Independence

Gregory M. Shinault

Goal for this Lecture

Learn about the concept of independence. Topics covered will include

1. the definition of independence for two events,
2. the definition of independence for many events, and
3. the definition of independence for random variables.

The material of this lecture roughly corresponds to Section 2.3 of the textbook.

Independence of 2 Events

Big Idea

Independent is the term we use for when conditioning tells us nothing:

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$

The Flaw: $\mathbb{P}(A|B)$ is not defined if $\mathbb{P}(B) = 0$.

Definition

We say the events A and B are *independent* if

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B).$$

Example

Prove the outcomes in a sample with replacement are always independent.

More rigorously, we have a bin with r red marbles and b blue marbles. We randomly select 1 marble and record its color, and place it back in the bin. Then we select another marble and record its color. Let R_1 be the event the first marble is red and B_2 be the event the second marble is blue. Determine if R_1 and B_2 are dependent for any choice of parameters r and b.

Examples

Is it ever possible for two distinct outcomes in a sample without replacement to be independent?

More rigorously, we have a bin with r red marbles and b blue marbles. We randomly select 1 marble and record its color. Then we select another marble and record its color. Let R_1 be the event the first marble is red and B_2 be the event the second marble is blue. Determine if R_1 and B_2 are independent for any choice of parameters r and b.

Useful Facts

The following are equivalent.

1. A and B are independent.
2. A and B^c are independent.
3. A^c and B^c are independent.

Exercise: Prove the three statements are equivalent.

Common Mistake

WARNING: Do not confuse *independent* and *disjoint*.

Independence of Many Events

Definition

We call the events $A_1, A_2, \ldots, A_n$ *(mutually) independent* if and only if for any subcollection $A_{j_1}, \ldots, A_{j_k}$ we have

$$\mathbb{P}(A_{j_1} \cdots A_{j_k}) = \mathbb{P}(A_{j_1})\mathbb{P}(A_{j_2}) \cdots \mathbb{P}(A_{j_k}).$$

Exercise: Calculate how many equations must be checked to determine if n events are independent.

WARNING:
$A_1, A_2, \ldots, A_n$ are pairwise independent
$\not\Rightarrow A_1, A_2, \ldots, A_n$ are independent

Equivalent Definition

Fact: The events $A_1, A_2, \ldots, A_n$ are *(mutually) independent* if and only if

$$\mathbb{P}(A_1^* A_2^* \cdots A_n^*) = \mathbb{P}(A_1^*)\mathbb{P}(A_2^*) \cdots \mathbb{P}(A_n^*).$$

where A_j^* is either A_j or A_j^c.

(So that is 2^n equations to check!)

Example

The perfect-use failure rate of the combined oral contraceptive pill is 0.3% That means for a woman that uses the pill exactly as prescribed there is a 0.003 probability that she will become pregnant in one year of use. Let A_j denote the event that the woman becomes pregnant in the j-th year of use, and assume $A_1, \ldots, A_{10}$ are independent.

What is the probability (under these possibly flawed assumptions) a woman using the combined oral contraceptive pill experiences an unplanned pregnancy in the next 10 years?

Now update this probability for the typical-use failure rate of 9%.

Comments

I suspect the assumptions of the previous problem are flawed, but even if they are not these types of figures often lead our mind into the *gambler's fallacy*.

There is a good visualization of these numbers under the same assumptions in an old NYT article. `https://www.nytimes.com/interactive/2014/09/14/sunday-review/unplanned-pregnancies.html?_r=0`

An astute blogger quickly pointed out the flaws in the probability calculations. `https://andrewwhitby.com/2014/09/15/averages-deceive-birth-control-is-better-than-the-nyt-c`

Independence of Random Variables

Definition

Let $X_1, X_2, \ldots, X_n$ be random variables on the same sample space. Then $X_1, X_2, \ldots, X_n$ are *independent* if

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{k=1}^{n} \mathbb{P}(X_k \in B_k)$$

for all subsets $B_1, B_2, \ldots, B_n$ of $\mathbb{R}$.

Discrete Case

Fact: The discrete RVs $X_1, X_2, \ldots, X_n$ are independent if

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{k=1}^{n} \mathbb{P}(X_k = x_k)$$

for all choices $x_1, \ldots, x_n$.

Example

Suppose we take a size k sample from the set $\{1, 2, \ldots, n\}$. Let $X_1, \ldots, X_k$ denote the outcomes. Determine if the random variables are independent for

- a sample with replacement and
- a sample without replacement.

The Wrap Up

Summary

1. Independent $\Leftrightarrow \mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$.
2. Do not confuse independence and mutual disjointness.
3. Increasing independence to multiple events is complicated, but you should still know the definitions.
4. Independence of discrete RVs requires only looking at the individual outcomes.
5. The outcomes of a sample
   a. with replacement are independent,
   b. without replacement are dependent.

Next step

We take what we know from independence and look at a few classes of random variables that occur frequently.