**Lecture 1 Summary: July 27 2016.**
**Introduction to Probability**

**Overview:**

(1) Random Variables
(2) Probability distributions
(3) Expected value
(4) Probability (Joint, Marginal, and Conditional)
(5) Probability of a an event (number in $[0, 1]$)
(6) Independence

**What is Probability?** - A measure of how likely an event is.

The probability $P$ of some event $A$, denoted $P(A)$, is usually defined such that $P(\cdot)$ satisfies the Kolmogorov axioms, named after the Russian mathematician Andrey Kolmogorov:

(1) $0 \leq P(A) \leq 1$
(2) $P(\text{not } A) = P(A^{\mathrm{c}}) = 1 - P(A)$ ; either $A$ happens or not
(3) $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A, B)$
   **If** events $A$ and $B$ are mutually exclusive; (the realization of $A$ prevents the realization of $B$ and vice versa), we have that $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$ since $P(A, B) = 0$

**Note**: In set theory the compliment of set $A$, $A^{\mathrm{c}}$ stands for (not $A$) or event A is not realized. Additionally the union of $A$ and $B$, $(A \cup B)$ stands for ($A$ or $B$) as we discussed it in class.

Let's take a look at figure 1 on page 2: We can talk about Conditional and Joint Probability.

(1) For women who do not get screened, the probability of dying from breast cancer is 4/15.
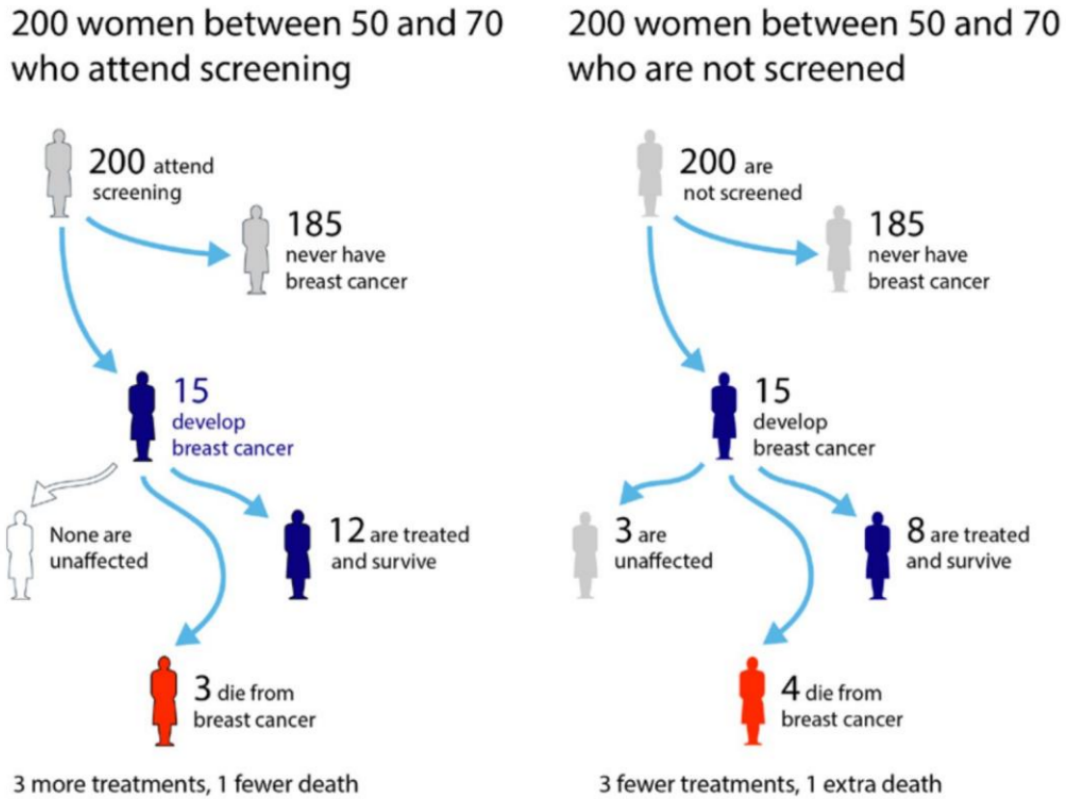(2) For women who get screened, the probability of dying from breast cancer is 3/15.

FIGURE 1. probability tree

**Conditional probability of** $A$ **given** $B$ is defined as the ratio
$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(A,B)}{P(B)} \quad (1).$$
Somewhat intuitively, in order to find out how often $A$ occurs, given that we know $B$ has occurred, we need to find the occurrences of $A$ and divide by the total number of possibilities. Since $B$ occurred, the occurrences of $A$ are exactly those situations in which both $A$ and $B$ occurred $(A, B)$, and since we're assuming $B$ occurred, the total number of possibilities are reduced to only those where (1) holds since the total number of possibilities in the expressions $P(A,B)$, $P(B)$ cancel.

Note: Conditional probability is **not symmetric**, meaning that $\mathbf{P(A|B) \neq P(B|A)}$. This can also be illustrated by the following example. If $A$ is the event of a person training hard and $B$ the event of a person being an NBA player. The common sense agrees in that given event $B$; a person being an NBA player, the probability that this person is training hard $P(A|B)$ is almost one, whereas given event $A$ a person practicing hard, the probability that this person is an NBA player $P(B|A)$

is basically zero.

**Joint probability** $P(A, B)$ of $A$ and $B$ is the probability of both events happening.

For example in the probability tree above, the probability of a woman developing breast cancer and surviving is $\frac{12}{200}$ (or $\frac{15}{200} \cdot \frac{12}{15}$).

So: conditional probability and joint probability are connected by the formula $P(A|B) = \frac{P(A,B)}{P(B)}$.

**Terminologies**:

**Random variable:** We can classify as random variables practically anything we are not sure about. For example the value of a stock index tomorrow, the value of a retirement portfolio in 35 years or the name of the next person that walks by.

We can describe the uncertainty about the random variables using a probability distribution. **Probability distribution** most intuitively is a list of possible outcomes accompanied by their probabilities; we could think of it as a representation of our uncertainty variable.

The list of possible outcomes is called **sample space**. The sample space can be **categorical** (heads/tails), **discrete** (anything that is countable, for example the integers from one to 6 ) or **continuous** (anything that can be between two real numbers).

**Expected value** $E(X) = \sum_{i=1}^{n} x_i P(X = x_i)$ is the weighted average of the possible outcomes ($x_i$ is the $i$th element of the sample space) where the weights are the corresponding probabilities . In the case of a continuous random variable, an integral is used in the place of sum.

**Joint Distributions and Marginal Probabilities**: Remember the World War II Example-story:

The army was looking for patterns to learn where to reinforce air bombers. The personnel reviewed some data and found some interesting patterns. At first they were mistaken because they used the wrong conditional probability than the one they needed and also assumed it was symmetric.

Let's go over quickly this example in order to remember the joint distributions and marginal probabilities we discussed about.

| Location | Number of planes |
|---|---|
| Engine | 53 |
| Cockpit area | 65 |
| Fuel system | 96 |
| Wings, fuselage, etc. | 434 |

FIGURE 2

| Location | Probability of hit |
|---|---|
| Engine | 0.08 |
| Cockpit area | 0.10 |
| Fuel system | 0.15 |
| Wings, fuselage, etc. | 0.67 |

FIGURE 3

| | Returned | Shot down |
|---|---|---|
| Engine | 53 | 57 |
| Cockpit area | 65 | 46 |
| Fuel system | 96 | 16 |
| Wings, fuselage, etc. | 434 | 33 |

FIGURE 4

We can convert the data listed in figure 2 to a probability distribution by making the percentages to sum to 1.

Then we will get the following table illustrated in figure 3 above.

These patterns unfortunately were of the planes that actually returned; not of the planes that were taken down. We can see Wald's reconstruction of the joint distribution in the figure above. Wald argued that the Navy was using the wrong probability and he was right!

| | Returned | Shot down | Marginal |
|---|---|---|---|
| Engine | 0.066 | 0.071 | 0.137 |
| Cockpit area | 0.081 | 0.058 | 0.139 |
| Fuel system | 0.120 | 0.020 | 0.140 |
| Wings, fuselage, etc. | 0.542 | 0.042 | 0.584 |
| Marginal | 0.809 | 0.191 | 1 |

FIGURE 5

**Marginal Probabilities**:

Again, in the same context, we can compute the marginal probabilities through summing across the relevant margin. (See figure 5 ) That is how marginal probability got its name.

Conditional Probability (think in term of frequencies rather than probabilities) and Independence :

We have already defined formally the conditional probability $P(A|B) = \dfrac{P(A,B)}{P(B)}$.

In the context of the previous example,

$$P(\text{returns}|\text{ engine hit}) = \frac{\text{Fraction taking engine hit and returning safely}}{\text{Fraction taking engine hit}}$$

(1)
$$= \frac{53/800}{110/800}$$
$$= \frac{0.066}{0.137}$$
$$\approx 0.48$$

**Independence**:

$A$ and $B$ are independent if $P(A) = P(A|B) = P(A|B^c)$.
That means knowing whether $B$ happened does not affect the probability of $A$ happening.
Be careful, a lot of fallacies can be boiled down to misunderstandings of this notion. For example:

Let $E$: the event that an American works for Enron and $L$: the event that an American works for Lehman Brothers.
What is the (joint) probability that an American chosen at random works for both firms? In other words, what is $P(E, L)$?

If we begin assuming that these two events $E, L$ are independent we might come up with the number $P(E, L) = P(E) \cdot P(L) \approx \frac{20,000}{200,000,000} \cdot \frac{26,000}{200,000,000} \approx 1.3 \times 10^{-8}$. This result is very alarming. There can't be true that there are only 2 people in the entire country who work in both firms. What led us to this false conclusion was our initial hypothesis about the independence of $L, E$. If they are not assumed independent (and indeed they are not independent of each other, they are part of the same industry), the correct calculation will come from the formula: $P(E, L) = P(E) \cdot P(L|E) = \frac{20,000}{200,000,000} \cdot \frac{26,000}{2,000} \approx 1.3 \times 10^{-6}$ which yields as an answer one American in a million, which makes much more sense.

Bottom line, we must be careful when making assumptions about independence.

**Synopsis of the formulas and/or important points:**

(1) $P(A^c) = 1 - P(A)$

(2) $P(A \cup B) = P(A) + P(B) - P(A, B)$ in general, unless $A$ and $B$ are mutually exclusive hence $P(A \cap B) = 0$

(3) $P(A|B) = \dfrac{P(A, B)}{P(B)} = \dfrac{P(A, B)}{P(B)}$

(4) $E(X) = \displaystyle\sum_{i=1}^{n} x_i P(X = x_i)$

(5) Joint probability $P(A, B)$ of $A$ and $B$ is the probability of both events happening

(6) Always remember that the probability of a an event is a number in $[0, 1]$ !