Probability 1

Lecture 4: Discrete Random Variables - Part 1

Dr. Daniel Flores-Agreda,

(based on the notes of Prof. Davide La Vecchia)

Spring Semester 2021

Objectives

- Define the concept of a Random Variable
- Explore the features of Discrete Random Variables
  - Distribution and Probability Mass Function (PMF)
  - Cumulative Distribution Function (CDF)
  - Expectation and Variance
- (If time allows it) Start presenting some important Discrete Distributions.

Outline

What is a Random Variable?

Let's say we have a Random Experiment with different outcomes.

Definition (Informal)

A **Random Variable** X is a variable that takes on different **numerical values** according to the outcomes of a random experiment.
The probability of the numerical values will result from the probabilities of the outcomes.

To define a random variable, we need:

1. a list of **all possible numerical values**
2. the **probability of each numerical value**

What is a Random Variable?

Example (Rolling the dice - Again)

- Roll a single die, and record the number of dots on the top side.
- The list of all possible outcomes is the number shown on the die.
  - i.e. the possible outcomes are 1, 2, 3, 4, 5 and 6
- If we say each outcome is equally likely, then the probability of each outcome must be $1/6$

What is a Random Variable?

Example (Flipping coins - Again)

- Flip a coin 10 times, and record the number of times T (tail) occurs
- The possible outcomes are

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9 \text{ and } 10$$

- For each number we associate a probability
- The probabilities are determined by the assumptions made about the coin flips, e.g.
  - what is the probability of a 'tail' on a single coin flip
  - whether this probability is the same for every coin flip
  - whether the 10 coin flips are 'independent' of each other

What is a Random Variable?

Example

- Measure the time taken by school students to complete a test.
- Every student has a maximum of 2 hours to finish the test.
- Let $X =$ completion time (in minutes).
- The possible values of the random variable X are contained in the interval

$$(0, 120] = \{x : 0 < x \leq 120\}.$$

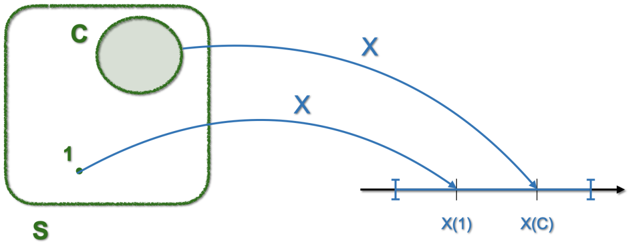
- We then need to associate probabilities with all events we may wish to consider, such as

$$P(\{X \leq 15\}) \quad \text{or} \quad P(\{X > 60\}).$$

What is a Random Variable?

A more formal definition

- Suppose we have:
  a. A sample space S "for the events"
  b. A probability measure (Pr) "for the events" in S
- Let $X(s)$ be a function that takes an element $s \in S$ to a number x



What is a Random Variable?

Example (Rolling two dice)

Experiment: We already know that the sample space S is given by:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

- let $i \in \{1, \ldots, 6\}$ denote the outcome of Die 1
- let $j \in \{1, \ldots, 6\}$ denote the outcome of Die 2

Every pair $(i, j) = s_{ij} \in S$ has a probability $1/36$

For every element or subset of S we can compute a probability with $Pr(\cdot)$

What is a Random Variable?

Example (continued)

Let us define $X(s_{ij})$ as the sum of the outcomes in both dice:

$$X(s_{ij}) = X(i,j) = i + j, \quad \textit{for} \quad i = 1, ..., 6, \textit{and } j = 1, ..., 6$$

Consequences:

- $X(\cdot)$ maps S into D.
- The sample space D is given by

$$D = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

where, for instance:

2 is related to the pair $(1, 1)$,
3 is related to the pairs $(1, 2)$ and $(2, 1)$, etc etc.
Every element or subset of D we can compute a probability with $P(\cdot)$

What is a Random Variable?

Example (continued)

- To each element (event) in D we can attach a probability, using the probability of the corresponding event(s) in S. For instance,

$$P(2) = Pr(1,1) = 1/36, \quad \text{or} \quad P(3) = Pr(1,2) + Pr(2,1) = 2/36.$$

- How about the $P(7)$?

$$P(7) = Pr(3,4) + Pr(2,5) + Pr(1,6) + Pr(4,3) + Pr(5,2) + Pr(6,1) = 6/36.$$

- The latter equality can also be re-written as

$$P(7) = 2(Pr(3,4) + Pr(2,5) + Pr(1,6)) = 6\ Pr(3,4),$$

Exercise

What is $P(9)$? What is $P(13)$? [Hint: does 13 belong to D?]

What is a Random Variable?

A(n) even more formal characterisation

Let us formalise all these ideas:

- Let D be the set of all values x that can be obtained by $X(s)$, for all $s \in S$:

$$D = \{x : x = X(s),\ s \in S\}$$

- D is a **list of all possible numbers** x that can be obtained, and thus is a **sample space for** X. *Remark that the random variable is* X *while* x *represents its realization (non random).*
- D can be either an **uncountable interval**
  - X is a **continuous** random variable, or
- D can be **discrete** or **countable**
  - X is a **discrete** random variable

What is a Random Variable?

Moreover, because P is defined from Pr, it is also a probability measure on D. For each A:

$$P(A) = Pr(\{s \in S : X(s) \in A\})$$

where P and Pr stand for "probability" on D and on S, respectively. Hence:

1. $P(A) \geq 0$
2. $P(D) = Pr(\{s \in S : X(s) \in D\}) = Pr(S) = 1$
3. If $A_1, A_2, A_3 ...$ is a sequence of events such that

   $$A_i \cap A_j = \varnothing$$

   for all $i \neq j$ then

   $$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

(From now on, we'll be dropping the colors.)

Example (Geometric random variable)

Experiment: rolling a die until a 6 appears.

- Let $X =$ "number of rolls until we get a 6"
- $D = \{1, 2, 3, \dots, n, \dots\} \equiv \mathbb{N}$.

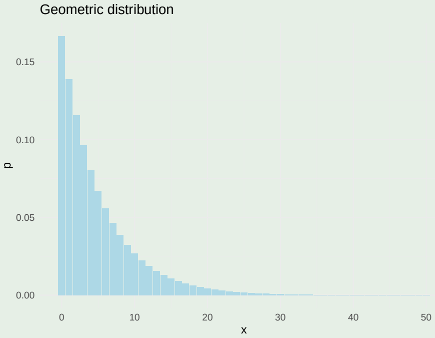
$$P(\{X=1\}) = Pr(\text{'6' on the 1st roll}) = \frac{1}{6}$$

$$P(\{X=2\}) = Pr\left(\text{no '6' on the 1st roll and '6' on the 2nd roll}\right) = \frac{5}{6}\cdot\frac{1}{6} = \frac{5}{36}$$

$$\begin{aligned}P(\{X=3\}) &= Pr\left(\text{no '6' on either the 1st nor 2nd roll and '6' on the third roll}\right)\\ &= \frac{5}{6}\cdot\frac{5}{6}\cdot\frac{1}{6} = \frac{25}{216}\\ &\vdots\\ P(\{X=n\}) &= Pr(\text{no '6' on the first } n-1 \text{ rolls and '6' on the last roll})\\ &= \left(\frac{5}{6}\right)^{n-1}\cdot\frac{1}{6}\\ &\vdots\end{aligned}$$

What is a Random Variable?

Example (continued)



Example (continued)

- Rather than list the possible values of X along with the associated probabilities in a table, we can provide a formula that gives the required probabilities.

$$P(\{X = n\}) = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6} \quad \text{for} \quad n = 1, 2, \ldots$$

Exercise

Show that

$$\sum_{n=1}^{\infty} \left(\frac{5}{6}\right)^{n-1} \frac{1}{6} = 1.$$

Discrete random variables

Discrete random variables

Discrete random variables are often associated with the process of counting. More generally:

Definition (Probability of a Discrete Random Variable)

Suppose X can take the values $x_1, x_2, x_3, \ldots, x_n$.
The probability of x_i is

$$p_i = P(\{X = x_i\})$$

and we must have $p_1 + p_2 + p_3 + \cdots + p_n = 1$ and all $p_i \geq 0$. These probabilities may be put in a table

x_i	$P(\{X = x_i\})$
x_1	p_1
x_2	p_2
x_3	p_3
$\vdots$	$\vdots$
x_n	p_n
Total	1

Discrete random variables

- For a **discrete random variable** X, any table listing all possible nonzero probabilities provides the entire **Probability Distribution**.
- The **probability mass function** $p(a)$ of X is defined by

$$p_a = p(a) = P(\{X = a\})$$

and this is positive for at most a countable number of values of a.

For instance, $p_1 = P(\{X = x_1\})$, $p_2 = P(\{X = x_2\})$, and so on. That is, if X must assume one of the values $x_1, x_2, ...$, then

$$\begin{aligned} p(x_i) &\geq 0 \quad \text{for} \quad i = 1, 2, ... \\ p(x) &= 0 \qquad \text{otherwise.} \end{aligned} \tag{1}$$

Clearly, we must have

$$\sum_{i=1}^{\infty} p(x_i) = 1.$$

Cumulative Distribution Function

The **Cumulative Distribution Function** (**CDF**) is a table listing the values that X can take, along with

$$F_X(a) = P(\{X \leq a\}) = \sum_{\text{all} \quad x \leq a} p(x).$$

If the random variable X takes on values $x_1, x_2, x_3, \ldots, x_n$ *listed in increasing order* $x_1 < x_2 < x_3 < \cdots < x_n$, the CDF is a step function, that it its value is constant in the intervals $(x_{i-1}, x_i]$ and takes a step/jump of size p_i at each x_i:

x_i	$F_X(x_i) = P(\{X \leq x_i\})$
x_1	p_1
x_2	$p_1 + p_2$
x_3	$p_1 + p_2 + p_3$
$\vdots$	$\vdots$
x_n	$p_1 + p_2 + \cdots + p_n = 1$

Cumulative Distribution Function

Example

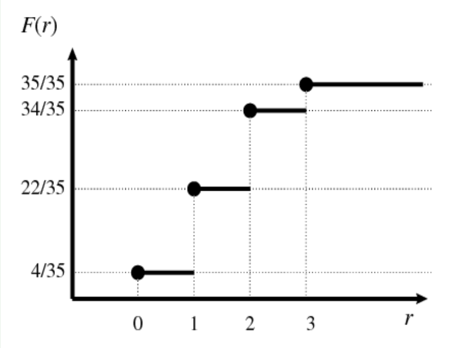
x_i	0	1	2	3	Tot
$P(\{X = x_i\})$	$4/35$	$18/35$	$12/35$	$1/35$	1
$P(\{X \leq x_i\})$	$4/35$	$22/35$	$34/35$	$35/35$	

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 4/35 & 0 \leq x < 1 \\ 22/35 & 1 \leq x < 2 \\ 34/35 & 2 \leq x < 3 \\ 1 & x \geq 3. \end{cases}$$

Cumulative Distribution Function

Example (continued)

... or graphically, you get *a step function* ...



Cumulative Distribution Function

Remark

Suppose $a \leq b$*. Then, because the event* $\{X \leq a\}$ *is contained in the event* $\{X \leq b\}$*, namely*

$$\{X \leq a\} \subseteq \{X \leq b\},$$

it follows that

$$F_X(a) \leq F_X(b),$$

so, the probability of the former is less than or equal to the probability of the latter.

In other words, $F_X(x)$ is a nondecreasing function of x.

Cumulative Distribution Function

Definition (Quantiles)

The CDF can be inverted to define the value x of X that corresponds to a given probability α, namely $\alpha = P(X \leq x)$, for $\alpha \in [0, 1]$.

The inverse CDF $F_X^{-1}(\alpha)$ or **quantile of order** α, $Q(\alpha)$, is the smallest realisation of X associated to a CDF greater or equal to α
In formula, the α-quantile $Q(\alpha)$ is the smallest number satisfying:

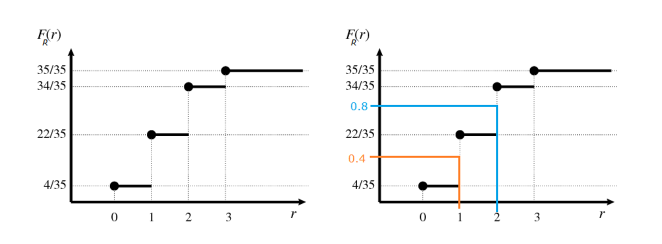
$$F_X[F_X^{-1}(\alpha)] = P[X \leq \underbrace{F_X^{-1}(\alpha)}_{Q(\alpha)}] \geq \alpha, \quad \text{for} \quad \alpha \in [0, 1].$$

By construction, a quantile of a discrete random variable is a realization of X[a]

[a] More to come in future chapters

Cumulative Distribution Function

Example (cont'd, graphically)



...calling (only for this slide) R the rv, r its realizations and $F_R(r)$ its CDF at r...

Distributional Summaries

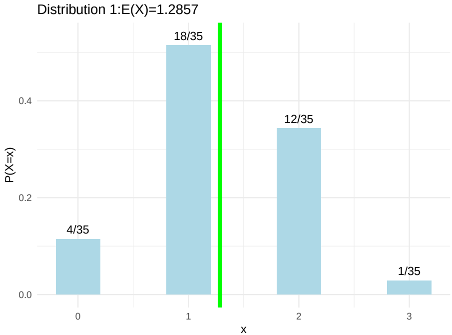
For a discrete random variable, it is useful to provide values that describe some **attributes** of its distribution

Definition

The **Expectation**, a.k.a. **Expected** or **Mean** value, of the distribution is (roughly speaking) its *center*.

$$E[X] = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n = \sum_{i=1}^{n} p_i x_i,$$

and constitutes a measure of *location*.



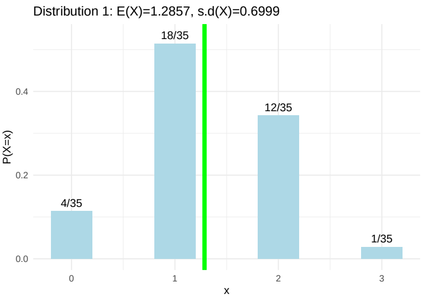
Distributional Summaries

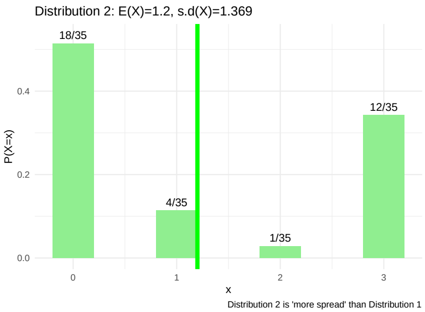
Definition

The **Square root of the Variance**, or **Standard Deviation,** of the distribution

$$
\begin{aligned}
s.d(X) &= \sqrt{Var(X)} \\
&= \sqrt{p_1 (x_1 - E[X])^2 + p_2 (x_2 - E[X])^2 + \cdots + p_n (x_n - E[X])^2}
\end{aligned}
$$

is a measure of *spread* a.k.a. 'variability' or 'dispersion').





Distributional Summaries

If X is a discrete random variable and a is any real number, then:

$$
\begin{aligned}
E[\alpha X] &= \alpha E[X] \\
E[\alpha + X] &= \alpha + E[X] \\
Var(\alpha X) &= \alpha^2 Var(X) \\
Var(\alpha + X) &= Var(X)
\end{aligned}
$$

Exercise

Let us very the first property: $E[\alpha X] = \alpha E[X]$. *From the intro lecture we know that, for every* $\alpha_i \in \mathbb{R}$,

$$\sum_{i=1}^{n} \alpha_i X_i = \alpha_1 X_1 + \alpha_2 X_2 + ... + \alpha_n X_n.$$

So, the required result follows as a special case, setting $\alpha_i = \alpha$, *for every* i, *and applying the definition of Expectation. Verify this and the other properties as an exercise. [Hint: set* $\alpha_i = \alpha p_i$.*]*

Definition

Consider two discrete random variables Then, X and Y are **independent** if

$$P(\{ X = x\} \cap \{Y = y\}) = P(\{X = x\}) \cdot P(\{Y = y\})$$

for all values x that X can take and all values y that Y can take.

- If X and Y are two discrete random variables, then

$$E\left[X+Y\right]=E\left[X\right]+E\left[Y\right]$$

- If X and Y are also *independent*, then

$$Var\left(X+Y\right)=Var\left(X\right)+Var\left(Y\right) \tag{2}$$

Remark

Note that Eq. (2) does not (typically) hold if X and Y are NOT independent—more to come on this later on...

Distributional Summaries

Recall that the expectation of X was defined as

$$E[X] = \sum_{i=1}^{n} p_i x_i$$

Now, suppose we are interested in a function m of the random variable X, say $m(X)$. We define

$$E[m(X)] = p_1 m(x_1) + p_2 m(x_2) + \cdots p_n m(x_n).$$

Remark

Notice that the variance is a special case of expectation where,

$$m(X) = (X - E[X])^2.$$

Indeed,

$$Var(X) = E\left[(X - E[X])^2\right].$$

Exercise

Show that

$$Var(X) = E\left[X^2\right] - E[X]^2.$$

Wrap-up (to this point)

- A **Random Variable** *maps* events from the **Sample Space of the Events** onto a set of numerical values.
- This entails a new sample space and a new probability measure $P(\cdot)$
- When the mapping is onto a **countable set** the Random Variable is **Discrete**.
- A Discrete Random Variable is endowed of a **Probability Distribution**, which is a **listing of the probabilities of each value**.
- This Distribution can be displayed as a **table** or as **functions** : Probability Mass Function (PMF) or Cumulative Distribution Function (CDF).
- A Distribution has a "gravity center" The **Expectation**.
- A Distribution has a measure of "spread" with respect to this center: the Standard Deviation, which depends on the **Variance**.

Thank You for your Attention!
“See you” Next Week