Lecture 2

Jussi Klemelä

January 27, 2009

1 Concepts of regression function estimation

Conditional density. Conditional expectation can be written in terms of the conditional density as

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y \, f_{Y \mid X=x}(y) \, dy,$$

where the conditional density can be defined as

$$f_{Y \mid X=x}(y) = \frac{f_{Y,X}(y,x)}{f_X(x)}, \qquad y \in \mathbf{R},$$

when $f_X(x) > 0$ and $f_{Y \mid X=x}(y) = 0$ otherwise, where $f_{Y,X} : \mathbf{R}^{d+1} \to \mathbf{R}$ is the joint density of (Y, X) and $f_X : \mathbf{R}^d \to \mathbf{R}$,

$$f_X(x) = \int_{\mathbf{R}} f_{Y,X}(y,x) \, dy, \qquad x \in \mathbf{R}^d,$$

is the density of X.

Heteroscedastic noise. In the regression setting we try to estimate a function which is observed under additive noise. We can write the response variable as

$$Y = f(X) + \epsilon,$$

where $\epsilon \in \mathbf{R}$ is an error term and we assume that

$$E(\epsilon \mid X = x) = 0.$$

If ϵ and X are independent, then the regression model is homosekedastic and when ϵ and X are dependent, then the regression model is heteroskedastic. An example of a heteroscedastic regression model is

$$Y = f(X) + \sigma(X) \, \xi,$$

where $\sigma : \mathbf{R}^d \to (0, \infty)$ and $\xi \in \mathbf{R}$ is independent from X.

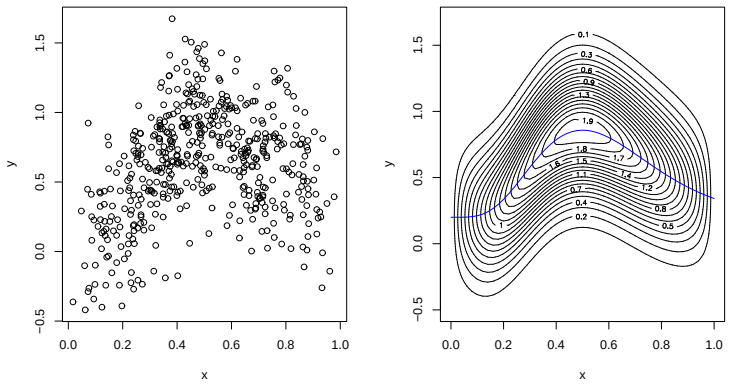

Figure 1: *(Scatter plot of regression data.)* Frame a) shows a scatter plot of simulated data. Frame b) shows a contour plot of the joint density of an explanatory variable and a response variable. The regression function is shown as the solid line.

Fixed design. We have assumed that the design points $X_1, \ldots, X_n \in \mathbf{R}^d$ are random variables (realizations of random variables). In some cases the design points can be chosen by the conducter of the experiment, and they are not chosen by a random mechanism. In this case we say that the regression is fixed design regression, in contrast to random design regression.

Illustration. Figure 1 shows a simulated data of size $n = 500$ and a contour plot of the density function of $(X, Y) \in [0, 1]^2$. The density of X is a scaled and translated Bartlett density: $f_X(x) = 3(1 - [(x - 0.5)/h)]^2)_+/(4h)$, where $h = 0.5$. The distribution of the errors ξ is standard Gaussian and $\epsilon = 0.3$. The regression function is $g(x) = 0.2 + \exp\{-2(x/h)\}(x/h)^5$, where $h = 0.2$. The regression function is shown in the figure as a blue line.

2 Kernel estimator

Introduction. Let us first consider the problem where we want to estimate the value of a function $f : \mathbf{R}^d \to \mathbf{R}$ at point $z \in \mathbf{R}^d$, when we have available only the values $f(x_1), \ldots, f(x_n)$, for some collection of points $x_1, \ldots, x_n \in \mathbf{R}^d$. The classical interpolation methods, including piecewise constant and polynomial interpolation of the values $f(x_1), \ldots, f(x_n)$, can be used to obtain an approximation of the value $f(z)$ at any point $z \in \mathbf{R}^d$. For example, we

can estimate

$$f(z) \approx f(x_{i(z)}),$$

where $i(z)$ is such that

$$\|z - x_{i(z)}\| = \min\{\|z - x_i\| : i = 1, \ldots, n\}.$$

In the setting of regression function estimation we do not observe the exact values of the function, but only values with are corrupted with noise:

$$Y_i = f(x_i) + \epsilon_i,$$

where ϵ_i, $i = 1, \ldots, n$, are random errors. We could use the estimate

$$f(z) \approx Y_{i(z)},$$

where $i(z)$ is as before, but this estimator would contain too much random variation and it is a better idea to take a local average over several observations:

$$\hat{f}(z) = \sum_{i=1}^{n} p_i(z) Y_i, \tag{1}$$

where $p_i(z) \geq 0$, $\sum_{i=1}^{n} p_i(z) = 1$, are weights which are close to zero when x_i is distant from z.

Regressogram. We define a piecewise constant regression function estimator as

$$\hat{f}(x) = \sum_{R \in \mathcal{P}} \hat{Y}_R I_R(x), \qquad x \in \mathbf{R}^d,$$

where $\mathcal{P}$ is a partition of $\mathbf{R}^d$ and

$$\hat{Y}_R = \frac{1}{n_R} \sum_{i: X_i \in R} Y_i,$$

with

$$n_R = \#\{X_i : X_i \in R,\ i = 1, \ldots, n\}.$$

If $x \in R$, then the value of the regressogram is

$$\hat{f}(x) = \hat{Y}_R.$$

It can happen that x is near the boundary of R and it would seem to be better to use a moving average over those values of the response variable where

the corresponding points of the explanatory variables are in a symmetric neigborhood of x:

$$f(x) \approx \hat{Y}_{R_x},$$

where

$$R_x = [x-h, x+h] = [x_1-h, x_1+h] \times \cdots \times [x_d-h, x_d+h],$$

where $h > 0$. This moving average is equal to the kernel estimator, to be defined in (2), when the kernel function is $K = I_{[-1,1]}$.

Kernel estimator. The kernel estimator of the regression function is defined as

$$\hat{f}(x) = \sum_{i=1}^{n} p_i(x)\, Y_i, \tag{2}$$

where

$$p_i(x) = \frac{K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)}, \qquad i = 1, \ldots, n, \tag{3}$$

$K : \mathbf{R}^d \to \mathbf{R}$ is the kernel function, $K_h(x) = K(x/h)/h^d$, and $h > 0$ is the smoothing parameter. The kernel estimator is called also a Nadaraya-Watson estimator, since it was defined by Nadaraya (1964) and Watson (1964).

Moving average The kernel regression estimator is quite similar than a moving average in the time series setting. If we have a time series $U_1, \ldots, U_T$, then the k period moving average is

$$\mathrm{ma}(t) = \sum_{i=t-k+1}^{t} p_t(i)\, U_i, \qquad t = k, \ldots, T,$$

where $p_t(i) \geq 0$, $\sum_{i=t-k+1}^{t} p_t(i) = 1$. The exponential moving average is obtained by taking

$$p_t(i) = w(t-i), \qquad w(i) = \frac{\exp(-i)}{\sum_{i=0}^{k-1} \exp(-i)}, \qquad i = 0, \ldots, k-1.$$

A moving average of time series can be a regression function estimator (in the fixed design setting), if we have a model $U_t = f(t) + \epsilon_t$ for the times series U_t. A moving average can also be used as an explanatory variable in portfolio selection. See for example Aït-Sahalia and Brandt (2001) and Franke, Härdle and Hafner (2004), Section 18.4.

3 Illustration

We look at the following code in

http://cc.oulu.fi/ jklemela/finatool/

```
# we obtain returns of the DAX stock index

ticker<-c("^GDAXI")
destfile<-"~/pois"
ry<-read.yahoo(ticker, source="web", destfile=destfile)
#save(file="/home/jsk/Arti/statfina/Dax.var",list=c("ry"))
#load(file="/Users/jsk/Karhu/Arti/statfina/DaxMdax.var")
dm<-data.manip(ry,ticker)
method<-"return"
df<-data.final(dm,ticker,method=method)
n<-dim(df)[1]
S<-matrix(df[1:n,1],n,1)
plot(S,type="l")

# we calculate volatilities for the 5 day periods

perlen<-5
pernum<-floor(n/perlen)
volas<-matrix(0,pernum,1)
for (i in 1:pernum){
    beg<-(i-1)*perlen+1
    end<-(i-1)*perlen+perlen
    period<-S[beg:end]
    volas[i]<-sqrt(sum(period^2)/perlen)*sqrt(252)
}
plot(volas,type="l")

# 1D kernel estimation

# we try to pedict the volatility of a 5 day period with the
# help of the volatility of the previous 5 day period

dendat<-matrix(0,pernum-1,2)
for (i in 1:(pernum-1)){
    dendat[i,1]<-volas[i]
```

```
    dendat[i,2]<-volas[i+1]
}
plot(dendat)

cor(dendat[,1],dendat[,2])
x<-seq(0,1,0.05)
y<-as.real(cor(dendat[,1],dendat[,2]))*x
matplot(dendat[,1],dendat[,2])
matplot(x,y,add=TRUE,type="l")

# we estimate the regression function

x<-dendat[,1]
y<-dendat[,2]
h<-0.05
N<-60
pcf<-pcf.kernesti(x,y,h,N)
dp<-draw.pcf(pcf)          # requires package "denpro"
matplot(dp$x,dp$y,type="l",xlim=c(0,1.1),ylim=c(0,1.1))
matplot(x,y,add=TRUE)

# we make a logarithmic transform

x<-log(dendat[,1])
y<-dendat[,2]
plot(x,y)

h<-0.2
N<-80
pcf<-pcf.kernesti(x,y,h,N)
dp<-draw.pcf(pcf)
matplot(dp$x,dp$y,type="l",xlim=c(-3.5,0.2),ylim=c(0,1.2))
matplot(x,y,add=TRUE)
```

References

Aït-Sahalia, Y. and Brandt, M. W. (2001), 'Variable selection for portfolio choice', *J. Finance* **56**(4), 1297–1351.

Franke, J., Härdle, W. and Hafner, C. M. (2004), *Statistics of Financial Markets*, Springer.

Nadaraya, E. A. (1964), 'On estimating regression', *Theory Probab. Applications* **10**, 186–190.

Watson, G. S. (1964), 'Smooth regression analysis', *Sankhya Ser. A* **26**, 359–372.