Published in final edited form as:

Stat Biopharm Res. 2012 October 1; 4(4): 348–356. doi:10.1080/19466315.2012.698945.

Comparing Two Exponential Distributions Using the Exact Likelihood Ratio Test

Gang Han¹, Michael J. Schell^{1,2}, and Jongphil Kim^{1,2}

Gang Han: Gang.Han@moffitt.org; Michael J. Schell: Michael.Schell@moffitt.org; Jongphil Kim: Jongphil.Kim@moffitt.org

¹Department of Biostatistics, H. Lee Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, Tampa, FL, 33612

²Oncologic Sciences, University of South Florida, 4202 E. Fowler Ave Tampa, FL, 33620

Abstract

The exact two-sided likelihood ratio test for testing the equality of two exponential means is proposed and proved to be the uniformly most powerful unbiased test. This exact test has advantages over two alternative approaches in that it is unbiased and more powerful while maintaining the type I error. The use of the proposed test is demonstrated in a non-small cell lung cancer clinical trial design.

Some Keywords

Survival analysis; exponential family; uniformly most powerful unbiased test; power calculation

1 Introduction

The exponential distribution has been shown to have high inference efficiency in survival analysis (Miller (1983) and Meier et al. (2004)). The problem of testing the equality of two exponential means is commonly seen in practice. For example, Bui et al. (2011) compared the survival distributions of sarcoma before and after metastasis as well as for patients with low and high biomarker Cx43 scores, where the survival distribution is believed to be exponential.

There have been existing approaches for comparing two exponential distributions. For example, Lee (1992) stated an asymptotic test. An exact F-test, which was originally given in Cox (1953), has been widely employed by statistical software packages including both noncommercial packages such as STPLAN and commercial packages such as *Power Analysis and Sample Size* (PASS). These two widely used approaches, however, suffer their weaknesses. As shown in Section 4, the ALRT is an asymptotic test requiring a relatively large sample size to maintain type I error. The F-test is a biased test and with the bias, its power is typically low, which may cause misleading interpretation of the data.

In this paper, we propose a uniformly most powerful unbiased (UMPU) test that has advantages over the existing tests. We discuss a likelihood ratio test in Section 2, and prove it is the UMPU test in Section 3. In Section 4, we compare the proposed test with the two existing tests. Concluding remarks are given in Section 5. Proofs of a remark, two lemmas, and a theorem are provided in the appendix.

2 Testing the Equivalence of Two Exponential Distributions

Suppose we have two groups of observations following exponential distributions. In group 1, we let $\{t_{1,i}\}_{i=1, \dots, n_1}$ and $\{c_{1,i}\}_{i=1, \dots, n_1}$ denote the event times and the censoring

indicator, respectively, where n_1 is the number of observations, $c_{1,i} = 1$ if the i th observation is an event, and $c_{1,i} = 0$ if censored. Similarly, we let $\{t_{2,i}\}_{i=1, \dots, n_2}$ and $\{c_{2,i}\}_{i=1, \dots, n_2}$ denote the event/censoring times and the censoring status in group 2. We let d_1 and d_2 denote the numbers of events in groups 1 and 2, respectively. We let $\{t_{1,1}, t_{1,2}, \dots, t_{1,d_1}\}$ denote the event times in group 1, and the corresponding $c_{1,1} = c_{1,2} = \dots = c_{1,d_1} = 1$ the indicators of the d_1 events. Similarly, $\{t_{2,1}, t_{2,2}, \dots, t_{2,d_2}\}$ are the event times in group 2, and $c_{2,1} = c_{2,2} = \dots = c_{2,d_2} = 1$ are indicators of d_2 events. We assume that the survival times in group 1 and 2 follow exponential distributions with means λ_1 and λ_2 , respectively, and that the censoring mechanisms in two groups are both type II. For $m \in \{1, 2\}$, the maximum likelihood estimates (MLE) of λ_m is

$$\hat{\lambda}_m = \frac{x_m}{d_m}, \quad (1)$$

where x_m is the total time on test (TTOT) between time 0 and time t_{m,d_m}

$$x_m = \text{TTOT}(0, t_{m,d_m}) = \sum_{i=1}^{d_m} t_{m,i} + (n_m - d_m) \times t_{m,d_m}.$$

The α level confidence interval of λ_m is

$$\lambda_m \in \left(\frac{2d_m \hat{\lambda}_m}{\chi_{2d_m, \alpha/2}^2}, \frac{2d_m \hat{\lambda}_m}{\chi_{2d_m, 1-\alpha/2}^2} \right),$$

where $\chi_{2d_m, \alpha/2}^2$ is the lower quantile at probability $\alpha/2$ of the central chi-square distribution with $2d_m$ degrees of freedom (Epstein and Sobel 1954). As discussed in Kalbfleisch and Prentice (1980), page 41, if the survival distribution is exponential in group 1, then x_1 is a realization from $\text{Gamma}(d_1, \lambda_1)$, the gamma distribution with parameters (d_1, λ_1) and having mean $d_1 \lambda_1$ and variance $d_1 \lambda_1^2$. Similarly, x_2 is a realization from $\text{Gamma}(d_2, \lambda_2)$ if the survival distribution is exponential in group 2. We let X_1 and X_2 denote two random variables with their distributions $X_1 \sim \text{Gamma}(d_1, \lambda_1)$ and $X_2 \sim \text{Gamma}(d_2, \lambda_2)$.

The null and alternative hypotheses for testing the equivalence between λ_1 and λ_2 are

$$H_0: \lambda_1 = \lambda_2 \text{ vs. } H_1: \lambda_1 \neq \lambda_2. \quad (2)$$

Under the null hypothesis, we let $\lambda = \lambda_1 = \lambda_2$ denote the model parameter. We derive the likelihood ratio test (LRT) in Remark 1.

Remark 1—The likelihood ratio test statistic for testing hypothesis (2) can be written as

$$\varphi(x_1, x_2) = \left(\frac{x_1}{x_1 + x_2} \right)^{d_1} \left(\frac{x_2}{x_1 + x_2} \right)^{d_2}.$$

The level α likelihood ratio test rejects H_0 if $\phi(x_1, x_2) \leq C_\alpha$, where C_α is a real value having $P(\phi(X_1, X_2) \leq C_\alpha | H_0) = \alpha$ for all $\alpha \in (0, 1)$.

Next we quantify the p-value of this LRT. Under H_0 , $Y = X_1/(X_1 + X_2)$ has a beta distribution, $\text{Beta}(d_1, d_2)$, with the *probability density function* (p.d.f.)

$$f_\beta(y|d_1, d_2) = \frac{(d_1 + d_2 - 1)!}{(d_1 - 1)!(d_2 - 1)!} y^{d_1-1} (1-y)^{d_2-1}, \quad (3)$$

for all $0 < y < 1$. Given x_1 and x_2 , the p-value of the LRT is $P(Y(1-Y) \leq \phi(x_1, x_2) | H_0)$. Because the function $y^{d_1}(1-y)^{d_2} \propto f_\beta(y|d_1+1, d_2+1)$ is strictly increasing for $y \in [0, d_1/(d_1+d_2)]$ and decreasing for $y \in [d_1/(d_1+d_2), 1]$, there must exist two real numbers $A_1 \in [0, d_1/(d_1+d_2)]$ and $A_2 \in [d_1/(d_1+d_2), 1]$ satisfying

$$Y^{d_1}(1-Y)^{d_2} \leq \phi(x_1, x_2) \text{ iff } Y \in [0, A_1] \cup [A_2, 1], \quad (4)$$

where $0 \leq A_1 \leq d_1/(d_1+d_2) \leq A_2 \leq 1$, and

$$f_\beta(A_1|d_1+1, d_2+1) = f_\beta(A_2|d_1+1, d_2+1) = \phi(x_1, x_2) \times \frac{(d_1+d_2+1)!}{d_1!d_2!}. \quad (5)$$

The p-value of the LRT can be computed as

$$p\text{-value} = P(Y \leq A_1 | H_0) + P(Y \geq A_2 | H_0). \quad (6)$$

We illustrate this LRT for $d_1 = d_2 = 1$. In this case, the test statistic is $\phi(x_1, x_2) = x_1 x_2 / (x_1 + x_2)^2 = y(1-y)$, where $y = x_1 / (x_1 + x_2)$. So $\phi(x_1, x_2) \leq C_\alpha$ iff

$$y \in \left[0, 0.5 - \sqrt{0.25 - C_\alpha}\right] \cup \left[0.5 + \sqrt{0.25 - C_\alpha}, 1\right].$$

Because $Y = X_1/(X_1 + X_2)$ has a uniform distribution on $[0, 1]$,

$P(y \in [0, 0.5 - \sqrt{0.25 - C_\alpha}] \cup [0.5 + \sqrt{0.25 - C_\alpha}, 1]) = \alpha$ leads to $C_\alpha = 0.25 \times (2\alpha - \alpha^2)$. If $\alpha = 0.05$, the LRT rejects H_0 if $x_1/(x_1 + x_2) \leq 1/40$ or $x_1/(x_1 + x_2) \geq 39/40$.

There exist A_1 and A_2 to achieve any level $\alpha \in (0, 1)$ because the function $y^{d_1}(1-y)^{d_2}$ is continuous and monotone for $y \in [0, d_1/(d_1+d_2)]$ and $y \in [d_1/(d_1+d_2), 1]$, and the *cumulative density function* (c.d.f.) of $\text{Beta}(d_1, d_2)$ is continuous as well. To compute the p-value of the test (2) given x_1 and x_2 , we search (A_1, A_2) numerically for arbitrary positive integers d_1 and d_2 with (5) and (6). Based on (3)–(5), $x_1/(x_1 + x_2) = A_1$ if $x_1/(x_1 + x_2) \leq d_1/(d_1 + d_2)$, and $x_1/(x_1 + x_2) = A_2$ if $x_1/(x_1 + x_2) \geq d_1/(d_1 + d_2)$. We implement the bisection search algorithm to determine $A_2 \in [d_1/(d_1 + d_2), 1]$ (or $A_1 \in [0, d_1/(d_1 + d_2)]$). The exact p-value in (6) can be computed using a beta distribution table or a beta c.d.f. in software packages such as MATLAB (function “betacdf”) and SAS (function “cdf”). Similarly, given a level α , the critical values A_1 and A_2 can be computed using (5) and the equation

$$P(A_1 < Y < A_2 | H_0) = 1 - \alpha. \quad (7)$$

3 UMPU Property of the Likelihood Ratio Test

Lehmann (1986), page 188–192, provided a general form of the UMPU test for the exponential family. For a two parameter model having parameters (θ_1, θ_2) , the level α UMPU test for

$$H_0^*: \theta_2 = 0 \text{ vs. } H_1^*: \theta_2 \neq 0 \quad (8)$$

is to reject H_0^* if $V(\mathbf{x}) < C_1$ or $V(\mathbf{x}) > C_2$ if the following five conditions hold:

1. The likelihood function can be written as

$$L(\theta_1, \theta_2 | \mathbf{x}) = C(\theta_1, \theta_2) H(\mathbf{x}) \exp\{T(\mathbf{x})\theta_1 + U(\mathbf{x})\theta_2\}, \quad (9)$$

where $C(\theta_1, \theta_2)$ is a function of (θ_1, θ_2) and $H(\mathbf{x})$, $T(\mathbf{x})$, and $U(\mathbf{x})$ are functions of $\mathbf{x}$;

2. $T(\mathbf{x})$ is sufficient for θ_1 and $U(\mathbf{x})$ is sufficient for θ_2 ;
3. $V(\mathbf{X})$ is independent of $T(\mathbf{X})$ under $\theta_2 = 0$;
4. $V(\mathbf{x})$ can be written as

$$V(\mathbf{x}) = a(T(\mathbf{x}))U(\mathbf{x}) + b(T(\mathbf{x})), \quad (10)$$

where $a(\cdot)$ and $b(\cdot)$ are two functions with $a(T(\mathbf{x})) > 0$;

5. $V(\mathbf{x})$ is continuous in $\mathbf{x}$ and $L(\theta_1, \theta_2 | \mathbf{x})$ is continuous in θ_1 and θ_2 .

The value of (C_1, C_2) is determined by

$$P(V(\mathbf{X}) \leq C_1) + P(V(\mathbf{X}) \geq C_2) = \alpha \quad (11)$$

and

$$E(V(\mathbf{X}) \times (1 - I_{(C_1, C_2)}(V(\mathbf{X})))) = \alpha \times E(V(\mathbf{X})), \quad (12)$$

where $I_{(C_1, C_2)}(\cdot)$ is an indicator function such that $I_{(C_1, C_2)}(V(\mathbf{X})) = 1$ if $V(\mathbf{X}) \in (C_1, C_2)$ and 0 otherwise.

Applying this general form to the two sample exponential setting, we can derive Lemma 1.

Lemma 1—The UMPU test of H_0 vs. H_1 in (2) rejects H_0 if $y = x_1/(x_1 + x_2) < C_1$ or $y > C_2$ where C_1 and C_2 are determined by

$$\int_{C_1}^{C_2} f_{\beta}(y | d_1, d_2) dy = 1 - \alpha \quad (13)$$

and

$$\int_{C_1}^{C_2} f_{\beta}(y|d_1+1, d_2)dy = 1 - \alpha. \quad (14)$$

We propose Lemma 2 to prove Theorem 1.

Lemma 2—Suppose that X_n is a random variable having binomial distribution $Bin(n, p)$, where the probability mass function (p.m.f.) of X_n can be written as

$$P(X_n=x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \quad (15)$$

for all $x = 0, 1, \dots, n$, and Z_{n+1} is a random variable with $Bin(n+1, p)$. If Z_{n+1} is independent with X_n , then

$$P(Z_{n+1} \geq k) = P(X_n \geq k) + P(X_n = k-1) \times p, \quad (16)$$

for any positive integer k .

Following Lemma 1 and Lemma 2, Theorem 1 provides the UMPU condition for the test of H_0 vs. H_1 .

Theorem 1—For any level $\alpha \in (0, 1)$, the LRT in Remark 1 is the UMPU test.

4 Method Comparison

In this section, we compare two existing tests with the exact LRT. Prior to the comparison in Section 4.3, we introduce the alternative tests in Section 4.1 and derive the power of the three tests in Section 4.2.

4.1 Alternative Tests

The first alternative is an asymptotic likelihood ratio test (ALRT) described in Lee (1992), page 233–236. The α level ALRT rejects H_0 in (2) if $-2\log(\Lambda) > \chi_{1,1-\alpha}^2$, where

$$\Lambda = \frac{(d_1+d_2)^{d_1+d_2}}{d_1^{d_1} d_2^{d_2}} \times \frac{x_1^{x_1} x_2^{x_2}}{(x_1+x_2)^{x_1+x_2}}$$

is the LRT test statistic. Following the notation $y = x_1/(x_1 + x_2)$ and the p.d.f. in (3), the α level ALRT rejects H_0 if $f_{\beta}(y|d_1+1, d_2+1) \leq C_{\alpha}^A$, where

$$C_{\alpha}^A = \frac{d_1^{d_1} d_2^{d_2} (d_1+d_2+1)!}{(d_1+d_2)^{d_1+d_2} d_1! d_2!} \exp \{-\chi_{1,1-\alpha}^2/2\}.$$

As discussed in Section 2, $f_{\beta}(y|d_1+1, d_2+1) \propto y^{d_1} (1-y)^{d_2}$ is monotonically increasing for $y \in [0, d_1/(d_1+d_2)]$ and monotonically decreasing for $y \in [d_1/(d_1+d_2), 1]$. The rejection region of the ALRT for testing the two-sided hypothesis (2) at level α is

$$y \in [0, A_1^A] \cup [A_2^A, 1], \quad (17)$$

where $0 \leq A_1^A \leq d_1/(d_1+d_2) \leq A_2^A \leq 1$ and $f_\beta(A_1^A|d_1+1, d_2+1) = f_\beta(A_2^A|d_1+1, d_2+1) = C_\alpha^A$ for all $\alpha \in [0, 1]$. The bisection search algorithm mentioned in Section 2 can be used to compute A_1^A and A_2^A numerically. Lee (1992) pointed out three limitations of the ALRT. First, it does not suit one-sided hypotheses. Secondly, this asymptotic test requires a large sample size. Thirdly, the power of ALRT is typically lower than the exact tests.

The second alternative is an F-test originally proposed by Cox (1953) for comparing the rates of occurrence of two Poisson samples, which is the same as comparing the means of two exponential samples of waiting times. Based on the fact that $F' = (X_1/d_1)/(X_2/d_2)$ has an exact F distribution with degrees of freedom $2d_1$ and $2d_2$ under H_0 , the F-test rejects H_0 in (2) at level α if $f' \leq F_{2d_1, 2d_2, \alpha/2}$ or $f' \geq F_{2d_1, 2d_2, 1-\alpha/2}$, where $f' = (x_1/d_1)/(x_2/d_2)$ and $F_{2d_1, 2d_2, \alpha}$ is the $100 \times \alpha\%$ lower quantile of the F distribution with degrees of freedom $2d_1$ and $2d_2$. Because of the functional relation $Y = X_1/(X_1 + X_2) = 1 - 1/(F'd_2 + 1)$, the F-test and the LRT are identical when testing a one-sided hypothesis. For the two-sided hypothesis in (2), the level α F-test has the rejection region

$$y \in [0, I_\beta^{-1}(\alpha/2; d_1, d_2)] \cup [I_\beta^{-1}(1 - \alpha/2; d_1, d_2), 1] \quad (18)$$

The difference between the F-test and the proposed LRT lies in that the F-test has the equal tail probability rejection region in (18) but the rejection region of the LRT is defined by equations (5) and (7). The F-test and the LRT are identical if and only if $A_1 = I_\beta(\alpha/2; d_1, d_2)$ and $A_2 = I_\beta(1 - \alpha/2; d_1, d_2)$, which occurs when $d_1 = d_2$.

4.2 Rejection Probabilities of the Three Tests

To quantify the power, we first derive the p.d.f. of $y = x_1/(x_1 + x_2)$ given λ_1 and λ_2 . Define $z = x_2$. Then $x_1 = yz/(1-y)$ and $x_2 = z$. Thus,

$$\begin{aligned} f(y|\lambda_1, \lambda_2) &= \int_0^\infty f(y, z|\lambda_1, \lambda_2) dz \\ &= \int_0^\infty f\left(\frac{yz}{1-y}|\lambda_1\right) f_2(z|\lambda_2) \times \left| \frac{\frac{dz_1}{dz} \frac{dz_2}{dz}}{\frac{dx_1}{dy} \frac{dx_2}{dy}} \right| dz \\ &= \int_0^\infty f\left(\frac{yz}{1-y}|\lambda_1\right) f_2(z|\lambda_2) \times \left| \frac{z}{(1-y)^2} \right| dz \\ &= \frac{y^{d_1-1} (d_1+d_2-1)!}{(d_1-1)!(d_2-1)! \lambda_1^{d_1} \lambda_2^{d_2} (1-y)^{d_1+1}} \left[\frac{\lambda_1 \lambda_2 (1-y)}{\lambda_2 y + \lambda_1 (1-y)} \right]^{d_1+d_2} \\ &= f_\beta(y|d_1, d_2) \times \frac{\lambda_1^{d_2} \lambda_2^{d_1}}{[(\lambda_2 - \lambda_1)y + \lambda_1]^{d_1+d_2}}. \end{aligned} \quad (19)$$

Thus, given two generic real numbers $0 \leq g_1 \leq g_2 \leq 1$, we can quantify the probability of $Y \in [0, g_1] \cup [g_2, 1]$ as

$$\begin{aligned} P(Y \in [0, g_1] \cup [g_2, 1]|\lambda_1, \lambda_2) &= 1 - P(Y \in (g_1, g_2)|\lambda_1, \lambda_2) \\ &= 1 - \int_{A_1}^{A_2} f(y|\lambda_1, \lambda_2) dy \\ &= 1 - \int_{A_1}^{A_2} f_\beta(y|d_1, d_2) \times \frac{\lambda_1^{d_2} \lambda_2^{d_1}}{[(\lambda_2 - \lambda_1)y + \lambda_1]^{d_1+d_2}} dy. \end{aligned} \quad (20)$$

Combining (20) with (5), (7), (17), and (18), the rejection probabilities of all three tests at level α has the form (20) with $(g_1, g_2) = (A_1^A, A_2^A)$ for the ALRT,

$(g_1, g_2) = (I_{\beta}^{-1}(\alpha/2; d_1, d_2), I_{\beta}^{-1}(1 - \alpha/2; d_1, d_2))$ for the F-test, and $(g_1, g_2) = (A_1, A_2)$ for the proposed LRT. Numerical integration methods (e.g., importance sampling, Laplace approximation, and Riemann sum approximation) can be used to approximate the integration in (20). We use Riemann sum approximation for the three examples in Section 4.3.

4.3 Comparing Three Tests

We compare the three tests in Examples 1–3 by investigating the type I error in Example 1 and the power in Examples 2 and 3.

Example 1—Let $d_1 = d_2 = d$, $\lambda_1 = \lambda_2$, and level $\alpha = 0.05$. Rejection probabilities of the three tests are calculated for $d = 1, \dots, 100$ using (20). Figure 1(a) plots the type I error against d_1 . We can see that both the F-test and the LRT have their type I errors equal to α . The type I error of the ALRT is larger than 0.05 and converges to 0.05 asymptotically as the sample size increases. A general rule given by Lee (1992) suggests that the ALRT works only if $d_1 + d_2 \geq 25$, which can be violated in engineering experiments and medical research if the sample size is small. The rejection probability under $\alpha = 0.05$ for $d = 13$ (the smallest value of d to satisfy $d_1 + d_2 \geq 25$) is roughly 0.053, which can be still too high for controlling the type I error.

Example 2—Let $(d_1, d_2) = (30, 4)$ and $(\lambda_1, \lambda_2) = (12, 11)$. We compare the power of the F-test with the power of the LRT for level $\alpha \in (0, 1)$. Following the power computation in (20), we depict in Figure 1(b) the difference between the power and level α . We see that the proposed LRT is unbiased because the solid curve is positive. The F-test, however, is biased for all $\alpha \in (0, 1)$. Further, the LRT is more powerful than the F-test. For example, at level $\alpha = 0.1$, power of the F-test is about 0.097, and the power from the LRT is about 0.104 in this example.

Example 3—We investigate the relation between the power and sample size of the F-test and LRT under two scenarios. In this example, $\alpha = 0.1$, $(\lambda_1, \lambda_2) = (10, 10.5)$ for scenario 1, and $(\lambda_1, \lambda_2) = (10, 15)$ for scenario 2. We compute the power of the F-test and LRT for $d_1 = \{1, 2, \dots, 50\}$ and $d_2 = 4 \times d_1$, which mimics a typical sample size increase when moving from the phase II to the phase III of a clinical trial. Figure 1(c) and (d) plot the power against d_1 for scenarios 1 and 2, respectively. We can see that the power is increasing with the sample size d_1 . The power of the LRT is higher than the power of the F-test for all $d_1 \in \{1, 2, \dots, 50\}$. For instance, 2 more patients are needed for the F-test to be as powerful as the LRT for all $d_1 \in \{1, \dots, 50\}$ in scenario 2. Moreover, bias from the F-test occurs under both scenarios. In scenario 1 where the difference between λ_1 and λ_2 is small ($\lambda_2 = 1.05 \times \lambda_1$), the F-test is biased for all $d_1 < 11$; for scenario 2 with a larger difference between λ_1 and λ_2 ($\lambda_2 = 1.5 \times \lambda_1$), the F-test is biased for $d_1 < 2$. This result indicates that the bias in the F-test occurs mostly for small sample size with a moderate difference between λ_1 and λ_2 .

4.4 Application in the Design of a Comparative Lung Cancer Clinical Trial

We implement the F-test and LRT to design a non-small cell lung cancer (NSCLC) clinical trial. Simon, Schell, Begum, Haura, Antonia and Beppler (2011) introduced a personalized NSCLC therapy along with three alternative standard (non-personalized) NSCLC therapies, and showed that the personalized therapy had statistically significant improvements over standard therapies in response rate, overall survival, and progression free survival. We attempt to design a phase III trial to further compare the first 2 year survival distributions of the personalized therapy and that of a standard therapy (recorded as MCC 13303 in Simon et al. (2011)). The time span of first two years was chosen because it was where the difference

in survival was found and where exponential assumption was believed to be valid. We assume that more patients can be recruited for the personalized therapy compared with the alternative therapy because 1, the personalized therapy showed a significant improvement in the Phase II trial (Simon et al. 2011), and 2, non-small cell lung cancer patients typically prefer the most promising treatment. Specifically, we assume that 75% of the events are from the personalized therapy and the other 25% are from the alternative therapy.

Figure 2 plots the Kaplan-Meier curves of overall survival corresponding to the two therapies from month 0 to 24. Assuming that $\lambda_1 = 24$ for personalized therapy group and $\lambda_2 = 15$ for the alternative group, we applied the exponential test in Hollander and Proschan (1979) to check the exponential assumption with patients alive at 2 years being censored. The p-value of 0.853 (test statistic 0.185) of personalized therapy group and 0.561 (test statistic 0.582) of the other group indicate that exponential assumption was valid for both groups.

We compute the power of the F-test and LRT based on the MLE of (λ_1, λ_2) that were estimated using the data in Simon et al. (2011), i.e., $\hat{\lambda}_1 = 22.25$ and $\hat{\lambda}_2 = 13.52$. Table 1 lists the exact power from the two tests for the two-sided hypothesis in (2) with significance level 0.1. We can see that in the first row, where only 4 patients were recruited, the power of the F-test is less than the significance level 0.1, which indicates that the F-test is biased. The proposed LRT is unbiased. Furthermore, the minimum sample size to achieve 80% of power is 136 for the LRT, less than that for the F-test with difference of 4 patients. Similarly, the minimum sample size to achieve 90% of power is 184 for the LRT, which is 4 patients less than that of the F-test.

5 Summary and Concluding Remarks

The test for comparing two sample exponential distributions is used extensively in survival studies and engineering experiments. A UMPU test is necessary to provide correct interpretation of the two example exponential data in scientific research. In this paper we have proposed a two-sided exact likelihood ratio test to compare two exponential parameters. We have proved the LRT is the UMPU test. This test can also be implemented to compare the occurrence rates of two Poisson samples (Cox 1953).

Because of the UMPU property, the acceptance regions of $Y = X_1/(X_1 + X_2)$ for all level α are the uniformly most accurate unbiased (UMAU) confidence sets under H_0 (Lehmann 1986). Similar form of the acceptance region has been discussed in literature. For instance, Tate and Klett (1959) provided the optimal confidence interval of the variance parameter from a normal distribution. Their confidence interval was defined using two conditions similar to (5) and (7).

To implement the proposed LRT (as well as the other two tests), it is crucial to verify the exponential assumption. Given that the focus of this manuscript is rather on the exact UMPU test for comparing two exponential samples, we briefly discuss violation and verification of the exponential assumption next. Violation of the exponential assumption may be due to either non-exponential survival or censoring (e.g., different follow-up time or lost to follow-up). Verifying the exponential distribution with censored observations can be challenging, especially for small samples. Three approaches could be applied to check the exponential assumption: The first approach is the test derived in Hollander and Proschan (1979) that was shown in Section 4.4 for the lung cancer data with censoring. The second is a graphical inspection approach given in Lee (1992). Relying on subjective judgement only, this approach can be difficult to apply in practice. The third approach, which is currently under development, is a reduced piecewise exponential modeling approach that builds on the proposed LRT and a backward elimination procedure to model the failure rate by a

piecewise constant function with significant change points. Exponential distribution is believed to be valid if no significant change point in the failure rate could be identified. According to our findings, this approach is more objective than the graphical inspection, and can be more powerful than the test in Hollander and Proschan (1979).

Current software packages (e.g., PASS and STPLAN) implement the F-test, which, as shown in Section 4.3, may fail to reject significant differences due to its bias and lower power, and thus result in misleading interpretation of the data. The weaknesses of the F-test are more pronounced in situations where the sample size of the test is relatively small, which can be seen in clinical trials of rare diseases and high cost industrial experiments. We suggest that the proposed LRT shall be used instead for comparing two exponential distributions. A software package on the two-sided UMPU test has been developed in MATLAB and is available upon request.

Acknowledgments

The authors would like to thank the editor, Dr. Steven Snapinn, the associate editor, and two referees for their constructive suggestions that improved the quality of this paper. The authors give thanks Gerold Bepler and George R. Simon for providing the non-small cell lung cancer data. This research was sponsored, in part, by the National Institute of Health and the National Cancer Institute, grant 1RC2CA14833201.

References

- Basu D. On Statistics Independent of a Sufficient Statistic. *Sankhyā*. 1958; 20:223–226.
- Bui M, Han G, Acs G, Gonzalez R, Reed D, Pasha IL, Zhang P. Connexin 43 Is a Potential Prognostic Biomarker for Ewing Sarcoma (EWS)/Primitive neuroectodermal tumor (PNET). *Sarcoma*. 2011:Article ID 971050.10.1155/2011/971050
- Cox DR. Some Simple Approximate Tests for Poisson Variates. *Biometrika*. 1953; 40:354–360.
- Epstein B, Sobel M. Some Theorems Relevant to Life Testing from an Exponential Distribution. *Annals of Mathematical Statistics*. 1954; 25:373–381.
- Hollander M, Proschan F. Testing to Determine the Underlying Distribution Using Randomly Censored Data. *Biometrics*. 1979; 35:393–401.
- Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. Wiley; New York: 1980.
- Lee, ET. *Statistical Methods for Survival Data Analysis*. Wiley; New York: 1992.
- Lehmann, EL. *Testing Statistical Hypotheses*. Springer; New York: 1986.
- Meier P, Karrison T, Chappell R, Xie H. The Price of Kaplan-Meier. *Journal of the American Statistical Association*. 2004; 99:890–896.
- Miller RG. What Price KaplanMeier? *Biometrics*. 1983; 39:1077–1081. [PubMed: 6671119]
- Simon GR, Schell MJ, Begum M, Haura JKE, Antonia SJ, Bepler G. Preliminary indication of survival benefit from ERCC1 and RRM1-tailored chemotherapy in patients with advanced nonsmall cell lung cancer: Evidence from an individual patient analysis. *Cancer*. 2011;10.1002/cncr.26522
- Tate RF, Klett GW. Optimal Confidence Intervals for the Variance of a Normal Distribution. *Journal of the American Statistical Association*. 1959; 54:674– 682.

APPENDIX. PROOFS

Proof of Remark 1

Proof

The test statistic of the LRT is

$$\begin{aligned}
\Lambda &= \frac{L(\widehat{\lambda})\{t_{1,i}, c_{1,i}\}_{i=1}^{n_1}}{L(\widehat{\lambda}_1, \widehat{\lambda}_2)\{t_{2,j}, c_{2,j}\}_{j=1}^{n_2}} \\
&= \frac{\prod_{i=1}^{d_1} f(t_{1,i}|\widehat{\lambda})(1-F(t_{1,d_1}|\widehat{\lambda}))^{n_1-d_1} \times \prod_{j=1}^{d_2} f(t_{2,j}|\widehat{\lambda})(1-F(t_{2,d_2}|\widehat{\lambda}))^{n_2-d_2}}{\prod_{i=1}^{d_1} f(t_{1,i}|\widehat{\lambda}_1)(1-F(t_{1,d_1}|\widehat{\lambda}_1))^{n_1-d_1} \times \prod_{j=1}^{d_2} f(t_{2,j}|\widehat{\lambda}_2)(1-F(t_{2,d_2}|\widehat{\lambda}_2))^{n_2-d_2}} \\
&= \frac{(d_1+d_2)^{d_1+d_2}}{d_1^{d_1} d_2^{d_2}} \times \frac{x_1^{x_1} x_2^{x_2}}{(x_1+x_2)^{x_1+x_2}} \\
&\propto \left(\frac{x_1}{x_1+x_2}\right)^{d_1} \left(\frac{x_2}{x_1+x_2}\right)^{d_2} = \varphi(x_1, x_2).
\end{aligned}$$

So the LRT rejects H_0 if $(x_1/(x_1+x_2))^{d_1} (x_2/(x_1+x_2))^{d_2} > C_\alpha$, where $P((X_1/(X_1+X_2))^{d_1} (X_2/(X_1+X_2))^{d_2} > C_\alpha) = \alpha$ under H_0 .

Proof of Lemma 1

Proof

Define $\theta_1 = 1/\lambda_1 + 1/\lambda_2$ and $\theta_2 = 1/\lambda_2 - 1/\lambda_1$. The likelihood function of (θ_1, θ_2) can be written in the form of (9), where

$$\begin{aligned}
C(\theta_1, \theta_2) &= I_{(0,\infty)}(\theta_1) I_{(-\theta_1, \theta_1)}(\theta_2) / \left[(\theta_1 + \theta_2)^{d_1} (\theta_1 - \theta_2)^{d_2} \right], \\
H(x_1, x_2) &= 2^{d_1+d_2} x_1^{d_1-1} x_2^{d_2-1} / [(d_1-1)!(d_2-1)!], \\
T(x_1, x_2) &= -(x_1+x_2)/2, \\
U(x_1, x_2) &= (x_1-x_2)/2.
\end{aligned}$$

Let $V(x_1, x_2) = x_1/(x_1+x_2) = y$. Then the distribution of Y , Beta(d_1, d_2), is free of θ_1 under H_0 . Thus $V(x_1, x_2)$ is an ancillary statistic of θ_1 . For exponential family, $T(X_1, X_2)$ is minimum sufficient and complete for θ_1 . (See Lehmann (1986), page 142.) Thus, Basu's theorem (Basu 1958) guarantees that $V(X_1, X_2)$ and $T(X_1, X_2)$ are independent. On the other hand, $V(x_1, x_2)$ can be written in the form of (10) with $a(T(x_1, x_2)) = -1/[2T(x_1, x_2)] > 0$ and $b(T(x_1, x_2)) = 0.5$. Thus, (9) and (10) both hold.

Note that testing the hypothesis in (2) is equivalent to testing $H_0^*: \theta_2=0$ vs. $H_1^*: \theta_2 \neq 0$.

Following the five conditions in Section 3, the UMPU test rejects H_0 (or H_0^*) if $y < C_1$ or $y > C_2$ where the values of C_1 and C_2 are determined by (11) and (12). Since Y has Beta(d_1, d_2) distribution, (11) can be written as (13), and (12) can be written as

$$\int_{C_1}^{C_2} y f_\beta(y|d_1, d_2) dy = (1-\alpha) \times \frac{d_1}{d_1+d_2}. \quad (21)$$

By (3), (21) is equivalent to (14). Thus (13) and (14) are equivalent to (11) and (12). The test in Lemma 1 is the UMPU test.

Proof of Lemma 2

Proof

Let $X_n = \sum_{i=1}^n S_i$ and $Z_{n+1} = \sum_{j=1}^{n+1} R_j$ denote two mutually independent random variables distributed Bin(n, p) and Bin($n+1, p$), respectively, where $\{S_i\}_{i=1, \dots, n}$, $\{R_j\}_{j=1, \dots, n+1}$ are $2n+1$ mutually independently and identically distributed (i.i.d.) random variables having

Bernoulli distribution with parameter p . Using the i.i.d. condition and the definition of X_n and Z_{n+1} ,

$$\begin{aligned}
 P(Z_{n+1} \geq k) &= P\left(\sum_{i=1}^{n+1} R_i \geq k\right) \\
 &= P\left(\sum_{i=1}^n R_i \geq k \text{ and } \sum_{i=1}^{n+1} R_i \geq k\right) + P\left(\sum_{i=1}^n R_i < k \text{ and } \sum_{i=1}^{n+1} R_i \geq k\right) \\
 &= P\left(\sum_{i=1}^n R_i \geq k\right) + P\left(\sum_{i=1}^n R_i = k-1 \text{ and } R_{n+1} = 1\right) \\
 &= P\left(\sum_{i=1}^n R_i \geq k\right) + P\left(\sum_{i=1}^n R_i = k-1\right) \times P(R_{n+1} = 1) \\
 &= P\left(\sum_{i=1}^n S_i \geq k\right) + P\left(\sum_{i=1}^n S_i = k-1\right) \times p \\
 &= P(X_n \geq k) + P(X_n = k-1) \times p.
 \end{aligned}$$

Equation (16) holds for all positive integer k .

Proof of Theorem 1

Proof

Conditional on (13), (14) is equivalent to $\int_{C_1}^{C_2} f_\beta(y|d_1+1, d_2)dy = \int_{C_1}^{C_2} f_\beta(y|d_1, d_2)dy$, which can be written as $I_\beta(C_2; d_1+1, d_2) - I_\beta(C_1; d_1+1, d_2) = I_\beta(C_2; d_1, d_2) - I_\beta(C_1; d_1, d_2)$, where the function $I_\beta(\cdot; d_1, d_2)$ is the *incomplete Beta function* with parameters d_1 and d_2 , and $I_\beta(C_1; d_1, d_2) = P(Y < C_1 | Y \sim \text{Beta}(d_1, d_2))$. Thus, conditional on (13), (14) can be written as

$$I_\beta(C_2; d_1, d_2) - I_\beta(C_2; d_1+1, d_2) = I_\beta(C_1; d_1, d_2) - I_\beta(C_1; d_1+1, d_2). \quad (22)$$

By integrating beta p.d.f. (3) by parts, $I_\beta(C_1; d_1, d_2)$ can be written as

$$I_\beta(C_1; d_1, d_2) = \sum_{j=d_1}^{d_1+d_2-1} \frac{(d_1+d_2-1)!}{j!(d_1+d_2-1-j)!} C_1^j (1-C_1)^{d_1+d_2-1-j}. \quad (23)$$

Let $n = d_1 + d_2 - 1$ and $p = C_1$. Following the notation in Lemma 2, $X_n \sim \text{Bin}(d_1 + d_2 - 1, C_1)$ and $Z_{n+1} \sim \text{Bin}(d_1 + d_2, C_1)$.

By (15) and (23), $I_\beta(C_1; d_1, d_2) = P(X_n \geq d_1)$ and $I_\beta(C_1; d_1+1, d_2) = P(Z_{n+1} \geq d_1+1)$. Using Lemma 2, the right hand side of (22) can be written as

$$\begin{aligned}
 I_\beta(C_1; d_1, d_2) - I_\beta(C_1; d_1+1, d_2) &= P(X_n \geq d_1) - P(Z_{n+1} \geq d_1+1) \\
 &= P(X_n \geq d_1) - [P(X_n \geq d_1+1) + C_1 \times P(X_n = d_1)] \\
 &= [P(X_n \geq d_1) - P(X_n \geq d_1+1)] - C_1 \times P(X_n = d_1) \\
 &= P(X_n = d_1) - C_1 \times P(X_n = d_1) \\
 &= (1 - C_1) \times P(X_n = d_1) \\
 &= \frac{(d_1+d_2-1)!}{d_1!(d_2-1)!} C_1^{d_1} (1 - C_1)^{d_2}.
 \end{aligned} \quad (24)$$

Similarly, the left hand side of (22) can be derived as

$$I_{\beta}(C_2; d_1, d_2) - I_{\beta}(C_2; d_1+1, d_2) = \frac{(d_1+d_2-1)!}{d_1!(d_2-1)!} C_2^{d_1} (1-C_2)^{d_2}. \quad (25)$$

By (22), (24), and (25), the joint condition of (13, 14) is equivalent to the joint condition of (13) and

$$C_1^{d_1} (1-C_1)^{d_2} = C_2^{d_1} (1-C_2)^{d_2}. \quad (26)$$

Notice that if $(A_1, A_2) = (C_1, C_2)$, the joint condition of (5, 7) can be written as (13, 26). Thus, A_1 and A_2 in the LRT can guarantee (13) and (14) if we let $A_1 = C_1$ and $A_2 = C_2$. This completes the proof.

Figure 1.

(a) Plot of type I error vs. sample size per group for the three tests in Example 1 ($d = d_1 = d_2$, $\alpha = 0.05$); (b) Plot of power minus type I error vs. level α for the F-test and LRT in Example 2 ($d_1 = 30$, $d_2 = 4$, $\lambda_1 = 12$, $\lambda_2 = 11$); (c) and (d) Plots of power vs. d_1 for the F-test and LRT in Example 3 ($d_2 = 4 \times d_1$, $\alpha = 0.1$, $(\lambda_1, \lambda_2) = (10, 10.5)$ in (c), $(\lambda_1, \lambda_2) = (10, 15)$ in (d)).

Figure 2. Kaplan-Meier curves of the first 2 year survival data from the personalized therapy and an alternative therapy described in Simon et al. (2011).

Table 1

Total sample size (left column), power from the LRT (middle column), and power from the F-test (right column) for testing the equivalence of λ_1 and λ_2 with $(\hat{\lambda}_1, \hat{\lambda}_2) = (22.25, 13.52)$ and proportion of patients being 3: 1 in the two groups.

Total Sample Size	LRT Power	F-Test Power
4	0.112	0.098
8	0.148	0.124
12	0.177	0.152
100	0.687	0.673
104	0.703	0.689
108	0.718	0.704
112	0.732	0.719
116	0.745	0.733
120	0.758	0.747
124	0.771	0.760
128	0.783	0.772
132	0.794	0.784
136	0.805	0.795
140	0.815	0.806
144	0.825	0.816
148	0.834	0.826
152	0.843	0.835
156	0.852	0.844
160	0.860	0.852
164	0.867	0.860
168	0.875	0.868
172	0.882	0.875
176	0.888	0.882
180	0.894	0.889
184	0.900	0.895
188	0.906	0.901