Pakistan Journal of Statistics and Operation Research

On nonparametric estimation of the ratio of two reliability function

Zohreh Mohammadi[1*]

*Corresponding author

1. Department of Statistics, Jahrom University, Jahrom, Iran, z.mohammadi@jahromu.ac.ir

Abstract

One of the most important topics in reliability analysis is the comparisons of coherent systems. To this end, various methods and criteria have been proposed by researchers. This study is concerned with comparing such systems using the conditional reliability function. A nonparametric estimator for the conditional reliability function is derived and its properties are investigated. Then, for the comparison of the two independent systems, the ratio of their conditional reliability function is considered. A nonparametric estimator for this parameter is obtained and its asymptotic distribution is established. The asymptotic confidence interval and the corresponding hypothesis testing procedure is obtained. Finally, the applicability of the proposed method is investigated through simulation and an example of real data.

Key Words: Nonparametric estimation; Residual lifetime; Conditional Reliability; Maximum likelihood Estimator; Asymptotic properties.

Mathematical Subject Classification: 62G05, 62F12, 62N05.

1. Introduction

In reliability analysis, the concept of aging is described in terms of the survival function of residual lifetime, failure rate, mean residual lifetime, etc. Bryson and Siddiqui (1969) considered several criteria for the concept of aging. Deshpande et al. (1986) derived other criteria and related this concept to the stochastic dominance of first and higher orders. Let T represent the lifetime of an item or a system with the cumulative distribution function (CDF) $F_T(t)$. The reliability function of T at t is defined by $R_T(t) = 1 - F_T(t)$. Then, the random variable T_t, which is defined as $T - t|T > t$ shows the residual lifetime of a system operating at time $t > 0$. The reliability function of T_t at $x > 0$ is named the conditional reliability function, and defined as

$$R_{T_t}(x) = R(x|t) = \frac{R_T(x+t)}{R_T(t)} \quad \text{for } t > 0, \tag{1}$$

where for $t = 0$ we define: $R(x|0) = R_T(x)$. Siddiqui and Çağlar (1994) discussed the properties of the random variable T_t, including its mean, variance, and percentiles.

The function in (1) plays an important role in reliability analysis. Moreover, in many practical problems, researchers and practitioners are interested in comparing the reliability of several systems that are subject to failure. Various probability tools have been introduced to compare the reliability of different systems. These tools allow engineers and other decision makers to compare different products, maintenance and allocation policies. Navarro and Rychlik (2007) compared coherent systems using their lifetime expectations. Kochar et al. (1999) and Shaked and Suarez-Llorens (2003) used stochastic ordering for the comparisons of systems. Samaniego (2007) proposed to use stochastic precedence to compare the lifetime of two independent coherent systems. Navarro and Rubio (2010) obtained some new expressions to compare systems using stochastic precedence. The most important drawback of expected lifetime-based comparisons is their dependence on the distribution of components. The most important drawback of stochastic order-based comparisons is the lack of order in some systems. Zardasht and Asadi (2010) introduced a time-dependent criterion to compare the residual lifetime of two systems. This criterion was obtained based on the probability that the residual lifetime random variable of the one system is greater than that of the other system.

In this study, we take the conditional reliability function approach. We also consider the ratio of the conditional reliability function of two systems instead of their difference. In some cases, the ratio is more informative than the difference. For example, assume that the conditional reliability functions of two systems are 0.01 and 0.001, respectively. The ratio index suggests that the first system is ten times more reliable than the second one, while the corresponding difference is less informative.

The rest of the paper is organized as follows. In section 2, we formulate the problem and obtain a nonparametric estimator for the ratio of the conditional reliability function of two independent systems. We also establish the asymptotic distribution of the proposed estimator. Then, we use the asymptotic distribution to derive the asymptotic confidence interval and to test the hypothesis for the ratio of the system's conditional reliability function. In section 3, we investigate the performance of the proposed method through the Monte Carlo simulation. Finally, we analyze a real data set to illustrate the usefulness of the proposed method.

2. Nonparametric estimator and its properties

To start with, for a given t, we first derive a nonparametric estimator for the reliability function of the residual lifetime random variable T_t at x, i.e $R(x|t)$, and find its asymptotic distribution. Then, we use the results to estimate the ratio of the conditional reliability function of two independent systems.

Let $T_1, T_2, \ldots, T_n$ be a random sample from a distribution $F_T(t)$. Also, let Z_1 and Z_2 denote the number of failures in the random sample during the intervals $(0, t]$ and $(t, x+t]$, respectively. It is now possible to express the random variables Z_1 and Z_2 as below:

$$Z_1 = \sum_{j=1}^{n} I_{(0,t]}(T_j), \quad Z_2 = \sum_{j=1}^{n} I_{(t,t+x]}(T_j),$$

where

$$I_A(T_j) = \begin{cases} 1 & T_j \in A \\ 0 & otherwise. \end{cases}$$

Therefore, Z_1 and Z_2 have a multinomial distribution with the following probability function

$$f_{Z_1,Z_2}(z_1, z_2) = \frac{n!}{z_1! z_2! (n - z_1 - z_2)!} p_1^{z_1} p_2^{z_2} (1 - p_1 - p_2)^{(n - z_1 - z_2)} \quad z_1 + z_2 \leq n, \quad p_1 + p_2 \leq 1,$$

where $p_1 = p(T \leq t)$ and $p_2 = p(t < T \leq x + t)$. As a result, for a particular t, the reliability function of T_t at x can be written as

$$R(x|t) = 1 - \frac{p_2}{1 - p_1}. \tag{2}$$

It is now possible to obtain the maximum likelihood estimator (MLE) of p_i via $\hat{p}_i = \frac{Z_i}{n}$ $(i = 1, 2)$. By the invariance principle of MLEs, Zehna et al. (1966) and Berk and Zehna (1967), for a fixed x and t, the MLE of parameter $R(x|t)$ is

$$\hat{R}(x|t) = 1 - \frac{Z_2}{n - Z_1}. \tag{3}$$

This estimator is a nonparametric estimator for $R(x|t)$, because it does not depend on the distribution of F. We proceed with the derivation of several theoretical properties of $\hat{R}(x|t)$; these are outlined in Theorem 2.1, Theorem 2.2, Theorem 2.3, and Corollary 2.1.

Theorem 2.1. *$\hat{R}(x|t)$ is an unbiased estimator for $R(x|t)$.*

Proof.

$$\begin{aligned} E(\hat{R}(x|t)) =& 1 - E(\frac{Z_2}{n-Z_1}) = 1 - \sum_{z_1=0}^{n}\sum_{z_2=0}^{n-z_1}\frac{z_2}{n-z_1}f_{Z_1,Z_2}(z_1,z_2) \\ =& 1 - \sum_{z_1=0}^{n}\frac{n!}{z_1!(n-z_1)}\,p_1^{z_1}\sum_{z_2=1}^{n-z_1}\frac{1}{(z_2-1)!(n-z_1-z_2)!}p_2^{z_2}(1-p_1-p_2)^{(n-z_1-z_2)} \\ =& 1 - p_2\sum_{z_1=0}^{n}\frac{n!}{z_1!(n-z_1)!}\,p_1^{z_1}(1-p_1)^{n-z_1-1} = 1 - \frac{p_2}{1-p_1} \end{aligned} \quad (4)$$

□

Theorem 2.2. *Under the above assumptions,*

$$\hat{R}(x|t) - R(x|t) \xrightarrow{\mathcal{L}} N(0,\sigma^2) \quad as \;\; n \to \infty,$$

where

$$\sigma^2 = \frac{p_2(1-p_1-p_2)}{n(1-p_1)^3}.$$

Proof. Let $\mathbf{p} = (p_1,p_2)^{'}$ and $\hat{\mathbf{p}} = (\hat{p}_1,\hat{p}_2)^{'}$, by large sample properties of Maximum Likelihood Estimators (MLEs) (see Zacks, 2012), we obtain

$$(\hat{\mathbf{p}} - \mathbf{p}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, V(\mathbf{p})),$$

where $V(\mathbf{p}) = I^{-1}(\mathbf{p})$ is the inverse of Fisher information matrix for (Z_1, Z_2) which is equal to

$$I(\mathbf{p}) = \begin{bmatrix} \frac{n}{p_1} + \frac{n}{1-p_1-p_2} & \frac{n}{1-p_1-p_2} \\ \frac{n}{1-p_1-p_2} & \frac{n}{p_2} + \frac{n}{1-p_1-p_2} \end{bmatrix}$$

Now, we apply Cramer's Theorem (see page 45 Ferguson, 1996), for the function

$$R(x|t) = 1 - \frac{p_2}{1-p_1}.$$

We define $g : \mathbb{R}^2 \to \mathbb{R}$ as $g(x_1,x_2) = 1 - \frac{x_2}{1-x_1}$. Then, the gradient function with respect to g in point (p_1,p_2) is

$$\nabla g(p_1,p_2) = (\frac{-p_2}{(1-p_1)^2}, \frac{-1}{1-p_1}).$$

Finally, we complete the proof by utilizing the Delta Method. Specifically, it holds that

$$var(\hat{R}(x|t)) = \nabla g(p_1,p_2)I(\mathbf{p})^{-1}(\nabla g(p_1,p_2))^{'}$$

which leads to

$$\begin{aligned} var(\hat{R}(x|t)) =& \nabla g(p_1,p_2)I(\mathbf{p})^{-1}(\nabla g(p_1,p_2))^{'} \\ =& \frac{p_2(1-p_2)}{n(1-p_1)^2} - \frac{p_1p_2^2}{n(1-p_1)^3} \\ =& \frac{p_2(1-p_1-p_2)}{n(1-p_1)^3}. \end{aligned} \quad (5)$$

□

Now, we consider two independent systems with reliability functions $R_1(t)$ and $R_2(t)$, respectively. We want to compare them through the ratio of their conditional reliability functions, i.e. via the following parameter

$$\gamma = \frac{R_1(x|t)}{R_2(x|t)}. \quad (6)$$

Suppose that $T_{11}, T_{12}, \ldots, T_{1n_1}$ and $T_{21}, T_{22}, \ldots, T_{2n_2}$ are random samples with sizes n_1 and n_2 from two independent populations, respectively. Let

$$Z_{i1} = \sum_{j=1}^{n_i} I_{(0,t]}(T_{ij}), \;\; Z_{i2} = \sum_{j=1}^{n_i} I_{(t,t+x]}(T_{ij}), \quad i = 1, 2.$$

Since $\hat{R}_i(x|t) = 1 - \frac{Z_{i2}}{n_i - Z_{i1}}$ is MLE for $R_i(x|t)$ $i = 1, 2$, $\hat{\gamma} = \frac{\hat{R}_1(x|t)}{\hat{R}_2(x|t)}$ is an estimator for the parameter γ. In the sequence, without loss of generality, we suppose that $n_1 = n_2$. In the case that $n_1 \neq n_2$, we can let $n = min(n_1, n_2)$. The asymptotic distribution of the proposed estimator $\hat{\gamma}$ is given in the following theorem.

Theorem 2.3. *Under the above assumptions and if $p_{i1} = p(T_{i1} \leq t)$ and $p_{i2} = p(t < T_{i1} \leq x + t)$ for $i = 1, 2$.*

$$\hat{\gamma} - \gamma \xrightarrow{\mathcal{L}} N(0, \delta^2) \quad as \;\; n \to \infty, \tag{7}$$

where

$$\delta^2 = \frac{\sigma_1^2}{R_2^2(x|t)} + \frac{\sigma_2^2 R_1^2(x|t)}{R_2^4(x|t)} \tag{8}$$

and

$$\sigma_i^2 = \frac{p_{i2}(1 - p_{i1} - p_{i2})}{n(1 - p_{i1})^3} \tag{9}$$

Proof. By using Theorem 2.2, we have

$$\hat{R}_1(x|t) - R_1(x|t) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2), \quad as \;\; n \to \infty$$

and

$$\hat{R}_2(x|t) - R_2(x|t) \xrightarrow{\mathcal{L}} N(0, \sigma_2^2), \quad as \;\; n \to \infty$$

Since the samples are independent, we have

$$\begin{bmatrix} \hat{R}_1(x|t) - R_1(x|t) \\ \hat{R}_2(x|t) - R_2(x|t) \end{bmatrix} \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right).$$

Now, define $g : \mathbb{R}^2 \to \mathbb{R}$ as $g(x_1, x_2) = \frac{x_1}{x_2}$. Then the gradient function with respect to g in point $(R_1(x|t), R_2(x|t))$ is

$$\nabla g(R_1(x|t), R_2(x|t)) = \left(\frac{1}{R_2(x|t)}, -\frac{R_1(x|t)}{R_2^2(x|t)}\right)$$

From Cramer's Theorem Ferguson (1996), we have the following formula

$$var(\hat{\gamma}) = \left(\frac{1}{R_2(x|t)}, -\frac{R_1(x|t)}{R_2^2(x|t)}\right) \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1} \left(\frac{1}{R_2(x|t)}, -\frac{R_1(x|t)}{R_2^2(x|t)}\right)'$$

This completes the proof. □

Note that δ in (7) depends on the unknown parameters, so the asymptotic distribution in Theorem 2.3 cannot be used to construct an asymptotic confidence interval and hypothesis testing.

Corollary 2.1. *Under the assumptions stated in Theorem 2.3*

$$Q = \frac{\hat{\gamma} - \gamma}{\hat{\delta}} \xrightarrow{\mathcal{L}} N(0, 1) \quad as \;\; n \to \infty \tag{10}$$

where

$$\hat{\delta}^2 = \frac{\hat{\sigma}_1^2}{\hat{R}_2^2(x|t)} + \frac{\hat{\sigma}_2^2 \hat{R}_1^2(x|t)}{\hat{R}_2^4(x|t)},$$

$$\hat{R}_i(x|t) = 1 - \frac{Z_{i2}}{n - Z_{i1}}, \quad \hat{p}_{i1} = \frac{Z_{i1}}{n}, \ \ \hat{p}_{i2} = \frac{Z_{i2}}{n},$$

and $\hat{\sigma}_i^2$ *which is the same of* σ_i^2 *in (9), in which* p_{i1} *and* p_{i2} *are replaced by* $\hat{p}_{i1}$ *and* $\hat{p}_{i2}$*, respectively* $(i = 1, 2)$.

Proof. By the consistency of MLE, we have

$$\hat{\gamma} \xrightarrow{p} \gamma \qquad as \quad n \to \infty.$$

The proof is completed by combining Slutsky's Theorem with Theorem 2.3. □

Now, Q can be used as a pivotal quantity to construct asymptotic confidence interval for γ,

$$(\hat{\gamma} - z_{\frac{\alpha}{2}}\hat{\delta}, \hat{\gamma} + z_{\frac{\alpha}{2}}\hat{\delta}). \tag{11}$$

As previously mentioned, researchers are interested in comparing two independent systems as the ratio of the conditional reliability functions in many studies. This comparison can be made by hypothesis testing about the parameter γ. For instance, the assumption $\gamma = 1$ is equivalent to the assumption that two systems have equal residual lifetime. In general, to test $H_0 : \gamma = \gamma_0$, the test statistic is

$$Z = \frac{\hat{\gamma} - \gamma_0}{\hat{\delta}}. \tag{12}$$

By a similar methodology applied in Corollary 2.1, under the null hypothesis, Z has an asymptotic standard normal distribution.

3. Empirical study

In this section, the applicability of the proposed estimator was investigated using simulation methods. Also, the obtained results were used on a real data set.

3.1. Simulation study

The data sets were generated from different distributions and values of (n_1, n_2) and γ. The Exponential, Gamma and Weibull distributions with the following parameters

$$(n_1, n_2) = \{(70, 100), (100, 200), (200, 300), (500, 700)\}$$

and

$$(R_1(x|t), R_2(x|t)) = \{(0.2, 0.4), (0.2, 0.8), (0.4, 0.6), (0.6, 0.8), (0.8, 0.8)\}$$

were considered. The Monte Carlo method was used to investigate the accuracy of equations (11) and (12). For each parameter setting, the percentage of runs in which Equation (11) contains true γ was estimated based on 1000 simulation runs using R software. These values were reported as the empirical coverage probability in Table 1. Also, the Kolmogorov Smirnov normality test was applied to verify the normality of test statistic (12). The p-values are reported in Table 2.

In Figure 1, some typical simulation studies are reported by plotting the normal Q-Q plots of the test statistic (12) for different values of (n_1, n_2) and $(R_1(x|t), R_2(x|t))$.

Table 1 shows that the empirical coverage probability of the proposed estimator gets very close to the nominal level (0.95) as the sample size grows. Therefore, we can accept Equation (11) as the asymptotic confidence interval for γ. In addition, Figure 1 and Table 2 confirm that the asymptotic approximation seems to be quite satisfactory in all of the cases considered (P-Value is more than 0.05). Therefore, our approach is a good alternative for constructing a confidence interval (CI) and performing a test of the hypothesis for the ratio of residual lifetime reliability of two independent populations.

Table 1: The empirical coverage probability of proposed estimator.

Distribution	$(R_1(x\|t), R_2(x\|t))$	(n_1, n_2) $(70, 100)$	$(100, 200)$	$(200, 300)$	$(500, 700)$
Exponential	(0.2,0.4)	0.924	0.933	0.946	0.947
	(0.2,0.8)	0.933	0.937	0.946	0.956
	(0.4,0.6)	0.94	0.936	0.958	0.954
	(0.6,0.8)	0.949	0.955	0.956	0.95
	(0.8,0.8)	0.952	0.961	0.943	0.954
Gamma	(0.2,0.4)	0.94	0.945	0.941	0.935
	(0.2,0.8)	0.929	0.942	0.948	0.948
	(0.4,0.6)	0.953	0.948	0.961	0.947
	(0.6,0.8)	0.959	0.952	0.954	0.957
	(0.8,0.8)	0.948	0.946	0.951	0.956
Weibull	(0.2,0.4)	0.946	0.948	0.936	0.955
	(0.2,0.8)	0.93	0.942	0.949	0.952
	(0.4,0.6)	0.943	0.941	0.939	0.949
	(0.6,0.8)	0.959	0.955	0.94	0.957
	(0.8,0.8)	0.954	0.956	0.957	0.961

Table 2: Kolmogorov-Smirnov's normality test p-value for the test statistics.

Distribution	$(R_1(x\|t), R_2(x\|t))$	(n_1, n_2) $(70, 100)$	$(100, 200)$	$(200, 300)$	$(500, 700)$
Exponential	(0.2,0.4)	0.0512	0.2426	0.5816	0.4456
	(0.2,0.8)	0.0531	0.0991	0.3404	0.3620
	(0.4,0.6)	0.3399	0.8767	0.6615	0.9784
	(0.6,0.8)	0.8688	0.9062	0.9697	0.9746
	(0.8,0.8)	0.6111	0.8537	0.8411	0.9117
Gamma	(0.2,0.4)	0.0019	0.1660	0.4952	0.6131
	(0.2,0.8)	0.1193	0.2555	0.8801	0.9193
	(0.4,0.6)	0.5481	0.6938	0.7594	0.9901
	(0.6,0.8)	0.6299	0.7471	0.9665	0.9649
	(0.8,0.8)	0.4452	0.6807	0.9297	0.9548
Weibull	(0.2,0.4)	0.0429	0.0462	0.1622	0.3264
	(0.2,0.8)	0.1233	0.4615	0.4324	0.7112
	(0.4,0.6)	0.2482	0.6029	0.8915	0.9461
	(0.6,0.8)	0.9015	0.9812	0.9888	0.9909
	(0.8,0.8)	0.8992	0.8801	0.9548	0.9774

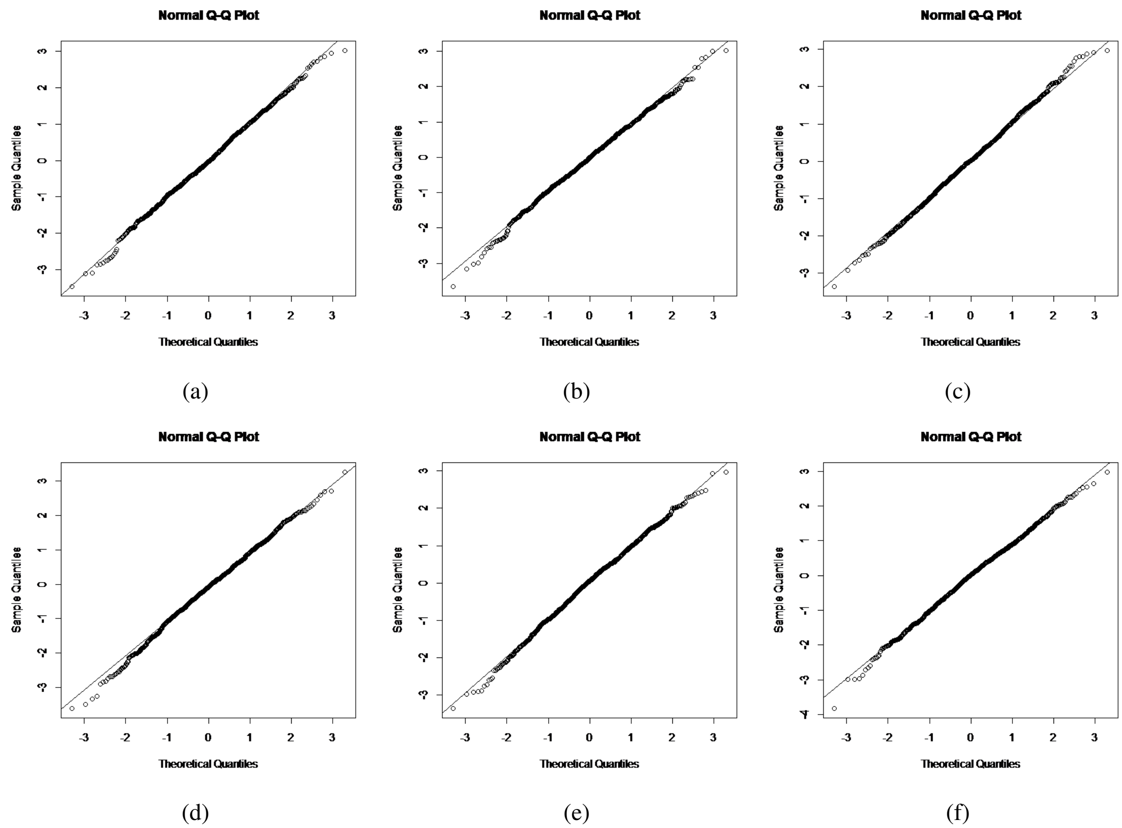

Figure 1: The Q-Q plots against the standard normal distribution. **a**. $(n_1, n_2) = (70, 100)$ and $(R_1(x|t), R_2(x|t)) = (0.6, 0.8)$; **b**.$(n_1, n_2) = (100, 200)$ and $(R_1(x|t), R_2(x|t)) = (0.4, 0.6)$; **c**.$(n_1, n_2) = (100, 200)$ and $(R_1(x|t), R_2(x|t)) = (0.8, 0.8)$; **d**.$(n_1, n_2) = (200, 300)$ and $(R_1(x|t), R_2(x|t)) = (0.2, 0.8)$; **e**.$(n_1, n_2) = (200, 300)$ and $(R_1(x|t), R_2(x|t)) = (0.4, 0.6)$; **f**.$(n_1, n_2) = (500, 700)$ and $(R_1(x|t), R_2(x|t)) = (0.2, 0.4)$.

3.2. Real data

As an application, a real data set was considered, which consisted of 228 patients with advanced lung cancer from the North Central Cancer Treatment Group. (This data set is available in the "`survival`" package in `R` software, where it is named "`lung`".) We considered the survival times of females and males, denoted by groups 1 and 2, respectively. In Table 3, for different values of x and t, the estimate and 95% CI for the parameter γ are shown. Also, the difference of $R_1(x|t)$ and $R_2(x|t)$ is reported. It can be observed that the parameter $\hat{\gamma}$ reveals the difference between the two groups better than the parameter $\hat{R}_1(x|t) - \hat{R}_2(x|t)$.

Table 3: Estimated some parameters of the Lung cancer data.

(t, x)	$\hat{\gamma}$	CI	$\hat{R}_1(x|t) - \hat{R}_2(x|t)$
$(0, 180)$	1.3082	(1.0079,1.6085)	0.1733
$(150, 30)$	1.1202	(0.9721,1.2682)	0.0996
$(10, 50)$	1.1324	(1.0394,1.2255)	0.1147
$(200, 470)$	1.9886	(-0.2192,4.1963)	0.0852

References

1. Berk, R. and Zehna, P. (1967). Invariance of maximum likelihood estimators. *Mathematical Reviews*, 33:342–343.

2. Bryson, M. C. and Siddiqui, M. (1969). Some criteria for aging. *Journal of the American Statistical Association*, 64(328):1472–1483.
3. Deshpande, J. V., Kochar, S. C., and Singh, H. (1986). Aspects of positive ageing. *Journal of Applied Probability*, 23:748–758.
4. Ferguson, T. S. (1996). *A course in large sample theory*. Chapman Hall.
5. Kochar, S., Mukerjee, H., and Samaniego, F. J. (1999). The “signature” of a coherent system and its application to comparisons among systems. *Naval Research Logistics (NRL)*, 46(5):507–523.
6. Navarro, J. and Rubio, R. (2010). Comparisons of coherent systems using stochastic precedence. *Test*, 19(3):469–486.
7. Navarro, J. and Rychlik, T. (2007). Reliability and expectation bounds for coherent systems with exchangeable components. *Journal of Multivariate Analysis*, 98(1):102–113.
8. Samaniego, F. J. (2007). *System signatures and their applications in engineering reliability*, volume 110. Springer Science & Business Media.
9. Shaked, M. and Suarez-Llorens, A. (2003). On the comparison of reliability experiments based on the convolution order. *Journal of the American Statistical Association*, 98(463):693–702.
10. Siddiqui, M. and Çağlar, M. (1994). Residual lifetime distribution and its applications. *Microelectronics Reliability*, 34(2):211–227.
11. Zacks, S. (2012). *Introduction to reliability analysis: probability models and statistical methods*. Springer Science & Business Media.
12. Zardasht, V. and Asadi, M. (2010). Evaluation of $p(\mathbf{X}_t > \mathbf{Y}_t)$ when both $\mathbf{X}_t$ and $\mathbf{Y}_t$ are residual lifetimes of two systems. *Statistica Neerlandica*, 64(4):460–481.
13. Zehna, P. W. et al. (1966). Invariance of maximum likelihood estimators. *Annals of Mathematical Statistics*, 37(3):744.