Mathematical Statistics

1 Introduction to the materials to be covered in this course

1. Univariate & Multivariate r.v's

2. Borl-Cantelli Lemma

   Large Deviations. e.g. $X_1, \cdots, X_n$ are iid r.v's,

   $$P(\frac{X_1 + \cdots + X_n}{n} \in A) \leq e^{-nI(A)}$$

   where $I(A)$ is a number depending on A (event).

3. Extreme value theory

   X_i iid $N(0,1)$ with $W_n = \max_{1 \leq i \leq n} \mid X_i \mid$ We will prove

   - $\frac{W_n}{\sqrt{logn}} \to C$ a.s.
   - $P(W_n \leq C_n + x) \to F(x)\ \forall$ x

4. Multivariate Normal Distribution

   $$\begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} \sim N(\mu, \Sigma)$$

   $$Ee^{it^T X} = exp[it^T\mu - \frac{1}{2}t^T\Sigma t]$$

5. Exponential Family

   $$p_\theta(x) = exp[\sum_{i=1}^{s} \eta_i(\theta)T_i(x) - B(\theta)]h(x)$$

6. Sufficiency & Factorization Thm

   $p_\theta, \theta \in \Omega$, $T(x)$ is a statistic. If $P(X \in A \mid T = t)$ doe not depend on θ, then we say T is a sufficient statistic. If $p_\theta(x) = g_\theta(T(x))h(x)$, then T is sufficient.

7. Rao-Blackwell Theorem

   Unbiasedness, uniqueness, Basu's thoerem, completeness and MVUE (i.e. minimum variance unbiased estimators).

8. If time permits, we will study some of the following:

   - Weak convergence of probability measures. i.e. μ_n with $n = 1, 2, 3, \cdots$ are measures, $\mu_n \to \mu$
   - Empirical Processes

     Let $X_i; i \geq 1$ be iid random variables with distribution $N(0,1)$. Then $\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \to F(x)$ where $\delta_{x_i} = \{ \begin{array}{cc} 1 & x \in A \\ 0 & o.w. \end{array}$ So

     $$\sup_{A \in \mathcal{A}} \mid \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}(A) - \mu(A) \mid \to 0$$

     If $\mathcal{A}$ is VC class, then garanteed to go to 0.
   - Random Matrices $\begin{bmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & & \vdots \\ X_{n1} & \cdots & X_{nn} \end{bmatrix}$

     If $X_{ij}'s$ are random, what's the distribution of the eigenvalues? How about as $n \to \infty$?

2 Probability

2.1 Probability measure and probability spaces

9. Ω is a set.

   $\mathcal{F}$ is a set of subsets of Ω.

10. If

    (i) $\Omega \in \mathcal{F}$

    (ii) $A^c \in \mathcal{F}$ if $A \in \mathcal{F}$

    (iii) If $A_1 \in \mathcal{F}, A_2 \in \mathcal{F}, \cdots$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

    Then $\mathcal{F}$ is called a σ-algebra or σ-field.

11. If $\mathcal{F}$ is a σ-algebra, then

    (i) $\phi \in \mathcal{F}$

    (ii) If $A_1, A_2, \cdots$ are in $\mathcal{F}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$

(iii) If $B_1, B_2, \cdots, B_n$ are in $\mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$, $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

12. Eg. $\Omega = 1, 2, \cdots$. $\mathcal{F}$ =all subsets of Ω. Then $\mathcal{F}$ is σ-algebra. It is generated by the set $\mathcal{F}'$=finite subset of Ω. i.e. $\mathcal{F}$ is the smallest σ-algebra containing $\mathcal{F}'$.

13. Eg. $\Omega = \Re^1$. $\mathcal{F}' = (a, b), -\infty < a < b < \infty$. The smallest σ-algebra containing $\mathcal{F}'$, denoted by $\mathcal{B}(\Re^1)$, is called the Borel σ-algebra of $\Re^1$.

14. For any $F \in \mathcal{F}$, assign a number to F, call it $P(F)$ and satisfies the following properties

    (i) $P(\Omega = 1)$

    (ii) $P(A^c) = 1 - P(A)$ for any $A \in \mathcal{F}$

    (iii) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ provided $A_i's$ are disjoint, or mutually exclusive, i.e. $A_i \bigcap A_j = \phi$ for any $i \neq j$.

    If (i),(ii) and (iii) are true, then $P : \mathcal{F} \to [0, 1]$ is called **probability**. Right now, $(\Omega, \mathcal{F}, P)$ is a triplet which is called a **probability space**.

15. Eg. $\Omega = 1, 2, \cdots$, $\mathcal{F}$ ={all subsets of Ω}. Define $P(F) = \sum_{i \in F} \frac{1}{2^i}$ for $\forall F \in \mathcal{F}$. Then P is a probability.

16. Eg. $\Omega = [0, 1]$, $\mathcal{B} = \sigma-$algebra generated by $\{9a, b], [a, b), 0 < a < b \leq 1\}$. One can verify that $\mathcal{B} = [0, 1] \bigcap \mathcal{B}(\Re^1)$. $P(B) =$ the Lebesque measure of B for any $B \in \mathcal{B}$.(The r.v. generates this P is uniform.), defines a probability over $([0, 1], \mathcal{B})$

17. If $X : \Omega \to \Re$, i.e. $X(\omega)$ is a real number for any $\omega \in \Omega$, satisifies that

    $$X^{-1}(B) = \{w \in \Omega, X(\omega) \in B\} \in \mathcal{F} \text{ for any } B \in \mathcal{B}(\Re^1)$$

    then we say X is a **random variable**.

18. Eg. For any $F \in \mathcal{F}$, define $1_F(\omega) = \{ \begin{array}{cc} 1 & \text{if } \omega \in F \\ 0 & o.w. \end{array}$. Then 1_F, the indicator function of F is a random variable.

19. $(\Omega, \mathcal{F}, P)$ and X is a r.v. Define $\mu(B) = P(X \in B) = P(\omega \in \Omega; X(\omega) \in B)$ for $B \in \mathcal{B}(Re^1)$. Note that μ is a probability on $(\Re^1, \mathcal{B}(\Re^1))$, call μ the **distribution** of X.

20. Verify: X is a r.v. iff $\{X \leq x\} \in \mathcal{F}$ for any $x \in \Re^1$. (HW)

21. Def. If r.v. X & Y generate the same distribution, then we say X & Y are identically distributed, denoted by $X \stackrel{d}{=} Y$, or $\mathcal{L}(\mathcal{X}) = \mathcal{L}(\mathcal{Y})$ (law).

22. Eg. $X \sim U[0,1]$, then $X \stackrel{d}{=} 1 - X$. Actually, $P(1 - X \in B) = P(X \in 1 - B)$ =Leb measure of (1-B)=Leb measure of B.

Problem 1. Let X be a random variable and p is a positive number such that $P(|X| \geq x) \leq 1/x^p$ for any $x > 0$. Show that $E(|X|^\alpha) \leq p/(p - \alpha)$ for any $0 < \alpha < p$.

Sep. 17, Wednesday, 2003

1. Let X be a random variable, then

$$\sum_{n=1}^{\infty} P(|X| \geq n^{1/r}) \leq E|X|^r \leq \sum_{n=0}^{\infty} P(|X| > n^{1/r}), \text{where}\, r > 0.$$

2. In general, if $g(X), X \geq 0$ is a strictly increasing continuous function, let $a_n = g(n)$,$n = 0, 1, 2, \ldots$, then

$$\sum_{n=1}^{\infty} P(X \geq a_n) \leq Eg^{-1}(X) \leq \sum_{n=0}^{\infty} P(X > a_n).$$

Notice that the first statement is just a general case of the second case, so we are going to prove the second one.

Proof: Let $\Phi(X) = g^{-1}(X)$ and

$$Y = \sum_{j=1}^{\infty} j 1_{j \leq \Phi(X) < j+1}, Z = \sum_{j=0}^{\infty} (j+1) 1_{(j < \Phi(X) \leq j+1)}.$$

$$\begin{aligned}
Y &\leq \sum_{j=1}^{\infty} \Phi(X) 1_{(j \leq \Phi(X) < j+1)} \\
&= \Phi(X) \sum_{j=1}^{\infty} 1_{(j \leq \Phi(X) < j+1)} \\
&= \Phi(X) 1_{(1 \leq \Phi(X))} \\
&\leq \Phi(X) \\
Z &\geq \Phi(X) \sum_{j=0}^{\infty} 1_{(j < \Phi(X) \leq j+1)} \\
&= \Phi(X) 1_{\Phi(X) > 0} \\
&= \Phi(X)
\end{aligned}$$

First,

$$\begin{aligned}
\sum_{n=1}^{\infty} P(X \geq a_n) &= \sum_{n=1}^{\infty} P(\Phi(X) \geq n) \\
&= \sum_{n=1}^{\infty} \sum_{j=n}^{\infty} P(j \leq \Phi(X) < j+1) \\
&= \sum_{j=1}^{\infty} \sum_{n=1}^{j} P(j \leq \Phi(X) < j+1) \\
&= \sum_{j=1}^{\infty} jP(j \leq \Phi(X) < j+1) \\
&= EY
\end{aligned}$$

Similarly, $\sum_{n=0}^{\infty} P(X > a_n) = EZ$. **3.** $E|X|^p = p\int_0^{\infty} P(|X| > t)t^{p-1}dt = p\int_0^{\infty} P(|X| \leq$

$t)t^{p-1}dt, for p > 0.$

Proof: Notice that

$$\begin{aligned}
|X|^p &= p\int_o^{|X|} t^{p-1}dt \\
&= p\int_0^{\infty} t^{p-1}1_{(t<|X|)}dt \\
&= p\int_0^{\infty} t^{p-1}1_{(t\leq|X|)}dt
\end{aligned}$$

Take expectation on both sides, we have:

$$\begin{aligned}
E|X|^p &= pE(\int_o^{|X|} t^{p-1}dt) \\
\text{using Fubini's Theorem,} \quad &= p\int_0^{\infty} P(|X| > t)t^{p-1}dt
\end{aligned}$$

4. Chernoff's bound (simple example of large deviations.) Let $X_1, X_2, \ldots, X_n$ be iid r.v.

with mean μ. $S_n = \sum_{i=1}^{n}, a > \mu$. Then we like to study $P(\frac{S_n}{n} > a)$.

$$\begin{aligned}
P(\frac{S_n}{n} \geq a) &\leq P(tS_n \geq tan), \text{ for all} t > 0 \\
&= P(e^{tS_n} \geq e^{tan}) \\
&\leq e^{-tan} E e^{tS_n}, \qquad \text{by Markov's Inequality} \\
&= e^{-tan} (E e^{tx_1})^n \\
&= e^{-n(at - log E e^{tx_1})}
\end{aligned}$$

Thus,

$$\begin{aligned}
P(\frac{S_n}{n} \geq a) &\leq Inf_{(t>0)} e^{-n(at - log E e^{tx_1})} \\
&= e^{-nSup_{(t>0)}(at - log M_X(t))} \qquad \text{let } M_X(t) = Ee^{tX_1},
\end{aligned}$$

Thus, $P(\frac{S_n}{n} \geq a) \leq e^{-nI(a)}$, where $I(a) = Sup_{(t>0)}(at - log M_X(t))$.

Compare to Chevychev's inequality:

$$\begin{aligned}
P(\frac{S_n}{n} > a) &\leq P(|\frac{S_n}{n} - \mu| \geq a - \mu) \\
&\leq \frac{Var(X_1)}{n(a-\mu)^2}
\end{aligned}$$

09/22

Chernoff's bound

Let $X, X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim}$ r.v's with $M(t) = Ee^{tX} < \infty, \forall\ |t| < t_0, t_0 \in R^+$, $S_n = \sum_1^n X_i$. Then

$$P(\frac{S_n}{n} \geq a) \leq e^{-nA} \quad a > \mu = EX$$

$$P(\frac{S_n}{n} \leq b) \leq e^{-nB} \quad b < \mu = EX$$

Where

$$A = I(a) = \sup_{t \geq 0}\{ta - \log M(t)\}$$

$$B = I(b) = \sup_{t \leq 0}\{tb - \log M(t)\}$$

Define: $I(x) = \sup_{t \in R}\{tx - \log M(t)\}$, then

(1)$I(x) \nearrow$ on $[\mu,\ +\infty]$ and $\searrow$ on$(-\infty,\ \mu)$

(2)$I(\cdot)$ is convex

(3)$I(x) = \sup_{t \geq 0}\{tx - \log M(t)\}$ if $x \geq \mu$

$I(x) = \sup_{t \leq 0}\{tx - \log M(t)\}$ if $x \leq \mu$

(4)$I(x) = 0 \Leftrightarrow x = \mu$

Proof:

(4)Assume $I(x_0) = 0$,

$$tx \le \log M(t)\ \forall t > 0$$

Let $t \nearrow 0,\ x \le \lim_{t\nearrow 0} \frac{\log Ee^{tX}}{t} \overset{Lopitale}{=} \lim_{t\nearrow 0} \frac{Ee^{tX}X}{Ee^{tx}}$ Similarly, $tx \le \log M(t)\ \forall t < 0$, so $x \ge \frac{\log M(t)}{t}$

$\to \mu$ as $t \nearrow 0$, and $x \ge \mu$

So we get $x = \mu$.

(3)

$$M(t) = Ee^{tX} \ge e^{t\mu}$$

$$tx - \log M(t) \le t(x - \mu)$$

$$I(x) = \sup_t \{tx - \log M(t)\}$$

$$\text{If } x > \mu \text{ and } t < 0,\ \text{then } tx - \log M(t) < 0$$

But $I(x) \ge 0$, so $I(x) = \sup_{t\ge 0}\{tx - \log M(t)\}$

Chernoff's bound: General case

$$P(\frac{S_n}{n} \in A) \le 2e^{-nI(A)}\ \forall \text{ closed set } A,\ \text{where } I(A) = \inf_{x\in A} I(x)$$

Proof:

$\exists\ a, b\ st.\ b \le \mu \le a,\ a \in A\ , b \in A$

and $A \subset (-\infty,\ b] \cup [a,\ +\infty)$

$P(\frac{S_n}{n} \in A) \le P(\frac{S_n}{n} \ge a) + P(\frac{S_n}{n} \le b) \le e^{-nI(a)} + e^{-nI(b)} \le 2exp\{-n \min(I(a), I(b))\}$

$\sup_{allsuch(a,b)pairs} \min\{I(a), I(b)\} = \inf_{x\in A} I(x)$

Comment:

(1)Lower bound:

$$\liminf_{n\to\infty} \frac{1}{n} \log P(\frac{S_n}{n} \in A) \ge -\inf_{x\in A} I(x)\quad \forall A \text{ open}$$

(2)

$$P(\frac{S_n}{n} \ge a) \sim \frac{C}{\sqrt{n}} e^{-nI(a)} \text{ as } n \to \infty \text{ where } a \ge \mu \text{ for some } C$$

What's large deviation?

$\mu_n\ n \ge 1$ are probability measures, if $\exists$ a function $I(x)$ satisfies

(1)$I(x) \ge 0$

(2)$\{x : I(x) \le l\}$ is b.d.d. and closed set $\forall l$

(3) $\limsup_n \frac{1}{n}\log\mu_n(A) \le -\inf_{x\in A} I(x)\ \forall A$ closed

$\liminf_n \frac{1}{n}\log\mu_n(B) \ge -\inf_{x\in B} I(x)\ \forall B$ open

Then we say $\{\mu_n\}$ satisfies Large Deviation Principle (LDP) with rate function $I(x)$.

Note 10 of Stat8111, 9/24/Fall 2003

Characteristic Function:

X is a random variable, then the characteristic function of X is

$$\varphi_X(t) = Ee^{itx} = E(\cos(tx) + i\sin(tx))$$

where t is a real number, $i = \sqrt{-1}$, $e^{is} = \cos(s) + i\sin(s)$.

Example 1. $X \sim \text{Ber}(p)$, $P(X=1) = p, P(X=0) = q = 1-p$.

$$\varphi_X(t) = Ee^{itx} = q + pe^{it}$$

Example 2. $X \sim \text{Bin}(n,p)$. There are two methods to calculate the characteristic function.

Method 1: $P(X=k) = \binom{n}{k}p^k q^{n-k}$, so

$$\begin{aligned}\varphi_X(t) &= Ee^{itx} \\ &= \sum_{k=0}^{n} e^{itk}\binom{n}{k}p^k q^{n-k} \\ &= \sum_{k=0}^{n}\binom{n}{k}(pe^{it})^k q^{n-k} \\ &= (q+pe^{it})^n\end{aligned}$$

Method 2: $X = X_1 + \cdots + X_n$, X_i's are i.i.d Ber(p), then

$$\begin{aligned}\varphi_X(t) &= Ee^{itx} \\ &= E(e^{itx_1}\cdot e^{itx_2}\cdots e^{itx_n}) \\ &= Ee^{itx_1}\cdot Ee^{itx_2}\cdots Ee^{itx_n} \\ &= (Ee^{itx_1})^n \\ &= (q+pe^{it})^n\end{aligned}$$

Example 3. $X \sim N(\mu, \sigma^2)$, then

$$\varphi_X(t) = e^{(it\mu - \frac{\sigma^2}{2}t^2)}, t \in \mathbb{R}$$

Example 4. $X \sim Unif[-1,1]$, then

$$\begin{aligned}
\varphi_X(t) &= \frac{1}{2}\int_{-1}^{1} e^{itx}dx \\
&= \frac{1}{2it}(e^{it} - e^{-it}) \\
&= \frac{1}{2it}(\cos t + i\sin t - (\cos t - i\sin t)) \\
&= \frac{\sin t}{t}
\end{aligned}$$

Levy's Inversion Formular

Theorem: X is a random variable with characteristic function $\varphi(t)$. Then, for any $a < b$,

$$\lim_{c\to+\infty}\left[\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}\varphi(t)dt\right] = P(a < X < b) + \frac{P(X=a)+P(X=b)}{2}$$

Proof: Let

$$\begin{aligned}
I(c) &= \frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}\varphi(t)dt \\
&= \frac{1}{2\pi}E\left(\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}e^{itx}dt\right) \\
&= \frac{1}{2\pi}E\left(\int_{-c}^{c}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}dt\right)
\end{aligned}$$

First,

$$\begin{aligned}
\int_{-c}^{c}\frac{e^{it(x-a)}}{it}dt &= \frac{1}{i}\int_{-c}^{c}\frac{1}{t}[\cos(t(x-a)) + i\sin(t(x-a))]dt \\
&= \frac{1}{i}\int_{-c}^{c} i\cdot\frac{\sin(t(x-a))}{t}dt \\
&= 2\int_{0}^{c}\frac{\sin(t(x-a))}{t}dt \\
&= 2\int_{0}^{c}\frac{\sin(t(x-a))}{t(x-a)}dt(x-a) \\
&= 2\int_{0}^{c(x-a)}\frac{\sin t}{t}dt
\end{aligned}$$

Similarly,

$$\int_{-c}^{c}\frac{e^{it(x-b)}}{it}dt = 2\int_{0}^{c(x-b)}\frac{\sin t}{t}dt$$

Then,

$$I(c) = \frac{1}{\pi}E\left[\int_{0}^{c(x-a)}\frac{\sin t}{t}dt - \int_{0}^{c(x-b)}\frac{\sin t}{t}dt\right]$$

Let $J_c(x) = \int_{c(x-a)}^{c(x-b)} \frac{\sin t}{t} dt$
(i) Suppose $x > a$ and $x < b$, then

$$J_c(x) \to \int_{-\infty}^{\infty} \frac{\sin x}{x} dx = \pi$$

(ii) Suppose $x > b$, then

$$J_c(x) \to 0 \text{ as } c \to +\infty$$

(iii) Suppose $x < a$, then

$$J_c(x) \to 0 \text{ as } c \to +\infty$$

(iv) Suppose $x = a$, then

$$J_c(x) \to \int_{-\infty}^{0} \frac{\sin x}{x} dx = \frac{\pi}{2}$$

(v) Suppose $x = b$, then

$$J_c(x) \to \int_{0}^{\infty} \frac{\sin x}{x} dx = \frac{\pi}{2}$$

Let's summarize,

$$\lim_{c\to\infty} J_c(x) = \pi 1(a < x < b) + \frac{\pi}{2}(1(x = a) + 1(x = b))$$

Since $\int_{-\infty}^{\infty} \frac{\sin x}{x} dx = \pi$, we have $\sup_c J_c(x) < \infty$. By dominant convergence theorem,

$$\lim_{c\to\infty} I(c) = \frac{1}{\pi} \lim_{c\to\infty} EJ_c(x) = P(a < x < b) + \frac{P(x = a) + P(x = b)}{2}$$

Theorem: If $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$, then the distribution function of X has a bounded probability density function given by

$$f(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \varphi(t) dt$$

09-26-03 Friday

1. **Levy's Inversion Formula**

   Suppose X has characteristic function $\phi(t)$, then

   $$\lim_{c\to\infty} [\frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \cdot \phi(t) dt] = P(a < X < b) + \frac{P(X = a) + P(X = b)}{2}$$

   Given $\phi(t)$, how do we get distribution of X? Note: we can not recover the random variable X because different random variables could have the same distribution. (e.g. $X \sim U[0,1]$, then $-X \sim U[0,1]$).

2. **Discussion**

   Given $\phi(t)$, we recover $F(x) = P(X < x)$. From the inversion formula, I know that $P(a < X < b)$ for a and b such that $P(X = a) = 0, P(X = b) = 0$.

   Note:

   $$D = a : P(X = a) \neq 0$$

   is countable, i.e. discontinous points are countable. So I know the value of $F(b) - F(a)$ for a, $b \in R - D$.

   - Let $a_n \in R - D$, and $a_n \to -\infty$, then I get $F(b)$ for any $b \in R - D$.
   - Now for any $x \in D$, choose $b_n \to x$, define

   $$F(x) = \lim_{b_n \to x} F(b_n)$$

   So I obtain $F(x)$ for any $x \in R$

3. **Proposition**

   If cdf $F(x)$ and $G(x)$ are identical on H, where the closure of H is $\Re$, then $F(x) \equiv G(x), \forall x \in \Re$.

   Proof: $\forall x_0 \in H$, choose $x_n \in H$ and $x_n \to x_0$. Since $F(x_n) = G(x_n), n \geq 1$. Now let $n \to \infty$, we then have $F(x_0) = G(x_0)$, because $F(x)$ and $G(x)$ are right-continuous.

4. **Theroem**

   If $\int_{-\infty}^{\infty} dt < \infty$, then X doesn't have any discontinuous points and the density of X is given by

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

   Also, f(x) is bounded.

   Note: density of $f(x)$ can be unbounded. For example,

   $$f(x) = \{ \begin{array}{cc} 0 & x < 0 \\ \frac{1}{2\sqrt{x}} & 0 < x < 1 \end{array}$$

5. **Proof:**

   Note:

   $$\mid \frac{e^{-ita} - e^{-itb}}{it} \mid = \mid \int_a^b e^{-itx} dx \mid \leq \int_a^b \mid e^{-itx} \mid dx = b - a$$

Claim: No discontinuous points.

In fact, by the inversion formula,

$$\begin{aligned}\frac{P(X=a)+P(X=b)}{2} &\leq \frac{P(X=a)+P(X=b)}{2}+P(a<X<b)\\ &= \lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}\phi(t)dt\\ &\leq \frac{1}{2\pi}(b-a)\lim_{c\to\infty}\int_{-c}^{c}\mid\phi(t)\mid dt\\ &\leq (b-a)\int_{-\infty}^{\infty}\mid\phi(t)\mid dt\end{aligned}$$

If $P(X=a)>0$, then pick $b_n=a+\epsilon_n, \epsilon_n>0$ and $\epsilon_n\to 0$, and b_n is a continuous point of $F(x)$. Let $n\to\infty$, then

$$P(X=a)\leq 0 \Rightarrow P(X=a)=0$$

6. So far the inversion formula becomes

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-ita}-e^{-itb}}{it}\phi(t)dt=F(b)-F(a)$$

Now,

$$\begin{aligned}f(a) &= F'(a)\\ &= \lim_{b\to a}\frac{F(b)-F(a)}{b-a}\\ &= \frac{1}{2\pi}\lim_{b\to a}\int_{-\infty}^{\infty}\frac{1}{it}\frac{e^{-ita}-e^{-itb}}{b-a}\phi(t)dt\end{aligned}$$

Since $\frac{1}{it}\frac{e^{-ita}-e^{-itb}}{b-a}$ is bounded and $\int\mid\phi(t)\mid dt<\infty$, so by Dominated Convergence Therorem,

$$\begin{aligned}F'(a) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{it}(-e^{-ita})'\phi(t)dt\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-ita}\phi(t)dt\end{aligned}$$