binomials

# 1. Compare Binomial and Hypergeometric

par(mfrow=c(3,3))
n=20 ; k=c(0:n)
p=.125;pk = dbinom(k, n, p);plot(k,pk,type="h",main="B(20,1/8)");abline(h=0)
p=.5;pk = dbinom(k, n, p);plot(k,pk,type="h",main="B(20,1/2)");abline(h=0)
p=.875;pk = dbinom(k, n, p);plot(k,pk,type="h",main="B(20,7/8)");abline(h=0)

N=1000
p=.125;N1=round(N*p);pk = dhyper(k,
N1,N-N1,n);plot(k,pk,type="h",main="Hyper(20,1000p,1000(1-p))");abline(h=0)
p=.5;N1=round(N*p);pk = dhyper(k,
N1,N-N1,n);plot(k,pk,type="h",main="Hyper(20,1000p,1000(1-p))");abline(h=0)
p=.875;N1=round(N*p);pk = dhyper(k,
N1,N-N1,n);plot(k,pk,type="h",main="Hyper(20,1000p,1000(1-p))");abline(h=0)

N=30
p=.125;N1=round(N*p);pk = dhyper(k,
N1,N-N1,n);plot(k,pk,type="h",main="Hyper(20,30p,30(1-p))");abline(h=0)
p=.5;N1=round(N*p);pk = dhyper(k,
N1,N-N1,n);plot(k,pk,type="h",main="Hyper(20,30p,30(1-p))");abline(h=0)
p=.875;N1=round(N*p);pk = dhyper(k,
N1,N-N1,n);plot(k,pk,type="h",main="Hyper(20,30p,30(1-p))");abline(h=0)


# 2. Negative Binomials

r=1    # Geomertric
n=200 ; k=c(0:n)
p=.125;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,1/8)");abline(h=0)
p=.5;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,1/2)");abline(h=0)
p=.875;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,7/8)");abline(h=0)

r=10
n=200 ; k=c(0:n)
p=.125;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,1/8)");abline(h=0)
p=.5;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,1/2)");abline(h=0)
p=.875;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,7/8)");abline(h=0)


r=15
n=200 ; k=c(0:n)
p=.125;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,1/8)");abline(h=0)
p=.5;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,1/2)");abline(h=0)
p=.875;pk = dnbinom(k, r, p);plot(k,pk,type="h",main="NB(r,7/8)");abline(h=0)
```

NegBinomial package:base R Documentation

The Negative Binomial Distribution

Description:

Density, distribution function, quantile function and random generation for the negative binomial distribution with parameters 'size' and 'prob'.

Usage:

```
dnbinom(x, size, prob, mu, log = FALSE)
pnbinom(q, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
qnbinom(p, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
rnbinom(n, size, prob, mu)
```

Arguments:

x: vector of (non-negative integer) quantiles.

q: vector of quantiles.

p: vector of probabilities.

n: number of observations. If 'length(n) > 1', the length is taken to be the number required.

size: target for number of successful trials, or dispersion parameter (the shape parameter of the gamma mixing distribution).

prob: probability of success in each trial.

mu: alternative parametrization via mean: see Details

log, log.p: logical; if TRUE, probabilities p are given as log(p).

lower.tail: logical; if TRUE (default), probabilities are P[X <= x], otherwise, P[X > x].

Details:

The negative binomial distribution with 'size' = n and 'prob' = p has density

p(x) = Gamma(x+n)/(Gamma(n) x!) p^n (1-p)^x

for x = 0, 1, 2, ...

This represents the number of failures which occur in a sequence of Bernoulli trials before a target number of successes is reached.

A negative binomial distribution can arise as a mixture of Poisson distributions with mean distributed as a gamma ('pgamma') distribution with scale parameter '(1 - prob)/prob' and shape parameter 'size'. (This definition allows non-integer values of 'size'.) In this model 'prob' = 'scale/(1+scale)', and the mean is 'size * (1 - prob)/prob'

The alternative parametrization (often used in ecology) is by the _mean_ 'mu', and 'size', the _dispersion parameter_, where 'prob' = 'size/(size+mu)'. In this parametrization the variance is 'mu + mu^2/size'.

If an element of 'x' is not integer, the result of 'dnbinom' is zero, with a warning.

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

$n = 100$
$p = .0278$

k	Binomial Probability	Poisson Approximation $\leftarrow m = 2.78$
0	.0596	.0620
1	.1705	.1725
2	.2414	.2397
3	.2255	.2221
4	.1564	.1544
5	.0858	.0858
6	.0389	.0398
7	.0149	.0158
8	.0050	.0055
9	.0015	.0017
10	.0004	.0005
11	.0001	.0001

The approximation is quite good. ■

The Poisson frequency function can be used to approximate binomial probabilities for large n and small p. This suggests how Poisson distributions can arise in practice. Suppose that X is a random variable that equals the number of times some event occurs in a given interval of time. Heuristically, let us think of dividing the interval into a very large number of small subintervals of equal length, and let us assume that the subintervals are so small that the probability of more than one event in a subinterval is negligible relative to the probability of one event, which is itself very small. Let us also assume that the probability of an event is the same in each subinterval and that whether an event occurs in one subinterval is independent of what happens in the other subintervals. The random variable X is thus nearly a binomial random variable, with the subintervals consitituting the trials, and, from the limiting result above, X has nearly a Poisson distribution.

The preceding argument is not formal, of course, but merely suggestive. But, in fact, it can be made rigorous. The important assumptions underlying it are (1) what happens in one subinterval is independent of what happens in any other subinterval, (2) the probability of an event is the same in each subinterval, and (3) events do not happen simultaneously. The same kind of argument can be made if we are concerned with an area or a volume of space rather than with an interval on the real line.

The Poisson distribution is of fundamental theoretical and practical importance. It has been used in many areas, including the following:

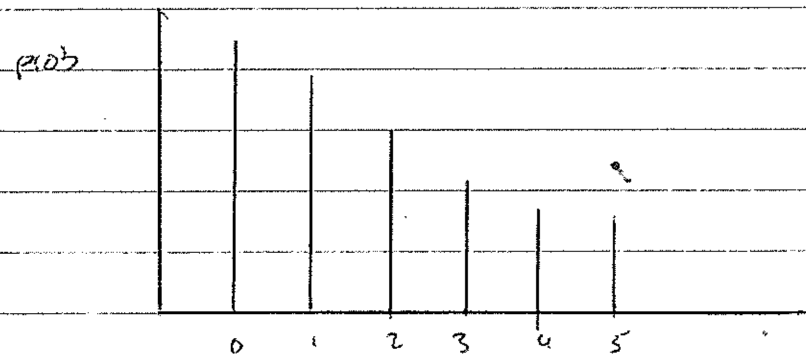
- The Poisson distribution has been used in the analysis of telephone systems. The number of calls coming into an exchange during a unit of time might be modeled as a Poisson variable if the exchange services a large number of customers who act more or less independently.
- One of the earliest uses of the Poisson distribution was to model the number of alpha particles emitted from a radioactive source during a given period of time.
- The Poisson distribution has been used as a model by insurance companies. For example. the number of freak acidents, such as falls in the shower. for a large

Poisson Probability Distribution

X = # occasions of a rare event

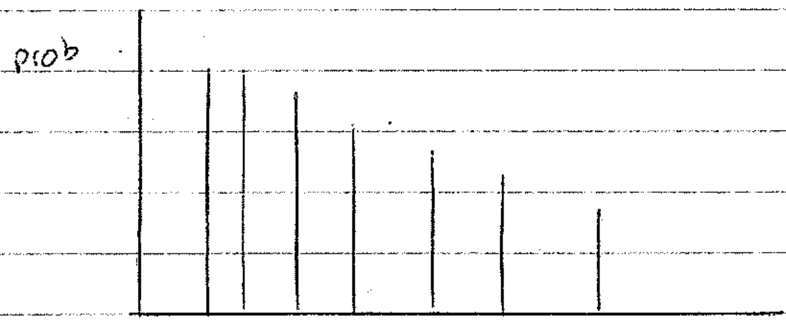
We'll see that $P(x=k) = \frac{m^k e^{-m}}{k!}$ $\left(\frac{\lambda^k e^{-\lambda}}{k}\right)$

(m is the parameter).

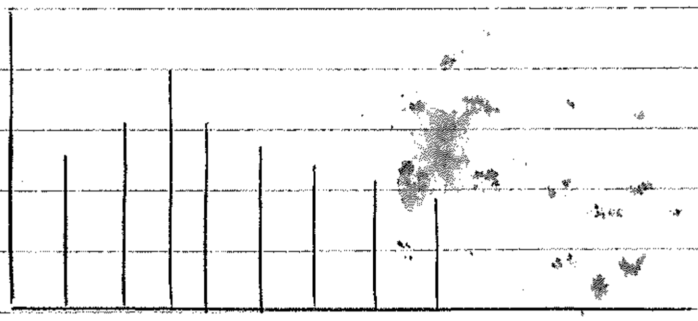


depends on m

$m < 1$



$m = 1$



$m > 1$

41

λ has an interpretation as Average # of occurances per unit time.

$\lambda t = m$

$$P(x=k) = \frac{m^k e^{-m}}{k!}$$

Example: Find $P(x \geq 5)$ with $m = 5$.

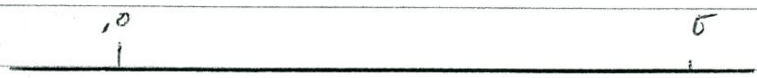
$= .559$

$.1088 = 1 - P(X<5) = 1 - \left(P[X=0] + P(X=1) + \dots P(X=4)\right)$

<u>Poission Distribution is an approximation of Binomial dist.</u>

In certain circumstances we can approximate the Binomial dist. using the poisson Distribution.

consider poission again.



Divide into n short time invervals each t/n.

Assume we have a Poisson Distribution $\frac{\lambda t}{n}$

For each short time interval, we can get 1 occurance will probability $\sim \lambda h \to \frac{\lambda t}{n}$

and prob of >1 occurance is negligable.

Thus we have n short intervals for each of which we get 1 occurance with prob $\sim \lambda\left(\frac{t}{n}\right)$

or 0 occurances with prob $\sim 1 - \lambda\left(\frac{t}{n}\right)$

Cut dey ✓ non-overlapping intervals the # of occurances are all indep.

Each short time interval corresponds to a bernoulli trial

X_t = # occurances in the entire interval $(0, t)$

We get $(X_t = k)$ when k of the n short intervals results in success (ie an occurance)

Thus:

$$P(X_t = k) = P[\, k \text{ successes over the } n \text{ short intervals}\,]$$

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

Now let n get larger $(\to \infty)$ and the approximate of the poisson Process (by this collection of n short intervals) improves:

We want to find

$$\lim_{n\to\infty} \binom{n}{k}\left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k}$$

$$\lim_{n\to\infty} \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} \frac{(\lambda t)^k}{n^k} \frac{\left(1 - \lambda t/n\right)^n}{\left(1 - \lambda t/n\right)^k}$$

switch

$$= \lim_{n\to\infty} \left[\frac{n(n-1)\ldots(n-k+1)}{\underbrace{(n)(n)\ldots(n)}_{k \text{ times}}}\right] \frac{(\lambda t)^k}{k!} \frac{\left(1 - \frac{\lambda t}{n}\right)^n}{\left(1 - \frac{\lambda t}{n}\right)^k}$$

$$\frac{(\lambda t)^k}{k!} \lim_{n\to\infty} \left[\quad\right] \lim_{n\to\infty} \frac{\left(1 - \frac{\lambda t}{n}\right)^n}{\left(1 - \frac{\lambda t}{n}\right)^k}.$$

$$\frac{(\lambda t)^k}{k!} \; [1] \; \frac{\lim_{n\to\infty} \left(1 - \frac{\lambda t}{n}\right)^n}{\lim_{n\to\infty} \left(1 - \frac{\lambda t}{n}\right)^k \to 1} \quad \text{---} \quad \text{note: } \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$$

$$= P(X_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

so:
$$\lim_{n\to\infty} \binom{n}{k} \left(1 - \frac{\lambda t}{n}\right)^{n-k} \left(\frac{\lambda t}{n}\right)^k = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad (*)$$

$$\frac{\lambda t}{n} = p \Rightarrow \lambda t = np.$$

Since λt is fixed then as $n \to \infty$ must have $p \to 0$ (in such a way that np remains fixed)

Thus $(*)$ equation means that

$$\lim_{\substack{n\to\infty \\ (np \text{ fixed})}} \binom{n}{k} p^k (1-p)^{n-k} = \frac{(np)^k e^{-np}}{k!}$$

Consequence of this:

When n is large

p is small

we can use a poisson Prob function to approximate Binomial Prob function

rule: $n \geq 100$ $\quad p \leq .01 \quad np \leq 20$

population of people in a given time period might be modeled as a Poisson distribution, since the accidents would presumably be rare and independent (provided there was only one person in the shower.)

- The Poisson distribution has been used by traffic engineers as a model for light traffic. The number of vehicles that pass a marker on a roadway during a unit of time can be counted. If traffic is light, the individual vehicles act independently of each other. In heavy traffic, however, one vehicle's movement may influence another's, so the approximation might not be good.

EXAMPLE B This amusing classical example is from von Bortkiewicz (1898). The number of fatalities that resulted from being kicked by a horse was recorded for 10 corps of Prussian cavalry over a period of 20 years, giving 200 corps-years worth of data. These data and the probabilities from a Poisson model with $\lambda = .61$ are displayed in the following table. The first column of the table gives the number of deaths per year, ranging from 0 to 4. The second column lists how many times that number of deaths was observed. Thus, for example, in 65 of the 200 corps-years, there was one death. In the third column of the table, the observed numbers are converted to relative frequencies by dividing them by 200. The fourth column of the table gives Poisson probabilities with the parameter $\lambda = .61$. In chapters 8 and 9 we discuss how to choose a parameter value to fit a theoretical probability model to observed frequencies and methods for testing goodness of fit. For now we will just remark that the value $\lambda = .61$ was chosen to match the average number of deaths per year.

hardly

Number of Deaths per Year	*Observed*	*Relative Frequency*	*Poisson Probability*
0	109	.545	.543
1	65	.325	.331
2	22	.110	.101
3	3	.015	.021
4	4	.005	.003

■

The Poisson distribution often arises from a model called a **Poisson process** for the distribution of random events in a set S, which is typically one-, two-, or three-dimensional, corresponding to time, a plane, or a volume of space. Basically, this model states that if $S_1, S_2, \ldots, S_n$ are disjoint subsets of S, then the numbers of events in these subsets, $N_1, N_2, \ldots, N_n$, are independent random variables that follow Poisson distributions with parameters $\lambda|S_1|, \lambda|S_2|, \ldots, \lambda|S_n|$, where $|S_i|$ denotes the measure of S_i (length, area, or volume, for example). The crucial assumptions here are that events in disjoint subsets are independent of each other and that the Poisson parameter for a subset is proportional to the subset's size. Later, we will see that this latter assumption implies that the average number of events in a subset is proportional to its size.

EXAMPLE C Suppose that an office receives telephone calls as a Poisson process with $\lambda = .5$ per min. The number of calls in a 5-min. interval follows a Poisson distribution with

(Cumulative) Distribution Function.

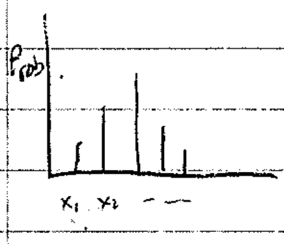
This is $P[X \leq x]$

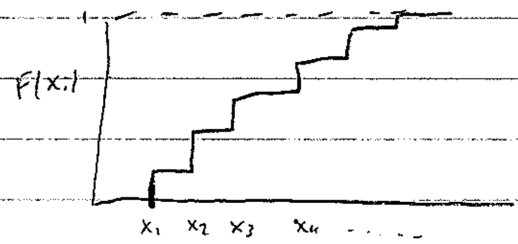
For x discrete $(\in \mathbb{Z}^+)$, C.D.F look

Prob function | Distribution function.

we specify $P(X = x_i)$ for each possible x_i

we specify $P(X \leq x_i)$ for each possible x_i "$F(x)$"





Note: Tables of binomial & poisson give $1 - F(x_{i-1})$ for each x_i

Continuous Random variable.

Here the values of X are in some continuous range

eg. 0 to 1

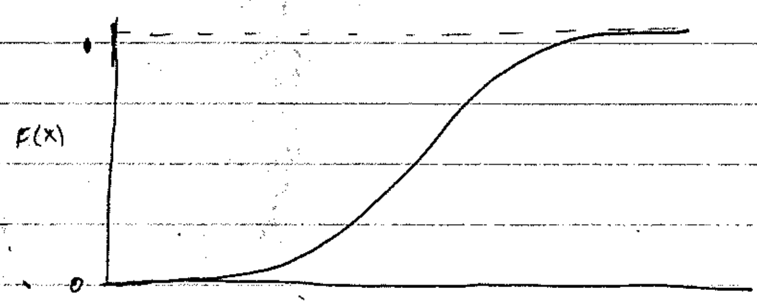
or $0 \rightarrow \infty$

or $-\infty \rightarrow +\infty$

Continuous → can take any value

Discrete → only values $\in \mathbb{N}$

Shape of the $F(x)$ is quite different



$F(x)$ is continuous.

Our interest is in D.F's (dist Functions) which are differentiable everywhere

Prob. Desity Functions

(except for a finite # of points in any finite interval)

Denote the derivative of $F(x)$ by $f(x)$

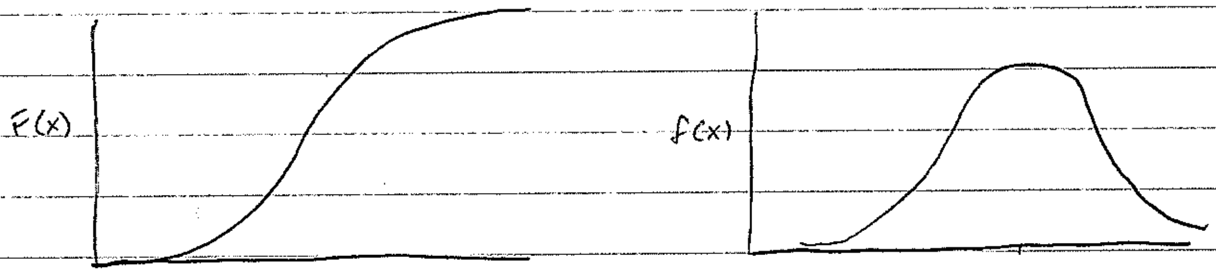
$f(x)$ is called

Probability Desity Function (pdf)

of random variable x or "distribution"

$$f(x) = F'(x)$$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

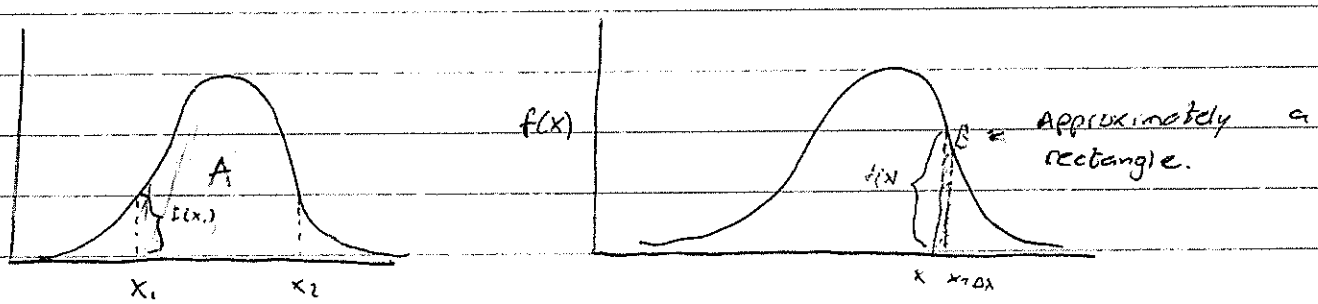


Why "density" in density function?

Consider $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$

$$= \int_{x_1}^{x_2} f(t)\,dt \quad - \underline{\text{Area } A}.$$

$\left.\begin{array}{l} x_1 \to x \\ x_2 \to x + \Delta x \end{array}\right\}$ $P(x < X \leq x + \Delta x) = \int_{x}^{x+\Delta x} f(t)\,dt \quad - \underline{\text{Area } B}$



$F(x_1)^*$ $\qquad$ $f(x)^*$

$$P[x < X \leq x + \Delta x] \simeq f(x)\,\Delta x$$

Approximation improves for smaller Δx

$\frac{dF(x)}{dx} = P.D.F$ X

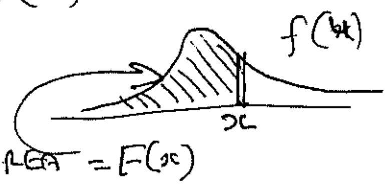
$\int_{-\infty}^{y} f(y)dy = P(Y \leq y)$

$= F(y)$

pdf $f(x)$

(cdf)(DF) $F(x) = \int_{-\infty}^{x} f(u)\,du$

$P(X \leq x)$.



519.2 Mull - Concepts of prob.

Welcome to business class

519.2 Mill - Problems in prob.

519 Sves - Problems in prob theory

519.2 Lars.

Thus $f(x)$ has the dimensions of prob per unit interval

Thus $f(x)$ is giving us the density of probability at the value x

From * $F(x_i)$ gives us the prob that X is $\leq x_i$

$f(x)$ is the density of the prob.

Properties of PDF.

PDF

i) $f(x) \geq 0$

ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

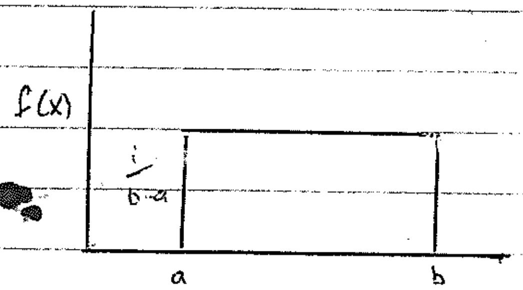
Prob (mass) Function.

i) $f(x_i) \geq 0$

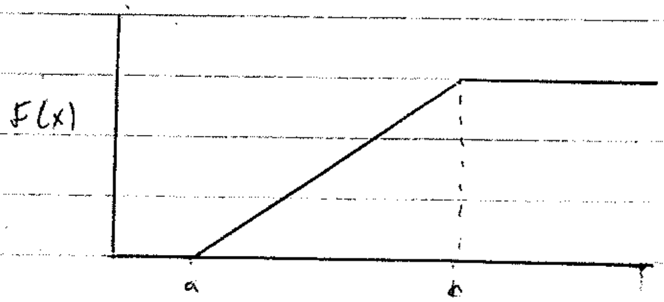
ii) $\sum_{x_i} f(x_i) = 1$

---

1) Uniform Random Variable



$X \sim R(a,b)$



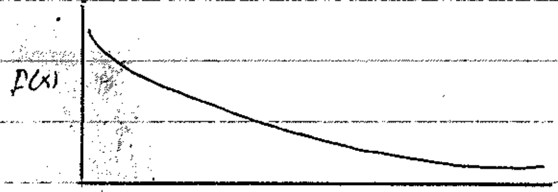
$f(x) = \dfrac{1}{b-a}$ for $a \leq x \leq b$

$= 0$ for everything else

2) (Negative) Exponential Random variable

$f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$

$= 0$ for everything else



$X \sim NE(\lambda)$

Easy to see: $\int_0^\infty f(x)dx = 1$

Parameter is λ $(\lambda > 0)$

– Connection with Poisson

let L = time up to first occurance. (for poisson process)

Events $(L > t)$ and $(X_t = 0)$ are equivilant events;

X_t = # of occurances so if $L > t$ no of occurances inside $t = 0$

we found that $P(X_t = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$ *

Thus $P(L > t) = P(X_t = 0)$

$= e^{-\lambda t}$ (from *)

$P(L \leq t) = 1 - e^{-\lambda t}$

$F(t)$ = D. F for L

Cut dey r non-overlapping intervals the # of occurances are all indep.

Each short time interval corresponds to a bernoulli trial

X_t = # occurances in the entire interval $(0,t)$

We get $(X_t = k)$ when k of the n short intervals result in success (ie an occurance)

Thus:

$$P(X_t = k) = P[\,k \text{ successes over the } n \text{ short intervals}\,]$$

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

Now let n get larger $(\to \infty)$ and the approximate of the poisson Process (by this collection of n short intervals) improves:

We want to find

$$\lim_{n\to\infty} \binom{n}{k}\left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k}$$

$$\lim_{n\to\infty} \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} \frac{(\lambda t)^k}{n^k} \frac{(1-\lambda t/n)^n}{(1-\lambda t/n)^k}$$

Switch

$$= \lim_{n\to\infty}\left[\frac{n(n-1)\ldots(n-k+1)}{\underbrace{(n)(n)\ldots(n)}_{k \text{ times}}}\right] \frac{(\lambda t)^k}{k!} \frac{\left(1-\frac{\lambda t}{n}\right)^n}{\left(1-\frac{\lambda t}{n}\right)^k}$$

$$\frac{(\lambda t)^k}{k!} \lim_{n\to\infty}\Big[\quad\Big] \lim_{n\to\infty} \frac{\left(1-\frac{\lambda t}{n}\right)^n}{\left(1-\frac{\lambda t}{n}\right)^k}$$

$$\frac{(\lambda t)^k}{k!} \; [1] \; \frac{\lim_{n\to\infty} \left(1 - \frac{\lambda t}{n}\right)^n}{\lim_{n\to\infty} \left(1 - \frac{\lambda t}{n}\right)^k \to 1} \; \cdots$$

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

$$= P(X_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

so: $$\lim_{n\to\infty} \binom{n}{k} \left(1 - \frac{\lambda t}{n}\right)^{n-k} \left(\frac{\lambda t}{n}\right)^k = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad (*)$$

$$\frac{\lambda t}{n} = p \implies \lambda t = np.$$

Since λt is fixed then as $n \to \infty$ must have $p \to 0$ (in such a way that np remains fixed)

Thus $(*)$ equation means that

$$\lim_{\substack{n\to\infty \\ (np \text{ fixed})}} \binom{n}{k} p^k (1-p)^{n-k} = \frac{(np)^k e^{-np}}{k!}$$

Consequence of this:

When n is large

p is small

we can use a poisson Prob function to approximate Binomial Prob function

rule: $n \geq 100$ $\quad p \leq .01 \quad np \leq 20$

population of people in a given time period might be modeled as a Poisson distribution, since the accidents would presumably be rare and independent (provided there was only one person in the shower.)

- The Poisson distribution has been used by traffic engineers as a model for light traffic. The number of vehicles that pass a marker on a roadway during a unit of time can be counted. If traffic is light, the individual vehicles act independently of each other. In heavy traffic, however, one vehicle's movement may influence another's, so the approximation might not be good.

EXAMPLE B This amusing classical example is from von Bortkiewicz (1898). The number of fatalities that resulted from being kicked by a horse was recorded for 10 corps of Prussian cavalry over a period of 20 years, giving 200 corps-years worth of data. These data and the probabilities from a Poisson model with $\lambda = .61$ are displayed in the following table. The first column of the table gives the number of deaths per year, ranging from 0 to 4. The second column lists how many times that number of deaths was observed. Thus, for example, in 65 of the 200 corps-years, there was one death. In the third column of the table, the observed numbers are converted to relative frequencies by dividing them by 200. The fourth column of the table gives Poisson probabilities with the parameter $\lambda = .61$. In chapters 8 and 9 we discuss how to choose a parameter value to fit a theoretical probability model to observed frequencies and methods for testing goodness of fit. For now we will just remark that the value $\lambda = .61$ was chosen to match the average number of deaths per year.

hardly

Number of Deaths per Year	*Observed*	*Relative Frequency*	*Poisson Probability*
0	109	.545	.543
1	65	.325	.331
2	22	.110	.101
3	3	.015	.021
4	4	.005	.003

■

The Poisson distribution often arises from a model called a **Poisson process** for the distribution of random events in a set S, which is typically one-, two-, or three-dimensional, corresponding to time, a plane, or a volume of space. Basically, this model states that if $S_1, S_2, \ldots, S_n$ are disjoint subsets of S, then the numbers of events in these subsets, $N_1, N_2, \ldots, N_n$, are independent random variables that follow Poisson distributions with parameters $\lambda|S_1|, \lambda|S_2|, \ldots, \lambda|S_n|$, where $|S_i|$ denotes the measure of S_i (length, area, or volume, for example). The crucial assumptions here are that events in disjoint subsets are independent of each other and that the Poisson parameter for a subset is proportional to the subset's size. Later, we will see that this latter assumption implies that the average number of events in a subset is proportional to its size.

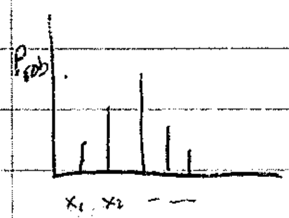
EXAMPLE C Suppose that an office receives telephone calls as a Poisson process with $\lambda = .5$ per min. The number of calls in a 5-min. interval follows a Poisson distribution with

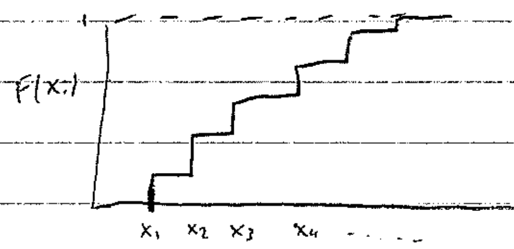
(Cumulative) Distribution Function

This is $P[X \leq x]$

For x discrete ($\in \mathbb{Z}^+$), CDF look

Prob function	Distribution function.
we specify $P(X = x_i)$ for each possible x_i	we specify $P(X \leq x_i)$ for each possible x_i $F(x)$





Note: Tables of binomial & poisson give $1 - F(x_{i-1})$ for each x_i

Continuous Random variable.

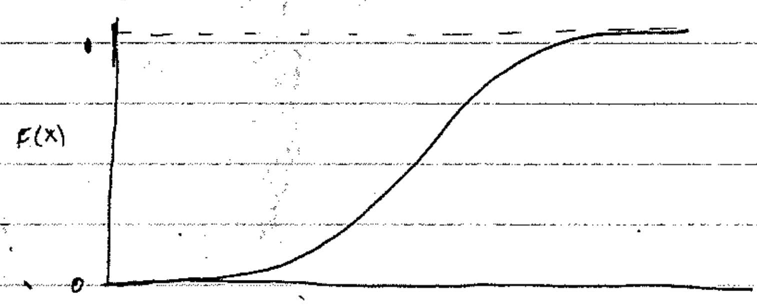
Here the values of X are in some continuous range

eg. 0 to 1

or $0 \to \infty$ Continuous $\to$ can take any value

or $-\infty \to +\infty$ Discrete $\to$ only values $\in \mathbb{N}$

Shape of the $F(x)$ is quite different



$F(x)$ is continuous.

Our interest is in D.F's (dist Functions) which are differentiable everywhere

Prob. Desity Functions

(except for a finite # of points in any finite interval)

Denote the derivative of $F(x)$ by $f(x)$

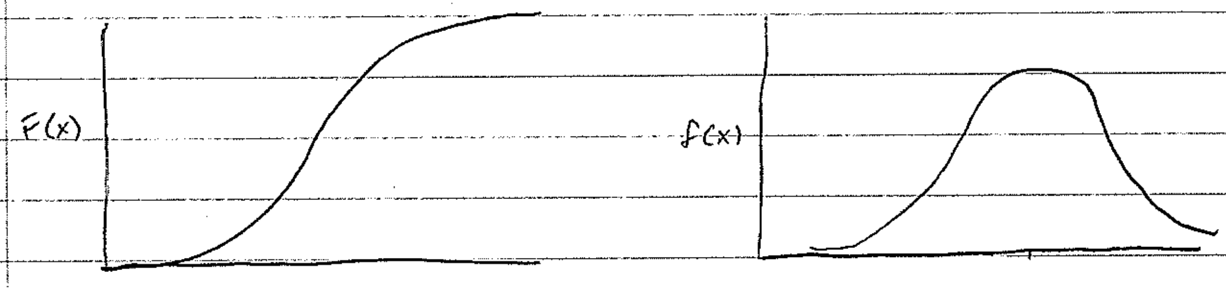
$f(x)$ is called

Probability Desity Function (pdf)

of random variable x or "distribution"

$$f(x) = F'(x)$$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

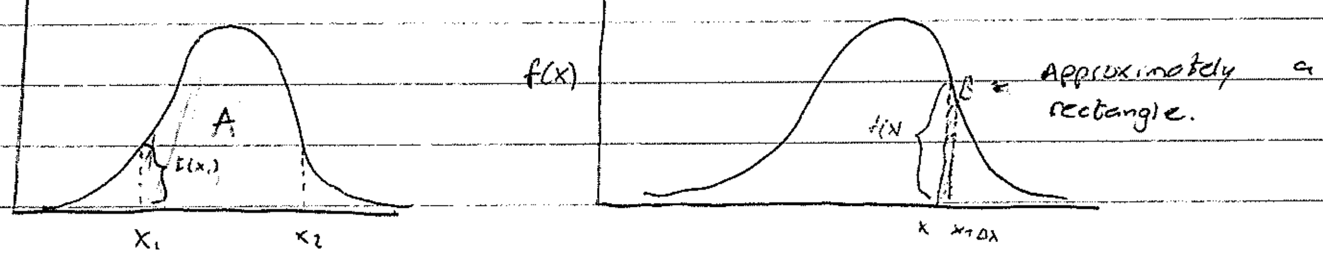


Why "density" in density function?

Consider $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$

$$= \int_{x_1}^{x_2} f(t)\,dt \quad - \ \underline{\text{Area } A}.$$

$\left.\begin{array}{l} x_1 \to x \\ x_2 \to x + \Delta x \end{array}\right\}$ $P(x < X \leq x + \Delta x) = \int_{x}^{x+\Delta x} f(t)\,dt$ - $\underline{\text{Area } B}$



$F(x_1)^*$ $f(x)^*$

$$P[x < X \leq x + \Delta x] \simeq f(x)\,\Delta x$$

Approximation improves for smaller Δx

45

$\frac{d F(x)}{dx} = \text{P.D.F}$ ~~X~~

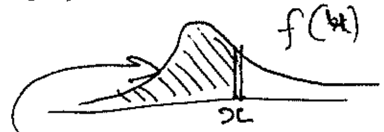
$\int_{-\infty}^{y} f(y)\,dy = P(Y \le y)$

$= F(Y)$

pdf $f(x)$

(cdf)(DF) $F(x) = \int_{-\infty}^{x} f(u)\,du$

$P(X \le x)$



AREA $= F(x)$

519.2 Mull - Concepts of prob.

Welcome to business class

519.2 Mill - Problems in prob.

519 Sves - Problems in prob theory

519.2 Lars.

Deloitte.
www.deloitte.com/ie

$I_{[Y \le 4]}$

$P[Z \le z \mid Y \le 4]$

$\min(Y, 4)$

$P[Z \le z]$

$= P[\min(Y,4) \le z] \quad z \ge 4$

$= 1$

$=$

Thus $f(x)$ has the dimensions of prob per unit interval

Thus $f(x)$ is giving us the density of probability at the value x

From * $F(x_i)$ gives us the prob that X is $\le x_i$

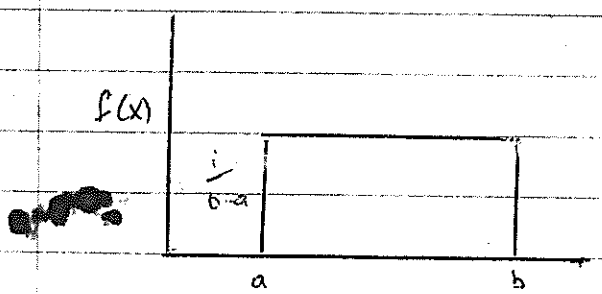
$f(x)$ is the density of the prob.

Properties of PDF

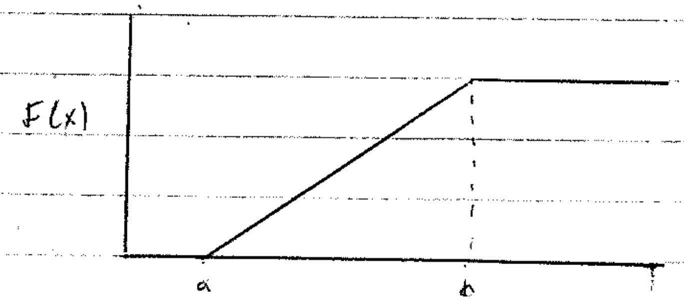
PDF	Prob (mass) Function.
i) $f(x) \geq 0$	i) $f(x_i) \geq 0$
ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$	ii) $\sum_{x_i} f(x_i) = 1$

---

1) Uniform Random Variable



$X \sim R(a,b)$

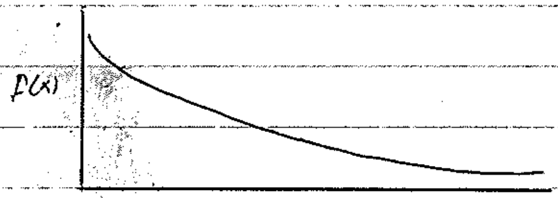


$$f(x) = \frac{1}{b-a} \quad \text{for } a \le x \le b$$
$$= 0 \quad \text{for everything else}$$

21) (Negative) Exponential Random variable

$f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$

$= 0$ for everything else



$X \sim NE(\lambda)$

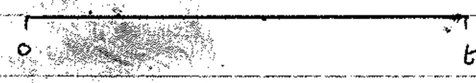
Easy to see: $\int_0^\infty f(x)dx = 1$

Parameter is λ $(\lambda > 0)$

– Connection with Poisson

let L = time up to first occurance. (for poisson process)

Events $(L > t)$ and $(X_t = 0)$ are equivilant events



X_t = # of occurances so if $L > t$ no of occurances inside $t = 0$

we found that $P(X_t = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$ *

Thus $P(L > t) = P(X_t = 0)$

$= e^{-\lambda t}$ (from *)

$P(L \leq t) = 1 - e^{-\lambda t}$

$F(t)$ = D. F for L

$\alpha > 1$

$$\int \overset{u}{t^{\alpha-1}} \overset{dv}{e^{-t}}\, dt = \left[-t^{\alpha-1} e^{-t}\right]_0^{\infty} + (\alpha-1)\int_0^{\infty} t^{\alpha-2} e^{-t}\, dt$$

$\cancel{(\alpha-1)t^{\alpha-2}}$ $\alpha > 1$

$$\Gamma(\alpha) = (\alpha-1)\,\Gamma(\alpha-1)$$

Differentiate $F(t)$ to get PDF for L

$\Rightarrow f(t) = \lambda e^{-\lambda t}$

(we recognize as an Negative Exp pdf)

☆ This pdf has the unusual property referred to as "<u>No memory</u>" (Think of coin (Heads / tails)) if throw is heads, $P(\text{tails})$ still $= \frac{1}{2}$

$$P\left(L > t+s \mid L > s\right) = P\left[L > t\right]$$

ie (magnitude of S has no effects on the prob)

$$\frac{P\left[(L > t+s) \cap (L > s)\right]}{P(L > s)}$$

$$= \frac{P\left(L > t+s\right]}{P(L>s)}$$

$$= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} \rightarrow P(L > t)$$

---

3) $X \sim \text{Gamma}(\alpha, \beta)$

$$f(x) = \frac{1}{\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} \frac{1}{\beta}$$

Shape: if $\alpha = 1 \rightarrow Ne(1/\beta)$ $\quad f(x) = \frac{1}{\beta} e^{-x/\beta}$

what about $\Gamma(\alpha) \overset{?}{=} (\alpha-1)!$

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$$

TO SHOW $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

$$\Gamma\left(\tfrac{1}{2}\right) = \int_0^\infty x^{-\frac{1}{2}} e^{-x}\, dx$$

$$= 2\int_0^\infty e^{-u^2}\, du \qquad u^2 = x, \quad 2u\,du = dx$$

$$\Gamma^2\left(\tfrac{1}{2}\right) = 4\left(\int_0^\infty e^{-u^2} du\right)\left(\int_0^\infty e^{-v^2} dv\right)$$

$$= 4\int_0^\infty \int_0^\infty e^{-(u^2+v^2)}\, du\, dv \qquad u = r\cos\theta, \quad v = r\sin\theta$$

$$= 4\int_0^\infty \int_0^{\frac{\pi}{2}} e^{-r^2}\, r\, dr\, d\theta \qquad u^2+v^2 = r^2, \quad du\,dv \longrightarrow r\,dr\,d\theta$$

$$= 4\left(\int_0^\infty e^{-r^2} r\, dr\right)\left(\int_0^{\frac{\pi}{2}} d\theta\right)$$

$$= 4\left(\tfrac{1}{2}\right)\left(\tfrac{\pi}{2}\right)$$

$$\Gamma^2\left(\tfrac{1}{2}\right) = \pi$$

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}$$

NOTICE $\Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right) = \frac{3}{4}\sqrt{\pi}$

DISTRIB.

Figure 2-11 shows several gamma densities. Gamma densities provide a fairly flexible class for modeling nonnegative random variables.

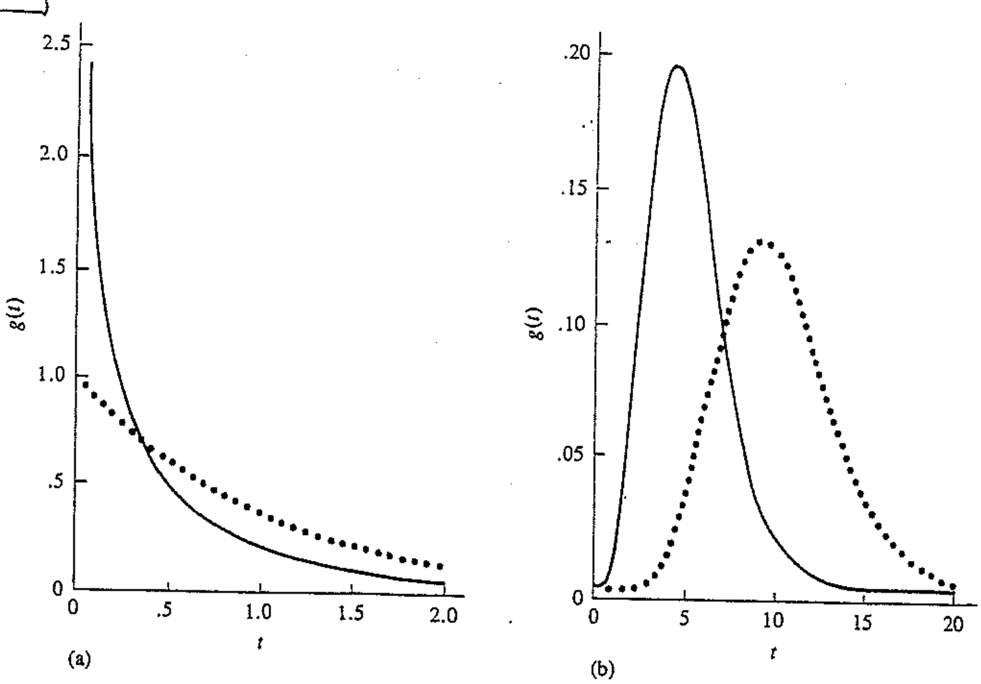


Figure 2-11. Gamma densities, (a) $\alpha = .5$ (solid) and $\alpha = 1$ (dotted) and (b) $\alpha = 5$ (solid) and $\alpha = 10$ (dotted); $\lambda = 1$ in all cases.

EXAMPLE A. The patterns of occurrence of earthquakes in terms of time, space, and magnitude are very erratic, but attempts are sometimes made to construct probabilistic models for these events. The models may be used in a purely descriptive manner or, more ambitiously, for purposes of predicting future occurrences and consequent damage.

Figure 2-12 shows the fit of a gamma density and an exponential density to the observed times separating a sequence of small earthquakes (Udias and Rice, 1975). The gamma density clearly gives a better fit ($\alpha = .509$ and $\lambda = .00115$). Note that an exponential model for interoccurrence times would be memoryless; that is, knowing that an earthquake had not occurred in the last t time units would tell us nothing about the probability of occurrence during the next s time units. The gamma model does not have this property. In fact, although we will not show this, the gamma model with these parameter values has the character that there is a large likelihood that the next earthquake will immediately follow any given one and this likelihood decreases monotonically with time. □

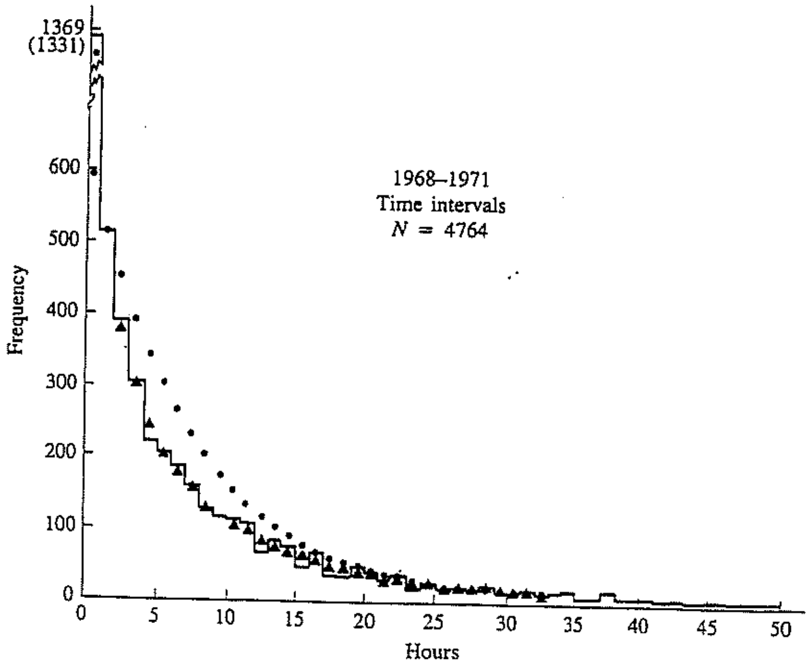


Figure 2-12. Fit of gamma distribution (triangles) and of exponential distribution (circles) to times between microearthquakes.

We can show $\Gamma(\alpha) = (\alpha-1)\,\Gamma(\alpha-1)$

case 1) If $\alpha = n$ $\quad \Gamma(\alpha) = (n-1)!$

case 2) $\alpha = \frac{n}{2}$ $\quad \Gamma\left(\frac{n}{2}\right) = \left(\frac{n}{2}-1\right)\Gamma\left(\frac{n}{2}-1\right)$

$$= \left(\frac{n}{2}-1\right)\left(\frac{n}{2}-2\right) \cdots \Gamma\left(\frac{1}{2}\right)$$

Can show $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

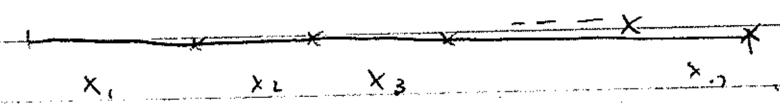
Connection with Poission Process (Gamma dist)

Although the gamma dist has many uses that are unconnected with the Poisson Process But there is an incresting connection with Poisson.

* see H 10

Note: L_n = (time to 1st occ) + (time from 1st to 2nd) + ... + (time from (n-1)th to nth)

$$L_n = X_1 + X_2 + \ldots + X_n$$



Each x_i is neg exp (λ).

For $\alpha = n$ (integer) then

L_n is $= \sum_{i=1}^{n} x_i$

when x_i = Time up to next occurance (for a poisson process)

15 dist funcs.

L_n is sum of negative exponential random variables and has a gamma Distribution

gamma is prob density function.

4) Normal Distribution

P.D.F

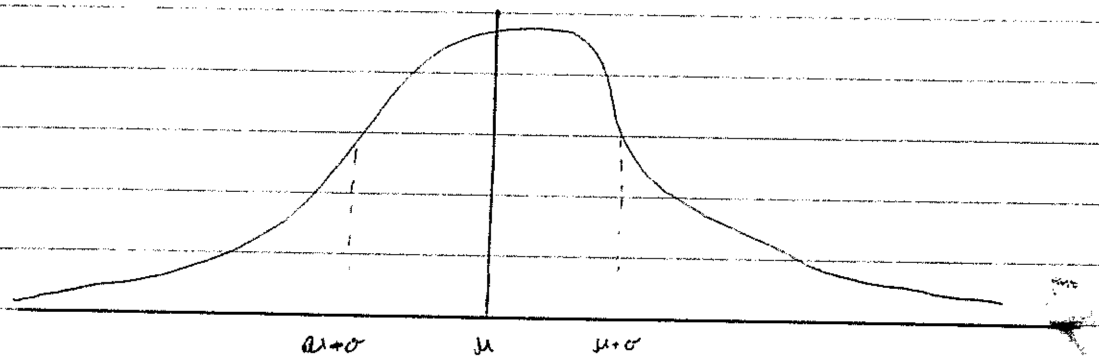
$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

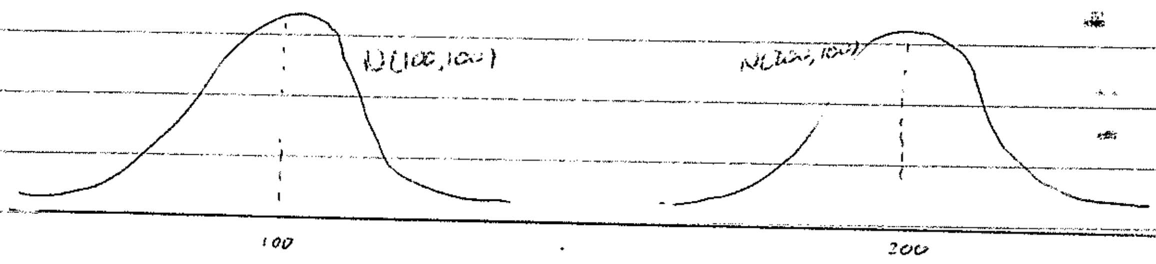
. μ = mean

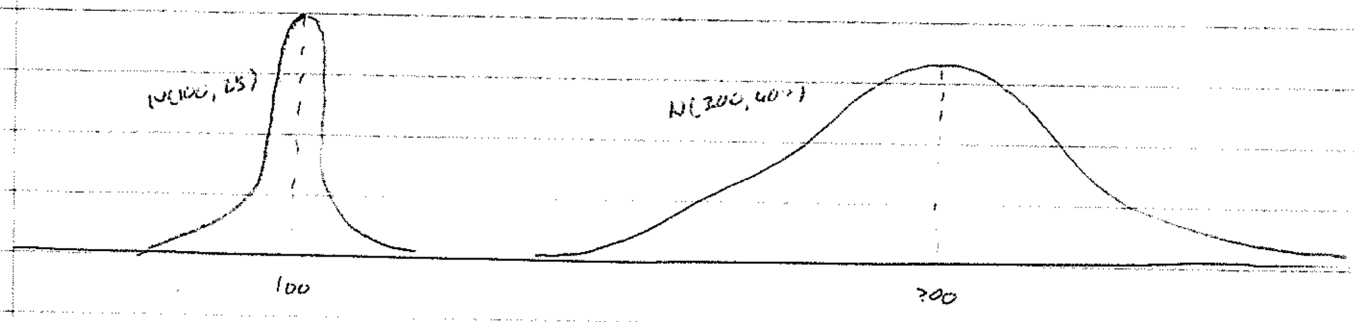
σ = Standard diviation

Shape : depends on μ and σ

$N(\mu, \sigma^2)$







Why important?

1) In nature (and man made phenomena) the normal curve can closely describe the shapes of many frequency distributions.

(eg distribution of heights (of plants, people etc)

- distribution of IQ measurements.

2) The distribution of averages ($\bar{x}$) follows a Normal Curve (Central limit theorem)

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$\sigma > 0$

$\mu \in \mathbb{R}$

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

let $z = \dfrac{x-\mu}{\sigma}$ $\qquad dz = \dfrac{1}{\sigma} dx$

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

This is normal P.D.F $\mu = 0$ $\sigma = 1$ known as the standard normal P.D.F

let $t = \dfrac{z^2}{2}$ $\qquad dt = z\,dz \Rightarrow dz = \dfrac{1}{\sqrt{2}} t^{-\frac{1}{2}} dt$

$$I = 2\int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad \text{- Symmetry.}$$

$$I = 2 \int_0^\infty \frac{1}{\sqrt{2}\sqrt{\pi}} e^{-t} \cdot \frac{1}{\sqrt{2}} t^{-\frac{1}{2}} \, dt$$

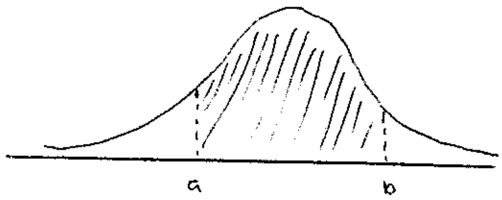
$$= \frac{1}{\sqrt{\pi}} \underbrace{\int_0^\infty t^{-\frac{1}{2}} e^{-t} \, dt}_{\Gamma(\frac{1}{2})}$$

$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \, dt$

$$I = 1/\sqrt{\pi} \cdot \sqrt{\pi} = 1$$

$X = N(\mu, \sigma^2)$

$P[a < X \leq b]$



$= P(X \leq b) - P(X \leq a)$

$= F(b) - F(a)$

See sheet from 1st year.

$$P(X > x) = P\left(\frac{X - \mu}{\sigma} > \frac{x - \mu}{\sigma}\right)$$

$\uparrow$ $N(\mu, \sigma^2)$ $\qquad$ $\uparrow$ $N(0, 1)$

Example: $X \sim N(100, 100)$

Find $P(X > 120)$.

$$= P\left(\frac{X - 100}{10} > \frac{120 - 100}{10}\right).$$

$P(Z > 2) = .02275.$

Beta Distribution:

Random variable (continuous) on $(0, 1)$

The Prob Density Function (PDF):

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad \text{for} \quad 0 < x < 1$$

where: $B(\alpha, \beta)$ is such that $\int_0^1 f(x)\,dx = 1$ and can be shown that $B(\alpha, \beta) = \dfrac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

Note: α, β are the parameters of the Beta distribution. They control the shape of the P.D.F. see H10 on shape.

Weibull Distribution:

The CDF (cumulative Dist Function)

$$F(x) = 1 - e^{-(x/\theta)^\beta}$$

Differentiate w.r.t x, we'll get P.D.F

$$f(x) = F'(x) = \beta\left(\frac{x}{\theta}\right)^{\beta-1} \frac{1}{\theta} e^{-(x/\theta)^\beta}$$

$$= \beta x^{\beta-1} \cdot \frac{1}{\theta^\beta} e^{-(x/\theta)^\beta}$$

2 parameters β, θ: β = shape parameter

θ = scale parameter

SHAPES OF THE BETA DENSITY ST2054

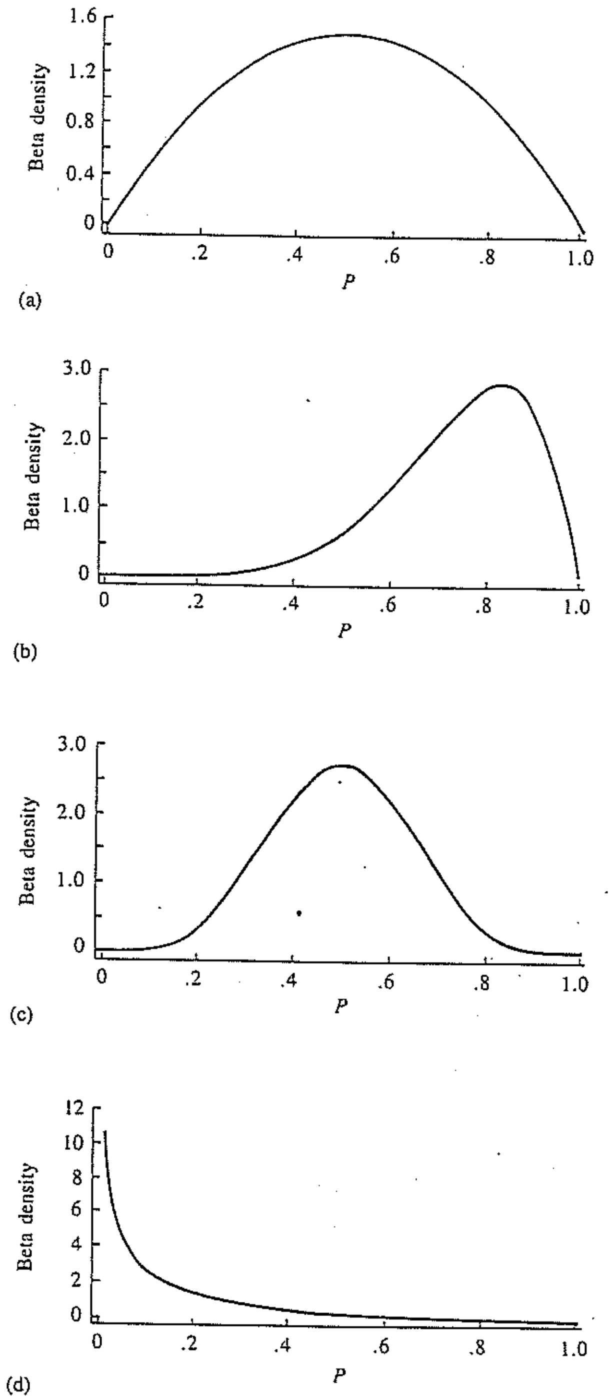
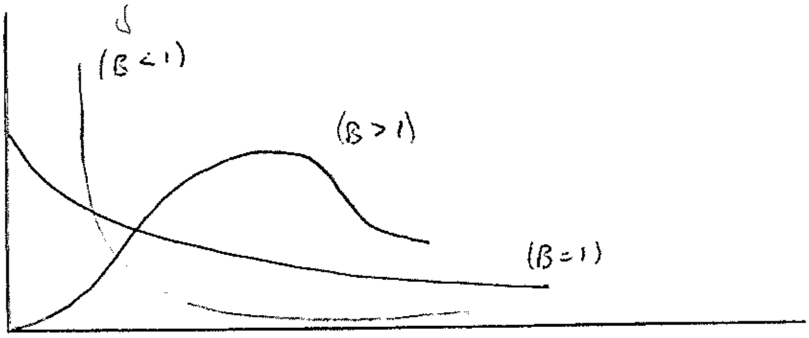


Figure 15-6. Beta density functions for various values of a and b: (a) $a = 2, b = 2$; (b) $a = 6, b = 2$; (c) $a = .5, b = 4$; and (d) $a = 6, b = 6$.

Shapes :



Log - Normal Distribution

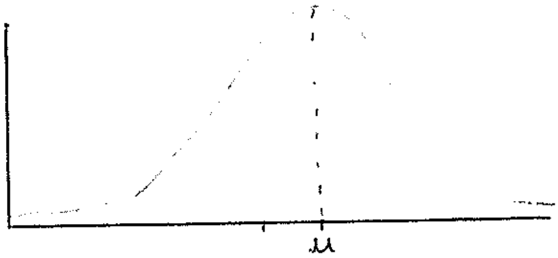
Suppose Z is $N(\mu, \sigma^2)$

$$P[Y \le y] = P[\log Y \le \log y] = \Phi\left(\frac{\log y - \mu}{\sigma}\right).$$

Then $Y = e^Z$ will have a log - normal distribution.

Range of values for Y is $0 \to \infty$

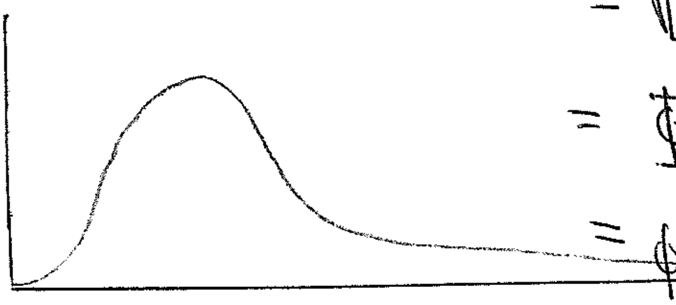
Z is $N(\mu, \sigma^2)$



$Y = \log N(\mu, \sigma^2)$

$f(y)$

used for claim size distribution.



$Y \sim LN(-2, 4)$

$$P[Y \le .32]$$
$$= P[\log Y \le \log(.32)]$$
$$= \Phi\left[\frac{\log(.32)+2}{2}\right]$$
$$= \Phi\left[\frac{-1.14+2}{2}\right]$$
$$= \Phi(.43) = .667$$

Location parameters & Scale Parameters

Consider a random variable with Dist Funct $F(x)$

if we can write $F(x)$ as $F(x - \gamma)$

eg : For $X \sim N(\mu, \sigma^2)$

μ = location Parameter

$$F(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$f(x) = h(x - \gamma)$ $\quad$ $h(x)$ indep of γ

Note: An equivalent condition is that the P.D.F can be written as $f(x - \gamma)$ ... γ = location parameter

Scale Parameter:

If the P.D.F can be written in form

$$f(x) = \frac{1}{\theta} g\left(\frac{x}{\theta}\right) \qquad \theta = \text{scale parameter}$$

Examples: Normal $\text{PDF} = \dfrac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$\Rightarrow \sigma$ = scale Parameter.

Gamma $\text{PDF} = \dfrac{1}{\Gamma(\gamma)} \left(\dfrac{x}{\beta}\right)^{\gamma - 1} e^{-(x/\beta)} \dfrac{1}{\beta}$

$\Rightarrow \beta$ = scale parameter.

Weibull Dist Funct $= F(x) = 1 - e^{-(x/\theta)^\beta}$

PDF $\Rightarrow f(x) = \beta \left(x/\theta\right)^{\beta-1} \, 1/\theta \; e^{-(x/\theta)^\beta}$

$\Rightarrow \theta$ = scale parameter.

Whatever is "over" = scale Parameter.

EXPECTED VALUES for Random Variables.

Vague def: Exp. value is a long term average value for the random variable (if one cud observe it many times)

Discrete Case First: x_i has Prob $f(x_i)$

$$\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \\ f(x_1) & f(x_2) & & f(x_n) \end{array}$$

Example: Game of chance: charge €1 for entry

Die is thrown → Prizes won

G = Net Gain from 1 play.

G	-1	0	4
Prob	4/6	1/6	1/6

Play over & over again. Ask what is long term outlook.

Average values for G is ~0

The reason for this is cuz Exp. value of G = 0

$$E(G) = (4/6)(-1) + 0(1/6) + 4(1/6) = 0$$

Definition of expected value for Random Variables.

X is discrete with prob function $f(x_i)$ then Exp. value of X is given by

$$E(x) = \sum_{\text{all } x_i} x_i \, p(x_i)$$

Provided the sum converges absolutely

ie $\sum_{\text{all } x_i} |x_i| \, |p(x_i)| < \infty$

If we don't have absolute convergence then we say $E(x)$ does not exist

$$F_x(x_i) = P[X \le x_i]$$

Lets also define $E[g(X)]$ where $g(X)$ is a function of random var X

Definition

$$E[g(X)] = \sum_{\text{all } x_i} g(x_i)\, P[X = x_i] \quad (*)$$

exists ~~prid~~ provided $\sum_{\text{all } x_i} |g(x_i)|\, P[X = x_i] < \infty$

If X cont. $E[g(x)] = \int g(x) f(x)dx$ prov. $\int |g| f(x) dx < \infty$

$E[a g_1(x) + b g_2(x)] = a E[g_1(x)] + b E[g_2(x)]$

Note: lets denote $g(x)$ by Y. Y is also a random variable.

- so we don't need a definition for $E(Y)$

what we could do is to get the Prob Distribution for Y and then apply the definition of $E(Y)$ with that prob dist.

However: its often quite difficult to find the Prob Distribution for $Y = g(x)$

- and because of this its easier to get $E[g(x)]$ using $(*)$

ii) later we look at methods for finding the prob dist. of Y (a function of X)

Important Example of $g(x) = [X - E(x)]^2$

$\mu = E(x)$

we will need to find $E[(X-\mu)^2]$

variance of a random variable

Denote $E[(x-\mu)^2] = V(x) = \sigma^2$

Related quantity is Standard Div of X

$SD(X) = \sqrt{V(x)} = \sigma$

— Example: Find The variance of $X \sim b(n,p)$

Soln.: In evaluating $V(x)$ its useful to know

$V(x) = E[(x-\mu)^2] = E(x^2) - [E(x)]^2$

Proof: $V(x) = E[(x-\mu)^2] = E[x^2 - 2x\mu + \mu^2]$

$= \sum (x_i^2 - 2x_i\mu + \mu^2)\, P(X=x_i)$

$= \sum_{\text{all } x_i} x_i^2\, P(X=x_i) - 2\mu \sum x_i\, P(X=x_i) + \mu^2 \sum P(X=x_i)$

$= E(x^2) - 2\mu\, E(x) + \mu^2 (1)$

$= E(x^2) - 2\mu \cdot \mu + \mu^2$

$= E(x^2) - \mu^2$

$v(x) = E(x^2) - [E(x)]^2$

Lets find $E(x)$ for $x \sim b(n,p)$

$E(x) = np$

$$E(x) = \sum_{k=0}^{n} k\, P(x=k)$$

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

First term = 0

$$= \sum_{\underline{k=1}}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \qquad \Big| \; Q = 1-p$$

$$= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k Q^{n-k} \quad \text{.. by cancellation.}$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)![(n-1)-(k-1)]!} p^{k-1} Q^{(n-1)-(k-1)}$$

$$\Big| \; k-1 = j$$

$$= np \sum_{j=0}^{n-1} \frac{(n-1)!}{j![(n-1)-j]!} p^j Q^{(n-1)-j}$$

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

$$= np\,[p+Q]^{n-1} \quad \text{.. but } p+Q = 1$$

$$\Rightarrow \underline{\underline{np. = E(x)}}$$

NOTE! Expected values for discrete var

$= \sum(x_i).\, P(x_i)$ .. always.

cont: 1 var $E(g(x)) = \int g(x)\, f(x)\,dx$.

2 var: $E(g(x,y)) = \iint g(x,y)\,[\text{particular pdf}]\,dy\,dx$

eg $E(xy) = \iint xy\, f(x,y)\,dy\,dx$.

lets get $V(x)$ = $X \sim b(n,p)$.

$E(x) = np$.

$$E(x^2) \Big| = E[x(x-1)] = \sum_{k=0}^{n} k(k-1)\,P(x=k)$$

$$= \sum_{k=2}^{n} k(k-1)\left\{\frac{n!}{k!(n-k)!}\right.\,p^k(1-p)^{n-k}$$

Just manipulation

$$= n(n-1)\sum_{k=2}^{n}\frac{(n-2)!}{(k-2)!\,[(n-2)-(k-2)]!}\,p^2 . p^{k-2}(1-p)^{(n-2)-(k-2)}$$

$J = k-2$.

$$n(n-1)p^2\sum_{j=0}^{n-2}\binom{n-2}{j}p^j(1-p)^{n-2-j}$$

$$= n(n-1)p^2\,[p+1-p]^{n-2}$$

$$E[x(x-1)] = n(n-1)p^2 \;*$$

$$E[x^2-x] = Ex^2 - Ex$$

$$\therefore\; E(x^2) = E[(x(x-1))]^* + Ex$$

Thus $E(x^2) = n(n-1)p^{2*} + Ex$

$$= n(n-1)p^2 + np.$$

Then:

$$V(x) = E(x^2) - [E(x)]^2$$

$$= n(n-1)p^2 + np - (E(x))^2$$

$= n(n-1)p^2 + np - (np)^2$

$= -np^2 + np$

$= V(x) = np(1-p)$

$E(x) = np.$

Learn these.

---

like $E(x) = x.\,p(x=x_i)$

Example: $X \sim$ Poisson (m) : $\dfrac{m^k e^{-m}}{k!}$

$E(x) = \sum_{\text{all } k} k\, P(x=k)$

$$E(x) = \sum_{k=0}^{\infty} k \frac{m^k e^{-m}}{k!}$$

$$= \sum_{k=1}^{\infty} \frac{m^k e^{-m}}{(k-1)!}$$

$$= m \sum_{k=1}^{\infty} \frac{m^{k-1} e^{-m}}{(k-1)!} \qquad k-1 = j$$

$$= m \sum_{j=0}^{\infty} \underbrace{\frac{m^j e^{-m}}{j!}}_{\text{Prob}(x=j)} = m(1)$$

$$= m.$$

$E(x) = m.$ for poisson.

Thus the parameter (m) of the poisson Dist is the mean value.

Note: Remember the poisson process.

we found $p(X_t = k) = \dfrac{(\lambda t)^k e^{-\lambda t}}{k!}$

Thus $\underline{E(X_t) = \lambda t}$

(λt) is the average # occurances in time t
and λ must b the average # occurances per unit time.

λ often called Rate of poisson process.

Next: $V(X)$ for $X \sim \text{Poisson}(m)$

usefu to consider $E(x(x-1))$

$$\sum_{k=0}^{\infty} k(k-1) \frac{m^k e^{-m}}{k!}$$

First 2 terms = 0

$$= \sum_{k=2}^{\infty} \frac{m^k e^{-m}}{(k-2)!} \qquad j = k-2$$

$$m^2 \sum_{j=0}^{\infty} \frac{m^j e^{-m}}{j} \quad . \text{ think abt it } \sum_{j=0}^{\infty} P(x=j) = 1$$

$$= m^2 \, (1)$$

$$E[x(x-1)] = m^2 =$$

$$E(x^2) = E(x^2 - x) + E(x)$$
$$= m^2 + E(x)$$

$= m^2 + m.$

$$V(x) = E(x^2) - [E(x)]^2$$

$$= m^2 + m - (m)^2$$

$$= m$$

$$\left.\begin{array}{l} V(x) = m. \\ E(x) = m. \end{array}\right\} \text{ For poisson.}$$

Next : $X \sim$ Geometric.

X = # trials up to 1^{st} success.

$$P(X=k) = \underbrace{(1-p)^{k-1}}_{k-1 \text{ fails}} . \underbrace{p}_{1 \text{ success}}$$

$$E(x) = 1/p$$

Why?

$$E(x) = \sum_{k=1}^{\infty} k . P(X=k)$$

$$= \sum_{k=1}^{\infty} k (1-p)^{k-1} p.$$

$$E(x) = p + 2(1-p)p + 3(1-p)^2 p + 4(1-p)^3 p + \ldots$$

Multiply across by $(1-p)$.

$$(1-p)E(x) = (1-p)p + 2(1-p)^2 p + 3(1-p)^3 p + \dots$$

Subtract.

$$\big(1-(1-p)\big)E(x) = p + (1-p)p + (1-p)^2 p + (1-p)^3 p + \dots$$

= Sum of geometric

$$= \frac{p}{1-(1-p)}$$

$$\big(1-(1-p)\big)E(x) = \frac{p}{1-(1-p)} \quad 1.$$

$$pE(x) = 1$$

$$\Rightarrow \underline{E(x) = \frac{1}{p}.} \quad \Big\} \text{ for geometric.}$$

$V(x)$? we leave this till we do "Generating function"

$X \sim$ HYPERGEOMETRIC.

reminder of HYPERGEOMETRIC: Population of N items

TYPE 1	TYPE 2
R	$N-R$

When we take a sample of 'n' items (selected without replacement)

count up number of <u>TYPE 1</u> in this sample

'X'

$$P(X=k) = \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}$$

NOTE: not like Bernoulli trials cuz prob of success of each trial is <u>not</u> the same! without replacement so 2nd trial = $\frac{R-1}{N-1}$

$$E(x) = n\left(\frac{R}{N}\right)$$

SWITCH TO CONTINUOUS RANDOM VARIABLES.

we will find $E(x), V(x)$ for many of the standard random variables.

- Suppose X has P.D.F (Prob density function) $f(x)$
  ie Prob dist of X is described by the P.D.F $f(x)$

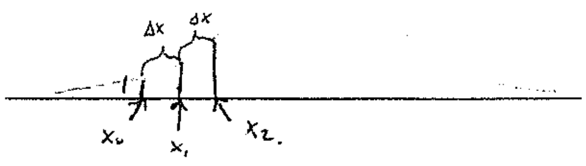
$$E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$$

provided that $\int |x| f(x)\, dx < \infty$

if not $\Rightarrow$ $E(x)$ doesn't exist.

ote: Connection with discrete case $\left[E(x) = \sum_{i=1}^{\infty} x_i P(X = x_i) \right.$

Continuous: $E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$



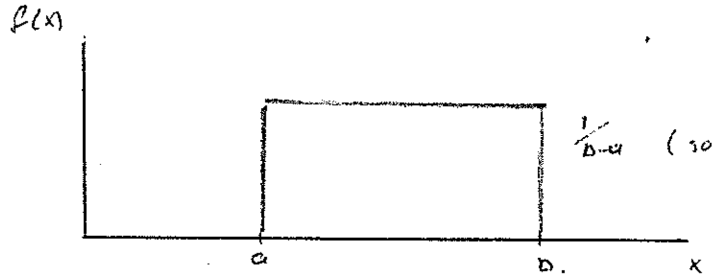
With numerical integration, we could try to evaluate

$\int x f(x)\, dx$ using $\sum x_i \underbrace{\left[f(x_i)\, \Delta x \right]}_{\text{Area of each strip}}$

For very small Δx

$$\lim_{\Delta x \to 0} \sum_{i=1}^{\infty} x_i \left[f(x_i)\,\Delta x \right] = \int x f(x)\,dx$$

Examples: $X \sim R(a,b)$ or $U(a,b)$ uniform



$\frac{1}{b-a}$ (so Area of □ = 1)

for (0, 1) eg.

$E(x) = x$

$f(x) = 1$

$$E(x) = \int_a^b x \frac{1}{(b-a)}\,dx$$

$$= \frac{1}{b-a} \int_a^b x\,dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right|_a^b$$

$$= \frac{1}{b-a} \left[\frac{b^2 - a^2}{2} \right] = \frac{b+a}{2} = E(x)$$

$$\int \left[x - \frac{b+a}{2} \right]^2 dx$$

$$\int x^2 \quad \frac{b^3 - a^3}{3}$$

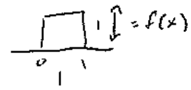
$$V(X) = E(x^2) - \left[E(x) \right]^2$$

N.B

$f(x) = \frac{1}{b-a}$

eg for $R(0,1)$

$$\left.\begin{aligned} V(x) &= \frac{(b-a)^2}{12}. \quad ✓ \\ E(x) &= \frac{b+a}{2} \end{aligned}\right\} R(a,b).$$



$$\frac{b^3 - a^3}{3(b-a)} - \frac{(b+a)^2}{4}$$

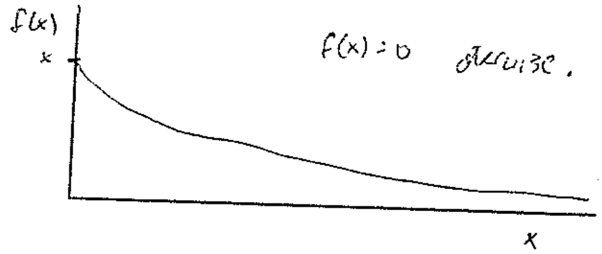
$$b^3 - a^3 = (b - a)^2 ($$

2 $X \sim$ Negative Exp (λ).

Time up to 1st occurance of poisson success (Time is continuous)

$\Rightarrow f(x) := \lambda e^{-\lambda x} \quad x \geq 0$

$f(x) = 0$ otherwise.



$E(x) = \int_0^\infty x f(x)\, dx$ we only cons from $0 \to \infty$ ie $x \in (0,\infty)$

$= \int_0^\infty x \lambda e^{-\lambda x}\, dx$

$= \int_0^\infty x \; d(-e^{-\lambda x})$ --- By parts

$= x(-e^{-\lambda x})\Big|_0^\infty - \int (-e^{-\lambda x})\, dx$

$= 0 + \int_0^\infty e^{-\lambda x}\, dx$.

$= -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^\infty$

$= -\frac{1}{\lambda}\left[0 - 1\right]$

$\Rightarrow E(x) = \frac{1}{\lambda}$

$V(x) = E(x^2) - [E(x)]^2$

$E(x^2) = \int_0^\infty x^2 f(x)\, dx$

$= \int_0^\infty x^2 \lambda e^{-\lambda x}\, dx$

use integration by Parts [Twice]

- lets try to use Gamma Function

$$\Gamma(\gamma) = \int_0^\infty t^{\gamma-1} e^{-t}\, dt.$$

- let $T = \lambda x$ $\quad dt = \lambda dx$

$$E(x^2) = \int_0^\infty \left(\frac{t}{\lambda}\right)^2 \lambda e^{-t} \frac{dt}{\lambda}$$

$$= \frac{1}{\lambda^2} \int_0^\infty t^2 e^{-t}\, dt.$$

$$= \frac{1}{\lambda} \Gamma(3) \quad \ldots \text{ ie } \gamma = 3.$$

But: $\Gamma(3) = 2[\Gamma(2)] = 2[(1)\Gamma(1)]$

$\therefore \Gamma(3) = 2$

$\Rightarrow E(x^2) = 2/\lambda^2$

$$V(x) = E(x^2) - [E(x)]^2.$$

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2}$$

$$V(x) = \frac{1}{\lambda^2}.$$

$$\left.\begin{aligned} E(x) &= \frac{1}{\lambda} \\ V(x) &= \frac{1}{\lambda^2} \end{aligned}\right\} \text{Negative exp.}$$

3) $X \sim \text{Gamma}(\alpha, \beta)$

$$E(x) = \int_0^\infty x \underbrace{\frac{1}{\Gamma(\gamma)} \left(\frac{x}{\beta}\right)^{\gamma-1} e^{-x/\beta} \frac{1}{\beta}}_{f(x)} \, dx$$

- Let $t = \dfrac{x}{\beta}$ $\quad dt = \dfrac{dx}{\beta}$.

$$E(x) = \int_0^\infty \frac{\beta}{\Gamma(\gamma)} t^{\alpha} e^{-t} \, dt$$

$$= \frac{\beta}{\Gamma(\gamma)} \int_0^\infty t^{\alpha} e^{-t} \, dt$$

$$= \frac{\beta}{\Gamma(\lambda)} \Gamma(\alpha+1)$$

$$= \frac{\gamma \beta \;\Gamma(\alpha)}{\Gamma(\alpha)} = E(x)$$

$$E(x) = \alpha\beta.$$

$$\text{Variance} = E(x^2) - [E(x)]^2$$
$$= E[x - E(x)]^2$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx. \quad f(x).$$

$$= \int_{-\infty}^{\infty} x^2 \cdot \overbrace{\frac{1}{\Gamma(\gamma)} \left(\frac{x}{\beta}\right)^{\gamma-1} e^{-x/\beta} \cdot \frac{1}{\beta}}^{f(x)}\, dx$$

$$= \frac{1}{\Gamma(\gamma)} \int_0^{\infty} \beta^2 \left(\frac{x}{\beta}\right)^{\gamma+1} e^{-x/\beta}\, \frac{1}{\beta}\, dx$$

$$T = x/\beta.$$

$$\frac{\beta^2}{\Gamma(\gamma)} \underbrace{\int_0^{\infty} t^{\gamma+1} e^{-t}\, dt}_{\Gamma(\gamma+2)}$$

N.B

$$\Gamma(\gamma) = \int_0^{\infty} t^{\gamma-1} e^{-t}\, dt$$

$$\Gamma(\gamma+1) = \gamma\, \Gamma(\gamma).$$

$$\Gamma(\gamma+2) = \gamma+1\, \Gamma(\gamma+1) = (\gamma+1)(\gamma)\, \Gamma(\gamma)$$

$$\frac{\beta^2}{\Gamma(\gamma)} (\gamma+1)(\gamma)\, \Gamma(\gamma) = \gamma(\gamma+1)\beta^2.$$

Thus: $V(x) = E(x^2) - [E(x)]^2$

$$= \gamma(\gamma+1)\beta^2 - (\gamma\beta)^2$$

$$V(x) = \gamma\beta^2.$$

$$\left.\begin{aligned} E(x) &= \gamma\beta. \\ V(x) &= \gamma\beta^2 \end{aligned}\right\} \text{gamma.}$$

$E(x) = \mu$

$E(x) = \sigma^2$

} Normal distribution,

MOMENTS FOR A RANDOM VARIABLE

These are more descriptive measures for a random var (and its distribution).

Consider $\underbrace{E[x-a]^k}_{k^{th} \text{ moment about the constant } a.}$ $k \in \mathbb{Z}^+$, $a \in \mathbb{R}$

condition: $\int |x-a|^k f(x)\,dx < \infty$

Can make choices for a and k

$A = 0$: we have $E[x]^k$

These are called moments about the origin

Denoted by μ_k'

but μ_1' is denoted by μ

$A = \mu$: we have $E(x-\mu)^k$

These are known as moments about the mean (μ) or central moments

Notation: $E(x-\mu)^k$ or μ_k

Most important of these is μ_2 which we have already seen

$V(x) = \mu_2 = Var(x) = \sigma^2$

Higher moments: when $k > 2$.

only $k = 3$ and (to a lesser extent) $k = 4$ are of interest.

$E(x-\mu)^3$ is supposed to tell us about the skewness of the dis of x.

$\dfrac{E(x-\mu)^3}{\sigma^3}$: coefficient of skewness.

Prob distributions for which $E(x)$ doesn't exist.

Discrete example.

1) $x = 1, 2, 3, 4, \ldots, n, \ldots \infty$

with $Prob[x = i] = k \dfrac{1}{i^2}$ for $i = 1, 2, \ldots$

Find k to ensure that $\sum_{i=1}^{\infty} Prob[x = i] = 1$

$\Rightarrow k \sum_{i=1}^{\infty} 1/i^2 = 1$

$k\left[\dfrac{\pi^2}{6}\right] = 1$

$k = \dfrac{6}{\pi^2}$

$E(x) = \sum x_i \, Prob[x = x_i]$

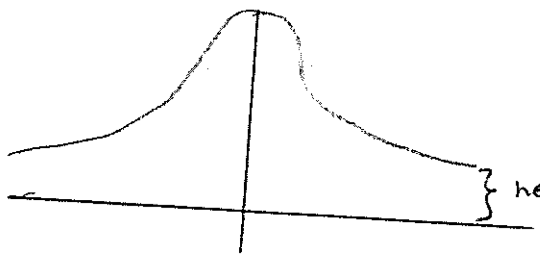
$E(x) = \sum_{i=1}^{\infty} i \left(\dfrac{6}{\pi^2}\right) \dfrac{1}{i^2}$

$= \dfrac{6}{\pi^2} \sum_{i=1}^{\infty} \dfrac{1}{i} = \infty \Rightarrow E(x)$ doesn't exist.

2 nice Thats the prob.

2) Continuous: X having cauchy distribution

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad \text{for} \quad -\infty < x < \infty$$



} heavy tailed dist. - more area under tails than $N(0,1)$.

Check Area under curve = 1.

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \qquad = \int_{-\infty}^{\infty} \frac{1}{\pi} \cdot \frac{1}{1+x^2}\,dx$$

let $x = \tan\theta \qquad d\theta = \sec^2\theta\, d\theta = 1 + \tan^2\theta$

$$\int_{-\pi/2}^{\pi/2} \frac{1}{\pi} \frac{1}{1+\tan^2\theta} \sec^2\theta\, d\theta \quad = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} d\theta = 1$$

$$E(x) = \int_{-\infty}^{\infty} |x|\, f(x)\, dx$$

$$= \int_{-\infty}^{\infty} |x| \frac{1}{\pi(1+x^2)}\, dx$$

cuz of symmetry

$$= 2\int_{0}^{\infty} \frac{x}{\pi(1+x^2)}\, dx$$

$$= 2 \cdot \frac{1}{2} \int_{0}^{\infty} \frac{d(x^2)}{\pi(1+x^2)} \quad = \frac{1}{\pi}\int_{0}^{\infty} \frac{d(x^2)}{(1+x^2)}$$

$$= \frac{1}{\pi} \ln(1+x^2) \Big]_0^\infty \rightarrow \infty$$

$\Rightarrow E(x)$ does not exist.

What does $E(x)$ not existing imply?

$E(x)$ is long term average value if exp. is observed a large number of time.

In the cauchy dist. large + or large − values occur quite often
(This is cuz of large area under tail).

---

note need lecture notes on this.

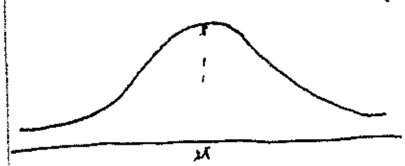
Linear Function of a Random Variable

Random Var. X

$y = a + bx$ (a, b constants)

We showed $E(ax+b) = aE(x) + b$

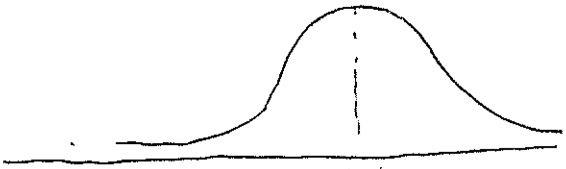
$$V(ax+b) = a^2 V(x)$$



μ

P. Den. F.

$Y = (1)X + b$.



$\mu + b$.

Same shape but shifted.

APPLICATION : n Bernoulli trials

x = # of successes

$x = b(n, p)$

$\hat{p}$ = Sample proportion

Consider $\hat{p} = \frac{x}{n}$ ← x is random. : $ax + 0$ where $a = \frac{1}{n}$.

random too

Possible values of $\hat{p}$: $0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}$.

we want $E(\hat{p})$, $V(\hat{p})$

$E(\hat{p}) = \frac{1}{n} E(x)$

$\Downarrow$

$\frac{1}{n}(np) = p$

So the expected value of our estimate proportion is the actual proportion unbiased

$V(\hat{p}) = \left(\frac{1}{n}\right)^2 V(x)$

$\Downarrow$

$\frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$.

Variance is $\div n$. so if we have more n, variance decreases x is more accurate we control the accuracy.

- If E(estimate) = actual thing we're estimating :
"unbiased" estimator.

- If we want to estimate p, we can think of using $\hat{p} = \frac{x}{n}$ as our estimator for p

$\because$ coz $E(\hat{p}) = p$ we see that $\hat{p}$ is an unbiased Estimator for p

- and also that the reliability of $\hat{p}$ as an estimator will improve as n is made large.

CONVERGENCE of a sequence of Random Var.

Consider $\hat{p}$ $\left(= \frac{x}{n}\right)$

This depends on n, so lets change notation
$\hat{p}$ now $= \hat{p}_n$

How does $\hat{p}_n$ behave as n gets larger.

This leads us to define:

$P|\,|X| > \varepsilon\,| \le \frac{E(|X|)}{a}$

convergence in probability:

Consider a sequence of random variables $\{X_n\}$
We say that the sequence converges in probability to the value c (a constant) if for every $\varepsilon > 0$ $\lim_{n \to \infty} \text{Prob}\left[\,|X_n - c| < \varepsilon\,\right] = 1$

For large enuf n, the difference between X_n and c can be as small as we want

With regard to the sequence $\{\hat{p}_n\}$, it turns out that $\hat{p}_n$ converges in probability to p

To show this we need a few more results.

In caught:

i) Markov Inequality:

Consider some random variable X and some function $g(x) \geq 0$ and constant $c > 0$

g operating on rand var.

rand variable.

$$\text{Prob}\left[g(X) \geq c \right] \leq \frac{E[g(X)]}{c}$$

rember: $E(Y) = \int_{\text{all } y} |y| f(y)\,dy$

PROOF: $E[g(X)] = \int_{\text{all } x} g(x) f(x)\,dx$

$$= \int_{\{\text{all } x \text{ such that } g(x) \geq c\}} g(x) f(x)\,dx + \int_{\{\text{all } x \text{ such that } g(x) < c\}} g(x) f(x)\,dx$$

~~both~~ 2^{nd} term is ≥ 0 since $g(x) \geq 0$

$$\Rightarrow E[g(X)] \geq \int_{\{\text{all } x \text{ such that } g(x) \geq c\}} g(x) f(x)\,dx$$

$$\geq \int_{\{\text{all } x \text{ s.t } g(x) \geq c\}} c\, f(x)\,dx = c \int_{\text{all } x \,:\, g(x) \geq c} f(x)\,dx \; *$$

<u>But</u> when we evaluate $\int_{\text{all } x \text{ s.t } g(x) \geq c} f(x)\,dx$ we are actually getting

$$P[g(X) \geq c]$$

Ex: To get $P(X > a)^{**} = \int_a^{\infty} f(x)\,dx$.

You integrate over interval that $**$ is true. $a \to \infty$

$$E[g(x)] \geq c P[g(x) \geq c]$$

$$P[g(x) \geq c] \leq \frac{1}{c} E[g(x)]$$

APPLICATION: Suppose X is possion $m = 5$

choose $g(x) = x$.

Note that (for x - poisson) $g(x) \geq 0$

lets choose $c = 10$

Then from Markov inequality

$$P[x \geq 10] \leq \frac{E[x]}{10}$$

from tables

$.0318 \leq .5$ / correct but not very sharp.

CHEBYSHEV INEQUALITY.

Suppose we have some Random variable X for which $V(x) = \sigma^2$ exists.

Then $$P[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$

where $\mu = E(x)$

$k \in \mathbb{Z}^+$ usually $k \geq 1$

"The prob that the rand. var differs from its mean by at least $k\sigma$ is at most $\frac{1}{k^2}$

For many distributions the actual value of $P[|x-\mu| \geq k\sigma]$ is considerably less than the $\frac{1}{k^2}$ given by the inequality.

<u>But</u> There are examples where the chebyshev bound is sharp:

x	80	100	120
Prob	$1/8$	$3/4$	$1/8$

$\underline{E(x) = 100} = \mu$ (by chance) $= 100$

$10 + 75 + 15 = 100$

$V(x) = E[x-\mu]^2$

$= \frac{1}{8}(80-100)^2 + \frac{3}{4}(100-100)^2 + \frac{1}{8}(120-100)^2$

$= \frac{400}{8} + 0 + \frac{400}{8}$

$\underline{V(x) = 100}$

$\underline{\sigma = 10}$ $\qquad \underline{k=2}$

lets get: $P[|x-\mu| \geq 2\sigma]$ for the dist

$= P[|x-100| \geq 20]$

$= P[(x=80) \text{ or } (x=120)] = \frac{1}{8} + \frac{1}{8} = 1/4$

Bound $= 1/k^2 = 1/4$

What's important about the chebyshev Inequality is that it applies to <u>all</u> probability distributions (as long as $E(x)$ & $V(x)$ exist).

Our principle application is to prove

Weak LAW of LARGE numbers.

Consider a seq of Bernoulli trials with Prob (succ) - p

Consider n trials, let $\hat{p} = \frac{x}{n}$ (x = # successes in n trials)

$\hat{p}$ estimate for p (prob of success) Think of heads / Tails

We will show that $\hat{p}_n$ converges in probability to p as $n \to \infty$

Remark: convergance in Prob.

we say that $\hat{p}_n$ converges in probability to p if for any positive $\varepsilon > 0$

$$\lim_{n \to \infty} P[\,|\hat{p}_n - p| < \varepsilon\,] = 1$$

ie $$\lim_{n \to \infty} \text{Prob}[\,p - \varepsilon < \hat{p}_n < p + \varepsilon\,] = 1$$

Proof: Our rand. var is $\hat{p}_n$ & weve already shown

$$E(\hat{p}_n) = p = \mu$$

$$V(\hat{p}_n) = \frac{p(1-p)}{n} = \sigma^2$$

Use chebyshev : $x = \hat{p}_n$

so.

$$P\left(\,|\hat{p}_n - p| \geq k\sqrt{\frac{p(1-p)}{n}}\,\right] \leq \frac{1}{k^2}$$

get complemntary event

$P(A) \le 1/k^2$

$P(A^c) = 1 - P(A)$

$- P(A) \ge -\frac{1}{k^2}$

$1 - P(A) \ge 1 - 1/k^2$

$$P\left[|\hat{P}_n - P| < k\sqrt{\frac{P(1-P)}{n}} \right] \ge 1 - 1/k^2$$

$$\varepsilon = k\sqrt{\frac{P(1-P)}{n}}$$

$$\varepsilon^2 = k^2\left(\frac{P(1-P)}{n}\right) \qquad \ge 1 - \frac{P(1-P)}{n\varepsilon^2}$$

$$1/k^2 = \frac{P(1-P)}{n\varepsilon^2}$$

as $n \to \infty$

$$\lim_{n\to\infty} P[|\hat{P}_n - P| < \varepsilon] \ge 1$$ cant be greater than 1

$$\lim_{n\to\infty} P[|\hat{P}_n - P| < \varepsilon] = 1$$

convergence in Prob.

Weak LAW of large numbers. See M10.

μ = Population mean.

$\bar{X}_n$: mean of sample of size n.

$$\bar{X}_n \xrightarrow{P} \mu$$

Biveriate & Multivariate Distributions

We consider 2 (or more) random variables simultaniously.

We want to know about their joint Probability Distribution

One can describe this using a bivariate distribution function.

$$F(x_1, x_2) = \text{Prob}[X_1 \le x_1 \cap X_2 \le x_2]$$

Proof: use markov Inequality.

- choose $g(x) = (x-\mu)^2$

note $g(x) \geq 0$ as required by markov.

- Choose $c = k^2\sigma^2$

Note that $E[g(X)] = E[(X-\mu)^2] = \sigma^2$

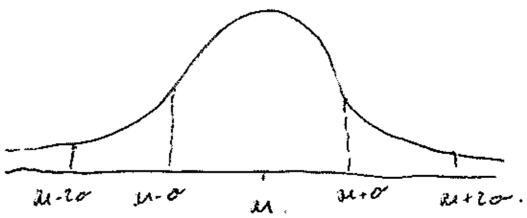
Consider markov for these choices:

$$P[(x-\mu)^2 \geq k^2\sigma^2] \leq \frac{E[g(x)]}{c} = \frac{\sigma^2}{k^2\sigma^2}$$

$$P[|x-\mu| \geq k\sigma] \leq \frac{1}{k^2}$$ - remember

Example: Suppose x is $N(\mu, \sigma^2)$ & ~~$k=2$~~.

Thus $P[|x-\mu| \geq 2\sigma] \leq \frac{1}{4}$ ← chebyshev bound.



$\mu-\sigma$, $\mu+\sigma$ at pts of inflection.

$|x-\mu| \geq 2\sigma$ corresponds to x values below $(\mu-2\sigma)$ and above $(\mu+2\sigma)$

$N(0,1)$ Tables Tell us its actually .0456

WEAK LAW OF LARGE NUMBERS

(A RESULT REQUIRED IN THE COURSE [illegible])

FIRST, WE LOOK AT THE CONCEPT OF CONVERGENCE IN PROBABILITY:

DEFN LET $Z_1, Z_2, \ldots$ BE A SEQUENCE OF RANDOM VARIABLES. THE SEQUENCE $\{Z_n\}$ CONVERGES IN PROBABILITY TO b IF FOR ANY GIVEN NUMBER $\varepsilon > 0$

$$\lim_{n \to \infty} P(|Z_n - b| < \varepsilon) = 1$$

(ROUGHLY SPEAKING, THIS SAYS THAT $\{Z_n\}$ CONVERGES IN PROBABILITY TO b IF THE PROB. DISTRIB OF Z_n BECOMES MORE AND MORE CONCENTRATED AROUND b AS $n \to \infty$.)

NOTATION: $\operatorname{plim}_{n \to \infty} Z_n = b$ $\boxed{\text{OR}}$ $Z_n \xrightarrow{P} b$

WEAK LAW OF LARGE NUMBERS (KHINTCHINE'S THEOREM)

IF THE COMMON DISTRIBUTION OF THE INDEPENDENT, IDENTICALLY DISTRIBUTED VARIABLES $X_1, X_2, \ldots$ HAS A FINITE FIRST MOMENT μ, THEN FOR THE SEQUENCE OF AVERAGES $\overline{X}_n = \sum_1^n X_i$, WE HAVE

$$\operatorname{plim}_{n \to \infty} \overline{X}_n = \mu$$

PROOF: WE SHALL ASSUME THAT THE DISTRIBUTION OF THE X_i HAS A FINITE VARIANCE σ^2. (THEOREM ALSO HOLDS WHEN VARIANCE DOES NOT EXIST, I.E. INFINITE VARIANCE).

WE USE THE CHEBYSHEV INEQUALITY:

$$P[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$

$$\Rightarrow P[|X - \mu| < k\sigma] \geq 1 - \frac{1}{k^2}$$

{R. VAR X, WITH MEAN μ, VAR σ^2

NOW FOR THE R. VAR $\overline{X}_n$, $E\overline{X}_n = \mu$

AND $V(\overline{X}_n) = \sigma^2/n$ {SINCE X_i ARE iid

FOR ANY $\varepsilon > 0$, CHOOSE k SO THAT $\frac{k\sigma}{\sqrt{n}} = \varepsilon$ [I.E. $k = \frac{\varepsilon\sqrt{n}}{\sigma}$

$$\Rightarrow P[|\overline{X}_n - \mu| < \varepsilon] \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$$

$$\Rightarrow \lim_{n \to \infty} P[|\overline{X}_n - \mu| < \varepsilon] = 1$$

(I.E. $\operatorname{plim}_{n \to \infty} \overline{X}_n = \mu$)