Fundamentals of Statistical Modeling (VT20)

Andrea Discacciati
Karolinska Institutet
Stockholm, Sweden

Lab 3 (Extra material on flexible modeling with splines)

Load the dataset and the `mlci` command

{{1}}

Assume that $f(age)$ follows a generalized extreme values distribution. Estimate the parameters μ and σ. Constrain σ to be positive. Inflate the probability of death during the first year of life, while constraining it to be between 0 and 1.

{{2}}

Generate the estimated density and the transform $u1 = \hat{F}(y)$ (we'll use it to assess the goodness-of-fit).

Remember: we're assuming that, for $age \geq 1$, the variable age is Standard-Exponential–distributed after we apply the transform $G(y)$. The CDF of a Standard Exponential is $F(y) = 1 - \exp(-y)$

{{3}}

Now, let's include a spline transformation of age with 4 degrees of freedom and let's see whether this improves the fit of our generalized extreme values model. Jointly test the 3 parameters η_1, η_2, η_3 to assess whether adding the 3 RCS transforms improves the fit of this model with respect to the "base" model (see above).

We need to help Stata a little by providing reasonable initial values for the model's parameters.

{{4}}

Generate the estimated density and the transform $u2 = \hat{F}(y)$.

{{5}}

Plot the 2 estimated densities over the sample histogram and the quantile plot.

{{6}}