Legen-*wait for it*-dary Papers

Ahan Chakraborty

May 2025

1 Probability

Problem 1. *(a)$Y \sim N_n(0, I)$ and P and A are matrices of order $n \times p$ and $n \times n$ respectively and A is symmetric. Then prove that the random variables $P^T Y$ and $Y^T AY$ are independent if and only if $AP^T = O$.*

(b) Find necessary and sufficient conditions for the independence of $Y^T AY$ and $Y^T BY$ with A, B are real symmetric matrices of order n. Note that the mentioned random variables do not follow the χ^2 distributions necessarily. *(10+15)*

Problem 2. *X and Y follow bivariate normal distribution with means μ_x, μ_Y, variances σ_X^2, σ_Y^2 and correlation-coefficient ρ.*

(a)Find the probability that

$$\frac{1}{1-\rho^2} \cdot [\left(\frac{X-\mu_x}{\sigma_X}\right)^2 - 2\rho\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right) + \left(\frac{Y-\mu_Y}{\sigma_Y}\right)^2] \leq r^2$$

in closed form of r and known mathematical constants.

(b) Prove that $\rho = \cos(q\pi)$ where $q = P[(X-\mu_X)(Y-\mu_Y) < 0]$

(c) Find correlation between X^2 and Y^2. *(10+10+5)*

Problem 3. *Suppose X is a non-negative random variable which has density f which is devreasing and is convex downwards. Determine the signs (positive or negative) of $E(\sin X)$ and $E(\cos X)$.* *(10)*

Problem 4. *Consider the two definitions of Multivariate-Normal distribution :*

(i)Y is a random vector of order n. It is said to follow a n-dimensional Multivariate Normal Distribution if any linear combination of the components of Y follows Normal Distribution.

(ii) For any $\mu \in \mathbb{R}^m$ and any NND $\Sigma_{m\times m}$, we define $N_m(\mu, \Sigma)$, the m -dimensional multivariate normal distribution as the distribution of $Y = AX + \mu$, where $X_{m\times 1}$ has IID N(0,1) components, and A is a full column rank matrix such that $\Sigma = AA^T$.

Prove that this two definitions are equivalent.

(Hint: for (i) $\implies$ (ii), first try to find the cf of the distribution.) *(20)*

2 Statistical Methods

Problem (1). *One boring morning, finding nothing to do, and to forget a certain someone, Gublu Sanyal decides that he will no longer use R. So, he sits to write a new statistical language named GSSC(Gublu Sanyal's Statistical Calculator), and all he with his semester 1 coding skills makes a simple calculator which can do Addition, Subtraction, Multiplication and Division and then he copies the source code of the function 'rnorm' from R itself, but forgets to copy the code of 'dnorm'. Now he has gone to his evil statistics professor to tell him about his new language (and says R is now obsolete), but the professor doesn't like Gublu making a new language and has told Gublu, if Gublu, with his language, fails to execute the class-assignment codes, will be penalized. Now the first task Gublu has to do is to find the normal density at point 1.223. Can you help him doing that ? (Gublu was just clever enough to include the option of running a for loop and some conditionals(if, else) in his language but not clever enough to carry a piece of Freedman, Pissani, Purves last page and also he couldn't recall the 1st 10 decimal digits of* π *and and so his calculator cannot calculate normal density directly from 1.223.)*

So, all Gublu can do, is to run some loops and some conditionals, use the basic 4 operations and the function rnorm to generate random numbers from the standard normal distribution, as many as he wants and he has to find the dnorm(1.223) or the normal density at 1.223. *(20)*

Problem (2). *By using direct or indirect methods,*

(a) Generate 100 independent samples from $Be(3,5)$ *distribution starting from* $Unif(0,1)$.

(b) Generate 100 independent samples from $Be(3.4,5.7)$ *distribution not using any random number generating function other than runif.* *(10+15)*

Problem (3). X, Y *are random variables and* Z *is a random vector, find the partial correlation of* X *and* Y *after controlling for* Z. *If* X, Y *have strong positive correlation, and* Z *is a random variable, is it possible to have the partial correlation of* X *and* Y *after correcting for the linear effect of* Z, *to be negative?*

Also show that if joint distribution of X, Y, Z *is* $(k+2)-$*dimensional Multivariate Normal, where* Z *is a* $k-$*dimensional random vector, then correcting for the linear effect of* Z *is same as considering the conditional distribution of* X *and* Y *given* Z, *i.e. Find the Dispersion matrix of the joint distribution of* X *and* Y *given* Z *and show that it is equal to the Dispersion Matrix of the residuals of* X *and* Y *when regressed upon* Z.

(15 +15)

Problem (4). *Let* $X_1, X_2, ...X_n$ *be a random sample from a* $Binom(k,p)$ *distribution where* k *is unknown and* p *is known. Provide a method for calculating the MLE for* k. *(10)*

Problem (5). *Give an example in each case, where the MLE fails to*

(a) be unique

(b) exist. *(5+10)*