# Probability–The Science of Uncertainty and Data

## edX Course: 6.431x

semester="3T2018" info_sidebar_name="Related Links" display_coursenumber="6.431x" course_organization="MITx"

November 4, 2018

## Contents

# Exam 1

## Exam 1

## 1. Exam Rules

### Exam Rules

1. You have opened a timed exam with a **48 hours** time limit. Please use the timer to see the time remaining. If you had opened this exam too close to the exam **closing time, October 9 23:59UTC** , you will not have the full 48 hours, and the exam will close at the closing time.

2. This is an **open book exam** and you are allowed to refer back to all course material and use (online) calculators. However, you must abide by the honor code, and not ask for answers directly from any aide.

3. You will be given **no feedback** during the exam. This means that unlike in the problem sets, you will not be shown whether any of your answers are correct or not. This is to test your understanding, to prevent cheating, and to encourage you to try your very best before submitting. Solutions will be available after the exam closes.

4. You will be given **2 attempts** for each, multipart problem, **except for True or False problems**. Since you will be given no feedback, the extra attempt will be useful only in case you hit the "submit" button in a haste and wish to reconsider. With no exception, **your last submission will be the one that counts** .

5. While the exam is open, you are **not allowed to post on the discussion forum on anything related to the exam** , except to report bugs/platform difficulties. If you think you have found a bug, please only state what needs to be checked. You can still go through Unit 5, and post on its contents, but the post must not shed any light on the contents or concepts in the exam. **Violators will receive a 0 score in Exam 1** .

## 2. Selecting different days from a year

### Problem 1. Selecting different days from a year

A set of 60 different days is selected from a given year. Assume that all sets of cardinality 60 are equally likely. Also, for simplicity, assume that the year has only 360 days, divided into twelve 30-day months. Evaluate the probabilities of the following events.

1. Exactly 5 days are selected from each of the 12 months.
2. None of the selected days is from January.
3. There exist 3 different months such that exactly 20 days are selected from each one of these months.

**Solution:**

1. The number of ways of selecting 60 different days out of 360 days is $\binom{360}{60}$.

   The number of ways of selecting 5 days from each month is

   $$\left(\binom{30}{5}\right)^{12},$$

   since from, say, January, 5 different days can be selected in $\binom{30}{5}$ different ways; the same holds for February, March, etc. Therefore, the answer is

   $$\frac{\left(\binom{30}{5}\right)^{12}}{\binom{360}{60}}.$$

2. We have to exclude the 30 days from January and select 60 days among the remaining 330 days. This can be done in $\binom{330}{60}$ different ways. Hence, the answer is

   $$\frac{\binom{330}{60}}{\binom{360}{60}},$$

   where the denominator is the total number of ways of choosing 60 different days from a given year.

3. We first select 3 different months, out of the 12 months. This can be done in $\binom{12}{3}$ different ways. For these 3 months, we select 20 days from each, which can be done in

$$\left(\binom{30}{20}\right)^3$$

different ways. Hence, the answer is

$$\frac{\binom{12}{3}\cdot\left(\binom{30}{20}\right)^3}{\binom{360}{60}}.$$

# 3. Independence: True or false

## Problem 2. Independence: True or false

Determine whether each of the following statements about events $A, B, C$ is always true or not.

1. Suppose that $A, B$, and $C$ are independent events; then $A^c$ and $B \cup C^c$ are independent.

2. From now on, we do not assume that $A, B, C$ are independent. Suppose that $A$ is independent of $B$, given $C$.

   (a) $A^c$ is independent of $B$, given $C$. (b) $A \cap C$ is independent of $B \cap C$, given $C$.
   (c) $A$ is independent of $B$, given $C^c$.

**Solution:**

1. True. This follows from the intuitive meaning of event independence: when $A$, $B$, and $C$ are independent, the occurrence or non-occurrence of some of these events does not provide any information on the occurrence or non-occurrence of the remaining ones.

   A formal proof is also possible but is somewhat tedious.

2. (a) True. We know that if $A$ and $B$ are independent, then $A^c$ and $B$ are independent. We now apply this fact to the conditional universe, and obtain the validity of the statement in question.

   (b) True. Within the conditional universe where $C$ is known to have occurred, the events $A \cap C$ and $B \cap C$ coincide with the events $A$ and $B$, respectively. The truth of the statement follows because we assumed that $A$ and $B$ are independent given $C$.

   (c) False. Let $X$ and $Y$ be independent binary random variables and let $A = \{X = 1\}$, $B = \{Y = 1\}$, $C = A \cap B = \{X = Y = 1\}$. Given $C$, the random variables $X$ and $Y$ are deterministic, hence trivially independent. On the other hand, given $C^c = A^c \cup B^c = \{X = 0\} \cup \{Y = 0\}$, the event $X = 0$ allows for both $Y = 1$ or $Y = 0$ to be true, whereas the event $X = 1$ conveys the information that $Y = 0$ must be true. In particular, $X$ and $Y$ are not independent given $C$. Equivalently, $A$ and $B$ are not independent given $C$.

# 4. Indicator notation

## Problem 3. Indicator notation

Let $A, B, C$ be three events, and let $X = I_A$, $Y = I_B$, and $Z = I_C$ be the associated indicator random variables.
We already know that $X \cdot Y$ is the indicator random variable of the event $A \cap B$. In the same spirit, give an algebraic expression, involving $X, Y$ and $Z$, for the indicator random variable of the following events.

**Note:** Express your answers in terms of $X, Y$, and $Z$ (the answer box is case sensitive) using standard notation.

1. The event $A^c \cap B \cap C$.
2. At most two of the events $A, B, C$ occurred.

`./xml/standardnotation.xml`

**Solution:**

1. The indicator random variable for $A^c$ is $1 - X$. Hence, the indicator random variable for the event, $A^c \cap B \cap C$ is $(1 - X)YZ$.
2. The event of interest is the complement of the event that all three events have occurred, and it is the same as
$$(A \cap B \cap C)^c,$$
Thus, the associated indicator random variable is,
$$1 - XYZ.$$

# 5. Expectation values

## Problem 4. Expectation practice

Let $X, Y, Z$ be independent discrete random variables with

$$\mathbf{E}[X] = 2 \qquad \mathbf{E}[Y] = 0 \qquad \mathbf{E}[Z] = 0,$$

$$\mathbf{E}[X^2] = 20 \qquad \mathbf{E}[Y^2] = \mathbf{E}[Z^2] = 16,$$

and

$$\mathsf{Var}(X) = \mathsf{Var}(Y) = \mathsf{Var}(Z) = 16.$$

Let $A = X(Y + Z)$ and $B = XY$.

1. Find $\mathbf{E}[B]$. $\mathbf{E}[B] =$
2. Find $\mathsf{Var}(B)$. $\mathsf{Var}(B) =$
3. Find $\mathbf{E}[AB]$. $\mathbf{E}[AB] =$

**Solution:**

1. $\mathbf{E}[B] = \mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y] = 0$, using the fact that $X$ and $Y$ are independent.
2. Note that
$$\mathsf{Var}(B) = \mathbf{E}[B^2] - \mathbf{E}[B]^2 = \mathbf{E}[B^2],$$
using the answer in the previous part. Next, $\mathbf{E}[B^2] = \mathbf{E}[X^2]\mathbf{E}[Y^2]$, using independence. Since, $\mathbf{E}[Y^2] = 16$, and since $\mathbf{E}[X^2] = 20$, we obtain $\mathsf{Var}(B) = 320$.
3. We have
$$\mathbf{E}[AB] = \mathbf{E}[X^2 Y(Y + Z)] = \mathbf{E}[X^2]\mathbf{E}[Y(Y + Z)]$$
using independence. We know that, $\mathbf{E}[X^2] = 20$. Now, for the second term,
$$\mathbf{E}[Y(Y + Z)] = \mathbf{E}[Y^2] + \mathbf{E}[YZ] = \mathbf{E}[Y^2] + \mathbf{E}[Y]\mathbf{E}[Z],$$
using linearity of expectation, and independence. Note that, $\mathbf{E}[Y]\mathbf{E}[Z] = 0$. Thus $\mathbf{E}[AB] = \mathbf{E}[X^2]\mathbf{E}[Y^2] = 320$.

**Problem 5. Expectation: True or False**

With $A, B, X$ defined as before, determine whether the following statements are true or false:

1. $A$ and $B$ are independent.
2. $A$ and $B$ are conditionally independent, given $X = 0$.
3. $A$ and $B$ are conditionally independent, given $X = 1$.

**Solution:**

1. Note that
$$\mathbf{E}[AB] = 320 \neq \mathbf{E}[A]\mathbf{E}[B] = 0.$$
Hence, $A$ and $B$ cannot be independent. Intuitively, this is because they are both affected by $X$.

2. Given $X = 0$ both $A$ and $B$ are identically equal to 0. Hence, they are independent. (Deterministic random variables are always independent; one does not carry any new information on the other.)

3. Given $X = 1$, we have
$$\begin{aligned} \mathbf{E}[AB \mid X = 1] = & \quad \mathbf{E}[X(Y+Z)XY \mid X = 1] = \mathbf{E}[(Y+Z)Y \mid X = 1] \\ = & \quad \mathbf{E}[(Y+Z)Y] = \mathbf{E}[Y^2] + \mathbf{E}[Z]\mathbf{E}[Y] \\ = & \quad \mathbf{E}[Y^2] = 16 \neq 0. \end{aligned}$$
On the other hand,
$$\mathbf{E}[A \mid X = 1] = \mathbf{E}[X(Y+Z) \mid X = 1] = \mathbf{E}[(Y+Z) \mid X = 1] = \mathbf{E}[Y+Z] = 0.$$
Thus,
$$\mathbf{E}[AB|X = 1] \neq \mathbf{E}[A|X = 1]\mathbf{E}[B|X = 1],$$
and they are not conditionally independent, given $X = 1$. Intuitively, this is because both $A$ and $B$ are affected by $Y$.

# 6. Bob and his coins

## Problem 6. Bob and his coins

Bob has two coins, $A$ and $B$, in front of him. The probability of Heads at each toss is $p = 0.5$ for coin $A$ and $q = 0.9$ for coin $B$.

Bob chooses one of the two coins at random (both choices are equally likely).

He then continues with 5 tosses of the chosen coin; these tosses are conditionally independent given the choice of the coin.

Let:

$H_i$ : the event that Bob's $i$th coin toss resulted in Heads;

$N$ : the number of Heads in Bob's coin tosses.

1. For $i \in \{0, 1, \ldots, 5\}$, $p_N(i)$, the pmf of $N$, is of the form

$$\frac{1}{2}\binom{5}{a} b^5 + c\binom{5}{d} q^e (1-q)^f.$$

   Find the coefficients $a, b, \ldots, f$. Your answer can be either a number or an expression involving $i$.

   $a =$

   $b =$

   $c =$

   $d =$

   $e =$

   $f =$

2. Find $\mathbf{E}[N]$. $\mathbf{E}[N] =$

3. Find the conditional variance of $N$, in a conditional model where we condition on having chosen coin $A$ **and** the first two tosses resulting in Heads. $\mathsf{Var}(N \mid A, H_1, H_2) =$

4. Are the events $H_1$ and $\{N = 5\}$ independent?

5. Given that the 3rd toss resulted in Heads, what is the probability that coin $A$ was chosen?

`./xml/standardnotation.xml`

**Solution:**

1. Using the total probability theorem,

$$\begin{aligned} \mathbb{P}(N=i) &= \mathbb{P}(N=i \mid A)\mathbb{P}(A) + \mathbb{P}(N=i \mid B)\mathbb{P}(B) \\ &= \tfrac{1}{2}\tbinom{5}{i}p^5 + \tfrac{1}{2}\tbinom{5}{i}q^i(1-q)^{5-i} \end{aligned}$$

2. Using the total expectation theorem,

$$\mathbf{E}[N] = \frac{1}{2}(\mathbf{E}[N|A] + \mathbf{E}[N|B]) = \frac{1}{2}(5p+5q) = \frac{7}{2}.$$

3. Conditioned on $H_1, H_2$, we have $N = 2 + M$ where $M$ is the number of Heads in the last three tosses. Adding a constant does not change the variance. Thus,

$$\mathsf{Var}(N|A, H_1, H_2) = \text{var}(M|A) = 3p(1-p) = 0.75.$$

4. No. It suffices to show $\{N = 5\}$ and $H_1^c$ are not independent. By part 1, we have

$$\mathbb{P}(N=5) = \frac{1}{2}p^5 + \frac{1}{2}q^5 \neq 0,$$

whereas

$$\mathbb{P}(N = 5|H_1^c) = 0.$$

Since $\mathbb{P}(N = 5) \neq \mathbb{P}(N = 5|H_1^c)$, the events $N = 5$ and $H_1^c$ are not independent.

5. By Bayes' rule we have

$$\mathbb{P}(A|H_3) = \frac{\mathbb{P}(H_3|A)\mathbb{P}(A)}{\mathbb{P}(H_3)}.$$

We have $\mathbb{P}(A) = 0.5$ and $\mathbb{P}(H_3|A) = 0.5$. Also,

$$\mathbb{P}(H_3) = \mathbb{P}(H_3|A)\mathbb{P}(A) + \mathbb{P}(H_3|B)\mathbb{P}(B) = \frac{1}{2}(0.5 + 0.9) = 0.7.$$

Therefore,

$$\mathbb{P}(A|H_3) = \frac{0.5 * 0.5}{0.7} = \frac{5}{14} \sim 0.357.$$

# 7. Friendship and happiness

## Problem 7. Friendship and happiness

Consider a group of $n \geq 4$ people, numbered from 1 to $n$. For each pair $(i, j)$ with $i \neq j$, person $i$ and person $j$ are friends, with probability $p$. (Assume that friendship is a symmetric relation, i.e. if $i$ is friends with $j$, then $j$ is also friends with $i$.) Friendships are independent for different pairs. These $n$ people are seated around a round table. For convenience, assume that the chairs are numbered from 1 to $n$, clockwise, with $n$ located next to 1, and that person $i$ seated in chair $i$. In particular, person 1 and person $n$ are seated next to each other.

If a person is friends with both people sitting next to him/her, we say this person is **happy**. Let $H$ be the total number of happy people.

We will find $\mathbf{E}[H]$ and $\mathsf{Var}(H)$ by carrying out a sequence of steps. Express your answers below in terms of $p$ and/or $n$ using standard notation (or click on "STANDARD NOTATION" button below). Remember to use "*" for multiplication and to include parentheses where necessary.

We first work towards finding $\mathbf{E}[H]$.

1. Let $I_i$ be a random variable indicating whether the person seated in chair $i$ is **happy** or not (i.e., $I_i = 1$ if person $i$ is happy and $I_i = 0$ otherwise). Find $\mathbf{E}[I_i]$.

   For $i = 1, 2, \ldots, n$, $\mathbf{E}[I_i] =$

2. Find $\mathbf{E}[H]$.
   (Note: The notation $a \triangleq \mathbf{E}[H]$ means that $a$ is defined to be $\mathbf{E}[H]$. The simpler variable names will be used in the last question of this problem.)

   $a \triangleq \mathbf{E}[H] =$

   Since $I_1, I_2, \ldots, I_n$ are not independent, the variance calculation is more involved.

3. For any $k \in \{1, 2, \ldots, n\}$, find $\mathbf{E}[I_k^2]$.

   $b \triangleq \mathbf{E}[I_k^2] =$

4. For any $i \in \{1, 2, \ldots, n\}$, and under the convention $I_{n+1} = I_1$, find $\mathbf{E}[I_i I_{i+1}]$.

   $c \triangleq \mathbf{E}[I_i I_{i+1}] =$

5. Suppose that $i \neq j$ and that persons $i$ and $j$ are not seated next to each other. Find $\mathbf{E}[I_i I_j]$.

   $d \triangleq \mathbf{E}[I_i I_j] =$

6. Give an expression for $\mathsf{Var}(H)$, in terms of $n$, and the quantities $a, b, c, d$ defined in earlier parts. $\mathsf{Var}(H) =$

./xml/standardnotation.xml

**Solution:**

As the seating is circular, it is convenient to use the following notational convention: $I_0 = I_n$, $I_1 = I_{n+1}$, $I_2 = I_{n+2}$.

1. Recall that $I_i = 1$, namely, the $i^{\text{th}}$ person is happy, if and only if he/she is friends with both neighbors, which happens with probability $p^2$.

2. The total number of happy people is

$$H = \sum_{i=1}^{n} I_i.$$

   Therefore,

$$a = \mathbf{E}[H] = \sum_{i=1}^{n} \mathbf{E}[I_i] = np^2.$$

3. Since $I_k$ is either 0 or 1, we have $I_k^2 = I_k$. Therefore, $\mathbf{E}[I_k^2] = \mathbb{P}(I_k = 1) = p^2$.

4. The random variable $I_k I_{k+1}$ is 1, if and only if both persons $k$ and $k+1$ are happy. Equivalently, the pairs $(k-1, k)$, $(k, k+1)$, and $(k+1, k+2)$ are pairs of friends. As each of these events are independent, the probability that $I_k I_{k+1} = 1$ is $p^3$.

5. We note that $I_i I_j = 1$ if and only if both $i$ and $j$ are happy, i.e., when the four pairs

$$(i, i-1),\ (i, i+1),\ (j, j-1),\ (j, j+1)$$

   are pairs of friends, which happens with probability $p^4$.

6. We have $\mathsf{Var}(H) = \mathbf{E}[H^2] - \mathbf{E}[H]^2 = \mathbf{E}[(I_1 + \cdots + I_n)^2] - a^2$. When we expand the product, $(I_1 + \cdots + I_n)^2$, we obtain $n^2$ terms, of different types:
   (i) $n$ terms of the form $I_i^2$.
   (ii) $2n$ terms of the form $I_i I_{i+1}$, including terms such as $I_n I_{n+1}$, which is interpreted as $I_n I_1$.
   (iii) The remaining $n^2 - 3n$ terms of the form $I_i I_j$, where $i$ and $j$ are not seated next to each other.

   The expected value of a term of the form (i), (ii), (iii), is $b, c, d$, respectively. Therefore,

$$\mathbf{E}[(I_1 + \cdots + I_n)^2] = nb + 2nc + (n^2 - 3n)d,$$

   which gives,

$$\mathsf{Var}(H) = nb + 2nc + (n^2 - 3n)d - a^2.$$