Math Revision Session

Statistics (2): Discrete Random Variables and Famous Distributions

Jukina HATAKEYAMA

The University of Osaka, Department of Economics

April 7, 2026

Discrete Random Variable

A **discrete random variable** takes a finite or countable number of possible values.

Examples:

- the number of heads in repeated coin flips,
- the number of customers arriving at a store,
- the outcome of a die roll.

Unlike continuous random variables, a discrete random variable does not take every possible value in an interval.

Probability Mass Function (PMF)

Let X be a discrete random variable. Its **probability mass function** (PMF) is

$$p_X(x) = \Pr(X = x).$$

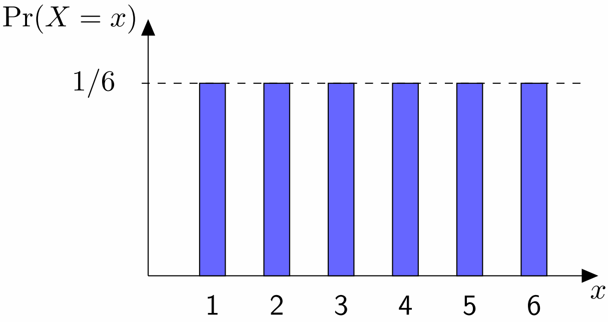
Example: if X is the outcome of a fair die, then

$$\Pr(X = x) = \begin{cases} \frac{1}{6}, & x \in \{1, 2, 3, 4, 5, 6\}, \\ 0, & \text{otherwise.} \end{cases}$$

The PMF satisfies:

$$0 \leq p_X(x) \leq 1, \qquad \sum_x p_X(x) = 1.$$

Example: PMF of a Fair Die



Cumulative Distribution Function (CDF)

The **cumulative distribution function** (CDF) of X is

$$F_X(x) = \Pr(X \leq x).$$

For a discrete random variable, the CDF is a non-decreasing step function.

Properties:

- $0 \leq F_X(x) \leq 1$,
- $F_X(-\infty) = 0$,
- $F_X(\infty) = 1$,
- if $x_1 < x_2$, then $F_X(x_1) \leq F_X(x_2)$.

Example: CDF of a Fair Die

For a fair die,

$$F_X(x) = \Pr(X \leq x).$$

For example,

$$F_X(2) = \Pr(X \leq 2) = \Pr(X = 1) + \Pr(X = 2) = \frac{2}{6}.$$

More generally,

$$F_X(x) = \begin{cases} 0, & x < 1, \\ 1/6, & 1 \leq x < 2, \\ 2/6, & 2 \leq x < 3, \\ 3/6, & 3 \leq x < 4, \\ 4/6, & 4 \leq x < 5, \\ 5/6, & 5 \leq x < 6, \\ 1, & x \geq 6. \end{cases}$$

Expectation and Variance

For a discrete random variable X with PMF $p_X(x)$,

$$E[X] = \sum_x x \, p_X(x),$$

and

$$\mathrm{Var}(X) = E[(X - E[X])^2] = \sum_x (x - E[X])^2 p_X(x).$$

The standard deviation is

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

Examples of Expectation and Variance

Fair die:

$$E[X] = \sum_{x=1}^{6} x \cdot \frac{1}{6} = 3.5, \qquad \mathrm{Var}(X) = \frac{35}{12} \approx 2.92.$$

Bernoulli trial: If $X \in \{0, 1\}$ with

$$\Pr(X = 1) = p, \qquad \Pr(X = 0) = 1 - p,$$

then

$$E[X] = p, \qquad \mathrm{Var}(X) = p(1 - p).$$

Joint Probability Function

For two discrete random variables X and Y, the **joint probability function** is

$$p_{X,Y}(x, y) = \Pr(X = x, Y = y).$$

Example: if X and Y are the outcomes of two fair dice, then

$$\Pr(X = 1, Y = 3) = \frac{1}{36}.$$

Marginal Probability Function

The **marginal probability function** of X is obtained by summing the joint probability over all possible values of Y:

$$p_X(x) = \Pr(X = x) = \sum_y p_{X,Y}(x, y).$$

Similarly,

$$p_Y(y) = \sum_x p_{X,Y}(x, y).$$

Example:

$$\Pr(X = 1) = \sum_{y=1}^{6} \Pr(X = 1, Y = y) = \frac{6}{36} = \frac{1}{6}.$$

Covariance

The covariance between X and Y is

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

It measures whether the two variables tend to move together.

- If $\mathrm{Cov}(X, Y) > 0$, large values of X tend to be associated with large values of Y.
- If $\mathrm{Cov}(X, Y) < 0$, large values of X tend to be associated with small values of Y.

Correlation Coefficient

The correlation coefficient between X and Y is

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

It measures the strength and direction of the linear relationship between X and Y.

- $\rho = 1$: perfect positive linear relationship,
- $\rho = -1$: perfect negative linear relationship,
- $\rho = 0$: no linear relationship.

Independence of Random Variables

Two random variables X and Y are **independent** if

$$\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y) \quad \text{for all } x, y.$$

If X and Y are independent, then

$$E[XY] = E[X]E[Y].$$

Also, independence implies

$$\mathrm{Cov}(X, Y) = 0.$$

Correlation and Independence

If X and Y are independent, then

$$\mathrm{Cov}(X,Y)=0 \qquad \text{and hence} \qquad \rho(X,Y)=0.$$

However, the converse is not true:

$$\rho(X,Y)=0 \quad \not\Rightarrow \quad X \text{ and } Y \text{ are independent.}$$

Zero correlation means no **linear** relationship, but nonlinear dependence may still exist.

Conditional Probability

For events A and B with $\Pr(B) > 0$,

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

Example: if we know that the first die shows 4, then

$$\Pr(\text{second die} = 6 \mid \text{first die} = 4) = \frac{\Pr(\text{first die} = 4, \text{second die} = 6)}{\Pr(\text{first die} = 4)} = \frac{1/36}{1/6} = \frac{1}{6}.$$

Conditional Expectation

For discrete random variables X and Y, the conditional expectation of X given $Y = y$ is

$$E[X \mid Y = y] = \sum_x x \, \Pr(X = x \mid Y = y).$$

It gives the expected value of X when we know the value of Y.

Example of Conditional Expectation

Suppose X and Y are the outcomes of two independent fair dice. Then

$$E[X \mid Y = 6] = \sum_{x=1}^{6} x \Pr(X = x \mid Y = 6).$$

Because X and Y are independent,

$$\Pr(X = x \mid Y = 6) = \Pr(X = x) = \frac{1}{6}.$$

Hence

$$E[X \mid Y = 6] = \sum_{x=1}^{6} x \cdot \frac{1}{6} = 3.5.$$

Linearity of Expectation

For random variables X and Y, and constants a, b, c,

$$E[aX + bY + c] = aE[X] + bE[Y] + c.$$

This rule holds whether or not X and Y are independent.

Expectation of a Product

If X and Y are independent, then

$$E[XY] = E[X]E[Y].$$

This does **not** necessarily hold when X and Y are dependent.

Conditional Expectation Rules

Some useful rules:

- $$E[X+Y \mid Z] = E[X \mid Z] + E[Y \mid Z].$$
- **Law of total expectation (tower property):**
  $$E[X] = E[E[X \mid Z]].$$
- If X and Y are conditionally independent given Z, then
  $$E[XY \mid Z] = E[X \mid Z]E[Y \mid Z].$$

Variance Rules

Recall

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2.$$

Also:

$$\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X),$$

and

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y).$$

If X and Y are independent, this simplifies to

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Examples of Variance Rules

If $Y = 3X + 2$, then

$$\mathrm{Var}(Y) = 9\,\mathrm{Var}(X).$$

If $\mathrm{Var}(X) = 4$, then

$$\mathrm{Var}(Y) = 36.$$

If X and Y are independent with

$$\mathrm{Var}(X) = 5, \qquad \mathrm{Var}(Y) = 3,$$

then

$$\mathrm{Var}(X + Y) = 5 + 3 = 8.$$

Bernoulli Distribution

A random variable X follows a **Bernoulli distribution** with parameter p if

$$\Pr(X = 1) = p, \qquad \Pr(X = 0) = 1 - p, \qquad 0 \leq p \leq 1.$$

We write

$$X \sim \mathrm{Bernoulli}(p).$$

Its PMF is

$$\Pr(X = x) = p^x (1 - p)^{1-x}, \qquad x \in \{0, 1\}.$$

CDF, Expectation, and Variance of Bernoulli

If $X \sim \text{Bernoulli}(p)$, then the CDF is

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - p, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

Also,

$$E[X] = p, \qquad \text{Var}(X) = p(1 - p).$$

Relation to the Binomial Distribution

The Bernoulli distribution is a special case of the binomial distribution:

$$X \sim \mathrm{Bernoulli}(p) \quad \Longleftrightarrow \quad X \sim \mathrm{Bin}(1, p).$$

Example:

- one coin flip $\rightarrow$ Bernoulli,
- number of heads in 10 flips $\rightarrow$ Binomial.

Binomial Distribution

A random variable X follows a **binomial distribution** if it counts the number of successes in n independent Bernoulli trials with success probability p.
We write

$$X \sim \mathrm{Bin}(n, p).$$

Its PMF is

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

Why Do We Use Combinations?

In the binomial distribution, we count the number of successes, not the order in which they occur.
For example, when flipping a coin 5 times, the outcomes

$$HHTTT, \quad HTHTT, \quad TTHHT$$

all correspond to 2 heads.
So we multiply by the number of possible arrangements:

$$\binom{n}{k}.$$

Example of a Binomial Probability

Suppose $X \sim \mathrm{Bin}(5, 0.6)$. Then

$$\Pr(X = 3) = \binom{5}{3}(0.6)^3(0.4)^2.$$

Since

$$\binom{5}{3} = 10,$$

we get

$$\Pr(X = 3) = 10 \times 0.216 \times 0.16 = 0.3456.$$

Expectation and Variance of the Binomial Distribution

If $X \sim \mathrm{Bin}(n, p)$, then

$$E[X] = np, \qquad \mathrm{Var}(X) = np(1-p).$$

The binomial distribution is widely used for count data, such as:

- number of correct answers,
- number of defective products,
- number of people with a certain characteristic.

Poisson Distribution

The **Poisson distribution** models the number of events occurring in a fixed interval of time or space.
If

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

then its PMF is

$$\Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \qquad k = 0, 1, 2, \ldots,$$

where $\lambda > 0$ is the average rate of occurrence.

Expectation and Variance of the Poisson Distribution

If $X \sim \mathrm{Poisson}(\lambda)$, then

$$E[X] = \lambda, \qquad \mathrm{Var}(X) = \lambda.$$

So the mean and variance are equal in the Poisson distribution.

Example of a Poisson Probability

Suppose $X \sim \text{Poisson}(3)$. Then

$$\Pr(X = 2) = \frac{3^2 e^{-3}}{2!} = \frac{9e^{-3}}{2} \approx 0.2241.$$

Relation to the Binomial Distribution

The Poisson distribution can be obtained as an approximation to the binomial distribution when:

- n is large,
- p is small,
- $np = \lambda$ remains fixed.

In that case,

$$\mathrm{Bin}(n, p) \approx \mathrm{Poisson}(\lambda).$$

Example:

$$\mathrm{Bin}(1000, 0.005) \approx \mathrm{Poisson}(5).$$

Example Problem

A machine produces defective products 5 times out of every 1000 items.
If the machine produces 50 items, approximate the probability of producing no defective items.

Let X be the number of defective items. Then the rate is

$$\lambda = 50 \times \frac{5}{1000} = 0.25.$$

Using a Poisson approximation,

$$X \sim \mathrm{Poisson}(0.25).$$

Therefore,

$$\Pr(X = 0) = \frac{0.25^0 e^{-0.25}}{0!} = e^{-0.25} \approx 0.7788.$$

Summary

In this lecture, we studied:

- PMF and CDF,
- expectation and variance of discrete random variables,
- joint and marginal distributions,
- covariance, correlation, and independence,
- conditional probability and conditional expectation,
- rules for expectation and variance,
- Bernoulli, Binomial, and Poisson distributions.