A primer on Asymptotics

2 main concepts in asymptotic theory:

(1) Consistency

(2) Asymptotic Normality

Intuition

Consistency: As we get more data we eventually know the truth

asymptotic normality: averages of random variables behave "normally" if there are lots of variables to be averaged

Motivating Example

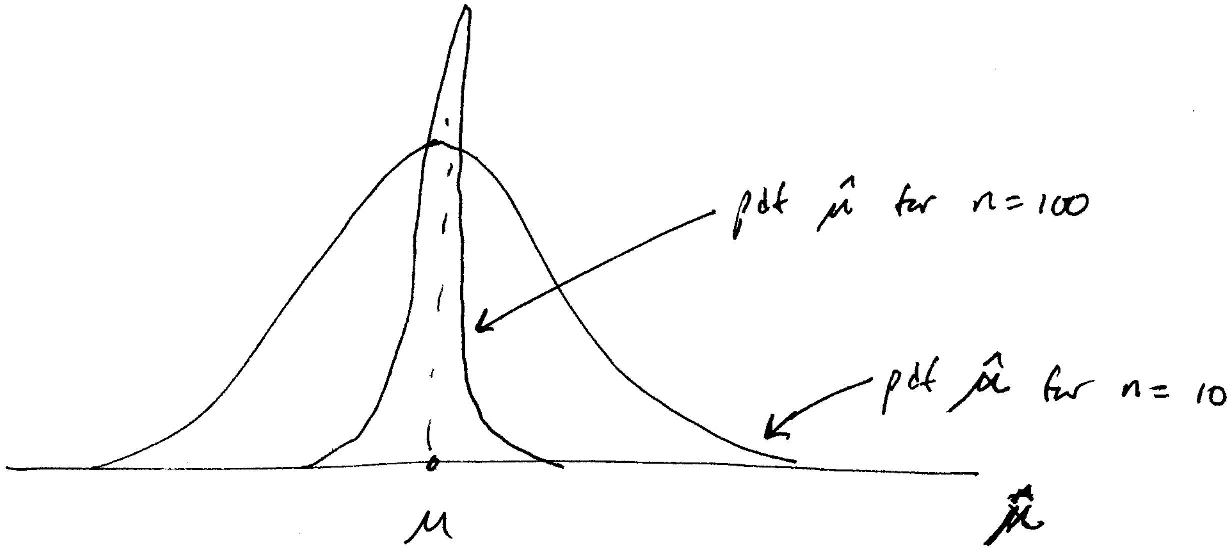
$X_1, \ldots, X_n$ iid with $E\{X_i\} = \mu$ and $\text{var}(X_i) = \sigma^2$.

Don't know $f(x_i, \theta)$. Suppose we know σ^2.

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} \quad , \quad E\{\hat{\mu}\} = \mu$$

$$\cancel{\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1} (x_i - \bar{x})^2}$$

$$\text{var}(\hat{\mu}) = \text{var}(\bar{x}) = \frac{\sigma^2}{n}$$



We don't know the pdf for $\hat{\mu}$ but we know

$$E\{\hat{\mu}\} = \mu$$

$$var(\hat{\mu}) = \frac{\sigma^2}{n}$$

So that as $n \to \infty$ $var(\hat{\mu}) \to 0$ and the pdf collapses at μ. Intuitively then as $n \to \infty$

$$\hat{\mu} \xrightarrow{P} \mu.$$

~~Note,~~ Furthermore, consider

$$z = \frac{\hat{\mu} - \mu}{\sqrt{var(\hat{\mu})}} = \frac{\hat{\mu} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \sqrt{n}\left(\frac{\hat{\mu} - \mu}{\sigma}\right)$$

For any value of n, $E[z] = 0$ and variance $(z) =$ but we don't know the pdf of z since we don't know $f(x)$; ~~The~~ Asymptotic normality says that as n

gets large the p.d.f of Z is well approximated by a $N(0,1)$:

$$\sqrt{n}\left(\frac{\hat{\mu}-\mu}{\sigma}\right) \overset{A}{\sim} N(0,1)$$

$$\Rightarrow \quad \hat{\mu} \overset{A}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

$\Rightarrow$ p.d.f. of $\hat{\mu}$ is well approximated by a Normal r.v. with mean μ and variance $\frac{\sigma^2}{n}$. Here the asymptotic variance is $\frac{\sigma^2}{n}$.

$$\sqrt{n}\left(\frac{\hat{\mu}-\mu}{\sigma}\right) = Z \overset{A}{\sim} N(0,1) \quad \text{for } n \text{ large}$$

Now

$$\hat{\mu} - \mu = \frac{1}{\sqrt{n}}\sigma \cdot Z$$

$$\Rightarrow \hat{\mu} = \mu + \frac{1}{\sqrt{n}}\sigma \cdot Z$$

$$\overset{A}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{for } n \text{ large.}$$

Remark: Often have to estimate the asymptotic variance

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \xrightarrow{?} \sigma^2$, $\hat{\sigma}^2$ = estimate of ...

Asymptotic Theory for Econometricians by Halbert White, Academic Press 1984.

Statistical Tools

(1) Laws of Large Numbers are used to establish consistency results

(2) Central Limit Theorems are used to establish asymptotic normality results.

Convergence in Probability & Laws of Large Numbers

Let $x_1, \ldots, x_n$ denote an iid sample with pdf $f(x;\theta)$.

Let

$$Y_1 = g(x_1)$$

$$Y_2 = g(x_1, x_2)$$

$$\vdots$$

$$Y_n = g(x_1, \ldots, x_n)$$

denote a sequence of random variables based on the sample. For example, let $x \sim N(\mu, \sigma^2)$ so that $\theta = (\mu, \sigma^2)'$ and define $Y_n = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \hat{\mu}$

Defn: Convergence in probability

Let $Y_1, \dots, Y_n$ be a sequence of rv's. We say Y_n converges in probability to a constant c and write

$$Y_n \xrightarrow{P} c \quad \text{as } n \to \infty$$

or

$$\operatorname*{plim}_{n \to \infty} Y_n = c$$

if $\forall \epsilon > 0$, $\lim_{n\to\infty} \Pr\left(|Y_n - c| > \epsilon \right) = 0$.

Defn. If $\hat{\theta}$ is an estimator of θ, then $\hat{\theta}$ is consistent for θ if

$$\hat{\theta} \xrightarrow{P} \theta$$

i.e. $\lim_{n \to \infty} \Pr\left(|\hat{\theta} - \theta| > \epsilon \right) = 0$.

All consistency proofs are based on a particular Law of Large Numbers (LLNs). A ~~law of~~ LLN is a result that states the conditions under which a sample average of random variables "converges" to a population expectation.

Example (Chebychev's weak LLN)

Let $X_1, \dots, X_n$ be iid r.v's with $E[X_1] = \mu < \infty$ and $\text{var}(X_1) = \sigma^2 < \infty$. Then

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\text{Sample average}} = \bar{X} \xrightarrow{P} \underbrace{E[X_1]}_{\text{population expectation}} = \mu .$$

The proof is based on the so-called Chebychev's inequality.

Let X be a r.v. with $E[X] = \mu$ and $\text{var}(X) = \sigma^2 < \infty$.

Then for every $\epsilon > 0$

$$\Pr\left(|X - \mu| \geq \epsilon \right) \leq \frac{\text{var}(X)}{\epsilon^2}$$

This result applied to $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ gives

$$\Pr\left(|\bar{X}-\mu| \geq \epsilon\right) \leq \frac{\mathrm{var}(\bar{X})}{\epsilon^2}$$

$$= \frac{\sigma^2/n}{\epsilon^2}$$

$$= \frac{\sigma^2}{n\epsilon^2}$$

and so

$$\lim_{n\to\infty} \Pr\left(|\bar{X}-\mu| \geq \epsilon\right) = \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0$$

Hence, $\bar{X} \xrightarrow{P} \mu = E[X_i]$.

<u>Remark</u>: Chebychev's LLN is an example of "convergence in MSE"; That is,

$MSE(\hat{\mu}, \mu) = \frac{\sigma^2}{n} \longrightarrow 0$ as $n \to \infty$

In general, if $MSE(\hat{\mu},\mu) \to 0$ as $n \to \infty$ then $\hat{\mu} \xrightarrow{P} \mu$. However, it can be the case that $\hat{\mu} \xrightarrow{P} \mu$ but $MSE(\hat{\mu},\mu) \not\to 0$.

The Weakest LLN is due to Kolmogorov:

Theorem (Kolmogorov's LLNs)

Let $X_1, \dots, X_n$ be iid random variables with $E[|X_i|] < \infty$ and $E[X_i] = \mu$. Then

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} E[X_i] = \mu.$$

<u>Results for the manipulation of probability limits</u>

Let b and c be constants

1. If $Y_n \xrightarrow{P} c$ then $b \cdot Y_n \xrightarrow{P} b \cdot c$

2. Let c & d be constants. If $Y_n \xrightarrow{P} c$ and $Z_n \xrightarrow{P} d$ then $Y_n + Z_n \xrightarrow{P} c + d$

3. Let c & d be constants ~~s.t. d≠0~~. If $Y_n \xrightarrow{P} c$ and $Z_n \xrightarrow{P} d$ then

   (i) $\dfrac{Y_n}{Z_n} \xrightarrow{P} \dfrac{c}{d}$ provided $d \neq 0$

   (ii) $Y_n \cdot Z_n \xrightarrow{P} c \cdot d$

4. If $Y_n \xrightarrow{P} c$ and $h(\cdot)$ is a continuous function ~~[illegible]~~ then $h(Y_n) \xrightarrow{P} h(c)$

(Slutsky's

example: Consistency of sample variance standard deviation

Let $X_1, \dots, X_n$ be an iid sample with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2 < \infty$. The sample mean and variance are

$$\bar{X} = \frac{1}{n} \Sigma_1^n X_i$$

$$S^2 = \frac{1}{n} \Sigma_1^n (X_i - \bar{X})$$

Claim: $S^2 \xrightarrow{P} \sigma^2$ as $n \to \infty$, $S \xrightarrow{P} \sigma$ as $n \to \infty$

Now

$$S^2 = \frac{1}{n} \Sigma_1^n (X_i - \bar{X})^2 = \frac{1}{n} \Sigma_1^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2)$$

$$= \frac{1}{n} \Sigma_1^n X_i^2 - 2\bar{X} \frac{1}{n} \Sigma_1^n X_i + \bar{X}^2$$

$$= \frac{1}{n} \Sigma_1^n X_i^2 - 2\bar{X}^2 + \bar{X}^2$$

$$= \frac{1}{n} \Sigma_1^n X_i^2 - \bar{X}^2$$

$$= \frac{1}{n} \Sigma_1^n X_i^2 - \left(\frac{1}{n} \Sigma_1^n X_i \right)^2$$

Next

(i) $\frac{1}{n} \Sigma_1^n X_i \xrightarrow{P} E[X_i] = \mu$ by Chebychev's LLN

(ii) $\frac{1}{n} \Sigma_1^n X_i^2 = \frac{1}{n} \Sigma_1^n W_i \xrightarrow{P} E[W_i]$ by Kolmogorov's LLNS since $E[|W_i|] < \infty$

Further, $E[W_i] = E[X_i^2]$

$$= Var(X_i) + E[X_i]^2$$

$$= \sigma^2 + \mu^2$$

iii) $\left(\frac{1}{n} \Sigma_1^n X_i \right)^2 \xrightarrow{P} (\mu)^2$ by Slutsky's theorem

(iv) $\frac{1}{n} \Sigma_1^n X_i^2 - \left(\frac{1}{n} \Sigma_1^n X_i \right)^2 \xrightarrow{P} (\sigma^2 + \mu^2) - \mu^2 = \sigma^2$

by additive property of convergence in probability

(v) let $s = \sqrt{s^2}$. Since

$$s^2 \xrightarrow{P} \sigma^2$$

Slutsky's theorem gives

$$(s^2)^{1/2} \xrightarrow{P} (\sigma^2)^{1/2}$$

i.e.

$$s \xrightarrow{P} \sigma.$$

Remark 1 Let $\underset{\sim}{Y}_1, \underset{\sim}{Y}_2, \ldots, \underset{\sim}{Y}_n$ be a sequence of random vectors in $\mathbb{R}^K$. We say that

$$\underset{\sim}{Y}_n \xrightarrow{P} \underset{\sim}{c} \quad , \; c \in \mathbb{R}^K$$

vector of constants

if

$$(\underset{\sim}{Y}_n)_j \xrightarrow{P} (\underset{\sim}{c})_j \quad \text{for } j = 1, \ldots, K$$

jth element of Y_n

j th element of $\underset{\sim}{c}$

Example. Let $X_1, \ldots, X_n$ be iid with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$. Define $\bar{X} = \frac{1}{n}\sum_1^n X_i$, $S^2 = \frac{1}{n}\sum_1^n (X_i - \bar{X})$

Then $\underset{\sim}{Y}_n = \begin{pmatrix} \bar{X} \\ S^2 \end{pmatrix}$ and using previous results we know that

$$\begin{pmatrix} \bar{X} \\ S^2 \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$$

~~Remark 2 (Generalization of Slutsky's theorem)~~

- A LLN that is particularly useful for regression is due to Markov

<u>Theorem (Markov's LLNs)</u>

Let $X_1, \ldots, X_n$ be a sequence of ~~independent~~ uncorrelated (but not identically distributed) random variables with finite means $\mu_i = E[X_i]$ and uniformly bounded

- variances $\sigma_i^2 = var(X_i) \leq M < \infty$ $i = 1, \ldots, n$. Then

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{n} \sum_{i=1}^{n} \mu_i$$
$$= \frac{1}{n} \left(\sum_{i=1}^{n} X_i - \mu_i \right) \xrightarrow{P} 0.$$

Alternatively,

$$\tilde{X} \xrightarrow{P} \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mu_i$$

This LLN is used in the HW.

-