Mitman, Jeffrey

From: Ferrante, Fernando
Sent: Monday, March 08, 2010 1:13 PM
To: Laur, Steven; James, Lois; Vail, James; Mitman, Jeffrey
Subject: RE: Dam statistics

Steven,

Thanks, this really will help in the following sections in NUREG/CR-6823 which I was also working on.

Also, thanks for the discussion with Jeff Mitman and I. We talked to Jim Vail and we believe we have a path forward to complete the revisions to the document as soon as possible.

Fernando

From: Laur, Steven
Sent: Monday, March 08, 2010 12:42 PM
To: Ferrante, Fernando; James, Lois; Vail, James; Mitman, Jeffrey
Subject: RE: Dam statistics

I agree with Fernando that equation 8.2 in NUREG/CR-6823 is in error.

FYI, I believe I see how it was "derived" and where the error must be.

We are looking for $\Psi(\alpha + xi) - \Psi(\alpha)$, which is defined as $d/dx\{[\ln(\Gamma(\alpha + xi)) - \ln(\Gamma(\alpha))]\}$.

Now $\Gamma(x) = (x - 1)^* \Gamma(x - 1)$, and xi is an integer in the above, so:

$\ln(\Gamma(\alpha + xi)) - \ln(\Gamma(\alpha)) = \ln\{\Gamma(\alpha + xi) / \Gamma(\alpha)\}$

The numerator inside the logarithm may be written as:

$(\alpha + xi - 1)^* (\alpha + xi - 2) * \ldots * (\alpha + xi - xi)^* \Gamma(\alpha)$ or $\Pi(j = 1 \text{ to } xi) \{(\alpha + xi - j)^* \Gamma(\alpha)\}$

The $\Gamma(\alpha)$ "cancels out" since it is the denominator as well.

So: since the log of a product is the sum of logs, so we get:

$\ln(\Gamma(\alpha + xi)) - \ln(\Gamma(\alpha)) = \Sigma(j= 1 \text{ to } xi)\{\ln (\alpha + xi - j)\}$

Unfortunately, we have no function left: just a bunch of constants (whose derivative would be zero). But, thinking $d(\ln x)/dx$ is $1/x$, the authors wrote the sum of the reciprocals of the above, shown in equation 8.2.

Ah well, it was a neat looking formula ...

Steven A. Laur
NRR Division of Risk Assessment
OWFN 10-C15
(301) 415-2889
steven.laur@nrc.gov

B-91

From: Ferrante, Fernando
Sent: Friday, March 05, 2010 5:55 PM
To: Laur, Steven; James, Lois; Vail, James; Mitman, Jeffrey
Subject: RE: Dam statistics

Steven,

Jeff Mitman and I took a stab at incorporating and/or adressing some of your comments (see attached).

We also noticed the issue with respect to Dale Rasmussen's calculations and discussed it with Selim Sancaktar in RES. Our understanding is the same as what you expressed. Additionally, I personally agree that it would be better to use numbers we can explain and reproduce clearly (although I suspect that what Dale did is input the data into SAS and use empirical Bayes with some kind of adjustment).

I used the method in NUREG/CR-6823 to derive the prior and implemented it in both MATLAB and Mathcad. All that is really needed is a package that has some basic mathematical functions. I am attaching a few files exported from Mathcad that demonstrate the method using either (i) the numbers from example 8.1 in NUREG/CR-6823 (which replicates the results on page 8-9), and (ii) the dam failure numbers and dam-years used as input in the spreadsheet from RES. I may be able to do this in EXCEL too (macros would have to be developed).

The results are not the same (alpha = 2.0362, beta = 7023.7 instead of alpha = 2.4026, beta = 10094.7) which is why I believe some adjustment was made in SAS by Dale beyond what is indicated in NUREG/CR-6823. NUREG/CR-6823 also suggests the use of a Kass-Stefey adjustment as explained in 8.2.4.1 but this would only affect the posterior distributions, not the priors. Once the prior is obtained, it is straight-forward to obtain the individual posterior distributions exactly as shown in the spreadsheet from RES, we can add a discussion if this is still seen as pertinent.

The other comments that we thought needed further discussion are:

- Shouldn't this number change somewhat as the cut-off year is changed? There were 6 failures since 1900, 4 failures since 1940. How many since the other years shown? (This could be shown on the table by making the point estimate **bold**, for example.) Also, since 6 is the max, why not include 7 failures in this table (e.g., show 2 – 7 failures instead of 1 – 6)?

- Although counterintuitive, the change from a cut-off from 1940 to 1900 actually reflects the data. Because extending the cut-off from 1940 to 1900 adds both failures in the numerator as well as dam-years in the denominator, the change in the failure rate may actually not be significant depending on the numbers. To illustrate this, a file titled 'Failure Rate Difference.rtf' is attached that shows how the change appears consistent with the additional failures and dam-years resulting from 1940-2005 to 1900-2005 (i.e., the delta between failure rates is not significant). This is a convoluted way of explaining this, please let me know if it does not make sense.

- There are almost a dozen dam categories. This implies you have a prior distribution for each of the categories. If so, where are the alphas and betas for them? If this is for ALL dams, it should say that.
- For repeatability, you should provide the number of failures and dam-years assumed in the prior.
- This appears to say there were 10 failures in 7765 dam years, since your reference says the posterior alpha and beta are found by adding # of failures to alpha prior and dam-years to beta prior. Please define in this paragraph how the posterior parameters were derived.

- Some of the discussion above on the Bayesian analysis has implications on these comments, and I believe your discussion with Lois helped clarify somewhat what was done. We can discuss this further next week although I believe the real issues are (i) can the SAS results be repeated (and are they correct), or (ii) should

they be substituted by something that is at least repeatable? I did not adress the issue on your previous e-mail regarding the discrepancy with respect to your spreadsheet analysis but I will take a look at it next week.

- Should we add another sentence to the effect: "Dam-specific failure frequencies outside the range presented in this report would likely depend on a non-data approach, such as event tree/fault tree modeling." (?)

- Our intent was to adress specifically the use of data on dam failures to derive a failure rate for a large dam such as Jocassee. Although event tree approaches are available for dams, we believe that discussing this issue here may provide the impression that a more "refined" method automatically translates into a way to obtain a failure rate lower than the range presented (especially when there is already an intent to disregard the data results as too conservative). We believe the data used has value to the analysis and the results reflect a range that several other references also obtained. To prevent this misperception, we would rather not include a sentence as suggested. A potential interesting method to address some of this issues is discussed in the 1982 report from Los Alamos, referenced in the document (but this is beyond what we intended to discuss in the analysis).

I will be available to discuss the issue on Monday, please let me know if you are available.

Have a nice weekend,

Fernando Ferrante, Ph.D.
Office of Nuclear Reactor Regulation (NRR)
Division of Risk Assessment (DRA)
Operational Support and Maintenance Branch (APOB)
Mail Stop: 0-10C15
Phone: 301-415-8385
Fax: 301-415-3577

From: Laur, Steven
Sent: Thursday, March 04, 2010 7:09 PM
To: James, Lois; Vail, James; Mitman, Jeffrey; Ferrante, Fernando
Subject: Dam statistics

Jim and Lois:

I believe Lois is correct that Dale Rasmussen's analysis of all dams (top part of his spreadsheet) used the "Constrained Noninformative Prior" (paragraph 6.2.2.5.3 of NUREG/CR-6823).

The lower part (dams over 50') uses the empirical bayes method - see section 8.2.2, "MLE Equations for the Gamma-Poisson Model," in NUREG/CR-6823. It tells you how to estimate the alpha and beta for the prior Gamma using the poisson parameter lambda. However, it implies (to me at least) that a software package is the only practical way to accomplish it!

Just to let you know I'm thinking about it ...

Steven A. Laur
Sr Tech Adv - Risk Informed Initiatives
steven.laur@nrc.gov
(301) 415-2889