Min And Max Exponential Extreme Interval Values And Statistics

Marsha Jance, Illinois Institute of Technology Stuart School of Business, USA
Nick Thomopoulos, Illinois Institute of Technology Stuart School of Business, USA

ABSTRACT

The extreme interval values and statistics (expected value, median, mode, standard deviation, and coefficient of variation) for the smallest (min) and largest (max) values of exponentially distributed variables with parameter λ = 1 are examined for different observation (sample) sizes. An extreme interval value g_α is defined as a numerical bound where a specified percentage α of the data is less than g_α. A procedure for finding the extreme interval values when λ > 0, an analysis of the extreme interval values and statistics, and an application of this research are provided.

Keywords: Min and Max Exponential, Extreme Interval Values

INTRODUCTION

This paper presents a method for determining the extreme interval values and statistics for the min (minimum) and max (maximum) observations in a sample of size n. The observations in the sample come from an exponential distribution with parameter λ = 1. An extreme interval value g_α is defined as a numerical bound where a specified percentage α of the data is less than g_α. For example, if the probability is $P(g < g_{0.25}) = 0.25$, then α = 0.25 and 25% of the data is below $g_{0.25}$. 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 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 normally, exponentially, and uniformly distributed variables. Jance developed Excel VBA (Visual Basic for Applications) programs to find the min and max extreme interval values and statistics for these distributions. The dissertation includes tables, graphs, and applications of this research. Jance's tables display the min and max extreme interval values and statistics for the normal, exponential, and uniform distributions for different observation sizes.

MIN AND MAX VALUES

Suppose one continually takes samples of size n from a continuous population that has a probability density function f(x) and a cumulative distribution function F(x). Then, from each sample one selects the largest (max) and smallest (min) observations. The smallest and largest values will vary from sample to sample. As a result, the min and max values will have a probability density function and a corresponding cumulative distribution function associated with them.

If the variable $g = \min(x_1, \ldots\ldots, x_n)$ is the minimum of the n observations, then the min probability density function is $h(g) = nf(g)(1 - F(g))^{(n-1)}$ (Hines, Montgomery, Goldsman, and Borror 215). Now, if the n observations come from an exponentially distributed population with parameter λ = 1, the min probability density function is $h(g) = ne^{-g}(e^{-g})^{(n-1)} = n(e^{-g})^n$. In addition, the min cumulative distribution function is $H(g) = \int_0^g h(g)\,dg$, and the min expected value is $E(g) = \mu = \int_0^\infty g h(g) dg$. Finally, the min variance is $\sigma^2 = E(g^2) - E(g)^2$ where $E(g^2) = \int_0^\infty g^2\, h(g) dg$.

If the variable g = max ($x_{1,}......, x_n$) is the maximum of the n observations, then the max probability density function is $h(g) = nf(g)F(g)^{(n-1)}$ (Hines, Montgomery, Goldsman, and Borror 215). Now, if the n observations come from an exponentially distributed population with parameter λ = 1, the max probability density function is $h(g) = ne^{-g}(1 - e^{-g})^{(n-1)}$. In addition, the max cumulative distribution function is $H(g) = \int_0^g h(g)\, dg$, and the max expected value is $E(g) = \mu = \int_0^\infty gh(g)dg$. Finally, the max variance is
$\sigma^2 = E(g^2) - E(g)^2$ where $E(g^2) = \int_0^\infty g^2\, h(g)dg$.

MIN AND MAX EXTREME INTERVAL VALUES

An Excel VBA application was developed to find the exponential min and max extreme interval values and statistics. First, the min and max cumulative distribution functions: H(g) are evaluated beginning with g = 0 to the g value where H(g) is slightly greater than 0.99. The VBA application works in conjunction with MATLAB's Excel Link and integration function int() to evaluate the max expected value, max variance, and max cumulative distribution function.

Next, interpolation is 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 = 70 OBSERVATIONS

If g = min ($x_{1,}......., x_{70}$), where x is exponential with λ = 1, the min probability density function is $h(g) = 70e^{-g}(e^{-g})^{69} = 70e^{-70g}$, and the min cumulative distribution function is $H(g) = \int_0^g h(g)d(g)$. The min expected value is μ = 0.014286, the min standard deviation is σ = 0.014286, the min mode is 0.000000, and the min coefficient of variation is 1.000000. The min median is the min extreme interval value $g_{0.50}$ = 0.009902.

If g = max ($x_{1,}......., x_{70}$), where x is exponential with λ = 1, the max probability density function is $h(g) = 70e^{-g}(1 - e^{-g})^{69}$, and the max cumulative distribution function is $H(g) = \int_0^g h(g)\, dg$. The max expected value is μ = 4.832837, the max standard deviation is σ = 1.277008, the max mode is 4.250000, and the max coefficient of variation is 0.264236. The max median is the max extreme interval value $g_{0.50}$ = 4.619955.

The min and max extreme interval values for an observation size of n = 70 and for different probabilities α can be found in the following table. For example, given the probability $P(g < g_{0.99})$ = 0.99, the min extreme interval value is $g_{0.99}$ = 0.065788, and the max extreme interval value is $g_{0.99}$ = 8.848722.

$\mathbf{P(g < g_\alpha) = \alpha}$	**Min g_α**	**Max g_α**
0.01	0.000144	2.753985
0.05	0.000733	3.172609
0.10	0.001505	3.430860
0.20	0.003188	3.784077
0.30	0.005095	4.071455
0.40	0.007298	4.342455
0.50	0.009902	4.619955
0.60	0.013090	4.923875
0.70	0.017200	5.281977
0.80	0.022992	5.750029
0.90	0.032894	6.499617
0.95	0.042796	7.219061
0.99	0.065788	8.848722

OTHER λ VALUES

The inverse-transform method for generating an exponential variable is $x = \frac{-\ln(1-u)}{\lambda}$ where u is uniformly distributed with parameters a = 0 and b = 1 (Law and Kelton 460). When λ = 1, the exponential variable x is –ln(1-u). Suppose there is another exponential variable with parameter λ > 0. Thus, this exponential variable is $-\frac{\ln(1-u)}{\lambda} = \frac{x}{\lambda}$.

Using the inverse-transform method, the min and max exponential extreme interval values can be found for parameter values other than λ = 1 provided that λ > 0. First, find the extreme interval value when λ = 1. Then, the extreme interval value for λ > 0 is $g'_{\alpha} = \frac{g_{\alpha}}{\lambda}$. In the following example, the min and max extreme interval values are found for an observation size of n = 70, α = 0.10, and λ = 10.

Let $g' = \min(x_1, \ldots\ldots, x_{70})$ where x is exponentially distributed with parameter λ = 10. The min extreme interval value is $g_{0.10}$ = 0.001505 when n = 70, α = 0.10, and λ = 1. Thus, when n = 70, α = 0.10, and λ = 10, the min extreme interval value is $g'_{0.10} = \frac{g_{0.10}}{\lambda} = \frac{0.001505}{10}$ = 0.000151.

The following shows that $g'_{0.10}$ = 0.000151 is the min extreme interval value when n = 70, α = 0.10, and λ = 10. When λ = 10 and n = 70, the min probability density function is
$h(g') = 70\left(10e^{-10g'}\right)\left(1-\left(1-e^{-10g'}\right)\right)^{69} = 700e^{-700g'}$, and the min cumulative distribution function is

$H(g') = \int_0^{g'} h(g')dg'$. Hence, $H(g') = \int_0^{0.000151} h(g')dg'$ is approximately equal to 0.10 when g' = 0.000151.

Let $g' = \max(x_1, \ldots\ldots, x_{70})$ where x is exponentially distributed with parameter λ = 10. The max extreme interval value is $g_{0.10}$ = 3.430860 when n = 70, α = 0.10, and λ = 1. Thus, when n = 70, α = 0.10, and λ = 10, the max extreme interval value is $g'_{0.10} = \frac{g_{0.10}}{\lambda} = \frac{3.430860}{10}$ = 0.343086.

The following shows that $g'_{0.10}$ = 0.343086 is the max extreme interval value when n = 70, α = 0.10, and λ = 10. When λ = 10 and n = 70, the max probability density function is $h(g') = 70\left(10e^{-10g'}\right)\left(1-e^{-10g'}\right)^{69}$, and the max cumulative distribution function is $H(g') = \int_0^{g'} h(g')dg'$. Hence, $H(g') = \int_0^{0.343086} h(g')dg'$ is approximately equal to 0.10 when g' = 0.343086.

ANALYSIS

The min extreme interval values, expected value, and median shift closer to zero, and the min standard deviation decreases as the observation size increases. In addition, the min mode is zero and the min mode < min median < min expected value for all observation sizes. The min coefficient of variation is one for all observation sizes since the min expected value and the min standard deviation are equal to one another.

As the observation size becomes larger, the max extreme interval values, expected value, median, mode, and standard deviation increase in value; whereas, the max coefficient of variation decreases in value. The max mode < max median < max expected value for all observation sizes.

The following tables display the min and max extreme interval values for α = 0.05 and α = 0.95, expected values, standard deviations, modes, medians, and coefficient of variations (CV) for the observation sizes of n = 1, 5, 10, 50, 70, 100, 500, and 1000. The min and max extreme interval values and statistics are the same when the observation size is n = 1.

Min Extreme Interval Values and Statistics

n	α = 0.05	α = 0.95	μ	σ	Mode	Median	CV = σ/μ
1	0.051293	2.995732	1.000000	1.000000	0.000000	0.693147	1.000000
5	0.010259	0.599146	0.200000	0.200000	0.000000	0.138629	1.000000
10	0.005129	0.299573	0.100000	0.100000	0.000000	0.069315	1.000000
50	0.001026	0.059915	0.020000	0.020000	0.000000	0.013863	1.000000
70	0.000733	0.042796	0.014286	0.014286	0.000000	0.009902	1.000000
100	0.000513	0.029957	0.010000	0.010000	0.000000	0.006931	1.000000
500	0.000103	0.005991	0.002000	0.002000	0.000000	0.001386	1.000000
1000	0.000051	0.002996	0.001000	0.001000	0.000000	0.000693	1.000000

Max Extreme Interval Values and Statistics

n	α = 0.05	α = 0.95	μ	σ	Mode	Median	CV = σ/μ
1	0.051293	2.995732	1.000000	1.000000	0.000000	0.693147	1.000000
5	0.796885	4.584770	2.283333	1.209798	1.610000	2.044470	0.529839
10	1.351433	5.275356	2.928968	1.244897	2.300000	2.703559	0.425029
50	2.844617	6.882741	4.499205	1.274807	3.910000	4.285463	0.283340
70	3.172609	7.219061	4.832837	1.277008	4.250000	4.619955	0.264236
100	3.522902	7.575634	5.187378	1.278665	4.610000	4.975151	0.246495
500	5.120410	9.184866	6.792823	1.281771	6.210000	6.581816	0.188695
1000	5.812048	9.877984	7.485471	1.282160	6.910000	7.274619	0.171287

APPLICATION

Suppose one wants to determine the minimum number of components running in parallel for there to be at least a 95% chance that a system will still be running after four hours. One is seeking the minimum n when $g' = \max(x_1, \ldots\ldots, x_n)$ and $P(g' > T) = 0.95$ for T = 4.00 hours. Let x be the failure time of one component where x is exponential with λ = 1/5 per hour or E(x) = 1/λ = 5 hours. The probability $P(g' > 4) = 0.95$ is equivalent to $P(g' < 4) = 0.05$.

One wants to find the minimum n for λ = 1/5 per hour and $P(g' < 4) = 0.05$. The n for $P(g' < 4) = 0.05$ and λ = 1/5 is equivalent to the n for λ = 1 and $P(g < g_{0.05}) = 0.05$ where $g_{0.05} = T\lambda = 4 \times (1/5) = 0.80$.

Jance's max exponential extreme interval value table can be used to find n when $P(g < 0.80) = 0.05$ and λ = 1. In the table the closest entry for $P(g < 0.80) = 0.05$ is when n = 5. Hence, if there are at least five components there is a 95% chance that the system will still be running after four hours.

CONCLUSION

This paper presents a method for determining the exponential extreme interval values and statistics for the min (minimum) and max (maximum) observations in a sample of size n. An example, showing some min and max extreme interval values for an observation size of n = 70 is provided. In addition, a procedure for finding the extreme interval values when the exponential parameter is different from λ = 1 (but λ > 0) is discussed. Finally, an analysis of the min and max extreme interval values and statistics and an application of this research are presented.

AUTHOR INFORMATION

Marsha Jance received her Ph.D. in Management Science from Illinois Institute of Technology Stuart School of Business. Her research interests include developing new statistical and mathematical programming techniques to solve complex business problems.

Professor Thomopoulos has degrees in Business and in Mathematics from the University of Illinois (Urbana) and in Industrial Engineering (Ph.D.) from IIT. Nick was Supervisor of Operations Research at International Harvester Company, Senior Scientist at IIT Research Institute and Professor at IIT. He is the author of three books: Assembly

Line Systems (Hayden Books), Applied Forecasting Methods (Prentice Hall) and Strategic Inventory Management and Planning (Hitchcock Publishing Company).

REFERENCES

1. 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.
2. 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.
3. Law, Averill M., and W. David Kelton. *Simulation Modeling and Analysis Third Edition*. McGraw-Hill, 2000.
4. MATLAB: Version 7.1, The MathWorks, 2005.
5. Microsoft Excel: Professional Edition 2003, Redmond, WA: Microsoft Corporation, 2003.

NOTES

NOTES