DISCRIMINATING BETWEEN NORMAL AND GUMBEL DISTRIBUTIONS *

Authors: Abdelaziz Qaffou
– Department of Applied Mathematics,
Faculty of Sciences and Techniques, Sultan Moulay Slimane University,
Beni Mellal. Morocco (aziz.qaffou@gmail.com)

Abdelhak Zoglat
– Department of Mathematics,
Faculty of Sciences, Mohammed V University,
Rabat. Morocco (azoglat@gmail.com)

Abstract:

- The normal and Gumbel distributions are much alike in practical engineering, in flood frequency and they are similar in appearance, especially for small samples. So the aim of this paper is to discriminate between these two distributions. Considering the logarithm of the ratio of the maximized likelihood (RML) as test statistic, its asymptotic distribution is found under both normal and Gumbel distributions, which can be used to compute the probability of correct selection (PCS). Finally, Monte Carlo (MC) simulations are performed to examine how the asymptotic results work for finite sample sizes.

Key-Words:

- *Normal and Gumbel distributions; probability of correct selection; ratio of maximized likelihood, Monte Carlo simulation.*

AMS Subject Classification:

- 62H10, 62H12, 62H15, 65C05.

---

*The opinions expressed in this text are those of the authors and do not necessarily reflect the views of any organization.

1. INTRODUCTION

In engineering practice, risk criteria and economic considerations are important parts of a project design. These criteria are crucial, for example, in the design of an urban sewer network, the sizing of a hydraulic structure, or the conception of a storage capacity system. The adequate knowledge of design events (e.g., design flood magnitudes) is often helpful for the proper sizing of a project to avoid the high initial investments associated with the oversizing of the project and the large future failure costs resulting from its undersizing.

To estimate these design events, statistical frequency analysis of hydrological data is often used; it consists of fitting a probability distribution to a set of recorded hydrological values (e.g., annual maximum flood series) and obtaining estimated results concerning the underlying population. Estimates are often needed for such quantities as the magnitude of an extreme event (quantile) x_T, corresponding to a return period T. Evidently, the reliability of the estimates depends largely on the quality of the data as well as the length of the period of record.

The aim of this paper is to discriminate between normal and Gumbel distributions. These two distributions are widely applied in engineering and often used as a model for hydrologic data sets. Some of its recent application areas include flood frequency analysis, network and software reliability engineering, nuclear engineering and epidemic modeling.

There are many practical applications where Gumbel and normal distributions are similar in appearance and the two distributions cannot be distinguished from one another. Normal and Gumbel distributions belong to the location scale family. Discriminating between any two general probability distribution functions from the location scale family was widely investigated in the literature. See, for instance, [1], [2], [4], [5], [6], [7], [15] and [9] who studied the discrimination problem in general between the two models. Besides, [16], [19] and [22] studied the discrimination problem between lognormal and gamma distributions. [3] and [10] studied the discrimination problem between Weibull and gamma distributions. Recently, Gupta and Kundu considered the discrimination problem between Weibull and generalized exponential distributions, between gamma and generalized exponential distributions and between lognormal and generalized exponential distributions (see, [12], [13], [18]).

Among the discrimination problems, the one for Weibull and lognormal distributions is particularly important and has received much attention; this is because the two distributions are the most popular ones for analyzing the lifetime of electronic products. [8] adopted the ratio of maximized likelihood (RML) in discriminating between the two distributions for complete data and provided the percentile points for some sample sizes by simulation. Recently, [17] considered the discrimination problem for complete data using the RML procedure.

In the present work, to discriminate between normal and Gumbel distributions, we consider the ratio of maximized likelihood (RML) as test statistic. Based on the result of [21], the asymptotic distribution of the logarithm of the RML is found under both normal and Gumbel distributions, which can be used to compute the probability of correct selection (PCS). For small sample size, maximum likelihood estimators (MLE) of Gumbel parameters are biased; henceforth, we will use a correction for the bias introduced by [11] and [14].

The rest of the paper is organized as follows. Section 2 is dedicated to the mathematical notations that we use in this paper. In Section 3 we describe the logarithm of RML as test statistic, their asymptotic distributions under both normal and Gumbel distributions are obtained. Monte Carlo simulations are presented in Section 4 to examine how the asymptotic results work for finite samples. Finally, we conclude the paper in Section 5.

2. NOTATION

To facilitate the analysis that follows, we use the following notations. A normal distribution with mean μ and variance σ^2, denoted by $N(\mu, \sigma^2)$, has a probability density function (pdf) given by

$$f_N(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(x-\mu)^2}{2\sigma^2}, \; x \in \mathbb{R}.$$

The maximum likelihood estimators of μ and σ^2 are respectively given by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i := \bar{X} \text{ and } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2. \tag{2.1}$$

A Gumbel distribution with location parameter α and scale parameter β, denoted by $G(\alpha, \beta)$, has a pdf given by

$$f_G(x, \alpha, \beta) = \frac{1}{\beta} \exp\left[-\frac{x-\alpha}{\beta} - \exp -\frac{x-\alpha}{\beta}\right], \; x \in \mathbb{R},$$

and the maximum likelihood estimators of its parameters satisfy the following equations

$$\hat{\beta} = \bar{X} - \frac{\sum_{i=1}^{n} X_i \exp -\frac{X_i}{\hat{\beta}}}{\sum_{i=1}^{n} \exp -\frac{X_i}{\hat{\beta}}} \text{ and } \hat{\alpha} = -\hat{\beta} \ln\left[\frac{1}{n}\sum_{i=1}^{n} \exp -\frac{X_i}{\hat{\beta}}\right]. \tag{2.2}$$

These estimators, obtained as numerical solutions to the above equations, are known to be biased when the sample size is small. [11] proposed a correction for that bias:

$$\hat{\beta}_c^* = \frac{\hat{\beta}}{1 - 0.8/n} \text{ and } \hat{\alpha}_c^* = -\hat{\beta}_c \ln \left[\frac{1}{n} \sum_{i=1}^{n} \exp - \frac{X_i}{\hat{\beta}_c} \right] - 0.7 \frac{\hat{\beta}_c}{n}.$$

Using a rather theoretical analysis, [14] made more accurate corrections leading to the following estimators:

$$\hat{\beta}_c = \hat{\beta} \left(1 + \frac{0.7716}{n} \right) \text{ and } \hat{\alpha}_c = -\hat{\beta}_c \ln \left[\frac{1}{n} \sum_{i=1}^{n} \exp - \frac{X_i}{\hat{\beta}_c} \right] - 0.3698 \frac{\hat{\beta}_c}{n}.$$

It is to be noted that in instances when a non-negative random variable is needed, it is the discrimination between the lognormal and the Weibull distributions that might be of interest, but in such a case the results of the present study remain applicable because the normal and the lognormal (also the Gumbel and the Weibull distributions) are linked by a simple logarithmic transformation. The discrimination between lognormal and Weibull has been proposed by [17].

3. THE TEST STATISTIC AND ITS ASYMPTOTIC DISTRIBUTION

Assume that the random sample X_1,...,X_n is known to come from either a normal distribution, $X \sim N(\mu, \sigma^2)$, or a Gumbel distribution, $X \sim G(\alpha, \beta)$. The log-likelihood ratio statistic, T, is defined as the logarithm of the ratio of two maximized likelihood functions:

$$T = \ln \left(\frac{L_N(\hat{\mu}, \hat{\sigma}^2)}{L_G(\hat{\alpha}, \hat{\beta})} \right)$$

where $L_N(\mu, \sigma^2)$ and $L_G(\alpha, \beta)$ the likelihood functions under a normal distribution and a Gumbel distribution, respectively. The decision rule for discriminating between the normal and the Gumbel distributions is to choose the normal if $T > 0$, and to reject the normal in favor of the Gumbel, otherwise. Because both of these two distributions are of the location scale type, one important property of the T statistic is that it is independent of the parameters from both distributions (see, [8]).

Let us look at the expressions of T in terms of the corresponding MLEs.

Note that

$$\begin{aligned} T &= \ln L_N(\hat{\mu}, \hat{\sigma}^2) - \ln L_G(\hat{\alpha}, \hat{\beta}) \\ &= \left[-n \ln \hat{\sigma} - n \ln \sqrt{2\pi} - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n} (X_i - \hat{\mu})^2 \right] \\ &\quad - \left[-n \ln \hat{\beta} - \sum_{i=1}^{n} \left[\frac{X_i - \hat{\alpha}}{\hat{\beta}} + \exp - \frac{X_i - \hat{\alpha}}{\hat{\beta}} \right] \right] \\ &= -n \ln \hat{\sigma} - n \ln \sqrt{2\pi} - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n} (X_i - \hat{\mu})^2 + n \ln \hat{\beta} \\ &\quad + \frac{1}{\hat{\beta}} \sum_{i=1}^{n} X_i - \frac{n\hat{\alpha}}{\hat{\beta}} + \sum_{i=1}^{n} \exp - \frac{X_i}{\hat{\beta}} \exp \frac{\hat{\alpha}}{\hat{\beta}}. \end{aligned} \tag{3.1}$$

Using (2.2), we get

$$n \exp \frac{-\hat{\alpha}}{\hat{\beta}} = \sum_{i=1}^{n} \exp - \frac{X_i}{\hat{\beta}}. \tag{3.2}$$

If we replace the MLE finding in the equations (2.1) and the last equation (3.2) in (3.1), we obtain

$$T = -n \ln \frac{\hat{\sigma}}{\hat{\beta}} + n \frac{\hat{\mu} - \hat{\alpha}}{\hat{\beta}} + \frac{n}{2}(1 - \ln 2\pi).$$

We denote T_c, the new test statistic which introduces a correction for bias of maximum likelihood estimators proposed by [14]. Therefore, T_c can be written as:

$$T_c = -n \ln \frac{\hat{\sigma}}{\hat{\beta}_c} + n \frac{\hat{\mu} - \hat{\alpha}_c}{\hat{\beta}_c} + \frac{n}{2}(1 - \ln 2\pi).$$

Note that T and T_c are asymptotically equivalent, then we state the following lemma:

Lemma 3.1. *The test statistics T and T_c have the same asymptotic distribution.*

Proof: We have $\dfrac{T_c}{n} = -\ln \dfrac{\hat{\sigma}}{\hat{\beta}_c} + \dfrac{\hat{\mu} - \hat{\alpha}_c}{\hat{\beta}_c} + \dfrac{1}{2}(1 - \ln 2\pi)$ and

$$-\ln \frac{\hat{\sigma}}{\hat{\beta}_c} = -\ln \frac{\hat{\sigma}}{\hat{\beta}} + \ln \left(1 + \frac{0.7716}{n} \right) = -\ln \frac{\hat{\sigma}}{\hat{\beta}} + o(1). \tag{3.3}$$

In addition, $\hat{\beta}_c = \hat{\beta} + o(1)$ and $\hat{\alpha}_c = \hat{\alpha} + o_p(1)$ lead to

$$\frac{\hat{\mu} - \hat{\alpha}_c}{\hat{\beta}_c} = \frac{\hat{\mu} - \hat{\alpha}}{\hat{\beta}} + o_p(1). \tag{3.4}$$

From (3.3) and (3.4) we obtain $\frac{T_c}{n} = \frac{T}{n} + o_p(1)$. Then $\frac{T_c}{n}$ and $\frac{T}{n}$ have the same limit distribution, thus for $\epsilon > 0$ and n sufficiently large, we have

$$\left| P\left[\frac{T_c}{n} < \frac{t}{n}\right] - P\left[\frac{T}{n} < \frac{t}{n}\right] \right| < \epsilon.$$

Immediately $|P\,[T_c < t] - P\,[T < t]| < \epsilon$, for $\epsilon > 0$ and n is sufficiently large. Finally, if $\lim\limits_{n\to+\infty} P[T < t]$ exists, then $\lim\limits_{n\to+\infty} P[T_c < t] = \lim\limits_{n\to+\infty} P[T < t]$. □

3.1. Asymptotic distribution of T_c under the normal distribution

Suppose data are coming from a normal distribution $N(\mu, \sigma^2)$. Based on [18], the following theorem can be stated:

Theorem 3.1. *Assume that the sample $X_1,...,X_n$ follows $N(\mu, \sigma^2)$, then the test statistic T_c is asymptotically normally distributed with mean $E_N(T)$ and variance $Var_N(T)$.*

Proof: The proof of this theorem is based on the Lemma 3.1, the following Lemma 3.2 and the Central Limit Theorem (CLT). □

Lemma 3.2. *Denote $\tilde{T} = \ln\left(\dfrac{L_N(\mu, \sigma^2)}{L_G(\tilde{\alpha}, \tilde{\beta})}\right)$, where $\tilde{\alpha}$ and $\tilde{\beta}$ are given by the following equation and may depend on μ and σ,*

$$E_N[\ln f_G(X, \tilde{\alpha}, \tilde{\beta})] = \max_{\alpha,\beta} E_N[\ln f_G(X, \alpha, \beta)],$$

then $\hat{\alpha} \to \tilde{\alpha}$ a.s, $\hat{\beta} \to \tilde{\beta}$ a.s and $\frac{T - E_N(T)}{\sqrt{n}}$ is asymptotically equivalent to $\frac{\tilde{T} - E_N(\tilde{T})}{\sqrt{n}}$.

The proof of this lemma is similar to that of Theorem 1 presented by White in [21], then the proof of Theorem 3.1 is established by proving that $\frac{\tilde{T} - E_N(\tilde{T})}{\sqrt{n}}$ is asymptotically normal based on the central limit theorem. As for the needed quantities $\widetilde{\alpha}$ and $\widetilde{\beta}$ in Lemma 3.2, $E_N(T)$ and variance $Var_N(T)$ in Theorem 3.1, they are derived by first referring to Lemma 3.2 and performing the following calculation:

$$\begin{aligned} E_N[\ln f_G(X, \alpha, \beta)] &= -\ln\beta - E_N\left(\frac{X-\alpha}{\beta}\right) - E_N\left(\exp\left(-\frac{X-\alpha}{\beta}\right)\right) \\ &= -\ln\beta - \frac{\mu-\alpha}{\beta} - \exp\left(-\frac{\mu-\alpha}{\beta} + \frac{\sigma^2}{2\beta^2}\right). \end{aligned}$$

We maximize with respect to α and β, we get $\tilde{\alpha} = \mu - \frac{\sigma}{2}$ and $\tilde{\beta} = \sigma$. By the second point of Lemma 3.2, $E_N(T)$ and $Var_N(T)$ are calculated.

$$\begin{aligned}
E_N(T) &\simeq E_N\left[\ln\left(\frac{L_N(\mu,\sigma^2)}{L_G(\tilde{\alpha},\tilde{\beta})}\right)\right] \\
&= nE_N[\ln f_N(X,\mu,\sigma^2) - \ln f_G(X,\widetilde{\alpha},\widetilde{\beta})] \\
&= nE_N[\ln f_N(X,\mu,\sigma^2)] - nE_N[\ln f_G(X,\widetilde{\alpha},\widetilde{\beta})] \\
&= nE_N\left[-\ln\sigma - \ln\sqrt{2\pi} - \frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right] \\
&\quad -nE_N\left[-\ln\tilde{\beta} - \frac{X-\tilde{\alpha}}{\tilde{\beta}} - \exp - \frac{X-\tilde{\alpha}}{\tilde{\beta}}\right] \\
&= n\left(-\ln\sigma - \ln\sqrt{2\pi} - \frac{1}{2} - (-\ln\sigma - \frac{3}{2})\right) \\
&= n\left(1 - \ln\sqrt{2\pi}\right),
\end{aligned}$$

for n sufficiently large, we obtain

$$\lim_{n\to+\infty} \frac{E_N(T)}{n} = 0.081016.$$

In addition, $Var_N[\ln f_N(X,\mu,\sigma^2)] = Var_N\left[-\frac{1}{2\sigma^2}\left(X-\mu)^2\right)\right] = \frac{1}{2}$ and taking into account that $e^{-\frac{1}{2}}\int z^2e^{-z}\phi(z)dz = 2$ and $e^{-\frac{1}{2}}\int ze^{-z}\phi(z)dz = -1$ where $\phi(.)$ is the standard normal probability density function, then we have

$$\begin{aligned}
Var_N\left[\ln f_G(X,\widetilde{\alpha},\widetilde{\beta})\right] &= Var_N\left[-\frac{X-\widetilde{\alpha}}{\widetilde{\beta}} - \exp - \frac{X-\tilde{\alpha}}{\widetilde{\beta}}\right] \\
&= Var_N\left(\frac{X-\mu}{\sigma}\right) + Var_N\left[e^{-\frac{1}{2}}\exp\left(-\frac{X-\mu}{\sigma}\right)\right] \\
&\quad +2e^{-\frac{1}{2}}Cov_N\left[\frac{X-\mu}{\sigma}; \exp - \frac{X-\mu}{\sigma}\right] \\
&= e-2
\end{aligned}$$

and

$$\begin{aligned}
&Cov_N\left[\ln f_N(X,\mu,\sigma^2), \ln f_G(X,\tilde{\alpha},\tilde{\beta})\right] \\
&= \frac{1}{2}Cov_N\left[\left(\frac{X-\mu}{\sigma}\right)^2, \frac{X-\mu}{\sigma}\right] \\
&\quad +\frac{1}{2}e^{-\frac{1}{2}}Cov_N\left[\left(\frac{X-\mu}{\sigma}\right)^2, \exp\left(-\frac{X-\mu}{\sigma}\right)\right] \\
&= \frac{1}{2},
\end{aligned}$$

thus,

$$\begin{aligned}\frac{Var_N(T)}{n} &\simeq Var_N[\ln f_N(X,\mu,\sigma^2) + Var_N[\ln f_G(X,\tilde{\alpha},\tilde{\beta})] \\ &\quad -2Cov_N[\ln f_N(X,\mu,\sigma^2);\ln f_G(X,\tilde{\alpha},\tilde{\beta})] \\ &\simeq e - \frac{5}{2}.\end{aligned}$$

Then

$$\lim_{n\to+\infty}\frac{Var_N(T)}{n} = 0.218282.$$

Finally, $\lim\limits_{n\to+\infty}\frac{E_N(T)}{n}$ and $\lim\limits_{n\to+\infty}\frac{Var_N(T)}{n}$ are independent of μ and σ, then the asymptotic distribution of T is independent of μ and σ. Then from Theorem 3.1, the test statistic T_c is asymptotically normally distributed with mean $0.081016\times n$ and variance $0.218282 \times n$.

3.2. Asymptotic distribution of T_c under the Gumbel distribution

Now we turn to the case where the sample comes from a Gumbel distribution $G(\alpha,\beta)$. As before, based on Kundu, Gupta, and Manglick [18], the following theorem can be stated:

Theorem 3.2. *We suppose that the sample $X_1,...,X_n$ follows $G(\alpha,\beta)$, then the test statistic T_c is asymptotically normally distributed with mean $E_G(T)$ and variance $Var_G(T)$.*

Once again, the proof of this theorem is straightforward from the central limit theorem and the following lemma.

Lemma 3.3. *Denote $\tilde{T}' = \ln\left(\frac{L_N(\tilde{\mu},\tilde{\sigma}^2)}{L_G(\alpha,\beta)}\right)$, where $\tilde{\mu}$ and $\tilde{\sigma}$ are given by the following equation and may depend on α and β:*

$$E_N[\ln f_G(X,\tilde{\mu},\tilde{\sigma}^2)] = \max_{\mu,\sigma} E_N[\ln f_G(X,\mu,\sigma^2)]$$

then $\hat{\mu}\to\tilde{\mu}$ a.s, $\hat{\sigma}\to\tilde{\sigma}$ a.s and $\frac{T-E_G(T)}{\sqrt{n}}$ is asymptotically equivalent to $\frac{\tilde{T}'-E_G(\tilde{T}')}{\sqrt{n}}$.

It is now possible to evaluate $\widetilde{\mu}$ and $\widetilde{\sigma}$ by referring to Lemma 3.3 and performing the following calculation:

$$E_G[\ln f_N(X,\mu,\sigma^2)] = E_G\left[-\frac{1}{2}\ln 2\pi - \ln\sigma - \frac{(X-\mu)^2}{2\sigma^2}\right].$$

Since X follows $G(\alpha,\beta)$, it is immediate that $E_G(X)=\alpha+\beta\gamma$ and $Var_G(X)=\frac{\pi^2}{6}\beta^2$, where $\gamma\simeq 0.5772...$(the Euler constant). Therefore,

$$\begin{aligned}
E_G[\ln f_N(X,\mu,\sigma^2)] &= -\frac{1}{2}\ln 2\pi-\ln\sigma-\frac{1}{2\sigma^2}E_G(X^2-2\mu X+\mu^2)\\
&= -\frac{1}{2}\ln 2\pi-\ln\sigma\\
&\quad -\frac{1}{2\sigma^2}\left[\frac{\pi^2\beta^2}{6}+(\alpha+\beta\gamma)^2-2\mu(\alpha+\beta\gamma)+\mu^2\right].
\end{aligned}$$

Maximizing with respect to μ and σ yields $\tilde\mu=\alpha+\beta\gamma$ and $\tilde\sigma=\frac{\pi}{\sqrt6}\beta$. The quantities $E_G(T)$ and $Var_G(T)$ can be derived using again Lemma 3.3,

$$\begin{aligned}
E_G(T) &\simeq nE_G\left[\ln f_N(X,\tilde\mu,\tilde\sigma^2)-\ln f_G(X,\alpha,\beta)\right]\\
&\simeq nE_G\left[-\ln\tilde\sigma-\ln\sqrt{2\pi}-\frac{1}{2}\left(\frac{X-\tilde\mu}{\tilde\sigma}\right)^2\right]\\
&\quad +nE_G\left[\ln\beta+\frac{X-\alpha}{\beta}+\exp-\frac{X-\alpha}{\beta}\right]\\
&\simeq nE_G\left[-\ln\frac{\pi\beta}{\sqrt6}-\frac{1}{2}\ln 2\pi-\frac{1}{2}\left(\frac{X-(\alpha+\beta\gamma)}{\frac{\pi\beta}{\sqrt6}}\right)^2\right]\\
&\quad +nE_G\left[\ln\beta+\frac{X-\alpha}{\beta}+\exp-\frac{X-\alpha}{\beta}\right]\\
&\simeq n\left(-\frac{3}{2}\ln\pi+\frac{1}{2}\ln 3\right)\\
&\quad +nE_G\left[-\frac{3}{\pi^2}\left(\frac{X-\alpha}{\beta}-\gamma\right)^2+\frac{X-\alpha}{\beta}+e^{-\frac{X-\alpha}{\beta}}\right]
\end{aligned}$$

we put $Z=\dfrac{X-\alpha}{\beta}$, then Z follows $G(0,1)$ and we obtain

$$\begin{aligned}
E_G(T) &\simeq -\frac{3n}{2}\ln\pi+\frac{n}{2}\ln 3+nE_G\left[-\frac{3}{\pi^2}(Z-\gamma)^2+Z+\exp-Z\right]\\
&\simeq -\frac{3n}{2}\ln\pi+\frac{n}{2}\ln 3-\frac{3n}{\pi^2}E_G[(Z-\gamma)^2]+E_G[Z]+E_G[\exp-Z]\\
&\simeq n\left(-\frac{3}{2}\ln\pi+\frac{1}{2}\ln 3-\frac{3}{\pi^2}\frac{\pi^2}{6}+\gamma+1\right)
\end{aligned}$$

for n sufficiently large, we obtain

$$\lim_{n\to+\infty}\frac{E_G(T)}{n}=-0.090573.$$

Similarly,

$$\begin{aligned}\frac{Var_G(T)}{n} &\simeq Var_G\left[\ln f_N(X,\tilde{\mu},\tilde{\sigma}^2)-\ln f_G(X,\alpha,\beta)\right]\\&\simeq Var_G\left[-\ln\tilde{\sigma}-\ln\sqrt{2\pi}-\frac{1}{2}\left(\frac{X-\tilde{\mu}}{\tilde{\sigma}}\right)^2+\ln\beta+\frac{X-\alpha}{\beta}\right.\\&\qquad\left.+\exp-\frac{X-\alpha}{\beta}\right]\\&\simeq Var_G\left[-\frac{3}{\pi^2}\left(\frac{X-\alpha}{\beta}-\gamma\right)^2+\frac{X-\alpha}{\beta}+\exp-\frac{X-\alpha}{\beta}\right]\\&\simeq Var_G\left[-\frac{3}{\pi^2}(Z-\gamma)^2+Z+\exp-Z\right],\end{aligned}$$

then

$$\lim_{n\to+\infty}\frac{Var_G(T)}{n}=0.283408.$$

Since both $\lim\limits_{n\to+\infty}\frac{E_G(T)}{n}$ and $\lim\limits_{n\to+\infty}\frac{Var_G(T)}{n}$ do not depend on α and β, the asymptotic distribution of T is independent of α and β. Then from Theorem 3.2, the test statistic T_c is asymptotically normally distributed with mean $-0.090573\times n$ and variance $0.283408\times n$.

4. PCS AND MC SIMULATION

It is assumed that the data have been generated from one of the two distributions: $N(\mu,\sigma^2)$ or $G(\alpha,\beta)$. Then the discrimination procedure based on a random sample $X=X_1,...,X_n$ is as follows.

Choose normal distribution if $T_c>0$ and Gumbel distribution if $T_c<0$. If the data were originally coming from $N(\mu,\sigma^2)$, the PCS_N can be written as follows: $PCS_N=P(T_c>0|$ data follow a normal distribution). Similarly, if the data were originally coming from $G(\alpha,\beta)$, the PCS_N can be written as follows: $PCS_G=P(T_c<0|$ data follow Gumbel distribution). Since for normal distribution

$$\begin{aligned}PCS_N=P[T_c>0]\ &\simeq\ \Phi\left(\frac{E_N(T)}{\sqrt{Var_N(T)}}\right)\\&=\ \Phi\left(\frac{0.081016\times n}{\sqrt{0.218282\times n}}\right)\\&=\ \Phi(0.1734\sqrt{n})\end{aligned}$$

where Φ is the distribution function of the standard normal distribution. In the same manner, we have for Gumbel distribution

$$\begin{aligned} PCS_G = P[T_c < 0] &= 1 - P[T_c > 0] \\ &\simeq 1 - \Phi\left(\frac{E_G(T)}{\sqrt{Var_G(T)}}\right) \\ &= 1 - \Phi\left(\frac{-0.090573 \times n}{\sqrt{0.283408 \times n}}\right) \\ &= \Phi\left(\frac{0.090573 \times n}{\sqrt{0.283408 \times n}}\right) \\ &= \Phi(0.1701\sqrt{n}). \end{aligned}$$

We use Monte-Carlo simulations to examine how the asymptotic results work for small sizes. All computations are performed using the statistical freeware R [20]. We compute the PCS based on simulations and those based on the asymptotic normality results. Since the distribution of T_c is independent of the location and scale parameters, we take the location and scale parameters to be zero and one respectively in all cases. We consider different sample sizes, namely $n = 20$, 30, 40, 50, 60 and 100. First we consider the case when the data comes from normal distribution. In this case we generate a random sample of size n from $N(0,1)$, we compute T_c and check whether T_c is positive or negative. We replicate the process 10 000 times and obtain an estimate of PCS. Similarly, we obtain the results when the data comes from Gumbel distribution. The results are reported in Table 1.

Sample size (n)	**MC**	**Asymptotic results**
10	0.62 (0.70)	0.71 (0.70)
20	0.75 (0.79)	0.78 (0.77)
30	0.81 (0.84)	0.84 (0.82)
40	0.86 (0.88)	0.86 (0.85)
50	0.90 (0.91)	0.89 (0.88)
60	0.91 (0.92)	0.91 (0.90)
70	0.93 (0.94)	0.93 (0.92)
80	0.94 (0.95)	0.94 (0.94)
90	0.95 (0.96)	0.95 (0.95)
100	0.96 (0.97)	0.96 (0.95)

Table 1: PCS's based on Monte Carlo simulations (MC) with 10 000 replications and those based on the asymptotic results (AR) when the data come from the normal (Gumbel) distribution respectively.

The comparison between the MC simulation and the asymptotic results shows that the asymptotic approximation works quite well even for small samples. Results also reveal that it is easy to discriminate between normal and Gumbel distributions even for a small sample as 20. For example, the comparison of the results of Table 1 with those of Kundu and Manglick [17] shows that the selection between the normal and Gumbel distributions gives an asymptotic approximation more accurate even for a small sample size when the data comes from Gumbel distribution. Table 1 shows that the minimum sample size needed to choose between normal and Gumbel distributions is less than 50; it is also clear that the power of the test varies between 0.62 and 0.96 as the sample size varies between 10 and 100.

5. CONCLUSION

The normal and Gumbel distributions are often considered as competing models when the variable of interest takes values from $-\infty$ to $+\infty$. In this work we consider the statistic based on the RML and obtain asymptotic distributions of the test statistics under null hypothesis. Using MC simulations we compare the probability of correct selection with these asymptotic result and it is observed that even when the sample size is as small as 20, these asymptotic results work quite well for a wide range of the parameter space. Therefore, these asymptotic results can be used to estimate the PCS. Our method can be used for discriminating between any two members of the different location and scale families.

ACKNOWLEDGMENTS

The authors would like to thank the Reviewer for his valuable comments which had improved the earlier version of the paper and one Associate Editor for some very constructive suggestions.

REFERENCES

[1] Atkinson, A. (1969). A test for discriminating between models, *Biometrika*, **56**, 337–347.

[2] Atkinson, A. (1970). A method for discriminating between models (with discussions), *Journal of Royal Statistical Society, Series B*, **32**, 323–353.

[3] Bain, L.J., and Engelhardt, M. (1980). Probability of correct selection of Weibull versus Gamma based on likelihood ratio, *Statit.-Theor.Meth*, **9**, 375–381.

[4] Chambers, E.A., Cox, D.R. (1967). Discriminating between alternative binary response models, *Biometrika*, **54**, 573–578.

[5] Chen, W.W. (1980). On the tests of separate families of hypotheses with small sample size, *Journal of Statistical Computations and Simulations*, **2**, 183–187.

[6] Cox, D.R. (1961). Tests of separate families of hypotheses, *Proceedings of the Fourth Berkeley Symposium in Mathematical Statistics and Probability, Berkeley, University of California Press*, 105–123.

[7] Cox, D.R. (1962). Further results on tests of separate families of hypotheses, *J. Roy. Statist. Soc. Ser. B*, **24**, 406–424.

[8] Dumonceaux, R., Antle, C.E. and Haas, G. (1973). Likelihood ratio test for discriminating between two models with unknown location and scale parameters, *Technometrics*, **15**, 19–31.

[9] Dyer, A.R. (1973). Discriminating procedure for separate families of hypotheses, *Journal of American Statistics Association*, **68**, 970–974.

[10] Fearn, D.H., Nebenzahl, E. (1991). On the maximum likelihood ratio method of deciding between the Weibull and gamma distribution, *Communications in Statistics - Theory and Methods*, **20**, 579–593.

[11] Fiorentino, M. ,and Gabriele, S. (1984). A correction for the bias of maximum-likelihood estimators of Gumbel parameters, *Journal of Hydrology*, **73**, 39–49.

[12] Gupta, R.D., Kundu, D. (2003). Discriminating between the Weibull and generalized exponential distributions, *Computational Statistics and Data Analysis*, **43**, 179–196.

[13] Gupta, R.D., Kundu, D. (2004). Discriminating between the gamma and generalized exponential distributions, *Journal of Statistical Computations and Simulations*, **74**, 1, 107–121.

[14] Hosking, J.R.M. (1985). A correction for the bias of maximum-likelihood estimators of Gumbel parameters-comment, *Journal of Hydrology*, **78**, 393–396.

[15] Jackson, O.A.Y. (1968). Some results on tests of separate families of hypotheses, *Biometrika*, **55**, 355–363.

[16] Jackson, O.A.Y. (1969). Fitting a gamma or log-normal distributions to fibre-diameter measurements of wool tops, *Applied Statistics*, **18**, 70–75.

[17] Kundu, D., Manglick, A. (2004). Discriminating between the Weibull and lognormal distributions, *Naval Research Logistics*, **51**, 893–905.

[18] Kundu, D., Gupta, R. D., Manglick, A. (2005). Discriminating between the lognormal and generalized exponential distributions, *Journal of Statistical Planning and Inference*, **127**, 213–227.

[19] Quesenberry, C. P., Kent, J. (1982). Selecting among probability distributions used in reliability, *Technometric*, **24**, 59–65.

[20] R Development Core Team. (2009). R : A language and environment for statistical computing, *ISBN 3-900051-07-0, Vienna, Austria.*

[21] White, H. (1982). Regularity conditions for Cox's test of non-nested hypotheses, *Journal of Econometrics*, **19**, 301–318.

[22] Wiens, B. L. (1999). When lognormal and gamma models give different results: A case study, *American Statistics*, **53**, 89–93.