Random Vector

Normal, Multivariate Normal, χ^2, t and F Distributions

ZHANG Wei

March 5, 2024

6.0.1 Mutual vs Pairwise Independence

Definition 6.0.1. *The random variables X_1, X_2, $\cdots$, X_n are* **mutually independent** *iff*

$$f(x_1, \cdots, x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n) = \prod_{i=1}^{n} f_i(x_i) \tag{6.1}$$

for the continuous case, and

$$p(x_1, \cdots, x_n) = p_1(x_1)p_2(x_2)\cdots p_n(x_n) = \prod_{i=1}^{n} p_i(x_i) \tag{6.2}$$

for the discrete case.

Theorem 6.0.2. *(Generalization of Theorem **??**) Let $X_1, \cdots, X_n$ be n mutually independent random variables. Let $g_1, \cdots, g_n$ be real-valued functions such that $g_i(x_i)$ is a function only of x_i, $i = 1, \cdots, n$. Then*

$$E\big[g_1(X_1)g_2(X_2)\cdots g_n(X_n)\big] = E\big[g_1(X_1)\big]E\big[g_2(X_2)\big]\cdots E\big[g_n(X_n)\big]. \tag{6.3}$$

Definition 6.0.3. *(Generalization of Definition **??**) The* **moment-generating function** *(mgf) of the joint distribution of n random variables X_1, X_2, $\cdots$, X_n is defined as*

$$M(t_1, t_2, \cdots, t_n) = E\big[e^{t_1X_1+t_2X_2+\cdots t_nX_n}\big] = E\big[e^{\mathbf{t}'\mathbf{X}}\big], \quad \mathbf{t} \in B \subset \mathbb{R}^n, \tag{6.4}$$

where $B = \{\mathbf{t} : -h_i < t_i < h_i, i = 1, \cdots, n\}$.

Likewise, this **mgf is unique and uniquely determines the joint distribution** of the n variables (and hence all marginal distributions).

Theorem 6.0.4. *(Generalization of Theorem **??**) Random variables $X_1, \cdots, X_n$ with marginal mgfs $M(0, \cdots, 0, t_i, 0, \cdots, 0)$, for $i = 1, 2, \cdots, n$, are mutually independent, if and only if,*

$$M(t_1, t_2, \cdots, t_n) = \prod_{i=1}^{n} M(0, \cdots, 0, t_i, 0, \cdots, 0). \tag{6.5}$$

The following theorem gives the mgf of a linear combination of independent random variables.

Theorem 6.0.5. *Suppose X_1, X_2, $\cdots$, X_n are n mutually independent random variables, each with mgfs $M_i(t)$, for $-h_i < t < h_i$, where $h_i > 0$, $i = 1, 2, \cdots, n$. Let $Z = c_1X_1 + c_2X_2 + \cdots + c_nX_n$, where $c_1, c_2, \cdots, c_n$ are constants. Then the mgf of the new random variable Z is given by*

$$M_Z(t) = E(e^{tZ}) = \prod_{i=1}^{n} M_i(c_it), \qquad -\min_i\{h_i\} < t < \min_i\{h_i\}. \tag{6.6}$$

Remarks:

1. Mutual independence $\Rightarrow$ Pairwise independence. But the converse is not always true.
2. Unless there is a possible misunderstanding between *mutual* and *pairwise* independence, we usually drop "*mutual*".
3. If several random variables are mutually independent and have the same distribution, we say that they are **independent and identically distributed**, also known as **iid**.

Corollary 6.0.6. *Suppose $X_1, X_2, \cdots, X_n$ are iid random variables with the common mgf $M(t)$, for $h < t < h$, where $h > 0$. Let $T = \sum_i X_i$. Then T has the mgf given by*

$$M_T(t) = \big[M(t)\big]^n, \qquad -h < t < h. \tag{6.7}$$

6.0.2 Multivariate Variance-Covariance Matrix

For a random vector $\mathbf{X}$, we first define

$$\boldsymbol{\mu} = [E(X_i)]_{n\times 1} = E(\mathbf{X}),$$

which is the **mean** of the random vector $\mathbf{X}$. Then the random matrix

$$\begin{aligned}(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})' &= \begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \\ \vdots \\ X_n-\mu_n \end{pmatrix}\begin{pmatrix} X_1-\mu_1 & X_2-\mu_2 & \cdots & X_n-\mu_n \end{pmatrix} \\ &= \begin{pmatrix} (X_1-\mu_1)^2 & \cdots & (X_1-\mu_1)(X_n-\mu_n) \\ (X_2-\mu_2)(X_1-\mu_1) & \cdots & (X_2-\mu_2)(X_n-\mu_n) \\ \vdots & \ddots & \vdots \\ (X_n-\mu_n)(X_1-\mu_1) & \cdots & (X_n-\mu_n)^2 \end{pmatrix},\end{aligned}$$

has expectation $E[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})']$ and we define this expectation matrix as the **variance-covariance matrix** of $\mathbf{X}$

$$\boldsymbol{\Sigma} = E[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'] = [\sigma_{ij}],$$

whose elements $\sigma_{ij} = Cov(X_i, X_j)$ are the covariances between random variables X_i and X_j, with the diagonal elements being equal to the variances.

The linearity of the expectation operator also easily follows from this definition:

Theorem 6.0.7. *Let $\mathbf{W_1}$ and $\mathbf{W_2}$ be $m \times n$ matrices of random variables, let $\mathbf{A_1}$ and $\mathbf{A_2}$ be $k \times m$ matrices of constants, and let $\mathbf{B}$ be an $n \times l$ matrix of constants. Then*

$$E\big[\mathbf{A_1W_1} + \mathbf{A_2W_2}\big] = \mathbf{A_1}E\big[\mathbf{W_1}\big] + \mathbf{A_2}E\big[\mathbf{W_2}\big] \tag{6.8}$$

$$E\big[\mathbf{A_1W_1B}\big] = \mathbf{A_1}E\big[\mathbf{W_1}\big]\mathbf{B}. \tag{6.9}$$

Theorem 6.0.8. *Let $\mathbf{X} = (\mathbf{X_1}, \cdots, \mathbf{X_n})'$ be an n-dimensional random vector, such that $\sigma_i^2 = \sigma_{ii}^2 = Var(X_i) < \infty$. Let $\mathbf{A}$ be an $m \times n$ matrix of constants. Then*

$$Cov(\mathbf{X}) = E\big[\mathbf{XX}'\big] - \boldsymbol{\mu}\boldsymbol{\mu}' \tag{6.10}$$

$$Cov(\mathbf{AX}) = \mathbf{A}Cov(\mathbf{X})\mathbf{A}'. \tag{6.11}$$

Proof. Use Theorem 6.0.7 to derive (6.10); i.e.,

$$\begin{aligned} Cov(\mathbf{X}) &= E\big[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\big] \\ &= E\big[\mathbf{XX}' - \boldsymbol{\mu}\mathbf{X}' - \mathbf{X}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}'\big] \\ &= E\big[\mathbf{XX}'\big] - \boldsymbol{\mu}E\big[\mathbf{X}'\big] - E\big[\mathbf{X}\big]\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}', \end{aligned}$$

which is the desired result. □

Exercise 6.0.9. Prove the equality (6.11). (Hint: Use (6.10) and Theorem 6.0.7.) ✓

Remark: All variance-covariance matrices are **positive semi-definite**; that is, $\mathbf{a}'Cov(\mathbf{X})\mathbf{a} \geq 0$, for all vectors $\mathbf{a} \in \mathbb{R}^{\mathbf{n}}$. This can be shown by denoting a new random variable $Y = \mathbf{a}'\mathbf{X}$ and considering

$$\begin{aligned} \mathbf{a}'Cov\big(\mathbf{X}\big)\mathbf{a} &= E\big[\mathbf{a}'(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\mathbf{a}\big] \\ &= E\big[(Y - \mathbf{a}'\boldsymbol{\mu})^2\big] \qquad [\text{Since } Y = \mathbf{a}'\mathbf{X} = \mathbf{X}'\mathbf{a} \text{ and } E(Y) = \mathbf{a}'\boldsymbol{\mu}.] \\ &= Var(Y) \geq 0 \end{aligned}$$

6.0.3 Transformations for Several RVs

The nature of transformation is the change of variables in n-fold integrals.

For *continuous* case: Consider an integral of the form

$$\int \cdots \int_A f(x_1, x_2, \cdots, x_n) \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n$$

taken over a subset A of an n-dimensional space $\mathcal{S}$. Let

$$y_1 = u_1(x_1, x_2, \cdots, x_n),\ y_2 = u_2(x_1, x_2, \cdots, x_n), \cdots,\ y_n = u_n(x_1, x_2, \cdots, x_n),$$

together with the inverse functions

$$x_1 = w_1(y_1, y_2, \cdots, y_n),\ x_2 = w_2(y_1, y_2, \cdots, y_n), \cdots,\ x_n = w_n(y_1, y_2, \cdots, y_n)$$

define a **one-to-one** transformation that maps $\mathcal{S}$ **onto** $\mathcal{T}$ in the $y_1,\ y_2,\ \cdots,\ y_n$ space and, hence, maps the subset A of $\mathcal{S}$ onto a subset B of $\mathcal{T}$.

Define the **Jacobian** as the $n \times n$ determinant

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

which is not identically zero in $\mathcal{T}$. Then

$$\begin{aligned}&\int \cdots \int_A f(x_1, x_2, \cdots, x_n) \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n \\ &= \int \cdots \int_B f\big[w_1(y_1, y_2, \cdots, y_n), w_2(y_1, y_2, \cdots, y_n), \cdots, w_n(y_1, y_2, \cdots, y_n)\big] |J|\ \mathrm{d}y_1 \mathrm{d}y_2 \cdots \mathrm{d}y_n\end{aligned}$$

In order for this equality to hold for any region A, the integrands must be equal everywhere,

$$g(y_1, y_2, \cdots, y_n) = f\big[w_1(y_1, y_2, \cdots, y_n), \cdots, w_n(y_1, y_2, \cdots, y_n)\big] |J|, \tag{6.12}$$

where $(y_1, y_2, \cdots, y_n) \in \mathcal{T}$, and is zero elsewhere.

Example 6.0.10. Let X_1, X_2, X_3 have the joint pdf

$$f(\mathbf{x}) = \begin{cases} 120x_1x_2 & 0 < x_1, x_2, x_3, x_1 + x_2 + x_3 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Find the joint pdf of $Y_1 = \frac{X_1}{X_1+X_2+X_3}$, $Y_2 = \frac{X_2}{X_1+X_2+X_3}$, $Y_3 = X_1 + X_2 + X_3$.

Solution: The inverse transformation is given by

$$x_1 = y_1y_3, \quad x_2 = y_2y_3, \quad x_3 = y_3(1 - y_1 - y_2).$$

Inequalities defining the support are equivalent to

$$0 < y_1y_3,\ y_2y_3,\ y_3(1-y_1-y_2),\ y_3 < 1 \quad \to \quad 0 < y_1, y_2 < \frac{1}{y_3},\ 1-\frac{1}{y_3} < y_1+y_2 < 1,\ 0 < y_3 < 1.$$

The support of $\mathbf{Y}$ is

$$\mathcal{T} = \{(y_1, y_2, y_3) : y_1, y_2 > 0,\ 0 < y_1 + y_2 < 1,\ 0 < y_3 < 1\}.$$

The Jacobian is given by

$$J = \begin{vmatrix} y_3 & 0 & y_1 \\ 0 & y_3 & y_2 \\ -y_3 & -y_3 & 1 - y_1 - y_2 \end{vmatrix} = y_3^2.$$

Hence the joint pdf of Y_1, Y_2, Y_3 is

$$g(y_1, y_2, y_3) = 120(y_1y_3)(y_2y_3)|y_3^2| = 120y_1y_2y_3^4, \quad \text{for } \boldsymbol{y} \in \mathcal{T},$$

and zero elsewhere. ꝏ

Exercise 6.0.11. In the last example, find the joint marginal pdf of Y_1 and Y_2 and the marginal pdf of Y_2.

Ans: $g_{1,2} = \int_{\mathbb{R}} g(y_1, y_2, y_3)\mathrm{d}y_3 = 24y_1y_2, \quad y_1, y_2 > 0,\ 0 < y_1 + y_2 < 1$; zero elsewhere.

$g_2 = \int_{\mathbb{R}} g_{1,2}\ \mathrm{d}y_1 = 6y_2^2 - 8y_2^3 + 3y_2^4, \quad y_2 \in (0, 1)$; zero elsewhere. ✓

6.0.4 Linear Combinations of RVs

In this section, we consider linear combinations of these variables, i.e., $Z = \sum_{i=1}^n a_iX_i$, for a constant vector $\mathbf{a} = (a_1, \cdots, a_n)'$.

Theorem 6.0.12. *(The* **mean** *of Z) Provide $E\big[|X_i|\big] < \infty$, for $i = 1, \cdots, n$,*

$$E(Z) = \sum_{i=1}^n a_iE(X_i). \tag{6.13}$$

For the variance of Z, we first state a very general result involving covariances.

Theorem 6.0.13. *Let $Z = \sum_{i=1}^n a_iX_i$ and let $W = \sum_{j=1}^m b_jY_j$. If $E\big[X_i^2\big] < \infty$ and $E\big[Y_j^2\big] < \infty$, for $i = 1, \cdots, n$ and $j = 1, \cdots, m$, then*

$$Cov(Z, W) = \sum_{i=1}^n \sum_{j=1}^m a_ib_jCov(X_i, Y_j). \tag{6.14}$$

To obtain $Var(Z)$, simply replace W by Z in Theorem 6.0.13. We state the result as below:

Corollary 6.0.14. *(The* **variance** *of Z) Provide $E\big[X_i^2\big] < \infty$, for $i = 1, \cdots, n$,*

$$Var(Z) = Cov(Z, Z) = \sum_{i=1}^n a_i^2Var(X_i) + 2\sum_{i<j} a_ia_jCov(X_i, X_j). \tag{6.15}$$

Corollary 6.0.15. *If X_1, $\cdots$, X_n are independent random variables with finite variances, then $Cov(X_i, X_j) = 0$ and*

$$Var(Z) = \sum_{i=1}^n a_i^2Var(X_i). \tag{6.16}$$

6.1 The Normal Distribution

The **normal distribution** (sometimes called the *Gaussian distribution*) plays a central role in a large body of statistics. Based on the Central Limit Theorem (we will talk about it in Chapter 4), which shows that, under mild conditions, the normal distribution can be used to approximate a large variety of distributions in large samples. A random variable X has a normal distribution if its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad \text{for } -\infty < x < \infty.$$

The parameters μ and σ^2 are the mean and variance of X, respectively. We often denote it as $X \sim N(\mu, \sigma^2)$. Let $Z = (X-\mu)/\sigma$, then the distribution of Z is easily established by writing

$$\begin{aligned} P(Z \le z) &= P\left(\frac{X-\mu}{\sigma} \le z\right) \\ &= P(X \le \mu + z\sigma) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\mu+z\sigma} e^{-(x-\mu)^2/(2\sigma^2)} \mathrm{d}x \\ &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \mathrm{d}y, \qquad \left(\text{substitute } y = \frac{x-\mu}{\sigma}\right) \end{aligned}$$

we get the pdf of Z as

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\}, \quad \text{for } -\infty < z < \infty.$$

which is a normal distribution with $\mu = 0$ and $\sigma^2 = 1$, a.k.a the **standard normal distribution**. To show that the total area under the curve is equal to 1, i.e.

$$P(Z < \infty) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \mathrm{d}z = 1,$$

We only need to show $\int_0^\infty e^{-u^2}\, \mathrm{d}u = \frac{1}{2}\sqrt{\pi}$, by taking $u = z/\sqrt{2}$. It is proved in section **??**.

The **mgf of the standard normal distribution** is

$$M_Z(t) = \exp\left\{\frac{t^2}{2}\right\}.$$

By simple transformation $X = \mu + \sigma Z$, the mean, variance and the **mgf of** X are:

$$E(X) = \mu, \quad Var(X) = \sigma^2,$$

and

$$M_X(t) = E[e^{t\mu + t\sigma Z}] = e^{\mu t} E[e^{t\sigma Z}] = \exp\left\{\mu t + \sigma^2 \frac{t^2}{2}\right\},$$

for $-\infty < t < \infty$.

Example 6.1.1. If X has the mgf

$$M(t) = e^{2t+32t^2}$$

then the distribution of X is a normal distribution with $\mu = 2$ and $\sigma = 8$. ♾

Example 6.1.2. Let X be $N(\mu, \sigma^2)$, then

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = P\Big(-2 < \frac{X-\mu}{\sigma} < 2\Big) = 2P\Big(0 < \frac{X-\mu}{\sigma} < 2\Big) = 0.9544.$$

♾

Theorem 6.1.3. *If $X \sim N(\mu, \sigma^2)$, $\sigma^2 > 0$, then $Z^2 = (X-\mu)^2/\sigma^2 \sim \chi^2(1)$.*

Theorem 6.1.4. *Let $X_1, \cdots, X_n$ be independent with $X_i \sim N(\mu_i, \sigma_i^2)$. Let $Y = \sum_{i=1}^n a_i X_i$, where $a_i, \cdots, a_n$ are constants. Then the distribution of Y is*

$$N(\sum_{i=1}^n a_i\mu_i, \sum_{i=1}^n a_i^2\sigma_i^2).$$

Corollary 6.1.5. *Let $X_1, ..., X_n$ be iid with $X_i \sim N(\mu, \sigma^2)$. Let $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$. Then $\overline{X} \sim N(\mu, \sigma^2/n)$.*

6.2 The Multivariate Normal Distribution

Consider the random vector $\mathbf{Z} = (Z_1, \cdots, Z_n)'$, where $Z_1, \cdots, Z_n$ are iid $N(0,1)$ random variables. The pdf of $\mathbf{Z}$ is

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

for $\mathbf{z} \in \mathbb{R}^n$. The mean and covariance matrix of $\mathbf{Z}$ are

$$E[\mathbf{Z}] = \mathbf{0}, \quad Cov[\mathbf{Z}] = \mathbf{I}_n.$$

where $\mathbf{I}_n$ denotes the identity matrix of order n.

The **mgf** of $\mathbf{Z}$ is

$$\begin{aligned} M_{\mathbf{Z}}(\mathbf{t}) = E\big[\exp\{\mathbf{t}'\mathbf{Z}\}\big] &= E\Big[\prod_{i=1}^n \exp\{t_i Z_i\}\Big] = \prod_{i=1}^n E\big[\exp\{t_i Z_i\}\big] \\ &= \exp\left\{\frac{1}{2}\sum_{i=1}^n t_i^2\right\} = \exp\left\{\frac{1}{2}\mathbf{t}'\mathbf{t}\right\} \end{aligned}$$

for all $\mathbf{t} \in \mathbb{R}^n$. We say that $\mathbf{Z}$ has a **multivariate normal distribution** with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}_n$, or $\mathbf{Z} \sim N_n(\mathbf{0}, \mathbf{I}_n)$.

Let $\mathbf{Z}$ have an $N_n(\mathbf{0}, \mathbf{I}_n)$ distribution. Let $\mathbf{\Sigma}$ be a positive semi-definite, symmetric matrix and let $\boldsymbol{\mu}$ be an $n \times 1$ vector of constants. Define the random vector $\mathbf{X}$ by

$$\mathbf{X} = \mathbf{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}$$

Then by Theorem 6.0.8, we have

$$E[\mathbf{X}] = \boldsymbol{\mu} \quad , \quad Cov[\mathbf{X}] = \mathbf{\Sigma}^{1/2}\mathbf{I}_n(\mathbf{\Sigma}^{1/2})' = \mathbf{\Sigma},$$

and

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{t}) &= E\big[\exp\{\mathbf{t}'\mathbf{X}\}\big] = E\big[\exp\{\mathbf{t}'\mathbf{\Sigma}^{1/2}\mathbf{Z} + \mathbf{t}'\boldsymbol{\mu}\}\big] \\ &= \exp\{\mathbf{t}'\boldsymbol{\mu}\}\exp\left\{\frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}\right\} \end{aligned}$$

An n-dimensional random vector $\mathbf{X}$ has a multivariate normal distribution iff. its **mgf** is

$$M_{\mathbf{X}}(\mathbf{t}) = \exp\left\{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}\right\} \tag{6.17}$$

for all $\mathbf{t} \in \mathbb{R}^n$, where $\mathbf{\Sigma}$ is a positive semi-definite, symmetric matrix and $\boldsymbol{\mu} \in \mathbb{R}^n$. We say that $\mathbf{X}$ has an $N_n(\boldsymbol{\mu}, \mathbf{\Sigma})$ distribution.

The pdf of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}, \quad \text{for } \mathbf{x} \in \mathbb{R}^n.$$

Theorem 6.2.1. *Suppose $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \mathbf{\Sigma})$. Let $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Then $\mathbf{Y} \sim N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}')$.*

Proof. For $\mathbf{t} \in \mathbb{R}^m$, the mgf of $\mathbf{Y}$ is

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= E\left[\exp\{\mathbf{t}'\mathbf{Y}\}\right]' \\ &= E\left[\exp\{\mathbf{t}'(\mathbf{A}\mathbf{X} + \mathbf{b})\}\right] \\ &= \exp\{\mathbf{t}'\mathbf{b}\}E\left[\exp\{(\mathbf{A}'\mathbf{t})'\mathbf{X}\}\right] \\ &= \exp\{\mathbf{t}'\mathbf{b}\}\exp\{(\mathbf{A}'\mathbf{t})'\boldsymbol{\mu} + (1/2)(\mathbf{A}'\mathbf{t})'\mathbf{\Sigma}(\mathbf{A}\mathbf{t})\} \qquad \text{from (6.17)} \\ &= \exp\{\mathbf{t}'(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}) + (1/2)\mathbf{t}'\mathbf{A}\mathbf{\Sigma}\mathbf{A}'\mathbf{t}\}, \end{aligned}$$

which is the mgf of an $N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}')$ distribution. ∎

A simple corollary to this theorem gives marginal distributions of a multivariate normal random variable.

Let $\mathbf{X}_1$ be any subvector of $\mathbf{X}$, of dimension $m < n$. WLOG, we write $\mathbf{X}$ as

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$$

where $\mathbf{X}_2$ is of dimension $p = n - m$. In the same way, partition the mean and covariance matrix of $\mathbf{X}$:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Define $\mathbf{A}$ to be the matrix

$$\mathbf{A} = \left(\begin{array}{c|c} \mathbf{I}_m & \mathbf{O}_{mp} \end{array} \right)$$

where $\mathbf{O}_{mp}$ is an $m \times p$ matrix of zeros. Then $\mathbf{X}_1 = \mathbf{A}\mathbf{X}$. Hence, we have the corollary:

Corollary 6.2.2. *Suppose $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, partitioned as before. Then $\mathbf{X}_1 \sim N_n(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.*

Example 6.2.3. Explore the multivariate normal case when $n = 2$. The distribution in this case is called the bivariate normal.

Solution: Suppose $(X, Y) \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{bmatrix} \quad \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \qquad\qquad \end{bmatrix},$$

where $\sigma_{12} = \sigma_{21} = Cov(X, Y)$. Recall that $Cov(X, Y) =$ ________, where ρ is the correlation coefficient between X and Y.

Recall that $\rho^2 \leq 1$. For the remainder of this example, assume that $\rho^2 < 1$. Then

$$|\boldsymbol{\Sigma}| = \sigma_1^2\sigma_2^2 - \sigma_{12}\sigma_{21} = \sigma_1^2\sigma_2^2(1 - \rho^2).$$

Using this expression, the pdf of (X, Y) can be written as

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \exp^{-q/2}, \qquad \text{for } x, y \in \mathbb{R},$$

where

$$q = (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

Recall that if X and Y are independent, then their correlation coefficient is 0.

Based on the joint pdf of (X, Y), we see that if $\rho = 0$, then X and Y are independent.

$\rightarrow$ That is, for the bivariate normal case, independence $\Leftrightarrow \rho = 0$. ∞

Exercise 6.2.4. Prove that, in general, if X and Y are independent, then they must be uncorrelated. (Hint: Use definition) ✓

However, the converse is not always true. But it's true for the multivariate normal distribution.

Theorem 6.2.5. *Suppose* $\mathbf{X}$ *has a* $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *distribution, partitioned as before. Then* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are independent if and only if* $\boldsymbol{\Sigma}_{12} = \mathbf{O}$.

Proof. First note that $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}'_{12}$. The joint mgf of $\mathbf{X}_1$ and $\mathbf{X}_2$ is given by

$$M_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{t}_1, \mathbf{t}_2) = \exp\left\{\mathbf{t}'_1\boldsymbol{\mu}_1 + \mathbf{t}'_2\boldsymbol{\mu}_2 + \frac{1}{2}(\mathbf{t}'_1\boldsymbol{\Sigma}_{11}\mathbf{t}_1 + \mathbf{t}'_2\boldsymbol{\Sigma}_{22}\mathbf{t}_2 + \mathbf{t}'_2\boldsymbol{\Sigma}_{21}\mathbf{t}_1 + \mathbf{t}'_1\boldsymbol{\Sigma}_{12}\mathbf{t}_2)\right\}$$

where $\mathbf{t}' = (\mathbf{t}'_1, \mathbf{t}'_2)$ is partitioned the same as $\boldsymbol{\mu}$. By Corollary (6.2.2), $\mathbf{X}_1$ has a $N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ distribution and $\mathbf{X}_2$ has a $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ distribution. Hence, the product of their marginal mgfs is

$$M_{\mathbf{X}_1}(\mathbf{t}_1)M_{\mathbf{X}_2}(\mathbf{t}_2) = \exp\left\{\mathbf{t}'_1\boldsymbol{\mu}_1 + \mathbf{t}'_2\boldsymbol{\mu}_2 + \frac{1}{2}(\mathbf{t}'_1\boldsymbol{\Sigma}_{11}\mathbf{t}_1 + \mathbf{t}'_2\boldsymbol{\Sigma}_{22}\mathbf{t}_2)\right\}$$

By (6.0.4), $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if their joint mgf equals the product of their marginal mgfs, that is, the above two expressions are the same.

If $\boldsymbol{\Sigma}_{12} = \mathbf{O}$ and, hence, $\boldsymbol{\Sigma}_{21} = \mathbf{O}$, then the expressions are the same and $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent. If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, then the covariances between their components are all 0; i.e., $\boldsymbol{\Sigma}_{12} = \mathbf{O}$ and $\boldsymbol{\Sigma}_{21} = \mathbf{O}$. □

Theorem 6.2.6. *Suppose* $\mathbf{X}$ *has a* $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *distribution, partitioned as before. Assume that* $\boldsymbol{\Sigma}$ *is positive definite. Then the conditional distribution of* $\mathbf{X}_1|\mathbf{X}_2$ *is*

$$N_m(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

Theorem 6.2.7. *Suppose* $\mathbf{X}$ *has a* $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *distribution, where* $\boldsymbol{\Sigma}$ *is positive definite. Then the random variable* $W = (\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ *has a* $\chi^2(n)$ *distribution.*

6.3 Chi-Square (χ^2) Distribution

A random variable X of the continuous type that has the pdf

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(r/2)2^{r/2}}x^{r/2-1}e^{-x/2}, & x \geq 0 \\ 0, & \text{elsewhere.} \end{cases}$$

and the mgf

$$M(t) = (1 - 2t)^{-r/2}, \quad t < \frac{1}{2}$$

is said to have a **chi-square distribution**, denoted as $\chi^2(r)$. The parameter r is the number of **degrees of freedom** of the chi-square distribution. The chi-square distribution is actually a special case of the gamma distribution, Gamma$(\frac{r}{2}, \frac{1}{2})$, $r \in \mathbb{Z}^+$.

For a chi-square distribution,

$$E(X) = \alpha/\beta = (r/2)2 = r, \qquad Var(X) = \alpha/\beta^2 = (r/2)2^2 = 2r.$$

Remark: The χ^2 distribution has another definition. Also see in Theorem 6.1.3.

Let $X_1, \cdots, X_r$ be i.i.d. each following the standard normal distribution, $N(0,1)$. Then

$$X = X_1^2 + X_2^2 + \cdots + X_r^2 \sim \chi^2(r).$$

Theorem 6.3.1. *Let $X \sim \chi^2(r)$. If $k > -r/2$, then $E(X^k)$ exists and*

$$E(X^k) = \frac{2^k \Gamma(\frac{r}{2} + k)}{\Gamma(\frac{r}{2})}, \quad k > -r/2.$$

Corollary 6.3.2. *Let $X_1, \cdots, X_n$ be independent with $X_i \sim \chi^2(r_i)$. Let $Y = \sum_{i=1}^n X_i$. Then $Y \sim \chi^2(\sum_{i=1}^n r_i)$.*

6.4 t- and F- Distributions

6.4.1 The t-distribution

If $Z \sim N(0,1)$ and $V \sim \chi^2(r)$ with r degrees of freedom, and Z and V are independent, then

$$T = \frac{Z}{\sqrt{V/r}}$$

is a random variable following a **t-distribution** with r **degrees of freedom**. Its pdf is

$$f_T(t) = \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\,\Gamma(r/2)} \frac{1}{(1+t^2/r)^{(r+1)/2}}, \qquad -\infty < t < \infty.$$

Example 6.4.1. Let the random variable T have a t-distribution with r degrees of freedom. Then its mean and variance can be found as,

$$E(T) = E(Z)E(\sqrt{r/V}) = 0, \quad Var(T) = E(T^2) - 0 = E(Z^2)E(\frac{r}{V}) = \frac{r}{r-2}, \quad \text{for } r > 2,$$

by Theorem 6.3.1. ꝏ

6.4.2 The F-distribution

If U is a $\chi^2(r_1)$ random variable and V is a $\chi^2(r_2)$ random variable, and U and V are independent, then

$$F = \frac{U/r_1}{V/r_2}.$$

is a random variable having an F**-distribution** with r_1 and r_2 degrees of freedom. Its pdf is

$$f_F(x) = \begin{cases} \frac{\Gamma[(r_1+r_2)/2](r_1/r_2)^{r_1/2}}{\Gamma(r_1/2)\Gamma(r_2/2)} \frac{x^{r_1/2-1}}{(1+r_1x/r_2)^{(r_1+r_2)/2}}, & 0 < x < \infty \\ 0 & \text{elsewhere} \end{cases}$$

6.4.3 Students Theorem

Theorem 6.4.2. *Let $X_1, \cdots, X_n$ be iid with each $X_i \sim N(\mu, \sigma^2)$. Define the random variables*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{and} \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Then

(a) $\overline{X}$ has a $N(\mu, \sigma^2/n)$ distribution.

(b) $\overline{X}$ and S^2 are independent.

(c) $(n-1)S^2/\sigma^2$ has a $\chi^2(n-1)$ distribution.

(d) The random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a Student t-distribution with $n-1$ degrees of freedom.

Proof. (a) It can be easily seen by Thm 6.1.4.

(b) Let $\boldsymbol{X} = (X_1, \cdots, X_n)'$. Define the random vector: $\boldsymbol{Y} = (X_1 - \overline{X}, \cdots, X_n - \overline{X})'$. Consider the following transformation:

$$\boldsymbol{W} = \begin{bmatrix} \overline{X} \\ \boldsymbol{Y} \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \cdot \mathbf{1}_{1\times n} \\ \boldsymbol{I} - \frac{1}{n}\mathbf{1}_{n\times n} \end{bmatrix} \boldsymbol{X},$$

Because $\boldsymbol{W}$ is a linear transformation of multivariate normal random vector, by Thm 6.2.1, it has a multivariate normal distribution with mean

$$E[\boldsymbol{W}] = \begin{bmatrix} \mu \\ \mathbf{0}_{n\times 1} \end{bmatrix}, \quad Var[\boldsymbol{W}] = \sigma^2 \begin{bmatrix} \frac{1}{n} & \mathbf{0}_{1\times n} \\ \mathbf{0}_{n\times 1} & \boldsymbol{I} - \frac{1}{n}\mathbf{1}_{n\times n} \end{bmatrix}$$

Because the covariances are 0, $\overline{X}$ is independent of $\boldsymbol{Y}$. Since $S^2 = (n-1)^{-1}\boldsymbol{Y}'\boldsymbol{Y}$. Hence, by Thm 6.2.5, $\overline{X}$ is independent of S^2.

(c) We can prove that $\sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \frac{(\mu-\overline{X})^2}{\sigma^2/n}$ because

$$\begin{aligned}
(n-1)S^2/\sigma^2 &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu+\mu-\overline{X})^2 \\
&= \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} + 2\sum_{i=1}^{n}\frac{(X_i-\mu)(\mu-\overline{X})}{\sigma^2} + \frac{n(\mu-\overline{X})^2}{\sigma^2} \\
&= \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} + 2(\mu-\overline{X})\sum_{i=1}^{n}\frac{(X_i-\mu)}{\sigma^2} + \frac{n(\mu-\overline{X})^2}{\sigma^2} \\
&= \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} - 2(\overline{X}-\mu)\frac{n(\overline{X}-\mu)}{\sigma^2} + \frac{n(\mu-\overline{X})^2}{\sigma^2} \\
&= \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} - \frac{n(\mu-\overline{X})^2}{\sigma^2}
\end{aligned}$$

Since the mgfs of a $\chi^2(n)$ distribution and a $\chi^2(1)$ distribution are:

$$M_{\chi^2(n)}(t) = \frac{1}{(1-2t)^{n/2}}, \quad M_{\chi^2(1)}(t) = \frac{1}{(1-2t)^{1/2}}.$$

Due to independency of $\overline{X}$ and S^2 in (b), the mgf of $\frac{(n-1)S^2}{\sigma^2}$ should be $M_{\chi^2(n)}(t)/M_{\chi^2(1)}(t) = (1-2t)^{-\frac{n-1}{2}}$. Thus we can see $\frac{(n-1)S^2}{\sigma^2}$ has a $\chi^2(n-1)$ distribution.

(d) Finally, part (d) follows immediately from parts (a)-(c) upon writing T as

$$T = \frac{(\overline{X}-\mu)/(\sigma/\sqrt{n})}{\sqrt{S^2/\sigma^2}}.$$

$\square$

6.5 Summary of distributions

○ *Discrete random variables*

1. **Bernoulli with** θ

    - $\{X|\ 1 \text{ for success}, 0 \text{ for failure}\}$, with $P(\text{success}) = \theta$.
    - $p_X(x) = \begin{cases} \theta, & x = 1 \\ 1-\theta, & x = 0, \end{cases}$ and zero elsewhere.
    - $E(X) = \theta$
    - $Var(X) = \theta(1-\theta)$
    - $M_X(t) = 1 - \theta + \theta e^t$

2. **Binomial** (n, θ)

    - X: The # of successes in n independent Bernoulli trials(IBT), with $P(\text{success}) = \theta$.
    - $p_X(x) = C_n^x\ \theta^x(1-\theta)^{n-x}, x = 0, 1, \cdots, n$ and zero elsewhere.
    - $E(X) = n\theta$
    - $Var(X) = n\theta(1-\theta)$
    - $M_X(t) = \big(1 - \theta + \theta e^t\big)^n$

3. **Multinomial** $(n, x_i's, \theta_i's)$**,** $i = 1, \cdots, k-1$

    - X_i: The # of outcomes from C_i, with $P(C_i) = \theta_i$, where $\displaystyle\sum_{i=1}^k x_i = n$ and $\displaystyle\sum_{i=1}^k \theta_i = 1$.
    - $p(\mathbf{x}) = \dfrac{n!}{x_1! \cdots x_{k-1}! x_k!}\ \theta_1^{x_1} \cdots \theta_{k-1}^{x_{k-1}} \theta_k^{x_k}$, for $x_i \in \{0, 1, \cdots\}$, and zero elsewhere.
    - $M(t_1, \cdots, t_{k-1}) = \big(\theta_1 e^{t_1} + \cdots + \theta_{k-1} e^{t_{k-1}} + \theta_k\big)^n$

4. **Geometric** (θ) (Different)

    - X: The # of trials when the first success appears in a sequence of IBTs, $P(\text{success}) = \theta$.
    - $p_X(x) = \theta(1-\theta)^{x-1}, x = 1, 2, \cdots,$ and zero elsewhere.
    - $E(X) = 1/\theta$
    - $Var(X) = (1-\theta)/\theta^2$
    - $M_X(t) = \dfrac{\theta e^t}{1-(1-\theta)e^t}$, for $t < -\ln(1-\theta)$

5. **Negative Binomial** (r, θ) (Different)

- X: The # of trials when the rth success in a sequence of IBTs, $P(\text{success}) = \theta$.
- $p_X(x) = C_{x-1}^{r-1}\theta^r(1-\theta)^{x-r}, x = r, r+1, \cdots$, and zero elsewhere.
- $E(X) = r/\theta$
- $Var(X) = r(1-\theta)/\theta^2$
- $M_X(t) = \left[\dfrac{\theta e^t}{1-(1-\theta)e^t}\right]^r, \qquad$ for $t < -\ln(1-\theta)$

6. **Hypergeometric** (r, θ)

- X: The # of red in n draws without replacement, from an urn of N balls of which D are red.
- $p_X(x) = \dfrac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}$, $x = 0, 1, \cdots, n$, and zero elsewhere.
- $E(X) = n\frac{D}{N}$
- $Var(X) = n\frac{D}{N}\frac{N-D}{N}\frac{N-n}{N-1}$

7. **Poisson** (m)

- X: The # of events occur within an interval.

  m: Average occurrence during that interval.
- $p_X(x) = \dfrac{m^x e^{-m}}{x!}, x = 0, 1, 2, \cdots$, and zero elsewhere.
- $E(X) = m$
- $Var(X) = m$
- $M_X(t) = e^{m(e^t-1)}$

○ *Continuous random variables*

1. **Gamma** (α, β)

- $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, \quad x > 0$ and zero elsewhere.
- $E(X) = \alpha/\beta$
- $Var(X) = \alpha/\beta^2$
- $M_X(t) = \dfrac{1}{(1-t/\beta)^\alpha}$

2. **Exponential** (λ) = **Gamma** $(1, \lambda)$

- X: The waiting time until the next event happens in a Poisson process, with the rate of events' occurance λ.

- $f_X(x) = \lambda e^{-\lambda x}, \quad x > 0$ and zero elsewhere.
- $E(X) = 1/\lambda$
- $Var(X) = 1/\lambda^2$
- $M_X(t) = \dfrac{\lambda}{(\lambda - t)}$

3. $\chi^2(r) =$ **Gamma** $(\frac{r}{2}, \frac{1}{2})$ **with** $df = r$

   - $f(x) = \dfrac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} e^{-x/2}, \quad x \geq 0$, and zero elsewhere.
   - $E(X) = r$
   - $Var(X) = 2r$
   - $M_X(t) = \dfrac{1}{(1-2t)^{\frac{r}{2}}}$
   - $X = X_1^2 + X_2^2 + \cdots + X_r^2 \sim \chi^2(r)$ if each X_i is iid with $N(0,1)$.

4. **Normal** $N(\mu, \sigma^2)$

   - $f(x) = \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right\}, \quad$ for $-\infty < x < \infty$.
   - $E(X) = \mu$
   - $Var(X) = \sigma^2$
   - $M_X(t) = \exp\left\{\mu t + \sigma^2 \dfrac{t^2}{2}\right\}$

5. **Standard normal** $N(0,1)$

   - $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$ iff. $X \sim N(\mu, \sigma^2)$
   - $E(Z) = 0$
   - $Var(Z) = 1$
   - $M_Z(t) = \exp\left\{\dfrac{t^2}{2}\right\}$

6. **Beta distribution**

   - $Y = \frac{X_1}{X_1+X_2}$, where $X_1 \sim$ Gamma $(\alpha, 1)$ and $X_2 \sim$ Gamma $(\beta, 1)$

7. t**-distribution with** $df = r$

   - $T = \dfrac{Z}{\sqrt{V/r}}, \quad$ if $Z \sim N(0,1)$ and $V \sim \chi^2(r)$, are independent
   - Student's Theorem

8. F**-distribution with** $df = (r_1, r_2)$

   - $F = \dfrac{U/r_1}{V/r_2}, \quad$ if $U \sim \chi^2(r_1)$ and $V \sim \chi^2(r_2)$ are independent.