Colin Gay

Probability & Statistics

1) Probabilities + sets

2) Conditional Probability
   - Bayes Thm

3) Probability Density fn

4) Expectation Operator
   - Mean, Variance

5) Examples with 1 variable
   - Binomial
   - Gaussian
   - Poisson

6) P.d.f.'s with many variables

7) Law of Propagation of Errors

8) χ^2 disn

9) Estimation of Parameters

10) Least Squares Fit

11) Maximum Likelihood Fit

What we can do

Probability

Statistics

Probability Density fn
(p.d.f.)

If x is a continuous variable, the probability of throwing x s.t.

$a < x < a + da$ is

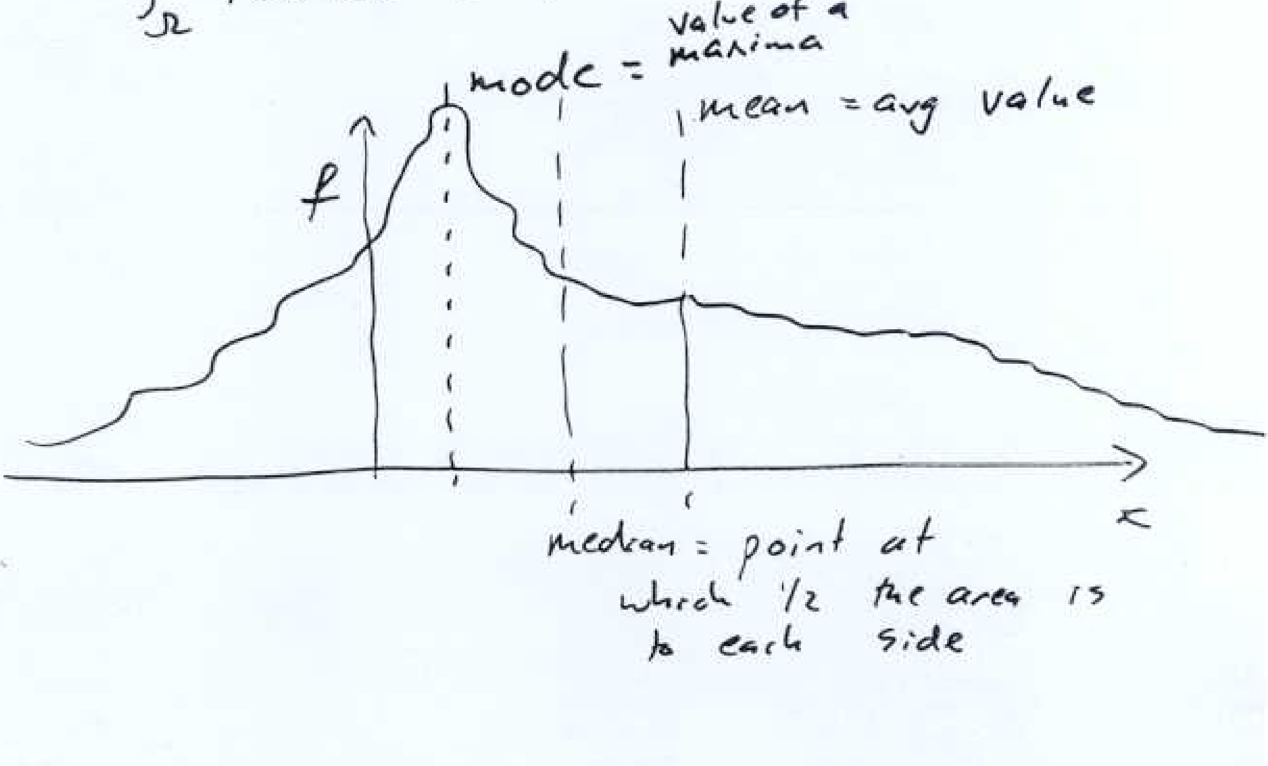
$P(a < x < a + da) = f(a)\,da$

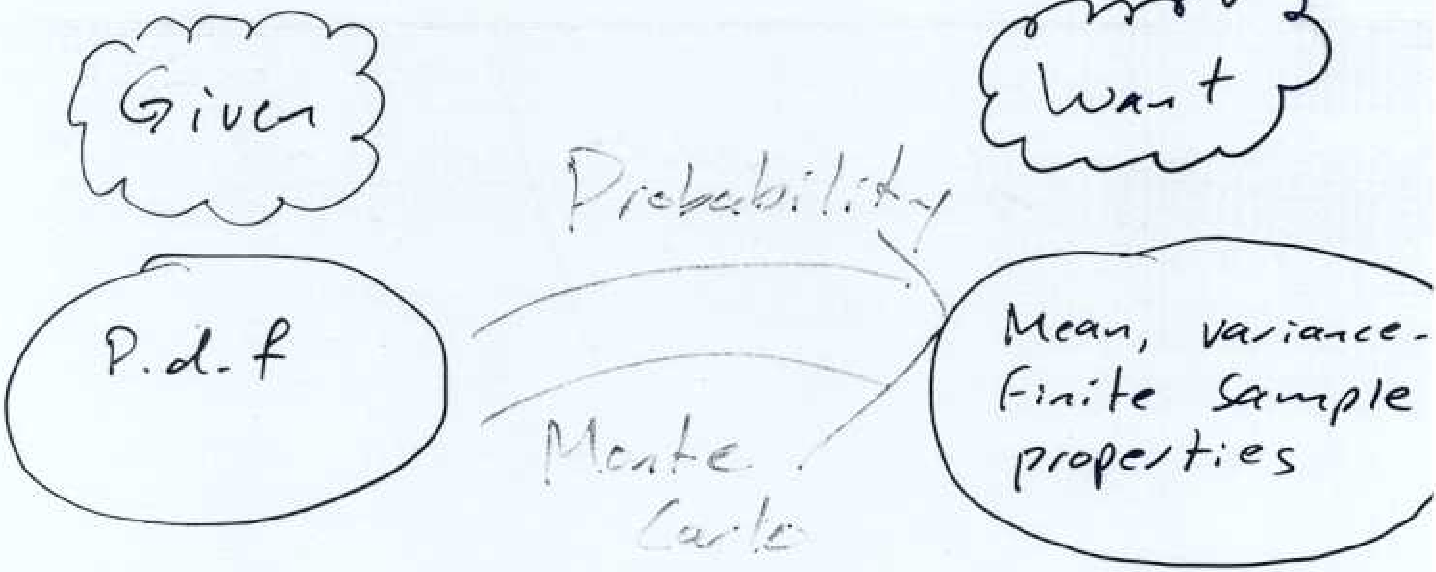
f is called the probability density function

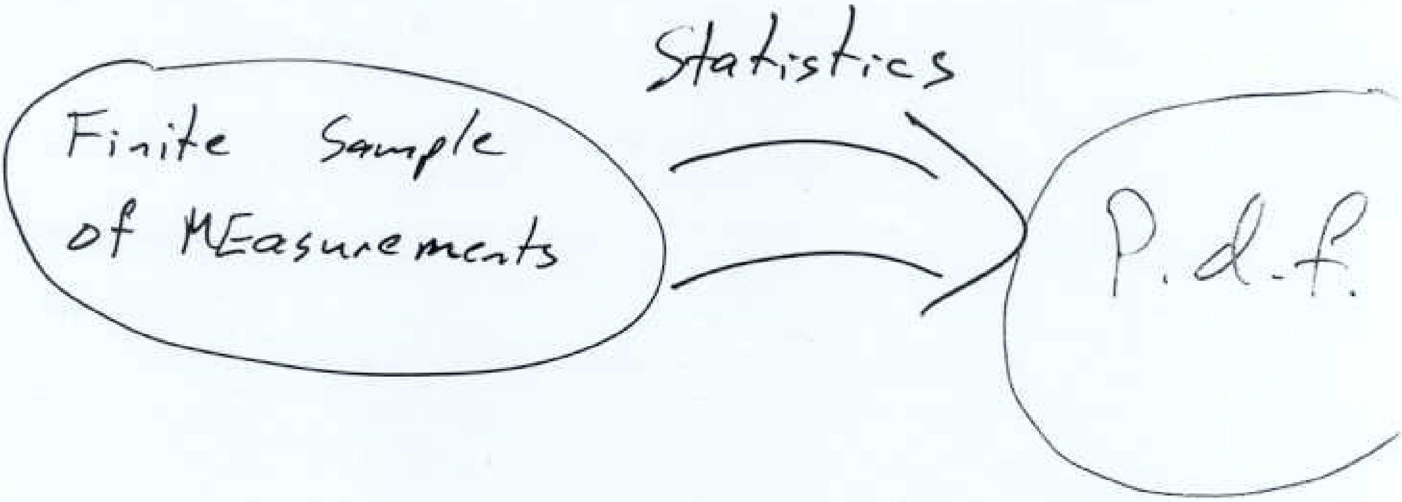
It has the properties

$f(x) \geq 0 \quad \forall x$

$\int_{\Omega} f(x)\,dx = 1$




Given
Want
Probability
P.d.f
Mean, variance-
Finite sample
properties
Monte
Carlo


Statistics
Finite Sample
of MEasurements
P.d.f.

Expectation Operator

Let g be any function and f a p.d.f.
The expectation value of g over f is

$$E(g(x)) = \int g(x) f(x) \, dx$$

Note that E is a linear operator, so

$$E(ah + bg) = aE(h) + bE(g)$$

Also, never forget that $E(g)$ is a <u>number</u>

Famous Expectations

Let x be a random variable

1) Mean:

The <u>mean</u> of a random variable x with p.d.f. $f(x)$ is

$$\mu \equiv E(x) = \int x f(x)\, dx$$

2) Variance:

The <u>variance</u> of x above is defined by

$$\sigma^2 \equiv V(x) = E((x - E(x))^2)$$

Expanding this out gives

$$\begin{aligned}\sigma^2 &= E(x^2 - 2x E(x) + E(x)^2) \\ &= E(x^2) - 2E(x)E(x) + E(x)^2 E(1) \\ &= E(x^2) - E(x)^2\end{aligned}$$

But I'm Discrete

If k is a discrete random variable with probability distribution $f(k)$, then the mean of k is

$$\mu = \sum k\, f(k)$$

and the variance is

$$\sigma^2 = \sum (k-\mu)^2 f(k)$$
$$= \sum k^2 f(k) - \mu^2$$

Binomial Dis^n

- Let some experiment have 2 possible outcomes, A + $\bar{A}$.
  Define $p = Prob(A)$ and. $q = 1-p = Prob(\bar{A})$

  After n independent trials the probability to have exactly r outcomes of A is

$$B(r; n, p) = \binom{n}{r} p^r (1-p)^{n-r}$$

This is called the Binomial distribution

- Mean:

$$\mu = E(r) = \sum_{r=0}^{n} r \, B(r; n, p)$$

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

$$= \sum_{r=1}^{n} \frac{r \, n!}{(n-r)! \, r!} p^r (1-p)^{n-r}$$

$$= \sum_{r=1}^{n} \frac{n!}{(n-r)! \, (r-1)!} p^r (1-p)^{n-r}$$

$$\Rightarrow \quad \mu = np \sum_{r=1}^{n} \frac{(n-1)!}{(n-r)!\,(r-1)!} p^{r-1} (1-p)^{n-r}$$

$$= np \sum_{r=1}^{n} \binom{n-1}{r-1} p^{r-1} (1-p)^{n-r}$$

$$= np \underbrace{\sum_{r=0}^{n-1} \binom{n-1}{r} p^{r} (1-p)^{n-1-r}}_{=1}$$

Recall that

$$(p+q)^{n-1} = \sum_{r=0}^{n-1} \binom{n-1}{r} p^r q^{n-1-r}$$

& in our case $p+q=1$, so

$$\boxed{\mu = np = E(r)}$$

— Variance:

First calculate something easier:

$$E(r(r-1)) = E(r^2) - E(r)$$

$$= E(r^2) - np$$

$$E(r(r-1)) = \sum_{r=0}^{n} r(r-1)\frac{n!}{(n-r)!\,r!}\, p^r(1-p)^{n-r}$$

$$= n(n-1)p^2 \sum_{r=2}^{n} \frac{(n-2)!}{(n-r)!\,(r-2)!}\, p^{r-2}(1-p)^{n-r}$$

$$= n(n-1)p^2 \underbrace{\sum_{r=0}^{n-2} \binom{n-2}{r} p^r (1-p)^{n-2-r}}_{1}$$

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

$$\Rightarrow E(r^2) = n(n-1)p^2 + np$$

$$= n^2p^2 - np^2 + np$$

$$V(r) = E(r^2) - E(r)^2$$

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$\boxed{V(r) = np(1-p)}$$

So what?

Histograms are binomial!

Let $A =$ entry falls in a given bin

$\bar{A} =$ entry is anywhere else

$n =$ total # of events

$r =$ # events in bin

so,

$$P(r) = B(r; n, p)$$

But what is p?

Best guess or estimate, is

$p = \frac{r}{n}$ (more later...)

With this value of p,

the mean number of entries in the bin is $np = n\frac{r}{n} = r$ (no kidding)

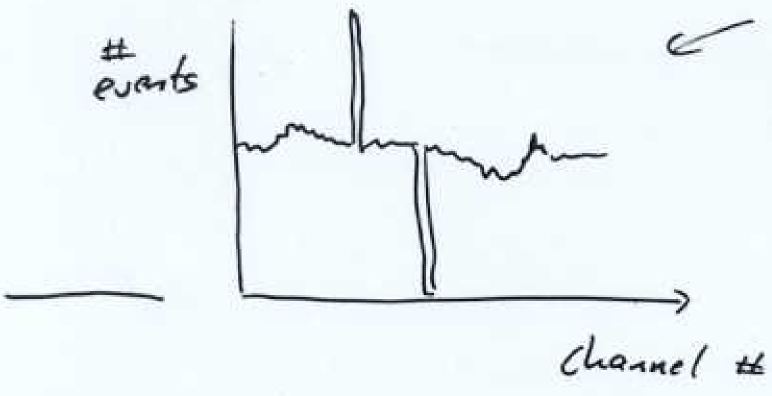
and

$$V(r) = np(1-p) = r\left(1 - \frac{r}{n}\right) \equiv \sigma^2$$

$$\Rightarrow \sigma = \sqrt{r\left(1 - \frac{r}{n}\right)}$$

Usually n is very large
(often total # events in experiment)
and $r \ll n$, so we approximate
$\sigma \sim \sqrt{r}$
Be very careful with this
approximation!

Eg 1: Occupancy of a silicon vertex detector
channel



each event, if channel hit, put entry in histogram

If a particular channel has
r entries,
$V(r) = r\left(1 - \frac{r}{n}\right)$
n = total # events.
Usually the occupancy $\frac{r}{n}$ is ~ few %
$\Rightarrow V(r) \simeq r$

If O = occupancy $= \frac{r}{n}$

$$V(O) = V\left(\frac{r}{n}\right) = \frac{1}{n^2} V(r) = \frac{r}{n^2}$$

prove this

$$\Rightarrow \sigma(O) = \frac{\sqrt{r}}{n} = \frac{\sqrt{O}}{\sqrt{n}}$$

Eg 2: Efficiency of a cut

Binomial - it passes or it doesn't

Let n = # events before cut

r = # events which pass cut

$$\varepsilon = \frac{r}{n}$$

$$V(\varepsilon) = V\left(\frac{r}{n}\right) = \frac{1}{n^2} V(r) = \frac{r}{n^2}\left(1 - \frac{r}{n}\right)$$

$$= \frac{\varepsilon(1-\varepsilon)}{n}$$

$$\Rightarrow \sigma(\varepsilon) = \frac{\sqrt{\varepsilon(1-\varepsilon)}}{\sqrt{n}}$$

clearly both ε, $1-\varepsilon$ nec in general

ubiquitous $\sqrt{n}$ improvement

Poisson Dis^n

Consider a binomial dis^n as

$n \to \infty$ but $np = \mu$ constant

Use Stirling's Approx

$$n! \approx \sqrt{2\pi n}\, n^n e^{-n}$$

$$B(r;n,p) = \frac{n!}{r!\,(n-r)!}\, p^r (1-p)^{n-r}$$

$$\approx \frac{1}{r!} \frac{\sqrt{2\pi n}}{\sqrt{2\pi(n-r)}} \frac{n^n e^{-n}}{(n-r)^{n-r} e^{-(n-r)}} \left(\frac{\mu}{n}\right)^r (1-$$

$$= \frac{1}{r!} \sqrt{\frac{n}{n-r}} \left(\frac{n}{n-r}\right)^{n-r} \frac{\mu^r}{e^r} \frac{\left(1-\frac{\mu}{n}\right)^n}{\left(1-\frac{\mu}{n}\right)^r}$$

$$= \frac{1}{r!} \sqrt{\frac{n}{n-r}} \frac{\left(1-\frac{r}{n}\right)^r}{\left(1-\frac{r}{n}\right)^n} \frac{\mu^r}{e^r} \frac{\left(1-\frac{\mu}{n}\right)^n}{\left(1-\frac{\mu}{n}\right)^r}$$

As $n \to \infty$ $\sqrt{\frac{n}{n-r}} \to 1$,

$\left(1-\frac{\mu}{n}\right)^n \to e^{-\mu}$, $\left(1-\frac{\mu}{n}\right)^r \to 1$ e

$$\lim_{\substack{n\to\infty,\\ np=\mu}} B(r;n,p) = \frac{1}{r!} \mu^r e^{-\mu}$$

$\equiv P(r,\mu)$, the Poisson Dis^n

Mean:

$$E(r) = \sum_{r=0}^{\infty} r P(r,\mu)$$

$$= \sum_{r=0}^{\infty} r \frac{\mu^r}{r!} e^{-\mu}$$

$$= \mu e^{-\mu} \sum_{r=1}^{\infty} \frac{\mu^{r-1}}{(r-1)!}$$

$$= \mu e^{-\mu} \sum_{r=0}^{\infty} \frac{\mu^r}{r!}$$

$$= \mu$$

Variance:

$$V(r) = E(r^2) - \mu^2$$

$$E(r(r-1)) = E(r^2) - \mu$$

$$= \sum_{r=2}^{\infty} r(r-1) \frac{\mu^r}{r!} e^{-\mu}$$

$$= \mu^2 e^{-\mu} \sum_{r=0}^{\infty} \frac{\mu^r}{r!}$$

$$= \mu^2$$

$$\Rightarrow E(r^2) - \mu^2 = \mu = V(r)$$

i.e. $\boxed{E(r) = V(r) = \mu}$

So What?

Another way of looking at Poisson!

Consider the radioactive decay of some atoms

Make 3 assumptions:

1) Any time interval $[t, t+dt]$ contains at most 1 decay
2) Prob. of the decay occurring in this interval is prop. to dt
3) Whether or not the atom decays in the interval $[t, t+dt]$ is independent of any other (non-overlapping) interval

1) + 2) $\Rightarrow$ $P_d(dt) = \lambda dt$

+ prob. of no decay in this interval is $P_0(dt) = 1 - \lambda dt$

3) $\Rightarrow$ prob of no decay by time $t+dt$

$$P_0(t+dt) = P_0(t)\, P_0(dt)$$
$$= P_0(t)(1 - \lambda dt)$$

$$\Rightarrow \quad P_0(t+dt) - P_0(t) = -\lambda dt$$

$$\Rightarrow \quad \frac{dP_0}{dt} = -\lambda$$

$$\boxed{P_0(t) = e^{-\lambda t}}$$

$(P_0(0) = 1)$

Prob of getting r decays in time $t+dt$ is

$$P_r(t+dt) = P_r(t)\, P_0(dt) + P_{r-1}(t)\, P_d(dt)$$

(see 1)

$$P_r(t+dt) = P_r(t)(1-\lambda dt) + P_{r-1}(t)\, \lambda dt$$

$$\Rightarrow \quad \frac{dP_r}{dt} = -\lambda P_r(t) + \lambda P_{r-1}(t)$$

Sol^n is $\quad P_r(t) = \frac{1}{r!}(\lambda t)^r e^{-\lambda t}$

In particle production we often have an analagous situation

N_{beam} = # beam particles is very large, so we approximate as continuous stream

If a process has x-section σ, + the beams have instantaneous luminosity $\mathcal{L}$ the rate of this process is

$$R = \sigma \mathcal{L}$$

Hence the # ocurrences ~~[illegible]~~ in time T is a Poisson variable with distribution

$$P(r, RT)$$

$$RT = \sigma \int \mathcal{L} \, dt$$

e.g. Top quark production

Normal (Gaussian) Dist^n

The p.d.f.

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

is called the <u>normal</u> or <u>gaussian</u> distribution

<u>Mean:</u>

Recall $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$

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

set $y = \frac{x-\mu}{\sigma}$, so $dx = \sigma dy$

$$E(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} (\sigma y + \mu) \, e^{-\frac{1}{2}y^2} \sigma dy$$

$$= \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy$$

$$= \mu$$

Variance

$$V(x) = E\left((x-\mu)^2\right)$$

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

set $y = \frac{x-\mu}{\sigma}$

$$V(x) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 \, e^{-\frac{1}{2}y^2} dy$$

$$= \sigma^2 \qquad \text{(integrate by parts)}$$

So What?

The normal distribution is the single most important probability density, because of the Central Limit Theorem.

This roughly states that if a variable $y = \sum_{i=1}^{n} x_i$, with each x_i being independent random variables with finite variances, then $f(y)$ is ~~asym~~ asymptotically normal as $n \to \infty$

Usually, an error on a measurement is made up of a very many small effects, and so most measurement errors end up being gaussian.

Addition of Normal variables

If x is distributed as $N(\mu_1, \sigma_1)$ + y is $N(\mu_2, \sigma_2)$, then

$z = x + y$ is distributed as

$N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Pf Left as an exercise

Important Eg:

Mean of n independent random variables

$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$; x_i is $N(\mu_i, \sigma_i^2)$

Then

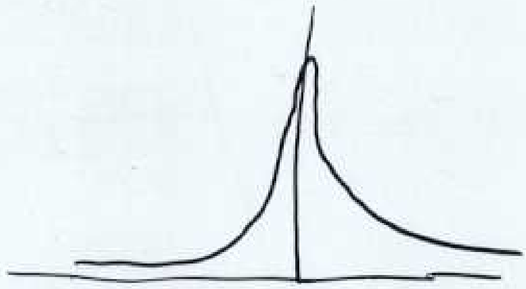
$\bar{x}$ is distributed as

$N(\frac{1}{n} \sum \mu_i, \frac{1}{n^2} \sum \sigma_i^2)$

In particular, if all x_i's distributions are the same $N(\mu, \sigma^2)$ (measure same quantity n times) then $\bar{x}$ is distributed as

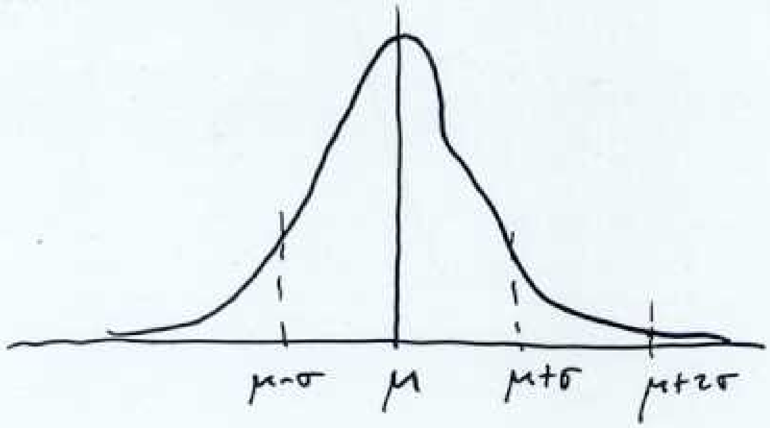
$N(\mu, \frac{\sigma^2}{n})$

$\Rightarrow$ n independent measurements give



$\bar{\sigma} = \frac{\sigma}{\sqrt{n}} \leftarrow$ again!

Probability Intervals



Some common intervals:

$P(\mu-\sigma \le x \le \mu+\sigma) = 0.683$ "1σ"

$P(\mu-2\sigma \le x \le \mu+2\sigma) = 0.955$ "2σ"

$P(\mu-3\sigma \le x \le \mu+3\sigma) = 0.997$ "3σ

$P(\mu-1.645\sigma \le x \le \mu+1.645\sigma) = 0.90$