Certain Identities in FGMC

Jolien Creighton

Created: January 19, 2017; Revised: October 9, 2019

1 Review

This is a brief summary of the counting method developed by Farr, Gair, Mandel, and Cutler, hereafter FGMC [1].

Given a set of event data $\{x_1, x_2, \ldots, x_M\}$, and probability distributions $p_0(x)$ and $p_1(x)$ for background (noise) and foreground (signal) models, the FGMC likelihood for the count of background events, Λ_0, and foreground events, Λ_1, is

$$L(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\}) \propto \exp(-\Lambda_0 - \Lambda_1) \prod_{i=1}^{M} [\Lambda_0 p_0(x_i) + \Lambda_1 p_1(x_i)]. \tag{1}$$

With a (possibly improper) prior distribution, $p(\Lambda_0, \Lambda_1)$, the posterior distribution is

$$p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\}) \propto p(\Lambda_0, \Lambda_1) \exp(-\Lambda_0 - \Lambda_1) \prod_{i=1}^{M} [\Lambda_0 p_0(x_i) + \Lambda_1 p_1(x_i)]. \tag{2}$$

For an event x, the probability that it is a foreground event given Λ_0 and Λ_1 is

$$P_1(x \mid \Lambda_0, \Lambda_1) = \frac{\Lambda_1 p_1(x)}{\Lambda_0 p_0(x) + \Lambda_1 p_1(x)}. \tag{3}$$

For any particular event in the set of observed events, $x_i \in \{x_1, x_2, \ldots, x_M\}$, the probability that that event is a foreground event can be obtained by marginalizing over Λ_0 and Λ_1 using Eq. (2):

$$P_1(x_i \mid \{x_1, x_2, \ldots, x_M\}) = \iint_0^\infty d\Lambda_0 \, d\Lambda_1 \, p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\}) \frac{\Lambda_1 p_1(x_i)}{\Lambda_0 p_0(x_i) + \Lambda_1 p_1(x_i)}. \tag{4}$$

If sources are uniformly distributed in comoving spacetime intervals and a search is sensitive to sources within a spacetime volume (VT) then the rate (number per unit co-moving volume per unit source-frame time) is

$$R = \frac{\Lambda_1}{(VT)}. \tag{5}$$

A change of variables in Eq. (2) gives the posterior distribution for R:

$$p(R \mid (VT), \{x_1, x_2, \ldots, x_M\}) = (VT) \int_0^\infty d\Lambda_0 \, p(\Lambda_0, \Lambda_1 = R(VT) \mid \{x_1, x_2, \ldots, x_M\}). \tag{6}$$

Proof of Eq. (1). The proof follows Loredo & Wasserman (1995) [2] and Abbott et al. (2016) [3]. Suppose that we divide the total observation time T into a large number N of subintervals of duration $\delta t = T/N$. If these subintervals are small enough then each one will contain either 0 events or 1 event. There are M intervals that have 1 event, $\{x_i : i = 1, 2, \ldots, M\}$, and $N - M$ intervals with 0 events, $\{\varnothing_j : j = M+1, M+2, \ldots, N\}$ where

$\varnothing_j$ indicates that no event was detected in interval j. The joint probability of these independent propositions is

$$L(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M, \varnothing_{M+1}, \varnothing_{M+2}, \ldots, \varnothing_N\}) = \binom{N}{M} \left(\prod_{i=1}^{M} p(x_i \mid \Lambda_0, \Lambda_1) \right) \left(\prod_{j=M+1}^{N} p(\varnothing_j \mid \Lambda_0, \Lambda_1) \right). \tag{7}$$

Here, I've put in the combinatoric factor $\binom{N}{M}$ representing the number of ways to distribute the M events in the N intervals. I'm not sure about this factor, but it doesn't matter for our purposes.

The first set of factors involves the probability of observing the event x_i in the ith interval, $p(x_i \mid \Lambda_0, \Lambda_1)$. The rate at which the experiment produces events with values between x and $x + dx$ is

$$\frac{d\mathcal{N}}{dt\,dx} = \dot{\Lambda}_0 p_0(x) + \dot{\Lambda}_1 p_1(x) \tag{8}$$

where $\dot{\Lambda}_0 = \Lambda_0/T$ is the rate of background events and $\dot{\Lambda}_1 = \Lambda_1/T$ is the rate of foreground events. The overall rate of events is

$$\frac{d\mathcal{N}}{dt} = \int dx \, \frac{d\mathcal{N}}{dt\,dx} = \dot{\Lambda}_0 + \dot{\Lambda}_1 \tag{9}$$

Therefore, $p(x_i \mid \Lambda_0, \Lambda_1)$ is the differential probability of a single event in the ith interval (we are free to choose δt to be arbitrarily small, and by so doing make the probability of getting *two* events in the same interval vanishingly small). We can write this as

$$p(x_i \mid \Lambda_0, \Lambda_1) = p(x_i \mid 1, \Lambda_0, \Lambda_1) p(1 \mid \Lambda_0, \Lambda_1) \tag{10}$$

where

$$p(1 \mid \Lambda_0, \Lambda_1) = \left(\frac{d\mathcal{N}}{dt} \delta t \right) \exp\left(-\frac{d\mathcal{N}}{dt} \delta t \right) = [(\dot{\Lambda}_0 + \dot{\Lambda}_1)\delta t] \exp\{(\dot{\Lambda}_0 + \dot{\Lambda}_1)\delta t\} \tag{11}$$

is the probability of getting a single event in the interval δt and

$$p(x_i \mid 1, \Lambda_0, \Lambda_1) = \left. \frac{d\mathcal{N}}{dt\,dx}(x_i) \middle/ \frac{d\mathcal{N}}{dt}(x_i) \right. = \frac{\dot{\Lambda}_0 p_0(x_i) + \dot{\Lambda}_1 p_1(x_i)}{\dot{\Lambda}_0 + \dot{\Lambda}_1} \tag{12}$$

is the probability density function for one event over x evaluated at x_i. Therefore,

$$p(x_i \mid \Lambda_0, \Lambda_1) = \left(\frac{d\mathcal{N}}{dt\,dx} \delta t \right) \exp\left(-\frac{d\mathcal{N}}{dt} \delta t \right) = \delta t [\dot{\Lambda}_0 p_0(x_i) + \dot{\Lambda}_1 p_1(x_i)] \exp\{-(\dot{\Lambda}_0 + \dot{\Lambda}_1)\delta t\}. \tag{13}$$

The second set of factors involves the probability of observing no events in the jth interval, $p(\varnothing_j \mid \Lambda_0, \Lambda_1)$:

$$p(\varnothing_j \mid \Lambda_0, \Lambda_1) = \exp\left(-\frac{d\mathcal{N}}{dt} \delta t \right) = \exp\{-(\dot{\Lambda}_0 + \dot{\Lambda}_1)\delta t\}. \tag{14}$$

Since $\delta t = T/N$, $\dot{\Lambda}_0 \delta t = \Lambda_0/N$ and $\dot{\Lambda}_1 \delta t = \Lambda_1/N$, so we have

$$\begin{aligned} &L(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M, \varnothing_{M+1}, \varnothing_{M+2}, \ldots, \varnothing_N\}) \\ &\quad = \binom{N}{M} \left(\prod_{i=1}^{M} \delta t [\dot{\Lambda}_0 p_0(x_i) + \dot{\Lambda}_1 p_1(x_i)] \exp\{-(\dot{\Lambda}_0 + \dot{\Lambda}_1)\delta t\} \right) \left(\prod_{j=M+1}^{N} \exp\{-(\dot{\Lambda}_0 + \dot{\Lambda}_1)\delta t\} \right) \\ &\quad = \binom{N}{M} N^{-M} \exp(-\Lambda_0 - \Lambda_1) \prod_{i=1}^{M} [\Lambda_0 p_0(x_i) + \Lambda_1 p_1(x_i)]. \end{aligned} \tag{15}$$

As $N \to \infty$, we have $N \gg M$ and

$$\binom{N}{M} = \frac{N(N-1)\cdots(N-M+1)}{M!} \simeq \frac{N^M}{M!} \tag{16}$$

so

$$L(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M, \varnothing_{M+1}, \varnothing_{M+2}, \ldots, \varnothing_N\}) = \frac{1}{M!} \exp(-\Lambda_0 - \Lambda_1) \prod_{i=1}^{M} [\Lambda_0 p_0(x_i) + \Lambda_1 p_1(x_i)]. \tag{17}$$

(Again, I am not sure about the combinatoric factor.)

□

2 Useful Identities

Lemma 1. *Suppose that the prior distribution takes the form*

$$p(\Lambda_0, \Lambda_1) \propto p(\Lambda_0)\Lambda_1^\alpha \tag{18}$$

for some power α. Then the sum of the foreground probabilities of the events is related to the expectation value of Λ_1 by

$$\sum_{i=1}^{M} P_1(x_i \mid \{x_1, x_2, \ldots, x_M\}) = \langle \Lambda_1 \rangle - \alpha - 1. \tag{19}$$

Proof. Consider

$$\begin{aligned} \frac{d}{d\Lambda_1}[\Lambda_1 p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\})] &= (\alpha+1) p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\}) - \Lambda_1 p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\}) \\ &\quad + p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\}) \sum_{i=1}^{M} \frac{\Lambda_1 p_1(x_i)}{\Lambda_0 p_0(x_i) + \Lambda_1 p_1(x_i)}. \end{aligned} \tag{20}$$

Integrate over Λ_0 and Λ_1 to obtain

$$0 = \alpha + 1 - \langle \Lambda_1 \rangle + \sum_{i=1}^{M} P_1(x_i \mid \{x_1, x_2, \ldots, x_M\}). \tag{21}$$

□

Lemma 2. *These same things supposed, then*

$$\sum_{i=1}^{M} \iint_0^\infty d\Lambda_0 \, d\Lambda_1 \, p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\}) \frac{\Lambda_1^2 p_1(x_i)}{\Lambda_0 p_0(x_i) + \Lambda_1 p_1(x_i)} = \langle \Lambda_1^2 \rangle - (\alpha+2)\langle \Lambda_1 \rangle. \tag{22}$$

Proof. Consider

$$\begin{aligned} \frac{d}{d\Lambda_1}[\Lambda_1^2 p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\})] &= (\alpha+2)\Lambda_1 p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\}) - \Lambda_1^2 p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\}) \\ &\quad + p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\}) \sum_{i=1}^{M} \frac{\Lambda_1^2 p_1(x_i)}{\Lambda_0 p_0(x_i) + \Lambda_1 p_1(x_i)}. \end{aligned} \tag{23}$$

Integrate over Λ_0 and Λ_1 to obtain

$$0 = (\alpha+2)\langle \Lambda_1 \rangle - \langle \Lambda_1^2 \rangle + \sum_{i=1}^{M} \iint_0^\infty d\Lambda_0 \, d\Lambda_1 \, p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\}) \frac{\Lambda_1^2 p_1(x_i)}{\Lambda_0 p_0(x_i) + \Lambda_1 p_1(x_i)}. \tag{24}$$

□

Lemma 3. *If $p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\})$ is the posterior distribution for the M events $\{x_1, x_2, \ldots, x_M\}$, then if a new event x_{M+1} is recorded, the posterior is updated to be*

$$p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M, x_{M+1}\}) = p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\}) \frac{\Lambda_0 p_0(x_{M+1}) + \Lambda_1 p_1(x_{M+1})}{\langle\Lambda_0\rangle_M p_0(x_{M+1}) + \langle\Lambda_1\rangle_M p_1(x_{M+1})} \tag{25}$$

where

$$\langle\Lambda_0\rangle_M = \iint_0^\infty d\Lambda_0\, d\Lambda_1\, p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\})\, \Lambda_0 \tag{26}$$

and

$$\langle\Lambda_1\rangle_M = \iint_0^\infty d\Lambda_0\, d\Lambda_1\, p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\})\, \Lambda_1 \tag{27}$$

are the mean values of Λ_0 and Λ_1 prior to the inclusion of event x_{M+1}.

Proof. From the form of the FGMC posterior, we see that

$$p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M, x_{M+1}\}) = A p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\})[\Lambda_0 p_0(x_{M+1}) + \Lambda_1 p_1(x_{M+1})] \tag{28}$$

where A is a normalization constant. Integrating both sides over Λ_0 and Λ_1 yields

$$1 = A[\langle\Lambda_0\rangle_M p_0(x_{M+1}) + \langle\Lambda_1\rangle_M p_1(x_{M+1})]. \tag{29}$$

Solve this for A and insert into the original equation to get the desired result. □

Theorem 1. *When a new event x_{M+1} is added to an existing set of events $\{x_1, x_2, \ldots, x_M\}$, the mean value of Λ_1 is increased as*

$$\langle\Lambda_1\rangle_{M+1} = \langle\Lambda_1\rangle_M + \frac{\mathrm{cov}_M(\Lambda_0, \Lambda_1) p_0(x_{M+1}) + \mathrm{Var}_M(\Lambda_1) p_1(x_{M+1})}{\langle\Lambda_0\rangle_M p_0(x_{M+1}) + \langle\Lambda_1\rangle_M p_1(x_{M+1})} \tag{30}$$

where

$$\mathrm{Var}_M \Lambda_1 = \iint_0^\infty d\Lambda_0\, d\Lambda_1\, p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\})\, (\Lambda_1 - \langle\Lambda_1\rangle_M)^2 \tag{31}$$

and

$$\mathrm{cov}_M(\Lambda_0, \Lambda_1) = \iint_0^\infty d\Lambda_0\, d\Lambda_1\, p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \ldots, x_M\})\, (\Lambda_0 - \langle\Lambda_0\rangle_M)(\Lambda_1 - \langle\Lambda_1\rangle_M) \tag{32}$$

are the variance of Λ_1 and the covariance of Λ_0 and Λ_1 prior to the inclusion of event x_{M+1}.

Proof. Use the result of Lemma 3:

$$\begin{aligned}
&\langle\Lambda_1\rangle_{M+1} - \langle\Lambda_1\rangle_M \\
&\qquad = -\langle\Lambda_1\rangle_M + \iint_0^\infty d\Lambda_0\, d\Lambda_1\, p(\Lambda_0,\Lambda_1 \mid \{x_1,x_2,\ldots,x_M,x_{M+1}\})\,\Lambda_1 \\
&\qquad = -\langle\Lambda_1\rangle_M + \iint_0^\infty d\Lambda_0\, d\Lambda_1\, p(\Lambda_0,\Lambda_1 \mid \{x_1,x_2,\ldots,x_M\})\frac{\Lambda_0 p_0(x_{M+1}) + \Lambda_1 p_1(x_{M+1})}{\langle\Lambda_0\rangle_M p_0(x_{M+1}) + \langle\Lambda_1\rangle_M p_1(x_{M+1})}\,\Lambda_1 \\
&\qquad = -\langle\Lambda_1\rangle_M + \frac{\langle\Lambda_0\Lambda_1\rangle_M p_0(x_{M+1}) + \langle\Lambda_1^2\rangle_M p_1(x_{M+1})}{\langle\Lambda_0\rangle_M p_0(x_{M+1}) + \langle\Lambda_1\rangle_M p_1(x_{M+1})} \\
&\qquad = \frac{(\langle\Lambda_0\Lambda_1\rangle_M - \langle\Lambda_0\rangle_M\langle\Lambda_1\rangle_M) p_0(x_{M+1}) + (\langle\Lambda_1^2\rangle_M - \langle\Lambda_1\rangle_M^2) p_1(x_{M+1})}{\langle\Lambda_0\rangle_M p_0(x_{M+1}) + \langle\Lambda_1\rangle_M p_1(x_{M+1})} \\
&\qquad = \frac{\mathrm{cov}_M(\Lambda_0,\Lambda_1) p_0(x_{M+1}) + \mathrm{Var}_M(\Lambda_1) p_1(x_{M+1})}{\langle\Lambda_0\rangle_M p_0(x_{M+1}) + \langle\Lambda_1\rangle_M p_1(x_{M+1})}. \qquad (33)
\end{aligned}$$

Here we have

$$\langle\Lambda_1^2\rangle_M = \iint_0^\infty d\Lambda_0\, d\Lambda_1\, p(\Lambda_0,\Lambda_1 \mid \{x_1,x_2,\ldots,x_M\})\,\Lambda_1^2 \qquad (34)$$

and

$$\langle\Lambda_0\Lambda_1\rangle_M = \iint_0^\infty d\Lambda_0\, d\Lambda_1\, p(\Lambda_0,\Lambda_1 \mid \{x_1,x_2,\ldots,x_M\})\,\Lambda_0\Lambda_1 \qquad (35)$$

from which it is seen that

$$\mathrm{Var}_M\Lambda_1 = \langle\Lambda_1^2\rangle_M - \langle\Lambda_1\rangle_M^2 \qquad (36)$$

and

$$\mathrm{cov}_M(\Lambda_0,\Lambda_1) = \langle\Lambda_0\Lambda_1\rangle_M - \langle\Lambda_0\rangle_M\langle\Lambda_1\rangle_M. \qquad (37)$$

□

Remark. In the case when $\langle\Lambda_1\rangle_M p_1(x_{M+1}) \gg \langle\Lambda_0\rangle_M p_0(x_{M+1})$, i.e., the new event is almost certainly a foreground signal, then

$$\langle\Lambda_1\rangle_{M+1} \approx \langle\Lambda_1\rangle_M + \frac{\mathrm{Var}_M\Lambda_1}{\langle\Lambda_1\rangle_M}. \qquad (38)$$

If $\mathrm{Var}_M\Lambda_1 > \langle\Lambda_1\rangle_M$ then the new event contributes *more than one count* to Λ_1, even though $P_1(x_{M+1} \mid \{x_1,x_2,\ldots,x_M,x_{M+1}\}) \approx 1$ (see Theorem 2 below). This seems peculiar in light of Lemma 1, which seemingly implies that the new event should contribute just one count; however, the new event also has the effect of updating the foreground probabilities $P_1(x_i \mid \{x_1,x_2,\ldots,x_M,x_{M+1}\})$ for all other events $x_i \in \{x_1,x_2,\ldots,x_M\}$.

Theorem 2. *When a new event x_{M+1} is added to an existing set of events $\{x_1,x_2,\ldots,x_M\}$, its probability of being a foreground event is*

$$P_1(x_{M+1} \mid \{x_1,x_2,\ldots,x_M,x_{M+1}\}) = \frac{\langle\Lambda_1\rangle_M p_1(x_{M+1})}{\langle\Lambda_0\rangle_M p_0(x_{M+1}) + \langle\Lambda_1\rangle_M p_1(x_{M+1})}. \qquad (39)$$

Proof. Again use Lemma 3:

$$
\begin{aligned}
&P_1(x_{M+1} \mid \{x_1, x_2, \dots, x_M, x_{M+1}\}) \\
&\quad = \iint_0^\infty d\Lambda_0\, d\Lambda_1\, p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \dots, x_M, x_{M+1}\}) \frac{\Lambda_1 p_1(x_{M+1})}{\Lambda_0 p_0(x_{M+1}) + \Lambda_1 p_1(x_{M+1})} \\
&\quad = \iint_0^\infty d\Lambda_0\, d\Lambda_1\, p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \dots, x_M\}) \frac{\Lambda_0 p_0(x_{M+1}) + \Lambda_1 p_1(x_{M+1})}{\langle\Lambda_0\rangle_M p_0(x_{M+1}) + \langle\Lambda_1\rangle_M p_1(x_{M+1})} \frac{\Lambda_1 p_1(x_i)}{\Lambda_0 p_0(x_i) + \Lambda_1 p_1(x_i)} \\
&\quad = \iint_0^\infty d\Lambda_0\, d\Lambda_1\, p(\Lambda_0, \Lambda_1 \mid \{x_1, x_2, \dots, x_M\}) \frac{\Lambda_1 p_1(x_i)}{\langle\Lambda_0\rangle_M p_0(x_{M+1}) + \langle\Lambda_1\rangle_M p_1(x_{M+1})} \\
&\quad = \frac{\langle\Lambda_1\rangle_M p_1(x_{M+1})}{\langle\Lambda_0\rangle_M p_0(x_{M+1}) + \langle\Lambda_1\rangle_M p_1(x_{M+1})}.
\end{aligned}
\tag{40}
$$

□

3 Approximate 1-Dimensional Posterior

The FGMC likelihood induces correlations between Λ_0 and Λ_1 when there are events for which $p_0(x_i) \approx p_1(x_i)$. This is expected: as events accumulate, we expect both Λ_0 and Λ_1 to grow. For a stationary experiment, we would also expect that the ratio of the rate of growth of Λ_1 to the rate of growth of Λ_0 to be constant. Furthermore, FGMC is intended to be applied when the rate of background events is larger than the rate of foreground events so that most of the observed events are background. This motivates the change of variables $\Lambda_0 = Mu$ and expressing the FGMC posterior in the form

$$
\begin{aligned}
&p(u, \Lambda_1 \mid \{k_1, k_2, \dots, k_M\}) \\
&\quad \propto p(u) p(\Lambda_1) M^M u^M \exp(-Mu) \exp(-\Lambda_1) \prod_{i=1}^{M} (1 + \Lambda_1 k_i / u) \\
&\quad = p(\Lambda_1) \exp(-\Lambda_1) h(u) \exp[M g(u)]
\end{aligned}
\tag{41}
$$

where we have supposed the prior distribution factorizes, $p(\Lambda_0, \Lambda_1) = p(\Lambda_0) p(\Lambda_1)$, have introduced the functions

$$
g(u) = \log(Mu) - u = \log M - 1 - \frac{1}{2}(u-1)^2 + \frac{1}{3}(u-1)^3 - \frac{1}{4}(u-1)^4 + \cdots, \tag{42}
$$

$$
h(u) = p(u) \prod_{i=1}^{M} (1 + \Lambda_1 k_i / u), \tag{43}
$$

and have defined

$$
k_i = \frac{K_i}{M} \qquad \text{with} \qquad K_i = \frac{p_1(x_i)}{p_0(x_i)} \tag{44}
$$

to be the *reduced Bayes factor* for the ith event, i.e., the ratio of the likelihood of the signal model to the noise model, known as the *Bayes factor* K_i, divided by M. Next we assume that $M \gg 1$ and that $h(u)$ is a slowly varying function of u; then, since the factor $\exp[Mg(u)]$ is sharply peaked about $u = 1$, we can marginalize over Λ_0 using the method of Laplace to evaluate the integral:

$$
\begin{aligned}
&p(\Lambda_1 \mid \{k_1, k_2, \ldots, k_M\}) \\
&\quad = \int_0^\infty du\, p(u, \Lambda_1 \mid \{k_1, k_2, \ldots, k_M\}) \\
&\quad \propto p(\Lambda_1) \exp(-\Lambda_1) \int_0^\infty du\, h(u) \exp[M g(u)] \\
&\quad \approx p(\Lambda_1) \exp(-\Lambda_1) \sqrt{\frac{2\pi}{M|g''(1)|}}\, h(1) \exp[M g(1)] \\
&\quad \propto p(\Lambda_1) \exp(-\Lambda_1) \prod_{i=1}^{M} (1 + \Lambda_1 k_i).
\end{aligned} \tag{45}
$$

Several results from the previous section carry over. In particular, the probability that a particular event is a foreground event is

$$
P_1(k_i \mid \{k_1, k_2, \ldots, k_M\}) = \int_0^\infty d\Lambda_1\, p(\Lambda_1 \mid \{k_1, k_2, \ldots, k_M\}) \frac{\Lambda_1 k_i}{1 + \Lambda_1 k_i} \tag{46}
$$

for $i \in [1, M]$ or, if $i = M + 1$ is a new event, Theorem 2 gives

$$
P_1(k_{M+1} \mid \{k_1, k_2, \ldots, k_M\}) = \frac{\langle \Lambda_1 \rangle_M k_{M+1}}{1 + \langle \Lambda_1 \rangle_M k_{M+1}}, \tag{47}
$$

while the mean value of Λ_1 is

$$
\langle \Lambda_1 \rangle = 1 + \alpha + \sum_{i=1}^{M} P_1(k_i \mid \{k_1, k_2, \ldots, k_M\}). \tag{48}
$$

for $i \in [1, M]$ or, if $i = M + 1$ is a new event, the revised version of Theorem 1 describes how the mean foreground count increases with the inclusion of a new event:

$$
\langle \Lambda_1 \rangle_{M+1} = \langle \Lambda_1 \rangle_M + \frac{\mathrm{Var}_M(\Lambda_1) k_{M+1}}{1 + \langle \Lambda_1 \rangle_M k_{M+1}}. \tag{49}
$$

3.1 Application

Suppose that our zerolag analysis has yielded M events, $\{k_1, k_2, \ldots, k_M\}$, from which we derive the posterior distribution on Λ_1, $p(\Lambda_1 \mid \{k_1, k_2, \ldots, k_M\})$. An injection campaign, in which N_{inj} events were injected, has produced M' *recorded* events, $\{k'_1, k'_2, \ldots, k'_{M'}\}$. We wish to determine how many of the injections are found.

Suppose that we take a single injection event, k'_j, $1 \le j \le M'$, and add it to the zerolag events so that our events are now $\{k_1, k_2, \ldots, k_M, k_{M+1} = k'_j\}$. By Theorem 1 we have

$$
\Delta \langle \Lambda_1 \rangle = \frac{\mathrm{Var}_M(\Lambda_1) k'_j}{1 + \langle \Lambda_1 \rangle_M k'_j} \tag{50}
$$

where

$$
\Delta \langle \Lambda_1 \rangle = \langle \Lambda_1 \rangle_{M+1} - \langle \Lambda_1 \rangle_M \tag{51}
$$

is the additional number of counts that injection would have contributed to the zerolag search. For our injection set, the average number of counts that the injections would have contributed to our search is

$$
\overline{\Delta \langle \Lambda_1 \rangle} = \frac{1}{N_{\mathrm{inj}}} \sum_{j=1}^{M'} \frac{\mathrm{Var}_M(\Lambda_1) k'_j}{1 + \langle \Lambda_1 \rangle_M k'_j} = \frac{\mathrm{Var}_M(\Lambda_1)}{N_{\mathrm{inj}}} \sum_{j=1}^{M'} \frac{k'_j}{1 + \langle \Lambda_1 \rangle_M k'_j}. \tag{52}
$$

This represents the efficiency of the search at detecting the injections. (Note that, if $N_{\rm inj} > M'$, then there are injections that did not produce any recorded event; these injections are presumed to be missed entirely.) Specifically, if the injection set surveyed a spacetime volume $(VT)_0$, then the search was sensitive to a spacetime volume

$$\overline{VT} = \overline{\Delta\langle\Lambda_1\rangle}(VT)_0 = N_{\rm rec}\Delta(VT) \tag{53}$$

where

$$\Delta(VT) = \frac{(VT)_0}{N_{\rm inj}} \tag{54}$$

is the incremental increase in VT per found injection and

$$N_{\rm rec} = {\rm Var}_M(\Lambda_1)\sum_{j=1}^{M'}\frac{k'_j}{1+\langle\Lambda_1\rangle_M k'_j} \tag{55}$$

is the average number of recovered injections.

The posterior on the rate $R = \Lambda_1/\overline{VT}$ is therefore

$$p(R\,|\,\{k_1,k_2,\ldots,k_M\}) \propto p(R)\exp(-\overline{VT}R)\prod_{i=1}^{M}(1+\overline{VT}Rk_i). \tag{56}$$

To include uncertainties in the spacetime volume count and in such things as calibration, marginalize the result with a log normal distribution:

$$p(R\,|\,\{k_1,k_2,\ldots,k_M\},S) = \frac{1}{S\sqrt{2\pi}}\int_0^\infty dx\,\exp[-(\ln x)^2/2S^2]\,p(xR\,|\,\{k_1,k_2,\ldots,k_M\}) \tag{57}$$

where S is the fractional error. This fractional error can consist of two components:

- A statistical error from counting uncertainty in the injection campaign. The fractional uncertainty is

$$S_{\rm stat}^2 = {\rm Var}\left(\log\overline{VT}\right) = {\rm Var}(\log N_{\rm rec}) = \frac{1}{N_{\rm rec}}. \tag{58}$$

- A systematic fractional error from calibration uncertainty is

$$S_{\rm syst} = \frac{\Delta(h^3)}{(h^3)} = 3\frac{\Delta h}{h}. \tag{59}$$

- The total fractional error is

$$S = \sqrt{S_{\rm stat}^2 + S_{\rm syst}^2} = \sqrt{\frac{1}{N_{\rm rec}} + \left(3\frac{\Delta h}{h}\right)^2}. \tag{60}$$

4 Implementation Details

Many of the integrals that need to be computed are of the form

$$\int_0^\infty x^\alpha e^{-x}dx\,P(x) \quad (\alpha > -1) \tag{61}$$

where $P(x)$ is a polynomial in x. Recall the 1-dimensional posterior, Eq. (45),

$$p(\Lambda_1\,|\,\{k_1,k_2,\ldots,k_M\}) \propto p(\Lambda_1)\exp(-\Lambda_1)\prod_{i=1}^{M}(1+\Lambda_1 k_i)$$

and note that, if $p(\Lambda_1) \propto \Lambda_1^\alpha$, then moments of this distribution will be of the aforementioned form. In particular, $\langle\Lambda_1\rangle$ and $\mathrm{Var}\Lambda_1$ require integrals of the form of Eq. (61). Indeed, even the integral in Eq. (46),

$$P_1(k_i \,|\, \{k_1, k_2, \ldots, k_M\}) = \int_0^\infty d\Lambda_1\, p(\Lambda_1 \,|\, \{k_1, k_2, \ldots, k_M\}) \frac{\Lambda_1 k_i}{1 + \Lambda_1 k_i}$$

is of the form of Eq. (61) since the divisor is one of the factors in the probability density function. Such integrals can be readily evaluated using *generalized Gauss-Laguerre quadrature*:

$$\int_0^\infty x^\alpha \mathrm{e}^{-x} dx\, P(x) = \sum_{j=1}^{n} w_j P(x_j) \tag{62}$$

where x_j is the jth root of the Generalized Laguerre function $L_n^{(\alpha)}(x)$, $\alpha > -1$, and w_j are the weighting factors

$$w_j = \frac{\Gamma(n+\alpha) x_j}{n!(n+\alpha)![L_{n-1}^{(\alpha)}(x_j)]^2}. \tag{63}$$

The integral is *exactly* equal to the sum if the degree of the polynomial $P(x)$ is $2M-1$ or less. In practice, however, only a handful of events will contribute significantly to the posterior distribution and it is bound that $M = 20$ is sufficient for evaluation of the integrals of interest.

The abscissas and weights for the generalized Gauss-Laguerre quadratures are obtained with the routine `scipy.special.la_roots`

```
>>> from scipy.special import la_roots
>>> x, w = la_roots(n = 20, alpha = -0.5)
>>> x
array([ 3.04632393e-02, 2.74444716e-01, 7.63887558e-01,
        1.50180150e+00, 2.49283015e+00, 3.74341804e+00,
        5.26205585e+00, 7.05962774e+00, 9.14989831e+00,
        1.15501983e+01, 1.42824037e+01, 1.73743670e+01,
        2.08620752e+01, 2.47930399e+01, 2.92319102e+01,
        3.42704289e+01, 4.00468158e+01, 4.67888464e+01,
        5.49315556e+01, 6.55899320e+01])
>>> w
array([ 6.77286555e-01, 5.31456504e-01, 3.26757465e-01,
        1.56949212e-01, 5.86251311e-02, 1.69217760e-02,
        3.74299366e-03, 6.27707189e-04, 7.87386796e-05,
        7.26315230e-06, 4.82228833e-07, 2.24247217e-08,
        7.05124158e-10, 1.43130561e-11, 1.76114153e-13,
        1.20167176e-15, 3.97836202e-18, 5.13518673e-21,
        1.70881139e-24, 5.18208743e-29])
```

Integrals of the form

$$\int_0^\infty p(\Lambda_1 \,|\, \{k_1, k_2, \ldots, k_M\})\, d\Lambda_1\, f(\Lambda_1)$$

can be accurately evaluated as

$$\int_0^\infty p(\Lambda_1 \,|\, \{k_1, k_2, \ldots, k_M\})\, d\Lambda_1\, f(\Lambda_1) = \sum_{j=1}^{n} p_j f(x_j) \tag{64}$$

where

$$p_j \propto w_j \prod_{i=1}^{M} (1 + x_j k_i). \tag{65}$$

The normalization is determined by

$$1 = \int_0^\infty p(\Lambda_1 \mid \{k_1, k_2, \ldots, k_M\})\, d\Lambda_1 = \sum_{j=1}^{n} p_j \tag{66}$$

whence

$$p_j = \frac{w_j \prod_{i=1}^{M}(1 + x_j k_i)}{\sum_{j=1}^{n} \left\{ w_j \prod_{i=1}^{M}(1 + x_j k_i) \right\}}. \tag{67}$$

The abscissas and probability-weights, $\{(x_1, p_1), (x_2, p_2), \ldots, (x_n, p_n)\}$ can be pre-computed one time and then used to evaluate multiple integrals.

For example, to evaluate the probability that an event is a foreground event for *every* event, we compute

$$P_1(k_i \mid \{k_1, k_2, \ldots, k_M\}) = 1 - \int_0^\infty d\Lambda_1\, p(\Lambda_1 \mid \{k_1, k_2, \ldots, k_M\}) \frac{1}{1 + \Lambda_1 k_i} = 1 - \sum_{j=1}^{n} \frac{p_j}{1 + x_j k_i}. \tag{68}$$

This can be vectorized as follows: let

$$\mathbf{k} = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_M \end{bmatrix}, \quad \mathbf{P_1} = \begin{bmatrix} P_1(k_1 \mid \mathbf{k}) \\ P_1(k_2 \mid \mathbf{k}) \\ \vdots \\ 1 - P_1(k_M \mid \mathbf{k}) \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{and} \quad \mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}; \tag{69}$$

then

$$\mathbf{P_1} = 1 - \frac{1}{1 + \mathbf{k} \otimes \mathbf{x}} \cdot \mathbf{p} \tag{70}$$

or, in Python code,

```
>>> from numpy import dot, logspace, outer, prod
>>> from scipy.special import la_roots
>>> K = logspace(-2, 2)                                 # faked Bayes factors
>>> k = K / len(K)                                      # Eq. (44)
>>> x, w = la_roots(n = 20, alpha = -0.5)
>>> p = w * prod(1.0 + outer(k, x), axis = 0)           # Eq. (65)
>>> p /= sum(p)                                         # Eq. (66)
>>> P1 = 1.0 - dot(1.0 / (1.0 + outer(k, x)), p)       # Eq. (70)
>>> P1
array([ 0.00413129, 0.00498108, 0.00600455, 0.00723669, 0.00871931,
        0.01050226, 0.01264486, 0.01521744, 0.01830312, 0.02199965,
        0.02642145, 0.0317015 , 0.03799315, 0.04547152, 0.05433415,
        0.06480062, 0.07711039, 0.09151845, 0.108288 , 0.12767956,
        0.14993601, 0.17526363, 0.20380961, 0.23563752, 0.27070322,
        0.30883437, 0.34971767, 0.39289726, 0.43778697, 0.48369712,
        0.52987384, 0.57554655, 0.61997765, 0.66250771, 0.70259073,
        0.73981599, 0.77391547, 0.80475829, 0.83233517, 0.85673671,
        0.87812911, 0.89673032, 0.91278882, 0.92656592, 0.93832222,
        0.94830784, 0.95675604, 0.96387959, 0.96986915, 0.97489318])
>>> 1 - 0.5 + sum(P1)                                   # mean of counts posterior using Eq. (48)
20.747642701543494
>>> dot(p, x)                                           # mean of counts posterior using Eq. (64)
20.747642701543498
>>> dot(p, x**2) - dot(p, x)**2                         # variance of counts posterior using Eq. (64)
27.752127849039539
```

Notice that the computation of M values of P_1 requires only $O(nM)$ operations.

References

[1] Will M. Farr, Jonathan R. Gair, Ilya Mandel, and Curt Cutler (2015) "Counting and confusion: Bayesian rate estimation with multiple populations" Phys. Rev. D **91**, 023005 DOI: 10.1103/PhysRevD.91.023005

[2] Thomas J. Loredo and Ira M. Wasserman (1995) "Inferring the spatial and energy distribution of gamma-ray burst sources. I. Methodology" Astrophys. J Supp. Ser. **96**, 261 DOI: 10.1086/192119

[3] B. P. Abbott et al. (2016) "Supplement: The rate of binary black hole mergers inferred from advanced LIGO observations surrounding GW150914" Astrophys. J Supp. Ser. **227**, 14 DOI: 10.3847/0067-0049/227/2/14