A model-independent likelihood function for the Belle II $B^+ \to K^+ \nu\bar{\nu}$ analysis

Lorenz Gärtner[1,2] *on behalf of Belle II,*

in collaboration with

Danny van Dyk[3], Lukas Heinrich[2,4], Méril Reboud[3]

[1]LMU Munich, [2]Excellence Cluster ORIGINS,
[3]IPPP Durham, [4]TU Munich

26.10.2023

Reusability in HEP

Publishing statistical models: Getting the most out of particle physics experiments

Kyle Cranmer[1*], Sabine Kraml[2‡], Harrison B. Prosper[3§] (editors), Philip Bechtle[4], Florian U. Bernlochner[4], Itay M. Bloch[5], Enzo Canonero[6], Marcin Chrzaszcz[7], Andrea Coccaro[8], Jan Conrad[9], Glen Cowan[10], Matthew Feickert[11], Nahuel Ferreiro Iachellini[12,13] Andrew Fowlie[14], Lukas Heinrich[15], Alexander Held[1], Thomas Kuhr[13,16], Anders Kvellestad[17], Maeve Madigan[18], Farvah Mahmoudi[15,19], Knut Dundas Morå[20], Mark S. Neubauer[11], Maurizio Pierini[15], Juan Rojo[8], Sezen Sekmen[22], Luca Silvestrini[23], Veronica Sanz[24,25], Giordon Stark[26], Riccardo Torre[8], Robert Thorne[27], Wolfgang Waltenberger[28], Nicholas Wardle[29], Jonas Wittbrodt[30]

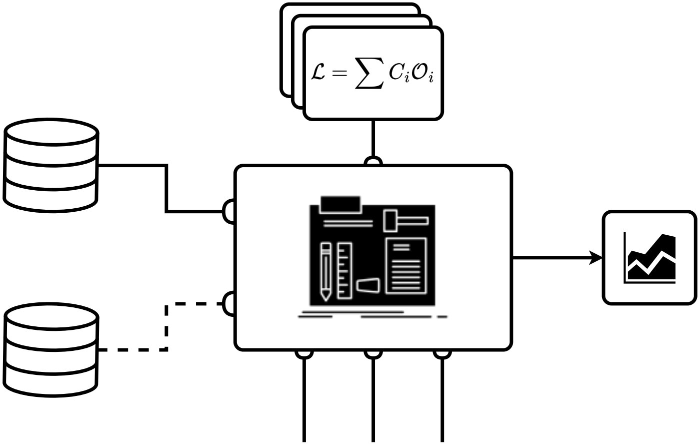
Forum [8], with the current status and updated recommendations presented in Ref. [6]. This paper takes these decade-long efforts to what we argue is the logical conclusion: if we wish to maximize the scientific impact of particle physics experiments, decades into the future, we should make the publication of full statistical models, together with the data to convert them into likelihood functions, standard practice. A statistical model provides the complete mathematical description of an experimental analysis and is, therefore, the appropriate starting

[arXiv:2109.04981 [hep-ph]]

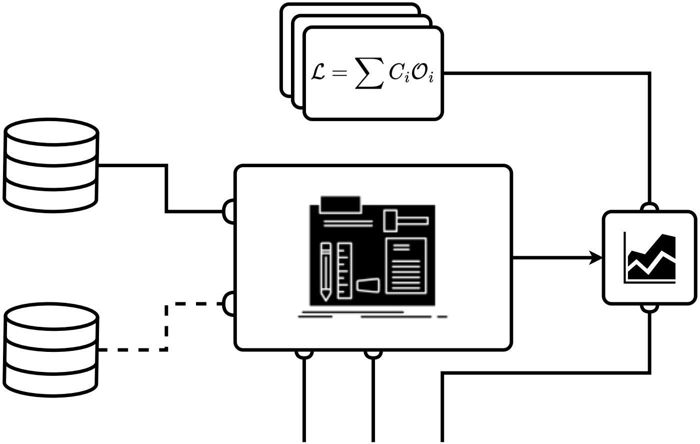
Shortcomings of model-dependencies in analyses

- **Limited interpretability** in terms of any model with different kinematic predictions.

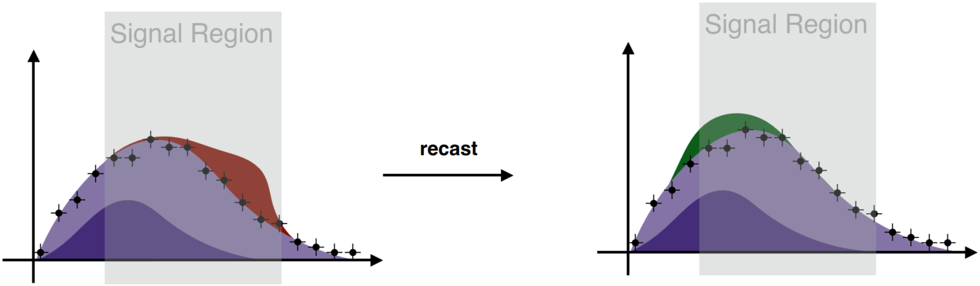
A general analysis



A better analysis



Idea behind reinterpretation



original analysis (w.r.t **model A**)   original analysis (recast to **model B**)

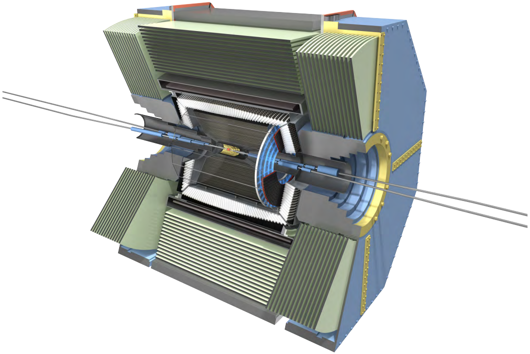
[source]

Analysis

Where is the model dependence?

SuperKEKB & Belle II introduction

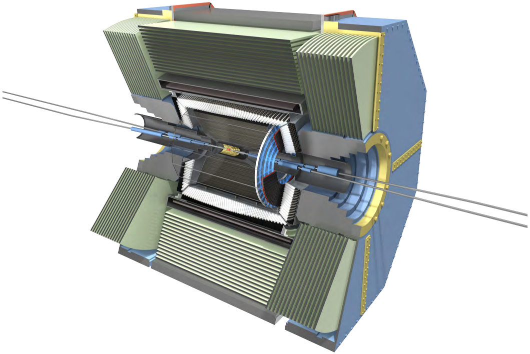
- **Luminosity** vs. energy frontier
  - Current total $\int L\, dt = 428fb^{-1}$



arXiv:1808.10567 [hep-ex]

SuperKEKB & Belle II introduction

- **Luminosity** vs. energy frontier
  - Current total $\int L \, dt = 428fb^{-1}$
- **SuperKEKB**
  - asymmetric e^{-}(7 GeV) $-$ e^{+}(4 GeV)
- **Belle II**
  - Hermetic, longitudinally asymmetric detector

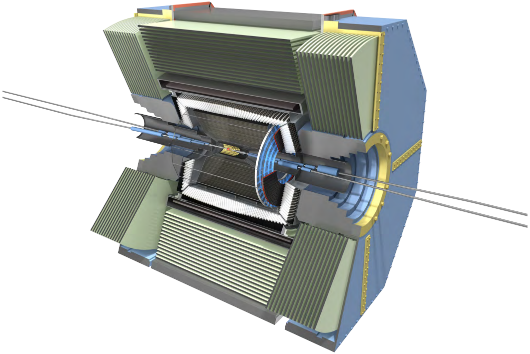


arXiv:1808.10567 [hep-ex]

SuperKEKB & Belle II introduction

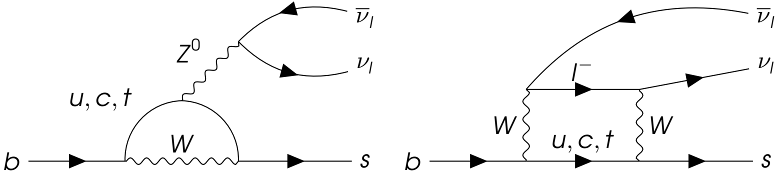
- **Luminosity** vs. energy frontier
  - Current total $\int L\,dt = 428 fb^{-1}$
- **SuperKEKB**
  - asymmetric e^{-}(7 GeV) $-$ e^{+}(4 GeV)
- **Belle II**
  - Hermetic, longitudinally asymmetric detector

➜ Missing mass analyses possible



arXiv:1808.10567 [hep-ex]

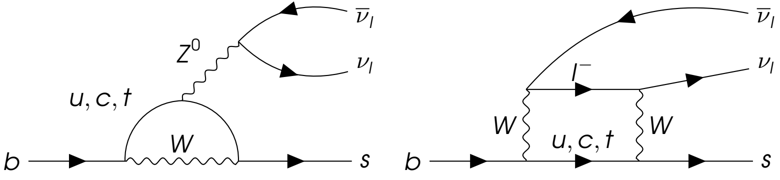
Why reinterpret $B^+ \to K^+ \nu\bar{\nu}$?



- Suppression of FCNCs in the SM.

➜ Tree level BSM effects could substantially affect observables.

Why reinterpret $B^+ \to K^+ \nu\bar{\nu}$?



- Suppression of FCNCs in the SM.

➜ Tree level BSM effects could substantially affect observables.

- **Benefits of reinterpretation**
  - **Sensitivity** to any current or future (B)SM prediction.
  - **Exclusion** limits in BSM parameter space inferable.
  - **Combinations** with other measurements possible.

The Belle II $B^+ \to K^+ \nu \bar{\nu}$ analysis

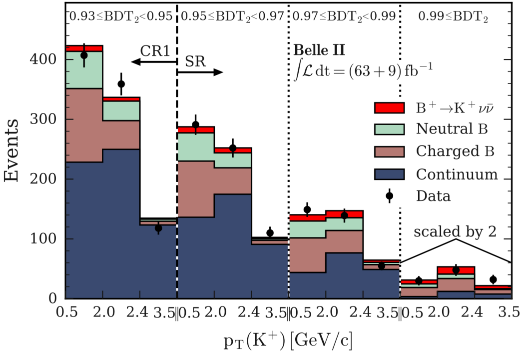
1. Machine learning methods ($BDT_1 + BDT_2$): separate signal from background.
2. Signal MC weighted according to SM kinematic prediction $\to$ **model dependence**
3. Max. likelihood fit in bins of $p_T(K^+) \times BDT_2$.

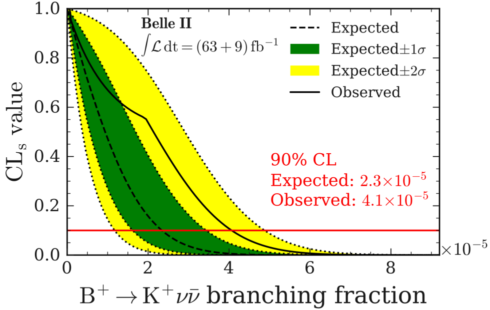
[hepdata.130199]
[Phys.Rev.Lett.127.181802]

The Belle II $B^+ \to K^+ \nu\bar{\nu}$ analysis

1. Machine learning methods ($BDT_1 + BDT_2$): separate signal from background.
2. Signal MC weighted according to SM kinematic prediction $\to$ **model dependence**
3. Max. likelihood fit in bins of $p_T(K^+) \times BDT_2$.

[hepdata.130199]
[Phys.Rev.Lett.127.181802]





$\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) < 4.1 \times 10^{-5}$@90%$CL$

Reweighting approach

How do we obtain new signal templates?

Reweighting recipe

Recipe

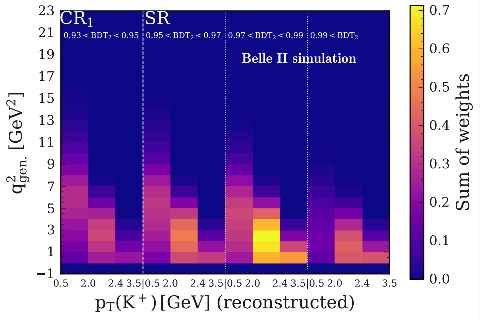
1. Find kinematic dependence of measured observable: $\Gamma(q^2)$, $q^2 = (p_\nu + p_{\bar{\nu}})^2$

Reweighting recipe

Recipe

1. Find kinematic dependence of measured observable: $\Gamma(q^2)$, $q^2 = (p_\nu + p_{\bar{\nu}})^2$
2. Get distributions of kinematic d.o.f (q^2)

$$N_{k\,l\,m} = \underbrace{p_T \times BDT_2}_{\text{analysis binning}} \times \underbrace{q^2_{gen.}}_{\text{kinematic d.o.f}}$$



Reweighting approach

Recipe

1. Find kinematic dependence of theoretical prediction: $\Gamma(q^2)$, $q^2 = (p_\nu + p_{\bar{\nu}})^2$
2. Get distributions of kinematic d.o.f (q^2)

$$N_{k\,l\,m} = \underbrace{p_T \times BDT_2}_{\text{analysis binning}} \times \underbrace{q^2_{gen.}}_{\text{kinematic d.o.f}}$$

3. Apply weights in bins of kinematic d.o.f. (phase space (PHSP): $\mathcal{M} = 1$)

$$N_{k\,l} = \sum_{m \in q^2} N^{\text{PHSP}}_{k\,l\,m} w_m = \sum_{m \in q^2} N^{\text{PHSP}}_{k\,l\,m} \int_{\text{bin } m} dq^2 \frac{d\Gamma^{(B)SM}}{dq^2} \left(\frac{d\Gamma^{\text{PHSP}}}{dq^2} \right)^{-1}$$

Reweighting approach

Recipe

1. Find kinematic dependence of theoretical prediction: $\Gamma(q^2)$, $q^2 = (p_\nu + p_{\bar{\nu}})^2$
2. Get distributions of kinematic d.o.f (q^2)

$$N_{k\,l\,m} = \underbrace{p_T \times BDT_2}_{\text{analysis binning}} \times \underbrace{q^2_{gen.}}_{\text{kinematic d.o.f}}$$

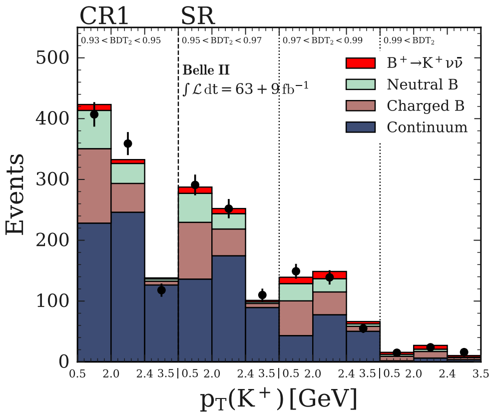
3. Apply weights in bins of kinematic d.o.f. (phase space (PHSP): $\mathcal{M} = 1$)

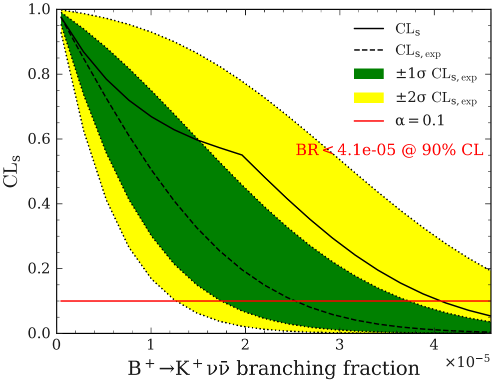
$$N_{k\,l} = \sum_{m \in q^2} N^{\text{PHSP}}_{k\,l\,m}\, w_m = \sum_{m \in q^2} N^{\text{PHSP}}_{k\,l\,m} \int_{\text{bin } m} dq^2 \frac{d\Gamma^{(B)SM}}{dq^2} \left(\frac{d\Gamma^{\text{PHSP}}}{dq^2} \right)^{-1}$$

Benefits

+ Fast
+ Versatile
+ Easily publishable

Validation: Reproducing the upper limit





$\mathcal{B}(B^+ \to K^+ \nu\bar{\nu}) < 4.1 \times 10^{-5}$@90%CL

pyhf.readthedocs.io

Theory

How can we parametrize our model dependence?

Weak Effective Theory for $B \to K\nu\bar{\nu}$

Contribution operators

The effective Lagrangian is

$$\mathcal{L}^{WET} = \sum_{X=L,R} C_{\mathrm{VX}} \, \mathcal{O}_{\mathrm{VX}} + \sum_{X=L,R} C_{\mathrm{SX}} \, \mathcal{O}_{\mathrm{SX}} + C_{\mathrm{TL}} \, \mathcal{O}_{\mathrm{TL}} + \text{h.c.}$$

The $d = 6$ contributing operators in and beyond the SM are given by

$$\mathcal{O}_{\mathrm{VL}} = (\overline{\nu_L}\gamma_\mu \nu_L)\,(\overline{s_L}\gamma^\mu b_L) \qquad \mathcal{O}_{\mathrm{VR}} = (\overline{\nu_L}\gamma_\mu \nu_L)\,(\overline{s_R}\gamma^\mu b_R)$$

$$\mathcal{O}_{\mathrm{SL}} = \left(\overline{\nu_L^c}\,\nu_L\right)(\overline{s_R} b_L) \qquad \mathcal{O}_{\mathrm{SR}} = \left(\overline{\nu_L^c}\,\nu_L\right)(\overline{s_L} b_R)$$

$$\mathcal{O}_{\mathrm{TL}} = \left(\overline{\nu_L^c}\,\sigma_{\mu\nu}\nu_L\right)(\overline{s_R}\sigma^{\mu\nu} b_L)$$

[arXiv:2111.04327 [hep-ph]]

(B)SM theory predictions

Capture BSM physics above electroweak symmetry breaking scale with the Wilson coefficients

$C_{SL} + C_{SR}$ $\quad$ $C_{VL} + C_{VR} = C_{VL}^{SM} + C_{VL}^{NP} + C_{VR}$ $\quad$ C_{TL}

EOS

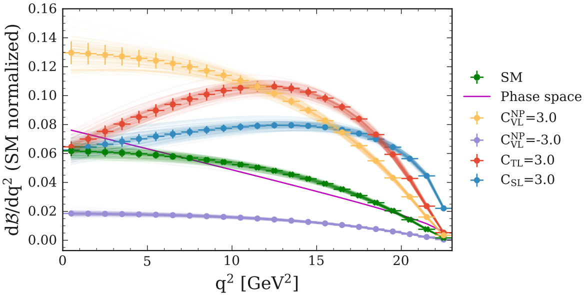
eos.github.io

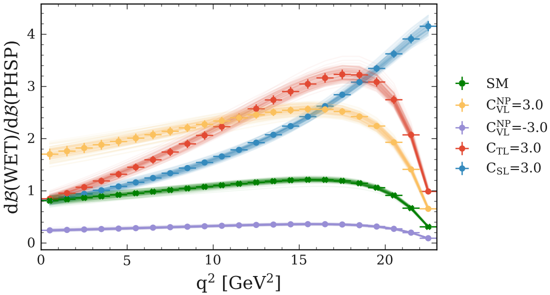
(B)SM theory predictions

Capture BSM physics above electroweak symmetry breaking scale with the Wilson coefficients

$C_{SL} + C_{SR}$ $\quad$ $C_{VL} + C_{VR} = C_{VL}^{SM} + C_{VL}^{NP} + C_{VR}$ $\quad$ C_{TL}

eos.github.io





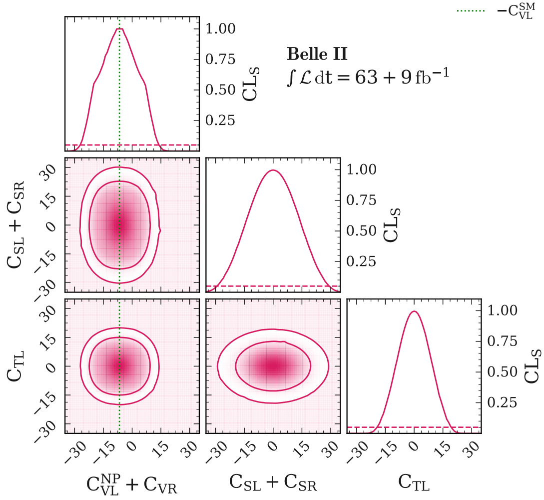
Fruits of reinterpretation

What do we get from all this?

Wilson coefficient exclusion limits

- pyhf (differentiable Likelihoods) + EOS → fast & reliable
- Exclusion limits @95%CL

$$|C_{VL}^{SM} + C_{VL}^{NP} + C_{VR}| < 20.6$$
$$|C_{SL} + C_{SR}| < 29.3$$
$$|C_{TL}| < 19.4$$



Analysis update: $B^+ \to K^+ \nu\bar{\nu}$ @ 362 fb^{-1}

Inclusive $\mathcal{B} = \left(2.8^{+0.5}_{-0.5}{}^{+0.5}_{-0.5}\right) \times 10^{-5}$

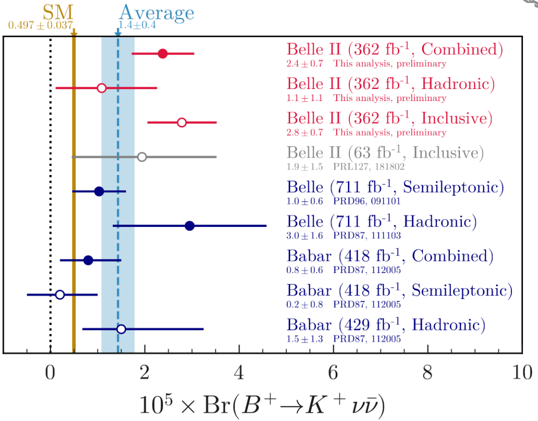
Hadronic $\mathcal{B} = \left(1.1^{+0.9}_{-0.8}{}^{+0.8}_{-0.5}\right) \times 10^{-5}$

Combined $\mathcal{B} = \left(2.4^{+0.5}_{-0.5}{}^{+0.5}_{-0.4}\right) \times 10^{-5}$

Significance of the **combined** result:

- 3.6σ wrt. null hypothesis
- 2.8σ wrt. SM

First evidence of $B^+ \to K^+ \nu\bar{\nu}$



Model-independent likelihood method will be applied and published once paper is accepted.

Presented at EPS 2023

Summary

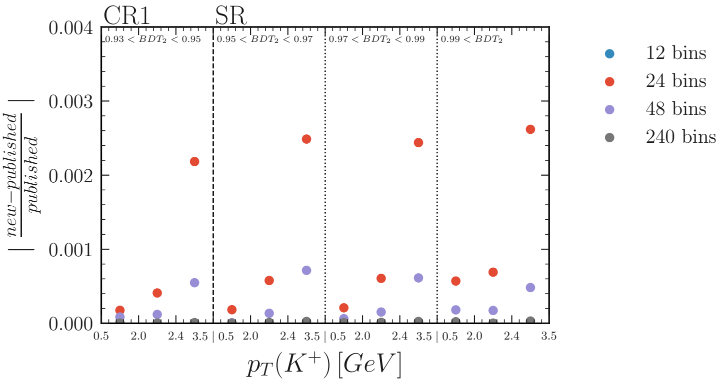
- **Challenge**
  - Neutrino-induced experimental complexities
  - ➜ **model-dependent results**
- **Solution**
  - **Model-independent likelihood function**
    Maximum likelihood fits for any given (B)SM signal prediction.
- **Tool integration (more in backup)**
  - Extend `pyhf` and interface it with `EOS` for run-time template updating.
  - Method fully applicable to other decay channels and results.
- **Scientific benefits**
  - **Exclusions in BSM parameter space.**
  - **Combinations** with other channels and/or experiments.
  - ...

Publishing such likelihoods is crucial for a full exploitation of experimental results.

lorenz.gaertner@physik.uni-muenchen.de

Binning choice

We compared the relative accuracy of the binned weighting (new) with the event-by-event weighting (published).

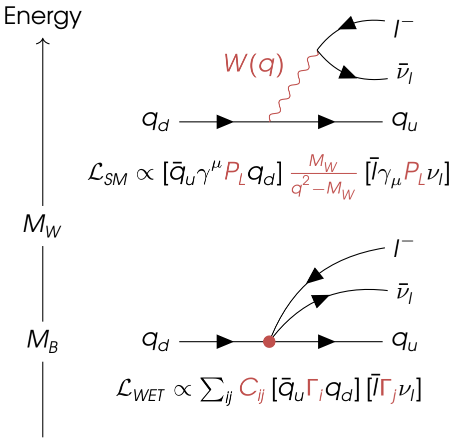


Effective field theory

- **High energy collisions**
  Enough energy to radiate off an on-shell (massive) W boson.
- **Lower energy quark decays**
  W boson is always off-shell.
- **Weak effective theory**
  W is integrated out – effects are encoded in new couplings

$$\mathcal{L}_{SM} \to \mathcal{L}_{WET} = \sum C_i \, \mathcal{O}_i$$

➜ Model independent **parametrization**, constrained only by Wilson coefficients C_i.



Weak Effective Theory for $B \to K\nu\bar{\nu}$

Decay width

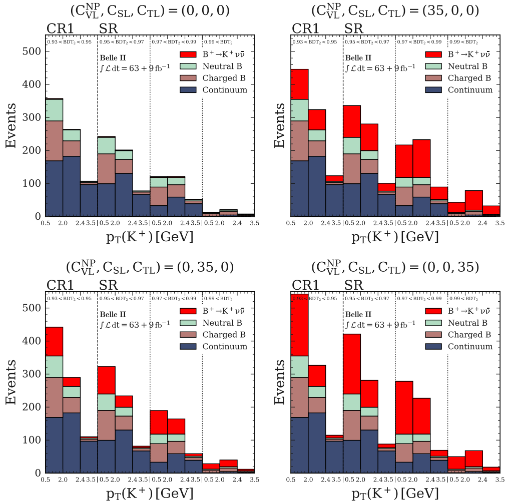
Decay width dependence on the Wilson coefficients is given by

$$\frac{d\Gamma(B \to K\nu\bar{\nu})}{dq^2} = \frac{\sqrt{\lambda_{BK}}q^2}{(4\pi)^3 m_B^3}\Bigg[\frac{\lambda_{BK}}{24q^2}\left|f_+(q^2)\right|^2 \left|C_{\mathrm{VL}} + C_{\mathrm{VR}}\right|^2 + \frac{\left(m_B^2 - m_K^2\right)^2}{8\left(m_b - m_s\right)^2}\left|f_0(q^2)\right|^2 \left|C_{\mathrm{SL}} + C_{\mathrm{SR}}\right|^2 + \frac{2\lambda_{BK}}{3\left(m_B + m_K\right)^2}\left|f_T(q^2)\right|^2 \left|C_{\mathrm{TL}}\right|^2\Bigg]$$

valid for $J^P = 0^-$ kaon states.

[arXiv:2111.04327 [hep-ph]]

Expected yields

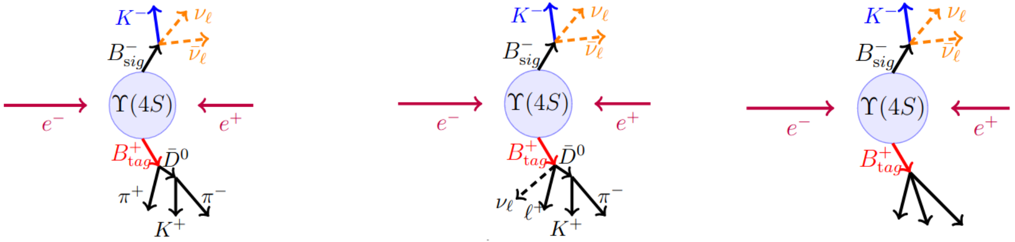


Reconstruction techniques

Efficiency

$\epsilon \sim 0.1 - 1\,\%$ | $\epsilon \sim 1 - 3\,\%$ | $\epsilon \sim 1 - 100\,\%$

Exclusive hadronic | Exclusive semileptonic | Inclusive



Purity, Resolution

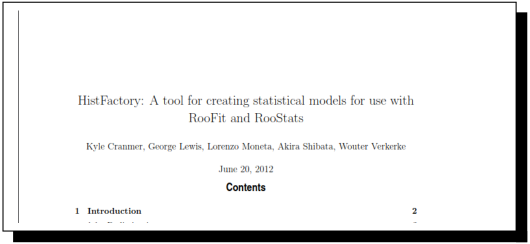
Different reconstruction techniques lead to nearly orthogonal data samples

Implementation

How do we realize this?

A statistical model for multi-bin histogram-based analysis and its interval estimation.

pyhf = **py**thonic **H**ist**F**actory


HistFactory: A tool for creating statistical models for use with RooFit and RooStats

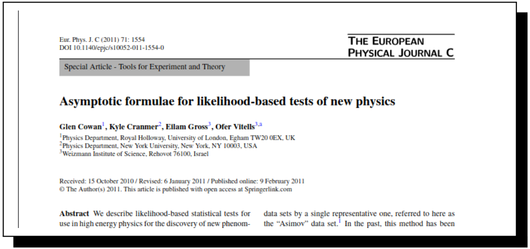
Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, Wouter Verkerke

June 20, 2012

Contents

1 **Introduction** 2

Interval estimation based on


Eur. Phys. J. C (2011) 71: 1554
DOI 10.1140/epjc/s10052-011-1554-0

THE EUROPEAN PHYSICAL JOURNAL C

Special Article - Tools for Experiment and Theory

Asymptotic formulae for likelihood-based tests of new physics

Glen Cowan[1], **Kyle Cranmer**[2], **Eilam Gross**[3], **Ofer Vitells**[3,a]
[1]Physics Department, Royal Holloway, University of London, Egham TW20 0EX, UK
[2]Physics Department, New York University, New York, NY 10003, USA
[3]Weizmann Institute of Science, Rehovot 76100, Israel

Received: 15 October 2010 / Revised: 6 January 2011 / Published online: 9 February 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract We describe likelihood-based statistical tests for use in high energy physics for the discovery of new phenom-

data sets by a single representative one, referred to here as the "Asimov" data set.[1] In the past, this method has been

HistFactory model

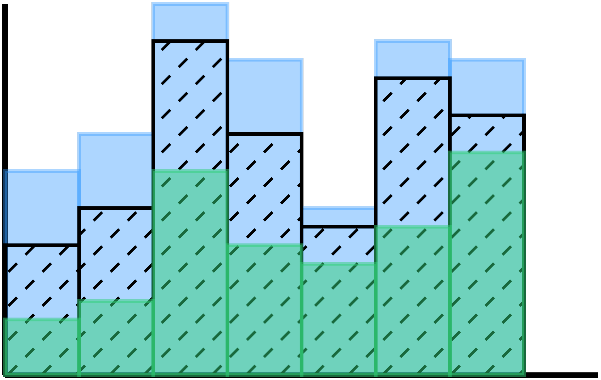
Likelihood function for observed event counts $\boldsymbol{n}$ is

$$L(\boldsymbol{n}, \boldsymbol{a} \mid \boldsymbol{\eta}, \boldsymbol{\chi}) = \underbrace{\prod_{c \in \text{channels}} \prod_{b \in \text{bins}} \text{Pois}\left(n_{cb} \mid \nu_{cb}(\boldsymbol{\eta}, \boldsymbol{\chi})\right)}_{\text{multiple channels}} \underbrace{\prod_{\chi \in \boldsymbol{\chi}} c_{\chi}\left(a_{\chi} \mid \chi\right)}_{\text{constraint terms}}$$

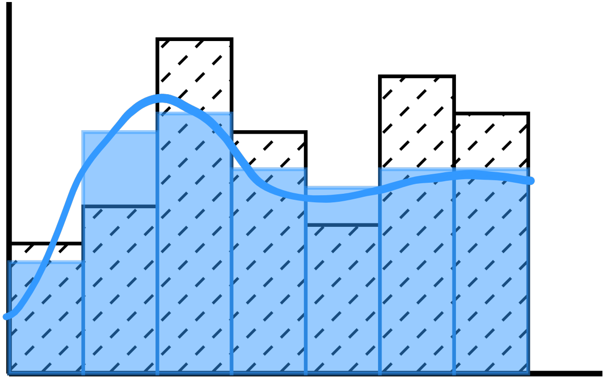
Expected number of events per channel per bin are

$$\nu_{cb}(\boldsymbol{\eta}, \boldsymbol{\chi}) = \sum_{s \in \text{samples}} \underbrace{\prod_{\kappa \in \boldsymbol{\kappa}} \kappa_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi})}_{\text{multiplicative modifiers}} \left(\nu^{0}_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi}) + \underbrace{\sum_{\Delta \in \boldsymbol{\Delta}} \Delta_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi})}_{\text{additive modifiers}}\right).$$

Modifiers



Custom modifiers



Modifiers and constraints

Description	Modification	Constraint Term c_χ	Input
Uncorrelated Shape	$\kappa_{scb}(\gamma_b) = \gamma_b$	$\prod_b \mathrm{Pois}\left(r_b = \sigma_b^{-2} \middle| \rho_b = \sigma_b^{-2}\gamma_b\right)$	σ_b
Correlated Shape	$\Delta_{scb}(\alpha) = f_p\left(\alpha \middle| \Delta_{scb,\alpha=-1}, \Delta_{scb,\alpha=1}\right)$	$\mathrm{Gaus}\left(a = 0 \middle| \alpha, \sigma = 1\right)$	$\Delta_{scb,\alpha=\pm1}$
Normalisation Unc.	$\kappa_{scb}(\alpha) = g_p\left(\alpha \middle| \kappa_{scb,\alpha=-1}, \kappa_{scb,\alpha=1}\right)$	$\mathrm{Gaus}\left(a = 0 \middle| \alpha, \sigma = 1\right)$	$\kappa_{scb,\alpha=\pm1}$
MC Stat. Uncertainty	$\kappa_{scb}(\gamma_b) = \gamma_b$	$\prod_b \mathrm{Gaus}\left(a_{\gamma_b} = 1 \middle| \gamma_b, \delta_b\right)$	$\delta_b^2 = \sum_s \delta_{sb}^2$
Luminosity	$\kappa_{scb}(\lambda) = \lambda$	$\mathrm{Gaus}\left(l = \lambda_0 \middle| \lambda, \sigma_\lambda\right)$	$\lambda_0, \sigma_\lambda$
Normalisation	$\kappa_{scb}(\mu_b) = \mu_b$		
Data-driven Shape	$\kappa_{scb}(\gamma_b) = \gamma_b$		