Asymptotic Statistics

Lecture 1

What is asymptotic statistics, why study it?

Asymptotic statistics: study the behaviour of statistical procedures as sample size $n \to \infty$.

Two main reasons:

- To obtain useful approximations of the distribution of test statistics, pivots used for confidence sets, etcetera.
- To assess the asymptotic optimality of statistical procedures.

Example: t-test (1)

Have i.i.d. sample $X_1, \dots, X_n$ with unknown mean μ. Want to test the hypothesis $H_0 : \mu = \mu_0$ for some given $\mu_0 \in \mathbb{R}$.

If the X_i are $N(\mu, \sigma^2)$ for some (unknown) σ^2, we can use the classical t-test, which rejects the null for large values of the statistic

$$T_n = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{S_n},$$

where

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

Example: t-test (2)

We know in the normal case that under $H_0 : \mu = \mu_0$, it holds that T_n has a t-distribution with $n-1$ degrees of freedom. Hence, to get a test of level α, we reject H_0 if $|T_n| \geq t_{n-1,\alpha/2}$. [1]

What if the X_i are not normal? Then in general, for fixed n the distribution of T_n is unknown. We will be able to prove however that as $n \to \infty$, if the X_i have finite variance, then we have the convergence in distribution

$$T_n \rightsquigarrow N(0,1),$$

i.e. for all $x \in \mathbb{R}$, under H_0,

$$\mathbb{P}(T_n \leq x) \to \Phi(x),$$

where Φ is the distribution function of the $N(0,1)$-distribution.

[1] $t_{n,\alpha}$ is the upper α-quantile of the t-distribution with n degrees of freedom.

Example: t-test (3)

Hence, whatever the distribution of the X_i is, we have, for large n, that T_n approximately has a $N(0,1)$-distribution under H_0. So we get a test of approximate level α if we reject H_0 if $|T_n| \geq \xi_{\alpha/2}$. [2]

Of course, this reasoning is only sensible if the sample size n is large enough!

[2] ξ_α is the upper α-quantile of the $N(0,1)$-distribution. (Book: z_α.)

Example: asymptotic optimality of the MLE (1)

Suppose we have an i.i.d. sample $X_1, \ldots, X_n$ from a positive density p_θ that depends smoothly on a parameter $\theta \in \mathbb{R}$.

The Cramér-Rao lower bound asserts that the variance of an unbiased estimator for θ is, under regularity conditions, bounded from below by $(ni_\theta)^{-1}$, where

$$i_\theta = \mathbb{V}\mathrm{ar}_\theta \frac{\partial}{\partial\theta} \log p_\theta(X_1)$$

is the Fisher information in a single observation.

Example: asymptotic optimality of the MLE (2)

It is typically difficult/impossible to find estimators that are optimal (for instance in the sense of minimal variance) for fixed sample size n. It is often possible however to exhibit estimators that are asymptotically optimal.

Under regularity conditions, it holds for the MLE $\hat{\theta}_n$ that if θ is the true parameter, then as $n \to \infty$,

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow N(0, i_\theta^{-1}).$$

Hence for large n, we have the approximations,

$$\mathbb{E}_\theta \hat{\theta}_n \approx \theta, \qquad \mathbb{V}\mathrm{ar}_\theta \hat{\theta}_n \approx (n i_\theta)^{-1}.$$

In this sense, the MLE is often asymptotically optimal in smooth parametric models.

Crucial tools: stochastic convergence

Chapter 1 of the notes. . .