CMPT 210: Probability and Computation

Lecture 14

Sharan Vaswani

June 28, 2022

Collect your Midterm exams

Recap

A **distribution** can be specified by its probability density function (PDF) (denoted by f).

Bernoulli Distribution: If random variable R follows the Bernoulli distribution i.e. $R \sim \text{Ber}(p)$, then $f_p(0) = 1 - p$, $f_p(1) = p$.

Uniform Distribution: If random variable $R : \mathcal{S} \to V$ follows the Uniform distribution i.e. $R \sim \text{Uniform}(1, |V|)$, then for all $v \in V$, $f(v) = 1/|V|$.

Binomial Distribution: If random variable R follows the Binomial distribution i.e. $R \sim \text{Bin}(n, p)$, then $f_{n,p}(k) = \binom{n}{k} p^k (1 - p)^k$.

Geometric Distribution: If random variable R follows the Geometric distribution i.e. $R \sim \text{Geo}(p)$, then $f_p(k) = (1 - p)^{k-1} p$.

Recap

Expectation/mean of a random variable R is denoted by $\mathbb{E}[R]$ and "summarizes" its distribution. Formally, $\mathbb{E}[R] := \sum_{\omega \in \mathcal{S}} \Pr[\omega]\, R[\omega]$

Example: When throwing a standard dice, if R is the random variable equal to the number on the dice. $\mathbb{E}[R] = \sum_{i \in \{1,2,\ldots,6\}} \frac{1}{6}[i] = \frac{7}{2}$.

Alternate definition of expectation: $\mathbb{E}[R] = \sum_{x \in \mathrm{Range}(R)} x \Pr[R = x]$.

This definition does not depend on the sample space.

Example: If $\mathcal{I}_A$ is the indicator random variable for event A, then
$\mathbb{E}[\mathcal{I}_A] = \Pr[\mathcal{I}_A = 1](1) + \Pr[\mathcal{I}_A = 0](0) = \Pr[A]$. For $\mathcal{I}_A$, the expectation is equal to the probability that event A happens.

Linearity of Expectation: For n random variables $R_1, R_2, \ldots, R_n$ and constants $a_1, a_2, \ldots, a_n$, $\mathbb{E}\left[\sum_{i=1}^{n} a_i R_i\right] = \sum_{i=1}^{n} a_i\, \mathbb{E}[R_i]$.

Recap

If $R \sim \text{Bernoulli}(p)$, $\mathbb{E}[R] = p$. *Example*: When tossing a coin, if R is the random variable equal to 1 if we get a heads.

If $R \sim \text{Uniform}(v_1, v_n)$, $\mathbb{E}[R] = \frac{v_1 + v_2 + \ldots + v_n}{n}$. *Example*: When throwing an n-sided dice with numbers $v_1, \ldots v_n$, if R is the random variable equal to the number.

If $R \sim \text{Bin}(n, p)$, $\mathbb{E}[R] = np$. *Example*: When tossing n independent coins, if R is the random variable equal to the number of heads.

If $R \sim \text{Geo}(p)$, $\mathbb{E}[R] = \frac{1}{p}$. *Example*: When tossing a coin repeatedly, if R is the random variable equal to the number of tosses required to get the first heads.

Expectation - Examples

Q: We throw a standard dice, and define a random variable R which is equal to 1 if we get an even number and 0 otherwise. What is the distribution of R? What is $\mathbb{E}[R]$?

Q: We throw 10 independent dice and define R to be the random variable equal to the number of dice that have an even number. What is the distribution of R? What is $\mathbb{E}[R]$?

Q: We repeatedly and independently throw the dice until we get an even number. We define a random variable R equal to the number of throws we need to get an even number. What is the distribution of R? What is $\mathbb{E}[R]$?

Expectation - Examples - Coupon Collector Problem

Q: In a game started by a coffee shop, each time we buy a coffee, we get a coupon. Each coupon has a color (amongst n different colors) and each time, the color of the coupon is selected uniformly at random from amongst the n colors. If we collect at least one coupon of each color, we can claim a free coffee. On average, how many coupons should we collect (coffees we should buy) to claim the prize?

Suppose we get the following sequence of coupons:

$$blue, green, green, red, blue, orange, blue, orange, gray$$

Let us partition this sequence into segments such that a segment ends when we collect a coupon of a new color we did not have before. For this example,

$$\underbrace{blue}_{S_1}\underbrace{green}_{S_2}\underbrace{green, red}_{S_3}\underbrace{blue, orange}_{S_4}\underbrace{blue, orange, gray}_{S_5}$$

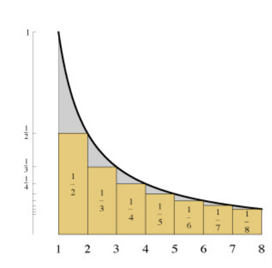
If the number of segments is equal to n, by definition, we will have collected coupons of the n different colors. Define X_k to be the random variable equal to the length of segment S_k and T to be the total number of coupons required to have at least one coupon per color.

Expectation - Examples - Coupon Collector Problem

$T = X_1 + X_2 + \dots X_n$. We wish to compute $\mathbb{E}[T]$. By linearity of expectation, $\mathbb{E}[T] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n]$.

Let us calculate $\mathbb{E}[X_k]$. If we are on stage k, we have seen $k-1$ colors before. Hence, the probability of seeing a new (one that we have not seen before) colored coupon in S_k is $\frac{n-(k-1)}{n}$. $X_k \sim \text{Geo}\left(\frac{n-(k-1)}{n}\right)$, and we know that $\mathbb{E}[X_k] = \frac{n}{n-k+1}$.

$$\begin{aligned}\mathbb{E}[T] &= \sum_{k=1}^{n} \frac{n}{n-k+1} = n\left[\frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{1}\right] \\ &\leq n\left[1 + \int_1^n \frac{dx}{x}\right] = n\left[1 + \ln(n)\right] \leq 2n\ln(n)\end{aligned}$$

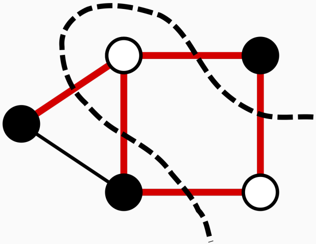


We also know that $\mathbb{E}[T] \geq n\ln(n+1)$. Hence, $\mathbb{E}[T] = O(n\ln(n))$, meaning that we need to buy $O(n\ln(n))$ coffees to collect coupons of n colors and get a free coffee.

Questions?

Max Cut

Given a graph $G = (\mathcal{V}, \mathcal{E})$, partition the graph's vertices into two complementary sets $\mathcal{S}$ and $\mathcal{T}$, such that the number of edges between the set $\mathcal{S}$ and the set $\mathcal{T}$ is as large as possible.



Max Cut has applications to VLSI circuit design.

Equivalently, find a set $\mathcal{U} \subseteq \mathcal{V}$ of vertices that solve the following

$$\max_{\mathcal{U} \subseteq \mathcal{V}} |\delta(\mathcal{U})| \text{ where } \delta(\mathcal{U}) := \{(u, v) \in \mathcal{E} | u \in \mathcal{U} \text{ and } v \notin \mathcal{U}\}$$

Here, $\delta(\mathcal{U})$ is referred to as the "cut" corresponding to the set $\mathcal{U}$.

Max Cut

- Max Cut is NP-hard (Karp, 1972), meaning that there is no polynomial (in $|\mathcal{E}|$) time algorithm that solves Max Cut exactly.
- We want to find an approximate solution $\mathcal{U}$ such that, if OPT is the size of the optimal cut, then, $|\delta(\mathcal{U})| \geq \alpha \, OPT$ where $\alpha \in (0,1)$ is the multiplicative approximation factor.
- Randomized algorithm that guarantees an approximate solution with $\alpha = \frac{1}{2}$ with probability close to 1 (Erdos, 1967).
- Algorithm with $\alpha = 0.878$. (Goemans and Williamson, 1995).
- Under some technical conditions, no efficient algorithm has $\alpha > 0.878$ (Khot et al, 2004).

We will use Erdos' randomized algorithm and first prove the result in expectation. We wish to prove that for $\mathcal{U}$ returned by Erdos' algorithm,

$$\mathbb{E}[|\delta(\mathcal{U})|] \geq \frac{1}{2} OPT$$

. **Algorithm**: Select $\mathcal{U}$ to be a random subset of $\mathcal{V}$ i.e. for each vertex v, choose v to be in the set $\mathcal{U}$ independently with probability $\frac{1}{2}$ (do not even look at the edges!).

Max Cut

Claim: For Erdos' algorithm, $\mathbb{E}[|\delta(\mathcal{U})|] \geq \frac{1}{2}OPT$.

Proof: For each edge $(u, v) \in \mathcal{E}$, let $X_{u,v}$ be the indicator random variable equal to 1 iff the event $E_{u,v} = \{(u, v) \in \delta(\mathcal{U})\}$ happens.

$$\mathbb{E}[|\delta(\mathcal{U})|] = \mathbb{E}\left[\sum_{(u,v)\in\mathcal{E}} X_{u,v}\right] = \sum_{(u,v)\in\mathcal{E}} \mathbb{E}\left[X_{u,v}\right] = \sum_{(u,v)\in\mathcal{E}} \Pr[E_{u,v}]$$

$$\Pr[E_{u,v}] = \Pr[(u, v) \in \delta(\mathcal{U})] = \Pr\left[(u \in \mathcal{U} \cap v \notin \mathcal{U}) \cup (u \notin \mathcal{U} \cap v \in \mathcal{U})\right]$$

$$= \Pr\left[(u \in \mathcal{U} \cap v \notin \mathcal{U})\right] + \Pr\left[(u \notin \mathcal{U} \cap v \in \mathcal{U})\right]$$

$$\Pr[E_{u,v}] = \Pr[u \in \mathcal{U}]\Pr[v \notin \mathcal{U}] + \Pr[u \notin \mathcal{U}]\Pr[v \in \mathcal{U}] = \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{1}{2}.$$

$$\implies \mathbb{E}[|\delta(\mathcal{U})|] = \sum_{(u,v)\in\mathcal{E}} \Pr[E_{u,v}] = \frac{|\mathcal{E}|}{2} \geq \frac{\text{OPT}}{2}.$$

Later in the course, we will prove that $|\delta(\mathcal{U})| \geq \frac{\text{OPT}}{2}$ with probability close to 1.

Questions?

Conditional Expectation

Similar to probabilities, expectations can be conditioned on some event.

For random variable R, the expected value of R conditioned on an event A is given by:

$$\mathbb{E}[R|A] = \sum_{x \in \text{Range}(R)} x \Pr[R = x|A]$$

Q: If we throw a standard dice and define R to be the random variable equal to the number that comes up, what is the expected value of R given that the number is at most 4? A is the event that the number is at most 4. $\Pr[R = 1|A] = \frac{\Pr[(R=1) \cap A]}{\Pr[A]} = \frac{\Pr[R=1]}{\Pr[A]} = \frac{1/6}{4/6} = 1/4$. Similarly, $\Pr[R = 2|A] = \Pr[R = 3|A] = \Pr[R = 4|A] = \frac{1}{4}$ and $\Pr[R = 5|A] = \Pr[R = 6|A] = 0$.

$$\mathbb{E}[R|A] = \sum_{x \in \{1,2,3,4\}} x \Pr[R = x|A] = \frac{1}{4}[1 + 2 + 3 + 4] = \frac{5}{2}.$$

Q: What is the expected value of R given that the number is at least 4?

Law of Total Expectation

If R is a random variable $\mathcal{S} \to V$ and events $A_1, A_2, \ldots A_n$ form a partition of the sample space, then,

$$\mathbb{E}[R] = \sum_i \mathbb{E}[R|A_i] \Pr[A_i]$$

$$\begin{aligned}
\mathbb{E}[R] &= \sum_{x \in \text{Range}(R)} x \Pr[R = x] = \sum_{x \in \text{Range}(R)} x \sum_i \Pr[R = x|A_i] \Pr[A_i] \\
&= \sum_i \Pr[A_i] \sum_{x \in \text{Range}(R)} x \Pr[R = x|A_i] \\
&= \sum_i \Pr[A_i] \, \mathbb{E}[R|A_i].
\end{aligned}$$

Conditional Expectation - Examples

Q: Suppose that 49.6% of the people in the world are male and the rest female. If the expected height of a randomly chosen male is 5 feet 11 inches, while the expected height of a randomly chosen female is 5 feet 5 inches, what is the expected height of a randomly chosen person?

Conditional Expectation - Examples

Recall that if $R \sim \text{Geo}(p)$, $\mathbb{E}[R] = 1/p$. To derive this, we computed the following sum $\mathbb{E}[R] = \sum_{k=1} k\,(1-p)^{k-1} p$. Let's use conditional expectation to do it in a simpler way.

For our coin tossing example, define R to be the random variable equal to the number of coin tosses required to get the first heads. Let A be the event that we get a heads in the first toss. Using the law of total expectation,

$$\mathbb{E}[R] = \mathbb{E}[R|A]\,\Pr[A] + \mathbb{E}[R|A^c]\,\Pr[A^c]$$

We know that, $\mathbb{E}[R|A] = 1$ ($R = 1$ if we get a heads in the first coin toss). $\Pr[A] = p$. Hence,

$$\mathbb{E}[R] = (1)(p) + \mathbb{E}[R|A^c](1-p)$$

$\mathbb{E}[R|A^c]$ is the expected number of tosses required to get the first heads *if* we do not get a heads on the first toss. Hence, $\mathbb{E}[R|A^c] = \mathbb{E}[R] + 1$.

$$\mathbb{E}[R] = (1)(p) + [1 + \mathbb{E}[R]](1-p) \implies \mathbb{E}[R] = 1 + \mathbb{E}[R] - p\mathbb{E}[R] \implies \mathbb{E}[R] = \frac{1}{p}.$$

Questions?

Randomized Quick Select

Given an array A of n distinct numbers, return the k^{th} smallest element in A for $k \in [1, n]$.

Algorithm 1 Randomized Quick Select

function QuickSelect(A, k)
If Length(A) $= 1$, return A[1].
Select $p \in A$ uniformly at random.
Construct sets Left $:= \{x \in A | x < p\}$ and Right $:= \{x \in A | x > p\}$.
$r = |\text{Left}| + 1$ {Element p is the r^{th} smallest element in A.}
if $k = r$ **then**
    return p
end if
if $k < r$ **then**
    QuickSelect(Left, k)
else
    QuickSelect(Right, $k - r$)
end if

Randomized QuickSelect

If $A = \{2, 7, 0, 1, 3\}$ and we wish to find the 2^{nd} smallest element meaning that $k = 2$. According to the algorithm, $p \sim \text{Uniform}(A)$. Say $p = 3$.

Then after step 1, Left $= \{0, 1, 2\}$ and Right $= \{7\}$. $r := |\text{Left}| + 1 = 3 + 1 = 4$. Since $r > k$, we recurse on the left-hand side by calling the algorithm on $\{0, 1, 2\}$ with $k = 2$.

$p \sim \text{Uniform}(\{0, 1, 2\})$. Say $p = 1$. After step 2, Left $= \{0\}$ and Right $= \{2\}$. $r := |\text{Left}| + 1 = 1 + 1 = 2$. Since $r = k$, we terminate the recursion and return $p = 1$ as the second-smallest element in A.

Q: Run the algorithm if $p = 0$ in the first step?

Q: Run the algorithm if $p = 1$ in the first step?

Randomized Quick Select – Analysis

Alternate way: Sort the elements in A and return the k^{th} element in the sorted list. Uses $O(n\log(n))$ comparisons.

Q: Can Randomized Quick Select do better – what is the maximum number of comparisons required by Randomized Quick Select (i) in the worst-case and (ii) in expectation (over the pivot selection)?

Claim: For any array A with n distinct elements, and for any $k \in [n]$, Randomized Quick Select performs fewer than $8n$ comparisons in expectation.

In order to prove this claim, we will need to prove the following Lemma.

Lemma: The child sub-problem's array (either Left or Right) after the partitioning (in Line 4 of the algorithm) has expected size smaller than $\frac{7n}{8}$.

Randomized Quick Select – Analysis

Let us define a "good" event $\mathcal{E}$ that the randomly chosen pivot splits the array roughly in half. Formally, if n is the length of the array, then $\mathcal{E}$ is the event that $r \in \left(\frac{n}{4}, \frac{3n}{4}\right]$ (for simplicity, let us assume that n is divisible by 4.) Since r is chosen randomly, $\Pr[\mathcal{E}] = \frac{3n/4 - n/4}{n} = \frac{1}{2}$.

Recall that $|\text{Left}| = r - 1$ and $|\text{Right}| = n - r$. Hence if event $\mathcal{E}$ happens, then $|\text{Left}| < \frac{3n}{4}$ and $|\text{Right}| < \frac{3n}{4}$. Hence, $|\text{Child}| < \frac{3n}{4}$. If event $\mathcal{E}$ does not happen, in the worst-case, $|\text{Child}| < n$. By using the law of total expectation,

$$\begin{aligned} \mathbb{E}[|\text{Child}|] &= \mathbb{E}[|\text{Child}| \mid \mathcal{E}] \Pr[\mathcal{E}] + \mathbb{E}[|\text{Child}| \mid \mathcal{E}^c] \Pr[\mathcal{E}^c] \\ &< \frac{3n}{4}\frac{1}{2} + (n)\frac{1}{2} = \frac{7n}{8}. \end{aligned}$$

Hence on average, the size of the child sub-problem is smaller than $\frac{7n}{8}$, proving the lemma.

Randomized Quick Select – Analysis

In order to upper-bound the total number of comparisons, we use the Lemma with an induction on n. Recall that we need to prove that Randomized QuickSelect requires fewer than $8n$ comparisons in expectation.

Base case: If $n = 1$, then we require $0 < 8$ comparisons. Hence the base case is satisfied.

Inductive Step:

$$
\begin{aligned}
&\mathbb{E}[\text{Total number of comparisons for size } n \text{ array}] \\
&= \mathbb{E}[(n-1) + \text{Total number of comparisons in child sub-problem}] \\
&= (n-1) + \mathbb{E}[\text{Total number of comparisons in child sub-problem}] && \text{(Linearity of expectation)} \\
&= (n-1) + 8\,\mathbb{E}[|\text{Child}|] && \text{(Induction hypothesis)} \\
&< (n-1) + 8\frac{7n}{8} < 8n. && \text{(Lemma)}
\end{aligned}
$$

Hence we have proved our claim that for any $k \in [n]$, on average, Randomized Quick Select requires fewer than $8n$ comparisons.

Questions?

Independence of random variables

We define two random variables R_1 and R_2 to be independent if for *all* $x_1 \in \text{Range}(R_1)$ and $x_2 \in \text{Range}(R_2)$, events $[R_1 = x_1]$ and $[R_2 = x_2]$ are independent. More formally, we require,

$$\Pr[(R_1 = x_1) \cap (R_2 = x_2)] = \Pr[(R_1 = x_1)] \Pr[(R_2 = x_2)]$$

Q: Suppose we toss three independent, unbiased coins. Let C be r.v. equal to the number of heads that appear and M be the r.v. that is equal to 1 if all the coins match. Are random variables C and M independent?

$\text{Range}(C) = \{0, 1, 2, 3\}$ and $\text{Range}(M) = \{0, 1\}$. $\Pr[C = 3] = \frac{1}{8}$ and $\Pr[M = 1] = \frac{1}{4}$. $\Pr[(C = 3) \cap (M = 1)] = \frac{1}{8} \neq \frac{1}{32} = \Pr[C = 3] \Pr[M = 1]$. Hence, C and M are not independent.

Independence - Examples

Q: If H_1 is the indicator r.v. equal to one if the first toss is a heads, are H_1 and M independent?

Similar to events, random variables $R_1, R_2, \ldots, R_n$ are mutually independent if for all $x_1, x_2, \ldots, x_n$, events $[R_1 = x_1], [R_2 = x_2], \ldots [R_n = x_n]$ are mutually independent.

Independance - Examples

Q: Suppose that the successive daily changes of the price of a given stock are assumed to be independent and identically distributed random variables – for each day i, the PDF is:

$$
\begin{aligned}
\Pr[\text{Daily change on day } i] &= -3 && \text{(With } p = 0.1,\text{)} \\
&= -2 && \text{(With } p = 0.1\text{)} \\
&= -1 && \text{(With } p = 0.2\text{)} \\
&= 0 && \text{(With } p = 0.3\text{)} \\
&= 1 && \text{(With } p = 0.2\text{)} \\
&= 2 && \text{(With } p = 0.1\text{)}
\end{aligned}
$$

If E is the event that the stocks price will increase successively by 1, 2, and 0 points in the next three days, compute $\mathbb{E}[\mathcal{I}_E]$.

Independence of random variables

Q: If R_1 and R_2 are not independent, is $\mathbb{E}[R_1 + R_2] = \mathbb{E}[R_1] + \mathbb{E}[R_2]$?

Q: If R_1 and R_2 are independent, is $\mathbb{E}[R_1 R_2] = \mathbb{E}[R_1]\,\mathbb{E}[R_2]$?

Expectation - Examples

Q: Suppose there is a dinner party where n people check in their coats. The coats are mixed up during dinner, so that afterward each person receives a random coat. In particular, each person gets his own coat with probability $\frac{1}{n}$. What is the expected number of people who get their own coat?

Q: If G_i is the indicator r.v. that person i gets their own coat, are the random variables $G_1, G_2, \ldots G_n$ mutually independent?

Questions?

Joint distribution

For a given experiment, we are often interested not only in the PDFs of individual random variables but also in the relationships between two or more random variables. For example, in we might be interested in the mean time of failure and its connection with different number of components in the system.

A joint distribution between r.v's X and Y can be specified by its joint PDF as follows:

$$\mathrm{PDF}_{X,Y}[x,y] = \Pr[X = x \cap Y = y]$$

If X and Y are independent random variables, $\mathrm{PDF}_{X,Y}[x,y] = \mathrm{PDF}_X[x]\,\mathrm{PDF}_Y[y]$.

If $\mathrm{Range}[X] = \{x_1, x_2, \ldots x_n\}$, $\mathrm{Range}[Y] = \{y_1, y_2, \ldots y_n\}$, then for $x \in \mathrm{Range}(X)$, $[X = x] = [X = x, y = y_1] \cup [X = x, y = y_2] \cup \ldots \cup [X = x, y = y_n] \implies \Pr[X = x] = \Pr[X = x, y = y_1] + \Pr[X = x, y = y_2] + \ldots + \Pr[X = x, y = y_n]$.

$$\mathrm{PDF}_X[x] = \sum_i \mathrm{PDF}_{X,Y}[x, y_i]$$

Hence, we can obtain the distribution for each r.v. from the joint distribution by "marginalizing" over the other r.v's.

Joint distribution - Examples

Q: Suppose that 3 batteries are randomly chosen from a group of 3 new, 4 used but still working, and 5 defective batteries. If we let X and Y denote, respectively, the number of new and used but still working batteries that are chosen, completely specify $\text{PDF}_{X,Y}$. For $i \in [3], j \in [3]$, $\text{PDF}_{X,Y}[i,j] = \Pr[X = i \cap Y = j] = \frac{\binom{3}{i}\binom{4}{j}\binom{5}{3-i-j}}{\binom{12}{3}}$. $\text{PDF}_{X,Y}[0,0] = \frac{\binom{5}{3}}{\binom{12}{3}} = 10/220$,

Table 4.1 $P\{X = i, Y = j\}$.

i \ j	0	1	2	3	Row Sum $= P\{X = i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Column Sums = $P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

$$\text{PDF}_{X,Y}[1,2] = \frac{\binom{3}{1}\binom{4}{2}\binom{5}{2}}{\binom{12}{3}} = 18/220.$$

Questions?