MATH 2030 3.00 – Elementary Probability Course Notes
Part II: Independence and Conditional Probabilities

Tom Salisbury – salt@yorku.ca

York University, Dept. of Mathematics and Statistics

Original version April 2010. Thanks are due to
E. Brettler, V. Michkine, and R. Shieh for many corrections

January 18, 2018

Independence and Conditional Probabilities

- Events A and B are said to be *independent* if

$$P(A \cap B) = P(A)P(B).$$

- If $P(B) \neq 0$ then the *conditional probability of A given B* is

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

Interpretation: If we repeat our experiment many times, one sees that $P(A \mid B)$ is the relative frequency with which A will occur, *out of just those repetitions in which B occurs*. So it is the likelihood that A will occur, if we're given the information that B will occur.

Likewise, if $P(B) \neq 0$ then independence $\Rightarrow P(A \mid B) = P(A)$. That is, knowing that B will occur doesn't affect the likelihood that A will occur. In other words, independence means that B occurring has no influence on A occuring.

Independence and Conditional Probabilities

Properties:

- Independent – $P(A \cap B) = P(A)P(B)$
  Mutually exclusive – $P(A \cup B) = P(A) + P(B)$.
  These are different; don't mix them up.
- $P(A \mid A) = 1$
- $P(A^c \mid A) = 0$
- $P(A^c \mid B) = 1 - P(A \mid B)$
  [pf: $P(B) = P(A \cap B) + P(A^c \cap B)$. Now divide by $P(B)$.]
- $P(B) = 0$ or $1 \Rightarrow B$ is independent of every event A.
- sampling with replacement $\Rightarrow$ independence.
  sampling without replacement $\Rightarrow$ dependence.

Probabilities $\longmapsto$ Conditional Prob's.

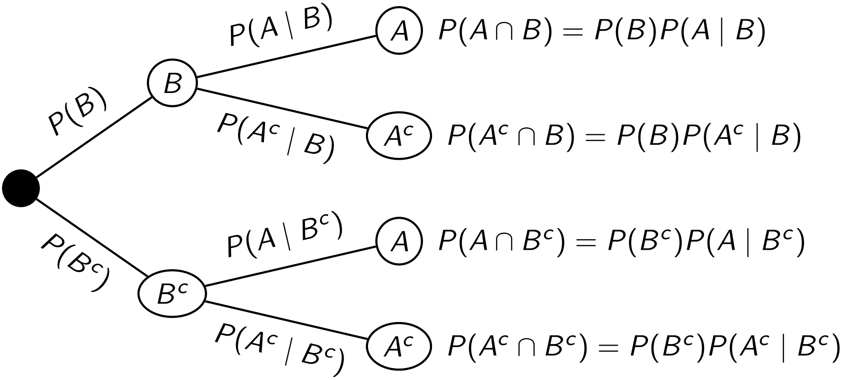
- **Eg:** Draw 2 cards in order, without replacement.
  $$P(\text{2nd is a } \heartsuit \mid \text{1st is a } \heartsuit) = \frac{P(\text{both are } \heartsuit)}{P(\text{1st is a } \heartsuit)} = \frac{\frac{13\times 12}{52\times 51}}{\frac{13\times 51}{52\times 51}} = \frac{12}{51}$$
  [which is what we'd have guessed anyway].
- **Roll 2 dice:** $P(\text{1st is a 3}) = \frac{1}{6} = P(\text{2nd is a 3})$, and $P(\text{both are 3's}) = \frac{1}{36} = \frac{1}{6} \times \frac{1}{6}$, so they are independent [as we'd have guessed]. Likewise
  $$P(\text{2nd is a 3} \mid \text{1st is a 3}) = \frac{P(\text{both are 3's})}{P(\text{2nd is a 3})} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}.$$
- **Eg:** Flip 4 coins. A is that all 4 flips agree.
  B is that flips 1 & 2 are H. C is that flip 1 is H.
  Then A and B are dependent, but A and C are independent.
  [Calculate: $P(A \mid C) = \frac{1}{8} = P(A) \neq P(A \mid B) = \frac{1}{4}$.
  The probability that the last 3 flips agree with the 1st doesn't depend on whether the 1st is H or T.]

Conditional Prob's $\longmapsto$ Probabilities.

- $P(A \cap B) = P(B)P(A \mid B)$
- $P(A \cap B \cap C) = P(C)P(B \mid C)P(A \mid B \cap C)$
- **Eg.** 2 urns. 1st has 3 red balls, 5 yellow balls. 2nd has 2 red and 3 yellow. Pick an urn at random and then a ball at random from that urn. What is the probability that it's red?

  A: ball is red. B: pick first urn. We know that $P(B) = P(B^c) = \frac{1}{2}$, $P(A \mid B) = \frac{3}{8}$, $P(A \mid B^c) = \frac{2}{5}$.
  So $P(A) = P(A \cap B) + P(A \cap B^c) = \frac{1}{2} \times \frac{3}{8} + \frac{1}{2} \times \frac{2}{5} = \frac{31}{80}$.

Tree Diagrams

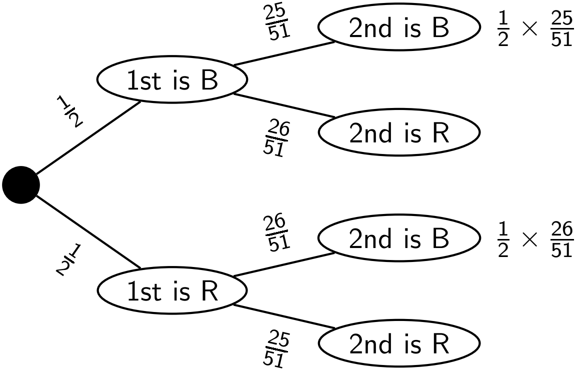


The probability for a node of the tree is the product of the conditional probabilities along the branches leading to that node.

Tree Diagrams

Eg: Draw 2 cards without replacement. $P(\text{2nd is black}) = ?$

1st method: By symmetry, the answer must $= \frac{1}{2}$.
2nd method: Go back to a model and count.
3rd method: Conditional probabilities:

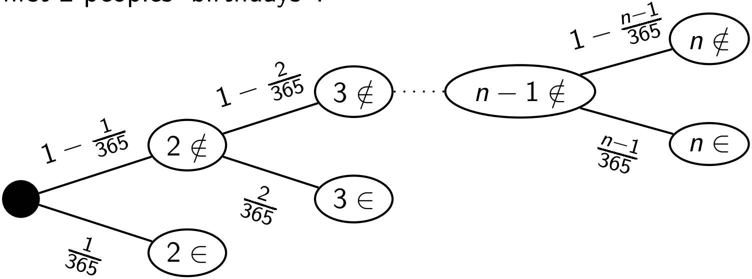


Adding the probabilities for these 2 nodes gives $\frac{1}{2}(\frac{25}{51} + \frac{26}{51}) = \frac{1}{2}$.

Birthday problem

Eg: n people. $P(\exists\ 2$ people with the same birthday$) = ?$

Put the people in some order and write eg. "$3 \notin$" as an abbreviation for "the 3rd person's birthday is different from the first 2 peoples' birthdays".



Then the probability of having shared birthdays is $1 - (1 - \frac{1}{365})(1 - \frac{2}{365}) \cdots (1 - \frac{n-1}{365})$. If $n = 30$ this is ≈ 0.7063, while if $n = 65$ it is ≈ 0.9977, and for $n = 80$ is ≈ 0.9999

Independence for 2 events

If A and B are indep., then the following are also indep.:

- A^c and B
  [pf: $P(A^c \cap B) = P(B) - P(A \cap B)$
  $= P(B) - P(A)P(B) = (1 - P(A))P(B) = P(A^c)P(B)$]
- A and B^c [likewise]
- A^c and B^c [likewise]

We'll see that this is the right way to generalize the notion of independence to 3 (or more) events.

Independence of 3 events

Three events A, B, C are said to be *independent* if
$P(A_1 \cap B_1 \cap C_1) = P(A_1)P(B_1)P(C_1)$
for $A_1 = A$ or A^c, $B_1 = B$ or B^c, and $C_1 = C$ or C^c.
In other words, if the following 8 conditions hold:

- $P(A \cap B \cap C) = P(A)P(B)P(C)$
- $P(A \cap B \cap C^c) = P(A)P(B)P(C^c)$
- $P(A \cap B^c \cap C) = P(A)P(B^c)P(C)$
- $P(A^c \cap B \cap C) = P(A^c)P(B)P(C)$
- $P(A \cap B^c \cap C^c) = P(A)P(B^c)P(C^c)$
- $P(A^c \cap B \cap C^c) = P(A^c)P(B)P(C^c)$
- $P(A^c \cap B^c \cap C) = P(A^c)P(B^c)P(C)$
- $P(A^c \cap B^c \cap C^c) = P(A^c)P(B^c)P(C^c)$

Independence of 3 events

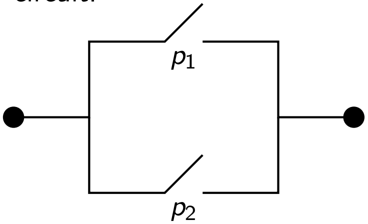
Consequences:

- A, B, C indep. $\Rightarrow$ pairwise independence.
  [eg. A and B are independent, because
  $P(A \cap B) = P(A \cap B \cap C) + P(A \cap B \cap C^c)$
  $= P(A)P(B)P(C) + P(A)P(B)P(C^c) = P(A)P(B)$]
- But A, B, C pairwise indep. $\not\Rightarrow$ A, B, C indep.
  [You'll work out a counterexample on an assignment]
- A, B, C indep. $\Rightarrow$ A indep. of any event got from B, C.
  [eg. $P(A \mid B \cap C) = P(A)$]
- Therefore all conditional probabilities in the tree diagram for A, B, C equal the corresponding probabilities.
- A, B, C indep. $\Leftrightarrow$
  A, B, C pairwise indep. and $P(A \cap B \cap C) = P(A)P(B)P(C)$.

Switches

Eg: 2 switches in parallel.

Switch 1 is closed (ie current flows through it), with probability p_1.
Switch 2 is closed, with probability p_2.
Assume the switches are open/closed independently of each other.
Problem: Find the probability that current flows through the circuit.

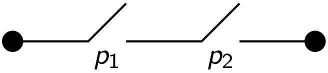


Solution: Write using $\cap$'s:

Let A_i be the event that switch i is closed. P(current flows)
$= P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = p_1 + p_2 - p_1 p_2$.
Alternatively, $P(A_1 \cup A_2) = 1 - (P((A_1 \cup A_2)^c)$
$= 1 - P(A_1^c \cap A_2^c) = 1 - (1 - p_1)(1 - p_2) = p_1 + p_2 - p_1 p_2$.

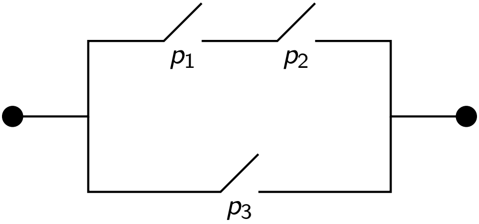
Switches

Eg: 2 switches in series.



A_i: event that switch i is closed.
$P(\text{current flows}) = P(A_1 \cap A_2) = p_1 p_2$

Eg: 3 switches.



3 switch example, cont'd

$A =$ event that current flows.
$A_i =$ event that switch i is closed.
$A_{12} = A_! \cap A_2$ is the event that current flows along the top.

- Inclusion/Exclusion: $P(A) = P(A_{12} \cup A_3)$
  $= P(A_{12}) + P(A_3) - P(A_{12} \cap A_3) = p_1 p_2 + p_3 - p_1 p_2 p_3$
- Complements: $P(A) = 1 - P(A^c) = 1 - P(A_{12}^c \cap A_3^c)$
  $= 1 - P(A_{12}^c)P(A_3^c) = 1 - (1 - p_1 p_2)(1 - p_3)$.
- Enumeration of alternatives:
  $A = A_3 \cup (A_{12} \cap A_3^c)$ and the latter are disjoint. So
  $P(A) = P(A_3) + P(A_{12} \cap A_3^c) = p_3 + p_1 p_2 (1 - p_3)$.

With each approach, the point is to write things in terms of intersections, so that independence applies.

Bayes rule: 2 alternatives

- Problem: Compute $P(B \mid A)$, knowing $P(B), P(A \mid B), P(A \mid B^c)$.
- Bayes rule: $\boxed{P(B \mid A) = \frac{P(B)P(A \mid B)}{P(B)P(A \mid B) + P(B^c)P(A \mid B^c)}}$
- Proof: $= \frac{P(A \cap B)}{P(A)} = \frac{P(A \cap B)}{P(A \cap B) + P(A \cap B^c)}$.
- **Eg:** 2 urns. Urn 1 has 3 Red & 5 Green balls. Urn 2 has 2 Red & 3 Green. Pick an urn at random and then a ball. If it's red, what are the chances we had picked the 1st urn? They will no longer be $\frac{1}{2}$, since the evidence favours urn 2, which has a higher % of reds.

  A: get a red ball. B: pick 1st urn. $P(B) = \frac{1}{2}$, $P(B^c) = \frac{1}{2}$, $P(A \mid B) = \frac{3}{8}$, $P(A \mid B^c) = \frac{2}{5}$.

  So $P(B \mid A) = \frac{\frac{1}{2} \times \frac{3}{8}}{\frac{1}{2} \times \frac{3}{8} + \frac{1}{2} \times \frac{2}{5}} = \frac{15}{31} < \frac{1}{2}$

Bayes rule: multiple alternatives

- Let $B_1, \ldots, B_n$ partition Ω. ie the B's are nonempty and disjoint, with $\cup B_i = \Omega$.
- Bayes rule:

$$\boxed{P(B_1 \mid A) = \frac{P(B_1)P(A \mid B_1)}{P(B_1)P(A \mid B_1) + \cdots + P(B_n)P(A \mid B_n)}}$$

- Proof: same as before.
  2-alternative case is $B_1 = B$, $B_2 = B^c$.
- **Eg: Medical screening test**
  A medical condition affects 1 person in 1,000. A test is 98% effective on healthy people and 99% effective on infected ones (ie it gives the "correct" answer that % of the time). If you test positive, what's the likelihood you have the condition?

Medical test, cont'd

- Define events A – test positive. B – are ill. B^c – are healthy.
- Problem asks for $P(B \mid A)$
- Information given is that $P(B) = 0.001$, $P(A^c \mid B^c) = 0.98$, $P(A \mid B) = 0.99$
- Therefore we compute $P(B^c) = 0.999$ and $P(A \mid B^c) = 0.02$ and apply Bayes.
- $P(B \mid A) = \frac{0.001 \times 0.99}{0.001 \times 0.99 + 0.999 \times 0.02} \approx 0.0472$ – ie. rather small.
- Point is that Bayes $\Rightarrow$ false positives can swamp true ones, if the disease is rare. That is, such a screening test is only useful as a trigger for further tests. Doctors need to be able to explain this to patients.

Bayes rule cont'd

Other examples:

- Legal use of DNA testing:
  A – DNA match; B – guilty; B^c – innocent.
  Judge wants $P(B \mid A)$. Have same false positive issue as before: when a big database is tested, a match means much less than when an actual suspect is tested.
- Bayesian statistics:
  Suppose there is good evidence for a "prior probability" $P(B_k)$ for each alternative k. One then gathers data (ie observes an event A) and one revises the prior to get "posterior probabilities" $P(B_k \mid A)$. That is, likelihoods for the alternatives k, given the data.