Lab1: Estimating the survival function using targeted MLE

September 18, 2006

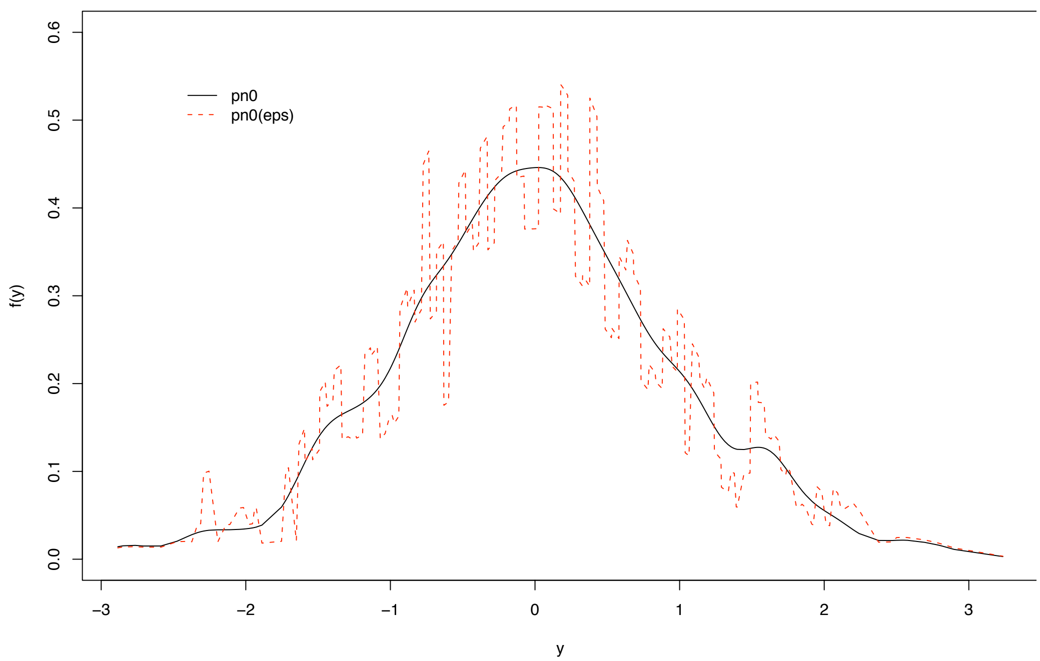
1 Introduction

This lab involves using a targeted MLE approach to estimating the empirical survival distribution using an initial smooth density estimate. The goal is to implement, in R an algorithm that.

- Provides an initial density estimate using kernel density estimation with arbitrary bandwidth, p_n^0.
- Derive the estimate of the survival function at several (say 10) points (t) using numerical integration $S_n^0(t) = \int I(u > t)p_n^0(u)du$.
- Derive a targeted MLE, by first posing one-dimensional sub-models: $p_n^0(\epsilon)(u) = [1+\epsilon^T D(p_n^0(u))]p_n^0(u)$ and find the MLE for ϵ, say ϵ_n. Use the following closed form estimator, $\epsilon_n = \Sigma(p_n^0)^{-1} P_n D(p_n^0)$, where $\Sigma(p_n^0) = E_{p_n^0} D(p_n^0) D(p_n^0)^T$ and $D(p_n^0)(t) = I(Y > t) - S_n^0(t)$.
- Find the new targeted MLE of the survival function as: $S_n^0(\epsilon_n)(t) = \int I(u > t)p_n^0(\epsilon_n)(u)du$.

For now, pick an arbitrary bandwidth (say 0.15) and a sample size of 1000 (we are going to follow-up with methods to choose the bandwidth in this setting). Show the original density (p_n^0) and the density resulting from targeted MLE ($p_n^0(\epsilon_n)$) on one plot (see figure below as an example). As well, show the survival function at all original data points from the original smoothed density ($S_n^0(t)$), the same but based on the targeted estimate ($S_n^0(\epsilon_n)(t)$) and

Figure 1: Original Smoothed density and the density after targeted MLE



the empirical survival distribution (you should see that the latter matches $S_n^0(\epsilon_n)(t)$ at the 10 chosen times).

Please provide the code you used as well as a short write-up that describes the results.