

CMPT 210: Probability and Computing

Lecture 14

Sharan Vaswani

March 5, 2024

Probability density function (PDF): Let R be a r.v. with codomain V . The probability density function of R is the function $\text{PDF}_R : V \rightarrow [0, 1]$, 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 $\text{CDF}_R : \mathbb{R} \rightarrow [0, 1]$, 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 $R : \mathcal{S} \rightarrow V$, 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. $V = \{1, 2, \dots, n\}$. In this case, R follows the Uniform distribution i.e. $R \sim \text{Uniform}(1, n)$.

Binomial Distribution: $f_{n,p}(k) = \binom{n}{k} p^k (1 - p)^{n-k}$. *Example:* When tossing n independent coins such that $\Pr[\text{heads}] = p$, random variable R is the number of heads in n coin tosses. In this case, R follows the Binomial distribution i.e. $R \sim \text{Bin}(n, p)$.

Geometric Distribution: $f_p(k) = (1 - p)^{k-1} p$. *Example:* When repeatedly tossing a coin such that $\Pr[\text{heads}] = p$, random variable R is the number of tosses needed to get the first heads. In this case, R follows the Geometric distribution i.e. $R \sim \text{Geo}(p)$.

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?

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?

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 $\Pr[E_i] = \Pr[\text{Package } i \text{ has more than 1 defective disk}] \approx 0.05$.

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

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

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

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

$$\Pr[F] = \Pr[Y = 1] \approx \binom{3}{1}(0.05)^1(0.95)^2 \approx 0.15.$$

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. (a) What is the probability that you win after exactly k throws? (b) What is the probability you win in less than k throws?

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. (a) What is the probability that you win after exactly k throws? (b) What is the probability you win in less than k throws?

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

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. (a) What is the probability that you win after exactly k throws? (b) What is the probability you win in less than k throws?

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

(b) **Answer 1:** If E is the event that we win in less than k throws,
$$\Pr[E] = \Pr[T < k] = \sum_{i=1}^{k-1} \Pr[T = i] = p \sum_{i=1}^{k-1} (1 - p)^{i-1} = 1 - (1 - p)^{k-1}.$$

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. (a) What is the probability that you win after exactly k throws? (b) What is the probability you win in less than k throws?

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

(b) **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:

$$\Pr[E] = 1 - \Pr[E^c] = 1 - \Pr[\text{do not hit the bullseye in } k - 1 \text{ throws}] = 1 - (1 - p)^{k-1}.$$

Number Guessing Game

Q: We have two envelopes. Each contains a distinct number in $\{0, 1, 2, \dots, 100\}$. 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?

Number Guessing Game

Q: We have two envelopes. Each contains a distinct number in $\{0, 1, 2, \dots, 100\}$. 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?

Number Guessing Game

Q: We have two envelopes. Each contains a distinct number in $\{0, 1, 2, \dots, 100\}$. 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.

Number Guessing Game

Q: We have two envelopes. Each contains a distinct number in $\{0, 1, 2, \dots, 100\}$. 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.

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 $\{0.5, 1.5, 2.5, \dots, n - 1/2\}$ according to the uniform distribution i.e. $\Pr[x = 0.5] = \Pr[1.5] = \dots = 1/n$. 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.

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 $\{0.5, 1.5, 2.5, \dots, n - 1/2\}$ according to the uniform distribution i.e. $\Pr[x = 0.5] = \Pr[1.5] = \dots = 1/n$. 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.

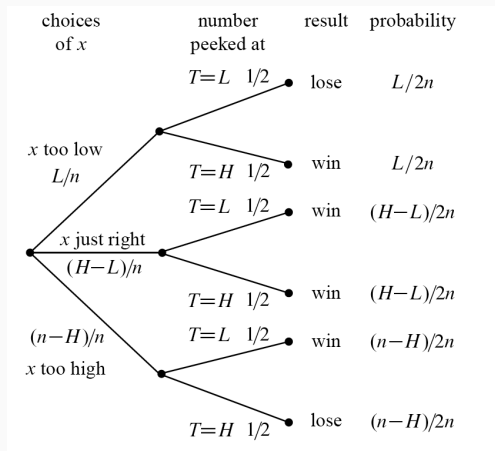
Q: But does it have better than 50% chance of winning?

Number Guessing Game

Let the numbers in the two envelopes be L (lower number) and H (the higher number).

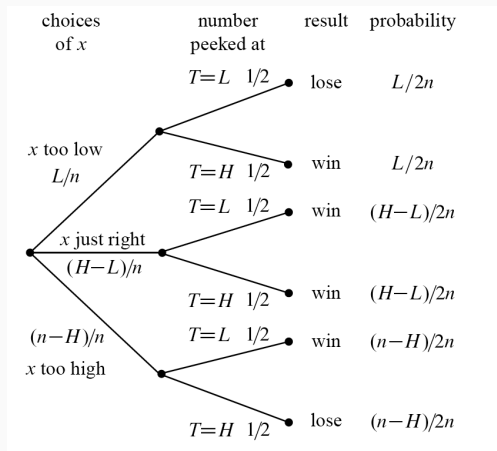
Number Guessing Game

Let the numbers in the two envelopes be L (lower number) and H (the higher number).



Number Guessing Game

Let the numbers in the two envelopes be L (lower number) and H (the higher number).

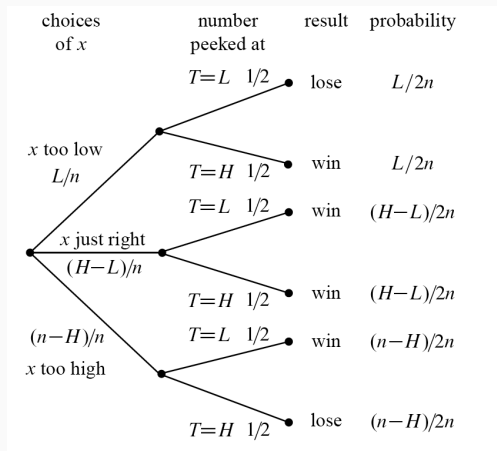


$$\begin{aligned}\Pr[\text{win}] &= \frac{L}{2n} + \frac{H-L}{2n} + \frac{H-L}{2n} + \frac{n-H}{2n} \\ &= \frac{1}{2} + \frac{H-L}{2n} \geq \frac{1}{2} + \frac{1}{2n} > \frac{1}{2}\end{aligned}$$

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

Number Guessing Game

Let the numbers in the two envelopes be L (lower number) and H (the higher number).



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

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?