# Frequentist v Bayesian Inference and the AIHA Bayesian Decision Analysis (BDA)

Yuma Pacific Meeting 23-25 Jan 2013

James C. Rock, PhD, PE,CIH
jcelrock@hotmail.com

## Industrial Hygiene Elevator Discussion

- IH - The profession Dedicated to Making Possible
- the Safe and Healthful Use
- of Necessary Hazardous Materials in
- Necessary Hazardous Processes

# Frequentist (Deduction) v Bayesian (Inference) Data Analysis

- Frequentists & Bayesians use different theorems of Probability Theory
- The following statements are generalizations, and subject to limitations thereof
- Frequentist analysis relies on the Law of Large Numbers
  - As an experiment is performed an increasing number of times,
  - the average outcome approaches the Expected Value
  - In a long run of throwing a 6 sided die, the mean approaches 3.5
- Bayesian analysis relies on Bayes Theorem
  - A single experiment results in data and in IH a small data set
  - Parameters and their uncertainty can be estimated
  - Earlier data can be used to inform interpretation of new data
  - Data can be used to select the best of alternative models



# Bayesian Inference for Typically Small IH data sets

- In IH we believe our data and it is often sparse.
- Bayesian Inference IS the choice for estimating PDF Parameter Values.
- Bayes Rule allows us to combine prior data with new data to determine:
  - does our new data show a change in the workplace?
  - do we need to collect additional data to make a decision?

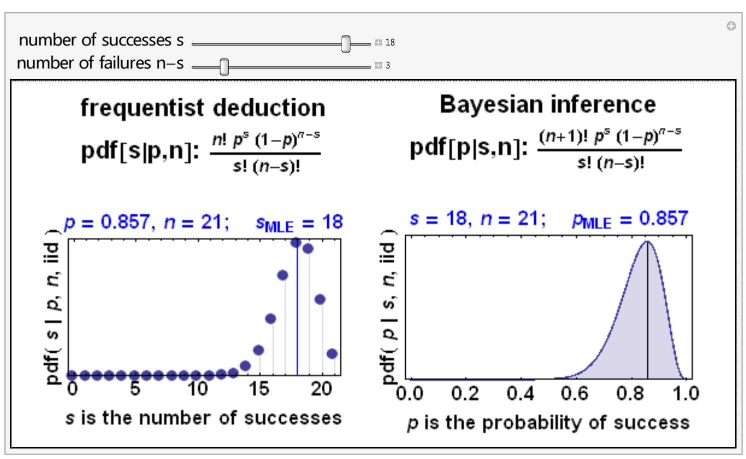
| Analysis | Data Set | Data | Parameters | Central Region | Meaning |
|---|---|---|---|---|---|
| Frequentist | Large | Uncertain | Known | Confidence Interval | p−Value usually Prob [ data do not fit Parameter ] |
| Bayesian | Small<br>Large | Known | Uncertain | Credible Region | CR gives Prob [ unknown parameter value is in CR ] |

# Likelihood v. Probability – based on Fisher (1920s - 60s)

- First, an example of likelihood and probability while we defer definitions of those terms.
- Forward Problem: Use *probability ( 0 ≤ P ≤ 1 ), a function of the outcome, given fixed parameter values.*
  - Given 100 flips of a fair coin, find the *probability that it landed heads-up 51 times.*
  - The discrete Binomial Distribution is the solution to this problem in terms experimental design parameters, {n, p}.
  - It is useful to Casinos and Insurance Actuarial Problems.
- Inverse Problem: Use *likelihood ( 0 < LH ), a function of parameters, given a fixed outcome.*
  - Given 100 flips of a coin which landed heads-up 87 times, find the *likelihood that the coin is fair.*
  - The continuous Beta distribution gives the shape of this likelihood function in terms of outcome data {n, s}.
  - A function of a Maximum Likelihood Estimate (MLE) is its MLE value



# Frequentist v Bayesian Analysis of the Compliance Problem




ONLINE < http : // demonstrations.wolfram.com/FrequentistVersusBayesianPDFForBinaryDecisionsLikeCoinTossin/ >

# Bayesian Introduction: Posterior LH = Prior LH * Data LH

- Bayes rule is one theorem in Probability Theory, equal weight with all others
- It is simple to write, but often requires challenging computations
- Bayes Rule has been deprecated by academics for 25 decades
- It has been used to solve real world problems throughout its lifetime
- In WWII, Allies used it to break multiple versions of the German Enigma Code
- It found Nuclear Weapons sunk off coast of Palomaris Spain after a B-52 collision
- It found a Soviet submarine lost in the central Pacific
- It finds/counts spectral peaks in noisy spectra; for trace analysis and astronomy
- IH Bayesian Decision Analysis (BDA) portrays IH data to non-IH executives



# Desirable Properties of a Probability Theory

- Probability Measures are represented by real numbers.
  - For more detail, see Phil Gregory, Bayesian Logical Data analysis for the Physical Sciences, p30
- Probability Measures must have qualitative agreement with rational intuition.
  - Probability must increase as evidence supporting the truth of a proposition accumulates
  - When the deductive limit is reached, Probability Theory must exhibit formal logic's syllogism
- Probability Measures must be consistent - same info always gives same value.
  - Structural - every possible path to a conclusion must produce the same probability measure
  - Propriety - all available evidence must be used while estimating every probability measure
  - Jaynes Equivalence - Equivalent information must produce the same probability measure for all analysts
  - Example: If **(A && B) | C = B | C, then p[ (A && B) | C ] = p[ B | C ]**
- These three lead uniquely to the axioms of Probability Theory

# Unique Theorems from Those Three Desirable Properties

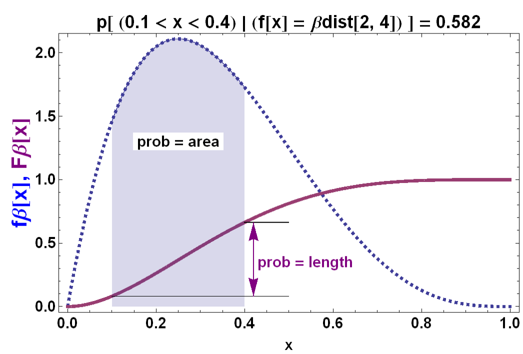
- $0 \le p \le 1$; Impossible Event for p = 0; Certain for p = 1
- In Bayesian Inference, All Probabilities are conditional
- p[ A | B ] = probability A is True GIVEN that B is True
- Sum Rule (with NOT = ~)
- p[ A | C ] + p[~ A | C ] = 1
- Product Rule
- p[ A, B | C ] = p[ A | B , C ] p[ A | C ] = p[ B | A , C ] p[ B | C ]
- Bayes Rule
- p[ A | B , C ] p[ A | C ] = p[ B | A , C ] p[ B | C ]; from the Product Rule
- lh[ A | B , C ] = p[ B | A , C ] p[ B | C ]; 0 < lh
- Marginal Rule, to eliminate the effect of a nuisance parameter *x*

$$p[\, a \mid C \,] = \int_{-\infty}^{\infty} p[\, a, x \mid C \,]\, dx$$



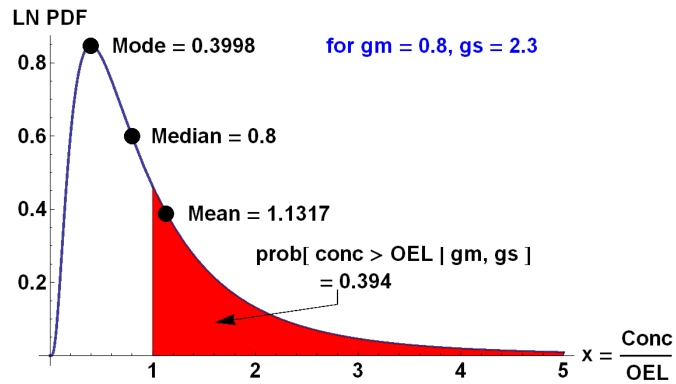
# Beta Distribution Illustrates PDF and CDF

- Probability & Cumulative Distribution Functions (PDF = f[x], CDF = F[x])
- Pro Prob[ a< x< b] = $\int_a^b f[x]\, dx$ = F[b] - F[a]; where F[a] ≡ $\int_{-\infty}^{a} f[x]\, dx$ bility is high or low
  - A PDF has unit area (unit volume for multidimensional PDFs)
- Cumulative Distribution Function (CDF) gives probability x < X



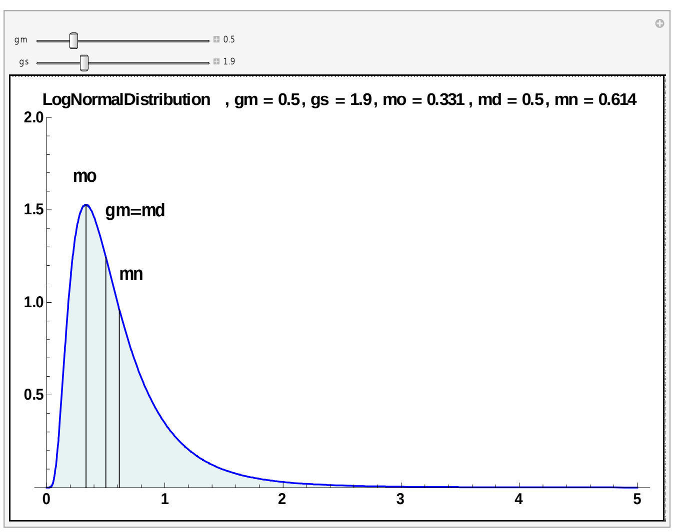
## Review - LogNormal Distribution for Normalized Concentration

- A single parameter, for the normalized exposure, may be misleading
- For example, Maximum Likelihood Estimate (MLE) < Action Level
  - MODE = MLE
- Note: CONC MLE < 0.4 OEL, Mean Conc > 1.13 OEL, P[CONC > OEL] > 0.39






## A Positively Skewed PDF has Mode < Median < Mean

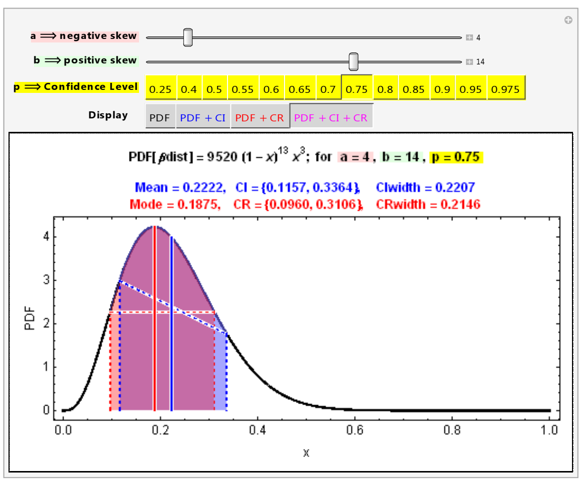


## Confidence Interval (frequentist) v Credible Region (Bayesian)

- Frequentist CI surrounds the mean with equal area in its tails
  - Some probabilities in tails exceed some probabilities in CI
  - Two tails have equal areas, sometimes called a/2, and Confidence Level (CL) is called 1-a.
- Bayesian CR surrounds the mode and may have unequal area in its tails
  - All probabilities in tails are smaller than any probability in the CR
  - Two tails may have unequal areas whose sum is often called a for (1-a) CL.
- The next slide Illustrates CI & Mean v CR & Mode
  - Examine the PDF, its mean and mode by clicking the first button (black)
  - Examine the CI and mean by clicking the second button (blue)
  - Examine the CR and mode by clicking the third button (red)
  - Compare all by clicking the last button (maroon)
  - Adjust a and b to change the amount and direction of skew
  - Adjust the Confidence Level from 0.25 to 0.975 to see its effect on CI, CR, mean & mode
  - Note that for small Confidence Levels, the Mean is outside CR and Mode is outside CI



## Mean in CI and Median In CR



## Parameter Estimation

- Use Prior Information and New Data to Build (Joint) Likelihood Function of Model Parameter(s)
  - Normalize (Joint) Likelihood Function to the Posterior (Joint) PDF, which has unit volume
  - Marginalize the (Joint) Likelihood Function to obtain the posterior PDF for each parameter
  - The Mode of a posterior PDF is the Maximum Likelihood Estimate for that parameter
  - NOTE: Any function of MLE parameter values returns the MLE value for that function

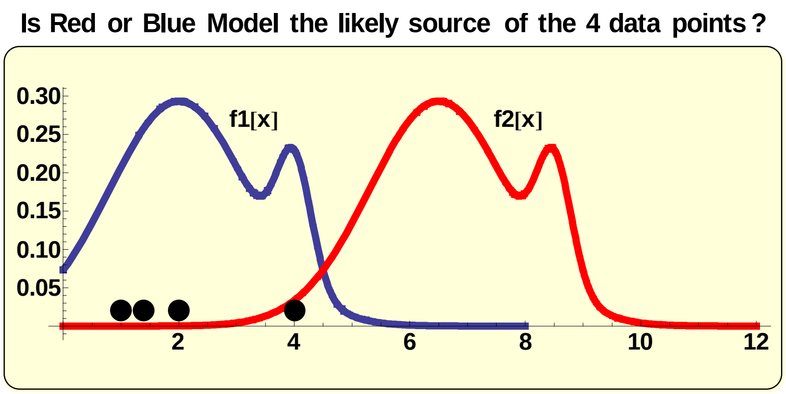


## Model Selection

- Compute the (Joint) Bayesian Likelihood of each model, by multiplying PDF(s) for each data & parameter value
  - The model with the highest likelihood is the best of those tested
  - Bayesian model selection mechanizes Ockham's Razor
  - Bayesian Model Selection favors simple over complex models AND tight over loose fitting models
  - Bayesian model selection weights both simplicity and goodness of fit when choosing "best" model

# Intuitive Likelihood for Model Selection

- Data are represented by small black discs.






## Likelihood (LH), and Likelihood Function (LHF) are both products of probabilities

- For a data set, d = {d1, d2, ... , dn}
- and a LogNormalPDF = f [ gm, gs | x ]
- The product is **LH** when gm & gs are numbers,

  **LH[ d ] = f [ d1 ] * f [ d2 ] * ... * f [ dn ]**

- The product is a **LHF** when gm & gs are variables

  **LHF [ gm, gs | d ] = f[gm, gs | d1] * f[gm, gs | d2] * ... * f[gm, gs | dn]**

## Likelihoods in Practice

- A Likelihood (LH) is the product of the probabilities of independent random variables
- Here are 3 probabilities. Their product equals their likelihood.

  **prob = {0.2, 0.5, 0.3}**

  **lh = 0.03**
- In general, LH tend to be very small numbers so that LH << probability



## A Likelihood Function ( LHF ) with Algebra

- LHF = product of the PDF[ di ], for independent data {d1, d2, ... dn}
- The Peak of LHF gives the locus of the MLE parameters

$$\mathrm{pdf}[x] = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}, \quad \mathrm{data} = \{1.1, 2.9, 3.1, 5\}$$

$$\mathrm{lhf}[\mu, \sigma] = \frac{e^{-\frac{(1.1-\mu)^2}{2\sigma^2}-\frac{(2.9-\mu)^2}{2\sigma^2}-\frac{(3.1-\mu)^2}{2\sigma^2}-\frac{(5-\mu)^2}{2\sigma^2}}}{4\sigma^4\pi^2} = \frac{e^{\frac{-22.115+12.1\mu-2.\mu^2}{\sigma^2}}}{4\sigma^4\pi^2}$$

$$\mathrm{lhf}_{MAX} = 0.0009428 \text{ at } \mu_{MLE} = 3.025 \text{ and } \sigma_{MLE} = 1.381$$

## The LHF and PDF have same shape, but PDF has unit volume

- The LH is the area (volume) of the LHF
- LH > 0, is a positive real number
- The parent PDF has unit area (volume)
- Thus,  PDF =  LHF / LH

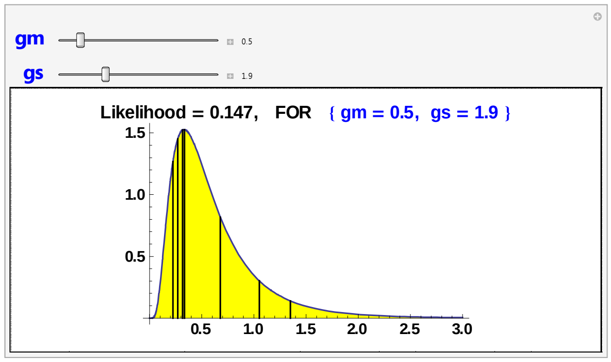


## Parameter Estimation using Data Likelihood

- The next slide Illustrates the Calculations for the illustrated PDF Prior using dat7.
- Data shown as Vertical Lines  whose lengths are the relative probabilities from PDF.
- Model Likelihood is the product of line lengths for independent samples.
- Next two slides illustrate LogNormal and Normal PDFs with interactive graphics.
- Find Maximum Likelihood by moving sliders to adjust { gm and gs } or { m and s }.

# MLE Parameter Estimation: LogNormal PDF

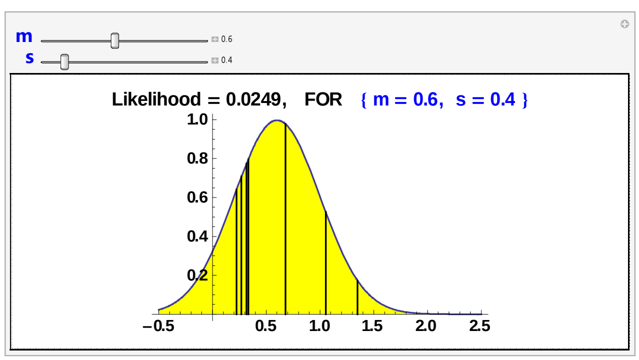
- $gm_{MLE}$ & $gs_{MLE}$ Define the Maximum Likelihood for dat7 when prior = LogNor PDF
- Line heights are proportional to probability for each data value;
- LH = product of line heights
- Adjust the sliders to see the changes in the line heights with various LogNormal Models
- Note that the initial setting shows parameters for the maximum likelihood



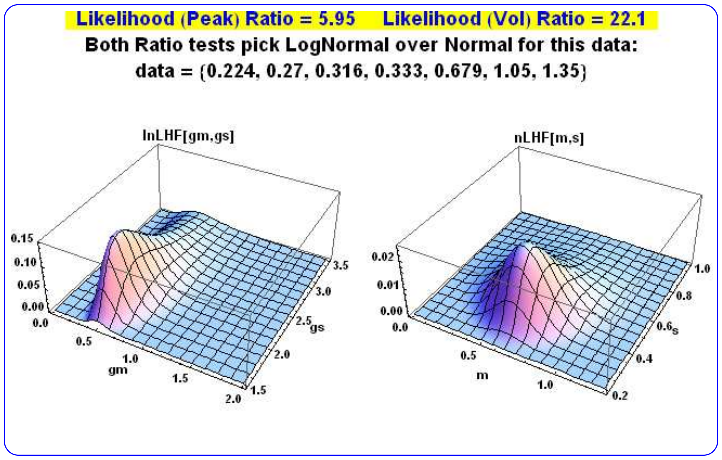



# Parameter Estimation: Normal PDF

- $m_{MLE}$ & $s_{MLE}$ Define the Maximum Likelihood for dat7 when prior = Nor PDF
- Adjust the sliders to see the change in line heights and in their product, the Likelihood
- Note again that the initial setting shows parameters for the maximum likelihood



- Illustration of dat7 LHFs as function of their Parameters
- Moving a parameter away from the peak of the Jt LHF decreases the associated LH



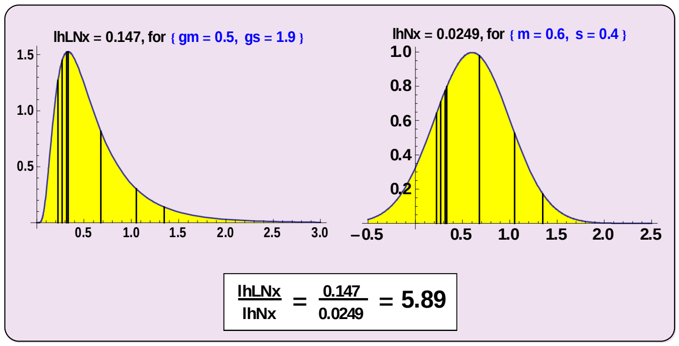



# Parameter Estimation with Model Selection

- Extend to our Earlier Parameter Estimation with dat7 to Model Selection
- Consider dat7, a data set with 7 independent values
- **dat7 = { 0.224, 0.27, 0.316, 0.333, 0.679, 1.05, 1.35 }**
- Use our Computed Likelihoods for both the Normal and a LogNormal Models
- Normal Likelihood Function is a function of two parameters, m and s
- LogNormal Likelihood Function is a function of two parameters, gm and gs
- The model with the larger Maximum Likelihood is a better model for dat7

# Model Selection

- Maximum Likelihood ratio for dat7 favors
  - LogNor over Nor by > 5.8 : 1
- lnLNx = maximum value for LogNormal prior PDF,
- lhNx = maximum value for Normal prior PDF



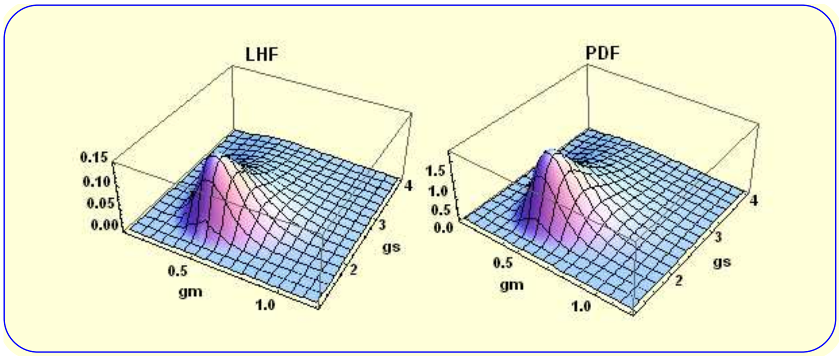



# LHF Ratios replace trial and error parameter estimates

Frequentists' use LHFpeak, Bayesians' use LHFvolume.

| LHF[ central , width \| dt7, *dist* ] | LHF Peak | LH = LHF Volume | Model |
|---|---|---|---|
| $(0.263\ \exp(-(10.2\ \log(gm)+7.\ \log^2(gm)+6.74)/(2\log^2(gs))))/\log^7(gs)$ | 0.149 | 0.0834 | LogNormal |
| $\exp(-(7.\ m^2+3.72-8.45\ m)/(2 s^2))/(\sqrt{2}\ 8\pi^{7/2}\ s^7)$ | 0.025 | 0.00378 | Normal |
| Likelihood (Peak) Ratio = | 5.95 | | $\frac{LNpeak}{Npeak}$ |
| Likelihood (Vol) Ratio = | | 22.1 | $\frac{LNvol}{Nvol}$ |

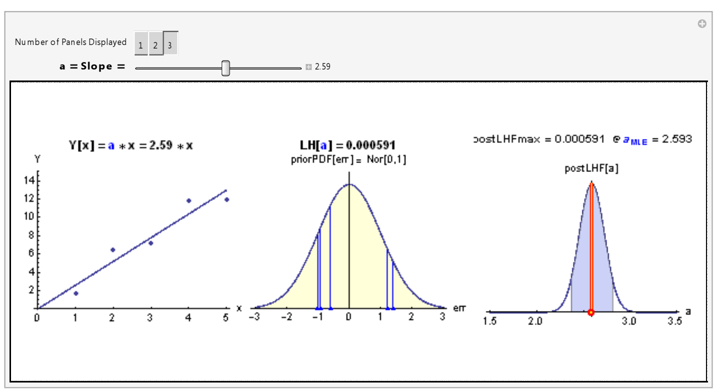
LogNormal Joint PDF = $\frac{\text{Joint LHF}}{\text{Joint LH}}$, and it has unit volume




With dat7, the PDF is ~ 10x larger than the LHF



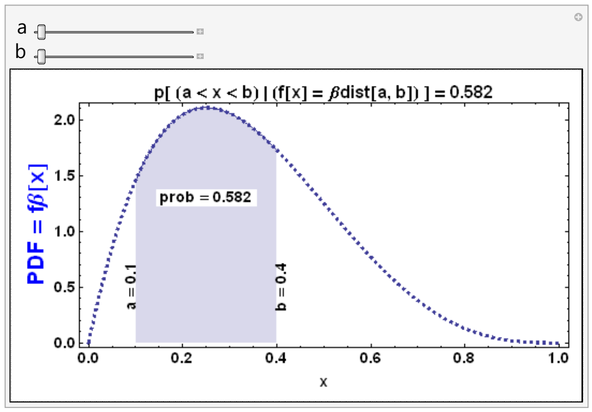
# MLE Parameter Estimation for Calibration Curve Slope




- The buttons display the left panel, the left & middle panels, or all three panels
- Move the slider to change the slope (left), the errors (middle) and LH (right)
- The initial value, a = 2.59, is its maximum likelihood,
  - It represents the best fit to the data.

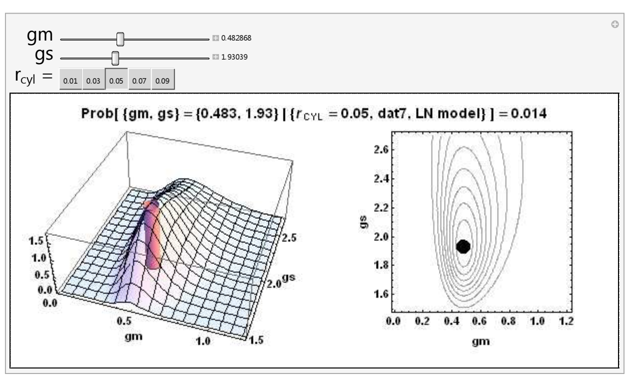
# Probability (p) from PDF

## Limit [ p ] → 0 as a → b



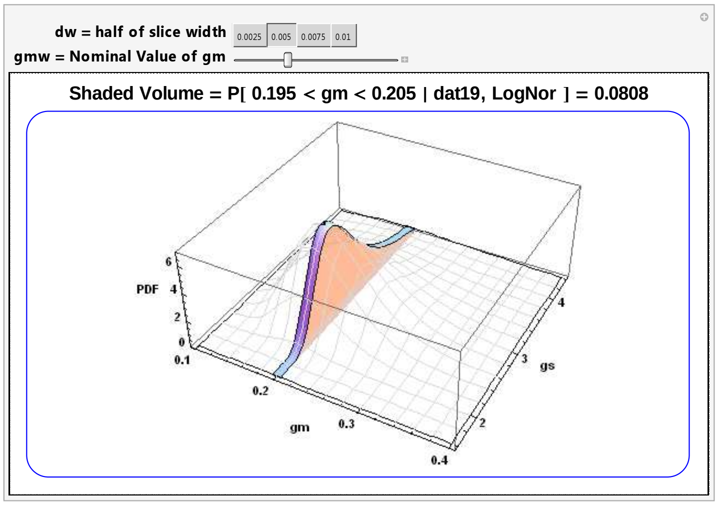



- From dat7 Joint PDF to Probability of specific values for { gm, gs }
- Maximum Likelihood Parameter Estimates are actually quite UNLIKELY
- Cylinder Volume = Probability that {gm,gs} lie in that defined region of dat7
  - As the radius is reduced, the probability the model parameters are inside the cylinder decreases.
  - As the sliders move the cylinder around the {gm, gs} plane, the probability decreases.
  - The initial position indicates the maximum likelihood values for gm and gs.



## Illustration of the MARGINAL INTEGRAL

Probability of gm given dat19, by averaging over gs from the Jt PDF

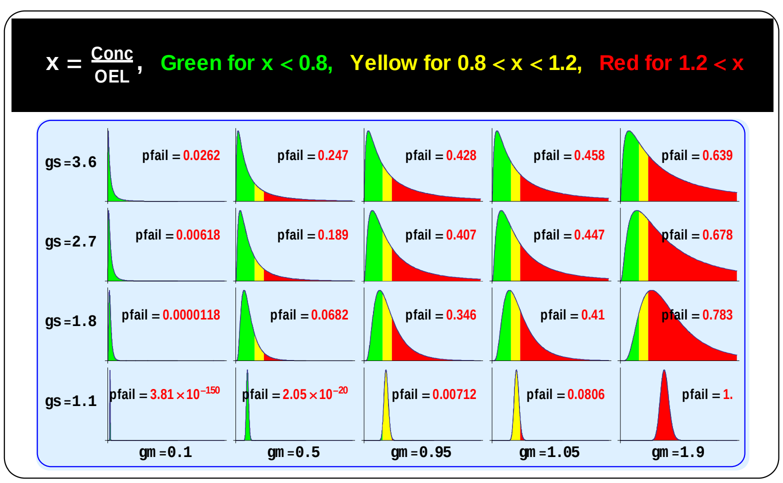





## Lognormal PDF is a parametric function of *gm and gs*

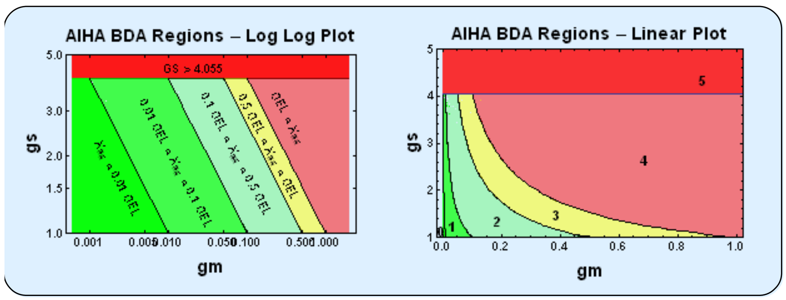
Each pair, {gm, gs}, represents a unique LogNormal PDF.

Probability of non-compliance (or "fail") increases moving away from the origin



# Regions in the {gm, gs} plane, colored for Compliance

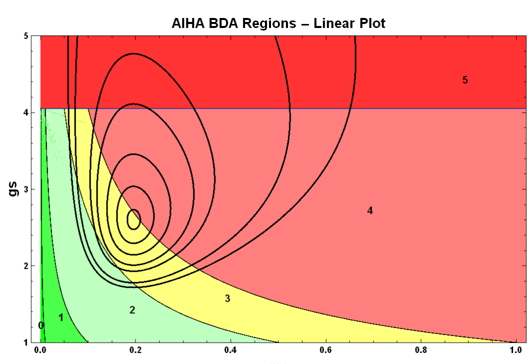
- NIOSH in 1977 suggested 95% confidence as a goal for compliance decisions.
- Let X95 = the 95th percentile of the LogNormalDistribution[ gm, gs ]
- **X95 = gm gs$^{1.64485}$**
- The contours for x95 = {0.01, 0.1, 0.5, 1} and gs > 4.055 are now easily plotted.
- Avoid Environments with gs > 4.055 .OR. x95 > OEL
  - They represent dangerous uncontrolled exposures






# Overlay contour Plot for dat19PDF onto AIHA BDA Regions

- The volume of the Jt PDF over the gm, gs plane in each region is the probability
- Each point, {gm, gs} represents one possible model for the exposures
- Each such discrete model has p = 0
- Probabilities are computed for regions of the {gm, gs} plane



# Decision Regions On PDF

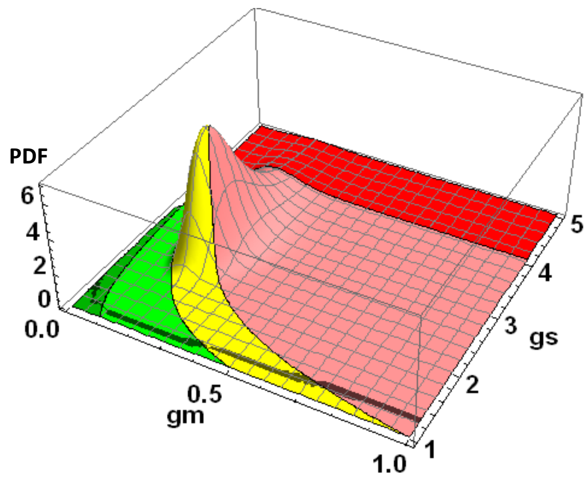
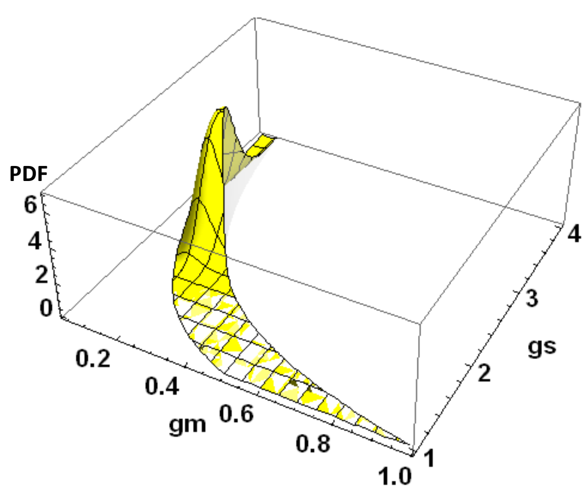


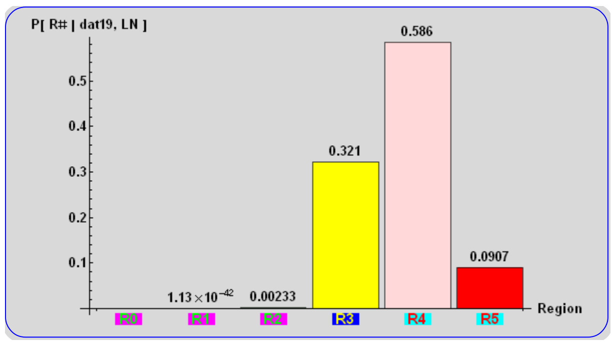



Illustration of Volume (P) over Region 3 using the dat19PDF
This is another type of Probability Integral,
It sums over the portion of the PDF that is of interest.



## BDA Bar Chart for dat19, which has 2 values > OEL;

- $\{gm, gs\}_{PEAK} = \{0.196, 2.60\}$, but $P[x > X_{95} \mid dat19, LN] = 0.676$
- No need for p-value, ANOVA, F-test, etc. etc. ... etc.
- Bar height = probability for Regions 0 thru 5
  - Given dat19 & LogNormal model






## Summary

- Probability Theory is useful for both Deduction (FREQUENTIST) and Inference (BAYESIAN)
- Bayes Rule enables Inference using both prior and new data
- Likelihood (LH) is product of probabilities
- Likelihood function (LHF) is product of probabilities w unknown parameters
- LH = area (volume) of LHF and PDF = LHF/LH
- BDA defines regions in terms of compliance goals
- Probability that data are in a region = Area (volume) of that region of PDF
- BDA is a useful communications tool

# Further Reading

- *Information Processing: Boolean Algebra, Classical Logic, Cellular Automata, and Probability Manipulation.* by David J. Blower
- *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy.* by Sharon Bertsch McGrayne
- *Probability Theory - The Logic of Science.* by Edwin T. Jaynes
- *Data Analysis - A Bayesian tutorial*. by D. S. Sivia with J. Skilling
- Paul Hewett, Perry Logan, John Mulhausen, Gurumurthy Ramachandran, and Sudipto Banerjee. "Rating Exposure Control Using Bayesian Decision Analysis." *Journal of Occupational and Environmental Hygiene, (2006) 3 : 568–581*