accepted for publication in *Risk Analysis*

A LogNormal Distribution for Time Spent Showering

David E. Burmaster
Alceon Corporation
PO Box 382669 Harvard Square Station
Cambridge, MA 02238-2669
tel: 617-864-4300; fax: 617-864-9954
deb@Alceon.com

Abstract

In 1987, James and Knuiman published their analysis of a comprehensive domestic water use study conducted in Perth, Western Australia to quantify the components of water usage in approximately 3,000 households. This manuscript corrects errors and omissions about James and Knuiman's study in the US EPA's *Exposure Factors Handbook*, and it shows James and Knuiman's results in a form and notation more readily used in Monte Carlo simulations.

Introduction

In 1987, James and Knuiman (1987) published their analysis of a comprehensive domestic water use study conducted in Perth, Western Australia to quantify the components of water usage in approximately 3,000 households for a 2-week period during the year July 1981 to June 1982. They used Bayesian methods to estimate the usage for each in-house water use activity for each house for each day of participation. Their estimation procedure allowed for inaccuracies in the diary records and water flow measurements, and they incorporated prior information from a sample of approximately 150 houses into the distribution for certain rates and volumes.

James and Knuiman published their results for the time (min/d) spent showering by an individual in one of 2,550 households in terms of this LogNormal distribution (page 710):

$$\hat{\beta}_1 = 1.23\ \hat{\phi}_1^{\ 0.9} \qquad \text{Eqn 1}$$

where $\hat{\beta}_1$ is their fitted marginal LogNormal distribution for time (min/d) spent showering by an individual and $\hat{\phi}_1$ is a fitted marginal LogNormal distribution with arithmetic mean = 7.78 and an arithmetic standard deviation = 3.5 in appropriate units.

Even though James and Knuiman published their results for the time spent showering in terms of a LogNormal distribution for an individual, the US EPA has published only a table of certain percentiles for this distribution in different editions of its *Exposure Factors Handbook*. (*EFH*; US EPA, 1990; 1995; 1996). For example, all editions of the *EFH* (1990; 1995; 1996) state that the "100.00 Cumulative frequency (percentage)" for the time spent by an individual showering equals 20 minutes, an assertion clearly in error because James and Knuiman state "[A] further 50 households [beyond 2,500 households] had durations exceeding 20 minutes" (from the caption to their Figure 5, page 710). This error alone casts serious doubt on all the other percentiles for this distribution in the *EFH* s.

This paper recasts James and Knuiman's original LogNormal distribution in a parametric form more easily and efficiently used in deterministic and probabilistic exposure assessments.

Methods and Results

To present James and Knuiman's results in a form more familiar to exposure assessors, I modify their notation so that T_S (min/d) is a LogNormal random variable expressed in this parameterization (Aitchison & Brown, 1957; Crow & Shimizu, 1988):

$$T_S \sim \exp[\text{ Normal}[\mu, \sigma]\,] \quad ; \quad T_S \text{ in min/d} \qquad \text{Eqn 2}$$

This is equivalent to $\ln[T_S] \sim \text{Normal}[\mu, \sigma]$. Here, exp[•] represents the exponential function, ln[•] represents the Napierian (or natural) logarithm function, and Normal[μ, σ] represents the Normal or Gaussian distribution with mean μ and standard deviation σ (with $\sigma > 0$). James and Knuiman's results may be converted to the notation in Eqn 2 to the final result in Eqn 4 as follows:

$$\begin{aligned} T_S &\sim \hat{\beta}_1 \quad ; \quad T_S \text{ in min/d} \qquad \text{Eqn 3} \\ &\sim 1.23\ \hat{\phi}_1^{\ 0.9} \\ &\sim 1.23 \text{ LogNormal[AMean = 7.78; AStdDev = 3.5]}^{0.9} \\ &\sim 1.23 \exp[\text{ Normal }[1.9594, 0.42931]\,]^{0.9} \end{aligned}$$

$$\sim \quad \exp[\ \text{Normal}[\ 0.9 \bullet 1.9594 + \ln[\ 1.23\ ],\ 0.9 \bullet 0.42931\ ]\ ]\ ]$$

$$T_S \quad \sim \quad \exp[\ \text{Normal}[\ 1.9705,\ 0.3864\ ]\ ] \quad ; \quad T_S \text{ in min/d} \qquad \text{Eqn 4}$$

This derivation uses key identities for LogNormal distributions: AMean = $\exp[\ \mu + 0.5\ \sigma^2\ ]$ and AStdDev = $\exp[\ \mu\ ] \sqrt{\exp[\ \sigma^2\ ]\ (\ \exp[\ \sigma^2\ ] - 1\ )}$ (Evans et al, 1993).

Figures 1 and 2, respectively, show the PDF and the CDF for T_S in Eqn 4 plotted in Mathematica® (Wolfram, 1991; Wickham-Jones, 1994). In these plots, the dots represent the arithmetic mean (AMean) and the 95th percentile:

Statistic	Expression	Value (min/d)	Percentile
Mode	$= \exp[\ \hat{\mu} - \hat{\sigma}^2\ ]$	6.18	~35th
Median	$= \exp[\ \hat{\mu}\ ]$	7.17	50th
AMean	$= \exp[\ \hat{\mu} + \frac{1}{2}\hat{\sigma}^2\ ]$	7.73	~58th
95th percentile	$= \exp[\ \hat{\mu} + z_{0.95}\ \hat{\sigma}\ ]$	13.54	95th

Table 1 shows incorrect cumulative frequencies for T_S as published by US EPA, along with the correct cumulative frequencies from James & Knuiman's results and with columns for the absolute error and the relative error between the incorrect and the correct distributions. The largest absolute error of some 6 percentage points occurs near the Mode of the correct distribution, and the largest relative errors occur in the first few minutes of the correct distribution.

Conclusions and Discussion

Even though people living in Perth, Western Australia may not have exactly the same showering behaviors as people living in the United States, the US Environmental Protection Agency has recommended the results by James and Knuiman in three different editions of the *Exposure Factors Handbook* since 1990. I agree that James and Knuiman's results are suitable and appropriate for use in exposure assessments in the

US until such time as new results become available. Unfortunately, the Agency has misinterpreted the James and Knuiman's study. With James and Knuiman's results re-expressed in Eqn 4 as a parametric distribution, many more exposure assessors can model the variability in individual shower times.

Acknowledgments

Alceon Corporation funded this research.

Trademarks

Mathematica ® is a registered trademark of Wolfram Research, Inc: http://www.wri.com
Alceon ® is a registered trademark of Alceon Corporation: http://www.alceon.com

References

Aitchison & Brown, 1957
Aitchison, J. and J.A.C. Brown, 1957, The Lognormal Distribution, Cambridge University Press, Cambridge, UK

Burmaster & Hull, 1997
Burmaster, D.E. and D.A. Hull, 1997, Using LogNormal Distributions and LogNormal Probability Plots in Probabilistic Risk Assessment, Human and Ecological Risk Assessment, Volume 3, Number 2, pp 235 - 255

Crow & Shimizu, 1988
Crow, E.L. and K. Shimizu, Eds., 1988, Lognormal Distributions, Theory and Applications, Marcel Dekker, New York, NY

Evans et al, 1993
Evans, M., N. Hastings, and B. Peacock, 1993, Statistical Distributions, Second Edition, John Wiley & Sons, New York, NY

James & Knuiman, 1987
James, I.R. and M.W. Knuiman, 1987, An Application of Bayes Methodology to the Analysis of Diary Records From a Water Use Study, Journal of the American Statistical Association, Volume 832, Number 399, pp 705 - 711

US EPA, 1990
US Environmental Protection Agency, 1990, Exposure Factors Handbook, Office of Health and Environmental Assessment, EPA/600/8-89/043, March 1990

US EPA, 1995
US Environmental Protection Agency, 1995, Exposure Factors Handbook, External Review Draft, Exposure Assessment Group, EPA/600/P-95/002A, Washington, DC, June 1995

US EPA, 1996
US Environmental Protection Agency, 1996, Exposure Factors Handbook, Science Advisory Board Review Draft, Exposure Assessment Group, EPA/600/P-96/002B a, b, and c, Washington, DC, August 1996

Wickham-Jones, 1994
Wickham-Jones, T., 1994, Mathematica Graphics, Techniques & Applications, Springer-Verlag, Telos, Santa Clara, CA

Wolfram, 1991
Wolfram, S., 1991, Mathematica®, A System for Doing Mathematics by Computer, Second Edition, Addison-Wesley, Redwood City, CA

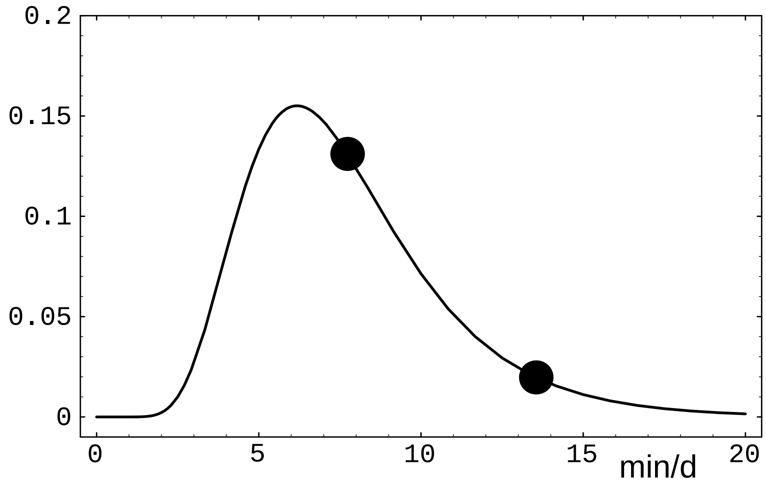

Figure 1
PDF for T_S showing the AMean and the 95th Percentile

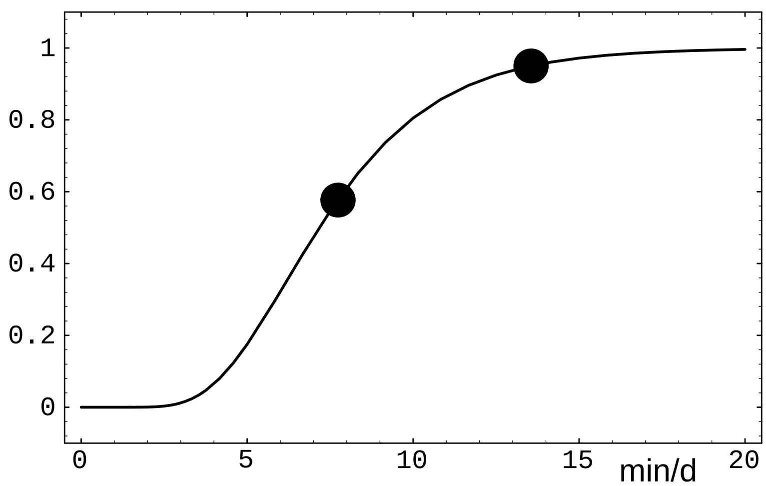

Figure 2
CDF for T_S showing the AMean and the 95th Percentile

deb 4 April 1997
proofed

Table 1
Comparison of the Cumulative Distributions

Shower Duration (min/d) •••••	Incorrect Cumulative Frequency § •••••	Correct Cumulative Frequency §§ •••••	Absolute Error §§§ •••••	Relative Error §§§§ •••••
1	0.002	1.6950E-07	-0.0020	-11798.4100
2	0.008	0.0005	-0.0075	-15.9008
3	0.032	0.0120	-0.0200	-1.6667
4	0.098	0.0653	-0.0327	-0.5008
5	0.226	0.1750	-0.0510	-0.2914
6	0.382	0.3218	-0.0602	-0.1871
7	0.526	0.4747	-0.0513	-0.1081
8	0.638	0.6110	-0.0270	-0.0442
9	0.734	0.7213	-0.0127	-0.0176
10	0.810	0.8050	-0.0050	-0.0062
11	0.862	0.8657	0.0037	0.0043
12	0.902	0.9085	0.0065	0.0072
13	0.924	0.9380	0.0140	0.0149
14	0.942	0.9582	0.0162	0.0169
15	0.956	0.9719	0.0159	0.0164
16	0.968	0.9811	0.0131	0.0134
17	0.976	0.9872	0.0112	0.0113
18	0.986	0.9914	0.0054	0.0054
19	0.994	0.9941	0.0001	0.0001
20	1.000	0.9960	-0.0040	-0.0040
21	1.000	0.9973	-0.0027	-0.0027
22	1.000	0.9981	-0.0019	-0.0019
23	1.000	0.9987	-0.0013	-0.0013
24	1.000	0.9991	-0.0009	-0.0009
25	1.000	0.9993	-0.0007	-0.0007

§ tabulated in US EPA, 1990, 1995 & 1996
§§ computed from James & Knuiman, 1987
§§§ computed as Correct - Incorrect
§§§§ computed as (Correct - Incorrect) / Correct