Instructions:

1. Do not turn this page until instructed to do so.

2. Start each new question on a new sheet of paper.

3. This is a closed book exam.

4. Show your working and explain your reasoning as fully as possible.

5. Switch off all electronic devices.

6. The time limit is 3 hours.

7. The exam has 6 questions – do as much as you can but credit will be given for only 5 answers.
**Question 1.** Consider the linear regression model given by

\[ y = X\beta + \sigma \varepsilon, \]

where \( y \) is a \( n \times 1 \) column vector, consisting of \( (y_1, \ldots, y_n) \), \( X = (x_{ij}) \) is a \( n \times p \) design matrix, \( \beta \) is a \( p \times 1 \) unknown column vector of coefficients, \( \sigma \) is a known standard deviation and the \( \varepsilon \) is a \( n \times 1 \) vector of independent standard normal random variables.

A ridge estimator for \( \beta \) is of the form

\[ \hat{\beta}_R = (X'X + \lambda I)^{-1} X'y \]

for some \( \lambda > 0 \).

(i) Show, by considering a normal prior distribution for \( \beta \), that the ridge estimator can be seen as a Bayes estimator.

(ii) Find the variance–covariance matrix of \( \hat{\beta}_R \).

(iii) If \( p = 1 \), and \( \sum_{i=1:n} x_i^2 = 1 \), find

\[ E[(\hat{\beta}_R - \beta)^2]. \]

(iv) In the case just looked at, what is the variance of the maximum likelihood estimator \( \hat{\beta} \)?

(v) Are there values of \( \lambda \) for which

\[ E[(\hat{\beta}_R - \beta)^2] \leq E[(\hat{\beta} - \beta)^2], \]

and if so what is the problem with using such a value for \( \lambda \)?
**Question 2.** The density function for random variable $X$ has exponential family form:

$$f(x; \theta) = c(x) \exp\{x\theta - b(\theta)\}, \quad x \in \mathcal{X},$$

with $\mathcal{X}$ being a subset of the real line.

(i) By considering the first two moments of $X$, or otherwise, prove that $b(\theta)$ is a convex function.

(ii) Find the maximum likelihood estimator of $\theta$, written as $\hat{\theta}$, based on a sample of size $n$; i.e. $X_1, \ldots, X_n$.

(iii) Find $E b'(\hat{\theta})$ and $\text{Var} b'(\hat{\theta})$, where $b'$ is the first derivative of $b$.

(iv) Suppose $T(X_1, \ldots, X_n)$ is an estimator of $\theta$ for which

$$E T(X_1, \ldots, X_n) = b'(\theta).$$

Show that

$$E \{T(X_1, \ldots, X_n) \bar{X}\} = b''(\theta)/n + \{b'(\theta)\}^2.$$

(v) Following on from part (iv), show that

$$\text{Var} T(X_1, \ldots, X_n) \geq b''(\theta)/n$$

and what is the significance of this result.
Question 3. Suppose \( \pi(x) \) is a target density for sampling using a Markov chain. The transition density taking \( x \) to \( y \) is written as \( p(y \mid x) \). A necessary condition for \( \pi \) to be the stationary density is that

\[
\pi(y) = \int p(y \mid x) \pi(x) \, dx.
\]

(i) Show that the stationary condition is satisfied if the reversible condition

\[
p(y \mid x) \pi(x) = p(x \mid y) \pi(y)
\]

is satisfied for all \( x \) and \( y \).

(ii) If \( q(y \mid x) \) is an arbitrary conditional density, known as a proposal density; what is the condition on \( \alpha(x, y) \) for

\[
p(y \mid x) = \alpha(x, y) q(y \mid x) + (1 - r(x)) \mathbf{1}(y = x)
\]

to satisfy the reversible condition with respect to \( \pi \). Here

\[
r(x) = \int \alpha(x, y) q(y \mid x) \, dy.
\]

(iii) Show that the reversible condition is satisfied if

\[
\alpha_M(x, y) = \min \left\{ 1, \frac{\pi(y) q(x \mid y)}{\pi(x) q(y \mid x)} \right\}
\]

(iv) An alternative \( \alpha \) is given by

\[
\alpha_B(x, y) = \frac{\pi(y) q(x \mid y)}{\pi(y) q(x \mid y) + \pi(x) q(y \mid x)}.
\]

Show that \( r_M(x) \geq r_B(x) \) for all \( x \) and what are the implications of this result?

(v) Show that by taking a \( y^* \) from \( q(\cdot \mid x) \) and setting \( y = y^* \) with probability \( \alpha(x, y^*) \) or \( y = x \) with probability \( 1 - \alpha(x, y^*) \), the \( y \) is a sample from \( p(\cdot \mid x) \).
**Question 4.** Let \( f(x \mid \theta) \) be a family of density functions indexed by \( \theta \in \Theta \) and \( x \in X \subset (-\infty, +\infty) \). Define

\[
l(\theta; x) = -\log f(x \mid \theta)
\]

and suppose \( l(\theta; x) \) is strictly convex in \( \theta \) for all \( x \); and let \( l'(\theta; x) = \partial l(\theta; x) / \partial \theta \).

(i) From a sample \((x_1, \ldots, x_n)\) of size \( n \) from \( f(\cdot \mid \theta) \), let \( \hat{\theta} \) be the maximum likelihood estimator of \( \theta \). Show that

\[
P(\hat{\theta} \leq z) = P(l'(z) \geq 0)
\]

where

\[
l'(z) = \sum_{i=1}^{n} l'(z; x_i).
\]

(ii) Assuming a central limit theorem holds for \( l'(z)/n \), for all \( z \), find an approximate distribution for \( \hat{\theta} \).

(iii) Suppose \( f(x \mid \theta) = \theta e^{-x^\theta} \), with \( x > 0 \) and \( \theta > 0 \). Show that \(-\log f(x \mid \theta)\) is convex for all \( \theta \).

(iv) Find the normal approximation to \( l'(z)/n \).

(v) Following from (iii) and (iv), find an approximate distribution for \( \hat{\theta} \).
**Question 5.** Suppose $X(t)$, for $t \geq 0$, is a discrete valued Markov process in continuous time with $X(0) = 0$. For any $n \geq 0$, and small $h$,

$$P(X(t + h) = n + 1 | X(t) = n) = h\lambda_n + o(h),$$

$$P(X(t + h) = n | X(t) = n) = 1 - h\lambda_n + o(h),$$

and

$$P(X(t + h) > n + 1 | X(t) = n) = o(h).$$

Define $p_t(n) = P(X(t) = n)$.

(i) Show that $p_t'(0) = -\lambda_0 p_t(0)$ and $p_t'(n) = -\lambda_n p_t(n) + \lambda_n p_t(n - 1)$ for $n = 1, 2, \ldots$, where $'$ denotes differentiation with respect to $t$.

(ii) Find the matrix $G = (g_{ij})_{0 \leq i,j < \infty}$ such that

$$P_t' = P_t G,$$

where $P_t' = (p_t'(0), p_t'(1), \ldots)$ and $P_t = (p_t(0), p_t(1), \ldots)$.

(iii) If $\lambda_n = \lambda$ for all $n$; show that the solution to the differential equation in (i) and (ii) is

$$p_t(n) = \frac{(t\lambda)^n}{n!} e^{-t\lambda}.$$

(iii) Explain why $p_0(0) = 1$.

(iv) Show that the solution in (iii) can be written as

$$p_t(n) = [\exp(tG)]_n,$$

where $\exp(tG)$ is the matrix

$$\sum_{l=0}^{\infty} \frac{t^l G^l}{l!},$$

and pay special attention to the meaning of $G^0$, and $[.]_n$ denotes the $(0, n)$ element of the matrix.
Question 6. Suppose $p(x_t|x_{t-1})$ is a conditional density for a first order time series $(x_t)$. Further suppose, for some copula density $c(u,v)$,

$$p(y|x) = f(y) c(F(y), F(x))$$  \hspace{1cm} (1)

for some distribution function $F$ with corresponding density function $f$.

Recall a copula density is a density on $(0,1)^2$ with marginals the uniform density on $(0,1)$. Note that if $C(u,v)$ is the copula distribution function then

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v}.$$

(i) Show that $f(x)$ is the stationary density for the process $(x_t)$.

(ii) Find the form for the distribution function

$$P(x_t \leq y | x_{t-1} = x).$$

(iii) If

$$C(u,v) = \frac{uv}{1-\rho(1-u)(1-v)}$$

for some $0 \leq \rho < 1$, find the distribution function in (ii).

(iv) If $\rho = 0$, what can you say about the process?

(v) If $(x_t)$ is a first order process with stationary density $f(x)$, show that it is possible to write the conditional transition density $p(y|x)$ as (1) for some copula density $c(u,v)$. 

7
A growth model for a rare event over time behaves as

\[ g(t) = a + b e^{ct}, \]

where \( t > 0 \) indicates time. A statistical model for individual \( i \) with observed growth \( y_{ij} \) at time points \( (t_j) \) is given by

\[ y_{ij} = \mu + z_i e^{\theta_i t_j} + \sigma \varepsilon_{ij}, \quad i = 1, \ldots, 1000, \quad j = 1, \ldots, 10. \]

Here \( \mu \) is a grand mean and \((z_i, \theta_i)\) are individual effects; \( \sigma \) denotes an error and the \((\varepsilon_{ij})\) are assumed zero mean and unit variance independent variables.

Most of the \((\theta_i)\) are sparse; i.e. all but a few of them are 0, so most of the individuals experience no growth. For those that do experience growth, the \( \theta_i \) is positive.

The data are \((y_{ij}, t_j)\), for \( i = 1, \ldots, 1000 \) and \( j = 1, \ldots, 10 \), and the \( t_j \) are fixed and \( t_j = j/10 \). The goal of the data analysis is to estimate the set

\[ S = \{ i : \theta_i > 0 \}. \]

Whatever \( \hat{S} \) you come up with, also provide some notion of uncertainty with it; and for those \( i \in \hat{S} \), provide estimates of the \( \theta_i \), again with some notion of associated uncertainty.