Probability Theory Oral Exam Abridged Notes 2008

Chaitanya Ekanadham

1 Distributions, characteristic functions, weak convegence

1. **Characterizations of measures**

   (a) **Statement:** Each probability measure P on $\mathbb{R}$ is equivalently characterized its distribution function F (monotonic/R-cnts/onto/$[0,1]$-valued) or characteristic function ϕ (cnts and equal to 1 at $t = 0$/pos. def./$\mathbb{C}$-valued).

   (b) **Depends on:** The relationship is $F(x) = P((-\infty, x])$. Countable additivity of P follows by using right continuity to make a compact under-approximation and using the FIP. Char fns and dbns are related via the Fourier transform (and inverse transform).

   (c) **Used for:** Dbn fns and char fns are easier to deal with than probability measures (set fns), especially when proving weak convergence, convergence of sums, the list goes on.

2. **Weak convergence**

   (a) **Statement:** Weak convergence of measures is equivalent to pointwise convergence of the dbn fns (at points of continuity). This is equivalent to convergence in integrals of any bounded cnts fns (weak * convergence on the space of cnts fns). This is also equivalent to pointwise convergence of the char fns. Furthermore, any limit of char fns that is cnts at 0 is the char fn of some dbn.

   (b) **Depends on:** Cvgnce of integrals from pointwise cvgnce of dbn fns follows from Riemann integrals on a compact interval. We can show weak cvgnce from pointwise cgnce of char fns by constructing a distribution fn from its skeleton on the rationals ('subdistribution fn').

   (c) **Used for:** WLLN, CLT (any time we want to prove cvgnce in dbn for things expressed more readily by char fns e.g., sums of indep. r.v.'s). In particular, the definition using integrals against cnts functions shows the relationship between *weak convergence* in probability theory with *weak * convergence* in the space of continuous function on Ω (since the dual of $C(\Omega)$ is the set of nonneg. measures on Ω).

2 Sums of random variables and limit theorems

1. **Chebyshev's inequality**

   (a) **Statement:** An upper bound of on the probability that a random variable X exceeds M (in abs. val) is $\frac{\|X\|_p^p}{M^p}$

   (b) **Depends on:** property of measure and integral - just estimate.

   (c) **Used for:** easily showing a WLLN for i.i.d. r.v.'s *with finite variance*. Kolmogorov's inequality and Doob's inequality specialize this bound to sums of indep. r.v.'s and martingales, respectively.

2. **WLLN**

   (a) **Statement:** Given a sequence of i.i.d. r.v.'s with finite first moment, their sample mean converges in probability to the distribution mean.

   (b) **Depends on:** taylor-expansions of char fns and the fact that weak cvgnce to the point mass measure is equivalent to cvgnce in probability to that constant.

   (c) **Used for:** (obvious applications) In fact from the proof with char fns we see that we do not really need a finite first moment - just differentiability of the char fn at 0 - see examples. The first moment is *necessary* for the SLLN.

3. **Kolmogorov's consistency theorem**

   (a) **Statement:** Suppose we have an infinite product (measurable) space (where the space is separable metric space) with a sequence of probability measures defined on the finite-dimensional product spaces that are consistent with each other. Then there exists a c.a. prob. measure on the infinite product space.

   (b) **Depends on:** Extension theorem and defining a c.a. measure on the field generated by finite dimensional cylinder sets. We need tightness of measure (for compact under-approximations), diagonalization, and the FIP to show countable additivity.

(c) **Used for:** This is a very technical theorem but is crucial since otherwise any talk of limits of sums of random variables wont make any sense (since these require the existence of infinite product measures that agree with the natural intuition on finite-dimensional cylinders). In particular, Borel Cantelli, the 1/2/3-series, and SLLN would not make sense.

4. **Borel-Cantelli Lemma**

   (a) **Statement:** If we have an infinite sequence of events, whose probabilities sum up to a finite value, then the probability that $\omega \in A_n$ for infintely many n is 0. Conversely, if the events are independent and the probabilities sum up to ∞, then the probability that $\omega \in A_n$ for infinitely many n is 1.

   (b) **Depends on:** Interpretation of the statement in terms of $\limsup A_n$ and $\liminf A_n^c$. The identity $1 - x \leq e^{-x}$ is also useful for the second part.

   (c) **Used for:** Proving 'almost sure' thoerems. In particular, it is used in proving the 3-series theorem and SLLN in showing that sequence of random variables is equal (a.s.) to their truncated versions eventually (either sufficiently later in the sequence or for sufficiently large truncation values). We can also reinterpret Borel-Cantelli as saying that on the *infinite-product space with the infinite-product measure* (generated by copying the orignal space/measure), the probability that $\omega(n) = x_n \in A_n$ infinitely many times is 0 or 1 and so on. In other words, taking an ω in the single space and testing membership in each A_n is equivalent to taking an ω in the product space (an infinite sequence) and testing membership of the n'th member in the sequence in A_n for each n.

5. **Kolmogorov's inequality and Levy's inequality**

   (a) **Statement:** Consider a sequence of indep. random variables with 0 mean and finite variances σ_i^2. Let E_M be the event that the random walk defined by summing the variables exceeded a value M at any time j, $1 \leq j \leq n$. Then $P(E_M)$ is bounded above by $\{\sum_{j=1}^{n} \sigma_j^2\}/M^2$.

Suppose also that we know that the probability that any tail of the random walk (up to time n) exceeds $M/2$ is bounded above by δ. Then $P(E_M)$ is bounded above by $\frac{\delta}{1-\delta}$.

(b) **Depends on:** Partitioning E_M into n events (the random walk exceeds M for the first time at time j) and exploiting independence of $S_n - S_m$ and S_m for any $1 \leq m \leq n$. To show the first we use Chebyshev and complete a square to bound each part. To show the second we condition E_M on whether $S_n \leq M/2$ or not.

(c) **Used for:** Kolmogorov's inequality is like Chebyshev's inequality specialized to sums of indep. r.v.'s. It tells us that taking the sup over the random walk costs nothing more. It is used to prove the 1-series convergence theorem. Levy's inequality is similar but tells you more given stronger assumptions - it is used to prove that cvgnce in prob. of a sum of indep. r.v.'s implies cvgnce almost surely.

6. **Sufficient conditions for a.s. cvgnce**

   (a) **Statement:** If a sequence of functions (f_n) satisfies a stronger version of cauchy-in-measure (probability) convergence which involves the variable $\sup_{m \leq k \leq n} |f_k - f_m|$, then the sequences converges a.e. to a limit function.

   (b) **Depends on:** Similar to proving a subsequence cvgs a.e. if the sequence cvgs in measure. However, in this case we construct a special subsequence with the extra sup condition. The rest follows the same format.

   (c) **Used for:** This is the link we need to establish a.s. convergence for sums of indep. r.v.'s. The sup assumption of this result is guaranteed by using either Kolmogorov's inequality (for finite variance r.v.'s) or assuming cvgnce in probability and applying Levy's inequality.

7. **Levy's theorem for sums of r.v.'s**

   (a) **Statement:** For sums of independent r.v.'s weak convergence, convergence in probability, and a.s. convergence are all equivalent.

   (b) **Depends on:** Going from weak cvgnvce to cvgnce in prob. uses cvgnce of the char fn and expressing it as an infinite product to show that the sequence must be cauchy-in-prob. Going from cvgnce in prob. to a.s. cvgnce uses Levy's inequality.

(c) **Used for:** This tells us how special sums of indep. r.v.'s are in terms of their convergence.

8. **Kolmogorov 1-series, 2-series, 3-series**

   (a) **Statement:** If a sequence of indep. 0-mean r.v.'s have finite variances σ_i^2 that are summable, then the sum cvgs a.s. to some limit variable (1-series). This generalizes to r.v.'s with nonzero mean if the means are summable (2-series). This generalizes further to r.v.'s with possibly infinite means/variances if there is a cutoff C for which the sum over n that the $|X_n|$ exceeds C is summable.

   (b) **Depends on:** 1-series is basically Kolmogorov's inequality and the sup-cauchy condition. 2-series is just subtracting the mean and applying 1-series. 3-series is applying Borel-Cantelli Lemma to the sequence of events that $|X_n|$ exceeds C (so that a.s. convergence of the truncations is equivalent to a.s. convergence of the original series) and applying 2-series.

   (c) **Used for:** SLLN, useful in their own right since here the r.v.'s are not necessarily identically distributed.

9. **SLLN**

   (a) **Statement:** If a sequence of i.i.d. r.v.'s has finite first moment, then their sample mean converges a.s. to the 'true' mean (WLOG $= 0$ since we can subtract off the mean).

   (b) **Depends on:** The 3-series applied to special truncated versions of the sequences (with truncation relaxing as n increases). Borel Cantelli to show that convergence between these versions and a variant of the sample mean is equivalent. Kronecker's lemma to show that convergence of the variant of the sample mean implies convergence of the sample mean.

   (c) **Used for:** (obvious applications) Note that the finite first moment assumption is necessary. To see why, note that $\mathbb{E}[|X|]$ is finite iff $\sum P(|X| \geq n)$ is (rearrangement of sums). If that sum is infinite, then by Borel Cantelli part 2 we have that $|X_n| \geq n$ infinitely many times a.s. , which makes it impossible for S_n/n to converge a.s. (it is not cauchy).

10. **Kolmogorov 0/1 Law**

(a) **Statement:** Events in the tail field of an infinite product space (intersection of the sigma fields generated by tails) have probability 0 or 1 wrt the infinite product measure.

(b) **Depends on:** Show that the event is independent of all measurable events (including itself) using the monotone class theorem.

(c) **Used for:** Gives us alternative intuition of the SLLN (sum converges to point mass measure), Borel-Cantelli (prob is 0/1), which tell us things about events in the tail field. Now we can say things about the probability that the sample mean is asymptotically $O(n^\alpha)$ for any α.

11. **Central limit theorem**

(a) **Statement:** Suppose we have a sequence of i.i.d. r.v.'s with finite first and second moments. Then the distribution of $\frac{X_1+\ldots X_n}{\sqrt{n}}$ cvgs weakly to the normal distribution with mean and variance same as that of X_1.

(b) **Depends on:** Char fns, taylor expansions, and limit-proprety of e.

(c) **Used for:** (obvious applications)

12. **Lindeberg's theorem**

(a) **Statement:** Generalization of CLT to non-identically distributed indep. r.v.'s. If, when considering each S_n, the fraction of the total variance coming from values outside an ϵ standard deviation of 0 goes to 0 as $n \to \infty$ for arbitrarily small ϵ, then the CLT still holds.

(b) **Depends on:** The proof is long and painful but the general approach using triangular arrays is useful. See text or other notes.

(c) **Used for:** This tells us to what extent we can weaken the the indentically distributed assumption. The CLT can fail in 2 notable cases. If the variances are summable, it can cvg to a possibly non-Gaussian r.v. (1-series). Even if the variances are not summable, if the probability of each summand being nonzero is small enough to be summable but the total variance diverges to ∞, then the sum divided by $\sqrt{n}$ goes to 0 a.s (not normal).

13. **Law of iterated logarithm**

(a) **Statement:** The lim sup / lim inf of the sum of i.i.d mean-0 variance-1 r.v.'s divided by $\sqrt{n}$ is $\pm\infty$ (respectively) with probability 1. In particular, they grow asymptotically on the order of $\pm\sqrt{2\log\log n}$ with probability 1.

(b) **Depends on:** The Kolmogorov 0/1 law and CLT together immediately imply the first part (this is a tail event). The proof of the second part is complicated- it involves Borel Cantelli on the events where the maximum sum between two specified indices (in a special index sequence) exceeds a constant times $\sqrt{x\log\log x}$ evalauted at the lower index. Each summand is bounded using Levy's inequality and choosing the index seqience to be ρ^n for some $\rho > 1$.

(c) **Used for:** Not that many applications. But it is important to know that the sum divded by $\sqrt{n}$ will never converge a.s. to another r.v. We need to normalize by a power of n biiger then $\frac{1}{2}$.

3 Dependent random variables

1. **Existence of conditional expectation**

(a) **Statement:** Any random variable with finite first moment has a conditional expectation random variable (also integrable) wrt any sub-sigma-field (defined uniquely up to sets of 0-measure). In addition, conditional expectations satisfy, the law of iterated conditional expectation (tower property), conditional nonnegativity, linearity, and conditional Jensen's inequality.

(b) **Depends on:** Existence/uniqueness is basically taking a Radon-Nikodym derivative of the measure defined by integrating the random variable (*restricted* to the sub sigma-field) with respect to the original prob. measure (again, the restricted to the sub sigma-field). Jensen's can be proved using the property that convex functions are sup's of linear functions.

(c) **Used for:** Developing the whole theory of dependent random variables, Martingales and markov chains, etc. Conditional expectation also gives a notion of 'information' of 1 random variable coming from another random variable.

4 Markov chains

1. **MC's and the Chapman Kolmogorov equations**

   (a) **Statement:** MC's are essentially special classes of distributions on an infinite product measure space, in which the cylinder set probabilities depend only on the penultimate fixed point in the sequence (which may in general depend on the size of the cylinder's base) (1-step transition probs). These measures induce l-step transition probabilities by just repeatedly integrating the 1-steps. Any n-step transition probability from time k can be arrived at by 'integrating out' any intermediate state (just by conditioning). In addition, these agree with the conditional probabilities coming from the infinite product measure defining the MC (a.e.).

   (b) **Depends on:** formal manipulation and definition of conditional probability.

   (c) **Used for:** Not sure - this theorem is a 'common sense' fact. Our natural intuition of 'm-step transition' probabilities is now guaranteed to make sense with respect to the infinite product measure coming from Kolmogorov's consistency theorem. This is especially obvious for the time-homogeneous case and finite state space in which the transition probabilities is just a stochastic matrix.

2. **Strong markov property for MC**

   (a) **Statement:** Given a MC and a stopping time τ, the conditional distribution of the sequence following a stopping time given 'information up to the stopping time' (this is the stopping time sigma-field - take any measurable set, intersect it with event that $\tau \leq n$, and see if it is $\mathcal{F}_n$-measurable) is the same as the original chain but starting at the state attained at the stopping time (X_τ) - a.e. in the set where τ is finite.

   (b) **Depends on:** The standard procedure for proving conditional probabilities - take any τ-measurable set, intersect with any event pertaining to a finite sequence following the stopping time. By conditioning on $\tau = k$ and summing over k, we can exploit the definition of τ-measurable sets.

   (c) **Used for:** This shows that our intuitive notion of the Markov property (i.e. If we know the sequence at some *fixed* point in time is x, the distribution on values thereafter is like starting the sequence over at x) now extends to stopping times. When τ is just the hitting time of some state, they are interpreted as renewal times.

3. **Irreducible markov chains**

   (a) **Statement:** If an MC is irreducible and has a countable state space, then the states are either all transient, all positive recurrent, or all null recurrent.

   (b) **Depends on:** If there is communication between 2 states, it follows that one is recurrent iff the other is. If x is positive recurrent, we can bound the expected recurrence time of y by the sum of the expected hitting time of x starting at y and the reverse expected hitting time, which in turn is bounded by $2/p$ times the expected recurrence time of x, where p is the probability of hitting y before x starting at x.

   (c) **Used for:** ?

4. **Necessary/sufficient conditions for transcience**

   (a) **Statement:** If an MC is irreducible and has a countable state space, then it is transient iff the expected number of hits (including possible the first) to state y starting at state x is finite for *some* pair of states (x, y). In addition, this value is equal to the probability that you hit y eventually starting at x, multiplied by the expected hits to y starting at y. Finally, the expected number of hits to x starting from x the reciprocal of probability that you never return to x starting from x.

   (b) **Depends on:** The number of returns is a random variable with geometric distribution and parameter equal to the probability that you will return eventually. To get the expected number of visits to y starting from x, we just compute the probability of getting to y eventually, and using the renewal property we just multiply this by the expected number returns to y starting from y.

   (c) **Used for:** For specific MC's, we can figure out whether it is transient or not just by analyzing the n-step transition probabilities on *emphone* pair of states. Usually we can pick $x = y = 0$ and check this.

5. **SSRW on $\mathbb{Z}^d$**

   (a) **Statement:** Consider the simple symmetric random walk on $\mathbb{Z}^d$ where you add ± 1 to some coordinate at each time step w.p $1/2d$.

The $2n$-step transition probability from 0 to 0 is asymptotically order of $1/(n^{d/2})$ (and 0 for *odd*-step transitions). It follows that for $d \leq 2$ the MC is null recurrent and for $d > 2$, it is transient.

(b) **Depends on:** Once we show the estimate, just apply the previous theorem. Showing the estimate involves interpreting the random walk as fourier coefficients of an L_2 function on the d-dimensional torus $[-\pi, \pi]^d$.

(c) **Used for:** ?

6. **Positive recurrence and invariant distributions**

   (a) **Statement:** Consider an (irreducible) recurrent MC with countable state space. If the MC is null recurrent, then the n-step transition between any state pair goes to 0. If the MC is positive recurrent, then the n-step transition from x to y tends to the reciprocal of the expected recurrence time of y (and is indep. of x). In particular, this limit is an invariant distribution for the MC.

   (b) **Depends on:** This corresponds to basic intuitition that for large n, the n-step transition should be the asymptotic fraction of time that the chain spends in state y, i.e. one over the mean recurrence time.

   (c) **Used for:**

5 Martingales

1. **Doob's inequality - analog of chebychev's for martingales**

   (a) **Statement:** If we we have submartingale sequence, the probability that the maximum size of the sequence up to time n exceeds a number l is bounded above by the p'th moment of X_n divided by l^p for any $p \geq 1$. There is also the following corollary - the p'th moment of the sup of the sequence (up to time n) is bounded above by $(\frac{p}{p-1})^p$ times the p'th moment of the n'th element in the sequence. Finally, this also naturally generalizes to continuous-time martingales that have cnts sample paths.

   (b) **Depends on:** Similar conditioning trick as in Kolmogorov's inequality. Bound the probability of the event that the sequence exceeds l for the first time at time j using Markov's inequality and the fact that any convex function *e.g.*, $|.|$ of a (sub)martingale is a submartingale (this is basically Jensen's inequality). To show the corollary is manipulations using Fubini/Holders.

(c) **Used for:** This gives us a chebychev-type bound on the sup of the sequence that depends only on the last element of the sequence. This gives us hope for proving some types of convergence of martingales.

2. **Martingales with L_2 differences**

   (a) **Statement:** Suppose a martingale sequence has square-integrable differences, and assume also that the sequence is uniformly bounded in the L_2-norm. Then the sequence converges in L_2 to some L_2 random variable.

   (b) **Depends on:** Since the covariance of 2 distinct difference is 0, we can write the second moment of a martingale element, as the sum of the second moments of the differences. Then we can use the cauchy-criterion in L_2.

   (c) **Used for:** This is a convergence result specialized for L_2 - however if we assume the martingale is uniformly bounded in ω and time, then all types of L_p convergence are equivalent. This will be used in proving more general convergence results (see below).

3. **'Projected' martingales converge in L_p and a.s.**

   (a) **Statement:** Suppose a martingale sequence is generated by taking conditional expectations (projections) of a random variable X wrt a filtration. Then if X is in L_p for some $p \geq 1$, the martingale converges to X in L_p and a.s.

   (b) **Depends on:** L_p convergence is easier - show first for bounded X using the previous theorem. Then if X is L_p we can approximate it with a bounded function and generate a martingale with respect to that function, then apply Minkowski's inequality. A.s. cvgnce is hard - see book.

   (c) **Used for:** This is the first of the convergence results for martingale. It's not so useful because we have to already know there is a random variable 'generating' the martingale sequence. The following theoerems address this.

4. **Sufficient conditions for 'projection limit' and L_p convergence**

(a) **Statement:** Suppose a martingale sequence is uniformly bounded in the L_p norm for some $p > 1$. Then there is a r.v. in L_p which generates the martingale via projections wrt the filtration, and the martingale converges to this variable in L_p.

(b) **Depends on:** The last part follows from the previous theorem. To show the first part, use L_p duality (need $p > 1$) and weak compactness of bounded sets in L_p to get a weakly convergent subsequence of the martingale in L_p (as an L_q fn take the char fn of some $\mathcal{F}_n$-measurable set).

(c) **Used for:** ?

5. **Sufficient conditions for 'projection limit' and L_1 convergence**

   (a) **Statement:** Suppose a martingale sequence is uniformly integrable. Then there is a r.v. in L_1 which generates the martingale via projections wrt the filtration, and the martingale converges to this variable in L_1.

   (b) **Depends on:** The uniform integrability gives us weak compactness in L_1, and so we can follow the strategy in the previous theorem.

   (c) **Used for:** ?

6. **Stopping time properties**

   (a) **Statement:** Constant times are stoppping times. Stopping times are preserved under nondec. functions that lie above the diagonal. They are also preserved under max and min operations. If we have a sequence measurable wrt filtration, the sequence at a stopping time is measurablwe wrt the stopping time field.

   (b) **Depends on:** easy.

   (c) **Used for:** If a stopping time is finite a.s., we take $\tau = \lim \min\{tau, n\}$ (a limit of bounded stopping times).

7. **Optional stopping theorem**

   (a) **Statement:** Given a (sub/super)martingale sequence and 2 non-neg. bounded and ORDERED stopping times ($0 \leq \tau_1 \leq \tau_2 \leq C$) (a.s.), the conditional expectation of the sequence at the later stopping time wrt the sigma-field generated by the former stopping time is equal to (more/less) the value of the sequence at the former stopping time (a.s.).

(b) **Depends on:** Reduce it to showing that for any time greater than a stopping time (a.s.), the conditional expectation of the sequence at that time wrt the stopping time field is just the sequence at the stopping time (a.s.). Then use iterated conditional expectation.

(c) **Used for:** A common application is to use τ_1 as constantly 0 and τ_2 some bounded stopping time to get that the expected value of the martingale at the stopping time is just the mean at time 0. Typically if a stopping time is unbounded (e.g., the hitting time of some value) we can take a limit of bounded stopping times. This can be used to derive the 'Gambler's ruin' result - for SSRW, what is the probability that you hit a before $-b$?

8. **Doob decomposition theorem**

(a) **Statement:** Any submartingale can be written as a sum of a martingale and a sequence A_n of (a.s.) nondecreasing r.v.'s starting at 0, and where A_n is $\mathcal{F}_{n-1}$-measurable for $n \geq 2$. The decomposition is unique.

(b) **Depends on:** Just define the nondecreasing sequence inductively by adding the conditional expectation of the n-to-$(n-1)$ martingale differences wrt the $\mathcal{F}_{n-1}$. The rest is just verification.

(c) **Used for:** Extending results on martingales to results on sub/supermartingales. For instance, the optional stopping time can be extended to submartingales with equality replaced by an inequality.

9. L_1 **martingales**

(a) **Statement:** A martingale sequence uniformly bounded in the L_1 norm can be written as the difference of 2 nonneg. martingales (with the same filtration). In addition a nonnegative uniformly bnded martingale (wrt L_1 norm) has an almost sure limit (this limit may not in general 'generate' the martingale).

(b) **Depends on:** Doobs inequality and the L_1 bound imply that the probability that martingale exceeds l is less than K/l, which in turn implies that this maximum is finite w.p. 1. Restricting to this set where the sequence is uniformly bounded, any ω for which the limit DNE must 'upcross' some pair of rationals infinitely often. The upcrossing inequality says that the prob. of this happening for any pair of rationals is 0.

(c) **Used for:** ?

10. **Doob's upcrossing inequality**

   (a) **Statement:** The expected number of upcrossings of an interval (by a martingale sequence)up till time n is bounded above by the expected value of the rectified difference between the left endpoint and the n'th value of the martingale, divided by the length of the interval.

   (b) **Depends on:** Optional stopping theorem - create stopping times associated with the martingale entering the area above or below the interval in question (use a minimum trick). Then look at the sum of the difference in the sequence values at this time (each difference is more than the length of the interval) and how it relates to the number of upcrossings. Show that the expected value of this sum (except the last term) is 0 by Optional stopping time theorem.

   (c) **Used for:** Showing that L_1-bounded martingales have a.s. limits.

6 Ergodic theorems

1. **Equivalence of stationary stochastic processes and measure-preserving maps**

   (a) **Statement:** Stationary (doubly-infinite) sequences are essentially equivalent to measure-preserving maps on the probability space.

   (b) **Depends on:** To go forward, we interpret the sequence as random variables on the product space of infinite sequences. The measure-preserving map is just the shift operator. To go backward, take any r.v. X and define a sequence of other r.v.'s $X_n = X(T^n\omega)$ where T is the given measure-preserving map. The resulting sequence is then stationary.

   (c) **Used for:** This basically says we can look at infinite sequences as moving from ω to ω via repeated application of a measure-preserving operator on a simple space (this is like moving in space) OR think of the entire sequence as one ω in a product space (this is like moving in time).

2. **Ergodic theorem**

   (a) **Statement:** Let P be a T-invariant measure. Then for any integrable function f the 'time' average of f (by repeated application of T) converges a.s. to the conditional expectation of f wrt the sigma-field of T-invariant sets. If P is ergodic for T then this is just the expectation ('space' avg) of f. If f is in L_p then we have convergence in L_p also.

(b) **Depends on:** For $p = 2$ we can use Hilbert space techniques involving orthogonal decomposition. The idea is to consider the operator $(Uf)(\omega) = f(T\omega)$ and consider the subspace $Ker(U-I)$. The RHS is just a projection onto this subspace

(c) **Used for:** This is a generalization of the SLLN to dependent random variables, namely ergodic sequences. For these sequences, taking a time average (over several samples) is asymptotically equivalent to taking a space average (i.e. an expectation integral).

3. **Extremal measures**

   (a) **Statement:** The set of T-invariant measures on a probability space is a convex set. Furthermore, a T-invariant measure is ergodic for T iff it is an extremal point in this set. Any 2 distinct ergodic measures are orthogonal.

   (b) **Depends on:** Convexity is easy to show. If P is not extremal (i.e. a convex combo), we can apply the ergodic theorem to its two consitutent T-invariant measures and use ergodicity to show that the constitutents must be identical measures. If P is not ergodic, then there is a T-invariant set on which they differ. We can define 2 measures that linearly combine to get the original measure (just restrict to the set and its complement).

   (c) **Used for:** This is another characterization of what ergodic measures for a map T are. Notice that these are exactly the measures for which the space/time average are asymptotically equivalent.

4. **Invariant measures for stationary MC's**

   (a) **Statement:** Consider an MC with state space $(X, \mathcal{B})$ and transitions π. The set of dbns $\mathcal{M}$ on $(X, \mathcal{B})$ that are invariant for π is a convex set. Furthermore, μ is extremal in $\mathcal{M}$ iff the induced measure P_μ is ergodic in the set of shift-invariant measures on the infinite product space $(\Omega, \mathcal{F})$

   (b) **Depends on:**

   (c) **Used for:**

5. **CLT for martingales**

(a) **Statement:** The CLT holds for stationary ergodic sequences that are uniformly L_2-bounded martingale differences (that is, the conditional expectation of the n'th element wrt the $n-1$'th sigma-field is 0).

(b) **Depends on:** characteristic functions

(c) **Used for:** Proving CLT's for special sequences of dependent random variables. For instance we can apply this (nontrivially) to the sequence generated by applying a function f to an MC.