STA 103 Lecture 3: Discrete Random Variables

Instructor: Wookyeong Song

Department of Statistics, University of California, Davis

Aug 6th, 2025

Announcement

Update on Assignment Due Dates

Hi everyone,

To give you time to review solutions before each exam, all assignments will now be due **one day earlier** than originally scheduled. This allows me to post solutions **24 hours before the exam**, so you can use them as an additional study resource.

To balance the shorter timeline, I've removed one problem from Assignment 1:

- **Before:** 6 problems in 7 days
- **Now:** 5 problems in 6 days (similar to having 12 days in a regular quarter as summer session moves twice as fast as a regular quarter)

The removed problem will **not** appear on the exam.

Revised Due Dates:

- Assignment 1: **Aug 12 (Tue) Noon**
- Assignment 2: **Aug 25 (Mon) Noon**
- Assignment 3: **Sep 8 (Mon) Noon**

The syllabus and assignment have been updated accordingly, **please check the latest versions**. Thanks for your understanding, and I hope this helps you better prepare for the exams!

Bernoulli Random Variables

- **Definition**: A Bernoulli random variable takes on only two values: 0 and 1, with probabilities $1 - p$ and p, respectively.

- PMF in table form:

x	$p(x)$
0	$1 - p$
1	p

- PMF in Function Form:

$$p(x) = \begin{cases} p^x (1-p)^{1-x} & \text{if } x = 0 \text{ or } x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

- The Bernoulli random variable can be denoted by $X \sim \text{Bernoulli}(p)$.

Bernoulli Random Variables

For Bernoulli random variables:

- Expectation:

$$E(X) = \sum_x xp(x) = 0 \cdot (1-p) + 1 \cdot p = p.$$

- Variance:

$$E(X^2) = \sum_x x^2 p(x) = 0^2 \cdot (1-p) + 1^2 \cdot p = p,$$
$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p).$$

Binominal Random Variables

- **Definition**: A binominal random variable X counts the number of successes in n independent Bernoulli trials.
- Each trial has only two outcomes: "success" with probability p, and "failure" with probability $1 - p$.
- The total number of trials is fixed and denoted by n.
- The random variable X can take values $x = 0, 1, 2, \dots, n$.

Binomial Random Variables

- A binomial random variable is the sum of n independent and identically distributed **(i.i.d.)** Bernoulli random variables with the same success probability p:

$$Y = X_1 + X_2 + \cdots + X_n, \quad X_i \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p).$$

- The PMF of Binomial RV, getting exactly x successes in n trials, is

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}.$$

- The term $\binom{n}{x} = \frac{n!}{(n-x)!x!}$ counts the number of different ways to choose x successes out of n trials.

- The expression $p^x (1 - p)^{n-x}$ is the probability of any particular sequence with x successes and $n - x$ failures.

Binominal Random Variables

- How to calculate $E(X)$ and $\mathrm{Var}(X)$? Isn't it too complicated?

- Luckily here is the trick.
  - **Expectation**:

$$\begin{aligned} E(Y) &= E(X_1 + X_2 + \cdots + X_n) \\ &= E(X_1) + E(X_2) + \cdots + E(X_n). \end{aligned}$$

    The summation expansion of expectation always holds.

  - **Variance**:

$$\begin{aligned} \mathrm{Var}(Y) &= \mathrm{Var}(X_1 + X_2 + \cdots + X_n) \\ &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n). \end{aligned}$$

    The summation expansion of variance holds only if $X_1, \ldots, X_n$ are independent. (We will discuss independence in the next lecture!)

Binominal Random Variables

Then,

- **Expectation**:

$$\begin{aligned} E(Y) &= E(X_1 + X_2 + \cdots + X_n) \\ &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= n \times E(X_1) = np. \end{aligned}$$

- **Variance**:

$$\begin{aligned} \mathrm{Var}(Y) &= \mathrm{Var}(X_1 + X_2 + \cdots + X_n) \\ &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) \\ &= n \times \mathrm{Var}(X_1) = np(1-p). \end{aligned}$$

- We can denote it by $Y \sim$ Binomial(n, p). The Bernoulli(p) is a special case of Binomial(n, p) when $n = 1$.

Binominal Random Variables

- **Example** (Market Research Analyst): You are analyzing customer response behavior in an online survey conducted by a retail economics team. From past data, it is known that the probability a randomly selected customer responds to the survey is $p = 0.2$. You randomly contact $n = 50$ customers. Let the random variable X denote the number of customers who respond.

- **Question 1**: Identify the distribution of X. Justify your answer.

Binominal Random Variables

- **Question 2**: What is the expected number of responses and the standard deviation?
- **Question 3**: What is the probability that exactly 10 customers respond?
- **Question 4**: What is the probability that at most 2 customers respond?

Poisson Random Variable

- **Definition**: A Poisson random variable models the number of occurrences of an event in a fixed interval of time or space, under the assumptions that:
  - Events occur independently.
  - The average rate (events per interval) is constant.
  - Two events cannot occur at exactly the same time.

- Let X denote the number of events in a given interval and let $\lambda > 0$ be the expected number of events in that interval. Then X is said to follow a Poisson distribution with parameter λ, and its probability mass function is

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \text{ for } x = 0, 1, 2, \ldots$$

Poisson Random Variable

- We can denote it by $X \sim \text{Poisson}(\lambda)$.

- **Expectation and Variance**:

$$E(X) = \text{Var}(X) = \lambda.$$

Poisson Random Variable

Identifying a Poisson Random Variable: A random variable X can be modeled using a Poisson distribution if it satisfies the following conditions

- Counts of Events: X represents the number of times an event occurs in a fixed interval of time or space.
- Non-negative Integer Values: $X = 0, 1, 2, \ldots$
- Constant Rate: Events occur at a constant average rate λ over time or space.
- Rare Events: In a very small interval, the probability of more than one event occurring is negligible.

Poisson Random Variable

Examples that likely follow a Poisson distribution:

- Number of customer arrivals at a bank per hour.
- Number of website clicks or page views per second.
- Number of insurance claims filed by a policyholder per year.

Poisson Random Variable

- **Example**: A local bank observes that, on average, 6 customers arrive at the counter per hour. Assume customer arrivals follow a Poisson process. Let X be the number of customer arrivals in each hour. Then $X \sim \text{Poisson}(\lambda = 6)$.

- **Question 1**: What is the probability that exactly 4 customers arrive in one hour?

- **Question 2**: What is the probability that at most 2 customers arrive?

Poisson Random Variable

- **Example**: A local bank observes that, on average, 6 customers arrive at the counter per hour. Assume customer arrivals follow a Poisson process. Let X be the number of customer arrivals in each hour. Then $X \sim \text{Poisson}(\lambda = 6)$.

- **Question 3**: What is the probability that more than 1 customers arrive?

- **Question 4**: What are the mean, standard deviation and variance of X?

Conditional PMF

Suppose X is a random variable with PMF:

- PMF in a table form:

x	$p(x)$
1	0.5
2	0.1
3	0.1
4	0.2
5	0.1
	1

- Equivalently, we have the following box model:

5×	1×	1×	2×	1×
$\boxed{1}$	$\boxed{2}$	$\boxed{3}$	$\boxed{4}$	$\boxed{5}$

Box Model (10 tickets total)

Conditional PMF

- What if I sampled X but did not tell you which number it was but instead told you that $X \geq 3$.

- Then, the actual value of X is still random but now the possible values are $3, 4, 5$ instead of $1, 2, 3, 4, 5$.

- The conditional PMF given $X \geq 3$, denoted $p(x \mid X \geq 3)$, characterizes the randomness in X after observing $X \geq 3$.

- To find $p(x \mid X \geq 3)$, just zero out $p(x)$ for x's that are not observed and renormalize to 1.

Conditional PMF

- The original PMF in a table form:

x	$p(x)$
1	0.5
2	0.1
3	0.1
4	0.2
5	0.1
	1

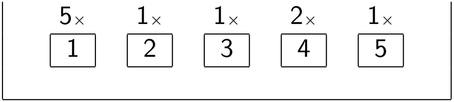
- After observing $X \geq 3$,

x	$p(x \mid X \geq 3)$
1	0
2	0
3	$0.1/0.4$
4	$0.2/0.4$
5	$0.1/0.4$
	$(0.1 + 0.2 + 0.1)/0.4$

Conditional PMF

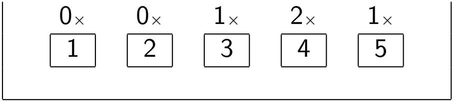
Equivalently,

- The original box model:



Box Model (10 tickets total)

- After observing $X \geq 3$,



Box Model (4 tickets total)

Conditional PMF

- **Question 1**: Find $P(X \geq 4)$?

x	$p(x)$
1	0.5
2	0.1
3	0.1
4	0.2
5	0.1
	1

- **Question 2**: Find $P(X \geq 4 \mid X \geq 3)$?

x	$p(x \mid X \geq 3)$
1	0
2	0
3	$0.1/$0.4
4	$0.2/$0.4
5	$0.1/$0.4
	$(0.1 + 0.2 + 0.1)/$0.4

Conditional PMF

- We can also compute the "new" expected value $E(X)$ after observing $X \geq 3$.

x	$p(x \mid X \geq 3)$
1	0
2	0
3	$0.1/0.4$
4	$0.2/0.4$
5	$0.1/0.4$
	$(0.1 + 0.2 + 0.1)/0.4$

$$
\begin{aligned}
E(X \mid X \geq 3) &= \sum_x x p(x \mid X \geq 3) \\
&= 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{2} + 5 \cdot \frac{1}{4} = \frac{3 + 8 + 5}{4} = 4.
\end{aligned}
$$

- **Interpretation** of $E(X \mid X \geq 3)$: Adjusted prediction after given new information that $X \geq 3$.

Conditional PMF

- Box average after observing $X \geq 3$.

$$\begin{array}{ccccc} 0\times & 0\times & 1\times & 2\times & 1\times \\ \boxed{1} & \boxed{2} & \boxed{3} & \boxed{4} & \boxed{5} \end{array}$$

Box Model (4 tickets total)

$$\text{Box Average} = \frac{3+4+4+5}{4} = 4.$$

- **Interpretation**: Box Average after observing $X \geq 3$ is identical to $E(X \mid X \geq 3)$.

Conditional PMF

- Similarly, the $\mathrm{Var}(X)$, and $\mathrm{sd}(X)$ after observing $X \geq 3$ are:

x^2	x	$p(x \mid X \geq 3)$
1^2	1	0
2^2	2	0
3^2	3	$0.1/0.4 = 1/4$
4^2	4	$0.2/0.4 = 1/2$
5^2	5	$0.1/0.4 = 1/4$
		$(0.1 + 0.2 + 0.1)/0.4$

$$\begin{aligned}\mathrm{Var}(X \mid X \geq 3) &= E(X^2 \mid X \geq 3) - (E(X \mid X \geq 3))^2 \\ &= 3^2 \cdot \frac{1}{4} + 4^2 \cdot \frac{1}{2} + 5^2 \cdot \frac{1}{4} - (4)^2 \\ &= \frac{9 + 32 + 25}{4} - 16 = 1/2.\end{aligned}$$

$$\mathrm{sd}(X \mid X \geq 3) = \sqrt{\mathrm{Var}(X \mid X \geq 3)} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2} \approx 0.7071.$$