MAT 377: Probability
Midterm Assessment, Nov. 22 2021

(Four problems, 80 points in total. Time: 1 hour 50 minutes

1. (20pts) Suppose we throw n balls into m boxes at random. Let N be the number of boxes with exactly three balls in them.

   (a) Compute its expectation, EN.

   (b) Find the limit $\lim_{n\to\infty} EN/n$ when $m = \alpha n$ for fixed $\alpha > 0$.

   (c) What does Azuma's inequality say about $|N - EN|$?
   (Remember to verify the hypotheses!)

2. (20pts) Let $X \sim \mathrm{Geo}(p)$ be a random variable with geometric distribution.

   (a) Write down the distribution, mean, and variance of X.

   (b) Prove that $EX^{10} < \infty$ but $Ee^{X^2} = \infty$.

   (c) Prove the *memoryless property*: For any pair of integers $k, \ell \geq 1$,

$$\mathbb{P}\left(X > k + \ell \middle| X > k\right) = \mathbb{P}\left(X > \ell\right) .$$

3. (20pts) On $\Omega = [0, \pi] \times [0, 2\pi]$, consider the probability measure

$$\mathbb{P}(E) = \int_0^{2\pi} \int_0^{\pi} I_E(\phi, \theta) \, \sin\phi \, d\phi d\theta \,, \quad \text{for } E \subset \Omega \,.$$

   Define three random variables on Ω by $X = \sin\phi\cos\theta,\ Y = \sin\phi\sin\theta,\ Z = \cos\phi$.

   (a) Find the probability density p_Z of Z (formula and name).

   (b) Argue that the vector-valued random variable (X, Y, Z) is *uniformly distributed* on the unit sphere $S^2 \subset \mathbb{R}^3$, that is,

$$\mathbb{P}\big((X, Y, Z) \in A\big) \;=\; \frac{\text{Area}\,(A)}{4\pi} \,, \quad \text{for } A \subset S^2 \,.$$

   Conclude that Y and Z have the same distribution as X.

   (c) With as little computation as possible, find the means, variances, and covariances of X, Y, and Z. Please explain! (Use that $X^2 + Y^2 + Z^2 = 1$.)

4. (20pts) The goal in this problem is to adapt the proof of Azuma's inequality so that it yields a bound on the variance of a random variable $f(X_1, \ldots, X_n)$,

   Assume that f satisfies the stability condition for Azuma's inequality, with stability constants $a_1 = \ldots = a_n = a$ (for simplicity). Let $X_1, \ldots, X_n$ be independent.

   For $i = 0, \ldots, n$, and any choice of $x_1, \ldots, x_n$, let

   $$E_i f(x_1, \ldots, x_i) := E(x_1, \ldots, x_i, X_{i+1}, \ldots, X_n) \, .$$

   (Here, $x_1, \ldots, x_n$ are fixed while $X_{i+1}, \ldots, X_n$ are random variables.) Set

   $$Y_i := E_i f(X_1, \ldots, X_n) - E_{i-1}(X_1, \ldots, X_n) \, .$$

   (a) Argue that $Y_1 + \cdots + Y_n = f(X_1, \ldots, X_n) - Ef(X_1, \ldots, X_n)$.

   (b) Prove that $\mathrm{Var}(f(X_1, \ldots, X_n)) \leq na^2$.
   Hint: Show that Y_i and Y_j are uncorrelated for $i \neq j$, and $\mathrm{Var}(Y_i) \leq a^2$.

   (c) *Discuss:* What does Chebyshev's inequality say about $\mathbb{P}(|f - Ef| \geq t)$?
   How does it compare with the conclusion of Azuma's inequality in this case?

   Remark: In Part (b), the best estimate is $\mathrm{Var}(Y_i) \leq \frac{1}{4}a^2$ (you are not asked to show this).