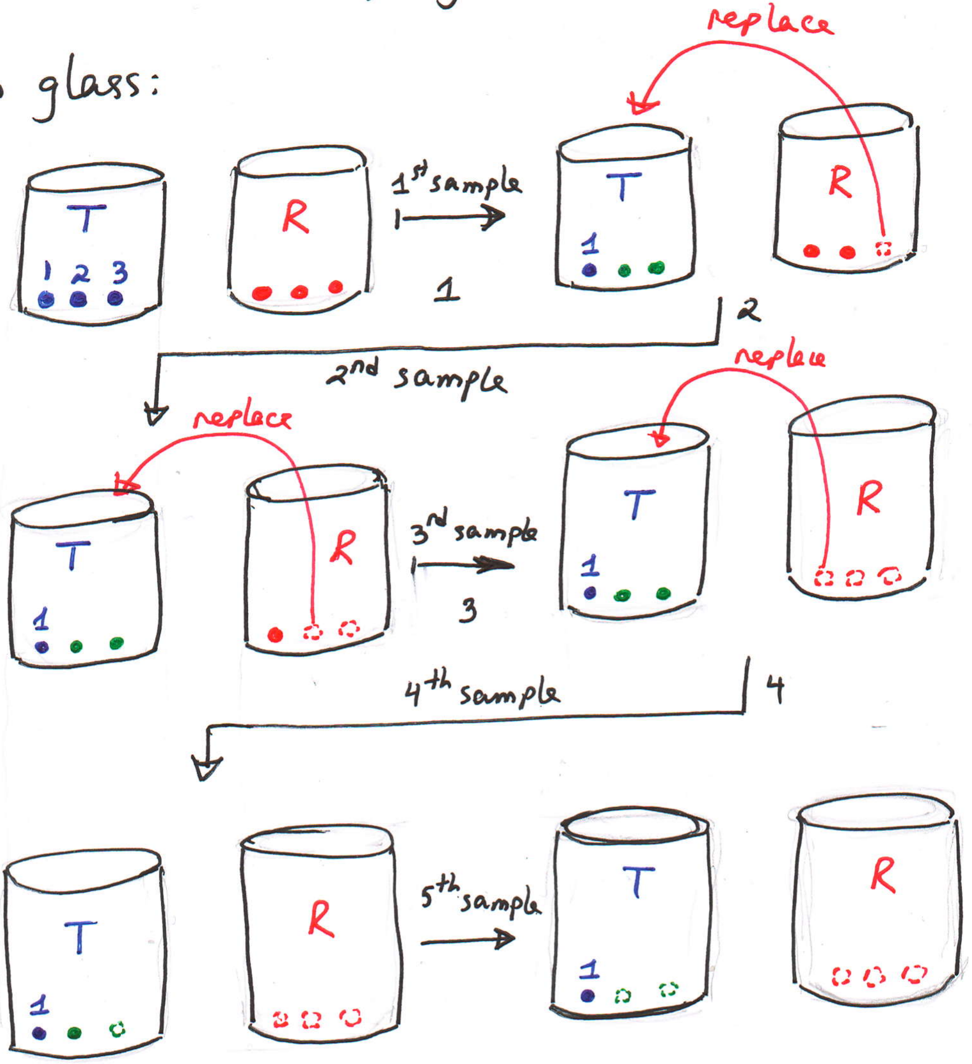
Conditional Probability Lecture 3

A few curious Problems

Ex. A mathematician likes to drink his tea with rum as follows: He fills up one full glass of tea and another identical glass with rum. He takes a sip of tea and then pours in the rum until the tea glass is full again. He stirs the mixture thoroughly and takes another sip. He Pours in more rum and repeats the process until both glasses are empty.
What is the probability that his last sip only contains tea?

Solution: To make the problem a little more precise, imagine that each sip is a random sampling of one molecule without replacement.

Let's see how it plays out with 3 molecules in each glass:



Let T - event last molecule to be sampled is tea.

Then $T = T_1 \cup T_2 \cup T_3$, where T_k - event last molecule is tea molecule k.

Clearly $P(T) = P(T_1) + P(T_2) + P(T_3) = 3\,P(T_1)$

$$= 3 \cdot \underbrace{\frac{2}{3}}_{\text{1 survived 1st sample}} \cdot \underbrace{\frac{2}{3}}_{\text{1 survived 2nd sample}} \cdot \underbrace{\frac{2}{3}}_{\text{3rd sample}} \cdot \underbrace{\frac{2}{3}}_{\text{4th sample}} \cdot \underbrace{\frac{1}{2}}_{\text{5th sample}} = \left(\frac{2}{3}\right)^3$$

Convince yourself that if there are 4 molecules in each beaker then the probability that a molecule of tea survives to the last is $P(T) = \left(\frac{3}{4}\right)^4$, with 5 molecules in each beaker, the probability is $p(T) = \left(\frac{4}{5}\right)^5$ and so on.

In general, there will be many molecules in each beaker so the approximation $\lim_{n \to \infty} \left(\frac{n-1}{n}\right)^n$ is

excellent... What is this limit?

Notice that $\lim_{n \to \infty} \left(\frac{n-1}{n}\right)^n = \lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^n$

$= e^{-1}$.

<u>Ex.</u> A jewlery display window was shattered in some trinket store on Madison Ave. and a shiny golden ring was stolen by one and only one of n suspects. Each of them is initially assumed to be guilty with probability $p = \frac{1}{n}$. Vasya Voronov is one of the suspects. An eye witness testifies that she saw the thief escaping in a black car.

An innocent person is likely to drive a black car with probability P_0 (ratio of black cars to all cars in the city perhaps). Furthermore, assume that the eye witness testimony is reliable with probability $P_1 > P_0$. Let $a = \frac{P_0}{P_1}$ and $b = \frac{1-P_1}{1-P_0}$

(a) If camera evidence shows that Vasya Voronov has driven away from the crime location, what is the probability that he is the perpetrator?

(b) Further investigation shows that Vasya is the only one from among the n suspects who drove off in a black car. What is the probability that he is guilty now?

Solution: Let G_k be the event suspect k is guilty for $k = 1, 2, \dots, n$, where suspect #n is Mr. Voronov.

Let B_k be the event suspect k is driving a black car (i.e. observed driving from crime scene in a black car).

$$(a)\quad P(G_n \mid B_n) = \frac{P(B_n \mid G_n)\, P(G_n)}{P(B_n \mid G_n)\, P(G_n) + P(B_n \mid G_n^c)\, P(G_n^c)}$$

$$= \frac{P_1 \frac{1}{n}}{P_1 \frac{1}{n} + P_0 \frac{n-1}{n}} = \frac{1}{1 + a(n-1)}$$

(b) $P(G_n \mid B_1^c B_2^c \dots B_{n-1}^c B_n) =$

$$= \frac{P(G_n B_1^c B_2^c \dots B_{n-1}^c B_n)}{P(B_1^c B_2^c \dots B_{n-1}^c B_n)}$$

Observe that $P(G_n B_1^c B_2^c \dots B_{n-1}^c B_n) =$

$$= P(G_n) P(B_1^c \mid G_n) P(B_2^c \mid G_n B_1^c) \dots P(B_n \mid G_n B_1^c B_2^c \dots B_{n-1}^c)$$

$$= \frac{1}{n} (1-P_0)^{n-1} P_1$$

And $P(B_1^c B_2^c \dots B_{n-1}^c B_n) = \sum_{k=1}^{n} P(G_k B_1^c B_2^c \dots B_{n-1}^c B_n)$

$$= \frac{1}{n}(1-P_0)^{n-1} P_1 + \sum_{k=1}^{n-1} P(G_k B_1^c B_2^c \dots B_{n-1}^c B_n)$$

$$= \frac{1}{n}(1-P_0)^{n-1} P_1 + (n-1) \boxed{P(G_1 B_1^c B_2^c \dots B_{n-1}^c B_n)}$$

where $P(G_1 B_1^c B_2^c \dots B_{n-1}^c B_n) =$

$$= P(G_1) P(B_1^c \mid G_1) P(B_2^c \mid G_1 B_1^c) \dots P(B_{n-1}^c \mid G_1 B_1^c \dots B_{n-2}^c) \cdot$$

$$\cdot P(B_n \mid G_1 B_1^c B_2^c \dots B_{n-1}^c) = \boxed{\frac{1}{n}(1-P_1)(1-P_0)^{n-2} P_0}$$

Thus $P(G_n \mid B_1^c B_2^c \ldots B_{n-1}^c B_n) =$

$$= \frac{\frac{1}{n}(1-P_0)^{n-1} P_1}{\frac{1}{n}(1-P_0)^{n-1} P_1 + \frac{n-1}{n}(1-P_1)(1-P_0)^{n-2} P_0}$$

$$= \frac{1}{1 + (n-1)ab}$$

<u>Ex.</u> A kidney study is looking at how well two different treatments (A and B) work on small and large kidney stones. Here is the succes rate that was found:

	Treatment A	Treatment B
$K =$ Small stones	81 out of 87 93%	234 out of 270 87%
$K^c =$ Large stones	192 out of 263 73%	55 out of 80 69%
Total	273 of 350 78%	289 of 350 83%

Which of the two treatments is better?

Solution: Notice that $P(A) = 0.78$ and $P(B) = 0.83$

It may seem that treatment A is inferior to treatment B. However $P(A|k) = 0.93$ v.s. $P(B|k) = 0.87$ indicates that treatment A is more effective than treatment B against small stones.

Similarly $P(A|k^c) = 0.73$ v.s. $P(B|k^c) = 0.69$ indicates that treatment A is more effective than treatment B against large kidney stones as well.

Thus $\boxed{\text{treatment A is better.}}$

Remark: The reason I labeled small kidneys with k is that kleine Nierensteine is German for small kidney stones.

Remark: While we're at it, I highly recommend reading Drei Kameraden (Three Comrades) by

Erich Maria Remarque. I am yet to meet a German who has read this novel, even though it is very popular in the Russian-speaking world. It is one of my favorite books and has become the favorite novel among two of my English speaking friends. Both of them have read more of his works by now than I have!

I trust that by now you have finished reading Dostoyevsky's The Gambler and are well on the way to Kafka's Das Schloss (The Castle).

The fallacy of assuming that treatment B is superior is called Simpson's Paradox.

To understand it, imagine we have two students. Student A is better at solving both easy and hard math questions as compared to student B.

However, suppose that student B is better at solving easy questions than student A is at solving hard ones. In symbols

$$P(A|H) < P(B|H^c)$$

Let P_A be the frequency with which A is given hard questions to solve. Let the corresponding frequency for B be P_B.

Then $P(A) = P(A|H)\,P_A + P(A|H^c)(1-P_A)$

Taking $\lim_{P_A \to 1} P(A)$ we obtain $P(A|H)$

Similarly $\lim_{P_B \to 0} P(B) = \lim_{P_B \to 0} P(B|H)P_B + P(B|H^c)(1-P_B)$

$= P(B|H^c)$

Thus, for suitable P_A, P_B we shall find

$P(A) < P(B)$.

In words, we are killing the good student with excruciating problems, while making life easy for the "poor student". Reminds you of something?

<u>Ex.</u> Under ancient Jewish law, if a suspect on trial was found guilty by all the judges (unanimously), then the suspect was acquitted.

Let's try to see why.

There are n judges deciding a case. The suspect has prior probability p of being guilty. Each judge votes whether to convict or acquit the suspect.

With probability s, a systemic error occurs (e.g. the defense is incompetent). If a systemic error occurs, then the judges unanimously vote to convict (i.e. all n judges vote to convict)

Whether a systemic error occurs is independent of whether the suspect is guilty. Given that a systemic error doesn't occur and that the suspect is guilty, each judge has probability c of voting to convict, independently. Given that a systemic error doesn't occur and that the suspect is not guilty, each judge has probability w

of voting to convict, independently. Suppose that

$$0 < p < 1, \quad 0 < s < 1, \quad \text{and} \quad 0 < w < \tfrac{1}{2} < c < 1$$

(a) Suppose that exactly k out of n judges vote to convict, where $k < n$. Given this information, find the probability that the suspect is guilty.

(b) Now suppose that all n judges vote to convict. Given this information, find the probability that the suspect is guilty.

(c) Is the answer to (b), viewed as a function of n, an increasing function?

<u>Solution:</u> Let C_k - event exactly k out of n judges vote to convict. Let J_i - event judge #i voted to convict, B - event of systemic bias, and G - event suspect is guilty.

(a) $$P(G|C_k) = \frac{P(C_k|G)\,P(G)}{P(C_k|G)\,P(G) + P(C_k|G^c)\,P(G^c)}$$

Now $$P(C_k|G) = \binom{n}{k} P(J_1 \dots J_k J_{k+1}^c \dots J_n^c | G)$$

$$= \binom{n}{k} c^k (1-c)^{n-k}$$

and $$P(C_k|G^c) = \binom{n}{k} \omega^k (1-\omega)^{n-k}$$

Hence $$P(G|C_k) = \frac{p\,c^k(1-c)^{n-k}}{p\,c^k(1-c)^{n-k} + (1-p)\,\omega^k(1-\omega)^{n-k}}$$

(b) $$P(G|C_n) = \frac{P(GBC_n) + P(GB^cC_n)}{P(GBC_n) + P(GB^cC_n) + P(G^cBC_n) + P(G^cB^cC_n)}$$

Now $$P(GBC_n) = P(G)\,P(B|G)\,P(C_n|GB)$$

$$= p \cdot s \cdot 1.$$

$$P(GB^cC_n) = P(G)\,P(B^c|G)\,P(C_n|GB^c)$$

$$= p(1-s)c^n$$

$$p(G^c B C_n) = p(G^c)\,p(B|G^c)\,p(C_n|G^c B)$$

$$= (1-p)\cdot s \cdot 1$$

$$p(G^c B^c C_n) = p(G^c)\,p(B^c|G^c)\,p(C_n|G^c B^c)$$

$$= (1-p)(1-s)\,\omega^n$$

Thus $$P(G|C_n) = \frac{p\left(s+(1-s)c^n\right)}{p\left(s+(1-s)c^n\right) + (1-p)\left(s+(1-s)\omega^n\right)}$$

(c) Notice that as n is increased both

$s+(1-s)c^n \approx s$ and $s+(1-s)\omega^n \approx s$, making

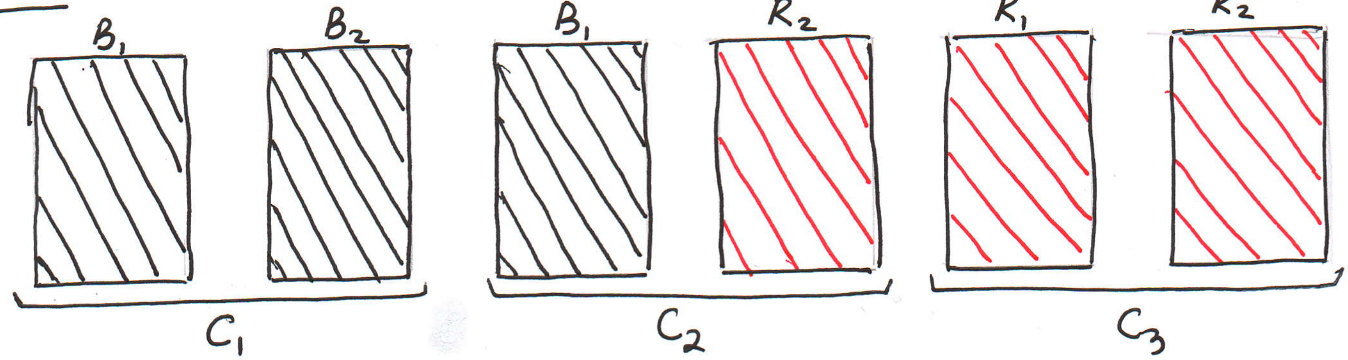
systemic bias more and more likely.

Thus $$\lim_{n\to\infty} P(G|C_n) = \frac{ps}{ps+(1-p)s}$$

$$= \frac{ps}{(p+1-p)s} = p = P(G).$$

One wonders if the converse indicates that the suspect is guilty. That is, when the suspect is unanimously acquitted (or no investigation is ever undertaken)

Ex.



3 cards are shuffled. One card is picked at random and placed on table. If upturned side is red, what is the probability that both sides are red?

Solution: Let R - event upturned side is red.

We want $p(C_3|R) = \dfrac{p(R|C_3)\,p(C_3)}{p(R|C_3)p(C_3) + p(R|C_2)\,p(C_2)}$

$$= \frac{p(R|C_3)}{p(R|C_3) + p(R|C_2)} = \frac{1}{1+p}$$

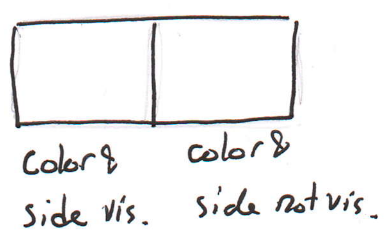
If $p = \frac{1}{2}$ we get $\dfrac{1}{1+\frac{1}{2}} = \dfrac{2}{2+1} = \dfrac{2}{3}$

Note that we are assuming that the person drawing the card does not generate a bias when choosing the side of the card to be displayed. For example, if Jennifer does not like red (because it reminds her of blood) and will always put the card black side up when she has a choice, then

$$p = p(R|C_2) = 0.$$

and $P(C_3|R) = 1.$

Assuming that the card is drawn and placed on table blindly, the probability is $P(C_3|R) = \frac{2}{3}.$ We can also see this by working with a restricted sample space. Imagine the sides of each card are labled by 1 and 2 respectively and let the Kafka protocol be

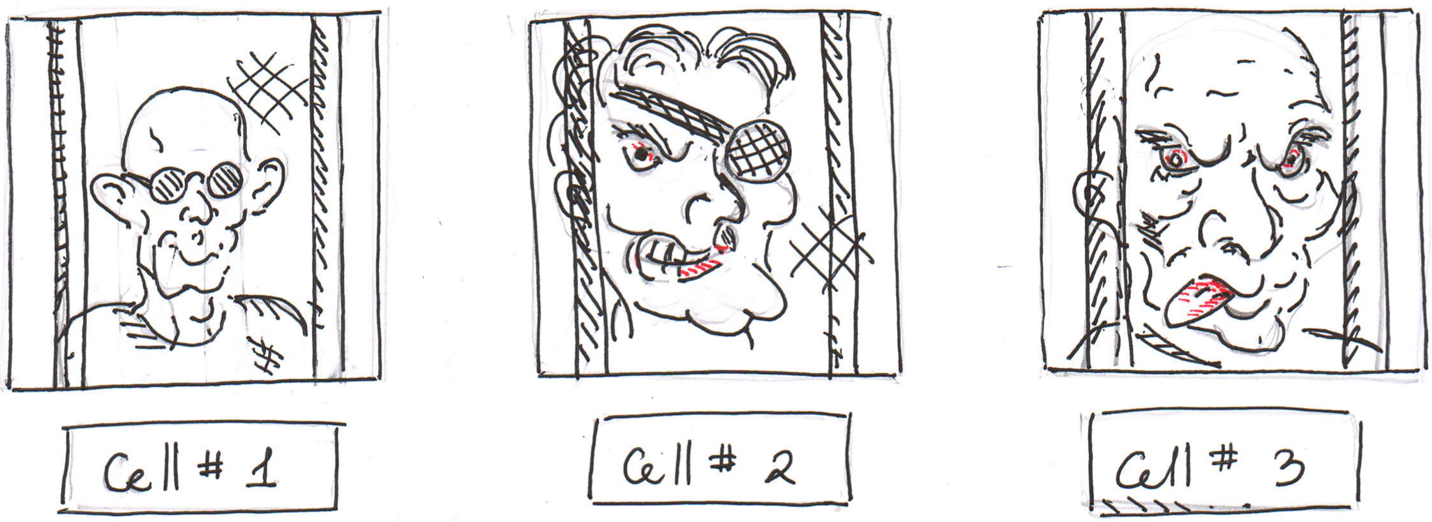


Then the event $R = \{(R_2, B_1), \overbrace{(R_1, R_2), (R_2, R_1)}^{C_3}\}$

if each outcome is assumed equally likely

$$P(C_3|R) = \frac{2}{3}.$$

Ex. Out of 3 prisoners, one was chosen to be executed at random. Prisoner in cell #1 asks the jailer to let him know one name from the list of the other 2 prisoners that will not be executed. Would such information affect the probability that prisoner in cell #1 is executed?



Solution: Let E_i - event prisoner i is executed.

$P(E_i) = \frac{1}{3}$. Let N-event jailor names prisoner 2 as the one pardoned.

$$P(E_1|N) = \frac{P(N|E_1)P(E_1)}{P(N|E_1)P(E_1) + P(N|E_2)P(E_2) + P(N|E_3)P(E_3)}$$

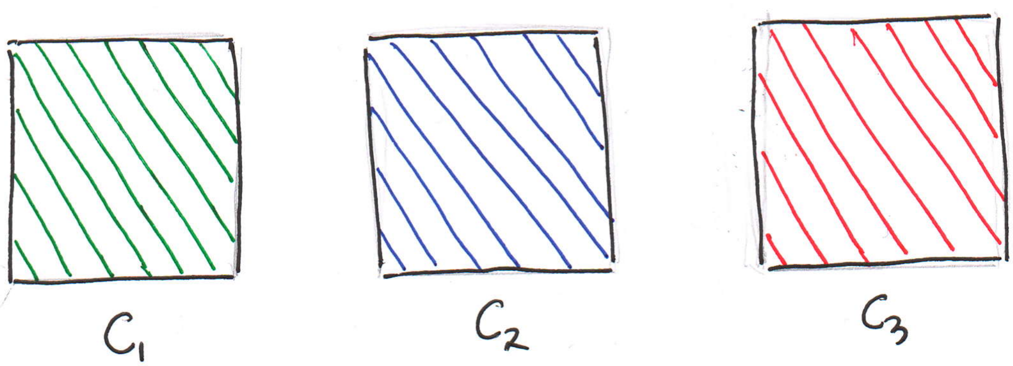
$$p(E_1|N) = \frac{p(N|E_1)}{p(N|E_1) + p(N|E_2) + p(N|E_3)}$$

$$= \frac{p}{p+0+1} = \frac{p}{p+1}$$

so if the jailer has no bias, $p = \frac{1}{2}$ and

$$p(E_1|N) = \frac{\frac{1}{2}}{\frac{1}{2}+1} = \frac{1}{3} = p(E_1).$$

Ex. Prize is behind one of 3 curtains. Contestant picks a curtain and then host of show opens one of the two remaining curtains where the prize isn't found. What should the contestant do?



Solution: Let W_s - event contestant wins by switching and C_k - event prize behind curtain k.

$$p(W_s) = p(W_s|C_1)p(C_1) + p(W_s|C_2)p(C_2) + $$
$$+ p(W_s|C_3)p(C_3) = 0\cdot\frac{1}{3} + 1\cdot\frac{1}{3} + 1\cdot\frac{1}{3}$$
$$= \frac{2}{3}$$

Thus the strategy of switching improves the odds of winning.

The following example was relayed to me through the book "The Grapes of Math" via a very helpful dwarf:

Ex. Write any two distinct numbers on two separate sheets of paper. These can be any two distinct real numbers (e.g. π, $\sqrt{2}$)
I will pick one of the sheets and look at the number. If I can guess which of the two numbers is biggest, I win. Otherwise you win. Does anyone have an advantage?

Solution: Believe it or not, but I have the advantage. Here is a strategy:

1) Generate a random number k (You can use the normal distribution or any other number generating distribution capable of generating any real number)

2) If the observed number is less than k, I claim that the observed number is the biggest of the two that you wrote.

If the observed number is bigger than k, I claim that the unobserved number is the biggest.

Let A, B, C be the events that k is smaller than both numbers, k is between the two numbers, and k is bigger than both numbers respectively.

Then $P(A) + P(B) + P(C) = 1$. Let W be the event that I win with my strategy.

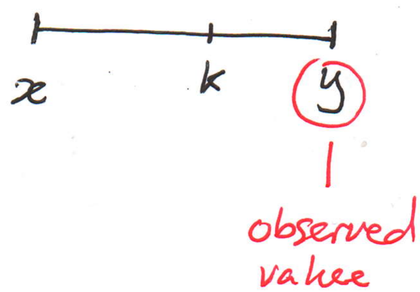
$$P(W) = P(W|A)P(A) + P(W|B)P(B) + P(W|C)P(C)$$

Observe that $P(w|A) = \frac{1}{2}$ because I stay with my choice and I have no way of knowing if the number I looked at is the biggest.

Similarly $P(w|C) = \frac{1}{2}$, because I switch to the unknown value blindly.

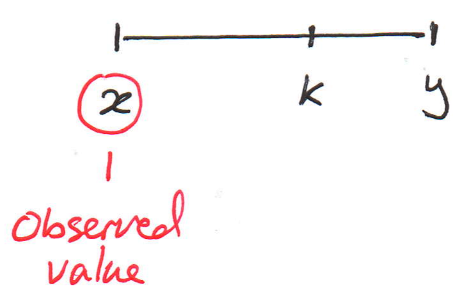
However $P(w|B) = 1$:

- If the generated number k is smaller than the observed value, then I am in this situation



staying with my original choice insures that I win.

- If the generated number k is bigger than the observed value, then I am in this situation



switching to the unobserved value insures that I win.

In particular

$$P(W) = \frac{1}{2}P(A) + P(B) + \frac{1}{2}P(C)$$

$$= \frac{1}{2}\big(P(A) + P(B) + P(C)\big) + \frac{1}{2}P(B)$$

$$= \frac{1}{2}\big(1 + P(B)\big) > \frac{1}{2}.$$