Lecture 2
Discrete Random Variables

Chao Song

College of Ecology
Lanzhou University

September 15, 2023

Motivating examples

Example 1: Let the random experiment be throwing a die. The sample space associated with this experiment is $S = \{1, 2, 3, 4, 5, 6\}$, with elements of S indicating the number of spots on the side facing up. Let X be a quantity that denotes the number on the die. X is a quantity whose number is determined by the outcome of a random experiment.

Example 2: A rat is selected at random from a cage and its sex is determined. The sample space is thus $S = \{\textit{female}, \textit{male}\}$. Let X be a quantity whose value is 1 if the rat is male and 0 if the rate is female. Same as above, X is a quantity whose value is determined by the outcome of a random experiment.

Definition of random variables

Definition: Given a random experiment with an sample space S, a function X that assigns one and only one real number $X(s) = x$ to each element s in S is called a random variable. The space of X is the set of real number $\{x : X(s) = x, s \in S\}$

Example 1: Let the random experiment be throwing a die. The sample space associated with this experiment is $S = \{1, 2, 3, 4, 5, 6\}$, with elements of S indicating the number of spots on the side facing up. Let $X(s) = s$, the space of the random variable X is $\{1, 2, 3, 4, 5, 6\}$.

Example 2: A rat is selected at random from a cage and its sex is determined. The sample space is $S = \{\mathit{female}, \mathit{male}\}$. Let X be a function such that $X(F) = 0$ and $X(M) = 1$. X is a random variable with space $\{0, 1\}$

Definition of random variables

A few remarks on the definition of random variable:

- Intuitively, we may view random variable as a quantity whose value is determined by the outcome of an random experiment. For practical purpose, this intuitive interpretation of random variable is sufficient;
- Rigorously, a random variable is a function that maps the outcome of a random experiment to real numbers. This is mainly for mathematical rigor.
- Roughly speaking, because probability is a measure mapping events to unit interval. The argument of probability is events. Thus, if we are going to define probability for random variable, we must be able to interpret $\{X \leqslant x\}$ as an event.
- How to map outcome of random experiment to a real number is not a trivial mathematical question. In practice, the choice is often made based on intuition or convenience.

Discrete random variables

Discrete random variable: a random variable is discrete if it only takes values that are in some countable subsets $\{x_1, x_2, \ldots\}$ of real number.

- Number of heads in 10 coin flips;
- Number of coin flips until we have two heads;
- Species richness in a country;
- Number of students late to this class each week.

Probability mass function

Definition: The probability mass function (PMF) of a discrete random variable X is the function $f(x) : \mathbb{R} \to [0, 1]$ given by $f(x) = P(X = x)$.

Properties of probability mass function:

- $0 \leqslant f(x) \leqslant 1$ for all x;
- $f(x) = 0$ if $x \notin \{x_1, x_2, \ldots\}$;
- $\sum_x f(x) = 1$.

Example: Let X be the number of heads when tossing two fair coins. What is the probability mass function for random variable X?

Answer: possible number of heads are 0, 1, 2. The PMF of X is

- $P(X = 0) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$;
- $P(X = 1) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$;
- $P(X = 2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$;

Cumulative distribution function

Definition: The cumulative distribution function (CDF) of a random variable X is the function $F(x) : \mathbb{R} \to [0, 1]$ given by $F(x) = P(X \leqslant x)$

Properties of cumulative distribution function:

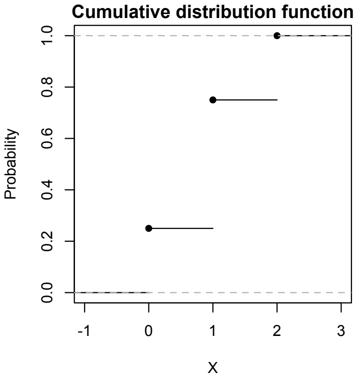
- $F(x)$ is a non-decreasing function: if $x < y$, then $F(x) \leqslant F(y)$;
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to -\infty} F(x) = 1$;
- $F(x)$ is right-continuous.

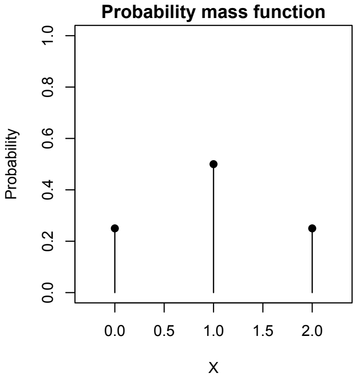
Proposition: Consider real numbers x and y with $x < y$, then

- $P(X > x) = 1 - F(x)$;
- $P(x < X \leqslant y) = F(y) - F(x)$;

Visualize PMF and CDF

Example: Let X be the number of heads when tossing two fair coins. Possible values for X are 0, 1, and 2. The PMF and CDF of X are:





Mathematical expectation

In addition to PMF and CDF, which fully characterize the distribution of a random variable, **mathematical expectation** is an important concept in summarizing characteristics of distribution of probability.

Definition: if $f(x)$ is the probability mass function of the discrete random variable X with space S, and if the summation $\sum_{x \in X} u(x)f(x)$ exists, then the sum is called the mathematical expectation or the expected value of $u(x)$, and it is denoted $E[u(x)]$

Example: Let X be the number of heads when tossing two coins. What is the expected value of X? If one gets two points for each head, what is the expected value of points?
Answer: The expected value of number of heads is

$$E(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1.$$

Let $u(x)$ be the points one get after tossing two coins, $u(x) = 2x$, then

$$E[u(x)] = (2 \times 0) \times \frac{1}{4} + (2 \times 1) \times \frac{1}{2} + (2 \times 2) \times \frac{1}{4} = 2.$$

Properties of mathematical expectation

When exists, the mathematical expectation satisfies the following properties:

- If c is a constant, then $E(c) = c$;
- If c is a constant, $E[cu(x)] = cE[u(x)]$;
- if c_1 and c_2 are constant, $E[c_1 u_1(x) + c_2 u_2(x)] = c_1 E[u_1(x)] + c_2 E[u_2(x)]$.

The above properties arise from the fact that **mathematical expectation is a linear operation**. Thus nonlinear operations cannot be applied the same way. For example, $E(x^2) \neq [E(x)]^2$ in general.

Mean and variance

Mean and **variance** are special cases of the mathematical expectation. Let X be a discrete random variable with probability mass function $f(x)$

- Mean: $\mu = E(X) = \sum_{x \in S} x f(x)$;
- Variance: $\sigma^2 = Var(X) = E[(X - \mu)^2] = \sum_{x \in S} (x - \mu)^2 f(x)$

Variance can be calculated in another way:

$$\begin{aligned}\sigma^2 &= E[(x - \mu)^2] = E[X^2 - 2\mu X + \mu^2] \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2\end{aligned}$$

Mean and variance

Properties of mean and variance: Let X be a random variable with mean μ and variance σ^2. Let a and b be constants. What is the mean and variance of $aX + b$?

Based on the property of mathematical expectation, we have

- $E(aX + b) = aE(X) + b = a\mu + b$;
- $Var(aX + b) = E[(aX + b - a\mu - b)^2] = E[a^2(X - \mu)^2] = a^2\sigma^2$

Bernoulli distribution

A **Bernoulli trial** is a random experiment, the outcome of which can be classified in on of the two mutually exclusive and exhaustive ways–say, success of failure. Let X be a random variable associated with a Bernoulli trial such that $X = 1$ for success and $X = 0$ for failure, X follows a **Bernoulli distribution**.

The probability mass function of X following a Bernoulli distribution is

$$f(x) = \begin{cases} p, & X = 1 \\ 1 - p, & X = 0 \end{cases},$$

Or more concisely, $f(x) = p^x(1-p)^{1-x}$.

The mean and variance of a Bernoulli distribution is

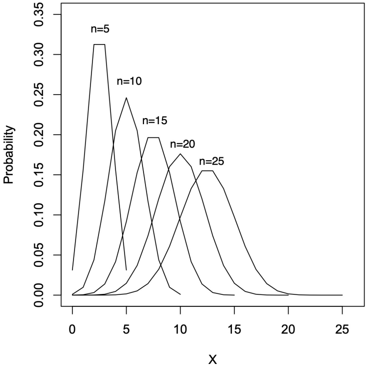
- $E(X) = 1 \times p + 0 \times (1-p) = p$
- $Var(X) = (1-p)^2 p + (0-p)^2(1-p) = p(1-p)$

Binomial distribution

Binomial distribution: If a random variable X denotes the number of successes in n independent Bernoulli trials, X follows a binomial distribution and its PMF is

$$P(X = k) = \mathbf{C}_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n$$

The mean and variance of X are np and $np(1-p)$ respectively.

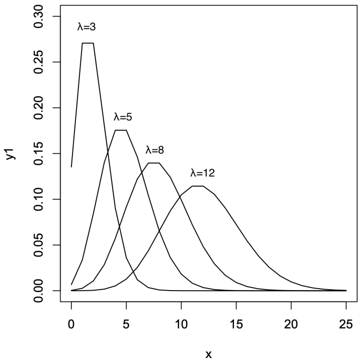


Poisson distribution

Poisson distribution: Let λ be a positive number. A random variable is said to have a Poisson distribution if its probability mass function is

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

The mean and variance of a Poisson distribution are both λ.



Poisson distribution

What does a Poisson distributed variable model?

Poisson distribution models the number of events occurring in a time interval t.

- Divide t into n segments such that at most one event occur within a segment;
- Probability of occurrence is $\mu t/n$;
- Number of occurrence is modeled with a binomial distribution.

$$\begin{aligned}
P(X=k) &= \lim_{n\to\infty} \mathbf{C}_n^k p^k (1-p)^{n-k} \\
&= \lim_{n\to\infty} \frac{n!}{k!(n-k)!} (\frac{\mu t}{n})^k (1-\frac{\mu t}{n})^{n-k} \\
&= \lim_{n\to\infty} \frac{(\mu t)^k}{k!} \frac{n(n-1)\ldots(n-k+1)}{n^k} (1-\frac{\mu t}{n})^{-k} (1-\frac{\mu t}{n})^{n} \\
&= \frac{(\mu t)^k}{k!} e^{-\mu t}
\end{aligned}$$

Poisson distribution is a limiting case of a binomial distribution. Here, $\lambda = \mu t$ is often referred to as the rate parameter of the Poisson distribution.

Negative binomial and geometric distribution

Negative binomial distribution: In a sequence of independent Bernoulli trials with success probability p, let X be the number of failure until r successes. Then X has a negative binomial distribution with probability mass function

$$P(X = k) = \mathbf{C}_{k+r-1}^{k}(1-p)^k p^r$$

The mean of X is $\frac{r(1-p)}{p}$ and the variance of X is $\frac{r(1-p)}{p^2}$. An important feature of negative binomial distribution is that the variance is larger than the mean.

Geometric distribution: In a sequence of independent Bernoulli trials with success probability p, let X be the total number of trials until we have1 successes, X has a geometric distribution with probability mass function:

$$P(X = k) = (1-p)^{k-1}p$$

The mean and variance of the geometric distribution is $\frac{1}{p}$ and $\frac{1-p}{p^2}$, respectively.

Summary of common discrete distributions

Distribution	Probability mass function	Mean	Variance
Bernoulli	$p^x(1-p)^{1-x}$	p	$p(1-p)$
Binomial	$\mathbf{C}_n^k p^k (1-p)^{n-k}$	np	$np(1-p)$
Poisson	$\frac{\lambda^k}{k!} e^{-\lambda}$	λ	λ
Negative binomial	$\mathbf{C}_{k+r-1}^k (1-p)^k p^r$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
Geometric	$(1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$