Distributions and Matrix Algebra

1 Mean square error

Define the mean square error of an estimator $\hat{\theta}$ as $MSE(\hat{\theta}) = \mathbb{E}(\hat{\theta} - \theta)^2$.

First, we derive the MSE for $\hat{\theta}$. Assume that $\hat{\theta}$ is the parameter estimate from a set of random variables $\mathbf{X} = X_1, ..., X_n$ which are IID, and where we assume that θ is unknown. Our estimate for θ is a random variable since it inherits random fluctuations from those of $\mathbf{X}$. Then, the MSE for $\hat{\theta}$ is given by:

$$\begin{aligned}
MSE(\hat{\theta}) &= \mathbb{E}(\hat{\theta} - \theta)^2 \\
&= \mathbb{E}\left[(\hat{\theta} - E(\hat{\theta}) + \mathbb{E}(\hat{\theta}) - \theta)^2\right] \\
&= \mathbb{E}\left[(\hat{\theta} - \theta)^2 + 2\left((\hat{\theta} - \mathbb{E}(\hat{\theta}))(\mathbb{E}(\hat{\theta}) - \theta)\right) + (\mathbb{E}(\hat{\theta}) - \theta)^2\right] \\
&= \mathbb{E}\left[(\hat{\theta} - E(\hat{\theta}))^2\right] + 2\mathbb{E}\left[(\hat{\theta} - \mathbb{E}(\hat{\theta}))(\mathbb{E}(\hat{\theta}) - \theta)\right] + \mathbb{E}\left[(\mathbb{E}(\hat{\theta}) - \theta))^2\right] \\
&= \mathbb{E}\left[(\hat{\theta} - \mathbb{E}(\hat{\theta}))^2\right] + 2(\mathbb{E}(\hat{\theta}) - \mathbb{E}(\hat{\theta}))(\mathbb{E}(\hat{\theta}) - \mathbb{E}(\theta)) + \mathbb{E}\left[(\mathbb{E}(\hat{\theta}) - \theta))^2\right] \\
&= \mathbb{E}\left[(\hat{\theta} - \mathbb{E}(\hat{\theta}))^2\right] + (\mathbb{E}(\hat{\theta}) - \theta)^2.
\end{aligned}$$

In the derivation, we use the linear properties of the expectations operator. The final line is an expression for the variance of $\hat{\theta}$ and its bias term squared.

1.1

Show $\mathbb{E}(\hat{\theta} - \mathbb{E}(\hat{\theta}))(\mathbb{E}(\hat{\theta}) - \theta) = 0$.

This was shown in lines 4 and 5 in the second term.

1.2

Show $MSE(\hat{\theta}) = Var(\hat{\theta}) + [\mathbb{E}(\hat{\theta}) - \theta]^2$.

This was shown above in the final line of the derivation.

1.3

Compare MSE for the residual variance estimators

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2, \text{ and}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

You can assume that

$$\sum_{i=1}^{n}(Y_i - \bar{Y}) \overset{d}{\sim} \sigma^2\chi^2_{n-1},$$

and that χ^2_{n-1} *has an expectation* $n-1$ *and variance* $2(n-1)$.

First, compute the MSE of the sample estimator for population variance. In the case of s^2 we have $MSE(s^2)$:

$$MSE(s^2) = \mathbb{E}(s^2 - \sigma^2)^2.$$

Using the information we are given:

$$\begin{aligned}
MSE(s^2) &= \text{Var}(s^2) + \text{Bias}(s^2, \sigma^2)^2 \\
&= \text{Var}\left[\frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2\right] + \mathbb{E}\left[\left(\mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2\right] - \sigma^2\right)^2\right] \\
&= \frac{1}{(n-1)^2}\text{Var}\left[\sum_{i=1}^n (Y_i - \bar{Y})^2\right] + \mathbb{E}\left[\left(\frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^n (Y_i - \bar{Y})^2\right] - \sigma^2\right)^2\right] \\
&= \frac{1}{(n-1)^2}\text{Var}\left[\sigma^2\chi^2_{n-1}\right] + \mathbb{E}\left[\left(\frac{1}{n-1}\mathbb{E}\left[\sigma^2\chi^2_{n-1}\right] - \sigma^2\right)^2\right] \\
&= \frac{\sigma^4}{(n-1)^2}\text{Var}\left[\chi^2_{n-1}\right] + \mathbb{E}\left[\left(\frac{1}{n-1}(n-1)\mathbb{E}\left[\sigma^2\right] - \sigma^2\right)^2\right] \\
&= \frac{\sigma^4}{(n-1)^2}(2(n-1)) \\
&= \frac{2\sigma^4}{n-1}.
\end{aligned}$$

This result tells us that s^2 is unbiased, and that essentially $MSE(s^2) = \text{Var}(s^2)$. Next we compute $MSE(\hat{\sigma}^2)$. Using our results above, $MSE(\hat{\sigma}^2)$ is expressed as:

$$\begin{aligned}
MSE(\hat{\sigma}^2) &= \mathbb{E}\left[\hat{\sigma}^2 - \sigma^2\right]^2 \\
&= \text{Var}(\hat{\sigma}^2) + \mathbb{E}\left[(\mathbb{E}(\hat{\sigma}^2) - \sigma^2)^2\right],
\end{aligned}$$

and so we compute $\mathbb{E}(\hat{\sigma}^2)$ and $\text{Var}(\hat{\sigma}^2)$. By a LLN, $n^{-1}\sum_{i=1}^n (Y_i - \bar{Y})^2 \to \sigma^2$, and so

$$\begin{aligned}
\mathbb{E}(\hat{\sigma}^2) &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})^2\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sigma^2\chi^2_{n-1}\right] \\
&= \frac{\sigma^2}{n}\mathbb{E}\left[\chi^2_{n-1}\right] \\
&= \frac{n-1}{n}\sigma^2,
\end{aligned}$$

and variance of $\hat{\sigma}^2$ is given by:

$$\begin{aligned}
\text{Var}(\hat{\sigma}^2) &= \text{Var}\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right] \\
&= \text{Var}\left[\frac{1}{n}\sigma^2\chi^2_{n-1}\right] \\
&= \frac{\sigma^4}{n^2}\text{Var}\left[\chi^2_{n-1}\right] \\
&= \frac{2(n-1)\sigma^4}{n^2}.
\end{aligned}$$

Putting it together:

$$\begin{aligned}
MSE(\hat{\sigma}^2) &= \frac{2(n-1)\sigma^4}{n^2} + \mathbb{E}\left[\left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2\right] \\
&= \frac{2(n-1)\sigma^4}{n^2} + \mathbb{E}\left[\frac{\sigma^4}{n^2}\right] \\
&= \frac{2(n-1)\sigma^4}{n^2} + \frac{\sigma^4}{n^2}
\end{aligned}$$

Comparing the MSE terms for s^2 and $\hat{\sigma}^2$ we see a bias-efficiency tradeoff. $\hat{\sigma}^2$ is biased but more efficient compared to s^2. The key takeaway here is that s^2 adjust for the degrees of freedom and is unbiased, unlike $\hat{\sigma}^2$.

2 The exponential distribution

In this exercise we discuss the distribution of an average of IID variables. Once again, we use the exponential distribution. Pay particular attention to which assumptions are used where. Suppose $Y_1, Y_2, \ldots$ are IID Exponential(2) so

$$f_Y(y) = \frac{1}{2}\exp(-y/2),\ y > 0.$$

2.1

Prove that $\mathbb{E}(Y_1)$ and $Var(Y_1) = 4$. Hint: use partial integration. Recall that

$$\int_a^b u\ dv = [uv]_a^b - \int_a^b v\ du$$

Assuming that $f_Y(y)$ satisfies all the requirements of a continuous distribution, we can compute the expectation and variance of Y_1 as the following:

$$\begin{aligned}
\mathbb{E}(Y_1) &= \int_0^\infty y_1 \frac{1}{2}\exp\left(\frac{-y_1}{2}\right) dy_1 \\
&= \left[-y_1 \exp\left(\frac{-y_1}{2}\right)\right]_0^\infty + \int_0^\infty \exp\left(\frac{-y_1}{2}\right) dy_1 \\
&= [0+0] + 2\int_0^\infty \frac{1}{2}\exp\left(\frac{-y_1}{2}\right) dy_1 \\
&= 2
\end{aligned}$$

since the second term on the RHS of the second line is essentially double the mass of the distribution. For the variance we first define it as

$$\text{Var}(Y_1) = \mathbb{E}(Y_1^2) - \mathbb{E}(Y_1)^2,$$

and so

$$\begin{aligned}
\mathbb{E}(Y_1^2) &= \int_0^\infty y_1^2 \frac{1}{2}\exp\left(\frac{-y_1}{2}\right) dy_1 \\
&= \left[-y_1^2 \exp\left(\frac{-y_1}{2}\right)\right]_0^\infty + \int_0^\infty 2y_1 \exp\left(\frac{-y_1}{2}\right) dy_1 \\
&= 0 + 2\int_0^\infty y_1 \exp\left(\frac{-y_1}{2}\right) dy_1 \\
&= 4\int_0^\infty y_1 \frac{1}{2}\exp\left(\frac{-y_1}{2}\right) dy_1 \\
&= 8
\end{aligned}$$

where we use the result for $\mathbb{E}(Y_1)$. The variance is therefore:

$$\text{Var}(Y_1) = 8 - 2^2 = 4.$$

2.2

Argue that $Y_1 \overset{d}{\sim} \chi^2_2$*. Hint: See Example 2.12.*

This argument relies on a simple theorem:

$$Y_1 \overset{d}{\sim} \chi^2_n$$

if

$$f_Y(y) = \frac{1}{c_n} y^{\frac{n}{2}-1} \exp\left(\frac{-y}{2}\right), \quad y \geq 0,$$

for some normalisation constant, c_n such that $\int_0^\infty f_{Y_1}(y) = 1$. The normalisation constant can be expressed in terms of the gamma integral:

$$c_n = \Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}.$$

As we wish to argue $Y_1 \overset{d}{\sim} \chi^2_2$, for $n = 2$, we have a normalisation constant of 2. These imply:

$$f_Y(y) = \frac{1}{2} y^{\frac{2}{2}-1} \exp\left(\frac{-y}{2}\right) = \frac{1}{2} \exp\left(\frac{-y}{2}\right),$$

which completes our argument.

2.3

Find the exact distribution of $\bar{Y} = n^{-1} \sum_{i=1}^n Y_i$*. Hint: See Example 4.4.*

This is most easily done by finding the moment generating function (MGF) of Y_i since we assume they are IID Exponential(2):

$$M_{Y_i}(t) = \frac{1}{1-2t} = (1-2t)^{-1}.$$

We can check the MGF by differentiating it and then evaluating the derivative at $t = 0$:

$$\mathbb{E}(Y_i) = M'_{Y_i}(t = 0) = 2(1 - 2(0))^{-2} = 2.$$

We can attain the MGF of the sample mean due to our IID assumption too. Here the MGF is a function of t and n:

$$M_{\bar{Y}}(t) = \left[M_{Y_i}\left(\frac{t}{n}\right)\right]^n = \left[\frac{1}{1 - \frac{2}{n}t}\right]^n,$$

which we note is a MGF of a gamma distribution. Thus, the sample mean's exact distribution is $\Gamma(n, (2/n))$.

2.4

Apply the CLT to approximate the distribution of $\bar{Y}$*. What does this say about the shape of the* χ^2 *distribution?*

By the central limit theorem,

$$\frac{\sqrt{n}(\bar{Y}-\mu)}{\sigma} \sim N(0,1)$$
$$\implies \bar{Y} \sim N(\mu, \frac{\sigma^2}{n}).$$

We know that the χ^2 distribution is a special case of the gamma distribution when a random variable is distributed as χ^2 with 2 degrees of freedom and when the scale parameter is 2. With what we know from from 2a) we know that $\mathbb{E}(Y) = 2$ and $\text{Var}(Y) = 4$. Therefore

$$\bar{Y} \sim N(2, \frac{4}{n}).$$

This implies that the χ^2 distribution tends to a normal distribution for sufficiently large values of n.

3 Marginal densities

Table 1: Marginal densities for X, Y

$Y \backslash X$	-1	0	1	$f(y)$
-1	-	d	-	1/4
0	c	-	a	1/2
1	-	b	-	1/4
$f(x)$	1/4	1/2	1/4	1

3.1

Find the marginal densities for X, Y.

$$f_X(x) = \sum_y f_{XY}(x,y) = (0,-1) + (-1,0) + (0,1) + (1,1),$$
$$f_Y(y) = \sum_x f_{XY}(x,y) = (-1,0) + (0,-1) + (0,1) + (1,0).$$

3.2

Find $\mathbb{E}(XY)$, $\mathbb{E}(X)$, and $\mathbb{E}(Y)$, and check that $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

$$\mathbb{E}(X|y) = \sum_x f_{X|y}(x) = \frac{1}{4}(-1) + \frac{1}{2}0 + \frac{1}{4}1 = 0$$
$$\mathbb{E}(Y|x) = \sum_y f_{Y|x}(y) = \frac{1}{4}(-1) + \frac{1}{2}0 + \frac{1}{4}1 = 0$$

To find $E(XY)$, we use the covariance formula:

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\ \implies \mathbb{E}(XY) &= \mathrm{Cov}(X,Y). \end{aligned}$$

$\mathbb{E}(XY) = 0$ since $XY = 0$ at all four outcomes. $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ since the marginal distributions are symmetric.

3.3

Show that X, Y are not independent.

Shown in 3a) as independence is given by $P(A \cap B) = P(A)P(B)$. This clearly does not hold for the case where X and Y eventuate to 0.

4 Matrix operations

(see question for matrices)

4.1

When $\mathbf{X} = [X_1, X_2]$:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}, \ (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & -1/4 & 1/4 & -1/4 \end{bmatrix}$$

and so

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y} = \begin{bmatrix} 5/2 \\ -1/2 \end{bmatrix}.$$

This gives

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} 2 \\ 3 \\ 2 \\ 3 \end{bmatrix},$$

and

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}.$$

Which also gives the orthogonality condition:

$$\mathbf{X}^\top \hat{\boldsymbol{\epsilon}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

When $\mathbf{X} = [X_1, X_3]$:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}, \ (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top = \begin{bmatrix} 1/4 & -1/4 & 1/4 & 3/4 \\ 0 & 1/2 & 0 & -1/2 \end{bmatrix}$$

and so

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y} = \begin{bmatrix} -7/2 \\ -1 \end{bmatrix}.$$

This gives

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} 5/2 \\ 3/2 \\ 5/2 \\ 7/2 \end{bmatrix},$$

and

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} -3/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}.$$

Which also gives the orthogonality condition:

$$\mathbf{X}^\top \hat{\boldsymbol{\epsilon}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Note, we could've used projection and elimination matrices.

4.2

Given above.

4.3

The matrices do not conform (3×1 matrix with a 3×3 matrix) and you cannot divide matrices, respectively.

5 Block matrix inversion

(see question paper for full details)

5.1

Let

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

and

$$E = M_{11\cdot 2} = A - BD^{-1}C.$$

Matrices A and D must be square and invertible. B and C matrices are not guaranteed to be square, so we cannot invert them. Under a symmetry assumption $B^\top = C$. To get MM^{-1}:

$$MM^{-1} = I = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} E^{-1} & -E^{-1}BD^{-1} \\ -D^{-1}CE^{-1} & D^{-1} + D^{-1}CE^{-1}BD^{-1} \end{bmatrix} = \begin{bmatrix} I & O \\ O & I \end{bmatrix},$$

where

$$\begin{aligned} AE^{-1} - BD^{-1}CE^{-1} &= [A - BD^{-1}C]E^{-1} \\ &= EE^{-1} \\ &= I, \end{aligned}$$

$$\begin{aligned} -AE^{-1}BD^{-1} + BD^{-1} + BD^{-1}CE^{-1}BD^{-1} &= BD^{-1} + [-A + BD^{-1}C]E^{-1}BD^{-1} \\ &= BD^{-1} - EE^{-1}BD^{-1} \\ &= O, \end{aligned}$$

$$\begin{aligned} CE^{-1} - DD^{-1}CE^{-1} &= CE^{-1} - CE^{-1} \\ &= O, \end{aligned}$$

$$\begin{aligned} -CE^{-1}BD^{-1} + DD^{-1} + DD^{-1}CE^{-1}BD^{-1} &= -CE^{-1}BD^{-1} + I + CE^{-1}BD^{-1} \\ &= I. \end{aligned}$$

Now, suppose that we have

$$\begin{aligned} L &= \begin{bmatrix} I & -BD^{-1} \\ O & I \end{bmatrix}, \text{ and,} \\ L^\top &= \begin{bmatrix} I & O \\ -(D^{-1})^\top B^\top & I \end{bmatrix}. \end{aligned}$$

Write $M = L^{-1}(LML^\top)(L^{-1})^\top$ and compute $LML^\top$. Write M^{-1} in terms of $(LML^\top)^{-1}$. We can do this by either assuming M is symmetric or asymmetric. If we assume asymmetry, then we have

$$\begin{aligned} M &= L^{-1}LML^\top(L^\top)^{-1} \\ &= L^{-1}LMRR^{-1}, \end{aligned}$$

say, where

$$R = \begin{bmatrix} I & O \\ -D^{-1}C & I \end{bmatrix}.$$

So

$$M^{-1} = R(LMR)^{-1}L = RR^{-1}M^{-1}L^{-1}L.$$

If M is symmetric then $-(D^{-1})^\top B^\top = -D^{-1}C$, and we have

$$\begin{aligned} LML^\top &= \begin{bmatrix} I & -BD^{-1} \\ O & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & O \\ -D^{-1}C & I \end{bmatrix} \\ &= \begin{bmatrix} A - BD^{-1}C & B - BD^{-1}D \\ C & D \end{bmatrix} \begin{bmatrix} I & O \\ -D^{-1}C & I \end{bmatrix} \\ &= LM \begin{bmatrix} I & O \\ -D^{-1}C & I \end{bmatrix} \\ &= \begin{bmatrix} E & O \\ C & D \end{bmatrix} \begin{bmatrix} I & O \\ -D^{-1}C & I \end{bmatrix} \\ &= \begin{bmatrix} E & O \\ C - DD^{-1}C & D \end{bmatrix} \\ &= \begin{bmatrix} E & O \\ O & D \end{bmatrix}. \end{aligned}$$

Therefore

$$(LML^\top)^{-1} = \begin{bmatrix} E^{-1} & O \\ O & D^{-1} \end{bmatrix},$$

and we can compute $M^{-1} = L^\top(LML^\top)^{-1}L$:

$$\begin{aligned} M^{-1} &= \begin{bmatrix} I & O \\ -D^{-1}C & I \end{bmatrix} \begin{bmatrix} E^{-1} & O \\ O & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ O & I \end{bmatrix} \\ &= \begin{bmatrix} E^{-1} & O \\ -D^{-1}CE^{-1} & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ O & I \end{bmatrix} \\ &= \begin{bmatrix} E^{-1} & -E^{-1}BD^{-1} \\ -D^{-1}CE^{-1} & D^{-1}CE^{-1}BD^{-1} + D^{-1} \end{bmatrix} \end{aligned}$$

and subbing in for $E = A - BD^{-1}C$ yields

$$M^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}.$$

5.2

Let $M_{22\cdot 1} = D - CA^{-1}B$. Show $M_{22\cdot 1}^{-1} = D^{-1} + D^{-1}CM_{11\cdot 2}^{-1}BD^{-1}$.

If

$$\begin{aligned} MM^{-1} &= I \\ \implies \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} U & V \\ X & Y \end{bmatrix} &= \begin{bmatrix} I & O \\ O & I \end{bmatrix}, \end{aligned}$$

so

$$\begin{aligned} AU + BX &= I && (1) \\ AV + BY &= O && (2) \\ CU + DX &= O && (3) \\ CV + DY &= I, && (4) \end{aligned}$$

which would then imply from (2)

$$\begin{aligned} AV &= -BY \\ V &= -A^{-1}BY. \end{aligned}$$

Substituting this value for V into (4) gives

$$\begin{aligned} C(-A^{-1}BY) + DY &= I \\ (\underbrace{D - CA^{-1}B}_{M_{22\cdot 1}})Y &= I \\ \implies M_{22\cdot 1}Y &= I \\ \therefore M_{22\cdot 1}^{-1} &= Y. \end{aligned}$$

Then, we wish to show $M_{22\cdot 1}^{-1} = D^{-1} + D^{-1}CM_{11\cdot 2}^{-1}BD^{-1}$, so begin with (3):

$$\begin{aligned} CU + DX &= O \\ X &= -D^{-1}CU. \end{aligned}$$

Then, substitute X into (1) to get

$$\begin{aligned} AU + B(-D^{-1}CU) &= I \\ (\underbrace{A - BD^{-1}C}_{M_{11\cdot 2}})U &= I \\ \implies M_{11\cdot 2}^{-1} &= U. \end{aligned}$$

Finally, putting it together, recall that

$$\begin{aligned} M^{-1} &= \begin{bmatrix} U & V \\ X & Y \end{bmatrix} \\ &= \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix} \\ \implies M^{-1} &= \begin{bmatrix} M_{11\cdot 2}^{-1} & V \\ X & M_{22\cdot 1}^{-1} \end{bmatrix}, \end{aligned}$$

finishing our proof.

6 Numerical issues of matrix inversion

(see problem set for details)

$$\begin{aligned}
\det(A) &= 1.1 - 1 = 1/10 \\
\det(B) &= \frac{1}{1,000,000,000} \\
\det(C) &= \frac{1}{10,000,000,000,000},
\end{aligned}$$

and

$$\begin{aligned}
A^{-1} &= \begin{bmatrix} 11 & -10 \\ -10 & 10 \end{bmatrix} \\
B^{-1} &= \begin{bmatrix} 1000000001 & -1000000000 \\ -1000000000 & 1000000000 \end{bmatrix},
\end{aligned}$$

and a similar expression can be found for C^{-1}. Multiplying the determinants of the inverse by the determinant of the matrix itself yields:

$$\det(A^{-1}) = 10, \implies \det(A^{-1})\det(A) = 1,$$

where the same can be shown for A and B.