§7

Frequentist Techniques for Parameter Estimation

Observed model response

$$y = f(x, q)$$

Observations

$$y_i = f(x_i, q) + \overset{\text{measurement errors}}{\varepsilon_i}, \quad i = 1 \cdots n$$

Determine q in a stable manner from y
inverse uncertainty quantification

Evolution process

$$\frac{\partial u}{\partial t} = \mathcal{N}(u, q) + F(q), \quad x \in D, \; t \geq t_0$$

$$B(u, q) = G(q) \qquad x \in \partial D, \; t \geq t_0$$

$$u(t_0, x, q) = I(q) \qquad x \in D$$

$$\Rightarrow f(q) = (f(t_1, q) \cdots f(t_n, q))^T$$

$$Y = f(q_0) + \overset{\text{model errors}}{\delta} + \varepsilon$$, random vector with realization y

Frequentist assumption: parameters are fixed q_0 but possibly unknown

q_0 given parameter value, generating y

Calibration of model with q such that $f(q)$ is close to data

estimator $\hat{q}$

$$\underset{\uparrow \text{random vector}}{\hat{q}_{OLS}} = \underset{q \in Q}{\arg\min} \sum_{i=1}^{n} (Y_i - f_i(q))^2$$ Ordinary Least Squares

Require $E(\hat{q}) = q_0$, covariance quantifies variability of errors does not provide distribution for model parameters, q_0 is not a random variable

Linear regression

$$Y = Xq_0 + \varepsilon \quad , X \in \mathbb{R}^{n\times p} \quad , \overset{\text{measurements}}{\overset{\downarrow}{n}} > \overset{\text{parameters}}{\overset{\swarrow}{p}}$$

rank $X = p \Rightarrow q_0$ is identifiable, $\mathcal{R}(X^T) = \mathbb{R}^p$

Errors ε are unbiased and iid with variance σ_0^2

$\Rightarrow$ 1. $E(\varepsilon_i) = 0$

2. $Var(\varepsilon_i) = \underset{\text{fixed, unknown}}{\sigma_0^2}$, $Cov(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$

Goal: unbiased estimators $\hat{q}$, $\hat{\sigma}^2$ for q_0, σ_0^2

$$J(q) = (Y - Xq)^T (Y - Xq)$$

$$\min J(q) \Rightarrow \nabla_q J = 2\Big(\underbrace{\nabla_q (Y - Xq)^T}_{-\nabla_q q^T X^T = -X^T}\Big)(Y - Xq) = 0$$

$$\Rightarrow X^T Y = X^T X q \Rightarrow \hat{q} = \underbrace{(X^T X)^{-1} X^T}_{A} Y = AY$$

with realization

$$q = (X^T X)^{-1} X^T y$$

mean value

1. $E(\hat{q}) = E((X^TX)^{-1}X^TY) = (X^TX)^{-1}X^T E(Y) =$
$\{ E(Y) = E(Xq_0) + E(\varepsilon) = Xq_0 \} = q_0$

covariance matrix

2. $V(\hat{q}) = E((\hat{q} - q_0)(\hat{q} - q_0)^T) =$
$\{ \hat{q} = AY = AXq_0 + A\varepsilon = q_0 + A\varepsilon \}$
$= A \underbrace{E(\varepsilon\varepsilon^T)}_{\sigma_0^2 I} A^T = \sigma_0^2 (X^TX)^{-1} X^T X (X^TX)^{-1} = \sigma_0^2 (X^TX)^{-1}$

Unbiased covariance estimator

$$\hat{\sigma}^2 = \frac{1}{n-p} \hat{R}^T \hat{R}, \quad \underset{\uparrow \text{residual}}{\hat{R}} = Y - X\hat{q} = Y - X(X^TX)^{-1}X^TY = (I - X(X^TX)^{-1}X^T)Y$$

cf. standard deviation (sample variance) $p=1$

Sampling Distribution for $\hat{q}$

Assume errors normally distributed or large number of samples to apply central limit theorem

Errors are iid and $\varepsilon_i \sim N(0, \sigma_0^2)$

$\Rightarrow \hat{q} \sim N(q_0, \sigma_0^2 (X^TX)^{-1})$

Normal Distribution specified by $E(\hat{q}), V(\hat{q})$

Errors may be iid with variance σ_0^2 but not normally distributed

For large sample sizes use central limit theorem

$\Rightarrow$ as above

Confidence intervals for mean of $\hat{q}$

$p=1, \; X = (1, 1 \cdots 1)^T, \quad Y_i \sim N(\mu_0, \sigma_0^2)$

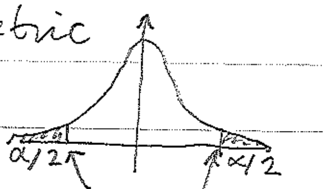
$$\hat{q} = \frac{1}{n} \sum_{i=1}^{n} Y_i, \; E[\hat{q}] = \mu_0, \; V(\hat{q}) = \frac{\sigma_0^2}{n}, \quad \hat{q} \sim N(\mu_0, \frac{\sigma_0^2}{n})$$

$$q = \frac{\sqrt{n}(\hat{q} - \mu_0)}{\sigma_0} \sim N(0,1), \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \hat{q})^2$$

$$S^2 \sim \frac{\sigma_0^2}{n-1} \chi^2(n-1) \Rightarrow Z = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$$

$$\Rightarrow T = \frac{q}{\sqrt{Z/(n-1)}} = \frac{\sqrt{n}(\hat{q} - \mu_0)}{S}$$ is $\overset{\text{Student's}}{\downarrow}$ t-Distributed, $n-1$

$\uparrow$ symmetric



$$P(\underset{\uparrow}{\overset{-t_{n-1,1-\alpha/2}}{a}} < \frac{\sqrt{n}(\hat{q} - \mu_0)}{S} < \overset{t_{n-1,1-\alpha/2}}{b}) = 1 - \alpha$$

$$P(\hat{q} - \frac{t_{n-1,1-\alpha/2} S}{\sqrt{n}} < \mu_0 < \hat{q} + \frac{t_{n-1,1-\alpha/2} S}{\sqrt{n}}) = 1 - \alpha$$

↗ probability that μ_0 is in the interval

Nonlinear models

$$Y = f(q_0) + \varepsilon$$

ε_i iid, $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma_0^2$

estimator

$$\hat{q} = \arg\min_{q \in Q} \sum_{i=1}^{n} (Y_i - f_i(q))^2$$

estimate

$$q = \arg\min_{q \in Q} \sum_{i=1}^{n} (y_i - f_i(q))^2$$

nonlinear least squares problem $\Rightarrow$ $\chi^T(q_0)(y - f(q_0)) = 0$

Linearization around q_0, $\chi_{ij}(q) = \frac{\partial f_i}{\partial q_j}(q)$

$V(\hat{q}) \approx \sigma_0^2 (\chi^T(q_0)\chi^T(q_0))^{-1}$, $\hat{q} = q_0 + \delta q$

1. $\chi_{ij}(q) \approx \frac{f_i(q + h_j) - f_i(q)}{\underset{\nwarrow h}{|h_j|}}$, $h_j = h \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \end{pmatrix} \leftarrow j$
2. Automatic differentiation (AD)

Sampling distribution

$$\hat{q} \sim N(q_0, \sigma_0^2 (\chi^T(q_0)\chi(q_0))^{-1}), \quad q_0 = E(\hat{q})$$

Confidence intervals in the same manner

Exercise: 7.6

$$Y = f(q_0 + \delta q) + \varepsilon = f(q_0) + \chi\,\delta q + \varepsilon + O(\|\delta q\|^2)$$

$$Y - f(q_0) = \chi\,\delta q + \varepsilon$$

Linearize

$$J(q_0+\delta q) = (Y - f(q_0) - \chi_0 \delta q)^T (Y - f(q_0) - \chi_0 \delta q)$$

$$\nabla_{\delta q} J = -2\chi_0^T (Y - f(q_0) - \chi_0 \delta q) = 0$$

$$\Rightarrow \quad \chi_0^T \chi_0 \delta q = \chi_0^T (Y - f(q_0))$$

$$\delta q = \underbrace{(\chi_0^T \chi_0)^{-1} \chi_0^T}_{A} (\underbrace{Y - f(q_0)}_{\varepsilon})$$

$$E(\delta q) = E(A\varepsilon) = A\,E(\varepsilon) = 0$$

$$V(\delta q) = \sigma_0^2 (\chi'\chi)^{-1}$$

$$\hat{q} = q_0 + \delta q \Rightarrow E(\hat{q}) = q_0 , \quad \hat{q} = q_0 + A\varepsilon$$

$$V(\hat{q}) = E((\hat{q}-q_0)(\hat{q}-q_0)^T) = A\,E(\varepsilon\varepsilon^T)A^T = \sigma_0^2 (\chi^T\chi)^{-1}$$

UQ5

Bayesian techniques for parameter estimation

$Y_i = f_i(Q) + \varepsilon_i$, $i = 1 \cdots n$, ε_i unbiased iid
Y_i, ε_i, Q random variables

Parameters random variables, realizations $q = Q(\omega)$
Q has density, updated by incoming data
⇒ posterior density for Q based on sample(d) observations

prior density, likelihood ⇒ posterior density
$\pi_0(q)$ $\quad \pi(y|q) \quad$ $\pi(q|y)$

$\pi_0(q)$ knowledge obtained before observations y
noninformative prior $\pi_0(q) = \chi_{(0,\infty)}(q)$

$$\chi_{(a,b)}(q) = \begin{cases} 1, & q \in (a,b) \\ 0, & \text{otherwise} \end{cases}$$ indicator function

$\pi(y|q) = L(q|y)$ based on information from samples

$\pi(q,y)$ joint density of Q and Y

$$\pi(y|q) = \frac{\pi(q,y)}{\pi_0(q)}$$

$$\Rightarrow \quad \pi(q|y_{obs}) = \frac{\pi(q,y_{obs})}{\pi(y_{obs})}$$ posterior density

$$\pi(y_{obs}) = \int_{\mathbb{R}^p} \pi(q, y_{obs})\,dq = \int_{\mathbb{R}^p} \pi(y_{obs}|q)\pi_0(q)\,dq$$

Q has prior density $\pi_0(q)$

$$\pi(q|y_{obs}) = \frac{\pi(y_{obs}|q)\pi_0(q)}{\pi(y_{obs})} = \frac{\pi(y_{obs}|q)\pi_0(q)}{\int_{\mathbb{R}^p}\pi(y_{obs}|q)\pi_0(q)\,dq}$$

$$\Rightarrow \int_{\mathbb{R}^p}\pi(q|y_{obs})\,dq = 1$$

like lect 4

Likelihood function (y = yobs from now on)

$\varepsilon_i \sim N(0,\sigma^2)$, iid, σ^2 fixed

$$\pi(y|q) = L(q,\sigma^2|y) = \frac{1}{(2\pi\sigma^2)^{n/2}}\underbrace{e^{-SS_q/2\sigma^2}}_{\prod_{i=1}^{n} e^{-\underbrace{(y_i-f_i(q))}_{\varepsilon_i}/2\sigma^2}}$$

$$SS_q = \sum_{i=1}^{n}(y_i - f_i(q))^2$$

Maximum a posteriori (MAP) estimate

$$q_{MAP} = \arg\max_q \pi(q|y)$$

normalization $\pi(y)$ does not change the max

$$\Rightarrow q_{MAP} = \arg\max \pi(y|q)\pi_0(q)$$

$\pi_0(q)$ uniform on $\mathbb{R}$ $\Rightarrow$

$$q_{MAP} = q_{MLE} = \arg\max_q \pi(y|q)$$

Take $\pi_0(q) = \chi_{[0,\infty)}(q)$

$$\Rightarrow \pi(q|y) = \frac{e^{-SS_q/2\sigma_0^2}}{\int_0^\infty e^{-SS_\zeta/2\sigma_0^2}\,d\zeta} = \frac{1}{\int_0^\infty e^{-(SS_\zeta - SS_q)/2\sigma_0^2}\,d\zeta}$$