Homework 6

Problem 1. *Given a sequence* $(d_1, d_2, \dots, d_n)$ *of positive integers (where* $n \geq 1$*):*

(i) There exists a tree with score $(d_1, d_2, \dots, d_n)$*.*

(ii) $\sum_{i=1}^{n} d_i = 2n - 2$*.*

Prove that (i) and (ii) are equivalent.

Solution.

1. $(i) \Rightarrow (ii)$ is obvious.

2. To prove $(ii) \Rightarrow (i)$:

   By induction on the number n.

   For $n = 1, 2$ the implication holds trivially, so let $n > 2$. Suppose the implication holds for any $n-1$ long positive sequence $(d_1, d_2, \dots, d_{n-1})$ with $\sum_{i=1}^{n-1} d_i = 2(n-1) - 2$.

   For the induction step, consider an length n positive sequence $\ell = (d_1, d_2, \dots, d_n)$ with $\sum_{i=1}^{n} d_i = 2n - 2$:

   Since the sum of the d_i is smaller than $2n$, there exists an i with $d_i = 1$. w.l.o.g. we assume $d_1 = 1$. With a similar argument we can also conclude that there must exist some index j such that $d_j \geq 2$. We take $k = \min\{j \mid d_j \geq 2\}$.
   . Now the sequence $\ell = (d_1, d_2, \dots, d_k, \dots, d_n) = (1, d_2, \dots, d_k - 1 + 1, \dots, d_n)$, we can derive a new sequence $\ell' = (d_2, \dots, d_k - 1, \dots, d_n)$. Obviously ℓ' is a $n-1$ length sequence (all positive) with the summation to be $2n - 2 - 1 - 1 = 2(n-1) - 1$. Then according to the induction hypothesis, there exists a tree $\mathcal{T}'$ which corresponds to ℓ'.
   Then $\mathcal{T} = (V(\mathcal{T}') \cup \{v_1\}, E(\mathcal{T}') \cup \{v_1, v_k\})$ is the tree which witnesses the validity of the sequence ℓ.

BE CAREFUL: Why is the following 'proof' of the implication (ii)⇒(i) insufficient (or, rather, makes no sense)? We proceed by induction on n. The base case $n = 1$ is easy to check, so let us assume that the implication holds for some $n \geq 1$. We want to prove it for $n + 1$.If $D = (d_1, d_2, ..., d_n)$ is a sequence of positive integers with $\sum_{i=1}^{n} d_i = 2n - 2$, then we already know that there exists a tree T on n vertices with D as a score. Add another vertex v to T and connect it to any vertex of T by an edge, obtaining a tree T' on $n+1$ vertices. Let D' be the score of T'. We know that the number of vertices increased by 1, and the sum of degrees of vertices increased by 2 (the new vertex has degree 1 and the degree of one old vertex increased by 1). Hence the sequence D' satisfies condition (ii) and it is a score of a tree, namely of T'. This finishes the inductive step. □

Problem 2. *Find an example to verify the claim that '(pairwise) independence does not imply mutual independence'. Pls give a detailed proof.*

Solution. (by S. Bernstein)

Suppose X and Y are two independent tosses of a fair coin, where we designate 1 for heads and 0 for tails. Let the third random variable $Z = (X + Y) \bmod 2$.

Then jointly the triple $\langle X, Y, Z\rangle$ has the following probability distribution:

$$\langle X, Y, Z\rangle = \begin{cases} \langle 0,0,0\rangle & \text{with probability } 1/4 \\ \langle 0,1,1\rangle & \text{with probability } 1/4 \\ \langle 1,0,1\rangle & \text{with probability } 1/4 \\ \langle 1,1,0\rangle & \text{with probability } 1/4 \end{cases}$$

$i, j, k \in \{0, 1\}$.

It is easy to verify that $Pr(X = i) = Pr(Y = j) = Pr(Z = k) = 1/2$ and $Pr(X = i, Y = j) = Pr(X = i, Z = k) = Pr(Y = j, Z = k) = 1/4$. i.e., X, Y, Z are pairwise independent.

However, $Pr(X = i, Y = j, Z = k) \neq Pr(X = i) \cdot Pr(Y = j) \cdot Pr(Z = k)$. For example, the left side equals $1/4$ for $\langle x, y, z\rangle = \langle 0, 0, 0\rangle$ while the right side equals $1/8$.

In fact, any of $\langle X, Y, Z\rangle$ is completely determined by the first two components. That is as far from independence as random variables can get.

□

Problem 3. *Show that, if $E_1, E_2, \ldots, E_n$ are mutually independent, then so are $\overline{E_1}, \overline{E_2}, \ldots, \overline{E_n}$.*

Solution. (sketch) It will be enough to prove that for any $2 \leq k \leq n$, and $\{F_1, F_2, \ldots, F_k\} \subseteq \{E_1, E_2, \ldots, E_n\}$

$$Pr\Big(\bigcap_{i=1}^{k} \overline{F_i}\Big) = \prod_{i=1}^{k} Pr(\overline{F_i})$$

Let $Pr(F_i) = f_i$, then

$$Pr(\bigcap_{i=1}^{k} \overline{F_i}) = 1 - Pr(\bigcup_{i=1}^{k} F_i) = 1 - \sum_{i=1}^{k} f_i + \sum_{1 \le i < j \le k} f_i f_j - \sum_{1 \le i < j < l \le k} f_i f_j f_l + \cdots$$

The right hand side of the above equation is $(1 - f_1)(1 - f_2)\cdots(1 - f_k) = \prod_{i=1}^{k} Pr(\overline{F_i})$. □

Problem 4. *The problem on the 37^st page of slide on* 'Probability: a quick review' *(i.e.,* the more complicated example*). (What is $Pr(U|W)$?)*

Solution.

$$\begin{aligned} Pr(U|W) = \frac{Pr(U \cap W)}{Pr(W)} &= \frac{Pr(U \cap W)}{Pr(R) \cdot Pr(W|R) + Pr(\neg R) \cdot Pr(W|\neg R)} && //\text{Law of total probability} \\ &= \frac{\mathrm{Pr}(R) \cdot Pr(U \cap W|R) + \mathrm{Pr}(\neg R) \cdot Pr(U \cap W|\neg R)}{Pr(R) \cdot Pr(W|R) + Pr(\neg R) \cdot Pr(W|\neg R)} && //\text{Law of total probability} \end{aligned}$$

□

Problem 5. *Suppose X and Y are two independent random variables, show that*

$$E(X \cdot Y) = E[X] \cdot E[Y]$$

.

Solution.

$$\begin{aligned} E[X \cdot Y] &= \Sigma_i \Sigma_j (i \cdot j) \cdot Pr((X = i) \cap (Y = j)) \\ &= \Sigma_i \Sigma_j (i \cdot j) \cdot Pr(X = i) \cdot Pr(Y = j)) \\ &= (\Sigma_i i \cdot Pr(X = i)) \left(\Sigma_j j \cdot Pr(Y = j)\right) \\ &= E(X) \cdot E(Y) \end{aligned}$$

□

Problem 6. *A monkey types on a 26 -letter keyboard that has lowercase letters only. Each letter is chosen independently and uniformly at random from the alphabet. If the monkey types 1,000,000 letters. what is the expected number of times the sequence "proof" appears?*

Solution. By the linearity of expectation:

$$E[X] = (1/26)^5 \times (1000000 - 4)$$

□