Applied Mathematics, 2012, 3, 2098-2100
http://dx.doi.org/10.4236/am.2012.312A289 Published Online December 2012 (http://www.SciRP.org/journal/am)

A Note on Generalized Inverses of Distribution Function and Quantile Transformation

Changyong Feng[1*], Hongyue Wang[1], Xin M. Tu[1], Jeanne Kowalski[2]
[1]Department of Biostatistics and Computational Biology, University of Rochester, Rochester, USA
[2]Department of Biostatistics & Bioinformatics, Rollins School of Public Health,
Winship Cancer Institute, Atlanta, USA
Email: *feng@bst.rochester.edu

Received September 7, 2012; revised October 7, 2012; accepted October 14, 2012

ABSTRACT

In this paper we study the relations of four possible generalized inverses of a general distribution functions and their right-continuity properties. We correct a right-continuity result of the generalized inverse used in statistical literature. We also prove the validity of a new generalized inverse which is always right-continuous.

Keywords: Distribution Function; Quantile Function; Right Continuity

1. Introduction

A *distribution function* defined on $\mathbb{R}$ is a map $F:\mathbb{R}\to\mathbb{R}$ such that F is nondecreasing and right continuous (see for example [1], p. 22). It is called a *probability distribution function* (PDF) if $\lim_{x\to\infty} F(x)=1$ and $\lim_{x\to-\infty} F(x)=1$. Suppose X is a random variable with distribution function F which is continuous, then $F(X)$ has standard uniform distribution. Furthermore, if F is also strictly increasing with inverse F^{-1}, and U is a standard uniform random variable, then $F^{-1}(U)$ has distribution function F. This fact is the basis for generating random numbers given a distribution function. Hence if F is a PDF, F_2^{-1} is also called the *quantile function* of F [2].

The distribution function dose not, in general, have an inverse (in strict sense) as it may be not strictly increasing, for example, the PDF of a discrete random variable. In statistics, the empirical distribution function (EDF) from a random sample is a step function. If we want to (nonparametrically) estimate the population quantiles from the sample data, we need to find an appropriate 'inverse' of the EDF. Unfortunately there is no universally accepted definition of sample quantiles given the data. For example, Langford [3] compares many methods proposed in literatures to calculate quantiles from data and finds that none of them is uniformly better than others.

There are four meaningful ways to define a *generalized inverse* of a distribution function F as follows:

$$F_1^{-1}(x)=\sup\{y|F(y)<x\},$$
$$F_2^{-1}(x)=\inf\{y|F(y)\ge x\},$$
$$F_3^{-1}(x)=\inf\{y|F(y)>x\},$$
$$F_4^{-1}(x)=\sup\{y|F(y)\le x\}.$$

It's easy to see that these methods are different in dealing with endpoints of flat parts if F is not strictly increasing. In fact, F_2^{-1} is the definition of generalized inverse in literatures, for example, [4-7]. It's obvious that all these generalized inverse functions are nondecreasing and equal the inverse of F if F is strictly monotone increasing.

In this manuscript we study some properties of these four functions and their relations to the *quantile transformation* in probability theory. The quantile transformation is the theoretical basis for random number generation in simulation studies in statistics. In simulation studies we usually need to generate random samples from a given distribution. The general method is first to generate uniform random variables on $[0,1]$ and then to use the quantile transformation to transform the uniform random variables to the random sample we need. Our results shows that any one of the generalized inverses defined above will work as the quantile transformation.

2. Relations between the Four Generalized Inverses

Four generalized inverse of a distribution function were introduced in last section. In this section we prove the following relations between them.

*Corresponding author.

Theorem 1. *The four generalized inverses of F defined above satisfy*

$$F_1^{-1}(x) = F_2^{-1}(x) \le F_3^{-1}(x) = F_4^{-1}(x), \text{ for all } x \in \mathbb{R}.$$

Proof. For any $x \in \mathbb{R}$, if $y_0 \in \{y | F(y) \ge x\}$, then $y_0 \notin \{y | F(y) < x\}$, therefore $y_0 \ge F_1^{-1}(x)$, which means $F_2^{-1}(x) \ge F_1^{-1}(x)$. For any $\varepsilon > 0$, from the definition of $F_2^{-1}(\cdot)$, we have $F(F_2^{-1}(x) - \varepsilon) < x$. Therefore $F_2^{-1}(x) - \varepsilon \le F_1^{-1}(x)$, which means $F_2^{-1}(x) \le F_1^{-1}(x)$. It's obvious that $F_3^{-1}(x) \ge F_2^{-1}(x)$. If $y_0 \in \{y | F(y) > x\}$, then $y_0 \notin \{y | F(y) \le x\}$, therefore $y_0 \ge F_4^{-1}(x)$, which means $F_3^{-1}(x) \ge F_4^{-1}(x)$. For any $\varepsilon > 0$, from the definition of F_4^{-1}, we have $F(F_4^{-1}(x) + \varepsilon) > x$. Then $F_4^{-1}(x) + \varepsilon \ge F_3^{-1}(x)$, which means $F_4^{-1}(x) \ge F_3^{-1}(x)$. □

Remark. Theorem 1 shows that there are actually only two distinct versions of the generalized inverse of F defined in Section 1. The generalized inverse $(F_2^{-1}(\cdot))$ widely used in literature (e.g. [4], p. 113) is the smaller one. The asymptotic property of sample quantiles based on $F_2^{-1}(\cdot)$ have been studied extensively in statistical literatures (e.g. [4], p. 113). However, the asymptotic property of sample quantiles based on $F_4^{-1}(\cdot)$ has not been reported. It's reasonably to conjecture that it should have the same asymptotic properties as that based on $F_2^{-1}(\cdot)$.

3. Right Continuity

The distribution function *F* is right continuous. We want to know if its generalized inverses are also right continuous. Here is the result for its two versions of generalized inverses.

Theorem 2. F_4^{-1} *is right continuous. Generally* F_1^{-1} *is not right continuous.*

Proof. We first prove that $F_4^{-1}(\cdot)$ is right continuous by contradiction. Suppose not. Then there exist x_0 and h such that

$$F_4^{-1}(x_0^+) > h > F_4^{-1}(x_0).$$

Then $F(h) > x_0$. Hence there exist x_1 such that $F(h) > x_1 > x_0$. Therefore $h \ge F_4^{-1}(x_1) \ge F_4^{-1}(x_0^+)$. A contradiction.

As for F_1^{-1}, let F be defined as

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } 0 \le x < 2, \\ 2 & \text{if } 2 \le x. \end{cases}$$

Then for any $\varepsilon \in (0,1), F_1^{-1}(1) = 0 < F_1^{-1}(1+\varepsilon) = 2$. □

Remark 1. F_4^{-1} is called the right-continuous version inverse of F. The right-continuity property of both the distribution function and its quantile transform based on F_4^{-1} shows a symmetric property between these two functions. Marshall and Olkin [8] gave an nice introduction to the generalized inverse of a distribution function and prove that F_4^{-1} was right continuous in a different way. However, they did not give the inequalities in our Theorem 1.

Remark 2. There are some mistakes in statistical literatures about the continuity properties of generalized inverse of distribution functions. For example, Andersen *et al.* (1993, p. 274) stated that F_2^{-1} was the right-continuous inverse of *F*. According to our Theorem 2, their claim is incorrect.

4. Generalized Inverse and Quantile Transformation

In this section we assume that *F* is a PDF. It's well known that if *U* is uniformly distributed on $[0,1]$, then the random variable $F_1^{-1}(U)$ has distribution function *F*. Durrett [9] gives a nice proof. In his proof, he constructed a probability space $(\Omega, \mathcal{F}, P)$, where $\Omega = [0,1]$, $\mathcal{F}$ is the Borel σ-field on Ω, and *P* is the Lebesgue measure. For each *x*, define two sets $A_x = \{\omega | F_1^{-1}(\omega) \le x\}$ and $B_x = \{\omega | \omega \le F(x)\}$. It's easy to prove that $A_x = B_x$. Then $P(A_x) = P(B_x) = F(x)$. We have similar result for F_3^{-1}.

Theorem 3. $F_3^{-1}(U)$ *has distribution function* F.

Proof. Following the same idea of Durrett (2010), for any *x*, define $A_x = \{\omega | F_3^{-1}(\omega) \le x\}, B_x = \{\omega | \omega \le F(x)\}$. It's easy to prove that $A_x \subset B_x$. In general, $A_x \ne B_x$. However, if we define $B_x^- = \{\omega | \omega < F(x)\}$, then $B_x^- \subset A_x$. For if $\omega \in B_x^-$, then $F(x) > \omega$, *i.e.* $x \ge F_3^{-1}(\omega)$. As $P(B_x) = P(B_x^-) = F(x)$, we have $P(A_x) = F(x)$. □

5. Conclusion

In this paper we study the relations of four popular generalized inverses of a general distribution functions and their right-continuity properties. Our results indicate that the generalized inverse (F_2^{-1}) widely used in literature may be not right continuous. We also prove that for a PDF F, F_3^{-1} is a valid quantile transformation which has one more property (right continuity) than the quantile transformation F_2^{-1} which is currently used. One remaining problem is to show that the sample quantile based on F_3^{-1} has the same asymptotic properties as that based on F_2^{-1}. Since both F_2^{-1} and F_3^{-1} as reasonable generalized inverse of *F*, their average $(F_2^{-1} + F_3^{-1})/3$ should also be a good candidate of generalized inverse. The properties of this now function deserves further exploration.

6. Acknowledgements

This research was supported by grants 5U19AI056390-05 and 3 UL1 RR024160-02S1 from the National Institutes of Health.

REFERENCES

[1] R. B. Ash, "Probability and Measure Theory," 2nd Edition, Academic Press, San Diego, 2000.

[2] A. W. Van der Vaart, "Asymptotic Statistics," Cambridge University Press, New York, 1998.

[3] E. Langford, "Quartiles in Elementary Statistics," *Journal of Statistics Education*, Vol. 14, No. 3, 2006. www.amstat.org/publications/jse/v14n3/langford.html

[4] P. K. Andersen, Ø. Borgan, R. D. Gill and N. Keiding, "Statistical Models Based on Counting Processes," Springer, New York, 1993. doi:10.1007/978-1-4612-4348-9

[5] S. I. Resnick, "Extreme Values, Regular Variation, and Point Processes," Springer, New York, 1987.

[6] P. Embrechts, C. Klüppelberg and T. Mikosch, "Modeling Extremal Events for Insurance and Finance," Springer, New York, 1997.

[7] A. J. McNeil, R. Frey and P. Embrechts, "Quantitative Risk Management: Concepts, Techniques, Tools," Princeton University Press, Princeton, 2005.

[8] A. W. Marshall and I. Olkin, "Life Distributions," Springer, New York, 2007.

[9] R. Durrett, "Probability: Theory and Examples," 4th Edition, Cambridge University Press, New York, 2010. doi:10.1017/CBO9780511779398