Review for the Final Examination

Ch 3. Integration
Ch 4. MCMC Methods
Ch 5. Bootstrap Methods

1). Arrangement of 2021 Final Examination

- **Date and Time**
  06 Jan 2022 (Thursday),
  08:00am – 10:00am (2 hours)
- **Venue**
  Room 101, Building 1, Lychee Hills
- **Range**
  Chapters 1,2,3,4,5 (Except for Sections 3.1, 3.4, 5.2)
- **Assessment**
  50% weighting

2). The Policy of Closed Book Final Examination

- Please bring two pens in case one is not available.
- You can prepare anything on one side of an A4 paper and bring it with you to the test venue.
- You are not allowed to bring any other material (including iPhone/iPad) to the test venue. The calculator is not required this time.

3). The Distribution and Marks of the Questions in the Examination

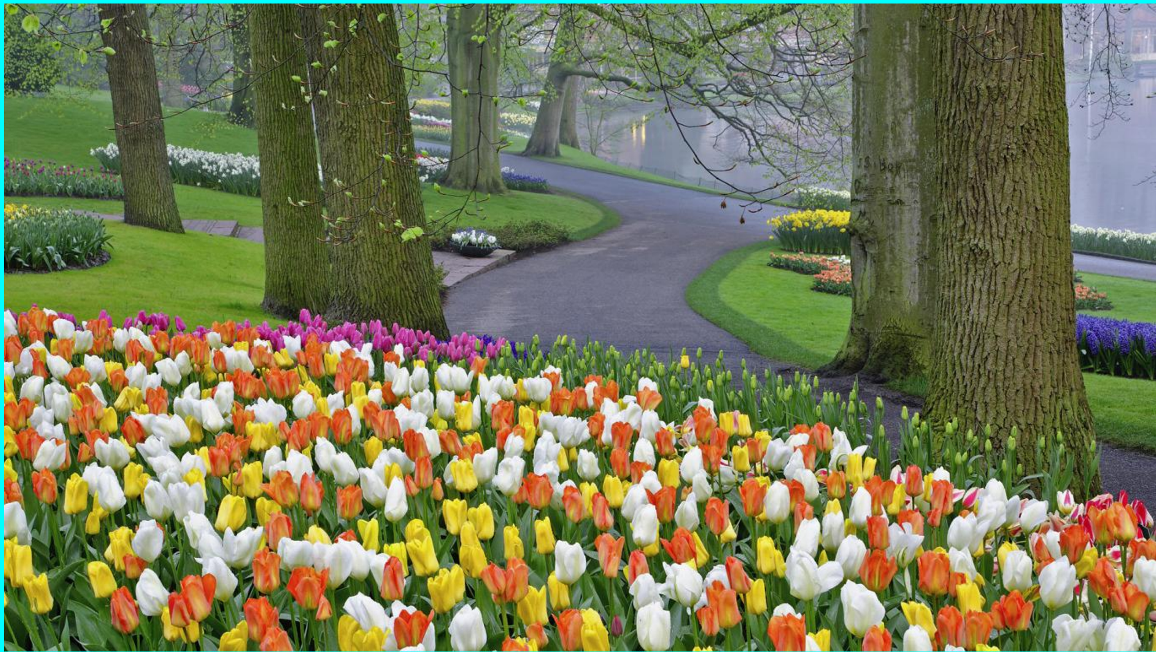
- **Fill-in-the-blanks.** **[30-marks]**
- **1 question on the inversion method in Chapter 1 with 2 sub-questions (one is discrete distribution and the other is continuous distribution).** **[10-marks]**
- **1 question on the mixture representation method for generating a continuous random variable.** **[10-marks]**

- **1 question on the Newton–Raphson, Fisher scoring in Chapter 2. [15-marks]**

- **1 question on SR, EM, DA, MM algorithms in Ch 4. [35-marks]**

Ch 3. Integration

- §3.2 Riemannian Simulation
- §3.3 The Importance Sampling Method



§3.2 Riemannian Simulation

1. Classical Monte Carlo integration

- **The Monte Carlo integration** of

$$\mu = E\{h(X)\} = \int_{\mathbb{X}} h(x) \cdot f(x)\, dx, \qquad (3.8)$$

is

$$\bar{\mu}_m = \frac{1}{m} \sum_{i=1}^{m} h(X^{(i)}). \qquad (3.9)$$

where $X^{(1)}, \ldots, X^{(m)} \overset{\text{iid}}{\sim} f(x)$.

2. The Rieman sum

Let $\mathbb{X} = [a, b]$ **and** $a = a_1 < \cdots < a_{m+1} = b$**, then when** $m \to \infty$**, the Rieman sum**

$$\sum_{i=1}^{m} h(a_i) f(a_i)(a_{i+1} - a_i) \to \int_a^b h(x) \cdot f(x)\, dx.$$

3. The Riemannian sum estimator

Replacing the fixed points $\{a_i\}_{i=1}^{m+1}$ by s-tochastic points $\{X_{(i)}\}_{i=1}^{m+1}$, the ordered s-tatistics of i.i.d. samples $X^{(1)}, \ldots, X^{(m+1)}$ from $f(x)$, the approach of Riemannian simulation approximates the integral (3.8) by

$$\hat{\mu}^R = \sum_{i=1}^{m} h(X_{(i)}) f(X_{(i)}) \{X_{(i+1)} - X_{(i)}\}. \quad (3.14)$$

We call $\hat{\mu}^R$ the Riemannian sum estimator of μ.

§3.3 The Importance Sampling Method

Step 1. **Generate i.i.d. samples** $\{X^{(i)}\}_{i=1}^{m}$ **from a proposal density** $g(\cdot)$ **with the same support of** $f(\cdot)$**;**

Step 2. **Let** $H(x) = h(x)f(x)$ **. Calculate ratios** $\mathbf{w}(X^{(i)}) = H(X^{(i)})/g(X^{(i)})$ **for** $1 \leqslant i \leqslant m$ **& approximate** $\mu = \int_{\mathbb{X}} H(x)\,dx = \int_{\mathbb{X}} \mathbf{w}(x)g(x)\,dx$ **by**

$$\tilde{\mu}_m = (1/m)\sum_{i=1}^{m}\mathbf{w}(X^{(i)}), \quad (3.15)$$

Which is called the importance sampling estimator.

Ch 4. MCMC Methods

- §4.1 Bayes Formula and IBF
- §4.2 The Bayesian Methodology
- §4.3 The DA Algorithm
- §4.4 The Gibbs Sampler
- §4.5 The Exact IBF Sampler
- §4.6 The IBF Sampler

§4.1 Three Points to Bayes Formula and IBF

- **1. The core of Bayes Formula:**

$$f_{(X|Y)}(x|y)f_Y(y) = f_{(Y|X)}(y|x)f_X(x), \qquad (4.1)$$

where $(x, y) \in \mathcal{S}_{(X,Y)}$**.**

- **2. Given two conditional densities, how to derive three IBF under the assumption of the product space (please review (4.3) – (4.6) on pages 154–155, and Ex 4.1 – Ex 4.3).**

- **3. Given three full conditional densities, how to determine the joint density** $f(x_1, x_2, x_3)$ **(please review §4.1.3).**

§4.2 The Bayesian Methodology

- **Posterior $\propto$ Likelihood $\times$ Prior:**

$$p(\boldsymbol{\theta}|Y_{\mathrm{com}}) = \frac{f(Y_{\mathrm{com}}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{f(Y_{\mathrm{com}})} \propto f(Y_{\mathrm{com}}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{4.14}$$

- **What is the conjugate prior distribution? (See page 166)**
- **Please review Exs 4.4, 4.5, and 4.6.**

§4.3 The DA Algorithm

- **1. The original DA algorithm:**
  Please review I-step and P-step in page 179.

- **2. The modern DA algorithm:**
  – **I-step: Draw** $z^{(t)} \sim f(z|Y_{\rm obs}, \boldsymbol{\theta}^{(t)})$
  – **P-step: Draw** $\boldsymbol{\theta}^{(t+1)} \sim p(\boldsymbol{\theta}|Y_{\rm obs}, z^{(t)})$

- **3. The connection with the IBF:**
  Please review §4.3.4, and Example 4.8.

§4.4 The Gibbs Sampler

- **1. The general sampling procedure:**
  - **Draw** $x_1^{(t+1)} \sim f(x_1 | x_2^{(t)}, \ldots, x_d^{(t)})$**;**
  - **Draw** $x_2^{(t+1)} \sim f(x_2 | x_1^{(t+1)}, x_3^{(t)}, \ldots, x_d^{(t)})$**;**

    $\ldots \qquad \ldots \qquad \ldots$
  - **Draw** $x_d^{(t+1)} \sim f(x_d | x_1^{(t+1)}, \ldots, x_{d-1}^{(t+1)})$**.**
- **2. The two-block Gibbs sampling:**
  - **Draw** $\boldsymbol{x}_1^{(t+1)} \sim f(\boldsymbol{x}_1 | \boldsymbol{x}_2^{(t)})$**;**
  - **Draw** $\boldsymbol{x}_2^{(t+1)} \sim f(\boldsymbol{x}_2 | \boldsymbol{x}_1^{(t+1)})$**.**

  Please review Exs 4.9 and 4.10.

§4.5 The Exact IBF Sampling

- **1. Basic assumptions for the exact IBF sampling:**
  $p(\boldsymbol{\theta}|Y_{\text{obs}}, z)$ **and** $f(z|Y_{\text{obs}}, \boldsymbol{\theta})$ **are given, and** Z **is a discrete r.v.**

- **2. The algorithm:**

  Step 1. Identify $\mathcal{S}_{(Z|Y_{\text{obs}})} = \{z_1, \ldots, z_K\}$ **from** $f(z|Y_{\text{obs}}, \boldsymbol{\theta})$ **and calculate** $\{p_k\}_1^K$ **according to (4.32) and (4.31);**

§4.5 The Exact IBF Sampling (Cont's)

Step 2. **Generate i.i.d. samples** $\{z^{(\ell)}\}_{\ell=1}^{L}$ **of** Z **from the pmf** $f(z|Y_{\text{obs}})$ **with probabilities** $\{p_k\}_1^K$ **on** $\{z_k\}_1^K$**;**

Step 3. **Generate** $\boldsymbol{\theta}^{(\ell)} \sim p(\boldsymbol{\theta}|Y_{\text{obs}}, z^{(\ell)})$ **for** $\ell = 1, \ldots, L$**, then** $\{\boldsymbol{\theta}^{(\ell)}\}_1^L$ **are i.i.d. samples from the observed posterior distribution** $p(\boldsymbol{\theta}|Y_{\text{obs}})$**.**

§4.6 The IBF Sampler

- **1. Basic assumptions for the IBF sampler:**
  The posterior mode $\tilde{\boldsymbol{\theta}}$ can be obtained by an EM algorithm based on $p(\boldsymbol{\theta}|Y_{\mathrm{obs}}, z)$ and $f(z|Y_{\mathrm{obs}}, \boldsymbol{\theta})$, and set $\boldsymbol{\theta}_0 = \tilde{\boldsymbol{\theta}}$;
- **2. The algorithm:**

  Step 1. Draw J i.i.d. samples $\{z^{(j)}\}_{j=1}^J$ of Z from $f(z|Y_{\mathrm{obs}}, \boldsymbol{\theta}_0)$;

Step 2. **Calculate the reciprocals of the augmented posterior densities to obtain the weights**

$$\omega_j = \frac{p^{-1}(\boldsymbol{\theta}_0|Y_{\text{obs}}, z^{(j)})}{\sum_{\ell=1}^{J} p^{-1}(\boldsymbol{\theta}_0|Y_{\text{obs}}, z^{(\ell)})}, \quad (4.35)$$

for $j = 1, \ldots, J$**;**

Step 3. **Choose a subset from** $\{z^{(j)}\}_1^J$ **via resampling without replacement from the discrete distribution on** $\{z^{(j)}\}$ **with probabilities** $\{\omega_j\}$ **to obtain an i.i.d. sample of size** $I\,(< J)$ **approximately from** $f(z|Y_{\rm obs})$**, denoted by** $\{z^{(k_i)}\}_1^I$**;**

Step 4. **Generate** $\boldsymbol{\theta}^{(i)} \sim p(\boldsymbol{\theta}|Y_{\rm obs}, z^{(k_i)})$ **for** $i = 1, \ldots, I$**, then** $\{\boldsymbol{\theta}^{(i)}\}_1^I$ **are i.i.d. samples from** $p(\boldsymbol{\theta}|Y_{\rm obs})$**.**

Ch.5 Bootstrap Methods

- §5.1.1 Parametric Bootstrap
- §5.1.2 Non-parametric Bootstrap

§5.1.1 Parametric Bootstrap

- **1. Several important notions:** **Bootstrap sample; Bootstrap replication; Normality-based Bootstrap CI; Non-normality-based Bootstrap CI.**
- **2. The Bootstrap Approach for Constructing CIs:** **Please review the six steps in pages 205–206 and Table 5.1**
- **3. Applications. (Please review Examples 5.1 and 5.3)**

§5.1.2 Non-parametric Bootstrap

- **1. The empirical distribution function:** **Based on the iid observations** $x_1, \dots, x_n \overset{\text{iid}}{\sim} F$, **the empirical cdf is**

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(x_i \leqslant x)$$

where $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$**.**

- **2. The Bootstrap Approach for Constructing CIs:** **Please review the six steps in pages 214–215 and Table 5.2**

- **3. Applications. (Please review Examples 5.4 and 5.5)**

Please email me or come to my office

During January 3–5, 2022, if you have any questions.

End of the Review