A Demonstration of the Lognormal Distribution

Donald L. Smith

Argonne National Laboratory

CSEWG-2003

Properties of Physical Constants

- Generally defined to be inherently positive
- Repeated measurements lead to varying results – both at an individual lab as well as between different labs (viewed collectively)
- Variations are generally presumed to be due to random perturbations
- Variations are expressed in terms of errors (uncertainty) and are governed by probability

The Problem with Gaussians

- Normal (Gaussian) functions are commonly used to describe uncertainties due to random perturbations
- Based on the Central Limit theorem and the assumption of additive (or subtractive) perturbations
- When these perturbations are large (large errors) sampling with a Gaussian function can generate <u>negative values</u>! … Physically unacceptable

Lognormal to the Rescue

- The lognormal function can be used for positive variables since in random sampling it never yields negative values
- Bayesian probability theory recommends use of the lognormal function for a positive variable if its knowledge is restricted to the mean value and standard deviation
- But, can this be justified more intuitively?

Key Formulas for the Lognormal Distribution

- $p(Q) = (2\pi S^2 Q^2)^{-1/2} \exp[-(\mathit{ln}\ Q - M)^2/2S^2] \quad (Q>0)$
- $m = \exp[M + (S^2/2)]$ (mean value)
- $s^2 = m^2[\exp(S^2) - 1]$ (standard deviation)
- $S^2 = \mathit{ln}[1 + (s^2/m^2)]$
- $M = \mathit{ln}\ m - (S^2/2)$

- Given the parameters of the lognormal function, the mean and standard deviation can be calculated
- Given the mean and standard deviation, the parameters of the lognormal function can be determined

A Simplified Model of Measurement

- Suppose Q_0 is the "true" value of parameter Q and that attempts at its determination (measurement) are influenced by a number of multiplicative – rather than additive (or subtractive) – random perturbations centered on unity, but limited to prescribed ranges

- $Q = Q_0 \Pi_{i=1,n} f_i$
- $1-\alpha_i < f_i < 1+ \alpha_i \quad (0 < \alpha_i < 1)$

- What probability distribution emerges from this?

Assumptions About the Individual Perturbations

- We will assume that the f_i are distributed uniformly within their assigned ranges
- Outcome does not depend critically on this assumption (analogous to demonstrating emergence of the Gaussian by assuming additive perturbations and the Central Limit Theorem)

- $m_{0i} = \langle f_i \rangle = 1$ (mean value)
- $\sigma_i = [\alpha_i^2/3]^{1/2}$ (standard deviation)

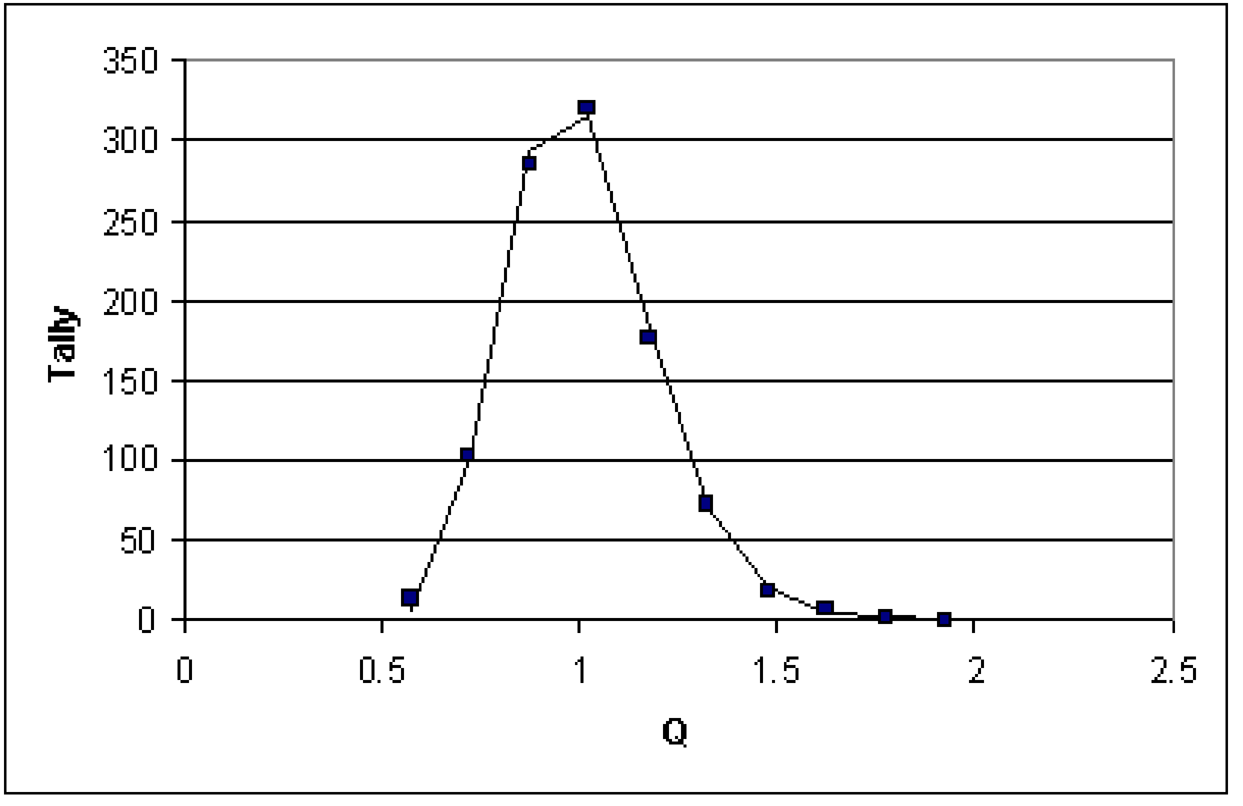
Demonstration by Simulation

- The emergence of the lognormal function has been demonstrated by Monte Carlo simulation using the formulas and assumptions indicated above
- While the indicated model is simple, the same outcome can be shown to occur for even more complicated functional relationships between the variable Q and non-negative perturbations f_i – as long as individual perturbations are not too large

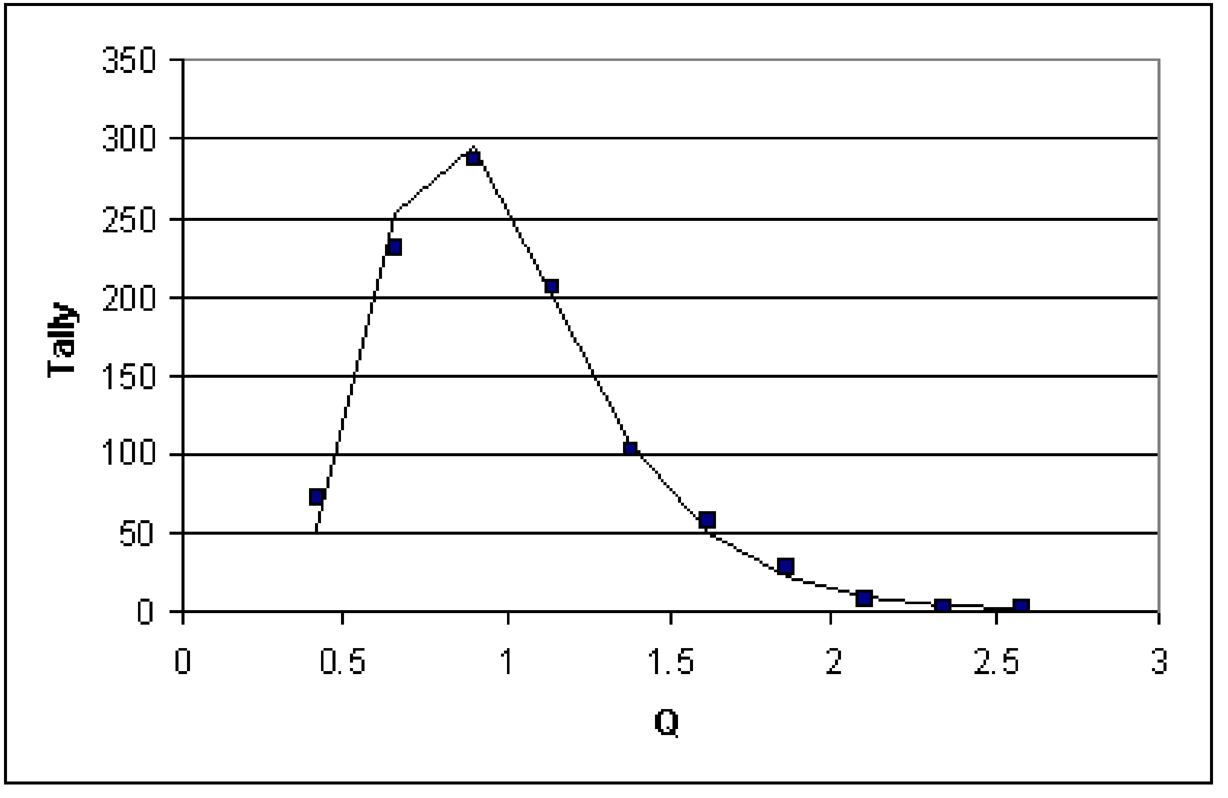
Numerology

- Ten random multiplicative perturbations f_i were treated, with equal $\alpha_i = \alpha = 0.1.\ 0.2, 0.3$, and 0.5
- Q_0 was taken to be unity for convenience
- One thousand values of Q were generated and allocated among ten equally-spaced "bins"
- Computations were performed using EXCEL
- The bin content distributions (histograms) were compared to appropriately normalized lognormal functions whose parameters were calculated from mean values and standard deviations that were deduced directly from the sampling data

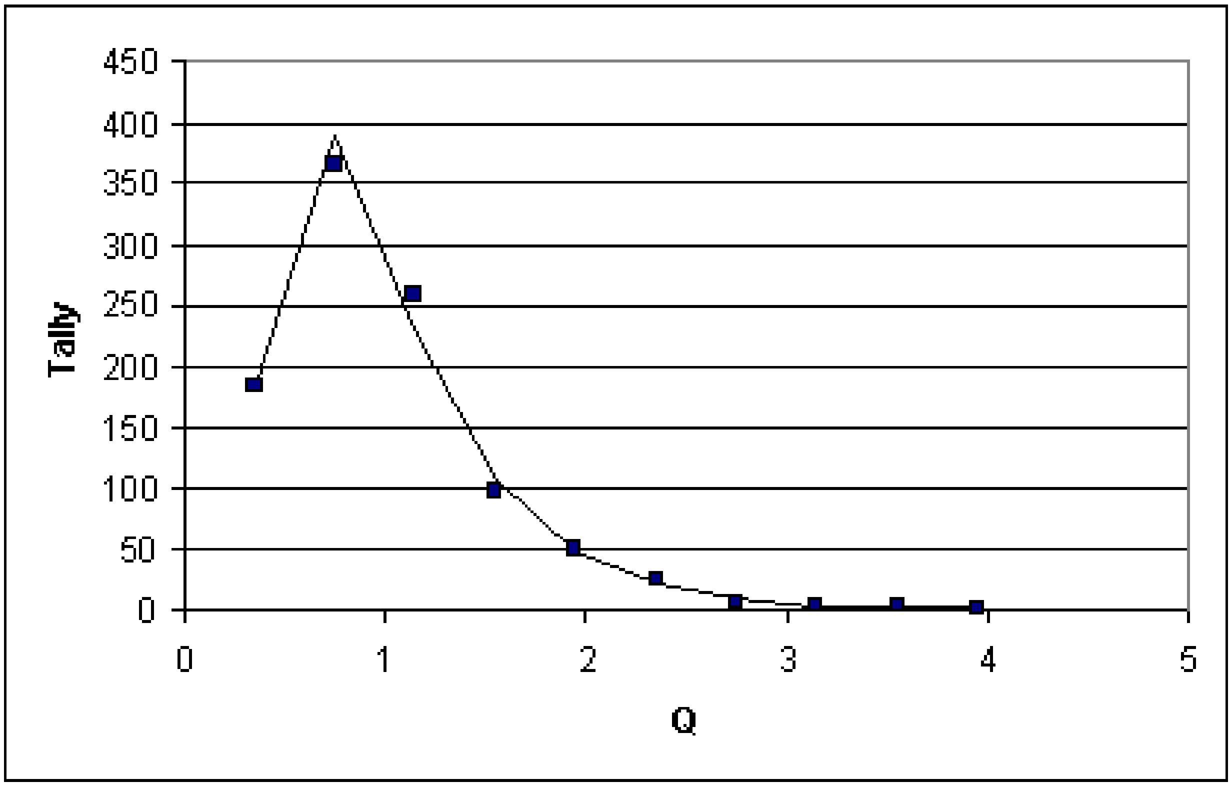
Results for $\alpha = 0.1$



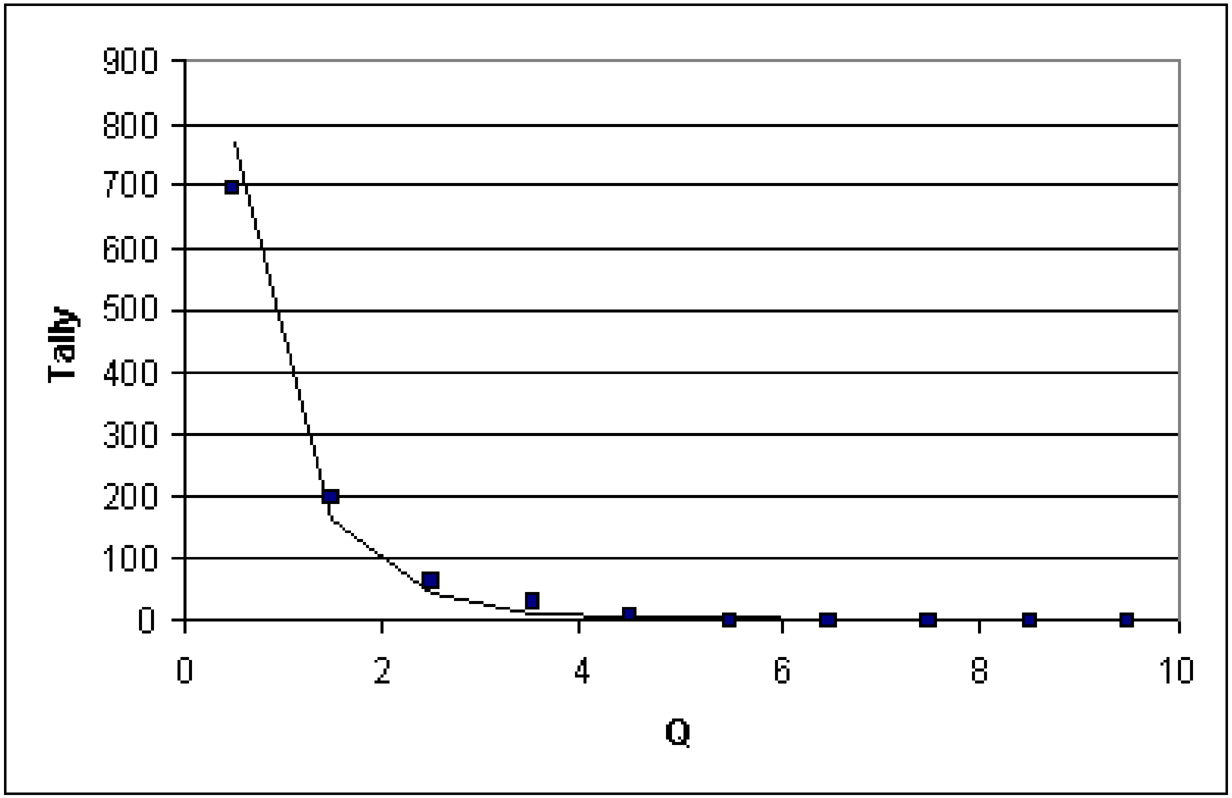
Results for $\alpha = 0.2$



Results for $\alpha = 0.3$



Results for $\alpha = 0.5$



Lognormal Standard Deviations

α	Standard Deviation
0.1	18.6%
0.2	37.0%
0.3	56.3%
0.5	92.1%

Conclusions

- The histograms from Monte Carlo simulation compare very well with the corresponding lognormal functions over the considered range
- The standard deviations for the lognormal functions obey the quadratic summing law
- The present demonstration strongly supports use of the lognormal function to describe inherently positive variables, but it does not provide a mathematically rigorous proof
- See Report ANL/NDM-156 for further details