What is NP Bayes?

Peter Hoff

Statistics, Biostatistics and the CSSS
University of Washington
What is NP Bayes?

Is NP Bayes
  • Bayesian?
What is NP Bayes?

Is NP Bayes

- Bayesian?
- nonparametric?
What is NP Bayes?

Is NP Bayes

- Bayesian?
- nonparametric?
- a good idea?
Bayesian inference is the change from prior to posterior information:

- the prior information can be your information;
Bayesian inference is the change from prior to posterior information:

- the prior information can be your information;
- it can be someone else’s;
Bayesian inference is the change from prior to posterior information:

- the prior information can be your information;
- it can be someone else’s;
- it can even be “really small.”
Is it Bayesian?

Bayesian inference is the change from prior to posterior information:

- the prior information can be your information;
- it can be someone else’s;
- it can even be “really small.”

It should at least approximate information that someone could possibly have. Otherwise

\[ p(\theta|y) \neq \text{posterior information} \]

because

\[ p(\theta) \neq \text{prior information} \]
Is it nonparametric?

If “nonparametric” means no parameters, then clearly not.

If “nonparametric” means $p$ grows with $n$, then sort of:

- Parametric asymptotics: $p_n / \sqrt{n} \to 0$;
- DPM: $p_n / \log n \to c$, $p_n / \sqrt{n} \to 0$.

If “nonparametric” means consistency for any population, then sometimes:

- DPM: $p(y|q) = \int f(y|\theta)q(d\theta) \Rightarrow p(y|q) \in \mathcal{H}\{f(y|\theta), \theta \in \Theta\}$
- Examples of inconsistency
Is it a good idea?

Yes:
- clustering/mixture modeling (although: clusters of what?)
- density estimation
- prediction

No:
- a huge modeling effort
- need large amounts of data
- parameters can be hard to interpret
A few more miscellaneous complaints

Entropy reduction

- Consequences for mixture modeling, hierarchical modeling.
- Hill, Lane Sudderth (1986?) negative result on urn processes

Data analysis

- \( y \rightarrow \text{data analysis} \rightarrow t(y) \)
- \( y \rightarrow \text{NPBayes} \rightarrow t(y), \theta_1, \theta_2, \ldots \)
- How to summarize a posterior distribution over partitions?
Misspecified models

What’s so bad about misspecified models?

\[
\arg \max_{\theta} \prod_{i=1}^{n} f_{\theta}(y_i) = \hat{\theta} \rightarrow \theta_0 = \arg \min_{\theta} \int \log \frac{p_0(y)}{f_\theta(y)} p_0(y) dy.
\]
Misspecified models

What’s so bad about misspecified models?

\[
\arg\max_{\theta} \prod_{i=1}^{n} f_{\theta}(y_i) = \hat{\theta} \rightarrow \theta_0 = \arg\min_{\theta} \int \log \frac{p_0(y)}{f_{\hat{\theta}}(y)} p_0(y) dy.
\]

Suppose we are interested in estimating \( \lambda_0 \in \mathbb{R}^p \) where

\[
\lambda_{0j} = \int t_j(y) p_0(y) \, dy , j \in \{1, \ldots, p\}
\]

for some nice functions \( t_1, \ldots, t_p \).
Misspecified models

What’s so bad about misspecified models?

$$\arg \max_\theta \prod_{i=1}^n f_\theta(y_i) = \hat{\theta} \rightarrow \theta_0 = \arg \min_\theta \int \log \frac{p_0(y)}{f_\theta(y)} p_0(y) dy.$$ 

Suppose we are interested in estimating $$\lambda_0 \in \mathbb{R}^p$$ where

$$\lambda_{0j} = \int t_j(y) p_0(y) \, dy , j \in \{1, \ldots, p\}$$

for some nice functions $$t_1, \ldots, t_p$$.

We don’t necessarily need that $$p_0 \in \{ f_\theta : \theta \in \Theta \}$$, we just need

$$\int t_j(y) f_{\hat{\theta}}(y) \, dy = \hat{\lambda}_j \rightarrow \lambda_{0j}.$$
Exponential families

\[ f_\theta(y) = \exp\{\theta_1 t_1(y) + \cdots \theta_p t_p(y) - c(\theta)\} \]

\[ \hat{\lambda}_j = \int t_j(y)f_\hat{\theta}(y) \, dy \rightarrow \int t_j(y)f_{\theta_0}(y) = \lambda_{0j} \]

Useful functions of functionals:

- means, variances covariances
- “treatment effects” \( (\mu_1 - \mu_2)/\sigma \)
- solutions to (smooth) equations \( h(\psi, \lambda) = 0 \), solve \( h(\psi, \hat{\lambda}) \).
A simple example

Interest is in

\[ \lambda = \int t(y)p_0(y)dy \]

Then using

\[ f(y|\theta) \propto \exp\{\theta_1 t(y) + \theta_2 t^2(y)\} \]

gives

- consistent estimation of \( \lambda = \mathbb{E}[t(y)] \)
- asymptotically correct confidence intervals.

Another example:

\[ f(y|\theta) \propto \exp\{\theta_1 y + \theta_2 y^2 + \gamma_1 \log y + \gamma_2 (\log y)^2\} \]

- can vary smoothly between (truncated) normal, lognormal model.
- valid confidence intervals for \( \mathbb{E}[y], \mathbb{E}[\log y] \)
What does Bayes have to offer

“Sandwich estimation” for exponential families is method of moments:

\[ \sqrt{n}(\hat{\lambda} - \lambda_0) = \text{multivariate normal}(0, \text{Cov}[t_1(y), \ldots, t_p(y)]) \]

The covariance term needs to be plugged in.

Bayes offers:
- shrinkage
- integration over nuisance parameters

Which performs better?
- Sandwich estimation with plug-in variance
- Misspecified exponential family model, integrating over nuisance/variance parameters, with a proper, informative prior on the parameter of interest?
What is NP Bayes?

Is NP Bayes

- Bayesian?
What is NP Bayes?

Is NP Bayes

- Bayesian? no.
- nonparametric?

NP Bayes does several things that can't be done easily any other way. But in some problems we can use "Bayesian nonparametric" methods that

- are Bayesian, where the prior (over the parameter of interest) can be considered "real";
- are nonparametric, in that parameters for things that you (a) are not interested in and (b) have no information about need not be specified or estimated.
What is NP Bayes?

Is NP Bayes

- Bayesian? no.
- nonparametric? sort of.
- a good idea?
What is NP Bayes?

Is NP Bayes

- Bayesian? no.
- nonparametric? sort of.
- a good idea? sometimes.

NP Bayes does several things that can’t be done easily any other way.

But in some problems we can use “Bayesian nonparametric” methods that

- are Bayesian, where the prior (over the parameter of interest) can be considered “real”;
- are nonparametric, in that parameters for things that you
  (a) are not interested in and
  (b) have no information about

need not be specified or estimated.