Bayesian Estimation of $\log N - \log S$

Irina S Udaltsova

Department of Statistics
University of California, Davis

February 17th, 2015

Flux Distribution

Probabilistic Connection: Under independent sampling, linearity on the $\log N - \log S$ scale is equivalent to the flux distribution being a Pareto distribution.

$$\log_{10}(1 - F_G(s)) = \beta_0 + \theta \log_{10}(S)$$

(Follows from log-linearity of the survival function)

The Pareto representation for the flux distribution now allows to build a framework that accounts for many sources of uncertainty and can be embedded into a hierarchical model.

Broken Power-Law Model

Correspondence between power-law and flux distribution extends to broken power-law:

$$\log_{10}\left(1 - F_G(s)\right) = \begin{cases} \alpha_0 - \theta_0 \log_{10}(s) & s \leq \tau \\ \alpha_1 - \theta_1 \log_{10}(s) & s > \tau \end{cases}, \qquad (1)$$

subject to a continuity constraint.

$$Y \sim I \cdot X_1 + (1 - I) \cdot X_2,$$

where:

$$I \sim \mathrm{Bin}\left(1, \left[1 - \left(\frac{\tau}{\tau}\right)^{-\theta_1}\right]\right)$$

$$X_1 \sim \text{Truncated-Pareto}\left(\tau, \theta_1, \tau\right), \qquad X_2 \sim \text{Pareto}\left(\tau, \theta_2\right).$$

Multiple Broken Power-Law Model

The broken power-law model can be generalized to a piece-wise linear relationship with arbitrary number, $m-1$, of break-points:

$$F_G(s) = \begin{cases} 1-\alpha_1^* s^{-\theta_1} & \tau_1 \leq s < \tau_2 \\ \vdots & \vdots \\ 1-\alpha_{m-1}^* s^{-\theta_{m-1}} & \tau_{m-1} \leq s < \tau_m \\ 1-\alpha_m^* s^{-\theta_m} & s \geq \tau_m \end{cases}.$$

Constraints on the mixture probabilities give the multiple broken Pareto:

$$Y \sim I_1 X_1 + I_2 X_2 + \cdots + I_m X_m$$

$$I_j \sim \mathrm{Multinomial}\left(1; p_1, p_2, \ldots, p_m\right), \quad X_j \sim \mathrm{Truncated\text{-}Pareto}\left(\tau_j, \theta_j, \tau_{j+1}\right).$$

Additional Pieces

The previous discussion centered around the flux distribution.

Many sources of uncertainty:

- We only observe photon counts from the source with intensity proportional to the flux
- There is background contamination for all sources
- Different sensitivities across the detector
- Some sources will not be observed to detector limitations
- We do not know how many sources there actually are
- Some 'sources' extracted from the image may not actually be sources

To Model

Missing Data

There are many potential causes of missing data in astronomical data:

- Low-count sources (below detection threshold)
- Detector schedules (source not within detector range)
- Background contamination (e.g., total=source+background)
- Foreground contamination (other objects between the source and detector)
- etc.

Important: Whether a source is observed is a function of its source count (intensity), which is unobserved for unobserved sources – missing data mechanism is non-ignorable, and needs to be carefully accounted for in the analysis.

Non-Ignorable Missingness

Let $Y_{com} = (Y_{obs}, Y_{mis})$, and the missing data indicators be M. The missing data mechanism is defined to be $p(M|Y_{com}, \psi)$.

The observed data likelihood is based on:

$$p(Y_{obs}|\theta) = \int p(Y_{com}|\theta) dY_{mis}. \tag{2}$$

The complete data likelihood is based on:

$$p(Y_{obs}, M|\theta, \psi) = \int p(Y_{com}|\theta) p(M|Y_{com}, \psi) dY_{mis}. \tag{3}$$

Inference based on (2) is valid only if inference about θ agrees with that from (3). In these cases, the missing data mechanism is called ignorable.

Main condition for ignorability is the data be missing at random (MAR):

$$p(M|Y_{com}, \psi) = p(M|Y_{obs}, \psi) \qquad \forall \quad Y_{mis}, \psi. \tag{4}$$

Here this is not true! Missingness depends on the unobserved flux.

The Data

The data is a list of photon counts – with some extra information about the background and detector properties.

Src_ID	Counts	Bgr_intensity	Src_area	Off_axis	Effective_area
1	1093	47.38195	466	6.18	383.609
2	927	16.40961	180	5.75	392.709
3	31	12.66816	126	4.43	396.570
4	5	1.155294	12	0.48	278.892
5	286	17.50082	190	5.82	345.492
6	469	44.74188	436	5.36	358.845

- ...and an incompleteness function, specifying the probability of source detection under a range of conditions:

$$\mathbb{P}(\text{ Detecting a source with flux S, background intensity B, location L and effective area E }) \equiv g(S, B, L, E)$$

The Model

Standard power-law flux distribution:

$$S_i|\tau,\theta \overset{iid}{\sim} \mathrm{Pareto}\left(\theta,\tau\right), i = 1,\ldots,N.$$

Source and background photon counts:

$$Y_i^{tot}|S_i,B_i,L_i,E_i \overset{indep}{\sim} \mathrm{Pois}\left(\lambda(S_i,B_i,L_i,E_i) + k(B_i,L_i,E_i)\right), i = 1,\ldots,N,$$

Incompleteness, missing data indicators:

$$I_i \sim \mathrm{Bernoulli}\left(g\left(S_i,B_i,L_i,E_i\right)\right).$$

Prior distributions:

$$\begin{aligned} N &\sim \mathrm{NegBinom}\left(a_N,b_N\right), \\ \theta &\sim \mathrm{Gamma}(a,b), \\ \tau &\sim \mathrm{Gamma}(a_m,b_m). \end{aligned}$$

Broken Power-Law Model

Broken power-law flux distribution:

$$S_i|\tau, \theta \overset{iid}{\sim} \text{Broken-Pareto}\left(\vec{\theta}, \vec{\tau}\right), i = 1, \ldots, N.$$

Source and background photon counts:

$$Y_i^{tot}|S_i, B_i, L_i, E_i \overset{indep}{\sim} \text{Pois}\left(\lambda(S_i, B_i, L_i, E_i) + k(B_i, L_i, E_i)\right), i = 1, \ldots, N,$$

Incompleteness, missing data indicators:

$$I_i \sim \text{Bernoulli}\left(g\left(S_i, B_i, L_i, E_i\right)\right).$$

Prior distributions:

$$N \sim \text{NegBinom}\left(a_N, b_N\right),$$

$$\theta_j \overset{indep}{\sim} \text{Gamma}(a_j, b_j), j = 1, \ldots, m,$$

$$\tau_1 \sim \text{Gamma}(a_m, b_m)$$

$$\tau_j = \tau_1 + \sum_{k=2}^{j} e^{\eta_k}, \ \ \eta_j \overset{indep}{\sim} \text{Normal}(\mu_j, c_j), j = 2, \ldots, m.$$

Model Overview

Unusual points and important notes:

- The dimension of the missing data is unknown (care must be taken with conditioning)
- Incompleteness function g can take any form and is problem-specific
- The flux lower limit and break-points, $\tau_j, j = 1, \ldots, m$, can be estimated
- Prior parameters can be science-based, i.e., 'weakly informative'

Posterior Inference (Single Pareto)

Inference about θ, N, S, τ is based on the *observed data* posterior distribution. Care must be taken with the variable dimension marginalization over the unobserved fluxes.

The single power-law posterior can be shown to be:

$$p\left(N, \theta, \tau, S_{obs}, Y_{obs}^{src} | n, Y_{obs}^{tot}, B_{obs}, L_{obs}, E_{obs}\right)$$

$$\propto \begin{pmatrix} N \\ n \end{pmatrix} \mathbb{I}_{\{n \leq N\}} \cdot (1 - \pi(\theta, \tau))^{(N-n)} \cdot \begin{pmatrix} N + a_N - 1 \\ a_N - 1 \end{pmatrix} \left(\frac{1}{1 + b_N}\right)^N \left(\frac{b_N}{1 + b_N}\right)^{a_N} \mathbb{I}_{\{N \in \mathbb{Z}^+\}} \cdot$$
$$\cdot \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \mathbb{I}_{\{\theta > 0\}} \cdot \frac{b_m^{a_m}}{\Gamma(a_m)} \tau^{a_m - 1} e^{-b_m \tau} \mathbb{I}_{\{\tau > 0\}} \cdot \left[\prod_{i=1}^{n} p\left(B_i, L_i, E_i\right) \cdot \theta \tau^{\theta} S_i^{-(\theta+1)} \mathbb{I}_{\{\tau < S_i\}} \cdot\right.$$
$$\cdot g(S_i, B_i, L_i, E_i) \cdot \frac{(\lambda_i + k_i)^{Y_i^{tot}}}{Y_i^{tot}!} e^{(\lambda_i + k_i)} \mathbb{I}_{\{Y_i^{tot} \in \mathbb{Z}^+\}}$$
$$\left.\cdot \begin{pmatrix} Y_i^{tot} \\ Y_i^{src} \end{pmatrix} \left(\frac{\lambda_i}{\lambda_i + k_i}\right)^{Y_i^{src}} \left(1 - \frac{\lambda_i}{\lambda_i + k_i}\right)^{Y_i^{tot} - Y_i^{src}} \mathbb{I}_{\left\{Y_i^{src} \in \{0, 1, \ldots, Y_i^{tot}\}\right\}}\right]$$

with $\lambda_i \equiv \lambda(S_i, B_i, L_i, E_i)$ and $k_i \equiv k(B_i, L_i, E_i)$.

Posterior Inference (Broken-Pareto)

The broken power-law posterior can be shown to be:

$$p\left(N, \theta, \tau, S_{obs}, Y_{obs}^{src} | n, Y_{obs}^{tot}, B_{obs}, L_{obs}, E_{obs}\right)$$

$$= \frac{1}{p(n, Y_{obs}^{tot}, B_{obs}, L_{obs}, E_{obs})} \cdot \left[\binom{N}{n} \mathbb{I}_{\{n \leq N\}} \cdot (1 - \pi(\theta, \tau))^{(N-n)} \right]$$

$$\cdot \left[\binom{N + a_N - 1}{a_N - 1} \left(\frac{1}{1 + b_N}\right)^N \left(\frac{b_N}{1 + b_N}\right)^{a_N} \mathbb{I}_{\{N \in \mathbb{Z}^+\}} \right] \cdot \left[\prod_{j=1}^{m} \frac{b_j^{a_j}}{\Gamma(a_j)} \theta_j^{a_j - 1} e^{-b_j \theta_j} \mathbb{I}_{\{\theta_j > 0\}} \right]$$

$$\cdot\, p\left(\tau_1, \ldots, \tau_m\right) \mathbb{I}_{\{0 < \tau_1 < \tau_1 < \cdots < \tau_m\}} \cdot \left[\prod_{i=1}^{n} p\left(B_i, L_i, E_i\right) \cdot g(S_i, B_i, L_i, E_i) \right.$$

$$\cdot \sum_{j=1}^{m} \left\{ \prod_{l=1}^{j-1} \left(\frac{\tau_{l+1}}{\tau_l}\right)^{-\theta_l} \right\} \left(\frac{\theta_j}{\tau_j}\right) \left(\frac{S_i}{\tau_j}\right)^{-(\theta_j + 1)} \cdot \mathbb{I}_{\left\{\tau_j \leq S_i < \tau_{j+1}\right\}}$$

$$\cdot \frac{(\lambda_i + k_i)^{Y_i^{tot}}}{Y_i^{tot}!} e^{(\lambda_i + k_i)} \mathbb{I}_{\{Y_i^{tot} \in \mathbb{Z}^+\}}$$

$$\left. \cdot \binom{Y_i^{tot}}{Y_i^{src}} \left(\frac{\lambda_i}{\lambda_i + k_i}\right)^{Y_i^{src}} \left(1 - \frac{\lambda_i}{\lambda_i + k_i}\right)^{Y_i^{tot} - Y_i^{src}} \mathbb{I}_{\left\{Y_i^{src} \in \{0, 1, \ldots, Y_i^{tot}\}\right\}} \right]$$

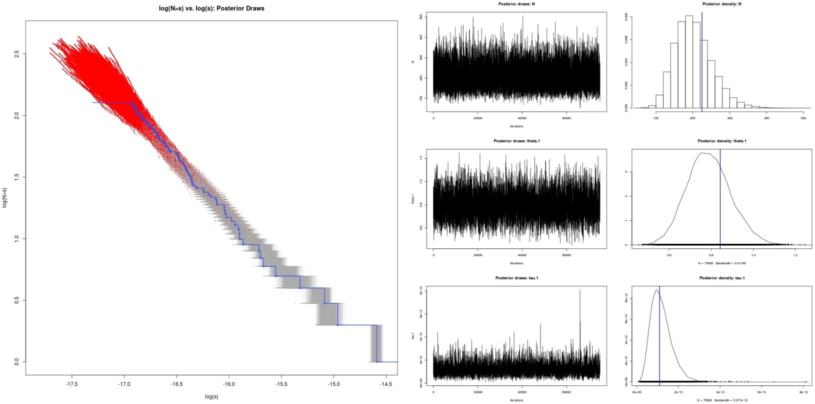
with $\tau_{m+1} = +\infty$, $\lambda_i \equiv \lambda(S_i, B_i, L_i, E_i)$, $k_i \equiv k(B_i, L_i, E_i)$, and $\prod_{i=1}^{0} \left(\frac{\tau_{i+1}}{\tau_i}\right)^{-\theta_i} = 1$.

Computation Strategy: Gibbs Sampler

The Gibbs sampler consists of five steps:

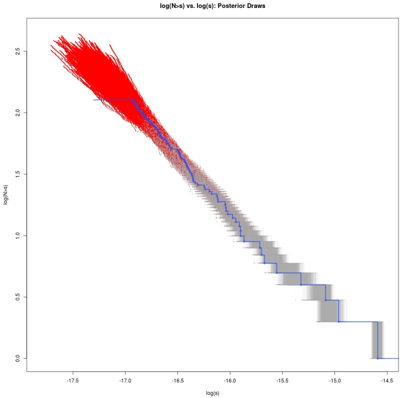
$$[N|n,\theta], \quad [\theta|n,N,S_{obs},\tau], \quad [\tau|n,N,\theta,S_{obs},B_{obs},L_{obs},E_{obs}],$$
$$\left[S_{obs}|N,\theta,\tau,I_{obs},Y^{tot}_{obs},Y^{src}_{obs},B_{obs},L_{obs},E_{obs}\right],$$
$$\left[Y^{src}_{obs}|Y^{tot}_{obs},B_{obs},L_{obs},E_{obs},I_{obs},S_{obs}\right].$$

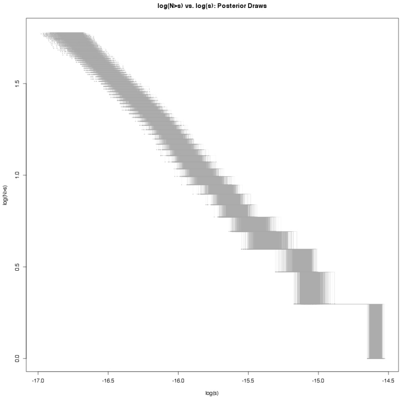
Example MCMC Output



(L) Posterior logN-logS (red: missing, gray: observed), truth (blue).
(R) Posterior distributions for N, θ, τ

Non-Ignorable Missingness





(L) Nonignorable (full) analysis: $\hat{\theta} = 0.990,\ (0.803, 1.192)$

(R) Ignorable analysis: $\hat{\theta} = 0.784,\ (0.520, 0.978)$

Truth: $\theta = 0.986$

▸ To Model Checking

Computational Details

The Gibbs sampler consists of five steps:

$$[N|n,\theta]\,,\quad [\theta|n,N,S_{obs},\tau]\,,\quad [\tau|n,N,\theta,S_{obs},B_{obs},L_{obs},E_{obs}]\,,$$
$$\left[S_{obs}|N,\theta,\tau,I_{obs},Y_{obs}^{tot},Y_{obs}^{src},B_{obs},L_{obs},E_{obs}\right]\,,$$
$$\left[Y_{obs}^{src}|Y_{obs}^{tot},B_{obs},L_{obs},E_{obs},I_{obs},S_{obs}\right]\,.$$

- The marginal probability of observing a source $\pi(\theta,\tau)$ is pre-computed via the numerical integration.

$$\pi(\theta,\tau)=\int g(S_i,B_i,L_i,E_i)\cdot p(S_i|\theta,\tau)\cdot p(B_i,L_i,E_i)\ dS_i\ dB_i\ dL_i\ dE_i.$$

- Sample the total number of sources, N, (Numerical Integration):

$$\begin{aligned} p\left(N|\cdot\right) &\propto \binom{N}{n}\mathbb{I}_{\{n\le N\}}\cdot(1-\pi(\theta,\tau))^{(N-n)}\cdot p\left(N\right)\cdot p\left(S_{obs}|N,\theta,\tau\right) \\ &\propto \frac{\Gamma(N+a_N)}{\Gamma(N-n+1)}\cdot\left(\frac{1}{b_N+1}\right)^{N}\cdot\,(1-\pi(\theta,\tau))^{(N-n)}\,\mathbb{I}_{\{n\le N\}} \end{aligned}$$

Computational Details cont...

- Sample the power-law slope, θ, (Metropolis Hastings using a Normal-proposal):

$$p\left(\theta|\cdot\right) \propto p\left(\theta\right)\cdot p\left(S_{obs}|N,\theta,\tau\right)\cdot\ \left(1-\pi(\theta,\tau)\right)^{(N-n)}$$
$$\propto \left(1-\pi(\theta,\tau)\right)^{(N-n)}\cdot \mathrm{Gamma}\left(\theta; a+n, b+\sum_{i=1}^{n}\log\left(\frac{S_i}{\tau}\right)\right)$$

- Sample the observed photon counts $Y^{src}_{obs,i}$, $i=1,\ldots,n$:

$$p\left(Y_i^{src}|\cdot\right) \propto p(Y_i^{src}|Y_i^{tot},S_i,B_i,L_i,E_i)$$
$$\sim \mathrm{Bin}\left(Y_i^{src}; Y_i^{tot}, \frac{\lambda(S_i,B_i,L_i,E_i)}{\lambda(S_i,B_i,L_i,E_i)+k(B_i,L_i,E_i)}\right)$$

Computational Details cont...

- Sample the minimum flux τ (Metropolis Hastings using a Truncated-Normal-proposal after log-transformation):

$$
\begin{aligned}
p(\tau|\cdot) &\propto p(\tau)\cdot p(\theta)\cdot p(N)\cdot p\left(B_{obs}, L_{obs}, E_{obs}\right) \\
&\quad \cdot p\left(n, S_{obs}, I_{obs}|N, \theta, \tau, B_{obs}, L_{obs}, E_{obs}\right) \\
&\propto \mathrm{Gamma}\left(\tau; a_m + n\theta, b_m\right)\cdot\left(1-\pi(\theta,\tau)\right)^{N-n}\cdot \mathbb{I}_{\{0<\tau<c_m\}}
\end{aligned}
$$

$$
\begin{aligned}
p(\eta = \log(\tau)|\cdot) &= e^{\eta}\cdot p(\tau = e^{\eta}|\cdot) \\
&\propto e^{\eta(n\theta+a_m+1)}\cdot e^{-b_m e^{\eta}}\cdot\left(1-\pi(\theta,\tau=e^{\eta})\right)^{N-n}\cdot\mathbb{I}_{\{\eta<\log(c_m)\}}.
\end{aligned}
$$

- Sample the fluxes $S_{obs,i}, i = 1,\ldots,n$ (Metropolis Hastings using a Normal-proposal):

$$
\begin{aligned}
p\left(S_i|\cdot\right) &\propto p\left(S_i|N,\theta,\tau\right)\cdot p\left(I_i = 1|S_i, B_i, L_i, E_i\right)\cdot p\left(Y_i^{tot}|S_i, B_i, L_i, E_i\right)\cdot \\
&\quad \cdot p\left(Y_i^{src}|Y_i^{tot}, S_i, B_i, L_i, E_i\right) \\
&\sim Pareto(S_i;\theta,\tau)\cdot g(S_i, B_i, L_i, E_i)\cdot Pois(Y_i^{tot};\lambda(S_i, B_i, L_i, E_i)+k(B_i, L_i, E_i)) \\
&\quad \cdot \mathrm{Bin}\left(Y_i^{src}; Y_i^{tot}, \frac{\lambda(S_i, B_i, L_i, E_i)}{\lambda(S_i, B_i, L_i, E_i)+k(B_i, L_i, E_i)}\right)
\end{aligned}
$$

Computational Details cont...

- Sample $\theta = (\theta_1, \ldots, \theta_m)^T$: (Metropolis-Hastings using a Normal-proposal)

$$p(\theta | \cdot) \propto \left[(1 - \pi(\theta, \tau))^{N-n} \right] \cdot$$

$$\cdot \prod_{j=1}^{m} \mathrm{Gamma} \left(\theta_j;\ a_j + n(j) - 1,\ b_j + \mathbb{I}_{\{j \neq m\}} \log \left(\frac{\tau_{j+1}}{\tau_j} \right) \sum_{i=1}^{m} \left[n(i) \mathbb{I}_{\{i \geq j+1\}} \right] + \sum_{i \in \mathcal{I}(j)} \log \left(\frac{s_i}{\tau_j} \right) \right),$$

where $\mathcal{I}(j) = \{i : \tau_j \leq s_i < \tau_{j+1}\}$ and $n(j)$ is the cardinality of $\mathcal{I}(j)$ i.e., $\mathcal{I}(j)$ ($n(j)$) denotes the set (number) of source indices whose flux is contained in the interval corresponding to the j-th mixture component.

- Sample the break-points $\tilde{\tau} = (\tau_2, \ldots, \tau_m)^T$ (Metropolis-Hastings based on original transformed scale $\eta_j = h(\tilde{\tau} | \tau_1) = \log(\tau_j - \tau_{j-1}), j = 2, \ldots, m$):

$$p(\eta | \cdot) = p(h(\tilde{\tau} | \tau_1) | \cdot) \propto \left[(1 - \pi(\theta, \tau))^{(N-n)} \right] \cdot \exp \left[-\frac{1}{2} \sum_{j=2}^{m} \{ c_j (\eta_j - \mu_j) \}^2 \right] \cdot \mathbb{I}_{\{\tau_1 < \tau_2 < \cdots < \tau_m\}}$$

$$\cdot \left[\prod_{j=1}^{m} \left\{ \prod_{l=1}^{j-1} \left(\frac{\tau_{l+1}}{\tau_l} \right)^{-\theta_l} \right\}^{n(j)} \prod_{i \in \mathcal{I}(j)} \left(\frac{\theta_j}{\tau_j} \right) \left(\frac{s_i}{\tau_j} \right)^{-(\theta_j + 1)} \mathbb{I}_{\{\tau_1 < \min(s_1, \ldots, s_n)\}} \right],$$

Computational Details cont...

Fluxes of missing sources can (optionally) be imputed to produce posterior draws of a 'corrected' $\log N - \log S$

- Impute missing fluxes $S_{mis,i}, i = 1, \ldots, n$ (Rejection Sampling):

$$
\begin{aligned}
&(B_i, L_i, E_i) \sim p(B_i, L_i, E_i) \\
&S_i | n, N, \theta, \tau, B_i, L_i, E_i, I_i = 0 \\
&\qquad \sim (1 - g(S_i, B_i, L_i, E_i)) \cdot Pareto(S_i; \theta, \tau).
\end{aligned}
$$

Model Checking

Bayesian model checking is designed around posterior predictive distribution:

$$p(y^*|y) = \int p(y^*, \theta|y) d\theta = \int p(y^*|\theta) \cdot p(\theta|y) d\theta$$

where the second identity follows only if the predictive distribution of y^* depends only on θ (which usually holds).

Posterior predictive $p-$value, Rubin (1984), is an important tool for assessing the adequacy of the model fit for Bayesian models. It is derived based on posterior predictive distribution.

Posterior Predictive Checking

Idea:

(Assuming conditional independence) we expect the predictive distribution of new data to look ‘similar’ to the empirical distribution of the observed data.

Extension:

(Assuming conditional independence) we expect the predictive distribution of functions (e.g., test statistics) of the new data to look ‘similar’ to the empirical distribution of functions of the observed data.

Posterior Predictive $p-$values

Consider testing the hypothesis:

$$\mathcal{H}_0: \text{ The model is correctly specified }, vs.,$$
$$\mathcal{H}_1: \text{ The model is not correctly specified }.$$

Select a test statistic $T(x)$ to perform the test; then define the posterior predictive $p-$value:

$$p_b = \mathbb{P}\left(T(y^*) \geq T(y) | y, \mathcal{H}_0\right).$$

In practice we use:

$$p_b^* = 2 \cdot \min\left\{p_b, 1 - p_b\right\}.$$

Posterior Predictive $p-$values

Notes:

- Very easy to compute since $p(y^*|y)$ is easy to sample from:
  1. Sample θ from the posterior distribution $p(\theta|y)$
  2. Given θ, sample y^* from $p(y^*|\theta)$

  Use MC to compute the posterior predictive $p-$values.

- Minimal extra work once posterior samples have been obtained

- Choice of test statistic: mean? median? max? min? other?

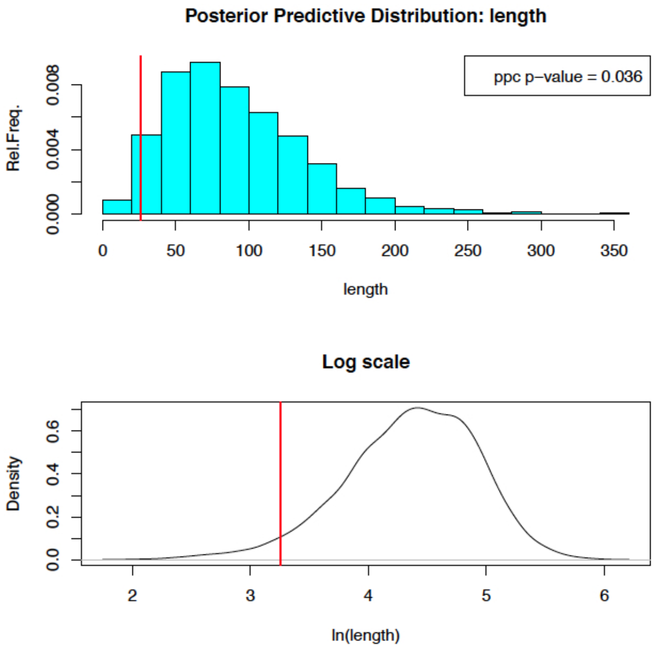
- Test statistic must be a function of data: photon counts

Posterior Predictive Check: Example

For example, in the logN-logS example our predictive data are the photon counts for 'replicate datasets'.

Take the number of observed sources, n, in the replicate dataset as our statistic.

This will test part of the assumptions about the missing data mechanism. . .



Posterior Predictive Distribution: length
ppc p−value = 0.036
Rel.Freq.
0.000
0.004
0.008
0
50
100
150
200
250
300
350
length
Log scale
Density
0.0
0.2
0.4
0.6
2
3
4
5
6
ln(length)

Extended Posterior Predictive $p-$values

Multivariate test statistics: further extend the idea behind posterior predictive $p-$values.
First, rewrite:

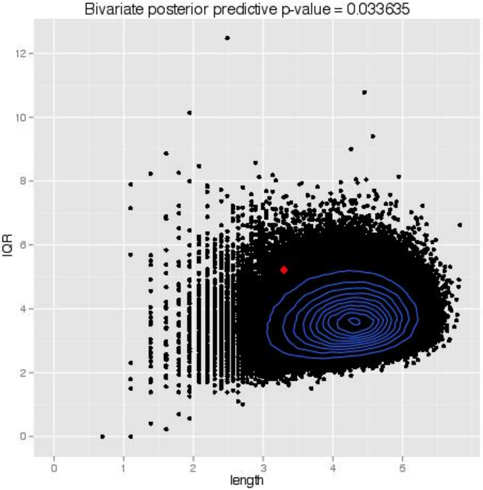
$$p_b = \mathbb{P}\left(T(y^*) \geq T(y) | y, \mathcal{H}_0\right)$$

as the tail probability:

$$p_b = \int I_{\{T(y^*) \geq T(y)\}} p\left(y^* | y, \mathcal{H}_0\right) dy^*$$

The multivariate posterior predictive $p-$values can yield extra insight when the dimensions of the test statistic are highly correlated.

Multivariate PP p–values



To Example

Validating Bayesian Computation

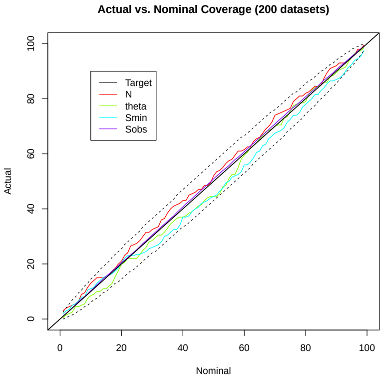
Given the complexity of the hierarchical model and computation, it is important to validate that everything works correctly.

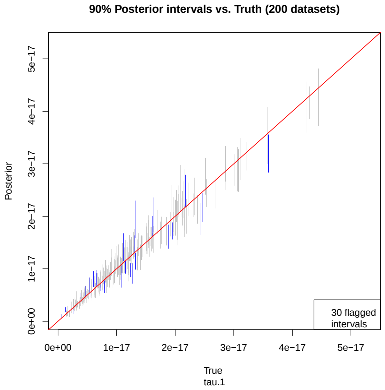
Bayesian methods admit a self-consistency check:

1. Simulate parameters from the prior, and data from the model, given those parameters
2. Fit the model to obtain posterior intervals
3. Record whether or not the 'true' value of the parameter was within the interval
4. Repeat steps 1 & 2 a large number of times, and calculate the average coverage

$\Rightarrow$ The average and nominal coverages should be equal.

These validation checks are extremely important when dealing with complex procedures.


Actual vs. Nominal Coverage (200 datasets)
Target
N
theta
Smin
Sobs
Actual
Nominal
0
20
40
60
80
100


90% Posterior intervals vs. Truth (200 datasets)
30 flagged intervals
Posterior
True
tau.1
0e+00
1e−17
2e−17
3e−17
4e−17
5e−17

Conclusions

1. Probabilistic insight allows us to build statistical procedures that correspond to more physically realistic models
2. Hierarchical modeling allows for us to account for multiple types of uncertainties
3. Allows crucial handling the non-ignorable missing data mechanism
4. Provides a recipe for assessing goodness-of-fit (posterior predictive checks)
5. Provides a way to include prior information
6. Allows for bias-corrected inference (Incompleteness)
7. Flexible framework for computation (e.g., distributional assumptions for fluxes)

Future Work

1. False sources (allowing that 'observed' sources might actually be background/artificial)
2. Field contamination (allowing a mixture of a source population with known parameters)
3. Estimation of the number of breakpoints in multiple power-law setting
4. Extension to non-Poisson regimes

References

- P.D. Baines, I.S. Udaltsova, A. Zezas, V.L. Kashyap (2011) Bayesian Estimation of $\log N - \log S$, *Proc. of Statistical Challenges in Modern Astronomy V*
- P.D. Baines, I.S. Udaltsova, A. Zezas, V.L. Kashyap (2011) Hierarchical Modeling for $\log N - \log S$: Theory, Methods and Applications *Working Paper*
- J. Jiang. et. al (2008) Fence methods for mied model selection. Ann Stat, 36:1669-1692.
- J. Jiang (2014) The Fence Methods. Adv Stat, 2014:830821.
- R.J.A. Little, D.B. Rubin. (2002) *Statistical analysis with missing data*, Wiley.
- R.K.W. Wong, P. Baines, A. Aue, T.C.M. Lee, V.L. Kashyap (2014) Automatic estimation of flux distributions of astrophysical source populations. Ann Appl Stat, 8:1690?1712.
- A. Zezas et al. (2004) Chandra survey of the 'Bar' region of the SMC *Revista Mexicana de Astronomia y Astrofisica (Serie de Conferencias)* Vol. 20. IAU Colloquium 194, pp. 205-205.