CSCI-GA.1180:
Mathematical Techniques for Computer Science Applications
New York University, Fall 2016

Instructor: Margaret Wright

Grader: Abhishek Nilesh Shah

Homework Assignment 5

Assigned 15 November 2016; due 11:59pm, 29 November 2016

Exercise 5.1. There are five distinct paths that lead from the base of a mountain to the top, and each of these five paths can also lead from the top of the mountain to the bottom. A "round trip" up and down the mountain is a pair of path numbers (p_1, p_2), where p_1 is the number of the path taken from the bottom to the top, and p_2 is the number of the path taken from the top back down to the bottom.

The question is: how many round trips are there? Here are four answers, all correct under certain circumstances.

1. $20 = 5 \times 4$
2. $25 = 5^2$
3. $\binom{5}{2} = 10$
4. $\binom{6}{2} = 15.$

For each of these four cases, explain why the given number is the correct answer under specific assumptions about acceptable round trips, state those assumptions (in words), and give the complete list of pairs that constitute acceptable round trips.

Exercise 5.2. Consider a family with two children who have different ages. Assume that, in any birth of one baby, the probability is $\frac{1}{2}$ that the baby is a boy, and $\frac{1}{2}$ that the baby is a girl. Let A be the event that the older of the two children is a girl, and let B be the event that at least one of the two children is a girl.

1. What is the conditional probability that both children are girls, given that the older child is a girl? Explain how you derived your answer.
2. What is the conditional probability that both children are girls, given that at least one of the children is a girl? Explain how you derived your answer.

Exercise 5.3. Assume that you have a farm whose main crop is carrots. Over the very long term, 14% of the carrots harvested are unacceptable because they are too sour. (They should be sweet.) Consider a sample of 11 carrots, harvested from your farm at random. In each answer to the following questions, please give the formula you used as well as the numerical value. (Suggestion: use Matlab.)

(a) What is the probability, shown to at least four decimal figures, that all of the carrots in the sample will be sweet?

(b) What is the probability that 1 or more carrots in the sample are sour?

(c) What is the probability that at most 1 carrot in the sample is sour?

Exercise 5.4. Write a program that generates samples consisting of n components, where each component is either 1 ("success") or 0 ("failure"). Suppose that "success" means that a coin toss has produced heads, where the coin being used is a biased coin in which the probability of tossing heads is p such that $p > \frac{1}{2}$.

In each sample, the n components should be defined using a random number generator and the value of p, as described next. To obtain the value of any one component `x` knowing p, use the Matlab uniform random number generator `rand`, which produces random numbers uniformly distributed in $(0, 1)$:

```
z = rand
if z <= p,
  x = 1
else
  x = 0
end
```

Assume that $p = 0.75$. Your program should produce and print 8 samples, each containing 12 components. For each sample, count (i) the number of successes and (ii) the ratio of the number of successes to n. Comment on how closely the computed fraction of successes corresponds to p.

Exercise 5.5. Let e denote the m-vector whose components are all equal to 1, and let b denote a specific m-vector, so that

$$e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

Suppose that we want to find one number, denoted by β, that minimizes various norms of the residual vector r whose i-th component is $b_i - \beta$:

$$r = b - \beta e = \begin{pmatrix} b_1 - \beta \\ b_2 - \beta \\ \vdots \\ b_m - \beta \end{pmatrix},$$

where

$$\|r\|_1 = \sum_{i=1}^{m} |r_i|, \qquad \|r\|_2^2 = \sum_{i=1}^{m} r_i^2, \qquad \text{and} \qquad \|r\|_\infty = \max_i |r_i|.$$

(Note that the middle formula gives the squared two-norm.)

(a) Let β_2 be the scalar that produces the smallest value of the squared two-norm $\|r\|_2^2 = \|b - \beta e\|_2^2$. Give a mathematical expression for β_2 in terms of the elements of b. (Hint: use the normal equations.)

(b) Let β_∞ be the scalar that produces the smallest value of the infinity norm of the residual, $\|r\|_\infty = \max_i |r_i|$. Express β_∞ in terms of the elements of b and explain why the infinity norm is minimized for this value.

(c) [Optional.] For the one-norm, the scalar β_1 that minimizes $\|b - \beta e\|_1$ is the *median* of b, i.e., the value separating the values $\{b_i\}$ into a higher half and a lower half. (Consider a set of ordered real scalars $\{y_i\}$, $i = 1, \ldots,$ with $y_i \le y_{i+1}$. If there are $2k + 1$ values, the median is y_{k+1}; if there are $2k$ values, the median is $\frac{1}{2}(y_k + y_{k+1})$. The Matlab command `median(y)` returns the median when `y` is a vector.) This result (c) about the one-norm is not easy to prove, but please give a proof if you feel so inclined. In any case, you should use this result in part (d).

(d) Let $m = 7$. For each of the following vectors b, calculate β_1, β_2, and β_∞ (suggestion: use Matlab), based on the results of (a)–(c).

(i) $b = (-150,\ 25,\ 0,\ -70,\ 70,\ 150,\ -25)^T$;

(ii) $b = (1,\ 500,\ 1,\ 250,\ 1,\ 300,\ 1)^T$;

(iii) $b = (1,\ -2,\ -10,\ 4,\ 7,\ -5,\ 100000)^T$.

For each of these three cases, print the vector b, the three values β_1, β_2, and β_∞ for that b, and the three residual vectors and their norms:

- for case (i), print $r_1 = b - \beta_1 e$ and $\|r_1\|_1$;
- for case (ii), print $r_2 = b - \beta_2 e$ and $\|r_2\|_2$; and
- for case (iii), print $r_\infty = b - \beta_\infty e$ and $\|r_\infty\|_\infty$.

In each case, comment on how well (or how badly) the single value β "represents" the nature of the original vector.

(e) Which norm(s) appear to be least sensitive to "outliers" (a small number of elements of b that differ widely in magnitude from the other elements)? Explain your answer.