GAUSSIAN CENTERED L-MOMENTS

Hyowon An

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research (Statistics).

Chapel Hill
2017

Approved by:

Edward Carlstein

Nicolas Fraiman

Jan Hannig

J. S. Marron

Kai Zhang

ABSTRACT

HYOWON AN: Gaussian Centered L-moments
(Under the direction of J. S. Marron and Kai Zhang)

As various types of media currently generate so-called *big data*, data visualization faces the challenge of selecting a few representative variables that best summarize important structure inherent in data. The conventional moments-based summary statistics can be useful for the purpose of variable screening. In particular, they can find important distributional features such as bimodality and skewness. However, their sensitivity to outliers can lead to selection based on a few extreme outliers rather than distributional shape. To address this type of non-robustness, we consider the L-moments. But, describing a marginal distribution with the L-moments has an intuitive limitation in practice because these moments take zero values at the uniform distribution; the interest usually lies in the shape of the marginal distribution compared to the Gaussian, but the sign and magnitude of the L-moments are not as useful as expected for this purpose. As a remedy, we propose the *Gaussian Centered L-moments* with zeros at the Gaussian distribution while sharing robustness of the L-moments. The Gaussian Centered L-moments can be especially useful for gene expression data in which variable screening corresponds to finding biologically meaningful genes. The mixtures of Gaussian distributions seems to be underlying mechanism generating gene expression profiles, and this suggests moments that are sensitive to departure from Gaussianity.

This dissertation deeply investigates theoretical properties of the Gaussian Centered L-moments in various ways. First, by the means of Oja's criteria, the first four terms of the Gaussian Centered L-moments are shown to describe the shape of a distribution in a physically meaningful fashion. Second, comparison between robustness of the conventional, L- and Gaussian Centered L-moments is made based on asymptotic behavior of their influence functions on Tukey's h distributions. Third, the efficiencies of these moments in capturing

departure from Gaussianity are compared by developing Jarque-Bera type goodness-of-fit test statistics for Gaussianity. While developing such test statistics, a method for obtaining optimal balance between skewness and kurtosis estimators is introduced. Finally, comprehensive performances including both the robustness and efficiency of the different moments on high dimensional gene expression data are analyzed by the *Gene Set Enrichment Analysis*.

To Eun Young Oh

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

$\mathbb{R}$	Real line.
$\mathbb{R}_+$	Nonnegative part of the real line, i.e. $\{x \in \mathbb{R} \mid x \geq 0\}$. First random
$X_1, X_2, \cdots, X_n$	sample.
$Y_1, Y_2, \cdots, Y_m$	Second random sample.
F	Distribution of $X_1, X_2, \cdots, X_n$.
G	Distribution of $Y_1, Y_2, \cdots, Y_m$.
$X_{i:n}$	i-th order statistic of the random sample $X_1, X_2, \cdots, X_n$.
$F_{a,b}$	Distribution of $aX_1 + b, aX_2 + b, \cdots, aX_n + b$ for $a \neq 0, b \in \mathbb{R}$.
F^{-1}, G^{-1}	Quantile functions of F, G, respectively.
$\mathcal{F}$	Family of absolutely continuous and strictly increasing distributions.
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with the mean μ and the variance σ^2.
$\mathcal{U}(a, b)$	Uniform distribution with the lower bound a and upper bound b.
$\mathcal{T}(g, h)$	Tukey's distribution with the parameters g and h.
γ_1, γ_2	Conventional skewness and kurtosis, respectively.
P_r^*	r-th order shifted Legendre polynomial.
H_r	r-th order Hermite polynomial.
θ	Functional from $\mathcal{F}$ to $\mathbb{R}$.
λ_r	r-th L-moment.
η_r	r-th Hermite L-moment.
ρ_r	r-th rescaled L-moment.

CHAPTER 1
Introduction

Data quality is an issue that is currently not receiving as much attention as it deserves in the age of big data. Traditional careful analysis of small data sets involves a study of marginal distributions, which easily finds data quality challenges such as skewness and suggests remedies such as data transformation. Furthermore, unusual marginal distributional structure can suggest potential new scientific investigation. Direct implementation of this type of operation is difficult with high dimensional data, as there are too many marginal distributions to individually visualize. This hurdle can be overcome by using summary statistics to select a representative set for visualization and potential remediation. Traditional summaries such as the sample mean, variance, skewness and kurtosis can be very useful for this process. However, as seen in Chapter 5, those have some limitations for this purpose, e.g. they can be strongly influenced by outliers. In some situations outliers are important and well worth finding, but in other cases summaries that are dominated by them (such as those based on the moments) can miss more important features of some variables, such as bimodality.

For this purpose, this dissertation proposes, and deeply studies, some new univariate distributional summaries. The starting point is the *L-moments* (Hosking, 1990), which are known to have good robustness properties against outliers, while being interpretable because of their intuitive definition in terms of expected order statistics. Linear combinations of order statistics, so-called *L-statistics* are typically chosen estimators of the L-moments. L-statistics were first proposed in the general research area of robust statistics, in particular robustness against outliers; see Hampel et al. (2011), Staudte and Sheather (2011) and Huber and Ronchetti (2009). With the goal of screening for non-Gaussianity, a limitation of the classical L-moments is that they are not centered at the typically expected (and frequently appearing in real data) Gaussian distribution, but instead are zero at the uniform

distribution. Zero at the uniform distribution hinders interpreting the signs and magnitudes of the L-moments, especially in terms of the critical notion of *kurtosis*, which should be negative for bimodal distributions and highly positive for distributions with high peaks and heavy tails.

This dissertation focuses on skewness and kurtosis as directions of departure from Gaussianity. Those distributional aspects have gained relatively little attention from the robustness community. The theoretical *skewness* γ_1 and (*excess*) *kurtosis* γ_2 of a random variable X are defined as

$$\gamma_1 = \frac{E\left(X - EX\right)^3}{\left(E\left(X - EX\right)^2\right)^{3/2}}, \quad \gamma_2 = \frac{E\left(X - EX\right)^4}{\left(E\left(X - EX\right)^2\right)^2}. \tag{1.1}$$

Given a random sample $X_1, X_2, \cdots, X_n$ and their sample mean $\bar{X}$, those are often estimated by the *sample skewness* $\hat{\gamma}_1$ and *sample kurtosis* $\hat{\gamma}_2$ defined as

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^3}{\left(\sum_{i=1}^n \left(X_i - \bar{X}\right)^2\right)^{3/2}}, \quad \hat{\gamma}_2 = \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^4}{\left(\sum_{i=1}^n \left(X_i - \bar{X}\right)^2\right)^2}. \tag{1.2}$$

Each observation X_i influences both the sample skewness and kurtosis with polynomial degree. If these statistics are used to select a few representative variables in high dimensional data, variables with outliers will tend to be selected rather than variables with more meaningful notions of skewness and kurtosis in their distributional bodies.

The limitation of conventional summary statistics in practice is demonstrated using a modern high dimensional data set from cancer research in Chapter 5. These data are part of the TCGA project, and were first studied in (Ciriello et al., 2015) and (Hu et al., 2015). The precise version of the data here was used in (Feng et al., 2015). The data consist of 16,615 genes and 817 breast cancer patients, each of which is labelled according to 5 subtypes. While much is known about this data, as discussed in (Feng et al., 2015), the sheer data size means there have only been preliminary studies of the marginal (individual gene) distributions. In this study we do a much deeper search for genes with unexpected marginal structure. This provides a nonstandard, but very useful basis for the comparison of marginal distributional summaries.

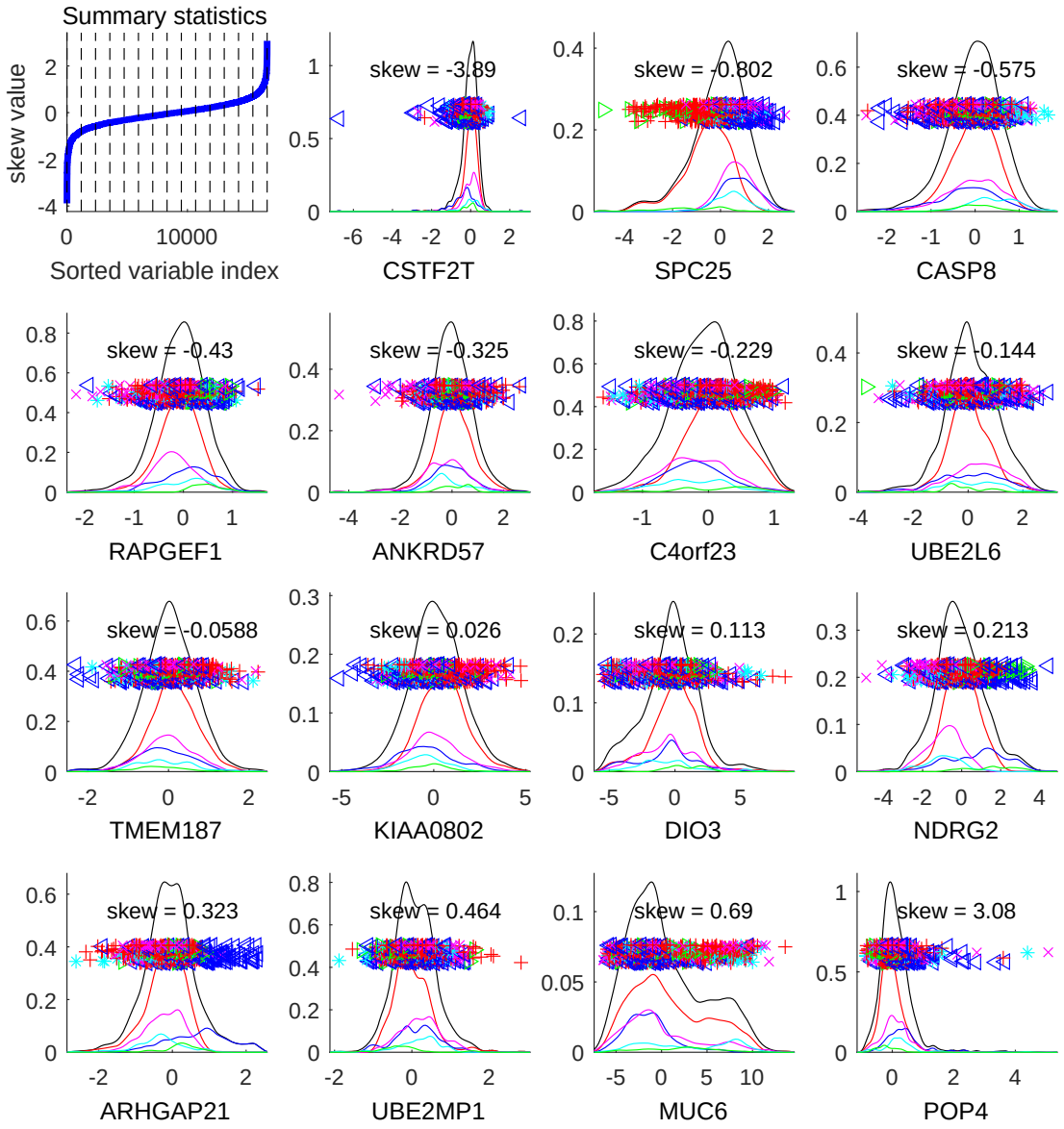

Figure 1.1: The marginal distribution plots of the 15 genes with equally spaced sample conventional skewness values. The upper left most plot is a quantile plot. The first and last genes should show skewness to the left and right sides, respectively, but they actually have a couple of strong outliers instead.

Subtype	LumA	LumB	Her2	Basal	Normal-like
Symbol	+	×	*	◁	▷

Table 1.1: The symbols and colors corresponding to the 5 breast cancer subtypes in the marginal distribution plots. The upper left most plot is a quantile plot. For detailed explanation, see Chapter 1.

Figure 1.1 shows the marginal distributions of the 15 variables (or genes) with the equally spaced sample skewness values. The upper left most plot shows the distribution of the sorted values of the summary statistics, as the quantile function. The dashed vertical lines indicate the locations of the displayed genes in the sorted list. The remaining plots are a selected subset of the marginal distribution plots that correspond to equally spaced sample quantiles of the sample skewness values. Each symbol represents a breast cancer patient by colors and symbol based on subtypes; which are very important to determining which modern treatments are best for which patients. See Table 1.1 for reference. The height of each symbol provides visual separation, based on order in the data set. The black solid line is a kernel density estimate, i.e. smooth histogram, of the marginal distribution, with colored sub-densities corresponding to subtypes. The first and last genes selected by the sample skewness have a couple of strong outliers on its left and right sides, respectively. This does not realize the goal of finding genes with strong distributional body skewness, for example driven by differing subtype behavior.

On the other hand, Figure 1.2 shows the marginal distributions of the 15 genes with the equally spaced sample kurtosis values. The genes with low sample kurtosis values have bi- or multimodality. It seem that those genes do not have strong outliers. This is a natural result of the property of kurtosis that its negative side indicates large flanks, with light tails and a low peak (i.e. bimodality) of a distribution. Since outliers affect measures of heavy-tailedness rather than light-tailedness, the negative side of sample kurtosis are not affected by outliers. On the contrary, the genes with high sampled kurtosis values, TDRD3 and CSTF2T, have a couple of outliers on one side or both sides of their marginal distributions. This shows that the conventional kurtosis is not effective at picking up genes possessing distributional kurtosis.

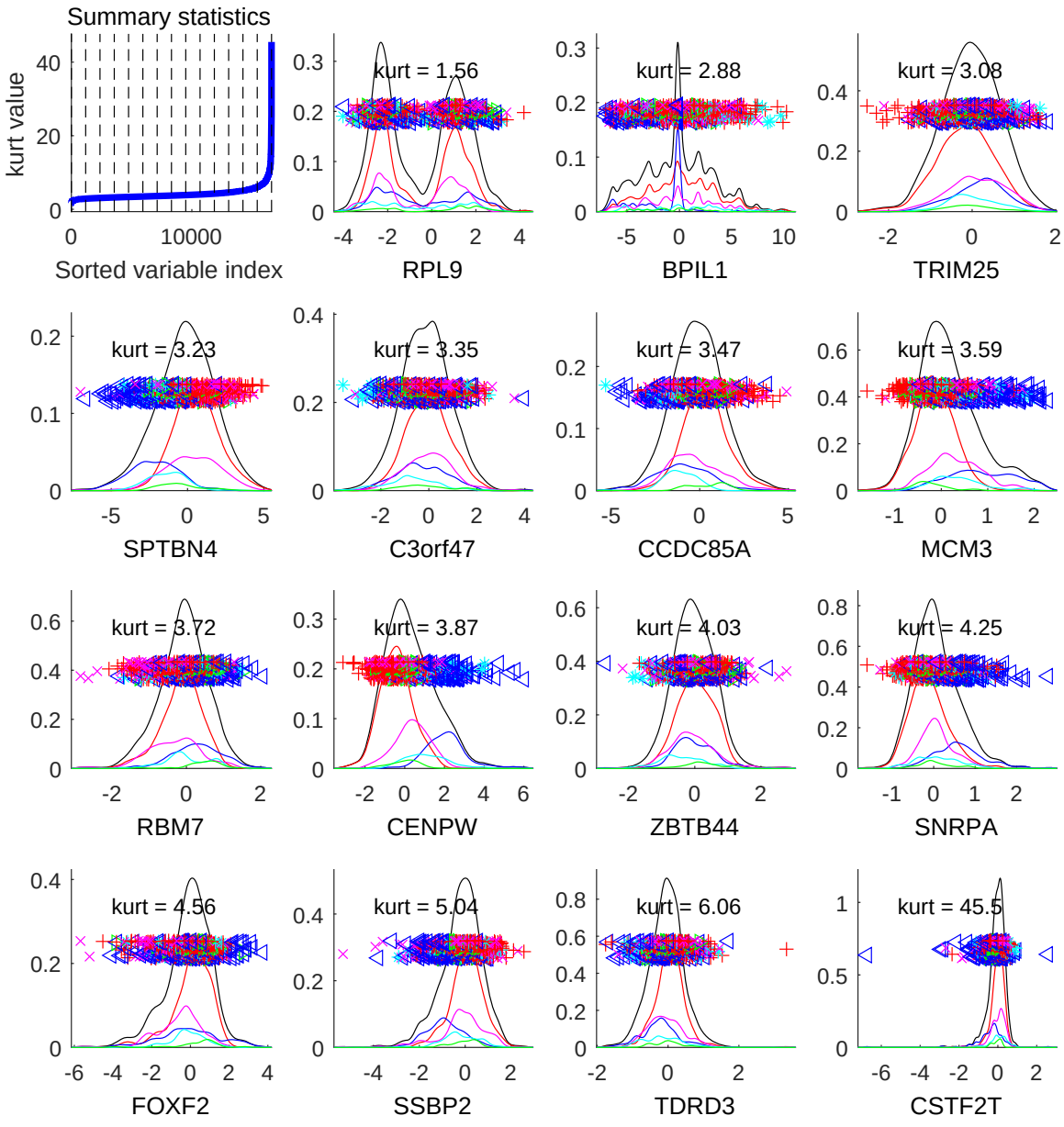

Figure 1.2: The marginal distribution plots of the 15 genes with equally spaced sample kurtosis values. The upper left most plot is a quantile plot. The last gene should have heavy tailed-ness on both sides of its marginal distribution, but the gene CSTF2T actually has a couple of strong outliers on the both sides.

1.1 Preliminaries

Throughout this dissertation, we assume that we have two random variables X and Y following absolutely continuous cumulative distribution functions (CDF) F and G with probability density functions f and g respectively. Also, a random sample $X_1, X_2, \cdots, X_n$ is assumed to be generated from F, and $X_{i:n}$ denotes the i-th order statistic of the random sample. Following Oja (1981), the cumulative distribution function F is said to be *strictly increasing* if it is strictly increasing on its *support*, $S(F) = \overline{\{x|0 < F(x) < 1\}}$ where $\overline{S}$ indicates the closure of $S \subset \mathbb{R}$. In this dissertation we consider only the family $\mathcal{F}$ of absolutely continuous and strictly increasing cumulative distribution functions. The quantile function F^{-1} is then the inverse of $F \in \mathcal{F}$ defined on $(0, 1)$.

Various kinds of orthogonal polynomials are introduced and used throughout this paper. One of the most famous sequences of orthogonal polynomials is the *Legendre polynomials* $\{P_r\}_{r\geq 0}$ which have been comprehensively investigated in Szegö (1959). The Legendre polynomials are orthogonal to each other on the interval $[-1, 1]$ with respect to the weight function $w(x) = 1$, i.e.

$$\int_{-1}^{1} P_{r_1}(u)P_{r_2}(u)\,\mathrm{d}u = 0$$

for all $r_1, r_2 = 0, 1, \cdots$ such that $r_1 \neq r_2$. The *shifted Legendre polynomials* $\{P_r^*\}_{r\geq 0}$ are linear transformations of the Legendre polynomials such that

$$P_r^*(u) = P_r(2u - 1) \text{ for } 0 \leq u \leq 1$$

for all $r = 0, 1, \cdots$. The shifted Legendre polynomials are orthogonal to each other on the unit interval $[0, 1]$ with respect to the weight function $w(x) = 1$ which is the uniform density function. That is,

$$\int_{0}^{1} P_{r_1}^*(u)P_{r_2}^*(u)\,\mathrm{d}u = 0$$

for all $r_1, r_2 = 0, 1, \cdots$ such that $r_1 \neq r_2$.

Another sequence of orthogonal polynomials of interest given in Szegö (1959) is the *Hermite polynomials* $\{H_r'\}_{r\geq 0}$. The Hermite polynomials are orthogonal to each other on

the real line $\mathbb{R}$ with respect to the weight function $w(x) = e^{-x^2}$, i.e.

$$\int_{-\infty}^{\infty} e^{-x^2} H'_{r_1}(x) H'_{r_2}(x)\,\mathrm{d}x = 0$$

for all $r_1, r_2 = 0, 1, \cdots$ such that $r_1 \neq r_2$. The sequence $\{H'_r\}_{r\geq 0}$ is sometimes called the *physicists' Hermite polynomials*. The primary focus of this dissertation is a variation of physicists' Hermite polynomials, which is called the *probabilists' Hermite polynomials*. These have a different weight function $w(x) = e^{-x^2/2}$ which is proportional to the standard Gaussian density such that

$$\int_{-\infty}^{\infty} e^{-x^2/2} H_{r_1}(x) H_{r_2}(x)\,\mathrm{d}x = 0$$

for all $r_1, r_2 = 0, 1, \cdots$ such that $r_1 \neq r_2$. When we refer to the Hermite polynomials in this paper, we indicate the probabilists' Hermite polynomials.

The following is the list of mathematical notations used in this paper.

- If J is a function, then $J(\cdot)$ denotes the function itself while $J(x)$ denotes the value of J evaluated at x.
- If $\hat{\theta}$ is a statistic, $E_F\left(\hat{\theta}\right)$ and $\mathrm{Var}_F\left(\hat{\theta}\right)$ indicate the mean and variance of $\hat{\theta}$ with respect to the distribution F. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two statistics, their covariance with respect to the distribution F is denoted by $\mathrm{Cov}_F\left(\hat{\theta}_1, \hat{\theta}_2\right)$.
- The function $F_{a,b}$ denotes the cumulative distribution function of $aX + b$.
- The function $\phi\left(\cdot|\mu, \sigma^2\right)$ and $\Phi\left(\cdot|\mu, \sigma^2\right)$ are the probability density function and cumulative distribution function of $\mathcal{N}\left(\mu, \sigma^2\right)$, the Gaussian distribution with mean μ and variance σ^2, respectively.

1.2 L-statistics and L-moments

The term *L-statistic* is used as a term indicating a statistic in the form of a *linear combination of order statistics* (Andrews et al., 1972). An L-statistic is generally expressed

as follows

$$\sum_{i=1}^{n} c_{ni} X_{i:n} \tag{1.3}$$

where c_{ni} is a function of both i and n, and $X_{i:n}$ is the i-th order statistic such that $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ is a reordering of the random sample $X_1, X_2, \cdots, X_n$. Theoretical properties and abundant examples of the order statistics are given in Chapter 8 of David and Nagaraja (2003). Equation (1.3) is more often expressed in the form

$$\tilde{\theta}_n = \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) X_{i:n}. \tag{1.4}$$

where $J : (0,1) \to \mathbb{R}$ is a measurable function. Section 8.2 of Serfling (1980) and Chapter 19 of Shorack and Wellner (2009) enumerate various sets of conditions on the function J and distribution F that guarantee that $\tilde{\theta}_n$ converges, under various modes of convergence in the limit as $n \to \infty$, to the quantity

$$\theta(F) = \int_{-\infty}^{\infty} x f(x) J(F(x))\, \mathrm{d}x = \int_0^1 F^{-1}(u) P_{r-1}^*(u) \mathrm{d}u. \tag{1.5}$$

We call the functional $\theta : \mathcal{F} \to \mathbb{R}$ in the form (1.5) an *L-functional*.

A connection between L-statistics and location, scale, skewness and kurtosis of a distribution has been made by Hosking (1990). That paper presented a way to use expected order statistics in describing the shape of a distribution and called the moments defined in such a way the theoretical *L-moments*. The r-th L-moment of a random variable X is usually defined as

$$\lambda_r = \int_{-\infty}^{\infty} x f(x) P_{r-1}^*(F(x)) \mathrm{d}x = \int_0^1 F^{-1}(u) P_{r-1}^*(u) \mathrm{d}u \tag{1.6}$$

for $r = 1, 2, \cdots$ where F^{-1} is the quantile function of F and P_r^* is the r-th order shifted Legendre polynomial which is explained in Chapter 4 of (Szegö, 1959). The r-th *L-moment ratio* is defined as $\lambda_r^* = \lambda_r / \lambda_2$ for $r = 3, 4, \cdots$. Some nice properties of the L-moments presented in Hosking (1990) are as follows.

- Every integrable random variable has finite L-moment values.

- A distribution whose mean exists is identified by its sequence of L-moments.
- The first four L-moment based measures $\lambda_1, \lambda_2, \lambda_3^*$ and λ_4^* satisfy Oja's criteria (Section 1.3) for measures of location, scale, skewness and kurtosis.
- The L-moment ratios are bounded between -1 and 1.

The L-moments can be estimated from a random sample $X_1, X_2, \cdots, X_n$ in various ways by the empirical L-statistics. A natural estimator in the form of Equation (1.4) is

$$\tilde{\lambda}_{n,r} = \sum_{i=1}^{n} P_{r-1}^{*}\left(\frac{i}{n+1}\right) X_{i:n}.$$

However, this estimator suffers from being biased, i.e. $E\Big(\tilde{\lambda}_r\left(X_1, X_2, \cdots, X_n\right)\Big) \neq \lambda_r(F)$ for many distributions F. Hosking (1990) adopted the U-statistics based estimators

$$\hat{\lambda}_{n,r} = \begin{pmatrix} n \\ r \end{pmatrix}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_r \leq n} \cdots \sum \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \begin{pmatrix} r-1 \\ k \end{pmatrix} X_{i_{r-k}:n} \tag{1.7}$$

which is unbiased. Throughout this dissertation, the U-statistics based estimators $\hat{\lambda}_r$ will be used as the *sample L-moments*. The *sample L-moment ratios* are defined as $\hat{\lambda}_r^* = \hat{\lambda}_r / \hat{\lambda}_2$ accordingly.

1.3 Oja's criteria

When defining new measures of location, scale, skewness and kurtosis, a challenge is to ensure that those new measures reflect the intuitive meanings of corresponding distributional properties. This challenge was elegantly addressed by the framework of Oja (1981) using stochastic dominance ideas, which is applied here. Intuitively, a functional being a measure of any property of a distribution means that the functional should preserve a partial ordering among distributions in the direction of that property. For example, for a functional θ to be a measure of location, its value at a distribution F, $\theta(F)$, should be smaller than or equal to its value at another distribution G, $\theta(G)$, whenever the distribution F is stochastically dominated by the distribution G.

To describe Oja's criteria we first say that a function $J : I \to \mathbb{R}$ is *convex of order* k if $J^{(k)}(x) \geq 0$ for all $x \in I$ where I is an open interval and $J^{(k)}$ is the k-th order derivative of h. Note that

- If J is convex of order 0, then J is a nonnegative function,
- If J is convex of order 1, then J is a nondecreasing function,
- If J is convex of order 2, then J is a convex function.

The relationship between two distributions in terms of distributional shape is defined based on the convexity of the following inverse composition function defined by those distributions.

Definition 1.1(Oja, 1981). Let $\Delta_{F,G} : \mathbb{R} \to \mathbb{R}$ be a function such that $\Delta_{F,G}(x) = G^{-1}(F(x)) - x$ for all $x \in \mathbb{R}$. Then for $k = 0, 1, \cdots$, we write

$$F \lesssim_k G$$

if $\Delta_{F,G}$ is convex of order k for $0, 1, \cdots$. ■

For example, $F \lesssim_0 G$ if

$$G^{-1}(F(x)) - x \geq 0 \quad \forall x \in \mathbb{R}$$

which is equivalent to

$$F(x) \geq G(x) \quad \forall x \in \mathbb{R}.$$

This coincides with the well-known statement that F is *stochastically dominated by* G, which is an important sense in which F does not lie to the right of G. In particular, the function $G^{-1} \circ F$ provides a natural link between the distributions F and G via the Probability Integral Transform. If a random variable X follows the distribution F, then $F(X)$ follows the uniform distribution $\mathcal{U}(0, 1)$ and thus $G^{-1}(F(X))$ follows the distribution G.

(Oja, 1981) imposed a unique meaning to the inequality $\lesssim_k$ so that the relationship between the two distributions F and G can be named in terms of the distributional aspect corresponding to the order k. The functional $p_s : \mathcal{F}_s \to \mathbb{R}$ for a family of absolutely

continuous and symmetric distributions $\mathcal{F}_s$ is called the *symmetry point* if

$$f\left(p_s(F)+x\right)=f\left(p_s(F)-x\right) \tag{1.8}$$

for all $F \in \mathcal{F}_s$ and $x \geq 0$. It was shown in Bickel and Lehmann (1975) that there exists $p_s(F)$ satisfying (1.8) for every absolutely continuous and symmetric distribution.

Definition 1.2(Oja, 1981). We say

a. *F is not to the right of G* if $F \lesssim_0 G$,

b. *F has scale not larger than G* if $F \lesssim_1 G$,

c. *F is not more skew to the right than G* if $F \lesssim_2 G$,

d. *F does not have more kurtosis than G* if F and G are symmetric distributions and $F_s \lesssim_2 G_s$ where $F_s(x)=F(x)-F(-x)$ for all $x \geq 0$. ■

Oja's criteria are given as follows.

Definition 1.3(Oja, 1981). Let F be the distribution of a random variable X and $F_{a,b}$ be the distribution of a random variable $aX+b$. Then the functional $\psi : \mathcal{F} \to \mathbb{R}$ is a

a. *measure of location* in $\mathcal{F}$ if $\psi\left(F_{a,b}\right)=a\psi(F)+b$ for all $a,b \in \mathbb{R}, F \in \mathcal{F}$ and $\psi(F) \leq \psi(G)$ when *F is not to the right of G*.

b. *measure of scale* in $\mathcal{F}$ if $\psi\left(F_{a,b}\right)=|a|\psi(F)$ for all $a,b \in \mathbb{R}, F \in \mathcal{F}$ and $\psi(F) \leq \psi(G)$ when *F has scale not larger than G*.

c. *measure of skewness* in $\mathcal{F}$ if $\psi\left(F_{a,b}\right)=\text{sign}(a)\psi(F)$ for all $a \neq 0, b \in \mathbb{R}, F \in \mathcal{F}$ and $\psi(F) \leq \psi(G)$ when *F is not more skew to the right than G*.

d. *measure of kurtosis* in a family of symmetric distributions $\mathcal{F}_s$ if $\psi\left(F_{a,b}\right)=\psi(F)$ for all $a \neq 0, b \in \mathbb{R}, F \in \mathcal{F}$ and $\psi(F) \leq \psi(G)$ when *F does not have more kurtosis than G*. ■

Oja (1981) showed that the conventional moments based measures of location, scale, skewness and kurtosis satisfy Oja's criteria. As indicated in Section 1.2, the L-moments satisfy

the criterion as well, enabling comparison between those two moments on data sets. The results of comparison are given in Chapters 4 and 5.

CHAPTER 2

Gaussian Centered L-moments

2.1 Motivation

As mentioned in Chapter 1, investigation into distributional shape is often performed relative to the Gaussian distributions. Since observations in many data are aggregations of small independent errors, they often have an approximately Gaussian shape by the Central Limit Theorem. However, in high dimensional data such as the TCGA data shown in Figure 1.1, marginal distributions often have strong departure from Gaussianity. A suitable transformation can be adopted to yield bell shape distributions but this incurs loss of useful information such as skewness and multimodality. This suggests that measures of departure from Gaussianity can better reveal meaningful structure in high dimensional data than transformation methods. The term (*excess*) *kurtosis* (Pearson, 1905) is an example of the importance of measuring the difference between the shapes of a distribution and the Gaussian distribution. Based on the sign of the excess kurtosis, distributions are classified into *platykurtic* and *leptokurtic* distributions if they have positive and negative conventional kurtosis values, respectively.

However, as discussed in Hosking (1990), the L-moments satisfy

$$\lambda_r(\mathcal{U}(a,b)) = \int_0^1 \{(b-a)x + a\} P^*_{r-1}(x)\,\mathrm{d}x = 0 \quad \forall r = 3, 4, \cdots \tag{2.1}$$

where $\mathcal{U}(a,b)$ is the uniform distribution with the lower bound a and the upper bound b and P^*_r is the r-th order shifted Legendre polynomial given in Section 1.1. This implies that the sign and magnitude of the L-moments measure the direction and magnitude of departure from the uniform rather than the Gaussian distributions. For example, a positive value of

the L-kurtosis λ_4^* implies that a distribution has heavier tails than the uniform distributions, which is not useful since the uniform distributions have abnormally light tails.

We introduce a definition.

Definition 2.1. A sequence of functionals $\{\theta_r; r = 1, 2, \cdots\}$ is *centered at the family of distributions* $\mathcal{F}$ when it satisfies $\theta_r(F) = 0$ for all $r = 3, 4, \cdots$ and $F \in \mathcal{F}$. ■

Important functionals centered at the Gaussian distributions are the cumulants $\{\kappa_r; r = 1, 2, \cdots\}$. The Marcinkiewicz theorem (Marcinkiewicz, 1939) showed that the family of Gaussian distributions is the unique center of the cumulants. One of the main goal of this dissertation is to develop different types of moments with their distributional centers at the Gaussian family. For that purpose, we introduce the following definition.

Definition 2.2. Suppose that $\mathcal{F}$ contains the family of Gaussian distributions, i.e. $\Phi(\cdot|\mu, \sigma^2) \in \mathcal{F}$ for all $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. We call functionals $\{\theta_r : \mathcal{F} \to \mathbb{R}; r = 1, 2, \cdots\}$ *Gaussian Centered L-moments* in $\mathcal{F}$ if they are L-functionals and centered at the Gaussian distributions. ■

The letter 'L' was originally for the linear combination of expected order statistics as mentioned in Hosking (1990), but here is generalized to any L-functionals in the form (1.5).

2.2 Hermite L-moments

As pointed out Equation (2.1), the L-moments are centered at the uniform distributions due to orthogonality property of the shifted Legendre polynomials. This motivates us to consider adopting another sequence of orthogonal polynomials to locate the center of new moments at the Gaussian distributions. In particular, the L-functional

$$\theta_r(F) = \int_{-\infty}^{\infty} x f(x) J_{r-1}(F(x)) \, \mathrm{d}x$$

where $J_r : (0, 1) \to \mathbb{R}$ is an r-th order polynomial should satisfy

$$\theta_r(\Phi) = \int_{-\infty}^{\infty} x \phi(x) J_{r-1}(\Phi(x)) \, \mathrm{d}x = 0$$

to be the Gaussian Centered L-moments. This results in one possible solution

$$\eta_r = \int_{-\infty}^{\infty} x f(x) H_{r-1}\left(\Phi^{-1}(F(x))\right)\,\mathrm{d}x = \int_0^1 F^{-1}(u) H_{r-1}\left(\Phi^{-1}(u)\right)\,\mathrm{d}u \tag{2.2}$$

where H_r is the r-th order Hermite polynomial which was introduced in Chapter 5 of (Szegö, 1959). Note that

$$\eta_r(\Phi(\cdot|\mu,\sigma^2)) = \int_{-\infty}^{\infty} (\mu+\sigma x)\phi(x) H_{r-1}(x)\,\mathrm{d}x = 0$$

for all $\mu \in \mathbb{R}, \sigma > 0$ and $r = 3, 4, \cdots$ by orthogonality of the Hermite polynomials. We call $\{\eta_r; r = 1, 2, \cdots\}$ the *Hermite L-moments* (*HL-moments*).

Recall from Definition 1.3 that a measure of skewness or kurtosis should be invariant under linear transformation of a random variable. The need for such invariance motivates us to introduce *Hermite L-moment ratios* (*HL-moment ratios*) defined as $\eta_r^* = \eta_r/\eta_2$ for $r = 3, 4, \cdots$. The *HL-skewness* and *HL-kurtosis* are defined as η_3^* and η_4^* respectively. A central issue is whether or not the HL-skewness and kurtosis actually measure the skewness and kurtosis of a distribution in the sense of Oja's criteria.

Theorem 2.1. The HL-moment based measures η_1, η_2, η_3^* and η_4^* satisfy Oja's criteria for measures of location, scale, skewness and kurtosis respectively.

Proof . See Chapter 7. ■

Both the HL-moments and cumulants $\{\kappa_r | r = 1, 2, \cdots\}$ are centered at the Gaussian distribution, and there is a relation between them. Substitution of $F = \Phi\left(\cdot|\mu,\sigma^2\right)$ in the definition of HL-moments (2.2) yields that the two functionals coincide with each other at the family of Gaussian distributions. At other families, those two functionals are related to each other based on the *Cornish-Fisher expansion* (Cornish and Fisher, 1938). Suppose that the moment generating function of F exists. Then $F^{-1}(u)$ can be approximated for all $0 < u < 1$ by

$$\begin{aligned} F^{-1}(u) &= \mu + \sigma\left\{\Phi^{-1}(u) + \gamma_1 h_1(\Phi^{-1}(u)) + \gamma_2 h_2(\Phi^{-1}(u)) + \gamma_1^2 h_{11}(\Phi^{-1}(u))\right. \\ &\quad \left. + \gamma_3 h_3(\Phi^{-1}(u)) + \gamma_1\gamma_2 h_{12}(\Phi^{-1}(u)) + \gamma_1^3 h_{111}(\Phi^{-1}(u)) + \cdots\right\} \end{aligned} \tag{2.3}$$

where $\gamma_{r-2} = \kappa_r/\kappa_2^{r/2}$ for $r = 3, 4, \cdots$, κ_r is the r-th cumulant and

$$\begin{aligned}
h_1(x) &= \tfrac{1}{6}H_2(x), & h_2(x) &= \tfrac{1}{24}H_3(x), \\
h_{11}(x) &= -\tfrac{1}{36}\left\{2H_3(x) + H_1(x)\right\}, & h_3(x) &= \tfrac{1}{120}H_4(x), \\
h_{12}(x) &= -\tfrac{1}{24}\left\{H_4(x) + H_2(x)\right\}, & h_{111}(x) &= \tfrac{1}{324}\left\{12H_4(x) + 19H_2(x)\right\}.
\end{aligned}$$

By substituting (2.3) for $F^{-1}(u)$ in the definition of the HL-moments (2.2), it can be seen that η_r is a linear combination of powers of the cumulant ratios $\gamma_{r-2} = \kappa_r/\kappa_2^{r/2}$. For example, the third HL-moment can be expressed as

$$\begin{aligned}
\eta_3 &= \int_0^1 F^{-1}(u) H_2\left(\Phi^{-1}(u)\right) \mathrm{d}u \\
&= \sigma\left\{\gamma_1 - \frac{1}{24}\gamma_1\gamma_2 + \frac{19}{324}\gamma_1^3 + \cdots\right\} \\
&= \sigma\left\{\frac{\kappa_3}{\kappa_2^{3/2}} - \frac{1}{24}\frac{\kappa_3}{\kappa_2^{3/2}}\frac{\kappa_4}{\kappa_2^{4/2}} + \frac{19}{324}\left(\frac{\kappa_3}{\kappa_2^{3/2}}\right)^3 + \cdots\right\}.
\end{aligned}$$

A similar expression can be derived for the fourth HL-moment.

2.3 Rescaled L-moments

Additional insights come from another view of why the L-moments are centered at the uniform. Note that

$$\begin{aligned}
\lambda_3 &= \frac{1}{3}\left\{E\left(X_{3:3} - X_{2:3}\right) - E\left(X_{2:3} - X_{1:3}\right)\right\}, \\
\lambda_4 &= \frac{1}{4}\left\{E\left(X_{4:4} - X_{3:4}\right) - 2E\left(X_{3:4} - X_{2:4}\right) + E\left(X_{2:4} - X_{1:4}\right)\right\}.
\end{aligned}$$

These expressions indicate that if F has equally spaced expected order statistics, then its third and higher order L-moments are zero. Figure 2.1 shows the four expected order statistics of $\mathcal{U}(-1, 1)$ and $\mathcal{N}(0, 1)$ as four vertical dashed lines. Note that the vertical lines of the uniform distribution are equally spaced, e.g. if $F \sim$ Uniform$(0, 1)$, then $EX_{i:4} = i/5$. However, for the standard Gaussian distribution, the space between the inner pair of expected order statistics is smaller than the spaces between the two outer pairs.

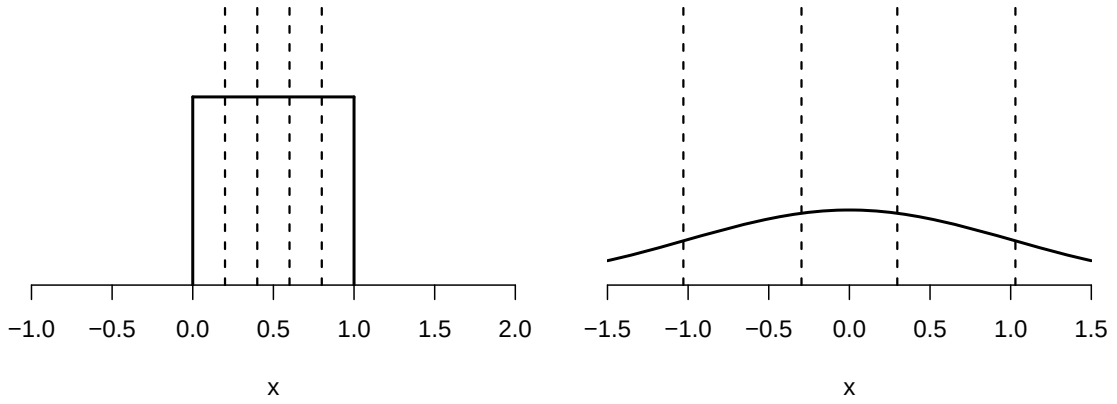

Figure 2.1: The four expected order statistics of the uniform distribution $\mathcal{U}(-1,1)$ (left plot) and the standard Gaussian distribution $\mathcal{N}(0,1)$ (right plot). In each plot, a vertical dashed line indicates the first, second, third and fourth expected order statistics from the leftmost to the rightmost.

This motivates us to rescale the space between adjacent expected order statistics by the corresponding space of the standard Gaussian distribution. The following new theorem shows another definition of the L-moments which re-expresses Equation (2.1) of Hosking (1990) in terms of expected spaces between order statistics.

Theorem 2.2. The r-th L-moments λ_r can be expressed as

$$\lambda_r = \frac{1}{r}\sum_{k=0}^{r-2}(-1)^k \begin{pmatrix} r-2 \\ k \end{pmatrix} E\left(X_{(r-k):r} - X_{(r-k-1):r}\right)$$

for $r = 2, 3, \cdots$.

Proof . See Chapter 7. ■

We first show that when any symmetric distribution is used for rescaling, the resulting measures of location, scale, skewness and kurtosis satisfy Oja's criteria (Definition 1.3). This opens up a broad new family of potential distributional summaries, whose general study could provide interesting future work. Let $\delta_{i,j:k}(F) = E\left(X_{j:k} - X_{i:k}\right)$ for $i < j$ be the expected spacing between the i-th and j-th order statistics. Then we can consider the

rescaled L-moments based on the distribution F_0 defined as

$$\rho_{F_0,r} = \frac{1}{r}\sum_{k=0}^{r-2} \frac{(-1)^k}{\delta_{(r-k-1),(r-k):k}(F_0)} \begin{pmatrix} r-2 \\ k \end{pmatrix} E\left(X_{(r-k):r} - X_{(r-k-1):r}\right) \tag{2.4}$$

for $r = 2, 3, \cdots$. We let $\rho_{F_0,1} = \lambda_1$. The corresponding *rescaled L-moment ratios based on the distribution* F_0 are defined as $\rho^*_{F_0,r} = \rho_{F_0,r}/\rho_{F_0,2}$ for $r = 3, 4, \cdots$.

Theorem 2.3. Suppose that F_0 is a symmetric distribution. Then the rescaled L-moments based measures $\rho_{F_0,1}, \rho_{F_0,2}, \rho^*_{F_0,3}$ and $\rho^*_{F_0,4}$ satisfy Oja's criteria for a measure of location, scale, skewness and kurtosis, respectively.

Proof . See Chapter 7. ■

Based on this theorem, we define the r-th *Rescaled L-moment* (*RL-moment*) as $\rho_r = \rho_{\Phi,r}$. Using Equation (2.4), it can easily be shown that $\rho_r(\Phi(\cdot|\mu,\sigma^2)) = 0$ for all $\mu \in \mathbb{R}$, $\sigma^2 > 0$ and $r = 3, 4, \cdots$.

It is hard to obtain exact coefficients of R_r for general order r, but the first four polynomials can be obtained using the results in Hosking (1986) as

$$\begin{aligned} &R_0(u) = P^*_0(u), \qquad R_1(u) = c_1 P^*_1(u), \\ &R_2(u) = c_2 P^*_2(u), \quad R_3(u) = (6c_3+2)u^3 - 3(3c_3+1)u^2 + (3c_3+3)u - 1. \end{aligned} \tag{2.5}$$

where $c_1 \approx 0.8862$, $c_2 \approx 1.1816$ and $c_3 \approx 3.4658$. Based on these equations, it can be seen that the first three terms the RL-moments and L-moments coincide with each other up to a constant multiple while possibly from the fifth order they deviate from each other. This observation has an impact on the analysis of Chapters 4 and 5.

CHAPTER 3
Robustness

Since one of the main reasons for developing the Gaussian Centered L-moments is their robustness, we analyze that first. We use the *influence function* as a primary tool for robustness analysis. Note from Huber and Ronchetti (2009) that the influence function of a functional θ evaluated at a distribution F is defined as the functional derivative

$$\mathrm{IF}(x; F, \theta) = \lim_{\epsilon \downarrow 0} \frac{\theta\left(F_{\epsilon,x}\right) - \theta(F)}{\epsilon} \tag{3.1}$$

based on the x point mass contaminated version of F, $F_{\epsilon,x} = (1-\epsilon)F + \epsilon\delta_x$, where δ_x is a degenerate distribution putting mass 1 at the point x. Hence, if a functional is sensitive to an outlier, its influence function should have large absolute values at extreme values of x. If the distribution F changes, the same outlier x can affect the functional in a different way. In this dissertation, we compare the robustness of various measures of skewness and kurtosis based on their influence functions evaluated at a family of distributions.

The papers Groeneveld (1991) and Ruppert (1987) compared various measures of skewness and kurtosis, respectively, using the influence function. As a criterion of comparison, both papers used the degree of polynomials that are *asymptotic tight bounds* of the influence functions. Suppose that $J_1, J_2 : \mathbb{R} \to \mathbb{R}_+$ are two functions where $\mathbb{R}_+ = \{x \in \mathbb{R} | x \geq 0\}$. In this dissertation, we indicate that the asymptotic behavior of J_1 and J_2 are the same by

$$J_1(x) = \Theta\left(J_2(x)\right)$$

to mean that there exist $a_1, a_2 > 0$ and $x' > 0$ such that $a_1 J_2(x) \leq J_1(x) \leq a_2 J_2(x)$ for all $|x| \geq x'$. Using such asymptotic tight bounds, those papers compared the robustness of different measures. For example, if two functionals θ_1 and θ_2 satisfy $|\mathrm{IF}\left(x; F, \theta_1\right)| = \Theta(|x|)$

and $|\mathrm{IF}(x; F, \theta_2)| = \Theta\left(x^2\right)$, then θ_1 was considered to be more robust than θ_2 for the distribution F.

An interesting family of distributions for evaluation of the influence functions is *Tukey's g and h distributions*, see Jorge and Boris (1984) for good discussion, which contain all the transformed normal random variables of the form

$$\left(\frac{e^{gZ}-1}{g}\right)\exp\left[\frac{hZ^2}{2}\right]$$

where $g \in \mathbb{R}, h \geq 0$ and Z is the standard Gaussian random variable. By convention, the case when $g = 0$ is defined using the limit $g \to 0$ as the random variable $Z \exp\left(hZ^2/2\right)$. We denote the distribution function of Tukey(g, h) by $\mathcal{T}^{g,h}$. An important special case, where the distributions are symmetric, $\mathcal{T}^{0,h}$ is called *Tukey's h distributions*. Tukey's g and h family is ideal for our study because it allows direct application of Oja's criteria as seen below.

Theorem 3.1. If $g > 0$ and $h = 0$, then Φ is not more skew to the right than $T^{g,0}$. On the contrary, if $g < 0$ and $h = 0$, then $T^{g,0}$ is not more skew to the right than Φ. If we have $g = 0$ and $h > 0$, then, Φ is not more kurtotic than $T^{0,h}$.

Proof . See Chapter 7. ■

A particular case of interest is when $g = 0, h > 0$ since this case corresponds to distributions with heavier tails than the Gaussian distributions. Note that a heavier tail of a distribution indicates a higher chance of existence of extreme outliers. This suggests checking the influence functions not only for the standard Gaussian distribution but also for Tukey's h distributions with $h > 0$.

Before we derive the influence functions of measures of skewness and kurtosis based on the Gaussian Centered L-moments, we introduce the previous results for the conventional skewness and kurtosis. Note that Ruppert (1987) used the notion of *symmetric influence function* defined as

$$\mathrm{SIF}(x; F, \theta) = \lim_{\epsilon \downarrow 0} \frac{\theta\left(\left(F_{\epsilon,x} + F_{\epsilon,-x}\right)/2\right) - \theta(F)}{\epsilon}.$$

The symmetric influence function measures the sensitivity of a functional to symmetric contamination by points $-x$ and x so it is better suited for comparison of kurtosis measures.

Theorem 3.2 (Groeneveld, 1991), (Ruppert, 1987). Suppose that F is a symmetric distribution such that $\mu(F) = 0$ and $\sigma^2(F) = 1$. Then we have

$$\text{IF}\,(x; F, \gamma_1) = x^3 - 3x = H_3(x),\ \text{SIF}\,(x; F, \gamma_2) = x^4 - 6x^2 + 3 = H_4(x). \qquad \blacksquare$$

Before we derive the influence functions of the measures of skewness and kurtosis based on the HL- and RL-moments at various distributions, we show the relation between their influence functions and symmetric influence functions.

Theorem 3.3. If F is a symmetric distribution, we have

$$\text{SIF}\,(x; \Phi, \eta_4^*) = \text{IF}\,(x; \Phi, \eta_4^*)\,,\ \text{SIF}\,(x; \Phi, \rho_4^*) = \text{IF}\,(x; \Phi, \rho_4^*)\,.$$

Proof . See Chapter 7. $\blacksquare$

The following theorem shows the influence functions evaluated at the standard Gaussian distribution.

Theorem 3.4. We have

$$\begin{aligned} \text{IF}\,(x; \Phi, \eta_r^*) &= \frac{1}{r} H_r(x), \\ \text{IF}\,(x; \Phi, \rho_r^*) &= l_{\Phi, R_{r-1}} + \int_0^x R_{r-1}(\Phi(y))\,\mathrm{d}y \end{aligned}$$

for $x \in \mathbb{R}$ and $r = 3, 4, \cdots$ where

$$l_{F,J} = \int_{-\infty}^{0} F(y) J(F(y))\,\mathrm{d}y - \int_0^{\infty} \{1 - F(y)\} J(F(y))\,\mathrm{d}y$$

for a distribution $F \in \mathcal{F}$ and measurable function $J : (0, 1) \to \mathbb{R}$.

Proof . See Chapter 7. $\blacksquare$

Note that we have

$$\mathrm{IF}\left(x;\Phi,\eta_3^*\right)=\frac{1}{3}\mathrm{IF}\left(x;\Phi,\gamma_1\right),\ \mathrm{SIF}\left(x;\Phi,\eta_4^*\right)=\frac{1}{4}\mathrm{SIF}\left(x;\Phi,\gamma_2\right). \tag{3.2}$$

The influence function can be understood in some sense as description of local behavior of a functional since it is the directional derivative of a functional with respect to contamination of a distribution by a single point. Equation (3.2) implies that the HL- and conventional skewness share the same local behavior up to a constant multiple, and the same holds between the HL- and conventional kurtosis. This observation coincides with the observation made in Section 2.2 where the HL- and conventional moments are related to each other by the Cornish-Fisher expansion.

Note that Theorem 3.5 does not show distinction between robustness of measures based on the conventional and HL-moments. It can be seen from Theorems 3.2, 3.3 and 3.4 that we have

$$\begin{aligned}
&|\mathrm{IF}(x;\Phi,\gamma_1)|=\Theta\left(|x|^3\right), \quad |\mathrm{IF}(x;\Phi,\eta_3^*)|=\Theta\left(|x|^3\right),\\
&|\mathrm{SIF}(x;\Phi,\gamma_2)|=\Theta\left(|x|^4\right), \quad |\mathrm{SIF}(x;\Phi,\eta_4^*)|=\Theta\left(|x|^4\right).
\end{aligned}$$

As noted above, we adopt distributions with heavier tails as other bases on which the influence functions are compared.

Theorem 3.5. We have

$$\begin{aligned}
|\mathrm{IF}(x;F,\rho_r^*)| &= \Theta(|x|),\\
\left|\mathrm{IF}\left(x;T^{0,h},\eta_r^*\right)\right| &= \Theta\left(|x|\left\{\log(|x|+1)\right\}^{(r-1)/2}\right)
\end{aligned}$$

for all $F\in\mathcal{F}$ and $h>0$.

Proof . See Chapter 7. ■

It can simply be checked from Ruppert (1987) and Groeneveld (1991) that even though the distribution F does not satisfy $\sigma(F)=1$, the influence functions of γ_1 and γ_2 satisfy $|\mathrm{IF}\left(x;F,\gamma_1\right)|=\Theta\left(|x|^3\right)$ and $|\mathrm{SIF}\left(x;F,\gamma_2\right)|=\Theta\left(|x|^4\right)$, respectively. Theorem 3.5 implies that the new measures are much more robust than the conventional skewness and kurtosis on Tukey's h distributions. The RL-moment based measures, ρ_r^*, are somewhat more robust

than the the HL-moment based moments. Note that the influence function of η_r^* does not depend on the parameter h. This indicates that even slightly heavier tails of distributions than the standard Gaussian distribution can result in better robustness of the HL-moment based measures than the conventional moment based measures.

CHAPTER 4

Goodness-of-fit test for Gaussianity

4.1 Estimation of the Gaussian Centered L-moments

One of main strengths of the Gaussian Centered L-moments is their interpretability; their signs and absolute values indicate direction and magnitude of departure from the Gaussian distributions, respectively. Consistency, which shows that estimators converge to the true underlying functionals in the limit as the sample size goes to infinity, is a reassuring property. In addition, their asymptotic distributions are useful in hypothesis testing and general statistical inferences. Based on the relationship between an L-statistic (1.4) and L-functional (1.5), L-statistics based estimators of the r-th HL- and RL-moments are naturally derived as

$$\begin{aligned} \tilde{\eta}_{n,r} &= \frac{1}{n}\sum_{i=1}^{n} H_{r-1}\left(\Phi^{-1}\left(\frac{i}{n+1}\right)\right) X_{i:n}, \\ \tilde{\rho}_{n,r} &= \frac{1}{n}\sum_{i=1}^{n} R_{r-1}\left(\frac{i}{n+1}\right) X_{i:n}. \end{aligned} \tag{4.1}$$

As mentioned in Section 1.2, there are multiple ways to estimate L-functionals by L-statistics. We illustrate potential improvements using the HL-moments. Motivation comes from the following approximation

$$\begin{aligned} \eta_r = E\left(H_{r-1}\left(\Phi^{-1}(F(X))\right)X\right) &\approx \frac{1}{n}\sum_{i=1}^{n} H_{r-1}\left(\Phi^{-1}\left(F\left(X_i\right)\right)\right)X_i \\ &= \frac{1}{n}\sum_{i=1}^{n} H_{r-1}\left(\Phi^{-1}\left(F\left(X_{i:n}\right)\right)\right)X_{i:n} \end{aligned}$$

where the approximation can be replaced by the almost sure convergence when suitable assumptions are made on the distribution F. For the last expression to actually play the role of an estimator, the terms including F should be estimated. The L-statistics $\tilde{\eta}_{n,r}$ given

in Equation (4.1) originate from the following approximation

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} H_{r-1}\left(\Phi^{-1}\left(\underline{F\left(X_{i:n}\right)}\right)\right) X_{i:n} &\approx \frac{1}{n}\sum_{i=1}^{n} H_{r-1}\left(\Phi^{-1}\left(\underline{E\left(F\left(X_{i:n}\right)\right)}\right)\right) X_{i:n} \\ &= \frac{1}{n}\sum_{i=1}^{n} H_{r-1}\left(\Phi^{-1}\left(E\left(U_{i:n}\right)\right)\right) X_{i:n} \\ &= \frac{1}{n}\sum_{i=1}^{n} H_{r-1}\left(\Phi^{-1}\left(\frac{i}{n+1}\right)\right) X_{i:n} = \tilde{\eta}^{*}_{n,r} \end{aligned} \tag{4.2}$$

where the underlined expressions present approximated and approximating terms and $U_{i:n}$ is the i-th uniform order statistic. This implies that commonly used L-statistics base their performance on how well the coefficient $i/(n+1)$ approximates $F\left(X_{i:n}\right)$.

Another estimator can be obtained from a different approximation in Equation (4.2) as

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} \underline{H_{r-1}\left(\Phi^{-1}\left(F\left(X_{i:n}\right)\right)\right)} X_{i:n} &\approx \frac{1}{n}\sum_{i=1}^{n} \underline{E\left(H_{r-1}\left(\Phi^{-1}\left(F\left(X_{i:n}\right)\right)\right)\right)} X_{i:n} \\ &= \frac{1}{n}\sum_{i=1}^{n} E\left(H_{r-1}\left(Z_{i:n}\right)\right) X_{i:n} \end{aligned} \tag{4.3}$$

where $Z_{i:n}$ is the i-th standard Gaussian order statistic. The key idea is that careful choice of location of the expectation can increase accuracy of approximation of an L-functional by an L-statistic. Since the quantile function Φ^{-1} is a highly nonlinear function, taking expectation outside Φ^{-1} can yield better approximation in Equation (4.3). The *sample Hermite L-moments* (*sample HL-moments*) are defined as

$$\hat{\eta}_{n,r} = \frac{1}{n}\sum_{i=1}^{n} E\left(H_{r-1}\left(Z_{i:n}\right)\right) X_{i:n}. \tag{4.4}$$

This can be understood as the inner product between the order statistics $X_{i:n}$ and polynomials of the expected order statistics of the standard Gaussian distribution. By changing the degree of the polynomial, r, different distributional aspects of F are compared with the standard Gaussian distribution Φ. For example, the third and fourth sample HL-moments

are

$$\hat{\eta}_{n,3} = \frac{1}{n}\sum_{i=1}^{n}\left\{E\left(Z_{i:n}^2\right) - 1\right\} X_{i:n},$$
$$\hat{\eta}_{n,4} = \frac{1}{n}\sum_{i=1}^{n}\left\{E\left(Z_{i:n}^3\right) - 3E\left(Z_{i:n}\right)\right\} X_{i:n}.$$

We have the following theorem on asymptotic Gaussianity of the sample Gaussian Centered L-moments based on Shorack (1972) and Li et al. (2001).

Theorem 4.1. Let $r_1, r_2 = 3, 4, \cdots$ such that $r_1 \neq r_2$. If $E\left|X_1\right|^2 < \infty$, then we have

$$n^{1/2}\left(\begin{pmatrix} \hat{\rho}_{n,r_1}^* \\ \hat{\rho}_{n,r_2}^* \end{pmatrix} - \begin{pmatrix} \rho_{r_1}^* \\ \rho_{r_2}^* \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(0, \Psi^R\right)$$

where

$$\Psi_{i,j}^R = \left(\sigma_{r_i r_j}^R - \rho_{r_i}^*\sigma_{2r_i}^R - \rho_{r_j}^*\sigma_{2r_j}^R + \rho_{r_i}^*\rho_{r_j}^*\sigma_{22}^R\right)/\rho_2^2,$$
$$\sigma_{k_1k_2}^R = \int_0^1\int_0^1 (u \wedge v - uv) R_{k_1-1}(u) R_{k_2-1}(v)\,\mathrm{d}F^{-1}(u)\,\mathrm{d}F^{-1}(v).$$

for $i, j \in \{1, 2\}$ and for $k_1, k_2 \in \{2, r_1, r_2\}$. If we further have $E\left|X_1\right|^{2+\epsilon} < \infty$ for some $\epsilon > 0$, then we have

$$n^{1/2}\left(\begin{pmatrix} \hat{\eta}_{n,r_1}^* \\ \hat{\eta}_{n,r_2}^* \end{pmatrix} - \begin{pmatrix} \eta_{r_1}^* \\ \eta_{r_2}^* \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(0, \Psi^H\right)$$

where

$$\Psi_{i,j}^H = \left(\sigma_{r_i r_j}^H - \eta_{r_i}^*\sigma_{2r_i}^H - \eta_{r_j}^*\sigma_{2r_j}^H + \eta_{r_i}^*\eta_{r_j}^*\sigma_{22}^H\right)/\eta_2^2,$$
$$\sigma_{k_1k_2}^H = \int_0^1\int_0^1 (u \wedge v - uv) H_{k_1-1}\left(\Phi^{-1}(u)\right) H_{k_2-1}\left(\Phi^{-1}(v)\right)\,\mathrm{d}F^{-1}(u)\,\mathrm{d}F^{-1}(v).$$

for $i, j \in \{1, 2\}$ and $k_1, k_2 \in \{2, r_1, r_2\}$.

Proof . See Chapter 7. ■

4.2 Goodness-of-fit test for Gaussianity

As seen in Chapter 3, the Gaussian Centered L-moments exhibit better robustness than the conventional moments. Since robustness often comes at a price of loss of efficiency, we compare the conventional and Gaussian Centered L-moments in those terms. To this end, we consider the goodness-of-fit test for Gaussianity defined by the null and alternative hypotheses

$$H_0 : F = \Phi\left(\cdot|\mu, \sigma^2\right) \text{ for some } \mu \in \mathbb{R}, \sigma^2 > 0, \quad H_1 : \text{Not } H_0. \tag{4.5}$$

Various test statistics for these hypotheses were introduced and compared in terms of their powers under various alternative hypothetical distributions in Romão et al. (2010).

The *Kolmogorov-Smirnov test* (*KS*) is one of the most popular goodness-of-fit tests in the statistics literature. The KS test statistic measures the maximum departure of the probability integral transform of the distribution of data from the uniform distribution. Given the sample mean $\bar{X}$ and the sample variance S_X^2 of the random sample $\{X_1, X_2, \cdots, X_n\}$, the KS test statistic is

$$D^{\mathrm{KS}} = \max_{1 \le i \le n} \max \left\{ \Phi(X_{i:n}|\bar{X}, S^2) - \frac{i-1}{n}, \frac{i}{n} - \Phi(X_{i:n}|\bar{X}, S^2) \right\}.$$

The Gaussianity null hypothesis is rejected for large values of D^{KS}. Since the power of this test in important directions has been questioned by several studies including (Gan and Koehler, 1990), this method is not included in our analysis.

Rather than applying the probability integral transform to data, Anderson and Darling (1954) suggested comparing the *empirical distribution function* (*EDF*) with the standard Gaussian distribution based on L_2 distance. The *Anderson-Darling* (*AD*) test statistic is given as

$$D^{\mathrm{AD}} = n \int_{-\infty}^{\infty} \{F_n(x) - \Phi(x)\}^2 \, w(x) \, \mathrm{d}F(x)$$

where F_n is the EDF given by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1(Y_i \leq x) \tag{4.6}$$

where $Y_i = (X_i - \bar{X})/S_X$ and w is a weight function given by

$$w(x) = \frac{1}{\Phi(x)\left\{1 - \Phi(x)\right\}}.$$

The null hypothesis of (4.5) is rejected for large values of the test statistics. It was shown in (Anderson and Darling, 1954) that D^{AD} can be rewritten as

$$D^{\mathrm{AD}} = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left\{\ln\left(p_i\right) + \ln\left(1 - p_{n+1-i}\right)\right\}$$

where the $p_i = \Phi(Y_{i:n})$. It can be seen from this equation that if there is a strong outlier in data, then that data point will significantly increase the value of $\log\left(p_i\right) + \log\left(1 - p_{n+1-i}\right)$ resulting in rejection.

The *Shapiro-Wilk* (SW) *test* (Shapiro and Wilk, 1965) is a well-established test for Gaussianity which is based on the analysis of variance of a Q-Q plot. In such a plot, one can consider regression of order statistics of a random sample on the expected order statistics of a hypothesized distribution, and that is the Gaussian distribution in the Shapiro-Wilk test. Application of analysis of variance to the regression yields the ratio of the squared slope of the regression line to the residual mean square about the regression line. The test statistic is formally defined as

$$D^{\mathrm{SW}} = \frac{\left(\sum_{i=1}^{n} a_i X_{i:n}\right)^2}{\sum_{i=1}^{n}\left(X_{i:n} - \bar{X}\right)^2}$$

where the weight vector a is obtained by

$$(a_1, \cdots, a_n) = mV^{-1}\left(mV^{-1}V^{-1}m^T\right)^{-1/2}$$

in which $m = (EZ_{1:n}, EZ_{2:n}, \cdots, EZ_{n:n})^T$ is the mean vector of standard Gaussian order statistics and V is their covariance matrix such that $V_{i,j} = \mathrm{Cov}\,(Z_{i:n}, Z_{j:n})$. It was established in (Shapiro and Wilk, 1965) that the Gaussian null hypothesis should be rejected for small values of D^{SW}.

Joint use of conventional skewness and kurtosis has been investigated in the literature as another method for measuring non-Gaussianity. D'Agostino and Pearson (1973) first suggested an omnibus test statistic in the form of a weighted average of transformed sample skewness and kurtosis into approximately Gaussian random variables

$$\delta_{n,1} \sinh^{-1}\left(\frac{\hat{\gamma}_1}{\delta_{n,2}}\right) + \delta_{n,3} \sinh^{-1}\left(\frac{\hat{\gamma}_2}{\delta_{n,4}}\right) \tag{4.7}$$

where $\delta_{n,1}$ and $\delta_{n,2}$ are the functions of conventional variance and kurtosis of the distribution of the conventional sample skewness $\hat{\gamma}_{n,1}$, and $\delta_{n,3}$ and $\delta_{n,4}$ are functions of those of conventional sample kurtosis $\hat{\gamma}_{n,2}$. The reasoning behind using the hyperbolic sine function sinh is that the distribution of $\hat{\gamma}_{n,1}$ is closely approximated by Johnson's symmetric S_U distributions (p.22, Johnson and Kotz (1970)) whose natural link with the standard Gaussian distribution is given as the inverse hyperbolic sine function. Later, Bowman and Shenton (1975) derived the distribution of the statistic (4.7) under the Gaussianity assumption by simulation and claimed that its finite sample distribution is far from its asymptotic distribution which is the chi squared distribution with 2 degrees of freedom χ^2_2.

Jarque and Bera (1980) showed that instead of searching for suitable transformations of the conventional sample skewness and kurtosis, we can adopt their asymptotic means and variances but still obtain good properties like an asymptotic χ^2_2 distribution and being asymptotically locally most powerful. Their test is the score test (Chapter 9, Cox and Hinkley (1974)) on the Pearson family of distributions (p.148, Kendall and Stuart (1977)) whose densities f satisfy

$$\frac{\mathrm{d}f(u)}{\mathrm{d}u} = \frac{(c_1 - u)\, f(u)}{(c_0 - c_1 u + c_2 u^2)}$$

where $u \in \mathbb{R}$, for the null hypothesis $H_0 : c_1 = c_2 = 0$. This results in the *Jarque-Bera test statistic*

$$D^{\mathrm{JB}} = n\frac{\hat{\gamma}_{n,1}^2}{6} + n\frac{(\hat{\gamma}_{n,2} - 3)^2}{24}. \tag{4.8}$$

Jarque and Bera (1980) pointed out that this statistic can be understood from asymptotic joint Gaussianity of the sample conventional skewness and kurtosis

$$n^{1/2}\left(\begin{pmatrix} \hat{\gamma}_{n,1} \\ \hat{\gamma}_{n,2} \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} 6 & 0 \\ 0 & 24 \end{pmatrix}\right) \tag{4.9}$$

which can also be derived from Theorem 2.2.3.B of Serfling (1980). Note from this equation that the sample skewness and kurtosis are asymptotically independent of each other. This motivates investigation of joint distributions of the sample Gaussian Centered L-moments. A similar but stronger fact can be proven for the HL-moments, which is that all of them are asymptotically independent of each other. This is done by showing the off-diagonal terms of Ψ^H in Theorem 4.1 are zero at the Gaussian distributions.

Theorem 4.2. Suppose that $X_1, X_2, \cdots, X_n \sim \mathcal{N}(0,1)$. Then for the sample HL-moments, we have $\lim_{n\to\infty} \mathrm{Cov}\left(n^{1/2}\tilde{\eta}_{r_1}^*, n^{1/2}\tilde{\eta}_{r_2}^*\right) = \Psi_{r_1 r_2}^H = 0$ for all $r_1, r_2 = 3, 4, \cdots$ such that $r_1 \neq r_2$. For the sample RL-moments, we have $\lim_{n\to\infty} \mathrm{Cov}\left(n^{1/2}\tilde{\rho}_3^*, n^{1/2}\tilde{\rho}_4^*\right) = \Psi_{34}^R = 0$.

Proof . See Chapter 7. ■

Theorem 4.2 shows that for Gaussian data, the sample HL-moments are asymptotically uncorrelated with each other. This is a substantial improvement over the sample conventional moments. For example, the fourth and sixth sample moments are not independent of each other at a Gaussian distribution, which can be checked by Monte-Carlo simulation. Similarly, numerical calculation yields that the fourth and sixth sample RL-moments are not independent of each other.

Based on the asymptotic Gaussianity shown in Section 4.1, we can define Gaussian Centered L-moments based goodness-of-fit test statistics for Gaussianity. The *L-moments*

based goodness-of-fit test statistic for Gaussianity was first defined in Henderson (2006) as

$$D^{\mathrm{L}} = n\frac{\left(\hat{\lambda}_3^* - E_\Phi\left(\hat{\lambda}_3^*\right)\right)^2}{\mathrm{Var}_\Phi\left(\hat{\lambda}_3^*\right)} + n\frac{\left(\hat{\lambda}_4^* - E_\Phi\left(\hat{\lambda}_4^*\right)\right)^2}{\mathrm{Var}_\Phi\left(\hat{\lambda}_4^*\right)} \tag{4.10}$$

where the sample L-skewness $\hat{\lambda}_3^*$ and sample L-kurtosis $\hat{\lambda}_4^*$ were defined in Section 1.2 and their means and variances are numerically computed by Monte Carlo simulation with 10,000 replications at the standard Gaussian distribution. The reason that Henderson (2006) used Monte Carlo simulation is that the closed form expressions of the variances $\mathrm{Var}_\Phi\left(\hat{\lambda}_3^*\right)$ and $\mathrm{Var}_\Phi\left(\hat{\lambda}_4^*\right)$ do not have known closed forms. In a similar manner with the L-moments, the *HL- and RL-moments based goodness-of-fit test statistics for Gaussianity* are defined as

$$\begin{aligned} D^{\mathrm{HL}} &= n\frac{(\hat{\eta}_{n,3}^* - E_\Phi(\hat{\eta}_{n,3}^*))^2}{\mathrm{Var}_\Phi\left(\hat{\eta}_{n,3}^*\right)} + n\frac{\left(\hat{\eta}_{n,4}^* - E_\Phi(\hat{\eta}_{n,4}^*)\right)^2}{\mathrm{Var}_\Phi\left(\hat{\eta}_{n,4}^*\right)}, \\ D^{\mathrm{RL}} &= n\frac{\left(\hat{\rho}_{n,3}^* - E_\Phi(\hat{\rho}_{n,3}^*)\right)^2}{\mathrm{Var}_\Phi\left(\hat{\rho}_{n,3}^*\right)} + n\frac{\left(\hat{\rho}_{n,4}^* - E_\Phi(\hat{\rho}_{n,4}^*)\right)^2}{\mathrm{Var}_\Phi\left(\hat{\rho}_{n,4}^*\right)}. \end{aligned} \tag{4.11}$$

Here the reason that we use the approximate values for the means and variances is that the estimators $\hat{\eta}_{n,r}$ and $\hat{\rho}_{n,r}$ are biased. For example, we have $E_\Phi(\hat{\eta}_{20,4}) \approx 0.2833$ and $E_\Phi(\hat{\eta}_{50,4}) \approx -0.1733$ which can significantly affect data analysis. Explicit correction for the biases remains as future work. By simulation, we checked that using the finite sample means and variances for the conventional skewness and kurtosis did not bring us any benefit, we stick to the original Jarque-Bera test statistic given in Equation (4.8).

4.3 Optimal balance between skewness and kurtosis estimators

There are two assumptions that underlie the test statistics in Equations (4.8) and (4.11). To illustrate, we restrict our attention to the HL-moments based test statistic D^{HL}. First, equal weights are given to the skewness estimator $\hat{\eta}_{n,3}^*$ and the kurtosis estimator $\hat{\eta}_{n,4}^*$ no matter what distribution the random sample $X_1, X_2, \cdots, X_n$ follows. If the distribution of that random sample, F, is a symmetric distribution, then there is no signal in the skewness estimator. Secondly, the covariance between $\hat{\eta}_{n,3}^*$ and $\hat{\eta}_{n,4}^*$ is not taken into account by the

test statistic. It was noted in Equation (2.7) of Hosking (1990) that the L-skewness λ_3^* and L-kurtosis λ_4^* are related to each other by

$$\frac{1}{4}\left(5\left(\lambda_3^*\right)^2-1\right) \leq \lambda_4^* < 1.$$

This suggests that the sample L-skewness and L-kurtosis may have some correlation with each other. To improve the power of a test statistic, we investigate an adaptive approach that incorporates the covariance between those skewness and kurtosis estimators with the goal of resolving the two limitations given above.

To begin with, consider a convex combination of the sample HL-skewness and kurtosis

$$D^{\mathrm{HL},\alpha} = \alpha\left\{n\frac{(\hat{\eta}_{n,3}^* - E_\Phi(\hat{\eta}_{n,3}^*))^2}{\mathrm{Var}_\Phi\left(\hat{\eta}_{n,3}^*\right)}\right\} + (1-\alpha)\left\{n\frac{\left(\hat{\eta}_{n,4}^* - E_\Phi(\hat{\eta}_{n,4}^*)\right)^2}{\mathrm{Var}_\Phi\left(\hat{\eta}_{n,4}^*\right)}\right\} \tag{4.12}$$

where $0 \leq \alpha \leq 1$. If the alternative distribution is not more skew to the left or right than the null Gaussian distribution but more kurtotic than that distribution, the test statistic that purely depends on the sample HL-kurtosis, $D^{\mathrm{HL},0}$, will yield a more powerful test. For example, $D^{\mathrm{HL},0}$ will have higher power than the equal weight test statistic $D^{\mathrm{HL},0.5}$ when the alternative distribution is Tukey's h distribution with $h > 0$. Recall from Theorem 3.1 that the Gaussian distributions are not more kurtotic than Tukey's h distributions.

The test statistic (4.12) can be more intuitively understood as the following Mahalanobis distance,

$$\begin{aligned} D^{\mathrm{HL}} = &\left(n^{1/2}\hat{\eta}_{n,3} - n^{1/2}E_\Phi\left(\hat{\eta}_{n,3}\right), n^{1/2}\hat{\eta}_{n,4} - n^{1/2}E_\Phi\left(\hat{\eta}_{n,4}\right)\right) \\ &\times \begin{pmatrix} \mathrm{Var}_\Phi\left(\hat{\eta}_{n,3}^*\right)/\alpha & 0 \\ 0 & \mathrm{Var}_\Phi\left(\hat{\eta}_{n,4}^*\right)/(1-\alpha) \end{pmatrix}^{-1} \begin{pmatrix} n^{1/2}\hat{\eta}_{n,3} - n^{1/2}E_\Phi\left(\hat{\eta}_{n,3}\right) \\ n^{1/2}\hat{\eta}_{n,4} - n^{1/2}E_\Phi\left(\hat{\eta}_{n,4}\right) \end{pmatrix}. \end{aligned} \tag{4.13}$$

This motivates us to focus on the family of projections

$$w_1\left(n^{1/2}\hat{\eta}_{n,3}^* - n^{1/2}E_\Phi\left(\hat{\eta}_{n,3}^*\right)\right) + w_2\left(n^{1/2}\hat{\eta}_{n,4}^* - n^{1/2}E_\Phi\left(\hat{\eta}_{n,4}^*\right)\right)$$

indexed by $\boldsymbol{w} = (w_1, w_2)^T \in \mathbb{R}^2$ subject to $\|\boldsymbol{w}\| = 1$. Let the joint distribution of $\left(n^{1/2}\hat{\eta}^*_{n,3}, n^{1/2}\hat{\eta}^*_{n,4}\right)^T$ under the null Gaussian hypothesis be F_0 and under the alternative distribution F be F_1. Then choosing the weight vector $\boldsymbol{w}$ is equivalent to finding a direction in which separation between the bivariate distributions F_0 and F_1 generated by Φ and F, respectively, becomes maximal in $\mathbb{R}^2$. A simple approach to this is *Fisher's linear discriminant* (*FLD*) introduced in Fisher (1936).

The FLD aims at a projection direction in which two distributions are "best" separated from each other. Suppose that $\boldsymbol{\mu}_{F_0}$ and $\boldsymbol{\mu}_{F_1}$ are the means of the distributions F_0 and F_1, respectively, and Ψ_{F_0} and Ψ_{F_1} are their covariance matrices, respectively. Then the FLD seeks the projection direction $\boldsymbol{w} \in \mathbb{R}^2$ which maximizes the gain function

$$\mathcal{G}(\boldsymbol{w}) = \frac{\boldsymbol{w}^T \Psi^B \boldsymbol{w}}{\boldsymbol{w}^T \Psi^W \boldsymbol{w}}$$

subject to $\|\boldsymbol{w}\| = 1$ where Ψ^B and Ψ^W are the between-class and within-class covariance matrices defined as

$$\Psi^B = \left(\boldsymbol{\mu}_{F_1} - \boldsymbol{\mu}_{F_0}\right)\left(\boldsymbol{\mu}_{F_1} - \boldsymbol{\mu}_{F_0}\right)^T, \quad \Psi^W = \Psi_{F_0} + \Psi_{F_1},$$

respectively. The solution vector $\boldsymbol{w}^{\mathrm{FLD}}$ maximizing the gain $\mathcal{G}$ is given as

$$\boldsymbol{w}^{\mathrm{FLD}} = \boldsymbol{w}_{F_0,F_1} / \left\|\boldsymbol{w}_{F_0,F_1}\right\|$$

where

$$\boldsymbol{w}_{F_0,F_1} = \left(\Psi^W\right)^{-1}\left(\boldsymbol{\mu}_{F_1} - \boldsymbol{\mu}_{F_0}\right) = \left(\Psi_{F_0} + \Psi_{F_1}\right)^{-1}\left(\boldsymbol{\mu}_{F_1} - \boldsymbol{\mu}_{F_0}\right).$$

For applications, $\boldsymbol{\mu}_{F_0}, \boldsymbol{\mu}_{F_1}$ and Ψ_{F_0}, Ψ_{F_1} are usually substituted for by suitable estimators. For a broad range of underlying distributions of the random sample $X_1, X_2, \cdots, X_n$, their sample HL-skewness and kurtosis have an asymptotically joint Gaussian distribution. This enables us to expect somewhat stable performance of the FLD.

Based on the FLD, the optimal weights are given by

$$\boldsymbol{w}^{\mathrm{FHL}} = \boldsymbol{w}^{\mathrm{HL}} / \left\| \boldsymbol{w}^{\mathrm{HL}} \right\| \text{ where } \boldsymbol{w}^{\mathrm{HL}} = \left(\Psi_{F_0} + \Psi_{F_1} \right)^{-1} \left(\boldsymbol{\mu}_{F_1} - \boldsymbol{\mu}_{F_0} \right) \tag{4.14}$$

where

$$\begin{aligned}
\boldsymbol{\mu}_{F_0} &= \begin{pmatrix} n^{1/2} E_{\Phi}(\hat{\eta}^*_3) \\ n^{1/2} E_{\Phi}(\hat{\eta}^*_4) \end{pmatrix}, \quad \boldsymbol{\mu}_{F_1} = \begin{pmatrix} n^{1/2} E_{F}(\hat{\eta}^*_3) \\ n^{1/2} E_{F}(\hat{\eta}^*_4) \end{pmatrix}, \\
\Psi_{F_0} &= \begin{pmatrix} \mathrm{Var}_{\Phi}\left(n^{1/2}\hat{\eta}^*_{n,3}\right) & \mathrm{Cov}_{\Phi}\left(n^{1/2}\hat{\eta}^*_{n,3}, n^{1/2}\hat{\eta}^*_{n,4}\right) \\ \mathrm{Cov}_{\Phi}\left(n^{1/2}\hat{\eta}^*_{n,3}, n^{1/2}\hat{\eta}^*_{n,4}\right) & \mathrm{Var}_{\Phi}\left(n^{1/2}\hat{\eta}^*_{n,4}\right) \end{pmatrix}, \\
\Psi_{F_1} &= \begin{pmatrix} \mathrm{Var}_{F}\left(n^{1/2}\hat{\eta}^*_{n,3}\right) & \mathrm{Cov}_{F}\left(n^{1/2}\hat{\eta}^*_{n,3}, n^{1/2}\hat{\eta}^*_{n,4}\right) \\ \mathrm{Cov}_{F}\left(n^{1/2}\hat{\eta}^*_{n,3}, n^{1/2}\hat{\eta}^*_{n,4}\right) & \mathrm{Var}_{F}\left(n^{1/2}\hat{\eta}^*_{n,4}\right) \end{pmatrix}.
\end{aligned} \tag{4.15}$$

The mean vector $\boldsymbol{\mu}_{F_0}$ and covariance matrix Ψ_{F_0} can be approximated by Monte Carlo simulation under the Gaussian distributions. Note that the off-diagonal terms of Ψ_{F_0} can be nonzero since those are finite sample covariances of the sample HL-moments even though Theorem 4.2 indicates their asymptotic independence.

Suitable estimators should be chosen for $\boldsymbol{\mu}_{F_1}$ and Ψ_{F_1}. We first use the sample HL-moment ratio $\hat{\eta}^*_{n,r}$ as an estimator of its expectation $E_F\left(\hat{\eta}^*_{n,r}\right)$. For the covariance matrix, we note from Theorem 4.1 that the asymptotic covariance matrix of the random variables $n^{1/2}\hat{\eta}^*_{n,3}$ and $n^{1/2}\hat{\eta}^*_{n,4}$ is Ψ^H whose (i,j)-th element is

$$\begin{aligned}
\Psi^H_{i,j} &= \left(\sigma^H_{r_i r_j} - \eta^*_{r_i}\sigma^H_{2r_i} - \eta^*_{r_j}\sigma^H_{2r_j} + \eta^*_{r_i}\eta^*_{r_j}\sigma^H_{22} \right) / \eta^2_2, \\
\sigma^H_{k_1 k_2} &= \int_0^1 \int_0^1 (u \wedge v - uv) H_{k_1-1}\left(\Phi^{-1}(u)\right) H_{k_2-1}\left(\Phi^{-1}(v)\right) \, \mathrm{d}F^{-1}(u) \, \mathrm{d}F^{-1}(v)
\end{aligned}$$

for $i,j \in \{1,2\}$ and $k_1, k_2 \in \{2, r_1, r_2\}$. Once we find an estimator of covariances, $\hat{\sigma}^H_{k_1 k_2}$, the estimator of the covariance matrix $\hat{\Psi}^H$ can be obtained by

$$\hat{\Psi}^{\mathrm{HL}}_{i,j} = \left(\hat{\sigma}^{\mathrm{HL}}_{r_i r_j} - \hat{\eta}^*_{r_i}\hat{\sigma}^{\mathrm{HL}}_{2r_i} - \hat{\eta}^*_{r_j}\hat{\sigma}^{\mathrm{HL}}_{2r_j} + \hat{\eta}^*_{r_i}\hat{\eta}^*_{r_j}\hat{\sigma}^{\mathrm{HL}}_{22} \right) / \hat{\eta}^2_2.$$

One way to estimate $\sigma^H_{k_1k_2}$ is to use the *jackknife* method. Parr and Schucany (1982) used the jackknife method to obtain a consistent estimator of the variance of the random variable $n^{1/2}\tilde{\theta}_n$ where

$$\tilde{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) X_{i:n}$$

in the case when J is trimmed, i.e. $J(u) = 0$ for $u \in (0, \alpha)$ and $u \in (1-\alpha, 1)$. Later, Sen (1984) showed the same result when the function J is bounded by integrable functions and broadened the range of possible constants used inside J rather than used a fixed constant $i/(n+1)$.

We first define a jackknife estimator for the covariances of the HL-moments as

$$\hat{\sigma}^{\mathrm{HL,JK}}_{n,r_1r_2} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{\eta}^{\mathrm{JK}}_{i:n,r_1r_2} - \bar{\eta}^{\mathrm{JK}}_{n,r_1r_2}\right)^2$$

for $r_1, r_2 = 3, 4, \cdots$ where

$$\begin{aligned}
\bar{\eta}^{\mathrm{JK}}_{n,r_1r_2} &= \frac{1}{n}\sum_{i=1}^{n}\hat{\eta}^{\mathrm{JK}}_{i:n,r_1r_2} \\
\hat{\eta}^{\mathrm{JK}}_{i:n,r_1r_2} &= \sum_{i=1}^{n} E\left(H_{r_1-1}\left(Z_{i:n}\right) + H_{r_2-1}\left(Z_{i:n}\right)\right) X_{i:n} \\
&\quad - \sum_{j=1}^{i-1} E\left(H_{r_1-1}\left(Z_{j:(n-1)}\right) + H_{r_2-1}\left(Z_{j:(n-1)}\right)\right) X_{j:n} \\
&\quad - \sum_{j=i+1}^{n} E\left(H_{r_1-1}\left(Z_{(j-1):(n-1)}\right) + H_{r_2-1}\left(Z_{(j-1):(n-1)}\right)\right) X_{j:n}.
\end{aligned}$$

Furthermore, we can define a jackknife estimator for the covariances of the L-moments as

$$\hat{\sigma}^{\mathrm{L,JK}}_{n,r_1r_2} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{\lambda}^{\mathrm{JK}}_{i:n,r_1r_2} - \bar{\lambda}^{\mathrm{JK}}_{n,r_1r_2}\right)^2$$

for $r_1, r_2 = 3, 4, \cdots$ where

$$\bar{\lambda}^{\mathrm{JK}}_{n,r_1r_2} = \frac{1}{n}\sum_{i=1}^{n}\hat{\lambda}^{\mathrm{JK}}_{i:n,r_1r_2}$$

and $\hat{\lambda}^{\mathrm{JK}}_{i:n,r_1r_2}$ is defined in the same way with $\hat{\eta}^{\mathrm{JK}}_{i:n,r_1r_2}$ by removing the i-th observation from the observed sample. The closed form expression is hard to obtain for $\hat{\lambda}^{\mathrm{JK}}_{i:n,r_1r_2}$ since it is defined as a U-statistic. Parr and Schucany (1982) pointed out that the jackknife estimators perform well for symmetric distributions.

Another way to estimate the covariance $\sigma^{\mathrm{HL}}_{r_1r_2}$ is to use the EDF defined in (4.6) as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\left(X_i \leq x\right).$$

Gardiner and Sen (1979) showed that plugging the EDF into F in the expression of the asymptotic variance of an L-statistic yields a consistent estimator of that variance. Referring to Equation (2.9) of that paper, we obtain the EDF estimator of the covariance $\sigma^{\mathrm{HL}}_{r_1r_2}$ as

$$\hat{\sigma}^{\mathrm{HL,EDF}}_{n,r_1r_2} = \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\left(\frac{i\wedge j}{n} - \frac{i}{n}\frac{j}{n}\right) H_{r_1-1}\left(\Phi^{-1}\left(\frac{i}{n}\right)\right) H_{r_2-1}\left(\Phi^{-1}\left(\frac{j}{n}\right)\right) \left(X_{(i+1):n} - X_{i:n}\right)\left(X_{(j+1):n} - X_{j:n}\right). \tag{4.16}$$

Similarly, the EDF estimator of the covariance $\sigma^{\mathrm{L}}_{r_1r_2}$ is given as

$$\hat{\sigma}^{\mathrm{L,EDF}}_{n,r_1r_2} = \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\left(\frac{i\wedge j}{n} - \frac{i}{n}\frac{j}{n}\right) P^{*}_{r_1-1}\left(\frac{i}{n}\right) P^{*}_{r_2-1}\left(\frac{j}{n}\right) \left(X_{(i+1):n} - X_{i:n}\right)\left(X_{(j+1):n} - X_{j:n}\right). \tag{4.17}$$

Based on these results, we proceed with finalization of a new test statistic. Let $\hat{\sigma}^{\mathrm{HL}}_{n,k_1k_2}$ be either the jackknife or EDF estimator of $\sigma^{\mathrm{HL}}_{k_1k_2}$. Then the estimator of the covariance matrix Ψ^{HL}_{ij} can accordingly be defined as

$$\hat{\Psi}^{\mathrm{HL}}_{i,j} = \left(\hat{\sigma}^{\mathrm{HL}}_{r_ir_j} - \hat{\eta}^{*}_{r_i}\hat{\sigma}^{\mathrm{HL}}_{2r_i} - \hat{\eta}^{*}_{r_j}\hat{\sigma}^{\mathrm{HL}}_{2r_j} + \hat{\eta}^{*}_{r_i}\hat{\eta}^{*}_{r_j}\hat{\sigma}^{\mathrm{HL}}_{22}\right)/\hat{\eta}^{2}_{2}.$$

Using Equation (4.14) and suitable estimators suggested so far, we can derive the *FLD type HL-moments based test statistic for Gaussianity* (*FHL test statistic*) as

$$
\begin{aligned}
D^{\mathrm{FHL}} &= \left(\hat{\boldsymbol{w}}^{\mathrm{HL}}\right)^T \begin{pmatrix} \hat{\eta}^*_{n,3} - E_\Phi\left(\hat{\eta}^*_{n,3}\right) \\ \hat{\eta}^*_{n,4} - E_\Phi\left(\hat{\eta}^*_{n,4}\right) \end{pmatrix} \\
&= n\left(\hat{\eta}^*_{n,3} - E_\Phi\left(\hat{\eta}^*_{n,3}\right), \hat{\eta}^*_{n,4} - E_\Phi\left(\hat{\eta}^*_{n,4}\right)\right) \\
&\quad \times \begin{pmatrix} \left(\psi^{\mathrm{HL}}_{\Phi,3}\right)^2 + \left(\hat{\psi}^{\mathrm{HL}}_{n,3}\right)^2 & \psi^{\mathrm{HL}}_{\Phi,34} + \hat{\psi}^{\mathrm{HL}}_{n,34} \\ \psi^{\mathrm{HL}}_{\Phi,34} + \hat{\psi}^{\mathrm{HL}}_{n,34} & \left(\psi^{\mathrm{HL}}_{\Phi,4}\right)^2 + \left(\hat{\psi}^{\mathrm{HL}}_{n,4}\right)^2 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\eta}^*_{n,3} - E_\Phi\left(\hat{\eta}^*_{n,3}\right) \\ \hat{\eta}^*_{n,4} - E_\Phi\left(\hat{\eta}^*_{n,4}\right) \end{pmatrix}. \quad (4.18)
\end{aligned}
$$

By adopting Fisher's linear discriminant method, we are able to exploit the covariance between the skewness and kurtosis estimators in an adaptive fashion. This gives a data adaptive optimal balance between the skewness and kurtosis estimators yielding a test statistic which is the Mahalanobis distance between the estimators $(\hat{\eta}_{n,3}, \hat{\eta}_{n,4})$ and their null hypothesis means $(E_\Phi(\hat{\eta}_{n,3}), E_\Phi(\hat{\eta}_{n,4}))$.

In the same way, we can derive a *FLD type L-moments based test statistic for Gaussianity* (*FL test statistic*) as follows

$$
\begin{aligned}
D^{\mathrm{FL}} &= n\left(\hat{\lambda}^*_{n,3} - E_\Phi\left(\hat{\lambda}^*_{n,3}\right), \hat{\lambda}^*_{n,4} - E_\Phi\left(\hat{\lambda}^*_{n,4}\right)\right) \\
&\quad \times \begin{pmatrix} \left(\psi^{\mathrm{L}}_{\Phi,3}\right)^2 + \left(\hat{\psi}^{\mathrm{L}}_{n,3}\right)^2 & \psi^{\mathrm{L}}_{\Phi,34} + \hat{\psi}^{\mathrm{L}}_{n,34} \\ \psi^{\mathrm{L}}_{\Phi,34} + \hat{\psi}^{\mathrm{L}}_{n,34} & \left(\psi^{\mathrm{L}}_{\Phi,4}\right)^2 + \left(\hat{\psi}^{\mathrm{L}}_{n,4}\right)^2 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\lambda}^*_{n,3} - E_\Phi\left(\hat{\lambda}^*_{n,3}\right) \\ \hat{\lambda}^*_{n,4} - E_\Phi\left(\hat{\lambda}^*_{n,4}\right) \end{pmatrix}. \quad (4.19)
\end{aligned}
$$

From simulation results, we confirmed that the RL-moments perform exactly the same with the L-moments in the goodness-of-fit test. Hence, we do not further develop a test based on the RL-moments. The reason of the same performance of the L- and RL-moments remains a future research topic.

The same reasoning can be applied to conventional moments. To derive, we first define the standardized moments μ_k and standardized sample moments $m_{n,k}$ as

$$
\mu_k = E\left(X - \mu\right)^k, \quad m_{n,k} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^k
$$

for $k = 1, 2, \cdots$. Note that the standardized sample moments can be understood as the EDF based estimators of their parallel population moments such that

$$m_{n,k} = E_{F_n}\left(X - E_{F_n}(X)\right)^k. \tag{4.20}$$

By Theorem 2.2.3.A of Serfling (1980), we have the almost sure convergence of the standardized sample moments to the standardized moments,

$$m_{n,k} \xrightarrow{\text{a.s.}} \mu_k \tag{4.21}$$

when $E|X_1|^k < \infty$. Direct application of Theorems 2.2.3.B and 3.3.A of Serfling (1980) yields

$$n^{1/2}\begin{pmatrix} \frac{m_{n,3}}{m_{n,3}^{3/2}} - \frac{\mu_3}{\mu_2^{3/2}} \\ \frac{m_{n,4}}{m_{n,2}^2} - \frac{\mu_4}{\mu_2^2} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, D^{\mathrm{M}}\Sigma^{\mathrm{M}}\left(D^{\mathrm{M}}\right)^T\right)$$

where $\mathbf{0} = (0,0)^T$ and

$$D^{\mathrm{M}} = \begin{pmatrix} -\frac{3}{2}\frac{\mu_3}{\mu_2^{5/2}} & \frac{1}{\mu_2^{3/2}} & 0 \\ -2\frac{\mu_4}{\mu_2^3} & 0 & \frac{1}{\mu_2^2} \end{pmatrix},$$

$$\Sigma^{\mathrm{M}}_{i,j} = \mu_{i+j+2} - \mu_{i+1}\mu_{j+1} - (i+1)\mu_i\mu_{j+2} - (j+1)\mu_{i+2}\mu_j + (i+1)(j+1)\mu_i\mu_j\mu_2$$

for $i, j \in \{1, 2, 3\}$. Based on Equation (4.21), we can obtain the estimators of covariances of the sample conventional moments as

$$\hat{D}^{\mathrm{M}} = \begin{pmatrix} -\frac{3}{2}\frac{m_{n,3}}{m_{n,2}^{5/2}} & \frac{1}{m_{n,2}^{3/2}} & 0 \\ -2\frac{m_{n,4}}{m_{n,2}^3} & 0 & \frac{1}{m_{n,2}^2} \end{pmatrix},$$

$$\hat{\Sigma}^{\mathrm{M}}_{i,j} = m_{n,i+j+2} - m_{n,i+1}m_{n,j+1} - (i+1)m_{n,i}m_{n,j+2} - (j+1)m_{n,i+2}m_{n,j} \tag{4.22}$$

$$+ (i+1)(j+1)m_{n,i}m_{n,j}m_{n,2}. \tag{4.23}$$

As a result, we obtain a *FLD type Jarque-Bera test statistic for Gaussianity* (*FJB test statistic*) as follows

$$D^{\mathrm{FJB}} = n\left(\hat{\gamma}^*_{n,1} - E_\Phi\left(\hat{\gamma}^*_{n,1}\right), \hat{\gamma}^*_{n,2} - E_\Phi\left(\hat{\gamma}^*_{n,2}\right)\right) \times \begin{pmatrix} 6 + \left(\hat{\psi}^{\mathrm{JB}}_{n,3}\right)^2 & \hat{\psi}^{\mathrm{JB}}_{n,34} \\ \hat{\psi}^{\mathrm{JB}}_{n,34} & 24 + \left(\hat{\psi}^{\mathrm{JB}}_{n,4}\right)^2 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\gamma}^*_{n,1} - E_\Phi\left(\hat{\gamma}^*_{n,1}\right) \\ \hat{\gamma}^*_{n,2} - E_\Phi\left(\hat{\gamma}^*_{n,2}\right) \end{pmatrix}. \tag{4.24}$$

Note that the asymptotic distribution of the sample conventional skewness and kurtosis in (4.9) is used in the equation.

4.4 Simulation results

Based on the goodness-of-fit test statistics developed in Sections 4.1 and 4.3, we compare their efficiencies in terms of their powers against various alternative hypothetical distributions. The Shapiro-Wilk (SW) and Anderson-Darling (AD) test statistics introduced in Section 4.1 are used as baseline tests. Many papers including Shapiro et al. (1968) pointed out that those test statistics perform better than distance-based statistics such as the Kolmogorov-Smirnov statistic for a wide range of alternative distributions, so we do not consider the Kolmogorov-Smirnov statistic here. For moments based tests, we consider the Jarque-Bera (4.8), FLD-type Jarque-Bera (4.24), HL-moments based (4.11), FLD-type HL-moments based (4.18), L-moments based (4.10) and FLD-type L-moments based (4.19) test statistics as main focuses of comparison. The abbreviations for those test statistics in the upcoming figures are presented in Table 4.1. As mentioned in Section 4.3, the RL-moments are not considered in these experiments since their power curves were exactly the same as those of the L-moments. For the covariance estimation method in the FLD-type improvement, we use the EDF method because the jackknife method made performances of the moments-based statistics worse. This seems to originate from the observation made in Parr and Schucany (1982) that the jackknife method performs worse than the EDF method for asymmetric distributions. Since many distributions that we deal with in this section and coming sections have skewness, the jackknife method is not included in our analysis.

Ab.	Test Name	Ab.	Test Name
AD	Anderson-Darling	SW	Shapiro-Wilk
JB	Jarque-Bera	FJB	FLD-type Jarque-Bera
HL	HL-moments based	FHL	FLD type HL-moments based
L	L-moments based	FL	FLD type L-moments based

Table 4.1: Abbreviations used in the legend of Figures 4.2 and 4.1. 'Ab.' stands for an abbreviation.

As alterative hypothetical distributions, we first focus on 2-component mixture distributions. The reason is that the TCGA lobular freeze data, which was introduced in Chapter 1 and will be deeply investigated in Chapter 5, consist of 5 subtypes each of which often forms its own cluster. There can be more than or equal to 5 clusters in data, but capturing many clusters with only skewness and kurtosis estimators is limited. Oja's criteria (Theorem 1.3) for a measure of kurtosis are related to bimodality of a distribution as shown in Oja (1981), but none of these criteria which are equivalent to tri- or multi-modality have been suggested so far. To compare the performances of kurtosis estimators to capture bimodality, we choose 2-component location-mixtures of Gaussian distributions.

Next, we consider Tukey's g and h distributions introduced in Section 3. Based on Theorem 3.1, the parameters g and h enable us to study a broad range of skewness and kurtosis of these distributions. We consider a few pairs of values of g and h in the simulation. Romão et al. (2010) used various types of 2-component scale-mixtures of the Gaussian distributions to cover alternative distributions with heavier tails than a single Gaussian distribution. However, we do not investigate them in the following simulations since Gleason (1993) pointed out that those distributions are not comparable with a single Gaussian distribution in Oja's sense (Theorem 1.3). That is, if $F = \Phi$ and $G = \alpha\Phi(\cdot|0,1)+(1-\alpha)\Phi(\cdot|0,10)$ for $0 < \alpha < 1$, then neither F does not have more kurtosis than G or vice versa (Theorem 1.2). Hence, we substitute Tukey's h distributions for those scale-mixtures.

For fair comparison, we do not use the critical regions given in the original papers but perform simulation to obtain them. For each test, 20,000 values of test statistics were computed from simulated samples, each of which has $n(= 15, 20, 25, \cdots, 100)$ observations generated from $\mathcal{N}(0,1)$. The significance level was set at $\alpha = 0.05$. Then 5,000 repetitions

of simulation from alternative hypothetical distributions were performed where the same number of observations n was generated.

The first family of alternative distributions is the family of 2-component mixtures of the Gaussian distributions which are shown in Figure 4.1. The plots in the left column show the power curves of test statistics and those in the right column show the densities of the null and alternative hypotheses. The colored lines in the left plots represent moments-based, JB, HL-moments based and L-moments based test statistics and their FLD-type variations while the gray-level lines represent the baseline test statistics, the Anderson-Darling (AD) and Shapiro-Wilk (SW).

The mixture model given in the top row gives a small proportion to one of its two components, so it is skew like the Tukey distribution $\mathcal{T}(g, 0)$ with positive g. The HL-moments based statistic performs better than the other moments based statistics. After the FLD improvement is applied, the FJB statistic performs better than the FHL statistic, and the FL statistic performs the worst. The SW statistic performs the best among all the statistics, while the AD statistic performs the worst. This indicates that even though we only incorporate skewness and kurtosis estimators, we can achieve higher power than conventionally used goodness-of-fit test statistics.

The second mixture model in the middle row has mild skewness and bimodality. In this case, the FLD variations significantly improve the performances of moments-based tests. The JB test has very low power against this alternative distribution, but it performs quite well after being adjusted by the FLD and catches up with other tests after the sample size 70. Among the moments-based tests, the L-moments perform better than the other two moments. This coincides with the observation made in Romão et al. (2010) where the L-moments based test performed the best for almost all the 2-component mixture models considered therein. Interestingly, the AD test performs the best among all tests and the SW test is not the best any more beaten by the FL statistic. Again this implies that the skewness and kurtosis based statistic can outperform conventional goodness-of-fit statistics.

The mixture model in the bottom row shows similar patterns. The FLD greatly improves the performances of all the moments-based tests. Among the moments, the L-moments based statistic and its FLD variation again achieve the best power outperforming

the two baseline tests, AD and SW. The JB test again achieves significant power improvement after being tuned by the FLD method.

Figure 4.2 shows the power comparison results for Tukey's g and h distributions. The top row presents a distribution which is obviously more skew to the right than the standard Gaussian distribution. The middle row presents a distributions with more skewness and heavy-tailedness than the standard Gaussian. The bottom row corresponds to a more asymmetric distribution than the distribution in the first row. The upper row studies the $\mathcal{T}(0.3, 0)$ as the alternative distribution. It can be seen from the upper right plot that $\mathcal{T}(0.3, 0)$ (black curve) is more skew to the right than $\mathcal{N}(0, 1)$ (gray curve). As can be seen from the upper left plot, the SW and FHL statistics perform the best for all the sample sizes. The FJB test catches up with those two statistics from the sample size 50, and then competes with them. Unlike 2-component mixtures of Gaussian distributions presented in 4.1, the L-moments based statistic and its FLD variation perform worse than the other moments based statistics. This implies that the direction of powers of moments can depend on an alternative distribution under consideration.

The middle row corresponds to Tukey$(0.3, 0.1)$ which is more skew to the right and has heavier tails on both sides than the Gaussian distributions. Similarly to the first row, the SW and FJB statistics perform the best for all the sample sizes and the FHL statistic catches up with them from the sample size 70. Again the L-moments based statistic and its FLD variation generally do not perform well. The lower row corresponds to Tukey$(4, 0)$ when the alternative distribution is asymmetric with heavier skewness than the first row. The patterns of power curves are very similar to those in the first row. The SW and FHL statistics perform the best and the FJB statistic catches up with them in the middle.

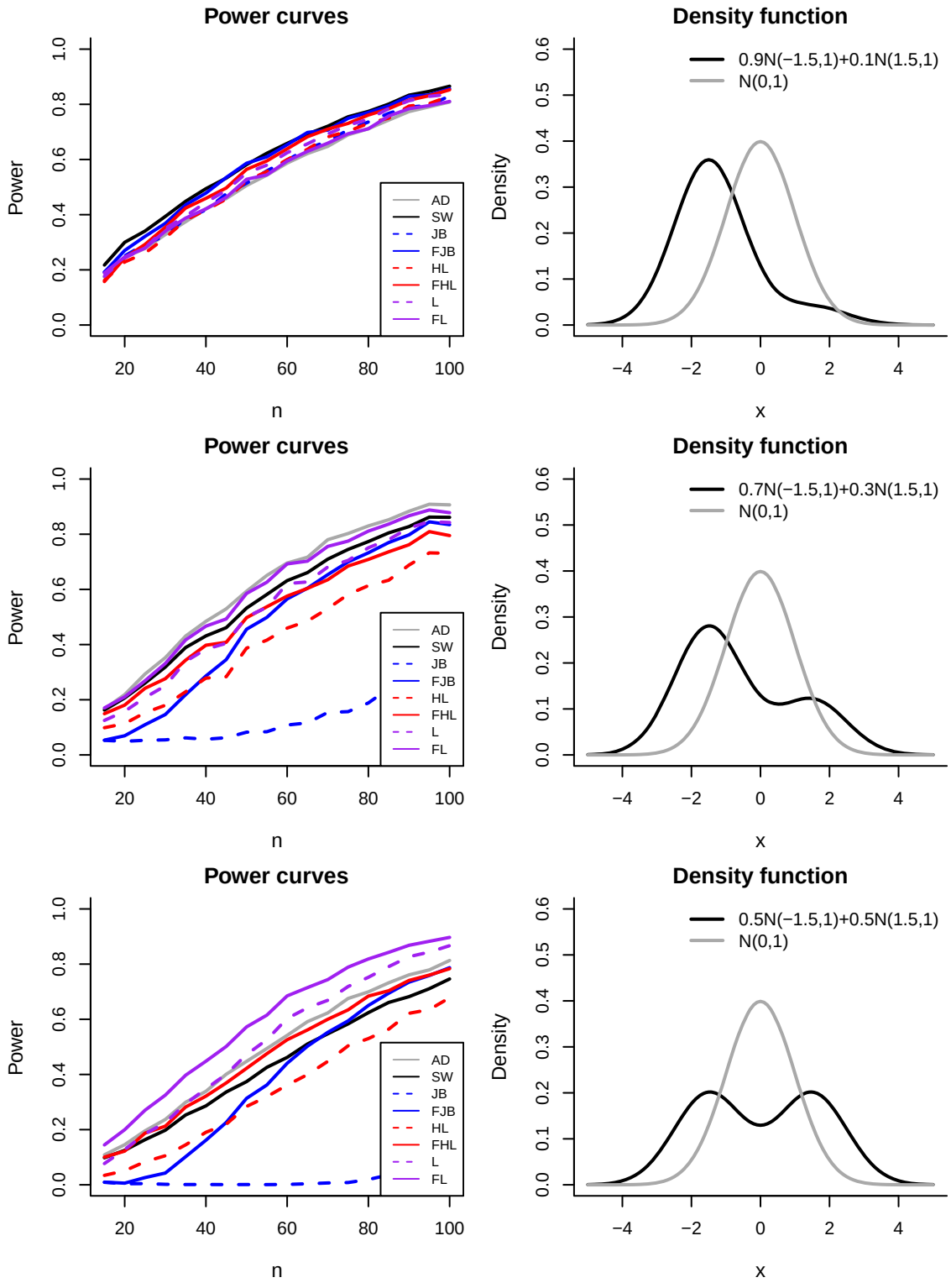

Figure 4.1: Power curves of test statistics when an alternative distribution is a 2-component mixture of Gaussian distributions. The FLD significantly increases powers for the alternative distributions in the middle and bottom rows. Among the moments-based statistics, the L-moments based statistic and its FLD variation perform the best, while the conventional moments based statistic and its variation perform the worst.

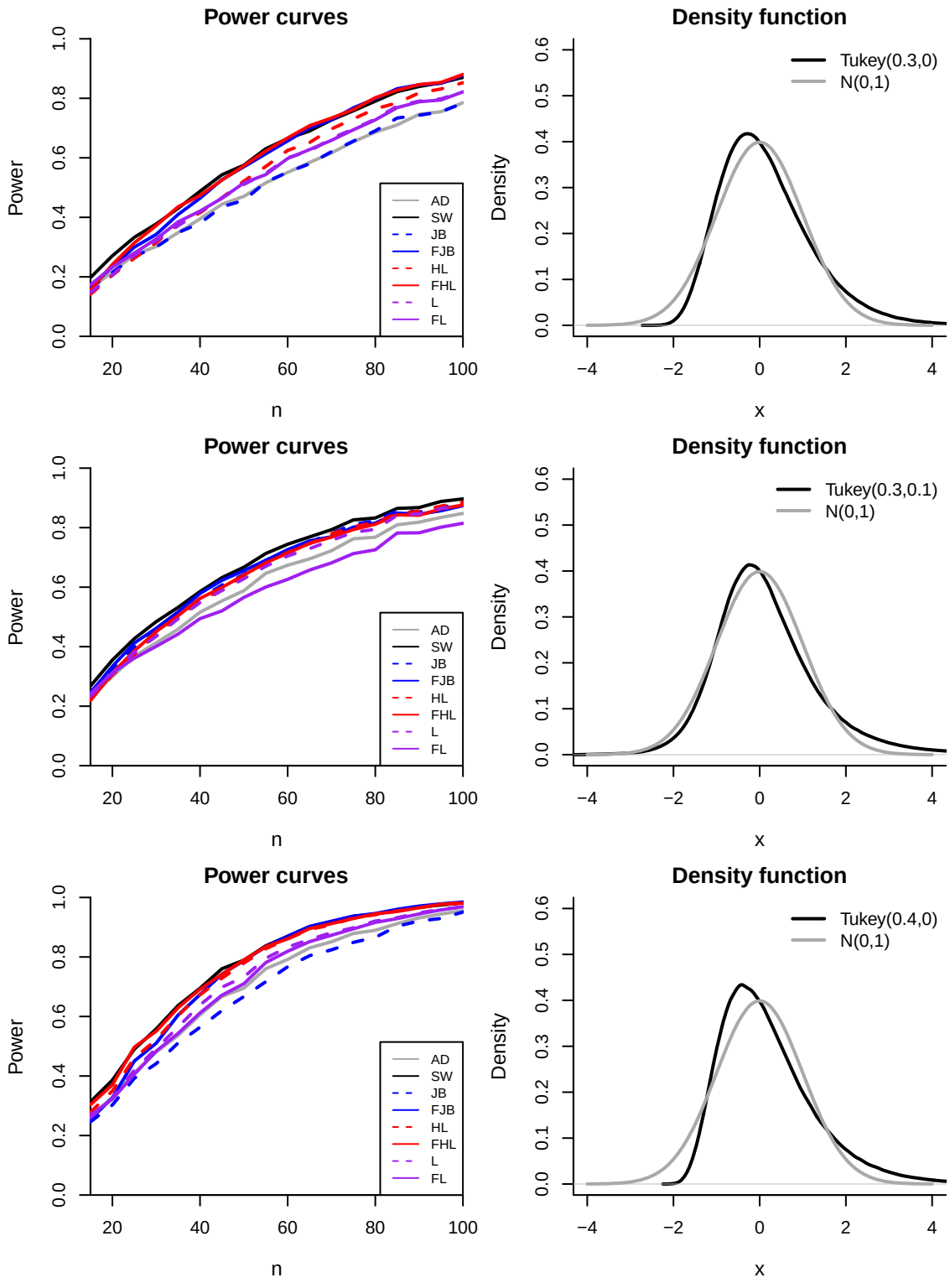

Figure 4.2: Power curves of test statistics when the alternative distribution is the Tukey. The legend in the right plots indicate which of Tukey's distributions is used.

CHAPTER 5

Variable screening analysis of TCGA lobular freeze data

5.1 Marginal distribution plots

As introduced in Chapter 1, the TCGA data contain 16,615 genes each of which has 817 expression profiles. The goal of our analysis is to discover interesting genes, i.e. variables, in terms of the shape of their marginal distributions. In high dimensional data such as the TCGA data, visualizing all the variables is usually infeasible. As discussed in Chapter 1, this motivates us to use summary statistics to sort and choose representative variables to look at. As shown below, such interesting variables often have skewness or multimodality in their marginal distributions. We investigate whether L-statistics based skewness and kurtosis estimators capture more interesting and genetically useful departures from the Gaussian distributions. The sorted list of genes generated by the L- and RL-moments were essentially the same, so we omit marginal distribution analysis of the RL-moments.

5.1.1 Comparison among skewness and kurtosis estimators

Figure 5.1 shows the 7 genes with the smallest conventional and HL-skewness. The format of this figure is the same as Figure 1.1 discussed in Chapter 1 except that the left arrow in the upper left plot indicates that the presented genes are at the bottom of the sorted genes. That is, the 7 genes presented have the smallest values of the current measure which is the conventional skewness for the upper panel of Figure 5.1. The upper two rows correspond to the conventional skewness while the lower two rows correspond to the negative HL-skewness. For color specification, see Table 1.1. Note that the 7 genes in the upper panel are all selected because of strong outliers on their left sides. On the other hand, the 7 genes in the lower panel have a much more biologically relevant type of skewness that is

clearly related to subtypes. This coincides with Theorem 3.5 in which the HL-skewness was shown to have more robustness than the conventional skewness.

Figure 5.2 shows the 7 genes selected by the negative L-skewness. It can be seen that the negative L-skewness selects a similar set of genes with the negative HL-skewness. Especially, the genes 'FOXA1', 'GSTT1', 'SPDEF', 'AGR3' and 'MLPH' are selected by both skewness measures. This implies that even though the HL-skewness was shown to have less robustness than the L-skewness as in Theorems 3.4 and 3.5, both share the level of robustness in real data. Comparison based on the genes with the largest skewness estimates shows a similar pattern among different measures, so it is omitted.

The next direction of sorting genes is kurtosis of distributions. As mentioned so far, high kurtosis of a distribution implies its central peakedness and heavy-tailedness on its both sides while low kurtosis implies the distribution's light-tailedness which is often related to bimodality. In the case of skewness, all the three measures of skewness, conventional, HL- and L-skewness, are zero at the Gaussian distributions. However, the measures of kurtosis presented herein have different zeros from each other. On top of that, the different kurtosis measures have different robustness as shown in Chapter 3. This motivates us to compare the genes selected by different kurtosis measures.

Figure 5.3 shows the 7 genes with the largest conventional and HL-kurtosis values which are presented at the upper and lower panels, respectively. The fact that we are looking at the top of the sorted list of genes is implied by the right arrow in the quantile plot in the upper left. Like the conventional skewness, the conventional kurtosis screens genes with at least one strong outlier on either the left or right side of their marginal distributions. This implies that conventional moment based measures are driven more by a couple of outliers than the shapes of underlying distributions. The HL-kurtosis, on the contrary, finds the genes with heavy-tailedness in their distributional bodies. Especially, the gene 'MTAP' has a concentrated region of Basal type samples ($\triangleleft$) on its right tail, and the gene 'CBLC' has a cluster of Her2 type samples ($*$) on its light tail. However, the gene with the largest HL-kurtosis value, 'CSTF2T', is driven by two outliers on its left and right sides. This implies that even though the HL-kurtosis is based on an L-statistic, it can sometimes be driven by outliers.

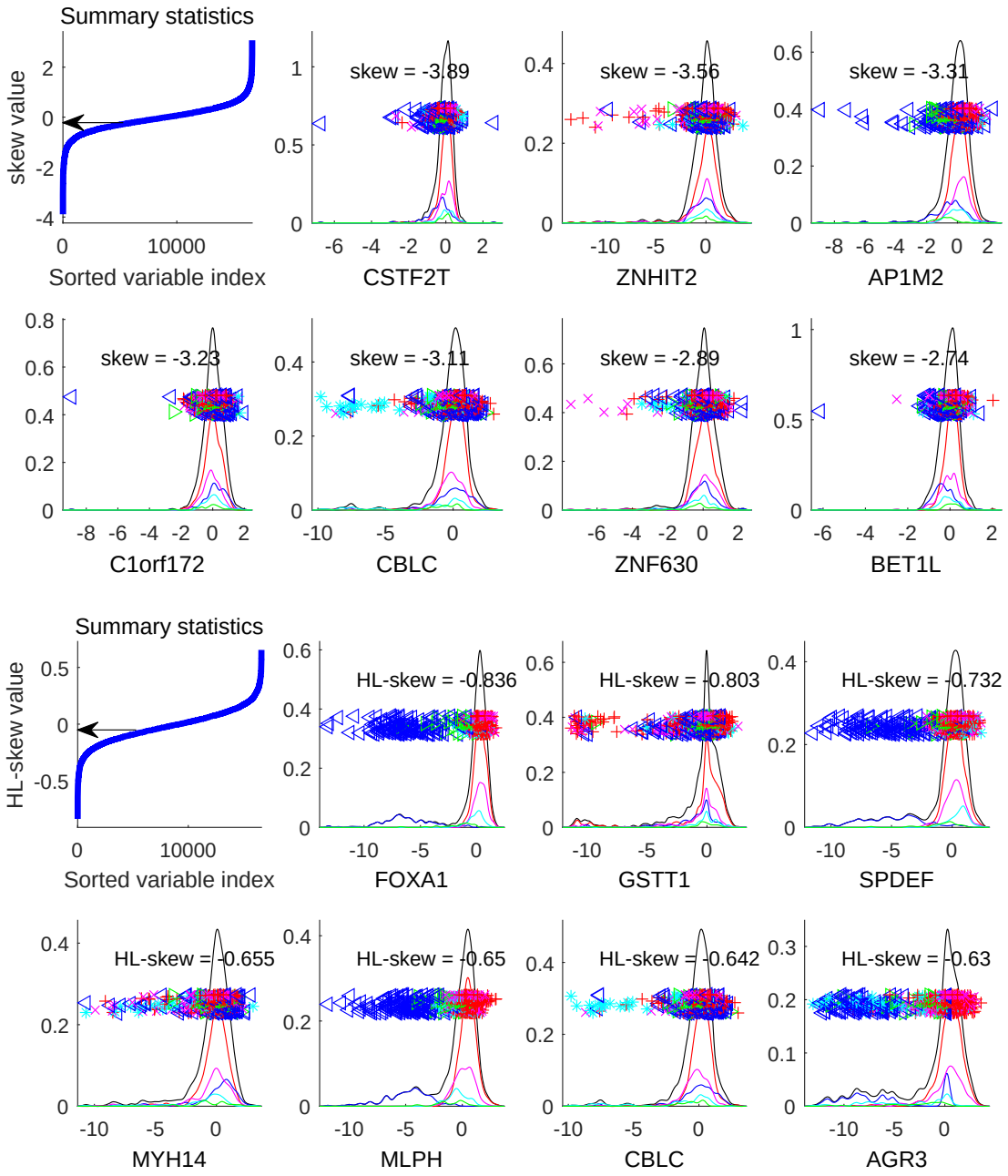

Figure 5.1: The 7 genes with smallest conventional skewness (upper panel) and HL-skewness (lower panel). The upper left plot in each panel shows the quantile plot of the statistics. The genes selected by conventional skewness have strong outliers on their left sides while the genes of the HL-skewness have strong subtype driven skewness.

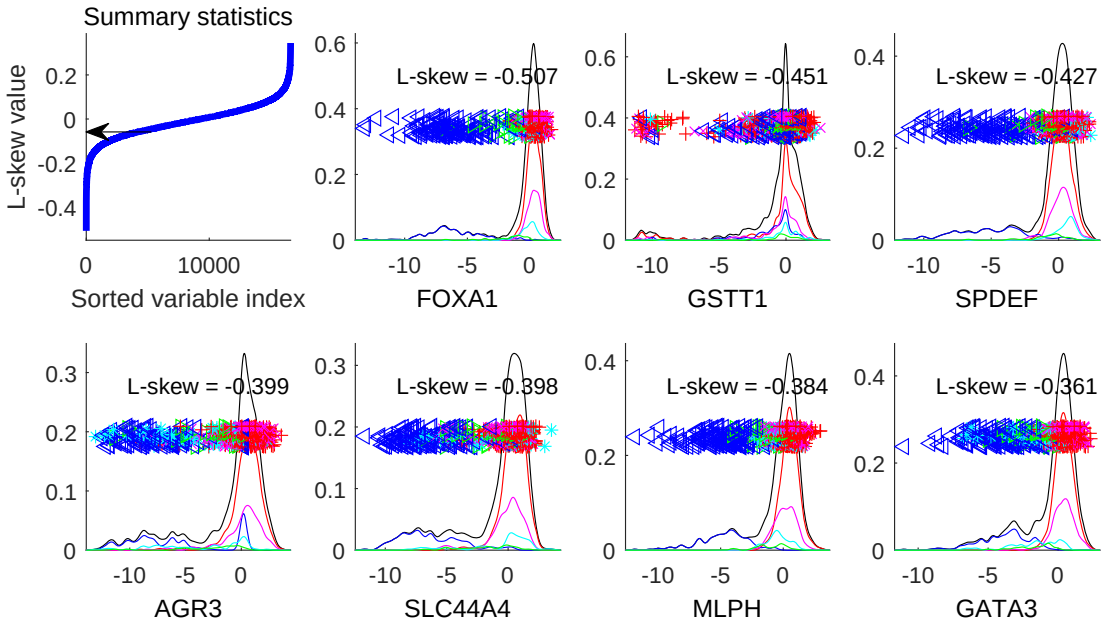

Figure 5.2: The 7 genes with smallest L-skewness. All the genes clearly have skewness to the left side in their distributional bodies.

As seen in Chapter 3, the L-kurtosis has better robustness than the other two kurtosis measures considered in this dissertation, which can be assured by Figure 5.4. All the 7 genes picked up by the HL-kurtosis have heavy-tailedness in thier distributional bodies. Unlike the HL-kurtosis, the L-kurtosis finds the gene 'CBLC' as the most highly kurtotic gene which has a heavy left tail. This implies that the L-kurtosis has better robustness than the HL-kurtosis in actual data as well.

Figure 5.5 shows the 7 genes with the smallest values of the conventional and HL-kurtosis. The upper pannel corresponds to the conventional kurtosis. All the 7 genes have multimodality, i.e. they do not have unimodal shape. Even though the negative kurtosis value is typically thought to indicate bimodality, this figure shows that it can also capture multimodality of data. Interestingly, the sets of 7 genes selected by conventional and HL-kurtosis are exactly the same. The 7 genes, 'RPL9', 'GSTM1', 'PRAME', 'SLC7A4', 'BMPR1B', 'RPS27' and 'C10orf82', appear in both panels of Figure 5.5. This seems to be because at the negative kurtosis end, both methods find bimodal genes. The genes screened by the L-kurtosis showed a similar pattern, so we omit presenting those genes.

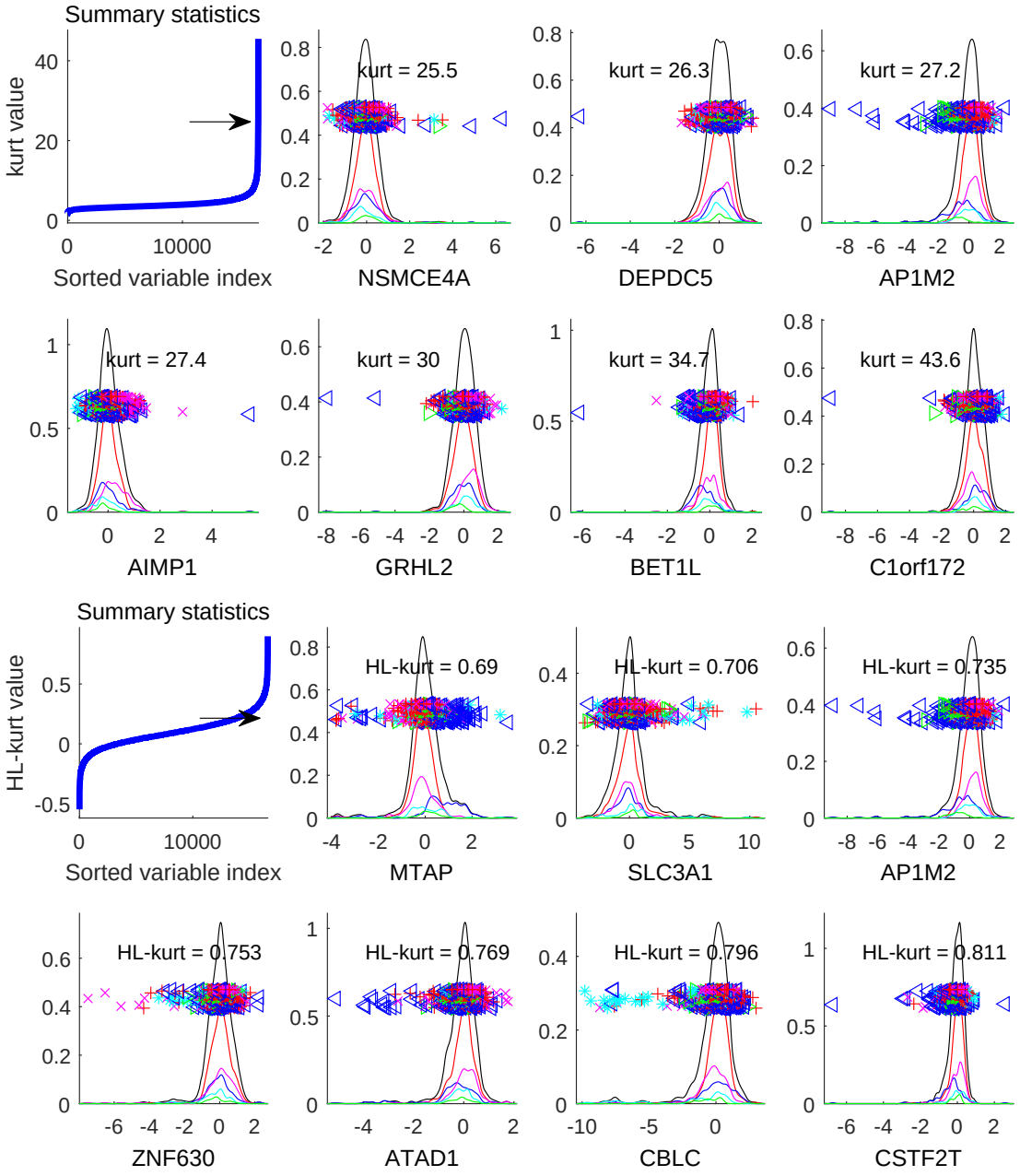

Figure 5.3: The 7 genes with largest conventional kurtosis (upper panel) and HL-kurtosis (lower panel). The upper left plot in each panel shows the quantile plot of the statistics. The genes selected by conventional kurtosis have strong outliers on their left or right sides while the genes of the HL-kurtosis mostly have heavy-tailed distributions in their bodies.

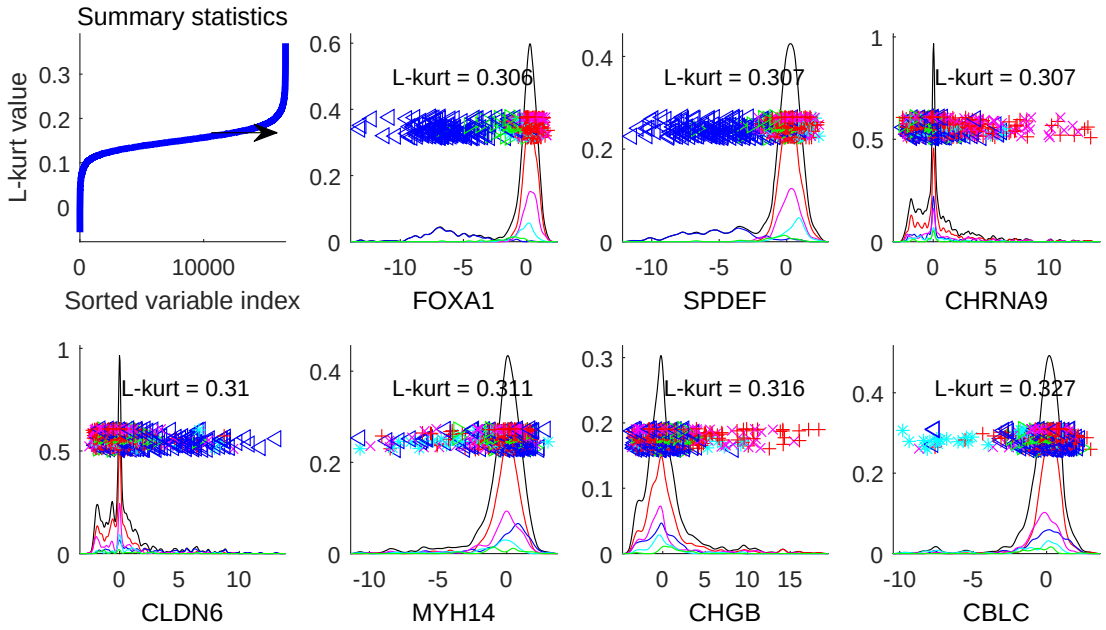

Figure 5.4: The 7 genes with largest L-kurtosis. All the genes clearly have skewness to the left side in their bodies.

5.1.2 Comparison among goodness-of-fit test statistics

The two goodness-of-fit tests of Gaussianity presented in Chapter 4 are the Anderson-Darling (AD) and Shapiro-Wilk (SW) tests. Since one of the main assumptions that we make in TCGA data analysis is that biologically meaningful genes have departure from Gaussianity in their distributional bodies, we check whether those two goodness-of-fit test statistics actually screen such interesting genes. Figure 5.6 shows the 7 genes with the highest AD test statistics and lowest SW test statistics. Recall from Section 4.2 that the AD test rejects the Gaussianity hypothesis for large test statistic values while the SW test rejects the hypothesis for small statistic values.

As can be seen from Figure 5.6, the AD statistic screened 7 genes with at least one strong outlier. It was mentioned in Section 4.2 that the AD test statistic can be highly affected by outliers. On the contrary, the SW test statistic generally screened the genes with skewness to the left or right side of their marginal distributions. Among those genes, the genes 'FOXA1' and 'SPDEF' especially exhibit distinction between different subtypes implying that those are biologically related to breast cancer subtypes.

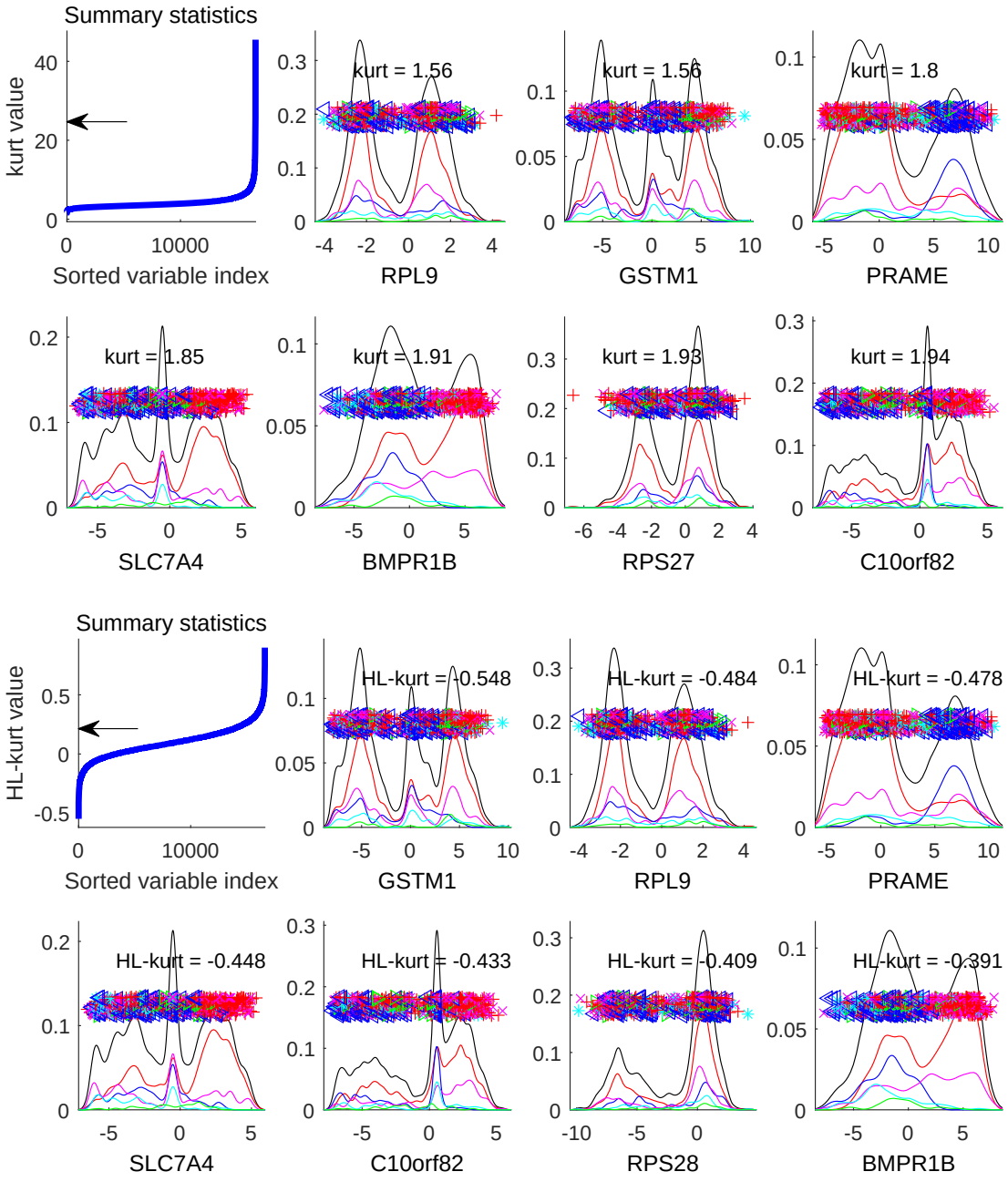

Figure 5.5: The 7 genes with the smallest conventional kurtosis (upper panel) and HL-kurtosis (lower panel). The sets of 7 genes in both panels have multimodality and no outlier.

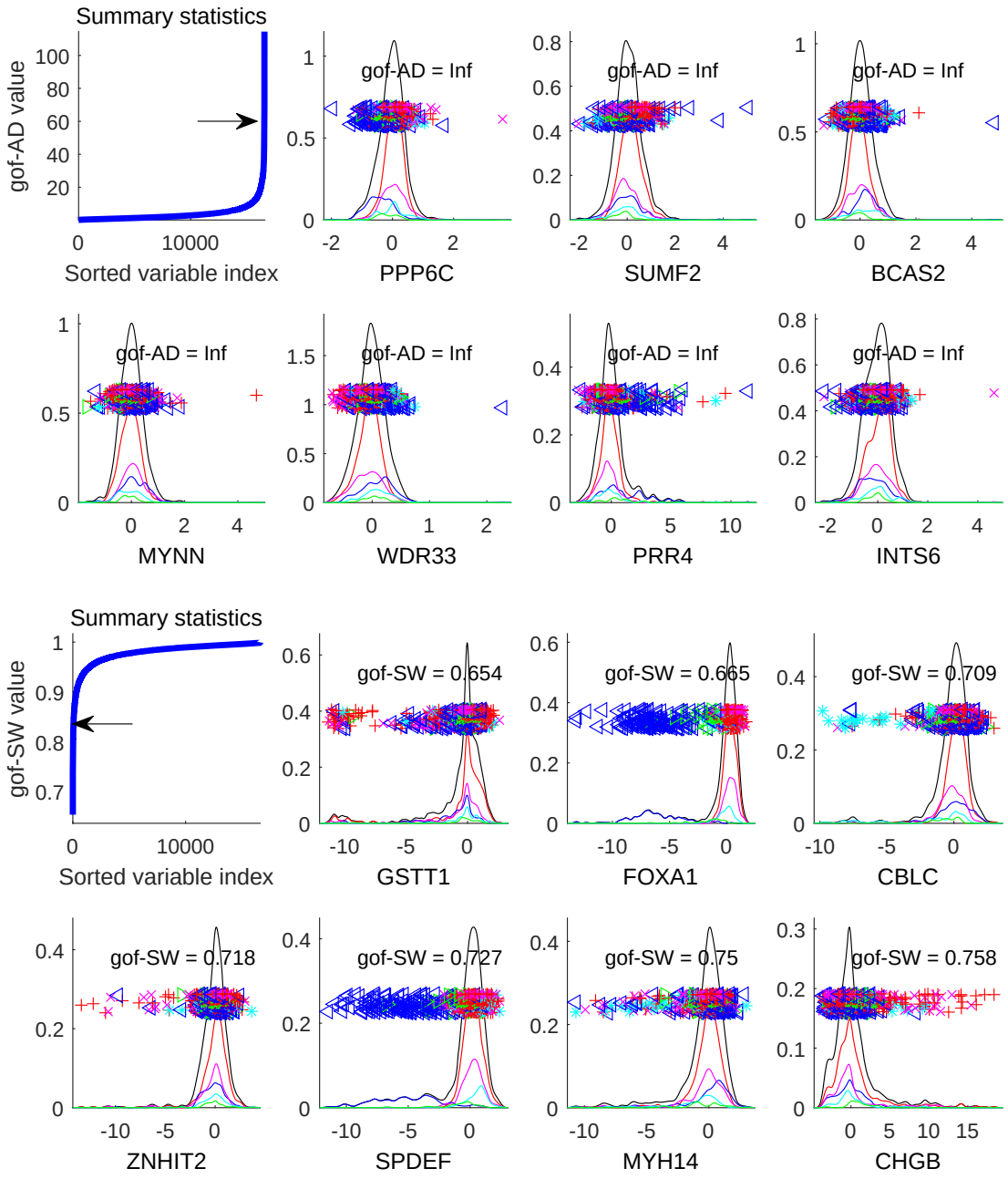

Figure 5.6: The 7 genes with the largest AD test statistics (upper panel) and smallest SW test statistics (lower panel). The 7 genes screened by the AD statistic were driven by outliers while those screened by the SW test statistic have strong skewness in their distributional bodies.

Figure 5.7 shows the comparison result between the JB and FJB test. As shown in Chapter 3, the JB test statistic is highly sensitive to outliers. Here, all the 7 genes screened by the JB statistic have at least one strong outlier on either side of their distributions. However, the FJB statistic finds the genes with skewness in their distributional bodies rather than a couple of outliers. perhaps surprising result can be explained as follows. The FJB statistic formula given in (4.24) as

$$D^{\mathrm{FJB}} = n\left(\hat{\gamma}^*_{n,1} - E_\Phi\left(\hat{\gamma}^*_{n,1}\right), \hat{\gamma}^*_{n,2} - E_\Phi\left(\hat{\gamma}^*_{n,2}\right)\right) \times \begin{pmatrix} 6 + \left(\hat{\psi}^{\mathrm{JB}}_{n,3}\right)^2 & \hat{\psi}^{\mathrm{JB}}_{n,34} \\ \hat{\psi}^{\mathrm{JB}}_{n,34} & 24 + \left(\hat{\psi}^{\mathrm{JB}}_{n,4}\right)^2 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\gamma}^*_{n,1} - E_\Phi\left(\hat{\gamma}^*_{n,1}\right) \\ \hat{\gamma}^*_{n,2} - E_\Phi\left(\hat{\gamma}^*_{n,2}\right) \end{pmatrix},$$

implies that outliers affect not only the skewness and kurtosis measures but also their covariances, with some cancellation effect on the statistic D^{FJB}.

As shown in Figure 5.8, the 7 genes with the smallest HL statistic values have more skewness in their distributional bodies than those with the smallest JB statistic values. Especially, the genes 'SPDEF' and 'FOXA1' have a cluster of Basal type ($\triangleleft$) samples on their right side. Robustness of the HL-skewness relative to the conventional skewness shown in Section 5.1.1 is confirmed here. The FHL statistic seems to screen more subtype relevant genes than the HL statistic. All the 7 genes screened by the FHL, except the gene 'TDRD12', have a subtype driven cluster of samples.

The L-moments based statistic screens more subtype related genes than the HL-moments based statistic as can be seen from the upper panel of Figure 5.9. All the genes except the gene 'GSTT1' have a cluster of either Basal ($\triangleleft$) or LumA ($+$) type samples. The FL test statistic screens mostly the same genes with the L test statistic. The 5 genes, 'SLC44A4', 'MLPH', 'AGR3', 'SPDEF' and 'GSTT1', were screened by both the L and FL statistics. This implies that there might not be much difference between performances of L and FL statistics, which will be confirmed in the next section.

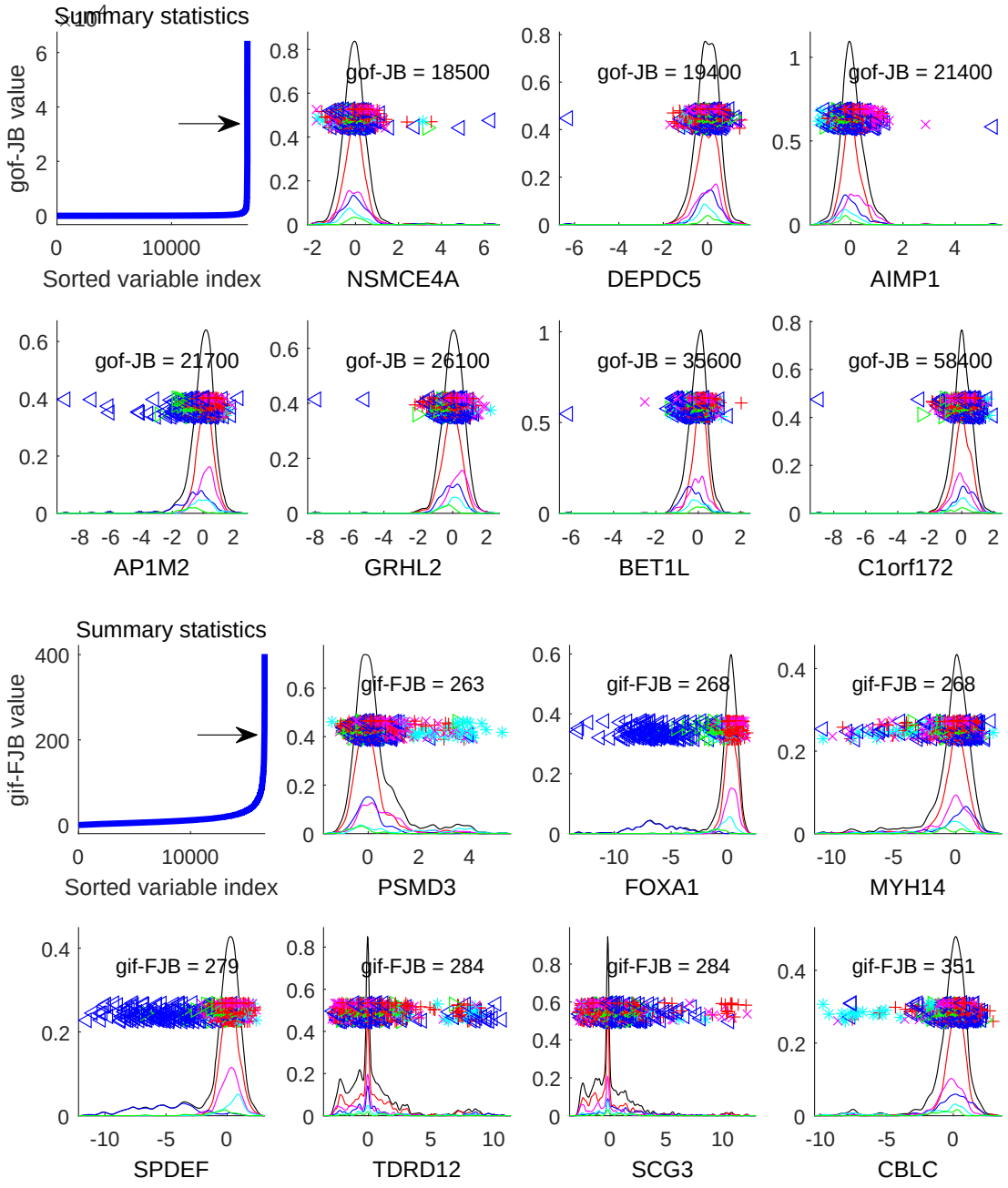

Figure 5.7: The 7 genes with the largest JB statistics (upper panel) and FJB statistics (lower panel). All the 7 genes screened by the JB statistic have at least one outlier while the 7 genes screened by the FJB statistic have skewness and multimodality in their distributional bodies.

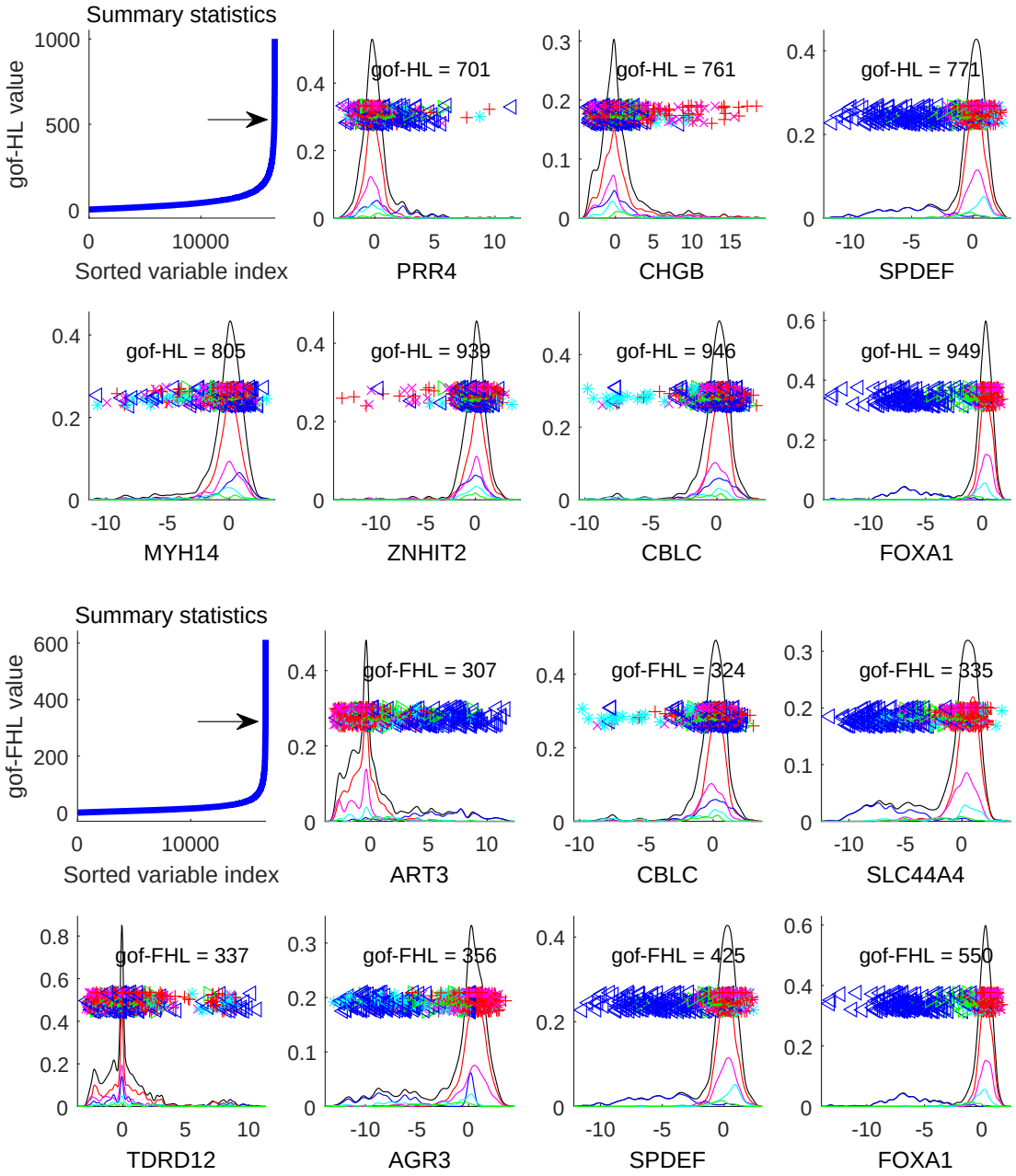

Figure 5.8: The 7 genes with the largest HL statistic values (upper panel) and FHL statistic values (lower panel). The FHL seems to screen more subtype relevant genes than the HL.

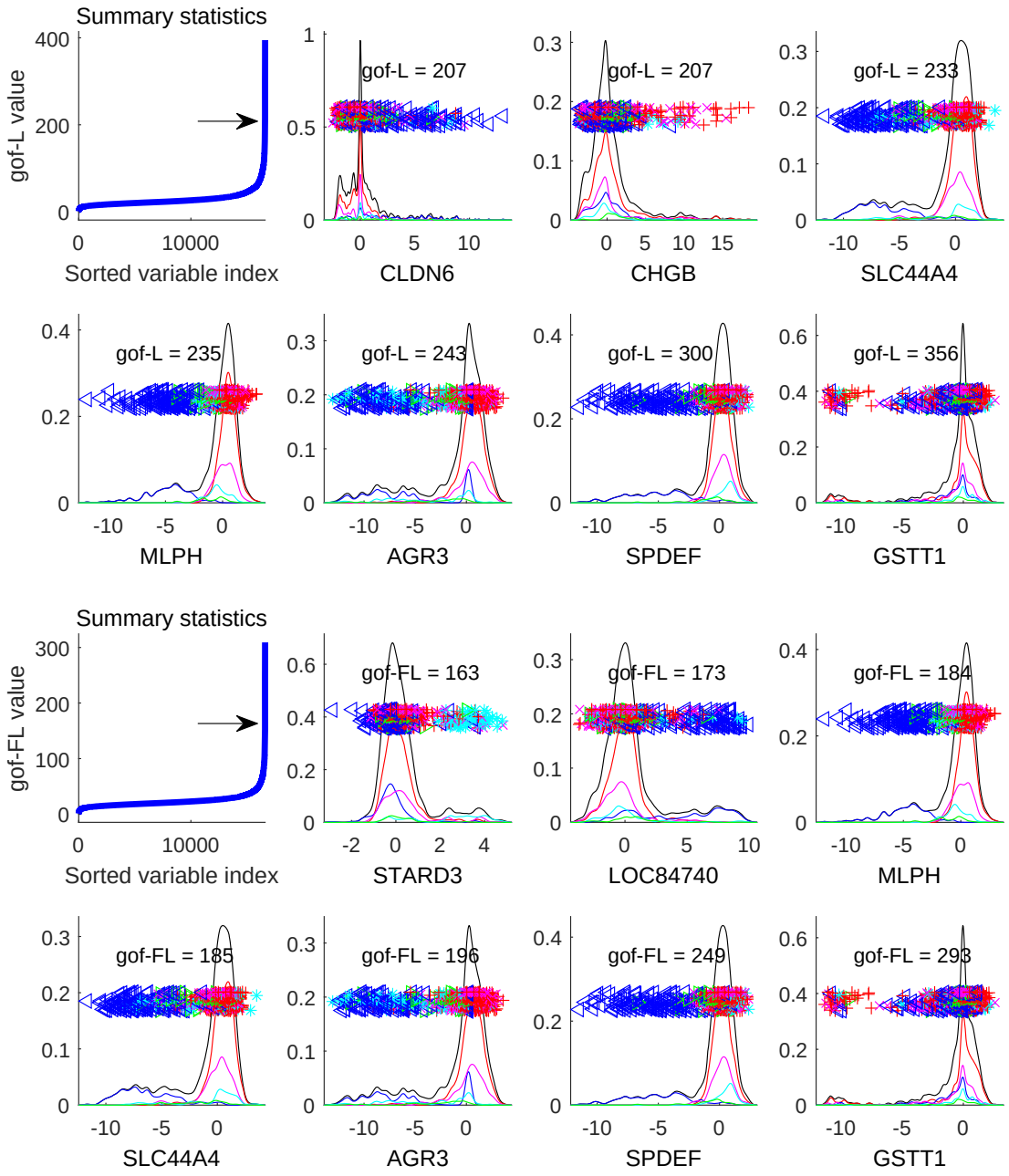

Figure 5.9: The 7 genes with the largest L test statistic values (upper panel) and FL statistic values (lower panel). Unlike the JB and HL, there is not much difference between the sets of genes screened by the L and FL test statistics.

5.2 Gene Set Enrichment Analysis

A main goal of the TCGA data analysis done here is to assess the performances of various skewness and kurtosis measures by their abilities in screening biologically meaningful genes. Marginal distributions shaped similarly with the standard Gaussian distributions or having outliers on either side tend not to be helpful for understanding subtypes. This suggests use of robust skewness and kurtosis measures which can efficiently capture more relevant departures from Gaussianity to screen for biologically meaningful genes. To assess such abilities, we focus on the *Gene Set Enrichment Analysis* (*GSEA*) introduced in Subramanian et al. (2005).

The GSEA deals with a matrix D of mRNA expression profiles in which rows represent genes and columns represent samples. Samples belong to one of two phenotype classes, e.g. tumors resistant to a drug or not. For a ranked list L of genes, the GSEA assesses whether interesting genes in an independent gene set S are randomly distributed throughout L or primarily found at the top or bottom of it. Figure 5.10 adopted from Figure 1 in Subramanian et al. (2005) shows an example of an expression profile matrix and an independent gene set. The heatmap given on the left side shows the levels of expression values by color. The rectangle on the right side indicates the location of genes in an independent gene set presented by horizontal lines. The GSEA assesses whether the interesting genes corresponding to the horizontal lines gather more closely to the top, bottom or both of the list.

To statistically evaluate goodness of a ranked list, we first define a score which is a function of true and false positive rates. Suppose that we rank the N genes in D to form $L = \{g_1, g_2, \cdots, g_N\}$ according to their correlations r_j for $j = 1, 2, \cdots, N$ with the 2-class phenotype. We first compute the probabilities of hitting and missing interesting genes in S in the top n list of genes extracted from the ranked list L as

$$P_{\text{hit}}(S, n) = \frac{\sum_{j \le n, g_j \in S} 1}{\sum_{g_j \in S} 1}, \quad P_{\text{miss}}(S, n) = \frac{\sum_{j \le n, g_j \notin S} 1}{\sum_{g_j \notin S} 1}. \tag{5.1}$$

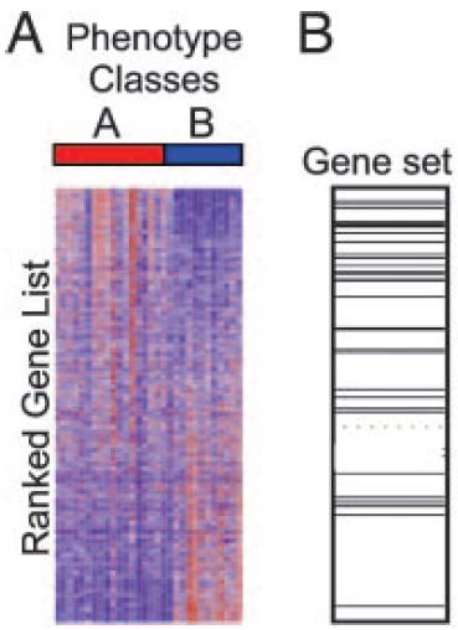

Figure 5.10: Figure 1 of Subramanian et al. (2005). The left heatmap represents a sorted list of genes whose expression values are colored based on their values, and the right rectangle represents the locations of biologically meaningful genes in that sorted list by the horizontal lines.

The probabilities P_{hit} and P_{miss} can be understood as true and false positive rates, respectively. Note that Subramanian et al. (2005) suggested using a correlation value instead of 1 given in Equation 5.1. However, using the values of different skewness and kurtosis measures can cause a problem in our case since different measures have different scales. The largest problem is the conventional kurtosis which ranges from 0 to the infinity. Computing the ES based on the values of the conventional kurtosis can yield unpredictable results.

As we increase the size n of the top-n list, both $P_{\text{hit}}(S, i)$ and $P_{\text{miss}}(S, i)$ become functions of n whose maximum or minimum value is of interest. We define the *enrichment score* (*ES*) of the gene set S as

$$\text{ES}(S) = \max_{1 \le n \le N} \left(P_{\text{hit}}(S, n) - P_{\text{miss}}(S, n) \right)$$

to assess the significance of difference between them. That is, we compute the maximum distance of a random walk from zero in which we step up when we encounter an interesting gene and step down when we miss it. If interesting genes are gathered at the top of the ranked list, the enrichment score will be highly positive, and if they are gathered at the

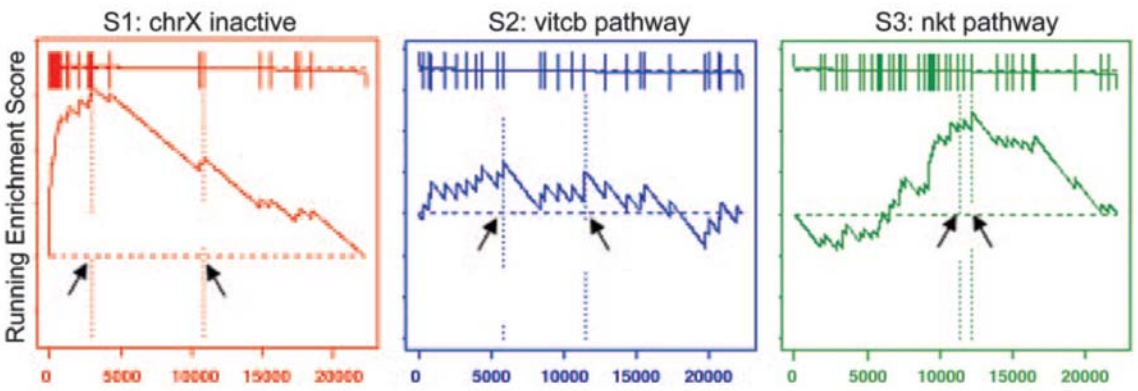

Figure 5.11: Figure 2 of Subramanian et al. (2005). Each plot shows the trace of the random walk generated by Equation 5.1 for three different independent gene sets.

bottom, the score will be highly negative. Since both of those cases are of interest, we perform 2-sided tests to obtain the p-value of ES(S).

Figure 5.11 shows the three examples of the random walk generated by the probabilities in (5.1). Each plot shows the trace of the random walk for three different independent gene sets. The vertical bars presented on top of the plots show the locations of independent genes in the ranked list of genes. In the left plot, the independent genes gather at the top of the ranked list yielding highly positive ES. On the other hand, the middle plot does not have highly positive or negative ES since the independent genes are equally spread throughout the ranked list of genes.

To obtain a p-value, we do a permutation test. We randomly permute the ranks of genes in the ranked list, and for the k-th permuted list, we compute the k-th enrichment score $\mathrm{ES}_k(S)$. Repeat the permutation for, say, $K = 1,000$ times and obtain the null distribution of $\mathrm{ES}_k(S)$. This gives a nominal p-value of ES(S). Since there are many different sets of interesting genes, we adjust the p-values for multiple comparisons so that multiple gene sets can be jointly used to assess the significance of L. For multiple gene sets S_m for $m = 1, 2, \cdots, M$, we control the False Discovery Rate (FDR) by computing

$$\mathrm{FDR}(S) = \frac{\sum_{m=1}^{M} \sum_{k=1}^{K} I\left(|\mathrm{ES}_k\left(S_m\right)| \geq |\mathrm{ES}(S)|\right)}{KM}.$$

If the FDR(S) is less 0.05, then it is said that the corresponding gene set S is significantly enriched. We also investigate the FDR level 0.25 since the GSEA User Guide (`http://software.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html`) recommends that both the FDR levels 0.05 and 0.25 are worth being investigated.

5.2.1 Comparison among skewness and kurtosis estimators

On top of the conventional and Gaussian Centered L-moments, we consider also the quantile-based measures which have been known to be robust as baseline measures. Bowley's skewness measure (Bowley, 1920) is a typically used quantile-based measure of skewness defined as

$$\gamma_p = \frac{F^{-1}(1-p) - F^{-1}(1/2) - \left\{F^{-1}(1/2) - F^{-1}(p)\right\}}{F^{-1}(1-p) - F^{-1}(p)}$$

where p is usually set to 0.75. On the other hand, Ruppert's interfractile range ratio (Ruppert, 1987) is frequently used as a measure of kurtosis and defined as

$$\gamma_{p_1,p_2} = \frac{F^{-1}\left(1-p_1\right) - F^{-1}\left(p_1\right)}{F^{-1}\left(1-p_2\right) - F^{-1}\left(p_2\right)}$$

where the parameter p_1 and p_2 are usually set to 0.9 0.7, respectively. The sample quantiles are used to estimate the quantiles in γ_p and γ_{p_1,p_2} for data. Both Bowley's and Ruppert's estimators were shown to satisfy Oja's criteria for measures of skewness and kurtosis, respectively in those papers. The papers Ruppert (1987) and Groeneveld (1991) showed that their influence functions evaluated at symmetric distributions are bounded while we showed in Theorems 3.4 and 3.5 that the HL- and L-moments have unbounded influence functions for some symmetric distributions. This indicates that those quantile based measures can play the role of baseline robust estimators.

The analysis was performed using the Gene Set Enrichment v2.2.4 software released by the Broad Institute (`http://software.broadinstitute.org/gsea/index.jsp`). The independent gene sets in MSigDB v5.2 were downloaded, and the gene sets with the minimum 15 and maximum 5,000 genes were used in the analysis. This left us 15,470 gene sets for the analysis including FDR computation.

FDR	Skewness	HL-skewness	p-value	Kurtosis	HL-kurtosis	p-value
0.05	3022	**3147**	0.078	1520	**2037**	**0**
FDR	Skewness	HL-skewness	p-value	Kurtosis	HL-kurtosis	p-value
0.25	6220	**6349**	0.14	4088	**5009**	**0**

Table 5.1: The numbers of genes screened by the conventional and HL-moments with FDR less than 0.05 and 0.25. Generally, the HL-moment screened for interesting genes better than the conventional moments with larger superiority for the direction of kurtosis.

Table 5.1 shows the comparison results of the conventional and HL-moments when the FDR is fixed as 0.05 and 0.25. The values in the first, second, fourth and fifth columns are the numbers of independent gene sets with the given FDR screened by different measures. The p-values are obtained by Fisher's exact test in which screened genes are treated as positive samples and the other genes are treated as negative samples. By performing this test, we can measure the statistical significance of the differences between different measures in screening genes. About comparison between the conventional and HL-moments, the HL-skewness performs better than the conventional skewness but its degree is not statistically significant at either FDR level 0.05 or 0.25. For kurtosis, HL-kurtosis significantly performs better than the conventional kurtosis at both levels. This indicates that the robustness of the HL-moments enables themselves to screen meaningful genes without being much affected by outliers.

About comparison between the HL- and L-moments in Table 5.2, the L-skewness performs significantly better than the HL-skewness at the FDR level 0.05 but the HL-skewness performs slightly, but not significantly, better than the L-skewness at the level FDR 0.25. For the kurtosis, HL-kurtosis performs significantly better than the L-kurtosis at both FDR levels. This shows that relative superiority of the HL- and L-moments depends on the direction of departure from Gaussianity. Their abilities in the skew direction can depend on the size of the test that we look at, but the HL-moments clearly screen for genes better in the heavy-tailed and bimodal direction.

From comparison results between the L-moments and quantile-based measures given in 5.3, the L-moments perform better than the quantile-based measures in every direction. This shows that even though L-statistics have less robustness than the quantile-based mo-

FDR	HL-skewness	L-skewness	p-value	HL-kurtosis	L-kurtosis	p-value
0.05	3147	**3303**	**0.03**	**2037**	1145	**0**
FDR	HL-skewness	L-skewness	p-value	HL-kurtosis	L-kurtosis	p-value
0.25	**6349**	6287	0.48	**5009**	3349	**0**

Table 5.2: The numbers of genes screened by the HL- and L-moments with FDR less than 0.05 and 0.25. Generally, the HL-moments perform better than the L-moments in the direction of kurtosis whatever FDR level is given. In the direction of skewness, their relative performances depend on the FDR levels with statistically significant superiority of the L-moments when the FDR is 0.05.

FDR	L-skewness	Q-skewness	p-value	L-kurtosis	Q-kurtosis	p-value
0.05	**3303**	2016	**0**	**1445**	681	**0**
FDR	L-skewness	Q-skewness	p-value	L-kurtosis	Q-kurtosis	p-value
0.25	**6287**	4318	**0**	**3349**	2700	**0**

Table 5.3: The numbers of genes screened by the L-moments and quantile-based moments with FDR less than 0.05 and 0.25. Generally, the HL-moments perform better than the L-moments in the direction of kurtosis whatever FDR level is given. For both skewness and kurtosis, the L-moments based estimators screen for meaningful genes better than the quantile based estimators.

ments, they are actually more able to screen for important variables in high dimensional data. The mild balance between robustness and efficiency that L-statistics possess seems to fit to the goal of TCGA data analysis.

5.2.2 Comparison among goodness-of-fit test statistics

Since our main claim in TCGA data analysis is that biologically meaningful genes often have asymmetry and bimodality in their marginal distributions, we check whether goodness-of-fit test statistics for Gaussianity based on skewness and kurtosis estimators screen such genes well. We use the AD and SW test statistics as baselines. We first check whether the skewness and kurtosis based test statistics perform better than those baseline statistics, then compare relative performances among different moments based statistics. Unlike Subsection 5.2.1, only the direction of positive enrichment scores is of our interest since goodness-of-fit test statistics put Gaussian shape genes at the bottom of their ranked lists. Hence, the FDR is fixed as 0.25 in this subsection to offset reduction of the critical region.

FDR	AD	JB	p-value	AD	FJB	p-value
0.25	**933**	623	**0**	933	**1334**	**0**
FDR	SW	JB	p-value	SW	FJB	p-value
0.25	**1081**	623	**0**	1081	**1334**	**0**

Table 5.4: The numbers of genes screened by the baseline statistics, AD and SW, and the JB and FJB with FDR 0.25. The JB statistic is inferior to the baseline statistics, but the FJB statistic screens biologically meaningful genes better than the baseline statistics.

FDR	AD	HL	p-value	AD	FHL	p-value
0.25	933	**972**	0.369	933	**1340**	**0**
FDR	SW	HL	p-value	SW	FHL	p-value
0.25	**1081**	972	**0.0136**	1081	**1340**	**0**

Table 5.5: The numbers of genes screened by the baseline statistics, AD and SW, and the HL and FHL with FDR 0.25. The HL statistic competes with the baseline statistics, but the FJB statistic outperforms those baseline statistics.

The result of comparing the JB and FJB with the AD and SW is given in Table 5.4. As can be seen from the table, the JB significantly performs worse than the AD and SW. However, after Fisher type improvement is applied to the JB statistic, its performance is dramatically enhanced and significantly better than both the SW and AD. This shows that the FLD type adjustment introduced in Section 4.3 improves not only the power of the test statistic in the goodness-of-fit test but also its ability in screening biologically meaningful genes in our data. In addition, the SW and AD statistics seem not to be appropriate for the purpose of variable screening in our context, since their directions of departure from Gaussianity are too broadly spread to capture interesting shape in marginal distributions. The fact that the FJB outperforms both the AD and SW shows that we are studying the right direction of departure from Gaussianity, which are skewness and bimodality.

Similarly to the JB statistic, the HL-moments based statistic does not outperform the baseline statistics, but the FLD type improvement makes its performance surpass the AD and SW statistics as can be seen in Table 5.5. The HL-moments based statistic is better than the AD statistic but the degree is not significant, and it performs worse than the SW statistic. However, the FHL statistic performs significantly better than the AD and SW. The comparison between the FJB and FHL statistics is presented later.

FDR	AD	L	p-value	AD	FL	p-value
0.25	933	**1218**	**0**	933	**1437**	**0**
FDR	SW	L	p-value	SW	FL	p-value
0.25	1081	**1218**	**0.003**	1081	**1437**	**0**

Table 5.6: The numbers of genes screened by the AD, SW and L and FL with FDR 0.25. Both the L and FL statistics outperform the AD and SW statistics.

FDR	JB	HL	p-value	HL	L	p-value
0.25	623	**972**	**0**	972	**1218**	**0**
FDR	FJB	FHL	p-value	FHL	FL	p-value
0.25	1334	**1340**	0.919	1340	**1437**	0.0562

Table 5.7: The numbers of genes screened by the AD, SW and L and FL with FDR 0.25. The HL test statistic outperforms the JB test statistic, and the L test statistic outperforms the HL test statistic.

In Subsection 5.1.2, we saw that the L-moments based statistics seem to best capture biologically meaningful genes in data among all statistics based on marginal distributions. This can be confirmed in Table 5.6 in which both the L and FL test statistics perform significantly better than the baseline test statistics. The degree to which the L test statistic without FLD improvement performs better than the baseline statistics is significant. This implies that the L-moments themselves have enough power to capture important structure in data without covariance adjustment.

The most interesting part is comparison among different skewness and kurtosis based statistics, which is presented in Table 5.7. Before the Fisher improvement is applied, the order of performances is JB $<$ HL $<$ L which coincides with our observation made in Subsection 5.1.2. This also coincides with robustness result shown in Chapter 3 where the order of robustness was shown to be JB $<$ HL $<$ L. After the Fisher improvement is applied, the differences between performances of different statistics get much narrower, but the order of relative performances remains the same. The difference between the FHL and FL is slightly non-significant. On the contrary, the p-value of the difference between the FJB and FL was slightly significant ($p = 0.0423$), which is not shown in the table. The numbers of genes screened by the FJB and FHL are almost the same. This shows that the HL-moments are on the border between the sample conventional moments and L-statistics.

CHAPTER 6
Future work

6.1 Including more moments in goodness-of-fit test statistics

The goodness-of-fit test statistics developed in Chapter 4 have skewness and kurtosis estimators as building blocks. However, there is no limit on the number of terms that can be included in those test statistics. For example, the FHL statistic can be defined as

$$
\begin{aligned}
D^{\mathrm{FHL},6} = n & \left(\hat{\eta}^*_{n,3} - E_\Phi\left(\hat{\eta}^*_{n,3}\right), \hat{\eta}^*_{n,4} - E_\Phi\left(\hat{\eta}^*_{n,4}\right), \hat{\eta}^*_{n,5} - E_\Phi\left(\hat{\eta}^*_{n,5}\right), \hat{\eta}^*_{n,6} - E_\Phi\left(\hat{\eta}^*_{n,6}\right)\right) \\
& \times \begin{pmatrix}
\left(\psi^{\mathrm{HL}}_{\Phi,3}\right)^2 + \left(\hat{\psi}^{\mathrm{HL}}_{n,3}\right)^2 & \psi^{\mathrm{HL}}_{\Phi,34} + \hat{\psi}^{\mathrm{HL}}_{n,34} & \cdots & \psi^{\mathrm{HL}}_{\Phi,36} + \hat{\psi}^{\mathrm{HL}}_{n,36} \\
\psi^{\mathrm{HL}}_{\Phi,34} + \hat{\psi}^{\mathrm{HL}}_{n,34} & \left(\psi^{\mathrm{HL}}_{\Phi,4}\right)^2 + \left(\hat{\psi}^{\mathrm{HL}}_{n,4}\right)^2 & \cdots & \psi^{\mathrm{HL}}_{\Phi,46} + \hat{\psi}^{\mathrm{HL}}_{n,46} \\
 & & \ddots & \\
\psi^{\mathrm{HL}}_{\Phi,36} + \hat{\psi}^{\mathrm{HL}}_{n,36} & \psi^{\mathrm{HL}}_{\Phi,46} + \hat{\psi}^{\mathrm{HL}}_{n,46} & \cdots & \left(\psi^{\mathrm{HL}}_{\Phi,6}\right)^2 + \left(\hat{\psi}^{\mathrm{HL}}_{n,6}\right)^2
\end{pmatrix}^{-1} \\
& \times \begin{pmatrix}
\hat{\eta}^*_{n,3} - E_\Phi\left(\hat{\eta}^*_{n,3}\right) \\
\hat{\eta}^*_{n,4} - E_\Phi\left(\hat{\eta}^*_{n,4}\right) \\
\hat{\eta}^*_{n,5} - E_\Phi\left(\hat{\eta}^*_{n,5}\right) \\
\hat{\eta}^*_{n,6} - E_\Phi\left(\hat{\eta}^*_{n,6}\right)
\end{pmatrix}.
\end{aligned}
$$

by including the fifth and sixth HL-moments. As mentioned in Section 7 of Oja (1981), the sixth moment term is usually related to the tri-modality of a distribution. This implies that by incorporating more terms in the test statistic, we can measure departure from Gaussianity in more various directions. However, that can result in poor efficiency in each of directions of departure from Gaussianity. For example, $D^{\mathrm{FHL},6}$ can result in poorer performance than D^{FHL} in discriminating between a Gaussian distribution and a bimodal distribution. In the TCGA data analysis, it seems appropriate to use the terms between

η_3^* and η_{10}^* to screen penta-modalities since there are five subtypes in the data which often form their own groups.

6.2 Centering L-functionals at other distributions

An approach to setting a non-uniform distribution as the center of the L-moments has been explored earlier in Hosking (2007). That paper developed the theory of *trimmed L-moments* (*TL-moments*) which had originally been proposed by Elamir and Seheult (2003). The trimmed L-moments are basically linear combinations of expected order statistics but have a different form from the L-moments (1.6) given as

$$\lambda_r^{(s,t)} = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k \begin{pmatrix} r-1 \\ k \end{pmatrix} EX_{r+s-k:r+s+t}.$$

Note that the s smallest order statistics and t largest order statistics are excluded from the random sample of size $r+s+t$. Even though the TL-moments were proposed for better robustness, an interesting discovery made in the paper was that appropriately setting the coefficient before expected order statistics makes a family of logistic distributions the distributional center of the TL-moments.

However, Hosking (2007) did not show the possibility of systematically shifting the distributional center of the L-moments to an arbitrary distribution. As seen in Theorem 2.3, one of the interesting properties of the RL-moments is that those can be centered at any symmetric distribution using its expected order statistics. This suggests that the HL-moments can also be centered at another family of distributions different from the Gaussian distributions.

For the HL-moments, note that their Gaussian centering property

$$\eta_r = \int_{-\infty}^{\infty} x f(x) H_{r-1}\left(\Phi^{-1}(F(x))\right)\, \mathrm{d}x$$

comes from the orthogonality of the Hermite polynomials H_r with respect to the weight function ϕ. We can consider a similar type of moments defined as

$$\eta^{\alpha}_{\Gamma,r}(F) = \int_{-\infty}^{\infty} x f(x) L^{\alpha}_{r-1}\left(\Gamma^{-1}(F(x)|\alpha,1)\right)\,\mathrm{d}x$$

where L^{α}_r is the r-th order Laguerre polynomial and $\Gamma(\cdot|\alpha,\beta)$ is the CDF of the gamma distribution with the parameters α and β. It can be seen that $\eta^{\alpha}_{\Gamma,r}\left(\Gamma(\cdot|\alpha,1)\right) = 0$ for $r = 3, 4, \cdots$ from the equation

$$\int_{-\infty}^{\infty} x g(x|\alpha,1) L^{\alpha}_{r-1}(x)\,\mathrm{d}x = 0$$

since

$$x = -(1+\alpha-x) + (1+\alpha) = -L^{\alpha}_1(x) + (1+\alpha)L^{\alpha}_0(x),$$

i.e. x is a linear combination of the first two Laguerre polynomials of order α. Since the limit of the Gamma distributions is the Gaussian distribution, the relationship between the moments $\{\eta^{\alpha}_{\Gamma,r} r = 1, 2, \cdots\}$ and the HL-moments $\{\eta_r | r = 1, 2, \cdots\}$ might shed light on the relationship between the Laguerre and Hermite polynomials.

CHAPTER 7

Proofs

The following lemma can be derived from Section 3.1 of David and Nagaraja (2003).

Lemma 7.1. Suppose that $E|X|^k < \infty$ for some k, then we have

$$\lim_{u\to 1} \left|F^{-1}(u)\right|^k (1-u) = 0 \text{ and } \lim_{u\to 0} \left|F^{-1}(u)\right|^k u = 0. \tag{7.1}$$

This further implies that

$$\lim_{u\to 1} \left|F^{-1}(u)\right|^{sk} (1-u)^s = 0 \text{ and } \lim_{u\to 0} \left|F^{-1}(u)\right|^{sk} u^s = 0 \quad \forall s > 0$$

Moreover, if the CDF F has a MGF, then we have

$$\lim_{u\to 1} \left|F^{-1}(u)\right|^{s} (1-u)^t = 0 \text{ and } \lim_{u\to 0} \left|F^{-1}(u)\right|^{s} u^t = 0 \quad \forall s, t > 0. \tag{7.2}$$

□

The following lemma in Chapter 22 of Abramowitz and Stegun (1964) is also used throughout this chapter.

Lemma 7.2. The Hermite polynomials $\{H_r | r = 0, 1, 2, \cdots\}$ satisfy the recursion formula $(r+1)H_r(x) = \frac{\partial}{\partial x} H_{r+1}(x)$ for $x \in \mathbb{R}$ and $r = 0, 1, \cdots$. In addition, if r is odd, then H_r is an odd function. If r is even, then H_r is an even function. □

To obtain the influence functions of the HL- and RL-moment ratios, the following lemma is needed.

Lemma 7.3. For a functional $\theta = \theta_1/\theta_2$ such that $\theta_1, \theta_2 : \mathcal{F} \to \mathbb{R}$, we have

$$\begin{aligned} \mathrm{IF}(x; F, \theta) &= \frac{\theta_2(F)\mathrm{IF}(x; F, \theta_1) - \theta_1(F)\mathrm{IF}(x; F, \theta_2)}{\theta_2(F)^2}, \\ \mathrm{SIF}(x; F, \theta) &= \frac{\theta_2(F)\mathrm{SIF}(x; F, \theta_1) - \theta_1(F)\mathrm{SIF}(x; F, \theta_2)}{\theta_2(F)^2} \end{aligned}$$

since both the influence and symmetric influence functions are right-hand derivatives of a function. □

We present a useful definition and a lemma for deriving the symmetric influence functions given below. We say a functional $\theta : \mathcal{F} \to \mathbb{R}$ is *symmetric* if $\theta(F) = \theta\left(F_{-1,0}\right)$ for all $F \in \mathcal{F}$.

Lemma 7.4. Let $\theta : \mathcal{F} \to \mathbb{R}$ be an L-functional in the form (1.5) of the main paper. If θ is a symmetric L-functional and F is a symmetric distribution, then we have $\mathrm{SIF}(x; F, \theta) = \mathrm{IF}(x; F, \theta)$ for all $x \in \mathbb{R}$.

Proof . From Equation (5.35) of Huber and Ronchetti (2009), if both the functional θ and distribution F are symmetric, then we have

$$\mathrm{IF}(x; F, \theta) = \mathrm{IF}(-x; F, \theta). \tag{7.3}$$

Let $Q_u : \mathcal{F} \to \mathbb{R}$ be a functional such that $Q_u(F) = F^{-1}(u)$. Then it can be seen from Equations (3.46) and (3.47) of Huber and Ronchetti (2009) that

$$\mathrm{SIF}\left(x, F, Q_u\right) = \frac{1}{2}\left\{\mathrm{IF}\left(-x; F, Q_u\right) + \mathrm{IF}\left(x; F, Q_u\right)\right\}.$$

It can be seen using this result and the first equality in Equation (3.49) of Huber and Ronchetti (2009) that

$$\mathrm{SIF}(x, F, \theta) = \frac{1}{2}\{\mathrm{IF}(-x; F, \theta) + \mathrm{IF}(x; F, \theta)\}.$$

Combining this equation and Equation (7.3), we obtain the desired result. □

Proof of Theorem 2.1. It can be seen from

$$\eta_1 = \int_0^1 F^{-1}(u) H_0\left(\Phi^{-1}(u)\right)\, \mathrm{d}u = \int_0^1 F^{-1}(u)\, \mathrm{d}u$$

that the first HL-moment is the mean. It was shown in Oja (1981) that the mean satisfies Oja's criterion for a measure of location.

To check whether η_2 satisfies the first condition of Oja's criterion for a measure of scale (Definition 1.3.b), we let $Y = aX + b$ and F, G be the cumulative distribution functions of X, Y respectively. First, assume that $a > 0$. Then we have $G^{-1}(u) = aF^{-1}(u) + b$ for $0 < u < 1$. Now we have

$$\begin{aligned}
\eta_2(G) &= \int_0^1 G^{-1}(u)\Phi^{-1}(u)\, \mathrm{d}u \\
&= \int_0^1 \left\{aF^{-1}(u) + b\right\} \Phi^{-1}(u)\, \mathrm{d}u \\
&= a\int_0^1 F^{-1}(u)\Phi^{-1}(u)\, \mathrm{d}u + b\int_0^1 \Phi^{-1}(u)\, \mathrm{d}u \\
&= a\int_0^1 F^{-1}(u)\Phi^{-1}(u)\, \mathrm{d}u \\
&= a\eta_2(F). \qquad (7.4)
\end{aligned}$$

Hence, we have $\eta(G) = a\eta(F)$ when $a > 0$. If we assume that $a < 0$, then we have $G^{-1}(u) = aF^{-1}(1-u) + b$. Following the same steps of derivation as Equation (7.4), we can obtain $\eta_2(G) = -a\eta_2(F)$. Combining these two results, we obtain $\eta_2(G) = |a|\eta_2(F)$.

To check the second condition, we first have

$$\begin{aligned}
G^{-1}(F(x)) - x \text{ is nondecreasing in } x &\Leftrightarrow \frac{f(x)}{g(G^{-1}(F(x)))} \geq 1 \\
&\Leftrightarrow \frac{1}{g(G^{-1}(u))} - \frac{1}{f(F^{-1}(u))} \geq 0 \\
&\Leftrightarrow G^{-1}(u) - F^{-1}(u) \text{ is nondecreasing in } u.
\end{aligned}$$

This yields $G^{-1}(u)-F^{-1}(u)-G^{-1}(1/2)+F^{-1}(1/2) \leq 0$ for $u \leq 1/2$ and $G^{-1}(u)-F^{-1}(u)-G^{-1}(1/2)+F^{-1}(1/2) \geq 0$ for $u \geq 1/2$. Now we have

$$\begin{aligned}
\eta_2(G)-\eta_2(F) &= \int_0^1 \left\{G^{-1}(u)-F^{-1}(u)\right\}\Phi^{-1}(u)\,\mathrm{d}u \\
&= \int_0^1 \left\{G^{-1}(u)-F^{-1}(u)-G^{-1}\left(\frac{1}{2}\right)+F^{-1}\left(\frac{1}{2}\right)\right\}\Phi^{-1}(u)\,\mathrm{d}u \\
&\quad + \int_0^1 \left\{G^{-1}\left(\frac{1}{2}\right)-F^{-1}\left(\frac{1}{2}\right)\right\}\Phi^{-1}(u)\,\mathrm{d}u \\
&\geq 0
\end{aligned}$$

where the last inequality results from the same signs of two functions inside the integral.

We first prove the following lemma.

Lemma 7.5. If $G^{-1}\circ F$ is convex on the support of F and $\mu(F)=\mu(G)$, then there exist two points $0<u_1<u_2<1$ such that

$$G^{-1}(u)-F^{-1}(u) \geq 0 \text{ for } 0<u\leq u_1 \text{ and } u_2<u<1,$$
$$G^{-1}(u)-F^{-1}(u) \leq 0 \text{ for } u_1<u\leq u_2.$$

Proof . Since the function $G^{-1}\circ F$ is convex, it meets the function $y=x$ at most twice. Suppose that the two functions meet less than twice. Then we have $G^{-1}(F(x))-x>0$ for all $x\in\mathbb{R}$ except at most one point x' which implies $G^{-1}(u)>F^{-1}(u)$ for all $0<u<1$ except at most one point u'. However, this implies that

$$\mu(G)=\int_0^1 G^{-1}(u)\,\mathrm{d}u > \int_0^1 F^{-1}(u)\,\mathrm{d}u=\mu(F)$$

which contradicts the assumption $\mu(F)=\mu(G)$. Hence there exist two points $x_1,x_2\in\mathbb{R}$ such that $G^{-1}(F(x))-x\geq 0$ for all $x<x_1$ or $x>x_2$ and $G^{-1}(F(x))-x\leq 0$ for all $x_1\leq x\leq x_2$. This implies that there exist two points u_1 and u_2 such that $0\leq u_1<u_2\leq 1$ and $G^{-1}(u)-F^{-1}(u)\geq 0$ for all $u\leq u_1$ or $u\geq u_2$ and $G^{-1}(u)-F^{-1}(u)\leq 0$ for all $u_1\leq u\leq u_2$. □

To check that η_3^* satisfies the first condition of Definition 1.3.c, let $G = F_{a,b}$. We first assume that $a > 0$. Then we have

$$\begin{aligned}
\eta_3(G) &= \int_0^1 G^{-1}(u)\left\{\Phi^{-1}(u)^2 - 1\right\}\,\mathrm{d}u \\
&= \int_0^1 \left\{aF^{-1}(u) + b\right\}\left\{\Phi^{-1}(u)^2 - 1\right\}\,\mathrm{d}u \\
&= a\int_0^1 F^{-1}(u)\left\{\Phi^{-1}(u)^2 - 1\right\}\,\mathrm{d}u + b\int_0^1 \left\{\Phi^{-1}(u)^2 - 1\right\}\,\mathrm{d}u \\
&= a\int_0^1 F^{-1}(u)\left\{\Phi^{-1}(u)^2 - 1\right\}\,\mathrm{d}u \\
&= a\eta_3(F). \qquad (7.5)
\end{aligned}$$

Hence, we have $\eta_3(G) = a\eta_3(F)$ when $a > 0$. Following similar steps of derivation, it can be shown that $\eta_3(G) = a\eta_3(F)$ when $a < 0$. Combining these two results and $\eta_2(G) = |a|\eta_2(F)$, we obtain the desired result $\eta_3^*(G) = \mathrm{sign}(a)\eta_3^*(F)$.

To check the second condition, we first assume that $\eta_1(F) = \eta_1(G) = 0$ and $\eta_2(F) = \eta_2(G) = 1$. Note that

$$\begin{aligned}
\eta_3(G) - \eta_3(F) &= \int_0^1 \left\{G^{-1}(u) - F^{-1}(u)\right\}\left\{\Phi^{-1}(u)^2 - 1\right\}\,\mathrm{d}u \\
&= \int_{-\infty}^{\infty} \left\{G^{-1}(\Phi(x)) - F^{-1}(\Phi(x))\right\}\phi(x)\left\{x^2 - 1\right\}\,\mathrm{d}x
\end{aligned}$$

By Lemma 7.5 and the monotonic increasing property of $\Phi(x)$, we know that there exist two points $x_1 < x_2$ such that $G^{-1}(\Phi(x)) - F^{-1}(\Phi(x)) \geq 0$ for $x \leq x_1$ or $x \geq x_2$ and $G^{-1}(\Phi(x)) - F^{-1}(\Phi(x)) \leq 0$ for $x_1 \leq x \leq x_2$. Now consider a polynomial $K(x|a,b) = H_2(x) + aH_1(x) + bH_0(x) = x^2 + ax + b - 1$ for $a \neq 0, b \in \mathbb{R}$. By equating $x^2 + ax + b - 1 = (x - x_1)(x - x_2)$ for all $x \in \mathbb{R}$, we can find two constants $a_{F,G}$ and $b_{F,G}$ such that $K(x|a_{F,G}, b_{F,G}) \geq 0$ for $x < x_1$ or $x > x_2$ and $K(x|a_{F,G}, b_{F,G}) \leq 0$ for $x_1 \leq x \leq x_2$.

Now we have

$$\begin{aligned}
0 &\leq \int_{-\infty}^{\infty} \phi(x) \left\{ G^{-1}(\Phi(x)) - F^{-1}(\Phi(x)) \right\} K(x|a_{F,G}, b_{F,G}) \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \phi(x) \left\{ G^{-1}(\Phi(x)) - F^{-1}(\Phi(x)) \right\} \left\{ H_2(x) + a_{F,G} H_1(x) + b_{F,G} H_0(x) \right\} \, \mathrm{d}x \\
&= \left\{ \eta_3(G) - \eta_3(F) \right\} + a_{F,G} \left\{ \eta_2(G) - \eta_2(F) \right\} + b_{F,G} \left\{ \eta_1(G) - \eta_1(F) \right\} \\
&= \eta_3(G) - \eta_3(F)
\end{aligned}$$

where the first inequality holds since the two functions $G^{-1}(\Phi(x)) - F^{-1}(\Phi(x))$ and $K(x|a_{F,G}, b_{F,G})$ have the same sign for all $x \in \mathbb{R}$, and the last equality comes from the assumption that $\eta_1(F) = \eta_1(G) = 0$ and $\eta_2(F) = \eta_2(G) = 1$.

Now assume that two distributions F and G have arbitrary first and second HL-moment coefficients. Then we have

$$\eta_3^*(F) = \eta_3 \left(F_{1/\eta_2(F), -\eta_1(F)/\eta_2(F)} \right) \leq \eta_3 \left(G_{1/\eta_2(G), -\eta_1(G)/\eta_2(G)} \right) = \eta_3^*(G)$$

where the first and last equality holds since we showed that the first condition of Oja's criterion is satisfied.

To prove the theorem, we need the following lemmas.

Lemma 7.6. If F is a symmetric distribution, then $\eta_3(F) = 0$.

Proof . Since F^{-1} is symmetric with respect to the point $(1/2, m(F))$, we have

$$F^{-1}(u) - m(F) = m(F) - F^{-1}(1-u) \tag{7.6}$$

for all $1/2 \leq u < 1$. We have

$$\begin{aligned}
\eta_3(F) &= \int_0^{1/2} F^{-1}(u) \left\{ \Phi^{-1}(u)^2 - 1 \right\} \, \mathrm{d}u + \int_{1/2}^{1} F^{-1}(u) \left\{ \Phi^{-1}(u)^2 - 1 \right\} \, \mathrm{d}u \\
&= \int_0^{1/2} F^{-1}(u) \left\{ \Phi^{-1}(u)^2 - 1 \right\} \, \mathrm{d}u + \int_0^{1/2} F^{-1}(1-v) \left\{ \Phi^{-1}(v)^2 - 1 \right\} \, \mathrm{d}v \\
&= 2m(F) \int_0^{1/2} \Phi^{-1}(u)^2 - 1 \, \mathrm{d}u \\
&= 0
\end{aligned}$$

where the second equation results from the change of variable $v = 1 - u$ and the second to last equation results from Equation (7.6). □

Lemma 7.7. Let F and G be symmetric distributions with the symmetry points $m(F)$ and $m(G)$ such that $\eta_1(F) = \eta_1(G) = 0$ and $\eta_2(F) = \eta_2(G) = 1$. If $G^{-1} \circ F$ is concave on $\{x|x < 0\}$ and convex on $\{x|x > 0\}$, then there exists two points $0 < u_1 < 1/2 < u_2 < 1$ such that

$$G^{-1}(u) - F^{-1}(u) \leq 0 \quad \text{for} \quad 0 < u \leq u_1 \text{ and } 1/2 < u \leq u_2,$$
$$G^{-1}(u) - F^{-1}(u) \geq 0 \quad \text{for} \quad u_1 < u \leq 1/2 \text{ and } u_2 < u \leq 1.$$

Proof . By the convexity assumption on $G^{-1} \circ F$, this function meets the function $y = x$ either once at $x = 0$ or three times at $x = x_1, 0, x_2$ such that $x_1 < x_2$ on the real line $\mathbb{R}$. Suppose that these two functions meet each other once. Then we have $G^{-1}(F(x)) - x < 0$ for $x < 0$ and $G^{-1}(F(x)) - x > 0$ for $x > 0$ which implies that

$$G^{-1}(u) - F^{-1}(u) < 0 \text{ for } u < \frac{1}{2} \text{ and } G^{-1}(u) - F^{-1}(u) > 0 \text{ for } u > \frac{1}{2} \tag{7.7}$$

since we assumed that $m(F) = 0$. We have

$$\begin{aligned}
\eta_2(G) &= \int_0^1 G^{-1}(u)\Phi^{-1}(u)\,\mathrm{d}u \\
&= \int_0^{1/2} G^{-1}(u)\Phi^{-1}(u)\,\mathrm{d}u + \int_{1/2}^1 G^{-1}(u)\Phi^{-1}(u)\,\mathrm{d}u \\
&> \int_0^{1/2} F^{-1}(u)\Phi^{-1}(u)\,\mathrm{d}u + \int_{1/2}^1 F^{-1}(u)\Phi^{-1}(u)\,\mathrm{d}u \\
&= \eta_2(F)
\end{aligned}$$

where the strict inequality holds owing to Equation (7.7). This contradicts the assumption $\eta_2(F) = \eta_2(G)$. Hence, $G^{-1}(F(x)) - x \leq 0$ for $x < x_1, 0 < x < x_2$ and $G^{-1}(F(x)) - x \geq 0$ for $x_1 < x < 0, x > x_2$. This indicates that there exist two points u_1, u_2 such that $0 < u_1 < 0 < u_2 < 1$ and $G^{-1}(u) - F^{-1}(u) \leq 0$ for $0 < u \leq u_1, 1/2 < u \leq u_2$ and $G^{-1}(u) - F^{-1}(u) \geq 0$ for $u_1 < u \leq 1/2, u_2 < u < 1$. □

To check whether η_4^* satisfies the first condition of Oja's criterion (Definition 1.3.d), we let $G = F_{a,b}$. First, assume that $a < 0$. Then we have

$$\begin{aligned}
\eta_4(G) &= \int_0^1 G^{-1}(u)\left\{\Phi^{-1}(u)^3 - 3\Phi^{-1}(u)\right\}\,\mathrm{d}u \\
&= \int_0^1 \left\{aF^{-1}(1-u)+b\right\}\left\{\Phi^{-1}(u)^3 - 3\Phi^{-1}(u)\right\}\,\mathrm{d}u \\
&= a\int_0^1 F^{-1}(1-u)\left\{\Phi^{-1}(u)^3 - 3\Phi^{-1}(u)\right\}\,\mathrm{d}u + b\int_0^1 \Phi^{-1}(u)^3 - 3\Phi^{-1}(u)\,\mathrm{d}u \\
&= -a\int_0^1 F^{-1}(u)\left\{\Phi^{-1}(u)^3 - 3\Phi^{-1}(u)\right\}\,\mathrm{d}u \\
&= -a\eta_4(F).
\end{aligned}$$

The case when $a > 0$ can be derived in a similar and easier way yielding $\eta_4(G) = a\eta_4(F)$. Combining these two results and $\eta_2(G) = |a|\eta_2(F)$, we obtain the desired result $\eta_4^*(G) = \eta_4^*(F)$.

To check the second condition, we first assume that $\eta_1(F) = \eta_1(G) = 0$ and $\eta_2(F) = \eta_2(G) = 1$. Since we have assumed that F and G are symmetric distributions, we have $\eta_3(F) = \eta_3(G) = 0$ by Lemma 7.6. Note that

$$\begin{aligned}
\eta_4(G) - \eta_4(F) &= \int_0^1 \left\{G^{-1}(u) - F^{-1}(u)\right\}\left\{\Phi^{-1}(u)^3 - 3\Phi^{-1}(u)\right\}\,\mathrm{d}u \\
&= \int_{-\infty}^{\infty} \left\{G^{-1}(\Phi(x)) - F^{-1}(\Phi(x))\right\}\phi(x)\left\{x^3 - 3x\right\}\,\mathrm{d}x
\end{aligned}$$

By Lemma 7.7 and the monotonic increasing property of $\Phi(x)$, we know that there exist two points $x_1 < 0 < x_2$ and $G^{-1}(\Phi(x)) - F^{-1}(\Phi(x)) \geq 0$ for $x_1 < x < 0$ and $x > x_2$, and $G^{-1}(\Phi(x)) - F^{-1}(\Phi(x)) \leq 0$ for $x < x_1$ and $0 \leq x \leq x_2$. Now consider a polynomial $K(x|a,b,c) = H_3(x) + aH_2(x) + bH_1(x) + cH_0(x) = x^3 - 3x + a(x^2-1) + bx + c = x^3 + ax^2 + (b-3)x + c - a$ for some a, b and c. By equating

$$x^3 + ax^2 + (b-3)x + c - a = x(x - x_1)(x - x_2) \tag{7.8}$$

for all $x \in \mathbb{R}$, we can find $a_{F,G}, b_{F,G}$ and $c_{F,G}$ such that $K(x|a_{F,G}, b_{F,G}, c_{F,G}) \geq 0$ for $x_1 \leq x \leq 0$ and $x \geq x_2$, $K(x|a_{F,G}, b_{F,G}, c_{F,G}) \leq 0$ for $x \leq x_1$ and $0 \leq x \leq x_2$. Note that

$a_{F,G} = c_{F,G}$ should hold from the equation (7.8). Now we have

$$\begin{aligned}
0 &\leq \int_{-\infty}^{\infty} \phi(x) \left\{ G^{-1}(\Phi(x)) - F^{-1}(\Phi(x)) \right\} K(x|a_{F,G}, b_{F,G}, c_{F,G}) \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \phi(x) \left\{ G^{-1}(\Phi(x)) - F^{-1}(\Phi(x)) \right\} \\
&\qquad \times \left\{ H_3(x) + a_{F,G} H_2(x) + b_{F,G} H_1(x) + c_{F,G} H_0(x) \right\} \, \mathrm{d}x \\
&= \left\{ \eta_4(G) - \eta_4(F) \right\} + a_{F,G} \left\{ \eta_3(G) - \eta_3(F) \right\} + b_{F,G} \left\{ \eta_2(G) - \eta_2(F) \right\} \\
&\quad + c_{F,G} \left\{ \eta_1(G) - \eta_1(F) \right\} \\
&= \eta_4(G) - \eta_4(F).
\end{aligned}$$

Now assume that two distributions F and G have arbitrary first and second HL-moment values. Then we have

$$\eta_4^*(F) = \eta_4 \left(F_{1/\eta_2(F), -\eta_1(F)/\eta_2(F)} \right) \leq \eta_4 \left(G_{1/\eta_2(G), -\eta_1(G)/\eta_2(G)} \right) = \eta_4^*(G)$$

where the first and last equality holds since we showed that the first condition of Oja's criterion is satisfied. ■

Proof of Theorem 2.2. Note that

$$\begin{aligned}
&\frac{1}{r} \sum_{k=0}^{r-2} (-1)^k \binom{r-2}{k} E \left(X_{(r-k):r} - X_{(r-k-1):r} \right) \\
&= \frac{1}{r} \sum_{k=1}^{r-2} \left\{ (-1)^k \binom{r-2}{k} - (-1)^{k-1} \binom{r-2}{k-1} \right\} E X_{(r-k):r} \\
&\quad + \frac{1}{r} E X_{r:r} - \frac{1}{r} (-1)^{r-2} E X_{1:r}. \qquad (7.9)
\end{aligned}$$

We have

$$(-1)^k \binom{r-2}{k} - (-1)^{k-1} \binom{r-2}{k-1} = (-1)^k \binom{r-1}{k}.$$

Substituting this result into (7.9) yields

$$\frac{1}{r}\sum_{k=0}^{r-2}(-1)^k \begin{pmatrix} r-2 \\ k \end{pmatrix} E\left(X_{(r-k):r} - X_{(r-k-1):r}\right) = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k \begin{pmatrix} r-1 \\ k \end{pmatrix} EX_{(r-k):r} = \lambda_r$$

where the last equality results from Equation (2.1) of Hosking (1990). ■

Proof of Theorem 2.3. Following the same steps of the proof of Lemma 7.7, we can show the following lemma.

Lemma 7.8. Let F and G be symmetric distributions with the symmetry points $m(F)$ and $m(G)$ such that $\lambda_1(F) = \lambda_1(G) = 0$ and $\lambda_2(F) = \lambda_2(G) = 1$. If $G^{-1} \circ F$ is concave on $\{x|x < 0\}$ and convex on $\{x|x > 0\}$, then there exist two points $0 < u_1 < 1/2 < u_2 < 1$ such that

$$G^{-1}(u) - F^{-1}(u) \leq 0 \text{ for } 0 < u \leq u_1 \text{ and } 1/2 < u \leq u_2,$$
$$G^{-1}(u) - F^{-1}(u) \geq 0 \text{ for } u_1 < u \leq 1/2 \text{ and } u_2 < u < 1. \qquad \square$$

Since $\rho_{F_0,r}$ is basically defined in terms of expected order statistics (2.4), it has an integral representation similar with that of the L-moments (1.6). Hence, there exists a polynomial $R_{F_0,r}$ with degree r such that

$$\rho_{F_0,r} = \int_{-\infty}^{\infty} x f(x) R_{F_0,r-1}(F(x)) \mathrm{d}x = \int_0^1 F^{-1}(u) R_{F_0,r-1}(u) \mathrm{d}u. \tag{7.10}$$

Since F_0 is symmetric, we have

$$\begin{aligned}
\rho_{F_0,1} &= \lambda_1 \\
\rho_{F_0,2} &= \frac{1}{2\delta_{1,2:2}(F_0)} E\left(X_{2:2} - X_{1:2}\right) = \frac{1}{\delta_{1,2:2}(F_0)}\lambda_2 \\
\rho_{F_0,3} &= \frac{1}{3\delta_{2,3:3}(F_0)} \left\{E\left(X_{3:3} - X_{2:3}\right) - E\left(X_{2:3} - X_{1:3}\right)\right\} = \frac{1}{\delta_{2,3:3}(F_0)}\lambda_3 \\
\rho_{F_0,4} &= \frac{1}{4\delta_{3,4:4}(F_0)} \left\{E\left(X_{4:4} - X_{3:4}\right) - \frac{2\delta_{3,4:4}(F_0)}{\delta_{2,3:4}(F_0)} E\left(X_{3:4} - X_{2:4}\right) + E\left(X_{2:4} - X_{1:4}\right)\right\} \\
&= \frac{1}{4\delta_{3,4:4}(F_0)} \left\{E\left(X_{4:4} - X_{1:4}\right) - \frac{\delta_{2,3:4}(F_0) + 2\delta_{3,4:4}(F_0)}{\delta_{2,3:4}(F_0)} E\left(X_{3:4} - X_{2:4}\right)\right\}. \qquad (7.11)
\end{aligned}$$

The first rescaled L-moment $\rho_{F_0,1}$ is the mean, so it satisfies Oja's criterion for a measure of location by Oja (1981). Since the second and third rescaled L-moments $\rho_{F_0,2}$ and $\rho_{F_0,3}$ are constant multiples of the second and third L-moments respectively, those two functionals satisfy Oja's criterion for a measure of scale and skewness respectively by Hosking (1989). For the fourth rescaled L-moment, we first define $\alpha = \{\delta_{2,3:4}(F_0) + 2\delta_{3,4:4}(F_0)\}/\delta_{2,3:4}(F_0)$ and let

$$\rho_{\alpha,4} = \frac{1}{4}\{E(X_{4:4} - X_{1:4}) - \alpha E(X_{3:4} - X_{2:4})\}, \tag{7.12}$$

then show that $\rho^*_{\alpha,4} = \rho_{\alpha,4}/\lambda_2$ satisfies Oja's criterion for a measure of kurtosis.

To check whether $\rho^*_{\alpha,4}$ satisfies the first condition of Oja's criterion (Definition 1.3.d), we let $G = F_{a,b}$ with $a < 0$. Then we have

$$\begin{aligned}
\rho^*_{\alpha,4}(G) &= \frac{EY_{4:4} - \alpha EY_{3:4} + \alpha EY_{2:4} - EY_{1:4}}{2(EY_{2:2} - EY_{1:2})} \\
&= \frac{aEX_{1:4} - a\alpha EX_{2:4} + a\alpha EX_{3:4} - aEX_{4:4}}{2(aEX_{1:2} - aEX_{2:2})} \\
&= \frac{EX_{4:4} - \alpha EX_{3:4} + \alpha EX_{2:4} - EX_{1:4}}{2(EX_{2:2} - EX_{1:2})} \\
&= \rho^*_{\alpha,4}(F).
\end{aligned}$$

The case when $a > 0$ can be derived in a similar and easier way yielding $\rho^*_{\alpha,4}(G) = \rho^*_{\alpha,4}(F)$.

To check the second condition, we first assume that $\lambda_1(F) = \lambda_1(G) = 0$ and $\lambda_2(F) = \lambda_2(G) = 1$. Since we assumed that F and G are symmetric distributions, we have $\lambda_3(F) = \lambda_3(G) = 0$ from Hosking (1990). From (7.12) we have

$$\begin{aligned}
\rho_{\alpha,4} &= \int_0^1 F^{-1}(u)\left\{u^3 - 3\alpha u^2(1-u) + 3\alpha u(1-u)^2 - (1-u)^3\right\} \mathrm{d}u \\
&= \int_0^1 F^{-1}(u)\left\{(6\alpha+2)u^3 - 3(3\alpha+1)u^2 + (3\alpha+3)u - 1\right\} \mathrm{d}u.
\end{aligned}$$

We let $R_{\alpha,3}(u) = (6\alpha+2)u^3 - 3(3\alpha+1)u^2 + (3\alpha+3)u - 1$. By Lemma 7.8, we know that there exist two points $0 < u_1 < 1/2 < u_2 < 1$ such that $G^{-1}(u) - F^{-1}(u) \le 0$ for $0 < u \le u_1$ and $1/2 < u \le u_2$, and $G^{-1}(u) - F^{-1}(u) \ge 0$ for $u_1 < u \le 1/2$ and $u_2 < u < 1$. Now consider a polynomial $K(u|a,b,c) = R_{\alpha,3}(u) + aP^*_2(u) + bP^*_1(u) + cP^*_0(u)$ for $a, b, c \in \mathbb{R}$.

Then there exist $a_{F,G}, b_{F,G}$ and $c_{F,G}$ such that

$$K\left(u|a_{F,G}, b_{F,G}, c_{F,G}\right) = (6\alpha + 2)\left(u - \frac{1}{2}\right)(u - u_1)(u - u_2)$$

for all $u \in (0, 1)$. Then we have

$$\begin{aligned} 0 &\leq \int_0^1 \left\{G^{-1}(u) - F^{-1}(u)\right\} K\left(u|a_{F,G}, b_{F,G}, c_{F,G}\right) \mathrm{d}u \\ &= \left\{\rho_{\alpha,4}(G) - \rho_{\alpha,4}(F)\right\} + a_{F,G}\left\{\lambda_3(G) - \lambda_3(F)\right\} + b_{F,G}\left\{\lambda_2(G) - \lambda_2(F)\right\} \\ &\quad + c_{F,G}\left\{\lambda_1(G) - \lambda_1(F)\right\} \\ &= \rho_{\alpha,4}(G) - \rho_{\alpha,4}(F). \end{aligned}$$

since the two functions $G^{-1}(u) - F^{-1}(u)$ and $K\left(u|a_{F,G}, b_{F,G}, c_{F,G}\right)$ have the same signs on $(0, 1)$ due to $\alpha > 0$.

Now assume that two distributions F and G have arbitrary λ_1 and λ_2 values. Then we have

$$\rho^*_{\alpha,4}(F) = \rho_{\alpha,4}\left(F_{1/\lambda_2(F), -\lambda_1(F)/\lambda_2(F)}\right) \leq \rho_{\alpha,4}\left(G_{1/\lambda_2(G), -\lambda_1(G)/\lambda_2(G)}\right) = \rho^*_{\alpha,4}(G)$$

where the first and last equality holds since we showed that the first condition of Oja's criterion is satisfied. Note from (7.11) that

$$\rho^*_{F_0,4} = \frac{\delta_{1,2:2}\left(F_0\right)}{\delta_{3,4:4}\left(F_0\right)}\rho^*_{\alpha,4},$$

i.e. $\rho^*_{F_0,4}$ is a constant multiple of $\rho^*_{\alpha,4}$. Hence, the functional $\rho^*_{F_0,4}$ satisfies Oja's criterion for a measure of kurtosis. ■

Proof of Theorem 4.1. The paper Serfling (1980) presented asymptotic distributions of functions of a random vector which asymptotically follows the multivariate Gaussian distribution.

Theorem 7.1(Serfling, 1980). Suppose that $\boldsymbol{X}_n = (X_{n1}, X_{n2}, \cdots, X_{nk})^T$ converges in distribution to $\mathcal{N}(\boldsymbol{\mu}, b_n^2\Sigma)$ where $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_k)$ and $b_n \to 0$ as $n \to \infty$. Let $\boldsymbol{g} : \mathbb{R}^k \to$

$\mathbb{R}^m$ be a vector-valued function such that $\boldsymbol{g}(\boldsymbol{x}) = (g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), \cdots, g_k(\boldsymbol{x}))$ with a nonzero derivative at $\boldsymbol{\mu}$. Then $\boldsymbol{g}(\boldsymbol{X}_n)$ converges in distribution to

$$\mathcal{N}\left(\boldsymbol{g}(\boldsymbol{\mu}), b_n^2 D\Sigma D^T\right) \tag{7.13}$$

where D is a matrix of which the (i,j)-th element is $\mathrm{d}g_i/\mathrm{d}x_j|_{x_j=\mu_j}$. □

The paper Shorack (1972) showed asymptotic Gaussianity of L-statistics in the form (1.4) with some boundedness and smoothness conditions on F and J. We present Example 1 in that paper as the following theorem.

Theorem 7.2(Shorack, 1972). Let $X_1, X_2, \cdots, X_n$ be a random sample generated by the distribution F such that $E|X_1|^k < \infty$ for some positive real number k. Let

$$\tilde{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} J(t_{ni})\, X_{i:n}$$

be the L-statistic of interest. Assume that J, t_{ni} and F satisfy the following conditions;

1. $n \max_{1\le i\le n} \left|t_{ni} - \frac{i}{n}\right| = O(1)$.

2. There exists $a > 0$ such that

$$a\left\{\left(\frac{i}{n}\right) \wedge \left(1 - \frac{i}{n}\right)\right\} \le t_{ni} \le 1 - a\left\{\left(\frac{i}{n}\right) \wedge \left(1 - \frac{i}{n}\right)\right\} \tag{7.14}$$

   for all $1 \le i \le n$.

3. J is continuous except at a finite number of points at which F^{-1} is continuous and there exist $0 < M < \infty$ and $\delta > 0$ such that $|J(t)| \le M\{t(1-t)\}^{-1/2+1/k+\delta}$ for $0 < t < 1$.

4. The derivative of J, say J', exists and is continuous on $(0,1)$, and there exist $0 < M < \infty$ and $\delta > 0$ such that $|J'(t)| \le M\{t(1-t)\}^{-3/2+1/k+\delta}$ for $0 < t < 1$.

Then we have

$$n^{1/2}\left(\tilde{\theta}_n - \mu\right) \to \mathcal{N}\left(0, \sigma^2\right)$$

as $n \to \infty$ where

$$\mu = \int_0^1 J(u)F^{-1}(u)\,\mathrm{d}u, \quad \sigma^2 = \int_0^1\int_0^1 (u \wedge v - uv)J(u)J(v)\,\mathrm{d}F^{-1}(u)\,\mathrm{d}F^{-1}(v). \qquad \square$$

We first show asymptotic Gaussianity of a linear combination of the sample HL-moments $\tilde{\eta}_r$.

Lemma 7.9. Suppose that $E|X_1|^{2+\epsilon} < \infty$ for some $\epsilon > 0$. Let $c_1, c_2, \cdots, c_r \in \mathbb{R}$ be given and let

$$\bar{\eta}_{n,r} = \sum_{k=1}^{r} c_k \tilde{\eta}_{n,k} = \frac{1}{n}\sum_{i=1}^{n} \bar{J}_r\left(\frac{i}{n+1}\right) X_{i:n}$$

where $\bar{J}_r(t) = \sum_{k=1}^{r} c_k H_{k-1}\left(\Phi^{-1}(t)\right)$. Then $\bar{\eta}_{n,r}$ satisfies

$$n^{1/2}\left(\bar{\eta}_{n,r} - \sum_{k=1}^{r} c_k \eta_k\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

as $n \to \infty$ for all $r = 1, 2, \cdots$ where

$$\bar{\sigma}^2 = \int_0^1\int_0^1 (s \wedge t - st)\bar{J}_r(s)\bar{J}_r(t)\,\mathrm{d}F^{-1}(s)\,\mathrm{d}F^{-1}(t).$$

Proof . We show that $\bar{\eta}_{n,r}$ satisfies the conditions of Theorem 7.2. We only show the parts 3 and 4 since the parts 1 and 2 can easily be obtained from algebra.

3. Since $\bar{J}_r = \sum_{k=1}^{r} c_k H_{k-1} \circ \Phi^{-1}$ is a sum of compositions of continuous functions H_{k-1} and Φ^{-1}, J_r is continuous.

Let $K_s : (0,1) \to \mathbb{R}$ be a function for $s = 1, 2, \cdots, r-1$ such that $K_s(t) = \left\{\Phi^{-1}(t)\right\}^s$ for $0 < t < 1$. Since we have $k = 2 + \epsilon$ in Theorem 7.2, there exists $\delta > 0$ such that

$$-\frac{1}{2} + \frac{1}{2+\epsilon} + \delta = -\frac{\epsilon}{2(2+\epsilon)} + \delta < 0.$$

Let $\nu(\delta, \epsilon) = \epsilon/\{2(2+\epsilon)\} - \delta > 0$ It can be seen from (7.2) that

$$\lim_{t\to 0} t^{\nu(\delta,\epsilon)}\,|K_s(t)| = 0 \text{ and } \lim_{t\to 1}(1-t)^{\nu(\delta,\epsilon)}\,|K_s(t)| = 0.$$

This implies that there exist two points $0 < l_r < u_r < 1$ such that

$$\begin{aligned} t^{\nu(\delta,\epsilon)} \left|K_s(t)\right| &\leq (1-t)^{-\nu(\delta,\epsilon)} \text{ for } 0 < t < l_s \\ (1-t)^{\nu(\delta,\epsilon)} \left|K_s(t)\right| &\leq t^{-\nu(\delta,\epsilon)} \text{ for } u_s < t < 1. \end{aligned}$$

Since the function $\left|K_s(t)\right| t^{\nu(\delta,\epsilon)}(1-t)^{\nu(\delta,\epsilon)}$ is continuous, there exists a constant $M_s' < \infty$ such that $\left|K_s(t)\right| t^{\nu(\delta,\epsilon)}(1-t)^{\nu(\delta,\epsilon)} \leq M_s'$ for $l_s \leq t \leq u_s$. Now letting $M_s = \max\{M_s', 1\}$ yields $\left|K_s(t)\right| t^{\nu(\delta,\epsilon)}(1-t)^{\nu(\delta,\epsilon)} \leq M_s$ for $0 < t < 1$. Since there exist $c_{1r}, c_{2r}, \cdots, c_{rr}$ such that

$$\bar{J}_r(t) = \sum_{k=1}^{r} c_k H_k\left(\Phi^{-1}(t)\right) = c_{1r}\Phi^{-1}(t) + c_{2r}\Phi^{-1}(t)^2 + \cdots + c_{rr}\Phi^{-1}(t)^r,$$

we have

$$\begin{aligned} \left|\bar{J}_r(t)\right| &\leq |c_{1r}| \left|\Phi^{-1}(t)\right| + |c_{2r}| \left|\Phi^{-1}(t)\right|^2 + \cdots + |c_{rr}| \left|\Phi^{-1}(t)\right|^r \\ &= |c_{1r}| \left|K_1(t)\right| + |c_{2r}| \left|K_2(t)\right|^2 + \cdots + |c_{rr}| \left|K_r(t)\right|^r \\ &\leq \left(|c_{1r}| M_1 + |c_{2r}| M_2 + \cdots + |c_{rr}| M_r\right) t^{\nu(\delta,\epsilon)}(1-t)^{\nu(\delta,\epsilon)} \end{aligned}$$

for all $0 < t < 1$.

4. Note that

$$\begin{aligned} \frac{1}{x}\phi(x) \geq 1 - \Phi(x) \text{ for } x \geq 0 &\Leftrightarrow \frac{1}{x\left\{1-\Phi(x)\right\}} \geq \frac{1}{\phi(x)} \text{ for } x \geq 0 \\ &\Leftrightarrow \frac{1}{\Phi^{-1}(t)(1-t)} \geq \frac{1}{\phi(\Phi^{-1}(t))} \text{ for } 1/2 < t < 1. \end{aligned}$$

Let K' be the derivative of K. Then we have

$$\begin{aligned} \left|K_s'(t)\right| &= s\left|\Phi^{-1}(t)\right|^{s-1} \cdot \frac{1}{\phi(\Phi^{-1}(t))} \leq s\left|\Phi^{-1}(t)\right|^{s-1} \cdot \frac{1}{(1-t)\left|\Phi^{-1}(t)\right|} \\ &= s\left|\Phi^{-1}(t)\right|^{s-2} \cdot \frac{1}{1-t} \end{aligned}$$

for $1/2 \leq t < 1$ which implies that

$$(1-t)^{1+\nu(\delta,\epsilon)}\left|K_s'(t)\right| \leq s(1-t)^{\nu(\delta,\epsilon)}\left|\Phi^{-1}(t)\right|^{s-2} \cdot \frac{1}{t^{1+\nu(\delta,\epsilon)}} \to 0$$

as $t \to 1$. Similarly, it can be seen that

$$-\frac{1}{x}\phi(x) \geq \Phi(x) \Leftrightarrow -\frac{1}{x\Phi(x)} \geq \frac{1}{\phi(x)} \Leftrightarrow -\frac{1}{t\Phi^{-1}(t)} \geq \frac{1}{\phi(\Phi^{-1}(t))}$$

for $0 < t < 1/2$. Hence we have

$$\left|K_s'(t)\right| \leq s\left|\Phi^{-1}(t)\right|^{s-1} \cdot \frac{1}{t\left|\Phi^{-1}(t)\right|} = s\left|\Phi^{-1}(t)\right|^{s-2} \cdot \frac{1}{t}$$

which implies that

$$t^{1+\nu(\delta,\epsilon)}\left|K_s'(t)\right| \leq st^{\nu(\delta,\epsilon)}\left|\Phi^{-1}(t)\right|^{s-2} \cdot \frac{1}{(1-t)^{1+\nu(\delta,\epsilon)}} \to 0$$

as $t \to 0$. This implies that there exist two points $0 < l_r < u_r < 1$ such that

$$\begin{aligned} t^{1+\nu(\delta,\epsilon)}\left|K_s'(t)\right| &\leq (1-t)^{-1-\nu(\delta,\epsilon)} \text{ for } 0 < t < l_r \\ (1-t)^{1+\nu(\delta,\epsilon)}\left|K_s(t)\right| &\leq t^{-1-\nu(\delta,\epsilon)} \text{ for } u_r < t < 1. \end{aligned}$$

Since the function $|K_s'(t)|\, t^{1+\nu(\delta,\epsilon)}(1-t)^{1+\nu(\delta,\epsilon)}$ is continuous, there exists a constant $M_s' < \infty$ such that $|K_s'(t)|\, t^{1+\nu(\delta,\epsilon)}(1-t)^{1+\nu(\delta,\epsilon)} \leq M_s'$ for $l_r \leq t \leq u_r$. Now letting $M_s = \max\{M_s', 1\}$ yields $|K_s'(t)|\, t^{1+\nu(\delta,\epsilon)}(1-t)^{1+\nu(\delta,\epsilon)} \leq M_s$ for $0 < t < 1$. Since there exist constants $c_{1r}, \cdots, c_{rr}$ such that

$$\bar{J}_r'(t) = c_{1r}K_1'(t) + c_{2r}K_2'(t) + \cdots + c_{rr}K_r'(t)$$

we have

$$\begin{aligned} \left|\bar{J}_r'(t)\right| &\leq |c_{1r}|\left|K_1'(t)\right| + |c_{2r}|\left|K_2'(t)\right|^2 + \cdots + |c_{rr}|\left|K_r'(t)\right|^r \\ &\leq \left(|c_{1r}|\, M_1 + |c_{2r}|\, M_2 + \cdots + |c_{rr}|\, M_r\right) t^{-1-\nu(\delta,\epsilon)}(1-t)^{-1-\nu(\delta,\epsilon)} \end{aligned}$$

for all $0 < t < 1$. □

By Lemma 7.9 and the Cramér-Wold Theorem, we have

$$n^{1/2}\left(\begin{pmatrix}\tilde{\eta}_{n,2}\\ \tilde{\eta}_{n,r_1}\\ \tilde{\eta}_{n,r_2}\end{pmatrix}-\begin{pmatrix}\eta_2\\ \eta_{r_1}\\ \eta_{r_2}\end{pmatrix}\right)\xrightarrow{d}\mathcal{N}\left(0,\begin{pmatrix}\sigma^H_{22} & \sigma^H_{2r_1} & \sigma^H_{2r_2}\\ \sigma^H_{2r_1} & \sigma^H_{r_1r_1} & \sigma^H_{r_1r_2}\\ \sigma^H_{2r_2} & \sigma^H_{r_2r_1} & \sigma^H_{r_2r_2}\end{pmatrix}\right)$$

as $n\to\infty$ where

$$\begin{aligned}
\sigma^H_{r_ir_j} &= \mathrm{Cov}\left(n^{1/2}\tilde{\eta}_{r_i}, n^{1/2}\tilde{\eta}_{r_j}\right)\\
&= \frac{1}{2}\left\{\mathrm{Var}(\tilde{\eta}_{r_i}+\tilde{\eta}_{r_j})-\mathrm{Var}(\tilde{\eta}_{r_i})-\mathrm{Var}(\tilde{\eta}_{r_j})\right\}\\
&= \frac{1}{2}\left\{\int_0^1\int_0^1(u\wedge v-uv)\left\{H_{r_i-1}\left(\Phi^{-1}(u)\right)+H_{r_j-1}\left(\Phi^{-1}(u)\right)\right\}\right.\\
&\qquad\qquad\times\left\{H_{r_i-1}\left(\Phi^{-1}(v)\right)+H_{r_j-1}\left(\Phi^{-1}(v)\right)\right\}\,\mathrm{d}F^{-1}(u)\,\mathrm{d}F^{-1}(v)\\
&\quad-\int_0^1\int_0^1(u\wedge v-uv)H_{r_i-1}\left(\Phi^{-1}(u)\right)H_{r_i-1}\left(\Phi^{-1}(v)\right)\,\mathrm{d}F^{-1}(u)\,\mathrm{d}F^{-1}(v)\\
&\quad\left.-\int_0^1\int_0^1(u\wedge v-uv)H_{r_j-1}\left(\Phi^{-1}(u)\right)H_{r_j-1}\left(\Phi^{-1}(v)\right)\,\mathrm{d}F^{-1}(u)\,\mathrm{d}F^{-1}(v)\right\}\\
&= \int_0^1\int_0^1(u\wedge v-uv)H_{r_i-1}\left(\Phi^{-1}(u)\right)H_{r_j-1}\left(\Phi^{-1}(v)\right)\,\mathrm{d}F^{-1}(u)\,\mathrm{d}F^{-1}(v)
\end{aligned}$$

where $r_i, r_j\in\{2,r_1,r_2\}$. Now let the function $\boldsymbol{g}=(g_1,g_2)^T$ in Theorem 7.1 be such that $g_1(x_1,x_2,x_3)=x_2/x_1$ and $g_2(x_1,x_2,x_3)=x_3/x_1$. Then we have

$$D=\begin{pmatrix}-\frac{\eta_{r_1}}{\eta_2^2} & \frac{1}{\eta_2} & 0\\ -\frac{\eta_{r_2}}{\eta_2^2} & 0 & \frac{1}{\eta_2}\end{pmatrix}$$

Substituting this equation into (7.13) yields

$$n^{1/2}\left(\begin{pmatrix}\tilde{\eta}^*_{n,r_1}\\ \tilde{\eta}^*_{n,r_2}\end{pmatrix}-\begin{pmatrix}\eta^*_{r_1}\\ \eta^*_{r_2}\end{pmatrix}\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\Psi^H\right)$$

where $\Psi^H_{i,j}=\left(\sigma^H_{r_ir_j}-\eta^*_{r_i}\sigma^H_{2r_i}-\eta^*_{r_j}\sigma^H_{2r_j}+\eta^*_{r_i}\eta^*_{r_j}\sigma^H_{22}\right)/\eta_2^2$ for $i,j=1,2$. ■

Proof of Theorem 4.2. Note that

$$\begin{aligned}\sigma_{r_1r_2} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{\Phi(x\wedge y)-\Phi(x)\Phi(y)\}\,H_{r_1}(x)H_{r_2}(y)\,\mathrm{d}x\,\mathrm{d}y\\ &= \int\int_{-\infty<x<y<\infty}\Phi(x)\,\{1-\Phi(y)\}\,H_{r_1}(x)H_{r_2}(y)\,\mathrm{d}x\,\mathrm{d}y\\ &\quad+\int\int_{-\infty<y<x<\infty}\Phi(y)\,\{1-\Phi(x)\}\,H_{r_1}(x)H_{r_2}(y)\,\mathrm{d}x\,\mathrm{d}y. \qquad (7.15)\end{aligned}$$

First, suppose that one of r_1 and r_2 is even and the other is odd. Then we have

$$\begin{aligned}&\int\int_{-\infty<x<y<\infty}\Phi(x)\,\{1-\Phi(y)\}\,H_{r_1}(x)H_{r_2}(y)\,\mathrm{d}x\,\mathrm{d}y\\ =\;&\int\int_{-\infty<t<s<\infty}\Phi(-s)\,\{1-\Phi(-t)\}\,H_{r_1}(-s)H_{r_2}(-t)\,\mathrm{d}s\,\mathrm{d}t\\ =\;&-\int\int_{-\infty<t<s<\infty}\Phi(t)\,\{1-\Phi(s)\}\,H_{r_1}(s)H_{r_2}(t)\,\mathrm{d}s\,\mathrm{d}t. \qquad (7.16)\end{aligned}$$

where the first equality results from the change of variables $s=-x, t=-y$ and the second equality results from Lemma 7.2. From (7.15), we have $\sigma_{r_1r_2}=0$.

Next, suppose both r_1 and r_2 are even numbers. Following the same steps of derivation as Equation (7.16), it can be seen that

$$\begin{aligned}&\int\int_{-\infty<x<y<\infty}\Phi(x)\,\{1-\Phi(y)\}\,H_{r_1}(x)H_{r_2}(y)\,\mathrm{d}x\,\mathrm{d}y\\ =\;&\int\int_{-\infty<y<x<\infty}\Phi(y)\,\{1-\Phi(x)\}\,H_{r_1}(y)H_{r_2}(x)\,\mathrm{d}x\,\mathrm{d}y.\end{aligned}$$

but this time without a negative sign in front of the right hand side expression because both H_{r_1} and H_{r_2} are even. This implies that

$$\begin{aligned}\sigma_{r_1r_2} &= 2\int\int_{-\infty<x<y<\infty}\Phi(x)\,\{1-\Phi(y)\}\,H_{r_1}(x)H_{r_2}(y)\,\mathrm{d}x\,\mathrm{d}y\\ &= 2\int_{-\infty}^{\infty}\int_{-\infty}^{y}\Phi(x)H_{r_1}(x)\,\mathrm{d}x\,\{1-\Phi(y)\}\,H_{r_2}(y)\,\mathrm{d}y. \qquad (7.17)\end{aligned}$$

Performing integration by parts yields

$$\begin{aligned}\int_{-\infty}^{y}\Phi(x)H_{r_1}(x)\,\mathrm{d}x &= \left[\frac{1}{r_1+1}\Phi(x)H_{r_1+1}(x)\right]_{-\infty}^{y}-\frac{1}{r_1+1}\int_{-\infty}^{y}\phi(x)H_{r_1}(x)\,\mathrm{d}x \\ &= \frac{1}{r_1+1}\Phi(y)H_{r_1+1}(y)-\frac{1}{r_1+1}\int_{-\infty}^{y}\phi(x)H_{r_1}(x)\,\mathrm{d}x \qquad (7.18)\end{aligned}$$

where the first equality comes from Lemma 7.2 and the second equality comes from Lemma 7.1. Substituting this equation into (7.17) yields

$$\begin{aligned}\sigma_{r_1r_2} &= 2\int_{-\infty}^{\infty}\left\{\frac{1}{r_1+1}\Phi(y)H_{r_1+1}(y)-\frac{1}{r_1+1}\int_{-\infty}^{y}\phi(x)H_{r_1}(x)\,\mathrm{d}x\right\}\{1-\Phi(y)\}H_{r_2}(y)\,\mathrm{d}y \\ &= \frac{2}{r_1+1}\int_{-\infty}^{\infty}\Phi(y)\{1-\Phi(y)\}H_{r_1+1}(y)H_{r_2}(y)\,\mathrm{d}y \\ &\quad -\frac{2}{r_1+1}\int_{-\infty}^{\infty}\int_{-\infty}^{y}\phi(x)\{1-\Phi(y)\}H_{r_1}(x)H_{r_2}(y)\,\mathrm{d}x\,\mathrm{d}y. \qquad (7.19)\end{aligned}$$

Note that

$$\Phi(-y)\{1-\Phi(-y)\}H_{r_1+1}(-y)H_{r_2}(-y) = -\Phi(y)\{1-\Phi(y)\}H_{r_1+1}(y)H_{r_2}(y)$$

holds for $y \geq 0$ since H_{r_1+1} is an odd function and H_{r_2} is an even function. Hence, we have

$$\frac{2}{r_1+1}\int_{-\infty}^{\infty}\Phi(y)\{1-\Phi(y)\}H_{r_1+1}(y)H_{r_2}(y)\,\mathrm{d}y = 0. \qquad (7.20)$$

Note also that

$$\begin{aligned}&\int_{-\infty}^{\infty}\int_{-\infty}^{y}\phi(x)\{1-\Phi(y)\}H_{r_1}(x)H_{r_2}(y)\,\mathrm{d}x\,\mathrm{d}y \\ =\ &\frac{2}{r_1+1}\int_{-\infty}^{\infty}\int_{x}^{\infty}\{1-\Phi(y)\}H_{r_2}(y)\,\mathrm{d}y\phi(x)H_{r_1+1}(x)\,\mathrm{d}x \\ =\ &-\frac{2}{(r_1+1)(r_2+1)}\int_{-\infty}^{\infty}\phi(x)\{1-\Phi(x)\}H_{r_1+1}(x)H_{r_2+1}(x)\,\mathrm{d}x \\ &+\frac{2}{(r_1+1)(r_2+1)}\int_{-\infty}^{\infty}\int_{x}^{\infty}\phi(x)\phi(y)H_{r_1+1}(x)H_{r_2+1}(y)\,\mathrm{d}y\,\mathrm{d}x\end{aligned}$$

where the second equality comes from the same steps of derivation as Equation (7.18). Note that

$$\begin{aligned}
&\int_{-\infty}^{\infty} \phi(x) \left\{1 - \Phi(x)\right\} H_{r_1+1}(x) H_{r_2+1}(x) \, \mathrm{d}x \\
&quad \int_{-\infty}^{\infty} \phi(x) \left\{\frac{1}{2} - \Phi(x)\right\} H_{r_1+1}(x) H_{r_2+1}(x) \, \mathrm{d}x \\
&quad + \frac{1}{2} \int_{-\infty}^{\infty} \phi(x) H_{r_1+1}(x) H_{r_2+1}(x) \, \mathrm{d}x \\
&= 0
\end{aligned}$$

where the second equality results from the fact that

$$\phi(-x) \left\{\frac{1}{2} - \Phi(-x)\right\} H_{r_1+1}(-x) H_{r_2+1}(-x) = -\phi(x) \left\{\frac{1}{2} - \Phi(x)\right\} H_{r_1+1}(x) H_{r_2+1}(x)$$

which holds since H_{r_1+1}, H_{r_2+1} both are odd functions. Note also that

$$\begin{aligned}
&\int_{-\infty}^{\infty} \int_{x}^{\infty} \phi(x)\phi(y) H_{r_1+1}(x) H_{r_2+1}(y) \, \mathrm{d}y \, \mathrm{d}x \\
= &\int_{-\infty}^{\infty} \int_{-\infty}^{s} \phi(-s)\phi(-t) H_{r_1+1}(-s) H_{r_2+1}(-t) \, \mathrm{d}t \, \mathrm{d}s \\
= &\int_{-\infty}^{\infty} \int_{-\infty}^{s} \phi(s)\phi(t) H_{r_1+1}(s) H_{r_2+1}(t) \, \mathrm{d}t \, \mathrm{d}s
\end{aligned} \tag{7.21}$$

where the first equality comes from the change of variables $s = -x, t = -y$ and the second equality comes from the fact that both H_{r_1+1} and H_{r_2+1} are odd functions. This implies that

$$\begin{aligned}
0 &= \left(\int_{-\infty}^{\infty} \phi(x) H_{r_1+1}(x) \, \mathrm{d}x \right) \left(\int_{-\infty}^{\infty} \phi(y) H_{r_2+1}(y) \, \mathrm{d}y \right) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x)\phi(y) H_{r_1+1}(x) H_{r_2+1}(y) \, \mathrm{d}x \, \mathrm{d}y \\
&= \int_{-\infty}^{\infty} \int_{x}^{\infty} \phi(x)\phi(y) H_{r_1+1}(x) H_{r_2+1}(y) \, \mathrm{d}y \, \mathrm{d}x \\
&\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{x} \phi(x)\phi(y) H_{r_1+1}(x) H_{r_2+1}(y) \, \mathrm{d}y \, \mathrm{d}x \\
&= 2 \int_{-\infty}^{\infty} \int_{x}^{\infty} \phi(x)\phi(y) H_{r_1+1}(x) H_{r_2+1}(y) \, \mathrm{d}y \, \mathrm{d}x
\end{aligned}$$

where the first equality results from the orthogonality of the Hermite polynomials and the last equality results from Equation (7.21). Hence, it can be seen that we have

$$\frac{2}{r_1+1}\int_{-\infty}^{\infty}\int_{-\infty}^{y}\phi(x)\left\{1-\Phi(y)\right\}H_{r_1}(x)H_{r_2}(y)\,\mathrm{d}x\,\mathrm{d}y=0.$$

Being combined with the result (7.20), this indicates that $\sigma_{r_1r_2}=0$. The case when both r_1 and r_2 are odd numbers can be proved in the same manner. Substituting $\sigma_{r_1r_2}=0$ and $\eta^*_{r_1}(\Phi)=\eta^*_{r_2}(\Phi)=0$ yields

$$\Psi^H_{r_1,r_2}=\frac{\sigma^H_{r_1r_2}-\eta^*_{r_1}\sigma^H_{2r_1}-\eta^*_{r_2}\sigma^H_{2r_2}+\eta^*_{r_1}\eta^*_{r_2}\sigma^H_{22}}{\eta^2_2}=0. \qquad \blacksquare$$

BIBLIOGRAPHY

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.

Anderson, T. W. and Darling, D. A. (1954). A test of goodness of fit. *J. Amer. Statist. Assoc.*, 49:765–769.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972). *Robust estimates of location: Survey and advances*. Princeton University Press, Princeton, N.J.

Bickel, P. J. and Lehmann, E. L. (1975). Descriptive statistics for nonparametric models. II. Location. *Ann. Statist.*, 3(5):1045–1069.

Bowley, A. L. (1920). *Elements of Statistics, 4th Edition*. New York: Scribner's.

Bowman, K. O. and Shenton, L. R. (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. *Biometrika*, 62(2):243–250.

Ciriello, G., Gatza, M. L., Beck, A. H., Wilkerson, M. D., Rhie, S. K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., et al. (2015). Comprehensive molecular portraits of invasive lobular breast cancer. *Cell*, 163(2):506–519.

Cornish, E. A. and Fisher, R. A. (1938). Moments and cumulants in the specification of distributions. *Revue de l'Institut international de Statistique*, pages 307–320.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical statistics*. Chapman and Hall, London.

D'Agostino, R. and Pearson, E. S. (1973). Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$. *Biometrika*, 60:613–622.

David, H. A. and Nagaraja, H. N. (2003). *Order statistics*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition.

Elamir, E. A. H. and Seheult, A. H. (2003). Trimmed L-moments. *Comput. Statist. Data Anal.*, 43(3):299–314.

Feng, Q., Hannig, J., and Marron, J. (2015). Non-iterative joint and individual variation explained. *arXiv preprint arXiv:1512.04060*.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188.

Gan, F. and Koehler, K. (1990). Goodness-of-fit tests based on p-p probability plots. *Technometrics*, 32(3):289–303.

Gardiner, J. C. and Sen, P. K. (1979). Asymptotic normality of a variance estimator of a linear combination of a function of order statistics. *Z. Wahrsch. Verw. Gebiete*, 50(2):205–221.

Gleason, J. R. (1993). Understanding elongation: The scale contaminated normal family. *Journal of the American Statistical Association*, 88(421):327–337.

Groeneveld, R. A. (1991). An influence function approach to describing the skewness of a distribution. *The American Statistician*, 45(2):97–102.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (2011). *Robust statistics: the approach based on influence functions*, volume 114. John Wiley & Sons.

Henderson, A. R. (2006). Testing experimental data for univariate normality. *Clinica Chimica Acta*, 366(1):112–129.

Hosking, J. (1986). *The theory of probability weighted moments*. IBM Research Division, TJ Watson Research Center.

Hosking, J. R. M. (1989). *Some theoretical results concerning L-moments*. IBM Thomas J. Watson Research Division.

Hosking, J. R. M. (1990). *L*-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Statist. Soc. Ser. B*, 52(1):105–124.

Hosking, J. R. M. (2007). Some theory and practical uses of trimmed *L*-moments. *J. Statist. Plann. Inference*, 137(9):3024–3039.

Hu, Y.-J., Sun, W., Tzeng, J.-Y., and Perou, C. M. (2015). Proper use of allele-specific expression improves statistical power for cis-eqtl mapping with rna-seq data. *Journal of the American Statistical Association*, 110(511):962–974.

Huber, P. J. and Ronchetti, E. M. (2009). *Robust statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, second edition.

Institute, T. B. (2010). Gene set enrichment analysis. `http://software.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html`.

Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econom. Lett.*, 6(3):255–259.

Johnson, N. L. and Kotz, S. (1970). *Distributions in statistics. Continuous univariate distributions. 1*. Houghton Mifflin Co., Boston, Mass.

Jorge, M. and Boris, I. (1984). Some properties of the tukey g and h family of distributions. *Communications in Statistics-Theory and Methods*, 13(3):353–369.

Kendall, M. and Stuart, A. (1977). *The advanced theory of statistics. Vol. 1*. Macmillan Publishing Co., Inc., New York, fourth edition. Distribution theory.

Li, D., Rao, M. B., and Tomkins, R. J. (2001). The law of the iterated logarithm and central limit theorem for *L*-statistics. *J. Multivariate Anal.*, 78(2):191–217.

Marcinkiewicz, J. (1939). Sur une propriété de la loi de Gauß. *Math. Z.*, 44(1):612–618.

Oja, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scand. J. Statist.*, 8(3):154–168.

Parr, W. C. and Schucany, W. R. (1982). Jackknifing L-statistics with smooth weight functions. *J. Amer. Statist. Assoc.*, 77(379):629–638.

Pearson, K. (1905). "das fehlergesetz und seine verallgemeinerungen durch fechner und pearson." a rejoinder. *Biometrika*, 4(1/2):169–212.

Romão, X., Delgado, R., and Costa, A. (2010). An empirical power comparison of univariate goodness-of-fit tests for normality. *Journal of Statistical Computation and Simulation*, 80(5):545–591.

Ruppert, D. (1987). What is kurtosis? an influence function approach. *The American Statistician*, 41(1):1–5.

Sen, P. K. (1984). Jackknifing L-estimators: affine structure and asymptotics. *Sankhyā Ser. A*, 46(2):207–218.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality: Complete samples. *Biometrika*, 52:591–611.

Shapiro, S. S., Wilk, M. B., and Chen, H. J. (1968). A comparative study of various tests for normality. *J. Amer. Statist. Assoc*, 63:1343–1372.

Shorack, G. R. (1972). Functions of order statistics. *Ann. Math. Statist.*, 43:412–427.

Shorack, G. R. and Wellner, J. A. (2009). *Empirical processes with applications to statistics*, volume 59 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Reprint of the 1986 original [MR0838963].

Staudte, R. G. and Sheather, S. J. (2011). *Robust estimation and testing*, volume 918. John Wiley & Sons.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550.

Szegö, G. (1959). *Orthogonal polynomials*. American Mathematical Society Colloquium Publications, Vol. 23. Revised ed. American Mathematical Society, Providence, R.I.