다변량 확률변수의 성질

서울시립대 통계학과 이용희

FALL 2019

1 일변량분포

일변량 확률변수 X가 확률밀도함수 $f(x)$를 가지는 분포를 따를때 기대값과 분산은 다음과 같이 정의된다.

$$E(X) = \int x(x)dx = \mu, \quad V(X) = E[X - E(X)]^2 = \int (x-\mu)^2 f(x)dx = \sigma^2$$

새로운 확률변수 Y가 확률변수 X의 선형변환으로 표시된다면 (a와 b는 실수)

$$Y = aX + b$$

그 기대값(평균)과 분산은 다음과 같이 계산된다.

$$\begin{aligned}
E(Y) &= E(aX+b) \\
&= \int (ax+b)f(x)dx \\
&= a\int xf(x)dx + b \\
&= aE(X) + b \\
&= a\mu + b \\
V(Y) &= Var(aX+b) \\
&= E[aX+b-E(aX+b)]^2 \\
&= E[a(X-\mu)]^2 \\
&= a^2E(X-\mu)^2 \\
&= a^2\sigma^2
\end{aligned}$$

2 확률벡터와 분포

확률벡터 $\boldsymbol{X}$가 p 차원의 다변량분포를 따른다고 하고 결합확률밀도함수 $f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_p)$를 를 가진다고 하자.

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ .. \\ X_p \end{pmatrix}$$

다변량 확률벡터의 기대값(평균벡터)과 공분산(행렬)은 다음과 같이 계산된다.

$$\boldsymbol{E}(\boldsymbol{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ E(X_3) \\ .. \\ E(X_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ .. \\ \mu_p \end{pmatrix} = \boldsymbol{\mu}$$

$$V(\boldsymbol{X}) = Cov(\boldsymbol{X}) = E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^t = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \ldots & \sigma_{2p} \\ & \ldots & \ldots & \\ \sigma_{1p} & \sigma_{2p} & \ldots & \sigma_{pp} \end{pmatrix} = \boldsymbol{\Sigma}$$

여기서 $\sigma_{ii} = V(X_i)$, $\sigma_{ij} = Cov(X_i, X_j) = Cov(X_j, X_i)$이다. 따라서 공분산 행렬 $\boldsymbol{\Sigma}$는 대칭행렬(symmetric matrix)이다. 다음 공식은 유용한 공식이다.

$$\boldsymbol{\Sigma} = E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^t = E(\boldsymbol{X}\boldsymbol{X}^t) - \boldsymbol{\mu}\boldsymbol{\mu}^t$$

두 확률변수의 상관계수 ρ_{ij}는 다음과 같이 정의된다.

$$\rho_{ij} = \frac{Cov(X_i, X_j)}{\sqrt{V(X_i)V(X_j)}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

새로운 확률벡터 $\boldsymbol{Y}$가 확률벡터 $\boldsymbol{X}$ 의 선형변환라고 하자.

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$$

단 여기서 $\boldsymbol{A} = \{a_{ij}\}$는 $p \times p$ 실수 행렬이고 $\boldsymbol{b} = (b_1 b_2 \ldots b_p)^t$는 $p \times 1$ 실수 벡터이다.

확률벡터 $\boldsymbol{Y}$의 기대값(평균벡터)과 공분산은 다음과 같이 계산된다.

$$\begin{aligned} E(\boldsymbol{Y}) &= E(\boldsymbol{AX}+\boldsymbol{b}) \\ &= \boldsymbol{A}E(\boldsymbol{X})+\boldsymbol{b} \\ &= \boldsymbol{A\mu}+\boldsymbol{b} \\ V(\boldsymbol{Y}) &= Var(\boldsymbol{AX}+\boldsymbol{b}) \\ &= E[\boldsymbol{AX}+\boldsymbol{b}-E(\boldsymbol{AX}+\boldsymbol{b})][\boldsymbol{AX}+\boldsymbol{b}-E(\boldsymbol{AX}+\boldsymbol{b})]^t \\ &= E[\boldsymbol{AX}-\boldsymbol{A\mu}][\boldsymbol{AX}-\boldsymbol{A\mu}]^t \\ &= E[\boldsymbol{A}(\boldsymbol{X}-\boldsymbol{\mu})][\boldsymbol{A}(\boldsymbol{X}-\boldsymbol{\mu})]^t \\ &= \boldsymbol{A}E[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^t]\boldsymbol{A}^t \\ &= \boldsymbol{A\Sigma A}^t \end{aligned}$$

만약 표본 $\boldsymbol{X}_i, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ 이 독립적으로 평균이 $\boldsymbol{\mu}$ 이고 공분산이 $\boldsymbol{\Sigma}$ 인 분포에서 추출되었다면 표본의 평균벡터 $\bar{\boldsymbol{X}}$ 는 평균이 $\boldsymbol{\mu}$ 이고 공분산이 $\frac{1}{n}\boldsymbol{\Sigma}$ 인 분포를 따른다.

$$\bar{\boldsymbol{X}} = \begin{pmatrix} \sum_{i=1}^n X_{i1}/n \\ \sum_{i=1}^n X_{i2}/n \\ \sum_{i=1}^n X_{i3}/n \\ .. \\ \sum_{i=1}^n X_{ip}/n \end{pmatrix}$$

여기서 X_{ij} 는 i번째 표본벡터 $\boldsymbol{X}_i = (X_{i1}X_{i2}\ldots X_{ip})^t$의 j번째 확률변수이다.

3 다변량 정규분포

일변량 확률변수 X가 평균이 μ 이고 분산이 σ^2인 정규분포를 따른다면 다음과 같이 나타내고

$$X \sim N(\mu, \sigma^2)$$

확률밀도함수 $f(x)$ 는 다음과 같이 주어진다.

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

p-차원 확률벡터 $\boldsymbol{X}$가 평균이 $\boldsymbol{\mu}$ 이고 공분산이 $\boldsymbol{\Sigma}$인 다변량 정규분포를 따른다면 다음과 같이 나타내고

$$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

확률밀도함수 $f(\boldsymbol{x})$ 는 다음과 같이 주어진다.

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

다변량 정규분포 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$를 따르는 확률벡터 $\boldsymbol{X}$를 다음과 같이 두 부분으로 나누면

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}, \quad \boldsymbol{X}_1 = \begin{bmatrix} \boldsymbol{X}_{11} \\ \boldsymbol{X}_{12} \\ \vdots \\ \boldsymbol{X}_{1p} \end{bmatrix}, \quad \boldsymbol{X}_2 = \begin{bmatrix} \boldsymbol{X}_{21} \\ \boldsymbol{X}_{22} \\ \vdots \\ \boldsymbol{X}_{2q} \end{bmatrix}$$

각각 다변량 정규분포를 따르고 다음과 같이 나타낼 수 있다.

$$\begin{bmatrix} E(\boldsymbol{X}_1) \\ E(\boldsymbol{X}_2) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \begin{bmatrix} V(\boldsymbol{X}_1) & Cov(\boldsymbol{X}_1, X_2) \\ Cov(\boldsymbol{X}_2 X_1) & V(\boldsymbol{X}_2) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^t & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \sim N_{p+q}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \Sigma_{12} \\ \boldsymbol{\Sigma}_{12}^t & \Sigma_{22} \end{bmatrix}\right)$$

확률벡터 $\boldsymbol{X}_2 = \boldsymbol{x}_2$가 주어진 경우 $\boldsymbol{X}_1$의 조건부 분포는 p-차원 다변량 정규분포를 따르고 평균과 공분산은 다음과 같다.

$$E(\boldsymbol{X}_1|\boldsymbol{X}_2 = \boldsymbol{x}_2) = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{x}_2), \quad V(\boldsymbol{X}_1|\boldsymbol{X}_2 = \boldsymbol{x}_2) = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^t$$

예를 들어 2-차원 확률벡터 $\boldsymbol{X} = (X_1, X_2)^t$가 평균이 $\boldsymbol{\mu} = (\mu_1, \mu_2)^t$ 이고 공분산 $\boldsymbol{\Sigma}$가 다음과 같이 주어진

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

이변량 정규분포를 따른다면 확률밀도함수 $f(\boldsymbol{x})$에서 exp함수의 인자는 다음과 같이 주어진다.

$$(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^t =$$
$$-\frac{1}{2(1-\rho^2)}\left[\left(\frac{(x_1-\mu_1)^2}{\sigma_{11}}\right) + \left(\frac{(x_2-\mu_2)^2}{\sigma_{22}}\right) - 2\rho\left(\frac{(x_1-\mu_1)}{\sqrt{\sigma_{11}}}\right)\left(\frac{(x_2-\mu_2)}{\sqrt{\sigma_{22}}}\right)\right]$$

그리고 $p = 2$인 경우 확률밀도함수의 상수부분은 다음과 같이 주어진다.

$$(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2} = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho^2)}}$$

여기서 $\rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$

만약 $X_2 = x_2$가 주어졌을 때 X_1의 조건부 분포는 정규분포이고 평균과 분산은 다음과 같이 주어진다.

$$E(X_1|X_2 = x_2) = \mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(\mu_2 - x_2) = \mu_1 + \rho\frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}(\mu_2 - x_2)$$

$$V(X_1|X_2 = x_2) = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} = \sigma_{11}(1 - \rho^2)$$

다변량 정규분포에서 공분산이 0인 두 확률 변수는 독립이다.

$$\sigma_{ij} = 0 \leftrightarrow X_i \text{ and } X_j \text{ are independent}$$

4 표준정규분포로의 변환

일변량 확률변수 X가 평균이 μ 이고 분산이 σ^2인 경우 다음과 같은 선형변환을 고려하면.

$$Z = \frac{X - \mu}{\sigma} = (\sigma^2)^{-1/2}(X - \mu)$$

확률변수 Z 는 평균이 0 이고 분산이 1인 분포를 따른다.

p차원 확률벡터 $\boldsymbol{X}$ 가 평균이 $\boldsymbol{\mu}$ 이고 공분산이 $\boldsymbol{\Sigma}$인 분포를 가진다고 가정하자. 공분산 행렬 $\boldsymbol{\Sigma}$는 양정치 행렬(positive definite matrix)이며 다음과 같은 행렬의 분해가 가능하다.

$$\Sigma = \boldsymbol{C}\boldsymbol{C}^t$$

여기서 $\boldsymbol{C}$ 는 정칙행렬이며 역행렬 $\boldsymbol{C}^{-1}$가 존재한다. 위와 같은 행렬의 분해는 스펙트럴 분해(spectral decomposition)을 이용하여 구할 수 있다. 공분산 행렬 $\boldsymbol{\Sigma}$는 양정치 행렬이므로 고유치(eigen value) $(\lambda_1, \lambda_2, \ldots, \lambda_p)$가 모두 양수이고 정규직교 고유벡터(orthonormal eigen vector)의 행렬 $\boldsymbol{P}$을 이용하여 다음과 같은 분해가 가능하다.

$$\Sigma = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^t = \boldsymbol{P}\boldsymbol{\Lambda}^{1/2}\Lambda^{1/2}\boldsymbol{P}^t$$

여기서 $\boldsymbol{\Lambda}$는 고유치 $(\lambda_1, \lambda_2, \ldots, \lambda_p)$를 대각원소로 가지는 대각행렬이며 $\boldsymbol{\Lambda}^{1/2}$는 고유치의 제곱근을 대각원소로 가지는 대각행렬이다. 따라서 $\boldsymbol{C} = \boldsymbol{P}\boldsymbol{\Lambda}^{1/2}$로 하면 위와 같은 행렬의 분해가 가능하다. 정규직교 고유벡터(orthonormal eigen vector)의 행렬 $\boldsymbol{P}$는 직교행렬이므로

$$\boldsymbol{C}^{-1} = (\boldsymbol{P}\boldsymbol{\Lambda}^{1/2})^{-1} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{P}^t$$

p차원 확률벡터 $\boldsymbol{X}$의 다음과 같은 선형변환을 고려하면.

$$\boldsymbol{Z} = \boldsymbol{C}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{P}^t(\boldsymbol{X} - \boldsymbol{\mu})$$

확률벡터 $\boldsymbol{Z}$ 는 평균이 $\boldsymbol{0}$ 이고 공분산이 $\boldsymbol{I}$인 분포를 따른다 (why?).

확률벡터 $\boldsymbol{X}$가 정규분포를 따른다면 선형변환한 확률벡터 $\boldsymbol{Z}$도 정규분포를 따른다.

5 예제

예를 들어 이변량확률벡터 $\boldsymbol{X}$가 다음과 같은 평균벡터와 공분산을 가진 정규분포를 따른다고 하자

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

공분산행렬 $\boldsymbol{\sigma}$의 고유치는 $|\boldsymbol{\sigma} - \lambda \boldsymbol{I}| = 0$의 방정식을 풀어 구할 수 있다.

$$|\boldsymbol{\sigma} - \lambda \boldsymbol{I}| = \begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = \lambda^2 - 4\lambda + 3 = 0$$

방정식을 풀면 고유치는 $(\lambda_1, \lambda_2) = (3, 1)$이다. 각 고유치에 대한 고유벡터 $\boldsymbol{p} = (p_1, p_2)^t$는 $\boldsymbol{\Sigma p} = \lambda \boldsymbol{p}$ 으로 구할 수 있다. 각 고유치에 대하여 방정식을 구하면 다음 두 개의 방정식을 얻을 수 있다.

$$p_1 - p_2 = 1 \text{ and } p_1 + p_2 = 0$$

정규직교 벡터의 조건을 만족 시키기 위해서 $p_1^2 + p_2^2 = 1$의 조건을 적용하면 다음과 같은 정규직교 고유행렬을 얻을 수 있다.

$$\boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

또한

$$\boldsymbol{\Lambda} = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \quad \boldsymbol{\Lambda}^{1/2} = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix}$$

따라서 $C^{-1} = \Lambda^{-1/2} \boldsymbol{P}^t$ 이며

$$\boldsymbol{C}^{-1} = \boldsymbol{\Lambda}^{-1/2} \boldsymbol{P}^t = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

위의 계산을 R 프로그램으로 다음과 같이 구현할 수 있다.

```r
mu <- c(1,2)
S <- matrix(c(2,1,1,2),2,2)
res<- eigen(S)
res

## eigen() decomposition
## $values
## [1] 3 1
##
## $vectors
```

```
##             [,1]       [,2]
## [1,] 0.7071068 -0.7071068
## [2,] 0.7071068  0.7071068
```

```
L <- res$values
P <- res$vectors
Lsqrt <- diag(sqrt(L))
C <- P %*% Lsqrt
C
```

```
##          [,1]       [,2]
## [1,] 1.224745 -0.7071068
## [2,] 1.224745  0.7071068
```

```
Cinv <- solve(C)
Cinv
```

```
##            [,1]      [,2]
## [1,]  0.4082483 0.4082483
## [2,] -0.7071068 0.7071068
```

```
Cinv %*% S %*% t(Cinv)
```

```
##      [,1] [,2]
## [1,]    1    0
## [2,]    0    1
```