Review of Linear Algebra

Trace: $tr(A+B) = tr(A) + tr(B)$ and $trace(AB) = trace(BA)$

Determinants: $|A^{-1}| = |A|^{-1}$. $|CA| = C^n \cdot |A|$, $|AB| = |A||B|$

Eigenvalues: If $Ax = \lambda x$ where $x \neq 0$, then λ is eigenvalue of A and x is a corresponding eigenvector.

Properties: for any symmetric matrix A with eigenvalues $\lambda_1, \cdots, \lambda_n$,

1. Spectral Decomposition: There is an orthonormal matrix T such that $T^T A T = \Lambda = diag(\lambda_1, \cdots, \lambda_n)$
2. $r(A) = \#$ of non-zero λ_i.
3. $tr(A) = \sum_i \lambda_i$
4. $|A| = \prod_i \lambda_i$

P.D. and p.s.d:

Def: A symmetric matrix A is called positive semi definite if $x^T A x \geq 0$ for All non-zero x.

properties:
1. Diagonal elements a_{ii} are all non-negative.
2. All eigenvalues of A are non-negative.
3. $tr(A) \geq 0$

$\Rightarrow$ for p.s.d. matrix

Def: A symmetric matrix A is called positive definite if $x^T A x > 0$ for all non-zero x.

properties: 1. $tr(A) > 0$ 2. a_{ii} are all positive 3. $|A| > 0$ 4. there exists non-singular R s.t. $A = RR^T$

5. A^{-1} is also p.d.

Idempotent and Projection Matrices:

Def: A matrix P is idempotent if $P^2 = P$. A symmetric and idempotent matrix is called a projection matrix.

properties:
1. Let P be a symmetric matrix, P is idempotent and of rank r iff it has r eigenvalues equal to 1 and $n-r$ eigenvalues equal to 0.
2. Projection matrices have $tr(P) = rank(P)$
3. Projection matrices are positive semidefinite.

Random Vectors and Matrices

properties:
1. $E(X+Y) = E(X) + E(Y)$
2. $E(AX) = AE(X)$
3. $E(AXB+C) = AE(X)B + C$

$\Rightarrow$ Vector Expectation.

properties:
1. Symmetry: $Cov(X) = [Cov(X)]^T$
2. $Cov(X+a) = Cov(X)$ for constant vector a.
3. $Cov(AX) = A\,cov(X)A^T$ for constant matrix A.

$\Rightarrow$ Vector Covariance Matrices.

4. $Cov(X)$ is p.s.d.
5. $Cov(X)$ is p.d. as long as no linear combination of X_i is constant.
6. $Cov(X) = E(XX^T) - E(X)E(X^T)$

Theorem: $Cov(AX, BY) = A\,cov(X,Y)B^T$ for constant matrices A and B.

Theorem: $E((X-\mu)^T A (X-\mu)) = tr(A\Sigma)$ if $E(X) = \mu$ and $Cov(X) = \Sigma$ and A is a constant matrix.

Corollary: $E(X^T A X) = tr(A\Sigma) + \mu^T A \mu$

Theorem: For $X \sim N(0, \Sigma)$ and constant matrices A and B, $X^T A X$ and BX are independently distributed IFF $B\Sigma A = 0$. Example: $\bar{x}$ and s^2.

Matrix Derivative: $\frac{\partial \beta^T A}{\partial \beta} = A$, $\frac{\partial \beta^T A \beta}{\partial \beta} = (A + A^T)\beta$

Multivariate Normal Distribution

1. Definition: Given a vector μ and p.s.d. matrix Σ. $y \sim N_n(\mu, \Sigma)$ if

① $f(y) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y-\mu)^T \Sigma^{-1}(y-\mu)\right)$

② MGF of Y is: $M_Y(t) = E(\exp(t^T y)) = \exp(\mu^T t + \frac{1}{2} t^T \Sigma t)$

③ Y has the same distribution as $AZ+\mu$, $Z := (Z_1, \cdots, Z_n)$ are iid $N(0,1)$ R.V. and $A_{n \times k}$ satisfies $AA^T = \Sigma$

2. Linear Transformations of MVN Vectors:

1. If $Y \sim N_n(\mu, \Sigma)$, $C_{p\times n}$ is a matrix of rank p, then $CY \sim N_p(C\mu, C\Sigma C^T)$
2. Y is MVN iff $a^T Y$ is normally distributed for all non-zero vectors a.

3. Orthogonal Transformations of MVN Vectors.

let $Y \sim N_n(\mu, \sigma^2 I)$, let $T_{n\times n}$ be an orthogonal matrix

①. Orthogonal matrices correspond to rotations and reflections about the origin. They preserve vector length.

②. If $Y \sim N_n(\mu, \sigma^2 I)$ and $w = T(Y-\mu)+\mu$ for some orthogonal transformation T, then $w \sim N_n(\mu, \sigma^2 I)$

3. χ^2 distribution: For any positive integer d, χ^2_d is the distribution of $\sum_{i=1}^d Z_i^2$, where $Z_1, \cdots Z_d$ are iid $N(0,1)$ random variables.

4. Conditional Distribution: $Y_1 | Y_2 = y_2 \sim N_p\left(\mu_1 + (\Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2)), \ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$

5. Multiple correlation coefficient: $\rho_{Y,X} = \operatorname{corr}(y, \hat{y}(x)) = \sqrt{\dfrac{\sigma_{YX}\Sigma_{XX}^{-1}\sigma_{XY}}{\sigma_y^2}}$. y is $n \times 1$

6. χ^2 distribution - continued:

1. Let $Y \sim N_n(\theta, \sigma^2 I)$ be $n\times n$ matrix of rank r. Then $Q \equiv (Y-\theta)^T P (Y-\theta)/\sigma^2 \sim \chi^2_r$ iff P is projection.
2. If $Y \sim N_n(\theta, \Sigma)$ and Σ is p.d, then $(Y-\theta)^T \Sigma^{-1}(Y-\theta) \sim \chi^2_n$
3. let $Y \sim N(0, \sigma^2 I)$, let A and B be symmetric matrices. Then Y^TAY and Y^TBY are independent if $AB=0$
4. Non-central χ^2: $\chi^2(\eta)$ is defined as the distribution of $\sum_{i=1}^n Z_i^2$ where $Z_1, \cdots, Z_n$ are ind. $N(\mu_i, 1)$. and $\eta = \sum_{i=1}^n \mu_i^2/2$

Facts: 1. If $Y \sim \chi^2_n(\eta)$, then $E(Y) = n + 2\eta$

2. If $Y \sim \chi^2_n(0)$, then $E(1/Y) = 1/(n-2)$

3. If $Y \sim N_n(\mu, \sigma^2 I)$ and P is projection matrix, $r(P) = r$, iff $Y^T P Y/\sigma^2 = \chi^2_r(\mu^T P \mu / 2\sigma^2)$

4. If $Y \sim N_n(\mu, \Sigma)$, A is a constant matrix with rank(r). then $Y^TAY \sim \chi^2_r(\mu^T A \mu/2)$ iff $A\Sigma$ is idempotent.

Theorem: let $Y \sim N_n(\theta, \sigma^2 I)$, $Q_1 = (Y-\theta)^T P_1 (Y-\theta)/\sigma^2$, $Q_2 = (Y-\theta)^T P_2 (Y-\theta)/\sigma^2$. P_1, P_2 are symmetric. If $Q_i \sim \chi^2_{r_i}$ and $Q_1 - Q_2 \geq 0$, then $Q_1 - Q_2$ and Q_2 are independent and $Q_1 - Q_2 \sim \chi^2_{r_1 - r_2}$

Full Rank LSE:

$\hat{\beta} = (X^TX)^{-1}X^TY$. $X(X^TX)^{-1}X^TY = \hat{Y}$ $\Rightarrow$ $P = X(X^TX)^{-1}X^T$ is hat matrix.

Lemma: 1. P and $I-P$ are projection matrices.

2. $r(I-P) = tr(I-P) = n-p$

3. $PX = X$, $(I-P)X = 0$

Less than full rank LSE: $r(X) < p$. $X(X^TX)^-X^T$, $(X^TX)^-$ is the generalised inverse.

1. P and $I-P$ are projection matrices
2. $r(I-P) = tr(I-P) = n-r$
3. $X^T(I-P) = 0$

Properties of LSE:

Assumption: 1. Errors are unbiased: $E(\varepsilon)=0$

2. Errors are uncorrelated with common variance: $cov(\varepsilon)=\sigma^2 I$.

Results of full rank case:

1. $E(\hat{\beta})=\beta$

2. $Cov(\beta)=(X^TX)^{-1}\sigma^2$

Theorem: (BLUE). Let $\hat{\theta}$ be the LSE of θ, for any linear combination $C^T\theta$, $C^T\hat{\theta}$ is the (unique) estimate with minimum variance among all linear unbiased estimates. We call $C^T\hat{\theta}$ the best linear unbiased estimate of $C^T\theta$.

Variance Estimation:

Let $rank(X)=r$, $S^2=(Y-X\hat{\beta})^T(Y-X\hat{\beta})/(n-r)=RSS/_{n-r}$ is an unbiased estimate of σ^2.

With normality assumption:

1. $\hat{\beta}\sim N_p(\beta,\sigma^2(X^TX)^{-1})$

2. $\hat{\beta}$ is independent from S^2

3. $\frac{RSS}{\sigma^2}=(n-p)S^2/\sigma^2\sim\chi^2_{n-p}$

4. $\frac{(\hat{\beta}-\beta)^TX^TX(\hat{\beta}-\beta)}{\sigma^2}\sim\chi^2_p$

Less than Full Rank LSE - Continued:

Generalized Inverse:

1. If $PX^TX=QX^TX$, then $PX^T=QX^T$

2. If G is a generalized inverse of X^TX, then G^T is also a generalized inverse of X^TX.

3. GX^TXG^T is a symmetric reflexive generalized inverse of X^TX

4. $X^TXGX^T=X^T$, $XGX^TX=X$

5. $XGX^T=XHX^T$ for any other generalized inverse H.

6. XGX^T is symmetric.

LSE: Method I: Reduce model to be full rank: $X=(X_1,X_2)=(X_1,X_1F)=X_1(I_{r\times r},F)$

Let $X_1=K$, $n\times r$, $(I_{r\times r},F)=L$, $r\times p$. $X=KL$.

$E(Y)=X\beta=KL\beta=K\alpha$, $\alpha=L\beta$

$\Rightarrow\hat{\alpha}=(K^TK)^{-1}K^TY\Rightarrow\hat{Y}=K\hat{\alpha}=K(K^TK)^{-1}K^TY$.

Method II: Impose Constraints for Identifiability.

$X\hat{\beta}=\hat{Y}$, $H\hat{\beta}=0\Rightarrow\begin{pmatrix}\hat{Y}\\0\end{pmatrix}=\begin{pmatrix}X\\H\end{pmatrix}\hat{\beta}=G\hat{\beta}$

Method III: Compute a generalized inverse.

$$(X^TX)^-=\begin{pmatrix}(X_1^TX_1)^{-1}&0\\0&0\end{pmatrix},$$ X_1 is r independent columns.

Estimable Functions:

For any linear combination $C^T\theta$, $C^T\hat{Y}$ is the BLUE of $C^T\theta$, where $\hat{Y}$ is the least squares orthogonal projection of Y onto $R(X)$.

If X is full rank, for $\forall a$, $a^T\hat{\beta}$ is the BLUE of $a^T\beta$.

Estimable Functions:

A linear combination $a^T\beta$ is estimable if it has a linear unbiased estimate.

Lemma: 1. $a^T\beta$ is estimable iff $a \in R(X^T)$

2. If $a^T\beta$ is estimable, there's a unique $b^* \in R(X)$, s.t. $a = X^T b^*$

Gauss-Markov Theorem: If $a^T\beta$ is estimable, then $a^T\hat{\beta}$ is unique, $a^T\hat{\beta}$ is the BLUE of $a^T\beta$.

Lemma: If $a^T\beta$ is estimable, then $a^T(X^TX)^-X^TX = a^T$ for any generalized inverse $(X^TX)^-$.

Also, if $a^T\beta$ is estimable, $var(a^T\hat{\beta}) = \sigma^2 \cdot a^T(X^TX)^- a$.

Generalized Least Squares:

The case when $var(Y) = \sigma^2 \cdot V$, $K^{-1}Y = K^{-1}X\beta + K^{-1}\varepsilon$, $(K^TK^T) = V$.

$$\Rightarrow \beta^* = (X^TV^{-1}X)^{-1}X^TV^{-1}Y.$$

Properties: 1. Unbiased: $E(\beta^*) = \beta$

2. $Var(\beta^*) = \sigma^2(X^TV^{-1}X)^{-1}$

3. $RSS = (Y - X\beta^*)^T V^{-1}(Y - X\beta^*)$

Thm: 1. If $E(Y) = X\beta$, $cov(Y) = \sigma^2 \cdot V$, $a^T\beta^*$ is the BLUE of $a^T\beta$ for $\forall a$, β^* is the GLS.

2. GLS and OLS are equal iff $R(V^{-1}X) = R(X)$