Regression Analysis For A Bottom-up Approach To Analyzing Delayed Fission Gamma Yields

M.E. Monville[1,3], E. Padovani[2], S. A. Pozzi[1], J. T. Mihalczo[1]

[1] Oak Ridge National Laboratory, Oak Ridge, TN

[2] University of Aberdeen Business School, King's College

[3] Washington University, St. Louis, MO

[4] *Politecnico, Milano, Italy*

October 2005 Institute of Nuclear Material Management

Objective & Motivation

- Model gamma rays from de-excitation of fission fragments in the [20,958] N time range
- Fission gamma measurements hard to find
- Available data are censored due to instrumentation acquisition limits
  - Censored data point are those whose measured properties are not known precisely, but are known precisely, but are known to lie above or below some limiting sensitivity
- Available data are aggregate
  - Need more data for faithful simulations of fission
  - Need more simulations for reliable predictions at devising fission-based application
  - Need more data to get more insight in the de-excitation processes of fission fragments

Original Measurements for U^{233}

Verbinski et al, 1974

A. Number of resolved and continuum gamma rays per fission

Energy Group (KeV)	Time Group (nsec) 20-45	45-100	100-215	215-458	458-958	**20-958**
140-380	4.50E-002	4.00E-002	2.90E-002	1.85E-002	1.11E-002	**1.43E-001**
380-620	1.63E-002	1.06E-002	5.90E-003	4.10E-003	2.70E-003	**4.00E-002**
620-860	3.10E-003	3.00E-003	2.60E-003	1.93E-003	1.32E-003	**1.19E-002**
860-1100	3.90E-003	2.70E-003	2.10E-003	1.03E-003	6.10E-004	**1.04E-002**
1100-1340	5.40E-003	7.60E-003	1.00E-002	9.90E-003	6.00E-003	**3.89E-002**
140-1340	**7.40E-002**	**6.40E-002**	**4.91E-002**	**3.50E-002**	**2.20E-002**	**2.40E-001**

B. Energy (KeV) per fission for the resolved and continuum gamma rays

140-380	1.08E+001	9.20E+000	6.40E+000	4.40E+000	2.60E+000	**3.34E+001**
380-620	8.50E+000	5.50E+000	3.00E+000	2.00E+000	1.26E+000	**2.04E+001**
620-860	2.90E+000	2.20E+000	2.00E+000	1.39E+000	9.70E-001	**9.90E+000**
860-1100	3.90E+000	2.70E+000	2.10E+000	1.02E+000	5.90E-001	**1.03E+001**
1100-1340	6.50E+000	9.30E+000	1.24E+001	1.22E+001	7.50E+000	**4.80E+001**
140-1340	**3.30E+001**	**2.89E+001**	**2.59E+001**	**2.10E+001**	**1.29E+001**	**1.21E+002**

C. Ratio of energy of the resolved gamma rays to the energy of the Resolved and continuum gamma rays

140-380	8.90E-001	9.10E-001	8.90E-001	9.40E-001	9.40E-001	9.00E-001
380-620	7.30E-001	7.10E-001	5.40E-001	5.30E-001	7.00E-001	6.70E-001
620-860	2.70E-001	4.70E-001	5.70E-001	6.60E-001	5.50E-001	4.70E-001
860-1100	4.70E-001	6.40E-001	5.40E-001	6.70E-001	3.20E-001	5.40E-001
1100-1340	7.50E-001	9.00E-001	9.20E-001	9.50E-001	9.70E-001	9.10E-001
140-1340	7.10E-001	8.10E-001	8.10E-001	8.80E-001	8.90E-001	8.00E-001

Isolate Continuum Energy Contribution

- Total photons spectrum is given by A continuum energy spectrum plus A set of peaks characterist of the isotope
- We devised an algorithm to extract the contribution of the continuum energy spectrum from the total spectrum given in the data
- The continuum energy spectrum is representative of the common physical properties of fissioning nuclides

Continuum Energy Data for U^{235}

A. Number of resolved and continuum gamma rays per fission

Energy Group (KeV)	Time Group (nsec) 20-45	45-100	100-215	215-458	458-958	**20-958**
140-380	4.50E-002	4.00E-002	2.90E-002	1.85E-002	1.11E-002	**1.43E-001**
380-620	1.63E-002	1.06E-002	5.90E-003	4.10E-003	2.70E-003	**4.00E-002**
620-860	3.10E-003	3.00E-003	2.60E-003	1.93E-003	1.32E-003	**1.19E-002**
860-1100	3.90E-003	2.70E-003	2.10E-003	1.03E-003	6.10E-004	**1.04E-002**
1100-1340	5.40E-003	7.60E-003	1.00E-002	9.90E-003	6.00E-003	**3.89E-002**
140-1340	**7.40E-002**	**6.40E-002**	**4.91E-002**	**3.50E-002**	**2.20E-002**	**2.40E-001**

B. Energy (KeV) per fission for the resolved and continuum gamma rays

140-380	1.08E+001	9.20E+000	6.40E+000	4.40E+000	2.60E+000	**3.34E+001**
380-620	8.50E+000	5.50E+000	3.00E+000	2.00E+000	1.26E+000	**2.04E+001**
620-860	2.90E+000	2.20E+000	2.00E+000	1.39E+000	9.70E-001	**9.90E+000**
860-1100	3.90E+000	2.70E+000	2.10E+000	1.02E+000	5.90E-001	**1.03E+001**
1100-1340	6.50E+000	9.30E+000	1.24E+001	1.22E+001	7.50E+000	**4.80E+001**
140-1340	**3.30E+001**	**2.89E+001**	**2.59E+001**	**2.10E+001**	**1.29E+001**	**1.21E+002**

C. Ratio of energy of the resolved gamma rays to the energy of the Resolved and continuum gamma rays

140-380	8.90E-001	9.10E-001	8.90E-001	9.40E-001	9.40E-001	9.00E-001
380-620	7.30E-001	7.10E-001	5.40E-001	5.30E-001	7.00E-001	6.70E-001
620-860	2.70E-001	4.70E-001	5.70E-001	6.60E-001	5.50E-001	4.70E-001
860-1100	4.70E-001	6.40E-001	5.40E-001	6.70E-001	3.20E-001	5.40E-001
1100-1340	7.50E-001	9.00E-001	9.20E-001	9.50E-001	9.70E-001	9.10E-001
140-1340	7.10E-001	8.10E-001	8.10E-001	8.80E-001	8.90E-001	8.00E-001

Finding a Model to Fit the Data

- Several problems trying to fit A model to this data
  - Gamma emission times values below 20 ns and above 958 ns from emission are missing
  - Gamma energy values below 140 kev or above 1240 kev are missing
- Data are aggregated
  - From table cannot tell what time and energy was recorded for each observed gamma
- Data are partially censored
  - Time/energy ranges are what could be recorded not actually what happened
  - Some low/high energy photons and slow/and events are likely to be missing

Linear Regression Model

- Lm(formula = log(frequency) ~ kev + time, data = table)
- Coefficients:
- Estimate std. Error t value pr(>|t|)
- (intercept) -5.6283382 0.1324165 -42.505 < 2e-16 ***
  - Kev -0.0008405 0.0001670 -5.035 4.84e-05 ***
  - Time -0.0014073 0.0002298 -6.124 3.66e-06 ***
- Residual standard error: 0.2833 on 22 degrees of freedom
- Multiple r-squared: 0.7407, adjusted r-squared: 0.7171
- F-statistic: 31.42 on 2 and 22 DF, p-value: 3.561e-07
- Using the lower bound for kev ranges instead of the midpoint.
  - Makes sense in terms of the model because we have:

Interpreting the Regression Model Statistics

- **Not the best adjusted**
- **Best model matching the regression assumptions reasonably**
- **Does not fail A *heteroskedasticity* test**
  - **Replace the dependent variable in the model by the sum of squares of the standardized residuals**
  - **Check the regression sum of A squares against A distribution**
- ***Heteroskedasticity* is A test for model internal consistency**
  - ***It occurs where the variance or spread of results changes over an independent variable***
  - ***The standard errors of the estimates are biased, if we have heteroskedasticity.***
  - ***In the model we chose the average residual (difference between observed and predicted data) does not increase as time or energy increases***

What the Regression Model Suggests

- The log in the frequency term suggests that most of the photons observed in any energy interval will be very close to the bottom of the interval
- The model does not suggest that there is A ‘peak’ time for photon emissions for any given energy
- Tried several quadratic models that might reveal peaks if any, but in all of them one of the time variables was non-significant
- The data suggest that A peak is plausible, but may be insufficient to show where there is one

Assumptions

- The data plausibly fit A reasonably smooth distribution
- When the peaks are removed we can reasonably infer that what is left has A distribution that is unimodal with continuous derivatives
- The distribution function is plausibly asymptotic to zero as energy or time goes to infinity
- Do not expect A specific cutoff energy or time beyond which no gammas may occur
- For large times or energies the probability of observing A gamma becomes vanishingly small
- Plausibly the photons density function should be zero at energy zero
- Plausibly energy and time may be independent variables

Choice of a Density Function

- Can use A discrete distribution:
  - Bernoulli
  - Poisson
  - Binomial
  - Geometric
  - Hypergeometric
  - Negative binaomial

  To predict how many of A total of N photons will appear in each interval
- Can use A continuous distribution:
  - Negative exponential
  - Gamma
  - Beta

  To model the exact distributions of times
- The two approaches "*discrete versus continuous*" can be seen as complementary
- If the exact times are exponential
  - The numbers of photons out of N in A particular interval will be described by A poisson distribution

Choice of a Density Function

- **This limits the choice of functions:**
  - **Normal, beta, uniform distributions are ruled out**
    - **We choose to fit A bivariate gamma distributions f(t,e) =g(t)*h(e) with probability density function:**

$$f(t,E) = \frac{t^{\alpha-1}e^{-\frac{t}{\theta}}}{\Gamma(\alpha)\theta^{\alpha}} \frac{E^{\beta-1}e^{-\frac{E}{\lambda}}}{\Gamma(\beta)\lambda^{\beta}}$$

  - **Where t is time and e is energy**
  - **Need to choose parameters A, Θ, B, Λ**
  - **Use A *goodness-of-fit* test procedure choosing the values of the four parameters that minimize the value of the X^2 test statistic**

X^2 Statistics

- The X^2 test is A goodness of fit test
- Tests how well A set of empirical data fits A given distribution by calculating A single number: X^2
- *Use both the empirical data and the given distribution*
- *Use* X^2 test of distribution not A X^2 test of association

Example of X^2 Test of Association

- **Data for 30 females and 6 males is in the following table**
- **Frequency with which males and females come from small, medium, large cities**

	Small	Medium	Large	Totals		
Female	10	14	6	30		
Male	4	1	1	6		
Totals	14	15	7	36		
	30					
	Female					
	10					
	Male					
	4					

- **Expected frequency for a cell = (column total * row total)/grand total**

The formula for chi-square

$$\chi^2 = \frac{\sum (O-E)^2}{E}$$

O is the observed frequency

E is the expected frequency

The degrees of freedom for the two-dimensional chi-square statistic is:

df = (C - 1)(R - 1)

C is the number of colums of the first variable

R is the number of rows of the second variable.

Algorithm

- **Choose parameters A, Θ, B, Λ**
- **Use cumulative distribution function for the joint bivariate gamms distribution to generate the proportions of photons that occur in each rectangle**

$$\int_{t_1}^{t_2}\int_{\varepsilon_1}^{\varepsilon_2} \frac{t^{\alpha-1}e^{-\frac{t}{\theta}}}{\Gamma(\alpha)\theta^{\alpha}} \frac{E^{\beta-1}e^{-\frac{E}{\lambda}}}{\Gamma(\beta)\lambda^{\beta}} dtdE$$

$$\int_{t_1}^{t_2} \frac{t^{\alpha-1}e^{-\frac{t}{\theta}}}{\Gamma(\alpha)\theta^{\alpha}} dt \int_{\varepsilon_1}^{\varepsilon_2} \frac{E^{\beta-1}e^{-\frac{E}{\lambda}}}{\Gamma(\beta)\lambda^{\beta}} dE$$

- **Calculate the proportion of gammas that will appear in the whole region of obseravation**
- **Calculate the proportion of photons that fall outside the range of observations**

Algorithm

- Calculate the expected number of photons in each rectangle. (Proportion from the distribution)* (the total observed number)* (proportion of photons that appear in range of observations)
- Verify the total observed number is equal to the total expected number
- Calculate the X^2 statistics. That the sum over all available rectangles:

$$\sum 2\frac{(o-e)^2}{e}$$

- Use the nelder-mead simplex method for multidimensional minimization of the χ^2 statistics
- State the null hypothesis and the alternative hypothesis based on the research question

$$H_0 : o = e \qquad H_1 : o \neq e$$

- Set the alpha level: $\alpha = 0.02$
- Calculate the value of the appropriate statistic: $X^2 = 0.033008$

Algorithm

- Indicate the degree of freedom given 25 observed values and 4 parameters df = 24 -4 -1 = 20
- Read out the critical value for the given degrees of freedom and α_value: $\chi^2_{20.02}$ = 3502

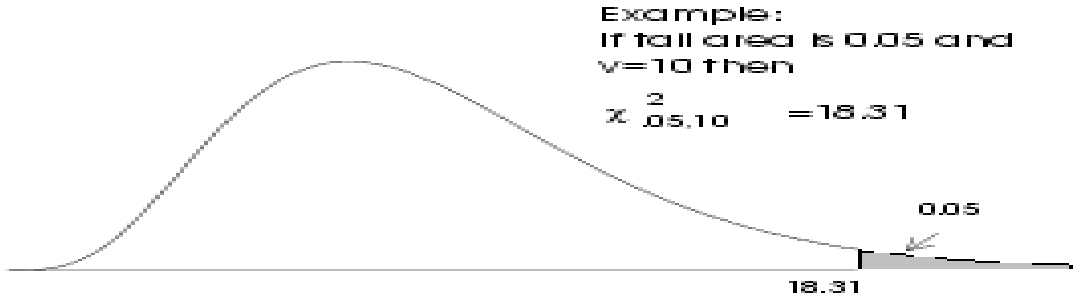
$\mathcal{P}rob\{\chi^2 > \chi^2_{20,.02}\} = 0.02$

- Write the decision rule for rejecting the null hypothesis
- Reject H_0 if $\chi^2 > \chi^2_{20.02}$

$\chi^2 = 0.033008$

$\chi^2_{20,.02} = 35.02$

The chi-square distribution



Values of χ^2 for specified right tail areas

Degrees of freedom v	Right-tail area				
	0.2	0.1	0.05	0.02	0.01
1	1.64	2.71	3.84	5.41	6.63
2	3.22	4.61	5.99	7.82	9.21
3	4.64	6.25	7.81	9.84	11.34
4	5.99	7.78	9.49	11.67	13.28
5	7.29	9.24	11.07	13.39	15.09
6	8.56	10.64	12.59	15.03	16.81
7	9.80	12.02	14.07	16.62	18.48
8	11.03	13.36	15.51	18.17	20.09
9	12.24	14.68	16.92	19.68	21.67
10	13.44	15.99	18.31	21.16	23.21
11	14.63	17.28	19.68	22.62	24.73
12	15.81	18.55	21.03	24.05	26.22
13	16.98	19.81	22.36	25.47	27.69
14	18.15	21.06	23.68	26.87	29.14
15	19.31	22.31	25.00	28.26	30.58
16	20.47	23.54	26.30	29.63	32.00
17	21.61	24.77	27.59	31.00	33.41
18	22.76	25.99	28.87	32.35	34.81
19	23.90	27.20	30.14	33.69	36.19
20	25.04	28.41	31.41	35.02	37.57
21	26.17	29.62	32.67	36.34	38.93
22	27.30	30.81	33.92	37.66	40.29
23	28.43	32.01	35.17	38.97	41.64
24	29.55	33.20	36.42	40.27	42.98
25	30.68	34.38	37.65	41.57	44.31
26	31.79	35.56	38.89	42.86	45.64
27	32.91	36.74	40.11	44.14	46.96
28	34.03	37.92	41.34	45.42	48.28
29	35.14	39.09	42.56	46.69	49.59
30	36.25	40.26	43.77	47.96	50.89

Comments and Results

- The distribution provides values for regions in which we have no data
- The test is simple to implement for bivariate data
- Alternative more powerful tests (anderson—darling, kolmogorov—smirnov) need to sensibly deal with the missing data
- The fitting procedure produces the following values for the parameters
  $\alpha = 0.99468$, $\theta = 142.849$, $\beta = 3.51131$, $\lambda = 160.341$
- Resultant probability density function (rounded to three significant figures):

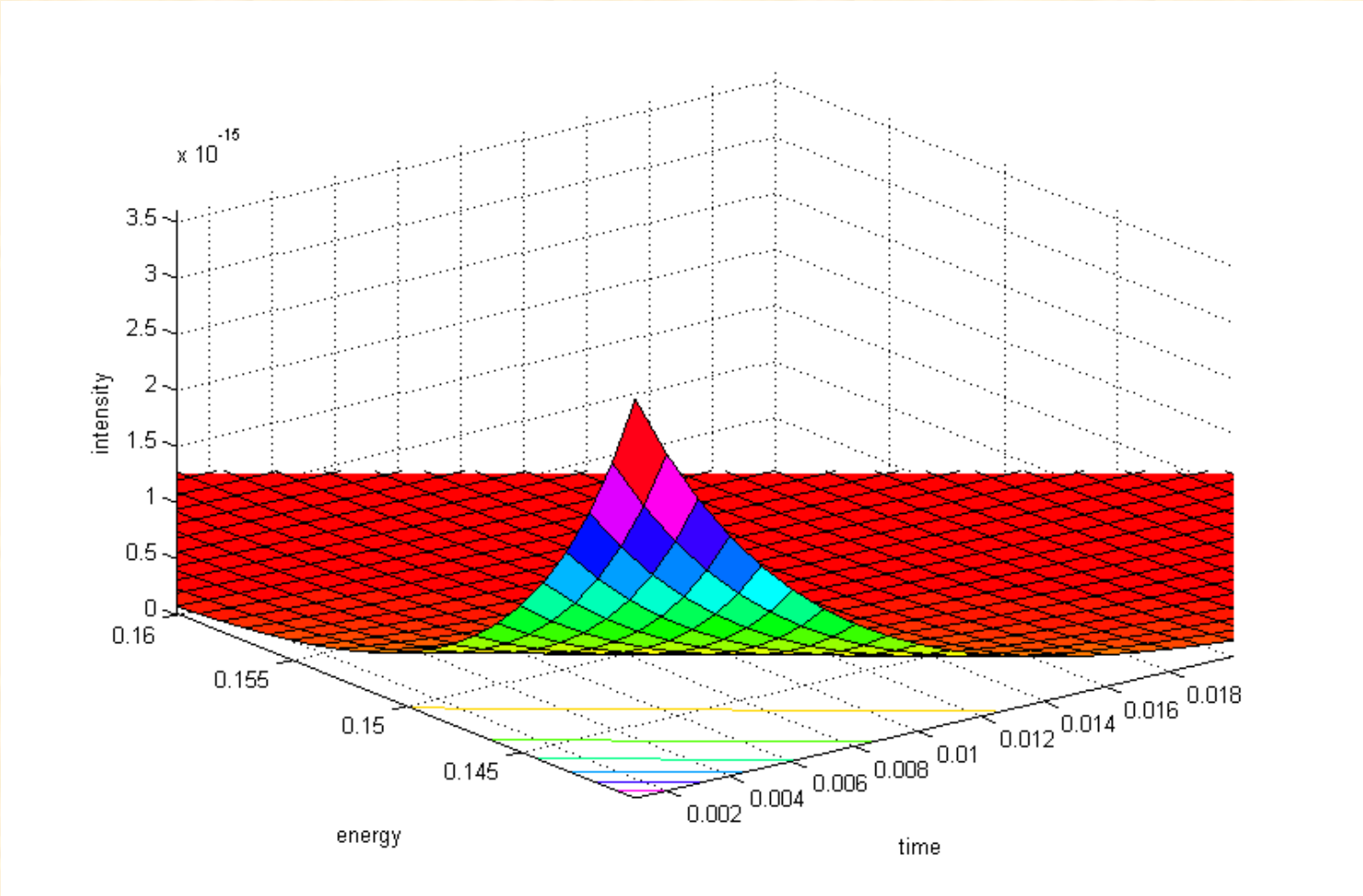
$$f(t, E) = \frac{t^{-0.005} e^{-\frac{t}{143}}}{\Gamma(0.995) 143^{0.995}} \frac{E^{2.51} e^{-\frac{E}{160}}}{\Gamma(3.51) 160^{3.51}}$$

- An intriguing question is whether the true value of α is in fact 1. This would produce:

$$f(t, E) = \frac{e^{-\frac{t}{143}}}{143} \frac{E^{2.51} e^{-\frac{E}{160}}}{1.329 * 160^{3.51}}$$

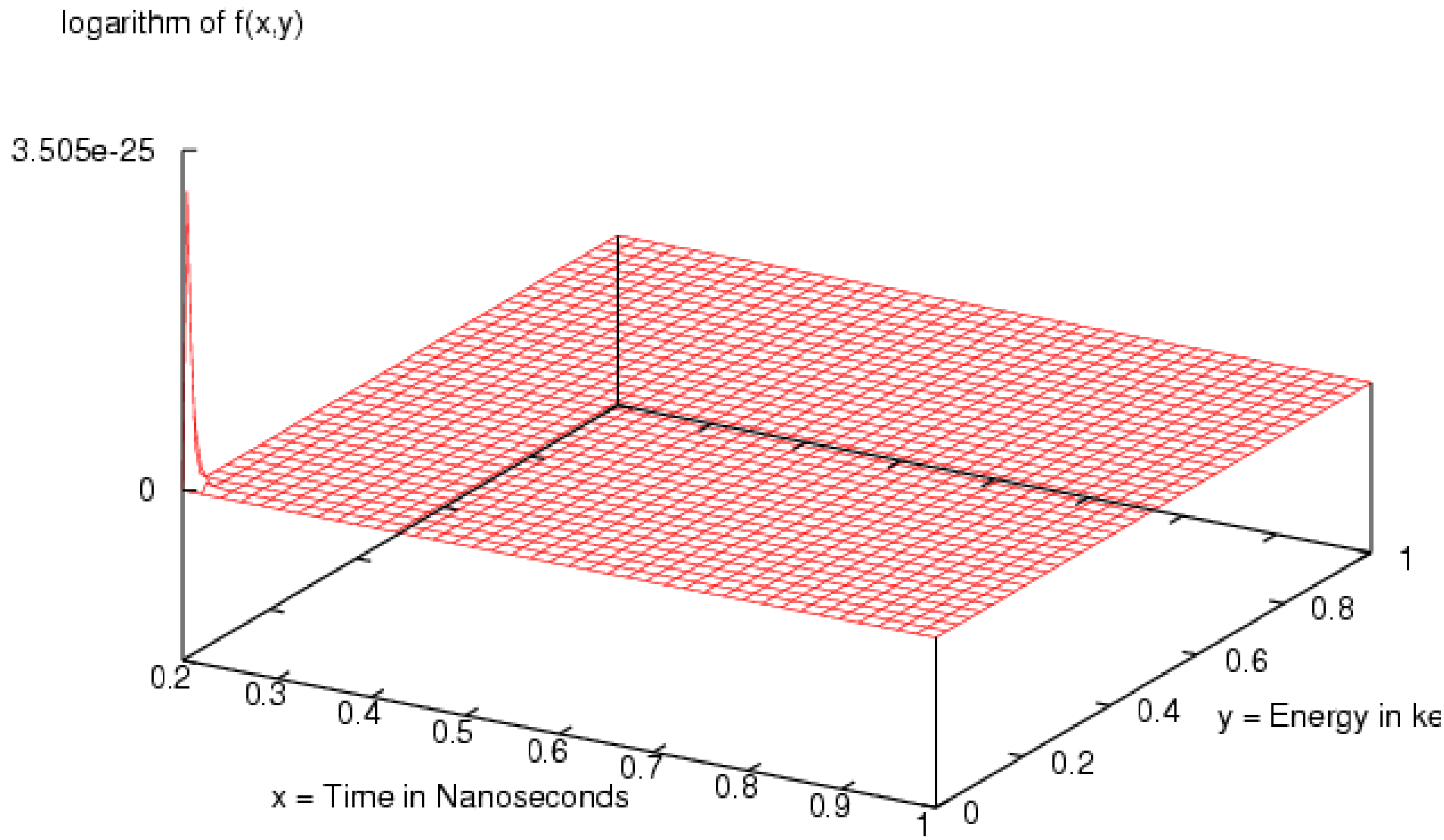
- The marginal distribution giving proportion of gammas as A function of time only exponential

Fitted Functional Form for Delayed Gamma Distribution – U^{235}



Semi-log plot of f(x,y)

y**2*exp(-169.45*x-215*y)/(169.45*2*215) -



Comments and Results

- **Similar Results Were Obtained For** P_u^{239}
- **The Probability Density Function Is Approximately:**

$$g(t,E) = \frac{e^{-\frac{t}{169.45}}}{169.45} \frac{E^{1.97} e^{-\frac{E}{215}}}{2 * 215^{2.97}}$$

α = 0.993787, β = 2.97057, θ = 169.45 , λ = 215

α and β are particularly interesting.

α should be 1.

this value makes the distribution in time *exponential*

β should be 3

This value makes the distribution in energy an Erlang

Comments and Results

- **Would like to express the range of plausible values for the parameters in form of confidence interval**
  - **Would require conditional distributions of parameter values given A test statistic:**

    Prob. **{Parameters lie in region R | calculated χ^2 *value is C}***
  - **We have instead positive value C and choice of parameters:**

    Prob. **{ X^2 <** C **|** parameters α, β, θ, λ }
  - **We don't know if it is even practicable to estimate such A distribution**

Can lower and upper bounds on the parameter values

These are *<u>not confidence interval limits</u>* but parameter values such that any more extreme value of A parameter would yield a χ^2 test statistic greater than the calculated $\chi^2_{20,.02}$

Parameters Bounds

Individual bounds: if any one parameter was at any bound all other parameters would need to be exactly equal to the optimum value.

parameter	*Value*	*Lower bound*	*Upper bound*
α	*0.99468*	*0.0351799*	*4.21268*
θ	*142.849*	*41.8488*	*10289.4*
β	*3.51131*	*0.413607*	*9.13711*
λ	*160.341*	*63.0409*	*893.841*

Conclusion

- We developed A model to describe A functional relationship among the physical quantities characterizing gamma rays emission in fission
- Gamma ray measurements are given in time intervals spanning 20 to 958 ns and in energy intervals spanning 140 to 1340 kev
- This model can be used in place of large data tables in radiation transport monte carlo codes as well as other programs