Math 107, Review for the Final Exam

Final Exam: A cumulative final exam will be held on Friday, April 30 8:30 A.M. - 11:00 A.M.

Honor Code: The Emory Honor Code will be in effect during all exams.

Note: You can bring one legal pad sheet (2-sided) containing formulas. You will be allowed to use calculators on the exams.

Part I, 25%

Probability experiment, sample space, classical and emipirical probability:

\[ P(E) = \frac{|E|}{|S|} \quad P(E) = \frac{f_E}{n}, \]

where \( f_E \) is the frequency of event \( E \) among \( n \) trials;
the complement \( P(\overline{E}) = 1 - P(E) \), mutually exclusive events \( A \cap B = \emptyset \),
addition rule:
\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

independent events, multiplication rule for independent events: \( P(A \cap B) = P(A)P(B) \), conditional probability \( P(A|B) \), general multiplication rule: \( P(A \cap B) = P(A)P(B|A) \), formula for conditional probability
\[ P(B|A) = \frac{P(A \cap B)}{P(A)} \]

Counting rules: multiplication rule \( n_1 \times n_2 \times \cdots \times n_k \); in particular, the number of selections of \( r \) objects from \( n \) objects when order matters and repetitions are allowed (with replacement) is \( n^r \); same thing without repetitions (without replacement) \( \frac{n!}{(n-r)!} \); the number of permutations \( n! \), the number of combinations (no repetitions, no order) is \( \frac{n!}{r!(n-r)!} \).

Part II, 25%

Random variables \( X : S \to R \), probability distribution (tables and graphs), discrete and continuous (countable and measurable); The mean (expected value): \( \mu = E(X) = \sum_X XP(X) \); problems about fair games and lottery wins/gains; the variance
\[ \sigma^2 = \sum_X (X - \mu)^2 P(X) = \sum_X X^2 P(X) - \mu^2, \]

standard deviation: \( \sigma = \sqrt{\sigma^2} \); Chebyshev’s inequality:
\[ P(|X - \mu| > k\sigma) \leq \frac{1}{k^2} \quad P(|X - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}. \]
Distributions:

- binomial: \( n \) identical trials, prob. of success \( p \), \( X = \# \) of successes, 
  \[ E(X) = np, \quad \sigma^2 = npq \]
  \[ P(X = k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}, \quad q = 1 - p \]

- multinomial: \( k \) outcomes of each trial 
  \[ P(X_1 = n_1, \ldots, X_k = n_k) = \frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k} \]
  each \( X_i \) alone is binomial with \( p = p_i \)

- hypergeometric: \( a \) white balls, \( b \) red balls, draw \( n \) without replacement, \( X = \# \) of red balls selected
  \[ P(X = k) = \binom{a}{k} \binom{b}{n-k} \binom{a+b}{n} \]
  \[ E(X) = \frac{a}{a+b} n; \]

- Normal \( X \sim N(\mu, \sigma) \), \( Z = \frac{X-\mu}{\sigma} \sim N(0,1) \) using Table E in two ways: from \( z \) to area, from area to \( z \); 
  \[ P(X \leq a) = p(Z \leq \frac{a-\mu}{\sigma}) \]

Part III, 25%

A random sample: \( X_1, \ldots, X_n \) (iid – independent, identically distributed), 
sample mean \( \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \), \( E(\overline{X}) = \mu \), \( \sigma^2(\overline{X}) = \frac{\sigma^2}{n} \), so \( \sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}} \); if \( X \sim N(\mu, \sigma) \) then \( \overline{X} \sim N(\mu, \sigma/\sqrt{n}) \);

Central Limit Theorem (CLT): For any \( X \), \( \overline{X} \rightarrow N(\mu, \sigma/\sqrt{n}) \) as \( n \rightarrow \infty \), \( n \geq 30 \); when \( X \) is 0-1 (proportions), then \( \sum_i X_i \) is binomial and use the continuity correction (\( \pm \frac{1}{2} \));

Confidence Intervals (C.I), confidence level \((1-\alpha)100\%\), \( \alpha \) – probability that the estimated parameter is outside the C.I.;

- For \( \mu, \sigma \) is known: \( \overline{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \), error (= half of the length) \( E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \),
  \[ n = \left( \frac{z_{\alpha/2}}{E} \right)^2 \]

- For \( \mu, \sigma \) is unknown: \( \overline{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \), where \( s^2 \) is the sample variance and \( t \) is the Student \( t \) distribution with \( n-1 \) degrees of freedom (d.f.);

- For \( p \): \( \hat{p} = Y/n \) (sample proportion), C.I : \( \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \); \( n = \hat{p}\hat{q} \left( \frac{z_{\alpha/2}}{E} \right)^2 \),
  where \( \hat{p} \hat{q} \) come from past, if no previous knowledge then use \( \hat{p} = \hat{q} = 1/2 \);
  if partial knowledge, use the nearest value to 1/2
For $\sigma^2$ or $\sigma$: use $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$, then C.I is 
\[
\left( \frac{(n-1)s^2}{\chi^2_{right}}, \frac{(n-1)s^2}{\chi^2_{left}} \right).
\]

Hypotheses testing: $H_0 : \mu = \mu_0$ versus $H_0 : \mu \neq \mu_0$, $H_0 : \mu > \mu_0$ or $H_0 : \mu < \mu_0$. Two types of error: I with prob. $\alpha = P(\text{reject } H_0|H_0 \text{ true})$, II with prob. $\beta = P(\text{do not reject } H_0|H_0 \text{ false})$.

- $z$ test for $\mu$ (known) $z = \frac{x - \mu_0}{\sigma/\sqrt{n}}$, $n \geq 30$ or population normal
- P-value: $P(Z \geq z)$ (times two if 2-tailed)
- $t$ test for $\mu$ (unknown) $t = \frac{x - \mu_0}{s/\sqrt{n}}$, $n \geq 30$ or population normal
- $z$ test for proportion $p$: $z = \frac{\hat{p} - p_0}{\sqrt{p_0q_0/n}}$
- $\chi^2$ test for $\sigma^2$: $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$, d.f. = $n - 1$
- Relation between C.I and H.T.: $\mu_0$ outside C.I if and only if $H_0$ is rejected against a 2-tailed alternative

Part IV, 25%

Comparing means of two populations: $H_0 : \mu_1 = \mu_2$, $\mu_D = \mu_1 - \mu_2$, $n_1, n_2 \geq 30$ or populations normal

- samples independent, $\sigma_1, \sigma_2$ - known: $z = \frac{x_1 - x_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$; C.I. $x_1 - x_2 \pm z_{\alpha/2}\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$
- samples independent, $\sigma_1, \sigma_2$ - unknown: $t = \frac{xbar_1 - xbar_2}{s_D/\sqrt{n}}$, d.f. = min$(n_1 - 1, n_2 - 1)$, C.I. $x_1 - x_2 \pm t_{\alpha/2}\sqrt{s_1^2/n_1 + s_2^2/n_2}$
- samples dependent: $n_1 = n_2 = n$, $t = \frac{\bar{x}_1 - \bar{x}_2}{s_D/\sqrt{n}}$, where $s_D = \sqrt{\frac{n(\sum D^2 - (\sum D)^2)}{n(n-1)}}$
- proportions $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(1/n_1 + 1/n_2)}}$, where $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$, C.I. $\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}(1/n_1 + 1/n_2)}}$.

Comparing variances of two populations: $H_0 : \sigma_1^2 = \sigma_2^2$, indep. samples, $F = \frac{s_1^2}{s_2^2}$ (larger value on top), a pair of degrees of freedom: $n_1 - 1, n_2 - 1$, even for a 2-tailed just one critical region (to the right):

Nonparametric tests for medians and differences $H_0 : MD = m$
• Sign test for the median: \( X = \min(#+, #-) \) when comparing data with \( m \), for \( n \leq 25 \) use Table J, reject when \( X \leq C.V. \), for \( n \geq 26 \), use 
\[ z = \frac{X+1/2-n/2}{\sqrt{n}/2} \]
• Sign Test for comparison of two dependent samples: \( X_B, X_A; H_0: \) no difference, \( H_1: \) there is difference, \( X_B - X_A \), again take \( X = \min(#+, #-) \)
• Wilcoxon Rank Sum Test for comparison of two independent samples, \( 10 \leq n_1 \leq n_2 \); combine the two samples and rank the data from low to high, let \( R \) be the smaller of the sums of ranks over each sample, 
\[ \mu_R = n_1(n_1 + n_2 + 1)/2, \quad \sigma^2_R = n_1n_2(n_1 + n_2 + 1)/12, \quad z = \frac{R - \mu_R}{\sigma_R} \]
• Wilcoxon Signed-Rank Test for comparison of two dependent samples, \( D = X_B - X_A, w_s = \min(\sum_+ |D|, \sum_- |D|) \), for \( n \geq 29 \) use Table K, reject if \( w_s \leq C.V. \), for \( n \geq 30 \), use 
\[ z = \frac{w_s-n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \]