Lecture 13

Poisson Distribution

Text: A Course in Probability by Weiss 5.5

STAT 225 Introduction to Probability Models
February 16, 2014

Whitney Huang
Purdue University
Agenda

1. Motivation
2. Poisson Distribution
3. Summary
Review

So far, we have seen discrete probability distributions of the number of successes in a sequence of random experiments with specified sample size.

- Bernoulli distribution: independent trial (sampling with replacement), sample size $= 1$
Review

So far, we have seen discrete probability distributions of the number of successes in a sequence of random experiments with specified sample size.

- **Bernoulli distribution**: independent trial (sampling with replacement), sample size $= 1$
- **Binomial distribution**: independent trials (sampling with replacement), sample size $= n$
Review

So far, we have seen discrete probability distributions of the number of successes in a sequence of random experiments with specified sample size.

- **Bernoulli distribution**: independent trial (sampling with replacement), sample size $= 1$
- **Binomial distribution**: independent trials (sampling with replacement), sample size $= n$
- **Hypergeometric distribution**: dependent trials (sampling without replacement), sample size $= n$
Review

So far, we have seen discrete probability distributions of the number of successes in a sequence of random experiments with specified sample size.

- **Bernoulli distribution**: independent trial (sampling with replacement), sample size \(= 1\)
- **Binomial distribution**: independent trials (sampling with replacement), sample size \(= n\)
- **Hypergeometric distribution**: dependent trials (sampling without replacement), sample size \(= n\)
Review

So far, we have seen discrete probability distributions of the number of successes in a sequence of random experiments with specified sample size.

- **Bernoulli distribution**: independent trial (sampling with replacement), sample size $= 1$
- **Binomial distribution**: independent trials (sampling with replacement), sample size $= n$
- **Hypergeometric distribution**: dependent trials (sampling without replacement), sample size $= n$

The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space ⇒ **does not have a (fixed) sample size**
Poisson Distribution:

Characteristics of the Poisson Distribution:
Let $X$ be a Poisson r.v.

- The definition of $X$: The number of successes
Poisson Distribution:

Characteristics of the Poisson Distribution:
Let $X$ be a Poisson r.v.

- The definition of $X$: The number of successes
- The support: $x = 0, 1, 2, \cdots$
Poisson Distribution:

Characteristics of the Poisson Distribution:
Let $X$ be a Poisson r.v.

- The definition of $X$: The number of successes
- The support: $x = 0, 1, 2, \cdots$
- Its parameter(s) and definition(s): $\lambda$: the average number of successes
Poisson Distribution:

Characteristics of the Poisson Distribution:
Let $X$ be a Poisson r.v.

- The definition of $X$: The number of successes
- The support: $x = 0, 1, 2, \ldots$
- Its parameter(s) and definition(s): $\lambda$: the average number of successes
- The probability mass function (pmf): $p_X(x) = \frac{e^{-\lambda} \lambda^x}{x!}$
Poisson Distribution:

Characteristics of the Poisson Distribution:
Let $X$ be a Poisson r.v.

- The definition of $X$: The number of successes
- The support: $x = 0, 1, 2, \cdots$
- Its parameter(s) and definition(s): $\lambda$: the average number of successes
- The probability mass function (pmf): $p_X(x) = \frac{e^{-\lambda} \lambda^x}{x!}$
- The expected value: $\mathbb{E}[X] = \lambda$
Poisson Distribution:

Characteristics of the Poisson Distribution:
Let $X$ be a Poisson r.v.

- The definition of $X$: The number of successes
- The support: $x = 0, 1, 2, \cdots$
- Its parameter(s) and definition(s): $\lambda$: the average number of successes
- The probability mass function (pmf): $p_X(x) = \frac{e^{-\lambda} \lambda^x}{x!}$
- The expected value: $\mathbb{E}[X] = \lambda$
- The variance: $\text{Var}(X) = \lambda$
If $X \sim \text{Binomial}(n, p)$ and $n > 100$ and $p < .01$ then we can approximate the distribution $X$ by using $X^* \sim \text{Poisson}(n \times p)$
Example 33

Let us say a certain disease has a 0.14% of occurring. Let us sample 1,000 people. Find the exact and approximate probabilities that 0 people have the disease and at most 5 people have the disease.

Solution.

Set-up:
Let $X$ be the number of people have the disease in the sample ($n = 1000$).

Which distribution to use?
- The sample size is fixed ($n = 1000 \Rightarrow \text{Binomial or Hypergeometric}$

What are the parameters? $n = 1000, p = .0014$
Example 33

Let us say a certain disease has a 0.14% of occurring. Let us sample 1,000 people. Find the exact and approximate probabilities that 0 people have the disease and at most 5 people have the disease.

Solution.

Set-up:
Let \( X \) be the number of people have the disease in the sample \((n = 1000)\).

Which distribution to use?
1. The sample size is fixed \((n = 1000) \Rightarrow \text{Binomial or Hypergeometric}\)
2. One person have the disease is independent to others \(\Rightarrow \text{Binomial}\)

What are the parameters? \(n = 1000, p = .0014\)
Example 33 cont’d

Any approximation?
Since \( n = 1000 > 100 \) and \( p = .0014 < .01 \). We can use Poisson \( X^* \) to approximate Binomial distribution \( X \)

Exact probabilities
\[
X \sim Binomial(n = 1000, p = .0014)
\]
\[
P(X = 0) = \binom{1000}{0}(.0014)^0(1 − .0014)^{1000−0} = .2464
\]
\[
P(X \leq 5) = \sum_{x=0}^{5} \binom{1000}{x}(.0014)^x(1 − .0014)^{1000−x} = .9986
\]

Approximate probabilities \( X^* \sim Poisson(\lambda = n \times p = 1.4) \)
\[
P(X = 0) \approx P(X^* = 0) = \frac{e^{-1.4}1.4^0}{0!} = .2466
\]
\[
P(X \leq 5) \approx P(X^* \leq 5) = \sum_{x=0}^{5} \frac{e^{-1.4}1.4^x}{x!} = .9986
\]
Example 34

Suppose earthquakes occur in the western US with a rate of 2 per week. Let $X$ be the number of earthquakes in the western US this week. Let $Y$ be the number of earthquakes in the western US this month (assume a 4 week period of time). Find the probability that $X$ is 3 and $Y$ is 12. Let $Z$ be the number of weeks in a 4 week period that have a week with 3 earthquakes in the western US. Find the probability that $Z$ is 4. Is this the same as the probability that $Y$ is 12? Does this make sense?

Solution.

**Set-up:**

$X$ be the number of earthquakes in the western US this week. $Y$ be the number of earthquakes in the western US in a 4 week period. $Z$ be the number of weeks in a 4 week period that have a week with 3 earthquakes in the western US.
Example 34 cont’d

Solution.

Which distribution to use?
X and Y

- No set sample size

Z

What are the parameters?
X : \( \lambda = 2 \) per week
Y : \( \lambda = 8 \) per 4 weeks
Z : \( n = 4, p = \Pr(X = 3) \)
Example 34 cont’d

Solution.

Which distribution to use?

X and Y

1. No set sample size
2. Number of events in a given time interval with specified occurrence rate ⇒ Poisson

Z

What are the parameters?

X : \( \lambda = 2 \) per week
Y : \( \lambda = 8 \) per 4 weeks
Z : \( n = 4, p = \Pr(X = 3) \)
Example 34 cont’d

Solution.

Which distribution to use?

$X$ and $Y$

1. No set sample size
2. Number of events in a given time interval with specified occurrence rate $\Rightarrow$ Poisson

$Z$

1. The sample size is fixed $n = 4$ $\Rightarrow$ Binomial or Hypergeometric

What are the parameters?

$X : \lambda = 2$ per week
$Y : \lambda = 8$ per 4 weeks
$Z : n = 4, p = P(X = 3)$
Example 34 cont’d

Solution.

Which distribution to use?

\( X \) and \( Y \)
1. No set sample size
2. Number of events in a given time interval with specified occurrence rate ⇒ Poisson

\( Z \)
1. The sample size is fixed \( n = 4 \) ⇒ Binomial or Hypergeometric
2. Whether in a given week with 3 earthquakes is independent to other weeks ⇒ Binomial

What are the parameters?

\( X : \lambda = 2 \) per week
\( Y : \lambda = 8 \) per 4 weeks
\( Z : n = 4, p = \mathbb{P}(X = 3) \)
Example 34 cont’d

\( X \sim \text{Poisson}(\lambda = 2) \) \( Y \sim \text{Poisson}(\lambda = 8) \)

\[ P(X = 3) = \frac{e^{-2}2^3}{3!} = .1804 \]

\[ P(Y = 12) = \frac{e^{-8}8^{12}}{12!} = .0481 \]

\[ P(Z = 4) = \binom{4}{4}(.1804)^4(1 - .1804)^{4-4} = .0011 \]

The probability that \( Z \) is 4 is much lower than the probability that \( Y \) is 12. This makes sense since there are more options for \( Y \) being 12. The probability that \( Z \) is 4 means that 4 weeks in a row there were 3 earthquakes. However with \( Y \) being 12 that just means 12 earthquakes happened in 4 weeks. We could have had 12, 0, 0, 0 or 7, 2, 1, 2, etc.
Example 35

PRP has on average 4 telephone calls per minute. Let \( X \) be the number of phone calls in the next minute. Find the probability that \( X \) is at least 3

Solution.

Which distribution to use?

- No set sample size

What are the parameters?

\( \lambda = 4 \) per minute

\( X \sim \text{Poisson}(\lambda = 4) \)

\[
P(X \geq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)] = 1 - \left[ \frac{e^{-4}4^0}{0!} + \frac{e^{-4}4^1}{1!} + \frac{e^{-4}4^2}{2!} \right] = 1 - .0183 - .0733 - .1465 = .7619\]
Example 35

PRP has on average 4 telephone calls per minute. Let $X$ be the number of phone calls in the next minute. Find the probability that $X$ is at least 3

Solution.

Which distribution to use?

1. No set sample size
2. Number of events in a given time interval with specified occurrence rate $\Rightarrow$ Poisson

What are the parameters?

$\lambda = 4$ per minute

$X \sim \text{Poisson}(\lambda = 4)$

$P(X \geq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)] =
1 - \left[ \frac{e^{-4}4^0}{0!} + \frac{e^{-4}4^1}{1!} + \frac{e^{-4}4^2}{2!} \right] =
1 - .0183 - .0733 - .1465 = .7619$
Poisson Distribution

Motivation

Summary

In today’s lecture, we

- Introduced the **Poisson distribution**
  - Used when measuring rate of “success"
  - Different from **Bernoulli, Binomial, and Hypergeometric distributions**, the sample size is not fixed
- Poisson approximation to Binomial
  - Appropriate when $n > 100$ and $p < .01$