15. Parametric Model Construction

15.1. Parameter Estimation

1. Method of Moments

The MOM estimate of θ is any solution of the *p* equations that match the raw moments:

$\mu_k'(\theta) = \hat{\mu}_k', k = 1,2,\ldots p$

2. Percentile-Matching

For a percentile *q*,

$VaR_q = \lfloor q \times (n+1) \rfloor$

The *smoothed empirical estimate* disregards the largest lower-integer function.

Likelihood Ratio Test

The likelihood ratio (LR) is a test-stat that follows the chi-squared distribution with $(df_1 - df_0)$ degrees of freedom, i.e. number of parameters in the alternative and null models respectively.

$$LR = 2\left(\frac{\lg L(M_0)}{\lg L(M_1)}\right) = 2(\lg L_0 - \lg L_1)$$

$$LR \sim \chi^2_{df_1 - df_0}$$

If LR > χ^2 at (1- α) significance level, then reject H_0.

3. Likelihood Functions

Complete Data

Discrete

$$L(\theta) = \prod_{i=1}^{n} f_{X_i}(x_i|\theta)$$

$$\lg L(\theta) = \sum_{i=1}^{n} \lg f_{X_i}(x_i|\theta)$$

Continuous

n_i = #obs in the interval $(c_{i-1}, c_i]$, $i = 1\ldots k$:

$$L(\theta) = \prod_{i=1}^{k} \left[F(c_i|\theta) - F(c_{i-1}|\theta)\right]^{n_i}$$

$$\lg L(\theta) = \sum_{i=1}^{k} n_i \lg\left[F(c_i|\theta) - F(c_{i-1}|\theta)\right]$$

Right-Censoring, *u*

If there are n_1 observed losses, and n_2 observed censored points *u*:

$$\lg L(\theta) = \sum_{i=1}^{n_1} \lg f_{X_i}(x_i|\theta) + n_2 \lg\left[1 - F(u)\right]$$

Left-Truncation, *d*

NB.: Left truncation is to be used whenever the question says "there is no information about loses below d", or something to that effect.

$$L(\theta) = \prod_{i=1}^{n} \frac{f_X(x_i|\theta)}{1 - F_X(d)}$$

$$\lg L(\theta) = \sum_{i=1}^{n} \lg f_{X_i}(x_i|\theta) - n \lg\left[1 - F(d)\right]$$

Interval Estimation for MLEs

Properties of MLEs

Asymptotic Properties of MLEs:

$$\begin{aligned} &\textit{Consistency}: && \hat{\theta}_{mle} \xrightarrow{p} \theta \\ &\textit{Normality}: && \hat{\theta}_{mle} \overset{d}{\sim} N(\theta, \mathrm{I}^{-1}) \\ &\textit{Efficiency}: && \textit{Cramer-Rao} \end{aligned}$$

Variance of an MLE Parameter Estimation

$$Var(\hat{\theta}) = I^{-1}(\theta) = -E\left[\frac{\partial^2 \lg L(\theta)}{\partial \theta^2}\right]$$

Using the delta method, with MLE covariance matrix Σ:

$$\hat{Var}(\hat{\theta}_1, \hat{\theta}_2) = \begin{bmatrix} g'_{\theta_1} & g'_{\theta_2} \end{bmatrix} \Sigma \begin{bmatrix} g'_{\theta_1} \\ g'_{\theta_2} \end{bmatrix}$$

Covariance Matrix of MLEs

$$\begin{pmatrix} Var(\hat{\alpha}) & Cov(\hat{\alpha}, \hat{\beta}) & Cov(\hat{\alpha}, \hat{\theta}) \\ Cov(\hat{\alpha}, \hat{\beta}) & Var(\hat{\beta}) & Cov(\hat{\beta}, \hat{\theta}) \\ Cov(\hat{\alpha}, \hat{\theta}) & Cov(\hat{\beta}, \hat{\theta}) & Var(\hat{\theta}) \end{pmatrix}$$

Cramer-Rao Inequality:

$$Var(\hat{\theta}) \geq \frac{1}{nI}$$

Equality holds in the case of MLEs since they are *efficient*.

Approximating a 2nd Order Derivative

$$\frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \approx \frac{1}{h_i h_j}\begin{bmatrix} f\left(\theta + \frac{1}{2}h_i e_i + \frac{1}{2}h_j e_j\right) - f\left(\theta + \frac{1}{2}h_i e_i - \frac{1}{2}h_j e_j\right) \\ -f\left(\theta - \frac{1}{2}h_i e_i + \frac{1}{2}h_j e_j\right) + f\left(\theta - \frac{1}{2}h_i e_i - \frac{1}{2}h_j e_j\right) \end{bmatrix}$$

- $h_i = \frac{\theta_i}{10^v}$
- Where *v* is a third of the number of significant digits being used. e_i is a vector of zeros except for a 1 in the *i*-th position.

Delta Method

The Delta method derives an approximate pdf for a function of an asymptotically normal estimator, from knowledge of the limiting variance of that estimator. Commonly:

1. θ_{MLE} is derived for an assumed pdf on the data.
2. $\theta_{MLE} \sim N$, then $F(\theta) \sim N(F(\theta), F'(\theta)^2 Var(\theta))$

Fisher Information

1-Parameter Case

The quantity *I* is the *Fisher's information*, where:

$$\hat{\theta}_{mle} \overset{d}{\sim} N(\theta, \mathrm{I}^{-1})$$

$$\mathrm{I} = -nE\left[\left(\frac{\partial}{\partial \theta} \lg f(x;\theta)\right)^2\right]$$

$$= nE\left[\left(\frac{\partial^2}{\partial \theta^2} \lg f(x;\theta)\right)\right]$$

Multivariate Case

$$\mathrm{I}_{r,s} = -E\left[\frac{\partial^2}{\partial \theta_s \partial \theta_r} \lg f(X;\theta)\right]$$

$$= nE\left[\left(\frac{\partial}{\partial \theta_s} \lg f(X;\theta)\right)\left(\frac{\partial}{\partial \theta_r} \lg f(X;\theta)\right)\right]$$

$$\mathrm{I}_{r,s}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Where *I* is a covariance matrix, or matrix of 2nd order derivatives of the parameters.

15.4 – Non-normal C.I.

Definition

A confidence region would be the set of all parameters that exceeds some choice of *c*:

$$\{\theta : l(\theta) \geq c\}$$

$$c = l(\hat{\theta}) - 0.5\chi^2_{\alpha/2}$$

Where the second term is the (1 - α) percentile from the chi-square distribution with degrees of freed equal to the number of estimated parameters.

15.5 – Bayesian Estimation

Terminology	
Prior A d.f. of the parameters.	$\pi(\theta)$
Model/Likelihood A d.f. of the for the data collected given a particular parameter value.	$f_{X\mid\Theta}(x\mid\theta)=\prod_{i=1}^{n} f_{X\mid\Theta}(x_i\mid\theta)$
Joint Joint = Model * Prior	$f_{X,\Theta}=f_{X\mid\Theta}(x\mid\theta)\pi(\theta)$
Marginal Marginal = Integral of Joint	$f_X=\int f_{X\mid\Theta}(x\mid\theta)\pi(\theta)d\theta$
Posterior Conditional d.f. of parameter values, given observed data	$\pi_{\Theta\mid X}=\frac{f_{X\mid\Theta}\pi_\Theta}{\int f_{X\mid\Theta}\pi_\Theta\, d\theta}=\frac{f_{X,\Theta}}{f_X}$
Predictive	$f_{Y\mid X=x}=\int f_{Y\mid\Theta}(X\mid\theta)\pi_{\Theta\mid X}\, d\theta$

15.5b – Inference & Prediction

3 Loss Functions

Name	Loss Function	Property of Posterior DF
Squared-Error	$(\hat{\theta}-\theta)^2$	Mean
Absolute-Value	$\lvert\hat{\theta}-\theta\rvert$	Median
Zero-One	$\begin{cases} 0 & \hat{\theta}=\theta \\ 1 & otherwis \end{cases}$	Mode

Bayesian Confidence Intervals

The BCI of an *estimated parameter* is the probability that the estimated parameter lies in [a, b] is 1 - α.

$$P(a\le\theta\le b\mid x)\ge 1-\alpha$$

Highest Posterior Density (HPD) Set

This is the *shortest* possible interval [a, b] enclosing the posterior probability mass 1 - α. Formally: HPD is the set of parameter values $\{C\}$ s.t.:

$P(\theta_i\in\{C\})\le 1-\alpha$

$C:=\{\theta_i:\pi_{\Theta\mid X}(\theta_i\mid x)\ge c\}$

for some *c*, where *c* is the largest value for which the 1st inequality holds.

The CI is defined on the *posterior* as the smallest unique interval [a, b] such that:

$\int_a^b \pi_{\Theta\mid X}(\theta\mid x)d\theta=1-\alpha$

$\pi_{\Theta\mid X}(a\mid x)=\pi_{\Theta\mid X}(b\mid x)$

Bayesian CLT

To make life easier, given certain conditions and asumptions, the posterior d.f. is asymptotically normal.