How instabilities in covariance matrices affect the interpretation of LHC data

Ultimate Precision at hadron colliders, Institut Pascal (Paris-Saclay)

Zahari Kassabov,

in collaboation with Emanuele Nocera and Michael Wilson

November 27, 2019

Cavendish Laboratory, University of Cambridge

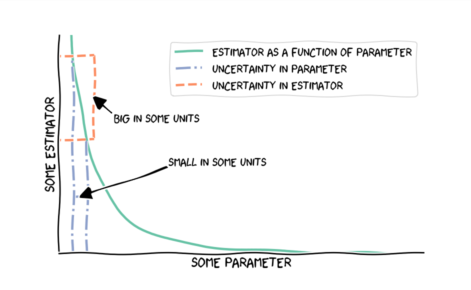
UNIVERSITY OF CAMBRIDGE

European Research Council

What are the causes of disagreement between data and theory?

- Inaccurate theory
    - Fixed order
    - Insufficient parametrization of degrees of freedom
- Inaccurate parameters of the theory
    - E.g. α_S.
- Inaccurate data
    - Underestimated experiment systematics.
    - Other problems with the measurement
- Instabilities on statistical estimators (this talk)

Instability



- Uncertainties on a parameter, even when *small*, can affect critically some function of them.
- This talk: Effect of uncertainties in covariance matrices on uncertainties on the χ^2 statistic.
- Still need to specify which parameter is varied and what are the small and big units.

The χ^2 statistic

$$\chi^2 = \sum_{i}^{N}\sum_{j}^{N}(\text{data}_i - \text{prediction}_i)\Sigma_{ij}^{-1}(\text{data}_j - \text{prediction}_j) = \delta^T\Sigma^{-1}\delta$$

- Predictions supplied by the theoretical model.
- Central measurement of data and covariance matrix Σ supplied by experiments.
- Used as:
    - Figure of merit in fits.
    - Assessment of agreement between data and theory.
- Underlying assumption: Experimental uncertainties distributed as a multivariate Gaussian.

Matrix of uncertainties

- Consider any matrix A such that

$$\Sigma = AA^t$$

- A can be chosen to have physical meaning:
  - $\mathbf{v} = \mathbf{f}(\mathbf{p})$ vector of N unknown interesting quantities.
  - $\mathbf{p}$ vector of M measurements, with central values $\mathbf{p^0}$ and independent Gaussian uncertainties $\mathbf{s}$. Assuming linear error propagation.
  - Then:

$$A_{ij} = \frac{\partial f_i}{\partial p_j}\Big|_{\mathbf{p}=\mathbf{p^0}} s_j$$

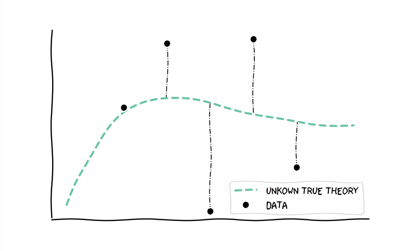
and $\mathbf{v}$ is multivariate Gaussian with mean $\mathbf{f}(\mathbf{p^0})$ and covariance AA^t

$$\mathbf{v} \sim \mathcal{N}(\mathbf{f}(\mathbf{p^0}), AA^t)$$

- Or else, A can be obtained e.g. using the Cholesky decomposition of Σ.

Sampling uncertainties around true theory

- We expect that experimental deviations around the true theory (which we do not know) to be $\delta \sim \mathcal{N}(0, \Sigma)$
  - To sample, generate M independent numbers with $\mathbf{n} \sim \mathcal{N}(0, I)$ and do $\delta = A\mathbf{n}$.



- The quantity $\chi^2 = \left\| A^+ \delta \right\|^2 = \left\| A^+ A \mathbf{n} \right\|^2$ is a random variable following a χ^2 distribution with N degrees of freedom.
  - $\langle \chi^2 \rangle = \left\| A^+ A \right\|_F^2 = N$
    - $\left(\left\| A \right\|_F = \mathrm{tr}(AA^t)^{1/2} = \sqrt{\sum_i \sum_j A_{ij}^2} \right)$.
  - Standard deviation $\sqrt{2N}$

Defining stability upon mistakes in the covariance matrix

- Imagine the residuals are sampled using A, but we are given a different matrix, $\bar{A}$ to estimate the χ^2 statistic. The expected value of the χ^2, with the *wrong* matrix, $\bar{\chi}^2$, is

$$\langle \bar{\chi}^2 \rangle = \left\| \bar{A}^+ A \right\|_F^2$$

$$\Delta\chi^2 = \langle \bar{\chi}^2 \rangle - \langle \chi^2 \rangle = \left\| \bar{A}^+ A \right\|_F^2 - N$$

- We assert that the χ^2 statistic is stable upon making this mistake, if it results in differences that are smaller than its statistical fluctuations.

$$\left| \Delta\chi^2 \right| < \sqrt{2N}$$

Toy model example

- Experiments have reached an impressive level of statistical precision.
  - Statistical component of the uncertainty (typically uncorrelated across bins) less important.
  - Systematic uncertainties (correlated across bins) tend to dominate.
- A somewhat realistic toy model for a matrix of uncertainties from HepData:

$$A = \begin{pmatrix} \epsilon & 0 & 0 & 0 & 1 \\ 0 & \epsilon & 0 & 0 & 1 \\ 0 & 0 & \epsilon & 0 & 1 \\ 0 & 0 & 0 & \epsilon & 1 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \epsilon^2 + 1 & 1 & 1 & 1 \\ 1 & \epsilon^2 + 1 & 1 & 1 \\ 1 & 1 & \epsilon^2 + 1 & 1 \\ 1 & 1 & 1 & \epsilon^2 + 1 \end{pmatrix}$$

with $\epsilon^2 \ll 1$.

  - Assumes 4 data points, and uncorrelated error of size ϵ and one completely correlated systematic of size 1.

Model for uncertainties in the correlations

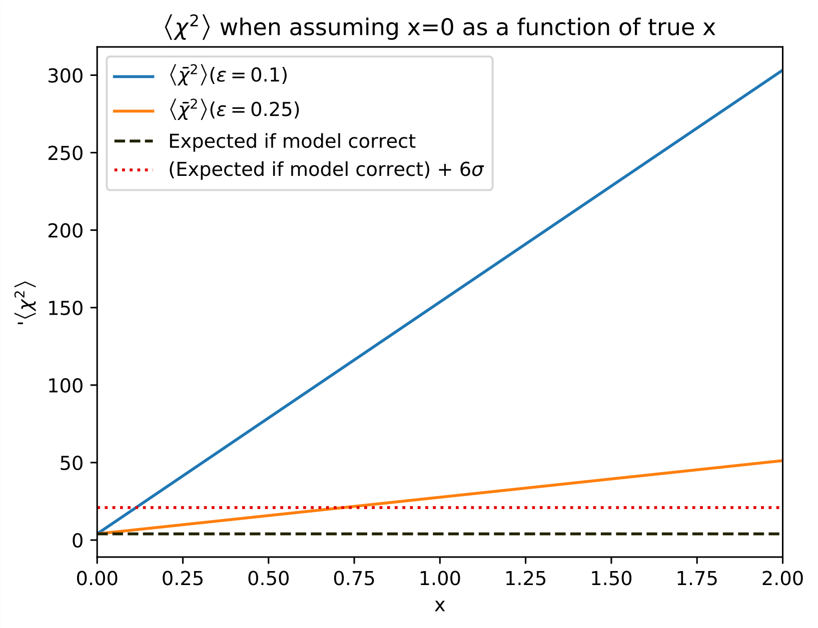
- Add unknown parameter $x \in [0, 2]$ controlling the correlations of the last bin.

$$A(x) = \begin{pmatrix} \epsilon & 0 & 0 & 0 & 1 & 0 \\ 0 & \epsilon & 0 & 0 & 1 & 0 \\ 0 & 0 & \epsilon & 0 & 1 & 0 \\ 0 & 0 & 0 & \epsilon & 1-x & \sqrt{1-(1-x)^2} \end{pmatrix}$$

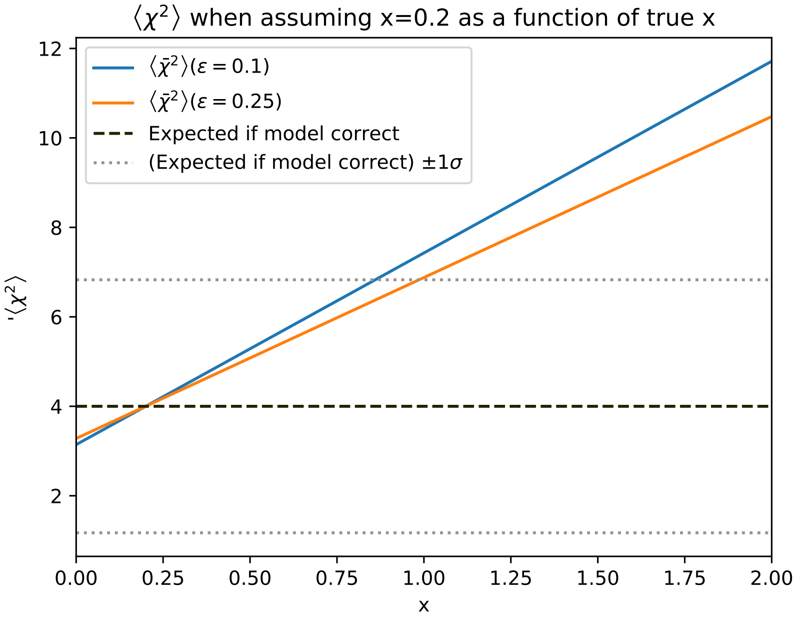
$$\Sigma(x) = \begin{pmatrix} \epsilon^2+1 & 1 & 1 & 1-x \\ 1 & \epsilon^2+1 & 1 & 1-x \\ 1 & 1 & \epsilon^2+1 & 1-x \\ 1-x & 1-x & 1-x & \epsilon^2+1 \end{pmatrix}$$

  - We are keeping the total variance fixed. It is realistic to think that x could be anywhere in the range.
- We have $\langle \bar{\chi}^2 \rangle = \left\| A^+(x_{\text{experimental}}) A(x_{\text{true}}) \right\|_F^2$
- Experimental results often presented by default with the highest correlation (i.e. $x_{\text{experimental}} = 0$).

Stability when assuming the highest correlation



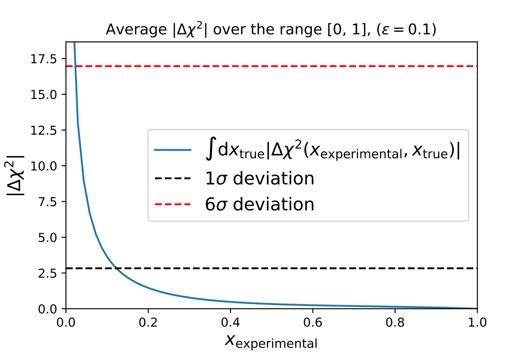
Stability when assuming a lower correlation



Optimizing for stability

- Given some prior probability density on x, $P(x_{\text{true}})$, we can find the value for $x_{\text{experimental}}$ that optimizes for stability

$$x^* = \underset{x_{\text{experimental}}}{\operatorname{argmin}} \int \left| \left\| A^+(x_{\text{experimental}}) A(x_{\text{true}}) \right\|_F^2 - N \right| P(x_{\text{true}}) \mathrm{d}x_{\text{true}}$$



Toy model, assuming $P(x_{\text{true}}) = \text{Uniform}[0, 1]$

Toy model summary and tentative conclusions

- Assuming smaller correlation wherever they are unknown seems like a good rule of thumb.
    - This is consistent with other studies, e.g. ATLAS Jets at 7 TeV (arxiv: 1410.8857): Enormous sensitivity to correlations studied in detail in [Harland-Lang, Martin, Thorne arxiv:1711.05757].

	Full	21	62	21,62
$\chi^2/N_{\text{pts.}}$	2.85	1.58	2.36	1.27

Table 1: χ^2 per number of data points ($N_{\text{pts}} = 140$) for fit to ATLAS jets data [23], with the default systematic error treatment ('full') and with certain errors, defined in the text, decorrelated between jet rapidity bins.

- In practice, not enough information to compute $\bar{A}^{+}A$ available from public data. Need to make some simplifications and assumptions.

Solving a different problem

- We typically have no information at all regarding the uncertainties of the experimental uncertainties.
- But unstable covariance matrices will lead to artificial discrepancies.

We want to solve a different problem that

- Avoids yielding data-theory discrepancies wherever those are likely due to instabilities.
- Gives the same answers when the answers are not affected by instabilities.
- Does not result in decreased uncertainties anywhere.

In practice, find a new, regularized covariance matrix.

- Avoids the instabilities we assume to be problematic.
- We do need to make assumptions state what these are.
- General principle: Come up with covariance matrices that are, in all likelihood, compatible with the original ones within their precision.

Upper bound to instabilities

- We have:

$$\langle \bar{\chi}^2 \rangle = \left\| \bar{A}^+ A \right\|_F^2$$

- Write

$$A = \bar{A} + \delta F$$

with δ a scalar parameter and F a matrix.

- Then

$$\langle \bar{\chi}^2 \rangle \leq 2\sqrt{N}\delta \|A^+\|_2 \|F\|_F$$

$$\left(\|A\|_2 = \max_{\{\mathbf{x} \in \mathbb{R}^M : \|\mathbf{x}\| = 1\}} \|A\mathbf{x}\| = \max \text{singular value}(A) \right)$$

$$\left(\|A^+\|_2 = \frac{1}{\min \text{singular value}(A)} \right)$$

- Hence the condition

$$\delta \left\| \bar{A}^+ \right\|_2 \|F\|_F < 1$$

is sufficient to avoid overestimating χ^2.

- Problem reduced to defining a value for δ and a model for $\|F\|_F$.

Correlation matrix regularization

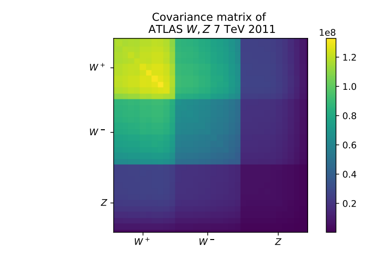
- Experimental covariance matrices can relate data with different magnitudes and even different units.
    - $\|F\|_F$ not particularly meaningful (units?).
    - The upper bound is a worst case. We don't want to include mislabelling of uncertainties in the analysis.
- On the other hand, estimating experimental correlations is well known to be challenging.
- Resolution: Assume the diagonal uncertainties are correct for the purposes of the regularization. Regularize the correlation matrix instead.
    - Note that the correlation matrix is the covariance of

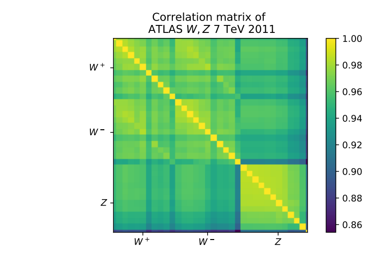
$$\frac{(\text{data} - \text{theory})}{\text{diagonal uncertainty}}$$

    so everything so far applies to these reduced variables.

Example: ATLAS WZ rapidity 2011

- The data from ATLAS W/Z production at 7 TeV [arxiv 1612.03016] is a representative example.
- Bad fit quality ($\chi^2/N = 75/34$ for NNPDF 3.1) has attracted some discussion.





- Correlation matrix clearly unstable, and not dissimilar to the toy model.

Assumptions on δ

- δ measures the size of the uncertainties on the uncertainties.
- It is not possible to retrieve that information from the public analysis.
- Therefore there is a fundamental ambiguity.
- In practice choose so that the resulting regularized covariance matrices differ little from the original ones.
    - E.g. given some regularization (to be described), impose δ such that diagonal elements change less than ~10% in the very worst case.

Assumptions on $\|F\|_F$

- Because we are regularizing correlations, $\|F\|_F$ is dimensionless.
- Need to specify how $\|F\|_F$ behaves as a function of N. This is important in a PDF fit because we have datasets of many different size (between 3 and 416 points for NNPDF 3.1).
- In practice choose so that the resulting regularized covariance matrices differ little from the original ones. This corresponds to assuming that $\|F\|_F = \text{const}(N)$.
  - Assumption same amount of *wrongness* irrespective of the number of data points.
  - We set

$$\|F\|_F = 1$$

Regularization procedure

The stability condition is finally

$$\delta \left\| \bar{A}_{\text{corr}}^{+} \right\|_2 \|F\|_F < 1 \Rightarrow \left\| \bar{A}_{\text{corr}}^{+} \right\|_2 < \frac{1}{\delta}$$

We regularize A by clipping the singular values of A_{corr} from below, so the condition is satisfied.

- We compute the Singular Value Decomposition of A_{corr}

$$A = DA_{\text{corr}} = DUSV^t$$

- Find regularized singular values

$$s_i^{\text{reg}} = \begin{cases} s_i & \text{if } s_i > \frac{1}{\delta} \\ \frac{1}{\delta} & \text{otherwise} \end{cases}$$

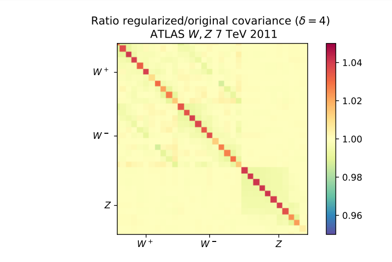
- D Diagonal matrix of standard deviations.
- U and V orthogonal matrices.
- S Diagonal matrix of singular values of A_{corr}.

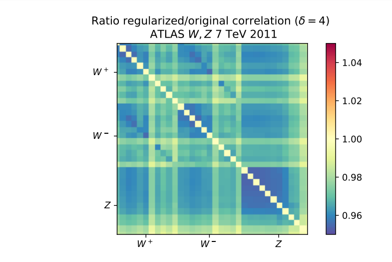
Finally

$$A_{\text{reg}} = DUS^{\text{reg}}V^t$$

Regularization on ATLAS WZ rapidity

δ	χ^2/N (NNPDF3.1)	Max change in diagonal uncertainties
∞	2.2	0
5	1.6	2%
4	1.2	4%
3	0.77	8.5%





Regularization on ATLAS WZ rapidity: Tentative conclusion

- Correlation matrix highly unstable.
  - Reasonable to hypothesize that discrepancies measured by the χ^2 are spurious.
- Can find an almost indistinguishable covariance matrix that gives perfect agreement.
- Note χ^2 with fixed PDF that included the unstable data does not have to coincide with the result including regularized data in the fit instead.

Regularized global PDF fits

- Made full NNPDF 3.1 NNLO-like fits, for several choices of thresholds.
- Only few dataset affected. Rest already “stable”.
- PDFs themselves hardly change, in terms of distance between functions.
- χ^2 estimators improve substantially.

Regularized datasets and pre-fit χ^2

δ	7			5			3		
	(chi2 - N)/sqrt(2N))	**cov diag diff (%)**	**corr diff (abs)**	**(chi2 - N)/sqrt(2N))**	**cov diag diff (%)**	**corr diff (abs)**	**(chi2 - N)/sqrt(2N))**	**cov diag diff (%)**	**corr diff (abs)**
BCDMSP	3.288	0	0	3.288	0	0	2.892	5.616	4.971E-2
BCDMSD	0.9470	0	0	0.9470	0	0	0.9233	2.409	2.025E-2
CHORUSNU	1.951	0	0	1.949	0.2726	2.779E-3	0.7943	7.180	6.703E-2
CDFZRAP	1.644	0	0	1.112	1.076	1.138E-2	-0.3105	8.222	7.975E-2
ATLASWZRAP36PB	-0.1758	0	0	-0.1758	0	0	-0.3405	2.378	2.277E-2
ATLASZHIGHMASS49FB	0.8799	0	0	0.8799	0	0	0.8660	0.4398	7.369E-3
ATLASLOMASSDY11EXT	-0.1779	0	0	-0.1779	0	0	-0.2287	2.564	3.012E-2
ATLASWZRAP11	4.133	0.5015	6.516E-3	2.175	2.160	2.473E-2	-1.238	9.438	9.489E-2
ATLAS1JET11	-0.1607	1.073	1.466E-2	-0.7020	2.724	3.338E-2	-1.543	9.854	0.1018
CMSDY2D11	1.984	0.4218	5.386E-3	1.936	1.534	1.941E-2	1.475	6.079	7.643E-2
CMSWMU8TEV	-1.209	1.046	1.355E-2	-1.945	2.891	3.253E-2	-2.614	10.42	0.1053
CMSJETS11	-0.3777	0.6244	6.024E-3	-0.7274	2.421	2.492E-2	-2.474	9.573	9.234E-2
CMSZDIFF12	1.153	0	0	1.153	0	0	0.6198	3.198	3.520E-2

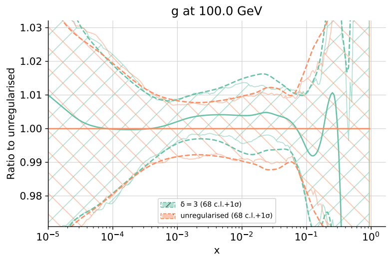
Results from global fits

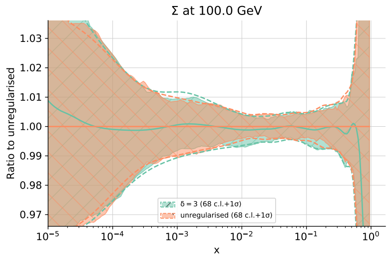
- Global χ^2 improved by up to 2 sigma.
- Combined ATLAS + CMS χ^2 can be made order 1.

Threshold	Global χ^2/(3979 datapoints)	ATLAS χ^2/(211 datapoints)	CMS χ^2/(328 datapoints)
∞	1.16	1.17	1.17
5	1.15	1.06	1.03
4	1.13	1.00	0.96
3	1.10	0.89	0.85

Changes in PDF themselves

- We observe few differences in the PDF themselves.





Conclusions

- Instabilities in statistical estimators affect notably description of the data.
- Regularization remedies best applied by experimentalists, since useful information is available in the experimental analysis only.
- Proposed a method to avoid instabilities on χ^2.
    - Using minimal information.
    - Independent on what the theory is.
    - Little change in PDFs, but notable change in the interoperation of the results.

Thank you!