Including theoretical uncertainties in global (flavor) fits

Luiz Vale Silva

Jožef Stefan Inst.

May 15th, 2017

GDR "QCD": *Progresses in algorithms and numerical tools for QCD*
Orsay, France

Outline

Statistical and theoretical uncertainties

$\rightarrow$ Statistical uncertainties result from the intrinsic variability of data, typically distributed normally

Statistical and theoretical uncertainties

$\rightarrow$ Statistical uncertainties result from the intrinsic variability of data, typically distributed normally

$\rightarrow$ Theoretical uncertainties are different in nature: they are modeling parameters (ξ), fixed and unknown, that incorporate our incomplete knowledge about the properties of a distribution [Punzi '01] (Ex.: truncation of a perturbative series)

Statistical and theoretical uncertainties

$\rightarrow$ Though *a prori* theoretical uncertainties are a universal issue, in the context of quark **flavor physics** they are particularly important, due to the **strong dynamics**

(Cf., e.g., EW global fit extraction of $\{M_Z, \alpha_s(M_Z), \ldots\}$: statistical uncertainties dominate)

Statistical and theoretical uncertainties

→ Though *a prori* theoretical uncertainties are a universal issue, in the context of quark **flavor physics** they are particularly important, due to the **strong dynamics**

(Cf., e.g., EW global fit extraction of $\{M_Z, \alpha_s(M_Z), \ldots\}$: statistical uncertainties dominate)

→ **Here**: discuss theoretical uncertainties, more specifically in the context of quark **flavor physics**

J. Charles, S. Descotes-G., V. Niess, LVS
Eur.Phys.J. C77 (2017), 214
[hep-ph/1611.04768]

Outline

CKMfitter and required inputs

→ **CKMfitter**: global fit package ▸ CKMfitter

Web interface version (comments are welcomed!) ▸ CKMlive

- Examples of classes of processes that require non-perturbative theoretical inputs:

Meson-mixing	$B_{(s)}\overline{B}_{(s)}$, $K\overline{K}$: bag-parameters $\widehat{B}_{B_s}$, $\widehat{B}_{B_s}/\widehat{B}_{B_d}$, $\widehat{B}_K$
(semi-)leptonic decays	$\pi \to \ell\nu$, $K \to \pi\ell\nu$, etc.: decay constants, form factors Ex.: f_π, $f_+^{K\to\pi}(0)$

CKMfitter and required inputs

→ **CKMfitter**: global fit package ▸ CKMfitter

Web interface version (comments are welcomed!) ▸ CKMlive

- Examples of classes of processes that require non-perturbative theoretical inputs:

Meson-mixing	$B_{(s)}\overline{B}_{(s)}$, $K\overline{K}$: bag-parameters $\widehat{B}_{B_s}, \widehat{B}_{B_s}/\widehat{B}_{B_d}, \widehat{B}_K$
(semi-)leptonic decays	$\pi \to \ell\nu$, $K \to \pi\ell\nu$, etc.: decay constants, form factors Ex.: f_π, $f_+^{K\to\pi}(0)$

- Nowadays, extraction of non-pert. parameters: Lattice QCD
- Dominance of **systematic uncertainties** (continuum extrapolation, finite volume, mass inter/extrapolations, etc.)

Statistical approach

- **CKMfitter**: Frequentist statistic based on a χ^2 analysis
- χ^2_{min}: **goodness-of-fit** under SM or NP, **estimators** for V_{CKM}
- $\Delta\chi^2$ (χ^2-distributed): **Confidence Level** (CL) intervals
- *Range* fit (*R*fit) scheme incorporates theoretical uncertainties

$$\mathcal{L} \overset{Rfit}{=} \mathcal{L}_{stat} \times \mathcal{L}_{theo}, \; \chi^2 = -2\,\ell\mathrm{n}\,\mathcal{L}$$

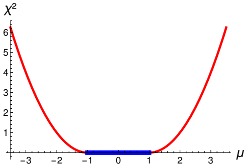
$\mathcal{L}_{stat}$: agreement of data & prediction

$\mathcal{L}_{theo}$: accuracy of QCD parameters

theo. uncertainties strictly contained in a range. Ex.: $\xi \in [-\Delta, \Delta]$

Example in 1D,

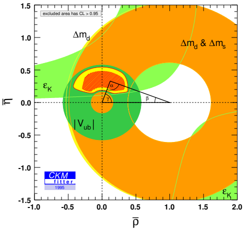
$0 \pm 1_{stat} \pm 1_{theo}$



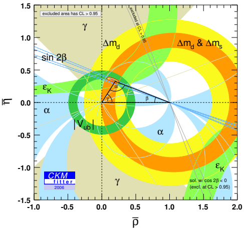
flat bottom, quadratic walls

Extraction of the CKM matrix elements

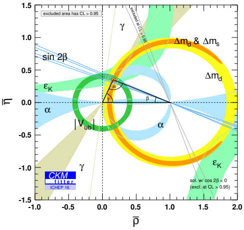
$\rightarrow$ Better theoretical control (Lattice QCD), and
more accurate data (LEP, KTeV, NA48, BaBar, Belle, CDF, DØ, LHCb, CMS, ...)



1995



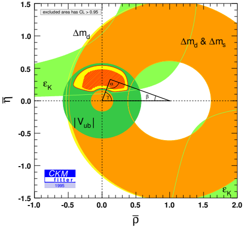
2006



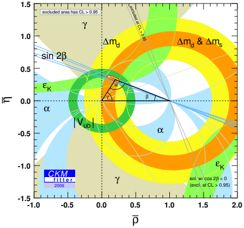
2016

Extraction of the CKM matrix elements

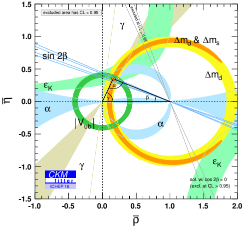
→ Better theoretical control (Lattice QCD), and more accurate data (LEP, KTeV, NA48, BaBar, Belle, CDF, DØ, LHCb, CMS, ...)



1995



2006



2016

→ **Question**: is there a more appropriate statistical approach than *R*fit to incorporate theoretical uncertainties, given the present and expected progresses?

Outline

Basic concepts in *frequentist* statistic

Gaussian case, without theoretical uncertainty

→ Consider an apparatus designed to measure the true value "x_t" of an observable (the one that is actually realized in nature)

$g(X; x_t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(X-x_t)^2}{2\sigma^2}\right]$ Ex.: LEP $\Rightarrow$ $\widehat{M}_Z^{(1)}, \ldots, \widehat{M}_Z^{(N)}$

→ Test statistic T: arbitrary as long as small values attest the agreement between the data and the predicted value(s) (under a certain hypothesis, $\mathcal{H}_\mu : x_t = \mu$ Ex.: $\mu = M_Z^{SM}$)

→ Maximum Likelihood Ratio (**MLR**): [Neyman, Pearson]

$$T(\overrightarrow{X}_0; \mu) = \sum_{i=1}^{N} \frac{(X_0^{(i)} - \mu)^2}{\sigma^2}$$

→ From the distribution of the i.i.d. random variable X, we determine the distribution of the test statistic, seen as a function of X

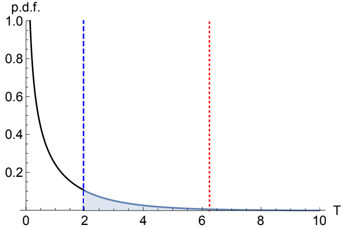
Example: $T(X; \mu) \sim \chi^2(N)$ for each μ

Basic concepts in *frequentist* statistic

$\rightarrow$ Given the real data X_0 from a *single* experiment, the probability of measuring a new value of the test statistic in worsen agreement is called the p-value

$$p(X_0;\mu) = \mathcal{P}[T \geq T(X_0;\mu)]$$

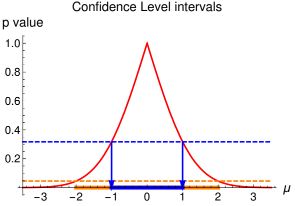
$$p(X_0;\mu) \simeq 0.32 \overset{1D}{\leftrightarrow} T(X_0;\mu) = 1 \equiv (1\,\sigma)^2,$$
$$p(X_0;\mu) \simeq 0.05 \overset{1D}{\leftrightarrow} T(X_0;\mu) = 4 \equiv (2\,\sigma)^2,$$
$$\vdots$$



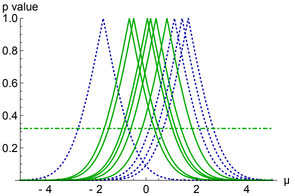
In formulas:

$$h(T|\mathcal{H}_\mu) = \int dX\, \delta[T - T(X;\mu)]\, g(X;\mu)$$
$$\mathcal{P}[T < T(X_0;\mu)] = \int_0^{T(X_0;\mu)} dT\, h(T|\mathcal{H}_\mu)$$

Basic concepts in *frequentist* statistic



Ex.: $x_t = 0,\ \sigma = 1$, w/ $\{X_0^{(1)}, \ldots, X_0^{(10)}\}$



→ Interpretation of $\alpha\%$ CL intervals: asymptotically, a fraction $\alpha\%$ of the CLs include the true value x_t

→ A *p*-value that respects this property (called coverage) is said to be exact

- CLs $\ni x_t$ in $\alpha\%$ times: *exact*
- ... in $> \alpha\%$ times: *conservative*
- ... in $< \alpha\%$ times: *aggressive*

Modeling theoretical uncertainties

Given the one-dimensional (1D) case: $X \sim X_0 \pm \underbrace{\sigma}_{\textit{statistical}} \pm \underbrace{\Delta}_{\textit{theoretical}}$

→ The true value of the theo. uncertainty ξ is fixed and unknown

→ Being unknown, one quotes a range $\xi \in \Omega$ and **vary** ξ

→ Usually, one has in mind that $\Omega = [-\Delta, \Delta]$, but this may miss an unexpectedly large value of ξ

→ Were ξ known, we would quote instead $(X_0 + \xi) \pm \sigma$

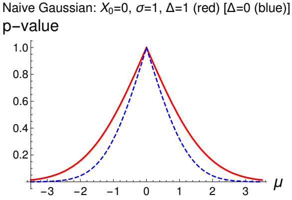
Modeling theoretical uncertainties: random

Random approach

→ Different techniques of calculation lead to different predictions around the exact one (pseudo-randomly distributed)

→ Naive Gaussian (nG): $\xi \sim \mathcal{N}_{(0,\Delta)}$

→ MLR ($\mathcal{H}_\mu : x_t = \mu$): $T(X;\mu) = \frac{(X-\mu)^2}{\sigma^2+\Delta^2}$



Modeling theoretical uncertainties: external

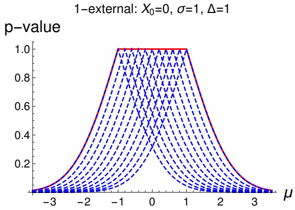
n-external approach [Scan: Dubois-Felsmann et al.]

$\rightarrow$ In a first step, assume that ξ is known

$\rightarrow$ Family of hypotheses, $\mathcal{H}_\mu^{(\xi)} : x_t = \mu + \xi$

$\rightarrow$ MLR ($\mathcal{H}_\mu^{(\xi)}$): $T(X;\mu) = \frac{(X-\mu-\xi)^2}{\sigma^2}$

$\rightarrow$ Combine the p_ξ, for $\xi \in n \times [-\Delta, \Delta]$



Close to what some experiments interpret as theo. uncertainties

Simple 1D case: **Rfit** and **1-external** are equivalent

Modeling theoretical uncertainties: nuisance

$$\text{MLR: } T(X_0;\mu) = -2\,\ell\text{n}\,\underbrace{\frac{\mathcal{L}_{X_0}(\mu)}{\max_\mu \mathcal{L}_{X_0}(\mu)}}_{\textit{Simple hypothesis}} \to -2\,\ell\text{n}\,\underbrace{\frac{\max_\xi \mathcal{L}_{X_0}(\mu,\xi)}{\max_{\mu,\xi} \mathcal{L}_{X_0}(\mu,\xi)}}_{\textit{Composite hypothesis}}$$

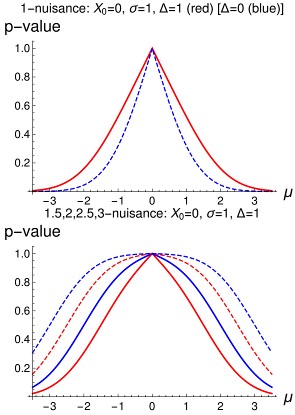
Fixed-n nuisance

$\to$ ~MLR ($\mathcal{H}_\mu : x_t = \mu$): $T(X;\mu) = \frac{(X-\mu)^2}{\sigma^2+\Delta^2}$

$\to$ ξ strictly found in $n \times [-\Delta, \Delta]$

$\to$ Small n may lead to reasonable CLs, but possibly uncovering

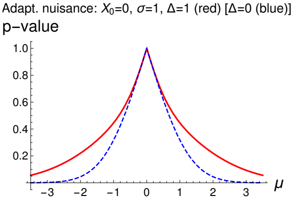
$\to$ Large n avoid uncovering, but lead to large CLs



Modeling theoretical uncertainties: nuisance

Adaptive nuisance [Charles, Descotes-G., Niess, LVS]

$\rightarrow \sim$MLR ($\mathcal{H}_\mu : x_t = \mu$): $T(X;\mu) = \frac{(X-\mu)^2}{\sigma^2+\Delta^2}$



$\rightarrow$ The interval where we look for ξ grows w/ the CL interval we want to quote

$\rightarrow$ n CL intervals: $\xi \in n \times [-\Delta, \Delta]$

Designed to deal with:

- Metrology/extraction of parameters ($1-2\ \sigma$ intervals)
- Minimizing Type-II (false positive) errors (above $\sim 5\ \sigma$)

adapt. nuisance for a significance $n \equiv n$-fixed nuisance

Incorporating theoretical uncertainties

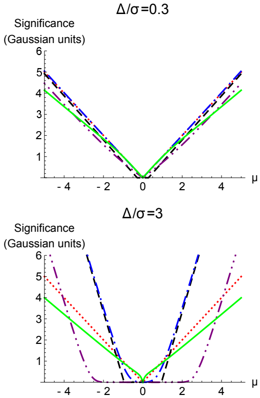
Approaches for dealing with theoretical uncertainties (some guiding principles for choosing a scheme):

- Good coverage properties (at least for the CL significances we are interested in)
- *Useful* metrology: reasonable size of CL intervals
- Propagation of uncertainties: clear separation of statistical and theoretical uncertainties

Illustration of confidence level intervals, 1D

→ Consider $0 \pm \sigma \pm \Delta$, w/ fixed $\sigma^2 + \Delta^2 = 1$

→ Gaussian units: $\sqrt{2}\,\mathrm{Erf}^{-1}(1 - p(X_0; \mu))$

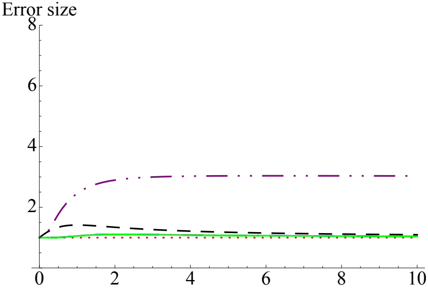


(**red**) naive Gaussian (nG); (**black**) fixed-1 external/Rfit; (**blue**) fixed-1 nuisance;

(**purple**) fixed-3 nuisance; (**green**) adaptive nuisance

Illustration of confidence level intervals, 1D

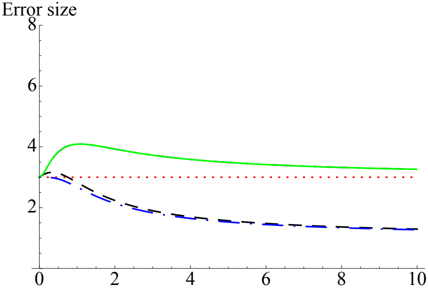
CL interval size vs. Δ/σ; $1\,\sigma$ significance



(**red**) naive Gaussian (nG); (**black**) fixed-1 external/Rfit $\xi \in [-\Delta, \Delta]$;
(**blue**) fixed-1 nuisance $\xi \in [-\Delta, \Delta]$; (**purple**) fixed-3 nuisance $\xi \in 3[-\Delta, \Delta]$;
(**green**) adaptive nuisance

Illustration of confidence level intervals, 1D

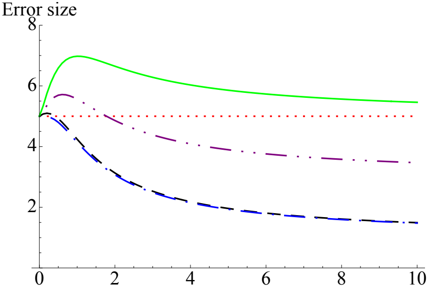
CL interval size vs. Δ/σ; $3\,\sigma$ significance



(**red**) naive Gaussian (nG); (**black**) fixed-1 external/Rfit $\xi \in [-\Delta, \Delta]$;
(**blue**) fixed-1 nuisance $\xi \in [-\Delta, \Delta]$; (**purple**) fixed-3 nuisance $\xi \in 3[-\Delta, \Delta]$;
(**green**) adaptive nuisance

Illustration of confidence level intervals, 1D

CL interval size vs. Δ/σ; $5\,\sigma$ significance



(**red**) naive Gaussian (nG); (**black**) fixed-1 external/Rfit $\xi \in [-\Delta, \Delta]$; (**blue**) fixed-1 nuisance $\xi \in [-\Delta, \Delta]$; (**purple**) fixed-3 nuisance $\xi \in 3[-\Delta, \Delta]$; (**green**) adaptive nuisance

Frequency of coverage of x_t

Limit case: the simulated ξ is at the edge of $[-\Delta, \Delta]$

$\Delta/\sigma = 3,\ \xi/\Delta = 1$	68.27% CL	95.45% CL	99.73% CL
nG	56.3%	100.0%	100.0%
1-nuisance	68.1%	95.5%	99.7%
adaptive nuisance	68.2%	100.0%	100.0%
1-external/Rfit	84.1%	97.7%	99.9%

Unfortunate case: the simulated ξ is outside $[-\Delta, \Delta]$

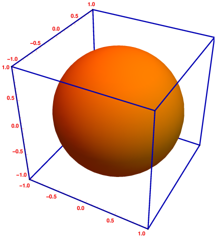
$\Delta/\sigma = 3,\ \xi/\Delta = 3$	68.27% CL	95.45% CL	99.73% CL
nG	0.00%	0.35%	68.7%
1-nuisance	0.00%	0.00%	0.07%
adaptive nuisance	0.00%	9.60%	99.8%
1-external/Rfit	0.00%	0.00%	0.13%

Multi-dimensional case

$$X_0^{(i)} \pm \sigma_i \pm \Delta_i,\ \xi_i \in [-\Delta_i, \Delta_i] \quad (\text{or } X_0 \pm \sigma \pm \Delta_1 \pm \ldots \pm \Delta_N)$$

Average: $\xi = \sum_{i=1}^N w_i\, \xi_i$, w/ weights $\sum_{i=1}^N w_i = 1\,, w_i \geq 0$

Edges: $[-\Delta_i, \Delta_i]$



Interval where the bias ξ_i is varied

Hyper-cube: assuming extreme values simultaneously $\widehat{\Delta} = \sum_{i=1}^N w_i \Delta_i$

Hyper-ball: $\widehat{\Delta} = \sqrt{\sum_{i=1}^N (w_i \Delta_i)^2}$

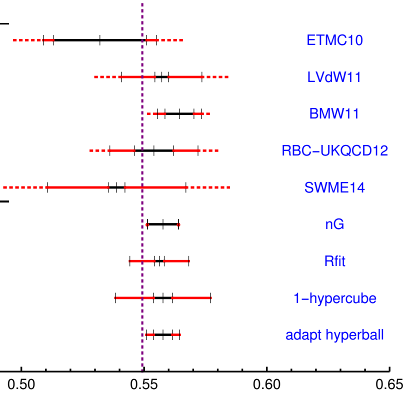
$$\Rightarrow \widehat{X}_0 \pm \widehat{\sigma} \pm \widehat{\Delta}$$

Further issue: "correlated" theoretical uncertainties lead to *deformed hyper-cubes* and *hyper-ellipsoids* (Ex.: 100 % $\Rightarrow$ single ξ)

Combining data

Example: combination of different extractions of $B_K^{\overline{\mathrm{MS}}}(2\ \mathrm{GeV})$

	CV	**stat**	(theo) quad	(theo) lin
ETMC10	0.532	(19)	(12)	(26)
LVdW11	0.557	(03)	(15)	(26)
BMW11	0.564	(06)	(06)	(10)
RBC-UKQCD12	0.554	(08)	(14)	(22)
SWME14	0.539	(03)	(27)	(44)
nG	0.5577	(63)	-	
*R*fit	0.556	(02)	(10)	
1-hypercube	0.558	(04)	(18)	
adapt hyperball	0.5577	(38)	(50)	

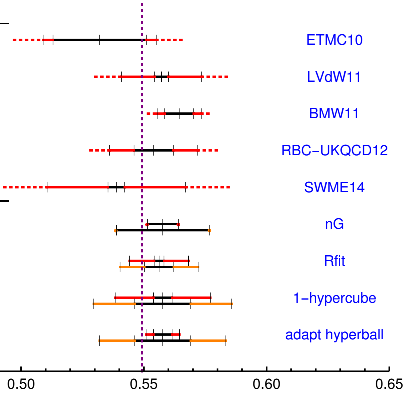


(red edges: $1\,\sigma$; purple: "naive" average of the CVs)

Combining data

Example: combination of different extractions of $B_K^{\overline{\mathrm{MS}}}(2\ \mathrm{GeV})$

	CV	**stat**	(theo) quad	(theo) lin
ETMC10	0.532	(19)	(12)	(26)
LVdW11	0.557	(03)	(15)	(26)
BMW11	0.564	(06)	(06)	(10)
RBC-UKQCD12	0.554	(08)	(14)	(22)
SWME14	0.539	(03)	(27)	(44)
nG	0.5577	(63)	-	
*R*fit	0.556	(02)	(10)	
1-hypercube	0.558	(04)	(18)	
adapt hyperball	0.5577	(38)	(50)	

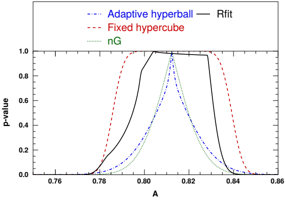


(orange edges: $3\,\sigma$; purple: "naive" average of the CVs)

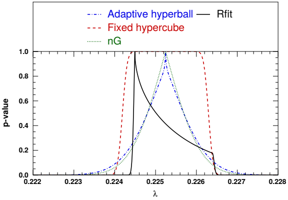
Illustrative global fit extraction of $A, \lambda, \overline{\rho}, \overline{\eta}$

$\underbrace{|V_{ud}|, |V_{ub}|, |V_{cb}|, \Delta m_d, \Delta m_s}_{\text{theo. dominated}}, \underbrace{\alpha, \sin(2\beta), \gamma}_{\text{stat. dominated}}$

[Inputs & details: Charles et al. '16]



A Method	$1\ \sigma$	$3\ \sigma$
nG	0.812 ± 0.011	0.812 ± 0.033
*R*fit	$0.804^{+0.029}_{-0.014}$	$0.804^{+0.038}_{-0.030}$
1-hypercube	0.812 ± 0.029	0.812 ± 0.038
adapt. hyperball	0.812 ± 0.012	0.812 ± 0.042

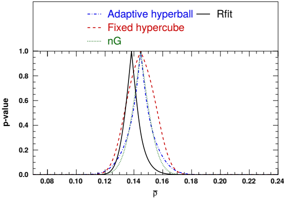


λ Method	$1\ \sigma$	$3\ \sigma$
nG	0.2252 ± 0.0007	0.2252 ± 0.0020
*R*fit	$0.2245^{+0.0011}_{-0.0001}$	$0.2245^{+0.0020}_{-0.0001}$
1-hypercube	0.2252 ± 0.0011	0.2252 ± 0.0013
adapt. hyperball	0.22525 ± 0.00070	0.2252 ± 0.0022

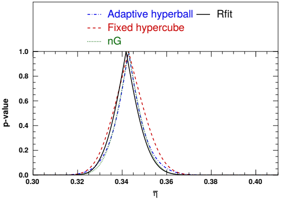
Illustrative global fit extraction of $A, \lambda, \overline{\rho}, \overline{\eta}$

$$\underbrace{|V_{ud}|, |V_{ub}|, |V_{cb}|, \Delta m_d, \Delta m_s}_{\textit{theo. dominated}}, \underbrace{\alpha, \sin(2\beta), \gamma}_{\textit{stat. dominated}}$$

[Inputs & details: Charles et al. '16]



$\bar{\rho}$ Method	$1\ \sigma$	$3\ \sigma$
nG	0.145 ± 0.009	0.145 ± 0.027
*R*fit	0.138 ± 0.007	$0.138^{+0.028}_{-0.020}$
1-hypercube	0.145 ± 0.015	0.145 ± 0.031
adapt. hyperball	0.145 ± 0.009	0.145 ± 0.036



$\bar{\eta}$ Method	$1\ \sigma$	$3\ \sigma$
nG	0.343 ± 0.008	0.343 ± 0.023
*R*fit	0.342 ± 0.008	$0.342^{+0.024}_{-0.022}$
1-hypercube	0.343 ± 0.011	0.343 ± 0.027
adapt. hyperball	0.343 ± 0.008	0.343 ± 0.028

Outline

Conclusions

→ Theoretical uncertainties are omnipresent in flavor analyses and deserve a careful look

→ Reported progresses in the modeling of theoretical uncertainties, introducing the adaptive nuisance approach

→ The choice of the scheme has an impact on:

- confidence level intervals,
- metrology,
- significance of a tension, etc.

CKMfitter: adaptive nuisance, candidate for further investigation (coverage properties, clear separation of stat. and theo. uncertainties)

Merci !

CKMfitter collaboration

Jérôme Charles	Theory	CPT Marseille (France)
Olivier Deschamps	LHCb	LPC Clermont-Ferrand (France)
Sébastien Descotes-Genon	Theory	LPT Orsay (France)
Heiko Lacker	ATLAS/BABAR	Humboldt-Universität Berlin (Germany)
Stéphane Monteil	LHCb	LPC Clermont-Ferrand (France)
José Ocariz	ATLAS/BABAR	LPNHE Paris (France)
Jean Orloff	Theory	LPC Clermont-Ferrand (France)
Alejandro Perez	BABAR	IPHC Strasbourg (France)
Luis Pesantez	Belle/Belle II	Melbourne University (Australia)
Wenbin Qian	LHCb	Warwick University (UK)
Vincent Tisserand	LHCb/BABAR	LAPP Annecy-Le-Vieux (France)
Karim Trabelsi	Belle/Belle II	KEK Tsukuba (Japan)
Philip Urquijo	Belle/Belle II	Melbourne University (Australia)
Luiz Vale Silva	Theory	IJS Ljubljana (Slovenia)

Different significances

$$(a_{muon}^{exp} - a_{muon}^{SM}) \times 10^{11} = 288 \pm 63_{exp} \pm 49_{SM}$$

significance of the tension	
nG	$3.6\,\sigma$
1-external/Rfit	$3.8\,\sigma$
1-nuisance	$3.9\,\sigma$
adapt. nuisance	$2.7\,\sigma$

CKMfitter as a numerical tool

Package and computational resources:

- Modular structure: set of files representing different observables $\mathcal{O}$ in a given model (SM or specific NP) in terms of CKM matrix elements, QCD inputs, etc.;

  $+$ experimental data files, etc.
- $\mathcal{O}$ may have a non-linear dependence on the parameters we want to extract; the numerical step (extremization, etc.) is facilitated by calculating the symbolic expressions for the derivatives
- Both steps may be computationally very demanding and may require the use of a cluster (while simple fits can easily be done in a terminal)

CKMlive

[A. Claude and J. Charles, S. Descotes-G., S. Monteil]

▸ CKMlive

- Run dedicated CKM fits from **CKMfitter** package through a web interface
- Global fit (in the SM scenario at this moment) for the extraction of V_{CKM}
- Given set of observables in terms of a given set of parameters
- User chooses the set of observables, and the values of the theoretical and experimental inputs, plus fitting parameters