Covariance and Correlation

Suppose X, Y are rv's with

$$EX = \mu_X, \quad \operatorname{Var} X = \sigma_X^2 < \infty,$$
$$EY = \mu_Y, \quad \operatorname{Var} Y = \sigma_Y^2 < \infty.$$

$\rightarrow$ $\operatorname{Cov}(X,Y) = E(X-\mu_X)(Y-\mu_Y)$ (defn.)

$= EXY - \mu_X\mu_Y$ (lemma)

Proof of lemma:

$$E\underbrace{(X-\mu_X)(Y-\mu_Y)}_{XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y}$$

$$= EXY - \mu_X \underbrace{EY}_{\mu_Y} - \mu_Y \underbrace{EX}_{\mu_X} + \mu_X\mu_Y$$

$$= EXY - \mu_X\mu_Y$$

$\rightarrow$ $$\rho_{XY} = \frac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Properties of Covariance

$$\mathrm{Cov}(X,X) = \mathrm{Var}(X)$$

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$$

$$\mathrm{Cov}(a,X) = 0 \quad (a \text{ is a constant})$$

$$\mathrm{Cov}(aX,Y) = a\,\mathrm{Cov}(X,Y)$$

$$\mathrm{Cov}(X+Y,Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)$$

By combining these properties we get

$$\mathrm{Cov}\left(\sum_i a_i X_i + c, \sum_j b_j Y_j + d\right)$$

$$= \sum_i \sum_j a_i b_j \,\mathrm{Cov}(X_i, Y_j)$$

and its special case (take $X_i = Y_i$, $a_i = b_i$, $c = d$)

$$\mathrm{Var}\left(\sum_{i=1}^n a_i X_i + c\right)$$

$$= \sum_{i=1}^n a_i^2 \,\mathrm{Var}(X_i) + 2\sum_{i<j} a_i a_j \,\mathrm{Cov}(X_i, X_j)$$

Properties of Correlation

$-1 \le \rho_{XY} \le 1$

Proof: Cauchy–Schwarz inequality implies $|E\,ST| \le (ES^2)^{1/2}(ET^2)^{1/2}$.
Now take $S = X - \mu_X$, $T = Y - \mu_Y$.

$\rho_{XY} = \pm 1$ iff $Y = aX + b$
for some constants a, b.

$\rho_{XY} = +1$ implies $a > 0$ in the above.

$\rho_{XY} = -1$ implies $a < 0$.

Definition: If $\rho_{XY} = 0$, we say X and Y are <u>uncorrelated</u>
(If $\ne 0$, then <u>correlated</u>.)

Fact: If X and Y are independent, then they are <u>un</u>correlated.
(But the converse is false!)

Counterexample

Suppose X and Z are independent and $Z = \begin{cases} +1 & \text{with prob } 1/2 \\ -1 & \text{with prob } 1/2 . \end{cases}$

Define $Y = ZX$.

Then $EY \underset{\substack{\uparrow \\ \text{(by independence)}}}{=} (EZ)(EX) = 0 \cdot EX = 0$

$$\operatorname{cov}(X,Y) = E\,XY - (EX)\underbrace{(EY)}_{0}$$

$$= E\,X(ZX) = E\,ZX^2$$

$$= (EZ)(EX^2) \text{ by independence}$$

$$= 0 \cdot EX^2 = 0.$$

Thus $\rho_{XY} = 0$, but X and Y are clearly <u>not</u> independent since $Y = \pm X$ with probability one.

<u>Formal Proof</u>: Choose $a > 0$ such that $p \equiv P(|X| < a)$ satisfies $0 < p < 1$.

Since $|X| = |Y|$ with probability one,

$$P(|X| < a, |Y| \geq a) = 0.$$

But

$$P(|X| < a)\, P(|Y| \geq a)$$
$$= P(|X| < a)\, P(|X| \geq a)$$
$$= p(1-p) > 0.$$

Thus X and Y are <u>not</u> independent.

Many other similar examples can be constructed. For example, if

$$X \sim N(0,1) \text{ and } Y = |X|$$

then $\mathrm{Cov}(X,Y) = 0$, but X and Y are not independent.

<u>Comment</u>: ρ_{XY} measures the strength (and direction) of the linear relationship between X and Y. ρ_{XY} tells the extent to which the joint distn. follows a straight line with nonzero slope.

Example: (Return to Leaf/Bug situation)
X = Area of Leaf, Y = # of bugs on leaf.
$X \sim \text{Gamma}(\alpha, \beta)$, $Y|X \sim \text{Poisson}(\theta X)$

Find $\text{Cov}(X,Y)$ and ρ_{XY}.

$$\text{Cov}(X,Y) = EXY - \underbrace{EX}_{\alpha\beta}\,\underbrace{EY}_{\theta\alpha\beta} \quad \text{(from before)}$$

$EXY = ?$

Useful General Facts:

(a) $E(g(X)+h(Y)|X) = g(X) + E(h(Y)|X)$

(b) $E(g(X)h(Y)|X) = g(X)E(h(Y)|X)$

Using (b) we have

$$E(XY) = E\,E(XY|X) = E[X \underbrace{E(Y|X)}_{\theta X}]$$

$$= E[\theta X^2] = \theta(EX^2)$$

$$= \theta[\text{Var}X + (EX)^2] = \theta[\alpha\beta^2 + (\alpha\beta)^2].$$

$$\text{Cov}(X,Y) = \theta[\alpha\beta^2 + \alpha^2\beta^2] - \theta\alpha^2\beta^2$$

$$= \theta\alpha\beta^2$$

$$\rho_{XY} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}X \cdot \text{Var}Y}} = \frac{\theta\alpha\beta^2}{\sqrt{\alpha\beta^2 \cdot (\theta\alpha\beta + \theta^2\alpha\beta^2)}}$$

$$= \left(\frac{1}{\theta\beta} + 1\right)^{-1/2}$$

<u>Find $E[(Y-\theta X)\sin X]$.</u>

$$E(Y-\theta X)\sin X$$

$$= E\, E[(Y-\theta X)\sin X \mid X]$$

$$= E[(\sin X)\, \underbrace{E(Y-\theta X \mid X)}_{E(Y|X) - \theta X}]$$

$E(Y|X) - \theta X$ using (a)

$$= \theta X - \theta X = 0$$

$$= E[(\sin X)\cdot 0] = E\,0 = 0$$

We close this section by introducing a very important bivariate distribution in which the correlation coefficient arises naturally as a parameter.

DEFINITION 4.5.3: Let $-\infty < \mu_X < \infty, -\infty < \mu_Y < \infty$, $0 < \sigma_X$, $0 < \sigma_Y$, and $-1 < \rho < 1$ be five real numbers. The *bivariate normal pdf with means μ_X and μ_Y, variances σ_X^2 and σ_Y^2, and correlation ρ* is the bivariate pdf given by

$$f(x, y) = \left(2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}\right)^{-1}$$

$$\times \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right)$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$.

Although the formula for the bivariate normal pdf looks formidable, this bivariate distribution is one of the most frequently used. (In fact, the derivation of the formula need not be formidable at all. See Exercise 4.42.)

The many nice properties of this distribution include these:

a. The marginal distribution of X is $\mathrm{n}(\mu_X, \sigma_X^2)$.
b. The marginal distribution of Y is $\mathrm{n}(\mu_Y, \sigma_Y^2)$.
c. The correlation between X and Y is $\rho_{XY} = \rho$.
d. For any constants a and b, the distribution of $aX + bY$ is $\mathrm{n}(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y)$.

We will leave the verification of properties (a), (b), and (d) as exercises (Exercise 4.41). Assuming (a) and (b) are true, we will prove (c). We have by definition

$$\begin{aligned}
\rho_{XY} &= \frac{\mathrm{Cov}(X, Y)}{\sigma_X\sigma_Y} \\
&= \frac{\mathrm{E}(X-\mu_X)(Y-\mu_Y)}{\sigma_X\sigma_Y} \\
&= \mathrm{E}\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) f(x, y)\, dx\, dy.
\end{aligned}$$

Other properties:

$$(Y|X=x) \sim N\Big(\underbrace{\mu_Y + \rho\sigma_Y\left(\frac{x-\mu_x}{\sigma_X}\right)}_{\text{linear regression}},\ \underbrace{\sigma_Y^2(1-\rho^2)}_{\substack{\text{constant}\\ \text{variance}\\ \text{(not}\\ \text{involving } x)}}\Big)$$

If $(X,Y) \sim$ Biv. Normal, then

(U,V) defined by $U = aX + bY + e$

$\qquad\qquad\qquad\qquad V = cX + dY + f$

is Biv. Normal

(so long as $ad - bc \neq 0$).

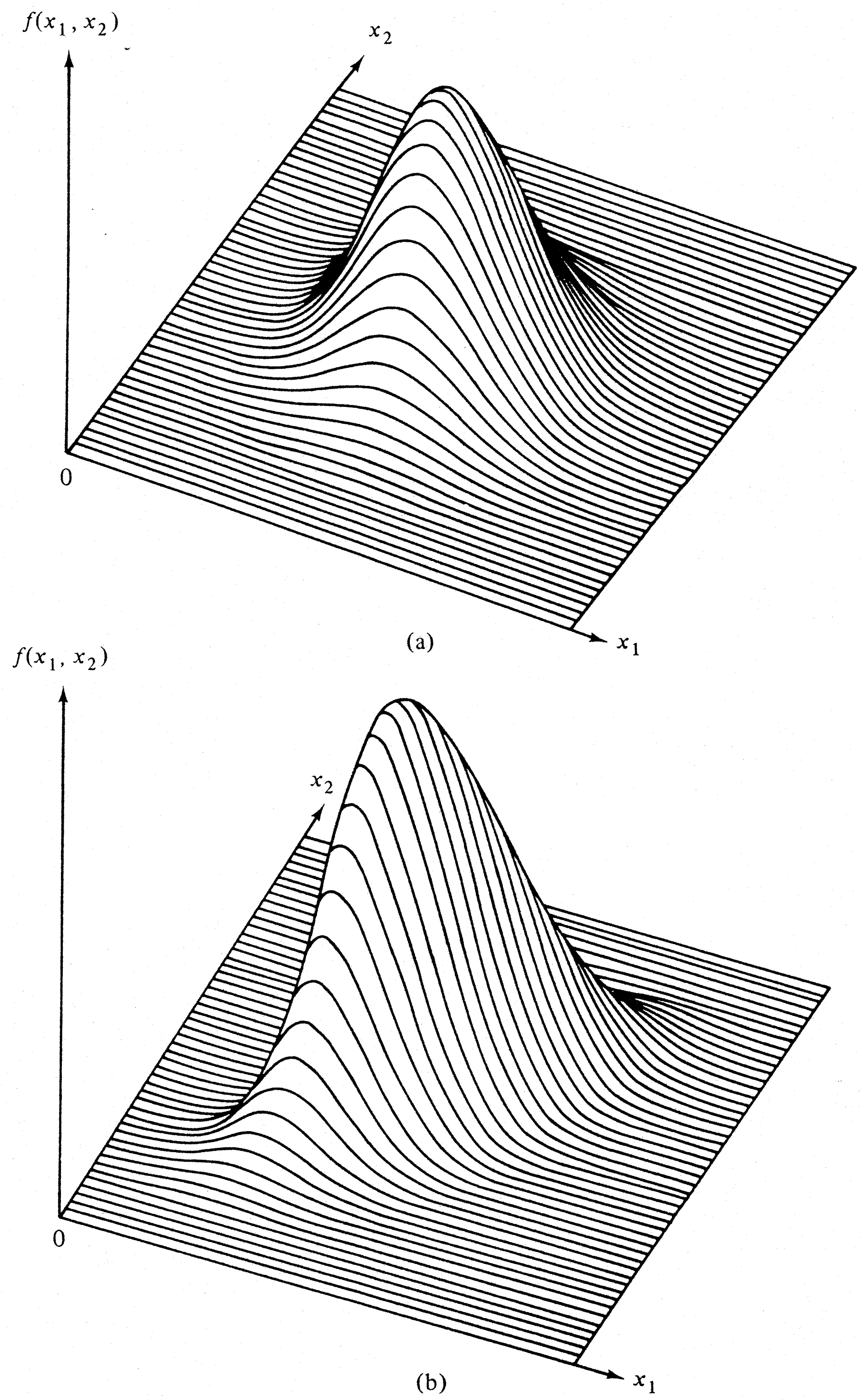

Figure 4.2 Two bivariate normal distributions. (a) $\sigma_{11} = \sigma_{22}$ and $\rho_{12} = 0$. (b) $\sigma_{11} = \sigma_{22}$ and $\rho_{12} = .75$.

The following summarizes these concepts.

> Contours of constant density for the p-dimensional normal distribution are ellipsoids defined by $\mathbf{x}$ such that
>
> $$(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c^2 \tag{4-7}$$
>
> These ellipsoids are centered at $\boldsymbol{\mu}$ and have axes $\pm c\sqrt{\lambda_i}\mathbf{e}_i$, where $\boldsymbol{\Sigma}\mathbf{e}_i = \lambda_i\mathbf{e}_i$, $i = 1, 2, \ldots, p$.

Handy Fact:

If $f(x,y)$ is a pdf and

$$f(x,y) \propto e^{(ax^2+by^2+cxy+dx+ey)} \quad (*)$$

for all x, y,

then $f(x,y)$ is a bivariate normal pdf; it can be re-expressed in the standard form for a bivariate normal pdf with $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$ given in terms of a, b, c, d, e.

Note: $\exp(ax^2+by^2+cxy+dx+ey)$ is proportional to a pdf (can be normalized) if it has a finite integral. This happens iff $a<0, b<0, c^2<4ab$.

Bivariate Normal obtained from hierarchical model

Fact: If $X \sim N(\mu, \xi)$ and $Y|X \sim N(\alpha + \beta X, \lambda)$, then (X, Y) has a bivariate normal distribution for any values of μ, α, β, $\xi > 0$, and $\lambda > 0$.

Proof: $f_{X,Y}(x, y) = f_X(x) f_{Y|X}(y|x)$

$$= \frac{1}{\sqrt{2\pi\xi}} \exp\left(-\frac{(x-\mu)^2}{2\xi}\right) \cdot \frac{1}{\sqrt{2\pi\lambda}} \exp\left(-\frac{(y-\alpha-\beta x)^2}{2\lambda}\right)$$

$$\propto \exp\left(-\frac{(x-\mu)^2}{2\xi} - \frac{(y-\alpha-\beta x)^2}{2\lambda}\right) \quad \left(\begin{array}{c}\text{expand the squares}\\ \text{and collect terms}\end{array}\right)$$

$$\propto \exp\left(\text{quadratic function of } x \text{ and } y\right).$$

This is a pdf (by construction) and has the form required in the "Handy Fact". QED.

Special case in detail:
If $X \sim N(0, 1)$ and $Y|X \sim N(\beta X, 1 - \beta^2)$ for $-1 < \beta < 1$, then (X, Y) has a bivariate normal distribution with $\mu_X = 0$, $\mu_Y = 0$, $\sigma_X^2 = 1$, $\sigma_Y^2 = 1$, and $\rho = \beta$.

Proof: $f_{X,Y}(x, y) = f_X(x) f_{Y|X}(y|x)$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \cdot \frac{1}{\sqrt{2\pi(1-\beta^2)}} \exp\left(-\frac{(y-\beta x)^2}{2(1-\beta^2)}\right)$$

$$= \frac{1}{2\pi\sqrt{1-\beta^2}} \exp\left\{-\frac{1}{2}\left(x^2 + \frac{y^2 - 2\beta xy + \beta^2 x^2}{1-\beta^2}\right)\right\}$$

$$= \frac{1}{2\pi\sqrt{1-\beta^2}} \exp\left\{-\left(\frac{x^2 - 2\beta xy + y^2}{2(1-\beta^2)}\right)\right\} \quad \text{QED}$$

Manipulating Joint Distributions

Obtaining Marginal Density from Joint Density (continuous case)

$$f_W(w) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{W,X,Y,Z}(w,x,y,z)\,dx\,dy\,dz$$

$$f_{W,Y}(w,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{W,X,Y,Z}(w,x,y,z)\,dx\,dz$$

$$f_{X,Y,Z}(x,y,z) = \int_{-\infty}^{\infty} f_{W,X,Y,Z}(w,x,y,z)\,dw$$

Conditional Densities

$$f_{W|X,Y,Z}(w|x,y,z) = \frac{f_{W,X,Y,Z}(w,x,y,z)}{f_{X,Y,Z}(x,y,z)}$$

$$f_{X,Z|W,Y}(x,z|w,y) = \frac{f_{W,X,Y,Z}(w,x,y,z)}{f_{W,Y}(w,y)}$$

$$f_{X,Y,Z|W}(x,y,z|w) = \frac{f_{W,X,Y,Z}(w,x,y,z)}{f_W(w)}$$

Joint density as Product

$$\begin{aligned} &f_{W,X,Y,Z}(w,x,y,z) \\ &\quad = \quad f_W(w) f_{X|W}(x|w) f_{Y|W,X}(y|w,x) f_{Z|W,X,Y}(z|w,x,y) \end{aligned}$$