Georgia Southern University

Digital Commons@Georgia Southern

Mathematical Sciences Faculty Publications

Mathematical Sciences, Department of

2007

A Note on Relations for Reliability Measures in Zero-Adjusted Models

Broderick O. Oluyede Georgia Southern University, boluyede@georgiasouthern.edu

Mavis Pararai Indiana University of Pennsylvania, mavis.pararai@iup.edu

Follow this and additional works at: https://digitalcommons.georgiasouthern.edu/math-sci-facpubs

Part of the Education Commons, and the Mathematics Commons

Recommended Citation

Oluyede, Broderick O., Mavis Pararai. 2007. "A Note on Relations for Reliability Measures in Zero-Adjusted Models." *Applied Mathematical Sciences*, 1 (56): 2789-2798. source: http://www.m-hikari.com/ams/ams-password-2007/ams-password53-56-2007/oluyedeAMS53-56-2007.pdf https://digitalcommons.georgiasouthern.edu/math-sci-facpubs/621

This article is brought to you for free and open access by the Mathematical Sciences, Department of at Digital Commons@Georgia Southern. It has been accepted for inclusion in Mathematical Sciences Faculty Publications by an authorized administrator of Digital Commons@Georgia Southern. For more information, please contact digitalcommons@georgiasouthern.edu.

Applied Mathematical Sciences, Vol. 1, 2007, no. 56, 2789 - 2798

A Note on Relations for Reliability Measures in Zero-Adjusted Models

Broderick O. Oluyede

Department of Mathematical Sciences Georgia Southern University, Statesboro, GA 30460, USA Boluyede@GeorgiaSouthern.edu

Mavis Pararai

Department of Mathematics Indiana University of Pennsylvania Indiana, PA 15705, USA

Abstract

In this note we examine and study relations in zero-adjusted models. Relations for reliability measures in the adjusted and unadjusted models are established and appropriate comparisons including the relative error are presented. The relative error is shown to be a decreasing function of the counts. Some inequalities and comparisons for weighted zeroadjusted models are established.

Mathematics Subject Classifications: 62N05, 60B10

Keywords: Weight functions, Zero-adjusted distribution, Survival function, Inequalities

1 Introduction

Statistical analysis for data that are often overdispersed are of tremendous practical importance. Mixed distributions including Poisson, negative binomial, Neyman Type A, to mention just a few distributions have been widely used to model overdispersed data. Areas of application include road safety (Miaou, 1994), patent application (Crepon and Duguet, 1997), medical consultations (Gurma, 1997), manufacturing defects (Lambert, 1992) to mention just a few. Lambert (1992) presented results on zero-inflated Poisson regression model. Heilborn (1994) discussed the zero altered Poisson and negative binomial regression models. Gupta et al (1996) investigated the zero-inflated form of the generalized Poisson distribution. Mullahy (1997) showed that the probability of a zero in the Poisson distribution is less than the probability of zero in mixed Poisson distribution with the same mean. Weighted distributions occur frequently in research related to reliability, bio-medicine, ecology, and several other areas. These distributions arise naturally as a result of observations generated from a stochastic process and recorded with some weight function. Length-biased distributions are weighted distributions (see Patil and Rao (1978)). A variety of methods have been developed for the estimation of the survival function, density function and other related functions under length-biased sampling. Weighted distributions in general and lengthbiased distributions in particular, are of tremendous practical importance in a wide variety of areas in probability and statistics. Length-biased sampling is widely used for the collection and analysis of wildlife data (Patil and Rao, 1978), fiber data (Daniels, 1942) or lifetime data (Zelen and Feinleib, 1969, Gupta and Keating, 1985).

In this paper, relations and stochastic inequalities for zero-adjusted models are established when the adjusted random variable is length-biased. Reliability measures of zero-adjusted and unadjusted models are compared and inequalities as well as bounds are presented. In section 2, models and utility notions are presented. In section 3, relations for reliability measures for zero-adjusted models are given. Section 4 deals with length-biased zero-adjusted and unadjusted models. Some relations and stochastic inequalities are presented. An application is presented in section 5. Concluding remarks are given in section 6.

2 Some Basic Results and Utility Notions

Let X be a non-negative random variable with distribution function F and probability density function (pdf) f. The weighted random variable X_W has a survival or reliability function given by

$$\overline{F}_W(x) = \frac{E_F[W(X)|X>x]}{E_F[W(X)]}\overline{F}(x).$$
(1)

Note that the survival or reliability function can also be expressed as

$$\overline{F}_W(x) = \overline{F}(x) \{ W(x) + M_F(x) \} / E(W(X)), \qquad (2)$$

where $M_F(x) = \int_x^\infty \{\overline{F}(t)W'(t)dt\}/\overline{F}(x)$, assuming $W(x)\overline{F}(x) \longrightarrow 0$ as $x \longrightarrow \infty$. The corresponding pdf of the weighted random variable X_W is

$$f_W(x) = W(x)f(x)/E(W(X)),$$
 (3)

 $x \ge 0$, where $0 < E(W(X) < \infty$. We give some basic and important definitions.

Definition 2.1 Let X and Y be two random variables with distribution functions F and G respectively. We say $F <_{st} G$, stochastically ordered, if $\overline{F}(x) \leq \overline{G}(x)$, for $x \geq 0$ or equivalently, for any increasing function $\Phi(x)$,

$$E(\Phi(X)) \le E(\Phi(Y)). \tag{4}$$

Definition 2.2 . A distribution function F is said to have increasing (decreasing) hazard rate or failure rate on $[0, \infty)$, denoted by IHR (DHR) or IFR (DFR), if F(0-) = 0, F(0) < 1 and $P(X > x + t|X > t) = \overline{F}(x+t)/\overline{F}(t)$ is decreasing (increasing) in $t \ge 0$ for each x > 0.

Let X be a discrete random variable, preferably discrete lifetime of a component and $f_X(k)$ the probability that failure will occur at time k. It is well known that the discrete failure rate function is given by

$$\gamma_X(k) = \frac{\overline{F}_X(k-1) - \overline{F}_X(k)}{\overline{F}_X(k)}, \qquad k = 1, 2, \dots$$
(5)

Roy and Gupta (1999), (see Xie et al (2002)) have shown that for discrete distributions, the definition of the hazard function $\gamma_X(k) = \frac{f_X(k)}{F(k)}$, the cumulative hazard function $H_X(k) = -ln(\overline{F}_X(k))$, $k \ge 1$, and related reliability functions can be problematic. This problem arises from the fact that the discrete failure rate function $\gamma_X(k) = \frac{\overline{F}_X(k-1)-\overline{F}_X(k)}{\overline{F}_X(k)}$ has the interpretation of a probability while it is known that the failure rate is not a probability in the continuous case. The discrete failure rate function is bounded and cannot be convex. These authors discussed other definitions of discrete failure rate and related functions that address this problem.

The corresponding cumulative hazard function and mean residual life functions are $~~\sim$

$$H_X(k) = \sum_{s=1}^{\infty} h_X(s), \tag{6}$$

and

$$M_X(k) = 1 + \sum_{s=k+1}^{\infty} exp(-\sum_{j=k+1}^{\infty} h_X(j))$$
(7)

respectively. We examine the definition of the discrete failure rate given by Roy and Gupta (1999) (see also Xie et al (2002)) and how it relates to the zero-adjusted and the length-biased adjusted models.

Definition 2.3 Let F_X be a discrete distribution function. The hazard rate function denoted by $h_X(x)$ is

$$h_X(k) = ln\left(\frac{\overline{F}_X(k-1)}{\overline{F}_X(k)}\right),$$

for $k \geq 1$.

Note that $h_X(k)$ is not bounded, and $\gamma_X(k) = 1 - exp\{-h_X(k)\}$, where $\gamma_X(k) = \frac{\overline{F}_X(k-1) - \overline{F}_X(k)}{\overline{F}_X(k)}$, is the traditional definition of the hazard rate function of a discrete random variable X.

Now we turn to the zero-inflated distribution. Suppose the discrete random variable X with mass concentrated on the integers is such that X = 0 is observed with lower (higher) probability than the assumed model, then the distribution of the adjusted random variable Y, for $0 \le \omega < 1$, is given by

$$P(Y=y) = \begin{cases} \omega + (1-\omega)P(X=0) & \text{if } y = 0, \\ (1-\omega)P(X=y) & \text{if } y > 0. \end{cases}$$

The distribution function of Y is given via the random variable X as

$$F_Y(k) = F_X(k) + \omega \overline{F}_X(k), \tag{8}$$

where $\overline{F}_X(k) = 1 - F_X(k), k > 0$, is the survival or reliability function of X.

Example: Ridout, Demetrio and Hinde (1998) presented a particular example in which the number of roots, Y, produced by a plant cutting during a period in a propagation environment has a zero-inflated Poisson (ZIP) distribution given by

$$P(Y = y) = \begin{cases} \omega + (1 - \omega)exp(-\lambda) & \text{if } y = 0, \\ (1 - \omega)exp(-\lambda)\lambda^y/y! & \text{if } y > 0. \end{cases}$$

In this case, the mixture distribution that arises when a proportion ω , $0 \leq \omega < 1$, of the cuttings are unable to root, is given by the ZIP distribution, where the remainder of the Poison parameter takes a value λ . For the ZIP distribution, the mean and variance are given by $E(Y) = (1-\omega)\lambda$ and $Var(Y) = \mu + \frac{\omega}{1-\omega}\mu^2$ respectively, where $\mu = (1-\omega)\lambda$.

3 Relations for Reliability Measures

In this section, we establish stochastic inequalities, bounds and relations for reliability measures for the zero-adjusted and unadjusted models. From equation (8), it follows immediately that

$$\overline{F}_Y(k) > \overline{F}_X(k),\tag{9}$$

if $\omega < 0$, and

$$\overline{F}_Y(k) \le \overline{F}_X(k),\tag{10}$$

if $\omega \geq 0$. Using the traditional definition of the hazard rate function of X we have $\gamma_F(k) = f_X(k)/\overline{F}_X(k) = \gamma_X(k), k \geq 1$, and

$$\gamma_Y(0) = \gamma_X(0) + \omega(1 - \gamma_X(0)).$$
 (11)

Clearly, $\gamma_Y(0) \ge \gamma_X(0)$ if $\omega \ge 0$ and $\gamma_Y(0) < \gamma_X(0)$ if $\omega < 0$. The corresponding survival function of Y is given by

$$\overline{F}_Y(k) = (1 - \omega)\overline{F}_X(k), \qquad (12)$$

 $k \geq 0$. The mean residual life function (MRLF) of Y is given by

$$M_Y(k) = (\overline{F}_Y(k))^{-1} \sum_{s=k}^{\infty} \overline{F}_Y(s)$$

= $((1-\omega)\overline{F}_X(k))^{-1} \sum_{s=k}^{\infty} (1-\omega)\overline{F}_X(s)$
= $(\overline{F}_X(k))^{-1} \sum_{s=k}^{\infty} \overline{F}_X(s)$
= $M_X(k),$ (13)

for $k \ge 1$. For k = 0, we have

$$M_Y(0) = (\overline{F}_Y(0))^{-1} \sum_{s=0}^{\infty} \overline{F}_Y(s) = \sum_{s=0}^{\infty} \overline{F}_Y(s) = (1-\omega)M_X(0),$$

where

$$M_X(0) = (\overline{F}_X(0))^{-1} \sum_{s=0}^{\infty} \overline{F}_X(s).$$
(14)

It follows therefore that $M_Y(0) \ge M_X(0)$ if $\omega \ge 0$, and $M_Y(0) < M_X(0)$ if $\omega < 0$.

Note that from definition 2.3, the hazard rate function of Y is

$$h_Y(k) = ln\left(\frac{(1-\omega)\overline{F}_X(k-1)}{(1-\omega)\overline{F}_X(k)}\right) = h_X(k),$$

and $\overline{F}_X(k) = \overline{F}_X(k-1)exp(-h_X(k), k \ge 1)$. For the zero-adjusted random variable Y, we have $\overline{F}_Y(k) = (1-\omega)\overline{F}_X(k-1)exp(-h_X(k))$, for k = 1, 2, The cumulative hazard rate function is given by

$$H_Y(k) = \sum_{s=1}^k h_Y(s) = \sum_{s=1}^k h_X(s) = H_X(k).$$
 (15)

4 Zero-Adjusted Length-Biased Distributions

In this section, some inequalities and reliability results for the zero-adjusted length-biased distributions and related measures are established. The lengthbiased probability function is a weighted probability density function with weight function W(k) = k. The corresponding length-biased reliability function is given by

$$\overline{F}_{X_l}(k) = \overline{F}_X(k) V_F(k) / \mu_F, \qquad (16)$$

where $V_F(k) = E(X|X > k)$ is the vitality function and $\mu_F = \mu = \sum_{k=0}^{\infty} \overline{F}(k)$. Note that $f_X(k)/f_{X_l}(k) = \mu_F/k \longrightarrow 0$ as $k \to \infty$. That is, the length-biased distribution F_{X_l} has a heavier tail than the original distribution F_X . Indeed

$$\overline{F}_{X_l}(k) \ge \overline{F}(k) \tag{17}$$

for all $k \ge 0$. The length-biased probability function is given by $f_{X_l}(k) = kf_X(k)/\mu$, for $k \ge 0$, where $\mu_F = \mu = \sum_{k=0}^{\infty} \overline{F}(k) < \infty$. The probability function for the adjusted random variable Y is given by

$$f_Y(k) = \begin{cases} \omega + (1 - \omega) f_{X_l}(0) & \text{if } k = 0, \\ (1 - \omega) f_{X_l}(k) & \text{if } k > 0. \end{cases}$$

The length-biased distribution function of the zero-adjusted random variable Y is given by

$$F_Y(k) = \omega + \frac{1-\omega}{\mu} \sum_{s=1}^k s P(X=s).$$
 (18)

The reliability or survival function of the adjusted random variable is

$$\overline{F}_Y(k) = (1-\omega)\overline{F}_X(k)(k+M_X(k))/\mu,$$
(19)

and the corresponding hazard rate function is given by

$$\gamma_Y(k) = \begin{cases} \frac{\omega \gamma_X(k)}{k + M_X(k)} & \text{if } k > 0, \\ \omega & \text{if } k = 0. \end{cases}$$

Using definition 2.3, the hazard rate function of Y is given by

$$h_Y(k) = ln\left(\frac{\overline{F}_Y(k-1)}{\overline{F}_Y(k)}\right)$$

$$= ln\left(\frac{(1-\omega)\overline{F}_X(k-1)(k-1+M_X(k-1))/\mu}{(1-\omega)\overline{F}_X(k)(k+M_X(k))/\mu}\right)$$

$$= ln\left(\frac{\overline{F}_X(k-1)(k-1+M_X(k-1))}{\overline{F}_X(k)(k+M_X(k))}\right)$$
(20)

$$= ln\left(\frac{\overline{F}_{X}(k-1)(k-1+\frac{\sum_{s=k-1}^{\infty}\overline{F}_{X}(s)}{\overline{F}(k-1)})}{\overline{F}_{X}(k)(k+\frac{\sum_{s=k}^{\infty}\overline{F}_{X}(s)}{\overline{F}_{X}(k)})}\right)$$

$$= ln\left(\frac{(k-1)\overline{F}_{X}(k-1)+\sum_{s=k-1}^{\infty}\overline{F}_{X}(s)}{k\overline{F}_{X}(k)+\sum_{s=k}^{\infty}\overline{F}_{X}(s)}\right)$$

$$= ln\left(\frac{(k-1)\overline{F}_{X}(k-1)+\overline{F}_{X}(k-1)+\sum_{s=k}^{\infty}\overline{F}_{X}(s)}{k\overline{F}_{X}(k)+\sum_{s=k}^{\infty}\overline{F}_{X}(s)}\right)$$

$$= ln\left(\frac{(k-1)\overline{F}_{X}(k-1)+\sum_{s=k}^{\infty}\overline{F}_{X}(s)}{k\overline{F}_{X}(k)+\sum_{s=k}^{\infty}\overline{F}_{X}(s)}\right)$$

$$= h_{X}(k), \qquad (21)$$

for $k \geq 1$. We have the following results:

Theorem 4.1 Let Y and X_l be the zero-adjusted and length-biased random variables respectively. The hazard rate functions $\gamma(k)$ and h(k) are affected only at the zero, that is

$$\gamma_Y(k) = \gamma_{X_l}(k) \qquad and \qquad h_Y(k) = h_{X_l}(k), \tag{22}$$

for $k \geq 1$.

Theorem 4.2 Let Y be the zero-adjusted random variable, then

$$\lim_{k \to \infty} \left(\frac{\overline{F}_Y(k)}{\overline{F}_X(k)} \right) = 1 - \omega.$$
(23)

Proof: Note that

$$\frac{\overline{F}_Y(k)}{\overline{F}_X(k)} = \frac{\overline{F}_X(k) - \omega \overline{F}_X(k)}{\overline{F}_X(k)}$$

Consequently for $0 \leq \omega < 1$, we have $\overline{\frac{F_Y(k)}{F_X(k)}} = 1 - \omega$.

Theorem 4.3 Let

$$\delta(k) = \left| \frac{\overline{F}_Y(k) - \overline{F}_X(k)}{\overline{F}_X(k)} \right|,\tag{24}$$

then $\delta(k+1) = \delta(k)$ for all $k \ge 0$. Also

$$\beta(k) = \left| \frac{\overline{F}_Y(k) - \overline{F}_X(k)}{F_X(k)} \right|,\tag{25}$$

is a decreasing function of k.

Proof: Note that

$$\frac{\overline{F}_Y(k) - \overline{F}_X(k)}{\overline{F}_X(k)} = \frac{\overline{F}_X(k) - \omega \overline{F}_X(k) - \overline{F}_X(k)}{\overline{F}_X(k)}.$$
(26)

The relative survival error is

$$\delta(k) = \left| \frac{\overline{F}_Y(k) - \overline{F}_X(k)}{\overline{F}_X(k)} \right| = |-\omega|, \qquad (27)$$

and $\delta(k+1) - \delta(k) = 0$. Similar considerations show that $\beta(k)$ is a decreasing function of k. See Gupta et al (1996).

5 Applications

Consider the zero-adjusted Poisson distribution presented earlier and given by

$$P(Y = y) = \begin{cases} \omega + (1 - \omega)exp(-\lambda) & \text{if } y = 0, \\ (1 - \omega)exp(-\lambda)\lambda^y/y! & \text{if } y > 0. \end{cases}$$

Using the following relation

$$\sum_{k=0}^{y} \frac{\lambda^k}{k!} = \frac{e^{\lambda}(1+y)\Gamma(1+y,\lambda)}{\Gamma(y+2)},$$
(28)

where $\Gamma(1+y,\lambda)$ is the incomplete gamma function. The corresponding cumulative distribution function is given by

$$F_Y(y) = \begin{cases} \omega + (1-\omega)exp(-\lambda) & \text{if } y = 0, \\ \omega + (1-\omega)exp(-\lambda)\left(\frac{exp(\lambda)(1+y)\Gamma(1+y,\lambda)}{\Gamma(y+2)} - 1\right) & \text{if } y \ge 1. \end{cases}$$

This reduces to

$$F_Y(y) = \begin{cases} \omega + (1-\omega)exp(-\lambda) & \text{if } y = 0, \\ \omega + \frac{(1-\omega)\Gamma(1+y,\lambda)}{\Gamma(y+1)} & \text{if } y \ge 1. \end{cases}$$

Note that $\overline{F}_Y(k) = (1-\omega) \frac{\Gamma(k+1) - \Gamma(k+1,\lambda)}{\Gamma(k+1)}$, for $k \ge 1$ so that

$$h_{Y}(k) = ln\left(\frac{\overline{F}_{Y}(k-1)}{\overline{F}_{Y}(k)}\right)$$

= $ln\left(\frac{(1-\omega)(\Gamma(k)-\Gamma(k,\lambda))}{\Gamma(k)}\cdot\frac{\Gamma(k+1)}{(1-\omega)(\Gamma(k)-\Gamma(k+1,\lambda))}\right)$
= $ln\left(\frac{k(\Gamma(k)-\Gamma(k,\lambda))}{\Gamma(k+1)-\Gamma(k+1,\lambda)}\right)$
= $h_{X}(k),$ (29)

for $k \geq 1$. Also,

$$\gamma_Y(k) = \frac{\lambda^k exp(-\lambda)}{\Gamma(k+1) - \Gamma(k+1,\lambda)} = \gamma_X(k), \tag{30}$$

for $k \ge 1$. The probability function for the length-biased zero-inflated Poisson (LBZIP) random variable Y is given by

$$f_Y(k) = \begin{cases} \omega & \text{if } k = 0, \\ (1 - \omega) \frac{exp(-\lambda)\lambda^{k-1}}{(k-1)!} & \text{if } k > 0. \end{cases}$$

Note that

$$\frac{f_Y(k+1)}{f_Y(k)} = \begin{cases} 1 & \text{if } k = 0, \\ \frac{\lambda}{k} & \text{if } k > 0. \end{cases}$$

Consequently, $f_Y(k)$ is log-concave, equivalently $\{\frac{f_Y(k+1)}{f_Y(k)}\}_{k\geq 1}$ is decreasing and hence IFR.

6 Concluding Remarks

In this paper, we obtained and presented results on reliability measures for weighted and unweighted zero-adjusted and unadjusted models. We examined the problem of interpretation of the hazard rate function that arises from the use of the traditional definition of hazard functions in the discrete distributions as it relates to the zero-adjusted and length-biased models. The definition of the hazard rate function that avoids interpretation as a probability is used, since it is known that the hazard rate function is not a probability. This definition is used to establish relations for various reliability measures including the mean residual life function in the zero-adjusted and unadjusted models. Applications of the results as it relates to weighted (length-biased) distributions are also presented.

References

- B. Crepon, and E. Duguet, Research and Development, Competition and Innovation - Pseudo-Maximum Likelihood and Simulated Maximum Likelihood Method Applied to Count Data Models with Heterogeneity, *Jour*nal of Econometrics, **79** (1984), 355-378.
- [2] H. E. Daniels, A New Technique for the Analysis of Fiber Length Distribution in Wool, Journal of Textile Institute, 33 (1942), 137-150.

- [3] R. C. Gupta, and J. P. Keating, relations for Reliability Measures Under Length-Biased Sampling, *Scandinavian Journal of Statistics*, **13** (1985), 49-56.
- [4] P. Gupta, R. Gupta, and R. Tripathi, Analysis of Zero-Adjusted Count Data, *Computational Statistics and Data Analysis*, **23** (1996), 207-218.
- [5] S. Gurmu, Semi-Parametric Estimation of Hurdle Regression Models with Applications to Medicaid Utilization, *Journal of Applied Econometrics*, 12 (1997), 225-242.
- [6] D. Heilbron, Zero-Altered and Other Regression Models for Count Data with Added Zeros, *Biometrical Journal*, 36 (1994), 531-547.
- [7] D. Lambert, Zero-Inflated Poisson Regression with Applications to Defects in Manufacturing, *Technometrics*, 43 (1992), 1-14.
- [8] S. P. Miaou, The Relations Between Truck Accidents and Geometric Designs of Road Sections. Poisson versus Negative binomial Regressions, *Accident Analysis and Prevention*, 26 (1994), 471-482.
- [9] J. Mullahy, Heterogeneity, Excess Zeros and the Structure of Count Data Models, Journal of Applied Econometrics, 12 (1997), 337-350.
- [10] G. P. Patil, and C. R. Rao, Weighted Distributions and Size-Biased Sampling with Applications to Wildlife and Human Families, *Biometrics*, 34 (1978), 179-189.
- [11] M. Ridout, C. Demetrio, and J. Hinde, Models for Count Data with Many Zeros, International Biometric Conference, IBC 98 (1998), 179-192.
- [12] D. Roy, and R. P. Gupta, Characterization and Model Selection Through Reliability Measures in the Discrete case, *Statistics and Probability Letters*, 43 (1999), 197-206.
- [13] Y. Vardi, Nonparametric Estimation in the Presence of Length Bias, Annals of Statistics, 10 (1969), 616-620.
- [14] Y. Vardi, Empirical Distributions in Selection Bias Models, Annals of Statistics, 13 (1985), 178-203.
- [15] M. Xie, O. Gaudion, and C. Bracquenomg, Redefining Failure Rate Function for Discrete Distributions, *International Journal of Reliability, Quality and Safety Engineering*, **19 (3)** (2002), 275-285.
- [16] M. Zelen, and M. Feinleib, On the Theory of Screening for Chronic Diseases, *Biometrika*, 56 (1969), 601-614.

Received: May 11, 2007