2 1999 Exam

Question 1

(a) The linear mixed model includes more than 1 random component (eg effect due to subject as well as measurement error) and is defined by:-

$$\begin{aligned}
Y_{n,1} &= X_{n,p}\beta_{p,1} + Z_{n,q}U_{q,1} + E_{n,1} \quad , \\
E &\sim N(0, V_e), \quad V_e = \sigma_e^2 \times I_p, \\
U &\sim N(0, V_u) \quad V_u = \sigma_u^2 \times I_q, \\
\text{covariance}(E, U) &= 0 \ .
\end{aligned}$$

(i) What is the expected value of Y? *[2 marks]*

(ii) What is the density of $(Y|U)$? *[2 marks]*

(iii) What is the joint density of (Y, U)? *[2 marks]*

(b) If $Y \sim N(0, \sigma^2 I_p)$ and

$$\begin{aligned}
Y^T Y &= \sum_{i=1}^{k} Y^T B_i Y \quad , \\
\text{rank}(B_i) &= r_i, \quad \sum_{i=1}^{k} r_i = p \\
B_i B_j &= 0, \quad \forall i, j \text{ where } i \neq j \quad ,
\end{aligned}$$

what is the distribution of $Y^T B_i Y$? *[3 marks]*

(c) If

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 3 \\ 7 \end{bmatrix}, \begin{bmatrix} 2 & \\ 4 & 12 \end{bmatrix} \right) ,$$

what is the density function for $x_2|x_1 = 4$? *[3 marks]*

Question 2

(a) For a sample of size n (n being an odd integer) of a continuous random variable x whose density and distribution functions are $f(x)I_{(a,b)}(x)$ and $F(x)$ respectively, write formulae for the densities of

(i) the nth order statistic, *[2 marks]*

(ii) the 1st order statistic, *[2 marks]*

(iii) the median *[2 marks]*

(b) Loads endured by a cable are assumed to be from a Weibull distribution with c.d.f

$$F(x) = 1 - \exp(-\lambda x^{\alpha}) \quad , x \in (0, \infty) \ .$$

A sample of loads was

$$\{6, 6, 7, 7, 7, 8, 8, 9, 10, 12, 17\} \ .$$

and the parameters were estimated as $\hat{\alpha} = 2$, $\hat{\lambda} = 0.1$.

Use the information in this sample to derive formulae, without evaluating the integrals) for calculating the following probabilities:-

(i) the maximum load is at least 15, *[4 marks]*

(ii) the minimum load is no more than 6, *[4 marks]*

(iii) the median load is between 7 and 9. *[4 marks]*

Question 3

(a) Let X and Y be independent $N(0, 1)$ normal random variables. The pair (X, Y) defines a point in two dimensions by Cartesian coordinates. The transformation to polar-coordinates is given by

$$\begin{aligned} X &= R\cos\theta \\ Y &= R\sin\theta \end{aligned}$$

(i) What is the joint density $f_{R\theta}(r, \theta)$. *[3 marks]*

(ii) What are the marginal densities $f_R(r)$ and $f_\theta(\theta)$. *[6 marks]*

(iii) What are the ranges for r and θ? *[2 marks]*

(b) Suppose X_1 and X_2 are independent exponential random variables, each with p.d.f. $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. Let $S = X_1 + X_2$. What is the density of S. *[4 marks]*

(c) Let X be a random variable with density

$$f_X(x) = C \cdot \left(1 - \frac{x^2}{5}\right) \quad -\sqrt{5} \leq x \leq \sqrt{5}$$

where C is a constant.

Find C. *[3 marks]*

Question 4

(a) The exponential family of distributions are those distributions for which the density function can be written in the form,

$$f(x;\theta) = B(\theta)h(x)\exp\{p(\theta)k(x)\}I_{(a,b)}(x)I_{(\gamma,\delta)}(\theta).$$

By expanding the qadratic term in the exponent and dividing each term by the divisor, writing the exponential term as the product of 3 separate exponential terms show the inverse Gaussian distribution,

$$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}}e^{-\lambda(x-\mu)^2/2\mu^2 x}I_{(0,\infty)}(x)$$

is a member of the exponential family by rewriting the density in the above form and identifying each component. (λ is a scale parameter and the mean is a function of μ.) *[5 marks]*

(b) State the result for finding a sufficient statistic for a parameter from the exponential family and give a sufficient statistic for μ in the Inverse gaussian distribution? *[3 marks]*

(c) A sample $X_1, \cdots X_n$ is taken from $f(x;\theta)$, and $t(\mathbf{x})$ is an estimate of θ. State the factorization theorem which would ascertain if $t(x)$ was a sufficient statistic for θ. *[3 marks]*

(d) Show that the sum of observations from a random sample of size n from a gamma distribution has the p.d.f.

$$f(x;\theta) = \frac{1}{\theta}e^{-x/\theta} \cdot I_{(0,\infty)}(x) \cdot I_{(0,\infty)}(\theta)$$

is a sufficient statistic for θ. *[5 marks]*

Question 5

(a) A nonlinear model relating observations, Y to covariates X is given by

$$Y = \alpha x^\beta + \epsilon \quad , \epsilon \sim N(0,\sigma^2) \ .$$

(i) What is $E[Y|X]$? *[1 mark]*

(ii) What is $\text{var}[Y|X]$? *[1 mark]*

(iii) What is the log-likelihood of $(y,\alpha,\beta|x)$ for a sample $(x_1,y_1)\ldots(x_n,y_n)$? *[3 marks]*

(iv) What is the general form of the information matrix for (α,β)? *[3 marks]*

(b) Let $Y_1 < Y_2 < Y_3$ be the order statistics from a sample of size 3 from the uniform distribution,

$$f(x) = \frac{1}{\theta} \cdot I_{(0,\theta)}(x) \cdot I_{(0,\infty)}(\theta) \ .$$

Show that the following are unbiased estimators of θ.

(i) $4Y_1$ *[4 marks]*

(ii) $2Y_2$ *[4 marks]*

(iii) $\frac{4}{3}Y_3$ *[4 marks]*

Question 6 A sample of i.i.d. data, $(x_1, y_1) \ldots (x_n, Y_n)$ may be represented by either

$$\begin{aligned} y_i &= a + bx_i + \epsilon_i \text{ or} \\ y_i &= a + bx_i + cx_i^2 + \epsilon_i \end{aligned}$$

where $\epsilon_i \sim N(0, \sigma^2)$ in both cases.

The data analyst has to decide whether the quadratic term should be included in the model.

(a) What is the null hypothesis to be tested for making the above decision? *[2 marks]*

(b) Derive a generalized likelihood test to assess the significance of the quadratic term. *[10 marks]*

(c) If the sample size n is large, how would you assess the probability that the quadratic term should or should not be included in the model? *[4 marks]*

FORMULAE

a. If $U = X + Y$, $\quad f_U(u) = \int_{-\infty}^{+\infty} f_X(u-v) f_Y(v) dv$.

b. $f_{Y_r}(y_r) = \frac{n!}{(n-r)!(r-1)!} f(y_r) F(y_r)^{r-1} \left[1 - F(y_r)\right]^{n-r}$.

c. $(x_2|x_1) \sim N(\mu_{2.1}, \Sigma_{22.1})$
$\mu_{2.1} = \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)$
$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$