Bayesian Non-parametric Modeling With Skewed and Heavy-tailed Data

David Draper
(joint work with Milovan Krnjajić, Thanasis Kottas and John Wallerius)

Department of Applied Mathematics and Statistics
University of California, Santa Cruz

draper@ams.ucsc.edu
www.ams.ucsc.edu/~draper

WNAR 2008 (UC Davis)

23 June 2008
Outline

Two case studies:

- **In-home geriatric assessment**: Parametric and Bayesian non-parametric modeling of skewed and over-dispersed count data (Krnjajić, Kottas and Draper (2008), *Computational Statistics and Data Analysis*, 52, 2110–2128)

  — Parametric random-effects Poisson modeling fails to capture features like bimodality and skewness on latent variable scale

- **Risk assessment for nuclear waste disposal**: Frequentist and Bayesian non-parametric modeling of long-tailed continuous data (in progress)

  — Mixture of point mass at 0 and very long-tailed distribution; non-parametric bootstrap fails miserably at drawing inferences about mean and variance functionals

In both cases Dirichlet process mixture modeling produces better-calibrated inferential and predictive statements than parametric and frequentist non-parametric approaches.
572 elderly people randomized (Hendriksen et al. 1984), 287 to control (C) group (standard care) and 285 to treatment (T) group (standard care plus in-home geriatric assessment (IHGA): preventive medicine in which each person’s medical/social needs assessed, acted upon individually).

One important outcome was number of hospitalizations (in two years): $y^T_i, y^C_j =$ numbers of hospitalizations for treatment person $i$, control person $j$, respectively.

| Distribution of number of hospitalizations in IHGA study over two-year period: |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                  | 0               | 1               | 2               | 3               | 4               | 5               | 6               | 7               |
| Group            | n              | Mean            | SD              | n              | Mean            | SD              |
| Control          | 138            | 77             | 46             | 12             | 8               | 4               | 0               | 2               | 287            | 0.944           | 1.24           |
| Treatment        | 147            | 83             | 37             | 13             | 3               | 1               | 1               | 0               | 285            | 0.768           | 1.01           |

Evidently IHGA lowered mean hospitalization rate (for these elderly Danish people, at least) by $(0.944 - 0.768) = 0.176$, which is about $100 \left( \frac{0.768 - 0.944}{0.944} \right) = 19\%$ reduction from control level, a difference that’s large in clinical terms; also large statistically?
**Parametric Modeling**

- **A natural starting-point** would be fixed-effects Poisson modeling:

\[
(C_i | \lambda_C) \overset{\text{IID}}{\sim} \text{Poisson}(\lambda_C), \quad (T_j | \lambda_T) \overset{\text{IID}}{\sim} \text{Poisson}(\lambda_T),
\]

for \(i = 1, \ldots, n_C = 287\) and \(j = 1, \ldots, n_T = 285\), with, say, a **diffuse prior** for \((\lambda_C, \lambda_T)\).

But the last two columns of previous table reveal that the **sample variance-to-mean ratios** in the **control** and **treatment groups** are 1.63 and 1.33, respectively, indicating **substantial Poisson over-dispersion**.

- **Parametric random-effects Poisson** (PREP) model is natural next choice for \(C\) and \(T\) data sets (in parallel): changing notation,

\[
(y_i | \theta_i) \overset{\text{ind}}{\sim} \text{Poisson}[\exp(\theta_i)]
\]

\[
(\theta_i | G) \overset{\text{iid}}{\sim} G
\]

\[
G \equiv \text{N}(\mu, \sigma^2)
\]

assuming a parametric CDF \(G\) for latent variables \(\theta_i\) (random effects).
Dirichlet Process Mixture Model

- What if this assumption is wrong?
- Want to remove the parametric assumption on distribution of random effects by building a prior model on CDF $G$ that can be centered on $N(\mu, \sigma^2)$, but permits adaptation (learning from data).
- Specifying prior for an unknown distribution requires a stochastic process with realizations (sample paths) that are CDFs.
- We use Dirichlet process (DP), in notation $G \sim DP(\alpha, G_0)$, where $G_0$ is the center or base distribution of the process and $\alpha$ a precision parameter (Ferguson 1973, Antoniak 1974).
- Poisson DP mixture model:

\[
(y_i \mid \theta_i) \overset{iid}{\sim} \text{Poisson}(\exp(\theta_i)) \\
(\theta_i \mid G) \overset{iid}{\sim} G \\
G \sim DP(\alpha G_0), \quad G_0 \equiv G_0(\cdot; \psi),
\]

where $i = 1, \ldots, n$ (we refer to (2) as BNP model 1).
• **Equivalent formulation** of the Poisson DP mixture model:

\[
(y_i \mid G) \overset{iid}{\sim} f(\cdot; G) = \int \text{Poisson}(y_i; \exp(\theta))dG(\theta), \ G \sim \text{DP}(\alpha G_0),
\]

where \( i = 1, \ldots, n \) and \( G_0 = \text{N}(\mu, \sigma^2) \).

• MCMC implemented for a **marginalized** version of DP mixture.

**Key idea:** \( G \) is integrated out over its prior distribution (Antoniak 1974, Escobar and West 1995), resulting in \([\theta_1, \ldots, \theta_n \mid \alpha, \psi]\) that follows **Pólya urn** structure (Blackwell and MacQueen, 1973).

• **Specifically,** \([\theta_1, \ldots, \theta_n \mid \alpha, \psi]\) is

\[
g_{r0}(\theta_{r1} \mid \mu_r, \sigma_r^2) \prod_{i=2}^{n_r} \left\{ \frac{\alpha_r}{\alpha_r + i - 1} g_{r0}(\theta_{ri} \mid \mu_r, \sigma_r^2) + \frac{1}{\alpha_r + i - 1} \sum_{\ell=1}^{i-1} \delta_{\theta_{r\ell}}(\theta_{ri}) \right\}.
\]
DP Mixture Model with Stochastic Order

- There are cases when the treatment always has an effect, only the extent of which is unknown.

This can be expressed by introducing stochastic order for the random effects distributions: \( G_1(\theta) \geq G_2(\theta), \theta \in R \), denoted by \( G_1 \leq_{st} G_2 \).

- Posterior predictive inference can be improved under this assumption if we incorporate stochastic order in the model.

To that end we introduce a prior over the space \( \mathcal{P} = \{(G_1, G_2) : G_1 \leq_{st} G_2 \} \).

- A convenient way to specify such a prior is to work with subspace \( \mathcal{P}' \) of \( \mathcal{P} \), where \( \mathcal{P}' = \{(G_1, G_2) : G_1 = H_1, G_2 = H_1H_2 \} \), with \( H_1 \) and \( H_2 \) CDFs on \( R \), and then place independent DP priors on \( H_1 \) and \( H_2 \).

- Note: to obtain a sample \( \theta \) from \( G_2 = H_1H_2 \), independently draw \( \theta_1 \) from \( H_1 \) and \( \theta_2 \) from \( H_2 \), and then set \( \theta = \max(\theta_1, \theta_2) \).

- Specifying independent DP priors on the mixing distributions \( H_1 \) and \( H_2 \) we obtain the following model:
DPMM with Stochastic Order (continued)

\[
\begin{align*}
Y_{1i} \mid \theta_i & \overset{\text{ind}}{\sim} \text{Poisson}(\exp(\theta_i)), i = 1, n_1 \\
Y_{2k} \mid \theta_{1,n_1+k}, \theta_{2k} & \overset{\text{ind}}{\sim} \text{Poisson}(\exp(\max(\theta_{1,n_1+k}, \theta_{2k}))), k = 1, n_2 \\
\theta_{1i} \mid H_1 & \overset{\text{iid}}{\sim} H_1, i = 1, n_1 + n_2 \\
\theta_{2k} \mid H_2 & \overset{\text{iid}}{\sim} H_2, k = 1, n_2 \\
H_r \mid \alpha_r, \mu_r, \sigma_r^2 & \sim \text{DP}(\alpha_r H_{r0})
\end{align*}
\]

where the base distributions of the Dirichlet processes, \(H_{10}\) and \(H_{20}\), are again Normal with parametric priors on hyperparameters; we refer to (4) as BNP model 2.

- We implement a standard MCMC with an extension for stochastic order (Gelfand and Kottas, 2002).

- To create a level playing field to compare quality of PREP and BNP models we compute predictive distributions for future data, based on predictive distribution for latent variables and posterior parameter samples.
For BNP model 1 the posterior predictive for a future $Y^{\text{new}}$ is

$$[Y^{\text{new}} \mid \text{data}] = \int \int \text{Poisson}(Y^{\text{new}}; \exp(\theta^{\text{new}}))[\theta^{\text{new}} \mid \eta][\eta \mid \text{data}],$$

(5)

where $\theta^{\text{new}}$ is associated with $Y^{\text{new}}$ and $\eta$ collects all model parameters except $\theta$s (we use bracket notation of Gelfand and Smith (1990) to denote distributions).

The posterior predictive for latent variables, induced by the Pólya urn structure of the DP, is

$$[\theta^{\text{new}} \mid \eta] = \frac{\alpha}{\alpha + n} G_{r0}(\theta^{\text{new}} \mid \mu_r, \sigma^2) + \frac{1}{\alpha + n} \sum_{\ell=1}^{n} n_{\ell} \delta_{\theta_{\ell}}(\theta^{\text{new}}).$$

(6)
Simulation datasets for control (C) and treatment (T) (n = 300 observations in each), and distributions of latent variables (D1: C and T both Gaussian, D2: C skewed, T bimodal, D3: C bimodal, T Gaussian).

Bayesian Non-parametric Modeling With Skewed and Heavy-Tailed Data.
Prior (lower [blue] circles) and posterior (upper [red] circles) predictive distributions for PREP model (top) and BNP model 1 (bottom) for data set $D_3$ with bimodal random effects.

The PREP model cannot adapt to the bimodality (without remodeling as, e.g., a mixture of Gaussians on the latent scale), whereas the BNP modeling smoothly adapts to the data-generating mechanism.
• Perhaps more interestingly, using generic approach for inference about random mixing distribution, we can obtain \([G \mid \text{data}]\), from which we can compute the posterior of any linear functional of \(G\), e.g. \([E(y\mid G)]\).

• With \(G \sim DP(\alpha G_0)\), following Ferguson (1973) and Antoniak (1974),

\[
[G \mid \text{data}] = \int [G \mid \theta, \alpha, \psi] \text{d}[\theta, \alpha, \psi \mid \text{data}].
\]  

(7)

where \([G \mid \theta, \alpha, \psi]\) is also a DP with parameters \(\alpha' = \alpha + n\) and

\[
G'_0(\cdot|\psi) = \frac{\alpha}{\alpha + n} G_0(\cdot|\psi) + \frac{1}{\alpha + n} \sum_{i=1}^{n} 1_{(-\infty, \theta_i]}(\cdot),
\]  

(8)

where \(\theta = (\theta_1, ..., \theta_n)\) and \(\psi\) collects parameters of \(G_0\).

• Using (7), (8) and the definition of the DP we develop a computationally efficient approach to obtaining posterior sample paths from \([G \mid \text{data}]\).
Normal random effects (data set $D_1$): Posterior MCMC estimates of the random effects distributions for PREP model (first row) and BNP model 1 (second row); first column $C$, second column $T$.

When PREP is correct it (naturally) yields narrower uncertainty bands (but see below).
Skewed and bimodal random effects (data set $D_2$): Posterior MCMC estimates of random effects distributions for PREP model (first row) and BNP model 1 (second row); first column $C$, second column $T$.

When PREP is incorrect it continues to yield narrower uncertainty bands that unfortunately fail to include the truth, whereas BNP model 1 adapts successfully to the data-generating mechanism.
Bimodal random effects in $T$ (data set $D_3$): Posterior MCMC estimates of random effects distributions for BNP model 1 (first row) and BNP model with stochastic order (second row); first column $C$, second column $T$.

Extra assumption of stochastic order, when true, yields narrower uncertainty bands (as it should).


\(LS_{CV}\) and \(LS_{FS}\) For PREP and BNP Models

\(LS_{CV}\) (left panel) and full-sample log-score \(LS_{FS}\) (right panel) for PREP and BNP models for all 3 data sets (\(C\) and \(T\), \(D_{1,C}, \ldots, D_{3,T}\)).

When PREP is correct (1C, 1T, 3C), it has small advantage in \(LS_{CV}\) and \(LS_{FS}\) over BNP (as it should), but when PREP is incorrect (2C, 2T, 3T) both kinds of \(LS\) give a clear preference for BNP model 1 (also as they should).
Results in IHGA Case Study

Results on IHGA data in case study: posterior mean and 90% intervals for random-effects distribution $G$ (first column is $C$ sample, second column is $T$; first row is PREP model, second row is BNP model 1).

Uncertainty bands are wider from BNP model 1, but direct comparison not fair because PREP model arrived at via data-analytic search on entire data set.
Results on **IHGA data** in case study: first row is posterior for $C$ mean, second row is posterior for $T$ mean, third row is posterior for ratio of means; first column is PREP model, second column is BNP model 1: on the actual Hendriksen et al. data, PREP and BNP lead to **similar results** for quantity of greatest interest (ratio of $T$ and $C$ means).
**A Parametric Pitfall**

**Warning:** the Gaussian assumption on the latent variables scale in the PREP model can make the model look plausible when it’s not:

Top panel: bimodal mixture of Gaussians as true latent-variable distribution of $\theta_i = \log(\lambda_i)$; bottom panel: posterior means of $\theta_i$ values from PREP model ($n = 300$ observations).

Diagnostic checking of PREP model would make it look appropriate when it’s not; by contrast BNP correctly picks up the bimodality.
Case Study 2: Risk Assessment

Best plan (so far) for disposal of waste from nuclear power plants: deep underground storage (e.g., Yucca Mountain NV); need to estimate risk of contamination at Earth’s surface if something goes wrong.

Goal: Predicting outcomes, including radioactive dose for people on the Earth’s surface, as a function of factors like time, how far the disposal chamber is underground, ...

Radioactive dose is estimated by computer simulation models, which numerically solve systems of partial differential equations modeling diffusion of radionuclides through rock fissures.

Output of such models is deterministic given inputs, but inputs not fully known; standard approach is to simulate from (prior) distributions of inputs and study output dose distributions.

Regulatory bodies insist on summarizing dose distribution $f$ at a given point in space-time by its mean $\theta = \int y f(y) \, dy$ and variance $\nu = \int (y - \theta)^2 f(y) \, dy$ (even though these may be highly unstably estimated quantities).
Risk Assessment (continued)

Technical challenge: $f$ is typically extremely (positively) skewed, with many zeros and a few comparatively huge values, and the number of Monte Carlo repetitions $N$ is constrained by time and money (often $\leq 10,000$, sometimes $\leq 500–1000$).

With relatively small $N$, the concern is that you haven’t seen all of the right tail yet.

Consider $N = 10,000$ dose values from computer program at $t = 100$ years, based on scenario permitting relatively large doses of Strontium 90 (Sr–90) with relatively low probability; the outcome examined is total dose from three nuclides including Sr–90.

9864 (98.6%) of the 10,000 values are 0; 134 of the other 136 (1.36%) range smoothly from $1.059e^{-14}$ to $8.552e{-01}$; the two largest values are 3.866 and 189.3 (!); the sample mean is 0.01964.

(The true mean at 100 years, obtained from another computer program (in this special case), is $9.382e{-4}$ (21 times smaller); the sample mean omitting the largest observation is $7.138e{-4}$.)
Normal qqplot of positive log dose values shows that plausible parametric model is mixture:

\[
(y_i | \pi, \mu, \sigma^2) \sim \begin{cases} 
0 & \text{with probability } \pi \\
LN(\mu, \sigma^2) & \text{with probability } (1 - \pi)
\end{cases}
\]
Non-Parametric Bootstrap

Standard **inferential** technology in use in **European regulatory work** is **non-parametric bootstrap**; this got me interested in **performance** of bootstrap confidence intervals for **population variance** when data-generating mechanism is log-normal.

**Bootstrap propaganda:**

One of the principal goals of bootstrap theory is to produce good confidence intervals automatically. Good means that the bootstrap intervals should closely match exact confidence intervals in those special situations where statistical theory yields an exact answer, and should give dependably accurate coverage probabilities in all situations. ... [The BC$_a$ and ABC intervals] come close to [these] criteria of goodness (Efron and Tibshirani, 1993).

**Non-parametric bootstrap** based on **empirical CDF** $\hat{F}_N$; if substantial tail mass is unobserved in sample of size $N$, bootstrap cannot perform well for functionals like **means** and **variances**; **surprise** is how badly it performs even with substantial $N$. 
Consider sample \( y = (y_1, \ldots, y_n) \) from model \( F = LN(0, 1), (Y_1, \ldots, Y_n | F) \overset{IID}{\sim} F \), and suppose functional of \( F \) of interest is
\[ V(F) = \int [y - E(F)]^2 dF(y), \text{ where } E(F) = \int y dF(y). \]

Usual unbiased sample variance \( s^2 \) is (almost) nonparametric MLE of \( V(F) \), and it serves as basis of ABC bootstrap intervals; population value for \( V(F) \) with \( LN(0, 1) \) is \( e(e - 1) = 4.67 \).

Distribution of sample variance for \( LN(0, 1) \) data, based on 1000 simulation repetitions (simulation SE in parentheses):

<table>
<thead>
<tr>
<th>( n )</th>
<th>mean</th>
<th>median</th>
<th>% &lt; 4.67</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4.66 (0.49)</td>
<td>1.88 (0.09)</td>
<td>77.1 (1.2)</td>
<td>9.43 (0.75)</td>
</tr>
<tr>
<td>20</td>
<td>4.68 (0.34)</td>
<td>2.52 (0.09)</td>
<td>74.0 (1.4)</td>
<td>9.37 (0.60)</td>
</tr>
<tr>
<td>50</td>
<td>4.68 (0.21)</td>
<td>3.20 (0.08)</td>
<td>70.4 (1.5)</td>
<td>8.59 (0.43)</td>
</tr>
<tr>
<td>100</td>
<td>4.67 (0.15)</td>
<td>3.62 (0.08)</td>
<td>67.6 (1.5)</td>
<td>7.98 (0.31)</td>
</tr>
<tr>
<td>500</td>
<td>4.68 (0.07)</td>
<td>4.23 (0.05)</td>
<td>62.6 (1.4)</td>
<td>6.64 (0.13)</td>
</tr>
</tbody>
</table>

\( s^2 \) achieves unbiasedness by being too small most of the time and much too large some of the time; does not bode well for bootstrap.
ABC method, nominal 90\% intervals, $LN(0, 1)$ data:

<table>
<thead>
<tr>
<th>$n$</th>
<th>actual cov. (%)</th>
<th>mean length</th>
<th>median length</th>
<th>% on left</th>
<th>% on right</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>36.0 (1.5)</td>
<td>8.08 (0.68)</td>
<td>2.43 (1.13)</td>
<td>61.1 (1.5)</td>
<td>2.9 (0.5)</td>
</tr>
<tr>
<td>20</td>
<td>49.4 (1.6)</td>
<td>9.42 (0.77)</td>
<td>3.79 (1.32)</td>
<td>48.6 (1.6)</td>
<td>2.0 (0.4)</td>
</tr>
<tr>
<td>50</td>
<td>61.9 (1.5)</td>
<td>10.1 (0.61)</td>
<td>4.56 (0.75)</td>
<td>35.4 (1.5)</td>
<td>2.7 (0.5)</td>
</tr>
<tr>
<td>100</td>
<td>68.6 (1.5)</td>
<td>8.76 (0.49)</td>
<td>4.63 (0.72)</td>
<td>27.4 (1.4)</td>
<td>4.0 (0.6)</td>
</tr>
<tr>
<td>500</td>
<td>76.8 (1.3)</td>
<td>5.43 (0.21)</td>
<td>3.51 (0.27)</td>
<td>18.5 (1.2)</td>
<td>4.7 (0.7)</td>
</tr>
</tbody>
</table>

With $n = 10$, nominal 90\% intervals only cover 36\% of the time, and even with $n = 500$ actual coverage is only up to 77\%!

Mistakes are almost always from interval lying entirely to left of true population variance; to improve must bring in prior information about tail weight and skewness.
Naive Parametric Bayesian Approach

Problem can of course be solved \textbf{parametrically}: consider model

\[(\mu, \sigma^2) \sim p(\mu, \sigma^2)\]

\[(Y_1, \ldots, Y_n | \mu, \sigma^2) \overset{\text{IID}}{\sim} LN(\mu, \sigma^2),\]

and take proper but highly diffuse prior on \((\mu, \sigma^2)\); easy to use Gibbs sampling to show that Bayesian intervals are \textbf{well-calibrated} (but note interval lengths!):

\textit{Lognormal model, }N(0, 10^4)\textbf{ prior for }\mu, \Gamma(0.001, 0.001)\textbf{ prior for }\tau = \frac{1}{\sigma^2}, \textbf{ nominal 90\%}, LN(0, 1)\textbf{ data (simulation SEs in parentheses)}

<table>
<thead>
<tr>
<th>n</th>
<th>actual cov. (%)</th>
<th>mean length</th>
<th>median length</th>
<th>% on left</th>
<th>% on right</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>88.7 (1.0)</td>
<td>6 \cdot 10^5 (2 \cdot 10^5)</td>
<td>194.6 (29.3)</td>
<td>4.9 (0.7)</td>
<td>6.4 (0.8)</td>
</tr>
<tr>
<td>20</td>
<td>89.3 (1.0)</td>
<td>145.2 (14.9)</td>
<td>39.6 (2.0)</td>
<td>4.9 (0.7)</td>
<td>5.8 (0.7)</td>
</tr>
<tr>
<td>50</td>
<td>89.1 (1.0)</td>
<td>17.8 (0.5)</td>
<td>12.6 (0.5)</td>
<td>5.3 (0.7)</td>
<td>5.6 (0.7)</td>
</tr>
<tr>
<td>100</td>
<td>90.6 (0.9)</td>
<td>8.9 (0.2)</td>
<td>7.6 (0.2)</td>
<td>4.0 (0.6)</td>
<td>5.4 (0.7)</td>
</tr>
<tr>
<td>500</td>
<td>89.9 (1.0)</td>
<td>3.0 (0.02)</td>
<td>3.0 (0.02)</td>
<td>5.5 (0.7)</td>
<td>4.6 (0.7)</td>
</tr>
</tbody>
</table>
But parametric Bayesian inference based on LN distribution is horribly non-robust (here naive means using Lognormal model when data-generating mechanism is far from Lognormal):

Lognormal model, $N(0, 10^4)$ prior for $\mu$, $\Gamma(0.001, 0.001)$ prior for $\tau = \frac{1}{\sigma^2}$, nominal 90%, $N(0, 10)$ data (simulation SEs in parentheses)

<table>
<thead>
<tr>
<th>$n$</th>
<th>actual cov. (%)</th>
<th>actual mean</th>
<th>mean length</th>
<th>% on left</th>
<th>% on right</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0 (0.0)</td>
<td>6.851 (0.03)</td>
<td>0.0 (0.0)</td>
<td>100.0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>5.1 (0.7)</td>
<td>2.542 (0.01)</td>
<td>0.0 (0.0)</td>
<td>94.9 (0.7)</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>44.2 (1.6)</td>
<td>0.990 (0.004)</td>
<td>0.0 (0.0)</td>
<td>55.8 (1.6)</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>64.1 (1.5)</td>
<td>0.586 (0.002)</td>
<td>0.0 (0.0)</td>
<td>35.9 (1.5)</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>85.3 (1.2)</td>
<td>0.221 (0.0004)</td>
<td>1.1 (0.3)</td>
<td>13.6 (1.1)</td>
<td></td>
</tr>
</tbody>
</table>

Need to bring in tail-weight and skewness nonparametrically.
Dirichlet Process Mixture Modeling

We’ve recently begun to experiment with using the Location Normal Dirichlet Process Mixture Model (LNDPMM) to create interval estimates for the mean and variance functionals (on the data scale, not on the scale of the latent variables) which we hope will be reasonably well-calibrated across a broad range of data-generating mechanisms.

Specifically, with \((w_i, i = 1, \ldots, n)\) as the positive sample of radiation doses and \(y_i = \ln(w_i)\), the model is

\[
(y_i | \theta_i, \phi) \overset{\text{indep}}{\sim} N(\theta_i, \phi) \\
(\theta_i | G) \overset{\text{IID}}{\sim} G \\
(G | \alpha, \mu, \tau^2) \sim DP(\alpha G_0), \quad G_0 = N(\mu, \tau^2) \\
(\alpha, \mu, \tau^2, \phi) \sim p(\alpha) p(\mu) p(\tau^2) p(\phi),
\]

with the standard (in most cases conditionally-conjugate) priors:

\[
\alpha \sim \Gamma(a_\alpha, b_\alpha), \quad \mu \sim N(a_\mu, b_\mu), \quad \tau^2 \sim \Gamma^{-1}(a_{\tau^2}, b_{\tau^2}), \quad \phi \sim \Gamma^{-1}(a_\phi, b_\phi).
\]
One approach to **posterior inference** about **mean** and **variance** functionals on **raw** (radioactive dose) scale $W$ would be to use quantities from standard stick-breaking constructive definition of DP to derive interval estimates for mean and variance on **log scale** $Y = \ln(W)$, but **not straightforward** to transform these back to raw scale.

Instead: LNDPMM produces samples in **standard way** from density $f_Y(y)$ on log scale, and (by usual **change-of-variables**) **raw-scale density** is $f_W(w) = f_Y[\ln(W)] \cdot \frac{1}{W}$; can therefore MCMC-approximate $f_W$ at **grid** of points $\{w_1, \ldots, w_M\}$ and use this as **discretized approximation** to compute

$$E(f_W) = \sum_{j=1}^{M} w_j f_W(w_j) \quad \text{and} \quad V(f_W) = \sum_{j=1}^{M} w_j^2 f_W(w_j) - [E(f_W)]^2.$$  

Some **care** required in choosing **grid** to **span data scale correctly** and **minimize** (what amounts to) numerical integration error.

LNDPMM is like **machine** with **eight dials** — governed by hyperparameter values $(a_\alpha, b_\alpha, a_\mu, b_\mu, a_\tau, b_\tau, a_\phi, b_\phi)$ — that (we hope) can be **tuned** to produce **good calibration** across many underlying population shapes.
LNDPMM Results for Variance Functional

Standard practice to use data to gently help LNDPMM with center and scale: in all runs so far we used \((a_\mu = \bar{y}, b_\mu = \left(\frac{y_r}{4}\right)^2\) (sample mean and range \(\bar{y}\) and \(y_r\), respectively), \((a_\tau = 2, b_\tau = \frac{y_r}{4}\)) and \((a_\phi = 2, b_\phi = \frac{y_r}{4}\)); \(R = \) number of simulation replications; recall that bootstrap coverage with LN data was 36\% with \(n = 10\) and 62\% with \(n = 50\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(R)</th>
<th>Population</th>
<th>(a_\alpha)</th>
<th>(b_\alpha)</th>
<th>Actual Coverage (%)</th>
<th>Mean Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>250</td>
<td>LN(0,1)</td>
<td>2.0</td>
<td>4/n = 0.40</td>
<td>88</td>
<td>36.9</td>
</tr>
<tr>
<td>10</td>
<td>250</td>
<td>LN(1,1)</td>
<td>2.0</td>
<td>4/n = 0.40</td>
<td>89</td>
<td>365.8</td>
</tr>
<tr>
<td>50</td>
<td>250</td>
<td>LN(0,1)</td>
<td>2.0</td>
<td>4/n = 0.08</td>
<td>95</td>
<td>20.0</td>
</tr>
<tr>
<td>50</td>
<td>250</td>
<td>LN(1,1)</td>
<td>2.0</td>
<td>4/n = 0.08</td>
<td>95</td>
<td>126.0</td>
</tr>
<tr>
<td>50</td>
<td>145</td>
<td>LN(0,1)</td>
<td>4.0</td>
<td>8/n = 0.16</td>
<td>95</td>
<td>20.7</td>
</tr>
<tr>
<td>50</td>
<td>112</td>
<td>LN(0,1)</td>
<td>1.0</td>
<td>8/n = 0.16</td>
<td>96</td>
<td>20.5</td>
</tr>
<tr>
<td>50</td>
<td>121</td>
<td>N(200,400)</td>
<td>2.0</td>
<td>4/n = 0.40</td>
<td>99</td>
<td>2447</td>
</tr>
</tbody>
</table>

Plots show that population distribution is recovered well on average; next step: hierarchically weaken dependence of prior on \(\bar{y}\) and \(y_r\).