5-11-04

PICK A GOOD ONE

Looking for an edge in Pick Six?

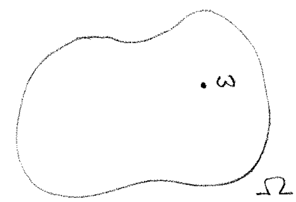
Most online random-number generators actually offer "pseudo-random" numbers because computers aren't good at doing anything by chance. To generate numbers that are truly random requires a source of entropy, or disorder, outside the computer itself.

A new site, **randomnumbers.info**, locates such a source in quantum physics – the reflection of a light particle on a semitransparent mirror. The site exploits this optical process to generate up to 1,000 random numbers on demand.

Other sites also offer true random numbers. **Random.org**, uses atmospheric noise from a radio as a source of disorder; the random numbers at HotBits (**www.fourmilab.ch/hotbits**) are generated by radioactive decay; and LavaRnd (**www.lavarnd.org**) taps the unpredictability of lava lamps.

In probability theory, you assume that all parameters are known; there is no need to estimate any numbers from data. The main task is to begin with a description of some "experiment" in which all parameters and distributions and whatever are known, and to derive from that certain probabilities or expected values that are not immediately obvious.

Probability setup:



Ω sample space. all possible outcomes

$A \subseteq \Omega$ event

Ex B = "right now, there are two students in the library who have the same birthday"

It's difficult to compute the probability of this!

$\mathbb{P}$ assigns a number to each event

axioms
$$\begin{cases} 0 \le \mathbb{P}(A) \le 1 \\ \mathbb{P}(\Omega) = 1 \\ \text{if } A_1, A_2, \ldots \text{ are disjoint, then } \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) \end{cases}$$

Law of total probability:

Suppose $A_1, A_2, \ldots$ are disjoint and $\bigcup_{n=1}^{\infty} A_n = \Omega$.

Then $\mathbb{P}(B) = \sum_{n=1}^{\infty} \mathbb{P}(B \cap A_n)$.

Use for this:

B = " .... same birthday".

This would be easier to tell if we knew <u>how many</u> people were in the library.

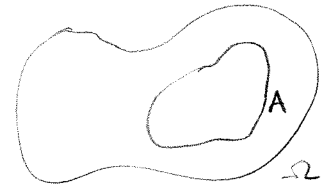
Let N = number of students in the library.

N can be $0, 1, 2, \ldots$.

$$B = (B \cap \{N=0\}) \cup (B \cap \{N=1\}) \cup \ldots$$

$$\mathbb{P}(B) = \sum_{n=0}^{\infty} \mathbb{P}(B \cap \{N=n\})$$

This is going to help.



Conditional probability

Suppose $\mathbb{P}(A) > 0$.

$$\mathbb{P}(B \mid A) \equiv \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$$

The idea is this: suppose we are conducting the experiment, and we learn that event A has occured, but we know no more than that.

How does this affect our understanding of the probability of B?

Compute relative to the set A.

For the birthday problem,

$$\mathbb{P}(B) = \sum_{n=0}^{\infty} \mathbb{P}(B \cap \{N=n\})$$

$$= \sum_{n=0}^{\infty} \mathbb{P}(B \mid N=n) \cdot \mathbb{P}(N=n)$$

This is very helpful!

$\mathbb{P}(B \mid N=n)$ you may have computed in a probability course. It is a matter of making some assumptions about how birthdays are distributed, then doing some counting.

$\mathbb{P}(N=n)$ has to do only with typical library usage. You can imagine knowing these. If you know some applied prob, you could argue that N has a Poisson distribution, which means that you only need to know the average number of people in the library at this time of day. You can estimate this from data — or just guess. Either way you'll gain some understanding.

Independence

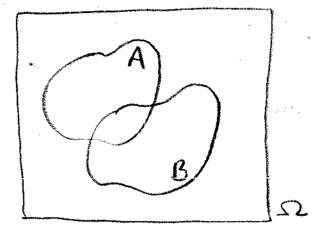
Events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Suppose $\mathbb{P}(A) > 0$. (Otherwise A and B are always independent!)

omit

Probability of B given A is defined to be

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$$



put yourself in A, with no further information, and ask what the prob. of B is now.

A and B are <u>independent</u> if $\mathbb{P}(B \mid A) = \mathbb{P}(B)$ or if $\mathbb{P}(A) = 0$.

(then $\dfrac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \mathbb{P}(B) \quad \ldots$)

<u>Example</u> The phone rings.

B = "caller wants to sell you something"

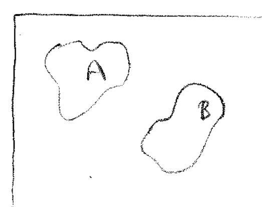
A = " caller says, 'Hey, how's it going?' "

$\mathbb{P}(B)$ is some appreciable number, like 0.2, before the phone rings.

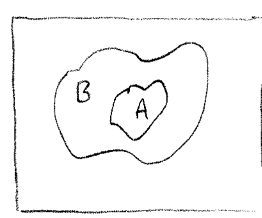
$\mathbb{P}(B \mid A)$ is very small, like 0.01, so they are dependent.

C = "you pick up the phone on the second ring"

<u>Example</u> Venn diagrams



$\mathbb{P}(B \mid A) = 0$



$\mathbb{P}(B \mid A) = 1$

Random variables

In this class, a random variable is a numerical quantity X whose value is not known ahead of time, but for which we know the probabilities that X takes on various values, (or at least we pretend to know them)

Contrast: prob. vs. stat.

Example Every morning, I walk to school.

T = amount of time between when I leave home and when I arrive at my building

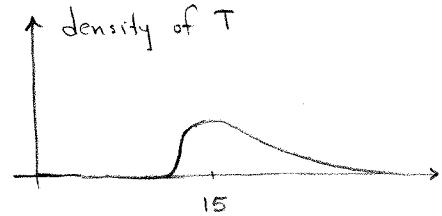
sensible to think of this as a continuous random variable,

possible values are $(0, \infty)$

$\mathbb{P}(T < 10) = 0$

$\mathbb{P}(10 < T < 20) = 0.8$

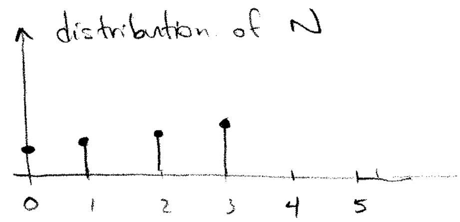
$\mathbb{E}T = 16$



N = number of people I meet out walking, on my way

possible values 0, 1, 2, 3, ...

discrete



Independence of random variables

Definition
Let X and Y be r.v.s
X and Y are indep. if for all x, y, ...
x_1, x_2, y_1, y_2.

What would it mean that T and N are independent?

Intuitively, knowledge of one gives you no information about the value of the other.

Definition: For all t, n, events "$T \le t$" and "$N \le n$" are independent.
Or, for all intervals, "$t_1 < T \le t_2$" "$n_1 < N \le n_2$" indep.

Are N and T independent?

A = "$N = 5$" B = "$T < 15$".

should have $\mathbb{P}(B \mid A) = \mathbb{P}(B)$ for all such events.

However, $\mathbb{P}(T < 15 \mid N = 5000)$ is less than $\mathbb{P}(T < 15)$ — how can I walk so quickly with 5000 people out walking?

Not strictly independent.

I might stop and talk w/ someone

Can we model them as indep.?

Independent and identically distributed?

Collections of random variables are often organized by an index.

For example, in $X_1, X_2, \ldots, X_n$, $1, 2, \ldots n$ are the indices, and the collection $\{X_1, \ldots, X_n\}$ is called a stochastic process indexed by $\{1, \ldots, n\}$ (see p. 5)

Stochastic to mean random, process because n is often time or a count.

Example

Begin with Sep. 1, 1998, a Friday.

Let $n = 1$ be that day, $n = 2$ is Sep. 2, ...

Let W_n be the number of 'wrong numbers' at your home on day n.

These are discrete random variables, values $0, 1, 2, \ldots$

Model: $\mathbb{P}(W_n = 0) = 0.98$ ...

The collection $\{W_1, W_2, \ldots\}$ is a stochastic process indexed by $\{1, 2, \ldots\}$

Are these random variables identically distributed?

That is, does $\mathbb{P}(W_n = a)$ depend on n?

More on Saturday?

Fewer over breaks when students are gone?

Are these random variables independent?

Suppose $W_1 = 0$.

Suppose $W_{10} = 5$. Why? Is W_{11} likely to be large?

Suppose we use the model that the W_n are iid,

$\mathbb{P}(W_n = 0) = 0.98$, $\mathbb{P}(W_n = 1) = 0.019$ $\mathbb{P}(W_n = 2) = 0.001$.

How long between wrong numbers?

Toss coin, prob. $0.02 = p$

N = day number of 1st wrong number.

$\mathbb{P}(N = 1) = 0.02 = p$

$\mathbb{P}(N = 2) = (0.98)(0.02) = q \cdot p$

$\vdots$

$$\sum_{n=1}^{\infty} n \cdot q^{n-1} \cdot p = p \cdot \left(\sum_{n=0}^{\infty} q^n\right)' = p\left(\frac{1}{1-q}\right)' = \frac{p}{p^2} = \frac{1}{p} = 50.$$

Example Go sit along the interstate somewhere.
Ignore the trucks, and look at traffic headed in one direction only.
Let n represent the number of each car (1^{st}, 2^{nd}, ...)
Let A_n be the model year of the car - 1989, 1997, ...

Are these identically distributed?
What could change?
What if you wait a whole year?!

Are they independent?

There goes a 1952 ... and a 1947 ....

maybe not one after another, but it may well be that people are headed for a car show, or it's Sunday, ...

Example I_n = number of people watching the evening national news on day n.

Identically distributed?

Independent?

Example S_n = price of Microsoft stock at the end of day n.

?

?

C_n = change in value on day n,
$= S_n - S_{n-1}$

What if these are not independent?
Then I could make money if I understood the dependence.

If they are independent, then I understand the structure of S_n better

They can't be independent, or the stock value could go under 0!

$R_n = \dfrac{S_n}{S_{n-1}}$ works better.

Collections of random variables

Starting on July 1, let M_n be the number of mosquito bites I receive on day n after July 1.

Are these random variables indep?

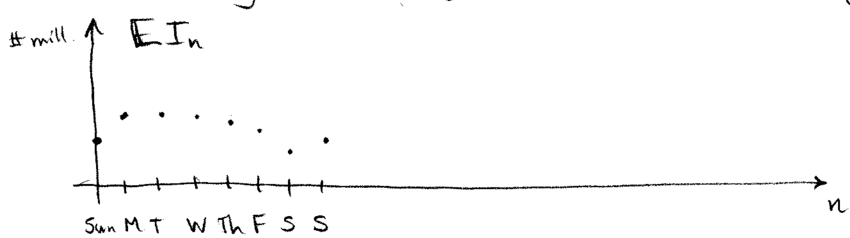
① W_n = # wrong numbers per day, dist'n may change, still indep.

② A_n = model years of cars passing certain spot on interstate. (ignore trucks)

caravans of old cars

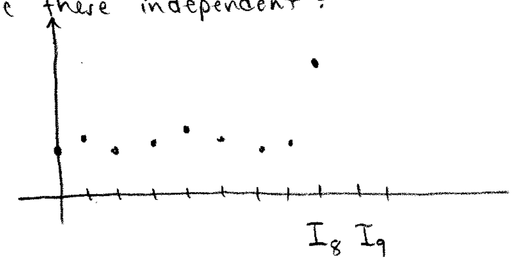
W_n = amount of water you drink on day n

I_n = # people watching the evening national news on day n.



take these as given

Are these independent?



S_n = trading price of Microsoft stock at the end of day n.

C_n = change in trading price on day