Moments of the Multivariate Gaussian (1)

$$\mathbb{E}[\mathbf{x}] = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \int \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}\mathbf{x}\,\mathrm{d}\mathbf{x}$$

$$= \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \int \exp\left\{-\frac{1}{2}\mathbf{z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{z}\right\}(\mathbf{z}+\boldsymbol{\mu})\,\mathrm{d}\mathbf{z}$$

$=0$

$\mu \int e^{-\frac{1}{2} z^T \Sigma^{-1} z} \, dz_1 \cdots dz_p$

$z_1 \cdots z_p$

$= 1$

since a valid density.

thanks to symmetry of z

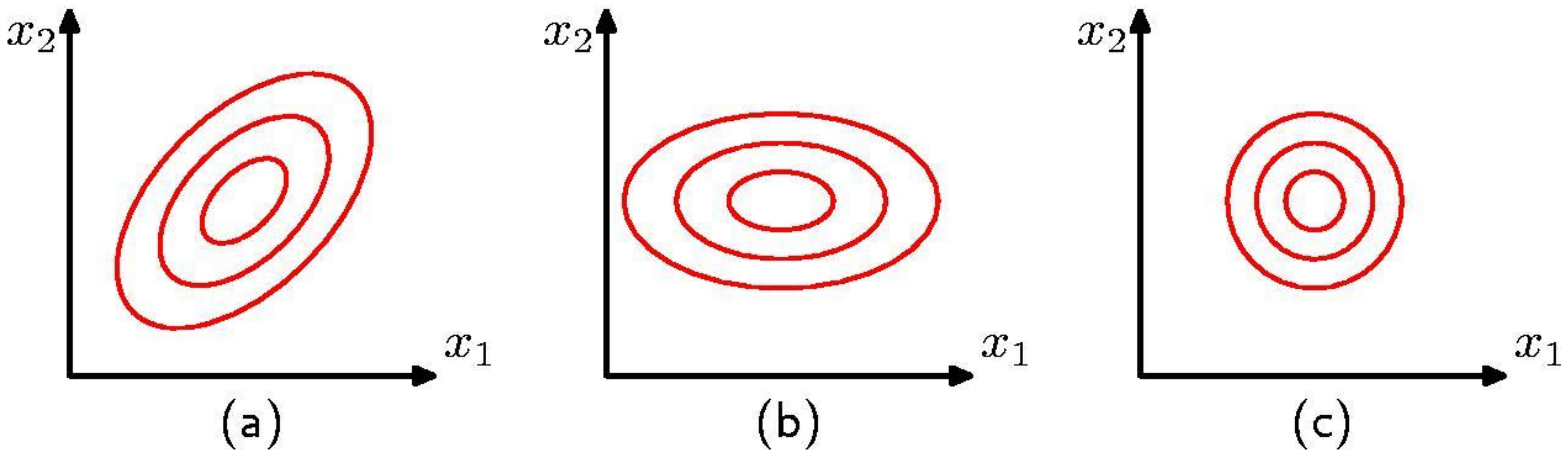
$$\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}$$

Moments of the Multivariate Gaussian (2)

$$\begin{bmatrix} \mathrm{cov}(x_i, x_j) \end{bmatrix} = \mathrm{Cov}(X)$$

$$\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathrm{T}}] = \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} + \boldsymbol{\Sigma}$$

$$\mathrm{cov}[\mathbf{x}] = \mathbb{E}\left[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^{\mathrm{T}}\right] = \boldsymbol{\Sigma}$$



Properties of Gaussian Distribution

- Let X be a p-dimensional random vector following a Normal distribution

$X \sim N(\mu_{p\times 1}, \Sigma_{p\times p})$ $\quad X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$

Linear combination of components of X are also normal.

- Let $c_1, c_2 \ldots c_p$ be arbitrary constants

$$Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p = \sum_{j=1}^{p} c_jX_j = \mathbf{c}'\mathbf{X}.$$

$c^T X$ $\quad c = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$

- Then Y also follows a normal distribution

$$Y \sim N(\mathbf{c}'\mu, \mathbf{c}'\Sigma\mathbf{c})$$

$Y \in R$

$\mu(Y) = c^T \mu$

$var(Y) = c^T \Sigma \mu$

[Reading material: Lesson 2: Linear Combinations of Random Variables | STAT 505 (psu.edu)]

Example 2-2: Monthly Employment Data

Another example where we might be interested in linear combinations is in the Monthly Employment Data. Here we have observations on 6 variables:

- X_1 Number people laid off or fired
- X_2 Number of people resigning
- X_3 Number of people retiring
- X_4 Number of jobs created
- X_5 Number of people hired
- X_6 Number of people entering the workforce

Net employment ~~decrease~~: increase

In looking at the net job increase, which is equal to the number of jobs created, minus the number of jobs lost.

$$Y = X_4 - X_1 - X_2 - X_3$$

In this case, we have the number of jobs created, (X_4), minus the number of people laid off or fired, (X_1), minus the number of people resigning, (X_2), minus the number of people retired, (X_3). These are all of the people that have left their jobs for whatever reason.

In this case

$$c_1 = c_2 = c_3 = -1 \text{ and } c_4 = 1$$

$c^T x = Y$

Mean and variance of Y

The population mean of a linear combination is equal to the same linear combination of the population means of the component variables. If

$$Y = c_1X_1 + c_2X_2 + \dots c_pX_p = \sum_{j=1}^p c_jX_j = \mathbf{c}'\mathbf{X}$$

then

$$E(Y) = c_1\mu_1 + c_2\mu_2 + \dots + c_p\mu_p = \sum_{j=1}^p c_j\mu_j = \mathbf{c}'\mu$$

$$var(Y) = \sum_{j=1}^p \sum_{k=1}^p c_jc_k\sigma_{jk} = c'\Sigma c$$

Applies for any set of p R.Vs. irrespective of whether they are jointly Gaussian

$f(x_1 \cdots x_p)$

$$Var(Y) = E[(Y - E(Y))^2] -$$

$$\int_y (y - E(y)) f(y)\, dy = \int_{x_1 \cdots x_p} \left(\sum_{j=1}^p c_j x_j - \sum_{i=1}^p c_j \mu_j \right)^2 f(x_1 \cdots x_p)\, dx_{1 \cdots p}$$

Proof

$$f(x_1 \cdots x_p) = f(x_j, x_k) f(\cdots | x_j, x_k)$$

$$\int_{x_1 \cdots x_p} \left(\sum_{j=1}^{p} c_j (x_j - \mu_j) \right)^2 f(x_1 \cdots x_p) \, dX_{1 \cdots p}$$

$$= \sum_{j=1}^{p} \sum_{k=1}^{p} \int_{x_1 \cdots x_p} c_j c_k \underbrace{(x_j - \mu_j)(x_k - \mu_k)}_{\int_{x_j, x_k} G(x_j, x_k) f(x_j, x_j)} \underbrace{f(x_1 \cdots x_p)}_{\int_{X - \{x_j, x_k\}} f(X - x_j - x_k | x_j, x_k) = 1} dX_{1 \cdots p}$$

$$= \sum_{j=1}^{p} \sum_{k=1}^{p} c_j c_k \int_{x_j, x_k} \underbrace{(x_j - \mu_j)(x_k - \mu_k) f(x_j, x_k) dx_j dx_k}_{\sigma_{j,k}}$$

$$\sigma_{j,k} \quad \mathrm{cov}(x_j, x_k)$$

$$\sum_{j=1}^{p} \sum_{k=1}^{p} c_j c_k \sigma_{k,j}$$

Example 2-1: Women's Health Survey (Linear Combinations)

The Women's Health Survey data contains observations for the following variables:

- X_1 calcium (mg)
- X_2 iron (mg)
- X_3 protein(g)
- X_4 vitamin A(μg)
- X_5 vitamin C(mg)

In addition to addressing questions about the individual nutritional component, we may wish to address questions about certain combinations of these components. For instance, we might want to ask what is the total intake of vitamins A and C (in mg). We note that in this case, Vitamin A is measured in micrograms while Vitamin C is measured in milligrams. There are a thousand micrograms per milligram so the total intake of the two vitamins, *Y*, can be expressed as the following:

$$Y = 0.001X_4 + X_5$$

Y is in milligrams.

In this case, our coefficients c_1 , c_2 and c_3 are all equal to 0 since the variables X_1, X_2 and X_3 do not appear in this expression. In addition, c_4 is equal to 0.001 since each microgram of vitamin A is equal to 0.001 milligrams of vitamin A. In summary, we have

$$c_1 = c_2 = c_3 = 0, c_4 = 0.001, c_5 = 1$$

1985, the USDA commissioned a study of women's nutrition. Nutrient intake was measured for a random sample of 737 women aged 25-50 years.

Example 2-3: Women's Health Survey (Population Mean)

The following table shows the sample means for each of the five nutritional components that we computed in the previous lesson.

Variable	Mean
Calcium	624.0 mg
Iron	11.1 mg
Protein	65.8 g
Vitamin A	839.6 μg
Vitamin C	78.9 mg

If, as previously, we define *Y* to be the total intake of vitamins A and C (in mg) or:

$$Y = 0.001X_4 + X_5$$

Then we can work out the estimated mean intake of the two vitamins as follows:

$$\bar{y} = 0.001\bar{x}_4 + \bar{x}_5 = 0.001 \times 839.6 + 78.9248 = 0.8396 + 78.9248 = 79.7680 \text{ mg.}$$

Example 2-4: Women's Health Survey (Population Variance)

Looking at the Women's Nutrition survey data we obtained the following variance/covariance matrix as shown below from the previous lesson.

$$S = \begin{pmatrix} 157829.4 & 940.1 & 6075.8 & 102411.1 & 6701.6 \\ 940.1 & 35.8 & 114.1 & 2383.2 & 137.7 \\ 6075.8 & 114.1 & 934.9 & 7330.1 & 477.2 \\ 102411.1 & 2383.2 & 7330.1 & 2668452.4 & 22063.3 \\ 6701.6 & 137.7 & 477.2 & 22063.3 & 5416.3 \end{pmatrix}$$

If we wanted to take a look at the total intake of vitamins A and C (in mg) remember we defined this earlier as:

$$Y = 0.001X_4 + X_5$$

Therefore the sample variance of Y is equal to $(0.001)^2$ times the variance for X_4, plus the variance for X_5, plus 2 times 0.001 times the covariance between X_4 and X_5. The next few lines carry out the mathematical calculations using these values.

$$\begin{aligned} s_Y^2 &= 0.001^2 s_4^2 + s_5^2 + 2 \times 0.001 s_{45} \\ &= 0.000001 \times 2668452.4 + 5416.3 + 0.002 \times 22063.3 \\ &= 2.7 + 5416.3 + 44.1 \\ &= 5463.1 \end{aligned}$$

More examples

- [4.1 - Comparing Distribution Types | STAT 505 (psu.edu)]