S Ü Fen Ed Fak Fen Derg Sayı 25 (2005) 1- 7, KONYA

Çok Değişkenli Veri Analizinde Derinliklere Dayalı Yüzdelikler

İhsan KARABULUT[1]

Özet: Bu çalışmada çok değişkenli dağılımlarda yüzdeliklerin tanımlanmasında önem kazandığı düşünülen çalışmalar özetlenecek ve derinliklere dayalı olarak yapılan yüzdelik tanımlaması ve yüzdelik süreci üzerine durulacaktır. Derinliklere dayalı yüzdelik sürecine ilişkin olarak çok değişkenli normal dağılıma uygunluğun grafiksel olarak değerlendirilebileceği ve güven bantlarının yer aldığı bir uygulama sunulacaktır.

Anahtar Kelimeler: Çok değişkenli dağılıma uygunluk, derinlik fonksiyonu, eliptik dağılımlar, güven bandı, yüzdelik süreci.

Quantiles Based on Depth Functions in Multivariate Data Analysis

Abstract: In this study, some studies, which are thought that imporant for the definition of quantiles for multivariate distributions , are summarized and concentred on the definition of quantiles based on depth functions and quantile processes. It has been presented a graphical evaluation with a confidence band of multivariate goodness of fit for a multivariate normal distribution in relation with the quatile processes based on depth functions.

Key Words: Goodness of fit for a multivariate distribution, depth function, elliptical distributions, confidence band , quantile processes.

Giriş

$X_1, X_2, X_3, ..., X_n$ birbirinden bağımsız ve aynı F dağılım fonksiyonuna sahip kitleden bir örneklem olduğunda F_n örneklem dağılım fonksiyonuna dayalı olarak F'nin $0 < p < 1$ olmak üzere p'inci yüzdeliği

$$\xi_p = \inf\{x : F(x) \geq p\}$$

$\hat{\xi}_{pn} = \min\{x_n : F_n(x_n) \geq p\}$ tahmin edicisi kullanılarak tahmin edilebilir. Sözkonusu örnekleme ilişkin $X_{1:n} < X_{2:n} < X_{3:n} < ... < X_{n:n}$ sıra istatistiklerinin bu tahmin edici ile ilişkisi kaçınılmazdır. Çünkü $\hat{\xi}_{pn}$ tahmin edicisi aşağıdaki gibi ifade edilebilir:

$$\hat{\xi}_{pn} = \begin{cases} X_{np} & ,np \text{ tamsayı} \\ X_{[np]+1} & ,np \text{ diger halde} \end{cases}$$

Yüzdeliklerin uygulamada karşımıza çıkışı ise çok değişik nedenlerle ortaya çıkabilmektedir. Dağılıma uygunluğun görsel değerlendirmesi olarak $Q-Q$ çizitleri, güven aralıkları , parametre tahminleri , hipotez testleri bunlar arasında sayılabilir.

Eğer $X_1, X_2, X_3, ..., X_n$ örnekleminde rasgele değişkenler d boyutlu rasgele vektörler ve F bunların dağılım fonksiyonu olduğunda bu dağılım fonksiyonuna ait p'inci yüzdeliklerin

[1] Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06100 Tandoğan Ankara. E-mail: kbulut@science.ankara.edu.tr

tanımlanmasında ise güçlüklerle karşılaşırız. Bu güçlüklerin nedenlerinden birisi de rasgele değişkenler için doğal bir sıralama sonucu sıra istatistikleri tanımlanabiliyorken rasgele vektörler için bu doğallıkta bir sıralamanın sözkonusu olmayışıdır diyebiliriz. Derinlik kavramının çok değişkenli veri analizinde bu ihtiyacı bir ölçüde karşıladığı düşünülmektedir. Son yıllarda uygulama alanı genişleyen derinlik kavramına dayalı olarak çok değişkenli veri analizinde yüzdelikler merkezden–dışa yüzdelik yüzeyleri olarak tanımlanabilmektedir. $I(\alpha_p, D, F)$ derinliklere dayalı p'inci merkezi bölgesini göstermek üzere bu bölgenin sınırı $\partial I(\alpha_p, D, F)$, p'inci yüzdelik yüzeyi olarak adlandırılır. $X_{[r]}$, rasgele vektörler büyükten küçüğe derinlik sıralamasına sokulduğunda r'inci sıra sayısını alan rasgele vektörü göstermek üzere $\partial I(\alpha_p, D, F)$'nin bir tahmini aşağıdaki gibi verilir:

$$I(\alpha_p, D, F_n) = K\{X_{[1]}, X_{[2]}, X_{[3]}, ..., X_{[np]}\}$$

burada, $[np]$ tamsayı ise np olarak, değilse (np'nin tam kısmı+1) olarak yazılmak üzere K, $\{X_{[1]}, X_{[2]}, X_{[3]}, ..., X_{[np]}\}$ kümesini içeren en küçük konveks bölge anlamında kullanılmaktadır.

Bu çalışmada çok değişkenli veri analizinde derinliklere dayalı olarak tanımlanan yüzdelik yüzeyleri ve tanıtılacak, bu alandaki sonuçların çok değişkenli dağılıma uygunluk için uygulamasına bir örnek verilecektir.

Çok Değişkenli Dağılımlarda Yüzdelikler

John H. J. Einmahl ve David M. Mason[1] R^d' de değer alan rasgele vektör X için R^d'de Borel kümeleri $B(R^d)$'nin bir alt kümesi $\mathbf{A}$ anlamında çok değişkenli p. yüzdelik kavramını, bir λ reel değerli fonksiyona bağlı olarak olasılığı p ve daha büyük olan Borel kümesi olarak tanımlamışlardır:

$$U(p) = \inf\{\lambda(A) : P_X(A) \geq p, A \in \mathbf{A}\}$$

burada $0 < p < 1$ ve P_X X'in olasılık ölçüsüdür. Örneklem yüzdelik fonksiyonu da yine $0 < p < 1$ reel sayısı ve örneklem olasılık ölçüsü $B \in B(R^d)$ kümeleri için

$$P_n(B) = \frac{1}{n}\sum_{i=1}^{n} 1_B(X_i)$$

tanımlanmak üzere $\inf \varnothing = \infty$ olduğunu dikkate alarak

$$U_n(p) = \inf\{\lambda(A) : P_n(A) \geq p, A \in \mathbf{A}\}$$

genelleştirilmiş yüzdelikler süreci olarak tanımlanır. Amaca göre $\mathbf{A}$ sınıfı ve λ'nın belirlemesi yapılır. λ için doğal bir seçim R^d'de λ_d Lebesgue ölçüsü olabilir ki bu durumda $U_n(p)$, $\mathbf{A}$ sınıfında eldeki verinin en az p yüzdeliği kadarını bulunduran en küçük kümenin "hacmi" olacaktır.

$d = 1$ durumunda $\mathbf{A} = \{[-\infty, x) : x \in R\}$ ve $\lambda([-\infty, x)) = x$ olarak belirlenirse yukarıdaki tanımlamalar R'deki bildik yüzdelik ve örneklem yüzdelik fonksiyonlarına denk olacaktır.

λ ve $\mathbf{A}$ üzerinde ölçülebilirlik ve rasgele elementler uzayında işlemleri kolaylaştırıcı gerekli varsayımlar yapılarak yukarıda tanımlanan yüzdelikler süreci için limit teoremleri verilmiştir. Uygun bir normalleştirme fonksiyonu $g(p)$ tanımlandığında

$$\beta_n(p) = g(p)\sqrt{n}\left(U_n(p) - U(p)\right)$$

genelleştirilmiş sürecinin dağılımda bir limit sürecine yakınsayacağı gösterilmiştir.

Chen, L.-A. ve Welsh, A.H. [2] İki değişkenli yüzdelikleri iki değişkenli dağılım fonksiyonuna dayalı olarak tanımlamışlardır. Tanımlama R^d' de de geçerlidir. Rasgele örneğin çekildiği yığının yüzdeliklerini dikkate alan bir tanımlamadır.

İki değişkenli rasgele $X=(X_1,X_2)$ vektörünün aldığı değerlerlerden (a,b) noktası, R^2'yi $A_1=\left\{(x_1,x_2)' : x_2 \ge b\right\}$, $A_2=\left\{(x_1,x_2)' : x_1 \le a, x_2 \le b\right\}$ ve $A_3=\left\{(x_1,x_2)' : x_1 \ge a, x_2 \le b\right\}$ gibi üç parçaya ayırır. Bu durumda (a,b) $\left(P(A_2),P(A_3)\right)$'ıncı iki değişkenli yüzdelik noktası olarak düşünülebilecektir. $F(x_1,x_2)=P(X_1 \le x_1, X_2 \le x_2)$, X'in iki değişkenli ortak dağılım fonksiyonu F_2, X_2'nin marjinal dağılım fonksiyonu olmak üzere tek değişkenli yüzdelik kavramına benzetilerek (p_1,p_2)'inci *Kuzey-Güney* iki değişkenli yüzdelik noktası aşağıdaki gibi tanımlanır:

$p_1,p_2 \ge 0,\quad p=p_1+p_2 \in [0,1]$ olmak üzere (p_1,p_2)'inci *Kuzey-Güney* iki değişkenli yüzdelik noktası bir vektör olup

$$\xi(p_1,p_2)=\left(F_{12}^{-1}(p_1,p_2),\ F_2^{-1}(p_1+p_2)\right)$$

olarak tanımlanır. $0 \le p \le 1$ olmak üzere $\xi(p)=\xi(\tfrac{1}{2}p,\tfrac{1}{2}p)$ p'inci Kuzey-Güney yüzdelik noktası ve özel olarak $\xi(\tfrac{1}{2})$ Kuzey- Güney *medyan (ortanca)* noktası olarak adlandırılır. Bu tanımda

$$F_2^{-1}(p_1+p_2)=\inf\left\{x_2 : F(x_2) \ge p_1+p_2\right\}$$

ve

$$F_{12}^{-1}(p_1,p_2)=\inf\{x_1 : F\left(x_1,F_2^{-1}(p_1+p_2)\right) \ge p_1\}$$

olarak verilmiş olup ikinci bileşen $F_2^{-1}(p_1+p_2)$, X_2'nin $p=p_1+p_2$'ıncı yüzdeliği olup $p_1=1-p_2=p$ verildiğinde de $F_{12}^{-1}(p,1-p)=F_1^{-1}(p)$, X_1'in p'inci yüzdeliği olmaktadır. $F_{12}^{-1}(p_1,p_2)$, X_2'nin marjinal dağılımının p_1+p_2'inci yüzdeliğinin verilmesi koşulu altında $F\left(F_{12}^{-1}(p_1,p_2),F_2^{-1}(p_1+p_2)\right)=p_1$ olacak şekilde bulunan değeridir.

Bir Kuzey-Güney yüzdelik noktasının tanımı bir doğrultu gözetilerek tanımlanmış olup tek başına bir anlam ifade etmekte yetersizdir çünkü keyfi bir doğrultu içindir. En uygun doğrultuyu seçmenin gerekliliğinden hareketle aşağıdaki tanım geliştirilir:

$X=(X_1,X_2)$ vektörünün μ gibi bir konum vektörü ile pozitif tanımlı bir saçılım matrisi Σ'ya sahip olduğunu varsayıldığında $\Sigma=P\Lambda P'$ olacak ortagonal P matrisi bulunabilecektir, burada Λ, $\lambda_1 \le \lambda_2$ özdeğerlerinin ana köşegende yer aldığı köşegen matrisidir. $\Sigma^{\frac{1}{2}}=P\Lambda^{\frac{1}{2}}$ olarak tanımlayıp X vektörünün

$$Y=(\Sigma^{-\frac{1}{2}})'(X-\mu)$$

biçimindeki küreye(diske) dönüşümü gözönüne alındığında

$$\eta(p_1,p_2)=\mu+\Sigma^{\frac{1}{2}}\xi^*(p_1,p_2)$$

iki bileşenli vektörü X 'in dağılımı için bir (p_1,p_2)'ıncı iki değişkenli yüzdelik noktasıdır. Burada $\xi^*(p_1,p_2)$ Y 'nin dağılımı için (p_1,p_2)'ıncı Kuzey-Güney iki değişkenli noktasıdır. $p \in [0,1]$ için $\eta(p)=\eta(p/2,p/2)$ p'ıncı yüzdelik , $\eta(1/2)$'de iki değişkenli ortanca olarak adlandırılır.

Bunlara karşılık gelen örneklem yüzdelikleri de tanımlarda yer alan F,F_{12},F_2 örneklem tahmin edicileri $\hat{F},\hat{F}_{12},\hat{F}_2$ ile değiştirilerek elde edilir. Chen, L.-A. ve Welsh, A.H. [2] çeşitli dağılım varsayımları altında yüzdeliklere ilişkin sonuçlar yanında örneklem tahmin edicileri için de sonuçlar vermişlerdir.

Chaudhuri, P. [3] $X_1,X_2,...,X_n$ reel değerli rasgele değişkenler söz konusu olduğunda $0<p<1$ ve $u=2p-1$ için

$$\sum_{i=1}^{n}\left(|X_i-Q|+u(X_i-Q)\right)$$

toplamının Q'nun $X_1,X_2,...,X_n$ rasgele gözlemlerine dayalı p'ıncı örneklem yüzdeliği olduğunda en küçük olduğu gerçeğini Öklid iç çarpımının tanımlı olduğu $\{\mathbf{u} \mid \mathbf{u} \in \mathbf{R}^d, |u|<1\}$ ve $t \in R^d$ ' de geometrik yüzdelik tanımı ile vermiştir.

R^d'de değer alan $X_1,X_2,...,X_n$ rasgele vektörleri ne dayalı olarak ilgili $\mathbf{u}$ için geometrik yüzdelik tahmini

$$\hat{Q}_n(\mathbf{u}) = \arg\min_{Q \in R^d} \sum_{i=1}^{n} \left(\|X_i - Q\| + \mathbf{u}(X_i - Q) \right)$$

olarak verilir. Varlığı, tekliği ve hesaplama problemi ile büyük örneklem özellikleri çalışılmıştır. Sonsuz boyutlu rasgele vektörler için de geçerli olup döndürmeye karşı değişmezdir.

Parametrik olmayan çok değişkenli istatistiksel çıkarım için bir araç ve R^d'de verinin sıralanmasına ilişkin bir kavram olarak derinliğin yüzdelikler için kullanımı da kaçınılmaz görülmektedir. Çok değişkenli verilerin geldikleri yığına ait dağılım özellikleri derinlikleri aracılığıyla tek boyutlu eğrilerle görselleştirilebilmekte; ancak bunlara ait dağılım teorisi yanı cevapsız kalmakta; seçilen fonksiyonellerin örneklem eğrilerine güven sınırları yerleştirme ihtiyacı doğmaktadır.

Bu ihtiyacı karşılama amacıyla Serfling, R. [4] , Einmahl, J.H.J. ve Mason, D. M. [1]'ın genelleştirilmiş yüzdelikler sürecini λ fonksiyonu ile Borel kümeleri $B(R^d)$'nin bir alt sınıfı A **verilen** bir derinlik ölçüsü $D(x;F)$ ile uygun olarak yeniden şekillendirerek derinlik temeline dayalı genelleştirilmiş yüzdelikler süreci olarak elde etmiştir.

$D(x;F)$ için kolayca kontrol edilebilir kimi koşullar altında , [1]' in varsayımları sağlanıp limitsel sürecinin bir Brownian süreç olduğu gösterilmiştir. Sonuçlar yarı-uzay ve simpleks derinlikleri için gözden geçirilmiştir. Bulguları özetlemek için aşağıdaki gösterimler kullanılacaktır:

α derinlikli iç bölge $I(\alpha,D,F) = \{x \in R^d : D(x;F) \geq \alpha\}$ ($I(0,D,F) = R^d$ olmak üzere) ile, bu bölgenin sınırları α derinlikli kontur $\partial I(\alpha,D,F)$ ile gösterilecektir. Olasılığı p ve p'den daha fazla olan derinlik bağlantılı en büyük iç derinlikli bölgenin sınır derinliği

$$\alpha_p = \sup\{\alpha : P(I(\alpha,D,F)) \geq p\}$$

ve $\sup_x D(x;F) = \alpha^*$ ile gösterilecektir. $I(\alpha_p,D,F)$'nin varlığı ise rasgele fonksiyon $D(X;F)$'nin dağılım fonksiyonunun sürekli olması halinde gösterilebilir.

Genel hatlarıyla derinlik fonksiyonu $D(x;F)$ kabaca sonsuzda sıfırlanan, x bakımından sürekli , $0 < \alpha < \alpha^*$ için $\{x : D(x;F) = \alpha\} \neq \emptyset$ olduğunda ve derinlik fonksiyonu bakımından A sınıfı

$$\mathbf{A} = \{I(\alpha,D,F) : \alpha \in (0,\alpha^*)\}$$

ve λ_d fonksiyonu $\alpha \in (0,\alpha^*)$ için $\lambda_d(I(\alpha,D,F))$ sonlu, kesin azalan ve sürekli türeve sahipse ilgili genelleştirilmiş yüzdelik fonksiyonu $0 < p < 1$ için

$$\begin{aligned} U(p) &= \inf_{\alpha}\{\lambda_d(I(\alpha,D,F)) : P(I(\alpha,D,F)) \geq p\} \\ &= \lambda_d(I(\alpha_p,D,F)) \end{aligned}$$

olacak ve $g(p) = 1/U'(p)$ seçildiğinde $n \to \infty$ iken

$$\beta_n(p) = g(p)\sqrt{n}\left(U_n(p) - U(p)\right) \xrightarrow{d} B(p)$$

olup burada $B(.)$ Brownian köprü yani $C[0,1]$, $[0,1]$'da tanımlı sürekli fonksiyonlar sınıfı ve $\|.\|$ ile üretilmiş Borel kümelerinin sınıfı Θ üzerinde tanımlı Gauss sürecine dağılımda yakınsar.

Örneğin $I(\alpha_p,D,F)$ bölgesindeki derinlik temeline dayalı çok değişkenli saçılım ölçüsü $I(\alpha_p,D,F)$'nin hacmi olarak tanımlanır, $V(p,D,F) = Hacim(I(\alpha_p,D,F))$. $U(p) = V(p,D,F)$ olarak tanımlanıp $v(p)$ bunun türevini göstermek üzere yukarıdaki sonucun bir uygulaması olarak bir $p \in (0,1)$ için $n \to \infty$ iken

$$v(p)^{-1} n^{\frac{1}{2}} (V(p,D,F_n) - V(p,D,F)) \xrightarrow{d} N(0, p(1-p))$$

dir. Sonuçların uygulamaya geçirilmesi aşamasında $v(p)$ bilinmediğinde bunun düzgün tutarlı bir tahmin edicisi $\hat{v}(p)$ bulunup $v(p)$ yerine kullanılabilir. [4]'de $\hat{v}(p)$'nin örneklemden gözlemlenen $V(p,D,F_n)$'nin düzgünleştirilmiş (smooth) halinin kullanılarak elde edilebileceği ifade edilmektedir. $V(p,D,F_n)$'nin bir p için büyük örneklem $\gamma \in (0,1)$ güven düzeyli bir güven bandı

$$V(p,D,F_n) \pm k\frac{\hat{v}(p)}{\sqrt{n}}$$

kullanılarak oluşturulabilecektir. Burada k bir sabit olup $[a,b]\subset(0,1)$ aralığı için $P\left(\|B(.)\|_a^b \le k\right)=\gamma$ olacak şekilde elde edilir.

Uygulama

Yukarıda verilen sonucun uygulamalarından birisi dağılıma uygunluk testleri yerine kullanılan $Q-Q$ çizitlerine güven bantları yerleştirilerek görsel yönü öne çıkan, değerlendirme ve yorumlamada başka kolaylıkları da sağlayabilen, dağılıma uygunluk testlerine bir alternatif geliştirmek olabilir. Rasgele değişkenlerin dağılımlarına uygunluk testlerine bir alternatif olarak Rosenkrantz[5]tarafından yapılan bu öneri çok değişkenli dağılımlara uygunluk için Liu, Parelius ve Singh [6] tarafından önerilen ve derinliklere dayalı bazı çizitler için uyarlanabilir.

Bu uyarlama eliptik dağılımlar için dağılıma uygunluğun grafik değerlendirilmesine yönelik olarak Karabulut ve Öztürk[7]'de verilen uygulama için de geçerlidir.

Eliptik dağılıma sahip bir rasgele vektörün olasılık yoğunluk fonksiyonu, $x\in R^d$ olmak üzere c_d, d'ye bağlı bir sabit; h, $R^d \to R^+\cup\{0\}$ olan bir fonksiyon ve Σ_0 pozitif tanımlı bir matris olmak üzere

$$f(x)=\frac{c_d}{|\Sigma_0|^{1/2}}h\left((x-\mu)'\Sigma_0^{-1}(x-\mu)\right)$$

biçimindedir, [8] . İlgili rasgele vektörün bileşenlerinin ikinci dereceden momentlerinin var olması halinde saçılım matrisi varyans-kovaryans matrisi olup Σ ile gösterildiğinde bir α sabiti için $\Sigma_0=\alpha\Sigma$ ve $EX=\mu$ olacaktır.

[6] ' da açıklandığı üzere X eliptik dağılıma sahip d boyutlu rasgele bir vektör ve $p\in(0,1)$ bir sabit olsun. η_p, p ve d'ye bağlı bir sabit olmak üzere

$$\Sigma(p)=\eta_p\Sigma$$

dir. $Y=(X-\mu)'\Sigma_0^{-1}(X-\mu)$ ve ξ_p bu rasgele değişkene ait p. yüzdeliği göstermek üzere

$$\eta_p=\frac{E(Y\mid Y\le\xi_p)}{E(Y)}$$

dır. Normal dağılımın özel bir eliptik dağılım olduğu açıktır. Eğer, d boyutlu normal dağılımlı rasgele bir vektör sözkonusu ise yukarıdaki sonuçta ifade edilen Y rasgele değişkeni şekil parametresi $d/2$ ve ölçek parametresi 2 olan gamma dağılımına sahip, yani $Y\sim Gamma(d/2,2)$ olacaktır. Özel olarak $d=2$ alındığında da üstel, yani $Y\sim Üstel(2)$ dağılımlı olup

$$\eta_p=\frac{1-p}{p}\ln(1-p)+1$$

olacaktır. [7] 'de uygulamada yapıldığı gibi derinliklere dayalı olarak $\hat{\Sigma}(p)=\hat{\eta}_p\hat{\Sigma}$ sağlayacak $\hat{\eta}_p$ tahminleri ile η_p değerleri birbirlerine ne kadar yakın konumlanmaları elde ki örneklemin sözkonusu çok değişkenli eliptik dağılıma sahip olabileceğini destekleyici bir kanıt olacaktır. Burada $\hat{\Sigma}$ sözkonusu dağılımın saçılım matrisi varyans-kovaryans matrisinin bir tahmin edicisi $\hat{\Sigma}(p)$ ise α_p derinlikli kontur (derinliklere ait izdüşüm eğrisi) içinde kalan p. merkezi bölgesi $I(\alpha_p,D,F_n)=K\{X_{[1]},X_{[2]},X_{[3]},...,X_{[np]}\}$ içinde kalan gözlemlerle elde edilen $\Sigma(p)$ tahminidir.

Yukarıdaki koşulları sağlayan verilen bir derinlik fonksiyonu D için $\alpha^*\in(0,1)$ olmak üzere derinlik değerleri $\alpha\in(0,\alpha^*)$ için $\mathbf{A}=\{I(\alpha,D,F):\alpha\in(0,\alpha^*)\}$ kümeler sınıfında

$$
\begin{aligned}
U(p) &= \inf\{\lambda(I(\alpha,D,F)) : P(I(\alpha,D,F)) \geq p\} \\
&= \inf\{\lambda(I(\alpha,D,F)) : 0 < \alpha \leq \alpha_p\} \\
&= \lambda(I(\alpha_p,D,F)) \\
&= \eta(I(\alpha_p,D,F)) \\
&= \eta_p
\end{aligned}
$$

bir yüzdelikler süreci oluşturur. Çünkü burada η_p α'ya göre azalan, sonlu ve ele alınan çok değişkenli normal dağılım sınıfı için sürekli türeve sahiptir. Derinlik fonksiyonu D eğer yarı uzay derinliği veya simpleks derinliği ise

[4]'te Teorem 3.1'e göre her $p \in (0,1)$ için $n \to \infty$ iken

$$\frac{1}{\eta'_p} n^{\frac{1}{2}} (\hat{\eta}_p - \eta_p) \xrightarrow{d} N(0, p(1-p))$$

olup $\hat{\eta}_p$ ortalaması η_p ve varyansı $\dfrac{p(1-p)\eta_p^2}{n}$ asimptotik normal dağılımlıdır. Buradan her $p \in [a,b] \subset (0,1)$ için

$$\hat{\eta}_p \pm 1.96 \frac{\eta_p}{\sqrt{n}}$$

değerleriyle oluşturulacak $\gamma = 0.95$ güven düzeyli güven bantları sözkonusu çizite eklenip çok değişkenli normal dağılıma uygunluğun testine bir alternatif sunulmuş olur.

Sonuç

Yukarıdaki uygulamada [4]'te bilinmeyen $\nu(p)$ yerine bir tahminin konulduğu değil, kendisinin de tahmin edilmesine gerek duyulmayacak duruma bir örnek sunulmuştur. Çok değişkenli normal dağılıma uygunluğun testinin grafiksel alternatiflerinden birisi olarak [9], [10] ve [11]'de önerilen diğer test ve grafiksel değerlendirmelere göre uygulama ve değerlendirme kolaylığı olduğu öne sürülebilir.

Kaynaklar

[1] Einmahl, J.H.J. and Mason, D.M., Generalized quantile processes, Ann.Statist. Vol.**20**, (1992).

[2] Chen, L.-A. and Welsh, A.H., Distribution-Function-Based Bivariate Quantiles, Journal of Multivariate Analysis , Vol.82, (2002).

[3] Chaudhuri, P., On a geometric notion of quantiles for multivariate data, J. Amer. Statist. Assoc. Vol. 91, (1996)

[4] Serfling, R., Generalized quantile processes based on multivariate depth functions, with applications in nonparametric multivariate analysis, Journal of Multivariate Analysis, Vol. 83, (2002).

[5] RozenKrantz, W., Confidence bands for quantile functions: A parametric and graphic alternative for testing goodnes of fit, The American Statistician, Vol.54, No:3, (2000)

[6] Liu Regina Y., PareliusJ. M. and Singh K.,Multivariate analysis by data depth: Descriptive statistics, graphics and inference(with discussions). The Annals of Statistics, Vol. 27, No. 3, (1999).

[7] Karabulut, İ ve Öztürk, F., Derinlik ölçüleri ve çok değişkenli normal dağılıma uygunluğun grafikle değerlendirilmesi, Gazi Üniversitesi Fen Bilimleri Enstitüsü Dergisi, Cilt:14, NO:2, Ankara, (2001).

[8] Muirhead, Robb J., Aspects of Multivariate Statistical Theory. John Wiley & Sons, Inc., New York, (1982).

[9] [9] Öztürk, A., Romeu, Jorge L., A new method for assessing multivariate normality with graphical applications, Communications in Statistics-Simulation, 20(1), (1992).

[10] [10] Henze, N. and Wagner, T. A new approach to the BHEP tests for the multivariate normality, Journal of Multivariate Analysis, Vol. 62, (1997).

[11] [11] Yanqin, Fan, Goodness-of-fit tests for a multivariate distribution by the empirical characteristic function, Journal of Multivariate Analysis, Vol.62, (1997)