16. Vectors of Random Variables

When working with several random variables X_1, X_2, ..., X_n, it is often convenient to arrange them in vector form

$$\mathbf{x} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

We can then make use of matrix algebra to help us organize and manipulate large numbers of random variables simultaneously. We **define** the expectation of a random vector as element-by-element expectation:

$$E[\mathbf{x}] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}.$$

If $\mathbf{X}$ is an $(m \times n)$ matrix of random variables, then $E[\mathbf{X}]$ is the $(m \times n)$ matrix where the (i, j)th element is the mean of the (i, j)th element of $\mathbf{X}$, i.e.,

$$\text{if } \mathbf{X} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix}, \text{ then } E[\mathbf{X}] = \begin{pmatrix} E[X_{11}] & E[X_{12}] & \cdots & E[X_{1n}] \\ E[X_{21}] & E[X_{22}] & \cdots & E[X_{2n}] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_{m1}] & E[X_{m2}] & \cdots & E[X_{mn}] \end{pmatrix}.$$

These definitions provide a neat way for computing the variances and covariances of the variables in $\mathbf{X}$ "all at once":

$$\text{var}[\mathbf{x}] = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])']$$

$$= E\begin{bmatrix} (X_1 - E[X_1])(X_1 - E[X_1]) & (X_1 - E[X_1])(X_2 - E[X_2]) & \cdots & (X_1 - E[X_1])(X_n - E[X_n]) \\ (X_2 - E[X_2])(X_1 - E[X_1]) & (X_2 - E[X_2])(X_2 - E[X_2]) & \cdots & (X_2 - E[X_2])(X_n - E[X_n]) \\ \vdots & \vdots & \ddots & \vdots \\ (X_n - E[X_n])(X_1 - E[X_1]) & (X_n - E[X_n])(X_2 - E[X_2]) & \cdots & (X_n - E[X_n])(X_n - E[X_n]) \end{bmatrix}$$

$$= \begin{bmatrix} \text{var}[X_1] & \text{cov}[X_1, X_2] & \cdots & \text{cov}[X_1, X_n] \\ \text{cov}[X_2, X_1] & \text{var}[X_2] & \cdots & \text{cov}[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}[X_n, X_1] & \text{cov}[X_n, X_2] & \cdots & \text{var}[X_n] \end{bmatrix}$$

We call $\text{var}[\mathbf{x}]$ the variance-covariance matrix of $\mathbf{x}$.

The formula $\text{var}[\mathbf{x}] = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])']$ can be viewed as the matrix version of the variance formula $\text{var}[X] = E[(X - E[X])^2]$ for a single variable.

Sometimes we want to compute a 'covariance matrix' between two vectors of random variables $\mathbf{x}$ and $\mathbf{y}$. We can compute

$$\text{cov}[\mathbf{x},\mathbf{y}] = E[(\mathbf{x}-E[\mathbf{x}])(\mathbf{y}-E[\mathbf{y}])']$$

$$= E\begin{bmatrix} (X_1-E[X_1])(Y_1-E[Y_1]) & (X_1-E[X_1])(Y_2-E[Y_2]) & \cdots & (X_1-E[X_1])(Y_n-E[Y_n]) \\ (X_2-E[X_2])(Y_1-E[Y_1]) & (X_2-E[X_2])(Y_2-E[Y_2]) & \cdots & (X_2-E[X_2])(Y_n-E[Y_n]) \\ \vdots & \vdots & \ddots & \vdots \\ (X_n-E[X_n])(Y_1-E[Y_1]) & (X_n-E[X_n])(Y_2-E[Y_2]) & \cdots & (X_n-E[X_n])(Y_n-E[Y_n]) \end{bmatrix}$$

$$= \begin{bmatrix} \text{cov}[X_1,Y_1] & \text{cov}[X_1,Y_2] & \cdots & \text{cov}[X_1,Y_n] \\ \text{cov}[X_2,Y_1] & \text{cov}[X_2,Y_2] & \cdots & \text{cov}[X_2,Y_n] \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}[X_n,Y_1] & \text{cov}[X_n,Y_2] & \cdots & \text{cov}[X_n,Y_n] \end{bmatrix}$$

<u>Rules for dealing with the mean vector and the variance-covariance matrix</u>

If $\mathbf{x}$ is an $(n\times 1)$ vector of random variables, $\mathbf{X}$ is an $(m\times n)$ matrix of random variables, $\mathbf{b}$ is an $(m\times 1)$ vector of constants, and $\mathbf{A}$ is an $(m\times n)$ matrix of constants, then

1. $E[\mathbf{Ax}+\mathbf{b}] = \mathbf{A}E[\mathbf{x}]+\mathbf{b}$
2. $\text{var-cov}[\mathbf{Ax}+\mathbf{b}] = \mathbf{A}\,\text{var}[\mathbf{x}]\mathbf{A}'$.

In particular,

$$\text{var}[c_1X_1+c_2X_2+\ldots+c_nX_n] = \text{var}[\mathbf{c}'\mathbf{x}] = \mathbf{c}'\,\text{var}[\mathbf{x}]\mathbf{c} = \sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_j\,\text{cov}[X_i,X_j]$$

3. A useful result is

$$\begin{aligned} E[tr[\mathbf{X}]] &= E[X_{11}+X_{22}+\ldots+X_{nn}] \\ &= E[X_{11}]+E[X_{22}]+\ldots+E[X_{nn}]. \\ &= tr[E[\mathbf{X}]] \end{aligned}$$

The first of these is straightforward to show by simply writing out the expression $\mathbf{Ax}+\mathbf{b}$ in full and taking expectations. This formula is the matrix version of the usual single variable result

$$E[aX+b] = aE[X]+b$$

To show (2), plug $\mathbf{Ax}+\mathbf{b}$ into the variance formula:

$$\begin{aligned} \text{var}[\mathbf{Ax}+\mathbf{b}] &= E[(\mathbf{Ax}+\mathbf{b}-E[\mathbf{Ax}+\mathbf{b}])(\mathbf{Ax}+\mathbf{b}-E[\mathbf{Ax}+\mathbf{b}])'] \\ &= E[(\mathbf{Ax}+\mathbf{b}-\mathbf{A}E[\mathbf{x}]-\mathbf{b})(\mathbf{Ax}+\mathbf{b}-\mathbf{A}E[\mathbf{x}]-\mathbf{b})'] \\ &= E[(\mathbf{Ax}-\mathbf{A}E[\mathbf{x}])(\mathbf{Ax}-\mathbf{A}E[\mathbf{x}])'] \\ &= E[\mathbf{A}(\mathbf{x}-E[\mathbf{x}])(\mathbf{A}(\mathbf{x}-E[\mathbf{x}]))'] \\ &= E[\mathbf{A}(\mathbf{x}-E[\mathbf{x}])(\mathbf{x}-E[\mathbf{x}])'\mathbf{A}'] \\ &= \mathbf{A}E[(\mathbf{x}-E[\mathbf{x}])(\mathbf{x}-E[\mathbf{x}])']\mathbf{A}' \\ &= \mathbf{A}\,\text{var}[\mathbf{x}]\mathbf{A}' \end{aligned}$$

This is the matrix version of the single variable result

$$\text{var}[aX+b] = a^2\,\text{var}[X].$$

Note that a variance-covariance matrix must be "positive-definite":

$$\text{var}[c_1X_1 + c_2X_2 + ... + c_nX_n] = \text{var}[\mathbf{c}'\mathbf{x}] = \mathbf{c}'\,\text{var}[\mathbf{x}]\mathbf{c}$$

has to be positive for all $\mathbf{c} \neq \mathbf{0}$, since variances must be positive.

Exercise Let $\mathbf{x} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ be a vector of random variables, and

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

be constants. Write out $\mathbf{Ax} + \mathbf{b}$ in full, and take expectations to show that

$$E[\mathbf{Ax} + \mathbf{b}] = \mathbf{A}E[\mathbf{x}] + \mathbf{b}$$

The Multivariate Normal Distribution

The random vector $\mathbf{x}$ follows the **multivariate normal distribution** with

$$\text{mean } E[\mathbf{x}] = \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \text{ and variance-covariance matrix } \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_{nn}^2 \end{bmatrix}$$

if its distribution has the form

$$f(\mathbf{x}) = (2\pi)^{-n/2} \,|\,\boldsymbol{\Sigma}\,|^{-1/2} \exp\{-(1/2)(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\}\,.$$

We denote this by $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. This is analogous to the univariate normal pdf:

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

Exercise Write the joint pdf out without matrix notation for the bivariate case

$$\mathbf{x} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}.$$

Then show that $f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$ when $\sigma_{12} = \sigma_{21} = 0$. What does this say?

Exercise Use computer software (say matlab) to plot the distribution of the bivariate normal for various parameter values.

The following are important properties of the multivariate normal:

4. The conditional distributions are also normal. In particular, if

$$\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right)$$

where $\mathbf{x}_1$ and $\boldsymbol{\mu}_1$ are $(n_1 \times 1)$, $\mathbf{x}_2$ and $\boldsymbol{\mu}_2$ are $(n_2 \times 1)$, $\boldsymbol{\Sigma}_{11}$ is $(n_1 \times n_1)$, $\boldsymbol{\Sigma}_{12}$ is $(n_1 \times n_2)$, $\boldsymbol{\Sigma}_{21}$ is $(n_2 \times n_1)$, and $(n_2 \times n_2)$, then

   a. the marginal distribution for $\mathbf{x}_1$ is $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$

   b. the conditional distribution for $\mathbf{x}_1$ given $\mathbf{x}_2$ is $N(\boldsymbol{\mu}_{1|2}\,,\ \boldsymbol{\Sigma}_{11|2})$ where

      $\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ and $\boldsymbol{\Sigma}_{11|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$

Exercise In the bivariate case,

$$\begin{bmatrix}X_1\\ X_2\end{bmatrix} \sim N\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11}^2 & \sigma_{12}\\ \sigma_{21} & \sigma_{22}^2\end{bmatrix}\right)$$

(4a) says that $X_1 \sim N(\mu_1, \sigma_{11}^2)$ and $X_2 \sim N(\mu_2, \sigma_{22}^2)$. Write out the expressions in (4b). Note that the conditional mean of X_1 given X_2 is a linear function of X_2.

Exercise Using the expressions for the conditional mean and conditional variance of X_2 given X_1, show that

$$f_{X_1,X_2}(x_1,x_2) = f_{X_2|X_1}(x_2 \mid x_1) f_{X_1}(x_1).$$

(You can make a similar argument for the general $\mathbf{x}_1$ and $\mathbf{x}_2$ case.)

5. If $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{Ax} + \mathbf{b} \sim N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$

The expression $\mathbf{Ax} + \mathbf{b}$ is normal because linear combinations of normal random variables remain normal. The formulae for the mean and variance-covariance matrix are the usual ones.

6. If $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2)$, then the random variables in $\mathbf{x}$ are independent.

The following make use of the fact that

   The square of a standard normal variable has a χ_1^2 distribution;

   The sum of n independent χ_1^2 is a χ_n^2;

   If $X \sim N(0,1)$, $Y \sim \chi_n^2$ and X and Y are independent, then $\dfrac{X}{\sqrt{Y/n}} \sim t_n$

   If $Y_1 \sim \chi_n^2, Y_2 \sim \chi_m^2$, and Y_1 and Y_2 are independent, then $\dfrac{Y_1/n}{Y_2/m} \sim F_{(n,m)}$

We have:

7. If $\mathbf{x} \sim N(\mathbf{0},\mathbf{I})$ and $\mathbf{A}$ is symmetric, and idempotent with rank J, then the scalar random variable $\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi^2_{(J)}$. In particular, $\mathbf{x}'\mathbf{x} \sim \chi^2_{(n)}$.

   Proof

   Because $\mathbf{A}$ is symmetric, we can write $\mathbf{A} = \mathbf{C}\mathbf{\Lambda}\mathbf{C}'$, with $\mathbf{C}'\mathbf{C} = \mathbf{I}$. Note that $\mathbf{C}'\mathbf{x} \sim N(\mathbf{0},\mathbf{I})$ because $\text{var}(\mathbf{C}'\mathbf{x}) = \mathbf{C}'\mathbf{I}\mathbf{C} = \mathbf{C}'\mathbf{C} = \mathbf{I}$. That is, $\mathbf{C}'\mathbf{x}$ is a vector of independently distributed standard normal variables.

   Write

   $$\mathbf{x}'\mathbf{A}\mathbf{x} \;=\; \mathbf{x}'\mathbf{C}\mathbf{\Lambda}\mathbf{C}'\mathbf{x} \;=\; \mathbf{y}'\mathbf{\Lambda}\mathbf{y} \;=\; \sum_{i=1}^{n} \lambda_i y_i^2$$

   where $\mathbf{y} = \mathbf{C}'\mathbf{x}$. Each y_i^2 is an independent chi-sq degree one, since the y_i's are independent standard normal variables. Because $\mathbf{A}$ is idempotent, there are J λ_i's equal to one, and $(n-J)$ λ_i's that are zero. Relabeling the y_i's so that the first λ_i's are equal to one, we have

   $$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{J} y_i^2$$

   which is a sum of J independent χ^2_J. Therefore $\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi^2_{(J)}$.

8. If $\mathbf{x} \sim N(\boldsymbol{\mu},\mathbf{\Sigma})$, then $(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \sim \chi^2_{(n)}$.

   Proof

   $\mathbf{\Sigma}$ is positive definite, symmetric, and full rank. Therefore we can write $\mathbf{\Sigma}^{-1} = \mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2}$. Note that $\mathbf{z} = \mathbf{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}) \sim N(\mathbf{0},\mathbf{I})$, therefore

   $$(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \;=\; (\mathbf{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}))'\mathbf{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}) \;=\; \mathbf{z}'\mathbf{z} \sim \chi^2_{(n)}.$$

9. If $\mathbf{x} \sim N(\mathbf{0},\mathbf{I})$, and $\mathbf{A}$ and $\mathbf{B}$ are symmetric and idempotent, then $\mathbf{x}'\mathbf{A}\mathbf{x}$ and $\mathbf{x}'\mathbf{B}\mathbf{x}$ are independent if $\mathbf{A}\mathbf{B} = \mathbf{0}$.

   Proof

   Because $\mathbf{A}$ and $\mathbf{B}$ are symmetric and idempotent, we have $\mathbf{A}'\mathbf{A} = \mathbf{A}$ and $\mathbf{B}'\mathbf{B} = \mathbf{B}$. Therefore we can write the quadratic forms as $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})'(\mathbf{A}\mathbf{x})$. Because $\mathbf{x}$ is normal with mean $\mathbf{0}$, $\mathbf{A}\mathbf{x}$ also normal with mean $\mathbf{0}$. For vectors of zero mean random variables, $\text{cov}[\mathbf{x},\mathbf{y}] = E[\mathbf{x}\mathbf{y}']$ (why?). We have

   $$\text{cov}\left[\mathbf{A}\mathbf{x},\mathbf{B}\mathbf{x}\right] = E\left[\mathbf{A}\mathbf{x}\mathbf{x}'\mathbf{B}'\right] = \mathbf{A}E[\mathbf{x}\mathbf{x}']\mathbf{B}' = \mathbf{A}\mathbf{B}' = \mathbf{A}\mathbf{B}.$$

   Therefore, $\mathbf{A}\mathbf{B} = \mathbf{0}$ implies that $\mathbf{A}\mathbf{x}$ and $\mathbf{B}\mathbf{x}$ are normally distributed, with covariance $\mathbf{0}$. This implies that $\mathbf{A}\mathbf{x}$ and $\mathbf{B}\mathbf{x}$ are independent (why?), and therefore the quadratic forms $\mathbf{x}'\mathbf{A}\mathbf{x}$ and $\mathbf{x}'\mathbf{B}\mathbf{x}$ are also independent.

   It follows from (7) that $\dfrac{[\mathbf{x}'\mathbf{A}\mathbf{x}/rank(\mathbf{A})]}{[\mathbf{x}'\mathbf{B}\mathbf{x}/rank(\mathbf{B})]}$ is distributed $F_{(rank(\mathbf{A}),\, rank(\mathbf{B}))}$.

10. If $\mathbf{x} \sim N(\mathbf{0},\mathbf{I})$, and $\mathbf{A}$ is symmetric and idempotent, then $\mathbf{L}\mathbf{x}$ and $\mathbf{x}'\mathbf{A}\mathbf{x}$ are independent if $\mathbf{L}\mathbf{A} = \mathbf{0}$.

    Proof

    Same idea as in (9).

We use these results to prove some standard results in statistics. Suppose $X_1, X_2, ..., X_n$ are n independent draws from a $N(\mu, \sigma^2)$ distribution, i.e.

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I}).$$

where $\boldsymbol{\mu}$ is an $(n \times 1)$ vector $\boldsymbol{\mu} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} = \mu \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \mu \mathbf{i}$.

We know that the sample mean

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is normally distributed (because a linear combination of normal variables is normal) with mean

$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] = \frac{1}{n} n\mu = \mu$$

and variance

$$\operatorname{var}[\bar{X}] = \operatorname{var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{var}\left[X_i\right] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Furthermore, $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ implies that

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2 / n}} \sim N(0,1)$$

Unfortunately, this result is not very helpful if, for instance, you want to test a hypothesis on μ, since σ^2 is general unknown, and must be estimated. An unbiased estimator is

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2$$

To show this, note that $\sum_{i=1}^{n} (X_i - \bar{X})^2 = \left(\sum_{i=1}^{n} X_i^2\right) - n\bar{X}^2$. Each X_i has variance

$$\operatorname{var}[X_i] = E[X_i^2] - \mu^2 = \sigma^2,$$

so

$$E\left[\sum_{i=1}^{n} X_i^2\right] = \sum_{i=1}^{n} E\left[X_i^2\right] = \sum_{i=1}^{n} \left(\sigma^2 + \mu^2\right) = n\sigma^2 + n\mu^2.$$

Also, $E\left[\bar{X}^2\right] = \operatorname{var}\left[\bar{X}\right] + E\left[\bar{X}\right]^2 = \frac{\sigma^2}{n} + \mu^2$.

Putting all this together, we have:

$$\begin{aligned}
E\left[\hat{\sigma}^2\right] &= \frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right] \\
&= \frac{1}{n-1}E\left[\left(\sum_{i=1}^{n}X_i^2\right) - n\bar{X}^2\right] \\
&= \frac{1}{n-1}\left(n\sigma^2+n\mu^2-n\left(\frac{\sigma^2}{n}+\mu^2\right)\right) \\
&= \frac{1}{n-1}(n-1)\sigma^2 \\
&= \sigma^2.
\end{aligned}$$

One idea, then, is to substitute σ^2 with $\hat{\sigma}^2$ in $\dfrac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}$ to get the 't-statistic'

$$t = \frac{\bar{X}-\mu}{\sqrt{\hat{\sigma}^2/n}}.$$

Unfortunately, the t-statistic does not have a standard normal distribution.

We use the results discussed earlier to derive the distribution of the t - statistic. We begin by deriving a matrix expression for $\hat{\sigma}^2$. Observe first that

$$(\mathbf{I}-\mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}')\mathbf{x} = \left\{\begin{bmatrix}1&0&\cdots&0\\0&1&\cdots&0\\ \vdots&\vdots&\ddots&\vdots\\0&0&\cdots&1\end{bmatrix} - \begin{bmatrix}1\\1\\ \vdots\\1\end{bmatrix}\left\{\begin{bmatrix}1&1&\cdots&1\end{bmatrix}\begin{bmatrix}1\\1\\ \vdots\\1\end{bmatrix}\right\}^{-1}\begin{bmatrix}1&1&\cdots&1\end{bmatrix}\right\}\begin{bmatrix}X_1\\X_2\\ \vdots\\X_n\end{bmatrix}$$

$$= \left\{\begin{bmatrix}X_1\\X_2\\ \vdots\\X_n\end{bmatrix} - \begin{bmatrix}1\\1\\ \vdots\\1\end{bmatrix}(1/n)\sum_{i=1}^{n}X_i\right\}$$

$$= \begin{bmatrix}X_1-\bar{X}\\X_2-\bar{X}\\ \vdots\\X_n-\bar{X}\end{bmatrix}$$

An interesting property of the matrix $\mathbf{M}=\mathbf{I}-\mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'$ is that it is symmetric and idempotent, with rank $n-1$. A matrix $\mathbf{M}$ is symmetric if $\mathbf{M}=\mathbf{M}'$. It is idempotent if $\mathbf{MM}=\mathbf{M}$.

Symmetric: $\mathbf{M}' = \left(\mathbf{I}-\mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'\right)' = \mathbf{I}'-\mathbf{i}''\left((\mathbf{i}'\mathbf{i})^{-1}\right)'\mathbf{i} = \mathbf{I}-\mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}' = \mathbf{M}$

Idempotent:

$$\begin{aligned}
\mathbf{MM} &= \left(\mathbf{I}-\mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'\right)\left(\mathbf{I}-\mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'\right) \\
&= \mathbf{II} - \mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'\mathbf{I} - \mathbf{Ii}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}' + \mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\underbrace{\mathbf{i}'\mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}}_{\text{cancels out}}\mathbf{i}' \\
&= \mathbf{I} - \mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}' - \mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}' + \mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}' \\
&= \mathbf{M}
\end{aligned}$$

Therefore

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \\
&= \frac{1}{n-1}\left\{\begin{bmatrix} X_1-\bar{X} & X_2-\bar{X} & & X_n-\bar{X}\end{bmatrix}\begin{bmatrix} X_1-\bar{X} \\ X_2-\bar{X} \\ \vdots \\ X_n-\bar{X}\end{bmatrix}\right\} \\
&= \frac{1}{n-1}\mathbf{x}'\mathbf{M}'\mathbf{Mx} \\
&= \frac{1}{n-1}\mathbf{x}'\mathbf{Mx}
\end{aligned}$$

Because $\mathbf{M}$ symmetric and idempotent, $rank(\mathbf{M}) = trace(\mathbf{M})$, and

$$trace\left(\mathbf{I}-\mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'\right) = trace(\mathbf{I})-trace(\mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}') = n-trace(\mathbf{i}'\mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}) = n-1$$

Furthermore, note that

$$\frac{1}{\sigma}(\mathbf{x}-\boldsymbol{\mu}) \sim N(\mathbf{0},\mathbf{I})$$

$\mathbf{Mx}=\mathbf{M}(\mathbf{x}-\boldsymbol{\mu})$ since $\mathbf{M}\boldsymbol{\mu}=\mathbf{0}$ (why?)

Together with the fact that $\mathbf{M}$ is symmetric and idempotent with rank $n-1$, result (7) implies that

$$\frac{1}{\sigma^2}\mathbf{x}'\mathbf{Mx} = \frac{1}{\sigma^2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{M}(\mathbf{x}-\boldsymbol{\mu}) \sim \chi^2_{n-1}.$$

This result is consistent with the fact that $E[\hat{\sigma}^2]=\sigma^2$. The mean of a χ^2_{n-1} is $n-1$, i.e.

$$\frac{1}{\sigma^2}E[\mathbf{x}'\mathbf{Mx}] = n-1.$$

Since $\hat{\sigma}^2 = \frac{1}{n-1}\mathbf{x}'\mathbf{Mx}$, the result follows. The fact that

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\mathbf{x}'\mathbf{Mx} \sim \chi^2_{n-1},$$

which we have just shown, is of course a much stronger one. The result that $E[\hat{\sigma}^2]=\sigma^2$ does not depend on the normality of $\mathbf{x}$. If we have normality of $\mathbf{x}$, then we have the χ^2 result.

Finally, note that

$$\frac{\sqrt{n}}{\sigma}\left(\bar{X}-\mu\right) = \frac{1}{\sigma}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'(\mathbf{x}-\boldsymbol{\mu}) \sim N(0,1),$$

and that

$$(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'\mathbf{M} = \left[(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'\right]\left[\mathbf{I}-\mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'\right] = (\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}' - (\mathbf{i}'\mathbf{i})^{-1}\underbrace{\mathbf{i}'\mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}}_{\text{cancels out}}\mathbf{i}' = \mathbf{0}$$

which says that

$$\frac{\sqrt{n}}{\sigma}\left(\bar{X}-\mu\right) = \frac{1}{\sigma}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'(\mathbf{x}-\boldsymbol{\mu}) \sim N(0,1) \quad \text{and} \quad \frac{(n-1)\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\mathbf{x}'\mathbf{M}\mathbf{x} \sim \chi^2_{n-1}$$

are independent. Therefore,

$$t = \frac{\bar{X}-\mu}{\sqrt{\hat{\sigma}^2/n}} = \frac{\frac{\sqrt{n}}{\sigma}(\bar{X}-\mu)}{\sqrt{\left(\frac{(n-1)\hat{\sigma}^2}{\sigma^2}\right)\Big/(n-1)}} = \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} \sim t_{n-1}.$$

Of course, if X_i's are not random draws from $N(\mu,\sigma^2)$, then all of these results do not hold (except $E[\bar{X}]=\mu$, $\mathrm{var}[\bar{X}]=\sigma^2/n$ and $E[\hat{\sigma}^2]=\sigma^2$, which does not require the normality assumption). Under reasonable conditions, the t-statistic will converge to the normal as the sample size grows.