NEXT 5 LECTURES : how we use statistics

Some references

- BARLOW, A Guide to the use of Stat. Methods in the Phys. Sci
- G. COWAN, Statistical Data Analysis
- K. CRANMER, "Practical Statistics for the LHC"

THESE NOTES FOLLOW Peskin (SLAC STATISTICS WEEK 2017)

messy, uncertain, imperfect

HOW DO WE USE STATISTICS?

BRIDGE BETWEEN EXPERIMENTAL DATA & THEORETICAL MODELS

given a set of measurements,

... what is the value of a parameter?
... what is the allowed range of a parameter?
... is one model more consistent than another?

did i discover something?

PROBABILITY

INTIMATELY TIED TO PHYSICS
- QUANTUM
- STATISTICAL
- DATA

$P(A)$ is the probability for A

eg result of a coin toss, whether there is a higgs boson in your data, whether Schr's cat is dead or alive, ...

$0 \le P(A) \le 1$

DEFINITELY NOT TRUE (0), DEFINITELY TRUE (1)

& only the Sith deal in absolutes

$$P(A \& B) = P(A)\,P(B)$$

if A & B ARE MUTUALLY EXCLUSIVE, ie have nothing to do w/ each other

compare to entangled states

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

$|\uparrow\rangle_A \quad |\downarrow\rangle_B$

IF THERE ARE N POSSIBLE VALUES FOR A,

$A \in \{a_1, \ldots, a_N\}$

eg $\{1, 2, \ldots, 6\}$ for a 6 sided die

THEN: $\boxed{\sum_i P(a_i) = 1}$ "conservation of probability"

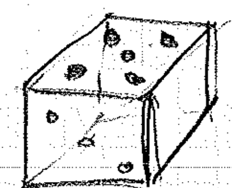
sum over all mutually exclusive outcomes.

CONDITIONAL PROBABILITY ← important

SUPPOSE A & B ARE NOT INDEPENDENT.

eg. ROLL A DIE. OBSERVE
① # ON TOP
② # ON SIDE CLOSEST TO YOU

if this is 5



BOTTOM is 2

if you know top face is a 5,
then side can be $\{1, 3, 4, 6\}$
w/ equal prob.;

CANNOT BE 2.

THE PROBABILITY OF A GIVEN INFORMATION B IS
OR ASSUMPTION

$$\boxed{P(A|B) = \frac{P(A \,\&\, B)}{P(B)}}$$

← DIVIDING BY THE PROB of B – WE'RE ASSUMING B IS TRUE.

P(A) GIVEN THAT B IS TRUE

then if $\{b_i\}$ IS A SET OF MUTUALLY EXCLUSIVE POSSIBILITIES

$$\sum_i P(A|b_i)\, P(b_i) = \sum_i P(A \,\&\, b_i) = P(A)$$

Bayes' Thm: $P(A|B)P(B) = P(A \,\&\, B) = P(B \,\&\, A)$
$= P(B|A)\, P(A)$

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

WHY THIS IS IMPORTANT: THIS RELATES

$$\underbrace{P(\text{DATA} \mid \text{THEORY})}$$ to $P(\text{THEORY} \mid \text{DATA})$

something we can usually calculate

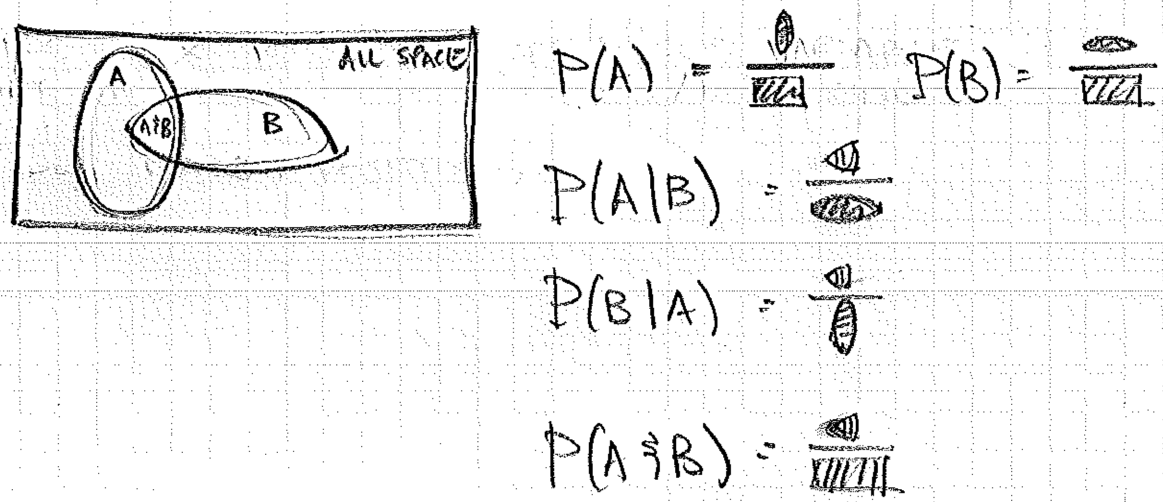
GIVEN PHYSICAL THEORY & INSTRUMENTAL EFFICIENCIES, CAN DETERMINE PROBABILITIES OF DIFF. EXPERIMENTAL OUTCOMES

we usually have this

& want to test this

but we want to answer: HOW LIKELY IS MY HYPOTHESIS?

VISUAL REPRESENTATION (from Bob Cousins)



$P(\text{USES ROOT} \mid \text{PARTICLE PHYSICIST}) \sim 50\%$ ← theorists don't use root

$P(\text{PARTICLE PHYSICIST} \mid \text{USES ROOT}) \sim 100\%$ ← only HEP (ex) uses ROOT

$P(\text{PREGNANT} \mid ♀) \sim 3\%$

$P(♀ \mid \text{PREGNANT}) \sim 100\%$

CONTINUOUS VARIABLES : x eg mass, position, ...

$\mathbb{P}(x=3)$ doesn't make sense .... formally 0.

$x = 3.000\ldots$
means $x \neq 3.\underbrace{000\ldots0}_{\text{many decimals}}27$

only thing that does make sense is to define a probability distribution function

$$\underbrace{\mathbb{P}(x \in [3, 3+dx])}_{\text{PROBABILITY THAT } x \text{ IS BETWEEN } 3 \text{ \& } (3+dx)} = \underbrace{p(3)}_{\text{probability distribution}} dx \quad \text{scales w/ } dx$$

then you can ask about finite ranges

$$\mathbb{P}(x \in [a,b]) = \int_a^b p(x)\, dx$$

NORMALIZATION:

$$\boxed{\int_{-\infty}^{\infty} p(x)\, dx = 1} \longleftrightarrow \sum_i \mathbb{P}(a_i) = 1$$

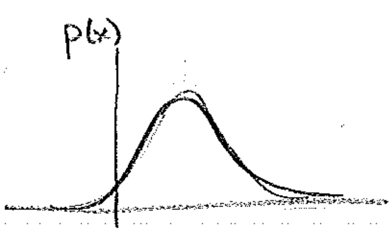
nb. pdf IS A DISTRIBUTION IN THE SAME SENSE THAT THE δ-FUNCTION IS A DISTRIBUTION

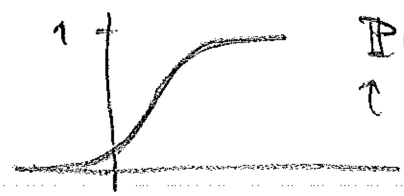
$p(x)$ MAKES AS LITTLE SENSE AS $\delta(x)$

they make sense only when integrated over:

$p(x)\, dx$ & $\delta(x)\, dx$

like 100% probability (COLLAPSED WAVEFUNCTION)





$\mathbb{P}(x) = \int_{-\infty}^{x} p(y)\, dy$

CUMULATIVE DIST. FUNC.

MOMENTS OF A DISTRIBUTION

$\langle x \rangle = \int dx \; x \; p(x)$ remember from moments of a Gaussian?

"expectation value of x"

similarly for $x^2, x^3, \ldots$

s.t. the EXPECTATION VALUE OF A FUNCTION IS

$$\langle f(x) \rangle = \int dx \; f(x) \; p(x)$$

eg can Taylor expand & take each term as a moment

SOME IMPORTANT EXPECTATION VALUES:

$\bar{x} \equiv \langle x \rangle$ is the mean / average

if you took a large # of events from the prob. dist. func. then took the avg, you'd expect this value.

$$\begin{aligned}
Var(x) &= \langle (x-\bar{x})^2 \rangle \\
&= \langle x^2 - 2x\bar{x} + \bar{x}^2 \rangle \\
&= \langle x^2 \rangle - 2\langle x \rangle \bar{x} + \bar{x}^2 \quad \text{just \#s} \\
&= \langle x^2 \rangle - \bar{x}^2
\end{aligned}$$

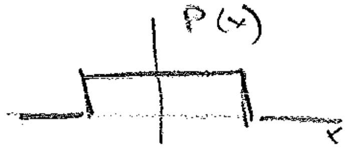
LIKE SQUARE OF STANDARD DEVIATION

but we haven't said anything about data

SOME COMMON DISTRIBUTIONS

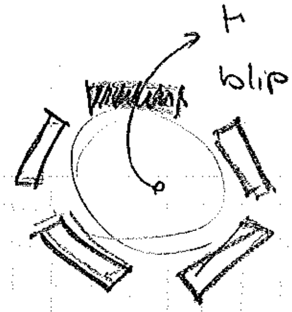
- BOX DISTRIBUTION

$$P(x) = \begin{cases} 1/L & \text{if } -L/2 < x < L/2 \\ 0 & \text{other.} \end{cases}$$



equal probability across a finite range.

eg. I have a detector that is made of a number of discrete elements (like CCD)



blip! → PARTICLE PASSED THROUGH SOMEWHERE ON THIS ELEMENT.

where?

$$\langle x \rangle = 0$$

$$\mathrm{Var}(x) = \langle (x-0)^2 \rangle = \langle x^2 \rangle$$

$$= \int_{-L/2}^{L/2} dx \; x^2 \; \frac{1}{L}$$

$$= \frac{1}{3L} x^3 \Big|_{-L/2}^{L/2}$$

$$= \frac{1}{3L} \cdot \left(\frac{L^3}{8} + \frac{L^3}{8} \right) = \boxed{\frac{L^2}{12}}$$

- (DISCRETE) BINOMIAL DISTRIBUTION : given a binary event / trial (like a coin toss) prob of a given # of positive trials

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

PROBABILITY OF K POSITIVE RESULTS

PROB. OF POSITIVE RESULT FOR EA TRIAL

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

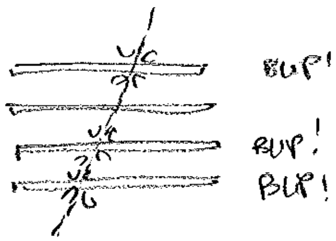
$\langle k \rangle = np$ $\quad \mathrm{Var}(k) = np(1-p)$ ← HW? you check

BARLOW §3.2

eg. DETECTOR EFFICIENCIES

SUPPOSE 'DETECTOR' IS COMPOSED OF SOME # OF REDUNDANT LAYERS, EACH w/ EFFICIENCY $p = 0.95$

↑ PROBABILITY OF BLIP GIVEN ACTUAL EVENTS



SUPPOSE 3 BLIPS ARE REQUIRED TO DEFINE A TRAJECTORY ("TRACK")

HOW MANY LAYERS SHOULD WE BUILD THE DETECTOR OUT OF?

if only 3:

$$P_{\substack{n=3\\k=3}} = \binom{3}{3}(.95)^3(0.05)^0 \approx 85.7\%$$

if 4 layers:

$$P_{\substack{n=4\\k=3}} + P_{\substack{n=4\\k=4}} = \binom{4}{3}(.95)^3(0.05)^1 + \binom{4}{4}(.95)^4(0.05)^0$$

$$= 0.171 + 0.815$$

$$= \boxed{98.6\%}$$

big improvement

<u>POISSON DISTRIBUTION</u> : discrete

Like binomial: still asking about probability of a given # of discrete events

<u>BUT</u>: no idea the # of trials.

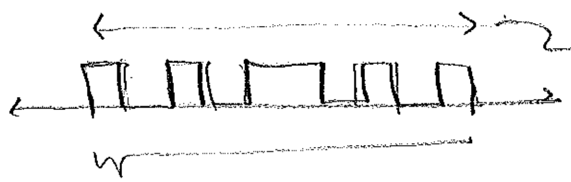
instead: sharp events occurring in a continuum

eg: THUNDERSTORM: definite # of flashes ... no sense to ask how many times there was no flash.

eg. IF YOU EXPECT TO SEE AN <u>AVERAGE</u> OF 10 EVENTS / day BUT SEE 13, IS THAT SIGNIFICANT?

IF WE EXPECT λ EVENTS IN <u>SOME INTERVAL</u>

SPLIT THIS INTO n SMALL INTERVALS



n BINS, n LARGE ENOUGH THAT CHANCE OF 2 IN ONE BIN IS IMPROBABLE

interval : <u>expect</u> 10

THEN THIS IS A BINARY DISTRIBUTION

$$P(k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \boxed{\left(1-\frac{\lambda}{n}\right)^{n-k}} \rightarrow e^{-\lambda} \text{ AS } n \to \infty$$

$p = \lambda/n$

$$\frac{1}{k!} \boxed{\frac{n!}{(n-k)!}} \rightarrow n^k \quad \text{for } n \to \infty$$

$$\boxed{P(k) = \frac{1}{k!} e^{-\lambda} \lambda^k}$$

PROBABILITY OF FINDING k EVENTS IF THE MEAN EXPECTATION IS λ

$\langle k \rangle = \lambda$ $\qquad$ $Var(k) = \lambda$

A SPECIAL DISTRIBUTION : the GAUSSIAN / NORMAL DIST

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

$Var(x) = \sigma^2$ — this is the "σ" when we say "5σ"

$\langle x \rangle = \mu$

why: we saw this in the last lecture

↳ PARTITION FUNCTIONS ARE GAUSSIAN TO LEADING ORDER

- EASY TO INTEGRATE & TAKE MOMENTS
- CENTRAL LIMIT THEOREM
  - eg: for large λ, POISSON → GAUSSIAN

if we make many measurements of something — no matter what the distribution that you draw from —

the measurements will be distributed according to a Gaussian.

(distribution of averages)

MOTIVATION: LET x_i be random variables w/ PDF $p_i(x_i)$
ASSUME $\langle x_i \rangle = 0$ (shift variables as necessary)

LET $X = \sum_{i=1} x_i$; HAS PDF

$$p(X) = \int dx_1 \cdots dx_N \prod_i p_i(x_i) \; \delta(X - \sum_i^N x_i)$$

$$p_i(x_i) = \int dk_i e^{-ik_i y_i} \tilde{p}_i(k_i) \quad \text{(FOURIER)}$$

$$\boxed{\tilde{p}_i(k_i) = \int dx_i e^{ik_i x_i} p(x_i)}$$

$\tilde{p}_i(0) = \langle 1 \rangle = 1$

MOMENTS

$$\left(-i \frac{d}{dk_i}\right)^n \tilde{p}_i(k)\Big|_0 = \langle y_i^n \rangle$$

WITH THESE MOMENTS

$$\tilde{p}_i(k_i) = e^{a_i + ib_ik - \frac{1}{2}c_ik^2 + \cdots} \qquad c_i = \langle x_i^2 \rangle$$

$a_i = b_i = 0$
from $\langle x_i \rangle = 0$
$\langle 1 \rangle = 0$

$$\tilde{p}(k) = \prod_i \tilde{p}_i(k) = e^{-\frac{1}{2}Ck^2 + iDk^3 + \cdots}$$

$P(X) = \int dk\, e^{-ikX}\tilde{p}(k)$ $\qquad C = \sum_i c_i = \sum_i \langle x_i^2 \rangle$

$D = \sum_i d_i$

So: $C, D, \ldots$ ARE $O(N)$ ← large N "dimensional analysis"

s.t. when $k \gtrsim O(1/\sqrt{N})$, $\tilde{p}(k)$ becomes small

BUT @ THAT SCALE OF k, $Dk^3 \sim 1/\sqrt{N}$ SMALLER THAN QUADRATIC TERM

$$\Rightarrow \boxed{P(X) = \frac{1}{\sqrt{2\pi\langle X^2 \rangle}}\, e^{-\frac{1}{2\langle X^2 \rangle}X^2}}$$

Gaussian!