CMPT 210: Probability and Computing

Lecture 13

Sharan Vaswani

February 16, 2023

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 $C : \mathcal{S} \to \{0, 1, 2, 3\}$. $C(HHT) = 2$.

An random variable partitions the sample space into several blocks. For r.v. R, for all $i \in \text{Range}(R)$, the event $[R = i] = \{\omega \in \mathcal{S} | R(\omega) = i\}$. For any r.v. R, $\sum_{i \in \text{Range(R)}} \Pr[R = i] = 1$.

Example: For the above r.v. C, $[C = 2] = \{HHT, HTH, THH\}$ and $\Pr[C = 2] = \frac{3}{8}$. $\sum_{i \in \text{Range(C)}} \Pr[C = i] = \Pr[C = 0] + \Pr[C = 1] + \Pr[C = 2] + \Pr[C = 3] = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$.

Recap

Indicator Random Variable: An indicator random variable corresponding to an event E is denoted as $\mathcal{I}_E$ and is defined such that for $\omega \in E$, $\mathcal{I}_E[\omega] = 1$ and for $\omega \notin E$, $\mathcal{I}_E[\omega] = 0$.

Example: When throwing two dice, if E is the event that both throws of the dice result in a prime number, then $\mathcal{I}_E((2,4)) = 0$ and $\mathcal{I}_E((2,3)) = 1$.

Probability density function (PDF): Let R be a r.v. 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 $x \notin \text{Range(R)}$.

Cumulative distribution function (CDF): The cumulative distribution function of R 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: If we flip three coins, and C counts the number of heads, then
$\text{PDF}_C[0] = \Pr[C = 0] = \frac{1}{8}$, and
$\text{CDF}_C[2.3] = \Pr[C \leq 2.3] = \Pr[C = 0] + \Pr[C = 1] + \Pr[C = 2] = \frac{7}{8}$.

Recap

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

Bernoulli Distribution: $f_p(0) = 1 - p$, $f_p(1) = p$. *Example*: When tossing a coin such that $\Pr[\text{heads}] = p$, random variable R is equal to 1 if we get a heads (and equal to 0 otherwise). In this case, R follows the Bernoulli distribution i.e. $R \sim \text{Ber}(p)$.

Uniform Distribution: If $R : \mathcal{S} \to V$, then for all $v \in V$, $f(v) = 1/|V|$. *Example*: When throwing an n-sided die, random variable R is the number that comes up on the die. $V = \{1, 2, \ldots, n\}$. In this case, R follows the Uniform distribution i.e. $R \sim \text{Uniform}(\{1, 2, \ldots, n\})$.

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.

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

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

$$= \Pr[A_1] \Pr[A_2] \Pr[A_k] \Pr[A_{k+1}^c] \Pr[A_{k+2}^c] \ldots \Pr[A_n^c] + \ldots \qquad \text{(Independence of tosses)}$$

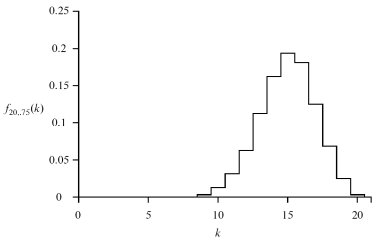
$$= p^k (1-p)^{n-k} + p^k (1-p)^{n-k} + \ldots$$

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

(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: $F : \mathbb{R} \to [0,1]$:

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

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.

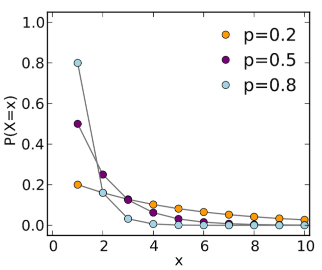
$$\begin{aligned} E_k &= A_1^c \cap A_2^c \cap \ldots \cap A_k \\ \Pr[E_k] &= \Pr[A_1^c \cap A_2^c \cap \ldots \cap A_k] = \Pr[A_1^c]\Pr[A_2^c]\ldots\Pr[A_k] \quad \text{(Independence of tosses)} \\ \implies \Pr[E_k] &= (1-p)^{k-1} p \end{aligned}$$

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, $\text{PDF}_R(k) = (1-p)^{k-1}p$.



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

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

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

Questions?

Distributions - Examples

Q: It is known that disks produced by a certain company will be defective with probability 0.01 independently of each other. The company sells the disks in packages of 10 and offers a money-back guarantee that at most 1 of the 10 disks is defective (the package can be returned if there is more than 1 defective disk). What proportion of packages is returned? If someone buys three packages, what is the probability that exactly one of them will be returned?

Let X be the random variable corresponding to the number of defective disks in a package. Let E be the event that the package is returned. We wish to compute $\Pr[E] = \Pr[X > 1]$. X follows the Binomial distribution $\text{Bin}(10, 0.01)$. Hence,

$$
\begin{aligned}
\Pr[E] = \Pr[X > 1] = 1 - \Pr[X \leq 1] &= 1 - \Pr[X = 0] - \Pr[X = 1] \\
&= 1 - \binom{10}{0}(0.99)^{10} - \binom{10}{1}(0.99)^{9}(0.01)^{1} \approx 0.05
\end{aligned}
$$

Distributions - Examples

Q: It is known that disks produced by a certain company will be defective with probability 0.01 independently of each other. The company sells the disks in packages of 10 and offers a money-back guarantee that at most 1 of the 10 disks is defective (the package can be returned if there is more than 1 defective disk). If someone buys three packages, what is the probability that exactly one of them will be returned?

Let F be the event that someone bought 3 packages and exactly one of them is returned.

Answer 1: Let E_i be the event that package i is returned. From the previous question, we know that $\Pr[E_i] = \Pr[\text{Package } i \text{ has more than 1 defective disk}] \approx 0.05$.

$$F = (E_1 \cap E_2^c \cap E_3^c) \cup (E_1^c \cap E_2^c \cap E_3) \cup (E_1^c \cap E_2 \cap E_3^c)$$

$$\Pr[F] = \Pr[E_1](1 - \Pr[E_2])(1 - \Pr[E_3]) + (1 - \Pr[E_1])(1 - \Pr[E_2]) \Pr[E_3] + \dots$$

$$\Pr[F] \approx 3 \times (0.05)(0.95)(0.95) \approx 0.15.$$

Answer 2: Let Y be the random variable corresponding to the number of packages returned. Y follows the Binomial distribution $\text{Bin}(3, 0.05)$ and we wish to compute $\Pr[F] = \Pr[Y = 1] \approx \binom{3}{1}(0.05)^1(0.95)^2 \approx 0.15.$

Distributions - Examples

Q: A communications system consists of n components, each of which will independently function with probability p. The total system will be able to operate effectively if at least one of its components functions. What is the probability that the total system functions?

Answer 1: Let E_i be the event that component i functions. $\Pr[E_i] = p$. Let F be the event that system functions. $\Pr[F] = \Pr[\cup_i E_i] = 1 - \Pr[\cap_i E_i^c] = 1 - (1-p)^n$.

Answer 2: If X is the number of functioning components, X follows the Binomial distribution $\text{Bin}(n, p)$, $\Pr[F] = \Pr[X \geq 1] = 1 - \Pr[X < 1] = 1 - \Pr[X = 0] = 1 - \binom{n}{0} p^0 (1-p)^n$.

Q: The total system will be able to operate effectively if at least 3 of its 5 components function. What is the probability that the total system functions?

In this case, $\Pr[F] = \Pr[X \geq 3] = \binom{n}{3} p^3 (1-p)^2 + \binom{n}{4} p^4 (1-p)^1 + \binom{n}{5} p^5 (1-p)^0$.

Distributions - Examples

Q: You are randomly and independently throwing darts. The probability that you hit the bullseye in throw i is p. Once you hit the bullseye you win and can go collect your reward. What is the probability that you win after exactly k throws?

The number of throws (T) to hit the bullseye follows a geometric distribution $\text{Geo}(p)$ and we wish to compute $\Pr[T = k] = (1-p)^{k-1}\, p$.

Q: What is the probability you win in less than k throws?

Answer 1: If E is the event that we win in less than k throws,
$\Pr[E] = \Pr[T < k] = \sum_{i=1}^{k-1} \Pr[T = i] = p \sum_{i=1}^{k-1} (1-p)^{i-1} = 1 - (1-p)^{k-1}$.
Answer 2:
$\Pr[E] = 1 - \Pr[E^c] = 1 - \Pr[\text{do not hit the bullseye in } k-1 \text{ throws}] = 1 - (1-p)^{k-1}$.

Distributions - Examples

Q: Suppose we throw a standard die and R is the random variable corresponding to the number on the die. We define a new random variable $X = 2R + 1$. What is the PDF_X?

Since R is a uniform random variable and the domain of $\text{PDF}_R = \{1, 2, \dots, 6\}$. .

The domain of PDF_X is $\{3, 5, 7, 9, 11, 13\}$.
$\text{PDF}_X[3] = \Pr[X = 3] = \Pr[2R + 1 = 3] = \Pr[R = 1] = \frac{1}{6}$. Similarly, $\text{PDF}_X[5] = \frac{1}{6}$. And we can conclude that X follows the uniform distribution on $\{3, 5, 7, 9, 11, 13\}$.

Q: Suppose $X = \max\{R - 3, 0\}$. What is the PDF_X?

In general, if $X = g(R)$, then for $x \in \text{Domain}(\text{PDF}_X)$,
$\text{PDF}_X[x] = \sum_{r \in \text{Domain}(\text{PDF}_R) | g(r) = x} \text{PDF}_R[r]$.

Questions?

Number Guessing Game

Q: 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).

Q: What is the probability that we win with this strategy?

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

But the numbers in the envelopes need 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!

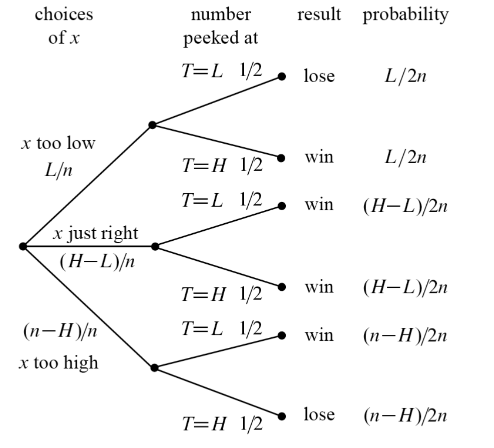
Strategy 3: Choose a random number x from $\{0.5, 1.5, 2.5, \ldots n - 1/2\}$ according to the uniform distribution i.e. $\Pr[x = 0.5] = \Pr[1.5] = \ldots = 1/n$. Then we peek at the number (denoted by T) in one envelope, and if $T > x$, we choose that envelope, else we choose the other envelope.

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).



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

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

Q: For $n = 100$, if $L = 23$ and $H = 54$, compute $\Pr[\text{ guessing too low} \mid \text{we win }]$

Questions?