

Change of Measure Applications in Nonparametric Statistics



Mayer Alvo

Abstract Neyman [7] was the first to propose a change in measure in the context of goodness of fit problems. This provided an alternative density to the one for the null hypothesis. Hoeffding introduced a change of measure formula for the ranks of the observed data which led to obtaining locally most powerful rank tests. In this paper, we review these methods and propose a new approach which leads on the one hand to new derivations of existing statistics. On the other hand, we exploit these methods to obtain Bayesian applications for ranking data.

Keywords Ranks · Change of measure · Bayesian methods

Mathematics Subject Classification (2010) 62F07 · 62G86 · 62H11

1 Introduction

In a landmark paper, [7] considered the nonparametric goodness of fit problem and introduced the notion of smooth tests of fit by proposing a parametric family of alternative densities to the null hypothesis. In this article, we describe a number of applications of this change of measure. Hence, we obtain a new derivation of the well-known Friedman statistic as the locally most powerful test in an embedded family of distributions.

2 Smooth Models

Suppose that the probability mass function of a discrete k -dimensional random vector X is given by

M. Alvo (✉)

Department of Mathematics, Statistics University of Ottawa, Ottawa, ON, Canada
e-mail: malvo@uottawa.ca

© Springer Nature Switzerland AG 2020

M. La Rocca et al. (eds.), *Nonparametric Statistics*, Springer Proceedings
in Mathematics & Statistics 339, https://doi.org/10.1007/978-3-030-57306-5_2

$$\pi(\mathbf{x}_j; \boldsymbol{\theta}) = \exp(\boldsymbol{\theta}' \mathbf{x}_j - K(\boldsymbol{\theta})) p_j, \quad j = 1, \dots, m, \quad (1)$$

where \mathbf{x}_j is the j th value of \mathbf{X} and $\mathbf{p} = (p_j)'$ denotes the vector of probabilities when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Here $K(\boldsymbol{\theta})$ is a normalizing constant for which

$$\sum_j \pi(\mathbf{x}_j; \boldsymbol{\theta}) = 1.$$

We see that the model in (1) prescribes a change of measure from the null to the alternative hypothesis. Let $\mathbf{T} = [\mathbf{x}_1, \dots, \mathbf{x}_m]$ be the $k \times m$ matrix of possible vector values of \mathbf{X} . Then under the distribution specified by \mathbf{p} ,

$$\boldsymbol{\Sigma} \equiv \text{Cov}_p(\mathbf{X}) = E_p[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])'] \quad (2)$$

$$= \mathbf{T}(\text{diag}(\mathbf{p}))\mathbf{T}' - (\mathbf{T}\mathbf{p})(\mathbf{T}\mathbf{p})', \quad (3)$$

where the expectations are with respect to the model (1). This particular situation arises often when dealing with the nonparametric randomized block design. Define

$$\boldsymbol{\pi}(\boldsymbol{\theta}) = (\pi(\mathbf{x}_1; \boldsymbol{\theta}), \dots, \pi(\mathbf{x}_m; \boldsymbol{\theta}))'$$

and suppose that we would like to test

$$H_0 : \boldsymbol{\theta} = \mathbf{0} \text{ vs } H_1 : \boldsymbol{\theta} \neq \mathbf{0}.$$

Letting \mathbf{N} denote a multinomial random vector with parameters $(n, \boldsymbol{\pi}(\boldsymbol{\theta}))$, we see that the log likelihood as a function of $\boldsymbol{\theta}$ is, apart from a constant, proportional to

$$\begin{aligned} \sum_{j=1}^m n_j \log(\pi(\mathbf{x}_j; \boldsymbol{\theta})) &= \sum_{j=1}^m n_j (\boldsymbol{\theta}' \mathbf{x}_j - K(\boldsymbol{\theta})) \\ &= \boldsymbol{\theta}' \left(\sum_{j=1}^m n_j \mathbf{x}_j \right) - nK(\boldsymbol{\theta}). \end{aligned}$$

The score vector under the null hypothesis is then given by

$$\begin{aligned} U(\boldsymbol{\theta}; \mathbf{X}) &= \sum_{j=1}^m N_j \left(\frac{1}{\pi_j(\boldsymbol{\theta})} \frac{\partial \pi_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \\ &= \mathbf{T}(\mathbf{N} - n\mathbf{p}). \end{aligned}$$

Under the null hypothesis,

$$E[U(\boldsymbol{\theta}; \mathbf{X})] = \mathbf{0},$$

and the score statistic is given by

$$\frac{1}{n} [\mathbf{T}(N - n\mathbf{p})]' \boldsymbol{\Sigma}^{-1} [\mathbf{T}(N - n\mathbf{p})] = \frac{1}{n} (N - n\mathbf{p})' (\mathbf{T}' \boldsymbol{\Sigma}^{-1} \mathbf{T}) (N - n\mathbf{p}) \xrightarrow{\mathcal{L}} \chi_r^2, \quad (4)$$

where $r = \text{rank}(\mathbf{T}' \boldsymbol{\Sigma}^{-1} \mathbf{T})$.

In the one-sample ranking problem whereby a group of judges are each asked to rank a set of t objects in accordance with some criterion, let $\mathcal{P} = \{\nu_j, j = 1, \dots, t!\}$ be the space of all $t!$ permutations of the integers $1, 2, \dots, t$ and let the probability mass distribution defined on \mathcal{P} be given by

$$\mathbf{p} = (p_1, \dots, p_{t!}),$$

where $p_j = \Pr(\nu_j)$. Conceptually, each judge selects a ranking ν in accordance with the probability mass distribution \mathbf{p} . In order to test the null hypothesis that each of the rankings are selected with equal probability, that is,

$$H_0 : \mathbf{p} = \mathbf{p}_0 \text{ vs } H_1 : \mathbf{p} \neq \mathbf{p}_0, \quad (5)$$

where $\mathbf{p}_0 = \frac{1}{t!} \mathbf{1}$, define a k -dimensional vector score function $\mathbf{X}(\nu)$ on the space \mathcal{P} and following (1), let its smooth probability mass function be given as

$$\pi(\mathbf{x}_j; \boldsymbol{\theta}) = \exp(\boldsymbol{\theta}' \mathbf{x}_j - K(\boldsymbol{\theta})) \frac{1}{t!}, \quad j = 1, \dots, t! \quad (6)$$

where $\boldsymbol{\theta}$ is a t -dimensional vector, $K(\boldsymbol{\theta})$ is a normalizing constant and \mathbf{x}_j is a t -dimensional score vector to be specified in (8). Since

$$\sum_{j=1}^{t!} \pi(\mathbf{x}_j; \boldsymbol{\theta}) = 1$$

it can be seen that $K(\mathbf{0}) = 0$ and hence the hypotheses in (5) are equivalent to testing

$$H_0 : \boldsymbol{\theta} = \mathbf{0} \text{ vs } H_1 : \boldsymbol{\theta} \neq \mathbf{0}. \quad (7)$$

It follows that the log likelihood function is proportional to

$$l(\boldsymbol{\theta}) \sim n [\boldsymbol{\theta}' \hat{\boldsymbol{\eta}} - K(\boldsymbol{\theta})],$$

where

$$\hat{\boldsymbol{\eta}} = \left[\sum_{j=1}^{t!} \mathbf{x}_j \hat{p}_{nj} \right], \quad \hat{p}_{nj} = \frac{n_j}{n}$$

and n_j represents the number of observed occurrences of the ranking ν_j . The Rao score statistic evaluated at $\theta = \mathbf{0}$ is

$$\begin{aligned} U(\theta; \mathbf{X}) &= n \frac{\partial}{\partial \theta} [\theta' \hat{\boldsymbol{\eta}} - K(\mathbf{0})] \\ &= n \left[\hat{\boldsymbol{\eta}} - \frac{\partial}{\partial \theta} K(\mathbf{0}) \right], \end{aligned}$$

whereas the information matrix is

$$\mathbf{I}(\theta) = -n \left[\frac{\partial^2}{\partial \theta^2} K(\mathbf{0}) \right].$$

The test then rejects the null hypothesis whenever

$$n^2 \left[\hat{\boldsymbol{\eta}} - \frac{\partial}{\partial \theta} K(\mathbf{0}) \right]' \mathbf{I}^{-1}(\mathbf{0}) \left[\hat{\boldsymbol{\eta}} - \frac{\partial}{\partial \theta} K(\mathbf{0}) \right] > \chi_f^2(\alpha),$$

where $\chi_f^2(\alpha)$ is the upper $100(1 - \alpha)\%$ critical value of a chi square distribution with $f = \text{rank}(\mathbf{I}(\theta))$ degrees of freedom. We note that the test just obtained is the locally most powerful test of H_0 .

Specializing this test statistic to the Spearman score function of adjusted ranks

$$\mathbf{x}_j = \left(\nu_j(1) - \frac{t+1}{2}, \dots, \nu_j(t) - \frac{t+1}{2} \right)', \quad j = 1, \dots, t!, \quad (8)$$

we can show that the Rao score statistic is the well-known Friedman test [5].

$$W = \frac{12n}{t(t+1)} \sum_{i=1}^t \left[\bar{R}_i - \frac{t+1}{2} \right]^2, \quad (9)$$

where \bar{R}_i is the average of the ranks assigned to the i th object.

2.1 The Two-Sample Ranking Problem

The approach just described can be used to deal with the two-sample ranking problem assuming again the Spearman score function. Let X_1, X_2 be two independent random vectors whose distributions as in the one sample case are expressed for simplicity as

$$\pi(\mathbf{x}_j; \theta_l) = \exp \{ \theta_l' \mathbf{x}_j - K(\theta_l) \} p_l(j), \quad j = 1, \dots, t!, l = 1, 2,$$

where $\boldsymbol{\theta}_l = (\theta_{l1}, \dots, \theta_{lt})'$ represents the vector of parameters for population l . We are interested in testing

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \text{ vs } H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2.$$

The probability distribution $\{p_l(j)\}$ represents an unspecified null situation. Define

$$\hat{\boldsymbol{p}}_l = \left(\frac{n_{l1}}{n_l}, \dots, \frac{n_{lt}}{n_l} \right)',$$

where n_{ij} represents the number of occurrences of the ranking ν_j in sample l .

Also, for $l = 1, 2$, set $\sum_j n_{ij} \equiv n_l$, $\boldsymbol{\gamma} = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$ and

$$\boldsymbol{\theta}_l = \boldsymbol{m} + b_l \boldsymbol{\gamma},$$

where

$$\boldsymbol{m} = \frac{n_1 \boldsymbol{\theta}_1 + n_2 \boldsymbol{\theta}_2}{n_1 + n_2}, b_1 = \frac{n_2}{n_1 + n_2}, b_2 = -\frac{n_1}{n_1 + n_2}.$$

Let $\boldsymbol{\Sigma}_l$ be the covariance matrix of \boldsymbol{X}_l under the null hypothesis defined as

$$\boldsymbol{\Sigma}_l = \boldsymbol{\Pi}_l - \boldsymbol{p}_l \boldsymbol{p}_l',$$

where $\boldsymbol{\Pi}_l = \text{diag}(p_l(1), \dots, p_l(t!))$ and $\boldsymbol{p}_l = (p_l(1), \dots, p_l(t!))'$. The logarithm of the likelihood L as a function of $(\boldsymbol{m}, \boldsymbol{\gamma})$ is proportional to

$$\log L(\boldsymbol{m}, \boldsymbol{\gamma}) \sim \sum_{l=1}^2 \sum_{j=1}^{t!} n_{lj} \{(\boldsymbol{m} + b_l \boldsymbol{\gamma})' \boldsymbol{x}_j - K(\boldsymbol{\theta}_l)\}.$$

In order to test

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \text{ vs } H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$$

we calculate the Rao score test statistic which is given by

$$n (\boldsymbol{T}_S \hat{\boldsymbol{p}}_1 - \boldsymbol{T}_S \hat{\boldsymbol{p}}_2)' \hat{\boldsymbol{D}} (\boldsymbol{T}_S \hat{\boldsymbol{p}}_1 - \boldsymbol{T}_S \hat{\boldsymbol{p}}_2). \quad (10)$$

It can be shown to have asymptotically a χ_f^2 whenever $n_l/n \rightarrow \lambda_l > 0$ as $n \rightarrow \infty$, where $n = n_1 + n_2$. Here $\hat{\boldsymbol{D}}$ is the Moore–Penrose inverse of $\boldsymbol{T}_S \hat{\boldsymbol{\Sigma}} \boldsymbol{T}_S'$ and $\hat{\boldsymbol{\Sigma}}$ is a consistent estimator of $\boldsymbol{\Sigma} = \frac{\boldsymbol{\Sigma}_1}{\lambda_1} + \frac{\boldsymbol{\Sigma}_2}{\lambda_2}$ and f is the rank of $\hat{\boldsymbol{D}}$, as required.

2.2 The Use of Penalized Likelihood

In the previous sections, it was possible to derive test statistics for the one and two-sample ranking problems by means of the change of measure paradigm. This paradigm may be exploited to obtain new results for the ranking problems. Specifically, we consider a negative penalized likelihood function defined to be the negative log likelihood function subject to a constraint on the parameters which is then minimized with respect to the parameter. This approach yields further insight into ranking problems.

For the one-sample ranking problem, let

$$\Lambda(\boldsymbol{\theta}, c) = -\boldsymbol{\theta}' \left[\sum_{j=1}^{t!} n_j \mathbf{x}_j \right] + nK(\boldsymbol{\theta}) + \lambda \left(\sum_{i=1}^t \theta_i^2 - c \right) \quad (11)$$

represent the penalizing function for some prescribed values of the constant c . We shall assume for simplicity that $\|\mathbf{x}_j\| = 1$. When t is large (say $t \geq 10$), the computation of the exact value of the normalizing constant $K(\boldsymbol{\theta})$ involves a summation of $t!$ terms. [6] noted the resemblance of (6) to the continuous von Mises-Fisher density

$$f(\mathbf{x}; \boldsymbol{\theta}) = \frac{\|\boldsymbol{\theta}\|^{\frac{t-3}{2}} \exp(\boldsymbol{\theta}'\mathbf{x})}{2^{\frac{t-3}{2}} t! I_{\frac{t-3}{2}}(\|\boldsymbol{\theta}\|) \Gamma(\frac{t-1}{2})},$$

where $\|\boldsymbol{\theta}\|$ is the norm of $\boldsymbol{\theta}$ and \mathbf{x} is on the unit sphere and $I_\nu(z)$ is the modified Bessel function of the first kind given by

$$I_\nu(z) = \sum_{k=0}^{\infty} \frac{1}{\Gamma(k+1)\Gamma(\nu+k+1)} \left(\frac{z}{2}\right)^{2k+\nu}.$$

This seems to suggest the approximation of the constant $K(\boldsymbol{\theta})$ by

$$\exp(-K(\boldsymbol{\theta})) \approx \frac{1}{t!} \cdot \frac{\|\boldsymbol{\theta}\|^{\frac{t-3}{2}}}{2^{\frac{t-3}{2}} I_{\frac{t-3}{2}}(\|\boldsymbol{\theta}\|) \Gamma(\frac{t-1}{2})}.$$

In [1], penalized likelihood was used in ranking situations to obtain further insight into the differences between groups of rankers.

3 Bayesian Models for Ranking Data

The fact that the model in (1) is itself parametric in nature leads one to consider an extension to Bayesian considerations. Let $\mathbf{R} = (R(1), \dots, R(t))'$ be a ranking of t items, labeled $1, \dots, t$ and define the standardized rankings as

$$\mathbf{y} = \left(\mathbf{R} - \frac{t+1}{2} \mathbf{1} \right) / \sqrt{\frac{t(t^2-1)}{12}},$$

where \mathbf{y} is the $t \times 1$ vector with $\|\mathbf{y}\| \equiv \sqrt{\mathbf{y}'\mathbf{y}} = 1$. We consider the following more general ranking model:

$$\pi(\mathbf{y}|\kappa, \boldsymbol{\theta}) = C(\kappa, \boldsymbol{\theta}) \exp \{ \kappa \boldsymbol{\theta}' \mathbf{y} \},$$

where the parameter $\boldsymbol{\theta}$ is a $t \times 1$ vector with $\|\boldsymbol{\theta}\| = 1$, parameter $\kappa \geq 0$, and $C(\kappa, \boldsymbol{\theta})$ is the normalizing constant. This model has a close connection to the distance-based models considered in [3]. Here, $\boldsymbol{\theta}$ is a real-valued vector, representing a consensus view of the relative preference of the items from the individuals. Since both $\|\boldsymbol{\theta}\| = 1$ and $\|\mathbf{y}\| = 1$, the term $\boldsymbol{\theta}' \mathbf{y}$ can be seen as $\cos \phi$ where ϕ is the angle between the consensus score vector $\boldsymbol{\theta}$ and the observation \mathbf{y} . The probability of observing a ranking is proportional to the cosine of the angle from the consensus score vector. The parameter κ can be viewed as a concentration parameter. For small κ , the distribution of rankings will appear close to a uniform whereas for larger values of κ , the distribution of rankings will be more concentrated around the consensus score vector. We call this new model an *angle-based ranking model*.

To compute the normalizing constant $C(\kappa, \boldsymbol{\theta})$, let \mathcal{P}_t be the set of all possible permutations of the integers $1, \dots, t$. Then

$$(C(\kappa, \boldsymbol{\theta}))^{-1} = \sum_{\mathbf{y} \in \mathcal{P}} \exp \{ \kappa \boldsymbol{\theta}' \mathbf{y} \}. \tag{12}$$

Notice that the summation is over the $t!$ elements in \mathcal{P} . When t is large, say greater than 15, the exact calculation of the normalizing constant is prohibitive. Using the fact that the set of $t!$ permutations lie on a sphere in $(t-1)$ -space, our model resembles the continuous von Mises-Fisher distribution, abbreviated as $vMF(\mathbf{x}|\mathbf{m}, \kappa)$, which is defined on a $(p-1)$ unit sphere with mean direction \mathbf{m} and concentration parameter κ :

$$p(\mathbf{x}|\kappa, \mathbf{m}) = V_p(\kappa) \exp(\kappa \mathbf{m}' \mathbf{x}),$$

where

$$V_p(\kappa) = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)},$$

and $I_a(\kappa)$ is the modified Bessel function of the first kind with order a . Consequently, we may approximate the sum in (12) by an integral over the sphere:

$$C(\kappa, \boldsymbol{\theta}) \simeq C_t(\kappa) = \frac{\kappa^{\frac{t-3}{2}}}{2^{\frac{t-3}{2}} t! I_{\frac{t-3}{2}}(\kappa) \Gamma(\frac{t-1}{2})},$$

where $\Gamma(\cdot)$ is the gamma function. In ([9], it is shown that this approximation is very accurate for values of κ ranging from 0.01 to 2 and t ranging from 4 to 11. Moreover, the error drops rapidly as t increases. Note that this approximation allows us to approximate the first and second derivatives of $\log C$ which can facilitate our computation in what follows.

3.1 Maximum Likelihood Estimation (MLE) of Our Model

Let $Y = \{y_1, \dots, y_N\}$ be a random sample of N standardized rankings drawn from $p(y|\kappa, \theta)$. The log likelihood of (κ, θ) is then given by

$$l(\kappa, \theta) = n \log C_t(\kappa) + \sum_{i=1}^n \kappa \theta' y_i. \quad (13)$$

Maximizing (13) subject to $\|\theta\| = 1$ and $\kappa \geq 0$, we find that the maximum likelihood estimator of θ is given by $\hat{\theta}_{MLE} = \frac{\sum_{i=1}^N y_i}{\|\sum_{i=1}^N y_i\|}$, and $\hat{\kappa}$ is the solution of

$$A_t(\kappa) \equiv \frac{-C'_t(\kappa)}{C_t(\kappa)} = \frac{I_{t-1}(\kappa)}{I_{t-3}(\kappa)} = \frac{\left\| \sum_{i=1}^N y_i \right\|}{N} \equiv r. \quad (14)$$

A simple approximation to the solution of (14) following [4] is given by

$$\hat{\kappa}_{MLE} = \frac{r(t-1-r^2)}{1-r^2}.$$

A more precise approximation can be obtained from a few iterations of Newton's method. Using the method suggested by [8], starting from an initial value κ_0 , we can recursively update κ by iteration:

$$\kappa_{i+1} = \kappa_i - \frac{A_t(\kappa_i) - r}{1 - A_t(\kappa_i)^2 - \frac{t-2}{\kappa_i} A_t(\kappa_i)}, \quad i = 0, 1, 2, \dots$$

3.2 One-Sample Bayesian Method with Conjugate Prior

Taking a Bayesian approach, we consider the following conjugate prior for (κ, θ) as

$$p(\kappa, \theta) \propto [C_t(\kappa)]^{\nu_0} \exp \{ \beta_0 \kappa m'_0 \theta \}, \quad (15)$$

where $\|\mathbf{m}_0\| = 1, \nu_0, \beta_0 \geq 0$. Given \mathbf{y} , the posterior density of $(\kappa, \boldsymbol{\theta})$ can be expressed by

$$p(\alpha, \boldsymbol{\theta}|\mathbf{y}) \propto \exp\{\beta\kappa\mathbf{m}'\boldsymbol{\theta}\} V_t(\beta\kappa) \cdot \frac{[C_t(\kappa)]^{N+\nu_0}}{V_t(\beta\kappa)},$$

where $\mathbf{m} = (\beta_0\mathbf{m}_0 + \sum_{i=1}^N \mathbf{y}_i) \beta^{-1}$, $\beta = \|\beta_0\mathbf{m}_0 + \sum_{i=1}^N \mathbf{y}_i\|$. The posterior density can be factored as

$$p(\kappa, \boldsymbol{\theta}|\mathbf{y}) = p(\boldsymbol{\theta}|\kappa, \mathbf{y})p(\kappa|\mathbf{y}), \quad (16)$$

where $p(\boldsymbol{\theta}|\kappa, \mathbf{y}) \sim vMF(\boldsymbol{\theta}|\mathbf{m}, \beta\kappa)$ and

$$p(\kappa|\mathbf{y}) \propto \frac{[C_t(\kappa)]^{N+\nu_0}}{V_t(\beta\kappa)} = \frac{\kappa^{\frac{t-3}{2}(\nu_0+N)} I_{\frac{t-2}{2}}(\beta\kappa)}{\left[I_{\frac{t-3}{2}}(\kappa) \right]^{\nu_0+N} (\beta\kappa)^{\frac{t-2}{2}}}.$$

The normalizing constant for $p(\kappa|\mathbf{y})$ is not available in closed form. For reasons explained in [9], we approximate the posterior distribution using the method of variational inference (abbreviated VI from here on). Variational inference provides a deterministic approximation to an intractable posterior distribution through optimization. We first adopt a joint vMF- Gamma distribution as the prior for $(\kappa, \boldsymbol{\theta})$:

$$\begin{aligned} p(\kappa, \boldsymbol{\theta}) &= p(\boldsymbol{\theta}|\kappa)p(\kappa) \\ &= vMF(\boldsymbol{\theta}|\mathbf{m}_0, \beta_0\kappa) \text{Gamma}(\kappa|a_0, b_0), \end{aligned}$$

where $\text{Gamma}(\kappa|a_0, b_0)$ is the Gamma density function with shape parameter a_0 and rate parameter b_0 (i.e., mean equal to $\frac{a_0}{b_0}$), and $p(\boldsymbol{\theta}|\kappa) = vMF(\boldsymbol{\theta}|\mathbf{m}_0, \beta_0\kappa)$. The choice of $\text{Gamma}(\kappa|a_0, b_0)$ for $p(\kappa)$ is motivated by the fact that for large values of κ , $p(\kappa)$ in (15) tends to take the shape of a Gamma density. In fact, for large values of κ , $I_{\frac{t-3}{2}}(\kappa) \simeq \frac{e^\kappa}{\sqrt{2\pi\kappa}}$, and hence $p(\kappa)$ becomes the Gamma density with shape $(\nu_0 - 1)\frac{t-2}{2} + 1$ and rate $\nu_0 - \beta_0$:

$$p(\kappa) \propto \frac{[C_t(\kappa)]^{\nu_0}}{V_t(\beta_0\kappa)} \propto \kappa^{(\nu_0-1)\frac{t-2}{2}} \exp(-(\nu_0 - \beta_0)\kappa).$$

Using the variational inference framework, [9] showed that the optimal posterior distribution of $\boldsymbol{\theta}$ conditional on κ is a von Mises-Fisher distribution $vMF(\boldsymbol{\theta}|\mathbf{m}, \kappa\beta)$ where

$$\beta = \left\| \beta_0\mathbf{m}_0 + \sum_{i=1}^N \mathbf{y}_i \right\| \quad \text{and} \quad \mathbf{m} = \left(\beta_0\mathbf{m}_0 + \sum_{i=1}^N \mathbf{y}_i \right) \beta^{-1}.$$

The optimal posterior distribution of κ is a $\text{Gamma}(\kappa|a, b)$ with shape a and rate b with

$$a = a_0 + N \left(\frac{t-3}{2} \right) + \beta \bar{\kappa} \left[\frac{\partial}{\partial \beta \bar{\kappa}} \ln I_{\frac{t-2}{2}}(\beta \bar{\kappa}) \right], \quad (17)$$

$$b = b_0 + N \frac{\partial}{\partial \bar{\kappa}} I_{\frac{t-3}{2}}(\bar{\kappa}) + \beta_0 \left[\frac{\partial}{\partial \beta_0 \bar{\kappa}} \ln I_{\frac{t-2}{2}}(\beta_0 \bar{\kappa}) \right]. \quad (18)$$

Finally, the posterior mode $\bar{\kappa}$ can be obtained from the previous iteration as

$$\bar{\kappa} = \begin{cases} \frac{a-1}{b} & \text{if } a > 1, \\ \frac{a}{b} & \text{otherwise.} \end{cases} \quad (19)$$

3.3 Two-Sample Bayesian Method with Conjugate Prior

Let $\mathbf{Y}_i = \{y_{i1}, \dots, y_{iN_i}\}$ for $i = 1, 2$, be two independent random samples of standardized rankings each drawn, respectively, from $p(y_i|\kappa_i, \theta_i)$. Taking a Bayesian approach, we assume that conditional on κ , there are independent von Mises conjugate priors, respectively, for (θ_1, θ_2) as

$$p(\theta_i|\kappa) \propto [C_i(\kappa)]^{\nu_{i0}} \exp \{ \beta_{i0} \kappa \mathbf{m}_{i0}^T \theta_i \},$$

where $\|\mathbf{m}_{i0}\| = 1$, $\nu_{i0}, \beta_{i0} \geq 0$. We shall be interested in computing the Bayes factor when considering two models. Under model 1, denoted M_1 , $\theta_1 = \theta_2$ whereas under model 2, denoted M_2 , equality is not assumed. The Bayes factor comparing the two models is defined to be

$$\begin{aligned} B_{21} &= \frac{\int p(\mathbf{y}_1|\kappa, \theta_1) p(\mathbf{y}_2|\kappa, \theta_2) p(\theta_1|\kappa) p(\theta_2|\kappa) d\theta_1 d\theta_2 d\kappa}{\int p(\mathbf{y}_1|\kappa, \theta) p(\mathbf{y}_2|\kappa, \theta) p(\theta|\kappa) d\theta d\kappa} \\ &= \frac{\int [\int p(\mathbf{y}_1|\kappa, \theta_1) p(\theta_1|\kappa) d\theta_1] [\int p(\mathbf{y}_2|\kappa, \theta_2) p(\theta_2|\kappa) d\theta_2] d\kappa}{\int p(\mathbf{y}_1|\kappa, \theta) p(\mathbf{y}_2|\kappa, \theta) p(\theta|\kappa) d\theta d\kappa}. \end{aligned}$$

The Bayes factor enables us to compute the posterior odds of model 2 to model 1. We first deal with the denominator in B_{21} . Under M_1 , we assume a joint von Mises-Fisher prior on θ and a Gamma prior on κ :

$$p(\theta, \kappa) = v M F(\theta|m_0, \beta_0 \kappa) G(\kappa|a_0, b_0).$$

Hence,

$$\begin{aligned} \int p(\mathbf{y}_1|\kappa, \theta) p(\mathbf{y}_2|\kappa, \theta) p(\theta|\kappa) d\theta d\kappa &= \int C_i^N(\kappa) \exp \{ \beta \kappa \theta^T \mathbf{m} \} V_i(\beta_0 \kappa) G(\kappa|a_0, b_0) d\theta d\kappa \\ &= \int C_i^N(\kappa) V_i(\beta_0 \kappa) V_i^{-1}(\beta \kappa) G(\kappa|a_0, b_0) d\kappa, \end{aligned}$$

where $N = N_1 + N_2$ and

$$\mathbf{m} = \left[\beta_0 \mathbf{m}_0 + \sum_{i=1}^2 \sum_{j=1}^{N_i} \mathbf{y}_{ij} \right] \beta^{-1}, \beta = \|\mathbf{m}\|.$$

Now,

$$\begin{aligned} \int C_t^N(\kappa) V_t(\beta_0 \kappa) V_t^{-1}(\beta \kappa) G(\kappa | a_0, b_0) d\kappa &= C \left(\frac{\beta_0}{\beta} \right)^{\frac{t-2}{2}} \int \left[\frac{\kappa^{a_0+N(\frac{t-3}{2})-1} e^{-b_0 \kappa} I_{\frac{t-2}{2}}(\beta \kappa)}{I_{\frac{t-2}{2}}(\beta_0 \kappa) I_{\frac{t-3}{2}}(\kappa)} \right] d\kappa \\ &\approx C \left(\frac{\beta_0}{\beta} \right)^{\frac{t-2}{2}} \int \kappa^{a-1} e^{-b \kappa} d\kappa, \end{aligned}$$

where in the last step, we used the method of variational inference as an approximation, with

$$\begin{aligned} C &= \frac{b_0^{a_0}}{\Gamma(a_0)} \left(2^N \left(\frac{t-3}{2} \right)! \Gamma^N \left(\frac{t-1}{2} \right) \right)^{-1} \\ a_1 &= a_0 + N \left(\frac{t-3}{2} \right) + \beta \bar{\kappa} \left[\frac{\partial}{\partial \beta \kappa} \ln I_{\frac{t-2}{2}}(\beta \bar{\kappa}) \right], \\ b_1 &= b_0 + N \frac{\partial}{\partial \kappa} I_{\frac{t-3}{2}}(\bar{\kappa}) + \beta_0 \left[\frac{\partial}{\partial \beta_0 \kappa} \ln I_{\frac{t-2}{2}}(\beta_0 \bar{\kappa}) \right] \end{aligned}$$

and the posterior mode $\bar{\kappa}$ is

$$\bar{\kappa} = \begin{cases} \frac{a_1-1}{b_1} & \text{if } a_1 > 1, \\ \frac{a_1}{b_1} & \text{otherwise.} \end{cases}$$

It follows that the denominator of B_{21} is

$$C \left(\frac{\beta_0}{\beta} \right)^{\frac{t-2}{2}} \frac{\Gamma(a_1)}{b_1^{a_1}}.$$

For the numerator, we shall assume that conditional on κ , there are independent von Mises conjugate priors, respectively, for θ_1, θ_2 given by

$$p(\theta_i | \kappa) \propto [C_t(\kappa)] \exp \{ \beta_0 \kappa \mathbf{m}_0^T \theta_i \}, i = 1, 2$$

where $\|\mathbf{m}_0\| = 1, \beta_0 \geq 0$. Hence,

$$\begin{aligned}
& \int \left[\int p(\mathbf{y}_1|\kappa, \boldsymbol{\theta}_1) p(\boldsymbol{\theta}_1|\kappa) d\boldsymbol{\theta}_1 \right] \left[\int p(\mathbf{y}_2|\kappa, \boldsymbol{\theta}_2) p(\boldsymbol{\theta}_2|\kappa) d\boldsymbol{\theta}_2 \right] d\kappa \\
&= \int C_t^N(\kappa) \exp\{\beta_1 \kappa \boldsymbol{\theta}_1^T \mathbf{m}_1\} \exp\{\beta_2 \kappa \boldsymbol{\theta}_2^T \mathbf{m}_2\} V_t^2(\beta_0 \kappa) G(\kappa|a_0, b_0) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 d\kappa \\
&= \int C_t^N(\kappa) V_t^{-1}(\beta_1 \kappa) V_t^{-1}(\beta_2 \kappa) V_t^2(\beta_0 \kappa) G(\kappa|a_0, b_0) d\kappa \\
&= C \left(\frac{\beta_0}{\beta_1} \right)^{\frac{t-2}{2}} \left(\frac{\beta_0}{\beta_2} \right)^{\frac{t-2}{2}} \int \left[\frac{\kappa^{a_0+N(\frac{t-3}{2})-1} e^{-b_0 \kappa} I_{(\frac{t-2}{2})}(\beta_1 \kappa) I_{(\frac{t-2}{2})}(\beta_2 \kappa)}{I_{(\frac{t-2}{2})}^2(\beta_0 \kappa) I_{(\frac{t-3}{2})}^N(\kappa)} \right] d\kappa \\
&= C \left(\frac{\beta_0}{\beta_1} \right)^{\frac{t-2}{2}} \left(\frac{\beta_0}{\beta_2} \right)^{\frac{t-2}{2}} \int \kappa^{a_2-1} e^{-b_2 \kappa} d\kappa
\end{aligned}$$

where for $i = 1, 2$,

$$\begin{aligned}
\mathbf{m}_i &= \left[\beta_0 \mathbf{m}_0 + \sum_{j=1}^{N_i} \mathbf{y}_{ij} \right] \beta_i^{-1} = \|\mathbf{m}_i\| \\
a_2 &= a_0 + N \left(\frac{t-3}{2} \right) + \sum_i \beta_i \bar{\kappa} \left[\frac{\partial}{\partial \beta_i \kappa} \ln I_{\frac{t-2}{2}}(\beta_i \bar{\kappa}) \right] \\
b_2 &= b_0 + N \frac{\partial}{\partial \kappa} \ln I_{\frac{t-3}{2}}(\bar{\kappa}) + 2\beta_0 \left[\frac{\partial}{\partial \beta_0 \kappa} \ln I_{\frac{t-2}{2}}(\beta_0 \bar{\kappa}) \right]
\end{aligned}$$

and the posterior mode $\bar{\kappa}$ is given recursively:

$$\bar{\kappa} = \begin{cases} \frac{a_2-1}{b_2} & \text{if } a_2 > 1, \\ \frac{a_2}{b_2} & \text{otherwise.} \end{cases}$$

It follows that the numerator of the Bayes factor is

$$C \left(\frac{\beta_0}{\beta_1} \right)^{\frac{t-2}{2}} \left(\frac{\beta_0}{\beta_2} \right)^{\frac{t-2}{2}} \frac{\Gamma(a_1)}{b_1^{a_1}}.$$

The Bayes factor is then given by the ratio

$$\begin{aligned}
 B_{21} &= \frac{\left(\frac{\beta_0}{\beta}\right)^{\frac{t-2}{2}} \frac{\Gamma(a_1)}{b_1^{a_1}}}{\left(\frac{\beta_0}{\beta_1}\right)^{\frac{t-2}{2}} \left(\frac{\beta_0}{\beta_2}\right)^{\frac{t-2}{2}} \frac{\Gamma(a_2)}{b_2^{a_2}}} \\
 &= \left(\frac{\beta_1\beta_2}{\beta\beta_0}\right)^{\frac{t-2}{2}} \frac{\Gamma(a_1) b_2^{a_2}}{\Gamma(a_2) b_1^{a_1}}.
 \end{aligned}$$

4 Conclusion

In this article, we have considered a few applications of the change of measure paradigm. In particular, it was possible to obtain a new derivation of the Friedman statistic. As well, extensions to the Bayesian models for ranking data were considered. Further applications as, for example, to the sign and Wilcoxon tests are found in [2].

Acknowledgements Work supported by the Natural Sciences and Engineering Council of Canada, Grant OGP0009068.

References

1. Alvo, M., Xu, H.: The analysis of ranking data using score functions and penalized likelihood. *Austrian J. Stat.* **46**, 15–32 (2017)
2. Alvo, M., Yu, Philip, L.H.: *A Parametric Introduction to Nonparametric Statistics*. Springer, Berlin (2018)
3. Alvo, M., Yu, Philip, L.H.: *Statistical Methods for Ranking Data*. Springer, Berlin (2014)
4. Banerjee, A., Dhillon, IS., Ghosh, J., Sra, S.: Clustering on the unit hypersphere using von mises-fisher distributions. *J. Mach. Learn. Res.* **6**, 1345–1382 (2005)
5. Friedman, M.: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Assoc.* **32**, 675–701 (1937)
6. McCullagh, P.: Models on spheres and models for permutations. In: Fligner, M.A., Verducci, J.S. (eds.) *Probability Models and Statistical Analyses for Ranking Data*, p. 278283. Springer, Berlin
7. Neyman, J.: Smooth test for goodness of fit. *Skandinavisk Aktuarietidskrift* **20**, 149–199 (1937)
8. Sra, S.: A short note on parameter approximation for von Mises-Fisher distributions: and a fast implementation of is (x). *Comput. Stat.* **27**(1), 177–190 (2012)
9. Xu, H., Alvo, M., Yu, Philip, L.H.: Angle-based models for ranking data. *Comput. Stat. Data Anal.* **121**, 113–136 (2018)