Appendix A

SURE estimator for the risk

We noted in chapter 2 that the risk $R$ depends on the unknown regression function $f$, and hence cannot be evaluated directly. It thus needs to be estimated. In this appendix, we drive a SURE estimator for the inverse-noise-weighted risk function (Eq. 2.34). SURE estimators originated in the context of what is known as the normal mean problem [53].

The simplest form of the normal means problem is as follows. We have series of observations $X_{ij}$ such that

$$X_{ij} = \theta_i + N(0, \sigma^2)$$  \hspace{1cm} (A.1)

That is each set of $X_{ij}$ for $j = 1,..,n$ are random variables with mean $\theta_i$ and variance $\sigma^2$. The aim is estimating the vector $\theta = (\theta_1, .., \theta_n)$ assuming that $\sigma^2$ is known [53].

From the central limit theorem, it could be argued that the average $Z_i = n^{-1} \sum_{j=1}^{n} X_{ij}$ has the following relation to unknown vector $\theta$

$$Z \sim N(\theta, \sigma_n)$$ \hspace{1cm} (A.2)

where $Z = (Z_1, ..., Z_n)$ and $\sigma_n = \sigma/\sqrt{n}$. 

108
Appendix A. SURE estimator for the risk

We thus see that the $Z$ is an unbiased estimator of $\theta$. It also happens to be maximum likelihood estimator for $\theta$. However it can be shown that when the number of unknown means ($n$) is more than two, then the risk $R = E[(Z - \theta)^2] = n\sigma^2$ for vector $Z$ is not minimum [119] [53]. In fact, $Z$ is a poor estimator of $\theta$ because its risk ($n\sigma^2_n$) which increases with $n$ [53]. The James-Stein estimator [120] is a better estimator of $\theta$ because it has the minimum risk. It contains a shrinkage factor which shrinks the $Z$ towards zero [53]. This result is often referred as Stein’s paradox [119].

The risk of $\hat{\theta}$ ($\theta$ estimator) can be estimated by Stein’s theorem [53]. In the rest of this section, we provide an alternative derivation of the SURE estimator for the inverse-noise-weighted risk (Eq. 2.34).

In chapter 2 we have shown

$$Z \sim N(\beta, B) \quad \text{(Eq. 2.19)} \quad \text{(A.3)}$$
$$\hat{\beta}_j = \lambda_j Z_j \quad \text{(Eq. 2.20)} \quad \text{(A.4)}$$

Therefore,

$$\hat{\beta} = DZ \quad \text{(A.5)}$$
$$E(\hat{\beta}) = D\beta \quad \Rightarrow \quad E(\hat{\beta}_j) = \lambda_j \beta_j \quad \text{(A.6)}$$
$$\text{Cov}(\hat{\beta}_j, \hat{\beta}_k) = \lambda_j \lambda_k B_{jk}, \quad \text{(A.7)}$$

where by definition the diagonal matrix $D \equiv \text{diag}(\lambda_0, \ldots, \lambda_{N-1})$ and hence $D_{ij} = \lambda_i \delta_{ij}$.

From the definition of covariance and Eq(s). (A.6) & (A.7), we have

$$E(\hat{\beta}_j\hat{\beta}_k) = \text{Cov}(\hat{\beta}_j, \hat{\beta}_k) + E(\hat{\beta}_j)E(\hat{\beta}_k) \quad \text{(A.8)}$$
$$= \lambda_j \lambda_k B_{jk} + \lambda_j \beta_j \lambda_k \beta_k. \quad \text{(A.9)}$$
The loss function can be written as

\[
L(f(x), \hat{f}(x)) \approx N^{-1} \sum_{j,k=0}^{N-1} (\hat{\beta}_j - \beta_j)(\hat{\beta}_k - \beta_k) W_{jk} \quad \text{(Eq. F.4) (A.10)}
\]

\[
= N^{-1} \sum_{j,k=0}^{N-1} (\beta_j \beta_k - \beta_j \hat{\beta}_k - \beta_k \hat{\beta}_j + \hat{\beta}_j \hat{\beta}_k) W_{jk}. \quad \text{(A.11)}
\]

The risk function which is expected value of loss, can be expanded as

\[
R = \mathbb{E}(L) = N^{-1} \sum_{j,k=0}^{N-1} \left( \beta_j \beta_k - \beta_j \mathbb{E}(\hat{\beta}_k) - \beta_k \mathbb{E}(\hat{\beta}_j) + \mathbb{E}(\hat{\beta}_j \hat{\beta}_k) \right) W_{jk} \quad \text{(A.12)}
\]

\[
= \beta^T W \beta - \beta^T W \mathbb{E}(\beta) - \beta \mathbb{E}(\beta^T \beta) + \mathbb{E}(\beta^T \beta) + \sum_{k=0}^{N-1} (DWDB)_{kk}
\]

\[
= \beta^T W (I - D) \beta - \beta^T DW (I - D) \beta + \text{tr}(DWDB)
\]

\[
= \beta^T W \bar{D} \beta - \beta^T DW \bar{D} \beta + \text{tr}(DWDB)
\]

\[
= (\beta^T W - \beta^T DW) \bar{D} \beta + \text{tr}(DWDB)
\]

\[
= \beta^T (I - D) W \bar{D} \beta + \text{tr}(DWDB)
\]

\[
= \beta^T \bar{D} W \bar{D} \beta + \text{tr}(DWDB)
\]

where \( \bar{D} \equiv I_N - D \) and \( I_N \) is the is the \( N \times N \) identity matrix.

Therefore, the risk expression can be written as

\[
R = \beta^T \bar{D} W \bar{D} \beta + \text{tr}(DWDB). \quad \text{(A.13)}
\]

As we can see, the first term depends on the coefficient vector \( \beta \), which is an unknown quantity. Therefore, to obtain a risk estimator, we need to find a good
estimator for this term. By expanding this term, we get

\[
(\beta^T \bar{D} W \bar{D} \beta)_{np} = \sum_{j,k,l,m=0}^{N-1} (\beta_{nj}^T (\bar{\lambda}_j \delta_{jl}) W_{lk} (\bar{\lambda}_k \delta_{km}) \beta_{mp}).
\] (A.14)

Since \( \beta \) is a vector, therefore \( n = p = 1 \), and hence the result of this expression would be a single scalar and not a matrix (which we expect since risk value is a scalar).

Therefore,

\[
\beta^T \bar{D} W \bar{D} \beta = \sum_{j,k=0}^{N-1} \beta_j \bar{\lambda}_j W_{jk} \bar{\lambda}_k \beta_k = \sum_{j,k=0}^{N-1} \beta_j \beta_k \bar{\lambda}_j \bar{\lambda}_k W_{jk}.
\] (A.15)

From the definition of covariance and Eq. A.3 we have

\[
\E(Z_j Z_k) = \E(Z_j) \E(Z_k) + \Cov(Z_j, Z_k) = \beta_j \beta_k + B_{jk},
\] (A.16)

and therefore

\[
\beta_j \beta_k = \E(Z_j Z_k - B_{jk}),
\] (A.18)

which shows that \((Z_j Z_k - B_{jk})\) is an unbiased estimator for \(\beta_j \beta_k\). By plugging this into A.15 we get
Appendix A. SURE estimator for the risk

\[ \beta^T \hat{D}W \hat{D} \beta = \sum_{j,k=0}^{N-1} (Z_j Z_k \bar{\lambda}_j \bar{\lambda}_k W_{jk} - B_{jk} \bar{\lambda}_j \bar{\lambda}_k W_{jk}) \]

\[ = \sum_{j,k,m,l=0}^{N-1} Z_j (\bar{\lambda}_j \delta_{jm}) W_{mk} (\bar{\lambda}_k \delta_{kl}) Z_l + \sum_{j,k=0}^{N-1} \bar{\lambda}_j W_{jk} \bar{\lambda}_k B_{kj} \]

\[ = \sum_{j,k,m,l=0}^{N-1} Z_j^T \bar{D}_{jm} W_{mk} \bar{D}_{kl} Z_l + \sum_{j,k=0}^{N-1} (\bar{D}W)_{jk} (\bar{D}B)_{kj} \]

\[ = Z^T \bar{D}W \bar{D}Z + \sum_{j=0}^{N-1} (\bar{D}W \bar{D}B)_{jj} \]

\[ = Z^T \bar{D}W \bar{D}Z + \text{tr}(\bar{D}W \bar{D}B) \] (A.19)

By substituting Eq. A.19 in risk expression (Eq. A.13), the Stein’s unbiased risk estimator [60] can be written as

\[ \hat{R}(\lambda) = Z^T \bar{D}W \bar{D}Z + \text{tr}(DWB) - \text{tr}(\bar{D}W \bar{D}B), \] (A.20)

Further manipulating the last two terms leads to

\[ \hat{R}(\lambda) = Z^T \bar{D}W \bar{D}Z + \text{tr}(DWB - \bar{D}W \bar{D}B) \]

\[ = Z^T \bar{D}W \bar{D}Z + \text{tr}((\bar{D}WD - \bar{D}W \bar{D})B) \]

\[ = Z^T \bar{D}W \bar{D}Z + \text{tr}((\bar{D}WD - (I - D)W(I - D))B) \]

\[ = Z^T \bar{D}W \bar{D}Z + \text{tr}((\bar{D}WD - (W - WD - DW + DWD))B) \]

\[ = Z^T \bar{D}W \bar{D}Z - \text{tr}((W - WD - DW)B) \] (A.21)

which is useful in deriving the expression for \( \tau^2 \) (Appendix B).
Appendix B

Estimating of $\hat{\tau}^2$

To arrive at an estimator $\hat{\tau}^2$ for $\tau^2$, we first rearrange the loss function as

$$L(f(x), \hat{f}(x)) \approx (\hat{\beta} - \beta)^T W (\hat{\beta} - \beta)$$

$$= \hat{\beta}^T W \hat{\beta} - \hat{\beta}^T W \beta - \beta^T W \hat{\beta} + \beta^T W \beta. \quad (B.1)$$

By substituting $\hat{\beta} = DZ$ (Eq. A.5) into the first term above, we get

$$\hat{\beta}^T W \hat{\beta} = (DZ)^T W (DZ) = Z^T DWDZ. \quad (B.2)$$

Here we used fact that $D$ is diagonal matrix and hence symmetric.

The second term in Eq. B.1 can be written as

$$-(\hat{\beta}^T W \beta) = -(DZ)^T W \beta = -Z^T DW \beta$$

$$-(Z^T DW \beta)_{im} = - \sum_{j,k,l=0}^{N-1} (Z_{ij}^T D_{jk} W_{kl} \beta_{lm}) = - \sum_{j,k,l=0}^{N-1} (Z_{ji} \lambda_j \delta_{jk} W_{kl} \beta_{lm}).$$

Since $Z, \beta$ are vectors, therefore $i = m = 1$ and the result of this expression would be a single scalar and not a matrix.
Therefore, for the second term we have

\[-Z^T DW \beta = - \sum_{j,k,l=0}^{N-1} (Z_j \lambda_j \delta_{jk} W_{kl} \beta_l) = - \sum_{j,l=0}^{N-1} (Z_j \lambda_j W_{jl} \beta_l). \tag{B.3}\]

The third term in Eq. B.1 would be

\[-\beta^T W \hat{\beta} = -\beta^T W (DZ) = -\beta^T W DZ\]
\[-(\beta^T W DZ)_{im} = - \sum_{j,k,l=0}^{N-1} (\beta_l W_{lk} \lambda_k \delta_{kj} Z_{jm}) = - \sum_{j,l=0}^{N-1} (\beta_l W_{lj} \lambda_j Z_j). \tag{B.4}\]

Again, since $Z, \beta$ are vectors, therefore $i = m = 1$ and the result of this expression would be a single scalar and not a matrix. Therefore for the third term we have

\[-\beta^T W DZ = - \sum_{j,k,l=0}^{N-1} (\beta_l W_{lk} \lambda_k \delta_{kj} Z_j) = - \sum_{j,l=0}^{N-1} (\beta_l W_{lj} \lambda_j Z_j). \tag{B.4}\]

From equations (B.3), (B.4) and symmetry of $W$, it follows that the second and third term are equal. Therefore we rewrite the loss function expression (B.1) as

\[L = Z^T DW DZ - 2Z^T DW \beta + \beta^T W \beta \tag{B.5}\]

\[\tau^2 \text{ expression}\]

$\tau^2$ is defined as the variance of $\sqrt{N}(L - \hat{R})$ (Eq. 2.40). Therefore we first need to simplify the expression for $(L - \hat{R})$. From equations B.5 & A.21, we have

\[L - \hat{R} = Z^T DW DZ - 2Z^T DW \beta + \beta^T W \beta - Z^T DW DZ + \text{tr}((W - WD - DW)B) \tag{B.6}\]
Let us expand and simplify the forth term:

\[-Z^T \tilde{D}W \tilde{D}Z = -Z^T (I - D)W(I - D)Z\]
\[-Z^T WZ + Z^T WDWZ + Z^T DWZ = Z^T (DW + WD - W)Z - Z^T DWZ.\]

(B.7)

If we substitute Eq. B.7 in Eq. B.6 we get

\[L - \hat{R} = \beta^T W \beta - 2Z^T DW \beta + Z^T (DW + WD - W)Z + \text{tr}((W - WD - DW)B).\]

(B.8)

To derive \(\tau^2\) (and then \(\hat{\tau}^2\)), we must find the variance of \((L - \hat{R})\). In Eq. B.8, the first and the last term are not random variables. Hence their variance is zero.

Therefore,

\[
\text{Var}(L - \hat{R}) = \text{Var}(Z^T (DW + WD - W)Z - 2Z^T DW \beta)
\]

\[
= \text{Var}(Z^T AZ - Z^T C)
\]

\[
= \text{Var}(Z^T AZ) + \text{Var}(Z^T C) - 2\text{Cov}(Z^T AZ, Z^T C),\]

(B.9)

where \(A = DW + WD - W\) and \(C = 2DW \beta\).

Let us deal with Eq. B.9 term by term.

The first term:

\[
\text{Var}(Z^T AZ) = \text{Var}\left(\sum_{i,j=0}^{N-1} A_{ij}Z_i Z_j\right) = \sum_{i,j,k,l=0}^{N-1} (A_{ij} A_{kl} \text{Cov}(Z_i Z_j, Z_k Z_l)).\]

(B.10)

From Eq. D.19, we know

\[
\text{Cov}(Z_i Z_j, Z_k Z_l) = \beta_i \beta_j B_{jk} + \beta_i \beta_k B_{ik} + \beta_j \beta_k B_{ij} + \beta_j \beta_k B_{il} + B_{ik} B_{kl} + B_{il} B_{jk}.\]

(B.11)
Appendix B. Estimating of $\widehat{\tau^2}$

Substituting B.11 in B.10 we have

$$\text{Var}(Z^T AZ) = \sum_{i,j,k,l=0}^{N-1} (A_{ij} A_{kl} (\beta_i \beta_l B_{jk} + \beta_i \beta_j B_{ik} + \beta_j \beta_k B_{il} + B_{ik} B_{jl} + B_{il} B_{jk}))$$

(B.12)

If we rearrange the indices and use the symmetry of $A$ and $B$, we see that the first four terms are equal to one another, and so are the last two terms. Therefore,

$$\text{Var}(Z^T AZ) = \sum_{i,j,k,l=0}^{N-1} (4A_{ij} A_{kl} \beta_i \beta_k B_{jl} + 2A_{ij} A_{kl} B_{ik} B_{jl})$$

$$= 4\beta^T ABA \beta + 2\text{tr}(ABAB).$$

(B.13)

The second term:

$$\text{Var}(Z^T C) = \text{Var}\left(\sum_{i=0}^{N-1} C_i Z_i\right) = \sum_{i,j=0}^{N-1} C_i C_j \text{Cov}(Z_i, Z_j) = \sum_{i,j=0}^{N-1} C_i C_j B_{ij}$$

$$= C^T B C.$$ 

(B.14)

The third term:

$$-2 \text{Cov}(Z^T AZ, Z^T C) = -2 \text{Cov}\left(\sum_{j,k=0}^{N-1} A_{jk} Z_j Z_k, \sum_{i=0}^{N-1} C_i Z_i\right)$$

$$= -2 \sum_{i,j,k=0}^{N-1} C_i A_{jk} \text{Cov}(Z_j Z_k, Z_i).$$

(B.15)

From Eq. D.19, we conclude that

$$\text{Cov}(Z_j Z_k, Z_i) = \beta_j B_{ki} + \beta_k B_{ji}.$$ 

(B.16)
If we substitute B.16 into B.15, we have

\[-2 \text{Cov}(Z^T AZ, Z^T C) = -2 \sum_{i,j,k=0}^{N-1} C_i A_{jk}(\beta_j B_{ki} + \beta_k B_{ji})\]

\[= -2 \sum_{i,j,k=0}^{N-1} (C_i A_{jk}\beta_j B_{ki} + C_i A_{jk}\beta_k B_{ji}) \quad (B.17)\]

By rearranging the indices and using the symmetry of the $A, B$ matrices, we can see that both the terms in Eq. B.17 are equal. Therefore,

\[-2 \text{Cov}(Z^T AZ, Z^T C) = -4 \beta^T ABC. \quad (B.18)\]

If we substitute B.13, B.14 and B.18 into B.9 we get

\[\text{Var}(L - \hat{R}) = 4 \beta^T ABA\beta + 2\text{tr}(ABAB) + C^T BC - 4 \beta^T ABC. \quad (B.19)\]

We substitute $C = 2DW\beta$ and rearrange terms; then:

\[\text{Var}(L - \hat{R}) = 2\text{tr}(ABAB) + 4 \beta^T ABA\beta + (2DW\beta)^T B(2DW\beta) - 4 \beta^T AB(2DW\beta)\]

\[= 2\text{tr}(ABAB) + 4 \beta^T ABA\beta + 4 \beta^T WDBDW\beta - 8 \beta^T ABDW\beta\]

\[= 2\text{tr}(ABAB) + 4 \beta^T (ABA + WDBDW - 2ABDW)\beta\]

\[= 2\text{tr}(ABAB) + \beta^T Q\beta, \quad (B.20)\]

where $Q = 4(ABA + WDBDW - 2ABDW)$.

Writing $\tau^2$ as

\[\tau^2 = \text{Var}(\sqrt{N}(L - \hat{R})) = N\text{Var}(L - \hat{R}),\]

we conclude that

\[\tau^2/N = 2\text{tr}(ABAB) + \beta^T Q\beta \quad (B.21)\]
\[ \hat{\tau}^2, \text{ an estimator for } \tau^2 \]

The expression of \( \tau^2 \) (Eq. B.21) depends on \( \beta \) which is unknown. Therefore we need to find a good estimator for \( \tau^2 \). We first expand the second term of Eq. B.21 as

\[
\frac{\tau^2}{N} = 2\text{tr}(ABAB) + \sum_{i,j=0}^{N-1} \beta_i \beta_j Q_{ij}(Z_i Z_j - B_{ij})
\]

(B.22)

By plugging \( \beta_i \beta_j \) estimator from Eq. A.18 the \( \tau \) estimator can be written as

\[
\hat{\tau}^2/N = 2\text{tr}(ABAB) + \sum_{i,j=0}^{N-1} Q_{ij}(Z_i Z_j - B_{ij})
= 2\text{tr}(ABAB) + \sum_{i,j=0}^{N-1} Q_{ij}Z_i Z_j - Q_{ij}B_{ij})
\]

Which can be expanded in a compact form of

\[
\hat{\tau}^2/N = 2\text{tr}(ABAB) + Z^T Q Z - \text{tr}(QB)
\]
Appendix C

Canonical quadratic form of the SURE risk estimator

The risk estimator $\hat{R}$ (Eq. 2.35) is clearly quadratic. However, it is not in the canonical quadratic form, $X^TAX$. To obtain a canonical quadratic form of $\hat{R}$ for minimization purposes, we first find the gradient of $\hat{R}$ with respect to $\lambda$ and then integrate it back to get the canonical quadratic form. Any constants that eliminated in the process are irrelevant as far as the minimization of $\hat{R}$ is concerned.

To find the gradient of the risk estimator, we start from risk estimator (Eq. 2.35):

$$\hat{R}(\lambda) = Z^T \bar{D}W \bar{D}Z + \text{tr}(DWDB) - \text{tr}(\bar{D}WDB).$$  \hspace{1cm} (C.1)

We first expand the first term as

$$(Z^T \bar{D}W \bar{D}Z)_{pn} = \sum_{j,k,l,m=0}^{N-1} (Z^T_{pj} \bar{D}_{jk} W_{kl} \bar{D}_{lm} Z_{mn}).$$  \hspace{1cm} (C.2)
Appendix C. Canonical quadratic form of the SURE risk estimator

Since the $Z$ is a vector, therefore $p = n = 1$ hence

$$Z^T \bar{D}W \bar{D}Z = \sum_{j,k,l,m=0}^{N-1} (Z_j \bar{D}_{jk} W_{kl} \bar{D}_{lm} Z_m) = \sum_{j,k,l,m=0}^{N-1} (Z_j \bar{\lambda}_j \delta_{jk} W_{kl} \bar{\lambda}_l \delta_{lm} Z_m) = \sum_{j,l=0}^{N-1} (Z_j \bar{\lambda}_j W_{jl} \bar{\lambda}_l Z_l). \tag{C.3}$$

Here, we define $\bar{D}_{ij} = \bar{\lambda}_i \delta_{ij}$, where $\bar{\lambda}_i = 1 - \lambda_i$.

From definition of $\bar{\lambda}_i$ it turns out

$$\frac{\partial}{\partial \lambda_i} = \sum_{j=0}^{N-1} \frac{\partial \bar{\lambda}_j}{\partial \lambda_i} \frac{\partial}{\partial \lambda_j} = \frac{\partial \bar{\lambda}_i}{\partial \lambda_i} \frac{\partial}{\partial \lambda_i} = - \frac{\partial}{\partial \lambda_i}. \tag{C.4}$$

Therefore from C.3 and C.4, we get the partial derivative of the first term to be

$$\frac{\partial}{\partial \lambda_i} (Z^T \bar{D}W \bar{D}Z) = - \frac{\partial}{\partial \lambda_i} (Z^T \bar{D}W \bar{D}Z) = - \sum_{j=0}^{N-1} Z_j \bar{\lambda}_j W_{ji} Z_i - \sum_{l=0}^{N-1} Z_l W_{il} \bar{\lambda}_l Z_l = -2 \sum_{j=0}^{N-1} Z_j \bar{\lambda}_j W_{ji} Z_i = -(H \bar{\lambda})_i = - (H(1_N - \lambda))_i = (H 1_N + H\lambda)_i, \tag{C.5}$$

where $H_{jk} = 2Z_j W_{jk} Z_k$, $\bar{\lambda} = 1_N - \lambda$ and $1_N \equiv (1, 1, ..., 1)$.

For the second term, we have

$$\text{tr}(DWDB) = \sum_{m=0}^{N-1} (DWDB)_{mm} = \sum_{m=0}^{N-1} \left( \sum_{j,k,l=0}^{N-1} (D_{mj} W_{jk} D_{kl} B_{lm}) \right) = \sum_{m,j,k,l=0}^{N-1} \lambda_m \delta_{mj} W_{jk} \lambda_l \delta_{kl} B_{lm} = \sum_{m,l=0}^{N-1} \lambda_m W_{ml} \lambda_l B_{lm}. \tag{C.6}$$
Therefore

\[
\frac{\partial}{\partial \lambda_i} \left[ \text{tr}(DWDB) \right] = \sum_{l=0}^{N-1} W_{il} \lambda_l B_{li} + \sum_{m=0}^{N-1} \lambda_m W_{mi} B_{im}
\]

\[
= 2 \sum_{l=0}^{N-1} W_{il} \bar{\lambda}_l
\]

\[
= \sum_{l=0}^{N-1} V_{il} \bar{\lambda}_l
\]

\[
= (V\lambda)_i
\]

(C.7)

where \( V_{il} = 2W_{il} B_{li} \).

For the third term, we have

\[
- \text{tr}(\tilde{D}W\tilde{D}B) = - \sum_{m=0}^{N-1} (\tilde{D}W\tilde{D}B)_{mm} = - \sum_{m=0}^{N-1} \left( \sum_{j,k,l=0}^{N-1} (\tilde{D}_{mj} W_{jk} \tilde{D}_{kl} B_{lm}) \right)
\]

\[
= - \sum_{m,j,k,l=0}^{N-1} \bar{\lambda}_m \delta_{mj} W_{jk} \bar{\lambda}_l \delta_{kl} B_{lm} = - \sum_{m,l=0}^{N-1} \bar{\lambda}_m W_{ml} \bar{\lambda}_l B_{lm}
\]

(C.8)

Then

\[
\frac{\partial}{\partial \lambda_i} \left[ - \text{tr}(\tilde{D}W\tilde{D}B) \right] = - \frac{\partial}{\partial \lambda_i} \left[ - \text{tr}(\tilde{D}W\tilde{D}B) \right]
\]

\[
= \sum_{l=0}^{N-1} W_{il} \bar{\lambda}_l B_{li} + \sum_{m=0}^{N-1} \bar{\lambda}_m W_{mi} B_{im}
\]

\[
= 2 \sum_{l=0}^{N-1} W_{il} \bar{\lambda}_l
\]

\[
= \sum_{l=0}^{N-1} V_{il} \bar{\lambda}_l
\]

\[
= (V\lambda)_i
\]

\[
= (V(1_N - \lambda))_i
\]

\[
= (V1_N - V\lambda)_i
\]

(C.9)
Putting every thing together we get

\[
\frac{\partial}{\partial \lambda_i} \hat{R} = (-H1_N + H\lambda)_i + (V\lambda)_i + (V1_N - V\lambda)_i. \\
= (H\lambda + V1_N - H1_N)_i
\]

(C.10)

Hence, the gradient of risk estimator $\nabla \hat{R}(\lambda)$ can be seen to have the simple form

$$\nabla \hat{R}(\lambda) = H\lambda - h,$$

where $h = (H - V)(1, 1, \ldots, 1)^T$.

Integrating this back, one gets a canonical quadratic form for the $\hat{R}$, disregarding terms that do not depend on $\lambda$ as

$$\hat{R}(\lambda) = \frac{1}{2} \lambda^T H\lambda - \lambda^T h,$$

(C.11)
Appendix D

Important formulae (with proofs)

D.1 \( \text{Var} \left( \sum_{i=0}^{N-1} a_i X_i \right) \)

Here we derive an expression for the variance of a linear combination of \( n \) random variables, i.e.,

\[
\text{Var} \left( \sum_{i=0}^{N-1} a_i X_i \right), \tag{D.1}
\]

where \( a_i \)s are constant and \( X_i \)s are random numbers.

From the definition of variance, we have

\[
\text{Var} \left( \sum_{i=0}^{N-1} a_i X_i \right) = \mathbb{E} \left[ \left( \sum_{i=0}^{N-1} a_i X_i - \left( \sum_{i=0}^{N-1} a_i X_i \right) \right)^2 \right]
\]

\[
= \mathbb{E} \left[ \sum_{i,j=0}^{N-1} \left( (a_i X_i - \langle a_i X_i \rangle) (a_j X_j - \langle a_j X_j \rangle) \right) \right]
\]

\[
= \sum_{i,j=0}^{N-1} a_i a_j \mathbb{E} \left[ \left( X_i - \langle X_i \rangle \right) \left( X_j - \langle X_j \rangle \right) \right]
\]

\[
= \sum_{i,j=0}^{N-1} a_i a_j \text{Cov}(X_i, X_j) \tag{D.2}
\]

\[\blacksquare\]
D.2 Cov\(\left(\sum_{i=0}^{N-1} a_i X_i, \sum_{j=0}^{N-1} b_i Y_i\right)\)

Here, we show to derive the covariance between linear combinations of two different sets of random variables, i.e.,

\[
\text{Cov}\left(\sum_{i=0}^{N-1} a_i X_i, \sum_{j=0}^{N-1} b_i Y_i\right),
\]

(D.3)

where \(a_i\)'s, \(b_i\)'s are constant and \(X_i\)'s and \(Y_i\)'s are random numbers.

Starting from the definition of covariance, we get

\[
\begin{align*}
\text{Cov}\left(\sum_{i=0}^{N-1} a_i X_i, \sum_{j=0}^{N-1} b_i Y_i\right) &= \mathbb{E}\left[\left(\sum_{i=0}^{N-1} a_i X_i - \langle \sum_{i=0}^{N-1} a_i X_i \rangle\right)\left(\sum_{j=0}^{N-1} b_j Y_j - \langle \sum_{j=0}^{N-1} b_j Y_j \rangle\right)\right] \\
&= \sum_{i,j=0}^{N-1} a_i b_j \mathbb{E}\left[(X_i - \langle X_i \rangle)(Y_j - \langle Y_j \rangle)\right] \\
&= \sum_{i,j=0}^{N-1} a_i b_j \text{Cov}(X_i, Y_j).
\end{align*}
\]

(D.4)

\[\blacksquare\]

D.3 \(\mathbb{E}[Y_i Y_j Y_k Y_l]\)

The data model of Eq. 2.9 and Eq. 2.16 assumes that the relationship between the dependent variable \(Y\), the regression function \(F\) & noise is of the form

\[
Y = F + \varepsilon
\]

(D.5)

where \(F \equiv (f(x_0), \ldots, f(x_{N-1}))^T\) is the vector of truth function values, and \(\varepsilon \equiv (\epsilon_0, \ldots, \epsilon_{N-1})^T\) is a mean-0 normal random noise vector. Here, we want to calculate the expected value of product of four different components of \(Y\), i.e.,

\[
\mathbb{E}[Y_i Y_j Y_k Y_l].
\]

(D.6)
First, we substitute Eq. D.5 into D.6, then we get

\[
\mathbb{E}[Y_i Y_j Y_k Y_l] = \mathbb{E}[F_i F_j F_k F_l + F_i F_j F_k \varepsilon_k + F_i F_k F_j \varepsilon_j + F_i F_k F_l \varepsilon_l + F_j F_k F_l \varepsilon_i + F_j F_k \varepsilon_i \varepsilon_k + F_j F_k \varepsilon_i \varepsilon_l + F_j F_k \varepsilon_j \varepsilon_l + F_j F_k \varepsilon_j \varepsilon_l + F_j F_k \varepsilon_l \varepsilon_i + F_j F_k \varepsilon_l \varepsilon_j + F_j F_k \varepsilon_l \varepsilon_k + \varepsilon_i \varepsilon_j \varepsilon_k \varepsilon_l]
\]

Here, \(F_i, F_j, F_k, F_l\) are fixed values and not random variables. Therefore,

\[
\mathbb{E}[Y_i Y_j Y_k Y_l] = F_i F_j F_k F_l + F_i F_k F_j \varepsilon_k + F_i F_k \varepsilon_l \varepsilon_j + F_i F_l \varepsilon_i \varepsilon_j + F_i \varepsilon_k \varepsilon_l + F_j F_k \varepsilon_i \varepsilon_k + F_j F_k \varepsilon_i \varepsilon_l + F_j F_k \varepsilon_j \varepsilon_l + F_j \varepsilon_k \varepsilon_l + F_k \varepsilon_i \varepsilon_k + F_k \varepsilon_i \varepsilon_l + F_k \varepsilon_j \varepsilon_k + F_k \varepsilon_j \varepsilon_l + \varepsilon_i \varepsilon_j \varepsilon_k \varepsilon_l
\]

According to Isserlis’ theorem [121, 122], all the even-order moments of a mean-0 multivariate normal random vector can be written as a sum of the product of covariances, and all the odd moments are zero identically. Hence,

\[
\begin{align*}
\mathbb{E}[\varepsilon_i] &= 0 \\
\mathbb{E}[\varepsilon_i \varepsilon_j] &= \Sigma_{ij} \\
\mathbb{E}[\varepsilon_i \varepsilon_j \varepsilon_k] &= 0 \\
\mathbb{E}[\varepsilon_i \varepsilon_j \varepsilon_k \varepsilon_l] &= \Sigma_{ij} \Sigma_{kl} + \Sigma_{ik} \Sigma_{jl} + \Sigma_{il} \Sigma_{jk}
\end{align*}
\]

Therefore,

\[
\mathbb{E}[Y_i Y_j Y_k Y_l] = F_i F_j F_k F_l + F_i F_l \Sigma_{ij} + F_j F_l \Sigma_{ik} + F_k F_l \Sigma_{ij} + F_j F_l \Sigma_{ik} + F_i F_k \Sigma_{il} + F_j F_k \Sigma_{il} + \Sigma_{ij} \Sigma_{kl} + \Sigma_{ik} \Sigma_{jl} + \Sigma_{il} \Sigma_{jk} \quad (D.7)
\]
Appendix D. Important formulae (with proofs)

D.4 \( \mathbb{E}[Z_i Z_j Z_k Z_l] \)

We have defined \( Z \) in Eq. 2.13, 2.15 as

\[
Z = \frac{1}{\sqrt{N}} U^T Y
\]
\[
Z_i = \frac{1}{\sqrt{N}} \sum_{i'=0}^{N-1} U_{ii'} Y_{i'}
\] \hspace{1cm} (D.8)

and \( Z \) is normally distributed as

\[
Z \sim N(\beta, B) \quad \text{(Eq. 2.19)} \] \hspace{1cm} (D.9)

where \( B = U^T \Sigma U / N \) is the covariance of \( Z \).

Now we want to calculate the expected value of a product of four components of \( Z \), i.e.,

\[
\mathbb{E}[Z_i Z_j Z_k Z_l]
\] \hspace{1cm} (D.10)
Appendix D. Important formulae (with proofs)

Substituting Eq. D.8 in D.10 and using D.7, we get

\[
\mathbb{E}[Z_i Z_j Z_k Z_l] = \left( \frac{1}{\sqrt{N}} \right)^4 \sum_{i'j'k'l'=0}^{N-1} \left( U_{ii'}^T U_{jj'}^T U_{kk'}^T U_{ll'}^T \mathbb{E}[Y_i Y_j Y_k Y_l] \right) \\
= \left( \frac{1}{\sqrt{N}} \right)^4 \sum_{i'j'k'l'=0}^{N-1} \left( U_{ii'}^T U_{jj'}^T U_{kk'}^T U_{ll'}^T \left( F_{i'i'} F_{j'j'} F_{k'k'} F_{l'l'} + F_{i'i'} F_{k'k'} \Sigma_{j'j'} + F_{i'i'} F_{l'l'} \Sigma_{k'k'} + \Sigma_{i'j'} \Sigma_{k'l'} \right) \right) \\
+ F_{i'i'} F_{k'k'} \Sigma_{j'j'} + F_{i'i'} F_{l'l'} \Sigma_{k'k'} + F_{j'j'} F_{k'k'} \Sigma_{i'i'} + \Sigma_{i'j'} \Sigma_{k'l'} \\
+ \Sigma_{i'k'} \Sigma_{j'l'} + \Sigma_{i'l'} \Sigma_{j'k'} \right) \\
= \sum_{i'j'k'l'=0}^{N-1} \left( \frac{1}{\sqrt{N}} U_{ii'}^T F_{i'} \right) \left( \frac{1}{\sqrt{N}} U_{jj'}^T F_{j'} \right) \left( \frac{1}{\sqrt{N}} U_{kk'}^T F_{k'} \right) \left( \frac{1}{\sqrt{N}} U_{ll'}^T F_{l'} \right) \\
+ \sum_{i'j'k'l'=0}^{N-1} \left( \frac{1}{\sqrt{N}} U_{ii'}^T F_{i'} \right) \left( \frac{1}{\sqrt{N}} U_{jj'}^T F_{j'} \right) \left( \frac{1}{\sqrt{N}} U_{kk'}^T F_{k'} \right) \left( \frac{1}{\sqrt{N}} U_{ll'}^T F_{l'} \right) \left( \frac{1}{\sqrt{N}} U_{ii'}^T U_{jj'}^T \Sigma_{i'j'} \right) \\
+ \sum_{i'j'k'l'=0}^{N-1} \left( \frac{1}{\sqrt{N}} U_{ii'}^T F_{i'} \right) \left( \frac{1}{\sqrt{N}} U_{jj'}^T F_{j'} \right) \left( \frac{1}{\sqrt{N}} U_{kk'}^T F_{k'} \right) \left( \frac{1}{\sqrt{N}} U_{ll'}^T \Sigma_{i'j'} \right) \\
+ \sum_{i'j'k'l'=0}^{N-1} \left( \frac{1}{\sqrt{N}} U_{ii'}^T F_{i'} \right) \left( \frac{1}{\sqrt{N}} U_{jj'}^T F_{j'} \right) \left( \frac{1}{\sqrt{N}} U_{kk'}^T F_{k'} \right) \left( \frac{1}{\sqrt{N}} U_{ll'}^T \Sigma_{i'j'} \right) \\
+ \sum_{i'j'k'l'=0}^{N-1} \left( \frac{1}{\sqrt{N}} U_{ii'}^T F_{i'} \right) \left( \frac{1}{\sqrt{N}} U_{jj'}^T F_{j'} \right) \left( \frac{1}{\sqrt{N}} U_{kk'}^T F_{k'} \right) \left( \frac{1}{\sqrt{N}} U_{ll'}^T \Sigma_{i'j'} \right) \\
+ \sum_{i'j'k'l'=0}^{N-1} \left( \frac{1}{\sqrt{N}} U_{ii'}^T U_{jj'}^T \Sigma_{i'j'} \right) \left( \frac{1}{\sqrt{N}} U_{kk'}^T \Sigma_{k'l'} \right) \\
+ \sum_{i'j'k'l'=0}^{N-1} \left( \frac{1}{\sqrt{N}} U_{ii'}^T U_{kk'}^T \Sigma_{i'k'} \right) \left( \frac{1}{\sqrt{N}} U_{jj'}^T U_{ll'}^T \Sigma_{j'j'} \right) \\
+ \sum_{i'j'k'l'=0}^{N-1} \left( \frac{1}{\sqrt{N}} U_{ii'}^T U_{ll'}^T \Sigma_{i'l'} \right) \left( \frac{1}{\sqrt{N}} U_{jj'}^T U_{kk'}^T \Sigma_{j'k'} \right) \right) \\
\tag{D.11}
From Eq. 2.18 and D.9 we know that

\[ F = \sqrt{NU\beta} \Rightarrow \beta = \frac{1}{\sqrt{N}}U^TF \Rightarrow \beta_i = \frac{1}{\sqrt{N}}U_{ii}'F_i' \quad (D.12) \]

\[ B = \frac{1}{N}U^T\Sigma U \Rightarrow B_{ij} = \frac{1}{N}U_{ii}'\Sigma_{i'j'}U_{j'j} = \frac{1}{N}U_{ii}'\Sigma_{i'j'}U_{j'j}' \quad (D.13) \]

Now, by considering D.12, D.13, the expression in D.11 can be written as

\[
E[Z_iZ_jZ_kZ_l] = \beta_i\beta_j\beta_k\beta_l + \beta_i\beta_kB_{jk} + \beta_j\beta_lB_{ik} + \beta_k\beta_lB_{il} + \beta_i\beta_B_{lk} + \beta_j\beta_B_{ik} + \beta_k\beta_B_{lj} + B_{ij}B_{kl} + B_{ik}B_{jl} + B_{il}B_{jk} \quad (D.14)
\]

\[\blacksquare\]

### D.5 \textbf{Cov}(Z_iZ_j, Z_kZ_l)

We want to calculate the covariance of product of two components of \(Z\) variable; i.e.,

\[
\text{Cov}(Z_iZ_j, Z_kZ_l) \quad (D.15)
\]

From the definition of covariance, we have

\[
\text{Cov}(Z_iZ_j, Z_kZ_l) = E[Z_iZ_jZ_kZ_l] - E[Z_iZ_j]E[Z_kZ_l] \quad (D.16)
\]

From D.14 we conclude that

\[
E[Z_iZ_j] = \beta_i\beta_j + B_{ij} \quad (D.17)
\]

\[
E[Z_kZ_l] = \beta_k\beta_l + B_{kl} \quad (D.18)
\]

By substituting D.14, D.17 and D.18 in D.16, we get

\[
\text{Cov}(Z_iZ_j, Z_kZ_l) = \beta_i\beta_lB_{jk} + \beta_i\beta_lB_{ik} + \beta_i\beta_kB_{ij} + \beta_j\beta_kB_{il} + B_{ik}B_{jl} + B_{il}B_{jk} \quad (D.19)
\]
Appendix E

The distribution of angular power spectrum of cosmic microwave background

E.1 Introduction

The angular power spectrum contains all the information about CMB anisotropy and its measurement is an important step in CMB analysis. In principle, the $C_l$ can be measured from $a_{lm}$’s provided different realization of CMB anisotropy is available. In reality we have only one CMB map therefore each $C_l$ is estimated by average value of $|a_{lm}|^2$. The estimated angular power spectrum $C_l^{est}$ is a random quantity and depends on CMB realization. Finding the distribution of $C_l^{est}$ is useful in several aspects such as angular power spectrum and cosmological parameter estimation and to simulate the angular power spectrum data.

Knox [109], argued that the $C_l^{est}$ has $\chi^2$ distributed with degree of freedom $2l + 1$. We start from his proof which shows $C_l^{est}$ is a linear transformation of the $\chi^2$ distributed random number. But we show that the linear transformation of $\chi^2$ distributed random numbers is Amoroso distribution. Hence, we show that the distribution of angular power spectrum of CMB is also a special case of Amoroso
distribution which is wrongly indicated as $\chi^2$ distribution in some literature. Basically the difference is in their mean value and this special case of Amoroso distribution has a shape equivalent to $\chi^2$ distribution with mean value at $C_l$ whereas $\chi^2_{2l+1}$ has the mean at $2l + 1$. For completeness we also show the linear transformation of $\Gamma$ distribution is a kind of Amoroso distribution.

E.2 $C_l^{est}$ and its relation to $\chi^2$ distributed random numbers

We can show the measured/estimated angular power spectrum $C_l^{est}$, has the relation of ([109], Appendix. Eq. A.9)

$$V = (2l + 1) \frac{C_l^{est} + w^{-1}e^{2\sigma_b^2}}{C_l + w^{-1}e^{2\sigma_b^2}}$$  \hspace{1cm} (E.1)

where $V$ is a $\chi^2_{2l+1}$ distributed random number and $C_l$ is true angular power spectrum (for the definition of rest of the parameters refer to above reference).

From the above equation we can see the $C_l^{est}$ is a linear transformation of $V$ i.e

$$C_l^{est} = \frac{C_l + w^{-1}e^{2\sigma_b^2}}{2l + 1} V - w^{-1}e^{2\sigma_b^2}$$  \hspace{1cm} (E.2)

Therefore to obtain the PDF\footnote{Probability Density Function} of $C_l^{est}$ we have to find the PDF of linear transformation of $\chi^2$ distributed random variables.

E.3 Linear transformation of $\chi^2$

The $\chi^2$ probability density function with $k$ degrees of freedom is defined as
Appendix E. The distribution of angular power spectrum of CMB

\[ f_X(x; k) = \frac{1}{2^{k/2}} \frac{1}{\Gamma(k/2)} x^{k/2 - 1} e^{-\frac{x}{2}} \]  

(E.3)

for \( x > 0 \) and \( k > 0 \), which is a special case of \( \Gamma(k, \theta) \) distribution

\[ f_X(x; k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} \]  

(E.4)

for \( x > 0 \) and \( k, \theta > 0 \).

Let say \( X \sim \chi^2_k \) then, the linear transformation of \( X \), \( Y = aX + b \) (\( a, b \) are constant and \( a > 0 \)) has the probability density function in the form of

\[
\begin{align*}
    f_Y(y) &= f_X(x) \left| \frac{dx}{dy} \right| \\
    &= \frac{1}{2^{k/2}} \frac{1}{\Gamma(k/2)} \left( \frac{y - b}{a} \right)^{k/2 - 1} e^{-\frac{y-b}{2a}} \times \frac{1}{a} \quad \text{(E.6)} \\
    f_Y(y) &= \frac{1}{2a} \frac{1}{\Gamma(k/2)} \left( \frac{y - b}{2a} \right)^{k/2 - 1} e^{-\frac{y-b}{2a}} \quad \text{(E.7)} 
\end{align*}
\]

By comparing the last equation and \( \chi^2 \) distribution (Eq. E.3) we see the linear transformation is not \( \chi^2 \) distributed any more! But if instead of linear transformation we have scaling i.e. \( Y = aX \) \( (b = 0) \) then the Eq. E.7 becomes

\[
\begin{align*}
    f_Y(y) &= \frac{1}{2a} \frac{1}{\Gamma(k/2)} \left( \frac{y}{2a} \right)^{k/2 - 1} e^{-\frac{y}{2a}} \quad \text{(E.8)} \\
    f_Y(y) &= \frac{1}{(2a)^{k/2}} \frac{1}{\Gamma(k/2)} y^{k/2 - 1} e^{-\frac{y}{2a}} \quad \text{(E.9)} 
\end{align*}
\]

which is \( \Gamma(k/2, 2a) \) (Eq. E.4). Therefore scaling of \( \chi^2_k \) distributed random variables converts into \( \Gamma(k/2, 2a) \) (where \( a \) is scaled factor).
Appendix E. The distribution of angular power spectrum of CMB

**Amoroso distribution**

*Amoroso distribution*\(^2\) [123] is a continuous distribution with probability density function as

\[
\text{Amoroso}(x|\nu, \theta_A, \alpha, \beta) = \left| \frac{\beta}{\theta_A} \right| \frac{1}{\Gamma(\alpha)} \left( \frac{x - \nu}{\theta_A} \right)^{\alpha \beta - 1} e^{-\left( \frac{x - \nu}{\theta_A} \right)^{\alpha \beta}} \quad (E.10)
\]

for \(-\infty \leq \nu, \theta_A, \beta \leq +\infty, \ \alpha > 0,\)

\[
x \geq \nu \quad (\theta_A > 0) \quad \text{and} \quad x \leq \nu \quad (\theta_A < 0).
\]

With

\[
\text{mean} = \nu + \theta_A \frac{\Gamma(\alpha + \frac{1}{\beta})}{\Gamma(\alpha)} \quad \alpha + \frac{1}{\beta} \geq 0 \quad (E.11)
\]

\[
\text{variance} = \theta_A^2 \left[ \frac{\Gamma(\alpha + \frac{3}{2})}{\Gamma(\alpha)} - \frac{\Gamma(\alpha + 1)^2}{\Gamma(\alpha)^2} \right] \quad \alpha + \frac{2}{\beta} \geq 0 \quad (E.12)
\]

By comparing the Amoroso distribution with Eq. E.7 we realize that the linear transformation of \(\chi^2_k\) is a special case of Amoroso distribution with

\[
\begin{align*}
x &= y \\
nu &= b \\
\theta_A &= 2a \\
\alpha &= k/2 \\
\beta &= 1
\end{align*}
\]  
\[\text{(E.13)}\]

Then the mean and variance (Eq(s). E.18, E.23) become

mean = \( b + 2a \frac{\Gamma(k/2+1)}{\Gamma(k/2)} \)  \hspace{1cm} (E.14)
\[ = b + 2a \frac{k}{2} \hspace{1cm} (E.15) \]
\[ = b + ak \hspace{1cm} (E.16) \]

which is exactly what we expect i.e.

\( \mathbf{E}(Y) = \mathbf{E}(aX + b) = a\mathbf{E}(X) + b \)  \hspace{1cm} (E.17)

Because \( X \sim \chi^2_k \), its mean is equal to \( k \) then

\( \mathbf{E}(Y) = ak + b \)  \hspace{1cm} (E.18)

And

\begin{align*}
\text{variance} &= (2a)^2 \left[ \frac{\Gamma(k/2+2)}{\Gamma(k/2)} - \frac{\Gamma(k/2+1)^2}{\Gamma(k/2)^2} \right] \hspace{1cm} (E.19) \\
&= (2a)^2 \left[ (\frac{k}{2} + 1)\frac{k}{2} - \frac{k^2}{4} \right] \hspace{1cm} (E.20) \\
&= 2ka^2 \hspace{1cm} (E.21)
\end{align*}

which is exactly equal to

\( \mathbf{V}(Y) = \mathbf{V}(aX + b) = a^2\mathbf{E}(X) \)  \hspace{1cm} (E.22)

Because \( X \sim \chi^2_k \), its variance is equal to \( 2k \) then

\( \mathbf{V}(Y) = 2ka^2 \)  \hspace{1cm} (E.23)

By comparing Eq. E.2 and \( Y = aX + b \) we realize that

\[ Y = C_l^{est}, \hspace{1cm} X = V, \hspace{1cm} k = 2l + 1 \]
\[ a = \frac{C_l + \frac{1}{w}e^{2}\sigma_{k}^2}{2l + 1}, \hspace{1cm} b = -\frac{1}{w}e^{2}\sigma_{k}^2 \]  \hspace{1cm} (E.24)
Appendix E. The distribution of angular power spectrum of CMB

Then by substituting above expression in Eq.(s) E.18 and E.23, the mean and variance of $C_{l}^{est}$ can be given as

$$E(C_{l}^{est}) = (2l + 1) \times \frac{C_l + w^{-1}l^2\sigma_b^2}{2l + 1} - w^{-1}l^2\sigma_b^2$$
$$= C_l \quad (E.25)$$

$$V(C_{l}^{est}) = 2(2l + 1) \times \left( \frac{C_l + w^{-1}l^2\sigma_b^2}{2l + 1} \right)^2$$
$$= \frac{2}{2l + 1}(C_l - w^{-1}l^2\sigma_b^2) \quad (E.26)$$

As expected ([109], Appendix. Eq. A.11).

### E.4 Linear transformation of Gamma distribution

For the completeness of this exercise we obtain the PDF of linear transformation of $\Gamma$ distributed random numbers. Suppose $X \sim \Gamma(k, \theta)$ (Eq. E.4) and $Y$ is the linear transformation of $X$, i.e $Y = aX + b$ ($a, b$ are constant and $a > 0$). Then the PDF of $Y$ would be

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$$
$$= \frac{1}{\theta^k \Gamma(k)} \left( \frac{y - b}{a} \right)^{k-1} e^{-\frac{y-b}{\theta a}} \times \frac{1}{a} \quad (E.27)$$

$$f_Y(y) = \frac{1}{a\theta \Gamma(k)} \left( \frac{y - b}{a\theta} \right)^{k-1} e^{-\frac{y-b}{\theta a \theta}} \quad (E.28)$$

We can see the above distribution is equivalent to Amoroso distribution (Eq. E.10) if we substitute
Appendix E. The distribution of angular power spectrum of CMB

$$\begin{align*}
x &= y \\
\nu &= b \\
\theta_A &= a\theta \\
\alpha &= k \\
\beta &= 1
\end{align*}$$  \quad (E.30)

\[\]

E.5 Conclusion

We show the $C_l^{est}$ is not $\chi^2_{2l+1}$ distributed whereas it is a linear transformation of $\chi^2_{2l+1}$ distributed random numbers. Since the mean value of $\chi^2_{2l+1}$ is $2l + 1$, it cannot represent the distribution of $C_l^{est}$ which is expected to have mean equal to $C_l$. Hence we prove the linear transformation of $\chi^2$ distributed random numbers has a specific Amoroso distribution. We calculate the mean value and variance of this kind of Amoroso distribution and show it is equal to mean and variance of $C_l^{est}$. As a result, $C_l^{est}$ is a special case of Amoroso distribution which is equivalent to $\chi^2_{2l+1}$ distribution except for its mean value.
Appendix F

Efficient weight matrix calculation for the inverse-noise-weighted loss function

The risk function $R(\lambda)$ (Eq. 2.34), unfortunately, depends on the unknown regression function $f$, and therefore needs to be estimated. To find an estimator for $R(\lambda)$ we need to represent the weighted loss function (Eq. 2.33) in more applicable form. We show that the effect of the weight term in weighted loss function can be expressed in terms of a weight matrix $W$ which will be defined below. In this appendix we present two different ways of calculating the weight matrix. The first method discussed below was developed in [1]. The second method developed by us has some advantages in comparison with first one. We will point these advantages out at the end of this section.

First Method

We expand the weight term as $(1/\sigma^2(x)) \approx \sum_{i=0}^{N-1} w_i \phi_i(x)$ for the cosine basis (Eq. 2.10) and insert it together with Eq. 2.11 and Eq. 2.12 in loss function (Eq. 2.33).
we get

\[
L(f(x), \hat{f}(x)) \approx \int_0^1 \left( \sum_{j=0}^{N-1} (\beta_j - \hat{\beta}_j) \phi_j(x) \right)^2 \left( \sum_{l=0}^{N-1} w_l \phi_l(x) \right) \, dx \tag{F.1}
\]

\[
= \sum_{j,k=0}^{N-1} (\beta_j - \hat{\beta}_j)(\beta_k - \hat{\beta}_k) \sum_{l=0}^{N-1} w_l \int_0^1 \phi_j(x) \phi_k(x) \phi_l(x) \, dx \tag{F.2}
\]

We define the weight matrix \( W \)

\[
W_{jk} = \sum_l \Delta_{jkl} w_l, \tag{F.3}
\]

where \( \Delta_{jkl} = \int_0^1 \phi_j(x) \phi_k(x) \phi_l(x) \, dx \). For the cosine basis (Eq. 2.10), we have

\[
\Delta_{jkl} = \begin{cases} 
1, & \text{if } \#\{j,k,l \} = 3, \\
0, & \text{if } \#\{j,k,l \} = 2, \\
\delta_{jk} \delta_{kl} + \delta_{jl} \delta_{kl} + \delta_{kl} \delta_{0j}, & \text{if } \#\{j,k,l \} = 1, \\
\frac{1}{\sqrt{2}} (\delta_{l,j+k} + \delta_{l,|j-k|}), & \text{if } \#\{j,k,l \} = 0.
\end{cases}
\]

Finally, the weighted loss function takes the form of

\[
L(f(x), \hat{f}(x)) \approx \sum_{j,k=0}^{N-1} (\hat{\beta}_j - \beta_j)(\hat{\beta}_k - \beta_k) W_{jk} = (\hat{\beta} - \beta)^T W (\hat{\beta} - \beta) \tag{F.4}
\]

where \( \hat{\beta} \equiv (\hat{\beta}_0, \ldots, \hat{\beta}_{N-1})^T \) and \( \beta \equiv (\beta_0, \ldots, \beta_{N-1})^T \).
Second Method

We first simplify the weighted loss function (Eq. 2.33) as

\[
L(\lambda) = \int_0^1 \left( \frac{f(x) - \hat{f}_N(x)}{\sigma(x)} \right)^2 dx \\
= \int_0^1 \left( f(x) - \hat{f}_N(x) \right)^2 \sigma^{-2}(x) dx \\
\approx \frac{1}{N} \sum_{j=0}^{N-1} \left( \hat{f}(x_j) - f(x_j) \right)^2 \sigma^{-2}(x_j) \\
= \frac{1}{N} \sum_{j=0}^{N-1} \left( \hat{f}(x_j)\sigma^{-1}(x_j) - f(x_j)\sigma^{-1}(x_j) \right)^2 
\]  
(F.5)

From Eq. 2.12 and the definition of the orthonormal matrix, i.e., \( U_{ij} = \phi_j(x_i) / \sqrt{N} \),

\[
\hat{F} = \sqrt{N} U \hat{\beta}, 
\]

where \( \hat{F} \equiv (\hat{f}(x_0), \ldots, \hat{f}(x_{N-1}))^T \). Therefore,

\[
\hat{F}_j = \hat{f}(x_j) = \sqrt{N} (U \hat{\beta})_j = \sqrt{N} \sum_{l=0}^{N-1} U_{jl} \hat{\beta}_l 
\]  
(F.6)

Similarly from Eq. 2.18 we have

\[
F = \sqrt{N} U \beta 
\]

where \( F \equiv (f(x_0), \ldots, f(x_{N-1}))^T \).

Hence,

\[
F_j = f(x_j) = \sqrt{N} (U \beta)_j = \sqrt{N} \sum_{l'=0}^{N-1} U_{jl'} \beta_{l'} 
\]  
(F.7)
Appendix F. Efficient weight matrix calculation for the inverse-noise-weighted loss function

By Substituting Eq(s. F.6 and F.7 in Eq. F.5 we have

\[
L(\lambda) = \left( \sum_{j=0}^{N-1} \left( \sum_{l=0}^{N-1} U_{jl} \hat{\beta}_l \sigma^{-1}(x_j) - \sum_{l'=0}^{N-1} U_{j,l'} \beta_{l'} \sigma^{-1}(x_j) \right) \right)^2 \\
= \sum_{j=0}^{N-1} \left( \sum_{l=0}^{N-1} R_{jl} \hat{\beta}_l - \sum_{l'=0}^{N-1} R_{jl'} \beta_{l'} \right)^2 \\
= \sum_{j=0}^{N-1} \left( (R\hat{\beta})_j - (R\beta)_j \right)^2 \\
= (R\hat{\beta} - R\beta)^T (R\hat{\beta} - R\beta) \\
= (\hat{\beta} - \beta)^T R^T R (\hat{\beta} - \beta) \quad (F.8)
\]

where \( R_{jl} = U_{jl} \sigma^{-1}(x_j) \).

Therefore, the weight matrix \( W \) can be written as

\[
W = R^T R \quad (F.9)
\]

which is another expression for \( W \) matrix. Substituting \( R_{jl} = U_{jl} \sigma^{-1}(x_j) \),

\[
W_{ij} = \sum_{l=0}^{N-1} R_{il}^T R_{lj} = \sum_{l=0}^{N-1} U_{il}^T S_{lk} U_{lj} \\
= \sum_{l=0}^{N-1} \left( U_{il} \sigma^{-2}(x_i) U_{lj} \right) = \sum_{l=0}^{N-1} \left( U_{il} \sigma^{-2}(x_i) U_{lj} \right) \\
= \sum_{l=0}^{N-1} \left( U_{il}^T \sigma^{-2}(x_i) U_{lj} \right) = \sum_{l=0}^{N-1} U_{il}^T \left( \sum_{k=0}^{N-1} \sigma^{-2}(x_i) \delta_{lk} U_{kj} \right) \\
= \sum_{l=0}^{N-1} U_{il}^T \left( \sum_{k=0}^{N-1} \sigma^{-2}(x_i) \delta_{lk} U_{kj} \right) = \sum_{k,l=0}^{N-1} U_{il}^T S_{lk} U_{kj} \quad (F.10)
\]

where \( S_{lk} = \sigma^{-2}(x_i) \delta_{lk} \) and \( \delta_{lk} \) is Kronecker delta.

Therefore

\[
W = U^T S U \quad (F.11)
\]

We see the weight matrix \( W \) can be calculated from the basis matrix \( U \) and a diagonal matrix \( S \equiv \text{diag}(\sigma^{-2}(x_0), \ldots, \sigma^{-2}(x_{N-1})) \).
Eq. F.11 has some advantages in comparison to Eq. F.3. It needs less computational efforts which reduces computational time and roundoff error and it can be easily calculated for any basis whereas in Eq. F.3, $\Delta_{jkl}$ must be calculated for every new basis. Also Eq. F.11 can be used not only for equispaced data based on the cosine basis, but can be easily used for any non-equispaced data (refer to Sec. 2.2.1) as well.