Statistics 205a: Probability Theory **Fall 2002**

Lecture 14: Continuity Theorem

Lecturer: *Jim Pitman* Scribes: *Songhwai Oh* `<sho@eecs.berkeley.edu>`

We first make a few remarks about the characteristic functions. We have proven the uniqueness of a characteristic function of a random variable in $\mathbb{R}$. We can extend the same result to random vectors in $\mathbb{R}^d$ by applying the same argument. Consider a random vector $X = (X_1, X_2, \ldots, X_d) \in \mathbb{R}^d$. The Cramér-Wold device shown below implies that the distribution of X is uniquely identified by $\mathbb{E}(e^{i\theta \cdot X})$. Since the characteristic function of X is

$$\varphi(t) = \mathbb{E}(e^{it \cdot X}) = \mathbb{E}(e^{i \sum t_k X_k}),$$

where $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$, the characteristic function of X determines the distribution of X. Furthermore the characteristic function φ is determined by distributions of $\sum t_k X_k$, so is the distribution of X.

Theorem 14.1 (Cramér-Wold device) Let X_n, $1 \le n \le \infty$ be random vectors with characteristic function φ_n. A sufficient condition for $X_n \xrightarrow{d} X_\infty$ is that $\theta \cdot X_n \xrightarrow{d} \theta \cdot X_\infty$ for all $\theta \in \mathbb{R}^d$ (see Durrett [1], p.170).

Let us now discuss some issues with repect to the characteristic functions. Suppose you have a sequence of random variables X_n with characteristic function φ_n and

$$\varphi_n(t) \to \varphi(t) \qquad \text{for all } t \in \mathbb{R},$$

for some function $\varphi(t)$. Thus $\varphi(t)$ is a pointwise limit of $\varphi_n(t)$ but it may not be a characteristic function. So what are the conditions required to ensure that φ is a characteristic function of some X? If φ is indeed a characteristic function, then

$$\mathbb{E}(e^{itX_n}) \to \mathbb{E}(e^{itX}) \qquad \text{for all } t \in \mathbb{R}.$$

That is,

$$\mathbb{E}f(X_n) \to \mathbb{E}f(X) \qquad \text{for } f(x) = e^{itx}.$$

Hence

$$\mathbb{E}f(X_n) \to \mathbb{E}f(X)$$

for every bounded continuous function f. Thus $X_n \xrightarrow{d} X$ (weak convergence).

Theorem 14.2 (Continuity theorem) *(due to Paul Lévy)* Assume we have X_n with $\mathbb{E}(e^{itX_n}) \to \varphi(t)$ as $t \to \infty$, for all $t \in \mathbb{R}$ and some function $\varphi(t)$. Then the followings are equivalent:

i. (X_n) is tight, i.e. $\lim_{x\to\infty} \sup_n P(|X_n| > x) = 0$;

ii. $X_n \xrightarrow{d} X$ for some $X \in \mathbb{R}$;

iii. φ is a characteristic function of some $X \in \mathbb{R}$, i.e. $\varphi(t) = Ee^{itX}$;

iv. φ is a continuous function of t;

v. φ is continuous at $t = 0$.

If all the conditions (i)-(v) hold, then $X_n \xrightarrow{d} X$ for X as in (iii).

We first study two examples before proving the theorem. The first example illustrates the significance of the condition (v) of Theorem 14.2. The second example shows the tightness of the i.i.d. sequence under the setting of the central limit theorem for the i.i.d. case. So the alternative proof of the central limit theorem using characteristic functions is an application of the continuity theorem.

Example 14.1 Let Z be a r.v. with the standard normal distribution. Let $X_n = nZ$. Then

$$\mathbb{E}\left(e^{itX_n}\right) = e^{-\frac{1}{2}n^2t^2} \to \mathbf{1}(t=0) \quad \text{as } n \to \infty,$$

where t is held fixed (see Figure 14.1). But $\mathbf{1}(t=0)$ is not a characteristic function of any r.v. since it is not continuous at 0 (see the proof of Theorem 14.2). So X_n does not weakly converge to any r.v. $X \in \mathbb{R}$. However $X_n \xrightarrow{d} X \in \bar{\mathbb{R}}$. Now $P(X=\infty) = 1/2$ and $P(X=-\infty) = 1/2$. It shows that X_n is not tight and gives us an insight into the connection between the continuity at $t=0$ and tightness.

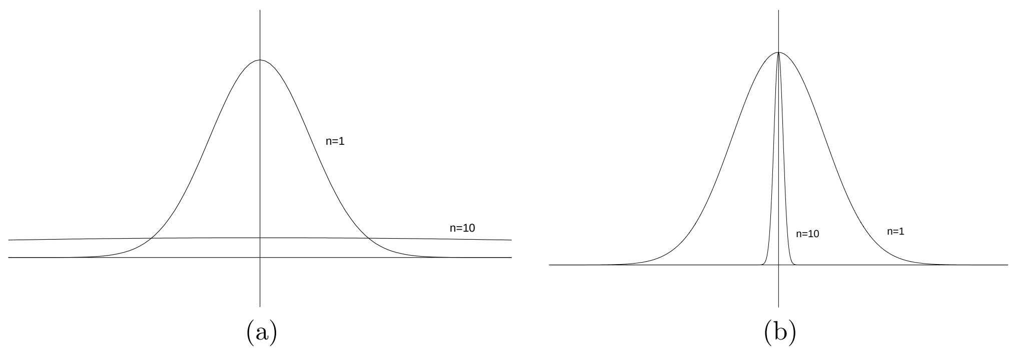

Figure 14.1: (a) Density function of X_n for $n=1$ and $n=10$; (b) Characteristic function of X_n for $n=1$ and $n=10$

Example 14.2 Let $X_1, X_2, \ldots, X_n$ be i.i.d. with $\mathbb{E}X_i = 0$ and $\mathbb{E}X_i^2 = \sigma^2 < \infty$. If $S_n = X_1 + \cdots + X_n$, then

$$\mathbb{E}\left(\left(\frac{S_n}{\sqrt{n}}\right)^2\right) = \sigma^2.$$

So by Chebyshev's inequality

$$P\left(\left|\frac{S_n}{\sqrt{n}}\right| > x\right) \leq \frac{\sigma^2}{x^2}.$$

Since it is true for all n,

$$\sup_n P\left(\left|\frac{S_n}{\sqrt{n}}\right| > x\right) \leq \frac{\sigma^2}{x^2} \to 0 \qquad \text{as } x \to \infty,$$

hence the sequence (X_n) is tight.

Proof Sketch: (Theorem 14.2)

- **(i) implies (ii)**: The complex exponentials of the form e^{itx} are bounded and continuous and the uniqueness theorem of characteristic functions implies that they are the determining class. Hence by Helly's selection theorem (Durrett [1] p.88) the tightness implies the existence of a distribution for a r.v. X such that $X_n \xrightarrow{d} X$.

- **(ii) implies (iii)**: Assume (ii) holds. Then $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all bounded continuous f. If we take $f(x) = e^{itx}$, then $\varphi_n(t) \to \mathbb{E}(e^{itX})$. Hence $\mathbb{E}(e^{itX}) = \varphi(t)$. Here we have assumed the uniqueness of a limit.
- **(iii) implies (iv)**: Notice that $|\varphi(t+h) - \varphi(t)| \le E|e^{ihX} - 1|$ and e^{ihX} goes to 1 as $h \to 0$. So by the bounded convergence theorem, $E|e^{ihX} - 1| \to 0$, so $\varphi(t)$ is a continuous function of t.
- **(iv) implies (v)**: If $\varphi(t)$ is continuous everywhere, it is continuous at $t = 0$.
- **(v) implies (i)**: The idea is to get a bound using the continuity of φ at $t = 0$ and show the sequence in (i) is tight. The complete proof is shown in p.99 of Durrett [1].

In conclusion, the uniqueness theorem and tightness imply the continuity theorem. ■

Example 14.3 (Cauchy processes) Let C_1 be a r.v. with the Cauchy distribution. Then the probability measure of C_1 is given by

$$P(C_1 \in dx) = \frac{dx}{\pi(1+x^2)}.$$

Notice that $E|C_1| = \infty$ and the Cauchy distribution has a heavy tail compared to other distributions. Using the inversion formula the characteristic function of C_1 is computed as

$$\varphi(\theta) = \mathbb{E}(e^{i\theta C_1}) = e^{-|\theta|}.$$

See Figure 14.2. Now let $C_1, \ldots, C_n$ be i.i.d. with the Cauchy distribution and $A_n = \frac{1}{n}(C_1 + \cdots + C_n)$. Then the characteristic function of A_n is

$$\begin{aligned} \mathbb{E}\Big(e^{i\theta A_n}\Big) &= \prod_{i=1}^{n} \mathbb{E}\Big(e^{i\theta C_i/n}\Big) \\ &= \prod_{i=1}^{n} e^{-|\frac{\theta}{n}|} \\ &= e^{-|\theta|}. \end{aligned}$$

Hence A_n has the same distribution as C_1. Recall that with the Gaussian distribution the same property holds with $\sqrt{n}$.

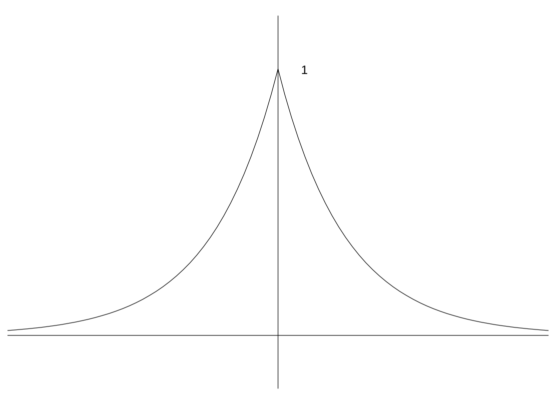

Figure 14.2: $e^{-|\theta|}$, the characteristic function of the Cauchy distribution

Theorem 14.3 (Polya's criterion) Every convex, symmetric, continuous function φ with $\varphi(0) = 1$ is $\varphi(t) = \mathbb{E}(e^{itX})$.

Proof Sketch: Here we give a graphical proof. See Durrett [1] for the formal proof of this theorem.

Let X be a r.v. uniformly distributed on $(-1, 1)$. Its density function is shown in Figure 14.3 (a) and the characteristic function of X is shown in Figure 14.3 (b). Let Y be another r.v. uniformly distributed on $(-1, 1)$ and independent of X. Then the density function for $X + Y$ can be computed by convolution and it is shown in Figure 14.3 (c)[1]. The characteristic function of $X + Y$ is shown in Figure 14.3 (d). Now the characteristic function shown in Figure 14.3 (d) is nonnegative and integrable so it can be defined as a density function with appropriate normalizing constant, namely π. Then by the inversion formula the tent function shown in Figure 14.3 (c) is the corresponding characteristic function upto a scaling factor. By (3.1g) of Durrett [1], a finite mixture of tents is a characteristic function. For example, if φ_1 and φ_2 are two different tent-shaped characteristic functions, then $\alpha_1\varphi_1 + \alpha_2\varphi_2$ with $\alpha_1 + \alpha_2 = 1$ is also a characteristic function (Figure 14.4). Since any convex and symmetric function is a limit of mixtures of tents, the result follows. ■

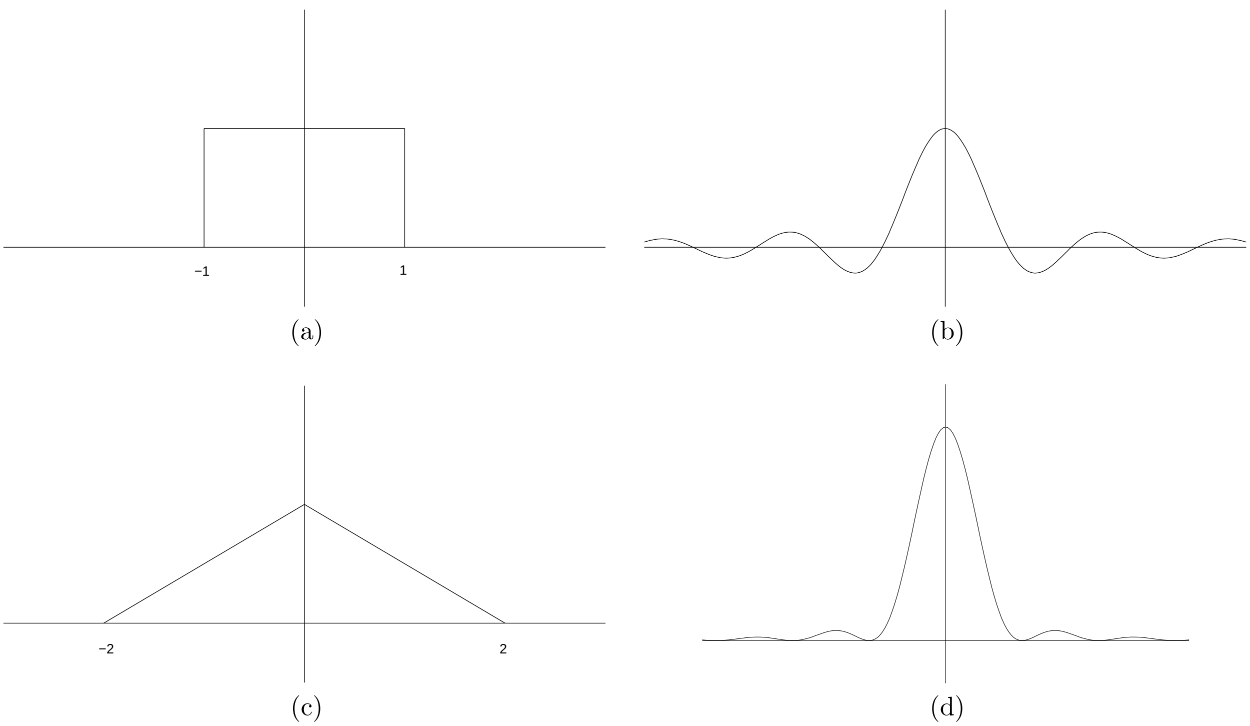

Figure 14.3: (a) Density function of a r.v. uniformly distributed on $(-1, 1)$; (b) $\frac{\sin t}{t}$, the characteristic function for (a); (c) density function of the sum of two independent r.v.'s, each uniformly distributed on $(-1, 1)$; (d) $\frac{\sin^2 t}{t^2}$, the characteristic function for (c).

[1] A function of this shape is known as a *tent* function.

$$tent_T(x) = \begin{cases} 1 - \frac{|x|}{T} & \text{if } |x| < T, \\ 0 & \text{otherwise} \end{cases}$$

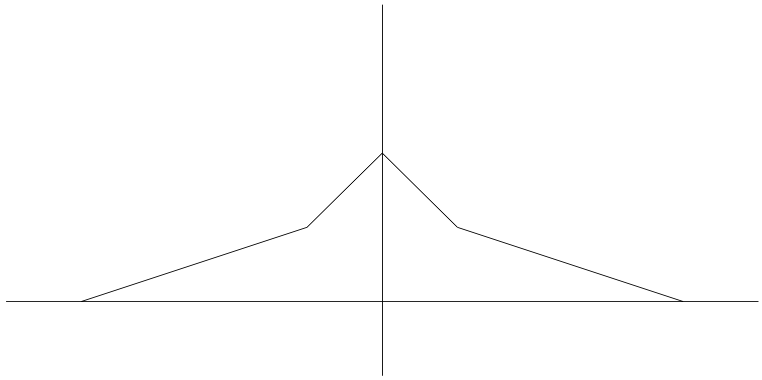

Figure 14.4: Mixture of two tents

References

[1] R. Durrett. *Probability: theory and examples.* Duxbury Press, Belmont, CA, second edition, 1996.