On fits to correlated and autocorrelated data

Mattia Bruno
in collab. with R. Sommer

based on *Comp. Phys. Comms (2022) 108643, 2209.14188*

SM&FT 2022 - The XIX Workshop on Statistical Mechanics and nonpertubative Field Theory, Bari, Italy, December 20th, 2022

Motivations

(Particle) physics physical information often from fits to data

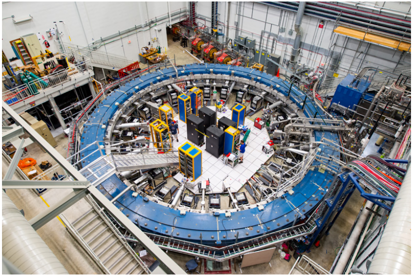
absence of direct signals for new physics → intensity frontier

precision is key word

Experiment

→ obtain data w/ stat.syst. errs

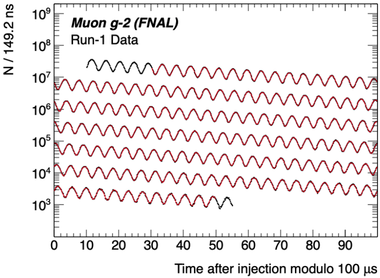
[Muon g-2 @ FermiLab]



Fit to data

driven by theory understanding

[Muon g-2 collab.]



Motivations

(Particle) physics physical information often from fits to data

absence of direct signals for new physics → intensity frontier

precision is key word

HPC

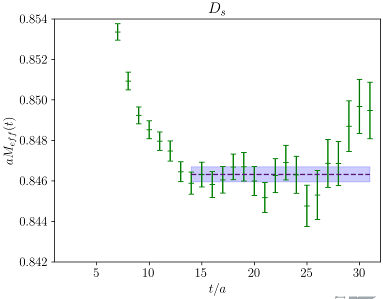
→ obtain data w/ stat.syst. errs

[Leonardo @ CINECA]

Fit to data

driven by theory understanding

[Gambino et al '22]



Lattice field theories

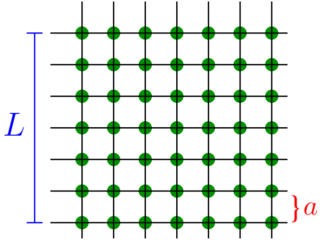
Due to confinement → non-perturbative formulation is necessary

lattice spacing a → regulate UV divergences

finite size L → infrared regulator

Continuum theory $a \to 0$, $L \to \infty$

Euclidean metric → Boltzman interpretation of path integral



$$\langle O \rangle = \mathcal{Z}^{-1} \int [DU] e^{-S[U]} O(U) \approx \frac{1}{N} \sum_{i=1}^{N} O[U_i]$$

Very high dimensional integral → Monte-Carlo methods

Markov Chain of gauge field configs $U_0 \to U_1 \to \cdots \to U_N$

0. $N_{\rm x}$ true expectation values Y_i
1. run a simulation with N configurations
2. measure estimators $y_i(t)$ at Monte Carlo time t

   e.g. a two-point correlator with i labelling source-sink separation
3. calculate averages $\bar{y}_i = \frac{1}{N}\sum_t y_i(t)$

   $\lim_{N\to\infty} \bar{y}_i = Y_i$

Central-limit theorem: $\bar{y}_i$ normally distributed around Y_i, $\delta\bar{y} = \bar{y} - Y$

$$P_C(\bar{y}) = (2\pi)^{-N_{\rm x}/2} (\det C)^{-1/2} \exp\Big(-\frac{1}{2}(\delta\bar{y}, C^{-1}\delta\bar{y})\Big)$$

Cov. matrix $\langle \delta\bar{y}_i \delta\bar{y}_j \rangle \equiv \int \mathrm{d}\bar{y}\ P_C(\bar{y})\ \delta\bar{y}_i\delta\bar{y}_j = C_{ij}$

$C = O(1/N)$ and $\delta\bar{y} = O(N^{-\frac{1}{2}})$

Notation: $(z, y) = z_i y^i$ and $||y||^2 = (y, y)$

χ^2 DISTRIBUTION

Null Hypothesis (NH): Y_i described by $\bar{y}_i$
statistical tests based on $\chi^2 = (\delta\bar{y}, C^{-1}\delta\bar{y}) = ||C^{-1/2}\delta\bar{y}||^2$

1. what is $\langle\chi^2\rangle$? $\rightarrow$ reduced χ^2
   $\langle\chi^2\rangle = N_{\mathrm{x}}$ degrees of freedom
   criterion 1) if $\chi^2 \simeq \langle\chi^2\rangle$ then NH valid (more later..)

2. probability of finding χ^2 larger than observed χ^2_{obs}? $\rightarrow$ p-value
   $Q(\chi^2_{\mathrm{obs}}) = \int \mathrm{d}\bar{y}\ P_C(\bar{y})\ \theta(\chi^2(\bar{y}) - \chi^2_{\mathrm{obs}}) = \gamma(k/2, \chi^2_{\mathrm{obs}}/2)$
   γ incomplete Γ-function
   criterion 2) if $Q(\chi^2_{\mathrm{obs}}) > 0.05$ then NH valid

Exact cancellation of C in P_C and χ^2 $\rightarrow$ simple analytic results

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Problems

In practical calculations we estimate C

1. Lattice QCD (fermions) simulations expensive
   $O(100)$ independent samples

2. Limit $N \to N_{\mathrm{x}}$ cov. matrix singular [Michael '94]

3. Markov chains induce autocorrelations
   consecutive configurations correlated along Monte-Carlo time
   statistical independent information reduced [Madras-Sokal '88]
   error $\sigma^2 = 2\tau_{\mathrm{int}}/N \mathrm{var}$

For fits we need C^{-1}, but hard to estimate in practice
$\to$ uncorrelated fits, or SVD-cuts/regularized C^{-1}

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Model function

Null-hypothesis: Y_j described by model function $\Phi(x_j, A) = \phi_j(A)$

Notation: A_α parameters, $\alpha = 1, \ldots N_{\mathrm{A}}$

Best fit parameters $\bar{a}$ from correlated fits

minimize $\chi^2(a) = ||C^{-\frac{1}{2}}(\bar{y} - \phi(a))||^2 \rightarrow \left.\frac{\partial \chi^2(a)}{\partial a_\alpha}\right|_{a=\bar{a}} = 0$

at minimum $\bar{a}(\bar{y})$ and $\chi^2(\bar{a})$

If C ill-conditioned, $\chi^2(a) = ||W(\bar{y} - \phi(a))||^2$

e.g. $W_{ij} = \delta_{ij}/\sqrt{C_{ii}}$ uncorrelated fits, or SVD cuts...

We want robust statistical tests to judge quality of fit

1. what is $\langle \chi^2(\bar{a}) \rangle$? $\rightarrow$ reduced χ^2
2. how likely finding χ^2 larger than χ^2_{obs}? $\rightarrow$ p-value

UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

$\langle \chi^2(\bar{a}) \rangle$

We evalute $\langle ||W(\bar{y} - \phi(\bar{a}))||^2 \rangle$ [MB, Sommer '22]

notation: $\delta\bar{a} = \bar{a} - A$ and $\phi^\alpha(a) = \partial\phi(a)/\partial a_\alpha$

1. Use minimum condition to define projector $\mathcal{P}$ span $W\phi^\alpha(\bar{a})$

   $\big(W\phi^\alpha(\bar{a}), W(\bar{y} - \phi(\bar{a}))\big) = 0 \quad \rightarrow \quad \mathcal{P}W(\bar{y} - \phi(\bar{a})) = 0$

   $\chi^2(\bar{a}) = ||(1 - \mathcal{P})W(\bar{y} - \phi(\bar{a})||^2$

2. Expand $\phi(\bar{a})$ about A

   $\phi(\bar{a}) = \phi(A) + \phi^\alpha(\bar{a})\delta\bar{a}_\alpha + O(\delta\bar{a}^2)$

   $\bar{y} - \phi(\bar{a}) = \bar{y} - \phi(A) + \phi^\alpha\delta\bar{a}_\alpha + O(\delta\bar{a}^2) = \delta\bar{y} + \phi^\alpha\delta\bar{a}_\alpha + O(\delta\bar{y}^2)$

3. $\chi^2(\bar{a}) = ||(1 - \mathcal{P})W\delta\bar{y}||^2 + O(\delta\bar{y}^3) + O(\delta\bar{y}^4)$

Taking $W = O(N^{1/2})$, i.e. a function of $C^{-1/2}$

   $\langle O(\delta\bar{y}^3) \rangle = 0$ up to corrections

$$\langle \chi^2(\bar{a}) \rangle = \mathrm{tr}[(1 - \mathcal{P})WCW] + O(N^{-1})$$

P-VALUE

$Q(\chi^2_{\rm obs}) = \int \mathrm{d}\bar{y}\ P_C(\bar{y})\ \theta(\chi^2(\bar{y}) - \chi^2_{\rm obs})$

1. a useful relation is given by change of variables $z = C^{-1/2}\delta\bar{y}$

$$\langle f(\delta\bar{y})\rangle = \int \mathrm{d}\bar{y}\ P_C(\bar{y})\ f(\delta\bar{y}) = \int \mathrm{d}z\ (2\pi)^{-N_{\rm x}/2} e^{-\frac{1}{2}||z||^2} f(C^{1/2}z)$$

2. replace $\chi^2(\bar{a})$ with $||(1-\mathcal{P})W\delta\bar{y}||^2 = (z, \nu z)$
   matrix $\nu = C^{1/2}W(1-\mathcal{P})WC^{1/2}$

$$Q(\chi^2_{\rm obs}) = \int \mathrm{d}z\ (2\pi)^{-N_{\rm x}/2} e^{-\frac{1}{2}||z||^2} \theta((z,\nu z) - \chi^2_{\rm obs})$$

Integral in Q evaluated numerically (easy)

Summary

[MB, Sommer '22]

For arbitrary W, i.e. uncorrelated fits, SVD cuts, correlated fits ...

robust statistical tests based on p-value (and reduced χ^2)

C and $C^{1/2}$, i.e. large (not small) eigenvalues of C

For correlated fits $[W = C^{-1/2}] + [\mathrm{tr} P = N_\mathrm{A}]$, imply:

$\mathrm{tr}[(1-\mathcal{P})WC] = \mathrm{tr}[1-\mathcal{P}] = N_\mathrm{x} - N_\mathrm{A} =$ degrees of freedom

similarly $\mathrm{tr}\nu = N_\mathrm{x} - N_\mathrm{A}$, so p-value takes standard form

C never known, only its estimator $\bar{C}$: consequences?

estimator of $\langle\chi^2\rangle$ with error $O(N^{-1/2})$

estimator of p-value, no closed-form for error

bootstrap? to be explored [MB, Kelly in prep]

Autocorrelations - I

[Madras-Sokal '88][Wolff '03][Schaefer et al. '11]

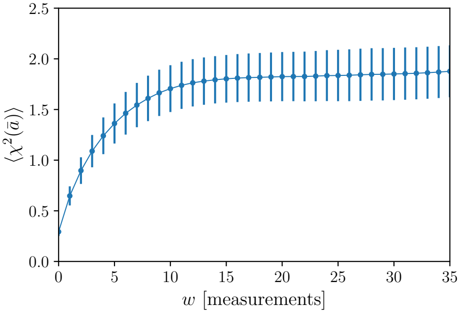
Assume $y_i(t)$ measured on single ensemble at Monte Carlo time t

autocorrelation function $\Gamma_{ij}(t) = \langle \Delta y_i(t+t_0)\, \Delta y_j(t_0) \rangle$

covariance matrix $C_{ij} = \frac{1}{N} \sum_{t=-\infty}^{\infty} \Gamma_{ij}(t)$

Expected χ^2 in presence of autocorrelations

$\langle \chi^2(\bar{a}) \rangle = \frac{1}{N} \sum_{t=-\infty}^{\infty} \mathrm{tr}\big[\Gamma(t)\, W(1-\mathcal{P})W\big]$



[MB, Sommer '22]

Estimator E of $\langle \chi^2 \rangle$

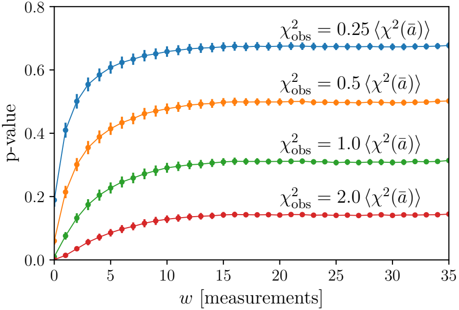
1. estimator of Γ
2. truncate sum (W) 3.

$\Delta E^2 \simeq \frac{2}{N}(2W+1)E^2$

example in toy model

Autocorrelations - II

Autocorrelation effects from matrix $\nu = C^{1/2} W (1 - \mathcal{P}) W C^{1/2}$



Uncorrelated fit

Error on p-value from several replicas

Example in toy model

P-VALUE

Correlated fits

large d.o.f.

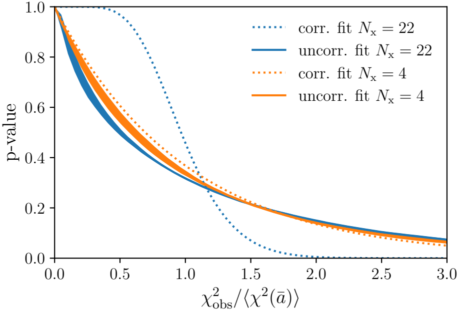
p-value sharp drop for $\frac{\chi^2}{\text{d.o.f}} \simeq 1$

$\rightarrow$ $\chi^2/\text{d.o.f} \simeq 1$ valid criterion

small d.o.f.

$\chi^2/\text{d.o.f} \simeq 1$ not good

$\rightarrow$ p-value preferable



Uncorrelated fits

$\chi^2/\text{d.o.f} \simeq 1$ meaningless; $\chi^2/\langle\chi^2\rangle \simeq 1$ not good criterion

$\rightarrow$ p-value from our method preferred choice

Conclusions

Lattice QCD

- predictions often involve fits to correlated and autocorrelated data
- correlated fits always preferred, but estimator of C is often singular
- alternatives: uncorrelated fits, SVD cuts...

Our work novel analytic control of $\langle \chi^2 \rangle$ and p-value

- unlocks robust statistical tests (only C,$C^{1/2}$ involved)

Autocorrelations are easily incorporated

- Γ-method, jackknife/bootstrap with binning

Thanks for your attention

A DIFFERENT METHOD

Bootstrap generates new 'fake' resampled ensembles

for each ensemble minimize χ^2

build histogram $\rightarrow$ well-defined "probability" density of χ^2? No

re-centering χ^2 allows to build proper density [Kelly Lattice '19]

well-defined p-value, valid for arbitrary W

More work under preparation [MB, Kelly in prep]

formal proof of equivalence of two ideas

extension of recentering to jackknife

study of $1/N$ neglected terms