Linear models with applications in R

PUBHLTH 744: Handout 8(Distribution Theory)

Instructor: Andrea S. Foulkes

Division of Biostatistics and Epidemiology
UMass School of Public Health and Health Sciences

Fall 2007

Sampling distribution of estimators

Recall in the usual linear model, when $Cov(\epsilon) = \sigma^2 I$ and $\Lambda'\beta$ is estimable where $\Lambda' = P'X$, the BLUE and UMVUE of $P'X\beta$ is $P'MY$. We have the following results:

$$\begin{aligned} E(P'MY) &= P'ME(Y) \\ &= P'MX\beta \\ &= P'X\beta = \Lambda'\beta \end{aligned}$$

and

$$\begin{aligned} Cov(P'MY) &= (P'M)Cov(Y)(P'M)' \\ &= (P'M)\sigma^2 I(MP) \\ &= \sigma^2 P'MP \\ &= \sigma^2 \Lambda'(X'X)^{-}\Lambda \end{aligned}$$

If we additionally assume $\epsilon \sim MVN(0, \sigma^2 I)$ then we have $P'MY \sim MVN(\Lambda'\beta, \sigma^2\Lambda'(X'X)^-\Lambda)$. For X full rank, β is estimable and the unique least squares estimate is $\widehat{\beta} = (X'X)^{-1}X'Y$. We then have the following results:

$$\begin{aligned} E(\widehat{\beta}) &= E((X'X)^{-1}X'Y) \\ &= (X'X)^{-1}X'X\beta = \beta \end{aligned}$$

and

$$\begin{aligned} Cov(\widehat{\beta}) &= Cov((X'X)^{-1}X'Y) \\ &= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

Therefore, $\widehat{\beta} \sim MVN(\beta, \sigma^2(X'X)^{-1})$.

Theory of distributions

- Central and non-central Chi-square distributions

Definition: A random variable, X has a central chi-square distribution with n degrees of freedom, written $X \sim \chi_n^2$ if the density of X is given by:

$$f(x) = \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} x^{n/2-1} e^{-x/2}$$

Definition: A random variable, X is said to have a non-central chi-square distribution with n degrees of freedom and non-centrality parameter γ, written $X \sim \chi^2_{n,\gamma}$, if the density of X is given by:

$$\begin{aligned} f(x) &= \sum_{i=0}^{\infty} \left(\frac{\gamma^i e^{-\gamma}}{i!} \right) \frac{1}{\Gamma((2i+n)/2)} \left(\frac{1}{2} \right)^{(2i+n)/2} x^{(2i+n)/2-1} e^{-x/2} \\ &= \sum_{i=0}^{\infty} \left(\frac{\gamma^i e^{-\gamma}}{i!} \right) g_i(x) \end{aligned}$$

where $g_i(\cdot)$ is a central chi-square density with $2i+n$ degrees of freedom.

- That is, the non-central chi-square distribution is an infinite Poisson mixture of central chi-square densities. Equivalently, the non-central chi-square distribution is a central chi-square distribution with $2P + n$ degrees of freedom where $P \sim Poisson_{\gamma}$.
- Non-central χ^2 distributions are important for characterizing the distribution of tests statistics under alternative hypotheses.
- Using moment generating functions ICBST if $X \sim \chi^2_{n,\gamma}$, then $E(X) = n + 2\gamma$ and $Var(X) = 2n + 8\gamma$.

- Normal and multivariate normal distributions

Definition: A random variable, X is said to have a normal distribution with mean μ and variance σ^2, written $X \sim N(\mu, \sigma^2)$ if the density of X is given by:

$$f(x) = (2\pi)^{-1/2}\sigma^{-1}exp\left\{\frac{-1}{2\sigma^2}(x-\mu)^2\right\}$$

If $\mu = 0$ and $\sigma^2 = 1$ then $X \sim N(0,1)$ and we say X has a standard normal distribution.

Theorem: Suppose $Z_1, \ldots, Z_n$ are independent and identically distributed (i.i.d) $N(0,1)$ random variables. Further suppose $X = \sum_{i=1}^{n} Z_i^2$, then $X \sim \chi_n^2$.

Theorem: Suppose $Y_1, \ldots, Y_n$ are independent random variables and $Y_i \sim N(\mu_i, \sigma^2)$. Further suppose $X = \frac{1}{\sigma^2} \sum_{i=1}^{n} Y_i^2$, then $X \sim \chi_{n,\gamma}^2$ where $\gamma = \frac{1}{\sigma^2} \sum_{i=1}^{n} \mu_i^2$.

<u>Definition 8.4</u>: Suppose $X = (X_1, \ldots, X_n)'$, then X is said to have an n-dimensional <u>multivariate normal distribution</u> with mean μ and covariance matrix Σ if the density of X is given by

$$f(X) = (2\pi)^{-n/2}|\Sigma|^{-1/2}exp\left\{\frac{-1}{2\sigma^2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}$$

Note this requires Σ to be positive definite (because a non-positive definite matrix has 0 determinant.)

- Properties of MVN
  - The moment generating function of $X \sim MVN_n(\mu, \Sigma)$ is given by:

$$\Phi_X(t) = E(e^{t'X}) = exp\left\{t'\mu + \frac{1}{2}t'\Sigma t\right\}$$

  - A linear combination of MVNs is MVN. That is, suppose $X \sim MVN_n(\mu, \Sigma)$ and define $Y = AX + b$ where $A_{r\times n}$ and $b_{r\times 1}$ are constants. Then $Y \sim MVN_r(A\mu + b, A\Sigma A')$.

- Properties of MVN continued
  - Proof of the last result is straightforward using moment generating functions.

$$
\begin{aligned}
\Phi_Y(t) &= E(e^{t'Y}) = E(e^{t'(AX+b)}) \\
&= e^{t'b}E(e^{t'AX}) \\
&= e^{t'b}E(e^{(A't)'X}) \\
&= e^{t'b}\Phi_X(A't) \\
&= e^{t'b}exp\left\{(A't)'\mu + \frac{1}{2}(A't)'\Sigma A't\right\} \\
&= exp\left\{t'(A\mu + b) + \frac{1}{2}t'A\Sigma A't\right\}
\end{aligned}
$$

But we know this is the m.g.f. for a random variable with a MVN distribution, mean $A\mu + b$ and Covariance $A\Sigma A'$. $\square$

- Properties of MVN continued
  - A linear combination of multivariate normals is multivariate normal. That is, if $X_i \sim MVN_n(\mu_i, \Sigma_i)$, $i = 1, \ldots, k$ and $X_i \perp X_j$ for $i \neq j$ and $Y = a_1 X_1 + \ldots + a_k X_k$ for a_i scalars, then $Y \sim MNV_n\left(\sum_{i=1}^k a_i \mu_i, \sum_{i=1}^k a_i^2 \Sigma_i\right)$.
  - Marginal distributions of MVNs are MVN. Consider $Y \sim MVN_n(\mu, \Sigma)$. Partition Y into $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$, where Y_1 is $r \times 1$ and Y_2 is $(n-r) \times 1$. Further write $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Then we have $X_1 \sim MVN_r(\mu_1, \Sigma_{11})$ and we say the marginal distribution of X_1 is MVN.
  - Conditional distributions of MVNs are MVN. Again suppose $Y \sim MVN_n(\mu, \Sigma)$. ICBST:

$$Y_1 | Y_2 = y_2 \sim MVN_r(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

__**Example**__: Suppose $X = \begin{pmatrix} X_1 \\ X_2 \\ X_2 \end{pmatrix} \sim MVN_3(\mu, \Sigma)$ where

$\mu = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix}$. The marginal distribution of $\begin{pmatrix} X_1 \\ X_3 \end{pmatrix}$ is given by

$\begin{pmatrix} X_1 \\ X_3 \end{pmatrix} \sim MVN_2\left(\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\right)$. The conditional distribution of

$\begin{pmatrix} X_1 \\ X_3 \end{pmatrix} | X_2 = x_2 \sim MVN_2\left(\begin{pmatrix} 3 - x_2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\right)$. Note that the dependency on X_2 only enters into the mean.

- T and non-central t-distributions

Definition: Suppose $X \sim N(0,1)$, $Y \sim \chi^2_n$ and $X \perp Y$ (independent.) Define:

$$T = \frac{X}{\sqrt{Y/n}}$$

We say T has a t-distribution with n degrees of freedom, written $T \sim t_n$.

Definition: Suppose $X \sim N(\mu, 1)$, $Y \sim \chi^2_n$ and $X \perp Y$ (independent.) Define:

$$W = \frac{X}{\sqrt{Y/n}}$$

We say W has a non-central t-distribution with n degrees of freedom and noncentrality parameter μ, written $W \sim t_{n,\mu}$.

- F distribution

Definition 8: Suppose $X_1 \sim \chi^2_{n_1,\gamma_1}$, $X_2 \sim \chi^2_{n_2,\gamma_2}$ and $X_1 \perp X_2$. The random variable:

$$F = \frac{X_1/n_1}{X_2/n_2}$$

is said to have a doubly noncentral F distribution with (n_1, n_2) degrees of freedom and noncentrality parameters (γ_1, γ_2), written $F \sim F_{(n_1,n_2),(\gamma_1,\gamma_2)}$. If $\gamma_2 = 0$ then F is said to have a noncentral F distribution. If $\gamma_1 = 0$ and $\gamma_2 = 0$ then F is said to have a central F distribution.

```
####################################
# Code for generating distributions
####################################

# Generating N=100 i.i.d. standard normals
> y <- rnorm(100,0,1)
> sort(y)
  [1] -2.61282437 -1.93248953 -1.91647985 -1.90535191 -1.87379699 -1.85172423
  [7] -1.57188648 -1.56764143 -1.48920480 -1.43521924 -1.40972669 -1.30731356
 [13] -1.24213179 -1.20008587 -1.12694106 -1.05213596 -1.02603976 -1.02133102
 [19] -0.98590201 -0.94432424 -0.93197011 -0.92907528 -0.92652735 -0.91377382
 [25] -0.90444338 -0.90118078 -0.84471359 -0.84443916 -0.80769821 -0.75108615
 [31] -0.74057986 -0.68905114 -0.66863188 -0.53769744 -0.49253338 -0.49098705
 [37] -0.46194700 -0.44662389 -0.44574891 -0.40259803 -0.35271715 -0.34973077
 [43] -0.21722945 -0.17554987 -0.17416975 -0.13131978 -0.08058516 -0.08028538
 [49] -0.01661600  0.01639520  0.04962701  0.07707743  0.10450690  0.13701769
 [55]  0.14759295  0.16442863  0.18587766  0.20088112  0.20847476  0.22740782
 [61]  0.30893409  0.44687767  0.45041816  0.47092395  0.48156957  0.49345912
 [67]  0.49606419  0.50178278  0.54506238  0.54556739  0.56798922  0.65041410
 [73]  0.72237934  0.72530236  0.73179586  0.75601810  0.81251957  0.81726919
 [79]  0.84045309  0.84668725  0.86686046  0.93475729  0.96138360  0.99377979
 [85]  1.00860093  1.02494443  1.03877227  1.04428168  1.20802666  1.24500979
 [91]  1.27398517  1.29533138  1.32427580  1.38077266  1.47040024  1.50033783
 [97]  1.58458201  1.61202978  1.62907528  2.18069509

# Plotting a histogram of the data
> postscript("histy.ps")
> hist(y)
> dev.off()

# Plotting the kernel density estimate of the data
> postscript("kdens.ps")
> d <- density(y)
> plot(d)
```

```
> dev.off()

# Determing the 1 and 2-sided p-values for a given quantile of the normal density
> pnorm(1.96,mean=0,sd=1)
[1] 0.9750021
> pnorm(1.645)
[1] 0.950015
> pnorm(1.96,lower.tail=FALSE)
[1] 0.02499790

# Determining p-value based on our simulation
> sum(sort(y)>1.96)/100
[1] 0.01
> sum(sort(abs(y))>1.96)/100
[1] 0.02

# Determining quantile based on alpha level
> qnorm(0.05)
[1] -1.644854
> qnorm(0.025)
[1] -1.959964

# Generating data from a central chi-square distribution with 1 d.f.
> z <- rchisq(100,df=1,ncp=0)
> sort(z)
  [1] 0.002048204 0.003072549 0.004346618 0.005647811 0.005720099 0.009924723
  [7] 0.011183900 0.013720423 0.017178320 0.018038249 0.023177999 0.023788501
 [13] 0.035781405 0.041095608 0.041922183 0.042268903 0.049436583 0.053171318
 [19] 0.056679650 0.057686002 0.060986913 0.061056146 0.063447509 0.070116520
 [25] 0.080560111 0.092825148 0.094776029 0.100120143 0.107496821 0.120582615
 [31] 0.130370445 0.144391901 0.176185342 0.177759035 0.184747385 0.193150146
 [37] 0.216652179 0.253571460 0.272744319 0.289773540 0.304056789 0.311123494
 [43] 0.387358885 0.399875881 0.407663212 0.413443179 0.417868405 0.450482325
 [49] 0.568759145 0.575315078 0.578377233 0.581684325 0.599035322 0.619210770
 [55] 0.625979017 0.691638015 0.711938921 0.740787290 0.748446265 0.764227808
 [61] 0.800950642 0.869531755 0.922785616 0.923210868 0.935517751 0.949999452
 [67] 0.957246214 1.053018552 1.079146476 1.092246818 1.126413971 1.144025531
```

```
 [73] 1.166079491 1.186530011 1.193458213 1.281009129 1.296192680 1.321850292
 [79] 1.382650102 1.391320240 1.428253314 1.497552383 1.532693308 1.607659159
 [85] 1.610683975 1.837818853 2.263741748 2.296067607 2.361217204 2.498717055
 [91] 2.539244569 2.609111708 2.921222928 3.072399920 3.105024880 3.259695121
 [97] 4.410998643 4.585548827 6.961171162 9.475800952

# Plotting density of z
> postscript("kdensz.ps")
> plot(density(z))
> dev.off()

# Plotting density from a non-central chi-square distribution
> postscript("kdensz2.ps")
> par(mfrow=c(2,2))
> plot(density(rchisq(500,df=1,ncp=.2)),xlab="ncp=.2",xlim=c(0,30),ylim=c(0,.6),main="")
> plot(density(rchisq(500,df=1,ncp=1)),xlab="ncp=1",xlim=c(0,30),ylim=c(0,.6),main="")
> plot(density(rchisq(500,df=1,ncp=2)),xlab="ncp=2",xlim=c(0,30),ylim=c(0,.6),main="")
> plot(density(rchisq(500,df=1,ncp=4)),xlab="ncp=4",xlim=c(0,30),ylim=c(0,.6),main="")
> dev.off()

# Determing p-value based for a given quantile
> pchisq(3.84,df=1)
[1] 0.9499565
> 1-pchisq(3.84,df=1)
[1] 0.05004352

# Determing p-value based on simulated data
> n <- 1000
> sum(rchisq(n,df=1)>3.84)/n
[1] 0.047

# Determining quantile for a given alpha -- note: need to enter 1-alpha
> qchisq(0.05,df=1)
[1] 0.00393214
> qchisq(0.95,df=1)
[1] 3.841459
```

```
# Simulating data from a central t-distribution with 2 d.f.
> sort(rt(100,df=2))
  [1] -6.246461762 -5.317224454 -4.531164926 -4.013382296 -3.593475750
  [6] -3.204597085 -3.179384553 -3.092508553 -2.928580481 -2.878289046
 [11] -2.771532911 -2.645335360 -2.264994127 -2.166634696 -2.011764516
 [16] -1.862491568 -1.649647139 -1.429138728 -1.275194502 -1.181035099
 [21] -1.136329168 -1.122737125 -1.117262515 -1.114338428 -1.018336300
 [26] -0.923179040 -0.836226167 -0.834977129 -0.829707392 -0.787124604
 [31] -0.772015830 -0.686251317 -0.659417862 -0.642090510 -0.634601172
 [36] -0.609144916 -0.599406425 -0.584614940 -0.539935050 -0.531136182
 [41] -0.420285445 -0.412296917 -0.405819852 -0.396164353 -0.383440893
 [46] -0.381616441 -0.360612612 -0.339735203 -0.332170546 -0.251955796
 [51] -0.238610579 -0.222041939 -0.180468479 -0.179519030 -0.173642595
 [56] -0.172425342 -0.160739788 -0.125710363 -0.092072712 -0.082759760
 [61] -0.009044007  0.044135917  0.052781865  0.085668389  0.151307672
 [66]  0.181307907  0.248851536  0.307217103  0.456290735  0.463244432
 [71]  0.530907512  0.556292116  0.558363987  0.569792133  0.576083043
 [76]  0.649849399  0.696181966  0.829130428  0.975534175  0.987076682
 [81]  0.995273843  1.006501414  1.114137911  1.145065576  1.162006371
 [86]  1.224393157  1.505457601  1.552928897  1.568869650  1.600674147
 [91]  1.625088758  1.834835946  1.933023032  1.968315919  2.130377958
 [96]  2.407508973  2.490672457  3.234760869  5.218534943  8.982601125

# Simulating data from a central F-distribution with df=(2,3)
> sort(rf(100,2,3))
  [1]  0.001228302  0.009621800  0.010747406  0.022579176  0.124469369
  [6]  0.125144249  0.132488251  0.134378779  0.135415599  0.156903413
 [11]  0.161863880  0.178158533  0.188767036  0.190652890  0.214303023
 [16]  0.245444030  0.259742733  0.279520462  0.285467527  0.286858659
 [21]  0.295895599  0.314610248  0.314613674  0.323596480  0.327502219
 [26]  0.328043574  0.334318782  0.347996478  0.360714019  0.388461927
 [31]  0.396380418  0.422209832  0.442343556  0.496892053  0.524252601
 [36]  0.548205023  0.622559862  0.649776788  0.664546045  0.667500328
 [41]  0.698160990  0.726124658  0.849576487  0.860215345  0.873134497
 [46]  0.903560076  0.938705110  0.980633582  1.056599252  1.074504111
 [51]  1.143767528  1.230344837  1.266528526  1.343770554  1.381751093
 [56]  1.393710810  1.400970431  1.430892842  1.449967085  1.496058941
```

```
[61]  1.505332087  1.515132140  1.625205893  1.634019336  1.643285923
[66]  1.681349426  1.754967864  1.808022764  1.920208365  1.975493975
[71]  2.050614758  2.166247803  2.290221120  2.303282595  2.321105995
[76]  2.365206342  2.432308343  2.475553358  2.765348959  2.862669262
[81]  2.952878867  2.959168702  3.077979040  3.119200066  3.277362654
[86]  3.855603259  4.293713033  4.401970656  5.738844767  5.917317852
[91]  6.769096253  7.139919289  7.177774652  7.416241410  8.656145543
[96]  9.387461008 11.969124197 12.752708337 15.149284430 37.254368858
```

Quadratic forms

Motivation (more in section on hypothesis testing): Consider the linear model $Y = X\beta + \epsilon$. where $\epsilon \sim MVN_n(0, \sigma^2 I)$. We are generally interested in testing a linear model against a reduced model. We call our starting model the "full model" and consider whether a simpler, more parsimonious model is reasonable. That is, we consider the "reduced model" given by $Y = X_0\gamma_0 + \epsilon$ where $\epsilon \sim MVN_n(0, \sigma^2 I)$ and $C(X_0) \in C(X)$. For example in the regression setting, we may have the full model given by

$$E(Y) = X_1\beta_1 + X_2\beta_2$$

and the reduced model given by

$$E(Y) = X_2\beta_2$$

Our null hypothesis might be that $E(Y) \in C(X_0)$ and the (disjoint) alternative is $E(Y) \in C(X)$ and $C(X) \notin C(X_0)$. If we let M and M_0 be the orthogonal projection operators onto $C(X)$ and $C(X_0)$ respectively, then under the full model the UMVUE of $E(Y)$ is MY and under the reduced model the UMVUE of $E(Y)$ is M_0Y.

Suppose the reduced model is true. This implies that MY and M_0Y are estimating the same quantity, $E(Y)$. On the other hand, if MY and M_0Y are different then they are not estimating the same quantity and the models must differ and so the reduced model is not correct. That is, the correctness of the reduced model depends on the quantity $(M - M_0)Y$. We measure this by the squared length given by $Y'(M - M_0)Y$.

Definition: Suppose $Y_{n\times 1}$ is a random vector and $A_{n\times n}$ is a matrix of constants. A quadratic form is a random variable given by $Y'AY$. We assume A is symmetric.

Theorem: Suppose $Y \sim MVN_n(0, \sigma^2 I)$. We have:

$$\frac{1}{\sigma^2}(Y'MY) \sim \chi^2_r$$

if and only if M is an orthogonal projection operator of rank r.

Theorem: Suppose $Y \sim MVN_n(\mu, \sigma^2 I)$. We have:

$$\frac{1}{\sigma^2}(Y'MY) \sim \chi^2_{r,\gamma}$$

if and only if M is an orthogonal projection operator of rank r and $\gamma = \frac{\mu' M \mu}{2\sigma^2}$.

Theorem: Suppose $Y \sim MVN_n(\mu, \sigma^2 I)$, then:

1. $Y'AY$ and BY are independent if and only if $AB' = 0$ where A is symmetric.
2. $Y'AY$ and $Y'BY$ are independent if and only if $AB = 0$ where A and B are symmetric.