Upper Limits for source detection in the Three-Poisson Model

Xiao-Li Meng

Representing:
Paul Baines, Paul Edlefsen, Alan Lenarcic, Yaming Yu and the CHASC team

February 6, 2007

The Three-Poisson Problem

The basic setting:

$$\begin{aligned} n &\sim \mathrm{Pois}(\epsilon s + b) \\ y &\sim \mathrm{Pois}(tb) \\ z &\sim \mathrm{Pois}(u\epsilon) \end{aligned}$$

Observation: The triplet (n, y, z)

Constants: (t, u) **known** constants

Interest parameter: s

Nuisance parameters: b, ϵ.

Goal: Find $\hat{s}_p$ such that: $\mathbb{P}(s \leq \hat{s}_p) = p$ (e.g. $p = 0.90$, $p = 0.99$)

Why do we care about the problem?

This problem comes from High Energy Physics, specifically, the data expected to come from the Large Hadron Collider (LHC) at CERN, Switzerland.

Why do we care about the problem?

This problem comes from High Energy Physics, specifically, the data expected to come from the Large Hadron Collider (LHC) at CERN, Switzerland.

- **Motivation:** Detection (or otherwise) of Higgs-Boson particles, and (possibly) their masses.

Why do we care about the problem?

This problem comes from High Energy Physics, specifically, the data expected to come from the Large Hadron Collider (LHC) at CERN, Switzerland.

- **Motivation:** Detection (or otherwise) of Higgs-Boson particles, and (possibly) their masses.
- This could either support (or violate) the **Standard Model** of particle physics

The Three-Poisson Problem

In fact, the particle may decay into one of many (say, m) 'channels':

$$\begin{aligned} n_i &\sim \mathrm{Pois}(\epsilon_i s + b_i) \quad & i = 1, \ldots, m \\ y_i &\sim \mathrm{Pois}(t_i b_i) \quad & i = 1, \ldots, m \\ z_i &\sim \mathrm{Pois}(u_i \epsilon_i) \quad & i = 1, \ldots, m \end{aligned}$$

- ϵ_i is the decay rate for channel i
- b_i is the background rate for channel i
- (Y_i, Z_i) are collected from separate experiments designed to estimate b_i and ϵ_i
- The goal remains to find confidence limits for (the common) s

Challenges

The problem look really easy, right? Well...

Challenges

The problem look really easy, right? Well...

(1) Extremely low signal/noise ratio

Challenges

The problem look really easy, right? Well...

(1) Extremely low signal/noise ratio

(2) Dimensionality of nuisance parameter grows with m

Challenges

The problem look really easy, right? Well...

(1) Extremely low signal/noise ratio

(2) Dimensionality of nuisance parameter grows with m

(3) Specifying non-informative priors for high-dimensional nuisance parameters is tricky
(Note the $s\epsilon_i$ term: sensitive to prior on ϵ_i's)

Challenges

The problem look really easy, right? Well...

(1) Extremely low signal/noise ratio

(2) Dimensionality of nuisance parameter grows with m

(3) Specifying non-informative priors for high-dimensional nuisance parameters is tricky
(Note the $s\epsilon_i$ term: sensitive to prior on ϵ_i's)

(4) Turns out that the actual coverage can be very different from nominal coverage

Prior Specification: Single Level

Conjugate priors do not exist, instead have 'term-wise conjugate' priors:

$$
\begin{aligned}
s &\sim \mathrm{Gamma}(\alpha_s, \beta_s) && (1)\\
b_i &\sim^{iid} \mathrm{Gamma}(\alpha_b, \beta_b) \quad i = 1, \ldots, m && (2)\\
\epsilon_i &\sim^{iid} \mathrm{Gamma}(\alpha_\epsilon, \beta_\epsilon) \quad i = 1, \ldots, m && (3)
\end{aligned}
$$

Where $X \sim \Gamma(\alpha, \beta)$ has density:

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-x\beta\} \qquad \forall x \geq 0$$

We allow this specification to include improper priors:
(e.g. $(\alpha, \beta) = (1, 0)$ corresponds to $f(x) \propto 1 \ \forall\ x \geq 0$).

Simulation Results

Clearly such a strategy is unlikely to succeed (else there wouldn't be much to talk about!) and this is indeed the case. Here is a 'typical' result.

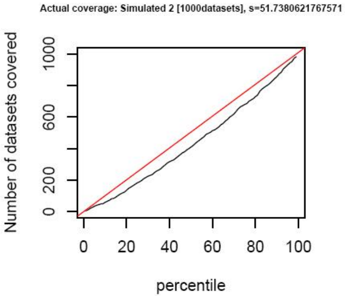

FIGURE: An example of undercoverage: $s = 51.7$, $m = 10$, $p(s) \propto 1 \quad p(b_i) \propto b_i^{-1/2} \quad p(\epsilon_i) \propto \epsilon_i^{-1/2}$.

Key Points

Nominal coverage varies drastically as s varies (other parameters remain of the same magnitude):

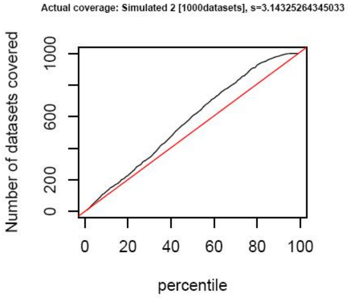

Figure: An example of overcoverage: $s = 51.7$, $m = 10$,
$p(s) \propto 1 \quad p(b_i) \propto b_i^{-1/2} \quad p(\epsilon_i) \propto \epsilon_i^{-1/2}$.

Single-Level Summary

Extensive simulation results yielded the following conclusions:

Single-Level Summary

Extensive simulation results yielded the following conclusions:

(1) Amongst this class of priors the following performed best: $(\alpha_s, \beta_s, \alpha_b, \beta_b, \alpha_\epsilon, \beta_\epsilon) = (1, 0, 0.5, 0, 0.5, 0)$ i.e.:

$$p(s) \propto 1 \qquad p(b_i) \propto b_i^{-1/2} \qquad p(\epsilon_i) \propto \epsilon_i^{-1/2}$$

Single-Level Summary

Extensive simulation results yielded the following conclusions:

(1) Amongst this class of priors the following performed best: $(\alpha_s, \beta_s, \alpha_b, \beta_b, \alpha_\epsilon, \beta_\epsilon) = (1, 0, 0.5, 0, 0.5, 0)$ i.e.:

$$p(s) \propto 1 \qquad p(b_i) \propto b_i^{-1/2} \qquad p(\epsilon_i) \propto \epsilon_i^{-1/2}$$

(2) Actual coverage is almost exact for single-channel $m = 1$

Single-Level Summary

Extensive simulation results yielded the following conclusions:

(1) Amongst this class of priors the following performed best: $(\alpha_s, \beta_s, \alpha_b, \beta_b, \alpha_\epsilon, \beta_\epsilon) = (1, 0, 0.5, 0, 0.5, 0)$ i.e.:

$$p(s) \propto 1 \qquad p(b_i) \propto b_i^{-1/2} \qquad p(\epsilon_i) \propto \epsilon_i^{-1/2}$$

(2) Actual coverage is almost exact for single-channel $m = 1$

(3) Actual coverage differs increasingly from nominal coverage as the number of channels m grows

Single-Level Summary

Extensive simulation results yielded the following conclusions:

(1) Amongst this class of priors the following performed best: $(\alpha_s, \beta_s, \alpha_b, \beta_b, \alpha_\epsilon, \beta_\epsilon) = (1, 0, 0.5, 0, 0.5, 0)$ i.e.:

$$p(s) \propto 1 \qquad p(b_i) \propto b_i^{-1/2} \qquad p(\epsilon_i) \propto \epsilon_i^{-1/2}$$

(2) Actual coverage is almost exact for single-channel $m = 1$

(3) Actual coverage differs increasingly from nominal coverage as the number of channels m grows

(4) More complicated Empirical Bayes schemes performed poorly and essentially 'overfit' the data

Single-Level Summary

Extensive simulation results yielded the following conclusions:

(1) Amongst this class of priors the following performed best: $(\alpha_s, \beta_s, \alpha_b, \beta_b, \alpha_\epsilon, \beta_\epsilon) = (1, 0, 0.5, 0, 0.5, 0)$ i.e.:

$$p(s) \propto 1 \qquad p(b_i) \propto b_i^{-1/2} \qquad p(\epsilon_i) \propto \epsilon_i^{-1/2}$$

(2) Actual coverage is almost exact for single-channel $m = 1$

(3) Actual coverage differs increasingly from nominal coverage as the number of channels m grows

(4) More complicated Empirical Bayes schemes performed poorly and essentially 'overfit' the data

(5) Large **overcoverage** exhibited when s small ($s < 1$) [Note: $\epsilon \in (0.04, 0.3)$ approx.]

Single-Level Summary

Extensive simulation results yielded the following conclusions:

(1) Amongst this class of priors the following performed best: $(\alpha_s, \beta_s, \alpha_b, \beta_b, \alpha_\epsilon, \beta_\epsilon) = (1, 0, 0.5, 0, 0.5, 0)$ i.e.:

$$p(s) \propto 1 \qquad p(b_i) \propto b_i^{-1/2} \qquad p(\epsilon_i) \propto \epsilon_i^{-1/2}$$

(2) Actual coverage is almost exact for single-channel $m = 1$

(3) Actual coverage differs increasingly from nominal coverage as the number of channels m grows

(4) More complicated Empirical Bayes schemes performed poorly and essentially ‘overfit’ the data

(5) Large **overcoverage** exhibited when s small ($s < 1$) [Note: $\epsilon \in (0.04, 0.3)$ approx.]

(6) Large **undercoverage** exhibited when s large ($s > 60$)

How did it actually do?

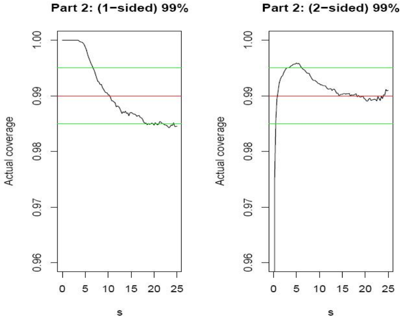

FIGURE: Actual coverage of the 99^{th} percentile (l) and the equal-tailed 99% posterior interval (r) for the single level model, with $m = 10$

How did it actually do?

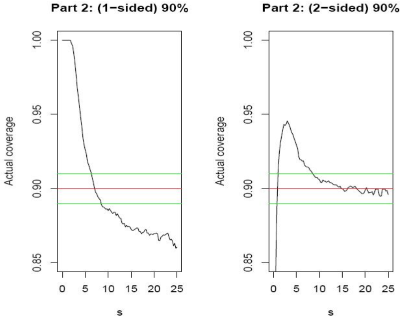

Figure: Actual coverage of the 90^{th} percentile (l) and the equal-tailed 90% posterior interval (r) for the single level model, with $m = 10$

Hierarchical Models

The multi-dimensional ‘vague’ prior on the nuisance parameter is the primary problem. Next step: **Hierarchical Model**:

$$
\begin{aligned}
n_i &\sim & \mathrm{Pois}(\epsilon_i s + b_i) & \quad i = 1, \ldots, m \\
y_i &\sim & \mathrm{Pois}(t_i b_i) & \quad i = 1, \ldots, m \\
z_i &\sim & \mathrm{Pois}(u_i \epsilon_i) & \quad i = 1, \ldots, m \\
s &\sim & \mathrm{Gamma}(\alpha_s, \beta_s) & \\
b_i &\sim^{iid} & \mathrm{Gamma}(\alpha_b, \beta_b) & \quad i = 1, \ldots, m \\
\epsilon_i &\sim^{iid} & \mathrm{Gamma}(\alpha_\epsilon, \beta_\epsilon) & \quad i = 1, \ldots, m \\
p(\alpha_s) \propto 1 & & p(\alpha_b) \sim 1 & \quad p(\alpha_\epsilon) \propto 1 \\
p(\beta_s) &\sim & \mathrm{Gamma}(\alpha_{\beta_s}, \beta_{\beta_s}) & \\
p(\beta_b) &\sim & \mathrm{Gamma}(\alpha_{\beta_b}, \beta_{\beta_b}) & \\
p(\beta_\epsilon) &\sim & \mathrm{Gamma}(\alpha_{\beta_\epsilon}, \beta_{\beta_\epsilon}) &
\end{aligned}
$$

Discussion

Some points to mention:

Discussion

Some points to mention:

(1) Flat priors can be replaced with vague proper priors if desired

Discussion

Some points to mention:

(1) Flat priors can be replaced with vague proper priors if desired

(2) No longer possible to integrate out nuisance parameters, sampling-based MCMC approach was used

Discussion

Some points to mention:

(1) Flat priors can be replaced with vague proper priors if desired

(2) No longer possible to integrate out nuisance parameters, sampling-based MCMC approach was used

(3) MCMC implementation is problematic for large-scale simulations

Discussion

Some points to mention:

(1) Flat priors can be replaced with vague proper priors if desired

(2) No longer possible to integrate out nuisance parameters, sampling-based MCMC approach was used

(3) MCMC implementation is problematic for large-scale simulations

(4) Hierarchical model retains physical interpretation of the parameters

Performance

The hierarchical model produces consistently larger $100(1-\alpha)\%$ upper limits, although actual coverage remains sensitive to s.

How did it actually do?

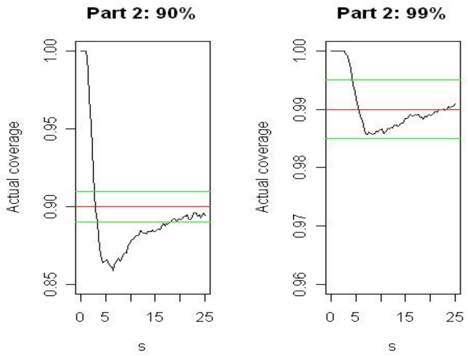

FIGURE: Actual coverage as a function of s for the hierarchical model.

Length Comparisons

It is also very important that the intervals be as short as possible whilst retaining excellent coverage properties. For simplicity we shall compare lengths of the 99% intervals, although the same conclusions hold for 90% intervals too.

Single-level vs. Hierarchical Bayes

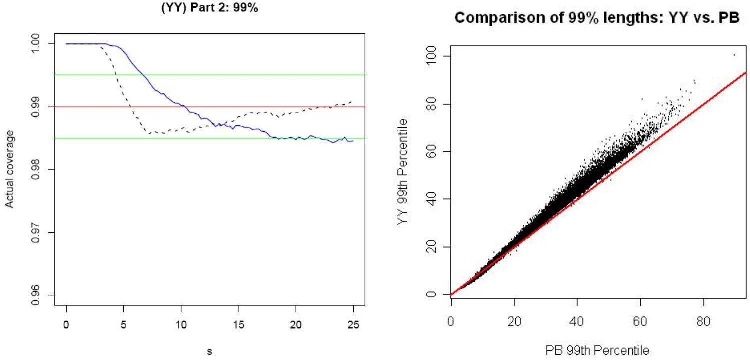

Figure: (L) Coverage (blue=single-level, dash=YY). (R) Comparison of lengths of the 99^{th} percentiles from the single-level and hierarchical Bayes models. Datasets are ordered according to the single-level lengths, $m = 10$.

Single-level vs. Dempster-Schafer

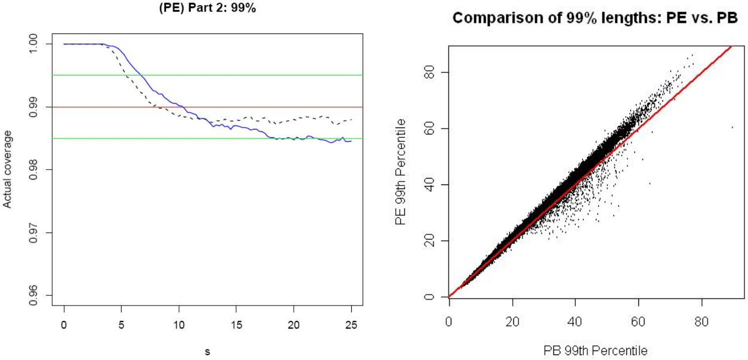

Figure: (L) Coverage (blue=single-level, dash=PE). (R) Comparison of lengths of the 99^{th} percentiles from the single-level and Dempster-Schafer models. Datasets are ordered according to the single-level lengths, $m = 10$.

Single-level vs. Profile Likelihood

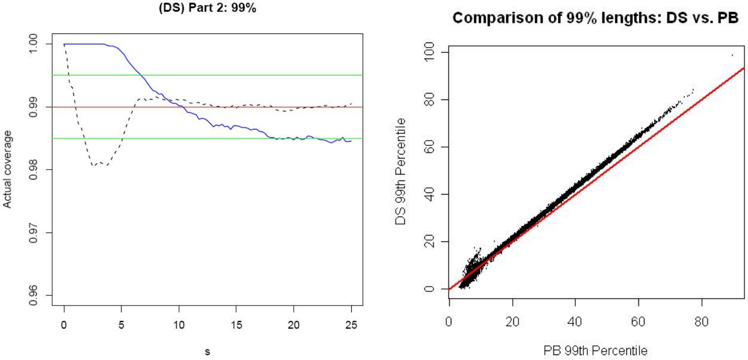

Figure: (L) Coverage (blue=single-level, dash=DS). (R) Comparison of lengths of the 99^{th} percentiles from the single-level model and profile likelihood approach. Datasets are ordered according to the single-level lengths, $m = 10$.

Single-level vs. Wolfgang Rolke

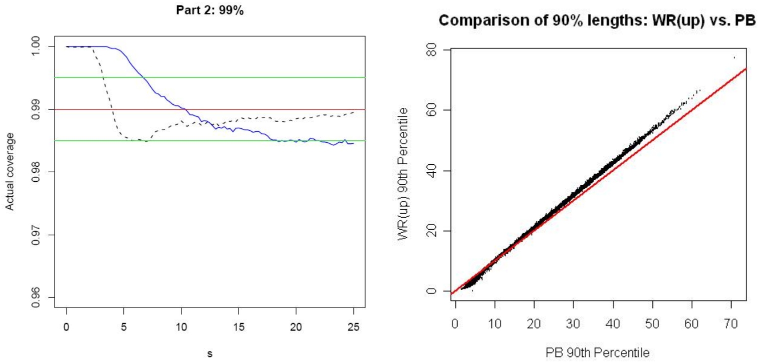

Figure: (L) Coverage (blue=single-level, dash=WR). (R) Comparison of lengths of the 99^{th} percentiles from the single-level model and one of Wolfgang Rolke's four entries (unknown method). Datasets are ordered according to the single-level lengths, $m = 10$

Single-level vs. Luc Demortier

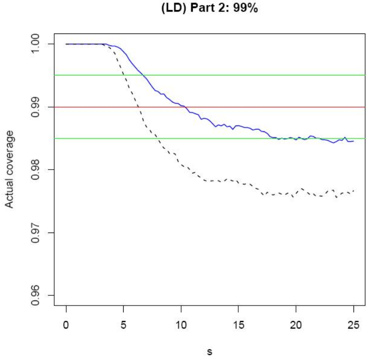

Figure: (Key: blue=single-level, dash=LD) Comparison of the coverage of the 99^{th} percentiles from the single-level model and Luc Demortier's entry, $m = 10$. Unable to compare lengths due to the file format.

Challenges

Problem remains a work in progress...

Challenges

Problem remains a work in progress. . .

(1) More work is needed to fully understand the properties of the hierarchical three-Poisson model; robustness etc.

Challenges

Problem remains a work in progress. . .

(1) More work is needed to fully understand the properties of the hierarchical three-Poisson model; robustness etc.

(2) Significant improvements in implementation are required in the MCMC scheme to permit large-scale application

Challenges

Problem remains a work in progress. . .

(1) More work is needed to fully understand the properties of the hierarchical three-Poisson model; robustness etc.

(2) Significant improvements in implementation are required in the MCMC scheme to permit large-scale application

(3) 'Matching priors' are theoretically available; implementation? interpretation?

Conclusions

In summary:

Conclusions

In summary:

- The problem is deceptively challenging and further work is needed to investigate issues such as hyperparameter specification and robustness

Conclusions

In summary:

- The problem is deceptively challenging and further work is needed to investigate issues such as hyperparameter specification and robustness
- The Bayesian approach should outperform others (e.g. profile likelihood), but only when we find the right prior. . .

Conclusions

In summary:

- The problem is deceptively challenging and further work is needed to investigate issues such as hyperparameter specification and robustness
- The Bayesian approach should outperform others (e.g. profile likelihood), but only when we find the right prior. . .
- Computational challenges remain for large-scale applications

Conclusions

In summary:

- The problem is deceptively challenging and further work is needed to investigate issues such as hyperparameter specification and robustness
- The Bayesian approach should outperform others (e.g. profile likelihood), but only when we find the right prior...
- Computational challenges remain for large-scale applications
- Many theoretical questions still need to be addressed (e.g. one-vs.two-sided, nuisance parameters, validity of Poisson assumptions)