ECON 481
LECTURE 7: CONTIGUITY

Ivan A. Canay
Northwestern University
ECON 481-3

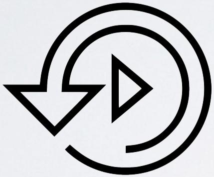
Past & Future

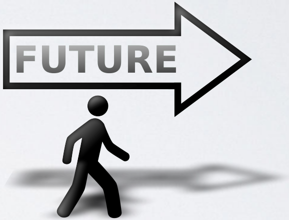
So Far

- Naive Power Approximations
- Local Power Approximations
- Symmetric Location Model
- *t*-test vs sign test

TODAY

- Absolute Continuity and LR
- Contiguity and Le Cam's 1st Lemma
- Le Cam's 3rd Lemma
- Wilcoxon Signed Ranked Test





Absolute continuity and likelihood ratios

- **Today**: about a technique to obtain the limit distribution of a sequence of statistics under underlying laws Q_n from a limiting distribution under laws P_n.

- Particularly useful to compute local asymptotic power of different statistics.

- First, let's start with a non-asymptotic analog

Definition

Let P and Q be measures on a measurable space $(\Omega, \mathcal{A})$. We say Q is **absolutely continuous** with respect to P if for every measurable set A we have that

$$P\{A\} = 0 \text{ implies } Q\{A\} = 0 \ .$$

Absolute continuity is denoted by $Q << P$.

Furthermore, P and Q are **orthogonal** if Ω can be partitioned as $\Omega = \Omega_P \cup \Omega_Q$ with $\Omega_P \cap \Omega_Q = \emptyset$ and $P\{\Omega_Q\} = Q\{\Omega_P\} = 0$. Orthogonality is denoted by $P \perp Q$.

Absolute continuity and likelihood ratios

Theorem (Radon-Nikodym)

Suppose Q and P are probability measures on $(\Omega, \mathcal{A})$. Then $Q << P$ if and only if there exists a measurable function $L(x)$ such that,

$$Q\{A\} = \int_A L(x)dP, \text{ for all } A \in \mathcal{A} \, .$$

The function $L(x) \equiv dQ(x)/dP(x)$ is called the Radon-Nikodym derivative (or density) or ***likelihood ratio****.*

Properties

- **Note!**: two measures P and Q need be neither absolutely continuous nor orthogonal.

- Suppose these measures have densities p and q wrt a measure μ. Then, $\Omega_P = \{p > 0\}$ and $\Omega_Q = \{q > 0\}$. The measure Q can be written as the sum $Q = Q^a + Q^\perp$ of the measures,

$$Q^a\{A\} = Q\{A \cap \{p > 0\}\}; \quad Q^\perp\{A\} = Q\{A \cap \{p = 0\}\}.$$

  This decomposition is called the **Lebesgue decomposition of Q** with respect to P.

- The likelihood ratio is a random variable $dQ/dP : \Omega \mapsto [0, \infty)$ and we want to study its law under P.

Lemma

Let P and Q be probability measures with densities p and q wrt a measure μ. Then,

1. $Q = Q^a + Q^\perp$, $Q^a << P$, $Q^\perp \perp P$.
2. $Q^a\{A\} = \int_A (q/p) dP$ *for every measurable set* A
3. $Q << P$ **if and only if** $Q\{p = 0\} = 0$ **if and only if** $\int (q/p) dP = 1$

IMPLICATIONS

- The function q/p is a density of Q^a with respect to P. It is denoted dQ/dP (not dQ^a/dP), so that $dQ/dP = q/p$, P-a.s.

- **Question**: Suppose that $T = f(X)$ is an estimator or test statistic. How can we compute the distribution of T under Q if we know how to compute probabilities under P?

- **Answer**: If Q is absolutely continuous wrt P, then the Q-law of a random variable X can be calculated from the P-law of the pair $(X, q/p)$ through the formula:

- **Remark**: The validity of this formula depends essentially on the absolute continuity of Q with respect to P, because a part of Q that is orthogonal to P cannot be recovered from any P-law.

QUESTIONS?

Contiguity

- We wish to consider an **asymptotic version** of the problem.
- Let $(\Omega_n, \mathcal{A}_n)$ be measurable spaces, each equipped with a pair of probabilities P_n and Q_n.
- Let T_n be some random vector and suppose the asymptotic distribution of T_n under P_n is easily obtained, but the behavior of T_n under Q_n is also required.
- **Example**: if T_n represents a test function for testing P_n versus Q_n, the power of T_n is the expectation under Q_n.
- **Question**: Under what conditions can a Q_n-limit law of random vectors T_n be obtained from suitable P_n-limit laws? The concept is called **contiguity** and essentially denotes a notion of "asymptotic absolute continuity".

Absolute Continuity for all n is not enough

Example

Let $P_n = N(0,1)$ and $Q_n = N(\xi_n, 1)$ with $\xi_n \to \infty$.

Contiguity

Definition (Contiguity)

Let Q_n and P_n be sequences of measures. We say that Q_n is **contiguous** w.r.t. to P_n, denoted $Q_n \triangleleft P_n$, if for each sequence of measurable sets A_n, we have that

$$P_n\{A_n\} \to 0 \Rightarrow Q_n\{A_n\} \to 0 \,.$$

We saw that absolute continuity does not imply contiguity. The following example provides an extension.

Example (cont)

Suppose P_n is the joint distribution of n i.i.d. observations $X_1, \ldots, X_n$ from $N(0,1)$ and Q_n is the joint distribution of n i.i.d. observations from $N(\xi_n, 1)$. Unless $\xi_n \to 0$, P_n and Q_n cannot be contiguous.

Le Cam's First Lemma

- For probability measures P and Q, Lemma (3) implies that the following are equivalent,

$$Q << P, \quad Q\left(\frac{dP}{dQ} = 0\right) = 0, \quad E_P\left[\frac{dQ}{dP}\right] = 1\ .$$

Le Cam: this equivalence persists if the three statements are replaced by their asymptotic counterparts.

- **Notation**: $\overset{P_n}{\rightsquigarrow}$ to denote $\overset{d}{\to}$ under P_n.

Lemma (Le Cam's first lemma)

Let P_n and Q_n be sequences of probability measures on measurable spaces $(\Omega_n, \mathcal{A}_n)$. Then the following statements are equivalent:

1. $Q_n \triangleleft P_n$.
2. *If $dP_n/dQ_n \overset{Q_n}{\rightsquigarrow} U$ along a subsequence, then* $\Pr\{U > 0\} = 1$.
3. *If $dQ_n/dP_n \overset{P_n}{\rightsquigarrow} V$ along a subsequence, then $E[V] = 1$.*
4. *For any statistic $T_n : \Omega_n \to \mathbf{R}^k$: If $T_n \overset{P_n}{\to} 0$, then $T_n \overset{Q_n}{\to} 0$.*

Corollary

Corollary

Let $dQ_n/dP_n \overset{P_n}{\rightsquigarrow} V$ *and suppose* $\log(V) \sim N(\mu, \sigma^2)$ *(this is,* V *has a log normal distribution). Then* Q_n *and* P_n *are mutually contiguous* **if and only if** $\mu = -\frac{1}{2}\sigma^2$*, which follows from* $E[V] = \exp(\mu + \frac{1}{2}\sigma^2)$*.*

Example (Contiguity does not imply absolute continuity)

Let $P_n = U[0,1]$, $Q_n = U[0, \theta_n]$, $\theta_n \to 1$, $\theta_n > 1$.

Examples

Example

Let $P_n = N(0,1)$ and $Q_n = N(\xi_n, 1)$. Then,

$$\log(L_n(X)) = \log\left(\frac{dQ_n}{dP_n}\right) = \xi_n X - \frac{1}{2}\xi_n^2 \, .$$

Example

Suppose P_n is the joint distribution of n i.i.d. observations $X_1, \ldots, X_n$ from $N(0,1)$ and Q_n is the joint distribution of n i.i.d. observations from $N(\xi_n, 1)$. Then,

$$\log(L_n(X_1, \ldots, X_n)) = \xi_n \sum_{i=1}^{n} X_i - \frac{n\xi_n^2}{2} ,$$

and so

$$\log(L_n(X_1, \ldots, X_n)) \sim N\left(-\frac{1}{2}n\xi_n^2, n\xi_n^2\right) \qquad \text{under} \quad P_n.$$

By the same arguments as before, Q_n is contiguous to P_n if and only if $n\xi_n^2$ remains bounded, i.e.

$$\xi_n = O(n^{-\frac{1}{2}}) .$$

Comments

- **Contiguity**. The sequences of measures P_n and Q_n do not separate asymptotically: given data from P_n or Q_n it is impossible to tell with certainty from which of the two sequences the data is generated, at least in an asymptotic sense, as $n \to \infty$.
- **Much more**: contiguity makes possible to derive asymptotic probabilities computed under Q_n from those computed under P_n. This is the content of Le Cam's third lemma.

Application

A popular application of contiguity is the comparison of statistical tests where one is given a sequence of tests ϕ_n concerning a parameter θ attached to a statistical model $(P_{n,\theta} : \theta \in \Theta)$ and corresponding power functions

$$\pi_n(\theta) = E_{P_{n,\theta}}[\phi_n] \, .$$

If P_{n,θ_0} and P_{n,θ_1} are asymptotically separated, then any "good" sequence of tests of the null hypothesis θ_0 versus the alternative θ_1 will have $\pi_n(\theta_0) \to 0$ and $\pi_n(\theta_1) \to 1$.

Contiguous alternatives will not allow this type of degeneracy, and hence may be used to pick a best test, or compute a relative efficiency of two given sequences of tests.

QUESTIONS?

Le Cam's Third Lemma

Lemma (Le Cam's third lemma)

Suppose that

$$\left(X_n, \log\left(\frac{dQ_n}{dP_n}\right)\right) \overset{P_n}{\rightsquigarrow} N\left(\left(\begin{array}{c} \mu \\ -\frac{1}{2}\sigma^2 \end{array}\right), \left(\begin{array}{cc} \Sigma & \tau \\ \tau' & \sigma^2 \end{array}\right)\right).$$

Then,

$$X_n \overset{Q_n}{\rightsquigarrow} N(\mu + \tau, \Sigma).$$

- **Result**: under the alternative distribution Q_n, the limiting distribution of the test statistic X_n is also normal but has mean shifted by

$$\tau = \lim_{n\to\infty} \text{Cov}\left(X_n, \log\left(\frac{dQ_n}{dP_n}\right)\right).$$

- **Testing**: with asymptotically normal test statistics X_n, a change from a null hypothesis to a contiguous alternative induces a change of asymptotic mean in the test statistics equal to the asymptotic covariance between X_n and $\log \frac{dQ_n}{dP_n}$ and no change of variance.

- It follows that good test statistics have a large (asymptotic) covariance with the log likelihood ratios.

Wilcoxon signed rank statistic

- **Application**: analyze the local asymptotic power of the Wilcoxon signed rank statistic.
- **Example**: Suppose P_θ is the distribution with density $f(x-\theta)$ on the real line. Suppose further that $f(x-\theta)$ is symmetric about θ. We observe $X_1, \ldots, X_n$ from f and wish to test the null $H_0 : \theta = 0$.
- **Wilcoxon signed rank statistic** serves to test this null and takes the form

$$W_n = n^{-3/2} \sum_{i=1}^{n} R^+_{i,n} \, \mathrm{sign}(X_i) \ ,$$

where

$$\mathrm{sign}(X_i) = \begin{cases} 1 & \text{if } X_i \geqslant 0 \\ -1 & \text{otherwise} \end{cases} \qquad \text{and} \qquad R^+_{i,n} = \sum_{j=1}^{n} I\{|X_j| \leqslant |X_i|\}$$

is the rank of $|X_i|$ among $|X_1|, \ldots, |X_n|$.

Wilcoxon signed rank: null hypothesis

$$W_n = n^{-3/2} \sum_{i=1}^{n} R_{i,n}^{+} \, \mathsf{sign}(X_i)$$

Wilcoxon signed rank: Le Cam's 3rd Lemma

Le Cam's third lemma: suggest that we look at

$$(W_n, \log(dP_{\theta_n}/dP_0)) \ .$$

Simplification: $P_{\theta_n} = N(\theta_n, 1)$ and $P_0 = N(0,1)$. In this case,

$$p_{\theta_n}(X_1, \ldots, X_n) = \prod_{i=1}^{n} (2\pi)^{-1/2} \exp[-\frac{1}{2}(X_i - \theta_n)^2]$$

and then,

$$\log L_n = \log(dP_{\theta_n}/dP_0) = \log \frac{e^{-\frac{1}{2}\sum_{i=1}^{n}(X_i^2 - 2X_i\theta_n + \theta_n^2)}}{e^{-\frac{1}{2}\sum_{i=1}^{n} X_i^2}}$$

$$= \theta_n \sum_{i=1}^{n} X_i - \frac{n}{2}\theta_n^2$$

$$= h \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i - \frac{1}{2} h^2 \ .$$

Wilcoxon signed rank: Local Alternative

$$\left(W_n, \log(dP_{\theta_n}/dP_0)\right) = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i \,\mathrm{sign}(X_i)\ ,\ h\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i - h^2/2\right) + o_p(1)\ ,$$

THE END!