Probability 2 - Notes 11

The bivariate and multivariate normal distribution.

Definition. Two r.v.'s (X,Y) have a bivariate normal distribution $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ if their joint p.d.f. is

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} e^{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]} \qquad (1)$$

for all x,y. The parameters μ_1,μ_2 may be any real numbers, $\sigma_1 > 0$, $\sigma_2 > 0$, and $-1 \le \rho \le 1$.

It is convenient to rewrite (1) in the form

$$f_{X,Y}(x,y) = ce^{-\frac{1}{2}Q(x,y)}, \text{ where } c = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \text{ and}$$

$$Q(x,y) = (1-\rho^2)^{-1}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right] \qquad (2)$$

Statement. *The marginal distributions of $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ are normal with r.v.'s X and Y having density functions*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, \quad f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}.$$

Proof. The expression (2) for $Q(x,y)$ can be rearranged as follows:

$$Q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1} - \rho\frac{y-\mu_2}{\sigma_2}\right)^2 + (1-\rho^2)\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right] = \frac{(x-a)^2}{(1-\rho^2)\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2}, \qquad (3)$$

where $a = a(y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)$. Hence

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = ce^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} \times \int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2(1-\rho^2)\sigma_1^2}}dx = \frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}},$$

where the last step makes use of the formula $\int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}}dx = \sqrt{2\pi}\sigma$ with $\sigma = \sigma_1\sqrt{1-\rho^2}$. □

Exercise. Derive the formula for $f_X(x)$.

Corollaries.

1. Since $X \sim N(\mu_1,\sigma_1^2)$, $Y \sim N(\mu_2,\sigma_2^2)$, we know the meaning of four parameters involved into the definition of the normal distribution, namely

$$E(X) = \mu_1, \quad Var(X) = \sigma_1^2, \quad E(Y) = \mu_2, \quad Var(X) = \sigma_2^2.$$

2. $X|(Y=y)$ **is a normal r.v.** To verify this statement we substitute the necessary ingredients into the formula defining the relevant conditional density:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_1}e^{-\frac{(x-a(y))^2}{2\sigma_1^2(1-\rho^2)}}.$$

In other words, $X|(Y=y) \sim N(a(y),(1-\rho^2)\sigma_1^2)$. Hence:

3. $E(X|Y=y)=a(y)$ or, equivalently, $E(X|Y)=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(Y-\mu_2)$. In particular, we see that $E(X|Y)$ is a linear function of Y.

4. $E(XY)=\sigma_1\sigma_2\rho+\mu_1\mu_2$.

Proof. $E(XY)=E[E(XY|Y)]=E[YE(X|Y)]=E[Y(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(Y-\mu_2)]=\mu_1E(Y)+$ $\rho\frac{\sigma_1}{\sigma_2}[E(Y^2)-\mu_2E(Y)]=\mu_1\mu_2+\rho\frac{\sigma_1}{\sigma_2}[E(Y^2)-\mu_2^2]=\mu_1\mu_2+\rho\frac{\sigma_1}{\sigma_2}Var(Y)=\sigma_1\sigma_2\rho+\mu_1\mu_2$. □

5. $Cov(X,Y)=\sigma_1\sigma_2\rho$. This follows from Corollary 4 and the formula $Cov(X,Y)=E(XY)-E(X)E(X)$.

6. $\rho(X,Y)=\rho$. In words: ρ is the correlation coefficient of X, Y. This is now obvious from the definition $\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)var(Y)}}$.

Exercise. Show that X and Y are independent iff $\rho=0$. (We proved this in the lecture; it is easily seen from either the joint p.d.f.)

Remark. It is possible to show that the m.g.f. of X,Y is

$$M_{X,Y}(t_1,t_2)=e^{(\mu_1t_1+\mu_2t_2)+\frac{1}{2}(\sigma_1^2t_1^2+2\rho\sigma_1\sigma_2t_1t_2+\sigma_2^2t_2^2)}$$

Many of the above statements follow from it. (To actually do this is a very useful exercise.)

The Multivariate Normal Distribution.

Using vector and matrix notation. To study the joint normal distributions of more than two r.v.'s, it is convenient to use vectors and matrices. But let us first introduce these notations for the case of two normal r.v.'s X_1,X_2. We set

$$\mathbf{X}=\begin{pmatrix}X_1\\X_2\end{pmatrix};\ \mathbf{x}=\begin{pmatrix}x_1\\x_2\end{pmatrix};\ \mathbf{t}=\begin{pmatrix}t_1\\t_2\end{pmatrix};\ \mathbf{m}=\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix};\ \mathbf{V}=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$$

Then $\mathbf{m}$ is the vector of means and $\mathbf{V}$ is the variance-covariance matrix. Note that $|\mathbf{V}|=\sigma_1^2\sigma_2^2(1-\rho^2)$ and

$$\mathbf{V}^{-1}=\frac{1}{(1-\rho^2)}\begin{pmatrix}\frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2}\\\frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2}\end{pmatrix}$$

Hence $f_{\mathbf{X}}(\mathbf{x})=\frac{1}{(2\pi)^{2/2}|\mathbf{V}|^{1/2}}e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m})}$ for all $\mathbf{x}$. Also $M_{\mathbf{X}}(\mathbf{t})=e^{\mathbf{t}^T\mathbf{m}+\frac{1}{2}\mathbf{t}^T\mathbf{V}\mathbf{t}}$.

We again use matrix and vector notation, but now there are n random variables so that $\mathbf{X}$, $\mathbf{x}$, $\mathbf{t}$ and $\mathbf{m}$ are now n-vectors with i^{th} entries X_i, x_i, t_i and μ_i and $\mathbf{V}$ is the $n\times n$ matrix with ii^{th} entry σ_i^2 and ij^{th} entry (for $i\neq j$) σ_{ij}. Note that $\mathbf{V}$ is symmetric so that $\mathbf{V}^T=\mathbf{V}$.

The joint p.d.f. is $f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{V}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m})}$ for all $\mathbf{x}$. We say that $\mathbf{X} \sim N(\mathbf{m},\mathbf{V})$.

We can find the joint m.g.f. quite easily.

$$M_{\mathbf{X}}(\mathbf{t}) = E\left[e^{\sum_{j=1}^n t_j X_j}\right] = E[e^{\mathbf{t}^T\mathbf{X}}] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2}|\mathbf{V}|^{1/2}} e^{-\frac{1}{2}\left((\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m})-2\mathbf{t}^T\mathbf{x}\right)} dx_1 \ldots dx_n$$

We do the equivalent of completing the square, i.e. we write

$$(\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}) - 2\mathbf{t}^T\mathbf{x} = (\mathbf{x}-\mathbf{m}-\mathbf{a})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}-\mathbf{a}) + b$$

for a suitable choice of the n-vector $\mathbf{a}$ of constants and a constant b. Then

$$M_{\mathbf{X}}(\mathbf{t}) = e^{-b/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2}|\mathbf{V}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}-\mathbf{a})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}-\mathbf{a})} dx_1 \ldots dx_n = e^{-b/2}.$$

We just need to find $\mathbf{a}$ and b. Expanding we have

$$\begin{aligned} & ((\mathbf{x}-\mathbf{m})-\mathbf{a})^T\mathbf{V}^{-1}((\mathbf{x}-\mathbf{m})-\mathbf{a}) + b \\ = \; & (\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}) - 2\mathbf{a}^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}) + \mathbf{a}^T\mathbf{V}^{-1}\mathbf{a} + b \\ = \; & (\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}) - 2\mathbf{a}^T\mathbf{V}^{-1}\mathbf{x} + \left[2\mathbf{a}^T\mathbf{V}^{-1}\mathbf{m} + \mathbf{a}^T\mathbf{V}^{-1}\mathbf{a} + b\right] \end{aligned}$$

This has to equal $(\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}) - 2\mathbf{t}^T\mathbf{x}$ for all $\mathbf{x}$. Hence we need $\mathbf{a}^T\mathbf{V}^{-1} = \mathbf{t}^T$ and $b = -\left[2\mathbf{a}^T\mathbf{V}^{-1}\mathbf{m} + \mathbf{a}^T\mathbf{V}^{-1}\mathbf{a}\right]$. Hence $\mathbf{a} = \mathbf{V}\mathbf{t}$ and $b = -\left[2\mathbf{t}^T\mathbf{m} + \mathbf{t}^T\mathbf{V}\mathbf{t}\right]$. Therefore

$$M_{\mathbf{X}}(\mathbf{t}) = e^{-b/2} = e^{\mathbf{t}^T\mathbf{m} + \frac{1}{2}\mathbf{t}^T\mathbf{V}\mathbf{t}}$$

Results obtained using the m.g.f.

1. Any (non-empty) subset of multivariate normals is multivariate normal. Simply put $t_j = 0$ for all j for which X_j is not in the subset. For example $M_{X_1}(t_1) = M_{X_1,\ldots,X_n}(t_1,0,\ldots,0) = e^{t_1\mu_1 + t_1^2\sigma_1^2/2}$. Hence $X_1 \sim N(\mu_1, \sigma_1^2)$. A similar result holds for X_i. This identifies the parameters μ_i and σ_i^2 as the mean and variance of X_i. Also

$$M_{X_1,X_2}(t_1,t_2) = M_{X_1,\ldots,X_n}(t_1,t_2,0,\ldots,0) = e^{t_1\mu_1 + t_2\mu_2 + \frac{1}{2}(t_1^2\sigma_1^2 + 2\sigma_{12}t_1t_2 + \sigma_2^2t_2^2)}$$

Hence X_1 and X_2 have bivariate normal distribution with $\sigma_{12} = Cov(X_1,X_2)$. A similar result holds for the joint distribution of X_i and X_j for $i \neq j$. This identifies $\mathbf{V}$ as the variance-covariance matrix for $X_1,\ldots,X_n$.

2. $\mathbf{X}$ is a vector of independent random variables iff $\mathbf{V}$ is diagonal (i.e. all off-diagonal entries are zero so that $\sigma_{ij} = 0$ for $i \neq j$).

Proof. From (1), if the $X's$ are independent then $\sigma_{ij} = Cov(X_i, X_j) = 0$ for all $i \neq j$, so that $\mathbf{V}$ is diagonal.

If $\mathbf{V}$ is diagonal then $\mathbf{t}^T\mathbf{V}\mathbf{t} = \sum_{j=1}^n \sigma_j^2 t_j^2$ and hence

$$M_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}^T\mathbf{m}+\frac{1}{2}\mathbf{t}^T\mathbf{V}\mathbf{t}} = \prod_{j=1}^n \left(e^{\mu_j t_j + \frac{1}{2}\sigma_j^2 t_j^2/2}\right) = \prod_{j=1}^n M_{X_j}(t_j)$$

By the uniqueness of the joint m.g.f., $X_1, \ldots, X_n$ are independent.

3. Linearly independent linear functions of multivariate normal random variables are multivariate normal random variables. If $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$, where $\mathbf{A}$ is an $n \times n$ non-singular matrix and $\mathbf{b}$ is a (column) n-vector of constants, then $\mathbf{Y} \sim N(\mathbf{Am} + \mathbf{b}, \mathbf{AVA}^T)$.

Proof. Use the joint m.g.f.

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= E[e^{\mathbf{t}^T\mathbf{Y}}] = E[e^{\mathbf{t}^T\mathbf{AX}+\mathbf{b}}] = e^{\mathbf{t}^T\mathbf{b}}E[e^{(\mathbf{A}^T\mathbf{t})^T\mathbf{X}}] = e^{\mathbf{t}^T\mathbf{b}}M_{\mathbf{X}}(\mathbf{A}^T\mathbf{t}) \\ &= e^{\mathbf{t}^T\mathbf{b}}e^{(\mathbf{A}^T\mathbf{t})^T\mathbf{m}+\frac{1}{2}(\mathbf{A}^T\mathbf{t})^T\mathbf{V}(\mathbf{A}^T\mathbf{t})} = e^{\mathbf{t}^T(\mathbf{Am}+\mathbf{b})+\frac{1}{2}\mathbf{t}^T(\mathbf{AVA}^T)\mathbf{t}} \end{aligned}$$

This is just the m.g.f. for the multivariate normal distribution with vector of means $\mathbf{Am} + \mathbf{b}$ and variance-covariance matrix $\mathbf{AVA}^T$. Hence, from the uniqueness of the joint m.g.f, $\mathbf{Y} \sim N(\mathbf{Am} + \mathbf{b}, \mathbf{AVA}^T)$.

Note that from (2) a subset of the $Y's$ is multivariate normal.

NOTE. The results concerning the vector of means and variance-covariance matrix for linear functions of random variables hold regardless of the joint distribution of $X_1, \ldots, X_n$.

We define the expectation of a vector of random variables $\mathbf{X}$, $E[\mathbf{X}]$ to be the vector of the expectations and the expectation of a matrix of random variables $\mathbf{Y}$, $E[\mathbf{Y}]$, to be the matrix of the expectations. Then the variance-covariance matrix of $\mathbf{X}$ is just $E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T]$.

The following results are easily obtained:

(i) Let $\mathbf{A}$ be an $m \times n$ matrix of constants, $\mathbf{B}$ be an $m \times k$ matrix of constants and $\mathbf{Y}$ be an $n \times k$ matrix of random variables. Then $E[\mathbf{AY} + \mathbf{B}] = \mathbf{A}E[\mathbf{Y}] + \mathbf{B}$.

Proof. The ij^{th} entry of $E[\mathbf{AY} + \mathbf{B}]$ is $E[\sum_{r=1}^n A_{ir}Y_{rj} + B_{ij}] = \sum_{r=1}^n A_{ir}E[Y_{rj}] + B_{ij}$, which is the ij^{th} entry of $\mathbf{A}E[\mathbf{Y}] + \mathbf{B}$. The result is then immediate.

(ii) Let $\mathbf{C}$ be a $k \times m$ matrix of constants and $\mathbf{Y}$ be an $n \times k$ matrix of random variables. Then $E[\mathbf{YC}] = E[\mathbf{Y}]\mathbf{C}$.

Proof. Just transpose the equation. The result then follows from (i).

Hence if $\mathbf{Z} = \mathbf{AX} + \mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix of constants, $\mathbf{b}$ is an m-vector of constants and $\mathbf{X}$ is an n-vector of random variables with $E[\mathbf{X}] = \mu$ and variance-covariance matrix $\mathbf{V}$, then

$$E[\mathbf{Z}] = E[\mathbf{AX} + \mathbf{b}] = \mathbf{A}E[\mathbf{X}] + \mathbf{b} = \mathbf{A}\mu + \mathbf{b}$$

Also the variance-covariance matrix for $\mathbf{Y}$ is just

$$E[(\mathbf{Y} - E[\mathbf{Y}])(\mathbf{Y} - E[\mathbf{Y}])^T] = E[\mathbf{A}(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T \mathbf{A}^T] = \mathbf{A}E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T]\mathbf{A}^T = \mathbf{AVA}^T$$

Example. Suppose that $E[X_1] = 1$, $E[X_2] = 0$, $Var(X_1) = 2$, $Var(X_2) = 4$ and $Cov(X_1, X_2) = 1$. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 + aX_2$. Find the means, variances and covariance and hence find a so that Y_1 and Y_2 are uncorrelated.

Writing in vector and matrix notation we have $E[\mathbf{Y}] = \mathbf{Am}$ and the variance-covariance matrix for $\mathbf{Y}$ is just $\mathbf{AVA}^T$ where

$$\mathbf{m} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \mathbf{V} = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} \quad \mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & a \end{pmatrix}$$

Therefore

$$\mathbf{Am} = \begin{pmatrix} 1 & 1 \\ 1 & a \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\mathbf{AVA}^T = \begin{pmatrix} 1 & 1 \\ 1 & a \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & a \end{pmatrix} = \begin{pmatrix} 8 & 3 + 5a \\ 3 + 5a & 2 + 2a + 4a^2 \end{pmatrix}$$

Hence Y_1 and Y_2 have means 1 and 1, variances 8 and $2 + 2a + 4a^2$ and covariance $3 + 5a$. They are therefore uncorrelated if $3 + 5a = 0$, i.e. if $a = -\frac{3}{5}$.