Lecture 7: Random Variables and Probability Distributions

Ye Tian

Department of Statistics, Columbia University
Calculus-based Introduction to Statistics (S1201)

July 14, 2022

Recap: independence and conditional independence

Independence:

- Two **events** A and B are **independent** if $\mathbb{P}(A|B) = \mathbb{P}(A)$, denoted as $A \perp\!\!\!\perp B$. They are **dependent** otherwise, denoted as $A \not\perp\!\!\!\perp B$.
- **Events** $A_1, \ldots, A_n$ are **(mutually) independent** if for every $k = 2, 3, \ldots, n$ and every subset of indices $i_1, i_2, \ldots, i_k$,

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k}).$$

Conditional independence:

- Two **events** A and B are **independent** conditioning on event C if $\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C)$, denoted as $(A \perp\!\!\!\perp B)|C$. They are **dependent** otherwise, denoted as $(A \not\perp\!\!\!\perp B)|C$.
- **Events** $A_1, \ldots, A_n$ are **(mutually) independent** conditioning on event C if for every $k = 2, 3, \ldots, n$ and every subset of indices $i_1, i_2, \ldots, i_k$,

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}|C) = \mathbb{P}(A_{i_1}|C)\mathbb{P}(A_{i_2}|C) \cdots \mathbb{P}(A_{i_k}|C).$$

Recap: probability calculation

$\mathbb{P}(\cdot) : A \in$ a set of all events $\rightarrow$ a number $\mathbb{P}(A)$ (a mapping/function which maps an event to a number)

Three axioms:

- $0 \leq \mathbb{P}(A) \leq 1$ for any event $A \subseteq \Omega$
- $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$
- If A_1, A_2, A_3, ..., A_n is a collection of disjoint (mutually exclusive) events, then

$$\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Theorem: In an experiment consisting of N outcomes **with equal probability**, for any event A,

$$\mathbb{P}(A) = \frac{N(A)}{N},$$

where we use counting techniques to calculate $N(A)$ and N.

Recap: probability calculation

General idea to calculate probability:

(1) Translate the conditions and the event of interest by probability language

(2) Calculate.

- For equally likely outcomes, consider $\mathbb{P}(A) = \frac{N(A)}{N}$ combined with counting techniques
- If conditional probability is easier to calculate or it is given, consider multiplication rule, rule of total probability, and Bayes' Theorem
- See whether there are independence/conditional independence or not, which may help us to do the calculation

Today's goal

- Understand the random variable and know how to use it in practice
- Understand the probability distribution, probability mass function (pmf), probability density function (pdf), and cumulative distribution function (cdf)
- Know the difference between discrete and continuous random variables
- See some examples of discrete and continuous random variables

Random Variables

Why do we need random variables

It seems that we have covered everything about probability...

- Frequently, we are interested in some **numerical aspects** of the outcome
  Example:
  - In political poll, the number of people voting for Trump among 100
  - The number of heads when flipping a coin 10 times
- And sometimes we are interested in the probability of many events instead of a single one, where we want to systematically give a formula for every case instead of studying the event separately
  Example:
  - The number of people voting for Trump among 100 (denoted as X): $\mathbb{P}(\{\text{the number} = 20\}) = ?$, $\mathbb{P}(\{\text{the number} = 50\}) = ?$, $\mathbb{P}(\{\text{the number} = 100\}) = ?$
  - The number of heads when flipping a coin 10 times (denoted as X): $\mathbb{P}(\{\text{the number} = 5\}) = ?$, $\mathbb{P}(\{\text{the number} = 0\}) = ?$, $\mathbb{P}(\{\text{the number} = 10\}) = ?$

Random variables

Definition: Given an experiment and the sample space Ω, a **random variable** is a **function** mapping a outcome ($\omega \in \Omega$) into a real number, i.e.

$$X : \omega \in \Omega \mapsto X(\omega) \in (-\infty, +\infty).$$

Example 1: Toss a coin 3 times: the sample space is $\Omega = \{\text{H, T}\} \times \{\text{H, T}\} \times \{\text{H, T}\}$. Random variable X = the number of heads.

Outcomes (w)	HHH	HTH	THH	HHT	HTT	THT	TTH	TTT
$X(\omega) =$	3		2			1		0

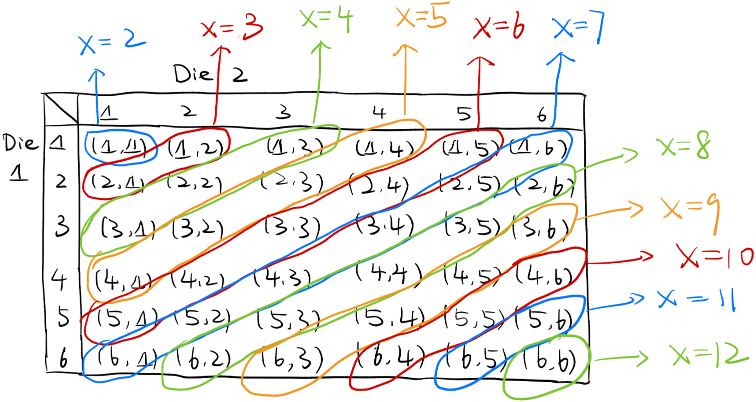
For instance: $X(\{\text{HHH}\}) = 3$, $X(\{\text{THH}\}) = X(\{\text{HHT}\}) = 2$, $X(\{\text{TTT}\}) = 0$

Random variables

Definition: Given an experiment and the sample space Ω, a **random variable** is a **function** mapping a outcome ($\omega \in \Omega$) into a real number, i.e.

$$X : \omega \in \Omega \mapsto X(\omega) \in (-\infty, +\infty).$$

Example 2: Toss two dice: the sample space is $\Omega = \{1, 2, \ldots, 6\} \times \{1, 2, \ldots, 6\}$. Random variable X = the sum of dice.



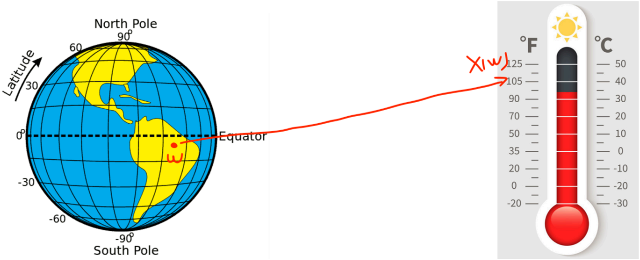
For instance: $X(\{(1,3)\}) = 4$, $X(\{(4,5)\}) = 9$, $X(\{(6,6)\}) = 12$

Random variables

Definition: Given an experiment and the sample space Ω, a **random variable** is a **function** mapping a outcome ($\omega \in \Omega$) into a real number, i.e.

$$X : \omega \in \Omega \mapsto X(\omega) \in (-\infty, +\infty).$$

Example 3: Suppose that we select a location at random (defined by latitude and longitude) and define X to be the temperature at that location at the current time.



Discrete and continuous random variables

- When the possible values of of a random variable are countable[1], the random variable is **discrete**.
  Examples: the number of heads/tails of coin flipping, the number of dice etc.
- When *both* of the following apply, the random variable is **continuous**.
  - The range is uncountable (e.g.: an interval on the number line)
  - No possible value of the variable has positive probability, i.e. $\mathbb{P}(X = c) = 0$ for any number c.

  Examples: the temperature at a random location

[1]either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on ("countably" infinite)

Exercise

Describe the set of possible values for the variable, and state whether the variable is discrete.

(1) $X =$ the number of unbroken eggs in a randomly chosen standard egg carton

(2) $Y =$ the number of students on a class list for a particular course who are absent on the first day of classes

(3) $U =$ the number of times a duffer has to swing at a golf ball before hitting it

(4) $X =$ the length of a randomly selected rattlesnake

(5) $Z =$ the sales tax percentage for a randomly selected Amazon purchase

(6) $Y =$ the pH of a randomly chosen soil sample

(7) $X =$ the tension (psi) at which a randomly selected tennis racket has been strung

(8) $X =$ the total number of times three tennis players must spin their rackets to obtain something other than UUU or DDD (to determine which two play next)

Random variables and random events

Compare their definitions:

- A **random variable** is a **function** mapping a outcome ($\omega \in \Omega$) into a real number, i.e. $X : \omega \in \Omega \mapsto X(\omega) \in (-\infty, +\infty)$.
- An **event** is a set (collection) of outcomes.

Furthermore, for any subset $\mathcal{B}$ on the number line [2], $\{\omega : X(\omega) \in \mathcal{B}\}$ is a random event, which is a set (collection) of outcomes. And we can calculate the corresponding probability.

Therefore, $\mathbb{P}(X \in \mathcal{B}) = \mathbb{P}(\{\omega : X(\omega) \in \mathcal{B}\})$.

[2] Actually not "any" subset, but the current statement is enough and correct for this course. You will learn more in a PhD-level probability course in the future. Currently, $\mathcal{B}$ can be any union/intersection of intervals/points on the real line. E.g., $\mathcal{B}$ can be $(0,1)$, $[-2,+\infty)$, $(5,5.5]$, $\{1\}$, $\{-1,2.5\}$, $(-3,-1) \cup (9,10]$ etc.

Random variables and random events

Let's recall our previous example:

Toss a *fair* coin 3 times: the sample space is $\Omega = \{\text{H, T}\} \times \{\text{H, T}\} \times \{\text{H, T}\}$. Random variable $X =$ the number of heads.

Outcomes (w)	HHH	HTH	THH	HHT	HTT	THT	TTH	TTT
$X(\omega) =$	3		2			1		0

Therefore,

$$\mathbb{P}(X = 3) = \mathbb{P}(\{\text{HHH}\}) = \frac{1}{8},$$

$$\mathbb{P}(X = 2) = \mathbb{P}(\{\text{HTH, THH, HHT}\}) = \frac{3}{8},$$

$$\mathbb{P}(X = 1) = \mathbb{P}(\{\text{HTT, THT, TTH}\}) = \frac{3}{8},$$

$$\mathbb{P}(X = 0) = \mathbb{P}(\{\text{TTT}\}) = \frac{1}{8}.$$

Distribution of Random Variables

Distribution

Definition: The **(probability) distribution** of a random variable X describes how the total probability of 1 is **distributed** among all possible values of X. It tells us $\mathbb{P}(X \in \mathcal{B}) = \mathbb{P}(\{\omega : X(\omega) \in \mathcal{B}\})$ for any subset $\mathcal{B}$ of number line [3].

Definition: **Cumulative distribution function (cdf)** of a r.v. X is defined as

$$F(x) = \mathbb{P}(X \leq x)$$

for any number x (including $-\infty$ and $+\infty$).

Proposition: The **cdf** can describe the distribution of random variables.

Why? (Not need to know): Because any subset $\mathcal{B}$ of the real line can be expressed as the union/intersection/difference of intervals like $(-\infty, x]$.
E.g.: $(5, 10] = (-\infty, 10] \backslash (-\infty, 5])$
$\Rightarrow$ Then $\mathbb{P}(X \in (5, 10]) \overset{\text{why?}}{=} \mathbb{P}(X \leq 10) - \mathbb{P}(X \leq 5) = F(10) - F(5)$.

[3] Currently, $\mathcal{B}$ can be any union/intersection of intervals/points on the real line. E.g., $\mathcal{B}$ can be $(0, 1)$, $[-2, +\infty)$, $(5, 5.5]$, $\{1\}$, $\{-1, 2.5\}$, $(-3, -1) \cup (9, 10]$ etc.

Distribution

Definition: Cumulative distribution function (cdf) of a r.v. X is defined as

$$F(x) = \mathbb{P}(X \leq x)$$

for any number x (including $-\infty$ and $+\infty$).

Remark:

- $F(+\infty) = 1$
  because $(-\infty, +\infty] \supseteq \Omega \Rightarrow F(+\infty) = \mathbb{P}((-\infty, +\infty]) \geq \mathbb{P}(\Omega) = 1$
- $F(-\infty) = 0$
  because $(-\infty, -\infty)$ behaves like an empty set [4]
  $\Rightarrow F(-\infty) = \mathbb{P}((-\infty, -\infty)) = 0$
- Therefore $0 \leq F(x) \leq 1$ for any number x
- $F(x)$ is an increasing function, i.e. for $x_1 \leq x_2$, $F(x_1) \leq F(x_2)$ (why?)
- (Not required to know) $F(x)$ is right-continuous, i.e. $\lim_{z \to x+0} F(z) = F(x)$

[4] not accurate, but enough for this course

Distribution

Definition: For a discrete r.v. X, its distribution can also be described by **probability mass function (pmf)**

$$p(x) = \mathbb{P}(X = x) = \mathbb{P}(\{\omega : X(\omega) = x\})$$

for any number x (including $-\infty$ and $+\infty$).

- For discrete r.v., suppose $S = \{z_1, z_2, z_3, \ldots\}$ including all possible values of X, then:
  - $p(x) > 0$ only when $x \in S$, and $p(x) = 0$ elsewhere
  - $F(x) = \sum_{i: z_i \leq x} p(z_i)$
  - $p(z_i) = F(x_2) - F(x_1)$ for any x_1 and x_2 with $(x_1, x_2] \cap S = \{z_i\}$
- Two conditions for a valid pmf:

  (1) $p(x) \geq 0$ for any x;

  (2) $\sum_{x \in S} p(x) = 1$.
- It's *senseless* to talk about pmf of *continuous* r.v., because $\mathbb{P}(X = x) = 0$ for any number x if X is continuous! (will see that from the view of integral)

cdf and pmf of discrete random variables

cdf: $F(x) = \mathbb{P}(X \leq x)$
pmf: $p(x) = \mathbb{P}(X = x)$

For discrete r.v., suppose $S = \{z_1, z_2, z_3, \ldots\}$ including all possible values of X, then:

- $p(x) > 0$ only when $x \in S$, and $p(x) = 0$ elsewhere
- $F(x) = \sum_{i: z_i \leq x} p(z_i)$
- $p(z_i) = F(x_2) - F(x_1)$ for any x_1 and x_2 with $(x_1, x_2] \cap S = \{z_i\}$
- $\mathbb{P}(x_1 < X \leq x_2) = F(x_2) - F(x_1) = \sum_{i: x_1 < z_i \leq x_2} p(z_i)$

Example: The pmf of a discrete r.v. is

x	0	1	2	3	4	5	6
$p(x)$	.05	.10	.15	.25	.20	.15	.10

$F(2) = \mathbb{P}(X \leq 2) = p(0) + p(1) + p(2) = 0.05 + 0.1 + 0.15 = 0.3$, $F(1) = \mathbb{P}(X \leq 1) = p(0) + p(1) = 0.05 + 0.1 = 0.15$,
$\mathbb{P}(1 \leq X \leq 3) = p(1) + p(2) + p(3) = 0.1 + 0.15 + 0.25 = 0.5$

Example of discrete distribution: binomial distribution

Toss a *unfair* coin 10 times. Suppose each time the probabilities of heads and tails are p and $1-p$, respectively. Random variable $X =$ the number of heads.

- $\mathbb{P}(X=0) = \mathbb{P}(\{\text{TTTTTTTTTT}\}) = (1-p)^{10}$
- $\mathbb{P}(X=1) = \mathbb{P}(\{9 \text{ T's and } 1 \text{ H}\}) = \binom{10}{1} p(1-p)^9$
- ...
- In general, the pmf
  $p(x) = \mathbb{P}(X=x) = \mathbb{P}(\{(10-x) \text{ T's and } x \text{ H's}\}) = \binom{10}{x} p^x (1-p)^{10-x}$, $x = 0, \ldots, 10$.
- The cdf $F(x) = \sum_{k:0 \le k \le x} p(k) = \sum_{k:0 \le k \le x} \binom{10}{k} p^k (1-p)^{10-k}$

We call such a variable X as **binomial random variable** and its distribution as the **binomial distribution**.

cdf and pdf of continuous random variables

Definition: For a continuous r.v. X, its distribution can also be described by a **non-negative probability density function (pdf)** $f(x)$ which satisfies

$$\mathbb{P}(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$$

for any two numbers a and b with $a \leq b$ (including $-\infty$ and $+\infty$).

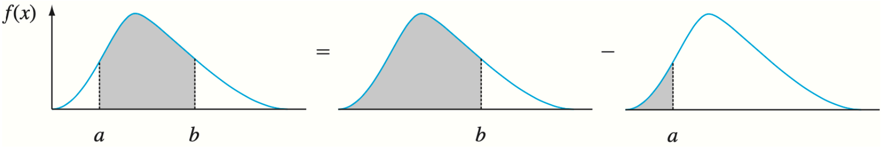
- By letting $a = -\infty$: cdf $F(b) = \int_{-\infty}^{b} f(x)dx$
- By Fundamental theorem of calculus (Newton-Leibniz Theorem): the cdf F of a continuous variable is differentiable and $F'(x) = f(x)$
- For continuous r.v., the single point doesn't matter, i.e. $\mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b)$ (why?)
- An appropriate pdf should satisfy two conditions:
  (1) $f(x) \geq 0$ for any number x
  (2) $\int_{-\infty}^{+\infty} f(x)dx = 1$

cdf and pdf of continuous random variables

Definition: For a continuous r.v. X, its distribution can also be described by **probability density function (pdf)** $f(x)$ which satisfies

$$\mathbb{P}(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$$

for any two numbers a and b with $a \leq b$ (including $-\infty$ and $+\infty$).



pdf cdf

The probability of a r.v. falling into a region is the **area** of shaded region under pdf $f(x)$, which connects to the physical meaning of the integral!

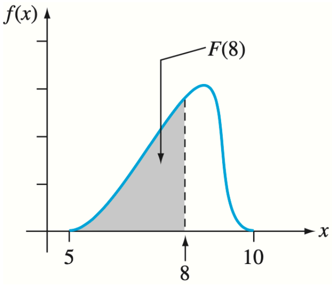
cdf and pdf of continuous random variables

<u>Definition:</u> **Probability density function (pdf)** $f(x)$ satisfies

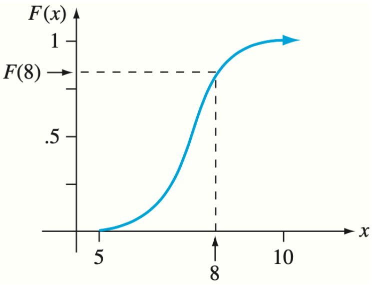
$$\mathbb{P}(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$$

for any two numbers a and b with $a \leq b$ (including $-\infty$ and $+\infty$).

- By letting $a = -\infty$: cdf $F(x) = \int_{-\infty}^{x} f(x)dx$
- $F'(x) = f(x)$



pdf



cdf

Exercise: continuous variables

Given the pdf, write down the corresponding cdf:

(1) $f(x) = 1, 0 \leq x \leq 1$ and $f(x) = 0$ elsewhere

(2) $f(x) = \frac{3}{2}x^2, -1 \leq x \leq 1$ and $f(x) = 0$ elsewhere

(3) $f(x) = 2e^{-2x}, x \geq 0$ and $f(x) = 0$ elsewhere

Given the cdf, write down the corresponding pdf:

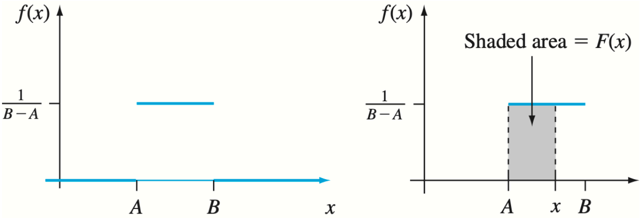
(1) $F(x) = x, 0 \leq x \leq 1$

(2) $F(x) = 1 - e^{-x}, x \geq 0$

Example of continuous distribution: uniform distribution

We say X follow a **uniform distribution** on $[A, B]$, if:

- Its pdf is $f(x) = \begin{cases} \frac{1}{B-A}, & A \le x \le B \\ 0, & \text{elsewhere} \end{cases}$
- Its cdf is $F(x) = \begin{cases} 0, & x \le A \\ \frac{x-A}{B-A}, & A < x \le B \\ 1, & x > B \end{cases}$

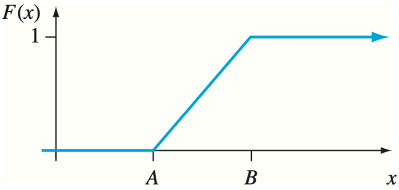


pdf

Example: uniform distribution

We say X follow a **uniform distribution** on $[A, B]$, if:

- Its pdf is $f(x) = \begin{cases} \frac{1}{B-A}, & A \le x \le B \\ 0, & \text{elsewhere} \end{cases}$
- Its cdf is $F(x) = \begin{cases} 0, & x \le A \\ \frac{x-A}{B-A}, & A < x \le B \\ 1, & x > B \end{cases}$



cdf

Comparison: discrete and continuous variables

Underlying intuition:

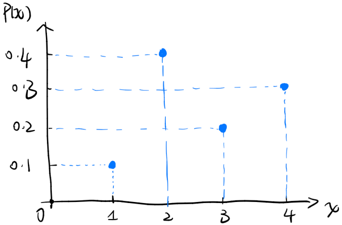
- The probability "mass" of discrete variables concentrates at a few points
- The probability "mass" of continuous variables spreads out in a dense region

Characterization of their distributions:

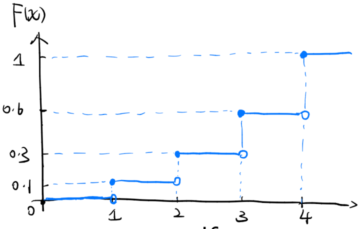
- cdf is available for both of them: $F(x) = \mathbb{P}(X \leq x)$
- pmf only works for discrete variables: $p(x) = \mathbb{P}(X = x)$
- pdf only works for continuous variables: $f(x) = F'(x)$ and $F(x) = \int_{-\infty}^{x} f(t)dt$

Comparison: discrete and continuous variables

Discrete distribution:

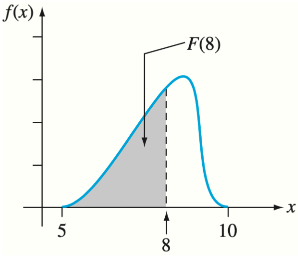


pmf

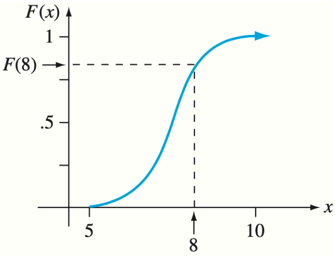


cdf

Continuous distribution:



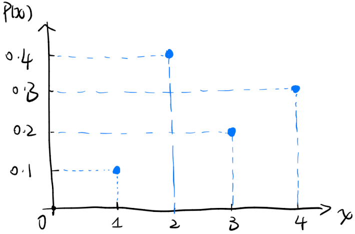
pdf



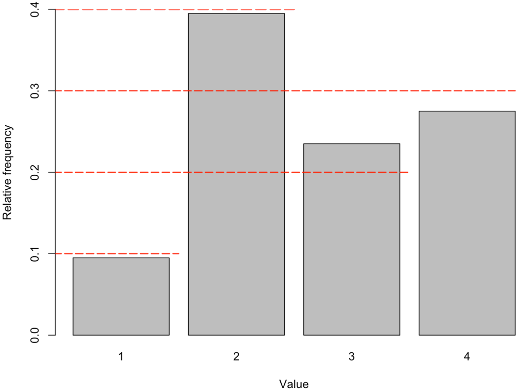
cdf

Relative frequency bar chart and pmf

Suppose a r.v. X has distribution with this pmf. I sampled X_1, X_2, ..., X_{1000} **independently** from this distribution.



pmf

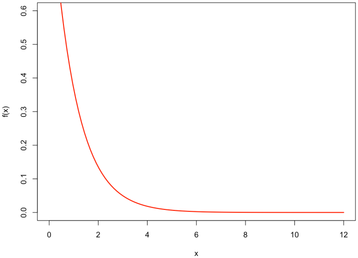


bar chart

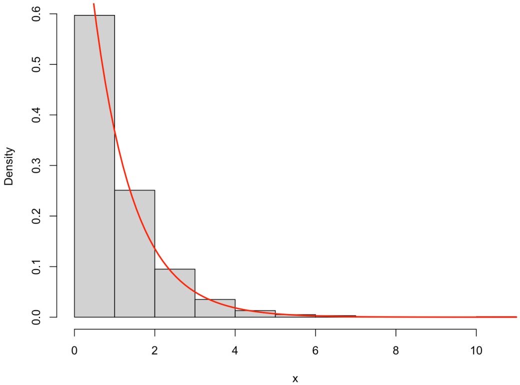
- Empirical relative frequency is an approximation of pmf.
- If we sample **infinite** points, the relative frequency will equal pmf.
- We will discuss more on this next week.

Density histogram and pdf

Suppose a r.v. X has distribution with this pdf. I sampled X_1, X_2, ..., X_{1000} **independently** from this distribution.



pdf



histogram

- Empirical density histogram is an approximation of pdf.
- If we sample **infinite** points and the bin width is **infinitely small**, the density histogram will be the same as the pdf curve.

Reading list (optional)

- "Probability and Statistics for Engineering and the Sciences" (9th edition):
  - Chapter 3.1, 3.2, 4.1 and 4.2 (skip the part of expectations)
- "OpenIntro statistics" (4th edition, free online, download `[here]`):
  - Chapter 3.4 and 3.5 (It's ok if you feel difficult to understand the expectation and variance. We will cover them next week.)

Many thanks to

- Yang Feng
- Joyce Robbins
- Chengliang Tang
- Owen Ward
- Wenda Zhou
- And all my teachers in the past 25 years