

# Assignment 1

---

Per Magnus Veierland  
permve@stud.ntnu.no

February 18, 2016

## 1 COUNTING AND BASIC LAWS OF PROBABILITY

### 1.1 5-CARD POKER HANDS

*Consider the domain of dealing 5-card poker hands from a standard deck of 52 cards, under the assumption that the dealer is fair.*

- a) *How many atomic events are there in the joint probability distribution (i.e., how many 5-card hands are there)?*

The number of 5-card hands in poker can be found with combinatorics by noting that the ordering of the cards does not matter:

$$\binom{52}{5} = \frac{52!}{5!(52-5)!} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \frac{311875200}{120} = 2598960 \quad (1)$$

- b) *What is the probability of each atomic event?*

The probability of each atomic event, i.e. the probability of any distinct poker hand, is:

$$\frac{1}{2598960} \approx 3.84769 \cdot 10^{-7} \quad (2)$$

- c) • *What is the probability of being dealt a royal straight flush?*

A royal straight flush consists of five cards of the same suit in a consecutive sequence, where the top card is an ace. There are four possible royal straight flush hands; one for each suit. The probability of being dealt a royal straight flush is:

$$\frac{4}{2598960} = \frac{1}{649740} \approx 1.53908 \cdot 10^{-6} \quad (3)$$

- *What is the probability of being dealt four of a kind?*

Four of a kind is a poker hand where four of the cards have the same rank, and the fifth card can be any other rank and suit. The number of possible four of a kind hands is the number of ranks; 13, multiplied by the number of possible fifth cards; which is any of the twelve remaining ranks from one of four suits:

$$\frac{13 \cdot 12 \cdot 4}{2598960} = \frac{1}{4165} \approx 2.40096 \cdot 10^{-4} \quad (4)$$

## 1.2 TWO CARDS IN A DECK

*Two cards are randomly selected from a deck of 52 playing cards.*

- a) *What is the probability they constitute a pair (that is, that they are the same denomination)?*

The number of pairs in a deck of 52 playing cards is given by the number of ranks times the number of ways to form a pair from four suits within a rank:

$$13 \cdot \binom{4}{2} = 13 \cdot \frac{4!}{2!(4-2)!} = 13 \cdot \frac{4 \cdot 3}{2} = 13 \cdot \frac{12}{2} = 78 \quad (5)$$

The number of ways to draw two cards from a deck of 52 playing cards is given by:

$$\binom{52}{2} = \frac{52!}{2!(52-2)!} = \frac{52 \cdot 51}{2} = 1326 \quad (6)$$

The probability that drawing two cards from a deck of 52 playing cards yields a pair is given by the number of pairs in a deck, divided by the number of ways to draw two cards from a deck:

$$\frac{78}{1326} = \frac{1}{17} \approx 0.05882 \quad (7)$$

The solution can also be found by reasoning that the probability of the second card drawn will be of the same rank as the first card drawn is given by:

$$\frac{4-1}{52-1} = \frac{3}{51} = \frac{1}{17} \approx 0.05882 \quad (8)$$

- b) *What is the conditional probability they constitute a pair given that they are of different suits?*

The number of ways to draw two cards of different suits from a deck is given by the number of ways to draw two cards from a deck, minus the number of suits times the number of ways to draw two cards from the same suit:

$$\binom{52}{2} - 4 \cdot \binom{13}{2} = \frac{52!}{2!(52-2)!} - 4 \cdot \frac{13!}{2!(13-2)!} = \frac{52 \cdot 51}{2} - 4 \cdot \frac{13 \cdot 12}{2} = 1326 - 312 = 1014 \quad (9)$$

The number of possible pairs is given by equation 5. The number of ways to draw a suit from a deck, given that the cards drawn will be from different suits, is given by:

$$\frac{78}{1014} = \frac{1}{13} \approx 0.07692 \quad (10)$$

The solution can also be found by reasoning that the probability of the second card drawn will be the same rank as the first card drawn, given that the second card drawn will be from a different suit, is given by:

$$\frac{4-1}{52-13} = \frac{3}{39} = \frac{1}{13} \approx 0.07692 \quad (11)$$

### 1.3 CONDITIONAL PROBABILITY

*If the occurrence of B makes A more likely, does the occurrence of A make B more likely? Why?*

We begin with the knowledge that the occurrence of B makes A more likely:

$$P(A|B) > P(A) \quad (12)$$

The **product rule** states that:

$$P(A \wedge B) = P(A|B) \cdot P(B) \quad \text{and} \quad P(A \wedge B) = P(B|A) \cdot P(A) \quad (13)$$

Equating the two right-hand sides and dividing by  $P(B)$  gives us **Bayes' rule**:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (14)$$

Replacing the left-hand side of the known inequality from equation 12 with the right-hand side of equation 14 gives us the inequality:

$$\frac{P(B|A) \cdot P(A)}{P(B)} > P(A) \quad (15)$$

Multiplying both sides by  $P(B)$ :

$$P(B|A) \cdot P(A) > P(A) \cdot P(B) \quad (16)$$

Dividing both sides by  $P(A)$ :

$$P(B|A) > P(B) \quad (17)$$

This derivation shows that if it is known that the probability of A given B is greater than the prior probability of A (equation 12), it is also true that the probability of B given A is greater than the prior probability of B (equation 17).

Intuitively this makes sense, since knowledge that a cause is more probable given the presence of a symptom would also make the probability of the symptom more likely given the presence of the cause.

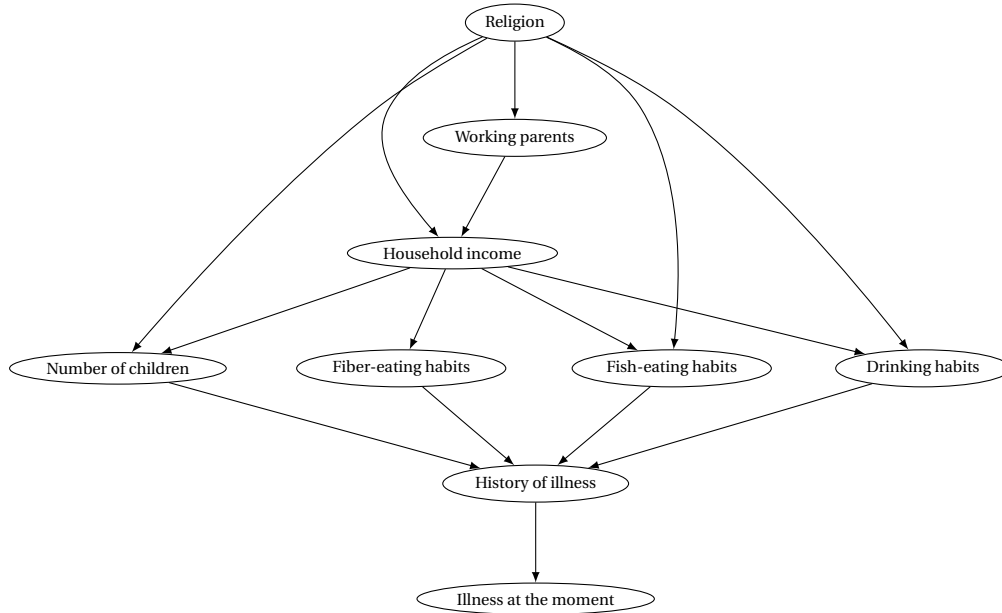


Figure 1: Bayesian network showing causal links between random variables in problem II.

## 2 BAYESIAN NETWORK CONSTRUCTION

Figure 2 shows a Bayesian network for the random variables 1) *Illness at the moment*, 2) *History of illness*, 3) *Number of children*, 4) *Working parents*, 5) *Religion*, 6) *Household income*, 7) *Fish-eating habits*, 8) *Fiber-eating habits*, and 9) *Drinking habits*.

Modeling the network based on causality leads to specifying fewer conditional probability values, as well as the values being easier to find since they are more intuitive.

The conditional independence properties of the network is as follows:

1. *Illness at the moment* is conditionally independent of all other random variables, given its parent *History of illness*.
2. *History of illness* is conditionally independent of *Household income*, *Working parents*, and *Religion* – given *Number of children*, *Fiber-eating habits*, *Fish-eating habits*, and *Drinking habits*.
3. *Number of children* is conditionally independent of *Working parents*, *Fiber-eating habits*, *Fish-eating habits*, and *Drinking habits* – given its parents *Religion* and *Household income*. Two causal relationships which are poorly modeled is the relationship with *History of illness* and the relationship with *Working parents*. The model shows that *Number of children* causally affects *History of illness*, but it does not show that *History of illness* may causally affect *Number of children* as well. The model shows that *Working parents* indirectly causally affects *Number of children* through *Household income*, but it does not show that *Number of children* may causally affect *Working parents*.

4. *Working parents* is conditionally independent of *Number of children*, *Fiber-eating habits*, *Fish-eating habits*, *Drinking habits*, *History of illness*, and *Illness at the moment* – given its parent *Religion* and its child *Household income*. Not causally modeled is how *Illness at the moment* may affect *Working parents*.
5. *Religion* is conditionally independent of *Fiber-eating habits*, *History of illness*, and *Illness at the moment* – given its children *Number of children*, *Household income*, *Working parents*, *Fish-eating habits*, and *Drinking habits*. Not causally modeled is how *History of illness* may affect *Religion*, i.e. not “mental health issues” – but how religion may serve as a coping mechanism for serious illnesses.
6. *Household income* is conditionally independent of *History of illness* and *Illness at the moment* – given its parents *Religion* and *Working parents*, and its children *Number of children*, *Fiber-eating habits*, *Fish-eating habits*, and *Drinking habits*.
7. *Fish-eating habits* is conditionally independent of *Working parents* and *Illness at the moment* – given its parents *Religion* and *Household income*, and its child *History of illness*, and the parents of its child; *Number of children*, *Fiber-eating habits* and *Drinking habits*.
8. *Fiber-eating habits* is conditionally independent of *Religion*, *Working parents*, and *Illness at the moment* – given its parent *Household income*, its children *History of illness*, and the parents of its child; *Number of children*, *Fish-eating habits*, and *Drinking habits*.
9. *Drinking habits* is conditionally independent of *Working parents* and *Illness at the moment* – given its parents *Religion* and *Household income*, its children *History of illness*, and the parents of its child; *Number of children*, *Fiber-eating habits* and *Fish-eating habits*.

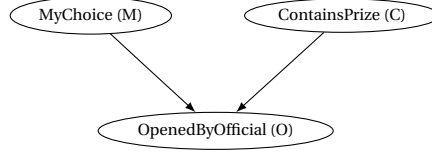


Figure 2: Bayesian network modeling problem III.

$C$	$M$	$P(O = a C, M)$	$P(O = b C, M)$	$P(O = c C, M)$
a	a	0	1/2	1/2
a	b	0	0	1
a	c	0	1	0
b	a	0	0	1
b	b	1/2	0	1/2
b	c	1	0	0
c	a	0	1	0
c	b	1	0	0
c	c	1/2	1/2	0

Table 1: Conditional Probability Table (CPT) for  $\mathbf{P}(O|C, M)$ .

### 3 BAYESIAN NETWORK APPLICATION

The described scenario is shown modeled in Figure 2 as a Bayesian network with three random variables; *ContainsPrize* (C), *MyChoice* (M), and *OpenedByOfficial* (O). In the game, the action made by the official is a direct consequence of the player's chosen door and the prize door. This makes it intuitive to model the network with *OpenedByOfficial* being caused by *MyChoice* and *ContainsPrize*.

The prior probability of *MyChoice* is  $\mathbf{P}(M) = \langle 1/3, 1/3, 1/3 \rangle$ , and the prior probability of *ContainsPrize* is  $\mathbf{P}(C) = \langle 1/3, 1/3, 1/3 \rangle$ . Table 1 shows the CPT for the random variable *OpenedByOfficial*.

Shown in Equation 18 is the scenario where the player chooses door *a* and the official chooses door *b*. This results in  $\mathbf{P}(C|M = a \wedge O = b) = \langle 1/3, 0, 2/3 \rangle$ . Interpreting this result shows that there is a  $1/3$  chance that the originally chosen door is the prize door, that there is zero chance that the door chosen by the official is the prize door, and that there is a  $2/3$  chance that door *c* is the prize door.

$$\begin{aligned}
\mathbf{P}(C|M = a \wedge O = b) &= \alpha \cdot \mathbf{P}(C) \cdot \mathbf{P}(M = a) \cdot \mathbf{P}(O = b|M = a, C) \\
&= \alpha \cdot \langle \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \rangle \cdot \frac{1}{3} \cdot \langle \frac{1}{2}, 0, 1 \rangle \\
&= \alpha \cdot \langle \frac{1}{18}, 0, \frac{1}{9} \rangle \\
&= \langle \frac{1}{3}, 0, \frac{2}{3} \rangle
\end{aligned} \tag{18}$$

Equation 19 shows the symmetric scenario where the player chooses door  $a$  and the official chooses door  $c$ . This results in  $\mathbf{P}(C|M = a \wedge O = c) = \langle 1/3, 2/3, 0 \rangle$ . Interpreting this result shows that there is a  $1/3$  chance that the originally chosen door is the prize door, that there is a  $2/3$  chance that door  $b$  is the prize door, and that there zero chance that the door chosen by the official is the prize door.

$$\begin{aligned}
 \mathbf{P}(C|M = a \wedge O = c) &= \alpha \cdot \mathbf{P}(C) \cdot P(M = a) \cdot \mathbf{P}(O = c|M = a, C) \\
 &= \alpha \cdot \langle \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \rangle \cdot \frac{1}{3} \cdot \langle \frac{1}{2}, 1, 0 \rangle \\
 &= \alpha \cdot \langle \frac{1}{18}, \frac{1}{9}, 0 \rangle \\
 &= \langle \frac{1}{3}, \frac{2}{3}, 0 \rangle
 \end{aligned} \tag{19}$$

Based on Equation 18 and Equation 19 it can be concluded that given the knowledge of which door the player picked, and which door the official opened; there is a posterior probability of  $1/3$  that the door picked by the player is the prize door, and a posterior probability of  $2/3$  that the door picked by the player is not the prize door. This means that the best strategy for the player will always be to change doors after the official opens one door; since this will double his chances of winning the prize!