Sec. 6.2 --- Lilliefors Goodness-of-Fit Tests

- **With the Kolmogorov test, the hypothesized distribution must be completely specified, including parameter values.**

- **In some cases, we may want to test whether the data may come from some distribution (e.g., normal, exponential, etc.) without knowing what the specific parameter values may be.**

- **Lilliefors introduced Kolmogorov-type tests to allow testing goodness of fit when the parameters are not specified, but rather estimated from the data.**

Lilliefors Test for Normality

- **Assume that we have a random sample $X_1, X_2, \ldots, X_n$.**

Hypotheses: H_0: The X_i's follow a normal distribution with unspecified mean and variance

vs. H_1: The X_i's follow a non-normal distribution

- **Compute the normalized sample values**

$$Z_i = \frac{X_i - \bar{X}}{S}, \quad i = 1, \ldots, n$$

where

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

• Let $S(x)$ be the e.d.f. of the Z_i's, and let $F^*(x)$ be the c.d.f. of the standard normal distribution; then the Lilliefors test statistic is:

$$T_1 = \sup_x \left| F^*(x) - S(x) \right|$$

• If T_1 is excessively large, then H_0 is rejected.

• The null distribution of T_1 is unknown and has been approximated by random number generation.

Using Table A14, reject H_0 if

$$T_1 > W_{1-\alpha}$$

The P-value is approximated via interpolation in Table A14, or by using R.

Example 1: The data on page 445 consist of 50 observations. At $\alpha = 0.05$, is it reasonable to claim that the data follow a normal distribution?

Test Statistic: $T_1 = 0.081$ from R

Decision Rule: Reject H_0 if $T_1 > \dfrac{.895}{\sqrt{50} - .01 + .83/\sqrt{50}} = .125$

P-value $\approx .567$ from R

Conclusion: Since $T_1 < .125$, we fail to reject H_0. We have no reason to doubt that the data follow a normal distribution

See examples on simulated data on course web page.

Lilliefors Test for Exponential Distribution

- **Assume that we have a random sample $X_1, X_2, \ldots, X_n$.**

Hypotheses: **H_0:** The X_i's follow an exponential distribution with unspecified mean

vs. H_1: The X_i's do not follow an exponential distribution

- **Compute the standardized sample values**

$$Z_i = \frac{X_i}{\overline{X}}, \quad i = 1, \ldots, n$$

- **Let $S(x)$ be the e.d.f. of the Z_i's, and let $F^*(x)$ be the c.d.f. of the standard exponential distribution:**
- **Then the Lilliefors test statistic is:**

$$T_2 = \sup_x \left| F^*(x) - S(x) \right|$$

- **If T_2 is excessively** large**, then H_0 is rejected.**

- **The null distribution of T_2 is complicated, but has been tabulated in Table A15.**

Using Table A15, reject H_0 if

$$T_2 > W_{1-\alpha}$$

The P-value is approximated via interpolation in Table A15.

Example 2: The exponential distribution is a common model for waiting times between occurrences of some random phenomenon. A built-in data set in R gives 272 waiting times between eruptions of the Old faithful geyser. At $\alpha = 0.05$, is it reasonable to claim that the waiting times follow an exponential distribution?

Test Statistic: $T_2 =$ 0.466 from R

bottom Table A15

Decision Rule:

Reject H_0 if $T_2 > \frac{1.0753}{\sqrt{272}} = .065$

Conclusion: Since $.466 > .065$, reject H_0 and conclude the waiting times do <u>not</u> follow an exponential distribution.

- **Lilliefors also presented a similar goodness-of-fit test for a gamma distribution with unknown parameter values. More such tests could be developed similarly.**

- **The Shapiro-Wilk test (`shapiro.test` in R) is another test for normality, based not on the e.d.f. but on the correlation between observed order statistics and expected order statistics under normality.**

- **The Shapiro-Wilk test tends to have better power than the Lilliefors test for detecting departures from normality.**

Sec. 6.3 --- Smirnov Test for Two Samples

• Suppose we have independent random samples (denoted $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$) from two populations.

• The Smirnov test (also called the Kolmogorov-Smirnov two-sample test) uses the e.d.f. of each sample to test whether the two samples come from the same distribution.

• Let $F(\cdot)$ represent the c.d.f. of the X_i's and let $G(\cdot)$ represent the c.d.f. of the Y_i's.

• Assume the samples are mutually independent and the measurement scale is at least ordinal.

• Let $S_1(x)$ represent the e.d.f. of the X_i's and let $S_2(x)$ represent the e.d.f. of the Y_i's.

Possible Hypotheses and Decision Rules

H$_0$: $F(x) = G(x)$ for all x	**H$_0$:** $F(x) \le G(x)$ for all x	**H$_0$:** $F(x) \ge G(x)$ for all x
H$_1$: $F(x) \ne G(x)$ for some x	**H$_1$:** $F(x) > G(x)$ for some x	**H$_1$:** $F(x) < G(x)$ for some x

These alternative hypotheses can also be stated as:

H$_1$: The X's and the Y's do not have the same distribution	**H$_1$:** The X's tend to be smaller than the Y's.	**H$_1$:** The X's tend to be larger than the Y's.

Test Statistic (depends on the alternative hypothesis):

$H_1: F(x) \neq G(x)$ for some x	$H_1: F(x) > G(x)$ for some x	$H_1: F(x) < G(x)$ for some x
$T_1 = \sup_x \lvert S_1(x) - S_2(x) \rvert$	$T_1^+ = \sup_x (S_1(x) - S_2(x))$	$T_1^- = \sup_x (S_2(x) - S_1(x))$

• **The null distributions of T_1, T_1^+ and T_1^- are based on the fact that all orderings of X's and Y's are equally likely if H_0 is true. The null distributions are tabulated in Table A19 for $n = m$ and in Table A20 if $n \neq m$.**

• **The corresponding rejection rules in each case are:**

Reject Ho if $T_1 > W_{1-\alpha}$ ← two-sided	Reject Ho if $T_1^+ > W_{1-\alpha}$ ← one-sided	Reject Ho if $T_1^- > W_{1-\alpha}$ ← one-sided

• **The P-values for each case are approximated by interpolation within the tables or found using R.**

• **If $n = m$, equation (5) on page 458 can be used to obtain a more exact p-value.**

• **If n is too large for the tables, an approximation given at the end of the tables can be used.**

Example 1: Consider the 2 samples given on page 460:

X_i's: 7.6, 8.4, 8.6, 8.7, 9.3, 9.9, 10.1, 10.6, 11.2

Y_i's: 5.2, 5.7, 5.9, 6.5, 6.8, 8.2, 9.1, 9.8, 10.8, 11.3, 11.5, 12.3, 12.5, 13.4, 14.6

Is it reasonable to believe that the two samples follow identical distributions?

$H_0: F(x) = G(x)$ for all x $\qquad$ $H_1: F(x) \neq G(x)$ for some x

$n=9, \quad m=15$

Test statistic: $T_1 = 0.4$ ← from R **P-value** $\approx .2653$ from R

Conclusion: Since $T_1 < W_{.95} = \frac{8}{15} = .533$, we fail to reject H_0. The two samples may follow identical distributions.

Example 2: In a test to compare gasoline types, a driver measured gas mileages for four tanks of "Unleaded"
X → (21.7, 21.4, 23.3, 22.8) and for four tanks of "Premium"
Y → (23.1, 23.5, 22.9, 23.4). Is there evidence that mileage tends to be less for "Unleaded" than for "Premium"?

$n=m=4$

H_0: $F(x) \le G(x)$ for all x **H_1:** $F(x) > G(x)$ for some x

Test statistic: $T_1^+ = 0.75$ **P-value** ≈ 0.1054 from R

Conclusion: $T_1^+ \not> W_{.95} = \frac{3}{4} = 0.75$ ← Table A19, so we fail to reject H_0. There is not significant evidence that the unleaded mileages tend to be less than the premium mileages.

- **This Smirnov test is** exact **if the data are continuous and** conservative **if the data are discrete.**

- **The Mann-Whitney test was another test used to compare two independent samples, but that was sensitive to differences in center.**

- **The Smirnov test is designed to detect any sort of difference in distribution, and is in fact** consistent **against any type of departure from the null hypothesis.**