Important examples for the oral exams

Chaitanya Ekanadham

1 Real analysis

1.1 Measure theory

1. **Nonmeasurable set:** Take $[0,1)/\mathbb{Q}$ i.e.2 reals are equivalent iff they differ by some rational in $[0,1]$. Construct a set P containing exactly one element from each equivalence class (axiom of choice). By translation invariance we know that $P_i = P + q_i$ has the same measure as P, and one can show the P_i's are disjoint using the def. of P. But now $[0,1)$ is inside the disjoint union of the P_i's. If P were measurable then $\mu([0,1))$ is bounded above $\sum_{j=1}^{\infty} \mu(P)$ which is either 0 or ∞. We know that span of the P_i's cannot cover more than the interval $[-1,2]$ - so the sum must be 0. This is a contradiction.

2. **Uncountable set of measure 0:** The Cantor set (obtained by deleting middle-thirds) has measure 0 since the part that was deleted has measure $\frac{1}{3}\sum(\frac{2}{3})^n = 1$. To see that it is uncountable we relate it to a certain ternary expansion of reals in $[0,1]$. In view of this representation, we see that the middle third $(1/3, 2/3)$ are exactly those reals with '1' as the first digit in their ternary expansion (by convention use 0.0222... as the expansion for 1/3). Similarly, those removed in the second step are exactly those reals with '1' as their second digit and so on. So the Cantor set can be thought of as all the reals in $[0,1]$ who don't have any 1's in their ternary expansion. Using the diagonal process (assume countability, enumerate and then construct a ternary expansion with diagonal elements), it follows that the set is uncountable.

3. **Failure of Egoroff's theorem/BCT for infinite measures:** Take $f_n(x) = 1_{[-n,n]} \to f \equiv 0$. The point is that we cannot use the notion of countable additivity asserting that if a sequence of sets decreases to ϕ, then the measures decreases to 0. On the contrary in this case although $S_n = (\infty, n) \cap (n, \infty) \downarrow \phi$, the measure is always infinite.

1.2 Weird functions

1. **A bounded measurable nowhere-cnts fn:** $f(x) = 1_{x\in\mathbb{Q}}$.

   A cnts monotone fn with derivative 0 a.e.: Take the cantor function on $[0,1]$. Define it by first constructing it on the Cantor set - take any number in this set and take its ternary expansion (it has no 1's) - change all the 2's to 1's and interpret this as a binary expansion. The number corresponding to this binary expansion will be the function value. If we look at any 2 endpoints of a 'middle-third' interval, we see that they must have the same function value. So we can 'fill in' the rest of the values to be just constant lines on these middle-thirds. This function is cnts/differentiable except on the Cantor set, is uniformly but not absolutely continuous.

2. **A differentiable function with a discontinuous derivative:** $f(x) = 1_{x\neq 0}x^2 sin(1/x)$ has derivative $1_{x\neq 0}2xsin(1/x) - cos(1/x)$, which is discnts at 0.

3. **A bounded differentiable fn with unbounded derivative on a compact set:** $f(x) = 1_{n\neq 0}x^2 sin(1/x^2)$ is bounded on $[-1,1]$ but its derivative is unbounded.

4. **An everywhere-cnts nowhere-differentiable fn:** (Exists by Baire category - for any n we can construct a class G_n of cnts 'sawtooth' fns that are dense in the space of cnts fns - Baire category says there is some fn in the countable intersection of G_n's). Consider

5. **A C^∞ 'bump' fn:** $f(x) = e^{-\frac{1}{1-x^2}}1_{[-1,1]}$. This relies on the more basic smooth fn $e^{-1/x}1_{(0,\infty)}$ which has all derivatives 0 at 0.

1.3 Integration and convergence

1. **Uniform cvgnce but not cvgnce in measure (infinite measure):** $f = 1_{[-n,n]}/n$ (this is stupid).

2. **Failure of Fatou without nonnegativity assumption:** Fatou fails for $f_n = -n1_{[1/n,2/n]} \to 0$ a.e.

3. **Domination is not necessary for integrals to converge - uniform integrability is:** Take $f_n = n1_{1/n^2,2/n^2}$. This converges a.e. and in L_1 to 0, but there is no integrable dominating function! However, the $\{f_n\}$ are uniformly integrable - for $\epsilon > 0$, take n_0 large s.t. $1/n^2 < \epsilon$

and then take δ small enough s.t. $\mu(A) < \delta$ implies that the integrals are less than ϵ. This allows us to conclude L_1 convergence.

4. **A.e. cvgnce/measure cvgnce but not L_p cvgnce:** Just need a non-domination effect - take $f_n = n1_{[1/n,2/n]} \to 0$ a.e and in measure

5. **Measure/L_p cvgnce but not a.e. cvgnce:** Take f_n's to be a strip at value 1 that cycles around the unit interval and gets thinner and thinner (but remains to be at height 1).

6. **Failure of Tonelli's theorem without nonnegativity (this is also a failure of Fubini's theorem without product-integrability):** Let $f(x,y) = \frac{x^2-y^2}{x^2+y^2}$ on $[0,1] \times [0,1]$. One can show that integrating $|f|$ on the product measure is ∞, but the iterated integrals are $\pm\frac{\pi}{4}$, respectively.

1.4 L_p spaces

1. **Something in $L_1([0,1]) \backslash L_2([0,1])$:** $f(x) = 1/\sqrt{x}$

2. **Something in $L_2(\mathbb{R}) \backslash L_1(\mathbb{R})$:** $f(x) = \frac{1}{x}1_{[1,\infty)}$

3. **Something in $\cap L_p([0,1]) \backslash L_\infty$:** $f(x) = \log(x)$

4. **Something in $L_\infty * \backslash L_1$:** One existence proof uses the Hahn-Banach theorem. Take an σ-finite partition A_j of the space. Consider the subspace of L_∞ consisting of functions that are some constant value a_j on each A_j. This subspace is isomorphic to the space of bounded sequences l_∞. Inside l_∞ look at the subspace consisting of all convergent sequences. Define a linear functional on this space to be the limit of the sequence. Extend this to l_∞ by Hahn-Banach (it must be sandwiched between $\liminf$ and $\limsup$ of the sequence), then transplant back to L_∞ and extend again by Hahn-Banach. Now if this linear functional was in L_1, then there would be a l_1 sequence b_j obtained by integrating this L_1 fn on each A_j. But this sequence would be in l_∞, and one can show this is impossible.

1.5 Other areas

1. **Failure of Radon-Nikodym without σ-finite assumption:** Consider the counting measure μ that assigns finite sets their cardinality and ∞ otherwise. Then the lebesgue measure $\lambda << \mu$. But if there were a Radon-Nikodym derivative f then we have $0 = \lambda(\{a\}) =$

$\int_{\{a\}} f d\mu = f(a)$, $\forall a \in \mathbb{R}$, which is a contradiction since then that means $\lambda \equiv 0$.

2 Complex analysis

These are mostly examples of how things that hold for complex-analytic functions do not hold for real-analytic functions.

1. **Failure of Cauchy's theorem for non-simply-connected regions:** $f(z) = 1/z$ on $D(0,1)\backslash\{0\}$ has integral $2\pi i$ on $dD(0,1)$.

2. **A bounded nonconstant real-analytic C^∞ fn:** $f = \frac{1}{1+x^2}$

3. **Failure of Weierstrass theorem in $\mathbb{R}$:** Uniform limits of real-analytic functions are not necessarily even smooth - take

4. **Infinitely many zeros converging on the boundary:** $f(z) = sin(\frac{1}{1-z})$ has zeroes at $1 - \frac{1}{n\pi}$, $\forall n$. This converges to the point $(1,0)$ on the boundary of the unit disk.

5. **Failure of Hurwitz theorem in $\mathbb{R}$:** Nowhere-vanishing real-analytic functions can converge to a nonconstant real-analytic function with a 0 - take $f_n(x) = x^2 + 1/n \to x^2$. If these functions were 'complexified', the f_n's would not be nowhere-vanishing (they would have 2 complex roots).

6. **Non-removable singularity for a bounded C^∞ function:** $f(z) = sin(\frac{1}{|z|}$ is bounded around 0 but cannot be extended even continuously to 0.

7. **Failure of Montel's theorem in $\mathbb{R}$:** $f_n(x) = sin(nx)$ does not even have a pointwise-converging subsequence on any compact set around 0.

8. **Standard residue problem:** We can show the identity $\sum_{n=0}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$ by integrating the function $f(z) = \frac{1}{z^2 sin(\pi z}$ along the boundary of a square box centered at the origin. I think the box integrates to 0, so the answer is the negative of the residue at 0 divided by 2 (?).

9. **Some Riemann surfaces:** The idea is to create a surface on which a possibly non-injective analytic function on the regular plane can be

made 1-1 on the surface. Just create a new copy of the plane to account for 'double images.' Typically such non-1-1 functions have a 'fundamental region' which they map to the whole plane. Consider the following examples:

(a) $f(z) = z^n$. The fundamental region is a sector with angle $(2\pi/n)$. The surface is just n sheets (each sheet corresponds to the image of one sector) connected by criss-cross cuts at the positive real-axis. The lower 'lip' of the cut on the top sheet is connect back to the upper 'lip' of the bottom sheet. The branch point is 0.

(b) $f(z) = e^z$. The fundamental region is a horizontal infinite strip of height 2π. The surface is an infinite 'screw' of sheets connected same way as above, except 0 is not on this surface unless we are considering f on the extended plane.

(c) $f(z) = \cos z$. The fundamental region is a vertical semi-infinite strip of width π. Here the sheets are made with cuts along the x-axis from $(-\infty, 1)$ and $(1, \infty)$.

3 ODE

1. **Multiple solutions:** $\dot{y}(t) = |y(t)|^{1/2}$ on $t \in [0,1]$ with $y(0) = 0$. $y(t) \equiv 0$ and $y(t) = \frac{t^2}{4}$ are both solutions. In fact, for any $c \geq 0$, $y_c(t) = 1_{t \leq c}\frac{1}{4}(t-c)^2$ is a solution.

2. **Finite-time blow up (where continuation fails):** $[\dot{y} = y^2]$ with $y(1) = -1$. The solution on $(0, \infty)$ is $y(t) = -t^{-1}$. But clearly $f(t, y) = y^2$ is not bounded on $(t, y) \in (0, 1) \times \mathbb{R}$, so the theorem cannot apply, which makes sense since the solutin blows up at $t = 0$.

3. **ODE for pendulum motion:** $[\frac{d^2}{dt^2}\theta + \omega_0^2(\theta) = 0]$ Solution: $y(t) = A\cos(\omega_0 t + \phi)$

4. **Volterra predator/prey equation - corredict prediction from linearization:** $[\dot{x} = zx - bxy, \dot{y} = -cy + bxy]$ for $a, b, c > 0$ ($x =$ prey, $y =$ predator).If we linearize around 0 we get a saddle, and around $(c/b, a/b$, we get a centre. (Here the linear dynamics correctly predict the nonlinear dynamics).

5. **Incorrect prediction from linearization:** Alter the system $[\dot{x} = y, \dot{y} = -x]$ to $[\dot{x} = y - x\sqrt{x^2+y^2}, \dot{y} = -x - y\sqrt{x^2+y^2}]$. The linearized problem (original system) has a centre at the origin, whereas

the nonlinear system has a focus. The 'centre' is a very precise condition that is unstable wrt nonlinear perturbations.

6. **First integral:** For the ODE $[\frac{d^2}{dt^2}x + f(x) = 0]$ a first integral is $F(x, \dot{x}) = g(x) = \dot{x}^2/2$ where $g'(x) = f(x)$.

7. **A linearly indep.** $A(t)$ **with 0 determinant** Consider $\varphi_1(t) = (t, 0)^T$ and $\varphi_2(t) = (t^2, 0)^T$. The point is that such matrices cannot be fundamental matrices for ODE's, i.e. a matrix is an f.m. for an ODE iff its columns are solutions and they are linearly independent iff the determinant is nonzero at some time τ.

8. **Perturbed constant coeff linear system:** $[x'' + x + \mu x' + x^2 = 0]$ where μ is a 'damping term'. The nonlinear term is just $[x^2, 0]^T$ and the matrix in the linear term has eigenvalues $\frac{-\mu \pm \sqrt{\mu^2 - 4}}{2}$ which is ± 1 for $\mu = 0$.

9. **Floquet exponents and orbital stability:** Consider the orbital stability theorem on the 2D ODE system defined by the equation $[x'' + f(x)x' + g(x) = 0]$, assuming there is some T-periodic solution $\varphi(t)$. The system is autonomous and we know that the sum of the char. exponenents $\lambda_1 + \lambda_2 = -\frac{1}{T}\int_0^T f(\varphi(s))ds$ and $\lambda_1 = 0$. It follows that if $\int_0^T f(\varphi(s))ds > 0$, then $\varphi(t)$ is asymptotically orbitally stable.

10. **Bessel functions:** Consider the ODE $t^2 x''(t) + tx'(t) + (t^2 - \alpha^2)x(t) = 0$. The solutions $x_\alpha(t)$ (for given α) are the same as those of Laplace's PDE $\delta x = 0$ in cylindrical coordinates. For $\alpha \in \mathbb{Z}$, they can be expressed in integral form $x_\alpha(t) = \frac{1}{\pi}\int_0^\pi \cos(\alpha\tau - t\sin\tau)d\tau$. on $[0, \infty)$ they look like oscillating sin/cosine functions with the oscillations decreasing in amplitude at roughly the rate $1/\sqrt{x}$.

4 Probability theory

4.0.1 Weak convergence and limit theorems

1. **Char fn diff at 0 but no finite first moment:** consider the density function $g_b(x) = C(x^2(ln|x|)^b)^{-1}1_{|x|>2}$. Then one can show that for $b \in (0, 1]$, the char fn corresponding to $g_b(x)$ is diff. at 0 but $\mathbb{E}[|X|] = \infty$.

2. **Dependence of weak cvgnce on parameters:** Consider the char fn $\phi(t) = e^{-c|t|^\alpha}$ for i.i.d. r.v.'s X_i. One can show that S_n/n cvgs in probability to 0 if $\alpha \in (1, 2]$, cvgs weakly to the Cauchy dbn for $\alpha = 1$ and does not cvg weakly to any dbn for $\alpha \in (0, 1)$. Furthermore, for $\alpha \in (0, 2]$, we can always choose a β s.t. S_n/n^β cvgs to a nondegenerate dbn (e.g., $\beta = 1/alpha$).

3. **Continuity at 0 is essential for char fns:** Consider the sequence of char fns $\phi_n(t) = (-n|t| + 1)1_{[-1/n,1/n]}$. They converge to $1_{t=0}$, which is not a char fn of any sort.

4. **Weak cvgnce does not imply cvgnce of expected values:** Suppose $\alpha_n \to \alpha$. If f is not bounded it is not necessarily true that $lim_{n\to\infty} \int f(x)d\alpha_n = \int f(x)d\alpha$. Let $C_n = \sum_{j=1}^{n} \frac{1}{j^2}$ and consider the measure $\alpha = C_\infty \sum_{j=1}^{\infty} \frac{1}{j^2}\delta_j$ and the sequence of measures $\alpha_n = (1 - \frac{C_n}{C_\infty})\delta_{-M_n} + \frac{1}{C_\infty}\sum_{j=1}^{n} \frac{1}{j^2}\delta_j$ with M_n chosen s.t. the mean of the dbn's corresp. to α_n is always 1. However, the mean of α is ∞.
   Another example is $\alpha_n = (1 - \frac{1}{n})\delta_1 + \frac{1}{n}\delta_{n^2}$ and $\alpha = \delta 1$. Then the mean of α_n is finite but $\to \infty$ as $n \to \infty$. However, the mean of α is 1.

5. **An important tail event:** One 'tail event' would be $A_k = \{\omega \in \Omega : \limsup_{n\to\infty} \frac{S_n(\omega)}{\sqrt{n}} \geq k\}$ for arbitrarily large k. The 0/1 law tells us that $P(A_k)$ is 1 or 0, but by the CLT it cannot be 0, so it is 1. It follows that the $\limsup \frac{S_n}{\sqrt{n}} = +\infty$ a.s. (and similarly for the lim inf). This is directly related to the law of iterated logarithm which says that we need to divide further by $\log(\log n)$ in order to get a constant lim sup.

6. **Failure of the CLT/Motivation of Lindeberg's condition:**

   (a) Suppose Y_j are independent and $Y_j = \pm 1$ each w.p. $1/2$ and let $X_j = \sigma_j Y_j$ with $\sigma_j = 1/j$. Let $S_n = \sum_{j=1}^{n} X_j$ and let $s_n^2 = \sum_{j=1}^{n} \sigma_j^2$. Then since $s_\infty^2 = \lim_{n\to\infty} s_n^2 < \infty$, by the 1-series theorem, S_n converges a.s. Therefore, $S_n/s_n \to \sum_{j=1}^{\infty} \sigma_j Y_j / s_\infty$ a.s., which is *not* normally distributed (can show this with char fns), so the CLT cannot apply. In general if the sum of the variances is finite, the CLT may not apply.

(b) Even if the sum of the variances is infinite, the CLT can stil fail. Consider $X_j = \pm j$ each w.p. $p_j/2$ and 0 w.p. $1-p_j$ with $p_j = 1/j^2$. Then since $\sum p_j < \infty$, X_j is eventually 0 all the time, i.e. $\sum X_j$ converges a.s. to a finite value. But $s_\infty^2 = \sum Var[X_j] = \infty \Rightarrow S_n/s_n \to 0$ a.s. Intuitively, even though $s_n \to \infty$, the contribution to s_n for large n comes from too large values of X_j with small probabilities. This motivates Lindeberg's condition.

(c) **Uniform samples and Borel Cantelli:** Consider a series of r.v.'s X_n uniformly distributed on the d-dimensinal cube in $\mathbb{R}^n$ with side-length $2n$. What's the probability that $\lim \|X_n(\omega)\|_d = +\infty$? For $d = 1$ it is 0 but for $d > 1$ it is 1. The reason is because $P(An = \{\omega : \|X_n(\omega)\|_d \leq C\}) \approx O(\frac{k_d C^d}{(2n)^d}$ which is summable iff $d > 1$.

7. **The WLLN may still hold when the SLLN does not (infinite first moment)**: We know the Cauchy distribution (density $\frac{C}{(1+x^2)}$, char fn $e^{-|t|}$), has infinite first moment (the char fn is not diff'ble at 0). In particular the sum of n i.i.d. Cauchy r.v.'s divided by n is again has a Cauchy distribution, so the WLLN/SLLN do not hold. However, if we 'dampen' the tails a little bit (make the density $\frac{C}{(1+x^2)\log(1+x^2)}$)) then the first moment is the integral of something of order $O(\frac{1}{xlogx})$ which is not integrable - thus the SLLN cannot hold. However, the char fn of this variant becomes differentiable at 0, allowing the WLLN to hold.

8. **Chernoff bounds/Hoeffding inequality:** Consider the sum S_n of n i.i.d. r.v.'s (X_j). We can derive bounds on $P(S_n \leq f(n))$ by equating it to $P(e^{tS_n} \leq e^{tf(n)}$ and optimizing over $t > 0$. For example, for an SSRW on $\mathbb{Z}$ we get $P(S_n \geq a) \leq e^{\frac{-a^2}{2n}}$ for $a > 0$. and for a binomial B with prob. of success 1/2 we get $P(B \leq (1-\delta)n/2) \leq e^{-\delta^2 n/2}$ for $\delta \in (0,1)$.

4.0.2 Markov chains

1. **HHH vs. HTH:** Consider coin-tossing until the sequence 'HTH' or 'HHH' is attained. What's the probability p that 'HHH' wins? This can be expressed as an MC with 6 states.Writing out the probabilities and solving the recurrence equation for the steady-state probabilities gives $p = \frac{3}{16} + \frac{11}{16}p$, so $p = \frac{3}{5}$.

2. **Walk on $\mathbb{Z}^+$:** Consider the walk along the nonneg. integers (starting at 0), jumping forward w.p. p and backward w.p. $(1-p)$ and always going forward from 0. This 'dominates' a regular RW on $\mathbb{Z}$ with drift p. Now, we can use Chernoff bounds to show that the RW goes to $+\infty$ a.s. if $p > 1/2$. In particular we get the bound: $P(\xi_n \leq a) \leq exp[tC + n((t - p + pe^{-2t}))]$. Choosing $t = \epsilon$ small gives the approx bound $e^{n\epsilon(1-2p)}$, which is summable over n. So we can apply Borel-Cantelli to see that the RW stays above C after sufficient time.

3. **SSRW on $\mathbb{Z}$:** We can show the chain is recurrent by using stirling's formula to estimate $n!$. For $\mathbb{Z}^d$, can use a Fourier transform argument to show recurrence iff $d \leq 2$. Let $p(\xi) : \mathbb{R}^n \to [0,1]$ be the probability of incrementing the walk by ξ. Thus $p(\xi) = \frac{1}{2d}$ for the 2d nearest neighbors of 0 and 0 otherwise.One can show that the corresp. char fn is $\hat{p}(t) = \frac{1}{d}\sum\cos(t_j) \leq 1 - c\sum t_j^2 \leq e^{-c\sum t_j^2}$ for some $c > 0$, so $\hat{p}(t)^{2n} \leq e^{-2nc\sum t_j^2}$. Thus one can estimate $\pi^{(2n)}(0,0) \approx \frac{C}{d^{n/2}}$.

4. **A markov process that does not have the Strong Markov Property:** Consider regular brownian motion but if it starts at 0 it remains 0. This introduces 'discontinuity' wrt the starting point of the process, thus violating the strong markov property (Take the stopping time τ as the hitting time of 0 when starting at a nonzero value). However one can verify this is still a markov process since we need only look at the current state to say anything about where the process might go.

4.0.3 Martingales

1. **'Backwards martingale':**: Consider a sequence of i.i.d. r.v.'s (X_i) with finite first moment. Consider the decreasing sequence of sigma fields $\mathcal{F}_n = \sigma(S_n, S_{n+1}, ...)$. Then $Y_n = \mathbb{E}[X_1|\mathcal{F}_n]$ is a martingale wrt $\mathcal{F}_n$. By symmetry we know that $\mathbb{E}[X_k|\mathcal{F}_n]$ is the same for all $1 \leq k \leq n$. Adding them up and dividing by n gives $Y_n = \mathbb{E}[X_1|\mathcal{F}_n] = \frac{S_n}{n}$ a.s. By the SLLN, we know that the Y_n's converge a.s. to some value (namely $\mathbb{E}[\mathbb{X}_{\mathbb{K}}]$ or in this context $\mathbb{E}[X_1|\sigma(\cap\mathcal{F}_n)]$).

5 Statistics

1. **An unbiased estimator for variance:** If X_i are i.i.d. samples with unknown mean μ and unknown variance σ^2, the estimator $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})$ is an unbiased estimator for σ^2. The proof relies on

the sum-of-squares identity:
$\mathbb{E}[\hat{\sigma}^2] = \dfrac{1}{n-1}[\sum_{j=1}^{n} \mathbb{E}[(X_i - \mu)^2] - n\mathbb{E}[(\overline{X} - \mu)^2]] = \dfrac{1}{n-1}[n\sigma^2 - \sigma^2] = \sigma^2.$

2. **A complete statistic:**

3. **A minimal statistic:**

4. **Failure of regularity conditions for the uniform distribution:** In general, $0 = \frac{d}{d\theta}\displaystyle\int_0^\theta \frac{1}{\theta}dx \neq \int_0^\theta \frac{d}{d\theta}\frac{1}{\theta}dx.$

5. **Failure of Cramer-Rao lower bound for biased estimator:** Take n i.i.d. samples X_i from the $Unif(0,\theta)$ distribution. Let $\hat{\theta} = \max_j X_j$. Then $P(\hat{\theta} \leq x) = \frac{x}{\theta}^n 1_{(0,\theta]}(x)$, so $\hat{\theta}$ has the density $f(x) = nx^{n-1}\theta^{-n}$ for $x \in (0,\theta]$. Therefore, $\hat{\theta}$ has mean $\frac{n\theta}{n+1}$ and variance $\frac{\theta^2}{n(n+2)} = O(\frac{1}{n^2})$.

6. **Confidence intervals:**

   (a) **Mean of n i.i.d normal samples with known variance:** $\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ contains μ w.p. $(1-\alpha)$.

   Show first for $n = 1$. Then interpret $\overline{X}$ as a single sample from $\mathcal{N}(\mu, \frac{\sigma^2}{n})$. **If the population is not normally distributed, this is an approximate C.I. for large n by the CLT.**

   (b) **Mean of n i.i.d normal samples with unknown variance:** $\overline{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ contains μ w.p. $(1-\alpha)$.

   This motivates the t-distribution. We have that $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ is $\mathcal{N}(0,1)$ but replacing σ with an estimate s yields the t-statistic $\frac{\overline{X}-\mu}{s/\sqrt{n}} = \frac{(\overline{X}-\mu)/\sigma}{\sqrt{((n-1)s^2/((n-1)\sigma^2)}} = \frac{\mathcal{N}(\ell,\infty}{\sqrt{\chi^2_{n-1}/(n-1)}} = t_{n-1}$.

   (c) **p-quantiles using order statistics:** $[X_{(i)}, X_{(j)}]$ contains ξ_p w.p. $(1-\alpha)$ where $i = np - z_{\alpha/2}\sqrt{np(1-p)}$ and $j = np + z_{\alpha/2}\sqrt{np(1-p)} + 1$.

   This follows from working backwards to solve for i and j - use the 'bernoulli trial' interpretation of order statistics and the normal approximation to the binomial.

(d) **correlation coefficient ρ for a normal bivariate sample:** $\tanh^{-1} r \pm \frac{z_{\alpha/2}}{\sqrt{n}}$ contains $\tanh^{-1}\rho$ w.p. $(1-\alpha)$.

This follows from applying propagation of errors (using $f(x) = \tanh x$) to the limiting normal distribution of the sample correlation.

7. **Hypothesis testing:**

(a) **For a $\mathcal{N}(\theta, \sigma^2)$ sample X, is $\theta \leq \theta_0$?** If we reject the hypothesis for $X \geq \theta_0 + z_{\alpha/2}\sigma$, then probability of type I error (size) is $(1-\alpha)$.

Work backwards from our condition $\sup\limits_{\theta \leq \theta_0} P_\theta(X \geq k) = \alpha$ we get that the expression inside the sup on the LHS is $1 - \Phi(\frac{k-\theta}{\sigma})$ which is an increasing function of θ and so the sup over $\theta \leq \theta_0$ is attained at $\theta = \theta_0$.

(b) **Is the mean of a normal sample μ_0?** An admissible strategy with size α is to reject the hypothesis if $T^2 = \frac{n(\overline{X}-\mu_0)^2}{s^2} > t_{\alpha/2}$.

This is derived from the likelihood ratio test $\lambda \leq \lambda_0$ (admissible by Neyman Pearson). The likelihood ratio λ is a decreasing function T^2. We can choose the right threshold k for T^2 by working backwards from the size condition α and noting that the positive root of T^2 has a t_{n-1}-dbn.

(c) **Is the dbn P_θ a good fit for the data?** Partition the sample space into $\{A_i\}_{i=1}^k$. Apply the chi-squared test to the probability vector p with $p_i = P_\theta(A_i)$, i.e. if we reject θ if $\hat{S} = \sum_{j=1}^{n} \frac{f_i - np_i)^2}{np_i} \geq \chi^2_{k-1,\alpha}$, then probability of type I error is α (for n large).

The key is to interpret $\hat{S}$ as $\|S_n\|_2^2 = \|\frac{1}{\sqrt{n}} \sum_{j=1}^{n} Y_j\|_2^2$ with $Y_j = P^{-1/2}(X_j - p)$. By the CLT, the inside of the norm brackets has a limiting $\mathcal{N}(0, \Sigma)$ where $\Sigma = I - \sqrt{p}\sqrt{p}^T$. Using an orthogonal transformation $\Sigma^{1/2} W_n = S_n$, expanding as inner products, and taking another orthogonal transformation $V = Q^T W$ with the first row of Q being $\sqrt{p}$, we get the result.

(d) **Is the family of dbns $\mathcal{P} = \{P_\theta\}$ a good fit for the data?** Take the MLE $\hat{\theta}$ and apply the previous test. Then $\hat{S}$ has a limiting

χ^2_{k-2} dbn instead of χ^2_{k-1}. One special case is when for each sample we have 2 attributes (A_i, B_j) w.p. p_{ij}. Then independence would assume that p_{ij} is the product of marginals - the Chi-squared test for independence is $E_{ij} = np_{i*}p_{*j}$. The number of DOF is $(m-1)(n-1)$.

6 Special topics