Theoretical Statistics. Lecture 21.

Peter Bartlett

1. Motivation: Asymptotics of tests
2. Recall: Contiguity
3. Le Cam's lemmas. [vdV6]

Motivating example: asymptotic testing

Consider the asymptotics of a test. We have

- A parametric model P_θ for $\theta \in \Theta$.
- A null hypothesis $\theta = \theta_0$.
- An alternative hypothesis $\theta = \theta_0 + h$.

Test: compute the log likelihood ratio,

$$\lambda = \log \prod_{i=1}^{n} \frac{dP_{\theta_0+h}(X_i)}{dP_{\theta_0}(X_i)},$$

and reject the null hypothesis if it is sufficiently large.

Asymptotic testing

For a fixed alternative, this typically has trivial asymptotics. For example, suppose $P_\theta = N(\theta, \sigma^2)$. Then

$$\begin{aligned}
\lambda &= \log \prod_{i=1}^{n} \frac{dP_{\theta_0+h}}{dP_{\theta_0}}(X_i) \\
&= \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left((X_i - \theta_0)^2 - (X_i - \theta_0 - h)^2 \right) \\
&= \frac{1}{2\sigma^2} \sum_{i=1}^{n} (2h(X_i - \theta_0) - h^2) \\
&= \frac{nh}{\sigma^2}(\bar{X} - \theta_0) - \frac{nh^2}{2\sigma^2}.
\end{aligned}$$

Asymptotic testing

Under the null hypothesis, $\bar{X} \sim N(\theta_0, \sigma^2/n)$, so the log likelihood ratio is

$$\lambda \overset{\theta_0}{\sim} N\left(-\frac{nh^2}{2\sigma^2}, \frac{nh^2}{\sigma^2}\right).$$

[Notice that the mean is half the negative variance!]

Clearly (consider, for example, Chebyshev's inequality), for a fixed $h \neq 0$, we have $\lambda \xrightarrow{P} -\infty$ (that is, for all c, $\Pr(\lambda < c) \to 1$). So the asymptotics are rather trivial: asymptotically, we do not reject the null hypothesis.

Asymptotic testing

Consider instead a shrinking alternative: Replace h with $h_n \to 0$. Then

$$\begin{aligned}\lambda &= \log \prod_{i=1}^{n} \frac{dP_{\theta_0+h_n}}{dP_{\theta_0}}(X_i) \\ &= \frac{nh_n}{\sigma^2}(\bar{X}-\theta_0) - \frac{nh_n^2}{2\sigma^2} \\ &\overset{\theta_0}{\sim} N\left(-\frac{nh_n^2}{2\sigma^2}, \frac{nh_n^2}{\sigma^2}\right).\end{aligned}$$

So for $\sqrt{n}h_n \to h \neq 0$, its parameters approach $(-h^2/(2\sigma^2), h^2/\sigma^2)$.

And provided $h^2/(2\sigma^2) \gg h/\sigma$ (that is, $h/(2\sigma) \gg 1$, or $h_n/(2\sigma) \gg n^{-1/2}$), we do not reject the null hypothesis.

Asymptotic testing

These asymptotics are typical. Another example: The exponential family with sufficient statistic T has density $p_\theta(x) = \exp\left(T(x)\theta - A(\theta)\right)$. We have

$$\begin{aligned}
\lambda &= \log \prod_{i=1}^{n} \frac{dP_{\theta_0+h_n}}{dP_{\theta_0}}(X_i) \\
&= h_n \sum_{i=1}^{n} T(X_i) - n\left(A(\theta_0+h_n) - A(\theta_0)\right) \\
&= h_n \sum_{i=1}^{n} T(X_i) - n\left(A'(\theta_0)h_n + \frac{1}{2}A''(\theta_0)h_n^2 + o(h_n^2)\right) \\
&= h_n \sum_{i=1}^{n} \left(T(X_i) - P_{\theta_0}T(X_i)\right) - \frac{n}{2}A''(\theta_0)h_n^2 + o(nh_n^2).
\end{aligned}$$

Asymptotic testing

So if $h_n = h/\sqrt{n}$, and $T(X_i)$ has finite variance, then we have

$$\lambda = \frac{h}{\sqrt{n}} \sum_{i=1}^{n} (T(X_i) - P_{\theta_0} T(X_i)) - \frac{h^2}{2} A''(\theta_0) + o(1)$$

$$\overset{\theta_0}{\sim} N\left(-\frac{h^2 \operatorname{var}(T(X_1))}{2}, h^2 \operatorname{var}(T(X_1)) \right).$$

Contiguity

For these examples, the distributions are absolutely continuous wrt each other. In general, we need to make sure that the **likelihood ratios**

$$\frac{dQ_n}{dP_n}$$

make sense (at least asymptotically). We need an analogous asymptotic condition to absolute continuity ($P(A) = 0$ only if $Q(A) = 0$): **contiguity**.

Recall: Absolute Continuity

Definition:

1. $Q \ll P$ ("Q is **absolutely continuous** wrt P") means $\forall A$,

$$P(A) = 0 \implies Q(A) = 0.$$

2. $P \perp Q$ ("P and Q are **orthogonal**") means $\exists \Omega_P, \Omega_Q$,

$$P(\Omega_P) = 1, \qquad Q(\Omega_P) = 0,$$
$$Q(\Omega_Q) = 1, \qquad P(\Omega_Q) = 0.$$

Recall: Absolute Continuity

Suppose that P and Q have densities p and q wrt some measure μ. Define

$$Q^a(A) = Q\left(A \cap \{p > 0\}\right), \qquad Q^\perp(A) = Q\left(A \cap \{p = 0\}\right).$$

Lemma:

1. $Q = Q^a + Q^\perp$, with $Q^a \ll P$ and $Q^\perp P$ (Lebesgue decomposition)
2. $Q^a(A) = \int_A \frac{q}{p} dP \left(= \int_A \frac{dQ}{dP} dP \right).$
3. $Q \ll P \;\Leftrightarrow\; Q = Q^a \;\Leftrightarrow\; Q(p = 0) = 0 \;\Leftrightarrow\; \int \frac{q}{p} dP = 1.$

If $Q \ll P$ then $Qf(X) = P\left(f(X) \frac{dQ}{dP} \right).$

Contiguity

Definition: $Q_n \triangleleft P_n$ ("Q_n is contiguous wrt P_n") means, $\forall A_n$,

$$P_n(A_n) \to 0 \Longrightarrow Q_n(A_n) \to 0.$$

Contiguity: Examples

Example:

1. $P_n = N(0,1)$, $Q_n = N(\mu_n, \sigma^2)$ with $\sigma^2 > 0$ and $\mu_n \to \mu \in \mathbb{R}$.
   Then $Q_n \lhd P_n$ and $P_n \lhd Q_n$.
2. $P_n = N(0,1)$, $Q_n = N(\mu_n, \sigma^2)$ with $\sigma^2 > 0$ and $\mu_n \to \infty$.
   Then $A_n = [\mu_n, \mu_n + 1]$ shows that we do not have $Q_n \lhd P_n$.
   (But notice that we have $Q_n \ll P_n$ for all n.)
3. P_n is uniform on $[0,1]$, Q_n is uniform on $[\theta_n^0, \theta_n^1]$, $\theta_n^0 < \theta_n^1$, $\theta_n^0 \to 0$, $\theta_n^1 \to 1$.
   Then $P_n \lhd Q_n$ and $Q_n \lhd P_n$.
   (But notice that we have neither $P_n \ll Q_n$ nor $Q_n \ll P_n$.)

Contiguity

Lemma: [Le Cam's first lemma] The following are equivalent:

1. $Q_n \triangleleft P_n$.
2. $\dfrac{dP_n}{dQ_n} \overset{Q_n}{\rightsquigarrow} U$ along a subsequence $\Longrightarrow P(U > 0) = 1$.
3. $\dfrac{dQ_n}{dP_n} \overset{P_n}{\rightsquigarrow} V$ along a subsequence $\Longrightarrow \mathbf{E}V = 1$.
4. $T_n \overset{P_n}{\to} 0 \Longrightarrow T_n \overset{Q_n}{\to} 0$.

Contiguity

Notice that $\frac{dP_n}{dQ_n}, \frac{dQ_n}{dP_n}$ are non-negative and

$$\mathbf{E}_{P_n} \frac{dQ_n}{dP_n} \leq 1, \qquad \mathbf{E}_{Q_n} \frac{dP_n}{dQ_n} \leq 1.$$

So the likelihood ratios are uniformly tight, and therefore have a weakly converging subsequence (Prohorov's theorem). Le Cam's first lemma shows that the limits characterize contiguity. These characterizations are analogous to the characterizations we saw for absolute continuity:

$$Q \ll P \;\Leftrightarrow\; Q\left(\frac{dP}{dQ} = 0\right) = 0 \;\Leftrightarrow\; \mathbf{E}_P \frac{dQ}{dP} = 1.$$

Aside: Recall normal asymptotic testing

For $P_\theta = N(\theta, \sigma^2)$,

$$\begin{aligned}
\lambda &= \log \prod_{i=1}^{n} \frac{dP_{\theta_0+h_n}(X_i)}{dP_{\theta_0}(X_i)} \\
&= \frac{nh_n}{\sigma^2}(\bar{X} - \theta_0) - \frac{nh_n^2}{2\sigma^2} \\
&\overset{\theta_0}{\sim} N\left(-\frac{nh_n^2}{2\sigma^2}, \frac{nh_n^2}{\sigma^2}\right).
\end{aligned}$$

And for $\sqrt{n}h_n \to h \neq 0$, its parameters approach $(-h^2/(2\sigma^2), h^2/\sigma^2)$.

Contiguity

Here is an important example. **Local asymptotic normality**: log likelihood ratio of local alternative to true parameter is asymptotically normal.

Example: Suppose

$$\log \frac{dP_n}{dQ_n} \overset{Q_n}{\rightsquigarrow} N(\mu, \sigma^2).$$

Then $\frac{dP_n}{dQ_n} \overset{Q_n}{\rightsquigarrow} U$ implies $U = \exp(N(\mu, \sigma^2)) > 0$, so part (2) of the lemma shows that $Q_n \triangleleft P_n$.

Conversely, part (3) of the lemma shows that $P_n \triangleleft Q_n$ iff $\mathbf{E} \exp(N(\mu, \sigma^2)) = 1$.

Contiguity

Example: (Continued)

This is true iff

$$\begin{aligned} 1 &= \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left(x - \frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left(-\frac{(x-(\mu+\sigma^2))^2}{2\sigma^2}\right) dx \exp\left(\frac{(\mu+\sigma^2)^2 - \mu^2}{2\sigma^2}\right), \end{aligned}$$

which is equivalent to $\mu = -\sigma^2/2$.

(Alternatively, $\mathbf{E}\exp(Z) = M_Z(1)$ for $Z \sim N(\mu, \sigma^2)$. And $M_Z(t) = \exp(\mu t + \sigma^2 t/2)$, so $M_Z(1) = 1$ for $\mu = -\sigma^2/2$.)

Contiguity

Example: (Continued)

That is, for

$$\log \frac{dP_n}{dQ_n} \overset{Q_n}{\rightsquigarrow} N(\mu, \sigma^2),$$

$P_n \triangleleft Q_n$ iff $\mu = -\sigma^2/2$.

Contiguity and change of measure

Recall that, if $Q \ll P$ then we can write the Q-law of X in terms of the P-law of the pair $(X, dQ/dP)$.

Le Cam's third lemma shows an asymptotic version:
If Q_n is contiguous wrt P_n, then we can write the limit of the Q_n-law of a weakly converging random variable X_n in terms of the limit of the P_n-law of the pair $(X_n, dQ_n/dP_n)$.

Contiguity and change of measure

Theorem: [Le Cam's third lemma] If $Q_n \triangleleft P_n$ and

$$\left(X_n, \frac{dQ_n}{dP_n}\right) \overset{P_n}{\rightsquigarrow} (X, V),$$

then we can write $X_n \overset{Q_n}{\rightsquigarrow} L$ where the distribution L satisfies

$$\mathbf{E}_L f = \mathbf{E} f(X) V,$$

$$L(X \in A) = \mathbf{E}\left[1[X \in A] V\right] = \int_{A \times \mathbb{R}} v \, dP_{X,V}(x, v).$$

Contiguity and change of measure

Corollary: Suppose, for $X_n \in \mathbb{R}^k$,

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

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

Think of X_n as some test statistic, which approaches a normal under P_n. Think of Q_n as an alternative distribution, for which the asymptotic distribution of the log likelihood ratio is normal, with $\mu = -\sigma^2/2$. Under the alternative distribution Q_n, the asymptotic distribution of the statistic X_n also approaches a normal, but with the variance shifted by the limiting covariance between X_n and $\log(dQ_n/dP_n)$ under P_n.

Contiguity and change of measure: Proof

Let (X, Z) have the limiting distribution, so

$$\left(X_n, \log \frac{dQ_n}{dP_n}\right) \overset{P_n}{\rightsquigarrow} (X, Z),$$

hence

$$\left(X_n, \frac{dQ_n}{dP_n}\right) \overset{P_n}{\rightsquigarrow} (X, \exp(Z)).$$

Because $Z \sim N(-\sigma^2/2, \sigma^2)$, $Q_n \triangleleft P_n$. By Le Cam's lemma, $X_n \overset{Q_n}{\rightsquigarrow} L$, where $\int f(x)L(dx) = \mathbf{E}f(X)\exp(Z)$.

Contiguity and change of measure: Proof

Thus, the characteristic function of L is

$$\begin{aligned}\phi_L(t) &= \mathbf{E}\exp(it^T X + Z)\\ &= \phi_{X,Z}\left(\begin{pmatrix} t \\ -i \end{pmatrix}\right),\end{aligned}$$

But the normal distribution of (X, Z) implies its characteristic function is

$$\phi_{X,Z}\left(\begin{pmatrix} t \\ u \end{pmatrix}\right) = \exp\left(it^T\mu - \frac{iu\sigma^2}{2} - \frac{1}{2}\begin{pmatrix} t^T & u \end{pmatrix}\begin{pmatrix} \Sigma & \tau \\ \tau^T & \sigma^2 \end{pmatrix}\begin{pmatrix} t \\ u \end{pmatrix}\right).$$

Contiguity and change of measure: Proof

Substituting gives

$$\phi_L(t) = \exp\left(it^T(\mu+\tau) - \frac{1}{2}t^T\Sigma t\right),$$

which implies that L is $N(\mu+\tau,\Sigma)$.