# Homework 1

1. (book #1.3) We shuffle a standard deck of cards, obtaining a permutation that is uniform over all 52! possible permutations. Find the probability of the following events. All events refer to inspecting the cards *without replacement*.

**1.** $A =$ "the first 2 cards include at least one ace":

$p(A) =$

**2** $B =$ "the first 5 cards include at least one ace:"

$p(B) =$

**3.** $C =$ "the first 2 cards are a pair of the same rank:"

$p(C) =$

**4.** $D =$ "the first 5 cards are all diamonds."

$p(D) =$

**5.** $E =$ "the first 5 cards form a full house (3 of one rank + 2 of another rank)."

$p(E) =$

6. (book #1.6) Consider the following balls-and-bin game. We start with one black ball and one white ball in a bin. We repeatedly do the following: choose one ball from the bin uniformly at random, and then put the ball back in the bin with another ball of the same color. We repeat until there are $n$ balls in the bin. Let $X$ = number of white balls, and let $p_X(j)$ denote the probability function, $j = 1, \ldots, n-1$. Find $p_X(j)$.

   (a) $\frac{2j}{(n-1)n}$ (b) $\frac{2(n-j)}{(n-1)n}$ (c) $\frac{1}{n-1}$ (d) $\frac{6j^2}{(n-1)n(2n-1)}$ (e) none of these

7. (book #1.10) I have a fair coin and a two-headed coin. I choose one of the two coins randomly with equal probability and flip it. Given that the flip was heads, what is the probability $p$ that I flipped the two-headed coin?

$p =$

8. (book #1.13) A medical company touts its new test for a certain genetic disorder. The false negative rate is small: if you have the disorder, the probability that the test returns a positive result is 0.999. The false positive rate is also small: if you do not have the disorder, the probability that the test returns a positive result is only 0.005. Assume that 2% of the population has the disorder. If a person chosen uniformly from the population is tested and the result comes back positive, what is the probability $p$ that the person has the disorder?

$p =$

9. (book #1.15:) Suppose that we roll $n = 10$ standard six-sided dice. What is the probability $p$ that their sum will be divisible by 6, assuming that the rolls are independent? (Hint: Use the principle of deferred decisions, and consider the situation after rolling all but one of the dice.)

$p =$

10. (book # 1.19) Using a sample space $S = \{1, 2, 3, 4\}$ with $\Pr(i) = 1/4$ and $A = \{1, 2\}$ find examples of events $B, C$ and $D$ where $\Pr(A \mid B) < \Pr(A)$, $\Pr(A \mid C) = \Pr(A)$, and $\Pr(A \mid D) > \Pr(A)$.

Let $E = \{1, 2, 3\}$, $F = \{2, 3, 4\}$ and $G = \{2, 3\}$. Then valid examples for $B, C$ and $D$ are:
(a) $E, F$ and $G$ (b) $F, E$ and $G$ (c) $F, G$ and $E$ (d) $G, E$ and $F$ (e) none of these

11. (book 1.20) Show that, if $E_1, E_2, \ldots, E_k$ are mutually independent, then so are $\bar{E}_1, \bar{E}_2, \ldots, \bar{E}_k$.

Which of the following arguments proves the claim? ($A \perp B$ means $A, B$ are independent).

(a) $\prod_{i\in I} \Pr(\bar{E}_i) = \prod_I (1 - \Pr(E_i)) = 1 - \Pr\left(\bigcup_I E_i\right) = \Pr\left(\overline{\bigcup_I E_i}\right) = \Pr\left(\bigcap_I \bar{E}_i\right)$
(b) $E_i \perp E_j \implies \bar{E}_i \perp \bar{E}_j$ for $i, j \in I \implies$ claim
(c) $\Pr(\bar{E}_i) = (1 - \Pr(E_i)) \implies \prod_{i\in I} \Pr(\bar{E}_i) = \Pr(\bigcap_i E_i)$
(d) $\Pr(\bigcup_{i\in I} E_i) = \sum_{i\in I} \Pr(E_i)$ for disjoint events $\implies \prod_{i\in I} \Pr(\bar{E}_i) = \Pr\left(\bigcap_i \bar{E}_i\right)$
(e) none of these

12. (book #1.21) Using a sample space $S = \{1, 2, 3, 4\}$ with $\Pr(i) = 1/4$ and $A = \{1, 2\}$, give an example of events $B, C$ such that any pair of events $A, B$ and $C$ are independent but all three are not mutually independent.

Let $E = \{2, 3\}$, $F = \{3, 4\}$, $G = \{1, 3\}$ and $H = \{1, 4\}$. Then valid examples for $B$ and $C$ are:
(a) $E$ and $F$ (b) $E$ and $G$ (c) $F$ and $G$ (d) $E$ and $H$ (e) none of these

13. Recall the union bound from Lemma 1.2. Let $\bar{E}_i$ denote the complement of $E_i$. Show

$$\underbrace{\Pr\left(\bigcap_{i\geq 1} \bar{E}_i\right)}_{LHS} \geq \underbrace{1 - \sum_{i\geq 1} \Pr(E_i)}_{RHS}.$$

*Hint:* you may use the identity $\bar{A} \cap \bar{B} = \overline{A \cup B}$, and a similar result for $E_1, \ldots, E_k$.

Which of the following arguments proves the claim?

(a) $LHS = \prod(1 - \Pr(E_i)) \geq RHS$
(b) $LHS = \sum_{i\geq q} \Pr(E_i) - \sum_{i_1<i_2} \Pr(E_{i_1} \cap E_{i_2}) + \sum_{i_1<i_2<i_3} \cdots \geq RHS$
(c) $LHS = 1 - \Pr\left(\bigcup_{i\geq 1} E_i\right) \geq RHS$
(d) $LHS = 1 - \prod_{i\geq 1} \Pr(E_i) \geq RHS$
(e) none of these

14. Consider a complete graph $K_n$ with $n$ nodes. That is a graph with nodes 1 through $n$, and all possible $\binom{n}{2}$ edges, i.e., all pairs of nodes are connected with an edge. Let $C(n, m) = \binom{n}{m}$. Show that for any integer $k < n$ with

$$\binom{n}{k} 2^{-C(k,2)+1} < 1 \tag{1}$$

it is possible to color the edges of $K_n$ with two colors so that it has no single-color subgraph $K_k$ of size $k$ (i.e., any $k$ nodes $\subset \{1, \ldots, n\}$).

*Hint:* Consider a chance experiment consisting of randomly assigning one of two colors, say red or blue, to each edge. There are $2^{C(n,2)}$ possible color assignments, and there are $C(n,k)$ size $k$ subgraphs. We index them $i = 1, \ldots, C(n,k)$. Let $E_i$ = "$i$-th subgraph has a single color". Now show

$$\Pr\left(\bigcap_{i=1}^{C(n,k)} \bar{E}_i\right) > 0 \tag{2}$$

and conclude that therefore there is at least one graph $K_n$ with *no* size $k$ subgraph of single-color.

Let NSG refer to the conclusion "There is at least one graph with no single-color size $k$ subgraph". Which of the following arguments proves the claim?

(a) $\Pr(E_i) < 1 \implies \prod(1 - \Pr(E_i)) = \prod \Pr(\bar{E}_i) > 0 \implies NSG$

(b) $\Pr(E_i) = \binom{n}{2} 2^{-C(k,2)+1} < 1 \implies \Pr(\bar{E}_i) > 0 \implies NSG$

(c) $\Pr(E_i) = 2^{-C(k,2)+1}$ for coloring edge $2, \ldots, C(k,2)$ the same color as the 1st edge in the $i$-th subgraph. Then $\Pr(NSG) = \Pr\left(\bigcap_{i \geq 1} \bar{E}_i\right) \geq 1 - \sum \Pr(E_i) = 1 - \binom{n}{k} 2^{-C(k,2)+1} > 0$ by (1). Therefore there is at least one outcome, i.e., graph with NSG

(d) By Markov inequality $\Pr(C(n,k) > 1) \geq \frac{E(C(n,k))}{1} = \binom{n}{k} 2^{-C(k,2)+1} > 0$ by (1).

(e) none of these

*Hint:* Read Section 6.1 in the book Although Section 6.1. is not strictly needed. The problem uses only methods from Section 1. If you just follow the hint and instructions in the HW problem, that's enough.

15. **Simulation:** Refer to problem **4.** (= book #1.13), above. Let $A$ ="positive" and $B$ ="genetic disorder". Carry out the following simulation. The instructions below include possible R code fragments.

Nothing to turn in for this problem. Please just carry out the following instructions.

Generate $n = 10000$ hypothetical subjects, generating indicators $B_i \in \{0, 1\}$, $i = 1, \ldots, n$ for subject $i$ having the genetic disorder, for example using

```
n = 1000
p = 0.02
B = sample(c(0,1),n, prob=c(1-p,p),replace=T)
```

Next we simulate test results for the $n$ subjects, using, for example

```
sens = 0.999
spec = 1-0.005
u = runif(n)                       # n Unif(0,1) r.v.'s
A = ifelse(B==1,u<sens,u<1-spec)   # p(A|B=1)=sens, p(A|B=0)=1-spec
```

Then the (empirical) frequency of disease given positive is

```
nBA = sum(A & B)
nA = sum(A)
pBgivenA = nBA/nA
```

Now create a macro (or function) that takes as an argument $p = Pr(B)$, the prob of the disorder, and returns `pBgivenA`. For example,

```
f <- function(p=0.02){
   n = 10000
   B = sample(c(0,1),n, prob=c(1-p,p),replace=T)
...
   pBgivenA = nBA/nA
   return(pBgivenA)
}
```

Finally, create a grid `pgrid=seq(from=0.01,to=0.5,length=100)`, evaluate $\Pr(B \mid A)$ on the grid, and plot it. For example

```
pgrid <- seq(from=0.01,to=0.5,length=100)
y <- rep(0,100)
for(i in 1:100)
   y[i] <- f(pgrid[i])
plot(pgrid,y,type="l",bty="l",xlab="Pr(B)",ylab="Pr(B | A)")
sens = 0.999
abline(h=sens,lty=2,col=2)
```

Interpret the plot.