STATS 310A: Theory of Probability I **Autumn 2016/17**

Lecture 20: December 7

Lecturer: Persi Diaconis *Scribes: Kenneth Tay*

20.1 Characteristic Functions

20.1.1 Inversion and Uniqueness

Let μ be a probability on $\mathbb{R}$ with characteristic function $\phi(t)$.

Theorem 20.1 (Inversion Formula) *If $a < b$ with $\mu\{a\} = \mu\{b\} = 0$ (i.e. not atoms of the measure), then*

$$\mu(a,b] = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{itb}}{it}\phi(t)dt.$$

Proof: Let $I_T := \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{itb}}{it}\phi(t)dt$. Define $\text{sinc}(T) := \int_0^T \frac{\sin x}{x}dx$. Observe that

- (by change of variables)
$$\int_0^T \frac{\sin(\theta x)}{x}dx = \text{sgn}(\theta)\text{sinc}(T|\theta|).$$
- $\lim_{T\to\infty} \text{sinc}(T) = \frac{\pi}{2}$. (Aside: Note that $\int_0^\infty \frac{\sin x}{x}dx$ doesn't exist as a Lebesgue integral.)

Thus, writing out the characteristic function and using Fubini,

$$\begin{aligned}
I_T &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}dt\mu(dx) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-T}^{T} \frac{\sin[t(x-a)] - \sin[t(x-b)]}{t}dt\mu(dx) \\
&= \int_{-\infty}^{\infty} \frac{\text{sgn}(x-a)\text{sinc}(T|x-a|) - \text{sgn}(x-b)\text{sinc}(T|x-b|)}{\pi}\mu(dx) \\
&=: \int_{-\infty}^{\infty} \psi_{a,b}^T(x)\mu(dx).
\end{aligned}$$

Note that

$$\lim_{T\to\infty} \psi_{a,b}^T(x) = \begin{cases} 0 & x < a \\ \frac{1}{2} & x = a \\ 1 & a < x < b \\ \frac{1}{2} & x = b \\ 0 & x > b. \end{cases}$$

Hence, by the Dominated Convergence Theorem,

$$\begin{aligned}\lim_{T\to\infty} I_T &= \lim_{T\to\infty}\int_{-\infty}^{\infty}\psi_{a,b}^T(x)\mu(dx)\\ &= \int_{-\infty}^{\infty}\lim_{T\to\infty}\psi_{a,b}^T(x)\mu(dx)\\ &= \int_{-\infty}^{\infty}1_{(a,b]}(x)\mu(dx)\\ &= \mu(a,b].\end{aligned}$$

■

Corollary 20.2 *If $\phi_\mu(t) = \phi_\nu(t)$ for all t, then $\mu = \nu$.*

Proof: The family $\mathcal{P} = \{(a,b] : \mu\{a\} = \mu\{b\} = \nu\{a\} = \nu\{b\} = 0\}$ is π-system which generates the Borel sets on $\mathbb{R}$.

By Theorem 20.1, μ and ν agree on $\mathcal{P}$, thus they also agree on $\sigma(\mathcal{P}) \supseteq \mathcal{B}(\mathbb{R})$.

■

Remarks:

1. If $\phi(t)$ is integrable, and if F is the corresponding distribution function, then

$$\frac{F(x+h)-F(x)}{h} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{itx}-e^{it(x+h)}}{ith}\phi(t)dt.$$

   Taking limits on both sides and switching limit and integral by the Dominated Covergence Theorem, F has density

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\phi(t)dt.$$

   This is known as the **Fourier Transform**.

2. World of questions: What can we deduce about a measure from its transform? Best source is Feller, "Introduction to Probability & Applications Vol II," 2nd ed, Chapter 15.

20.1.2 Examples of Characteristic Functions

Here are some examples of characteristic functions:

	Density	Characteristic Function
Normal	$\dfrac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}}$	$e^{it\mu-\sigma^2t^2/2}$
Uniform	$\dfrac{1}{b-a}$ for $a<x<b$	$\dfrac{e^{itb}-e^{ita}}{it(b-a)}$
"Tent"	$1-\lvert x\rvert$ for $\lvert x\rvert<1$	$\dfrac{4\sin^2(t/2)}{t^2}$ or $\dfrac{2(1-\cos t)}{t^2}$
"Tent" Converse	$\dfrac{1}{\pi}\dfrac{1-\cos x}{x^2}$	$1-\lvert t\rvert$ for $\lvert t\rvert<1$

Note that if $\{\phi_i\}_{i=1}^k$ are characteristic functions and we have weights $p_j \geq 0$ such that $\sum p_j = 1$, then $\sum_{j=1}^k p_j\phi_j$ is also a characteristic function.

Apply this to the family of "tent" functions: for every $a_j > 0$ and $\{p_j\}$ with $\sum p_j = 1$, $\sum_{j=1}^k p_j \phi\left(\frac{t}{a_j}\right)$ is a characteristic function, where ϕ is the "tent" function. We use this fact to prove the following:

Corollary 20.3 (Pólya's Criteria) *If ϕ is a continuous function such that:*

1. *$\phi(t) \geq 0$ for all t,*
2. *$\phi(0) = 1$,*
3. *$\phi(t) = \phi(-t)$ (i.e. ϕ is even), and*
4. *ϕ is convex on $(0,\infty)$ and $(-\infty,0)$,*

then ϕ is a characteristic function.

Proof: If ϕ is given, we can make a finite approximation to it by picking points on $y = \phi(t)$ symmetric about the y-axis and drawing straight lines between them. This finite approximation is a characteristic function of the form $\sum_{j=1}^k p_j \phi\left(\frac{t}{a_j}\right)$. By taking increasingly fine approximations and using the continuity theorem, ϕ is also a characteristic function.

∎

We state the following fact without proof:

Proposition 20.4 *If ϕ is a characteristic function with compact support, then the periodic repetition of ϕ is the characteristic function of a lattice distribution.*

This fact allows us to prove the following theorem, which is useful in finding counterexamples:

Theorem 20.5 *Two characteristic functions can agree in a neighborhood of 0 without having the same measure.*

Proof: Just consider a characteristic function ϕ with compact support and its periodic repetition. ∎

Some notes:

- It is possible to have μ_1, μ_2, μ_3 probabilities on $\mathbb{R}$ such that $\mu_1 * \mu_2 = \mu_1 * \mu_3$ but $\mu_2 \neq \mu_3$ (i.e. cannot "cancel" μ_1 from both sides).

  (Take $\mu_1 =$ tent, $\mu_2 = \mu_1$, $\mu_3 =$ periodic continution of μ_1.)

- μ corresponding to the "tent" function has density $\propto \frac{\sin^2 x}{x^2}$, and so does not have a mean. We can make a characteristic function ϕ with compact support and as many moments as we want by convolving μ with itself (and renormalizing).

- However, we can't have a characteristic function with compact support and is analytic near 0.

20.1.3 Characteristic functions in $\mathbb{R}^d$

Characteristic functions also make sense in $\mathbb{R}^d$. If μ is a probability on $\mathbb{R}^d$, then for each vector $t \in \mathbb{R}^d$ we define

$$\phi_\mu(t) := \mathbb{E}_\mu[e^{it \cdot X}].$$

All theorems for characteristic functions in $\mathbb{R}^d$ are essentially the same as the $\mathbb{R}^1$ case.

Proposition 20.6 (Cramer-Wold Device) *If $X \in \mathbb{R}^d$ is a random vector and if we know the law of $\sum_{i=1}^{d} a_i X_i$ for all a_i, then we know the law of X.*

Proof: Since we know that law of $t \cdot X$ for all $t \in \mathbb{R}^d$, we know $\mathbb{E}[e^{it \cdot X}]$ for all t. ∎