Testing Goodness-of-Fit (continued)

- Statistical Tests
  - The chi-square test
  - The Kolmogorov-Smirnov test
  - The Anderson-Darling test

The Chi-square Test

- Split the range of X into k adjacent intervals
- Let

$$I_i = [a_{i-1}, a_i) = \text{ ith interval}$$

$$O_i = \text{number of observations in interval } i$$

$$E_i = \text{ expected number of observations in interval } i$$

$$= n[\hat{F}(a_i) - \hat{F}(a_{i-1})]$$

CDF of fitted distribution

The Chi-square Test (continued)

- The null hypothesis is rejected (at level α) if

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} > \chi^2_{k-s-1,\alpha}$$

where s is the number of parameters replaced by their MLEs

  - One should use $E_i \geq 5$
  - The test has maximum power if the E_i are equal (the intervals are equiprobable)

The Kolmogorov-Smirnov Test

- Is applicable to continuous distributions only
- It generally assumes that all parameters are known
- Sort the data and define the empirical CDF

$$\bar{F}(x) = \frac{\text{number of } X_i \le x}{n}$$

$$= \begin{cases} 0 & \text{if } x < X_{(1)} \\ \dfrac{i}{n} & \text{if } X_{(i)} \le x < X_{(i+1)},\ 1 \le i \le n-1 \\ 1 & \text{if } x > X_{(n)} \end{cases}$$

The Kolmogorov-Smirnov Test (continued)

- The null hypothesis is rejected (at level α) if

$$D_n = \sup \left| \hat{F}(x) - \bar{F}(x) \right|$$

$$= \max \left\{ \max \left[\frac{i}{n} - \hat{F}(X_{(i)}) \right], \max \left[\hat{F}(X_{(i)}) - \frac{i-1}{n} \right] \right\} > \underbrace{d_{n,\alpha}}_{\text{tabulated}}$$

The Kolmogorov-Smirnov Test (continued)

- We usually simplify the above inequality by computing an adjusted test statistic and a modified critical value c_α:

$$\text{Adjusted Test Statistic} > \underbrace{c_\alpha}_{\text{tabulated}}$$

- When parameters are replaced by MLEs modified K-S test statistics exist for the following distributions:
  - Normal
  - Exponential
  - Weibull
  - Log-logistic

The Kolmogorov-Smirnov Test (continued)

Modified critical values for adjusted K-S test statistics

		Type I error α				
Case	Adjusted Test Statistic	0.150	0.100	0.050	0.025	0.001
All parameters known	$\left(\sqrt{n}+0.12+\frac{0.11}{\sqrt{n}}\right)D_n$	1.138	1.224	1.358	1.480	1.628
$\text{Nor}(\bar{X}_n, S_n^2)$	$\left(\sqrt{n}-0.01+\frac{0.85}{\sqrt{n}}\right)D_n$	0.775	0.819	0.895	0.955	1.035
$\text{Expo}(1/\bar{X}_n)$	$\left(D_n-\frac{0.2}{n}\right)\left(\sqrt{n}+0.26+\frac{0.5}{\sqrt{n}}\right)$	0.926	0.990	1.094	1.190	1.308

Example

- The following observations are times-to-failure (in days) for a piece of equipment: 0.83, 0.32, 4.35, 2.34, 0.75
- We wish to test the fit of the exponential distribution
- Since the parameter of the distribution has not been specified, we compute the MLE

$$\hat{\lambda} = 1 / \bar{X}_5 = 0.582$$

- The fitted CDF is

$$\hat{F}(x) \equiv F(x; \hat{\lambda}) = 1 - e^{-0.582x}, x > 0$$

- We sort the data in increasing order:

$$0.32 < 0.75 < 0.83 < 2.34 < 4.35$$

Example (continued)

$X_{(i)}$	0.32	0.75	0.83	2.34	4.35
$\hat{F}(X_{(i)})$	0.170	0.354	0.383	0.744	0.921
$\frac{i}{5} - \hat{F}(X_{(i)})$	0.030	0.046	**0.217**	0.056	0.079
$\hat{F}(X_{(i)}) - \frac{i-1}{5}$	**0.170**	0.154	–	0.144	0.121

The test statistic is $D_5 = 0.217$ and the adjusted test statistic is

$$\left(D_5 - \frac{0.2}{5}\right)\left(\sqrt{5} + 0.26 + \frac{0.5}{5}\right) = 0.332$$

Since $0.332 \le c_\alpha$ for $\alpha \le 0.15$, we fail to reject the hypothesis that the data come from the exponential distribution