Computer Science 351

Random Variables and Expectation

Instructor: Wayne Eberly

Department of Computer Science
University of Calgary

Lecture #20

Learning Goals

Learning Goals:

- Introduce ***random variable*** and their ***expected values*** as a way to consider numerical information that can be considered as part of an experiment (and may be the main reason why you want to consider the experiment, at all).

Random Variables

When we consider an experiment there is often a numerical value that we wish to ***count*** or ***bound.***

Examples:

- When tossing a sequence of coins, how many coins are tossed before a ***head*** is tossed for the first time?
- When tossing a sequence of n coins, ***how many times*** is a “head” tossed?
- When shuffling a deck of cards, what is the highest rank of the first five cards (where an Ace has rank 1, a Jack has rank 11, a Queen has rank 12, and King has rank 13, and the rank of any numbered card is its number)?
- When inserting a sequence of keys into a hash table, what is the length of the linked list of values at position 0 of the table?

Random Variables

Definition: Let Ω be a sample space. A ***random variable over*** $\boldsymbol{\Omega}$ is a (total) function $X : \Omega \rightarrow \mathbb{R}$.

- We will often shorten this phrase from "random variable over Ω" to "random variable" when the context makes it clear what sample space, Ω, is being considered.
- Any ***probability distribution*** $\mathrm{P} : \Omega \rightarrow \mathbb{R}$ is an example of a "random variable over Ω".

Random Variables

We will often be interested in random variable whose ranges are particular *subsets* V of $\mathbb{R}$ — so that these functions can also be viewed as functions $X : \Omega \to V$ (as well as functions $X : \Omega \to \mathbb{R}$).

- For example, an ***integer-valued random variable*** is a random variable $X : \Omega \to \mathbb{R}$ such that $X(\sigma) \in \mathbb{Z}$ for all $\sigma \in \Omega$ — so that, in effect, $X : \Omega \to \mathbb{Z}$.
- The (even more special) case that $X : \Omega \to \mathbb{N}$ will often be of interest too.

Random Variables

So will be the (even more special) case that $X(\sigma) \in \{0, 1\}$ for all $\sigma \in \Omega$, so that $X : \Omega \to \{0, 1\}$.

- This kind of random variable is often called an ***indicator random variable*** because it "indicates" an event, namely the event
$$\{\sigma \in \Omega \mid X(\sigma) = 1\} \subseteq \Omega.$$

Random Variables

Example: Consider the experiment of tossing a sequence of three coins — so that

$$\Omega = \{(\mathtt{H},\mathtt{H},\mathtt{H}),(\mathtt{H},\mathtt{H},\mathtt{T}),(\mathtt{H},\mathtt{T},\mathtt{H}),(\mathtt{H},\mathtt{T},\mathtt{T}),(\mathtt{T},\mathtt{H},\mathtt{H}),(\mathtt{T},\mathtt{H},\mathtt{T}),(\mathtt{T},\mathtt{T},\mathtt{H}),(\mathtt{T},\mathtt{T},\mathtt{T})\}.$$

The ***random variable*** "number of heads tossed" is the function $X : \Omega \to \mathbb{N}$ such that

- $X((\mathtt{H},\mathtt{H},\mathtt{H})) = 3.$
- $X((\mathtt{H},\mathtt{H},\mathtt{T})) = X((\mathtt{H},\mathtt{T},\mathtt{H})) = X((\mathtt{T},\mathtt{H},\mathtt{H})) = 2.$
- $X((\mathtt{H},\mathtt{T},\mathtt{T})) = X((\mathtt{T},\mathtt{H},\mathtt{T})) = X((\mathtt{T},\mathtt{T},\mathtt{H})) = 1.$
- $X((\mathtt{T},\mathtt{T},\mathtt{T})) = 0.$

Random Variables

Once again, let Ω be a sample space and let $X : \Omega \to \mathbb{R}$.

- We will write "$X = r$" as the name of the event

$$\{\sigma \in \Omega \mid X(\sigma) = r\} \subseteq \Omega.$$

- We will write "$X \geq r$" as the name of the event

$$\{\sigma \in \Omega \mid X(\sigma) \geq r\} \subseteq \Omega.$$

- "$X \leq r$", "$X > r$", $X < r$", and $X \neq r$" can be used as the names for (corresponding) events in the same way.

Random Variables

Continuing the previous example,

$$\text{“}X = 3\text{”} = \{(\mathtt{H}, \mathtt{H}, \mathtt{H})\},$$

$$\text{“}X = 2\text{”} = \{(\mathtt{H}, \mathtt{H}, \mathtt{T}), (\mathtt{H}, \mathtt{T}, \mathtt{H}), (\mathtt{T}, \mathtt{H}, \mathtt{H})\},$$

and

$$\text{“}X \geq 2\text{”} = \{\mathtt{H}, \mathtt{H}, \mathtt{H}), (\mathtt{H}, \mathtt{H}, \mathtt{T}), (\mathtt{H}, \mathtt{T}, \mathtt{H}), (\mathtt{T}, \mathtt{H}, \mathtt{H})\}.$$

Expectation

Let Ω be a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$, and let $X : \Omega \to \mathbb{R}$ be a random variable over Ω.

Suppose that

$$\sum_{\sigma \in \Omega} \mathrm{P}(\sigma) \times |X(\sigma)|$$

is finite — that is, " less than $+\infty$".[1]

Then the ***expected value of X, with respect to probability distribution P***, is the value

$$\mathrm{E}[X] = \sum_{\sigma \in \Omega} \mathrm{P}(\sigma) \times X(\sigma).$$

[1] This is a "technical restriction" that you will not need to worry about whenever Ω is a finite set.

Expectation

- The phrase “with respect to probability distribution P” will be dropped when it is clear, from context, which probability distribution is being used.
- This value has other names in the literature including
  - the ***mean*** of X,
  - the ***expectation*** of X, and
  - the ***first moment*** of X.

Expectation

Continuing this example — with the uniform distribution $\mathsf{P} : \Omega \to \mathbb{R}$ —

$$
\begin{aligned}
\mathsf{E}[X] &= \sum_{\sigma \in \Omega} \mathsf{P}(\sigma) \times X(\sigma) \\
&= \mathsf{P}((\mathtt{H},\mathtt{H},\mathtt{H})) \times X((\mathtt{H},\mathtt{H},\mathtt{H})) + \mathsf{P}((\mathtt{H},\mathtt{H},\mathtt{T})) \times X((\mathtt{H},\mathtt{H},\mathtt{T})) \\
&\quad + \mathsf{P}((\mathtt{H},\mathtt{T},\mathtt{H})) \times X((\mathtt{H},\mathtt{T},\mathtt{H})) + \mathsf{P}((\mathtt{H},\mathtt{T},\mathtt{T})) \times X((\mathtt{H},\mathtt{T},\mathtt{T})) \\
&\quad + \mathsf{P}((\mathtt{T},\mathtt{H},\mathtt{H})) \times X((\mathtt{T},\mathtt{H},\mathtt{H})) + \mathsf{P}((\mathtt{T},\mathtt{H},\mathtt{T})) \times X((\mathtt{T},\mathtt{H},\mathtt{T})) \\
&\quad + \mathsf{P}((\mathtt{T},\mathtt{T},\mathtt{H})) \times X((\mathtt{T},\mathtt{T},\mathtt{H})) + \mathsf{P}((\mathtt{T},\mathtt{T},\mathtt{T})) \times X((\mathtt{T},\mathtt{T},\mathtt{T})) \\
&= \frac{1}{8} \times 3 + \frac{1}{8} \times 2 + \frac{1}{8} \times 2 + \frac{1}{8} \times 1 + \frac{1}{8} \times 2 + \frac{1}{8} \times 1 \\
&\qquad + \frac{1}{8} \times 1 + \frac{1}{8} \times 0 \\
&= \frac{1}{8} \times 12 = \frac{3}{2}.
\end{aligned}
$$

Conditional Expectation

Recall, from the lecture on “Conditional Probability”, that if Ω is a sample space, $P : \Omega \to \mathbb{R}$ is a probability distribution, and $B \subseteq \Omega$ is an event such that $\mathrm{P}(B) > 0$, then a ***conditional probability distribution*** $P_B : \Omega \to \mathbb{R}$ can be defined by setting

$$\mathrm{P}_B(\sigma) = \begin{cases} \frac{\mathrm{P}(\sigma)}{\mathrm{P}(B)} & \text{if } \sigma \in B, \\ 0 & \text{if } \sigma \notin B \end{cases}$$

for every outcome $\sigma \in \Omega$.

Definition: If X is a random variable then the ***conditional expectation of X given B*** is the expected value of X with the respect to the conditional probability P_B:

$$\mathrm{E}[X \mid B] = \sum_{\sigma \in \Omega} \mathrm{P}_B(\sigma) \times X(\sigma).$$

Conditional Expectation

Continuing this example, let us consider the event

$$B = \text{“First toss is } \mathtt{H}\text{”},$$

that is, the event

$$B = \{(\mathtt{H},\mathtt{H},\mathtt{H}),(\mathtt{H},\mathtt{H},\mathtt{T}),(\mathtt{H},\mathtt{T},\mathtt{H}),(\mathtt{H},\mathtt{T},\mathtt{T})\}.$$

Now

$$\mathrm{P}_B((\mathtt{H},\mathtt{H},\mathtt{H})) = \frac{\mathrm{P}((\mathtt{H},\mathtt{H},\mathtt{H}))}{\mathrm{P}(B)} = \frac{1/8}{1/2} = \frac{1}{4}$$

and

$$\mathrm{P}_B((\mathtt{H},\mathtt{H},\mathtt{T})) = \frac{\mathrm{P}((\mathtt{H},\mathtt{H},\mathtt{T}))}{\mathrm{P}(B)} = \frac{1/8}{1/2} = \frac{1}{4}$$

since $\mathrm{P}((\mathtt{H},\mathtt{H},\mathtt{H})) = \mathrm{P}((\mathtt{H},\mathtt{H},\mathtt{T})) = \frac{1}{8}$.

Conditional Expectation

Indeed,

$$\mathsf{P}_B((\mathtt{H},\mathtt{T},\mathtt{H})) = \mathsf{P}_B((\mathtt{H},\mathtt{T},\mathtt{T})) = \frac{1}{4}$$

as well.

On the other hand,

$$\begin{aligned}\mathsf{P}_B((\mathtt{T},\mathtt{H},\mathtt{H})) = \mathsf{P}_B((\mathtt{T},\mathtt{H},\mathtt{T}))& \\ = \mathsf{P}_B((\mathtt{T},\mathtt{T},\mathtt{H})) &= \mathsf{P}_B((\mathtt{T},\mathtt{T},\mathtt{T})) = 0,\end{aligned}$$

since none of $(\mathtt{T},\mathtt{H},\mathtt{H})$, $(\mathtt{T},\mathtt{H},\mathtt{T})$, $(\mathtt{T},\mathtt{T},\mathtt{H})$ or $(\mathtt{T},\mathtt{T},\mathtt{T})$ belong to B.

Conditional Expectation

It now follows that

$$
\begin{aligned}
&\mathrm{E}[X \mid B] \\
&\quad = \mathrm{P}_B((\mathtt{H},\mathtt{H},\mathtt{H})) \times X((\mathtt{H},\mathtt{H},\mathtt{H})) + \mathrm{P}_B((\mathtt{H},\mathtt{H},\mathtt{T})) \times X((\mathtt{H},\mathtt{H},\mathtt{T})) \\
&\quad + \mathrm{P}_B((\mathtt{H},\mathtt{T},\mathtt{H})) \times X((\mathtt{H},\mathtt{T},\mathtt{H})) + \mathrm{P}_B((\mathtt{H},\mathtt{T},\mathtt{T})) \times X((\mathtt{H},\mathtt{T},\mathtt{T})) \\
&\quad + \mathrm{P}_B((\mathtt{T},\mathtt{H},\mathtt{H})) \times X((\mathtt{T},\mathtt{H},\mathtt{H})) + \mathrm{P}_B((\mathtt{T},\mathtt{H},\mathtt{T})) \times X((\mathtt{T},\mathtt{H},\mathtt{T})) \\
&\quad + \mathrm{P}_B(\mathtt{T},\mathtt{T},\mathtt{H})) \times X((\mathtt{T},\mathtt{T},\mathtt{H})) + \mathrm{P}_B((\mathtt{T},\mathtt{T},\mathtt{T})) \times X((\mathtt{T},\mathtt{T},\mathtt{T})) \\
&\quad = \frac{1}{4} \times 3 + \frac{1}{4} \times 2 + \frac{1}{4} \times 2 + \frac{1}{4} \times 1 + 0 \times 2 + 0 \times 1 \\
&\qquad\qquad + 0 \times 1 + 0 \times 0 \\
&\quad = \frac{1}{4} \times 8 + 0 = 2.
\end{aligned}
$$

Conditional Expectation

Exercise: Prove that Ω is a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$, $B \subseteq \Omega$ is an event such that $\mathrm{P}(B) > 0$, and X is a random variable, then

$$\mathrm{E}[X \mid B] = \frac{1}{\mathrm{P}(B)} \times \sum_{\sigma \in B} (\mathrm{P}(\sigma) \times X(\sigma)) \, .$$

Conditional Expectation

Recall that — since P_B is also a probability distribution for the sample space Ω whenever B is an event such that $\text{P}(B) > 0$, properties of *probability distributions* could be applied to P_B, in order to establish corresponding results for *conditional probabilities*.

- Since a *conditional expectation* is just the expected value of a random variable, defined using the conditional probability distribution P_B, properties of expectations can be used to establish corresponding properties of conditional expectations, in essentially the same way.

Conditional Expectation

The following result is useful just as the "Law of Total Probability" (Claim #3 from Lecture #19) is: It describes a way to compute an expectation by considering cases (corresponding to whether, or not, an event has occurred). Recall that if $B \subseteq \Omega$, then

$$B^C = \{\sigma \in \Omega \mid \sigma \notin B\}.$$

Claim #1: Let Ω be a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$, let $B \subseteq \Omega$ be an event such that $\mathrm{P}(B) > 0$ and $\mathrm{P}(B^C) > 0$, and let X be a random variable. Then

$$\mathrm{E}[X] = \mathrm{E}[X \mid B] \times \mathrm{P}(B) + \mathrm{E}[X \mid B^C] \times \mathrm{P}(B^C).$$

Proof: ***Exercise.***

Linearity of Expectation

Let Ω be a sample space with a probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$, and let $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ be random variables, for some positive integer n. "$X_1 + X_2 + \cdots + X_n$" denotes a random variable such that

$$(X_1 + X_2 + \cdots + X_n)(\sigma) = X_1(\sigma) + X_2(\sigma) + \cdots + X_n(\sigma)$$

for each outcome $\sigma \in \Omega$. Similarly, if $X : \Omega \to \mathbb{R}$ and $a, b \in \mathbb{R}$ then "$a \cdot X + b$" denotes a random variable such that

$$(aX + b)(\sigma) = a \cdot X(\sigma) + b$$

for every outcome $\sigma \in \Omega$, as well.

Linearity of Expectation

Claim #2 (Linearity of Expectation): Let Ω be a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$.

(a) If $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ are random variables over Ω, for a positive integer n, then

$$\mathrm{E}[X_1 + X_2 + \cdots + X_n] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n].$$

(b) If $X : \Omega \to \mathbb{R}$ is a random variable over Ω and $a, b \in \mathbb{R}$ then

$$\mathrm{E}[a \cdot X + b] = a \cdot \mathrm{E}[X] + b.$$

Proof: Another ***exercise***.

Linearity of Expectation

Once again, consider the random variable

$$X = \text{“Number of Heads Tossed”}.$$

Note that

$$X = X_1 + X_2 + X_3$$

where, for $1 \leq i \leq 3$ and $\sigma = (\sigma_1, \sigma_2, \sigma_3) \in \Omega$,

$$X_i(\sigma) = X_i((\sigma_1, \sigma_2, \sigma_3)) = \begin{cases} 1 & \text{if } \alpha_i = \mathtt{H}, \\ 0 & \text{if } \alpha_i = \mathtt{T}. \end{cases}$$

Linearity of Expectation

Note that

$$
\begin{aligned}
\mathsf{E}[X_1] &= \sum_{\sigma \in \Omega} \mathsf{P}(\sigma) \times X_1(\sigma) \\
&= \mathsf{P}((\mathtt{H},\mathtt{H},\mathtt{H})) \times X_1((\mathtt{H},\mathtt{H},\mathtt{H})) + \mathsf{P}((\mathtt{H},\mathtt{H},\mathtt{T})) \times X_1((\mathtt{H},\mathtt{H},\mathtt{T})) \\
&\quad + \mathsf{P}((\mathtt{H},\mathtt{T},\mathtt{H})) \times X_1((\mathtt{H},\mathtt{T},\mathtt{H})) + \mathsf{P}((\mathtt{H},\mathtt{T},\mathtt{T})) \times X_1(\mathtt{H},\mathtt{T},\mathtt{T})) \\
&\quad + \mathsf{P}((\mathtt{T},\mathtt{H},\mathtt{H})) \times X_1((\mathtt{T},\mathtt{H},\mathtt{H})) + \mathsf{P}((\mathtt{T},\mathtt{H},\mathtt{T})) \times X_1((\mathtt{T},\mathtt{H},\mathtt{T})) \\
&\quad + \mathsf{P}((\mathtt{T},\mathtt{T},\mathtt{H})) \times X_1((\mathtt{T},\mathtt{T},\mathtt{H})) + \mathsf{P}((\mathtt{T},\mathtt{T},\mathtt{T})) \times X_1((\mathtt{T},\mathtt{T},\mathtt{T})) \\
&= \frac{1}{8} \times 1 + \frac{1}{8} \times 1 + \frac{1}{8} \times 1 + \frac{1}{8} \times 0 + \frac{1}{8} \times 0 \\
&\qquad + \frac{1}{8} \times 0 + \frac{1}{8} \times 0 \\
&= \frac{4}{8} = \frac{1}{2}.
\end{aligned}
$$

Linearity of Expectation

Exercise: Confirm that

$$E[X_2] = E[X_3] = \frac{1}{2}$$

as well.

It now follows that

$$\begin{aligned} E[X] &= E[X_1 + X_2 + X_3] && \text{(since } X = X_1 + X_2 + X_3\text{)} \\ &= E[X_1] + E[X_2] + E[X_3] && \text{(by Linearity of Expectation)} \\ &= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = \frac{3}{2} \end{aligned}$$

— as previously noted. “Linearity of Expectation” can give a way to compute the expected value of a complicated random variable by considering simpler ones, instead.

Independent Random Variables

It follows by Claim #2 that if X_1 and X_2 are random variables over Ω, then $E[X + Y] = E[X] + E[Y]$.

- It might be reasonable to *wonder whether*

$$E[X \times Y] = E[X] \times E[Y]$$

  as well?

- It turns out that ***this is not generally true***.

Independent Random Variables

Consider, again, the random variable

$$X_1 : \Omega \to \mathbb{R}$$

— recalling that

$$X_1((\mathtt{H}, \mathtt{H}, \mathtt{H})) = X_1((\mathtt{H}, \mathtt{H}, \mathtt{T})) = X_1((\mathtt{H}, \mathtt{T}, \mathtt{H})) = X_1((\mathtt{H}, \mathtt{T}, \mathtt{T})) = 1$$

and

$$X_1((\mathtt{T}, \mathtt{H}, \mathtt{H})) = X_1((\mathtt{T}, \mathtt{H}, \mathtt{T})) = X_1((\mathtt{T}, \mathtt{T}, \mathtt{H})) = X_1((\mathtt{T}, \mathtt{T}, \mathtt{T})) = 0.$$

Independent Random Variables

Let $Y = X = X_1$ — noting, in this case, that

$$X \times Y(\sigma) = \begin{cases} 1 & \text{if } X_1(\sigma) = 1, \\ 0 & \text{if } X_1(\sigma) = 0. \end{cases}$$

This can be used to establish that

$$\mathrm{E}[X \times Y] = \mathrm{E}[X_1] = \frac{1}{2},$$

while

$$\mathrm{E}[X] \times \mathrm{E}[Y] = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

so that

$$\mathrm{E}[X \times Y] \neq \mathrm{E}[X] \times \mathrm{E}[Y]$$

in this case.

Independent Random Variables

Once again, suppose that Ω is a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$. Consider random variables $X : \Omega \to V_X$ and $Y : \Omega \to V_Y$, where $V_X, V_Y \subseteq \mathbb{R}$. Recall that, for $a \in V_X$ and $b \in V_Y$,

$$\text{“}X = a\text{”} = \{\sigma \in \Omega \mid X(\sigma) = a\}$$

and

$$\text{“}Y = b\text{”} = \{\tau \in \Omega \mid Y(\tau) = b\},$$

so that

$$\text{“}X = a \wedge Y = b\text{”} = \{\mu \in \Omega \mid X(\mu) = a \text{ and } Y(\mu) = b\}.$$

Independent Random Variables

Definition: The above random variables X and Y are ***independent*** if

$$P(X = a \wedge Y = b) = P(X = a) \times P(Y = b)$$

for all values $a \in V_X$ and $b \in V_Y$.

In other words, X and Y are "independent" if *all* events corresponding to choices of values for X and for Y, respectively, are independent.

Independent Random Variables

Continuing the above example, consider the random variables

$$X_1 : \Omega \to \mathbb{R} \text{ and } X_2 : \Omega \to \mathbb{R}$$

as previously described — so that

$$X_1(\sigma) \in \{0, 1\} \text{ and } X_2(\sigma) \in \{0, 1\}$$

for all $\sigma \in \Omega$.

Exercise: Confirm that

$$\mathrm{P}(X_1 = 0) = \mathrm{P}(X_1 = 1) = \mathrm{P}(X_2 = 0) = \mathrm{P}(X_2) = \frac{1}{2}.$$

Independent Random Variables

Confirm, as well, that

$$P(X_1 = i \wedge X_2 = j) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2}$$

for all $i, j \in \{0, 1\}$.

Thus

$$P(X_1 = i \wedge P(X)_2 = j) = P(X_i = j) \times P(X_2 = j)$$

for all $i, j \in \{0, 1\}$ — so that the random variables X_1 and X_2 are independent.

Pairwise Independent Random Variables

Once again, let Ω be a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$. Let k be a positive integer and let

$$X_1 : \Omega \to V_1, X_2 : \Omega \to V_2, \ldots, X_k : \Omega \to V_k$$

be random variables over Ω (so that $V_1, V_2, \ldots, V_k \subseteq \mathbb{R}$).

Definition: The random variables $X_1, X_2, \ldots, X_k$ are ***pairwise independent*** if the random variables X_i and X_j are independent, for every pair of numbers i and j such that $1 \leq i < j \leq k$.

Mutually Independent Random Variables

Finally, let Ω be a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$. Let k be a positive integer and let

$$X_1 : \Omega \to V_1, X_2 : \Omega \to V_2, \ldots, X_k : \Omega \to V_k$$

be random variables over Ω (so that $V_1, V_2, \ldots, V_k \subseteq \mathbb{R}$).

Definition: The random variables $X_1, X_2, \ldots, X_k$ are ***mutually independent*** if the following condition is satisfied: ***For every*** subset $S \subseteq \{1, 2, \ldots, k\}$ and ***for all*** combinations of $a_i \in V_i$, for $i \in S$,

$$\mathrm{P}\left(\bigwedge_{i \in S} (X_i = a_i)\right) = \prod_{i \in S} \mathrm{P}(X_i = a_i).$$

With a bit of work one can show that this is *equivalent* to the definition of the independence of a pair of random variables, X_1 and X_2, when $k = 2$.

Mutually Independent Random Variables

- Returning to the example of tossing three coins once again — with quite a bit more work, it can be shown that the random variables X_1, X_2 and X_3, included in the example, are mutually independent.
- If $k \geq 3$ and $X_1, X_2, \ldots, X_k$ are random variables, as above, and $X_1, X_2, \ldots, X_k$ are ***mutually independent***, then $X_1, X_2, \ldots, X_k$ must be ***pairwise independent*** as well.
- However, it is possible that $X_1, X_2, \ldots, X_k$ are *pairwise independent*, but ***not*** *mutually independent*.