Paper 4, Section II

27K Principles of Statistics

Assuming only the existence and properties of the univariate normal distribution, define $\mathcal{N}_p(\underline{\mu}, \Sigma)$, the *multivariate normal* distribution with mean (row-)vector $\underline{\mu}$ and dispersion matrix Σ; and $W_p(\nu; \Sigma)$, the *Wishart* distribution on integer $\nu > 1$ degrees of freedom and with scale parameter Σ. Show that, if $\underline{X} \sim \mathcal{N}_p(\underline{\mu}, \Sigma)$, $S \sim W_p(\nu; \Sigma)$, and $\underline{b}$ $(1 \times q)$, A $(p \times q)$ are fixed, then $\underline{b} + \underline{X}A \sim \mathcal{N}_q(\underline{b} + \underline{\mu}A, \Phi)$, $A^{\mathrm{T}}SA \sim W_p(\nu; \Phi)$, where $\Phi = A^{\mathrm{T}}\Sigma A$.

The random $(n \times p)$ matrix X has rows that are independently distributed as $\mathcal{N}_p(\underline{\mathrm{M}}, \Sigma)$, where both parameters $\underline{\mathrm{M}}$ and Σ are unknown. Let $\overline{\underline{X}} := n^{-1}\mathbf{1}^{\mathrm{T}}X$, where $\mathbf{1}$ is the $(n \times 1)$ vector of 1s; and $S^c := X^{\mathrm{T}}\Pi X$, with $\Pi := I_n - n^{-1}\mathbf{1}\mathbf{1}^{\mathrm{T}}$. State the joint distribution of $\overline{\underline{X}}$ and S^c given the parameters.

Now suppose $n > p$ and Σ is positive definite. *Hotelling's* T^2 is defined as

$$T^2 := n(\overline{\underline{X}} - \underline{\mathrm{M}}) \left(\overline{S}^c\right)^{-1} (\overline{\underline{X}} - \underline{\mathrm{M}})^{\mathrm{T}}$$

where $\overline{S}^c := S^c/\nu$ with $\nu := (n-1)$. Show that, for any values of $\underline{\mathrm{M}}$ and Σ,

$$\left(\frac{\nu - p + 1}{\nu p}\right) T^2 \sim F^p_{\nu-p+1},$$

the F distribution on p and $\nu - p + 1$ degrees of freedom.

[You may assume that:

1. If $S \sim W_p(\nu; \Sigma)$ and $\mathbf{a}$ is a fixed $(p \times 1)$ vector, then
$$\frac{\mathbf{a}^{\mathrm{T}}\Sigma^{-1}\mathbf{a}}{\mathbf{a}^{\mathrm{T}}S^{-1}\mathbf{a}} \sim \chi^2_{\nu-p+1}.$$

2. If $V \sim \chi^2_p$, $W \sim \chi^2_\lambda$ are independent, then
$$\frac{V/p}{W/\lambda} \sim F^p_\lambda.$$

]

Paper 3, Section II

27K Principles of Statistics

What is meant by a *convex decision problem*? State and prove a theorem to the effect that, in a convex decision problem, there is no point in randomising. [You may use standard terms without defining them.]

The sample space, parameter space and action space are each the two-point set $\{1,2\}$. The observable X takes value 1 with probability $2/3$ when the parameter $\Theta = 1$, and with probability $3/4$ when $\Theta = 2$. The loss function $L(\theta, a)$ is 0 if $a = \theta$, otherwise 1. Describe all the non-randomised decision rules, compute their risk functions, and plot these as points in the unit square. Identify an inadmissible non-randomised decision rule, and a decision rule that dominates it.

Show that the minimax rule has risk function $(8/17, 8/17)$, and is Bayes against a prior distribution that you should specify. What is its Bayes risk? Would a Bayesian with this prior distribution be bound to use the minimax rule?

Paper 1, Section II

28K Principles of Statistics

When the real parameter Θ takes value θ, variables $X_1, X_2, \ldots$ arise independently from a distribution P_θ having density function $p_\theta(x)$ with respect to an underlying measure μ. Define the *score variable* $U_n(\theta)$ and the *information function* $I_n(\theta)$ for estimation of Θ based on $\boldsymbol{X}^n := (X_1, \ldots, X_n)$, and relate $I_n(\theta)$ to $i(\theta) := I_1(\theta)$.

State and prove the Cramér–Rao inequality for the variance of an unbiased estimator of Θ. Under what conditions does this inequality become an equality? What is the form of the estimator in this case? [You may assume $\mathbb{E}_\theta\{U_n(\theta)\} = 0$, $\mathrm{var}_\theta\{U_n(\theta)\} = I_n(\theta)$, and any further required regularity conditions, without comment.]

Let $\widehat{\Theta}_n$ be the maximum likelihood estimator of Θ based on $\boldsymbol{X}^n$. What is the asymptotic distribution of $n^{\frac{1}{2}}(\widehat{\Theta}_n - \Theta)$ when $\Theta = \theta$?

Suppose that, for each n, $\widehat{\Theta}_n$ is unbiased for Θ, and the variance of $n^{\frac{1}{2}}(\widehat{\Theta}_n - \Theta)$ is exactly equal to its asymptotic variance. By considering the estimator $\alpha\widehat{\Theta}_k + (1-\alpha)\widehat{\Theta}_n$, or otherwise, show that, for $k < n$, $\mathrm{cov}_\theta(\widehat{\Theta}_k, \widehat{\Theta}_n) = \mathrm{var}_\theta(\widehat{\Theta}_n)$.

Paper 2, Section II

28K Principles of Statistics

Describe the *Weak Sufficiency Principle* (WSP) and the *Strong Sufficiency Principle* (SSP). Show that Bayesian inference with a fixed prior distribution respects WSP.

A parameter Φ has a prior distribution which is normal with mean 0 and precision (inverse variance) h_Φ. Given $\Phi = \phi$, further parameters $\boldsymbol{\Theta} := (\Theta_i : i = 1, \ldots, I)$ have independent normal distributions with mean ϕ and precision h_Θ. Finally, given both $\Phi = \phi$ and $\boldsymbol{\Theta} = \boldsymbol{\theta} := (\theta_1, \ldots, \theta_I)$, observables $\boldsymbol{X} := (X_{ij} : i = 1, \ldots, I; j = 1, \ldots, J)$ are independent, X_{ij} being normal with mean θ_i, and precision h_X. The precision parameters (h_Φ, h_Θ, h_X) are all fixed and known. Let $\overline{\boldsymbol{X}} := (\overline{X}_1, \ldots, \overline{X}_I)$, where $\overline{X}_i := \sum_{j=1}^{J} X_{ij}/J$. Show, directly from the definition of sufficiency, that $\overline{\boldsymbol{X}}$ is sufficient for $(\Phi, \boldsymbol{\Theta})$. [You may assume without proof that, if $Y_1, \ldots, Y_n$ have independent normal distributions with the same variance, and $\overline{Y} := n^{-1} \sum_{i=1}^{n} Y_i$, then the vector $(Y_1 - \overline{Y}, \ldots, Y_n - \overline{Y})$ is independent of $\overline{Y}$.]

For data-values $\boldsymbol{x} := (x_{ij} : i = 1, \ldots, I; j = 1, \ldots, J)$, determine the joint distribution, Π_ϕ say, of $\boldsymbol{\Theta}$, given $\boldsymbol{X} = \boldsymbol{x}$ and $\Phi = \phi$. What is the distribution of Φ, given $\boldsymbol{\Theta} = \boldsymbol{\theta}$ and $\boldsymbol{X} = \boldsymbol{x}$?

Using these results, describe clearly how Gibbs sampling combined with Rao–Blackwellisation could be applied to estimate the posterior joint distribution of $\boldsymbol{\Theta}$, given $\boldsymbol{X} = \boldsymbol{x}$.