

Complementary Gamma Zero-Truncated Poisson Distribution and Its Application

Ausaina niyomdecha and Patchanok Srisuradetchai

Department of Mathematics and Statistics, Faculty of Science and Technology,
Thammasat University, Pathum Thani , Thailand

IOCMA2023
The 1st International Online Conference on Mathematics and Applications
1-15 May 2023

Outline

- Introduction
- The Complementary Gamma Zero-Truncated Poisson Distribution
- Properties of the Distribution
- Parameter Estimation
- Simulation Study
- Application

Introduction

- The gamma distribution is widely used in modeling lifetime data, but gamma distribution gave a monotone hazard function
- Many physical phenomena have non-monotone hazard functions, such as bathtub and upside-down bathtub hazard functions
- Some new distributions to model lifetime data have appeared by compounding existing lifetime models with several discrete distributions.

Introduction

- Suppose that N is discrete random variable
- X_1, X_2, X_3, \dots are iid random variables and also independent of N
- The pdf or pmf of

$$Z = g(X_1, X_2, \dots, X_N)$$

is called compound distribution

(e.g., $Z = \sum_{i=1}^N X_i$, $Z = \min\{X_1, X_2, \dots, X_N\}$, $Z = \max\{X_1, X_2, \dots, X_N\}$)

Complementary Gamma Zero-Truncated Poisson Distribution (CGZTP)

- Let X_1, X_2, \dots, X_N be iid random variables from gamma distribution

$$f(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad x > 0$$

- N is a random variable with a zero-truncated Poisson distribution and independent of X_i

$$P(N = n) = \frac{e^{-\lambda} \lambda^n}{n! (1 - e^{-\lambda})}, \quad n = 1, 2, \dots \text{ and } \lambda > 0,$$

- We define

$$Z = \max \{X_1, X_2, \dots, X_N\}$$



The distribution of Z is called the Complementary Gamma Zero-Truncated Poisson (CGZTP) distribution

Complementary Gamma Zero-Truncated Poisson Distribution (CGZTP)

- Then $g(z|n) = n[F(z)]^{n-1} f(z)$
- The joint distribution between Z and N : $g(z,n) = g(z|n)p(N=n)$

$$= \frac{\lambda e^{-\lambda} f(z)}{(1-e^{-\lambda})} \frac{[\lambda F(z)]^{n-1}}{(n-1)!}$$

- The marginal distribution for Z :

$$\begin{aligned} g(z; \alpha, \beta, \lambda) &= \sum_{n=1}^{\infty} g(z, n) \\ &= \sum_{n=1}^{\infty} \frac{\lambda e^{-\lambda} f(z)}{(1-e^{-\lambda})} \frac{[\lambda F(z)]^{n-1}}{(n-1)!} = \frac{\lambda e^{-\lambda} f(z)}{(1-e^{-\lambda})} e^{\lambda F(z)} \end{aligned}$$

- The pdf of CGZTP distribution is

$$g(z; \theta) = \frac{\lambda e^{-\lambda}}{(1-e^{-\lambda})} \left(\frac{\beta^\alpha z^{\alpha-1} e^{-\beta z}}{\Gamma(\alpha)} \right) e^{\lambda \left(1 - \frac{\Gamma(\alpha, \beta z)}{\Gamma(\alpha)} \right)}$$

Complementary Gamma Zero-Truncated Poisson Distribution (CGZTP)

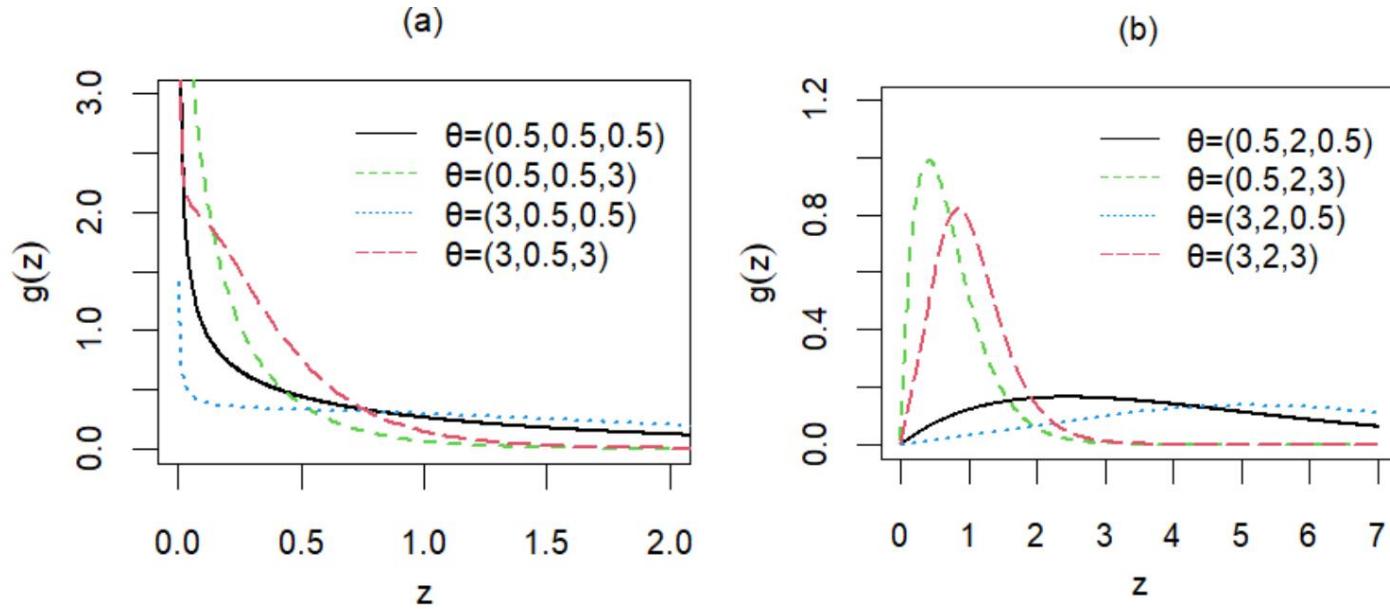


Figure 1. Probability density functions of the CGZTP distribution for (a) $\alpha = 0.5$ and (b) $\alpha = 2$.

CGZTP: Properties of the distribution

- The cumulative distribution function of the CGZTP distribution is given by

$$\begin{aligned} G(z; \theta) &= \int_0^z g(z) dz \\ &= \frac{\lambda \beta^\alpha}{(1 - e^{-\lambda}) \Gamma(\alpha)} \int_0^z z^{\alpha-1} e^{-\beta z - \lambda \frac{\Gamma(\alpha, z)}{\Gamma(\alpha)}} dz. \end{aligned}$$

$$\begin{aligned} \text{Since } \int z^{\alpha-1} e^{-\beta z - \lambda \frac{\Gamma(\alpha, z)}{\Gamma(\alpha)}} dz &= \frac{\Gamma(\alpha) e^{-\frac{\lambda \Gamma(\alpha, \beta z)}{\Gamma(\alpha)}}}{\lambda \beta^\alpha}, \quad G(z; \theta) = \left[\frac{\lambda \beta^\alpha}{(1 - e^{-\lambda}) \Gamma(\alpha)} \left(\frac{\Gamma(\alpha) e^{-\frac{\lambda \Gamma(\alpha, \beta z)}{\Gamma(\alpha)}}}{\lambda \beta^\alpha} + C \right) \right]_0^z \\ &= \frac{\left(e^{-\frac{\lambda \Gamma(\alpha, \beta z)}{\Gamma(\alpha)}} - e^{-\lambda} \right)}{(1 - e^{-\lambda})}. \end{aligned}$$

CGZTP: Properties of the distribution

- The r th quantile for this distribution is defined as the value z such that

$$\Gamma(\alpha, \beta z) = -\frac{\Gamma(\alpha)}{\lambda} \ln(r + (1-r)e^{-\lambda})$$

- The moment generating function

$$\begin{aligned} M_Z(t) &= E(e^{tZ}) = \int_0^{\infty} e^{tz} g(z; \theta) dz \\ &= \frac{\lambda \beta^\alpha}{\Gamma(\alpha)(1-e^{-\lambda})} \int_0^{\infty} z^{\alpha-1} e^{tz-\beta z-\lambda} \frac{\Gamma(\alpha, \beta z)}{\Gamma(\alpha)} dz \end{aligned}$$

- The k raw moments

$$E(Z^k) = \frac{\lambda \beta^\alpha}{\Gamma(\alpha)(1-e^{-\lambda})} \int_0^{\infty} z^{\alpha-1+k} e^{-\beta z-\lambda} \frac{\Gamma(\alpha, \beta z)}{\Gamma(\alpha)} dz , k \in \mathbb{N}$$

CGZTP: Properties of the distribution

- Survival function

$$\begin{aligned} S(z; \theta) &= 1 - G(z; \theta) \\ &= 1 - \frac{\left(e^{-\frac{\lambda \Gamma(\alpha, \beta z)}{\Gamma(\alpha)}} - e^{-\lambda} \right)}{\left(1 - e^{-\lambda} \right)} = \frac{\left(1 - e^{-\frac{\lambda \Gamma(\alpha, \beta y)}{\Gamma(\alpha)}} \right)}{\left(1 - e^{-\lambda} \right)} \end{aligned}$$

- Hazard function

$$\begin{aligned} H(z; \theta) &= \frac{g(z; \theta)}{s(z; \theta)} \\ &= \frac{\lambda \beta^\alpha z^{\alpha-1} e^{-\beta z - \frac{\lambda \Gamma(\alpha, \beta z)}{\Gamma(\alpha)}}}{\Gamma(\alpha) \left(1 - e^{-\frac{\lambda \Gamma(\alpha, \beta z)}{\Gamma(\alpha)}} \right)} \end{aligned}$$

CGZTP: Properties of the distribution

We define the function $\eta(z) = -\frac{g'(z; \theta)}{g(z; \theta)}$, then

$$\eta(z) = -\frac{1}{z} \left[\alpha - 1 - \beta z + \frac{\lambda(\beta z)^\alpha e^{-\beta z}}{\Gamma(\alpha)} \right]$$

and $\eta'(z) = \frac{1}{\Gamma(\alpha)z^2} \left[(\alpha - 1)\Gamma(\alpha) + \lambda(\beta z)^\alpha (\beta z - \alpha + 1)e^{-\beta z} \right].$

For $\alpha = 1$, $\eta'(z) < 0$ for all z . The CGZTP distribution has an increasing hazard function that follows from Glaser [14].

CGZTP: Properties of the distribution

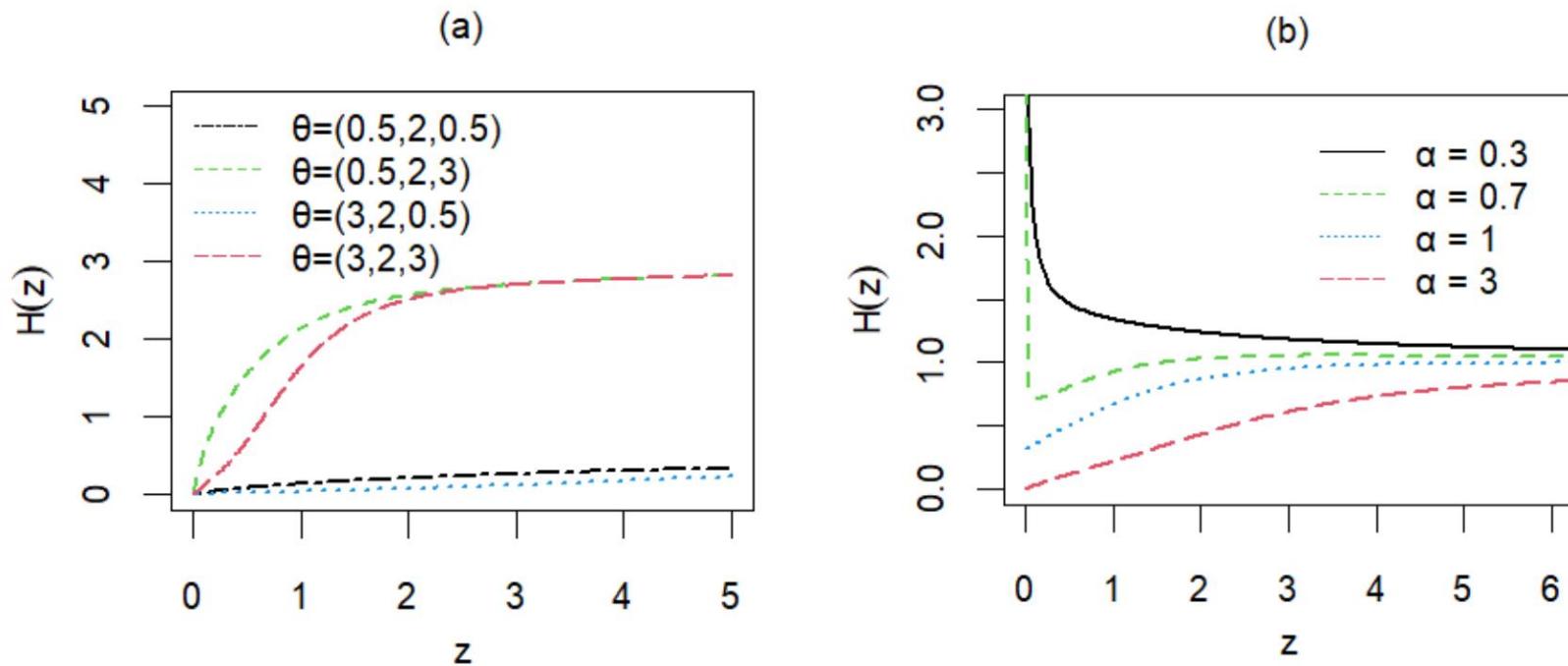


Figure 2. Hazard functions of the CGZTP distribution for (a) $\alpha = 2$ and (b) $\lambda = 2, \beta = 1$.

CGZTP: Parameter Estimation

□ Method of Maximum Likelihood

The likelihood function based on the observed random sample of size n , $w_{obs} = (z_1, z_2, \dots, z_n)$ is given by

$$L(\theta; w_{obs}) = \left(\frac{\lambda}{1 - e^{-\lambda}} \right)^n \left(\frac{\beta^{n\alpha}}{(\Gamma(\alpha))^n} \right) \left(\prod_{i=1}^n z_i \right)^{\alpha-1} e^{-\beta \left(\sum_{i=1}^n z_i \right) - \frac{\lambda}{\Gamma(\alpha)} \sum_{i=1}^n \Gamma(\alpha, \beta z_i)}.$$

The corresponding log-likelihood function is

$$\begin{aligned} l(\theta; w_{obs}) &= n \left(\log \lambda - \log (1 - e^{-\lambda}) \right) + n\alpha \log \beta - n \log \Gamma(\alpha) + (\alpha - 1) \sum_{i=1}^n \log z_i \\ &\quad - \beta \left(\sum_{i=1}^n z_i \right) - \lambda \sum_{i=1}^n \Gamma(\alpha, \beta z_i) / \Gamma(\alpha). \end{aligned}$$

CGZTP: Parameter Estimation

The first derivative of log-likelihood function are following:

$$\frac{\partial l(\boldsymbol{\theta}; w_{obs})}{\partial \lambda} = n \left(1/\lambda - e^{-\lambda} (1 - e^{-\lambda})^{-1} \right) - \sum_{i=1}^n \Gamma(\alpha, \beta z_i) / \Gamma(\alpha),$$

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta}; w_{obs})}{\partial \alpha} &= n \log \beta - n \psi_0(\alpha) + \sum_{i=1}^n \log z_i \\ &\quad - \lambda \sum_{i=1}^n \left[G_{2,3}^{3,0} \left(\beta z_i \middle| \begin{matrix} 1,1 \\ 0,0, \alpha \end{matrix} \right) + \Gamma(\alpha, \beta z_i) (\log(\beta z_i) - \psi_0(\alpha)) \right] / \Gamma(\alpha), \end{aligned}$$

$$\frac{\partial l(\boldsymbol{\theta}; w_{obs})}{\partial \beta} = \frac{n\alpha}{\beta} - \sum_{i=1}^n z_i + \frac{\lambda \beta^{\alpha-1}}{\Gamma(\alpha)} \sum_{i=1}^n z_i^\alpha e^{-\beta z_i}.$$

The last equation could be solved exactly for λ . Hence the MLE of λ is

$$\hat{\lambda} = \frac{\Gamma(\hat{\alpha})}{\hat{\beta}^{\hat{\alpha}-1} \sum_{i=1}^n z_i^{\hat{\alpha}} e^{-\hat{\beta} z_i}} \left(\sum_{i=1}^n z_i - \frac{n\hat{\alpha}}{\hat{\beta}} \right)$$

CGZTP: Parameter Estimation

Theorem 1 Let $l_1(\lambda; \alpha, \beta, w_{obs}) = \frac{\partial l(\theta; w_{obs})}{\partial \lambda}$ and α, β are known, then $\hat{\lambda}$ is the uniquely exist root of $l_1(\lambda; \alpha, \beta, w_{obs}) = 0$ if $\sum_{i=1}^n \Gamma(\alpha, \beta z_i) / \Gamma(\alpha) < \frac{n}{2}$

Proof. $\lim_{\lambda \rightarrow 0} l_1(\lambda; \alpha, \beta, w_{obs}) = \frac{n}{2} - \frac{\sum_{i=1}^n \Gamma(\alpha, \beta z_i)}{\Gamma(\alpha)} > 0$ if $\frac{\sum_{i=1}^n \Gamma(\alpha, \beta z_i)}{\Gamma(\alpha)} < \frac{n}{2}$.

$\lim_{\lambda \rightarrow \infty} l_1(\lambda; \alpha, \beta, w_{obs}) = -\frac{\sum_{i=1}^n \Gamma(\alpha, \beta z_i)}{\Gamma(\alpha)} < 0$ because $\frac{\Gamma(\alpha, \beta z_i)}{\Gamma(\alpha)} > 0$.

Therefore, there exist at least one solution of $l_1(\lambda; \alpha, \beta, w_{obs}) = 0$.

CGZTP: Parameter Estimation

Theorem 1 Let $l_1(\lambda; \alpha, \beta, w_{obs}) = \frac{\partial l(\theta; w_{obs})}{\partial \lambda}$ and α, β are known, then $\hat{\lambda}$ is the uniquely exist root of $l_1(\lambda; \alpha, \beta, w_{obs}) = 0$ if $\sum_{i=1}^n \Gamma(\alpha, \beta z_i) / \Gamma(\alpha) < \frac{n}{2}$

Proof. The first derivative of l_1 is considered and given by

$$l_1'(\lambda; \alpha, \beta, w_{obs}) = -\frac{ne^\lambda(e^{-\lambda} + e^\lambda - (\lambda^2 + 2))}{e^\lambda}$$

Consider $e^\lambda = 1 + \lambda + \frac{1}{2}\lambda^2 + \frac{1}{3!}\lambda^3 + \dots$ and $e^{-\lambda} = 1 - \lambda + \frac{1}{2}\lambda^2 - \frac{1}{3!}\lambda^3 + \dots$,

then $e^{-\lambda} + e^\lambda = 2 + \lambda^2 + \frac{2}{4!}\lambda^4 + \dots > \lambda^2 + 2$.

Therefore, $l_1'(\lambda; \alpha, \beta, w_{obs}) < 0$. This means l_1 is strictly decreasing function.

Then $\hat{\lambda}$ is the uniquely exist root of $l_1(\lambda; \alpha, \beta, w_{obs}) = 0$.

CGZTP: Parameter Estimation

□ Variance-Covariance Matrix of the MLEs

The MLE of $\boldsymbol{\theta}$ is approximately multivariate normal with a mean $\boldsymbol{\theta}$ and variance-covariance matrix which is the inverse of Fisher information matrix, i.e.,

$$\hat{\boldsymbol{\theta}} \sim N_3\left(\boldsymbol{\theta}, J(\hat{\boldsymbol{\theta}})^{-1}\right) \quad \text{or} \quad \hat{\boldsymbol{\theta}} \sim N_3\left(\boldsymbol{\theta}, I(\hat{\boldsymbol{\theta}})^{-1}\right),$$

where $J(\boldsymbol{\theta}) = E[I(\boldsymbol{\theta})]$ and $I(\boldsymbol{\theta})$ is the observed Fisher information matrix.

The asymptotic distribution of the i th component of $\hat{\boldsymbol{\theta}}$ is

$$\hat{\theta}_i \sim N\left(\theta_i, J^{ii}\right) \quad \text{or} \quad \hat{\theta}_i \sim N\left(\theta_i, I^{ii}\right),$$

where $J^{ii} = [J(\hat{\boldsymbol{\theta}})]_{ii}$ and $I^{ii} = [I(\hat{\boldsymbol{\theta}})]_{ii}$.

Then, the corresponding $(1 - \alpha)100\%$ Wald confidence intervals for $\hat{\theta}_i$ are

$$\hat{\theta}_i \pm z_{1-\alpha/2} \sqrt{J^{ii}} \quad \text{or} \quad \hat{\theta}_i \pm z_{1-\alpha/2} \sqrt{I^{ii}}.$$

CGZTP: Parameter Estimation

□ Variance-Covariance Matrix of the MLEs

The elements of the observed Fisher information matrix are found as follow:

$$I_{11} = ne^\lambda(e^{-\lambda} + e^\lambda - (\lambda^2 + 2))/e^\lambda,$$

$$I_{22} = n\psi^{(1)}(\alpha) + \lambda \sum_{i=1}^n \left[\frac{1}{\Gamma(\alpha)} \begin{aligned} & \left. 2G_{3,4}^{4,0} \left(\beta z_i \middle| \begin{matrix} 1,1,1 \\ 0,0,0,\alpha \end{matrix} \right) \right. \\ & + 2(\log(\beta z_i) - \psi^{(0)}(\alpha)) G_{2,3}^{3,0} \left(\beta z_i \middle| \begin{matrix} 1,1 \\ 0,0,\alpha \end{matrix} \right) \\ & \left. + \Gamma(\alpha, \beta z_i) (-2\psi^{(0)}(\alpha) \log(\beta z_i) + \psi^{(0)}(\alpha)^2 - \psi^{(1)}(\alpha) + \log^2(\beta z_i)) \right) \end{aligned} \right],$$

$$I_{33} = \frac{n\alpha}{\beta^2} - \frac{\lambda\beta^{\alpha-2}}{\Gamma(\alpha)} \sum_{i=1}^n z_i^\alpha e^{-\beta z_i} (\alpha - 1 - \beta z_i),$$

$$I_{12} = I_{21} = \left(1/\Gamma(\alpha)\right) \sum_{i=1}^n G_{2,3}^{3,0} \left(\beta z_i \middle| \begin{matrix} 1,1 \\ 0,0,a \end{matrix} \right) + \Gamma(\alpha, \beta z_i) (\log(\beta z_i) - \psi_0(\alpha)),$$

$$I_{13} = I_{31} = -\lambda\beta^{\alpha-1} \left(1/\Gamma(\alpha)\right) \sum_{i=1}^n z_i^\alpha e^{-\beta z_i},$$

$$I_{23} = I_{32} = -\frac{n}{\beta} - \lambda \sum_{i=1}^n e^{-\beta z_i} \frac{\partial}{\partial \alpha} \left[\frac{z_i^\alpha \beta^{\alpha-1}}{\Gamma(\alpha)} \right] = -\frac{n}{\beta} - \lambda \sum_{i=1}^n e^{-\beta z_i} \left[\frac{z_i^\alpha \beta^{\alpha-1} (-\psi^{(0)}(\alpha) + \log(\beta) + \log(z_i))}{\Gamma(\alpha)} \right].$$

CGZTP: Simulation Study

- Simulate 1,000 samples of sizes 50, 100, and 1,000
- Use the simulated annealing method to estimate parameters when all parameters are unknown
- To calculate the averages of MLEs and their MSEs
- To construct Wald confidence interval using the variance from observed Fisher information
- Use Monte Carlo simulations with 1,000 repetitions to estimate the coverage probability (CP) and average length (AL) of the confidence intervals

CGZTP: Simulation Study

Table 1. Mean estimates and mean-squared errors of λ , α , and β .

Distribution	n	Mean estimate			MSE		
		$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$
CGZTP(1,2,1)	50	1.7122	1.9225	1.0064	5.3198	0.4782	0.0519
	100	1.6464	1.8901	0.9864	4.5325	0.3659	0.0272
	1000	1.0225	2.0003	0.9965	0.3992	0.0523	0.0025
CGZTP(3,1,0.5)	50	2.2571	1.4061	0.5503	4.4487	0.5402	0.0193
	100	2.4622	1.2759	0.5294	3.4684	0.3293	0.0092
	1000	2.9382	1.0504	0.5049	0.8841	0.0597	0.0015
CGZTP(3,0.5,0.5)	50	2.3268	0.7150	0.5386	4.5848	0.1518	0.0165
	100	2.5139	0.6553	0.5265	3.6790	0.0981	0.0086
	1000	2.7430	0.5561	0.5098	0.7855	0.0159	0.0010

- As sample sizes increase, estimates become more accurate and MSE values decrease
- Among three estimates, β tends to have the smallest MSE

CGZTP: Simulation Study

Table 2. Coverage probabilities and average lengths of Wald CIs

$\theta = (\lambda, \alpha, \beta)$		n	CP	AL
$\lambda = 0.5, \beta = 3$	$\alpha = 1$	50	0.9130	1.4147
		100	0.9090	1.0495
		1,000	0.9530	0.3411
	$\alpha = 2$	50	0.9020	2.6400
		100	0.8880	1.9795
		1,000	0.9550	0.6626
$\alpha = 1, \beta = 3$	$\lambda = 0.5$	50	0.9800	6.0251
		100	0.9610	4.5462
		1,000	0.9520	1.3820
	$\lambda = 1$	50	0.9840	6.2418
		100	0.9580	5.2028
		1,000	0.9560	1.7814
$\lambda = 1, \alpha = 0.5$	$\beta = 0.5$	50	0.9670	0.5587
		100	0.9640	0.3832
		1,000	0.9470	0.1144
	$\beta = 1$	50	0.9710	1.0991
		100	0.9630	0.7861
		1,000	0.9530	0.2311

- In most cases, coverage probabilities are close to 0.95
- When sample size increases, the CPs will be close to the nominal coverage probability, 0.95, and the ALs will decrease.
- If λ has a small value, n is required to be 1,000 to achieve 0.95

CGZTP: Application

Illustrating example

- Dataset: The number of successive failures for the air conditioning system of each member in a fleet of 13 Boeing 720 jet airplanes was reported with 213 observations. The dataset is obtained from Proschan [15].
- The MLE of θ is $\hat{\theta} = (\hat{\lambda}, \hat{\alpha}, \hat{\beta}) = (0.11, 0.84, 0.01)$
- The Kolmogorov-Smirnov (K-S) statistic is 0.05611 with p-value of 0.5230
- The CGZTP distribution is useful for this data

References

1. Adamidis, K.; Loukas, S. A Lifetime Distribution with Decreasing Failure Rate. *Stat. Probab. Lett.* **1998**, *39*, 35–42, doi:10.1016/s0167-7152(98)00012-1.
2. Louzada, F.; Roman, M.; Cancho, V.G. The Complementary Exponential Geometric Distribution: Model, Properties, and a Comparison with Its Counterpart. *Comput. Stat. Data Anal.* **2011**, *55*, 2516–2524, doi:10.1016/j.csda.2011.02.018.
3. Barreto-Souza, W.; de Morais, A.L.; Cordeiro, G.M. The Weibull-Geometric Distribution. *J. Stat. Comput. Simul.* **2011**, *81*, 645–657, doi:10.1080/00949650903436554.
4. Tojeiro, C.; Louzada, F.; Roman, M.; Borges, P. The Complementary Weibull Geometric Distribution. *J. Stat. Comput. Simul.* **2014**, *84*, 1345–1362, doi:10.1080/00949655.2012.744406.
5. Zakerzadeh, H.; Mahmoudi, E. A New Two Parameter Lifetime Distribution: Model and Properties. *arXiv [stat.CO]* **2012**.
6. Gui, W.; Zhang, H.; Guo, L. The Complementary Lindley-Geometric Distribution and Its Application in Lifetime Analysis. *Sankhya B* **2017**, *79*, 316–335, doi:10.1007/s13571-017-0142-1.
7. Kuş, C. A New Lifetime Distribution. *Comput. Stat. Data Anal.* **2007**, *51*, 4497–4509, doi:10.1016/j.csda.2006.07.017.
8. Hemmati, F.; Khorram, E.; Rezakhah, S. A New Three-Parameter Ageing Distribution. *J. Stat. Plan. Inference* **2011**, *141*, 2266–2275, doi:10.1016/j.jspi.2011.01.007.
9. Lu, W.; Shi, D. A New Compounding Life Distribution: The Weibull–Poisson Distribution. *J. Appl. Stat.* **2012**, *39*, 21–38, doi:10.1080/02664763.2011.575126.
10. Ismail, E. The Complementary Compound Truncated Poisson-Weibull Distribution for Pricing Catastrophic Bonds for Extreme Earthquakes. *Br. J. Econ. Manag. Trade* **2016**, *14*, 1–9, doi:10.9734/bjemt/2016/26775.
11. Alkarni, S.; Oraby, A. A Compound Class of Poisson and Lifetime Distributions. *J. Stat. Appl. Probab.* **2012**, *1*, 45–51, doi:10.12785/jsap/010106.
12. Gui, W.; Zhang, S.; Lu, X. The Lindley-Poisson Distribution in Lifetime Analysis and Its Properties. *Hacet. J. Math. Stat.* **2014**, *43*, 1–1, doi:10.15672/hjms.201427453.
13. Cancho, V.G.; Louzada-Neto, F.; Barriga, G.D.C. The Poisson-Exponential Lifetime Distribution. *Comput. Stat. Data Anal.* **2011**, *55*, 677–686, doi:10.1016/j.csda.2010.05.033.
14. Glaser, R.E. Bathtub and Related Failure Rate Characterizations. *J. Am. Stat. Assoc.* **1980**, *75*, 667–672, doi:10.1080/01621459.1980.10477530.
15. Proschan, F. Theoretical Explanation of Observed Decreasing Failure Rate. *Technometrics* **2000**, *42*, 7, doi:10.2307/1271427.