Theoretical Statistics. Lecture 20.

Peter Bartlett

1. Recall:

   Functional delta method, differentiability in normed spaces, Hadamard derivatives. [vdV20]

2. Quantile estimates. [vdV21]

3. Contiguity. [vdV6]

Recall: Differentiability of functions in normed spaces

Definition: $\phi : D \to E$ is *Hadamard differentiable* at $\theta \in D$ tangentially to $D_0 \subseteq D$ if

$$\exists \phi'_\theta : D_0 \to E \text{ (linear, continuous)}, \forall h \in D_0,$$

$$\text{if } t \to 0, \|h_t - h\| \to 0, \text{ then}$$

$$\left\| \frac{\phi(\theta + th_t) - \phi(\theta)}{t} - \phi'_\theta(h) \right\| \to 0.$$

Recall: Functional delta method

Theorem: Suppose $\phi : D \to E$, where D and E are normed linear spaces. Suppose the statistic $T_n : \Omega_n \to D$ satisfies $\sqrt{n}(T_n - \theta) \rightsquigarrow T$ for a random element T in $D_0 \subset D$.
If ϕ is *Hadamard differentiable at* θ *tangentially to* D_0 then

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) \rightsquigarrow \phi'_\theta(T).$$

If we can extend $\phi' : D_0 \to E$ to a continuous map $\phi' : D \to E$, then

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) = \phi'_\theta(\sqrt{n}(T_n - \theta)) + o_P(1).$$

Recall: Quantiles

Definition: The *quantile function* of F is $F^{-1} : (0,1) \to \mathbb{R}$,

$$F^{-1}(p) = \inf\{x : F(x) \geq p\}.$$

- *Quantile transformation*: for U uniform on $(0,1)$,

$$F^{-1}(U) \sim F.$$

- *Probability integral transformation*: for $X \sim F$, $F(X)$ is uniform on [0,1] iff F is continuous on $\mathbb{R}$.
- F^{-1} is an inverse (i.e., $F^{-1}(F(x)) = x$ and $F(F^{-1}(p)) = p$ for all x and p) iff F is continuous and strictly increasing.

Empirical quantile function

For a sample with distribution function F, define the *empirical quantile function* as the quantile function F_n^{-1} of the empirical distribution function F_n.

$$F_n^{-1}(p) = \inf\{x : F_n(x) \geq p\} = X_{n(i)},$$

where i is chosen such that

$$\frac{i-1}{n} < p \leq \frac{i}{n},$$

and $X_{n(1)}, \ldots, X_{n(n)}$ are the order statistics of the sample, that is, $X_{n(1)} \leq \cdots \leq X_{n(n)}$ and

$$\big(X_{n(1)}, \ldots, X_{n(n)}\big)$$

is a permutation of the sample $(X_1, \ldots, X_n)$.

Quantiles

Define $\phi : D[a,b] \to \mathbb{R}$ as the pth quantile function $\phi(F) = F^{-1}(p)$.

Here, $D[a,b]$ is the set of *cadlag* functions on $[a,b]$, considered as a subset of $\ell^\infty[a,b]$:

$$\begin{aligned} \textit{cadlag} &= \textit{continue à droite, limite à gauche} \\ &= \text{right continuous, with left limits.} \end{aligned}$$

Quantiles

Theorem: If

- $F \in D[a, b]$,
- x_p satisfies $F(x_p) = p$,
- F is differentiable at x_p, with $F'(x_p) > 0$,

then ϕ is Hadamard-differentiable at F tangentially to

$$\{h \in D[a, b] : \ h \text{ is continuous at } x_p\} .$$

Its Hadamard derivative is the (continuous) function

$$\phi'_F(h) = -\frac{h(x_p)}{F'(x_p)}.$$

Quantiles

Recall:

$$\phi'_F(s_x - F) = \frac{p - s_x(F^{-1}(p))}{f(F^{-1}(p))}$$
$$= -\frac{(s_x - F)(x_p)}{F'(x_p)}.$$

Quantiles

Theorem: For

- $0 < p < 1$,
- F differentiable at $F^{-1}(p)$,
- $F'(F^{-1}(p)) = f(F^{-1}(p)) > 0$,

$$\sqrt{n}\left(F_n^{-1}(p) - F^{-1}(p)\right) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1[X_i \leq F^{-1}(p)] - p}{f(F^{-1}(p))} + o_P(1)$$

$$\rightsquigarrow N\left(0, \frac{p(1-p)}{f^2(F^{-1}(p))}\right).$$

Quantiles

Proof:

Because $\{x \mapsto 1[x \leq a] : a \in \mathbb{R}\}$ is Donsker, $\text{conv}\{x \mapsto 1[x \leq a] : a \in \mathbb{R}\}$ is Donsker, hence $\mathbb{G}_{n,F} = \sqrt{n}(F_n - F)$ converges weakly in $D[-\infty, \infty]$ to an F-Brownian bridge process $\mathbb{G}_F = \mathbb{G}_\lambda \circ F$.

[Recall that $\mathbb{G}_\lambda$ is the standard uniform Brownian bridge.]

The sample paths of $\mathbb{G}_F$ are continuous at points where F is continuous.

Now, $\phi : F \mapsto F^{-1}(p)$ is Hadamard-differentiable tangentially to the set D_0 of cadlag functions that are continuous where F is continuous. And the limiting process $\mathbb{G}_F$ takes its values in this set D_0. Furthermore, ϕ'_F is defined and continuous everywhere in $\ell^\infty[-\infty, \infty]$.

Quantiles

Hence, we can use the functional delta method:

$$\begin{aligned}\sqrt{n}(\phi(F_n) - \phi(F)) &= \phi'_F\left(\sqrt{n}(F_n - F)\right) + o_P(1) \\ &= \phi'_F\left(\mathbb{G}_{n,F}\right) + o_P(1).\end{aligned}$$

We can extend this result to the process $\sqrt{n}(F_n^{-1} - F^{-1})$, provided the differentiability conditions are satisfied over a set...

Quantiles

Theorem: Suppose

- $0 < p_1 < p_2 < 1$,
- $[a, b] = [F^{-1}(p_1) - \epsilon, F^{-1}(p_2) + \epsilon]$, for some $\epsilon > 0$,
- F continuously differentiable and with positive derivative f on $[a, b]$,
- $\phi : D[a, b] \to \ell^\infty[p_1, p_2]$ is defined by $\phi(G) = G^{-1}$.

Then ϕ is Hadamard differentiable at F tangentially to $C[a, b]$, with

$$\phi'_F(h) = -\left(\frac{h}{f}\right) \circ F^{-1}.$$

(Recall: $C[a, b]$ is the set of continuous functions on $[a, b]$.)

Quantiles

Theorem: For

- $0 < p_1 < p_2 < 1$,
- $[a, b] = [F^{-1}(p_1) - \epsilon, F^{-1}(p_2) + \epsilon]$, for some $\epsilon > 0$, and
- F continuously differentiable and with positive derivative f on $[a, b]$,

$$\sqrt{n}\left(F_n^{-1} - F^{-1}\right) \rightsquigarrow \frac{\mathbb{G}_\lambda}{f \circ F^{-1}},$$

where the convergence is in $\ell^\infty[p_1, p_2]$, and $\mathbb{G}_\lambda$ is the standard Brownian bridge.

(Recall: weak convergence in a metric space of functions is defined in terms of expectations of bounded continuous functions in the space.)

Contiguity

Motivation:
Suppose we wish to study the asymptotics of statistics T_n. Under the null hypothesis, say, $T_n \sim P_n$, we can show that $T_n \rightsquigarrow T$. What happens when the null hypothesis is not true? For instance, under the alternative hypothesis, $T_n \sim Q_n$. What can we say about the asymptotics?

We can relate them through **likelihood ratios** (also called **Radon-Nikodym derivatives**):

$$\frac{dP_n}{dQ_n} = \frac{p_n}{q_n},$$

where p_n and q_n are the corresponding densities. For these to make sense, we need the ratio to exist, and in particular we need $q_n = 0$ only if $p_n = 0$, at least asymptotically.

Absolute Continuity

Definition:

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

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

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

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

Absolute Continuity: Examples

Example:

1. $P = N(0,1)$, $Q = N(\mu, \sigma^2)$ with $\sigma^2 > 0$.
   Then $P(A) = 0 \Leftrightarrow Q(A) = 0$. Hence, $P \ll Q$ and $Q \ll P$.
2. $P = N(0,1)$, Q is uniform on $[0,1]$. Then $Q \ll P$ but not $P \ll Q$.
3. $P = N(0,1)$, Q is a mixture of a normal and a point mass at x: $Q(x) > 0$. Then $P \ll Q$ but not $Q \ll P$.
4. P is uniform on $[-1/2, 1/2]$, Q is uniform on $[0,1]$. Then neither $Q \ll P$ nor $P \ll Q$.

Absolute Continuity

We can always decompose Q into a part that is absolutely continuous wrt P and a part that is orthogonal (singular):
Suppose that P and Q have densities p and q wrt some measure μ.

Define

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

Absolute Continuity

Lemma:

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

Absolute Continuity

Proof:

(1) is immediate from the definitions.

(2):

$$
\begin{aligned}
Q^a(A) &= \int_{A \cap \{p>0\}} q \, d\mu \\
&= \int_{A \cap \{p>0\}} \frac{q}{p} p \, d\mu \\
&= \int_{A \cap \{p>0\}} \frac{q}{p} \, dP.
\end{aligned}
$$

Absolute Continuity

(3):

$$\begin{aligned} Q \ll P \Leftrightarrow Q &= Q^a \\ \Leftrightarrow Q^{\perp} &= 0 \\ \Leftrightarrow Q(p = 0) &= 0, \qquad \text{from the definitions.} \end{aligned}$$

Absolute Continuity

Also,

$$\int dQ = \int dQ^a + \int dQ^\perp$$
$$= \int \frac{q}{p}\, dP + \int dQ^\perp,$$

so

$$\int q/p\, dP = 1 \qquad \Leftrightarrow \qquad Q^\perp = 0.$$

Likelihood ratios

Write the likelihood ratio (= Radon-Nikodym derivative):

$$\frac{dQ}{dP} = \frac{q}{p}.$$

This is defined on $\Omega_P = \{p > 0\}$, and it is P-almost surely unique. It does not depend on the choice of dominating measure μ that is used to define the densities p and q.

Likelihood ratios: Change of measure

If $Q \ll P$ then $Q = Q^a$, so we can write the Q-law of $X : \Omega \to \mathbb{R}^k$ in terms of the P-law of the random pair $(X, dQ/dP)$, via

$$\mathbf{E}_Q f(X) = \mathbf{E}_P f(X) \frac{dQ}{dP},$$
$$Q(X \in A) = \mathbf{E}_P \left[1[X \in A] \frac{dQ}{dP} \right] = \int_{A \times \mathbb{R}} v \, dP_{X,V}(x, v),$$

where we have written the distribution under P of $(X, V) = (X, dQ/dP)$ as $P_{X,V}$.

This change of measure requires that Q is absolutely continuous wrt P.

Contiguity

We are interested in an asymptotic version of this change of measure. That is, we know the asymptotics of $T_n \sim P_n$, and we'd like to infer the asymptotics under an alternative sequence Q_n. When can we do that? We clearly need an asymptotic version of absolute continuity. This is called **contiguity**.

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

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