Economic Review (Otaru University of Commerce),
Vol. 66, No. 1, 1-10, July, 2015.

An expository supplement to the paper "Optimization of the Gaussian and Jeffreys power priors with emphasis on the canonical parameters in the exponential family"

Haruhiko Ogasawara

This article gives an expository supplement to Ogasawara (2014) for the estimators of the multinomial logits in the categorical distribution, which is the generalization of the Bernoulli distribution.

1. The Fisher information matrix

Define the likelihood of the vector of the canonical parameters $\boldsymbol{\theta}$ when n observations are given as:

$$L=\prod_{i=1}^{n}\prod_{j=1}^{K}p_j^{y_{ij}}, \qquad \text{(S.1)}$$

where $y_{ij}(i=1,...,n;j=1,...,K)$ are given observations. Let $\bar{l}=n^{-1}\log L$. Then,

$$\frac{\partial\bar{l}}{\partial\theta_j}=\frac{\partial\bar{l}}{\partial(\boldsymbol{\theta})_j}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}=n^{-1}\sum_{i=1}^{n}\sum_{a=1}^{K}\frac{y_{ia}}{p_a}\frac{\partial p_a}{\partial\theta_j},$$

$$\frac{\partial p_a}{\partial\theta_j}=-\frac{e^{\theta_a}e^{\theta_j}}{\left(1+\sum_{b=1}^{K-1}e^{\theta_b}\right)^2}=-p_jp_a \quad (a=1,...,K-1;a\neq j), \qquad \text{(S.2)}$$

$$\frac{\partial p_j}{\partial\theta_j}=p_j(1-p_j)\equiv p_jq_j,\ \frac{\partial p_K}{\partial\theta_j}=-p_jp_K \quad (j=1,...,K-1).$$

From (S.2),

$$\begin{aligned}\frac{\partial \bar{l}}{\partial \theta_j} &= n^{-1}\sum_{i=1}^{n}\left(-\sum_{a=1}^{K-1}\frac{y_{ia}}{p_a}p_a p_j + \frac{y_{ij}}{p_j}p_j - \frac{y_{iK}}{p_K}p_j p_K\right)\\ &= n^{-1}\sum_{i=1}^{n}y_{ij} - n^{-1}\sum_{i=1}^{n}\sum_{a=1}^{K}y_{ia}p_j\\ &\equiv \bar{y}_j - p_j,\\ \frac{\partial^2 \bar{l}}{\partial \theta_j^2} &= -\frac{\partial p_j}{\partial \theta_j} = -p_j q_j \quad (j=1,\ldots,K-1),\\ \frac{\partial^2 \bar{l}}{\partial \theta_j \partial \theta_{k^*}} &= p_j p_{k^*} \quad (j,k^*=1,\ldots,K-1;\ j\neq k^*).\end{aligned} \tag{S.3}$$

Then, the population information matrix per observation becomes

$$\mathbf{I}_0 = \begin{bmatrix} p_1q_1 & -p_1p_2 & \cdots & -p_1p_{K-1} \\ -p_2p_1 & p_2q_2 & \cdots & -p_2p_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{K-1}p_1 & -p_{K-1}p_2 & \cdots & p_{K-1}q_{K-1} \end{bmatrix} = \mathrm{cov}(\mathbf{Y}_{(K-1)}) \tag{S.4}$$

with $\mathbf{Y}_{(K-1)} = (Y_1,\ldots,Y_{K-1})'$. Note that (S.4) is also given from the property of the canonical parameters in the exponential family. The $(K-1)\times(K-1)$ matrix $\mathbf{I}_0$ is also denoted by $\mathbf{I}_{0(K-1)}$ for clarity.

Lemma 1.

$$|\mathbf{I}_{0(K-1)}| = p_1p_2\cdots p_K. \tag{S.5}$$

Proof. The result is derived by induction. Assume that $p_j > 0\ (j=1,\ldots,K)$. When $K=2$, $\mathbf{I}_{0(1)} = p_1q_1 = p_1p_2$, which shows that (S.5) holds. Suppose that when $K=J$, (S.5) holds. Write $\mathbf{I}_{0(J)} = \begin{bmatrix} \mathbf{I}_{0(J-1)} & -\mathbf{p}_{(J-1)}p_J \\ -\mathbf{p}_{(J-1)}'p_J & p_Jq_J \end{bmatrix}$, where $\mathbf{p}_{(J-1)} = (p_1,\ldots,p_{J-1})'$. Since $|\mathbf{I}_{0(J-1)}| = p_1p_2\cdots p_J > 0$ by assumption, $\mathbf{I}_{0(J-1)}$ has its inverse. Consequently, using the formula of the determinant of a partitioned matrix,

$\det\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}' & c \end{bmatrix} = |\mathbf{A}|(c - \mathbf{b}'\mathbf{A}^{-1}\mathbf{b})$ when $\mathbf{A}$ is nonsingular, it follows that

$$|\mathbf{I}_{0(J)}| = |\mathbf{I}_{0(J-1)}|(p_J q_J - p_J^2 \mathbf{p}_{(J-1)}'\mathbf{I}_{0(J-1)}^{-1}\mathbf{p}_{(J-1)}) \tag{S.6}$$

where

$$\begin{aligned}\mathbf{I}_{0(J-1)}^{-1} &= \{\mathrm{diag}(\mathbf{p}_{(J-1)}) - \mathbf{p}_{(J-1)}\mathbf{p}_{(J-1)}'\}^{-1} \\ &= \mathrm{diag}^{-1}(\mathbf{p}_{(J-1)}) + \frac{\mathrm{diag}^{-1}(\mathbf{p}_{(J-1)})\mathbf{p}_{(J-1)}\mathbf{p}_{(J-1)}'\mathrm{diag}^{-1}(\mathbf{p}_{(J-1)})}{1 - \mathbf{p}_{(J-1)}'\mathrm{diag}^{-1}(\mathbf{p}_{(J-1)})\mathbf{p}_{(J-1)}} \\ &= \mathrm{diag}^{-1}(\mathbf{p}_{(J-1)}) + \frac{\mathbf{1}_{(J-1)}\mathbf{1}_{(J-1)}'}{1 - \mathbf{p}_{(J-1)}'\mathbf{1}_{(J-1)}},\end{aligned} \tag{S.7}$$

which gives

$$\begin{aligned}|\mathbf{I}_{0(J)}| &= |\mathbf{I}_{0(J-1)}|\left[p_J q_J - p_J^2\left\{\mathbf{p}_{(J-1)}'\mathbf{1}_{(J-1)} + \frac{(\mathbf{p}_{(J-1)}'\mathbf{1}_{(J-1)})^2}{1 - \mathbf{p}_{(J-1)}'\mathbf{1}_{(J-1)}}\right\}\right] \\ &= |\mathbf{I}_{0(J-1)}|\left(p_J q_J - \frac{p_J^2 \mathbf{p}_{(J-1)}'\mathbf{1}_{(J-1)}}{1 - \mathbf{p}_{(J-1)}'\mathbf{1}_{(J-1)}}\right) \\ &= |\mathbf{I}_{0(J-1)}| p_J \frac{1 - \mathbf{p}_{(J-1)}'\mathbf{1}_{(J-1)} - p_J}{1 - \mathbf{p}_{(J-1)}'\mathbf{1}_{(J-1)}} \\ &= p_1 p_2 \cdots p_J (1 - p_1 - \ldots - p_J),\end{aligned} \tag{S.8}$$

which shows that (S.5) holds when $K = J+1$. When $p_1 p_2 \cdots p_J = 0$, from (S.4) with $K = J+1$, at least one of the rows/columns of $\mathbf{I}_{0(J)}$ is zero and consequently, $|\mathbf{I}_{0(J)}| = 0$, which shows that (S.5) holds also for the singular case. Q.E.D.

2. The Jeffreys prior

The log prior derivatives evaluated at the population values are

$$\mathbf{q}_0^* = \frac{1}{2}\frac{\partial \log|\mathbf{I}_{0(K-1)}|}{\partial \boldsymbol{\theta}_0} = \frac{1}{2}\frac{\partial}{\partial \boldsymbol{\theta}_0}\sum_{j=1}^{K}\log p_j = \frac{1}{2}\sum_{j=1}^{K}\frac{1}{p_j}\frac{\partial p_j}{\partial \boldsymbol{\theta}_0}, \tag{S.9}$$

where

$$\frac{\partial p_j}{\partial \boldsymbol{\theta}_0} = (-p_1 p_j, -p_2 p_j, \ldots, p_j q_j, \ldots, -p_{K-1} p_j)'$$
$$= -p_j \mathbf{p}_{(K-1)} + (\mathbf{0}', p_j, \mathbf{0}')' \quad (j = 1, \ldots, K-1),$$
$$\frac{\partial p_K}{\partial \boldsymbol{\theta}_0} = (-p_1 p_K, -p_2 p_K, \ldots, -p_j p_K, \ldots, -p_{K-1} p_K)' \qquad \text{(S.10)}$$
$$= -p_K \mathbf{p}_{(K-1)}.$$

Consequently, (S.9) becomes

$$\mathbf{q}_0^* = \frac{1}{2}(\mathbf{1}_{(K-1)} - K\mathbf{p}_{(K-1)}). \qquad \text{(S.11)}$$

When $K = 2$, (S.11) becomes $(1-2p_1)/2$, which is also given by $\frac{\partial \log(p_1 q_1)^{1/2}}{\partial \theta_1} = \frac{1}{2p_1 q_1}\frac{\partial p_1 q_1}{\partial \theta_1} = \frac{1-2p_1}{2}$. Note that $\mathbf{q}_0^* = \mathbf{0}$ when $\mathbf{p}_{(K-1)} = 1/K$ and $p_K = 1/K$ or when the proportions $p_j (j = 1, \ldots, K)$ are equal. When $K = 2$, this holds with $p_1 = 0.5$.

Lemma 2.

$$\boldsymbol{\alpha}_{\text{ML1}} = -0.5(p_1^{-1}, \ldots, p_{K-1}^{-1})' + 0.5 p_K^{-1} \mathbf{1}_{(K-1)}. \qquad \text{(S.12)}$$

Proof. Since

$$\hat{\boldsymbol{\theta}}_{\text{ML}} = \{\log(\hat{p}_1 / \hat{p}_K), \ldots, \log(\hat{p}_{K-1} / \hat{p}_K)\}' \qquad \text{(S.13)}$$

with $\hat{p}_j = n^{-1}\sum_{i=1}^{n} y_{ij} \ (j = 1, \ldots, K)$,

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\text{ML}} &= \boldsymbol{\theta}_0 + \frac{\partial \boldsymbol{\theta}_0}{\partial \mathbf{p}_{(K-1)}'}(\hat{\mathbf{p}}_{(K-1)} - \mathbf{p}_{(K-1)}) \\ &+ \frac{1}{2}\frac{\partial^2 \boldsymbol{\theta}_0}{(\partial \mathbf{p}_{(K-1)}')^{<2>}}(\hat{\mathbf{p}}_{(K-1)} - \mathbf{p}_{(K-1)})^{<2>} \\ &+ \frac{1}{6}\frac{\partial^3 \boldsymbol{\theta}_0}{(\partial \mathbf{p}_{(K-1)}')^{<3>}}(\hat{\mathbf{p}}_{(K-1)} - \mathbf{p}_{(K-1)})^{<3>} + O_p(n^{-2}),\end{aligned} \qquad \text{(S.14)}$$

where $\hat{\mathbf{p}}_{(K-1)} = (\hat{p}_1,\dots,\hat{p}_{K-1})'$,

$$\left(\frac{\partial \boldsymbol{\theta}_0}{\partial \mathbf{p}_{(K-1)}{}'}\right)_{jk^*} = \frac{\delta_{jk^*}}{p_j} + \frac{1}{1-\mathbf{p}_{(K-1)}{}'\mathbf{1}_{(K-1)}} \quad (j,k^* = 1,\dots,K-1)$$

which gives

$$\frac{\partial \boldsymbol{\theta}_0}{\partial \mathbf{p}_{(K-1)}{}'} = \mathrm{diag}(p_1^{-1},\dots,p_{K-1}^{-1}) + \frac{\mathbf{1}_{(K-1)}\mathbf{1}_{(K-1)}{}'}{1-\mathbf{p}_{(K-1)}{}'\mathbf{1}_{(K-1)}} = \mathbf{I}_{0(K-1)}^{-1}, \quad \text{(S.15)}$$

which also comes from Subsection A.1 of the appendix.

$$\begin{aligned}\frac{\partial^2 \boldsymbol{\theta}_0}{\partial p_j \partial p_{k^*}} &= (\mathbf{0}', -\delta_{jk^*} p_j^{-2}, \mathbf{0}')' + \frac{\mathbf{1}_{(K-1)}}{(1-\mathbf{p}_{(K-1)}{}'\mathbf{1}_{(K-1)})^2} \\ &= -\delta_{jk^*} p_j^{-2}\mathbf{e}_{(j)} + p_K^{-2}\mathbf{1}_{(K-1)},\end{aligned} \quad \text{(S.16)}$$

$$\frac{\partial^3 \boldsymbol{\theta}_0}{\partial p_j \partial p_{k^*} \partial p_{l^*}} = \delta_{jk^*}\delta_{k^*l^*} 2p_j^{-3}\mathbf{e}_{(j)} + 2p_K^{-3}\mathbf{1}_{(K-1)} \quad (j,k^*,l^* = 1,\dots,K-1).$$

From the above results,

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\mathrm{ML}} &= \boldsymbol{\theta}_0 + \{p_1^{-1}(\hat{p}_1 - p_1),\dots,p_{K-1}^{-1}(\hat{p}_{K-1} - p_{K-1})\}' \\ &\qquad + p_K^{-1}\mathbf{1}_{(K-1)}{}'(\hat{\mathbf{p}}_{(K-1)} - \mathbf{p}_{(K-1)})\mathbf{1}_{(K-1)} \\ &+ \frac{1}{2}[-\{p_1^{-2}(\hat{p}_1 - p_1)^2,\dots,p_{K-1}^{-2}(\hat{p}_{K-1} - p_{K-1})^2\}' \\ &\qquad + p_K^{-2}(\mathbf{1}_{(K-1)}{}')^{<2>}(\hat{\mathbf{p}}_{(K-1)} - \mathbf{p}_{(K-1)})^{<2>}\mathbf{1}_{(K-1)}] \\ &+ \frac{1}{3}[\{p_1^{-3}(\hat{p}_1 - p_1)^3,\dots,p_{K-1}^{-3}(\hat{p}_{K-1} - p_{K-1})^3\}' \\ &\qquad + p_K^{-3}(\mathbf{1}_{(K-1)}{}')^{<3>}(\hat{\mathbf{p}}_{(K-1)} - \mathbf{p}_{(K-1)})^{<3>}\mathbf{1}_{(K-1)}] + O_p(n^{-3}).\end{aligned} \quad \text{(S.17)}$$

From (S.17),

$$(\boldsymbol{\alpha}_{\mathrm{ML1}})_j = \frac{1}{2}\left\{-p_j^{-2}(\mathbf{I}_{0(K-1)})_{jj} + p_K^{-2}\sum_{a,b=1}^{K-1}(\mathbf{I}_{0(K-1)})_{ab}\right\}$$
$$= \frac{1}{2}\left(-\frac{1-p_j}{p_j} + \frac{1-p_K}{p_K}\right) \tag{S.18}$$
$$= \frac{1}{2}\left(-\frac{1}{p_j} + \frac{1}{p_K}\right) = \frac{p_j - p_K}{2p_j p_K} \quad (j = 1, \ldots, K-1),$$

where

$$\sum_{a,b=1}^{K-1}(\mathbf{I}_{0(K-1)})_{ab} = \mathbf{1}_{(K-1)}{}'\mathbf{I}_{0(K-1)}\mathbf{1}_{(K-1)}$$
$$= \mathbf{1}_{(K-1)}{}'(\mathbf{p}_{(K-1)} - \mathbf{p}_{(K-1)}\mathbf{p}_{(K-1)}{}'\mathbf{1}_{(K-1)}) \tag{S.19}$$
$$= \mathbf{1}_{(K-1)}{}'p_K\mathbf{p}_{(K-1)} = p_K(1-p_K)$$

is used. (S.18) gives (S.12). Q.E.D.

Note that the term $(1/3)[\cdot]$ in (S.17) is unnecessary to have the result, but is included for illustration.

From Lemma 2, we have

$$n\ \mathrm{acov}\{(\hat{\boldsymbol{\alpha}}_{\mathrm{ML1}})_j, \hat{\boldsymbol{\theta}}_{\mathrm{ML}}{}'\} = \frac{1}{2}\left\{\frac{\partial}{\partial \mathbf{p}_{(K-1)}{}'}\left(-\frac{1}{p_j} + \frac{1}{p_K}\right)\right\}\mathbf{I}_{0(K-1)}\frac{\partial \boldsymbol{\theta}_0{}'}{\partial \mathbf{p}_{(K-1)}}$$
$$= \frac{1}{2}\{(\mathbf{0}', p_j^{-2}, \mathbf{0}') + p_K^{-2}\mathbf{1}_{(K-1)}{}'\}\mathbf{I}_{0(K-1)}\mathbf{I}_{0(K-1)}^{-1} \tag{S.20}$$
$$= \frac{1}{2}(p_j^{-2}\,\mathbf{e}_{(j)} + p_K^{-2}\mathbf{1}_{(K-1)})' > \mathbf{0}',$$

which gives

$$n\ \mathrm{acov}(\hat{\boldsymbol{\alpha}}_{\mathrm{ML1}}, \hat{\boldsymbol{\theta}}_{\mathrm{ML}}{}') = \frac{1}{2}\{\mathrm{diag}(p_1^{-2}, \ldots, p_{K-1}^{-2}) + p_K^{-2}\mathbf{1}_{(K-1)}\mathbf{1}_{(K-1)}{}'\}. \tag{S.21}$$

2.1 Linear predictor with the Jeffreys prior

Using Lemma 2 and (S.20), Result 5 gives

$$k_{\min} = (\mathbf{p}^{*}{}'\boldsymbol{\alpha}_{\mathrm{ML1}})^{-2}\mathbf{p}^{*}{}'n\ \mathrm{acov}(\hat{\boldsymbol{\alpha}}_{\mathrm{ML1}}, \hat{\boldsymbol{\theta}}_{\mathrm{ML}}{}')\mathbf{p}^{*} + 1$$
$$= 2\{-\mathbf{p}^{*}{}'(p_1^{-1},\ldots,p_{K-1}^{-1})' + p_K^{-1}\mathbf{p}^{*}{}'\mathbf{1}_{(K-1)}\}^{-2}$$
$$\times\{\mathbf{p}^{*}{}'\mathrm{diag}(p_1^{-2},\ldots,p_{K-1}^{-2})\mathbf{p}^{*} + p_K^{-2}(\mathbf{p}^{*}{}'\mathbf{1}_{(K-1)})^2\} + 1 \qquad \text{(S.22)}$$
$$> 1.$$

2.2 TMSE with the Jeffreys prior

Similarly, Result 6 yields

$$k_{\min} = (\boldsymbol{\alpha}_{\mathrm{ML1}}{}'\boldsymbol{\alpha}_{\mathrm{ML1}})^{-1}\mathrm{tr}\{n\ \mathrm{acov}(\hat{\boldsymbol{\alpha}}_{\mathrm{ML1}}, \hat{\boldsymbol{\theta}}_{\mathrm{ML}}{}')\} + 1$$
$$= 2\left\{\sum_{j=1}^{K-1}\left(-\frac{1}{p_j} + \frac{1}{p_K}\right)^2\right\}^{-1}\left(\sum_{j=1}^{K-1}\frac{1}{p_j^2} + \frac{K-1}{p_K^2}\right) + 1 \qquad \text{(S.23)}$$

> 3 (for the inequality see Result 9).

3. The Gaussian prior

The density of the Gaussian prior is defined as

$$f(\boldsymbol{\theta}) \propto \exp(-\boldsymbol{\theta}'\boldsymbol{\theta}/2) \qquad \text{(S.24)}$$

giving $\mathbf{q}_0^{*} = -\boldsymbol{\theta}_0$.

Under correct model specification, from (2.3) and (2.4) $\mathbf{A}_{\mathrm{G|ML\Delta2}}$ or $\mathbf{A}_{\mathrm{W|ML\Delta2}}$ by the Gaussian prior is

$$\mathbf{A}_{\mathrm{G|ML\Delta2}} = n\mathrm{E}_{\theta}(\mathbf{L}^{(\mathrm{W})}\mathbf{l}_0^{(1)}{}'\boldsymbol{\Lambda}^{(1)} + \boldsymbol{\Lambda}^{(1)}\mathbf{l}_0^{(1)}\mathbf{L}^{(\mathrm{W})}{}'), \qquad \text{(S.25)}$$

where

$$\mathbf{L}^{(\mathrm{W})} = -(-\hat{\mathbf{I}}_{(K-1)}^{-1}\hat{\mathbf{q}}^{*} + \mathbf{I}_{0(K-1)}^{-1}\mathbf{q}_0^{*}) = \hat{\mathbf{I}}_{(K-1)}^{-1}\hat{\mathbf{q}}^{*} - \mathbf{I}_{0(K-1)}^{-1}\mathbf{q}_0^{*}, \qquad \text{(S.26)}$$

$\hat{\mathbf{I}}_{(K-1)}$ and $\hat{\mathbf{q}}^{*}$ are $\mathbf{I}_{0(K-1)}$ and $\mathbf{q}_0^{*}$, respectively with $\boldsymbol{\theta}_0$ replaced by $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$. (S.25) become

$$\mathbf{A}_{\mathrm{G|ML\Delta2}} = \frac{\partial\mathbf{I}_{0(K-1)}^{-1}\mathbf{q}_0^{*}}{\partial\boldsymbol{\theta}_0{}'}\mathbf{A}_{\mathrm{ML2}} + \mathbf{A}_{\mathrm{ML2}}\frac{\partial\mathbf{q}_0^{*}{}'\mathbf{I}_{0(K-1)}^{-1}}{\partial\boldsymbol{\theta}_0} \qquad \text{(S.27)}$$

where $\mathbf{A}_{\mathrm{ML2}} = \mathbf{I}_{0(K-1)}^{-1}$ and

$$\frac{\partial \mathbf{I}_{0(K-1)}^{-1}\mathbf{q}_0^*}{\partial(\boldsymbol{\theta}_0)_j} = -\frac{\partial \mathbf{I}_{0(K-1)}^{-1}\boldsymbol{\theta}_0}{\partial(\boldsymbol{\theta}_0)_j}$$

$$= -\left[\frac{\partial}{\partial(\boldsymbol{\theta}_0)_j}\{\mathrm{diag}(p_1^{-1},\ldots,p_{K-1}^{-1}) + p_K^{-1}\mathbf{1}_{(K-1)}\mathbf{1}_{(K-1)}'\}\right]\boldsymbol{\theta}_0 - (\mathbf{I}_{0(K-1)}^{-1})_{\cdot j}$$

$$= \left\{\mathrm{diag}\left(p_1^{-2}\frac{\partial p_1}{\partial(\boldsymbol{\theta}_0)_j},\ldots,p_{K-1}^{-2}\frac{\partial p_{K-1}}{\partial(\boldsymbol{\theta}_0)_j}\right) + p_K^{-2}\frac{\partial p_K}{\partial(\boldsymbol{\theta}_0)_j}\mathbf{1}_{(K-1)}\mathbf{1}_{(K-1)}'\right\}\boldsymbol{\theta}_0$$

$$- (\mathbf{I}_{0(K-1)}^{-1})_{\cdot j}$$

$$= [\mathrm{diag}\{-p_1^{-2}p_1p_j,\ldots,p_j^{-2}p_j(1-p_j),\ldots,-p_{K-1}^{-2}p_{K-1}p_j\}$$

$$- p_K^{-2}p_Kp_j\mathbf{1}_{(K-1)}\mathbf{1}_{(K-1)}']\boldsymbol{\theta}_0 - (\mathbf{I}_{0(K-1)}^{-1})_{\cdot j}$$

$$= \{-p_j\mathrm{diag}(p_1^{-1},\ldots,p_{K-1}^{-1}) + \mathrm{diag}(\mathbf{0}', p_j^{-1}, \mathbf{0}')$$

$$- p_K^{-1}p_j\mathbf{1}_{(K-1)}\mathbf{1}_{(K-1)}'\}\boldsymbol{\theta}_0 - (\mathbf{I}_{0(K-1)}^{-1})_{\cdot j} \quad \text{(S.28)}$$

$(j = 1,\ldots,K-1)$.

3.1 Linear predictor with the Gaussian prior

Result 3 with the above results gives

$$k_{\min} = -\frac{\mathbf{p}^{*\prime}\mathbf{A}_{\mathrm{G|ML\Delta2}}\mathbf{p}^*}{2(\mathbf{p}^{*\prime}\mathbf{I}_{0(K-1)}^{-1}\boldsymbol{\theta}_0)^2} + \frac{\mathbf{p}^{*\prime}\boldsymbol{\alpha}_{\mathrm{ML1}}}{\mathbf{p}^{*\prime}\mathbf{I}_{0(K-1)}^{-1}\boldsymbol{\theta}_0}. \quad \text{(S.29)}$$

3.2 TMSE with the Gaussian prior

Similarly, Result 4 gives

$$k_{\min} = \frac{1}{2}(\boldsymbol{\theta}_0'\mathbf{I}_{0(K-1)}^{-2}\boldsymbol{\theta}_0)^{-1}\{-\mathrm{tr}(\mathbf{A}_{\mathrm{G|ML\Delta2}}) + 2\boldsymbol{\theta}_0'\mathbf{I}_{0(K-1)}^{-1}\boldsymbol{\alpha}_{\mathrm{ML1}}\}. \quad \text{(S.30)}$$

4. The largest variance of the non-negative quantities whose sum is fixed

Let $p_j\,(j = 1,\ldots,K)$ with fixed K be non-negative quantities, which vary with their sum being fixed. Suppose that the sum is 1 without loss of generality. Then, the variance of $p_1,\ldots,p_K$ denoted by var(p) is given by

$$K\,\mathrm{var}(p) = \sum_{j=1}^{K} p_j^2 - K^{-1}\left(\sum_{j=1}^{K} p_j\right)^2 = \sum_{j=1}^{K} p_j^2 - K^{-1}. \qquad \text{(S.31)}$$

So, the problem of maximizing the var(p) reduces to that of $\sum_{j=1}^{K} p_j^2$. In the case of $K = 2$, $\sum_{j=1}^{2} p_j^2 = p_1^2 + p_2^2 = p_1^2 + (1 - p_1)^2$, whose largest value is given when $p_1 = 0$ or 1. The smallest variance 0 is given when $p_1 = p_2 = 0.5$. Note that $p_1^2 + p_2^2$ is the square of the radius of the circle whose center is the origin in the (p_1, p_2) plane. Since $p_1 \geq 0$, $p_2 \geq 0$ and $p_1 + p_2 = 1$, possible values of $p_1^2 + p_2^2$ are those on the line segment connecting (0, 1) and (1, 0) in Figure 1. The value $p_1^2 + p_2^2$ is given by the square of the radius of the circle which has point(s) on the line segment in the first quadrant including (0, 1) and (1, 0). From Figure 1, it is obvious that the largest value is given when (p_1, p_2) = (0, 1) or (1, 0) and that the smallest value is given when $p_1 = p_1 = 0.5$.

Generalizing the above result to the K-dimensional space with $K \geq 3$, the line segment becomes a portion of the $(K-1)$-dimensional hyperplane satisfying $p_1 + \ldots + p_K = 1$ and $p_j \geq 0$ $(j = 1, \ldots, K)$. The problem is to have the largest radius of the ball whose center is the origin and whose surface has point(s) on the plane. Since the plane is restricted to be in the first (generalized) quadrant $(1 \geq p_j \geq 0; j = 1, \ldots, K)$, the ball with the largest radius has K points $(p_1, p_2, \ldots, p_K)$ = (1,0,...,0), (0,1,0,...,0),...,(0,...,0,1) on the plane. This gives the largest variance

$$K^{-1}\sum_{j=1}^{K} p_j^2 - K^{-2} = K^{-1} - K^{-2}. \qquad \text{(S.32)}$$

The smallest variance 0 is given when $p_1 = \ldots = p_K = K^{-1}$.

When the sum is c rather than 1, the largest variance becomes $c^2(K^{-1} - K^{-2})$.

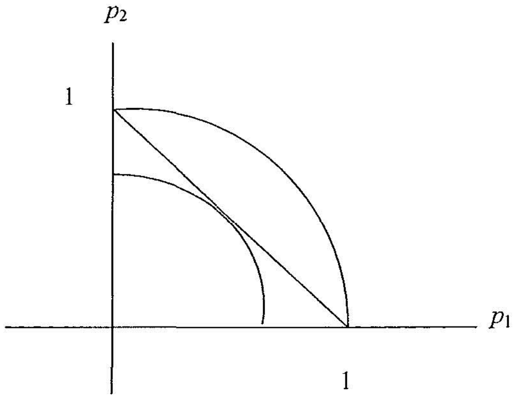

Figure 1. A geometric interpretation of the problem maximizing the variance of non-negative quantities when the mean is fixed

Reference

Ogasawara, H. (2014). Optimization of the Gaussian and Jeffreys power priors with emphasis on the canonical parameters in the exponential family. *Behaviormetrika*, *41*, 195-223.