Chapter 7

Functions of Random Variables

Section 7.2

Transformation of Variables

Theorem 7.1

Suppose that X is a **discrete** random variable with probability distribution $f(x)$. Let $Y = u(X)$ define a one-to-one transformation between the values of X and Y so that the equation $y = u(x)$ can be uniquely solved for x in terms of y, say $x = w(y)$. Then the probability distribution of Y is

$$g(y) = f[w(y)].$$

example

Example 7.1: Let X be a geometric random variable with probability distribution

$$f(x) = \frac{3}{4}\left(\frac{1}{4}\right)^{x-1}, \qquad x = 1, 2, 3, \ldots .$$

Find the distribution of $Y = X^2$

Theorem 7.2

Suppose that X_1 and X_2 are **discrete** random variables with joint probability distribution $f(x_1, x_2)$. Let $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$ define a one-to-one transformation between the points (x_1, x_2) and (y_1, y_2) so that the equations

$$y_1 = u_1(x_1, x_2) \qquad \text{and} \qquad y_2 = u_2(x_1, x_2)$$

may be uniquely solved for x_1 and x_2 in terms of y_1 and y_2, say $x_1 = w_1(y_1, y_2)$ and $x_2 = w_2(y_1, y_2)$. Then the joint probability distribution of Y_1 and Y_2 is

$$g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)].$$

example

Example 7.2: Let X_1 and X_2 be two independent random variables having Poisson distributions with parameters μ_1 and μ_2, respectively. Find the distribution of the random variable $Y_1 = X_1 + X_2$.

Theorem 7.3

Suppose that X is a **continuous** random variable with probability distribution $f(x)$. Let $Y = u(X)$ define a one-to-one correspondence between the values of X and Y so that the equation $y = u(x)$ can be uniquely solved for x in terms of y, say $x = w(y)$. Then the probability distribution of Y is

$$g(y) = f[w(y)]|J|,$$

where $J = w'(y)$ and is called the **Jacobian** of the transformation.

example

Example 7.3: Let X be a continuous random variable with probability distribution

$$f(x) = \begin{cases} \frac{x}{12}, & 1 < x < 5, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability distribution of the random variable $Y = 2X - 3$.

Derivative of function

- Consider a function $f(x)$ with $f : \mathbb{R}^n \to \mathbb{R}$
- When $n = 1$, we know the derivative

$$\frac{df(x)}{dx}$$

  or $f'(x)$, is the rate of change in function f
- This can be viewed also as “slope” of line tangent to f
- Or, the first-order approximation of f around $x = x_0$ is given by

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

Partial derivative

- Given a multivariate function f(x,y)

$$\frac{\partial f}{\partial x} = \lim_{\epsilon \to 0} \frac{f(x+\epsilon, y) - f(x,y)}{\epsilon}$$
$$\frac{\partial f}{\partial y} = \lim_{\epsilon \to 0} \frac{f(x, y+\epsilon) - f(x,y)}{\epsilon}$$

are the partial derivative of f with respect to x and y

- Rate of change of *f* in only one direction

Partial derivative

- Finding partial derivative: treat other variables **as constant**
- Example: $f(x,y)=x^2 + xy$

$$\frac{\partial f}{\partial x} = 2x + y, \ \frac{\partial f}{\partial y} = x$$

For $x \in \mathbb{R}^n, \ f(x) = f(x_1, x_2, ..., x_n)$,

$\frac{\partial f}{\partial x_i}$ can be defined similarly

Gradient

- now consider case $n \geq 1$

- consider first-order approximation of $f(x)$ with small change $\Delta x \in \mathbb{R}^n$
- By chain rule,

$$\begin{aligned} f(x + \Delta x) &\approx f(x) + \frac{\partial f}{\partial x_1} \Delta x_1 + \cdots + \frac{\partial f}{\partial x_n} \Delta x_n \\ &= f(x) + \nabla f(x)^T \Delta x \end{aligned}$$

Here

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)$$

is called gradient of f at x

Gradient Example

- Let $f(x) = c^T x$. Find ∇f
- Since $f(x) = \sum_{i=1}^{n} c_i x_i$, we have $\frac{\partial f}{\partial x_i} = c_i$ thus

$$\nabla f = (c_1, c_2, \ldots, c_n) = c$$

Jacobian

- Consider vector function $f(x) = (f_1(x), f_2(x), \ldots, f_m(x))$. Consider small change Δx.
- $$f(x+\Delta x) = \begin{bmatrix} f_1(x+\Delta x) \\ \ldots \\ f_m(x+\Delta x) \end{bmatrix} \approx f(x) + \begin{bmatrix} \nabla f_1(x)^T \Delta x \\ \ldots \\ \nabla f_m(x)^T \Delta x \end{bmatrix}$$
  $$= f(x) + \begin{bmatrix} \nabla f_1(x)^T \\ \ldots \\ \nabla f_m(x)^T \end{bmatrix} \Delta x$$
  The Jacobian of f at x is
  $$J_f = \begin{bmatrix} \nabla f_1(x)^T \\ \ldots \\ \nabla f_m(x)^T \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \ldots & \frac{\partial f_1}{\partial x_n} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \ldots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$
- So the first order change is $J_f \Delta x$

Jacobian

- Examples: R^m to R^n
- $f(x) = (f_1(x), f_2(x)) = (x_1^2, x_1 + 2x_1x_2)$

$$J_f = \begin{bmatrix} \nabla f_1(x)^T \\ \nabla f_2(x)^T \end{bmatrix} = \begin{bmatrix} 2x_1 & 0 \\ 1 + 2x_2 & 2x_1 \end{bmatrix}$$

Jacobian

- Examples: R^m to R^n
- y = Ax

$$\frac{\partial y}{\partial x} = A$$

- Often, Jacobian is considered as a (partial) derivative of multidimensional mapping

Summary: derivatives

- Meaning of derivatives

$$f : \mathbb{R} \to \mathbb{R} \qquad f(x+\delta) \approx f(x) + f'(x)\delta$$

$$f : \mathbb{R}^n \to \mathbb{R} \qquad f(x+\delta) \approx f(x) + \nabla f(x)^T \delta$$

$$f : \mathbb{R}^n \to \mathbb{R}^m \qquad f(x+\delta) \approx f(x) + J_f \delta$$

Theorem 7.4

Suppose that X_1 and X_2 are **continuous** random variables with joint probability distribution $f(x_1, x_2)$. Let $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$ define a one-to-one transformation between the points (x_1, x_2) and (y_1, y_2) so that the equations $y_1 = u_1(x_1, x_2)$ and $y_2 = u_2(x_1, x_2)$ may be uniquely solved for x_1 and x_2 in terms of y_1 and y_2, say $x_1 = w_1(y_l, y_2)$ and $x_2 = w_2(y_1, y_2)$. Then the joint probability distribution of Y_1 and Y_2 is

$$g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)]|J|,$$

where the Jacobian is the 2×2 determinant

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

and $\frac{\partial x_1}{\partial y_1}$ is simply the derivative of $x_1 = w_1(y_1, y_2)$ with respect to y_1 with y_2 held constant, referred to in calculus as the partial derivative of x_1 with respect to y_1. The other partial derivatives are defined in a similar manner.

Proof sketch

- Check equivalent probability
- dv_x, dv_y are infinitesimal volumes

$$g(y)dv_y = f(x)dv_x$$

- Let y = u(x) imply x = w(y)
- w(y) is a vector function

Proof sketch

- Using definition of Jacobian,

$$dx = J_w dy$$

- By using the property of linear transformation,

$$dv_x = |J_w| dv_y$$

- we get

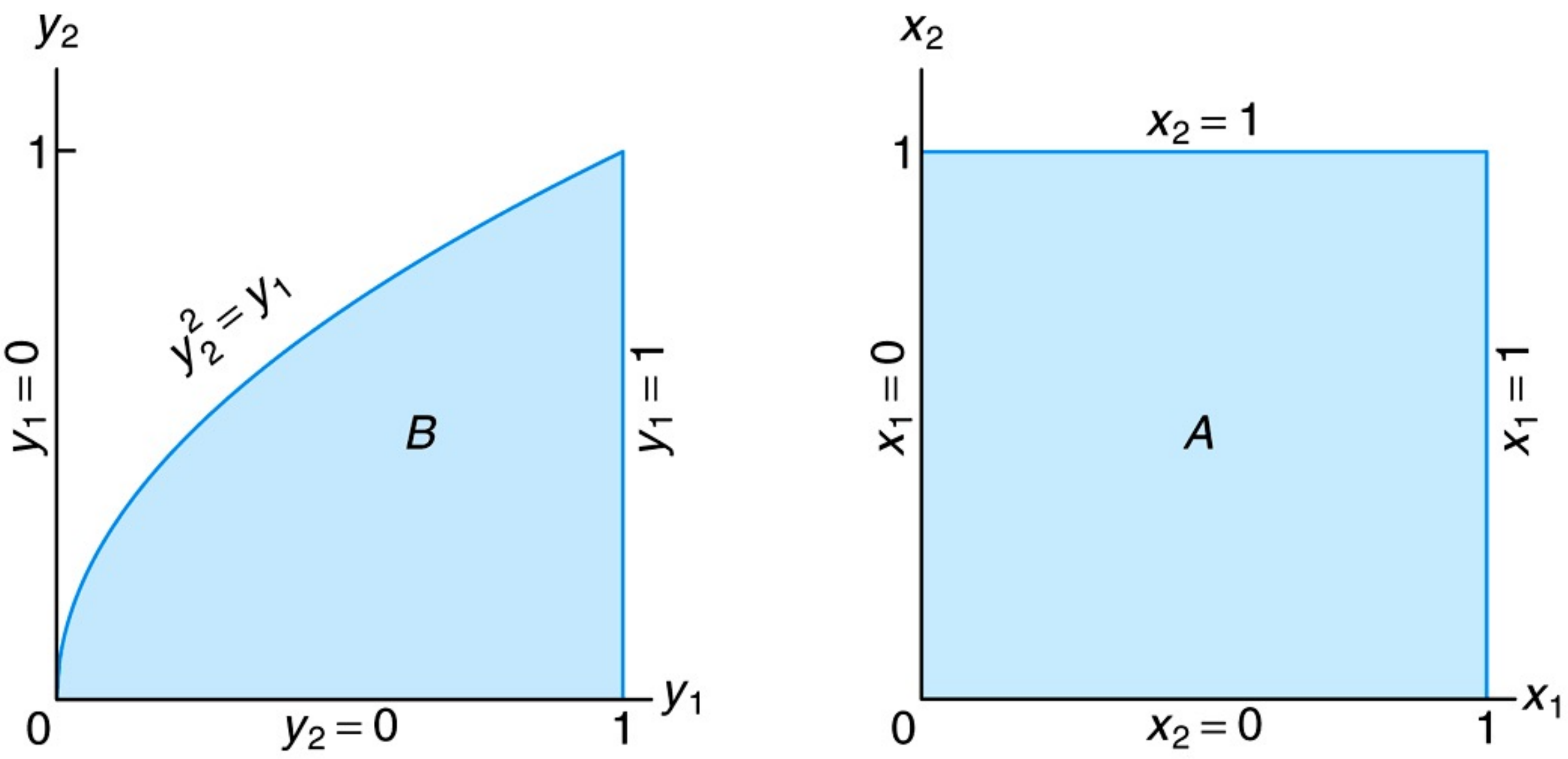
$$g(y) = f(w(y))|J_w|$$

Example 7.4: Let X_1 and X_2 be two continuous random variables with joint probability distribution

$$f(x_1, x_2) = \begin{cases} 4x_1x_2, & 0 < x_1 < 1,\ 0 < x_2 < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the joint probability distribution of $Y_1 = X_1^2$ and $Y_2 = X_1X_2$.

Figure 7.1 Mapping set *A* into set *B*



Theorem 7.5

Suppose that X is a **continuous** random variable with probability distribution $f(x)$. Let $Y = u(X)$ define a transformation between the values of X and Y that is not one-to-one. If the interval over which X is defined can be partitioned into k mutually disjoint sets such that each of the inverse functions

$$x_1 = w_1(y), \quad x_2 = w_2(y), \quad \ldots, \quad x_k = w_k(y)$$

of $y = u(x)$ defines a one-to-one correspondence, then the probability distribution of Y is

$$g(y) = \sum_{i=1}^{k} f[w_i(y)]|J_i|,$$

where $J_i = w_i^{'}(y)$, $i = 1, 2, \ldots, k$.

example

Example 7.5: Show that $Y = (X-\mu)^2/\sigma^2$ has a chi-squared distribution with 1 degree of freedom when X has a normal distribution with mean μ and variance σ^2.

example

- Show that lognormal distribution for X where Y=log(X) and Y~N(μ,σ^2) has pdf

$$f(x;\mu,\sigma) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2\sigma^2}[\ln(x)-\mu]^2}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Section 7.3

Moments and Moment-Generating Functions

Definition 7.1

The rth **moment about the origin** of the random variable X is given by

$$\mu_r' = E(X^r) = \begin{cases} \sum_x x^r f(x), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x^r f(x)\,dx, & \text{if } X \text{ is continuous.} \end{cases}$$

Definition 7.2

The **moment-generating function** of the random variable X is given by $E(e^{tX})$ and is denoted by $M_X(t)$. Hence,

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum\limits_x e^{tx} f(x), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\, dx, & \text{if } X \text{ is continuous.} \end{cases}$$

Theorem 7.6

Let X be a random variable with moment-generating function $M_X(t)$. Then

$$\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = \mu_r'.$$

example

Example 7.6: Find the moment-generating function of the binomial random variable X and then use it to verify that $\mu = np$ and $\sigma^2 = npq$.

example

Example 7.7: Show that the moment-generating function of the random variable X having a normal probability distribution with mean μ and variance σ^2 is given by

$$M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

example

- Find the MGF of X~χ^2(1)

$$M_X(t) = \mathbb{E}[e^{tZ^2}] = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} e^{(tz^2 - z^2/2)} dz$$

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

- MGF of X~χ^2(v) with d.o.f of v,

$$M_X(t) = (1-2t)^{-v/2}$$

Theorem 7.7

(Uniqueness Theorem) Let X and Y be two random variables with moment-generating functions $M_X(t)$ and $M_Y(t)$, respectively. If $M_X(t) = M_Y(t)$ for all values of t, then X and Y have the same probability distribution.

Theorem 7.8

$$\boxed{M_{X+a}(t) = e^{at} M_X(t).}$$

Theorem 7.9

$$\boxed{M_{aX}(t) = M_X(at).}$$

Theorem 7.10

If $X_1, X_2, \ldots, X_n$ are independent random variables with moment-generating functions $M_{X_1}(t), M_{X_2}(t), \ldots, M_{X_n}(t)$, respectively, and $Y = X_1 + X_2 + \cdots + X_n$, then

$$M_Y(t) = M_{X_1}(t) M_{X_2}(t) \cdots M_{X_n}(t).$$

Theorem 7.11

If $X_1, X_2, \ldots, X_n$ are independent random variables having normal distributions with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, respectively, then the random variable

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

has a normal distribution with mean

$$\mu_Y = a_1 \mu_1 + a_2 \mu_2 + \cdots + a_n \mu_n$$

and variance

$$\sigma_Y^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2.$$

Theorem 7.12

If $X_1, X_2, \ldots, X_n$ are mutually independent random variables that have, respectively, chi-squared distributions with $v_1, v_2, \ldots, v_n$ degrees of freedom, then the random variable

$$Y = X_1 + X_2 + \cdots + X_n$$

has a chi-squared distribution with $v = v_1 + v_2 + \cdots + v_n$ degrees of freedom.

Corollary 7.1

If $X_1, X_2, \ldots, X_n$ are independent random variables having identical normal distributions with mean μ and variance σ^2, then the random variable

$$Y = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$$

has a chi-squared distribution with $v = n$ degrees of freedom.

Corollary 7.2

If $X_1, X_2, \ldots, X_n$ are independent random variables and X_i follows a normal distribution with mean μ_i and variance σ_i^2 for $i = 1, 2, \ldots, n$, then the random variable

$$Y = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2$$

has a chi-squared distribution with $v = n$ degrees of freedom.

Random Vectors and Jointly Gaussian RVs

Random Vector

- Random vector $X = (X_1, X_2, \ldots, X_n)$ is a n-dimensional vector of RVs $X_1, X_2, \ldots, X_n$
- Random vector X is treated as a column vector

- Probabilities associated with random vectors can be found by joint distribution of $X_1, X_2, \ldots, X_n$

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

Random Vector: mean

- Mean of random vector $X = (X_1, X_2, \ldots, X_n)$ is simply a n-dimensional vector of means of RV $X_1, X_2, \ldots, X_n$

$$E[X] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

Random Vector: mean

- Let c be a vector in R^n. Consider random variable

$$Y = c^T X.$$

Its mean is

$$E[c^T X] = E[\sum_{i=1}^{n} c_i X_i] = \sum_{i=1}^{n} c_i E[X_i] = c^T \mu$$

where μ is the mean vector

Covariance Matrix

- covariance matrix of random vector $X = (X_1, X_2, \ldots, X_n)$ is n-by-n matrix S where

$$S_{i,j} = \sigma(X_i, X_j)$$

where $\sigma(X,Y)$ is the covariance of X and Y

$$\sigma(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)]$$

Covariance Matrix: properties

$$S = \begin{bmatrix} \sigma^2(X_i) & \sigma(X_1, X_2) & \dots & \sigma(X_1, X_n) \\ \sigma(X_1, X_2) & \sigma^2(X_2) & \dots & \sigma(X_2, X_n) \\ \dots & \dots & \dots & \dots \\ \sigma(X_n, X_1) & \sigma(X_n, X_2) & \dots & \sigma^2(X_n) \end{bmatrix}$$

- Diagonal elements of covariance matrix are variances of RVs

Covariance Matrix: properties

$$S = E[(X-\mu)(X-\mu)^T]$$

- Note $(X-\mu)(X-\mu)^T$ is a symmetric rank-1 matrix!
- Again shows S must be symmetric

Covariance Matrix: properties

$$S = E[(X-\mu)(X-\mu)^T]$$

- S is positive semi-definite (psd)
  - Symmetric matrix A is said to be psd iff $c^T A c$ is non-negative for any vector c

Covariance Matrix: properties

$$S = E[(X-\mu)^T(X-\mu)]$$

- S is positive semi-definite (psd)

$$\begin{aligned} c^T S c &= c^T E[(X-\mu)(X-\mu)^T]c \\ &= c^T[\sum_i p_i(x_i-\mu)(x_i-\mu)^T]c \\ &= \sum_i p_i[c^T(x_i-\mu)]^2 \end{aligned}$$

Random Vector: mean

- Let c be a vector in R^n. Consider random variable

$$Y = c^T X.$$

Its variance is

$$c^T S c$$

Random Vector: mean

- Let c be a vector in R^n. Consider random variable

$$Y = c^T X.$$

Its variance is

$$\begin{aligned} E[(Y - c^T\mu)^2] &= E[\{c^T(X-\mu)\}^2] \\ &= E[c^T(X-\mu)(X-\mu)^T c] \\ &= c^T E[(X-\mu)(X-\mu)^T]c \\ &= c^T S c \end{aligned}$$

Jointly Gaussian RVs

- Consider RVs $X_1, X_2, \ldots, X_n$ where

$$f_{X_1,\ldots,X_n}(x) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

for some $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{S}^n_{++}$ with $x = (x_1, \ldots, x_n)$

- We say $X_1, X_2, \ldots, X_n$ are jointly Gaussian RVs

$$\boxed{(X_1, X_2, \ldots, X_n) \sim N(\mu, \Sigma)}$$

Jointly Gaussian RVs

- $X_1, X_2, \ldots, X_n$ are jointly Gaussian

$$(X_1, X_2, \ldots, X_n) \sim N(\mu, \Sigma)$$

iff

$$X = AZ + \mu$$

1. Z is a random vector of independent standard Gaussian RVs
2. $A = U^T \Lambda^{1/2}$ for spectral decomposition $\Sigma = U^T \Lambda U$, so $\Sigma = AA^T$

Jointly Gaussian RVs

- Proof. Joint PDF of Z is

$$f_{Z_1,\ldots,Z_n}(z) = \frac{1}{\sqrt{(2\pi)^n}} \exp\left[-\frac{1}{2}z^T z\right]$$

from transformation of RVs

$$\begin{aligned} f_{X_1,\ldots,X_n}(x) &= f_{Z_1,\ldots,Z_n}(z)|J| \\ &= f_{Z_1,\ldots,Z_n}(A^{-1}(x-\mu))|J| \end{aligned}$$

Jointly Gaussian RVs

- Jacobian is

$$J_{i,j} = \frac{\partial z_i}{\partial x_j}$$

or $J=A^{-1}$.

- also

$$\begin{aligned} f_{X_1,\ldots,X_n}(x) &= f_{Z_1,\ldots,Z_n}(A^{-1}(x-\mu))|A|^{-1} \\ &= f_{Z_1,\ldots,Z_n}(A^{-1}(x-\mu))|\Sigma|^{-\frac{1}{2}} \end{aligned}$$

$$|\Sigma| = |AA^T| = |A|^2$$

Jointly Gaussian RVs

- So $f_{X_1,\ldots,X_n}(x)$ is given by

$$\frac{1}{\sqrt{(2\pi)^n}} \exp\left[-\frac{1}{2}\{A^{-1}(x-\mu)\}^T A^{-1}(x-\mu)\right] |\Sigma|^{-\frac{1}{2}}$$

$$= \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^T A^{-T} A^{-1}(x-\mu)\right]$$

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

Jointly Gaussian RVs

- $X=(X_1,X_2,\ldots,X_n)$ is a Gaussian random vector with mean μ and covariance matrix Σ

$$E[X] = E[AZ] + \mu = AE[Z] + \mu = \mu$$

$$\begin{aligned} E[(X-\mu)(X-\mu)^T)] &= E[AZZ^TA^T] \\ &= AE[ZZ^T]A^T = AA^T = \Sigma \end{aligned}$$

Jointly Gaussian RVs

- $X=(X_1,X_2,\ldots,X_n)$ is a Gaussian random vector with mean μ and covariance matrix Σ

- Individually, X_i is Gaussian (linear combination of independent Gaussian RVs) with $X_i \sim N(\mu_i, \sigma_i^2)$

Jointly Gaussian RVs

- Two-variable case: suppose (X,Y) is a Gaussian random vector, then the pdf $f_{X,Y}(x,y)$ is given by

$$\frac{1}{\sqrt{(2\pi)^2(\sigma_X^2\sigma_Y^2 - \sigma_{X,Y}^2)}} \exp\left[-\frac{1}{2}\begin{bmatrix} x-\mu_X & y-\mu_Y \end{bmatrix} \begin{bmatrix} \sigma_X^2 & \sigma_{X,Y} \\ \sigma_{X,Y} & \sigma_Y^2 \end{bmatrix}^{-1} \begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix}\right]$$

Jointly Gaussian RVs

- suppose (X,Y) is a Gaussian random vector, where X,Y are zero-mean, unit-variance then the pdf $f_{X,Y}(x,y)$ is given by

$$\frac{1}{2\pi\sqrt{(1-\rho^2)}}\exp\left[-\frac{1}{2(1-\rho^2)}\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}\right]$$

where ρ is the correlation coefficient

Jointly Gaussian RVs

- suppose (X,Y) is a Gaussian random vector, where X,Y are zero-mean, unit-variance then the pdf $f_{X,Y}(x,y)$ is given by

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

Jointly Gaussian RVs

- suppose (X,Y) is a Gaussian random vector and they are not correlated, $\rho=0$
- Also implies that X,Y are independent!
- In jointly Gaussian RVs, 'uncorrelated' and 'independent' are equivalent!
  - in general, false statement!

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

Jointly Gaussian RVs

- If jointly Gaussian RVs $X=(X_1,X_2,\dots,X_n)$ are uncorrelated, $\sigma(X_i,X_j)=0$ then pdf is given by

$$\frac{1}{\sqrt{(2\pi)^n}\prod_{i=1}^{n}\sigma_i}\exp\left[-\sum_{i=1}^{n}\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right]$$

- covariance matrix is diagonal
- $X_1,X_2,\dots,X_n$ are independent Gaussian
  - pdf is product of marginal Gaussians

Jointly Gaussian RVs

- $X=(X_1,X_2,\ldots,X_n)$ is a Gaussian random vector with mean μ and covariance matrix Σ.

- We write

$$X \sim N(\mu, \Sigma)$$

Jointly Gaussian RVs

- $X=(X_1,X_2,\ldots,X_n)$ is a Gaussian random vector with mean μ and covariance matrix Σ.

- Consider RV $Y = c^T X$ then

- $Y \sim N(c^T\mu, c^T\Sigma c)$

Jointly Gaussian RVs

- Why?

$$Y = c^T X = c^T(AZ + \mu) = c^T AZ + c^T \mu$$

- Y is Gaussian with mean $c^T\mu$, and variance

$$\|A^T c\|_2^2 = c^T AA^T c = c^T \Sigma c$$

- Linear combination of jointly Gaussian RVs is also Gaussian!

Conditional distribution for Jointly Gaussian RVs

- suppose (X,Y) is a Gaussian random vector, where X,Y are zero-mean, unit-variance
- pdf $f_{X,Y}(x,y)$ is given by

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

- What is the conditional distribution $f_{X|Y}(x|y)$?

Conditional distribution for Jointly Gaussian RVs

- What is the conditional distribution $f_{X|Y}(x|y)$?

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

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

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

- $X|Y=y \sim N(\rho y, 1-\rho^2)$
- Conditional distribution is also Gaussian!
  - Can be shown in general for jointly Gaussian RVs