Problem 1 – Probabilities of events. Warmup, not graded

Professor Chance Gardner is in a hurry to get to the class she is teaching. She can drive to campus either on I5 (denote this event by $I$) or the shorter way over the University Bridge (denote this event by $\bar{I}$). The latter can be either up (denote this event by $U$) or down. After she gets to campus she may find a parking spot (denote this event by $P$) or not (in which case she will have to park in the Montlake area). Finally, denote by $L$ the event “professor Gardner is late for class”. Below are given the probabilities of all possible individual outcomes of this random experiment.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$IPL$</td>
<td>0.05</td>
</tr>
<tr>
<td>$IP\bar{L}$</td>
<td>0.1</td>
</tr>
<tr>
<td>$I\bar{P}L$</td>
<td>0.15</td>
</tr>
<tr>
<td>$\bar{I}UPL$</td>
<td>0.18</td>
</tr>
<tr>
<td>$I\bar{U}PL$</td>
<td>0.07</td>
</tr>
<tr>
<td>$I\bar{U}\bar{P}L$</td>
<td>0.35</td>
</tr>
<tr>
<td>$I\bar{U}PL$</td>
<td>0.1</td>
</tr>
</tbody>
</table>

The outcomes not listed (like $I\bar{P}\bar{L}$ = "drives on I5, doesn’t find a parking space and is not late for class") have probability 0. Some “individual outcomes” in the above list are really events (for example $IPL$ which is the union of $IPLU$ and $IPL\bar{U}$). We still can call it an individual outcome assuming that if the route is over the I5 bridge then it doesn’t matter what is the state of University bridge (or we may not find out what this state is).

a. Make a neatly labeled drawing of this sample space, showing all the possible outcomes and their probabilities.

b. What is the probability that Chance drives over the University Bridge ?

c. What is the probability that Chance makes it to class in time?

d. What is the probability that Chance doesn’t find a parking spot and is late
for class?

e. Which of the following events is more probable: $A =$ “Chance drives over the University bridge and is not late for her class, or $B =$ “Chance drives over the I5 bridge and is not late for her class” ?

**Problem 2 – Probabilities of events. Warmup, not graded**

a. Show that for any two events $A, B$, $P(A \cup B) \leq P(A) + P(B)$.

b. The ACME company manufactures the Pogo Thunder personal jet plane. The plane consists of three major systems: the body, the engines, and the electronics with failure probabilities of 0.01, 0.02 and 0.03 respectively (the Pogo Thunder is the least reliable personal jet on the market). The plane will work only if no system fails. Prove that the probability that the plane has no failure is larger or equal to 0.94.

c. This problem is a special case of a very useful result in probability and measure theory called the union bound. Prove that for any events $A_1, A_2, \ldots, A_n \subseteq S$ the probability of the union is no larger than the sum of the events’ probabilities, i.e.

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) \leq P(A_1) + P(A_2) + \ldots + P(A_n)$$  \hspace{1cm} (1)

**Problem 3**

Let $S = \{0, 1, 2, \ldots, 9\}$ be the sample space of the 10 digits and $P(n)$ be an exponential distribution over this space, given by

$$P(n) = \frac{1}{Z} \gamma^n$$  \hspace{1cm} (2)

with $\gamma = 1/2$.

a. What is the value of the normalization constant $Z$? Give either a numerical answer or a formula in terms of $\gamma$.

b. What is the probability that $n < 5$? Give either a numerical answer or a formula in terms of $\gamma$.

c. What is the probability that $n$ is odd? Give either a numerical answer or a formula in terms of $\gamma$.

d. Let $S_3 = S^3$ be the sample space of all sequences of 3 elements from $S$. How many elements has $S_3$?

e. We assume that the sequences in $S_3$ are obtained by sampling independently
3 times from $S$ with the probability distribution $P$ defined in (2). Compute the probability of the sequences $(1,2,3)$, $(0,0,0)$ and $(0,1,1)$.

**f.** What is the sequence $\{n^{(1)}, n^{(2)}, n^{(3)}\}$ that has highest probability of occurrence? What is the sequence with the lowest probability?

**g.** For any outcome $\{n^{(1)}, n^{(2)}, n^{(3)}\}$ define

$$z(n^{(1)}, n^{(2)}, n^{(3)}) = 100n^{(1)} + 10n^{(2)} + n^{(3)},$$

i.e. $z(n^{(1)}, n^{(2)}, n^{(3)})$ is the decimal number whose digits are drawn randomly and independently from $P$. For instance $z(2,0,6) = 206$ and $z(0,1,2) = 12$.

Let $\{n^{(1)}, n^{(2)}, n^{(3)}\}$ and $\{\tilde{n}^{(1)}, \tilde{n}^{(2)}, \tilde{n}^{(3)}\}$ be two sequences with $z(n^{(1)}, n^{(2)}, n^{(3)}) > z(\tilde{n}^{(1)}, \tilde{n}^{(2)}, \tilde{n}^{(3)})$. Is it true that this implies

$$P(n^{(1)}, n^{(2)}, n^{(3)}) \leq P(\tilde{n}^{(1)}, \tilde{n}^{(2)}, \tilde{n}^{(3)})$$

In other words, is it true that the resulting probability distribution over 3 digit integers is monotonically decreasing?

Prove or give a counterexample.

FYI: The sum of the geometric progression

$$a + ax + ax^2 + ax^3 + \ldots + ax^{m-1} = \frac{1 - x^m}{1 - x}$$

**Problem 4 – Probabilities of events – Ungraded**

The “Little Amazon” company sells books on the internet. “Little Amazon” has the following titles for sale: 0 – “War and Peace”, 1 – “Harry Potter & the Deathly Hallows”, 2 – “Winnie the Pooh”, 3 – “Get rich NOW”, 4 – “Probability”. “Little Amazon” has collected data on the sales of each title over the last 3 months. This data is collected in the file hw2-little-amazon.dat which is available on the assignments web page.

For all the following questions, please give the “literal” expression of the answer as well as the numeric value.

**a.** Denote by $\theta_i$ the probability that a customer buys title $i$. Assume that each purchase of a book is independent of the other purchases by the same customer or by other customers. Estimate $\theta = (\theta_0, \ldots, \theta_4)$ from the data. What are the sufficient statistics?
b. A customer buys 3 books. What is the probability that he buys “War and Peace”, “Harry Potter”, “Probability” in this order?

c. A customer buys 4 books. What is the probability that she buys only non-fiction, that is, N={3, 4}?

d. A customer buys 2 “Probability” books and 3 fiction (i.e 0 or 1 or 2) books. What is the probability of this event?

e. A customer buys n books. What is the probability that he buys at least one “Probability”?

Problem 5 – Estimate letter probabilities from text

This problem requires you to use Maximum Likelihood estimation and the smoothing methods from Lecture 1 to estimate the probabilities of the letters in the English alphabet.

We assume that sentences in a language are generated by sampling letters independently from the alphabet \{A, B, C, ... Z\}. Spaces and punctuation are ignored. For instance, the probability of the sentence “’Who’s on first?’” is

$$\theta_W \theta_H \theta_O \theta_{’s} \theta_o \theta_{’s} \theta_n \theta_f \theta_r \theta_i \theta_t \theta_t$$

because the sentence contains (W, H, O, S, O, N, ... T) in this order. You will estimate the parameters $\theta_{A:Z}$ of this simple model from the text below (also available in hw1-mlk-letter-estimation.txt).

To save man from the morass of propaganda, in my opinion, is one of the chief aims of education. Education must enable one to sift and weigh evidence, to discern the true from the false, the real from the unreal, and the facts from the fiction. [...] The function of education, therefore, is to teach one to think intensively and to think critically. But education which stops with efficiency may prove the greatest menace to society.

Martin Luther King, Jr., The Purpose of Education

First, preprocess this text: Turn all letters to lower (or upper) case, eliminate spaces and punctuation. Then proceed with the questions of the homework.

a. Get the sufficient statistics: Count the number of times each letter appears in the sentence. These are the counts $n_a, n_b, ... n_z$. Print out the counts $n_{a:j}$ only.
What is the fingerprint $r_k$, $k = 0, \ldots$ of this data set?

For the following estimation questions, choose one letter for each type (i.e., for $k = 0, 1, \ldots$ choose a letter $i$ for which $n_i = k$), and display the estimate only for those selected letters.

b. Compute the ML estimates $\theta_{A;Z}^{ML}$ of the letter probabilities.

c. Compute now one of the Laplace $\theta_{A;Z}^{Lap}$ or Bayes (with $n_0 = 1$) estimates $\theta_{A;Z}^{B}$ of the same probabilities

d. Compute the Witten-Bell estimates $\theta_{A;Z}^{WB}$ of the same probabilities.

e. Compute the smoothed\footnote{These use the approximation of $E[r_{k+1}] / E[r_k]$.} Good-Turing estimates $\theta_{A;Z}^{GT}$ of the same probabilities.

f. Compute the Ney-essen estimates $\theta_{A;Z}^{NE}$ of the same probabilities, taking $\delta = 1$.

g. Now use the estimates you obtained to compute the (log-)probability of the text in either one of hw1-test-letter-estimation.txt or hw1-test-letter-estimation-large.txt.

Print out the results obtained by each method. Which method gives the highest probability?

**Problem 6 – ML estimation – MOVED TO Homework 2**

Sam rolls a die $n$ times, and observes a data set $\mathcal{D}$ with counts $n_1, \ldots n_6$. He is told that the die is not a fair one: the odd faces have the same probability of coming up, denoted by $\theta_o$, the even faces also have the same probability of coming up, denoted by $\theta_e$, but $\theta_o \neq \theta_e$, i.e. the distribution $P$ defined by the die is given by $\theta_1 = \theta_3 = \theta_5 = \theta_o$ and $\theta_2 = \theta_4 = \theta_6 = \theta_e$.

a. Write the expression of the probability $P(3, 2, 1, 1, 6)$.

b. Write the expression of $l(\theta_o, \theta_e)$ the log-likelihood of data set $\mathcal{D}$ as a function of $\theta_o, \theta_e$ and the counts $n_{1:6}$.

c. Transform $l(\theta_o, \theta_e)$ into a function of one variable, $l(\theta_e)$.

d. Now find the ML estimate of $\theta_e$ by equating the derivative of $l(\theta_e)$ with 0.

[e–Extra credit] Explain why the result above is intuitive/not surprising/natural.