FNAL STATISTICS LECTURE #2

LOUIS LYONS

PARAMETER DETERMINATION

METHOD OF MOMENTS

MAXIMUM LIKELIHOOD

(LEAST SQUARES NEXT TIME)

MOMENTS

e.g. $\frac{dn}{d\cos\theta} = N\left(1 + \frac{b}{a}\cos^2\theta\right)$

$$\overline{\cos^2\theta} = \frac{\frac{1}{3} + \frac{1}{5}\frac{b}{a}}{1 + \frac{1}{3}\frac{b}{a}}$$

$$\Rightarrow \quad \frac{b}{a} = \frac{5\left(3\,\overline{\cos^2\theta} - 1\right)}{\left(3 - 5\,\overline{\cos^2\theta}\right)}$$

Check $\overline{\cos^2\theta} = \frac{1}{3} \Rightarrow \frac{b}{a} = 0$

i.e. $\frac{b}{a}$ from averaging $\cos^2\theta$ over events

Error from error on $\overline{\cos^2\theta}$ by

i) spread of $\cos^2\theta_i$

OR ii) from expected distribution ← Better

[Error on $\frac{b}{a}$ asymmetric]

MOMENTS, contd

1) Very easy. No maximisation

2) No binning needed

3) Extendable to several parameters/variables

4) Constraints among params not readily incorporated

5) Params can be unphysical

e.g. $\overline{\cos^2\theta} = 0 \Rightarrow \frac{b}{a} = -5/3$

$\overline{\cos^2\theta} = 1 \Rightarrow \frac{b}{a} = -5$

$\overline{\cos^2\theta} = 3/5 + \varepsilon \Rightarrow \frac{b}{a} \sim -\infty$

6) Several possible moments

e.g. $\overline{\cos^4\theta}$ $\overline{|\cos\theta|}$

7) No check on "goodness of fit"

Useful $\Rightarrow$ starting values for more sophisticated methods.

MAXIMUM LIKELIHOOD

WHAT IT IS

HOW IT WORKS : RESONANCE

ERROR ESTIMATES :

STATUS OF $\Delta \ln L = -0.5$ RULE

DETAILED EXAMPLE : LIFETIME

LIKELIHOOD & PDF : TRANSFORMATION PROPS. OF $\mathcal{L}$

SEVERAL PARAMETERS

LIKELIHOOD & GOODNESS OF FIT

BINNED & UNBINNED $\mathcal{L}$

EXTENDED MAX LIKE

COMMENTS ON $\mathcal{L}$ METHOD

BIAS FROM INCORRECT $\mathcal{L}$

BAYESIAN SMEARING OF $\mathcal{L}$ OR $\ln \mathcal{L}$

MAXIMUM LIKELIHOOD

$$y = N\left(1 + \frac{b}{a}\cos^2\theta\right)$$

$$y_i = N\left(1 + \frac{b}{a}\cos^2\theta_i\right)$$

~ Probability density of observing θ_i, given b/a

$$\mathcal{L}\left(\frac{b}{a}\right) = \prod y_i$$

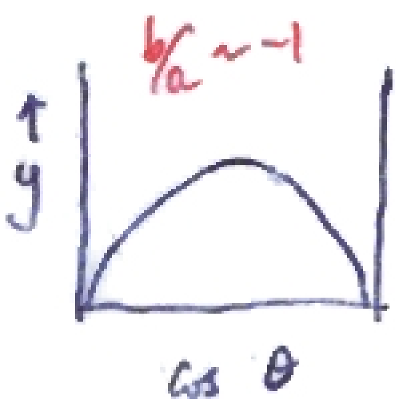
~ Probability of observing given set of θ_i for that b/a

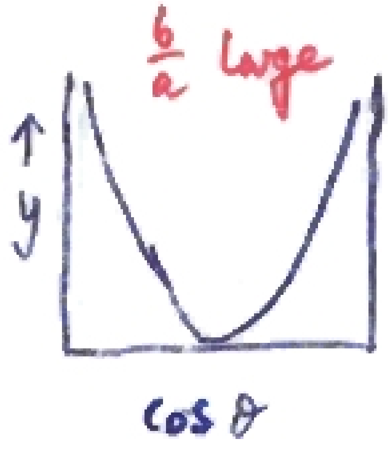
Best estimate of $\frac{b}{a}$ is that which maximises $\mathcal{L}$

Precision of $\frac{b}{a}$ from width of $\mathcal{L}$ distribution

CRUCIAL TO NORMALISE y

SHAPE DETERMINES PARAMS





1) NORMALISATION OF $\mathcal{L}$

$\int P(x|\mu)\,dx$ **MUST** BE INDEPENDENT OF μ

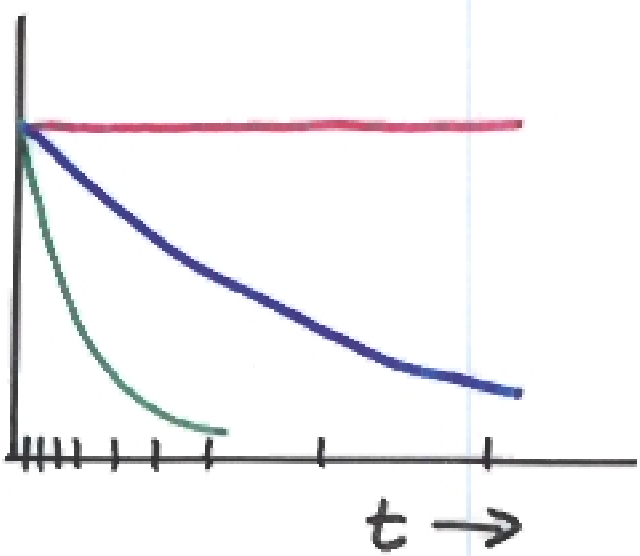
DATA PARAM

e.g. Lifetime fit to $t_1, t_2, \ldots\ldots t_n$

$$[\tau = \Sigma t_i / N]$$

Incorrect $P(t|\tau) = e^{-t/\tau}$

Missing $1/\tau$



— $\tau = \infty$

— τ too big

— Reasonable τ

Ì£8&é^y¯â6?xèâaJbúawâõKf¡üû¢ZÕÛ¶¢d NçÚPõLã= I8ªî1«@¹[QV;]BødJrâÎ°\

1) NORMALISATION OF $\mathcal{L}$

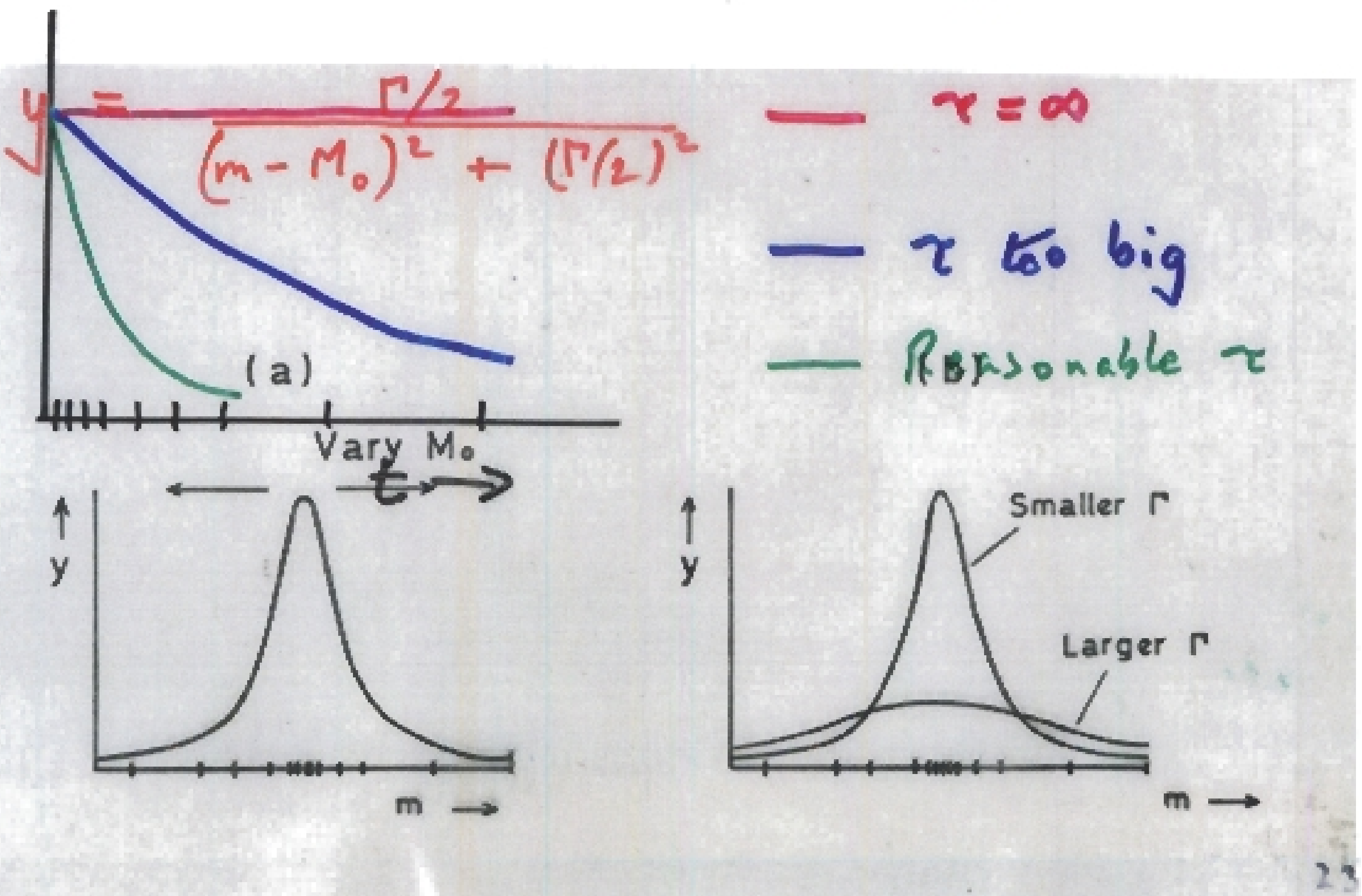
$\int P(x|\mu)\,dx$ **MUST** BE INDEPENDENT OF μ

x ← DATA, μ ← PARAM

e.g. Lifetime fit to $t_1, t_2, \ldots\ t_n$

$$[\tau = \Sigma t_i / N]$$

Incorrect $P(t|\tau) = e^{-t/\tau}$ ← Missing $1/\tau$



Conventional to consider

$$l = \ln(\mathcal{L}) = \sum \ln y_i$$

For large N, $\mathcal{L} \to$ Gaussian

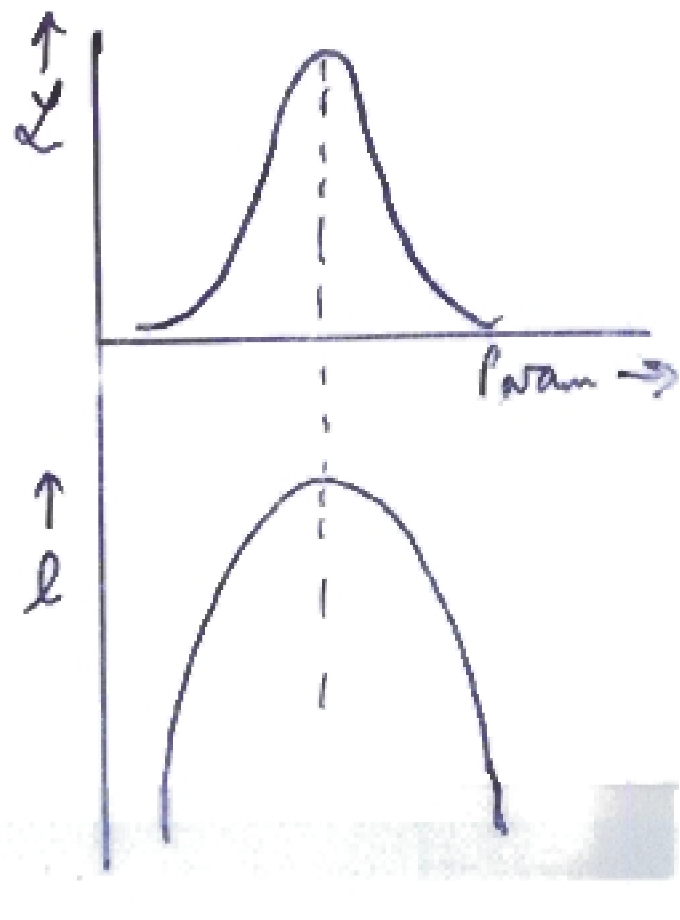
"Proof"

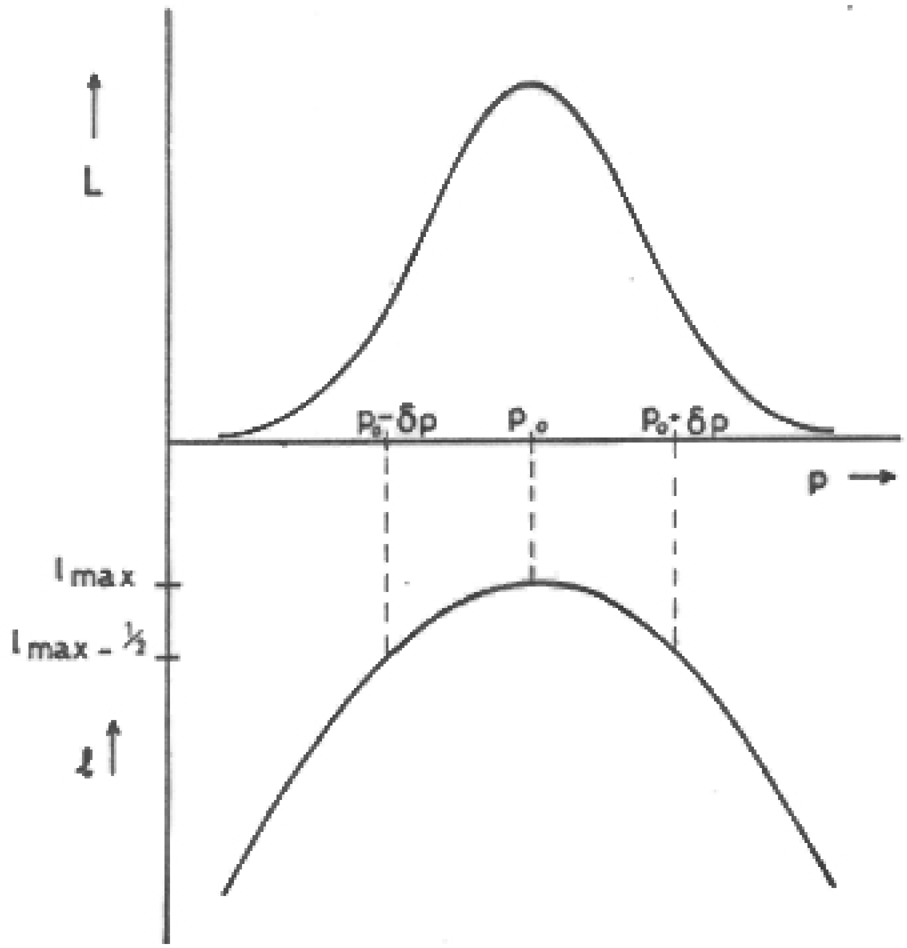
Taylor expand l about its maximum

$$l = l_{max} + \frac{1}{2!} l'' \left[\delta\left(\tfrac{b}{a}\right)\right]^2 + \cdots\cdots$$

$$= l_{max} - \frac{1}{2c}\delta^2 + \cdots \qquad c = -1/l''$$

$$\Rightarrow \mathcal{L} \sim \exp\left(-\frac{\delta^2}{2c}\right)$$




L
$p_0 - \delta p$
p_o
$p_0 + \delta p$
p
l_{max}
$l_{max} - \frac{1}{2}$
l

MAXIMUM LIKELIHOOD ERROR

Range of likely values of p = param from width of $\mathcal{L}$ or ℓ distributions

When $\mathcal{L}$ is Gaussian, following are equiv

1) RMS of $\mathcal{L}$ distribution

2) $\left(-\frac{\partial^2 \ell}{\partial p^2}\right)^{-1/2}$

3) Change in p so that $\ell = \ell_{max} - \frac{1}{2}$

If $\mathcal{L}$ is non-Gaussian, 3) still gives range of p that corresponds to 68% prob (though not usually the shortest)

Error usually asymmetric.

Asymmetric errors are messy, so try to choose parameters intelligently

e.g. λ or τ

b or $1/b$

COVERAGE:

HOW OFTEN DOES QUOTED RANGE FOR PARAM INCLUDE PARAM'S TRUE VALUE

N.B. COVERAGE IS PROPERTY OF <u>METHOD</u>, NOT OF A PARTICULAR LIMIT

COVERAGE CAN VARY WITH μ

— " —

STUDY COVERAGE OF DIFFERENT METHODS OF POISSON PARAMETER μ FOR OBSERVATION OF NUMBER OF EVENTS n

HOPE FOR:

COVERAGE

If true for all μ : "correct coverage"

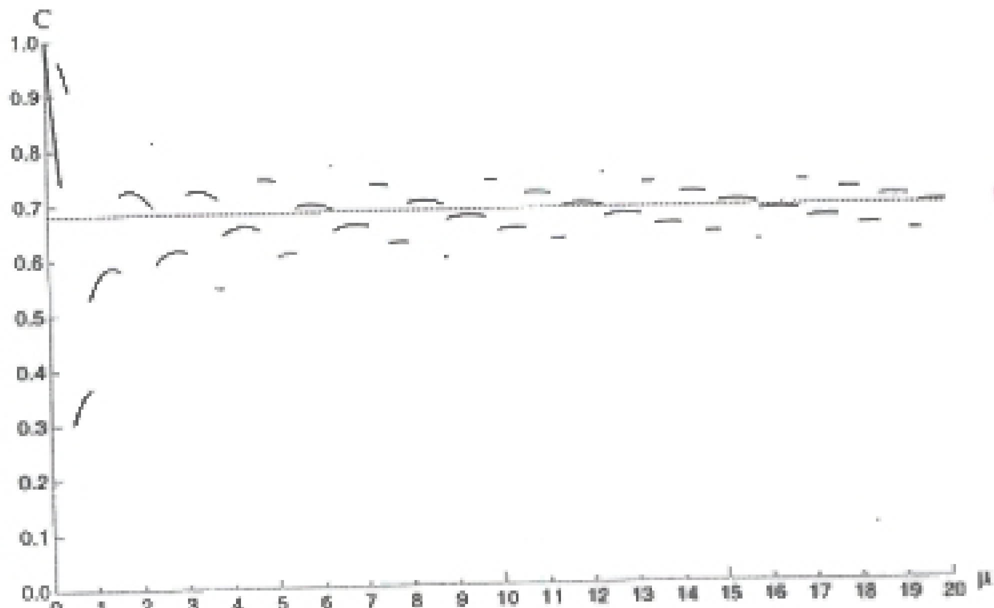
P< α for some μ: "undercoverage"
(this is serious !)

P> α for some μ: "overcoverage"

Conservative

Loss of rejection power

$\mathcal{L}$ approach
<u>NOT</u> frequentist



Coverage (C) vs μ: $-2\ln\lambda < 1$ ($C \to 0.6827$ as $\mu \to \infty$)

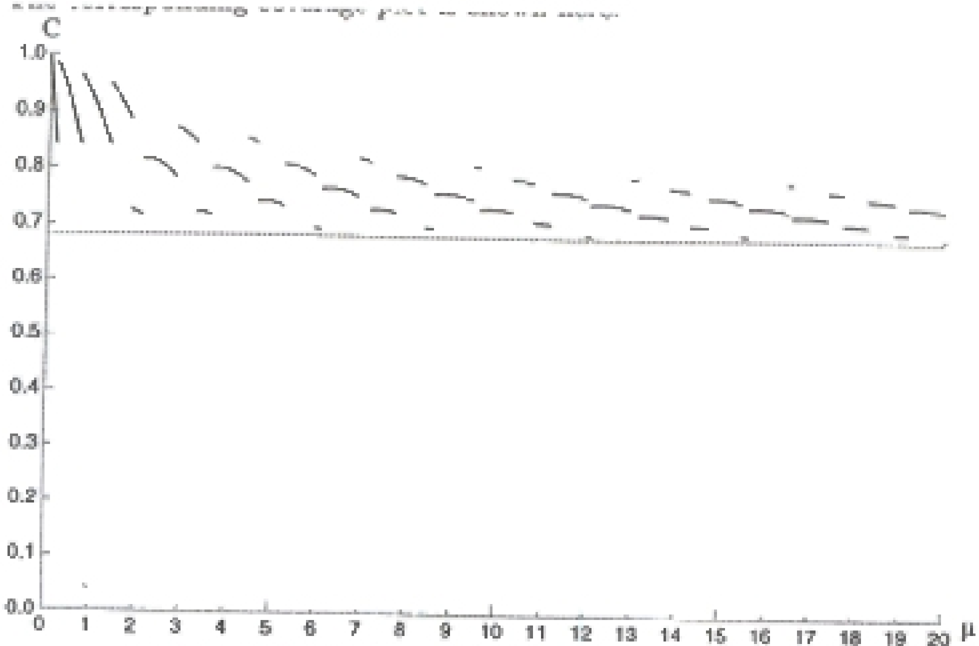
$$P(n,\mu) = e^{-\mu}\mu^{n}/n!$$

$$-2\ln\lambda < 1$$

$$\left[\lambda = P(n,\mu)/P(n,\mu_{best})\right]$$

COVERAGE OF ERROR BARS FOR POISSON DATA

JOEL HEINRICH
CDF 6438



Coverage (C) vs μ: Classical Central Intervals (C → 0.6827 as μ → ∞)

$$P(n,\mu) = e^{-\mu}\mu^n/n!$$

<u>Classical</u> central intervals at 68.3% coverage

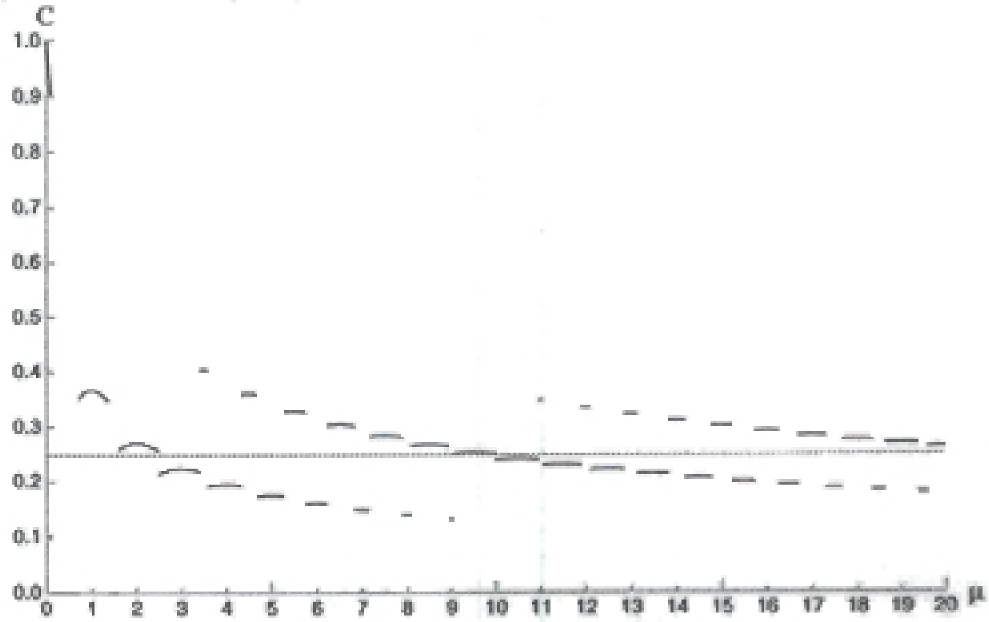
<u>Never</u> undercovers

(conservative at both ends)

COVERAGE OF ERROR BARS FOR POISSON DATA — JOEL HEINRICH

CDF 6438

plot of the coverage $C(\mu)$ for the case $\Delta = 0.1$.



Coverage (C) vs μ: $\chi^2 < 0.1$ ($C \to 0.2482$ as $\mu \to \infty$)

3

$$P(n, \mu) = e^{-\mu} \mu^n / n!$$

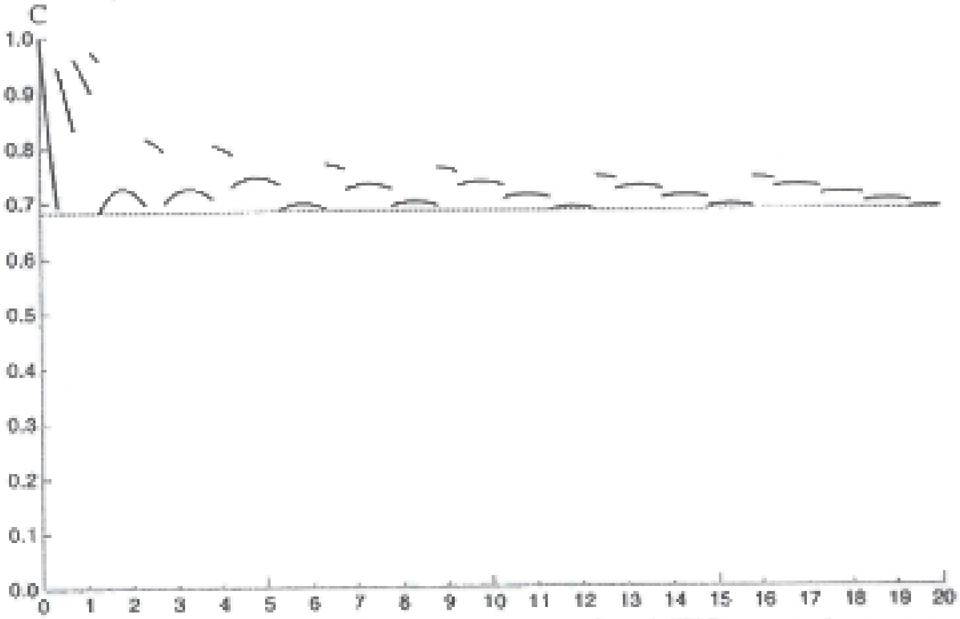
$$\chi^2 = \left(\frac{n-\mu}{\sqrt{\mu}}\right)^2$$

$\Delta\chi^2 = 0.1 \implies 24.8\%$ coverage

χ^2 Approach

NOT FREQUENTIST

Finally, we show the coverage of the 1σ unified intervals:

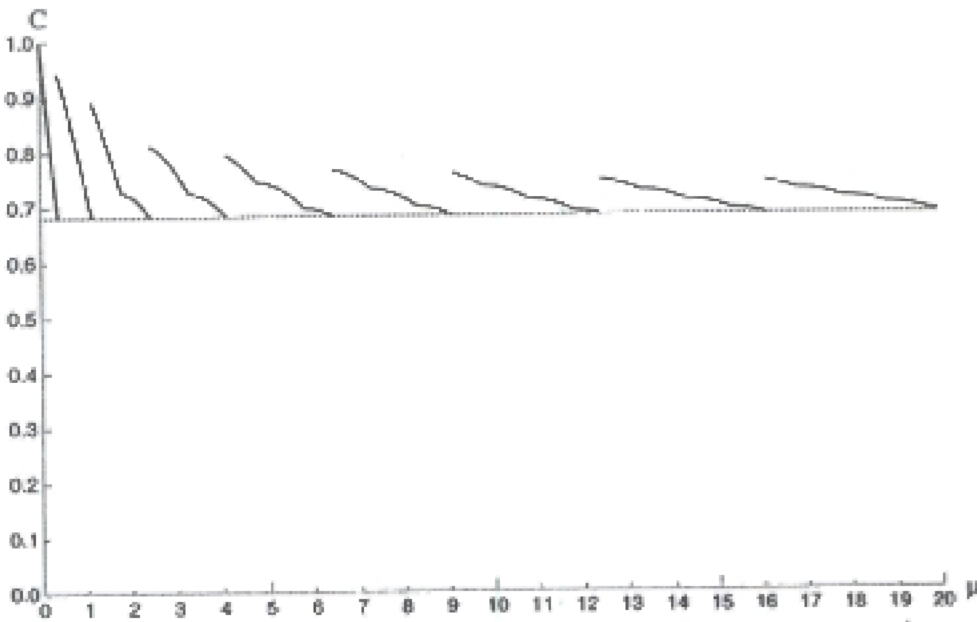


Coverage (C) vs μ: Unified Intervals (C → 0.6827 as μ → ∞)

$$P(n, \mu) = e^{-\mu} \mu^n / n!$$

Unified intervals at 68.3% coverage

Feldman Cousins



Coverage (C) vs μ: Probability Ordering Intervals (C → 0.6827 as μ → ∞)

$$P(n,\mu) = e^{-\mu}\mu^{n}/n!$$

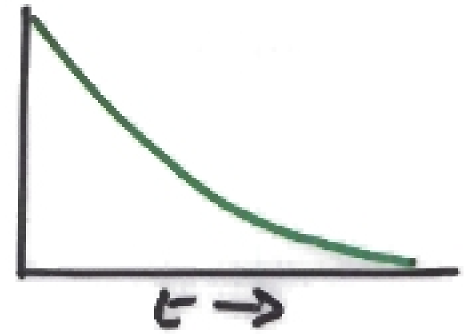
Probability ordering intervals

at 68.3% coverage

LIFETIME DETERMINATION

$$\frac{dn}{dt} = \frac{1}{\tau} e^{-t/\tau}$$

NORMALISATION



Observe $t_1, t_2 \ldots\ldots t_N$

Use pdf to construct

$$\mathcal{L} = \Pi \left(\frac{dn}{dt}\right)_i = \Pi \frac{1}{\tau} e^{-t_i/\tau}$$

$$\therefore \ \ell = \sum_i \left(-t_i/\tau - \ln \tau\right)$$

$$\frac{\partial \ell}{\partial \tau} = \sum \left(+t_i/\tau^2 - \frac{1}{\tau}\right) = 0 = \frac{\Sigma t_i}{\tau^2} - \frac{N}{\tau}$$

$$\Rightarrow \ \tau = \Sigma t_i / N = \overline{t_i}$$

"Obvious"

$$\frac{\partial^2 \ell}{\partial \tau^2} = -\sum \frac{2t_i}{\tau^3} + \sum \frac{1}{\tau^2} = -2\frac{N}{\tau^2} + \frac{N}{\tau^2} = -\frac{N}{\tau^2}$$

$$\Rightarrow \ \sigma_\tau = 1 \Big/ \sqrt{-\frac{\partial^2 \ell}{\partial \tau^2}} = \tau / \sqrt{N}$$

N.B. 1) Usual $1/\sqrt{N}$ behaviour

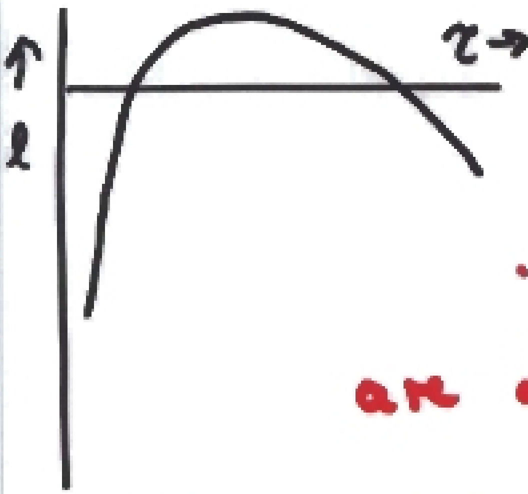
2) $\sigma_\tau \propto \tau_{est}$

BEWARE FOR AVERAGING RESULTS

$$\ln \tau - \ln \tau_{max} = \text{Universal Fn of } \tau/\tau_{max}$$

$$l(\tau) = \sum -t_i/\tau - N \ln \tau$$

$$l(\tau) - l(\tau_{max}) = -N\tau_{max}/\tau - N\ln\tau + N + N\ln\tau_{max}$$



$$= N\left[1 + \ln(\tau_{max}/\tau) - \tau_{max}/\tau\right]$$

$\therefore$ For given N, σ_+ & σ_- are defined $\left(\sim \dfrac{\tau_{max}}{\sqrt{N}} \text{ as } N \to \infty\right)$

For small N, $\sigma_+ > \sigma_-$

— " —

$$l(\tau_{max}) = -N(1 + \ln \bar{t})$$

N.B. $l(\tau_{max})$ depends only on $\bar{t}$, but <u>not</u> on distribution of t_i

Relevant for whether l_{max} is useful for testing goodness of fit

$\mathcal{L}$ AND pdf

EXAMPLE 1 POISSON

pdf = Probability distribution function for observing n, given μ, is

$$P(n;\mu) = e^{-\mu}\mu^n/n!$$

From this, construct $\mathcal{L}$ as

$$\mathcal{L}(\mu;n) = e^{-\mu}\mu^n/n!$$

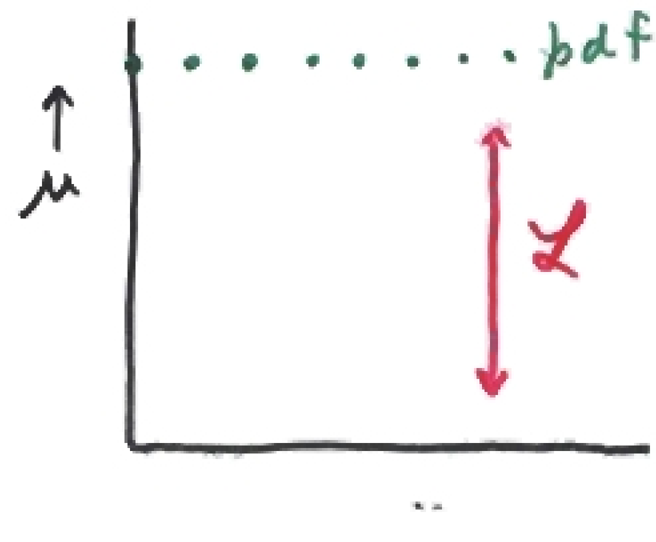
i.e. Use same function of μ + n, but

for pdf, μ is fixed

for $\mathcal{L}$, n is fixed

N.B. $P(n;\mu)$ exists only at integer $n \geq 0$

$\mathcal{L}(\mu;n)$ exists as continuous fn of $\mu \geq$



Example 2 Lifetime distribution

pdf $P(t;\lambda) = \lambda e^{-\lambda t}$

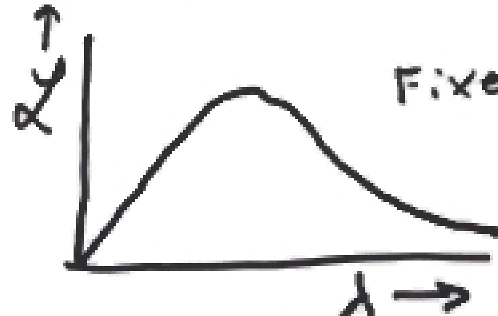
$\therefore \mathcal{L}(\lambda; t) = \lambda e^{-\lambda t}$ [Single observed t]

Now both t & λ are continuous

pdf maximises at $t=0$

$\mathcal{L}$ ... — " — $\lambda = t$

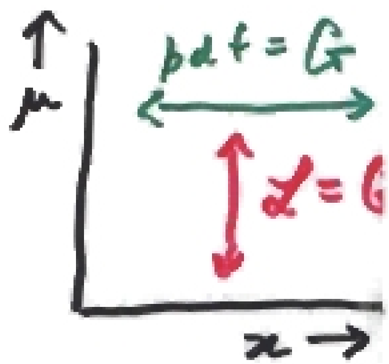
pdf



$\mathcal{L}$

N.B. Functional form of $P(t)$ & $\mathcal{L}(\lambda)$ are different

EXAMPLE 3 GAUSSIAN



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

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

N.B. In this case, <u>same</u> functional form

$\therefore$ If only consider Gaussians, can get confused between pdf & $\mathcal{L}$

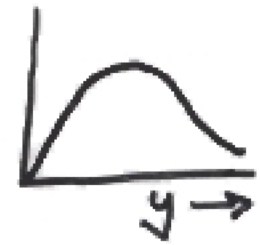
$\therefore$ Examples 1 & 2 are useful

TRANSFORMATION PROPERTIES OF $\mathcal{L}$ AND PROBABILITY DENSITIES

LIFETIME EXAMPLE $\quad \frac{dn}{dt} = \lambda e^{-\lambda t}$

CHANGE <u>OBSERVABLE</u> FROM t to $y = +\sqrt{t}$

$$\frac{dn}{dy} = \frac{dn}{dt}\frac{dt}{dy} = \lambda e^{-\lambda y^2} \cdot 2y$$



so (a) pdf changes, BUT

(b) $\int_{t_0}^{\infty} \frac{dn}{dt}\, dt = \int_{\sqrt{t_0}}^{\infty} \frac{dn}{dy}\, dy$

CONTRAST $\mathcal{L}$, WHICH IS <u>NOT</u> pdf for λ,
GIVEN OBSERVED

When parameter changed from λ to $\tau = 1/\lambda$

(a′) $\mathcal{L}$ DOES NOT CHANGE

$$\frac{dn}{dt} = \frac{1}{\tau} e^{-t/\tau}$$

$$\mathcal{L}(\tau; t) \equiv \mathcal{L}(\lambda = 1/\tau; t)$$

because identical numbers occur in evaluation of the two $\mathcal{L}$'s

BUT (b′) $\int_0^{\lambda_0} \mathcal{L}(\lambda; t)\, d\lambda \neq \int_{\tau_0 = 1/\lambda_0}^{\infty} \mathcal{L}(\tau; t)\, d\tau$

$\therefore$ It is <u>not</u> meaningful to integrate $\mathcal{L}$

	$pdf(t; \lambda)$	$\mathcal{L}(\lambda; t)$
VALUE OF FUNCTION	CHANGES WHEN OBSERVABLE IS TRANSFORMED	INVARIANT WRT TRANSFORMATION OF PARAMETER
INTEGRAL OF FN	INVARIANT W.R.T. TRANSFORMATION OF OBSERVABLE	CHANGES WHEN PARAM IS TRANSFORMED
CONCLUSION	MAX PROB DENSITY NOT VERY SENSIBLE	INTEGRATING $\mathcal{L}$ NOT VERY SENSIBLE

CONCLUSION:

"$\int_{p_L}^{p_U} \mathcal{L}\, dp = \alpha$" NOT RECOGNISED STATISTICAL PROCEDURE.

[METRIC DEPENDENT:–

τ RANGE AGREES WITH τ_{pred}

λ RANGE INCONSISTENT WITH $1/\tau_{pred}$]

BUT

1) COULD REGARD AS "BLACK BOX"

2) MAKE RESPECTABLE BY $\mathcal{L} \Rightarrow$ BAYES POSTERIOR

$$P_{posterior}(\lambda) \propto \mathcal{L}(\lambda) \times \pi(\lambda)$$

+ $\pi(\lambda)$ could be const

SEVERAL PARAMETERS

1 param p

p from $\frac{\partial \ell}{\partial p} = 0$

$\sigma_p^2 = 1 / \left(-\frac{\partial^2 \ell}{\partial p^2} \right)$

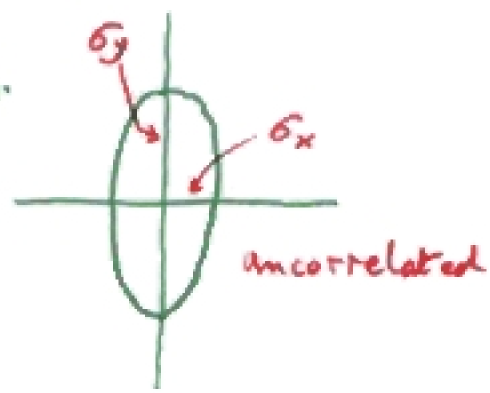
Many dimensions : $\ell(p_1, p_2, p_3 \cdots)$

$p_1 \; p_2 \; p_3 \cdots$ from $\frac{\partial \ell}{\partial p_i} = 0$

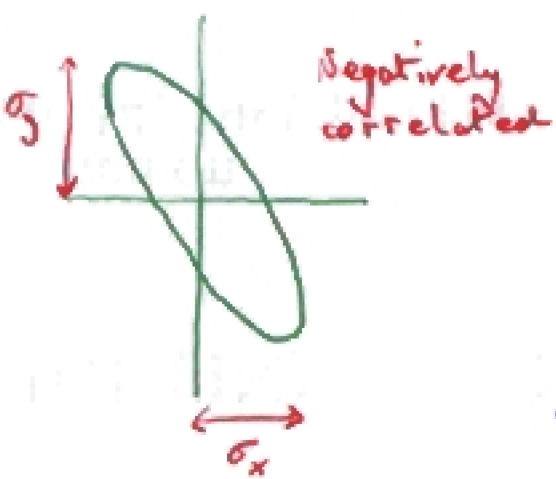
For errors, define $H_{ij} = -\frac{\partial^2 \ell}{\partial p_i \partial p_j}$ = Inverse Error Matrix

Error matrix $E_{ij} = (H^{-1})_{ij}$

e.g.



or



N.B. ERROR **NOT** GIVEN BY $\ell = \ell_{max} - \frac{1}{2}$ WHEN VARYING x FROM BEST VALUE WHILE KEEPING $y \ldots$ CONSTANT

ERROR **IS** GIVEN BY $\ell = \ell_{max} - \frac{1}{2}$ WHEN VARYING x FROM BEST VALUE WHILE . . . . . .

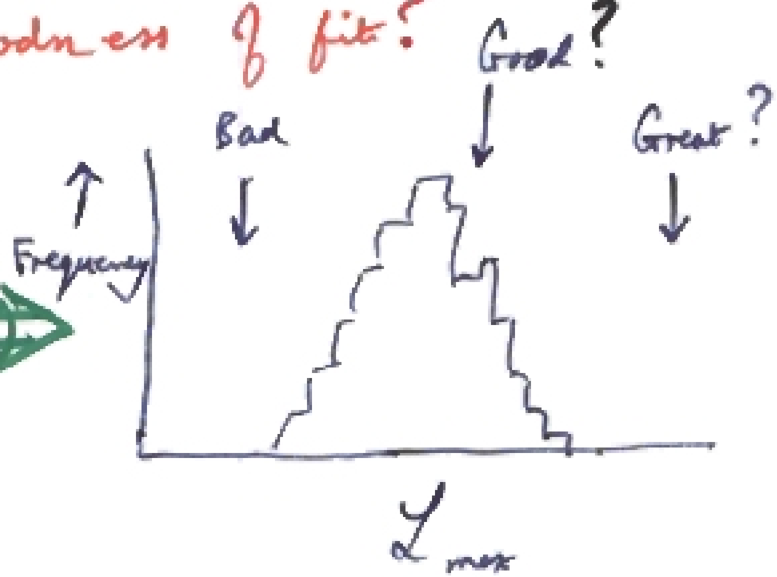
$\mathcal{L}_{max}$ and Goodness of Fit ?

Find parameters by maximising $\mathcal{L}$

$\therefore$ Larger $\mathcal{L}_{max}$ better than smaller $\mathcal{L}_{max}$

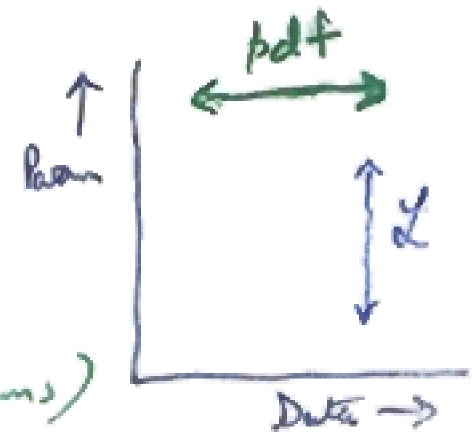
$\therefore$ Use $\mathcal{L}_{max}$ $\Rightarrow$ Goodness of fit?

M.C. distribution of unbinned $\mathcal{L}_{max}$ $\Longrightarrow$



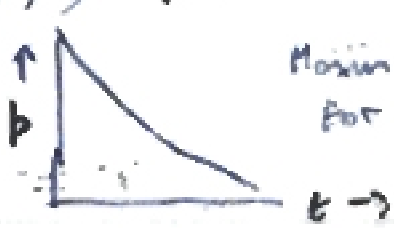
Not necessarily:

$$\mathcal{L}(\underset{\substack{\uparrow \\ \text{FIXED}}}{\text{data}}, \underset{\substack{\uparrow \\ \text{VARY}}}{\text{params}})$$

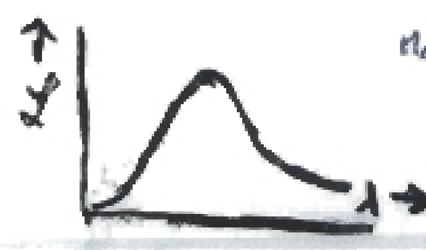


$$\text{Contrast} \quad \text{pdf} = p(\underset{\substack{\uparrow \\ \text{VARY}}}{\text{data}}, \underset{\substack{\uparrow \\ \text{FIXED}}}{\text{params}})$$

e.g. $p(t, \tau) = \lambda e^{-\lambda t}$



Maximises for $t = 0$



Maximises for $\lambda = 1/t$

Examples

(1) Fit exponential to times $t_1, t_2, t_3 \ldots$

[Joel Heinrich, CDF]

$$\mathcal{L} = \prod_i \lambda e^{-\lambda t_i}$$

$$\ln \mathcal{L}_{max} = -N(1 + \ln \bar{t})$$

i.e. Depends only on $\bar{t}$, not on

INDEPENDENT OF DISTRIBUTION OF t

(except for ----)

[$\bar{t}$ is SUFFICIENT STATISTIC]

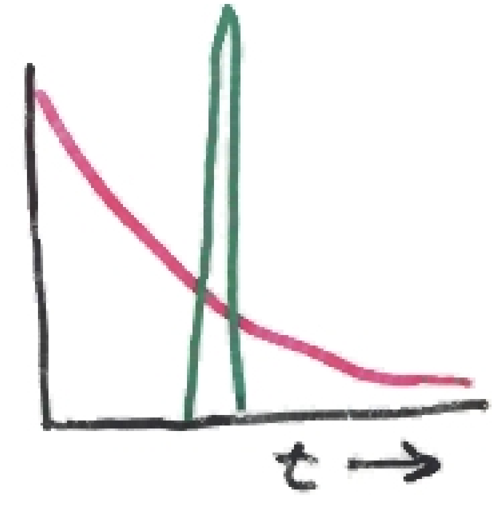
$\therefore$ Variation of $\mathcal{L}_{max}$ in M.C. due to

variations in sample $\bar{t}$

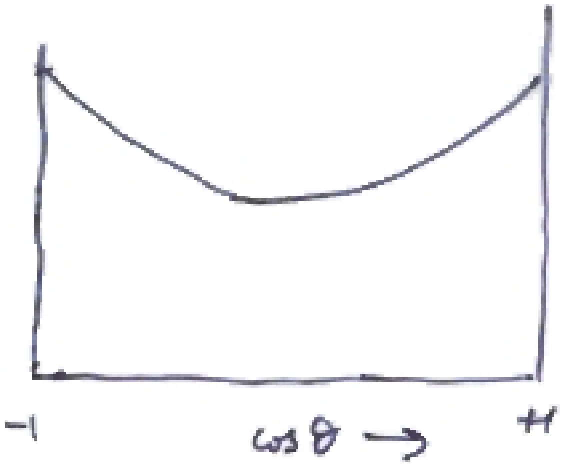
NOT TO BETTER OR WORSE FIT

Same $\bar{t}$

$\Rightarrow$ Same $\mathcal{L}_{max}$ $\longrightarrow$



(2) $\dfrac{dN}{d\cos\theta} = \dfrac{1 + \alpha\cos^2\theta}{1 + \frac{1}{3}\alpha}$



$$\mathcal{L} = \prod_i \frac{1 + \alpha\cos^2\theta_i}{1 + \frac{1}{3}\alpha}$$

DEPENDS ONLY ON $\cos^2\theta_i$

INSENSITIVE TO SIGN OF $\cos\theta_i$

$\therefore$ Data can be in very bad agreement with expected distribution (e.g. all data with $\cos\theta > 0$), + $\mathcal{L}_{max}$ does not know it.

Example of general principle.

3) Fit data to Gaussian, with variable μ
fixed σ

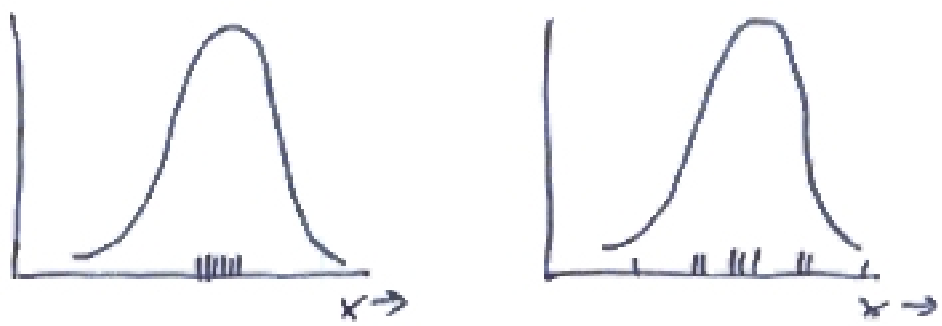
$$\text{pdf} = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right\}$$

$$\ln \mathcal{L}_{max} = \underbrace{N\left(-\frac{1}{2}\ln 2\pi - \ln \sigma\right)}_{\text{const}} - \frac{1}{2}\underbrace{\sum (x_i - \bar{x})^2 / \sigma^2}_{\frac{\text{Variance}(x)}{\sigma^2}}$$

i.e. $\mathcal{L}_{max}$ depends only on Variance(x),

which isn't even relevant for fitting μ $\quad (\hat{\mu} = \bar{x})$

Smaller than expected variance x $\Rightarrow$ larger $\mathcal{L}_{max}$



WORSE FIT
LARGER $\mathcal{L}_{max}$

BETTER FIT
LOWER $\mathcal{L}_{max}$

$\mathcal{L}$ has sensible properties wrt parameters
NOT wrt data

$\mathcal{L}_{max}$ & GOODNESS OF FIT ?

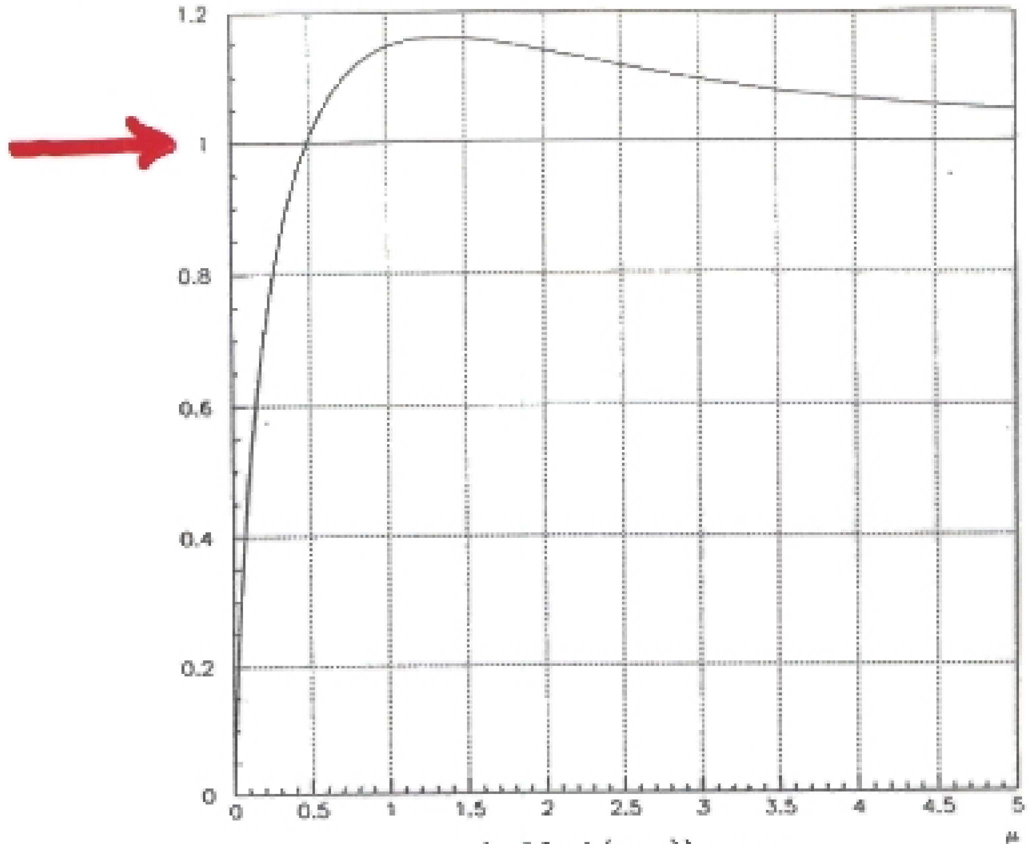
CONCLUSION:

$\mathcal{L}_{max}$ within M.C peak NECESSARY

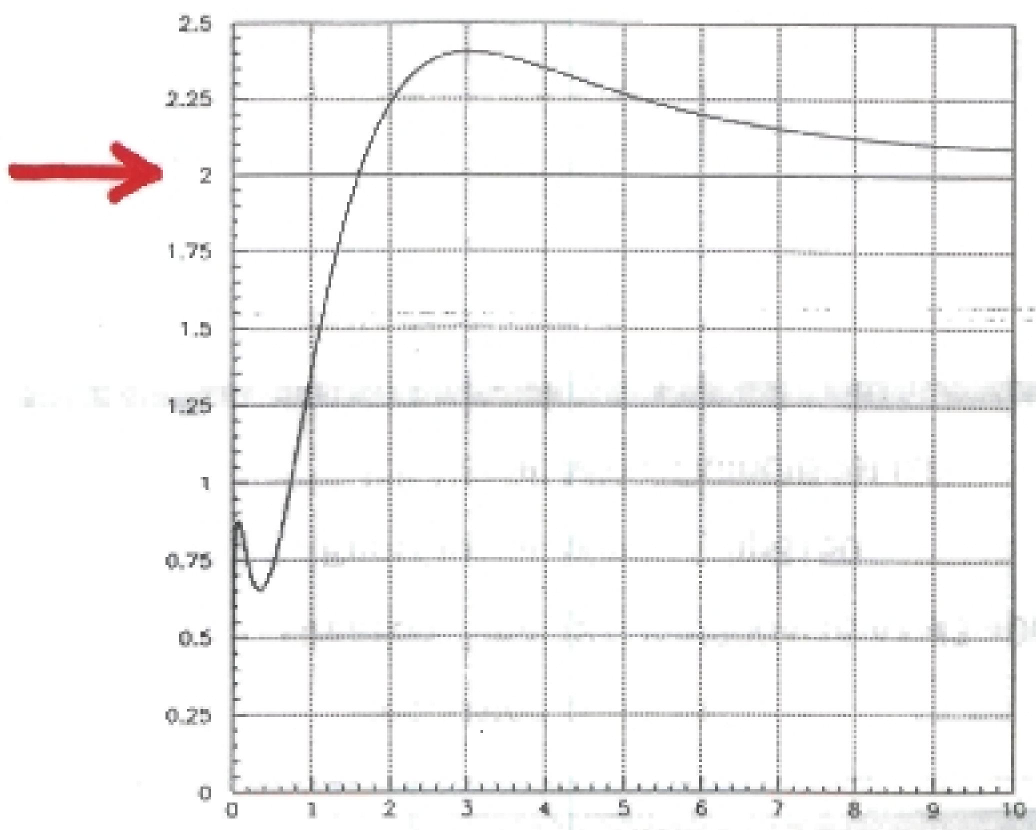
NOT SUFFICIENT

NECESSARY doesn't mean YOU HAVE TO DO IT!

HOW WELL DOES $\ln \mathcal{L}$-ratio approximate χ^2?



$\langle -2 \ln \lambda(\mu; n) \rangle$ vs μ

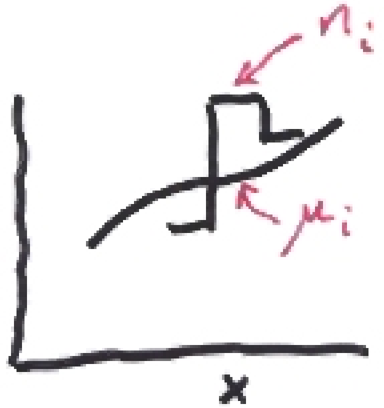


Variance of $-2 \ln \lambda(\mu; n)$ vs μ

JOEL HEINRICH

CDF 5718

BINNED $\mathcal{L}$ + GOODNESS OF FIT



USE $\mathcal{L}$-RATIO

$$\mathcal{L} = \prod_i P_{n_i}(\mu_i)$$

$$\mathcal{L}_{best} = \prod_i P_{n_i}(\mu_{i\,best})$$

$$= \prod_i P_{n_i}(n_i)$$

$$\ln[\mathcal{L}\text{-ratio}] = \ln[\mathcal{L}/\mathcal{L}_{best}]$$

$$\xrightarrow[\text{large } \mu_i]{} -\tfrac{1}{2}\chi^2$$

$\longrightarrow$ Goodness of fit

μ_{best} indep of parameters of fit

$\Rightarrow$ same parameter value from

$\mathcal{L}$ or $\mathcal{L}_{ratio}$

BAKER + COUSINS, NIM A221 (1984) 437

EXTENDED MAXIMUM LIKELIHOOD

Maximum Likelihood uses shape $\Rightarrow$ params

Extended Max Like uses shape + normalisation

i.e. EML uses prob of

1) observing sample size of N events

a 2) given distribution in x.....

$\Rightarrow$ shape parameters & normalisation

Example 1:

Angular distribution

Observe	N	events total	e.g	100
	F	forward		96
	B	backward		4

Rate estimates	ML	EML
Total	—	100 ± 10
Forward	96 ± 2	96 ± 10
Backward	4 ± 2	4 ± 2

ML & EML

Maximum Likelihood uses fixed normalisation

Extended Max Like has normalisation as parameter

e.g.1. Decay of resonance

Use M.L for Branching Ratios

Use EML for Partial Decay Rates

e.g.2 Cosmic ray experiment

See 96 protons & 4 heavy nuclei

M.L estimate 96 ± 2% protons 4 ± 2% heavy

EML estimate 96 ± 10 protons 4 ± 2 heavy

a) Max Like

Prob for fixed N = Binomial

Prob of forwards $$= f^F (1-f)^B \frac{N!}{F!\,B!} \quad \star$$

Maximise $\ln P_a$ wrt f $\Rightarrow \hat{f} = F/N$

Error on $\hat{f}$: $$1/\sigma^2 = -\frac{\partial^2 \ln P_a}{\partial f^2}$$

$$= \frac{N}{\hat{f}(1-\hat{f})} \qquad f = \hat{f}$$

$\Rightarrow$ Estimate of $\hat{F} = Nf = F \pm \sqrt{FB/N}$ ← Completely

- - - - - - . $\hat{B} = N(1-f) = B \pm \sqrt{FB/N}$ ← anti-corr

b) EML

expected overall rate

$$P_b = P_a \times \frac{e^{-\nu}\nu^N}{N!}$$ Poisson for overall rate

Maximise $\ln P_b(\nu, f)$

$\Rightarrow$ $\hat{\nu} = N \pm \sqrt{N}$

$\hat{f} = F/N \pm \sqrt{\frac{f(1-f)}{N}}$ uncorrelated

For $\hat{F}$ & $\hat{B}$, either propagate errors for $\hat{F} = \hat{\nu}\hat{f}$ & $\hat{B} = \hat{\nu}(1-\hat{f})$

or rewrite eqn $\star$ as product of 2 indep Poissons

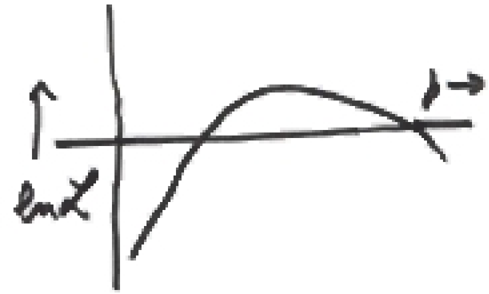
$$\left.\begin{aligned} \hat{F} &= F \pm \sqrt{F} \\ \hat{B} &= B \pm \sqrt{B} \end{aligned}\right\}$$

6) BAYESIAN SMEARING OF $\mathcal{L}$

"USE $\ln \mathcal{L}$ FOR $\hat{p}$ & σ_p

SMEAR IT TO INCORPORATE

SYSTEMATIC UNCERTAINTIES"



SCENARIO:

$$n = \text{Poisson}(\mu = s\epsilon + b)$$

PARAM OF INTEREST → s

ϵ: EFFIC / ACCEPTANCE / $\int \mathcal{L}$

$$\underbrace{\text{EFFIC / ACCEPTANCE / } \textstyle\int \mathcal{L}}_{\text{UNCERTAINTIES MEASURED IN 'SUBSIDIARY' EXPT}}$$

BACKGROUND → b

$$P(s,\epsilon|n) = \frac{\overbrace{P(n|s,\epsilon)}^{\mathcal{L}}\,\pi(s,\epsilon)}{\iint \cdots\cdots\cdots\, ds\, d\epsilon}$$

$$P(s|n) = \int P(s,\epsilon|n)\, d\epsilon$$

$$= \frac{\int \mathcal{L}\, \pi(s)\, \pi(\epsilon)\, d\epsilon}{\iint \cdots\cdots\cdots\, ds\, d\epsilon}$$

e.g. $\pi(s)$ = truncated const. $\quad \pi(\epsilon) \sim e^{-\frac{1}{2}\left(\frac{\epsilon-\epsilon_0}{\sigma}\right)}$

[BEWARE

i.e. SMEAR $\mathcal{L}$ (<u>not</u> $\ln \mathcal{L}$) by "prior" for ϵ

7) GETTING $\mathcal{L}$ WRONG

GIOVANNI PUNZI : PHYSTAT 2003

"COMMENTS ON $\mathcal{L}$ FITS WITH VARIABLE RESOLUTION"

SEPARATE SIGNAL FROM BGD

RESOLUTION VARIES EVENT BY EVENT

$\bar{\sigma}$ DIFFERENT FOR SIGNAL + BGD

e.g. 1) SIGNAL $1 + \cos^2\theta$

BGD ISOTROPIC or COSMIC RAYS

i.e. different parts of detector $\Rightarrow$ different σ

2) M (or τ)

Different number of tracks $\Rightarrow$ different σ_M (or σ_τ)

GIOVANNI'S MONTE CARLO FOR

$A: G(x, 0, \sigma_A)$
$B: G(x, 1, \sigma_B)$
$f_A = 1/3$

σ_A	σ_B	$\mathcal{L}_x$: f_A	$\mathcal{L}_x$: σ_f	$\mathcal{L}_\checkmark$: f_A	$\mathcal{L}_\checkmark$: σ_f
1·0	1·0	0·336 (3)	0·08		
1·0	1·1	0·374 (4)	0·08	0·333 (0)	0
1·0	2·0	0·645 (6)	0·12	0·333 (0)	0
1→2	1·5 – 3	0·514 (7)	0·14	0·335 (2)	0·03
1·0	1→2	0·482 (9)	0·09	0·333 (0)	0

1) $\mathcal{L}_x$ ok for $p(\sigma_A) = p(\sigma_B)$, but otherwise BIASSED

2) $\mathcal{L}_\checkmark$ gives smaller σ_f than $\mathcal{L}_x$

3) $\mathcal{L}_\checkmark$ unbiassed, but $\mathcal{L}_x$ biassed (enormously!)

Events characterised by x_i + σ_i

A events centred on $x=0$ } Gaussians

B events centred on $x=1$ } with σ_i

$$\mathcal{L}_x(f) = \prod_i \left[f\, G(x_i, 0, \sigma_i) + (1-f)\, G(x_i, 1, \sigma_i)\right]$$

Event by event σ_i

$$\mathcal{L}_v(f) = \prod_i \left[f\, p(x_i, \sigma_i | A) + (1-f)\, p(x_i, \sigma_i | B)\right]$$

$$p(S, T) = p(S|T)\, p(T)$$

$$p(x_i, \sigma_i | A) = p(x_i | \sigma_i, A)\, p(\sigma_i | A)$$

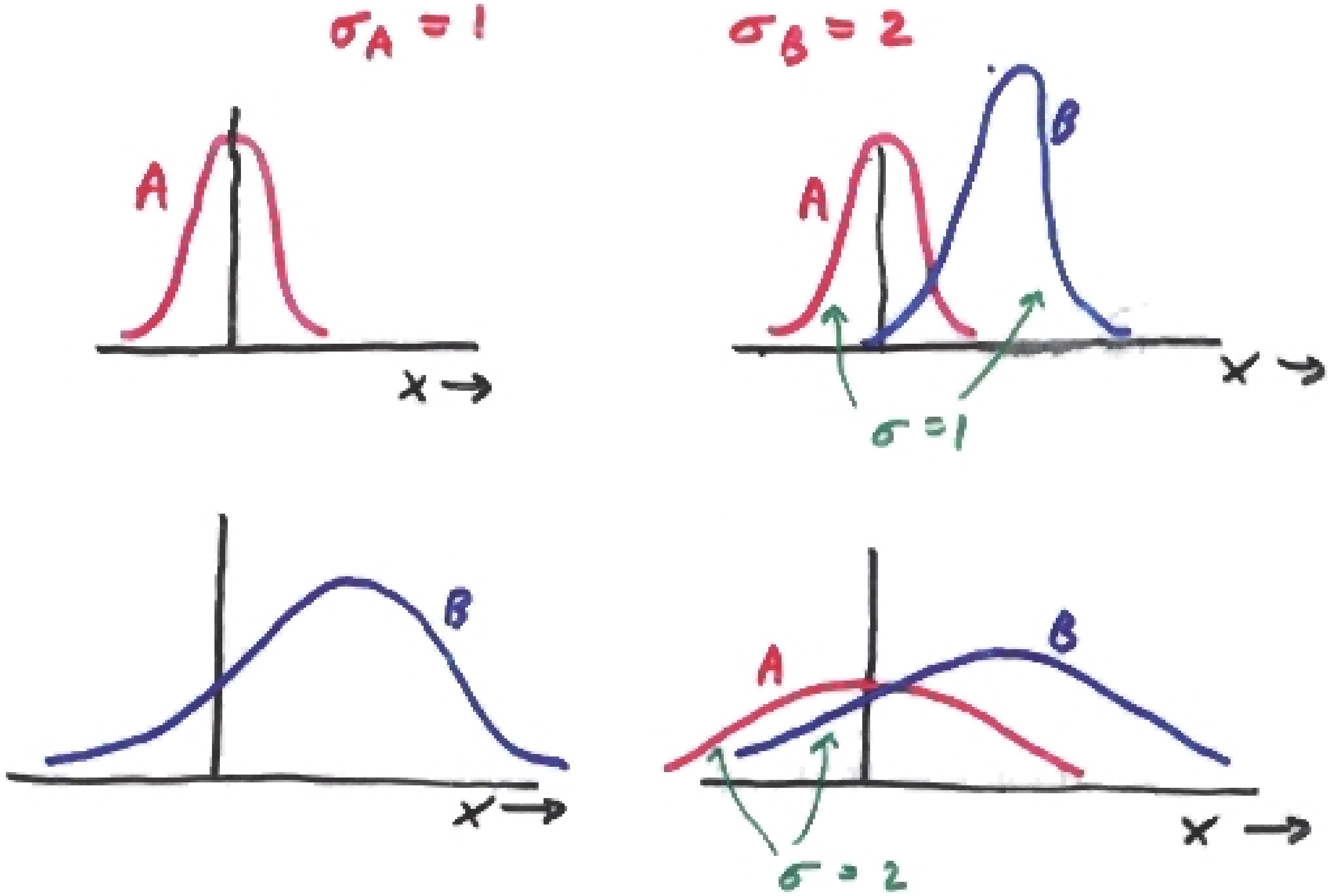
$$= G(x_i, 0, \sigma_i)\, p(\sigma_i | A)$$

$$\therefore \mathcal{L}_v(f) = \prod_i \left[f\, G(x_i, 0, \sigma_i)\, p(\sigma_i | A) + (1-f)\, G(x_i, 1, \sigma_i)\, p(\sigma_i | B)\right]$$

IF $p(\sigma|A) = p(\sigma|B)$, $\mathcal{L}_v \equiv \mathcal{L}_x$

BUT <u>NOT</u> OTHERWISE

EXPLANATION OF BIAS



ACTUAL DISTRIBUTION

FITTING FUNCTION
[N_A/N_B VARIABLE, BUT <u>SAME</u> FOR A + B EVENTS]

FIT GIVES <u>UPWARD</u> BIAS FOR N_A/N_B BECAUSE

a) THAT IS MUCH BETTER FOR A EVENTS, +

b) IT DOES NOT HURT TOO MUCH FOR B EVENTS

SOLUTION

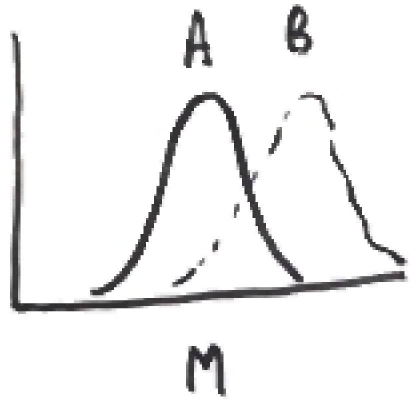
INCLUDE $p(\sigma|A)$ & $p(\sigma|B)$ IN FIT

OR

FIT EACH RANGE OF σ_i SEPARATELY, & ADD $(N_A)_i \rightarrow (N_A)_{tot}$ & $(N_B)_i \rightarrow (N_B)_{tot}$

INCORRECT METHOD USING $\mathcal{L}_x$ USES WEIGHTED AVERAGE OF $(f_A)_i$, ASSUMING INDEPENDENT OF i.

ANOTHER SCENARIO FOR PROBLEM

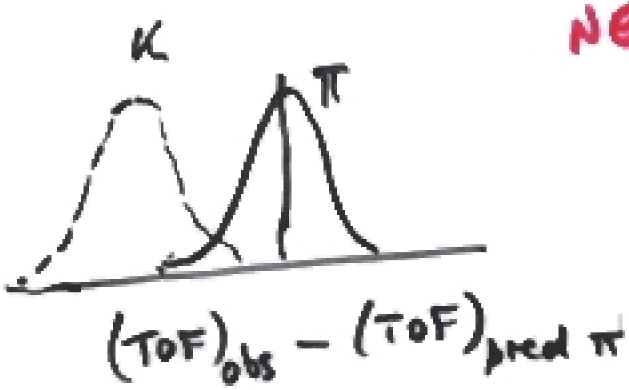


ORIGINALLY:

Positions of peaks = constant

σ_i variable

$(\overline{\sigma_i})_A \neq (\overline{\sigma_i})_B$



NEW SCENARIO : PID

$\sigma_i \sim$ constant

K peak $\rightarrow$ π peak at large momentum

$\overline{p}_K \neq \overline{p}_\pi$

Common feature : $\dfrac{\text{SEPARATION}}{\text{ERROR}} \neq$ CONSTANT

WHERE ELSE ???

MORAL: Beware of event-by-event variables whose pdf do not appear in $\mathcal{L}$

Comments on "Max $\mathcal{L}$" method

1) Uses individual events ⇒ no need to bin

2) Often most efficient method of analysing data

3) Unimportant which variable is used e.g. λ or $\tau = \frac{1}{\lambda}$

ie. maxima & errors correspond

if use $l_{max} - \frac{1}{2}$

4) Worst method computationally

Needs minimisation

Normalisation may need recomputation at each step

5) Limits on parameters
Constraints among params } easy to enforce

but max near boundary can cause trouble

6) Background subtraction difficult

7) Weighted events problematic for errors

8) Hypothesis testing not easy.

~~Best is MC, but not really sufficient.~~