Homework 1 — answer sheet

1. Prove the following theorems of probability from the basic axioms presented in the lecture. For the sake of simplicity, assume that all variables are binary variables.

   - Prove the **law of negation**: for all A,
   $$P(\neg A) = 1 - P(A).$$
   - Prove the **law of total probability**: for all A and B,
   $$P(A) = P(A \mid B)P(B) + P(A \mid \neg B)P(\neg B).$$
   - Prove the **law of disjunction**: for all A and B,
   $$P(A \vee B) = P(A) + P(B) - P(A \wedge B).$$
   [*Hint*: it may be useful to use the fact that '$A \vee B$' $= (A \wedge \neg B) \vee (B \wedge \neg A) \vee (A \wedge B)$, and then to also show that $P(A \wedge \neg B) = P(A) - P(A \wedge B)$.]

   Answer.

   - *Law of negation.* Note that '$A \vee \neg A$' is a sure proposition, so by Rule 2 we have $P(A \vee \neg A) = 1$. By Rule 3 we also have $P(A \vee \neg A) = P(A) + P(\neg A)$, since A and $\neg A$ are mutually exclusive. Therefore we have $P(A) + P(\neg A) = 1$, which we re-arrange as
   $$P(\neg A) = 1 - P(A). \qquad \blacksquare$$
   - *Law of disjunction.* We will first show that $P(A \wedge \neg B) = P(A) - P(A \wedge B)$, which will be useful in the proof. Consider that $A = (A \wedge \neg B) \vee (A \wedge B)$. Since $A \wedge \neg B$ and $A \wedge B$ are mutually exclusive, by Rule 3 we have $P(A) = P(A \wedge \neg B) + P(A \wedge B)$, which we re-write as
   $$P(A \wedge \neg B) = P(A) - P(A \wedge B).$$
   Now for $P(A \vee B)$. First observe that '$A \vee B$' $= (A \wedge \neg B) \vee (B \wedge \neg A) \vee (A \wedge B)$. Since these are mutually exclusive, by Rule 3 we have:
   $$P(A \vee B) = P(A \wedge \neg B) + P(B \wedge \neg A) + P(A \wedge B).$$
   Using the fact we proved at the start, $P(A \wedge \neg B)$ can be written as $P(A) - P(A \wedge B)$, and $P(B \wedge \neg A)$ as $P(B) - P(A \wedge B)$. So we have
   $$P(A \vee B) = \big[P(A) - P(A \wedge B)\big] + \big[P(B) - P(A \wedge B)\big] + P(A \wedge B) = P(A) + P(B) - P(A \wedge B). \qquad \blacksquare$$

- *Law of total probability.* Note first that $A = (A \wedge B) \vee (A \wedge \neg B)$. Since $A \wedge B$ and $A \wedge \neg B$ are mutually exclusive, by Rule 3 we have $P(A) = P(A \wedge B) + P(A \wedge \neg B)$; this is the law of marginalization that we saw in lectures. By the definition of conditional probability we then obtain

$$P(A) = P(A \mid B)P(B) + P(A \mid \neg B)P(\neg B). \qquad \blacksquare$$

2. Suppose we have an SCM with exogenous variables A, B, U_C, U_{ABC}, and endogenous variable C, with structural equation:

$$C := (A \wedge B \wedge U_{ABC}) \vee U_C$$

Write the Causal Bayes Net that describes the system, assuming that U_{ABC} and U_C are not observable (so the only variables in the model are A, B, and C — you will still need to use $P(U_C)$ and $P(U_{ABC})$ to fill the conditional probability table).

Answer. We have only one endogenous variable, so we need only one conditional probability table:

$$P(C \mid \neg A, B) = P(C \mid A, \neg B) = P(C \mid \neg A, \neg B) = P(U_C)$$

$$P(C \mid A, B) = P(U_C \vee U_{ABC}) = P(U_C) + P(U_{ABC}) - P(U_C)P(U_{ABC})$$

We also need the probability of the exogenous variables $P(A)$ and $P(B)$.

3. Suppose we have the following causal model:

$$C := f(A, B)$$

$$D := f(C)$$

$$E := f(D, A)$$

where f is some function (it doesn't matter which). Write the factorization of the joint distribution $P(A, B, C, D, E)$ defined by the causal model.

Answer.

$$P(A, B, C, D, E) = P(E \mid D, A)\, P(D \mid C)\, P(C \mid A, B)\, P(A)\, P(B)$$

4. Create a causal model where $P(A \mid \text{Do}(B)) \neq P(A \mid B)$, but where B is not a descendant of A. (B is a descendant of A if there is a series of $\rightarrow$ in the DAG of the causal model that lead from A to B)

Answer. For example: $A \leftarrow C \rightarrow B$.

5. In lecture we saw the **law of marginalization**, which says that:

$$P(A) = P(A, B) + P(A, \neg B).$$

Prove that the law of marginalization can be extended to conditional probabilities such that:

$$P(A \mid C) = P(A, B \mid C) + P(A, \neg B \mid C).$$

[*Hint*: use the fact that $A \wedge C = (A \wedge C \wedge B) \vee (A \wedge C \wedge \neg B)$.]

Answer. We have $P(A \mid C) = P(A, C)/P(C)$.

We also have $A \wedge C = (A \wedge C \wedge B) \vee (A \wedge C \wedge \neg B)$. Because the two disjuncts are mutually exclusive, we can use Rule 2 to write

$$P(A, C) = P(A, B, C) + P(A, \neg B, C).$$

This means that

$$P(A \mid C) = \frac{P(A, B, C) + P(A, \neg B, C)}{P(C)}.$$

Re-arranging, we get:

$$P(A \mid C) = \frac{P(A, B, C)}{P(C)} + \frac{P(A, \neg B, C)}{P(C)} = P(A, B \mid C) + P(A, \neg B \mid C). \quad \blacksquare$$

6. Suppose you have a causal model with DAG $C \rightarrow A \leftarrow B$. In lecture we saw that

$$P(A \mid C) = P(A \mid B, C)P(B) + P(A \mid \neg B, C)P(\neg B).$$

Prove this fact. [*Hint*: use the result derived in the previous exercise, and the factorization induced by the causal model.]

Answer. From the result in the previous exercise we have:

$$P(A \mid C) = P(A, B \mid C) + P(A, \neg B \mid C).$$

We can re-write the first term on the RHS as:

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

Using the factorization induced by the causal model, we re-write the joint distribution $P(A, B, C)$:

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

We apply the same reasoning to the second term on the RHS, and we're done. $\blacksquare$