ECON 5043: MANAGERIAL ECONOMICS II PROF. LE WANG

QUIZ #4

Name: ____________________________

ID (last 4 digits): ___________________

1. Suppose that a sentence S consists of n words, $w_1, w_2, \ldots, w_n$. Using what assumptions, can you reduce the probability of this sentence $\Pr(S) = \Pr(w_1, w_2, \ldots, w_n)$ to the Bigram Model of the following form?

$$\begin{aligned}\Pr(S) &= \Pr(w_1, w_2, \ldots, w_n) \\ &= \Pr(w_n|w_{n-1}) \cdot \Pr(w_{n-1}|w_{n-2}) \cdots \Pr(w_2|w_1) \cdot \Pr(w_1)\end{aligned}$$

2. Suppose that we observe the following *conditional* distribution of unemployment on job training program participation:

	Participation=1	Participation =0
Pr[Unemployment=1]	0.3	0.3
Pr[Unemployment=0]		

Table 1: Unemployment and Program Participation

(a) Fill in the empty cells.

(b) Is this job training program effective in reducing unemployment rates?

3. Suppose that we observe the following joint distribution for three variables for classification of whether an email is spam:

$Spam = 1$

	“On Sale”=1	“On Sale”=0
“Urgent”=1	0.1	0.1
“Urgent”=0	0.1	0.0

$Spam = 0$

	“On Sale”=1	“On Sale”=0
“Urgent”=1	0.0	0.2
“Urgent”=0	0.0	0.5

Table 2: Joint Distribution

What is the probability of a spam email if you observe an email contains both “on sale” and “urgent” in the subject line? Should you classify an email as a spam when you see “on sale” and “urgent” together in the subject line? Show your calculation. Note that you simply need to calculate $\Pr[\text{Spam} = 1 \mid \text{Urgent} = 1, \text{On Sale} = 1]$.