A.II. Kernel Estimation of Densities

Olivier Scaillet

University of Geneva and Swiss Finance Institute

Outline

Introduction

The moments of a random variable are a summary of its distributional behavior.

Full information is provided by its distribution.

The *cumulative distribution function* for a single asset i corresponds to

$$F_i(\zeta_i) = P(Y_{i,t} \leq \zeta_i),$$

while for two assets i and j, we have

$$F_{ij}(\zeta_i, \zeta_j) = P(Y_{i,t} \leq \zeta_i, Y_{j,t} \leq \zeta_j).$$

Introduction

A cdf may be expressed as an expectation

$$F_i(\zeta_i) = \int_{-\infty}^{\zeta_i} f(y_i)dy_i = \int_{-\infty}^{+\infty} 1_{y_i \leq \zeta_i} f(y_i)dy_i = E\left[1_{y_i \leq \zeta_i}\right],$$

where $1_{Y_{i,t} \leq \zeta_i}$ = indicator function of the set $\{Y_{i,t} : Y_{i,t} \leq \zeta_i\}$

$$1_{Y_{i,t} \leq \zeta_i} = \begin{cases} 1 \text{ if } Y_{i,t} \leq \zeta_i \\ 0 \text{ otherwise} \end{cases}$$

Problems with Empirical Averages

As previously in order to estimate expectations, we need to replace E by an empirical average:

$$\hat{F}_i(\zeta_i) = \frac{1}{T}\sum_{t=1}^{T} 1_{y_{i,t}\leq\zeta_i},$$

$$\hat{F}_{ij}(\zeta_i,\zeta_j) = \frac{1}{T}\sum_{t=1}^{T} 1_{y_{i,t}\leq\zeta_i, y_{j,t}\leq\zeta_j},$$

$\Longrightarrow$ We obtain step functions which are *not* differentiable.

$\Longrightarrow$ We cannot build empirical counterparts of densities, i.e., $f_i(\zeta_i) = \frac{dF_i(y_i)}{dy_i}\big|_{y_i=\zeta_i}$

Kernel Estimator

To build empirical counterparts of densities, we rely on kernel estimation.

Idea behind:
We start from the histogram,

$$\hat{f}_i(\zeta_i) = \frac{1}{T}\sum_{t=1}^{T} 1_{y_{i,t}=\zeta_i}$$

and replace bars by smooth bumps

$$\hat{f}_i(\zeta_i) = \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{y_{i,t}-\zeta_i}{h}\right)$$

The bump K is called a *Kernel.* It should be positive and integrate to one.

Gaussian Kernel Example

Gaussian Kernel = Gaussian density

$$K\left(\frac{y_{i,t}-\zeta_i}{h}\right)=\frac{1}{\sqrt{2\pi}}\exp-\frac{1}{2}\left(\frac{y_{i,t}-\zeta_i}{h}\right)^2$$

The smoothing parameter h is called the *bandwidth.*

The bandwidth h plays the same role as the class length for histograms.

Optimal Kernel Bandwidth

If h is too large (large class), we get *oversmoothing*.

If h is too small (small class), we get *undersmoothing*.

Rule of thumb to select the bandwidth:

$$h = \hat{\sigma} T^{-1/5}$$

where $\hat{\sigma}$ is empirical standard deviation of the data.

Kernel Estimation of a Bivariate Density

It is possible to extend to higher dimensions and to the conditional case.

$$\hat{f}_{ij}(\zeta_i, \zeta_j) = \frac{1}{Th^2} \sum_{t=1}^{T} K\left(\frac{y_{i,t} - \zeta_i}{h}\right) K\left(\frac{y_{j,t} - \zeta_j}{h}\right)$$

Note that the *curse of dimensionality* appears when we are above five dimensions.

We need a lot of information (data) to get an accurate estimation of the high dimensional object to be estimated.

Kernel Estimation of a Conditional Density

Recall the definition (Bayes Theorem)

$$f(\zeta_i \,| y_{j,t} = \zeta_j) = \frac{f_{ij}(\zeta_i, \zeta_j)}{f_j(\zeta_j)}$$

$\Longrightarrow$ we only need to replace the unknown quantities by their estimates

$$\hat{f}(\zeta_i \,| y_{j,t} = \zeta_j) = \frac{\hat{f}_{ij}(\zeta_i, \zeta_j)}{\hat{f}_j(\zeta_j)}$$

Extension I: Zero Boundary Kernel

Previous estimators have good properties when the data take values in $\Re$.

When data are bounded from below at zero (losses with a positive sign), they exhibit boundary bias (edge effect).

This *boundary bias* is due to weight allocation by the fixed symmetric kernel outside the density support when smoothing is carried out near the boundary.

Extension I: Zero Boundary Kernel

One of the remedy consists in replacing symmetric kernels by asymmetric kernels, which never assigns weight outside the support.

The form of the estimators is the same

$$\hat{f}_i(\zeta_i) = \frac{1}{Th} \sum_{t=1}^{T} K\left(y_{i,t}; \zeta_i, h\right)$$

but K is replaced by an asymmetric kernel.

Zero Boundary Kernel Examples

Gamma Kernel:

$$K\left(y;\zeta,h\right)=\frac{y^{\varsigma/h}e^{-y/h}}{h^{\varsigma/h+1}\Gamma(\varsigma/h+1)}$$

where $\Gamma(x)=\int_0^\infty e^{-u}u^{x-1}du$

Reciprocal Inverse Gaussian Kernel:

$$K\left(y;\zeta,h\right)=\frac{1}{\sqrt{2\pi hy}}\exp\left(-\frac{\varsigma-h}{2h}\left(\frac{y}{\varsigma-h}-2+\frac{\varsigma-h}{y}\right)\right)$$

Extension II: Compact Support

When the data are defined on $[0,1]$, we face two boundaries.

It is then useful to use a kernel whose support is also $[0,1]$, for example the Beta kernel:

$$K\left(y;\zeta,h\right)=\frac{1}{B(\varsigma/h+1,(1-\varsigma)/h+1)}y^{\varsigma/h}(1-y)^{(1-\varsigma)/h}$$

where $B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

Application: Default Recovery Rates

This estimator is useful to analyze the distribution of recovery rates at default.

There is a renewed interest in LGD (loss given default), which is mainly prompted by Basle II and the explosion of the credit derivatives market.

Data are scarce, in particular outside the US. The market standard to model LGD is a parametric assumption of beta distributed recoveries.

There are several measures of LGD
$\rightarrow$ ultimate recoveries
$\rightarrow$ trading price recoveries

Default Recovery Rates

These measures often give very different results. Which one should be used depends who you are and what you do with your defaulted positions.

The data concern 623 US defaulted bond issues spanning form 1981 to end 1999. These are trading price recoveries which are classified by industry and seniority.

The data comes from the S&P/PMD database.

The market assumption of a beta distribution is often severely wrong.This could lead to underestimation of risk measures.