INDIAN STATISTICAL INSTITUTE

Semestral Examination: (2019–2020)

M. Stat 2nd Year

Statistical Computing - I

Date:......... Full Marks: ..100.. Duration: ..3 hours..

Attempt all questions

1. (i) Suppose that $g(y|\theta)$ and $f(x|\theta)$ are the probability densities of the observed and complete data, respectively. Then show that the EM iterates obey

$$\log g(y|\theta_{n+1}) \geq \log g(y|\theta),$$

with strict inequality when $f(x|\theta_{n+1})/g(y|\theta_{n+1})$ and $f(x|\theta_n)/g(y|\theta_n)$ are different conditional densities or when the surrogate function $Q(\theta|\theta_n)$ satisfies $Q(\theta_{n+1}|\theta_n) > Q(\theta_n|\theta_n)$.

(ii) Suppose J items $Y_{i1}, \ldots, Y_{iJ}$ are recorded on each of $i = 1, \ldots, n$ test takers, where Y_{ij} are independent Bernoulli random variables with probability $P(Y_{ij} = 1 \mid \boldsymbol{\beta}_i) = \Phi(\mathbf{X}_j'\boldsymbol{\beta}_i)$; $j = 1, \ldots, J$. Here Φ denotes the cumulative distribution function of $N(0, 1)$ distribution, $\boldsymbol{\beta}_i$ is a $p \times 1$ vector, called the ability vector, signifying the test taker's proficiency in each attribute (item); $\mathbf{X}_j = (x_{j1}, \ldots, x_{jp})'$ is a vector of known binary covariates flagging the attributes tested in the j-th test item. Assume that $\boldsymbol{\beta}_i$ are *iid* p-variate normally distributed with mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$. Propose a suitable method based on EM algorithm to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

[10+15=25]

2. (i) In the context of maximum likelihood estimation in multinomial distribution show that your MM algorithm converges to the maximum likelihood estimate at a linear rate.

(ii) Multidimensional scaling attempts to represent q objects as faithfully as possible in p-dimensional space giving a weight $w_{ij} > 0$ and a dissimilarity measure y_{ij} for each pair of objects i and j. If $\boldsymbol{\theta}_i \in \mathbb{R}^p$ is the position of object i, then the $p \times q$ parameter matrix Θ with i-th column $\boldsymbol{\theta}_i$ is estimated by minimizing the stress

$$\sigma^2(\Theta) = \sum_{1 \leq i < j \leq q} w_{ij}(y_{ij} - \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|)^2,$$

where $\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|$ is the Euclidean distance between $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$. Construct an MM algorithm to estimate Θ. With reasons state any assumptions that you make.

[10+15=25]

3. Let V be the collection of functions f with $f'' \in L^2[0,1]$. Consider the subspace

$$W_2^0 = \{f(x) \in V : f, f' \text{ absolutely continuous and } f(0) = f'(0) = 0\}.$$

Define an inner product on W_2^0 as

$$< f, g >= \int_0^1 f''(t)g''(t)dt.$$

(i) Show that W_2^0 is a reproducing kernel Hilbert space.

(ii) Hence, obtain the reproducing kernel of W_2^0.

[15+10=25]

4. For some sequence $\{\theta_n : n = 1, 2, \ldots\}$, consider the following generalized accept-reject method:

At iteration n ($n \geq 1$)

(a) Generate $X_n \sim g_n$ and $U_n \sim Uniform(0, 1)$, independently.

(b) If $U_n \leq \theta_n f(X_n)/g_n(X_n)$, accept $X_n \sim f$;

(c) Otherwise, move to iteration $n + 1$.

Let Z be the random variable denoting the output of this algorithm.

(i) Show that Z has the cumulative distribution function

$$P(Z \leq z) = \sum_{n=1}^{\infty} p_n \int_{-\infty}^{z} f(x)dx$$

where $p_1 = \theta_1$ and $p_n = \theta_n \prod_{m=1}^{n-1}(1 - \theta_m)$ for $n = 2, 3, \ldots$.

(ii) Show that

$$\sum_{n=1}^{\infty} p_n = 1 \text{ if and only if } \sum_{n=1}^{\infty} \log(1 - \theta_n) \text{ diverges.}$$

[5+5=10]

5. Assume that the Markov transition kernel $P(x, A)$ satisfies a minorization condition on a small set C, in addition to the regularity conditions necessary for convergence to the target distribution π. Assume that two different chains are run, one started from an arbitrary fixed starting value x_0, and another started by drawing the initial value y_0 from the invariant distribution.

(i) Derive the coupling inequality

$$\|P^n(x_0,\cdot) - \pi(\cdot)\| = \sup_{A\in\mathcal{B}} |P^n(x_0,A) - \pi(A)| \leq Pr(T > n)$$

where $P^n(x_0, A)$ is the probability of hitting the set A in n iterations, beginning at x_0. T is the time at which coupling occurs and $\mathcal{B}$ is the appropriate Borel σ-algebra.

(ii) Using the coupling inequality, show that if the entire state space is small, then the Markov chain is uniformly ergodic.

[10+5=15]

INDIAN STATISTICAL INSTITUTE

Back Paper Examination: (2019–2020)

M. Stat 2nd Year

Statistical Computing

Date:..8/01/20.. Marks: ..100.. Duration: .3 hours.

Attempt all questions

1. (a) Let $x_1, \ldots, x_m$ be *iid* observations drawn from a mixture of two normal densities with means μ_1 and μ_2 and common variance σ^2, and let π and $1-\pi$ be the corresponding mixing proportions. Discuss an EM algorithm to estimate the parameters $(\mu_1, \mu_2, \sigma^2, \pi)$, up to identifiability.

(b) Suppose that each team i in a league of K teams is assigned a rank parameter $\theta_i > 0$. Assuming ties are impossible, suppose further that team i beats team j with probability $\theta_i/(\theta_i + \theta_j)$. If this outcome occurs y_{ij} times during a season of play, then devise an MM algorithm to estimate the rank parameters, stating clearly any further assumptions that you make.

[10+15=25]

2. (a) Suppose that the function $f(x)$ is twice continuously differentiable and $s(x)$ is the spline interpolating $f(x)$ at the nodes $x_0 < x_1 < \cdots < x_n$. If $h = \max\limits_{0 \le i \le n-1} (x_{i+1} - x_i)$, then prove that

(i) $\max\limits_{x_0 \le x \le x_n} |f(x) - s(x)| \le h^{\frac{3}{2}} \left[\int_{x_0}^{x_n} f''(y)^2 dy\right]^{\frac{1}{2}}.$

(ii) $\max\limits_{x_0 \le x \le x_n} |f'(x) - s'(x)| \le h^{\frac{1}{2}} \left[\int_{x_0}^{x_n} f''(y)^2 dy\right]^{\frac{1}{2}}.$

(b) Consider all linear functionals over $\mathbf{x} \in \mathbb{R}^p$ passing through the origin

$$H = \{f_{\boldsymbol{\theta}} : f_{\boldsymbol{\theta}}(\mathbf{x}) = \boldsymbol{\theta}'\mathbf{x}, \boldsymbol{\theta} \in \mathbb{R}^p\}.$$

Define

$$< f_{\boldsymbol{\theta}}, f_{\boldsymbol{\lambda}} >= \boldsymbol{\theta}'\boldsymbol{\lambda}.$$

Obtain the reproducing kernel of H.

[(5+5)+15=25]

3. (a) Consider a density of the form

$$f(x) \propto \exp\left(-\frac{x^2}{2}\right)\left\{\sin^2(6x) + 3\cos^2(x) + \sin^2(4x) + 1\right\}$$

Develop a rejection sampling algorithm for simulating *iid* random variates from the above density.

(b) Describe the adaptive rejection sampling (ARS) algorithm and prove that the random variates generated using the ARS algorithm have the desired target distribution.

[10+15=25]

4.

(a) Suppose that $X = \{1, 2, 3\}$, with $\pi\{1\} = \pi\{2\} = \pi\{3\} = 1/3$. Consider a Markov chain defined by $P(1, \{1\}) = P(1, \{2\}) = P(2, \{1\}) = P(2, \{2\}) = 1/2$, and $P(3, \{3\}) = 1$. In the above, $P(i, \{j\})$ denotes the one-step transition probability from state i to state j.

(i) Show that $\pi(\cdot)$ is stationary for the above Markov chain. Is it the unique stationary distribution?

(ii) Show that the Markov chain *does not* converge to the above stationary distribution π. Which convergence condition is violated here?

(b) Suppose again, that $X = \{1, 2, 3\}$, with $\pi\{1\} = \pi\{2\} = \pi\{3\} = 1/3$. Let $P(1, \{2\}) = P(2, \{3\}) = P(3, \{1\}) = 1$.

(i) Show that $\pi(\cdot)$ is stationary, and the Markov chain is irreducible.

(ii) Show that the chain *does not* converge to π. Which convergence condition is violated in this case?

[(6+6$\frac{1}{2}$)+(6+6$\frac{1}{2}$)= 25]

INDIAN STATISTICAL INSTITUTE

M.STAT. (2nd Year), 2019-2020

Semestral Examination

Subject: Time Series Analysis

Date:29.11.2019 F.M.-60 Duration: 3hrs

Attempt All Questions

1. Derive spectral density of the following time series and relate it with their properties.
   a) White Noise $(0,\sigma^2)$
   b) Moving Average process of order 1.
   c) Auto-regressive process of order 1. [4+4+4]

2. a) Explain time series with discrete spectra and time series with continuous spectra . Give example of each.
   b) Check by method of spectral analysis , if the following functions are Auto-covariance functions of some time series-
   (i) $\gamma(h) = 3$ if $h=0$
   $= 2/3$ if $|h| =1$
   $= 0$ o.w.
   (ii) $\gamma(h) =1$ if $h=0$
   $=0.2$ if $|h|=1$
   $=0.6$ if $|h| =2$
   $= 0$ o.w. [6+6]

3. a) $\gamma(h)$ is autocovariance function of some stationary time series with mean μ. $\overline{X}_n$ = mean of 1st n observations.
   Show that it is consistent for μ, if $\gamma(h)\to 0$,as $h\to \infty$
   b) Let $\{X_n\}$ be a Moving Average process of order 1 with equation
   $X_n= \theta Z_{n-1} +Z_n+\mu$, where $\{Z_n\} \sim$ IID (0,2) and it is given that $\theta=1$.
   Derive 95% confidence interval for μ, where $\overline{x}_n= 1.20$. [6+6]

4. Let $\{X_t\}$ be an Auto-regressive process of order 1 with equation

   $X_t = \varphi X_{t-1} + Z_t$ where $\{Z_t\} \sim WN(0,\sigma^2)$ and $|\varphi| \neq 1$,

   with $cov(Z_t, X_s) = 0$ for all $s<t$.

   For different values of φ, express $\{X_t\}$ as a linear process in $\{Z_t\}$ and show convergence of the process in root-mean square norm and almost sure convergence. [12]

5. a) State Wold Decomposition Theorem for a stationery time series. Discuss about deterministic and non-deterministic components of a stationary process in the decomposition.

   b) Determine the deterministic and non-deterministic components of the following series.

   (i) $X_t = U + Y_t$ where $\{Y_t\} \sim$ iid $N(0,1)$ and $U \sim N(0,1)$ independent of $\{Y_t\}$.

   (ii) $X_t = A \cos t + B \sin t$ where A, B are iid $N(0,1)$. [6+6]

XXXXXXXXXXXXXXXXXXXXXXXXX

INDIAN STATISTICAL INSTITUTE

M.STAT. (2nd Year), 2019-2020

Supplementary/Backpaper Examination

Subject: Time Series Analysis

Date:13.01.2020 F.M.-100 Duration: 3hrs

Attempt All Questions

1. a) Define and discuss with example all the four components of a time series.
   b) Discuss how trend component is estimated by method of curve fitting. [16+4]

2. Define ARMA(1,1) process $\{X_t\}^{\infty}_{t=-\infty}$.

   Write down X_t as a linear process in $\{Z_t\}$, the White Noise used to define the process.

   Discuss for what values of the parameters,the process is causal and for what values it is invertible. [2+8+10]

3. a) Derive $P_n(X_{n+1})$, the best linear predictor of X_{n+1} by X_n , X_{n-1} , ……, X_1, 1 for the stationary time series$\{X_n\}$.
   b) Show that $P_n(X_{n+1})$ is unique as a random variable. [14+6]

4. a) Discuss the result expressing the asymptotic distribution of sample autocorrelation function using Bartlett's formula.
   b) Using the above in (a) discuss how we can get confidence interval for $\rho(1)$ and $\rho(2)$ in a Moving Average process defined by $X_t = \theta Z_{t-1} + Z_t$ with $\{Z_t\} \sim IID(0,\sigma^2)$ and θ unknown, on the basis of observed x_1, x_2, ……, x_n . [10+10]

P.T.O.

5. a) State and prove the result where by using spectral analysis we can check if a function is Auto-covariance function of a time series or not.

b) Using (a) or otherwise check for which values of ρ, the following function is autocovariance function of some time series.

$$\gamma(h) = \begin{cases} 1 & \text{if } h=0 \\ \rho & \text{if } |h|=1 \\ 0 & \text{o.w.} \end{cases}$$

[13+7]

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

INDIAN STATISTICAL INSTITUTE
Semester Examination: 2019 – 20

M. Stat 2nd Year
Martingale Theory

Date: 18/11/2019 Maximum Marks: 60 Duration: 3hr

Try to write brief and precise answers.

1. **State true or false with reasons.**

   (a) Toss fair coins independently. The expected time of the first appearance of HTH-HTH is $2^6 + 2^3$.

   (b) Let $(S_n : n \geq 0)$ be the simple symmetric random walk on $\mathbb{Z}$. Then $E = \{S_n = 0 \text{ finitely often}\}$ is exchangeable and hence $\mathsf{P}(E) = 0$.

   (c) Let $\{X_i : i \geq 1\}$ be an iid sequence with $P(X_1 = 0) = P(X_1 = 2) = 0.5$. Let
   $$M_n = \prod_{i=1}^{n} X_i.$$
   Then $(M_n : n \geq 1)$ is uniformly integrable.

   (d) Suppose $f : \mathbb{Z}^2 \to [0, \infty)$ satisfies the following
   $$4f(x,y) = f(x+1,y) + f(x,y+1) + f(x-1,y) + f(x,y-1), \ \forall x, y \in \mathbb{Z}.$$
   Then f is constant.

   [$4 \times 5 = 20$]

2. **Answer any two:**

   (a) Let $X_1, X_2, \cdots, X_n$ be an exchangeable sequence of random variables with finite second moment and $\sigma^2 = \mathsf{Var}(X_i)$. Show that
   $$\text{Cov}(X_i, X_j) \geq -\frac{\sigma^2}{n-1}.$$
   What can you say about an infinite sequence of exchangeable random variables?

   (b) Consider the Polya urn with 1 red ball and 1 black ball, initially. A ball is drawn from the bag, its colour noted, and then returned to the bag together with a new ball of the same color. Let T be the number of balls until the first black ball appears. Show that
   $$\mathsf{E}\left[\frac{1}{T+2}\right] = \frac{1}{4}.$$

(c) Consider $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \otimes P_2)$ and $\mathcal{C} = \{Z \times \Omega_2 : Z \in \mathcal{A}_1\}$. Show that

$$Q^{\mathcal{C}}((\omega_1, \omega_2), M) = P_2(M_{\omega_1})$$

is an RCP (regular conditional probability) where $M_{\omega_1} = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in M\}$.

[$2 \times 5 = 10$]

3. Let $(\Omega, \mathcal{A}, P)$ be a probability space and $(\mathcal{F}_n)_{n\geq 0}$ be a filtration. Let $(X_n)_{n\geq 0}$ be an adapted process such that $X_n \in L^1$ for all $n \geq 0$. Suppose $\mathsf{E}[X_T] = \mathsf{E}[X_0]$ for any bounded stopping time T. Show that $(X_n)_{n\geq 0}$ is a martingale. [10]

4. Let $X_1, X_2, \cdots, X_n$ be independent standard Normal random variables and $g : \mathbb{R}^n \to \mathbb{R}$ be continuous and Lipschitz in each variable separately with Lipschitz constant 1. Set

$$Y = g(X_1, \cdots, X_n).$$

Using Azuma-Hoeffding inequality show that

$$P\left(|Y - \mathsf{E}[Y]| \geq t\right) \leq 2e^{-2t^2/n}.$$

Hint: To see the martingale differences, use an independent copy of X_k.

[10]

5. Let $(Y_{in} : i \in \mathbb{Z}, n \in \mathbb{N})$ be a collection of independent $\mathbb{Z}$-valued random variables with the following property:

$$\begin{aligned} Y_{in} &\stackrel{d}{=} Y_{i1}, \text{ for all } i, n\,, \\ \mathsf{E}(Y_{i1}) &= 0,\ i \in \mathbb{Z}\,, \\ \mathsf{E}(Y_{i1}^2) &= 1,\ i \in \mathbb{Z}\,, \\ \sup_{i\in\mathbb{Z}} \mathsf{E}[|Y_{i1}|^{2+\delta}] &< \infty \text{ for some } \delta > 0\,. \end{aligned}$$

Let $(X_n : n \geq 0)$ be a Markov chain with state space $\mathbb{Z}$ and $X_0 = 0$ evolving as follows:

$$X_{n+1} = X_n + Y_{X_n,n}\,.$$

Show that

$$\frac{X_n}{\sqrt{n}} \xrightarrow{d} N(0, 1).$$

[10]

Indian Statistical Institute. Kolkata
Final examination: Second semester 2019 - '20
M. Stat II year

Subject: **Functional Analysis** Date: **November 27, 2019**
Duration: **3 hours** Time: **2:30 PM to 5:30 PM**
Maximum score: **60**

Attempt all the problems. Justify every step in order to get full credit of your answers. All arguments should be clearly mentioned on the answerscript. Points will be deducted for missing or incomplete arguments.

(1) Let H be a Hilbert space and $T \in \mathcal{L}(H)$ such that $0 \notin \sigma(T)$. Show that $\sigma(T^{-1}) = \{\lambda^{-1} : \lambda \in \sigma(T)\}$.

[5 marks]

(2) Let K be an infinite compact subset of $\mathbb{C}$ and $H = l^2$. Construct a $T \in \mathcal{L}(H)$ such that $\sigma(T) = K$.

[5 marks]

(3) Consider the unilateral shift operator $S : l^2 \to l^2$ defined by

$$(S\underline{x})_n := \begin{cases} 0 & \text{if } n = 1 \\ x_{n-1} & \text{otherwise} \end{cases} \qquad \text{for } \underline{x} \in l^2.$$

(a) Evaluate S^*.
(b) Show that $\sigma_p(S^*)$ contains the open unit disc $\mathbb{D}$ in $\mathbb{C}$.
(c) Show that $\sigma(S) = \overline{\mathbb{D}}$.
(d) Find $\sigma_p(S)$.

[1+4+3+3=11 marks]

(4) Let H, K be Hilbert spaces, and $\{A_n\}_n$ be a sequence in $\mathcal{L}(H, K)$ such that $\{\langle A_n x, y\rangle\}_n$ is a Cauchy sequence in $\mathbb{C}$ for all $x \in H$ and $y \in K$. Show that there exists $A \in \mathcal{L}(H, K)$ such that A_n converges to A in WOT. (*Hint: Use UBP multiple times.*)

[8 marks]

(5) Let H be a Hilbert space and $T \in \mathcal{L}(H)$ such that $\langle Tx, x\rangle \geq 0$ for all $x \in H$. Show that $T \in \mathcal{L}_{\text{sa}}(H)$ and $\sigma(T) \subset [0, \infty)$.

[8 marks]

(6) Define $T : l^2 \to l^2$ by $(T\underline{x})_n = \dfrac{x_{n+1}}{n+1}$. Show that T is compact and find T^*.

[7 marks]

(7) Let S be a closed and convex subset of a Hilbert space H and $x \in H \setminus S$. Show that there exists unique $s_0 \in S$ such that $\mathrm{d}(x, s_0) = \mathrm{d}(x, S)$. (*Hint: Use parallelogram law.*)

[8 marks]

(8) Let H be the Hilbert space $L^2(\mathbb{R}, m)$ where m is the Lebesgue measure. For $t \in \mathbb{R}$, define $T_t : H \to H$ by $(T_t f)(x) := f(x - t)$ for all $x \in \mathbb{R}$.

(a) Show that $\{T_t\}_t$ is a (norm-) bounded net in $\mathcal{L}(H)$.

(b) Show that T_t converges to id_H in SOT as $t \to 0$. (*Hint: First test the convergence for compactly supported continuous functions.*)

(c) Show that $\|T_t - \mathrm{id}_H\| \geq \sqrt{2}$. (*Hint: Consider* $1_{[0,t]} \in H$.)

(d) Show that T_t converges to 0 in WOT as $t \to \infty$.

(e) Show that T_t does not converge in SOT as $t \to \infty$.

[1+4+2+5+2=14 marks]

INDIAN STATISTICAL INSTITUTE
Semester Examination 2019-20

M. Stat. 2nd Year
Statistical Inference II

Total marks 111. Answer as many as you can. Maximum you can score is 100. Date : 20.11.2019
Time: 3 hours
Note: Notations are as used in the class. No part-marking will be given for Question 1.

1. Suppose $X_1, \dots, X_n$ are n (> 2) independent and identically distributed observations from $Ber(\theta)$ for some $\theta \in [0,1]$ and consider the squared error loss function.

   (a) Considering uniform prior on θ, find out the posterior predictive probability of a new observation X_{n+1} being one given the sample data $(X_1, \dots, X_n)$. Compare the result with frequentist prediction (based on the MLE) when $X_i = 1$ for all $i = 1, \dots, n$.

   (b) Considering beta prior on θ, find out the Bayes Factor for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ for some pre-specified θ_0.

[(5+2)+8]=15

2. Prove or disprove the following statements.

   (a) The BVM theorem on posterior normality (for real parameter) also implies consistency of the posterior.

   (b) The posterior probability of the null hypotheses in any parametric hypotheses testing problem always differs from the corresponding (frequentist) p-value.

   (c) Jeffrey's prior maximizes the expected Kullback-Leibler divergence (appropriately defined) between the prior and the posterior.

[8×3]=24

3. Suppose $Y_1, \dots, Y_p$ are p independent random variables having distribution $N(\theta_i, V)$, respectively, for $i = 1, \dots, p$, where V is known and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$ depends on the given $p \times r$ covariate matrix $\boldsymbol{X}$ as $\boldsymbol{\theta} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim N_p(\mathbf{0}, A\boldsymbol{I}_p)$.

   (a) Find out the posterior of $\boldsymbol{\theta}$ given $\boldsymbol{\beta}$, A and the data.

   (b) Derive the parametric empirical Bayes estimate of $\boldsymbol{\theta}$ under squared error loss function.

   (c) In alternative Hierarchical Bayes (HB) approach, we put the priors on $\boldsymbol{\beta}$ and A as

   $$\boldsymbol{\beta}|A \sim N(\gamma_1, \gamma_2(\boldsymbol{X}^T\boldsymbol{X})^{-1}), \qquad A \sim \text{Inverse-Gamma}(a_0, b_0),$$

   with $\gamma_1, \gamma_2, a_0, b_0$ being known. Device a Gibbs sampler for computation of the HB estimator of $\boldsymbol{\theta}$ under squared error loss function.

[4+6+10]=20

4. Suppose that Bernoulli trials, each having probability of success $\theta \in [0, 1]$, are independently performed until a total of n successes is accumulated. Based on the outcome, we want to test for the hypotheses $H_0 : \theta = 0.5$ against $H_1 : \theta \neq 0.5$. Starting with Jeffrey's prior, derive an intrinsic prior for this problem. [10]

5. Suppose $X|\theta \sim N(\theta, \sigma^2)$ with some known $\sigma > 0$. We consider the following prior for the unknown location parameter θ:

$$\pi(\theta|\mu, \phi, c, \lambda) = k(c, \lambda)\sqrt{\phi}\exp\left[-\frac{c}{2}\rho_\lambda\left(1 + \frac{\phi(\theta-\mu)^2}{c-1}\right)\right],$$

where $\mu \in \mathbb{R}$, $\phi > 0$, $c > 1$ and $\lambda \geq 0$ are known hyperparameters, $k(c, \lambda)$ is the normalizing constant and $\rho_\lambda(\cdot)$ is given by

$$\begin{aligned} \rho_\lambda(x) &= \frac{x^\lambda - 1}{\lambda}, & \text{if } \lambda > 0, \\ &= \log(x), & \text{if } \lambda = 0. \end{aligned}$$

Our objective is to estimate θ with respect to the squared error loss function.

(a) Show that the family π contains both normal and student's t prior as its special cases.

(b) For $0 \leq \lambda \leq 1$, π can be written as a (scale) mixture of conjugate priors.

(c) Using (b), or otherwise, find out the Bayes estimator of θ. [4+6+10]=20

6. Suppose $X_1, \ldots, X_n$ are n (> 2) independent and identically distributed observations from a Weibull distribution, with both parameters (α, η) being unknown, having density

$$f(x|\alpha, \eta) \propto \alpha\eta x^{\alpha-1}e^{-\eta x^\alpha}, \quad x > 0,$$

(a) Describe the Metropolis-Hastings Algorithm to simulate from the joint posterior of (α, η) when the prior is given by

$$\pi(\alpha, \eta) \propto e^{-\alpha}\eta^{\beta-1}e^{-\zeta\eta}, \quad \beta, \zeta \text{ known.}$$

(b) Find out the class of probability matching priors if the shape parameter is of interest and verify if it contains the prior given in Part (a). [5+(5+2)]=12

7. Answer any ONE of the following questions.

(a) Explain why BIC can be considered as a model selection criterion following Bayesian paradigm. Any result used must be stated clearly along with the required assumptions.

(b) Explain how the Lasso regression solution can be considered as a Bayes estimator for appropriately chosen loss function and prior. How can it be computed using a Gibbs sampler for appropriately modified prior structure (to be specified by you)? [4+6]=10

INDIAN STATISTICAL INSTITUTE
Back-paper Examination 2019-20

M. Stat. 2nd Year
Statistical Inference II

January, 2020

Date : 09.1.2020

Total Marks: 100 Time: 3 hours

[Note: Notations are as used in the class. Answer as much as you can. Best of Luck!]

1. Explain the major differences between the Bayesian and the Frequentist paradigms, with examples, in the context of parametric estimation problem. Provide an example where Bayes approach yields more reasonable inference compared to the frequentist approach. [10+5]=15

2. Suppose that $X_1, \ldots, X_n$ are independent and identically distributed observations from $N(\mu, \sigma^2)$, where both μ, σ are unknown and consider the squared error loss function for the estimation of μ.

   (a) Find the Bayes estimator of μ with respect to the uniform prior $\pi(\mu) = 1$.

   (b) Find the Bayes estimator of μ with respect to the conjugate prior.

   (c) Consider the problem of testing $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$. Considering appropriate non-informative priors under these two hypotheses, derive the Bayes Factor, the Arithmetic intrinsic Bayes Factor and the Fractional Bayes Factors

[3+4+(6+6+6)]=25

3. Suppose $X_1, \ldots, X_n$ are independent and identically distributed observations from a parametric model density f_θ, for a real parameter θ, and θ has a prior density $\pi(\theta)$.

   (a) Write down the Bernstein-von Mises (BVM) theorem on posterior normality along with all the necessary regularity conditions.

   (b) Prove that all the regularity conditions hold when f_θ is the $N(\theta, 1)$ density, and simplify the BVM theorem for this particular case.

[(3+5)+(5+2)]=15

4. Suppose $Y_1, \dots, Y_k$ are k independent random variables having distribution $\text{Ber}(p_i)$, respectively, for $i = 1, \dots, k$, where p_i depends on the given value $\boldsymbol{x}_i$ of the covariates by the relation $p_i = \Phi(\boldsymbol{x}_i^T \boldsymbol{\beta})$. Assume a prior $\pi(\boldsymbol{\beta})$ on $\boldsymbol{\beta}$ and compute its posterior distribution. Device a Gibbs sampler for computation of the Bayes estimator of $\boldsymbol{\beta}$ under squared error loss function.

[5+15]=20

5. With appropriate set-up and assumptions, derive the James-Stein estimator of the mean of a multivariate normal (with known variance) as the solution to a Parametric Empirical Bayes problem. Explicitly prove all the steps of finding conditional distributions.

[5+10]=15

6. Prove that the Jeffrey's prior maximizes the expected Kullback-Leibler divergence (appropriately defined) between the prior and the posterior.

[10]

Indian Statistical Institute
M. Stat. II year, First Semester, 2019-20
Semestral Examination
Signal and Image Processing

Date: 30/11/2019 **Duration: 3 Hours** **Maximum Marks: 100**

Note: (i) Answer the two parts in two separate answer scripts.
(ii) Use of calculator is allowed.

Part A: Signal Processing

Answer any 5 questions.

1. a) Consider the following system function in Z-domain:

$$H(z) = \frac{3}{2}\cdot\left(\frac{1+5z^{-1}-3z^{-2}}{3z^{-1}+z^{-2}}\right).$$

Derive a possible realization of the above system as the cascade of an all-pass system and a non-minimum phase system.

b) A linear time-invariant system with frequency response $H(\omega)$ is excited with the following periodic input:

$$x(n) = \sum_{k=-\infty}^{\infty} \delta(n-kN).$$

Suppose that we compute the N-point DFT $Y(k)$ of the samples $y(n)$, $0 \le n \le N-1$ of the output sequence. How is $Y(k)$ related to $H(\omega)$? Explain mathematically.

c) Show that for a linear phase FIR filter, its coefficients are symmetrical around the central coefficient.

[4+3+3 = 10]

2. a) For the sequences: $x_1(n) = \cos\left(\frac{2\pi}{N}n\right)$ and $x_2(n) = \sin\left(\frac{2\pi}{N}n\right)$, $0 \le n \le N-1$, determine the N-point circular convolution of $x_1(n)$ and $x_2(n)$. Show your steps.

b) Let $X(k)$ be the N-point DFT of the sequence $x(n)$ for $0 < n < N-1$. We define a $2N$-point sequence $y(n)$ as:

PTO

$$y(n) = \begin{cases} x\left(\frac{n}{2}\right), & \text{if } n \text{ is even,} \\ 0, & \text{otherwise.} \end{cases}$$

Express the $2N$ point DFT of $y(n)$ in terms of $X(k)$.

[6+4 = 10]

3. Let $x(n)$ be a discrete time sequence with autocorrelation sequence given by $r_{xx}(l) = \left(\frac{1}{3}\right)^{|l|}$.

If $x(n)$ is applied at the input of an LTI system with impulse response:

$$h(n) = \delta(n) + 0.5\delta(n-1),$$

derive the autocorrelation sequence $r_{yy}(k)$ of the output sequence $y(n)$ and the output power spectral density $S_y(\omega)$. [4+6 = 10]

4. a) Consider the filter:

$$y(n) = 0.9x(n-1) + bx(n).$$

Let $H(\omega)$ denote the frequency response of the filter for discrete-time Fourier transform.

i) Determine b such that $|H(0)| = 1$.

ii) Determine the frequency for which $|H(\omega)| = \frac{1}{\sqrt{2}}$.

b) Convert the high-pass filter with system function:

$$H(z) = \frac{1 - z^{-1}}{1 - az^{-1}}, \text{ with } a < 1.$$

into a notch filter that rejects the frequency $\omega_0 = \frac{\pi}{4}$ and all its harmonics. Explain your steps. [(3+3)+4=10]

5. Compute the 16-point DFT of the following sequence:

$$x(n) = \cos\left(\frac{\pi}{2}n\right), \text{ with } 0 \le n \le 15$$

by using the radix-4 decimation-in-time FFT algorithm. Show all your steps. [10]

6. To estimate the parameter A from the observation of a signal:

$$x(n) = A\cos\left(\frac{\pi}{3}n\right) + w(n), \quad 0 \le n \le 7,$$

where w is the zero mean White Gaussian Noise (WGN) with unity variance, suppose one proposes the following estimator: $\hat{A} = \frac{1}{2}[x(1) - x(4)]$.

i) Determine the Cramér–Rao lower bound for the estimation of A.
ii) Is the estimator $\hat{A}$ unbiased? Give reasons.
iii) Is the estimator $\hat{A}$ efficient? Explain. [5+2+3 = 10]

Image Processing: Answer any ten (10) questions.

1. (a) A single channel 256×256, 8-bit per pixel, grey level image is stored in a file of size 6554 bytes. What is the compression ratio?

   (b) A 10×10 image has four grey level pixel values g_0, g_1, g_2 and g_3. There are 20, 30, 10 and 40 number of pixels in each pixel values g_0, g_1, g_2 and g_3 respectively. Find the Huffman code for these four pixel values. Comment on the savings in bits per pixel due to Huffman coding. [1+3+1=5]

2. Consider a 5×5 image matrix

2	1	2	0	1
0	2	1	1	2
0	1	2	2	0
1	2	2	0	1
2	0	1	0	1

   Find the grey level co-occurrence matrix for the configuration of pixels ***I*** and ***J*** shown below.

I	
	J

   Find the entropy of the co-occurrence configuration shown above. [3+2=5]

3. Derive level set based curve evolution equation. [5]

4. Explain region merge and split based image segmentation methods by using an example. [5]

5. Write the main steps and their functions of JPEG image compression. [5]

6. What is RLE? What is the advantage of RLE? How RLE can be used for 8-bit grey level image? [2+1+2=5]

7. What is grey level co-occurrence matrix (GLCM)? How entropy of GLCM can be evaluated? [3+2=5]

8. What is LBP? Explain briefly. What is uniform LBP? What is the advantage of uniform LBP? [3+1+1=5]

9. Show using an example that Huffman coding helps in bit level compression. [5]

10. State mathematical expressions for 2D Fourier transform and inverse Fourier transform and define each variable? State two utilities of frequency transformation of image. [2+2+1=5]

Statistical Genomics
M-Stat (2nd Year)
End Semester Examination 2019-20

Date:November 27, 2019 **Time: 3 hours**

Use separate answer booklets for each Group.
Calculators are allowed

Group A

This group carries 25 marks and is a closed notes examination. You need to submit your answer booklet before getting access to notes. Answer all questions.

1. You have sequenced the DNA of a new gene coding for an unknown protein.
(a) Describe how you would find possible similar genes in other organisms?
(b) Well-known scoring matrices are PAM and BLOSUM matrices. Which scoring matrix would you choose if you use translated-BLAST (searches protein database using a translated nucleotide query)? Justify your answer. [3+2]

2. Construct the BLOSUM scoring matrix using the following multiple sequence alignment of 5 protein sequences (S1, S2, S3, S4 and S5).

S1: A A I
S2: S A L
S3: T A L
S4: T A V
S5: A A L [7]

3. The tree metric distances among 4 taxa (A, B, C and D) are given below.

	A	B	C	D
A		5	4	8
B			3	5
C				6
D				

(a) Construct the rooted and unrooted metric tree for these 4 taxa. Show all the steps involved in producing the trees.
(b) If the trees produced are different, explain how and why they are different.

[(5+6)+2]

Group B

This group carries 25 marks and is an open notes examination. Answer all questions.

1. Consider the following data at a marker locus in an affected sib-pair study:

Sib-pair	Parental marker genotypes	Sib-pair marker genotypes
1	AC-BC	AC-AC
2	AB-AB	BB-BB
3	AB-BC	AB-AC
4	AC- *	AA-AB
5	AB-AC	AB-AB

(* denotes missing genotype)

Using Holmans' Triangle approach, test for linkage between the marker locus and a disease locus. [6]

2. Suppose you wish to perform the classical Transmission Disequilibrium Test at a biallelic marker locus with minor allele frequency 0.3.

(a) What is the expected value of the test statistic under no association but in the presence of linkage?
(b) What is the expected number of families that needs to be sampled so as to obtain 100 informative trios? [5]

3. Consider data on a quantitative trait on sib-pairs and genotypes at a marker locus for both parents as well as the sibs. Assuming the classical Haseman-Elston framework with no dominance, show that the squared sum in the sib-pair trait values conditioned on their i.b.d. score at the marker locus is a linear function of the i.b.d. scores. How would you test for linkage between a QTL and a marker locus based on the above property? [6]

4. Consider the following genotype data on 1000 randomly selected individuals at two autosomal biallelic loci with alleles (A,a) and (B,b) respectively. Explain how you would use the EM Algorithm to obtain the maximum likelihood estimate of the coefficient of linkage disequilibrium between alleles at the two loci? Show all computational steps with one cycle of iteration. [8]

	BB	*Bb*	*bb*
AA	240	162	34
Aa	58	270	160
aa	10	30	36

Statistical Genomics
M-Stat (2^{nd} Year)
Back paper Examination 2019-20

Date: 10.1.2020 **Time: 3 hours**

Use separate answer booklets for each Group. Calculators are allowed. The paper carries 100 marks. Answer all questions.

Group A

1. Consider the following data from an affected sib-pair study (both sibs are affected and affection status of parents are unknown) on schizophrenia. *CHRNA7* (cholinergic receptor nicotinic alpha polypeptide 7) on Chromosome 15 is believed to be a candidate gene for schizophrenia. Twelve sib-pairs comprising all affected siblings (along with their parents) were genotyped at a triallelic marker locus *D15S1012* near this candidate gene. Do these data provide evidence of linkage between *D15S1012* and a locus controlling schizophrenia? [25]

Sibship	Parental genotypes	Genotypes of affected sibs
1	AA,AB	AA,AB
2	AB,BC	AB,BC
3	AB,CC	AC,BC
4	BB,BC	BB,BB
5	AB,AB	AB,AB
6	AC,*	AA,AA
7	AB, BB	AB,BB
8	AA,AB	AA,AB
9	*, *	AC,AC
10	CC,CC	CC,CC
11	AB,AC	AC,BC
12	AC,AC	AA,AC

* denotes missing genotype

2. Consider genotype data at a marker locus on parent-offspring pairs where the offspring is affected with a dominant disorder. Show that the test for linkage disequilibrium between the marker locus and the disease locus is equivalent to a test for Hardy-Weinberg Equilibrium of the parental genotypes. Explain whether this test is protected against population stratification. [8 + 7]

3. Given genotype data at two biallelic loci for an unrelated set of individuals, explain how you would test for allelic association between the two loci. [10]

Group B

4. What are the important parameters in BLAST search? Write down their significances.
[7]

5. Suppose that for five species *a, b, c, d* and *e*, the pair-wise distances are given by:

	a	*b*	*c*	*d*	*e*
a		9	8	7	8
b			3	6	7
c				5	6
d					3
e					

Construct the tree using the neighbor-joining method. Show the partial tree in each case, together with the edge length. Now construct the tree using the UPGMA, together with the edge lengths. Compare your answers for the two reconstructions.
[10+10+5]

6. In BLOSUM80 you will find that matching score of tryptophan (W to W) is +11, while the same for alanine (A to A) is only +4. Explain why the matching score for W/W and A/A is different.
[8]

7. There are several different substitution matrices used in protein alignments. Why would you prefer one over another? [5]

8. What are the advantages and disadvantages of the Needleman-Wunsch alignment compared to a seeded alignment?
[5]

INDIAN STATISTICAL INSTITUTE
First Semester Examination: 2019-20

M.Stat Second Year
Pattern Recognition

Date: 25.11.2019 Maximum Marks: 100 Duration: 4 Hours

(All answers should be brief and to the point. Answer as much as you can. The maximum you can score is 100.)

1. Consider a two-class classification problem based on a single variable, where the densities of the two classes are given by

$$f_1(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases} \quad \text{and} \quad f_2(x) = \begin{cases} 1+\cos(2\pi x) & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Assuming that the prior probabilities of the two classes are equal, compute the average misclassification probability of the Bayes classifier. [5]

   (b) Show that the minimax classification rule is of the form

$$\delta(x) = \begin{cases} 1 & \text{if } c_0 \le x \le 1-c_0 \\ 2 & \text{otherwise,} \end{cases}$$

   where c_0 is a positive constant smaller than $1/4$. [5]

   (c) Show that the average misclassification probability of this minimax classifier cannot be smaller than $(\pi-1)/2\pi$. [2]

2. (a) A student was asked to generate three observations, one from the standard normal distribution, one from the Cauchy distribution and one from the standard double exponential distribution. The student generated three observations x_1, x_2 and x_3, but forgot which observation was generated from which distribution. Describe how you will help the student in labeling these observations? [4]

   (b) Consider a two-class classification problem, where we have d-dimensional observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ and $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ from the two classes. If $d > m+n$ and the observations are linearly independent, show that there exists a linear classifier, which will perfectly classify these observations. [4]

   (c) Prove or disprove:- "If two elliptically symmetric distributions differ only in their scales and they have the same prior probability, the Bayes classifier turns out to be a quadratic classifier". [4]

3. (a) Formulate the optimization problem used by the linear support vector machine (SVM) classifier when the observations from the two classes are not linearly separable. Also construct the corresponding dual problem. [3+3]

   (b) Assuming the observations to be two-dimensional, draw a scatter plot of the observations and identify the support vectors. [3]

   (c) Using your answers to 3(a), describe how the kernel trick can be used for constructing a nonlinear classifier based on SVM. [3]

4. (a) Show that in a two-class problem, asymptotic misclassification probability of the 1-nearest neighbor classifier cannot exceed twice the Bayes risk. [6]

   (b) Let $T_0 \supset T_1 \supset \ldots \supset T_k$ be a nested sequence of classification trees. Consider a cost function $\Delta_\alpha(T) = \Delta(T) + \alpha|\tilde{T}|$, where Δ stands for the resubstitution error (i.e., the error rate in the training sample) and $|\tilde{T}|$ is the cardinality of the leaf nodes in the tree T. For $i = 1, 2, \ldots, k$, define

$$\eta_i = (\Delta(T_i) - \Delta(T_0))/(|\tilde{T}_0| - |\tilde{T}_i|).$$

   If $\eta_1 > \eta_2$, show that for all choices of $\alpha \geq \eta_2$, $\Delta_\alpha(T_1)$ exceeds $\Delta_\alpha(T_2)$. [6]

5. (a) Write down Lloyd's algorithm used for k-means clustering. Does this algorithm always converge? Justify your answer. [3+3]

   (b) Give an example to show that this algorithm may not always lead to the global minimizer of the objective function. [3]

   (c) Briefly describe how EM algorithm can be used for model based semi-supervised classification (give the idea only, no need to write down the algorithm). [3]

6. Comment on the validity of the following statements. [10 × 3=30]

   (a) The maximum depth classifier based on half-space depth can perform better than the usual linear discriminant analysis (LDA) rule even when the Bayes classifier is linear.

   (b) A single-layer feed-forward neural network model (single-layer perceptron model) used for binary classification can be viewed as a generalized linear model.

   (c) Kernel discriminant analysis rule can outperform both linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) rules even when the underlying distributions have finite moments.

   (d) If the Gaussian kernel with the same bandwidth h is used for all competing classes, for a very small value of h, the kernel discriminant analysis rule behaves like the 1-nearest neighbor classifier.

   (e) If a large common bandwidth is used for both of the competing classes, a kernel discriminant analysis rule behaves like a linear classifier even when the prior probabilities are not equal.

   (f) In a classification problem involving three competing classes, for constructing a classification tree, $\psi(p_1, p_2, p_3) = p_1 p_2 p_3$ can be considered as an ideal choice for the impurity function.

   (g) If the impurity function used for constructing a classification tree is concave and symmetric in its arguments, it takes the maximum value when all classes have equal number of representatives.

   (h) In a high-dimensional classification problem, the 1-nearest neighbor classifier can classify all observations to a single class.

   (i) The integral of a k-nearest neighbor density estimate over the entire measurement space may not always be finite.

   (j) Popular k-means clustering algorithm may fail to properly identify two elliptic clusters even when they are well separated.

7. Computer Assignments [20]

M.Stat Second Year

Pattern Recognition (Backpaper) INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2019-20

Date: 08/01/2020 Maximum Marks: 100 Duration: $3\frac{1}{2}$ Hours

1. Let δ_π be a Bayes rule and $\mathcal{E}_\pi(\delta_\pi)$ be the corresponding Bayes risk for a two-class classification problem when π $(0 < \pi < 1)$ is the prior probability of the first class. If $R(\delta_{\pi_0}, 1) = R((\delta_{\pi_0}, 2)$ for some $\pi_0 \in (0,1)$, show that

   (a) δ_{π_0} is a minimax rule. [4]

   (b) π_0 is the maximizer of $g(\pi) = \mathcal{E}_\pi(\delta_\pi)$. [4]

2. Consider a classification problem, where each of the two competing classes is an equal mixture of two bivariate normal distributions. While one class is a mixture of $N(1,1,1,1,0)$ and $N(-1,-1,1,1,0)$, the other one is a mixture of $N(-1,1,1,1,0)$ and $N(1,-1,1,1,0)$. If the two classes have the same prior probability, find the Bayes classifier and the corresponding Bayes risk. [Here $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ denotes a bivariate normal with the location $(\mu_1, \mu_2)^\top$, marginal variances σ_1^2, σ_2^2 and correlation coefficient ρ. If required, use $\Phi(1) = 0.8413$ and $\Phi(2) = 0.9772$ for the calculation of the Bayes risk.] [4+4]

3. Suppose that $\mathbf{X}_1$ and $\mathbf{X}_2$ follow two elliptically symmetric distributions, which have the same scatter matrix $\mathbf{\Sigma}$, and differ only in their locations $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, show that

   (a) $\mathbf{X}_1 - \mathbf{X}_2$ follows an elliptically symmetric distribution. [6]

   (b) $P\{\boldsymbol{\alpha}'(\mathbf{X}_1 - \mathbf{X}_2) > 0\}$ is maximized when $\boldsymbol{\alpha} \propto \mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. [6]

4. (a) Consider a J class ($J > 2$) classification problem, where the competing classes have prior probabilities $\pi_1, \ldots, \pi_J$. If $\pi_1 > \pi_j$ for all $j > 1$, show that the error rate of the kernel discriminant analysis rule converges to 1 - π_1 as the common bandwidth h tends to infinity. [4]

   (b) Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed with a density $f(x)$, which is a decreasing function of $\|\mathbf{x}\|$. Let $\hat{f}_h$ be the kernel density estimate of f based on a spherically symmetric kernel. Show that $E[\hat{f}_h(x)]$ is also a decreasing function of $\|\mathbf{x}\|$. [8]

5. Consider a two-class classification problem, where we have observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ from the two classes. For $i = 1, 2, \ldots, n$, define $y_i = 1$ (respectively, -1) if $\mathbf{x}_i$ comes from the first (respectively, second) class. Assume that the observations from the two classes are linearly separable.

   (a) Formulate the optimization problem used by linear support vector machine (SVM) for constructing the separating hyperplane. [2]

   (b) Show that dual of the optimization problem can be expressed as [3]

$$\text{Maximize} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j \ \text{ subject to } \sum_{i=1}^{n} \alpha_i y_i = 0 \text{ and } \ \alpha_i \geq 0 \text{ for all } i = 1, 2, \ldots, n.$$

   (c) Show that the slope of the optimal separating hyperplane can be expressed as a linear combination of the support vectors. [2]

   (d) Describe how kernel trick can be used for nonlinear classification. [3]

[Turn over]

6. (a) Consider a two-class classification problem with unequal priors. If the average misclassification probability of the Bayes classifier is equal to the smaller of the two prior probabilities, show that it classifies all observations to a single class. [4]

   (b) Consider a classification problem between two Gaussian distributions $N(\mathbf{0}, \sigma_1^2\mathbf{I})$ and $N(\boldsymbol{\mu}, \sigma_2^2\mathbf{I})$, where $\boldsymbol{\mu} = (\mu, \mu, \ldots, \mu)^\top$. If the sample sizes from the two classes are fixed, study the high dimensional behavior of the 1-nearest neighbor classifier when (i) $\mu^2 > |\sigma_1^2 - \sigma_2^2|$ and (ii) $\mu^2 < |\sigma_1^2 - \sigma_2^2|$. [3+3]

7. (a) Describe the pairwise coupling algorithm used in a multi-class classification problem for aggregating the results of all pairwise classifications. [3]

   (b) Show that this algorithm can be viewed as a majorization-minimization algorithm. [4]

   (c) Does this algorithm always converge? Justify your answer. [3]

8. Consider a J-class classification problem. If a node t of a classification tree contains p_j proportion observations from the j-th class ($j = 1, \ldots, J$), its impurity function is defined as $i(t) = \psi(p_1, \ldots, p_J) = \sum_{i \neq j} p_i p_j$.

   (a) Show that for any split of the node t, the reduction in the impurity function is non-negative. [4]

   (b) Show that ψ is maximized when $p_j = 1/J$ for $j = 1, 2, \ldots, J$. [2]

   (c) Show that $\psi(p_1, p_2, \ldots, p_J) \geq \frac{1}{2}(1 - \max p_j)$. When does the equality hold ? [3+1]

9. (a) Describe how EM algorithm can be used for cluster analysis based on Gaussian mixture models. How will you choose the number of clusters in such cases? [4+2]

   (b) Describe how k-means clustering algorithm can be used for (i) image compression (ii) condensed nearest neighbor classification. [3+3]

   (c) Give a brief description of the k-medoids clustering algorithm. Does this algorithm always converge? Justify your answer. What are its advantages over the k-means clustering algorithm? [3+3+2]

INDIAN STATISTICAL INSTITUTE
Semestral Examination : 2019 – 20
MStat (2nd Year)
Quantitative Finance

Date: 27 November 2019 Maximum Marks: 100 Duration: 3 Hours

1. Explain the following with an example: [6 X 4 = 24]
   a) Dominant strategy and Law of one price
   b) Second order continuous parameter stochastic process
   c) Martingale property
   d) Use of Girsanov's theorem in finance

2. Prove the Put – Call parity of European option for the multi-period market. Is the same relation true for American options? – Prove or refute logically. [7 + 7 = 14]

3. Define the following option contracts:
   (i) Lookback
   (ii) Barrier
   (iii) Chooser

   For each of them, state the payoff function carefully, explaining all notation. [3 X 6 = 18]

4. Consider a two period model with $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and asset with time 0 price $A_0 = 100$.
   (a) If the bank process is deterministic with per period interest rate $r = 0.05$ and the asset can be sold short, then what is the time 0 forward price O_0 of the asset for delivery at time 2?
   (b) If the bank process is random with $B_1 = 1.05, B_2(\omega_1) = B_2(\omega_2) = 1.12$ and $B_2(\omega_3) = B_2(\omega_4) = 1.1$, and if the asset can be sold short, then what is O_0 now? Give an expression in terms of the risk neutral probability $Q(\{\omega_1, \omega_2\})$.
   (c) If the bank process is as in (b) but **short selling is not allowed,** then what is the largest value of O_0 consistent with no-arbitrage?
   (d) What happens if in (c), the asset has a carrying cost of 5 per period? [5 + 5 + 5 + 5 = 20]

5. Prove that $M_t = B_t^2 - t$ is a martingale where $\{B_t\}$ is a b.m. [12]

6. Let $x > 0$ be a constant and define $X_t = (x^{1/3} + \frac{1}{3}B_t)^3;\ t \geq 0$, where $\{B_t\}$ is an s.b.m.

   Show that $dX_t = \frac{1}{3} X_t^{1/3} dt + X_t^{2/3} dB_t; X_0 = x.$ [12]

INDIAN STATISTICAL INSTITUTE

Backpaper Examination : 2019 – 20

MStat (2nd Year)

Quantitative Finance

Date: 8 January 2020 Maximum Marks: 100 Duration: 3 Hours

1. Critically explain the concepts: [5 X 4 = 20]
   a) Mutual Fund Principle
   b) Law of One Price
   c) Barrier Options
   d) Reflection Principle

2. Let

$$A_{(K+1)X(K+2N)} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \\ \Delta S_1^*(\omega_1) & -\Delta S_1^*(\omega_1) & \Delta S_2^*(\omega_1) & \cdots & -\Delta S_N^*(\omega_1) & -1 & 0 & \cdots & 0 \\ \Delta S_1^*(\omega_2) & -\Delta S_1^*(\omega_2) & \Delta S_2^*(\omega_2) & \cdots & -\Delta S_N^*(\omega_2) & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Delta S_1^*(\omega_K) & -\Delta S_1^*(\omega_K) & \Delta S_2^*(\omega_K) & \cdots & -\Delta S_N^*(\omega_K) & 0 & 0 & \cdots & -1 \end{bmatrix}$$

and $b_{(K+1)} = (1, 0, \ldots, 0)'$. Show that

$$Ax = b, \;\; x \geq 0, x \in \mathbb{R}^{K+2N}$$

has a solution if and only if there exists an arbitrage opportunity in the securities market with N securities S_i (i = 1,…,N) and K states of nature ω_j (j = 1,…, K). S_i^*'s are discounted (by the bank process) values. [20]

3. Prove directly from the definition of Ito integrals that

$$\int_0^t B_s^2 dB_s = \frac{1}{3} B_t^3 - \int_0^t B_s ds,$$

where $\{B_t\}$ is the standard Brownian motion. [16]

4. In the two period model, explicitly solve the Consumption Investment problem for the utility function $u(w) = \ln w$, where w is the wealth. Compute the relevant expressions for the first order conditions and solve for the optimal trading strategy when N = 1, K = 2, r = 1/9, $S_0 = 5$, $S_1(\omega_1) = 20/3$, $S_1(\omega_2) = 40/9$ and $P(\omega_1) = 3/5$. [12 + 8 = 20]

5. (a) Suppose $S_0 = 12$, T = 3, $r = 0$, $u = 1.5 = 1/d$ are the parameters for a Binomial model. Compute the price of an Asian Call option with exercise price = 10.

   (b) Prove that $B_t^2 - t$ is a Martingale, where $\{B_t\}$ is an s.b.m. [12 + 12 = 24]