The Univariate Normal Distribution

- $Y \sim \mathsf{N}(\mu, \sigma^2)$, $\mathbb{E}(Y) = \mu, \mathsf{Var}(Y) = \sigma^2$, with pdf

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}.$$

- $Z \sim \mathsf{N}(0, 1)$: the standard normal rv. $\Phi(z)$ denotes its CDF.

$$\Phi(-z) = 1 - \Phi(z), \quad z > 0.$$

- Linear transformations of normals are still normal. $Y \sim \mathsf{N}(\mu, \sigma^2)$, then

$$aY + b \sim \mathsf{N}(a\mu + b, a^2\sigma^2), \quad \frac{1}{\sigma}(Y - \mu) \sim N(0, 1).$$

- Linear combinations of independent normal rv's are normal.

The Bivariate Normal Distribution

$(Y_1, Y_2) \sim \mathsf{N}_2(\boldsymbol{\mu}, \Sigma)$, where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

- μ_1 and σ_1^2 are the mean and variance, respectively, of Y_1;
- μ_2 and σ_2^2 are the mean and variance, respectively, of Y_2;
- $\sigma_{12} = \rho\sigma_1\sigma_2$ is the covariance of Y_1 and Y_2 with ρ being the correlation coefficient.

Assume $\rho^2 < 1$, and the joint pdf $f(y_1, y_2)$ is given by

$$
\begin{aligned}
& \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \\
& \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right) + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right]\right\} \\
= \quad & \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^t\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}.
\end{aligned}
$$

Note that $|\Sigma| = \sigma_1^2\sigma_2^2(1-\rho^2)$ and when $\rho^2 < 1$,

$$
\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}.
$$

- Marginals: $Y_i \sim \mathsf{N}(\mu_i, \sigma_i^2)$ where $i = 1, 2$.

- Conditionals:

$$Y_1 \mid Y_2 = y_2 \quad \sim \quad \mathsf{N}\Big(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y_2 - \mu_2), (1-\rho^2)\sigma_1^2\Big)$$

$$Y_2 \mid Y_1 = y_1 \quad \sim \quad \mathsf{N}\Big(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1), (1-\rho^2)\sigma_2^2\Big)$$

- Linear Combinations:

$$aY_1 + bY_2 \sim \mathsf{N}\big(a\mu_1 + b\mu_2, a^2\sigma_1^2 + 2ab\rho\sigma_1\sigma_2 + b^2\sigma_2^2\big)$$

- Uncorrelated = Independent: $\rho = 0$ implies Y_1 and Y_2 are independent.

- Note: All the statements above assume that the joint distribution of (Y_1, Y_2) is normal. However,

$$Y_1 \sim \text{Norm and } Y_2 \sim \text{Norm} \quad \text{do NOT imply} \quad (Y_1, Y_2) \sim \text{Norm.}$$

  So if Y_1 and Y_2 are marginally normally distributed with correlation 0, we cannot conclude that Y_1 and Y_2 are independent (see the homework problem).

The Multivariate Normal Distribution

- Let $\mathbf{Z} = (Z_1, \ldots, Z_n)$ where Z_i's are iid $\sim \mathsf{N}(0, 1)$ rv's. Then $\mathbf{Z}$ follows a multivariate normal distribution, denoted by $\mathsf{N}_n(\mathbf{0}, \mathbf{I}_n)$, with pdf

$$\begin{aligned} f(\mathbf{z}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-z_i^2} = \frac{1}{(2\pi)^{n/2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} z_i^2 \right\} \\ &= \frac{1}{(2\pi)^{n/2}} \exp\left\{ -\frac{1}{2} \mathbf{z}^t \mathbf{z} \right\}, \end{aligned}$$

and moment generating function

$$M(\mathbf{t}) = \mathbb{E}[\exp\{\mathbf{t}^t \mathbf{Z}\}] = \exp\left\{ \frac{1}{2} \mathbf{t}^t \mathbf{t} \right\},$$

and mean and covariance

$$\mathbb{E}(\mathbf{Z}) = \mathbf{0}, \qquad \mathsf{Cov}(\mathbf{Z}) = \mathbf{I}_n.$$

- $\mathbf{Y}$ has a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance Σ, denoted by $\mathsf{N}_n(\boldsymbol{\mu}, \Sigma)$ if its moment generating function is

$$M_Y(\mathbf{t}) = \exp\left\{\mathbf{t}^t\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^t\Sigma\mathbf{t}\right\}.$$

Why don't we define $\mathbf{Y}$ via its pdf? (It may not exist.)

- Recall the definition of the covariance matrix for a random vector $\mathbf{Y}$. Any covariance matrix Σ should be symmetric and *positive semi-definite* (psd), where psd means

$$\mathbf{a}^t \Sigma \mathbf{a} \geq 0.$$

This is because

$$0 \leq \mathsf{Var}(\mathbf{a}^t \mathbf{Y}) = \mathbf{a}^t \Sigma \mathbf{a}.$$

Any symmetric psd matrix has a *spectral decomposition*

$$\Sigma = \Gamma^t \Lambda \Gamma, \qquad \Lambda = \mathsf{diag}(\lambda_i)_{i=1}^n,$$

where $\Gamma_{n \times n}$ is an orthonormal matrix, i.e., $\Gamma\Gamma^t = \mathbf{I}_n$, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.

- If $\lambda_n > 0$, i.e., Σ is of full rank, then $|\Sigma| > 0$ and Σ^{-1} exists. The pdf of $Y \sim \mathsf{N}_n(\boldsymbol{\mu}, \Sigma)$ is given by

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^t\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}.$$

- If $\lambda_{m+1} = \cdots = \lambda_n = 0$ where $m < n$, then Σ is not of full rank, so the pdf does not exist (since Σ^{-1} is not defined). In fact, the support of $\mathsf{N}_n(\boldsymbol{\mu}, \Sigma)$ in this case is not $\mathbb{R}^n$.

  We can project $Y_{p\times 1}$ to the m-dim subspace spanned by the first m columns of Γ, which is a m-dim Gaussian random vector with a valid pdf.

  The projection of Y to the $(n-m)$-dim subspace spanned by remaining $(n-m)$ columns of Γ is a constant, equal to the corresponding projection of $\boldsymbol{\mu}$.

Properties of Multivariate Normals

- Affine transformations of a normal vector are still normal:

$$\mathbf{Y} \sim \mathsf{N}_n(\boldsymbol{\mu}, \Sigma) \Longrightarrow A_{m\times n}\mathbf{Y} + b_{m\times 1} \sim \mathsf{N}_m(A\boldsymbol{\mu} + b, A\Sigma A^t).$$

- Marginals of a normal are still normal.

- Conditionals of a normal are still normal.

$$\mathbf{Y}_1|\mathbf{Y}_2 \sim \mathsf{N}_m\Big(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{Y}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Big)$$

- For multivariate normals, uncorrelated = independent.

Gaussian Graphical Model

Suppose $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^t \sim \mathsf{N}_n(\mathbf{0}, \Sigma)$ where Σ is of full rank. Denote its precision matrix Σ^{-1} by Θ.

- (Marginal independence) Y_1 is independent of Y_2, if $\Sigma_{12} = 0$.
- (Conditional independence) Y_1 is independent of Y_2 conditioning on $Y_3, \ldots, Y_n$, if $\theta_{12} = 0$.

Graphical models are often used to describe the dependence among random variables: nodes corresponds to variables Y_i's and an edge between node i and node j indicates that Y_i and Y_j are conditionally **NOT** independent given all the other variables. With the Gaussian assumption, to retrieve the graph structure, all we need to do is to identify the zero patterns in the precision matrix Θ[a].

[a] Known as the sparse covariance matrix estimation.

Show that the conditional independence is related to entries in Θ: Y_1 is independent of Y_2 conditioning on $Y_3, \ldots, Y_n$, if $\theta_{12} = 0$.

One approach is to find the conditional distribution $p(y_1, y_2 \mid \mathbf{y}_{3:n})$, which is normal by the property of multivariate normal distributions. And then show the corresponding covariance is related to θ_{12}.

Another approach: show that $p(y_1, y_2 \mid \mathbf{y}_{3:n})$ can be factorized into two parts $f_1(y_1) \times f_2(y_2)$ if $\theta_{12} = 0$, therefore Y_1 and Y_2 are conditionally independent.

$$
\begin{aligned}
p(y_1, y_2 \mid \mathbf{y}_{3:n}) &\propto p(y_1, y_2, \ldots, y_n) \\
&= \frac{|\Theta|^{1/2}}{(2\pi)^{n/2}} \exp\Big\{ - \sum_{i,j=1}^{n} y_i y_j \theta_{ij} \Big\} \\
&\propto \exp\Big\{ - 2\theta_{12} y_1 y_2 \Big\} \times \exp\Big\{ - \theta_{11} {y_1}^2 - \sum_{j=3}^{n} y_1 y_j \theta_{1j} \Big\} \\
&\quad \times \exp\Big\{ - \theta_{22} {y_2}^2 - \sum_{j=3}^{n} y_2 y_j \theta_{2j} \Big\}
\end{aligned}
$$

Distributions Related to Normals

- Z_i iid $\sim N(0,1)$, then $Z_1^2 + \cdots + Z_n^2 \sim \chi_n^2$.

$$W \sim \chi_n^2, \qquad \mathbb{E}(W) = n, \qquad \mathsf{Var}(W) = 2n.$$

- $\mathbf{Z} \sim \mathsf{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, then

$$\|\mathbf{Z}\|^2 \sim \sigma^2 \chi_n^2.$$

- $\mathbf{Y} \sim \mathsf{N}_n(\boldsymbol{\mu}, \Sigma)$ and Σ^{-1} exists, then

$$(\mathbf{Y} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \sim \chi_n^2.$$

- If $\mathbf{W} \sim \mathsf{N}_n(\mathbf{0}, \mathbf{H})$ where $\mathbf{H}$ is a projection matrix, then

$$\|\mathbf{W}\|^2 \sim \chi^2_{\mathsf{tr}(\mathbf{H})}.$$

- $Z \sim N(0,1)$ and $W \sim \chi^2_n/n$ are independent, then

$$\frac{Z}{\sqrt{W}} \sim t_n \text{ (student } t\text{-dist).}$$

- $W_1 \sim \chi^2_n/n$, $W_2 \sim \chi^2_m/m$ and they are independent, then

$$\frac{W_1}{W_2} \sim F_{n,m}.$$

Distributions of LS Estimates

The normal assumption for the linear regression model: $\mathbf{y} \sim \mathsf{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$.

Since any affine transformation of $\mathbf{y}$ is still normally distributed[a], we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}\mathbf{y} &\sim \mathsf{N}_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^t\mathbf{X})^{-1}), \\
\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} &\sim \mathsf{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{H}), \\
\mathbf{r} = (\mathbf{I}_n - \mathbf{H})\mathbf{y} &\sim \mathsf{N}_n(\mathbf{0}, \sigma^2(\mathbf{I}_n - \mathbf{H})).
\end{aligned}$$

Note that

$$\mathbb{E}\hat{\mathbf{y}} = \mathbf{H}\mathbb{E}\mathbf{y} = \mathbf{H}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta} \qquad \mathsf{Cov}(\hat{\mathbf{y}}) = \mathbf{H}\sigma^2\mathbf{H}^t = \mathbf{H}$$

$$\mathbb{E}\mathbf{r} = (\mathbf{I}_n - \mathbf{H})\mathbf{X}\boldsymbol{\beta} = \mathbf{0} \qquad \mathsf{Cov}(\mathbf{r}) = (\mathbf{I}_n - \mathbf{H})\sigma^2(\mathbf{I}_n - \mathbf{H})^t = \mathbf{I}_n - \mathbf{H}$$

[a] They are also jointly normal.

- Although $\mathbf{r}$ is an n-dim vector, it always lies in a subspace of dim $(n-p)$. It behaves like $\mathsf{N}_{n-p}(\mathbf{0}, \sigma^2 \mathbf{I}_{n-p})$ and we have

$$\hat{\sigma}^2 = \frac{\|\mathbf{r}\|^2}{n-p} \sim \sigma^2 \frac{\chi^2_{n-p}}{n-p}.$$

- We can show that $\hat{\mathbf{y}}$ (or $\hat{\boldsymbol{\beta}}$) and $\mathbf{r}$ (or $\hat{\sigma}^2$) are uncorrelated, since they are in two orthogonal spaces. Then plus the joint normal assumption, we conclude $\hat{\mathbf{y}}$ (or $\hat{\boldsymbol{\beta}}$) and $\mathbf{r}$ (or $\hat{\sigma}^2$) are independent.