Mathematical statistics
Week 3: Asymptotic properties of estimators

Péter Elek and Ádám Reiff

October 1 and 3, 2013

Notation.

- θ is population parameter, to be estimated
- sample statistic (or estimator) to estimate θ: $\widehat{\theta}$
- in case of a sample size of n, write $\widehat{\theta}$ as $\widehat{\theta}_n$

1. Asymptotic unbiasedness

Asymptotic unbiasedness.

- $\widehat{\theta}_n$ is an asymptotically unbiased estimator of θ if its expected value converges to θ as $n \to \infty$: $\lim_{n\to\infty} E\left(\widehat{\theta}_n\right) = \theta$.
- (If this is not true, then $\widehat{\theta}_n$ is asymptotically biased.)
- Example: the sample variance s^2 as the estimator of population variance σ^2
  - As $E\left(s^2\right) = \frac{n-1}{n}\sigma^2$, s^2 is a biased estimator of σ^2.
  - But as $\lim_{n\to\infty} E\left(s^2\right) = \lim_{n\to\infty} \frac{n-1}{n}\sigma^2 = \sigma^2$, s^2 is an asymptotically unbiased estimator of σ^2.

2. Consistency

2.1. Consistency and probability limit

Consistency.

- $\widehat{\theta}_n$ is a *consistent* estimator of θ if for each $\varepsilon > 0$, $\lim_{n\to\infty} \Pr\left(\left|\widehat{\theta}_n - \theta\right| < \varepsilon\right) = 1$.

- The sampling distribution of $\widehat{\theta}_n$ becomes more and more "concentrated" about θ.
- If this is not true, then $\widehat{\theta}_n$ is inconsistent.

- In this case we say that the *probability limit (plim)* of $\widehat{\theta}_n$ is θ (as $n \to \infty$), or $\operatorname{plim}_{n\to\infty} \widehat{\theta}_n = \theta$.

Probability limit (convergence in probability).

- The plim definition can be applied to any sequence of random variables, not just to estimators. (And the limit may be a random variable, not just a constant.)
- Let Z_n be a sequence of random variables. We say that $\operatorname{plim}_{n\to\infty} Z_n = Z$ if for each $\varepsilon > 0$, $\lim_{n\to\infty} \Pr\left(|Z_n - Z| < \varepsilon\right) = 0$.
- Short notation of plim is simply $\xrightarrow{p}$.
- Note: the formal definition of $\lim$ in analysis is $\lim_{n\to\infty} x_n = A$ if for each $\varepsilon > 0$ there is a threshold N_ε such that $|x_n - A| < \varepsilon$ for $n > N_\varepsilon$.

Relation between consistency and asymptotic unbiasedness.

- Result: if $\lim_{n\to\infty} E\left(\widehat{\theta}_n\right) = \theta$ and $\lim_{n\to\infty} Var\left(\widehat{\theta}_n\right) = 0$, then $\operatorname{plim}_{n\to\infty} \widehat{\theta}_n = \theta$.
  - i.e. if the variance of an asymptotically unbiased estimator converges to 0, then the estimator is consistent.
  - Proof is based on Chebyshev's inequality.
- (The converse is not true in general: one may construct consistent estimators, which are not even asymptotically unbiased.)
- Hence, to prove consistency, either
  - prove $\lim_{n\to\infty} E\left(\widehat{\theta}_n\right) = \theta$ *and* $\lim_{n\to\infty} Var\left(\widehat{\theta}_n\right) = 0$
  - or prove directly $\operatorname{plim}_{n\to\infty} \widehat{\theta}_n = \theta$ (see later).

Example: estimators of μ.

- Sample mean
  - We know that $E\left(\overline{X}\right) = \mu$ and $Var\left(\overline{X}\right) = \frac{\sigma^2}{n}$.
  - Hence $\lim_{n\to\infty} E\left(\overline{X}\right) = \mu$ and $\lim_{n\to\infty} Var\left(\overline{X}\right) = 0$,
  - so $\overline{X}$ is a consistent estimator of the population mean μ.

- "Artificial" example: the first element of a sample of size n
  - Since $E(X_1) = \mu$, it is asymptotically unbiased.
  - But its distribution is unchanged as $n \to \infty$, so it does not become more and more concentrated, so it is inconsistent for μ.

2.2. Properties and theorems

Properties of plim**.**

- If c is a non-stochastic constant, then $\operatorname{plim}_{n\to\infty} c = c$.
- If $x_1, x_2, x_3, \ldots$ is a non-stochastic series, then $\operatorname{plim}_{n\to\infty} x_n = \lim_{n\to\infty} x_n$.
- If $\operatorname{plim}_{n\to\infty} \widehat{\theta}_{1n} = \alpha$ and $\operatorname{plim}_{n\to\infty} \widehat{\theta}_{2n} = \beta$, then
  - $\operatorname{plim}_{n\to\infty} \left(\widehat{\theta}_{1n} + \widehat{\theta}_{2n}\right) = \alpha + \beta$,
  - $\operatorname{plim}_{n\to\infty} \left(\widehat{\theta}_{1n}\widehat{\theta}_{2n}\right) = \alpha\beta$
  - and if $\beta \neq 0$, $\operatorname{plim}_{n\to\infty} \left(\widehat{\theta}_{1n}/\widehat{\theta}_{2n}\right) = \alpha/\beta$.

Continuous mapping theorem.

- *Continuous mapping theorem*: If $\operatorname{plim}_{n\to\infty} \widehat{\theta}_n = \theta$ and $g(\cdot)$ is a continuous function, then $\operatorname{plim}_{n\to\infty} g\left(\widehat{\theta}_n\right) = g(\theta)$.
  - Proof: based on the properties of $\lim$.
- Therefore, if $\widehat{\theta}_n$ is consistent for θ and g is continuous, then $g\left(\widehat{\theta}_n\right)$ is consistent for $g(\theta)$.
- A similar statement is not true for unbiasedness!

Law of large numbers (LLN).

- *(Weak) law of large numbers (LLN)*: If $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with a finite expected value μ, then $\operatorname{plim}_{n\to\infty} \overline{X} = E(X) = \mu$.
  - This is even true if the variance σ^2 does not exist.
  - (Earlier we proved this when the variance does exist.)

2.3. Applications

Ways to prove consistency.

- So we can prove consistency in two ways:
  - prove $\lim_{n\to\infty} E\left(\widehat{\theta}_n\right) = \theta$ *and* $\lim_{n\to\infty} Var\left(\widehat{\theta}_n\right) = 0$ (see earlier)
  - or for some average apply the law of large numbers, and if necessary, apply the continuous mapping theorem and/or other plim operations.

Estimators of σ^2.

- s^2_{ideal} is a consistent estimator of σ^2.
  - By LLN, $\text{plim}_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right) = E\left[(X_i-\mu)^2\right] = \sigma^2$.
- s^2 (uncorrected sample variance) is consistent.
  - We know that $s^2 = s^2_{ideal} - \left(\overline{X}-\mu\right)^2$. Then $\text{plim}_{n\to\infty}\left(s^2\right) = \text{plim}_{n\to\infty}\left(s^2_{ideal}\right) - \text{plim}_{n\to\infty}\left[\left(\overline{X}-\mu\right)^2\right]$.
  - Here $\text{plim}_{n\to\infty}\left(s^2_{ideal}\right) = \sigma^2$, and as $\text{plim}_{n\to\infty}\overline{X} = \mu$, we have $\text{plim}_{n\to\infty}\left(\overline{X}-\mu\right) = 0$, and $\text{plim}_{n\to\infty}\left[\left(\overline{X}-\mu\right)^2\right] = 0$ (by continuous mapping).
- s^{*2} (corrected sample variance) is consistent.
  - $s^{*2} = \frac{n}{n-1}s^2$, and $\text{plim}_{n\to\infty}\frac{n}{n-1} = 1$

Estimators of σ.

- s and s^* (uncorrected and corrected sample standard deviation) are consistent estimators of σ (the population standard deviation).
  - $\text{plim}_{n\to\infty}\sqrt{s^2_n} = \sqrt{\sigma^2}$, i.e. $\text{plim}_{n\to\infty} s_n = \sigma$.
- A similar statement is not true for unbiased estimators: s^* is biased although s^{*2} is unbiased.

Further examples.

- Example 1: $\widehat{\mu}_1 = \frac{\sum_{i=1}^{n}X_i}{n-1}$ is consistent.
  - $\text{plim}_{n\to\infty}\frac{\sum_{i=1}^{n}X_i}{n-1} = \text{plim}_{n\to\infty}\frac{n}{n-1}\overline{X} = \text{plim}_{n\to\infty}\frac{n}{n-1}\,\text{plim}_{n\to\infty}\overline{X} = \mu$.

- Example 2: $\widehat{\mu}_2 = wX_1 + (1-w)\frac{\sum_{i=2}^{n} X_i}{n-1}$ is not consistent, as first term does not converge.

- Example 3: $\Pr(X=1) = p$, $\Pr(X=0) = 1-p$, draw an i.i.d. sample $X_1, X_2, \ldots, X_n$. Estimate $\frac{p}{1-p}$ with $\frac{\overline{X}}{1-\overline{X}}$. Is it unbiased? Is it consistent?

Notes on consistency.

- The above examples show that
  - consistent estimators are NOT necessarily unbiased;
  - unbiased estimators are NOT necessarily consistent;
  - asymptotically unbiased estimators are NOT necessarily consistent.

3. Asymptotic normality

3.1. Convergence in distribution and asymptotic normality

Convergence in distribution.

- A sequence of random variables Z_n *converges in distribution* to a continuous random variable Z if for any real z, $\Pr(Z_n \leq z) \rightarrow \Pr(Z \leq z)$ as $n \rightarrow \infty$.

- Intuitively, the shape of the distribution of Z_n gets closer and closer to that of Z.

- Notation of convergence in distribution: $\xrightarrow{d}$.

- Of course, we can apply the definition to estimators $\widehat{\theta}_n$.

Asymptotic normality.

- If the limiting distribution is normally distributed: $Z \sim N\left(\mu, \sigma^2\right)$, then Z_n is said to have an *asymptotic normal distribution*.
  - We can use the notation $Z_n \overset{A}{\sim} N\left(\mu, \sigma^2\right)$.

- If $Z \sim N(0,1)$, then Z_n has an *asymptotic standard normal distribution*. Then, for large n we have the approximation $\Pr(Z_n \leq z) \approx \Phi(z)$ where $\Phi(z)$ is the standard normal cdf.

3.2. Properties and theorems

Properties of convergence in distribution.

- *Continuous mapping theorem*: If $Z_n \xrightarrow{d} Z$ and $g(\cdot)$ is a continuous function, then $g(Z_n) \xrightarrow{d} g(Z)$.
- If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} \alpha$ (where X is a distribution and α is a constant), then
  - $X_n + Y_n \xrightarrow{d} X + \alpha$,
  - $X_n Y_n \xrightarrow{d} \alpha X$,
  - and if $\alpha \neq 0$, $X_n / Y_n \xrightarrow{d} X/\alpha$.

Central limit theorem (CLT).

- *Central Limit Theorem*: Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from a distribution with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Then $Z_n = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$.
- Earlier result: if X_i is *normally distributed*, then $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ is standard normal (exactly!). According to CLT, $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$ for *any distribution*.
- Normal distribution arises as the "natural" limit distribution. This helps to establish asymptotic normality of many estimators.

3.3. Applications

Normal approximation of the binomial distribution.

- Let $\Pr(X=1) = p$, $\Pr(X=0) = 1-p$ and draw an i.i.d. sample $X_1, X_2, \ldots, X_n$.
- Then by CLT, $\frac{\overline{X}-p}{\sqrt{p(1-p)/n}} \xrightarrow{d} N(0,1)$, i.e. $\overline{X} \overset{A}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$.
- Also, since $Y = \sum_{i=1}^{n} X_i = n\overline{X}$, $Y \overset{A}{\sim} N(np, np(1-p))$. But we know that Y is distributed as Binomial (n,p), so the binomial distribution converges to the normal distribution as $n \to \infty$.

Asymptotic distribution of the sample t-ratio.

- By CLT $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$, and by LLN and continuous mapping $\frac{\sigma}{s^*} \xrightarrow{p} 1$.
- Then, by the properties of the convergence in distribution, we have $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \frac{\sigma}{s^*} = \frac{\overline{X}-\mu}{s^*/\sqrt{n}} \xrightarrow{d} N(0,1)$.
  - This means that the sample t-ratio is approximately normal for large n (even if the original distribution is not normal).
- Note: we know that for normal random variables $\frac{\overline{X}-\mu}{s^*/\sqrt{n}} \sim t_{n-1}$ even in finite samples. It follows that the t_n-distribution converges to $N(0,1)$ as $n \to \infty$.
  - This also follows by LLN from the definition of the t-distribution.

Asymptotic distribution of s^2_{ideal}.

- $s^2_{ideal} = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}$ is an average of i.i.d. random variables $(X_i - \mu)^2$ $(i = 1, \ldots, n)$. Then by CLT,

$$\frac{s^2_{ideal} - E\left[(X_i-\mu)^2\right]}{\sqrt{Var\left[(X_i-\mu)^2\right]/n}} \xrightarrow{d} N(0,1),$$

- which means that

$$s^2_{ideal} \overset{A}{\sim} N\left(\sigma^2, \frac{Var\left[(X_i-\mu)^2\right]}{n}\right).$$

Asymptotic distribution of s^2_{ideal} for normal samples.

- If X_i are normally distributed, then we know that $Var\left(s^2_{ideal}\right) = \frac{2\sigma^4}{n}$, so in this case we have

$$\frac{s^2_{ideal} - \sigma^2}{\sqrt{\frac{2\sigma^4}{n}}} = \sqrt{\frac{n}{2}}\frac{s^2_{ideal} - \sigma^2}{\sigma^2} \xrightarrow{d} N(0,1),$$

- which means that

$$s^2_{ideal} \overset{A}{\sim} N\left(\sigma^2, \frac{2\sigma^4}{n}\right).$$

Asymptotic distribution of s^2.

- Let us use the notation $H = \sqrt{Var\left[(X_i-\mu)^2\right]}$.
- Then

$$
\begin{aligned}
s^2 &= s^2_{ideal} - \left(\overline{X}-\mu\right)^2 \\
s^2-\sigma^2 &= s^2_{ideal} - \sigma^2 - \left(\overline{X}-\mu\right)^2 \\
\sqrt{n}\frac{s^2-\sigma^2}{H} &= \sqrt{n}\frac{s^2_{ideal}-\sigma^2}{H} - \frac{1}{\sqrt{n}}\frac{\sigma^2}{H}\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2.
\end{aligned}
$$

- Here first term $\xrightarrow{d} N(0,1)$ and second term $\xrightarrow{p} 0$.
- So $\sqrt{n}\frac{s^2-\sigma^2}{H} \xrightarrow{d} N(0,1)$ and hence $s^2 \overset{A}{\sim} N\left(\sigma^2, \frac{H^2}{n}\right)$.
- It has the same asymptotic distribution as s^2_{ideal}.

Asymptotic distribution of s^{*2}.

- $s^{*2} = \frac{n}{n-1}s^2$. Can be seen:

$$\sqrt{n}\frac{s^{*2}-\sigma^2}{H} = \frac{n}{n-1}\sqrt{n}\frac{s^2-\sigma^2}{H} + \sqrt{\frac{n}{(n-1)^2}}\frac{\sigma^2}{H}.$$

- Here $\frac{n}{n-1} \xrightarrow{p} 1$, $\sqrt{n}\frac{s^2-\sigma^2}{H} \xrightarrow{d} N(0,1)$, $\sqrt{\frac{n}{(n-1)^2}}\frac{\sigma^2}{H} \xrightarrow{p} 0$.
- So $\sqrt{n}\frac{s^{*2}-\sigma^2}{H} \xrightarrow{d} N(0,1)$ and hence the asymptotic distribution of s^{*2} and s^2 is the same.

Example: asymptotic distribution of the sample mean.

- Suppose X is a random variable with $E(X) = \mu = 80$ and $\sqrt{Var(X)} = \sigma = 5$. We have a random sample of $n = 25$: $x_1, x_2, \ldots, x_{25}$.
- Sketch the distribution of X and the sampling distribution of $\overline{X}$.
- Find the approximate probability that $\overline{x} > 82$.

Solution.

- We do not know the distribution of X. But the distribution of $\overline{X}$ is asymptotically normal, with mean $\mu = 80$ and standard deviation $\sigma/\sqrt{n} = 1$.
- $\Pr\left(\overline{X} > 82\right) = 1 - \Pr\left(\overline{X} < 82\right) = 1 - \Pr\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < \frac{82-80}{1}\right)$, which is approximately $1 - \Pr\left(N(0,1) < 2\right)$. The result is 0.0228.

Another example.

- A battery manufacturer claims that if X is a random variable about the length of life of a battery, then $E(X) = \mu = 54$ and $\sqrt{Var(X)} = \sigma = 6$. We want to test this by buying a random sample of size $n = 50$: $x_1, x_2, \ldots, x_{50}$.
- If the battery manufacturer does not lie, what is the sampling distribution of $\overline{X}$?
- If the battery manufacturer does not lie, what is the probability of $\overline{x} \leq 52$?

Solution.

- The sampling distribution of $\overline{X}$ is asymptotically normal: $\overline{X} \overset{A}{\sim} N\left(54, \frac{6^2}{50}\right)$.
- $\Pr\left(\overline{X} \leq 52\right) = \Pr\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < \frac{52-54}{6/\sqrt{50}}\right)$, which is approximately $\Pr\left(Z < -2.35\right)$, where Z is a standard normally distributed random variable. The result is 0.0094.
- (So if this is the case, then there are two possibilities: either we are extremely unlucky, or the manufacturer is a liar.)

Example: asymptotic distribution of sample proportions.

- Suppose X is a Bernoulli random variable with $p = 0.2$. (For instance we want to estimate the proportion of votes for a party.) We have a random sample of $n = 400$: $x_1, x_2, \ldots, x_{400}$. What is the probability that $\overline{x}$ (the sample proportion) will be between 0.17 and 0.23?

Solution.

- $E(X) = p = 0.2$ and $\sqrt{Var(X)} = \sqrt{p(1-p)} = 0.4$.
- $\Pr\left(0.17 < \overline{X} < 0.23\right) = \Pr\left(\frac{0.17-0.2}{0.4/\sqrt{400}} < \frac{\overline{X}-p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0.23-0.2}{0.4/\sqrt{400}}\right)$, which is approximately $\Pr\left(-1.5 < Z < 1.5\right)$, since $\frac{\overline{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \overset{A}{\sim} N(0,1)$ by the CLT. The result is 0.8664.

4. Summary of important estimators

Sample mean: $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$.

Property	normal sample	general sample
$E(\overline{X})$	μ	μ
$Var(\overline{X})$	σ^2/n	σ^2/n
plim $\overline{X}$	μ	μ
Fin. d. of $\overline{X}$	$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$	unknown
Fin. d. of $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	$\sim N(0,1)$	unknown
Asy. d. of $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	$\sim N(0,1)$	$\xrightarrow{d} N(0,1)$
Fin. d. of $\frac{\overline{X}-\mu}{s^*/\sqrt{n}}$	$\sim t_{n-1}$	unknown
Asy. d. of $\frac{\overline{X}-\mu}{s^*/\sqrt{n}}$	$\sim N(0,1)$	$\xrightarrow{d} N(0,1)$

Estimators of variance in normal samples.

stat	s^2_{ideal}	s^2	s^{*2}
	$\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$	$\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}$	$\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$
E	σ^2	$\frac{n-1}{n}\sigma^2$	σ^2
Var	$\frac{2\sigma^4}{n}$	$\frac{2\sigma^4(n-1)}{n^2}$	$\frac{2\sigma^4}{n-1}$
plim	σ^2	σ^2	σ^2
fin. d.	$\frac{ns^2_{ideal}}{\sigma^2} \sim \chi^2_n$	$\frac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$	$\frac{(n-1)s^{*2}}{\sigma^2} \sim \chi^2_{n-1}$
asy. d.	$\frac{s^2_{ideal}-\sigma^2}{\sigma^2\sqrt{2/n}} \xrightarrow{d} N(0,1)$	$\frac{s^2-\sigma^2}{\sigma^2\sqrt{2/n}} \xrightarrow{d} N(0,1)$	$\frac{s^{*2}-\sigma^2}{\sigma^2\sqrt{2/n}} \xrightarrow{d} N(0,1)$

Estimators of variance in general samples.

stat	s^2_{ideal}	s^2	s^{*2}
	$\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$	$\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}$	$\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$
E	σ^2	$\frac{n-1}{n}\sigma^2$	σ^2
Var	unknown	unknown	unknown
plim	σ^2	σ^2	σ^2
fin. d.	unknown	unknown	unknown
asy. d.	$\frac{s^2_{ideal}-\sigma^2}{H/\sqrt{n}} \xrightarrow{d} N(0,1)$	$\frac{s^2-\sigma^2}{H/\sqrt{n}} \xrightarrow{d} N(0,1)$	$\frac{s^{*2}-\sigma^2}{H/\sqrt{n}} \xrightarrow{d} N(0,1)$