Artificial Intelligence and Machine Learning

Barbara Caputo

Bernoulli Random Variables

Definition: Bernoulli Random Variable

Let X be a random variable taking only two possible values $\{0, 1\}$. Let $p = P(X = 1)$. Then X is said to be a **Bernoulli** random variable with parameter p, or equivalently $X \sim \text{Ber}(p)$.

Bernoulli Random Variables

Definition: Bernoulli Random Variable

Let X be a random variable taking only two possible values $\{0, 1\}$. Let $p = P(X = 1)$. Then X is said to be a **Bernoulli** random variable with parameter p, or equivalently $X \sim \text{Ber}(p)$.

Examples:

- Flipping a "fair" coin: $X \sim \text{Ber}(1/2)$
- Success starting Windows: $X \sim \text{Ber}(0.95)$

Let $X \sim \text{Ber}(p)$. Then

$$\mathbb{E}[X] = 0 \times P(X = 0) + 1 \times P(X = 1) = p$$

$$\mathbb{E}[X^2] = 0^2 \times P(X = 0) + 1^2 \times P(X = 1) = p$$

$$\mathrm{V}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p)$$

Bernoulli Trials

Independent Bernoulli random variables are important blocks to build more complicated random variables...

Bernoulli Trials

Independent Bernoulli random variables are important blocks to build more complicated random variables…

Consider the following examples:

- Flip a coin 10 times. Let X be the number of heads obtained.
- The packets received during a UDP connection arrive corrupted 5% of the times. Suppose you send 100 packets and let X be the number of packets received in error.
- In the next 30 births at an hospital, let X be the number of female babies.

Bernoulli Trials

Independent Bernoulli random variables are important blocks to build more complicated random variables…

Consider the following examples:

• Flip a coin 10 times. Let X be the number of heads obtained.
• The packets received during a UDP connection arrive corrupted 5% of the times. Suppose you send 100 packets and let X be the number of packets received in error.
• In the next 30 births at an hospital, let X be the number of female babies.

These are all examples of binomial random variables (provided you assume each of the Bernoulli trials are independent).

Repeated Bernoulli Trials

Example: Suppose we send 3 packets through a communication channel. Assume

- Each packet is received with probability 0.9.
- The events {packet 1 is received}, {packet 2 is received}, and {packet 3 is received} are all **independent**.
- Let X be the total number of received packets.

Repeated Bernoulli Trials

Example: Suppose we send 3 packets through a communication channel. Assume

- Each packet is received with probability 0.9.
- The events {packet 1 is received}, {packet 2 is received}, and {packet 3 is received} are all **independent**.
- Let X be the total number of received packets.

$$X_i \sim \text{Ber}(p) \rightarrow X_i = 1 \text{ iif } \{\text{packet i was received}\}, i = 1, 2, 3$$

$$X = X_1 + X_2 + X_3$$

x_1	x_2	x_3	$P(X_1 = x_1, X_2 = x_2, X_3 = x_3)$	$x = x_1 + x_2 + x_3$
0	0	0	0.001	0
0	0	1	0.009	1
0	1	0	0.009	1
0	1	1	0.081	2
1	0	0	0.009	1
1	0	1	0.081	2
1	1	0	0.081	2
1	1	1	0.729	3

Binomial Distribution

x_1	x_2	x_3	$P(X_1 = x_1, X_2 = x_2, X_3 = x_3)$	$x = x_1 + x_2 + x_3$
0	0	0	0.001	0
0	0	1	0.009	1
0	1	0	0.009	1
0	1	1	0.081	2
1	0	0	0.009	1
1	0	1	0.081	2
1	1	0	0.081	2
1	1	1	0.729	3

Binomial Distribution

x_1	x_2	x_3	$P(X_1 = x_1, X_2 = x_2, X_3 = x_3)$	$x = x_1 + x_2 + x_3$
0	0	0	0.001	0
0	0	1	0.009	1
0	1	0	0.009	1
0	1	1	0.081	2
1	0	0	0.009	1
1	0	1	0.081	2
1	1	0	0.081	2
1	1	1	0.729	3

Therefore

$$P(X = x) = \begin{cases} 0.001 & \text{if } x = 0 \\ 0.027 & \text{if } x = 1 \\ 0.243 & \text{if } x = 2 \\ 0.729 & \text{if } x = 3 \end{cases}$$

This is what is called a Binomial Random Variable!!!

Binomial Distribution

Definition: Binomial Random Variable

Consider a random experiment consisting of $n \in \mathbb{N}$ **independent** Bernoulli trials, where each trial can result in either success or failure.

Assume the probability of success for each trial is equal to $0 \leq p \leq 1$, and remains the same throughout the experiment.

The random variable X that equals the number of trials that result in success is a **Binomial Random Variable** with parameters n and p. The probability mass function of X is given by

$$P(X = k) = f(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n \ .$$

Binomial Distribution

Definition: Binomial Random Variable

Consider a random experiment consisting of $n \in \mathbb{N}$ **independent** Bernoulli trials, where each trial can result in either success or failure.

Assume the probability of success for each trial is equal to $0 \leq p \leq 1$, and remains the same throughout the experiment.

The random variable X that equals the number of trials that result in success is a **Binomial Random Variable** with parameters n and p. The probability mass function of X is given by

$$P(X = k) = f(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n \ .$$

We have seen $\binom{n}{k} = C_k^n$ before, the number of possible combinations of k out of n elements.

Binomial Distribution

The name of this distribution comes from the similarity with the *binomial expansion*: For any $a, b \in \mathbb{R}$

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i} \ .$$

Binomial Distribution

The name of this distribution comes from the similarity with the *binomial expansion*: For any $a, b \in \mathbb{R}$

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i} \ .$$

This immediately shows that the distribution is indeed valid, as

$$\sum_{k=0}^{n} f(k) = 1 \ .$$

The computation of the mean and variance of X can also be done directly from the p.m.f., but it is rather tedious and complicated... But there is a much easier way...

Binomial Distribution

Note that

$$X = X_1 + X_2 + \cdots + X_n \ ,$$

where $X_i \sim \text{Ber}(p)$ are independent Bernoulli trials.

As stated earlier, for independent random variables $X_1, \ldots, X_n$ we have

Binomial Distribution

Note that

$$X = X_1 + X_2 + \cdots + X_n \ ,$$

where $X_i \sim \mathsf{Ber}(p)$ are independent Bernoulli trials.

As stated earlier, for independent random variables $X_1, \ldots, X_n$ we have

$$\mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] \ ,$$

and

$$\mathsf{V}(X_1 + X_2 + \cdots + X_n) = \mathsf{V}(X_1) + \mathsf{V}(X_2) + \cdots + \mathsf{V}(X_n) \ .$$

Binomial Distribution

Note that

$$X = X_1 + X_2 + \cdots + X_n \ ,$$

where $X_i \sim \mathsf{Ber}(p)$ are independent Bernoulli trials.

As stated earlier, for independent random variables $X_1, \ldots, X_n$ we have

$$\mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] \ ,$$

and

$$\mathsf{V}(X_1 + X_2 + \cdots + X_n) = \mathsf{V}(X_1) + \mathsf{V}(X_2) + \cdots + \mathsf{V}(X_n) \ .$$

Let $X \sim \mathsf{Bin}(n, p)$ be a binomial random variable with parameters n and p. Then

$$\mathbb{E}[X] = np, \qquad \text{and} \qquad \mathsf{V}(X) = np(1-p)$$

Binomial Distribution Example

Samples from a certain water supply have a 10% chance of containing an organic pollutant. Suppose you collect 20 samples over time, and these are taking in such a way that you can assume these are independent. Let X be the number of contaminated samples.

$$X \sim \text{Bin}(20, 0.1)$$

Binomial Distribution Example

Samples from a certain water supply have a 10% chance of containing an organic pollutant. Suppose you collect 20 samples over time, and these are taking in such a way that you can assume these are independent. Let X be the number of contaminated samples.

$$X \sim \text{Bin}(20, 0.1)$$

$$P(X = 2) = \binom{20}{2} 0.1^2 (1 - 0.1)^{20-2} \approx 0.285$$

$$\begin{aligned} P(X \geq 2) = 1 - P(X < 2) &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - 0.1^0\ 0.9^{20} - 200.10.9^{20-1} = 0.608 \end{aligned}$$

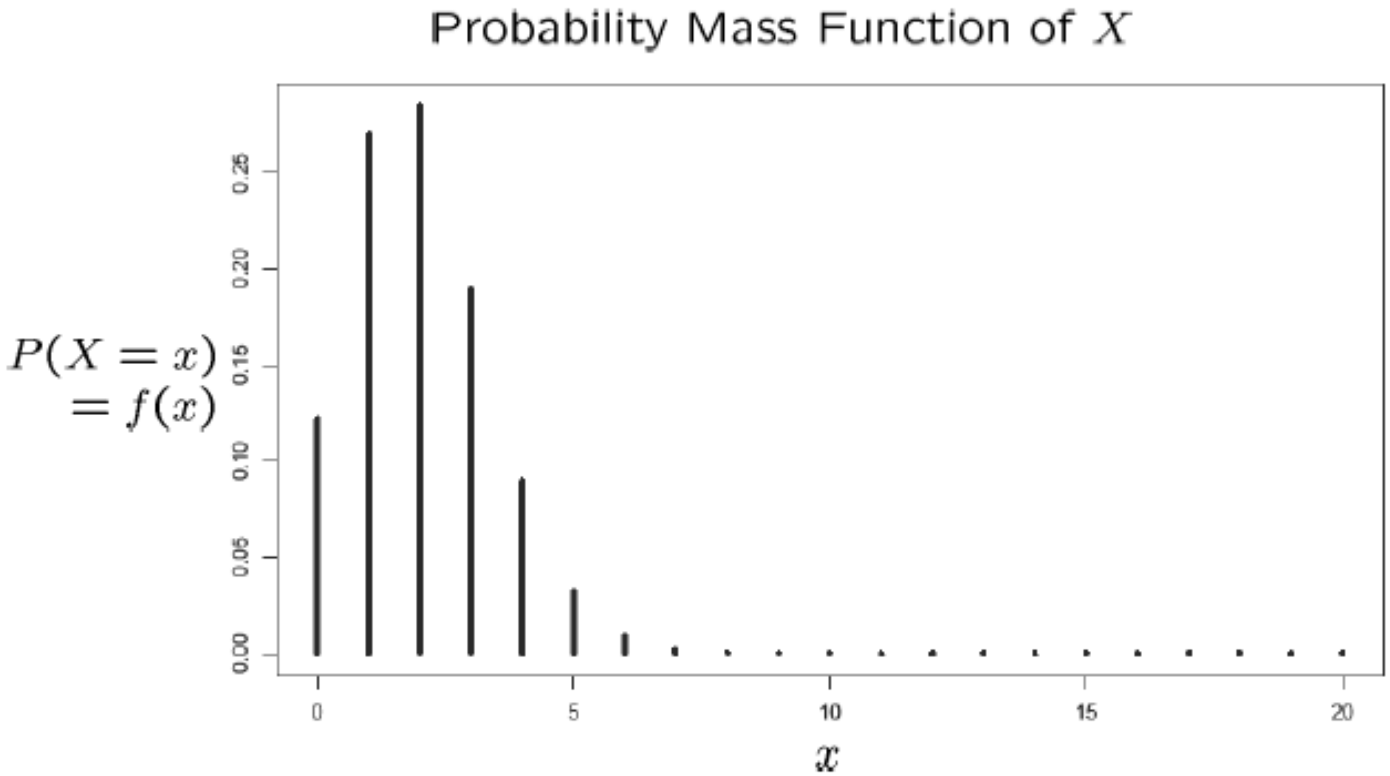
$$P(X \leq 7) = (\text{use table or computer}) = 0.9996$$

$$\begin{aligned} P(3 \leq X \leq 6) = P(X \leq 6) - P(X < 3) &= P(X \leq 6) - P(X \leq 2) \\ &= 0.3207 \end{aligned}$$

$$P(3 < X < 7) = 0.9976 - 0.8670 = 0.1306$$

$$\mathbb{E}[X] = 20 \times 0.1 = 2 \qquad \text{V}(X) = 20 \times 0.1(1 - 0.1) = 1.8$$

Binomial Distribution Example



I. Emily hits 60% of her free throws in basketball games. She had 25 free throws in last week's game. Use this information to answer the next **two** questions.

1. What is the average number of hits?
   (a) 10
   (b) 15
   (c) 20
   (d) 25

ANSWER: $\boxed{\text{(b)}}$

REASON:

The average (or the expected value) is $n \times p = 25 \times 0.60 = 15.$

2. What is the standard deviation of Emily's hit ?

   (a) 6
   (b) 3
   (c) 3.2
   (d) $\sqrt{6}$

ANSWER: $\boxed{\text{(d)}}$

REASON:

The standard deviation is $\sqrt{np(1-p)} = \sqrt{25 \times 0.60 \times (1 - 0.60)} = \sqrt{6} = 2.45$.

In the previous question, suppose Emily had 7 free throws in yesterday's game.

1. What is the probability that she made at least 5 hits?
   (a) 0.2613
   (b) 0.1306
   (c) 0.0280
   (d) 0.1586
   (e) 0.4199

ANSWER: $\boxed{\text{(e)}}$

REASON:

Denote Y the hits in Emily's 7 free throws. The event that she made at least 5 hits is then

$$(Y \geq 5) = (Y = 5 \text{ or } 6 \text{ or } 7).$$

So,

$$P(Y \geq 5) = P(Y = 5) + P(Y = 6) + P(Y = 7)$$

So,

$$P(Y \geq 5) = P(Y = 5) + P(Y = 6) + P(Y = 7)$$

$$\begin{aligned}
P(Y = 5) &= \frac{7!}{5!(7-5)!}(.6)^5(.4)^{7-5} = \frac{7 \times 6 \times 5!}{5! \times 2!}(.6)^5(.4)^2 \\
&= \frac{7 \times 6}{2 \times 1}(.6)^5(.4)^2 \ \ \text{(note that 5! was factored out)} \\
&= 21 \times (.6)^5(.4)^2 = 0.2613 \\
P(Y = 6) &= \frac{7!}{6!(7-6)!}(.6)^6(.4)^{7-6} = \frac{7 \times 6!}{6! \times 1!}(.6)^6(.4)^1 \\
&= \frac{7}{1}(.6)^6(.4)^1 \ \ \text{(note that 6! was factored out)} \\
&= 7 \times (.6)^6 \times .4 = 0.1306 \\
P(Y = 7) &= \frac{7!}{7!(7-7)!}(.6)^7(.4)^{7-7} = \frac{7!}{7! \times 0!}(.6)^7(.4)^0 \\
&= (.6)^7 \times 1 \ \ \text{(note that 7! was factored out)} \\
&= 0.0280.
\end{aligned}$$

Note that, $0! = 1$ and $(.4)^0 = 1$ above.
It follows that the probability of interest is $0.2613 + 0.1306 + 0.0280 = 0.4199$.