Problem 1

Let X be a random variable with CDF F, and $Y = \mu + \sigma X$, where μ and σ are real numbers with $\sigma > 0$. Find the CDF of Y, in terms of F.

Solution: Using the definition of the CDF, we have

$$F_Y(y) = \mathsf{P}(Y \le y) = \mathsf{P}(\mu + \sigma X \le y) = \mathsf{P}\left(X \le \frac{y-\mu}{\sigma}\right) = F\left(\frac{y-\mu}{\sigma}\right).$$

Problem 2

(a) Show that $p(n) = \left(\frac{1}{2}\right)^{n+1}$, for $n = 0, 1, 2, \dots$ is a valid PMF for a discrete random variable.

(b) Find the CDF of a random variable with PMF from (a).

Solution: (a) For p to be a valid PMF, we need that it is non-negative at all potential values, and that the sum of the probabilities on the support sum to one. The first part is evident from $\left(\frac{1}{2}\right)^{n+1} > 0$ for all $n = 1, 2, \dots$. For the second part, we note that the sum of the probabilities form a geometric series:

$$\sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^{n+1} = \frac{\frac{1}{2}}{1-\frac{1}{2}} = 1.$$

(b) Suppose that a random variable X has PMF p. Then the CDF of X is given by

$$\begin{aligned} F(x) &= \mathsf{P}(X \le x), \\ &= \sum_{n=0}^{x} \mathsf{P}(X = n), \\ &= \sum_{n=0}^{x} \left(\frac{1}{2}\right)^{n+1}, \\ &= \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^{n+1} - \sum_{n=x+1}^{\infty} \left(\frac{1}{2}\right)^{n+1}, \\ &= 1 - \frac{\left(\frac{1}{2}\right)^{x+2}}{1-\frac{1}{2}}, \\ &= 1 - \left(\frac{1}{2}\right)^{x+1}. \end{aligned}$$

Problem 3

Let X, Y and Z be discrete random variables such that X and Y have the same conditional distribution given Z, i.e., for all a and z we have

$$\mathsf{P}(X = a \mid Z = z) = \mathsf{P}(Y = a \mid Z = z).$$

Show that X and Y have the same distribution (unconditionally, not just when given Z).

Solution: We want to show that $\mathsf{P}(X = a) = \mathsf{P}(Y = a)$ for all a. By applying the law of total probability (LOTP) twice, we have

$$\begin{aligned} \mathsf{P}(X = a) &= \sum_z \mathsf{P}(X = a \mid Z = z) P(Z = z), \\ &= \sum_z \mathsf{P}(Y = a \mid Z = z) P(Z = z), \\ &= \mathsf{P}(Y = a). \end{aligned}$$

Problem 4

(a) Let $X \sim \mathsf{DUnif}(C)$, and B be a nonempty subset of C. Find the conditional distribution of X, given that X is in B.

(b) If $X \sim \mathsf{HGeom}(w, b, n)$, what is the distribution of $n - X$?

Solution: (a) It suffices to find the PMF of X conditional on $X \in B$. For any $b \notin B$,

$$\mathsf{P}(X = b \mid X \in B) = \frac{\mathsf{P}(X = b \text{ and } X \in B)}{\mathsf{P}(X \in B)} = \frac{0}{\mathsf{P}(X \in B)} = 0.$$

For any $b \in B$, we have

$$\mathsf{P}(X = b \mid X \in B) = \frac{\mathsf{P}(X = b \text{ and } X \in B)}{\mathsf{P}(X \in B)} = \frac{\mathsf{P}(X = b)}{\mathsf{P}(X \in B)} = \frac{\frac{1}{|C|}}{\frac{|B|}{|C|}} = \frac{1}{|B|}.$$

We used the fact that $\mathsf{P}(X \in B) = \frac{|B|}{|C|}$.

(b) Intuitively, since X represents the number of white balls sampled from n draws of an urn with w white and b black balls, $n - X$ must represent the number of black balls sampled in those same n draws. Thus, reversing the positions of the white and black balls, we should have

$$n - X \sim \mathsf{HGeom}(b, w, n).$$

We can also prove this mathematically as follows:

$$\mathsf{P}(n - X = k) = \mathsf{P}(X = n - k) = \frac{\binom{w}{n-k}\binom{b}{k}}{\binom{w+b}{n}},$$

and noting the last expression is just the PMF of $\mathsf{HGeom}(b, w, n)$.

Problem 5

A copy machine is used to make n pages of copies per day. The machine has two trays in which paper gets loaded, and each page is taken randomly and independently from one of the other trays. At the beginning of the day, the trays are refilled so that they each have m pages. Using simulations in R, find the smallest value of m for which there is at least a 95% chance that both trays have enough paper on a particular day, for $n = 10$, $n = 100$, $n = 1000$, and $n = 10000$.

Solution: See the code below. A summary of the results obtain by running that code is here:

	$n = 10$	$n = 100$	$n = 1000$	$n = 10000$
m	8	60	531	5098

```
# Estimate the probability that there are enough pages
enough_paper <- function(m, n, nsamples) {
   records <- rep(0, nsamples)
   for(i in 1:nsamples) {
      # Select trays
      copies = sample(c(0,1), n, replace = TRUE)

      # Enough pages for both trays?
      records[i] = sum(copies) <= m & (n - sum(copies)) <= m
   }
   mean(records)
}

#n = 10
enough_paper(m = 7, n = 100, nsamples = 100000)
enough_paper(m = 8, n = 100, nsamples = 100000)

#n = 100
enough_paper(m = 59, n = 100, nsamples = 100000)
enough_paper(m = 60, n = 100, nsamples = 100000)

#n = 1000
enough_paper(m = 530, n = 1000, nsamples = 100000)
enough_paper(m = 531, n = 1000, nsamples = 100000)

#n = 10000
enough_paper(m = 5097, n = 10000, nsamples = 100000)
enough_paper(m = 5098, n = 10000, nsamples = 100000)
```

Listing 1: Code implementing the simulations for Problem 5