Normal density for X $\longrightarrow$ Joint density for (X, Y)

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

quadratic form

$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}Q(x)} \longrightarrow \frac{1}{2\pi|\Sigma|^{1/2}} e^{-\frac{1}{2}Q(x,y)} = f(x,y) = f(\mathbf{x})$$

vector $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$

$|A| = \det A$ when A is a matrix.

Joint Distributions

A bivariate normal distribution generalizes the concept of normal distribution to bivariate random variables. It requires a matrix formulation of quadratic forms, and it is later studied in relation with a linear transformation of joint densities. An important application for bivariate normal distributions is discussed along with the introduction of conditional expectation and prediction.

Operations in matrices.

Matrices are usually denoted by capital letters, $A, B, C, \ldots$, while scalars (real numbers) are denoted by lower case letters, $a, b, c, \ldots$.

1. Scalar multiplication: $k \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ka & kb \\ kc & kd \end{bmatrix}$
2. Product of two matrices: $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix}$.
3. The transpose of a matrix: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$.

Determinant and inverse of a matrix.

To determine the existence of inverse A; If $\det A \neq 0$ then A^{-1} exists.

1. Determinant of a matrix: $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$

2. The inverse of a matrix: If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $\det A \neq 0$, then the inverse A^{-1} of A is given by

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix},$$

   and satisfies $AA^{-1} = A^{-1}A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. = I$

   Identity matrix.

3. Transpose of $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

   $A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$

Quadratic forms.

$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ Then $A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$. $A = A^T$ if and only if $b = c$.

If a 2-by-2 matrix A satisfies $A^T = A$, then it is of the form $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$, and is said to be *symmetric*. A 2-by-1 matrix $\boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ is called a *column vector*, usually denoted by boldface letters $\boldsymbol{x}, \boldsymbol{y}, \ldots$, and the transpose $\boldsymbol{x}^T = \begin{bmatrix} x & y \end{bmatrix}$ is called the *row vector*. Then, the bivariate function

$$Q(x, y) = \boldsymbol{x}^T A \boldsymbol{x} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + 2bxy + cy^2 \qquad (5.1)$$

is called a *quadratic form*.

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax+by & bx+cy \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x(ax+by) + y(bx+cy)$$

$$= ax^2 + \underbrace{bxy + bxy}_{2bxy} + cy^2$$

Cholesky decomposition.

optional problem

If $ac - b^2 \neq 0$ and $ac \neq 0$, then the symmetric matrix can be decomposed into the following forms known as *Cholesky decompositions*:

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} \sqrt{a} & 0 \\ b/\sqrt{a} & \sqrt{(ac-b^2)/a} \end{bmatrix} \begin{bmatrix} \sqrt{a} & b/\sqrt{a} \\ 0 & \sqrt{(ac-b^2)/a} \end{bmatrix} \tag{5.2}$$

$$= \begin{bmatrix} \sqrt{(ac-b^2)/c} & b/\sqrt{c} \\ 0 & \sqrt{c} \end{bmatrix} \begin{bmatrix} \sqrt{(ac-b^2)/c} & 0 \\ b/\sqrt{c} & \sqrt{c} \end{bmatrix} \tag{5.3}$$

Decomposition of quadratic form.

Corresponding to the Cholesky decompositions of the matrix A, the quadratic form $Q(x, y)$ can be also decomposed as follows:

$ax^2 + 2bxy + cy^2$

$\Downarrow$

$x^T A x$

$\Downarrow$

$[x \; y]\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$

$$Q(x, y) = \left(\sqrt{a}\,x + \frac{b}{\sqrt{a}}\,y\right)^2 + \left(\sqrt{\frac{ac - b^2}{a}}\,y\right)^2 \tag{5.4}$$

$$= \left(\sqrt{\frac{ac - b^2}{c}}\,x\right)^2 + \left(\frac{b}{\sqrt{c}}\,x + \sqrt{c}\,y\right)^2 \tag{5.5}$$

Bivariate normal density.

$f(x) = \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left[-\frac{1}{2}\underbrace{\frac{(x-\mu_x)^2}{\sigma_x^2}}_{Q(x)}\right]$ with parameters μ_x, σ_x^2

constant

quadratic form

The function

$$f(x,y) = \underbrace{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}}_{(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2}Q(x,y)\right) \tag{5.6}$$

with the quadratic form

$$Q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right] \tag{5.7}$$

gives the joint density function of a bivariate normal distribution.

$[x-\mu_x \;\; y-\mu_y]\, \Sigma^{-1} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}$

with parameters $\underbrace{\mu_x, \mu_y}_{\text{mean parameters}}$, $\underbrace{\sigma_x^2, \sigma_y^2}_{\text{variance parameters}}$, ρ

correlation coefficient

Quadratic form of bivariate normal density.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \longrightarrow (x-\mu)(\sigma^2)^{-1}(x-\mu) \xrightarrow[\sigma^2 \to \Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}]{x \to \boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix},\ \mu \to \boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}} (\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) = Q(x,y)$$

Note that the parameters σ_x^2, σ_y^2, and ρ must satisfy $\sigma_x^2 > 0$, $\sigma_y^2 > 0$, and $-1 < \rho < 1$. By defining the 2-by-2 symmetric matrix (also known as *covariance matrix*) and the two column vectors

$$\Sigma = \underbrace{\begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}}_{\text{symmetric matrix}}, \quad \underset{\text{vector}}{\boldsymbol{x}} = \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{and} \quad \underset{\text{vector}}{\boldsymbol{\mu}} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix},$$

the quadratic form can be expressed as

$$\begin{aligned} Q(x,y) &= (\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \\ &= \begin{bmatrix} x-\mu_x & y-\mu_y \end{bmatrix} \underbrace{\begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}}_{\Sigma}^{-1} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}. \end{aligned} \tag{5.8}$$

$$\Sigma^{-1} = \frac{1}{\underbrace{\det \Sigma}_{\sigma_x^2\sigma_y^2 - \rho^2\sigma_x^2\sigma_y^2 = (1-\rho^2)\sigma_x^2\sigma_y^2}} \begin{bmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{bmatrix} = \frac{1}{1-\rho^2} \begin{bmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix}$$

$$\underline{Q(x,y) = (\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

$$= [x-\mu_x \quad y-\mu_y] \left(\frac{1}{1-\rho^2}\right) \begin{bmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}$$

$$= \left(\frac{1}{1-\rho^2}\right) [\underbrace{x-\mu_x}_{X} \quad \underbrace{y-\mu_y}_{Y}] \underbrace{\begin{bmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix}}_{\begin{bmatrix} a & b \\ b & c \end{bmatrix}} \begin{array}{l} \left.\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}\right. \begin{array}{l} \}X \\ \}Y \end{array} \end{array}$$

$$\left\{ \begin{array}{l} [X \ Y] \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \\ = aX^2 + 2bXY + cY^2 \end{array} \right.$$

$$= \frac{1}{1-\rho^2}\left[\frac{1}{\sigma_x^2}(x-\mu_x)^2 - \frac{2\rho}{\sigma_x\sigma_y}(x-\mu_x)(y-\mu_y) + \frac{1}{\sigma_y^2}(y-\mu_y)^2\right]$$

$$\underline{= \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Extended with (x, y)

Marginal density functions.

$$f(x) = \frac{1}{\sqrt{2\pi}\,|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)'(\Sigma)^{-1}(x-\mu)\right]$$

$$(X,Y) \sim f(x,y) \Rightarrow \begin{cases} f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \\ f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx \end{cases}$$

$$f(x,y) = \frac{1}{2\pi(\det\Sigma)^{1/2}} \exp\left\{-\frac{1}{2}\begin{bmatrix} x-\mu_x & y-\mu_y \end{bmatrix} \Sigma^{-1} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}\right\}$$

Suppose that two random variables X and Y has the bivariate normal distribution (5.6). Then their marginal distributions $f_X(x)$ and $f_Y(x)$ respectively become

Problem 4

$$\int_{-\infty}^{\infty} f(x,y)\,dy = f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\}$$ ← Normal with μ_x and σ_x^2

and

$$\int_{-\infty}^{\infty} f(x,y)\,dx = f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}.$$ ← " with μ_y and σ_y^2

Thus, X and Y are normally distributed with respective parameters (μ_x, σ_x^2) and (μ_y, σ_y^2).

$$E[X] = \int_{-\infty}^{\infty} x\, f_X(x)\,dx = \mu_x$$

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu_x)^2 f_X(x)\,dx = \sigma_x^2$$

Independence condition for X and Y.

If X and Y are independent

$$f(x,y) = f_X(x)\, f_Y(y) = \frac{1}{\sigma_x \sigma_y (2\pi)} \exp\left[\underbrace{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}_{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 = Q(x,y) \text{ with } \rho = 0} \right]$$

If $\rho = 0$, then the quadratic form (5.7) becomes

$$Q(x,y) = \left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2,$$

and consequently we have $f(x,y) = f_X(x) f_Y(y)$. Thus, X and Y are independent if and only if $\rho = 0$.

Conditional density functions.

If X and Y are not independent (that is, $\rho \neq 0$), we can compute the conditional density functions $f_{Y|X}(y|x)$ given $X = x$ as

$$\frac{f(x,y)}{f_X(x)} = f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\left(\frac{y - \overbrace{\mu_y - \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)}^{\mu}}{\sigma_y\sqrt{1-\rho^2}}\right)^2\right\}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]$$

which is the normal density function with parameter $(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x), \sigma_y^2(1-\rho^2))$. Similarly, the conditional density function $f_{X|Y}(y|x)$ given $Y = y$ becomes

$$\frac{f(x,y)}{f_Y(y)} = f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\left(\frac{x - \overbrace{\mu_x - \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)}^{\mu}}{\sigma_x\sqrt{1-\rho^2}}\right)^2\right\}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

which is the normal density function with parameter $(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y), \sigma_x^2(1-\rho^2))$.

Covariance calculation.

Let (X, Y) be bivariate normal random variables with parameters $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$. Recall that $f(x,y) = f(x)f(y|x)$, and that $f(y|x)$ is the normal density with mean $\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x)$ and variance $\sigma_y^2(1-\rho^2)$. First we can compute

$$\int_{-\infty}^{\infty} y f(y|x)\,dy - \mu_y \underbrace{\int_{-\infty}^{\infty} f(y|x)\,dy}_{1} = \underbrace{\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)}_{\text{mean parameter}} - \mu_y$$

$$h(x) = \int_{-\infty}^{\infty} (y - \mu_y) f(y|x)\,dy = \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x).$$

μ_x μ_y

Then we can apply it to obtain

$$f(y|x) = \frac{f(x,y)}{f_x(x)} \Rightarrow f(x,y) = f(y|x)\, f_x(x)$$

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

$$= \int_{-\infty}^{\infty} (x - \mu_x)\left[\underbrace{\int_{-\infty}^{\infty} (y - \mu_y) f(y|x)\,dy}_{h(x)}\right] f(x)\,dx$$

$$= \rho\frac{\sigma_y}{\sigma_x}\underbrace{\int_{-\infty}^{\infty} (x - \mu_x)^2 f_x(x)\,dx}_{E[(X-E[X])^2] = Var(X) = \sigma_x^2} = \rho\frac{\sigma_y}{\sigma_x}\sigma_x^2 = \rho\sigma_x\sigma_y$$

If $f(y) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$ then

$$E[Y] = \mu \iff \int_{-\infty}^{\infty} y\, f(y)\, dy = \mu$$

Let $f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\,\underbrace{\sigma_y\sqrt{1-\rho^2}}_{\sigma}} \exp\left[-\frac{(y - \overbrace{\mu_y - \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)}^{\mu})^2}{2\underbrace{\sigma_y^2(1-\rho^2)}_{\sigma^2}}\right]$

$$\int_{-\infty}^{\infty} (y-\mu_y) f_{Y|X}(y|x)\, dy = \underbrace{\int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\, dy}_{\mu = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)} - \mu_y \underbrace{\int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dy}_{=1} = \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$$

Correlation coefficient ρ.

The *correlation coefficient* of X and Y is defined by

$$\rho = \frac{\overbrace{\mathrm{Cov}(X,Y)}^{= \rho \sigma_x \sigma_y}}{\sqrt{\underbrace{\mathrm{Var}(X)}_{\sigma_x^2}\underbrace{\mathrm{Var}(Y)}_{\sigma_y^2}}}.$$

It implies that the parameter ρ of bivariate normal distribution represents the correlation coefficient of X and Y.

Conditional expectation.

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

$\frac{f(x,y)}{f_X(x)}$

Let $f(x, y)$ be a joint density function for random variables X and Y, and let $f_{Y|X}(y|x)$ be the conditional density function of Y given $X = x$ (see Note #2: "conditional density functions"). Then the *conditional expectation* of Y given $X = x$, denoted by $E(Y|X = x)$, is defined as the function $h(x)$ of x:

$$E(Y|X = x) := h(x) = \int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\, dy.$$

In particular, $h(X)$ is the function of the random variable X, and is called the *conditional expectation* of Y given X, denoted by $E(Y|X)$.

Conditional expectation of a function $g(Y)$.

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

Similarly, we can define the conditional expectation for a function $g(Y)$ of the random variable Y by

$$E(g(Y)|X = x) := \int_{-\infty}^{\infty} g(y)\, f_{Y|X}(y|x)\, dy. = h(x)$$

Since $E(g(Y)|X = x)$ is a function, say $h(x)$, of x, we can define $E(g(Y)|X)$ as the function $h(X)$ of the random variable X.

Linearity properties of conditional expectation.

$E[aY + b] = aE[Y] + b$

density!!

Let a and b be constants. By the definition of conditional expectation, it clearly follows that

$$\int_{-\infty}^{\infty}(ay+b)f(y|x)\,dy = a\int_{-\infty}^{\infty} y\, f(y|x)\,dy + b\underbrace{\int_{-\infty}^{\infty} f(y|x)\,dy}_{=1}$$

$$E(aY + b|X = x) = aE(Y|X = x) + b.$$

Consequently, we obtain

$$E(aY + b|X) = aE(Y|X) + b.$$

Law of total expectation.

$E[E(Y|X)] = E[Y]$

Since $E(g(Y)|X)$ is a function $h(X)$ of random variable of X, we can consider "the expectation $E[E(g(Y)|X)]$ of the conditional expectation $E(g(Y)|X)$," and compute it as follows.

$E[h(X)]$

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

$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$

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

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

Thus, we have $\underbrace{E[E(g(Y)|X)]}_{\text{conditional expectation}} = \underbrace{E[g(Y)]}_{\text{expectation}}$.

Conditional variance formula.

Recall: $Var(Y) = E[(Y - E[Y])^2] = E[Y^2] - (E[Y])^2$ $\qquad E[Y] = \int_{-\infty}^{\infty} y f_Y(y)\,dy$

$\updownarrow$

$Var(Y|X) = E\left((Y - E(Y|X))^2 \middle| X\right)$ $\qquad E(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\,dy = h(x)$

The *conditional variance*, denoted by $\mathrm{Var}(Y|X = x)$, can be defined by

$$h(x) = \mathrm{Var}(Y|X = x) = E((Y - E(Y|X = x))^2|X = x) \longleftrightarrow E[(Y - E[Y])^2]$$
$$= E(Y^2|X = x) - (E(Y|X = x))^2, \longleftrightarrow E[Y^2] - (E[Y])^2$$

where the second equality can be obtained from the linearity property. Since $\mathrm{Var}(Y|X = x)$ is a function, say $h(x)$, of x, we can define $\mathrm{Var}(Y|X)$ as the function $h(X)$ of the random variable X.

Conditional variance formula, continued.

Now compute "the variance $\mathrm{Var}(E(Y|X))$ of the conditional expectation $E(Y|X)$" and "the expectation $E[\mathrm{Var}(Y|X)]$ of the conditional variance $\mathrm{Var}(Y|X)$" as follows.

$\mathrm{Var}(Z) = E[Z^2] - (E[Z])^2$

$$
\begin{aligned}
\mathrm{Var}(\overbrace{\underbrace{E(Y|X)}_{\text{conditional expectation}}}^{Z}) &= E[(E(Y|X))^2] - (E[E(Y|X)])^2 \\
&= E[(E(Y|X))^2] - (E[Y])^2 \qquad (1) \\
E[\underbrace{\mathrm{Var}(Y|X)}_{\text{conditional variance}}] &= E[E(Y^2|X)] - E[(E(Y|X))^2] \\
&= E[Y^2] - E[(E(Y|X))^2] \qquad (2)
\end{aligned}
$$

$\mathrm{Var}(Y|X) = E(Y^2|X) - (E(Y|X))^2$

By adding (1) and (2) together, we can obtain

$$\underline{\mathrm{Var}(E(Y|X))} + \underline{E[\mathrm{Var}(Y|X)]} = E[Y^2] - (E[Y])^2 = \underline{\mathrm{Var}(Y)}.$$

A function $g_1(X)$ in conditional expectation.

$E[a\, g_2(Y)] = a\, E[g_2(Y)]$

$E(g_1(x)\, g_2(Y) \mid X=x) = E(\underbrace{g_1(x)}_{\text{constant}}\, g_2(Y) \mid X=x)$

$= g_1(x)\, E(g_2(Y) \mid X=x)$

In calculating the conditional expectation $E(g_1(X)g_2(Y)|X=x)$ given $X=x$, we can treat $g_1(X)$ as the constant $g_1(x)$; thus, we have $\underline{E(g_1(X)g_2(Y)|X=x)} = \underbrace{g_1(x)}_{\text{constant}}E(g_2(Y)|X=x)$. This justifies

$$\underline{E(g_1(X)g_2(Y)|X)} = \underline{g_1(X)E(g_2(Y)|X)}.$$

This extends the formula: $E(c\, g_2(Y) \mid X) = c\, E(g_2(Y) \mid X)$

Prediction.

Predict Y given $X = x$

$E[(Y - g(x))^2]$ should be minimized at the best prediction of Y by $g(x)$

Suppose that we have two random variables X and Y. Having observed $X = x$, one may attempt to "predict" the value of Y as a function $g(x)$ of x. The function $g(X)$ of the random variable X can be constructed for this purpose, and is called a *predictor* of Y. The criterion for the "best" predictor $g(X)$ is to find a function g to minimize the expected square distance $E[(Y - g(X))^2]$ between Y and the predictor $g(X)$.

Prediction, continued.

$(A+B)^2 = A^2 + 2AB + B^2$

We can compute $E[(Y-g(X))^2]$ by applying all the properties of conditional expectation.

$$
\begin{aligned}
\boxed{E[(Y-g(X))^2]} &= E\left[(\overbrace{Y-E(Y|X)}^{A} + \overbrace{E(Y|X)-g(X)}^{B})^2\right] \\
&= E[(Y-E(Y|X))^2] \\
\#1 \quad &\quad + E[2(Y-E(Y|X))(E(Y|X)-g(X)) + (E(Y|X)-g(X))^2] \\
&= E[E((Y-E(Y|X))^2|X)] \\
&\quad + E\Big[2(E(Y|X)-g(X))E((Y-E(Y|X))|X)\Big] \\
&\quad + E[(E(Y|X)-g(X))^2] \\
&= E[\mathrm{Var}(Y|X)] + E[(\overset{=h(x)}{E(Y|X)}-g(X))^2], \quad = 0 \text{ when } g(x) = E(Y|X)
\end{aligned}
$$

which is minimized when $g(x) = E(Y|X=x)$. Thus, we can find $\underbrace{g(X) = E(Y|X)}_{=h(X)}$ as the best predictor of Y.

$$\#1 : \overbrace{E[(\underbrace{Y - E(Y|X)}_{Z})^2]}^{A^2} + \overbrace{2E[(Y - E(Y|X))(E(Y|X) - g(X))]}^{2AB} + \overbrace{E[(E(Y|X) - g(X))^2]}^{B^2}$$

$$= E[\underbrace{E((Y - E(Y|X))^2 | X)}_{Var(Y|X)}] + 2E\Big[\underbrace{E((Y - E(Y|X))(\underbrace{E(Y|X)}_{=h(X)} - g(X)) | X)}_{\underbrace{(E(Y|X) - g(X))}_{g_1(X)} \, \underbrace{E(Y - E(Y|X)|X)}_{= \underbrace{E(Y|X) - \underbrace{E(E(Y|X)|X)}_{E(Y|X)}}_{0}}}\Big] + \; "$$

$$E[Z] = E[E(Z|X)]$$

$$= E[Var(Y|X)] + E[(E(Y|X) - g(X))^2]$$

The best predictor for bivariate normal distributions.

$$E(Y|X=x) = \int_{-\infty}^{\infty} y \underbrace{f(y|x)} \, dx = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x) = h(x)$$

When X and Y has the bivariate normal distribution with parameter $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, we can find the best predictor

$$h(x) = E(Y|X) = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(X - \mu_x), = a_0 + a_1 X$$

which minimizes the expected square distance $E[(Y - g(X))^2]$ to $\sigma^2 = \sigma_y^2(1 - \rho^2)$.

$$E[(Y - g(x))^2] = E[\underbrace{Var(Y|X)}] = \sigma_y^2(1 - \rho^2)$$

$$Var(Y|X=x) = \sigma_y^2(1 - \rho^2)$$

The best linear predictor.

Except for the case of bivariate normal distribution, the conditional expectation $E(Y|X)$ could be a complicated nonlinear function of X. So we can instead consider the best linear predictor

$$g(X) = \alpha + \beta X$$

of Y by minimizing $E[(Y - (\alpha + \beta X))^2]$. First we can observe that

$$\begin{aligned} h(\alpha, \beta) = E[(Y - (\alpha + \beta X))^2] &= E[Y^2 - 2\alpha Y - 2\beta XY + \alpha^2 - 2\alpha\beta X + \beta^2 X^2] \\ &= E[Y^2] - 2\alpha\mu_y - 2\beta E[XY] + \alpha^2 + 2\alpha\beta\mu_x + \beta^2 E[X^2] \end{aligned}$$

where $\mu_x = E[X]$ and $\mu_y = E[Y]$.

The best linear predictor, continued.

we can obtain the desired values α and β by solving

$$\frac{\partial}{\partial \alpha} E[(Y - (\alpha + \beta X))^2] = -2\mu_y + 2\alpha + 2\beta\mu_x = 0$$
$$\frac{\partial}{\partial \beta} E[(Y - (\alpha + \beta X))^2] = -2E[XY] + 2\alpha\mu_x + 2\beta E[X^2] = 0$$

The solution can be expressed by using $\sigma_x^2 = \mathrm{Var}(X)$, $\sigma_y^2 = \mathrm{Var}(Y)$, and $\rho = \mathrm{Cov}(X, Y)/(\sigma_x \sigma_y)$, and obtain

$$\beta = \frac{E[XY] - E[X]E[Y]}{E[X^2] - (E[X])^2} = \rho \frac{\sigma_y}{\sigma_x}$$
$$\alpha = \mu_y - \beta\mu_x = \mu_y - \rho \frac{\sigma_y}{\sigma_x} \mu_x$$

In summary, the best linear predictor $g(X)$ becomes

$$g(X) = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(X - \mu_x).$$

$$\begin{cases} -\mu_y + \alpha + \beta\mu_x = 0 & \text{— } ① \\ -E[XY] + \alpha\mu_x + \beta E[X^2] = 0 & \text{— } ② \end{cases}$$

$$\Big\downarrow \; ② - \mu_x \times ①$$

$$-)\quad -\mu_x\mu_y + \alpha\mu_x + \beta\mu_x^2 = 0 \quad \text{— } ① \times \mu_x$$

$$\mu_x\mu_y - E[XY] + \beta(E[X^2] - \mu_x^2) = 0$$

$$\beta = \frac{E[XY] - \mu_x\mu_y}{E[X^2] - \mu_x^2} = \frac{Cov(X,Y)}{Var(X)} = \frac{\rho\sigma_x\sigma_y}{\sigma_x^2} = \frac{\rho\sigma_y}{\sigma_x}$$

By ①

$$\alpha = \mu_y - \beta\mu_x = \mu_y - \frac{\rho\sigma_y}{\sigma_x}\mu_x$$

So that

$$g(X) = \alpha + \beta X = \mu_y - \frac{\rho\sigma_y}{\sigma_x}\mu_x + \frac{\rho\sigma_y}{\sigma_x}X = \mu_y + \frac{\rho\sigma_y}{\sigma_x}(X - \mu_x)$$

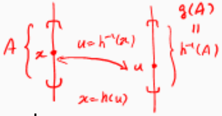
Transformation of variable.

For an interval $A = [a, b]$ on real line with $a < b$, the integration of a function $f(x)$ over A is formally written as

$$\int_A f(x)\,dx = \int_a^b f(x)\,dx. \tag{5.9}$$

Let $h(u)$ be differentiable and strictly monotonic function (that is, either strictly increasing or strictly decreasing) on real line. Then there is the inverse function $g = h^{-1}$, so that we can define the inverse image $h^{-1}(A)$ of $A = [a, b]$ by

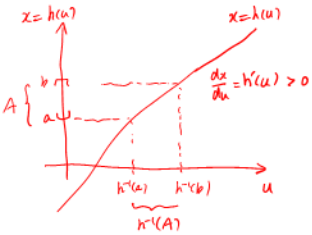
$$h^{-1}(A) = g(A) = \{g(x) : \ a \le x \le b\}.$$

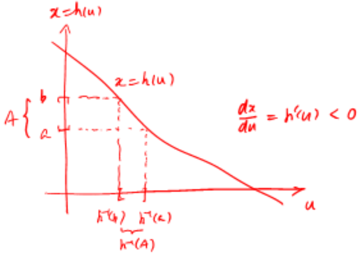


The change of variable in the integral (5.9) provides the formula

$$\boxed{\int_A f(x)\,dx = \int_{g(A)} f(h(u)) \left| \frac{d}{du} h(u) \right| \, du. \qquad (5.10)}$$

$$\int f(x)\,dx = \int f(x) \frac{dx}{du}\,du \overset{x = h(u)}{=} \int f(h(u)) \left[\frac{d}{du} h(u)\right] du$$





$$\int_A f(x)\,dx = \int_a^b f(x)\,dx = \int_{h^{-1}(a)}^{h^{-1}(b)} f(h(u))\,h'(u)\,du = \int_{h^{-1}(A)} f(h(u))\,h'(u)\,du$$

$$\int_A f(x)\,dx = \int_a^b f(x)\,dx = \int_{h^{-1}(a)}^{h^{-1}(b)} f(h(u))\,h'(u)\,du = -\int_{h^{-1}(b)}^{h^{-1}(a)} f(h(u))\,h'(u)\,du = \int_{h^{-1}(A)} f(h(u))\,[-h'(u)]\,du$$

$$\int_{h^{-1}(A)} f(h(u))\,|h'(u)|\,du$$

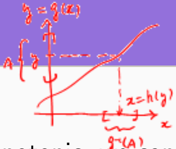
Transformation theorem of random variable.

Theorem

Let X be a random variable with pdf $f_X(x)$. Suppose that a function $g(x)$ is differentiable and strictly monotonic. Then the random variable $\boxed{Y = g(X)}$ *has the pdf*

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| . \tag{5.11}$$

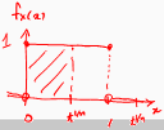
A sketch of proof.

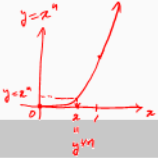


Since $g(x)$ is strictly monotonic, we can find the inverse function $h = g^{-1}$, and apply the formula (5.10). Thus, by letting $A = (-\infty, t]$, we can rewrite the cdf for Y as

$$\begin{aligned} F_Y(t) &= P(Y \le t) = P(Y \in A) \\ &= P(g(X) \in A) = P(X \in h(A)) \\ &= \int_{h(A)} f_X(x)\,dx = \int_{g(h(A))} f_X(h(y)) \left| \frac{d}{dy} h(y) \right| dy \\ &= \int_A f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| dy = \int_{-\infty}^{t} \underbrace{f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|}_{f_Y(y)} dy \end{aligned}$$

By comparing it with the form $\int_{-\infty}^{t} f_Y(y)\,dy$, we can obtain (5.11).





Example

Let X be a uniform random variable on $[0,1]$. Suppose that X takes nonnegative values [and therefore, $f(x)=0$ for $x \leq 0$]. Find the pdf $f_Y(y)$ for the random variable $Y = X^n$.

$$F_Y(t) = P(Y \le t) = P(0 \le X^n \le t) = P(0 \le X \le t^{1/n}) \quad \text{for } t \ge 0$$

$$= \int_0^{t^{1/n}} f_X(x)\,dx = \begin{cases} \int_0^{t^{1/n}} 1\,dx = t^{1/n} & \text{if } 0 \le t \le 1 \\ 1 & \text{if } t > 1 \end{cases}$$

$$f_Y(t) = \frac{d}{dt} F_Y(t) = \begin{cases} \frac{1}{n} t^{1/n - 1} & \text{if } 0 \le t \le 1 \\ 0 & \text{if } t > 1 \end{cases}$$

Example

Let X be a uniform random variable on $[0, 1]$. Suppose that X takes nonnegative values [and therefore, $f(x) = 0$ for $x \leq 0$]. Find the pdf $f_Y(y)$ for the random variable $Y = X^n$.

Apply Theorem!

Let $g(x) = x^n$. Then we obtain $h(y) = g^{-1}(y) = y^{1/n}$, and $h'(y) = \frac{1}{n} y^{(1/n)-1}$. Thus, we obtain

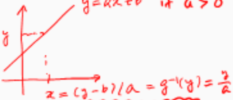
$$f_Y(y) = \frac{1}{n} y^{(1/n)-1} f_X(y^{1/n}) = \begin{cases} \frac{1}{n} y^{(1/n)-1} & \text{if } 0 \leq y \leq 1; \\ 0 & \text{otherwise.} \end{cases}$$

Linear transform of normal random variable.

We already know that: $X \sim N(\mu, \sigma^2)$ then $Y = aX + b \sim N(a\mu+b, a^2\sigma^2)$

Let's prove it again by applying theorem!

$Y = aX + b = g(X)$ with $g(x) = ax + b \leftarrow$ monotonic



$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$

$x = (y-b)/a = g^{-1}(y) = \frac{y}{a} - \frac{b}{a}$

Let X be a normal random variable with parameter (μ, σ^2). Then we can define the linear transform $Y = aX + b$ with real values a and b. To compute the density function $f_Y(y)$ of the random variable Y, we can apply the transformation theorem, and obtain

$$f_Y(y) = \frac{1}{|a|\sigma\sqrt{2\pi}} \exp\left[-\frac{(y - \overbrace{(a\mu + b)}^{\text{mean}})^2}{2\underbrace{(a\sigma)^2}_{\text{variance}}}\right].$$

This is the normal density function with parameter $(a\mu + b, (a\sigma)^2)$.

$$f_X(g^{-1}(y)) \left|\frac{d}{dy} g^{-1}(y)\right| = f_X\left(\frac{y-b}{a}\right)\left|\frac{1}{a}\right| = \left|\frac{1}{a}\right| \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\underbrace{-\frac{\left[\left(\frac{y-b}{a}\right)-\mu\right]^2}{2\sigma^2}}_{-\frac{[y-b-a\mu]^2}{2a^2\sigma^2}}\right]$$

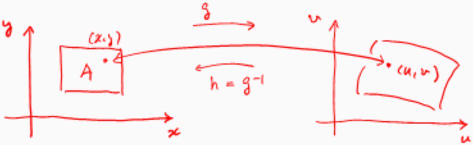
Transformation of two variables.

For a rectangle $A = \{(x, y) : \; a_1 \le x \le b_1, \; a_2 \le y \le b_2\}$ on a plane, the integration of a function $f(x, y)$ over A is formally written as

$$\iint_A f(x, y)\, dx\, dy = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x, y)\, dx\, dy \tag{5.12}$$

Suppose that a transformation $(g_1(x, y), g_2(x, y))$ is differentiable and has the inverse transformation $(h_1(u, v), h_2(u, v))$ satisfying

$$\begin{cases} x = h_1(u, v) \\ y = h_2(u, v) \end{cases} \quad \text{if and only if} \quad \begin{cases} u = g_1(x, y) \\ v = g_2(x, y) \end{cases}. \tag{5.13}$$



Transformation of two variables, continued.

We can state the change of variable in the double integral (5.12) as follows: (i) define the inverse image of A by

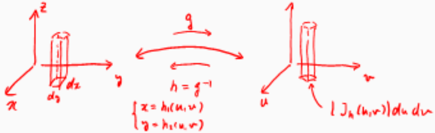
$$h^{-1}(A) = g(A) = \{(g_1(x,y), g_2(x,y)) : \ (x,y) \in A\},$$

(ii) calculate the *Jacobian*

$$J_h(u,v) = \det \begin{bmatrix} \frac{\partial}{\partial u} h_1(u,v) & \frac{\partial}{\partial v} h_1(u,v) \\ \frac{\partial}{\partial u} h_2(u,v) & \frac{\partial}{\partial v} h_2(u,v) \end{bmatrix}, \qquad (5.14)$$

(iii) finally, if it does not vanish for all x, y, then we can obtain

$$\int_A f(x,y)\, dx\, dy = \int_{g(A)} f(h_1(u,v), h_2(u,v))\, |J_h(u,v)|\ du\, dv. \qquad (5.15)$$



Theorem

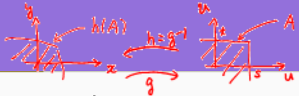
Suppose that (i) random variables X and Y has a joint density function $f_{X,Y}(x,y)$, (ii) a transformation $(g_1(x,y), g_2(x,y))$ is differentiable, and (iii) its inverse transformation $(h_1(u,v), h_2(u,v))$ satisfies (5.13). Then the random variables $U = g_1(X,Y)$ and $V = g_2(X,Y)$ has the joint density function

$$f_{U,V}(u,v) = f_{X,Y}(h_1(u,v), h_2(u,v)) \left| J_h(u,v) \right| . \tag{5.16}$$

Previous theorem : $Y = g(X)$ with $h = g^{-1}$

$f_Y(y) = f_X(h(y)) \, |h'(y)|$

A sketch of proof.



Define $A = \{(x, y) : x \le s,\ y \le t\}$ and the image of A by the inverse transformation $(h_1(x, y), h_2(x, y))$ by

$$h(A) = \{(h_1(u, v), h_2(u, v)) : (u, v) \in A\}.$$

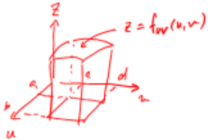
Then we can rewrite the joint distribution function $F_{U,V}$ for (U, V) by applying the formula (5.15).

$$\begin{aligned}
F_{U,V}(s, t) &= P((U, V) \in A) = P((g_1(X, Y), g_2(X, Y)) \in A) \\
&= P((X, Y) \in h(A)) = \iint_{h(A)} f_{X,Y}(x, y)\, dx\, dy \\
&= \int_{g(h(A))} f_{X,Y}(h_1(u, v), h_2(u, v))\, |J_h(u, v)|\, du\, dv \\
&= \int_{A} f_{X,Y}(h_1(u, v), h_2(u, v))\, |J_h(u, v)|\, du\, dv
\end{aligned}$$

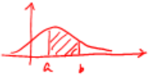
They must be equal

which is equal to $\int_{-\infty}^{t} \int_{-\infty}^{s} f_{U,V}(u, v)\, du\, dv$; thus, we obtain (5.16).

$$P(a \le U \le b,\ c \le V \le d) = \int_a^b \int_c^d f_{UV}(u,v)\,dv\,du$$



$$\left[\ P(a \le U \le b) = \int_a^b f_U(u)\,du \ \right]$$



$$P(-\infty < U \le s,\ -\infty < V \le t) = \int_{-\infty}^{s} \int_{-\infty}^{t} \boxed{f_{UV}(u,v)}\,dv\,du$$

$$\| $$

$$F_{UV}(s,t)$$

joint distribution

This must be a joint density function

Exercises

Problem

Verify the Cholesky decompositions by multiplying the matrices in each of (5.2) and (5.3).

Problem

Show that the right-hand side of (5.4) equals (5.1) by calculating

$$Q(x,y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \sqrt{a} & 0 \\ b/\sqrt{a} & \sqrt{(ac-b^2)/a} \end{bmatrix} \begin{bmatrix} \sqrt{a} & b/\sqrt{a} \\ 0 & \sqrt{(ac-b^2)/a} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Similarly show that the right-hand side of (5.5) equals (5.1).

Problem

Show that (5.8) has the following decomposition

$$[x-\mu_x \;\; y-\mu_y]\,\Sigma^{-1}\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix} = Q(x,y) = \left(\frac{x-\mu_x-\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)}{\sigma_x\sqrt{1-\rho^2}}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 \qquad \text{(E1)}$$

$$= \left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)}{\sigma_y\sqrt{1-\rho^2}}\right)^2 \qquad \text{(E2)}$$

by completing the following:

1. *Show that (5.7) is equal to (5.8).*
2. *Find the Cholesky decompositions of* Σ^{-1}.
3. *Verify (E1) and (E2) by finding the Cholesky decompositions of* Σ^{-1}.

$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}Q(x,y)\right] = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}(A^2+B^2)\right] = \underbrace{\exp\left[-\frac{1}{2}A^2\right]}_{\frac{1}{\sqrt{2\pi}\sigma_x}} \underbrace{\exp\left[-\frac{1}{2}B^2\right]}_{\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}}$

Problem

Compute $f_X(x)$ by completing (i) and (ii). Then find $f_{Y|X}(y|x)$ given $X = x$.

(i) Show that $\int_{-\infty}^{\infty} f(x,y)\,dy$ becomes

$f(x,y) = f(y|x)\,f_X(x)$

$$\int_{-\infty}^{\infty} f(x,y)\,dy = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\} = f_X(x)$$

$$\times \underbrace{\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)}{\sigma_y\sqrt{1-\rho^2}}\right)^2\right\} dy}_{= \int_{-\infty}^{\infty} f(y|x)\,dy \,=\, 1}.$$

by (ii)

$\left(\frac{y-\mu}{\sigma}\right)^2$

$E(Y|X=x) = \mu = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$

(ii) Show that

$$\frac{1}{\sqrt{2\pi}\,\sigma_y\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\left(\frac{y-\overbrace{\mu_y-\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)}^{\mu=\mu_y+\rho\frac{\sigma_x}{\sigma_y}(x-\mu_x)}}{\underbrace{\sigma_y\sqrt{1-\rho^2}}_{=\sigma}}\right)^2\right\}dy$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}u^2\right\}du=1.$$

$f_{Y|X}(y|x)$

$$\int_{-\infty}^{\infty}\underbrace{\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}}_{N(\mu,\sigma^2)}dy=1$$

Goal is to find $f_{Y|X}(y|x)$.

$$f_{Y|X}(x,y)=\frac{f(x,y)}{f_X(x)}=\frac{\cancel{f_X(x)}\,f_{Y|X}(x,y)}{\cancel{f_X(x)}}$$

$$\boxed{\text{(i) and (ii) allows to conclude } f_X(x)=\int_{-\infty}^{\infty}f(x,y)\,dy}$$

Problem

Suppose that the joint density function $f(x, y)$ of X and Y is given by

$$f(x,y) = \begin{cases} 2 & \text{if } 0 \leq x \leq y \leq 1; \\ 0 & \text{otherwise.} \end{cases}$$

1. *Find the covariance and the correlation of X and Y.*
2. *Find the conditional expectations $E(X|Y = y)$ and $E(Y|X = x)$.*
3. *Find the best linear predictor of Y in terms of X.*
4. *Find the best predictor of Y.*

Problem

Suppose that we have the random variables X and Y having the joint density

$$f(x, y) = \begin{cases} \frac{6}{7}(x + y)^2 & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1; \\ 0 & \text{otherwise.} \end{cases}$$

1. *Find the covariance and correlation of X and Y.*
2. *Find the conditional expectation $E(Y|X = x)$.*
3. *Find the best linear predictor of Y.*

Problem

Let X be the age of the pregnant woman, and let Y be the age of the prospective father at a certain hospital. Suppose that the distribution of X and Y can be approximated by a bivariate normal distribution with parameters $\mu_x = 28$, $\mu_y = 32$, $\sigma_x = 6$, $\sigma_y = 8$ and $\rho = 0.8$.

1. *Find the probability that the prospective father is over 30.*
2. *When the pregnant woman is 31 years old, what is the best guess of the prospective father's age?*
3. *When the pregnant woman is 31 years old, find the probability that the prospective father is over 30.*

Problem 8

Suppose that (X_1, X_2) *has a bivariate normal distribution with mean vector* $\boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$ *and covariance matrix* $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$.
Consider the linear transformation

$$Y_1 = g_1(X_1, X_2) = a_1X_1 + b_1;$$
$$Y_2 = g_2(X_1, X_2) = a_2X_2 + b_2.$$

Then answer (a)–(c) in the following.

1. Find the inverse transformation $(h_1(u,v), h_2(u,v))$, and compute the Jacobian $J_h(u,v)$.
2. Find the joint density function $f_{Y_1,Y_2}(u,v)$ for (Y_1, Y_2).
3. Conclude that the joint density function $f_{Y_1,Y_2}(u,v)$ is the bivariate normal density with mean vector $\boldsymbol{\mu} = \begin{bmatrix} a_1\mu_x + b_1 \\ a_2\mu_y + b_2 \end{bmatrix}$ and covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} (a_1\sigma_x)^2 & \rho(a_1\sigma_x)(a_2\sigma_y) \\ \rho(a_1\sigma_x)(a_2\sigma_y) & (a_2\sigma_y)^2 \end{bmatrix}$.

$$|J_h(u,v)|\, f_{XY}(h_1(u,v), h_2(u,v)) = \frac{1}{|a_1 a_2|} f_{XY}\left(\frac{u-b_1}{a_1}, \frac{u-b_2}{a_2}\right)$$

$$= \frac{1}{2\pi\sigma_u\sigma_v\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}[u-\mu_u \;\; v-\mu_v]\, \Sigma^{-1} \begin{bmatrix} u-\mu_u \\ v-\mu_v \end{bmatrix}\right\} \quad \text{with } \Sigma = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}$$

$\mu_u = E[U]$ $\quad \sigma_u^2 = Var(U)$ $\quad \rho = \dfrac{Cov(U,V)}{\sqrt{Var(U)\,Var(V)}}$

$\mu_v = E[V]$ $\quad \sigma_v^2 = Var(V)$

Problem

Let X_1 and X_2 be iid standard normal random variables. Consider the linear transformation

$$Y_1 = g_1(X_1, X_2) = a_{11}X_1 + a_{12}X_2 + b_1;$$
$$Y_2 = g_2(X_1, X_2) = a_{21}X_1 + a_{22}X_2 + b_2,$$

where $a_{11}a_{22} - a_{12}a_{21} \neq 0$. Find the joint density function $f_{Y_1,Y_2}(u, v)$ for (Y_1, Y_2), and conclude that it is the bivariate normal density with mean vector $\boldsymbol{\mu} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ and covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$ where $\sigma_x^2 = a_{11}^2 + a_{12}^2$, $\sigma_y^2 = a_{21}^2 + a_{22}^2$, and
$\rho = \dfrac{a_{11}a_{21} + a_{12}a_{22}}{\sqrt{(a_{11}^2 + a_{12}^2)(a_{21}^2 + a_{22}^2)}}$.

Problem

Suppose that X and Y are independent random variables with respective pdf's $f_X(x)$ and $f_Y(y)$. Consider the transformation

$$U = g_1(X, Y) = X + Y;$$
$$V = g_2(X, Y) = \frac{X}{X + Y}.$$

Then answer (a)–(c) in the following. The density function (5.17) for V is called the **beta distribution** *with parameter (α_1, α_2).*

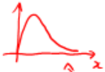
1. Find the inverse transformation $(h_1(u, v), h_2(u, v))$, and compute the Jacobian $J_h(u, v)$.
2. Find the joint density function $f_{XY}(x, y)$ $= f_X(x) f_Y(y)$ for (X, Y).
3. Now let $f_X(x)$ and $f_Y(y)$ be the gamma density functions with respective parameters (α_1, λ) and (α_2, λ). Then show that $U = X + Y$ has the gamma distribution with parameter $(\alpha_1 + \alpha_2, \lambda)$ and that $V = \dfrac{X}{X+Y}$ has the density function

$$f_V(v) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1 - 1}(1 - v)^{\alpha_2 - 1}. \tag{5.17}$$

Solution :

(1) $\begin{cases} u = g_1(x,y) = x+y \\ v = g_2(x,y) = \dfrac{x}{x+y} \end{cases}$ $\quad uv = x$, $\quad y = u - x = u - uv$ $\Rightarrow$ $\begin{cases} x = h_1(u,v) = \underline{uv} \\ y = h_2(u,v) = \underline{u - uv} \end{cases}$

$$J_h(u,v) = \det \begin{bmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{bmatrix} = \det \begin{bmatrix} v & u \\ 1-v & -u \end{bmatrix} = (v)(-u) - u(1-v) = -u$$



(2) $$f_{XY}(x,y) = f_X(x)\, f_Y(y) \underset{\text{If gamma}}{=} \left(\frac{1}{\Gamma(\alpha_1)} \lambda^{\alpha_1} x^{\alpha_1 - 1} e^{-\lambda x}\right)\left(\frac{1}{\Gamma(\alpha_2)} \lambda^{\alpha_2} x^{\alpha_2 - 1} e^{-\lambda x}\right), \quad x \geq 0,\ y \geq 0$$

$$f(x) = \frac{1}{\Gamma(\alpha)} \lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}$$

(3) $$f_{UV}(u,v) = |J_h(u,v)|\, f_X(\underbrace{h_1(u,v)}_{= uv})\, f_Y(\underbrace{h_2(u,v)}_{= u - uv = u(1-v)})$$

$$= \underline{\underline{u}}\, \frac{1}{\Gamma(\alpha_1)} \lambda^{\alpha_1} (\underline{\underline{u}}v)^{\alpha_1 - 1} e^{-\lambda u v} \frac{1}{\Gamma(\alpha_2)} \lambda^{\alpha_2} [\underline{\underline{u}}(1-v)]^{\alpha_2 - 1} e^{-\lambda u(1-v)}$$

$$= \lambda^{\alpha_1+\alpha_2} e^{-\lambda u}\, \underline{\underline{u}}^{\alpha_1+\alpha_2-1} \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1 - 1} (1-v)^{\alpha_2 - 1}$$

$$= \underbrace{\left(\frac{1}{\Gamma(\alpha_1+\alpha_2)} \lambda^{\alpha_1+\alpha_2} u^{\alpha_1+\alpha_2-1} e^{-\lambda u}\right)}_{\text{Gamma with } (\alpha_1+\alpha_2,\ \lambda)} \underbrace{\left(\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1-1} (1-v)^{\alpha_2-1}\right)}_{\text{Beta distribution}}$$

$$= f_U(u) \times f_V(v)$$

Problem

Let X and Y be random variables. The joint density function of X and Y is given by

$$f(x,y) = \begin{cases} xe^{-(x+y)} & \textit{if } 0 \leq x \textit{ and } 0 \leq y; \\ 0 & \textit{otherwise.} \end{cases}$$

1. *Are X and Y independent? Describe the marginal distributions for X and Y with specific parameters.*
2. *Find the distribution for the random variable $U = X + Y$.*

Problem

Consider an investment for three years in which you plan to invest \$500 in average at the beginning of every year. The fund yields the fixed annual interest of 10%; thus, the fund invested at the beginning of the first year yields 33.1% at the end of term (that is, at the end of the third year), the fund of the second year yields 21%, and the fund of the third year yields 10%. Then answer (a)–(c) in the following.

1. How much do you expect to have in your investment account at the end of term?
2. From your past spending behavior, you assume that the annual investment you decide is independent every year, and that its standard deviation is \$200. (That is, the standard deviation for the amount of money you invest each year is \$200.) Find the standard deviation for the total amount of money you have in the account at the end of term.
3. Furthermore, suppose that your annual investment is normally distributed. (That is, the amount of money you invest each year is normally distributed.) Then what is the probability that you have more than \$2000 in the account at the end of term?

Problem

$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$

Let X be an exponential random variable with parameter λ. And define

$$Y = X^n$$

where n is a positive integer.

Gamma distribution with (λ, m)

$f(x) = \frac{\lambda^m}{\Gamma(m)} x^{m-1} e^{-\lambda x}, \quad x \geq 0$

1. *Compute the pdf $f_Y(y)$ for Y.*
2. *Find c so that*

   $\frac{\lambda^{n+1}}{\Gamma(n+1)} = \frac{\lambda^{n+1}}{n!}$

$$f(x) = \begin{cases} cx^n e^{-\lambda x} & \text{if } x \geq 0; \\ 0 & \text{otherwise.} \end{cases}$$

   is a pdf. Hint: You should know the name of the pdf $f(x)$.
3. *By using (b), find the expectation $E[Y]$.*

$$E[Y] = E[X^n] = \int_0^\infty x^n \lambda e^{-\lambda x}\,dx = \frac{n!}{\lambda^n} \underbrace{\int_0^\infty \underbrace{\frac{\lambda^{n+1}}{n!} x^n e^{-\lambda x}}_{\text{Gamma density}}\,dx}_{1} = \frac{n!}{\lambda^n}$$

Problem 14

$f_{XY}(x,y) = f(x)f(y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}[x\ \ y]\underbrace{\begin{bmatrix}1/\sigma_x^2 & 0\\ 0 & 1/\sigma_y^2\end{bmatrix}}_{\Sigma^{-1}}\begin{bmatrix}x\\ y\end{bmatrix}\right]$ $\quad \Sigma = \begin{bmatrix}\sigma_x^2 & 0\\ 0 & \sigma_y^2\end{bmatrix}$

Let X and Y be independent normal random variables with mean zero and the respective variances σ_x^2 and σ_y^2. Then consider the change of variables via

$$\begin{bmatrix}U\\ V\end{bmatrix} = \underbrace{\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}}_{A}\begin{bmatrix}X\\ Y\end{bmatrix} \longleftrightarrow \begin{cases} U &= a_{11}X + a_{12}Y = g_1(X,Y) \\ V &= a_{21}X + a_{22}Y = g_2(X,Y) \end{cases} \xrightarrow{h=g^{-1}} \begin{bmatrix}X\\ Y\end{bmatrix} = A^{-1}\begin{bmatrix}U\\ V\end{bmatrix} \tag{5.18}$$

where $D = a_{11}a_{22} - a_{12}a_{21} \neq 0$. (i) Find the variances σ_u^2 and σ_v^2 for U and V, respectively. (ii) Find the inverse transformation $(h_1(u,v), h_2(u,v))$ so that X and Y can be expressed in terms of U and V by

$$\begin{cases} X &= h_1(U,V); \\ Y &= h_2(U,V). \end{cases}$$

And (iii) compute the Jacobian $J_h(u,v)$. Then continue to answer (a)–(d) in the following.

$= \det A^{-1} \in$ constant

(ii) $\begin{bmatrix} U \\ V \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}}_{A} \begin{bmatrix} X \\ Y \end{bmatrix} \longrightarrow \begin{bmatrix} X \\ Y \end{bmatrix} = A^{-1} \begin{bmatrix} U \\ V \end{bmatrix} = \frac{1}{\det A} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix}$ with $\det A = a_{11} a_{22} - a_{12} a_{21} = D$

$$\rightarrow \begin{cases} x = h_1(U,V) = \dfrac{a_{22} U - a_{12} V}{\det A} \\ y = h_2(U,V) = \dfrac{-a_{21} U + a_{11} V}{\det A} \end{cases} \qquad \begin{bmatrix} h_1(u,v) \\ h_2(u,v) \end{bmatrix} = A^{-1} \begin{bmatrix} u \\ v \end{bmatrix}$$

(iii) $J_h(u,v) = \det \begin{bmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{bmatrix} = \det \begin{bmatrix} a_{22}/D & -a_{12}/D \\ -a_{21}/D & a_{11}/D \end{bmatrix} = \dfrac{a_{11} a_{22} - a_{12} a_{21}}{D^2} = \dfrac{1}{D}$

(a) Apply the change of variables:

$[h_1(u,v)\ h_2(u,v)] = \left(A^{-1} \begin{bmatrix} u \\ v \end{bmatrix}\right)^T$

$\begin{bmatrix} h_1(u,v) \\ h_2(u,v) \end{bmatrix}^T$

$(AB)^T = B^T A^T$

$(A^{-1})^T = (A^T)^{-1}$

$= [u\ v](A^T)^{-1}$

$$f_{UV}(u,v) = |J_h(u,v)|\ f_{XY}(h_1(u,v), h_2(u,v))$$

$$= \frac{1}{|D|} \frac{1}{2\pi \sigma_x \sigma_y} \exp\left\{ -\frac{1}{2} \underbrace{[h_1(u,v)\ h_2(u,v)]\ \Sigma^{-1} \begin{bmatrix} h_1(u,v) \\ h_2(u,v) \end{bmatrix}}_{Q(u,v)} \right\}$$

$$Q(u,v) = [u,v]\,(A^T)^{-1}\, \Sigma^{-1} A^{-1} \begin{bmatrix} u \\ v \end{bmatrix} = [u\ v]\,(A \Sigma A^T)^{-1} \begin{bmatrix} u \\ v \end{bmatrix} = [u\ v]\, \Sigma_{new}^{-1} \begin{bmatrix} u \\ v \end{bmatrix}$$

$(AB)^{-1} = B^{-1} A^{-1}$

$$\Sigma_{new} = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}$$

$$f_{UV}(u,v) = \frac{1}{2\pi\sigma_u\sigma_v\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2} [u \ v] \Sigma_{new}^{-1} \begin{bmatrix} u \\ v \end{bmatrix} \right\}$$

(i) $\sigma_u^2 = Var(U) = Var(a_{11}X + a_{12}Y) = Var(a_{11}X) + Var(a_{12}Y)$

$= a_{11}^2 Var(X) + a_{12}^2 Var(Y) = a_{11}^2\sigma_x^2 + a_{12}^2\sigma_y^2$

$E[U] = E[a_{11}X + a_{12}Y] = 0$

$E[V] = 0$

$\sigma_v^2 = Var(V) = Var(a_{21}X + a_{22}Y) = a_{21}^2\sigma_x^2 + a_{22}^2\sigma_y^2$

$$\rho = \frac{Cov(U,V)}{\sqrt{Var(U)\,Var(V)}} = \frac{a_{11}a_{21}\sigma_x^2 + a_{12}a_{22}\sigma_y^2}{\sqrt{(a_{11}^2\sigma_x^2 + a_{12}^2\sigma_y^2)(a_{21}^2\sigma_x^2 + a_{22}^2\sigma_y^2)}}$$

$Cov(U,V) = E[UV] - E[U]E[V] = E[UV] = E[(a_{11}X + a_{12}Y)(a_{21}X + a_{22}Y)] = a_{11}a_{21}\sigma_x^2 + a_{12}a_{22}\sigma_y^2$

1. Find ρ so that the joint density function $f_{U,V}(u, v)$ of U and V can be expressed as

$$f_{U,V}(u, v) = \frac{1}{2\pi\sigma_u\sigma_v\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}Q(u, v)\right).$$

2. Express $Q(u, v)$ by using σ_u, σ_v, and ρ to conclude that the joint density is the bivariate normal density with parameters $\mu_u = \mu_v = 0$, σ_u, σ_v, and ρ.

3. Let $X' = X + \mu_x$ and $Y' = Y + \mu_y$. Then consider the change of variables via

$$U + \overbrace{a_{11}\mu_x + a_{12}\mu_y}^{\text{mean parameter for } \mu_{U'}} = U' = a_{11}X' + a_{12}Y' = \overbrace{a_{11}X + a_{12}Y}^{U} + a_{11}\mu_x + a_{12}\mu_y$$
$$V + \underbrace{a_{21}\mu_x + a_{22}\mu_y}_{\mu_{V'}} = V' = a_{21}X' + a_{22}Y' = \underbrace{a_{21}X + a_{22}Y}_{V} + a_{21}\mu_x + a_{22}\mu_y$$

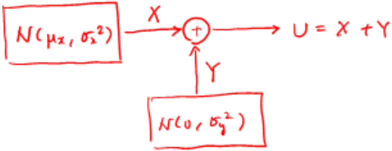
   Express U' and V' in terms of U and V, and describe the joint distribution for U' and V' with specific parameters.

4. Now suppose that X and Y are independent normal random variables with the respective parameters (μ_x, σ_x^2) and (μ_y, σ_y^2). Then what is the joint distribution for U and V given by (5.18)?

Problem 15

Suppose that a real valued data X is transmitted though a noisy channel from location A to location B, and that the value U received at location B is given by $U = X + Y$, where Y is a normal random variable with mean zero and variance σ_y^2. Furthermore, we assume that X is normally distributed with mean μ and variance σ_x^2, and is independent of Y.

1. *Find the expectation $E[U]$ and the variance $\mathrm{Var}(U)$ of the received value U at location B.*
2. *Describe the joint distribution for X and U with specific parameters.*
3. *Provided that $U = t$ is observed at location B, find the best predictor $g(t)$ of X.*
4. *Suppose that we have $\mu = 1$, $\sigma_x = 1$ and $\sigma_y = 0.5$. When $U = 1.3$ was observed, find the best predictor of X given $U = 1.3$.*



Best prediction for X given the observation $U=u$ is

$$E(X \mid U = u)$$

Proof is given by Problem 14

To complete Problem 15, you may use the following theorem:

Linear transformation theorem of bivariate normal random variables. *Let X and Y be independent normal random variables with the respective parameters (μ_x, σ_x^2) and (μ_y, σ_y^2). Then if $a_{11}a_{22} - a_{12}a_{21} \neq 0$, then the random variables*

$$U = a_{11}X + a_{12}Y$$
$$V = a_{21}X + a_{22}Y$$

has the bivariate normal distribution with parameter $\mu_u = a_{11}\mu_x + a_{12}\mu_y$, $\mu_v = a_{21}\mu_x + a_{22}\mu_y$, $\sigma_u^2 = a_{11}^2\sigma_x^2 + a_{12}^2\sigma_y^2$, $\sigma_v^2 = a_{21}^2\sigma_x^2 + a_{22}^2\sigma_y^2$, and

$$\rho = \frac{a_{11}a_{21}\sigma_x^2 + a_{12}a_{22}\sigma_y^2}{\sqrt{(a_{11}^2\sigma_x^2 + a_{12}^2\sigma_y^2)(a_{21}^2\sigma_x^2 + a_{22}^2\sigma_y^2)}}.$$

Optional problems

Objective of assignment.

A sequence of values, $X_1, \ldots, X_n$, are governed by the discrete dynamical system

$$X_k = a_k X_{k-1} + B_k, \quad k = 1, \ldots, n, \tag{5.19}$$

where a_k's are known constants, and B_k's are independent and normally distributed random variables with mean 0 and variance Q. It is assumed that initially the random variable X_0 is normally distributed with mean $\hat{x}_0$ and variance P_0. Otherwise, the information about the value X_k of interest is obtained by observing

$$Z_k = h_k X_k + W_k, \quad k = 1, \ldots, n, \tag{5.20}$$

where h_k's are known constants, and W_k's are independent and normally distributed random variables with mean 0 and variance R.

Objective of assignment, continued.

The data $Z_k = z_k$, $k = 1, \ldots, n$, are collected sequentially. By completing Problem 16 and 17 you will find an "algorithm" to recursively compute the "best" estimate $\hat{x}_1, \ldots, \hat{x}_n$ for the sequence of unobserved values, $X_1, \ldots, X_n$.

Problem

The density function

$$f(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} Q(y_1, y_2)\right)$$

with the quadratic form

$$\begin{aligned} &Q(y_1, y_2) \\ &= \frac{1}{1-\rho^2}\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\frac{(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2}\right] \end{aligned}$$

gives a bivariate normal distribution for a pair (Y_1, Y_2) *of random variables. Then answer (a)–(c) in the following.*

1. The marginal distribution for Y_2 has the normal density function

$$f_2(y_2) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{ -\frac{1}{2} \left(\frac{y_2 - \mu_2}{\sigma_2} \right)^2 \right\}$$

   Find the expectation $E[Y_2]$ and the variance $\mathrm{Var}(Y_2)$ for Y_2.

2. Find $f_1(y_1|y_2)$ satisfying

$$f(y_1, y_2) = f_1(y_1|y_2) f_2(y_2)$$

   The function $f_1(y_1|y_2)$ of y_1 is called a conditional density function given $Y_2 = y_2$.

3. Argue that $f_1(y_1|y_2)$ is a normal density function of y_1 given a fixed value y_2, and find the expectation $E(Y_1|Y_2 = y_2)$ and the variance $\mathrm{Var}(Y_2|Y_2 = y_2)$ given $Y_2 = y_2$.

Remark on Problem 16.

Note that the parameters σ_1^2, σ_2^2, and ρ must satisfy $\sigma_1^2 > 0$, $\sigma_2^2 > 0$, and $-1 < \rho < 1$. It is determined by the 2-by-2 symmetric matrix and the mean column vectors

$$\begin{bmatrix} \mathrm{Var}(Y_1) & \mathrm{Cov}(Y_2, Y_1) \\ \mathrm{Cov}(Y_1, Y_2) & \mathrm{Var}(Y_2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \text{ and } \begin{bmatrix} E[Y_1] \\ E[Y_2] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

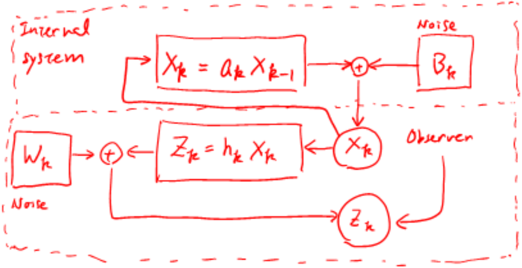
The quadratic form $Q(y_1, y_2)$ has the following decomposition

$$Q(y_1, y_2) = \left(\frac{y_1 - \mu_1 - \rho\frac{\sigma_1}{\sigma_2}(y_2 - \mu_2)}{\sigma_1\sqrt{1-\rho^2}} \right)^2 + \left(\frac{y_2 - \mu_2}{\sigma_2} \right)^2$$

Problem

Suppose that the variable X_{k-1} has a normal distribution with mean $\hat{x}_{k-1}$ and variance P_{k-1} at the $(k-1)$-th iteration of algorithm. In order to complete this problem you may use "linear transformation theorem of bivariate normal random variables" in Problem 15. Answer (a)–(e) in the following

1. Argue that X_k has a normal distribution, and find $\bar{x}_k = E[X_k]$ and $\bar{P}_k = \mathrm{Var}(X_k)$.
2. Find the joint density function for (X_k, Z_k) in terms of $\bar{x}_k$, $\bar{P}_k$, h_k and R.
   HINT: Let $U = X = X_k$, $V = Z_k$, and $Y = W_k$, and apply the linear transformation theorem with $a_{11} = a_{22} = 1$, $a_{12} = 0$, and $a_{21} = h_k$.
3. Find the "best" estimate $\hat{x}_k = E(X_k | Z_k = z_k)$ and the variance $P_k = \mathrm{Var}(X_k | Z_k = z_k)$ in terms of $\bar{x}_k$, $\bar{P}_k$, h_k and R.
4. Argue that X_k has a normal distribution with mean $\hat{x}_k$ and variance P_k.
5. Continue from (c). Find the formula K_k in order to express $\hat{x}_k = \bar{x}_k + K_k(z_k - h_k \bar{x}_k)$.



$$\begin{cases} X_k = a_k X_{k-1} + B_k \\ Z_k = h_k X_k + W_k \end{cases}$$

where $B_k \sim N(0, Q)$ and

$W_k \sim N(0, R)$.

$Z_0, Z_1, \cdots, Z_k \longrightarrow$ Best estimate $\hat{x}_k$ of X_k recursively.

$\longrightarrow$ Given the best estimate $\hat{x}_{k-1}$, construct the best estimate $\hat{x}_k$.

Concluding remark.

The formula K_k in Problem 17(e) is called the Kalman gain. Then the "best" estimate $\hat{x}_k$ and the variance P_k are recursively used and the computation procedure of Problem 17(a) and (c) is applied repeatedly. The resulting sequence $\hat{x}_1, \ldots, \hat{x}_n$ is the "best" estimate for $X_1, \ldots, X_n$. This iterative procedure became known as the discrete Kalman filter algorithm.

Answers to exercises

Problem 1 and 2.

Discussion in class should be sufficient in order to answer all the questions.

Problem 3 and 4.

The following decompositions of Σ^{-1} are sufficient in order to answer all the questions.

$$\begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}^{-1}$$

$$= \frac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)} \begin{bmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sigma_x\sqrt{1-\rho^2}} & 0 \\ -\frac{\rho}{\sigma_y\sqrt{1-\rho^2}} & \frac{1}{\sigma_y} \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_x\sqrt{1-\rho^2}} & -\frac{\rho}{\sigma_y\sqrt{1-\rho^2}} \\ 0 & \frac{1}{\sigma_y} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sigma_x} & -\frac{\rho}{\sigma_x\sqrt{1-\rho^2}} \\ 0 & \frac{1}{\sigma_y\sqrt{1-\rho^2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_x} & 0 \\ -\frac{\rho}{\sigma_x\sqrt{1-\rho^2}} & \frac{1}{\sigma_y\sqrt{1-\rho^2}} \end{bmatrix}$$

Problem 5.

1. $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = (1/4) - (1/3)(2/3) = 1/36$, and $\rho = (1/36)/\sqrt{(1/18)^2} = 1/2$.
2. Since $f(x|y) = \frac{1}{y}$ for $0 \leq x \leq y$ and $f(y|x) = \frac{1}{1-x}$ for $x \leq y \leq 1$, we obtain $E(X|Y = y) = \frac{y}{2}$ and $E(Y|X = x) = \frac{x+1}{2}$.
3. The best linear predictor $g(X)$ of Y is given by $g(X) = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(X - \mu_x) = \frac{1}{2}X + \frac{1}{2}$.
4. Then best predictor of Y is $E(Y|X) = \frac{X+1}{2}$. [This is the same as (c). Why?]

Problem 6.

1. $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = 17/42 - (9/14)^2 = -5/588$ and $\rho = \frac{(-5/588)}{(199/2940)} = -25/199$.
2. $E(Y|X = x) = \frac{6x^2+8x+3}{4(3x^2+3x+1)}$.
3. The best linear predictor $g(X)$ of Y is given by $g(X) = -\frac{25}{119}X + \frac{144}{199}$.

Problem 7.

1. $P(Y \geq 30) = P\left(\frac{Y-32}{8} \geq -\frac{1}{4}\right) = 1 - \Phi(-0.25) \approx 0.60.$
2. $E(Y \mid X = 31) = (32) + (0.8)\left(\frac{8}{6}\right)(31 - 28) = 35.2$
3. Since the conditional density $f_{Y|X}(y \mid x)$ has the normal density with mean $\mu = 35.2$ and variance $\sigma^2 = (8)^2(1 - (0.8)^2) = 23.04$, we can obtain

$$P(Y \geq 30 \mid X = 31) \approx P\left(\frac{Y - 35.2}{4.8} \geq -1.08 \,\middle|\, X = 31\right)$$
$$= 1 - \Phi(-1.08) \approx 0.86$$

Problem 8.

1. $h_1(u,v) = \dfrac{u-b_1}{a_1}$ and $h_2(u,v) = \dfrac{v-b_2}{a_2}$. Thus,
   $$J_h(u,v) = \det \begin{bmatrix} \frac{1}{a_1} & 0 \\ 0 & \frac{1}{a_2} \end{bmatrix} = \frac{1}{a_1 a_2}.$$
2. $f_{Y_1,Y_2}(u,v) = f_{X_1,X_2}\left(\dfrac{u-b_1}{a_1}, \dfrac{v-b_2}{a_2}\right) \cdot \left|\dfrac{1}{a_1 a_2}\right| =$
   $\dfrac{1}{2\pi(|a_1|\sigma_x)(|a_2|\sigma_y)\sqrt{1-\rho^2}} \exp\left(-\dfrac{1}{2}Q(u,v)\right)$, where $Q(u,v) =$
   $\frac{1}{1-\rho^2}\left[\left(\frac{u-(a_1\mu_x+b_1)}{a_1\sigma_x}\right)^2 + \left(\frac{v-(a_2\mu_y+b_2)}{a_2\sigma_y}\right)^2 - 2\rho\frac{(u-(a_1\mu_x+b_1))(v-(a_2\mu_y+b_2))}{(a_1\sigma_x)(a_2\sigma_y)}\right].$
3. From (b), $f_{Y_1,Y_2}(u,v)$ is a bivariate normal distribution with parameter $\mu_u = a_1\mu_x + b_1$, $\mu_v = a_2\mu_y + b_2$, $\sigma_u = |a_1|\sigma_x$, $\sigma_v = |a_2|\sigma_y$, and ρ if $a_1 a_2 > 0$ (or, $-\rho$ if $a_1 a_2 < 0$).

Problem 9.

(i) $h_1(u,v) = \dfrac{a_{22}(u-b_1) - a_{12}(v-b_2)}{a_{11}a_{22} - a_{12}a_{21}}$ and
$h_2(u,v) = \dfrac{-a_{21}(u-b_1) + a_{11}(v-b_2)}{a_{11}a_{22} - a_{12}a_{21}}$. Thus,
$J_h(u,v) = \dfrac{1}{a_{11}a_{22} - a_{12}a_{21}}$.

(ii) Note that $f_{X_1,X_2}(x,y) = \dfrac{1}{2\pi}\exp\left(-\dfrac{1}{2}(x^2+y^2)\right)$. Then we obtain

$$f_{Y_1,Y_2}(u,v) = f_{X_1,X_2}(h_1(u,v), h_2(u,v)) \cdot \left|\frac{1}{a_{11}a_{22} - a_{12}a_{21}}\right|$$
$$= \frac{1}{2\pi|a_{11}a_{22} - a_{12}a_{21}|}\exp\left(-\frac{1}{2}Q(u,v)\right),$$

where $Q(u,v) =$
$$\begin{bmatrix}(u-b_1) & (v-b_2)\end{bmatrix}\begin{bmatrix}a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} \\ a_{11}a_{21} + a_{12}a_{22} & a_{21}^2 + a_{22}^2\end{bmatrix}^{-1}\begin{bmatrix}u-b_1 \\ v-b_2\end{bmatrix}.$$

(iii) By comparing $f_{Y_1,Y_2}(u,v)$ with a bivariate normal density, we can find that $f_{Y_1,Y_2}(u,v)$ is a bivariate normal distribution with parameter $\mu_u = b_1$, $\mu_v = b_2$, $\sigma_u^2 = a_{11}^2 + a_{12}^2$, $\sigma_v^2 = a_{21}^2 + a_{22}^2$, and $\rho = \dfrac{a_{11}a_{21} + a_{12}a_{22}}{\sqrt{(a_{11}^2 + a_{12}^2)(a_{21}^2 + a_{22}^2)}}$.

Problem 10.

1. $h_1(u,v) = uv$ and $h_2(u,v) = u - uv$. Thus,
$$J_h(u,v) = \det \begin{bmatrix} v & u \\ 1-v & -u \end{bmatrix} = -u.$$
2. $f_{XY}(x,y) = f_X(x) f_Y(y)$.
3. We can obtain

$$\begin{aligned} f_{UV}(u,v) &= |-u| \times f_X(uv) \times f_Y(u-uv) \\ &= u \times \frac{1}{\Gamma(\alpha_1)} \lambda^{\alpha_1} e^{-\lambda(uv)} (uv)^{\alpha_1 - 1} \times \frac{1}{\Gamma(\alpha_2)} \lambda^{\alpha_2} e^{-\lambda(u-uv)} (u-uv)^{\alpha_2} \\ &= f_U(u) f_V(v) \end{aligned}$$

where $f_U(u) = \dfrac{1}{\Gamma(\alpha_1+\alpha_2)} \lambda^{\alpha_1+\alpha_2} e^{-\lambda u} u^{\alpha_1+\alpha_2-1}$ and $f_V(v) = \dfrac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1-1}(1-v)^{\alpha_2-1}$. Thus, $f_U(u)$ has a gamma distribution with parameter $(\alpha_1+\alpha_2, \lambda)$, and $f_V(v)$ has a beta distribution with parameter (α_1, α_2).

Problem 11.

1. Observe that $f(x, y) = (xe^{-x}) \cdot (e^{-y})$ and that xe^{-x}, $x \geq 0$, is the gamma density with parameter $(2, 1)$ and e^{-y}, $y \geq 0$, is the exponential density with $\lambda = 1$. Thus, we can obtain

$$f_X(x) = \int_0^\infty f(x, y)\, dy = xe^{-x} \int_0^\infty e^{-y}\, dy = xe^{-x}, \quad x \geq 0;$$
$$f_Y(y) = \int_0^\infty f(x, y)\, dx = e^{-y} \int_0^\infty xe^{-x}\, dx = e^{-y}, \quad y \geq 0.$$

   Since $f(x, y) = f_X(x) f_Y(y)$, X and Y are independent.
2. Notice that the exponential density $f_Y(y)$ with $\lambda = 1$ can be seen as the gamma distribution with $(1, 1)$. Since the sum of independent gamma random variables has again a gamma distribution, U has the gamma distribution with $(3, 1)$.

Problem 12.

Let X_1 be the amount of money you invest in the first year, let X_2 be the amount in the second year, and let X_2 be the amount in the third year. Then the total amount of money you have at the end of term, say Y, becomes

$$Y = 1.331X_1 + 1.21X_2 + 1.1X_3.$$

1. $E(1.331X_1 + 1.21X_2 + 1.1X_3) = 1.331E(X_1) + 1.21E(X_2) + 1.1E(X_3) = 3.641 \times 500 = 1820.5.$
2. $\mathrm{Var}(1.331X_1 + 1.21X_2 + 1.1X_3) = (1.331)^2\mathrm{Var}(X_1) + (1.21)^2\mathrm{Var}(X_2) + (1.1)^2\mathrm{Var}(X_3) = 4.446 \times 200^2 \approx 177800.$ Thus, the standard deviation is $\sqrt{\mathrm{Var}(Y)} \approx \sqrt{177800} \approx 421.7.$
3. $P(Y \geq 2000) = P\left(\frac{Y - 1820.5}{421.7} \geq \frac{2000 - 1820.5}{421.7}\right) \approx 1 - \Phi(0.43) \approx 0.334.$

Problem 13.

1. By using the change-of-variable formula with $g(x) = x^n$, we can obtain $f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = f_X(y^{\frac{1}{n}}) \cdot \frac{1}{n} y^{-\left(\frac{n-1}{n}\right)} = \frac{\lambda}{n} y^{-\left(\frac{n-1}{n}\right)} \exp\left(-\lambda y^{\frac{1}{n}}\right), \quad y \geq 0$
2. Since the pdf $f(x)$ is of the form of gamma density with $(n+1, \lambda)$, we must have $f(x) = \dfrac{\lambda^{n+1}}{\Gamma(n+1)} x^n e^{-\lambda x}$, $x \geq 0$. Therefore, $c = \dfrac{\lambda^{n+1}}{\Gamma(n+1)}$ $(= \frac{\lambda^{n+1}}{n!}$, since n is an integer).
3. Since $f(x)$ in (b) is a pdf, we have $\int_0^\infty f(x)\,dx = \int_0^\infty \frac{\lambda^{n+1}}{\Gamma(n+1)} x^n e^{-\lambda x}\,dx = 1$. Thus, we can obtain $E[Y] = E[X^n] = \int_0^\infty x^n \left(\lambda e^{-\lambda x}\right) dx = \frac{\Gamma(n+1)}{\lambda^n}$ $(= \frac{n!}{\lambda^n}$, since n is an integer).

Problem 14.

(i) $\sigma_u^2 = \mathrm{Var}(U) = a_{11}^2\sigma_x^2 + a_{12}^2\sigma_y^2$ and $\sigma_v^2 = \mathrm{Var}(V) = a_{21}^2\sigma_x^2 + a_{22}^2\sigma_y^2$.

(ii) $h_1(u, v) = \frac{1}{D}(a_{22}u - a_{12}v)$ and $h_2(u, v) = \frac{1}{D}(-a_{21}u + a_{11}v)$.

(iii) We have $J_h(u, v) = \frac{1}{D}$.

1. $\rho = \frac{a_{11}a_{21}\sigma_x^2 + a_{12}a_{22}\sigma_y^2}{\sqrt{(a_{11}^2\sigma_x^2 + a_{12}^2\sigma_y^2)(a_{21}^2\sigma_x^2 + a_{22}^2\sigma_y^2)}}$.
2. $Q(u, v)$ can be expressed as

$$Q(u, v) = \frac{1}{1 - \rho^2}\left[\left(\frac{u}{\sigma_u}\right)^2 + \left(\frac{v}{\sigma_v}\right)^2 - 2\rho\frac{uv}{\sigma_u\sigma_v}\right]$$

   Thus, the joint density function $f_{U,V}(u, v)$ is the bivariate normal density with parameters $\mu_u = \mu_v = 0$ together with σ_u, σ_v, ρ as calculated in (a) and (c).
3. $U' = U + a_{11}\mu_x + a_{12}\mu_y$ and $V' = V + a_{21}\mu_x + a_{22}\mu_y$. Then U' and V' has the bivariate normal density with parameters $\mu_u = a_{11}\mu_x + a_{12}\mu_y$, $\mu_v = a_{21}\mu_x + a_{22}\mu_y$, σ_u, σ_v, and ρ.
4. (U, V) has a bivariate normal distribution with parameter $\mu_u = a_{11}\mu_x + a_{12}\mu_y$, $\mu_v = a_{21}\mu_x + a_{22}\mu_y$, $\sigma_u^2 = a_{11}^2\sigma_x^2 + a_{12}^2\sigma_y^2$, $\sigma_v^2 = a_{21}^2\sigma_x^2 + a_{22}^2\sigma_y^2$, and $\rho = \frac{a_{11}a_{21}\sigma_x^2 + a_{12}a_{22}\sigma_y^2}{\sqrt{(a_{11}^2\sigma_x^2 + a_{12}^2\sigma_y^2)(a_{21}^2\sigma_x^2 + a_{22}^2\sigma_y^2)}}$.

Problem 15.

1. $E[U] = E[X] + E[Y] = \mu$ and
   $\mathrm{Var}(U) = \mathrm{Var}(X) + \mathrm{Var}(Y) = \sigma_x^2 + \sigma_y^2$.
2. By using **linear transformation theorem**, we can find that X and U have the bivariate normal distribution with means $\mu_x = \mu$ and $\mu_u = \mu$, variances $\sigma_x^2 = \sigma_x^2$ and $\sigma_u^2 = \sigma_x^2 + \sigma_y^2$, and $\rho = \dfrac{\sigma_x}{\sqrt{\sigma_x^2 + \sigma_y^2}}$.
3. $E(X \mid U = t) = \mu_x + \rho \dfrac{\sigma_x}{\sigma_u}(t - \mu_u) = \mu + \dfrac{\sigma_x^2}{\sigma_x^2 + \sigma_y^2}(t - \mu)$.
4. $E(X \mid U = 1.3) = 1 + \dfrac{(1)^2}{(1)^2 + (0.5)^2}(1.3 - 1) = 1.24$.