$0\nu\beta\beta$ UQ: Quick and Dirty Version

Kyle Wendt
LLNL

Witold Nazarewicz
MSU,FRIB
(Dated: February 2021)

I. A FEW QUICK DEFINITIONS

I just want to make some of the notation explicit. I am trying use as common notation as possible, however even within specific domains, statistical and probability notations are never consistent.

$$
\begin{aligned}
P(X=x|A) &= \text{Probability density that a stochastic variable } X \text{ has value } x \text{ given assumptions } A \\
&= f_{P(X|A)}(x) \\
P(X|A) &= \text{Shorthand to clean up notation for } P(X=X|A), \\
&\quad \text{where X represent both the stochastic variable and the value} \\
\boldsymbol{X} &= \text{a vector of numbers} \\
\boldsymbol{\Sigma} &= \text{a covariance matrix} \\
\boldsymbol{\Sigma}_{ab} &= \text{A sub-block of a covariance matrix} \\
\mathrm{E}[X|A] &= \int xP(X=x|A)dx = \text{The expectation value of X under the assumptions A} \\
\mathrm{Cov}[X,Y|A] &\equiv \mathrm{E}[XY|A] - \mathrm{E}[X|A]\ \mathrm{E}[Y|A] = \text{The covariance of } X \text{ and } Y \\
\mathrm{Cov}[\boldsymbol{X},\boldsymbol{Y}|A] &= \text{the matrix of covariances of the elements of the vectors } \boldsymbol{X} \text{ and } \boldsymbol{Y} \\
\mathrm{Cov}[\boldsymbol{X}|A] &\equiv \mathrm{Cov}[\boldsymbol{X},\boldsymbol{X}|A] \text{ (shorthand)}
\end{aligned}
$$

II. QUICK OVERVIEW OF BAYESIAN MODEL AVERAGING (BMA)

Let use assume we have N distinct models, with the ith model denoted M_i. Each model will have model specific parameters θ_i that are fixed externally, the number and nature of these parameters can vary from model to model. Each model M_i can predict a range of quantities denoted x_i with a common range of computable quantities between all models that we will denote as $\boldsymbol{X}$. We will further split $\boldsymbol{X}$ into $\boldsymbol{X}_p$ and $\boldsymbol{X}_e$ where we have some data for $\boldsymbol{X}_e$ and we want to make a robust prediction of $\boldsymbol{X}_p$. The idea behind Bayesian model averaging is that the collective information of many different (even bad) models can be greater than sum of information of individual models. More explicitly, a "best" collective estimator for $\boldsymbol{X}_p$ is distributed as:

$$P(\boldsymbol{X}_p|\boldsymbol{D}_e) = \sum_{i=1}^{N} P(\boldsymbol{X}_p|M_i,\boldsymbol{D}_e)P(M_i|\boldsymbol{D}_e)\,, \tag{1}$$

where $\boldsymbol{D}_e$ are the experimental data for the observables in $\boldsymbol{X}_e$.

The model evidence prior is defined using Bayes theorem as

$$P(M_i|\boldsymbol{D}_e) = \frac{P(\boldsymbol{D}_e|M_i)}{\sum_{i=l}^{N} P(\boldsymbol{D}_e|M_l)}\,. \tag{2}$$

The major challenge often comes from computing the marginal likelihood integrals:

$$P(\boldsymbol{D}_e|M_i) = \int P(\boldsymbol{D}_e|M_i,\theta_i)P(\theta_i|M_i)d\theta_i\,. \tag{3}$$

where the two ingredients are:

$$P(\theta_i|M_i) = \quad \text{Calibrated Model Parameter Posterior.} \tag{4}$$
(ex., the likelihoods from EFT papers)

$$P(\boldsymbol{D}_e|M_i,\theta_i) = \quad \text{Model Evidence Likelihood.} \tag{5}$$
(ex., a multivariate Gaussian in the simplest and most common cases in physics)

The relative normalization of each $P(\boldsymbol{D}_e|M_i)$ is important, so the common tricks we use in MCMC sampling won't work, and we either need to brute force these integrals or make some approximation, I will discuss two possibilities below.

Finally, our outputs, the estimated mean and covariance of the combined predictions can be computed as:

$$\hat{\boldsymbol{X}}_p \equiv \mathrm{E}[\boldsymbol{X}_p|\boldsymbol{D}_e] = \sum_{i=1}^{N} \hat{\boldsymbol{X}}_{p,i} P(M_i|\boldsymbol{D}_e) \tag{6}$$

$$\hat{\boldsymbol{\Sigma}}_{pp} \equiv \mathrm{Cov}[\boldsymbol{X}_p|\boldsymbol{D}_e] = \sum_{i=1}^{N} \left(\hat{\boldsymbol{\Sigma}}_{pp,i} + \hat{\boldsymbol{X}}_{p,i}\hat{\boldsymbol{X}}_{p,i}^{\intercal}\right) P(M_i|\boldsymbol{D}_e) - \hat{\boldsymbol{X}}_p\hat{\boldsymbol{X}}_p^{\intercal}\,, \tag{7}$$

where

$$\hat{\boldsymbol{X}}_{p,i} \equiv \mathrm{E}[\boldsymbol{X}_p|\boldsymbol{D}_e, M_i] = \text{Conditional Mean Vector from Model} \tag{8}$$

$$\hat{\boldsymbol{\Sigma}}_{pp,i} \equiv \mathrm{Cov}[\boldsymbol{X}_p|\boldsymbol{D}_e, M_i] = \text{Conditional Covariance Matrix from Model} \tag{9}$$

A. Concrete List of Needs and "Ingredients" for "Quick and *D*irty" BMA

For each model we need the following

$\bar{\boldsymbol{X}}_i = \begin{pmatrix} \bar{\boldsymbol{X}}_{p,i} \\ \bar{\boldsymbol{X}}_{e,i} \end{pmatrix}$	Mean vector of all predicted quantities from model M_i
$\boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_{pp,i} & \boldsymbol{\Sigma}_{pe,i} \\ \boldsymbol{\Sigma}_{ep,i} & \boldsymbol{\Sigma}_{ee,i} \end{pmatrix}$	Covariance matrix vector of all predicted quantities from model M_i
$P(\boldsymbol{D}_e\|M_i)$	Marginal Likelihood of evidence data given model M_i

In principle, for each model considered, we need the means and covariances for $\boldsymbol{X}_p$ conditioned on $\boldsymbol{D}_e$. This requires access to the full posterior for $\boldsymbol{X}$:

$$P(\boldsymbol{X}_p|\boldsymbol{D}_e, M_i) \equiv P(\boldsymbol{X}_p|\boldsymbol{X}_e = \boldsymbol{D}_e, M_i) = \frac{P(\boldsymbol{X}_p, \boldsymbol{X}_e|M_i)}{\int P(\boldsymbol{X}_p, \boldsymbol{X}_e|M_i) d\boldsymbol{X}_p}\,, \tag{10}$$

If we assume that $P(\boldsymbol{X}_p, \boldsymbol{X}_e|M_i)$ is close to a multivariate Gaussian near the mean vector, the we can compute/approximate these conditioned values as:

$$\hat{\boldsymbol{X}}_{p,i} = \mathrm{E}[\boldsymbol{X}_p|\boldsymbol{D}_e] \approx \bar{\boldsymbol{X}}_{p,i} + \boldsymbol{\Sigma}_{pe,i}\boldsymbol{\Sigma}_{ee,i}^{-1}\left(\boldsymbol{D}_e - \bar{\boldsymbol{X}}_{e,i}\right) \tag{11}$$

$$\hat{\boldsymbol{\Sigma}}_{pp,i} = \mathrm{Cov}[\boldsymbol{X}_p|\boldsymbol{D}_e, M_i] \approx \boldsymbol{\Sigma}_{pp,i} - \boldsymbol{\Sigma}_{pe,i}\boldsymbol{\Sigma}_{ee,i}^{-1}\boldsymbol{\Sigma}_{ep,i}\,. \tag{12}$$

Note that we recover the original means and covariances if our evidence and predicted quantities are uncorrelated. This procedure fails if the Gaussian approximation for $P(\boldsymbol{X}_p, \boldsymbol{X}_e|M_i)$ does not hold.

Along this same line, we need a reasonable approximation for $P(\boldsymbol{D}_e|M_i)$. In principle we need the exact normalized value for this quantity. That means performing the high dimensional integrals over model parameters. Instead, we can continue the multivariate Gaussian approximation and use:

$$P(\boldsymbol{D}_e|M_i) = f_{\mathcal{N}\left(\bar{\boldsymbol{X}}_{e,i}, \boldsymbol{\Sigma}_{ee,i}\right)}\left(\boldsymbol{D}_e\right)\,, \tag{13}$$

and compute all integrals and weights trivially.

If we have the model codes and resources, we can go one step further and generate samples from each models' true distribution without relying on the Gaussian ansatz. While these are likely not enough to compute the needed normalization integrals, we can instead turn to various statistical information criterion to approximate the Bayesian weights. To do this we replace the marginal likelihood with

$$P(\boldsymbol{D}_e|M_i) \to e^{-1/2\Delta IC_i}\,, \tag{14}$$

$$\Delta IC_i = IC_i - \min_j IC_j\,, \tag{15}$$

$$IC_i = \text{Information Criterion for } i\text{-th model.} \tag{16}$$

The information criterion is often the Watanabe–Akaike information criterion (or widely accepted information criterion), which packages like pymc3 can compute for us, though recently new metrics have gained favor such as Leave-one-out cross validation (again, we pass samples to pymc3 to compute for us). This information criterion driven approximation is often called pseudo-Bayesian model averaging and is typically a reasonable approximation to full Bayesian model averaging.

III. SUPER DIRTY HIERARCHICAL BAYESIAN MODEL AVERAGING (GAUSSIAN MODEL)

At this stage, we need to gather predictions, preferably with covariances from the many-body practitioners. Below is a table of what has been sent to Engel. We need to gather as much data around the nuclei of interest as possible, so masses, energy levels, radii, transition strengths (E1, M1, GT, etc) Even if the data is only weakly correlated to the desired transition strengths, it will at worst drop out and at best provide a small amount of additional information. The general workflow to producing the best informed prediction under this Gaussian Bayesian Model Averaging is presented below. I refer to this as a Hierarchical Bayesian Model Averaging since we will use the same BMA equations to combine different interaction models under the same many-body method into a mean and covariance prediction for the method as a whole, and then we use those method labels means and covariance to combine different methods into one global prediction.

1. Compute each observable needed (see below).
   - *Ab initio* models (i.e. too expensive to vary parameters or fully propagate covariances)
     (a) Compute for an many chiral interactions as possible
         – Non-converged with an estimate of convergence uncertainty is better than not computing anything.
     (b) Estimate systematic uncertainties
         – Method uncertainties from comparison to microscopic
         – Convergence error
     (c) When possible, estimate 1σ uncertainties and covariances.
   - Other models (i.e. possible to directly estimate covariances)
     (a) Compute every observable you can from matrix below
     (b) Estimate 1σ uncertainties and covariances when possible
         – Can fall back to same idea on chiral models, but it wont work as well[1].
2. (for *ab initio*, done by Nazarewicz or Wendt) Apply Quick and Dirty BMA on each method separately.
   - For each method (e.g. GCM IM-SRG), we combine all the different interactions models (e.g. "Magic" Hebeler potential) using BMA.
   - The resulting mean vector and covariance matrices will be considered a "model" for the global BMA

   (a) Compute mean vectors and covariance matrices conditioned on some subset of evidence data.
   (b) Compute Gaussian BMA weights on subset of evidence data
   (c) Compute BMA means and covariances

[1] More or less, the chiral models have more built in underlying physics correlations that we can lean upon

3. Compute mean vectors and covariance matrices conditioned on some global (all methods and models) full set of evidence data.

4. Compute Gaussian BMA weights on full subset of evidence data

5. Compute BMA means and covariances

IV. DATA TO BE COMPUTED

- In the $0\nu\beta\beta$ candidates ^{48}Ca, ^{76}Ge, ^{130}Te, ^{136}Xe (and particularly the first, which several ab initio methods have addressed):
  - Single β-decay rates in nearby nuclei, e.g., intermediate nucleus in decay
  - β^- strength distribution from initial nucleus
  - β^+ distribution from final nucleus
  - $2\nu\beta\beta$ matrix elements
  - Magnetic moments and M1's in three nuclei involved in decay
  - E2 to lowest 2^+ state in initial and final nuclei
  - Energies of lowest few excited states
  - Radii

  Other heavy nuclei, not necessarily experimental candidates, would also be useful.

  With ab initio methods, we should use three interactions:

  - Hebeler magic
  - N3LO L-NL
  - N3LO GO

  With the phenomenological shell model, a few different interactions would be good.

- The same quantities in lighter nuclei. The ab initio groups can use the interaction(s) they agree on (the one Ingo worked with?) and the phenomenological shell-model people can use their favorite.