smfl2.tex
Lecture 2. 21.1.2011

Proof. Go back to completing the square (or, return to (*) with $\int$ 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 2. So, $f_{Y|X}(y|x) = f(x,y)/f_1(x)$ is the second factor:

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

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*, ρ inches, and continuing in this way: *on average*, after n generations, each inch above/below average becomes *on average* ρ^n inches, and $\rho^n \to \infty$ as $n \to \infty$, giving *regression towards the mean.*
(A regression function is a *conditional mean* – see Section 5.)
Fact 6. The conditional variance of Y given $X = x$ is

$$var(Y|X=x) = \sigma_2^2(1-\rho^2).$$

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

Recall that the *covariance* is defined by

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

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

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

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

Fact 7. The correlation coefficient of X, Y is ρ.

Proof.

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

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

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

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

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

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 ρu ('normal density'). So

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

('normal variance'), as required.

This completes the identification of all five parameters in the bivariate normal distribution: two means μ_i, two variances σ_i^2, one correlation ρ.

Note 1. The above holds for $-1 < \rho < 1$; always, $-1 \le \rho \le 1$. In the limiting cases $\rho = \pm 1$, one of X, Y is a linear function of the other: $Y = aX + b$, say,

as with temperature (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.
Note 2. 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 $cov(X,Y)/varX = \sigma_{12}/\sigma_{11}$, or σ_{12}/σ_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*

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

from linear regression (Section 1).

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 $Q(x,y) = const$, which are ellipses (as Galton found).
Moment Generating Function (MGF). Recall $M(t) := E(e^{tX})$. For $X\ N(\mu, \sigma^2)$, $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$ [SP, Problems 5]. So $M_{X-\mu}(t) = \exp(\frac{1}{2}\sigma^2 t^2)$, and the CF is $\phi_{X-\mu}(t) = \exp(-\frac{1}{2}\sigma^2 t^2)$. Then (check) $\mu = EX = M'_X(0)$, $varX = E[(X-\mu)^2] = M''_{X-\mu}(0)$.

Similarly in the bivariate case: the MGF is

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

For the bivariate normal,

$$M(t_1, t_2) = E(\exp(t_1 X + t_2 Y)) = \int\int \exp(t_1 x + t_2 y) f(x,y)dxdy$$

$$= \int \exp(t_1 x) f_1(x)dx \int \exp(t_2 y) f(y|x)dy.$$

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 4, $c_x t_2 = [\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)]t_2$, so

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

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:
Fact 9. The joint MGF and joint CF of X, Y are

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

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

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 $f_X(x).f_Y(y)$. 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.
Note. X, Y independent implies X, Y uncorrelated ($\rho = 0$) in general (when the correlation exists). The converse if *false* in general, but *true*, by Fact 10, in the bivariate normal case.

3. THE MULTIVARIATE NORMAL DISTRIBUTION.

With one regressor, we used the bivariate normal distribution as above. Similarly for two regressors, we use the trivariate normal. With any number of regressors, as here, we need a general *multivariate normal* - or *'multinormal'* - distribution in n dimensions. We must expect that in n dimensions, to handle a random n-vector $\mathbf{X} = (X_1, \cdots, X_n)^T$, we will need
(i) a *mean vector* $\mu = (\mu_1, \cdots, \mu_n)^T$ with $\mu_i = EX_i$, $\mu = E\mathbf{X}$,
(ii) a *covariance matrix* $\mathbf{\Sigma} = (\sigma_{ij})$, with $\sigma_{ij} = cov(X_i, X_j)$: $\mathbf{\Sigma} = cov\mathbf{X}$.

First, note how mean vectors and covariance matrices transform under linear changes of variable:

PROPOSITION 1. If $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, with $\mathbf{Y}, \mathbf{b}$ m-vectors, $\mathbf{A}$ an $m \times n$ matrix and $\mathbf{X}$ an n-vector,
(i) the mean vectors are related by $E\mathbf{Y} = \mathbf{A}E\mathbf{X} + \mathbf{b} = \mathbf{A}\mu + \mathbf{b}$,
(ii) the covariance matrices are related by $\mathbf{\Sigma_Y} = \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T$.