CS 70 Discrete Mathematics and Probability Theory
Fall 2026 Hug, Sabin

DIS 12

Concentration Inequalities Intro

Markov's Inequality: For any nonnegative random variable X and $t > 0$,

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

Chebyshev's Inequality: For any random variable X and $c > 0$,

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq c] \leq \frac{\text{Var}(X)}{c^2}.$$

Law of Large Numbers: Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and variance σ^2. We have the following:

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu$$

$$\text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{\sigma^2}{n}.$$

Applying Chebyshev's inequality on the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$, we have that

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right] \leq \frac{\sigma^2}{n\varepsilon^2}$$

which means that as $n \to \infty$, the probability of the sample mean deviating from the true mean by any $\varepsilon > 0$ approaches zero.

1 Bounding the Lower Tail

Note 23

A bakery bakes 100 croissants every morning. Let X be the number of croissants sold on a random day, so $0 \leq X \leq 100$, and suppose $\mathbb{E}[X] = 80$ and $\text{Var}(X) = 100$. The owner wants an upper bound on $\mathbb{P}[X \leq 50]$, the probability of a slow day.

(a) Use Markov's inequality to bound $\mathbb{P}[X \leq 50]$. (*Hint:* Why can't you apply Markov's inequality directly to X? Try applying it to $Y = 100 - X$, the number of unsold croissants.)

(b) Use Chebyshev's inequality to bound $\mathbb{P}[X \leq 50]$. Which bound is better, and what information does it use that the other does not?

2 Inequality Practice

Note 23

You roll a die 100 times. Let Z be the sum of the numbers that appear on the die throughout the 100 rolls. Compute $\text{Var}(Z)$. Then use Chebyshev's inequality to bound the probability of the sum Z being greater than 400 or less than 300.

3 How Much Do Students Sleep?

Note 23

You want to estimate μ, the average number of hours Berkeley students sleep per night. You survey n students chosen independently and uniformly at random; let X_i be the hours slept by the i-th student, and let $A_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. You don't know $\text{Var}(X_i) = \sigma^2$ exactly, but you know $\sigma^2 \leq 4$.

For $n = 100$, use Chebyshev's inequality to find a value a such that $[A_{100} - a, A_{100} + a]$ is a 95% confidence interval for μ.

4 Vegas

Note 23

On the planet Vegas, everyone carries a coin. Many people are honest and carry a fair coin (heads on one side and tails on the other), but a fraction p of them cheat and carry a trick coin with heads on both sides. You want to estimate p with the following experiment: you pick a random sample of n people and ask each one to flip their coin. Assume that each person is independently likely to carry a fair or a trick coin.

(a) Let X be the proportion of coin flips which are heads. Find $\mathbb{E}[X]$.

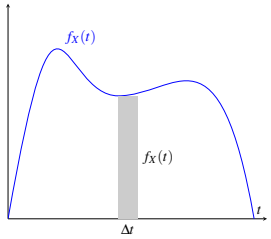
(b) Given the results of your experiment, how should you estimate p? (*Hint:* we're looking for an estimate $\hat{p}$ such that $\mathbb{E}[\hat{p}] = p$. Think about how you could find p by rewriting our expression for $\mathbb{E}[X]$.)

(c) Find the smallest n, the number of people you ask, for which Chebyshev's inequality *guarantees* that your estimate $\hat{p}$ is within 0.05 of the true value of p with probability greater than 0.95 — that is, guarantees that $\mathbb{P}[|\hat{p} - p| \leq 0.05] > 0.95$. (Note that this is not necessarily the smallest n for which the guarantee actually holds; Chebyshev's inequality is only an upper bound on the failure probability.)

Continuous Probability Intro I

In discrete probability, we are only concerned with RVs that take on countably many values; now, in continuous probability, we are interested in RVs that take on *uncountably* many values. Here, the most important difference is that $\mathbb{P}[X = k] = 0$ for any k, but we can have $\mathbb{P}[X \in (a,b)] > 0$.

This motivates the **probability density function (PDF)**, denoted as $f_X(t)$.



Recall from physics that in 1D, density = mass/length; we define *probability density* similarly, as probability/length. So the area of the rectangle, $f_X(t)\Delta t$ (a product of density and length), is a probability. Generalizing this idea to find the area under the curve, we can now find probabilities from the PDF:

$$\mathbb{P}[a \le X \le b] = \int_a^b f_X(t)\,\mathrm{d}t.$$

From here, we define the **cumulative distribution function (CDF)** as

$$F_X(t) = \mathbb{P}[X \le t] = \int_{-\infty}^{t} f_X(u)\,\mathrm{d}u.$$

From the fundamental theorem of calculus, we have $\frac{\mathrm{d}}{\mathrm{d}t}F_X(t) = f_X(t)$; the derivative of the CDF is the PDF.

Properties of PDF and CDF:

- $f_X(t) \ge 0$, and $\int_{-\infty}^{\infty} f_X(t)\,\mathrm{d}t = 1$
- $F_X(t)$ must be non-decreasing, with $\lim_{t\to\infty} F_X(t) = 1$ and $\lim_{t\to-\infty} F_X(t) = 0$

The **joint density** $f(x,y)$ satisfies $\mathbb{P}[a \le X \le b, c \le Y \le d] = \int_a^b \int_c^d f(x,y)\,\mathrm{d}y\,\mathrm{d}x$ and integrates to 1.

Other probability concepts follow naturally from these definitions; the only major difference from discrete probability is that sums turn into integrals, and $\mathbb{P}[X = t]$'s turn into $f_X(t)$'s. For example,

	Discrete	Continuous
Expectation	$\mathbb{E}[X] = \sum_t t \cdot \mathbb{P}[X = t]$	$\mathbb{E}[X] = \int_{-\infty}^{\infty} t \cdot f_X(t)\,\mathrm{d}t$
LOTUS	$\mathbb{E}[g(X)] = \sum_t g(t) \cdot \mathbb{P}[X = t]$	$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(t) \cdot f_X(t)\,\mathrm{d}t$
Total Probability	$\mathbb{P}[A] = \sum_{t=1}^{n} \mathbb{P}[A \mid X = t]\,\mathbb{P}[X = t]$	$\mathbb{P}[A] = \int_{-\infty}^{\infty} \mathbb{P}[A \mid X = t] f_X(t)\,\mathrm{d}t$
Marginal Distribution/Density	$\mathbb{P}[X = x] = \sum_y \mathbb{P}[X = x, Y = y]$	$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y$

Exponential Distribution: $X \sim \text{Exponential}(\lambda)$, the continuous analog to the geometric distribution; it models the amount of time needed to wait until a success, where the rate of success is λ.

$$f_X(t) = \lambda e^{-\lambda t} \qquad F_X(t) = 1 - e^{-\lambda t} \qquad \mathbb{E}[X] = \frac{1}{\lambda} \qquad \text{Var}(X) = \frac{1}{\lambda^2}$$

Similar to the geometric distribution, the exponential distribution also has the memoryless property:

$$\mathbb{P}[X > m + n \mid X > m] = \mathbb{P}[X > n].$$

5 What Does a Density Mean?

Note 24

Let X be uniform on $[0, 1/2]$ and Y be uniform on $[0, 1]$, with X and Y independent.

(a) Write the PDFs $f_X(x)$ and $f_Y(y)$, and the joint density $f_{X,Y}(x, y)$.

(b) For $0 < \varepsilon \leq 1/4$, compute $\mathbb{P}[1/4 \leq X \leq 1/4 + \varepsilon]$ in terms of $f_X(1/4)$ and ε. Using this, what is $\mathbb{P}[X = 1/4]$?

(c) Compute $\mathbb{P}[Y \leq X]$.

6 Max of Uniforms

Note 24

Let $X_1, \ldots, X_n$ be independent Uniform$(0,1)$ random variables (i.e., they are each uniformly distributed on $[0,1]$), and let $X = \max(X_1, \ldots, X_n)$. Compute each of the following in terms of n.

(a) What is the CDF of X?

(b) What is the PDF of X?

(c) What is $\mathbb{E}[X]$?

7 Darts Again

Note 24

Edward and Khalil are playing darts on a circular dartboard.

Edward's throws are uniformly distributed over the entire dartboard, which has a radius of 10 inches. Khalil has good aim (but his throws may land outside of the dartboard); the distance of his throws from the center of the dartboard follows an exponential distribution with parameter $\frac{1}{2}$.

Say that Edward and Khalil both throw one dart at the dartboard. Let X be the distance of Edward's dart from the center, and Y be the distance of Khalil's dart from the center of the dartboard. What is $\mathbb{P}[X < Y]$, the probability that Edward's throw is closer to the center of the board than Khalil's? Leave your answer in terms of an unevaluated integral.

[*Hint:* X is not uniform over $[0, 10]$. Solve for the distribution of X by first computing the CDF of X, $\mathbb{P}[X \leq x]$.]