STAT 131 - Intro to Probability Theory

Lecture 3: Random Variables and their Distributions

Suzana Șerboi, Ph.D.
Mathematics and Statistics, UCSC

Random variables

In many scenarios, we are interested in a function of the outcome as opposed to the actual outcome of an experiment. For example, we might be interested in the sum of two dice and not in the separate values of each die or, when we flip a coin, we want to know the number of tails.

Real-valued functions defined on the sample space are known as random variables. Examples include the number of customers visiting a store in a day or the total score in a basketball game.

> Given an experiment with sample space S, a **random variable** (r.v.) is a function from the sample space S to the real numbers $\mathbb{R}$.

It is common, but not required, to denote random variables by capital letters.

A random variable $X : S \to \mathbb{R}$ assigns a numerical value $X(s)$ to each possible outcome s of an experiment. The randomness comes from the fact that we have a random experiment (with probabilities described by a probability function P); the mapping X itself is deterministic.

Example. We are tossing 3 times a fair coin. Note there are $2^3 = 8$ possible outcomes.

If we call X the number of heads obtained, then it is a random variables such that

$$P(X = 0) = P(\{T, T, T\}) = \frac{1}{8},$$

$$P(X = 1) = P(\{T, T, H\}, \{T, H, T\}, \{H, T, T\}) = \frac{3}{8},$$

$$P(X = 2) = P(\{T, H, H\}, \{H, T, H\}, \{H, H, T\}) = \frac{3}{8},$$

$$P(X = 3) = P(\{H, H, H\}) = \frac{1}{8}.$$

Discrete Random Variables

There are two main types of random variables used in practice: discrete r.v.s and continuous r.v.s. For now, our focus is on discrete r.v.s.

A r.v. X is said to be **discrete** if there is a finite list of values $a_1, a_2, \ldots, a_n$ or an infinite list of values $a_1, a_2, \ldots$ such that

$$P(X = a_j \text{ for some j}) = 1$$

If X is a discrete r.v. then the finite or countably infinite set of values x such that $P(X = x) > 0$ is called support of X.

Let X be a discrete random variable with possible values $a_1, a_2, \ldots$ The **probability mass function (PMF)**, or **distribution** of a random variable tells us the probabilities of these possible values:

$$p_X(a_i) = P(X = a_i),$$

for all possible a_i's.

Remark about notation

In writing $P(X = x)$, we are using $X = x$ to denote an event, consisting of all outcomes s to which X assigns the number x. This event is also written as $\{X = x\}$; formally, $\{X = x\}$ is defined as $\{s \in S : X(s) = x\}$, but writing $\{X = x\}$ is shorter and more intuitive.

For example, if X is the number of Heads in two fair coin tosses, then $\{X = 1\}$ consists of the sample outcomes HT and TH, which are the two outcomes to which X assigns the number 1. Since $\{HT, TH\}$ is a subset of the sample space, it is an event.

So it makes sense to talk about $P(X = 1)$, or more generally, $P(X = x)$. If $\{X = x\}$ were anything other than an event, it would make no sense to calculate its probability!

It does not make sense to write $P(X)$; we can only take the probability of an event, not of an r.v.

Example (Sum of die rolls) - Discrete Uniform distribution.

We roll two fair 6-sided dice. Let $T = X + Y$ be the total of the two rolls, where X and Y are the individual rolls. The sample space of this experiment has 36 equally likely outcomes: $S = \{(1,1), (1,2), \ldots, (6,5), (6,6)\}$.

For example, 7 of the 36 outcomes s are shown in the table below, along with the corresponding values of X, Y, and T.

s	X	Y	$T = X + Y$
(1,2)	1	2	3
(1,6)	1	6	7
(2,5)	2	5	7
(3,1)	3	1	4
(4,3)	4	3	7
(5,4)	5	4	9
(6,6)	6	6	12

Since the dice are fair, the PMF of X is

$$P(X = j) = 1/6$$

for $j = 1, 2, \ldots, 6$ (and $P(X = j) = 0$ otherwise); we say that X has a **Discrete Uniform distribution** on $1, 2, \ldots, 6$.

Similarly, Y is also Discrete Uniform on $1, 2, \ldots, 6$.

Note that Y has the same distribution as X but is not the same random variable as X.

Let's now find the PMF of T. By the naive definition of probability,

$$P(T = 2) = P(T = 12) = 1/36, P(T = 3) = P(T = 11) = 2/36,$$

$$P(T = 4) = P(T = 10) = 3/36, P(T = 5) = P(T = 9) = 4/36,$$

$$P(T = 6) = P(T = 8) = 5/36, P(T = 7) = 6/36.$$

For all other values of t, $P(T = t) = 0$. We can see that the support of T is $\{2, 3, \ldots, 12\}$ just by looking at the possible totals for two dice, but note

$$P(T = 2) + P(T = 3) + \cdots + P(T = 12) = 1,$$

which shows that all possibilities have been accounted for.

**Valid PMF.** Let X be discrete r.v. which can take the values $\{x_1, x_2, \ldots\}$. The PMF p_X MUST satisfy the following two criteria:

- Nonnegative: $p_X(x) > 0$ if $x = x_j$ for some j, and $p_X(x) = 0$ otherwise;

- Sums to 1: $\displaystyle\sum_{j=1}^{\infty} p_X(x_j) = 1$

This R code simulates rolling two dice 10,000 times and plots the probability mass function (PMF) of the sums of the rolls, allowing us to visualize the likelihood of each possible outcome (from 2 to 12).

```r
sum_two_dice_rolls <- function(){
dice <- sample(1:6, size = 2, replace = TRUE)
return(sum(dice))
}
sims <- replicate(10000, sum_two_dice_rolls())
plot(table(sims)/length(sims), xlab = 't', ylab = 'PMF', main = '10,000 Rolls')
points(table(sims)/length(sims),col=1, pch=15)
abline(h=0,col=2)
```


10,000 Rolls
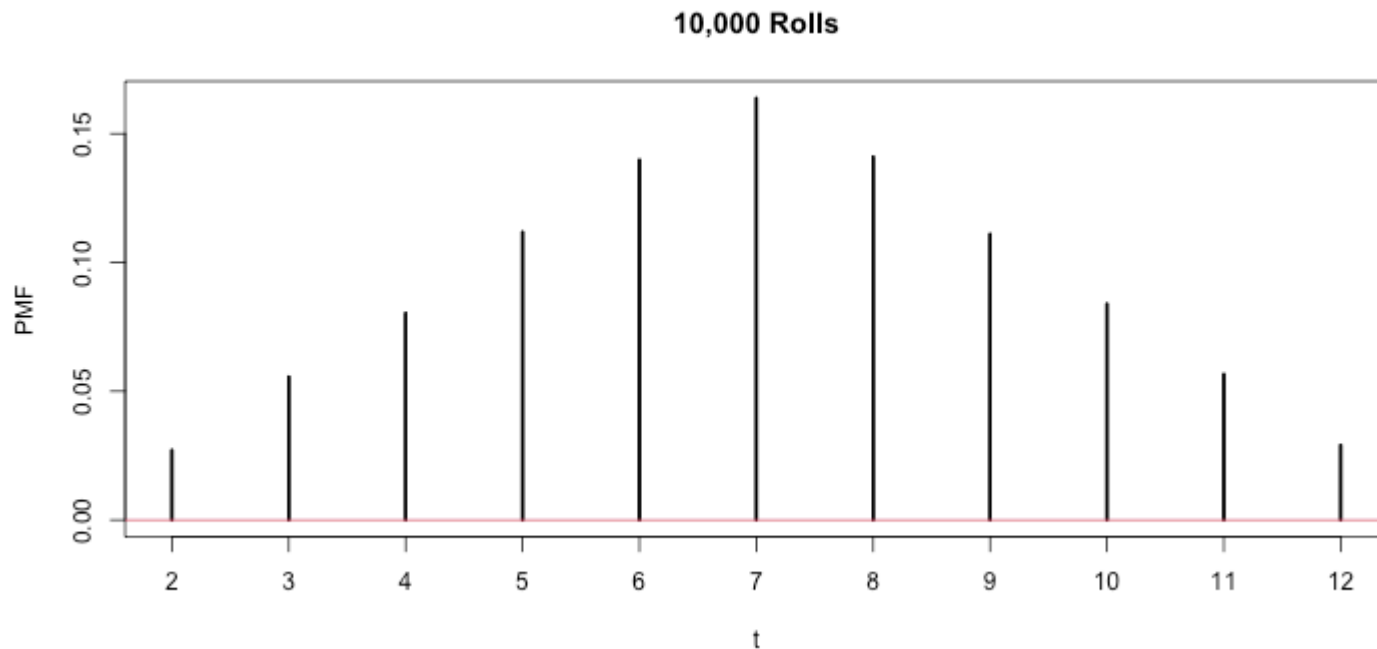
Let's break down the R code step by step:

The function `sum_two_dice_rolls()` simulates rolling two dice. It uses `sample()` to randomly select two numbers from 1 to 6 (with replacement). It then returns the sum of these two numbers:

```r
sum_two_dice_rolls <- function(){
  dice <- sample(1:6, size = 2, replace = TRUE)
  return(sum(dice))
}
```

Run `sum_two_dice_rolls()` function 10,000 times and stores the results in `sims`:

```r
sims <- replicate(10000, sum_two_dice_rolls())
```

Create a plot of the PMF of the simulation results: `table(sims)` counts the occurrences of each sum. Dividing by `length(sims)` converts these counts to probabilities. The plot is labeled with 'PMF' on the y-axis and 't' on the x-axis. The title '10000 Rolls' indicates that the simulation uses 10,000 rolls.

```r
plot(table(sims)/length(sims), xlab = 't', ylab = 'PMF', main = '10,000  Rolls')
```

`points()` adds points to the plot, using solid squares (pch=15) in black (col=1); `abline(h=0, col=2)` adds a horizontal line at `y=0` in red (col=2).

```r
points(table(sims)/length(sims),  col=1,  pch=15)
```
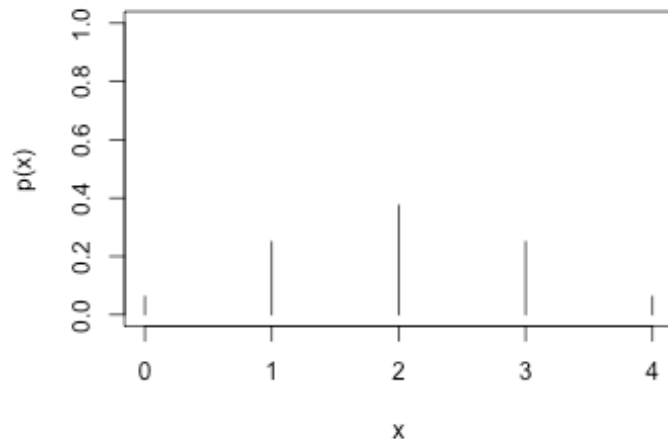
Cumulative distribution functions

Another function that describes the distribution of a r.v. is the cumulative distribution function (CDF). The CDF is defined for all r.v.s (discrete and continuous).
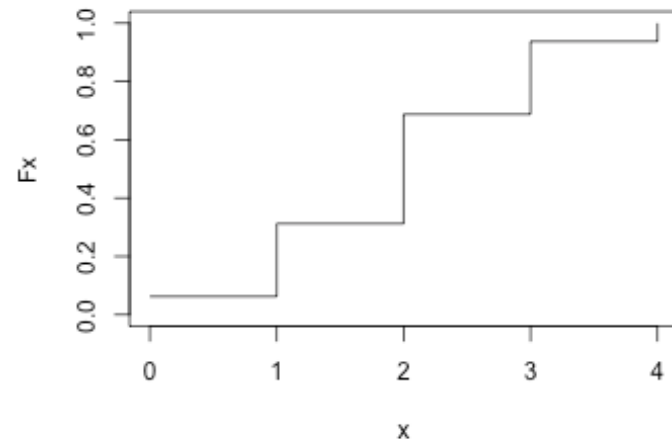
The cumulative distribution function (CDF) of an r.v. X is the function F_X given by

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

Probability mass function for Bin(4,0.5)

CDF for Bin(4,0.5)

Valid CDFs

Any CDF F has the following properties:

- Increasing: if $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.

- Right continuous: CDF is continuous except possibly for having some jumps. Wherever is a jump, the CDF is continuous from the right. That is:

$$F(a) = \lim_{x \to a^+} F(x)$$

- Convergence to 0 and 1 in the limits:

$$\lim_{x \to -\infty} F(x) = 0$$

$$\lim_{x \to \infty} F(x) = 1$$

Discrete Uniform distribution

Let C be finite, nonempty set of numbers. Choose one of these numbers uniformly at random (i.e., all values in C are equally likely). Call the chosen number X (a random variable), then X is said to have the Discrete Uniform distribution with parameter C; we denote this by $X \sim DUnif(C)$.

The PMF of $X \sim DUnif(C)$ is

$$P(X = x) = \frac{1}{|C|}$$

for $x \in C$, and 0 otherwise. The symbol $\sim$ is read "is distributed as".

As with questions based on the naive definition of probability, questions based on a Discrete Uniform distribution reduce to counting problems. Specifically, for $X \sim DUnif(C)$ and any $A \subseteq C$, we have

$$P(X \in A) = \frac{|A|}{|C|}.$$

Bernoulli distribution

A **Bernoulli trial** is an experiment in which there are only two possible outcomes, such as head or tail, success or failure, defective or nondefective, patient recovers or does not recover. It is convenient to designate the two possible outcomes of such an experiment as 0 and 1.

A random variable X has the **Bernoulli distribution** with parameter p if X can take only the values 0 and 1 and the probabilities

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

where $0 < p < 1$. We write this as $X \sim Bern(p)$.

If $X \sim Bern(p)$, the PMF of X can be written as:

$$P(X = k) = p^k(1 - p)^{1-k} \text{ for } k = 0, 1$$

(and $P(X = k) = 0$ otherwise).

Binomial distribution

Suppose that n independent Bernoulli trials are performed, each with the same success probability p. Let X be the number of successes. The distribution of X is called **Binomial distribution with parameters n and p**. We write $X \sim Bin(n, p)$ where n is a positive integer and $0 < p < 1$.

If $X \sim Bin(n, p)$, then the PMF of X is:

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

for $k = 0, 1, 2, \ldots, n$. $P(X = k) = 0$ otherwise (this is often implicit).

The cumulative distribution function (CDF) can be expressed as:

$$F(x) = P(X \le x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k},$$

where $\lfloor x \rfloor$ is the greatest integer less than or equal to x (the "floor" function).

```
n <- 50    # number of flips
p <- 0.5   # prob a of a success (let's say heads)
x <- 0:n   # possible values for our random variable
px <- dbinom(x, size = n, prob = p) # this calculates the PMF
plot(x, px, type="h",col=1, main = "Probability mass function for Bin(50,0.5)",
     xlab="x",ylab="p(x)")
```


Probability mass function for Bin(50,0.5)
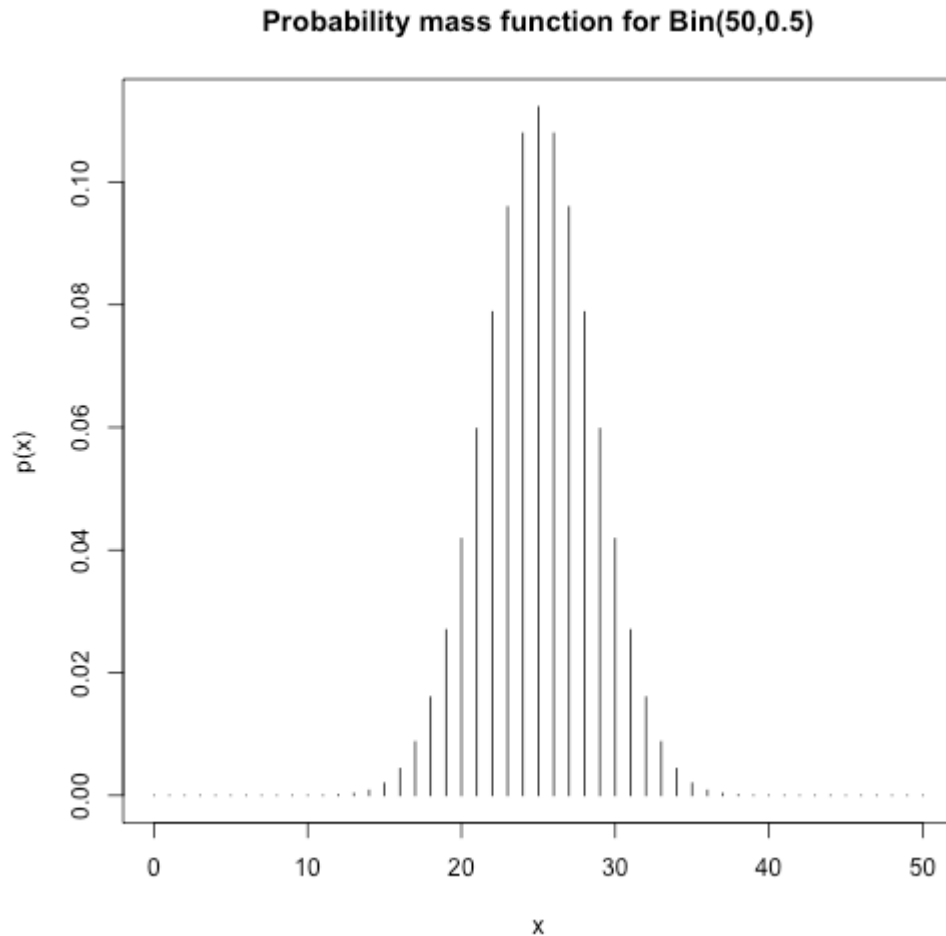
R Commands for the Binomial PMF and CDF

`dbinom(k, n, p)` gives the PMF $P(X = k)$ for $k = 0, 1, \ldots, n$

`pbinom(k, n, p)` gives the CDF $P(X \leq k)$ for $k = 0, 1, \ldots, n$

Example. The probability that a patient recovers from a rare blood disease is $p = 0.4$. If $n = 15$ people are known to have contracted this disease, what is the probability that:

(a) exactly 5 survive,
(b) from 3 to 8 survive, and
(c) at least 10 survive?

Solution: Let X be the number of people who survive. Then $X \sim Bin(15, 0.4)$.

$$(a)\ P(X = 5) = \binom{15}{5} (0.4)^5 (1 - 0.4)^{15-5} = 0.1859$$

```
dbinom(5, 15, 0.4)
```

```
## [1] 0.1859378
```

(b) $P(3 \leq X \leq 8) = P(X \leq 8) - P(X \leq 2)$

$$= \sum_{k=0}^{8} \binom{15}{k} (0.4)^k (0.6)^{15-k} - \sum_{k=0}^{2} \binom{15}{k} (0.4)^k (0.6)^{15-k}$$

$$= 0.9050 - 0.0271 = 0.8779$$

```
pbinom(8, 15, 0.4)
```

```
## [1] 0.9049526
```

```
pbinom(2, 15, 0.4)
```

```
## [1] 0.027114
```

(c) $P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{k=0}^{9} \binom{15}{k} (0.4)^k (0.6)^{15-k}$

$$= 1 - 0.9662 = 0.0338$$

```
pbinom(9, 15, 0.4)
```

```
## [1] 0.9661667
```

Hypergeometric distribution

The essential structure of the Hypergeometric story is that items in a population are classified using two sets of tags: in the urn story, each ball is either white or black (this is the first set of tags), and each ball is either sampled or not sampled (this is the second set of tags).

Consider an urn with w white balls and b black balls.
We draw n balls out of the urn at random without replacement, such that all $N = w + b$ samples are equally likely.
Let X be the number of white balls in the sample. Then X is said to have the Hypergeometric distribution with parameters w, b, and n. We denote this by $X \sim HGeom(w, b, n)$.

If $X \sim \mathrm{HGeom}(w, b, n)$, then the PMF of X is

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}} = \frac{\binom{w}{k}\binom{N-w}{n-k}}{\binom{N}{n}},$$

for integers k satisfying $0 \leq k \leq w$ and $0 \leq n - k \leq b$.

Proof.

To get $P(X = k)$, we first count the number of possible ways to draw exactly k white balls and $n - k$ black balls from the urn (without distinguishing between different orderings for getting the same set of balls).

If $k > w$ or $n - k > b$, then the draw is impossible. Otherwise, there are $\binom{w}{k}\binom{b}{n-k}$ ways to draw k white and $n - k$ black balls by the multiplication rule, and there are $\binom{w+b}{n}$ total ways to draw n balls.

Since all samples are equally likely, the naive definition of probability gives

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$$

for integers k satisfying $0 \le k \le w$ and $0 \le n - k \le b$.

This PMF is valid because the numerator, summed over all k, equals $\binom{w+b}{n}$ by Vandermonde's identity (Example 1.5.3), so the PMF sums to 1 .

Example (Fish Tagging and Recapture)

A small lake contains 50 fish. One day a fisherman catches 10 of these fish and tags them so that they can be recognized if they are caught again. The tagged fish are released back into the lake. The next day the fisherman goes out and catches 8 fish, which are kept in the fishing boat until they are all released at the end of the day.

The second day's fishing can be thought of as a sample of size eight taken without replacement from the fish stock. The sample is taken without replacement since the fish that are caught are kept in the fishing boat until all 8 fish have been caught (thereby eliminating the possibility of the same fish being caught twice on the second day).

Consequently, given that all 50 fish are equally likely to be caught, the number of tagged fish caught on the second day, denoted by X, has a hypergeometric distribution with $w = 10$, $b = N - w = 50 - 10 = 40$ and $n = 8$.

For example, the probability that 3 tagged fish are caught on the second day is:

$$P(X = 3) = \frac{\binom{10}{3}\binom{40}{8-5}}{\binom{10+40}{8}} = 0.147.$$

The Hypergeometric Distribution in R

The values from the previous example:

$k = 3$ the number of white balls drawn without replacement from an urn which contains both black and white balls

$w = 10$ the number of white balls in the urn

$b = 40$ the number of black balls in the urn

$n = 8$ the number of balls drawn from the urn

```
k <- 3
w <- 10
b <- 40
n <- 8
dhyper(k, w, b, n)
```

```
## [1] 0.1470741
```

Independence of random variables

Random variables X and Y are said to be **independent** if:

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y),$$

for all $x, y \in \mathbb{R}$.

In the discrete case, this is equivalent to the condition:

$$P(X = x, Y = y) = P(X = x)P(Y = y),$$

for all x, y with x in the support of X and y in the support of Y.

You can extend this definition to more than two random variables, and conditional independence.

Example. Roll two fair dice. Let X be the number on the first die, and Y the number on the second die, then $X + Y$ is not independent of $X - Y$ since

$$0 = P(X + Y = 12, X - Y = 1) \neq P(X + Y = 12)P(X - Y = 1) = \frac{1}{36} \cdot \frac{5}{36}.$$

Additional Practice Problems*

1. Suppose 70% of all purchases in a certain store are made with credit card. Let X denote the number of credit card uses in the next 10 purchases. Find the probability $P(5 \leq X \leq 8)$.

2 A company receives 60% of its orders online. Within a collection of 18 independently placed orders, what is the probability that

(a) between eight and ten of the orders are received online?

(b) no more than four of the orders are received online?

3. A jury of 12 people is selected at random from a group of 16 men and 18 women. What is the probability that the jury contains exactly 7 women?

4. Twelve refrigerators have been returned to the distributor because of a high-pitched oscillating noise. Suppose that four of the 12 have a defective compressor and the rest less-serious problems. Six refrigerators are selected at random for problem identification. Let X be the number of those found with a defective compressor. Find $P(X = 3)$.

*Solutions will be discussed in lecture, if time permits.