Statistics 522, Problem Set 4 Solutions

Wellner; 2/6/2004

1. Exercise 7.5, PfS page 290. For any df's F and G define

$$\lambda(F,G) \equiv \inf\{\epsilon > 0 : \ F(x-\epsilon) - \epsilon \le G(x) \le F(x+\epsilon) + \epsilon \ \text{ for all } x\} .$$

Show that λ is a metric and that the set of all df's under λ forms a complete and separable metric space. Also show that $F_n \to_d F$ ie equivalent to $\lambda(F_n, F) \to 0$.

Solution: Let $T(F,G)$ denote the set of ϵ's involved in the definition of λ.
(a) We first show that λ is a metric: It is trivially true that $\lambda(F,G) \ge 0$. The equality $\lambda(F,G) = \lambda(G,F)$ follows easily since $F(x-\epsilon) - \epsilon \le G(x) \le F(x+\epsilon) + \epsilon$ for all x if and only if $G(y-\epsilon) - \epsilon \le F(y) \le G(y+\epsilon) + \epsilon$ for all y (by first taking $y = x - \epsilon$, and then taking $y = x + \epsilon$). Furthermore $\lambda(F,G) = 0$ if and only if $F = G$, since $\lambda(F,F) = 0$ (note that $\epsilon \in T(F,F)$ for every $\epsilon > 0$), and if $\lambda(F,G) = 0$, then for every $\delta > 0$

$$F(x-\delta) - \delta \le G(x) \le F(x+\delta) + \delta \qquad \text{for all } \ x \in R \tag{0.1}$$

and

$$G(x-\delta) - \delta \le F(x) \le G(x+\delta) + \delta \qquad \text{for all } \ x \in R . \tag{0.2}$$

Letting $\delta \to 0$ on the right side of (0.1) and using the right continuity of F yields $G(x) \le F(x)$ for all x; doing the same in (0.2) yields $F(x) \le G(x)$ for all x; together we conclude that $F(x) = G(x)$ for all x, i.e. $F = G$.
Finally we show the triangle inequality for λ: if F, G, H are three df's, then

$$\lambda(F,G) \le \lambda(F,H) + \lambda(H,G) .$$

Note that for every $\delta > \lambda(F,H)$ we have

$$H(x-\delta) - \delta \le F(x) \le H(x+\delta) + \delta \qquad \text{for all } \ x \in R .$$

and for every $\epsilon > \lambda(H,G)$ we have

$$G(x-\epsilon) - \epsilon \le H(x) \le G(x+\epsilon) + \epsilon \qquad \text{for all } \ x \in R .$$

These inequalities imply that both

$$F(x) \le H(x+\delta) + \delta \le G(x+\delta+\epsilon) + \delta + \epsilon$$

and

$$F(x) \geq H(x-\delta) - \delta \geq G(x-\delta-\epsilon) - \delta - \epsilon$$

for all $x \in \mathbb{R}$. It follows that $\lambda(F,G) \leq \delta + \epsilon$ for every $\delta > \lambda(F,H)$ and $\epsilon > \lambda(H,G)$. Thus $\lambda(F,G) \leq \lambda(F,H) + \lambda(H,G)$ holds.

(b) Next we show that $F_n \to_d F$ if and only if $\lambda(F_n, F) \to 0$. First suppose that $\lambda(F_n, F) \to 0$, and fix $x \in C_F$. Then for every $\epsilon > 0$ there is a $\delta > 0$ such that $|F(y) - F(x)| < \epsilon$ if $|y - x| < \delta$. Choose N so large that $\lambda(F_n, F) < \epsilon \wedge \delta \equiv \xi$ for $n \geq N$. Then, for $n \geq N$ we have

$$F_n(x) \leq F(x+\xi) + \xi \leq F(x) + \epsilon + \epsilon = F(x) + 2\epsilon\,,$$

while, on the other hand,

$$F_n(x) \geq F(x-\xi) - \xi \geq F(x) - \epsilon - \epsilon = F(x) - 2\epsilon\,,$$

so that $|F_n(x) - F(x)| \leq 2\epsilon$ for $n \geq N$; thus $F_n(x) \to F(x)$ for $x \in C_F$. That is, $F_n \to_d F$.
Now suppose that $F_n \to_d F$. Let $\epsilon > 0$. Choose points $x_0, x_2, \ldots, x_k \in C_F$ such that $F(x_0) \leq \epsilon/2$, $F(x_k) \geq 1 - \epsilon$, and $x_j - x_{j-1} < \epsilon$ for $j = 1, \ldots, k$. Since $x_j \in C_F$ and $F_n \to_d F$, we can find a large N so that for $n \geq N$,

$$\max_{0 \leq j \leq k} |F_n(x_j) - F(x_j)| \leq \epsilon/2$$

for all $n \geq N$. Therefore, for fixed $j \in \{1, \ldots, k\}$ and $x_{j-1} \leq x \leq x_j$ we have

$$F_n(x) \leq F_n(x_j) \leq F(x_j) + \epsilon/2 \leq F(x+\epsilon) + \epsilon$$

and, on the other hand,

$$F_n(x) \geq F_n(x_{j-1}) \geq F(x_{j-1}) - \epsilon \geq F(x-\epsilon) - \epsilon\,.$$

For $x \in (-\infty, x_0]$,

$$F_n(x) \leq F_n(x_0) \leq F(x_0) + \epsilon/2 \leq \epsilon$$

and

$$F_n(x) \geq 0 \geq F(x_0) - \epsilon/2 \geq F(x-\epsilon) - \epsilon\,,$$

Similarly, for $x \in [x,\infty)$,

$$F(x-\epsilon) - \epsilon \leq F_n(x) \leq F(x+\epsilon) + \epsilon\,.$$

Combining these inequalities we conclude that (0.3) holds for all $x \in \mathbb{R}$ for $n \geq N$; i.e. $\lambda(F_n, F) \leq \epsilon$ for $n \geq N$. That is, $\lambda(F_n, F) \to 0$ as $n \to \infty$.

(c) Finally we show that $(\mathcal{F}, \lambda)$ is complete and separable where $\mathcal{F}$ is the collection of all distribution functions on $\mathbb{R}$. To show that $(\mathcal{F}, \lambda)$ is complete, suppose that $\{F_n\}$ is a Cauchy sequence for the Lévy metric λ: Then for every $\epsilon > 0$ there is an $N = N_\epsilon$ such that $m, n \ge N_\epsilon$ implies

$$F_n(x-\epsilon) - \epsilon \le F_m(x) \le F_n(x+\epsilon) + \epsilon \qquad \text{for all } \ x \in \mathbb{R}.$$

By taking lim sup on n and lim inf on m in the left inequality and vice-versa in the right inequality, we find that

$$\limsup_n F_n(x-\epsilon) - \epsilon \le \liminf_m F_m(x) \le \limsup_m F_m(x) \le \liminf_n F_n(x+\epsilon) + \epsilon\,.$$

Thus with $G(x) \equiv \liminf_n F_n(x)$, $H(x) \equiv \limsup F_n(x)$, and by letting $\epsilon \downarrow 0$ it follows that

$$H(x-) \le G(x) \le H(x) \le G(x+) \qquad \text{for all } \ x \in \mathbb{R}.$$

Since G and H are both non-decreasing, $H(x-) < G(x)$ can occur only at the (countably many) discontinuity points of H. Thus by taking

$$F(x) \equiv \begin{cases} \lim_n F(x) & \text{at continuity points of } H \\ H(x+) & \text{at discontinuity points of } \ H \end{cases}$$

it follows that $\lambda(F_n, F) \to 0$ as $n \to \infty$ where F is a proper distribution function (with $F(x) \to 1$ as $x \to \infty$, $F(x) \to 0$ as $x \to -\infty$).

2. Suppose that P_n and P are probability measures on $\mathbb{Z}$, the integers, and that $P_n \to_d P$. Show that $d_{TV}(P_n, P) \to 0$ where

$$d_{TV}(P_n, P) = \sup_{B \in \mathcal{B}} |P_n(B) - P(B)|\,.$$

Hint: You may use the fact that

$$d_{TV}(P_n, P) = \frac{1}{2} \int |p_n - p| d\mu$$

where $p_n = dP_n/d\mu$, $p = dP/d\mu$, and μ is any measure dominating P_n and P; recall Scheffé's theorem, Exercise 5.7, PfS, page 60.

Solution: Since P_n and P are all concentrated on the integers $\mathbb{Z}$ they have counting measure μ on $\mathbb{Z}$ as a common dominating measure, and the Radon-Nikodym derivatives $dP_n/d\mu$ and $dP/d\mu$ are given by the probability masses $p_n(k) \equiv P_n(\{k\})$, $p(k) \equiv P(\{k\})$. Since $P_n \to_d P$, it follows (by (i) of Proposition 11.2.2), $F_n(x) = P_n(-\infty, x] \to P(-\infty, x] \equiv F(x)$ at all continuity points of F. This implies that

$$p_n(k) \equiv P_n(\{k\}) = F_n(k+1/2) - F_n(k-1/2) \to F(k+1/2) - F(k-1/2) = P(\{k\}) \equiv p(k)$$

for all $k \in \mathbb{Z}$. Hence by Scheffé's theorem

$$d_{TV}(P_n, P) = \frac{1}{2}\int |p_n - p| d\mu = \frac{1}{2}\sum_{k\in\mathbb{Z}} |p_n(k) - p(k)| \to 0\,.$$

3. Let

$$p_n(x) = 2\sum_{k=1}^{2^{n-1}} 1_{[(2k-1)/2^n \le x < 2k/2^n]}$$

(draw a picture of p_n!), and let P_n be the probability measures on $\mathbb{R}$ having densities p_n with respect to Lebesgue measure λ.
(a) Show that $P_n \to_d P$.
(b) Show that $\int f dP_n \to \int f dP$ for all bounded measurable functions f on $[0,1]$. (Not just the continuous ones.)
(c) Find the total variation distance between P_n and $P = \lambda$ on $[0,1]$.

Solution: (a) Since p_n takes the value 2 on $[(2k-2)/2^n, (2k-1)/2^n)$ and the value 0 on $[(2k-2)/2^n, (2k-1)/2^n)$, the correspding distribution function is linear on $[(2k-1)/2^n, 2k/2^n)$ with slope 2, and constant on $[(2k-2)/2^n, (2k-1)/2^n)$ with height $(2k-2)/2^n$. Thus the distribution function $F_n(x) = \int_0^x p_n d\lambda$ has $F_n(2k/2^n) = 2k/2^n$ for $k = 1, \dots, 2^{n-1}$, and it is easily seen that, with $F(x) = x$ for $x \in [0,1]$, $\sup_{0\le x\le 1} |F_n(x) - F(x)| = 1/2^n \to 0$ as $n \to \infty$.
(b) Let f be a bounded measurable function. By Lusin's theorem (see e.g. Rudin (19xx), page 53, for each $\epsilon > 0$ there is a bounded continuous function $g = g_\epsilon$ with $\|g\|_\infty \le \|f\|_\infty$ such that

$$\lambda(D) \equiv \lambda(\{x :\ f(x) \ne g(x)\}) < \epsilon\,.$$

Then we have, since $P = \lambda$,

$$\begin{aligned} |P_n(f) - P(f)| &= |P_n(f-g) + P_n(g) - P(g) + P(g-f)| \\ &\le P_n(|f-g|) + |P_n(g) - P(g)| + P(|g-f|) \\ &\le 2\|f\|_\infty \int_D p_n d\lambda + |P_n(g) - P(g)| + 2\|f\|_\infty \lambda(D) \\ &\le 4\|f\|_\infty \epsilon + + |P_n(g) - P(g)| + 2\|f\|_\infty \epsilon\,, \end{aligned}$$

so that by (a) and the Portmanteau theorem since $g \in C_b(\mathbb{R})$,

$$\limsup_{n\to\infty} |P_n(f) - P(f)| \le 6\|f\|_\infty \epsilon\,.$$

But since $\epsilon > 0$ is arbitrary, this imlies that $P_n(f) \to P(f)$ for each bounded measurable function f.
(c) The total variation distance $d_{TV}(P_n, P)$ is given by

$$d_{TV}(P_n, P) = \frac{1}{2}\int |p_n - p| d\lambda\,.$$

Since p is the constant 1 on $[0,1]$ and p_n is either 0 or 2, it follows that $|p_n(x)-p(x)|=1$ for all $x\in[0,1]$. Hence $\int|p_n-p|d\lambda=1$ and $d_{TV}(P_n,P)=1/2$. Thus $P_n \not\to_{TV} P$.

4. For the following sequences of probability measures on $\mathbb{R}$ have densities p_n with respect to Lebesgue measure, which are uniformly tight? Explain why or why not.
(a) $p_n(x)=n^{-1}1_{[0,n]}(x)$.
(b) $p_n(x)=ne^{-nx}1_{[0,\infty)}(x)$.
(c) $p_n(x)=n^{-1}e^{-x/n}1_{[0,\infty)}(x)$.
(d) $p_n(x)=\sigma_n^{-1}\phi((x-\mu_n)/\sigma_n)$ with $\mu_n=3\cdot(-1)^n$ and $\sigma_n^2=\exp(-\sqrt{n})(2+\cos(\pi n))$.
(e) $p_n(x)=\sigma_n^{-1}\phi((x-\mu_n)/\sigma_n)$ with $\mu_n=3\cdot(-1)^n$ and $\sigma_n^2=n^{1/3}$.

Solution: (a) If $X_n\sim P_n$ with density p_n, then $E(X_n)=n/2\to\infty$, $F_n(x)=x/n\to 0$ for every $x\geq 0$, and $P(X_n>M)=1-M/n$ for $M\leq n$, which converges to 1 as $n\to\infty$. Thus $\{P_n\}$ is not tight.
(b) If $X_n\sim P_n$ with density p_n, then $E(X_n)=1/n\to 0$, $F_n(x)=1-e^{-nx}\to 1$, and $P(|X_n|>M)=e^{-nM}\to 0$ and $n\to\infty$ for every $M>0$. Thus $\{P_n\}$ is tight.
(c) If $X_n\sim P_n$ with density p_n, then $E(X_n)=n\to\infty$, $F_n(x)=1-e^{-x/n}\to 1-1=0$ for each $x\geq 0$, and $P(|X_n|>M)=e^{-M/n}\to 1$ as $n\to\infty$. Thus $\{P_n\}$ is not tight.
(d) Here the probability measures P_n are simply $N(3,\sigma_n^2)$ for n even and $N(-3,\sigma_n^2)$ for n odd with $\sigma_n^2\to 0$. Since $\sup_n\sigma_n^2\leq 3/e\leq 2$, it follows that

$$P(|X_n|>M)\leq 2P(3+\sqrt{2}Z>M)=2P(Z>(M-3)/\sqrt{2})\leq 2\sqrt{2}\phi((M-3)/\sqrt{2})/(M-3)$$

by Mills ratio, and hence M can be chosen to make this less than any $\epsilon>0$. Hence $\{P_n\}$ is tight.
(e) In this case the P_n's are $N(3,\sigma_n^2)$ and $N(-3,\sigma_n^2)$ for n even and odd respectively, but $\sigma_n^2\to\infty$. Thus the densities converge to 0 for all x and mass "escapes" to $\pm\infty$ as $n\to\infty$. Hence $\{P_n\}$ is not tight. (Note that $P(|X_n|>M)\to 1$ by the dominated convergence theorem.)

5. In the previous exercise, identify all the limiting probability distributions or sub-probability distributions (say in terms of the limiting distribution functions or sub-distribution functions) in each case.

Solution:
(a) In this case the mass is all escaping to infinity, no subseqeunces converges in distribution to a tight (proper) probability distribution, and the only limiting sub-distribution function is the function $H(x)\equiv 0$ for all x (with mass 1 at $+\infty$).
(b) In this case $F_n(x)\to 1$ for all $x>0$, (and $F_n(x)=0$ for $x<0$), so we conclude that $F_n\to_d 1_{[0,\infty)}$, the distribution function corresponding to the measure with all its mass at 0, δ_0.
(c) In this case the mass all escapes to $+\infty$ again, no subsequences converge in

distribution to tight limmiting distribution, and the only limiting sub-distribution function is the function $H(x) \equiv 0$ for all x (with mass 1 at $+\infty$).
(d) Here the probability measures are tight, and there are two subsequences converging in distribution to proper limits: $P_{2n} \to \delta_3$, while $P_{2n+1} \to_d \delta_{-3}$.
(e) In this case the probability measures are not tight: and there are two subsequences converging to the same subdistribution function, namely the subdistribution corresponding to $H(x) = 1/2$ for all x (with mass $1/2$ at $\pm\infty$.