Experimental Data

Many measurements deal in raw signals

- often from many detector elements
  - ADC counts giving ionization charge
  - signal giving transition radiation and ionization for particle identification

Even in simplest cases

- steps needed to arrive at most useful physical quantity

Energy lost in calorimeter cells

- sampling fractions, High Voltage correction, pulser calibration...

3D spacepoint in tracker

- drift time corrections, wire sag, magnetic field...

Reconstruction

1/2

Often, further calculations needed to "reconstruct" observables for physical particles

- Hits : track reconstruction : $\vec{p}$ track
- E in cells : cluster reconstruction : E, ϕ, η of jet

Requires proper use of knowledge of experimental uncertainties

- statistics of probability distributions
- optimization

Dealing with Experimental Data

- nature has inherently random processes

- measurements have uncertainties due essentially to a random element in making measurements

Statistics study of this randomness given ~~this~~ an underlying probability distribution

- also important in studying + optimizing detector design

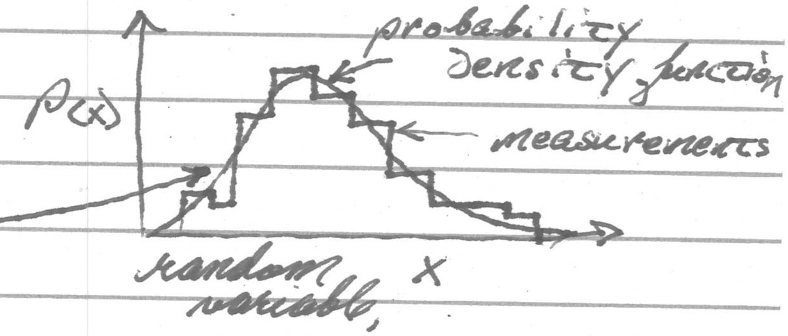
∴ statistics critical to study of particle physics

Probability Distributions

We consider random processes as described by a "probability distribution", $P(x)$

- this is the 'universe' of all cases of this process
- we 'sample' this distribution when we make measurements

$$\int P(x)\,dx = 1$$



→ if x discrete: $P(x)$ gives frequency for each case

→ if x continuous: speak of probability in window x to $x+dx$

Integral Distributions:

$$P(x_1 \le x \le x_2) = \int_{x_1}^{x_2} P(x)\,dx \quad \left(\text{or} = \sum_{i=1}^{2} P(x_i)\right)$$

if discrete x

- normalization

Expectation Values and Moments

Expected value of x

$$E(x) = \int x\, P(x)\, dx \qquad \left(= \sum_i x_i P(x_i)\right)$$

↪ all allowed values of x

The 'r'th moment of variable x around a point x_0 = $E((x-x_0)^r)$

Mean:

$$\boxed{\mu = E(x) = \int x\, P(x)\, dx}$$

- Technically the "theoretical mean" as $P(x)$ is function from universe
- experimental measurement is an estimate of this

Variance:

$$\boxed{\sigma^2 = E((x-\mu)^2) = \int (x-\mu)^2 P(x)\, dx}$$

σ is "standard deviation"

- gives "width" of distribution around mean

Covariance:

Consider process described by multiple variables

$$P(x, y, z \ldots)$$

eg. height + weight of individuals

- variables may be correlated partially or fully

Define covariance $\boxed{cov(x,y) = E\left((x-\mu_x)(y-\mu_y)\right)}$

means in x and y

- for each pairing of variables
  eg. $cov(x,y)$, $cov(y,z)$, $cov(z,x)$
  for x, y, z space

- a measure of the <u>linear</u> correlation between two variables
- correlation coefficient

$$\boxed{\rho = \frac{cov(x,y)}{\sigma_x \sigma_y}}$$

standard deviations of x and y

$$= -1 \ldots +1$$

$\rho = 0$ (independent) $\rho = -1$ (anticorrelated)

Binomial Distribution

Let us have a process which can only have one of two values (eg. coin toss)

Probability of r successes in N tries

$$P(r) = \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r}$$ number failures

where p is probability of success for each trial.

Gives mean, $\underline{\mu = Np}$ and variance

$$\boxed{\sigma^2 = Np(1-p)}$$

When N is large (>30), and $p \geq 0.05$, can approximate by a Gaussian distribution

When N large and $x \leq 0.05$, can approximate as a Poisson distribution

Efficiency Calculation

If you have a measurement where either detect or identify a particle : binomial statistics

- trigger efficiency
- tracking efficiency ...

detect/identify ⇒ # successes = r
don't " " ⇒ # failures = $N-r$

One's estimate of the mean efficiency, $\epsilon = r/N$

estimated Error on this vs. theoretical value, p

$$\delta\epsilon = \frac{\sigma}{N} = \boxed{\sqrt{\frac{\epsilon(1-\epsilon)}{N}}}$$

Example : muon trigger efficiency

100 events with μ's : 99 detected

$\epsilon = 99\%$; $\delta\epsilon = 1\%$ $(99 \pm 1)\%$

1000 events with 990 triggered

$\epsilon = 99\%$; $\delta\epsilon = 0.3\%$ $(99.0 \pm 0.3)\%$

Poisson Distribution

Special case of binomial distribution when
$p \to 0$ and $N \to \infty$
+ mean Np stays finite

Probability of observing r events

$$P(r) = \frac{\mu^r e^{-\mu}}{r!}$$

- only mean μ appears
- no explicit mention of N or p

This gives variance $\sigma^2 = \mu$

$$\therefore \boxed{\sigma = \sqrt{\mu}}$$

→ reason that measurements depending on counting have $\sqrt{N}$ behavior

Sensible to use for radioactive decay, or particle interactions (ie. very small probabilities)

Radioactive Decay

Consider 10^{-6} g of $^{137}Cs \to 10^{15}$ nuclei

- probability for each nucleus to decay

$$p = \underline{8 \times 10^{-10}/s}$$

$\to \mu = Np \sim 8 \times 10^{5}$ decays/s

In many cases, we know the mean counting rate $\to$ eg. # reactions/sec.

$\to$ Want to know probability of observing r events in τ units

$$P(r) = \frac{(\lambda\tau)^r}{r!} e^{-\lambda\tau}$$

where λ is rate in τ units

In above example, probability of seeing 1 event in 3s is

$$P(1) = 2.4 \times 10^{6} \, e^{-2.4 \times 10^{6}}$$
$$\sim \underline{\underline{0}} \quad \text{(will see many more)}$$

In typical experiment, if observe for 1 sec, error on rate

$$\frac{\sigma}{\mu_E} = \frac{1}{\sqrt{\mu}} = \sim \underline{\underline{10^{-3}}}$$

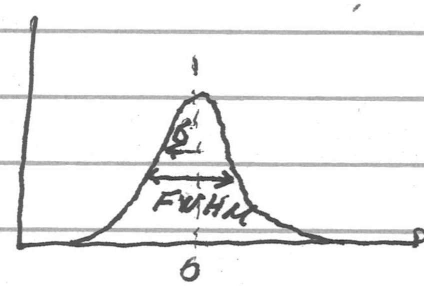
Gaussian Distribution

A distribution for a continuous variable

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

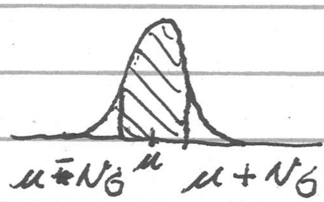
→ μ + σ^2 are mean and variance of distribution
- a symmetric distribution in x

FWHM (full-width at half maximum) $= 2.35\sigma$



Gaussian is a limiting case for Poisson if also switch to x being a continuous variable
→ as μ becomes large, Poisson gets more symmetric + adheres to Gaussian form

Integral intervals:



$N=1$: 68.3%
$N=2$: 95.5%
$N=3$: 99.7%

Chi Squared Distribution

To test goodness-of-fit, define for n independent random variables, x_i
→ have Gaussian probability densities

$$u \equiv \chi^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2$$

The probability distribution of u is

$$P(u)\,du = \frac{\left(\frac{u}{2}\right)^{\left(\frac{\nu}{2}-1\right)} \exp\left(-\frac{u}{2}\right)}{2\,\Gamma(\nu/2)}$$

where ν is # of independent variables in sum
Γ is gamma function.

$$\mu = \nu \qquad \left(\text{i.e. } \boxed{\mu/\text{dof} \approx 1}\right)$$
$$\sigma^2 = 2\nu \qquad \left(\text{i.e. } \sigma/\text{dof} = \sqrt{2/\nu}\right)$$