Further intuition for lecture 1 based on questions

Michael Curran

January 14, 2013

1 Empirical Distributions

The empirical distribution

$$P_N[(y,x) \in A] = \frac{1}{N}\sum_{i=1}^{N} 1[(y_i,x_i) \in A] \xrightarrow{as} P[(y,x) \in A]$$

converges almost surely by the SSLN. Rather than in convergence in probability (weak LLNs), strong LLNs imply almost sure convergence. Recall that convergence in probability relates to

$$\lim_{n \longrightarrow \infty} P(|X_n - X| > \epsilon) = 0$$

while almost sure / almost everywhere convergence relates to

$$P(\{\omega : \lim_{n \longrightarrow \infty} X_n(\omega) \neq X(\omega)\}) = 0$$

Note that $\lim_{n \longrightarrow \infty} X_n(\omega) \neq X(\omega)$ is related to ordinary convergence. Almost sure convergence allows for sets not converging – e.g. sets of measure zero (e.g. points). With convegence in probability, even as n gets larger, we can still get outside the ϵ interval with positive probability. However, with almost sure convergence, this cannot happen. So, almost sure convergence is stronger than convergence is probability (hence 'strong' in LLNs). It is important to go through this section on on empirical distributions as these kernel regressions will be demonstrated in practice through MATLAB during laboratory sessions in week 1.

2 Loss Functions

Our aim in conditional prediction is to predict $P(y|x)$. Suppress the conditioning on x and focus on $P(y)$. Let p be a predictor of y. It would be ideal for $p = y$ but we suffer a loss when we fail to predict y with p. So, we can look for instance at the distance $y - p$, how far our predictor is from y.

Sometimes, we are not too concerned with 'small' deviations, say when $y - p$ is less than one in absolute value, but we might be increasingly concerned with

larger deviations, say when $y-p$ is greater than one in absolute value. One such 'loss function' that captures this idea is the squared loss function, $L(y-p) = (y-p)^2$. Typically we let u denote $y-p$, i.e. $u \equiv y-p$ and write $L(u) = u^2$.

We may have different concerns however and feel that symmetric deviations should be equally weighted. We can therefore use a symmetric absolute loss function: $L(u) = |u|$.

To the extent that we must specify our predictor and loss function ex-ante, we necessarily have to use the *expected* loss function, i.e. $E(L(u))$. The aim is to minimise $E(L(u))$ by choosing a best predictor p.

One example of a loss function (in macro) is given by the inflation loss function of a central banker. Take for instance chapter 10 of Romer's Advanced Macroeconomics. There equation (10.11) specifies that a policymaker minimises:

$$L = \frac{1}{2}(y - y*)^2 + \frac{1}{2}a(\pi - \pi*)^2 \quad y* > \bar{y} \quad a > 0$$

where y is log of output, $\bar{y}$ is log of its flexible-price level, π is inflation and $y*$ and $\pi*$ denote the optimal log of output and optimal inflation rate, respectively; a reflects the relative importanc of output and inflation in social welfare (L). The further y deviates from $y*$ – and likewise π from $\pi*$ – the costlier it is in terms of social welfare losses. So optimal choices of $y*$ and $\pi*$ minimise L.

Specification of such functions is rather a formal approach to most problems in economics and researchers often neglect to specify their loss functions. Many papers do not make their assumptions clear, either. Authors typically report point estimates, but as from our study of identification, it is generally the case that to obtain such estimates, the researchers will have had to have made assumptions along the way. Sometimes, the assumptions are credible, but they should nonetheless be reported. Furthermore, there is no problem with reporting intervals or ranges rather than single points. Scientific reporting is certainly an area that can be much improved upon.