Bivariate Normal

Snehal M. Shekatkar

1 Derivation of the density formula

We will generalize the notion of a univariate normal distribution to two dimensions by demanding that the contours of constant density are ellipses. Consider a general equation of ellipse:

$$Ax_1^2 + 2Bx_1x_2 + Cx_2^2 = k$$

The LHS can be rewritten in a matrix form as:

$$\mathbf{x^T M x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} A & B \\ B & A \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

We will demand that the density should drop as exponential of this:

$$p(\mathbf{X}) = \frac{1}{Z} e^{-\mathbf{x^T M x}}$$

where $1/Z$ is the constant of proportionality to be determined. For the point $\mathbf{x}$ on a given ellipse, the quantity $\mathbf{x^T M x}$ is called *Mahalanobis distance*. Later we will see that for appropriate choice of matrix M, this represents the distance of a point on a circle from its center. Since we don't want it to be negative, we see that M must be a *positive-definite matrix*.

Throughtout the discussion of the bivariate normal, we will use the following Gaussian integral formula:

$$\int\limits_0^\infty x^m e^{-\beta x^n} dx = \frac{1}{n} \frac{\Gamma(\frac{m+1}{n})}{\beta^{\frac{m+1}{n}}} \tag{1}$$

where $\Gamma(n+1) = n\Gamma(n)$, and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Clearly,

$$\begin{aligned} Z &= \int\limits_{x_1=-\infty}^{\infty} \int\limits_{x_2=-\infty}^{\infty} \exp(-Ax_1^2 - 2Bx_1x_2 - Cx_2^2) dx_2 dx_1 \\ &= \int\limits_{x_1=-\infty}^{\infty} e^{-Ax_1^2} \left[\int\limits_{x_2=-\infty}^{\infty} e^{-2Bx_1x_2 - Cx_2^2} dx_2 \right] dx_1 \end{aligned} \tag{2}$$

Let us first consider the inner integral:

$$I_{x_2} = \int\limits_{x_2=-\infty}^{\infty} e^{-2Bx_1x_2 - Cx_2^2} dx_2 \tag{3}$$

To compute this integral, we can complete the square in x_2:

$$2Bx_1x_2 + Cx_2^2 = C\left(x_2^2 + 2\frac{B}{C}x_1x_2\right) = C\left(x_2 + \frac{B}{C}x_1\right)^2 - \frac{B^2}{C}x_1^2$$

$$I_{x_2} = e^{\frac{B^2}{C}x_1^2} \int\limits_{x_2=-\infty}^{\infty} \exp\left[-C\left(x_2 + \frac{B}{C}x_1\right)^2\right] dx_2 \tag{4}$$

To compute the integral, we can substitute $x_2' = x_2 + \frac{B}{C}x_1$. Hence:

$$\begin{aligned}
I_{x_2} &= e^{\frac{B^2}{C}x_1^2} \int\limits_{x_2'=-\infty}^{\infty} e^{-C{x_2'}^2}\, dx_2' = e^{\frac{B^2}{C}x_1^2} \times 2 \int\limits_{x_2'=0}^{\infty} e^{-C{x_2'}^2}\, dx_2' \\
&= e^{\frac{B^2}{C}x_1^2} \times 2 \times \frac{1}{2}\frac{\Gamma(1/2)}{\sqrt{C}} \\
&= \sqrt{\frac{\pi}{C}}\, e^{\frac{B^2}{C}x_1^2}
\end{aligned} \tag{5}$$

Putting this in 2, we get:

$$\begin{aligned}
Z &= \int\limits_{x_1=-\infty}^{\infty} e^{-Ax_1^2} \sqrt{\frac{\pi}{C}}\, e^{\frac{B^2}{C}x_1^2} dx_1 \\
&= \sqrt{\frac{\pi}{C}} \int\limits_{x_1=-\infty}^{\infty} e^{-\left(A-\frac{B^2}{C}\right)x_1^2} dx_1 \\
&= \sqrt{\frac{\pi}{C}} \times 2 \times \int\limits_{x_1=0}^{\infty} e^{-\left(A-\frac{B^2}{C}\right)x_1^2} dx_1 \\
&= \sqrt{\frac{\pi}{C}} \times 2 \times \frac{1}{2}\frac{\Gamma(1/2)}{\left(A-\frac{B^2}{C}\right)^{1/2}}
\end{aligned} \tag{6}$$

$$\boxed{\therefore Z = \frac{\pi}{\sqrt{AC - B^2}}} \tag{7}$$

Let us now compute the marginal distribution of x_1 by integrating over x_2:

$$\begin{aligned}
p_{X_1}(x) &= \int\limits_{x_2=-\infty}^{\infty} \frac{1}{Z}\, e^{-(Ax_1^2 + 2Bx_1x_2 + Cx_2^2)} dx_2 \\
&= \frac{1}{Z} e^{-Ax_1^2} \int\limits_{x_2=-\infty}^{\infty} e^{-(2Bx_1x_2 + Cx_2^2)} dx_2 = \frac{1}{Z} e^{-Ax_1^2} I_{x_2} \\
&= \frac{\sqrt{AC - B^2}}{\pi} e^{-Ax_1^2} \sqrt{\frac{\pi}{C}}\, e^{\frac{B^2}{C}x_1^2} \\
&= \sqrt{\frac{AC - B^2}{\pi C}} e^{-\left(A-\frac{B^2}{C}\right)x_1^2}
\end{aligned} \tag{8}$$

$$\boxed{\therefore p_{X_1}(x) = \sqrt{\frac{AC - B^2}{\pi C}} \exp\left[-\left(\frac{AC - B^2}{C}\right) x_1^2\right]} \tag{9}$$

But this is just the univariate Gaussian distribution with mean zero! Comparing the terms with the variance terms in the univariate Gaussian distribution formula, we see that:

$$\sigma_{x_1}^2 = \frac{C}{2(AC - B^2)} \tag{10}$$

Similarly, marginalizing over x_1, we see that the distribution of x_2 is also Gaussian:

$$\boxed{\therefore p_{X_2}(x) = \sqrt{\frac{AC - B^2}{\pi A}} \exp\left[-\left(\frac{AC - B^2}{A}\right) x_1^2\right]} \tag{11}$$

and hence:

$$\sigma_{x_2}^2 = \frac{A}{2(AC - B^2)} \tag{12}$$

Since the means of X_1 and X_2 are zero, we can also compute the correlation between X_1 and X_2 as:

$$\begin{aligned}
\mathrm{Cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] &= \frac{1}{Z} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} x_1 x_2 \exp(-Ax_1^2 - 2Bx_1x_2 - Cx_2^2) dx_2 dx_1 \\
&= \frac{1}{Z} \int\limits_{-\infty}^{\infty} x_1 e^{-Ax_1^2} \left[\int\limits_{-\infty}^{\infty} x_2 e^{-2Bx_1x_2 - Cx_2^2} dx_2 \right] dx_1
\end{aligned} \tag{13}$$

Now consider the inner integral:

$$I_{xx_2} = \int\limits_{-\infty}^{\infty} x_2 e^{-2Bx_1x_2 - Cx_2^2} dx_2 \tag{14}$$

Proceeding as before,

$$\begin{aligned}
I_{xx_2} &= e^{\frac{B^2}{C}x_1^2} \int\limits_{x_2=-\infty}^{\infty} x_2 \exp\left[-C \left(x_2 + \frac{B}{C}x_1 \right)^2 \right] dx_2 \\
&= e^{\frac{B^2}{C}x_1^2} \int\limits_{x_2=-\infty}^{\infty} \left(x_2 - \frac{B}{C}x_1 \right) e^{-Cx_2^2} dx_2 \\
&= e^{\frac{B^2}{C}x_1^2} \int\limits_{-\infty}^{\infty} x_2 e^{-Cx_2^2} dx_2 - e^{\frac{B^2}{C}x_1^2} \frac{B}{C} x_1 \int\limits_{-\infty}^{\infty} e^{-Cx_2^2} dx_2
\end{aligned} \tag{15}$$

The first of these integrals, being an odd-integral, is zero. Thus,

$$\begin{aligned}
I_{xx_2} &= -e^{\frac{B^2}{C}x_1^2} \frac{B}{C} x_1 \int\limits_{-\infty}^{\infty} e^{-Cx_2^2} dx_2 = -e^{\frac{B^2}{C}x_1^2} \frac{B}{C} x_1 \times 2 \times \int\limits_{0}^{\infty} e^{-Cx_2^2} dx_2 \\
&= -e^{\frac{B^2}{C}x_1^2} \frac{B}{C} x_1 \times 2 \times \frac{1}{2} \frac{\Gamma(1/2)}{C^{1/2}} \\
&= -x_1 \ e^{\frac{B^2}{C}x_1^2} \frac{B\sqrt{\pi}}{C^{3/2}}
\end{aligned} \tag{16}$$

Putting this in 13, we get:

$$\begin{aligned}
\mathrm{Cov}(X_1,X_2) &= \frac{1}{Z}\int\limits_{-\infty}^{\infty} x_1 e^{-Ax_1^2}\left(-x_1\ e^{\frac{B^2}{C}x_1^2}\frac{B\sqrt{\pi}}{C^{3/2}}\right)dx_1 \\
&= \frac{-1}{Z}\frac{B\sqrt{\pi}}{C^{3/2}}\int\limits_{-\infty}^{\infty} x_1 e^{-Ax_1^2}\left(x_1\ e^{\frac{B^2}{C}x_1^2}\right)dx_1 \\
&= \frac{-1}{Z}\frac{B\sqrt{\pi}}{C^{3/2}}\int\limits_{-\infty}^{\infty} x_1^2 e^{-\left(A-\frac{B^2}{C}\right)x_1^2}dx_1 \\
&= \frac{-1}{Z}\frac{B\sqrt{\pi}}{C^{3/2}}\times 2\times\int\limits_{0}^{\infty} x_1^2 e^{-\left(A-\frac{B^2}{C}\right)x_1^2}dx_1 \\
&= \frac{-1}{Z}\frac{B\sqrt{\pi}}{C^{3/2}}\times 2\times\frac{1}{2}\frac{\Gamma(\frac{3}{2})}{\left(A-\frac{B^2}{C}\right)^{3/2}} \\
&= \frac{-1}{Z}\frac{B\sqrt{\pi}}{C^{3/2}}\frac{(\sqrt{\pi}/2)C^{3/2}}{\left(AC-B^2\right)^{3/2}} \\
&= \frac{-B\pi}{2Z\left(AC-B^2\right)^{3/2}}
\end{aligned} \tag{17}$$

Using the value of Z from (7), we get:

$$\mathrm{Cov}(X_1,X_2) = \frac{\sqrt{AC-B^2}}{2\pi}\times\frac{-B\pi}{\left(AC-B^2\right)^{3/2}} \tag{18}$$

$$\boxed{\therefore \mathrm{Cov}(X_1,X_2) = \frac{-B}{2(AC-B^2)}} \tag{19}$$

Hence the full covariance matrix is:

$$\Sigma = \begin{bmatrix} \frac{C}{2(AC-B^2)} & \frac{-B}{2(AC-B^2)} \\ \frac{-B}{2(AC-B^2)} & \frac{A}{2(AC-B^2)} \end{bmatrix} = \frac{1}{2}M^{-1} \tag{20}$$

$$\therefore M = \frac{1}{2}\Sigma^{-1} \tag{21}$$

Also, then from Eq(7), we have:

$$Z = \frac{\pi}{\sqrt{AC-B^2}} = \frac{\pi}{\sqrt{\det(M)}} = \frac{\pi}{\sqrt{\det(\frac{1}{2}\Sigma^{-1})}} = \frac{\pi}{\sqrt{\frac{1}{2^2}\det(\Sigma^{-1})}} = 2\pi\sqrt{\det(\Sigma)}$$

$$\boxed{\therefore \frac{1}{Z} = \frac{1}{2\pi\sqrt{\det(\Sigma)}}} \tag{22}$$

Putting everything together, the bivariate density when the mean is $[0,0]^T$ is:

$$p(\mathbf{X}) = \frac{1}{Z}e^{-\mathbf{x^T M x}} = \frac{1}{2\pi\sqrt{\det(\Sigma)}}e^{-\frac{1}{2}\mathbf{x^T \Sigma^{-1} x}} \tag{23}$$

If we had started with an ellipse with center $\mu = [\mu_1,\mu_2]^T$, then the density would be:

$$\boxed{p(\mathbf{x}) = \frac{1}{2\pi\sqrt{\det(\Sigma)}}e^{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathbf{T}}\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu)}} \tag{24}$$

This can be generalized to n-dimensional normal distribution as:

$$\boxed{p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathbf{T}}\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu)}} \tag{25}$$

2 Bivariate normal and regression

We can write the Eq(24) in an algebraic form:

$$p(x_1, x_2) = \frac{1}{2\pi\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right) \tag{26}$$

where,

$$(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu) = \frac{1}{1-r^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_{x_1}^2} - 2r\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_{x_1}\sigma_{x_2}} + \frac{(x_2-\mu_2)^2}{\sigma_{x_2}^2}\right] \tag{27}$$

where r is the correlation coefficient of x_1 and x_2. Let us look at the conditional distribution of X_2 given X_1. To do this, let us not worry about the normalization constant, and any terms that depend only on X_1 as we are fixing that. Thus:

$$X_2|X_1 \sim K\exp\left(\frac{-1}{2(1-r^2)\sigma_{x_2}^2}\left[(x_2-\mu_2)^2 - 2r\frac{\sigma_{x_2}}{\sigma_{x_1}}(x_1-\mu_1)(x_2-\mu_2)\right]\right) \tag{28}$$

where K does not depend on x_2. Completing the squares in $(x_2-\mu_2)$, we get:

$$X_2|X_1 \sim K\exp\left(\frac{-1}{2(1-r^2)\sigma_{x_2}^2}\left[(x_2-\mu_2) - r\frac{\sigma_{x_2}}{\sigma_{x_1}}(x_1-\mu_1)\right]^2 + \frac{r^2}{2(1-r^2)\sigma_{x_1}^2}(x_1-\mu_1)^2\right) \tag{29}$$

The factor in the exponential (the last term) that depends on x_1 can be absorbed in K, and then rearranging the remaining factors,

$$X_2|X_1 \sim K\exp\left(\frac{-1}{2(1-r^2)\sigma_{x_2}^2}\left[x_2 - \left(\mu_2 - r\frac{\sigma_{x_2}}{\sigma_{x_1}}\mu_1 + r\frac{\sigma_{x_2}}{\sigma_{x_1}}x_1\right)\right]^2\right) \tag{30}$$

But earlier when we considered the linear regression model $X_2 = \beta_0 + \beta_1 X_1 + \eta$ where $\eta \sim \mathcal{N}(0, \sigma^2)$, we had proved that:

$$\begin{aligned} \beta_1 &= r\frac{\sigma_{x_2}}{\sigma_{x_1}} \\ \beta_0 &= \mu_2 - \beta_1\mu_1 = \mu_2 - r\frac{\sigma_{x_2}}{\sigma_{x_1}}\mu_1 \end{aligned} \tag{31}$$

Thus, we see that the conditional distribution of X_2 given X_1 is just Gaussian with $\mathbb{E}(X_2|X_1)$ varying linearly with $\mathbb{E}(X_1)$! This implies that when the joint density of (X_1, X_2) is bivariate Gaussian, the relationship between them is forced to be linear, and any other complex models (for example, quadratic) of the relationship cannot be used.