3. Probability and Statistics

- definitions, probability measures
- conditional expectations
- correlation and covariance
- some important random variables
- multivariate random variables

Definition

a random variable X is a *function* mapping an outcome to a real number

- the sample space, S, is the *domain* of the random variable
- S_X is the range of the random variable

example: toss a coin three times and note the sequence of heads and tails

$$S = \{\text{HHH,HHT,HTH,THH,HTT,THT,TTH,TTT}\}$$

Let X be the number of heads in the three tosses

$$S_X = \{0, 1, 2, 3\}$$

Probability measures

Cumulative distribution function (CDF)

$$F(a) = P(X \le a)$$

Probability mass function (PMF) for discrete RVs

$$p(k) = P(X = k)$$

Probability density function (PDF) for continuous RVs

$$f(x) = \frac{dF(x)}{dx}$$

Probability Density Function

Probability Density Function (PDF)

- $f(x) \geq 0$
- $P(a \leq X \leq b) = \int_a^b f(x)dx$
- $F(x) = \int\limits_{-\infty}^{x} f(u)du$

Probability Mass Function (PMF)

- $p(k) \geq 0$ for all k
- $\sum\limits_{k \in S} p(k) = 1$

Expected values

let $g(X)$ be a function of random variable X

$$\mathbf{E}[g(X)] = \begin{cases} \sum_{x \in S} g(x)p(x) & X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x)f(x)dx & X \text{ is continuous} \end{cases}$$

Mean

$$\mu = \mathbf{E}[X] = \begin{cases} \sum_{x \in S} xp(x) & X \text{ is discrete} \\ \int_{-\infty}^{\infty} xf(x)dx & X \text{ is continuous} \end{cases}$$

Variance

$$\sigma^2 = \mathbf{var}[X] = \mathbf{E}[(X - \mu)^2]$$

n^{th} **Moment**

$$\mathbf{E}[X^n]$$

Facts

Let $Y = g(X) = aX + b$, $\quad a, b$ are constants

- $\mathbf{E}[Y] = a\mathbf{E}[X] + b$
- $\mathbf{var}[Y] = a^2\,\mathbf{var}[X]$
- $\mathbf{var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$

Example of Random Variables

Discrete RVs

- Bernoulli
- Binomial
- Geometric
- Negative binomial
- Poisson
- Uniform

Continuous RVs

- Uniform
- Exponential
- Gaussian (Normal)
- Gamma, Chi-squared, Student's t, F
- Logistics

Joint cumulative distribution function

$$F_{XY}(a,b) = P(X \leq a, Y \leq b)$$

- a joint CDF is a nondecreasing function of x and y:

$$F_{XY}(x_1,y_1) \leq F_{XY}(x_2,y_2), \quad \text{if } x_1 \leq x_2 \text{ and } y_1 \leq y_2$$

- $F_{XY}(x_1,-\infty) = 0$, $F_{XY}(-\infty,y_1) = 0$, $F_{XY}(\infty,\infty) = 1$
- $P(x_1 < X \leq x_2, y_1 < Y \leq y_2)$

$$= F_{XY}(x_2,y_2) - F_{XY}(x_2,y_1) - F_{XY}(x_1,y_2) + F_{XY}(x_1,y_1)$$

Joint PMF for discrete RVs

$$p_{XY}(x,y) = P(X = x, Y = y), \quad (x,y) \in S$$

Joint PDF for continuous RVs

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

Marginal PMF

$$p_X(x) = \sum_{y \in S} p_{XY}(x,y), \quad p_Y(y) = \sum_{x \in S} p_{XY}(x,y)$$

Marginal PDF

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,z)dz, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(z,y)dz$$

Conditional Probability

Discrete RVs

the *conditional PMF of* Y *given* $X = x$ is defined by

$$\begin{aligned} p_{Y|X}(y|x) &= P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} \\ &= \frac{p_{XY}(x, y)}{p_X(x)} \end{aligned}$$

Continuous RVs

the *conditional PDF of* Y *given* $X = x$ is defined by

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

Conditional Expectation

the conditional expectation of Y given $X = x$ is defined by

Continuous RVs

$$\mathbf{E}[Y|X] = \int_{-\infty}^{\infty} y \; f_{Y|X}(y|x) dy$$

Discrete RVs

$$\mathbf{E}[Y|X] = \sum_{y} y \; p_{Y|X}(y|x)$$

- $\mathbf{E}[Y|X]$ is the center of mass associated with the conditional pdf or pmf
- $\mathbf{E}[Y|X]$ can be viewed as a function of random variable X
- $\mathbf{E}[\mathbf{E}[Y|X]] = \mathbf{E}[Y]$

in fact, we can show that

$$\mathbf{E}[h(Y)] = \mathbf{E}[\mathbf{E}[h(Y)|X]]$$

for any function $h(\cdot)$ that $\mathbf{E}[|h(Y)|] < \infty$

proof.

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

Independence of two random variables

X and Y are independent if and only if

$$F_{XY}(x,y) = F_X(x)F_Y(y), \quad \forall x, y$$

this is equivalent to

Discrete Random Variables

$$p_{XY}(x,y) = p_X(x)p_Y(y)$$
$$p_{Y|X}(y|x) = p_Y(y)$$

Continuous Random Variables

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$
$$f_{Y|X}(y|x) = f_{Y|X}(y)$$

If X and Y are independent, so are any pair of functions $g(X)$ and $h(Y)$

Expected Values and Covariance

the expected value of $Z = g(X, Y)$ is defined as

$$\mathbf{E}[Z] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{XY}(x,y)\, dx\, dy \qquad X, Y \text{ continuous}$$

$$\mathbf{E}[Z] = \sum_x \sum_y g(x,y)\, p_{XY}(x,y) \qquad X, Y \text{ discrete}$$

- $\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$
- $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$ if X and Y are independent

Covariance of X and Y

$$\mathbf{cov}(X,Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$$

- $\mathbf{cov}(X,Y) = 0$ if X and Y are independent (the converse is NOT true)

Correlation Coefficient

denote

$$\sigma_X = \sqrt{\mathbf{var}(X)}, \quad \sigma_Y = \sqrt{\mathbf{var}(Y)}$$

the standard deviations of X and Y

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

$$\rho_{XY} = \frac{\mathbf{cov}(X,Y)}{\sigma_X \sigma_Y}$$

- $-1 \leq \rho_{XY} \leq 1$
- ρ_{XY} gives the linear dependence between X and Y: for $Y = aX + b$,

$$\rho_{XY} = 1 \quad \text{if } a > 0 \quad \text{and} \quad \rho_{XY} = -1 \quad \text{if } a < 0$$

- X and Y are said to be **uncorrelated** if $\rho_{XY} = 0$

if X and Y are *independent* then X and Y are *uncorrelated*

but the converse is NOT true

Law of Total Variance

suppose that X and Y are random variables

$$\mathbf{var}(Y) = \mathbf{E}[\mathbf{var}(Y|X)] + \mathbf{var}(\mathbf{E}[Y|X])$$

aka Eve's Law; we say the unconditional variance equals EV plus VE

Proof. using $\mathbf{E}[\mathbf{E}[Y|X]] = \mathbf{E}[Y]$

$$\begin{aligned}
\mathbf{var}(Y) &= \mathbf{E}[Y^2] - (\mathbf{E}[Y])^2 \\
&= \mathbf{E}[\mathbf{E}[Y^2|X]] - (\mathbf{E}[\mathbf{E}[Y|X]])^2 \\
&= \mathbf{E}[\mathbf{var}(Y|X)] + \mathbf{E}[(\mathbf{E}[Y|X])^2] - (\mathbf{E}[\mathbf{E}[Y|X]])^2 \\
&= \mathbf{E}[\mathbf{var}(Y|X)] + \mathbf{var}[\mathbf{E}[Y|X]]
\end{aligned}$$

Moment Generating Functions

the moment generating function (MGF) $\Phi(t)$ is defined for all t by

$$\Phi(t) = \mathbf{E}[e^{tX}] = \begin{cases} \int_{-\infty}^{\infty} e^{tx} f(x) dx, & X \text{ is continuous} \\ \sum_x e^{tx} p(x), & X \text{ is discrete} \end{cases}$$

- except for a sign change, $\Phi(t)$ is the 2-sided Laplace transform of pdf
- knowing $\Phi(t)$ is equivalent to knowing $f(x)$
- $\mathbf{E}[X^n] = \frac{d^n \Phi(t)}{dt^n}\Big|_{t=0}$
- MGF of the sum of independent RVs is the product of the individual MGF

$$\Phi_{X+Y}(t) = \mathbf{E}[e^{t(X+Y)}] = \Phi_X(t)\Phi_Y(t)$$

Gaussian (Normal) random variables

- arise as the outcome of the *central limit theorem*
- the sum of a *large* number of RVs is distributed approximately normally
- many results involving Gaussian RVs can be derived in analytical form
- let X be a Gaussian RV with parameters mean μ and variance σ^2

Notation $$X \sim \mathcal{N}(\mu, \sigma^2)$$

PDF

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x-\mu)^2}{2\sigma^2}, \quad -\infty < x < \infty$$

Mean $\mathbf{E}[X] = \mu$ **Variance** $\mathbf{var}[X] = \sigma^2$

MGF $\Phi(t) = e^{\mu t + \sigma^2 t^2/2}$

let $Z \sim \mathcal{N}(0,1)$ be the normalized Gaussian variable

CDF of Z is

$$F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt \quad \triangleq \quad \Phi(z)$$

then CDF of $X \sim \mathcal{N}(\mu, \sigma^2)$ can be obtained by

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

in MATLAB, the error function is defined as

$$\mathsf{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

hence, $\Phi(z)$ can be computed via the `erf` command as

$$\Phi(z) = \frac{1}{2}\left[1 + \mathsf{erf}\left(\frac{z}{\sqrt{2}}\right)\right]$$

Gamma random variables

PDF

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

where $\Gamma(z)$ is the gamma function, defined by

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx, \quad z > 0$$

Mean $\mathbf{E}[X] = \frac{\alpha}{\lambda}$ **Variance** $\mathbf{var}[X] = \frac{\alpha}{\lambda^2}$

MGF $\Phi(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha-1}$

- if X_1 and X_2 are independent gamma RVs with parameters (α_1, λ) and (α_2, λ) then $X_1 + X_2$ is a gamma RV with parameters $(\alpha_1 + \alpha_2, \lambda)$

Properties of the gamma function

$$\begin{aligned}\Gamma(1/2) &= \sqrt{\pi} \\ \Gamma(z+1) &= z\Gamma(z) \quad \text{for } z > 0 \\ \Gamma(m+1) &= m!, \quad \text{for } m \text{ a nonnegative integer}\end{aligned}$$

Special cases

a Gamma RV becomes

- exponential RV when $\alpha = 1$
- m-Erlang RV when $\alpha = m$, a positive integer
- chi-square RV with n DF when $\alpha = n/2, \lambda = 1/2$

Chi-square random variables

if $Z_1, Z_2, \ldots, Z_n$ are independent normal RVs, then X defined by

$$X = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi_n^2$$

is said to have a **chi-square** distribution with n degrees of freedom

- $\Phi(t) = \mathbf{E}[\prod_{i=1}^n e^{tZ_i^2}] = \prod_{i=1}^n \mathbf{E}[e^{tZ_i^2}] = (1-2t)^{-n/2}$
- we recognize that X is a gamma RV with parameters $(n/2, 1/2)$
- sum of independent chi-square RVs with n_1 and n_2 DF is the chi-square with $n_1 + n_2$ DF

PDF

$$f(x) = \frac{(1/2)e^{-x/2}(x/2)^{n/2-1}}{\Gamma(n/2)}, \quad x > 0$$

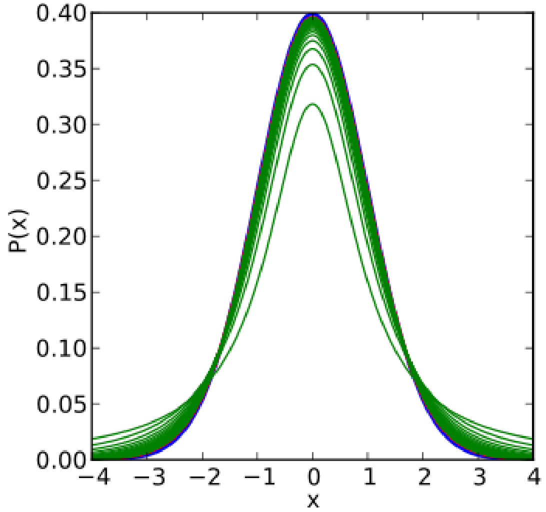
Mean $\mathbf{E}[X] = n$ **Variance** $\mathbf{var}(X) = 2n$

t random variables

if $Z \sim \mathcal{N}(0,1)$ and χ^2_n are independent then

$$T_n = \frac{Z}{\sqrt{\chi^2_n/n}}$$

is said to have a t-**distribution** with n degree of freedom



- t density is symmetric about zero
- t has greater variability than the normal
- $T_n \to Z$ for n large
- for $0 < \alpha < 1$ such that $P(T_n \geq t_{\alpha,n}) = \alpha$),

$$P(T_n \geq -t_{\alpha,n}) = 1 - \alpha \quad \Rightarrow \quad -t_{\alpha,n} = t_{1-\alpha,n}$$

F random variables

if χ_n^2 and χ_m^2 are independent chi-square RVs then the RV $F_{n,m}$ defined by

$$F_{n,m} = \frac{\chi_n^2/n}{\chi_m^2/m}$$

is said to have an F-**distribution** with n and m degree of freedoms

- for any $\alpha \in (0,1)$, let $F_{\alpha,n,m}$ be such that $P(F_{n,m} > F_{\alpha,n,m}) = \alpha$ then

$$P\left(\frac{\chi_m^2/m}{\chi_n^2/n} \geq \frac{1}{F_{\alpha,n,m}}\right) = 1 - \alpha$$

- since $\frac{\chi_m^2/m}{\chi_n^2/n}$ is another $F_{m,n}$ RV, it follows that

$$1 - \alpha = P\left(\frac{\chi_m^2/m}{\chi_n^2/n} \geq F_{1-\alpha,n,m}\right) \quad \Rightarrow \quad \frac{1}{F_{\alpha,n,m}} = F_{1-\alpha,m,n}$$

Logistics random variables

CDF

$$F(x) = \frac{e^{(x-\mu)/\nu}}{1 + e^{(x-\mu)/\nu}}, \quad -\infty < x < \infty, \quad \mu, \nu > 0$$

PDF

$$f(x) = \frac{e^{(x-\mu)/\nu}}{\nu(1 + e^{(x-\mu)/\nu})^2}, \quad -\infty < x < \infty$$

Mean $\mathbf{E}[X] = \mu$

- if $\mu = 0, \nu = 1$ then X is a standard logistic
- μ is the mean of the logistic
- ν is called the dispersion parameter

Multivariate Random Variables

- probabilities
- cross correlation, cross covariance
- Gaussian random vectors

Random vectors

we denote X a random vector

X is a function that maps each outcome ζ to a vector of real numbers

an n-dimensional random variable has n components:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

also called a *multivariate* or *multiple* random variable

Probabilities

Joint CDF

$$F(X) \triangleq F_X(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

Joint PMF

$$p(X) \triangleq p_X(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

Joint PDF

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

Marginal PMF

$$p_{X_j}(x_j) = P(X_j = x_j) = \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p_X(x_1, x_2, \ldots, x_n)$$

Marginal PDF

$$f_{X_j}(x_j) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(x_1, x_2, \ldots, x_n)\, dx_1 \ldots dx_{j-1} dx_{j+1} \ldots dx_n$$

Conditional PDF: the PDF of X_n given $X_1, \ldots, X_{n-1}$ is

$$f(x_n | x_1, \ldots, x_{n-1}) = \frac{f_X(x_1, \ldots, x_n)}{f_{X_1, \ldots, X_{n-1}}(x_1, \ldots, x_{n-1})}$$

Characteristic Function

the characteristic function of an n-dimensional RV is defined by

$$\begin{aligned}\Phi(\omega) = \Phi(\omega_1, \ldots, \omega_n) &= \mathbf{E}[e^{\mathrm{i}(\omega_1 X_1 + \cdots + \omega_n X_n)}] \\ &= \int_X e^{\mathrm{i}\omega^T X} f(X) dX\end{aligned}$$

where

$$\omega = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$\Phi(\omega)$ is the n-dimensional Fourier transform of $f(X)$

Independence

the random variables $X_1, \dots, X_n$ are **independent** if

the joint pdf (or pmf) is equal to the product of their marginal's

Discrete

$$p_X(x_1, \dots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$$

Continuous

$$f_X(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

we can specify an RV by the characteristic function in place of the pdf,

$X_1, \dots, X_n$ are *independent* if

$$\Phi(\omega) = \Phi_1(\omega_1) \cdots \Phi_n(\omega_n)$$

Expected Values

the expected value of a function

$$g(X) = g(X_1, \ldots, X_n)$$

of a vector random variable X is defined by

$$\mathbf{E}[g(X)] = \int_x g(x)f(x)dx \qquad \text{Continuous}$$

$$\mathbf{E}[g(X)] = \sum_x g(x)p(x) \qquad \text{Discrete}$$

Mean vector

$$\mu = \mathbf{E}[X] = \mathbf{E}\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{E}[X_1] \\ \mathbf{E}[X_2] \\ \vdots \\ \mathbf{E}[X_n] \end{bmatrix}$$

Correlation and Covariance matrices

Correlation matrix has the second moments of X as its entries:

$$R \triangleq \mathbf{E}[XX^T] = \begin{bmatrix} \mathbf{E}[X_1X_1] & \mathbf{E}[X_1X_2] & \cdots & \mathbf{E}[X_1X_n] \\ \mathbf{E}[X_2X_1] & \mathbf{E}[X_2X_2] & \cdots & \mathbf{E}[X_2X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{E}[X_nX_1] & \mathbf{E}[X_nX_2] & \cdots & \mathbf{E}[X_nX_n] \end{bmatrix}$$

with

$$R_{ij} = \mathbf{E}[X_iX_j]$$

Covariance matrix has the second-order central moments as its entries:

$$C \triangleq \mathbf{E}[(X-\mu)(X-\mu)^T]$$

with

$$C_{ij} = \mathbf{cov}(X_i, X_j) = \mathbf{E}[(X_i - \mu_i)(X_j - \mu_j)]$$

Properties of correlation and covariance matrices

let X be a (real) n-dimensional random vector with mean μ

Facts:

- R and C are $n \times n$ symmetric matrices
- R and C are positive semidefinite
- If $X_1, \ldots, X_n$ are independent, then C is diagonal
- the diagonals of C are given by the variances of X_k
- if X has zero mean, then $R = C$
- $C = R - \mu\mu^T$

Cross Correlation and Cross Covariance

let X, Y be vector random variables with means μ_X, μ_Y respectively

Cross Correlation

$$\mathbf{cor}(X, Y) = \mathbf{E}[XY^T]$$

if $\mathbf{cor}(X, Y) = 0$ then X and Y are said to be **orthogonal**

Cross Covariance

$$\begin{aligned} \mathbf{cov}(X, Y) &= \mathbf{E}[(X - \mu_X)(Y - \mu_Y)^T] \\ &= \mathbf{cor}(X, Y) - \mu_X \mu_Y^T \end{aligned}$$

if $\mathbf{cov}(X, Y) = 0$ then X and Y are said to be **uncorrelated**

Affine transformation

let Y be an affine transformation of X:

$$Y = AX + b$$

where A and b are deterministic matrices

- $\mu_Y = A\mu_X + b$

$$\mu_Y = \mathbf{E}[AX + b] = A\mathbf{E}[X] + \mathbf{E}[b] = A\mu_X + b$$

- $C_Y = AC_XA^T$

$$\begin{aligned} C_Y &= \mathbf{E}[(Y - \mu_Y)(Y - \mu_Y)^T] = \mathbf{E}[(A(X - \mu_X))(A(X - \mu_X))^T] \\ &= A\mathbf{E}[(X - \mu_X)(X - \mu_X)^T]A^T = AC_XA^T \end{aligned}$$

Gaussian random vector

$X_1, \ldots, X_n$ are said to be **jointly Gaussian** if their joint pdf is given by

$$f(X) \triangleq f_X(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} \text{ exp } -\frac{1}{2}(X - \mu)^T \Sigma^{-1} (X - \mu)$$

μ is the mean $(n \times 1)$ and $\Sigma \succ 0$ is the covariance matrix $(n \times n)$:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1n} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{n1} & \Sigma_{n2} & \cdots & \Sigma_{nn} \end{bmatrix}$$

and

$$\mu_k = \mathbf{E}[X_k], \quad \Sigma_{ij} = \mathbf{E}[(X_i - \mu_i)(X_j - \mu_j)]$$

example: the joint density function of X (not normalized) is given by

$$f(x_1, x_2, x_3) = \exp -\frac{x_1^2 + 3x_2^2 + 2(x_3-1)^2 + 2x_1(x_3-1)}{2}$$

- f is an exponential of *negative quadratic* in x so X must be a Gaussian

$$f(x_1, x_2, x_3) = \exp -\frac{1}{2}\begin{bmatrix} x_1 \\ x_2 \\ x_3 - 1 \end{bmatrix}^T \begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 - 1 \end{bmatrix}$$

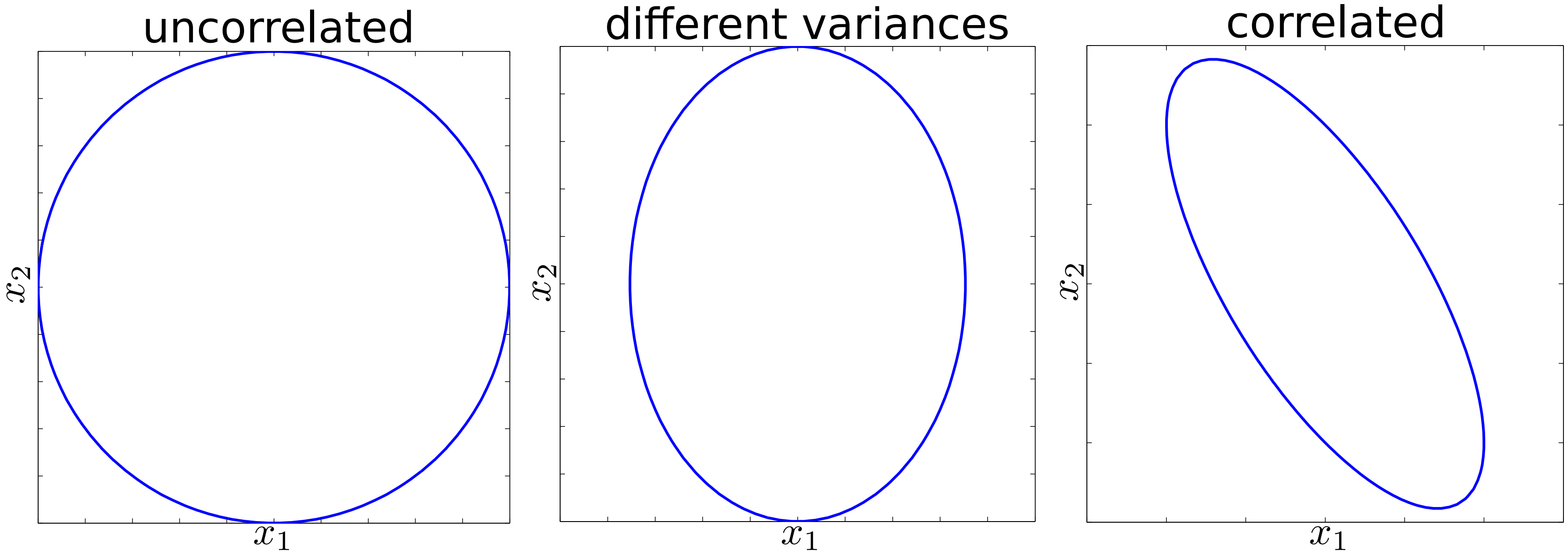
- the mean vector is $(0, 0, 1)$
- the covariance matrix is

$$C = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 1/3 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

- the variance of x_1 is highest while x_2 is smallest
- x_1 and x_2 are uncorrelated, so are x_2 and x_3

examples of Gaussian density contour (the exponent of exponential)

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 1$$



$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

Properties of Gaussian variables

many results on Gaussian RVs can be obtained analytically:

- marginal's of X is also Gaussian
- conditional pdf of X_k given the other variables is a Gaussian distribution
- uncorrelated Gaussian random variables are *independent*
- any affine transformation of a Gaussian is also a Gaussian

these are well-known facts

and more can be found in the areas of estimation, statistical learning, etc.

Characteristic function of Gaussian

$$\Phi(\omega) = \Phi(\omega_1, \omega_2, \ldots, \omega_n) = e^{\mathrm{i}\mu^T\omega} \; e^{-\frac{\omega^T\Sigma\omega}{2}}$$

Proof. By definition and arranging the quadratic term in the power of exp

$$\begin{aligned}
\Phi(\omega) &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \int_X e^{\mathrm{i}X^T\omega} \; e^{-\frac{(X-\mu)^T\Sigma^{-1}(X-\mu)}{2}} dx \\
&= \frac{e^{\mathrm{i}\mu^T\omega} \; e^{-\frac{\omega^T\Sigma\omega}{2}}}{(2\pi)^{n/2}|\Sigma|^{1/2}} \int_X e^{-\frac{(X-\mu-\mathrm{i}\Sigma\omega)^T\Sigma^{-1}(X-\mu-\mathrm{i}\Sigma\omega)}{2}} dx \\
&= \exp \; (\mathrm{i}\mu^T\omega) \exp \; (-\frac{1}{2}\omega^T\Sigma\omega)
\end{aligned}$$

(the integral equals 1 since it is a form of Gaussian distribution)

for one-dimensional Gaussian with zero mean and variance $\Sigma = \sigma^2$,

$$\Phi(\omega) = e^{-\frac{\sigma^2\omega^2}{2}}$$

Affine Transformation of a Gaussian is Gaussian

let X be an n-dimensional Gaussian, $X \sim \mathcal{N}(\mu, \Sigma)$ and define

$$Y = AX + b$$

where A is $m \times n$ and b is $m \times 1$ (so Y is $m \times 1$)

$$\begin{aligned}
\Phi_Y(\omega) &= \mathbf{E}[e^{\mathrm{i}\omega^T Y}] = \mathbf{E}[e^{\mathrm{i}\omega^T (AX+b)}] \\
&= \mathbf{E}[e^{\mathrm{i}\omega^T AX} \cdot e^{\mathrm{i}\omega^T b}] = e^{\mathrm{i}\omega^T b} \Phi_X(A^T\omega) \\
&= e^{\mathrm{i}\omega^T b} \cdot e^{\mathrm{i}\mu^T A^T \omega} \cdot e^{-\omega^T A \Sigma A^T \omega/2} \\
&= e^{\mathrm{i}\omega^T (A\mu + b)} \cdot e^{-\omega^T A \Sigma A^T \omega/2}
\end{aligned}$$

we read off that Y is Gaussian with mean $A\mu + b$ and covariance $A \Sigma A^T$

Marginal of Gaussian is Gaussian

the k^{th} component of X is obtained by

$$X_k = \begin{bmatrix} 0 & \cdots & 1 & 0 \end{bmatrix} X \quad \triangleq \quad \mathbf{e}_k^T X$$

($\mathbf{e}_k$ is a standard unit column vector; all entries are zero except the k^{th} position)

hence, X_k is simply a linear transformation (in fact, a projection) of X

X_k is then a Gaussian with mean

$$\mathbf{e}_k^T \mu = \mu_k$$

and covariance

$$\mathbf{e}_k^T \, \Sigma \, \mathbf{e}_k = \Sigma_{kk}$$

Uncorrelated Gaussians are independent

suppose (X, Y) is a jointly Gaussian vector with

$$\text{mean } \mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \quad \text{and covariance} \quad \begin{bmatrix} C_X & 0 \\ 0 & C_Y \end{bmatrix}$$

in otherwords, X and Y are *uncorrelated* Gaussians:

$$\mathbf{cov}(X, Y) = \mathbf{E}[XY^T] - \mathbf{E}[X]\mathbf{E}[Y]^T = 0$$

the joint density can be written as

$$\begin{aligned} f_{XY}(x, y) &= \frac{1}{(2\pi)^n |C_X|^{1/2} |C_Y|^{1/2}} \text{ exp } -\frac{1}{2} \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix}^T \begin{bmatrix} C_X^{-1} & 0 \\ 0 & C_Y^{-1} \end{bmatrix} \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix} \\ &= \frac{1}{(2\pi)^{n/2} |C_X|^{1/2}} e^{-\frac{1}{2}(x-\mu_x)^T C_X^{-1} (x-\mu_x)} \cdot \frac{1}{(2\pi)^{n/2} |C_Y|^{1/2}} e^{-\frac{1}{2}(y-\mu_y)^T C_Y^{-1} (y-\mu_y)} \end{aligned}$$

proving the independence

Conditional of Gaussian is Gaussian

let Z be an n-dimensional Gaussian which can be decomposed as

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_{yy} \end{bmatrix} \right)$$

the conditional pdf of X given Y is also Gaussian with conditional mean

$$\mu_{X|Y} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(Y - \mu_y)$$

and conditional covariance

$$\Sigma_{X|Y} = \Sigma_x - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^T$$

Proof:

from the **matrix inversion lemma**, Σ^{-1} can be written as

$$\Sigma^{-1} = \begin{bmatrix} S^{-1} & -S^{-1}\Sigma_{xy}\Sigma_{yy}^{-1} \\ -\Sigma_{yy}^{-1}\Sigma_{xy}^T S^{-1} & \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\Sigma_{xy}^T S^{-1}\Sigma_{xy}\Sigma_{yy}^{-1} \end{bmatrix}$$

where S is called the **Schur complement of** Σ_{xx} **in** Σ and

$$\begin{aligned} S &= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^T \\ \det \Sigma &= \det S \cdot \det \Sigma_{yy} \end{aligned}$$

we can show that $\Sigma \succ 0$ if any only if $S \succ 0$ and $\Sigma_{yy} \succ 0$

from $f_{X|Y}(x|y) = f_X(x,y)/f_Y(y)$, we calculate the exponent terms

$$\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix} - (y-\mu_y)^T \Sigma_{yy}^{-1}(y-\mu_y)$$

$$\begin{aligned}
&= (x-\mu_x)^T S^{-1}(x-\mu_x) - (x-\mu_x)^T S^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y) \\
&\quad -(y-\mu_y)^T \Sigma_{yy}^{-1}\Sigma_{xy}^T S^{-1}(x-\mu_x) \\
&\quad +(y-\mu_y)^T(\Sigma_{yy}^{-1}\Sigma_{xy}^T S^{-1}\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y) \\
&= [x-\mu_x-\Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y)]^T S^{-1}[x-\mu_x-\Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y)] \\
&\triangleq (x-\mu_{X|Y})^T \Sigma_{X|Y}^{-1}(x-\mu_{X|Y})
\end{aligned}$$

$f_{X|Y}(x|y)$ is an exponential of quadratic function in x

so it has a form of Gaussian

Standard Gaussian vectors

for an n-dimensional Gaussian vector $X \sim \mathcal{N}(\mu, C)$ with $C \succ 0$

let A be an $n \times n$ invertible matrix such that

$$AA^T = C$$

(A is called a **factor** of C)

then the random vector

$$Z = A^{-1}(X - \mu)$$

is a standard Gaussian vector, *i.e.*,

$$Z \sim \mathcal{N}(0, I)$$

(obtain A via eigenvalue decomposition or Cholesky factorization)

Quadratic Form Theorems

let $X = (X_1, \ldots, X_n)$ be a standard n-dimensional Gaussian vector:

$$X \sim \mathcal{N}(0, I)$$

then the following results hold

- $X^T X \sim \chi^2(n)$
- let A be a symmetric and *idempotent* matrix and $m = \mathbf{tr}(A)$ then

$$X^T A X \quad \sim \quad \chi^2(m)$$

Proof: the eigenvalue decomposition of A: $A = UDU^T$ where

$$\lambda(A) = 0, 1 \quad U^T U = UU^T = I$$

it follows that

$$X^T A X = X^T U D U^T X = Y^T D Y = \sum_{i=1}^{n} d_{ii} Y_i^2$$

- since U is orthogonal, Y is also a standard Gaussian vector
- since A is idempotent, d_{ii} is either 0 or 1 and $\mathbf{tr}(D) = m$

therefore $X^T A X$ is the m-sum of standard normal RVs

References

Chapter 1-3 in

J.M. Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, 2nd edition, the MIT press, 2010

Chapter 1-5 in

S.M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, 4th edition, Academic press, 2009

Appendix A in

A.C. Cameron and P.K. Trevedi, *Microeconometrics: Method and Applications*, Cambridge University Press, 2005