T.C.
ORDU ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİKSEL KORELASYON KAVRAMI

ESRA DEMÜR

Bu tez,
Matematik Anabilim Dalında
Yüksek Lisans
derecesi için hazırlanmıştır.

ORDU 2014

TEZ ONAY

Ordu Üniversitesi Fen Bilimleri Enstitüsü öğrencisi Esra DEMÜR tarafından hazırlanan ve Prof. Dr. Cemil YAPAR danışmanlığında yürütülen "İstatistiksel Korelasyon Katsayısı " adlı bu tez, jürimiz tarafından 19/06 /2014 tarihinde oy birliği / ~~oy çokluğu~~ ile Matematik Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman : Prof. Dr. Cemil YAPAR

Başkan : Prof. Dr. Cemil YAPAR İmza:
Matematik, Ordu Üniversitesi

Üye : Doç. Dr. Selahattin MADEN İmza:
Matematik, Ordu Üniversitesi

Üye : Yrd. Doç. Dr. Nurgül OKUR BEKAR İmza:
İstatistik, Giresun Üniversitesi

ONAY:

Bu tezin kabulü, Enstitü Yönetim Kurulu'nun 11.07.2014 tarih ve ...273....sayılı kararı ile onaylanmıştır.

11/07/2014

Enstitü Müdürü
Prof. Dr. Mehmet Fikret BALTA

TEZ BİLDİRİMİ

Tez yazım kurallarına uygun olarak hazırlanan bu tezin yazılmasında bilimsel ahlak kurallarına uyulduğunu, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezin içerdiği yenilik ve sonuçların başka bir yerden alınmadığını, kullanılan verilerde herhangi bir tahrifat yapılmadığını, tezin herhangi bir kısmının bu üniversite veya başka bir üniversitedeki başka bir tez çalışması olarak sunulmadığını beyan ederim.

İmza

E. Demir

Esra DEMÜR

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ÖZET

İSTATİSTİKSEL KORELASYON KAVRAMI

Esra DEMÜR

Ordu Üniversitesi
Fen Bilimleri Enstitüsü
Matematik Anabilim Dalı, 2014
Yüksek Lisans Tezi, 105s.

Danışman: Prof. Dr. Cemil YAPAR

Bu tez on bölümden ibarettir. Bu bölümlerde korelasyon katsayısı çeşitli yönlerden ele alınmıştır. Buna ek olarak, kanonik korelasyon hakkında açıklayıcı bilgiler verilmiştir. Konulara ilişkin geometrik yorumlar eklenmiştir. Tezde ihtiyaç duyulan bilgiler ekte sunulmuştur.

Anahtar Kelimeler : Pearson korelasyon katsayısı, Kanonik korelasyon, Kanonik korelasyon katsayısı, Singüler (tekil) değer ayrışımı, Salton kosinüs benzerliği.

ABSTRACT

CORRELATİON CONCEPT İN STATİSTİC

Esra DEMÜR

Ordu University
Institute of Science
Department of Mathematic, 2014
Master Thesis ,105p
Supervisor: Prof. Dr. Cemil YAPAR

This thesis consists of ten chapters. In these chapters, correlation coefficient has been discussed from various aspects. In addition to this, explanatory information on canonical correlation has been given. Geometric interpretations relating to the topics have been appended. Information needed for this thesis has been presented in the appendix.

Key Words: Pearson’s Correlation coefficient, Canonical correlation, Canonical correlation coefficient, Singular value decomposition, Salton’s cosines similarity.

TEŞEKKÜR

Bu tezin hazırlanması sürecinde, çok kıymetli zamanlarını ayırarak çalışmalarıma yardımcı olan engin bilgi ve deneyimleriyle bana yol gösteren değerli danışmanım Prof. Dr. Cemil YAPAR ‘a en içten teşekkürlerimi sunarım. Ayrıca, zaman zaman bilgi ve görüşlerine başvurduğum Ordu Üniversitesi, Fen-Edebiyat Fakültesi, Matematik Bölümü öğretim üye ve öğretim elemanlarına teşekkür ederim. Çalışmalarım boyunca yardımlarını eksik etmeyerek bana moral veren Ordu Üniversitesi, Fen Bilimleri Enstitüsü yüksek lisans öğrencisi kıymetli arkadaşım Fatma Buğlem YALÇIN’ a ve hayatım boyunca yanımda olan, maddi ve manevi yardımlarıyla ideallerimin gerçekleşmesini sağlayan değerli aileme yürekten teşekkürü bir borç bilirim.

İÇİNDEKİLER

ŞEKİLLER LİSTESİ

ÇİZELGELER LİSTESİ

SİMGELER VE KISALTMALAR

GKT	:	Genel kareler toplamı
RKT	:	Regresyon kareler toplamı
HKT	:	Hata kareler toplamı
SDA	:	Singüler değer ayrışımı
mnku	:	Mükemmel negatif korelasyondan uzaklık
mpku	:	Mükemmel pozitif korelasyondan uzaklık
Pu	:	Pozitif uzaklık
Nu	:	Negatif uzaklık
Köşeg	:	Köşegen
Kov	:	Kovaryans
Var	:	Varyans
E[ ]	:	Beklenen değer
TB	:	Temel bileşen
TBA	:	Temel bileşen analizi
ÇDND	:	Çok değişkenli normal dağılım
Y_{cos}	:	Yalancı kosinüs
J	:	Jaccard
P( )	:	Olasılık
o.y.f	:	Olasılık yoğunluk fonksiyonu

EK LİSTESİ

ÖNCEKİ ÇALIŞMALAR

İstatistiksel konular olarak tanımlanan deneysel ve teorik gelişmeler 1885'de Sir Francis Galton tarafından sunuldu. Bundan sonra 1895'de Karl Pearson, Pearson r sini yayımladı. Chatilon eliptik eş yoğunluk çevre eğrilerine sahip olan iki değişkenli dağılımların bir sınıfını verdi ve korelasyonun kabaca $r=\sqrt{1-\left(h/H\right)^2}$ olduğunu belirledi. . Marks (1982) basit bir hesapla $z_X = 0$ daki teğet doğrunun eğiminin korelasyon olduğunu gösterdi.

İki değişkenli normallik altında r'nin örneklem dağılımı ilk olarak 1915'de Fisher tarafından ortaya konan ve biraz karmaşıktır. Bununla beraber, sonunda anlaşıldı ki, n büyüdüğünde bu dağılım yaklaşık olarak ρ ortalamalı ve $V\left(r\right)\approx\frac{\left(1-\rho^2\right)^2}{n-1}$ varyanslı normal dağılım olduğunu gösterdi.

Jones ve Furnas (1987) Salton' un kosinüsü ve Pearson' un korelasyon katsayısı arasındaki farkı geometrik terimlere göre açıkladılar.

1.GİRİŞ

İstatistikte ve olasılıkta korelasyon katsayısı kavramı önemli bir yer tutar. İstatistikte, özellikle lineer modellerde ve lineer regresyon modellerinde bağımlı değişken ve açıklayıcı değişkenler arasındaki lineer ilişkilerin varlığının keşfinde, korelasyon katsayısı ve belirleme katsayısından yararlanılır. Olasılıkta ise olaylar arasındaki bağımlılık ve bağımsızlık ilişkileri çoğu kez olayların korelasyonlarının belirlenmesi ile ortaya konur. Korelasyon katsayısı lineer ilişkileri belirleyen bir katsayıdır. Lineer olmayan ilişkilerin belirlenmesi başka yöntemlerle ortaya konulur. Korelasyon katsayısı kavramını birçok yönden ele almak mümkündür. Burada bu kavrama bakışlardan önemli bir kısmın tanıtıldı. Bu tanıtımlardan sonra korelasyon katsayısının geometrik yorumları da ortaya kondu. Çünkü geometrik bakış, konunun anlaşılmasında önemli kolaylıklar getirmektedir. Özellikle kanonik korelasyon kavramı daha açık ve anlaşılabilir bir şekilde ortaya kondu. Bu kavramın açıklanmasında kullanılacak olan lineer cebir ifadelerine yer verildi. Tüm konuların incelenmesinde açıklayıcı örneklere başvuruldu. Özellikle geometrik yorumlar şekillerle açıklanmaya çalışıldı. Bütün bunlar yapılırken geniş bir kaynak taramasına yer verildi.

2.KORELASYON KAVRAMINA GENEL BİR BAKIŞ

Burada korelasyon katsayısı kavramına genel bir bakış yapılacaktır. Çok geniş olan bu kavram birbirinden on üç farklı bakışla ele alınacaktır.

2.1.Korelasyon Kavramı ve Korelasyon Katsayısı

Korelasyon iki rasgele değişken arasındaki ilişkinin (lineer ilişkinin) bir ölçüsüdür. Bu ilişkiyi ölçen katsayıya korelasyon katsayısı denir. Korelasyon katsayısı ölçekleme biriminden bağımsızdır ve $[-1,1]$ kapalı aralığında değerler alır. Kitle korelasyon katsayısı ρ ile, örneklem korelasyon katsayısı r ile gösterilir. Korelasyon katsayısının karesine belirleme katsayısı denir. Kitlede R^2 ile örneklemde r^2 ile gösterilir. r^2, R^2 nin örneklemden tahminidir. r nin sıfıra yakın olması X ve Y arasında zayıf bir korelasyon (ilişki) olduğunu ± 1 're yakın olması ise mükemmel bir korelasyonun olduğunu söyler. Korelasyonun negatif olması değişkenlerden birinin değeri artıyorken diğerinin azaldığını, pozitif olması biri artarken diğerinin arttığını (ya da biri azalırken diğerinin azaldığını) söyler.

Biz burada korelasyon katsayısına bakmanın 13 yolunu kısaca özetleyeceğiz ve sonra da kanonik korelasyon katsayısı üzerinde ayrıntılı bir şekilde duracağız.

2.2. Korelasyon Katsayısına Bakmanın On Üç Yolu

İstatistiksel konular olarak tanımlanan deneysel ve teorik gelişmeler 1885'de Sir Francis Galton tarafından sunuldu. Bundan sonra 1895'de Karl Pearson, Pearson 'un r sini yayımladı. Biz burada r nin istatistiği öğretenlere yararlı olacak olan hem geçmişine ve hem de bir takım kavramsallaştırmaları üzerinde duracağız. Aşağıda, korelasyon katsayısını yorumlamak için daha uzun bir inceleme biçimini sunacağız. Bu bölüm, korelasyonun bir geniş ve kavramsal farklı katsayıya geliştirildiğini, aynı zamanda bir yüzyıl önceki katsayının onun geçen zamanla önemli derecede etkilenmediğini gösterir.

2.3.Ham Sayıların ve Onların Ortalamalarının Bir Fonksiyonu Olarak Korelasyon

Pearson çarpım-moment korelasyon katsayısı her iki değişkenin lineer dönüşümüyle değişmeyen bir boyutsuz katsayıdır. Pearson ilk olarak 1895'de bu önemli ölçüm için matematisel bir formül ortaya koydu. Bu formül:

$$r = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\left[\sum\left(X_i - \overline{X}\right)^2 \sum\left(Y_i - \overline{Y}\right)^2\right]^{1/2}} \tag{2.1}$$

dir. Bu veya bazı basit cebirsel değişik biçim, istatistik ders kitaplarında bulunan formül için alışılagelen biçimdir. Pay kısmında ham sayılar her bir değişkenden ortalama çıkarılarak merkezileştirilir ve merkezileştirilen değişkenlerin çapraz-çarpımlarının toplamı yapılır. Paydada, aynı birimlere sahip olmak için değişkenlerin ölçekleri ayarlanır. Böylece (2.1) bağıntısı r yi iki değişkenin çapraz-çarpımının merkezileştirilmiş ve standartlaştırılmış toplamı olarak ifade eder. Cauchy-Schawarz eşitsizliğini kullanarak payın mutlak değerinin paydadan küçük veya eşit olduğu gösterilebilir. Bu nedenle r için ± 1 lik sınırlar belirlenir. Bu formülün birçok basit cebirsel dönüşümleri hesaplama amaçları için kullanılabilir.

2.4. Standartlaştırılmış Kovaryans Olarak Korelasyon

Korelasyon gibi, kovaryans da değişkenler arasındaki lineer ilişkinin bir ölçüsüdür. Kovaryans merkezileştirilen değişkenlerin çapraz-çarpımlarının toplamı üzerinde tanımlanır, değişkenlerin ölçeği için düzeltilmemiştir. İstatistik ders kitaplarında çoğu kez önemsenmemiş olmakla beraber varyans (kovaryans değil) kovaryansın özel bir durumudur, yani varyans bir değişkenin kendisi ile kovaryansıdır. İki değişkenin kovaryansı sonsuz kitlelerde sınırsızdır, örneklemde ise belirsiz sınırlara (ve kaba yoruma) sahiptir. Bu nedenle kovaryans çoğu kez ilişkinin yararlı bir tanımlayıcı ölçüsü değildir, çünkü onun değeri X ve Y için ölçümün ölçeklendirmelerine bağlıdır. s_{xy} örneklem kovaryansı s_x ve s_y örneklem standart sapmaları olmak üzere, korelasyon katsayısı yeniden ölçeklendirilmiş kovaryanstır. Yani;

$$r = s_{xy} / s_x s_y \tag{2.2}$$

dir. Kovaryans iki standart sapmayla bölündüğünde, kovaryansın değişim aralığı -1 ve $+1$ arasında yeniden ölçeklendirilir. Bu nedenle ilişkinin bir ölçümü olarak korelasyonun yorumu kovaryansınkinden genel olarak daha izlenebilirdir (ve farklı korelasyonlar daha kolay karşılaştırılırlar).

2.5 Regresyon Doğrusunun Standartlaştırılmış Eğimi Olarak Korelasyon

Korelasyon ve regresyon arasındaki ilişki

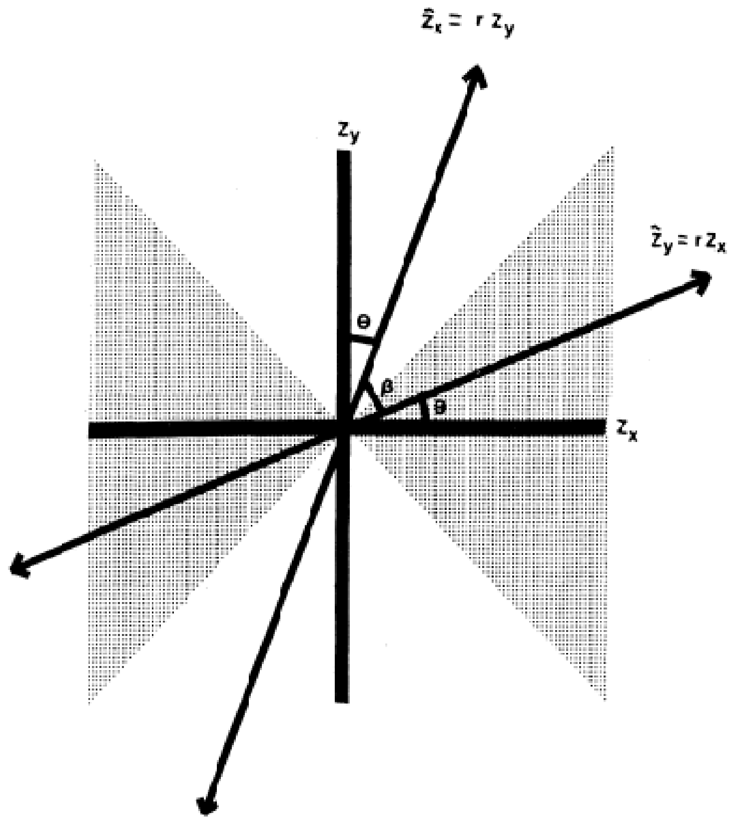
$$r = b_{Y.X}\left(s_X / s_Y\right) = b_{X.Y}\left(s_Y / s_X\right) \tag{2.3}$$

de çok daha kolay tanımlanır. Burada $b_{Y.X}$ ve $b_{X.Y}$ sırasıyla X den Y 'yi ve Y den X 'i tahmin etmek için ortaya konan regresyon doğrularının eğimleridir. Burada korelasyon her iki regresyon doğrusunun eğimleri ve iki değişkenin standart sapmalarının bir fonksiyonu olarak ifade edilir. Standart sapmaların oranı regresyon eğiminin birimlerini korelasyonun birimlerine yeniden ölçeklemenin etkisine sahiptir.

Benzer bir yorum korelasyonu standartlaştırılmış regresyon doğrusunun eğimi olarak içerir. İki ham değişkeni standartlaştırdığımızda, standart sapmalar birim olur ve regresyon doğrusunun eğimi korelasyon olur. Bu durumda, sabit terim 0 dır ve regresyon doğrusu

$$\hat{z}_y = r z_x \tag{2.4}$$

olarak kolayca ifade edilir. Bu yorumdan, korelasyonun, standartlaştırılmış Y değişkeninin birimlerini tahmin etmek için standartlaştırılmış X değişkenlerinin birimlerini yeniden ölçeklendirdiği açıktır. z_y nin z_x üzerindeki regresyonunun eğiminin Şekil 2.1 ün taralı bölgesinde gösterilen iki köşegen arasına düşecek regresyon doğrusuna sınırlayacağına dikkat edelim. Pozitif korelasyonlar doğrunun birinci ve üçüncü bölgelerden geçeceğini ifade eder, negatif korelasyonlar doğrunun ikinci ve dördüncü bölgelerden geçeceğini ifade eder. z_x in z_y üzerindeki regresyonunun y ekseni ile yaptığı açı z_y nin üzerindeki regresyonunun x ekseniyle yaptığı açının aynısıdır ve o gösterildiği gibi, Şekil 1 rin taralı olmayan bölgesine düşecektir.



Şekil 2.1 Standartlaştırılmış değişkenler için iki değişkenli korelasyonun geometrisi

2.6 İki Regresyon Eğiminin Geometrik Ortalaması Olarak Korelasyon

Korelasyon standartlaştırılmamış regresyon doğrularının iki eğiminin, yani $b_{Y.X}$ ve $b_{X.Y}$ nin eş anlı bir fonksiyonu olarak da ifade edilebilir. Gerçekten, fonksiyon geometrik ortalamadır ve bu fonksiyon r nin özel bir ortalama tipi olarak birkaç yorumundan ilkini gösterir. Yani:

$$r = \pm\sqrt{b_{Y.X}\,b_{X.Y}} \qquad (2.5)$$

dir. Bu ilişki r^2 yi ortaya koymak için eşitlikteki ikinci ve üçüncü terimleri çarparak, standart sapmaları yok ederek ve karekök alarak, (2.1) bağıntısından elde edilebilir.

2.7 İki Varyansın Oranının (Açıklanan Değişebilirliğin Oranının) Karekökü Olarak Korelasyon

Korelasyon çoğu kez birimlerinin için açık yoruma sahip olmadığı için eleştirilir. Bu eleştiri korelasyonun karesi alınarak hafifletilir. Karesi alınmış katsayıya çoğu kez belirleme katsayısı denir ve birimler bir değişkendeki değişimin diğerindeki farklılıklarla açıklanan oranı olarak yorumlanabilir. Y için genel kareler toplamını (GKT) yi regresyon kareler toplamı (RKT) ve hata kareler toplamı (HKT) na ayırabiliriz. Y 'deki farklılıklarla açıklanan X deki değişebilirlik (RKT)'nin (GKT) ye oranıdır ve r bu oranın kareköküdür, yani;

$$r = \sqrt{RKT/GKT} \tag{2.6}$$

dır. Eşdeğer olarak bu yorumun pay ve paydası $n-1$in karekökü ile bölünebilir ve r tahmin edilen ve gözlenen değişkenlerin varyansların oranının karekökü olur. Yani

$$r = \sqrt{\sum\left(Y_i - \hat{Y}_i\right)^2 \Big/ \sum\left(Y_i - \bar{Y}\right)^2} = \sqrt{RKT/GKT}$$

dır. Bu yorum katsayının daha önce harekete geçirilen Pearson kavramlaştırmalarından biridir. İki varyansın bir oranı olarak korelasyon iki ortalamanın oranı olarak korelasyonun başka bir yorumuyla karşılaştırılabilir. O yorumu da on üçüncü bakışta sunacağız.

2.8 Standartlaştırılmış Değişkenlerin Ortalama Çapraz Çarpımı Olarak Korelasyon

Korelasyonu bir ortalama olarak yorumlamak için başka bir yol onu standartlaştırılmış değişkenlerin ortalama çapraz çarpımı olarak ifade etmektir. Buna göre,

$$r = \sum z_X z_Y / N \tag{2.7}$$

dir. (2.5) bağıntısı, (2.7) bağıntısındaki pay ve paydanın her ikisinde iki örneklem standart sapmasının çarpımı ile bölmek suretiyle doğrudan doğruya elde edilebilir. Bir dağılımın ortalaması onun birinci momenti olduğundan, bu formül korelasyon katsayısının adlandırılmasında “çarpım momenti” nin anlamına ilişkin bilgi verir. Bundan sonraki iki tanımlama korelasyonun trigonometrik yorumlarını gösterir.

2.9 İki Standartlaştırılmış Regresyon Doğrusu Arasındaki Açının Bir Fonksiyonu Olarak Korelasyon

Daha önce de belirtildiği gibi, iki standartlaştırılmış regresyon doğrusu her iki köşegen etrafında simetriktir. İki doğru arasındaki açı β olsun.(bkz. Şekil 2.1) Bu takdirde,

$$r = \tan(\beta) \pm \sec(\beta) \tag{2.8}$$

dir. Bu ilişkinin basit bir ispatı hemen elde edilebilir.(2.8) bağıntısı sezgisel olarak açık değildir, dahası diğerlerinin bazısı gibi hesapsal ve kavramsal amaçlar için yararlı değildir. Bu yorumun değeri korelasyon ve iki regresyon doğrusu arasındaki açısal uzaklık arasında bir sistematik ilişkinin var olduğunu göstermektir. Bundan sonraki yorum bir trigonometrik yorumdur, esas olarak daha fazla kavramsal değere sahiptir.

2.10 İki Değişken Vektör Arasındaki Açının Bir Fonksiyonu Olarak Korelasyon

Değişkenler arasındaki ilişkiyi göstermek için standart geometrik model saçılım grafiğidir. Bu uzayda gözlemler değişken eksenlerle tanımlanan bir uzaydaki noktalar olarak gösterilirler. Bu uzayın genellikle "şahıs uzayı" denilen içi dışına dönmüş (ters yüz) versiyonu her bir eksenin bir gözlemi temsil etmesine izin vererek tanımlanabilir. Bu uzay büyük boyutlu uzaydaki vektörlerin uç noktalarını tanımlayan her bir değişken için bir-iki noktayı içerir. Her ne kadar bu uzayın çok boyutluluğu görsellemeyi engellese de, iki değişken vektör kolayca kavramsallaştırılan bir iki boyutlu alt uzayı tanımlar.

Eğer değişken vektörler merkezileştirilmiş değişkenlere dayandırılırsa bu takdirde korelasyon doğrudan doğruya değişken vektörler arasındaki tek bir α açısıyla anlaşılır bir ilişkiye sahiptir. Yani;

$$r = \cos\alpha \tag{2.9}$$

dır. Açı sıfır derece olduğunda vektörler aynı doğrular üzerine düşer ve $\cos\alpha = \pm 1$ dir. Açı doksan derece olduğunda ise vektörler dik olurlar ve $\cos\alpha = 0$ dır. Bununla beraber r nin bir açının kosinüsü olarak gösterilmesine imkan veren bu tersine çevrilmiş uzay bir yorum aracı olarak göreceli bir şekilde önemsenmez

2.11 Standartlaştırılmış Sayılar Arasındaki Farkların Yeniden Ölçeklendirilen Bir Varyansı Olarak Korelasyon

$z_Y - z_X$ her bir gözlem için standartlaştırılan X ve Y değişkenleri arasındaki fark olarak tanımlayalım. Bu takdirde,

$$r = 1 - s^2_{(z_Y - z_X)} \Big/ 2 \qquad (2.10)$$

dir. Bu, bir fark sayılarının varyansı ile başlamak suretiyle gösterilebilir. Buna göre $s^2_{Y-X} = s^2_X + s^2_Y - 2rs_Y s_X$ dir. Değişkenler standartlaştırıldığında standart sapmalar ve varyans bir olduğundan, $s^2_{Z_Y - Z_X} = s^2_{Z_Y} + s^2_{Z_X} - 2rs_{Z_y} s_{Z_x}$ yazar ve $s_{z_Y} = s_{z_X} = 1$ koyarak (2.8) bağıntısını kolayca elde ederiz. Bu denklemde korelasyon -1 den +1 e aralıkla sınırlandığından, bu fark sayısının varyansının 0 dan 4 e aralıkla sınırlandığını görmek ilginçtir. Bu nedenle standartlaştırılmış sayıların bir farkının varyansı 4 ü asla geçmez. Korelasyon -1 olduğunda varyans için üst sınıra ulaşılır. r yi standartlaştırılmış sayıların bir toplamının varyansı olarak da tanımlayabiliriz. Yani;

$$r = s^2_{z_Y + z_X} \Big/ 2 - 1 \qquad (2.11)$$

dir. Burada, toplamın varyansı da 0 dan 4 e değişir ve korelasyon $+1$ olduğunda, üst sınıra ulaşır. Bu dokuzuncu yorum, korelasyonun bir varyansın belli tipten bir lineer dönüşümü olduğunu göstermektedir. Bu nedenle korelasyon verildiğinde standartlaştırılmış değişkenlerin toplamının veya farkının varyansına doğrudan tanımlayabiliriz.

Korelasyon katsayısının yukardaki yorumlarının hepsi aslında cebirsel veya trigonometrikti. X ve Y nin bir değişkenli veya iki değişkenli dağılımların yapısı hakkında dağılımsal varsayımlar yapılmadı. Son yorumlarda iki değişkenli normallik kabul edilecek. r'nin kavramsal ve hesapsal versiyonları ile ilgileneceğiz fakat kitle dağılımı hakkındaki bu ortak varsayım üzerindeki yorumların son kümesine dayanacağız.

2.12 Balon Kuralından Tahmin Edilen Korelasyon

Bu yorum Chatilon' a aittir. O, bir iki değişkenli ilişkinin saçılım grafiği etrafında bir balon çizmeyi önerdi. Balon gerçekten yaklaşık bir elipstir, bu elipsten iki h ve H

ölçüleri elde edilir (bkz. şekil 2.2). h x ekseni üzerindeki dağılımın merkezinde elipsin dik çapıdır. H y ekseni üzerinde elipsin değişim aralığıdır. Chatilon korelasyonun kabaca

$$r = \sqrt{1-\left(h/H\right)^2} \tag{2.12}$$

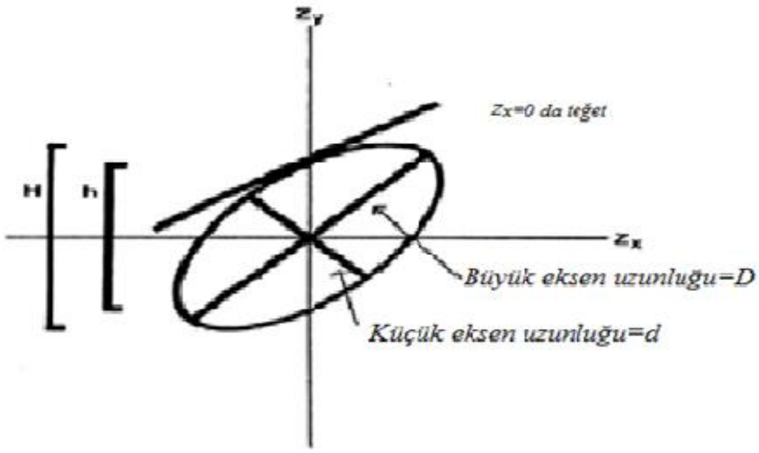
olarak hesaplanabileceğini gösterdi. O, hem iki değişkenli normalliği hem de iki değişkenli değişmezliği kabul ederek, bu kaba ama işe yarar hesapsal yönteminin etkiliğinin bir teorik doğrulamasını verdi. O teknik çalışmalarda oldukça iyi bir takım örnekleri de sundu. Onun yaptığı ilginç öneri bazı belirli korelasyonlarla bir, iki değişkenli ilişkiyi yaklaşık olarak kurmak için kullanılabilen "balon kuralı" dır. İstenilen r yi üreten bir elips çizeriz ve bundan sonra elipsi başından sonuna düzgün bir şekilde noktalarla doldururuz.

2.13 Eş Yoğunluğun İki Değişkenli Elipslerle İlişkisinde Korelasyon

r nin eş yoğunluklu iki değişkenli elipslere bağlı yorumunu farklı iki yazar önerdi. Bu elipslerin Kısım 2.10 daki balonun daha formal versiyonları olduğuna ve onların deneysel veride Galton(1984)' un gözlemlediği geometrik yapılar olduğuna dikkat edelim. Chatilon(1984a) eliptik eş yoğunluk çevre eğrilerine sahip olan iki değişkenli dağılımların bir sınıfını verdi. Kitle korelasyonu verildiğinde her pozitif sabit için bir elips vardır. Bir saçılım grafiği etrafında çizilmesi gereken balon büyük bir pozitif sabit için bu elipslerden birine yaklaşmalıdır. Eğer değişkenler standartlaştırılır ise bu elipsler orijinde merkezileşirler. $\rho > 0$ için büyük eksen pozitif köşegen üzerine düşer; $\rho < 0$ için negatif eksen üzerine düşer. Marks (1982) basit bir hesapla $z_X = 0$ daki teğet doğrunun eğiminin korelasyon olduğunu gösterdi. Şekil 2.2 eğimi r ye eşit olan bu teğet doğruyu gösterir. Korelasyon 0 olduğunda elips bir çemberdir ve teğet 0 eğimine sahiptir. Korelasyon 1 olduğunda elips bir doğruya yaklaşır ki o doğru köşegendir ve eğimi 1 dir. Eş yoğunluğun, elipslerinin tümü paralel olduğundan, yorumun elipslerin seçimine göre değişmediğine dikkat edelim. $z_x = 0$ da teğetin eğiminin standartlaştırılmış regresyon doğrusunun eğimiyle aynı olduğunu belirtmemiz gerekir. Schilling (1984) benzer bir ilişkiyi çıkarmak için bu yapıyı da kullandı. Değişkenler standartlaştırılmış olsun. Bu nedenle önceden olduğu gibi elipsler orijinde merkezileşirler. Eğer D eş yoğunluğun herhangi elipsinin büyük ekseninin uzunluğu ve d küçük eksenin uzunluğu ise, bu takdirde

$$r = \left(D^2 - d^2\right)\left(D^2 + d^2\right) \quad (2.13)$$

dir. Şekil 2.2 de bu eksenler de gösterildi ve önceden olduğu gibi yorum elipsin seçimine göre değişmez.



Şekil 2.2 Eş yoğunluk elipslerinin fonksiyonlarına bağlanan korelasyon

2.14 Tasarlanan Denemelerden Test İstatistiklerinin Bir Fonksiyonu Olarak Korelasyon

r nin önceki yorumları sayısal değişkenlere bağlanmıştı. Bu korelasyon yorumumuz değişkenlerden birinin (bağımsız değişken) bir kategorik değişken olduğu, tasarlanmış denemelerden test istatistiği ile korelasyonun ilişkisini gösterir. Bu da ele alınan deneysel tasarımda korelasyonun deneysel ayrımın yapaylığını gösterir. Gerçekten, Fisher (1925) ilk olarak varyans analizini sınıf içi korelasyona dayanarak sundu. İki işlem şartlı bir tasarlanmış denemeye sahip olduğumuzu farz edelim. Şartlar arasındaki farkı test etmek için standart istatistiksel model iki-bağımsız örneklem t testidir. X grup üyeliğini gösteren (grup 1 ise 0, grup 2 ise 1) bir iki değerli değişken olarak tanımlanırsa bu takdirde X ve Y bağımlı değişkeni arasındaki korelasyon

$$r = t\Big/\sqrt{t^2 + n - 2} \quad (2.14)$$

dir. Burada n iki işlem grubundaki gözlemlerin birleştirilmiş toplam sayısıdır. Bu korelasyon katsayısı bir etkinin önemine zıt olarak bir işlem etkisinin gücünün bir ölçüsü olarak kullanılabilir. Bu kurgulamada r nin öneminin testi alışılagelen t testi ile aynı testi ortaya koyar. Aşikar olarak, bu takdirde r gözlemsel kurgulamalardaki ilişkinin bir ölçüsünü vermesinin yanı sıra, tasarlanan bir denemede bir test istatistiği olarak hizmet görebilir. Çok gruplu ve çok faktörlü varyans analizi kurgulamalarında bu ilişkinin genişlemesi daha karmaşık denemelerde asıl etkilere ve etkileşimlere ilişkin çoklu korelasyon katsayılarını tanımlar.

2.15 İki Ortalamanın Oranı Olarak Korelasyon

Bu, ortalamaları içeren korelasyonun üçüncü yorumudur. X annenin IQ sunu gösteren bir değişken olsun ve Y onun en yaşlı çocuğunun IQ sunu göstersin. Bundan başka, $\mu(X)$ ve $\mu(Y)$ ortalamalarının 0 ve $\sigma(X), \sigma(Y)$ standart sapmalarının 1 olduklarını kabul edelim. Şimdi X in X_c diyeceğimiz, herhangi keyfi büyük değerini seçeceğiz. IQ su X den daha büyük olan annelerin ortalama IQ sunu hesaplayalım. Bu ortalama $\mu(X|X > X_c)$ ile yani IQ su X_c den daha büyük olan annelerin ortalama IQ su ile gösterilmiş olsun. Sonra da, bu istisna annelerin en yaşlı çocuğunun IQ sayılarının (yani Y lerin) ortalamasını alalım. Bu ortalamayı $\mu(Y|X > X_c)$ yani, IQ ları X_c den daha büyük olan annelerin en yaşlı çocuğunun ortalama IQ su ile gösterelim. Bu takdirde,

$$r = \frac{\mu(Y|X > X_c) - \mu_Y}{\mu(X|X > X_c) - \mu_X} = \frac{\mu(Y|X > X_c)}{\mu(X|X > X_c)} \tag{2.15}$$

olduğu gösterilebilir. Bu bağıntının ispatı standartlaştırılmış X ve Y nin bir iki değişkenli normallik varsayımına ihtiyaç duyar. İspat basittir sadece bu iki şartlı ortalamanın oranının yanı sıra z_X ve z_Y için r nin regresyon doğrusunun eğimi olduğu gerçeğini şarta bağlar. Örneğimiz özeldir, fakat yorum belli seçimin bir değişken üzerinde, dolaylı seçimlerin ikinci bir değişken üzerinde olduğu herhangi bir ortama uygulanır.

SONUÇ:

Hiç şüphesiz korelasyon katsayısını yorumlamak için diğer yollar da mevcuttur. Korelasyon problemine daha fazla istatistiksel ve daha az cebirsel bir yaklaşım alındığında, ek ilginç ve faydalı tanımlamalar mevcuttur. Çok dar tanımladığımız çerçeve içinde kalsa bile faydalı veya ilginç yaklaşımların tümünü özetlediğimizi asla kabul edemeyiz. Ama yine de bu ön üç yaklaşım, korelasyonu kullanan öğretmenler ve araştırmacılar için mevcut yorumların değişik türlerini açıklar. (Daha fazla ayrıntı için bkz. Joshep L. Rodgers; Alan Nicewander (1988)).

3. KANONİK KORELASYON KAVRAMI, KANONİK KORELASYON DEĞİŞKENLERİ VE KANONİK KORELASYON KATSAYISI

Burada kanonik korelasyonu basit olarak ele aldıktan sonra, buna ek olarak daha genel bir açıklamayı ele alacağız.

3.1 Kanonik Korelasyona Sade Bir Bakış

Şimdi kanonik korelasyona daha anlaşılabilir sade bir bakışı sunalım. $n_x + n_y = n$, $T > n$ olmak üzere, sırasıyla $T \times n_y$ ve $T \times n_x$ boyutlu X ve Y matrislerini ortaya koyan iki değişken kümesi üzerinde T sayıda gözlemin var olduğunu düşünelim. Bu matrislerin ortalamalarının sütun ortalamalarını çıkarmak suretiyle ayarlandıklarını ve matrislerin tam sütun ranklı olduklarını kabul edelim. Amaç $(v_1, w_1), (v_2, w_2), ..., (v_p, w_p)$ çiftleri dahilinde vektörlerin korelasyonları maksimum olacak şekilde, y nin sütunlarını karşılıklı olarak ilişkisiz lineer kombinasyonlarının bir kümesini yani

$$v_1 = Y\alpha_1, v_2 = Y\alpha_2, \ldots, v_2 = Y\alpha_2 \qquad (3.1)$$

kümesini ve X sütunlarının kombinasyonlarının benzer bir kümesini, yani

$$w_1 = X\beta_1, w_2 = X\beta_2, \ldots, w_p = X\beta_p \qquad (3.2)$$

yi bulmaktır. Bu şartlar sağlandığında $i \neq j$ olduğunda $v_i = Y\alpha_i$ ve $w_j = X\beta_j$ nin ilişkisiz olmaları şartı bir yan ürün olarak ortaya çıkar. v_j ve w_j vektörleri v yinci kanonik değişkenler olarak ifade edilirler. Değişkenlerin deneysel varyans-kovaryans matrisini

$$\begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix} = \frac{1}{T} \begin{bmatrix} Y'Y & Y'X \\ X'Y & X'X \end{bmatrix} \qquad (3.3)$$

ile gösterelim. Bu duruda kısıtlamalar: Her i için $\alpha_i' S_{yy} \alpha_i = 1$, her j için $\beta_j' S_{xx} \beta_j = 1$ ve $i \neq j$ olduğunda,

$$\alpha_i' S_{yx} \beta_j = 0 \qquad (3.4)$$

olur. Kanonik değişkenler v_1 ve w_1 ile başlayarak sıra ile bulunurlar.

$$r\left(\alpha,\beta\right)=\frac{\alpha' S_{yx}\beta}{\left(\alpha' S_{yy}\alpha\right)^{1/2}\left(\beta' S_{xx}\beta\right)^{1/2}} \tag{3.5}$$

korelasyonunu göz önüne lalım. (3.4) kısıtlamaları altında, maksimum

$$L\left(\alpha,\beta\right)=\left(\alpha' S_{yx}\beta\right)^2-\lambda\left(\alpha' S_{yy}\alpha-1\right)-\mu\left(\beta' S_{xx}\beta-1\right) \tag{3.6}$$

Lagrange fonksiyonunun değerini belirleyerek aranır. Bu fonksiyonu α ve β ya göre diferensiyelleme ve sonucu sıfır eşitleme:

$$\begin{aligned} &-\lambda S_{yy}\alpha+\left(\alpha' S_{yx}\beta\right)S_{yx}\beta=0 \\ &\left(\alpha' S_{yx}\beta\right)S_{xy}\alpha-\mu S_{xx}\beta=0 \end{aligned} \tag{3.7}$$

denklemlerini ortaya koyar. Birinci denklemi α' ve ikinci denklemi β' ile önden çarpma,

$$\lambda=\mu=\left(\alpha' S_{yx}\beta\right)^2 \tag{3.8}$$

olduğunu gösterir ki bu, verilen kısıtlamalar altında λ Lagrange çarpanının değerinin her α ve β için $r\left(\alpha,\beta\right)$ nin maksimum değeri olduğunu söylemektedir.

(3.8) bağıntısına göre (3.7) denklem sistemi

$$\begin{bmatrix} -\lambda^{1/2}S_{yy} & S_{yx} \\ S_{xy} & -\lambda^{1/2}S_{xx} \end{bmatrix}\begin{bmatrix}\alpha \\ \beta\end{bmatrix}=\begin{bmatrix}0 \\ 0\end{bmatrix} \tag{3.9}$$

olarak tekrar yazılabilir. α ve β yı ayrı veren denklemler ise;

$$\begin{aligned} &\left[S_{yx}S_{xx}^{-1}S_{xy}-\lambda S_{yy}\right]\alpha=0 \\ &\left[S_{xy}S_{yy}^{-1}S_{yx}-\lambda S_{xx}\right]\beta=0 \end{aligned} \tag{3.10}$$

denklemidir.(3.9) sisteminin $\alpha=0$ ve $\beta=0$ dan başka bir çözüme sahip olması için sistemin katsayılar matrisinin tekil olması veya eşdeğer olarak determinantının sıfır olması gerekir. Yani,

$$\begin{vmatrix} -\lambda^{1/2}S_{yy} & S_{yx} \\ S_{xy} & -\lambda^{1/2}S_{xx} \end{vmatrix}=0 \tag{3.11}$$

olmalıdır. Determinant n-yinci dereceden bir polinomdur. Matrisin simetrikliğinin sonucu olarak onun kökleri reel değildir ve azalan büyüklüğe göre $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ biçiminde sıralanabilirler. α_1 ve β_1, λ_1 karşılık gelen çözümler olsun. Bu takdirde $v_1 = Y\alpha_1$ ve $w = X\beta_1$ birinci (ilk) kanonik değişkenlerdir. $\sqrt{\lambda_1}$ onların kanonik korelasyonudur.

İkinci kanonik değişkenler, birinci v_1 ve w_1 kanonik değişkenlerle ilişkisiz olma şartına bağlı olarak maksimum korelasyona sahip olan maksimum korelasyona sahip olan $v_2 = Y\alpha_2$ ve $w_2 = X\beta_2$ lineer kombinasyonlarıdır. Tüm söylenen şartlar

(i) $r(\alpha,\alpha_1) = \alpha' S_{yy}\alpha_1 = 0$ (iii) $r(\beta,\beta_1) = \beta' S_{xx}\beta_1 = 0$

(ii) $r(\alpha,\beta_1) = \alpha' S_{xx}\beta_1 = \lambda_1\alpha' S_{yy}\alpha_1 = 0$ (iv) $r(\beta,\alpha_1) = \beta' S_{xy}\alpha_1 = \lambda_1\beta' S_{xx}\beta_1 = 0$

dır. Fakat (ii) ve (iv) şartının, (i) ve (iii) den otomatik olarak sezilmesi anlamında, lüzumsuz (gereksiz) oldukları açıktır. İkinci kanonik değişkenlerin korelasyonu

$$L(\alpha,\beta) = \left(\alpha' S_{yx}\beta\right)^2 - \lambda\left(\alpha' S_{yy}\alpha - 1\right) - \mu\left(\beta' S_{xx}\beta - 1\right) + \pi\alpha' S_{yy}\alpha_1 + \rho\beta' S_{xx}\beta_1 \quad (3.13)$$

Lagrange fonksiyonu vasıtasıyla maksimumlaştırır. Burada λ, μ, π ve ρ Lagrange çarpanlarıdır. α ve β ya göre diferensiyelleme ve sonucu sıfıra eşitleme aşağıdaki şartları verir.

$$\begin{aligned} -\lambda S_{yy}\alpha + \left(\alpha' S_{yx}\beta\right) S_{yx}\beta + \pi S_{yy}\alpha_1 = 0 \\ \left(\alpha' S_{yx}\beta\right) S_{xy}\alpha - \mu S_{xx}\beta + \rho S_{xx}\beta_1 = 0 \end{aligned} \quad (3.14)$$

Bu denklemleri sırasıyla $\alpha_1^{'}$ ve $\beta_1^{'}$ ile önden çarpma ve (3.12) nin şartlarını kullanma

$$\begin{aligned} \pi\alpha_1^{'} S_{yy}\alpha_1 = \pi = 0 \\ \rho\beta_1^{'} S_{xx}\beta_1 = \rho = 0 \end{aligned} \quad (3.15)$$

olduğunu ortaya koyar. Bu şartlar altında (3.14) denklemleri (3.7) nin biçimini kabul eder ve bu, denklemlerin α_1 ve β_1 den başka çözümlerinin (3.12) nin şartlarını otomatik olarak sağladığını gösterir. Eğer α_2 ve β_2 karşılık gelen çözümler ise, bu takdirde $v_2 = Y\alpha_2$ ve $w_2 = X\beta_2$ ikinci kanonik değişkenlerdir ve $\sqrt{\lambda_2}$ onların korelasyonudur.

Müteakip kanonik değişkenler (3.11) rin büyüklüğüne göre sıralanmış sıfır olmayan köklerine karşılık bulunabilirler. X ve Y biri veya diğerinin sahip olduğu en az sütun sayısından oluşturulabilen bağımsız lineer kombinasyonların maksimum sayısı olan $p = min\left(n_x, n_y\right)$ daha fazla kökünün olmadığı gösterilebilir.

Mesele (3.10) denklemlerine bağlı sonuca göre en iyi anlaşılabilir. $S_{yx}S_{xx}^{-1}S_{xy}$ ve $S_{xy}S_{yy}^{-1}S_{yx}$ sırasıyla Y nin X sütun uzayı üzerine ve X in Y nin sütun uzayı üzerine iz düşümünün deneysel moment matrisidirler. Bu moment matrislerinin ortak rankları X ve Y nin ranklarının ρ minimumuna eşittir. Eğer $n_x \neq n_y$ ise bu takdirde moment matrislerinden biri noksan ranklı olmaya mecburdur. (3.8) e göre S_{yy} ile tanımlanan metriğe göre $S_{yx}S_{xx}^{-1}S_{xy}$ ın sıfırdan farklı öz değerleri S_{xx} ile tanımlanan metriğe göre $S_{xy}S_{yy}^{-1}S_{yx}$ sıfır olmayan öz değerlerinin aynısıdır. Onların p sayısı moment matrislerinin ortak rankına eşittir.

3.2 Kanonik Korelasyona Genel Bir Bakış

Faktör analizi değişkenlerin bir kümesindeki ilişkiyi modelleme ve özetlememize yardım eder. Eğer değişkenlerin iki kümesi arasındaki ilişkiyi özetlemek istersek ne yapmalıyız? Örneğin, $x^{(1)}$ in bir mekânda (asidik) döküntüyü ıslatmaya ilişkin sekiz bileşenli bir vektör olduğunu farz edelim. Yani bileşenlerin: PH, SO_4, NO_3, CI, NH_4, Ca, Na ve Mg konsantrasyonları olduğunu farz edelim. Aynı zamanda $x^{(2)}$ nin de döküntüyü kurutmaya ilişkin üç bileşenli bir vektör olduğunu farz edelim. Bu bileşenler de: SO_2, O_3 ve $PMLO$ olsun.

$x^{(1)}$ hangi durumları $x^{(2)}$ ile en çok ilişkilidir? Bu güzel bir sorudur. Genel olarak $x^{(1)}$ ve $x^{(2)}$ nin hangi lineer kombinasyonlarını kabulleniriz. Yani $x^{(1)}$ ve $x^{(2)}$ nin en yüksek derecede ilişkili olan lineer kombinasyonlarını kabulleniriz. Yani, $x^{(1)}$ ve $x^{(2)}$ nin en yüksek derecede ilişkili olan lineer kombinasyonlarını, veya

$$\rho = kor\left(a'\mathrm{x}^{(1)}, b'\mathrm{x}^{(2)}\right)$$

maksimum olacak şekilde a ve b yi bulmak istiyoruz. ρ ya bir "kanonik korelasyon" denir ve $u = a'\mathrm{x}^{(1)}$, $v = b'\mathrm{x}^{(2)}$ bir kanonik değişken çiftidir. Şimdi bunu formüle bağlayalım. $x^{(1)}$ bir rasgele p-vektördür, ve $x^{(2)}$ bir rasgele q-vektördür. Basit olması için, $p \le q$ olduğunu kabul edelim. $a_i, i = 1, 2, ..., p$ ortonormal p-vektörler ve $b_i, i = 1, 2, ..., p$ ortonormal q-vektörler olsun. $\rho_1 = \left(a_1^{'} x^{(1)}, \mathrm{b}_1^{'} x^{(2)}\right)$ maksimum olacak şekilde a_1 ve b_1 seçelim. $1 \le i \le k$ için ilişkisiz olan $a_i^{'} x^{(1)}, a_k^{'} x^{(1)}$ ve $b_i^{'} x^{(2)}, b_k^{'} x^{(2)}$ çiftine bağlı olarak, $\rho_k = kor\left(a_k^{'} x^{(1)}, b_k^{'} x^{(2)}\right)$ maksimum olacak şekilde a_k ve b_k 'yı seçelim. ρ_i ler kitle kanonik korelasyonlarıdır ve $u_i = a_i^{'} x^{(1)}, v_i = b_i^{'} x^{(2)}$ kitle kanonik değişken çiftleridir. ρ ne kadar büyük olursa olsun, en fazla p tane sıfırdan farklı kanonik korelasyon vardır. $x^{(1)}, \Sigma_{11}$ varyansına, $x^{(2)}, \Sigma_{22}$ varyansına sahiptir. $Kov\left(x^{(1)}, \mathrm{x}^{(2)}\right) = \Sigma_{12}$ dir ve $x = \begin{bmatrix} x^{(1)} \\ x^{(2)} \end{bmatrix}$ alındığında, $Var(x) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ dir. Bu takdirde, $Kor\left(a_1^{'} x^{(1)}, b_1^{'} x^{(2)}\right) = \dfrac{a_1^{'} \Sigma_{12}\, b_1}{\sqrt{(a_1^{'} \Sigma_{11}\, a_1)(b_1^{'} \Sigma_{22}\, b_1)}}$ dir. G_1 tersi alınabilir(tersinir) bir $p \times p$ matris olmak üzere, $\tilde{x}^{(1)} = G_1 x^{(1)}$ olsun. Aynı şekilde G_2 tersinir bir $q \times q$ matris olmak üzere, $\tilde{x}^{(2)} = G_2 x^{(2)}$ olsun. Bu takdirde

$$Var\left(\tilde{x}^{(1)}\right) = G_1 \Sigma_{11} G_1^{'}$$
$$Var\left(\tilde{x}^{(2)}\right) = G_2 \Sigma_{22} G_2^{'}$$
$$Kov\left(\tilde{x}^{(1)}, \tilde{x}^{(2)}\right) = G_1 \Sigma_{12} G_2^{'}$$

olur. A kare matrisi için $(A^{-1})^{'}$, A nın tersinin transpozesi olmak üzere, $\left(A^{-1}\right)^{'} = \left(A^{'}\right)^{-1}$ dir. $\tilde{a}_1 = \left(G_1^{-1}\right)' a_1$ ve $\tilde{b}_1 = \left(G_2^{-1}\right)' b_1$ olsun. Bu takdirde aşağıdaki sonucu elde ederiz.

$$Kor\left(\tilde{a}_1^{'} \tilde{x}^{(1)}, \tilde{b}_1^{'} \tilde{x}^{(2)}\right) = \frac{\tilde{a}_1^{'} G_1 \Sigma_{12} G_2^{'} \tilde{b}_1}{\sqrt{(\tilde{a}_1^{'} G_1 \Sigma_{11} G_1^{'} \tilde{a}_1)(\tilde{b}_1^{'} G_2 \Sigma_{22} G_2^{'} \tilde{b}_1)}}$$

$$= \frac{a_1' G_1^{-1} G_1 \Sigma_{12} G_2' (G_2')^{-1} b_1}{\sqrt{(a_1' G_1^{-1} G_1 \Sigma_{11} G_1' (G_1')^{-1} a_1)(b_1' G_2^{-1} G_2 \Sigma_{22} G_2' (G_2')^{-1} b_1)}}$$

$$= \frac{a_1' \Sigma_{12} b_1}{\sqrt{(a_1' \Sigma_{11} a_1)(b_1' \Sigma_{22} b_1)}}$$

$$= kor\left(a_1' x^{(1)}, b_1' x^{(2)}\right)$$

Bu sonuç, yenileri ve vs. den orijinal kanonik değişkenleri yeniden ele geçirebildiğimizden dolayı, orijinal kanonik değişkenlerin tersinir lineer kombinasyonlarının kanonik değişkenleri değiştirmediğini söyler. $\Sigma_{11} = \Sigma_{11}^{1/2}\left(\Sigma_{11}^{1/2}\right)'$ ve $\Sigma_{22} = \Sigma_{22}^{1/2}\left(\Sigma_{22}^{1/2}\right)'$ olduğunu farz edelim. $G_1 = \left(\Sigma_{11}^{1/2}\right)^{-1} = \Sigma_{11}^{-1/2}$ ve $G_2 = \left(\Sigma_{22}^{1/2}\right)^{-1} = \Sigma_{22}^{-1/2}$ olsun. Bu takdirde,

$$\text{var}\begin{bmatrix} \tilde{x}^{(1)} \\ \tilde{x}^{(2)} \end{bmatrix} = \begin{bmatrix} \Sigma_{11}^{-1/2} \Sigma_{11} \left(\Sigma_{11}^{-1/2}\right)' & \Sigma_{11}^{-1/2} \Sigma_{12} \left(\Sigma_{22}^{-1/2}\right)' \\ \Sigma_{22}^{-1} \Sigma_{21} \left(\Sigma_{11}^{-1/2}\right)' & \Sigma_{22}^{-1/2} \Sigma_{22} \left(\Sigma_{22}^{-1/2}\right)' \end{bmatrix}$$

$$= \begin{bmatrix} I_p & \Sigma_{11}^{-1/2} \Sigma_{12} \left(\Sigma_{22}^{-1/2}\right)' \\ \Sigma_{22}^{-1/2} \Sigma_{22} \left(\Sigma_{11}^{-1/2}\right)' & I_p \end{bmatrix}$$

olur. Eğer $\|\tilde{a}\| = 1$ ve $\|\tilde{b}\| = 1$ seçersek, bu takdirde

$$\text{var}\left(\tilde{a}_1' \tilde{x}^{(1)}\right) = \tilde{a}_1' I \tilde{a}_1 = 1$$

ve

$$\text{var}\left(\tilde{b}_1' \tilde{x}^{(2)}\right) = \tilde{b}_1' I \tilde{b}_1 = 1$$

dır. Çünkü $\tilde{a}_1'$ ve $\tilde{a}_1$ ortonormal olduklarından normları 1 dir. Bu nedenle birim (bir) uzunluklu $\tilde{a}_1$ ve $\tilde{b}_1$ ile, en büyük korelasyon

$$\tilde{a}_1' \Sigma_{11}^{-1/2} \Sigma_{12} \left(\Sigma_{22}^{-1/2}\right)' \tilde{b}_1$$

maksimum olduğunda vuku bulur. Bu tamı tamına bir singüler değer ayrışımı (SDA) yapmadır. Şimdi burada tam singüler değer ayrışımını bir örnekle açıklayalım:

Örnek 3.1 (tam singüler değer ayrışımının bir örneği)

SDA; bir A dikdörtgen matrisinin, bir U ortogonal matrisi, bir S köşegen matrisi ve bir V ortogonal matrisinin transpozesinin çarpımı olarak yazılabileceğini söyleyen lineer cebirden bir teoreme dayanır. $U'U = I$, $V'V = I$, U -nun sütunları AA' nün ortonormal öz vektörleri, V nin sütunları $A'A$ nın ortonormal öz vektörleri ve S ; U dan veya V den azalan sırada öz değerlerin karekoklerini içeren bir köşegen matris olmak üzere, teorem genel olarak A nın

$$A_{mn} = U_{mm} S_{mn} V_{nn}^{'}$$

biçiminde yazılabileceğini söyler. Aşağıdaki örnek onun SDA sını hesaplamak için küçük bir matrise bu tanımın uygulanmasıdır.

$$A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

olsun. U yu bulmak için AA' ile işe başlarız. A nın transpozesi

$$A' = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}$$

matrisidir. Bu nedenle,

$$AA' = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$$

olur. Şimdi AA' nün karşılık gelen öz vektörlerini ve öz değerlerini bulalım. Öz vektörlerin $Av = \lambda v$ olarak tanımlandığını biliyoruz ve bunu AA' ye uygularsak,

$$\begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

elde ederiz. Bunu

$$11x_1 + x_2 = \lambda x_1$$
$$x_1 + 11x_2 = \lambda x_2$$

denklemlerinin bir kümesi olarak yeniden yazarız ve

$$(11-\lambda)x_1 + x_2 = 0$$
$$x_1 + (11-\lambda)x_2 = 0$$

elde etmek için yeniden düzenleriz. Katsayı matrisinin determinantını sıfıra eşitler ve λ için çözüm yapılırsa,

$$\begin{vmatrix} (11-\lambda) & 1 \\ 1 & (11-\lambda) \end{vmatrix} = 0 \ ,$$

$$(11-\lambda)(11-\lambda) - 1.1 = 0$$

$$(\lambda-10)(\lambda-12) = 0$$

$$\lambda = 10 \text{ ve } \lambda = 12$$

öz değerlerini buluruz. λ değerlerini $AA'v = \lambda v$ bağıntısında yerine koyarak öz vektörlerimizi elde ederiz. $\lambda = 10$ için;

$$(11-10)x_1 + x_2 = 0$$

$$x_1 = -x_2$$

olup bu x_2 nin her değerleri için doğrudur. Bu nedenle $x_1 = 1$ ve $x_2 = -1$ seçeceğiz. Çünkü bunlar küçüktür ve onlarla çalışmak daha kolaydır. Böylece biz $\lambda = 10$ öz değerlerine karşılık gelen $[1 \quad -1]$ öz vektörünü elde ederiz. $\lambda = 12$ için,

$$(11-12)x_1 + x_2 = 0$$

$$x_1 = x_2$$

elde ederiz ve öncekiyle aynı nedenle $x_1 = 1$ ve $x_2 = 1$ alacağız. Şimdi $\lambda = 12$ için $[1 \quad 1]$ öz vektörüne sahip oluruz. Bu öz vektörler karşılık gelen öz değerlerin büyüklüğüne göre düzenlenen bir matriste sütun vektörler olur. Başka bir deyişle, en büyük öz değerin öz vektörü birinci sütundur, bundan sonraki en büyük öz değerin öz vektörü ikinci sütundur ve en küçük öz değerin öz vektörü matrisimizin son sütunu oluncaya kadar devam ederiz. Aşağıdaki matriste $\lambda = 12$ için öz vektör birinci sütundur ve $\lambda = 10$ için öz vektör ikinci sütundur.

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Son olarak, Gram-Schmidt ortonormalleştirme yöntemini sütun vektörlere uygulayarak bu matrisi bir ortogonal matrise çevirmeliyiz. v_1'i normalleştirerek başlarız.

$$u_1 = \frac{v_1}{\|v_1\|} = \frac{[1 \quad 1]}{\sqrt{1^2+1^2}} = \frac{[1 \quad 1]}{\sqrt{2}} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

dır.

$$w_2 = v_2 - u_1 v_2 . u_1$$

$$\begin{aligned}
&= [1 \quad -1] - \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} . [1 \quad -1] . \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \\
&= [1 \quad -1] - [0 \quad 0] \\
&= [1 \quad -1]
\end{aligned}$$

hesaplar ve

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$$

matrisini elde etmek için $u_2 = \frac{w_2}{|w_2|} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$ normalleştirmesini yaparız. Burada $x.y$; x ve y vektörlerinin *iç çarpımını* veya *nokta çarpımını* gösterir. V nin hesaplanması da buna benzerdir. V matrisi $A'A$ ya dayanır, bu nedenle

$$A'A = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}$$

elde ederiz.

$$\begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

ile $A'A$ nın öz değerlerini buluruz. Bu

$$10x_1 + 2x_3 = \lambda x_1$$

$$10x_2 + 4x_3 = \lambda x_2$$

$$2x_1 + 4x_2 + (2-\lambda)x_3 = 0$$

olarak yeniden yazılır ve katsayılar determinantı sıfıra eşitleyerek

$$\begin{vmatrix} (10-\lambda) & 0 & 2 \\ 0 & (10-\lambda) & 4 \\ 2 & 4 & (2-\lambda) \end{vmatrix} = 0$$

dır. Buna göre,

$$\lambda(\lambda-10)(\lambda-12) = 0$$

dır. Buradan $\lambda = 0$, $\lambda = 10$ ve $\lambda = 12$ bulunur. Bunlar $A'A$ nın öz değerleridir.

λ yı $A'A\lambda = \lambda v$ bağıntısında yerine koyarak $\lambda = 12$ için öz vektör x_3 keyfi bir parametre olmak üzere ,$\begin{bmatrix} x_3 & 2x_3 & x_3 \end{bmatrix}$dır. Bu nedenle $x_3 = 1$ alındığında $\lambda = 12$ için, $v_1 = \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}$ dir. Benzer şekilde $\lambda = 10$ için öz vektör x_2 keyfi bir parametre olmak üzere $v_2 = \begin{bmatrix} -2x_2 & x_2 & 0 \end{bmatrix}$dır. $x_2 = -1$ alırsak $\lambda = 10$ için $v_2 = \begin{bmatrix} 2 & -1 & 0 \end{bmatrix}$ bir öz vektördür. Aynı şekilde $\lambda = 0$ için öz vektör x_1 bir parametre olmak üzere, $\begin{bmatrix} x_1 & 2x_1 & -5x_1 \end{bmatrix}$ dir. $x_1 = 1$ seçersek, $v_3 = \begin{bmatrix} 1 & 2 & -5 \end{bmatrix}$ $\lambda = 0$ için bir öz vektördür. Öz değerin büyüklüğüne göre v_1, v_2, v_3 vektörlerini bir matris içinde düzenleyerek

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & -1 & 2 \\ 1 & 0 & -5 \end{bmatrix}$$

elde ederiz ve bir ortonormal matrise çevirmek için Gram-Schmidt ortonormalleştirme yöntemini bu matrisin sütun vektörlerine uygulayarak,

$$V' = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{bmatrix}$$

elde ederiz.

S için sıfırdan farklı öz değerlerin kareköklerini alırız bunların en büyüğünü s_{11} ondan sonraki en büyüğünü s_{22} ve en küçüğünü de s_{mm} alarak onlar ile köşegeni oluştururuz. U nun ve V nin sıfır olmayan öz değerleri daima aynıdır. Bu nedenle onlardan birininkini almamız önemli değildir. İndirgenmiş singüler değer ayrışımı yerine tam singüler değer ayrışımı yapıyor olduğumuzdan, U ve V arasında çarpıma izin verecek uygun boyutlara sahip olacak şekilde, S ye bir sıfır sütun vektörü eklemeliyiz. S deki köşegen elemanlar A nın singüler değerleridir, U nun sütunlarına sol singüler vektörler denir ve V deki sütunlara sağ singüler vektörler denir.

$$S = \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix}$$

dır. Şimdi tüm parçaları elde ettik. Buna göre,

$$A_{mn} = U_{mm} S_{mn} V_{nn}^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

dır. Bu da görüldüğü gibi A nın kendisidir. Yani teori doğrulanmıştır.

Şimdi,

$$\Sigma_{11}^{-1/2} \Sigma_{12} \left(\Sigma_{22}^{-1/2} \right)' = UDV'$$

SDA ’nı yapalım. D’deki singüler değerler kanonik korelasyonlardır, ve U ve V nin sütunları a ve b katsayılarını verir.

$$U_i = \tilde{a}_i \quad , V_i = \tilde{b}_i .$$

$$a_i = \left(\Sigma_{11}^{-1/2}\right)\tilde{a}_i \quad , b_i = \left(\Sigma_{22}^{-1/2}\right)' \tilde{b}_i$$

ile orijinallerini elde ederiz. $i \neq j$ için

$$0 = \tilde{a}_i' \tilde{a}_j = a_i' \Sigma_{11} a_j$$

olduğuna dikkat edelim. $\tilde{a}_i$ ve $\tilde{a}_j$ nın ortogonalliği bize kanonik korelasyon için istenen $a_i' x^{(1)}$ ve $a_j' x^{(1)}$ ilişkisiz olduklarını söyler. Aynı şekilde, $b_i' x^{(2)}$ ile $b_j' x^{(2)}$ de ilişkisizdir. SDA nın özelliklerinden, $\tilde{a}_1$

$$\rho_1^2 \left(\Sigma_{11}^{1/2}\right)' \tilde{a}_1 = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-T/2} \left(\Sigma_{11}^{1/2}\right)' \tilde{a}_1$$

veya

$$\rho_1^2 \left(\Sigma_{11}^{1/2}\right)' a_1 = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-T/2} \left(\Sigma_{11}^{1/2}\right)' a_1$$

veya

$$\rho_1^2 a_1 = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} a_1$$

bağıntılarını sağlayan bir öz vektördür. Bu nedenle a_1, $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ in veya bir başka şekilde, $\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$in Σ_{11} ile ilgili bir öz vektörüdür. Benzer şekilde, b_1 $\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$ nin veya bir başka şekilde, $\Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$ nin Σ_{22} ile ilgili bir öz vektördür. A, i-yinci satırı a_i' olan bir matris olsun ve aynı şekilde, B, i-yinci satırı b_i' olan bir matris olsun. Bu takdirde $p = 3$ ve $q = 4$ için örneğimizdeki yolu izleyerek $\Sigma_{11}^{-1/2} \Sigma_{12} \left(\Sigma_{22}^{-1/2}\right)' = UDV'$ singüler değer ayrışımı yapıldığında,

$$kor\begin{bmatrix} Ax^{(1)} \\ Bx^{(2)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \rho_1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \rho_2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \rho_3 & 0 \\ \rho_1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \rho_2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \rho_3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

elde edilir. Buradan da görüldüğü gibi, ρ_1 en büyük korelasyondur, o halde ρ_1'i oluşturan çiftler kanonik korelasyon çiftleridir.

Notlar:

1.Kanonik korelasyon lineer kombinasyonları $x^{(1)}$ ve $x^{(2)}$ arasındaki korelasyonu (ilişkiyi) özetler; onlar $x^{(1)}$ ve $x^{(2)}$ deki değişimi özetlemenin kötü bir çalışmasını yapabilir.

2.Birinci kanonik korelasyon $x^{(1)}$ ve $x^{(2)}$ deki herhangi bir değişken çiftleri arasındaki ilişkiden daha büyüktür.

3.Örneklem kanonik korelasyonları, Σ nın yerine S ile çalışma hariç, aynı yapıya uyar.

4.KORELASYON KATSAYISI HAKKINDA EK YORUMLAR

Burada korelasyon katsayısı hakkında bazı yorumlara yer vereceğiz. Bu yorumlardan bazılarını geometrik olarak açıklayacağız.

4.1 Örneklem Korelasyonunun Bir Geometrik Yorumu

Korelasyon katsayısına geri dönerek, iki değişkenli gözlemlerin bir

$$\{(x_i, y_i) : i = 1, ..., n\}$$

kümesine sahip olduğumuzu farz edelim. Şimdi

$$x = (x_1, x_2, ..., x_n)'$$
$$y = (y_1, y_2, ..., y_n)'$$

vektörlerine n-boyutlu uzaydaki iki vektör olarak bakabiliriz. Şimdi n-boyutlu

$$j = (1, 1, ..., 1)'$$

birim vektörünü göz önüne alalım. Bu vektörlerle gerilen uzay yani $\Re(j)$ herhangi k reel sayısı için,

$$kj = (k, k, ..., k)'$$

biçimindeki tüm n-boyutlu vektörlerin bir topluluğudur. X vektörünün $\Re(j)$ üzerine izdüşümü (dik izdüşümü)

$$x^* = j(jj')^{-1} jx = j\left(\frac{1}{n}\right) j'\text{x} = \text{j}\overline{x} = \left(\overline{x}, \overline{x}, ..., \overline{x}\right)$$

dir. Burada $j'x = \sum x_i = n\overline{x}$ dir ve $\overline{x}$, $x_1, x_2, ..., x_n$ nin örneklem ortalamasıdır. Aynı şekilde, y nin $\Re(j)$ üzerinde izdüşümü

$$y^* = \overline{y} j = \left(\overline{y}, \overline{y}, ..., \overline{y}\right)'$$

olur. Burada $\overline{y}$; $y_1, y_2, ..., y_n$ nin örneklem ortalamasıdır. x ve onun izdüşümü arasındaki fark

$$x - x^* = \left(x_1 - \overline{x}, x_2 - \overline{x}, ..., x_n - \overline{x}\right)'$$

ve y ile onun ortalaması arasındaki fark

$$y-y^{*}=\left(y_1-\overline{y},y_2-\overline{y},...,y_n-\overline{y}\right)'$$

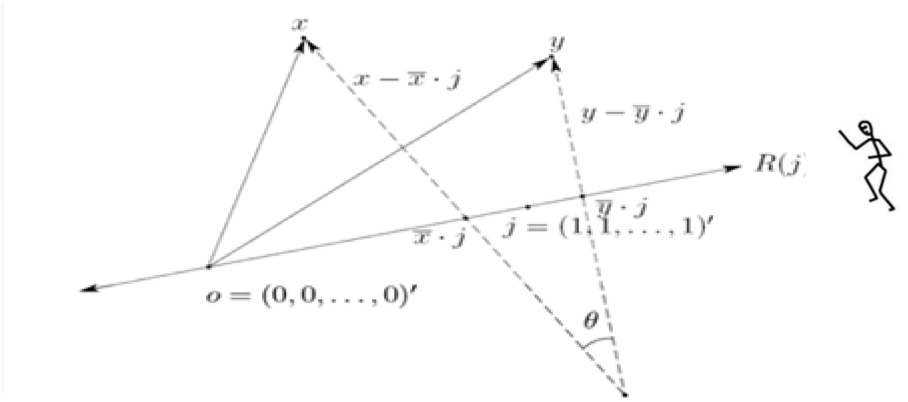
dir. Bu fark vektörleri ortalama sıfır olacak şekilde bir sabitle değiştirilen orijinal x ve y veri vektörleridir. Onlar, gözlemlerin onların ortalamasından sapmalarını gösterir. Şimdi örneklem korelasyonunu yorumlamaya hazırız. Korelasyon katsayısı sapmaların bu iki vektörün arasındaki açının kosinüsüdür. Görsel olarak x den $\Re(j)$ üzerine bir dik inildiğini ve y den $\Re(j)$ üzerine başka bir dik inildiğini düşünelim. Korelasyon katsayısı bu iki dik doğru yani $x-j\overline{x}$ ve $y-j\overline{y}$ arasındaki açının kosinüsüdür. $J=jj'$ olmak üzere bu iki doğrultuyu

$$x-j\overline{x}=x-\frac{1}{n}jj'x=x-\frac{1}{n}Jx=\left(I-\frac{1}{n}J\right)x$$

ve

$$y-j\overline{y}=y-\frac{1}{n}jj'y=y-\frac{1}{n}Jy=\left(I-\frac{1}{n}J\right)y$$

vektörlerinin doğrultuları ile temsil edebiliriz. Burada $\left(I-\frac{1}{n}J\right)$ matrisinin simetrik idempotent olduğuna dikkat edelim.



Şekil 4.1 Standartlaştırılmış değişkenler arasındaki korelasyonunun geometrik yorumu

Burada $R(A)$, A matrisinin sütunları tarafından gerilen vektör uzayını, yani A matrisinin sütun uzayını gösterir. Eğer x_i lerin onların ortalamasından sapmaları, y_i

lerin onların ortalamasından sapmaları ile birlikte iniş çıkışa meylederse, bu takdirde bu iki sapma vektörü aynı bir yönü gösterecek, o zaman arasındaki açı küçük olacak ve açının kosinüsü 1 e yakın olacak. Eğer iki değişken ilişkili değillerse sapma vektörlerinin her biri diğerine hemen hemen ortogonal ve açının kosinüsü sıfıra yakın olacak.

4.2 Kitle Korelasyonu Hakkında Çıkarımlar

r örneklem korelasyon katsayısı herhangi bir kabul edilebilir iki değişkenli dağılım altında ρ kitle korelasyon katsayısının bir tutarlı tahminidir. (Tutarlılık $n \to \infty$ iken r nin kesin olarak (1 olasılıkla) ρ ya yakınsamasını ifade eder. Bununla beraber, bunun ötesine de $H_0 : \rho = 0$ hipotezini test etmemiz veya ρ için bir güven aralığı hesaplamamız gerekir. $H_0 : \rho = 0$ hipotezini test etmek için bir yöntem r 'yi bir t-istatistiğine çevirmektir. Eğer y_i 'nin x_i üzerindeki bir basit lineer regresyonu uydursaydık, yani

$$y_i | x_i \sim N\left(\beta_0 + \beta_1 x_i, \sigma^2\right) \qquad (4.1)$$

olsaydı, bu takdirde eğimin sıfır olduğunu kabul eden sıfır hipotezinin, yani $H_0 : \beta_1 = 0$ hipotezinin bir testi sıfır korelasyonunun bir testine eşdeğer olacaktı. İki değişkenli normal modelin, yani;

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_{xx} & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_{yy} \end{pmatrix}\right)$$

nin,

$$\beta_1 = \rho\sigma_y / \sigma_x$$
$$\beta_0 = \mu_y - \beta_1\mu_x$$
$$\sigma^2 = \sigma_{yy}\left(1 - \rho^2\right)$$

olmak üzere x_i verildiğinde y_i nin şartlı dağılımının (4.1) ile ifade ettiğini biliyoruz. Bu nedenle $\beta_1 = 0$ olması için gerek ve yeter şart $\rho = 0$ olmasıdır. (4.1) basit lineer regresyon modelinin iki değişkenli normallikten daha genel olduğuna dikkat edelim.

Çünkü basit lineer regresyon modeli x_i açıklayıcısının normal dağılımlı olmadığı, yani o iki gösterge olduğunda, durumları kapsar.

Bir basit lineer regresyon modelinde $H_0 : \beta_1 = 0$ için test bir t-istatistiğine, yani

$$t = \hat{\beta}_1 \big/ SE\left(\hat{\beta}_1\right) \tag{4.2}$$

istatistiğine dayanır. Burada $\hat{\beta}_1$, β_1 rin alışılmış en küçük kareler tahmini ve $SE\left(\hat{\beta}_1\right)$ onun standart hatasıdır.

Bu istatistik $n-2$ serbestlik dereceli bir standart t-dağılımı ile karşılaştırılır. Şimdi sadece r örneklem korelasyon katsayısı ve (4.2) ile verilen t-istatistiği arasında bir bağıntının var olduğunu belirteceğiz. $sd = n-2$ olmak üzere, bu bağıntı

$$r^2 = \frac{t^2}{t^2 + sd} \tag{4.3}$$

dir. (4.3) bağıntısından

$$t^2 = \left(\frac{r^2}{1-r^2}\right) sd$$

elde ederiz. r ve β_1 aynı işarete sahip olduklarından, onu

$$t = işaret\left(r\right)\sqrt{\left(\frac{r^2}{1-r^2}\right) sd}$$

olarak da yazabiliriz. Eğer r yi bir t-istatistiğine çevirir ve t nin 2 den daha büyük, ve -2 den daha küçük olduğunu bulursak iki değişken arasındaki korelasyonun istatistiksel olarak anlamlı olduğunu söyleyebiliriz. Daha genel olarak, $-\left|t\right|$ de birikimli dağılım fonksiyonunu bulur ve iki-yanlı test için 2 ile çarparak, p-değerini hesaplayabiliriz.

4.3 ρ İçin Güven Aralıkları

$H_0 : \rho = 0$ sıfır hipotezini test etmek için $sd = n-2$ olmak üzere

$$t = işaret\left(r\right)\sqrt{\left(\frac{r^2}{1-r^2}\right) sd}$$

formülünü kullanarak, örneklem korelasyonunu bir t-istatistiğine çevirdik. $\rho = 0$ olduğunda, bu istatistik bir standart t-dağılımına sahiptir. Bir güven aralığı ortaya koymak için ρ nun keyfi değerleri için r nin dağılım özelliklerini anlamamız gerekir.

İki değişkenli normallik altında r nin örneklem dağılımı ilk olarak 1915'de Fisher tarafından ortaya konan ve biraz karmaşıktır. Bununla beraber, sonunda anlaşıldı ki, n büyüdüğünde bu dağılım yaklaşık olarak ρ ortalamalı ve

$$V(r) \approx \frac{\left(1-\rho^2\right)^2}{n-1}$$

varyanslı normal dağılımdır. ρ için bir güven aralığı oluşturmanın bir yolu, yukarıdaki varyans formülü içine r örneklem korelasyonunu koymak ve

$$r = \pm 1.96 \frac{\left(1-\rho^2\right)}{\sqrt{n-1}}$$

aralığını bir yaklaşık $\%95$ güven aralığı olarak olmaktır. Daha da iyi bir yöntem varyansını istikrara kavuşturmak için r ye bir dönüşüm uygulamaktır. r'nin varyansını gösteren yukarıdaki formül yaklaşık olarak $\left(1-E(r)^2\right)^2$ ile orantılıdır. Bu, r için varyans-istikrarlaştırma dönüşümünün ters hiperbolik tanjant, yani

$$g(r) = \tanh^{-1}(r) = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$$

olduğunu ifade eder. Bu dönüşüm ilk olarak Fisher (1921) tarafından sunuldu ve genellikle Fisher' in z dönüşümü olarak bilinir. Fisher iki değişkenli normallik altında, $g(r)$'nin yaklaşık olarak $g(\rho)$ ortalama ve $1/(n-3)$ varyans ile normal dağıldığını gösterdi. $g(\rho)$ için yaklaşık $\%95$ güven aralığı basit olarak

$$g(\rho) = \pm 1.96\sqrt{\frac{1}{n-3}}$$

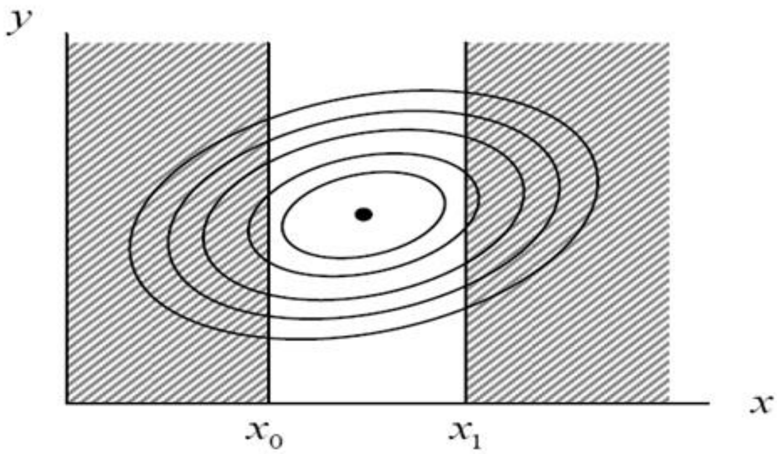
aralığıdır. Eğer $z = g(r)$ ise, bu takdirde ters dönüşüm

$$r = \tanh(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

dır. Bu nedenle, ρ için bir güven aralığı elde etmek için, $g(\rho)$ için aralığı bulabiliriz ve onun uç noktalarına $\tanh$ uygulayabiliriz. de hiperbolik tanjant ve onun tersi $\tanh$ ve $a\tanh$ olarak elde mevcuttur.

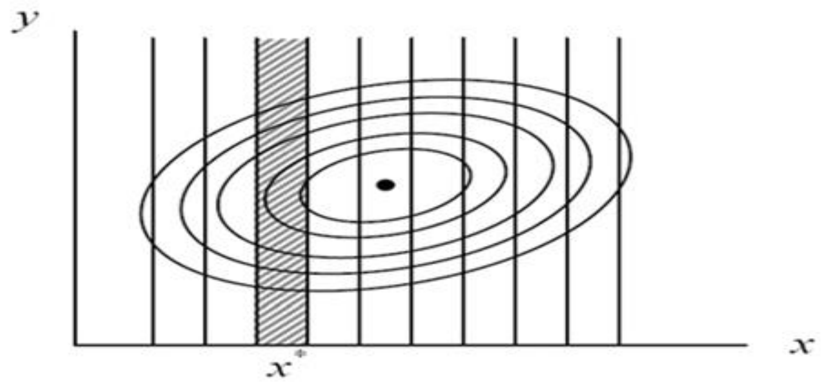
4.4 x in Korelasyonu ve Dağılımı

Eğer $(x_i, y_i), i = 1,...,n$ gözlemleri ilgili kitleden bir basit rasgele örneklem ise r örneklem korelasyonu ρ nun bir tutarlı tahminidir. Fakat örneklem kitlenin temsilcisi değilse, bu takdirde r çok yanlı bir tahmin olabilir. Özellikle eğer örnekleme mekanizması örneklemde kitledeki x lerden oldukça değişken x leri ortaya koyarsa r yanlı olacak. Örneğin, örneklemden $x < x_0$ veya olan herhangi bir birimi attığımızı farz edelim. Yani noktanın aşağıdaki gölgesiz bölgede bulunması kaydıyla (x_i, y_i) yi gözlemlediğimizi farz edelim. Bu takdirde örneklemden elde edilen r değeri ρ dan önemli derecede daha küçük olacaktır.



Şekil 4.2 x in korelasyonu ve dağılımı (1)

Eğer gölgeli bölgelerdeki her gözlemi belirleyici bir şekilde atmazsak fakat sadece onları alt örneklem yaparsak bu takdirde örneklemimizdeki x lerin varyansı yine x in kitledeki varyansından düşük olacak ve örneklem korelasyonu yine sıfıra yakın yanlı olacak. Öte yandan eğer örneklem seçim yöntemi kitledekinden daha büyük varyanslı x leri üretirse, bu takdirde örneklemdeki korelasyon kitledeki korelasyondan daha büyük olmaya meyledecek. Eğer bir birimin örneklemde içerilmesi olasılığı x e bağlı fakat y ye bağlı değilse, bu takdirde daha büyük kitlede ρ yu yanlı örneklemden tahmin etmek mümkündür. Aşağıdaki şekilde $x \approx x^*$ olduğu düşey şeridi düşünelim.



Şekil 4.3 x in korelasyonu ve dağılımı (2)

Eğer örneklem olasılıkları x e bağlı fakat y ye bağlı değilse, bu takdirde örneklemde bu şerit içine düşen noktaların oranı kitlede bu şerit içine düşen noktaların oranından çok farklı olabilir. Bununla beraber düşey şerit içinde y nin gözlenen değerleri $x = x^*$ verildiğinde y nin kitle şartlı dağılımından temsil edici örneklemdir. Eğer kitle iki değişkenli normal dağılım ise, bu takdirde $\beta_1 = \rho\sigma_y / \sigma_x$ ve $\beta_0 = \mu_y - \beta_1\mu_x$ olmak üzere, bu şartlı dağılımın ortalaması

$$E(y|x) = \beta_0 + \beta_1 x$$

ve şartlı varyans,

$$V(y|x) = \sigma_{yy.x} = \sigma_{yy}\left(1-\rho^2\right)$$

dir. Bu şartlı dağılımın üç parametresi yani β_0 sabiti β_1 eğimi ve $\sigma_{yy.x}$ hata varyansı örneklemde y nin x üzerindeki basit lineer regresyonundaki tutarlı olarak tahmin edilebilir. Bu tahminler bir kez elde mevcutsa ρ için kastedilen tahmini elde etmek için geriye doğru çözüm bulabiliriz. Küçük bir cebirsel işlemle σ_{xx} ilgili kitlede (örneklemde değil) x in varyansı olmak üzere,

$$\rho = \sqrt{\frac{\beta_1^2 \sigma_{xx}}{\sigma_{yy.x} + \beta_1^2 \sigma_{xx}}}$$

dir. Bu formülü uygulamak için kitlede x in değişkenliği hakkında bazı şeyleri bilmemiz gerekir. Bunu çoğu kez biliriz. y nin x üzerindeki lineer regresyonunda yani

$$y_i \,|\, x_i \sim N\left(\beta_0 + \beta_1 x_i, \sigma_{yy.x}\right)$$

dan $\left(\beta_0, \beta_1 \sigma_{yy\,x}\right)$ parametrelerinin tahminlerinin nasıl hesaplanacağı hakkında henüz bir şey demedik. Bu tahminler için formüllere vardır. Bununla beraber bu formülleri bilmeksizin istatistiksel yazılımı kullanarak bu tahminleri yine hesaplayabiliriz.

5. BİR DEĞİŞİK BAKIŞLA KORELASYON KATSAYISINI ELDE ETME

Burada bazı lineer cebir kavramlarını kullanarak korelasyon katsayısını elde edeceğiz.

5.1 Başka Bir Bakışla Korelasyon Katsayısının Ortaya Konması

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \;,\; \boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \; j_{n\times 1} = (1,1,\ldots,1)' \;,\; J = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}_{n\times n} = jj' \quad (j'j = n)$$

$$\boldsymbol{x}^* = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix} = \left(I - \frac{1}{n} J \right)\boldsymbol{x} = \boldsymbol{x} - J\bar{\boldsymbol{x}} \qquad \boldsymbol{y}^* = \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix} = \left(I - \frac{1}{n} J \right)\boldsymbol{y} = \boldsymbol{y} - J\bar{\boldsymbol{y}}$$

olmak üzere

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

dir.

Açıklama: $\lambda_1 \boldsymbol{x} + \lambda_2 \boldsymbol{y} = 0 \Rightarrow \lambda_1 \bar{\boldsymbol{x}} + \lambda_2 \bar{\boldsymbol{y}} = 0 \Rightarrow \lambda_1 \underbrace{(\boldsymbol{x} - \boldsymbol{j}\bar{x})}_{x^*} + \lambda_2 \underbrace{(\boldsymbol{y} - \boldsymbol{j}\bar{y})}_{y^*} = \boldsymbol{0}$

örneklem korelasyon katsayısıdır. Eğer, $\lambda_1 \boldsymbol{x}^* + \lambda_2 \boldsymbol{y}^* = \boldsymbol{0}$ denklemini sağlayan en az bir $\lambda_i \neq 0 \;, i = 1,2,\ldots,n$ varsa $\boldsymbol{x}^*$ ve $\boldsymbol{y}^*$ vektörleri lineer bağımlıdırlar. Bu denklem yalnız $\lambda_1 = \lambda_2 = 0$ için sağlanıyorsa, $\boldsymbol{x}^*$ ve $\boldsymbol{y}^*$ vektörleri lineer bağımsızdırlar. $\lambda_1 \boldsymbol{x}^* + \lambda_2 \boldsymbol{y}^* = \lambda_1 \boldsymbol{x} - \lambda_1 \boldsymbol{j}\bar{x} + \lambda_2 \boldsymbol{x} - \lambda_2 \boldsymbol{j}\bar{y} = \boldsymbol{0}$ denklemine denktir. Eğer $\boldsymbol{x}^*$ ve $\boldsymbol{y}^*$ vektörleri lineer bağımlı (ya da bağımsız) ise $\boldsymbol{x}$ ve $\boldsymbol{y}$ vektörleri de lineer bağımlı (ya da bağımsız) dır. Çünkü $\lambda_1 \boldsymbol{x} + \lambda_2 \boldsymbol{y} = \lambda_1 \boldsymbol{j}\bar{x} + \lambda_2 \boldsymbol{j}\bar{y}$ olur. (Özellikle $\lambda_1 = \lambda_2 = 0$ ise, $\boldsymbol{x}$ ve $\boldsymbol{y}$ lineer bağımsızdır)

Bu nedenle $\lambda_1 \boldsymbol{x}^* + \lambda_2 \boldsymbol{y}^* = 0 \Rightarrow \begin{bmatrix} \begin{bmatrix} x_1-\bar{x} \\ x_2-\bar{x} \\ \vdots \\ x_n-\bar{x} \end{bmatrix} & \cdots & \begin{bmatrix} y_1-\bar{y} \\ y_2-\bar{y} \\ \vdots \\ y_n-\bar{y} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ yani $A\lambda = 0$

homojen denklemi sistemine verilir. Bu sistem daima tutarlı yani, bir çözüme sahiptir. $\boldsymbol{A}$ katsayılar matrisi sütun ranklı, yani $r(\boldsymbol{A}) = r(\boldsymbol{A}'\boldsymbol{A}) = 2$ ise $\boldsymbol{x}^*$ ve $\boldsymbol{y}^*$ vektörleri lineer bağımsızdır. Çünkü bu durumda sistemi sağlayan $\lambda_1 = \lambda_2 = 0$ çözümünden başka bir çözüm yoktur.

$$\boldsymbol{A}'\boldsymbol{A} = \begin{bmatrix} \sum(x_i-\bar{x})^2 & \sum(x_i-\bar{x})(y_i-\bar{y}) \\ \sum(x_i-\bar{x})(y_i-\bar{y}) & \sum(x_i-\bar{x})^2 \end{bmatrix}$$

olduğundan, $\det(\boldsymbol{A}'\boldsymbol{A}) = \sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2 - \left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]^2$ dir. Eğer $\sum(x_i-\bar{x})(y_i-\bar{y}) = 0$ ise, kareler toplamları sıfır olamayacağından $\det(\boldsymbol{A}'\boldsymbol{A}) \neq \boldsymbol{0}$ ve $r(\boldsymbol{A}'\boldsymbol{A}) = 2$ olacağından $\boldsymbol{x}^*$ ve $\boldsymbol{y}^*$ vektörleri lineer bağımsız olurlar.

Çünkü bu durumda, $\det(\boldsymbol{A}'\boldsymbol{A}) = \sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2$ ye eşittir. Genel olarak

$$\det(\boldsymbol{A}'\boldsymbol{A}) = \sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2 - \left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]^2 = s_x^2 s_y^2 - s_{xy} = k \in \mathbb{R} \quad \text{dir.}$$

Buradan,

$$\sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2 - k = \left[\sum(x_i-\bar{x})(y_i-\bar{y})\right]^2$$

elde edilir. Bu ise,

$$1 - \frac{k}{s_x^2 s_y^2} = \left[\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{s_x^2 s_y^2}\right]^2 \Rightarrow r_{xy} = \mp\sqrt{1 - \frac{k}{s_x^2 s_y^2}} \quad , \quad \left(r_{xy} = \cos(\boldsymbol{x}^*, \boldsymbol{y}^*)\right)$$

olduğunu gösterir.

$0 \le r_{xy}^2 \le 1$ olduğundan, $0 \le 1 - \dfrac{k}{s_x^2 s_y^2} \le 1$ dır, yani $0 \le \dfrac{k}{s_x^2 s_y^2} \le 1$ dır. k ; pozitif yarı tanımlı bir matrisin determinantı olduğundan $k \ge 0$ dır. $k = 0$, yani $\det\left(\boldsymbol{A}'\boldsymbol{A}\right) = \boldsymbol{0}$ ya da $r\left(\boldsymbol{A}\right) < 2$ ise, $\boldsymbol{x}^*$ ve $\boldsymbol{y}^*$ arasında mükemmel bir lineer bağıntı (ilişki) vardır. $1 - \dfrac{k}{s_x^2 s_y^2} = 0$ yani $\dfrac{k}{s_x^2 s_y^2} = 1$ya da $\det\left(\boldsymbol{A}'\boldsymbol{A}\right) = k = s_x^2 s_y^2$ ise, bu takdirde $r\left(\boldsymbol{A}\right) = 2$ ya da $\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) = 0$ olduğu görülür. Bu durumda $\boldsymbol{x}^*$ ve $\boldsymbol{y}^*$ lineer bağımsızdırlar. $s_x^2 s_y^2 - \left[\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right]^2 = 0$ ise $\boldsymbol{x}^*$ ve $\boldsymbol{y}^*$ lineer bağımlıdır.

NOT: $\boldsymbol{x}$ ile $\boldsymbol{y}$ arasındaki lineer ilişki $\boldsymbol{x}$ üzerinde $\boldsymbol{y}$ nin regresyonundan, $\beta \ne 0$ olmak üzere, $y_i = \alpha + \beta x_i + \varepsilon_i$, $i = 1, 2, ..., \mathrm{n}$ basit lineer regresyon doğrusuyla belirlenir. $\boldsymbol{x} - j\bar{\boldsymbol{x}}$ ile $\boldsymbol{y} - \bar{\boldsymbol{y}}$ arasındaki lineer ilişkide de aynı $\beta \ne 0$ olmalıdır. Gerçekten, $y_i = \alpha + \beta x_i + \varepsilon_i$ $i = 1, 2, ..., \mathrm{n} \Rightarrow \bar{\boldsymbol{y}} = \alpha + \beta\bar{\boldsymbol{x}}$, $y_i - \bar{y_i} = \beta\left(x_i - \bar{x_i}\right) + \varepsilon_i$, $\boldsymbol{y} - \bar{\boldsymbol{y}} = \beta\left(\boldsymbol{x} - \bar{\boldsymbol{x}}\right)$, β aynı eğimdir.

$r_{xy} = \dfrac{\boldsymbol{x}^{*'}\boldsymbol{y}^*}{\left\|\boldsymbol{x}^*\right\|\left\|\boldsymbol{y}^*\right\|} = \cos\theta$ olduğundan, $-1 \le r_{xy} \le 1$ olacaktır. Burada θ , x^* ve y^* arasındaki pozitif yönlü açıdır. O halde $0 \le r_{xy}^2 \le 1$ dır. Bu ise, $0 \le 1 - \dfrac{k}{s_x^2 s_y^2} \le 1$ olduğunu gösterir. $\cos^2\theta = 1 - \dfrac{k}{s_x^2 s_y^2} \Rightarrow \dfrac{k}{s_x^2 s_y^2} = \sin^2\theta$ dir. $\theta = \pi$ ya da $\theta = 0$ ise bu durum gerçeklenir. O halde $\boldsymbol{x}^*$ ve $\boldsymbol{y}^*$ mükemmel lineer bağıntıya sahip olurlar. $\theta = \pi/2$ olduğunda, $\dfrac{k}{s_x^2 s_y^2} = 1 \Rightarrow k = s_x^2 s_y^2$ olduğunda $\boldsymbol{x}^*$ ve $\boldsymbol{y}^*$ lineer bağımsızdırlar.

Şimdi de, $s_x = \sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$, $s_y = \sqrt{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$ ve $s_{xy} = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)$ olmak üzere, $\dot{X}^* = \dfrac{1}{s_x}\left(X - j\bar{X}\right)$, $\dot{\boldsymbol{y}}^* = \dfrac{1}{s_y}\left(\boldsymbol{y} - \boldsymbol{j}\bar{\boldsymbol{y}}\right)$ alalım. Bu durumda,

$$\overset{\bullet}{\boldsymbol{x}}^{*\prime}\overset{\bullet}{\boldsymbol{x}}^{*}=\frac{1}{s_x^2}\underbrace{\boldsymbol{x}'\left(\boldsymbol{J}-\frac{1}{n}\boldsymbol{J}\right)\boldsymbol{x}}_{s_x^2}=1 \text{ ve } \overset{\bullet}{\boldsymbol{y}}^{*\prime}\overset{\bullet}{\boldsymbol{y}}^{*}=\frac{1}{s_y^2}\underbrace{\boldsymbol{y}'\left(J-\frac{1}{n}J\right)\boldsymbol{y}}_{s_y^2}=1$$

olacaktır. Yukarıdaki aynı düşünceyle $\lambda_1\overset{\bullet}{\boldsymbol{x}}^{*}+\lambda_2\overset{\bullet}{\boldsymbol{x}}^{*}=\boldsymbol{0}$, yani

$$\left[\frac{1}{s_x}\begin{bmatrix}x_1-\bar{x}\\x_2-\bar{x}\\\vdots\\x_n-\bar{x}\end{bmatrix}\frac{1}{s_y}\begin{bmatrix}y_1-\bar{y}\\y_2-\bar{y}\\\vdots\\y_n-\bar{y}\end{bmatrix}\right]\begin{bmatrix}\lambda_1\\\lambda_2\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix} \tag{5.1}$$

denklemi en az bir $\lambda_i\neq 0$, $i=1,2$ için sağlanırsa $\overset{\bullet}{\boldsymbol{x}}^{*}$ ve $\overset{\bullet}{\boldsymbol{y}}^{*}$ vektörleri lineer bağımlı ve $\lambda_1=\lambda_2=0$ için sağlanırsa lineer bağımsızdır. Şimdi,

$$\boldsymbol{B}=\left[\frac{1}{s_x}\begin{bmatrix}x_1-\bar{x}\\x_2-\bar{x}\\\vdots\\x_n-\bar{x}\end{bmatrix},\frac{1}{s_y}\begin{bmatrix}y_1-\bar{y}\\y_2-\bar{y}\\\vdots\\y_n-\bar{y}\end{bmatrix}\right]$$

alalım. Eğer $r(\boldsymbol{B})=2$ ise, (5.1) sistemi yalnız sıfır çözüme sahiptir ve vektörler lineer bağımsızdırlar. $r(\boldsymbol{B})=r(\boldsymbol{BB}')$ olduğundan, $\det(\boldsymbol{BB}')\neq 0$ ise $r(\boldsymbol{B})=2$ olacağından, bu vektörlerin lineer bağımsız olmaları için gerek ve yeter şart:

$$|\boldsymbol{BB}'|=\begin{vmatrix}1 & \frac{1}{s_xs_y}\sum(x_i-\bar{x})(y_i-\bar{y})\\ \frac{1}{s_xs_y}\sum(x_i-\bar{x})(y_i-\bar{y}) & 1\end{vmatrix}$$

$$=1-\underbrace{\frac{1}{s_x^2s_y^2}\left[\sum(x_i-\bar{x})(y_i-\bar{y})\right]^2}_{r_{xy}^2}\neq 0$$

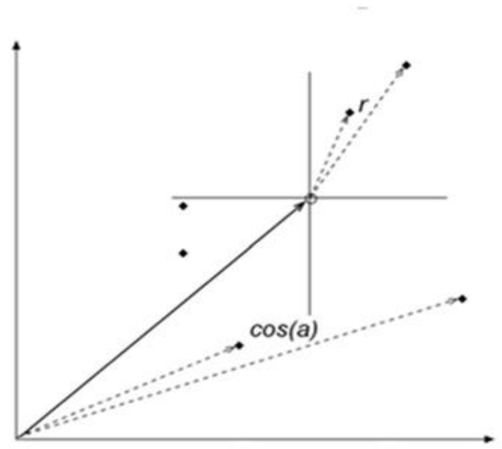
olmasıdır. r_{xy}, $\overset{\bullet}{\boldsymbol{x}}^{*}$ ve $\overset{\bullet}{\boldsymbol{y}}^{*}$ vektörleri arasındaki açının (φ diyelim) kosinüsü olduğundan, tam bağımsızlık için $1-\cos^2\varphi=\sin^2\varphi\neq 0$ olmalıdır. $\varphi=\frac{\pi}{2}$ olduğunda tam lineer bağımlılık vardır. $\varphi=0$ veya $\varphi=\pi$ olduğunda tam lineer bağımlıdır.

6.PEARSON'UN *r* KORELASYON KATSAYISI VE SALTON'UN KOSİNÜS ÖLÇÜSÜ ARASINDAKİ İLİŞKİ

Jones ve Furnas (1987) Salton' un kosinüsü ve Pearson' un korelasyon katsayısı arasındaki farkı geometrik terimlere göre açıkladılar ve her iki ölçüyü diğer birtakım benzerlik ölçütleri ile karşılaştırdılar.

6.1 Pearson' un *r* Korelasyon Katsayısı ve Salton'un Kosinüs Ölçüsü Arasındaki İlişkinin Geometrik Yorumu

Pearson korelasyonu, vektörlerin değerlerini onların ortalamalarına göre normalleştirir. Geometrik terimlere göre, kosinüs tüm vektörlerin bir sıfır değere sahip olduğu bir orijinden vektör uzayı oluştururken, bu; vektör uzayının orijininin kümenin ortasında yerleştirildiğini ifade eder. Bu durum Şekil 6.1 de gösterilmiştir



Şekil 6.1 Pearson' un *r* si ve Salton' un kosinüsü arasındaki ilişki

Sonuç olarak, kosinüs sadece bir tek dörtte bir dilimde sıfırdan bire değişirken Pearson korelasyonu -1 den 1 re değişebilir. (Eğer biri sadece pozitif değerleri kullanmayı isterse, $(r+1)/2$ yi kullanarak korelasyonun değerlerini lineer olarak dönüştürebilir.) Bununla beraber birçok uygulamalı durumlarda Pearson' un korelasyon katsayısı ve Salton' nun kosinüsünü kullanma arasındaki farklar önemsenmeyebilir.

6.2 Problemin Formüllemesi

$\boldsymbol{x}=(x_1,x_2,...,x_n)'$ ve $\boldsymbol{y}=(y_1,y_2,...,y_n)'$ tüm koordinatların pozitif olduğu iki vektör olsun. Bu iki vektörün Jaccard indeksi (sayısı) (bu vektörlerin benzerliğini ölçen)

$$J=\frac{\boldsymbol{x}.\boldsymbol{y}}{\|\boldsymbol{x}\|_2^2+\|\boldsymbol{y}\|_2^2-\boldsymbol{x}.\boldsymbol{y}} \tag{6.1}$$

$$J=\frac{\sum_{i=1}^{n}x_iy_i}{\sum_{i=1}^{n}x_i^2+\sum_{i=1}^{n}y_i^2-\sum_{i=1}^{n}x_iy_i} \tag{6.2}$$

olarak tanımlanır. Burada $\boldsymbol{xy}=\sum_{i=1}^{n}x_iy_i$, $\boldsymbol{x}$ ve $\boldsymbol{y}$ vektörlerin iç çarpımıdır. $\|\boldsymbol{x}\|_2=\sqrt{\sum_{i=1}^{n}x_i^2}$ ve $\|\boldsymbol{y}\|_2=\sqrt{\sum_{i=1}^{n}y_i^2}$ sırasıyla $\boldsymbol{x}$ ve $\boldsymbol{y}$ vektörlerinin Öklit normlarıdır (L^2 normlar da denir). Salton' un kosinüs ölçüsü, yukarıdaki aynı notasyonla

$$\cos(\theta)=\frac{\boldsymbol{x}.\boldsymbol{y}}{\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2} \tag{6.3}$$

$$\cos(\theta)=\frac{\sum_{i=1}^{n}x_iy_i}{\sqrt{\sum_{i=1}^{n}x_i^2}\sqrt{\sum_{i=1}^{n}y_i^2}} \tag{6.4}$$

olarak tanımlanır. Burada $\boldsymbol{\theta}$; $\boldsymbol{x}$ ve $\boldsymbol{y}$ vektörlerinin arasında pozitif yönlü açıdır. Diğer sonuçlar arasında eğer $\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2$ ise, bu takdirde deneysel bulgularla tamamen uyuşan basit bir

$$J=\frac{\cos(\theta)}{2-\cos(\theta)} \tag{6.5}$$

bağıntısını ispatlayabiliriz. Dice' nin E ölçüsü için: yani

$$E=\frac{2\boldsymbol{x}.\boldsymbol{y}}{\|\boldsymbol{x}\|_2^2+\|\boldsymbol{y}\|_2^2} \tag{6.6}$$

$$E = \frac{2\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2} \tag{6.7}$$

için eğer $\|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2$ ise, $E = \cos(\theta)$ elde ettiğimizi bile gösterebilirdik. Aynısı diğer birçok benzerlik ölçüsü için gösterilebilir. Diğer benzerlik ölçüleri için bkz. Boyce, Meadov ve Kraft (1955), Tague-Sutcliffe (1995), Grossman ve Frieder (1998); Loose (1998); Salton ve McGill (1987) ,Van Rijsebergen (1979) ,Egghe ve Michel (2002, 2003). Egghe (2008) diğer ölçülerle Pearson korelasyon katsayısının ilişkisi problemine değindi. r nin tanımı

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} \tag{6.8}$$

dir. Bu çalışmada Pearson' un korelasyon katsayısı ve Salton' un kosinüsü arasındaki ilişki hakkında bu geriye kalan bu soruya değineceğiz.

Problem, Pearson korelasyon katsayısının tanımında $\boldsymbol{x} = (x_1, x_2, ..., x_n)'$ ve $\boldsymbol{y} = (y_1, y_2, ..., y_n)'$ vektörlerinin L^2 normları ve bu vektörlerin L^1 normlarının eş anlı vukuu bulmasında yatar. L^1 normları aşağıdaki gibi tanımlanır.

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} x_i \tag{6.9}$$

$$\|\boldsymbol{y}\|_1 = \sum_{i=1}^{n} y_i \tag{6.10}$$

r ve $cos(\theta)$ arasındaki ilişkiyi belirlemek için temel, aşikar olarak, $\boldsymbol{x}$ ve $\boldsymbol{y}$ vektörlerinin L^1 ve L^2 normları arasındaki ilişki olacaktır. Bundan sonraki kısımda $a = \frac{\|\boldsymbol{x}\|_1}{\|\boldsymbol{x}\|_2}$ ve $b = \frac{\|\boldsymbol{y}\|_1}{\|\boldsymbol{y}\|_2}$ nin her sabit değerinin r ve $cos(\theta)$ arasındaki bir ilişkiyi verdiğini göstereceğiz.

6.3 r ve $\cos(\theta)$ Arasındaki İlişki için Matematiksel Model

$\boldsymbol{x} = (x_1, x_2, ..., x_n)'$ ve $\boldsymbol{y} = (y_1, y_2, ..., y_n)'$ n uzunluğa sahip iki vektör olsun.

$$a = \frac{\|\boldsymbol{x}\|_1}{\|\boldsymbol{x}\|_2} \tag{6.11}$$

ve

$$b = \frac{\|\boldsymbol{y}\|_1}{\|\boldsymbol{y}\|_2} \tag{6.12}$$

olsun. Aşikar olarak $a \geq 1$ ve $b \geq 1$ olduğuna dikkat edelim. Aynı zamanda $a < \sqrt{n}$ ve $b < \sqrt{n}$ olduğunu da elde ettik. Gerçekten, Cauchy-Schwartz eşitsizliğine göre (bkz. Hardy, Littlewood ve Polya, 1988)

$$\begin{aligned} \|\boldsymbol{x}\|_1 &= \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} 1x_i \\ &\leq \left(\sum_{i=1}^{n} 1\right)^{1/2} \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} \\ &= \sqrt{n}\,\|\boldsymbol{x}\|_2 \end{aligned}$$

elde ederiz. Bu nedenle,

$$a = \frac{\|\boldsymbol{x}\|_1}{\|\boldsymbol{x}\|_2} \leq \sqrt{n}$$

dir. Fakat eğer $\boldsymbol{x}$ nün sabit vektör olmadığını farz edersek, $a \neq \sqrt{n}$, bu nedenle yukardan, $a < \sqrt{n}$ elde ederiz. $\boldsymbol{y}$ vektörü için aynı muhakeme sürdürülerek, $b < \sqrt{n}$ elde edilir. Böylece aşağıdaki sonucu elde ederiz.

Yardımcı Teorem 6.3.1: (6.11) ve (6.12) verildiğinde ne $\boldsymbol{x}$ ne de $\boldsymbol{y}$ sabit vektörler ise aşağıdaki bağıntı genel olarak geçerlidir.

$$r = \frac{n}{\sqrt{n-a^2}\sqrt{n-b^2}}\left(\cos(\theta) - \frac{ab}{n}\right) \tag{6.13}$$

Yukarıya göre kök altındaki sayıların pozitif olduklarına dikkat edelim. (ve kesin olarak pozitif olan ne $\boldsymbol{x}$ ne de $\boldsymbol{y}$ sabit değildir.)

İspat:

$$Y\cos(\theta) = \frac{\sum_{i=1}^{n} x_i y_i}{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)} \tag{6.14}$$

"Yalancı Kosinüs" ($Y\cos$) ölçüsünü tanımlayalım. Jones ve Fumas (1987) 'deki önceki tanımları bulabiliriz. (6.2.3) formülünde (herkes tarafından bilinen $\boldsymbol{x}$ ve $\boldsymbol{y}$ vektörleri arasındaki açının reel kosinüsü) $\|\boldsymbol{x}\|_2$ ve $\|\boldsymbol{y}\|_2$ yerlerine sırasıyla $\|\boldsymbol{x}\|_1$ ve $\|\boldsymbol{y}\|_1$ koyduğumuzdan dolayı, ölçüye "Yalancı Kosinüs" denir. Bu nedenle (6.2.4) ve (6.3.4) den görüleceği gibi (6.3.1) ve (6.3.2)' yi kullanarak,

$$\frac{\cos(\theta)}{Y\cos(\theta)} = \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

$$\frac{\cos(\theta)}{Y\cos(\theta)} = \frac{\|\boldsymbol{x}\|_1 \|\boldsymbol{y}\|_1}{\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2} = ab \tag{6.15}$$

elde ederiz. Şimdi, ne $\boldsymbol{x}$ ne de $\boldsymbol{y}$ sabit olmadığından (gelecek ifadede $\frac{0}{0}$ dan sakınma) (6.11), (6.12) ve (6.14) den

$$\frac{r}{\cos(\theta)} = \frac{n - \dfrac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sum_{i=1}^{n} x_i y_i}}{\sqrt{n - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{\sum_{i=1}^{n} x_i^2}}\sqrt{n - \dfrac{\left(\sum_{i=1}^{n} y_i\right)^2}{\sum_{i=1}^{n} y_i^2}}}$$

$$\frac{r}{\cos(\theta)} = \frac{n - \dfrac{1}{Y\cos(\theta)}}{\sqrt{n-a^2}\sqrt{n-b^2}}$$

elde ederiz. Şimdi (6.15) den

$$\frac{r}{\cos(\theta)} = \frac{n - \dfrac{ab}{\cos}}{\sqrt{n-a^2}\sqrt{n-b^2}}$$

elde ederiz ki; buradan

$$\cos(\theta) = \frac{\sqrt{n-a^2}\sqrt{n-b^2}r + ab}{n} \tag{6.16}$$

olarak yeniden elde edilebilir. (6.16) dan r yi çözerek (6.13) ün doğru olduğunu görürüz. (6.13) ün r ve $\cos(\theta)$ arasında bir lineer ilişki olduğuna fakat a ve b parametrelerine bağlı olduğuna dikkat edelim. ($\boldsymbol{x}$ ve $\boldsymbol{y}$ vektörlerinin uzunluğu olan n nin sabit olduğuna dikkat edelim). $\cos(\theta) = 0$ olması için gerek ve yeter şartın

$$r = -\frac{ab}{\sqrt{n-a^2}\sqrt{n-b^2}} < 0 \tag{6.17}$$

ve $r = 0$ olması için gerek ve yeter şartın

$$\cos(\theta) = \frac{ab}{n} > 0 \tag{6.18}$$

olması gerektiğine dikkat edelim.

Formüllerin her ikisi de a ve b değişkenine göre değişirler, fakat (6.17) daima negatif ve (6.18) daima pozitiftir. Bu nedenle değişen a ve b için git gide artan bir doğru demeti elde ettik.

Bu nedenle, değişen a ve b için bir artan doğrular demeti elde ettik. Uygulamada a ve b kesin olarak değişeceğinden (yani $\dfrac{\|\boldsymbol{x}\|_1}{\|\boldsymbol{x}\|_2}$ sayıları tüm vektörler için aynı olmayacağından) burada r ve $\cos(\theta)$ arasındaki bağıntının bir fonksiyonel bağıntı olmadığını fakat artan bir noktalar bulutu gibi bir bağıntı olduğunu gösterdik. r=0 için noktalar bulutunun pozitif apsis değerli noktaların bir alanını kapsayacak iken $\cos(\theta)$=0 için noktalar bulutunun sıfırdan küçük ordinatlı noktaların bir alanını işgal edeceğini bekleyebiliriz. (Tüm vektör koordinatları pozitif iken $\cos(\theta) \geq 0$ olduğundan bu açıktır). Aynı zamanda, (6.17) (onun mutlak değeri) ve (6.18) in vektörün uzunluğu olan n ile (sabit a ve b için) azaldıklarına da dikkat edelim. Bu aynı zamanda hemen görüldüğü gibi, (sabit a ve b için) n nin büyük değeri için, (6.13) ün eğiminin 1 e gitmesi durumdur.

6.4 Kosinüs İlişkisinden Başka r ve Diğer Benzerlik Ölçüleri Arasındaki İlişki

Girişte cos ve Jaccard, Dice vs. gibi diğer benzerlik ölçüleri arasındaki fonksiyonel ilişkileri belirtti. L^2 norm ilişkilerine yani $\|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2$ ye (ancak genelleştirmeler Egghe (2008) de verilir) dayanarak (*J*=Jaccard)

$$J = \frac{\cos(\theta)}{2 - \cos(\theta)} \tag{6.19}$$

$E = \cos(\theta)$ olduğunu ve Egghe (2008) de ele alınan diğer benzerlik ölçüleri için aynısının gerçeklendiği gösterebiliriz. Bu takdirde (6.13) ile bu sonuçların birleştirilmesinin r ve diğer ölçüler arasındaki ilişkileri verdiği açıktır. $E = \cos(\theta)$ olduğundan L^2 norm eşitliğinin yukarıdaki varsayımları altında $\cos(\theta)$ yerine E koyarak (6.13) ün yine geçerli olduğunu görürüz. J için (6.13) ve (6.4.1)'u kullanarak

$$\cos(\theta) = \frac{2J}{J+1} \tag{6.20}$$

elde ederiz ve buradan da,

$$r = \frac{n}{\sqrt{n-a^2}\sqrt{n-b^2}}\left(\frac{2J}{J+1} - \frac{ab}{n}\right) \tag{6.21}$$

bağıntısı r ile Jaccard arasındaki bağıntıdır.

7.PEARSON'UN r 'SİNİN YENİ BİR YORUMU

Bu kısımda Pearson' un r 'sinin yeni bir yorumunu geometriye dayanarak vereceğiz.

7.1 *r* nin Geometrik Yorumu

$\boldsymbol{z}_1$ ve $\boldsymbol{z}_2$, beklenen değerleri sıfır ve varyansları 1 olan iki iki-değişkenli normal rasgele değişkenler olsun. Bu takdirde

$$(\boldsymbol{z}_1+\boldsymbol{z}_2)^2=(\boldsymbol{z}_1-(-\boldsymbol{z}_2))^2$$

ifadesine mükemmel negatif korelasyondan uzaklığın karesi ($(umnk)^2$) ve $(\boldsymbol{z}_1-\boldsymbol{z}_2)^2$ 'ye de mükemmel pozitif korelasyondan uzaklığın karesi ($(umpk)^2$) olarak bakılır.

$$E(\boldsymbol{z}_1+\boldsymbol{z}_2)^2=E(\boldsymbol{z}_1^2)+2E(\boldsymbol{z}_1\boldsymbol{z}_2)+E(\boldsymbol{z}_2^2)$$

$$=1+2\rho+1=2(1+\rho)$$

dur. Burada ρ korelasyon parametresidir. Aynı şekilde, $E(\boldsymbol{z}_1-\boldsymbol{z}_2)^2=2(1-\rho)$ olur. Negatif uzaklığı "nu" ile gösterelim. Buna göre $(nu)^2=(\boldsymbol{z}_1+\boldsymbol{z}_2)^2$ dir ve pozitif uzaklığı "pu" ile gösterelim. $(pu)^2=(\boldsymbol{z}_1-\boldsymbol{z}_2)^2$ dır. Yani $(nu)^2$ ve $(pu)^2$ sırasıyla, negatif ve pozitif korelasyondan uzaklıkların karesidir. Buna göre $\rho=\dfrac{E(nu)^2-E(pu)^2}{4}$ dur. Aşağıdaki tablo bunu açıklar.

Çizelge 7.1. ρ korelasyon katsayısı, nu ve pu arasındaki ilişkiler

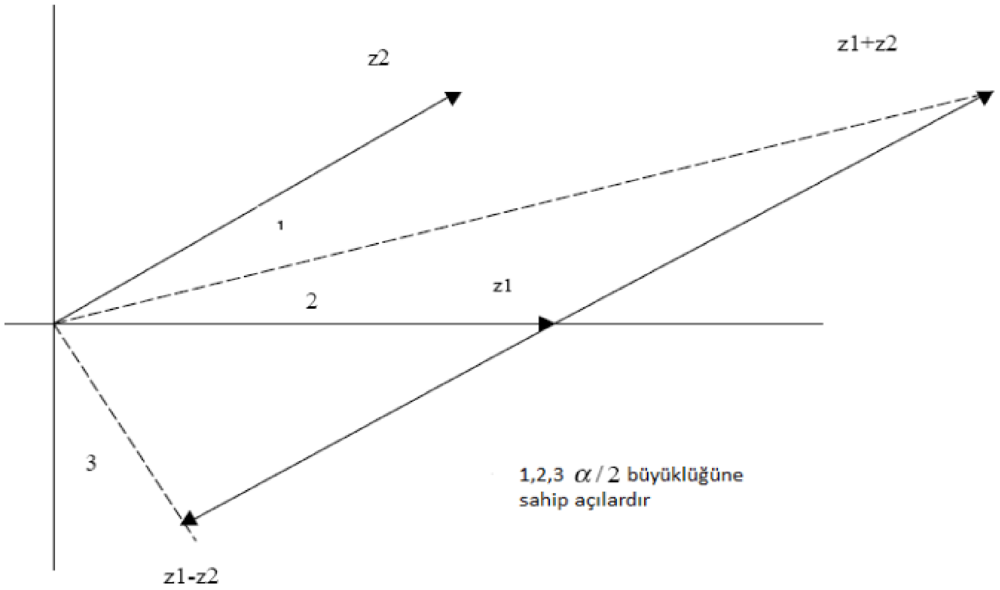
Korelasyon	$E\left[(nu)^2\right]=E(\boldsymbol{z}_1+\boldsymbol{z}_2)^2$	$E\left[(pu)^2\right]=E(\boldsymbol{z}_1-\boldsymbol{z}_2)^2$
-1	0	4
0	2	2
1	4	0

$\boldsymbol{x}, \boldsymbol{y}$ n-boyutlu veri vektörleri olsun. SS_x , SS_y ; ortalamalar etrafındaki yani $\bar{x}$, $\bar{y}$ etrafındaki, kareler toplamı olsun. Şimdi standartlaştırılmış veriyi $\boldsymbol{j}$; 1 'lerin bir n -boyutlu sütun vektörü olmak üzere,

$$\boldsymbol{z}_1 = \left(\boldsymbol{x} - \bar{x}\boldsymbol{j}\right) \big/ \left(SS_x\right)^{1/2} \ , \boldsymbol{z}_2 = \left(\boldsymbol{y} - \bar{y}\boldsymbol{j}\right) \big/ \left(SS_y\right)^{1/2}$$

olarak tanımlarız.

$\|\boldsymbol{z}_1\|^2 = \boldsymbol{z}_1^{'}\boldsymbol{z}_1 = 1$, $\|\boldsymbol{z}_2\|^2 = \boldsymbol{z}_2^{'}\boldsymbol{z}_2 = 1$, $\boldsymbol{z}_1^{'}\boldsymbol{j} = 0$ ve $\left(\boldsymbol{z}_1 - \boldsymbol{z}_2\right)'\left(\boldsymbol{z}_1 + \boldsymbol{z}_2\right) = 0$ dır. Eğer α n-boyutlu uzayda $\boldsymbol{z}_1$ ve $\boldsymbol{z}_2$ arasındaki açı ise $\cos\alpha = \boldsymbol{z}_1^{'}\boldsymbol{z}_2 = r$ Pearson un moment çarpım korelasyon katsayısıdır. Şimdi aşağıdaki şekli çizelim.



Şekil 7.1 Pearson un $r = \cos^2\alpha/2 - \sin^2\alpha/2$

Şekil 7.1 de $\boldsymbol{z}_1$ ve $\boldsymbol{z}_2$ sıfırda merkezileşen vektörlerdir ve α onların arasındaki açıdır. $a/2$ açısı $\alpha/2$ dir. Şimdi,

$$\cos\alpha = \cos^2\frac{\alpha}{2} - \sin^2\frac{\alpha}{2} = \frac{\|\boldsymbol{z}_1 + \boldsymbol{z}_2\|^2}{4} - \frac{\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|^2}{4}$$

olduğunu göstermek istiyoruz. Bunun için aşağıdaki trigonometrik özdeşlikleri kullanacağız.

$$\cos\alpha = \cos^2\frac{\alpha}{2} - \sin^2\frac{\alpha}{2}, \quad \cos\frac{\alpha}{2} = \sqrt{\frac{1+\cos\alpha}{2}} \quad \cos\left(\frac{3\pi}{2} + \frac{\alpha}{2}\right) = \sin\frac{\alpha}{2} = \sqrt{\frac{1-\cos\alpha}{2}}$$

β; $\boldsymbol{z}_1$ ve $\boldsymbol{z}_1 + \boldsymbol{z}_2$ arasındaki açı olsun. Bu takdirde

$$\cos\beta = \frac{z_1(\boldsymbol{z_1} + \boldsymbol{z_2})}{\|\boldsymbol{z_1}\|\|\boldsymbol{z_1} + \boldsymbol{z_2}\|} = \sqrt{\frac{1+\boldsymbol{z}_1'\boldsymbol{z}_2}{2}} = \sqrt{\frac{1+\cos\alpha}{2}} = \cos\frac{\alpha}{2}$$

dır. Bu nedenle, $\beta = \frac{\alpha}{2}$ dir. $\boldsymbol{z}_1 + \boldsymbol{z}_2$ ve $\boldsymbol{z_1}$ - $\boldsymbol{z_2}$ ortogonal vektörler olduğundan, $\boldsymbol{z_1}$ den $\boldsymbol{z_1}$ - $\boldsymbol{z_2}$ ye saat yönü tersindeki açı bir an için $\frac{3\pi}{2} + \frac{\alpha}{2} = \beta$ dır. Böylece

$$\cos\beta = \frac{\boldsymbol{z}_1'(\boldsymbol{z_1} - \boldsymbol{z_2})}{\|\boldsymbol{z}_1\|\|\boldsymbol{z_1} - \boldsymbol{z_2}\|} = \sqrt{\frac{1-\boldsymbol{z}_1'\boldsymbol{z}_2}{2}} = \sqrt{\frac{1-\cos\alpha}{2}} = \sin\frac{\alpha}{2}$$

dir. Şimdi,

$$\frac{\|\boldsymbol{z}_1 + \boldsymbol{z}_2\|^2}{4} = \frac{1+\boldsymbol{z}_1'\boldsymbol{z}_2}{2} = \cos^2\frac{\alpha}{2} \quad \text{ve} \quad \frac{\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|^2}{4} = \frac{1-\boldsymbol{z}_1'\boldsymbol{z}_2}{2} = \sin^2\frac{\alpha}{2}$$

olur. Bu nedenle, Pearson' un r sinin

$$r = \boldsymbol{z}_1'\boldsymbol{z}_2 = \frac{\|\boldsymbol{z}_1 + \boldsymbol{z}_2\|^2}{4} - \frac{\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|^2}{4} = \frac{1+\boldsymbol{z}_1'\boldsymbol{z}_2}{2} - \frac{1-\boldsymbol{z}_1'\boldsymbol{z}_2}{2} = \cos^2\frac{\alpha}{2} - \sin^2\frac{\alpha}{2}$$

olarak yazılabildiği görülür. Buna göre

$$\frac{E(nu)^2}{4} = \frac{\|\boldsymbol{z}_1 + \boldsymbol{z}_2\|}{4} = \cos^2\frac{\alpha}{2} \quad \text{ve} \quad \frac{E(pu)^2}{4} = \frac{\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|}{4} = \sin^2\frac{\alpha}{2}$$

dır. Standartlaştırılmış veriler için kovaryans matrisli korelasyon matrisidir ve bu 2×2 boyutlu matrisin öz değerleri $\lambda_1 = 1+r$ ve $\lambda_2 = 1-r$ dir. Bu nedenle,

$$\cos^2\frac{\alpha}{2} = \frac{1+r}{2} = \frac{\lambda_1}{2} \text{ ve } \sin^2\frac{\alpha}{2} = \frac{1-r}{2} = \frac{\lambda_2}{2}$$

olduğundan, r nin bir parçalanmasının başka bir biçimi veya yorumu $r = (\lambda_1 - \lambda_2)/2$ dir, ve öz değerler arasındaki farkın yarısı $\frac{E(nu)^2}{4}$ ve $\frac{E(pu)^2}{4}$ arasındaki farktır.

Şimdi, $\frac{E(nu)^2}{4}-\frac{E(pu)^2}{4}$ farkının yorumu için $\boldsymbol{z}_1=\boldsymbol{z}_2$ veya $\alpha=0$, $\cos^2\frac{\alpha}{2}=1$, $\sin^2\frac{\alpha}{2}=0$ olduğunda, mükemmel negatif korelasyonun maksimum olduğu kolaylıkla görülür. Bu durumda $\frac{E(nu)^2}{4}=1$ ve $\frac{E(pu)^2}{4}=0$ dır. Bu nedenle, $r=\frac{E(nu)^2}{4}-\frac{E(pu)^2}{4}=1-0=1$ dir. Mükemmel negatif korelasyon durumu için $\boldsymbol{z}_2=-\boldsymbol{z}_1$ ve $\alpha=\pi$, $\cos^2\frac{\alpha}{2}=0$, $\sin^2\frac{\alpha}{2}=1$ dir. Bu nedenle, $r=\frac{E(nu)^2}{4}-\frac{E(pu)^2}{4}=0-1=-1$ dir. Fazladan iki durumu da verelim. İlk olarak $\boldsymbol{z}_1$ ve $\boldsymbol{z}_2$ ortogonal olduğunda, $\alpha=\frac{\pi}{2}$, $\sin\frac{\alpha}{2}=\cos\frac{\alpha}{2}=\frac{\sqrt{2}}{2}$ ($\sin^2\frac{\alpha}{2}=\cos^2\frac{\alpha}{2}=\frac{1}{2}$) dir. Bu nedenle, $\frac{E(nu)^2}{4}=\frac{E(pu)^2}{4}$ dır ve $r=0$ dır. Son olarak $\alpha=\frac{3\pi}{4}$, $\cos^2\frac{3\pi}{8}=0{,}1464$ ve $\sin^2\frac{3\pi}{8}=0{,}8536$ olsun. Bu nedenle,

$$\frac{E(nu)^2}{4}-\frac{E(pu)^2}{4}=0{,}1464-0{,}8536=-0{,}7072$$

dir. Şekilden, $z_1=z_2$ olduğunda, $\frac{E(nu)^2}{4}=\|z_1+z_2\|/4$ ifadesinin maksimum ve $\frac{E(pu)^2}{4}=\|z_1-z_2\|/4$ ifadesinin minimum olduğu açık olarak görülebilir ve yukarıdaki örneklerin geometrisi kolaylıkla görselleştirilebilir. (Ayrıntılar için bkz. *Rudy A. Gideon, Universty of Montana,email:gideon@selway.umt.edu, Depertmant of Mathematical Sciences, " A Generalize İnterpretation of Pearson' s r ."*)

7.2 Bir Başka Bakışla Korelasyon Katsayısı

Bu kısımda daha önceden ele aldığımız bazı kavramları kullanarak korelasyon katsayısını başka bir biçimde inceleyeceğiz.

Korelasyon katsayısını tanımlamadan önce, Pearson un r si tanımları daha doğallaştıracak biçimde ifade edilecek. r nin bu yeni ifadesi mutlak değerlere dayanan parametrik ve parametrik olmayan korelasyonların bir doğal tanımını mümkün kılacak.

(x_i, y_i) , $i = 1,2,...,n$ bir iki değişkenli veri kümesi olsun. Alışılmış ortalama notasyonu kullanacağız ve $x_i^* = x_i - \bar{x}$, $y_i^* = y_i - \bar{y}$, $i = 1,2,...,n$ merkezileştirilmiş verilerdir. Örneklem kovaryansı $\sum x_i^* y_i^*$ ile orantılıdır. Daha sonraki tanımları hazırlamak için, bu kovaryans

$$\sum x_i^* y_i^* = \left(\sum \left(x_i^* + y_i^* \right)^2 - \sum \left(x_i^* - y_i^* \right)^2 \right) \Big/ 4$$

olarak yeniden yazılabilir. Merkezileştirilmiş notasyona göre, bu ifade

$$\left(\sum \left(x_i - \bar{x} + y_i - \bar{y} \right)^2 - \sum \left(x_i - \bar{x} - y_i + \bar{y} \right)^2 \right) \Big/ 4$$

olarak yazılabilir. Kovaryans fonksiyonunun bu biçimi, onların yeniden ölçeklendirilmiş bir varyans yorumları bir araya getirildiğinde, Rodgers ve Nicewater deki Pearson un r sinin bir yorumu olarak gözükür. $x - y$ verileri arasındaki ilişkinin bir ölçüsü olarak bu biçim için bir sezgisel dürtü (motivasyon) şimdi verilir. Bu, tanımlanacak olan tüm korelasyon katsayıları için geçerli olacak. Bir pozitif korelasyon var olduğunda $\left(x_i^* + y_i^* \right)^2 = \left(x_i - \bar{x} + y_i - \bar{y} \right)^2$, $i = 1,2,...,n$ terimleri aynı doğrultu ve yöne meyledecek ve 2 standart sapma kadar büyük olacaktır. Biri, bir negatif ilişkiden (korelasyondan) uzaklığın büyük olduğunu ve bu nedenle korelasyonun pozitif olması gerektiğini söyleyebilir. Öte yandan $\left(x_i^* - y_i^* \right)^2$, $i = 1,2,...,n$ herhangi yok etme etkisine sahip olacak, bu nedenle onlar küçük olmaya meyledecek ve net etki kovaryansın büyük olacağıdır. Biri, bir pozitif ilişkiden uzaklığın küçük olduğunu bu nedenle korelasyonun pozitif olması gerektiği söylenebilir. x ve y bağımsız değişkenler olduğunda, her iki terimde de benzer bir yok etme miktarı ortaya çıkacak ve kovaryans sıfır civarında değişecek. Negatif korelasyon var olduğunda, pozitif korelasyondan uzaklık büyümeye meyledecek. $\left(x_i^* - y_i^* \right)^2, i = 1,2,...,n$ terimleri kadar büyük olacak fakat hükümsüz kılma negatif korelasyondan yakınlık küçük olacak şekilde $\left(x_i^* + y_i^*\right)^2$ $, i = 1, 2, ..., n$ terimlerinde vukuu bulacaktır.

Şimdi bu kavramları n-boyutlu Öklid uzayına genişleteceğiz. Bu paragraf için x ve y merkezileştirilmiş verinin n-boyutlu vektörleri olsun. Her birisinin 1 uzunluğuna sahip olduğunu, yani $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$ olduğunu kabul edelim. n-boyutlu uzayda $\mathbf{x} + \mathbf{y}$ vektörünü göz önüne alalım. Bu vektör orijinden (bu vektör için orijin mükemmel negatif korelasyonu gösterir) ne kadar uzak ise, korelasyon o ölçüde pozitiftir. Mükemmel pozitif korelasyon için, $\cos(x, y) = 1$ ve $\|x + y\| = 2$ yani orijinden uzaklık 2 dir. Şimdi $x - y$ vektörünü göz önüne alalım. Bu vektör orijine ne kadar yakın ise, korelasyon o kadar pozitiftir. $x - y$ için orijin pozitif korelasyonu gösterir ve bu nedenle $\|x - y\|$ nin küçük olması mükemmel pozitif korelasyondan uzaklığın küçük olduğunu ifade eder. Bunu ifade etmek için başka bir yol, $x - y$ için 2 yarıçaplı n-boyutlu merkezi kürenin (topun) yüzeyinin büyük olması mükemmel pozitif korelasyondan uzaklığın büyük olduğunu ifade eder. Mükemmel negatif korelasyon için, $\cos(x, y) = 1$ ve $\|x - y\| = 0$ dır. Bu nedenle 2 yarıçaplı küreden uzaklık bir maksimumdur.

Parametrelere göre bunu ifade etmek için başka bir yol, $Var(X + Y) > Var(X - Y)$ olduğunda pozitif korelasyonun ve eşitsizlik ters yönde olduğunda negatif korelasyonun var olmasıdır. Negatif korelasyondan uzaklık ve $Var(X + Y)$ arasındaki ilişki şimdi gösterilir. Z_1 ve Z_2 için $E(Z_1 Z_2) = \rho = \left[Var(Z_1 + Z_2) - Var(Z_1 - Z_2)\right]/4$ olduğuna dikkat edelim. Bu takdirde, $Var(Z_1 + Z_2)$ mükemmel negatif korelasyondan uzaklığa eşittir ve bu ρ nun bir lineer fonksiyonudur yani $2 + 2\rho$ dur. $\rho = -1$ için bu uzaklık 0 dır, fakat $\rho = +1$ için bu uzaklık 4 tür. Şimdi $Var(Z_1 - Z_2)$ mükemmel pozitif korelasyondan uzaklıktır ve bu $2 - 2\rho$ dur. $\rho = -1$ için bu uzaklık 4 tür, fakat $\rho = +1$ için bu uzaklık 0 dır. Bu uzaklıkların ρ nun monoton fonksiyonları olduklarına ve bu durumda Pearson un korelasyon katsayısı için $Var(Z_1 + Z_2) - Var(Z_1 - Z_2)$ genel korelasyonu 4ρ ya eşittir. Bununla beraber, diğer korelasyon katsayısı için uzaklığın bir birleşimi basit değildir. Pearson korelasyon katsayısının bu durumda

$$\frac{1}{2}\ln\frac{Var\left(Z_1+Z_2\right)}{Var\left(Z_1-Z_2\right)}=\frac{1}{2}\ln\frac{1+\rho}{1-\rho}=\tanh^{-1}\rho=\ln\frac{\sqrt{Var\left(Z_1+Z_2\right)}}{\sqrt{Var\left(Z_1-Z_2\right)}}$$

olduğuna da dikkat edelim. Bu ise Fisher' in normalleştirme dönüşümüdür.

Bir korelasyon katsayısı negatif korelasyon için 1' den küçük, bağımsız rasgele değişkenler için 1 ve pozitif korelasyonlar için 1' den büyük olacak olan $Var\left(X+Y\right)/Var\left(X-Y\right)$ oranına dayandırılabilir

8. OLAYLAR ARASINDAKİ BAĞIMLILIK VE BAĞIMSIZLIK İLİŞKİLERİ İÇİN KORELASYON

Burada olasılık hesapta olaylar arasındaki bağımlılık ve bağımsızlık ilişkileri için korelasyon kavramını kullanacağız.

8.1 Olaylar Arasındaki Bağımsızlık

Hata veya olay ağaçları, güvenilirliği tahmin etmek için risk değerlendirmelerinde veya sistemin bileşenlerinin başarısızlıklarının olasılıklarına göre bir sistemin başarısızlığı gibi bir üst olayın riskinde, yaygın bir şekilde kullanılırlar. (Veseley ve ort.) Bir hata ağacındaki bileşik olaylar, daha basit olayların birlikte olmaları, birlikte olmamaları veya eksiklikleri olarak tanımlanırlar. B ile A olaylarının birlikte olması A&B ile gösterilir ve bazen buna "ve" olayı da denir. Bu hem A hem de B olayının olduğu olayıdır. Örneğin, bir güvenlik değerlendirmesinde, A olayı bir kapalı alanda bir yanıcı buhar üreten dikkatsizlikten kaynaklanan bir petrol sızıntısını gösterebilir. B olayı alandaki bir yerde bir elektrik devresinin kapatılmasından kaynaklanan bir kıvılcım gibi bir ateşleme kaynağının varlığını gösterebilir. Bu iki olayın birlikte olması bir patlama için gerekli şartı gösterecektir. İki olayın birlikte olmaması, A veya B nin biri veya diğerinin veya her ikisinin olduğu olaydır. Birlikte olmamaya bazen" veya olayı" da denir ve bu olay $A \vee B$ ile gösterilir. Örneğin, eğer olaylar normalden fazla güvenlik sistemlerinin tasarlanmış işleyişini gösterirse, bu takdirde olayların yalnız biri sakınılacak herhangi bir olumsuz sonuç için olmalıdır.

Bir hata ağacının amacı, daha basit olayların böyle birlikte olmaları ve birlikte olmamalarını içeren bir fonksiyon olarak üst olayı kendi kendini tekrarlar şekilde ifade etmektir. Alt olayların daha fazla ayrıştırılamayan bu tekrarların uç noktalarına temel olaylar denir. Onlar diğer olaylara bağlı olarak tanımlanmadığından, temel olayları deneysel gözlemleri veya teorik düşünceyi temsil eden analizlere girdilerle tanımlamak gerekir. Bu temel olaylar, çoğu kez bazen onları olasılık dağılımlarından ayırmak için toplam olasılıklar denen reel değerli olasılıklara karakterize edilirler.

Bu kısım, temel olaylar ve onların bir hata ağacı gibi mantıksal modeli vasıtasıyla yayılımı arasındaki bağımlılıkların gösterimini yeniden gözden geçirir. İki olayın bağımsız olmaları durumunda, onların birlikte olmalarının olasılığı iki olayın olasılıkları çarpımı olarak hesaplanabilir. Bu nedenle, $a = P(A)$ ve $b = P(B)$ olmak üzere,

$$P(A \& B) = ve_{bağımsız}(a,b) = ab$$

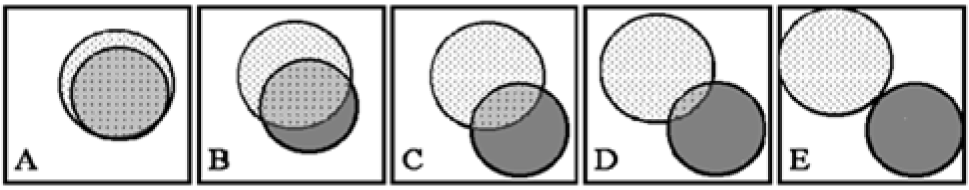
dir. Birlikte olmamanın olasılığı da,

$$P(A \vee B) = veya_{bağımsız}(a,b) = 1-(1-a)(1-b)$$

formülü ile ayrı ayrı olayların olasılıklarına bağlı olarak da hesaplanabilir. Uygun olabilmesine rağmen, bir hata ağacındaki olayların bir diğerinden bağımsız olduklarını kabul etmek daima mümkün değildir. Örneğin, Vesley ve ortakları (1981) olayların genel olarak bağımsız olmayacağı ortak-neden veya ortak-mod başarısızlıklarını içeren birçok durumu ifade etti. Örneğin, aynı anda bazen birkaç bileşenli başarısızlığı ortaya koyabilen bir tek neden olabilir. Ortak nedenlerin kategorileri; etki, yapım, tarih kullanımı, yer, titreşim, bulaştırma, rutubet, sel, sıcaklık, yangın vs. gibi, birçok şeyi içerebilir. Eğer bir hata ağacının minimal kesik kümesinde bileşenlerin tümü aynı nedenden etkilenirlerse, bu neden üst olayı başlatabilir. Bu biçimde üst olayın riski bu ortak nedenlerden birinin vukuu bulma riskine dejenere olabilir. Bir sistemin ortak neden başarısızlıklarına duyarlılığının belirlenmesi risk analizinde git gide artarak daha önemli olmaktadır. Mühendislik uygulamasında, ortak neden başarısızlıkları çoğu kez rasgele donanım başarısızlıklarına yön verebilir.

Bağımsızlık varsayımlarına güvenmeye gerek olmaksızın, riskleri ve güvenilirlikleri değerlendirebilmek için, böyle olayları içeren sistem verimliliğinin değerlendirmesini düzeltmek önemli olacak. Bundan başka, ateşlemeler gibi çevreleri anormal işlemlere tutma kapsamında sistemin planlayıcıları tarafından planlanmış bir sistemde, bileşenlerin bağımsız çalıştırılması gerçekten bağımlı davranışlara devrolabilir. Bu takdirde değerlendirmede tahmin edilmiş olan riskler ve güvenlikler üzerinde bağımsızlığın bir noksanlığının sonuçlarını tahmin edebilmek ciddi bir sorun olur.

Eğer olayları karakterize eden olasılıklar bir Venn diyagramında gösterilirlerse, olaylar arasındaki bağımlılık kümeler arasındaki örtüşmenin alanı ile tam olarak belirlenir. Şekil 8.1 de gösterilen beş venn diyagramını göz önüne alalım. İlki noktalı bir daire ile gösterilen ve ikincisi gri daire ile gösterilen her bir şekil, iki olayın olasılıklarını gösterir. Şekillerin daire olması önemli değildir, sadece onların alanları önemlidir. Benzer şekilde sadece onun alanı önemli olduğundan örtüşme bölgesinin şeklinin karmaşıklığı mesele değildir. Bu diyagramın hepsi aynı ölçekle çizilmiştir, bu nedenle beş venn diyagramının her birisi için çevreleyen karenin alanı birdir ve daha küçük gri dairenin alanı 0.22 iken daha büyük noktalı dairenin alanı 0.29 dur. A olayı tamı tamına noktalı olay içindeki gri olayı gösterir. Bu olayların marjinal olasılıkları verildiğinde bu, iki olay arasında en güçlü mümkün olan bağımlılığı gösterir. Bu mükemmel bağımlılıkta, eğer gri daire ile gösterilen olay vukuu bulursa, noktalı daire ile gösterilen diğer olayın da vukuu bulduğu garanti edilir. C durumu bağımsız durumu gösterir. Bu durumda örtüşmenin alanı 0,0638 olan 0.29×0.22 çarpımıyla verilir. Genel olarak, ortak olayın (her iki olayın da vukuu bulduğu) olasılığı olasılıkların çarpımıyla verilir. Bu sayısal ilişki sağlanmadıkça olaylar bağımsız değildirler. E durumu olayların karşılıklı olarak ayrık olduklarını gösterir. Onların örtüşen alanları sıfırdır. Bu, bir olayın vukuu bulmasının diğerinin vukuu bulmasını engellediğini ifade ettiğinden, bağımlılığın diğer ekstrem (uç) olası durumunu gösterir. B ve D durumları uç durumlar ve bağımsızlık arasında ortada olan bağımlılıkları gösterir.



Şekil 8.1 Alan olasılığı göstermek üzere Venn diyagramında gösterilen iki olay arasındaki bağımlılık

8.2. Uç Bağımlılık: Mükemmel ve Zıt

Temel olaylar arasındaki farklı bağımlılıkları göstermek için, Şekil 8.1 rin diyagramında gösterilen gri ve noktalanmış alanlar, alanlar değişmediği sürece, kare içindeki civarda hareket ettirilebilir ve onların şekilleri keyfi olarak bozulabilir. Alanlar verilen değerlere kısıtlandığında, belirli ilişkiler iki kümenin alanları ve onların bağımlılığı arasında onların örtüşme derecesiyle temsil edildiği gibi devam etmelidir. Örneğin en büyük örtüşme alanına götüren bağımlılık A diyagramında gösterilendir. Eğer alanlar sabit iseler, alanların nasıl konumlandırıldıkları mesele değildir, iki alan arasındaki böyle bir örtüşüm büyük olabilir. Bu nedenle, bu bağımlılığa "mükemmel" diyebiliriz. Alanlardan biri diğerini tam olarak içerdiğinden örtüşmenin alanı iki alanın minimumudur. Bu çeşit bağımlılık için, olaylar arasındaki aynı zamanda olma ve aynı zamanda olmamanın olasılıklarını hesaplamak çok basittir:

(Tam)

$$P(A \& B) = ve_{tam}(a,b) = \min(a,b)$$

$$P(A \vee B) = veya_{tam}(a,b) = \max(a,b)$$

formülleriyle ifade edilirler. Burada A ve B iki alandır. Birlikte olmanın (aynı zamanda olmanın) olasılığı, iki alanın kesişiminin alanı ile ölçülür. Onlar tam olarak örtüştüğünde, kesişimin alanı iki alandan daha küçüğü olmalıdır. Birlikte olmamanın olasılığı iki alanın birleşiminin alanı olduğundan bu olasılık alanların daha büyüğü olmalıdır.

Şekil 8.1 de örtüşmenin en küçük alanına götüren bağımlılık kalıbı E diyagramında gösterilene götürendir. Örtüşmenin alanı sıfır olduğundan, kümeler ayrıktırlar. Minimal örtüşmeye ilişkin bağımlılığa "zıt" bağımlılık diyeceğiz. Zıt bağımlılığa sahip olmanın ister istemez olayların karşılıklı olarak ayrık olmalarını ifade ettiğine dikkat edelim. Örneğin, her iki olay %50 den daha büyük olasılıklara sahip olabilir. Bununla beraber, böyle bir durumda, çizilen olayların alanları bir miktar örtüşmek zorundadır. Bu nedenle olayların karşılıklı olarak ayrık olduklarını söyleme onların bağımlılığı hakkında henüz bir iddia yapma değildir. O, olayların olasılıkları hakkında da bir şey söyler. Olayların zıt bağımlılığa sahip olduğunu söyleme sadece bağımlılık hakkında bir iddiadır.

Zıt bağımlılık altında birlikte olma ve birlikte olmama olasılıkları için formüller:

(Zıt)

$$P(A \& B) = ve_{zıt}(a,b) = \max(a+b-1,0)$$

$$P(A \vee B) = veya_{zıt}(a,b) = \min(1,\mathrm{a}+b)$$

dır. Bu formüller olasılıkların 1 den daha büyük olmaması gerektiği sınırlamasını açıklar.

Örnek: $P(A)=a=0.29$ ve $P(B)=b=0.22$ olduğunu ve A ve B olaylarının tam olarak bağımlı olduklarını farz edelim. Bu durum alanların maksimal olarak örtüşmesini gösteren Şekil 8.1 rin A diyagramında gösterilmiştir. $A\&B$ birlikte olmasının olasılığı min(0.29,0.22)=0.22 dir. Minimum operatörünün kullanımına rağmen, bu; bu marjinaller verildiğinde olasılığın mümkün olan en büyük değeridir. $A\vee B$ birlikte olmamanın olasılığı max(0.29, 0.22)=0.29 dur. Maksimuma rağmen bu herhangi bir olası bağımlılık için olasılığın mümkün olabilir en küçük değeridir. Şimdi bağımlılığın, olayların zıt olarak bağımlı oldukları E diyagramında gösterilene benzer olduğunu farz edelim. Şimdi alanlar minimal olarak örtüşüyorlar. Bu durumda, birlikte olmanın olasılığı max(1.029+0.22-1,0)=0 dır ve birlikte olmamanın olasılığı min(1,0.29+0.22)=0.51 dir.

Bu uç durumlar analizci, bağımlılık hakkında deneysel bilgi veya teorik düşünceye sahip olmadığında sınırlar olarak çoğunlukla yararlıdır, fakat onların kendilerinin sağındaki bir gerçek değerlendirmede kullanılabildikleri düşünülebilir. Örneğin, birçok mühendislik sistemlerinde temel bileşenlerin birkaçı çoğu kez tek bir bayii tarafından temin edilir veya aynı muayene, hizmet ve onarım tarihine maruz kalmıştır. Bundan başka, bileşenler, bir ateş veya birbirini izleyen geçici çevre şartı gibi, benzer bir anormal şarta maruz kalabilir. Böyle topluluklar bu bileşenlere ilişkin başarısızlık olasılıklarının bağımsızlıktan ziyade tam bağımlılığa daha yakın olabildiğini önermeye meyledebilirler. Öte yandan zıt bağımlılık değiş tokuşlarla kullanmayı önerebilir. Örneğin, gerekenden fazla emniyet sistemlerinin bir çiftinden biri daima bir özel uyarıcı çeşidine cevap vermede ilk harekete geçer ve uyarıcının çeşitleri sistemler vasıtasıyla rasgele tecrübe edilmezler. Eğer bir emniyet sisteminin işlemi sistem üzerinde normal durumda aşınıp eskimeye götürürse (sevk ederse) bu takdirde her iki sistemin ortak başarısızlığı onların bağımsız olduklarının bir varsayımından ziyade, zıt olarak bağımlı olan olaylarla daha iyi modellenebilirler. Böyle durumlarda, uç ekstrem-bağımlılık-bağımsızlık formüllerinin yerine kullanılabilir, çünkü uç bağımlılık bir dereceye kadar bağımsızlıktan daha mantıklı bir varsayımdır.

8.3 Olayların Arasındaki Korelasyon

Bağımlılık kavramı hem rasgele değişkenlere ve hem de basit olaylara uygulanmakla beraber korelasyon sözü çoğu kez sadece rastgele değişkenler ile kullanım için ayrılır. İlişkinin genel derecesini -1 ve +1 arasındaki değişen herhangi skalar nicelikle ölçme genel olarak risk belirlemelerinde çok yararlı olabilir. Bu nedenle, her biri bir toplam olasılıkla karakterize edilen olaylar ile kullanım için kavramın (yani sıfır ve bir arasında boyutsuz bir reel sayıya) genişletilmesi istenebilir. Olaylar bağlamında bağımlılığın basitliğinden dolayı, olayların ilişkisinin (korelasyonunun) böyle bir skalar ölçüsünü tanımlama bütün yönleriyle akla uygundur. Gerçekten, bu kullanım rastgele sayılar arasındaki genel bağımlılığın skalar ölçülerinin geleneksel kullanımından çok daha fazla mantıklı görülür. Rastgele sayılar arasındaki bağımlılık sonsuz-boyutludur, bu nedenle bir korelasyon katsayısının bir tek boyutu, bağımlılık fonksiyonunun potansiyel karmaşıklığını yansıtamaz. Bununla beraber olaylar arasındaki bağımlılık, tam olarak ve bir skalar ölçüyle bilgiyi kaybetmeksizin karakterize edilebilir.

Lucas (1955, bkz.Cheng 2003, Cui ve Blockley 1990; Davis ve Hall 2003) olaylar arasındaki korelasyonu onların gösterge fonksiyonlarının ilişkisi olarak tanımlamayı önerdiler. Bu düşünceyi açıklamak için, Şekil 8.1 deki gibi gösterilen bir Venn diyagramında rasgele oklar fırlatmayı ve her bir olay için bir ve sıfır ikili değerlerini kullanarak her bir fırlatmayı puanlamayı düşünelim. Eğer ok olayın alanını ıskalarsa (kaçırırsa) gösterge fonksiyonunun değeri sıfır olacak ve ok olayın alanına isabet ederse bir olacaktır. Rasgele fırlatılan oklar (dartlar) için böyle sayıların uzun bir dizisi için iki olay arasındaki bağımlılığın Lucas ölçüsü korelasyon katsayısı olacak. Eğer, rasgele fırlatılan oklar yerine birçok dişlinin bir sistemi benzer puan çiftlerinin düzgün ve sistematik olarak alıyor olsaydı ölçü aynı olacaktı. Bildiğimiz gibi, X ve Y rastgele değişkenleri arasındaki bir korelasyon (Pearson çarpım-moment korelasyonu) için formül

$$r = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

dir. Burada E beklenen değeri ve V varyansı gösterir. Bir A olayı için bir indikatör fonksiyonunun beklenen değeri olayın P(A) olasılığıdır. İndikatör fonksiyonu için varyans $P(A)(1\text{-}P(A))$ dır. Sonuç olarak, gösterge fonksiyonlarının çarpımının beklenen değeri iki olay arasındaki bağımlılığı veren birlikte olmanın olasılığı olduğundan, korelasyon için formül A ve B olayları için gösterge fonksiyonlarına (bkz. Ek1) uygulandığında, o;

$$r = \frac{P(A\&B) - P(A)P(B)}{\sqrt{P(A)(1-P(A))}\sqrt{P(B)(1-P(B))}}$$

olur. Yeniden bir düzenlemeyle, bu

$$P(A\&B) = ve_{Lucas}(a,b,r) = ab + r\sqrt{a(1-a)b(1-b)}$$

formüllemesine götürür. Burada $a = P(A)$ ve $b = P(B)$ dir. Bu formüle düzeltilmiş olaylar için Lucas modeli diyeceğiz. Örneğin $a = 0.29$ $,b = 0.22$ ve r korelasyonu $= 0.2$ olduğunu farz edelim.

Lucas modeli, A ve B nin birlikte olması olasılığının 0.101 olması gerektiğini önerir. Eğer r sıfıra eşit alınırsa, Lucas modeli 0.638 değerini verir. Bu ise, A ve B olaylarının bağımsız olduklarını kabul ederek birlikte olmanın olasılığıdır. r değişirken Lucas modeline göre birlikte olmanın olasılığı lineer olarak değişir. Birlikte olma için böyle olasılıkları sınırlayan Frechet sınırlarını dışlayan değerleri üretebilmesi hariç bu model istenilen model olarak göz önüne alınabilir. Gerçekten, bu model sıfırdan daha küçük veya birden daha büyük olan olasılıkları bile üretebilir. Örneğin, eğer $r = -1$ alırsak Lucas modelinin sonucu -0.124 dür. Hiç kimse negatif bir olasılıktan hoşlanmaz.

Problem, formülün kendisi ile değildir ancak korelasyonun pozitif ve negatif bir değer arasında herhangi bir değeri alabilmesi yanlış düşüncesiyledir. Pearson korelasyon katsayılarının daima bu tam aralığa yayılmadığı herkesçe bilinir. Bu durumda, gösterge fonksiyonları için en küçük olası korelasyon katsayısı -1 değildir fakat sadece -0.339 dur. Ayrıca en büyük değer +1 değildir fakat sadece 0.831 dir. Bu, sayıların iki sütununu düşünmek suretiyle gösterilebilir. A olayı için indikatör fonksiyonunu gösteren birinci sütun 290 tane 1 e ve 710 tane sıfıra sahiptir. B olayı için sayıların ikinci sütunu 220 tane bire ve 780 tane sıfıra sahiptir. Eğer sütunlardaki değerler tüm sıfırlar sütunların üs kısmında olacak şekilde sıralanırsa bu takdirde iki sütun arasındaki korelasyon 0.831 olacaktır. Bundan sonra, eğer sütunlardan birini ters sırada tüm sıfırlar altta olacak şekilde sıralarsak bu takdirde korelasyon -0.339 olacaktır. Sıfırların ve 1 lerin sütunlar içinde nasıl bir yerden alıp başka bir yere koyulması mesele değildir, sayıların bu iki sütunu arasındaki korelasyon katsayısı bu uç değerlerden hiçbir şekilde daha büyük veya daha küçük olamaz. Eğer r korelasyonu için girdi değerleri $\left[-0.339, 0.83\right]$ aralığına kısıtlanırsa bu takdirde Lucas modeliyle hesaplanan olasılıklar negatif olmayan olasılıkları veya 1 den daha büyük olasılıkları garanti eden mümkün olabilir aralığa uygun şekilde sınırlanırlar. Lucas modelini mantıki kılmak için yegane yöntem,

$$\underline{r} = \frac{\max\left(a+b-1,0\right)}{\sqrt{a\left(1-a\right)b\left(1-b\right)}}$$

$$\bar{r} = \frac{\min\left(a,b\right)-ab}{\sqrt{a\left(1-a\right)b\left(1-b\right)}}$$

olmak üzere giriş korelasyonlarını $\underline{r}$ den daha küçük ve $\bar{r}$ den daha büyük olmayacak şekilde sınırlamaktır. Olayların korelasyonu için başka bir formülleme ilk olarak Frank (1979) tarafından ortaya koyulan kopulaların (kopula: marjinal dağılımları birim aralıkta düzgün dağılımlar olan iki değişkenli olasılık dağılımlarının bir sınıfıdır. İki değişkenli dağılımların Frank ailesi buna bir örnektir.) Frank ailesinden elde edilebilir. Olaylar arasındaki korelasyonun Frank modelinde A ve B olaylarının birlikte olmasının olasılığı

$$P\left(A\& B\right) = ve_{Frank}\left(a,b,r\right) = \begin{cases} \min\left(a,b\right) & \text{eğer } r=+1 \text{ ise} \\ ab & \text{eğer } r=0 \text{ ise} \\ \max\left(a+b-1,0\right) & \text{eğer } r=-1 \text{ ise} \\ \log_s\left[1+\left(s^a-1\right)\left(s^b-1\right)/\left(s-1\right)\right] & \text{aksi takdirde} \end{cases}$$

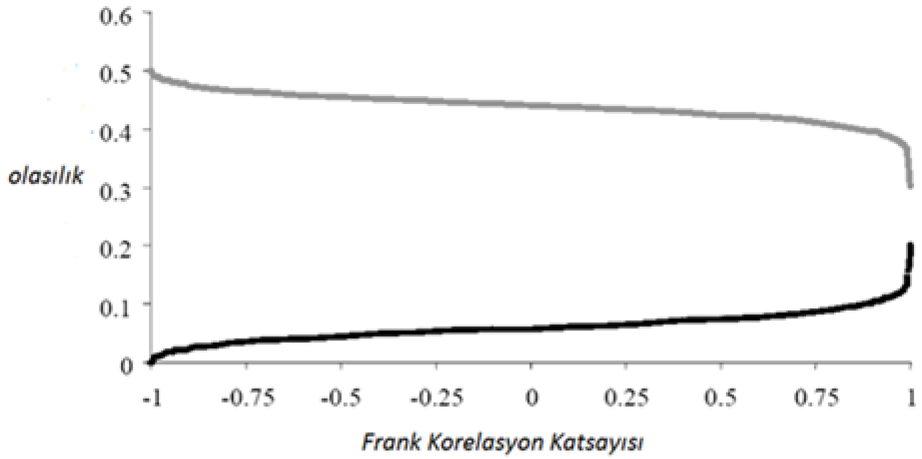
formülü ile verilir. Burada $s=\tan\left(\pi\left(1-r\right)/4\right)$, $a=P\left(A\right)$ ve $b=P\left(B\right)$ dir. Bu fonksiyon süreklidir; r, +1, 0 veya -1 olduğundaki özel durumlar, r sırasıyla bu değerlere gittiğinde bu formülün sağ yanındaki dip ifadenin limit değerleri olur. İlişkili olayların birlikte olmaması Frank ın ko-eş-kopula (co-copula) sı ile benzer şekilde tanımlanabilir. Bu nedenle,

$$P(A \vee B) = veya_{Frank}(a,b,r) = \begin{cases} \max(a,b) & \text{eğer } r = +1 \text{ ise} \\ 1-(1-a)(1-b) & \text{eğer } r = 0 \text{ ise} \\ \min(a+b,1) & \text{eğer } r = -1 \text{ ise} \\ 1-\log_s\left[1+\left(s^{1-a}-1\right)\left(s^{1-b}-1\right)/(s-1)\right] & \text{aksi takdirde} \end{cases}$$

dir.

Örnek: Eğer yine a=0.29 ve b=0.22 ve r korelasyonu $r = 0.2$ olduğunu farz edersek bağımsızlığın Frank modeli $A \& B$ nin olasılığının 0.0695 olduğunu ifade eder.

Bağımsızlığın Frank modelinde -1 ve +1 arasında değerin olmaması bir mümkün olmayan ilişkidir. Şekil 2, iki olay arasındaki ilişkinin Frank modeline göre korelasyonunun çeşitli değerleri için birlikte olmamanın (gri çizgi ile gösterilen) ve birlikte olmanın (siyah çizgi olarak gösterilen) olasılıklarını gösterir.



Şekil 8.2 Frank korelasyonunun bir fonksiyonu olarak 0.29 ve 0.22 olasılıklı iki ilişkili olayın birlikte olmaması (gri) ve birlikte olması (siyah) olasılığı

Frank modeli olaylar için bir korelasyon kavramını parametrelemek için birçok mümkün yoldan sadece birisidir. Birçok araştırmacı "ve" ve "veya" işlemleri için modeller olarak t-normlarını ve t-ko (eş) normlarını kullanmayı önerdi. t-normlarına genelleştirilmiş kesişim işlemcileri (operatörleri) de denir ve t-normları, olaylar arasındaki bağımlılığın çeşitli modelleri altında birlikte olmalarını değerlendirmek için ve_{Frank} ın yanı sıra $ve_{bağımsız}$, ve_{tam}, $ve_{zıt}$ fonksiyonlarını, özel durumlar olarak içerirler. Aynı şekilde, genelleştirilmiş birleşim operatörleri denen t-ko (eş) normları

$veya_{bağımsız}, veya_{zıt}$ ve $veya_{Frank}$ fonksiyonlarını içerirler. Bununla beraber, birlikte olmalar ve birlikte olmamaların olasılıklarını tahmin etmek için işlemleri tanımlamada t-normlarını ve t-eş normlarını kullanmak mantıklı görünmez. Nedeni bu fonksiyonların bazısının olasılık değerleri ile bağdaşmazlığıdır. Örneğin, uygun bir şekilde yeniden adlandırılan "güçlü kesişim" (Klir ve Yuan 1994) olasılıklarla mümkün olmayan sonuçları verir. Çünkü onlar Frechet sınırları dışındadırlar. Bu nedenle bağımlılık arasında ilişkileri karakterize eden bir model olarak kopulalara bakmak daha ihtiyatlı görünür. Frank modeli bir kopuladır. (Frank 1979; Nelsen 1999) . Nelsen (1999) bu amaç için de kullanılabilen kopullalar ve eş kopullaların diğer birçok ailesini ifade eder.

8.4 BİLİNMEYEN BAĞIMLILIK İÇİN HESAPLAMA

Sadece olayların olasılıkları hakkında marjinal bilgiyi kullanarak ve onların bağımlılığı hakkında hiçbir bilgiyi kullanmadan birlikte olmalar, birlikte olmamalar ve diğer müşterek (ortak) olaylar üzerindeki sınırları tahmin etmek mümkündür. Böyle hesaplamalar klasik

$$P(A\&B) = ve_{Frechet}(a,b) = \left[\max(0, a+b-1), \min(a,b)\right]$$

$$P(A\vee B) = veya_{Frechet}(a,b) = \left[\max(\mathrm{a},b), \min(1,\mathrm{a}+b)\right]$$

Frechet eşitsizliklerini kullanırlar. Burada $a = P(A), b = P(B)$ dir ve elde edilen köşeli parantezler a ve b kesin olsalar bile elde edilen olasılık tahminlerinin kesin reel değerlerden ziyade aralıklar olduklarını gösterirler.

Bu, gerçek olasılık ne olursa olsun, onun aralığın içinde olması gerektiğini söyler. Williamson (1989, s.131) bunlardan en azından birincisinin Boole (1854) 'e göre bilindiğini belirtti. Frechet (1935) onların sadece ekstrem (uç) durumlar olmadıklarını fakat aynı zamanda onların bağımlılığın mümkün olabilir tüm durumlar üzerindeki sınırlar da olduklarını ve ayrıca onların bağımlılık hakkındaki bilginin yokluğunda en iyi olası böyle sınırlar olduklarını gösterdi.

Frechet eşitsizliklerinin ispatları basittir. Birlikte olma için ispatı ele alalım. $P(A\vee B) = P(A) + P(B) - P(A\&B)$ tanımı $P(A\&B) = P(A) + P(B) - P(A\vee B)$

olduğunu ifade eder. $P(A \vee B) \leq 1$ olduğundan yani, tüm olasılıklar 1 den daha büyük olmadığından, $P(A)+P(B)-1 \leq P(A \& B)$ durumu gerçekleşmek zorundadır. Yine tüm olasılıklar sıfırdan daha küçük olamayacağından $0 \leq P(A \& B)$ olduğu doğrudur, bu nedenle $\max(0, P(A)+P(B)-1) \leq P(A \& B)$ eşitsizliğinin doğru olduğu da gerçekleşmiş olur. Bu, birlikte olma üzerindeki alt sınırı belirler. Üst sınırı elde etmek için $P(A \& B) = P(A|B)P(B) = P(B|A)P(A)$ olduğunu hatırlayalım. Tüm olasılıklar için, $P(A|B) \leq 1$ ve $P(B|A) \leq 1$ olduğundan $P(A \& B) \leq P(A)$ ve $P(A|B) \leq P(B)$ olduğu, bu nedenle $P(A \& B) \leq \min(P(A), P(B))$ olduğu görülür. Onların, A ve B olayları arasındaki herhangi bağımlılık ilişkisiyle gerçeklendiklerine dikkat etmek suretiyle bu sınırların en iyi olası yapısı görülür.

Örnek: $P(A) = a = 0.001$ ve $P(B) = b = 0.002$ olduğunu farz edelim. Bu takdirde, $P(A \& B)$ nin $[\max(0, 0.001+0.002-1), \min(0.001, 0.002)]$ aralığında olacağı kesindir. Aynı şekilde $P(A \vee B)$ de $[\max(0.001, 0.002), \min(1, 0.001+0.002)]$ arasında herhangi bir yerde olduğu kesindir. A ve B olayları arasındaki bağımlılık her ne olursa olsun bu aralıklar kesin ve doğrudur.

Frechet (1935) eşitsizlikleri çok değişkenli duruma indüksiyonla genelleşir. Sonuçta elde edilen formüller anlaşılır genişletmelerdir, bu nedenle, $a_i = P(A_i)$ olmak üzere,

$$P(A_1 \& A_2 \& \ldots \& A_n) = [\max(0, a_1 + a_2 + \ldots + a_n - (n-1)), \min(a_1, a_2, \ldots, a_n)]$$

$$P(A_1 \vee A_2 \vee \ldots \vee A_n) = [\max(a_1, a_2, \ldots, a_n), \min(1, a_1 + a_2 + \ldots + a_n)]$$

dır.

Örnek: $P(A_1) = a_1 = 0.001$, $P(A_2) = a_2 = 0.002$ ve $P(A_3) = a_3 = 0.003$ olduğunu farz edelim. Bu takdirde,

$$P\left(A_1 \,\&\, A_2 \,\&\, A_3\right) \in \left[\max\left(0, 0.001+0.002+0.003-(3-1)\right), \min\left(0.001, 0.002, 0.003\right)\right]$$
$$= [0, 0.001]$$
$$P\left(A_1 \vee A_2 \vee A_3\right) \in \left[\max\left(0.001, 0.002, 0.003\right), \min\left(1, 0.001+0.002+0.003\right)\right]$$
$$= [0.003, 0.006]$$

olur.

8.5 Bağımlılığın İşareti Hakkındaki Bilgiyi Kullanma

Eğer A ve B olayları pozitif olarak bağımlı olabilseydi bu takdirde,

(**Pozitif**)

$$P(A \,\&\, \mathrm{B}) = ve_{pozitif}\,(a,b) = \left[ab, \min(a,b)\right]$$

$$P(A \vee \mathrm{B}) = veya_{pozitif}\,(a,b) = \left[\max(a,b), 1-(1-a)(1-b)\right]$$

olacaktı. Eğer A ve B olayları sadece negatif olarak bağımlı olabilseydi bu takdirde

(**Negatif)**

$$P(A \vee \mathrm{B}) = veya_{negatif}\,(a,b) = \left[1-(1-a)(1-b), \min(1, a+b)\right]$$

olacaktı.

Örnek: A ve B nin kesin olarak pozitif bağımlı olduklarını ve $P(A) = a = 0.003$ ve $P(B) = b = 0.005$ olduklarını farz edelim. bu takdirde $P(A \,\&\, B)$ kesin olarak $\left[0.003 \times 0.005, \min(0.001, 0.002)\right] = [0.000015, 0.003]$ aralığında olacağı kesindir. Bununla beraber, eğer A ve B negatif bağımlı iseler, bu takdirde $P(A \vee B)$ nin $\left[1-(1-0.003)(1-0.005), \min(1, 0.003+0.005)\right] = [0.007985, 0.008]$ olacağı kesindir.

9.KORELASYON SINIRI

Bu bölümde korelasyon sınırı hakkında bazı bilgiler verilecektir. (bkz. *https://www.google.com.tr/#q=the+correlation +boundary ,“The correlation Boundary”.)*

9.1 Tanımlar

Bağımsız olarak aynı dağılan rasgele değişkenler olan X_1 ve X_2 servet getirileri onların ρ lineer bağımlılığı ve bir μ getirisi için aşağıdaki tanımları göz önüne alalım.

$$\mu_i \triangleq E[X_i]$$
$$\sigma_i^2 \triangleq V[X_i]$$
$$kor[X_1, X_2] \triangleq \rho$$
$$M \triangleq w_1 X_1 + w_2 X_2$$

$kor[M, X_i] > \rho$ olduğunu göstermek istiyoruz. X_i için X_1 ri kullanacağız. İndisleri değiştirerek ispatı X_2 ye genişleyeceğiz. Korelasyon için standart tanımlar ve dönüşümler ile işe koyulacağız.

$$kor[M, X_1] = \frac{kov[M, X_1]}{\sigma_M \sigma_1} = \frac{E[MX_1] - E[M]E[X_1]}{\sigma_M \sigma_1}$$

$$E[MX_1] = E[(w_1X_1 + w_2X_2)X_1]$$
$$E[MX_1] = E[w_1X_1^2] + E[w_2X_1X_2]$$
$$E[MX_1] = w_1E[X_1^2] + w_2E[X_1X_2]$$

dır. Burada $E[X^2] = V[X] + E[X]^2$ gerçeğini kullanırız.

$$E[MX_1] = w_1\sigma_1^2 + w_1\mu_1^2 + w_2\rho\sigma_1\sigma_2 + w_2\mu_1\mu_2$$

$$\Rightarrow kor[M, X_1] = \frac{w_1\sigma_1 + w_2\sigma_2\rho}{\sigma_M}$$

$$kor[M, X_1] = \frac{w_1\sigma_1 + w_2\sigma_2\rho}{\sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho\sigma_1\sigma_2}}$$

dır.

$$a = w_1^2\sigma_1^2$$
$$b = w_1^2\sigma_2^2$$
$$c = 2w_1w_2\sigma_1\sigma_2$$
$$\Rightarrow c = 2\sqrt{a}\sqrt{b}$$
$$a,b,c \geq 0$$

olsun. c üzerindeki son şart ağırlıklarımızın pozitif olduğu gerçeğinden çıkar. Şimdi $kor\left[M, X_1\right]$ ifadesini

$$kor\left[M, X_1\right] = \frac{\sqrt{a} + \rho\sqrt{b}}{\sqrt{a+b+c\rho}}$$

olarak yeniden yazabiliriz.

9.2 Hipotez

$kor\left[M, X_1\right] \geq kor\left[X_1, X_2\right]$ olduğunu varsayalım. Bunu tersini (yani $kor\left[M, X_1\right] < kor\left[X_1, X_2\right] = \rho$)olduğunu kabul edelim ve bu varsayımın ρ 'nun tüm durumları için çelişme ile ihlal edileceğini (bozulduğunu) göstererek ispatlayacağız. Çelişmemizi göstermek için aşağıdaki dört durumu araştıracağız.

$\rho = -1$, $\rho \in (-1,0)$, $\rho = 0$ ve $\rho \in (0,1]$

9.3 İspat

$\rho = -1$ için ispat:

ρ için-1'ri yerine koyarsak korelasyon -1're eşit ya da ondan daha büyük olması tanımıyla sınırlandığından dolayı bir çelişme olduğunu bildiğimiz

$$kor\left[M, X_1\right] = \frac{\sqrt{a} - \sqrt{b}}{\sqrt{a+b-c}} < -1$$

İfadesini elde ederiz.

$\boldsymbol{\rho} \in \left(\boldsymbol{-1,0}\right)$ için ispat:

$$cor\left[M, X_1\right] < \rho$$

$$cor\left[M, X_1\right]^2 > \rho^2$$

$$\frac{a + b\rho^2 + c\rho}{a + b + c\rho} > \rho^2$$

$$a + b\rho^2 + c\rho > a\rho^2 + b\rho^2 + c\rho^3$$

$$a - a\rho^2 + c\rho - c\rho^3 > 0$$

dır. $f\left(x\right) = a - ax^2 + cx - cx^3$ olsun. $\rho \in \left(-1,0\right)$ için $f\left(\rho\right) > 0$ olduğunu kabul edeceğiz. f ; 1 ya da 3 köke sahip ve köklerden ikisinin x=1 ve x=-1de yer aldığı bir kübik olacağından, denklemin bir üçüncü köke sahip olması gerektiğini biliyoruz. $g\left(p\right) = kor\left[M, X_1\right] = \dfrac{\sqrt{a} + \rho\sqrt{b}}{\sqrt{a + b + c\rho}}$ olarak tanımlayalım. $\lim_{x\to\infty} f\left(x\right) = -\infty$ ve $\lim_{x\to-\infty} f\left(x\right) = \infty$ iken, c nin işareti bize f nin şeklini verir. f nin şeklini bildiğimizden, eğer, $g\left(\rho^*\right) = 0$ olacak şekilde ρ^* için $f\left(\rho^*\right) < 0$ olduğunu ispatlarsak, $\rho \in \left(-1,0\right)$ için $f\left(\rho\right) \leq 0$ olduğunu görürüz. Eğer $\sqrt{a} + \rho\sqrt{b} = 0$ ise, $\rho^* = -\sqrt{\dfrac{a}{b}}$ alarak, $g\left(\rho\right) = \dfrac{\sqrt{a} + \rho\sqrt{b}}{\sqrt{a + b + c\rho}} = 0$ olur. $\left|\rho\right| < 1$ olduğundan, bu durumda $a < b$ almamız gerektiğini biliyoruz.

Şimdi $f\left(\rho^*\right) < 0$ olduğunu göstermeliyiz. Son olarak zıtlığı kabul edeceğiz ve bir çelişmeyi göstereceğiz. ρ için ρ^* alarak, $f\left(\rho^*\right) = -a + \dfrac{a^2}{b} > 0$ veya $\dfrac{a^2}{b} > a$ elde ederiz, ki bu $\dfrac{a}{b} > 1$ olduğunu ifade eder. Bu ise ρ^* için çözümle bulduğumuz sonuçla bir çelişmedir. Bu nedenle, $f\left(\rho^*\right) < 0$ olduğunu biliyoruz, ki bu $f\left(\rho\right) > 0$ varsayımıyla direkt bir çelişmedir.

$\rho = 0$ için ispat:

$kor\left[M, X_1\right]$ in tanımından, ρ için 0 koyarsak $\frac{\sqrt{a}}{\sqrt{a+b}} < 0$ eşitliğini ortaya koyarız. Hâlbuki a ve b nin her ikisi de pozitiftirler, bu nedenle bu bir çelişmedir.

$\rho \in (0,1]$ için ispat:

$$cor\left[M, X_1\right] < \rho$$

$$cor\left[M, X_1\right]^2 < \rho^2$$

$$\frac{a + b\rho^2 + c\rho}{a + b + c\rho} < \rho^2$$

$$a + b\rho^2 + c\rho < a\rho^2 + b\rho^2 + c\rho^3$$

$$a - a\rho^2 + c\rho - c\rho^3 < 0$$

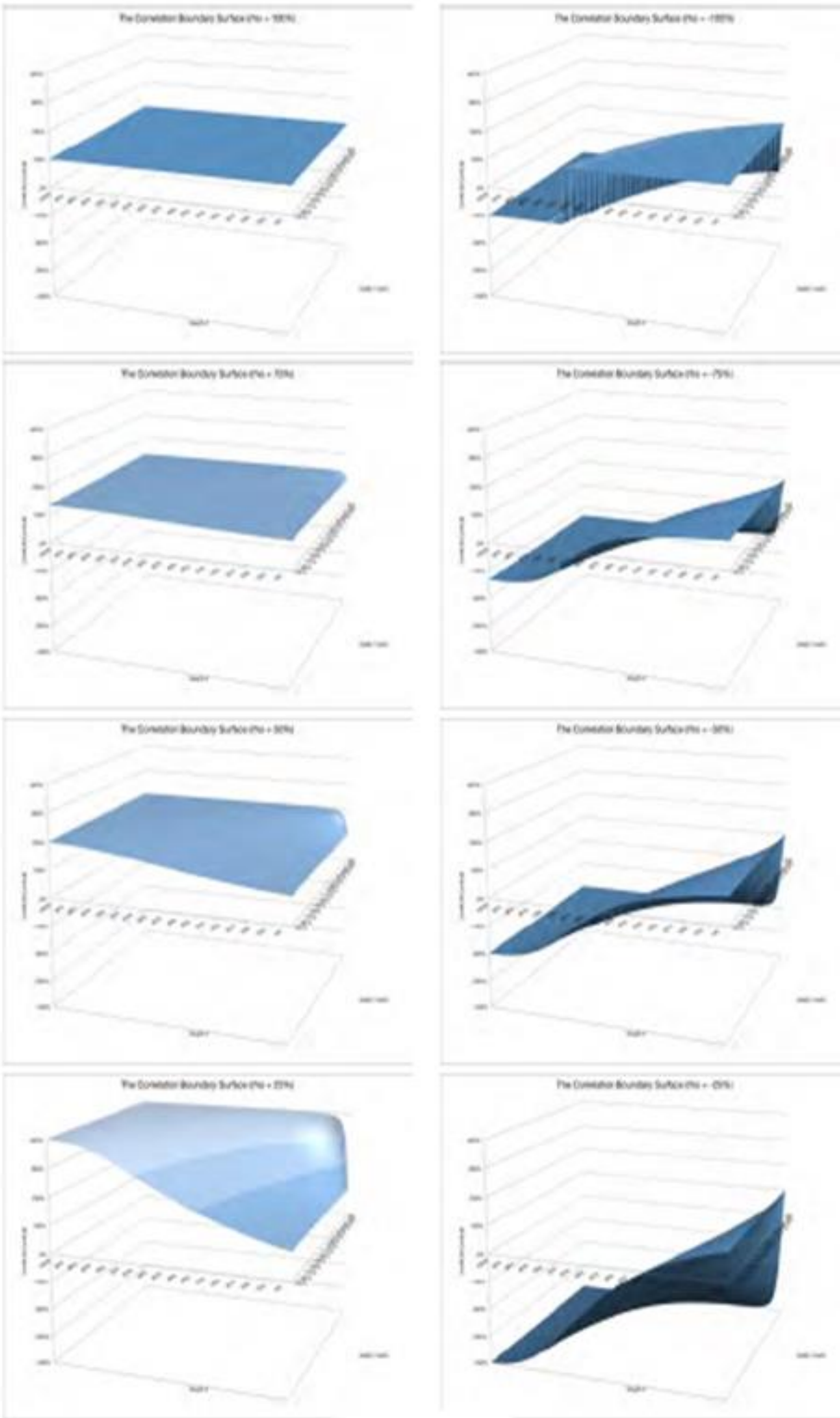
dır.

Tekrar $f(x) = a - ax^2 + cx - cx^3$ fonksiyonunu tanımlarız. Yukarıdaki ikinci ispatta belirtilen köklerden ikisi x=1 ve x=-1 de yer alır ve bu nedenle denklem bir üçüncü köke sahip olmalıdır. Sıfır koyarak, $f(0) = a > 0$ elde ederiz. ikinci ispatta tanımladığımız f nin şekli ile birleştirmeden kayıp kökün sıfırdan küçük olması gerektiğini biliyoruz. Bu nedenle, $0 < \rho \leq 1$ için, $f(\rho) \geq 0$ dır, ki bu varsayımla çelişir. (Bu fonksiyon için kayıp kök $= -\frac{a}{c}$ dir)Bu dört durum ilişkili X_1 ve X_2 servet getirilerinin bir lineer kombinasyonu olan bir M strateji getirileri için M ve X_1 ya da X_2 arasındaki korelasyonun X_1 ve X_2 arasındaki en az korelasyon olduğunu gösterir.

9.4 Görselleştirme

Korelasyonu, $kor\left[A, B\right]$ olan iki servetin, A ve B getirileri verildiğinde, $\frac{kor\left[M, A\right]}{kor\left[A, B\right]}$ nin yüzeyini veya w değişirken, stratejimiz ve A nın, A ve B nin korelasyonuna göre göreceli seviyesini ve $\frac{\sigma_B}{\sigma_A}$ değişirken, B nin A ya göre göreceli oynaklığını grafikleyebiliriz. $kor[A,B]$ nin farklı değerleri için yüzeyin nasıl değiştiğini görmek için bkz. Tablo 9.1

Çizelge 9.1 Değişken şartlar altındaki korelasyon sınır yüzeyi



9.5 Geometrik Sezgi

Korelasyon sınırı için, korelasyonun bir geometrik yorumu ile daha belirgin olacağı bir sezişi elde edeceğiz. İlk olarak Pearson' un r korelasyonu için geometrik sezişi belirleyeceğiz. Aşağıdaki tanımları göz önüne alalım.

$$\boldsymbol{v} = (v_1, \dots, v_n)$$

$$\boldsymbol{w} = (w_1, \dots, \mathrm{w}_n)$$

$$\Rightarrow \boldsymbol{vw} = (v_1 w_1 + \dots + v_n w_n)$$

$$\Rightarrow \sum v_i w_i = \|\boldsymbol{v}\| \|\boldsymbol{w}\| \cos(\theta)$$

dir. Burada,

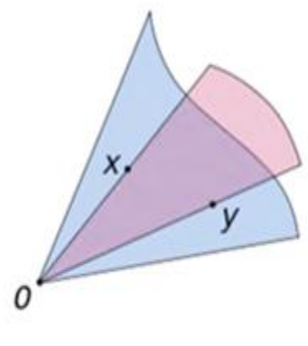
$$\|\boldsymbol{x}\| \equiv \sqrt{\sum \left(x_i - \bar{x}\right)^2}$$

$$\bar{x} \equiv \mu_x = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$$

dir. Terimleri basit olarak yeniden düzenleyerek,

$$\cos(\theta) = \frac{\boldsymbol{vw}}{\|\boldsymbol{v}\| \|\boldsymbol{w}\|}$$

elde ederiz.



Şekil 9.2 Koyu gölgeli alan *x* ve *y* nin tüm mümkün pozitif lineer kombinasyonları tarafından tanımlanan konveks konidir. Sağ üstteki yay koninin sonsuz genişlediğinin göstergesidir

$\cos(\theta)$ için tanımın iki normalleştirilmiş vektör arasındaki nokta çarpımı (skalar çarpım, iç çarpım) olarak da tanımlanabileceğine dikkat edelim. Şimdi Pearson un r örneklem korelasyon katsayısı için aşağıdaki tanımı göz önüne alalım.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$

Eğer vektörler $\boldsymbol{v} = (x_1 - \bar{x}, ..., x_n - \bar{x})$ ve $\boldsymbol{w} = (y_1 - \bar{y}, ..., y_n - \bar{y})$ ise, merkezileştirme $\bar{v} = \bar{w} = 0$ yapar. Bunları $\cos(\theta)$ tanımımızda yerine koyarak

$$\cos(\theta) = \frac{\vec{v}.\vec{w}}{\|\vec{v}\|\|\vec{w}\|} = \frac{\sum(v_i w_i)}{\sqrt{\sum v_i^2}\sqrt{\sum w_i^2}} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} = r$$

elde ederiz.

$r = \cos(\theta)$ olduğunu bulduk. Çoklu boyutta bu durumun zihinde canlandırılması çok zor olmakla beraber iki boyutlu bir durumu göz önüne alabiliriz.

R^2 de iki $\boldsymbol{v}$ ve $\boldsymbol{w}$ vektörünü göz önüne alalım ve θ bu vektörler arasındaki açı olsun. Ağırlıklarımızı, yani katsayılarımızı, pozitif olacak şekilde sınırlayarak bu vektörlerin herhangi bir $\boldsymbol{z}$ lineer kombinasyonu, $\{\boldsymbol{v}, \boldsymbol{w}\}$ kümesinin konveks koni kapanışının bir elemanı olacak. Eğer $\boldsymbol{v}$ ve $\boldsymbol{z}$ arasındaki açıyı ϕ ve $\boldsymbol{w}$ ile $\boldsymbol{z}$ arasındaki açıyı φ olacak şekilde tanımlarsak, $0 \le \phi \le \theta$ ve $0 \le \varphi \le \theta$ olduğunu biliyoruz. $\cos(x) = \cos(-x)$ olduğundan $\theta > \pi$ için $\theta = 2\pi - \theta$ yeniden tanımlayabiliriz.(Yani θ, $\boldsymbol{v}$ ve $\boldsymbol{w}$ arasındaki en küçük açıdır.)

Bu dönüşüm $0 \le \theta \le \pi$ sınırlamasını ortaya koyar ve $\cos x$ bu aralıkta kesin olarak azaldığından, $\cos(\phi) \ge \cos(\theta)$ ve $\cos(\varphi) \ge \cos(\theta)$ olduğunu biliyoruz. Bu nedenle $\cos(\theta)$ veya r korelasyon sınırımızdır.

Vektörlerin herhangi bir pozitif lineer kombinasyonu, vektörlerin kümesi tarafından tanımlanan konveks koni kapanışının bir elemanı olacağına göre, bu geometrik sezgi çoklu özellikleri (yani çoklu vektörleri) içerecek şekilde genişletilebilir.

10.BAZI YARARLI DÖNÜŞÜMLER VE DAĞILIMLAR

Burada konunun genel olarak anlaşılmasına yardımcı olan bazı yaralı dönüşüm ve dağılımları ele alacağız.

10.1Llineer Dönüşümler

$\boldsymbol{x} = \left(x_1, ..., x_p\right)'$ bir rastgele p-vektör olsun. Bazen, $\boldsymbol{x}$ in bileşenlerinin örneğin $y_1 = x_1 + x_2$ veya $y_2 = x_1 + 2x_3 - x_4$ lineer kombinasyonlarını düşünmek doğal ve yararlıdır. Genel olarak, p sayıda bileşene sahip $\boldsymbol{x}$ vektöründen

$$\boldsymbol{y} = \boldsymbol{Ax} + \boldsymbol{b} \tag{10.1}$$

ile verilen, q $\left(q < p\right)$ sayıda bileşene sahip bir $\boldsymbol{y}$ vektörüne bir dönüşümü göz önüne alalım. Burada $\boldsymbol{A}\left(q \times p\right)$ ve $\boldsymbol{b}\left(q \times 1\right)$ sabit matrislerdir

$\boldsymbol{y}$ için karşılık gelen ifadelerin $E\left(\boldsymbol{x}\right) = \boldsymbol{\mu}$ ve $V\left(\boldsymbol{x}\right) = \Sigma$ olduğunu farz edelim.

$$E\left(\boldsymbol{y}\right) = \boldsymbol{A\mu} + \boldsymbol{b} \tag{10.2a}$$

$$Var\left(\boldsymbol{y}\right) = \mathbf{A} \Sigma \boldsymbol{A}' \tag{10.2b}$$

dır. Bunlar beklenen değer operatörünün lineerliğinden,

$$\begin{aligned}
E\left(\boldsymbol{y}\right) &= E\left(\boldsymbol{Ax} + \boldsymbol{b}\right) \\
&= \boldsymbol{A}E\left(\boldsymbol{x}\right) + E\left(\boldsymbol{b}\right) \\
&= \boldsymbol{A\mu} + \boldsymbol{b} \\
&= \boldsymbol{\mu}_y \qquad \text{(diyelim)},
\end{aligned}$$

$$\begin{aligned}
Var\left(\boldsymbol{y}\right) &= E\left[\boldsymbol{yy}'\right] - \boldsymbol{\mu}_y \boldsymbol{\mu}_y' \\
&= E\left[\left(\boldsymbol{Ax} + \boldsymbol{b}\right)\left(\boldsymbol{Ax} + \boldsymbol{b}\right)'\right] - \left(\boldsymbol{A\mu} + \boldsymbol{b}\right)\left(\boldsymbol{A\mu} + \boldsymbol{b}\right)' \\
&= \mathbf{A}\,\mathrm{E}\left(\boldsymbol{xx}'\right)\boldsymbol{A}' + \boldsymbol{A}E\left(\boldsymbol{x}\right)\boldsymbol{b}' + \boldsymbol{b}E\left(\boldsymbol{x}'\right)\boldsymbol{A}' + \boldsymbol{bb}' - \boldsymbol{A\mu\mu}'\boldsymbol{A}' - \boldsymbol{A\mu b}' - \boldsymbol{b\mu A}' + \boldsymbol{bb}' \\
&= \boldsymbol{A}\left[E\left(\boldsymbol{xx}'\right) - \boldsymbol{\mu\mu}'\boldsymbol{A}'\right] \\
&= \boldsymbol{A}\Sigma\mathbf{A}'
\end{aligned}$$

dür.

10.2 Mahalonobis Dönüşümü

$E(\boldsymbol{x}) = \boldsymbol{\mu}$ ve $Var(\boldsymbol{x}) = \Sigma$ olmak üzere, p-değişkenli bir $\boldsymbol{x}$ rastgele vektörü verilmiş olsun. İlişkisiz değişkenlerin bir standartlaştırılmış kümesine bir dönüşüm; Mahalanobis dönüşümüyle verilir.

Σ nın pozitif tanımlı olduğunu, yani $\boldsymbol{x}$ de tam lineer bağımlılık olmadığını, farz edelim. Bu takdirde Σ^{-1} ters kovaryans matrisi

$$\Sigma^{-1/2} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}' \tag{10.3}$$

İle verilen bir $\Sigma^{-1/2}$ kareköküne sahiptir. Burada $\Sigma = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$ spektral (tayfsal) ayrışımdır, yani $\boldsymbol{V}$,sütunları Σ nin öz vektörleri olan bir ortogonal matristir yani $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}' = \boldsymbol{I}_p$ şeklindedir. $\boldsymbol{\Lambda} = k\ddot{o}\c{s}eg\left(\lambda_1, ..., \lambda_p\right)$ karşılık gelen öz değerlerdir. Mahalonobis dönüşümü

$$\boldsymbol{z} = \Sigma^{-1/2}\left(\boldsymbol{x} - \boldsymbol{\mu}\right) \tag{10.4}$$

biçimini alır. (10.2a) ve (10.2b) sonuçlarını kullanarak,

$$E(\boldsymbol{z}) = 0$$
$$V(\boldsymbol{z}) = \boldsymbol{I}_p$$

olduğunu gösterebiliriz. Gerçekten,

$$E(\boldsymbol{z}) = E\left[\Sigma^{-1/2}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)\right]$$

$$= \Sigma^{-1/2}\left[E(\boldsymbol{x}) - \boldsymbol{\mu}\right]$$
$$= \boldsymbol{0}$$

$$Var(\boldsymbol{z}) = \Sigma^{-1/2}\Sigma\Sigma^{-1/2}$$

$$= \boldsymbol{I}_p$$

olacaktır.

10.3 Örneklem Mahalanobis Dönüşümü

Verilen bir $\boldsymbol{X}' = (x_1, ..., x_n)$ veri matrisi için $\boldsymbol{S} = \boldsymbol{S}_x$ dönüştürülmüş bir $\frac{1}{n-1}\boldsymbol{X}'\boldsymbol{H}\boldsymbol{X}$ örneklem kovaryans matrisi olmak üzere, $i = 1, ..., n$ için $\boldsymbol{z}_i = \boldsymbol{S}^{-1/2}\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}\right)$ örneklem Mahalanobis dönüşümü bir $\mathbf{Z}' = (\boldsymbol{z}_1, ..., \boldsymbol{z}_n)$ dönüştürülmüş veri matrisini ortaya koyar. Şimdi veri matrisleri

$$\boldsymbol{Z}' = \boldsymbol{S}^{-1/2}\boldsymbol{X}'\boldsymbol{H} \quad \text{veya,}$$

$$\boldsymbol{Z} = \boldsymbol{H}\boldsymbol{X}\boldsymbol{S}^{-1/2} \tag{10.5}$$

ile bağlanırlar. Burada $\boldsymbol{H}$ merkezileştirme matrisidir. $\boldsymbol{Z}'$ nün merkezileştirilmiş olduğunu ve $\boldsymbol{S}_Z = \boldsymbol{I}_p$ olduğunu kolayca görebiliriz.

10.4 Örneklem Ölçekleme Dönüşümü

Her bir değişkeni sıfır ortalama ve bir varyansa ölçekleyen fakat korelasyon yapısını koruyan bir veri dönüşümü $\boldsymbol{D} = köşeg\left(s_1, ..., s_p\right)$ olmak üzere $i = 1, ..., n$ için $\boldsymbol{y}_i = \boldsymbol{D}^{-1/2}\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}\right)$ ile verilir. Böylece,

$$\boldsymbol{Y}' = \boldsymbol{D}^{-1}\boldsymbol{X}'\boldsymbol{H} \quad \text{veya}$$

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X}\boldsymbol{D}^{-1}$$

dir.

10.5 Bir Yararlı Matris Özdeşliği

$\boldsymbol{u}, \boldsymbol{v}$ n- vektörler ve $\boldsymbol{A} = \boldsymbol{u}\boldsymbol{v}'$ matris olsun. Bu takdirde

$$\left|1 + \boldsymbol{u}\boldsymbol{v}'\right| = 1 + \boldsymbol{v}'\boldsymbol{u}$$

dur.

İspat

İlk olarak $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{\lambda}\boldsymbol{v} \Rightarrow \left(I + \boldsymbol{A}\right)\boldsymbol{v} = \left(1 + \boldsymbol{\lambda}\right)\boldsymbol{v}$ olmasını gerektirdiğinden $\boldsymbol{A}$ ve $\boldsymbol{I} + \boldsymbol{A}'$ öz vektörlerinin ortak bir kümesini paylaştığına dikkat edelim. Bundan başka, $\boldsymbol{I} + \boldsymbol{A}'$ nın öz değerleri $1 + \lambda_i$ dir. Burada λ_i , $\boldsymbol{A}$ nın öz değerleridir.

Şimdi $\boldsymbol{uv'}$ 1 ranklı bir matristir, bu nedenle bir tek sıfır olmayan öz değerlere sahiptir. $(\boldsymbol{uv'})\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{v'u})$ olduğundan $\boldsymbol{I} + \boldsymbol{uv'}$ öz değerleri $1+\lambda$, 1 , . lerdir. Burada $\lambda = \boldsymbol{v'u}$ dur. $\boldsymbol{I} + \boldsymbol{uv'}$ nın determinantı öz değerlerin çarpımıdır, böylece sonuç elde edilir.

10.6 Temel Bileşenler Analizi

Tekniğin Ana Hatları:

$\boldsymbol{x}' = (x_1, x_2, ..., x_p)$, μ ortalamalı ve Σ kovaryans matrisli bir rastgele vektör olsun. Temel bileşenler analizi (TBA) p boyuttan $k < p$ boyuta boyutluluğun indirgenmesi için bir tekniktir. Bu analiz sıra ile, $y_1, y_2, ..., y_k$ değişkenlerinin bir kümesinin en bilgilendirici k sayıda lineer kombinasyonunu bulmaya çalışır. Burada bilgi Σ daki toplam değişimin bir yüzdesi olarak yorumlanacaktır. Bir $\boldsymbol{S}$ örneklem kovaryans matrisindeki toplam değişim %x ini açıklayan k tane örneklem temel bileşen benzer şekilde tanımlanabilir.

10.7 Formülleme

$\boldsymbol{y}_j = a_{1j}x_1 + a_{2j}x_2 + ... + a_{pj}x_p$, $\boldsymbol{a}_j^{'}\boldsymbol{a}_j = 1$ ve $j \neq k$ için, $\boldsymbol{a}_j^{'}\boldsymbol{a}_k = 0$ olacak şekilde, yani $\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_p$ ler p-vektörlerin bir ortonormal kümesini oluşturacak şekilde, $\boldsymbol{x}$ lerin standartlaştırılmış lineer kombinasyonlarının (SLK nın) bir dizisi olmak üzere,

$$
\begin{aligned}
y_1 &= \boldsymbol{a}_1^{'}\boldsymbol{x} \\
y_2 &= \boldsymbol{a}_2^{'}\boldsymbol{x} \\
&\vdots \\
y_p &= \boldsymbol{a}_p^{'}\boldsymbol{x}
\end{aligned}
$$

olsun. Eşdeğer olarak, $\{\boldsymbol{a}_j\}$ sütunlarından oluşan $\boldsymbol{A}$ $p \times p$ matrisini $\boldsymbol{A'A} = \boldsymbol{AA'} = \boldsymbol{I}_p$ olacak şekilde bir ortogonal matris olarak tanımlayabiliriz. $\boldsymbol{a}_1^{'}\boldsymbol{a}_1 = 1$ şartı altında

$$Var(y_1) = \boldsymbol{a}_1^{'}\Sigma\boldsymbol{a}_1$$

ifadesini maksimum yapacak $\boldsymbol{a}_1$ ri seçeriz.

Bundan sonra da, y_2^* ın y_1 ile ilişkisiz olacağını garantileyen $\boldsymbol{a}_2'\boldsymbol{a}_2 = 1$ ve $\boldsymbol{a}_2'\boldsymbol{a}_1 = 0$ şartı altında

$$\boldsymbol{Var}(\boldsymbol{y}_2) = \boldsymbol{a}_2' \Sigma \boldsymbol{\alpha}_2$$

ifadesini maksimum yapacak olan $\mathbf{a}_2$ yi seçeriz. Müteakip TB ler önceki TB ler ile ilişkisiz olmaya bağlı maksimum varyansa sahip olan standartlaştırılmış lineer kombinasyonlar olarak seçilirler. Bazen temel bileşenler $\boldsymbol{x}$ lerin "ortalama-düzeltilmiş" lineer dönüşümleri, yani temel bileşenlerin, bir dağılımın dağılmanın maksimumlaştığı merkezine göre p boyutlu uzaydaki doğrultu vektörleri olarak düşünülebildiğini vurgulayan,

$$\boldsymbol{y}_j = \boldsymbol{a}_j' (\boldsymbol{x} - \boldsymbol{\mu})$$

lineer dönüşümleri olarak alınırlar. Herhangi bir durumda hangi tanım kullanılırsa kullanılsın $Var(y_i)$ aynıdır.

10.8 Hesaplama

İlk olarak TB leri bulmak için, bir $f(\boldsymbol{x})$ fonksiyonunun bir $g(\boldsymbol{x}) = 0$ eşitlik kısıtlamasına bağlı maksimumunu bulmak için Lagrange çarpanları tekniğini kullanırız. λ bir Lagrange çarpanı olmak üzere,

$$L(\boldsymbol{a}_1) = \boldsymbol{a}_1' \Sigma \boldsymbol{a}_1 - \lambda(\boldsymbol{a}_1'\boldsymbol{a}_1 - 1)$$

Lagrange fonksiyonunu kullanırız. Diferansiyel alarak

$$\frac{\partial L}{\partial \boldsymbol{a}_1} = 2\Sigma \boldsymbol{a}_1 - 2\lambda \boldsymbol{a}_1 = 0$$

$$\Sigma \boldsymbol{a}_1 = \lambda \boldsymbol{a}_1$$

elde ederiz. Bu nedenle $\boldsymbol{a}_1$, Σ nın λ öz değerli bir öz vektörü olarak seçilmelidir. Σ nın öz değerlerinin farklı olduğunu ve $\lambda_1 > \lambda_2 > ... > \lambda_p > 0$ olacak şekilde azalan sırada sıralandığını farz edelim.

$$Var(y_1) = \boldsymbol{a}_1^{'} \Sigma \boldsymbol{a}_1$$
$$= \lambda \boldsymbol{a}_1^{'} \boldsymbol{a}_1$$
$$= \lambda$$

dır. Bu nedenle $\boldsymbol{a}_1, \Sigma$ nın en büyük öz değere karşılık gelen öz vektör olarak seçilmelidir.

İkinci Temel Bileşen

λ, μ Lagrange çarpanları olmak üzere Lagrange fonksiyonu

$$L(\boldsymbol{a}_2) = \boldsymbol{a}_2^{'} \Sigma \boldsymbol{a}_2 - \lambda\left(a_2^{'} a_2 - 1\right)$$
$$2 - \mu\left(\boldsymbol{a}_2^{'} \boldsymbol{a}_1\right)$$

dır. $\boldsymbol{a}_2^{'} \boldsymbol{a}_1 = 0$ olduğundan,

$$\frac{\partial L}{\partial \boldsymbol{a}_2} = 2\left(\Sigma - \lambda \boldsymbol{I}_p\right)\boldsymbol{a}_2 - \boldsymbol{\mu} \boldsymbol{a}_1 = 0$$

$$2\boldsymbol{a}_1^{'} \Sigma \boldsymbol{a}_2 - \mu = 0$$

dır. Bununla beraber,

$$\boldsymbol{a}_1^{'} \Sigma \boldsymbol{a}_2 = \boldsymbol{a}_2^{'} \Sigma \boldsymbol{a}_1$$

$$= \lambda \boldsymbol{a}_2^{'} \boldsymbol{a}_1 = 0$$

dır. Bu nedenle, $\mu = 0$ ve

$$\Sigma \boldsymbol{a}_2 = \lambda \boldsymbol{a}_2$$

dir. Bundan dolayı $\boldsymbol{a}_2$, Σ nın ikinci en büyük λ_2 öz değerlerine karşılık gelen öz vektörüdür.

Örnek:

Ölçeklendirilmiş (standartlaştırılmış) x_1 ve x_2 değişkenlerine karşılık gelen kovaryans matrisi

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

(gerçekten kovaryans matrisi) dır. Σ nın toplam değişimi 2 olan bir matris olduğuna dikkat edelim. Σ nin öz değerleri

$$|\Sigma - \lambda \boldsymbol{I}| = 0$$

$$\begin{vmatrix} 1-\lambda & \rho \\ \rho & 1-\lambda \end{vmatrix} = 0$$

$$(1-\lambda)^2 - \rho^2 = 0$$

determinant denkleminin kökleridir. Bu nedenle, $\lambda = 1+\rho$, $1-\rho$ dur. Eğer $\rho > 0$ ise, bu takdirde, $\lambda_1 = 1+\rho$, $\lambda_2 = 1-\rho$ dur. $\boldsymbol{a}_1$ ri bulmak için, $\Sigma \boldsymbol{a}_1 = \lambda \boldsymbol{a}_1$ de λ_1 i yerine koyarız.

Not : Bu, hemen $\boldsymbol{a}_1^{'} = (a_1, a_2)$ nın bileşenlerine bağlı bir

$$-\rho \boldsymbol{a}_1 + \rho \boldsymbol{a}_2 = 0$$

denklemini ortaya koyar. Bu nedenle, $a_1 = a_2$ dir.

$$\boldsymbol{a}_1^{'} \boldsymbol{a}_1 = a_1^2 + a_2^2 = 1$$

normalleştirmesini uygulayarak,

$$\boldsymbol{a}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

elde ederiz. Aynı şekilde,

$$\boldsymbol{a}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$$

elde edilir. Bu nedenle

$$y_1 = \frac{1}{\sqrt{2}}(x_1 + x_2)$$

$$y_2 = \frac{1}{\sqrt{2}}\left(x_1 - x_2\right)$$

sırasıyla toplam değişimin $\%\frac{100(1+\rho)}{2}$ ve $\%\frac{100(1-\rho)}{2}$ sini açıklayan TB lerdir. Her bir TB ile açıklanan toplam değişimin oranı ρ ya bağlı iken, TB lerin ρ dan bağımsız olduklarına dikkat edelim.

10.9 TBA ve Tayfi (Spektral) Ayrışım

Σ (aynı zamanda $\boldsymbol{S}$) bir reel simetrik matris olduğundan, onun

$$\Sigma = \boldsymbol{A\Lambda A}'$$

$$= \sum_{i=1}^{p} \lambda_i \boldsymbol{a}_i \boldsymbol{a}_i^{'}$$

spektral ayrışımına (öz analizine) sahip olduğunu biliyoruz. Burada $\boldsymbol{A}$ $(p \times p)$ matrisinin sütunları olarak girdiğimiz $\{\boldsymbol{a}_i\}$ ler, Σ nın öz vektörleridir ve $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ ler karşılık gelen öz değerleridir.

Eğer bazı öz değerler aynı iseler, şöyle ki $\lambda_k = \lambda_{k+1} = ... = \lambda_l = \lambda$ ise, öz vektörleri tek değildirler, fakat $l-k+1$ boyutlu bir alt uzayı germek için öz vektörlerin bir ortonormal kümesini seçebiliriz. ($b \to a$ iken bir $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ elipsinin büyük/ küçük eksenlerini karşılaştırınız.) Böyle bir durum eşit (eş) korelasyonlu matris ile ortaya çıkar.

Bir $\boldsymbol{x}$ p -rastgele vektörünün (onun μ ortalamasına göre düzeltilmiş) onun y p vektöründe içerilen temel bileşenlerinin (TB lerinin) kümesine bir dönüşümü

$$\boldsymbol{y} = \boldsymbol{A}'(\boldsymbol{x} - \boldsymbol{\mu})$$

dür. y_1 ; $\boldsymbol{x}$ in maksimum varyansa sahip olan lineer kombinasyonu (standartlaştırılmış lineer kombinasyonu) dur. y_2 ; y_1 ile ilişkisiz olma şartı altında maksimum varyansa sahip olan standartlaştırılmış lineer kombinasyondur v.s. $Var(y_1) = \lambda_1$,$Var(y_2) = \lambda_2, \ldots$olduğunu görürüz.

10.10 Varyansın Açıklanması

$(\boldsymbol{y})$ Temel bileşenlerinin, varyansın toplam değişimi, yani $(\boldsymbol{x})$ orijinal değişkenlerinin varyanslarının toplamını, açıklayan bileşenleri olarak yorumu aşağıdaki sonuçla açıklanır.

Sonuç:

Orijinal değişkenlerin ve onların TB lerinin varyanslarının toplamı aynıdır.

İspat:

$iz(\Sigma)$ üzerine bir not:

Bir Σ $(p\times p)$ kare matrisinin köşegen elemanlarının toplamı Σ nın izi olarak bilinir ve

$$iz(\Sigma)=\sum_{i=1}^{p}\sigma_{ii}$$

dır. Bu tanımdan $\boldsymbol{AB}$ ve $\boldsymbol{BA}$ tanımlı olduklarında, (yani $\boldsymbol{A}$ $m\times n$ ve $\boldsymbol{B}$ $n\times m$ olduğunda) $iz(\boldsymbol{AB})=iz(\boldsymbol{BA})$ olduğu, yani

$$\begin{aligned} iz(\boldsymbol{AB}) &= \sum_i\sum_j \boldsymbol{a}_{ij}\boldsymbol{b}_{ji} \\ &= \sum_j\sum_i \boldsymbol{b}_{ji}\boldsymbol{a}_{ij} \\ &= iz(\boldsymbol{BA}) \end{aligned}$$

olduğunu gösterebiliriz. TB ler için varyansların toplamı

$$\sum_i Var(y_i)=\sum_i \lambda_i = iz(\Lambda)$$

dır. Şimdi $\Sigma=\boldsymbol{A\Lambda A}'$ tayfsal ayrışımdır ve $\boldsymbol{A}$ ortonormaldir, bu nedenle $\boldsymbol{A}'\boldsymbol{A}=\boldsymbol{I}$ dır. Buradan,

$$\begin{aligned} iz(\Sigma) &= iz(\boldsymbol{A\Lambda A}') \\ &= iz(\boldsymbol{\Lambda A}'\boldsymbol{A}) \\ &= iz(\boldsymbol{\Lambda}) \end{aligned}$$

olur.

Σ, $\boldsymbol{X}$ in kovaryans matrisi olduğundan, onun köşegen elemanlarının toplamı orijinal değişkenlerin σ_{ii} varyanslarının toplamıdır. Böylece sonuç ispatlanmış olur. Bu nedenle, $\dfrac{\lambda_i}{\lambda_1+\lambda_2+\ldots+\lambda_p}$ i orijinal verideki toplam değişimin i-yinci temel bileşen tarafından açıklanan kısmı olarak yorumlamak ve

$$\frac{\lambda_1+\ldots+\lambda_k}{\lambda_1+\ldots+\lambda_p}$$

yi toplam değişiminin ilk k sayıda TB ler tarafından açıklanan kısmı olarak yorumlamak mümkündür. Bir $\boldsymbol{S}$ (10×10) örneklem kovaryans matrisi üzerindeki bir Temel Bileşen Analizinden örneğin ilk üç TB nin ($p=10$ tane TB nin bir toplamı dışında) verideki toplam değişimin %80 ini açıkladığı sonucunu çıkarabildik. Bu, verideki değişimin büyük ölçüde y_1, y_2, y_3 temel bileşenleri tarafından belirlenen bir 3-boyutlu alt uzaya bağlandığını (sınırlandığını) ifade etmelidir.

10.11 Ölçek (Ölçü) Değişmezliği

Bu maalesef TBA nın sahip olmadığı bir özelliktir. Uygulamada ekseriya farklı $\{\boldsymbol{x}_i\}$ değişkenlerimiz için ölçüm birimlerini seçmeliyiz ve özel bir $\boldsymbol{x}_i$ değişkeniyle açıklanan toplam değişimin miktarı bu seçime (ton, kg veya gram) bağlıdır.

Bir uygulama çalışmasında $\boldsymbol{x}$ veri vektörü ekseriya fiziksel olarak karşılaştırılamaz niceliklerden oluşur.(örneğin, boy, ağırlık, sıcaklık) Bu nedenle, benimsenecek aritmetik ölçekleme yoktur. Bir olabilirlik, bir korelasyon matrisi (her bir değişkeni birim örneklem varyansına sahip olacak şekilde seçerek) üzerinde Temel Bileşen Analizini uygulamaktır. Fakat bu hala ölçeklemenin dolaylı olarak anlaşılan bir seçimidir. Asıl mesele bir Temel Bileşen Analizinin benimsenen ölçeklemeye bağlı olmasıdır.

10.12 TEMEL BİLEŞEN SAYILARI

Bir $\boldsymbol{X}$ veri matrisi üzerinde örneklem TB dönüşümü r-yinci birey (örneklemin r-yinci satırı) için,

$$\boldsymbol{y}_r^{'} = \boldsymbol{A}'\left(\boldsymbol{x}_r - \overline{\boldsymbol{x}}\right)$$

biçimini alır. Burada $\boldsymbol{A}$ nın sütunları $\boldsymbol{S}$ örneklem kovaryans matrisinin öz vektörleridir. Birinci y_1 bileşeninin $x_r^{'}$ ile $\boldsymbol{A}$ nın birinci sütununun skalar çarpımına karşılık geldiğine v.s dikkat edelim. y_r nin bileşenleri, *r*-yinci birey için temel bileşen sayıları (ortalama- düzeltilmiş) olarak bilinirler.

$$\boldsymbol{y}_r = \boldsymbol{A}'\boldsymbol{x}_r$$

sayıları r-yinci birey için ham TB sayılarıdır. Geometrik olarak, TB sayıları her bir veri noktasının TB ler ile tanımlanan yeni eksenlere göre, yani bir döndürülmüş koordinat sistemine göre, koordinatlarıdır. Sayılar bireyler hakkında nitel bilgi verilebilir.

10.13 Orijinal değişkenler ile TB lerin Korelasyonu

x_i değişkeni ile *k*-yıncı TB nin $\rho\left(x_i, y_k\right)$ korelasyonları TB leri yorumlamak için bir yardımcıdır.

$$\boldsymbol{y} = \boldsymbol{A}'(\boldsymbol{x} - \boldsymbol{\mu})$$

olduğundan,

$$\begin{aligned} kov(\boldsymbol{x}, \boldsymbol{y}) &= E\left[(\boldsymbol{x} - \boldsymbol{\mu})\boldsymbol{y}'\right] \\ &= E\left[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{A}\right] \\ &= \Sigma \boldsymbol{A} \end{aligned}$$

ve tayfsal (tayfi) ayrışımdan,

$$\begin{aligned} \Sigma \boldsymbol{A} &= \left(\boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}'\right)\boldsymbol{A} \\ &= \boldsymbol{A}\boldsymbol{\Lambda} \end{aligned}$$

elde ederiz.

$\boldsymbol{A}$ yı bir $\boldsymbol{\Lambda}$ köşegen matrisiyle sondan çarpma onun sütunlarını ölçeklemenin etkisine sahiptir. Bu nedenle,

$$kov\left(x_i, y_k\right) = \lambda_k a_{ik}$$

i-yinci değişken ve *k*-yıncı TB arasındaki kovaryanstır.

$$\rho\left(x_i, y_k\right) = \frac{kov\left(x_i, y_k\right)}{Var\left(x_i\right) Var\left(y_k\right)}$$

$$= \frac{\lambda_k a_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\lambda_k}}$$

$$= a_{ik}\left(\frac{\lambda_k}{\sigma_{ii}}\right)^{1/2}$$

korelasyonu x_i deki değişimin k-yıncı TB tarafından açıklanan kısmı olarak yorumlanabilir. (Ayrıntılar için bkz. Ek-2)

SONUÇ VE ÖNERİLER

On bölümden oluşan bu çalışmada istatistikte korelasyon katsayısı tanıtıldıktan sonra bu katsayıya çeşitli yönlerden bakışlar ele alındı. Bazı geometrik yorumlarla korelasyon kavramı açıklanmaya çalışıldı. Bunun yanında olasılık teorisinde olayların bağımlılık ve bağımsızlığı korelasyon katsayısı kavramı ile ilişkilendirildi. Ayrıca kanonik korelasyon katsayısı ayrıntılı bir açıklamayla verildi. Mükemmel negatif korelasyon, mükemmel pozitif korelasyon, mükemmel negatif korelasyondan uzaklık ve mükemmel pozitif korelasyondan uzaklık kavramlarına da yer verildi. Pearson korelasyon katsayısı ve kosinüs benzerliği arasında ilişki kuruldu. Bunları yaparken bazı yararlı dönüşümler ve dağılımlar hakkında bilgi verildi.

Bu tezde sunulan bilgiler, korelasyon kavramı için istatistiksel uygulamalarda karşılaşılan birçok güçlüğü yenecektir.

KAYNAKLAR

Ahlgren, P. Jarneving, B. and Rousseau, R. (2003). Requirements for a cocitation similarity measure, with special reference to Pearson’s correlation coefficient. Journal of the American Society for Information Science and Technology 54(6), 550-560.

Ahlgren P, Jarneving B, and Rousseau R (2004). Autor cocitation and Pearson’s r. Journal of the American Society for Information Science and Technology 55(9), 843.

Bensman, S. J. (2004). Pearson’s r and Author Cocitation Analysis: A commentary on the controversy. Journal of the American Society for Information Science and Technology 55(10), 935-936.

Betro, B.(1993), "On the Distribution of Gini's Rank Correlation Association Coefficient",Communications in Statistics: Simulation and Computation, M. Dekker, 22,No. 2, 497-505.

Boyce, B.R., Meadow C.T. and. Kraft D.H (1995). Measurement in Information Science. Academic Press, New York, NY, USA.

Brandes, U. and Pich, C. (2007). Eigensolver Methods for Progressive Multidimensional Scaling of Large Data. In M. Kaufmann & D. Wagner (Eds.), Graph Drawing, Karlsruhe, Germany, September 18-20, 2006 (Lecture Notes in Computer Science, Vol. 4372, pp. 42-53). Berlin, Heidelberg: Springer.

Chatillon, G. ”The Balloon Rules for a Rough Estimate of the Correlation Coefficient”, The American Statistician, Vol. 38, No. 1. (Feb., 1984), p. 58

Durbin,J. (1960), "Estimation of parameters in time-series regression models," Journal of the Royal Statistical Society, Series B, 22 , 139-153.

Egghe, L. and Leydesdorff, L. Journal of the American Society for Information Science & Technology (forthcoming),”The relation between Pearson’s Correlation Coefficient r and Salton’s Cosine Measure.”

Egghe, L. and Rousseau, R. (1990). Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science. Elsevier, Amsterdam.

Egghe, L. and Rousseeau, R. (2001). Elementary Statistics for Effective Library and Information Service Management. Aslib imi, London, UK.

Egghe, L. and Michel, C. (2002). Strong similarity measures for ordered sets of documents in information retrieval. Information Processing and Management 38(6), 823-848.

Egghe, L. and Micheli, C. (2003). Construction of weak and strong similarity measures for ordered sets of documents using fuzzy set techniques. Information Processing and Management 39(5), 771-807.

Egghe, L. (2008). New relations between similarity measures for vectors based on vector norms. Preprint. 23

Frandsen, T.F. (2004). Journal diffusion factors – a measure of diffusion Aslib Proceedings: new Information Perspectives 56(1), 5-11.

Ferson, S. Roger B. Nelsen, Janos Hajagos, Daniel J. Berleant, Jianzhong Zhang, W. Troy Tucker, Lev R. Ginzburg and William L. Oberkampf " Dependence in Probabilistic modeling, Demster-Shafer theory, and probabbility bounds analysis " Sand 2004-3072, Unlimited release, Printed October 2004.

Gideon, R.A. Universty of Montana,email:gideon@selway.umt.edu, Depertmant of Mathematical Sciences, " A Generalize İnterpretation of Pearson' s r ."(PDF)

Gideon, R.A. and Hollister, R.A. (1987), "A Rank Correlation Coefficient Resistant to Outliers," Journal of the American Statistical Association 82,no.398, 656-666.

Gini, C. (1914), "L'Ammontare c la Composizione della Ricchezza della Nazioni", Bocca, Torino.

Gnanadesikan, R. (1977), Methods for Statistical Data Analysis of Multivariate Observations, John Wiley and Sons.

Gnanadesikan, R. and Kettenring, J.R. (1972), "Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data, Biometrics, 28, 81-124.

Grossman, D.A and Frieder, O (1998). Information Retrieval Algorithms and Heuristics. Kluwer Academic Publishers, Boston, MA, USA.

Hardy, G. , Littlewood, J.E and Pólya, G. (1988). Inequalities. Cambridge University Press, Cambridge, UK.

Huber, P.J. (1981), Robust Statistics John Wiley and Sons http: sites.stat.psu.edu/~jls/stat511/lectures/lec6.pdf

https://www.google.com.tr/#q=the+correlation +boundary ,"The correlation Boundary",Newfound Research, LLC 425 Boylston Street, 3rd Floor Boston, MA 02116

http://www.maths.manchester.ac.uk/~mkt/MT37332%20(MVA)/Ir , "İntroduction to Multivariate Data" Chatfeld, C. and A.J.Collins, Introduction to multivariate analysis. Chapman & Hall

Jaccard, P. (1901). Distribution de la flore alpine dans le Bassin des Drouces et dans quelques regions voisines. Bulletin de la Société Vaudoise des Sciences Naturelles 37(140), 241–272

Johnson, R.A.and Wichern, D.W. Applied multivariate statistical analysis. Prentice Hall T. Tanimoto (1957). Internal report: IBM Technical Report Series, November, 1957.

Jones, W. P. and Furnas, G. W. (1987). Pictures of relevance: a geometric analysis of similarity measures. Journal of the American Society for Information Science 36(6), 420-442.

Joshep L. Rodgers, Nicewander, W.A" Thirteen Ways to Look at the Correlation Coefficient", The American Statistician, Vol.42, No 1.(Feb.,1988) pp. 59-66

Kamada, T., and Kawai, S. (1989). An algorithm for drawing general undirected graphs. Information Processing Letters, 31(1), 7-15.

Kendall, M.G. and Gibbons, J.D. (1990), Rank Correlation Methods, 5th ed. Oxford University Press, or also Kendall, M.G. (1962), Rank Correlation Methods , 3rd ed. Hafner Publ. Co.

Kruskal, J. B., and Wish, M. (1978). Multidimensional Scaling. Beverly Hills, CA: Sage Publications.

Leydesdorff, L. (2007a). Visualization of the citation impact environments of scientific journals: an online mapping exercise. Journal of the American Society of Information Science and Technology 58(1), 207-222.

Leydesdorff, L. (2007b). Should co-occurrence data be normalized ? A rejoinder. Journal of the American Society for Information Science and Technology 58(14), 2411-2413.

Leydesdorff, L. (2008). On the normalization and visualization of author co-citation data: Salton's cosine versus the Jaccard index. Journal of the American Society for Information Science and Technology 59(1), 77-85. 24

Leydesdorff, L. and Cozzens, S.E. (1993). The delineation of specialties in terms of journals using the dynamic journal set of the Science Citation Index. Scientometrics 26, 133-154.

Leydesdorff, L. and Hellsten, I. (2006). Measuring the meaning of words in contexts: an automated analysis of controversies about 'Monarch butterflies,' 'Frankenfoods,' and 'stem cells'. Scientometrics 67(2), 231-258

Leydesdorff, L. and Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: extending ACA to the Web environment. Journal of the American Society for Information Science and Technology 57(12), 1616-1628.

Leydesdorff, L. and Zaal, R. (1988). Co-words and citations. Relations between document sets and environments. In L. Egghe and R. Rousseau (Eds.), Informetrics 87/88, 105-119, Elsevier, Amsterdam

Liang, K.Y. and Zeger, S.L. (1995), "Inference Based on Estimating Functions in the Presence of Nuisance Parameters," Statistical Science 10, no. 2, 158-172.

Losee, R.M. (1998). Text Retrieval and Filtering: Analytical Models of Performance. Kluwer Academic Publishers, Boston, MA, USA.

Marks, E."A Note on a Geometric Interpretation of the Correlation Coefficient",Journal of Educational Statistics, Vol. 7, No. 3. (Autumn, 1982), pp. 233-237.

Oehlert, G. email:gary@stat.umn.edu, "Cannonicial Correlation" (PDF)

Pearson, K. "Notes on the History of Correlation",Biometrika, Vol. 13, No. 1. (Oct., 1920), pp. 25-45.

Rodgers, J.L. and Nicewater, W.A. (1988), "Thirteen Ways to Look at the Correlation Coefficient," The American Statistician, 42, no. 1, 59-66.

Rousseeuw, P.J. and Croux, C. (1993), "Alternatives to the Median Absolute Deviation," Journal of the American Statistical Association , 88, 1273-1283.

Salton, G. and McGill, M.J (1987). Introduction to Modern Information Retrieval. McGraw-Hill, New York, NY, USA.

Scarsini, M. (1984), "On Measures of Condordance", Stochastica, 8, No.3, 201-218.

Schweizer, B. and Wolfe, E.F. (1981), "On Nonparametric Measures of Dependence for Random Variables," The Annals of Statistics , 9, 879-885.

Sen, P.K. (1968), "Estimates of the Regression Coefficient Based on Kendall's Tau," Journal of the American Statistical Association, 63, 1379-1389.

Small, H. (1973). Co corelatation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for Information Science 24(4), 265-269.

Spearman, C.(1906), " 'Footrule' for Measuring Correlations," British Journal of Psychology, 2,89-108.

Tague, J. Sutcliffe, J. (1995). Measuring Information: An Information Services Perspective. Academic Press, New York, NY, USA.

Van, C.J. Rijsbergen (1979). Information Retrieval. Butterworths, London, UK.

Waltman, L. and Van Eck, N.J. (2007). Some comments on the question whether co-occurrence data should be normalized. Journal of the American Society for Information Science and Technology 58(11), 1701-1703.

Wasserman, S. And Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press, New York, NY, USA.

White, H.D. (2003). Author cocitation analysis and Pearson’s r. Journal of the American Society for Information Science and Technology 54(13), 1250-1259

EK-1 GÖSTERGE FONKSİYONLARI

Ω bir örneklem uzayı olsun. $E \in \Omega$ bir olay olsun ve onun olasılığı $P(E)$ ile gösterilsin. Bir E olayının I_E ile gösterilen gösterge fonksiyonu (veya gösterge rasgele değişkeni), aşağıdaki gibi tanımlanan bir rasgele değişkendir.

$$I_{E(\omega)} = \begin{cases} 1 & \text{eğer } \omega \in E \text{ ise} \\ 0 & \text{eğer } \omega \notin \text{E ise} \end{cases}$$

Başka bir deyişle E olayının gösterge fonksiyonu E olduğunda 1 değerini ve E olmadığında 0 değerini alan bir rasgele değişkendir.

Örnek: düzgün bir oyun zarını attığımızda onun üste gelen yüzünde 1'den 6'ya kadar sayılar gözükür. Bu nedenle örnek uzayı $\Omega = \{1, 2, 3, 4, 5, 6\}$ dir. "üste gelen yüzde bir tek sayı gözükmesi" cümlesiyle ifade edilen

$$E = \{1, 3, 5\}$$

olayını tanımlayalım. Üste gelen yüzde tek bir sayı gözüktüğünde 1 değerini ve aksi takdirde 0 değerini alan bir rasgele değişken E olayının bir göstergesidir.

Yukarıdaki tanımda, I_E

$$R_{I_E} = \{0,1\}$$

değerli ise,

$$P_{I_E}(x) = \begin{cases} P(E) & \text{eğer x=1 ise,} \\ P(E^c) = 1 - P(E) & \text{eğer x=0 ise,} \\ 0 & \text{diğer durumlarda} \end{cases}$$

olasılık dağılım fonksiyonlu bir kesikli rasgele değişkendir. Gösterge fonksiyonları teoremleri ispatlamak ve kavramları basitleştirmek için olasılık teorisinde sık sık kullanılırlar.

Özellikler

Gösterge fonksiyonları aşağıdaki özelliklere sahiptir.

1.Kuvvetler

I_ε ,0 veya 1 den biri ve $0^n = 0$, $1^n = 1$ olduğundan I_n nin n-yinci kuvveti I_ε ye eşittir, yani

$$\left(I_{E(w)}\right)^n = I_{E(w)} \text{ , } \forall n, w$$

dır.

2.Beklenen Değer

I_E nin beklenen değeri $P(E)$ ye eşittir:

$$E[I_E] = \sum_{x \in R_{I_E}} xPI_E(x) = 1.PI_E(1) + 0P_{I_E}(0)$$

$$= 1.P(E) + 0.P(E^c) = P(E)$$

yazılabilir.

3.Varyans

I_E varyansı $P(E)(1-P(E))$ ye eşittir. Bilinen varyans formülü ve yukarıdaki kuvvetler özelliği nedeniyle

$$\begin{aligned}\operatorname{var}(I_E) &= E\left[(I_E)^2\right] - \left[E(I_E)\right]^2 \\ &= E(I_E) - \left[E(I_E)\right]^2 \\ &= P(E) - P(E)^2 \\ &= P(E)(1-P(E))\end{aligned}$$

elde ederiz.

4.Kesişimler

E ve F iki olay ise bu takdirde $I_{E\cap F}=I_E I_F$ dir. Çünkü eğer $w\in E\cap F$ ise bu takdirde $I_{E\cap F}(w)=1$ ve $w\in E$, $w\in F$ $\Rightarrow I_E(w)=1$, $I_F(w)=1$ $\Rightarrow I_E(w)I_F(w)=1$ dir.

5.Birleşimler

A ve B bir örnek uzayın herhangi iki olayı olsun. Her $A,B\in S$ (örnek uzay) için $I_{A\cap B}=I_A I_B$ olduğunu biliyoruz. Bu nedenle

$$
\begin{aligned}
I_{A\cup B} &= 1-I_{A^c\cap B^c}\\
&=1-I_{A^c}I_{B^c}\\
&=1-(1-A)(1-B)\\
&=I_A+I_B-I_A I_B
\end{aligned}
$$

dir. Her iki tarafın beklenen değerini alarak $P(A\cup B)=P(A)+P(B)-P(A\cap B)$ yani bildiğimiz formülü elde ederiz. Bunu genelleştirebiliriz. A_j $j=1,2,...,n$ olayların bir koleksiyonuna sahip olduğumuzu ve $B=\bigcup_{j=1}^{m}A_j$ olayının olasılığıyla ilgilendiğimizi farz edelim. $I_B=1-\prod_{j=1}^{n}\left(1-I_{Aj}\right)$ olduğuna dikkat edelim. Herhangi $a_1,a_2,....,a_n$ reel sayıları için kolayca gerçekleşen $\prod_{j=1}^{n}\left(1-a_j\right)=1-\sum_{1\le j\le n}a_j+\sum_{1\le i\le n}a_i a_j-\sum_{1\le i<j<k\le n}+...+(-1)^n a_1...a_n$ ifadesini elde ederiz. şimdi a_j yerine I_{A_j} koyar ve bundan sonra her iki tarafında beklenen değerlerini alırsak aşağıdaki bildiğimiz formülü de elde ederiz.

$$
P(B)=\sum_{i\le j<n}^{n}P(A_j)-\sum_{1\le i<j\le n}P(A_i\cap A_j)-\sum_{1\le n<j<k\le n}P(A_i\cap A_j\cap A_k)+...+(-1)^n P(A_1\cap...\cap A_n)
$$

$w\notin E\cap F$ ise bu takdirde $I_{E\cap F}(w)=0$ ve ya $w\notin E$ ya $w\notin F$ ya da w $w\notin E$ dir.

Aynı şekilde ya $I_E(w)=0$ ya $I_F(w)=0$ ya da $I_E(w)=0$ ve $I_F(w)=0$ dır.

EK-2

EK 2.1 Çok Değişkenli Normal Dağılım

Çok değişkenli normal dağılım (ÇDND) , μ dağılımının ortalaması ve σ^2 varyansı olmak üzere,

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} \ , \ -\infty < x < \infty$$

olasılık yoğunluk fonksiyonuna (o.y.f) sahip olan bir değişkenli normal dağılımın bir genellemesidir. p -boyutta yoğunluk

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{x}-\mu)' \Sigma^{-1} (\boldsymbol{x} - \mu) \right\} \qquad \text{(i)}$$

olur. Burada $\boldsymbol{\mu}$ ortalama vektörü içinde p tane (bağımsız) parametre vardır ve simetrik Σ kovaryans matrisi içinde $\frac{1}{2} p(p+1)$ tane bağımsız parametre (toplamda $\frac{1}{2} p(p+3)$ tane bağımsız parametre) vardır.

$$E(\boldsymbol{x}) = \boldsymbol{\mu}$$

beklenen değer vektörüne ve,

$$kov(\boldsymbol{x}) = \Sigma$$

kovaryans matrisine

sahip olan, ÇDND sahip bir x rastgele vektörünü göstermek için

$$\boldsymbol{x} \sim \boldsymbol{N}_p(\boldsymbol{\mu}, \Sigma) \qquad \text{(ii)}$$

notasyonunu kullanırız. ÇDND ların, dağılım birinci ve ikinci momentleriyle tam olarak belirlendiklerine dikkat edelim.

EK 2.2 Temel Özellikler

Eğer $\boldsymbol{x}$ $(p \times 1)$ rastgele vektörü $\boldsymbol{\mu}$ ortalamalı ve Σ kovaryans matrisli ÇDND a sahipse,

- $\boldsymbol{x}$ in herhangi bir lineer kombinasyonu ÇDND' a sahiptir.

$\boldsymbol{A}$ $(q \times p)$ ve $\boldsymbol{c}$ $(q \times 1)$ olmak üzere $\mathbf{y} = \mathbf{Ax} + \mathbf{c}$ olsun. Bu takdirde $\boldsymbol{\mu}_y = \mathbf{A\mu} + \mathbf{c}$ ve $\boldsymbol{\Sigma}_y = \boldsymbol{A}\boldsymbol{\Sigma}\mathbf{A}'$ olmak üzere

$$\boldsymbol{y} \sim \boldsymbol{N}_q\left(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y\right)$$

dir.

- $\boldsymbol{x}$' deki değişkenlerin herhangi bir alt kümesi bir ÇDND' a sahiptir.
- Eğer değişkenlerin bir kümesi ilişkisiz iseler, bu takdirde onlar bağımsız olarak dağılırlar. Özellikle,

i. $\sigma_{ij} = 0$ ise, bu takdirde $\boldsymbol{x}_i, \boldsymbol{x}_j$ bağımsızdırlar.

ii. Eğer $\boldsymbol{x}$, $\boldsymbol{\Sigma}$ kovaryans matrisli ÇDND' a sahipse, bu takdirde $\boldsymbol{Ax}$ ve $\boldsymbol{Bx}$ in bağımsız olması için gerek ve yeter şart

$$Cov\left(\boldsymbol{Ax}, \boldsymbol{Bx}\right) = \boldsymbol{A}\boldsymbol{\Sigma}\mathbf{B}' = \mathbf{0} \qquad \text{(iii)}$$

olmasıdır.

- Şartlı dağılımlar ÇDND lardır.

Sonuç Ek 2.1

ÇDND için, değişkenlerin ilişkisiz olması için gerek ve yeter şart değişkenlerin bağımsız olmasıdır.

İspat.

$\boldsymbol{x}$ $(p \times 1)$ rastgele vektörü

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \begin{matrix} q \\ p-q \end{matrix}$$

ortalama vektörü ve

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{matrix} q \\ p-q \end{matrix}$$

kovaryans matrisine sahip olmak üzere

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_{p-q}^{q}$$

şeklinde parçalanmış olsun.

i. Bağımsızlık ilişkisiz olmayı gerektirir. x_1, x_2 nin bağımsız olduklarını farz edelim. Bu takdirde $\Sigma_{12} = kov(x_1, x_2) = E\left[(x_1 - \mu_1)(x_2 - \mu_2)'\right]$, $E(x_1) = \mu_1$ ve $E(x_2) = \mu_2$ olduğundan her ikisi de sıfır olan $E\left[(x_1 - \mu_1)\right]$ ve $E\left[(x_2 - \mu_2)'\right]$ nin çarpımı şeklinde çarpanlarına ayrılır. Bu nedenle $\Sigma_{12} = 0$ dır.

ii. İlişkisizlik bağımsızlığı gerektirir. (ÇDND için)

$\Sigma_{12} = 0$ olduğunda bu sonuç (1) o.y.f 'nun nu çarpanlarına ayırmaya bağlıdır. Bu durumda $(x-\mu)'\Sigma^{-1}(x-\mu)$:

$$\begin{aligned}
&\left[x_1' - \mu_1' \quad x_2' - \mu_2'\right] \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \\
&= \left[x_1' - \mu_1' \quad x_2' - \mu_2'\right] \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \\
&= (x_1 - \mu_1)'\Sigma_{11}^{-1}(x_1 - \mu_1) + (x_2 - \mu_2)'\Sigma_{22}^{-1}(x_2 - \mu_2)
\end{aligned}$$

parçalanmış formuna sahiptir. Bu nedenle $\exp\{(x-\mu)'\Sigma^{-1}(x-\mu)\}$ $\exp\{(x_1-\mu_1)'\Sigma_{11}^{-1}(x_1-\mu_1)\}$ ve $\exp\{(x_2-\mu_2)'\Sigma_{22}^{-1}(x_2-\mu_2)\}$ çarpımı biçiminde çarpanlarına ayrılır. Böylece (o.y.f) x_1 ve x_2 nin bağımsız olduklarını gösteren o.y.f

$$f(x) = g(x_1)h(x_2)$$

olarak yazılabilir.

Sonuç Ek 2.2

$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_{p-q}^{q}$, $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}_{p-q}^{q}$ ortalamalı ve $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}_{p-q}^{q}$ kovaryans matrisli ÇDND' a sahip olsun. x_1 verildiğinde x_2 nin şartlı dağılımı

$$E(x_2 | x_1) = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1) \qquad \text{(iv)}$$

ortalamalı ve

$$Kov\left(\boldsymbol{x}_2 \,|\boldsymbol{x}_1\right) = \Sigma_{22} + \Sigma_{21}\,\Sigma_{11}^{-1}\,\Sigma_{12} \tag{v}$$

kovaryans matrisli ÇDND dır.

İspat: $\boldsymbol{x}_2^{'} = \boldsymbol{x}_2 - \Sigma_{21}\,\Sigma_{11}^{-1}\,\boldsymbol{x}_1$ olsun. İlk olarak $\boldsymbol{x}_2^{'}$ ve $\boldsymbol{x}_1$ rin bağımsız olduklarını gösteririz.

$$\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & 0 \\ -\Sigma_{21}\,\Sigma_{11}^{-1} & \mathbf{I} \end{bmatrix} \tag{vi}$$

$$= \boldsymbol{Ax} \text{ (diyelim)} \tag{vii}$$

lineer dönüşümünü göz önüne alalım. Bu lineer ilişki$\left(\boldsymbol{x}_1, \boldsymbol{x}_2^{'}\right)$ nün ortaklaşa ÇDND (ÇDND ın yukarıda ifade edilen birinci özelliğine göre) olduklarını gösterir.

$\boldsymbol{x}_1$ ve $\boldsymbol{x}_2$ nin ilişkisiz olduklarını iki şekilde gösterebiliriz. İlk olarak

$$\begin{aligned} kov\left(\boldsymbol{x}_1, \boldsymbol{x}_2^{'}\right) &= kov\left(\boldsymbol{x}_1, \boldsymbol{x}_2 - \Sigma_{21}\,\Sigma_{11}^{-1}\,\boldsymbol{x}_1\right) \\ &= kov\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) - kov\left(\boldsymbol{x}_1, \boldsymbol{x}_1\right)\Sigma_{11}^{-1}\,\Sigma_{12} \\ &= \Sigma_{12} - \Sigma_{11}\,\Sigma_{11}^{-1}\,\Sigma_{12} \\ &= 0 \end{aligned}$$

veya eğer $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{C} \end{bmatrix}$ yazarsak,

$$kov\left(\boldsymbol{x}_1, \boldsymbol{x}_2^{'}\right) = kov\left(\boldsymbol{Bx},\boldsymbol{Cx}\right)$$

$$= \boldsymbol{B}\Sigma\boldsymbol{C}'$$

$$kov\left(\boldsymbol{x}_1, \boldsymbol{x}_2^{'}\right) = \begin{bmatrix} \boldsymbol{I} & 0 \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \Sigma_{11}^{-1}\Sigma_{12} \\ \boldsymbol{I} \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \end{bmatrix} \begin{bmatrix} \Sigma_{11}^{-1}\Sigma_{12} \\ \boldsymbol{I} \end{bmatrix}$$

$$= \mathbf{0}$$

elde ederiz.

ÇDND ve ilişkisizlikten $\boldsymbol{x}_2^{'}$ ve $\boldsymbol{x}_1$ rin bağımsız olduklarını gösterdik. Bu nedenle

$$E\left(\boldsymbol{x}_2 | \boldsymbol{x}_1\right) = E\left(\boldsymbol{x}_2^{'}\right)$$

$$= E\left(\boldsymbol{x}_2 - \Sigma_{21} \Sigma_{11}^{-1} \boldsymbol{x}_1\right)$$

$$= \boldsymbol{\mu}_2 - \Sigma_{21} \Sigma_{11}^{-1} \boldsymbol{\mu}_1$$

dir. Şimdi $\boldsymbol{x}_2^{'} = \boldsymbol{x}_2 - \Sigma_{21} \Sigma_{11}^{-1} \boldsymbol{x}_1$ olduğundan,

$$E\left(\boldsymbol{x}_2^{'} | \boldsymbol{x}_1\right) = E\left(\boldsymbol{x}_2^{'} | \boldsymbol{x}_1\right) + \Sigma_{21} \Sigma_{11}^{-1} \boldsymbol{x}_1$$

$$= \boldsymbol{\mu}_2 - \Sigma_{21} \Sigma_{11}^{-1} \boldsymbol{\mu}_1 + \Sigma_{21} \Sigma_{11}^{-1} \boldsymbol{x}_1$$

$$= \boldsymbol{\mu}_2 + \Sigma_{21} \Sigma_{11}^{-1} \left(\boldsymbol{x}_1 - \boldsymbol{\mu}_1\right)$$

dir. $\boldsymbol{x}_1$ ve $\boldsymbol{x}_2^{'}$ bağımsız olduklarından

$$kov\left(\boldsymbol{x}_2^{'} | \boldsymbol{x}_1\right) = kov\left(\boldsymbol{x}_2^{'}\right)$$

$\boldsymbol{x}_1$ re verilen bir sabit olma şartı, $\boldsymbol{x}_2^{'} = \boldsymbol{x}_2 - \Sigma_{21} \Sigma_{11}^{-1} x_1$ olmasıdır, yani $\boldsymbol{x}_2^{'}$ ve $\boldsymbol{x}_2$ nin bir sabit kadar fark etmesidir. Bu nedenle,

$$kov\left(\boldsymbol{x}_2 | \boldsymbol{x}_1\right) = kov\left(\boldsymbol{x}_2^{'} | \boldsymbol{x}_1\right)$$

dır. Böylece $\boldsymbol{C} = \begin{bmatrix} -\Sigma_{21} \Sigma_{11}^{-1} & \boldsymbol{I} \end{bmatrix}$ olmak üzere,

$$kov\left(\boldsymbol{x}_2 | \boldsymbol{x}_1\right) = kov\left(\boldsymbol{x}_2^{'}\right)$$

$$= \boldsymbol{C} \Sigma \boldsymbol{C}'$$

dır. Bu nedenle de,

$$\begin{bmatrix} -\Sigma_{21} \Sigma_{11}^{-1} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \Sigma_{11}^{-1} \Sigma_{12} \\ \boldsymbol{I} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \end{bmatrix} \begin{bmatrix} \Sigma_{11}^{-1} \Sigma_{12} \\ \boldsymbol{I} \end{bmatrix}$$

$$= \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$$

dır.

Örnek:

$\boldsymbol{x}$, $\Sigma = \begin{bmatrix} 1 & \rho & \rho^2 \\ & 1 & 0 \\ & & 1 \end{bmatrix}$ kovaryans matrisi bir ÇDND a sahip olsun. $\boldsymbol{x}_3$ verildiğinde $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ nin şartlı dağılımının aynı zamanda

$$\boldsymbol{\mu} = \begin{bmatrix} \mu + \rho^2 (x_3 - \mu_3) \\ \mu_2 \end{bmatrix}$$

ortalamalı, ve

$$\begin{bmatrix} 1-\rho^4 & \rho \\ \rho & 1 \end{bmatrix}$$

kovaryans matrisli ÇDND dır.

EK 2.3 En Çok Olabilirlik Tahmini

$\boldsymbol{X}' = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)$, $N_p(\boldsymbol{\mu}, \Sigma)$ kitlesinden n-boyutlu bir bağımsız rasgele örneklemi içersin. $\boldsymbol{\mu}$, Σ nın en çok olabilirlik tahminleri (EÇOT leri)

$$\hat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}} \tag{viii}$$

ve

$$\hat{\Sigma} = \boldsymbol{S} \tag{ix}$$

dir. $\boldsymbol{X}$ veri matrisi verildiğinde

$$L(\boldsymbol{\mu}, \Sigma | \boldsymbol{X}) = \prod_{r=1}^{nn} f(\boldsymbol{x}_r | \boldsymbol{\mu}, \Sigma) \tag{x}$$

en çok olabilirlik fonksiyonu $\boldsymbol{\mu}$, Σ parametrelerin bir fonksiyonudur. Sağ yan $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$ farklı veri vektörlerinin $N_p(\boldsymbol{\mu}, \Sigma)$ nın olasılık yoğunluk fonksiyonunda sırayla yerlerine koyarak ve çarpımı alarak belirlenir. Buna göre,

$$\prod_{r=1}^{n} f(\boldsymbol{x}_r | \boldsymbol{\mu}, \Sigma) = (2\pi)^{\frac{-np}{2}} |\Sigma|^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{r=1}^{n} (\boldsymbol{x}_r - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x}_r - \boldsymbol{\mu}) \right\}$$

olacaktır. K , $\boldsymbol{\mu}$ ve Σ dan bağımsız bir vektör olmak üzere, L yi maksimumlaştırma

$$l = -2\log L$$
$$= \sum_{r=1}^{n} \log f\left(\boldsymbol{x}_r \,|\, \boldsymbol{\mu}, \Sigma\right)$$
$$= K + n\log|\Sigma| + \sum_{r=1}^{n}\left(\boldsymbol{x}_r - \boldsymbol{\mu}\right)'\Sigma^{-1}\left(\boldsymbol{x}_r - \boldsymbol{\mu}\right)$$

ifadesini minimumlaştırmaya eş değerdir. $\boldsymbol{x}_r - \boldsymbol{\mu} = \left(\boldsymbol{x}_r - \overline{\boldsymbol{x}}\right) + \left(\overline{\boldsymbol{x}} - \boldsymbol{\mu}\right)$ olduğuna dikkat ederek yukarıdaki son terim

$$\sum_{r=1}^{n}\left(\boldsymbol{x}_r - \overline{\boldsymbol{x}}\right)'\Sigma^{-1}\left(\overline{\boldsymbol{x}_r} - \overline{\boldsymbol{x}}\right) + \sum_{r=1}^{n}\left(\boldsymbol{x}_r - \overline{\boldsymbol{x}}\right)'\Sigma^{-1}\left(\overline{\boldsymbol{x}} - \boldsymbol{\mu}\right)$$
$$+\sum_{r=1}^{n}\left(\boldsymbol{x}_r - \overline{\boldsymbol{x}}\right)'\Sigma^{-1}\left(\overline{\boldsymbol{x}_r} - \overline{\boldsymbol{x}}\right) + n\left(\overline{\boldsymbol{x}} - \boldsymbol{\mu}\right)'\Sigma^{-1}\left(\overline{\boldsymbol{x}} - \boldsymbol{\mu}\right)$$

olarak yazılabilir. Bu nedenle,

$$l\left(\boldsymbol{\mu}, \Sigma\right) = iz\left(\Sigma^{-1}\,\mathrm{A}\right) + n\boldsymbol{d}'\Sigma^{-1}\boldsymbol{d}$$
$$= n\left\{iz\left(\Sigma^{-1}\boldsymbol{S}\right) + \boldsymbol{d}'\,\Sigma^{-1}\boldsymbol{d}\right\} \qquad \text{(xi)}$$

dir. Burada notasyonun kolaylığı için

$$\boldsymbol{A} = n\boldsymbol{S}$$
$$\boldsymbol{d} = \overline{\boldsymbol{x}} - \boldsymbol{\mu} \qquad \text{(xii)}$$

tanımını yaparız ve $\boldsymbol{S}$ (n ile bölenli) örneklem kovaryans matrisidir. $\boldsymbol{C}$,$n \times p$ merkezileştirilmiş veri matrisi

$$\boldsymbol{C}' = \left(\boldsymbol{x}_1 - \overline{\boldsymbol{x}}, \ldots, \boldsymbol{x}_n - \overline{\boldsymbol{x}}\right)$$

olmak üzere $n\boldsymbol{S} = \boldsymbol{C}'\boldsymbol{C}$ yi kulandık.

$$\sum_{r=1}^{n}\left(\boldsymbol{x}_r-\overline{\boldsymbol{x}}\right)'\Sigma^{-1}\left(\boldsymbol{x}_r-\overline{\boldsymbol{x}}\right)=iz\left(\boldsymbol{C}\,\Sigma^{-1/2}\,\boldsymbol{C}'\right)$$

$$=iz\left(\Sigma^{-1}\,\boldsymbol{C}'\boldsymbol{C}\right)$$

$$=iz\left(\Sigma^{-1}\,\boldsymbol{A}\right)$$

$$=(n)\,iz\left(\Sigma^{-1}\,\boldsymbol{S}\right)$$

olduğunu görürüz. $l=l(\boldsymbol{\mu},\Sigma)$ ve $\boldsymbol{\mu}$ üzerindeki bağımlılığın kesin olarak $\boldsymbol{d}$ ye dayandığına dikkat edelim. Şimdi Σ nın pozitif tanımlı olduğunu kabul edelim. Bu nedenle Σ^{-1} de pozitif tanımlıdır. Böylece $\forall\boldsymbol{d}\neq 0$ için $\boldsymbol{d}'\Sigma^{-1}\boldsymbol{d}>0$ elde etmemiz, $\boldsymbol{d}=0$ olduğundan sabit Σ için l nin $\boldsymbol{\mu}$ ye göre minimumlaştığını gösterir. Bu nedenle $\boldsymbol{\mu}=\overline{\boldsymbol{x}}$ dir.

$l=l\left(\boldsymbol{\mu},\Sigma\right)$ fonksiyonunun log -olabilirliğini Σ ya göre minimumlaştırmak için, keyfi bir ek bir sabite bağlı log -olabilirlik

$$l\left(\overline{\boldsymbol{x}},\Sigma\right)=nlog\left|\Sigma\right|+iz\left(\Sigma^{-1}\,\boldsymbol{A}\right)$$

$$=n\left(log\left|\Sigma\right|+iz\left(\Sigma^{-1}\,\boldsymbol{S}\right)\right)$$

dır.

$$\boldsymbol{\Phi}(\Sigma)=n\left\{log\left|\Sigma\right|+iz\left(\Sigma^{-1}\,\boldsymbol{S}\right)\right\} \qquad \text{(xiii)}$$

olsun.

$$\boldsymbol{\Phi}(\Sigma)-\boldsymbol{\Phi}(\boldsymbol{S})=n\left\{log\left|\Sigma\right|-log\left|\boldsymbol{S}\right|+iz\left(\Sigma^{-1}\,\boldsymbol{S}\right)-p\right\}$$

$$=n\left\{+iz\left(\Sigma^{-1}\,\boldsymbol{S}\right)-log\left|\Sigma^{-1}\,\boldsymbol{S}\right|-p\right\}$$

$$\geq 0$$

olduğunu gösteriniz.

Lemma 1: $\Sigma^{-1}\,\boldsymbol{S}$ pozitif tanımlıdır (İki pozitif tanımlı matrisin çarpımı pozitif tanımlı olduğundan).

Lemma 2: Pozitif sayıların herhangi sayıları için $A \geq logG+1$ dir. Burada A ve G sırası ile aritmetik ve geometrik ortalamadır.

İspat: Her $\boldsymbol{x}$ için $e^x \geq 1+\boldsymbol{x}$ elde ederiz. Bu nedenle, bir $i \in \{1,2,...,n\}$ kümesinin, her bir $y_i \geq 0$ sayısı için

$$y_i \geq 1+logy_i$$

$$\sum y_i \geq n+\sum logy_i$$

$$\boldsymbol{A} \geq 1+log\left(\prod y_i\right)^{\frac{1}{n}}$$

$$=1+logG \qquad \text{(xiv)}$$

elde ederiz. (xiv) bağıntısında $\Sigma^{-1}\boldsymbol{S}$ nin öz değerlerinin pozitif olduklarını kabul ederek herhangi bir $\boldsymbol{A}$ kare matrisini için, $iz(\boldsymbol{A})=\sum \lambda_i$,yani öz değerlerin toplamı ve $\boldsymbol{A}=\prod \lambda_i$ yani öz değerlerin çarpımı olduğunu hatırlayalım.

$\lambda_i\,(i=1,2,...,p)$, $\Sigma^{-1}\boldsymbol{S}$ nin öz değerleri olsun ve (xiv) de aşağıdaki ifadede yerine koyalım.

$$log\left|\Sigma^{-1}\boldsymbol{S}\right| = log\left(\prod \lambda_i\right)$$

$$=\underline{(p)logG}$$

ve

$$iz\left(\Sigma^{-1}\boldsymbol{S}\right)=\sum_{\underline{pA}} \lambda_i$$

dır. Burada , $\boldsymbol{\lambda_i}$ lerin aritmetik ortalamasıdır. Böylece,

$$\boldsymbol{\Phi}(\Sigma)-\boldsymbol{\Phi}(\boldsymbol{S})=np\{\boldsymbol{A}-logG\text{ - }1\} \geq 0$$

elde edilir.

EK 2.4 $\overline{X}$ ve s nin Örneklem Dağılımı

Wishart Dağılımı (Tanım)

$\boldsymbol{X}$ $m\times p$ $N_p(\mathbf{0},\Sigma)$ da bir veri matrisi olmak üzere, $\boldsymbol{M}$, $\boldsymbol{M}=\boldsymbol{X}'\boldsymbol{X}$ olarak yazılabilirse, bu takdirde $\boldsymbol{M}$ ye Σ ölçek matrisli ve m serbestlik dereceli bir Wishart dağılımına sahiptir denir. Bunu

$$\boldsymbol{M} \sim W_{\boldsymbol{p}}(\Sigma,m) \tag{xv}$$

şeklinde yazarız. $\Sigma=\boldsymbol{I}_{\boldsymbol{p}}$ olduğunda dağılımın standart formda olduğunu söyleriz.

Not: Wishart dağılımı χ^2 dağılımının çok değişkenli genelleştirmesidir.

EK 2.5 BİR WİSHART DAĞILIMINA SAHİP MATRİSLERİN BAŞKA BİR ÖZELLİĞİ

$\boldsymbol{M}_1,\boldsymbol{M}_2$ bağımsız olarak $\boldsymbol{M}_1\sim W_{\boldsymbol{p}}(\Sigma,m_1)$ ve $\boldsymbol{M}_2 \sim W_{\boldsymbol{p}}(\Sigma,m_2)$ Wishart dağılımına sahip olan matrisler olsun. Bu takdirde

$$\boldsymbol{M}_1+\boldsymbol{M}_2\sim W_{\boldsymbol{p}}(\Sigma,m_1+m_2)$$

dir. Eğer ,

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{bmatrix}$$

m_1+m_2 satırdan ibaret birleştirilmiş bir veri matrisi ise bu takdirde

$$\boldsymbol{X}'\boldsymbol{X}=\boldsymbol{X}_1'\boldsymbol{X}_1+\boldsymbol{X}_2'\boldsymbol{X}_2$$

$\boldsymbol{X}$ birleştirilmiş veri matrisinden oluşturulan veri matrisi (“Gram matris” olarak bilinir) olması anlamında veri matrisleri toplanabilir olduğundan bu özellik Wishart dağılımının tanımından görülür.

$p=1$ **durumu:**

$p=1$ olduğunda χ_r^2 nin dağılımının r sayıda bağımsız $N(0,1)$ değişkeninin karelerinin toplamı olarak dağılımından

$$\boldsymbol{M} = \sum_{\boldsymbol{i}=1}^{m} x_i^2 \sim \sigma^2 \chi_m^2$$

olduğunu bu nedenle de

$$W_1\left(\sigma^2, m\right) \equiv \sigma^2 \chi_m^2$$

olduğunu biliyoruz.

EK 2.5 Örneklem Dağılımları

$\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n$ $N_p(\boldsymbol{\mu}, \Sigma)$ dan n-boyutlu bir rastgele örneklem olsun. Bu takdirde

1. $\overline{\boldsymbol{x}}$ örneklem ortalaması

$$\overline{\boldsymbol{x}} \sim N_p\left(\boldsymbol{\mu}, \frac{1}{n}\Sigma\right)$$

normal dağılımına sahiptir.

2. $\boldsymbol{S}$ örneklem kovaryans matrisi (EÇOT: $\boldsymbol{S} = \frac{1}{n}\boldsymbol{C}'\boldsymbol{C}$)

$$n\boldsymbol{S} \sim W_p(\Sigma, n-1)$$

Wishart dağılımına sahiptir.

3. $\overline{\boldsymbol{x}}$ ve $\boldsymbol{S}$ nin dağılımları bağımsızdırlar.

EK 2.7 Verilen Bir Vektör İle Orantılı $\boldsymbol{\mu}$

Bazen, $\boldsymbol{\mu}$ nün, $\boldsymbol{\mu} = k\mu_0$ olacak şekilde, verilen bir vektöre orantılı olduğu bilinir. Örneğin eğer $\boldsymbol{x}$ tekrarlı ölçümlerin bir örneklemini gösterirse bu takdirde $\boldsymbol{j} = (1,1,...1)'$ 1 lerin p-vektörü olmak üzere $\boldsymbol{\mu} = k\boldsymbol{j}$ dir. Bu durum için k en çok olabilirlik tahminini buluruz. Σ nın bilindiği ve $\boldsymbol{\mu} = k\boldsymbol{\mu}_0$ olduğunu farz edelim. log-olabilirlik

$$\begin{aligned} l &= -2\log L \\ &= n\left\{\log|\Sigma| + iz\left(\Sigma^{-1}\boldsymbol{S}\right) + \left(\overline{x} - k\mu_0\right)'\Sigma^{-1}\left(\overline{x} - k\mu_0\right)\right\} \end{aligned}$$

dır. l yi k ya göre minimumlaştırmak için $\frac{\partial l}{\partial k} = 0$ koyarız, yani

$$\bar{\boldsymbol{x}}' \Sigma^{-1} \bar{\boldsymbol{x}} - 2k\boldsymbol{\mu}_0' \Sigma^{-1} \bar{\boldsymbol{x}} + k^2 \boldsymbol{\mu}_0' \Sigma^{-1} \boldsymbol{\mu}_0 = 0$$

koyarız ve buradan

$$k = \frac{\boldsymbol{\mu}_0' \Sigma^{-1} \bar{\boldsymbol{x}}}{\boldsymbol{\mu}_0' \Sigma^{-1} \boldsymbol{\mu}_0} \tag{xvi}$$

elde ederiz. k nın k nın bir yansız tahmin edicisi olduğunu gösterebiliriz ve k nın varyansını belirleyebiliriz. (xvi) bağıntısında $\boldsymbol{c}' = \boldsymbol{\mu}_0' \Sigma^{-1}$ ve $\alpha = \boldsymbol{\mu}_0' \Sigma^{-1} \boldsymbol{\mu}_0$ olmak üzere k ‘yı $\frac{1}{\alpha}\boldsymbol{c}'\bar{\boldsymbol{x}}$ formunu alır. Bu nedenle,

$$\begin{aligned} E\left[k\right] &= \frac{1}{\alpha}\boldsymbol{c}' E\left[\bar{\boldsymbol{x}}\right] \\ &= \frac{k}{\alpha}\boldsymbol{c}'\boldsymbol{\mu}_0 \end{aligned} \tag{xvii}$$

olur. Böylece, $E\left(k\right) = k$ dır. Bu ise da k nın bir yansız tahmin edicisi olduğunu gösterir. $Var\left[\bar{\boldsymbol{x}}\right] = \frac{1}{n}\Sigma$ ve bu nedenle $Var\left[\boldsymbol{c}'\bar{\boldsymbol{x}}\right] = \frac{1}{n}\boldsymbol{c}'\Sigma\boldsymbol{c}$ olduğuna dikkat edelim. Böylece,

$$\begin{aligned} Var\left(k\right) &= \frac{1}{n\alpha^2}\boldsymbol{c}'\Sigma\boldsymbol{c} \\ &= \frac{1}{n\boldsymbol{\mu}_0' \Sigma^{-1} \boldsymbol{\mu}_0} \end{aligned}$$

olduğunu görürüz.

EK 2.8 $\boldsymbol{\mu}$ Üzerinde Lineer Kısıtlama

A $(m \times p)$ olmak üzere $\boldsymbol{\mu}$ için $\boldsymbol{A\mu} = \boldsymbol{b}$ lineer kısıtlamasını sağlayan bir tahmin ediciyi belirleyeceğiz. m sayıda Lagrange çarpanının bir $\boldsymbol{\lambda}$ vektörünü ortaya koyalım.

$$l+2\boldsymbol{\lambda}'\left(\boldsymbol{A\mu}-\boldsymbol{b}\right)=n\left\{\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}\right)'\Sigma^{-1}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}\right)+2\boldsymbol{\lambda}'\left(\boldsymbol{A\mu}-\boldsymbol{b}\right)\right\}$$

minimumlaştırmaya çalışalım. $\boldsymbol{\mu}$ ‘ye göre diferansiyel alarak

$$-2\Sigma^{-1}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}\right)+2\boldsymbol{A}'\boldsymbol{\lambda}=0$$

$$\overline{\boldsymbol{x}}-\boldsymbol{\mu}=\Sigma \mathrm{A}'\boldsymbol{\lambda} \tag{xviii}$$

elde ederiz. Lagrange çarpanlarının değerini belirlemek için $\boldsymbol{A\mu}=\boldsymbol{b}$ kısıtlamasını kullanırız. $\boldsymbol{A}$ ile önden çarparak

$$\boldsymbol{A}\overline{\boldsymbol{x}}-\boldsymbol{b}=\boldsymbol{A}\Sigma \mathrm{A}'\boldsymbol{\lambda}$$

$$\boldsymbol{\lambda}=\left(\boldsymbol{A}\Sigma \mathrm{A}'\right)^{-1}\left(\boldsymbol{A}\overline{\boldsymbol{x}}-\boldsymbol{b}\right)$$

elde ederiz. Bunu (xviii) de yerine koyarak

$$\boldsymbol{\mu}=\overline{\boldsymbol{x}}-\Sigma \mathrm{A}'\left(\boldsymbol{A}\Sigma \mathrm{A}'\right)^{-1}\left(\boldsymbol{A}\overline{\boldsymbol{x}}-\boldsymbol{b}\right)$$

elde ederiz.

EK 2.9 Verilen Bir Matrise Orantılı Σ Kovaryans Matrisi

$\Sigma = k\Sigma_0$ olduğunda ve Σ_0 verildiğinde k yı tahmin etmeyi düşünelim. Bu,

$$l=n\left\{log\left|k\Sigma_0\right|+iz\left(\frac{1}{k}\Sigma_0^{-1}\boldsymbol{S}\right)\right\}$$

artı k yı içermeyen terimler aşağıdaki biçimi alır. Böylece

$$k=\frac{iz\left(\Sigma_0^{-1}\boldsymbol{S}\right)}{p}$$

yazılabilir.

ÖZGEÇMİŞ

Adı Soyadı : Esra DEMÜR

Doğum Yeri : Ordu

Doğum Tarihi : 18.09.1990

Yabancı Dili : İngilizce

E-mail : esra.demur.5225@windowslive.com

İletişim Bilgileri : Ordu Üniversitesi Fen Edebiyat Fakültesi

Öğrenim Durumu :

Derece	Bölüm/ Program	Üniversite	Yıl
Lisans	Matematik	Atatürk Üniversitesi	2012
Y. Lisans	Matematik	Ordu Üniversitesi	2014

İş Deneyimi:

Görev	Görev Yeri	Yıl
Öğretmen	Özel Ordu Beta-Açı dershaneleri	2012