ANIMAL BREEDING NOTES

CHAPTER 7

EXPECTATION, VARIANCES, AND COVARIANCES OF RANDOM VARIABLES AND RANDOM VECTORS

Expected value of a random variable

Discrete random variable: the expected value of a discrete random variable X, whose probability mass function is p(x), is denoted by E[X] and given by

$$E[X] \quad = \quad \sum_{x:p(x)>0} x\,p(x) \quad \equiv \quad \mu$$

i.e., the expected value of X is a **weighted average** of the possible values that X can have, **each value being weighted by the probability that X assumes it.**

Continuous random variable: if X is a continuous random variable having a probability density function f(x), then because $f(x)\,dx \approx P\{x \le X \le x + dx\}$ for dx small, it is reasonable to define E[X] as follows:

$$E[X] \quad = \quad \int_{-\infty}^{\infty} x\,f(x)\,dx \quad \equiv \quad \mu$$

Animal Breeding example (continued)

The expected value of X, i.e., the weighted average genetic value of the chromosomes of bull B is:

$$E[X] \quad = \quad 10\,(0.2) + 20(0.5) + 30(0.3)$$

$$= \quad 21$$

Similarly, E[Y], the weighted average genetic value of the chromosomes for cow C, is:

$$E[Y] \quad = \quad 10(0.3) + 20(0.5) + 30(0.2)$$

$$= \quad 19$$

Expected value of a function of a random variable

Let X be a **discrete random variable** with probability mass function p(x). Then, the expectation of a function g of X is:

$$E[g(X)] = \sum_{x:p(x)>0} g(x)\ p(x)$$

Let X be a **continuous random variable** with probability density function f(x). Then, the expectation of a function g of X is:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\ f(x)\ dx$$

Example:

The expected value of $g(X) = X^n$, $n \geq 1$, the n^{th} moment of X, is:

$$E[X^n] = \sum_{x:p(x)>0} x^n\ p(x) \qquad \text{if X is discrete}$$

$$E[X^n] = \int_{-\infty}^{\infty} x^n\ f(x)\ dx \qquad \text{if X is continuous}$$

The **expected value of the linear function aX + b with respect to X**, where a and b are constants, is:

$$E[aX + b] = aE[X] + b$$

Proof:

1) **Discrete case**:

$$\begin{aligned} E[aX + b] &= \sum_{x:p(x)>0} (ax + b)\, p(x) \\ &= a \sum_{x:p(x)>0} x\, p(x) + b \sum_{x:p(x)>0} p(x) \\ &= a\, E[X] + b\,(1) \\ &= a\, E[X] + b \end{aligned}$$

2) Continuous case:

$$E[aX + b] = \int_{-\infty}^{\infty} (ax + b)\, f(x)\, dx$$

$$= a \int_{-\infty}^{\infty} x\, f(x)\, dx + b \int_{-\infty}^{\infty} f(x)\, dx$$

$$= a\,E[X] + b\,(1)$$

$$= a\,E[X] + b$$

The **expected value of $(X–E[X])^2$** is the **variance of X**, i.e., **var(X)**, where X has density f(x). The **var(X)** is equal to:

$$\text{var}(X) = E[(X - E[X])^2]$$

$$= E[X^2] - (E[X])^2$$

Proof:

$$\text{var}(X) = E[(X - E[X])^2]$$

$$= E[X^2 - 2XE[X] + (E[X])^2]$$

$$= \int_{-\infty}^{\infty} x^2 f(x)\, dx \; - \; 2E[X] \int_{-\infty}^{\infty} x\, f(x)\, dx \; + \; (E[X]^2 \int_{-\infty}^{\infty} f(x)\, dx$$

$$= E[X^2] - 2(E[X])^2 + (E[X])^2$$

$$= E[X^2] - (E[X])^2$$

The **variance of the linear function aX + b**, where a and b are constants, is:

$$\text{var}(aX + b) = a^2\, \text{var}(X)$$

Proof:

$$\text{var}(aX + b) = E[(aX + b - (aE[X] + b))^2]$$

$$= E[(aX - aE[X])^2]$$

$$= E[a^2(X - E[X])^2]$$

$$= \quad a^2 \operatorname{var}(X)$$

Expected value of a sum of random variables: consider two random variables, X and Y. By the expectation of a function of a random variable,

$$E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)\, f(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f(x,y)\, dx\, dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} x\, f_X(x)\, dx + \int_{-\infty}^{\infty} y\, f_Y(y)\, dy$$

$$= E[X] + E[Y]$$

Thus,

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

If **X and Y are independent**, then the **expectation of the product of (any) functions g(X) and h(Y)** is:

$$E[g(X)\, h(Y)] = E[g(X)]\, E[h(Y)]$$

Proof:

Suppose that X and Y are jointly continuous with density f(x, y). Then,

$$E[g(X)h(Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)\, h(y)\, f(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)\, h(y)\, f(x)\, f(y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} g(x)\, f(x)\, dx \int_{-\infty}^{\infty} h(y)\, f(y)\, dy$$

$$= E[g(X)]\, E[h(Y)]$$

The **covariance of two random variables X and Y**, i.e., **cov(X, Y)**, is defined by:

$$\begin{aligned} \mathrm{cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - X\,E[Y] - E[X]\,Y + E[X]\,E[Y]] \\ &= E[XY] - E[X]\,E[Y] - E[X]\,E[Y] + E[X]\,E[Y] \\ &= E[XY] - E[X]\,E[Y] \end{aligned}$$

Remark: If **X and Y are independent, then cov(X,Y) = 0**. The converse is **not** true.

The **variance of a sum of random variables**, $X_1 + X_2 + \dots + X_n$, is equal to:

$$\mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{var}(X_i) + 2\sum_{i=1}^{n}\sum_{i=1}^{n} \mathrm{cov}(X_i, X_j)$$

Proof:

Consider two random variables, X and Y,

$$\begin{aligned} \mathrm{var}(X + Y) &= E[(X + Y - E[X + Y])^2] \\ &= E[((X - E[X]) + (Y - E[Y]))^2] \\ &= E[(X - E[X])^2] + E[(X - E[X])(Y - E[Y])] \\ &\quad + E[(Y - E[Y])(X - E[X])] + E[(Y - E[Y])^2] \\ &= \mathrm{var}(X) + 2\,\mathrm{cov}(X, Y) + \mathrm{var}(Y) \end{aligned}$$

By induction we get the result above.

Also, note that:

$$\mathrm{var}(X - Y) = \mathrm{var}(X) - 2\,\mathrm{cov}(X, Y) + \mathrm{var}(Y)$$

Conditional expectation

The **conditional expectation of X given Y = y** is :

$$E[X \mid Y = y] = \sum_{x:p(x)>0} x\,P\{X = x \mid Y = y\}$$

$$= \sum_{x:p(x)>0} x\, p_{X|Y}(x \mid y) \qquad \text{for all y such that } p_Y(y) > 0,$$

for the discrete case

and

$$E[X \mid Y = y] = \int_{-\infty}^{\infty} x\, f(x \mid y) f(x \mid y) dx \qquad \text{provided that } f_Y(y) > 0,$$

for the continuous case

The **conditional expectation of g(X) given Y = y** is:

$$E[g(X) \mid Y = y] = \sum_{x:p(x)>0} g(x)\, p_{X|Y}(x \mid y) \qquad \text{for the discrete case}$$

and

$$E[g(X) \mid Y = y] = \int_{-\infty}^{\infty} g(x)\, f_{X|Y}(x \mid y)\, dx \qquad \text{for the continuous case}$$

Remark: the conditional expectation of a sum of random variables is equal to the sum of the conditional expectations of the individual random variables, i.e.,

$$E\left[\sum_{i=1}^{n} X_i \mid Y = y\right] = \sum_{i=1}^{n} E\left[X_i \mid Y = y\right]$$

Computing probabilities by conditioning

Let E be an arbitrary event and define an indicator random variable X by

$$X = \begin{cases} 1 & \text{if E occurs} \\ 0 & \text{if E does not occur} \end{cases}$$

Then,

$$\begin{aligned} E[X] &= (1)*P(X = E) + (0)*P(X \neq E) \\ &= P(E) \end{aligned}$$

$$E[X \mid Y = y] = P\{X = E \mid Y = y\} + P\{X \neq E \mid Y = y\}$$

$$= P\{E \mid Y = y\}$$

Using the formulae for computation of expectations by conditioning above we get:

$$P(E) = \sum_{x:p(x)>0} P\{E \mid Y = y\}\, P\{Y = y\} \qquad \text{if Y is discrete}$$

and

$$P(E) = \int_{-\infty}^{\infty} P\{E \mid Y = y\}\, f_Y(y)\, dy \qquad \text{if Y is continuous}$$

Example 1: Let X and Y be **independent** random variables with densities $f_X(x)$ and $f_Y(y)$.

Compute $P\{X < Y\}$.

$$P\{X < Y\} = \int_{-\infty}^{\infty} P(X < Y \mid Y = y)\, f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} P(X < y \mid Y = y)\, f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} P(X < y)\, f_Y(y)\, dy \qquad \text{by independence}$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{y} f_X(x)\, dx \right] f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} F_X(y)\, f_Y(y)\, dy$$

Remark: If X and Y are **not** independent, then

$$P\{X < Y\} = \int_{-\infty}^{\infty} P(X < y \mid Y = y)\, f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{y} f_{X|Y}(x \mid y)\, dx \right] f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} F_{X|Y}(x \mid y)\, f_Y(y)\, dy$$

Example 2: Let X and Y be **independent** random variables. Find the distribution of X + Y, i.e., find P{X + Y < a}. Condition X on Y.

$$P\{X + Y < a\} = \int_{-\infty}^{\infty} P\{X + Y < a \mid Y = y\}\, f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} P\{X + y < a \mid Y = y\}\, f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} P\{X < a - y \mid Y = y\}\, f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} P\{X < a - y\}\, f_Y(y)\, dy \qquad \text{by independence}$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{a-y} f_X(x)\, dx \right] f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} F_X(a - y)\, f_Y(y)\, dy$$

If X and Y are **not** independent

$$P\{X + Y < a\} = \int_{-\infty}^{\infty} P\{X < a - y \mid Y = y\}\, f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{a-y} f_{X|Y}(x \mid y)\, dx \right] f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} F_{X|Y}(a - y \mid y)\, F_Y(y)\, dy$$

Computing expectations by conditioning

Sometimes it is easier to compute the expectation of a random variable by conditioning it on another. Let $E[X|Y]$ be the function of a random variable X, whose value at Y = y is $E[X|Y = y]$. Note that $E[X|Y]$ is itself a random variable.

Then,

$$E[X] = E_Y[E[X|Y]]$$

Thus, for a discrete random variable

$$E[X] = \sum_{x:p(x)>0} E[X|Y = y]\, P(Y = y)$$

and for a continuous random variable

$$E[X] = \int_{-\infty}^{\infty} E[X|Y = y]\, f_Y(y)\, dy$$

Proof: For X and Y continuous,

$$\begin{aligned}\int_{-\infty}^{\infty} E[X|Y = y]\, f_Y(y)\, dy &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, f_Y(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} x\left[\int_{-\infty}^{\infty} f(x, y)\, dy\right] dx \\ &= \int_{-\infty}^{\infty} x\, f_X(x)\, dx \\ &= E[X]\end{aligned}$$

Animal Breeding example (continued)

The variance of the genetic values of chromosomes from bull B, var(X), is:

$$\mathrm{var}(X) = E[X^2] - (E[X])^2$$

$$E[X^2] = \sum_{i=1}^{n} x^2\, p(x_i)$$

$$= (10)^2 (0.2) + (20)^2 (0.5) + (30)^2 (0.3)$$

$$= 490$$

and

$$E[X] = 21$$

$$\Rightarrow \quad var(X) = 490 - (21)^2$$

$$= 490 - 441$$

$$= 49$$

Similarly, the variance of the genetic values of chromosomes from cow C, var(Y), is:

$$E[Y^2] = (10)^2 (0.3) + (20)^2 (0.5) + (30)^2 (0.2)$$

$$= 410$$

and

$$E[Y] = 19$$

$$\Rightarrow \quad var(Y) = 410 - (19)^2$$

$$= 410 - 361$$

$$= 49$$

The cov(X,Y) is:

$$cov(X,Y) = E[XY] - E[X]E[Y]$$

$$E[XY] = \sum_{i=1}^{3}\sum_{j=1}^{3} x_i\, y_j\, p(x_i, y_j)$$

$$= (10)(10)(0.06) + (10)(20)(0.15) + (10)(30)(0.09)$$

$$+ (20)(10)(0.10) + (20)(20)(0.25) + (20)(30)(0.15)$$

$$+ (30)(10)(0.04) + (30)(20)(0.10) + (30)(30)(0.06)$$

$$= 399$$

$$\text{cov}(X,Y) = 399 - (21)(19)$$

$$= 399 - 399$$

$$= 0 \quad \text{as expected because of the independence of X and Y}$$

The var(X + Y) is:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\,\text{cov}(X, Y)$$

$$= 49 + 49 + 2(0)$$

$$= 98$$

and the var(X – Y) is:

$$\text{var}(X - Y) = 49 + 49 - 2(0)$$

$$= 98$$

$$= \text{var}(X + Y) \quad \text{because X and Y are independent}$$

The $E[X \mid Y = y]$ for y = 10, 20, 30, are computed using the formula:

$$E[X \mid Y = y] = \sum_{i=1}^{3} x_i\, p_{X|Y}(x_i \mid y)$$

The conditional probability mass function of X | Y is:

x_i	$p_{X\mid Y}(x \mid 10)$	$p_{X\mid Y}(x \mid 20)$	$p_{X\mid Y}(x \mid 30)$
10	(0.06/0.3) = 0.2	(0.10/0.5) = 0.2	(0.04/0.2) = 0.2
20	(0.15/0.3) = 0.5	(0.25/.05) = 0.5	(0.10/0.2) = 0.5
30	(0.09/0.3) = 0.3	(0.15/0.5) = 0.3	(0.06/0.2) = 0.3

$$E[X \mid Y = 10] = (10)(0.2) + (20)(0.5) + (30)(0.3)$$

$$= 21$$

$$= E[X \mid Y = 20]$$

$$= E[X \mid Y = 30]$$

$$= E[X] \qquad \text{because X does not depend on Y}$$

Similarly,

$$E[Y \mid X = 10] = E[Y \mid X = 20]$$

$$= E[Y \mid X = 30]$$

$$= E[Y]$$

$$= 19$$

$$\text{Var}(X \mid Y = 10) = E[(X \mid Y - E[X \mid Y])^2]$$

$$= E[(X \mid Y)^2] - (E[X \mid Y])^2$$

For instance, the **var(X | Y = 10)** is:

$$\text{var}(X \mid Y = 10) = E[(X \mid Y = 10)^2] - (E[X \mid y = 10])^2$$

$$= \sum_{i=1}^{3} x^2\, p_{X \mid Y}(x \mid 10) - (E[X \mid y = 10)^2$$

$$= [(10)^2\,(0.2) + (20)^2\,(0.5) + (30)^2\,(0.3)] - (21)^2$$

$$= 490 - 441$$

$$= 49$$

$$= \text{var}(X) \qquad \text{because X is independent of Y}$$

The E[X] computed as $E_Y[E[X \mid Y]]$ is:

$$E[X] = \sum_{j=1}^{3} E[X \mid y_j]\, p_Y(y_j)$$

$$= \sum_{j=1}^{3} \left[\sum_{i=1}^{3} x_i p_{X|Y}(x_i \mid y_j) \right] p_Y(y_j)$$

$$
\begin{aligned}
&= [(10)(0.2) + (20)(0.5) + (30)(0.3)](0.3) \\
&\quad + [(10)(0.2) + (20)(0.5) + (30)(0.3)](0.5) \\
&\quad + [(10)(0.2) + (20)(0.5) + (30)(0.3)](0.2) \\
&= [21](0.3) + [21](0.5) + [21](0.2) \\
&= [21][0.3 + 0.5 + 0.2] \\
&= 21
\end{aligned}
$$

The P(X = 10) computed as P(X = 10 | Y = y) is:

$$
\begin{aligned}
P(X = 10) &= \sum_{j=1}^{3} p_{X|Y}(x = 10)\, p_Y(y_j) \\
&= (0.06/0.3)(0.3) + (0.10/0.5)(0.5) + (0.04/0.2)(0.2) \\
&= 0.06 + 0.10 + 0.04 \\
&= 0.20
\end{aligned}
$$

Expectation and covariances of random vectors

1) The **expectation of a random vector $\mathbf{x}_{n\times 1}$** is defined to be the vector of expectations of its elements, i.e., E[each random variable in x],

$$
E[x] = E\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_n] \end{bmatrix} \equiv \mu
$$

2) Let x be a random vector with $E[x] = \mu$. Then, the covariance matrix of vector x is V, and it is defined as:

$$
V = E[(x-\mu)(x-\mu)']
$$

$$V = E\left[\begin{bmatrix}(x_1-\mu_1)\\(x_2-\mu_2)\\\vdots\\(x_n-\mu_n)\end{bmatrix}\begin{bmatrix}(x_1-\mu_1) & (x_2-\mu_2) & \cdots & (x_n-\mu_n)\end{bmatrix}\right]$$

$$V = \begin{bmatrix} E\left[(x_1-\mu_1)^2\right] & \cdots & E\left[(x_1-\mu_1)(x_n-\mu_n)\right] \\ \vdots & \ddots & \vdots \\ E\left[(x_n-\mu_n)(x_1-\mu_1)\right] & \cdots & E\left[(x_n-\mu_n)^2\right] \end{bmatrix}$$

$$V = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

Animal Breeding example (continued)

Let $x = \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

(a) $E[x] = \begin{bmatrix} E[X] \\ E[Y] \end{bmatrix} = \begin{bmatrix} E[x_1] \\ E[x_2] \end{bmatrix} = \begin{bmatrix} 21 \\ 19 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

(b) $V = \begin{bmatrix} E\left[(x_1-\mu_1)^2\right] & E\left[(x_1-\mu_1)(x_2-\mu_2)\right] \\ E\left[(x_2-\mu_2)(x_1-\mu_1)\right] & E\left[(x_2-\mu_2)^2\right] \end{bmatrix}$

$$V = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

$$V = \begin{bmatrix} 49 & 0 \\ 0 & 49 \end{bmatrix}$$

(c) $E[x_1 \mid x_2] = \sum_{x_1 : p(x_1) > 0} x_1\, p_{X_1|X_2}(x_1 \mid x_2)$ for the discrete case

$E[x_1 \mid x_2] = \int_{-\infty}^{\infty} x_1\, f_{X_1|X_2}(x_1 \mid x_2)\, dx_1$ for the continuous case

$E[x_1 \mid x_2 = 20] = (10)(0.2) + (20)(0.5) + (30)(0.3) = 21$

(d) $\mathrm{var}(x_1 | x_2 = 20) = \{[(10)^2 (0.2) + (20)^2 (0.5) + (30)^2 (0.3)] - (21)^2\}$

$= 490 - 441$

$= 49$

3) Let x be an n × 1 random vector, i.e., x = $[x_1, x_2, \ldots, x_n]$, where the $\{x_i\}$ are the realized values of the set of random variables $\{X_i\}$, then the **cumulative distribution function (c.d.f.) of the random vector x** is the joint c.d.f.

$$P\{X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n\} = F(x_1, x_2, \ldots, x_n)$$

$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(x_1, x_2, \ldots, x_n)\, dx_1, dx_2, \ldots, dx_n$$

where

$$f(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1\, \partial x_2 \cdots \partial x_n} F(x_1, x_2, \ldots, x_n)$$

and

$$f(x_1, x_2, \ldots, x_n) \ge 0 \quad \text{for } -\infty \le x_i \le \infty \text{ and for all } i$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_1, dx_2, \ldots, dx_n = 1$$

The **marginal density function of the last (n-k) x's** is $f(x_1, x_2, \ldots, x_n)$ after integrating out the first k x's, i.e., the marginal of $x_{k+1}, x_{k+2}, \ldots, x_n$ is:

$$g(x_{k+1}, x_{k+2}, \ldots, x_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n)\, dx_1 \ldots dx_k$$

The **conditional distribution of the first k x's given that last (n-k) x's** is the ratio of

$$f(x_1, \ldots, x_k \mid x_{k+1}, \ldots, x_n) = \frac{f(x_1, x_2, \cdots, x_n)}{g(x_{k+1}, x_{k+2}, \cdots, x_n)}$$

The **expected value of x_i^m**, i.e., $E[x_i^m]$, is

$$E[x_i^m] \;=\; x_i^m\, f(x_1, x_2, \ldots, x_n)\, dx_1, dx_2, \ldots, dx_n$$

If m = 1, then $E[X_i] = \mu_i$.

The **covariance between variables i and j**, i.e., $\sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$ is:

$$\sigma_{ij} \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} (x_i - \mu_i)(x_j - \mu_j)\, f(x_1, x_2, \ldots, x_n)\, dx_1, dx_2, \ldots, dx_n$$

Similar expressions for $E[x_i^m \mid x_{k+1}, \ldots, x_n]$ and $E[(x_i - \mu_i)(x_j - \mu_j) \mid x_{k+1}, \ldots, x_n]$ can be written using $f(x_1, \ldots, x_k \mid x_{k+1}, \ldots, x_n)$ instead of $f(x_1, x_2, \ldots, x_n)$ in the two previous formulae.

Expectations and covariances of normal random variables and vectors

A) Let X be a **normal random variable**. Then, X is normally distributed with parameters μ and σ^2.

Proof (Ross, 1976):

The density function of normal variable X is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty$$

The expectation of X is:

$$E[X] \;=\; \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} x\, e^{-(x-\mu)^2/2\sigma^2} dx$$

Replacing x by $[(x - \mu) + \mu]$ and letting $y = (x - \mu)$,

$$E[X] \;=\; \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} y\, e^{-(y)^2/2\sigma^2} dy \;+\; \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \mu\, e^{-(x-\mu)^2/2\sigma^2} dx$$

$$E[X] \;=\; \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} y\, e^{-(y)^2/2\sigma^2} dy \;+\; \mu\int_{-\infty}^{\infty} f(x)\, dx$$

where f(x) is the normal density. The first integral is zero by symmetry, and the second integral is equal to μ(1). Thus,

$$E[X] = 0 + \mu(1)$$

$$E[X] = \mu$$

The variance of X is:

$$E[(X - \mu)^2] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 \, e^{-(x-\mu)^2/2\sigma^2} dx$$

Letting y = (x – μ)/σ yields:

$$E[(X - \mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 \, e^{-y^2/2} \, dy$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \left[-y \, e^{-y^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2} \, dy \right] \qquad \text{by integration by parts}$$

$$= \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} \, dy$$

$$= \sigma^2$$

B) Let Z = (X – μ)/σ. Then, Z is a **standard normal random variable** with μ = 0 and $\sigma^2 = 1$, and its density function is:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty < x < \infty$$

The c.d.f. of Z is:

$$\Phi(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2} \, dz$$

and

$$\Phi(-Z) = 1 - \Phi(Z) \qquad -\infty < x < \infty$$

Remark:

$$F_Z(a) = P\left\{\frac{x-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right\}$$

$$= \Phi\left(\frac{a-\mu}{\sigma}\right)$$

C) **Multivariate normal random variables**

(c1) The random vector x = [x_1, x_2, ... , x_n] has a multivariate normal distribution with vector of means æ and covariance matrix V, i.e., x ~ MVN (μ, V), if its density function is:

$$f(x_1, x_2, \ldots, x_n) = \frac{e^{-\frac{1}{2}(x-\mu)' V^{-1} (x-\mu)}}{(2\pi)^{\frac{n}{2}} |V|^{\frac{1}{2}}}$$

where matrix V is positive definite.

Let

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$x_1' = [x_1 \ \ldots \ x_k]$$

$$x_2' = [x_{k+1} \ \ldots \ x_n]$$

Then,

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

and

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{12}' & V_{22} \end{bmatrix}$$

(c2) The **marginal density function of x_1** is:

$$g(x_1) = g(x_1, \ldots, x_k)$$

$$= \frac{\exp\left[-½\,(x_1-\mu_1)'\,V_{11}^{-1}(x_1-\mu_1)\right]}{(2\pi)^{\frac{k}{2}}\,|V_{11}|^{½}}$$

and the **marginal density function of x_2** is:

$$g(x_2) \quad = \quad g(x_{k+1}, \ldots, x_n)$$

$$= \frac{\exp\left[-½\,(x_2-\mu_2)'\,V_{22}^{-1}(x_2-\mu_2)\right]}{(2\pi)^{\frac{n-k}{2}}\,|V_{22}|^{½}}$$

Note that the **marginal densities** of the multivariate normal distribution are themselves **multivariate normal**.

(c3) The conditional density function of x_1 given x_2 is:

$$f(x_1 | x_2) \quad = \quad \frac{f(x)}{g(x_2)}$$

$$= \frac{\exp\left\{-½\,[(x-\mu)'\,V^{-1}(x-\mu) \;-\; (x_2-\mu_2)'\,V_{22}^{-1}(x_2-\mu_2)]\right\}}{(2\pi)^{\frac{k}{2}}\,\left(|V|\,/\,|V_{22}|\right)^{½}}$$

In terms of partitioned matrices, V^{-1} is equal to:

$$V^{-1} \quad = \quad \begin{bmatrix} W_{11} & -W_{11}V_{12}V_{22}^{-1} \\ -V_{22}^{-1}V_{12}'W_{11} & V_{22}^{-1}+V_{22}^{-1}V_{12}'W_{11}V_{12}V_{22}^{-1} \end{bmatrix}$$

where

$$W_{11} \quad = \quad (V_{11} - V_{12}V_{22}^{-1}V_{12}'\,)^{-1}$$

Then, the exponent in $f(x_1 | x_2)$ becomes:

$$[(x_1-\mu_1)' \quad (x_2-\mu_2)'] \begin{bmatrix} W_{11} & -W_{11}V_{12}V_{22}^{-1} \\ -V_{22}^{-1}V_{12}'W_{11} & V_{22}^{-1}+V_{22}^{-1}V_{12}'W_{11}V_{12}V_{22}^{-1} \end{bmatrix} \begin{bmatrix} (x_1-\mu_1) \\ (x_2-\mu_2) \end{bmatrix} - \; (x_2-\mu_2)'\,V_{22}^{-1}(x_2-\mu_2)$$

and simplifies to:

$$[(x_1-\mu_1)' \quad (x_2-\mu_2)'] \begin{bmatrix} I \\ -V_{22}^{-1}V_{12}' \end{bmatrix} W_{11} \begin{bmatrix} I & -V_{12}V_{22}^{-1} \end{bmatrix} \begin{bmatrix} (x_1-\mu_1) \\ (x_2-\mu_2) \end{bmatrix}$$

which is equal to:

$$[(x_1-\mu_1)\ -\ V_{12}V_{22}^{-1}(x_2-\mu_2)]'\ W_{11}\ [(x_1-\mu_1)\ -\ V_{12}V_{22}^{-1}(x_2-\mu_2)]$$

By the Laplace expansion of a determinant (Searle, 1966, pg. 74-76 and 95-96)

$$|V| = \begin{vmatrix} V_{11} & V_{12} \\ V_{12}' & V_{22} \end{vmatrix}$$

$$|V| = \begin{vmatrix} V_{11} & V_{12} \\ V_{12}' & V_{22} \end{vmatrix} \begin{vmatrix} I & 0 \\ -V_{22}^{-1}V_{12}' & I \end{vmatrix}$$

$$|V| = \left| \begin{bmatrix} V_{11} & V_{12} \\ V_{12}' & V_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ -V_{22}^{-1}V_{12}' & I \end{bmatrix} \right|$$

$$|V| = |V_{22}|\,|V_{11}-V_{12}V_{22}^{-1}V_{12}'|$$

$$\Rightarrow \quad |V| = |V_{22}|\,|W_{11}^{-1}|$$

$$\Rightarrow \quad \frac{|V|}{|V_{22}|} = |W_{11}^{-1}|$$

Thus,

$$f(x_1\,|\,x_2) = \frac{\exp\left\{[(x_1-\mu_1)-V_{12}V_{22}^{-1}(x_2-\mu_2)]'\,W_{11}\,[(x_1-\mu_1)-V_{12}V_{22}^{-1}(x_2-\mu_2)]\right\}}{(2\pi)^{\frac{k}{2}}\,|W_{11}^{-1}|^{½}}$$

$\Rightarrow$ The conditional distribution is also multivariate normal, i.e.,

$$x_1\,|\,x_2 \sim \text{MVN}\,[\mu_1 + V_{12}V_{22}^{-1}\,(x_2-\mu_2),\ W_{11}^{-1}]$$

or,

$$x_1\,|\,x_2 \sim \text{MVN}\,[\mu_1 + V_{12}V_{22}^{-1}\,(x_2-\mu_2),\ V_{11} - V_{12}V_{22}^{-1}V_{12}']$$

References

Ross, S. 1976. A First Course in Probability. Macmillan Publishing Co., Inc., NY.

Searle, S. R. 1971. Linear Models. John Wiley and Sons, Inc., NY.