



## Keywords

Peak Bone Mineral Density, Osteoporosis Diagnosis, Gumbel Maximum Distribution

Received: September 13, 2016 Accepted: October 3, 2016 Published: October 18, 2016

# Estimating Peak Bone Mineral Density in Osteoporosis Diagnosis by Maximum Distribution

## Loc Nguyen

Sunflower Soft Company, Ho Chi Minh City, Vietnam

## **Email address**

ng\_phloc@yahoo.com

## Citation

Loc Nguyen. Estimating Peak Bone Mineral Density in Osteoporosis Diagnosis by Maximum Distribution. *International Journal of Clinical Medicine Research*. Vol. 3, No. 5, 2016, pp. 76-80.

## Abstract

The T-score is very important to diagnosis of osteoporosis and its formula is calculated from two parameters such as peak bone mineral density (pBMD) and variance of pBMD. This research proposes a new method to estimate these parameters with recognition that pBMD conforms maximum distributions, for instance, Gumbel distribution. Firstly, my method models pBMD sample as a series of maximum values and such values are assumed to obey Gumbel distribution. Secondly, I apply moment technique to estimate mean and variance of Gumbel distribution with regard to the series of maximum values. These mean and variance are adjusted to become the best estimates of pBMD and pBMD variance. There is no normality assumption in this research because pBMD is essentially the extreme value with low frequency in population and so the estimate of pBMD will be more accurate if we take full advantage of specific characteristics of Gumbel maximum distribution.

## **1. Introduction to pBMD Estimation**

T-score is reference measure used by DXA instrument to diagnose osteoporosis but authors [1] indicates that DXA instrument tends to be over-diagnosing osteoporosis when DXA instrument is used in Vietnamese population. The issue is that T-score is calculated from peak bone mineral density (pBMD) and pBMD in Vietnamese population is different in other population where DXA instrument is produced. The authors [1] proposed a new approach to estimate Vietnamese pBMD by applying polynomial regression model. Author [2] used 4<sup>th</sup>-order polynomial fit model for estimating pBMD and she/he also used bootstrap method to obtain pBMD data like I do. Polynomial regression model in [1, p. 2] is very effective because it is appropriate to statistical cross-sectional and it is based on normality assumption but the research [1] meets the difficulty when determining the confident interval of pBMD. Originating from the research of authors [1], I propose another approach to estimate pBMD and variance of pBMD without normality assumption and help us to be easy to calculate confident interval of pBMD. This approach assumes that pBMD values conform a maximal distribution, for instance, Gumbel distribution; thus, pBMD and variance of pBMD are mean and variance of such maximal distribution, respectively. Before the way to estimate maximal mean and variance is described in next section, we should browse over some definitions about T-score and pBMD.

Bone mineral density (BMD) is amount of bone cell per square centimeter of bone [2]. There are many means to test BMD but dual-energy X-ray absorptiometry (DXA) mean is very popular [3]. DXA instrument uses two X-ray beams to project on bone and the BMD is measured by the absorption of each beam by bone [3]. When BMD is

determined, the relevant T-score is calculated according to following formula [1]:

$$T - score = \frac{BMD - pBMD}{standard deviation of pBMD}$$

Where pBMD is peak bone mineral density which is the main object in this research. World Health Organization "http://www.who.int" defines criteria of osteoporosis according to T-score. If T-score is greater than or equal to -1.0 then it is normal. If T-score is in interval [-2.5, -1.0] then that is osteoporosis. The most important parameters in T-score formula is relevant pBMD and its variance. According to [4], Peak bone mineral density (pBMD) is defined as BMD at the end of skeletal maturation and it lasts for a sufficient time interval. As aforementioned, pBMD varies from population to population, which causes lack of accuracy in determining

T-score in different populations. This research focuses on how to estimate pBMD and its variance from data sample collected in area where research is done in order to improve the accuracy of relevant pBMD.

## 2. A New Method to Estimate pBMD Based on Maximum Distribution

Given *m* independent identical variables  $\{X_1, X_2, ..., X_m\}$ with the same accumulative distribution  $F_1(X_1) = F_2(X_2)$  $= ... = F_m(X_m) = F(X)$  where *X* is the notation representing all variables  $X_i$  (s). Suppose *Y* is the maximum value among variables  $X_i$  (s) and *Y* is considered random variable. Let P(Y < y) is the accumulative distribution of *Y*, we have [5, p. 2]:

$$\begin{split} P(Y < y) &= P(X_1 < y \text{ and } X_2 < y \text{ and } \dots \text{ and } X_m < y) = P(X_1 < y)P(X_2 < y) \dots P(X_m < y) \\ & (\text{Due to } X_1, X_2, \dots, X_m \text{ mutually independent}) \\ &= F_1(X_1)F_2(X_2) \dots F_m(X_m) = F(X_1)F(X_2) \dots F(X_m) = F(X)^m \end{split}$$

It is proved that  $P(Y < y) = F(X)^m$  approaches to Gumbel maximum distribution denoted G(Y) when *m* approaches infinite limit [5, p. 2]. So *Y* conforms approximately to Gumbel distribution G(Y), we have [6]:

$$G \sim G(Y) = \exp\left(-\exp\left(-\frac{Y-\alpha}{\beta}\right)\right)$$

Where  $\alpha$  and  $\beta$  are location parameter and scale parameter, respectively. The notation "exp" denotes exponent function. The Gumbel density function is [6]:

$$g(Y) = \frac{\mathrm{d}G(Y)}{\mathrm{d}Y} = \frac{1}{\beta} \exp\left(-\frac{Y-\alpha}{\beta}\right) \exp\left(-\exp\left(-\frac{Y-\alpha}{\beta}\right)\right)$$

In general, equation (1) specifies cumulative function G(Y) and density function g(Y) of Gumbel distribution.

$$G(Y) = \exp\left(-\exp\left(-\frac{Y-\alpha}{\beta}\right)\right)$$
$$g(Y) = \frac{1}{\beta}\exp\left(-\frac{Y-\alpha}{\beta}\right)\exp\left(-\exp\left(-\frac{Y-\alpha}{\beta}\right)\right) \qquad (1)$$

Let  $\mu$  and  $\sigma^2$  be theoretical mean and variance of G(Y), respectively [6].

$$\mu = \alpha + \gamma \beta$$
  
$$\sigma^2 = \frac{\beta \pi}{\sqrt{6}}$$
(2)

Where  $\gamma$  is Euler – Mascheroni constant,  $\gamma \approx 0.5772$ .

According to equation (2), it is easy to recognize that  $\mu$  and  $\sigma^2$  represent pBMD and variance of pBMD. What we do now is to estimate theoretical  $\mu$  and  $\sigma^2$  given sample observations  $\{X_1, X_2, ..., X_m\}$  where  $X_i(s)$  are BMD

measures collected in research. In other words, parameters  $\alpha$  and  $\beta$  will be estimated via observations  $\{X_1, X_2, ..., X_m\}$  because  $\mu$  and  $\sigma^2$  are calculated based on  $\alpha$  and  $\beta$ . Now observations  $D = \{X_1, X_2, ..., X_m\}$  is considered as a population, we re-sample *m* samples from *D* with replacement at each time according to bootstrap technique. Suppose  $\{X_{i1}, X_{i2}, ..., X_{im}\}$  is a sample at time *i* where  $X_{ij}$  is  $j^{ih}$  observation taken from *D* at time *i*. Let  $Y_i = \max\{X_{i1}, X_{i2}, ..., X_{im}\}$  be the maximum value at time *i*. After re-sampling *n* times, we have *n* new maximal observations  $E = \{Y_1, Y_2, ..., Y_n\}$ . Parameters  $\alpha$  and  $\beta$  are estimated according to these maximal observations. Let E(Y) and  $E(Y^2)$  be theoretical first-order moment and second-order moment of g(Y), respectively, we have:

$$\frac{\beta\pi}{\sqrt{6}} = \sigma^2 = E(Y - E(Y))^2 = E(Y^2) - E(Y)^2 = E(Y^2) - \mu^2 = E(Y^2) - (\alpha + \gamma\beta)^2$$
$$\Rightarrow E(Y^2) = (\alpha + \gamma\beta)^2 + \frac{\beta\pi}{\sqrt{6}}$$

 $E(Y) = \alpha + \gamma \beta$ 

The sample first-order moment and second-order moment are  $\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$  and  $\frac{1}{n} \sum_{i=1}^{n} Y_i^2$ , respectively. According to moment method, we estimate parameters  $\alpha$  and  $\beta$  by equating theoretical moments to sample moments. Let  $\hat{\alpha}$  and  $\hat{\beta}$  be estimated values of respective parameters  $\alpha$  and  $\beta$ ,  $\hat{\alpha}$  and  $\hat{\beta}$ are solutions of following equations:

$$\begin{cases} \alpha + \gamma \beta = \bar{Y} \\ (\alpha + \gamma \beta)^2 + \frac{\beta \pi}{\sqrt{6}} = \frac{1}{n} \sum_{i=1}^n Y_i^2 \end{cases}$$

$$\Rightarrow \begin{cases} \hat{\alpha} = \bar{Y} - \frac{\gamma\sqrt{6}}{\pi} \left( \frac{1}{n} \sum_{i=1}^{n} Y_i^2 - \bar{Y}^2 \right) \\ \hat{\beta} = \frac{\sqrt{6}}{\pi} \left( \frac{1}{n} \sum_{i=1}^{n} Y_i^2 - \bar{Y}^2 \right) \end{cases}$$

The quality of statistical estimates such as  $\hat{\alpha}$  and  $\hat{\beta}$  is dependent on expectation of such estimates. Estimate gets good quality if its expectation is near to theoretical parameter. Estimate is the best one called unbiased estimate if its expectation is equal to theoretical parameter. Let  $E(\hat{\alpha})$ and  $E(\hat{\beta})$  be expectations of  $\hat{\alpha}$  and  $\hat{\beta}$ , respectively. We have:

 $E\left(\begin{smallmatrix}1\\ \Sigma^n & V^2 & \overline{V}^2\end{smallmatrix}\right) = \begin{smallmatrix}1\\ \Sigma^n & E(V^2)$ 

$$\left(\mu^2 + \frac{1}{n}\sigma^2\right) = \frac{n-1}{n}\sigma^2 = \frac{(n-1)\beta\pi}{n\sqrt{6}}$$
$$\left(\text{due to } E(Y_i^2) = \mu^2 + \sigma^2 \text{ and } E(\bar{Y}^2) = \mu^2 + \frac{1}{n}\sigma^2\right)$$
$$E(\hat{\alpha}) = E(\bar{Y}) - \frac{\gamma\sqrt{6}}{\pi}E\left(\frac{1}{n}\sum_{i=1}^n Y_i^2 - \bar{Y}^2\right) = \mu - \frac{(n-1)\gamma\beta}{n} = \alpha + \gamma\beta - \frac{(n-1)\gamma\beta}{n}$$
$$E(\hat{\beta}) = \frac{\sqrt{6}}{\pi}E\left(\frac{1}{n}\sum_{i=1}^n Y_i^2 - \bar{Y}^2\right) = \frac{(n-1)\beta}{n}$$

Both  $\hat{\alpha}$  and  $\hat{\beta}$  are biased estimates because their expectations shift away from parameters  $\alpha$  and  $\beta$ . Hence,  $\hat{\alpha}$  and  $\hat{\beta}$  are adjusted according to equation (3).

Where  $Y_i$  is maximum value at time *i* and  $\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$  is sample mean of  $Y_i$  (s).

Expectations of  $\hat{\alpha}$  and  $\hat{\beta}$  are re-calculated with regard to equation (2) as follows:

 $E(\overline{V}^2)$  -

$$E(\hat{\alpha}) = E(\bar{Y}) - \frac{\gamma\sqrt{6}}{(n-1)\pi} E\left(\sum_{i=1}^{n} Y_{i}^{2} - n\bar{Y}^{2}\right) = E(\bar{Y}) - \frac{\gamma\sqrt{6}}{\pi} \frac{n}{n-1} E\left(\frac{1}{n}\sum_{i=1}^{n} Y_{i}^{2} - \bar{Y}^{2}\right) = \alpha + \gamma\beta - \frac{\gamma\sqrt{6}}{\pi} \frac{n}{n-1} \frac{(n-1)\beta\pi}{n\sqrt{6}} = \alpha + \gamma\beta - \gamma\beta = \alpha$$

$$E(\hat{\beta}) = \frac{\sqrt{6}}{(n-1)\pi} E\left(\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2\right) = \frac{\sqrt{6}}{\pi} \frac{n}{n-1} E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \bar{Y}^2\right) = \frac{\sqrt{6}}{\pi} \frac{n}{n-1} \frac{(n-1)\beta\pi}{n\sqrt{6}} = \beta$$

Now adjusted  $\hat{\alpha}$  and  $\hat{\beta}$  are the best estimates because they are unbiased estimates. Note that the method of moment estimation for Gumbel distribution is not new; please read [7, pp. 5-6] in which authors mentioned both moment method and maximum likelihood method. Authors [8] proposed Expectation Maximization (EM) algorithm to estimate parameters of extreme value distributions from truncated data. However my contribution in Gumbel parameter estimation is to adjust moment method in order to produce unbiased estimate of parameter  $\beta$  according to equation (3).

Let  $\hat{\mu}$  and  $\hat{\sigma}^2$  be estimated values of mean  $\mu$  and variance  $\sigma^2$ , respectively. From equation (2), we have equation (4) to determine these estimates:

$$G(Y) = s_0 \Leftrightarrow \exp\left(-\exp\left(-\frac{Y-\alpha}{\beta}\right)\right) = s_0 \Rightarrow Y_0 = \alpha - \beta \ln\left(-\ln(s_0)\right) \Rightarrow Y_0 = \hat{\alpha} - \hat{\beta} \ln\left(-\ln(s_0)\right)$$

Note that the notation "ln" denotes natural logarithm function.

Thus, the lower-tail confident interval of pBMD given significant level  $s_0$  is  $\mu \leq Y_0$ , respectively. The upper-tail confident interval of pBMD is computed in the similar way when  $\mu \geq Y_0$  and  $Y_0 = \hat{\alpha} + \hat{\beta} \ln(\ln(1 - s_0))$  is solution of  $G(Y) = 1 - s_0$ . In general, equation (5) determines percentile point denoted  $Y_0$  given significant level  $s_0$ .

$$\hat{\mu} = \hat{\alpha} + \gamma \hat{\beta} = \bar{Y}$$

$$\hat{\sigma}^2 = \frac{\hat{\beta}\pi}{\sqrt{6}} = \frac{1}{n-1} (\sum_{i=1}^n Y_i^2 - n\bar{Y}^2) = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$
(4)
Where  $\hat{\alpha}$  and  $\hat{\beta}$  are specified by equation (3).

In general,  $\hat{\mu}$  and  $\hat{\sigma}^2$  are estimates of pBMD and its variance. Another facility of this method is to allow us to determine the confident interval of pBMD given significant level. Gumbel distribution is not symmetric and so we should calculate one-sided confident interval. Given significant level  $s_0$ , percentile point denoted  $Y_0$  of Gumbel distribution is solution of following equation:

$$\begin{cases} Y_0 = \hat{\alpha} - \hat{\beta} \ln(-\ln(s_0)) \text{ given lower } - \text{ tail} \\ Y_0 = \hat{\alpha} - \hat{\beta} \ln(-\ln(1 - s_0)) \text{ given upper } - \text{ tail } (5) \\ \text{Where } \hat{\alpha} \text{ and } \hat{\beta} \text{ are specified by equation } (3). \end{cases}$$

There is no pre-calculated table for accessing percentage point in the similar way of standard normal distribution lookup table due to lack of accuracy if Gumbel distribution is approximated to normal distribution. So it is slightly complicated to calculate pBMD confident interval due to a few of arithmetic operations for determining  $Y_0$ .

#### 3. A Case Study of pBMD Estimation

Table 1. Re-sampling and maximum values.

No	Re-sampling	Y
1	1.08428, 0.972234, 1.08428, 1.14306, 1.14306	$Y_1 = 1.14306$
2	1.08428, 0.972234, 1.14306, 0.972234, 0.972234	$Y_2 = 1.14306$
3	0.972234, 0.972234, 1.06506, 1.06506, 1.06506	$Y_3 = 1.06506$
4	1.08428, 0.972234, 1.08428, 1.14306, 1.06506	$Y_4 = 1.14306$
5	1.06506, 1.08428, 1.00824, 1.00824, 1.08428	$Y_5 = 1.08428$
6	1.14306, 1.00824, 0.972234, 0.972234, 1.06506	$Y_6 = 1.14306$
7	1.00824,1.08428, 1.14306, 0.972234, 1.00824	Y <sub>7</sub> =1.14306
8	1.00824, 0.972234, 1.06506, 1.00824, 1.00824	$Y_8 = 1.06506$
9	1.08428, 0.972234, 1.08428, 1.14306, 1.14306	Y <sub>9</sub> =1.14306
10	1.14306, 1.14306, 1.06506, 0.972234, 1.14306	$Y_{10} = 1.14306$

For convenience, suppose BMD population mean is approximated to  $1g/\text{cm}^2$  with deviation  $0.1g/\text{cm}^2$ . Suppose we have simulated BMD sample  $D = \{X_1=0.972234, X_2=1.06506, X_3=1.08428, X_4=1.00824, X_5=1.14306\}$  of size n=5 which conforms normal distribution with mean 1 and variance  $0.1^2$ . After re-sampling m=10 times without replacement according to bootstrap technique, we have 10 new maximal observations  $E = \{Y_1, Y_2, ..., Y_{10}\}$  where  $Y_i$  is maximum value of sample  $\{X_{i1}, X_{i2}, ..., X_{i5}\}$  be the maximum value at time *i*. It is easy to recognize that *E* is simulated pBMD sample. Note that author [2] also used bootstrap method to obtain pBMD data like I do. Table 1 shows these maximum values.

Suppose  $Y_i$  conforms Gumbel distribution, we need to estimate its mean  $\hat{\mu}$  and variance  $\hat{\sigma}^2$  where  $\hat{\mu}$  is the estimated pBMD. According to equation (4), the parametric estimates  $\hat{\alpha}$  and  $\hat{\beta}$  must be determined first with regard to equation (3). We have:

$$\bar{Y} = \frac{1}{10} \sum_{i=1}^{10} Y_i = 1.12158$$
$$\hat{\alpha} = \bar{Y} - \frac{\gamma\sqrt{6}}{(10-1)\pi} \left( \sum_{i=1}^{10} Y_i^2 - 10\bar{Y}^2 \right) = 1.12103$$
$$\hat{\beta} = \frac{\sqrt{6}}{(10-1)\pi} \left( \sum_{i=1}^{10} Y_i^2 - 10\bar{Y}^2 \right) = 0.000953833$$
$$\hat{\mu} = \hat{\alpha} + \gamma\hat{\beta} = \bar{Y} = 1.12158$$
$$\hat{\sigma}^2 = \frac{\hat{\beta}\pi}{\sqrt{6}} = 0.00122334$$

Finally, the estimated pBMD is 1.12158 with deviation  $\sqrt{0.00122334} = 0.0349763$ . Moreover, the lower-tail percentile point denoted  $Y_0$  at significant level 0.025 is:

$$Y_0 = \hat{\alpha} + \hat{\beta} \ln(\ln(s_0)) = 1.12103 - 0.000953833 \ln(-\ln(0.025)) = 1.11979$$

The upper-tail percentile point denoted  $Y'_0$  at significant level 0.025 is:

$$Y_0' = \hat{\alpha} + \hat{\beta} \ln(\ln(s_0)) = 1.12103 - 0.000953833 \ln(-\ln(1 - 0.025)) = 1.12454$$

According to equation (1), given estimates  $\hat{\alpha}$  and  $\hat{\beta}$ , Gumbel density function g(Y) is:

 $g(Y) = 1048.4\exp(1175.29 - 1048.4Y)\exp(-\exp(1175.29 - 1048.4Y))$ 

Figure 1 shows the graph of such Gumbel density function g(Y) as follows:

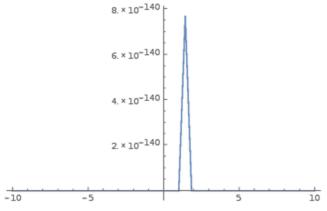
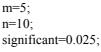


Figure 1. Graph of Gumbel density function.

As seen in figure 1, pBMD data concentrates highly on the peak of density function g(Y) at its mean  $\mu = 1.12158$ .

Although g(Y) is asymmetric, the lower-tail and upper-tail percentile points (1.11979 and 1.12454) at significant level 0.025 are very close to the peak. In other words, 95% confident interval containing the mean  $\mu = 1.12158$  is very narrow (0.00475 = 1.12454 - 1.11979), which indicates that such mean is the most likely pBMD.

Following is code snip for pBMD estimation mentioned above, written by Wolfram language built in Mathematica software [7].



(\* *Simulated sample D with mean* 1 *and deviation* 0.1 \*) x=RandomVariate[NormalDistribution[1,0.1],m]

(\* Producing maximal observations E by bootstrap technique \*)

yy=Table[RandomChoice[x,m], {i,n}]

y=Map[Max,yy]

(\* Estimating parameters such as alpha and beta \*) ymean=Mean[y] beta=Sqrt[6]/((n-1)Pi)(Sum[(y[[i]])^2, {i,n}]-n\*(ymean^2)) alpha=ymean-EulerGamma\*beta

(\* Estimating mean and variance \*) mu=alpha+EulerGamma\*beta var=beta\*Pi/Sqrt[6] Sqrt[var]

(\* Percentile points at significant level 0.025 \*) lowerpercentile=alpha-beta\*Log[-Log[significant]] upperpercentile=alpha-beta\*Log[-Log[1-significant]] g[z\_]:=1/beta\*Exp[-(z-alpha)/beta]\*Exp[-Exp[-(zalpha)/beta]]

g[z]

(\* *Plotting the Gumbel density function* \*) Image[Plot[g[z],{z,-10,10}]]

## 4. Conclusion

The ideology of this research is to apply maximum Gumbel distribution into estimating pBMD and its variance by moment technique, instead of using normal approximation to maximal values. This is my main contribution in the area of pBMD estimation. So Gumbel distribution ensures accuracy of estimation and this is the strong point of my method. The drawback is the complexity of estimation when it is required a lot of calculus and arithmetic operations. In the future, I will research how to make approximation of Gumbel distribution so that the derived distribution is simpler than Gumbel distribution.

## Acknowledgement

This research is the place to acknowledge Doctor Ho-Pham, Lan T. – Pham Ngoc Thach University of Medicine, Ho Chi Minh city, Vietnam who produces an excellent research [1] from which my research is inspired. Moreover she gave me valuable comments and advices that help me to improve this research.

## References

- [1] L. T. Ho-Pham, H. N. Pham, T. Q. Lai, U. D. T. Nguyen, N. D. Nguyen and T. V. Nguyen, "Reference Ranges for Bone Mineral Density and Prevalence of Osteoporosis in Vietnamese Men and Women," *BMC Musculoskeletal Disorders*, vol. 12, no. 182, pp. 1-7, 2011.
- [2] D. A. Banu, "Peak Bone Mineral Density Of Bangladeshi Men And Women," *International Journal of Scientific & Technology Research*, vol. 4, no. 11, pp. 8-13, November 2015.
- [3] Wikipedia, "Bone density," Wikimedia Foundation, 15 February 2016. [Online]. Available: https://en.wikipedia.org/wiki/Bone\_density.
- [4] Wikipedia, "Dual-energy X-ray absorptiometry," Wikimedia Foundation, 7 March 2016. [Online]. Available: https://en.wikipedia.org/wiki/Dual-energy\_Xray\_absorptiometry.
- [5] J. P. Bonjour, G. Theintz, F. Law, D. Slosman and R. Rizzoli, "Peak bone mass," *Osteoporosis international*, vol. 4, no. 1 Supplement, pp. 7-13, January 1994.
- [6] D. Veneziano, "Distribution of the maximum of independent identically-distributed variables," MIT OpenCourseWare, Cambridge, 2005.
- [7] Wikipedia, "Gumbel distribution," Wikimedia Foundation, 29 March 2016. [Online]. Available: https://en.wikipedia.org/wiki/Gumbel\_distribution.
- [8] J. Khemwong, T. Boonyawiwat, T. Kriengkomol, J. Songsiri and P. Hoisungwan, "Parameter estimation of Gumbel distribution for," Jitkomut Songsiri Homepage, Bangkok, 2015.
- [9] T. Kernane and Z. Raizah, "Estimation of the Parameters of Extreme Value Distributions from Truncated Data Via the EM Algorithm," The open archive HAL, 2010.
- [10] Wolfram, Mathematica, vol. 10, Wolfram Research.