CS 761: Randomized Algorithms Fall 2019

Lecture 1 — September 06, 2019

Prof. Gautam Kamath Scribe: Frédéric Bouchard, Patrick Li and Pranav Subramani

1 Administrivia

10% of the grade will be assigned for taking scribe notes from the lectures for which scribers will be chosen. The scribe notes have a template that is on the course website that should be followed. There are assignments worth 55% and a final project worth 35%. The final project can be done in one of two ways. The first type of final project is trying to come up with an original idea or solution to a problem. The second is a survey of literature in a field. Please refer to the course webpage for more information about this.

2 Overview of Probability

2.1 Basic Probability Review

Definition 1 (Random Variable)**.** *A function from the set of events, $\mathcal{E}$ to the real numbers is a random variable, $X : \mathcal{E} \to \mathbb{R}$*

A random variable is a function that allows us to take *events* to real numbers which allow us to perform mathematical operations on them.

An example of this is the following: Let X be a random variable and takes values in some space $S \subseteq \mathbb{R}^n$ for $S = \{1, 2, 3, ...\}$. This may be an example of a dice roll, where the possible values of X are $1, 2, 3, \cdots, 6$ and X maps each of those to a real number. In the case of a fair die, that number if $\frac{1}{6}$

Notation: $X \sim D$, standing for X is sampled from distribution D.

Example 2.1. *A fair dice roll where $S = \{1, 2, 3, 4, 5, 6\}$, is written as $X \sim$ Uniform$([6])$, where each outcome is equiprobable.*

Example 2.2. *A random variable X that follows a Bernoulli distribution is written as $X \sim$ Bernouilli(p) which means that $Pr[X = 1] = p$ (i.e. success) and $Pr[X = 0] = 1 - p$ (i.e. failure).*

2.1.1 Probability Mass Function (PMF)

Probability mass functions are defined for discrete distributions, i.e., distributions whose state space is discrete. An example of this is a die.

Definition 2. *The probability mass function of a discrete random variable ($\mathcal{E}$ is discrete) defined as $f_X(x) = Pr[X = x] = Pr[\{e \in \mathcal{E} : X(e) = x\}]$*

Remark 3. *An important note is that the sum over all measurable events must equal to* 1.

Example 2.3. *Consider the example where* $X \sim Bernoulli(p)$*, we have* $f_X(1) = p$ *and* $f_X(0) = 1-p$ *which indicate the probability that the random variable* ***realizes*** *the value* 1 *and* 0 *respectively.*

2.1.2 Cumulative Density Function (CDF)

Definition 4. *The cumulative distribution function of a* $F_X(x) = Pr[X \leq x_j] = \sum_{i \leq j} Pr[X = x_i] = \sum_{i \leq j} f_X(i)$.

2.1.3 Probability Density Function (PDF)

Definition 5. *The probability density function is defined as* $\frac{d}{dx} F_X(x) = f_X(x)$.

It is often viewed as the continuous analogue of the mass function, since the notion of mass functions over continuous sample spaces does not make sense.

Remark 6. *This equation has the following discrete equivalence:* $F_X(x) - F_X(x-1) = f_X(x)$.

2.1.4 Binomial Distribution

The binomial distribution is one of the most common discrete distributions.

Definition 7. $X \sim Binomial(n, p) \iff X = \sum_{i=1}^{n} X_i$ *subject to* $X_i \sim Bernoulli(p)$

The following are properties of the binomial distribution:

1. $X = 0$, with probability $(1-p)^n$.
2. $X = 1$, with probability $p \cdot (1-p)^{n-1}$.

The mass function for the binomial distribution is given as $f_X(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$.

2.1.5 Events

Definition 8. *An event is a set of realizations for random variables.*

Example 2.4. ξ_{even} *which can stand for the event that a fair dice comes up with an even number.*

$$Pr[\xi_{even}] = \frac{1}{2}$$

.

2.1.6 Conditional Probability

Intuitively, it is defined as the probability of an event happening, given another event.

Definition 9. *Given two events ξ_A, ξ_B, $Pr[\xi_A|\xi_B] = \frac{Pr[\xi_A \cap \xi_B]}{Pr[\xi_B]}$*

Example 2.5. *Let's apply conditional probability on a more concrete example. Suppose the event ξ_6 standing for obtaining 6 on a fair roll dice.*

$$Pr[\xi_6] = \frac{1}{6}$$

.

$$Pr[\xi_6|\xi_{even}] = \frac{Pr[\xi_6 \cap \xi_{even}]}{Pr[\xi_{even}]} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

.

2.1.7 Independence

Definition 10. *Two events, ξ_A, ξ_B are called independent $\iff$ $Pr[\xi_A \cap \xi_B] = Pr[\xi_A] \cdot Pr[\xi_B]$*

Remark 11. *An equivalent statement of the above definition is that $Pr[\xi_A|\xi_B] = Pr[\xi_A]$*

Intuitively, the remark means that if ξ_B does not influence the outcome of ξ_A, then conditioning on ξ_B should not change the outcome of ξ_A in any way.

More generally, we can say that $\xi_X \perp \xi_Y \leftrightarrow [X = x] \perp [Y = y] \forall x \in S_x, \forall y \in S_y$.

2.1.8 Expected Value

Definition 12. $E[g(X)] = \sum_{x \in S} g(x) \cdot f_X(x) = \int_{x \in S} g(x) \cdot f_X(x) dx$.

Example 2.6. *Suppose that $X \sim Uniform([6])$.*

$$E[X] = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6} = 3.5$$

Example 2.7. *Suppose that $X \sim Bernoulli(p)$.*

$$E[X] = 0 * (1 - p) + 1 * (p) = p$$

Example 2.8. *Suppose that $X \sim Binomial(n, p)$.*

We introduce linearity of expectations to show an easy way to compute $E[X]$, as direct computation may be tedious.

2.1.9 Linearity of Expectations

*The following result **ALWAYS HOLDS**, even when the random variables are correlated.*

Theorem 13. *Let X be a sum of random variables. If $X = \sum_{i=1}^{n} a_i X_i$, where $a_i \in \mathbb{R}$, and X_i is a random variable, then*

$$E[X] = E[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i E[X_i]$$

To go back to the previous example, suppose that $X \sim Binomial(n, p)$.

By linearity of expectations,

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np$$

2.1.10 Product of Expectations

Due to the convenience of linearity of expectations, we hope that a similar result exists when the random variable XY is the product of random variables X and Y. Unlike linearity of expectations, $E[XY] = E[X] * E[Y] \Leftrightarrow X$ and Y are independent.

Example 2.9. *The following is an example of the identity breaking down for correlated random variables.*

Let $X = Y \sim Bernoulli(\frac{1}{2})$

$$E[X] = \frac{1}{2}, E[Y] = \frac{1}{2}, E[XY] = \frac{1}{2} \neq \frac{1}{4} = E[X] * E[Y]$$

2.1.11 Independence and Correlation

Theorem 14. *If X and Y are independent random variables, then X and Y are uncorrelated.*

Proof.

$$\begin{aligned}
E[XY] &= \sum_{x \in S_x} \sum_{y \in S_y} xy * Pr[X = x \cap Y = y] \\
&= \sum_{x \in S_x} \sum_{y \in S_y} xy * Pr[X = x] Pr[Y = y] \\
&= \sum_{x \in S_x} x * Pr[X = x] \sum_{y \in S_y} y * Pr[Y = y] \\
&= \sum_{x \in S_x} x * Pr[X = x] * E[Y] \\
&= E[Y] * \sum_{x \in S_x} x * Pr[X = x] \\
&= E[X] * E[Y]
\end{aligned}$$

□

This theorem is not an if and only if. The following example shows a pair of uncorrelated random variables that are not independent.

Example 2.10. *Let $X \sim$ Uniform[-1,1], and $Y \sim X^2$*

$$E[X] = \int_{-1}^{1} \frac{1}{2} x f(x) dx = \frac{1}{2} \int_{-1}^{1} x f(x) dx = \frac{1}{2} \left[\frac{x^2}{2} \right]_{-1}^{1} = \frac{1}{4}(1^2 - (-1)^2) = 0$$

$$E[Y] = \int_{-1}^{1} \frac{1}{2} x^2 f(x) dx = \frac{1}{2} \int_{-1}^{1} x^2 f(x) dx = \frac{1}{2} \left[\frac{x^3}{3} \right]_{-1}^{1} = \frac{1}{6}(1^3 - (-1)^3) = \frac{1}{3}$$

$$E[XY] = \int_{-1}^{1} \frac{1}{2} x * x^2 f(x) dx = \frac{1}{2} \int_{-1}^{1} x^3 f(x) dx = \frac{1}{2} \left[\frac{x^4}{4} \right]_{-1}^{1} = \frac{1}{4}(1^4 - (-1)^4) = 0$$

Thus, $E[XY] = E[X]E[Y]$, which implies that X and Y are uncorrelated.

However, $f_Y(1) f_X(0) > 0$. But, $Pr[(X = 0) \cap (Y = 1)] = 0$.

2.1.12 Geometric Distribution

$X \sim \text{Geo}(p)$ denotes that X follows a geometric distribution. This is analogous to treating X as the number of tosses it takes before a weighted coin with probability p of landing heads lands on heads.

$$f_X(k) = (1 - p)^{k-1} p, \forall k \geq 1$$

$$
\begin{aligned}
E[X] &= \sum_{i=0}^{\infty}(1 - Pr[X \leq i]) \\
&= \sum_{i=0}^{\infty} Pr[X > i] \\
&= \sum_{i=1}^{\infty} Pr[X \geq i] \\
&= \sum_{i=1}^{\infty}\sum_{k=i}^{\infty}(1-p)^{k-1}p \\
&= \sum_{i=1}^{\infty}(1-p)^{i-1}\sum_{j=1}(1-p)^{j-i}p \\
&= \sum_{i=1}^{\infty}(1-p)^{i-1} \\
&= \sum_{i=1}^{\infty}(1-p)^{i-1}\frac{p}{p} \\
&= \frac{1}{p}\sum_{i=1}^{\infty}(1-p)^{i-1}p \\
&= \frac{1}{p}
\end{aligned}
$$

Alternatively, $\sum_{i=1}^{\infty}(1-p)^{i-1} = \frac{1}{1-(1-p)} = \frac{1}{p}$.

2.1.13 Dice Problem

Let $\xi_{\text{all even}}$ be the event that all dice rolls in a sequence are even.

Let T be the number of rolls it takes for a dice to roll a 6. What is $E[T|\xi_{\text{all even}}]$?

Conditional probability often leads to unintuitive results. $E[T|\xi_{\text{all even}}]$ is equivalent to finding the expected number of throws until the result is not a 2 or 4. Using the geometric distribution, we realize that $p = Pr[not\{2,4\}] = \frac{2}{3}$. Hence, $E[T|\xi_{\text{all even}}] = \frac{1}{p} = \frac{3}{2}$.

2.2 Quicksort

Algorithm 1 Deterministic Quicksort

procedure QUICKSORT($[x_1, x_2, ..., x_n]$)
  $pivot \leftarrow x_1$
  $\mathbf{S}_{smaller} \leftarrow [\,]$, $\mathbf{S}_{larger} \leftarrow [\,]$
  for `i in 1:n` **do**
    if $x_i \leq pivot$ **then**
      $\mathbf{S}_{smaller}.append(x_i)$
    else
      $\mathbf{S}_{larger}.append(x_i)$
  return[Quicksort($\mathbf{S}_{smaller}$), $pivot$, Quicksort($\mathbf{S}_{larger}$)]

In the Deterministic Quicksort presented in Algorithm 1, the runtime is dictated by arbitrary user input. For example, if a user were to feed the array $\{n, n-1, ..., 2, 1\}$, then there will be a total of $\frac{n*(n-1)}{2} = \Theta(n^2)$ comparisons.

Alternatively, we can select the pivot uniformly at random.

Theorem 15. *Let $y_1, ..., y_n$ be the correctly sorted list, and $X = \sum_{i<j} X_{ij}$ denote the number of comparisons for randomized quicksort where*

$$X_{ij} = \begin{cases} 1 & y_i, y_j \text{ are compared} \\ 0 & otherwise \end{cases}$$

then

$$E[X] = 2n \log n + O(n)$$

Intuition: we usually will separate the sequence fairly evenly ($\frac{n}{2}$, $\frac{n}{2}$), but even when the distribution is ($0.9n$,$0.1n$) which gives respectively a recursive sequence of

$$C(n) = n - 1 + 2C(n/2) \qquad \text{AND} \qquad C(n) = n - 1 + C(0.1n) + C(0.9n)$$

they lead to a run time of $\leq O(n \log n)$.

Proof. To determine the probability that y_i and y_j are compared, we need either y_i or y_j to be chosen as a pivot before any of $\{y_{i+1}, ..., y_{j-1}\}$. If one of the other pivots are chosen, then y_i and y_j are split into two different sets, and will never be compared.This probability equals $\frac{2}{j-i+1}$.

$$
\begin{aligned}
E[X] &= \sum_{i<j} E[X_{ij}] \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} \\
&= \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k} \\
&= \sum_{k=2}^{n-i+1} \sum_{i=1}^{n-1} \frac{2}{k} \\
&= \sum_{k=2}^{n} \sum_{i=1}^{n+1-k} \frac{2}{k} \\
&= \sum_{k=2}^{n} (n+1-k)\frac{2}{k} \\
&= (2n+2)\sum_{k=1}^{n} \frac{1}{k} - 4n
\end{aligned}
$$

Using the harmonic numbers,

$$H_n = \sum_{k=1}^{n} \frac{1}{k} = \Theta(\log(n)) = \log(n) + \gamma + \frac{1}{2n} + O\left(\frac{1}{n^2}\right)$$

Thus,

$$E[X] = (2n+2)H_n - 4n = 2n\log(n) + O(n)$$

□

2.2.1 Coupon Collector

Let $X_i \sim Uniform([n])$. How many samples T until $\cup_{i=1}(X_i) = [n]$?

Let $\mathcal{E}_i^R$ denote when item i is not observed in the first R draws. $Pr(\mathcal{E}_i^R) = (1 - \frac{1}{n})^R$

$$Pr(\text{not done in first R draws}) = Pr[\cup_{i=1}^{n} \mathcal{E}_i^R] \leq \sum_{i=1}^{n} Pr[\mathcal{E}_i^R] = \sum_{i=1}^{n} (1 - \frac{1}{n})^R = n(1 - \frac{1}{n})^R$$

Set $R = \beta n \log n$

$$n(1 - \frac{1}{n})^R \leq n(e^{-\frac{1}{n}})^{\beta n \log n} = ne^{-\beta \log n} = n(e^{\log n})^{-\beta} = n^{-\beta+1}$$

since

$$1 + x \leq e^x, \forall x \in \mathbb{R}$$

Thus,

$$Pr(\text{not done in first R draws}) \leq n^{-\beta+1}$$

Let T be the time to collect all n coupons.

Let t_i = time to collect the i^{th} unique coupon after collecting $i-1$ unique coupons. $t_i \sim$ Geometric(p).

$$p = 1 - \frac{i-1}{n} = \frac{n-i+1}{n}$$

Thus,

$$E[T] = \sum_{i=1}^{n} E[t_i] = \frac{n}{n} + \frac{n}{n-1} + ... + \frac{n}{1} = n(1 + \frac{1}{2} + ... + \frac{1}{n}) = nH_n$$

$$E[T] = n(\log n + \gamma)$$

Note: random variables may not be anywhere near its expectation. Example:

$$X = \begin{cases} n & \text{subject to } Pr[X = n] = \frac{1}{2} \\ -n & \text{subject to } Pr[X = -n] = \frac{1}{2} \end{cases}$$

$E[X] = \frac{1}{2}(n-n) = 0$, but $|X - E[X]| = n$, very large. To further formalize this notion of closeness to expectation, we introduce concentration inequalities.

2.3 Concentration Inequalities

If a random variable is 'nice', it will be near its expected value. For our first inequality 'nice' implies non-negative

Definition 16. ***Markov's Inequality*** *If X is a non-negative random variable, then*

$$Pr[X \geq a] \leq \frac{E[X]}{a}$$

In other words,

$$Pr[X \geq \beta E[X]] \leq \frac{1}{\beta}$$

Proof.

$$\begin{aligned} E[X] &= \int_{x \in S} x f(x) dx \\ &= \int_{x<a} x f(x) dx + \int_{x \geq a} x f(x) dx \\ &\leq 0 + \int_{x \geq a} f(x) dx \\ &= a \int_{x \geq a} f(x) dx \\ &= a Pr[X \geq a] \end{aligned}$$

Thus,

$$\frac{E[X]}{a} \geq Pr[X \geq a]$$

□

This inequality is tight. For example,

$$X = \begin{cases} a & \text{subject to } Pr[X = a] = \frac{t}{a} \\ 0 & \text{subject to } Pr[X = 0] = 1 - \frac{t}{a} \end{cases}$$

Then,

$$E[X] = t, Pr[X \geq a] = \frac{t}{a} = \frac{E[X]}{a}$$

Application of Markov's Inequality on the coupon collector problem

$$Pr[T > \beta n \log n] \leq \frac{1}{\beta}$$

2.3.1 Variance of random variable

The variance can be called the 'centered second moment'

$$Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

Proof.

$$\begin{aligned} Var(X) = E[(X - E[X])^2] &= E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - 2E[XE[X]] + E[E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X^2] \\ &= E[X^2] - E[X]^2 \end{aligned}$$

□