Fall

STAT 518 --- Chapter 6 --- Goodness-of-Fit Tests

• Often in statistics, we assume a sample comes from a particular distribution.

• Goodness-of-fit tests help us determine whether the assumed distribution is reasonable for the data we have.

Section 6.1: Kolmogorov Goodness-of-Fit Test

• Recall that the empirical distribution function (e.d.f.)

$$\hat{F}(x) = S(x) \longleftarrow \text{book's notation}$$

of a sample is an <u>estimate</u> of the cumulative distribution function (c.d.f.)

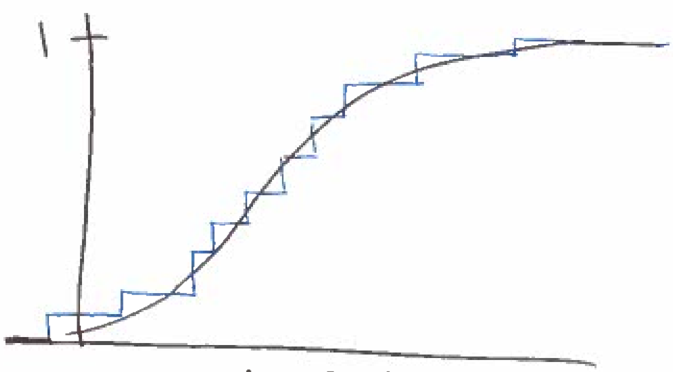
$$F(x) = P(X \le x) \qquad \text{for all } x$$

for the population that the sample came from.

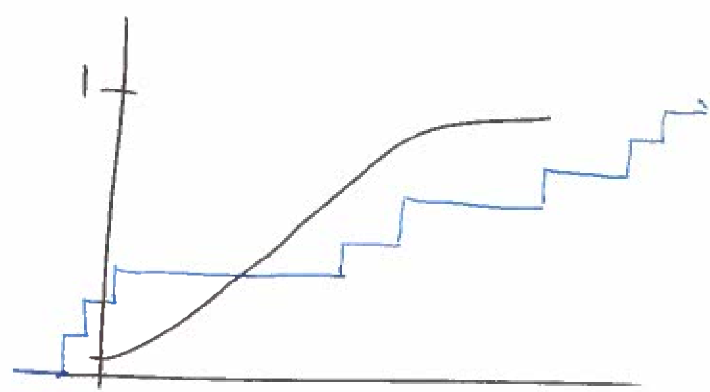
• If $S(x)$ is close to the c.d.f. $F^*(x)$ of our assumed distribution, then our assumption is reasonable.

• If $S(x)$ is far from the c.d.f. $F^*(x)$ of our assumed distribution, then our assumption should be rejected.

Picture:



Good fit



Poor fit

• **How to measure the distance between $S(x)$ and $F^*(x)$?**

• **Kolmogorov suggested using the maximum vertical discrepancy between $S(x)$ and $F^*(x)$ as a test statistic:**

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

• **In Chapter 6 we will see several tests that use a type of maximum vertical discrepancy.**

• **For the Kolmogorov Goodness-of-Fit test, we assume only that we have a random sample $X_1, X_2, \ldots, X_n$.**

Test Statistic (depends on the alternative hypothesis):

(Two-Sided) $\quad T = \sup_x \left| F^*(x) - S(x) \right|$

(One-Sided) $\quad T^+ = \sup_x \left(F^*(x) - S(x) \right)$

(One-Sided) $\quad T^- = \sup_x \left(S(x) - F^*(x) \right)$

• **The null distribution of T is tabulated in Table A13 for $n \leq 40$.**

• **This approach is** exact **if $F(x)$ is continuous and** conservative **if $F(x)$ is discrete.**

• **Pages 435-436 describe an adjustment to improve the test if $F(x)$ is discrete, but we will not cover this.**

• **If $n > 40$, an asymptotic null distribution can be used (see equation (5) on page 431).**

Possible Hypotheses and Decision Rules

$H_0: F(x) = F^*(x)$ for all x	$H_0: F(x) \geq F^*(x)$ for all x	$H_0: F(x) \leq F^*(x)$ for all x
$H_1: F(x) \neq F^*(x)$ for some x	$H_1: F(x) < F^*(x)$ for some x	$H_1: F(x) > F^*(x)$ for some x

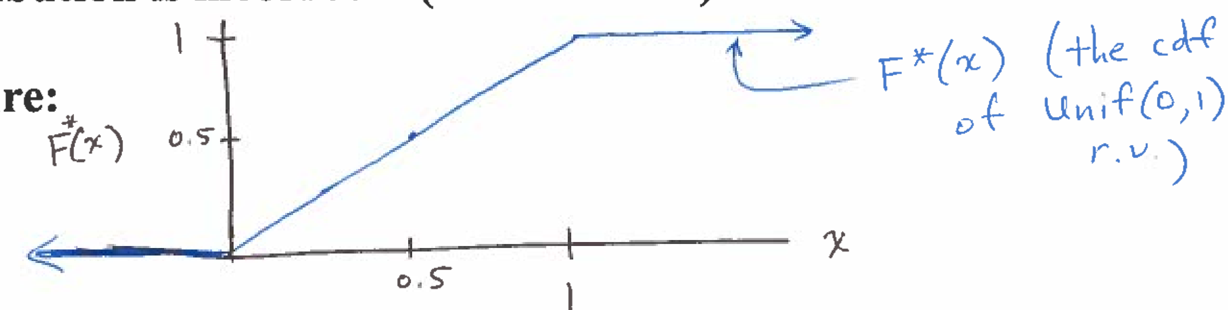
• The corresponding rejection rules in each case are:

Reject H_0 if $T > W_{1-\alpha}$	Reject H_0 if $T^+ > W_{1-\alpha}$	Reject H_0 if $T^- > W_{1-\alpha}$
Table A13, 2-sided	Table A13, One-sided	Table A13, One-sided

• The P-values for each case are approximated by interpolation within Table A13 or found using R.

Example 1: Ten observations are obtained in a sample which supposedly comes from a Uniform(0, 1) distribution. The sorted sample is: `0.203, 0.329, 0.382, 0.477, 0.480, 0.503, 0.554, 0.581, 0.621, 0.710`**. Is there evidence that the hypothesized distribution is incorrect? (Use $\alpha = 0.05$.)**

Picture:



$H_0: F(x) = F^*(x)$ for all x vs. $H_1: F(x) \neq F^*(x)$ for some x

Test statistic: $T = 0.29$ from R.

Decision rule: Reject H_0 if $T > W_{.95} = .409$ (Table A13, 2-sided)

P-value $\approx .307$ from R. Since $0.29 < .409$, we fail to reject H_0. Conclude that the Unif(0,1) distribution is reasonable for these data.

Example 2: A medical team collecting counts of tumors on kidneys has used a Poisson(1.75) distribution to model the counts in the past. They gather such counts on 18 kidneys and obtain this sample: `2,2,4,1,3,1,4,0,2,2,1,1,0,2,2,3,3,3`. Is there evidence that the count distribution is actually stochastically larger than previously thought? (Use $\alpha = .05$.)

"Stochastically larger than previous"

$\Rightarrow$ For any x, "Real $P(X > x)$" > "Previous (Assumed) $P(X > x)$"

$\Rightarrow$ "Real $P(X \le x)$" < "Assumed $P(X \le x)$"

$\Rightarrow$ H_1 is : $F(x) < F^*(x)$ The Pois (1.75) cdf

H_0: $F(x) \ge F^*(x)$ for all x vs. H_1: $F(x) < F^*(x)$ for some x.

Decision Rule: Reject H_0 if $T^+ > w_{.95} = 0.279$ (Table A13 one-sided, $n=18$)

Test Statistic: $T^+ = 0.4106$ from R

Since $.4106 > .279$, reject the H_0 and conclude the true distribution tends to produce larger counts than the Pois(1.75) distribution. P-value $\approx .0015$ from R.

In R, see the function `ks.test` to perform this test.

A Confidence Band for the True Population c.d.f.

- **From Table A13, it is easy to obtain an upper function and lower function that form a $(1 - \alpha)100\%$ confidence band.**

- **We can be, say, 95% confident that the entire true c.d.f. will fall within the 95% confidence band.**

• **To form the band, simply draw the e.d.f $S(x)$, and add** $W_{1-\alpha}$ from Table A13 (two-sided) **to each point to form the upper boundary of the band.**

• **Subtract** $W_{1-\alpha}$ **from $S(x)$ at each point to form the lower boundary.**

• **If** $S(x) - W_{1-\alpha}$ **goes below 0, simply draw the lower boundary at 0 in those places.**
• **If** $S(x) + W_{1-\alpha}$ **exceeds 1, simply draw the upper boundary at 1 in those places.**

• **See example in R using Example 1 data.**

Properties of the Kolmogorov Test

• **The two-sided Kolmogorov test is** consistent **against all differences between the hypothesized $F^*(x)$ and the true $F(x)$.**

• **However, the test is** biased **for finite sample sizes.**

• **The Kolmogorov test is** more powerful **than the chi-squared goodness-of-fit test (covered in Chapter 4) when the data are ordinal.**

• **The Kolmogorov test is exact even for small samples, while the chi-square test requires a large sample size.**