Department of Statistics
University of Wisconsin, Madison
PhD Qualifying Exam Part II
September 2, 2010
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 X, Y, and Z be random variables in a probability space, and $\sigma(W)$ denote the σ-field generated by the random variable W.

   (a) Show that

   $$P(A|Y,Z) = P(A|Y) \text{ a.s. for any } A \in \sigma(X) \tag{1}$$

   is equivalent to

   $$P(B|Y,X) = P(B|Y) \text{ a.s. for any } B \in \sigma(Z).$$

   (b) Show that (1) is equivalent to

   $$E[h(X)|Y,Z] = E[h(X)|Y] \text{ a.s. for any Borel function } h \text{ with } E|h(X)| < \infty.$$

   (c) Show that (1) is equivalent to

   $$P(C|Y,g(Z)) = P(C|Y) \text{ a.s. for any } C \in \sigma(h(X)) \text{ and Borel } h \text{ and } g.$$

2. An urn contains $2n$ balls with the following composition: two are numbered 1, two are numbered 2,..., and two are numbered n. A *draw* consists of selecting two balls from the urn at random *without replacement*. On a given draw, we are interested when the two balls *match*, that is when the two balls selected have the same number. Between each draw, the balls are replaced, so the successive draws will be independent and identically distributed.

   (a) Let M_k denote the number of draws out of k total draws where the selected balls match. For any positive integer k, show

   $$P(M_k = 0) = \left(\frac{2n-2}{2n-1}\right)^k .$$

   (b) Let T be the number of draws required until the first time the selected balls match. For any $u > 0$, find $\lim_{n\to\infty} P(T > un)$.

   (c) Show that the sequence of random variables $Y_n = T/n$ converges in distribution. What is the limiting distribution?

3. A chemist wishes to compare the effects of chlorine, bromine, and iodine (halogens) on the boiling points of some alkyl halides. Any alkyl group (e.g., C_4H_9) can combine with any halogen (here `Cl`, `Br`, or `I`) to make an alkyl halide (C_4H_9Cl, C_4H_9Br, C_4H_9I). The chemist wants to study how boiling points vary. The molecular weights and boiling points are given below.

Alkyl group	Mol. weight	Boiling point (°C) Cl	Br	I
C_2H_5	29	12.5	38	72.
n-C_3H_7	43	47.	71	102.
n-C_4H_9	57	78.5	102	130.
n-C_5H_{11}	71	108.	130	157.
n-C_6H_{13}	85	134.	156	180.
n-C_7H_{15}	99	160.	180	204.
n-C_8H_{17}	113	185.	202	225.5

Let

$$
\begin{aligned}
y &= \text{boiling point} \\
x_0 &= \text{molecular weight of alkyl group} \\
x_1 &= \begin{cases} 1 & \text{chlorine halogen} \\ 0 & \text{no chlorine halogen} \end{cases} \\
x_2 &= \begin{cases} 1 & \text{bromine halogen} \\ 0 & \text{no bromine halogen} \end{cases} \\
x_3 &= \begin{cases} 1 & \text{iodine halogen} \\ 0 & \text{no iodine halogen} \end{cases}
\end{aligned}
$$

(a) Find least squares estimates of the coefficients b_0, b_1, b_2, and b_3 in the regression model

$$y = b_0x_0 + b_1x_1 + b_2x_2 + b_3x_3 + \epsilon$$

where $E(\epsilon) = 0$.

(b) Find the variance of the estimate of b_0 in terms of $\sigma^2 = E(\epsilon^2)$.

[Note: To get full credit, your answers must be accurate to within two decimals.]

4. Consider a completely randomized design with $t = 5$ treatments, $r = 6$ experimental units for each treatment, and $n = 4$ replicate observations for each experimental unit. Thus the data follow the model

$$y_{ijk} = \mu + \tau_i + \varepsilon_{ij} + \eta_{ijk}$$

($i = 1, 2, \ldots, t$; $j = 1, 2, \ldots, r$; $k = 1, 2, \ldots, n$) where ε_{ij} i.i.d. $N(0, \sigma_\varepsilon^2)$ represents the experimental unit error, η_{ijk} i.i.d. $N(0, \sigma_\eta^2)$ the observation error, and μ and $\tau_1, \tau_2, \ldots, \tau_t$ are unknown constants.

Suppose the treatments are increasing amounts (x_i) of fertilizer applied to a crop and the following (partial) results are obtained:

x_i	0	2	4	6	8
$\bar{y}_{i\cdot\cdot}$	4.9	10.0	13.9	15.7	16.3

$$n \sum_{i=1}^{t} \sum_{j=1}^{r} (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot})^2 = 50, \quad \sum_{i=1}^{t} \sum_{j=1}^{r} \sum_{k=1}^{n} (y_{ijk} - \bar{y}_{ij\cdot})^2 = 60$$

where $\bar{y}_{ij\cdot} = n^{-1} \sum_{k=1}^{n} y_{ijk}$ and $\bar{y}_{i\cdot\cdot} = (nr)^{-1} \sum_{j=1}^{r} \sum_{k=1}^{n} y_{ijk}$.

Use the orthogonal polynomials given below to investigate whether

(a) the data exhibit linear and quadratic trends;

(b) first and second order terms provide an adequate fit to the data.

Fact: Orthogonal polynomials $P_l(z)$ are, apart from a constant factor, Chebyshev polynomials. For t treatments, they are, up to constant multiples,

$$\begin{aligned}
P_1(z) &= z \\
P_2(z) &= z^2 - (t^2 - 1)/12 \\
P_3(z) &= z^3 - (t^2 - 7)z/20 \\
P_4(z) &= z^4 - (3t^2 - 13)z^2/14 + 3(t^2 - 1)(t^2 - 9)/560.
\end{aligned}$$

[Note: To get full credit, your answers must be accurate to within two decimals.]