Department of Statistics
University of Wisconsin, Madison
PhD Qualifying Exam Part II
Thursday, January 24, 2013
1:00-4:00pm, Room 133 SMI

- There are a total of FOUR (4) problems in this exam. Please do a total of TWO (2) problems.
- Each problem must be done in a separate exam book.
- Please turn in TWO (2) exam books.
- Please write your code name and **NOT** your real name on each exam book.

1. Let $P_1, P_2, \cdots, P_k$ be probability measures on a measurable space $(\Omega, \mathcal{F})$, and define $Q(A) = [P_1(A) + \cdots + P_k(A)]/k$ for $A \in \mathcal{F}$.

   (a) Show that $P_1, \cdots, P_k$ are absolutely continuous with respect to Q.

   (b) Let $\mathcal{G}$ be a sub-sigma field of $\mathcal{F}$. Show that the statement:

   $$\frac{dP_j}{dQ}, j = 1, \cdots, k, \quad \text{are measurable with respect to } \mathcal{G}, \text{ a.s. } Q \tag{1}$$

   is equivalent to the statement

   $$P_j(A|\mathcal{G}) = Q(A|\mathcal{G}) \text{ a.s. } Q, \quad \text{for all } A \in \mathcal{F} \text{ and } j = 1, \cdots, k.$$

   (c) Show that (1) implies that for any sub-sigma field $\mathcal{G}_0$ of $\mathcal{G}$,

   $$P_j(A|\mathcal{G}_0) = Q(A|\mathcal{G}_0) \text{ a.s. } Q, \quad \text{for all } A \in \mathcal{F} \text{ and } j = 1, \cdots, k.$$

2. Let $X_1, X_2, ...$ be a sequence of i.i.d. random variables with a uniform distribution on $(0, 1)$. Define the events $A_1, A_2, ...$ by

$$A_n = \{X_n = \max(X_1, ..., X_n)\}.$$

(a) Prove or disprove that $A_1, A_2, ...$ are independent events.

(b) Define

$$R_n = \sum_{k=1}^{n} I(A_k),$$

where $I(A)$ denotes the indicator function for the event A. Let (m_n) be a sequence of positive numbers such that $\lim_{n\to\infty} m_n = \infty$. Compute the following limit

$$\lim_{n\to\infty} P\left(|R_n - \log n| > m_n\sqrt{\log n}\right).$$

(c) Find the asymptotic distribution of the following

$$\frac{R_n - \log n}{\sqrt{\log n}}.$$

3. The manufacturing of semiconductor (electronic) components often involves a number of steps, including "ion implantation" and "furnace annealing" applied to units called "wafers."

   An experiment was conducted to evaluate aspects of these two steps. First, for the ion implantation step, two different experimental factors were considered: the energy level used (either High or Low), and the "quantity" of ions (either High or Low) directed toward the target. Using standard notation, we describe the combinations of these two factors at their various levels as: EQ, Eq, eQ, and eq.

   The experiment was conducted as follows. Twelve wafers were selected from a a large supply of wafers and divided into 4 equal-sized batches. At random, each batch was assigned one of the treatment combinations EQ, Eq, eQ, or eq. All wafers in a batch were treated at the same time.

   The next step was furnace annealing. For the furnace annealing step, 3 different temperatures were studied: 750 C, 800 C, and 850 C. In this step, the original 12 wafers were divided into 3 annealing batches of size 4. The 3 annealing batches were randomly assigned to the three temperatures. All wafers in an annealing batch were treated at the same time.

   After this, each wafer was measured, and for our purposes we will focus on a quality measure which ranges from 0 to 100; 100 represents the highest quality.

   The entire experiment was performed twice, once in July (J) and once in August (A). In each month a new set of wafers was used, and the randomization of batches was redone.

   You may find the following output from R to be useful:

```
   quality month temp ener.quant
1       63     J  750         EQ
2       24     J  750         Eq
3       23     J  750         eQ
4       26     J  750         eq
5       65     J  800         EQ
6       24     J  800         Eq
7       25     J  800         eQ
8       26     J  800         eq
9       71     J  850         EQ
10      30     J  850         Eq
11      32     J  850         eQ
12      30     J  850         eq
13      69     A  750         EQ
14       8     A  750         Eq
15      11     A  750         eQ
16      20     A  750         eq
17      70     A  800         EQ
18      11     A  800         Eq
19      11     A  800         eQ
20      21     A  800         eq
21      84     A  850         EQ
22      24     A  850         Eq
```

```
23        23      A  850          eQ
24        31      A  850          eq


Analysis of Variance Table

Response: quality
                      Df  Sum Sq Mean Sq
month                  1   130.7   130.7
temp                   2   492.7   246.4
ener.quant             3 10524.2  3508.1
month:temp             2    58.6    29.3
month:ener.quant       3   390.3   130.1
temp:ener.quant        6    10.6     1.8
month:temp:ener.quant  6     7.4     1.2
Residuals              0     0.0
```

(a) Write down a model for this experiment, defining all terms in the model, and summarize the assumptions that accompany the model.

(b) Assume that Temperature has no effect at all. *Using only the data from July,* construct a useful plot to see whether there is evidence of an interaction between Energy and Quantity.

(c) Continue part (b). Perform a test of the null hypothesis that there is no evidence of an interaction between Energy and Quantity.

(d) Use all the data from both months and treat the four combinations of Energy and Quantity as four levels of a factor called EnQuan. Perform tests to determine whether there is evidence of a main effect for EnQuan, a main effect for Temperature, and an interaction of Temperature and EnQuan. (The factor EnQuan is coded as "ener.quant" in the R code above.)

4. Table 1 gives data from an experiment in England on the vegetative reproduction of root-stocks for plum trees from cuttings taken from the roots of older trees. The cuttings were taken between October 2007 and February 2008. Half of these were planted as soon as possible after they were taken while the other half were bedded in sand under cover until spring when they were planted. Two lengths of root cuttings were used at each planting time: long cuttings 12 cm in length and short cuttings 6 cm long. A total of 240 cuttings were taken for each of the four combinations of time of planting (at once or in spring) and length of cutting (short or long) and the condition of each plant in October, 2008 (alive or dead) was assessed.

Table 1: Survival rate of plum root-stock cuttings

Length of cutting	Time of planting	Number surviving out of 240	Proportion surviving
Short	At once	108	0.450
	In spring	30	0.125
Long	At once	156	0.650
	In spring	84	0.350

Let $x_1 = -1$ if the length of cutting was short and $x_1 = 1$ if it was long. Similarly, let $x_2 = -1$ if planting was done at once and $x_2 = 1$ if it was done in spring. Let $y = 1$ if the plant survived and $y = 0$ if it did not, and let p denote the probability that $y = 1$. Assume in this problem that the 960 cuttings were taken from 960 different parent trees.

(a) Find the maximum likelihood estimates of the regression coefficients and their standard errors in the model

$$\log\left(\frac{p}{1-p}\right) = \alpha_0 + \alpha_1 x_1$$

(b) Test at the 5% level of significance the goodness of fit of the model

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

You may use the following results.

```
Call:
glm(formula = y ~ x1 + x2, family = binomial)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.51054    0.07288  -7.005 2.47e-12 ***
x1           0.51054    0.07288   7.005 2.47e-12 ***
x2          -0.72452    0.07341  -9.870  < 2e-16 ***
---

    Null deviance: 1287.2  on 959  degrees of freedom
Residual deviance: 1135.6  on 957  degrees of freedom
AIC: 1141.6
```