Crater Counting/ v.77/October 11, 2016

We consider modeling for a "crater-counting" data set consisting of

- N_{2+} locations, sizes and records of exactly which 2 or more of J analysts counted each of the corresponding craters, and
- for $j = 1, 2, \ldots, J$, a number S_1^j of locations and sizes of possible craters counted only by analyst j, each of which may be either a real crater (missed by all other analysts) or a "phantom" non-crater seen only by analyst j (we'll call the number of actual craters in this group N_1^j and write $F_j = S_1^j - N_1^j$ for the number of phantoms seen by analyst j).

We'll do this in a way that is consistent with independent HPP models for locations of craters and phantoms across an area A, in particular with locations of real craters having intensity ρ and phantoms from analyst j having intensity ρ_j for $j = 1, 2, \ldots, J$. We'll further allow that there it is possible for analysts to miss craters in their counting and that there are N_0 craters missed by all in the counting (and that ultimately it is $N = N_0 + N_{2+} + \sum_{j=1}^{J} N_1^j$ that is Poisson with mean ρA).

So, suppose that analyst j misses a real crater of size s in his/her counting with probability

$$p(s \mid \boldsymbol{\gamma}_j)$$

(In more complex modeling, this probability could be a function of other things, like local conditions around the crater center under consideration perhaps specified in terms of a number of other centers within some distance, and perhaps an overall density/number of real craters.) Exactly what form to use for this is open to discussion. It could be something as simple as

$$p(s \mid \gamma_1, \gamma_2) = I[s < \gamma_1] + \gamma_2 I[s \geq \gamma_1]$$

a step function taking only the value 1 at and near 0 and some positive value γ_2 to the right of a cut point γ_1. Another possible form that looks like it might be tractable is

$$p(s \mid \gamma_1, \gamma_2, \gamma_3) = I[s < \gamma_1] + I[s \geq \gamma_1]\left(\gamma_2 + (1-\gamma_2)\exp\left(-\frac{s-\gamma_1}{\gamma_3}\right)\right)$$

This function is 1 at and below the threshold $s = \gamma_1$, it has limit γ_2 as $s \to \infty$, to which it decreases in exponential fashion with a rate parameter γ_3.

Next, suppose that real craters have sizes that are iid with marginal density

$$f(s \mid \boldsymbol{\theta})$$

Further, suppose that phantom craters for analyst j have recorded/perceived sizes that are iid with marginal density

$$h(s \mid \boldsymbol{\eta}_j)$$

Plots in the Robbins paper seem to make values of a cdf (or 1 minus the cdf) for crater size look roughly linear over some interval on log-log scales. If a continuous cdf has

$$\ln F(s \mid \theta_0, \theta_1) = (\theta_0 + \theta_1 \ln s) I[\theta_0 + \theta_1 \ln s < 0]$$

then

$$F(s \mid \theta_0, \theta_1) = I[\theta_0 + \theta_1 \ln s < 0] \exp(\theta_0 + \theta_1 \ln s) + I[\theta_0 + \theta_1 \ln s \geq 0]$$

and

$$\begin{aligned} f(s \mid \theta_0, \theta_1) &= I[\theta_0 + \theta_1 \ln s < 0]\left(\frac{\theta_1}{s}\right) \exp(\theta_0 + \theta_1 \ln s) \\ &= \theta_1 \exp(\theta_0) s^{\theta_1 - 1} I\left[0 < s < \exp\left(-\frac{\theta_0}{\theta_1}\right)\right] \end{aligned}$$

and perhaps this form for $\theta_1 > 0$ and $\theta_0 < 0$ would work as a marginal pdf of real crater diameters.

A plausible parametric form for the density for phantom diameters is up for discussion. Something as simple as the $\mathrm{U}(\eta_1, \eta_2)$ density might be used, as might the kind of form suggested above for real diameters.

In any event, the basic parameters of the modeling are

$$\boldsymbol{\theta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_J, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_J, \rho, \rho_1, \rho_2, \ldots, \text{ and } \rho_J$$

Various simplified versions of what follows can be had by assuming that some parameter(s) is (are) fixed across j, or are known, etc.

We proceed to consider some basic quantities associated with the mechanisms for generating sizes and observing craters of a given size. In what follows, let

$$\mathbf{I} = (I_1, I_2, \ldots, I_J)$$

be a vector of 0's and 1's. Define

$$q^{\mathbf{I}}(s \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J) = \prod_{j=1}^{J} \left(1 - p\left(s \mid \boldsymbol{\gamma}_j\right)\right)^{I_j} \left(p\left(s \mid \boldsymbol{\gamma}_j\right)\right)^{1-I_j}$$

= the probability that a crater of size s is seen by exactly those analysts j with $I_j = 1$

so that in particular, with $\mathbf{0}$ a vector of 0's,

$$q^{\mathbf{0}}(s \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J) = \prod_{j=1}^{J} p\left(s \mid \boldsymbol{\gamma}_j\right)$$

= the probability that a crater of size s is not seen by any analyst

We'll let

$$q_1^j(s \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J) = \left(1 - p\left(s \mid \boldsymbol{\gamma}_j\right)\right) \prod_{j' \neq j} p\left(s \mid \boldsymbol{\gamma}_{j'}\right)$$

= the probability that a crater of size s is seen only by analyst j

and abbreviate as

$$q_{2+}(s \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J) = 1 - q^{\mathbf{0}}(s \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J) - \sum_{j=1}^{J} q_1^j(s \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J)$$

= the probability that a crater of size s is seen by at least 2 analysts

These quantities are obviously all functions of size and depend upon the parameters $\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J$ that describe the analyst-specific crater-detection properties.

Versions of all the quantities q averaged across sizes of craters can be defined and depend upon both the parameters $\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J$ and upon the parameter $\boldsymbol{\theta}$ that describes the size distribution. That is, we define

$$q^{\mathbf{I}}(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}) = \int q^{\mathbf{I}}(s \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J)\, f(s \mid \boldsymbol{\theta})\, ds$$

= the probability of a crater detection pattern $\mathbf{I}$ across the analysts

and

$$q^{\mathbf{0}}(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}) = \int q^{\mathbf{0}}(s \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J)\, f(s \mid \boldsymbol{\theta})\, ds$$

= the probability a crater is seen by no analyst

and

$$q_1^j(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}) = \int q_1^j(s \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J)\, f(s \mid \boldsymbol{\theta})\, ds$$

= the probability a crater is seen only by analyst j

and

$$q_{2+}\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)=1-q^{\mathbf{0}}\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)-\sum_{j=1}^{J}q_1^j\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)$$

= the probability that a crater is seen by at least 2 analysts

These integrals are useful in defining some important conditional probabilities and conditional densities. In particular, for an $\mathbf{I}$ with at least 2 entries of 1,

$$\frac{q^{\mathbf{I}}\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)}{q_{2+}\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)}$$ = the conditional probability of crater-detection pattern $\mathbf{I}$ across analysts given that it is seen by at least 2 analysts

Also, the conditional pdf of crater size given detection pattern $\mathbf{I}$ across analysts is

$$\frac{q^{\mathbf{I}}\left(s\,|\,\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J\right)f\left(s\,|\,\boldsymbol{\theta}\right)}{q^{\mathbf{I}}\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)}$$

and in particular, the conditional pdf of crater size given detection by only analyst j is

$$\frac{q_1^j\left(s\,|\,\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J\right)f\left(s\,|\,\boldsymbol{\theta}\right)}{q_1^j\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)}$$

Conditional on the parameters $\boldsymbol{\theta},\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\ldots,\boldsymbol{\eta}_J,\rho,\rho_1,\rho_2,\ldots,$ and ρ_J, we suppose that there are independent random variables

$$N_0\sim\text{Poisson}\left(q^{\mathbf{0}}\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho A\right)$$

and

$$N_{2+}\sim\text{Poisson}\left(q_{2+}\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho A\right)$$

and for $j=1,2,\ldots,J$

$$S_1^j\sim\text{Poisson}\left(q_1^j\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho A+\rho_j A\right)$$

Then, for each $j=1,2,\ldots,J$ suppose that there are latent variables $T_1^j,T_2^j,\ldots,T_{S_1^j}^j$ that are (conditional on all before) independent Bernoulli variables with success probabilities

$$\frac{\rho q_1^j\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)}{\rho q_1^j\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)+\rho_j}$$

(indicators of a possible crater being real and not a phantom). Then

$$N_1^j = \sum_{l=1}^{S_1^j} T_l^j \text{ and } F^j = S_1^j - N_1^j$$

are independent Poisson variables, N_1^j (the number of real craters seen only by analyst j) with mean $q_1^j(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta})\rho A$ and F^j (the number of phantoms reported by analyst j) with mean $\rho_j A$. Notice then that

$$N_0 + N_{2+} + \sum_{j=1}^{J} N_1^j$$

(as a sum of independent Poisson variables) is Poisson with mean

$$q^{\mathbf{0}}(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta})\rho A + q_{2+}(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta})\rho A + \sum_{j=1}^{J} q_1^j(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta})\rho A = \rho A$$

So we have here a probability structure for the counts that is consistent with the Poisson process assumptions, the crater size generating assumptions, and the analyst crater-detection assumptions. It remains to build the part of the model that will attach patterns of observer detection $\mathbf{I}$ to the N_{2+} craters seen by 2 or more analysts and sizes to each of the $N_{2+} + \sum_{j=1}^{J} S_1^j$ observed potential craters and to the N_0 unobserved craters. This we can do using appropriate conditional probabilities and pdfs.

First, we suppose that conditional on all else, sizes of N_0 unobserved craters are iid according to the pdf

$$\frac{q^{\mathbf{0}}(s \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J) f(s \mid \boldsymbol{\theta})}{q^{\mathbf{0}}(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta})}$$

Then (still conditioned on all parameters and counts) independent of these sizes (independently for all $j = 1,2,\ldots,J$ and $l = 1,2,\ldots,S_1^j$) conditioned on $T_l^j = 1$ suppose that an lth potential crater seen only by analyst j has size from the density

$$\frac{q_1^j(s \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J) f(s \mid \boldsymbol{\theta})}{q_1^j(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta})}$$

while conditioned on $T_l^j = 0$ it has size from the density $h(s \mid \boldsymbol{\eta}_j)$.

Finally, consider modeling of vectors of crater-detection patterns and sizes for the N_{2+} craters seen by 2 or more analysts. Conditional on the parameters and all else, these pairs can be taken as iid according to the recipe that first $\mathbf{I}$ with at least 2 entries of 1 has pmf given by

$$m\left(\mathbf{I} \mid \boldsymbol{\gamma}_1, , \ldots, \boldsymbol{\gamma}_J, \boldsymbol{\theta}\right) = \frac{q^{\mathbf{I}}\left(\boldsymbol{\gamma}_1, , \ldots, \boldsymbol{\gamma}_J, \boldsymbol{\theta}\right)}{q_{2+}\left(\boldsymbol{\gamma}_1, , \ldots, \boldsymbol{\gamma}_J, \boldsymbol{\theta}\right)}$$

and subsequently a size is selected from the density

$$\frac{q^{\mathbf{I}}\left(s \mid \boldsymbol{\gamma}_1, , \ldots, \boldsymbol{\gamma}_J\right) f\left(s \mid \boldsymbol{\theta}\right)}{q^{\mathbf{I}}\left(\boldsymbol{\gamma}_1, , \ldots, \boldsymbol{\gamma}_J, \boldsymbol{\theta}\right)}$$

What must be sampled in order to do a Bayes analysis here? Of course, one must begin with a prior of some kind for the model parameters, say

$$g\left(\boldsymbol{\theta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_J, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_J, \rho, \rho_1, \rho_2, \ldots, \rho_J\right)$$

Then, the data and latent variables have "densities" conditional on the parameters that are as follows.

First, there are N_{2+} cases with at least 2 observers, and this has probability

$$\frac{\exp\left(-q_{2+}\left(\boldsymbol{\gamma}_1, , \ldots, \boldsymbol{\gamma}_J, \boldsymbol{\theta}\right) \rho A\right) \left(q_{2+}\left(\boldsymbol{\gamma}_1, , \ldots, \boldsymbol{\gamma}_J, \boldsymbol{\theta}\right) \rho A\right)^{N_{2+}}}{N_{2+}!}$$

Each such case l then has likelihood

$$\frac{q^{\mathbf{I}}\left(\boldsymbol{\gamma}_1, , \ldots, \boldsymbol{\gamma}_J, \boldsymbol{\theta}\right)}{q_{2+}\left(\boldsymbol{\gamma}_1, , \ldots, \boldsymbol{\gamma}_J, \boldsymbol{\theta}\right)} \cdot \frac{q^{\mathbf{I}}\left(s_l \mid \boldsymbol{\gamma}_1, , \ldots, \boldsymbol{\gamma}_J\right) f\left(s_l \mid \boldsymbol{\theta}\right)}{q^{\mathbf{I}}\left(\boldsymbol{\gamma}_1, , \ldots, \boldsymbol{\gamma}_J, \boldsymbol{\theta}\right)}$$

So ultimately, the contribution to the likelihood from those cases is proportional to

$$\exp\left(-q_{2+}\left(\boldsymbol{\gamma}_1, , \ldots, \boldsymbol{\gamma}_J, \boldsymbol{\theta}\right) \rho A\right) \prod_l q^{\mathbf{I}}\left(s_l \mid \boldsymbol{\gamma}_1, , \ldots, \boldsymbol{\gamma}_J\right) f\left(s_l \mid \boldsymbol{\theta}\right)$$

There is also a latent variable N_0 that has pmf

$$\frac{\exp\left(-q^{\mathbf{0}}\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho A\right)\left(q^{\mathbf{0}}\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho A\right)^{N_0}}{N_0\,!}$$

and this is the contribution to the "likelihood" from this variable.

Finally, there are S_1^J counts made by observer j alone. This count variable has pmf

$$\frac{\exp\left(-\left(q_1^j\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho A+\rho_j A\right)\right)\left(q_1^j\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho A+\rho_j A\right)^{S_1^j}}{S_1^j}$$

Each of the S_1^j cases (say, l) then has $T_l^j=1$ with probability

$$\frac{q_1^j\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho}{q_1^j\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho+\rho_j}$$

and, of course, $T_l^j=0$ with probability

$$\frac{\rho_j}{q_1^j\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho+\rho_j}$$

Conditioned on $T_l^j=1$ and all else, s_l has density

$$\frac{q_1^j\left(s\,|\,\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J\right)f\left(s\,|\,\boldsymbol{\theta}\right)}{q_1^j\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)}$$

and conditioned on $T_l^j=0$ it has density $h\left(s\,|\,\boldsymbol{\eta}_j\right)$. So the contribution to the likelihood for these cases is

$$\exp\left(-\left(q_1^j\left(\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho A+\rho_j A\right)\right)$$
$$\times\prod_l\left(\rho q_1^j\left(s_l\,|\,\boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J\right)f\left(s_l\,|\,\boldsymbol{\theta}\right)\right)^{T_l^j}\times\prod_l\left(\rho_j h\left(s_l\,|\,\boldsymbol{\eta}_j\right)\right)^{1-T_l^j}$$

There is a product of this type for every j included in the likelihood.

+++

Notice that in an MCMC algorithm, a Gibbs step update for a T_l^j will be made using the conditional probability that $T_l^j = 1$, i.e.

$$\frac{q_1^j\left(s_l \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho f\left(s_l \mid \boldsymbol{\theta}\right)}{q_1^j\left(s_l \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho f\left(s_l \mid \boldsymbol{\theta}\right)+\rho_j h\left(s_l \mid \boldsymbol{\eta}_j\right)}$$

and, of course, that $T_l^j = 0$, i.e.

$$\frac{\rho_j h\left(s_l \mid \boldsymbol{\eta}_j\right)}{q_1^j\left(s_l \mid \boldsymbol{\gamma}_1,,\ldots,\boldsymbol{\gamma}_J,\boldsymbol{\theta}\right)\rho f\left(s_l \mid \boldsymbol{\theta}\right)+\rho_j h\left(s_l \mid \boldsymbol{\eta}_j\right)}$$