$D_{23} = 0.307$. Comparing this statistic with the tabled critical value 0.199 at level of significance $\alpha = 0.10$ leads to rejecting the null hypothesis. The same conclusion is drawn as that in Example 8.16, where the Weibull distribution was chosen over the exponential distribution for modeling the ball bearing lifetimes. There is additional statistical evidence here indicating that the population of ball bearings is indeed wearing out, and an exponential model is not appropriate.

A similar pattern continues for conducting the Kolmogorov–Smirnov test for other population distributions when the parameters are estimated from the data. The previous discussion for complete data sets does not easily generalize to the case of random right censoring because the distribution of the test statistic becomes more complicated. Many researchers have devised approximate methods for determining the critical values for the Kolmogorov–Smirnov test with random right censoring and parameters estimated from data. Monte Carlo methods for determining $p$-values for the Kolmogorov–Smirnov goodness-of-fit test, which applies to all types of censoring, has been emphasized in this section.

## 11.4 A Test for Exponentiality

The exponential distribution plays a central role in reliability theory, and more generally, in stochastic processes. So rather than use an omnibus test such as the Kolmogorov–Smirnov test to assess the exponentiality of a data set, it is often advantageous to conduct a statistical hypothesis test that has been specifically tailored for the exponential distribution. Dozens of tests for exponentiality have been developed over the years; this section considers one such test.

Assume that a complete data set of $n$ lifetimes has been collected, and the interest is in conducting the hypothesis test

$$H_0 : F(t) = 1 - e^{-\lambda t}$$

versus

$$H_1 : F(t) \neq 1 - e^{-\lambda t}$$

for all $t \geq 0$, where $F(t)$ is the population cumulative distribution function and $\lambda$ is a positive rate parameter. The test for exponentiality described here was developed by Russian mathematician Boris Gnedenko, much of whose work was focused on reliability. This particular test was selected because it is an exact test and there is an intuitive development of the test statistic. We desire to conduct a goodness-of-fit test to determine whether a data set of lifetimes might have come from an exponential population.

Let $T_1, T_2, \ldots, T_n$ be a random sample of lifetimes drawn from an exponential($\lambda$) population. Let $T_{(1)}, T_{(2)}, \ldots, T_{(n)}$ be the associated order statistics. Define $T_{(0)} = 0$. Recall from Property 4.8 of the exponential distribution that the *gaps* between these order statistics, namely $G_i = T_{(i)} - T_{(i-1)}$ for $i = 1, 2, \ldots, n$, are independent random variables, and furthermore

$$G_i \sim \text{exponential}\big((n-i+1)\lambda\big)$$

for $i = 1, 2, \ldots, n$. A simple application of the transformation technique from probability theory indicates that

$$D_i = (n-i+1)G_i$$

for $i = 1, 2, \ldots, n$, yields *normalized gaps* $D_1, D_2, \ldots, D_n$ that are independent and identically distributed exponential($\lambda$) random variables. From Property 4.7 of the exponential distribution,

$$2\lambda \sum_{i=1}^{n} D_i \sim \chi^2(2n).$$

Since this relationship is true for $2\lambda$ times the summation of all $n$ normalized gaps, it is also true for any subset of the normalized gaps. The random variables $D_1, D_2, \ldots, D_n$ can be divided into two groups. For an integer $m$ satisfying $1 \le m \le n-1$, select the first $m$ of these values for group 1 and the remaining $n-m$ values for group 2. The random variables

$$2\lambda \sum_{i=1}^{m} D_i \sim \chi^2(2m) \qquad \text{and} \qquad 2\lambda \sum_{i=m+1}^{n} D_i \sim \chi^2\big(2(n-m)\big),$$

are independent under $H_0$. Since the ratio of two independent chi-square random variables divided by their associated degrees of freedom has the $F$ distribution,

$$G = \frac{2\lambda \sum_{i=1}^{m} D_i/(2m)}{2\lambda \sum_{i=m+1}^{n} D_i/\big(2(n-m)\big)} = \frac{\sum_{i=1}^{m} D_i/m}{\sum_{i=m+1}^{n} D_i/(n-m)} \sim F\big(2m,\, 2(n-m)\big).$$

So $G$ is a pivotal quantity that is independent of $\lambda$. It will be used as the test statistic in conducting Gnedenko's test for exponentiality. The null hypothesis is rejected for small and large values of the test statistic. Small values of the test statistic, along with a histogram having the appropriate shape, would lead one to conclude that the hazard function is decreasing. Large values of the test statistic, along with a histogram having the appropriate shape, would lead one to conclude that the hazard function is increasing. In either case, the exponential model would not be appropriate.

The next example returns to the ball bearing failure times one last time to conduct Gnedenko's test for exponentiality.

**Example 11.6** Use Gnedenko's test to determine whether an exponential distribution is an appropriate model for the ball bearing failure times from Example 8.1 or whether the ball bearings are wearing out.

We have already accumulated four bits of evidence that the ball bearings are wearing out, so we expect that Gnedenko's test for exponentiality will reject $H_0$.

- The histogram in Figure 8.2 is bell-shaped.
- The empirical and fitted survivor functions graphed in Figure 8.1 indicate that the exponential distribution is a poor fit to the ball bearing failure times.
- The 95% confidence region for the ball bearing failure times being fit to the Weibull($\lambda$, $\kappa$) distribution in Figure 8.18 does not include the line $\kappa = 1$.
- The Kolmogorov–Smirnov test conducted in Example 11.5 rejects exponentiality.

The first two of these were visual assessments of the fit and the last two were statistical assessments of the fit.

The manner in which the problem is stated indicates that this is a hypothesis test with a one-tailed alternative in which large values of the test statistic correspond to rejecting $H_0$. Recall that the $n = 23$ sorted ball bearing failure times $t_{(1)}, t_{(2)}, \ldots, t_{(n)}$ (in millions of revolutions) are

$$17.88,\ 28.92,\ 33.00,\ 41.52,\ 42.12,\ \ldots,\ 127.92,\ 128.04,\ 173.40.$$

The gaps between the sorted failure times $g_1, g_2, \ldots, g_n$ are

$$17.88,\ 11.04,\ 4.08,\ 8.52,\ 0.60,\ \ldots,\ 22.08,\ 0.12,\ 45.36.$$

The normalized gaps $d_1, d_2, \ldots, d_n$ are

$$411.24, 242.88, 85.68, 170.40, 11.40, \ldots, 66.24, 0.24, 45.36.$$

No advice is given as to the choice of $m$ in Gnedenko's test for exponentiality. It is an exact test (that is, the probability of rejecting $H_0$ is $\alpha$ when $H_0$ is true) for all values of $m$, but values of $m$ at the extremes ($m = 1$ and $m = n-1$) are more subject to random sampling variability, which results in a lack of power. So a Monte Carlo simulation experiment was conducted to determine which value of $m$ gave the most powerful test under a Weibull alternative (more detail is given in an exercise at the end of the chapter), and $m = 7$ was selected. Using $m = 7$, the test statistic is

$$g = \frac{\sum_{i=1}^{m} d_i/m}{\sum_{i=m+1}^{n} d_i/(n-m)} = 3.76.$$

Since large values of this test statistic lead to rejecting $H_0$, the $p$-value for the test is calculated with

$$p = P[F(14, 32) > 3.76] = 0.00094$$

where $F(14, 32)$ is an $F$ random variable with 14 and 32 degrees of freedom. This tiny $p$-value indicates that the null hypothesis $H_0$ should be rejected. There is overwhelming evidence in this sample that the ball bearing failure times are not drawn from an exponential population. Our earlier assessments are confirmed. There is statistical evidence in the data set that the ball bearings are indeed wearing out.

Example 10.12 considered using a nonhomogeneous Poisson process to model the failure times of copy machines over the first 10,000 actuations on $k = 20$ identical copiers. The estimated cumulative intensity function illustrated in Figure 10.11 is nearly linear. Should the nonhomogeneous Poisson process model be abandoned and a homogeneous Poisson process model be used instead? This question is taken up in the next example.

**Example 11.7** Use Gnedenko's test to determine whether an exponential distribution is an appropriate model for the time between copier failures for the superposition of the failure times in the first 10,000 actuations from Example 10.12.

Recall that the $n = 37$ sorted failure times in the superposition of the $k = 20$ realizations are

50, 102, 220, 415, 974, 1215, 1440, 1452, 1518, 1532, 2009, 2399, 2774,

2793, 2938, 3676, 3791, 3962, 4199, 5000, 5354, 6237, 6517, 6880, 6963,

6982, 7108, 7393, 7463, 7884, 8094, 8187, 8954, 9111, 9201, 9449, 9507.

The question being posed is whether the times between failures is well modeled by an exponential distribution using Gnedenko's test for exponentiality. Taking differences and sorting yields the $n = 37$ times between failures $t_{(1)}, t_{(2)}, \ldots, t_{(n)}$:

$$12, 14, 19, 19, 50, \ldots, 767, 801, 883.$$

The gaps between these values $g_1, g_2, \ldots, g_n$ are

$$12, 2, 5, 0, 31, \ldots, 29, 34, 82.$$

The normalized gaps $d_1, d_2, \ldots, d_n$ are

$$444, 72, 175, 0, 1023, \ldots, 87, 68, 82.$$

Using a Monte Carlo simulation to select $m = 13$ in order to maximize the power of the test under a Weibull alternative as in the previous example, the test statistic is

$$g = \frac{\sum_{i=1}^{m} d_i/m}{\sum_{i=m+1}^{n} d_i/(n-m)} = 1.08.$$

The two-tailed test is appropriate here because we are uncertain as to whether a monotone increasing or decreasing hazard function associated with the alternative hypothesis is appropriate. So the $p$-value for the test is

$$p = 2 \cdot \min\{P[F(26, 48) > 1.08], P[F(26, 48) < 1.08]\} = 0.81,$$

where $F(26, 48)$ is an $F$ random variable with 26 and 48 degrees of freedom. This large $p$-value leads us to fail to reject the null hypothesis of exponentiality using Gnedenko's test. We conclude that fitting the exponential distribution to the time between failures is a reasonable next step toward using a homogeneous Poisson process to model the time between copier failures over the first 10,000 actuations. This conclusion is reinforced by the empirical survivor function and the survivor function which corresponds to the exponential distribution fitted to the times between copier failures in Figure 11.8.

Gnedenko's test for exponentiality is one of dozens of such tests, and this section is only designed to introduce the development of one such test. Two final observations conclude this brief introduction of Gnedenko's test for exponentiality.

- Gnedenko's test for exponentiality is highly dependent on the choice of $m$ because this choice determines the partition of the normalized spacings. This is a weakness of the test. In both of the examples given in this section, a partition about $1/3$ of the way into the normalized spacings seems to give the highest power.

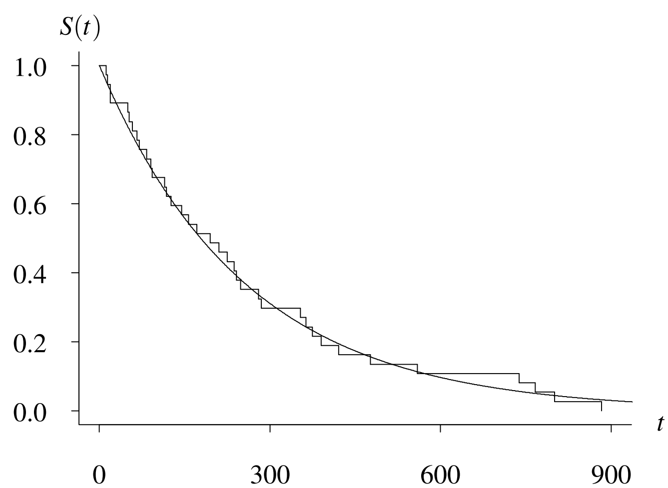


Figure 11.8: Empirical and exponential fitted survivor functions for the copier failure times.