Journal of the Korean Data & Information Science Society
2014, **25**(1), 237–244

http://dx.doi.org/10.7465/jkdi.2014.25.1.237
한국데이터정보과학회지

Some applications for the difference of two CDFs

Chong Sun Hong[1] · Yun Hwan Son[2]

[12]Department of Statistics, Sungkyunkwan University

Received 13 December 2013, revised 31 December 2013, accepted 6 January 2014

Abstract

It is known that the difference in the length between two location parameters of two random variables is equivalent to the difference in the area between two cumulative distribution functions. In this paper, we suggest two applications by using the difference of distribution functions. The first is that the difference of expectations of a certain function of two continuous random variables such as the differences of two kth moments and two moment generating functions could be defined by using the difference between two univariate distribution functions. The other is that the difference in the volume between two empirical bivariate distribution functions is derived. If their covariance is estimated to be zero, the difference in the volume between two empirical bivariate distribution functions could be defined as the difference in two certain areas.

Keywords: Area, bivariate, empirical, length, moment generating function, volume.

1. Introduction

Let us consider two continuous random variables, X_1 and X_2, and their corresponding cumulative distribution functions (CDFs), $F_1(\cdot)$ and $F_2(\cdot)$, respectively. It is known that the difference in the length between two location parameters of two random variables is equivalent to the difference in the area (gap) between two CDFs, that is, $E(X_2) - E(X_1) = \int [F_1(x) - F_2(x)]dx$ (Holland, 2002; Hong, 2013). And Hong (2013) showed that the difference in the area between the empirical two CDFs is the same as the difference value of the two sample means.

In this paper, we extend to

$$\int h'(x)[F_1(x) - F_2(x)]dx, \tag{1.1}$$

where $h'(x)$ is a derivative of a certain function $h(x)$ with respect to x. It is found that the integration in (1.1) could be defined as the difference between two expectation of some function $h(\cdot)$, that is $E[h(X_2)] - E[h(X_1)]$. For example, if a function $h(x)$ is X^k or e^{tX} for integer k and some appropriate interval of t, then integration in (1.1) turns to be the differences of two k^{th} moments or moment generating functions. These will be showed in Section 2 with some distribution examples.

[1] Corresponding author: Professor, Department of Statistics, Sungkyunkwan University, Seoul 110-740, Korea. E-mail: cshong@skku.edu
[2] Graduate student, Department of Statistics, Sungkyunkwan University, Seoul 110-740, Korea.

In Section 3, we consider two bivariate random variables (X_1, Y_1) and (X_2, Y_2), and their corresponding bivariate CDFs, $F_1(\cdot,\cdot)$ and $F_2(\cdot,\cdot)$, respectively. The difference in the volume between two empirical bivariate CDFs will be defined and explored. And the difference in the volume between two bivariate CDFs is also discussed in Section 3. Section 4 provides the conclusion.

2. The difference between expectations of $h(X_1)$ and $h(X_2)$

The difference of two expectations of X_1 and X_2 can be extended to that of $h(X_1)$ and $h(X_2)$ for a certain function $h(X)$ by using the difference of two CDFs.

Theorem 2.1 The difference between two expectations of a continuous and differentiable function $h(X)$ defined on real number is expressed as

$$E[h(X_2)] - E[h(X_1)] = \int_{-\infty}^{\infty} h'(x)[F_1(x) - F_2(x)]dx, \tag{2.1}$$

where $h'(x)$ is a derivative of $h(x)$.

Proof: Note that $E[h(X)] = h(0) + \int_0^\infty h'(x)[1 - F(x)]dx$, where a continuous function $h(x)$ is defined on $x \in (0, \infty)$ with the derivative $h'(x)$ (Hong, 2011, p. 81). Hence for a continuous function $h(x)$ on $x \in (-\infty, \infty)$, the RHS in (2.1) is obtained with ease. □

First, a function $h(X)$ is regarded as X^k for $k = 1, 2, \ldots$. Then the expectation of $h(X)$ turns to be the k^{th} moment of the random variable X.

Corollary 2.1 The difference between two k^{th} moments is expressed as

$$E[X_2^k] - E[X_1^k] = \int_{-\infty}^{\infty} kx^{k-1}[F_1(x) - F_2(x)]dx. \tag{2.2}$$

Proof: It is obtained that

$$\begin{aligned}
E[X_1^k] &= \int_{-\infty}^{\infty} x^k f_1(x)dx \\
&= \int_0^\infty \int_0^x ku^{k-1} f_1(x)dudx - \int_{-\infty}^0 \int_x^0 ku^{k-1} f_1(x)dudx \\
&= \int_0^\infty ku^{k-1} \int_u^\infty f_1(x)dxdu - \int_{-\infty}^0 ku^{k-1} \int_{-\infty}^u f_1(x)dxdu \\
&= \int_0^\infty ku^{k-1}[1 - F_1(u)]du - \int_0^\infty (-1)^{k-1} ku^{k-1} F_1(-u)du \\
&= \int_0^\infty kx^{k-1}[1 - F_1(x) + (-1)^k F_1(-x)]dx.
\end{aligned}$$

Therefore $E[X_2^k] - E[X_1^k]$ could be derived

$$\int_0^\infty kx^{k-1}[1 - F_1(x) + (-1)^k F_1(-x)]dx - \int_0^\infty kx^{k-1}[1 - F_2(x) + (-1)^k F_2(-x)]dx,$$

which is the same as the RHS in (2.2). □

We illustrate some examples of the difference between k^{th} moments of two random variables which follow normal, lognormal and chi-squared distributions. But these examples are not based on the proof process of Corollary 2.1 but obtained by using the exchange the order of the double integral.

Example 2.1 Consider the difference between 2^{nd} moments of two normal distributions such as $F_1(x) \equiv \Phi_1(x; \mu_1, \sigma_1)$ and $F_2(x) \equiv \Phi_2(x; \mu_2, \sigma_2)$. Then, the difference is

$$\begin{aligned} E[X_2^2] - E[X_1^2] &= \int_{-\infty}^{\infty} 2x[\Phi_1(x) - \Phi_2(x)]dx = \int_{-\infty}^{\infty} 2x \int_{(x-\mu_2)/\sigma_2}^{(x-\mu_1)/\sigma_1} \phi(z)dzdx \\ &= \int_{-\infty}^{\infty} \int_{(\mu_1+\sigma_1 z)}^{(\mu_2+\sigma_2 z)} 2xdx\phi(z)dz = (\mu_2^2 - \mu_1^2) + (\sigma_2^2 - \sigma_1^2). \end{aligned}$$

Example 2.2 For two log-normal distributions such as $F_1(x) \equiv \Phi((\ln x - \mu_1)/\sigma_1)$ and $F_2(x) \equiv \Phi((\ln x - \mu_2)/\sigma_2)$, the difference of two k^{th} moments is

$$\begin{aligned} E[X_2^k] - E[X_1^k] &= \int_0^{\infty} kx^{k-1}[\Phi(\frac{\ln x - \mu_1}{\sigma_1}) - \Phi(\frac{\ln x - \mu_2}{\sigma_2})]dx \\ &= \int_{-\infty}^{\infty} \int_{e^{\mu_1+\sigma_1 \cdot z}}^{e^{\mu_2+\sigma_2 \cdot z}} kx^{k-1}dx\phi(z)dz \\ &= \frac{1}{\sqrt{2\pi}} e^{k\mu_2 + \frac{1}{2}(k\sigma_2)^2} \int_{-\infty}^{\infty} e^{-\frac{(z-k\sigma_2)^2}{2}}dz - \frac{1}{\sqrt{2\pi}} e^{k\mu_1 + \frac{1}{2}(k\sigma_1)^2} \int_{-\infty}^{\infty} e^{-\frac{(z-k\sigma_1)^2}{2}}dz \\ &= exp(k\mu_2 + \frac{1}{2}(k\sigma_2)^2) - exp(k\mu_1 + \frac{1}{2}(k\sigma_1)^2). \end{aligned}$$

Note that $E[X^k] = exp[k\mu + (k\sigma)^2/2]$.

Example 2.3 Consider two gamma distributions such as $F_1(x) \equiv G(x; r, \lambda_1), F_2(x) \equiv G(x; r, \lambda_2)$. Then, the difference of two k^{th} moments is

$$\begin{aligned} E[X_2^k] - E[X_1^k] &= \int_0^{\infty} kx^{k-1}[G_1(x) - G_2(x)]dx = \int_0^{\infty} kx^{k-1} \int_{\frac{x}{\lambda_2}}^{\frac{x}{\lambda_1}} g(s; r, 1)dsdx \\ &= \int_0^{\infty} \int_{\lambda_1 s}^{\lambda_2 s} kx^{k-1}dxg(s; r, 1)ds = \int_0^{\infty} [(\lambda_2 s)^k - (\lambda_1 s)^k] \frac{1}{\Gamma(r)} s^{r-1}e^{-s}ds \\ &= \int_0^{\infty} (\lambda_2^k - \lambda_1^k) \frac{1}{\Gamma(r)} s^{k+r-1}e^{-s}ds \\ &= (\lambda_2^k - \lambda_1^k) \frac{\Gamma(k+r)}{\Gamma(r)} \int_0^{\infty} \frac{1}{\Gamma(k+r)} s^{k+r-1}e^{-s}ds \\ &= (\lambda_2^k - \lambda_1^k) \frac{\Gamma(k+r)}{\Gamma(r)}. \end{aligned}$$

Note that $E[X^k] = \lambda^k \Gamma(k+r)/\Gamma(r)$.

Consider $h(X)$ as e^{tX} for some interval of t. Then the expectation of $h(X)$ turns to be the moment generating function (MGF) of the random variable X. Then the difference of MGFs can be defined as the following.

Corollary 2.2 Suppose that there exist the MGFs of the random variable X_1 and X_2. The difference of MGFs is represented as

$$M_{X_2}(t) - M_{X_1}(t) = \int_{-\infty}^{\infty} te^{tx}[F_1(x) - F_2(x)]dx \tag{2.3}$$

Proof: Note that

$$\begin{aligned} M_{X_1}(t) &= E[e^{tX_1}] = \int_{-\infty}^{\infty} e^{tx} f_1(x)dx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{x} te^{tu} du f_1(x)dx = \int_{-\infty}^{\infty} te^{tu}\int_{u}^{\infty} f_1(x)dxdu \\ &= \int_{-\infty}^{\infty} te^{tx}[1 - F_1(x)]dx. \end{aligned}$$

Hence $M_{X_2}(t) - M_{X_1}(t)$ could be derived

$$\int_{-\infty}^{\infty} te^{tx}[1 - F_2(x)]dx - \int_{-\infty}^{\infty} te^{tx}[1 - F_1(x)]dx,$$

so that the RHS in (2.3) is obtained with ease. □

Examples of the difference between MGFs of two random variables whose distributions are normal and gamma are showed. But these examples are solved by exchanging the order of the double integral.

Example 2.4 Consider two normal distributions which are the same as those in Example 2.1. Then the difference of MGFs of normal distributions is

$$\begin{aligned} M_{X_2}(t) - M_{X_1}(t) &= \int_{-\infty}^{\infty} te^{tx}[\Phi_1(x) - \Phi_2(x)]dx = \int_{-\infty}^{\infty}\int_{\mu_1+\sigma_1\cdot z}^{\mu_2+\sigma_2\cdot z} te^{tx}dx\phi(z)dz \\ &= \int_{-\infty}^{\infty} [e^{t(\mu_2+\sigma_2 z)} - e^{t(\mu_1+\sigma_1 z)}]\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(z+t\sigma_2)^2}{2}+(t\mu_2+\frac{t^2\sigma_2^2}{2})}dz \\ &\quad -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(z+t\sigma_1)^2}{2}+(t\mu_1+\frac{t^2\sigma_1^2}{2})}dz \\ &= exp(t\mu_2 + t^2\sigma_2^2/2) - exp(t\mu_1 + t^2\sigma_1^2/2). \end{aligned}$$

Example 2.5 For the same gamma distributions as those in Example 2.3, the difference of MGFs of gamma distributions is

$$\begin{aligned} M_{X_2}(t) - M_{X_1}(t) &= \int_{0}^{\infty} te^{tx}[G_1(x) - G_2(x)]dx = \int_{0}^{\infty} te^{tx}\int_{\frac{x}{\lambda_2}}^{\frac{x}{\lambda_1}} g(s; r, 1)dsdx \\ &= \int_{0}^{\infty}\int_{\lambda_1 s}^{\lambda_2 s} te^{tx}dxg(s; r, 1)ds = \int_{0}^{\infty} (e^{\lambda_2 ts} - e^{\lambda_1 ts})g(s; r, 1)ds \\ &= \int_{0}^{\infty} \frac{1}{\Gamma(r)}s^{r-1}e^{-(1-\lambda_2 t)s}ds - \int_{0}^{\infty} \frac{1}{\Gamma(r)}s^{r-1}e^{-(1-\lambda_1 t)s}ds \\ &= 1/(1-\lambda_2 t)^r - 1/(1-\lambda_1 t)^r, \end{aligned}$$

for $t < 1/\lambda_i$ $(i = 1, 2)$. Note that the MGF of gamma distribution is $M_X(t) = 1/(1 - \lambda t)^r$.

3. The difference in volume between two bivariate CDFs

Let us extend our study to bivariate cases. Suppose that two independent bivariate random samples, $\{(X_{11}, Y_{11}), \ldots, (X_{1n}, Y_{1n})\}$ and $\{(X_{21}, Y_{21}), \ldots, (X_{2m}, Y_{2m})\}$, of size n and m are collected from bivariate CDFs $F_1(\cdot, \cdot)$ and $F_2(\cdot, \cdot)$, respectively. Let $(x_{(1n)}, y_{(1n)})$ and $(x_{(2m)}, y_{(2m)})$ be the largest order statistic values from each bivariate random samples. And for $i = 1, 2, \ldots, n+m$, $(x_{(pi)}, y_{(pi)})$ is defined as the ith order statistic value of the pooled bivariate random samples $\{(X_{11}, Y_{11}), \ldots, (X_{1n}, Y_{1n}), (X_{21}, Y_{21}), \ldots, (X_{2m}, Y_{2m})\}$. Then the difference in the volume between two empirical bivariate CDFs, $\hat{F}_1(x, y)$ and $\hat{F}_2(x, y)$, is expressed as the following:

$$\sum_{j=1}^{m+n-1} \sum_{i=1}^{m+n-1} [\hat{F}_1(x_{(pi)}, y_{(pj)}) - \hat{F}_2(x_{(pi)}, y_{(pj)})](x_{(p(i+1))} - x_{(pi)})(y_{(p(j+1))} - y_{(pj)}). \tag{3.1}$$

Then the difference defined in (3.1) can be obtained in Theorem 3.1.

Theorem 3.1 If $x_{(1n)} \le x_{(2m)}$ and $y_{(1n)} \le y_{(2m)}$, the difference in the volume between two empirical bivariate CDFs results in

$$\begin{aligned} &x_{(2m)}(\bar{Y}_2 - \bar{Y}_1) + y_{(2m)}(\bar{X}_2 - \bar{X}_1) + \bar{X}_1\ \bar{Y}_1 - \bar{X}_2\ \bar{Y}_2 \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}(x_{1i} - \bar{X}_1)(y_{1i} - \bar{Y}_1) - \frac{1}{m}\sum_{j=1}^{m}(x_{2j} - \bar{X}_2)(y_{2j} - \bar{Y}_2). \end{aligned}$$

Proof: The double summation in (3.1) is represented as the difference between the following two parts

$$\sum_{j=1}^{m+n-1} \sum_{i=1}^{m+n-1} \hat{F}_1(x_{(pi)}, y_{(pj)})(x_{(p(i+1))} - x_{(pi)})(y_{(p(j+1))} - y_{(pj)}) \ \ \textit{and}$$
$$\sum_{j=1}^{m+n-1} \sum_{i=1}^{m+n-1} \hat{F}_2(x_{(pi)}, y_{(pj)})(x_{(p(i+1))} - x_{(pi)})(y_{(p(j+1))} - y_{(pj)}).$$

When $x_{(1n)} \le x_{(2m)}$ and $y_{(1n)} \le y_{(2m)}$, the first part, $\sum_{j=1}^{m+n-1} \sum_{i=1}^{m+n-1} \hat{F}_1(x_{(pi)}, y_{(pj)})$ $(x_{(p(i+1))} - x_{(pi)})(y_{(p(j+1))} - y_{(pj)})$, is the volume under the sample distribution $\hat{F}_1(\cdot, \cdot)$ corresponding to $\{(x_{(11)}, y_{(11)}), \ldots, (x_{(1n)}, y_{(1n)}), (x_{(2m)}, y_{(2m)})\}$ among the pooled random samples, where $F_1(x_{(11)}, y_{(11)}) = 1/n$ and $F_1(x_{(1n)}, y_{(1n)}) = F_1(x_{(2m)}, y_{(2m)}) = 1$. Hence,

$$\begin{aligned} &\sum_{i=1}^{n-1} \frac{i}{n}[(x_{(2m)} - x_{(i)})(y_{(i+1)} - y_{(i)}) + (y_{(2m)} - y_{(i)})(x_{(i+1)} - x_{(i)}) \\ &-(x_{(i+1)} - x_{(i)})(y_{(i+1)} - y_{(i)})] + [(x_{(2m)} - x_{(1n)})(y_{(2m)} - y_{(1n)})] \\ = \ &-x_{(2m)}\bar{Y}_1 - y_{(2m)}\bar{X}_1 + \bar{X}_1\ \bar{Y}_1 + \frac{1}{n}\sum_{i=1}^{n}(x_{1i} - \bar{X}_1)(y_{1i} - \bar{Y}_1) + x_{(2m)}y_{(2m)}, \end{aligned} \tag{3.2}$$

where $\sum_{i=1}^{n} x_{(1i)}y_{(1i)} = \sum_{i=1}^{n}(x_{1i} - \bar{X}_1)(y_{1i} - \bar{Y}_1) + n\bar{X}_1\ \bar{Y}_1$, and where $\bar{X}_1 = \sum_{i=1}^{n} x_{(1i)}/n$, $\bar{Y}_1 = \sum_{i=1}^{n} y_{(1i)}/n$.

The second part, $\sum_{j=1}^{m+n-1}\sum_{i=1}^{m+n-1}\hat{F}_2(x_{(pi)}, y_{(pj)})(x_{p((i+1))} - x_{(pi)})(y_{(p(j+1))} - y_{(pj)})$, is the volume under the sample distribution $\hat{F}_2(\cdot,\cdot)$ corresponding to $\{(x_{(21)}, y_{(21)}), \ldots, (x_{(2m)}, y_{(2m)})\}$, where $F_2(x_{(21)}, y_{(21)}) = 1/m$ and $F_2(x_{(2m)}, y_{(2m)}) = 1$. Hence,

$$\begin{aligned} &\sum_{i=1}^{m-1}\frac{i}{m}[(x_{(2m)} - x_{(i)})(y_{(i+1)} - y_{(i)}) + (y_{(2m)} - y_{(i)})(x_{(i+1)} - x_{(i)}) \\ &-(x_{(i+1)} - x_{(i)})(y_{(i+1)} - y_{(i)})] + [(x_{(2m)} - x_{(1n)})(y_{(2m)} - y_{(1n)})] \\ = \ &-x_{(2m)}\bar{Y}_2 - y_{(2m)}\bar{X}_2 + \bar{X}_2\ \bar{Y}_2 + \frac{1}{m}\sum_{j=1}^{m}(x_{2j} - \bar{X}_2)(y_{2j} - \bar{Y}_2) + x_{(2m)}y_{(2m)}. \end{aligned} \tag{3.3}$$

Therefore, the difference between (3.2) and (3.3) leads to the results of Theorem 3.1. □

Alternatively, the result of Theorem 3.1 could be formulated as

$$x_{(2m)}(\bar{Y}_2 - \bar{Y}_1) + y_{(2m)}(\bar{X}_2 - \bar{X}_1) + \bar{X}_1\ \bar{Y}_1 - \bar{X}_2\ \bar{Y}_2 + \hat{Cov}(X_1, Y_1) - \hat{Cov}(X_2, Y_2). \tag{3.4}$$

Theorem 3.1 is defined under the condition $x_{(1n)} \leq x_{(2m)}$ and $y_{(1n)} \leq y_{(2m)}$. For other three types of conditions, Theorem 3.1 could be represented in Corollary 3.1.

Corollary 3.1 The difference in the volume between two empirical bivariate CDFs in (3.1) could be formulated for three types of conditions:

$$x_{(2m)}(\bar{Y}_2 - \bar{Y}_1) + y_{(1n)}(\bar{X}_2 - \bar{X}_1) + \bar{X}_1\ \bar{Y}_1 - \bar{X}_2\ \bar{Y}_2 + \hat{Cov}(X_1, Y_1) - \hat{Cov}(X_2, Y_2), \quad \text{for } x_{(1n)} \leq x_{(2m)} \text{ and } y_{(1n)} > y_{(2m)},$$

$$x_{(1n)}(\bar{Y}_2 - \bar{Y}_1) + y_{(2m)}(\bar{X}_2 - \bar{X}_1) + \bar{X}_1\ \bar{Y}_1 - \bar{X}_2\ \bar{Y}_2 + \hat{Cov}(X_1, Y_1) - \hat{Cov}(X_2, Y_2), \quad \text{for } x_{(1n)} > x_{(2m)} \text{ and } y_{(1n)} \leq y_{(2m)},$$

$$x_{(1n)}(\bar{Y}_2 - \bar{Y}_1) + y_{(1n)}(\bar{X}_2 - \bar{X}_1) + \bar{X}_1\ \bar{Y}_1 - \bar{X}_2\ \bar{Y}_2 + \hat{Cov}(X_1, Y_1) - \hat{Cov}(X_2, Y_2), \quad \text{for } x_{(1n)} > x_{(2m)} \text{ and } y_{(1n)} > y_{(2m)}.$$

□

If estimates of the covariance of two random variables are zero, i.e., $\hat{Cov}(X,\ Y) = 0$ in (3.4), then the difference between the two independent empirical bivariate CDFs in (3.1) is summarized briefly as

$$\begin{aligned} &x_{(2m)}(\bar{Y}_2 - \bar{Y}_1) + y_{(2m)}(\bar{X}_2 - \bar{X}_1) + \bar{X}_1\ \bar{Y}_1 - \bar{X}_2\ \bar{Y}_2 \\ &= (x_{(2m)} - \bar{X}_1)(y_{(2m)} - \bar{Y}_1) - (x_{(2m)} - \bar{X}_2)(y_{(2m)} - \bar{Y}_2). \end{aligned}$$

When we assume that $x_{(1n)} < x_{(2m)}$, $y_{(1n)} < y_{(2m)}$, and $\bar{X}_1 < \bar{X}_2$, $\bar{Y}_1 < \bar{Y}_2$, the difference in the volume between the two independent empirical bivariate CDFs in Corollary 3.2 can be interpreted as the difference in the area between rectangles A and B:

$$\begin{aligned} A &= [x_{(2m)} - \bar{X}_1] \times [y_{(2m)} - \bar{Y}_1] \\ B &= [x_{(2m)} - \bar{X}_2] \times [y_{(2m)} - \bar{Y}_2]. \end{aligned}$$

These two rectangles A and B could be drawn in Figure 3.1. And Figure 3.1 show that the difference in the area between rectangles A and B is equivalent to the difference in the volume between the two independent empirical bivariate CDFs. If the random variables are not independent, it is known that the difference in the volume between the two empirical bivariate CDFs is adjusted as the amount that $[\hat{C}ov(X_1, Y_1) - \hat{C}ov(X_2, Y_2)]$ from the difference in the area between rectangles A and B.

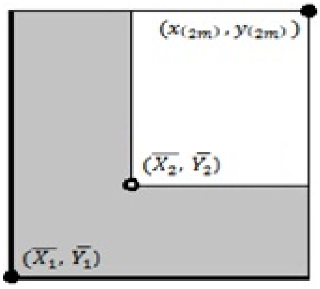

Figure 3.1 The difference of the two empirical CDFs when X, Y are independent

The double integral of the difference between the two bivariate CDFs, $F_1(\cdot, \cdot)$ and $F_2(\cdot, \cdot)$, is derived as follows:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F_1(x, y) - F_2(x, y)] dx dy$$
$$= (\mu_{y_2} - \mu_{y_1}) \int_0^{\infty} dx + (\mu_{x_2} - \mu_{x_1}) \int_0^{\infty} dy + \mu_{x_1}\mu_{y_1} - \mu_{x_2}\mu_{y_2} + Cov(X_1, Y_1) - Cov(X_2, Y_2).$$

Nonetheless, this double integration could not be well defined, since $(\mu_{y_2} - \mu_{y_1}) \int_0^{\infty} dx + (\mu_{x_2} - \mu_{x_1}) \int_0^{\infty} dy$ cannot be obtained as well as that $F_1(x, y) - F_2(x, y)$ does not converge to zero for some x and y.

4. Conclusion

Even though there are many statistical methods by using the difference of two CDFs, most of these methods are not to estimate their corresponding expectations and location parameters. In this paper, two methods using the difference of two CDFs are proposed. One is that the difference of expectations of a certain function of two random variables can be represented in terms of the difference between two corresponding CDFs. With this theory, it is possible to obtain the difference of two kth moments as well as the difference of two MGFs by using the difference of two CDFs. The second is for bivariate CDF. The difference in the volume between two empirical bivariate CDFs based on random samples is defined and derived. This difference could be expressed in terms of sample mean vector, covariance matrix, and the vector of largest order statistics. In particular, when their covariance is estimated to be zero, this difference in the volume is defined as the difference in the area of two rectangles.

References

Holland, P. W. (2002). Two measures of change in the gaps between the CDFs of test-score distributions. *Journal of Educational and Behavioral Statistics*, **27**, 3-17.

Hong, C. S. (2002). *Theory of statistical probability distributions*, 2nd Ed., Parkyoung Sa.

Hong, C. S. (2013). The difference between two distribution functions. *Journal of the Korean Data & Information Science Society*, **24**, 1449-1454.

Mood, A. M., Graybill, F. A. and Boes, D. C. (1974) *Introduction to the theory of statistics*, 3rd Ed, McGraw-Hill, International Edition, Singapore.