Quasi-optimal weights: a versatile tool of data analysis

Fyodor Tkachov

Institute for Nuclear Research, Russian Academy of Sciences

Best ν mass bound from Troitsk-ν-mass (arXiv:1108.5034).

Had been plagued by "Troitsk anomaly" for ~10 years.

Anomaly went away after a reanalysis that was made possible by **two enabling technologies**:

Part 1 (this talk) statistics: quasi-optimal weights

Part 2 (next talk) software: Oberon technologies

Reminder

Basic problem of math. statistics:

given

a parameterized distribution → $\pi_M(\mathbf{P})$

random sample → $\{\mathbf{P}_i\}_{i=1\ldots N}$ M_0

estimate

parameter and error → M_*, σ_*

- least squares (1800s) ← popular but ill-founded
- moments (K.Pearson, 1894) ← inferior (??) => neglected
- maximal likelihood (R.Fisher, 1912) ← best (??)

The FFRC Bound

- chiropractic (event selection, jet algorithms, etc.)

maximal likelihood (R.Fisher, 1912)

$$\sum_i \ln \pi_M(\mathbf{P}_i) \le \sum_i \ln \pi_{M_{\mathrm{ML}}}(\mathbf{P}_i)$$

$$\sigma^2_{\mathrm{ML}} = \frac{1}{N}\left[\text{Fisher's information}\right]^{-1}$$

absolute minimum

$$\sigma^2_{\mathrm{ML}} \le \sigma^2_*$$

The Bound

Fisher 1925
Frechet 1943
Rao 1945
Cramer 1946

NB in practice $\frac{\partial}{\partial M}\sum_i \ln \pi_M(\mathbf{P}_i) = 0$

NB what if $\pi_M(\mathbf{P})$ is unknown as a formula?

HEP: MC event generators

R.A.Fisher, Proc. Camb. Phil. Soc. 22 (1925) 700-725

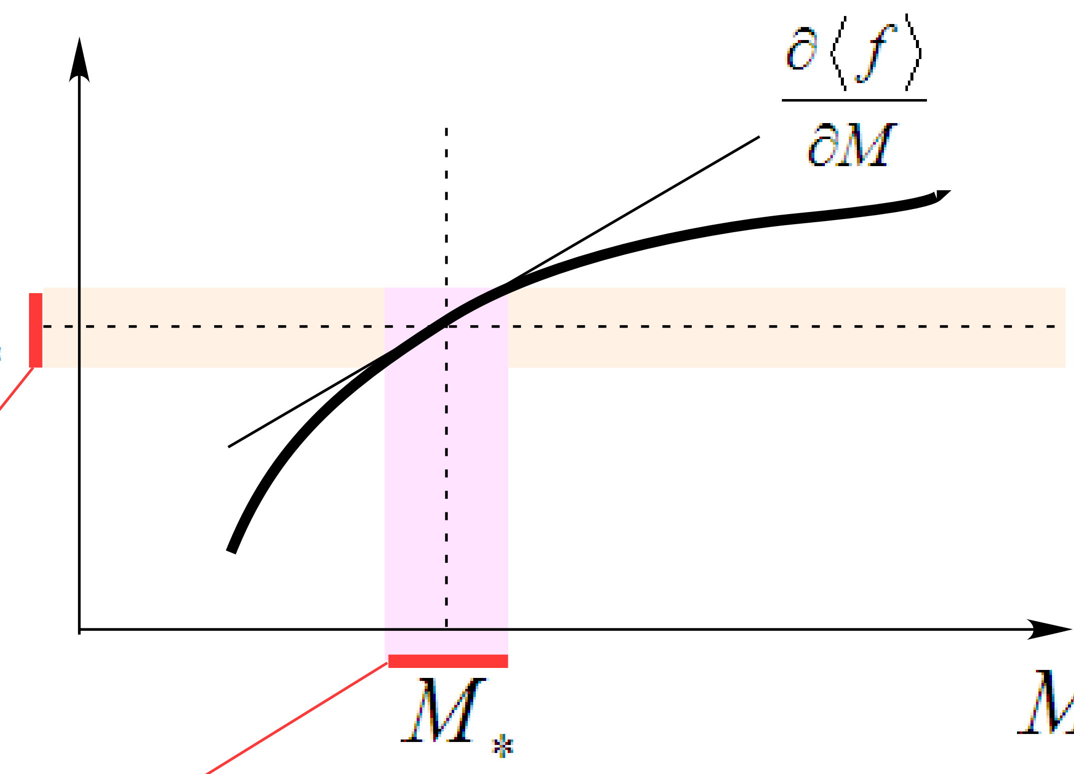
generalized moments

originally (Pearson 1894): $f(\mathbf{P}) = \mathbf{P}^n$

suboptimal compared with ML

$$\langle f \rangle = \int d\mathbf{P}\, f(\mathbf{P})\, \pi_M(\mathbf{P}) = F(M)$$

$$\langle f \rangle_{\exp} = \frac{1}{N} \sum_i f(\mathbf{P}_i) = F_*$$



$$\mathbf{Var}\langle f \rangle_{\exp} = \left\langle \left[f - \langle f \rangle_{\exp} \right]^2 \right\rangle_{\exp}$$

$$\sigma_*^2 = \mathbf{Var}\, M_* \sim \left(\frac{\partial \langle f \rangle}{\partial M} \right)^{-2} \mathbf{Var}\langle f \rangle_{\exp}$$

$$\mathbf{Var}\langle f \rangle_{\exp} \sim \frac{1}{N} \mathbf{Var}\langle f \rangle$$

$$\mathbf{Var}\langle f \rangle = \left\langle \left[f - \langle f \rangle \right]^2 \right\rangle$$

analytical transparency,
complete control
but suboptimal

But we do have explicit expression:

$$\sigma_*^2 \sim \frac{1}{N}\left\{\left(\frac{\partial\langle f\rangle}{\partial M}\right)^{-2}\left\langle\left[f-\langle f\rangle\right]^2\right\rangle\right\}$$

Let us minimize it!

$$\frac{\delta}{\delta f(\mathbf{P})}\{\ldots\}=0$$ whence $$f_{\rm opt}(\mathbf{P})=\frac{\partial}{\partial M}\ln\pi_M(\mathbf{P})$$

and the minimum value is Fisher's information!

could have been discovered long ago

FT, arXiv:physics/0108030

M is unknown   iterative procedure: $f_{{\rm opt},M_i}(\mathbf{P})\to M_{i+1}$

Equivalence to ML from: $\langle f_{\rm opt}\rangle=0$

---

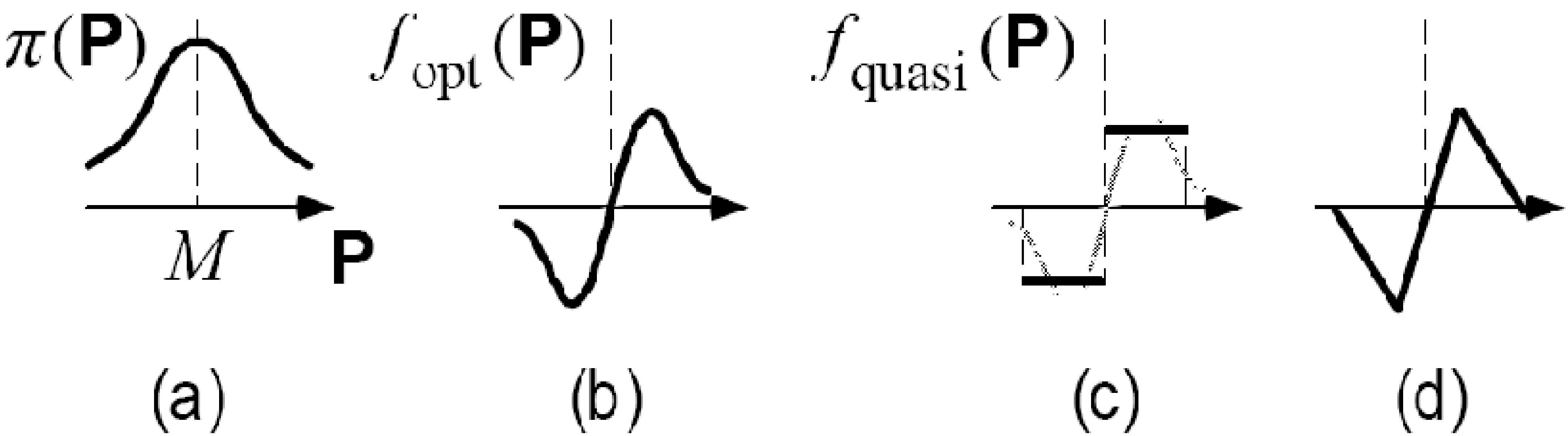
special case for linear dependence on M is known since 1992
--D.Atwood, A.Sony

$$\pi(\mathbf{P})=\pi_0(\mathbf{P})+g\pi_1(\mathbf{P}) \qquad f_{\rm opt}(\mathbf{P})=\frac{\pi_1(\mathbf{P})}{\pi(\mathbf{P})}$$

Example 1

Cauchy distribution, has **no** mean, yet weights work:

$$\pi(\mathbf{P}) \propto \frac{1}{(M-\mathbf{P})^2 + \Gamma^2} \qquad f_{M,\mathrm{opt}}(\mathbf{P}) \sim \frac{(M-\mathbf{P})}{(M-\mathbf{P})^2 + \Gamma^2}$$



Any of these shapes would yield a valid estimate for *M.*

Example 2

Poisson vs Normal

$$\pi_\mu(P) \sim \exp\left[-(P-\mu)^2/2\sigma^2\right], \quad \varphi_{\mathrm{opt},\mu}(P) \sim (P-\mu), \quad \varphi_{\mathrm{opt},\sigma}(P) \sim (P-\mu)^2$$

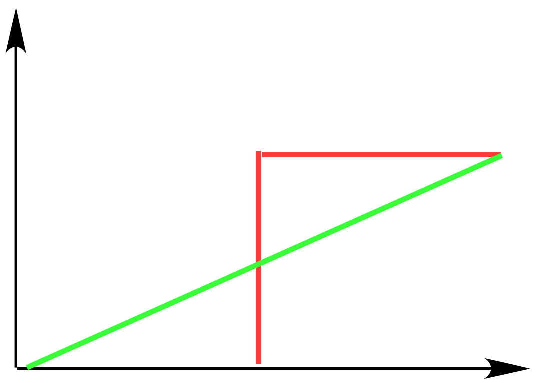
$$\pi_\mu(n) = e^{-\mu}\frac{\mu^n}{n!}, \quad \varphi_{\mathrm{opt},\mu}(n) = \left(\frac{n}{\mu}-1\right)$$

$$\pi_\mu(n) \sim \exp\left[-(n-\mu)^2/2\mu\right], \quad \varphi_{\mathrm{opt},\mu}(n) \sim \left(\frac{n}{\mu}-1\right) + \frac{1}{2}\left(\frac{n}{\mu}-1\right)^2$$

for $\mu \sim 0.15$-1.5, the effect is 10-15%

Example 3

Continuous weight *vs* Cuts



Ratio of variances = 3

Ratio of sigmas = 1.71 ($3\sigma \longrightarrow 5\sigma$)

Continuous weight in this case yields 1.71 times better variance than a cut.

Jorg Pretz: not bounded from above (!?!)

Jorg Pretz 1998, g-2
derived a special case of optimal weights

Example 4

Observation of top quark in all-hadrons channel by D0

P.C.Bhat, H.Prosper and S.S.Snyder,
Top quark physics at the Tevatron, hep-ex/9809011

... To make further progress, D0 performs a multivariate analysis using thirteen variables, described briefly in Table 10. ...
A particularly powerful (and unusual) variable is the average jet count, defined by

$$N^A_{\text{jets}} = \int_{15\text{ GeV}}^{55\text{ GeV}} E_T N(>E_T)\, dE_T \Big/ \int_{15\text{ GeV}}^{55\text{ GeV}} E_T\, dE_T\,,$$

where $N(>E_T)$ is the number of jets with transverse energy greater than E_T. This variable, inspired by the work of Tkachov [90] is interesting in that it assigns a nonintegral "number of jets" to the event. ...

a continuous weight

now, a guidance:

$$f_{\text{opt}}(\mathbf{P}) = \frac{\partial}{\partial M} \ln \pi_M(\mathbf{P})$$

- since the minimum is quadratic,
  an approximation to optimal *f* should be good enough

- therefore works even when no formula for density

data processing in HEP then reduces to finding $f_{\text{q-opt}}$ for specific situations

- great algorithmic flexibility

  e.g. take any MC integrator...

- light in various situations

analytical formula is a great force
e.g. to explore how ragged-ness affects quality

e.g. jet definition...

Non-uniform event sample

FT, arXiv:physics/0604127

$U_i \to \pi_{i,\theta}(P), \quad \{U_i, P_i\}_i$ $\quad U_i$ is a control parameter

If all P_i for different U_i are independent
then the situation reduces to the simple case by defining
a composite event and the corresponding probability density:

$$\otimes_i (P_i) \qquad \prod_i \pi_{i(\theta)}(P)$$

The opt. weight: $$\varphi_{\text{opt}}(\{P_i\}) = \sum_i \frac{\partial}{\partial\theta} \ln \pi_{i(\theta)}(P_i) = \sum_i \varphi_{\text{opt},i}(P_i)$$

Explicit formula for var: $$\textbf{Var}\,\varphi_{\text{opt}} = \sum_i \left[\varphi_{\text{opt},i}(P_i) - \left\langle \varphi_{\text{opt},i} \right\rangle^{\text{th}}_{\theta_{\text{exp}}} \right]^2$$

Straightforward to code.

Notes on usage from Troitsk-nu-mass:

Implementation is straightforward
solution of systems of equations instead of optimum search

MC tests: equivalent to ML

Churns out results where MINUIT collapses ("narrow valleys")

Can easily accomodate information about Poisson etc.
(actually, this was the initial concern)

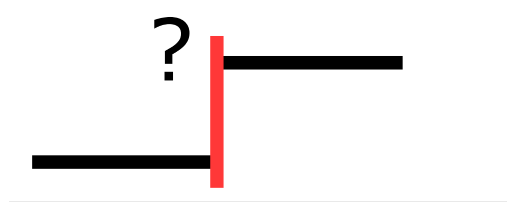
With poor statistics multidimensional systems of equations may not have a solution;
in practice this was not a problem but some research may be called for (same for Student-type correcting factors).

Deep mathematical reasons to consider weights.

~~Sets~~ → Functions and Functionals (+Categories)

---

conventional view of function: $x \to g(x)$

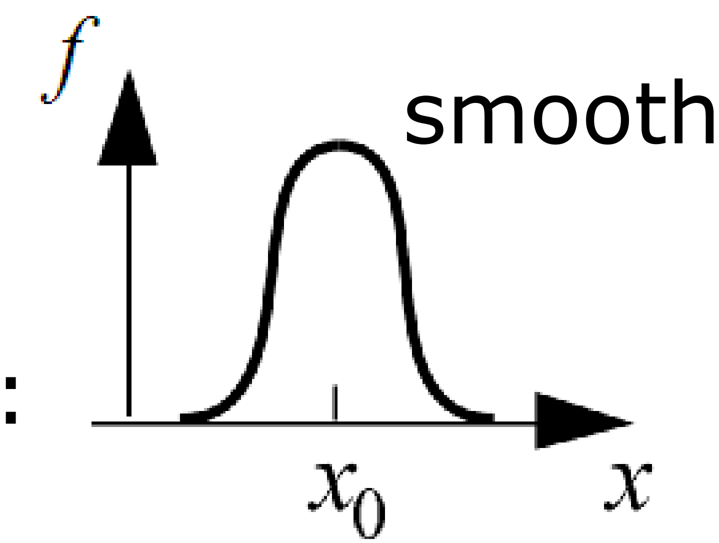


fully meaningful only for continuous $g(x)$

what is value at discontinuity?

more satisfactory description:

$$f \to \langle g\, f \rangle \quad \left[\!\left[= \int \mathrm{d}x\, g(x) f(x) \right]\!\right]$$



where f are the so-called "test functions":

OK for continuous functions
OK++ for discontinuous ("generalized") functions

instead of comparing point values,
one now compares weighted/smeared values

(recall Galerkin method from the geophysics talk by J.-P. Vilotte)

The same conceptual framework:

Optimal Jet Definition
O(N) behavior arXiv:hep-ph/0301185

Efficient multi-dimensional density modelling
from a random sample arXiv:physics/0401012

Quasi-optimal weights
+
Density modelling
+
Optimal Jet Definition
=
A systematic comprehensive framework
for HEP data processing