# Multivariate Normal Distribution

Two reasons for the practical use of the normal distributions:

1. a bona fide population model

2. sampling distribution (central-limit effect)

Univariate case

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

p-dimensional normal distribution:

$$f(x) = f(x_1, x_2, \ldots, x_p) = \frac{1}{(2\pi)^{p/2} \mid \Sigma \mid^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})'(\Sigma)^{-1}(\vec{x}-\vec{\mu})}$$

$$\approx N_p(\vec{\mu}, \Sigma)$$

$$\begin{array}{ll} \Sigma\text{: positive definite} & \Sigma^{-1}\text{: positive definite} \\ (\lambda_1, e_1), \ldots, (\lambda_p, e_p) & ({\lambda_1}^{-1}, e_1), \ldots, ({\lambda_p}^{-1}, e_p) \end{array}$$

**Result**: Constant-density contour of $f(x)$

ellipsoids: $(\vec{x} - \vec{\mu})'\Sigma^{-1}(\vec{x} - \vec{\mu}) = c^2$

center: $\vec{\mu}$

axes: $\pm c\sqrt{\lambda_1}e_1, \pm c\sqrt{\lambda_2}e_2, \ldots, \pm c\sqrt{\lambda_p}e_p$

**Properties:** Given that $\vec{X} \sim N_p(\vec{\mu}, \Sigma)$, and $\Sigma$ is positive definite.

1. $\vec{a}'\vec{X} = a_1X_1 + a_2X_2 + \cdots + a_pX_p \sim N(\vec{a}'\vec{\mu}, \vec{a}'\Sigma\vec{a})$

2.
$$A\vec{X} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{q1} & a_{q2} & \cdots & a_{qp} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} \sim N_q(A\vec{\mu}, A\Sigma A')$$

3.

$$\vec{X} = \begin{pmatrix} \vec{X}_1 \\ \vec{X}_2 \end{pmatrix}, \vec{\mu} = \begin{pmatrix} \vec{\mu}_1 \\ \vec{\mu}_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \Rightarrow, \vec{X}_1 \sim N_q(\vec{\mu}_1, \Sigma_{11})$$

4.

$$\vec{X} = \begin{pmatrix} \vec{X}_1 \\ \vec{X}_2 \end{pmatrix}, \vec{\mu} = \begin{pmatrix} \vec{\mu}_1 \\ \vec{\mu}_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

The conditional distribution of $\vec{X}_1$, given that $\vec{X}_2 = \vec{x}_2$, is normal with mean $\vec{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\vec{(x)}_2 - \vec{\mu}_2)$ and covariance matrix $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

5. $(\vec{X} - \vec{\mu})'\Sigma^{-1}(\vec{X} - \vec{\mu}) \sim \chi_p^2$ (Chi-square distribution with $p$ degrees of freedom)

$$P[(\vec{X} - \vec{\mu})'\Sigma^{-1}(\vec{X} - \vec{\mu}) \leq \chi_p^2(\alpha)] = 1 - \alpha$$

$\chi_p^2(\alpha)$ is the upper $100\alpha$ percentile of the $\chi_p^2$ distribution

**Estimation**

Suppose $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n$ is a random sample from $N_p(\vec{\mu}, \Sigma)$

Multivariate normal likelihood:

$$L(\vec{\mu}, \Sigma) = \left\{ \begin{array}{c} \text{joint density of} \\ \vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n \end{array} \right\} = \prod_{i=1}^{n} \{ \frac{1}{(2\pi)^{p/2} \mid \Sigma \mid^{1/2}} e^{-\frac{1}{2}(\vec{X}_i - \vec{\mu})'(\Sigma)^{-1}(\vec{X}_i - \vec{\mu})} \}$$

Maximum Likelihood Estimator and Estimate

$\hat{\vec{\mu}} = \bar{X}$ and $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\vec{X}_i - \bar{X})(\vec{X}_i - \bar{X})' = \frac{n-1}{n}S$

$\hat{\vec{\mu}} = \bar{x}$ and $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i - \bar{x})(\vec{x}_i - \bar{x})' = \frac{n-1}{n}S$

**Remark.** $\bar{X}$ and $S$ are sufficient statistics

**Sampling distributions**

1. $\bar{X} \sim N_p(\mu, \frac{1}{n}\Sigma)$
2. $(n-1)S \sim$ Wishart random matrix with $n-1$ degrees of freedom.

3. $\bar{X}$ and $S$ are independent.

**Large Samples**

$\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n$, $\sim$, distribution with mean $\mu$ and covariance matrix $\Sigma$

$\bar{X}$ and $S$ sample mean and sample variance-covariance matrix

**LLN**: (Law of Large Number)

$$\bar{X} \to \vec{\mu}, \text{ and } S \to \Sigma \text{ in probability when } n \to +\infty$$

**CLT**: (Central Limit Theorem)

$$\sqrt{n}(\bar{X} - \vec{\mu}) \approx N_p(0, \Sigma)$$

$$n(\bar{X} - \vec{\mu})' S^{-1} (\bar{X} - \vec{\mu}) \approx \chi^2_p$$

when $n \to +\infty$

**Checking for multivariate normality**

Univariate normality

Q-Q plot:

1. Order the original observations to get $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, and their corresponding probabilities values $(1 - 1/2)/n, (2 - 1/2)/n, \ldots, (n - 1/2)/n$.

2. Calculate the standard normal quantiles $q_{(1)}, q_{(2)}, \ldots, q_{(n)}$.

3. Plot the pairs of observations $(q_{(1)}, x_{(1)}), \ldots, (q_{(n)}, x_{(n)})$ and examine the straightness of the outcome.

Bivariate Normality

make scatter plot and look for an elliptical scatter

General Normality

Chi-square Plot:

1. Calculate the squared generalized distances ( or the Mahalanobis distances):

$$d_i^2 = (\vec{x}_i - \bar{x})' S^{-1} (\vec{x}_i - \bar{x})$$

where $i = 1, 2, \ldots, n$

2. Order the distances from smallest to largest as $d^2_{(1)} \leq d^2_{(2)} \leq \cdots \leq d^2_{(n)}$.

3. Graph the pairs $(q_{c,p}((i-1/2)/n), d^2_{(i)})$, where $q_{c,p}((i-1/2)/n)$ is the 100(i-1/2)/n quantile of the chi-square distribution with $p$ degrees of freedom.

**Transformations to Near Normality**

Special cases:

| Original Scale | Transformed scale |
|---|---|
| 1. Counts, $y$ | $\sqrt{y}$ |
| 2. Proportions, $\hat{p}$ | $\text{logit}(\hat{p}) = \frac{1}{2}\log\frac{\hat{p}}{1-\hat{p}}$ |
| 3. Correlations, $r$, | Fisher's$Z(r) = \frac{1}{2}\log\frac{1+r}{1-r}$ |

General power transformations:

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(x) & \lambda = 0 \end{cases}$$

$$\max_\lambda l(\lambda) = \max_\lambda \{-\frac{n}{2}\ln[\frac{1}{n}\sum_{i=1}^{n}({x_i}^{(\lambda)} - \bar{x}^{(\lambda)})^2] + (\lambda - 1)\sum_{i=1}^{n}\ln x_i\}$$

where

$$\bar{x}^{(\lambda)} = \frac{1}{n}\sum_{i}^{n} x_i^{(\lambda)}$$

Apply power transformation to each element of $x = (x_1, x_2, \ldots, x_n)$.