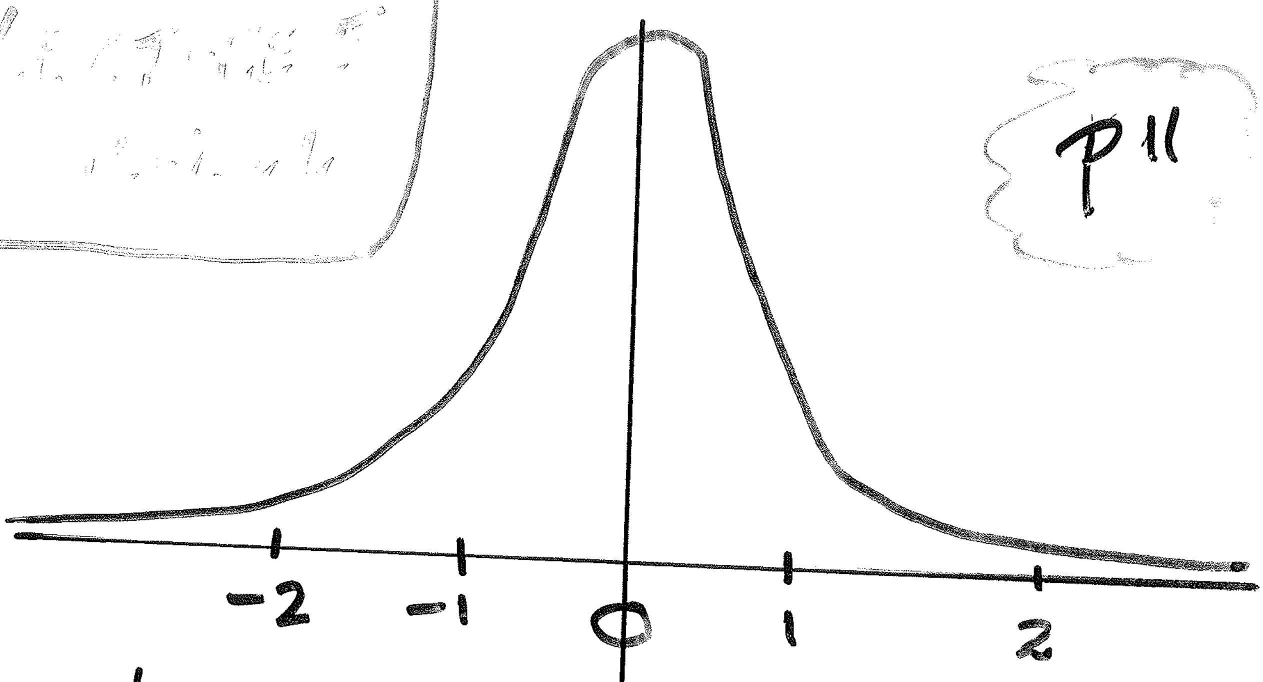
The Normal Distribution

The most important distribution in the whole of probability and statistics is the Normal Distribution

The random variable Z has pdf

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < Z < \infty$$

p 11



Properties

1. Symmetric about y axis.
2. Area under curve $= 1$.
3. Mean $= 0$, Variance $= 1$.

Mean.

$$\mu = \int_{-\infty}^{\infty} z\, f_Z(z)\, dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z\, e^{-z^2/2}\, dz$$

[Trivially the integrand is 'odd', but prove using integration by parts, setting $x = z^2/2$; $dx = z\,dz$.]

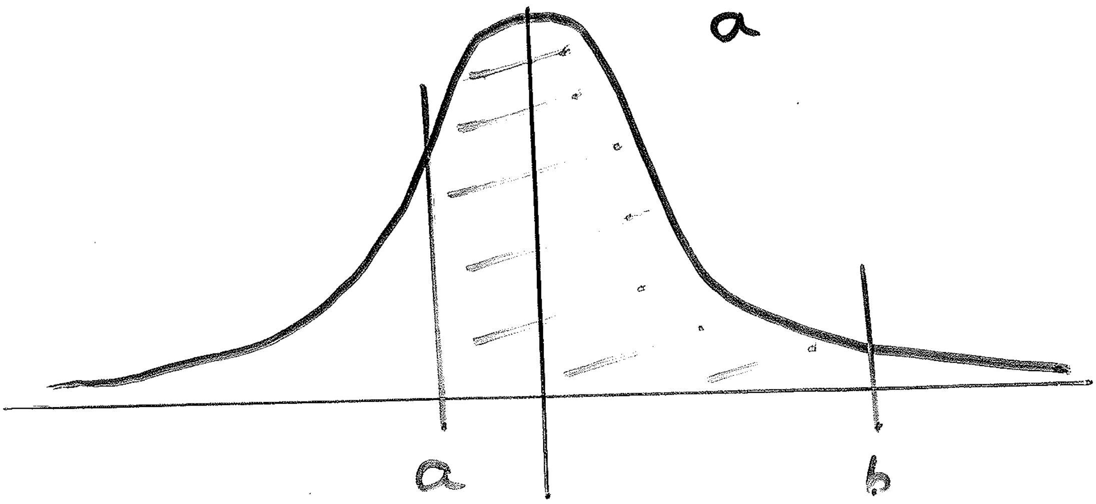
Variance

$$\sigma^2 = \int_{-\infty}^{\infty} (z-\mu)^2 f_Z(z)\, dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (z-0)^2 e^{-z^2/2}\, dz$$

$$= 1$$

[proof, more difficult]

$$P(a < Z < b) = \int_a^b f_Z(z)\, dz$$



needs to be tabulated

1.73

.958

NORMAL PROBABILITY TABLES

Table of $\varphi(x)$

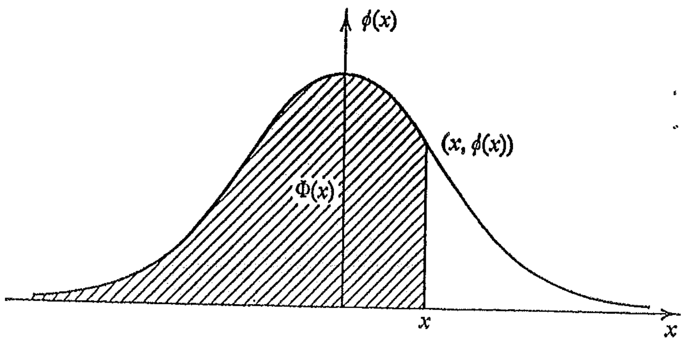
x	$\phi(x)$	x	$\phi(x)$	x	$\phi(x)$	x	$\phi(x)$
0·0	·399	0·8	·290	1·6	·111	2·4	·022
0·1	·397	0·9	·266	1·7	·094	2·5	·018
0·2	·391	1·0	·242	1·8	·079	2·6	·014
0·3	·381	1·1	·218	1·9	·066	2·7	·010
0·4	·368	1·2	·194	2·0	·054	2·8	·008
0·5	·352	1·3	·171	2·1	·044	2·9	·006
0·6	·333	1·4	·150	2·2	·035	3·0	·004
0·7	·312	1·5	·130	2·3	·028		

Table of $\Phi(x)$

x	0	1	2	3	4	5	6	7	8	9
0·0	·500	·504	·508	·512	·516	·520	·524	·528	·532	·536
0·1	·540	·544	·548	·552	·556	·560	·564	·567	·571	·575
0·2	·579	·583	·587	·591	·595	·599	·603	·606	·610	·614
0·3	·618	·622	·626	·629	·633	·637	·641	·644	·648	·652
0·4	·655	·659	·663	·666	·670	·674	·677	·681	·684	·688
0·5	·691	·695	·698	·702	·705	·709	·712	·716	·719	·722
0·6	·726	·729	·732	·736	·739	·742	·745	·749	·752	·755
0·7	·758	·761	·764	·767	·770	·773	·776	·779	·782	·785
0·8	·788	·791	·794	·797	·800	·802	·805	·808	·811	·813
0·9	·816	·819	·821	·824	·826	·829	·831	·834	·836	·839
1·0	·841	·844	·846	·848	·851	·853	·855	·858	·860	·862
1·1	·864	·867	·869	·871	·873	·875	·877	·879	·881	·883
1·2	·885	·887	·889	·891	·893	·894	·896	·898	·900	·901
1·3	·903	·905	·907	·908	·910	·911	·913	·915	·916	·918
1·4	·919	·921	·922	·924	·925	·926	·928	·929	·931	·932
1·5	·933	·934	·936	·937	·938	·939	·941	·942	·943	·944
1·6	·945	·946	·947	·948	·949	·951	·952	·953	·954	·954
1·7	·955	·956	·957	·958	·959	·960	·961	·962	·962	·963
1·8	·964	·965	·966	·966	·967	·968	·969	·969	·970	·971
1·9	·971	·972	·973	·973	·974	·974	·975	·976	·976	·977
2·0	·977	·978	·978	·979	·979	·980	·980	·981	·981	·982
2·1	·982	·983	·983	·983	·984	·984	·985	·985	·985	·986
2·2	·986	·986	·987	·987	·987	·988	·988	·988	·989	·989
2·3	·989	·990	·990	·990	·990	·991	·991	·991	·991	·992
2·4	·992	·992	·992	·992	·993	·993	·993	·993	·993	·994
2·5	·994	·994	·994	·994	·994	·995	·995	·995	·995	·995
2·6	·995	·995	·996	·996	·996	·996	·996	·996	·996	·996
2·7	·996	·997	·997	·997	·997	·997	·997	·997	·997	·997
2·8	·997	·997	·997	·997	·997	·997	·997	·998	·998	·998
2·9	·998	·998	·998	·998	·998	·998	·999	·999	·999	·999
3·0	·999	·999	·999	·999	·999	·999	·999	·999	·999	·999

The functions tabulated are
$$\phi(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} \quad \text{and} \quad \Phi(x) = \int_{-\infty}^{x} \phi(t)\,dt.$$

$\phi(x)$ is the ordinate of the Normal frequency curve. $\Phi(x)$ is the probability that a random variable having a Normal frequency density, with zero mean and unit variance, will be less than x.



General Normal Distribution

We call the standard normal distribution

$N(0, 1)$

mean variance

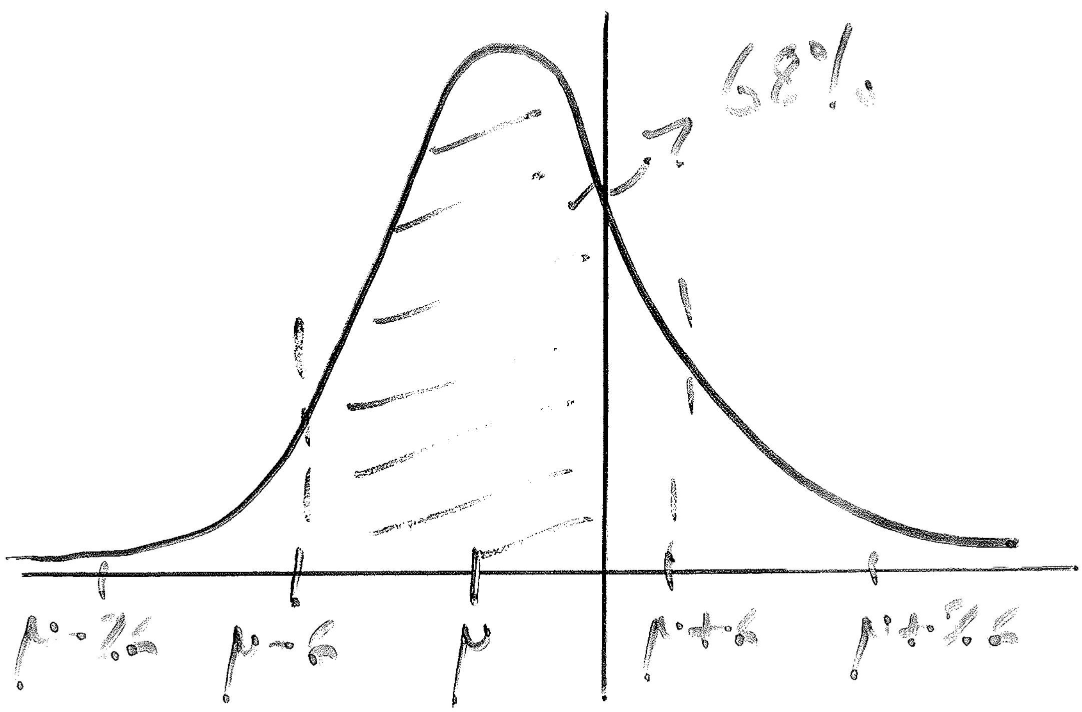
and the general distribution $N(\mu, \sigma^2)$.

It has pdf

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

but we can convert X to Z using

$$Z = \frac{X-\mu}{\sigma}$$



NORMAL PROBABILITY TABLES

Suppose the UK male population has an average height of adults over 21 of $\mu = 5$ feet 9 inches. $\sigma = 3$ inches.
What proportion lie between 5'7" and 6'2".

With $\mu = 69$, $\sigma = 3$, take

$$Z = \frac{X - 69}{3} \quad \left(\frac{X-\mu}{\sigma}\right)$$

We require $P(67 < X < 74)$

i.e. $P(-0.667 < Z < 1.667)$

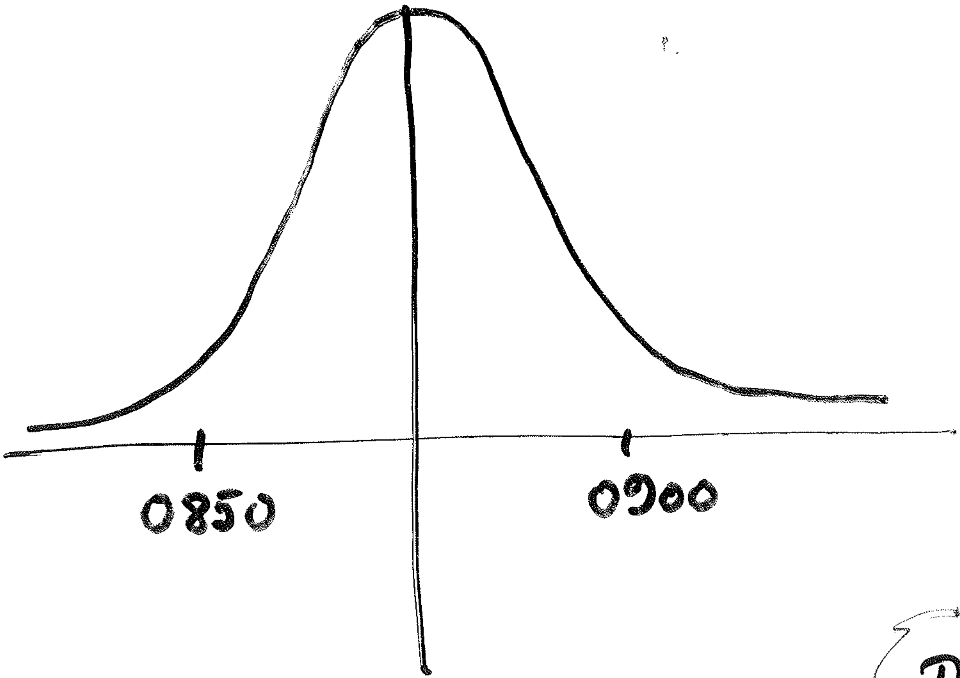
$= P(0 < Z < 0.667) + P(0 < Z < 1.667)$

$= 0.249 + 0.453 = 0.702$.

i.e. approx 70%.

P13

Commuter Exercise



p13

$$P(0850 < X < 0900)$$

$$Z = \frac{X - 0855}{2.5}$$

with $X_1 = 0850$, $X_2 = 0900$.

$$Z_1 = \frac{0850 - 0855}{2.5} = -2$$

$Z_2 = \quad 2$, similarly

$$P(-2 < Z < 2) = 0.954$$

Electrical Component

P 19

$\mu = 100, \quad \sigma = 5$

$$P(X > 105) = P\left(Z = \frac{X - 100}{5} > 1\right)$$
$$= 0.159$$

note $P(X > 90) = P(Z > -2)$
$$= 1 - .023 = .977$$

so $P(\overset{E}{X > 105} \mid \overset{F}{X > 90})$ $\quad P(E/F)$

$$= \frac{P(X > 105 \cap X > 90)}{P(X > 90)} \quad \frac{P(E \cap F)}{P(F)}$$

$$= \frac{P(X > 105)}{P(X > 90)} = \frac{.159}{.977}$$

$$\doteq .162$$

Normal Approximation to Binomial Distribution

The discrete Binomial distribution with mean np and variance npq can be approximated by the continuous Normal distribution, by setting

$$Z = \frac{X - np}{\sqrt{npq}}.$$

p15

n is large and neither p nor q, is close to zero.

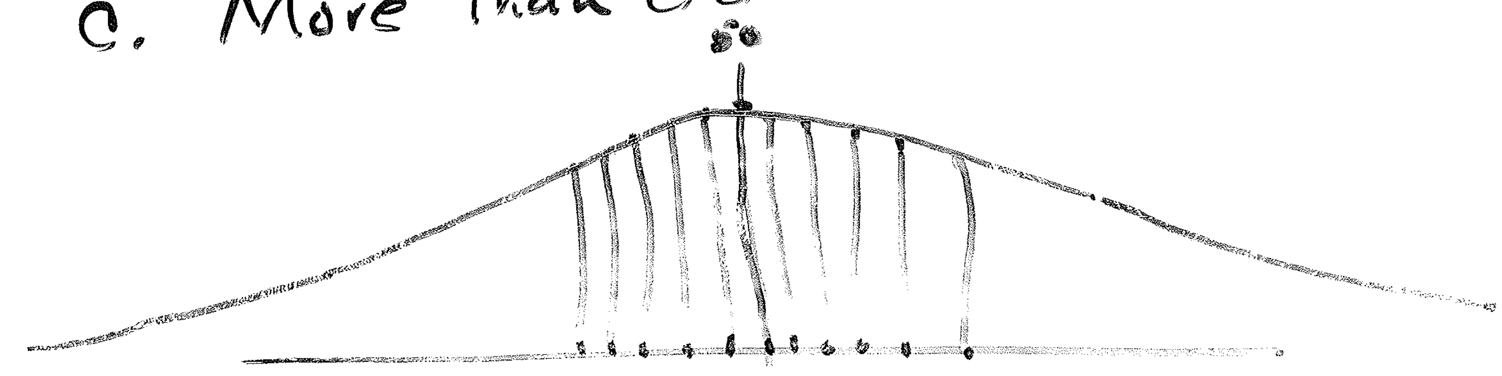
Example

A fair coin is tossed. Determine

a. The probability of exactly 5 heads in 10 tosses.

b. " " " exacty 50 heads in 100 tosses

c. More than 60 heads in 100 tosses.



(a) Binomial. $n = 10$ tosses, $p = q = 1/2$.

$$P(X=5) = \binom{10}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5$$

$$= 0.246$$

Normal, note $4.5 < X < 5.5$

$np = 5$, $npq = 2.5$

Equivalently, $Z_1 = \dfrac{-0.5}{\sqrt{2.5}} = -\dfrac{1}{\sqrt{10}}$

$\left(Z_2 = +\frac{1}{\sqrt{10}}\right)$

So we need $(Z = 1/\sqrt{10})$

$$2P\left(Z < \frac{1}{\sqrt{10}}\right) = 2P(Z \ldots .316)$$

$2 \times .124 = .248$

(b) We need $P(49.5 < X < 50.5)$

with $n = 100$, $p = q = 1/2$, $npq = 25$.

Reduces to $P(-0.1 < Z < 0.1)$

$= 0.0796$.

(c) $P(X > 60.5)$

$= P(Z > 2.1) = 0.01786$.

RANDOM NUMBERS (p.16)

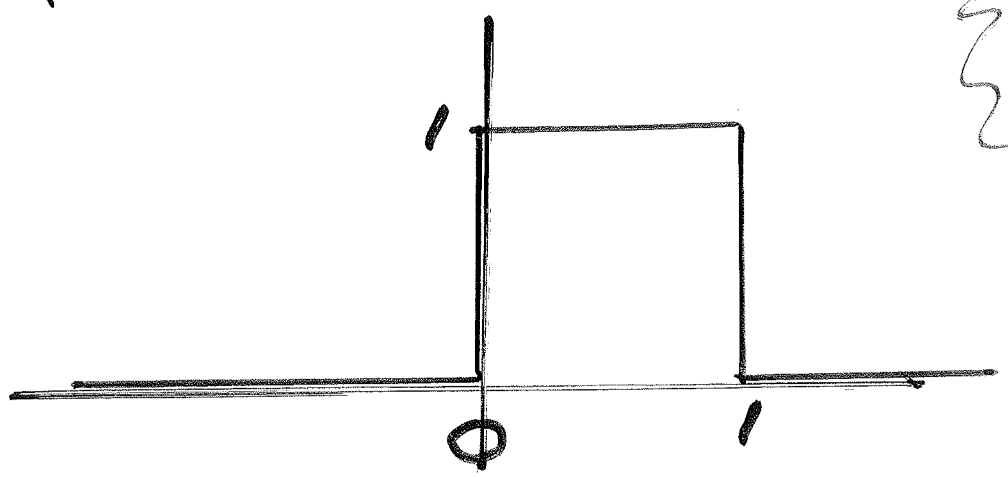
are used by engineers for the purpose of simulating random events. Applications are in signal processing and in the simulation of almost any random process.

If X is a random variable satisfying:

$$f_X(x) = 1, \quad 0 \le X \le 1$$
$$= 0, \quad \text{elsewhere}$$

then the pdf of X is:

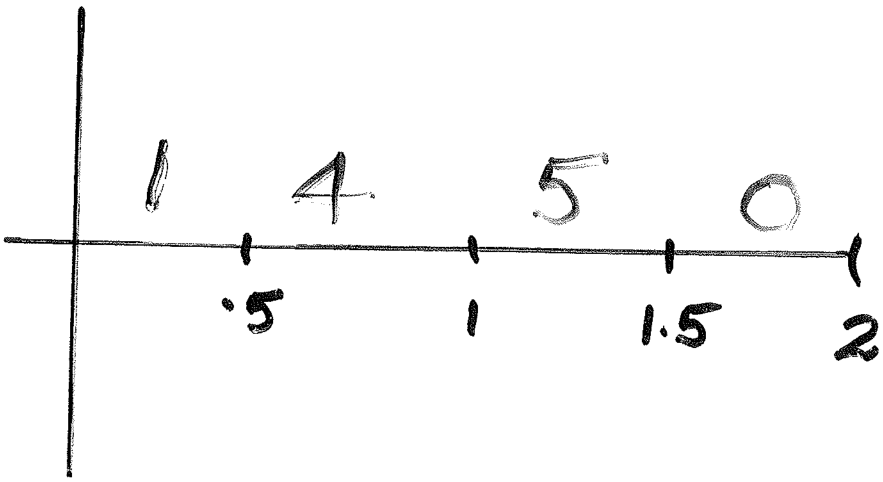
p16



Sum of Random Numbers: $Z = X + Y$

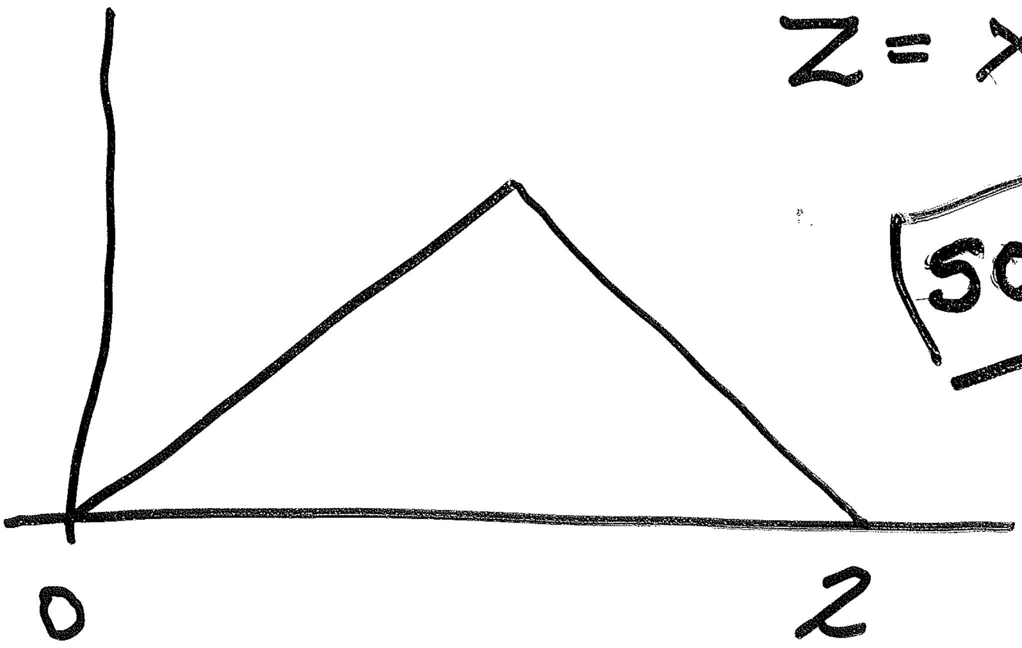
Note range of Z: $0 \le Z \le 2$

X	Y	Z
.226	.211	.437
.206	.871	1.077
.744	.036	.780
.129	.846	.975
.968	.461	1.429
.273	.606	.879
.281	.950	1.239
.481	.857	1.338
.064	.732	.796
.487	.842	1.329



$Z = X + Y$

$\boxed{\text{Scribbler}}$



L: $< .5$, H $> .5$

(LL, LH, HL, HH)

$E[X_i] = \mu$. $\qquad \sigma^2/n$.

$Var[X_i] = \sigma^2$. $\qquad n$

$E\left[\frac{X_i}{n}\right] = \frac{\mu}{n}$.

$E\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] = \frac{\mu}{n} \cdot n = \mu$.

$Var\left[\frac{X_i}{n}\right] = \sigma^2/n^2$.

$Var\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{\sigma^2}{\ } \cdot n^2$

$$E[aX] = \int_{\text{range}} a x f_X(x)\,dx = a\mu = a E[X]$$

Now $$Var[X] = \overset{*}{E}\left[(X-\mu)^2\right] = E[X^2] - \mu^2$$

$$= \int_{'r'} (x-\mu)^2 f_X(x)\,dx = \int_{'r'} (\underset{\mathrm{I}}{x^2} - \underset{\mathrm{II}}{2\mu x} + \underset{\mathrm{III}}{\mu^2}) f_X(x)\,dx$$

$$= \int_{'r'} x^2 \overset{\mathrm{I}}{f_X}(x)\,dx - 2\mu \int_{'r'} x \overset{\mathrm{II}}{f_X}(x)\,dx + \mu^2 \int_{'r'} \overset{\mathrm{III}}{f_X}(x)\,dx$$

$$E[X^2] \quad - \quad 2\mu\cdot\mu \quad + \quad \mu^2 \cdot 1 \quad *$$

$$= E[X^2] - \mu^2$$

$$Var[aX] = a^2\, Var[X] \quad +$$

$$\left(= E\left[(a(x-\mu))^2\right] \right)^+$$

So if

$$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$$

and each X_i has mean μ, variance σ^2

then $E[\bar{X}] = \mu$, $Var(\bar{X}) = \frac{\sigma^2}{n}$

Central Limit Theorem, pp 17-19

— note the theorm — pf.

Some results —

$$E[x] = \int_{\text{range}} x\, f_X(x)\, dx \quad (=\mu)$$

also $E[g(X)] = \int_{\text{range}} g(x) f_X(x)\, dx$

and $E[g+h] = \int (g+h) f_X(x)\, dx$

$$= E[g] + E[h]$$

So if $Z = X + Y$

$$E[Z] = E[X] + E[Y]$$

but

$$\mathrm{Var}[Z] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

iff X, Y _independent_ and thus _uncorrelated_.

Proof via — integral theorems.

From p 18-19, CLT says

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

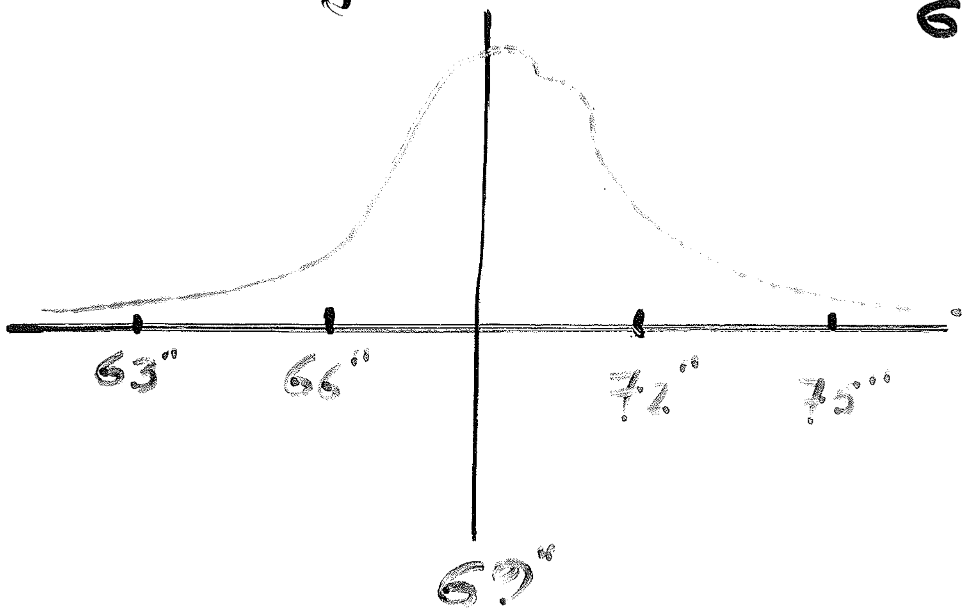
X – 'robust' distribution p. 18

$\bar{X}$ – 'normal' but pinched distribution p. 19.

Example

Take male population of UK, over 21.

Measure height $\mu = 69''$, $\sigma = 3''$



Now take samples, size 100

$\bar{X} \sim N\left(69, \left(\frac{3}{10}\right)^2\right)$ Mean 69", S/Dev 0.3"

Sample distributions:

$\bar{X}_1$, size n_1

from population (μ_1, σ_1^2) (X_1)

$\bar{X}_2$, size n_2,

from population (μ_2, σ_2^2) (X_2)

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Example (p. 24)

$$X \sim N(68, 3^2) = N(68, 9)$$

$\bar{X}$, size 25

$$\text{so } \bar{X} \sim N\left(68, \frac{9}{25}\right)$$

Standard deviation 0.6.