# PhD Comprehensive Written Examination in Statistics

Part I - Theory

August 21, 2017 - from 9:00 am till 12:00 noon

## Statistical Inference

1. Suppose that $X_1$, $X_2$, ..., $X_n$ is a random sample of size $n > 2$ from a distribution with probability density function (pdf)

$$f(x;\theta) = \theta\, x^{\theta-1}, \qquad 0 < x < 1, \quad \theta > 0,$$

and 0 elsewhere.

(a) Show that $X_1$, ..., $X_n$ has an exponential family distribution.

(b) Show that $X_1$, ..., $X_n$ has a regular exponential family distribution.

(c) Find a complete sufficient statistic for this model.

(d) Find the minimum variance unbiased estimator (MVUE) of $\theta$. [*Hint*: $-\sum_{i=1}^{n} \log X_i \sim$ Gamma$(n, \frac{1}{\theta})$]

**Note:**

If $X \sim \mathit{Gamma}(\alpha, \beta)$, the pdf of $X$ is

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0; \quad \alpha, \beta > 0,$$

and the expectation of $X$ is $E(X) = \alpha\beta$.

2. Let $X_1$, $X_2$, …, $X_n$ be a random sample of size $n > 2$ from a distribution with pdf

$$f(x;\theta) = \theta\, x^{\theta-1}, \qquad 0 < x < 1, \quad \theta > 0,$$

and 0 elsewhere.

(a) Find the maximum likelihood (ML) estimator of $\theta$.

(b) Find the likelihood ratio test statistic for testing $H_0 : \theta = 1$ against $H_1 : \theta \neq 1$.

3. Let $(X_1, ..., X_n)$ be a random sample from the distribution with pdf

$$f(x;\theta) = \frac{2\theta^2}{x^3}, \qquad x \geq \theta, \quad \theta > 0,$$

and 0 elsewhere.

(a) Find the complete sufficient statistic for this model.

(b) Find the minimum variance unbiased estimator (MVUE) of $\theta$.

(c) Find the maximum likelihood (ML) estimator $\hat{\theta}_n$ of $\theta$.

(d) Show that $\hat{\theta}_n$ is a consistent estimator of $\theta$.

(e) What is the ML estimator of $\tau(\theta) = 1/\theta$?

# Probability

4. The *Probability* part of this comprehensive exam is about an ancient story of stochastic processes, which you are very familiar with something called *Simple Random Walk*. Even though a simple random walk is one of the most elementary stochastic processes, it covers substantial amount of basic concepts of the theory. Therefore it is a good starting point for our understanding of stochastic processes. While reading this story, you are asked to show your talent/knowledge to a series of questions here (5 in total, equally weighted toward 100 marks). I encourage you do your best to answer them all.

   Let $X_i,\ i = 1, 2, \ldots$ be a sequence of i.i.d. random variables with common distribution:

   $$P\{X_i = 1\} = p,\ P\{X_i = -1\} = 1 - p \equiv q,\ \text{ for some } 0 < p < 1.$$

   We call $\{S_n, n = 1, 2, \ldots\}$ a *simple random walk* starting from state $i$ if

   $$S_0 = i,\ \text{for some integer } i,\ \ S_n = \sum_{i=1}^{n} X_i.$$

   If we let $i = 0$, then our process starts from origin. Have you ever been curious about how the simple random walk will behave as $n \to \infty$ in probability and almost surely? More precisely, what are the probabilities

   $$\lim_{n\to\infty} P\{S_n = 0\}^{[1]} \text{ and } P\{\lim_{n\to\infty} S_n = 0\}^{[2]}?$$

   We know that, origin 0 is one of the states in the process state space $\mathcal{S} = \{0, \pm1, \pm2, ...\}$ of the simple random walk $S_n$, it would be interesting to know what will happen with the above limits if we replace the state 0 by an arbitrary state $i \in \mathcal{S}^{[3]}$?

   I am sure you know Markov processes, can you give a small proof to show that the simple random walk $S_n$ is indeed a Markov process[4]? Well, let us assume that you have proved the simple random walk $\{S_n,\ n = 1, 2, \ldots\}$ is a Markov process, and let us define $P_{ij}$ to be the *one-step transition probability* of this Markov process from state $i$ to state $j$, $i, j \in \mathcal{S}$:

   $$P_{ij} = P\{S_{n+1} = j | S_n = i\}.$$

The last question I would ask is: what are the values of $P_{ij},\ i,j \in S^{[5]}$?

Please write your reasoning, calculation, mathematical proof in detail to ensure credits to be earned. Good luck.

# Multivariate Analysis

5. Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$, and $\boldsymbol{X}_4$ be independent $N_p(\boldsymbol{\mu}, \Sigma)$ random vectors.

   (a) Find the marginal distributions for each of the random vectors

   $$\boldsymbol{V}_1 = (1/4)(\boldsymbol{X}_1 - \boldsymbol{X}_2 + \boldsymbol{X}_3 - \boldsymbol{X}_4)$$

   and

   $$\boldsymbol{V}_2 = (1/4)(\boldsymbol{X}_1 + \boldsymbol{X}_2 - \boldsymbol{X}_3 - \boldsymbol{X}_4).$$

   (b) Find the joint density of the random vectors $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$.

   (c) Find the conditional distribution of $\boldsymbol{V}_1$ given $\boldsymbol{V}_2 = \boldsymbol{a}$

6. Let $\boldsymbol{X}' = (\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \ \ldots, \ \boldsymbol{X}_p)$ be a random vector with covariance matrix $\Sigma$. Let the eigenvalues of $\boldsymbol{\Sigma}$ be $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ with corresponding normalized eigenvectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_p$. The $i$th principal component of $\boldsymbol{X}$ is defined as

   $$Y_i = \boldsymbol{a}_i'\boldsymbol{X}, \quad i = 1, 2, \ldots, p,$$

   where $\boldsymbol{a}_i$ maximizes $\text{Var}(Y_i)$ subject to the constraint $\boldsymbol{a}_i'\boldsymbol{a}_i = 1$.

   (a) Show that $\boldsymbol{a}_1 = \boldsymbol{e}_1$ for the first principal component.

   (b) Show that $\text{Var}(Y_1) = \lambda_1$.

   (c) Suppose that $Y_i = \boldsymbol{e}_i'\boldsymbol{X}$ is the ith principal component, show that

   $$Cov(Y_i, Y_k) = 0, \quad i \neq k.$$

   (d) Find the correlation between $Y_i$ and $X_k$, $i \neq k$.