CMPT 210: Probability and Computation

Lecture 11

Sharan Vaswani

June 14, 2022

Recap

Sample (outcome) space $\mathcal{S}$: Nonempty (countable) set of possible outcomes.

Outcome $\omega \in \mathcal{S}$: Possible "thing" that can happen.

Event E: Any subset of the sample space.

Probability function on a sample space $\mathcal{S}$ is a total function $\Pr : \mathcal{S} \to [0, 1]$. For any $\omega \in \mathcal{S}$,

$$0 \leq Pr[\omega] \leq 1 \quad ; \quad \sum_{\omega \in \mathcal{S}} \Pr[\omega] = 1 \quad ; \quad \Pr[E] = \sum_{\omega \in E} \Pr[\omega]$$

Recap

Union: For mutually exclusive events $E_1, E_2, \ldots, E_n$,
$\Pr[E_1 \cup E_2 \cup \ldots E_n] = \Pr[E_1] + \Pr[E_2] + \ldots + \Pr[E_n]$.

Complement rule: $\Pr[E] = 1 - \Pr[E^c]$

Inclusion-Exclusion rule: For any two events E, F, $\Pr[E \cup F] = \Pr[E] + \Pr[F] - \Pr[E \cap F]$.

Union Bound: For any events $E_1, E_2, E_3, \ldots E_n$, $\Pr[E_1 \cup E_2 \cup E_3 \ldots \cup E_n] \leq \sum_{i=1}^{n} \Pr[E_i]$.

Uniform probability space: A probability space is said to be uniform if $\Pr[\omega]$ is the same for every outcome $\omega \in \mathcal{S}$. In this case, $\Pr[E] = \frac{|E|}{|\mathcal{S}|}$.

Recap

Conditional Probability: For events E and F, probability of event E conditioned on F is given by $\Pr[E|F]$ and can be computed as $\Pr[E|F] = \frac{\Pr[E \cap F]}{\Pr[F]}$.

Probability rules with conditioning: For the complement E^c, $\Pr[E^c|F] = 1 - \Pr[E|F]$.

Conditional Probability for multiple events:
$\Pr[E_1 \cap E_2 \cap E_3] = \Pr[E_1] \Pr[E_2|E_1] \Pr[E_3|E_1 \cap E_2]$.

Bayes rule: For events E and F if $\Pr[E] \neq 0$ and $\Pr[F] \neq 0$, $\Pr[F|E] = \frac{\Pr[E|F] \Pr[F]}{\Pr[E]}$.

Law of Total Probability: For events E and F, $\Pr[E] = \Pr[E|F] \Pr[F] + \Pr[E|F^c] \Pr[F^c]$.

Independent Events: Events E and F are said to be independent, if knowledge that F has occurred does not change the probability that E occurs, i.e. $\Pr[E|F] = \Pr[E]$ and $\Pr[E \cap F] = \Pr[E] \Pr[F]$.

Independent Events – Generalization to multiple events

Mutual Independence: 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 of the events has occurred.

For any selection of two or more of the events, the probability that all the selected events occur equals the product of the probabilities of the selected events.

Example: For events E_1, E_2 and E_3 to be mutually independent, all the following equalities should hold:

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

Can generalize this concept to n events – E_1, E_2, ..., E_n are mutually independent, if for *every* subset of events, the probability that all the selected events occur equals the product of the probabilities of the selected events. Formally, for every subset $S \subseteq \{1, 2, \ldots, n\}$, $\Pr[\cap_{i \in S} E_i] = \prod_{i \in S} \Pr[E_i]$.

Mutual independence vs Pairwise independence

Q: Suppose that we flip three fair, mutually-independent coins. Define the following events: E_1 is the event that coin 1 matches coin 2, E_2 is the event that coin 2 matches coin 3 and E_3 is the event that coin 3 matches coin 1. Are E_1, E_2 and E_3 mutually independent?

$\mathcal{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.

$\Pr[E_1] = \Pr[\{HHH, HHT, TTH, TTT\}] \implies \Pr[E_1] = \frac{4}{8} = \frac{1}{2}$. Similarly, $\Pr[E_2] = \Pr[E_3] = \frac{1}{2}$.

$\Pr[E_1 \cap E_2] = \Pr[\{HHH, TTT\}] = \frac{2}{8} = \frac{1}{4}$. Hence, $\Pr[E_1 \cap E_2] = \Pr[E_1] \cdot \Pr[E_2]$. Similarly, $\Pr[E_2 \cap E_3] = \Pr[E_2] \cdot \Pr[E_3]$ and $\Pr[E_1 \cap E_3] = \Pr[E_1] \cdot \Pr[E_3]$.

$\Pr[E_1 \cap E_2 \cap E_3] = \Pr[\{HHH, TTT\}] = \frac{2}{8} = \frac{1}{4} \neq \Pr[E_1]\ \Pr[E_2]\ \Pr[E_3] = \frac{1}{8}$.

Hence, three events (E_1, E_2 and E_3) can be pairwise independent, but not necessarily mutually independent!

Questions?

Random Variables

Definition: A random variable R on a probability space is a total function whose domain is the sample space $\mathcal{S}$.

The codomain is usually a subset of the real numbers.

Example: Suppose we toss three independent, unbiased coins. Let C be the number of heads that appear.

$\mathcal{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

C is a total function that maps each outcome in $\mathcal{S}$ to a number as follows: $C(HHH) = 3$, $C(HHT) = C(HTH) = C(THH) = 2$, $C(HTT) = C(THT) = C(TTH) = 1$, $C(TTT) = 0$.

C is a random variable that counts the number of heads in 3 tosses of the coin.

Back to throwing dice

Q: Suppose we throw two standard dice one after the other. Let us define R to be the random variable equal to the sum of the dice. What is the domain, codomain of R?

Indicator Random Variables: An indicator random variable is a random variable that maps every outcome to either 0 or 1.

Example: In the above dice throwing example, let us define M to be the random variable that is 1 iff both throws of the dice produce a prime number, else it is 0.

$M : \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\} \to \{0,1\}$. $M((2,3)) = 1$, $M((3,6)) = 0$.

Bernoulli random variables: Random variables with the codomain $\{0,1\}$ are called Bernoulli random variables. E.g. the indicator random variable.

Random Variables and Events

An indicator random variable partitions the sample space into those outcomes mapped to 1 and those outcomes mapped to 0.

Example: When throwing two dice, if E is the event that both throws of the dice result in a prime number, then random variable $M = 1$ iff event E happens.

The indicator random variable corresponding to an event E is denoted as $\mathcal{I}_E$, meaning that for $\omega \in E$, $\mathcal{I}_E[\omega] = 1$ and for $\omega \notin E$, $\mathcal{I}[\omega] = 0$.

In the above example, $M = \mathcal{I}_E$ and since $(2,4) \notin E$, $M((2,4)) = 0$ and since $(3,5) \in E$, $M((3,5)) = 1$.

We can define events corresponding to the value that the random variable takes. For example, F is the event that random variable $[R = 3]$, meaning that $F = \{(1,2),(2,1)\}$.

If F is the event that $[\mathcal{I}_E = 1]$, then $F = E$.

Random Variables and Events

In general, a random variable that takes on several values partitions $\mathcal{S}$ into several blocks.

For example, in the coin tossing example, random variable C partitions $\mathcal{S}$ as follows:
$\mathcal{S} = \{\underbrace{HHH}_{C=3}, \underbrace{HHT, HTH, THH}_{C=2}, \underbrace{HTT, , THT, TTH}_{C=1}, \underbrace{TTT}_{C=0}\}$.

Each block is a subset of the sample space and is therefore an event. For example, $[C = 2]$ is the event that the number of heads is two and consists of the outcomes $\{HHT, HTH, THH\}$.

Since it is an event, we can compute its probability i.e.
$\Pr[C = 2] = \Pr[\{HHT, HTH, THH\}] = \Pr[\{HHT\}] + \Pr[\{HTH\}] + \Pr[\{THH\}]$. Since this is a uniform probability space, $\Pr[\omega] = \frac{1}{8}$ for $\omega \in \mathcal{S}$ and hence $\Pr[C = 2] = \frac{3}{8}$.

Q: What is $\Pr[C = 0]$, $\Pr[C = 1]$ and $\Pr[C = 3]$?

Q: What is $\sum_{i=0}^{3} \Pr[C = i]$?

Since a random variable R is a total function that maps every outcome in $\mathcal{S}$ to some value in the codomain, $\sum_{i \in \text{codomain of R}} \Pr[R = i] = \sum_{i \in \text{codomain of R}} \sum_{\omega \text{ s.t. } R[\omega]=i} \Pr[\omega] = \sum_{\omega \in \mathcal{S}} \Pr[\omega] = 1$.

Back to throwing dice

Q: Suppose we throw two standard dice one after the other. Let us define R to be the random variable equal to the sum of the dice. What are the outcomes in the event $[R = 2]$?

Q: What is $\Pr[R = 4]$, $\Pr[R = 9]$?

Q: If M is the indicator random variable equal to 1 iff both throws of the dice produces a prime number, what is $\Pr[M = 1]$?

Random Variables - Example

Q: Suppose that an individual purchases two electronic components, each of which may be either defective or acceptable. In addition, suppose that the four possible results — (d, d), (d, a), (a, d), (a, a) — have respective probabilities 0.09, 0.21, 0.21, 0.49 [where (d, d) means that both components are defective, (d, a) that the first component is defective and the second acceptable, and so on]. If we let X be a random variable that denotes the number of acceptable components obtained in the purchase and E be the event that there was at least one acceptable component in the purchase,

- What is the domain, codomain of X?
- For every i in the codomain of X, compute $\Pr[X = i]$?
- What is the domain, codomain of $\mathcal{I}_E$?
- For every i in the codomain of $\mathcal{I}_E$, compute $\Pr[\mathcal{I}_E = i]$?
- How does X relate to $\mathcal{I}_E$? .

Questions?

Distribution Functions

Probability density function (PDF): Let R be a random variable with codomain V. The probability density function of R is the function $\text{PDF}_R : V \to [0, 1]$, such that $\text{PDF}_R[x] = \Pr[R = x]$ if $x \in \text{Range(R)}$ and equal to zero if $x \notin \text{Range(R)}$.

$\sum_{x \in V} \text{PDF}_R[x] = \sum_{x \in \text{Range(R)}} \Pr[R = x] = 1$.

Cumulative distribution function (CDF): If the codomain is a subset of the real numbers, then the cumulative distribution function is the function $\text{CDF}_R : \mathbb{R} \to [0, 1]$, such that $\text{CDF}_R[x] = \Pr[R \leq x]$.

Importantly, neither PDF_R nor CDF_R involves the sample space of an experiment.

Example: In the coin tossing example, if C counts the number of heads, then $\text{PDF}_C[0] = \Pr[C = 0] = \frac{1}{8}$, and
$\text{CDF}_C[2] = \Pr[C \leq 2] = \Pr[C = 0] + \Pr[C = 1] + \Pr[C = 2] = \frac{7}{8}$.

Q: What is $\text{CDF}_C[5]$? .

For a general random variable R, as $x \to \infty$, $\text{CDF}_R[x] \to 1$ and $x \to -\infty$, $\text{CDF}_R[x] \to 0$.

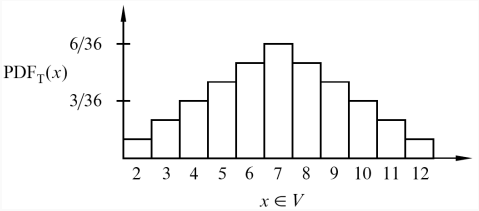
Back to throwing dice

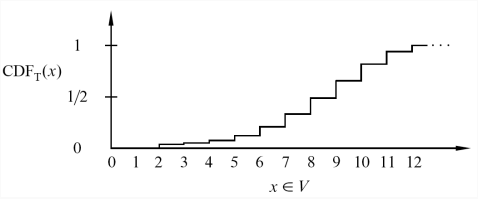
Q: Suppose we throw two standard dice one after the other. Let us define T to be the random variable equal to the sum of the dice. Plot PDF_T and CDF_T

Recall that $T : \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\} \to V$ where $V = \{2,3,4,\ldots 12\}$.

$\text{PDF}_T : V \to [0,1]$ and $\text{CDF}_T : \mathbb{R} \to [0,1]$.

For example, $\text{PDF}_T[4] = \Pr[T = 4] = \frac{3}{36}$ and $\text{PDF}_T[12] = \Pr[T = 12] = \frac{1}{36}$.





Distribution Functions - Example

Q: Suppose we toss three independent, unbiased coins. Let C be the number of heads that appear. What is PDF_C and CDF_C?

Q: What is $\Pr[1 \leq C \leq 3]$?

Q: If E is the event that three tosses have the same result, $\text{PDF}_{\mathcal{I}_E}$ and $\text{CDF}_{\mathcal{I}_E}$?

Distributions

Many random variables turn out to have the same PDF and CDF. In other words, even though R and S are different random variables on different probability spaces, it is often the case that $\text{PDF}_R = \text{PDF}_S$. Hence, by studying the properties of such PDFs, we can study different random variables and experiments.

A **distribution** over a random variable can be fully specified using the cumulative distribution function (CDF) (usually denoted by F). The corresponding probability density function (PDF) is denoted by f.

Common (Discrete) Distributions in Computer Science:

- Bernoulli Distribution
- Uniform Distribution
- Binomial Distribution
- Geometric Distribution

Bernoulli Distribution

We toss a biased coin such that the probability of getting a heads is p. Let R be the random variable such that $R = 0$ when the coin comes up heads and $R = 1$ if the coin comes up tails. R follows the Bernoulli distribution.

The Bernoulli distribution has the PDF $f\colon \{0,1\} \to [0,1]$ meaning that Bernoulli random variables take values in $\{0,1\}$. It can be fully specified by specifying the "probability of success" (of an experiment) p (probability of getting a heads in the example). Formally, PDF_R is given by:

$$f(0) = p \quad ; \quad f(1) = q := 1 - p.$$

In the example, $\Pr[R = 0] = f(0) = p = \Pr[\text{event that we get a heads}]$.

The corresponding CDF_R is given by $F : \mathbb{R} \to [0,1]$:

$$\begin{aligned} F(x) &= 0 && (\text{for } x < 0) \\ &= p && (\text{for } 0 \leq x < 1) \\ &= 1 && (\text{for } x \geq 1) \end{aligned}$$

Uniform Distribution

We roll a standard die. Let R be the random variable equal to the number that shows up on the die. R follows the uniform distribution.

A random variable R that takes on each possible value in its codomain V with the same probability is said to be uniform. The uniform distribution can be fully specified by $|V|$ and has PDF $f : V \to [0, 1]$ such that:

$$f(v) = 1/|V|. \qquad \text{(for all } v \in V\text{)}$$

In the example, $f(1) = f(2) = \ldots = f(6) = \frac{1}{6}$.

For n elements in V arranged in increasing order – $(v_1, v_2, \ldots, v_n)$, then the CDF is:

$$\begin{aligned} F(x) &= 0 && \text{(for } x < v_1\text{)} \\ &= k/n && \text{(for } v_k \leq x < v_{k+1}\text{)} \\ &= 1 && \text{(for } x \geq v_n\text{)} \end{aligned}$$

Q: If X has a Bernoulli distribution, when is X also uniform?

Questions?

Number Guessing Game

We saw an application of the Bernoulli distribution in Frievald's algorithm where we sampled each entry of x (the "probe" vector we multiplied the matrices by) according to a Bernoulli distribution with $p = 1/2$. Let us now study another randomized algorithm and use the uniform distribution.

We have two envelopes. Each contains a distinct number in $\{0, 1, 2, \ldots, 100\}$. To win the game, we must determine which envelope contains the larger number. We are allowed to peek at the number in one envelope selected at random. Can we devise a winning strategy?

Strategy 1: We pick an envelope at random and guess that it contains the larger number (without even peeking at the number). This strategy wins only 50% of the time.

Strategy 2: We peek at the number and if its below 50, we choose the other envelope.

But the numbers in the envelopes may not be random! The numbers are chosen "adversarially" in a way that will defeat our guessing strategy. For example, to "beat" Strategy 2, the two numbers can always be chosen to be below 50.

Q: Can we do better than 50% chance of winning?

Number Guessing Game

Suppose that we somehow knew a number x that was in between the numbers in the envelopes. If we peek in one envelope and see a number. If it is bigger than x, we know its the higher number and choose that envelope. If it is smaller than x, we know that is the smaller number and choose the other envelope.

Of course, we do not know such a number x. But we can guess it!

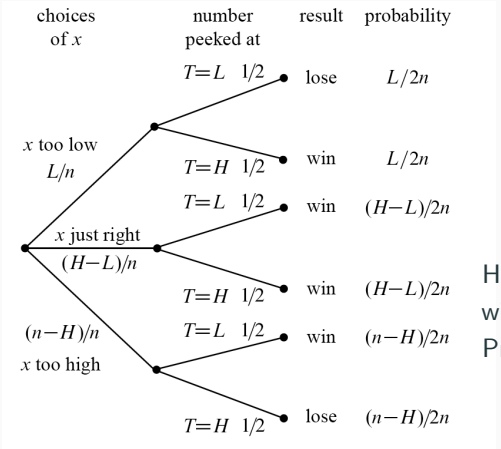
Strategy 3: Choose a random number x from $\{0.5, 1.5, 2.5, \ldots n - 1/2\}$ according to the uniform distribution. $\Pr[x = 3.5] = 1/n$. Then we peek at the number (denoted by T) in one envelope, and if $T > x$, we choose that envelope, else we choose the other envelope.

The advantage of such a randomized strategy is that the adversary cannot easily "adapt" to it.

Q: But does it have better than 50% chance of winning?

Number Guessing Game

Let the numbers in the two envelopes be L (lower number) and H (the higher number). Let us construct a tree diagram.



$$\begin{aligned}\Pr[\text{win}] &= \frac{L}{2n} + \frac{H-L}{2n} + \frac{H-L}{2n} + \frac{n-H}{2n} \\ &= \frac{1}{2} + \frac{H-L}{2n} \geq \frac{1}{2} + \frac{1}{2n} \geq \frac{1}{2}\end{aligned}$$

Hence our strategy has a greater than 50% chance of winning! If $n = 10$, the $\Pr[\text{win}] = 0.55$, if $n = 100$ then $\Pr[\text{win}] = 0.505$.

Questions?

Binomial Distribution

We toss a n biased coin independently. The probability of getting a heads for each coin is p. Let R be the random variable equal to the number of heads we get in the n tosses of the coin. R follows the Binomial distribution.

$\mathcal{S} = \{0, 1, 2, \ldots, n\}$. Hence the PDF_R is a function $f : \{0, 1, 2, \ldots, n\} \to [0, 1]$.

Let E_k be the event that we get k heads in n. Let A_i be the event that we get a heads in toss i.

$$E_k = (A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup (A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup \ldots$$

$$\Pr[E_k] = \Pr[(A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c)] + \Pr[A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap \ldots \cap] + \ldots$$

$$= \Pr[A_1] \Pr[A_2] \Pr[A_k] \Pr[A_{k+1}^c] \Pr[A_{k+2}^c] \ldots \Pr[A_n^c] + \ldots$$

$$\implies \Pr[E_k] = \binom{n}{k} p^k (1-p)^{n-k}$$

Sanity check: $E_0 \cup E_1 \cup E_2 \cup \ldots \cup E_n = \mathcal{S}$. Since $E_0, E_1, E_2, \ldots, E_n$ are mutually exclusive,

$$\sum_{i=0}^{n} \Pr[E_i] = \sum_{i=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1. \qquad \text{(Binomial Theorem)}$$

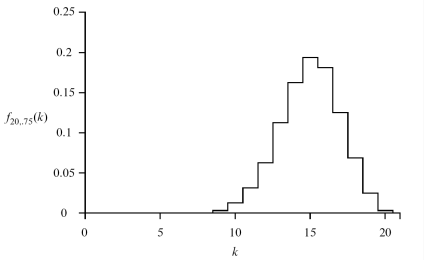
Binomial Distribution

The binomial distribution can be fully specified by n, p and has PDF $f : \{0, 1, \ldots n\} \to [0, 1]$:

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

The corresponding CDF is given by $F : \mathbb{R} \to [0, 1]$:

$$\begin{aligned} F(x) &= 0 && \text{(for } x < 0\text{)} \\ &= \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} && \text{(for } k \le x < k+1\text{)} \\ &= 1. && \text{(for } x \ge n\text{)} \end{aligned}$$



Q: If X has a Bernoulli distribution with parameter p, does it also follow the Binomial distribution? With what parameters?

Geometric Distribution

We toss a biased coin independently multiple times. The probability of getting a heads for each coin is p. Let R be the random variable equal to the number of tosses needed to get the first heads. R follows the geometric distribution.

$\mathcal{S} = \mathbb{N} = \{1, 2, \ldots, \}$. Hence the PDF_R is a function $f : \mathbb{N} \to [0, 1]$.

Let E_k be the event that we need k tosses to get the first heads. Let A_i be the event that we get a heads in toss i.

$$
\begin{aligned}
E_k &= A_1^c \cap A_2^c \cap \ldots \cap A_k \\
\Pr[E_k] &= \Pr[A_1^c \cap A_2^c \cap \ldots \cap A_k] = \Pr[A_1^c] \Pr[A_2^c] \ldots \Pr[A_k] \\
\implies \Pr[E_k] &= (1-p)^{k-1} p
\end{aligned}
$$

Sanity check: $\mathcal{S} = E_1 \cup E_2 \ldots E_\infty$. Since $E_1, E_2, \ldots, E_n$ are mutually exclusive,

$$
\sum_{i=1}^{\infty} \Pr[E_i] = \sum_{i=1}^{\infty} (1-p)^{k-1} p = \frac{p}{1-(1-p)} = 1. \qquad \text{(Sum of geometric series)}
$$

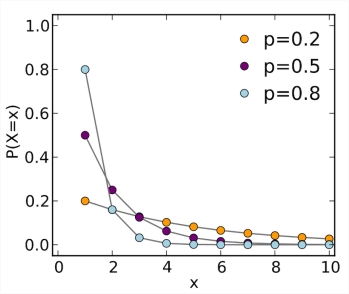
Geometric Distribution

The geometric distribution can be fully specified by p and has PDF $f : \{0, 1, \dots n\} \to [0, 1]$:

$$f(k) = (1-p)^{k-1}p.$$

The corresponding CDF is given by $F : \mathbb{R} \to [0, 1]$:

$$\begin{aligned} F(x) &= 0 && \text{(for } x < 1\text{)} \\ &= \sum_{i=0}^{k} (1-p)^{i-1}p && \text{(for } k \leq x < k+1\text{)} \end{aligned}$$



Q: We throw our standard dice multiple times until we get a 6. What is the probability that I get a 6 on the 4^{th} trial?