Week 2 — Probability & Statistics for Forecasting

Distributions, MLE, and Hypothesis Testing

Dr Juan Zurita

Economic & Business Forecasting
The University of Edinburgh · School of Economics

Key Distributions

Three Workhorses

Normal

$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Most common; foundation of classical econometrics.

Student-t

Heavier tails than Normal.

Robust to outliers; used in financial modelling.

Laplace

$f(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$

Even heavier tails; connection to L_1 estimation.

Tail Behaviour Matters

Financial returns are typically fat-tailed:

- Normal model *underestimates* the probability of extreme events
- Student-t with low ν provides a better fit
- Model choice affects **risk assessment** and **confidence intervals**

Maximum Likelihood Estimation

The MLE Principle

Given data $y_1, \ldots, y_T$ and model $f(y|\theta)$:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \sum_{t=1}^{T} \log f(y_t|\theta)$$

Properties (under regularity conditions):

- **Consistent**: $\hat{\theta} \xrightarrow{p} \theta_0$ as $T \to \infty$
- **Asymptotically efficient**: lowest variance among consistent estimators
- **Asymptotically normal**: $\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0))$

OLS as MLE

For $y_t = x_t'\beta + \varepsilon_t$ with $\varepsilon_t \sim N(0, \sigma^2)$:

$$\log L(\beta, \sigma^2) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - x_t'\beta)^2$$

Maximising w.r.t. $\beta \Leftrightarrow$ minimising $\sum(y_t - x_t'\beta)^2$ — OLS!

Hypothesis Testing

Likelihood Ratio, Wald, and LM Tests

Three asymptotically equivalent approaches to test $H_0 : g(\theta) = 0$:

LR Compare log-likelihoods: $\Lambda = 2[\ell(\hat{\theta}) - \ell(\tilde{\theta})]$

Wald How far is $g(\hat{\theta})$ from zero?

LM How steep is the log-likelihood at the restricted estimate?

All $\xrightarrow{d} \chi^2(r)$ under H_0 where $r =$ number of restrictions.

Information Criteria

Model selection without nested hypotheses:

$$\text{AIC} = -2\ell + 2k \tag{1}$$

$$\text{BIC} = -2\ell + k \log T \tag{2}$$

where k = number of parameters.

- BIC penalises complexity more $\Rightarrow$ simpler models
- BIC is consistent; AIC can overfit
- In practice: compare both

Summary

- Choose distributions that match the **tail behaviour** of your data
- **MLE** is the unifying estimation principle
- **AIC/BIC** guide model selection
- Next week: ARMA models — putting these tools to work