SOME NOTES ON HOTELLING TUBES

ROGER KOENKER

1. Introduction

We begin with the two examples of Johansen and Johnstone. The first is Hotelling's (1939) original example, and the second is closely related to our eventual goal of nonparametric applications.

1.1. **Hotelling's Original Example.** Consider the nonlinear regression model

$$Y_i = z_i^\top \alpha + \lambda_i(\tau)\beta + \varepsilon_i$$

where α, β, τ are unknown parameters, $\lambda_i(\cdot)$ are known functions and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. We want to test $H_0 : \beta = 0$. By the usual Frisch and Waugh (1933) trickery we can eliminate the α effect* and after redefining the notation, we are left with the likelihood ratio statistic

$$L = \inf_\tau \sum (Y_i - \hat{\beta}_\tau \lambda_i(\tau))^2 / \sum Y_i^2$$

Here $\hat{\beta}_\tau = Y^\top \lambda(\tau)/|\lambda(\tau)|^2$ so we can rewrite

$$\begin{aligned} L &= \inf_\tau |Y|^{-2}(|Y|^2 - 2(Y^\top \lambda)^2/|\lambda|^2 + (Y^\top \lambda)^2/|\lambda|^2) \\ &= 1 - \sup_\tau \left(\frac{Y^\top \lambda(\tau)}{|\lambda(\tau)||Y|} \right)^2 \\ &\equiv 1 - \sup_\tau (\gamma(\tau)^\top U)^2 \end{aligned}$$

Now $U = Y/|Y|$ is uniformly distributed on the sphere S^{n-1} and $\gamma(\tau) = \lambda(\tau)/|\lambda(\tau)|$ is a curve in S^{n-1}. Thus, the test rejects when $W = \sup_\tau \gamma(\tau)^\top U$ exceeds† some value $w = \cos\theta$ which is equivalent to

$$\begin{aligned} U \in \gamma^\theta &= \{u \in S^{n-1} : \sup_t u^\top \gamma(t) \geq \cos\theta\} \\ &= \{u \in S^{n-1} : d(u, \gamma) \leq (2(1-w))^{1/2}\} \end{aligned}$$

Version: May 4, 2012. These notes are intended as a readers guide to Johansen and Johnstone (1990) and thus to serve as a starting point for some work on uniform confidence bands for additive quantile regression models as currently implemented in `quantreg` by the function `rqss`.

*Hotelling obviously knew all about how to do this, and one doubts that he learned it from Frisch, but this would probably be hard to establish.

†Note that the original definition of L is such that we reject for *small* values, so $L < c$, implies we reject for $\sup_\tau \gamma(\tau)^\top U > w = \cos\theta$ for some critical value of θ.

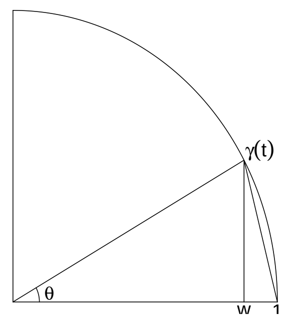

FIGURE 1. Angular distance from $\gamma(t)$ to $u = (1, 0)$.

This is illustrated by Figure 1 of Johansen and Johnstone, reproduced above. They call this the "angular or geodesic radius θ about γ:"

$$\begin{aligned} d^2(u,\gamma) &= \sin^2(\theta) + (1-\cos(\theta))^2 \\ &= 1 - 2\cos\theta + \cos^2\theta + \sin^2\theta \\ &= 2(1-\cos\theta). \end{aligned}$$

So when the distance is small, U falls *inside* tube, and we reject. This seems a bit counterintuitive, but is nonetheless correct. There are probably many ways to make this sound more intuitive. Let's try one possibility: Since it all boils down to a cosine, that is the simple correlation between $\lambda(\tau)$ and Y, we want to reject $H_0 : \beta = 0$ if this correlation/cosine is too large, but Y's that make it too large are the $Y's$ that fall *inside* the tube.

So how do we compute the critical w or equivalently the critical θ? Since $W > w \equiv \cos\theta$ is equivalent to $U \in$ tube, we need the volume of the tube. Let $|\gamma|$ denote the length of the arc $\gamma(\tau)$ on the sphere. (This can be approximated by the finite difference formula,

$$|\gamma| = \int \|\dot{\gamma}(\tau)\| d\tau \approx \sum_{i=2}^{m} \|(\gamma(\tau_i) - \gamma(\tau_{i-1}))\|,$$

with $\|\cdot\|$ the usual Euclidean norm, and the τ's on some relatively fine grid of m points. Note that in the finite difference approximation the $\tau_i - \tau_{i-1}$ that would normally appear in the difference quotient inside the norm expression cancels with the contribution of the $d\tau$.)

Theorem 1. *If γ is a nonclosed regular curve in S^{d-1} then for w near 1,*

$$\mathcal{P}(W \geq w) = \frac{|\gamma}{2\pi}(1-w^2)^{\frac{d-2}{2}} + \frac{1}{2}\mathcal{P}(B\left(\frac{1}{2}, \frac{d-1}{2}\right) \geq w^2) \tag{1}$$

where $B(1/2, (d-1)/2)$ is a beta random variable. If γ is closed then the second "cap" term is omitted.

This follows from a result of Hotelling (1939).

Theorem 2. *Let γ be a regular closed curve in S^{d-1} with length $|\gamma|$. And*

$$\begin{aligned}\gamma^\theta &= \{u \in S^{d-1} \big| \sup_t u^\top \gamma(t) \geq \theta\} \\ &= \{u \in S^{d-1} \big| d(u,\gamma) \leq (2(1-w))^{1/2}\}\end{aligned}$$

where $w = \cos\theta$. If θ is sufficiently small, then the volume of the tube $V(\gamma^\theta)$ is given by

$$V(\gamma^\theta) = |\gamma|\Omega_{d-2}\sin^{d-2}\theta \tag{2}$$

where $\Omega_{d-2} = \pi^{(d-2)/2}\Gamma(d/2)$ is the volume of the unit ball in R^{d-2}.

Heuristic: The formula is just

$$V(\gamma^\theta) = (\text{length of tube}) \cdot (\text{Volume of unit ball}) \cdot \text{ radius}^{d-2}$$

Recall that the volume of the unit ball in dimension d is $V = \pi^{d/2}/\Gamma((d+2)/2)$. When θ is larger, or γ is twisty, then the tube intersects itself and the formula needs some refinement.

When the curve isn't closed, then it needs "caps" on each end. These caps are given by

$$w_{d-2}\int_{\cos\theta}^1 (1-z^2)^{(d-3)/2}dz$$

where $w_{d-2} = 2\pi^{(d-1)/2}/\Gamma((d-1)/2)$ is the $(d-2)$-volume of S^{d-2}. (Note that the volume of the sphere, $V(S^{d-1}) = 2\pi^{d/2}/\gamma(d/2)$, is not the same as the volume of the ball. Note also that $(1-z^2)^{1/2}$ is again the radius and integrating out the r^{d-3} yields a $d-2$ dimensional volume.)

How do we get from (2) to (1)? Recall that U is uniform on the $(d-1)$ sphere so we need to divide by the volume of that sphere to evaluate the probability of being in the tube, so for closed curves,

$$\begin{aligned}\frac{V(\text{tube})}{V(\text{sphere})} &= \frac{|\gamma|\Omega_{d-2}\sin^{d-2}\theta}{2\pi^{d/2}/\Gamma(d/2)} \\ &= \frac{|\gamma|(\pi^{(d-2)/2}/\Gamma(d/2))\sin^{d-2}\theta}{2\pi(\pi^{(d-2)/2}/\Gamma(d/2))} \\ &= \frac{|\gamma|}{2\pi}(1-w^2)^{(d-2)/2}\end{aligned}$$

To include caps we also need to divide by the volume of the sphere. Note that

$$\begin{aligned}\mathcal{P}(B_{1/2,\frac{d-1}{2}} \geq w^2) &= \int_{w^2}^1 [x^{1/2-1}(1-x)^{\frac{d-1}{2}-1}/B(1/2,\frac{d-1}{2})]dx \\ &= \int_{w^2}^1 [x^{-1/2}(1-x)^{\frac{d-3}{2}}/B]dx\end{aligned}$$

Changing variables $x \to y^2$ we have

$$\begin{aligned}&= \int_{y_0}^1 [y^{-1}(1-y^2)^{\frac{d-3}{2}}/B]2ydy \\ &= 2\int_{y_0}^1 B^{-1}(1-y^2)^{\frac{d-3}{2}}dy\end{aligned}$$

It remains to show that $B^{-1} = w_{d-2}/\text{V(sphere)}$, which follows after a little simplification and recalling that $\Gamma(1/2) = \sqrt{\pi}$.

1.2. **Uniform confidence bands for nonparametric regression.** Consider the series expansion model

$$Y_i = \sum_{j=1}^{d} \beta_j a_j(t_i) + \varepsilon_i$$

with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ as before and $t \in I \subset \mathcal{R}$. Our objective is to find a positive c such that

$$P_{\beta,\sigma,\Sigma}(|\beta^\top a(t) - \hat{\beta}^\top a(t)| \le c\sigma(a(t)^\top \Sigma a(t))^{1/2} \; \forall \, t \in I) \approx 1 - \alpha$$

uniformly in β, σ. Johansen and Johnstone write this as

$$P_{\beta,\sigma,\Sigma}(T < c\sigma)$$

where

$$T = \sup_{a \in C} \frac{a^\top(\hat{\beta} - \beta)}{\sqrt{a^\top \Sigma a}}$$

Now consider $X \sim \mathcal{N}(\xi, \Sigma)$, so X plays the role of $\hat{\beta}$ and ξ of β. We'd like to make a confidence statement about $\{a^\top \xi | a \in C\}$ and C is some sort of "curve' ' . So now we write

$$T = T(X, \xi) = \sup_{a \in C} \frac{a^\top(X - \xi)}{\sqrt{a^\top \Sigma a}}$$

We want the distribution of T so we can obtain the confidence set

$$R_x = \{\{a^\top \xi\}_{a \in C} | T(X, \xi) < c_{1-\varepsilon}\}$$

where $P_{\xi,\Sigma}(T < c_{1-\varepsilon}) = 1 - \varepsilon$. Write $T = RW$ where

$$R^2 = (X - \xi)^\top \Sigma^{-1}(X - \xi) \sim X_d^2$$

and

$$\begin{aligned} W &= \sup_{a \in C} \frac{a^\top X - \xi)}{\sqrt{a^\top \Sigma a}\sqrt{(X - \xi)\Sigma^{-1}(X - \xi)}} \\ &= \sup_{a \in C} \frac{(\Sigma^{1/2} a)^\top \Sigma^{-1/2}(X - \xi)}{|\Sigma^{1/2} a||\Sigma^{-1/2}(X - \xi)|} \end{aligned}$$

Now to put things back into the earlier framework of γ and U we set

$$\begin{aligned} \gamma(a) &= \frac{\Sigma^{1/2} a}{|\Sigma^{1/2} a|} \\ U &= \Sigma^{-1/2}(X - \xi)/|\Sigma^{-1/2}(X - \xi)| \end{aligned}$$

So as before $\gamma = \gamma(C) \subset S^{d-1}$ and U is uniform on S^{d-1}. R and W don't depend on ξ, Σ or they do, but only via γ. $R^2 \perp\!\!\!\perp W$ and $R^2 \sim \chi_d^2$ so

$$\mathcal{P}(T > c) = \int_c^\infty \mathcal{P}(W > c/r)\mathcal{P}(R \in dr)$$

The random variable W has the same form as in the previous example so

$$\begin{aligned} \mathcal{P}(W > w) &= \frac{|\gamma|}{2\pi}(1 - w^2)^{(d-2)/2} + \frac{1}{2}\mathcal{P}(B \ge w^2) \\ &\equiv b_\gamma(w) \end{aligned}$$

Naiman (1986) bounds this probability by

$$\mathcal{P}(T > c) \leq \int_c^\infty \min\{b_\gamma(c/r), 1\} \mathcal{P}(R \in dr)$$

and Knowles (1987) ignores the $b_\gamma < 1$ constraint and integrates the bound to get,

$$\mathcal{P}(T > c) \leq \frac{|\gamma|}{2\pi} e^{-c^2/2} + 1 - \Phi(c)$$

This integral appears somewhat miraculous, but does actually work out provided that one carefully observes the $\mathcal{P}(R \in dr)$ term. Since $R^2 \sim \chi_d^2$, letting F denote the df of χ_d^2, we have,

$$\mathcal{P}(R \leq r) = \mathcal{P}(R^2 \leq r^2) = F(r^2)$$

so the corresponding density of R is

$$f_R(r) = 2rF'(r^2) = 2rf_{R^2}(r^2).$$

Once one has this bound then various other things follow easily. For example

$$\mathcal{P}(|T| > c) \leq 2\mathcal{P}(T > c).$$

Johansen and Johnstone (1990) give further details on the accuracy of the bounds and applications.

References

Frisch, R., and F. V. Waugh (1933): "Partial Time Regressions as Compared with Individual Trends," *Econometrica*, 1, 387–401.

Hotelling, H. (1939): "Tubes and Spheres in n-spaces, and a class of statistical problems," *Am J. Math*, 61, 440–460.

Johansen, S., and I. M. Johnstone (1990): "Hotelling's Theorem on the Volume of Tubes: Some Illustrations in Simultaneous Inference and Data Analysis," *The Annals of Statistics*, 18, 652–684.

Knowles, M. (1987): "Simultaneous confidence bands for random functions," Ph.D. thesis, Stanford University.

Naiman, D. (1986): "Conservative confidence bands in curvilinear regression," *Annals of Statistics*, 14, 896–906.