Prepared by:
Çağatay Candan

METU, Electrical Engin. Dept.
Dec. 2011
document version:1.2

On Gaussian Distribution

Gaussian distribution is defined as follows:

$$f_{\tilde{x}}(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\mu)^2}{2\sigma_x^2}}$$

The function $f_{\tilde{x}}(x)$ is clearly positive valued. Before calling this function as a probability density function, we should check whether the area under the curve is equal to 1 or not.

<u>R1.1: Area under Gaussian Distribution [Signal Analysis, Papoulis]</u>

Let $I = \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma_x^2}} dx$, then by a simple change of variables $I = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma_x^2}} dx$; then

$$\begin{aligned}
I^2 &= \left(\int_{-\infty}^{\infty} e^{-\frac{x_1^2}{2\sigma_x^2}} dx_1 \right)\left(\int_{-\infty}^{\infty} e^{-\frac{x_2^2}{2\sigma_x^2}} dx_2 \right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x_1^2}{2\sigma_x^2}} e^{-\frac{x_2^2}{2\sigma_x^2}} dx_1 dx_2 \\
&= \int_{0}^{\infty}\int_{-\pi}^{\pi} e^{-\frac{r^2}{2\sigma_x^2}} r dr d\phi \\
&= 2\pi\sigma_x^2
\end{aligned}$$

This shows that $f_{\tilde{x}}(x)$ is a valid probability density function.

<u>R1.2: Mean and Variance Calculation for Gaussian Distribution;</u>

Mean: It can be noted that the distribution is symmetric around $x = \mu_x$, that is $f_{\tilde{x}}(\mu_x + x) = f_{\tilde{x}}(\mu_x - x)$. We know that the mean value should be located at the center of symmetry then $\bar{x} = \mu_x$.

We can show this result as follows: By a simple change of variables, we get $\bar{x} = \int x f_{\tilde{x}}(x) dx = \int (\mu_x + x) f_{\tilde{x}}(\mu_x + x) dx = \int (\mu_x - x) f_{\tilde{x}}(\mu_x - x) dx$. But since we have

$f_{\underset{\sim}{x}}(\mu_x + x) = f_{\underset{\sim}{x}}(\mu_x - x)$, $\bar{x} = \int (\mu_x + x) f_{\underset{\sim}{x}}(\mu_x + x)dx = \int (\mu_x - x) f_{\underset{\sim}{x}}(\mu_x + x)dx$, then $2\bar{x} = \int (\mu_x + x) f_{\underset{\sim}{x}}(\mu_x + x)dx + \int (\mu_x - x) f_{\underset{\sim}{x}}(\mu_x + x)dx = 2\mu_x$.

Variance: This is more tricky, [Probability, Papoulis]. The area under the Gaussian distribution is equal to 1, that is $\int \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\mu)^2}{2\sigma_x^2}} dx = 1$. Now taking derivative of this relation wrt to μ, we get

$$\frac{d}{d\mu} \int \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\mu)^2}{2\sigma_x^2}} dx = 0$$

$$\frac{1}{\sqrt{2\pi\sigma_x^2}} \int \frac{(x-\mu)}{\sigma_x^2} e^{-\frac{(x-\mu)^2}{2\sigma_x^2}} dx = 0$$

The last equation can be written as $\int (x-\mu) f_{\underset{\sim}{x}}(x)dx = 0$, showing one more time that the mean of Gaussian distribution is μ. Taking derivative of $\int (x-\mu) f_{\underset{\sim}{x}}(x)dx = 0$ relation with respect to μ for a second time, we get $\int \frac{(x-\mu)^2}{\sigma_x^2} f_{\underset{\sim}{x}}(x)dx + \int (-1) f_{\underset{\sim}{x}}(x)dx = 0$, showing that the variance of Gaussian distribution is σ_x^2.

Taking derivative wrt to a "constant" (such as μ in here) is a powerful calculation/proof tool. This process can be done if the relation whose derivative is taken is valid for a continuum of μ values. In our case, the area under Gaussian distribution is 1 for any value of μ. Hence $I(\mu) = \int \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\mu)^2}{2\sigma_x^2}} dx = 1$ is valid for $\forall \mu$.

A second note is on the interchange of derivative an integral operators. This step looks innocent but it is indeed treacherous. We have to verify the interchange in general. If you have one sided or double sided integrals, (One sided integral has upper or lower limit as $\pm\infty$, double sided integrals have upper/lower limits as $-\infty$ and ∞.); the integrals may or may not converge. The interchange of integrals or derivatives operators is the interchange of two limiting operations. (Remember derivative is also defined at a limit of $h \to 0$ or $1/h \to \infty$). One has to be careful about this interchange. The integrals studied in the probability course in general behave nicely (positive and unit area functions). But we should always be careful when we exchange operators involving infinite or infinitesimal quantities.

R1.3: The distribution of y=ax +b where x is normal distributed:

The distribution of $\underset{\sim}{y} = a\underset{\sim}{x} + b$ can be expressed using the fundamental theorem for functions of random variables: $f_{\underset{\sim}{y}}(y) = \frac{1}{|a|} f_{\underset{\sim}{x}}(\frac{y-b}{a})$. When the definition for Gaussian density is substituted for $f_{\underset{\sim}{x}}(x)$ we get

$$f_{\underset{\sim}{y}}(y) = \frac{1}{|a|}\frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(\frac{y-b}{a}-\mu)^2}{2\sigma_x^2}} = \frac{1}{\sqrt{2\pi a^2\sigma_x^2}} e^{-\frac{(y-b-a\mu)^2}{2a^2\sigma_x^2}} .$$

From the last equation, we note that $\underset{\sim}{y}$ is also Gaussian distributed with mean $(a\mu_x + b)$ and variance $a^2\sigma_x^2$.

We reach an important conclusion that by scaling and biasing a Gaussian r.v., we get another Gaussian r.v. with different mean and variance.

The mean and variance of the new Gaussian random variable can be easily calculated using expectation operator, $\bar{y} = E\{y\} = E\{ax+b\} = a\mu_x + b$ (and a similar relation for variance). And the distribution of $\underset{\sim}{y}$ can be written by using the calculated mean and variance. We know that if a random variable is Gaussian distributed, we only need to calculate its mean and variance to write the distribution. As an example let $\underset{\sim}{z} : N(0,1)$ (Gaussian distributed or normal distributed with zero mean and unit variance) , then $\underset{\sim}{w} = 10\underset{\sim}{z} - 10$ is N(-10,100) since the mean of $10\underset{\sim}{z} - 10$ is -10 and its variance is $E\{(w-\overline{w})^2\} = E\{\left((10\underset{\sim}{z}-10)-(-10)\right)^2\} = E\{(10z)^2\} = 100E\{z^2\} = 100.$ s

R1.4: Moment Generating Function:

Let $\underset{\sim}{z}$ be zero mean, unit variance Gaussian distribution: $\underset{\sim}{z} : N(0,1)$. The moment generating function of $\underset{\sim}{z}$ is $\Phi(s) = E\{e^{sz}\} = \frac{1}{\sqrt{2\pi}}\int e^{sz} e^{-\frac{z^2}{2}} dz$. We can write *sz-z²/2* as a perfect square and a correction term as follows: $-(\frac{z^2}{2} - sz) = -(\frac{z^2-2sz}{2}) = -\frac{(z-s)^2}{2} + \frac{s^2}{2}$. This process is known as completing the square and frequently used. When this relation is substituted to $\Phi(s)$, we get

$\Phi(s)=\frac{1}{\sqrt{2\pi}}\int e^{sz}e^{-\frac{z^2}{2}}dz=e^{\frac{s^2}{2}}\frac{1}{\sqrt{2\pi}}\int e^{-\frac{(z-s)^2}{2}}dz=e^{\frac{s^2}{2}}\frac{1}{\sqrt{2\pi}}\int e^{-\frac{z^2}{2}}dz=e^{\frac{s^2}{2}}$. (The moment generating function is valid for all s values.)

When $\underset{\sim}{x}$ is a Gaussian random variable with mean μ and variance σ^2, $\underset{\sim}{x}:N(\mu,\sigma^2)$; then we know that $\underset{\sim}{x}$ can be written as $\underset{\sim}{x}=\sqrt{\sigma^2}\,\underset{\sim}{z}+\mu$ where $\underset{\sim}{z}:N(0,1)$. The moment generating function of $\underset{\sim}{x}$ is equal to $\Phi_x(s)=E\{e^{sx}\}=E\{e^{s(\sqrt{\sigma^2}\,\underset{\sim}{z}+\mu)}\}=e^{s\mu+\frac{s^2\sigma^2}{2}}$.

R1.5: Higher Order Moments of Gaussian Distribution:

Let $\underset{\sim}{x}:N(\mu,\sigma^2)$, we know that $\Phi_x(s)=e^{s\mu+\frac{s^2\sigma^2}{2}}$. Using power series expansion of e^z which is $e^z=\sum_{k=0}^{\infty}\frac{z^k}{k!}$, we can write $\Phi_x(s)$ as follows:

$$\Phi_x(s)=1+(s\mu+s^2\sigma^2/2)+\frac{(s\mu+s^2\sigma^2/2)^2}{2}+\frac{(s\mu+s^2\sigma^2/2)^3}{6}+\ldots$$

The coefficient of s^k in $\Phi_x(s)$ is equal to $E\{x^k\}/k!$. When $\underset{\sim}{x}:N(0,\sigma^2)$, we have

$$E\{x^k\}=m_k=\begin{cases}\frac{(2n)!\sigma^{2n}}{2^n n!}=1.3.5....(2n-1)\sigma^{2n} & k=2n\\ 0 & k=2n+1\end{cases}$$

R1.6: Addition of two independent Gaussian r.v.'s, z=x+y, is a Gaussian r.v.

When two Gaussian distributed independent random variables $\underset{\sim}{x}:N(\mu_x,\sigma_x^2)$ and $\underset{\sim}{y}:N(\mu_y,\sigma_y^2)$ are added, the resultant distribution is also Gaussian distributed $\underset{\sim}{z}:N(\mu_x+\mu_y,\sigma_x^2+\sigma_y^2)$, where $\underset{\sim}{z}=\underset{\sim}{x}+\underset{\sim}{y}$. . This is a property of great importance in practice.

This result can be easily shown using moment generating functions. $\Phi_z(s)=E\{e^{sz}\}=E\{e^{s(x+y)}\}=E\{e^{sz}\}E\{e^{sy}\}$ where we have used the independence of $\underset{\sim}{x}$ and $\underset{\sim}{y}$ in the last equality. From R1.4, we can substitute the moment generating function of Gaussian distribution and get

$\Phi_z(s) = e^{s\mu_x + \frac{s^2\sigma_x^2}{2}} e^{s\mu_y + \frac{s^2\sigma_y^2}{2}} = e^{s(\mu_x+\mu_y) + \frac{s^2(\sigma_x^2+\sigma_y^2)}{2}}$. Comparing this result with R1.4, we conclude that $\underset{\sim}{z} = \underset{\sim}{x} + \underset{\sim}{y}$ is indeed Gaussian distributed with mean $\mu_x + \mu_y$ and variance $\sigma_x^2 + \sigma_y^2$.

In the next section we define joint Gaussianity and show that arbitrary linear combination of Gaussian random variables result in Gaussian distribution.

Jointly Gaussian Random Variables

We first examine two random variables and then extend to random vectors which is the joint distribution of *n* random variables.

Two random variables $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are called jointly Gaussian if their joint density can be written in the following form:

$$f_{xy}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)\ (y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right)$$

The function $f_{xy}(x,y)$ is clearly positive valued and of finite area (since e^{-x^2} upper bounded by $e^{-|x|}$ for large x), therefore with proper scaling it can be utilized as density function. Note that $f_{xy}(x,y)$ is a function of two independent variables and has five parameters $\{\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho\}$. We will attach the labels of mean, variance and correlation coefficient to these variables after justifying why we do so.

An alternative (but equivalent) definition for joint Gaussianity can be given as follows: $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are called jointly Gaussian if $\underset{\sim}{z} = \varpi_x \underset{\sim}{x} + \varpi_y \underset{\sim}{y}$ is Gaussian distributed for any value of ϖ_x and ϖ_y. Here $\underset{\sim}{z}$ is the standard univariable Gaussian distribution that we have previously examined. (lso called univariate Gaussian) We take granted the equivalence of two definitions and proceed without a proof for now. (The function $g(\Theta_x, \Theta_y)$ and related descriptions given in R2.4 (given below) can be formulated as proof.)

R2.1: Definition in matrix form:

The definition can also be written as follows:

$$f_{xy}(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{\begin{bmatrix}(x-\mu_x) & (y-\mu_y)\end{bmatrix}\begin{bmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{bmatrix}^{-1}\begin{bmatrix}(x-\mu_x)\\(y-\mu_y)\end{bmatrix}}{2}\right)$$

Note that the argument of the exponential mimics the one dimension Gaussian distribution. If you try not to see the terms related to *y* variable then you have

$$\exp\left(\frac{\begin{bmatrix}(x-\mu_x) & \end{bmatrix}\begin{bmatrix}\sigma_x^2 & \\ & \end{bmatrix}^{-1}\begin{bmatrix}(x-\mu_x)\\ \end{bmatrix}}{2}\right)=\exp\left(-\frac{(x-\mu_x)^2}{2}\right)$$

From here we feel that multi-dimensional Gaussian distributions is linked to univariate (single variable) Gaussian distribution, which is a fact whose details are examined below.

The equivalency of matrix definition given in R2.1 to the original definition follows very simply by calculating the matrix inverse in the matrix definition:

$$\begin{bmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{bmatrix}^{-1}=\frac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)}\begin{bmatrix}\sigma_y^2 & -\rho\sigma_x\sigma_y\\ -\rho\sigma_x\sigma_y & \sigma_x^2\end{bmatrix}=\frac{1}{(1-\rho^2)}\begin{bmatrix}1/\sigma_x^2 & -\rho/\sigma_x\sigma_y\\ -\rho/\sigma_x\sigma_y & 1/\sigma_y^2\end{bmatrix}$$

R2.2: Definition for zero mean jointly distributed Gaussian r.v's.

When zero is substituted for the mean we get the following

$$f_{xy}(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{\begin{bmatrix}x & y\end{bmatrix}\begin{bmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{bmatrix}^{-1}\begin{bmatrix}x\\y\end{bmatrix}}{2}\right)$$

If new random variables $\underset{\sim}{x_2}$ and $\underset{\sim}{y_2}$ are defined from $\underset{\sim}{x}$ and $\underset{\sim}{y}$ where $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are jointly Gaussian r.v.'s with zero mean (the joint pdf is given above) such that

$$x_2 = x+\bar{x}$$
$$y_2 = y+\bar{y}$$

The distribution of new random variables is $f_{x_2y_2}(x,y)=f_{xy}(x-\bar{x},y-\bar{y})$ (by fundamental theorem) and we see that the distribution of $\underset{\sim}{x_2}$ and $\underset{\sim}{y_2}$ has the form of general joint Gaussian density. Since mean values only shift the center of symmetry to a point on (x,y) plane, sometimes we prefer to work with zero mean Gaussian distribution and say that this result is also valid for non-zero mean Gaussian distributions. This is due to the mentioned shift in the center of symmetry. We have some examples of this situation in these notes. s

R2.3: Two independent Gaussian r.v.'s are jointly Gaussian

Let $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are independent Gaussian distributed zero mean random variables. (As noted in R2.2, the value of the mean is not important in this argument. We prefer to take the mean values as zero to simply the presentation.) Then $f_{xy}(x,y) = f_x(x) f_y(y)$ and when the definition for the univariate Gaussian distribution is inserted, we get the definition of joint Gaussian distribution for zero mean variables. Since the joint density of two independent Gaussian random variables can be written in the form required by the the joint distribution of Gaussian densities, two independent Gaussian random variables are indeed jointly Gaussian. (This is not very surprising, the labels of the distributions give away the final conclusion.)

A more surprising result is the next one. If $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are known to be Gaussian distributed (but not independent), the joint distribution of $\underset{\sim}{x}$ and $\underset{\sim}{y}$ is *not* necessarily is Gaussian. That is if marginal densities of $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are Gaussian, the joint density of $\underset{\sim}{x}$ and $\underset{\sim}{y}$ is not necessarily Gaussian.
A simple example is as follows. Let $\underset{\sim}{x}$: N(0,1) and $\underset{\sim}{y} = \underset{\sim}{c}\,\underset{\sim}{x}$. Here $\underset{\sim}{c}$ can take the value of 1 and –1 with equal probability. It is not very difficult to show that (for EE230 students) that $\underset{\sim}{y}$ is also N(0,1). Clearly $\underset{\sim}{y}$ and $\underset{\sim}{x}$ are related to each other through a random mapping $\underset{\sim}{c}$. Therefore $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are not independent. Furthermore the condition density of $\underset{\sim}{y}$ given $\underset{\sim}{x}$ is $0.5\delta(y-1)+0.5\delta(y+1)$. We show in R2.8 that if $\underset{\sim}{y}$ and $\underset{\sim}{x}$ are jointly Gaussian distributed, then $\underset{\sim}{y}$ given $\underset{\sim}{x}$ is always an univariate Gaussian distribution. Therefore the distribution for $\underset{\sim}{y}$ given $\underset{\sim}{x}$ which is $0.5\delta(y-1)+0.5\delta(y+1)$ gives away that $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are <u>not</u> jointly Gaussian. For a constructive example see [Probability, Papoulis].

R2.4: Moment Generating Function of joint Gaussian r.v.'s x and y (zero mean)

Independent Case: $\Phi(s_x, s_y) = E\{e^{s_x x + s_y y}\} = E\{e^{\begin{bmatrix} s_x & s_y \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}}\}$ is equal to $E\{e^{s_x x}\}E\{e^{s_y y}\}$ by independence. Then

Prepared by: METU, Electrical Engin. Dept.
Çağatay Candan Dec. 2011
document version:1.2

$$\Phi(s_x, s_y) = e^{s_x \bar{x} + \frac{s_x^{\,2}\sigma_x^{\,2}}{2}} e^{s_y \bar{y} + \frac{s_y^{\,2}\sigma_y^{\,2}}{2}} = e^{s_x \bar{x} + s_y \bar{y}} e^{\frac{s_x^2 \sigma_x^{\,2} + s_y^2 \sigma_y^{\,2}}{2}}$$

Dependent Case: Let $\underset{\sim}{x}'$ and $\underset{\sim}{y}'$ be independent *zero mean* Gaussian random variables. Lets define the random variables $\underset{\sim}{x}$ and $\underset{\sim}{y}$ from $\underset{\sim}{x}'$ and $\underset{\sim}{y}'$ as follows:

$$\begin{aligned} \underset{\sim}{x} &= a\,\underset{\sim}{x}' \\ \underset{\sim}{y} &= c\underset{\sim}{x}' + d\,\underset{\sim}{y}' \end{aligned} \tag{1}$$

and lets define an arbitrary looking function $g(\Theta_x, \Theta_y)$ as follows:

$$\begin{aligned} g(\Theta_x, \Theta_y) &= E\{\exp(\Theta_x x + \Theta_y y)\} \\ &= E\{\exp(\Theta_x a x' + \Theta_y (c x' + d y'))\} \\ &= E\{\exp((\Theta_x a + \Theta_y c) x' + \Theta_y d y')\} \\ &= E\{\exp((\Theta_x a + \Theta_y c) x')\} E\{\exp(\Theta_y d y')\} \\ &= \Phi_{x'}(\Theta_x a + \Theta_y c) \Phi_{y'}(\Theta_y d) \end{aligned}$$

Here Θ_x, Θ_y are some scalars. When the moment generating functions for $\underset{\sim}{x}'$ and $\underset{\sim}{y}'$ is inserted in the last equation, we get:

$$\begin{aligned} g(\Theta_x, \Theta_y) &= \Phi_{x'}(\Theta_x a + \Theta_y c) \Phi_{y'}(\Theta_y d) \\ &= e^{\frac{1}{2}(\Theta_x a + \Theta_y c)^2} e^{\frac{1}{2}(\Theta_y d)^2} \\ &= e^{\frac{1}{2}(\Theta_x^2 a^2 + 2\Theta_x \Theta_y ac + \Theta_y^2 (c^2 + d^2))} \end{aligned}$$

Now we interpret $g(\Theta_x, \Theta_y)$ and establish its connection with $\Phi(s_x, s_y)$. The moment generating function $\Phi(s_x, s_y) = E\{e^{s_x x + s_y y}\}$ is defined for s_x and s_y values for which the expectation result is finite. For univariate Gaussian distributions such as $\underset{\sim}{x}'$, $\Phi_{x'}(s_{x'})$ function is defined for all complex $s_{x'}$ values. The same is also valid for $\underset{\sim}{y}'$ variable. We can then say that $g(\Theta_x, \Theta_y)$ function is defined for arbitrary complex valued pairs of (Θ_x, Θ_y).

Since the parameters of $g(\Theta_x, \Theta_y)$ function can be arbitrary, we choose to relabel them as $\Theta_x = s_x$ and $\Theta_y = s_y$ to get:

$$g(\Theta_x,\Theta_y)\downarrow_{\substack{\Theta_x=s_x\\ \Theta_y=s_y}} = E\{\exp(s_x x + s_y y)\} = \Phi(s_x,s_y)$$

and we get the following for the moment generating function of $\underset{\sim}{x}$ and $\underset{\sim}{y}$.

$$\Phi(s_x,s_y) = e^{\frac{1}{2}(s_x^2 a^2 + 2 s_x s_y ac + s_y^2 (c^2+d^2))} \tag{2}$$

Next we use the fundamental theorem to express the distribution of $\underset{\sim}{x}$ and $\underset{\sim}{y}$. We can also invert $\Phi(s_x,s_y)$ given in (2) to get the distribution of $\underset{\sim}{x}$ and $\underset{\sim}{y}$, but we prefer using the fundamental theorem instead of inverse Laplace transforms.

Rewriting equation (1) in the matrix form, we get the following:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a & 0 \\ c & d \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} \Rightarrow \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1/a & 0 \\ -c/ad & 1/d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

Then

$$f_{xy}(x,y) = \frac{1}{|J|} f_{x'y'}\left(\frac{x}{a}, \frac{1}{d}y - \frac{c}{ad}x\right) = A\exp\left(-\frac{1}{2}\left(\frac{x}{a}\right)^2\right)\exp\left(-\frac{1}{2}\left(\frac{1}{d}y - \frac{c}{ad}x\right)^2\right) \tag{3}$$

Here J is the Jacobian (which is $J=ad$) and A is a scalar whose exact value is not of interest for now.

Finally we fix (a,c,d) given in (1) to some special values.

$$a = \sigma_x$$
$$c = \rho\sigma_y$$
$$d = \sigma_y\sqrt{(1-\rho^2)}$$

Then (2) and (3) become

$$\Phi(s_x,s_y) = e^{\frac{1}{2}(s_x^2\sigma_x^2 + 2 s_x s_y \rho\sigma_x\sigma_y + s_y^2\sigma_y^2)} \tag{2'}$$

$$f_{xy}(x,y) = A\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{x^2}{\sigma_x^2} - 2\rho\frac{xy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2}\right]\right) \tag{3'}$$

From (2') and (3'), we get the moment generating function of jointly distributed Gaussian zero mean random variables $\underset{\sim}{x}$ and $\underset{\sim}{y}$.

<u>R2.5: Moment Generating Function of joint Gaussian r.v.'s x and y (non-zero mean)</u>

Let $\underset{\sim}{x}'$ and $\underset{\sim}{y}'$ be jointly distributed zero mean Gaussian random variables. The moment generating function of $\underset{\sim}{x}'$ and $\underset{\sim}{y}'$ is given in R2.4 equation (2'). Define new random variables as

$$\begin{aligned} x &= x' + \bar{x} \\ y &= y' + \bar{y} \end{aligned}$$

The moment generating function of $\underset{\sim}{x}$ and $\underset{\sim}{y}$ can be written as follows

$$\Phi(s_x, s_y) = E\{e^{s_x x + s_y y}\} = E\{e^{s_x(x'+\bar{x}) + s_y(y'+\bar{y})}\} = e^{s_x\bar{x} + s_y\bar{y}}\Phi_{x'y'}(s_x, s_y)$$

which is

$$\Phi(s_x, s_y) = e^{s_x\bar{x} + s_y\bar{y}} e^{\frac{1}{2}(s_x^2\sigma_x^2 + 2s_x s_y \rho\sigma_x\sigma_y + s_y^2\sigma_y^2)}$$

Note that when $\rho = 0$ the moment generating function given in the last equation reduces to the one found for the independent variables in R2.4.

<u>R2.6: Correlation coefficient between x and y is ρ.</u>

We know that the correlation coefficient between random variables $\underset{\sim}{x}$ and $\underset{\sim}{y}$ by its definition is not effected by the mean values of $\underset{\sim}{x}$ and $\underset{\sim}{y}$. We also know that $E\{xy\} = \frac{\partial^2}{\partial x \partial y}\Phi(s_x, s_y)\downarrow_{s_x = s_y = 0}$. We take the moment generating function of joint Gaussian random variables (with zero mean) and calculate its partial derivative as follows:

$$\frac{\partial^2}{\partial x \partial y}\Phi(s_x,s_y)=\frac{\partial^2}{\partial x \partial y}e^{\frac{1}{2}(s_x^2\sigma_x^2+2s_xs_y\rho\sigma_x\sigma_y+s_y^2\sigma_y^2)}$$

$$=\frac{\partial}{\partial y}(s_x\sigma_x^2+s_y\rho\sigma_x\sigma_y)e^{\frac{1}{2}(s_x^2\sigma_x^2+2s_xs_y\rho\sigma_x\sigma_y+s_y^2\sigma_y^2)}$$

$$=\rho\sigma_x\sigma_y e^{\frac{1}{2}(s_x^2\sigma_x^2+2s_xs_y\rho\sigma_x\sigma_y+s_y^2\sigma_y^2)}+(s_y\sigma_y^2+s_x\rho\sigma_x\sigma_y)e^{\frac{1}{2}(s_x^2\sigma_x^2+2s_xs_y\rho\sigma_x\sigma_y+s_y^2\sigma_y^2)}$$

When zero is substituted for s_x and s_y, we get $E\{xy\}=\rho\sigma_x\sigma_y$ and from here we see that $\rho=\frac{E\{xy\}}{\sigma_x\sigma_y}$. Note that right hand side of the last relation is the definition of correlation coefficient for zero mean random variables. Therefore ρ appearing in the definition of the joint density is the actual value of correlation coefficient between $\underset{\sim}{x}$ and $\underset{\sim}{y}$. We call this parameter as the correlation coefficient.

R2.7: If the correlation coefficient of jointly Gaussian distributed r.v.'s is zero, then random variables are independent.

In R2.5 we have noted that "Note that when $\rho=0$ the moment generating function given in the last equation reduces to the one found for the independent variables in R2.4.", we have also shown that ρ the correlation coefficient between $\underset{\sim}{x}$ and $\underset{\sim}{y}$ in R2.5. This important result follows from the combination of R2.4 and R2.5.

R2.7: The marginal density for x and y is Gaussian.

The moment generating function for the marginal density for $\underset{\sim}{x}$ is $E\{e^{s_x x}\}$. The same function can be expressed from the moment generating function of the joint distribution $\Phi(s_x,s_y)=E\{e^{s_x x+s_y y}\}=e^{s_x\bar{x}+s_y\bar{y}}e^{\frac{1}{2}(s_x^2\sigma_x^2+2s_xs_y\rho\sigma_x\sigma_y+s_y^2\sigma_y^2)}$ by taking $s_y=0$.

When this is done, we get $E\{e^{s_x x}\}=e^{s_x\bar{x}}e^{\frac{1}{2}s_x^2\sigma_x^2}$ which is the moment generating function of univariate Gaussian distribution with mean $\bar{x}$ and variance σ_x^2. Hence when we marginalize a joint Gaussian density, then we have an univariate Gaussian density.

R2.8: The conditional density of x given y for jointly Gaussian variables.

The density of $\underset{\sim}{x}$ conditioned on $\underset{\sim}{y}$ is also Gaussian. This result has significant importance in estimation theory.

Prepared by:
Çağatay Candan

METU, Electrical Engin. Dept.
Dec. 2011
document version:1.2

i. Case of scalar random variables.

Here $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are jointly Gaussian distributed with the following density.

$$f_{xy}(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}-2\rho\frac{(x-\mu_x)\ (y-\mu_y)}{\sigma_x\sigma_y}+\frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right)$$

The goal is to find $f_{y|x}(y\,|\,x)$, $f_{y|x}(y\,|\,x)=\dfrac{f_{xy}(x,y)}{f_x(x)}$. We explicitly calculate this ratio and show that the conditional density is the Gaussian distribution. Without any loss of generality, we assume $\mu_x=\mu_y=0$ to simplify the algebra. If $\mu_x=\mu_y=0$ is not true, then the following substitutions should be made in all expressions, $x\rightarrow(x-\mu_x)$ and $y\rightarrow(y-\mu_y)$.

Without a further ado, we start the calculation:

$$f_{y|x}(y\,|\,x)=\frac{f_{xy}(x,y)}{f_x(x)}=\frac{\dfrac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left(-\dfrac{1}{2(1-\rho^2)}\left[\dfrac{x^2}{\sigma_x^2}-2\rho\dfrac{xy}{\sigma_x\sigma_y}+\dfrac{y^2}{\sigma_y^2}\right]\right)}{\dfrac{1}{\sqrt{2\pi}\sigma_x}\exp\left(-\dfrac{x^2}{2\sigma_x^2}\right)}$$

Let's re-express the exponent in the numerator $\dfrac{1}{2(1-\rho^2)}\left[\dfrac{x^2}{\sigma_x^2}-2\rho\dfrac{xy}{\sigma_x\sigma_y}+\dfrac{y^2}{\sigma_y^2}\right]$:

$$\begin{aligned}
\frac{1}{2(1-\rho^2)}\left[\frac{x^2}{\sigma_x^2}-2\rho\frac{xy}{\sigma_x\sigma_y}+\frac{y^2}{\sigma_y^2}\right]&=\frac{1}{2(1-\rho^2)\sigma_y^2}\left[\frac{\sigma_y^2}{\sigma_x^2}x^2-2\rho\frac{\sigma_y xy}{\sigma_x}+y^2\right]\\
&=\frac{1}{2(1-\rho^2)\sigma_y^2}\left[y^2-2\rho\frac{\sigma_y xy}{\sigma_x}+\left\{x\rho\frac{\sigma_y}{\sigma_x}\right\}^2-\left\{x\rho\frac{\sigma_y}{\sigma_x}\right\}^2+\frac{\sigma_y^2}{\sigma_x^2}x^2\right]\\
&=\frac{1}{2(1-\rho^2)\sigma_y^2}\left[\left(y-\rho\frac{\sigma_y}{\sigma_x}x\right)^2-\left\{x\rho\frac{\sigma_y}{\sigma_x}\right\}^2+\frac{\sigma_y^2}{\sigma_x^2}x^2\right]\\
&=\frac{1}{2(1-\rho^2)\sigma_y^2}\left(y-\rho\frac{\sigma_y}{\sigma_x}x\right)^2+\frac{x^2}{2\sigma_x^2}
\end{aligned}$$

Note that in the second line we have added and subtracted the same term shown with curly brackets, i.e. $\{...\}^2$. This process is called completion to a square. Now $f_{y|x}(y\,|\,x)$ can be easily written as follows:

$$f_{y|x}(y\,|\,x)=\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)\sigma_y^2}\left(y-\rho\frac{\sigma_y}{\sigma_x}x\right)^2\right)$$

Hence the conditional distribution is $N\left(\rho\frac{\sigma_y}{\sigma_x}x,(1-\rho^2)\sigma_y^2\right)$.

ii. Case of Multiple Observations (Multivariate Gaussians)

Here $\underset{\sim}{y}$ and $\underset{\sim}{\mathrm{x}}$ is jointly Gaussian distributed. In this case, $\underset{\sim}{y}$ is a scalar Gaussian random variable and $\underset{\sim}{\mathrm{x}}$ is a N x 1 Gaussian vector. As in the previous case, both $\underset{\sim}{y}$ and $\underset{\sim}{\mathrm{x}}$ are assumed to be zero mean without any loss of generality. The goal is to express $f_{y|x}(y\,|\,x)=\frac{f_{xy}(x,y)}{f_x(x)}$, that is "update" the a-priori pdf for $\underset{\sim}{y}$, that is $\mathrm{N}\left(0,\sigma_y^2\right)$ given the observation vector $\underset{\sim}{\mathrm{x}}$:

$$f_{y|\mathrm{x}}(y\,|\,\mathrm{x})=\frac{f_{\mathrm{x}y}(\mathrm{x},y)}{f_{\mathrm{x}}(\mathrm{x})}=\frac{\dfrac{1}{\left(\sqrt{2\pi}\right)^{N+1}\left|\mathrm{C}_{\mathrm{z}}\right|^{1/2}}\exp\left(-\dfrac{1}{2}\mathrm{z}^T\mathrm{C}_{\mathrm{z}}^{-1}\mathrm{z}\right)}{\dfrac{1}{\left(\sqrt{2\pi}\right)^{N}\left|\mathrm{C}_{x}\right|^{1/2}}\exp\left(-\dfrac{1}{2}\mathrm{x}^T\mathrm{C}_{x}^{-1}\mathrm{x}\right)}$$

Here $\underset{\sim}{\mathrm{z}}$ is the concatenation of $\underset{\sim}{\mathrm{x}}$ and $\underset{\sim}{y}$, $\mathrm{z}=\begin{bmatrix}\mathrm{x}_{\mathrm{Nx1}}\\ y\end{bmatrix}_{(\mathrm{N}+1)\mathrm{x}1}$. C_z and C_{x} is the covariance matrix of $\underset{\sim}{\mathrm{z}}$ and $\underset{\sim}{\mathrm{x}}$, respectively. Then as in the scalar the problem, the exponent in the numerator can be expressed as follows:

$$\frac{1}{2}\mathrm{z}^T\mathrm{C}_{\mathrm{z}}^{-1}\mathrm{z}=\begin{bmatrix}\mathrm{x} & y\end{bmatrix}\begin{bmatrix}\mathrm{C}_{\mathrm{x}} & \mathrm{r}_{\mathrm{x}y}\\ \mathrm{r}_{\mathrm{x}y}^T & \sigma_y^2\end{bmatrix}^{-1}\begin{bmatrix}\mathrm{x}\\ y\end{bmatrix}$$

In the equation above, $\mathrm{r}_{\mathrm{x}y}=E\{\mathrm{x}y\}$ is the cross-correlation of observations and $\underset{\sim}{y}$. The matrix inverse in the equation above can be taken using *the partitioned matrix inversion lemma*.

There are various type of matrix inversion lemmas, they are abundant in books but for the sake of simplicity we present a screenshot of a webpage illustrating the form of lemma that we would like to implement.

Prepared by:
Çağatay Candan

METU, Electrical Engin. Dept.
Dec. 2011
document version:1.2

Special Case 1

Let a $(m+1)\times(m+1)$ matrix $\mathbf{M}$ be partitioned into a block form:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & c \end{bmatrix} \begin{matrix} \}m \\ \}1 \end{matrix}$$

Then the inverse of $\mathbf{M}$ is

$$\mathbf{M}^{-1} = \begin{bmatrix} \left(\mathbf{A} - \frac{1}{c}\mathbf{b}\mathbf{b}^T\right)^{-1} & -\frac{1}{k}\mathbf{A}^{-1}\mathbf{b} \\ -\frac{1}{k}\mathbf{b}^T\mathbf{A}^{-1} & \frac{1}{k} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{-1} + \frac{1}{k}\mathbf{A}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{A}^{-1} & -\frac{1}{k}\mathbf{A}^{-1}\mathbf{b} \\ -\frac{1}{k}\mathbf{b}^T\mathbf{A}^{-1} & \frac{1}{k} \end{bmatrix}$$

where $k = c - \mathbf{b}^T\mathbf{A}^{-1}\mathbf{b}$.

Partitioned Matrix Inversion Lemma

A screenshot from http://www.cs.nthu.edu.tw/~jang/book/addenda/matinv/matinv/

Substituting A with C_x and b with r_{xy}, we can express $\mathrm{C}_\mathrm{z}^{-1}$ as follows using the lemma:

$$\mathrm{C}_\mathrm{z}^{-1} = \begin{bmatrix} \mathrm{C}_\mathrm{x} & \mathrm{r}_{xy} \\ \mathrm{r}_{xy}^T & \sigma_y^2 \end{bmatrix}^{-1} = \begin{bmatrix} \mathrm{C}_\mathrm{x}^{-1} + \frac{1}{k}\mathrm{C}_\mathrm{x}^{-1}\mathrm{r}_{xy}\mathrm{r}_{xy}^T\mathrm{C}_\mathrm{x}^{-1} & -\frac{1}{k}\mathrm{C}_\mathrm{x}^{-1}\mathrm{r}_{xy} \\ -\frac{1}{k}\mathrm{r}_{xy}^T\mathrm{C}_\mathrm{x}^{-1} & \frac{1}{k} \end{bmatrix}$$

Then $f_{y|\mathrm{x}}(y\,|\,\mathrm{x})$ can be written as:

$$f_{y|\mathrm{x}}(y\,|\,\mathrm{x}) = \frac{f_{\mathrm{x}y}(\mathrm{x},y)}{f_\mathrm{x}(\mathrm{x})} = \frac{|\mathrm{C}_x|^{1/2}}{\sqrt{2\pi}|\mathrm{C}_\mathrm{z}|^{1/2}} \frac{\exp\left(-\frac{1}{2}\mathrm{z}^T\mathrm{C}_\mathrm{z}^{-1}\mathrm{z}\right)}{\exp\left(-\frac{1}{2}\mathrm{x}^T\mathrm{C}_x^{-1}\mathrm{x}\right)} = K\frac{\exp\left(-\frac{1}{2}\mathrm{z}^T\mathrm{C}_\mathrm{z}^{-1}\mathrm{z}\right)}{\exp\left(-\frac{1}{2}\mathrm{x}^T\mathrm{C}_x^{-1}\mathrm{x}\right)}$$

In the last equation, K denotes a constant scalar. The ratio of $\exp\left(-\frac{1}{2}\mathrm{z}^T\mathrm{C}_\mathrm{z}^{-1}\mathrm{z}\right)$ and $\exp\left(-\frac{1}{2}\mathrm{x}^T\mathrm{C}_x^{-1}\mathrm{x}\right)$ can be written more explicitly as follows:

$$\frac{\exp\left(-\frac{1}{2}\mathrm{z}^T\mathrm{C}_\mathrm{z}^{-1}\mathrm{z}\right)}{\exp\left(-\frac{1}{2}\mathrm{x}^T\mathrm{C}_x^{-1}\mathrm{x}\right)} = \frac{\exp\left(-\frac{1}{2}\begin{bmatrix}\mathrm{x}^T & y\end{bmatrix}\mathrm{C}_\mathrm{z}^{-1}\begin{bmatrix}\mathrm{x} \\ y\end{bmatrix}\right)}{\exp\left(-\frac{1}{2}\begin{bmatrix}\mathrm{x}^T & y\end{bmatrix}\begin{bmatrix}\mathrm{C}_x^{-1} & 0 \\ 0 & 0\end{bmatrix}\mathrm{C}_x^{-1}\begin{bmatrix}\mathrm{x} \\ y\end{bmatrix}\right)} = \exp\left\{-\frac{1}{2}\begin{bmatrix}\mathrm{x}^T & y\end{bmatrix}\left(\mathrm{C}_\mathrm{z}^{-1} - \begin{bmatrix}\mathrm{C}_x^{-1} & 0 \\ 0 & 0\end{bmatrix}\right)\begin{bmatrix}\mathrm{x} \\ y\end{bmatrix}\right\}$$

With the substitution of C_z^{-1}, we get the following for the matrix in the quadratic product:

$$\mathrm{C}_{\mathrm{z}}^{-1} - \begin{bmatrix} \mathrm{C}_x^{-1} & 0 \\ 0 & 0 \end{bmatrix} = \frac{1}{k} \begin{bmatrix} \mathrm{C}_{\mathrm{x}}^{-1}\mathrm{r}_{xy}\mathrm{r}_{xy}^T\mathrm{C}_{\mathrm{x}}^{-1} & -\mathrm{C}_{\mathrm{x}}^{-1}\mathrm{r}_{xy} \\ -\mathrm{r}_{xy}^T\mathrm{C}_{\mathrm{x}}^{-1} & 1 \end{bmatrix}$$

with $k = \sigma_y^2 - \mathrm{r}_{xy}^T\mathrm{C}_{\mathrm{x}}^{-1}\mathrm{r}_{xy}$. Then the density becomes:

$$f_{y|\mathrm{x}}(y\,|\,\mathrm{x}) = K\,\frac{\exp\left(-\frac{1}{2}\mathrm{z}^T\mathrm{C}_{\mathrm{z}}^{-1}\mathrm{z}\right)}{\exp\left(-\frac{1}{2}\mathrm{x}^T\mathrm{C}_x^{-1}\mathrm{x}\right)} = K\exp\left\{-\frac{1}{2k}\begin{bmatrix}\mathrm{x}^T & y\end{bmatrix}\frac{1}{k}\begin{bmatrix} \mathrm{C}_{\mathrm{x}}^{-1}\mathrm{r}_{xy}\mathrm{r}_{xy}^T\mathrm{C}_{\mathrm{x}}^{-1} & -\mathrm{C}_{\mathrm{x}}^{-1}\mathrm{r}_{xy} \\ -\mathrm{r}_{xy}^T\mathrm{C}_{\mathrm{x}}^{-1} & 1 \end{bmatrix}\begin{bmatrix}\mathrm{x} \\ y\end{bmatrix}\right\}$$

The last relation can be further reduced to

$$f_{y|\mathrm{x}}(y\,|\,\mathrm{x}) = K\exp\left\{-\frac{1}{2k}\left(\mathrm{x}^T\mathrm{C}_{\mathrm{x}}^{-1}\mathrm{r}_{xy}\mathrm{r}_{xy}^T\mathrm{C}_{\mathrm{x}}^{-1}\mathrm{x} - 2y\mathrm{r}_{xy}^T\mathrm{C}_{\mathrm{x}}^{-1}\mathrm{x} + y^2\right)\right\}$$

This not-so-friendly relation is our old friend in wolf's clothing. To recognize our friend, we just need to define $\hat{y} = \mathrm{r}_{xy}^T\mathrm{C}_{\mathrm{x}}^{-1}\mathrm{x}$, and rewrite the same expression as follows:

$$f_{y|\mathrm{x}}(y\,|\,\mathrm{x}) = K\exp\left\{-\frac{1}{2k}\left(\hat{y}^2 - 2y\hat{y} + y^2\right)\right\} = K\exp\left\{-\frac{1}{2k}(y-\hat{y})^2\right\}$$

Hence the conditional density is $\mathrm{N}\left(\mathrm{r}_{xy}^T\mathrm{C}_{\mathrm{x}}^{-1}\mathrm{x}, \sigma_y^2 - \mathrm{r}_{xy}^T\mathrm{C}_{\mathrm{x}}^{-1}\mathrm{r}_{xy}\right)$. (This is one of the most important results for the estimation theory. This shows that the linear minimum mean square error estimator (a topic of EE503) is the conditional mean which is the optimal estimator for the general minimum mean square error estimation having no linearity constraints.)