Chapter 5: Multivariate Probability Distributions

Shujia Wang

Contents

1 Basic Concepts for Multivariate Distributions

1.1 Joint, Marginal and Conditional Distributions

Joint Distributions

If there exists a function $f_{XY}(x, y)$ such that

$$f_{XY}(x, y) \geq 0,\ (x, y) \in \mathbb{R}^2 \text{ and } \iint_{\mathbb{R}^2} f_{XY}(x, y)\mathrm{d}x\mathrm{d}y = 1$$

then the function $f_{XY}(x, y)$ is called the **joint probability density function** (joint pdf) of X and Y.

For simplicity it will be denoted by $f(x, y)$.

The **joint cumulative distribution function** (cdf) of X and Y is

$$\begin{aligned} F_{XY}(x, y) &= \mathrm{P}(X \leq x, Y \leq y) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\mathrm{d}v\mathrm{d}u,\ \text{ for } (x, y) \in \mathbb{R}^2 \end{aligned}$$

Remark:

If X and Y are discrete, $f(x, y) = \mathrm{P}(X = x, Y = y)$ and called the *joint probability mass function* (pmf).

Marginal Distributions

If X and Y have the joint pdf $f(x, y)$, then the **marginal density** of X may be recovered by

$$f_X(x) = \int_{\mathbb{R}} f(x, y)\mathrm{d}y$$

Similarly,

$$f_Y(y) = \int_{\mathbb{R}} f(x, y)\mathrm{d}x$$

Remark:

If X and Y are discrete, the integration becomes summation, e. g.

$$f_X(x) = \sum_{y \in S_y} f_{XY}(x, y)$$

Conditional Distributions

If $x \in S_X$ is such that $f_X(x) > 0$, then we define the **conditional density** of $Y|X = x$ by

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

Also useful:

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

1.2 Expectation, Covariance and Correlation Coefficient

Joint and Marginal Expectation

Let $Z = g(X, Y)$ be a bivariate function of (X, Y), then the expectation of Z is

$$\mathrm{E}(Z) = \mathrm{E}[g(X, Y)] = \iint_{\mathbb{R}^2} g(x, y)f(x, y)\ \mathrm{d}x\mathrm{d}y$$

Marginal expectation: $Z = g(X, Y) = X$

$$\mathrm{E}(X) = \iint_{\mathbb{R}^2} xf(x, y)\ \mathrm{d}x\mathrm{d}y$$

$g(X, Y) = XY$:

$$\mathrm{E}(XY) = \iint_{\mathbb{R}^2} xyf(x, y)\ \mathrm{d}x\mathrm{d}y$$

Covariance and Correlation

The **covariance** of X and Y is

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}(X-\mu_X)(Y-\mu_Y)\\ &= \mathrm{E}(XY)-\mathrm{E}(X)\mathrm{E}(Y)\end{aligned}$$

The Pearson product moment **correlation coefficient** between X and Y is

$$\rho_{XY} = \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

Properties:

1. $|\rho_{XY}| \le 1$
2. $|\rho_{XY}| = 1$ if and only if Y is a linear function of X with probability one.

Independent Random Variables

If the joint pdf or pmf of (X,Y) happens that

$$f(x,y) = f_X(x)f_Y(y),\ \ x\in S_X, y\in S_Y$$

then we say that X and Y are **independent**.

Corollary 1. *1. If X and Y are independent, then X and Y are uncorrelated, i. e. Cov(X,Y) = 0, and consequently, Corr(X,Y) = 0.*

Note that the converse is not true.

2. If X and Y are independent, then $f(X)$ and $g(Y)$ are independent for any functions f and g.

1.3 Moment Generating Functions and Bivariate Transformations

Moment Generating Functions

Definitions 2. The **moment generating function** (mgf) of a r.v.X is defined by

$$M(t) = \mathrm{E}(e^{tX})$$

Definition 3. The mgf of the bivariate (X,Y) is

$$M_{XY}(t,s) = \mathrm{E}(e^{tX+sY})$$

Corollary 4. *If X and Y are independent, then $M_{XY}(t,s) = M_X(t)M_Y(s)$. The converse is also true.*

Bivariate Transformations

Theorem 5. *Soppose (X,Y) has joint pdf $f(x,y)$, let $U = g(X,Y),\ V = h(X,Y)$ be one to one functions. Then the joint pdf of (U,V) is*

$$f_{UV}(u,v) = f(x,y)\ \mathrm{abs}|J|$$

where abs$|J|$ *refers to the absolute value of Jacobian*

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

Example 6. If the joint pdf of (X,Y) is

$$f(x,y) = \begin{cases} 2-x-y, & 0<x<1, 0<y<1 \\ 0, & \text{Otherwise} \end{cases}$$

Find the distribution of $U = X+Y$.

Solution:
Let $U = X + Y,\ V = Y$, then $X = U - V,\ Y = V$,

1. Range: $S_{XY} = \{(x,y)|0 < x < 1, 0 < y < 1\},\ S_{UV} = \{(u,v)|v < u < v+1, 0 < v < 1\}$
2. The Jacobian
$$|J| = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$$

The joint pdf of (U,V)
$$f_{UV}(u,v) = f(x,y)|J| = \begin{cases} 2-u, & v < u < v+1, 0 < v < 1 \\ 0, & \text{Otherwise} \end{cases}$$

The pdf of U
$$\begin{aligned} f_U(u) &= \int_{-\infty}^{\infty} f_{UV}(u,v)\mathrm{d}v \\ &= \begin{cases} \int_0^u (2-u)\mathrm{d}v = (2-u)u, & 0 < u < 1 \\ \int_{u-1}^1 (2-u)\mathrm{d}v = (2-u)^2, & 1 \le u < 2 \\ 0, & \text{Otherwise} \end{cases} \end{aligned}$$

2 Multivariate Normal Distribution

2.1 Bivariate Normal

Bivariate Normal

If (X,Y) have a **bivariate normal** distribution with parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$, then the joint pdf $f(x,y)$ of X and Y is
$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_1}\right)^2\right]}$$

We write $(X,Y) \sim \mathrm{N}_2(\boldsymbol{\mu}, \Sigma)$ or
$$(X,Y) \sim \mathrm{N}_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$$

Bivariate Normal with Matrix

Denote
$$\boldsymbol{\mu} = (\mu_1, \mu_2)^T,\ \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

then $\boldsymbol{X} = (X_1, X_2)^T \sim \mathrm{N}_2(\boldsymbol{\mu}, \Sigma)$ with pdf
$$f(\boldsymbol{x}) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\}$$

where $\boldsymbol{x} = (x_1, x_2)^T$.

Marginal and Conditional Distribution

Theorem 7. *If $(X,Y) \sim \mathrm{N}_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, then*

1. *The marginal distribution $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$*
2. *The marginal distribution $Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$*
3. *The conditional density of Y given $X = x$ is $Y|X = x \sim \mathrm{N}(\mu_{Y|x}, \sigma_{Y|x}^2)$, where*
$$\mu_{Y|x} = \mathrm{E}(Y|x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$
$$\sigma_{Y|x}^2 = \sigma_Y^2(1 - \rho^2)$$
4. *Given any two constants a and b, then*
$$Z = aX + bY \sim \mathrm{N}(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y)$$

2.2 Representations of Bivariate Normal

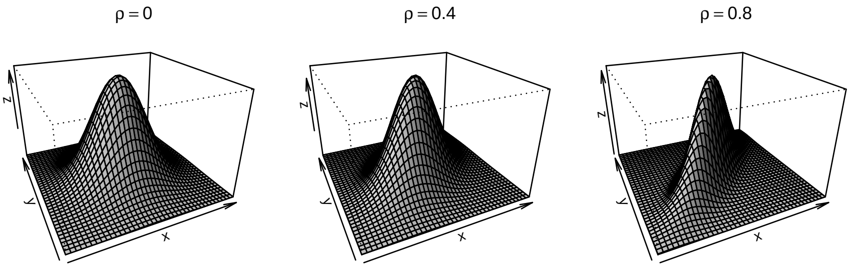
Plot Bivariate Normal: The R base functions

- Goal: Plot the bivariate normal densities with $\mu_X = \mu_Y = 0; \sigma_X = \sigma_Y = 1$ and $\rho = 0, 0.4, 0.8$ respectively
- Use R base function: **persp()**, **contour()**, and **image()**, which draw perspective plots, contour plots, and heat plots

```
> f1 <- function(x, y, p = 0){
+ exp( (x^2 - 2*p*x*y + y^2) / (-2*(1 - p^2)) ) / (2*pi*sqrt(1 - p^2))}
> f2 <- function(x, y, p = 0.4){
+ exp( (x^2 - 2*p*x*y + y^2) / (-2*(1 - p^2)) ) / (2*pi*sqrt(1 - p^2))}
> f3 <- function(x, y, p = 0.8){
+ exp( (x^2 - 2*p*x*y + y^2) / (-2*(1 - p^2)) ) / (2*pi*sqrt(1 - p^2))}
```

persp()

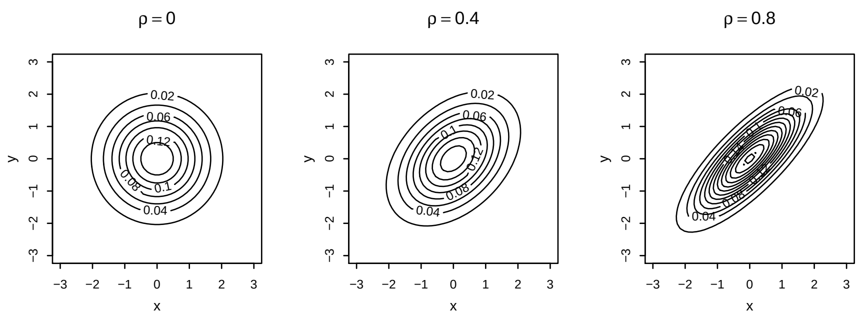
```
> opar <- par(no.readonly = TRUE) # copy of current settings
> par(mfrow = c(1, 3), mar = c(1.1, 1.1, 1.1, 1.1), pty = "s")
> x <- seq(-3, 3, length = 40); y <- x
> persp(x, y, outer(x, y, f1), zlab = "z", main = expression(rho == 0),
+        theta = -25, expand = 0.65, phi = 25, shade = 0.4)
> persp(x, y, outer(x, y, f2), zlab = "z", main = expression(rho == 0.4),
+     theta = -25, expand = 0.65, phi = 25, shade = 0.4)
> persp(x, y, outer(x, y, f3), zlab = "z", main = expression(rho == 0.8),
+     theta = -25, expand = 0.65, phi = 25, shade = 0.4)
```




```
> par(opar)
```

contour()

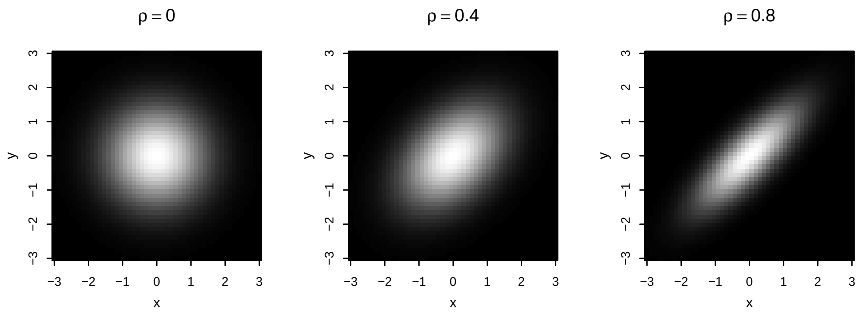
```
> opar <- par(no.readonly = TRUE) # copy of current settings
> par(mfrow = c(1, 3), mar = c(4.1, 4.1, 4.1, 1.1), pty = "s")
> x <- seq(-3, 3, length = 50); y <- x
> contour(x, y, outer(x, y, f1), nlevels = 10, xlab = "x", ylab = "y",
+         main = expression(rho == 0))
> contour(x, y, outer(x, y, f2), nlevels = 10, xlab = "x", ylab = "y",
+         main = expression(rho == 0.4))
> contour(x, y, outer(x, y, f3), nlevels = 10, xlab = "x", ylab = "y",
+         main = expression(rho == 0.8))
```



```
> par(opar)
```

image()

```
> opar <- par(no.readonly = TRUE) # copy of current settings
> par(mfrow = c(1, 3), mar = c(4.1, 4.1, 4.1, 1.1), pty = "s")
> x <- seq(-3, 3, length = 50); y <- x
> image(x, y, outer(x, y, f1), col  = gray((0:100)/100), xlab = "x",
+       ylab = "y", main = expression(rho == 0))
> image(x, y, outer(x, y, f2), col  = gray((0:100)/100), xlab = "x",
+      ylab = "y", main = expression(rho == 0.4))
> image(x, y, outer(x, y, f3), col  = gray((0:100)/100), xlab = "x",
+      ylab = "y", main = expression(rho == 0.8))
```




```
> par(opar)
```

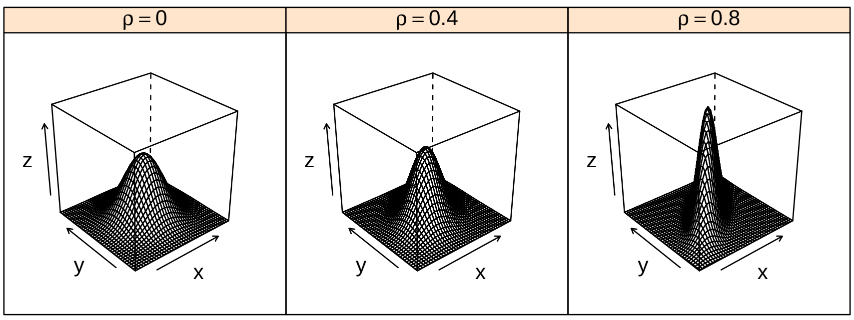
Bivariate Normal Plots: lattice functions

Use the **lattice** package: **wireframe(), contourplot(),** and **levelplot()**
prepare variable data and *expand.grid*

```
> x <- seq(-3, 3, length = 40)
> y <- x
> z1 <- outer(x, y, f1)
> z2 <- outer(x, y, f2)
> z3 <- outer(x, y, f3)
> Grid <- expand.grid(x = x, y = y)
> zp <- c(expression(rho == 0.0), expression(rho == 0.4),
+         expression(rho == 0.8))
```

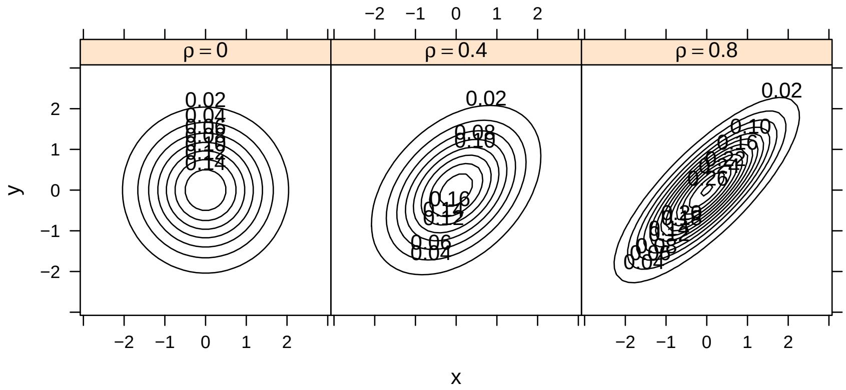
wireframe()

```
> library("lattice")
> wireframe( z1 + z2 + z3 ~ x * y, data = Grid, xlab = "x", ylab = "y",
+ zlab = "z", outer = TRUE, layout = c(3, 1),
+ strip = strip.custom(factor.levels = zp) )
```



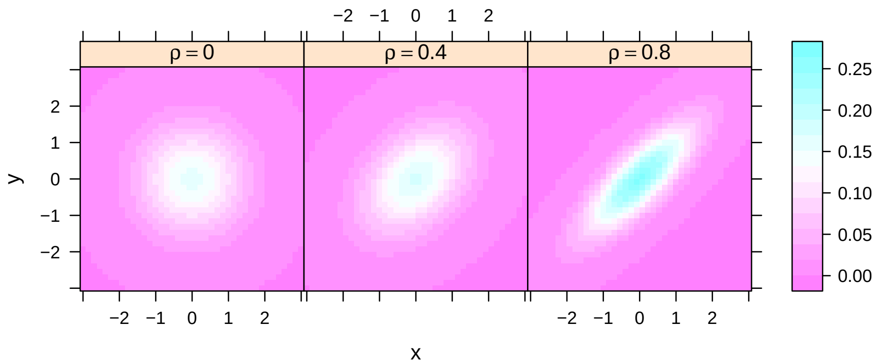
contourplot()

```
> contourplot(z1 + z2 + z3 ~ x * y, data = Grid, xlab = "x", ylab = "y",
+             outer = TRUE, layout = c(3, 1), aspect = "xy",
+             cuts = 11, strip = strip.custom(factor.levels = zp))
```



levelplot()

```
> levelplot(z1 + z2 + z3 ~ x * y, data = Grid, xlab = "x", ylab = "y",
+           outer = TRUE, layout = c(3, 1), aspect = "xy",
+           strip = strip.custom(factor.levels = zp))
```



Bivariate Normal Plots: ggplot2 functions

- ggplot2: **stat_contour()** and **geom_tile()** for contour plot
- **geom_raster()** for heat plot.
- ggplot2 does not currently render three-dimensional graphs

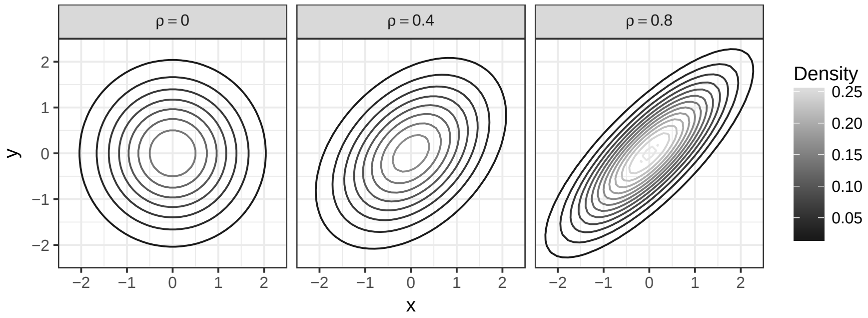
Prepare data for ggplot2:

```
> x <- seq(-3, 3, length = 50)
> y <- x
> z1 <- outer(x, y, f1)
> z2 <- outer(x, y, f2)
> z3 <- outer(x, y, f3)
> Grid <- expand.grid(x = x, y = y)
> DF1 <- data.frame(x = Grid$x, y = Grid$y, z = as.vector(z1))
> DF2 <- data.frame(x = Grid$x, y = Grid$y, z = as.vector(z2))
> DF3 <- data.frame(x = Grid$x, y = Grid$y, z = as.vector(z3))
```

```
> DF1$r = "rho == 0.0"
> DF2$r = "rho == 0.4"
> DF3$r = "rho == 0.8"
> BDF <- rbind(DF1, DF2, DF3)
```

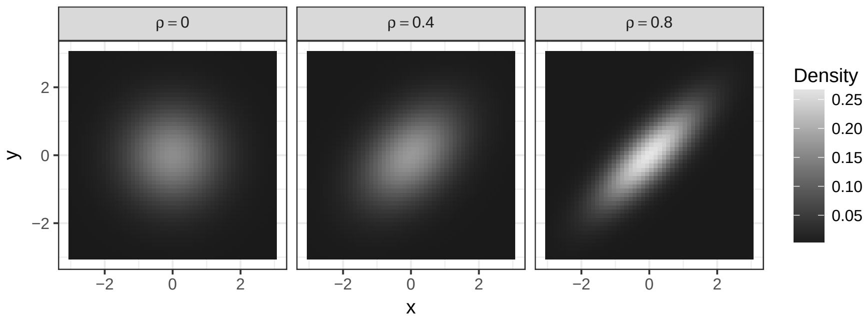
ggplot2: contour

```
> library("ggplot2")
> p <- ggplot(data = BDF, aes(x = x, y = y, z = z))
> p + stat_contour(aes(colour = ..level..)) + theme_bw() +
+     scale_colour_gradient(low = "gray10", high = "gray90") +
+     labs(colour = "Density", x = "x", y = "y") +
+     facet_grid(. ~ r, labeller = label_parsed) +
+     coord_fixed(ratio = 1)
```



ggplot2: heat maps

```
> p <- ggplot(data = BDF, aes(x = x, y = y, fill = z))
> p + geom_raster() + theme_bw() +
+     scale_fill_gradient(low = "gray10", high = "gray90") +
+     labs(fill = "Density", x = "x", y = "y") +
+     facet_grid(. ~ r, labeller = label_parsed) +
+     coord_fixed(ratio = 1)
```



2.3 Multivariate Normal

Multivariate Normal Distribution

If a p-variate random vector $\boldsymbol{X}$ has the joint pdf

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}$$

where the matrix Σ is nonsingular, then we say $\boldsymbol{X}$ has a **Multivariate Normal** distribution, and is denoted by

$$\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$$

The mgf of $\boldsymbol{X}$ is

$$M(\boldsymbol{t}) = \mathrm{E}(e^{\boldsymbol{t}^T\boldsymbol{X}}) = \exp\{\boldsymbol{\mu}^T\boldsymbol{t} + \frac{1}{2}(\boldsymbol{t}^T\Sigma\boldsymbol{t})\}$$

Marginal Distribution

The marginal distributions of a multivariate normal are all multivariate normal.

If

$$\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

then

$$\begin{aligned} \boldsymbol{X}_1 &\sim \mathrm{N}(\boldsymbol{\mu}_1, \Sigma_{11}) \\ \boldsymbol{X}_2 &\sim \mathrm{N}(\boldsymbol{\mu}_2, \Sigma_{22}) \end{aligned}$$

Conditional Distribution

The conditional distributions of a multivariate normal are also multivariate normal.

If

$$\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

then

$$\boldsymbol{X}_2 | \boldsymbol{X}_1 = \boldsymbol{x}_1 \sim \mathrm{N}(\boldsymbol{\mu}_{2.1}, \Sigma_{22.1})$$

where

$$\begin{aligned} \boldsymbol{\mu}_{2.1} &= \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1) \\ \Sigma_{22.1} &= \Sigma_{22} - \Sigma_{2.1}\Sigma_{11}^{-1}\Sigma_{12} \end{aligned}$$

General Linear Transformation

Theorem 8. *If $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ and $\boldsymbol{A}$ is an $n \times p$ real matrix, then the random vector $\boldsymbol{Y} = \boldsymbol{AX}$ is distributed $\boldsymbol{Y} \sim \mathrm{N}_n(\boldsymbol{A\mu}, \boldsymbol{A}\Sigma\boldsymbol{A}^T)$*

Proof: The mgf of $\boldsymbol{Y}$ is

$$\begin{aligned} M_Y(\boldsymbol{t}) &= \mathrm{E}(e^{\boldsymbol{t}^T\boldsymbol{Y}}) = \mathrm{E}(e^{(\boldsymbol{A}^T\boldsymbol{t})\boldsymbol{X}} = M_X(\boldsymbol{A}^T\boldsymbol{t}) \\ &= \exp\left\{\boldsymbol{\mu}^T\boldsymbol{A}^T\boldsymbol{t} + \frac{1}{2}(\boldsymbol{A}^T\boldsymbol{t})^T\Sigma(\boldsymbol{A}^T\boldsymbol{t})\right\} \\ &= \exp\left\{(\boldsymbol{A\mu})^T\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T(\boldsymbol{A}\Sigma\boldsymbol{A}^T)\boldsymbol{t}\right\} \end{aligned}$$

the last expression is the MGF of $\mathrm{N}(\boldsymbol{A\mu}, \boldsymbol{A}\Sigma\boldsymbol{A}^T)$ distribution.

3 Other Multivariate Distributions

Multinomial Distribution

The **multinomial** is a generalization of the binomial distribution.

Definition 9. Suppose there are n independent trials, each trial results in exactly one of fixed finite number k possible outcomes, with probabilities $p_1, ..., p_k$ (so that $p_i \geq 0$ and $\sum_{i=1}^{k} p_i = 1$). Let X_i be the number of times outcome number i is observed over the n trials, then the vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_k)^T$ follows a **multinomial** distribution with parameters n and $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)^T$, we write

$$\boldsymbol{X} \sim \mathrm{Multinom}(n, \boldsymbol{p})$$

The joint pmf

$$P(X_1 = x_1, \ldots, X_k = x_k) = \binom{n}{x_1 x_x \cdots x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

Some Properties

Fact 10.

1. *$\mathrm{E}(\boldsymbol{X}) = n\boldsymbol{p}$, i.e. $\mathrm{E}X_i = np_i (i = 1, 2, \ldots, k)$*

2. *The covarience matrix Σ is symmetric with diagonal entries $\sigma_i^2 = np_i(1 - p_i)$, and off-diagonal entries $\mathrm{Cov}(X_i, X_j) = -np_ip_j$,for $i \neq j$.*

3. *The correlation coefficients*

$$\mathrm{Corr}(X_i, X_j) = \frac{-np_ip_j}{\sqrt{np_i(1 - p_i) \cdot np_j(1 - p_j)}} = -\sqrt{\frac{p_ip_j}{q_iq_j}}$$

4. *The marginal distribution of $\boldsymbol{X}_{(k-1)} = (X_1, X_2, \ldots, X_{k-1})^T$ is Multinom$(n, \boldsymbol{p}_{(k-1)})$ with $\boldsymbol{p}_{(k-1)} = (p_1, \ldots, p_{k-2}, p_{k-1} + p_k)^T$. in particular, $X_i \sim$ Bin(n, p_i).*