CMPT 210: Probability and Computing

Lecture 18

Sharan Vaswani

November 12, 2024

Recap

- **Expectation**: For a r.v. R, $\mathbb{E}[R] = \sum_{x \in \text{Range}(R)} x \Pr[R = x]$.
- **Conditional Expectation**: For an event A and r.v. R, $\mathbb{E}[R|A] := \sum_{x \in \text{Range}(R)} x \Pr[R = x|A]$.
- **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 i.e. for all i, j, $A_i \cap A_j = \emptyset$ and $A_1 \cup A_2 \cup \ldots \cup A_n = \mathcal{S}$, then, $\mathbb{E}[R] = \sum_i \mathbb{E}[R|A_i] \Pr[A_i]$.
- **Pairwise Independence**: R.v's $R_1, R_2, R_3, \ldots R_n$ are *pairwise* independent iff for *any* pair R_i and R_j, for *all* $x \in \text{Range}(R_i)$ and $y \in \text{Range}(R_j)$, events $[R_i = x]$ and $[R_j = y]$ are pairwise independent implying that $\Pr[(R_i = x) \cap (R_j = y)] = \Pr[R_i = x] \Pr[R_j = y]$.

Independence of random variables

- Similar to events, random variables $R_1, R_2, \ldots, R_n$ are mutually independent if for *all* $x_1 \in \text{Range}(R_1), x_2 \in \text{Range}(R_2), \ldots, x_n \in \text{Range}(R_n)$, events $[R_1 = x_1], [R_2 = x_2], \ldots [R_n = x_n]$ are mutually independent.

Mutual Independence of events: A set of events is said to be mutually independent if the probability of each event in the set is the same no matter which subset of events has occurred. For events E_1, E_2 and E_3 to be mutually independent, all the following equalities should hold:

$$\Pr[E_1 \cap E_2] = \Pr[E_1] \Pr[E_2] \quad \Pr[E_1 \cap E_3] = \Pr[E_1] \Pr[E_3]$$
$$\Pr[E_2 \cap E_3] = \Pr[E_2] \Pr[E_3] \quad \Pr[E_1 \cap E_2 \cap E_3] = \Pr[E_1] \Pr[E_2] \Pr[E_3].$$

Alternatively, (i) $\forall i$ and $j \neq i$, $\Pr[E_i | E_j] = \Pr[E_i]$ and (ii) $\forall i$ and $j, k \neq i$, $\Pr[E_i | E_j \cap E_k] = \Pr[E_i]$.

- For 2 r.v's R_1 and R_2, mutual independence and pairwise independence are equivalent.
- For more than 2 r.v's $R_1, R_2, \ldots, R_n$, mutual independence implies pairwise independence.

Independence - 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 i.e. a person gets their own coat with probability $\frac{1}{n}$. What is the expected number of people who get their own coat?

Let G be the number of people that get their own coat. We wish to compute $\mathbb{E}[G]$. Define G_i to be the indicator r.v. that person i gets their own coat. Observe that $G = G_1 + G_2 + \ldots + G_n$ and by linearity of expectation $\mathbb{E}[G] = \mathbb{E}[G_1] + \mathbb{E}[G_2] + \ldots + \mathbb{E}[G_n]$. For each i, $\mathbb{E}[G_i] = \Pr[G_i] = \frac{1}{n}$. Hence, $\mathbb{E}[G] = 1$ meaning that on average one person will correctly receive their coat.

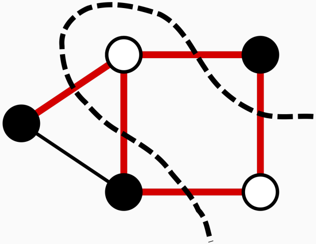
Q: If G_i is the indicator r.v. that person i gets their own coat, are the random variables $G_1, G_2, \ldots G_n$ mutually independent?

No. Since if $G_1 = G_2 = \ldots G_{n-1} = 1$, then,
$\Pr[G_n = 1 | (G_1 = 1 \cap G_2 = 1 \cap \ldots \cap G_{n-1} = 1)] = 1 \neq \frac{1}{n} = \Pr[G_n = 1]$. Conditioning on $(G_1, G_2, \ldots, G_{n-1})$ changes $\Pr[G_n]$, and hence the r.v's are not independent. Notice that we used the linearity of expectation even though these r.v's are not mutually independent.
Q: Are the random variables $G_1, G_2, \ldots G_n$ pairwise independent?

Questions?

Max Cut

- **Aim**: 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.



In this example, the set $\mathcal{U}$ consists of nodes colored black. $|\mathcal{U}| = 3$ and $|\delta(\mathcal{U})| = 5$.

Formal objective: Find a set $\mathcal{U} \subseteq \mathcal{V}$ of vertices that solve the following problem:

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

$\delta(\mathcal{U})$ is the “cut” corresponding to the set $\mathcal{U}$ and the aim is to find the cut with the largest size.

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).
- Complicated 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 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 (u, v) is in the cut, i.e. the event $E_{u,v} = \{(u, v) \in \delta(\mathcal{U})\}$ happens.

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

(Linearity of expectation, and Expectation of indicator r.v's.)

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

$$= \Pr\left[(u \in \mathcal{U} \cap v \notin \mathcal{U})\right] + \Pr\left[(u \notin \mathcal{U} \cap v \in \mathcal{U})\right] \quad \text{(Union rule for mutually exclusive events)}$$

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

(Independent events)

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

Questions?

Randomized Quick Select

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

Algorithm Randomized Quick Select

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

Randomized Quick Select

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

$p \sim \text{Uniform}(\{2, 0, 1\})$. Say $p = 1$. After step 2, $\text{Left} = \{0\}$ and $\text{Right} = \{2\}$.
$r := |\text{Left}| + 1 = 1 + 1 = 2$. 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 in the worst-case?

- In the worst case, Randomized Quick Select is worse than the naive strategy of sorting and returning the k^{th} element. What about the average (over the pivot selection) case?

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.

Randomized Quick Select – Analysis

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 $\frac{7n}{8}$.

Proof: 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 p is chosen uniformly at random, $\Pr[\mathcal{E}] = \frac{3n/4 - n/4}{n} = \frac{1}{2}$.

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,

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

- 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 a strong induction on n. Recall that we need to prove that Randomized Quick Select requires fewer than $8n$ comparisons in expectation.

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

Inductive Step: Assume that for all $m < n$,
$\mathbb{E}[\text{Total number of comparisons for size } m \text{ array}] < 8\,m$.

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

- Hence, for any $k \in [n]$, on average, Randomized Quick Select requires fewer than $8n$ comparisons, even though it might require $O(n^2)$ comparisons in the worst-case.

Questions?

Deviation from the Mean

- We have developed tools to calculate the mean of random variables. Getting a handle on the expectation is useful because it tell us what would happen on average.

- However, summarizing the PDF using the mean is typically not enough. We also want to know how "spread" the distribution is.

Example: Consider three random variables W, Y and Z whose PDF's can be given as:

$$
\begin{aligned}
W &= 0 && \text{(with } p = 1\text{)}\\
Y &= -1 && \text{(with } p = 1/2\text{)}\\
&= +1 && \text{(with } p = 1/2\text{)}\\
Z &= -1000 && \text{(with } p = 1/2\text{)}\\
&= +1000 && \text{(with } p = 1/2\text{)}
\end{aligned}
$$

Though $\mathbb{E}[W] = \mathbb{E}[Y] = \mathbb{E}[Z] = 0$, these distributions are quite different. Z can take values really far away from its expected value, while W can take only one value equal to the mean. Hence, we want to understand how much does a random variable "deviate" from its mean.

Deviation from the Mean

- Before we calculate the deviation of a r.v. from its mean, we need an additional definition.
- For a r.v. $X : \mathcal{S} \to V$ and a function $g : V \to \mathbb{R}$, we define $\mathbb{E}[g(X)]$ as follows:

$$\mathbb{E}[g(X)] := \sum_{x \in \text{Range}(X)} g(x) \Pr[X = x]$$

If $g(x) = x$ for all $x \in \text{Range}(X)$, then $\mathbb{E}[g(X)] = \mathbb{E}[X]$.

Q: For a standard dice, if X is the r.v. corresponding to the number that comes up on the dice, compute $\mathbb{E}[X^2]$ and $(\mathbb{E}[X])^2$

For a standard dice, $X \sim \text{Uniform}(\{1, 2, 3, 4, 5, 6\})$ and hence,

$$\mathbb{E}[X^2] = \sum_{x \in \{1,2,3,4,5,6\}} x^2 \Pr[X = x] = \frac{1}{6}\left[1^2 + 2^2 + \ldots + 6^2\right] = \frac{91}{6}$$

$$(\mathbb{E}[X])^2 = \left(\sum_{x \in \{1,2,3,4,5,6\}} x \Pr[X = x]\right)^2 = \left(\frac{1}{6}[1 + 2 + \ldots + 6]\right)^2 = \frac{49}{4}$$

Variance

Definition: *Variance* is the standard way to measure the deviation of a r.v. from its mean. Formally, for a r.v. X,

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in \mathrm{Range}(X)} (x - \mu)^2 \Pr[X = x] \qquad \text{(where } \mu := \mathbb{E}[X])$$

Intuitively, the variance measures the weighted (by the probability) average of how far (in squared distance) the random variable is from its mean μ.

Q: If $X \sim \mathrm{Ber}(p)$, compute $\mathrm{Var}[X]$.

Since X is a Bernoulli random variable, $X = 1$ with probability p and $X = 0$ with probability $1 - p$. Recall that $\mathbb{E}[X] = \mu = (0)(1-p) + (1)(p) = p$.

$$\mathrm{Var}[X] = \sum_{x \in \{0,1\}} (x - p)^2 \Pr[X = x] = (0-p)^2 \Pr[X = 0] + (1-p)^2 \Pr[X = 1]$$

$$= p^2(1-p) + (1-p)^2 p = p(1-p)[p + 1 - p] = p(1-p).$$

- For a Bernoulli r.v. X, $\mathrm{Var}[X] = p(1-p) \leq \frac{1}{4}$ and the variance is maximum when $p = 1/2$.

Variance

Alternate definition of variance: $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mu^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

$$\begin{aligned}
\textit{Proof}: \mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] &= \sum_{x \in \mathrm{Range}(X)} (x-\mu)^2 \Pr[X = x] \\
&= \sum_{x \in \mathrm{Range}(X)} (x^2 - 2\mu x + \mu^2) \Pr[X = x] \\
&= \sum_{x \in \mathrm{Range}(X)} (x^2 \Pr[X = x]) - (2\mu x \Pr[X = x]) + (\mu^2) \Pr[X = x] \\
&= \sum_{x \in \mathrm{Range}(X)} x^2 \Pr[X = x] - 2\mu \sum_{x \in \mathrm{Range}(X)} x \Pr[X = x] + \mu^2 \sum_{x \in \mathrm{Range}(X)} \Pr[X = x] \\
&\qquad \text{(Since μ is a constant does not depend on the x in the sum.)} \\
&= \mathbb{E}[X^2] - 2\mu\, \mathbb{E}[X] + \mu^2 \sum_{x \in \mathrm{Range}(X)} \Pr[X = x] \quad \text{(Definition of $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$)} \\
&= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 \qquad \text{(Definition of μ)} \\
\implies \mathrm{Var}[X] = \mathbb{E}[X^2] - \mu^2 &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2.
\end{aligned}$$

Back to throwing dice

Q: For a standard dice, if X is the r.v. equal to the number that comes up, compute $\text{Var}[X]$.

Recall that, for a standard dice, $X \sim \text{Uniform}(\{1, 2, 3, 4, 5, 6\})$ and hence,

$$\mathbb{E}[X^2] = \sum_{x \in \{1,2,3,4,5,6\}} x^2 \Pr[X = x] = \frac{1}{6}\left[1^2 + 2^2 + \ldots + 6^2\right] = \frac{91}{6}$$

$$(\mathbb{E}[X])^2 = \left(\sum_{x \in \{1,2,3,4,5,6\}} x \Pr[X = x]\right)^2 = \left(\frac{1}{6}\left[1 + 2 + \ldots + 6\right]\right)^2 = \frac{49}{4}$$

$$\implies \text{Var}[X] = \frac{91}{6} - \frac{49}{4} \approx 2.917$$

Q: If $X \sim \text{Uniform}(\{v_1, v_2, \ldots v_n\})$, compute $\text{Var}[X]$.

$$\mathbb{E}[X] = \sum_{i=1}^{n} v_i \Pr[X = v_i] = \frac{1}{n}\left[v_1 + v_2 + \ldots v_n\right] \quad ; \quad \mathbb{E}[X^2] = \frac{1}{n}\left[v_1^2 + v_2^2 + \ldots v_n^2\right].$$

$$\implies \text{Var}[X] = \frac{\left[v_1^2 + v_2^2 + \ldots v_n^2\right]}{n} - \left(\frac{\left[v_1 + v_2 + \ldots v_n\right]}{n}\right)^2$$

Variance - Examples

Q: Calculate Var[W], Var[Y] and Var[Z] whose PDF's are given as:

$$\begin{aligned}
W &= 0 && \text{(with } p = 1\text{)} \\
Y &= -1 && \text{(with } p = 1/2\text{)} \\
&= +1 && \text{(with } p = 1/2\text{)} \\
Z &= -1000 && \text{(with } p = 1/2\text{)} \\
&= +1000 && \text{(with } p = 1/2\text{)}
\end{aligned}$$

Recall that $\mathbb{E}[W] = \mathbb{E}[Y] = \mathbb{E}[Z] = 0$.

$\text{Var}[W] = \mathbb{E}[W^2] - (\mathbb{E}[W])^2 = \mathbb{E}[W^2] = \sum_{w \in \text{Range}(W)} w^2 \Pr[W = w] = 0^2(1) = 0$. The variance of W is zero because it can only take one value and the r.v. does not "vary".
$\text{Var}[Y] = \mathbb{E}[Y^2] = \sum_{y \in \text{Range}(Y)} y^2 \Pr[Y = y] = (-1)^2(1/2) + (1)^2(1/2) = 1$.

$\text{Var}[Z] = \mathbb{E}[Z^2] = \sum_{z \in \text{Range}(Z)} z^2 \Pr[Z = z] = (-1000)^2(1/2) + (1000)^2(1/2) = 10^6$.

- Hence, the variance can be used to distinguish between r.v.'s that have the same mean.