Independent random variables

MATH/STAT 394: Probability I
Summer 2021 A Term

Introduction to Probability
D. Anderson, T.Seppäläinen, B. Valkó

§ 2.4

Aaron Osgood-Zimmerman

Department of Statistics

Logistics

- HW2 due tonight (July 2) at 11:59
- HW3 will be available as soon as I finish it (by midnight)
- Mid-term feedback for is now available, and will be up through Tuesday at 3pm (I need time to review before Wed class)

Practice solution

Practice

Chose a number uniformly at random on $[0,1]$. Consider the events

$$A = [0, 1/2] \quad B = [1/3, 2/3], \quad C = [11/24, 17/24]$$

Are A, B, C independent (i.e. mutually independent)?

Solution You can check that $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. Similarly you have $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, yet,

$$\mathbb{P}(A \cap C) = \frac{1}{24} \neq \frac{1}{2} \cdot \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(C)$$

So they are not mutually independent.

Reminder to compute $\mathbb{P}(A \cap C)$, it suffices to identify $A \cap C = [11/24, 1/2]$ then $\mathbb{P}([a,b]) = \frac{b-a}{d-c}$ for a number chosen uniformly at random on $[c,d]$.

Recap

Independence

- A and B are independent iff $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- If A and B are independent, then so are A^*, B^* for $A^* = A$ or A^c.
- $A_1, \ldots, A_n$ are independent if

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

for any $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$, with $k \geq 2$.

Outline

Independent random variables

Classical random variables

Independent random variables

Recall that on probability space $(\Omega, \mathcal{F}, P)$,

A random variable X is a function from Ω to $\mathbb{R}$.

And a **discrete random variable** X is a function from Ω to a countable (or finite) set of values $\{k_1, k_2, \dots\}$.

Definition

Random variables $X_1, \dots, X_n$ on a prob. space $(\Omega, \mathcal{F}, \mathbb{P})$ are independent if

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_n \in B_n)$$

for any[1] *subsets $B_1, \dots, B_n \subseteq \mathbb{R}$*

Note:

This means that the distribution of the r.v. can be **factorized** in the distributions of each r.v.

[1] See APV §1.6

Independent discrete random variables

Proposition

Discrete random variables $X_1, \dots, X_n$ on a prob. space $(\Omega, \mathcal{F}, \mathbb{P})$ are independent if and only if

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n) \tag{1}$$

for any possible choices $x_1, \dots, x_n$ of the r.v.

Proof

- Clearly if they are independent, they satisfy (1) by choosing $B_k = \{i_k\}$
- Now suppose they satisfy (1).

$$\begin{aligned}
\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) &= \sum_{x_1 \in B_1, \dots, x_n \in B_n} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\
&= \sum_{x_1 \in B_1, \dots, x_n \in B_n} \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n) \\
&= \left(\sum_{x_1 \in B_1} \mathbb{P}(X_1 = x_1) \right) \dots \left(\sum_{x_n \in B_n} \mathbb{P}(X_n = x_n) \right) \\
&= \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_n \in B_n)
\end{aligned}$$

Independent Discrete Random Variables

Exercise

Two fair dies are rolled independently until a sum of 5 or 7 is obtained.

What is the probability that the trials end with a sum of 5?

Solution

- Let S_i be the sum of i-th roll of the two dies.
- Consider the event that the game ends in trial i with a sum of 5:
$$A_i = \{S_i = 5\} \bigcap_{j=1}^{i-1} \{S_j \notin \{5,7\}\}, \quad i = 1, 2, \ldots.$$
- Then,
$$P(\{\text{Game ends in } 5\}) = P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$
- Since random variables $S_1, S_2, \ldots$ are **independent**,
$$P(A_i) = P(\{S_1 \notin \{5,7\}\}) \cdots P(\{S_{i-1} \notin \{5,7\}\}) P(\{S_i = 5\}),$$
where $P(S_i = 5) = \frac{4}{36}$, $P(\{S_i = 7\}) = \frac{6}{36}$, $P(\{S_i \notin \{5,7\}\}) = \frac{26}{36}$.
-
$$P(\{\text{Game ends in } 5\}) = \sum_{i=1}^{\infty} \left(\frac{26}{36}\right)^{i-1} \frac{4}{36} = \left(\frac{1}{1 - \frac{13}{18}}\right) \frac{1}{9} = \frac{2}{5}.$$

Independent Discrete Random Variables

We often deal with repeated experiments that have the same prob.

Definition

Random variables $X_1, \ldots, X_n$ are ***identically distributed*** *if*

$$\mathbb{P}(X_i \in B) = \mathbb{P}(X_j \in B) \quad \textit{for any } i, j \in \{1, \ldots n\}$$

In the following, we consider r.v. $X_1, \ldots, X_n$ that are

independent and identically distributed (i.i.d.)

Independent and identically distributed

Example

Take a coin whose probability of $\mathtt{H}$ is p. Toss the coin n times. Let

$$X_i = \begin{cases} 1, & i\text{-th toss is } \mathtt{H} \\ 0, & i\text{-th toss is } \mathtt{T} \end{cases}.$$

then the r.v. $X_1, \ldots, X_n$ are iid.

Proof

- This is an experiment with equally likely outcomes on
$$\Omega = \{\text{all sequence of } \mathtt{H} \text{ and } \mathtt{T} \text{ of length } n\}.$$
- 1. Given any $i_1, \ldots, i_n \in \{0, 1\}$,
$$P(X_1 = i_1, \ldots X_n = i_n) = P(X_1 = i_1)P(X_2 = i_2)\ldots P(X_n = i_n).$$
  2. For $i = 1, \ldots, n$, by definition
$$P(X_i = 1) = p, \quad P(X_i = 0) = 1 - p.$$
- For example, the probability that all but the first toss is $\mathtt{H}$, is
$$P(X_1 = 0, X_2 = 1, \ldots, X_n = 1) = (1 - p)p^{n-1}.$$

Independent and identically distributed

Example
Consider sampling k times with replacement form an urn with n balls labeled $1, \ldots, n$. Denote

$$X_i = \text{label of the } i\text{th ball drawn.}$$

Then $X_1, \ldots, X_n$ are iid.

Proof

- Independence already shown for 2 balls in lecture 11
- Can easily be generalized to n balls
- They clearly have the same distribution

Outline

Bernoulli random variable

Reminder:

Definition

A r.v. X *has a* ***Bernoulli*** *dist. with param.* $p \in [0,1]$ *if it takes its values in* $\{0,1\}$ *and*

$$\mathbb{P}(X = 1) = p \quad \mathbb{P}(X = 0) = 1 - p$$

We denote it $X \sim \mathrm{Ber}(p)$

Binomial random variable

Many random variables *arise from repeated trials*.

Example
Take a coin whose probability of $\mathtt{H}$ is p. Toss the coin n times.

From

$$X_i = \begin{cases} 1, & i\text{-th toss is } \mathtt{H} \\ 0, & i\text{-th toss is } \mathtt{T} \end{cases},$$

define

$$S := X_1 + \cdots + X_n.$$

S is a discrete random variable taking value from $\{0, 1, \ldots, n\}$.

What is the p.m.f. of S?

Solution

$$P_S(k) = \mathbb{P}(S = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

- Each outcome with k $\mathtt{H}$'s has probability $p^k(1-p)^{n-k}$.
- There are $\binom{n}{k}$ such outcomes and use finite additivity.

Binomial random variable

Definition

*A r.v. X has a **Binomial** distribution with parameters $n \in \mathbb{N}$, $n > 0$, and $p \in [0,1]$, if the possible values of X are $\{0, \ldots, n\}$ and*

$$P_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

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

Alternative definition

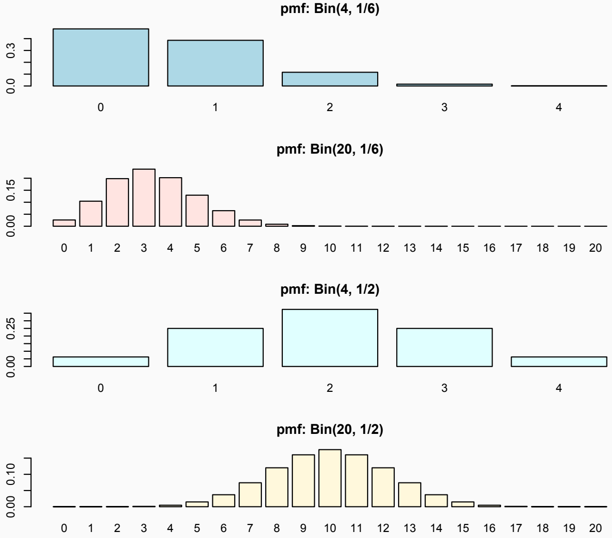
$X \sim \mathrm{Bin}(n,p)$ if and only if $X = Y_1 + \ldots + Y_n$ for $Y_i \overset{i.i.d.}{\sim} \mathrm{Ber}(p)$

where $Y_i \overset{i.i.d.}{\sim} \mathrm{Ber}(p)$ means that the Y_i are independent and identically distributed with a dist. $\mathrm{Ber}(p)$.

Notes:

- For $n = 1$, we retrieve the Bernoulli dist.
- Sanity check: $\sum_{k=0}^{n} P_S(k) = 1$ (by the binomial theorem)

Binomial random variable



Practice next lecture

Practice

What is the probability that 5 rolls of a fair die gives at least 2 sixes?

Hint:

1. Identify the fact that you get a six for each roll as a classical r.v.
2. Identify the number of sixes you got for 5 rolls as another r.v.
3. Note that if you have access to the p.m.f. of X that takes values in $\{0, \ldots, n\}$, for example you can decompose

$$\mathbb{P}(X < k) = \mathbb{P}(X = 0) + \ldots + \mathbb{P}(X = k - 1)$$

   where each element of the sum is given by the p.m.f. of the r.v.