Karadeniz Teknik Üniversitesi

Fen Fakültesi
İstatistik ve Bilgisayar Bilimleri Bölümü

Çok Değişkenli İstatistiksel Yöntemler

Çok Değişkenli Normal Dağılım

Dr. Öğr. Üyesi Uğur ŞEVİK

Ders İçeriği

- BÖLÜM 5 ÇOK DEĞİŞKENLİ NORMAL DAĞILIM

Ders Hedefleri

- Çok değişkenli normal dağılım ve örnekleri verilecektir.

BÖLÜM 5 ÇOK DEĞİŞKENLİ NORMAL DAĞILIM

- Çok değişkenli istatistiksel yöntemlerin çoğu, X veri matrisinin çok değişkenli normal dağıldığı varsayımı temeline dayandırılmıştır.
- Çok değişkenli normal dağılım, tek değişkenli normal dağılımın, değişken sayısı (p)> 2 için genelleştirilmiş durumudur.
- Bilindiği gibi, tek değişkenli normal dağılım:

x: İncelenen değişken

μ: İncelenen değişkenin ortalaması

σ: İncelenen değişkenin standart sapması

olmak üzere,

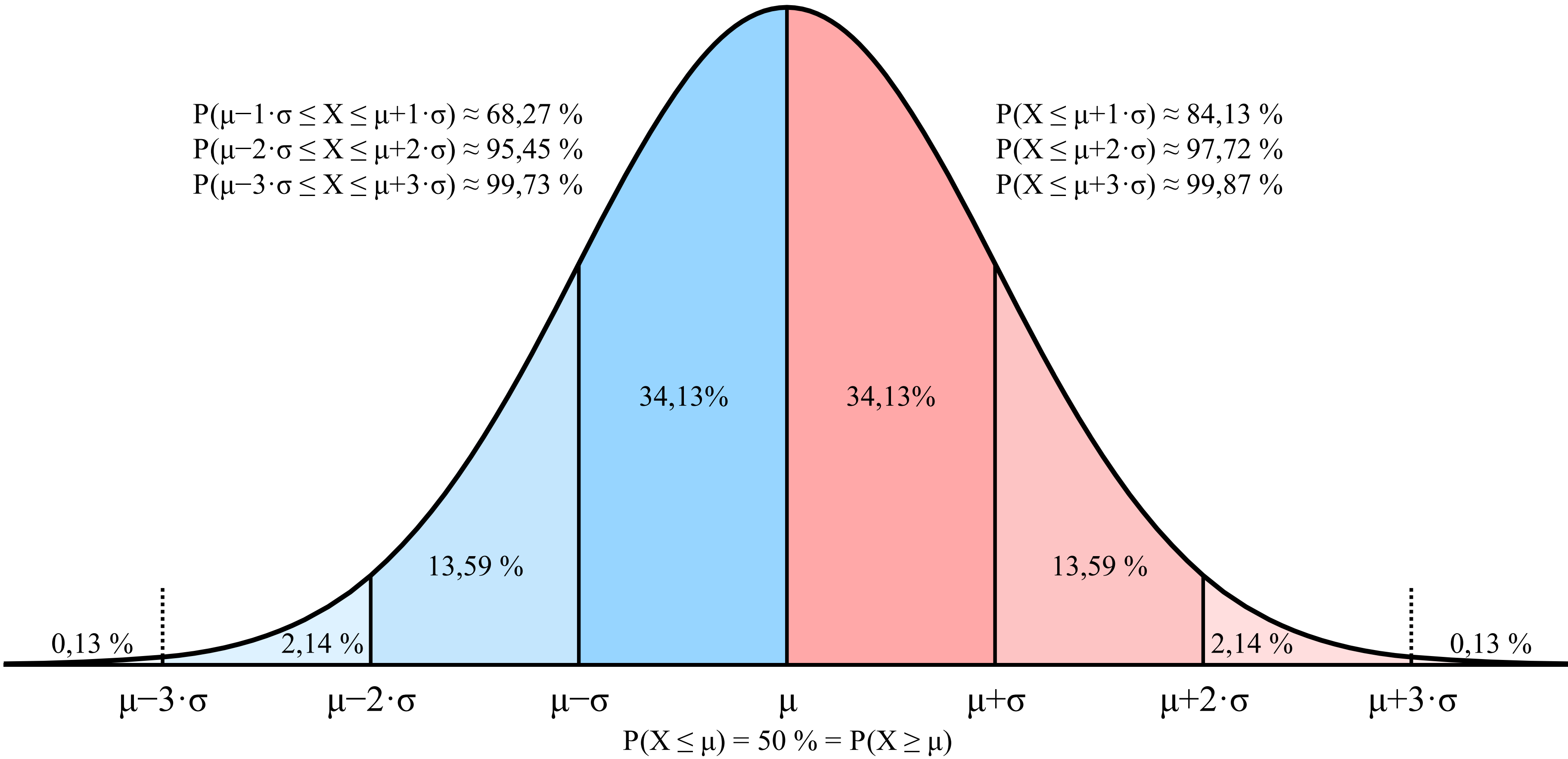
$$f(x) = \frac{1}{\sqrt{2\pi\,\sigma^2}} e^{(-1/2)\,[(x-\mu)/\sigma]^2} \qquad -\infty < x < +\infty$$

BÖLÜM 5 ÇOK DEĞİŞKENLİ NORMAL DAĞILIM

$$f(x) = \frac{1}{\sqrt{2\pi\,\sigma^2}} e^{(-1/2)\,[(x-\mu)/\sigma]^2} \qquad -\infty < x < +\infty$$

- ile verilen bir olasılık yoğunluk fonksiyonuna sahiptir.
- Bu fonksiyonda: ortalama (μ) ile ±1 standart sapma (±1σ) aralığında gözlemlerin % 68,26’sı, ±2 standart sapma (±2σ) aralığında gözlemlerin %95,44’ü ve ±3 standart sapma (±3σ) aralığında gözlemlerin yaklaşık %100’ü bulunur ve aşağıdaki gibi yazılır:
- $p(\mu - 1\sigma \leq x \leq \mu + 1\sigma) = 0{,}6826$
- $p(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = 0{,}9544$
- $p(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = 0{,}9974$
- Eğer bir x dağılımı μ ve σ^2 ile normal dağılım gösteriyor ise bu dağılımın normal dağıldığına ilişkin gösterim $x \sim N(\mu, \sigma^2)$ şeklinde olur.
- Örneğin, bir x dağılımı $\mu = 25$ ve $\sigma^2 = 4$ ile normal dağılım gösteriyor ise $x \sim N(25, 4)$ yazılır.

BÖLÜM 5 ÇOK DEĞİŞKENLİ NORMAL DAĞILIM



BÖLÜM 5 ÇOK DEĞİŞKENLİ NORMAL DAĞILIM

- Normal dağılım olasılık yoğunluk fonksiyonunda $[(x-\mu)/\sigma]^2$ terimi, x değişkenine ilişkin gözlemlerin ortalamaya olan uzaklığının standart sapma ile standartlaştırılmış halinin yani uzaklığın karesidir ve

$$\frac{(x-\mu)^2}{\sigma^2} = (x-\mu)\ \left(\frac{1}{\sigma^2}\right)\ (x-\mu) = (x-\mu)\ (\sigma^2)^{-1}(x-\mu)$$

- şeklinde ifade edilebilir. Bu yaklaşım p x 1 boyutlu x gözlem vektörü için genelleştirilirse;

Σ : $p \times p$ boyutlu ve p ranklı kovaryans matrisi

x : $p \times 1$ boyutlu gözlem vektörü

μ : $p \times 1$ boyutlu ortalama vektörü

olmak üzere;

$$(x-\mu)'\ \Sigma^{-1}(x-\mu)$$

şeklinde yazılabilir.

BÖLÜM 5 ÇOK DEĞİŞKENLİ NORMAL DAĞILIM

- Bu ifade kısaca, “x gözlem vektöründen p ortalama vektörüne olan genelleştirilmiş kare uzaklık” şeklinde tanımlanır.
- Yani $(x-\mu)'\,\Sigma^{-1}(x-\mu)$ Mahalanobis uzaklığı denir.
- Çok değişkenli normal yoğunluk fonksiyonunu elde ederken, tek değişkenli uzaklık

$$\frac{(x-\mu)^2}{\sigma^2} = (x-\mu)\ \left(\frac{1}{\sigma^2}\right)\ (x-\mu) = (x-\mu)\ (\sigma^2)^{-1}(x-\mu)$$

- Çok değişkenli genelleştirilmiş uzaklık $(x-\mu)'\,\Sigma^{-1}(x-\mu)$ ile yer değiştirirken

$$\sqrt{2\pi\,\sigma^2} = (2\pi)^{1/2}\ (\sigma^2)^{1/2},\ (2\pi)^{p/2}\,|\Sigma|^{1/2}$$

- ile yer değiştirir ve bir x gözlem vektörü için p boyutlu normal dağılım fonksiyonu;
- $$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-(x-\mu)'\Sigma^{-1}(x-\mu)/2}$$ ile verilir.

BÖLÜM 5 ÇOK DEĞİŞKENLİ NORMAL DAĞILIM

- Burada $-\infty \leq x_i \leq +\infty$ ve i=1,2,..,p olup, p boyutlu normal yoğunluk $N_p(\mu,\Sigma)$ ile gösterilir.
- Örneklemler için μ yerine $\bar{x}$, Σ yerine S yazılır.
- **Özet olarak çok değişkenli normal dağılım, ortalama vektörü ve varyans-kovaryans matrisi ile tanımlanır.**

Örnek: $x' = [x_1, x_2]$, $\mu' = [1, 3]$, $\Sigma = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$

$X \sim N(\mu, \Sigma)$ ise $f_X(x) = ?$

Formülü kullanırsak;

$$|\Sigma| = \begin{vmatrix} 2 & -3 \\ -3 & 5 \end{vmatrix} = 1, \quad \Sigma^{-1} = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}$$

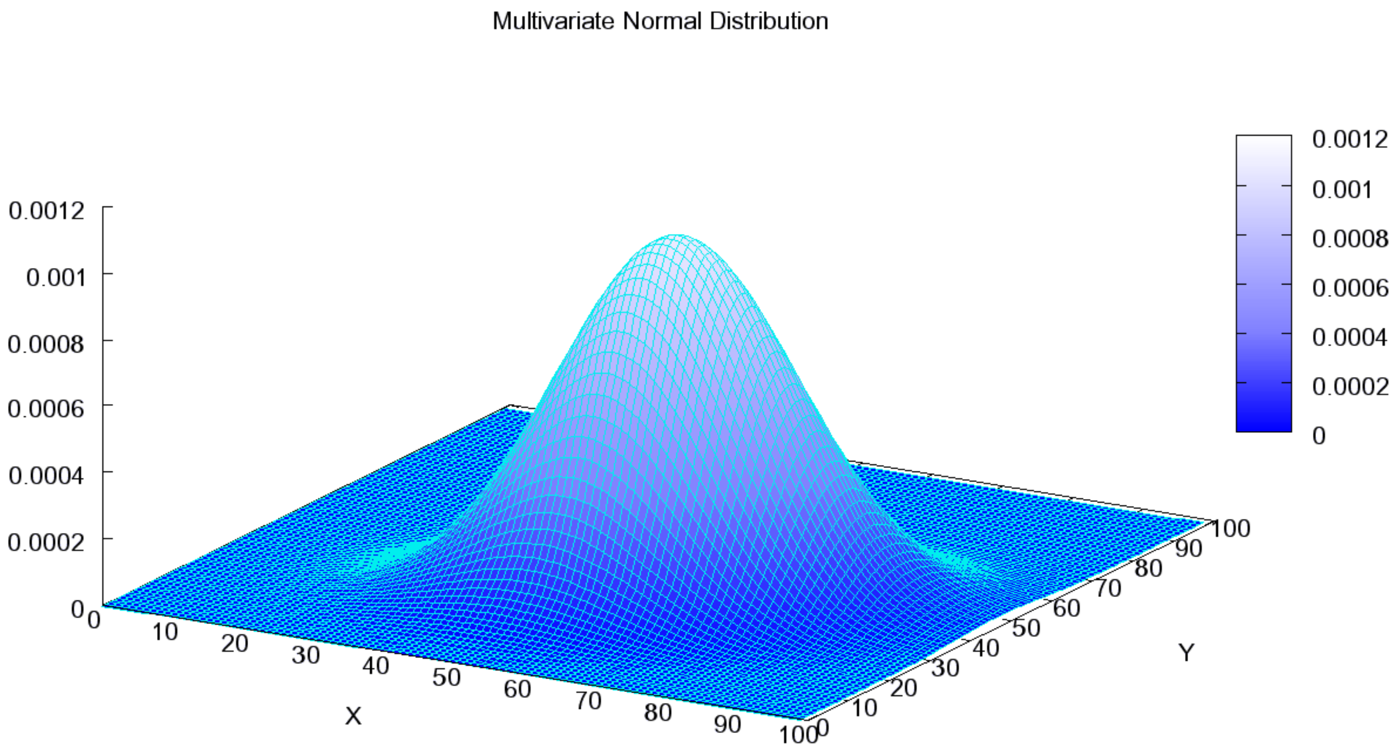
$$\exp = \left\{ -\frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{2} \right\} = -\frac{1}{2} \begin{bmatrix} x_1 - 1 \\ x_2 - 3 \end{bmatrix}' \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 - 3 \end{bmatrix}$$

$$= -\frac{1}{2}(5x_1^2 + 2x_2^2 + 6x_1x_2 + 28x_1 + 18x_2 + 41)$$

$$\Rightarrow f_X(x) = \frac{1}{2\pi} \exp\left\{ -\frac{1}{2}(5x_1^2 + 2x_2^2 + 6x_1x_2 + 28x_1 + 18x_2 + 41) \right.$$

BÖLÜM 5 ÇOK DEĞİŞKENLİ NORMAL DAĞILIM

- İki değişkenli normal dağılım (Bivariate normal joint density)



ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- Sunum hazırlanırken aşağıdaki kaynaktan yararlanılmıştır.