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?

Solution:

(a) Markov's inequality only bounds $\mathbb{P}[X \geq a]$, the upper tail; hence, it cannot be used to bound $\mathbb{P}[X \leq a]$, the lower tail. Using the hint, we apply it to $Y = 100 - X$. Y is always nonnegative (since $X \leq 100$), and $\mathbb{E}[Y] = 100 - 80 = 20$, so by Markov's inequality,

$$\mathbb{P}[X \leq 50] = \mathbb{P}[Y \geq 50] \leq \frac{\mathbb{E}[Y]}{50} = \frac{20}{50} = \frac{2}{5}.$$

(b) If $X \leq 50$, then $X - 80 \leq -30$, so $|X - 80| \geq 30$. By Chebyshev's inequality,

$$\mathbb{P}[X \leq 50] \leq \mathbb{P}[|X - 80| \geq 30] \leq \frac{\text{Var}(X)}{30^2} = \frac{100}{900} = \frac{1}{9}.$$

Chebyshev's bound of $1/9$ is better than Markov's $2/5$. Chebyshev uses the variance, which tells us how spread out X is around its mean; Markov only uses the mean (and the fact that $Y \geq 0$).

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.

Solution:

Let Z_i be the number on the die for the ith roll, for $i = 1, \ldots, 100$. Then, $Z = \sum_{i=1}^{100} Z_i$. By linearity of expectation, $\mathbb{E}[Z] = \sum_{i=1}^{100} \mathbb{E}[Z_i]$.

$$\mathbb{E}[Z_i] = \sum_{j=1}^{6} j \cdot \mathbb{P}[Z_i = j] = \sum_{j=1}^{6} j \cdot \frac{1}{6} = \frac{1}{6} \cdot \sum_{j=1}^{6} j = \frac{1}{6} \cdot 21 = \frac{7}{2}$$

Then, we have $\mathbb{E}[Z] = 100 \cdot (7/2) = 350$.

$$\mathbb{E}[Z_i^2] = \sum_{j=1}^{6} j^2 \cdot \mathbb{P}[Z_i = j] = \sum_{j=1}^{6} j^2 \cdot \frac{1}{6} = \frac{1}{6} \cdot \sum_{j=1}^{6} j^2 = \frac{1}{6} \cdot 91 = \frac{91}{6}$$

Then, we have

$$\text{Var}(Z_i) = \mathbb{E}[Z_i^2] - \mathbb{E}[Z_i]^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12},$$

Since the Z_is are independent, and therefore uncorrelated, we can add the $\text{Var}(Z_i)$s to get $\text{Var}(Z) = 100(35/12)$.

Finally, we note that we can upper bound $\mathbb{P}[|Z - 350| > 50]$ with $\mathbb{P}[|Z - 350| \geq 50]$.

Putting it all together, we use Chebyshev's to get

$$\mathbb{P}[|Z - 350| > 50] \leq \mathbb{P}[|Z - 350| \geq 50] \leq \frac{100(35/12)}{50^2} = \frac{7}{60}.$$

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.

Solution: Since the X_i are independent, $\text{Var}(A_{100}) = \sigma^2/100 \leq 4/100 = 0.04$, and $\mathbb{E}[A_{100}] = \mu$. By Chebyshev,

$$\mathbb{P}[|A_{100} - \mu| \geq a] \leq \frac{\text{Var}(A_{100})}{a^2} \leq \frac{0.04}{a^2}.$$

Setting $0.04/a^2 = 0.05$ gives $a = \sqrt{0.8} \approx 0.89$ hours. So $[A_{100} - 0.89, A_{100} + 0.89]$ is a 95% confidence interval.

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.)

Solution:

(a) Let X_i be the indicator that the ith person's coin flips heads. Then $X = \frac{1}{n}\sum_{i=1}^{n} X_i$. Applying linearity of expectation, we have

$$\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] = \mathbb{E}[X_i].$$

Since X_i is an indicator, $\mathbb{E}[X_i] = \mathbb{P}[i\text{th person's coin flips heads}]$. The ith person carries a trick coin with probability p (in which case the flip is heads for sure) and a fair coin with probability $1-p$ (in which case the flip is heads with probability $\frac{1}{2}$), so

$$\mathbb{E}[X_i] = \mathbb{P}[i\text{th person's coin flips heads}] = p \cdot 1 + (1-p) \cdot \frac{1}{2} = \frac{1}{2}(p+1).$$

(b) We want to construct an estimate $\hat{p}$ such that $\mathbb{E}[\hat{p}] = p$. Then, if we have a large enough sample, we'd expect to get a good estimate of p.

In our experiment, we measure X, the fraction of people whose coin flips heads. How can we use this observation to construct $\hat{p}$? From part (a), $\mathbb{E}[X] = \frac{1}{2}(p+1)$. First, we isolate p:

$$p = 2\,\mathbb{E}[X] - 1.$$

Then, applying linearity of expectation in reverse,

$$p = 2\,\mathbb{E}[X] - 1 = \mathbb{E}[2X - 1].$$

Thus, our estimator $\hat{p}$ should be $2X - 1$.

(c) We want to find n such that $\mathbb{P}[|\hat{p} - p| \leq 0.05] > 0.95$. Another way to state this is that we want

$$\mathbb{P}[|\hat{p} - p| > 0.05] < 0.05.$$

Notice that $\mathbb{E}[\hat{p}] = p$ by construction, so we can immediately apply Chebyshev's inequality on $\hat{p}$. What we get is:

$$\mathbb{P}[|\hat{p} - p| > 0.05] \leq \mathbb{P}[|\hat{p} - p| \geq 0.05] \leq \frac{\text{Var}[\hat{p}]}{0.05^2}$$

If $\frac{\text{Var}(\hat{p})}{0.05^2} < 0.05$, then we have $\mathbb{P}[|\hat{p} - p| > 0.05] < 0.05$ as desired. So, we want n such that $\text{Var}(\hat{p}) < 0.05^3$. Using properties of variance, as well as the fact that the variance of a sum of independent

random variables is the sum of their variances (the X_i are independent here),

$$\text{Var}(\hat{p}) = \text{Var}(2X-1) = 4\,\text{Var}(X) = \frac{4}{n^2}\,\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{4}{n}\,\text{Var}(X_1).$$

But X_i is an indicator (Bernoulli variable), so its variance is bounded by $\frac{1}{4}$: for a random variable with distribution Bernoulli(q), its variance $q(1-q)$ is maximized at $q = \frac{1}{2}$, yielding a value of $\frac{1}{4}$. Therefore we have

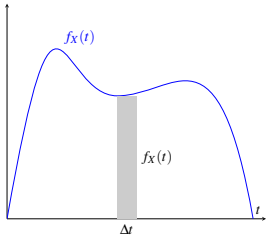
$$\text{Var}[\hat{p}] \leq \frac{4}{n}\frac{1}{4} = \frac{1}{n}.$$

So, we choose n such that $\frac{1}{n} < 0.05^3$, giving $n > \frac{1}{0.05^3} = 8000$. Since n must be an integer, the smallest such value is $n = 8001$.

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 $\mathbb{P}[Y \leq X]$.

Solution:

(a) Each density must be constant on its interval and integrate to 1, so

$$f_X(x) = \begin{cases} 2, & 0 \leq x \leq 1/2, \\ 0, & \text{otherwise}, \end{cases} \qquad f_Y(y) = \begin{cases} 1, & 0 \leq y \leq 1, \\ 0, & \text{otherwise}. \end{cases}$$

Since X and Y are independent, the joint density is the product of the marginals:

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) = \begin{cases} 2, & 0 \leq x \leq 1/2,\ 0 \leq y \leq 1, \\ 0, & \text{otherwise}. \end{cases}$$

(b) For $0 < \varepsilon \leq 1/4$,

$$\mathbb{P}[1/4 \leq X \leq 1/4+\varepsilon] = \int_{1/4}^{1/4+\varepsilon} 2\,dx = 2\varepsilon = f_X(1/4) \cdot \varepsilon.$$

As $\varepsilon \to 0$, this window shrinks to the single point $1/4$, so $\mathbb{P}[X = 1/4] = \lim_{\varepsilon \to 0} 2\varepsilon = 0$.

(c) Integrating the joint density over the region $\{0 \leq y \leq x \leq 1/2\}$,

$$\mathbb{P}[Y \leq X] = \int_0^{1/2} \int_0^x 2\,dy\,dx = \int_0^{1/2} 2x\,dx = \frac{1}{4}.$$

Alternate solution: Note that the region $\{0 \leq y \leq x \leq 1/2\}$ is a right triangle with legs of length $1/2$ and $1/2$, and that the joint density is constant ($= 2$) on this region. Hence, $\mathbb{P}[Y \leq X]$ = joint density $\times$ area $= 2 \cdot \left(\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}\right) = \frac{1}{4}$.

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 $\mathbb{E}[X]$?

Solution:

(a) $\mathbb{P}[X \leq x] = x^n$ for $0 \leq x \leq 1$ (and 0 for $x < 0$, 1 for $x > 1$), since in order for $\max(X_1, \ldots, X_n) < x$, we must have $X_i < x$ for all i. Since they are independent, we can multiply together the probabilities of each of them being less than x, which is x itself, as their distributions are uniform.

(b) The PDF is the derivative of the CDF, so we have $f_X(x) = nx^{n-1}$ for $0 \leq x \leq 1$ and 0 elsewhere.

(c) To find the expectation, we integrate $xf_X(x)$ over all values of x:

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^1 x f_X(x)\, \mathrm{d}x \\
&= \int_0^1 n x^n dx \\
&= \frac{n}{n+1}
\end{aligned}$$

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, $\mathbb{P}[X \leq x]$.]

Solution: We are given that $Y \sim \text{Exponential}(1/2)$. We now find the distribution of X by solving for the CDF of X, $\mathbb{P}[X \leq x]$. To get this, we'll consider the ratio of the area where the distance to the center is at most x, compared to the entire available area. This gives us the following expression:

$$\mathbb{P}[X \leq x] = \frac{\pi x^2}{\pi 10^2} = \frac{x^2}{100}.$$

for $x \in (0, 10)$. For $x < 0$, the CDF is 0, and for $x > 10$, the CDF is 1.

Differentiating gives us the PDF of X, which is given by $f_X(x) = \frac{x}{50}$ for $x \in (0, 10)$, and 0 elsewhere. Now, we solve for $\mathbb{P}[X < Y]$ with total probability:

$$\begin{aligned}
\mathbb{P}[X < Y] &= \int_0^{10} \mathbb{P}[Y > X \mid X = x] f_X(x)\, \mathrm{d}x \\
&= \int_0^{10} \mathbb{P}[Y > x] f_X(x)\, \mathrm{d}x \\
&= \int_0^{10} e^{-0.5x} \frac{x}{50}\, \mathrm{d}x
\end{aligned}$$

$\mathbb{P}[Y > x] = e^{-0.5x}$ comes from the (complement of the) exponential CDF. Evaluating this integral gives us $\mathbb{P}[X < Y] \approx 0.0767$.

Explanation of Integral: The integral may seem a bit confusing, but let's break it down. This is an expression of the total probability rule, where we condition on X. Recall that in discrete, we could calculate $\mathbb{P}[X < Y] = \sum_x \mathbb{P}[X = x]\mathbb{P}[Y > x]$. In continuous, it is pretty much analogous, just with an integral and $f_X(x)\, \mathrm{d}x$ instead of the summation of $\mathbb{P}[X = x]$.

Alternative Calculation of PDF: Another way we could've calculate the PDF of X is by noticing that the PDF corresponds to the likelihood of falling on a point with radius x. The portion of the circle that corresponds to the radius of x is equivalent to its circumference, which is linear with respect to the radius, thus the likelihood (PDF) should be linear with respect to the radius as well.
Thus, we have that $f_X(x) = cx$ for $x \in (0, 10)$, and 0 elsewhere. The PDF must integrate to 1, so $\int_0^{10} cx\, \mathrm{d}x = 50c = 1$, which means that $c = \frac{1}{50}$.

Alternative Setup of Integral: You may have noticed that we chose to condition on X in our setup for total probability. It happens to be that this is the easier way to setup the integral, because the bounds are simpler (since X is bounded by 0 to 10) and $\mathbb{P}[Y > X \mid X = x]$ has a simple, continuous expression. We could've

instead conditioned on Y, but it is more difficult. The interested reader may choose to follow along with the alternate integral setup as follows.

$$\begin{aligned}\mathbb{P}[X<Y] &= \int_0^\infty \mathbb{P}[X<Y \mid Y=y] f_Y(y)\,\mathrm{d}y \\ &= \int_0^{10} \frac{y^2}{100}\cdot 0.5e^{-y/2}\,\mathrm{d}y + \int_{10}^\infty 1\cdot 0.5e^{-y/2}\,\mathrm{d}y\end{aligned}$$

This is because $\mathbb{P}[X<Y \mid Y=y]$ is the CDF of X: $F_X(y)$, which changes expression past $X=10$. Thus, we must split the integral into two parts.