ANNOUNCEMENTS

- iEVAL on Piazza
- LATE HW & SOLUTIONS → DON'T WAIT TO VERY LAST MINUTE

this lec from Peskin SLAC STATS WEEK 2011

THE CENTRAL LIMIT THEOREM

in the limit of a large # of MEASUREMENTS, the distribution of a measured parameter is GAUSSIAN

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

μ: true val.

σ^2: Variance = (standard dev.)^2

standard dev.: measure of significance

SKETCH PROOF recall from LEC 17 (DISPERSION REL) the convolution theorem

$$\text{if } f(X) \equiv \int dx_1\, f_1(x_1)\, f_2(X-x_1) = \int dx_1 dx_2 f(x_1) f(x_2)\, \delta(X-x_1-x_2)$$

ARGUMENTS SUM TO X

- then the FOURIER TRANSFORMS ARE RELATED BY

$$\tilde{f}(k) = \tilde{f}_1(k)\, \tilde{f}_2(k)$$

SAME MOMENTUM VAR

π not for class: TRICK IS TO INSERT $1 = \int \delta(\cdots)\, dy$ everywhere

$$f(X) = \int dx_1 \int dk\, e^{-ikx_1} \tilde{f}_1(k) \int dk'\, e^{-ik'(X-x_1)} \tilde{f}_2(k')$$

do THIS INTEGRAL

$$= \int dk\, dk'\, e^{-ik'X} \tilde{f}_1(k) \tilde{f}_2(k') \boxed{\int e^{-i(k-k')x_1}\, dx_1}_{\ \delta(k-k')\,2\pi}$$

$$= \int dk\, e^{-ikX} \tilde{f}_1(k) \tilde{f}_2(k)$$ ✓

$\Box$

THIS RESULT GENERALIZES ("it is obvious")

$$f(X) \equiv \int dx_1 \cdots dx_N \, f_1(x_1) f_2(x_2) \cdots f_N(x_N) \, \delta\left(X - \sum_i^N x_i\right)$$

DIFFERENT X'S SUBJECT TO: SUM IS FIXED

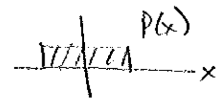
$$\Rightarrow \tilde{f}(k) = \prod_{i=1}^{N} \tilde{f}_i(k)$$

SAME k

SO LET'S PROVE IT. ASSUME SOME NICE pdf $p(x)\,dx$, NOT NEC. GAUSSIAN.

finite $\langle x^n \rangle$

eg



CHOOSE VARIABLES SUCH THAT $\langle x \rangle = 0$.
ie if $\langle x \rangle \neq 0$, then USE $y \equiv x - \langle x \rangle$; SHAPE OF pdf UNCHANGED.

NOW DRAW N numbers from this distribution.

$x_1, \ldots, x_N$ WITH SUM $X \equiv \sum_i^N x_i$

WANT: SHOW THAT $p(X)$ IS GAUSSIAN

from which you may deduce that $p(x_i)$ IS ALSO GAUSSIAN.

$$\mathbb{P}(X) = \int dx_1 \cdots dx_N \left| \prod_i^N p_i(x_i) \right| \delta\left(X - \sum_i^N x_i\right)$$

to dist. from p

PRODUCT OF INDIVIDUAL PROBABILITIES OF EACH MEASUREMENT

ALLOWING ANY VALUES OF THE $\{x_i\}$ SUBJECT TO SUMMING TO X.

this is just an N-dim. convolution
USE CONVOLUTION THM

$$p(x_i) = \int dk \, e^{-ikx_i} \tilde{p}(k)$$

$$\tilde{p}(k) = \int dx \, e^{ikx} p(x)$$

REMEMBER <u>MOMENTS</u> OF A PDF? (EXPECTATION VALS)

$\langle x^n \rangle = \int dx\, x^n\, p(x)$ cf. $\tilde{p}(k) = \int dx\, e^{ikx} p(x)$

<u>OBSERVE</u>: $\tilde{p}(0) = \langle 1 \rangle = 1$

that was boring. but we can write any exp. val:

$$\langle x \rangle = \left(-i\frac{d}{dk}\right)\tilde{p}(k)\Big|_{k=0} \leftarrow \langle x \rangle = 0 \text{ by assump.}$$

$$\langle x^n \rangle = \left(-i\frac{d}{dk}\right)^n \tilde{p}(k)\Big|_{k=0}$$

but these are just n^{th} derivatives of a function $\tilde{p}(k)$ (TAYLOR EXPANSION?!)

without loss of generality (given $\tilde{p}(0) = 1$)

instead of $\tilde{p}(k) = A + Bk + Ck^2 + \ldots$

write this as

$$\tilde{p}(k) = e^{a + ib_1 k - \frac{1}{2}c_1 k^2 + \ldots}$$

<u>then</u>: $\tilde{p}(0) = e^a \rightarrow \boxed{a = 1}$

$$-i\frac{d}{dk}\tilde{p}(k)\Big|_0 = \boxed{b_1 = \langle x \rangle = 0}$$

$$-\frac{d^2}{dk^2}\tilde{p}(k)\Big|_0 = \boxed{c_1 = \langle x^2 \rangle} \quad \text{etc.}$$

then the fourier transform of $\mathbb{P}(X)$ is

$$\tilde{\mathbb{P}}(k) = \prod_i^N p(k) = e^{-\frac{1}{2}C_1 k^2 - iDk^3 + \ldots}$$

BY CONVOLUTION THEOREM

$a = b = 0$

$C_1 = Nc$

"LARGE N" LIMIT : HOW DO THINGS SCALE W/ N?

$C, D, E, \ldots$ are all sum over N terms

$\Rightarrow$ all $\mathcal{O}(N)$

$$e^{-\frac{1}{2}CK^2 - iDK^3 + \cdots}$$

$\mathcal{O}(N)$

exp. suppress.

so this gets small when K is at least $\mathcal{O}(1/\sqrt{N})$ $\longrightarrow$ at this point,

CK^2 is $\mathcal{O}(1)$

these terms get super tiny $\left\{ \begin{array}{l} DK^3 \text{ is } \mathcal{O}(1/\sqrt{N}) \\ EK^4 \text{ is } \mathcal{O}(1/N) \quad \text{etc.} \end{array} \right.$

* this is where we use $\langle x^n \rangle$ is finite

** also nb that every other term is oscillatory, not exponential

AS $N \to \infty$,

$$\boxed{\tilde{\mathbb{P}}(k) = e^{-\frac{1}{2}CK^2}}$$ $\leftarrow$ Gaussian!

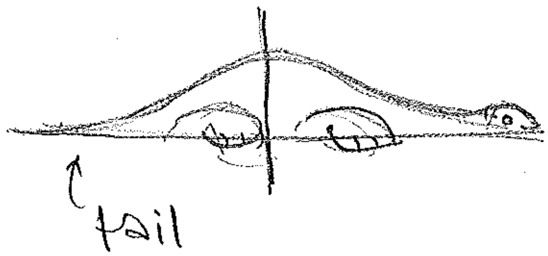
FACT: FOURIER TRANSFORM OF A GAUSSIAN IS ALSO GAUSSIAN

$$\boxed{\mathbb{P}(X) = \frac{1}{\sqrt{2\pi C}} e^{-X^2/2C}}$$ ✓

SO WHAT:

- CUTE PROOF, REMINDED US OF FOURIER-OLOGY
- Reminder of where CLT fails

Convergence to a Gaussian is slowest @ the tails of the distribution



STATISTICAL INFERENCE IS CRUCIALLY TIED TO BEHAVIOR @ TAILS.

This is why we sometimes fall into the trap of misquoting "sigmas"

WE DO EXPERIMENTS, then do a statistical test

↳ wave hands @ central limit theorem like it is some holy incantation,

then quote some value for σ

BUT σ only makes sense if the distribution is ACTUALLY described by a Gaussian

for that value of x
& the finite value of N
& the shape of $p(x)$

CHEAT: SIMULATE EXPERIMENTS ON A COMPUTER (pseudo experiments)

↳ do statistics & make sure it aligns w/ what you ~~expect~~.

(computer time is cheap, data isn't)

HYPOTHESIS TEST

given DATA x ← eg some vector of measurements
& THEORY (hypothesis) H ← eg some set of params.

HOW COMPATIBLE IS H w/ x? QUANTIFY.

$L(H) \equiv P(x|H)$ prob. of obtaining data ASSUMING H is true.

↑ LIKELIHOOD

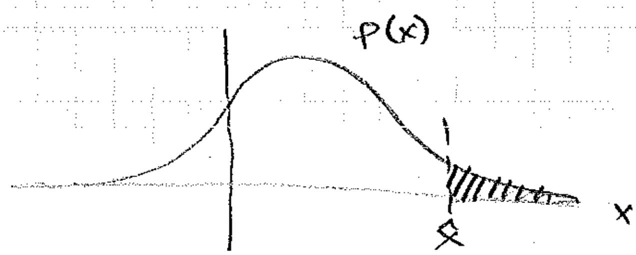
this is easy to calculate. small $L(H) \to H$ is unlikely... but how unlikely?

HARDER: how to use $L(H)$ to say something quantifiable about H.

WRITE: $\hat{x}$ AS EXPERIMENTAL VALUE (DATA)

in contrast to x, which is the quantity being measured

DEFINE A "TAIL MEASURE" $\alpha = \underbrace{\int_{\hat{x}}^{\infty} dx \, P(x|H)}_{\text{UPPER CUMULATIVE DIST.}}$

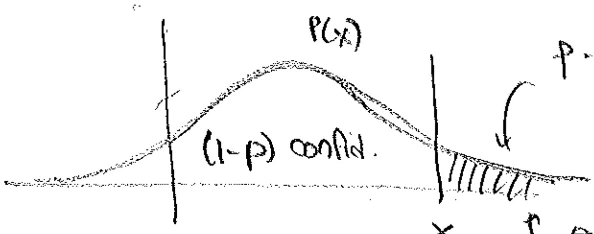


(assume $\hat{x}$ is on upper tail)

if H IS TRUE: α MEASURES PROBABILITY THAT THE OUTCOME OF AN EXPERIMENT IS AT LEAST AS CRAZY AS $\hat{x}$.

if $\alpha = 1\%$, then: H excluded @ 1% confidence.

CAN DEFINE CRITERION FOR EXCLUSION:



NB "p" in "p-value" is not a pdf.

↑ exclude H if $\hat{x} > x_*$

WHEN $p(x)$ IS GAUSSIAN
↑ i.e. "sufficiently close to bump" (vs tail)
& N is large enough that central limit thm. holds

then: can relate p-values to standard deviations

eg.	16%	1σ
	2.3%	2σ
	0.14%	3σ

nb: if you underestimate σ → too small p-value!

---

philosophical question:
what is the probability of H given $\hat{x}$?

FREQUENTIST: impossible to answer.
can give p-value, but this is not a measure of truth of H.

BAYESIAN:

$$P(H|\hat{x}) = \frac{P(\hat{x}|H)\,P(H)}{P(\hat{x})}$$

← what is this?!

← = 1 since we definitely measured $\hat{x}$

P(H): PRIOR PROBABILITY
"probability prior to measurement that H is correct"
↑ SUBJECTIVE AS HELL. ← incomplete knowledge

$P(H|\hat{x})$: POSTERIOR PROBABILITY
↑ euphemism for butt.

SUPPOSE YOU HAVE A <u>FAMILY</u> OF HYPOTHESES

$H(a)$

↑ theory depends on a parameter
eg lifetime of unstable particle

<u>prior</u>: no idea what a is.

$P(H(a))$ or $p(a)$ is constant.

<u>posterior</u>: $p(a) \sim p(\hat{x}\,|\,a)$, makes sense.

NORMALIZING:

$$p(a|\hat{x}) = \frac{P(\hat{x}|a)}{\int da'\, p(\hat{x}|a')}$$

← flat dist. is a const. that cancels.

<u>A PROBLEM</u>: WE ASSUMED a HAS FLAT PRIOR & THAT THIS IS CONSERVATIVE

BUT $b = a^2$ IS A PERFECTLY FINE ALTERNATIVE PARAMETER TO a THAT CONTAINS THE SAME INFO.

BUT if a HAS A FLAT PRIOR, b DOES NOT!!

↑ even choice of param is subjective.

↑

RELATED TO THE MEASURE PROBLEM IN MULTIVERSE THEORIES

PARTICLE THEORIST: want to answer
"What is prob. that $M_h \in [124, 126]$ GeV?"
$\quad\quad\quad\quad\quad\quad\quad\quad\uparrow$ A PARAM
$\Rightarrow$ WANT TO BE BAYESIAN

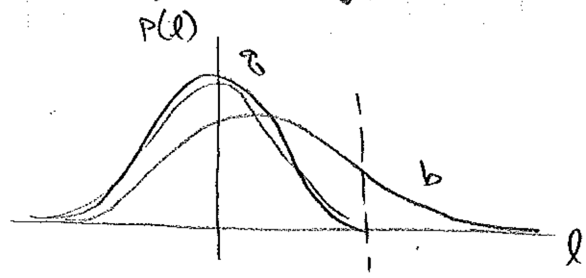
ASTROPHYSICS: so many uncertainties
... impossible to make useful conclusion w/o priors.
$\longrightarrow$ BAYESIAN (typically)

PARTICLE EXPERIMENT:
given $\mathcal{O}(1000)$ collaborators,
will never agree on a prior.
$\longrightarrow$ FREQUENTIST

---

eg. @ LHC, you produce jets from { LIGHT QUARK, q (or glue) ; BOTTOM QUARK, b }

distinguish by: position of 2NDARY vertex, ℓ



$\ell_* \leftarrow$ if $\ell > \ell_*$, PROBABLY b-jet!

EFFICIENCY: given a b-jet, how likely are we to tag it?

$$\varepsilon_b = \int_{\ell_*}^{\infty} d\ell \; p(\ell | b)$$

PURITY: given a tagged b-jet, how likely is it actually a q?

$$\varepsilon_\ell = \int_{\ell_*}^{\infty} d\ell \; p(\ell | q)$$

FREQUENTIST: choose l_+ according to balance of high ε_b, low ε_q.

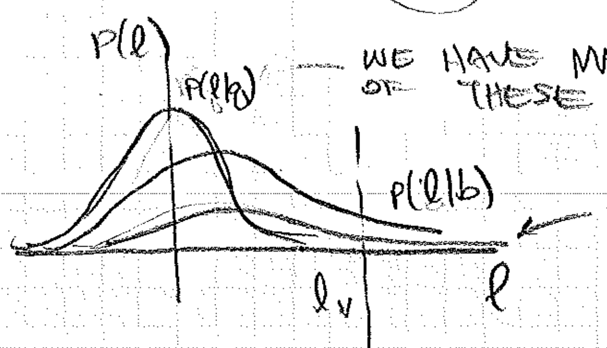
Bayesian: actually, it is much more likely that a given proton collision produces a q jet rather than b.

posterior: $p(q) \gg p(b)$

so RELATIVE PROBABILITIES ARE

$$\frac{P(l|b)\; P(b)}{P(l|q)\; P(q)}$$

because $p(b)/p(q)$ SMALL, WANT TO PICK l_+ LARGE.



$$P(l|b)\,\frac{P(b)}{P(q)}$$

RATIO OF PRIORS

SO THIS LINE SHOULD BE FURTHER TO RIGHT

↳ tails!!

29 NOV

BAYESIAN INFERENCE EXAMPLE

from <u>How Not to be Wrong</u>, Ellenberg p.167

Minority Report + Facebook

IMAGINE FB. CAN DEV. AN ALGORITHM TO SAY:

here's a list of users (SUBSET) ← 100.k (of 200M)

(people on this list are 2x as likely as avg user to be a terrorist

SUPPOSE one of your friends is on this list. ? how do you feel ?

LET'S BE <u>PARANOID</u> & MAKE UP A #: 10K "persons of interest"

	<u>on LIST</u>	<u>not on LIST</u>	
TERRORIST	10 *	9,990	} 10K
not terrorist	99,990	199,890,010	
	100K		

*: on list: 2x AS LIKELY TO BE TERRORIST

AVG USER: $\frac{10k}{200M} = \frac{1}{20k}$ so on list: $1/10k$

even if on list, 99.99% chance innocent.

<u>Better q</u>: IF <u>NOT</u> terrorist, what is chance of being on list?

$$\frac{99{,}900}{199{,}899{,}010} \approx \boxed{0.05\%}$$

$\boxed{1/2k}$

chance of wrong ID!

<u>TWO DIFF Q</u>

1. CHANCE PERSON PUT ON LIST GIVEN <u>NOT</u> TERRORIST (1/2000)

2. CHANCE <u>NOT</u> TERRORIST GIVEN ON LIST (1/10K)