

Parameter Estimation for Geometric-Gumbel Compound Extreme-value Distribution based on the pi-th Quantiles of Samples

GUAN Qingsong,

School of Mathematics and Statistics, Wuhan University, Hubei 430072, China

PENG Wei

School of Mathematics and Statistics, Wuhan University, Hubei 430072, China

Abstract. In this paper, the pi-th quantiles of samples on parameter estimation were applied to geometric-Gumbel compound extreme-value distribution. The asymptotic normal estimation of distribution parameter and scale parameter was proposed by linear regression model, based on k pi-th quantiles of extreme-value distribution simple sample. The asymptotic confidence interval was given for extreme-value distribution. Then, the effect was evaluated by numerical simulation, the result was good.

Keywords: Geometric-Gumbel Compound Extreme-value Distribution; Sample Quantile; Asymptotic Confidence Interval

1. Introduction

Extreme-value event refers to an event of a very low probability to occur, but it will have a huge impact once it occurs. For examples, once-in-a-century earthquakes, huge typhoon, and broken high jump record are extreme-value events. In order to explore the occurrence law of extreme-value events and evaluate the risks of extreme events, the extreme-value theory was proposed.

With the gradual development of the extreme-value theory and more accurately study extreme-value events, many compound extreme-value distributions were raised by people. It was assumed that $\xi_1, \dots, \xi_n, \dots$ was independent identically distributed random variables sequences, and their common

distribution function was $G(x)$; N was a random variable independent of $\{\xi_n\}$, and its distribution law was $p_k = P(N = k), k = 1, 2, \dots$. $X = \max_{1 \leq i \leq N} \xi_i$ was ordered, and then the distribution function of X was

$$F(x) = P(X \leq x) = \sum_{k=1}^{\infty} P(\max_{1 \leq i \leq N} \xi_i \leq x | N = k) = \sum_{k=1}^{\infty} p_k G(x)^k$$

According to the total probability formula, in which $F(x)$ was an extreme-value distribution mixed by N and $\{\xi_n\}$. PENG Wei [1] assumed $G(x)$ was Gumbel distribution, N obeyed geometric distribution, and then Geometric-Gumbel compound extreme-value distribution was obtained and its distribution function is as follows:

$$F(x) = \sum_{k=1}^{\infty} p_k G(x)^k = \frac{pG(x)}{1 - (1-p)G(x)}, x \in \mathbb{R},$$

Density function is shown as below:

$$f(x) = f(x; \mu, \sigma, p) = -\frac{pG(x) \ln G(x)}{\sigma [1 - (1-p)G(x)]^2}, x \in \mathbb{R},$$

Where, $G(x) = \exp\{-\exp(-\frac{x-\mu}{\sigma})\}, x \in \mathbb{R}$, and

parameters $\mu \in \mathbb{R}, \sigma > 0, 0 < p < 1$. Then, three methods of estimating the unknown parameters were given, including the maximum likelihood method, composite moment method, probability moment method.

The above three methods of parameter estimation were provided in the case of unknown parameters μ, σ, p . In practice, p could be estimated according to the sample observation data of random variable N , and then μ, σ were

further estimated. CHENG Weihu [2] estimated the parameters of the double extreme-value distribution using sample fractiles. In this paper, the methods of estimating parameters μ, σ using sample fractiles were mainly discussed in the cause of know p .

2. Model and methods

2.1 The asymptotic theory of sample fractiles

It was assumed that X_1, X_2, \dots, X_n was the simple random sample sourcing from totality X , and the distribution function of X was $F(x)$.

Definition [3]: X_1, X_2, \dots, X_n was arranged as $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, from small to large, and $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ was the sequential statistical magnitude of samples. For any $0 < \alpha < 1$, $X_{([n\alpha])}$ was α fractile quantile of samples, in which $[n\alpha]$ was the largest integer no greater than $n\alpha$.

Theorem [3]: It was assumed that $F(x)$ was absolutely continuous; for $0 < \alpha_1 < \alpha_2 < \dots < \alpha_k < 1$, a_i was the lower α_i fractile quantile of totality, and y_i was the lower α_i fractile quantile of samples, and $i = 1, 2, \dots, k$. Order,

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix}$$

If $f(x) = F'(x)$, they were continuous and not equal to 0 at points a_1, a_2, \dots, a_k , and also there was $n \rightarrow \infty$, thus

$$\sqrt{n}(Y - A) \xrightarrow{L} N(O, \Lambda)$$

Where, $O' = (0, 0, \dots, 0)_{1 \times k}$, $\Lambda = (\lambda_{ij})_{k \times k}$, $\lambda_{ij} = \frac{\alpha_i(1 - \alpha_j)}{f(\alpha_i)f(\alpha_j)}$, $i \leq j$, and $\lambda_{ji} = \lambda_{ij}$.

2.2 The linear statistical model of sample fractiles

For Geometric-Gumbel distribution, it was easy to demonstrate distribution function

$$F(x) = \frac{pG(x)}{1 - (1-p)G(x)}$$

the given $0 < \alpha_1 < \alpha_2 < \dots < \alpha_k < 1$, the lower α_i fractile quantile of totality was $a_i = \mu - \sigma \ln[\ln(\frac{p}{\alpha_i} + 1 - p)]$,

and the lower α_i fractile quantile of samples was $y_i = X_{([n\alpha_i])}$, and also

$$f(x) = F'(x) = -\frac{pG(x) \ln G(x)}{\sigma[1 - (1-p)G(x)]^2}$$

equal to 0 at point a_i , thus meeting the theorem conditions. The following equation was further demonstrated.

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} = \begin{pmatrix} 1 & -\ln[\ln(\frac{p}{\alpha_1} + 1 - p)] \\ \vdots & \vdots \\ 1 & -\ln[\ln(\frac{p}{\alpha_k} + 1 - p)] \end{pmatrix} \begin{pmatrix} \mu \\ \sigma \end{pmatrix}$$

Order,

$$U = Y - A, X = \begin{pmatrix} 1 & -\ln[\ln(\frac{p}{\alpha_1} + 1 - p)] \\ \vdots & \vdots \\ 1 & -\ln[\ln(\frac{p}{\alpha_k} + 1 - p)] \end{pmatrix}, \beta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}$$

$Y = X\beta + U$, and $\sqrt{n}U \xrightarrow{L} N(O, \Lambda)$ could be obtained according to the theorem. Thus, a linear model was established.

2.3 The least square estimation

For the above linear model, the least squares estimation of β was shown as below.

$$\hat{\beta} = (X'X)^{-1} X'Y = \beta + (X'X)^{-1} X'U$$

Where, $(X'X)^{-1}$ was the generalized inverse of $X'X$. Here, $\hat{\beta} = (\hat{\mu}, \hat{\sigma})'$ was estimated as β . The asymptotic properties of β were discussed.

Lemma 1 [4]: $\{X_n\}$ was random variable; X was random vector; $X_n \xrightarrow{L} X$; A was appropriate constant matrix.

$$AX_n \xrightarrow{L} AX$$

According to the above formula, $\hat{\beta} - \beta = (XX)^- X'U$ was obtained. Then, according to $\sqrt{n}U \xrightarrow{L} N(O, \Lambda)$ and Lemma 1, the following formula was obtained.

$$\sqrt{n}(XX)^- X'U \xrightarrow{L} N(O, (XX)^- X' \Lambda X (XX)^-)$$

Thus,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} N(O, (XX)^- X' \Lambda X (XX)^-) \quad (1)$$

The asymptotic normality of $\hat{\beta}$ was obtained according to (1). After calculation, the following equation was obtained.

$$\hat{\mu} = \frac{(\sum_{i=1}^k t_i^2)(\sum_{i=1}^k y_i) - (\sum_{i=1}^k t_i)(\sum_{i=1}^k t_i y_i)}{k(\sum_{i=1}^k t_i^2) - (\sum_{i=1}^k t_i)^2}$$

$$\hat{\sigma} = \frac{k(\sum_{i=1}^k t_i y_i) - (\sum_{i=1}^k t_i)(\sum_{i=1}^k y_i)}{k(\sum_{i=1}^k t_i^2) - (\sum_{i=1}^k t_i)^2}$$

Where,

$$t_i = -\ln[\ln(\frac{p}{\alpha_i} + 1 - p)], y_i = X_{(n\alpha_i+1)}, i = 1, 2, \dots, k.$$

2.4 Asymptotic confidence interval

According to the above asymptotic normality of $\hat{\beta}$, the asymptotic confidence interval of μ and σ was obtained. The composition of the asymptotic variance $(XX)^- X' \Lambda X (XX)^-$ of $\hat{\beta}$ was first discussed here. According to $F(a_i) = \alpha_i$,

$$G(a_i) = \frac{\alpha_i}{p + (1-p)\alpha_i} \text{ was obtained, so } f(a_i) = \frac{M_i}{\sigma}$$

which M_i was the constant unrelated to μ, σ . Thus,

element $\lambda_{ij} = \frac{\sigma^2 \alpha_i (1 - \alpha_j)}{M_i M_j}$ in Λ was only related

to σ^2 . Then, $(XX)^- X' \Lambda X (XX)^- = \sigma^2 \Sigma$ was assumed, in which $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ was a matrix of constants

unrelated to μ, σ . Then,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} N(O, \sigma^2 \Sigma) \quad (2)$$

Thus,

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{L} N(0, \sigma^2 \sigma_{11}) \quad (3)$$

$$\sqrt{n}(\hat{\sigma} - \sigma) \xrightarrow{L} N(0, \sigma^2 \sigma_{22}) \quad (4)$$

According to (3) and (4), the following formula was obtained.

$$\hat{\mu} \xrightarrow{P} \mu, \hat{\sigma} \xrightarrow{P} \sigma \quad (5)$$

Lemma 2 [4]: It was assumed that $\{Y_n\}$ and $\{Z_n\}$ were two random variables sequences, in which Y was a random variable, c was a non-zero constant, and also $Y_n \xrightarrow{L} Y$ and $Z_n \xrightarrow{P} c$. Thus, $Y_n / Z_n \xrightarrow{L} Y / c$.

According to (5) and Lemma 2, the following formulas were obtained.

$$\frac{\sqrt{n}(\hat{\mu} - \mu)}{\hat{\sigma} \sqrt{\sigma_{11}}} \xrightarrow{L} N(0, 1), \quad (6)$$

$$\frac{\sqrt{n}(\hat{\sigma} - \sigma)}{\sigma \sqrt{\sigma_{22}}} \xrightarrow{L} N(0, 1). \quad (7)$$

A given level of significance was $0 < \alpha < 1$. Order $Z_{\frac{\alpha}{2}}$ was the upper $\frac{\alpha}{2}$ fractile point of the standard normal distribution. According to (6) and (7), $U(\mu) = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\hat{\sigma} \sqrt{\sigma_{11}}}$ and $U(\sigma) = \frac{\sqrt{n}(\hat{\sigma} - \sigma)}{\sigma \sqrt{\sigma_{22}}}$ were used respectively as pivots to obtain the following formulas.

The confidence coefficient of μ was the asymptotic confidence interval of $1-\alpha$:

$$\left[\hat{\mu} - \sqrt{\frac{\sigma_{11}}{n}} Z_{\frac{\alpha}{2}} \hat{\sigma}, \hat{\mu} + \sqrt{\frac{\sigma_{11}}{n}} Z_{\frac{\alpha}{2}} \hat{\sigma} \right]$$

The confidence coefficient of σ was the asymptotic confidence interval of $1-\alpha$:

$$\left[\frac{\sqrt{n}\hat{\sigma}}{\sqrt{n} + \sqrt{\sigma_{22}} Z_{\frac{\alpha}{2}}}, \frac{\sqrt{n}\hat{\sigma}}{\sqrt{n} - \sqrt{\sigma_{22}} Z_{\frac{\alpha}{2}}} \right]$$

2.5 Numerical simulation

To verify the rationality of the above parameter estimation, numerical simulation was conducted using Monte Carlo's simulation method. The independent identically distributed samples (n) were generated and then simulated for 7 times to obtain simulated values (T). In this paper, $p=0.4$ was chosen; the values of μ, σ were selected for β ; fractiles were 0.3, 0.4, 0.5, and 0.6; the significance level was $\alpha=0.05$; n were 50, 100, 500, and 1000 respectively; simulation ($T=2000$) was conducted to obtain a group of values for estimator $\hat{\beta}$.

The mean value of T was $MEAN = \frac{1}{T} \sum_{i=1}^T \hat{\beta}_i$; the deviation of samples was $BIA = \beta - MEAN$; the error of

mean square (mse) of samples was $MSE = \frac{1}{T} \sum_{i=1}^T (\hat{\beta}_i - \beta)^2$;

the variance of samples was $VAR = \frac{1}{T} \sum_{i=1}^T (\hat{\beta}_i - MEAN)^2$.

It was easy to prove that MSE, BIA, VAR met the relational expression ($MSE = BIA^2 + VAR$). The estimation method was more effective if the mse and variance of samples were smaller. Here, the coverage of the confidence interval was 0 introduced besides the accuracy of the estimated results was investigated from the perspective of the mse and variance of samples. A given significance level was α ; each parameter was corresponding to an asymptotic confidence interval; the true value of each simulation parameter fell inside or outside the confidence interval. For a large enough capacity of samples, the probability of the true value to fall within the internal was about $1-\alpha$. After the simulation of 7 times, the total times of the true value to fall within the confidence interval were N , and the estimation method was more reasonable if $\frac{N}{T}$ was

closer to $1-\alpha$. In table 1, the simulation results of the sample mean, error of mean square, standard deviation and interval coverage under the condition that sample capacity and parameter true-value were different. γ_1, γ_2 were the coverages of the asymptotic confidence intervals of μ, σ , respectively.

Table 1: The simulation under different sample capacity and parameter true-value

n	(μ, σ)	μ			σ			γ_1	γ_2
		MEAN	MSE	VAR	MEAN	MSE	VAR		
50	(0,0.5)	0.0228	0.0192	0.0187	0.4954	0.0131	0.0131	0.9170	0.9530
	(0,1.5)	0.0594	0.1753	0.1717	1.4926	0.1143	0.1143	0.9150	0.9530
	(0,3)	0.1332	0.7489	0.7312	2.9734	0.4973	0.4966	0.9125	0.9470
	(-1,1)	-0.9512	0.0794	0.0770	0.9939	0.0539	0.0538	0.9090	0.9495
	(1,1)	1.0538	0.0840	0.0812	0.9880	0.0515	0.0513	0.9020	0.9585
100	(0,0.5)	0.0125	0.0100	0.0098	0.4988	0.0066	0.0066	0.9325	0.9485
	(0,1.5)	0.0253	0.0824	0.0818	1.5065	0.0616	0.0616	0.9415	0.9475
	(0,3)	0.0682	0.3599	0.3552	2.9966	0.2310	0.2310	0.9285	0.9585
	(-1,1)	-0.9840	0.0390	0.0388	0.9953	0.0259	0.0258	0.9290	0.9530
	(1,1)	1.0252	0.0397	0.0390	0.9977	0.0267	0.0267	0.9235	0.9530
500	(0,0.5)	0.0036	0.0020	0.0020	0.4990	0.0014	0.0014	0.9440	0.9445
	(0,1.5)	0.0001	0.0177	0.0177	1.5013	0.0114	0.0114	0.9530	.09530
	(0,3)	0.0004	0.0720	0.0720	3.0092	0.0492	0.0492	0.9450	0.9570
	(-1,1)	-0.9970	0.0079	0.0079	1.0004	0.0051	0.0051	0.9460	0.9555
	(1,1)	1.0041	0.0080	0.0080	0.9987	0.0052	0.0052	0.9450	0.9515
1000	(0,0.5)	0.0004	0.0009	0.0010	0.5004	0.0007	0.0007	0.9455	0.9420
	(0,1.5)	0.0050	0.0085	0.0085	1.4987	0.0058	0.0058	0.9465	0.9525
	(0,3)	0.0040	0.0340	0.0340	3.0044	0.0236	0.0236	0.9530	0.9540
	(-1,1)	-0.9989	0.0042	0.0042	1.0008	0.0026	0.0027	0.9415	0.9505
	(1,1)	1.0029	0.0039	0.0039	0.9994	0.0026	0.0026	0.9440	0.9585

3 CONCLUSIONS

The simulation found that the mse and variance of the estimator were with good accuracy regardless of sample capacity. With the increase of sample capacity, the mean value was much closer to the true value, and mse and variance decreased. Seen from the interval coverage, the coverage of μ was slightly poor under the small sample capacity, but approximated to 0.95 with the increase of the sample capacity, suggesting the covering situation was better. Sample capacity did not greatly affect the coverage of σ , because σ always had a coverage of approximated 0.95.

References

- [1] Wei PENG, Xiaoxing LV, Lvqin LIU. Geometric-Gumbel Compound Extreme Value Distribution Parameter Estimation [J]. Journal of Statistics and Decision, 2015.
- [2] Weihu CHENG. The Parameter Estimation of Extreme Value Distribution Using Sample Fractiles [J]. Journal of Beijing University of Technology, 2002, 28.
- [3] Xiru CHEN. An Introduction to Mathematical Statistics [M]. Beijing: Science Press, 1981.
- [4] Ming ZHENG. Mathematical Statistics Teaching Materials [M]. Shanghai: Fudan University Press, 2005.