TESTING IN NONPARAMETRIC ACCELERATED LIFE TIME MODELS

Hannelore Liero
Mathematical Institute, University of Potsdam
Potsdam, GERMANY
e-mail: liero@uni-potsdam.de

We consider a random life time T which depends on some explanatory variable or covariate X in such way that X reduces a basic life time, say T_0, by a factor $\psi(X)$. In other words, the life time T satisfies the equation

$$T = \frac{T_0}{\psi(X)}.$$

Examples for X are the dose of a drug, temperature, pressure or stress. The function ψ is called *acceleration function.* We will assume that T is an absolute continuous random variable and that the covariate X does not depend on the time; and for simplicity of presentation let X take one dimensional values.

The conditional survival function of T given $X = x$ has the form

$$S(t|x) = \mathsf{P}(T > t|X = x) = S_0(t\psi(x))$$

where S_0 is the survival function of T_0, the so-called *baseline survival function.*

The aim of the talk is to propose a test procedure for testing whether the function ψ belongs to a pre-specified parametric class of functions

$$\mathcal{F} = \{\psi \,|\, \psi(\cdot) = \psi(\cdot;\beta)\ \beta \in \mathbb{R}^d\}.$$

The test is based on the following transformation: The conditional expectation of $Y = \log T$ is given by

$$\mathsf{E}\,(Y|X = x) = \mu_0 - \log\psi(x) \quad \text{with} \quad \mu_0 = -\int \log z \,\mathrm{d}S_0(z) = \mathsf{E}\,(\log T_0),$$

and we can translate the considered test problem into a problem of testing the regression function in a nonparametric regression model

$$Y = m(X) + \varepsilon,$$

where

$$m(x) = \mu_0 - \log\psi(x) \quad \text{and} \quad \mathsf{E}\,(\varepsilon|x) = 0.$$

Now, define the set of hypothetical regression functions

$$\mathcal{M} = \{\, m(\cdot;\vartheta) = \mu_0 - \log\psi(\cdot;\beta),\ \vartheta = (\mu_0,\beta) \in \mathbb{R}^{d+1}\}.$$

Then the "new" test problem has the form

$$H: \quad m \in \mathcal{M} \qquad \text{versus} \qquad K: \quad m \notin \mathcal{M}.$$

Before we will give the test statistic let us describe the underlying data: We do not observe the pairs (X_i, T_i) but consider a model in which the life times T_i are subject to random right censoring, so that the observable random variables are given by

$$(X_i, T_i^*, \Delta_i) \quad \text{with} \quad T_i^* = \min(T_i, C_i) \quad \text{and} \quad \Delta_i = 1(T_i \leq C_i), \quad i = 1, \ldots, n.$$

Here C_i is a nonnegative random variable, representing the censoring time. To identify the conditional distribution of T given X we will assume that C and (X, T) are independent, or at least that T and C are conditionally independent given X.

The test statistic is based on the following idea: Since the alternative consists of "all" regression functions we compare a nonparametric estimator of m, say $\hat{m}_n$, with an estimator for the hypothetical parametric $m(\cdot; \vartheta)$, i.e. $m(\cdot; \hat{\vartheta}_n)$. But since the estimate $\hat{m}_n$ is a result of smoothing it seems to be appropriate to compare $\hat{m}_n$ with a smoothed version of $m(\cdot; \hat{\vartheta}_n)$, denoted by $\tilde{m}_n(\cdot; \hat{\vartheta}_n)$. As deviation measure we take the weighted integrated squared distance. Proving the asymptotic normality of this (properly standardized) statistic we obtain an asymptotic α-test.

This idea is realized as follows: Since

$$m(x) \;=\; -\int_0^\infty \log t \; \mathrm{d}S(t|x)$$

a nonparametric estimator of m is given by

$$m(x) \;=\; -\int_0^\infty \log t \; \mathrm{d}\hat{S}_n(t|x)$$

where $\hat{S}_n$ is the weighted Kaplan-Meier estimator for the conditional survival function. As weights, denoted by W_{ni}, we take kernel weights. The Kaplan-Meier estimator is the extension of the empirical survival function for censored data, and the weighted Kaplan-Meier estimator is the analog for conditional survival functions.

Further, let $\hat{\vartheta}_n$ be the least squares estimator of the unknown parameter in the hypothetical regression model, and set

$$\tilde{m}_n(x; \hat{\vartheta}_n) = \sum_i W_{ni}(x) m(X_i, \hat{\vartheta}_n).$$

The test statistic is defined by

$$Q_n \;=\; \int [\hat{m}_n(x) \;-\; \tilde{m}_n(x; \hat{\vartheta}_n)]^2 a(x)\,\mathrm{d}x,$$

here a is a known weight function which is introduced to control the region of integration. Using standard methods for deriving asymptotic normality of quadratic forms we obtain

$$\rho_n\,(Q_n \;-\; e_n) \xrightarrow{\mathcal{D}} \mathrm{N}(0,1)$$

for standardizing terms ρ_n and e_n, which depend on the underlying unknown distribution. Replacing these unknown terms by appropriate estimates $\hat{\rho}_n$ and $\hat{e}_n$, respectively, we get the desired asymptotic α-test: Reject H if

$$Q_n > z_\alpha \hat{\rho}_n + \hat{e}_n,$$

where z_α is the fractile of the standard normal distribution.

The proposed procedure will be completed by simulation results concerning this "limit-distribution-approach" and some resampling methods for the test statistic Q_n.