Beyond-the-standard-model contributions to rare B decays analyzed with variational-Bayes enhanced adaptive importance sampling

Stephan Jahn

March 17, 2015

Overview

Bayes' formula:

$$P(\boldsymbol{\theta}|\mathcal{D},\mathrm{M}) = \frac{P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})}{P(\mathcal{D}|\mathrm{M})} = \frac{P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})}{\int P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})d\boldsymbol{\theta}}$$

Overview

Bayes' formula:

$$P(\boldsymbol{\theta}|\mathcal{D},\mathrm{M}) = \frac{P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})}{P(\mathcal{D}|\mathrm{M})} = \frac{P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})}{\int P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})d\boldsymbol{\theta}}$$

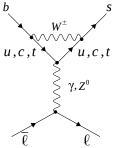
model independent search for new physics (effective theory):

Overview

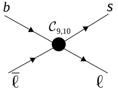
Bayes' formula:

$$P(\boldsymbol{\theta}|\mathcal{D},\mathrm{M}) = \frac{P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})}{P(\mathcal{D}|\mathrm{M})} = \frac{P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})}{\int P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})d\boldsymbol{\theta}}$$

model independent search for new physics (effective theory):



$\boldsymbol{\theta} = _{\text{effective}}\ \text{couplings}\ \mathcal{C}_i$

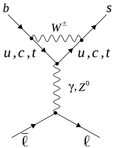


Overview

Bayes' formula:

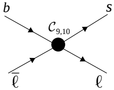
$$P(\boldsymbol{\theta}|\mathcal{D},\mathrm{M}) = \frac{P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})}{P(\mathcal{D}|\mathrm{M})} = \frac{P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})}{\int P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})d\boldsymbol{\theta}}$$

model independent search for new physics (effective theory):



$\boldsymbol{\theta}$ = effective couplings $\mathcal{C}_i$

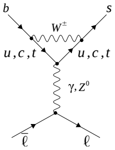
$\mathcal{D}$ = detector events



Bayes' formula:

$$P(\boldsymbol{\theta}|\mathcal{D},\mathrm{M}) = \frac{P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})}{P(\mathcal{D}|\mathrm{M})} = \frac{P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})}{\int P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})d\boldsymbol{\theta}}$$

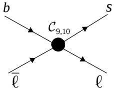
model independent search for new physics (effective theory):



$$\boldsymbol{\theta} = \text{effective couplings } \mathcal{C}_i$$

$$\mathcal{D} = \text{detector events}$$

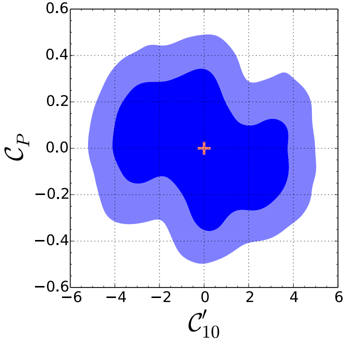
$$\mathrm{M} = \mathrm{EFT}, \mathrm{SM}, \ldots$$



Goals

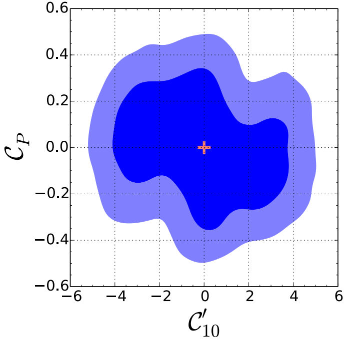
Goals

- draw marginal plots of the posterior



Goals

- draw marginal plots of the posterior



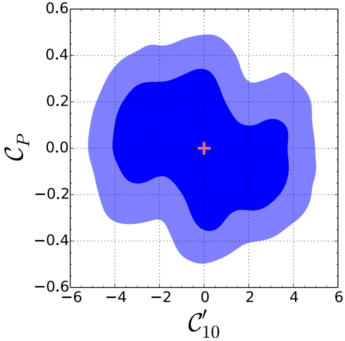
- compare models
  NP $\leftrightarrow$ SM

$$\frac{P(\mathrm{NP}|\mathcal{D})}{P(\mathrm{SM}|\mathcal{D})} = \frac{P(\mathcal{D}|\mathrm{NP})}{P(\mathcal{D}|\mathrm{SM})} \cdot \frac{P(\mathrm{NP})}{P(\mathrm{SM})}$$

$$P(\mathrm{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathrm{M})P(\mathrm{M})}{P(\mathcal{D})}$$

Goals

- draw marginal plots of the posterior



- compare models
  NP $\leftrightarrow$ SM

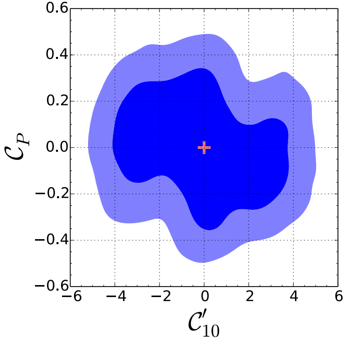
$$\frac{P(\mathrm{NP}|\mathcal{D})}{P(\mathrm{SM}|\mathcal{D})} = \frac{P(\mathcal{D}|\mathrm{NP})}{P(\mathcal{D}|\mathrm{SM})} \cdot \frac{P(\mathrm{NP})}{P(\mathrm{SM})}$$

$$P(\mathrm{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathrm{M})P(\mathrm{M})}{P(\mathcal{D})}$$

$\frac{P(\mathrm{NP}|\mathcal{D})}{P(\mathrm{SM}|\mathcal{D})} > 1$ new physics ☺

Goals

- draw marginal plots of the posterior



- compare models
  $\mathrm{NP} \leftrightarrow \mathrm{SM}$

$$\frac{P(\mathrm{NP}|\mathcal{D})}{P(\mathrm{SM}|\mathcal{D})} = \frac{P(\mathcal{D}|\mathrm{NP})}{P(\mathcal{D}|\mathrm{SM})} \cdot \frac{P(\mathrm{NP})}{P(\mathrm{SM})}$$

$$P(\mathrm{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathrm{M})P(\mathrm{M})}{P(\mathcal{D})}$$

$\frac{P(\mathrm{NP}|\mathcal{D})}{P(\mathrm{SM}|\mathcal{D})} > 1$ new physics ☺

$\frac{P(\mathrm{NP}|\mathcal{D})}{P(\mathrm{SM}|\mathcal{D})} < 1$ confirm SM ☹

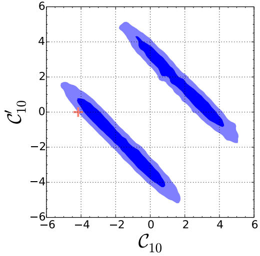
Difficulties

Difficulties

- curse of dimensionality

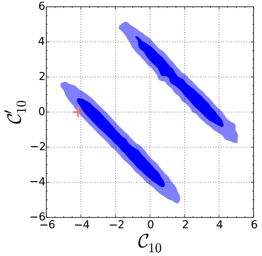
Difficulties

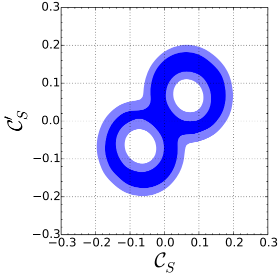
- curse of dimensionality
- multimodality



Difficulties

- curse of dimensionality
- multimodality
- degeneracies

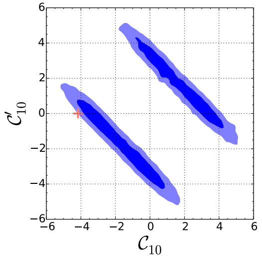


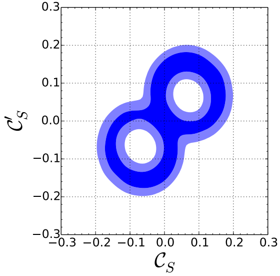


Difficulties

- curse of dimensionality
- multimodality
- degeneracies

no standard algorithm so far





Contents

Adaptive importance sampling

$$\int P(x)\mathrm{d}x = \int \frac{P(x)}{p(x)}p(x)\mathrm{d}x \approx \frac{1}{N}\sum_{n=1}^{N}\frac{P(x_n)}{p(x_n)} \equiv \mu^N \;\; \textit{where} \;\; x_n \sim p$$

Adaptive importance sampling

$$\int P(x)\mathrm{d}x = \int \frac{P(x)}{p(x)} p(x)\mathrm{d}x \approx \frac{1}{N}\sum_{n=1}^{N} \frac{P(x_n)}{p(x_n)} \equiv \mu^N \ \ where \ \ x_n \sim p$$

squared uncertainty (variance):

$$var(\mu^N) = \frac{1}{N}\left[\int \frac{P(x)}{p(x)} P(x)\mathrm{d}x - \left(\int P(x)\mathrm{d}x\right)^2\right]$$

Adaptive importance sampling

$$\int P(x)\mathrm{d}x = \int \frac{P(x)}{p(x)} p(x)\mathrm{d}x \approx \frac{1}{N}\sum_{n=1}^{N} \frac{P(x_n)}{p(x_n)} \equiv \mu^N \quad \textit{where} \quad x_n \sim p$$

squared uncertainty (variance):

$$var(\mu^N) = \frac{1}{N}\left[\int \frac{P(x)}{p(x)} P(x)\mathrm{d}x - \left(\int P(x)\mathrm{d}x\right)^2\right]$$

- minimize the uncertainty $var(\mu^N)$ with respect to p

Adaptive importance sampling

$$\int P(x)\mathrm{d}x = \int \frac{P(x)}{p(x)} p(x)\mathrm{d}x \approx \frac{1}{N}\sum_{n=1}^{N} \frac{P(x_n)}{p(x_n)} \equiv \mu^N \quad \textit{where} \quad x_n \sim p$$

squared uncertainty (variance):

$$var(\mu^N) = \frac{1}{N}\left[\int \frac{P(x)}{p(x)} P(x)\mathrm{d}x - \left(\int P(x)\mathrm{d}x\right)^2\right]$$

- minimize the uncertainty $var(\mu^N)$ with respect to p
  - it suffices to minimize: $\log\left(\int \frac{P(x)}{p(x)} P(x)\mathrm{d}x\right)$

Adaptive importance sampling

$$\int P(x)\mathrm{d}x = \int \frac{P(x)}{p(x)} p(x)\mathrm{d}x \approx \frac{1}{N}\sum_{n=1}^{N} \frac{P(x_n)}{p(x_n)} \equiv \mu^N \quad \textit{where} \quad x_n \sim p$$

squared uncertainty (variance):

$$var(\mu^N) = \frac{1}{N}\left[\int \frac{P(x)}{p(x)} P(x)\mathrm{d}x - \left(\int P(x)\mathrm{d}x\right)^2\right]$$

- minimize the uncertainty $var(\mu^N)$ with respect to p
  - it suffices to minimize: $\log\left(\int \frac{P(x)}{p(x)} P(x)\mathrm{d}x\right)$
  - by Jensen's inequality: $\geq \int \log\left(\frac{P(x)}{p(x)}\right) P(x)\mathrm{d}x$

Adaptive importance sampling

$$\int P(x)\mathrm{d}x = \int \frac{P(x)}{p(x)} p(x)\mathrm{d}x \approx \frac{1}{N} \sum_{n=1}^{N} \frac{P(x_n)}{p(x_n)} \equiv \mu^N \quad \textit{where} \quad x_n \sim p$$

squared uncertainty (variance):

$$var(\mu^N) = \frac{1}{N} \left[\int \frac{P(x)}{p(x)} P(x)\mathrm{d}x - \left(\int P(x)\mathrm{d}x \right)^2 \right]$$

- minimize the uncertainty $var(\mu^N)$ with respect to p
  - it suffices to minimize: $\log \left(\int \frac{P(x)}{p(x)} P(x)\mathrm{d}x \right)$
  - by Jensen's inequality: $\geq \int \log \left(\frac{P(x)}{p(x)} \right) P(x)\mathrm{d}x \equiv KL(P \| p)$
  - known as the *Kullback-Leibler divergence*

Adaptive importance sampling

Note:

Note:

$$0 \leq KL(P\|p) \quad \text{and} \quad KL(P\|p) = 0 \quad \Leftrightarrow \quad P = p$$

Adaptive importance sampling

Note:

$$0 \leq KL(P\|p) \quad \text{and} \quad KL(P\|p) = 0 \quad \Leftrightarrow \quad P = p$$

$$0 \leq var(\mu^N) \quad \text{and} \quad var(\mu^N) = 0 \quad \Leftrightarrow \quad P = p$$

Adaptive importance sampling

Note:

$$0 \leq KL(P \| p) \quad \text{and} \quad KL(P \| p) = 0 \quad \Leftrightarrow \quad P = p$$

$$0 \leq var(\mu^N) \quad \text{and} \quad var(\mu^N) = 0 \quad \Leftrightarrow \quad P = p$$

Remember:

$$var(\mu^N) - const = \log \left(\int \frac{P(x)}{p(x)} P(x) \mathrm{d}x \right) \geq KL(P \| p)$$

Adaptive importance sampling

Note:

$$0 \le KL(P\|p) \quad \text{and} \quad KL(P\|p) = 0 \quad \Leftrightarrow \quad P = p$$

$$0 \le var(\mu^N) \quad \text{and} \quad var(\mu^N) = 0 \quad \Leftrightarrow \quad P = p$$

Remember:

$$var(\mu^N) - const = \underbrace{\log\left(\int \frac{P(x)}{p(x)} P(x) \mathrm{d}x\right)}_{\text{minimization typically infeasible}} \ge KL(P\|p)$$

Adaptive importance sampling

Note:

$$0 \leq KL(P\|p) \quad \text{and} \quad KL(P\|p) = 0 \quad \Leftrightarrow \quad P = p$$

$$0 \leq var(\mu^N) \quad \text{and} \quad var(\mu^N) = 0 \quad \Leftrightarrow \quad P = p$$

Remember:

$$var(\mu^N) - const = \underbrace{\log\left(\int \frac{P(x)}{p(x)} P(x)\mathrm{d}x\right)}_{\text{minimization typically infeasible}} \geq KL(P\|p)$$

minimize $KL(P\|p)$ and hope to approach the unique global minimum $P = p$

restrict p to Gaussian mixtures

$$p(x_n|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k), \quad \boldsymbol{\theta} = \{\pi_k, \mu_k, \Sigma_k\}$$

restrict p to Gaussian mixtures

$$p(x_n|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k), \quad \boldsymbol{\theta} = \{\pi_k, \mu_k, \Sigma_k\}$$

rewrite with *latent variables* $\mathbf{Z}$:

$$z_{nk} = \begin{cases} 1 \text{ if } x_n \text{ from component k} \\ 0 \text{ else} \end{cases}$$

restrict p to Gaussian mixtures

$$p(x_n|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k), \quad \boldsymbol{\theta} = \{\pi_k, \mu_k, \Sigma_k\}$$

rewrite with *latent variables* $\mathbf{Z}$:

$$z_{nk} = \begin{cases} 1 \text{ if } x_n \text{ from component k} \\ 0 \text{ else} \end{cases}$$

$$p(x_n|\mathbf{Z}, \boldsymbol{\theta}) = \prod_{k=1}^{K} \mathcal{N}(x_n|\mu_k, \Sigma_k)^{z_{nk}}$$

Variational Bayes (Gaussian mixture model)

restrict p to Gaussian mixtures

$$p(x_n|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k), \quad \boldsymbol{\theta} = \{\pi_k, \mu_k, \Sigma_k\}$$

rewrite with *latent variables* $\mathbf{Z}$:

$$z_{nk} = \begin{cases} 1 \text{ if } x_n \text{ from component k} \\ 0 \text{ else} \end{cases}$$

$$p(x_n|\mathbf{Z}, \boldsymbol{\theta}) = \prod_{k=1}^{K} \mathcal{N}(x_n|\mu_k, \Sigma_k)^{z_{nk}}$$

$$p(z_{nk} = 1|\boldsymbol{\theta}) = \pi_k = \prod_{k'=1}^{K} \pi_{k'}^{z_{nk'}}$$

restrict p to Gaussian mixtures

$$p(x_n|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k), \quad \boldsymbol{\theta} = \{\pi_k, \mu_k, \Sigma_k\}$$

rewrite with *latent variables* **Z**:

$$z_{nk} = \begin{cases} 1 \text{ if } x_n \text{ from component k} \\ 0 \text{ else} \end{cases}$$

$$p(x_n|\mathbf{Z}, \boldsymbol{\theta}) = \prod_{k=1}^{K} \mathcal{N}(x_n|\mu_k, \Sigma_k)^{z_{nk}}$$

$$p(x_n|\boldsymbol{\theta}) = \sum_{\mathbf{Z}} p(x_n|\mathbf{Z}, \boldsymbol{\theta}) p(\mathbf{Z}|\boldsymbol{\theta})$$

$$p(z_{nk} = 1|\boldsymbol{\theta}) = \pi_k = \prod_{k'=1}^{K} \pi_{k'}^{z_{nk'}}$$

minimize $KL(P \| p)$

$$\text{minimize } KL(P\|p) \Leftrightarrow \text{maximize } \log p(\mathbf{X}|\boldsymbol{\theta}) \equiv \log \prod_{n=1}^{N} p(x_n|\boldsymbol{\theta})$$

impose a *prior* $p(\boldsymbol{\theta})$ and
rewrite with arbitrary probability distribution $q(\mathbf{Z}, \boldsymbol{\theta})$

$$\text{minimize } KL(P\|p) \Leftrightarrow \text{maximize } \log p(\mathbf{X}|\boldsymbol{\theta}) \equiv \log \prod_{n=1}^{N} p(x_n|\boldsymbol{\theta})$$

impose a *prior* $p(\boldsymbol{\theta})$ and
rewrite with arbitrary probability distribution $q(\mathbf{Z}, \boldsymbol{\theta})$

$$\log p(\mathbf{X}) = \underbrace{\int \mathrm{d}\mathbf{Z}\,\mathrm{d}\boldsymbol{\theta}\, q(\mathbf{Z},\boldsymbol{\theta}) \log\left[\frac{p(\mathbf{X},\mathbf{Z},\boldsymbol{\theta})}{q(\mathbf{Z},\boldsymbol{\theta})}\right]}_{\equiv \mathcal{L}[q]} + \underbrace{\int \mathrm{d}\mathbf{Z}\,\mathrm{d}\boldsymbol{\theta}\, q(\mathbf{Z},\boldsymbol{\theta}) \log\left[\frac{q(\mathbf{Z},\boldsymbol{\theta})}{p(\mathbf{Z},\boldsymbol{\theta}|\mathbf{X})}\right]}_{\equiv KL(q\|p)}$$

$$\text{minimize } KL(P\|p) \Leftrightarrow \text{maximize } \log p(\mathbf{X}|\boldsymbol{\theta}) \equiv \log \prod_{n=1}^{N} p(x_n|\boldsymbol{\theta})$$

impose a *prior* $p(\boldsymbol{\theta})$ and
rewrite with arbitrary probability distribution $q(\mathbf{Z}, \boldsymbol{\theta})$

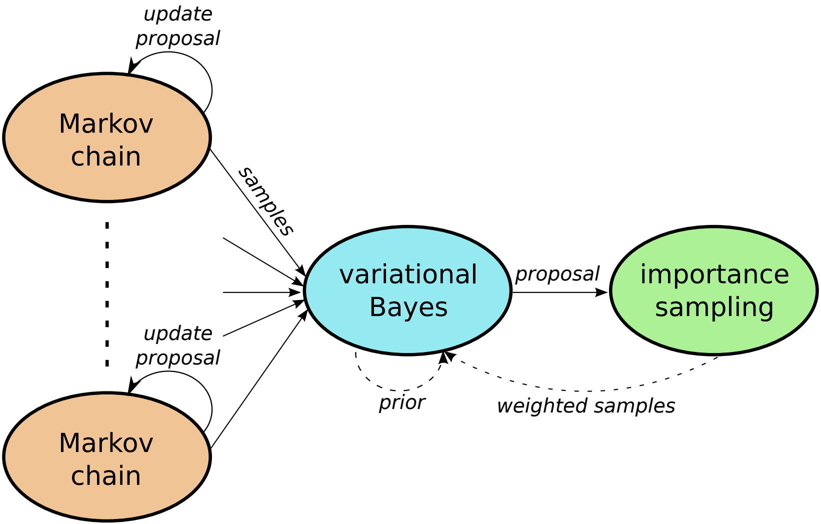
$$\log p(\mathbf{X}) = \underbrace{\int \mathrm{d}\mathbf{Z}\, \mathrm{d}\boldsymbol{\theta}\, q(\mathbf{Z}, \boldsymbol{\theta}) \log \left[\frac{p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta})}{q(\mathbf{Z}, \boldsymbol{\theta})}\right]}_{\equiv \mathcal{L}[q]} + \underbrace{\int \mathrm{d}\mathbf{Z}\, \mathrm{d}\boldsymbol{\theta}\, q(\mathbf{Z}, \boldsymbol{\theta}) \log \left[\frac{q(\mathbf{Z}, \boldsymbol{\theta})}{p(\mathbf{Z}, \boldsymbol{\theta}|\mathbf{X})}\right]}_{\equiv KL(q\|p)}$$

assumption $q(\mathbf{Z}, \boldsymbol{\theta}) = q(\mathbf{Z})q(\boldsymbol{\theta})$
$\Rightarrow$ iterative solution

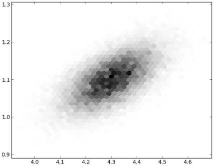
further reading:

Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer 2006, chapter 10

Adaptive importance sampling with the variational-Bayes approach



```python
'''This example illustrates how to run a Markov Chain using pypmc'''

import numpy as np
import pypmc

# define a proposal
prop_dof   = 50.
prop_sigma = np.array([[0.1 , 0.  ]
                      ,[0.  , 0.02]])
prop = pypmc.density.student_t.LocalStud

# define the target; i.e., the function
# In this case, it is a Gaussian with me
# covariance "target_sigma".
#
# Note that the target function "log_tar
# unnormalized gaussian density.
target_sigma = np.array([[0.01 , 0.003 ]
                        ,[0.003, 0.0025]])
inv_target_sigma = np.linalg.inv(target_sigma)
target_mean  = np.array([4.3, 1.1])

def unnormalized_log_pdf_gauss(x, mu, inv_sigma):
    diff = x - mu
    return -0.5 * diff.dot(inv_sigma).dot(diff)

log_target = lambda x: unnormalized_log_pdf_gauss(x, target_mean, i

# choose a bad initialization
start = np.array([-2., 10.])

# define the markov chain object
mc = pypmc.sampler.markov_chain.AdaptiveMarkovChain(log_target, prc

# run burn-in
mc.run(10**4)

# delete burn-in from history
mc.history.clear()
```



(effective samples).

- **rel_tol** –
  Relative tolerance ϵ. If two consecutive values of the log likelihood bound, L_t, L_{t-1}, are close, declare convergence. More precisely, check that

$$\left\| \frac{L_t - L_{t-1}}{L_t} \right\| < \epsilon.$$

- **abs_tol** –
  Absolute tolerance ϵ_a. If the current bound L_t is close to zero, $(L_t < \epsilon_a)$, declare convergence if

$$\|L_t - L_{t-1}\| < \epsilon_a.$$

- **verbose** – Output status information after each update.

`set_variational_parameters()`

Reset the parameters to the submitted values or default.

Use this function to set the prior value (indicated by the subscript 0 as in α_0) or the initial value (e.g., α) used in the iterative procedure to find the posterior value of the variational distribution.

Every parameter can be set in two ways:

1. It is specified for only one component, then it is copied to all other components.
2. It is specified separately for each component as a K vector.

The prior and posterior variational distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ for each component are given by

$$q(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = q(\boldsymbol{\mu}|\boldsymbol{\Lambda})q(\boldsymbol{\Lambda}) = \prod_{k=1}^{K} \mathcal{N}(\boldsymbol{\mu}_k|\boldsymbol{m_k}, (\beta_k \boldsymbol{\Lambda}_k)^{-1})\mathcal{W}(\boldsymbol{\Lambda}_k|\boldsymbol{W_k}, \nu_k),$$

where $\mathcal{N}$ denotes a Gaussian and $\mathcal{W}$ a Wishart distribution. The weights $\boldsymbol{\pi}$ follow a Dirichlet distribution

$$q(\boldsymbol{\pi}) = Dir(\boldsymbol{\pi}|\boldsymbol{\alpha}).$$

Warning

This function may delete results obtained by `update()`.

https://pypi.python.org/pypi/pypmc

The standard model (SM) of particle physics cannot explain:

- dark matter
- neutrino masses
- hierarchy problem
- strong CP problem
- ...

Motivation

The standard model (SM) of particle physics cannot explain:

- dark matter
- neutrino masses
- hierarchy problem
- strong CP problem
- ...

new physics (NP) required

The standard model (SM) of particle physics cannot explain:

- dark matter
- neutrino masses
- hierarchy problem
- strong CP problem
- ...

new physics (NP) required

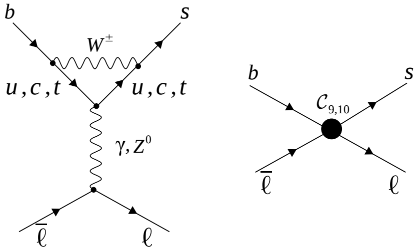
exact structure unknown $\Rightarrow$ model independent analysis

Effective theory

effective Lagrangian for $b \to s\ell^+\ell^-$ (SM):

$$\mathcal{L}_{int} = \frac{4G_F}{\sqrt{2}}\frac{\alpha_e}{4\pi}V_{tb}V_{ts}^*\sum_i \mathcal{C}_i\mathcal{O}_i + \ldots + \text{h.c.}$$

$$\mathcal{O}_9 = \left[\bar{s}\gamma_\mu P_L \quad b\right]\left[\bar{\ell}\gamma^\mu\ell\right] \qquad \mathcal{O}_{10} = \left[\bar{s}\gamma_\mu P_L \quad b\right]\left[\bar{\ell}\gamma^\mu\gamma_5\ell\right]$$



Effective theory

effective Lagrangian for $b \to s\ell^+\ell^-$ (beyond-SM):

$$\mathcal{L}_{int} = \frac{4G_F}{\sqrt{2}} \frac{\alpha_e}{4\pi} V_{tb} V_{ts}^* \sum_i \mathcal{C}_i \mathcal{O}_i + \ldots + \text{h.c.}$$

$$\mathcal{O}_9^{(\prime)} = \left[\bar{s}\gamma_\mu P_{L(R)} b\right]\left[\bar{\ell}\gamma^\mu \ell\right] \qquad \mathcal{O}_{10}^{(\prime)} = \left[\bar{s}\gamma_\mu P_{L(R)} b\right]\left[\bar{\ell}\gamma^\mu \gamma_5 \ell\right]$$

$$\mathcal{O}_S^{(\prime)} = \left[\bar{s} P_{R(L)} b\right]\left[\bar{\ell}\ell\right] \qquad \mathcal{O}_P^{(\prime)} = \left[\bar{s} P_{R(L)} b\right]\left[\bar{\ell}\gamma_5 \ell\right]$$

$$\mathcal{O}_T = \left[\bar{s}\sigma_{\mu\nu} b\right]\left[\bar{\ell}\sigma^{\mu\nu}\ell\right] \qquad \mathcal{O}_{T5} = \left[\bar{s}\sigma_{\mu\nu} b\right]\left[\bar{\ell}\sigma^{\mu\nu}\gamma_5 \ell\right]$$

Effective theory

- **B $\rightarrow$ K$\mu^+\mu^-$ angular distribution**

- **B$_s$ $\rightarrow$ $\mu^+\mu^-$ branching fraction**

Effective theory

- **B $\rightarrow$ K$\mu^+\mu^-$ angular distribution**

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\theta} = \frac{3}{4}(1 - F_H)\sin^2\theta + \frac{1}{2}F_H + A_{FB}\cos\theta$$

- **B$_s$ $\rightarrow$ $\mu^+\mu^-$ branching fraction**

Effective theory

- **$\mathbf{B \to K\mu^+\mu^-}$ angular distribution**

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\theta} = \frac{3}{4}(1 - F_H)\sin^2\theta + \frac{1}{2}F_H + A_{FB}\cos\theta$$

$$A_{FB} \propto \mathrm{Re}\Big[(\mathcal{C}_P + \mathcal{C}'_P)\,\mathcal{C}^*_{T5} + (\mathcal{C}_S + \mathcal{C}'_S)\,\mathcal{C}^*_T + \mathcal{O}(m_\ell/\sqrt{q^2})\Big]\Big/\Gamma$$

- **$\mathbf{B_s \to \mu^+\mu^-}$ branching fraction**

Effective theory

- **B → K$\mu^+\mu^-$ angular distribution**

$$\frac{1}{\Gamma}\frac{\mathrm{d}\Gamma}{\mathrm{d}\cos\theta} = \frac{3}{4}(1 - F_H)\sin^2\theta + \frac{1}{2}F_H + A_{FB}\cos\theta$$

$$A_{FB} \propto \mathrm{Re}\left[(\mathcal{C}_P + \mathcal{C}_P')\,\mathcal{C}_{T5}^* + (\mathcal{C}_S + \mathcal{C}_S')\,\mathcal{C}_T^* + \mathcal{O}(m_\ell/\sqrt{q^2})\right]\Big/\Gamma$$

$$F_H \propto \Big[\ldots(|\mathcal{C}_T|^2 + |\mathcal{C}_{T5}|^2) + \ldots(|\mathcal{C}_S + \mathcal{C}_S')|^2 + |\mathcal{C}_P + \mathcal{C}_P'|^2) + \mathcal{O}(m_\ell/\sqrt{q^2})\Big]\Big/\Gamma$$

- **B$_s$ → $\mu^+\mu^-$ branching fraction**

Effective theory

- **B $\to$ K$\mu^+\mu^-$ angular distribution**

$$\frac{1}{\Gamma}\frac{\mathrm{d}\Gamma}{\mathrm{d}\cos\theta} = \frac{3}{4}(1-F_H)\sin^2\theta + \frac{1}{2}F_H + A_{FB}\cos\theta$$

$$A_{FB} \propto \mathrm{Re}\left[(\mathcal{C}_P+\mathcal{C}'_P)\,\mathcal{C}^*_{T5} + (\mathcal{C}_S+\mathcal{C}'_S)\,\mathcal{C}^*_T + \mathcal{O}(m_\ell/\sqrt{q^2})\right]\Big/\Gamma$$

$$F_H \propto \Big[\ldots(|\mathcal{C}_T|^2+|\mathcal{C}_{T5}|^2) + \ldots(|\mathcal{C}_S+\mathcal{C}'_S)|^2+|\mathcal{C}_P+\mathcal{C}'_P|^2) + \mathcal{O}(m_\ell/\sqrt{q^2})\Big]\Big/\Gamma$$

$$F_H^{\mathrm{SM}} = \mathcal{O}(m_\ell^2/q^2) \qquad A_{FB}^{\mathrm{SM}} = 0$$

- **B$_s$ $\to \mu^+\mu^-$ branching fraction**

Effective theory

- **B $\to$ K$\mu^+\mu^-$ angular distribution**

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\theta} = \frac{3}{4}(1 - F_H)\sin^2\theta + \frac{1}{2}F_H + A_{FB}\cos\theta$$

$$A_{FB} \propto \mathrm{Re}\Big[(\mathcal{C}_P + \mathcal{C}'_P)\,\mathcal{C}^*_{T5} + (\mathcal{C}_S + \mathcal{C}'_S)\,\mathcal{C}^*_T + \mathcal{O}(m_\ell/\sqrt{q^2})\Big]\Big/\Gamma$$

$$F_H \propto \Big[\ldots(|\mathcal{C}_T|^2 + |\mathcal{C}_{T5}|^2) + \ldots(|\mathcal{C}_S + \mathcal{C}'_S)|^2 + |\mathcal{C}_P + \mathcal{C}'_P|^2) + \mathcal{O}(m_\ell/\sqrt{q^2})\Big]\Big/\Gamma$$

$$F_H^{\mathrm{SM}} = \mathcal{O}(m_\ell^2/q^2) \qquad A_{FB}^{\mathrm{SM}} = 0$$

- **B$_s$ $\to \mu^+\mu^-$ branching fraction**

$$\mathcal{B}(B_s \to \mu^+\mu^-) \propto |\mathcal{C}_S - \mathcal{C}'_S|^2 + |(\mathcal{C}_P - \mathcal{C}'_P) + \frac{2m_\ell}{M_{B_s}}(\mathcal{C}_{10} - \mathcal{C}'_{10})|^2$$

Methodology

we want:

$$P(\boldsymbol{\theta}|\mathcal{D},\mathrm{M}) \overset{\text{Bayes}}{\propto} P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})$$

Methodology

we want:

$$P(\boldsymbol{\theta}|\mathcal{D}, \mathrm{M}) \overset{\text{Bayes}}{\propto} P(\mathcal{D}|\boldsymbol{\theta}, \mathrm{M}) P(\boldsymbol{\theta}|\mathrm{M})$$

split theory and experiment - *observables* **O**:

$$P(\mathcal{D}|\boldsymbol{\theta}, \mathrm{M}) = P(\mathcal{D}|\mathbf{O}(\boldsymbol{\theta}, \mathrm{M}))$$

Methodology

we want:

$$P(\boldsymbol{\theta}|\mathcal{D},\mathrm{M}) \overset{\text{Bayes}}{\propto} P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})$$

split theory and experiment - *observables* **O**:

$$P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M}) = P(\mathcal{D}|\mathbf{O}(\boldsymbol{\theta},\mathrm{M}))$$

theory	experiment

Methodology

we want:

$$P(\boldsymbol{\theta}|\mathcal{D},\mathrm{M}) \overset{\mathrm{Bayes}}{\propto} P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})$$

split theory and experiment - *observables* **O**:

$$P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M}) = P(\mathcal{D}|\mathbf{O}(\boldsymbol{\theta},\mathrm{M}))$$

theory	experiment
calculate observables $\mathbf{O}(\boldsymbol{\theta},\mathrm{M})$	

Methodology

we want:

$$P(\boldsymbol{\theta}|\mathcal{D},\mathrm{M}) \overset{\text{Bayes}}{\propto} P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})$$

split theory and experiment - *observables* **O**:

$$P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M}) = P(\mathcal{D}|\mathbf{O}(\boldsymbol{\theta},\mathrm{M}))$$

theory

calculate observables

$\mathbf{O}(\boldsymbol{\theta},\mathrm{M})$

experiment

measure observables

$P\left(\mathcal{D}|\mathbf{O}\right)$

we want:

$$P(\boldsymbol{\theta}|\mathcal{D},\mathrm{M}) \overset{\text{Bayes}}{\propto} P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M})P(\boldsymbol{\theta}|\mathrm{M})$$

split theory and experiment - *observables* **O**:

$$P(\mathcal{D}|\boldsymbol{\theta},\mathrm{M}) = P(\mathcal{D}|\mathbf{O}(\boldsymbol{\theta},\mathrm{M}), \underbrace{\cancel{\boldsymbol{\theta},\mathrm{M}}}_{\text{assumption}})$$

theory

calculate observables

$\mathbf{O}(\boldsymbol{\theta},\mathrm{M})$

experiment

measure observables

$P\left(\mathcal{D}|\mathbf{O},\cancel{\boldsymbol{\theta},\mathrm{M}}\right)$

Measurements $P(\mathcal{D}|\mathbf{O})$

- $B \to K\mu^+\mu^-$: $\mathcal{B}, A_{FB}, F_H$
  - LHCb 2014 (arXiv:1403.8044 , arXiv:1403.8045)
  - CDF 2012

    (http://www-cdf.fnal.gov/physics/new/bottom/120628.blessed-b2smumu_96)
- $B_s \to \mu^+\mu^-$: $\mathcal{B}$
  - LHCb+CMS 2014 (arXiv:1411.4413)
- $B \to K^*\mu^+\mu^-$: $\mathcal{B}$
  - LHCb 2013 (arXiv:1304.6325)
  - CMS 2013 (arXiv:1308.3409)
  - CDF 2012

    (http://www-cdf.fnal.gov/physics/new/bottom/120628.blessed-b2smumu_96)

Parameters $\boldsymbol{\theta}$

scan parameters

- Wilson coefficients $\mathcal{C}_{10}^{(\prime)}, \mathcal{C}_S^{(\prime)}, \mathcal{C}_P^{(\prime)}, \mathcal{C}_T$, and $\mathcal{C}_{T5}$

Parameters θ

scan parameters

- Wilson coefficients $\mathcal{C}_{10}^{(\prime)}, \mathcal{C}_S^{(\prime)}, \mathcal{C}_P^{(\prime)}, \mathcal{C}_T$, and $\mathcal{C}_{T5}$

nuisance parameters

- CKM matrix (4 parameters)
- charm and bottom quark mass (2 parameters)
- form factors
  - $B \to K$ (5 parameters)
  - $B \to K^*$ (6 parameters)
- B_s decay constant f_{B_s}
- subleading corrections (11 parameters)

Parameters θ

scan parameters

- Wilson coefficients $\mathcal{C}_{10}^{(\prime)}, \mathcal{C}_{S}^{(\prime)}, \mathcal{C}_{P}^{(\prime)}, \mathcal{C}_{T}$, and $\mathcal{C}_{T5}$

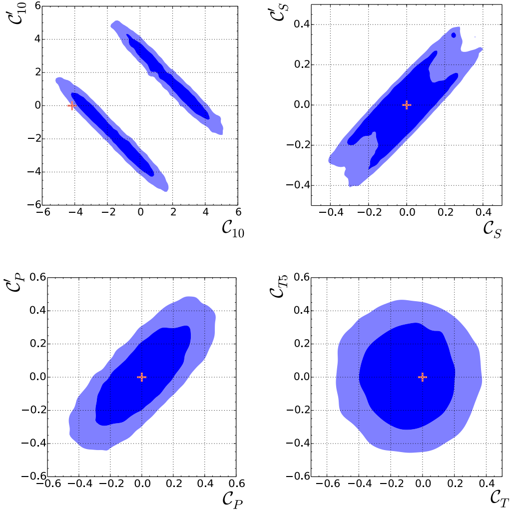
nuisance parameters

- CKM matrix (4 parameters)
- charm and bottom quark mass (2 parameters)
- form factors
  - $B \to K$ (5 parameters)
  - $B \to K^*$ (6 parameters)
- B_s decay constant f_{B_s}
- subleading corrections (11 parameters)

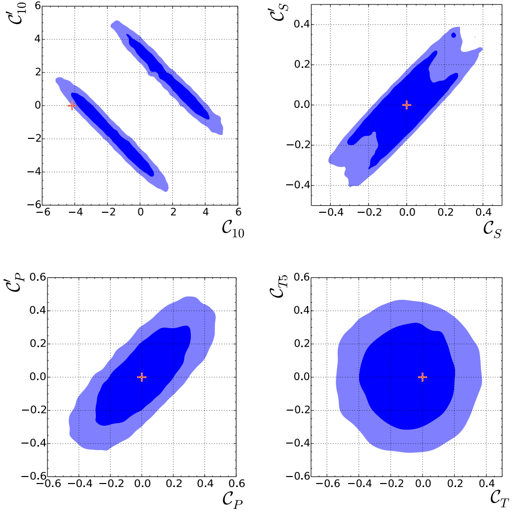
theory calculation $\mathbf{O}(\theta, \mathbf{M})$:

open-source implementation: EOS-package
`http://project.het.physik.tu-dortmund.de/eos/`

Joint fit of $\mathcal{C}_{10}^{(\prime)}, \mathcal{C}_S^{(\prime)}, \mathcal{C}_P^{(\prime)}, \mathcal{C}_T$, and $\mathcal{C}_{T5}$

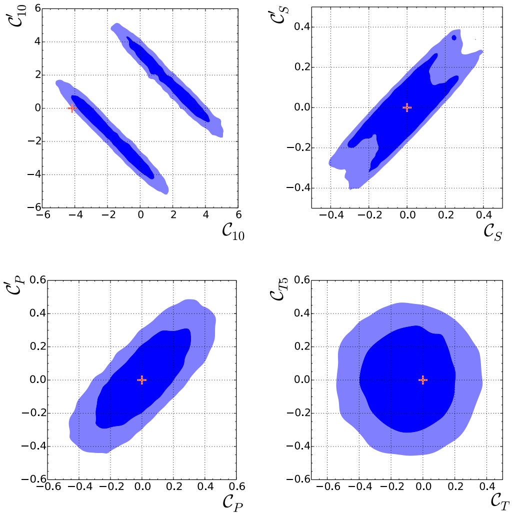


Joint fit of $\mathcal{C}_{10}^{(\prime)}, \mathcal{C}_S^{(\prime)}, \mathcal{C}_P^{(\prime)}, \mathcal{C}_T$, and $\mathcal{C}_{T5}$



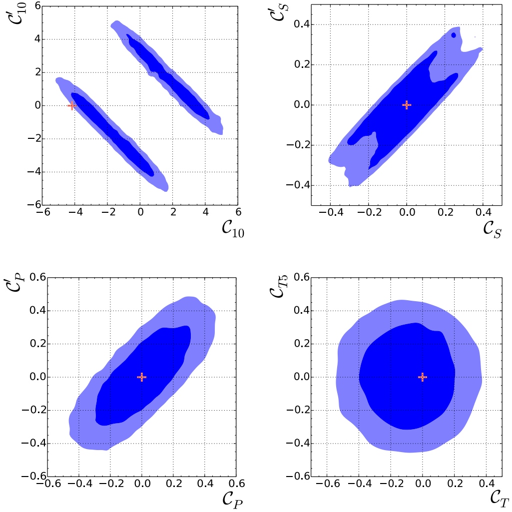
- first *simultaneous* fit

Joint fit of $\mathcal{C}_{10}^{(\prime)}, \mathcal{C}_S^{(\prime)}, \mathcal{C}_P^{(\prime)}, \mathcal{C}_T$, and $\mathcal{C}_{T5}$



- first *simultaneous* fit
- interference $\mathcal{C}_{10}^{(\prime)} \leftrightarrow \mathcal{C}_{S,P}^{(\prime)}$ in $\mathcal{B}(B_s \to \mu^+\mu^-)$

Joint fit of $\mathcal{C}_{10}^{(\prime)}, \mathcal{C}_S^{(\prime)}, \mathcal{C}_P^{(\prime)}, \mathcal{C}_T$, and $\mathcal{C}_{T5}$



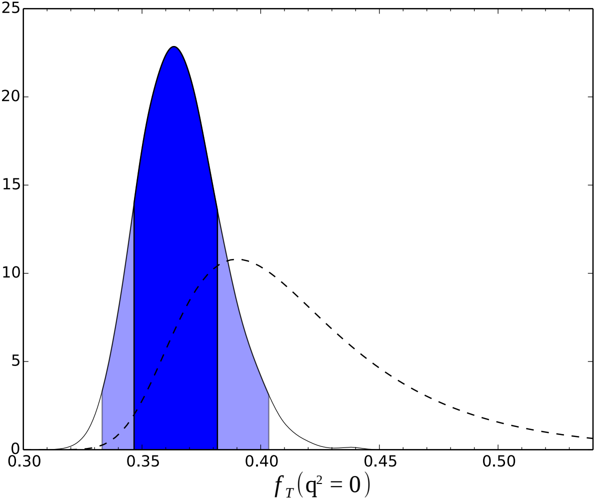
- first *simultaneous* fit
- interference $\mathcal{C}_{10}^{(\prime)} \leftrightarrow \mathcal{C}_{S,P}^{(\prime)}$ in $\mathcal{B}(B_s \to \mu^+\mu^-)$
  - $\Rightarrow$ larger uncertainty than obtained for fixed $\mathcal{C}_{10}^{(\prime)} = \mathcal{C}_{10}^{(\prime)\mathrm{SM}}$

arXiv:1205.5811,

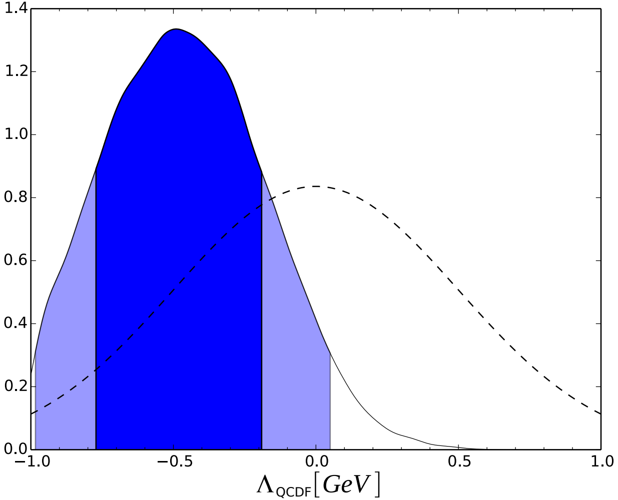
arXiv:1206.0273,

arXiv:1407.7044

Nuisance parameters



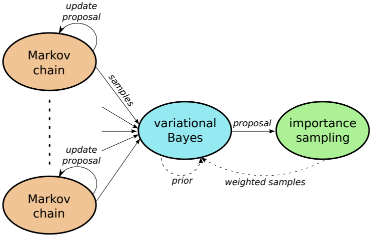
Nuisance parameters



Summary

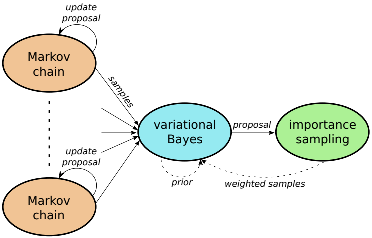
Summary

algorithm to sample and integrate in dim = $\mathcal{O}(40)$



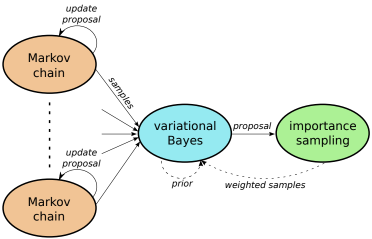
algorithm to sample and integrate in dim = $\mathcal{O}(40)$

model-independent search for new physics



Summary

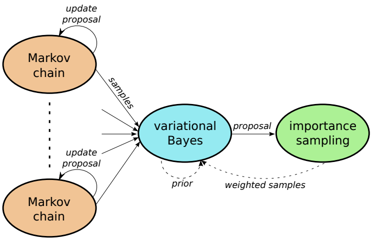
algorithm to sample and integrate in dim = $\mathcal{O}(40)$



model-independent search for new physics

- simultaneous fit of $\mathcal{C}_{10}^{(\prime)}, \mathcal{C}_S^{(\prime)}, \mathcal{C}_P^{(\prime)}, \mathcal{C}_T$, and $\mathcal{C}_{T5}$
  $\Rightarrow$ updated constraints

algorithm to sample and integrate in dim $= \mathcal{O}(40)$

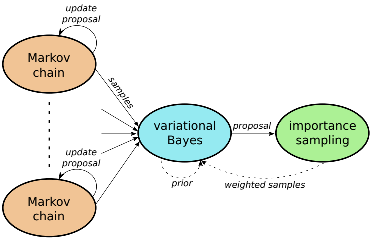


model-independent search for new physics

- simultaneous fit of $\mathcal{C}_{10}^{(\prime)}, \mathcal{C}_S^{(\prime)}, \mathcal{C}_P^{(\prime)}, \mathcal{C}_T$, and $\mathcal{C}_{T5}$
  $\Rightarrow$ updated constraints
- no significant deviation from the SM

Summary

algorithm to sample and integrate in dim = $\mathcal{O}(40)$



model-independent search for new physics

- simultaneous fit of $\mathcal{C}_{10}^{(\prime)}, \mathcal{C}_S^{(\prime)}, \mathcal{C}_P^{(\prime)}, \mathcal{C}_T$, and $\mathcal{C}_{T5}$
  $\Rightarrow$ updated constraints
- no significant deviation from the SM
- need better theoretical control (form factors, QCDF)