Due: Saturday, 08/08, 4:00 PM
Grace period until Saturday, 08/08, 6:00 PM
Remember to show your work for all problems!

Sundry

Before you start writing your final homework submission, state briefly how you worked on it. Who else did you work with? List names and email addresses. (In case of homework party, you can just describe the group.)

1 Estimating π

Note 19

In this problem, we discuss one way that you could probabilistically estimate π. We'll use a square dartboard of side length 2, and a circular target drawn inscribed in the square dartboard with radius 1. A dart is then thrown uniformly at random in the square. Let p be the probability that the dart lands inside the circle.

(a) What is the value of p?

(b) Suppose we throw n darts uniformly at random in the square. Let $\hat{p}$ be the proportion of darts that land inside the circle. Create an unbiased estimator X for π using $\hat{p}$ (in other words, create a random variable X such that $\mathbb{E}[X] = \pi$).

(c) Using Chebyshev's inequality, find a value of n such that your estimate X is within ε of π with $1-\delta$ confidence. Your answer should be in terms of ε and δ. Note that since we are estimating π, your answer should not have π in it, nor use any bounds on the value of π.

2 Tightness of Inequalities

Note 19

(a) Give an example where Markov's inequality is tight; that is, show that given any fixed $k > 0$, there exists a discrete non-negative random variable X which is **not always zero** such that $\mathbb{P}[X \geq k] = \mathbb{E}[X]/k$.

(b) Give an example where Chebyshev's inequality is tight; that is, show that given any fixed $k \geq 1$, there exists a random variable X such that $\mathbb{P}[|X - \mathbb{E}[X]| \geq k\sigma] = 1/k^2$, where $\sigma^2 = \text{Var}(X)$.

3 Short Answer

Note 20

(a) Let X be uniform on the interval $[0,2]$, and define $Y = 4X^2 + 1$. Find the PDF, CDF, expectation, and variance of Y.

(b) Let X and Y have joint distribution

$$f(x,y) = \begin{cases} cxy + \frac{1}{4} & x \in [1,2] \text{ and } y \in [0,2] \\ 0 & \text{otherwise.} \end{cases}$$

Find the constant c (Hint: remember that the PDF must integrate to 1). Are X and Y independent?

(c) Let $X \sim \text{Exp}(3)$.

(i) Find the probability that $X \in [0,\, 1]$.

(ii) Let $Y = \lfloor X \rfloor$, where the floor operator is defined as: $(\forall x \in [k, k+1))(\lfloor x \rfloor = k)$. For each $k \in \mathbb{N}$, what is the probability that $Y = k$? Write the distribution of Y in terms of one of the famous distributions; provide that distribution's name and parameters.

(d) Let $X_i \sim \text{Exp}(\lambda_i)$ for $i = 1, \ldots, n$ be mutually independent. It is a (very nice) fact that $\min(X_1, \ldots, X_n) \sim \text{Exp}(\mu)$. Find μ.

4 Predictable Gaussians

Note 20

Let Y be the result of a fair coin flip, and X be a normally distributed random variable with parameters dependent on Y. That is, if $Y = 1$, then $X \sim N(\mu_1, \sigma_1^2)$, and if $Y = 0$, then $X \sim N(\mu_0, \sigma_0^2)$.

(a) Sketch the two distributions of X overlaid on the same graph for the following cases:

(i) $\sigma_0^2 = \sigma_1^2, \mu_0 \neq \mu_1$

(ii) $\sigma_0^2 \neq \sigma_1^2, \mu_0 = \mu_1$

(b) Bayes' rule for mixed distributions can be formulated as $\mathbb{P}[Y = 1 \mid X = x] = \frac{\mathbb{P}[Y=1] f_{X|Y=1}(x)}{f_X(x)}$ where Y is a discrete distribution and X is a continuous distribution. Compute $\mathbb{P}[Y = 1 \mid X = x]$, and show that this can be expressed in the form of $\frac{1}{1+e^{\gamma}}$ for some expression γ. (Hint: any value z can be equivalently expressed as $e^{\ln(z)}$)

(c) In the special case where $\sigma_0^2 = \sigma_1^2$ find a simple expression for the value of x where $\mathbb{P}[Y = 1 \mid X = x] = \mathbb{P}[Y = 0 \mid X = x] = 1/2$, and interpret what the expression represents. (Hint: the identity $(a+b)(a-b) = a^2 - b^2$ may be useful)

5 Chebyshev's Inequality vs. Central Limit Theorem

Note 19
Note 20

Let n be a positive integer. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with the following distribution:

$$\mathbb{P}[X_i = -1] = \frac{1}{12}; \qquad \mathbb{P}[X_i = 1] = \frac{9}{12}; \qquad \mathbb{P}[X_i = 2] = \frac{2}{12}.$$

(a) Calculate the expectations and variances of X_1, $\sum_{i=1}^n X_i$, $\sum_{i=1}^n (X_i - \mathbb{E}[X_i])$, and

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}.$$

(b) Use Chebyshev's Inequality to find an upper bound b for $\mathbb{P}[|Z_n| \geq 2]$.

(c) Use b from the previous part to bound $\mathbb{P}[Z_n \geq 2]$ and $\mathbb{P}[Z_n \leq -2]$.

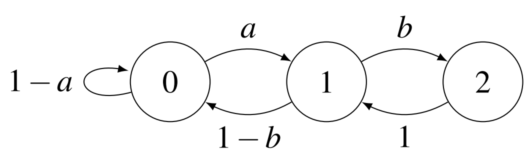
(d) As $n \to \infty$, what is the distribution of Z_n?

(e) We know that if $Z \sim \mathcal{N}(0,1)$, then $\mathbb{P}[|Z| \leq 2] = \Phi(2) - \Phi(-2) \approx 0.9545$. As $n \to \infty$, provide approximations for $\mathbb{P}[Z_n \geq 2]$ and $\mathbb{P}[Z_n \leq -2]$.

6 Analyze a Markov Chain

Note 21

Consider a Markov chain with the state diagram shown below where $a, b \in (0,1)$.



Here, we let $X(n)$ denote the state at time n.

(a) Is this Markov chain irreducible? Is this Markov chain aperiodic? Justify your answers.

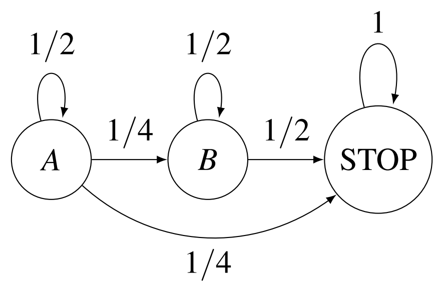
(b) Calculate $\mathbb{P}[X(1) = 1, X(2) = 0, X(3) = 1, X(4) = 2 \mid X(0) = 0]$.

(c) Calculate the invariant distribution. Do all initial distributions converge to this invariant distribution? Justify your answer.

7 A Tiny Language Model

Note 21

A *bigram language model* generates text one token at a time, where the distribution of each new token depends only on the token immediately before it. Consider a tiny model with vocabulary $\{A, B, \text{STOP}\}$. Every sentence begins with the token A, and generation ends when the model emits STOP. The next-token probabilities are shown below.



(a) (i) Briefly explain why the sequence of tokens produced by a bigram model satisfies the Markov property.

(ii) A modern transformer-based language model instead chooses each token based on all tokens generated so far, up to a fixed maximum of w most recent tokens (its *context window*). If we take the states to be single tokens from the vocabulary, is transformer generation a Markov chain? If not, describe a different (much larger, but still finite) state space that would make it one.

(b) What is the probability that the model generates exactly the sentence "ABSTOP"? What about "AASTOP"?

(c) Let L be the total number of tokens in a generated sentence, including the initial A and the final STOP. Use first-step analysis (hitting times) to compute $\mathbb{E}[L]$.

(d) Now suppose we run the model in an endless loop: immediately after emitting STOP, it starts a new sentence, which (as always) begins with the token A. This replaces the self-loop at STOP with a transition to A with probability 1. Show that the resulting Markov chain on $\{A, B, \text{STOP}\}$ is irreducible and aperiodic, and compute its stationary distribution π.

(e) In the long run, what fraction of all emitted tokens are STOP? Using the relationship between stationary probabilities and expected return times, explain how your answer is consistent with your value of $\mathbb{E}[L]$ from part (c).