Chapter 6

Normal errors

1 The multivariate normal and related distributions

Let $Z_1, Z_2, \ldots, Z_n$ be independent $N(0,1)$ random variables. When treated as the coordinates of a point in $\mathbb{R}^n$ they define a random vector $\mathbf{Z}$, whose (joint) density function is

$$f(\mathbf{z}) = (2\pi)^{-n/2} \exp\left(-\tfrac{1}{2}\sum\nolimits_i z_i^2\right) = (2\pi)^{-n/2} \exp\left(-\tfrac{1}{2}\|\mathbf{z}\|^2\right).$$

Such a random vector is said to have a ***spherical normal distribution***. That is, $\mathbf{Z} \sim N(0, I_n)$.

(i) The ***chi-square***, χ_n^2, is defined as the distribution of the sum of squares $Z_1^2 + \cdots + Z_n^2$ of independent $N(0,1)$ random variables. The ***noncentral chi-square***, $\chi_n^2(\gamma)$, with noncentrality parameter $\gamma \ge 0$ is defined as the distribution of the sum of squares $(Z_1+\gamma)^2 + Z_2 \cdots + Z_n^2$.

(ii) If $Z \sim N(0,1)$ is independent of $S_k^2 \sim \chi_k^2$ then

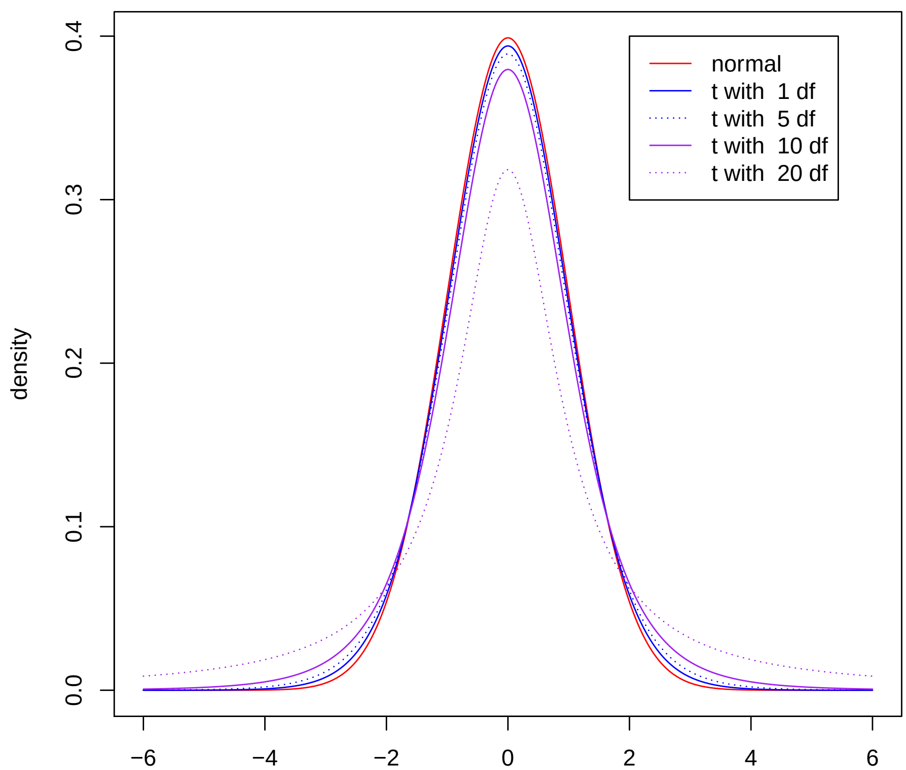
$$\frac{Z}{\sqrt{S_k^2/k}} \text{ has a } t\text{-distribution on } k \text{ degrees of freedom } (t_k)$$

(iii) If $S_k^2 \sim \chi_k^2$ is independent of $S_\ell^2 \sim \chi_\ell^2$ then

$$\frac{S_\ell^2/\ell}{S_k^2/k} \text{ has an } F\text{-distribution on } \ell \text{ and } k \text{ degrees of freedom } (F_{\ell,k})$$

The t disributions are actually not much different from the normal if the degrees of freedom and not too small.

```
xx <- seq(-6,6,by=0.01)
T1 <- dt(xx,1); T5 <- dt(xx,5); T10 <- dt(xx,10); T20 <- dt(xx,20); NN <- dnorm(xx)
plot(xx,NN,col="red",xlab="",ylab="density",type="l")
lines(xx,T20,col="blue"); lines(xx,T10,col="blue",lty=3)
lines(xx,T5,col="purple"); lines(xx,T1,col="purple",lty=3)
legend(2, 0.4, leg = c("normal",paste("t with ", c(1,5,10,20),"df")),lty=c(1,1,3,1,3),
col= c("red","blue","blue","purple","purple"),cex=1)
```



As the degrees of freedom increase, the density at zero increases to the value of the normal density at zero.

2 Rotation of axes

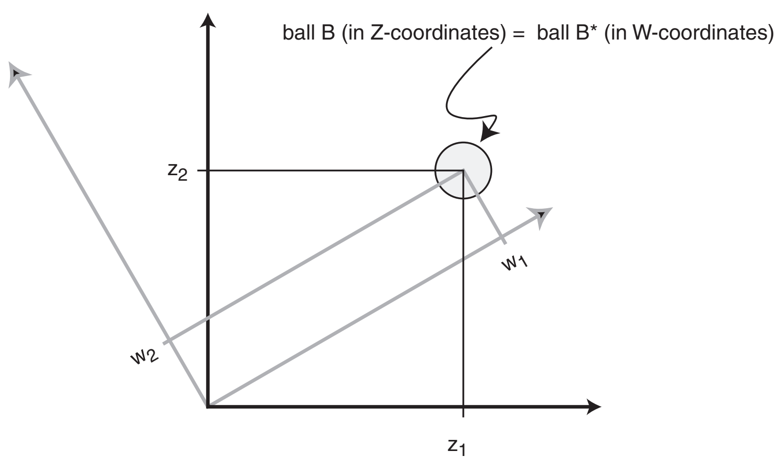
The spherical symmetry of the density $f(\cdot)$ is responsible for an important property of multivariate normals. Let $\mathbf{q}_1, \dots, \mathbf{q}_n$ be a new orthonormal basis for $\mathbb{R}^n$, and let

$$\mathbf{Z} = W_1 \mathbf{q}_1 + \cdots + W_n \mathbf{q}_n$$

be the representation for $\mathbf{Z}$ in the new basis.

<6.1> **Theorem.** *The $W_1, \dots, W_n$ are also independent $N(0,1)$ distributed random variables.*

If you know about multivariate moment generating functions this is easy to establish using the matrix representation $\mathbf{Z} = Q\mathbf{W}$, where Q is the orthogonal matrix with columns $\mathbf{q}_1, \dots, \mathbf{q}_n$.



A more intuitive explanation is based on the approximation

$$\mathbb{P}\{\mathbf{Z} \in B\} \approx f(\mathbf{z})(\text{volume of } B)$$

for a small ball B centered at $\mathbf{z}$. The transformation from $\mathbf{Z}$ to $\mathbf{W}$ corresponds to a rotation, so

$$\mathbb{P}\{\mathbf{Z} \in B\} = \mathbb{P}\{\mathbf{W} \in B^*\},$$

where B^* is a ball of the same radius, but centered at the point $\mathbf{w} = (w_1, \dots, w_n)$ for which $w_1 \mathbf{q}_1 + \cdots + w_n \mathbf{q}_n = \mathbf{z}$. The last equality implies $\|\mathbf{w}\| = \|\mathbf{z}\|$, from which we get

$$\mathbb{P}\{\mathbf{W} \in B^*\} \approx (2\pi)^{-n/2} \exp(-\tfrac{1}{2}\|\mathbf{w}\|^2)(\text{volume of } B^*).$$

That is, $\mathbf{W}$ has the asserted spherical normal density.

To prove results about the spherical normal it is often merely a matter of transforming to an appropriate orthonormal basis.

<6.2> **Theorem.** *Let $\mathcal{X}$ be an m-dimensional subspace of $\mathbb{R}^n$. Let $\mathbf{Z}$ be a vector of independent $N(0,1)$ random variables, and $\boldsymbol{\mu}$ be a vector of constants. Then*

(i) the projection $\widehat{\mathbf{Z}}$ of $\mathbf{Z}$ onto $\mathcal{X}$ is independent of the projection $\mathbf{Z}-\widehat{\mathbf{Z}}$ of $\mathbf{Z}$ onto $\mathcal{X}^\perp$, the orthogonal complement of $\mathcal{X}$.

(ii) $\left\|\widehat{\mathbf{Z}}\right\|^2$ has a χ^2_m distribution.

(iii) $\|\mathbf{Z}+\boldsymbol{\mu}\|^2$ has a noncentral $\chi^2_n(\gamma)$ distribution, with $\gamma = \|\boldsymbol{\mu}\|$.

(iv) $\left\|\widehat{\mathbf{Z}}+\boldsymbol{\mu}\right\|^2$ has a noncentral $\chi^2_m(\gamma)$ distribution, with $\gamma = \|\boldsymbol{\mu}\|$.

Proof Let $\mathbf{q}_1,\dots\mathbf{q}_n$ be an orthonormal basis of $\mathbb{R}^n$ such that $\mathbf{q}_1,\dots,\mathbf{q}_m$ span the space $\mathcal{X}$ and $\mathbf{q}_{m+1},\dots,\mathbf{q}_n$ span $\mathcal{X}^\perp$. If $\mathbf{Z} = W_1\mathbf{q}_1+\dots+W_n\mathbf{q}_n$ then

$$\begin{aligned}\widehat{\mathbf{Z}} &= W_1\mathbf{q}_1+\dots+W_m\mathbf{q}_m,\\ \mathbf{Z}-\widehat{\mathbf{Z}} &= W_{m+1}\mathbf{q}_{m+1}+\dots+W_n\mathbf{q}_n,\\ \|\mathbf{Z}\|^2 &= W_1^2+\dots+W_m^2,\end{aligned}$$

from which the first two asserted properties follow.

For the third and fourth assertions, choose the basis so that $\boldsymbol{\mu} = \gamma\mathbf{q}_1$. Then

$$\begin{aligned}\mathbf{Z}+\boldsymbol{\mu} &= (W_1+\gamma)\mathbf{q}_1+W_2\mathbf{q}_2+\dots \qquad\qquad +W_n\mathbf{q}_n\\ \widehat{\mathbf{Z}}+\boldsymbol{\mu} &= (W_1+\gamma)\mathbf{q}_1+W_2\mathbf{q}_2+\dots+W_m\mathbf{q}_m\end{aligned}$$

from which we get the noncentral chi-squares.

□

<6.3> **Example.** Suppose $X_1,\dots,X_n$ are independent random variables, each distributed $N(\mu,\sigma^2)$. Define $\overline{X} = n^{-1}\sum_{i\le n}X_i$ and $S^2 = \sum_{i\le n}(X_i-\overline{X})^2$. Many textbooks prove the following assertion in a gruesome way:

$$\overline{X}\sim N(\mu,\sigma^2/n) \text{ independent of } S^2/\sigma^2\sim\chi^2_{n-1}.$$

The clean proof uses the fact that the random variables $Z_i = (X_i-\mu)/\sigma$ are independent $N(0,1)$'s, so that $\mathbf{Z} = (Z_1,\dots,Z_n)\sim N(0,I_n)$. Define

$q_1 = \mathbb{1}/\sqrt{n}$ then find $q_2, \dots, q_n$ so that $\{q_i : 1 \le i \le n\}$ is an onb for $\mathbb{R}^n$. (Actually it is not necessary to calculate $q_2, \dots, q_n$ explicitly. It suffices to know that such q_i's exist.

From Theorem <6.1>, if

$$\mathbf{Z} = W_1 q_1 + \cdots + W_n q_n$$

then the W_i's are independent $N(0,1)$. In particular,

$$\overline{Z} = \mathbb{1}^T \mathbf{Z}/n = q_1^T \mathbf{Z}/\sqrt{n} = W_1/\sqrt{n} \sim N(0, 1/n)$$

so that

$$\overline{X} = \mu + \sigma \overline{Z} \sim N(\mu, \sigma^2/n).$$

Also $\mathbf{Z} - \overline{Z}\mathbb{1} = \sum_{i=2}^n W_i q_i$ so that

$$S^2 = \sigma^2 \sum\nolimits_{i \le n} (Z_i - \overline{Z})^2 = \sigma^2 \sum\nolimits_{2 \le i \le n} W_i^2.$$

The independence comes from the fact that $\overline{X}$ is a function of W_1 and S^2 is a function of $W_2, \dots, W_n$. Notice also that

$$\frac{\sqrt{n}(\overline{X} - \mu)}{\sqrt{S^2/(n-1)}} = \frac{\sqrt{n}\sigma\overline{Z}}{\sigma\sqrt{(Z_i - \overline{Z})^2/(n-1)}} = \frac{W_1}{\sqrt{\sum_{i \ge 2} W_i^2/(n-1)}} \sim t_{n-1}.$$

The final assertion comes from the fact that $W_1 \sim N(0,1)$ independently of $\sum_{i \ge 2} W_i^2 \sim \chi^2_{n-1}$.

Now suppose we were wondering if μ were really zero. If it were, then

$$T_{obs} = \frac{\sqrt{n}\overline{X}}{\sqrt{S^2/(n-1)}}$$

would be distributed t_{n-1}. We could then calculate a two-sided p-value, $p_{obs} = \text{tail}(T_{obs}, n-1)$ where

$$\text{tail}(x, n-1)) = \mathbb{P}\{|T| \ge x\} \qquad \text{for } T \sim t_{n-1}.$$

If p_{obs} is very small then we are faced with a choice between "$\mu = 0$ and we have just observed the occurrence of a rare event" or "$|T_{obs}|$ is large, perhaps because $|\mu|$ is a long way from zero."

□

3 Facts about the multivariate normal

Suppose $Z \sim N(0, I_n)$ and μ is an $m \times 1$ vector of constants. If A is an $m \times n$ matrix of constants then the random vector $X = \mu + AZ$ has expected value μ with variance matrix $V = AA'$, and moment generating function

$$\mathbb{E}\exp(t^T X) = \exp(t^T \mu + t^T A A^T t/2) = \exp(t^T \mu + t^T V t/2).$$

The distribution of X depends only on μ and V. The random vector X has a $N(\mu, V)$ distribution.

If γ is a $k \times 1$ vector of constants and B is a $k \times m$ matrix of constants then

$$\gamma + BX = (\gamma + B\mu) + BAZ \sim N(\gamma + B\mu, BVB').$$

4 Least squares

Much of the distribution theory for least squares has been worked out for the simple model where $y = \mu + \xi \sim N(\mu, \sigma^2 I_n)$, where the unknown μ is assumed to lie in some known p-dimensional subspace $\mathcal{X}$ of $\mathbb{R}^n$ and σ^2 is unknown.

Write ξ as $\sigma\mathbf{Z}$, where $\mathbf{Z} \sim N(0, I_n)$. Let $q_1, \dots, q_n$ be an onb for $\mathbb{R}^n$ such that $q_1, \dots, q_p$ are an onb for $\mathcal{X}$ and $q_{p+1}, \dots, q_n$ are an onb for $\mathcal{X}^\perp$. Then $\mathbf{Z} = \sum_{i \le n} W_i q_i$ with, by Theorem <6.1>, $\mathbf{W} \sim N(0, I_n)$.

The matrix

$$H = \sum\nolimits_{i \le p} q_i q_i^T$$

projects vectors orthogonally onto $\mathcal{X}$. Thus

$$\widehat{y} = H(\mu + \sigma\mathbf{Z}) = \mu + \sigma H\mathbf{Z} = \mu + \sigma \sum\nolimits_{i \le p} W_i q_i$$

$$y - \widehat{y} = \sigma \sum\nolimits_{i > p} W_i.$$

Independence of the W_i's implies that $\widehat{y}$ and $y - \widehat{y}$ are independent, with

$$y \sim N(\mu, \sigma^2 H) \quad \text{AND} \quad y - \widehat{y} \sim N(0, \sigma^2(I_n - H)).$$

Under the model, the residual sum of squares equals

$$\text{RSS} = \|y - \widehat{y}\|^2 = \sigma^2 \sum\nolimits_{i > p} W_i^2,$$

which implies that $\text{RSS}/\sigma^2 \sim \chi^2_{n-p}$. The estimate of σ^2 is $\widehat{\sigma}^2 = \text{RSS}/(n-p)$ is independent of $\widehat{y}$.

5 Some t-tests and p-values

Consider first the simplest case where X is an $n \times p$ matrix of rank p and the θ model posits that $y \sim N(X\theta, \sigma^2 I_n)$. That is, $\mathbb{E}_\theta y = X\theta$ and $y = X\theta + \xi$ where $\xi \sim N(0, \sigma^2 I_n)$.

The matrix X has a qr-decomposition $X = Q_1 R_1$ where Q_1 is an $n \times p$ matrix whose columns provide an onb for $\mathfrak{X}$ and R_1 is an $m \times m$ upper-triangular matrix of rank m, that R_1 has an $m \times m$ inverse S_1.

The orthogonal projection of y onto $\mathfrak{X}$ equals Hy, for hat matrix $H = Q_1 Q_1^T$. The least squares estimate $\widehat{\theta}$ is defined by $\widehat{y} = X\widehat{\theta}$. That is,

$$\widehat{y} = Q_1 Q_1^T y = Q_1 R_1 \widehat{\theta} \quad \text{AND} \quad \widehat{\theta} = S_1 Q_1^T \widehat{y} = S_1 Q_1^T y.$$

Under the model, $\widehat{\theta} \sim N(\theta, \sigma^2 S_1 S_1^T)$. In particular, $\widehat{\theta}_j \sim N(\theta_j, \sigma^2 v_1^2)$, where v_j^2 is the jth diagonal element of $S_1 S_1^T$.

By the independence of $\widehat{y}$ and RSS, under the θ model

$$\frac{\widehat{\theta}_j - \theta_j}{v_j \widehat{\sigma}} = \frac{\left(\widehat{\theta}_j - \theta_j\right)/(v_1\sigma)}{\sqrt{\text{RSS}/(n-p)\sigma^2}} \sim t_{n-p}.$$

If $\theta_j = 0$ then, under the model

$$T_{obs,j} = \widehat{\theta}_j/(v_j\widehat{\sigma}) \sim t_{n-p}.$$

We could then calculate a two-sided p-value, $p_{obs,j} = \text{tail}(T_{obs,j}, n-p)$ where

$$\text{tail}(x, n-p)) = \mathbb{P}\{|T| \geq x\} \qquad \text{for } T \sim t_{n-p}.$$

That is, the interpretation parallels the interpretation in Example <6.3>. For example, in the following summary table, each line gives the name corresponding to θ_j, the estimate $v_j\widehat{\sigma}$ for the square root of $\text{var}(\widehat{\theta}_j)$, the ratio $T_{obs,j}$, and $p_{obs,j}$. Formally the p-value corresponds to a test of the null hypothesis $\theta_j = 0$ under the θ model. If the model is badly wrong then the p-value has little meaning.

```
cath <- read.table("catheter.txt",header=T)
outHW <- lm(distance ~ height + weight,cath)
look(outHW)
```

```
## lm(formula = distance ~ height + weight, data = cath)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   21.008      8.751   2.401    0.040
## height         0.196      0.361   0.545    0.599
## weight         0.191      0.165   1.155    0.278
```

Now for the harder case where the matrix X has rank $m < p$. For example, for the Box-Cox data discussed in Contrasts.pdf, the conceptual design matrix prescribed by `lm( rate ~ treatment + poison,BC)` is a 48×8 matrix

$$X = (\mathbb{1}_{48}, F_1, F_2, F_3, F_4, G_1, G_2, G_3)$$

where $F = (F_1, F_2, F_3, F_4)$ is the matrix of summy variables for the factor `Ht` and $G = (G_1, G_2, G_3)$ is the matrix of summy variables for the factor `Hp`

By means of the (Helmert) contrasts for the two factors `Ht` and `Hp`, **R** replaces X by the 48×6 matrix

$$\widetilde{X} = X\mathbb{M} \qquad \text{where } \mathbb{M} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & C_4 & 0 \\ 0 & 0 & C_3 \end{pmatrix},$$

which has rank 6. The matrix $\widetilde{X}$ has qr-decomposition Q_1R_1 where Q_1 is a 48×6 matrix whose columns provide an onb for the 6-dimensional subspace $\mathcal{X}$ for $\mathbb{R}^{48}$ spanned by the columns of X. The 6×6 upper triangular matrix R_1

```
XX <- model.matrix(outBC)
C3 <- contr.helmert(3)
C4 <- contr.helmert(4)
MM <-bdiag(1,C3,C4)
print(MM)

## 8 x 6 sparse Matrix of class "dgCMatrix"
##
## [1,] 1  .  .  .  .  .
## [2,] . -1 -1  .  .  .
## [3,] .  1 -1  .  .  .
## [4,] .  .  2  .  .  .
## [5,] .  .  . -1 -1 -1
## [6,] .  .  .  1 -1 -1
## [7,] .  .  .  .  2 -1
## [8,] .  .  .  .  .  3

BCqr <- outBC$qr
R1 <- qr.R(BCqr); Q1 <- qr.Q(BCqr)
S1 <- solve(R1) # inverse of R1
round(R1,1)
```

```
##   (Intercept) Ht1 Ht2 Ht3  Hp1 Hp2
## 1        -6.9 0.0 0.0   0  0.0 0.0
## 2         0.0 4.9 0.0   0  0.0 0.0
## 3         0.0 0.0 8.5   0  0.0 0.0
## 4         0.0 0.0 0.0  12  0.0 0.0
## 5         0.0 0.0 0.0   0 -5.7 0.0
## 6         0.0 0.0 0.0   0  0.0 9.8
```

It has inverse S_1. Define $\Delta = Q_1 S_1$. Then

$$\Delta^T \widetilde{X} = S_1^T Q_1^T Q_1 R_1 = I_6$$

```
Delta <- Q1 %*% S1
round(t(Delta) %*% XX,3)

##   (Intercept) Ht1 Ht2 Ht3 Hp1 Hp2
## 1           1   0   0   0   0   0
## 2           0   1   0   0   0   0
## 3           0   0   1   0   0   0
## 4           0   0   0   1   0   0
## 5           0   0   0   0   1   0
## 6           0   0   0   0   0   1
```

Vectors in $\mathcal{X}$ have a unique representation as $X\theta$, with $\theta = \mathbb{M}t$ and $t \in \mathbb{R}^6$. The coefficients $\widehat{t}$ for which $\widehat{y} = \widetilde{X}\widehat{t}$ are contained in `outBC`.

```
round(outBC$coeff,3)

## (Intercept)         Ht1         Ht2         Ht3         Hp1         Hp2
##       2.622      -0.829       0.086      -0.154       0.234       0.587
```

The coefficients $\widehat{\theta} = \mathbb{M}\widehat{t}$ satisfy the sum constraints and $\widehat{y} = X\widehat{\theta}$.