STK4011 and STK9011 Autumn 2016

Point estimation

Covers (most of) the following material from chapter 7:

- Section 7.1: pages 311-312
- Section 7.2.1: pages 312-313
- Section 7.2.2: pages 315-320
- Section 7.2.3: pages 324-325
- Section 7.3.1: pages 330-332
- Section 7.3.2: pages 334-339

Ørnulf Borgan
Department of Mathematics
University of Oslo

Estimation

Let $X_1, X_2, ..., X_n$ be a random sample from the population $f(x|\boldsymbol{\theta})$, so $X_1, X_2, ..., X_n$ are iid and their pmf or pdf is $f(x|\boldsymbol{\theta})$

A point estimator $W(X_1, X_2, ..., X_n)$ is a statistic that we use to guess the value of the population parameter $\boldsymbol{\theta}$ (or a function $\tau(\boldsymbol{\theta})$ of the parameter)

When we evaluate the estimator at the observed values $x_1, x_2, ..., x_n$ of the random variables we obtain an estimate of $\boldsymbol{\theta}$

We will first consider methods for finding estimators, and then discuss how the estimators may be evaluated

Method of moments

Assume that $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$

The moment estimators of $\theta_1, ..., \theta_k$ are obtained by the equations one obtains by equating the first k sample moments and population moments

Sample moments: $m_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$

Population moments: $\mu_j'(\theta_1, ..., \theta_k) = \mathrm{E}X^j$

The moment estimators solve the equations

$$\mu_j'(\theta_1, ..., \theta_k) = m_j \qquad j = 1, ..., k$$

Example – gamma distribution

Let $X_1, X_2, ..., X_n$ be iid with pdf

$$f(x|\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad \text{for } x > 0$$

We have

$$\mu_1'(\alpha,\beta) = \mathrm{E}X = \alpha\beta$$

$$\mu_2'(\alpha,\beta) = \mathrm{E}X^2 = \mathrm{Var}X + (\mathrm{E}X)^2 = \alpha\beta^2 + (\alpha\beta)^2$$

The moment estimators are given by the equations

$$\bar{X} = \alpha\beta \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \alpha\beta^2 + (\alpha\beta)^2$$

This gives

$$\hat{\alpha} = \frac{\bar{X}^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad \text{and} \quad \hat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}{\bar{X}}$$

Maximum likelihood

Let $X_1, X_2, ..., X_n$ be an iid sample from a population with pmf or pdf $f(x \mid \theta_1, ..., \theta_k)$

Then the likelihood is given by

$$L(\boldsymbol{\theta} \mid \mathbf{x}) = L(\theta_1, ..., \theta_k \mid x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i \mid \theta_1, ..., \theta_k)$$

For each sample point $\mathbf{x}$, let $\hat{\boldsymbol{\theta}}(\mathbf{x})$ be the parameter value at which $L(\boldsymbol{\theta} \mid \mathbf{x})$ attains its maximum as a function of $\boldsymbol{\theta}$ with $\mathbf{x}$ held fixed

Then $\hat{\boldsymbol{\theta}}(\mathbf{X})$ is the maximum likelihood estimator (MLE) based on the sample $\mathbf{X} = (X_1, X_2, ..., X_n)$

Usually one finds the MLE by maximizing the log-likelihood $\log L(\boldsymbol{\theta} \mid \mathbf{x})$

Example – gamma distribution

Let $X_1, X_2, ..., X_n$ be iid with pdf

$$f(x \mid \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad \text{for } x > 0$$

The likelihood is given by

$$L(\alpha, \beta \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i \mid \alpha, \beta) = \left(\frac{1}{\beta^\alpha \Gamma(\alpha)} \right)^n \left(\prod_{i=1}^{n} x_i^{\alpha-1} \right) e^{-\sum_{i=1}^{n} x_i / \beta}$$

and the log-likelihood becomes

$$\log L(\alpha, \beta \mid \mathbf{x})$$

$$= -n\alpha \log \beta - n \log \Gamma(\alpha) + (\alpha - 1) \sum_{i=1}^{n} \log x_i - \frac{1}{\beta} \sum_{i=1}^{n} x_i$$

Differentiating the log-likelihood we obtain

$$\frac{\partial}{\partial \alpha} \log L(\alpha, \beta \mid \mathbf{x}) = -n \log \beta - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum_{i=1}^{n} \log x_i$$

$$\frac{\partial}{\partial \beta} \log L(\alpha, \beta \mid \mathbf{x}) = -\frac{n\alpha}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{n} x_i$$

If we set the partial derivatives equal to zero and solve the equations, we find that the MLEs are given by

$$\hat{\beta} = \frac{\bar{X}}{\hat{\alpha}}$$

and

$$n \log \hat{\alpha} - n \log \bar{X} - n \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} + \sum_{i=1}^{n} \log X_i = 0$$

An important property of ML-estimators is that they are invariant in the following sense:

Theorem 7.2.10 (invariance property of MLEs)

If $\hat{\theta}$ is the MLE of θ , then $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$

Assume that the function $\eta = \tau(\theta)$ is one-to-one

Let $\tau^{-1}(\eta) = \theta$ be the inverse function

Then $L^*(\eta \mid \mathbf{x}) = L(\tau^{-1}(\eta) \mid \mathbf{x})$ and

$$\sup_\eta L^*(\eta \mid \mathbf{x}) = \sup_\eta L(\tau^{-1}(\eta) \mid \mathbf{x}) = \sup_\theta L(\theta \mid \mathbf{x})$$

Thus the maximum of $L^*(\eta \mid \mathbf{x})$ is attained at $\hat{\eta} = \tau(\hat{\theta})$

The result is also valid when $\eta = \tau(\theta)$ is not one-to-one (cf. page 320)

Bayes estimators

In this course we mainly focus on the classical frequentist approach to statistics, but we will now have a brief look at the Bayesian approach

In the classical approach a population parameter θ is a fixed quantity, but its value is unknown to us

In the Bayesian approach θ is assumed to be a quantity whose variation may be described by the prior distribution $\pi(\theta)$

When a sample is taken from the population, we may update the prior distribution and obtain the posterior distribution $\pi(\theta \mid \mathbf{x})$

The joint pmf or pdf of θ and $\mathbf{X}$ is

$$f(\mathbf{x},\theta) = f(\mathbf{x} \mid \theta)\pi(\theta)$$

The marginal pmf or pdf of $\mathbf{X}$ is

$$m(\mathbf{x}) = \int f(\mathbf{x} \mid \theta)\pi(\theta)\, d\theta$$

(with sum instead of integral if $\pi(\theta)$ is a pmf)

Hence the posterior pmf or pdf of θ becomes

$$\pi(\theta \mid \mathbf{x}) = \frac{f(\mathbf{x} \mid \theta)\pi(\theta)}{m(\mathbf{x})}$$

The posterior is used to make statements about θ

We may e.g. use the posterior mean as a point estimate of θ

Example 7.2.14 – binomial Bayes estimator

Let $Y \sim \text{binomial}(n, p)$:

$$f(y \mid p) = \binom{n}{y} p^y (1-p)^{n-y}$$

Assume that the prior distribution of p is $\text{beta}(\alpha,\beta)$:

$$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, p^{\alpha-1}(1-p)^{\beta-1}$$

Then the joint distribution of Y and p is

$$f(y,p) = \binom{n}{y} p^y (1-p)^{n-y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, p^{\alpha-1}(1-p)^{\beta-1}$$

$$= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, p^{y+\alpha-1}(1-p)^{n-y+\beta-1}$$

Then marginal distribution of Y is

$$m(y) = \int_0^1 \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, p^{y+\alpha-1}(1-p)^{n-y+\beta-1}\, dp$$

$$= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 p^{y+\alpha-1}(1-p)^{n-y+\beta-1}\, dp$$

$$= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)}$$

Then posterior distribution of p becomes

$$f(p \mid y) = \frac{f(y,p)}{m(y)} = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)}\, p^{y+\alpha-1}(1-p)^{n-y+\beta-1}$$

Thus

$$p \mid Y = y \sim \text{beta}(y+\alpha, n-y+\beta)$$

We may use the posterior mean to estimate p , i.e.

$$\hat{p}_B = \frac{y+\alpha}{n+\alpha+\beta}$$

Note that we may write the estimator as

$$\hat{p}_B = \left(\frac{n}{n+\alpha+\beta}\right)\left(\frac{y}{n}\right) + \left(\frac{\alpha+\beta}{n+\alpha+\beta}\right)\left(\frac{\alpha}{\alpha+\beta}\right)$$

In this example both the prior and the posterior distributions are beta-distributions

The reason for this is that the beta distribution is a conjugate family for the binomial

Evaluation of estimators

One criterion for evaluating estimators is the mean squared error (MSE)

The MSE of an estimator $W = W(X_1,\ldots,X_n)$ of θ is the function of θ given by $\mathrm{E}_\theta(W-\theta)^2$

The bias of W is given as $\mathrm{Bias}_\theta\, W = \mathrm{E}_\theta\, W - \theta$

The estimator is unbiased if the bias is zero for all θ

Note that the MSE may be written

$$\mathrm{E}_\theta(W-\theta)^2 = \mathrm{Var}_\theta W + \left(\mathrm{Bias}_\theta\, W\right)^2$$

For an unbiased estimator, MSE equals its variance

Example 7.3.5 – MSE of binomial Bayes estimator

Let $Y \sim \mathrm{binomial}(n, p)$

The MLE of p is $\hat{p} = Y/n$

We know that this is unbiased, so its MSE becomes

$$\mathrm{E}_p(\hat{p}-p)^2 = \mathrm{Var}_p\,\hat{p} = \frac{p(1-p)}{n}$$

Then consider the Bayes estimator $\hat{p}_B = \frac{Y+\alpha}{n+\alpha+\beta}$

The variance of the Bayes estimator is

$$\mathrm{Var}_p\,\hat{p}_B = \mathrm{Var}_p\left(\frac{Y+\alpha}{n+\alpha+\beta}\right) = \frac{np(1-p)}{(n+\alpha+\beta)^2}$$

The bias of the Bayes estimator is

$$\mathrm{Bias}_p\,\hat{p}_B = \mathrm{E}_p\left(\frac{Y+\alpha}{n+\alpha+\beta}\right) - p = \frac{np+\alpha}{n+\alpha+\beta} - p$$

Hence the MSE of the Bayes estimator is

$$\mathrm{E}_p(\hat{p}_B-p)^2 = \mathrm{Var}_p\,\hat{p}_B + \left(\mathrm{Bias}_p\,\hat{p}_B\right)^2$$

$$= \frac{np(1-p)}{(n+\alpha+\beta)^2} + \left(\frac{np+\alpha}{n+\alpha+\beta} - p\right)^2$$

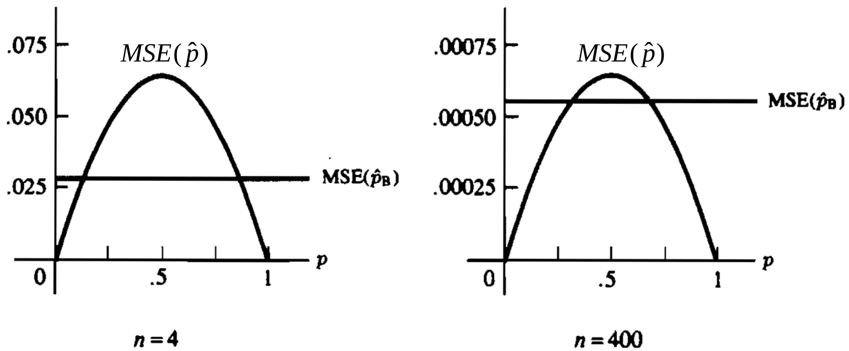
If we chose

$$\alpha = \beta = \sqrt{n/4}$$

the MSE will not depend on p

Then we obtain

$$\hat{p}_B = \frac{Y + \sqrt{n/4}}{n + \sqrt{n}} \quad \text{and} \quad \mathrm{E}(\hat{p}_B - p)^2 = \frac{n}{4(n+\sqrt{n})^2}$$



Best unbiased estimators

We are not able to find an estimator that minimizes the MSE for all values of θ

If we want to find «the best» estimator, we need to restrict the class of estimators we consider

If we restrict ourselves to the unbiased estimators, the best estimator is the one with smallest variance

Definition 7.3.7

An estimator W^* is a best unbiased estimator of $\tau(\theta)$ if it satisfies $\mathrm{E}_\theta W^* = \tau(\theta)$ for all θ and, for any other estimator W with $\mathrm{E}_\theta W = \tau(\theta)$, we have $\mathrm{Var}_\theta W^* \leq \mathrm{Var}_\theta W$ for all θ. We also say that W^* is a uniform minimum variance unbiased estimator (UMVUE) for $\tau(\theta)$

Example 7.3.12 – Poisson unbiased estimation

Let $X_1, X_2, ..., X_n$ be iid Poisson(λ)

Further let $\bar{X}$ and S^2 be the sample mean and the sample variance

Then $\mathrm{E}\bar{X} = \lambda$ and $\mathrm{E}S^2 = \lambda$

Thus $\bar{X}$, S^2 and $a\bar{X} + (1-a)S^2$ are all unbiased estimators of λ

Which one is best, and are there better unbiased estimators?

Information

Consider a sample $\mathbf{X} = (X_1, X_2, ..., X_n)$ with joint pdf (or pmf) $f(\mathbf{x} \mid \theta)$

Note that we do not assume that the X_i are iid

We assume that expressions of the form

$$\int_{\mathcal{X}} W(\mathbf{x}) f(\mathbf{x} \mid \theta) d\mathbf{x}$$

may be differentiated with respect to θ by changing the order of differentiation and integration (summation in the case of pmf)

Then the information number or Fisher information of the sample is

$$I(\theta) = \mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right)^2\right] = \int_{\mathcal{X}}\left(\frac{\partial}{\partial\theta}\log f(\mathbf{x}|\theta)\right)^2 f(\mathbf{x}|\theta)d\mathbf{x}$$

For the special case where $X_1, X_2, ..., X_n$ are iid with pdf (or pmf) $f(x|\theta)$ the information in the sample is given by $I(\theta) = nI_1(\theta)$, where $I_1(\theta)$ is the information in one observation and is given by

$$I_1(\theta) = \mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right] = \int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2 f(x|\theta)dx$$

If expressions of the form $\int W(\mathbf{x})f(\mathbf{x}|\theta)d\mathbf{x}$ may be differentiated twice with respect to θ by changing the order of differentiation and integration (summation for pmf), the information in the sample may also be given as

$$I(\theta) = -\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2}\log f(\mathbf{X}|\theta)\right) = -\int_{\mathcal{X}}\left(\frac{\partial^2}{\partial\theta^2}\log f(\mathbf{x}|\theta)\right)f(\mathbf{x}|\theta)d\mathbf{x}$$

while in the case of iid observations, the information in one observation may be given as

$$I_1(\theta) = -\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right) = -\int_{-\infty}^{\infty}\left(\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right)f(x|\theta)dx$$

The Cramér-Rao inequality

Consider a sample $\mathbf{X} = (X_1, X_2, ..., X_n)$ with joint pdf (or pmf) $f(\mathbf{x}|\theta)$ and assume that expressions of the form

$$\int W(\mathbf{x})f(\mathbf{x}|\theta)d\mathbf{x}$$

may be integrated with respect to θ by changing the order of differentiation and integration (summation in the case of pmf). Then for any estimator $W(\mathbf{X})$ with finite variance we have

$$\mathrm{Var}_\theta W(\mathbf{X}) \geq \frac{\left(\frac{d}{d\theta}\mathrm{E}_\theta W(\mathbf{X})\right)^2}{\mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right)^2\right]}$$

Example 7.3.8 – Poisson UMVUE

Let $X_1, X_2, ..., X_n$ be iid Poisson(λ)

Here

$$\log f(x|\lambda) = \log\left(\frac{\lambda^x}{x!}e^{-\lambda}\right) = x\log\lambda - \lambda - \log x!$$

$$\frac{\partial}{\partial\lambda}\log f(x|\lambda) = \frac{x}{\lambda} - 1 \qquad \frac{\partial^2}{\partial\lambda^2}\log f(x|\lambda) = -\frac{x}{\lambda^2}$$

Hence the information in one observation is

$$I_1(\lambda) = \mathrm{E}\left(\frac{X}{\lambda^2}\right) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

and the information in the sample is $I(\lambda) = n/\lambda$

By the Cramér-Rao inequality, we have for any unbiased estimator W for λ that

$$\text{Var}_\lambda W \geq \frac{1}{I(\lambda)} = \frac{1}{n/\lambda} = \frac{\lambda}{n}$$

Now $\bar{X}$ is an unbiased estimator for λ, and

$$\text{Var}_\lambda \bar{X} = \frac{\lambda}{n}$$

Hence $\bar{X}$ is an UMVUE for λ