Consider the following bivariate density:

\[ f(x, y) = c \exp\left\{ -\frac{1}{2}Q(x, y) \right\}, \]

where \( c \) is a constant, \( Q \) a positive definite quadratic form in \( x \) and \( y \). Specifically:

\[
c = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}},
\]

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

Here \( \sigma_i > 0, \mu_i \) are real, \(-1 < \rho < 1\). Since \( f \) is clearly non-negative, to show that \( f \) is a (probability density) function (in two dimensions), it suffices to show that \( f \) integrates to 1:

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1, \quad \text{or} \quad \int \int f = 1.
\]

Write

\[
f_1(x) := \int_{-\infty}^{\infty} f(x, y) \, dy, \quad f_2(y) := \int_{-\infty}^{\infty} f(x, y) \, dx.
\]

Then to show \( \int \int f = 1 \), we need to show \( \int_{-\infty}^{\infty} f_1(x) \, dx = 1 \) (or \( \int_{-\infty}^{\infty} f_2(y) \, dy = 1 \)). Then \( f_1, f_2 \) are densities, in one dimension. If \( f(x, y) = f_{X,Y}(x, y) \) is the joint density of two random variables \( X, Y \), then \( f_1(x) \) is the density \( f_X(x) \) of \( X \), \( f_2(y) \) the density \( f_Y(y) \) of \( Y \) (\( f_1, f_2 \), or \( f_X, f_Y \), are called the marginal densities of the joint density \( f \), or \( f_{X,Y} \)).

To perform the integrations, we have to complete the square. We have the algebraic identity

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

(reducing the number of occurrences of \( y \) to 1, as we intend to integrate out \( y \) first). Then (taking the terms free of \( y \) out through the \( y \)-integral)

\[
f_1(x) = \frac{\exp\left(-\frac{1}{2}(x-\mu_1)^2/\sigma_1^2\right)}{\sigma_1\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sigma_2\sqrt{2\pi\sqrt{1-\rho^2}}} \exp\left(-\frac{1}{2}(y-c_x)^2/\sigma_2^2(1-\rho^2)\right) \, dy,
\]

\( (*) \)
where  

\[ c_x := \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1). \]

The integral is 1 (‘normal density’). So

\[ f_1(x) = \exp \left( -\frac{1}{2} (x - \mu_1)^2 / \sigma_1^2 \right) \]

which integrates to 1 (‘normal density’), proving  
**Fact 1.** \( f(x, y) \) is a joint density function (two-dimensional), with marginal density functions \( f_1(x) \), \( f_2(y) \) (one-dimensional). So we can write

\[ f(x, y) = f_{X,Y}(x, y), \quad f_1(x) = f_X(x), \quad f_2(y) = f_Y(y). \]

**Fact 2.** \( X, Y \) are normal: \( X \) is \( N(\mu_1, \sigma_1^2) \), \( Y \) is \( N(\mu_2, \sigma_2^2) \). For, we showed \( f_1 = f_X \) to be the \( N(\mu_1, \sigma_1^2) \) density above, and similarly for \( Y \) by symmetry.

**Fact 3.** \( EX = \mu_1, EY = \mu_2, varX = \sigma_1^2, varY = \sigma_2^2 \).

This identifies four out of the five parameters: two means \( \mu_i \), two variances \( \sigma_i^2 \). Next, recall the definition of conditional probability:

\[ P(A|B) := P(A \cap B)/P(B). \]

In the discrete case, if \( X, Y \) take possible values \( x_i, y_j \) with probabilities \( f_X(x_i), f_Y(y_j) \), \( (X, Y) \) takes possible values \( (x_i, y_j) \) with probabilities \( f_{X,Y}(x_i, y_j) \):

\[ f_X(x_i) = P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j f_{X,Y}(x_i, y_j). \]

Then the conditional distribution of \( Y \) given \( X = x_i \) is

\[ f_{Y|X}(y_j|x_i) = P(Y = y_j \& X = x_i)/P(X = x_i) = f_{X,Y}(x_i, y_j)/\sum_j f_{X,Y}(x_i, y_j), \]

and similarly with \( X, Y \) interchanged.

In the density case, we have to replace sums by integrals. Thus the conditional density of \( Y \) given \( X = x \) is (see e.g. Haigh (2002), Def. 4.19, p. 80)

\[ f_{Y|X}(y|x) := f_{X,Y}(x, y)/f_X(x) = f_{X,Y}(x, y)/\int_{-\infty}^{\infty} f_{X,Y}(x, y)dy. \]

Returning to the bivariate normal:

**Fact 4.** The conditional distribution of \( y \) given \( X = x \) is \( N(\mu_2 + \rho \sigma_2 \sigma_1^{-1} (x - \mu_1), \sigma_2^2 (1 - \rho^2)) \).
Proof. Go back to completing the square (or, return to (*) with \( f \) and \( dy \) deleted):

\[
f(x, y) = \frac{\exp(-\frac{1}{2}(x - \mu_1)^2/\sigma_1^2)}{\sigma_1 \sqrt{2\pi}} \cdot \frac{\exp(-\frac{1}{2}(y - c_x)^2/(\sigma_2^2(1 - \rho^2)))}{\sigma_2 \sqrt{2\pi \sqrt{1 - \rho^2}}}.\]

The first factor is \( f_1(x) \), by Fact 1. So, \( f_{Y|X}(y|x) = f(x, y)/f_1(x) \) is the second factor:

\[
f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi} \sigma_2 \sqrt{1 - \rho^2}} \exp\left(\frac{- (y - c_x)^2}{2 \sigma_2^2(1 - \rho^2)}\right),
\]

where \( c_x \) is the linear function of \( x \) given below (*). //

This not only completes the proof of Fact 4 but gives

**Fact 5.** The conditional mean \( E(Y|X = x) \) is linear in \( x \):

\[
E(Y|X = x) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1).
\]

**Note.** This simplifies when \( X \) and \( Y \) are equally variable, \( \sigma_1 = \sigma_2 \):

\[
E(Y|X = x) = \mu_2 + \rho (x - \mu_1)
\]

(recall \( EX = \mu_1, EY = \mu_2 \)). Recall that in Galton’s height example, this says: for every inch of mid-parental height above/below the average, \( x - \mu_1 \), the parents pass on to their child, on average, \( \rho \) inches, and continuing in this way: on average, after \( n \) generations, each inch above/below average becomes on average \( \rho^n \) inches, and \( \rho^n \to 0 \) as \( n \to \infty \), giving regression towards the mean.

This line is the population regression line (PRL), the population version of the sample regression line (SRL).

The relationship in Fact 5 can be generalized: a population regression function – more briefly, a regression – is a **conditional mean**.

This also gives

**Fact 6.** The conditional variance of \( Y \) given \( X = x \) is

\[
\text{var}(Y|X = x) = \sigma_2^2(1 - \rho^2).
\]

Recall (Fact 3) that the variability (= variance) of \( Y \) is \( \text{var}Y = \sigma_2^2 \). By Fact 5, the variability remaining in \( Y \) when \( X \) is given (i.e., not accounted
for by knowledge of $X$ is $\sigma_Y^2(1 - \rho^2)$. Subtracting: the variability of $Y$ which is accounted for by knowledge of $X$ is $\sigma_Y^2\rho^2$. That is: $\rho^2$ is the proportion of the variability of $Y$ accounted for by knowledge of $X$. So $\rho$ is a measure of the strength of association between $Y$ and $X$.

Recall that the covariance is defined by

$$\text{cov}(X,Y) := E[(X - EX)(Y - EY)] = E[(X - \mu_1)(Y - \mu_2)],$$

and the correlation coefficient $\rho$, or $\rho(X,Y)$, defined by

$$\rho = \rho(X,Y) := \text{cov}(X,Y)/\sqrt{\text{var}X \text{var}Y} = E[(X - \mu_1)(Y - \mu_2)]/\sigma_1\sigma_2$$

is the usual measure of the strength of association between $X$ and $Y$ ($-1 \leq \rho \leq 1$; $\rho = \pm 1$ iff one of $X, Y$ is a function of the other).

**Fact 7.** The correlation coefficient of $X, Y$ is $\rho$.

*Proof.*

$$\rho(X,Y) := E\left[\left(\frac{X - \mu_1}{\sigma_1}\right)\left(\frac{Y - \mu_2}{\sigma_2}\right)\right] = \int \int \left(\frac{x - \mu_1}{\sigma_1}\right)\left(\frac{y - \mu_2}{\sigma_2}\right)f(x,y)dxdy.$$ 

Substitute for $f(x,y) = c \exp(-\frac{1}{2}Q)$, and make the change of variables $u := (x - \mu_1)/\sigma_1$, $v := (y - \mu_2)/\sigma_2$:

$$\rho(X,Y) = \frac{1}{2\pi \sqrt{1 - \rho^2}} \int \int uv \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2(1 - \rho^2)}\right)dudv.$$ 

Completing the square as before, $[u^2 - 2\rho uv + v^2] = (v - \rho u)^2 + (1 - \rho^2)u^2$. So

$$\rho(X,Y) = \frac{1}{\sqrt{2\pi}} \int u \exp\left(-\frac{u^2}{2}\right)du \cdot \frac{1}{\sqrt{2\pi \sqrt{1 - \rho^2}}} \int v \exp\left(-\frac{(v - \rho u)^2}{2(1 - \rho^2)}\right)dv.$$ 

Replace $v$ in the inner integral by $(v - \rho u) + \rho u$, and calculate the two resulting integrals separately. The first is zero (‘normal mean’, or symmetry), the second is $\rho u$ (‘normal density’). So

$$\rho(X,Y) = \frac{1}{\sqrt{2\pi}} \rho \int u^2 \exp\left(-\frac{u^2}{2}\right)du = \rho$$ 

(‘normal variance’), as required. //

This completes the identification of all five parameters in the bivariate
normal distribution: two means $\mu_i$, two variances $\sigma_i^2$, one correlation $\rho$.

Note. The above holds for $-1 < \rho < 1$; always, $-1 \leq \rho \leq 1$. In the limiting cases $\rho = \pm 1$, one of $X, Y$ is a linear function of the other: $Y = aX + b$, say, as in the temperature example (Fahrenheit and Centigrade). The situation is not really two-dimensional: we can (and should) use only one of $X$ and $Y$, reducing to a one-dimensional problem.

The slope of the regression line $y = c_x$ is $\rho \sigma_2 / \sigma_1 = (\rho \sigma_1 \sigma_2) / (\sigma_1^2)$, which can be written as $\text{cov}(X, Y) / \text{var} X = \sigma_{12} / \sigma_{11}$, or $\sigma_{12} / \sigma_1^2$: the line is

$$y - EY = \frac{\sigma_{12}}{\sigma_{11}} (x - EX).$$

This is the population version (what else?!) of the sample regression line

$$y - \bar{Y} = \frac{S_{XY}}{S_{XX}} (x - \bar{X}),$$

familiar from linear regression.

The case $\rho = \pm 1$ – apparently two-dimensional, but really one-dimensional – is singular; the case $-1 < \rho < 1$ - genuinely two-dimensional - is non-singular; or (see below) full rank.

We note in passing

**Fact 8.** The bivariate normal law has elliptical contours.

For, the contours are $Q(x, y) = \text{const}$, which are ellipses (as Galton found).

**Moment Generating Function (MGF).** Recall (see e.g. Haigh (2002), 102-6) $M(t)$, or $M_X(t) := E(e^{tX})$. For $X$ normal $N(\mu, \sigma^2)$,

$$M(t) = \frac{1}{\sigma \sqrt{2\pi}} \int e^{tx} \exp\left(-\frac{1}{2} (x - \mu)^2 / \sigma^2\right) dx.$$

Change variable to $u := (x - \mu) / \sigma$:

$$M(t) = \frac{1}{\sqrt{2\pi}} \int \exp(\mu t + \sigma u - \frac{1}{2} u^2) du.$$

Completing the square,

$$M(t) = e^{\mu t} \frac{1}{\sqrt{2\pi}} \int \exp\left(-\frac{1}{2} (u - \sigma t)^2\right) du e^{\frac{1}{2} \sigma^2 t^2},$$

5
or $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$ (recognising that the central term on the right is $1 - \text{‘normal density’}$). So $M_{X-\mu}(t) = \exp(\frac{1}{2}\sigma^2 t^2)$. Then (check) $\mu = EX = M'_X(0), \text{var}X = E[(X - \mu)^2] = M''_{X-\mu}(0)$.

Similarly in the bivariate case: the MGF is

$$M_{X,Y}(t_1, t_2) := E \exp(t_1 X + t_2 Y).$$

In the bivariate normal case:

$$M(t_1, t_2) = E(\exp(t_1 X + t_2 Y)) = \int \int \exp(t_1 x + t_2 y) f(x, y) dxdy = \int \exp(t_1 x) f_1(x) dx \int \exp(t_2 y) f(y|x) dy.$$

The inner integral is the MGF of $Y|X = x$, which is $N(c_x, \sigma_2^2, (1 - \rho^2))$, so is $\exp(c_x t_2 + \frac{1}{2} \sigma_2^2 (1 - \rho^2) t_2^2)$. By Fact 5 $c_x t_2 = [\mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1)] t_2$, so

$$M(t_1, t_2) = \exp(t_2 \mu_2 - t_2 \sigma_2 \frac{\mu_1}{\sigma_1} + \frac{1}{2} \sigma_2^2 (1 - \rho^2) t_2^2) \int \exp([t_1 + t_2 \rho \frac{\sigma_2}{\sigma_1}] x) f_1(x) dx.$$

Since $f_1(x)$ is $N(\mu_1, \sigma_1^2)$, the inner integral is a normal MGF, which is thus $\exp(\mu_1 [t_1 + t_2 \rho \frac{\sigma_2}{\sigma_1}] + \frac{1}{2} \sigma_1^2 [\ldots]^2)$. Combining the two terms and simplifying, we obtain

**Fact 9.** The joint MGF is

$$M_{X,Y}(t_1, t_2) = M(t_1, t_2) = \exp(\mu_1 t_1 + \mu_2 t_2 + \frac{1}{2} [\sigma_1^2 t_1^2 + 2 \rho \sigma_1 \sigma_2 t_1 t_2 + \sigma_2^2 t_2^2]).$$

**Fact 10.** $X, Y$ are independent if and only if $\rho = 0$.

Proof. For densities: $X, Y$ are independent iff the joint density $f_{X,Y}(x, y)$ factorises as the product of the marginal densities $f_X(x), f_Y(y)$ (see e.g. Haigh (2002), Cor. 4.17).

For MGFs: $X, Y$ are independent iff the joint MGF $M_{X,Y}(t_1, t_2)$ factorises as the product of the marginal MGFs $M_X(t_1).M_Y(t_2)$. From Fact 9, this occurs iff $\rho = 0$. Similarly with CFs, if we prefer to work with them. //

NHB