Lecture 11:

• Geometric Distribution
• Poisson Process
• Poisson Distribution
Geometric Distribution

The Geometric Distribution occurs when you count the number of independent and identically distributed Bernoulli trials until the first success.

Formally, if $Y \sim \text{Bernoulli}(p)$, and

$$X = \text{“The number of trials of } Y \text{ until the first success”}$$

then we say that $X$ is distributed according to the Geometric Distribution with parameter $p$, and write this as:

$$X \sim G(p)$$

where

$$R_X = \{1, 2, 3, \ldots, k, \ldots\}$$

$$S = \{S, FS, FFS, \ldots, FFF\ldots S, \ldots\}$$

$$f_X = \{p, (1-p)p, (1-p)^2p, \ldots, (1-p)^{k-1}p, \ldots\}$$

For $k$, we have $k-1$ failures and 1 success ($\text{FFF}... \ S$), which has probability $(1-p)^{k-1}p$. 
Geometric Distribution

Examples

An absent-minded professor has 6 keys on his key ring and does not always remember which of his keys opens his office door. He chooses keys randomly and with replacement to try to open his door.

(a) On average, how many keys would he normally try before his door opens.

(b) What is the probability that he opens it after only 3 tries?

Solution. This is $G(1/6)$.

(a) This is just $E(X) = 1/p = 6$

(b) $P(X=3) = (5/6)^2 (1/6) = 0.1157$
As you might expect, the standard formulae for infinite geometric series come in handy when working with the geometric distribution:

\[
\sum_{k=0}^{\infty} r^k = \frac{1}{1 - r} \quad \text{for } |r| < 1
\]

\[
\sum_{k=1}^{\infty} n x^{n-1} = \frac{1}{(1 - x)^2} \quad \text{for } |x| < 1
\]

First, we show that all the probabilities sum to 1.0 (so it actually IS a distribution):

\[
\sum_{n=1}^{\infty} f_X(n) = p + (1 - p)p + (1 - p)^2p + \ldots
\]

\[
= p \left( 1 + (1 - p) + (1 - p)^2 + \ldots \right)
\]

\[
= p \sum_{k=0}^{\infty} (1 - p)^k
\]

\[
= p \frac{1}{1 - (1 - p)} = 1.0
\]
Geometric Distribution

Similarly, we can derive the expected value:

\[
E(X) = 1p + 2(1 - p)p + 3(1 - p)^2p + ... \\
= p \left( 1 + 2(1 - p) + 3(1 - p)^2 + ... \right) \\
= p \sum_{k=1}^{\infty} k(1 - p)^{k-1} \\
= p \frac{1}{(1 - (1 - p))^2} \\
= \frac{1}{p}
\]

Deriving the variance is much more complicated, and we’ll just quote it...

So we have:

\[
E(X) = \frac{1}{p} \\
Var(X) = \frac{1 - p}{p^2}
\]
Geometric Distribution

Examples

On average, how many independent games of poker are required until a particular player is dealt a straight?

Solution: This is $G(0.003925)$. $E(X) = 1/0.003925 = 254.777$
**Geometric Distribution**

Fortunately, the probability mass function and the CDF are both easy to compute:

\[
R_X = \{1, 2, 3, \ldots\} \\
\]

\[
f_X = \{p, (1-p)p, (1-p)^2p, (1-p)^3p, \ldots\} \\
\]

\[
f_X(k) = (1-p)^{k-1}p \\
\]

\[
P(X > k) = (1-p)^k p + (1-p)^{k+1} p + (1-p)^{k+2} p + \ldots \\
= (1-p)^k \left( p + (1-p)p + (1-p)^2p + (1-p)^3p + \ldots \right) \\
= (1-p)^k \\
\]

\[
P(X \leq k) = F_X(k) = 1.0 - (1-p)^k \\
\]
Geometric Distribution

Examples

From an ordinary deck of 52 cards we draw cards at random and with replacement, and successively until an ace is drawn. What is the probability that at least 10 draws are needed?

Solution: The probability of an ace is $4/52 = 1/13$. $P(X>9) = (1-1/13)^9 = 0.4866$
Geometric Distribution: The Memoryless Property

The Geometric is one of two distributions that has the Memoryless Property, which we have already discussed informally before now as “the coin doesn’t remember its past flips” but which is actually a much stronger statement about the entire distribution.

Intuitive Version: Suppose we are about to start flipping a coin for which the probability of heads is p. Then the probability distribution of $X = \text{“how many flips until the first head?”}$ is $G(p)$.

Now suppose that the first $k$ flips are tails. Then the probability distribution of $Y = \text{“how many more flips until the first head?”}$ is still $G(p)$.

In other words, it doesn’t matter when you start to count or what the past history is: the exact theoretical distribution is always the same.

Intuitive Proof: Suppose you come into the room while someone is flipping the coin, and see him/her just about to flip the coin. How do you know how many flips have occurred before you came in, and why would it matter?
Geometric Distribution: The Memoryless Property

**Theorem**  A random variable $X$ is called memoryless if, for any $n, m \geq 0$,

$$P(X > n + m \mid X > m) = P(X > n)$$

For any probability $p$, $X \sim G(p)$ has the memoryless property.

(In fact, the Geometric is the only discrete distribution with this property; a continuous version of the Geometric, called the Exponential, is the other one.)
Geometric Distribution: The Memoryless Property

Proof:

\[
P(X > n + m \mid X > m) = \frac{P(X > n + m \text{ and } X > m)}{P(X > m)}
\]

\[
= \frac{P(X > n + m)}{P(X > m)}
\]

\[
= \frac{(1 - p)^{n+m}}{(1 - p)^m}
\]

\[
= (1 - p)^n
\]

\[
= P(X > n)
\]
The *Poisson Process* concept captures an important way of thinking about events randomly occurring through time (or space)... Two things to remember are

- **Events are discrete** (they happen or they don’t – you can think of it as a Bernoulli trial with an outcome of success or failure), but
- **Time and space are continuous**..... the random behavior here is the time of an event.

When an event has happened we say it has arrived. You can think of this as a sequence of real numbers giving the **arrival time** of an event:

\[
\text{Arrival Times} = \{ 0.4324\ldots, 0.734, 1.389\ldots, 1.453\ldots, 2.1546\ldots, \ldots \}
\]
Poisson Process

Examples in the time domain:

- Sneezes in a classroom
- Alpha particles emitted from U 238
- Email arriving in my inbox
- Accidents at an intersection
- Earthquakes, volcanoes, asteroids, ...

Yellowstone volcano eruption: NASA to SAVE the world from supervolcano erupting

NASA scientists are creating an ambitious plan to prevent an explosion of a Yellowstone volcano that could even end human life by drilling a hole.
Poisson Process

Formally, we have the following definition: suppose we have discrete events occurring through time as just described, and let

\[ N[s..t] = \text{the number of events arriving in the time interval } [s..t] \]

such that

1) The expected value of \( N[s..t] \) is proportional to the length \( (t - s) \) of the interval; in particular, for any two non-overlapping intervals of the same length, the mean number of occurrences in each is the same;
2) The number of arrivals in two non-overlapping intervals is independent; and
3) The probability of two events occurring at the same time is 0.

Then this random process is said to be a Poisson Process.

We shall be dealing only with discrete intervals of time for now, and so the important things to remember are that the intervals are independent and the mean number of arrivals in each time unit is the same.
Poisson Process

It is also possible that the continuous dimension is distance in space, in 1 dimension or more than 1. Examples include:

The occurrence of leaks in an undersea pipeline (1D):

Location of trees in a 1 square mile plot of land:

Location of novas in a cubic gigaparsec volume in space:

The important point is that events (discrete) occur along 1 or more (continuous) dimensions.
Poisson Random Variables

Suppose we have a Poisson Process and we fix the unit time interval we consider (say, 1 second or 1 year, etc.), where the mean number of arrivals in a unit interval is $\lambda$, and then each time we “poke” the random variable $X$ we return $N[0..1], N[1..2], N[3..4]$, etc.

Then we call $X$ a Poisson Random Variable with rate parameter $\lambda$, denoted

$$X \sim Poi(\lambda)$$

where

$$R_X = \{ 0, 1, 2, 3, ... \}$$

$$f_X(k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Poi(3)
Poisson Random Variables

Examples

Assume that arrivals of email in my Inbox are a Poisson Process with rate $\lambda = 10$ messages per hour. Then $X \sim \text{Poi}(10)$ returns the random number of emails which arrive within any particular hour.

What is the probability that I get no emails in the next hour?

$$P(X = 0) = f_X(0) = \frac{e^{-10} \lambda^0}{0!} = e^{-10} = 4.54 \times 10^{-5}$$

$$X \sim \text{Poi}(10)$$

$$R_X = \{0, 1, 2, 3, \ldots\}$$

$$P(X = k) = f_X(k) = \frac{e^{-10} 10^k}{k!}$$
Poisson Random Variables

Examples

Assume that arrivals of email in my Inbox are a Poisson Process with rate $\lambda = 10$ messages per hour. Then $X \sim \text{Poi}(10)$ returns the random number of emails which arrive within any particular hour.

What is the probability that I get exactly 10 emails in the next hour?

$$P(X = 10) = f_X(10) = \frac{e^{-10} \cdot 10^{10}}{10!} = 0.1251$$

$$X \sim \text{Poi}(10)$$

$$R_X = \{ 0, 1, 2, 3, \ldots \}$$

$$P(X = k) = f_X(k) = \frac{e^{-10} \cdot 10^k}{k!}$$
Poisson Random Variables

Examples

Unfortunately there is no way to compute the CDF or ranges except by simply adding together all the individual values.

What is the probability that I get between 5 and 15 emails (inclusive) emails in the next hour?

\[
P(5 \leq X \leq 15) = \sum_{k=5}^{15} \frac{e^{-10} \cdot 10^k}{k!} = 0.922
\]

\[X \sim Poi(10)\]

\[R_X = \{0, 1, 2, 3, \ldots\}\]

\[
P(X = k) = f_X(k) = \frac{e^{-10} \cdot 10^k}{k!}
\]