

# Asymptotic Distribution of Cramér-von Mises Statistic When Contamination Exists

Mayumi Naka<sup>1</sup> & Ritei Shibata<sup>2</sup>

<sup>1</sup> School of Fundamental Science and Technology, Keio University, Yokohama, Japan

<sup>2</sup> Department of Mathematics, Keio University, Yokohama, Japan

Correspondence: Mayumi Naka, School of Fundamental Science and Technology, Keio University, Yokohama 223-8522, Japan. Tel: 81-45-566-1641. E-mail: naka@datascience.jp

Received: November 11, 2015 Accepted: December 1, 2015 Online Published: December 23, 2015

doi:10.5539/ijsp.v5n1p90

URL: <http://dx.doi.org/10.5539/ijsp.v5n1p90>

## Abstract

In this paper, asymptotic distribution of Cramér-von Mises goodness-of-fit test statistic is investigated when contamination exists. We first derive the asymptotic distribution of the Cramér-von Mises statistic when the observations are contaminated with noise as a mixture. The result is extended to the case where the parameters are estimated by the minimum distance estimator, which minimizes the Cramér-von Mises statistic. In both cases the asymptotic distribution of the Cramér-von Mises statistic is given by that of the weighted infinite sum of non-central  $\chi_1^2$  variables and the effect of contamination appears only in the non-centrality of the variables. We also demonstrate the robustness of the goodness-of-fit test by Monte Carlo simulations when the parameters are estimated by the minimum distance estimator and the maximum likelihood estimator. Numerical experiments indicate that the use of the minimum distance estimator makes the test insensitive to contamination whereas the power is retained almost the same as that of the maximum likelihood estimator.

**Keywords:** contamination, Cramér-von Mises statistic, goodness-of-fit test, minimum distance estimator, robustness.

## 1. Introduction

Let  $X_1, X_2, \dots, X_n$  be independent and identically distributed observations, and  $X_{(j)}, j = 1, 2, \dots, n$  be their order statistics. In this paper, we assume that the distribution is contaminated as

$$F_\varepsilon(x, \theta) = \left(1 - \frac{\varepsilon}{\sqrt{n}}\right) F(x, \theta) + \frac{\varepsilon}{\sqrt{n}} G(x),$$

where  $F(x, \theta)$  is a continuous distribution with the parameter  $\theta = (\theta_1, \theta_2, \dots, \theta_m)^\top \in \Theta \subset \mathbb{R}^m$ , and  $G(x)$  is the distribution of the contamination, where the rate of contamination is  $\varepsilon \geq 0$ . We hereafter assume that both distributions  $F(x, \theta)$  and  $G(x)$  have bounded and smooth densities  $f(x, \theta)$  and  $g(x)$ , respectively. We are interested in the behavior of the goodness-of-fit test statistic,

$$W_n^2(\theta) = \sum_{j=1}^n \left( F(X_{(j)}, \theta) - \frac{j}{n+1} \right)^2.$$

The asymptotic behavior of  $W_n^2(\theta)$  when no contamination exists has been thoroughly investigated, including the case when an estimation of the parameter  $\theta$  is involved (for example, Shorack & Wellner, 1986). However, only a few works have been performed for cases where contamination exists. We first derive the asymptotic distribution of  $W_n^2(\theta)$  in a general frame work via an elementary matrix calculation. We next employ the minimum distance estimator  $\hat{\theta}$  which minimizes  $W_n^2(\theta)$  as an estimate of  $\theta$ . There are two reasons why we are interested in the use of the minimum distance estimator. The first is due to the fact that the estimator shares the same loss function with  $W_n^2(\theta)$ . It therefore seems natural to employ the same loss function for both parameter estimation and goodness-of-fit tests. The other reason is that  $\hat{\theta}$  is robust as an estimator, as shown in Millar (1981), for example. Woodward, Parr, Schucany & Lindsey (1984) demonstrated via numerical experiments that the minimum distance estimator is better than the maximum likelihood estimator under symmetric departures from normality for each component in normal mixture models. Since then, the minimum distance estimator is often used in practice for mixture models (Beutner & Bordes, 2011; Garcia-Dorado & Marin, 1998).

The rest of this paper is organized as follows. In Section 2 the asymptotic distribution of  $W_n^2(\theta)$  is derived. The effect of contamination appears only in the non-centrality of the weighted infinite sum of non-central  $\chi_1^2$  variables. In Section

3 the result is extended to the case where the parameters are estimated by the minimum distance estimator. The result is useful not only for knowing the effect of contamination to the asymptotic distribution, but also for obtaining the weights and non-centralities of  $\chi^2_1$  variables of finite sample sizes. The method described here is much simpler than solving the integral equation, as in the case of no contamination. The robustness of the test when the minimum distance estimator is used is demonstrated by computer simulations in Section 4. All proofs are given in the Appendix.

**2. Asymptotic Distribution When the Parameters Are Known**

We rewrite  $W_n^2(\theta)$  as  $W_n^2(\theta) = \|(n + 1)S_n \mathbf{U}_n\|^2$  by introducing an  $n \times (n + 1)$  matrix

$$S_n = \left( \frac{1}{n + 1} \left( 1_{j \geq k} - \frac{j}{n + 1} \right), 1 \leq j \leq n, 1 \leq k \leq n + 1 \right)$$

and an  $n + 1$  dimensional vector

$$\mathbf{U}_n = \left( F(X_{(j)}, \theta) - F(X_{(j-1)}, \theta); 1 \leq j \leq n + 1 \right)^\top.$$

Here we define  $F(X_{(0)}, \theta) = 0$  and  $F(X_{(n+1)}, \theta) = 1$  for convenience. We also define a diagonal matrix  $B$  with diagonal elements,  $b_1 = b_2$ , and

$$b_{j+1} = \frac{f(F_\varepsilon^{-1}(\frac{j}{n+1}, \theta), \theta)}{f_\varepsilon(F_\varepsilon^{-1}(\frac{j}{n+1}, \theta), \theta)}, \quad j = 1, 2, \dots, n,$$

where  $f_\varepsilon(x, \theta)$  and  $F_\varepsilon^{-1}(u, \theta) = x$  are the density function and the inverse function of  $F_\varepsilon(x, \theta)$ , respectively.

Taylor expansion yields an approximation of  $\mathbf{U}_n$  as

$$B \left( \mathbf{U}_n^* - \frac{1}{n + 1} (\mathbf{1} - \mathbf{c}_n) \right),$$

where  $\mathbf{U}_n^* = \left( F_\varepsilon(X_{(j)}, \theta) - F_\varepsilon(X_{(j-1)}, \theta); 1 \leq j \leq n + 1 \right)^\top$  with the convention  $F_\varepsilon(X_{(0)}, \theta) = 0$  and  $F_\varepsilon(X_{(n+1)}, \theta) = 1$  and the first element of  $\mathbf{c}_n$  is  $(n + 1)F(F_\varepsilon^{-1}(\frac{1}{n+1}, \theta), \theta)/b_1$  and all others are equal to 1.

We now see that it is enough to know the distribution of  $\sum_{j=1}^n \lambda_{nj}(V_{nj} + \mu_{nj})^2$  in place of  $W_n^2(\theta)$ , where

$$\mathbf{V}_n = (V_{n1}, V_{n2}, \dots, V_{nn})^\top = (n + 1)\Lambda_n^{-\frac{1}{2}} P_n^\top S_n B \left( \mathbf{U}_n^* - \frac{1}{n + 1} \mathbf{1} \right)$$

and

$$\boldsymbol{\mu}_n = (\mu_{n1}, \mu_{n2}, \dots, \mu_{nn})^\top = \Lambda_n^{-\frac{1}{2}} P_n^\top S_n B \mathbf{c}_n.$$

Here  $\Lambda_n$  is a diagonal matrix of eigenvalues  $\lambda_{n1} \geq \lambda_{n2} \geq \dots \geq \lambda_{nn}$ , and  $P_n$  is an orthogonal matrix of eigenvectors  $\mathbf{p}_j^{(n)}$ ,  $j = 1, 2, \dots$ , of  $S_n B^2 S_n^\top$ .

**Proposition 1.** For any fixed  $j > 0$ , as  $n$  tends to infinity  $\lambda_{nj}$  converges to  $\lambda_j = 1/(j\pi)^2$  and  $\sqrt{n}p_{[nu]j}^{(n)}$  converges to  $f_j(u) = \sqrt{2} \sin(\pi ju)$  for  $0 < u < 1$ , which are the eigenvalues and the eigenfunctions of the integral equation

$$\lambda f(u) = \int_0^1 \rho_0(u, v) f(v) dv,$$

where  $\rho_0(u, v) = \min(u, v) - uv$ ,  $p_{kj}^{(n)}$  is the  $k$ th element of  $\mathbf{p}_j^{(n)}$  and  $[x]$  is the minimum integer, which is greater than, or equal to,  $x$ .

Note that the eigenvalues and the eigenvectors become independent of the contamination in this limit.

**Proposition 2.** Any finite-dimensional random vector  $(V_{nj_1}, V_{nj_2}, \dots, V_{nj_p})^\top$  converges in distribution to a normally distributed random vector  $(V_{j_1}, V_{j_2}, \dots, V_{j_p})^\top$  with mean  $\mathbf{0}$  and variance  $I_p$  as  $n$  tends to infinity.

Combining these results, we see that  $\sum_{j=1}^p \lambda_{nj}(V_{nj} + \mu_{nj})^2$  converges to  $\sum_{j=1}^p \lambda_j(V_j + \mu_j)^2$ , where  $V_1, V_2, \dots, V_p$  are given in the proof of Proposition 2. On the other hand,  $\sum_{j=1}^n \lambda_{nj}\mu_{nj}^2$  converges to  $\sum_{j=1}^\infty \lambda_j\mu_j^2$  because of the boundedness of  $\sum_{j=1}^n \lambda_{nj}\mu_{nj}^2$ . We therefore have the following Theorem 1.

**Theorem 1.** *The distribution of  $W_n^2(\theta)$  converges in distribution to that of  $\sum_{j=1}^{\infty} \lambda_j(V_j + \mu_j)^2$ , where  $V_1, V_2, \dots$ , are independent and identically distributed random variables with a standard normal distribution and*

$$\mu_j = \varepsilon \lambda_j^{-\frac{1}{2}} \int_0^1 \int_0^1 f_j(u)(1_{u \geq v} - u) \left( 1 - \frac{g(F^{-1}(v, \theta))}{f(F^{-1}(v, \theta), \theta)} \right) dudv, \quad j = 1, 2, \dots$$

We see that contamination is only effective for  $\mu_j$ 's with the proportion  $\varepsilon$ . Theorem 1 reduces to the well-known result when no contamination exists, which is derived as an application of the theory of empirical processes. A closely related result to Theorem 1 is given by Guttorp & Lockhart (1988). They developed a general theory for an asymptotic distribution of quadratic forms of order statistics from a uniform distribution under contiguous alternatives, where densities are of the form  $1 + \delta\eta(u)/n^{\frac{1}{2}}$  under the condition  $\int_0^1 \eta(u)^2 du = 1$ . For  $\eta(F(x, \theta)) = g(x)/f(x, \theta) - 1$ , the same result as in Theorem 1 can be derived from their theory. However a major difference is that Theorem 1 is free from the constraint  $\int_0^1 \eta(u)^2 du < \infty$ . We have used only the fact that  $f(x, \theta)$  and  $g(x)$  are probability density functions for the proof.

### 3. Asymptotic Distribution When Parameters Are Estimated via the Minimum Distance Estimator

#### 3.1 Asymptotic Distribution

We need the following assumptions in order to derive the asymptotic distribution of  $W_n^2(\hat{\theta})$  when the parameters are estimated by the minimum distance estimator  $\hat{\theta}$ .

**Assumption 1** (Identifiability).

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} (F(x, \theta_n) - F(x, \theta))^2 dF(x, \theta) = 0$$

implies the convergence of  $\theta_n$  to  $\theta$  in  $\Theta$ .

**Assumption 2** (Regularity).

- $F(x, \theta)$  is differentiable with respect to  $\theta$ .
- $g_k(u, \theta) = \frac{\partial}{\partial \theta_k} F(x, \theta)$ ,  $k = 1, 2, \dots, m$ , are all continuous and square-integrable with respect to  $u \in (0, 1)$ , where  $u = F(x, \theta)$ .
- The matrix  $A = \left( \int_0^1 g_j(u, \theta) g_k(u, \theta) du, 1 \leq j, k \leq m \right)$  is of full rank.
- $\frac{\partial}{\partial \theta_k} g_j(u, \theta)$ ,  $j, k = 1, 2, \dots, m$  exist and are continuous for  $u$  in  $(0, 1)$ .
- $\sup_{0 < u < 1} \left| \frac{\partial}{\partial \theta_l} g_k(u, \theta) \right| < \infty$  for any  $k, l = 1, 2, \dots, m$ .
- $\sup_{0 < u < 1} \left| \frac{\partial}{\partial u} g_k(u, \theta) \right| < \infty$  for any  $k = 1, 2, \dots, m$ .

Define  $\bar{\lambda}_{n1}(\theta) \geq \bar{\lambda}_{n2}(\theta) \geq \dots \geq \bar{\lambda}_{n-n-m}(\theta)$  and  $\mathbf{q}_1^{(n)}, \mathbf{q}_2^{(n)}, \dots, \mathbf{q}_{n-m}^{(n)}$  are the eigenvalues and the eigenvectors of the matrix

$$D_n(\theta) = \Lambda_n^{\frac{1}{2}} \left[ I - P_n^T Z_n(\theta) \{ Z_n(\theta)^T Z_n(\theta) \}^{-1} Z_n(\theta)^T P_n \right] \Lambda_n^{\frac{1}{2}},$$

where

$$Z_n(\theta) = \left( \frac{1}{\sqrt{n}} g_k \left( \frac{j}{n+1}, \theta \right), 1 \leq j \leq n, 1 \leq k \leq m \right).$$

Without loss of generality, we may assume that  $Z_n(\theta)$  is of full rank for each  $n$ , in view of the third assumption in Assumption 2.

Since the minimum distance estimator  $\hat{\theta}$  is the solution of

$$\sum_{j=1}^n \left( F(X_{(j)}, \theta) - \frac{j}{n+1} \right) \mathbf{g}(X_{(j)}, \theta) \Bigg|_{\theta=\hat{\theta}} = 0,$$

for  $\mathbf{g}(u, \boldsymbol{\theta}) = (g_1(u, \boldsymbol{\theta}), g_2(u, \boldsymbol{\theta}), \dots, g_m(u, \boldsymbol{\theta}))^\top$ , the estimation error  $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$  is approximated as

$$-\sqrt{n} \left( \sum_{j=1}^n \mathbf{g}(X_{(j)}, \boldsymbol{\theta}) \mathbf{g}^\top(X_{(j)}, \boldsymbol{\theta}) \right)^{-1} \left( \sum_{j=1}^n \left( F(X_{(j)}, \boldsymbol{\theta}) - \frac{j}{n+1} \right) \mathbf{g}(X_{(j)}, \boldsymbol{\theta}) \right).$$

It is further approximated as

$$-(n+1) \left( Z_n(\boldsymbol{\theta})^\top Z_n(\boldsymbol{\theta}) \right)^{-1} Z_n(\boldsymbol{\theta})^\top S_n U_n.$$

If we note that  $W_n^2(\hat{\boldsymbol{\theta}})$  is asymptotically equivalent to

$$W_n^2(\boldsymbol{\theta}) - n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \left( Z_n(\boldsymbol{\theta})^\top Z_n(\boldsymbol{\theta}) \right) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}),$$

we see that it is enough to derive the asymptotic distribution of

$$(\mathbf{V}_n + \boldsymbol{\mu}_n)^\top D_n(\boldsymbol{\theta}) (\mathbf{V}_n + \boldsymbol{\mu}_n) = \sum_{j=1}^{n-m} \bar{\lambda}_{nj}(\boldsymbol{\theta}) (Y_{nj} + \bar{\mu}_{nj}(\boldsymbol{\theta}))^2,$$

where  $Y_{nj} = \mathbf{q}_j^{(n)\top} \mathbf{V}_n$  and  $\bar{\mu}_{nj}(\boldsymbol{\theta}) = \mathbf{q}_j^{(n)\top} \boldsymbol{\mu}_n$ ,  $j = 1, 2, \dots, n - m$ .

**Proposition 3.** For any fixed  $j > 0$ ,  $\bar{\lambda}_{nj}(\boldsymbol{\theta})$  converges to  $\bar{\lambda}_j$  and  $\mathbf{q}_j^{(n)}$  converges to  $\mathbf{q}_j$ , where  $\bar{\lambda}_j$  and  $\mathbf{q}_j = (q_{1j}, q_{2j}, \dots)^\top$  are the  $j$ th eigenvalue and eigenvector of the following infinite-dimensional matrix.

$$D_\infty(\boldsymbol{\theta}) = \left( \lambda_j^{\frac{1}{2}} \lambda_k^{\frac{1}{2}} \left( \int_0^1 \int_0^1 f_j(u) h(u, v) f_k(v) dudv \right), 1 \leq j, k < \infty \right),$$

where  $h(u, v) = \delta(u - v) - \mathbf{g}^\top(u, \boldsymbol{\theta}) A^{-1} \mathbf{g}(v, \boldsymbol{\theta})$  with the Dirac delta function  $\delta(u)$ .

**Proposition 4.** Any finite-dimensional random vector  $(Y_{nj_1}, Y_{nj_2}, \dots, Y_{nj_p})$  converges in distribution to a normally distributed random vector  $(Y_{j_1}, Y_{j_2}, \dots, Y_{j_p})$  with a mean  $\mathbf{0}$  and variance  $I_p$  as  $n$  tends to infinity.

If we note that  $\sum_{j=1}^{n-m} \bar{\lambda}_{nj}(\boldsymbol{\theta}) \bar{\mu}_{nj}^2(\boldsymbol{\theta}) \leq \sum_{j=1}^n \lambda_{nj} \mu_{nj}^2$ , we have the following theorem, which is similar to that seen in Theorem 1.

**Theorem 2.** The distribution of  $W_n^2(\hat{\boldsymbol{\theta}})$  converges to that of  $\sum_{j=1}^\infty \bar{\lambda}_j (Y_j + \bar{\mu}_j)^2$  in distribution as  $n$  tends to infinity, where  $Y_1, Y_2, \dots$ , are independent and identically random variables distributed as a standard normal distribution, and

$$\bar{\mu}_j = \sum_{l=1}^\infty q_{lj} \mu_l, \quad j = 1, 2, \dots,$$

where the  $\mu_l$  terms are the same as those in Theorem 1.

In the case where no estimation is involved, the asymptotic distribution derived in Theorem 2 is consistent with the classical result that is derived from the theory of a Brownian bridge (Shorack & Wellner, 1986). In fact, the eigenvalues  $\bar{\lambda}_{js}$  are consistent with those of the integral equation with the kernel function

$$\rho(s, t) = \int_0^1 \int_0^1 h(s, u) h(t, v) \rho_0(u, v) dudv.$$

However, it is well known that it is not trivial to solve such an integral equation accurately with any reasonable cost of computation. The result above gives us an alternative and simplified way to perform the calculation.

### 3.2 The Distribution for Finite $n$

To evaluate the distribution of  $W_n^2(\boldsymbol{\theta})$  or  $W_n^2(\hat{\boldsymbol{\theta}})$  for finite  $n$ , the proofs in the Appendix suggest a good way of performing the calculation. In the case of Theorem 1, finding the eigenvalues and the eigenvectors of  $S_n \mathbf{B}^2 S_n^\top$ , then  $\sum_{j=1}^p \lambda_{nj} (Y_j + \mu_{nj})^2$  gives us a good approximation of  $W_n^2(\boldsymbol{\theta})$  for an appropriate choice of  $p < n$ . In the case of Theorem 2, finding the eigenvalues and eigenvectors of  $D_n(\hat{\boldsymbol{\theta}})$ , then  $\sum_{j=1}^{n-m} \bar{\lambda}_{nj}(\hat{\boldsymbol{\theta}}) (Y_j + \bar{\mu}_{nj}(\hat{\boldsymbol{\theta}}))^2$  gives us a good approximation of  $W_n^2(\hat{\boldsymbol{\theta}})$  for an

appropriate value of  $p < n - m$ . The replacement of  $\theta$  by  $\hat{\theta}$  in the calculation is justified by the consistency of  $\hat{\theta}$  irrespective of the existence of contamination. In fact,

$$\sum_{j=1}^{n-m} \bar{\lambda}_{nj}(\hat{\theta})(Y_j + \bar{\mu}_j(\hat{\theta}))^2 - \sum_{j=1}^{n-m} \bar{\lambda}_{nj}(\theta)(Y_j + \bar{\mu}_j(\theta))^2$$

converges to 0 in probability as  $n$  tends to infinity. It follows from the strong consistency of  $\hat{\theta}$  and Lemma A 1 in the Appendix.

#### 4. Investigation of Robustness by Simulations

It is still unclear how advantageous the use of the minimum distance estimator is over other estimators, for example, the maximum likelihood estimator. In this section, via Monte Carlo simulations we demonstrate that the power curves remain almost identical when no contamination exists, but the rejection probability is significantly affected by the existence of contamination when the maximum likelihood estimator is used.

The sample size here is fixed at 200, and the significance level is 0.1. The solid line in each panel is for the case of the minimum distance estimator and the dotted line is for the case of the maximum likelihood estimator.

The power curves given in Figure 1 suggest that there is no significant difference between the minimum distance and the maximum likelihood estimator when no contamination exists. The left panel shows the power curves for Exponential against Gamma with the parameter  $\nu$ , and the right panel shows them for Normal against Student's  $t$  with the parameter  $\nu = 1/d$ , where  $d$  is the number of degrees of freedom.

On the other hand, if contamination exists, the rejection probability quickly increases for the maximum likelihood estimator, but it does not increase for the minimum distance estimator. Two examples are shown in Figure 2. The results indicate that the use of the minimum distance estimator makes the goodness-of-fit test robust to contamination.

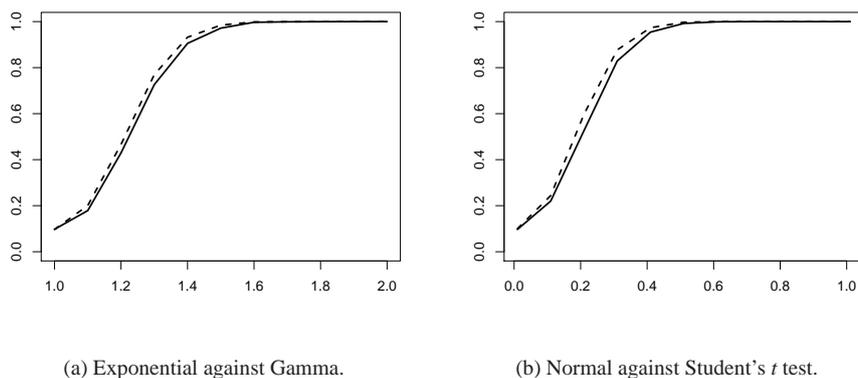


Figure 1. Power curves.

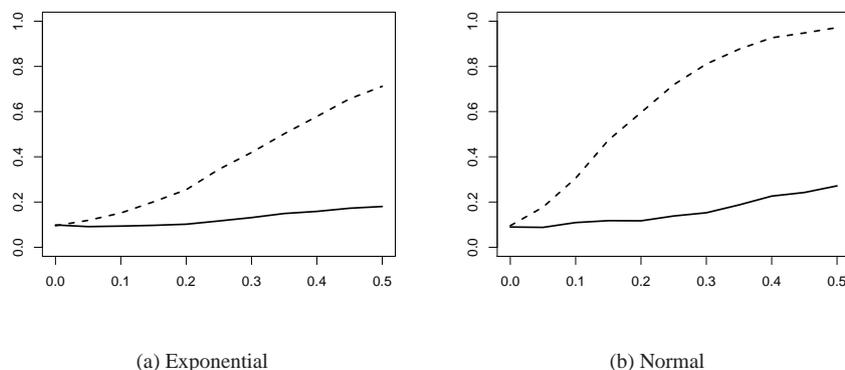


Figure 2. Rejection probability against the rate of contamination  $\epsilon$ .

## 5. Concluding Remarks

The asymptotic distributions of the Cramér-von Mises statistic are derived when the observations are contaminated. Numerical experiments indicate that the use of the minimum distance estimator makes the test less sensitive to contamination, although the power stays almost the same as that of the maximum likelihood estimator. Such insensitivity would be harmful when the aim of the test is to detect the existence of contamination. However, it becomes advantageous if the aim of the test is to check whether the underlying distributional model is usable or not. It often happens in practice that the hypothesis testing is not a goal but a beginning of the analysis. Then, a goodness-of-fit test statistic which is insensitive to small number of contaminations is preferred.

We leave any other problems for further investigations, including a more explicit analysis of sensitivity to the contamination and extension of the theorems for other types of estimators. Also extension to multivariate case would be interesting, although the distribution of the statistic becomes much more complicated even when the parameters are known (Rosenblatt, 1952).

## References

- Beutner, E., & Bordes, L. (2011). Estimators based on data-driven generalized weighted Cramér-von Mises distances under censoring—with applications to mixture models. *Scandinavian Journal of Statistics*, 38, 108-129. <http://dx.doi.org/10.1111/j.1467-9469.2010.00708.x>
- Garcia-Dorado, A., & Marin, JM. (1998). Minimum distance estimation of mutational parameters for quantitative traits. *Biometrics*, 54, 1097-1114. <http://dx.doi.org/10.2307/2533860>
- Guttorp, P., & Lockhart, R.A. (1988). On the asymptotic distribution of quadratic forms in uniform order statistics. *The Annals of Statistics*, 16, 433-449. <http://dx.doi.org/10.1214/aos/1176350713>
- Millar, P. W. (1981). Robust estimation via minimum distance methods. *Probability Theory and Related Fields*, 55, 73-89. <http://dx.doi.org/10.1007/bf01013462>
- Riesz, F., & Nagy, B. S. (1955). *Functional analysis* (translated from the 2d French ed. by Leo F. Boron) ch. 9, New York: Dover Publications, Inc.
- Rosenblatt, M. (1952). Limit Theorems Associated with Variants of the Von Mises Statistic. *The Annals of Mathematical Statistics*, 23, 617-623. <http://dx.doi.org/10.1214/aoms/1177729341>
- Shorack, G. R., & Wellner, J. A. (1986). *Empirical processes with applications to statistics*. New York: Wiley.
- Woodward, W. A., Parr, W. C., Schucany, W. R., & Lindsey, H. (1984). A comparison of minimum distance and maximum likelihood estimation of a mixture proportion. *Journal of the American Statistical Association*, 79, 590-598. <http://dx.doi.org/10.1080/01621459.1984.10478085>

## Appendix

*A lemma for the proofs.* Before giving the proofs of the propositions, we will make sure that the convergence of the eigenvalues and the eigenvectors appear there. The following lemma can be derived from the theorem on page 372 of Riesz & Nagy (1955), which states that eigenvalues and the eigenfunctions of an integral equation are continuous with respect to the kernel function as far as it belongs to the space

$$L_2(\nu \times \nu) = \left\{ k(x, y); \int \int k^2(x, y) d\nu(x) d\nu(y) < \infty \right\}$$

with the norm  $\|\cdot\|$ , where  $\nu$  is a sigma finite measure.

**Lemma A 1.** Let  $\lambda_{n_j}$  and  $f_n^{(j)}(u)$  be the  $j$ th eigenvalue and eigenfunction of the integral equation, respectively

$$\lambda f(u) = \int k_n(u, v) f(v) dv,$$

and  $\lambda_j$  and  $f^{(j)}(u)$  be the  $j$ th eigenvalue and eigenfunction of the integral equation, respectively

$$\lambda f(u) = \int k(u, v) f(v) dv.$$

If  $k_n(x, y)$  is a compact operator and  $\|k_n - k\|$  converges to 0 as  $n$  tends to infinity, then  $\lambda_{n_j}$  converges to  $\lambda_j$  and  $\|f_n^{(j)} - f^{(j)}\|$  converges to 0 as  $n$  tends to infinity for  $j = 1, 2, \dots$  when  $f_n^{(j)}$  and  $f^{(j)}$  are properly normalized, including their signs.

*Proof of Proposition 1.* We first rewrite the equation  $\lambda_{nj} \mathbf{p}_j^{(n)} = S_n B^2 S_n^\top \mathbf{p}_j^{(n)}$  as the integral equation

$$\lambda_{nj} f_n^{(j)}(u) = \int k_n(u, v) f_n^{(j)}(v) dv,$$

by using the kernel

$$k_n(u, v) = \frac{n}{(n+1)^2} \sum_{l=1}^{n+1} \left( 1_{\lceil nu \rceil \leq l} - \frac{\lceil nu \rceil}{n+1} \right) b_l^2 \left( 1_{\lceil nv \rceil \leq l} - \frac{\lceil nv \rceil}{n+1} \right)$$

and the eigenfunction  $f_n^{(j)}(u) = \sqrt{n} \rho_{\lceil nu \rceil}^{(n)}$ . Then, if  $k_n(u, v)$  is written in the form of an integral, the convergence to  $\rho_0(u, v) = \int_0^1 (1_{u \leq s} - u)(1_{v \leq s} - v) ds$  is clear from the fact  $|k_n(u, v)| \leq 1$  for any  $0 \leq u, v \leq 1$ . The desired result follows from Lemma A1. Note that the eigenvalues and eigenfunctions of the integral equation for the kernel  $\rho_0(u, v)$  are  $1/(j\pi)^2$  and  $\sqrt{2} \sin(\pi ju)$ , respectively,  $j = 1, 2, \dots$  □

*Proof of Proposition 2.* We first note that the  $k$ th element of  $\mathbf{U}_n$  can be replaced by  $E_k / \sum_{j=1}^{n+1} E_j$ , where  $E_j, j = 1, 2, \dots, n+1$ , are independent and identically distributed exponential variables with a mean of one. Let  $c_{jk}$  be the  $(j, k)$  element of  $\Lambda_n^{-\frac{1}{2}} P_n^\top S_n B$ . Then, it is enough to show that for any  $\mathbf{t} = (t_1, t_2, \dots, t_p) \in \mathbb{R}^p$ ,

$$\frac{n+1}{\sum_{j=1}^{n+1} E_j} \sum_{j=1}^p t_j \sum_{k=1}^{n+1} c_{jk} (E_k - 1) = \frac{n+1}{\sum_{j=1}^{n+1} E_j} \sum_{k=1}^{n+1} \left( \sum_{j=1}^p t_j c_{jk} \right) (E_k - 1)$$

converges to a normally distributed random variable  $V = \mathbf{t}^\top \mathbf{V}$  with a mean of 0 and variance  $\mathbf{t}^\top \mathbf{t}$ . It is easily seen that the Lindeberg condition is satisfied for

$$V_n = \sum_{k=1}^{n+1} \left( \sum_{j=1}^p t_j c_{jk} \right) (E_k - 1).$$

In fact, the inequality

$$\frac{1}{\mathbf{t}^\top \mathbf{t}} \sum_{k=1}^{n+1} \mathbb{E} \left[ V_n^2 1_{|V_n| > \varepsilon \sqrt{\mathbf{t}^\top \mathbf{t}}} \right] \leq H \left( \frac{\varepsilon}{\sqrt{p} c_n} \right)$$

implies the desired result, where the function

$$H(x) = \mathbb{E} \left[ (E_k - 1)^2 1_{|E_k - 1| > x} \right]$$

is a monotone decreasing to 0, and

$$c_n = \max_{1 \leq j \leq p, 1 \leq k \leq n} |c_{jk}| \leq \frac{b_k \lambda_{nj}^{-\frac{1}{2}}}{n+1} \left( \sum_{l=1}^n \left( 1_{l \geq k} - \frac{l}{n+1} \right) \right)^{\frac{1}{2}}.$$

□

*Proof of Proposition 3.* By taking  $\nu$  as a counting measure in Lemma A1,  $D_n(\boldsymbol{\theta})$  and  $D_\infty(\boldsymbol{\theta})$  can be considered as compact operators on  $L_2(\nu)$ . The first term on the right hand side of

$$\|D_n(\boldsymbol{\theta}) - D_\infty(\boldsymbol{\theta})\| \leq \sum_{j=1}^p \sum_{k=1}^p (d_{jk}^{(n)} - d_{jk})^2 + 2 \sum_{j=1}^n \sum_{k=p+1}^n (d_{jk}^{(n)} - d_{jk})^2 + 2 \sum_{j=1}^\infty \sum_{k=n+1}^\infty d_{jk}^2$$

converges to 0 for any fixed  $p$ , since the  $(j, k)$  element of  $D_n(\boldsymbol{\theta})$  converges to that of  $D_\infty(\boldsymbol{\theta})$ . The convergence of the second term to 0 can be shown as the following

$$\sum_{k=p+1}^n \left( \sum_{j=1}^n d_{jk}^{(n)2} + \sum_{j=1}^n d_{jk}^2 \right) \leq \left( \sum_{j=1}^n \lambda_{nj} \right) \sum_{k=p+1}^n \lambda_{nj} + \alpha \left( \sum_{j=1}^n \lambda_j \right) \sum_{k=p+1}^n \lambda_k, \tag{1}$$

where  $\alpha = \max_{j,k} \lambda_j^{-\frac{1}{2}} \lambda_k^{-\frac{1}{2}} |d_{jk}|$ . Here  $\sum_{j=1}^{\infty} \lambda_j = 1/6$  and

$$\sum_{j=1}^n \lambda_{nj} = \text{trace}(S_n B^2 S_n^T) = \sum_{j=1}^n \sum_{k=1}^{n+1} \left(1_{j \geq k} - \frac{j}{n+1}\right)^2 b_k^2 \leq \frac{1}{6} \left(1 - \frac{\varepsilon}{\sqrt{n}}\right)^{-2},$$

because  $0 \leq f(x, \theta)/f_\varepsilon(x, \theta) \leq (1 - \varepsilon/\sqrt{n})^{-1}$  for any  $x$  and  $\varepsilon^2 < n$ . Therefore, by taking a large enough value of  $p$ , the right hand side of (1) will be sufficiently small. The convergence of the last term is clear from the inequality

$$\sum_{j=1}^{\infty} \sum_{k=n+1}^{\infty} d_{jk}^2 \leq \frac{1}{6} \alpha^2 \left( \sum_{k=n+1}^{\infty} \lambda_k \right).$$

If we note that  $D_n(\theta)$  is a compact operator, the proof is complete from Lemma A1.

□

*Proof of Proposition 4.* Using a similar argument to that given in the proof of Proposition 2, it is enough to show that

$$\max_{1 \leq j \leq p, 1 \leq k \leq n} \left| \sum_{l=1}^n q_{lj}^{(n)} c_{lk} \right|$$

converges to 0 as  $n$  tends to infinity. Using the fact in A.3 that  $c_n = \max_{1 \leq l \leq p, 1 \leq k \leq n} |c_{lk}|$  converges to 0 as  $n$  tends to infinity, we see that the first term on the right hand side of

$$\left| \sum_{l=1}^n q_{lj}^{(n)} c_{lk} \right| \leq \left| \sum_{l=1}^{p'} q_{lj}^{(n)} c_{lk} \right| + \left( \sum_{l=p'+1}^n q_{lj}^{(n)2} \right)^{\frac{1}{2}} \left( \sum_{l=p'+1}^n c_{lk}^2 \right)^{\frac{1}{2}}$$

converges to 0 as  $n$  tends to infinity for any  $p' < n$ . If we note that  $\sum_{l=1}^n q_{lj}^{(n)2} = 1$ ,  $\sum_{l=1}^n c_{lk}^2 \leq 1$  and Lemma A1, the proof is complete. □

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).