MAST90104 - Lecture 3

Weichang Yu

Room 108, Old Geology South Bldg
School of Mathematics and Statistics, University of Melbourne

5 Aug, 2024

From random variables to random vectors

If Y is a random variable, then we can express its moments as $\mathbb{E}(Y)$ and $Var(Y)$. For example, if $Y \sim N(\mu, \sigma^2)$, then $\mathbb{E}(Y) = \mu$ and $Var(Y) = \sigma^2$.

A *random vector* is a vector of random variables.

$$\text{If} \quad \mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \text{then} \quad \mathbb{E}[\mathbf{y}] = \begin{bmatrix} \mathbb{E}[Y_1] \\ \mathbb{E}[Y_2] \\ \vdots \\ \mathbb{E}[Y_n] \end{bmatrix}.$$

From random variables to random vectors

- If $\mathbf{a}$ is a vector of constants, then $\mathbb{E}[\mathbf{a}] = \mathbf{a}$.

- If $\mathbf{a}$ is a vector of constants and $\mathbf{y}$ is a random vector, then $\mathbb{E}[\mathbf{a}^T\mathbf{y}] = \mathbf{a}^T\mathbb{E}[\mathbf{y}]$.

- If $\mathbf{A}$ is a matrix of constants, then $\mathbb{E}[\mathbf{A}\mathbf{y}] = \mathbf{A}\mathbb{E}[\mathbf{y}]$.

Expectation properties

Example. Let

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

and assume that $\mathbb{E}[Y_1] = 10$ and $\mathbb{E}[Y_2] = 20$.
Then

$$\mathbf{A}\mathbb{E}[\mathbf{y}] = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 10 \\ 20 \end{bmatrix} = \begin{bmatrix} 80 \\ 90 \end{bmatrix}.$$

Expectation properties

On the other hand,

$$
\begin{aligned}
\mathbb{E}[\mathbf{Ay}] &= \mathbb{E}\left[\begin{array}{c} 2Y_1 + 3Y_2 \\ Y_1 + 4Y_2 \end{array}\right] \\
&= \left[\begin{array}{c} \mathbb{E}[2Y_1 + 3Y_2] \\ \mathbb{E}[Y_1 + 4Y_2] \end{array}\right] \\
&= \left[\begin{array}{c} 2\mathbb{E}[Y_1] + 3\mathbb{E}[Y_2] \\ \mathbb{E}[Y_1] + 4\mathbb{E}[Y_2] \end{array}\right] \\
&= \left[\begin{array}{c} 80 \\ 90 \end{array}\right] = \mathbf{A}\mathbb{E}[\mathbf{y}].
\end{aligned}
$$

Variance

Defining the variance of a random vector is slightly trickier. We want to not just include the variance of the variables themselves, but also the covariance between pairs of variables.

Recall that the variance of a random variable Y with mean μ is defined to be $E[(Y - \mu)^2]$.
We define the variance (or *covariance matrix*) of the random vector $\mathbf{y}$ to be

$$Var(\mathbf{y}) = E[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^T]$$

where $\boldsymbol{\mu} = E[\mathbf{y}]$.

Variance

The diagonal elements of the covariance matrix are the variances of the elements of $\mathbf{y}$:

$$[Var(\mathbf{y})]_{ii} = Var(Y_i), \qquad i = 1, 2, \ldots, k.$$

The off-diagonal elements are the covariances of the elements:

$$[Var(\mathbf{y})]_{ij} = Cov(Y_i, Y_j) = E[(Y_i - \mu_i)(Y_j - \mu_j)],$$

where $\mu_i = \mathbb{E}(Y_i)$ and $\mu_j = \mathbb{E}(Y_j)$.

This means that all covariance matrices are symmetric.

Variance: The covariance matrix

$$Var(\mathbf{y}) = \begin{pmatrix} Var(Y_1) & Cov(Y_1, Y_2) & \dots & Cov(Y_1, Y_n) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(Y_n, Y_1) & Cov(Y_n, Y_2) & \dots & Var(Y_n) \end{pmatrix}$$

Variance properties

Example. Let

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

be a random vector with $Var(Y_i) = \sigma^2$ for all i, and the elements of $\mathbf{y}$ are independent.

Then the covariance matrix of $\mathbf{y}$ is

$$Var(\mathbf{y}) = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}.$$

Variance properties

Suppose that $\mathbf{y}$ is a random vector with $\mathrm{V}ar(\mathbf{y}) = \mathbf{V}$. Then:

- If $\mathbf{a}$ is a vector of constants, then $\mathrm{V}ar(\mathbf{a}^T\mathbf{y}) = \mathbf{a}^T\mathbf{V}\mathbf{a}$.
- If $\mathbf{A}$ is a matrix of constants, then $\mathrm{V}ar(\mathbf{A}\mathbf{y}) = \mathbf{A}\mathbf{V}\mathbf{A}^T$.
- $\mathbf{V}$ is positive semidefinite.

Example. Continuing the expectation example

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

and assume that $E[Y_1] = 10$, $E[Y_2] = 20$ and

$$\mathbf{V} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

.

Variance properties - example

Example. Then

$$
\begin{aligned}
Var[\mathbf{Ay}] &= \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \\
&= \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \\
&= \begin{bmatrix} 25 & 25 \\ 25 & 25 \end{bmatrix}.
\end{aligned}
$$

Variance properties - example

Example. Assume that $\mathbf{X}$ is a matrix of full rank (with more rows than columns), which implies that $\mathbf{X}^T\mathbf{X}$ is nonsingular. Let

$$\mathbf{Z} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{A}\mathbf{y},$$

where $\mathrm{V}ar(\mathbf{y}) = \sigma^2\mathbf{I}$.

Then

$$\begin{aligned}
\mathrm{V}ar(\mathbf{Z}) &= \mathbf{A}\mathrm{V}ar(\mathbf{y})\mathbf{A}^T = [(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]\sigma^2\mathbf{I}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]^T \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}^T)^T[(\mathbf{X}^T\mathbf{X})^{-1}]^T\sigma^2 \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}[(\mathbf{X}^T\mathbf{X})^T]^{-1}\sigma^2 \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}.
\end{aligned}$$

We will be using this quite a bit later on!

Matrix square root

The square root of a matrix $\mathbf{A}$ is a matrix $\mathbf{B}$ such that $\mathbf{B}^2 = \mathbf{A}$. In general the square root is not unique.

If $\mathbf{A}$ is symmetric and positive semidefinite, there is a unique symmetric positive semidefinite square root, called the principal root, denoted $\mathbf{A}^{1/2}$.

Suppose that $\mathbf{P}$ diagonalises $\mathbf{A}$, that is $\mathbf{P}^T\mathbf{A}\mathbf{P} = \mathbf{\Lambda}$. Then

$$\begin{aligned}
\mathbf{A} &= \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T \\
&= \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{P}^T \\
&= \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}^T\mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}^T \\
\mathbf{A}^{1/2} &= \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}^T.
\end{aligned}$$

Multivariate normal distribution

Definition 3.1

Let $\mathbf{Z}$ be a $k \times 1$ vector of independent standard normal random variables, $\mathbf{A}$ an $n \times k$ matrix of constants, and $\boldsymbol{\mu}$ an $n \times 1$ vector. We say that

$$\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$$

has (an n-dimensional) multivariate normal distribution, with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{AA}^T$.

We write $\mathbf{X} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ or just $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathbf{AA}^T$.

Multivariate normal distribution

For any $\boldsymbol{\mu}$ and any symmetric positive semidefinite matrix $\boldsymbol{\Sigma}$, let $\mathbf{Z}$ be a vector of independent standard normals. Then

$$\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

If $\mathbf{X} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}$ is $k \times k$ positive definite, then $\mathbf{X}$ has the density

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}.$$

Compare this with the univariate normal density

$$f(x) = \frac{1}{(2\pi)^{1/2}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}.$$

Multivariate normal distribution

Any linear combination of multivariate normals results in another multivariate normal: if $\mathbf{X} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is $k \times 1$, $\mathbf{A}$ is $n \times k$, and $\mathbf{b}$ is $n \times 1$, then

$$\mathbf{y} = \mathbf{AX} + \mathbf{b} \sim MVN(A\boldsymbol{\mu} + \mathbf{b}, \mathbf{A\Sigma A}^T).$$

Multivariate normal distribution

If $\mathbf{Z} = (Z_1, Z_2)^T$ is multivariate normal, then Z_1 and Z_2 are independent if and only if they are uncorrelated.

In general, if Z_1 and Z_2 are normal random variables, $\mathbf{Z} = (Z_1, Z_2)^T$ does not have to be multivariate normal. Moreover, Z_1 and Z_2 can be uncorrelated but not independent.

For example, suppose that $Z_1 \sim N(0, 1)$, $U \sim U(-1, 1)$, and $Z_2 = Z_1 \text{sign}(U) \sim N(0, 1)$, but $\mathbf{Z} = (Z_1, Z_2)^T$ is *not* multivariate normal (consider $Z_1 + Z_2$). Moreover Z_1 and Z_2 are uncorrelated, but clearly dependent.

Example

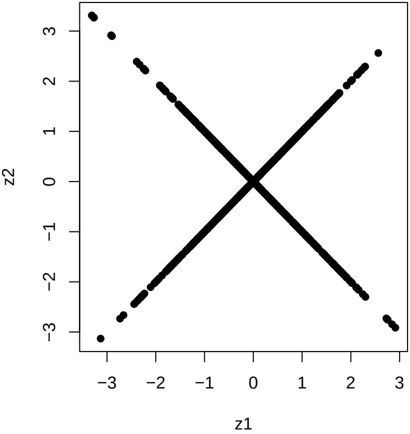

Figure: Joint range of Z_1 and Z_2

Generalisation to all distributions

Z_1 and Z_2 are independent random variables $\Rightarrow$ Z_1 and Z_2 are uncorrelated.

But the converse may not be true!

Multivariate normals in R

To generate a sample of size 100 with distribution

$$MVN\left(\begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}\right)$$

```
library(MASS)
a <- matrix(c(3, 1), 2, 1)
V <- matrix(c(1, .8, .8, 1), 2, 2)
y <- mvrnorm(100, mu = a, Sigma = V)
plot(y[,1], y[,2])
```

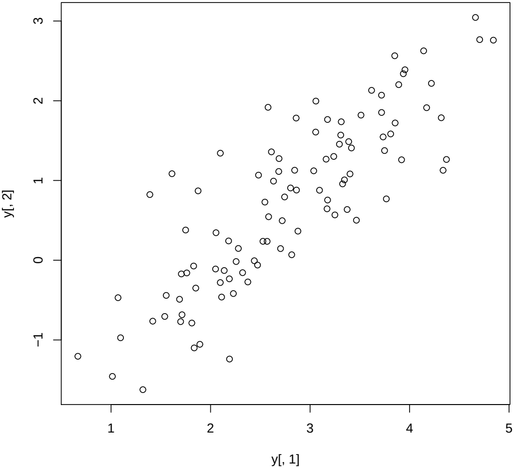

Figure: 100 bivariate normal random vectors using `mvrnorm`

Multivariate normals in R

Alternatively, note that if $\mathbf{Z}$ is a vector of independent standard normal then we can generate $\mathbf{y} \sim MVN(\boldsymbol{\mu}, \mathbf{V})$ by noting that $\mathbf{y} = \boldsymbol{\mu} + \mathbf{V}^{1/2}\mathbf{Z}$

We can find $\mathbf{V}^{1/2}$ from its eigenvalues. Let $\mathbf{P}$ be a matrix whose columns are eigenvectors of $\mathbf{V}$ and $\boldsymbol{\Lambda}$ is a diagonal matrix where the diagonal elements are the corresponding eigenvalues. From module 2 we know that $\mathbf{P}$ diagonalise $\mathbf{V}$, then

$$\mathbf{V}^{1/2} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P}^T$$

```
P <- eigen(V)$vectors
sqrtV <- P %*% diag(sqrt(eigen(V)$values)) %*% t(P)
z <- matrix(rnorm(200), 2, 100)
y_new <- sqrtV %*% z + rep(a, 100)
plot(y[,1], y[,2])
points(y_new[1,], y_new[2,], col = "red")
```

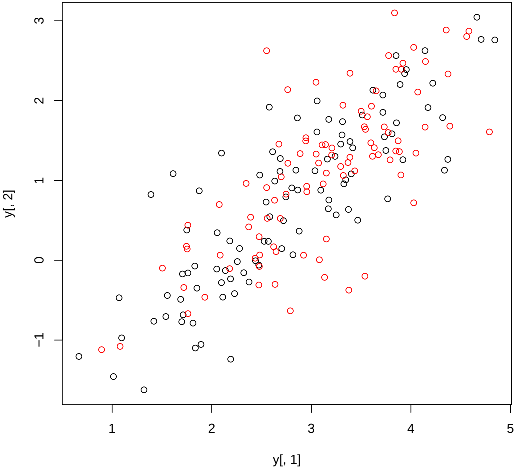

Figure: 100 extra with same distribution in red using definition and `rnorm`

Random quadratic forms

We have seen that a matrix induces a quadratic form (multivariate function).

What happens when the variables in a quadratic form are random variables?

The form becomes a scalar function of random variables, so it is itself a random variable.

Random quadratic forms will appear regularly in our theory of linear models - the sample variance for a random sample is an example. To analyse our models, their distributions are needed.

Random quadratic forms

Theorem 3.2

Let $\mathbf{y}$ *be a random vector with* $E[\mathbf{y}] = \boldsymbol{\mu}$ *and* $Var(\mathbf{y}) = \mathbf{V}$, *and let* $\mathbf{A}$ *be a matrix of constants. Then*

$$E[\mathbf{y}^T\mathbf{A}\mathbf{y}] = tr(\mathbf{A}\mathbf{V}) + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}.$$

Remark. The proof requires the following straightforward results

- For any random $\mathbf{X}$ (can be variable, vector, or matrix), we have $\mathbb{E}\{tr(\mathbf{X})\ \} = tr\{\mathbb{E}(\mathbf{X})\}$.
- For any appropriately-sized matrices of constants $\mathbf{A}$, $\mathbf{B}$, and random matrix $\mathbf{X}$, we have $\mathbb{E}(\mathbf{X}\mathbf{A}) = \mathbb{E}(\mathbf{X})\mathbf{A}$ and $\mathbb{E}(\mathbf{B}\mathbf{X}) = \mathbf{B}\mathbb{E}(\mathbf{X})$
- If c is a scalar, then $tr(c) = c$.

Proof.

$$\begin{aligned}\mathbb{E}\{\mathbf{y}^T\mathbf{A}\mathbf{y}\} &= \mathbb{E}\{tr(\mathbf{y}^T\mathbf{A}\mathbf{y})\} \\ &= \mathbb{E}\{tr(\mathbf{y}\mathbf{y}^T\mathbf{A})\} \\ &= tr\{\mathbb{E}(\mathbf{y}\mathbf{y}^T\mathbf{A})\} \\ &= tr\{\mathbb{E}(\mathbf{y}\mathbf{y}^T)\mathbf{A}\}\end{aligned}$$

Remember that $Var(\mathbf{y}) = \mathbf{V}$. Since

$$\begin{aligned}Var(\mathbf{y}) &= \mathbb{E}\{(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^T\} \\ &= \mathbb{E}(\mathbf{y}\mathbf{y}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T \\ &= \mathbb{E}(\mathbf{y}\mathbf{y}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{E}\{\mathbf{y}^T\mathbf{A}\mathbf{y}\} &= tr[\{\mathbf{V} + \boldsymbol{\mu}\boldsymbol{\mu}^T\}\mathbf{A}] \\ &= tr[\mathbf{V}\mathbf{A} + \boldsymbol{\mu}\boldsymbol{\mu}^T\mathbf{A}] \\ &= tr(\mathbf{A}\mathbf{V}) + tr(\boldsymbol{\mu}\boldsymbol{\mu}^T\mathbf{A}) \\ &= tr(\mathbf{A}\mathbf{V}) + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}\end{aligned}$$

Example. Let $\mathbf{y}$ be a 2×1 random vector with

$$\mathbb{E}(\mathbf{y}) = \boldsymbol{\mu} \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad Var(\mathbf{y}) = \mathbf{V} = \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix}.$$

Let

$$A = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}.$$

Consider the quadratic form

$$\mathbf{y}^T A \mathbf{y} = 4y_1^2 + 2y_1 y_2 + 2y_2^2.$$

Random quadratic forms

From Theorem 3.2,

$$
\begin{aligned}
E[\mathbf{y}^T A \mathbf{y}] &= tr(AV) + \boldsymbol{\mu}^T A \boldsymbol{\mu} \\
&= tr\left(\begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix} \right) + \begin{bmatrix} 1 & 3 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \\
&= tr\left(\begin{bmatrix} 9 & 9 \\ 4 & 11 \end{bmatrix} \right) + \begin{bmatrix} 1 & 3 \end{bmatrix} \begin{bmatrix} 7 \\ 7 \end{bmatrix} \\
&= 9 + 11 + 7 + 21 = 48.
\end{aligned}
$$

Random quadratic forms

```
mu <- c(1,3)
V <- matrix(c(2,1,1,5),2,2)
y <- t(mvrnorm(5000, mu = mu, Sigma = V))
A <- matrix(c(4,1,1,2),2,2)
sum(diag(A%*%V)) + t(mu)%*%A%*%mu

##      [,1]
## [1,]   48

quadform <- function(y, A) t(y) %*% A %*% y
mean(apply(y, 2, quadform, A = A))

## [1] 48.17805
```

Recap: χ^2 distribution

Definition 3.3

Let $\mathbf{y} \sim MVN(\mathbf{0}, I)$ be a $k \times 1$ random vector. Then

$$X = \mathbf{y}^T \mathbf{y} = \sum_{i=1}^{k} y_i^2$$

has a χ^2 *distribution* with k degrees of freedom and we write $X \sim \chi^2_k$.

Recall that $\mathbb{E}(X) = k$ and $Var(X) = 2k$.

Noncentral χ^2 distribution

Definition 3.4

Let $\mathbf{y} \sim MVN(\boldsymbol{\mu}, I)$ be a $k \times 1$ random vector. Then

$$X = \mathbf{y}^T\mathbf{y} = \sum_{i=1}^{k} y_i^2$$

has a *noncentral* χ^2 *distribution* with k degrees of freedom and noncentrality parameter $\lambda = \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\mu}$. We write $X \sim \chi^2_{k,\lambda}$.

In other words, X is the sum of the squares of k independent, normally distributed random variables with means μ_i and variances 1

Warning: some authors (including R!) define non-centrality parameter as 2λ.

Note that the distribution of X depends on $\boldsymbol{\mu}$ only through λ.

Noncentral χ^2 distribution

Example. Let

$$\mathbf{y} \sim MVN\left(\begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix}, I_3 \right).$$

Then $y_1^2 + y_2^2 + y_3^2$ has a noncentral χ^2 distribution with 3 degrees of freedom and noncentrality parameter

$$\lambda = \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\mu} = \frac{1}{2}\begin{bmatrix} 4 & 2 & -2 \end{bmatrix}\begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} = 12.$$

Noncentral χ^2 distribution

Suppose $\mathbf{y} \sim MVN(\boldsymbol{\mu}, I_k)$ and $x = \mathbf{y}^T\mathbf{y} \sim \chi^2_{k,\lambda}$. Then from Theorem 3.2,

$$E[X] = tr(I_k) + \boldsymbol{\mu}^T\boldsymbol{\mu} = k + 2\lambda.$$

The noncentrality parameter $\lambda = \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\mu}$ is zero if and only if $\boldsymbol{\mu} = \mathbf{0}$, in which case x is just the sum of squares of k independent standard normals.

In this case, we already know that X has an ordinary (central) χ^2 distribution with k degrees of freedom.

Noncentral χ^2 distribution

The density of the noncentral $X \sim \chi^2$ distribution is complicated; one expression is

$$f(x; k, \lambda) = \sum_{i=0}^{\infty} \frac{e^{-\lambda} \lambda^i}{i!} g(x; k + 2i)$$

where $g(\cdot; k)$ is the density of the χ^2 distribution

$$g(x; k) = \frac{1}{2^{k/2} \Gamma\left(\frac{k}{2}\right)} x^{k/2-1} e^{-x/2}.$$

The density is a weighted sum of χ^2 densities whose weights are Poisson probabilities.

Some densities are shown in Figure 4.

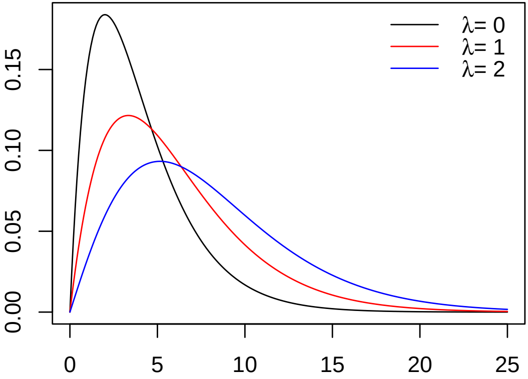

Figure: Densities of chisquare with 4 df and $\lambda = 0, 1, 2$

```r
y <- t(mvrnorm(100, mu = c(1,2,1,0), Sigma = diag(4)))
hist(apply(y,2,quadform,A=diag(4)), freq=FALSE)
curve(dchisq(x,4,6), add=TRUE, col="red")
```

Histogram and density shown in Figure 5.

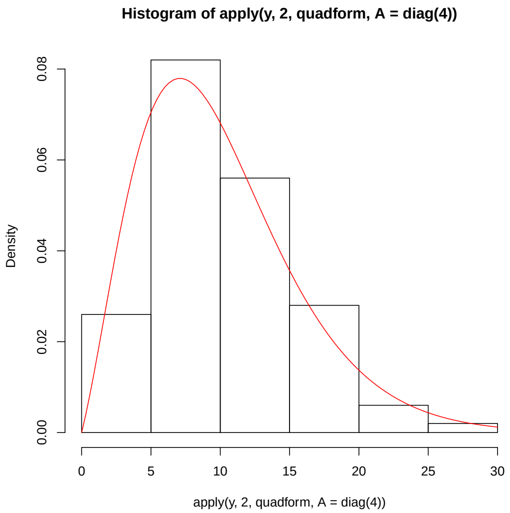

Figure: 100 Chisq rvs with df $= 4$, $\lambda = 3$

Adding noncentral χ^2s

Theorem 3.5

Let $X^2_{k_1,\lambda_1}, \ldots, X^2_{k_n,\lambda_n}$ be a collection of n independent noncentral χ^2 random variables, with $X^2_{k_i,\lambda_i} \sim \chi^2_{k_i,\lambda_i}$. Then

$$\sum_{i=1}^{n} X^2_{k_i,\lambda_i}$$

has a noncentral χ^2 distribution with $\sum_{i=1}^{n} k_i$ degrees of freedom and noncentrality parameter $\sum_{i=1}^{n} \lambda_i$.

If we set $\lambda_i = 0$ for all i, we get the result from MAST90105 that the sum of independent χ^2 variables is another χ^2 variable whose degrees of freedom is the sum of the degrees of freedom of the components.

Proof of Theorem 2.5:
Consider a sequence of independent random vectors $\{\mathbf{y}_i\}_{i=1}^n$, where $\mathbf{y}_i \sim MVN(\boldsymbol{\nu}_i, \mathbf{I}_{k_i})$, i.e., $\mathbf{y}_i$ and $\mathbf{y}_{i'}$ are independent for all $i \neq i'$ and $\|\boldsymbol{\nu}_i\|^2 = 2\lambda_i$. $\mathbf{I}_{k_i}$ denotes an identity matrix of order k_i.

Then, we can express $X^2_{k_i,\lambda_i} = \mathbf{y}_i^T \mathbf{y}_i$, where $\lambda_i = \frac{1}{2}\boldsymbol{\nu}_i^T \boldsymbol{\nu}_i$.

Put all the random vectors into a single column

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix} \sim MVN(\boldsymbol{\nu}, \mathbf{I}_K), \text{ where } K = \sum_{i=1}^n k_i \text{ and } \boldsymbol{\nu} = \begin{pmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \\ \vdots \\ \boldsymbol{\nu}_n \end{pmatrix}.$$

Adding noncentral χ^2s

Proof of Theorem 2.5 (cont'd):
Then, it can be shown that $\sum_{i=1}^{n} X^2_{k_i,\lambda_i} = \sum_{i=1}^{n} \mathbf{y}_i^T \mathbf{y}_i = \mathbf{y}^T \mathbf{y}$.

Since $\mathbf{y} \sim MVN(\boldsymbol{\nu}, \mathbf{I}_K)$, by definition of the non-central χ^2 distribution, we have

$$\mathbf{y}^T \mathbf{y} \sim \chi^2_{K,\lambda},$$

where $\lambda = \frac{1}{2}\boldsymbol{\nu}^T \boldsymbol{\nu}$.

Observe that $\boldsymbol{\nu}^T \boldsymbol{\nu} = \sum_{i=1}^{n} \boldsymbol{\nu}_i^T \boldsymbol{\nu}_i$. Hence we have our required result:

$$\sum_{i=1}^{n} X^2_{k_i,\lambda_i} = \mathbf{y}^T \mathbf{y} \sim \chi^2_{\sum_{i=1}^{n} k_i, \sum_{i=1}^{n} \lambda_i}$$

Adding noncentral χ^2s

Caveat to my proof:

- We require $k_i \in \mathbb{Z}^+$.
- Challenge yourself: try proving the theorem for $k_i \in \mathbb{R}^+$ on your own.

Adding noncentral χ^2s

```
x1 <- rchisq(100,3,0)
x2 <- rchisq(100,6,2)
x3 <- rchisq(100,5,5)
hist(x1+x2+x3,freq=FALSE)
curve(dchisq(x,14,7),add=TRUE,col='red')
```

Histogram and density are shown in Figure 6.

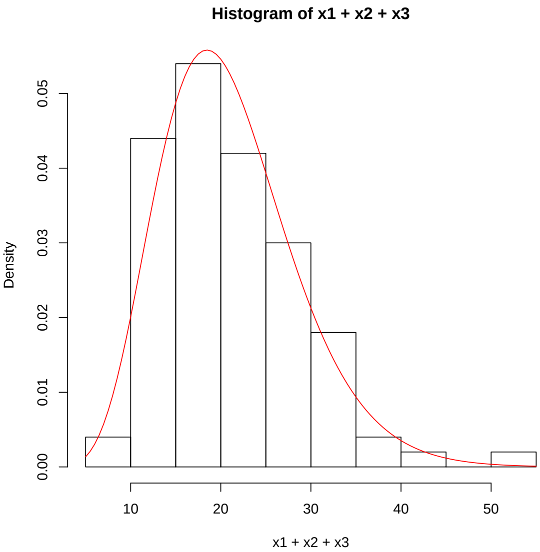

Figure: Adding 100 noncentral χ^2 rvs

Distribution of quadratic forms

Theorem 3.6

Let $\mathbf{y} \sim MVN(\boldsymbol{\mu}, I)$ be a $n \times 1$ random vector and let A be a $n \times n$ symmetric matrix. Then $\mathbf{y}^T A\mathbf{y}$ has a noncentral χ^2 distribution with k degrees of freedom and noncentrality parameter $\lambda = \frac{1}{2}\boldsymbol{\mu}^T A\boldsymbol{\mu}$ if and only if A is idempotent.

Proof of theorem

Proof. We only demonstrate $\Leftarrow$ as this is the one that will be most important for us. Assume that A is idempotent and let k denote its rank. Because it is symmetric, it can be diagonalised. Let the (orthogonal) diagonalising matrix be P.

Since A is symmetric and idempotent, all its eigenvalues are either 0 or 1 (Theorem 2.2). Moreover, Theorem 2.3 tells us that the sum of the eigenvalues is

$$tr(A) = r(A) = k.$$

Therefore, A must have k eigenvalues of 1 and $n - k$ eigenvalues of 0.

Proof of theorem

Recalling that the columns of P are the unit length eigenvectors corresponding to the eigenvalues, it is possible to arrange the columns of P so that all the 1 eigenvalues are first (using elementary matrix results which is beyond course's scope) Then we can partition the diagonalisation of A as

$$P^T AP = \left[\begin{array}{c|c} I_k & 0 \\ \hline 0 & 0 \end{array}\right].$$

Now define the random vector $\mathbf{z} = P^T\mathbf{y} \sim MVN(P^T\boldsymbol{\mu}, I_n)$. Partition the vectors/matrices as

$$\mathbf{z} = \left[\begin{array}{c} \mathbf{z}_1 \\ \hline \mathbf{z}_2 \end{array}\right], \quad P = \left[\begin{array}{c|c} P_1 & P_2 \end{array}\right]$$

where $\mathbf{z}_1$ is $k \times 1$ and P_1 is $n \times k$.

Proof of theorem

Then $\mathbf{z}_1 = P_1^T \mathbf{y} \sim MVN(P_1^T \boldsymbol{\mu}, I_k)$ and

$$\begin{aligned}
\mathbf{y}^T A \mathbf{y} &= (P\mathbf{z})^T A (P\mathbf{z}) = \mathbf{z}^T P^T A P \mathbf{z} \\
&= \left[\begin{array}{c|c} \mathbf{z}_1^T & \mathbf{z}_2^T \end{array}\right] \left[\begin{array}{c|c} I_k & 0 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{c} \mathbf{z}_1 \\ \hline \mathbf{z}_2 \end{array}\right] \\
&= \left[\begin{array}{c|c} \mathbf{z}_1^T & \mathbf{z}_2^T \end{array}\right] \left[\begin{array}{c} \mathbf{z}_1 \\ \hline 0 \end{array}\right] \\
&= \mathbf{z}_1^T \mathbf{z}_1.
\end{aligned}$$

Proof of theorem

Therefore, $\mathbf{y}^T A\mathbf{y} = \mathbf{z}_1^T\mathbf{z}_1$ has a noncentral χ^2 distribution with k degrees of freedom and noncentrality parameter

$$\lambda = \frac{1}{2}\boldsymbol{\mu}^T P_1 P_1^T \boldsymbol{\mu}.$$

Since

$$\begin{aligned} A &= \left[\begin{array}{c|c} P_1 & P_2 \end{array}\right] \left[\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{c} P_1^T \\ \hline P_2^T \end{array}\right] \\ &= \left[\begin{array}{c|c} P_1 & P_2 \end{array}\right] \left[\begin{array}{c} P_1^T \\ \hline 0 \end{array}\right] \\ &= P_1 P_1^T, \end{aligned}$$

we have

$$\lambda = \frac{1}{2}\boldsymbol{\mu}^T A\boldsymbol{\mu}.$$

Corollaries

Corollary 3.7

Let $\mathbf{y} \sim MVN(\mathbf{0}, I_n)$ *be a* $n \times 1$ *random vector and let* A *be a* $n \times n$ *symmetric matrix. Then* $\mathbf{y}^T A\mathbf{y}$ *has a (central)* χ^2 *distribution with* k *degrees of freedom if and only if* A *is idempotent and has rank* k.

Corollary 3.8

Let $\mathbf{y} \sim MVN(\boldsymbol{\mu}, \sigma^2 I_n)$ *be a* $n \times 1$ *random vector and let* A *be a* $n \times n$ *symmetric matrix. Then* $\frac{1}{\sigma^2}\mathbf{y}^T A\mathbf{y}$ *has a noncentral* χ^2 *distribution with* k *degrees of freedom and noncentrality parameter* $\lambda = \frac{1}{2\sigma^2}\boldsymbol{\mu}^T A\boldsymbol{\mu}$ *if and only if* A *is idempotent and has rank* k.

Example

Example. Let y_1 and y_2 be independent normal random variables with means 3 and -2 respectively and variance 1. Let

$$A = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

It is easy to verify that A is symmetric and idempotent, and has rank 1.

Example

Therefore

$$\mathbf{y}^T A\mathbf{y} = \frac{1}{2}\begin{bmatrix} y_1 & y_2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{2}y_1^2 + y_1 y_2 + \frac{1}{2}y_2^2$$

has a noncentral χ^2 distribution with 1 degree of freedom and noncentrality parameter

$$\lambda = \frac{1}{4}\begin{bmatrix} 3 & -2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \end{bmatrix} = \frac{1}{4}.$$

Distribution of quadratic forms

What happens if $\mathbf{y}$ does not have variance $\mathbf{I}$?

Theorem 3.9

Let $\mathbf{y} \sim MVN(\boldsymbol{\mu}, \mathbf{V})$ be a $n \times 1$ random vector, and let $\mathbf{A}$ be a $n \times n$ symmetric matrix. Then $\mathbf{y}^T \mathbf{A} \mathbf{y}$ has a noncentral χ^2 distribution with k degrees of freedom and noncentrality parameter $\lambda = \frac{1}{2} \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}$ if and only if $\mathbf{AV}$ is idempotent and has rank k.

Corollaries

Corollary 3.10

Let $\mathbf{y} \sim MVN(\mathbf{0}, V)$ *be a* $n \times 1$ *random vector and let* $\mathbf{A}$ *be a* $n \times n$ *symmetric matrix. Then* $\mathbf{y}^T\mathbf{A}\mathbf{y}$ *has a (central)* χ^2 *distribution with* k *degrees of freedom if and only if* $\mathbf{AV}$ *is idempotent and has rank* k*.*

Corollary 3.11

Let $\mathbf{y} \sim MVN(\boldsymbol{\mu}, \mathbf{V})$ *be a* $n \times 1$ *random vector where the variance* $\mathbf{V}$ *is nonsingular. Then* $\mathbf{y}^T\mathbf{V}^{-1}\mathbf{y}$ *has a noncentral* χ^2 *distribution with* n *degrees of freedom and noncentrality parameter* $\lambda = \frac{1}{2}\boldsymbol{\mu}^T\mathbf{V}^{-1}\boldsymbol{\mu}$*.*

R example: noncentral χ^2

Let

$$\mathbf{y} \sim MVN\left(\mathbf{a} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, V = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}\right).$$

```
y <- t(mvrnorm(100, mu=a, Sigma=V))
x <- apply(y, 2, quadform, A = solve(V))
(lambda <- t(a) %*% solve(V) %*% a / 2)

##          [,1]
## [1,] 2.277778

mean(x)

## [1] 6.56662
```

R example ctd: noncentral χ^2

```
2 + 2*lambda

##          [,1]
## [1,] 6.555556
```

```
hist(x, freq=F)
curve(dchisq(x, 2, 2*lambda), add = TRUE, col='red')
```

100 Bivariate normals generating non-central χ^2 rv's with density shown in Figure 7.

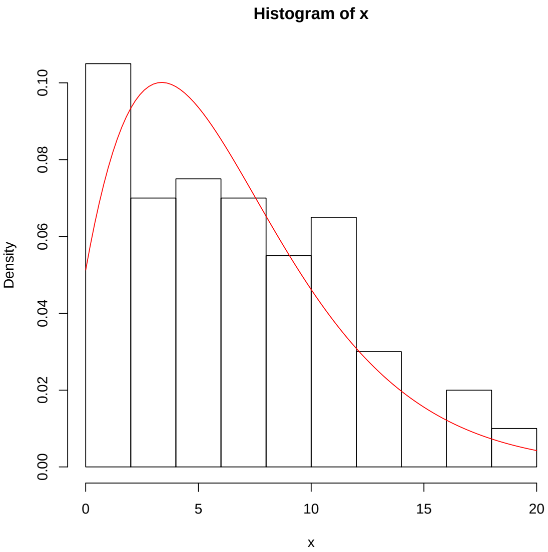

Figure: 100 Bivariate Normal Generate Noncentral χ^2

Noncentral χ^2

Example. Let y_1 and y_2 follow a multivariate normal distribution with means -1 and 4 respectively, and covariance matrix $V = \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix}$. Then

$$V^{-1} = \frac{1}{3 \times 2 - 2 \times 2} \begin{bmatrix} 2 & -2 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 3/2 \end{bmatrix},$$

and the quadratic form

$$\mathbf{y}^T V^{-1} \mathbf{y} = y_1^2 - 2y_1 y_2 + \frac{3}{2} y_2^2$$

has a noncentral χ^2 distribution with 2 degrees of freedom and noncentrality parameter

$$\lambda = \frac{1}{2} \begin{bmatrix} -1 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 3/2 \end{bmatrix} \begin{bmatrix} -1 \\ 4 \end{bmatrix} = \frac{33}{2}.$$

Sometimes we will want to know when two quadratic forms are independent. The next theorem tells us when this happens.

Theorem 3.12

Let $\mathbf{y} \sim MVN(\boldsymbol{\mu}, \mathbf{V})$ *be a* $n \times 1$ *random vector with nonsingular variance* $\mathbf{V}$*, and let* $\mathbf{A}$ *and* $\mathbf{B}$ *be symmetric matrices of order* n*. Then* $\mathbf{y}^T\mathbf{A}\mathbf{y}$ *and* $\mathbf{y}^T\mathbf{B}\mathbf{y}$ *are independent if and only if*

$$\mathbf{AVB} = 0.$$

Proof. Again we only prove $\Leftarrow$ because this is most important for us. Suppose that $\mathbf{AVB} = 0$. Since $\mathbf{V}$ is symmetric and positive definite we have $\mathbf{V} = \mathbf{C}^2$ for some symmetric, positive-definite $\mathbf{C}$, thus $\mathbf{ACCB} = \mathbf{0}$.

Let

$$\mathbf{R} = \mathbf{CAC}, \quad \mathbf{S} = \mathbf{CBC},$$

then

$$\mathbf{RS} = \mathbf{CACCBC} = \mathbf{0}.$$

Independence of quadratic forms

Because $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ are symmetric, $\mathbf{R}$ and $\mathbf{S}$ are also symmetric. Thus

$$\mathbf{SR} = \mathbf{S}^T\mathbf{R}^T = (\mathbf{RS})^T = \mathbf{0} = \mathbf{RS}.$$

By Theorem 2.5, we can find an orthogonal matrix $\mathbf{P}$ which diagonalises $\mathbf{R}$ and $\mathbf{S}$ simultaneously.

Since $\mathbf{C}$ is nonsingular, $r(\mathbf{R}) = r(\mathbf{CAC}) = r(\mathbf{A})$. Call the common rank r.

Independence of quadratic forms

Thus

$$\mathbf{P}^T\mathbf{R}\mathbf{P} = \left[\begin{array}{c|c} \mathbf{D}_1 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right]$$

where $\mathbf{D}_1$ has dimension $r \times r$.

The first r diagonal entries in $\mathbf{D}_1$ are the non-zero eigenvalues of R and the first r columns of $\mathbf{P}$ are the corresponding unit length eigenvectors.

The remaining $n - r$ diagonal entries are 0 since the rank, r, is the number of non-zero eigenvalues. The second $n - r$ columns of $\mathbf{P}$ are orthogonal unit length eigenvectors corresponding to the eigenvalue 0 for $\mathbf{R}$.

Independence of quadratic forms

Suppose $\mathbf{x}$ is one of the first r columns of $\mathbf{P}$, so that $\mathbf{Rx} = \lambda\mathbf{x}$ for some non-zero λ.

Then

$$\begin{aligned}\mathbf{Sx} &= \frac{1}{\lambda}\mathbf{S}\lambda\mathbf{x}\\ &= \frac{1}{\lambda}\mathbf{SRx}\\ &= \mathbf{0},\end{aligned}$$

since $\mathbf{SR} = \mathbf{0}$, so that $\mathbf{x}$ is an eigenvector of $\mathbf{S}$ with eigenvalue 0.

A similar argument applied to $\mathbf{S}$ shows that each of the non-zero eigenvalues of $\mathbf{S}$ has an eigenvector which is one of the second $n - r$ columns of $\mathbf{P}$ because it is a column corresponding to a $\mathbf{0}$ eigenvalue of $\mathbf{R}$.

Independence of quadratic forms

Hence

$$\mathbf{P}^T\mathbf{S}\mathbf{P} = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{D}_2 \end{array}\right]$$

where $\mathbf{D}_2$ is diagonal (possibly with some zero entries) and the partition has the same dimensions as that for $\mathbf{P}^T\mathbf{R}\mathbf{P}$ (ie $\mathbf{D}_2$ is $n - r \times n - r$).

Now define

$$\mathbf{z} = \mathbf{P}^T\mathbf{C}^{-1}\mathbf{y}.$$

Then $\mathbf{z}$ is multivariate normal with

$$E[\mathbf{z}] = \mathbf{P}^T\mathbf{C}^{-1}\boldsymbol{\mu}$$

and

$$Var(\mathbf{z}) = \mathbf{P}^T\mathbf{C}^{-1}\mathbf{V}\mathbf{C}^{-1}\mathbf{P} = \mathbf{P}^T\mathbf{P} = \mathbf{I}.$$

Thus the elements of $\mathbf{z}$ are independent.

Independence of quadratic forms

Partition $\mathbf{z}$ into

$$\mathbf{z} = \left[\begin{array}{c} \mathbf{z}_1 \\ \hline \mathbf{z}_2 \end{array}\right]$$

where $\mathbf{z}_1$ has dimension $r \times 1$.

By rewriting our equations, we see that

$$\begin{aligned} \mathbf{y} &= \mathbf{CPz} \\ \mathbf{A} &= \mathbf{C}^{-1}\mathbf{R}\mathbf{C}^{-1} \\ \mathbf{B} &= \mathbf{C}^{-1}\mathbf{S}\mathbf{C}^{-1} \end{aligned}$$

Independence of quadratic forms

So

$$\begin{aligned}
\mathbf{y}^T\mathbf{A}\mathbf{y} &= \mathbf{z}^T\mathbf{P}^T\mathbf{C}\mathbf{C}^{-1}\mathbf{R}\mathbf{C}^{-1}\mathbf{C}\mathbf{P}\mathbf{z} \\
&= \mathbf{z}^T\mathbf{P}^T\mathbf{R}\mathbf{P}\mathbf{z} \\
&= \left[\begin{array}{c|c} \mathbf{z}_1^T & \mathbf{z}_2^T \end{array}\right] \left[\begin{array}{c|c} \mathbf{D}_1 & 0 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{c} \mathbf{z}_1 \\ \hline \mathbf{z}_2 \end{array}\right] \\
&= \mathbf{z}_1^T\mathbf{D}_1\mathbf{z}_1
\end{aligned}$$

and similarly

$$\mathbf{y}^T\mathbf{B}\mathbf{y} = \mathbf{z}_2^T\mathbf{D}_2\mathbf{z}_2.$$

But $\mathbf{z}_1$ and $\mathbf{z}_2$ are mutually independent of each other, since all elements of $\mathbf{z}$ are independent. Therefore $\mathbf{y}^T\mathbf{A}\mathbf{y}$ and $\mathbf{y}^T\mathbf{B}\mathbf{y}$ are independent.

Example

Let $\mathbf{y} = (y_1, y_2)^T$ follow a multivariate normal distribution with covariance matrix

$$\mathbf{V} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

Consider the symmetric matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then

$$\mathbf{y}^T\mathbf{A}\mathbf{y} = y_1^2, \quad \mathbf{y}^T\mathbf{B}\mathbf{y} = y_2^2.$$

Independence of quadratic forms

These quadratic forms will be independent if and only if

$$\begin{aligned}
\mathbf{AVB} &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & b \\ 0 & c \end{bmatrix} \\
&= \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix}
\end{aligned}$$

is the $\mathbf{0}$ matrix.

But this means $b = 0$, i.e. y_1 and y_2 have zero covariance.

We've just shown that for multivariate random normals, uncorrelatedness gives independence of the squares of the components.

Independence of quadratic forms

Corollary 3.13

Let $\mathbf{y} \sim MVN(\boldsymbol{\mu}, \sigma^2 I)$ *be a random vector and let* $\mathbf{A}$ *and* $\mathbf{B}$ *be symmetric matrices. Then* $\mathbf{y}^T\mathbf{A}\mathbf{y}$ *and* $\mathbf{y}^T\mathbf{B}\mathbf{y}$ *are independent if and only if* $\mathbf{AB} = 0$.

Independence example

```r
A <- matrix(1, 2, 2)
B <- matrix(c(1,-1,-1,1), 2, 2)
A %*% B
```

```
##      [,1] [,2]
## [1,]    0    0
## [2,]    0    0
```

Independence example

```r
y <- t(mvrnorm(200, c(0, 0), diag(2)))
x1 <- apply(y, 2, quadform, A = A)
x2 <- apply(y, 2, quadform, A = B)
cor(x1, x2)
```

```
## [1] -0.08779417
```

```r
plot(x1, x2)
```

Plot in Figure 8

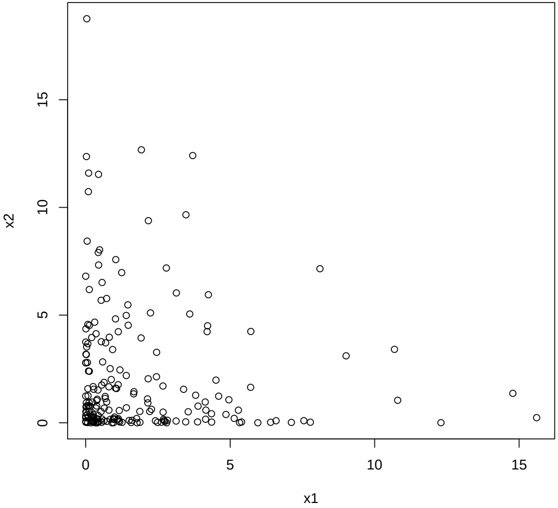

Figure: 2 sets of 200 Independent χ^2 rvs

Linear combination and quadratic form

Next we consider when a quadratic form is independent of a random vector. Firstly, we define a random variable to be independent of a random vector if and only if it is independent of all elements of that vector.

Theorem 3.14

Let $\mathbf{y} \sim MVN(\boldsymbol{\mu}, \mathbf{V})$ be a $n \times 1$ random vector, and let A be a symmetric matrix or order n and $\mathbf{B}$ a $m \times n$ matrix. Then $\mathbf{y}^T\mathbf{A}\mathbf{y}$ and $\mathbf{B}\mathbf{y}$ are independent if and only if $\mathbf{BVA} = 0$.

Lastly, we can combine several of the theorems we have seen before to tell when a group of quadratic forms (more than two) are independent.

m quadratic forms

Theorem 3.15

Let $\mathbf{y} \sim MVN(\boldsymbol{\mu}, I)$ be a random vector, and let $\mathbf{A}_1, \ldots, \mathbf{A}_m$ be a set of symmetric matrices. If any two of the following statements are true:

- *All $\mathbf{A}_i$ are idempotent;*
- *$\sum_{i=1}^{m} \mathbf{A}_i$ is idempotent;*
- *$\mathbf{A}_i \mathbf{A}_j = 0$ for all $i \neq j$;*

then so is the third, and

- *For all i, $\mathbf{y}^T \mathbf{A}_i \mathbf{y}$ has a noncentral χ^2 distribution with $r(\mathbf{A}_i)$ d.f. and noncentrality parameter $\lambda_i = \frac{1}{2} \boldsymbol{\mu}^T \mathbf{A}_i \boldsymbol{\mu}$;*
- *$\mathbf{y}^T \mathbf{A}_i \mathbf{y}$ and $\mathbf{y}^T \mathbf{A}_j \mathbf{y}$ are independent for $i \neq j$; and*
- *$\sum_{i=1}^{m} r(\mathbf{A}_i) = r\left(\sum_{i=1}^{m} \mathbf{A}_i\right)$.*

Cochran-Fisher Theorem

When $\sum_i \mathbf{A}_i = I$, the previous result is related to the following theorem (which we will not prove):

Theorem 3.16 (Cochran-Fisher Theorem)

Let $\mathbf{y} \sim MVN(\boldsymbol{\mu}, \sigma^2 I)$ be a $n \times 1$ random vector. Decompose the sum of squares of $\mathbf{y}/\sigma$ into the quadratic forms

$$\frac{1}{\sigma^2}\mathbf{y}^T\mathbf{y} = \sum_{i=1}^{m} \frac{1}{\sigma^2}\mathbf{y}^T\mathbf{A}_i\mathbf{y}.$$

Then the quadratic forms are independent and have noncentral χ^2 distributions with parameters $r(\mathbf{A}_i)$ and $\frac{1}{2\sigma^2}\boldsymbol{\mu}^T\mathbf{A}_i\boldsymbol{\mu}$, respectively, if and only if

$$\sum_{i=1}^{m} r(\mathbf{A}_i) = n.$$

Big picture - recall

You may recall that last semester there were a number of examples where χ^2 rv's were claimed to be independent (for example, the sums of squares in analysis of variance for the normal location problem (t-distribution), regression).

In each case, we separated the sum of squares of the response variable (around the parameters) into a sum of two quadratic forms:

$$\mathbf{y}^T\mathbf{y} = \mathbf{y}^T\mathbf{A}_1\mathbf{y} + \mathbf{y}^T\mathbf{A}_2\mathbf{y}.$$

Big picture - recall

Since $\mathbf{A}_1 + \mathbf{A}_2 = \mathbf{I}$, it is idempotent.

Then if we can establish one of the other two conditions given in the theorem (most commonly that $\mathbf{A}_i$ is idempotent), we can use the Theorem 3.15 to determine the distribution of the quadratic forms, and also conclude their independence.

This will justify the claims of independence in all cases so far, and many more that are central in techniques of machine learning.