LA-UR-11-01857

Approved for public release; distribution I unlimited.

Title: Modeling the Number of Ignitions Following an Earthquake: Developing Prediction Limits for Overdispersed Count Data

Authors: Elizabeth J. Kelly and Raymond N. Tell

Intended Use: Deliverable to SB-TS: Safety Basis Technical Services

Modeling the Number of Ignitions Following an Earthquake: Developing Prediction Limits for Overdispersed Count Data

ABSTRACT

This report describes an approach for modeling the number of ignitions (fires) following an earthquake. The modeling is not meant to be exact, but to provide a context for assessing the likelihood of various fire scenarios. The first component of the approach is a statistical model to predict the number of ignitions for a new earthquake event. This model is based on data for ignitions following earthquakes from 1906 to 1989 in Alaska and California. These U.S. fire data are taken from reports by fire departments on the fires they responded to immediately after the earthquakes and for several days thereafter. These data are for fires in the general built environment, including residential, commercial and industrial structures. The data contain estimates for the mean peak ground acceleration (*PGA*) for each earthquake, an estimate of the built area affected in million square feet (*MMSF*) for each earthquake, and the number of ignitions within the estimated affected area (*IGNS*). The statistical model uses negative binomial regression to estimate the expected number of ignitions as a function of the explanatory variables, *PGA* and *MMSF*. The associated upper confidence and prediction limits are derived from the statistical model using only spreadsheet technology. The upper prediction limit is used to determine a conservative estimate of the probability of a specified number of ignitions following a future earthquake event. The results from the spreadsheet technology are compared to more exact results based on numerical integration. The spreadsheet probability estimates are shown to be conservative.

However, these fire data are limited in two ways. First, there are no estimates of the number of fires that may not have been responded to by the fire department, e.g. unreported fires following an earthquake. Second, the terms "fire" and "ignition" are used interchangeably; there are no data on the number of ignitions causing the fire. The second component of the approach provides methods for adjusting the statistical model to account for these limitations of the data. This report also provides an example of an application of this approach to a large single structure.

BACKGROUND

This report is concerned with determining a conservative estimate of the probability of a specified number of ignition events in buildings (including a large single structure), which might occur because of an earthquake. Conservatism is meant in the sense that the estimate of the probability of a fire is greater than the true probability. These values are not meant to be exact, but to provide a context for assessing the likelihood of various fire scenarios.

The publication, “Fire Following Earthquake,” Technical Council on Lifeline Earthquake Engineering Monograph No. 26 (Scawthorn et. al.,2005) provides data on fires following earthquakes in the United States. Table 4-1 of the Monograph, *Fires Following U.S. Earthquakes – 1906 – 1989* contains the data from Alaska and California used in this report. These data are taken from reports by fire departments on the fires they responded to immediately after the earthquakes and for several days thereafter. The data are for fires in the general built environment, including residential, commercial and industrial structures.

Table 4-1 of the Monograph contains the estimate for the earthquake mean peak ground acceleration (*PGA*), the corresponding number of ignitions (*IGNS*) and the ratio of ignitions to million square feet of estimated affected building space (*IGNS*/*MMSF*). The PGA estimates are based upon a standard attenuation model or direct measurements where available (communication with John Eidinger). The MMSF estimates are based on definable geographic entities such as cities.

The *MMSF* values can be calculated from these data for all events with non-zero values of *IGNS*. The Monograph data and the calculated *MMSF* values are given in Table 1. For the two events with zero ignitions, the *MMSF* values were provided by Doug Honegger (D.G. Honegger Consulting).

There are no data in the Monograph describing the size of the individual burning structures. However, the Monograph does describe Single Family Equivalent Dwellings (SFED), which consist of 1500 sq. ft areas that can be thought of as detached houses. The Monograph discussion (page 105) indicates that thinking of the SFED as the basic unit for a fire is a reasonable approach for communication to fire departments on the number of separate structural fires to expect, and that “a large building of 1.5 *MMSF*” can be thought of as “1000 SFED.” The Monograph develops models to estimate ignitions per SFED. This conceptual model underlies the application of the statistical model to a single structure.

In the Monograph the terms “fire” and “ignition” are used interchangeably. The Monograph does not address the possibility of multiple ignitions in a single structure. Because of this, the number of separate ignitions may be underreported. Another possible limitation of the data is that there are no estimates of the number of fires that may not have been responded to by the fire department, e.g. unreported fires.

Table 1. *PGA*, Ignitions (*IGNS*), Ignitions per Million ft^2 (*IGNS/MMSF*) from Table 4-1 of the Monograph (USA 1906 – 1989) and the calculated *MMSF*.

PGA (g)	Ignitions (*IGNS*)	*IGNS* / *MMSF*	*MMSF*
0.36	1	0.3	3.33
0.12	3	0.05	60.00
0.71	7	0.24	29.17
0.44	1	0.16	6.25
0.07	1	0.013	76.92
0.21	7	0.16	43.75
0.15	9	0.13	69.23
0.15	128	0.09	1422.22
0.15	3	0.01	300.00
0.53	19	0.26	73.08
0.12	2	0.02	100.00
0.21	4	0.4	10.00
0.21	1	0.02	50.00
0.28	1	0.05	20.00
0.44	2	0.06	33.33
0.07	0	0	160
0.21	2	0.04	50.00
0.21	27	0.08	337.50
0.44	52	0.26	200.00
0.12	0	0	350
0.53	3	0.37	8.11
0.36	5	0.02	250.00
0.36	1	0.08	12.50
0.44	1	0.22	4.55
0.36	1	0.04	25.00
0.28	24	0.03	800.00
0.36	1	0.14	7.14
0.36	1	0.06	16.67
0.71	1	0.18	5.56
0.28	6	0.1	60.00

STATISTICAL MODEL[1] FOR THE MONOGRAPH DATA

This section uses the term "ignition", meaning a structural fire, as it is used in the Monograph.

---

[1] Statistical calculations in this report were produced using the commercial software TIBCO Spotfire S+ version 8.1.1 available at http://spotfire.tibco.com/products/s-plus/statistical-analysis-software.aspx. The calculations were verified using the comparable software in R available at http://www.r-project.org/ (open source software).

Monograph Ignitions Data

Figures 1 and 2 contain panel graphs of the Monograph data. Figure 1 shows the panel graph for the number of ignitions (*IGNS*) versus *PGA* for low values of *MMSF* (≤50 *MMSF*) and higher values (> 50 *MMSF*). (Note that one event *MMSF* = 1422.22, *PGA* = 0.15, *IGNS* = 128 is not shown in order to make the graph more readable.) From this figure it can be seen that for smaller areas the number of ignitions varies only slightly with *PGA*. However, for larger areas the increase in *IGNS* as a function of *PGA* is more pronounced, as is the variability of *IGNS*. Figure 2 contains the plots of *IGNS* versus *MMSF* for lower and higher values of *PGA* (≤0.28 and > 0.28). The variability in *IGNS* increases considerably as *MMSF* increases.

Figure 1. Panel graph of IGNS versus PGA for low values of MMSF (≤ 50 MMSF) and higher values (>50 MMSF)

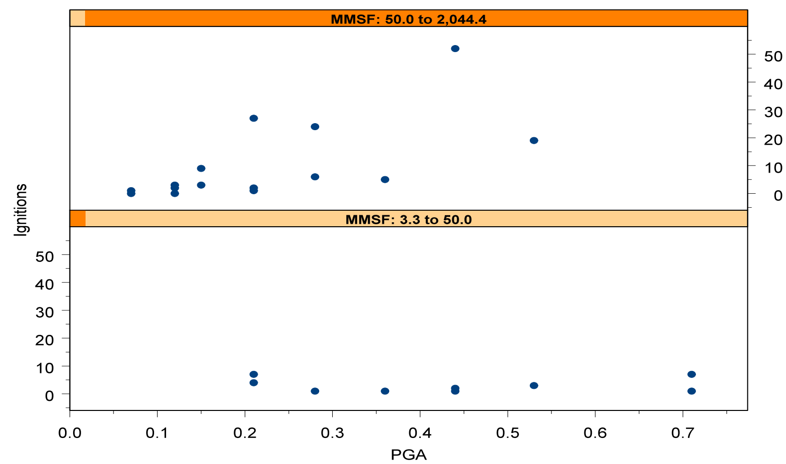
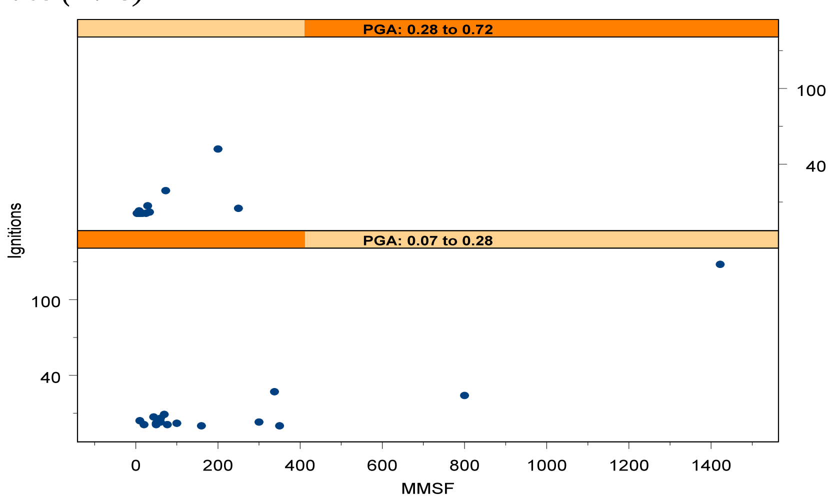


Figure 2. Panel graph of IGNS versus MMSF for low values of PGA ≤ 0.28) and higher values (>.28)



Modeling the Ignitions (IGNS) Data

The analysis approach used in the Monograph is to model the *IGNS*/*MMSF* data using classical polynomial regression. The problem with this approach is that in addition to using a ratio as the dependent variable in the regression, which can introduce error (Kronmal, R. A., 1993), this approach assumes constant variance for all *MMSF* and *PGA*. As a consequence this approach does not model the standard errors associated with the parameter estimates correctly (Hilbe, 2008).[2] For the Monograph data, the standard errors are underestimated for small areas and overestimated for large areas.

Another approach would be to use Poisson regression. This is a common approach for modeling count data (e.g., *IGNS*) as a function of independent variables (e.g., *PGA* and *MMSF*). In the case of the number of ignitions following an earthquake (IGNS), the Poisson model is

$$f(IGNS = y) = (\mu^{y} e^{-\mu}) / y!$$

where $f(IGNS = y)$ is the probability that there are exactly y ignitions, ($y = 0, 1, 2, \ldots$) and μ is the expected number of ignitions for specified values of *PGA* and *MMSF*. The probability that the number of ignitions is less than or equal to y (the cumulative distribution) is given by

$$\mathrm{F}(IGNS \le y) = \sum_{i=0}^{y} \frac{\mu^{i} e^{-\mu}}{i!}.$$

The Poisson regression then estimates μ as a function of *PGA* and *MMSF*. However, the Poisson regression model requires that the variance of *IGNS* for given *PGA* and *MMSF* equals μ. In many applications, including this one, the data are overdispersed, e.g. the variance of *IGNS* is greater than the expected value.[3] Negative binomial regression is typically used when there are signs of overdispersion in the Poisson regression. The negative binomial distribution results when the parameter of the Poisson is modeled as a random variable with a gamma distribution (Hilbe, 2008).[4] This is called a continuous mixture model and the gamma distribution is the mixing distribution (Cameron and Trivedi, 2001).

Negative Binomial Regression Model

---

[2] See also (http://www.jerrydallal.com/LHSP/Poisson.htm).

[3] The dispersion is calculated to be 5.92. This value should be approximately one for Poisson regression. The program call used to determine the dispersion parameter is glm(formula = Ignitions ~ log(PGA) + log(MMSF), family = quasi(link = "log", var = "mu"), data = IGN.data.alldata.splus)

[4] See also (http://www.jerrydallal.com/LHSP/Poisson.htm).

The use of the negative binomial to model the overdispersed IGNS data arises naturally when the Poisson parameter is considered to have earthquake-to-earthquake variability resulting from factors other than PGA and MMSF. This variability could result from site-specific fire vulnerability. In the case of the ignitions data, this site-specific vulnerability variability could be a result of differences in structure construction or usage. The vulnerability variability is captured by assuming the parameter of the Poisson (for a specified PGA and MMSF) is an observation from a random variable M conditioned on μ and *k*, $M(m|\mu,k) \sim \mu V(k, 1/k)$, where V (the vulnerability distribution) is a random variable with a gamma distribution [$g(v \mid k)$] with shape parameter *k* and scale parameter *1/k* (Venables and Ripley, 2002). V has mean equal to one and variance *1/k*. M also has a gamma distribution with mean μ and the variance μ^2/k. An observation from M is denoted *m* ($m = \mu v$). Using this notation, the distribution of the Poisson conditioned on *m* is given by

$$f(IGNS = y \mid m) = \frac{m^y e^{-m}}{y!} = \mathrm{f}(IGNS = y \mid \mu, v) = \frac{(\mu v)^y e^{-\mu v}}{y!}.$$

The marginal distribution of IGNS unconditional on V, but conditional on parameters μ and *k*, is obtained by integrating out V. This yields

$$\mathrm{h}(IGNS \mid \mu, k) = \int_{v>0} f(IGNS \mid \mu, v) g(v \mid k)) dv$$

It can be shown (Hilbe, 2008), that the resulting distribution for IGNS (unconditional on the random variable V) is negative binomial with mean μ and variance $\mu+\mu^2/k$. The distribution has the form

$$h(y \mid \mu, k) = \frac{\Gamma(k+y)}{\Gamma(k)\Gamma(y+1)} \left(\frac{k+\mu}{k}\right)^{-k} \left(\frac{\mu}{\mu+k}\right)^{y}.$$

The $\Gamma(\bullet)$ denotes the gamma integral, which specializes to a factorial for an integer argument.

The parameter μ is modeled as a function of PGA and MMSF (the regression equation). In this application, the regression equation is given by

$$\mu = \beta_0 PGA^{\beta_1} \mathrm{MMS}F^{\beta_2}$$

$$\ln(\mu) = \eta = \ln(\beta_0) + \beta_1 \ln(PGA) + \beta_2 \ln(MMSF).$$

The negative binomial regression for the ignitions data in Table 1 results in the following estimates for the β's [5]

$\ln(\hat{\mu}) = \hat{\eta}$ = -0.53183 + 1.08995 ln(*PGA*) + 0.89368 ln(*MMSF*)

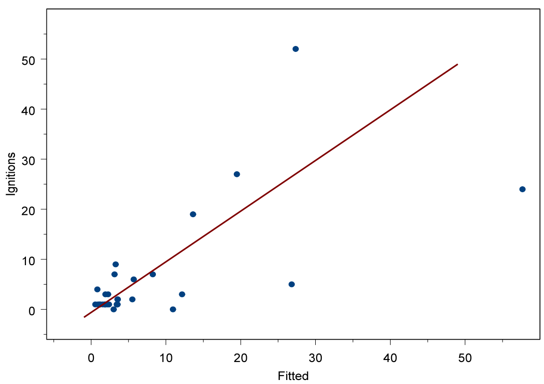
$\hat{\mu}$ = e $^{(-0.53183 + 1.08995 \ln(PGA) + 0.89368 \ln(MMSF))}$

k = 1.635

where $\hat{\mu}$ is the estímate for μ and $\hat{\eta}$ is the estimate for η. The value for *k* is also an estimate, but uncertainty in k is not considered in this paper, so the hat notation is not used.[6]

Figure 3 shows the plots of the observed number of ignitions versus the predicted number of ignitions from the regression fit to the data. This plot shows that the negative binomial regression for the overdispersed Poisson data provides a reasonable approach for modeling these data (particularly for small numbers of ignitions). Appendix A contains the diagnostic plots for the model fits. These plots also indicate that the negative binomial regression model is a reasonable approach for modeling these data, although in the region of high numbers of ignitions there is considerable variability (uncertainty).

Figure 3. Ignitions (IGNS) versus the negative binomial regression fitted values for μ (the expected number of ignitions) .



One observation is not shown in order to see more detail in the lower region: PGA (g) = 0.15, MMSF = 1422.22, Ignitions =128. The estimate of the expected number of ignitions in this case is 49.

[5] The routine used for implementing the negative binomial is glm.nb (Venables and Ripley, 2002, page 206) and is available in the MASS Library associated with S+. The expression used for the negative binomial fit is: glm.nb(formula = Ignitions ~ log(PGA) + log(MMSF), data = IGN.alldata.Splus) (S+ defines log as natural logarithm). This routine is also available in R.

[6] The glm.nb software uses a sequential iteratively reweighted least squares (IRLS) method to determine the regression parameters, β's, and *k*. First the IRLS is used for estimation of the β's, given a value for *k* determined from the method of moments, using these β's, the IRLS is used to determine *k* with the β's fixed. The algorithm then alternates between estimates of the β's and *k* until convergence (Venables and Ripley, 2002), (Hilbe, 2008).

Confidence and Prediction Limits

Confidence Limits. It is well known that $\hat{\eta}$ is asymptotically normal with variance derived from the inverse of the Fisher information matrix determined by maximum likelihood methods (Hilbe, 2008; Wood, 2005). Thus, an approximate upper 95% confidence interval (UCL) is given by

$$\hat{\eta} + 1.65\sqrt{Var(\hat{\eta})}\,.$$

Taking the exponential gives the approximate upper 95% confidence interval for $\mu = e^{\eta}$ (Zhou and Gao, 1997)

$$\text{UCL}(95\%, \mu) = e^{\left(\hat{\eta} + 1.65\sqrt{Var(\hat{\eta})}\right)}$$

The values for Var($\hat{\eta}$) are available from the software used for the negative binomial regression.[7] The general formula for Var($\hat{\eta}$) for the ignitions data is

$$Var(\hat{\eta}) = Var(\hat{\alpha}) + Var(\hat{\beta}_1)\ln(PGA)^2 + Var(\hat{\beta}_2)\ln(MMSF)^2 + 2Cov(\hat{\alpha},\hat{\beta}_1)\ln(PGA) + 2Cov(\hat{\alpha},\hat{\beta}_2)\ln(MMSF) + 2Cov(\hat{\beta}_1,\hat{\beta}_2)\ln(PGA)*\ln(MMSF)$$

where $\hat{\alpha}$ is the estimate of $\ln(\beta_0)$.

The variances and covariances in this formula can be derived from the standard errors of the β's and the correlation matrix reported in the negative binomial (glm.nb) output. Using these values in the equation above results in

$$Var(\hat{\eta}) = 0.30004 + 0.10844\ln(PGA)^2 + 0.01697\ln(MMSF)^2 + 0.11987\ln(PGA) - 0.08848\ln(MMSF) + 0.04111\ln(PGA)*\ln(MMSF).$$

Prediction Limits. The prediction limits of interest in this report are the upper prediction limits (UPL) for a new observation (*m*) from the gamma distribution [M(m|μ,k)] for the Poisson parameter for PGA and MMSF resulting from a new earthquake. The UPL is the limit such that there is a specified probability (e.g., 95%) that the Poisson parameter will be less than or equal to this value for a new earthquake. However, μ (the mean of the gamma distribution) is unknown and to determine prediction limits $\hat{\mu}$, the estimate for μ, must be used in the gamma distribution. Since $\hat{\mu}$ is a random variable with uncertainty, the prediction limits for *m* should incorporate this uncertainty. This means that the unconditional distribution of *m*, M(*m*|*k*) must be determined.

The approach taken in this paper to determine prediction limits for the unconditional *m*, is inspired by Wood (2005), who assumes a normal distribution for $\hat{\mu}$ and determines

[7] The Var($\hat{\eta}$) can be determined from using the *predict* function on the output of *glm.nb* in both S+ and R.

means and variances for the resulting unconditional distribution for *m*, which is the marginal of the continuous mixture of a gamma and a normal, i.e. Gamma($k, \frac{\hat{\mu}}{k}$) and mixing distribution $\hat{\mu}$ ~N(μ_o,Var[$\hat{\mu}$]). However, for the ignitions data in Table 1, prediction limits for *m* based on Wood's approach are too small (i.e., smaller than the confidence limits). The assumption of normality for $\hat{\mu}$ does not work for these data.

The approach taken in this paper is to note that $\hat{\eta}$ is asymptotically normal, thus $\hat{\mu}$ is asymptotically lognormal. Therefore, the mean and variance of $\hat{\mu}$ are given by

$$E(\hat{\mu}) = \theta = e^{\eta + \frac{1}{2}Var(\hat{\eta})} \text{ and } Var(\hat{\mu}) = \sigma_{\hat{\mu}}^2 = (e^{Var(\hat{\eta})} - 1)e^{2\eta + Var(\hat{\eta})}.$$

Using this approach, the continuous mixture consists of a gamma with a lognormal mixing distribution. Integrating over $\hat{\mu}$ gives the unconditional marginal for *m*

$$f_m = \int_{\hat{\mu}>0} \frac{e^{-\frac{(\ln(\hat{\mu})-\eta)^2}{2Var(\hat{\eta})}}}{\hat{\mu}\sqrt{2\pi Var(\hat{\eta})}} \frac{m^{k-1}e^{-\frac{km}{\hat{\mu}}}}{\left(\hat{\mu}/k\right)^k \Gamma(k)} d\hat{\mu} \quad .$$

The 95% UPL for *m* can be determined by solving for m_0, such that $\int_0^{m_0} f_m = 95\%$ with $\hat{\eta}$ substituted for η. However, this approach requires numerical integration. Prediction limits that require only spreadsheet calculations can be determined using the distribution of H = ln(M(*m*|*k*)). The mean and variance of H are given by (see Appendix B for the derivations).

E(H) = ln(θ)-(1/2)Var(H)   and

$$\text{Var(H)} = \ln\left(\frac{k+1}{k}\left[\frac{\theta^2 + \sigma_{\hat{\mu}}^2}{\theta^2}\right]\right)$$

Under the assumption that H is normal, a reasonable assumption (see Appendix C), the following approximate prediction limit estimates can be used

$$\text{UPL(95\%, m)} = e^{E(H)+1.645\sqrt{Var(H)}}.$$

The value for $\hat{\eta}$ is substituted for η in the equations for θ and $\sigma_{\hat{\mu}}^2$. The value of *k* produced by the software (1.635 for the Monograph data) is used in these equations. Uncertainty in *k* is not included in these approximate UPLs. Comparison of these spreadsheet UPLs to those determined from numerical integration show that they are very close (less than 3.6% differences for the 70 cases evaluated in this study). See Figure 6 for an example.

Wood (2005) determines approximate UPLs (conditioned only on *k*) based on the Chebyshev inequality. However for the ignitions data, these UPLs are extremely conservative. For larger values of $\hat{\eta}$, they are so conservative that they are not useful. Another spreadsheet approach is to use the UPL for *m* in the Poisson to determine the probability that $IGNS|k = y$. A more exact method, but one that requires numerical integration is to note that the unconditional distribution of IGNS|*k* is a mixture of a negative binomial and a lognormal. The density f_y (where $y = IGNS|k$) is shown below

$$f_y = \int_{\hat{\mu}>0} \frac{e^{-\frac{(\ln(\hat{\mu})-\eta)^2}{2Var(\hat{\eta})}}}{\hat{\mu}\sqrt{2\pi Var(\hat{\eta})}} \left(\frac{\hat{\mu}}{\hat{\mu}+k}\right)^y \left(1+\frac{\hat{\mu}}{k}\right)^{-k} \frac{\Gamma(y+k)}{\Gamma(y+1)\Gamma(k)} d\hat{\mu}$$

Numerical integration can be used to find the probabilities of various values of IGNS|k=*y*.

ADJUSTMENTS TO THE MODEL

Adjusting the Model to Include Potential Fire Incidents Not Reported in the Monograph

The Monograph data do not include fire incidents that did not involve a fire department response. For conservative estimates of ignitions, fire incidents without fire department response need to be considered.

It is assumed that fires attracting no attention from either the Fire Department or the populace are inconsequential and therefore irrelevant for ignition after earthquake scenarios. This leaves fires without Fire Department response that were attended by someone to be considered.

J.L. Bryan in the report *Smoke as a Determinant of Human Behavior in Fire Situations* (Bryan, 1977) describes the results of a study on the first three actions of the U.S. population during fire incidents, as reported in the Society of Fire Protection Engineers Handbook 4th Ed. Tbl. 3-11.14. Bryan reports that the percentage of the population studied that either “fought fire” or “tried to extinguish” as one of their first three actions is 27%. To adjust for the possibility of fire incidents following an earthquake not captured in the Monograph data (e.g., where the fire was attended by someone other than the fire department), the Monograph data are assumed to represent 73% of the potential fires following an earthquake. This provides conservatism, since some of the 27% in the Bryan report could have also summoned the fire department as one of their first three actions. To adjust these data for the possibility of fires not reported in the Monograph it is assumed that the actual expected number of fires is 1.37 (1/ 0.73) times what is determined from the Monograph data. Thus the model becomes

$f(IGNS^* = y \mid M^* = m_0) = \dfrac{m_0^y e^{-m_0}}{y!}$ and M* ~ 1.37μV(k, 1/k).

The resulting upper confidence and prediction intervals are

$UCL^*(95\%, 1.37\mu) = 1.37\ UCL(95\%, \mu)$

$UPL^*(95\%, m^*) = 1.37\ UPL(95\%, m)$

Adjusting the Model for the Possibility of More Than One Ignition[8] Being Reported as One Fire

There are a number of potential systemic causes within one structure that could lead to multiple independent ignitions after an earthquake. The authors propose dividing these systemic causes into two groups. One group is called "occupant behavior" and is intended to represent such activities as using candles or kerosene lanterns for light. The other group is called "code deficiency" and is intended to represent deficiencies in construction such as poor wiring connections or inadequate shear wall fastening. This source of potential under reporting of ignitions in the Monograph data becomes important when the approach is applied to a large single structure and the goal is to argue for conservatism in the estimation of the expected number of ignitions following an earthquake in the structure.

One approach is to follow the previous example assuming that the parameter of interest is $(1.37+R)\mu$, where R depends on occupant behavior and code deficiency for a large structure in the area impacted by the earthquake.

This results in final upper confidence and prediction intervals, denoted UCL^{**} and UPL^{**}

$UCL^{**}(95\%, [1.37 + R]\mu) = (1.37 + R)*UCL(95\%, \mu)$

$UPL^{**}(95\%, m^{**}) = (1.37 + R)*UPL(95\%, m)$

The authors do not suggest a value for R, but recommend consideration of its importance for the area and facility being modeled. It can be argued that R is inconsequential for a building constructed and operated in a highly regulated environment, where the general public is excluded and the occupants are highly trained in safe operations. On the other hand, in a facility with operations that are predisposed to ignitions, R could be large.

[8] The term "ignition" means an individual ignition regardless of the progression of a resulting fire.

CONSERVATIVE APPLICATIONS OF THE APPROACH

The UPL(95%, m) is an upper limit on the expected (average) number of ignitions for a given PGA and MMSF from a future earthquake at a random site with conditions (i.e., vulnerabilities) that are well-represented by the thirty cases reported in the Monograph data (Table 1). As shown in Figures 1 and 2, there is considerable variability in these data indicating that a wide range of vulnerabilities are represented by the data. The adjusted UPL, UPL**(95%, m), maintains conservatism in the event of possible underreporting of ignitions in the Monograph data. However, if it is not reasonable to think of the conditions for a new earthquake site as a random draw from the Monograph population, then, to maintain conservatism, one must argue that the new earthquake site is less (or at least not more) vulnerable than what would be expected of the population represented by the Monograph data. Given that one of these arguments can be made (e.g., same population as Monograph data adjusted appropriately or bounded by that population), a conservative approach for determining the probability of a specified number of ignitions (y) following an earthquake with given PGA and MMSF is to use

$$m_0 = \mathrm{UPL}^{**}(95\%, m)$$

as the parameter of the Poisson distribution. That is

$$f(IGNS = y) = \frac{m_o^y e^{-m_0}}{y!}.$$

In the case that the new earthquake is not considered a random draw from the same population as the Monograph data, conservatism of the UPL can be maintained if the new site includes buildings with a construction type and occupancy type that is demonstrably less susceptible to earthquake damage and earthquake caused ignitions than the majority of the buildings represented by the Monograph data. The Monograph provides conditions for judging whether a specific building is a candidate for conservative application of this method. For example

- Typical institutional construction/occupancies are good candidates for conservative predictions because the preponderance of the data is from single or multiple family residences. Residences are almost all wood frame construction with plastic sheathed wiring and are more susceptible to ignition from electrical sources than institutional occupancies. The Monograph indicates in Tbl. 4-10 that ignition rates per million square feet for commercial and industrial occupancies are much lower than for residential occupancies.
- Buildings having no gas or liquid fueled appliances are good candidates for conservative predictions because it is cited in the Monograph that 20% to 50% of total post-earthquake fire ignitions are expected from natural gas [from Improving Natural Gas Safety in Earthquakes by the ASCE and the California Seismic Safety Commission (2002, Report No. SSC-02-03)]
- Buildings constructed from concrete with no wood stud partitions

- Buildings with no industrial processes that would greatly increase the probability of ignitions, such as processes involving flammable or combustible liquids, or high temperatures

Example of a Conservative Application to a Large Single Structure

This example assumes an 80,000 ft^2 (*MMSF* = 0.08) cast-in-place concrete building that is classified as Type I (4,4,2) per National Fire Protection Association (NFPA) Standard 220 with a Industrial Occupancy per NFPA Code 101(Life Safety Code). This building can be thought of as consisting of multiple SFEDs (or rooms) that are independent in terms of ignitions. Consistent with the Monograph concept, these rooms are the areas of potential ignitions.

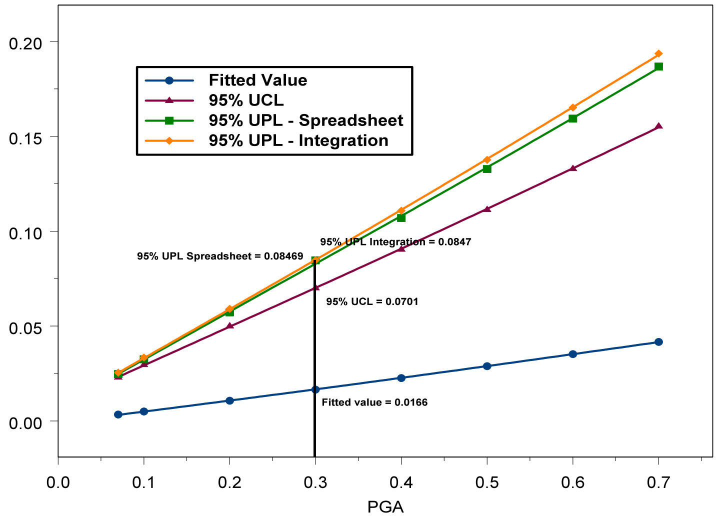
The building has all power wiring in metal conduit and no gas or liquid fuel service. Interior partitions are constructed with metal studs. Activities are light machining with little flammable or combustible liquids present. This is considered a conservative application structure using the criteria discussed above.

The value for *R* is considered to be insignificant (*R* = 0) for this application, because occupant behaviors and code deficiencies are not issues.

Figure 6 shows $\hat{\mu}$ (fitted) as a function of *PGA* for *MMSF* = 0.08. Figure 6 also contains the upper 95% confidence limit for μ (95% UCL). In addition the upper 95% UPL for a new observation, *m*, is given both for the spreadsheet calculation method and for the numerical integration method, which is more accurate. The values for *PGA* = 0.3, are marked with a line on Figure 6.

A conservative spreadsheet estimate of the probability that there are one or more ignitions following an earthquake with $PGA = PGA_0$ and *MMSF* = 0.08 is given by: 1-$F(0, m_0)$, where *F* is the cumulative Poisson distribution with parameter m_0 (the UPL adjusted for possible under reporting of fires in the Monograph data (see Kelly and Tell). For example for *PGA* = 0.3 and *MMSF* = 0.08, UPL(95%, *m*) is 0.08469 and m_0 = UPL*(95%, m) = 1.37*0.08469 = 0.1160, thus, the spreadsheet estimate of probability of one or more ignitions following an earthquake is 1-*F*(0,0.1160) = 0.1095. (Note that the probability based on the more accurate numerical integration technique is 0.0316). The spreadsheet estimate of the probability of two or more ignitions is 0.00623 (numerical integration value = 0.00159, the probability of three or more ignitions is 0.00024 (numerical integration technique = 0.000133). These probabilities are not meant to be exact, but to provide a context for assessing the conservatism of these fire scenarios following an earthquake.

Figure 6. Fitted values for μ ($\hat{\mu}$), the 95% UCLs for μ and the 95% UPLs for m (from the spreadsheet approach and from integration) as functions of PGA for MMSF = 0.08. The black line marks the values for PGA = 0.3



The spreadsheet estimate of the probability that IGNS is less than or equal to one (based on using the UPL for the Poisson parameter) is 99.4%. Using numerical integration produces a more accurate result and this approach shows that even with the 1.37 multiplier for possible under reporting, there is a 99.8% probability that IGNS is less than or equal to one and a 99.99% probability that IGNS is less than equal to two.

A seismic geological evaluation predicts a 0.3 g *PGA* seismic event with a return frequency of once every 2000 years. Under this assumption, the spreadsheet estimate of the frequency of one or more ignitions in a given year from an earthquake event with this PGA is 0.1095 x 1/2000 = 5.5E-5 note that the more accurate numerical integration result is 1.6E-05. The spreadsheet estimate of the frequency of two or more ignitions is 3.1E-6 and the result from numerical integration is 7.9E-07, the spreadsheet estimate of the frequency of three or more ignitions is 1.2E-7 and the result from numerical integration is 6.7E-08.

These estimated frequencies indicate that an earthquake of this magnitude followed by a fire in such a structure would be an infrequent event, and the frequencies of an earthquake of this magnitude followed by multiple fires in such a structure range from very rare to incredible events. Note that these frequency values are not the annual risk of ignition from all potential earthquake events. To determine this frequency, the distribution of ignitions unconditional on PGA must be evaluated.

REFERENCES

ASCE and the California Seismic Safety Commission. 2002. Improving Natural Gas Safety in Earthquakes, *ASCE Report No. SSC-02-03*.

Bryan, J.J., 1977. Smoke as a Determinant of Human Behavior in Fire Situations, U. of Maryland.

Cameron, A. C. and Trivedi, P.K. , 2001. The Essentials of Count Data Regression," in *Companion in Theoretical Econometrics*. B. Baltagi ed. Basil Blackwell.

Hilbe, J.M. 2008. *Negative Binomial Regression*. Cambridge University Press, Cambridge, U.K.

Kronmal, R. A., 1993. Spurious Correlation and the Fallacy of the Ratio Standard Revisited. *Journal of the Royal Statistical Societ*y. *Series A (Statistics in Society)*, Vol. 156, No. 3, pp. 379-392.

Piegorsch, W., 1990. Maximum Likelihood Estimation for the Negative Binomial Dispersion Parameter. *Biometrics* 46,863-867.

Scawthorn, C., Eidinger, J, Schiff, A., 2005. Fire Following Earthquake. *Technical Council on Lifeline Earthquake Engineering Monograph No. 26*. The American Society of Civil Engineers (ASCE) and jointly sponsored by ASCE and The National Fire Protection Association

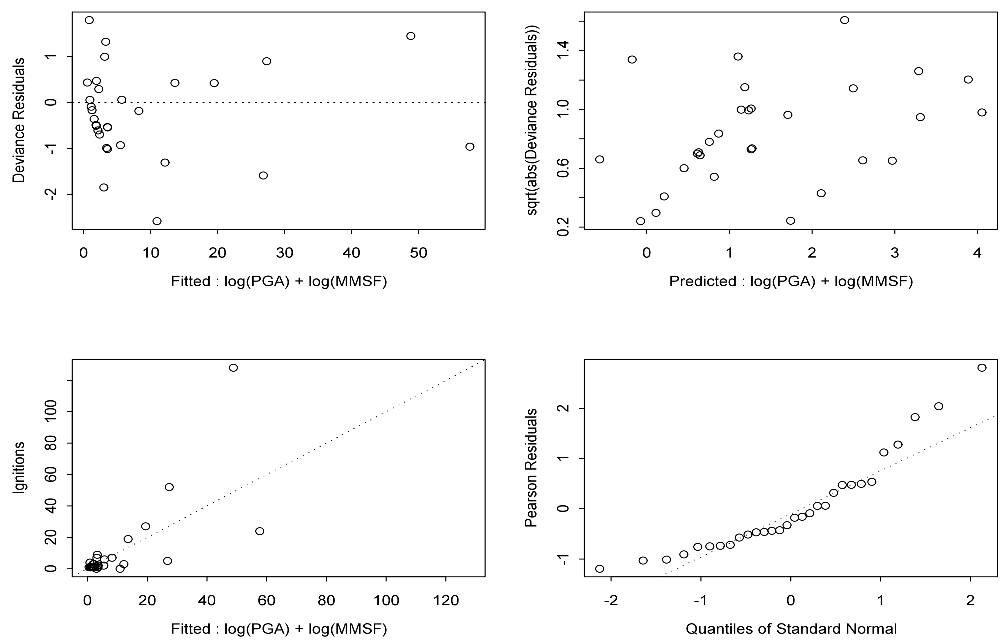
Venables W.N. and B.D. Ripley, 2002. *Modern Applied Statistics with S – 4th ed*. New York.

Wood, G.R., 2005. Confidence and prediction intervals for generalized linear accident models. *Accident Analysis and Prevention* 37, 267-273.

Zhou, X. and Gao S. ,1997. Confidence Intervals For The Log-Normal Mean. *Statistics In Medicine*, *Vol*. 16, 783-790.

APPENDIX A. Diagnostic Plots For The Negative Binomial Regression Fit To The Monograph Data.

If there were a problem with the model fit, the top two plots would show some kind of pattern (e.g., curvature or systematic changes). Neither plot indicates a problem. If the square root of the absolute value of the deviance residuals were greater than two, there could be a potential problem, but this is not the case (upper right graph). The Pearson residual Q-Q plot (lower right graph) is used to identify outliers. It shows skewness, which is not unusual for GLM, but does not indicate an outlier of concern. As seen in Figure 3, the plot of fitted values versus observed values (Ignitions) (lower left graph) indicates that predictions are reasonable for cases with small numbers of ignitions (which are associated with small areas [*MMSF*]). See Guide to Statistics Vol 1, S-Plus 7, Insightful Corporation, Seattle Washington, April 2005, page 412 for more discussion of these plots. More recent S+ documentation also contains discussion of regression diagnostics.



APPENDIX B: PARAMETERS OF THE DISTRIBUTIONS OF M, H AND IGNS

M given $\hat{\mu}$, is distributed gamma with parameters $\hat{\mu}/k$ and k,

M| $\hat{\mu}$ ~ Gamma(k, $\hat{\mu}/k$) , so E(M| $\hat{\mu}$) = $\hat{\mu}$ and Var(M| $\hat{\mu}$) = $\dfrac{\hat{\mu}^2}{k}$.

The quantity $\hat{\eta} = \ln(\hat{\mu})$ and $\hat{\eta} \approx N(\eta, \sigma^2)$, ($\sigma^2 = Var(\hat{\eta})$), therefore, $\hat{\mu} \approx$ lognormal(η, σ^2), and

$$E_{\hat{\mu}}(\hat{\mu}) = e^{\eta + \frac{1}{2}\sigma^2},\ Var_{\hat{\mu}}(\hat{\mu}) = (e^{\sigma^2} - 1)e^{2\eta + \sigma^2}.$$

If we denote $E_{\hat{\mu}}(\hat{\mu})$ as θ and $Var_{\hat{\mu}}(\hat{\mu})$ as $\sigma^2_{\hat{\mu}}$, then the mean and variance of the unconditional distribution of M = M(*m*|*k*) are

$$E(M) = E_{\hat{\mu}}(E(M \mid \hat{\mu})) = E_{\hat{\mu}}(\hat{\mu}) = \theta$$

$$Var(M) = E_{\hat{\mu}}Var(M \mid \hat{\mu}) + Var_{\hat{\mu}}(E(M \mid \hat{\mu})) = E_{\hat{\mu}}(\frac{\hat{\mu}^2}{k}) + Var_{\hat{\mu}}(\hat{\mu}) = \frac{k+1}{k}\sigma^2_{\hat{\mu}} + \frac{1}{k}\theta^2$$ [9]

M appears to be distributed lognormally for the data in this report (see Figure 4 in Appendix C). For M lognormal, the mean and variance of H are given by

E(H) = ln(E(M)) –(1/2) Var(H) = ln(θ)-(1/2)Var(H)

Var(H) = ln (1 +Var(M) /E(M)^2) = $\ln\left(\dfrac{k+1}{k}\left[\dfrac{\theta^2 + \sigma^2_{\hat{\mu}}}{\theta^2}\right]\right)$.

The distribution of $IGNS \mid \hat{\mu}$ has mean $\hat{\mu}$ and variance $\hat{\mu} + \dfrac{\hat{\mu}^2}{k}$. The mean and variance of the unconditional distribution of *IGNS* are given by

$$E(IGNS) = E_{\hat{\mu}}(E(IGNS \mid \hat{\mu})) = E_{\hat{\mu}}(\hat{\mu}) = \theta$$

$$Var(IGNS) = E_{\hat{\mu}}Var(IGNS \mid \hat{\mu}) + Var_{\hat{\mu}}(E(IGNS \mid \hat{\mu})) = E_{\hat{\mu}}(\hat{\mu} + \frac{\hat{\mu}^2}{k}) + Var_{\hat{\mu}}(\hat{\mu}) = \theta + \frac{k+1}{k}\sigma^2_{\hat{\mu}} + \frac{\theta^2}{k}.$$

[9] Note that $\hat{\mu}$ is a biased estimate of μ and E(M) $\neq$ μ, however, it is biased positively, which maintains conservatism for the application of interest in this paper. It does raise the issue of adjusting $\hat{\mu}$ to provide unbiased estimates. This is a topic for further study.

APPENDIX C: DENSITIES OF M|k = M(m|k), H|k = ln(M(m|k) and IGNS|k

The density of M|k is given in the paper. The density of H|k = [ln(M(m|k))] is given by

$$f_h = e^h f_m(e^h)$$

Numerical integration using Mathematica was performed to examine plots of the density of M|k (f_m) and of H|k (f_h) for the 70 cases (see Appendix D for example Mathematica notebooks). In all cases, the shapes of the density of M|k were approximately lognormal (see Figure 4 for an example) and H|k was approximately normal (see Figure 5 for an example), with H|k very slightly negatively skewed as compared to a normal.

Figure 4. The density plot for M|k based on numerical integration for the example in the report (PGA =0.3 and MMSF = 0.08). The density plots for other values of PGA and MMSF have the same shape.

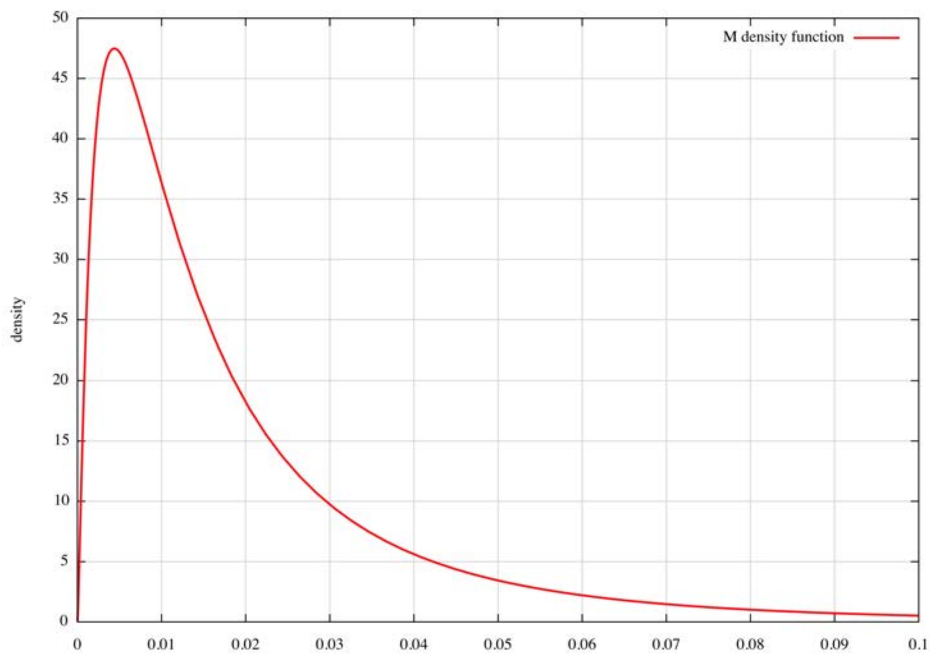
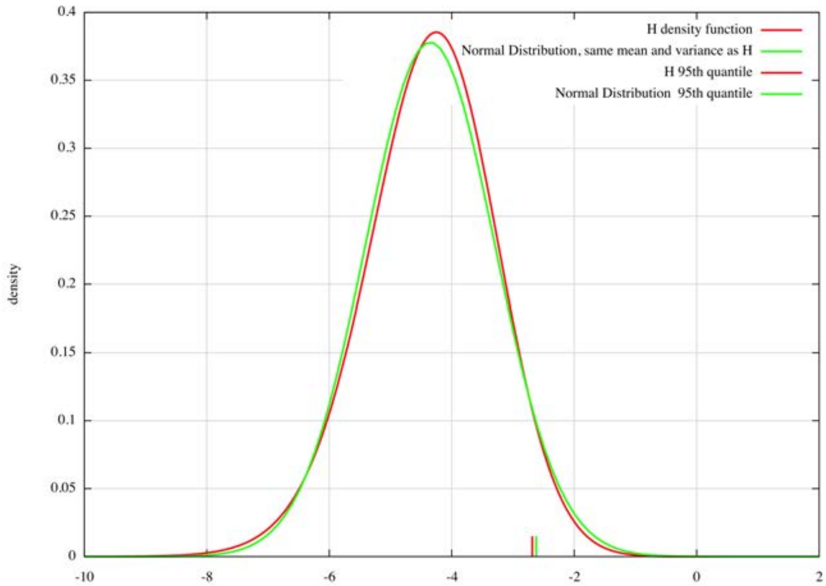


Figure 5. The density plot for H|k based on numerical integration for the example in the report (PGA =0.3 and MMSF = 0.08) is shown in red and a normal distribution with the same mean and variance is shown in green. The density plots for other values of PGA and MMSF have the same shape (although some cases are slightly more negatively skewed).



APPENDIX D. Mathematica Code for Numerical Integration

Below is an example of finding the distribution of M|k and of H|k using Mathematica to do numerical integration. The values of *eta*, *sig* and *k* are from the case study in the paper. This example finds means and variances and the 95th percentile. Note that the notation has changed, for example, $\hat{\mu}$ is denoted *u* and *h* is denoted *n*. In addition, the functions *fm*, *fh*, and *fy* are densities in the text, but it is the integrals of these quantities that are the densities in the Mathematica notation. (To determine results for the adjusted values simply add log(1.37) to eta.)

eta = -4.1013($\hat{\eta}$)
sig = 0.87729 (square root of Var($\hat{\eta}$))
k = 1.635

`fu = 1/(u sig Sqrt[2 Pi]) Exp[-(Log[u] - eta)^2/(2 sig^2)]` (lognormal dist of $\hat{\mu}$ (denoted u))

`gm = m^(k - 1) Exp[-m k/u]/((u/k)^k Gamma[k])` (gamma distribution of m| $\hat{\mu}$)

`fm = fu gm` (joint distribution)

`fh = Exp[n] fm /. m -> Exp[n]` (joint distribution for H (h is denoted n))

`plot1 = Plot[NIntegrate[fm, {u, 0, Infinity}], {m, 0.00001, .5}]` distribution of M (marginal e.g., integrating over u ($\hat{\mu}$))

`plot2 = Plot[NIntegrate[fh, {u, 0, Infinity}], {n, -10, .5}]` distribution of H (integrating over u ($\hat{\mu}$))

`mynorm = NIntegrate[fh, {u, .0000001, 10}, {n, -10, .5}, WorkingPrecision -> 10]` (checking to make sure integrates to one)

`mymean = NIntegrate[fh n, {u, .0000001, 10}, {n, -10, 0.5}, WorkingPrecision -> 10]` (finding the mean of H)

`myvariance = NIntegrate[fh (n - mymean)^2, {u, .0000001, 10}, {n, -10, 0.5}, WorkingPrecision -> 10]` (finding the variance of H)

`mysigma = Sqrt[myvariance]` (SD of H)

```
mytest[x_] := NIntegrate[fm, {u, .0000001, 10}, {m, .0000001, x},
WorkingPrecision -> 10]
FindRoot[mytest[x] == 0.95, {x, .06}]
```
(finding 95^{th} percentile of M)

Below is an example of finding the distribution of *IGNS|k* using Mathematica. The values of *eta*, *sig* and *k* are from the case study in the paper. In this case, addition must be used to find percentiles because of the discrete nature of *IGNS|k*. For larger values of eta, the approach used above for M|k can be applied.

eta = -4.1013($\hat{\eta}$)
sig = 0.87729 (square root of Var($\hat{\eta}$))
k = 1.635

`fu = 1/(u sig Sqrt[2 Pi]) Exp[-(Log[u] - eta)^2/(2 sig^2)]` (lognormal)

`ry = (u/(u + k))^y (1 + u/k)^(-k) Gamma[y + k]/(Gamma[y + 1] Gamma[k])` (negative binomial)

`fy = fu ry` (joint distribution)

`Sum[ NIntegrate[fy /. y -> i, {u, 0, Infinity}, WorkingPrecision -> 10] ,    {i, 0, upperlimit}]` (check to make sure sums to one)

```
NIntegrate[fy /. y -> 0, {u, 0, Infinity}, WorkingPrecision -> 10] +
NIntegrate[fy /. y -> 1, {u, 0, 1}, WorkingPrecision -> 10]
```
(Search for 95th percentile)