Multivariate Random Variables

Max Turgeon

STAT 4690–Applied Multivariate Analysis

Joint distributions

- Let X and Y be two random variables.
- The *joint distribution function* of X and Y is

$$F(x, y) = P(X \leq x, Y \leq y).$$

- More generally, let $Y_1, \ldots, Y_p$ be p random variables. Their *joint distribution function* is

$$F(y_1, \ldots, y_p) = P(Y_1 \leq y_1, \ldots, Y_p \leq y_p).$$

Joint densities

- If F is absolutely continuous almost everywhere, there exists a function f called the *density* such that

$$F(y_1, \ldots, y_p) = \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_p} f(u_1, \ldots, u_p) du_1 \cdots du_p.$$

- The *joint moments* are defined as follows:

$$E(Y_1^{n_1} \cdots Y_p^{n_p}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u_1^{n_1} \cdots u_p^{n_p} f(u_1, \ldots, u_p) du_1 \cdots du_p.$$

- **Exercise**: Show that this is consistent with the univariate definition of $E(Y_1^{n_1})$, i.e. $n_2 = \cdots = n_p = 0$.

Marginal distributions i

- From the joint distribution function, we can recover the *marginal distributions*:

$$F_i(x) = \lim_{\substack{y_j \to \infty \\ j \neq i}} F(y_1, \ldots, y_n).$$

- More generally, we can find the joint distribution of a subset of variables by sending the other ones to infinity:

$$F(y_1, \ldots, y_r) = \lim_{\substack{y_j \to \infty \\ j \neq r}} F(y_1, \ldots, y_n), \quad r < p.$$

Marginal distributions ii

- Similarly, from the joint density function, we can recover the *marginal densities*:

$$f_i(x) = \int_{-\infty}^{\infty} f(u_1, \ldots, u_p) du_1 \cdots \widehat{du_i} \cdots du_p.$$

- In other words, we are integrating *out* the other variables.

Conditional distributions

- Let f_1, f_2 be the densities of random variables Y_1, Y_2, respectively. Let f be the joint density.
- The *conditional density* of Y_1 given Y_2 is defined as

$$f(y_1|y_2) := \frac{f(y_1, y_2)}{f_2(y_2)},$$

  whenever $f_2(y_2) \neq 0$ (otherwise it is equal to zero).
- Similarly, we can define the conditional density in $p > 2$ variables, and we can also define a conditional density for $Y_1, \ldots, Y_r$ given $Y_{r+1}, \ldots, Y_p$.

Expectations

- Let $\mathbf{Y} = (Y_1, \ldots, Y_p)$ be a random vector.
- Its *expectation* is defined entry-wise:

$$E(\mathbf{Y}) = (E(Y_1), \ldots, E(Y_p)).$$

- **Observation**: The dependence structure has no impact on the expectation.

Covariance and Correlation i

- The multivariate generalization of the variance is the *covariance matrix*. It is defined as

$$\mathrm{Cov}(\mathbf{Y}) = E\left((\mathbf{Y} - \mu)(\mathbf{Y} - \mu)^T\right),$$

  where $\mu = E(\mathbf{Y})$.
- **Exercise**: The (i, j)-th entry of $\mathrm{Cov}(\mathbf{Y})$ is equal to

$$\mathrm{Cov}(Y_i, Y_j).$$

Covariance and Correlation ii

- Recall that we obtain the correlation from the covariance by dividing by the square root of the variances.
- Let V be the diagonal matrix whose i-th entry is $\mathrm{Var}(Y_i)$.
    - In other words, V and $\mathrm{Cov}(\mathbf{Y})$ have the same diagonal.
- Then we define the *correlation matrix* as follows:

$$\mathrm{Corr}(\mathbf{Y}) = V^{-1/2}\mathrm{Cov}(\mathbf{Y})V^{-1/2}.$$

- **Exercise**: The (i, j)-th entry of $\mathrm{Corr}(\mathbf{Y})$ is equal to

$$\mathrm{Corr}(Y_i, Y_j).$$

Example i

- Assume that

$$\mathrm{Cov}(\mathbf{Y}) = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{pmatrix}.$$

- Then we know that

$$V = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 25 \end{pmatrix}.$$

Example ii

- Therefore, we can write

$$V^{-1/2} = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.33 & 0 \\ 0 & 0 & 0.2 \end{pmatrix}.$$

- We can now compute the correlation matrix:

Example iii

$$\mathrm{Corr}(\mathbf{Y}) = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.33 & 0 \\ 0 & 0 & 0.2 \end{pmatrix} \begin{pmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{pmatrix} \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.33 & 0 \\ 0 & 0 & 0.2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0.17 & 0.2 \\ 0.17 & 1 & -0.2 \\ 0.2 & -0.2 & 1 \end{pmatrix}.$$

Measures of Overall Variability

- In the univariate case, the variance is a scalar measure of spread.
- In the multivariate case, the *covariance* is a matrix.
- No easy way to compare two distributions.
- For this reason, we have other notions of overall variability:

1. **Generalized Variance**: This is defined as the determinant of the covariance matrix.

$$GV(\mathbf{Y}) = \det(\mathrm{Cov}(\mathbf{Y})).$$

2. **Total Variance**: This is defined as the trace of the covariance matrix.

$$TV(\mathbf{Y}) = \mathrm{tr}(\mathrm{Cov}(\mathbf{Y})).$$

Examples i

```r
A <- matrix(c(5, 4, 4, 5), ncol = 2)

results <- eigen(A, symmetric = TRUE,
                 only.values = TRUE)

# Generalized variance
prod(results$values)
```

```
## [1] 9
```

Examples ii

```r
# Total variance
sum(results$values)
```

```
## [1] 10
```

```r
# Compare this with the following
B <- matrix(c(5, -4, -4, 5), ncol = 2)

# Generalized variance
# GV(A) = 9
det(B)
```

Examples iii

```
## [1] 9
```

```r
# Total variance
# TV(A) = 10
sum(diag(B))
```

```
## [1] 10
```

Measures of Overall Variability (cont'd)

- As we can see, we do lose some information:
  - In matrix B, we saw that the two variables are negatively correlated, and yet we get the same values
- But GV captures *some* information on dependence that TV does not.
  - Compare the following covariance matrices:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

- *Interpretation*: A small value of the sampled Generalized Variance indicates either small scatter in data points or multicollinearity.

Geometric Interlude i

- A random vector $\mathbf{Y}$ with positive definite covariance matrix Σ can be used to define a distance function on $\mathbb{R}^p$:

$$d(x, y) = \sqrt{(x - y)^T \Sigma^{-1} (x - y)}.$$

- This is called the *Mahalanobis distance* induced by Σ.
- **Exercise**: This indeed satisfies the definition of a distance:
    1. $d(x, y) = d(y, x)$
    2. $d(x, y) \geq 0$ and $d(x, x) = 0 \Leftrightarrow x = 0$
    3. $d(x, z) \leq d(x, y) + d(y, z)$

Geometric Interlude ii

- Using this distance, we can construct *hyper-ellipsoids* in $\mathbb{R}^p$ as the set of all points x such that

$$d(x, 0) = 1.$$

- Equivalently:

$$x^T \Sigma^{-1} x = 1.$$

- Since Σ^{-1} is symmetric, we can use the spectral decomposition to rewrite it as:

$$\Sigma^{-1} = \sum_{i=1}^{p} \lambda_i^{-1} v_i v_i^T,$$

where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of Σ.

Geometric Interlude iii

- We thus get a new parametrization if the hyper-ellipsoid:

$$\sum_{i=1}^{p} \left(\frac{v_i^T x}{\sqrt{\lambda_i}} \right) = 1.$$

- **Theorem**: The volume of this hyper-ellipsoid is equal to

$$\frac{2\pi^{p/2}}{p\Gamma(p/2)} \sqrt{\lambda_1 \cdots \lambda_p}.$$

- In other words, the Generalized Variance is proportional to the square of the volume of the hyper-ellipsoid defined by the covariance matrix.
    - *Note*: the square root of the determinant of a matrix (if it exists) is sometimes called the *Pfaffian*.

Example i

```r
Sigma <- matrix(c(1, 0.5, 0.5, 1), ncol = 2)

# First create a circle
theta_vect <- seq(0, 2*pi, length.out = 100)
circle <- cbind(cos(theta_vect), sin(theta_vect))
# Then turn into ellipse
ellipse <- circle %*% Sigma
```

Example ii

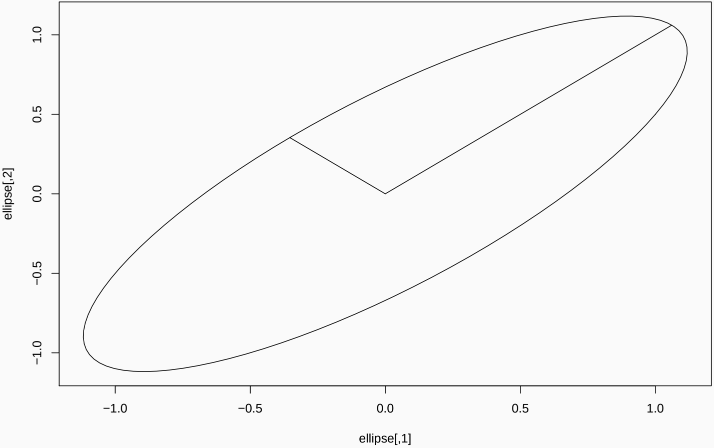
```r
# Principal axes
result <- eigen(Sigma, symmetric = TRUE)

first <- result$values[1]*result$vectors[,1]
second <- result$values[2]*result$vectors[,2]
```

Example iii

```r
# Plot results
plot(ellipse, type = 'l')
lines(x = c(0, first[1]),
      y = c(0, first[2]))
lines(x = c(0, second[1]),
      y = c(0, second[2]))
```

Example iv



Example (cont'd) i

```r
# Generalized Variance
det(Sigma)
```

```
## [1] 0.75
```

```r
# Predicted volume of the ellipse above
pi/(gamma(1))*sqrt(det(Sigma))
```

```
## [1] 2.720699
```

Example (cont'd) ii

```r
# How can we estimate the area?
# Monte Carlo simulation!
Sigma_inv <- solve(Sigma)

x_1 <- runif(1000, min = min(ellipse[,1]),
             max = max(ellipse[,1]))
x_2 <- runif(1000, min = min(ellipse[,2]),
             max = max(ellipse[,2]))

X <- cbind(x_1, x_2)
distances <- apply(X, 1, function(row) {
  sqrt(t(row) %*% Sigma_inv %*% row)
  })
```

Example (cont'd) iii

```r
# Estimate
length_x <- diff(range(ellipse[,1]))
length_y <- diff(range(ellipse[,2]))
area_rect <- length_x * length_y

area_rect * mean(distances <= 1)
```

```
## [1] 2.794066
```

Statistical Independence

- The variables $Y_1, \ldots, Y_p$ are said to be *mutually independent* if

$$F(y_1, \ldots, y_p) = F(y_1) \cdots F(y_p).$$

- If $Y_1, \ldots, Y_p$ admit a joint density f (with marginal densities $f_1, \ldots, f_p$), and equivalent condition is

$$f(y_1, \ldots, y_p) = f(y_1) \cdots f(y_p).$$

- **Important property**: If $Y_1, \ldots, Y_p$ are mutually independent, then their joint moments factor:

$$E(Y_1^{n_1} \cdots Y_p^{n_p}) = E(Y_1^{n_1}) \cdots E(Y_p^{n_p}).$$

Linear Combination of Random Variables

- Let $\mathbf{Y} = (Y_1, \ldots, Y_p)$ be a random vector. Let $\mathbf{A}$ be a $q \times p$ matrix, and let $b \in \mathbb{R}^q$.
- Then the random vector $\mathbf{X} := \mathbf{A}\mathbf{Y} + b$ has the following properties:
  - **Expectation**: $E(\mathbf{X}) = \mathbf{A}E(\mathbf{Y}) + b$;
  - **Covariance**: $\mathrm{Cov}(\mathbf{X}) = \mathbf{A}\mathrm{Cov}(\mathbf{Y})\mathbf{A}^T$

Transformation of Random Variables

- More generally, let $h : \mathbb{R}^p \to \mathbb{R}^p$ be a one-to-one function with inverse $h^{-1} = (h_1^{-1}, \ldots, h_p^{-1})$. Define $\mathbf{X} = h(\mathbf{Y})$.
- Let J be the *Jacobian matrix* of h^{-1}:

$$\begin{pmatrix} \frac{\partial h_1^{-1}}{\partial y_1} & \cdots & \frac{\partial h_1^{-1}}{\partial y_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_p^{-1}}{\partial y_1} & \cdots & \frac{\partial h_p^{-1}}{\partial y_p} \end{pmatrix}.$$

- Then the density of $\mathbf{X}$ is given by

$$g(x_1, \ldots, x_p) = f(h_1^{-1}(y_1), \ldots, h_p^{-1}(y_p))|\det(J)|.$$

- *This result is very useful for computing the density of transformations of normal random variables.*

Properties of Sample Statistics i

- Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be a random sample from a p-dimensional distribution with mean μ and covariance matrix Σ.
- **Sample mean**: We define the sample mean $\bar{\mathbf{Y}}$ as follows:

$$\bar{\mathbf{Y}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Y}_i.$$

- *Properties*:
  - $E(\bar{\mathbf{Y}}) = \mu$ (i.e. $\bar{\mathbf{Y}}$ is an unbiased estimator of μ);
  - $\mathrm{Cov}(\bar{\mathbf{Y}}) = \frac{1}{n}\Sigma$.

Properties of Sample Statistics ii

- **Sample covariance**: We define the sample covariance $\mathbf{S}$ as follows:

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^T.$$

- *Properties*:
  - $E(\mathbf{S}) = \frac{n-1}{n}\Sigma$ (i.e. $\mathbf{S}$ is a biased estimator of Σ);
  - If we define $\tilde{\mathbf{S}}$ with n instead of $n-1$ in the denominator above, then $E(\tilde{\mathbf{S}}) = \Sigma$ (i.e. $\tilde{\mathbf{S}}$ is an unbiased estimator of Σ).