Advanced Statistics
DS2003 (BDS-4A)
Lecture 03

Instructor: Dr. Syed Mohammad Irteza

Assistant Professor, Department of Computer Science, FAST

22 February, 2022

Previous Lecture

- IID
- Bernoulli Trials
- Permutations and Combinations
- Geometric Distribution
- Binomial Distribution

Topics for Today

- The Poisson Distribution
- Negative Binomial Distribution
- Normal Distribution

The Poisson Distribution

- The Poisson distribution is often useful for estimating the *number of rare events* in a large population over a *short unit of time for a fixed population* if the individuals within the population are *independent*.
- The ***rate*** for a Poisson distribution is the average number of occurrences in a mostly-fixed population per unit of time, and is typically denoted by λ.
- Using the *rate*, we can describe the probability of *observing exactly k rare events* in a single unit of time.

$$\text{P(observe } k \text{ rare events)} = \frac{\lambda^k e^{-\lambda}}{k!}$$

The Poisson Distribution

$$\text{P(observe } k \text{ rare events)} = \frac{\lambda^k e^{-\lambda}}{k!}$$

- where *k* may take a value *0, 1, 2, and so on*, and *k!* represents *k-factorial*.
- The letter $e \approx 2.718$ is the base of the natural logarithm.
- The mean and standard deviation of this distribution are λ and $\sqrt{\lambda}$, respectively

Poisson Distribution – Example 01

- Suppose that in a rural region of a developing country electricity power failures occur following a Poisson distribution with an average of *2 failures every week*. Calculate the probability that in a given week the electricity fails only once.

Poisson Distribution – Example 01

- Suppose that in a rural region of a developing country electricity power failures occur following a Poisson distribution with an average of *2 failures every week*. Calculate the probability that in a given week the *electricity fails only once*.
- *Given λ = 2*.

$$
\begin{aligned}
P(\text{only 1 failure in a week}) &= \frac{2^1 \times e^{-2}}{1!} \\
&= \frac{2 \times e^{-2}}{1} \\
&= 0.27
\end{aligned}
$$

Poisson Distribution – Example 02

- Suppose that in a rural region of a developing country electricity power failures occur following a Poisson distribution with an average of *2 failures every week*. Calculate the probability that on a ***given day*** the *electricity fails three times*.

  - *We are given the weekly failure rate, but to answer this question we need to first calculate the average rate of failure on a given day:*
  - $\lambda_{day} = 2/7 = 0.2857.$
  - *Note that we are assuming that the probability of power failure is the same on any day of the week, i.e. we assume independence.*

Poisson Distribution – Example 02

- Suppose that in a rural region of a developing country electricity power failures occur following a Poisson distribution with an average of *2 failures every week*. Calculate the probability that on a ***given day*** the *electricity fails three times*.
  - $\lambda_{day} = 2/7 = 0.2857.$

- $(0.2857^3 * \ e^{-0.2857})\ /\ (3!) = 0.00292$

Is it Poisson?

- A random variable may follow a Poisson distribution if the event being considered is rare, the population is large, and the events occur independently of each other
- However we can think of situations where the events are not really *independent*. For example, if we are interested in the *probability of a certain number of weddings over one summer*, we should take into consideration that *weekends* are more popular for weddings.
- In this case, a Poisson model may sometimes still be reasonable if we allow it to have a *different rate for different times*; we could model the rate as higher on weekends than on weekdays.
    - The idea of modeling rates for a Poisson distribution against a second variable (day of the week) forms the foundation of some more advanced methods called *generalized linear models*.

Practice

- A random variable that follows which of the following distributions can take on values other than positive integers?

(a) Poisson

(b) Negative binomial

(c) Binomial

(d) Normal

(e) Geometric

Practice

- A random variable that follows which of the following distributions can take on values other than positive integers?

(a) Poisson

(b) Negative binomial

(c) Binomial

(d) Normal

(e) Geometric

Poisson – Another Example

- Consider an experiment that consists of counting the number of α-particles given off in a one-second interval by one gram of radioactive material. If we know from past experience that, on the average, 3.2 such α-particles are given off, what is a good approximation to the probability that no more than 2 α -particles will appear?

Poisson – Another Example

- Consider an experiment that consists of counting the number of α-particles given off in a one-second interval by one gram of radioactive material. If we know from past experience that, on the average, 3.2 such α-particles are given off, what is a good approximation to the probability that no more than 2 α -particles will appear?
- Solution:
  - If we think of the gram of radioactive material as consisting of a large number n of atoms each of which has probability $\frac{3.2}{n}$ of disintegrating and sending off an α-particle during the second considered, then we see that, to a very close approximation, the number of α-particles given off will be a Poisson random variable with parameter $\lambda = 3.2$. Hence the desired probability is:

Poisson – Another Example

- If we think of the gram of radioactive material as consisting of a large number n of atoms each of which has probability $\frac{3.2}{n}$ of disintegrating and sending off an α-particle during the second considered, then we see that, to a very close approximation, the number of α-particles given off will be a Poisson random variable with parameter *λ = 3.2*. Hence the desired probability is:

- $P(X \le 2) = e^{-3.2} + 3.2 * e^{-3.2} + \left(\frac{3.2^2}{2}\right) * e^{-3.2} = 0.382$

Negative Binomial Distribution

- The negative binomial distribution describes the probability of observing the k^{th} success on the n^{th} trial.
- The following *four conditions* are useful for identifying a negative binomial case:
  - The trials are independent.
  - Each trial outcome can be classified as a success or failure.
  - The probability of success (p) is the same for each trial.
  - The last trial must be a success.

Note that the first three conditions are common to the binomial distribution.

Negative Binomial Distribution

Negative binomial distribution:

$$\text{P}(k^{th} \text{ success on the } n^{th} \text{ trial}) = \binom{n-1}{k-1} p^k \ (1-p)^{n-k},$$

where p is the probability that an individual trial is a success. All trials are assumed to be independent.

Practice

- A college student working at a psychology lab is asked to recruit 10 married couples to participate in a study. She decides to stand outside the PhD Scholar Offices and ask every 5th person leaving the building whether they are married and, if so, whether they would like to participate in the study with their spouse.
- Suppose the probability of finding such a person is 10%. What is the probability that she will need to ask 30 people before she hits her goal?

Practice

- Suppose the probability of finding such a person is 10%. What is the probability that she will need to ask 30 people before she hits her goal?
- *Given: p = 0.10, k = 10, n = 30. We are asked to find the probability of 10th success on the 30th trial, therefore we use the negative binomial distribution.*

$$\begin{aligned} P(10^{th} \text{ success on the } 30^{th} \text{ trial}) &= \binom{29}{9} \times 0.10^{10} \times 0.90^{20} \\ &= 10{,}015{,}005 \times 0.10^{10} \times 0.90^{20} \\ &= 0.00012 \end{aligned}$$

Binomial vs Negative Binomial

- How is the negative binomial distribution different from the binomial distribution?
  - In the binomial case, we typically have a fixed number of trials and instead consider the number of successes.
  - In the negative binomial case, we examine how many trials it takes to observe a fixed number of successes and require that the last observation be a success.

Practice

Which of the following describes a case where we would use the negative binomial distribution to calculate the desired probability?

(a) Probability that a 5 year old boy is taller than 42 inches.

(b) Probability that 3 out of 10 softball throws are successful.

(c) Probability of being dealt four aces out of a pack of cards (4 picked).

(d) Probability of missing 8 shots before the first hit.

(e) Probability of hitting the ball for the 3rd time on the 8th try.

Practice

Which of the following describes a case where we would use the negative binomial distribution to calculate the desired probability?

(a) Probability that a 5 year old boy is taller than 42 inches.

(b) Probability that 3 out of 10 softball throws are successful.

(c) Probability of being dealt four aces out of a pack of cards (4 picked).

(d) Probability of missing 8 shots before the first hit.

(e) Probability of hitting the ball for the 3rd time on the 8th try.

The Normal Distribution

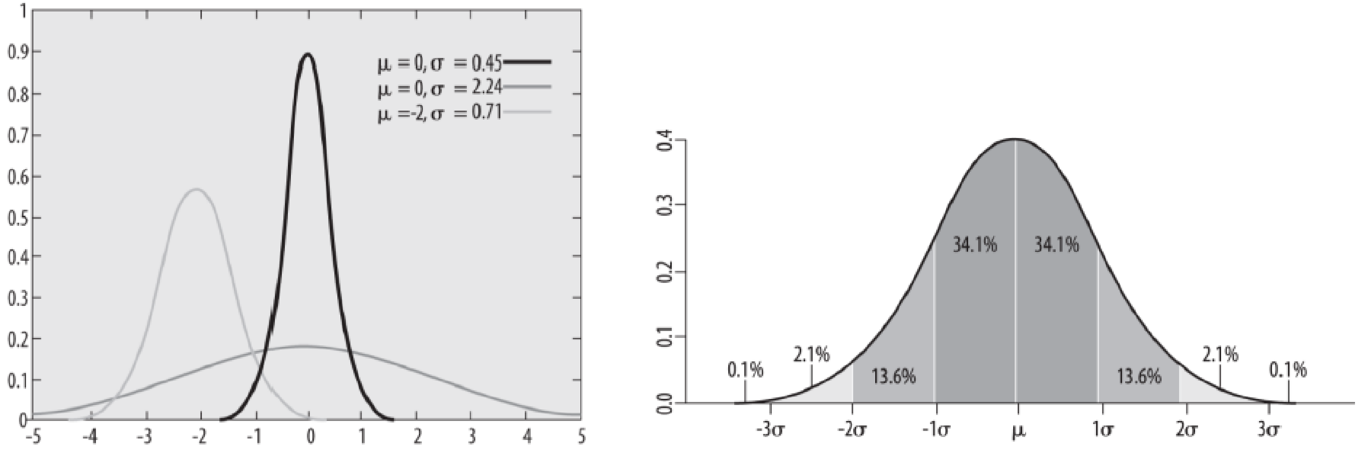

Figure 3.1: (left) All normal distributions have the same shape but differ to their μ and σ: they are shifted by μ and stretched by σ. (right) Percent of data failing into specified ranges of the normal distribution.

The Normal Distribution

- The amount of data that falls into the different regions:
  - $\mu \pm \sigma \rightarrow$ *68% data*
  - $\mu \pm 2\sigma \rightarrow$ *95% data*
  - $\mu \pm 3\sigma \rightarrow$ *99.7% data*

Sources

- openintro.org/os (Chapter 4)