Feedback — Homework 7

Help Center

You submitted this homework on **Mon 20 Apr 2015 8:59 PM PDT**. You got a score of **7.00** out of **7.00**.

Question 1

Week 1: combinatorial probabilities.

Six cups and six saucers come in pairs, two pairs are red, two are white, and two are blue. If cups are randomly assigned to saucers find the probability that no cup is upon a saucer of the same colour.

Your Answer		Score	Explanation
$\frac{1}{6}$			
$\frac{1}{120}$			
$\frac{1}{3}$			
$\frac{4}{45}$			
$\frac{1}{45}$			
$\frac{1}{9}$	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

It is simplest to systematically account for all possibilities of arrangements. For purposes of enumeration we may consider the saucers to be labelled from 1 through 6, the cups likewise. Formally, each permutation $(\Pi_1, \ldots, \Pi_6)$ of $(1, \ldots, 6)$ represents a sample point of the experiment and corresponds to the cup-saucer matchings $i \mapsto \Pi_i$. As $(\Pi_1, \ldots, \Pi_6)$ ranges over all permutations, we range over all $6!$ assignments of cups to saucers. Of these, those arrangements where all cups are on saucers of a different colour are of two types.

1. *Both cups of a given colour are placed on matching saucers of a different colour.* A little introspection shows that this means that each pair of cups of a given colour must be matched with a pair of saucers of a different colour. If, say, both red cups are placed on blue saucers then both blue cups must be placed on white saucers which means in turn that both white cups must be placed on blue saucers. The red cups may be placed on the blue saucers in $2!$ ways, the blue cups on the white saucers in $2!$ ways, and the white cups on the blue saucers in $2!$ ways, for a total of $2! \times 2! \times 2! = 8$ possibilites with the assignments $\mathrm{red} \to \mathrm{blue} \to \mathrm{white}$. Interchanging the rôles of blue and white, there are 8 more assignments of the form $\mathrm{red} \to \mathrm{white} \to \mathrm{blue}$. There are hence a total of 16 arrangements with both cups of each colour placed on matching saucers of a different colour.
2. *The cups of a given colour are placed on saucers of differing colours, neither of the original colour.* We may systematically enumerate the possibilities as follows. The first red cup may be placed on one of the four saucers that are not red; once a saucer is selected, the second red cup has two possibilites in the saucers of the remaining colour; there are thus 8 ways in which we may deploy the red cups on one blue and one white saucer. Now both red cups are deployed, one on a blue saucer and one on a white saucer. The two blue cups must now be placed one on the remaining white saucer and one on a red saucer. The red saucer may be selected in two ways and the blue cup to be placed on it may be selected in two ways; there are thus 4 ways in the blue cups may be placed on a white and a red saucer. Finally, there remain two white cups, one blue and one red saucer, and both deployments of cups on saucers give rise to valid arrangements. In total we have $8 \times 4 \times 2 = 64$ arrangements of the cups on the saucers so that the cups of a given colour are placed on saucers of differing colours, neither of the original colour.

Combining the two cases, there are $16 + 64 = 80$ arrangements of cups on saucers so that no cup is on a saucer of a matching colour. The associated probability is hence $(16 + 64)/6! = 1/9$, or about ten percent.

Those who managed to brave the dangerous bend Lecture 12.2: d will have seen a systematic way forward through the method of *inclusion and exclusion*. This has the virtue of extensibility to other settings though at some level there is no escaping the individual combinatorial calculations that have to be made. I will leave you to try it out: number the cups from, say, 1 through 6 and let A_j denote the event that the jth cup lands on a saucer of its own colour. You will now need to calculate the probabilities of intersections of the form: $\mathbf{P}(A_j)$, $\mathbf{P}(A_i \cap A_j)$, …, $\mathbf{P}(A_1 \cap \cdots \cap A_6)$. There's no getting around a tiresome breakdown of cases but the individual calculations are not hard.

Question 2

Week 6: The binomial distribution.

Let S_{10} denote the accumulated number of successes in ten tosses of a coin with unknown success probability p. The only other information you are given is that $\mathrm{Var}(S_{10}) = 0$. Estimate $\mathbf{P}\{1 \leq S_{10} \leq 9\}$.

Your Answer		Score	Explanation
◉ 0	✔	1.00	
○ 0.5			
○ 0.2461			
○ 0.0020			
○ 0.9980			
○ 1			
Total		1.00 / 1.00	

Question Explanation

The direct path:

As $S_{10} \sim \text{Binomial}(10, p)$ for some p, by Lecture 10.1: p, we know that $\text{Var}(S_n) = 10p(1-p)$. But then $\text{Var}(S_n)$ can be equal to zero if, and only if, either $p = 0$ or $p = 1$. In other words, the coin is either double-headed or double-tailed! Consequently, all the atomic mass of S_{10} is concentrated either at 0 or at 10. Spelt out, if $p = 0$, then $\mathbf{P}\{S_n = 0\} = b_{10}(0; 0) = 1$ while, if $p = 1$, then $\mathbf{P}\{S_n = 10\} = b_{10}(10; 1) = 1$. In either case S_{10} has no mass in the set of values $\{2, 3, \ldots, 9\}$ and so $\mathbf{P}\{1 \leq S_{10} \leq 9\} = 0$.

Via Chebyshev's inequality:

By a direct application of Chebyshev's inequality of Lecture 11.1: f, we obtain

$$0 \leq \mathbf{P}\{|S_{10} - 10p| \geq 1\} \leq \frac{\text{Var}(S_{10})}{1^2} = 0.$$

It follows that $\mathbf{P}\{|S_{10} - 10p| \geq 1\} = 0$ identically. As S_{10} is an arithmetic random variable with support only in the integers from 0 to 10, this is the same as saying that $\mathbf{P}\{S_{10} = 10p\} = 1$, that is to say, S_{10} has all its atomic mass concentrated at the point $10p$. But, if $\text{Var}(S_{10}) = 0$ then, as argued earlier, $p = 0$ or $p = 1$. In either case, S_{10} places no mass in the set $\{2, 3, \ldots, 9\}$ and so $\mathbf{P}\{1 \leq S_{10} \leq 9\} = 0$.

Question 3

Week 7: The de Moivre–Laplace theorem..

Let S_{10} be a binomial random variable denoting the accumulated number of successes in ten Bernoulli trials, each with success probability 0.7. Use the normal approximation to estimate $\mathbf{P}\left\{\left|\frac{S_{10}}{10} - 0.7\right| > 0.2\right\}$ to four decimal places. (There are a variety of free calculators on the web if you don't have access to a scientific calculator or mathematical software; just type in "normal distribution calculator" into a search engine.)

This portion is optional: Now compare with the exact binomial probability for this event. The purpose of this problem is to empahsise that sample sizes need to become sufficiently large before normal approximation gives accurate answers.

Your Answer		Score	Explanation
◉ 0.1675	✔	1.00	
○ 0.1245			
○ 0.0298			
○ 0.5143			
○ 0.0455			
○ 0.2234			
Total		1.00 / 1.00	

Question Explanation

By Lecture 11.3; d, we know that as n gets larger and larger an appropriately centred and scaled version of S_n behaves approximately normally. This is the content of the de Moivre–Laplace theorem of Lecture 11.3: d. Formally, if we set

$$S_n^* = \frac{S_n - np}{\sqrt{npq}}$$

then, for any $a < b$,

$$\mathbf{P}\{a < S_n^* < b\} \to \int_a^b \phi(x)\,dx = \Phi(b) - \Phi(a) \qquad (\text{as } n \to \infty).$$

Let's see how we can connect this fact to our problem: we will need to shape the desired inequality to get it into a form where the de Moivre–Laplace theorem can be used. For the given problem, $n = 10$, $p = 0.7$, and $q = 1 - p = 0.3$. Accordingly,

$$\mathbf{P}\left\{\left|\frac{S_{10}}{10} - 0.7\right| > 0.2\right\} = \mathbf{P}\left\{\left|S_{10} - 10 \times 0.7\right| > 2\right\} = \mathbf{P}\left\{\left|\frac{S_{10} - 10 \times 0.7}{\sqrt{10 \times 0.7 \times 0.3}}\right| > \frac{2}{\sqrt{10 \times 0.7 \times 0.3}}\right\}$$

$$> 1.38013\} = 1 - \mathbf{P}\{|S_n^*| \le 1.38013\} = 1 - \mathbf{P}\{-1.3013 \le S_n^* \le 1.38013\} \approx 1 - \int_{-1.38013}^{1.38013} \phi(x)\,dx = 0.1675$$

truncated to four decimal places.

The reader may well question the appropriateness of the normal approximation in this setting as $n = 10$. The exact calculation using the binomial probabilities gives

$$\mathbf{P}\left\{\left|\frac{S_{10}}{10} - 0.7\right| > 0.2\right\} = \sum_{k:k<5 \text{ or } k>9} b_{10}(k; 0.7) = b_{10}(0; 0.7) + b_{10}(1; 0.7) + b_{10}(2; 0.7) + b_{10}(3; 0.7) + b_{10}(4; 0.7) + b_{10}(10; 0.7) = 0.0756$$

truncated to four decimal places. The error in approximation is significant. The reason, of course, is that n is much too small in this setting for us to apply the de Moivre–Laplace theorem with any degree of confidence. As a rule of thumb one usually looks for n to be 30 or larger.

Question 4

Weeks 3, 4, and 6: conditional probabilities, independence, and the binomial distribution.

Let $X_1, \ldots, X_n$ be a sequence of Bernoullli trials with success probability p. Let $S_n = X_1 + \cdots + X_n$ denote the number of accumulated successes. For any $0 \le k \le n$, determine $\mathbf{P}\{X_n = 1 \mid S_n = k\}$ [*Hint*: Consider the decomposition $S_n = S_{n-1} + X_n$.]

(Once you obtain the answer look at it carefully. Is it intuitive? Do you see a direct path to the answer?)

Your Answer		Score	Explanation
$\frac{k}{2(n-1)}$			
$\frac{k-1}{n-1}$			
$\frac{k}{n-1}$			
$\frac{k-1}{n}$			
$\frac{k}{n}$	✔	1.00	
$\frac{k}{2n}$			
Total		1.00 / 1.00	

Question Explanation

Writing $S_n = S_{n-1} + X_n$, by the definition of conditional probability, we have

$$\mathbf{P}\{X_n = 1 \mid S_n = k\} = \frac{\mathbf{P}\{X_n = 1 \text{ and } S_n = k\}}{\mathbf{P}\{S_n = k\}} = \frac{\mathbf{P}\{X_n = 1 \text{ and } S_{n-1} + X_n = k\}}{\mathbf{P}\{S_n = k\}} = \frac{\mathbf{P}\{X_n = 1 \text{ and } S_{n-1} = k-1\}}{\mathbf{P}\{S_n = k\}}.$$

But $S_{n-1} = X_1 + \cdots + X_{n-1}$ is independent of X_n and so

$$\mathbf{P}\{X_n = 1 \mid S_n = k\} = \frac{\mathbf{P}\{X_n = 1\} \times \mathbf{P}\{S_{n-1} = k-1\}}{\mathbf{P}\{S_n = k}.$$

We know the mass functions of all the variables on the right: $X_n \sim \text{Bernoulli}(p)$, $S_{n-1} \sim \text{Binomial}(n-1, p)$, and $S_n \sim \text{Binomial}(n, p)$. Accordingly,

$$\mathbf{P}\{X_n = 1\} = p, \quad \mathbf{P}\{S_{n-1} = k-1\} = \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)}, \quad \mathbf{P}\{S_n = k\} = \binom{n}{k} p^k q^{n-k}.$$

Substituting the values and simplifying, we obtain

$$\mathbf{P}\{X_n = 1 \mid S_n = k\} = \frac{p \cdot \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)}}{\binom{n}{k} p^k q^{n-k}} = \frac{(n-1)^{k-1}}{(k-1)!} p^k q^{n-k} \Big/ \frac{n^{\underline{k}}}{k!} p^k q^{n-k} = \frac{k}{n}.$$

Surprising? The answer does not depend upon p. A little thought shows why. Given that there are exactly k successes in n trials, the symmetry of the problem tells us that every distribution of these k successes among the n trials is equally likely of occurrence. This implies

that any given trial has a conditional probability k/n of being a success.

Question 5

Week 2: Mass functions, probability measure.

The Martian alphabet has been deciphered! It consists of five symbols α, β, γ, δ, and ϵ (which appear to bear an uncanny resemblance to certain modern Greek letters). A study of Martian literature reveals that these symbols appear in the frequencies given in the following table.

Martian symbol	α	β	γ	δ	ϵ
Probability	0.25	0.15	0.15	0.25	0.20

Your Martian interlocutor selects a random letter from this alphabet in accordance with the given symbol atomic probabilities but does not reveal it to you. He will however answer yes or no truthfully to any question you pose him. Your task is to determine which symbol has been selected by asking him a series of yes/no questions. For example, one straightforward strategy that you could implement is to sequentially work through the list of symbols by asking him if the selected symbol is α and if he says no asking him if it is β and proceeding in this fashion through the alphabet until you get a yes answer. But this strategy may not be the best from the point of view of minimising the expected number of questions that you will have to ask before you know which symbol has been selected. Design a strategy (that is to say, a series of yes/no questions that culminate in an identification of the selected symbol) that will determine the selected symbol with the *smallest* expected number of queries and evaluate this expected value. [*Hint*: What is the probability space for this problem? Once you choose a strategy, the number of queries that your chosen strategy requires to identify the selected symbol is itself an arithmetic chance variable, say, Z which takes integer values. Evaluate its mass function $p(k) := \mathbf{P}\{Z = k\}$ and thence its expected value $\mathbf{E}(Z) = \sum_k k \cdot p(k)$. Different strategies will yield different expected values. Your task is to select a strategy with the smallest expected number of queries.]
This portion is optional: Once you have done the problem, think about why you think your strategy is optimal. Can you extend it? This problem is related to the venerable game of twenty questions. And it is at the heart of modern data compression; the fact that you can watch the streaming lecture videos for this course is a testament to the principles behind this problem!

Your Answer		Score	Explanation
○ 2			
○ 2.1			
◉ 2.3	✔	1.00	
○ 2.5			
○ 3			
○ 4			
Total		1.00 / 1.00	

Question Explanation

The student who has played the game *twenty questions* as a child will recognise in this problem a variant on the same theme. The objective? To identify a selected (but *a priori* unknown) ground truth with the smallest number of queries.

As always, we should begin with the sample space. What is chance-driven in the problem? Clearly, the selection of symbol. Accordingly, the sample space is $\Omega = \{\alpha, \beta, \gamma, \delta, \epsilon\}$. The outcome of the chance experiment is a symbol $S \in \Omega$. The associated atomic measure is provided for us.

In this setting, a *query strategy* $\mathscr{Q}$ is a fixed procedure which determines any selected symbol with a finite number of queries. The *number* of queries, $N = N_{\mathscr{Q}}$ is a function of S which, for each S identifies the number of queries needed by $\mathscr{Q}$ to identify it: $S \mapsto N_{\mathscr{Q}}$.

The simplest strategy (let's call it the *naïve strategy* $\mathscr{N}$) is to simply go sequentially through the symbols. The first question: "Is it α?". If the answer is no, then go forward with the second question: "Is it β?". If the answer is no, then ask the third question: "Is it γ?". If the answer is still stubbornly no, then ask the fourth question: "Is it δ?". And, if one gets another no answer ask, the final (fifth) question: "Is it ϵ?" which will reveal the answer. (Of course, the student has realised that the fifth question is superfluous: if we have received four negative responses to the first four questions, the symbol must be ϵ. But, this is the naïve strategy after all: no thinking is required for this one!) We see that the number of queries is a chance-driven entity which inherits its mass function from that of the original symbol alphabet. Indeed:

Martian symbol	α	β	γ	δ	ϵ
$N_{\mathscr{N}}$	1	2	3	4	5
Probability	0.25	0.15	0.15	0.25	0.20

It follows that the expected number of queries for the naïve strategy is given by

$$\mathbf{E}(N_{\mathscr{N}}) = 1 \cdot 0.25 + 2 \cdot 0.15 + 3 \cdot 0.15 + 4 \cdot 0.25 + 5 \cdot 0.20 = 3.$$

On average this strategy requires three queries to identify the right answer.

Of course, even a little thought shows that the naïve strategy is, well, naïve. To do better we need a guiding principle and it is this.

Slogan: *Expend very few queries on more likely symbols; be prepared to expend many queries on less likely symbols.* This is the basis behind a *divide and conquer strategy* $\mathscr{D}$. We attempt, at each stage, to reduce the remaining uncertainty by as close to one-half as we can. A computer scientist will be familiar with this approach.

1. The first step divides the set $\{\alpha, \beta, \gamma, \delta, \epsilon\}$ into two sets $\{\alpha, \delta\}$ and $\{\beta, \gamma, \epsilon\}$ whose probabilities are 0.5 and 0.5, apiece. This can be done, for example, by the simple query: "Is the symbol α or δ?".
   A. If the answer is yes, the next query splits $\{\alpha, \delta\}$ into two, thus identifying the symbol: the query "Is the symbol α?", for instance, will suffice.
      a. If the answer is yes then the symbol is α and two queries sufficed to identify it.
      b. If the answer is no then the symbol is δ and it is identified with two queries also.
   B. If the answer is no, the next query splits $\{\beta, \gamma, \epsilon\}$ into two as nearly equiprobable sets as possible via the query "Is the symbol ϵ?", for example.
      a. If the answer is yes then the symbol ϵ is identified with two queries.
      b. If the answer is no then the next query splits the remaining two symbols $\{\beta, \gamma\}$ into two equiprobable possibilities via, say, the query "Is it β?".
         i. If the answer is yes then the symbol is β and it required three queries to identify it.
         ii. If the answer is no then the symbol is γ and three queries also identified it.

In accordance with our slogan, the three higher probability symbols α, δ, and ϵ are identified with two queries apiece; the two low probability symbols β and γ are identified with three queries apiece. The number $N_{\mathscr{D}}$ of queries required by the divide-and-conquer strategy hence has mass function

$N_{\mathscr{D}}$	2	3
Probability	$0.25+0.25+0.2$	$0.15+0.15$

It follows that

$$\mathbf{E}(N_{\mathscr{D}}) = 2 \times (0.25 + 0.25 + 0.2) + 3 \times (0.15 + 0.15) = 1.4 + 0.9 = 2.3.$$

Thus, the divide-and-conquer query strategy requires only 2.3 queries on average to identify a Martian symbol. This may not seem like much of an improvement over the 3 queries required on average by the naïve strategy but imagine now that a Martian Tolstoy is intent on transmitting the Martian equivalent of *War and Peace* to us: a savings of 0.7 queries on average per symbol over billions of symbols results in a very substantial savings.
A powerful reinterpretation of this simple children's game is obtained by the realisation that the query sequence required to identify a given symbol uniquely identifies that symbol or, in other words, is a unique name or *codeword* for that symbol. For definiteness, write 1 for "yes" and 0 for "no". Then the two strategies we have discussed give the symbols in our fictional Martian alphabet the following names:

Martian symbol	α	β	γ	δ	ϵ
Naïve strategy codeword	1	01	001	0001	00001
Divide-and-conquer codeword	11	001	000	10	01
Probability	0.25	0.15	0.15	0.25	0.20

Each permissible query strategy hence provides a distinct codeword for each symbol. (This is clearly a minimal requirement.) Our question may hence be restated as follows: *Select a permissible query strategy for which the expected codeword length is minimal*. This is one of the central questions in coding theory. What good codes do is reduce redundancy by providing spare descriptions for data.

The simple, top-down, divide-and-conquer strategy we have outlined indeed turns out to be optimal in the setting we have provided. But it is difficult to determine what is the best way to divide large sets when probabilities don't exactly break into one-half each time. David A. Huffman discovered that working backwards from the least probable symbols is the right way to approach the problem in a general setting. The class of optimal procedures built on this observation is now called *Huffman coding* in honour of his pioneering work.

Question 6

Weeks 6 and 7: Polls, the binomial distribution, Chebyshev's inequality.

The following prompt should be used for Questions 6 and 7. The goal of this problem is to see whether reliable inferences can be made from unreliable data.

A frightening new epidemic is sweeping through a large population in a metropolitan area. You are charged with assessing the threat by determining an estimate of the number of affected people. You do a random draw of n people from the population and perform a (yet to be perfected) medical test on them to determine the presence (or lack thereof) of the disease agent in them. If the person is affected then the medical test will produce a positive result with probability 0.8 and a false negative result with probability 0.2. If the person is free from the disease agent, the test produces a negative result with probability 0.9 but produces a false positive result with probability 0.1.

Suppose that in a random sample of $n = 100$ test results you find that 40 have tested positive. Assuming that you know the false positive and false negative rates for the medical test, which of the following numbers provides the best estimate for the percentage of the population who harbor the disease? Round up or down to the nearest whole percent.

Your Answer		Score	Explanation
◉ 43%	✔	1.00	

- 50%
- 40%
- 32%
- 57%
- 28%

Total	1.00 / 1.00

Question Explanation

The student may recognise that the elements are similar to that of Question 1 in Homework 6: in both cases we are dealing with a sample that is unreliable. This suggests that an analysis along similar lines will be profitable.

The sample space: In our sanitised model we're dealing with a sequence of repeated independent trials so it will suffice to understand the sample space and measure attendant on each trial. What is the sample space for a single trial? Well, there are two chance elements: whether a randomly individual from the population is a disease carrier, and whether the result of the test on him is correct. Let the Bernoulli variable X denote the disease state of the selected individual: write 1 if he carries the disease and 0 if he is disease-free. Likewise, let Y denote the result of the test: write 1 if the test shows positive and 0 if it shows negative. The sample space hence has four elements with $(X, Y) \in \{(0,0), (0,1), (1,0), (1,1)\}$. Let p_0 denote the (fixed but unknown) fraction of individuals in the population with who are disease carriers and $q_0 = 1 - p_0$ the corresponding fraction of individuals who are disease-free. We wish to estimate p_0.

The probability measure: The Bernoulli variables X and Y are not independent but the mass function attached to the atoms may be deduced from the given conditional probabilities:

$$\mathbf{P}\{X = 1\} = p_0, \quad \mathbf{P}\{X = 0\} = q_0,$$
$$\mathbf{P}\{Y = 1 \mid X = 1\} = 0.8, \quad \mathbf{P}\{Y = 0 \mid X = 1\} = 0.2, \quad \mathbf{P}\{Y = 1 \mid X = 0\} = 0.1, \quad \mathbf{P}\{Y = 0 \mid X = 0\} = 0.9.$$

The observable is the test result Y which is a Bernoulli variable whose "success" probability p may now be inferred by the theorem of total probability:

$$p = \mathbf{P}\{Y = 1\} = \mathbf{P}\{Y = 1 \mid X = 1\}\mathbf{P}\{X = 1\} + \mathbf{P}\{Y = 1 \mid X = 0\}\mathbf{P}\{X = 0\} = 0.8p_0 + 0.1(1 - p_0) = 0.7p_0 + 0.1,$$
$$q = 1 - p = 0.9 - 0.7p_0.$$

Suppose $Y_1, \ldots, Y_n$ is a random sample of test results governed by a sequence of $\text{Bernoulli}(p)$ trials where $p = 0.7p_0 + 0.1$. Write $S_n = Y_1 + \cdots + Y_n$. Then $S_n \sim \text{Binomial}(n, p)$, whence $\mathbf{E}(S_n) = np$ or $\mathbf{E}(S_n/n) = p = 0.7p_0 + 0.1$. Equivalently,

$$p_0 = \frac{\mathbf{E}\left(\frac{S_n}{n}\right) - 0.1}{0.7}.$$

With $n = 100$ and $S_{100} = 40$, we see that we may estimate p_0 by

$$\frac{\frac{40}{100} - 0.1}{0.7} = \frac{3}{7} = 0.428571$$

or 43% rounded up.

Question 7

You are tasked by the Centre for Disease Control to generate a reliable estimate of the fraction of people carrying the disease in the metropolitan population. You are told that your estimate should have an error of no more than 1% with a confidence of at least 99%. After taking a hurried on-line course on probability you realise that a sample of size 100 is much too small to give you reliable results. Using Chebyshev's inequality, amongst the following options, select the smallest value of sample size n that is required to ensure that your procedure will provide an estimate with an error of no more than 1% and a confidence of 99%.

Your Answer		Score	Explanation
2000			
5556			
510205	✔	1.00	
1068			
7938			
1000000			
Total		1.00 / 1.00	

Question Explanation

With notation as introduced in the solution to Question 6, we know per the law of large numbers that S_n/n is concentrated at its expected value $p = 0.7p_0 + 0.1$. This suggest that we may estimate p_0 by the principled estimate

$$\widehat{p}_0 := \frac{\frac{S_n}{n} - 0.1}{0.7}\,.$$

We wish to estimate the probability that $\widehat{p}_0$ deviates from p_0 in absolute value by more than $\epsilon = 0.01$. As in our analysis of Question 1, Homework 6, we rephrase the problem in a form where we can leverage Chebyshev's inequality:

$$\begin{aligned}\mathbf{P}\{|\widehat{p}_0 - p_0| > \epsilon\} = \mathbf{P}\left\{\left|\frac{\frac{S_n}{n} - 0.1}{0.7} - p_0\right| > \epsilon\right\} = \mathbf{P}\left\{\left|\frac{S_n}{n} - (0.7p_0 + 0.1)\right| > 0.7\epsilon\right\} \\ = \mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > 0.7\epsilon\right\} \leq \frac{1}{4n(0.7\epsilon)^2} = \frac{1}{1.96n\epsilon^2}\,.\end{aligned}$$

To meet the given confidence requirement it will suffice if, with $\epsilon = \delta = 0.01$, we have

$$n > \frac{1}{1.96 \times \epsilon^2 \times \delta} = \frac{1}{1.96 \times (0.01)^2 \times (0.01)} = 510\,204.$$

The reader will have recognised that we can improve the estimate by using sharper bounds than Chebyshev's; the central limit theorem provides a tempting alternative. The good news is hence that we can overcome unreliability in data. But, as we saw in Question 1, Homework 6, however we cut it, the price we pay for our lack of reliability in the data is in an increase in the requisite sample size.