CMPT 210: Probability and Computing

Lecture 13

Sharan Vaswani

October 24, 2024

Recap

Random variable: A random "variable" R on a probability space is a total function whose domain is the sample space $\mathcal{S}$. The codomain is denoted by V (usually a subset of the real numbers), meaning that $R : \mathcal{S} \to V$.

Example: Suppose we toss three independent, unbiased coins. In this case, $\mathcal{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$. C is a random variable equal to the number of heads that appear such that $C : \mathcal{S} \to \{0, 1, 2, 3\}$. $C(HHT) = 2$.

Random Variables and Events

Indicator Random Variable: An indicator random variable maps every outcome to either 0 or 1.

Example: Suppose we throw two standard dice, and define M to be the random variable that is 1 iff both throws of the dice produce a prime number, else it is 0.

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

• 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, else $M = 0$.

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

Random Variables and Events

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

Example: When we toss a coin three times, and define C to be the r.v. that counts the number of heads, 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, $\sum_{i \in \text{Range of R}} \Pr[R = i] = \sum_{i \in \text{Range of R}} \sum_{\omega \text{ s.t. } R(\omega)=i} \Pr[\omega] = \sum_{\omega \in \mathcal{S}} \Pr[\omega] = 1$.

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 $\Pr[M = 1]$?

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 $\text{CDF}_R[x] = \Pr[R \le x]$.

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

Example: If we flip three coins, and C counts the number of heads, then
$\text{PDF}_C[0] = \Pr[C = 0] = \frac{1}{8}$, and
$\text{CDF}_C[2.3] = \Pr[C \le 2.3] = \Pr[C = 0] + \Pr[C = 1] + \Pr[C = 2] = \frac{7}{8}$.

Q: What is $\text{CDF}_C[5.8]$? .

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

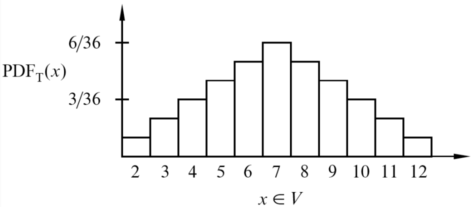
Back to throwing dice

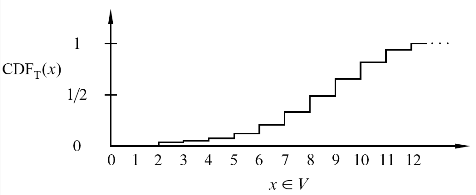
Q: Suppose we throw two standard dice one after the other. Let us define T to be the random variable equal to the sum of the dice. Plot PDF_T and CDF_T

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 $\text{PDF}_T[12] = \Pr[T = 12] = \frac{1}{36}$.





Questions?

Distributions

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

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

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

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

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

In the example, $\Pr[R = 1] = f(1) = p = \Pr[\text{event that we get a heads}]$.

CDF$_R$ for Bernoulli distribution: $F : \mathbb{R} \to [0,1]$:

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

Uniform Distribution

Canonical Example: 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.

PDF$_R$ for Uniform distribution: $f : V \to [0,1]$ such that for all $v \in V$, $f(v) = {}^1/_{|V|}$. In the example, $f(1) = f(2) = \ldots = f(6) = \frac{1}{6}$.

CDF$_R$ for Uniform distribution: For n elements in V arranged in increasing order – $(v_1, v_2, \ldots, v_n)$, 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?

Binomial Distribution

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

PDF$_R$ for Binomial distribution: $f : \{0, 1, 2, \ldots, n\} \to [0, 1]$. For $k \in \{0, 1, \ldots, n\}$, $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Proof: Let E_k be the event we get k heads. Let A_i be the event we get a heads in toss i.

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

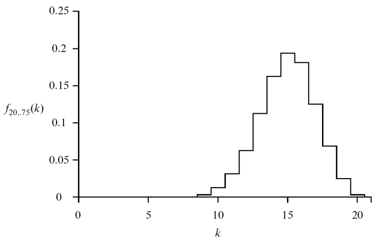
$$\begin{aligned}\Pr[E_k] &= \Pr[(A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c)] + \Pr[A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap \ldots \cap] + \ldots \\ &= \Pr[A_1] \Pr[A_2] \Pr[A_k] \Pr[A_{k+1}^c] \Pr[A_{k+2}^c] \ldots \Pr[A_n^c] + \ldots \qquad \text{(Independence of tosses)} \\ &= p^k (1-p)^{n-k} + p^k (1-p)^{n-k} + \ldots\end{aligned}$$

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

(Number of terms = number of ways to choose the k tosses that result in heads = $\binom{n}{k}$)

Binomial Distribution

For the Binomial distribution, $\mathrm{PDF}_R(k) = \binom{n}{k} p^k (1-p)^{n-k}$.



Q: Prove that $\sum_{k \in \mathrm{Range(R)}} \mathrm{PDF}_R[k] = 1$.
By the Binomial Theorem, $\sum_{k \in \mathrm{Range(R)}} \mathrm{PDF}_R[k] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$.

CDF$_R$ for Binomial distribution: $F : \mathbb{R} \to [0, 1]$:

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

Geometric Distribution

Canonical Example: We toss a biased coin independently multiple times. The probability of getting a heads 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.

PDF$_R$ for Geometric distribution: $f : \{1, 2, \ldots\} \to [0, 1]$. For $k \in \{1, 2, \ldots, \infty\}$, $f(k) = (1-p)^{k-1}\, p$.

Proof: 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.

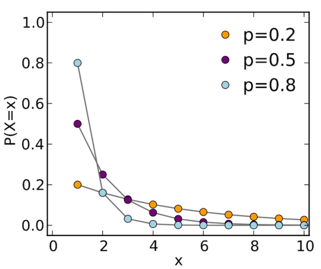
$$\begin{aligned} E_k &= A_1^c \cap A_2^c \cap \ldots \cap A_k \\ \Pr[E_k] &= \Pr[A_1^c \cap A_2^c \cap \ldots \cap A_k] = \Pr[A_1^c]\Pr[A_2^c]\ldots\Pr[A_k] \quad \text{(Independence of tosses)} \\ \implies \Pr[E_k] &= (1-p)^{k-1} p \end{aligned}$$

Q: Prove that $\sum_{k \in \text{Range(R)}} \text{PDF}_R[k] = 1$.

By the sum of geometric series, $\sum_{k \in \text{Range}(R)} \text{PDF}_R[k] = \sum_{k=1}^{\infty} (1-p)^{k-1} p = \frac{p}{1-(1-p)} = 1$.

Geometric Distribution

For the Geometric distribution, $\mathrm{PDF}_R(k) = (1-p)^{k-1}p$.



CDF_R for Geometric distribution: $F : \mathbb{R} \to [0,1]$:

$$F(x) = 0 \qquad \text{(for } x < 1\text{)}$$

$$= \sum_{i=1}^{k} (1-p)^{i-1}p \qquad \text{(for } k \leq x < k+1\text{)}$$

Questions?