BIOS 660/BIOS 672 (3 Credits):
Probability and Statistical Inference I

Jianwen Cai

https://sakai.unc.edu/portal/site/bios660-bios672-3-credits

Notes 15

Conditional Expectation

Suppose we have discrete rvs X and Y with conditional pmf $f_{Y|X}(y|x)$. The *conditional expectation* of $g(Y)$ given $X = x$ is

$$\mathsf{E}[g(Y)|X = x] = \sum_y g(y)\ f_{Y|X}(y|x)$$

Notice that this is a function $h(x)$ of x. We may define the random variable

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

In particular, if $X = x$ then $\mathsf{E}[g(Y)|X] = \mathsf{E}[g(Y)|x]$.

For continuous rvs:

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

Example

Let $X \sim Exp(\lambda)$, $Z \sim Exp(\lambda)$, independent, be the times to mitosis of a cell and its daughter. Define the total time $Y = X + Z$.

Given $X = x > 0$, $Y = x + Z$ is a shifted exponential rv

$$f_{Y|X}(y|x) = \lambda e^{-\lambda(y-x)}, \qquad y > x,$$

The conditional expectation of Y given $X = x$ is

$$\mathsf{E}[Y|X = x] = \int_x^{\infty} y\ \lambda e^{-\lambda(y-x)}\, dy = x + 1/\lambda$$

yielding the rv

$$\mathsf{E}[Y|X] = X + 1/\lambda$$

More directly

$$\mathsf{E}(Y|X) = \mathsf{E}(X + Z|X) = \mathsf{E}(X|X) + \mathsf{E}(Z|X) = X + 1/\lambda$$

Iterative expectation formula

If X and Y are any two random variables then

$$\mathsf{E}X = \mathsf{E}(\mathsf{E}(X|Y))$$

Proof:

$$\begin{aligned}
\mathsf{E}(\mathsf{E}(X|Y)) &= \int \mathsf{E}(X|y)f(y)dy \\
&= \int \left[\int xf(x|y)dx\right] f(y)dy \\
&= \int\int xf(x,y)dxdy \\
&= \mathsf{E}X
\end{aligned}$$

Notice that the expectations are with respect to different variables and densities. Careful!

Example

Back to the cells example:

$$\mathsf{E}(Y|X) = X + 1/\lambda$$

so

$$\mathsf{E}[\mathsf{E}(Y|X)] = \mathsf{E}[X + 1/\lambda] = \mathsf{E}X + 1/\lambda = 1/\lambda + 1/\lambda = 2/\lambda$$

This should be equal to $\mathsf{E}Y$, which is

$$\mathsf{E}Y = \mathsf{E}(X + Z) = \mathsf{E}X + \mathsf{E}Z = 1/\lambda + 1/\lambda = 2/\lambda$$

Hierarchical Model

Example: Suppose an insect lays eggs according to a Poisson(λ) and each egg survives with probability p. Assume that the survival of eggs is independent of each other, then what is the average number of eggs surviving?

Let's say $Y \sim$ Poisson(λ) and $X|Y \sim$ Binomial(Y, p) where X is the total number of eggs surviving.

$$
\begin{aligned}
P(X=x) &= \sum_{y=0}^{\infty} P(X=x, Y=y) \\
&= \sum_{y=0}^{\infty} P(X=x|Y=y)P(Y=y) \\
&= \sum_{y=x}^{\infty} \left[\binom{y}{x} p^x (1-p)^{y-x}\right] \left[\frac{e^{-\lambda}\lambda^y}{y!}\right] \\
&= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{y=x}^{\infty} \frac{((1-p)\lambda)^{y-x}}{(y-x)!}
\end{aligned}
$$

cont

$$
\begin{aligned}
&= \frac{(\lambda p)^x e^{-\lambda}}{x!} \exp((1-p)\lambda) \\
&= \frac{(\lambda p)^x}{x!} \exp(-\lambda p)
\end{aligned}
$$

So $X \sim$ Poisson (λp) and E$X = \lambda p$.

Using the iterative expectation formula

$$
\mathrm{E}X = \mathrm{E}(\mathrm{E}(X|Y)) = \mathrm{E}(Yp) = \lambda p
$$

Three-layer herarchical model

Example: In the previous example, suppose that there are several insect mothers, each with different average number of eggs. Model λ of the Poisson as being an exponential rv Λ with mean parameter β. Now what is the expected number of eggs surviving?

$$\begin{aligned} \mathsf{E}X &= \mathsf{E}(\mathsf{E}(X|Y)) \\ &= \mathsf{E}(pY) \\ &= \mathsf{E}(\mathsf{E}(pY|\Lambda)) \\ &= \mathsf{E}(p\Lambda) \\ &= p\beta \end{aligned}$$

Conditional Variance

Suppose we have rvs X and Y. Recall that the marginal variance of $g(Y)$ is

$$\mathsf{Var}[g(Y)] = \mathsf{E}[g(Y) - \mathsf{E}(g(Y))]^2$$

The *conditional variance* of $g(Y)$ given X is

$$\mathsf{Var}[g(Y)|X] = \mathsf{E}\big\{[g(Y) - \mathsf{E}(g(Y)|X)]^2|X\big\}$$

where both expectations are taken with respect to the conditional pmf or pdf $f_{Y|X}(y)$.
Like the conditional expectation, the conditional variance of Y given X is a random variable whose value depends on the rv X.

Conditional variance hierarchical formula

For any two random variables X and Y,

$$\mathsf{Var}X = \mathsf{E}(\mathsf{Var}(X|Y)) + \mathsf{Var}(\mathsf{E}(X|Y))$$

Proof

$$\begin{aligned}
\mathsf{Var}X &= \mathsf{E}\big\{[X - \mathsf{E}X]^2\big\} \\
&= \mathsf{E}\big\{[X - \mathsf{E}(X|Y) + \mathsf{E}(X|Y) - \mathsf{E}X]^2\big\} \\
&= \mathsf{E}\big\{[X - \mathsf{E}(X|Y)]^2\big\} + \mathsf{E}\big\{[\mathsf{E}(X|Y) - \mathsf{E}X]^2\big\} \\
&\quad +2\mathsf{E}\big\{[X - \mathsf{E}(X|Y)][\mathsf{E}(X|Y) - \mathsf{E}X]\big\}
\end{aligned}$$

Study the three terms separately. 1st term:

$$\begin{aligned}
\mathsf{E}\big\{[X - \mathsf{E}(X|Y)]^2\big\} &= \mathsf{E}\Big(\mathsf{E}\big\{[X - \mathsf{E}(X|Y)]^2|Y\big\}\Big) \\
&= \mathsf{E}(\mathsf{Var}(X|Y))
\end{aligned}$$

cont

2nd term:

$$\begin{aligned}
\mathsf{E}\big\{[\mathsf{E}(X|Y) - \mathsf{E}X]^2\big\} &= \mathsf{E}\big\{[\mathsf{E}(X|Y) - \mathsf{E}(\mathsf{E}(X|Y))]^2\big\} \\
&= \mathsf{Var}(\mathsf{E}(X|Y))
\end{aligned}$$

3rd term:

$$\begin{aligned}
&\mathsf{E}\big\{[X - \mathsf{E}(X|Y)][\mathsf{E}(X|Y) - \mathsf{E}X]\big\} \\
&\quad = \mathsf{E}\Big(\mathsf{E}\big\{[X - \mathsf{E}(X|Y)][\mathsf{E}(X|Y) - \mathsf{E}X]|Y\big\}\Big) \\
&\quad = \mathsf{E}\Big([\mathsf{E}(X|Y) - \mathsf{E}X]\ \mathsf{E}\big\{[X - \mathsf{E}(X|Y)]|Y\big\}\Big) \\
&\quad = \mathsf{E}\Big([\mathsf{E}(X|Y) - \mathsf{E}X]\ \big\{\mathsf{E}[X|Y] - \mathsf{E}(X|Y)\big\}\Big) \\
&\quad = 0
\end{aligned}$$

Example

In the the Poisson-Binomial hierarchical model, we had

$$Y \sim \text{Poisson}(\lambda), \qquad X|Y \sim \text{Binomial}(Y, p)$$

and showed that $\mathsf{E}X = \lambda p$. Using the conditional variance formula,

$$\begin{aligned} \text{Var}X &= \mathsf{E}(\text{Var}(X|Y)) + \text{Var}(\mathsf{E}(X|Y)) \\ &= \mathsf{E}(Yp(1-p)) + \text{Var}(Yp) \\ &= \lambda p(1-p) + \lambda p^2 \\ &= \lambda p \end{aligned}$$

This is consistent with the result that $X \sim$Poisson(λp).

Covariance and Correlation

Definitions

Let X and Y be two random variables with respective means μ_X, μ_Y and variances $\sigma_X^2 > 0$ and $\sigma_Y^2 > 0$, all assumed to exist.

- The *covariance* of X and Y is

$$\text{Cov}(X, Y) = \mathsf{E}[(X - \mu_X)(Y - \mu_Y)] = \sigma_{XY}$$

- The *correlation* between X and Y is

$$\begin{aligned} \rho_{XY} &= \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}X \ \text{Var}Y}} = \frac{\sigma_{XY}}{\sigma_X \, \sigma_Y} \\ &= \mathsf{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right] \end{aligned}$$

- X and Y are called *uncorrelated* iff

$$\text{Cov}(X, Y) = 0 \qquad \text{or equivalently} \qquad \rho_{XY} = 0$$

Properties of Covariance

1. $\text{Cov}(X, X) = \text{Var}(X)$
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
3. For any constant c, $\text{Cov}(X, c) = 0$
4. $\text{Cov}(X, Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y)$
5. If X and Y are independent and $\text{Cov}(X, Y)$ exists, then $\text{Cov}(X, Y) = 0$.

Note: If X and Y are uncorrelated, this does not imply that they are independent. Example:

$$X \sim U[-1, 1], \qquad Y = \begin{cases} X, & \text{prob. } 1/2 \\ -X, & \text{prob. } 1/2 \end{cases}$$

Linear Combinations

If X, Y, and Z are rvs each with a variance, and a and b are constants, then

$$\begin{aligned} \text{Cov}(aX + bY, Z) &= a\text{Cov}(X, Z) + b\text{Cov}(Y, Z) \\ \text{Var}(aX + bY) &= a^2\text{Var}X + b^2\text{Var}Y + 2ab\text{Cov}(X, Y) \\ \text{Corr}(aX + b, cY + d) &= sign(ac)\text{Corr}(X, Y) \end{aligned}$$

Proof:

Correlation Coefficient

Theorem C-B 4.5.7 For any rvs X and Y,

1. $-1 \leq \rho_{XY} \leq 1$
2. $|\rho_{XY}| = 1$ if and only if $\exists\, a \neq 0$ and b such that

$$P(Y = aX + b) = 1.$$

If $\rho_{XY} = 1 \Rightarrow a > 0$, and if $\rho_{XY} = -1 \Rightarrow a < 0$.

Proof: Let $\tilde{X} = (X - \mu_X)/\sigma_X$, $\tilde{Y} = (Y - \mu_Y)/\sigma_Y$, so that $\rho_{XY} = \mathsf{E}(\tilde{X}\tilde{Y})$.

1.
$$\begin{aligned} 0 \leq \mathsf{E}(\tilde{X} - \tilde{Y})^2 = 1 + 1 - 2\mathsf{E}(\tilde{X}\tilde{Y}) \quad &\Rightarrow \quad \mathsf{E}(\tilde{X}\tilde{Y}) \leq 1 \\ 0 \leq \mathsf{E}(\tilde{X} + \tilde{Y})^2 = 1 + 1 + 2\mathsf{E}(\tilde{X}\tilde{Y}) \quad &\Rightarrow \quad -1 \leq \mathsf{E}(\tilde{X}\tilde{Y}) \end{aligned}$$

2.
$$\begin{aligned} \rho_{XY} = 1 \text{ iff } P(\tilde{Y} = \tilde{X}) = 1 \quad &\Rightarrow \quad a > 0 \\ \rho_{XY} = -1 \text{ iff } P(\tilde{Y} = -\tilde{X}) = 1 \quad &\Rightarrow \quad a < 0 \end{aligned}$$

Bivariate Normal

Standard Bivariate Normal

Given a number $-1 \leq \rho \leq 1$, define the *standard bivariate normal density* of $(X, Y) \in \mathbb{R}^2$ by

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

Properties:

1. The marginal distribution of X is $N(0, 1)$.
2. The marginal distribution of Y is $N(0, 1)$.
3. The correlation of X and Y is ρ.
4. The conditional distributions are normal:

$$Y|X \sim N(\rho X, 1 - \rho^2), \qquad X|Y \sim N(\rho Y, 1 - \rho^2)$$

The means are the *regression lines* of Y on X and X on Y respectively.

Bivariate Normal

Let $\tilde{X}$ and $\tilde{Y}$ have a standard bivariate normal distribution with correlation ρ. Let

$$\begin{aligned} X &= \mu_X + \sigma_X \tilde{X}, \qquad \mu_X \in \mathbb{R}, \sigma_X > 0 \\ Y &= \mu_Y + \sigma_Y \tilde{Y}, \qquad \mu_Y \in \mathbb{R}, \sigma_Y > 0 \end{aligned}$$

Then (X, Y) has the *bivariate normal density*

$$\begin{aligned} f_{XY}(x,y) &= \left(2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}\right)^{-1} \\ &\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right.\right. \\ &- \left.\left. 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\} \end{aligned}$$

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

Properties

1. The marginal distribution of X is $N(\mu_X, \sigma_X^2)$.
2. The marginal distribution of Y is $N(\mu_Y, \sigma_Y^2)$.
3. The correlation between X and Y is ρ.
4. The conditional distributions are normal:

$$Y|X \sim N\left[\mu_Y + \rho(\sigma_Y/\sigma_X)(x-\mu_X),\ \sigma_Y^2(1-\rho^2)\right]$$

   The mean is the *regression line* of Y on X.
5. For any constants a and b, the distribution of $aX + bY$ is

$$N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y)$$

Proof: HW
(Suggestion: standardize the variables first)

Multivariate Distributions

The n-dimensional random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)^\top$ defined on the triplet $(S, \mathcal{B}, P)$ takes values on the hyperspace $\mathcal{R}^n$.

If the variables $X_1, \ldots, X_n$ are *discrete* then we have a discrete random vector, if the X's are *continuous*, then we have a continuous random vector.

If $\mathbf{X}$ is discrete then

$$P(\mathbf{X} \in A) = \sum_{\mathbf{x} \in A} f(\mathbf{x})$$

where $f(\mathbf{x})$ denotes the *joint pmf*.

If $\mathbf{X}$ is continuous then

$$P(\mathbf{X} \in A) = \int \ldots \int_A f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

where $f(\mathbf{x})$ denotes the *joint pdf*.

Marginals and Conditionals

Definition The *marginal pdf* or *pmf* of any subset of the coordinates of $(X_1, \ldots, X_n)$ can be computed by integrating or summing the joint pdf or pmf over all possible values of the other coordinates.

Definition The *conditional pdf* or *pmf* of a subset of the coordinates of $(X_1, \ldots, X_n)$ given the values of the remaining coordinates is obtained by dividing the full joint pdf or pmf by the joint pdf or pmf of the conditioning variates:

$$f(x_{k+1}, \ldots, x_n | x_1, \ldots, x_k) = \frac{f(x_1, \ldots, x_n)}{f(x_1, \ldots, x_k)}$$

Multinomial Distribution

Let n and m be positive integers and let $p_1, \ldots, p_n$ be probabilities summing to one. Then the random vector $(X_1, \ldots, X_n)$ has a *multinomial distribution with m trials and cell probabilities* $p_1, \ldots, p_n$ if its joint pmf is

$$\begin{aligned} f(x_1, \ldots, x_n) &= \binom{m}{x_1 \cdots x_n} p_1^{x_1} \ldots p_n^{x_n} \\ &= \frac{m!}{x_1! \ldots x_n!} p_1^{x_1} \ldots p_n^{x_n} = m! \prod_{j=1}^{n} \frac{p_j^{x_j}}{x_j!} \end{aligned}$$

for $x_i = 0, \ldots, m$, $i = 1, \ldots, n$, and $x_1 + \cdots + x_n = m$.

The proof that this is a pmf is called the Multinomial Theorem.

Multinomial Theorem 4.6.4

Let m and n be positive integers. Let $\mathcal{A}$ be the set of all vectors $\mathbf{x} = (x_1, \ldots, x_n)$ which are such that the sum of their nonnegative integer components is m, i.e. $\sum_{j=1}^{n} x_j = m$ and $x_j \geq 0$.Then, for any real numbers $p_1, \ldots, p_n$

$$(p_1 + \ldots + p_n)^m = \sum_{\mathbf{x} \in \mathcal{A}} \frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n}$$

E.g. for $n = 2$ we have the binomial theorem.

E.g. for $n = 3$:

$$(p_1 + p_2 + p_3)^m = \sum_{x_1=0}^{m} \sum_{x_2=0}^{m-x_1} \frac{m!}{x_1!\, x_2!\, (m - x_1 - x_2)!} p_1^{x_1}\, p_2^{x_2}\, p_3^{m-x_1-x_2}$$

Multinomial Distribution: Properties

- Marginals are multinomials. E.g.

$$\begin{aligned} X_1 &\sim \text{Binomial}(m, p_1) \\ (X_1, X_2, m - X_1 - X_2) &\sim \text{Multinomial}(m, p_1, p_2, 1 - p_1 - p_2) \end{aligned}$$

- Conditionals are multinomials. E.g.

$$(X_1, \ldots, X_{n-1}) | [X_n = x_n]$$
$$\sim \text{Multinomial}\left(m - x_n, \frac{p_1}{1 - p_n}, \ldots, \frac{p_{n-1}}{1 - p_n}\right)$$

- Variance and covariance:

$$\text{Cov}(X_j, X_k) = \text{E}[(X_j - mp_j)(X_k - mp_k)] = \begin{cases} mp_j(1 - p_j), & j = k \\ -mp_j p_k, & j \neq k \end{cases}$$

Multivariate Independence

Independent Random Vectors

Let $\mathbf{X_1}, \ldots, \mathbf{X_n}$ be random vectors with joint pdf or pmf $f(\mathbf{x_1}, \ldots, \mathbf{x_n})$. Let $f_{\mathbf{X_j}}(\mathbf{x_j})$ denote the marginal pdf or pmf of $\mathbf{X_j}$. Then $\mathbf{X_1}, \ldots, \mathbf{X_n}$ are called *mutually independent random vectors* if, for every $(\mathbf{x_1} \ldots, \mathbf{x_n})$,

$$f(\mathbf{x_1}, \ldots, \mathbf{x_n}) = \prod_{j=1}^{n} f_{\mathbf{X_j}}(\mathbf{x_j})$$

cont.

Theorem C-B 4.6.6 (Generalization of 4.2.10) Let $X_1, \ldots, X_n$ be independent rvs. Let $g_1, \ldots, g_n$ be real-valued functions such that $g(x_j)$ is only a function of x_j. Then

$$\mathsf{E}\left[g_1(X_1) \ldots g_n(X_n)\right] = \mathsf{E}g_1(X_1) \ldots \mathsf{E}g_n(X_n)$$

Theorem C-B 4.6.7 (Generalization of 4.2.12) Let $X_1, \ldots, X_n$ be mutually independent rvs with characteristic functions $\phi_{X_1}(\theta), \ldots, \phi_{X_2}(\theta)$. Let $Z = X_1 + \ldots + X_n$. Then the characteristic function of Z is

$$\phi_Z(\theta) = \prod_{j=1}^{n} \phi_{X_j}(\theta)$$

Note simplification if $\phi_{X_j}(\theta) = \phi_X(\theta)$.

Specific Characteristic Functions

	mgf	cf
Bernoulli(p)	$pe^t + q$	$pe^{it} + q$
Binomial(n,p)	$(pe^t + q)^n$	$(pe^{it} + q)^n$
Poisson(λ)	$e^{\lambda(e^{t-1})}$	$e^{\lambda(e^{it}-1)}$
Geometric(p)	$pe^t/(1-qe^t)$	$pe^{it}/(1-qe^{it})$
Negbin(n,p)	$\left[\frac{pe^t}{1-qe^t}\right]^n$	$\left[\frac{pe^{it}}{1-qe^{it}}\right]^n$
Uniform(a,b)	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$
Normal(μ,σ^2)	$e^{t\mu+\frac{1}{2}\sigma^2 t^2}$	$e^{it\mu-\frac{1}{2}\sigma^2 t^2}$
Exponential(λ)	$\frac{\lambda}{\lambda-t} = (1-\frac{t}{\lambda})^{-1}$	$(1-\frac{it}{\lambda})^{-1}$
Gamma(a,λ)	$(1-\frac{t}{\lambda})^{-a}$	$(1-\frac{it}{\lambda})^{-a}$

Multivariate moments

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ be a random vector.
The *mean vector* $\boldsymbol{\mu}$ is

$$\boldsymbol{\mu} = \mathsf{E}(\boldsymbol{X}) = \begin{pmatrix} \mathsf{E}X_1 \\ \vdots \\ \mathsf{E}X_n \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

We can also define the second moment matrix

$$\mathsf{E}(\boldsymbol{X}\boldsymbol{X}^\top) = \begin{pmatrix} \mathsf{E}X_1^2 & \mathsf{E}X_1X_2 & \cdots & \mathsf{E}X_1X_n \\ \mathsf{E}X_2X_1 & \mathsf{E}X_2^2 & \cdots & \mathsf{E}X_2X_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{E}X_nX_1 & \mathsf{E}X_nX_2 & \cdots & \mathsf{E}X_n^2 \end{pmatrix}$$

Multivariate covariance

The *variance-covariance matrix* $\boldsymbol{\Sigma}$ is defined as

$$\begin{aligned} \boldsymbol{\Sigma} &= \mathsf{Cov}\left[\boldsymbol{X} - \boldsymbol{\mu}, (\boldsymbol{X} - \boldsymbol{\mu})^\top\right] \\ &= \mathsf{E}\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^\top\right] \\ &= \begin{pmatrix} \mathsf{Var}(X_1) & \mathsf{Cov}(X_1, X_2) & \cdots & \mathsf{Cov}(X_1, X_n) \\ \mathsf{Cov}(X_2, X_1) & \mathsf{Var}(X_2) & \cdots & \mathsf{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{Cov}(X_n, X_1) & \mathsf{Cov}(X_n, X_2) & \cdots & \mathsf{Var}(X_n) \end{pmatrix} \end{aligned}$$

Notice that $\boldsymbol{\Sigma}$ is a symmetric matrix.

Bivariate normal

Let X and Y are bivariate normal, then

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$$

The joint density of the vector $\boldsymbol{X} = (X, Y)^\top$ can be written as

$$f_{\boldsymbol{X}}(\boldsymbol{X}) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left[-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right]$$

Proof:

Linear functions

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ be a random vector with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. For a vector $\boldsymbol{c} = (c_1, c_2, \ldots, c_n)^\top \in \mathbb{R}^n$, define

$$Y = \boldsymbol{c}^\top \boldsymbol{X} = \sum_{i=1}^{n} c_i X_i$$

Then

$$\begin{aligned} \mathsf{E}(Y) &= \boldsymbol{c}^\top \boldsymbol{\mu} \\ \mathsf{Var}(Y) &= \boldsymbol{c}^\top \Sigma \boldsymbol{c} \end{aligned}$$

Proof:

Positive definiteness

Definition:

- An $n \times n$ symmetric matrix $\mathbf{\Lambda}$ is called *positive semi-definite* or *nonnegative definite* iff for every vector $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{c}^\top \mathbf{\Lambda} \boldsymbol{c} \geq 0$.
- An $n \times n$ symmetric matrix $\mathbf{\Lambda}$ is called *positive definite* iff for every vector $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{c}^\top \mathbf{\Lambda} \boldsymbol{c} > 0$.

Properties:

- A positive definite matrix:
  - is invertible
  - its determinant is positive
  - all its eigenvalues are positive
- The variance-covariance matrix is positive semi-definite.

Multivariate linear transformations

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ have joint pdf $f_{\boldsymbol{X}}(\boldsymbol{X})$ and let

$$\begin{aligned}
Y_1 &= a_{11}X_1 + a_{12}X_2 + ...a_{1n}X_n \\
Y_2 &= a_{21}X_1 + a_{22}X_2 + ...a_{2n}X_n \\
&\vdots \\
Y_n &= a_{n1}X_1 + a_{n2}X_2 + ...a_{nn}X_n
\end{aligned}$$

Using vector and matrix notation, we can write this transformation as

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$$

where

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}$$

Multivariate linear transformations

The Jacobian of the transformation

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$$

is

$$\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}^\top} = \begin{pmatrix} \partial y_1/\partial x_1 & \dots & \partial y_1/\partial x_n \\ \vdots & \ddots & \vdots \\ \partial y_n/\partial x_1 & \dots & \partial y_n/\partial x_n \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \boldsymbol{A}$$

with determinant $\det(\boldsymbol{A})$.

The inverse transformation is

$$\boldsymbol{X} = \boldsymbol{A}^{-1}\boldsymbol{Y}$$

with Jacobian determinant $\det(\boldsymbol{A}^{-1}) = 1/\det(\boldsymbol{A})$.

Multivariate normal

Let $\boldsymbol{\mu}$ be a vector in $\mathbb{R}^n$ and $\boldsymbol{\Sigma}$ be an $n \times n$ symmetric positive definite matrix. The vector $\boldsymbol{X} = (X_1, \dots, X_n)$ has a *multivariate normal distribution* with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ if it has joint density

$$f_{\boldsymbol{X}}(\boldsymbol{X}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}} \exp\left[-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right]$$

In particular:

$$\begin{aligned} \boldsymbol{\mu} &= \mathsf{E}\boldsymbol{X} \\ \boldsymbol{\Sigma} &= \mathsf{E}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^\top] \end{aligned}$$

Construction

Let $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ be i.i.d. $N(0,1)$ variables. Then

$$f_{\boldsymbol{Z}}(z_1, \ldots, z_n) = \prod_{i=1}^{n} \frac{e^{-z_i^2/2}}{\sqrt{2\pi}} = \frac{e^{-\sum_{i=1}^{n} z_i^2/2}}{(2\pi)^{n/2}} = \frac{e^{-\boldsymbol{Z}^\top \boldsymbol{Z}/2}}{(2\pi)^{n/2}}$$

Now define

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{\mu}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ is a vector and $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is an invertible matrix.
The inverse transformation is

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

with Jacobian

$$J = \frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{X}^\top} = \boldsymbol{A}^{-1}$$

and determinant $|J| = 1/|\boldsymbol{A}|$.

Construction (cont.)

The joint pdf of $\boldsymbol{X}$ is

$$f_{\boldsymbol{X}}(\boldsymbol{X}) = \frac{e^{-(\boldsymbol{X}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})/2}}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}$$

where $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^\top$. We can confirm that

$$\mathsf{E}(\boldsymbol{X}) = \mathsf{E}(\boldsymbol{A}\boldsymbol{Z} + \boldsymbol{\mu}) = \boldsymbol{\mu}$$

and

$$\begin{aligned} \mathsf{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^\top] &= \mathsf{E}[(\boldsymbol{A}\boldsymbol{Z})(\boldsymbol{A}\boldsymbol{Z})^\top] = \mathsf{E}(\boldsymbol{A}\boldsymbol{Z}\boldsymbol{Z}^\top \boldsymbol{A}^\top) \\ &= \boldsymbol{A}\mathsf{E}(\boldsymbol{Z}\boldsymbol{Z}^\top)\boldsymbol{A}^\top = \boldsymbol{A}\boldsymbol{I}\boldsymbol{A}^\top = \Sigma \end{aligned}$$

Properties

Theorem:

1. Let X and Y be jointly normal random variables. Then X and Y are independent if and only if they are uncorrelated.
2. Let $X_1, \ldots, X_n$ be jointly normal random variables. Then they are mutually independent if and only if they are pairwise uncorrelated.

Theorem: If $\boldsymbol{X}$ has a multivariate normal distribution, then

1. All marginal distributions are normal.
2. All conditional distributions are normal.
3. For any constants $\boldsymbol{A} = (a_1, \ldots, a_n)^\top$ the random variable $Y = \boldsymbol{A}^\top \boldsymbol{X}$ has a normal distribution.