STAT 210B HWK #4 SOLUTIONS (DUE MARCH 29)

GARVESH RASKUTTI

(1) Given densities p_1 and p_2 with respect to some $\sigma-$finite measure μ, define that the *Hellinger distance* as follows:

$$h(p_1, p_2) = \left(\frac{1}{2}\int(\sqrt{p_1} - \sqrt{p_2})^2 d\mu\right)^{1/2},$$

and define the *variation distance* as follows:

$$||p_1 - p_2||_1 = \int |p_1 - p_2| d\mu.$$

Establish the following two inequalities:

$$h^2(p_1, p_2) \leq \frac{1}{2}||p_1 - p_2||_1 \leq h(p_1, p_2)[2 - h^2(p_1 - p_2)]^{1/2}$$

For the first inequality, we have

$$\begin{aligned} |p_1 - p_2| &= |(\sqrt{p_1} + \sqrt{p_2})(\sqrt{p_1} - \sqrt{p_2})| \\ &\geq (\sqrt{p_1} - \sqrt{p_2})^2, \end{aligned}$$

which implies

$$\begin{aligned} ||p_1 - p_2||_1 &= \int |p_1 - p_2| d\mu \\ &\geq \int (\sqrt{p_1} - \sqrt{p_2})^2 d\mu \\ &= 2h^2(p_1, p_2). \end{aligned}$$

Then the first inequality is established. For the second one,

$$\begin{aligned} ||p_1 - p_2||_1 &= \int |p_1 - p_2| d\mu \\ &= \int |\sqrt{p_1} - \sqrt{p_2}||\sqrt{p_1} + \sqrt{p_2}| d\mu \\ &\leq \left(\int (\sqrt{p_1} - \sqrt{p_2})^2 d\mu \int (\sqrt{p_1} + \sqrt{p_2})^2 d\mu\right)^{1/2} \\ &= (2h^2(p_1, p_2)(4 - 2h^2(p_1, p_2)))^{1/2} \\ &= 2h(p_1, p_2)[2 - h^2(p_1, p_2)]^{1/2}, \end{aligned}$$

which establishes the second inequality. The third line holds by Cauchy-Schwartz inequality, and the fourth line hold because of the fact that $\int(\sqrt{p_1} - \sqrt{p_2})^2 d\mu + \int(\sqrt{p_1} + \sqrt{p_2})^2 d\mu = 4$.

Date: Apr 2, 2011.

(2) Consider two probability measure P_1 and P_2 on a measurable space $(\mathcal{X}, \mathcal{A})$, dominated by a σ-finite measure μ (e.g. $\mu = P_1 + P_2$). Show that the Hellinger distance

$$\sqrt{\frac{1}{2}\int\left(\sqrt{\frac{dP_1}{d\mu}} - \sqrt{\frac{dP_2}{d\mu}}\right)^2 d\mu}$$

does not depend on the dominating measure μ. More generally, neither does

$$\int\left|\left(\frac{dP_1}{d\mu}\right)^{(1/a)} - \left(\frac{dP_2}{d\mu}\right)^{(1/a)}\right|^a$$

Say $P_1 \ll \lambda$ and $P_2 \ll \lambda$. We'll show that

$$\int\left|\left(\frac{dP_1}{d\lambda}\right)^{(1/a)} - \left(\frac{dP_2}{d\lambda}\right)^{(1/a)}\right|^a d\lambda \;=\; \int\left|\left(\frac{dP_1}{d\mu}\right)^{(1/a)} - \left(\frac{dP_2}{d\mu}\right)^{(1/a)}\right|^a d\mu$$

The strategy is to write both integrals w.r.t. the measure $\lambda + \mu$. Since $\mu \ll \lambda + \mu$, the Radon-Nikodym theorem implies the existence of a nonnegative function $d\mu/d(\lambda+\mu)$ such that

$$\begin{aligned}
&\int\left|\left(\frac{dP_1}{d\mu}\right)^{(1/a)} - \left(\frac{dP_2}{d\mu}\right)^{(1/a)}\right|^a d\mu \\
= &\int\left|\left(\frac{dP_1}{d\mu}\right)^{(1/a)} - \left(\frac{dP_2}{d\mu}\right)^{(1/a)}\right|^a \frac{d\mu}{d(\lambda+\mu)} d(\lambda+\mu) \\
= &\int\left|\left(\frac{dP_1}{d\mu}\frac{d\mu}{d(\lambda+\mu)}\right)^{(1/a)} - \left(\frac{dP_2}{d\mu}\frac{d\mu}{d(\lambda+\mu)}\right)^{(1/a)}\right|^a d(\lambda+\mu) \\
= &\int\left|\left(\frac{dP_1}{d(\lambda+\mu)}\right)^{(1/a)} - \left(\frac{dP_2}{d(\lambda+\mu)}\right)^{(1/a)}\right|^a d(\lambda+\mu)
\end{aligned}$$

The last equality uses a standard result about Radon-Nikodym derivatives. The exact same steps go through using μ as the dominating measure. Thus the Hellinger distance does not depend on the dominating measure.

(3) Suppose $||P_n - Q_n|| \to 0$. Show that P_n and Q_n are mutually contiguous

Note that

$$||P - Q|| = \sup_A |P(A) - Q(A)|.$$

For any A_n, if we have

$$P_n(A_n) \to 0,$$

then

$$\begin{aligned}
Q_n(A_n) &\leq P_n(A_n) + |Q_n(A_n) - P_n(A_n)| \\
&\leq P_n(A_n) + ||Q_n - P_n|| \\
&\to 0.
\end{aligned}$$

By definition, $Q_n \lhd P_n$. By symmetry, we have $Q_n \lhd\rhd P_n$.

(4) Suppose P_θ is the uniform distribution on $(0, \theta)$. Let P_θ^n denote the distribution of n iid draws from P_θ. Fix h and determine whether or not P_1^n and $P_{1+h/n}^n$ are mutually contiguous. Consider both $h > 0$ and $h < 0$.

(1) When $h > 0$, we have $P_{1+h/n}^n \ntriangleleft P_1^n$ but $P_1^n \triangleleft P_{1+h/n}^n$. In fact, write down the two densities:

$$p_1^n(x) = \begin{cases} 1, & \text{if } x \in [0,1]^n \\ 0, & \text{o.w.} \end{cases}$$

and

$$p_{1+h/n}^n(x) = \begin{cases} (1+h/n)^{-n}, & \text{if } x \in [0, 1+h/n]^n \\ 0, & \text{o.w.} \end{cases}$$

Use the fact that $(1 + h/n)^{-n} \to e^{-h}$, we have

$$\frac{p_1^n(x)}{p_{1+h/n}^n(x)} \overset{P_{1+h/n}^n}{\to} \begin{cases} e^h, & \text{if } x \in [0,1]^n \\ 0, & \text{if } x \in [0, 1+h/n]^n \backslash [0,1]^n \end{cases}$$

So

$$\frac{dP_1^n(x)}{dP_{1+h/n}^n(x)} \overset{P_{1+h/n}^n}{\rightsquigarrow} U,$$

and

$$P(U > 0) = \lim_{n\to\infty} P_{1+h/n}^n([0,1]^n) \to e^{-h} < 1.$$

By Le Cam's first lemma we have $P_{1+h/n}^n \ntriangleleft P_1^n$.

Meanwhile, we also have:

$$\frac{p_{1+h/n}^n(x)}{p_1^n(x)} \overset{P_1^n}{\to} e^{-h},$$

and Le Cam's first lemma tells us $P_1^n \triangleleft P_{1+h/n}^n$.

(2) When $h < 0$, the likelihood ratio becomes

$$\frac{p_1^n(x)}{p_{1+h/n}^n(x)} \overset{P_{1+h/n}^n}{\to} e^h,$$

and

$$\frac{p_{1+h/n}^n(x)}{p_1^n(x)} \overset{P_1^n}{\to} \begin{cases} e^{-h}, & \text{if } x \in [0, 1+h/n]^n \\ 0, & \text{if } x \in [0,1]^n \backslash [0, 1+h/n]^n \end{cases}$$

Using Le Cam's first lemma we have $P_1^n \ntriangleleft P_{1+h/n}^n$ but $P_{1+h/n}^n \triangleleft P_1^n$.

(5) Consider estimating the distribution function $P(X \le x)$ at a fixed point x based on a sample $X_1, ..., X_n$ from the distribution of X. A nonparametric estimator is $n^{-1} \sum_i 1(X_i \le x)$. If it is know that the true underlying distribution is $N(\theta, 1)$, another possible estimator is $\Phi(x - \bar{X})$. Calculate the relative efficiency of these estimators.

This is a problem on the relative efficiency of estimators (as opposed to tests in the next problem). The theory for this problem can be found in in VDV pages 108-111.

If we have two estimator sequences that converge to normal limit distributions at rate $\sqrt{n}$, their relative efficiency is defined to be the ratio of their asymptotic variances. So we only need to find the asymptotic distributions of $\hat{p}$ and $\tilde{p}$.

Consider $\tilde{p} = \frac{1}{n} \sum_{i=1}^n 1(X_i \le t)$.

Since

$$\mu(\theta) := E_\theta 1(X_i \le t) = P_\theta(X_i \le t) = \Phi_\theta(t) = \Phi(t - \theta)$$

and

$$\begin{aligned}\sigma^2(\theta) := \mathrm{Var}_\theta 1(X_i \le t) &= P_\theta(X_i \le t) - [P_\theta(X_i \le t)]^2 \\ &= \Phi(t-\theta)\,[1-\Phi(t-\theta)]\end{aligned}$$

We have by the CLT that $\frac{1}{\sqrt{n}}\sum_{i=1}^n [1(X_i \le t) - \mu(\theta)] \xrightarrow{d} N(0, \sigma^2(\theta))$. Equivalently,

$$\sqrt{n}\,(\tilde{p} - \mu(\theta)) \xrightarrow{d} N(0, \sigma^2(\theta))$$

Now consider $\hat{p} = \Phi(t - \bar{X})$. Since $X_1, \ldots, X_n$ are iid $N(\theta, 1)$, by the CLT we have $\sqrt{n}(\bar{X} - \theta) \xrightarrow{d} N(0,1)$. Now use the delta method (c.f. VDV p. 26) to find the asymptotic distribution of $\hat{p}$. Define $f : x \mapsto \Phi(t - x)$. Then the first order approximation to $f(\theta + a) - f(\theta)$ is

$$f'_\theta(a) = a\,[\partial_x \Phi(t-x)]_{x=\theta} = -a\phi(t-\theta)$$

where $\phi(x) = \Phi'(x)$. If we take $Z \sim N(0,1)$, then by the delta method,

$$\begin{aligned}\sqrt{n}\,[\hat{p} - \mu(\theta)] = \sqrt{n}\left[f(\bar{X}) - f(\theta)\right] &\xrightarrow{d} f'_\theta(Z) \\ &= -Z\phi(t-\theta) \\ &\stackrel{d}{=} N\left[0, (\phi(t-\theta))^2\right]\end{aligned}$$

Thus the asymptotic relative efficiency at θ is

$$\frac{(\phi(t-\theta))^2}{\Phi(t-\theta)\,[1-\Phi(t-\theta)]}.$$

As a result, the second estimator is asymptotically more efficient since it uses extra information about the distribution.