Testing exponentiality via Khmaladze transformation.
Calculation of Kolmogorov-Smirnov statistics.

Background

Suppose $X_1, X_2, \ldots, X_n$ is a sample (a sequence of i.i.d. random variables). Suppose we wish to test whether each X_i follows an exponential distribution function $P_\lambda(x) = 1 - e^{-\lambda x}$ with some unspecified $\lambda > 0$. Moreover, suppose we wish to test this hypothesis of exponentiality using a goodness of fit test, i.e. using a test which should be able to detect "all sorts" of alternatives to exponentiality.

One can do this, for example, by using Kolmogorov-Smirnov statistics from the process

$$w_n(x) = \sqrt{n}[P_n(x) - K(x, P_n)]$$

where

$$K(x, P_n) = \hat{\lambda} \int_0^\infty \left(2 + \frac{\hat{\lambda}}{2}\min(x,y) - \hat{\lambda} y\right) \min(x,y) P_n(dy)$$

$$= \hat{\lambda} \int_0^x \left(2 - \frac{\hat{\lambda}}{2} y\right) y P_n(dy) + \hat{\lambda}\left(2 + \frac{\hat{\lambda}}{2} x\right) x\{1 - P_n(x)\} - \hat{\lambda}^2 x \int_x^\infty y P_n(dy)$$

or

$$K(x, P_n) = \frac{\hat{\lambda}}{n} \sum_{i: X_i \le x} \left(2X_i - \frac{\hat{\lambda}}{2} X_i^2\right) + \hat{\lambda}\left(2 + \frac{\hat{\lambda}}{2} x\right) x\{1 - P_n(x)\} - x\frac{\hat{\lambda}^2}{n} \sum_{i: X_i > x} X_i\,.$$

This form of empirical process is based on the so called Khmaladze transformation.

Calculation of the Kolmogorov-Smirnov statistics

$$d_n^+ = \sup_x w_n(x), d_n^- = -\inf_x w_n(x) \;\text{ and }\; d_n = \sup_x |w_n(x)| = \max(d_n^+, d_n^-)$$

becomes easy and quick if we note the following: within each interval $(X_{(j)}, X_{(j+1)})$ formed by adjacent order statistics the compensator $K(x, P_n)$ is a quadratic function in x and its minimum is attained at the point

$$x_j^0 = \frac{\sum_{i: X_i > X_{(j)}} X_i}{n - j} - 2\bar{X}\ ,$$

while $P_n(x)$ stays constant. It follows that the maximum or minimum of the difference $P_n(x) - K(x, P_n)$ on each such interval is attained either at the end-points or in x_j^0, provided $x_j^0 \in (X_{(j)}, X_{(j+1)})$. In other words,

$$d_n^+ = \max(a_n, b_n)$$

where

$$a_n = \max_j\{\frac{j}{n} - K(X_{(j)}, P_n)\}, \text{ and } b_n = \max_{j: x_j^0 \in (X_{(j)}, X_{(j+1)})}\{\frac{j}{n} - K(x_j^0, P_n)\}$$

while

$$d_n^- = -\min_j\{(j-1)/n - K(X_{(j)}, P_n)\}.$$