Homework 5

Due May 6, in class

1. Conditional Expectation

The conditional expectation of a random variable $X$ given an event $A$ is defined naturally to be:

$$
E[X|A] := \sum_a a \cdot P(X = a|A).
$$

(a) Show that

$$
E[X] = E[X|A]P(A) + E[X|A^c]P(A^c). \tag{1}
$$

(Hint: what does this equation remind you of?)

(b) Using (1), compute the expectation and the variance of a geometric random variable with parameter $p$.

(Hint: condition on the event that the first flip is a Tails and think recursively.)

2. Misprints

A textbook has on average one misprint per page.

(a) What is the chance that you see exactly 4 misprints on page 1?

(b) What is the chance that you see exactly 4 misprints on some page in the textbook, if the textbook is 250 pages long?

[HINT: You may assume that misprints are “rare events” that obey the Poisson distribution.]

3. Those 3407 Votes

In the aftermath of the hotly contested 2000 US Presidential Election, many people claimed that the 3407 votes cast for independent candidate Pat Buchanan in Palm Beach County were statistically highly significant, and thus of dubious validity. In this problem, we will examine this claim from a statistical viewpoint.

The total percentage votes cast for each presidential candidate in the entire state of Florida were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Gore</th>
<th>Bush</th>
<th>Buchanan</th>
<th>Nader</th>
<th>Browne</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent</td>
<td>48.8%</td>
<td>48.9%</td>
<td>0.3%</td>
<td>1.6%</td>
<td>0.3%</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

In Palm Beach County, the actual votes cast (before the recounts began) were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Gore</th>
<th>Bush</th>
<th>Buchanan</th>
<th>Nader</th>
<th>Browne</th>
<th>Others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>268945</td>
<td>152846</td>
<td>3407</td>
<td>5564</td>
<td>743</td>
<td>781</td>
<td>432286</td>
</tr>
</tbody>
</table>

To model this situation probabilistically, we need to make some assumptions. Let’s model the vote cast by each voter in Palm Beach County as a random variable $X_i$, where $X_i$ takes on each of the six possible values (five candidates or “Others”) with probabilities corresponding to the Florida percentages. (Thus, e.g., $P(X_i = \text{Gore}) = 0.488$.) There are a total of $n = 432286$ voters, and their votes are assumed to be mutually independent. Let the r.v. $B$ denote the total votes cast for Buchanan in Palm Beach County (i.e., the number of voters $i$ for which $X_i = \text{Buchanan}$).

(a) Compute the expectation $E[B]$ and the variance $\text{Var}(B)$. 
(b) Use Chebyshev’s inequality to compute an upper bound \( b \) on the probability that Buchanan receives at least 3407 votes, i.e., find a number \( b \) such that

\[
P(B \geq 3407) \leq b.
\]

Based on this result, do you think Buchanan’s vote is significant?

(c) Now suppose that your bound \( b \) in part (b) is in fact sharp, i.e., assume that \( P(B \geq 3407) \) is equal to \( b \). [In fact the true value of this probability is quite a bit smaller than \( b \).] Suppose also that all 67 counties in Florida have the same number of voters as Palm Beach County, and that all behave independently according to the same statistical model as Palm Beach County. What is the probability that in at least one of the counties, Buchanan receives at least 3407 votes? How would this affect your judgement as to whether the Palm Beach tally is significant?