4 Parametric point estimation

4.1 General comments

Context:

- Parametric statistical model for X, $\mathcal{F} = \{f(\cdot \mid \theta) : \theta \in \Theta\}$
- $X_1, \ldots, X_n$ iid random sample of size n extracted from X
- Sample space denoted by $\mathcal{X}$

Problem:

- produce a point estimate of θ, that is, select an application $\mathbf{x} \in \mathcal{X} \mapsto T(\mathbf{x}) \in \Theta$ that to each observed sample associates a value for θ
- The application T is a statistic which we call an estimator of θ; the observed value $T(\mathbf{x})$ is called the estimate
- We may be interested in estimating a function of θ, $\tau(\theta)$

Example 4.1 *Assume that X, the number of claims per insurance policy per year, follows a Poisson distribution with parameter $\lambda > 0$.*

We may be interested in estimating λ. A natural estimator is $T(X_1, \ldots, X_n) = \bar{X}$; the associated estimate is $T(x_1, \ldots, x_n) = \bar{x}$.

Alternatively, instead of focusing our interest on the average number of claims, we may be interested in estimating $\tau(\lambda) = e^{-\lambda}$, which corresponds to the probability of zero claims in one year: $P(X = 0 \mid \lambda) = e^{-\lambda}$.

Plan

- The definition of estimator is extremely general
- What are some desirable characteristics for an estimator — optimality criteria
- In what circumstances can we be assured that optimal estimators exist?
- General methods for constructing estimators and associated properties

4.2 Optimality criteria

- In frequentist statistics, the quality of an estimator is assessed by looking at the population of estimates it produces, that is, at its sampling distribution
- We are evaluating the estimator and not the estimate - pre-experimental versus post-experimental precision

4.2.1 Unbiasedness

Definition 4.1 (Unbiased estimator) *An estimator T is said to be an unbiased estimator of $\tau(\theta)$ iff*

$$\forall\ \theta \in \Theta\ E_\theta[T] = \tau(\theta)\ .$$

■

Remarks

- In practice, this means that if we use T to estimate $\tau(\theta)$ a very large number of times, then the average of the estimates will be close to $\tau(\theta)$ no matter the true value of θ.
- The quantity $b(T) = E[T \mid \theta] - \tau(\theta)$ is known as the bias of the estimator of $\tau(\theta)$. An estimator which is not unbiased is said to be biased.

Remarks (cont'd)

- There are cases where an unbiased estimator cannot be found: suppose $X \sim \mathrm{B}(1, \theta)$ and $n = 1$. We seek an unbiased estimator of θ^2. Such an estimator, T, should satisfy

$$T(0) \times (1 - \theta) + T(1) \times \theta = \theta^2, \ \forall \theta \in (0, 1)$$

  which is impossible.

- The sample mean is unbiased estimator of the population mean as long as it exists. The bias-corrected variance is an unbiased estimator of the population variance as long as it exists.

- By restricting the class of interesting estimators to the class of unbiased estimators, we may miss interesting estimators — more on this later.

4.2.2 Most efficient estimation

Definition 4.2 *Let T and $T^\star$ be two unbiased estimators of $\tau(\theta)$. We say that T is more efficient than $T^\star$ in the estimation of $\tau(\theta)$ if*

$$Var_\theta(T) \leq Var_\theta(T^\star),\ \forall \theta \in \Theta\ .$$

Theorem 4.1 (Cramer-Rao inequality) *Consider a statistical model satisfying the regularity conditions with $\Theta \subset \mathbb{R}$, and let $\tau(\theta)$ be a differentiable function. Let T be an unbiased estimator of $\tau(\theta)$ with finite variance. Additionally, assume that, $\forall\ \theta \in \Theta$*

$$E_\theta\left[|T(X_1,\ldots,X_n)S(\theta \mid X_1,\ldots,X_n)|\right] < +\infty$$
$$E_\theta\left[T(X_1,\ldots,X_n)S(\theta \mid X_1,\ldots,X_n)\right] = \tau'(\theta)$$

in which case we say that T is a regular estimator. Then

$$Var_\theta(T) \geq \frac{[\tau'(\theta)]^2}{nI_{X_1}(\theta)}\ .$$

■

Remarks

- The quantity $[\tau'(\theta)]^2/[nI_{X_1}(\theta)]$ is known as the Cramer-Rao lower bound. It is only meaningful under the regularity conditions
- Even when the regularity conditions are satisfied, there may not exist an estimator whose variance equals the CR lower bound
- The ratio between the CR lower bound and the variance of an unbiased estimator of $\tau(\theta)$ is known as efficiency:

$$e(T) = \frac{[\tau'(\theta)]^2/[nI_{X_1}(\theta)]}{\mathrm{Var}_\theta(T)} \ .$$

  If the regularity conditions are satisfied, $0 \leq e(T) \leq 1$.

- If T is an unbiased estimator of $\tau(\theta)$ and $e(T) = 1$, then T is known as the most efficient estimator of $\tau(\theta)$
- There is also the notion of asymptotic efficiency: $\lim_{n\to+\infty} e(T)$ and also of asymptotically most efficient estimators

Example 4.2 *Let $X_1, \ldots, X_n$ be a random sample from a $Po(\lambda)$ population, $\lambda > 0$. It's easy to see that $I_X(\lambda) = 1/\lambda$, and hence any unbiased estimator of λ, T, satisfies*

$$Var(T \mid \theta) \geq \lambda/n \ .$$

Consequently, $\bar{X}$ is the most efficient of λ. Any unbiased estimator of $\tau(\lambda) = e^{-\lambda} = P(X = 0 \mid \lambda)$, W, satisfies

$$Var(W \mid \lambda) \geq \lambda e^{-2\lambda}/n \ .$$

■

Example 4.3 *Let* $X \sim U(0, \theta)$, $\theta > 0$ *and consider*

$$T = \frac{n+1}{n} \, X_{(n)} \, .$$

It's easy to check that T *is unbiased for* θ *and that* $Var(T \mid \theta) = \theta^2/[n(n+2)]$. *Additionally, the CR lower bound turns out to be* θ^2/n, *and hence*

$$Var(T \mid \theta) < CRLB \, .$$

Of course, the regularity conditions are not met. ■

Under what conditions can we be assured that most efficient estimators exist? Corollary of the CR theorem:

Theorem 4.2 *Let T be a regular and unbiased estimator of $\tau(\theta)$. Then, T is the most efficient estimator of $\tau(\theta)$ if and only iff there exists $a(\theta)$ such that*

$$S(\theta \mid x_1, \ldots, x_n) = a(\theta) \, [T(x_1, \ldots, x_n) - \tau(\theta)] \, .$$

■

- The CRLB in the estimation of $\tau(\theta)$ is attained by an estimator T iff T is a sufficient statistic in the one-parameter exponential family with density

$$f(\mathbf{x}) = h(\mathbf{x}) \; c(\theta) \; \exp[Q(\theta) \; T(\mathbf{x})]$$

  where $c(\theta) = \int a(\theta) \; \tau(\theta) \; d\theta$ and $Q(\theta) = \int a(\theta) \; d\theta$.

- If there is a most efficient estimator for $\tau(\theta)$ then T must be sufficient for θ — there aren't most efficient estimators in models that do not admit one-dimensional sufficient statistics

Example 4.4 *If* $X \sim B(1, \theta)$*, then*

$$S(\theta \mid x_1, \ldots, x_n) = \frac{n}{\theta(1-\theta)}(\bar{x} - \theta)$$

and we can immediately conclude that $\bar{X}$ *is the most efficient estimator of* θ*.* ■

Example 4.5 *Consider the Cauchy model with location parameter* θ *and scale parameter 1, that is* $f(x \mid \theta) = \left\{\pi \left[1 + (x-\theta)^2\right]\right\}^{-1}$*. Then,*

$$S(\theta \mid x_1, \ldots, x_n) = 2\sum_{i=1}^{n} \frac{x_i - \theta}{1 + (x_i - \theta)^2}$$

and clearly there is no most efficient estimator for θ*. Recall that the minimal sufficient statistic in this case is the ordered sample...* ■

Example 4.6 *If* $X \mid \lambda \sim Ex(\lambda)$*, then*

$$S(x_1, \ldots, x_n \mid \lambda) = -n(\bar{x} - 1/\lambda)$$

and as a consequence $\bar{X}$ *is the most efficient estimator of* $1/\lambda$*. There are no most efficient estimator for* λ*.* ■

4.2.3 Uniformly minimum-variance unbiased estimation

Definition 4.3 (Uniformly minimum-variance unbiased estimator) *Let T be an unbiased estimator of $\tau(\theta)$. If for any other unbiased estimator of $\tau(\theta)$, W, we have*

$$Var(T \mid \theta) \leq Var(W \mid \theta) \ , \quad \forall \ \theta \in \Theta$$

then T is the so-called uniformly minimum-variance unbiased estimator of $\tau(\theta)$, or UMVUE. ■

Theorem 4.3 (Rao-Blackwell) *Let T be a a sufficient statistic for θ and U an unbiased estimator of $\tau(\theta)$. Then, $E[U \mid T]$ is an unbiased estimator of $\tau(\theta)$ with variance that is never superior to that of U. The two variances coincide iff U is a function of T.* ■

Note: The process of computing $E[U \mid T]$ is called Rao-Blackwellization.

Example 4.7 *Let $X_1, \ldots, X_n$ be a random sample of size n from a $N(\mu, 1)$ population. We are interested in estimating $\tau(\mu) = \mu^2 + 1$. We know that $U = \sum X_i^2/n = S^2 + \bar{X}^2$ is an unbiased estimator of $\tau(\mu)$. Since $\bar{X}$ is sufficient for μ, and is independent of S^2, it follows that*

$$E_\mu[U \mid \bar{X}] = E_\mu[S^2] + \bar{X}^2 = \frac{n-1}{n} + \bar{X}^2$$

is an unbiased estimator of $\tau(\mu)$ and its variance is not larger than that of U.

Remarks

- The UMVUE should be a function of a sufficient statistic
- When we Rao-Blackwellize an unbiased estimator, we do not obtain necessarily the UMVUE because there is always the possibility that if we had started with another unbiased estimator we could have obtained a smaller variance
- That cannot happen if we start with a sufficient and complete statistic: let $g(T) = E[U \mid T]$ and $g^\star(T)$ another unbiased estimator which is also a function of T. The completeness of T implies that $g(T) = g^\star(T)$.

Theorem 4.4 (Lehmann-Scheffé) *If the statistical model admits a sufficient and complete statistic T and there is at least an unbiased estimator of $\tau(\theta)$ then there is an UMVUE for $\tau(\theta)$ that is unique and a function of T.* ■

Strategies to find UMVU estimators in models admitting sufficient and complete statistics:

1. Obtain an unbiased estimator and Rao-Blackwellize it using the sufficient and complete statistic
2. Directly identify an unbiased estimator that is a function of a complete and sufficient statistic

Example 4.8 *Let $X_1, \ldots, X_n$, $n > 1$ be a random sample from a $Po(\theta)$, $\theta > 0$. The model is part of the one-parameter exponential family with natural parameter $\alpha = \ln \theta \in \mathbb{R}$, so the sufficient statistic $T = \sum X_i$ is complete. The goal is to find an UMVU estimator of $\tau(\theta) = e^{-\theta} = P(X_i = 0 \mid \theta)$.*

Following strategy 1., note that $U = I_{\{0\}}(X_1)$ is an unbiased estimator of $\tau(\theta)$. Now,

$$\begin{aligned} E[U \mid T = t] &= P(X_1 = 0 \mid T = t) \\ &= P(X_1 = 0 \mid \theta)\ P(\sum_{i=2}^{n} X_i = t \mid \theta)/P(T = t \mid \theta) \\ &= [(n-1)/n]^t\ . \end{aligned}$$

Hence, $[(n-1)/n]^T$ is the UMVUE of $\tau(\theta)$.

Following 2., we want to find $g(T)$ *such that* $E[g(T) \mid \theta] = e^{-\theta}$*. That is,*

$$\sum_{t=0}^{+\infty} g(t) e^{-n\theta} (n\theta)^t / t! = e^{-\theta} \quad \forall\ \theta > 0$$

$$\Leftrightarrow \sum_{t=0}^{+\infty} g(t)\ n^t\ \theta^t / t! = \sum_{t=0}^{+\infty} (n-1)^t\ \theta^t / t! \quad \forall\ \theta > 0$$

$$\Leftrightarrow g(t)\ n^t = (n-1)^t, \quad t \in \mathbb{N}_0$$

and we arrive to the same conclusion.

Most efficient and UMVU:

- If the model is regular and there exists a most efficient estimator of $\tau(\theta)$, then this estimator is necessarily the UMVUE
- The variance of the UMVUE will not necessarily be equal to the CRLB:
  - if the model is regular, there may not exist most efficient estimators, in which case the UMVUE will have a variance strictly larger than the CRLB
  - the UMVU criterion does not depend on regularity conditions, and hence, if these are not met, the variance of the UMVUE may be smaller than the CRLB defined as if the model were regular

4.2.4 Mean squared error

How should we compare unbiased estimators in terms of the dispersion of their distribution around $\tau(\theta)$?

Definition 4.4 (Mean squared error) *The mean squared error of an estimator T of $\tau(\theta)$ is*

$$MSE(T) = E_\theta[(T - \tau(\theta))^2] \ .$$

Remarks:

- T will be superior to $T^\star$ in mean squared error in the estimation of $\tau(\theta)$ if $\mathrm{MSE}(T) \leq \mathrm{MSE}(T^\star)$, $\forall\ \theta$.
- Markov inequality: $P(|T - \tau(\theta)| > \varepsilon) \leq \mathrm{MSE}(T)/\varepsilon^2$
- We have $\mathrm{MSE}(T) = \mathrm{Var}_\theta(T) + [b(T)]^2$

Example 4.9 *In the context of a normal population,* $E[S'^2] = \sigma^2$ *and* $E[S^2] = (n-1)\sigma^2/n$. *Nevertheless,*

$$\frac{2\sigma^2}{n} = Var(S'^2) > MSE(S^2) = \frac{2\sigma^2}{n} - \frac{3\sigma^2}{n^2}$$

4.2.5 Consistency

We now look at the behaviour of the sampling distribution of an estimator as $n \to +\infty$. Let $T_n = T(X_1, \ldots, X_n)$

Definition 4.5 (Consistency) *The estimator T_n is said to be a (weakly) consistent estimator of $\tau(\theta)$ if, for all $\theta \in \Theta$,*

$$\forall\ \varepsilon > 0 \qquad \lim_{n \to +\infty} P(|T_n - \tau(\theta)| > \varepsilon) = 0$$

that is, if, for all $\theta \in \Theta$, $T_n \xrightarrow{P} \tau(\theta)$. ■

Definition 4.6 (Mean-square consistency) *The estimator T_n is said to be a mean-square consistent estimator of $\tau(\theta)$ if, for all $\theta \in \Theta$,*

$$\lim_{n \to +\infty} E_\theta[(T_n - \tau(\theta))^2] = 0$$

that is, if, for all $\theta \in \Theta$, $T_n \xrightarrow{ms} \tau(\theta)$. ■

Remarks

- We already know (e.g. via Markov inequality) that mean-square convergence implies convergence in probability, so mean-square consistency implies weak consistency
- Since $E_\theta[(T_n - \tau(\theta))^2] = \mathrm{MSE}(T_n) = \mathrm{Var}(T_n) + [b(T)]^2$, a sufficient condition for weak consistency of T_n is that

$$\lim_{n \to +\infty} E[T_n] = \tau(\theta)$$

$$\text{and}$$

$$\lim_{n \to +\infty} \mathrm{Var}[T_n] = 0$$

- Consistency is not a very restrictive property. It is useful to exclude estimators, though.
- As an example, the sample mean is a weakly consistent estimator of the population mean, if it exists.

4.3 Estimation Methods

4.3.1 Method of Moments

- Idea: estimate population moments by the corresponding sample moments
- Let $\theta = (\theta_1, \ldots, \theta_k)$ be the vector of unknown parameters of the population
- $\mu'_r = E[X^r]$ will necessarily be a function of θ: $\mu'_r = \psi_r(\theta)$
- Consider the corresponding sample moments, $M_r = \sum_{i=1}^{n} X_i^r/n$ and form the system of equations $M'_r = \psi_r(\theta)$
- The solution to this equation determines the method of moments estimator of θ: $\hat{\Theta} = (\phi_r(X_1, \ldots, X_n),\ r = 1, \ldots, k)$.

Example 4.10 *Seja $X_1, \ldots, X_n$ a random sample from a $N(\mu, \sigma^2)$ population. We know that $\mu = \mu'_1$ and that $\sigma^2 = \mu'_2 - (\mu'_1)^2$, and hence $\mu'_2 = \sigma^2 + \mu^2$. The system of equations that we need to solve is*

$$\begin{cases} M'_1 = \mu \\ M'_2 = \sigma^2 + \mu^2 \end{cases}$$

and its solution determines that the method of moments estimator of μ is $M'_1 = \bar{X}$ and that of σ^2 is $M'_2 - (M'_1)^2 = S^2$.

Example 4.11 *Consider a* $X \sim G(\alpha, \lambda)$ *population. We know that* $\mu_1' = \alpha/\lambda$ *and that* $Var(X) = \alpha/\lambda^2$*. The method of moments estimator of* (α, λ) *is the solution to*

$$\begin{cases} M_1' = \alpha/\lambda \\ M_2' = \alpha/\lambda^2 + (\alpha/\lambda)^2 \end{cases}$$

that is, $(\bar{X}^2/S^2, \bar{X}/S^2)$.

Properties

- Recall the sampling properties of the sampling moments:

$$E[M_r \mid \theta] = \mu'_r$$
$$\mathrm{Var}(M'_r \mid \theta) = [\mu'_{2r} - (\mu'_r)^r]/n \equiv v_{rr}/n$$
$$\mathrm{Cov}_\theta(M'_r, M'_s) = [\mu'_{r+s} - \mu'_r \ \mu'_s]/n \equiv v_{rs}/n$$

- The multivariate central limit theorem states that, with $M = (M'_1, \ldots, M'_r)$ and $\mu = (\mu'_1, \ldots, \mu'_k)$,

$$\sqrt{n}(M - \mu) \xrightarrow{d} N_k(0, V)$$

  where $V = [v_{rs}]$

- The method of moments estimator of θ is a function of M, and its asymptotic properties can be determined using the delta method (possibly its multivariate version)

- Recall that if $T = h(M'_r)$ then

$$\sqrt{n}(h(M'_r) - h(\mu'_r)) \xrightarrow{d} N(0, [h'(\mu'_r)]^2 \ v_{rr})$$

4.3.2 Maximum likelihood method

- Idea: propose as an estimate of θ the value that maximizes the likelihood function
- Formally, the ML estimate of θ, when it exists, is $\hat{\theta}$ such that

$$L(\hat{\theta} \mid x_1, \ldots, x_n) \geq L(\theta \mid x_1, \ldots, x_n) \quad \forall\ \theta \in \Theta\ .$$

- The corresponding random variable determines the ML estimator

Remarks

- In most cases, it's easier (and equivalent) to find the value of θ which maximizes the log-likelihood function
- In some cases, but not always, this can be done by finding the zeroes of the score function:

$$\frac{d}{d\theta} \ln L(\theta \mid x_1, \ldots, x_n) = 0$$

  and verifying that the stationary point is indeed a global maximum
- The maximum likelihood estimators may not be unique: consider a random sample of size n from a $\mathrm{U}(\theta - 1/2, \theta + 1/2)$, $\theta \in \mathbb{R}$, population
- In many cases there will not be closed form expression for the estimate, in which case we have to resort to numerical methods — e.g., package `maxLik` in `R`

Example 4.12 *$X_1, \ldots, X_n$ denotes a random sample from a $Po(\lambda)$, $\lambda > 0$. Then*

$$L(\lambda \mid x_1, \ldots, x_n) \propto e^{-n\lambda}\ \lambda^{\sum x_i}$$

$$\ln L(\lambda \mid x_1, \ldots, x_n) = c - n\lambda + n\bar{x} \ln \lambda$$

$$\frac{d}{d\lambda}\ \ln L = -n + n\bar{x}/\lambda$$

and the score function has a zero at $\hat{\lambda} = \bar{x}$. It is easy to see that the second derivative of $\ln L$ at $\hat{\lambda}$ is negative, and hence this is indeed a point of global maximum. The ML estimator of λ is $\bar{X}$.

Example 4.13 *$X_1, \ldots, X_n$ denotes a random sample from a $G(\alpha, \lambda)$ population, with λ known. Let's determine the ML estimate of α:*

$$L(\alpha \mid x_1, \ldots, x_n) \propto \frac{\lambda^{n\alpha}}{[\Gamma(\alpha)]^n} \left(\prod_{i=1}^{n} x_i \right)^{\alpha - 1}$$

$$\ln L(\alpha \mid x_1, \ldots, x_n) = c + n\alpha \ln \lambda - n \ln \Gamma(\alpha) + (\alpha - 1) \sum \ln x_i$$

$$\frac{d}{d\alpha} \ln L = -n \ln \lambda - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum \ln x_i$$

In this case, the zeroes of the score function have to computed numerically.

Properties of the ML estimators:

- Invariance: if $\hat{\theta}$ is the maximum likelihood estimate of θ, and τ is a one-to-one function of θ, then $\tau(\hat{\theta})$ is the ML estimate of $\tau(\theta)$. This can be generalized to situations where $\tau(\cdot)$ is not one-to-one
- If T is sufficient and there is a ML estimator of θ, then this estimator is a function of T
- The most efficient estimator of θ, if it exists, is also the ML estimator of θ

Properties of the ML estimators (ctd):

- Under suitable regularity conditions, the ML estimator of θ, $\hat{\theta}$, satisfies

$$\sqrt{n}\ (\hat{\theta} - \theta) \xrightarrow{d} \mathrm{N}(0, [I_X(\theta)]^{-1})$$

- It is possible to replace $nI_X(\theta) = I_{X_1,\ldots,X_n}(\theta)$ in the formula above by
  - $n\ I_X(\hat{\theta})$, the Fisher information evaluated at the MLE
  - the observed Fisher information:

$$H(\theta) = -\frac{\partial^2}{\partial\theta^2}\ \ln L(\hat{\theta} \mid X_1, \ldots, X_n)$$

- The result above shows that MLE are BAN: Best Asymptotically Normal. They are consistent and asymptotically most efficient.