Min and Max Normal Extreme Interval Values and Statistics

Marsha Jance (mjance@comcast.net)

Nick Thomopoulos (thomop@stuart.iit.edu)

Illinois Institute of Technology Stuart School of Business

Abstract

The method for determining the min and max normal extreme interval values and statistics: expected value, standard deviation, median, mode, and coefficient of variation is discussed. An extreme interval value g_α is defined as a numerical bound, where a specified percentage α of the data is less than or equal to g_α.

Introduction

This paper presents a method for determining the extreme interval values and statistics for the minimum (min) and maximum (max) observations in a sample of size "n". The observations in the sample are normally distributed with a mean of $\mu = 0$ and a variance of $\sigma^2 = 1$. An extreme interval value g_α is defined as a numerical bound, where a specified percentage α of the data is less than or equal to g_α. For example, if the probability is $P(g \le g_{0.75}) = \alpha = 0.75$, then 75% of the data is below or equal to $g_{0.75}$. The extreme interval values are found for probabilities ranging from α = 0.01 to α = 0.99 and observation sizes ranging from n = 1 to n = 1000.

The basis for this paper comes from the doctoral dissertation and research of *Calculating Min and Max Extreme Interval Values for Various Distributions* (Jance). In this dissertation research, Jance discusses the min and max extreme interval values and statistics for normal, exponential, and uniform variables. Jance developed Excel VBA (Visual Basic for Applications) programs to find the min and max extreme interval values and statistics. Jance's work includes tables, graphs, and applications of this research. The tables provide the min and max extreme interval values and statistics for the normal, exponential, and uniform distributions for various observation sizes.

The method and an example of finding the min and max normal extreme interval values and statistics are first presented. The statistics include the expected value, standard deviation, median, mode, and coefficient of variation (the standard deviation divided by the expected value). Then, it is shown how to find the min and max extreme interval values for normal variables with parameters other than $\mu = 0$ and $\sigma^2 = 1$. Next, an analysis of the min and max probability density functions, extreme interval values, and statistics is presented. Finally, an application of this research is provided.

Min and Max Values

Suppose that the minimum and maximum values are selected from a sample with "n" observations. The "n" observations, $x_1,\ldots,x_n$, come from a continuous distribution with probability density function f(x) and cumulative distribution function F(x). If one continued to take samples of size "n" observations (from the same continuous population), the minimum and maximum values will vary from sample to sample. Thus, the min and max values have a probability density function and a corresponding cumulative distribution function.

Let the variable g be the minimum or maximum of the "n" observations. If $g = \min(x_1,\ldots,x_n)$, then the min probability density function is $h(g) = nf(g)(1-F(g))^{(n-1)}$ (Hines, Montgomery, Goldsman, and Borror 215). If the variable $g = \max(x_1,\ldots,x_n)$, then the max probability density function is $h(g) = nf(g)F(g)^{(n-1)}$ (Hines, Montgomery, Goldsman, and Borror 215). In addition, the min and max cumulative distribution function is $H(g) = \int_{-\infty}^{g} h(g)\,dg$.

Now, if the "n" observations are normally distributed with a mean of $\mu = 0$ and a variance of $\sigma^2 = 1$, then the min probability density function is $h(g) = nf(g)(1-\Phi(g))^{(n-1)}$ and the max probability density function is $h(g) = nf(g)\Phi(g)^{(n-1)}$, where $f(g) = \frac{1}{\sqrt{2\pi}}e^{-g^2/2}$.

In the min and max probability density functions, $\Phi(g)$ represents the standard normal cumulative distribution function evaluated at g. Since a closed form solution is not available for $\Phi(g)$, an Excel VBA program, based on C. Hastings Jr.'s approximation, was written to find $\Phi(g)$ (United States Department of Commerce, National Bureau of Standards 932-933).

Min and Max Extreme Interval Values and Statistics

An Excel VBA application was developed to find the extreme interval values and statistics. In this application, the min and max cumulative distribution functions are first evaluated and then interpolation is used to find the extreme interval values.

The min and max cumulative distribution functions: $H(g) = \int_{-\infty}^{g} h(g)\,dg$, expected values: $E(g) = \int_{-\infty}^{\infty} gh(g)\,dg$, and standard deviations: $\sigma = \sqrt{E(g^2) - E(g)^2}$ do not have a closed form solution available. Thus, the Riemann Sums right-end-points method will be used to approximate these functions. Under this method, the functions are approximated by first dividing the area under the curve into rectangles with a base of Δg and a height of h(g), then calculating the area of each rectangle, and finally adding up all the rectangle areas. (Finney, Thomas, Demana, and Waits 361-364).

The Riemann Sums right-end-points approximation for the min and max cumulative distribution functions is $\mathbf{H}(g) = \sum_{j=-4.4995}^{g}(h(j)\Delta g)$. In addition, the approximation for the expected values is $\mathbf{E}(g) = \sum_{j=-4.4995}^{4.5000}(jh(j)\Delta g)$ and the approximation for the standard deviations is $\sigma = \sqrt{\sum_{j=-4.4995}^{4.5000}(j^2h(j)\Delta g) - \left(\sum_{j=-4.4995}^{4.5000}(jh(j)\Delta g)\right)^2}$. Although the g values fall within the interval (-∞, ∞), the functions will be approximated for g values in the interval [-4.50, 4.50]. This is done since most g values fall within this range. A base value of Δg = 0.0005 is used for the approximations.

Interpolation is then used to find the extreme interval value g_α for a given probability α. Extreme interval values are found for probabilities ranging from α = 0.01 to α = 0.99 and observation sizes ranging from n = 1 to n = 1000. The VBA application looks for the largest cumulative distribution function value below α and the smallest cumulative distribution function value above α. The interpolation formula: $g_\alpha = g_1 + \frac{(g_2 - g_1)\times(\alpha - H(g_1))}{(H(g_2) - H(g_1))}$, where $H(g_1) < \alpha < H(g_2)$ and $g_1 < g_\alpha < g_2$, is used to find the extreme interval value g_α (Law and Kelton 470).

Example: n = 100 Observations

If g = min ($z_1, \ldots\ldots, z_{100}$), where z_i is normal with $\mu = 0$ and $\sigma^2 = 1$, then the min probability density function is $h(g) = 100f(g)(1 - \Phi(g))^{99}$.

If g = max ($z_1, \ldots\ldots, z_{100}$), where z_i is normal with $\mu = 0$ and $\sigma^2 = 1$, then the max probability density function is $h(g) = 100f(g)\Phi(g)^{99}$.

The following table contains some of the min and max extreme interval values for an observation size of n = 100 and different probabilities (α). For example, given the probability $P(g \le g_{0.01})$ = α = 0.01, the min extreme interval value is $g_{0.01}$ = -3.70951 and the max extreme interval value is $g_{0.01}$ = 1.69507. The min extreme interval value $g_{0.50}$ = -2.46193 is the min median, and the max extreme interval value $g_{0.50}$ = 2.46179 is the max median.

$P(g \le g_\alpha) = \alpha$	Min g_α	Max g_α
0.01	-3.70951	1.69507
0.05	-3.28170	1.88775
0.10	-3.07403	1.99951
0.30	-2.69094	2.25792
0.50	-2.46193	2.46179
0.70	-2.25806	2.69089
0.90	-1.99940	3.07460
0.95	-1.88726	3.28316
0.99	-1.69209	3.71752

Other Parameter Values

It is also possible to find the min and max extreme interval values for normal variables with parameters other than $\mu = 0$ and $\sigma^2 = 1$. The extreme interval value g_α when $\mu = 0$ and $\sigma^2 = 1$ is first found. Then, apply the equation: $g'_\alpha = \mu + g_\alpha \sigma$ to find the min and max extreme interval values. The following example shows some min and max extreme interval values for an observation size of n = 50, a mean of $\mu = 30$, and a variance of $\sigma^2 = 49$.

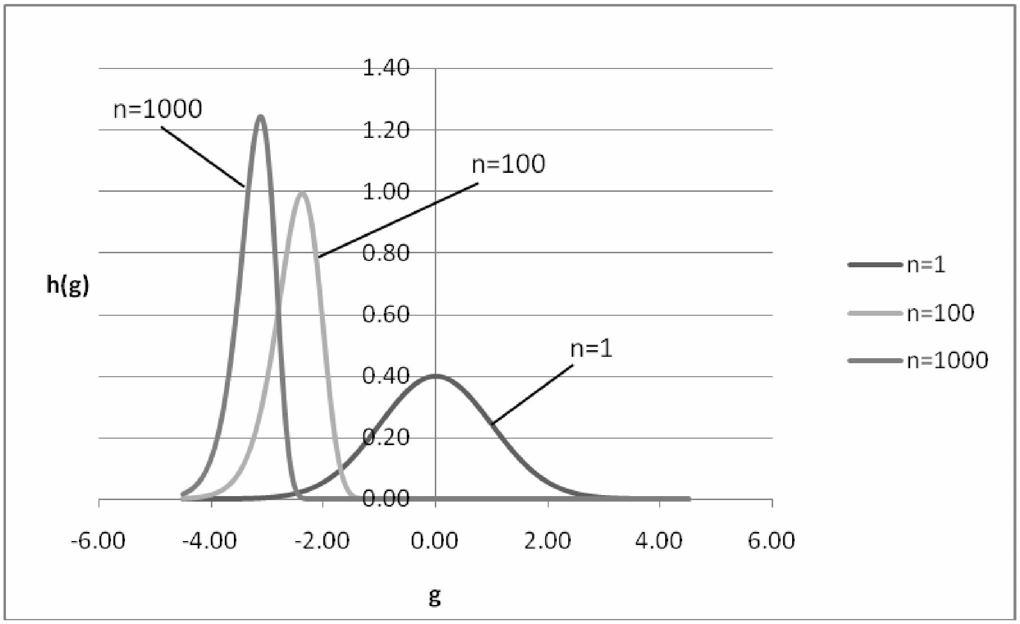
Let $g' = \min(x_1, \ldots\ldots, x_{50})$ or $g' = \max(x_1, \ldots\ldots, x_{50})$, where x_i is normally distributed with parameters $\mu = 30$ and $\sigma^2 = 49$. Suppose one wants to find the min and max extreme interval value $g'_{0.01}$ for the probability $P(g' \le g'_{0.01}) = \alpha = 0.01$. The min extreme interval value is $g'_{0.01} = \mu + g_\alpha \sigma = 30 + (-3.53456)(7) = 5.25808$ and the max extreme interval value is $g'_{0.01} = \mu + g_\alpha \sigma = 30 + (1.35299)(7) = 39.47093$. The min and max extreme interval values are listed below for α = 0.01, 0.05, 0.10, 0.30, 0.50, 0.70, 0.90, 0.95, and 0.99.

$P(g' \le g'_\alpha) = \alpha$	Min g_α when $\mu = 0$ and $\sigma^2 = 1$	Min g'_α when $\mu = 30$ and $\sigma^2 = 49$	Max g_α when $\mu = 0$ and $\sigma^2 = 1$	Max g'_α when $\mu = 30$ and $\sigma^2 = 49$
0.01	-3.53456	5.25808	1.35299	39.47093
0.05	-3.08201	8.42593	1.57020	40.99140
0.10	-2.86167	9.96831	1.69507	41.86549
0.30	-2.45176	12.83768	1.98082	43.86574
0.50	-2.20391	14.57263	2.20360	45.42520
0.70	-1.98112	16.13216	2.45150	47.16050
0.90	-1.69523	18.13339	2.86173	50.03211
0.95	-1.57015	19.00895	3.08254	51.57778
0.99	-1.35154	20.53922	3.53852	54.76964

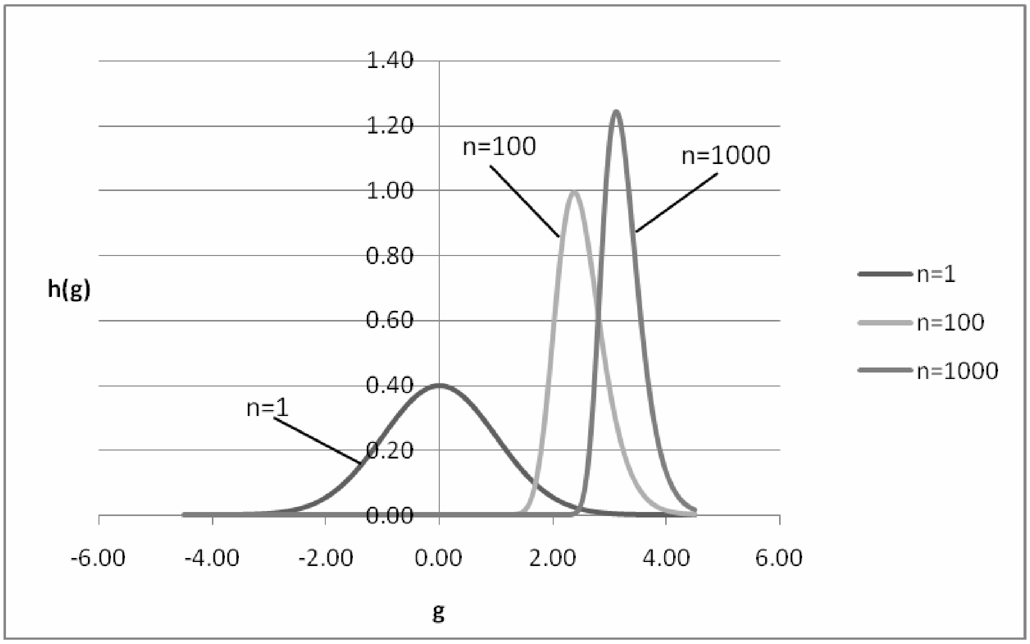
Analysis

As the observation size increases, the min probability density function shifts farther to the left and in a corresponding way, the max probability density function shifts farther to the right. The following graphs show the min and max probability density functions for observation sizes of n = 1, 100, and 1000.

Min Probability Density Function for Observation Sizes n = 1, 100, and 1000



Max Probability Density Function for Observation Sizes n = 1, 100, and 1000



The min extreme interval values, expected value, median, and mode become more negative as the observation size increases. The max extreme interval values, expected value, median, and mode become more positive as the observation size increases. As the observation size increases, the min and max standard deviations and the max coefficient of variation decrease and the min coefficient of variation increases.

The following tables display the min and max extreme interval values for α = 0.10 and α = 0.90, expected values, medians, modes, standard deviations, and coefficient of variations (CV) for the observation sizes of n = 1, 10, 50, 100, and 1000. Note, the min and max values are the same and the coefficient of variation is not available when the observation size is n = 1.

Min Extreme Interval Values and Statistics

n	α = 0.10	α = 0.90	μ	Median	Mode	σ	CV
1	-1.28178	1.28132	0.00000	-0.00024	0.00000	0.99992	NA
10	-2.30879	-0.82169	-1.53859	-1.49897	-1.42000	0.58659	-0.38125
50	-2.86167	-1.69523	-2.24827	-2.20391	-2.11650	0.46426	-0.20650
100	-3.07403	-1.99940	-2.50599	-2.46193	-2.37500	0.42999	-0.17158
1000	-3.69712	-2.82922	-3.22529	-3.19500	-3.11550	0.38993	-0.12090

Max Extreme Interval Values and Statistics

n	α = 0.10	α = 0.90	μ	Median	Mode	σ	CV
1	-1.28178	1.28132	0.00000	-0.00024	0.00000	0.99992	NA
10	0.82128	2.30843	1.53859	1.49852	1.42000	0.58659	0.38125
50	1.69507	2.86173	2.24828	2.20360	2.11650	0.46426	0.20650
100	1.99951	3.07460	2.50599	2.46179	2.37500	0.42998	0.17158
1000	2.83355	3.70571	3.22532	3.19737	3.11550	0.38984	0.12087

Application

There are many potential applications of this research. One application is a system that consists of "n" components. The system fails when one of the components stops running. An example of this type of system is a race car with four tires. The race car will stop running when one of the tires fails.

Suppose the number of miles that a race car tire will run before failing is normally distributed with a mean of 250 miles and a standard deviation of 50 miles. One wants to determine the number of miles M, where the probability that the race car is still running after M miles is 90%. The number of miles M can be determined by first finding the min normal extreme interval value g_α when n = 4, α = 0.10, and the parameters are $\mu = 0$ and $\sigma^2 = 1$. Then, the equation $M = \mu + g_\alpha \sigma$ is used to find M.

Let g = min($x_1,\ldots,x_4$), where g is the number of miles that the race car travels before it stops running. The variable x_i is the number of miles that a race car tire will run before failing. This variable is normally distributed with parameters $\mu = 250$ and $\sigma = 50$. One wants to determine M such that P (g > M) = 1 – P(g ≤ M) = 0.90. The min extreme interval value for n = 4 and α = 0.10 when the normal parameters are $\mu = 0$ and $\sigma^2 = 1$ is -1.94338. Thereby, $M = \mu + g_\alpha \sigma$ = 250 + (-1.94338)(50) = 152.831. Hence, the probability is 90% that the race car will still be running after 152.831 miles.

Conclusion

This paper presents a method for finding the min and max normal extreme interval values and statistics for different observation sizes. An example displaying some extreme interval values for an observation size of n = 100 and how to find the min and max extreme interval values for normal variables with parameters other than $\mu = 0$ and $\sigma^2 = 1$ are provided. In addition, an application of this research is presented along with an analysis of the min and max probability density functions, extreme interval values, and statistics.

References

Finney, Ross L., George B. Thomas Jr., Franklin D. Demana, and Bert K. Waits. *Calculus Graphical, Numerical, Algebraic*. Addison-Wesley Publishing Company Inc., 1994.

Hines, William W., Douglas C. Montgomery, David M. Goldsman, and Connie M. Borror. *Probability and Statistics in Engineering Fourth Edition*. John Wiley & Sons, Inc., 2003.

Jance, Marsha L. *Calculating Min and Max Extreme Interval Values for Various Distributions*. Doctoral dissertation. Chicago, IL: Illinois Institute of Technology Stuart School of Business, 2007.

Law, Averill M., and W. David Kelton. *Simulation Modeling and Analysis Third Edition*. McGraw-Hill, 2000.

Microsoft Excel: Professional Edition 2003, Redmond, WA: Microsoft Corporation, 2003.

United States Department of Commerce, National Bureau of Standards. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables National Bureau of Standards Applied Mathematics Series 55*. Edited by Milton Abramowitz and Irene A. Stegun, Washington D.C., 1964.