Lecture 4.2

Common Probability Distributions: Discrete Distribution

Discrete Distributions

The distribution of a discrete random variable is specified by its *probability mass function*.

Three common distributions for discrete variables are the *Bernoulli*, *Binomial* and *Poisson* distributions.

Bernoulli Distribution

A **Bernoulli trial** is a random process which has two possible outcomes, usually labelled "success" and "failure".

A random variable X is said to have a **Bernoulli** distribution with success probability p if X can only assume the values 0 and 1, with probabilities

$$\mathbb{P}(X = 1) = p \quad \text{and} \quad \mathbb{P}(X = 0) = 1 - p \, .$$

We write $X \sim \text{Ber}(p)$.

Can you think of some examples?

Properties of the Bernoulli Distribution

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

Expectation:

$$\mathbb{E}X = 0 \cdot (1-p) + 1 \cdot p = p.$$

Variance:

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}X^2 - (\mathbb{E}X)^2 \\ &= (0^2 \cdot (1-p) + 1^2 \cdot p) - p^2 = p(1-p)\end{aligned}$$

Moment generating function:

$$\begin{aligned}M(s) &= \mathbb{E}e^{sX} = e^{s\cdot 0} \cdot (1-p) + e^{s\cdot 1} \cdot p \\ &= 1 - p + pe^s\end{aligned}$$

for $s \in \mathbb{R}$.

Binomial Distribution

The **count of total successes** in a series of **independent** Bernoulli trials with a **constant probability of success**, has what is called the **Binomial distribution**.

If X is the total success count from n independent Bernoulli trials, each with success probability p, then we write $X \sim \text{Bin}(n, p)$.

Binomial Distribution

Suppose proportion p of all students live at home. If we take a random sample of 5 students, what is the probability that exactly 5 of them live at home?

1. $5p$
2. p^5 ✓
3. $(1-p)^5$
4. $5p^5$

Binomial Distribution

Suppose proportion p of all students live at home. If we take a random sample of 5 students, what is the probability that exactly 4 of them live at home?

1. p^4
2. $4p^5$
3. $4(1-p)^5$
4. $5p^4(1-p)$ ✓

Binomial Distribution

We can use the product rule to find that the probability of having a particular sequence with x successes and $n - x$ failures is $p^x(1-p)^{n-x}$.

There are $\binom{n}{x}$ sequences with x successes and $n - x$ failures.

The pmf of a **Binomial** distribution with parameters n and p is

$$\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n\,.$$

Properties of the Binomial Distribution

We could determine the expectation (mean), variance and moment generating function of the Binomial distribution directly from the definition. This can be done more easily once we see how to deal with multiple random variables.

Expectation: $\mathbb{E}X = np$.

Variance: $\mathrm{Var}(X) = np(1-p)$.

Moment generating function: $M(s) = (1 - p + pe^s)^n$, for $s \in \mathbb{R}$.

Questions

Suppose 70% of all students live at home. If we take a random sample of 5 students, what is the probability that 3 of them live at home?

Answer: $\mathbb{P}(X = 3) = \binom{5}{3} \times 0.7^3 \times (1 - 0.7)^{5-3} = 0.3087$

What is the expected number of students living at home from a sample of 5 students?

Answer: $\mathbb{E}X = 5 \times 0.7 = 3.5$

What is the standard deviation of the number of students living at home from a sample of 5 students?

Answer: $\mathrm{sd}(X) = \sqrt{5 \times 0.7 \times (1 - 0.7)} = \sqrt{1.05} = 1.024$

Questions

Suppose 70% of all students live at home. Let X be the number of students who live at home in a random sample of 10 students.

What is the mean and variance of $Y = \dfrac{X}{10}$?

(a) $\mathbb{E}(Y) = 7, \quad \mathrm{Var}(Y) = 0.21$

(b) $\mathbb{E}(Y) = 7, \quad \mathrm{Var}(Y) = 2.1$

(c) $\mathbb{E}(Y) = 0.7, \quad \mathrm{Var}(Y) = 2.1$

(d) $\mathbb{E}(Y) = 0.7, \quad \mathrm{Var}(Y) = 0.021$ ✓

Vinny from Vegas has received a "lucky" coin from "Tricky" Trish. Tricky has assured Vinny that the coin is biased towards heads. Vinny is skeptical, so decides to flip the coin 30 times with the aim to investigate Tricky's claim.

Vinny conducts his experiment and finds 21 heads out of 30 flips. Assuming the coin is fair, what is the probability of getting 21 or more heads. What should Vinny conclude?

Answer:

$$\mathbb{P}(X \geq 21) = 1 - \mathbb{P}(X \leq 20) = 1 - \overbrace{\sum_{x=0}^{20} \mathbb{P}(X = x)}^{F_X(20)} = 0.021.$$

There is moderate evidence that the coin is not fair.

Poisson Distribution

Many experiments consist of observing the number of times a random event occurs in a fixed interval of time or space, e.g.,

- arrivals of customers for service in a week,
- arrivals of calls at a switchboard overnight,
- the distribution of trees in a forest,
- the distribution of galaxies in a given region of the sky, and
- the number of fish caught in an afternoon in a certain region.

As long as these events occur at random and at a rate that does not change with time (or distance, area, volume, etc), the family of *Poisson distributions* can be used to model such random phenomena.

In essence, Poisson distribution describes the probability of a random event happening a certain number of times within a given interval of time or space.

Poisson Distribution

The Poisson distribution is used in many probability models and may be viewed as the "limit" of $\mathrm{Bin}(n; \lambda/n)$ for $\lambda > 0$ and large n.

Consider a coin tossing experiment where we toss a coin n times with success probability $\lambda/n \leq 1$. Let X be the number of successes. Then, as we have seen $X \sim \mathrm{Bin}(n, \lambda/n)$. In other words,

$$\mathbb{P}(X = x) = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \left(\frac{\lambda^x}{x!}\right) \left(\frac{n \times (n-1) \times \cdots \times (n-x+1)}{n^x}\right) \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}.$$

For a given x and taking the limit as $n \to \infty$, we see that the second and fourth terms in the second line converge to 1. The third term converges to $e^{-\lambda}$.

Poisson Distribution

A random variable X is said to have a **Poisson** distribution if its pmf is given by

$$\mathbb{P}(X = x) = \frac{\lambda^x e^{-\lambda}}{x!},$$

for $x \in \{0, 1, 2, \ldots\}$ and fixed $\lambda > 0$. We write $X \sim \text{Poisson}(\lambda)$.

Suppose $X \sim \text{Poisson}(\lambda)$. Then,

- $\mathbb{E}X = \lambda$
- $\text{Var}(X) = \lambda$
- $M(s) = e^{\lambda(e^s - 1)}, \quad s \in \mathbb{R}$

Poisson Distribution

We saw that Poisson distribution can be regarded as the limit of binomial distribution in some sense. According to this result, in any binomial experiment in which n is large and p is small, we can approximate $\text{Bin}(n, p)$ with $\text{Poisson}(np)$.

As a rule of thumb, this approximation can safely be applied if $n > 50$ and $np < 5$.