MATH 8651 - Key Content

Note Set 1:

- Definition of probability space, probability measure, sigma-field
- Definition of measurable function, random variable from $(\Omega, \mathcal{F}, P)$ to $(\Psi, \mathcal{G})$.
- Proposition 2.1: only need preimage of sets that generate a sigma-field to be measurable for a function to be measurable.
- Corollary 1: A continuous function between topological spaces is measurable.
- A random variable $X : (\Omega, \mathcal{F}, P) \to (\Psi, \mathcal{G})$ induces a probability measure Q on $(\Psi, \mathcal{G})$ by setting $Q(B) = P(X^{-1}(B))$ for all $B \in \mathcal{G}$. We call Q the distribution of X.
- Equal almost surely.
- Proposition 2.2: The limit of an increasing sequence of measurable $\overline{\mathbb{R}}$-valued functions defined on a common measurable space has a measurable limit.
- Corollary 1: For a sequence (does not need to be increasing) of measurable $\overline{\mathbb{R}}$-valued functions defined on a common measurable space, the following are measurable: $\sup_n X_n, \inf_n X_n, \limsup_{n\to\infty} X_n, \liminf_{n\to\infty} X_n$. Furthermore, if $\lim_{n\to\infty} X_n$ exists, it is measurable.
- Definition of a simple function.
- Corollary 2: Characterization of $\overline{\mathbb{R}}^+$-valued measurable functions as limits of simple functions.

Note set 2:

- Definition of a distribution function (not to be confused with a distribution).
- Proposition 3.1: Continuity of measure for countable nested sequence of measurable sets.
- Proposition 3.2: Let Q be a p.m. on $(\mathbb{R}, \mathcal{B})$. Then, $F(x) := Q((-\infty, x])$ is a distribution function.
- Proposition 3.3: Let F be a distribution function. There exists a unique p.m. Q on $(\mathbb{R}, \mathcal{B})$ such that $Q((-\infty, x]) = F(x)$ for all x. So together with Prop 3.2 this fully characterizes the probability measure by the distribution function.
- Examples of common distribution functions given.
- Definition of a density.
- Define expectation of a random variable (for simple functions). Linearity of expectation; expectation of positive RV is positive.
- Corollary 1: For simple functions (and later in general) $X \leq Y \implies E(X) \leq E(Y)$.
- Let X be a $\overline{\mathbb{R}}^+$-valued RV on p.s. $(\Omega, \mathcal{F}, P)$. Set $E(X) := \sup_Z E(Z)$ for all simple nonnegative RVs Z with $Z \leq X$. Extend to X with values in $\overline{\mathbb{R}}$ with $E(X) := E(X^+) - E(X^-)$; does not exist if both terms infinite.
- **Theorem 4.1 (Monotone Convergence)**: Let $X_1, X_2, ...$ be an *increasing* sequence of $\overline{\mathbb{R}}^+$ valued RVs on a common p.s. $(\Omega, \mathcal{F}, P)$, and set $X = \lim_{n\to\infty} X_n$. Then $E(X) = \lim_{n\to\infty} E(X_n)$.
- Corollary 1: Expectation of countable sum is the sum of expectations.
- Corollary 2: For $X_1, X_2, ...$ be an *increasing* sequence of $\overline{\mathbb{R}}$ (not necessarily positive), with same setup as MCT otherwise, if $E(X_1) > -\infty$ then $E(X_n) \to E(X)$ as $n \to \infty$.
- Proposition 4.1: For the reasonable setup, $E_P[\varphi \circ X] = E_Q[\phi]$.
- Corollary 1: $E[X] = \int_{-\infty}^{\infty} xF(dx)$, where $F(dx) = dF(x)$ is equal to $f(x)dx$ if the density exists. When $X \geq 0$, $E[X] = \int_0^{\infty} [1 - F(x)]dx$.
- Define variance, standard deviation, comp. form of variance.

- **Proposition 5.1 (Chebyshev Inequality):** Suppose X is an $\overline{\mathbb{R}}$-valued RV with finite mean, μ, and variance σ^2. Then, for all $z > 0$,

$$P(\{\omega : |X(\omega) - \mu| \geq z\}) \leq \sigma^2/z^2$$

.

- Markov inequality: Let Y be an $\mathbb{R}$-valued RV, and let f be an $\mathbb{R}^+$-valued function which is increasing on some interval $J \subset \mathbb{R}$ containing the support of Y. Then, for all $z \in J$ such that $f(z) > 0$,

$$P(\{\omega \ : \ Y(\omega) \geq z\}) \leq E(f \circ Y)/f(z).$$

- Proposition 5.2 (Cauchy-Schwarz Inequality): Assume that $E(X^2), E(Y^2) < \infty$. Then $E(XY)$ exists and

$$[E(XY)]^2 \leq E(X^2)E(Y^2),$$

  and equality holds iff one of the RVs X or Y a.s. equals a constant multiple of the other.
- Define independence for sets, pairwise independence (events pairwise independent $\not\Rightarrow$ independence), random variables.
- Definition and properties of covariance, correlation.
- **Theorem 5.1 (Weak Law of Large Numbers):** Let $X_1, X_2, \ldots$ be $\overline{\mathbb{R}}$ valued RVs, defined on a common probability space $(\Omega, \mathcal{F}, P)$. Assume they have the same distribution, with finite mean μ and variance σ^2. Also assume each pair is uncorrelated, or has negative correlation. Then, for each $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left\{\omega \ : \ \left|\frac{S_n(\omega)}{n} - \mu\right| > \varepsilon\right\}\right) = 0,$$

  where $S_n(\omega) = \sum_{k=1}^n X_k(\omega)$.

- Define convex function in two ways.
- Proposition 5.3 (Jensen Inequality): Let X be a RV with finite mean. Let φ be an $\mathbb{R}$-valued function which is convex on the interval J, which supports X. Then, $\varphi(E[X]) \leq E[\varphi \circ X]$.

Note set 3:

- Introduce notion of limit of a set (via indicator functions).
- Useful characterizations (these always exist) for subsets of Ω $A_1, A_2, \ldots$ we have
  - the points that are in infinately many A_n:

$$\limsup_{n\to\infty} = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m,$$

  - the points that will eventually be in all A_n:

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m.$$

- **Lemma 6.1 (Borel-Cantelli):** Let $A_1, A_2, \ldots$ be a sequence of events on a common p.s. $(\Omega, \mathcal{F}, P)$, and set $A = \limsup_{n\to\infty} A_n$.
  (a) If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A) = 0$ (this is generally the more useful one).
  (b) If $\sum_{n=1}^{\infty} P(A_n) = \infty$ and A_n are independent, then $P(A) = 1$.
- Define finite measure, infinite measure, σ-finite measure.
- Define Sierpinski class.

- Proposition 7.1 (Sierpinski class): With correct setup, the smallest Sierpinksi class of subsets of Ω that contains $\mathcal{E}$ equals $\sigma(\mathcal{E})$.
- Lemma 7.2: A Sierpinski class of subsets of Ω which is closed under finite intersections and contains Ω is a σ-field.
- **Theorem 7.1 (Uniqueness of Measure)**: Let P and Q be p.m. on the m.s $(\Omega, \sigma(\mathcal{E}))$, where $\mathcal{E}$ is a collection of sets closed under pairwise intersections. If $P(A) = Q(A)$ for all $A \in \mathcal{E}$ then $P = Q$.
  - Important application: Let F be a distribution function. Let $\mathcal{E}$ be the collection of all intervals $A_n = (a_n, b_n]$. Note that $\mathcal{E}$ is closed under pairwise intersections. If $Q((-\infty, x]) = F(x)$, then one must have $Q(A_n) = F(b_n) - F(a_n)$. It follows from Theorem 7.1 that there exists at most one p.m. Q on $(\mathbb{R}, \mathcal{B})$ with these values.
- **Theorem 7.2 (Existence of Measure):** Let $\mathcal{E}$ be a field (not σ-field) of subsets of a space Ω, and R a nonnegative finitely additive function defined on $\mathcal{E}$ such that $R(\Omega) = 1$, and $R(A_n) \to 0$ as $n \to \infty$ as $n \to \infty$ for each decreasing sequence $A_1, A_2, ...$ in $\mathcal{E}$ for which $\cap_n A_n = \emptyset$. Then there exists a p.m. P defined on $\sigma(\mathcal{E})$ with $P(A) = R(A)$ for every $A \in \mathcal{E}$.
- Lemma 7.3: Let R be a nonnegative *finitely* additive function on a field $\mathcal{E}$ of subsets of Ω with $R(\Omega) = 1$. Then R is countably additive on $\mathcal{E}$ if $R(A_n) \to 0$ for every decreasing sequence $(A_1, A_2, ...)$ in $\mathcal{E}$ for which $\cap_n A_n = \emptyset$.

- Proposition 7.2: Let $\mathcal{E}$ be a field of subsets of Ω and R a nonnegative *countably* additive function on $\mathcal{E}$ with $R(\Omega) = 1$. Then there exists a p.m. P defined on $\sigma(\mathcal{E})$ with $P(A) = R(A)$ for all $A \in \mathcal{E}$.
  - Importantly: Lemma 7.3 allows us to reduce Theorem 7.2 to Proposition 7.2.
- Definition of completeness (other given in HW), and the completion of a σ-field, and a probability.
- Example that $([0, 1], \mathcal{B}, \lambda)$ is not complete. Example that the completion of $\mathcal{B}$ here still does not contain all subsets of $(0, 1])$. So in a very standard setting there are ways to arrive at undefined probabilities!

Note set 4:

- Define integral of an extended real-valued measurable function f on general measure space $(\Omega, \mathcal{F}, \mu)$.
- **Theorem 8.1 (Monotone Convergence):** Let $0 \leq f_1 \leq f_2 \leq ...$ be $\overline{\mathbb{R}}^+$ valued measurable functions on a m.s. $(\Omega, \mathcal{F}, \mu)$ and set $f = \lim_{n\to\infty} f_n$. Then
$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$
- **Lemma 8 (Fatou's Lemma):** Let $f_1, f_2, ...$ be a sequence (doesn't need to be increasing) of $\overline{\mathbb{R}}^+$ valued measurble functions on m.s. $(\Omega, \mathcal{F}, \mu)$. Then,
$$\int \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int f_n d\mu.$$
- **Theorem 8.2 (Dominated Convergence):** Let $f_1, f_2, ...$ be a sequence (doesn't need to be increasing) of $\overline{\mathbb{R}}$ (no longer require positive values) valued measurable functions on m.s. $(\Omega, \mathcal{F}, \mu)$, and suppose $g \geq 0$ is a measurable function on the same m.s. with $|f_n| \leq g$ a.e. and $\int g d\mu < \infty$, and that $f = \lim_{n\to\infty} f_n$ exists a.e.. Then, $\int |f| d\mu < \infty$ and
$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

- Special case of DCT called **bounded convergence theorem** when $\mu(\Omega) < \infty$ and $g(\omega) \equiv M < \infty$ for some constant M.

- Define a function to be **uniformly integrable** if

$$\lim_{c \to \infty} \left(\sup_{t \in T} \int_{|f_t| \geq c} |f_t| d\mu \right) = 0.$$

- **Theorem 8.3 (Uniform Integrability Criterion)**: Let $f_1, f_2, ...$ be a sequence (note this can be extended to a sequence over an arbitrary set) of $\overline{\mathbb{R}}$ valued measurable functions on m.s. $(\Omega, \mathcal{F}, \mu)$, where $\mu(\Omega) < \infty$ (this wasn't required for the others). Assume $f = \lim_{n \to \infty} f_n$ exists a.e. and $\int |f_n| d\mu < \infty$ for all n. The following statements are equivalent:

  (i) $f_1, f_2, ...$ are uniformly integrable.
  (ii) $\int |f| d\mu < \infty$ and $\lim_{n \to \infty} \int |f_n - f| d\mu = 0$ (i.e. the sequence converges in the L_1 norm).
  (iii) $\lim_{n \to \infty} \int |f_n| d\mu = \int |f| d\mu < \infty$.
  (iv) Each of the above (i)-(iii) imply $\lim_{n \to \infty} \int f_n d\mu = \int f d\mu$.

- Define absolute continuity w.r.t some measure μ. For two measures on same p.s. $\nu << \mu$ if $\mu(A) = 0$ implies $\nu(A) = 0$ for $A \in \mathcal{F}$.

- Proposition 8.1 is converse of the following.

- **Theorem 8.2 (Radon-Nykodym):** Let μ and ν be σ-finite measures defined on a common m.s. and satisfying $\nu << \mu$. Then there exists an $\mathbb{R}^+$ measurable function f s.t.

$$\nu(A) = \int_A f d\mu, \text{ for all } A \in \mathcal{F}.$$

  We call f the R-N derivative, or density, of ν w.r.t. μ.

- Define independence for σ fields.

- Define σ-field generated by a RV X. Characterize independence of σ-fields via independence of RVs.

- Define measurable rectangles, product σ-field, product of measurable spaces.

- Proposition 9.1: For two σ-finite m.s., there exists a unique measure μ on the product of measurable spaces such that $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ for all $A_i \in \mathcal{F}_{\rangle}$, $i = 1, 2$. Also, μ is σ-finite, and is a p.m. if μ_1, μ_2 are p.m.s.

- Call μ above the product measure. Define product space.

- Extending some results to arbitrary product measures (i.e. infinite products).

- **Theorem 9.1 (Weaker version of Kolmogorov Extension Theorem):** Let $(\Omega_n, \mathcal{F}_n, \mu_n)$ for $n = 1, 2, ...$ be a sequence of p.s.. Let $\Omega = \otimes_{n=1}^{\infty} \Omega_n$ and $\mathcal{F} = \otimes_{n=1}^{\infty} \mathcal{F}_n$. Then there exists a unique p.m. P on $(\Omega, \mathcal{F})$ such that

$$P \left(\bigotimes_{n=1}^{\infty} A_n \right) = \prod_{n=1}^{\infty} P_n(A_n),$$

  for events $A_n \in \mathcal{F}_n$, $n = 1, 2, ...$

  - Note: Resulting Ω needs to be a p.s. if Ω is to be σ-finite.
  - Note: Can extend to arbitrary (not just countable) products.

- Proposition 9.2: If φ is a measurable function from a product measurable space to a measurable space, then the "slice" function is measurable when one variable is fixed.

- **Theorem 9.2 (Fubini):** Let $(\Psi, \mathcal{G}, \mu)$ and $(\Theta, \mathcal{H}, \nu)$ be two σ-finite measure spaces, and let φ be an $\overline{\mathbb{R}}$ valued measurable function defined on the product measure space $(\Psi, \mathcal{G}, \mu) \times (\Theta, \mathcal{H}, \nu)$. If $\int_{\Psi \times \Theta} \varphi(x, y) d(\mu \times \nu)$ exists, then $x \mapsto \int_{\Theta} \varphi(x, y) \nu(dy)$ is a μ-almost everywhere defined measurable function from $(\Psi, \mathcal{G}, \mu)$ to $\overline{\mathbb{R}}$, and

$$\int_{\Psi \times \Theta} \varphi(x, y) d(\mu \times \nu) = \int_{\Psi} \left(\int_{\Theta} \varphi(x, y) \nu(dy) \right) \mu(dx).$$

  - Note: If φ is $\overline{\mathbb{R}}^+$ valued, then the integral on the product space always exists.

- **Proposition 9.3:** Let X and Y be independent $\overline{\mathbb{R}}$ valued RVs with finite expectations. Then $E(XY) = E(X)E(Y)$.

- One can show usual product of densities characterization of independence for RVs.

Note set 5:

- Definition: Let $Y_1, Y_2, ...; Y$ be real valued RVs. Then $Y_n \to Y$ **in probability** as $n \to \infty$ if, for each $\varepsilon > 0$,

$$\lim_{n \to \infty} P(\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}) = 0 \text{ (i.p).}$$

  Equivalently, for all $\varepsilon > 0$, there exists n_0 s.t.

$$P(\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}) \leq \varepsilon \text{ for } n \geq n_0.$$

- Definition: With same conditions, $Y_n \to Y$ **almost surely** as $n \to \infty$ if,

$$P(\{\omega : \lim_{n \to \infty} Y_n(\omega) = Y(\omega)\}) = 1.$$

  Equivalently, for all $\varepsilon > 0$, there exists n_0 s.t.

$$P(\{\omega : |Y_n(\omega) - Y(\omega)| \geq \varepsilon \text{ for some } n \geq n_0\}) < \varepsilon.$$

- Proposition 12.1: If Y's above defined on same p.s. then almost sure convergence implies convergence in probability.

- Proposition 12.2: Same setup as above. If Y_n converges in probability to Y, then there exists a subsequence $n_1, n_2, ...$ so that $Y_{n_k} \to Y$ a.s.

- For measure spaces, use the term convergence in measure rather than i.p.

- Proposition 12.3 (DCT): Restated for general measure space. If $f_n \to f$ in measure, plus the usual conditions, then can push the limit through the integral.

- **Theorem 12.1 (Cantelli SLLN):** Let $X_1, X_2, ...$ be independent $\overline{\mathbb{R}}$ valued RVs on a p.s. $(\Omega, \mathcal{F}, P)$ satisfying $E(X_k) = \mu$, $E(X_k^4) \leq M < \infty$. Set $S_n = \sum_{k=1}^{n} X_k$. Then,

$$\frac{S_n}{n} \to \mu \text{ a.s.}$$

- Example of an application of Cantelli SLLN to prove Borel's theorem, i.e. almost all numbers on unit interval are simply normal.

- **Theorem 12.2 (Kolmogorov SLLN):** Let $X_1, X_2, ...$ be a sequence of IID (so adding identically distributed to requirements here) $\mathbb{R}$ valued RVs.

  (a) If $E(|X_1|) < \infty$, then $S_n/n \to E(X_1)$ a.s.

(b) If $E(|X_1|) = \infty$, then $\limsup_{n\to\infty} |S_n|/n = \infty$ a.s.

- Note: If $X_1, X_2, ...$ are independent (not necessarily identically distributed) RVs with mean μ and $\sum_{n=1}^{\infty} \mathrm{Var}(X_n/n) < \infty$, then $S_n/n \to \mu$ a.s. (did not prove this though.)

- **Theorem 12.3 ("Grand-daddy" WLLN):** Let $X_1, X_2, ...$ be a sequence of independent $\mathbb{R}$ valued RVs on a common p.s., with distribution functions F_n. Let $b_1, b_2, ...$ be a sequence of real numbers that are increasing to ∞. Suppose (i)

$$\lim_{n\to\infty} \sum_{k=1}^{n} \int_{|x|>b_n} dF_k(x) = 0,$$

and (ii)

$$\lim_{n\to\infty} \frac{1}{b_n^2} \sum_{k=1}^{n} \int_{|x|\le b_n} x^2 dF_k(x) = 0.$$

Set $a_n = \sum_{k=1}^{n} \int_{|x|\le b_n} x dF_k(x)$. Then,

$$\frac{1}{b_n}(S_n - a_n) \to 0 \text{ i.p..}$$

- The above conditions (i) and (ii) are also necessary in the following sense: Suppose there exists x_0 and a $\lambda > 0$ such that for all k,
$$P(X_k \ge x_0) \ge \lambda, \; P(X_k \le x_0) \ge \lambda.$$
Then $\frac{1}{b_n}(S_n - a_n) \to 0$ i.p. implies (i) and (ii) above.

- Example given with the above showing that WLLN can hold even if $E(|X_1|) = \infty$.

- Define empirical distribution function.

- **Theorem 12.4 (Glivenko-Cantelli):** Let $X_1, X_2, ...$ bet IID $\mathbb{R}$ valued RVs on a common p.s., with distribution function F. Then,

$$\lim_{k\to\infty} \sup_{x\in\mathbb{R}} |F_k(x,\omega) - F(x)| = 0 \text{ a.s. (in } \Omega).$$

- Intro to renewal theory, Proposition 12.4 gives strong law statement for number of renewals.

- Proposition 12.5: under same setup as above, $\lim_{n\to\infty} E[N(t)]/t = 1/m$.

- General useful result: If $X_t \to X$ a.s. and for all t, $E(X_t^2) \le M < \infty$, then $E(|X_t - X|) \to 0$ as $t \to \infty$.

- Lemma for expectation of square of renewals.

- Define convolution, Proposition 10.1 which gives convenient integral for convolutions.

Note set 6:

- Define the characteristic function of a real valued RV X to be $\varphi(t) = E[e^{itX}]$, $t \in \mathbb{R}$. Note: $E[e^{itX}] = E[\cos(tX)] + iE[\sin(tX)]$.

- Define related Laplace transform, MGF, and generating function.

- Some properties of CFs:

  1. $\varphi(0) = 1$.

2. $|\varphi(t)| \leq 1$.
3. $\overline{\varphi}(t) = \varphi(-t)$.
4. $\varphi(t)$ is uniformly continuous.
5. Shifts and scaling: $\varphi_{aX+b} = e^{itb}\varphi_X(at)$.
6. Convex combinations: If φ_i are CFs and $\sum a_i = 1,\ a_i \geq 0$, then so is $\sum a_i \varphi_i$.
7. Products: If φ_1, φ_2 are CFs, then so is $\varphi_1\varphi_2$. Moreover, if X_1, X_2 are independent ($\mathbb{R}$-valued) RVs with CFs φ_1, φ_2. Then $X_1 + X_2$ has CF $\varphi_1\varphi_2$.
8. Symmetrization: if φ is a CF, then so is $|\varphi|^2$. (Note: $|\varphi|$ is not a CF in general.)

- Uniqueness: If $\varphi_{X_1}(t) = \varphi_{X_2}(t)$ for all t, then X_1, X_2 have the same distribution.
- **Theorem 13.1 (Inversion Theorem):** If $x_1 < x_2$, then

$$\mu((x_1, x_2)) + \frac{1}{2}\mu(\{x_1\}) + \frac{1}{2}\mu(\{x_2\}) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-itx_1} - e^{-itx_2}}{it}\varphi(t)dt.$$

- Useful facts related to CFs in proof, e.g. $|e^{itX}| = 1$, and the integral trick.
- Corollary 1: If two distribution functions have the same characteristic function, they are equal.
- Corollary 2: A RV X has real-valued CF iff X is symmetric.
- Corollary 3: If $\int_{-\infty}^{\infty} |\varphi(t)|dt < \infty$, then F is continuously differentiable and

$$F'(x) = f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi(t)dt.$$

  - Other result 1: for all x,
  $$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e^{-itx}\varphi(t)dt = \mu(\{x\}).$$
  - Other result 2: for all x,
  $$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} |\varphi(t)|^2 dt = \sum_{x\in D} \mu(\{x\})^2,$$
  where D is the set of points $\{x :\ \mu \text{ is discontinuous at } x\}$, i.e. points that have mass.
  - Periodic version of 1:
  $$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itx}\varphi(t)dt = \mu(\{x\}).$$

- Define the rth absolute moment, $\mu_F^{(r)}$ and rth moment, $m_F^{(r)}$ of F.
- Proposition 13.2: If $\mu_F^{(n)} < \infty$ for a given $n \in \mathbb{Z}^+$, then $\varphi(t)$ has a continuous derivative of order n given by

$$\varphi^{(n)}(t) = \int_{-\infty}^{\infty} (ix)^n e^{itx} F(dx).$$

- Corollary 1: If F has finite absolute moment of order $n \in \mathbb{Z}^+$, then φ has the following expansion in a neighborhood of 0:

$$\varphi(t) = \sum_{k=0}^{n} \frac{i^k}{k!} m^{(k)} t^k + o(|t|^n) = \sum_{k=0}^{n-1} \frac{i^k}{k!} m^{(k)} t^k + \frac{\theta_n(t)}{n!}\mu^{(n)}|t|^n,$$

where $|\theta_n(t)| \leq 1$.

- Proposition 13.3: If φ has a finite derivative of *even* order n at $t = 0$, then F has a finite moment of order n.

- Note: with Proposition 13.2 this means that if a CF's *even* order derivative exists at 0, then it exists everywhere.

- Symmetric derivative formula:

$$\varphi^{(k+2)}(t) = \lim_{h\to 0} \frac{\varphi(t+h)^{(k)} - 2\varphi^{(k)}(t) + \varphi^{(k)}(t-h)}{h^2}.$$

- Note:

$$\sin(t) = \frac{e^{it} - e^{-it}}{2i} \text{ and } \cos(t) = \frac{e^{it} + e^{-it}}{2}.$$

- Definition of a complex valued function being positive definite.

- Theorem 13.2: The function φ is a characteristic function iff it is positive definite and is continuous at 0 with $\varphi(0) = 1$.

Note set 7:

- Define the set of continuity points for a distribution function $C_F = \{x : F(\cdot) \text{ is continuous at } x\}$.

- Proposition 14.1: If $X_n \to X$ i.p., then $\lim_{n\to\infty} F_n(x) = F(x)$ for all $x \in C_F$.
  - Note: the other direction is *very* false.

- Corollary: If $X_n \to X$ i.p., then $\{x : \lim_{n\to\infty} F_n(x) = F(x)\}$ is dense (in $\mathbb{R}$).

- Definition: A sequence of distribution functions F_n **converges weakly** to the distribution function F if

$$\lim_{n\to\infty} F_n(x) = F(x) \text{ for all } x \in C_F,$$

  or equivalently,

$$\lim_{n\to\infty} \mu_n((-\infty, x])) = \mu(-\infty, x])) \text{ for all } x \in C_F.$$

  - Note: by Proposition 14.1 this means converge i.p. implies F_n's converge weakly.
  - When $F_n \to F$ weakly, we say $X_n \to X$ **in distribution**.

- Proposition 14.2: Suppose that $X_n \to c$ in distribution as $n \to \infty$, where $c \in \mathbb{R}$. Then $X \to c$ i.p. (but not necessarily a.s.).

- Proposition 14.3: If $X_n \to X$ in distribution, then there exists a p.s. $(\Omega, \mathcal{F}, P)$ and RVs $X'_n, n = 1, 2, ...$ and X' defined on $(\Omega, \mathcal{F}, P)$ and having the same distribution functions as X_n and X such that $X'_n \to X'$ a.s. as $n \to \infty$.

- Proposition 14.4: The sequence of distribution functions F_n converge weakly to F iff for each bounded, continuous function g on $\mathbb{R}$,

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} g(x) F_n(dx) = \int_{-\infty}^{\infty} g(x) F(dx).$$

- Proposition 14.5 (Helly Selection Principle): Let $F_n(x)$, $n = 1, 2, ...$ be a sequence of distribution functions. Then there exists a subsequence $F_{n_k}(x)$ and an increasing right continuous function $F(x)$ with $\lim_{x\to-\infty} F(x) \geq 0$ and $\lim_{x\to\infty} F(x) = 1$, such that for all $x \in C_F$,

$$\lim_{k\to\infty} F_{n_k}(x) = F(x).$$

  - Note: This mean $F_n(x)$, $n = 1, 2, ...$ has a "limit", but this need not be a distribution function. Need the sequence to be tight in order to get a distribution function (see below).

- Basic fact about sequences in a compact space that is related here: Let $G_n(x)$ be functions with $G_n(x) \in [0,1]$ for all x. For each $x \in \mathbb{R}$, (pick x first, different than Prop 14.5) there exists a subsequence $\{m_k\}$ s.t. $\lim_{n\to\infty} G_{m_k}(x)$ exists. This suggests compactness is playing a key role in the HSP.

- Definition: A collection of measures $\{\mu_\alpha\}$ (or $F_\alpha(x)$) is **tight** if for every $\varepsilon > 0$, there exists a *finite* interval I s.t.
$$\inf_\alpha \mu_\alpha(I) > 1 - \varepsilon.$$

- Proposition 14.6: Every limit $F(x)$ in Proposition 14.5 is a distribution function iff $\{F_n(x)\}$ is tight.
  - Note: Taking Proposition 14.5 and 14.6 together, this means that if $\{F_n\}$ is tight, then there exists a subsequence $\{F_{n_k}\}$ such that $F_{n_k} \to F$ weakly, where F is a distribution function.

- Definition: The family of functions $f_\alpha : \mathbb{R} \to \mathbb{R}$ is **equicontinuous** if for every $\varepsilon > 0$, there exists a $\delta > 0$ s.t. for $|x_2 - x_1| < \delta$ and any α, then $|f_\alpha(x_2) - f_\alpha(x_1)| < \varepsilon$.

- **Theorem 14.1 (Levy-Cramer Convergence Theorem):**
  (i) Let F_n, $n = 1, 2, \ldots$ and F_∞ be distribution functions with characteristic functions φ_n, $n = 1, 2, \ldots$ and φ_∞. If F_n converges weakly to F, then $\varphi_n(t) \to \varphi_\infty(t)$ uniformly in every finite interval. Furthermore $\{\varphi_n\}$ is equicontinuous.
  (ii) (More useful) Let F_n, $n = 1, 2, \ldots$ be distribution functions with CFs φ_n, $n = 1, 2, \ldots$. Suppose that (a) $\lim_{n\to\infty} \varphi_n(t) = \varphi_\infty(t)$ for all t for some function φ_∞ (not necessarily a CF yet) and (b) φ_∞ is continuous at 0. Then, (a') F_n converges weakly to F_∞ for some distribution function F_∞, and (b') φ_∞ is the CF of F_∞.

- Lemma 14.1: For each $\delta > 0$, $\mu([-2/\delta, 2/\delta]) \geq \frac{1}{\delta}|\int_{-\delta}^{\delta} \varphi(t)dt| - 1$, where φ is the CF associated with μ.
  - This says that if φ is to close to 1 near $t = 0$, then μ has most of its mass not too far from 0.
  - Note on the LHS this is a *large* interval since δ is usually small in practice.

Note set 8:

- **Theorem 15.1 (Khintchine WLLN):** Let $X_1, X_2, \ldots$ be IID RVs with mean 0. Then $S_n/n \to 0$ i.p.

- Useful fact:
$$\lim_{n\to\infty} (1 + t/n + o(1/n))^n = e^t.$$

- **Theorem 15.2 (Polya CLT):** Let $X_1, X_2, \ldots$ be IID RVs with mean 0 and variance 1 (added assumption on variance here compared to the weak law above). Then
$$\frac{S_n}{\sqrt{n}} \to_D Z \text{ as } n \to \infty.$$
  - Note: this immediately generalizes the case where $X_1, X_2, \ldots$ are IID with mean μ and variance $\sigma^2 > 0$, then
$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \to_D Z \text{ as } n \to \infty.$$
  - Note: previous dominating techniques will not work to show this result because we ultimately want a distribution. So we need the CF machinery and to leverage the LCCT.

- **Theorem 15.3 (Liapounov CLT):** Let $X_1, X_2, \ldots$ be independent (no longer require IID) with $E(X_k) = 0$, $\text{Var}(X_k) = \sigma_k^2 < \infty$, and $E(|X_k|) = \gamma_k < \infty$. Set $\text{Var}(S_n) = \sum_{k=1}^n \sigma_k^2 := s_n^2$, and $\Gamma_n = \sum_{k=1}^n \gamma_k$. If $\lim_{n\to\infty} \Gamma_n / s_n^3 = 0$ (note the power of 3 here), then
$$\frac{S_n}{s_n} \to_D Z \text{ as } n \to \infty.$$

- Note the s_n in the denominator of final statement.

- Definition: A double array of RVs (idk if this is a general definition but we need below) can be written as $\{X_{n,j},\ n \geq 1,\ 1 \leq j \leq k_n\}$ and has the following properties:
  - RVs in the same row (same n) are independent.
  - Define $S_n = \sum_{j=1}^{k_n} X_{n,j}$ to be the sum (across columns) for a single row.
  - For all n, j we have $E(X_{n,j}) = 0$, and $s_n^2 = \text{Var}(S_n) = \sum_{j=1}^{k_n} \text{Var}(X_{n,j}) = 1$ (i.e. the sum of variances in a single row equals 1).

- **Theorem 15.4 (Lindeberg-Feller CLT):** Let $\{X_{n,j}\}$ be a double array as above. Assume that for all $\eta > 0$,
$$\lim_{n\to\infty} \sum_{j=1}^{k_n} \int_{|x|>\eta} x^2 F_{n,j}(dx) = 0, \quad (1)$$
(i.e. $\lim_{n\to\infty} \sum_{j=1}^{k_n} \int_{|x|\leq\eta} x^2 F_{n,j}(dx) = 1$). Then,
$$S_n \to_D Z \text{ as } n \to \infty.$$
Conversely, if $S_n \to_D Z$ and
$$\lim_{n\to\infty} \max_{1\leq j\leq k_n} P(|X_{n,j}| > \eta) = 0, \text{ for all } \eta > 0,$$
then (1) holds above.

- **Theorem 15.5 (Law of the Iterated Logarithm):** Let $X_k,\ k = 1, 2, ...$ be IID RVs with $E(X_1) = 0,\ \text{Var}(X_1) = 1$. Then,
$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = 1 \text{ a.s.},$$
and
$$\liminf_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = -1 \text{ a.s.}.$$

- **Theorem 15.6 (Alternative LIL):** Let $X_k,\ k = 1, 2, ...$ be independent (no londer require IID) RVs with $E(X_k) = 0,\ \text{Var}(X_k) = \sigma_k^2,\ \gamma_k = E(|X_k|^3) < \infty, s_n^2 = \sum_{k=1}^n \sigma_k^2, \Gamma_n = \sum_{k=1}^n \gamma_k$. Suppose that for some $\varepsilon > 0, A > 0$, we have $\Gamma_n / s_n^3 \leq A/(\log s_n)^{1+\varepsilon}$ for all n. Then,
$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2s_n^2 \log\log s_n}} = 1 \text{ a.s.},$$
and
$$\liminf_{n\to\infty} \frac{S_n}{\sqrt{2s_n^2 \log\log s_n}} = -1 \text{ a.s.}.$$

- **Theorem 13.3 (Another Rep of CFs):** Let $\varphi(t)$ be a continuous real valued function. If it satisfies the following five conditions then it is a characteristic function:
  1. $\varphi(0) = 1$.
  2. $\varphi(t) = \varphi(-t)$.
  3. $\varphi(t) \geq 0$.
  4. $\varphi(t)$ is nonincreasing on $[0, \infty)$.
  5. $\varphi(t)$ is convex on $[0, \infty)$.

- Note: For the above to hold the mean cannot exist, so this is a very restrictive set of functions.

- Note: The above shows that it is possible for two distinct distribution functions F_1 and F_2 to have CF which are identical on an interval.