310 Quals Strategy Compendium

July 14, 2025

1 Permutation and counting facts

Fact 1 (Number derangements of k-element set)
Derangements: D_n is the number of permutations with no fixed points.

Via inclusion exclusion.

$$D_k = k! \sum_{j=0}^{k} \frac{(-1)^j}{j!}.$$

Eg, if T is number of fixed points:

$$P(T = k) = \frac{1}{n!}\binom{n}{k} D_{n-k},$$

since the remaining $n-k$ must **not** be fixed.

Apply the above to "Distribution of number of fixed points"- type questions.

Fact 2 (Catalan Numbers)
Catalan Numbers:

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!n!} \quad n \geqslant 0$$

Dyck Paths

Definition 3 (Cycles)
Let $s \in S_n$ be a permutation. Then $L_i(s) = \inf\{j : s^j(i) = i\}$. Ie the first time we come back to i. See pg 101 Dembo remark. Durret Ex 2.2.4. Then the cycle is the collection $\{s^j(i) : 1 \leqslant j \leqslant L_i(s)\}$– ie the elements of $[n]$ before we come back to i.

If $T_n = \#\text{cycles}$, $(T_n - \log n)/\sqrt{\log n} \xrightarrow{d} \mathcal{N}(0, 1)$.

Definition 4 (Descents)

Reference: check Persi and Susan's paper

2 Distribution Facts

2.1 Gaussian Facts

Fact 5 (Max of Gaussians and fluctuations)
Max of Gaussians $\sqrt{2\log n}$ with fluctuations $1/\sqrt{\log n}$

Fact 6 (Max of sub-gaussian)
Expectation upper bound applies to correlated sub-gaussian reandom variables – see Vershynin.

Fact 7 (Mill's Ratio)

Fact 8 (Normal Conditional Distributions)

$$(X, Y) \sim MVN(\mu, \Sigma \implies X|Y \sim \dots$$

2.2 Poisson, Exponential Distribution

Superposition and thinning.

Fact 9 (Superposition)
See Poisson with integer mean? Try superposition:

$$Pois(n) = \sum_{i=1}^{n} Pois(1)$$

Fact 10 (Renyi Representation of Exponential)

Fact 11 (Maximum, minimum of Exponential)

3 Basic set theory and measure theory

Definition 12 (Sigma Algebra)

Definition 13 (Algebra/Field)

Definition 14 (Outer measure)
Defined by

1. A non-negative set function
2. ......

The typical outer measure is WRite this out

Definition 15 (Measurable sets)
E is μ^* measurable if for all $B \subset \Omega$:

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c).$$

Write about littlewood's principles here

Example 16 (Non-measurable sets)
Todo

4 Pi-Lambda and Good Sets

Definition 17 (π-system)
Collection of sets that is closed under **finite intersections**

Definition 18 (λ-system)
L lambda system if

1. $\Omega \in L$
2. closed under complements
3. closed under countable disjoint unions

Alternative definition:

1. $\Omega \in L$
2. $A, B \in L$ and $A \subset B$ then $B \setminus A \in L$
3. $A_1, A_2, \ldots, \in L$ an increasing sequence of sets, then $\bigcup_i A_i \in L$

Note that a collection is a σ-algebra $\iff$ it is both a pi system and a lambda system.

Theorem 19 ($\pi - \lambda$ Theorem)
If P is a π system and L is a λ-system with $P \subset L$, then $\sigma(P) \subset L$.
Use to proof uniqueness of extension from an algebra to the sigma field.

Example 20 (Quals 2017, Question 2)

Theorem 21 (Monotone Class Theorem)
Use def:

Definition 22 (Monotone Class)
M is a monotone class if

1. Closed under increasing unions
2. Closed under decreasing intersections

If an algebra A is contained in a monotone class M, then $\sigma(A) \subset M$.

Theorem 23 (Monotone Class for functions)
Double check this
M be a vector space of $\mathbb{R}$-valued functions on Ω such that

1. $1 \in M$
2. M is a vector space over $\mathbb{R}$. Ie closed under addition and scalar mult
3. If $h_n \geqslant 0 \in M$ and $h_n \uparrow h$, then $h \in M$.

Then if P is a pi system such that $\mathbf{1}_A \in M$ for all $M \in P$, then M contains all functions that are measurable with respect to $\sigma(P)$.

Ie want to prove something about some functions measurable on $\sigma(P)$...

Theorem 24 (Independent π systems generate independent sigma algebras)

$$\{C_i\}_{i \in I} \text{ independent } \implies \{\sigma(C_i)\} \text{ independent}$$

Example 25 (Prove two random variables independent)
Check that generating pi systems are independent, ie $P(X \leqslant b, Y \leqslant a) = \prod P(X \leqslant b)P(Y \leqslant a)$.

Definition 26 (Independence of (uncountable) collections of sets)
$\{A_\alpha\}_{\alpha \in I}$ each be a collection of sets. Independent if any **finite subcollection** are mutually independent.

5 Extension theorems

Theorem 27 (Caratheodory)
Extend a measure on an algebra to a σ algebra, uniquely if finite.

The important idea is make an outer measure, then prove the lemma is that the μ^* measurable sets in $\mathcal{F}$ for a sigma algebra on which μ^* is countably additive, ie a bonafide measure.

6 Random variables

Definition 28 (Random variable)
$X : \Omega \to \mathbb{S}$ is measurable if:

$$X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F} \quad \forall B \in \mathcal{S}$$

Fix an arbitrary set in the co-sigma algebra and check that its preimage is measurable-$\mathcal{F}$.

Recipe 29 (Measurability)
If $\mathcal{S} = \sigma(\mathcal{A})$ and $X^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{A}$, then X is measurable.

Note– nothing necessary about it being a pi system, just pick any convenient generators.

Proof. Such sets form a sigma algebra and $\mathcal{S} = \sigma(\mathcal{A})$. □

Definition 30 (Characterize distribution functions of real-valued random variable)
Dembo Theorem 1.2.37.

1. Non decreasing
2. $\lim_{x\to\infty} F(x) = 1$ and $\to -\infty$ gives 0
3. F right continuous

7 0-1 Laws

Definition 31 (Tail field)
Defining $\mathcal{T}_n = \sigma(X_r, r > n)$ and the tail sigma algebra of the process is $\mathcal{T} = \cap_n \mathcal{T}_n$.

Theorem 32 (Kolmogorov 0-1)
The tail sigma field is P-trivial. Dembo 1.4.10

8 Borel Cantelli

Example 33 (Longest head runs)
See Durrett.

A helpful idea is splitting into blocks.

Truncation arguments. Want to show something about $\{X_n\}$. Consider:

$$Y_n = X_n \mathbf{1}[|X_n| \leqslant c_n],$$

which will then be integrable. Then prove something about Y_n, you can transfer this knowledge to knowledge about X_n by the following idea:

$$P(X_n \neq Y_n \text{ io}) = 0 \text{ if } \sum_{i=1}^{n} P(|X_n| > c_n) < \infty$$

So if the above is zero, eventually $X_n = Y_n$, so the limiting behavior of Y_n is the same as that of X_n. The difficulty is picking a c_n such that the above holds.

Theorem 34 (Borel Cantelli 3)
Not as useful but still check these
If you have a filtration $\{F_n\}$ and $A_n \in \mathcal{F}_n$, then:

1. (analog of Borel Cantelli 1)

$$\sum_{k=1}^{\infty} P(A_k|\mathcal{F}_{k-1})(\omega) < \infty \implies \sum \mathbf{1}[\omega \in A_k\} < \infty$$

2. Analog of BC 2

$$\sum_k P(A_k|\mathcal{F}_{k-1})(\omega) = \infty \implies \frac{\sum \mathbf{1}[A_k(\omega)]}{\sum P(A_k|\mathcal{F}_{k-1})(\omega)|} \to 1$$

9 Modes of Convergence

Recipe 35 (Proving as convergence)
Some ideas:

1. BC 1: check that $P(|X_n - X| > \epsilon \text{ i.o}) = 0$. Consider *subsequence trick* below.
2. If $X_n \xrightarrow{p} X$, then $X_{n_k} \xrightarrow{\text{a.s.}} X$ for a subsequence. (Good for counter examples)
3. $X_n \xrightarrow{p} X$ and X_n is monotone (ie for all ω, $X_n(\omega)$ is increasing in n), then $X_n \xrightarrow{\text{a.s.}} X$.
4. Continuous mapping theorem
5. Skorohod's Representation. If $X_n \implies X$ weakly, then there exists a probability space and random variables $Z_n =^d X_n$ and $Z =^d X$ such that $Z_n \xrightarrow{\text{a.s.}} Z$. (Good for counter examples)

Theorem 36 (BC Subsequence trick)
Sometimes $\sum P(A_n) = \infty$.

1. Pick a subsequence such that it's finite.
2. Shows that A_{n_k} infinitely often occurs with probability 0.
3. Apply interpolation or some other argument for everything in between, ie to conclude that A_n infinitely often wp 0 also.

Eg for monotone X_n:

$$\frac{X_n}{\mathbf{E}X_n} \leqslant \frac{X_{n_{k+1}-1}}{\mathbf{E}X_{n_k}}$$

So $\limsup_k \frac{X_{n_{k+1}-1}}{\mathbf{E}X_{n_k}} \leqslant 1 \implies \limsup_n \frac{X_n}{\mathbf{E}X_n} \leqslant 1$. We can control the left hand side by controlling the right hand side along our choice of subsequence.

See Dembo notes subsequence section Examples– SLLN, Brownian LLN, LIL.

SLLN in Persi notes and Renewal theorem in Dembo

Recipe 37 (Proving almost sure limit is finite)
Try:

1. Take expectation - use MCT or something and show that finite, so random variable is finite wp 1
2. If $\sum_n \mathbf{Var}(X_n) < \infty$ and $\mathbf{E}X_n = 0$ (2.3.17 Dembo) (Kolmogorov 2 series theorem)
3. Kolmogorov 3 series

Recipe 38 (Proving conv in prob)
Some ideas:

1. Show almost sure convergence
2. Show convergence in L^p for $p \geqslant 1$ (in L^2 in particular useful):
    - Show that $\mathbf{E}X_n \to \mu$, $\mathbf{Var}X_n \to 0$, then $X_n \to \mu$ in L^2
3. Show $\mathbf{Var}X_n/b_n^2 \to 0$ and use Markov to get $b_n^{-1}(X_n - \mathbf{E}X_n) \xrightarrow{p} 0$ (and also in L^2).
4. CMT
5. $X_n \xrightarrow{d} c$ constant
6. Truncation tool: **Weak law for Triangular Array**. For when we don't have second moments.

Theorem 39 (Coupon Collector)
If T_n is time to get all n possible coupons. Then

$$T_n/(n \log n) \xrightarrow{p} 1$$

(and in L^2)– see Dembo 2.1.8.

Theorem 40 (Weak law triangular array - 2.1.11 Dembo)
Check this Suppose triangular array $\{X_{n,k}\}_{n\in\mathbb{N},k\leqslant n}$ of pairwise independent rv. Define a truncated array with the same *within-row* truncation:

$$\overline{X}_{n,k} = X_{n,k}\mathbf{1}[|X_{n,k}| < b_n]$$

. If we have two conditions:

$$\sum_{k=1}^{n} P(|X_{n,k}| > b_n) \to^n 0,$$

and,

$$b_n^{-2}\sum_{k=1}^{n} \mathbf{Var}(\overline{X}_{n,k}) \to 0,$$

Then:

$$b_n^{-1}(\sum_{k=1}^{n} X_{n,k} - a_n) \xrightarrow{p} 0 \quad \text{where } a_n = \sum_{k=1}^{n} \mathbf{E}\overline{X}_{n,k}.$$

In case of St Petersburg paradox, then show that $a_n/b_n \to 1/2$ so that we get $b_n^{-1}\sum_{k=1}^{n} X_k = (b_n^{-1}\sum_{k=1}^{n} X_{n,k}) \xrightarrow{p} 1/2$.

Recipe 41 (Prove L^p convergence)
Ideas - see session notes

1. $X_n \xrightarrow{\text{a.s.}} X$ and X_n uniformly bounded then $X_n \to^{L1} X$.
2. If X_n uniformly bounded then $X_n \xrightarrow{d} 0 \implies X_n \to^{L1} 0$.
3. If $|X_n|^p$ is UI then $X_n \xrightarrow{d} X \implies \mathbf{E}|X_n|^p \to E|X|^p$
4. Scheffe's Lemma

Recipe 42 (Proving UI)
Try

1. Sums of UI sequence of rvs are UI
2. Domination: If $E \sup |X_i| < \infty$ then $\{X_i\}$ is UI
3. If $\{X_i\}$ is UI, then $\sup E|X_i| < \infty$ but not the converse.
4. Prop 8.3.3: X_n UI $\iff$ $\sup E|X_i| < \infty$ and for all ε, there exists δ for all A such that $P(A) < delta$ then $\mathbf{E}[|X_n|\mathbf{1}[A]] < \varepsilon$
5. If $\sup \mathbf{E}|X_i|^r < \infty$ if $r > 1$, then UI
6. $X \in L^1$, then $\{\mathbf{E}[X|\mathcal{G}] : \mathcal{G} \subset \mathcal{F}\}$ is UI
7. L^1 convergence implies UI

10 Integration

Definition 43 (Lebesgue Integral)

$$\int_\Omega f(\omega) d\mu(\omega) = \sup \left(\sum_{i=1}^n \nu_i \mu(A_i) \right)$$

where $\nu_i = \inf_{\omega \in A_i} f(\omega)$ and the sup is over all partitions of Ω.

Theorem 44 (MCT)
Note that can also do for general functions (not necessarily non negative) so long as $f_n \geqslant g \in L^1$.

Theorem 45 (Fatou's Lemma)

$$\int \liminf_n f_n d\mu \leqslant \liminf_n \int f_n d\mu$$

Theorem 46 (DCT)
If $f_n \to f$ a.e ω, and there exists g such that $|f_n(\omega)| \leqslant g(\omega)$ a.e ω and $\int g d\mu < \infty$ then exchange integral and limit. Must dominate the sequence.

Theorem 47 (Scheffe's Lemma)
Is a statement about combining L^1 and as convergence. If $f_n \xrightarrow{\text{a.s.}} f \in L^1$, then:

$$\|f_n - f\|_{L^1} \to 0 \iff \int |f_n| d\mu \to \int |f| d\mu$$

We always have that convergence in L^1 implies convergence of the expectations (without any a.s. convergence assumption) , but the other direction is Scheffe's contribution.

Theorem 48 (Reverse Fatou)
If $f_n \leqslant g \in L^1$ then :

$$\limsup_n \int f_n d\mu \leqslant \int \limsup f_n d\mu$$

Fact 49 (L^p spaces nested)

$$\|Y\|_r \leqslant \|Y\|_q$$

Fact 50 (L^q convergence fact - Dembo 1.3.28)
$X_n \to^{L^q} X \implies \mathbf{E}|X_n|^q \to \mathbf{E}|X_\infty|^q$ for any q. (Minkowski)

Also for only $q \in \mathbb{N}$, $\mathbf{E}X_n^q \to \mathbf{E}X_\infty^q$. (some wild algebraic shit for odd q).

Theorem 51 (Holder's)
If $p, q > 1$ with $1/p + 1/q = 1$ then

$$E|XY| \leqslant \|X\|_p \|Y\|_q$$

Cauchy Schwarz is special case.

Theorem 52 (Minkowski)
Triangle inequality for the $\|\cdot\|_p$ norm

Definition 53 (Uniform integrability (UI))
Possibly uncountable collection $\{X_\alpha : \alpha \in I\}$ is called UI if

$$\lim_{M\to\infty} \sup_{\alpha \in I} \mathbf{E}[|X_\alpha| \mathbf{1}[|X_\alpha| > M] = 0$$

Fact 54 (Dominated implies UI)
If $|X_\alpha| \leqslant Y$ for integrable Y, then collection is UI.

As a corollary, any finite collection of integrable rv is UI.

Theorem 55 (Vitali Convergence Theorem)
Supposing that $X_n \xrightarrow{p} X$, then:

$$\{X_n\} \text{ is UI} \iff X_n \to^{L1} X \iff X_n \text{ is integrable for all } n \leqslant \infty \text{ and } E|X_n| \to E|X_\infty|.$$

11 Product σ-algebras

Existence of unique product measure of n $\sigma-$ finite measures.

Theorem 56 (Kolmogorov Extension)
Unique probability measure on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}_c)$ with correct FDDs.

Theorem 57 (Fubini's)
Conditions: $h \geqslant 0$ or $\int |h| d\mu < \infty$ where $\mu = \mu_1 \times \mu_2$.

12 Weak Convergence

12.1 Methods

Main tools:

1. **Direct**. Show that $F_n(x) \to F(x)$ for all continuity points.
2. **Characteristic functions-** show that $\phi_n(t) \to \phi(t)$ for all $t \in \mathbb{R}$. (ϕ is a characteristic function of a probability measure if $\psi(t) \to 1$ as $t \downarrow 0$.
3. **CLT**

Example 58 (Cycling of Random Number Generators (2007 Q2))
Similar to birthday problem - direct method.

$$
\begin{aligned}
P(T > k) &= \prod_{i=1}^{k} (1 - \frac{i}{n}) && (1)\\
&\approx \prod \exp(-1/n) && (2)\\
&\approx \exp(-k^2/n). && (3)
\end{aligned}
$$

So $P(T > x\sqrt{n}) \approx \exp(-x^2/2)$ should work.

Need to justify this rigorously. To do so, use $|\log(1 + x) - x| < Cx^2$ when $|x| < 1/2$. Ie, $\log(1 + x) = x + O(x^2)$.

See lecture 1 in 310a.

Backup plans

1. If there's a density, try to show that $f_n(x) \to f_\infty$ and check that f_∞ a valid pdf.
2. If $X_n \geqslant 0$, show that $\int_0^\infty \exp(-\lambda x) d\mu_n(x) \to L(\lambda) = \int_0^\infty \exp(-\lambda x) d\mu_\infty(x)$. Note that $L(\lambda)$ is a Laplace transform of some μ if $L(\lambda) \downarrow 1$ as $\lambda \downarrow 0$. (Just need for positive λ.

3. MGFs
4. Moment Method
5. Can you use a Prohorov argument?

Theorem 59 (Moment Method)
If can show that $\int X^k \mu_n(dx) \to \int X^k \mu_\infty(dx)$ and μ_∞ is the only measure with its particular moments, then $\mu_n \xrightarrow{d} \mu$.

To show the second thing, the following are sufficient:

1. Mgf exists
2. Power series condition?
3. Carleman's condition: $\sum_{k \geqslant 1} (m_{2k})^{-1/2k} < \infty$

Example 60 ((Non-)Examples defined by moments)
Examples: Normal, poisson, exponential.

Non-examples: Log-normal

Theorem 61 (Convergence of type)
If two normalizing sequences, then those sequences are basically the same up to some scaling.

Idea is you can't change the class of limniting distribution by rescaling.

13 CLT

Heuristic: "not too dependent", "no few terms dominate".

Theorem 62 (Lindeberg CLT)
Suppose we have a triangular array such that:

1. for fixed n, $\{X_{ni}\}_{i=1}^{k_n}$ are independent (ie independence within row).
2. Suppose also $\mathbf{E}(X_{ni}) = 0$ for all, and $\mathbf{Var}X_{ni} = \sigma_{ni}^2 < \infty$.
3. Define $S_n = \sum_{i=1}^{k_n} X_{ni}$ and $s_n^2 = \sum_{i=1}^{k_n} \sigma_{ni}^2$ (sum of rows)
4. **Lindeberg condition** holds ie for all $\varepsilon > 0$:

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{k_n} \int X_{ni}^2 \mathbf{1}[|X_{ni}| > \epsilon s_n] dP = 0$$

Then:

$$\frac{S_n}{s_n} \xrightarrow{d} \mathcal{N}(0,1).$$

Note that if not mean zero, subtract off the means:

$$\frac{S_n - \sum_{i=1}^{k_n} \mu_{kn}}{s_n} \xrightarrow{d} N(0,1)$$

Example 63 (2007 Problem 1)
X_k takes values $\pm k^a$ wp $\frac{1}{2}k^{-\alpha}$ each and ± 1 wp $\frac{1}{2}(1-k^{-a})$ each. Ie Rademacher plus something big. Want to know if S_n/c_n is Gaussian.

For $a > 1$, apply BC + CLT on iid Rademachers to get a normal limit.

For $a < 1$, try Lindeberg:

$$\mathbf{Var}X_k \sim \frac{k^{2a}}{k^a} \implies \sigma_n^2 \sim 2\sum_{k=1}^{n} k^a \sim n^{a+1}$$

And apply Lindeberg.

For $a = 1$ case use CF.

Example 64 (CLT failures: too wild)

$$X_i = \begin{cases} 0 \text{ wp } 1 - 1/i \\ 1 \text{ wp } 1/i \end{cases}$$

The issue is that some X_i's dominate– ie the big ones.

Example 65 (CLT Failures: too dependent)

Recipe 66 (CLT for non-square integrable)
Session 4 notes. Similar to convergence in probability strategy.

Assume $X_{n,k} \in L^1$ and exist c_n such that

1. $\sum_{k=1}^{\ell_n} P(|X_{n,k}| > c_n) = o(1)$
2. Lindeberg condition satisfied for truncated $Y_{n,k} = X_{n,k}\mathbf{1}[X_{n,k} \leqslant c_n]$
3. $\sum_{k=1}^{\ell_n}(\mathbf{E}X_{n,k} - \mathbf{E}Y_{n,k}) = o(s_n)$ where s_n^2 sum of variances of truncated in the n-th row.

Then

$$\frac{\sum_{k=1}^{\ell_n} X_{n,k} - \mathbf{E}X_{n,k}}{s_n} \xrightarrow{d} N(0,1)$$

For example, see Dembo 3.1.12.

Theorem 67 (Lyapunov CLT)
Lyapunov condition is sufficient for Lindeberg's condition. Same setup, check that:

$$s_n^{-2-\delta} \sum_{i=1}^{k_n} \mathbf{E}|X_{n,k} - \mathbf{E}(X_{n,k})|^{2+\delta} \to 0 \quad \text{for some } \delta > 0$$

Theorem 68 (Other CLTs for dependent rvs)
Look at Sourav/Quals notes

Theorem 69 (Kolmogorov 3 Series Theorem)
....

Some helpful CLT references:

1. Exponential approximation for the geometric - pg 105 dembo
2. Normal approx to Poisson - Dembo ch 3
3. Normal approx to Binomial

Some limit theorems for max of random variables (Ex 3.2.13 Dembo). The below three are the only possible type of limits for max of iid random variables.

1. Max of exponentials or normals is is Gumbel-type $F_\infty(y) = \exp(-e^{-y})$. (see also 3.2.14)
2. Frechet type $F_\infty(y) = \exp(-y^{-\alpha})$
3. Weibull type $F_\infty(y) = \exp(-|y|^\alpha)$

A few more examples of max of rvs and limiting distributions

Example 70 (Birthday problem limiting law)
T_n is the number needed to get a match with n possible birthdays (eg $n = 365$), then can show that $P(n^{-1/2}T_n > s) \to \exp(-s^2/2)$ - Ex 3.2.15. See also the RNG 2007 Q2 question below.

14 Characteristic Functions

Definition 71 (Characteristic Function)
Characteristic function is Fourier transform of μ.

$$\phi(t) = \mathbf{E}[\exp(itX)].$$

Recipe 72 (Characteristic Function Chaos)
Try to get $\phi(t) \approx (1 + \frac{f(t)}{n})^n$ form so that we can use exp limit.

Recipe 73 (Characteristic Function)
Given a triangular array, want to show weak convergence of $\sum_{k=1}^{\ell_n} X_{n,k}$.

1. Center the random variables if necessary
2. Try to justify approximation:

$$\log \prod_{k=1}^{\ell_n} \phi_{n,k}(t) \approx -\sum_{k=1}^{\ell_n}(1-\phi_{n,k}(t))$$

3. Use Lemma: Dembo 3.3.31, Chaterjee 8.10.4 which says:

   If $a_1, \ldots, a_n$ and $b_1, \ldots b_n$ complex with maximum modulus 1 then:

$$|\prod_{j=1}^{n} a_j - \prod_{j=1}^{n} b_j| \leqslant \sum_{j=1}^{n} |a_j - b_j|$$

   Then since Taylor Approx gives $|\exp(z) - 1 - z| \leqslant C|z|^2$ for all small $|z|$, then

$$|\prod_{k=1}^{\ell_n} \phi_{n,k}(t) - \exp\left[-\sum_{k=1}^{\ell_n}(1-\phi_{n,k}(t)\right]| \leqslant \sum_{k=1}^{\ell_n} |\phi_{n,k}(t) - \exp(-[1-\phi_{n,k}(t)])| \leqslant C\sum_{k=1}^{\ell_n} |1-\phi_{n,k}(t)|^2$$

   Want to show that this goes to 0. Which we can do by showing:

$$\sum_{k=1}^{\ell_n} |1-\phi_{n,k}(t)| = O_n(1) \quad \max_{k\leqslant \ell_n}(1-\phi_{n,k}(t)) = o_n(1)$$

   So that we can show the $C\sum|\cdot|^2$ term goes to 0.

4. See tim notes.....
5. ....
6. Apply Levy Continuity Theorem

Example 74 (Cauchy Examples 2017 Quals)
If can show that

$$\sum_{k=1}^{n} 1 - \phi_{1/X_i}(t_n) \to -c|t|$$

Because of iid case we're good

Theorem 75 (Levy's continuity Theorem)
If $\phi_{\mu_n}(t) \to \phi(t)$ pointwise and ϕ is continuous at $t = 0$, then μ_n is uniformly tight sequence and $\mu_n \implies \mu$ weakly where characteristic function of μ given by ϕ.

Conversely, weak convergence implies convergence of characteristic functions.

15 Stein's Method (Poisson)

Tool for weak convergence for Poisson approximation. Also for Gaussian approximation but we didn't cover it in class.

Gaussian heuristic: sums of weakly dependent random variables of roughly the same size.

Poisson heuristic: number of occurrences of rare events which are weakly dependent.

Idea behind Stein's method:

$$X \sim N(\mu, \sigma^2) \iff \forall \text{ nice } f, \mathbf{E}f'(X) = \mathbf{E}Xf(X)$$

So we define this Stein operator – $\mathcal{A}f(x) = f'(x) - xf(x)$, then $\mathbf{E}\mathcal{A}(f(X)) = 0$. Then vibes is that if $\mathcal{A}f(X) \approx 0$, then $X \approx$ Gaussian. Stein quantifies this.

Index set examples. The indicators should be rare.

1. Coupon collector: indicator that a coupon is **not** present. $|I| = n$.
2. ER Graph: indicator that three nodes form a triangle, $|I| = \binom{n}{3}$.
3.

15.1 Method 1 - Dependency Graphs

Define $N_i \subset I$ such that $i \in N_i \implies X_i \perp\!\!\!\perp \{X_j : j \notin N_i\}$. (Ie a dependency graph - if it's not in the neighborhood, they're independent).

$$\|P_W - \text{Poisson}(\lambda)\|_{\text{TV}} \leqslant \min\big(3, \lambda^{-1}\big)\left[\sum_{i\in I}\sum_{j\in N_i\setminus\{i\}} p_{ij} + \sum_{i\in I}\sum_{j\in N_i} p_i p_j\right].$$

Recipe 76 (Dependency Graphs)
If you can make a dependency graph that isn't complete, this is a good strat.

1. Write $W = \sum_{i\in I} X_i$ where X_i are indicators of some events.
2. Calculate p_i, λ.
3. How are the indicators correlated? Form a dependency graph.
4. Show the bound goes to 0.

Example 77 (Session 6- ER Graph Triangles)
$|I| = \binom{n}{3}$

Example 78 (Session 6 - Head Runs)
Consider n coin tosses with probability of heads p. $Y_n = \#$ head runs of at least k. Define $X_i = \mathbf{1}$[exists a head run of length at least k starting at i]. $Y_n = \sum_{i=1}^n X_i$.

$$P(X_i = 1) = \begin{cases} p^k & i = 1 \\ (1-p)p^k & i \in [2, n-k+1] \\ 0 & i > n-k+1 \end{cases} .$$

X_i and X_j are independent if we look far enough away, ie whenever $|i-j| > k$

15.2 Method 2 - Positive ASsociations

Suppose $Y = \sum_{i \in I} X_i$. Check this

15.3 Method 3 - Negative Association - More useful.

For all $i \in I$, we can construct $\{Y_j^{(i)} : j \neq i\}$ coupled with X_i such that

$$\{Y_j^{(i)} : j \neq i\} \overset{d}{=} \{X_j : j \neq i\} \mid X_i = 1 \quad \text{and} \quad Y_j^{(i)} \leqslant X_j \;\; \forall\, j \neq i.$$

In this case,

$$d_{\text{TV}}\big(Y, \text{Poisson}(\lambda)\big) \;\leqslant\; \left(1 \wedge \lambda^{-1}\right)\big(\lambda - \text{Var}(Y)\big).$$

Example 79 (Coupon Collector)
Number of missing coupons after many trials is approximately Poisson.

Isomorphic to balls into bins. N bins (N coupons). Then b balls - number of trials picking a new coupon.

$$X_i = \mathbf{1}[\textbf{box i empty}] \text{ are rare events.}$$

Moreover, they're negatively associated– if one box is empty, then the other boxes are more likely to be non-empty.

Example 80 (Baseball question - 2015, Q4)
5 cards per pack, each pack has unique cards randomly. 600 total players. How many packs do we need? This is like the coupon collector problem. For p packs,

$$p_i = (1 - 1/120)^p \approx \exp(-p/120).$$

So $W \approx Pois(\lambda)$ with $\lambda = 600 \exp(-p/120)$. Find p such that

$$P(W = 0) = 0.95).$$

Justifying the approximation. Negative association. Other players more likely to have been chosen if one player is missing. By exchangeability/symmetry: check that the negative association condition holds for $i = 1$. Ie need to show that there exists a suitable coupling. Construct as follows. Amongst all p draws, if I get a pack with player 1 in it, resample another player to substitute out player 1 (do it so that the pack is still distinct). Then:

$$Y_j^{(1)} = \mathbf{1}[\text{ player j is not in the "resampled" p packs defined above}].$$

Since whenever we get a player 1, we resample, it forces $Y_j^{(1)} \leqslant X_j$ for all $j \neq i$. CHECK THIS LATER

$$\mathbf{Var} Y = \sum \mathbf{Var} X_i + 2 \sum_{i<j} \mathbf{Cov}(X_i, X_j) = \sum p_i(1 - p_i) + 2 \sum (\mathbf{Cov}(X_i, X_j))$$

So tv distance is bounded above by:

$$(1 \wedge \lambda^{-1})[\sum p_i^2 - 2 \sum \mathbf{Cov}(X_i, X_j)]$$

15.4 Other Arratia Goldstein Method

Example 81 (Fixed Points - 310a HW8)

Let σ be a uniformly chosen permutation in the symmetric group S_n. Let $W = \#\{i : \sigma(i) = i\}$ (the number of fixed points in σ). Show that W has an approximate Poisson(1) distribution by using Stein's method to get an upper bound on $\|P_W - \text{Poisson}(1)\|$. (Hint: see section 4.5 of Arratia-Goldstein-Gordon.) Give details for this specific case.

Let $I = [n]$. We choose $B_\alpha = \{\alpha\}$ and use Theorem 1 from Arratia-Goldstein-Gordon.
For each $i \in I$, letasdf

$$X_i = \begin{cases} 1 \text{ if } \sigma(i) = i \\ 0 \text{ otherwise} \end{cases}.$$

Naturally, $P(X_i = 1) = \frac{1}{n}$. We let $W = \sum_{i \in I} X_i$ and $\lambda = E[W] = 1$. We now use Stein's method as given in Arratia-Goldstein-Gordon Theorem 1 to get an upper bound on $\|P_W - Pois(1)\|$.

$$b_1 = \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta = \sum_{\alpha \in I} p_\alpha^2 = \frac{1}{n}.$$

Next, because we let $B_\alpha = \{\alpha\}$,

$$b_2 = 0.$$

Finally, for the third term, by Lemma 2 (p 418) in Arratia et al,

$$b_3 \leqslant \min_{1<k<n} \left(\frac{2k}{n-k} + 2n2^{-k}e^e\right) \sim 2\frac{(2log_2 n + e/ln2)}{n},$$

due to the fact that $\lambda = 1$ in our problem, so $\lambda = o(n)$
Now note that as $n \to \infty$, $b_1 \to 0$ and $b_3 \to 0$, so, noting that the Arratia paper's definition of TV distance is twice our definition of TV distance:

$$\|P_W - Pois(1)\| \leqslant b_1 + b_2 + b_3 = \frac{1}{n} + \frac{4\log_2 n + 2\epsilon/\ln 2}{n} + o(1)$$

Now as $n \to \infty$, $b_1 \to 0$ and $b_3 \to 0$, so $\|P_W - Pois(1)\| \to 0$.

Example 82 (Near Fixed Points- 2004 Q2)

16 Approximations

$$1 - x \leqslant e^{-x} \quad 1 - x \geqslant e^{-2x} \quad \text{both for small } x?$$

$$\log(1+x) = x + O(x^2)) \quad \text{for small } x$$

$$-x - x^2 \leqslant \log(1-x) \leqslant -x \quad \text{for } x \in [0, 1/2].$$

16.1 Binomial Coeffs and Stirlings

$$(\frac{n}{k})^k \leqslant \binom{n}{k} \leqslant (\frac{ne}{k})^k$$

Stirlings

Theorem 83 (Stirlings)

$$(1-\epsilon)\sqrt{2\pi}k^{k+1/2}e^{-k} \leqslant k! \leqslant (1+\epsilon)\sqrt{2\pi}k^{k+1/2}e^{-k}$$

17 TV Distance

$$d_{TV}(P,Q) = \sup_B |P(B) - Q(B)| = \frac{1}{2}\int |p(x) - q(x)|dx = 1 - \int \min\{p(x), q(x)\}dx$$

18 Misc

Definition 84 (Metric)

18.1 Series convergence

Theorem 85 (Root Test)
If $S_n = \sum_{k=1}^n a_n$ then S_n converges if $L = \limsup_n |a_n|^{1/n} < 1$. If $= 1$, no info. > 1, diverges.

Theorem 86 (Kolmogorov's Maximal inequality)
Y_i mutually independent with $\mathbf{E}Y_i^2 < \infty$ and $\mathbf{E}Y_i = 0$, then for any $z > 0$:

$$z^2 P(\max_{k \leqslant n} |S_k| \geqslant z) \leqslant \mathbf{Var}(Z_n)$$

where $S_n = \sum_{k=1}^n Y_k$.