1 Deriving Target Parameters

Consider the non-parameteric Roy model:

$$
\begin{aligned}
Y &= DY_1 + (1-D)\, Y_0 \\
D &= 1\left[U \le \nu\left(Z\right)\right]
\end{aligned}
$$

$$
(Y_0, Y_1, U) \perp Z
$$

$$
\mathbb{E}\left[Y_d^2\right] < \infty, d \in \{0,1\}
$$

where (Y_0, Y_1, U) are unobserved.

Problem 1.1. What assumptions about distribution of U did you see in class? Are they restrictive?

Solution. We assumed that $U \sim Uniform\,[0,1]$. This restriction is not restrictive.

- ▷ To see this, let U be any random variable and F_U denote its cdf.
- ▷ Then we know that $U^* \equiv F_U\left(U\right)$, which is also a random variable, is uniformly distributed.
- ▷ Since F_U is an increasing function in its argument, this implies that we have

$$
D = 1\left[U \le \nu\left(Z\right)\right] = 1\left[F_U\left(U\right) \le F_U\left(\nu\left(Z\right)\right)\right] = 1\left[U^* \le \nu^*\left(Z\right)\right]
$$

  i.e. we can transform a non-uniform variable into a uniform variable with support on $[0,1]$ arbitrarily, and thus the provided assumption is not restrictive.

Problem 1.2. Define MTE and derive ATE, ATT, and ATUT as weighted averages of MTE.

Solution. We first define MTE as the following:

$$
MTE(u) = \mathbb{E}\left[Y_1 - Y_0 | U = u\right]
$$

and we express ATE, ATT, and ATUT as the following.

1. Average Treatment Effect (ATE)
   - ▷ Definition: $\mathbb{E}\left[Y_1 - Y_0\right]$
   - ▷ Expression as an weighted average of MTE
     - * Using Law of Iterated Expectations, we have

$$
\begin{aligned}
\mathbb{E}\left[Y_1 - Y_0\right] &= \mathbb{E}\left[\mathbb{E}\left[Y_1 - Y_0 | U = u\right]\right] \\
&= \int_0^1 MTE\left(u\right)\left(1\right) du
\end{aligned}
$$

       i.e. a simple average of the MTEs since u is uniformly distributed.

* In other words, the weights are all equal across different $MTE(u)$.

2. Average Treatment Effect on the Treated (ATT)

▷ Definition: $\mathbb{E}\left[Y_1 - Y_0|D=1\right]$

▷ Expression as an unweighted average of MTE

* Using LIE, Bayes rule and the fact that $Z \perp (Y_1, Y_0, U)$:

$$\begin{aligned}
\mathbb{E}\left[Y_1 - Y_0|D=1\right] &= \int_0^1 \mathbb{E}\left[Y_1 - Y_0|D=1, U=u\right] du \\
&= \int_0^1 \mathbb{E}\left[\left(Y_1 - Y_0\right) 1\left\{D=1\right\}|U=u\right] \frac{1}{P\left(D=1\right)} du \\
&= \int_0^1 \mathbb{E}\left[Y_1 - Y_0|U=u\right] \frac{P\left(D=1|U=u\right)}{P\left(D=1\right)} du
\end{aligned}$$

* Since $D = 1\left[U \leq \nu\left(Z\right)\right]$, we have

$$P\left(D=1|U=u\right) = P\left(\nu\left(Z\right) \geq u\right)$$

* Plugging in, we obtain:

$$\mathbb{E}\left[Y_1 - Y_0|D=1\right] = \int_0^1 MTE\left(u\right)\left[\frac{P\left(\nu\left(Z\right) \geq u\right)}{P\left(D=1\right)}\right] du$$

▷ Interpretation: Those with lower values of u are more highly weighted. This makes sense since these people are more likely to take treatment.

3. Average Treatment Effect on the UnTreated (ATUT)

▷ Definition: $\mathbb{E}\left[Y_1 - Y_0|D=0\right]$

▷ Expression as an unweighted average of MTE

* Using LIE, Bayes rule and the fact that $Z \perp (Y_1, Y_0, U)$:

$$\begin{aligned}
\mathbb{E}\left[Y_1 - Y_0|D=0\right] &= \int_0^1 \mathbb{E}\left[Y_1 - Y_0|D=0, U=u\right] du \\
&= \int_0^1 \mathbb{E}\left[\left(Y_1 - Y_0\right) 1\left\{D=0\right\}|U=u\right] \frac{1}{P\left(D=0\right)} du \\
&= \int_0^1 \mathbb{E}\left[Y_1 - Y_0|U=u\right] \frac{P\left(D=0|U=u\right)}{P\left(D=0\right)} du
\end{aligned}$$

* Since $D = 1\left[U \leq \nu\left(Z\right)\right]$, we have

$$P\left(D=0|U=u\right) = P\left(\nu\left(Z\right) < u\right)$$

* Plugging in, we obtain:

$$\mathbb{E}\left[Y_1 - Y_0|D=0\right] = \int_0^1 MTE\left(u\right)\left[\frac{P\left(\nu\left(Z\right) < u\right)}{P\left(D=0\right)}\right] du$$

▷ Interpretation: Those with higher values of u are more highly weighted. This makes sense since these people are less likely to take treatment.

Problem 1.3. Interpret the weights you received in the previous part.

Solution. For the ATE, the weights are equal across all types of u since we are simply looking for the average effect across all people. For ATT, those with lower values of u are more highly weighted, which makes sense since these people are more likely to be treated. For ATUT, those with higher values of u are more highly weighted, since these people are less likely to take treatment.

2 Switching Regressions

Consider the following setup for the Roy model:

$$Y_1 = U_1$$
$$Y_0 = U_0$$
$$\begin{pmatrix} U_1 \\ U_0 \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}\right)$$
$$D = 1\left[U_1 > U_0\right]$$
$$Y = DY_1 + (1-D)\,Y_0 = \underbrace{(Y_1 - Y_0)}_{\beta} D + Y_0$$

Problem 2.1. Derive the expression for β_{OLS}. What treatment effect does it correspond to in the case $D \perp (Y_1, Y_0)$?

Solution. The expression for β_{OLS} is given as

$$\beta_{OLS} = \frac{Cov\,(D, Y)}{Var\,(D)}$$

▷ Numerator: $Cov\,(D,Y) = \mathbb{E}\left[YD\right] - \mathbb{E}\left[Y\right]\mathbb{E}\left[D\right]$

▷ Denominator: $Var\,(D) = \mathbb{E}\left[D\right](1 - \mathbb{E}\left[D\right])$

▷ Combining the above results, this yields:

$$\beta_{OLS} = \frac{\mathbb{E}\left[YD\right] - \mathbb{E}\left[Y\right]\mathbb{E}\left[D\right]}{\mathbb{E}\left[D\right](1 - \mathbb{E}\left[D\right])}$$

▷ Note that $Y_0 - Y_1 = U_0 - U_1$ is distributed $\mathcal{N}\left(0, \sigma^2 + 1 - 2\rho\sigma\right)$. Furthermore, if $X \sim N\left(\mu, \sigma^2\right)$, then the truncated normal in $[a, b]$ has pdf

$$\frac{\phi\,(\xi)}{\sigma\left[\Phi\,(\beta) - \Phi\,(\alpha)\right]}$$

where ϕ is the pdf of standad normal, Φ is the cdf of standard normal, and ξ, α, β are standardized variables:

$$\xi = \frac{x-\mu}{\sigma}, \alpha = \frac{a-\mu}{\sigma}, \beta = \frac{b-\mu}{\sigma}$$

Note that this yields the mean to be

$$\mu + \frac{\phi\,(\alpha) - \phi\,(\beta)}{\Phi\,(\beta) - \Phi\,(\alpha)}\sigma$$

▷ Using this we compute the following quantitites:

* $\mathbb{E}\left[D\right]$: we can write:

$$\mathbb{E}\left[D\right] = P\,(D=1) = P\,(U_0 - U_1 < 0) = \frac{1}{2}$$

  * $\mathbb{E}[Y]$: Since $Y = DU_1 + (1-D)U_0$, write:

$$\mathbb{E}[Y] = \mathbb{E}[DU_1] + \mathbb{E}[(1-D)U_0]$$

  * $\mathbb{E}[YD]$: Since $Y = DU_1 + (1-D)U_0$, write:

$$\begin{aligned}\mathbb{E}[YD] &= \mathbb{E}\left[D^2U_1\right] + \mathbb{E}[D(1-D)U_0] \\ &= \mathbb{E}[DU_1]\end{aligned}$$

▷ Rewriting β:

$$\begin{aligned}\beta_{OLS} &= \frac{\mathbb{E}[YD] - \mathbb{E}[Y]\mathbb{E}[D]}{\mathbb{E}[D](1-\mathbb{E}[D])} \\ &= 4(\mathbb{E}[YD] - \mathbb{E}[Y]\mathbb{E}[D]) \\ &= 4\left(\mathbb{E}[DU_1] - \frac{1}{2}\{\mathbb{E}[DU_1] + \mathbb{E}[(1-D)U_0]\}\right) \\ &= 2(\mathbb{E}[DU_1] - \mathbb{E}[(1-D)U_0]) \\ &= 2(\mathbb{E}[D(U_1+U_0)]) \\ &= 2\mathbb{E}[U_1+U_0|D=1]P(D=1) \\ &= \mathbb{E}[U_1+U_0|U_1>U_0]\end{aligned}$$

Note that

$$\begin{pmatrix} U_1 \\ U_0 \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}\right)$$

which implies

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} U_1+U_0 \\ U_1-U_0 \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2+2\rho\sigma+1 & \sigma^2-1 \\ \sigma^2-1 & \sigma^2-2\rho\sigma+1 \end{bmatrix}\right)$$

and define

$$\rho^* = \frac{\sigma^2-1}{\sqrt{(\sigma^2+1-2\rho\sigma)(\sigma^2+1+2\rho\sigma)}} = \frac{\sigma^2-1}{\sqrt{\sigma^2+1-4\rho^2\sigma^2}}$$

▷ Writing out the conditional density:

$$\begin{aligned}X|Y &\sim \mathcal{N}\left(\rho^*\sqrt{\frac{\sigma^2+2\rho\sigma+1}{\sigma^2-2\rho\sigma+1}}Y, (1-\rho^*)^2\left(\sigma^2+2\rho\sigma+1\right)\right) \\ &= \mathcal{N}\left(\left(\frac{\sigma^2-1}{\sigma^2-2\rho\sigma+1}\right)Y, (1-\rho^*)^2\left(\sigma^2+2\rho\sigma+1\right)\right)\end{aligned}$$

▷ Using the Law of Iterated Expectations (LIE):

$$\mathbb{E}[X|Y>0] = \mathbb{E}[\mathbb{E}[X|Y=y>0]|Y>0]$$

Since $(Y>0, Y=y)$ and $(Y=y)$ gives rise to the same conditional distribution:

$$\begin{aligned}&= \mathbb{E}\left[\left(\frac{\sigma^2-1}{\sigma^2-2\rho\sigma+1}\right)y\cdot 1\{Y=y>0\}|Y>0\right] \\ &= \left(\frac{\sigma^2-1}{\sigma^2-2\rho\sigma+1}\right)\mathbb{E}[y|Y>0]\end{aligned}$$

Using the expression for the truncated normal:

$$= 2\phi(0)\left(\frac{\sigma^2 - 1}{\sigma^2 - 2\rho\sigma + 1}\right)\sqrt{\sigma^2 - 2\rho\sigma + 1}$$

Thus:

$$\begin{aligned}\beta_{OLS} &= 2\phi(0)\left(\frac{\sigma^2 - 1}{\sqrt{\sigma^2 - 2\rho\sigma + 1}}\right) \\ &= \sqrt{\frac{2}{\pi}}\left(\frac{\sigma^2 - 1}{\sqrt{\sigma^2 - 2\rho\sigma + 1}}\right)\end{aligned}$$

Now assume $D \perp (Y_1, Y_0)$.

▷ Plugging in the expression $Y = \beta D + Y_0$:

$$\begin{aligned}\beta_{OLS} &= \frac{\mathbb{E}[(\beta D + Y_0) D] - \mathbb{E}[(\beta D + Y_0)]\mathbb{E}[D]}{\mathbb{E}[D](1 - \mathbb{E}[D])} \\ &= \frac{\mathbb{E}\left[\beta D^2\right] + \mathbb{E}[DY_0] - \mathbb{E}[\beta D]\mathbb{E}[D] - \mathbb{E}[Y_0]\mathbb{E}[D]}{\mathbb{E}[D](1 - \mathbb{E}[D])}\end{aligned}$$

▷ Using the fact that $D^2 = D$ and $\mathbb{E}[DY_0] = \mathbb{E}[D]\mathbb{E}[Y_0]$:

$$\begin{aligned}&= \frac{\mathbb{E}[\beta D] - \mathbb{E}[\beta D]\mathbb{E}[D]}{\mathbb{E}[D](1 - \mathbb{E}[D])} \\ &= \frac{\mathbb{E}[\beta D]}{\mathbb{E}[D]} = \mathbb{E}[\beta] \equiv ATE\end{aligned}$$

Thus, the treatment effect corresponds to the ATE in the specialized case.

Problem 2.2. Derive the expression for ATT and ATUT, commenting on their relative magnitudes and signs.

Solution. Preliminaries:

▷ Note that $Y_0 - Y_1 = U_0 - U_1$ is distributed $\mathcal{N}\left(0, \sigma^2 + 1 - 2\rho\sigma\right)$. Thus:

$$D = 1\{U_1 > U_0\} = 1\{0 > U_0 - U_1\}$$

which implies:

$$\mathbb{E}[D] = P(D = 1) = \frac{1}{2}$$

▷ Truncated normal: If $X \sim N\left(\mu, \sigma^2\right)$, then the truncated normal in $[a, b]$ has pdf

$$\frac{\phi(\xi)}{\sigma[\Phi(\beta) - \Phi(\alpha)]}$$

where ϕ is the pdf of standad normal, Φ is the cdf of standard normal, and ξ, α, β are standardized variables:

$$\xi = \frac{x - \mu}{\sigma}, \alpha = \frac{a - \mu}{\sigma}, \beta = \frac{b - \mu}{\sigma}$$

Note that this yields the mean to be

$$\mu + \frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)}\sigma$$

Given this, we can compute expressions for ATT and ATUT.

▷ ATT can be written as:

$$\begin{aligned}\mathbb{E}\left[Y_1 - Y_0 | D = 1\right] &= -\mathbb{E}\left[U_0 - U_1 | D = 1\right] \\ &= -\mathbb{E}\left[U_0 - U_1 | U_0 - U_1 < 0\right]\end{aligned}$$

* Using our formula for the mean of the truncated normal:

$$\mathbb{E}\left[Y_1 - Y_0 | D = 1\right] = \sqrt{(\sigma^2 + 1 - 2\rho\sigma)}\frac{\phi(0)}{\Phi(0)}$$

* Since $\Phi(0) = 1/2$, we have

$$\begin{aligned}\mathbb{E}\left[Y_1 - Y_0 | D = 1\right] &= 2\phi(0)\sqrt{(\sigma^2 + 1 - 2\rho\sigma)} > 0 \\ &= \sqrt{\frac{2}{\pi}}\sqrt{(\sigma^2 + 1 - 2\rho\sigma)} > 0\end{aligned}$$

▷ ATUT can be written as:

$$\begin{aligned}\mathbb{E}\left[Y_1 - Y_0 | D = 0\right] &= -\mathbb{E}\left[U_0 - U_1 | D = 0\right] \\ &= -\mathbb{E}\left[U_0 - U_1 | U_0 - U_1 > 0\right]\end{aligned}$$

* Using our formula for the mean of the truncated normal:

$$\mathbb{E}\left[Y_1 - Y_0 | D = 0\right] = -\sqrt{(\sigma^2 + 1 - 2\rho\sigma)}\frac{\phi(0)}{\Phi(0)}$$

* Since $\Phi(0) = 1/2$, we have

$$\begin{aligned}\mathbb{E}\left[Y_1 - Y_0 | D = 0\right] &= -2\phi(0)\sqrt{(\sigma^2 + 1 - 2\rho\sigma)} < 0 \\ &= -\sqrt{\frac{2}{\pi}}\sqrt{(\sigma^2 + 1 - 2\rho\sigma)} < 0\end{aligned}$$

Note that the ATT and ATUT have equal magnitude but different signs. Their magnitudes are the same because $Y_1 = U_1$ and $Y_0 = U_0$ as opposed to a more traditional notation of

$$Y_1 = X\beta_1 + U_1, \quad Y_0 = X\beta_0 + U_0$$

in which case the magnitudes will be different.

Problem 2.3. What is ATE in this case?

Solution. The ATE can be written as

$$\mathbb{E}\left[Y_1 - Y_0\right] = \mathbb{E}\left[U_1 - U_0\right] = 0$$

i.e. ATE is equal to zero.

Problem 2.4. Derive $\frac{\partial ATT}{\partial \rho}, \frac{\partial ATUT}{\partial \rho}, \frac{\partial \beta_{OLS}}{\partial \rho}$ and provide intuitive explanation for your results. Derive similar partial derivatives with respect to σ. Is there any simple intuitive explanation for these results?

Solution. The results are shown below:

▷ ATT: $\mathbb{E}[Y_1 - Y_0 | D = 1] = 2\phi(0)\sqrt{(\sigma^2 + 1 - 2\rho\sigma)}$

* Taking the derivative with respect to ρ and σ:

$$\frac{\partial ATT}{\partial \rho} = \frac{-2\phi(0)\sigma}{\sqrt{\sigma^2 + 1 - 2\rho\sigma}} < 0, \quad \frac{\partial ATT}{\partial \sigma} = \frac{2\phi(0)(\sigma - \rho)}{\sqrt{\sigma^2 + 1 - 2\rho\sigma}} > 0 \text{ if} \sigma > \rho$$

▷ ATUT: $\mathbb{E}[Y_1 - Y_0 | D = 1] = -2\phi(0)\sqrt{(\sigma^2 + 1 - 2\rho\sigma)}$

* Taking the derivative with respect to ρ and σ:

$$\frac{\partial ATUT}{\partial \rho} = \frac{2\phi(0)\sigma}{\sqrt{\sigma^2 + 1 - 2\rho\sigma}} > 0, \quad \frac{\partial ATUT}{\partial \sigma} = \frac{-2\phi(0)(\sigma - \rho)}{\sqrt{\sigma^2 + 1 - 2\rho\sigma}} < 0 \text{ if } \sigma > \rho$$

▷ β_{OLS}: $2\phi(0)\left(\frac{\sigma^2 - 1}{\sqrt{\sigma^2 - 2\rho\sigma + 1}}\right)$

* Taking the derivative with respect to ρ and σ:

$$\frac{\partial \beta_{OLS}}{\partial \rho} = \frac{2\phi(0)\sigma\left(\sigma^2 - 1\right)}{\left(\sigma^2 - 2\rho\sigma + 1\right)^{3/2}} > 0 \text{ if } \sigma^2 > 1$$
$$\frac{\partial \beta_{OLS}}{\partial \sigma} = \frac{4\phi(0)\sigma}{\sqrt{\sigma^2 + 1 - 2\rho\sigma}} - \frac{2\phi(0)\left(\sigma^2 - 1\right)(\sigma - \rho)}{\left(\sigma^2 - 2\rho\sigma + 1\right)^{3/2}} > 0 \text{ if } \rho < \frac{\sigma\left(\sigma^2 + 3\right)}{3\sigma^2 + 1}$$

Interpretation with respect to ρ:

▷ Higher (more positive) correlation reduces the magnitude of $U_1 - U_0$ which lowers ATT and increases ATUT.

▷ For β_{OLS}, recall that

$$\begin{aligned}
\beta_{OLS} &= \mathbb{E}[Y_1 | D = 1] - \mathbb{E}[Y_0 | D = 0] \\
&= \mathbb{E}[Y_1 | D = 1] - \mathbb{E}[Y_0 | D = 1] + \mathbb{E}[Y_0 | D = 1] - \mathbb{E}[Y_0 | D = 0] \\
&= ATT + \text{ selection bias}
\end{aligned}$$

When $\sigma > 1$, we have that β_{OLS} is increasing in ρ so the selection bias increases sufficiently to offset the decrease in ATT that stems from the increase in ρ.

Interpretation with respect to σ:

▷ For $\sigma > \rho$, an increase in σ raises the variance of $U_1 - U_0$. This implies that the tails of the resulting distribution get fatter, thus implying an increased ATT and a lowered ATUT.

▷ Increasing σ that satisfies $\rho < \frac{\sigma(\sigma^2+3)}{3\sigma^2+1}$ also results in an increase in β_{OLS} by widening the spread between U_1 and U_0.

Problem 2.5. Set $\sigma = 2$ and $\rho = 0.5$. Draw $N = 10,000$ pairs (U_0, U_1) and compute ATE, ATT, ATUT and β_{OLS}. Compute $\mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0]$ – what parameter does it correspond to? Repeat the exercise for $\sigma = 2, \rho = 0$ and $\sigma = 2, \rho = -0.5$. Also try fixing $\rho = 0.5$ and vary σ to verify your conclusions from the previous part.

Solution. We report the values below.

▷ $\sigma = 2, \rho = 0.5$

  * ATE: -0.00388 (simulated) and 0.000 (analytical)
  * ATT: 1.384 (simulated) and 1.382 (analytical)
  * ATUT: -1.388 (simulated) and -1.382 (analytical)
  * β_{OLS}: 1.383 (simulated) and 1.382 (analytical)
  * $\mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0]$: 1.383 which corresponds to β_{OLS}

▷ $\sigma = 2, \rho = 0$

  * ATE: 0.003 (simulated) and 0.000 (analytical)
  * ATT: 1.788 (simulated) and 1.784 (analytical)
  * ATUT: -1.789 (simulated) and -1.784 (analytical)
  * β_{OLS}: 1.076 (simulated) and 1.070 (analytical)
  * $\mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0]$: 1.376 which corresponds to β_{OLS}

▷ $\sigma = 2, \rho = -0.5$

  * ATE: -0.001 (simulated) and 0.000 (analytical)
  * ATT: 2.112 (simulated) and 2.111 (analytical)
  * ATUT: -2.211 (simulated) and -2.111 (analytical)
  * β_{OLS}: 0.907 (simulated) and 0.905 (analytical)
  * $\mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0]$: 0.907 which corresponds to β_{OLS}

▷ From the results above, we do verify that for a given fixed σ, increasing ρ lowers the ATT, raises ATUT, and raises β_{OLS}. This is in line with our theoretical expectations from the previous part. Similarly, to verify the remaining parts, I fix$\rho = 0.5$ and vary σ from 1 to 3. I find that this increases the ATT and lowers the ATUT. This is because the given values of σ and ρ meets the condition specified in the previous step and indeed verifies our theoretical results from above.

Problem 2.6. Claim: In this setup, $D \perp (Y_1, Y_0) \Leftrightarrow \rho = 0$. Argue whether this claim is correct or not. Support your conclusion with result from the previous part.

Solution. The claim is incorrect.

▷ First, $\rho = 0$ does not imply $D \perp (Y_1, Y_0)$. This is because we have $\beta_{OLS} > 0$ even when $\rho = 0$ from the previous part.

▷ Second, $D \perp (Y_1, Y_0)$ does not imply $\rho = 0$. Specifically, you can have ρ to be any value when $\sigma = 1$, so $D \perp (Y_1, Y_0)$ does not necessarily guarantee $\rho = 0$.

3 Bootstrapping Problem

We showed earlier that under i.i.d. assumption we can receive the following results:

$$\hat{\beta} \xrightarrow{p} \bar{\beta} = \mathbb{E}\left[X_i X_i'\right]^{-1} \mathbb{E}\left[X_i Y_i\right]$$
$$\sqrt{N}\left(\hat{\beta} - \bar{\beta}\right) \xrightarrow{d} \mathcal{N}\left(0, V\right)$$
$$\exists \hat{V} \xrightarrow{p} V$$
$$se\left(\hat{\beta}_k\right) = \sqrt{\frac{1}{N} diag\left(\hat{V}\right)_k}$$

Monte Carlo Simulations Consider the model:

$$Y_i = X_i'\beta + U_i$$
$$U_i | X_i \sim^{iid} \mathcal{N}\left(0, \sigma^2\right)$$

Problem 3.1. Define $\beta = (2,3)^\top, \sigma^2 = 4$. Generate $N = 10,000$ values for $X \in \mathbb{R}^2$ (constant and one more covariate). Using your value for σ^2 draw Us (you can make it completely independent of X) . Finally, compute Ys. Estimate $\hat{\beta}$ and its standard errors from your data using standard OLS formulas.

Solution. We implement the system in Python and estimate $\hat{\beta}$ as well as the standard errors using the standard OLS formula. We obtain:

$$\hat{\beta}_{OLS} = \left[\begin{array}{cc} 2.0515 & 2.9774 \end{array}\right]^\top, \quad \hat{se}\left(\hat{\beta}_k\right) = \left[\begin{array}{cc} 0.039338 & 0.0691162 \end{array}\right]^\top$$

Problem 3.2. Using X, β, and σ^2 from the previous part and draw $S = 10,000$ of $U^{(s)}$ and corresponding $Y^{(s)}$. For each $Y^{(s)}$, compute:

$$\hat{\beta}\left(s\right) = \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \left(\sum_{i=1}^{N} X_i Y_i^{(s)}\right)$$
$$\sqrt{\hat{Var}\left[\hat{\beta}_k^{(s)} | X_1, ..., X_N\right]} = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\hat{\beta}_k^{(s)}\right)^2 - \left(\frac{1}{S}\sum_{s=1}^{S}\hat{\beta}_k^{(s)}\right)^2} \underbrace{\xrightarrow{p}}_{?} se\left(\hat{\beta}_k | X_1, ..., X_N\right)$$

Justify the ? step in the last linte. Plot a histogram for the first component of $\beta^{(s)}$.

Solution. We implement the system in Python and compute the relevant numbers to obtain:

$$\frac{1}{S}\sum_{s=1}^{S}\hat{\beta}_{OLS}^{(s)} = \left[\begin{array}{cc} 2.00003 & 3.00002 \end{array}\right]^\top$$
$$\sqrt{\hat{Var}\left[\hat{\beta}_k^{(s)} | X_1, ..., X_N\right]} = \left[\begin{array}{cc} 0.03952962 & 0.06982055 \end{array}\right]^\top$$

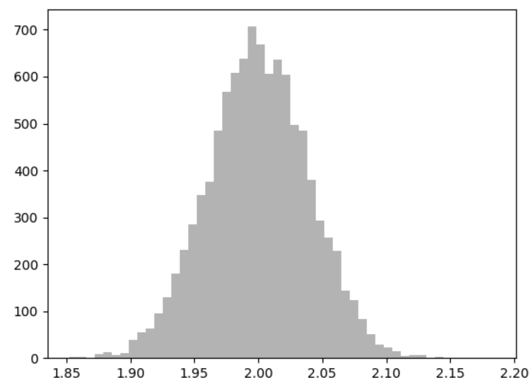
The ? step can be justified as the following:

▷ By Law of Large Numbers, we have

$$\frac{1}{S}\sum_{s=1}^{S}\left(\hat{\beta}_k^{(s)}\right)^2 \xrightarrow{p} \mathbb{E}\left[\left(\hat{\beta}_k^{s}\right)^2 | X_1, ..., X_N\right]$$
$$\frac{1}{S}\sum_{s=1}^{S}\hat{\beta}_k^{(s)} \xrightarrow{p} \mathbb{E}\left[\hat{\beta}_k^{s} | X_1, ..., X_N\right]$$

▷ Using the Continuous Mapping Theorem three times for $f(x) = x^2$, $f(x, y) = x - y$ and $f(x) = \sqrt{x}$ respectively, we arrive at the convergence in probability for the parameter of interest.

The histogram is plotted below:



Nonparameteric Bootstrap Roughly, the idea behind the bootstrap procedure is that we expect large sample of observed data to behave similarly to the population. Then, following the logic of Monte Carlo simulations, we want to draw sample from this population and conduct the inference (so we do not have to specify the data generating process as we did in the previous part).

Let us work through an example. Consider the RCT setup:

$$Y = DY_1 + (1 - D)\,Y_0$$
$$D \perp (Y_1, Y_0)$$

Problem 3.3. Define constant values for $Y_1 = 5$ and $Y_0 = 2$, assign $D \in \{0, 1\}$ randomly (define probability with which $D = 1, 0.5$ would be a good choice) to $N = 10,000$ individuals. Note we can write the initial outcome equation:

$$Y = Y_0 + D\underbrace{(Y_1 - Y_0)}_{\beta}$$

Estimate β using a standard OLS procedure, compute the standard errors, argue that OLS gives consistent coefficient estimates.

Solution. For the OLS to give consistent coefficient estimates, we need the exogeneity condition to hold. In this case, it does hold since:

$$\begin{aligned}\mathbb{E}\left[D\epsilon\right] &= \mathbb{E}\left[DY_0\right] - \mathbb{E}\left[D\right]\mathbb{E}\left[Y_0\right] \\ &= \mathbb{E}\left[D\right]\mathbb{E}\left[Y_0\right] - \mathbb{E}\left[D\right]\mathbb{E}\left[Y_0\right] = 0\end{aligned}$$

from the $D \perp (Y_1, Y_0)$ condition. We estimate β using the standard OLS procedure and obtain the following estimates:

$$\begin{aligned}\hat{\beta}_{OLS} &= \left[\begin{array}{cc} 1.994717 & 3.01641 \end{array}\right] \\ \hat{se}\left(\hat{\beta}_k\right) &= \left[\begin{array}{cc} 0.014199 & 0.020072 \end{array}\right]\end{aligned}$$

Problem 3.4. Now, we will create bootstrap samples. From the initial date that you generated draw $N = 10,000$ pairs (Y_i, D_i) choosing each of the original data pairs with probability $1/N$ (with replacement). Repeat this procedure a total of $S = 10,000$ times. Now you should have $S = 10,000$ samples of $N = 10,000$ observations each generated from the original sample. Repeat the computations from the Monte Carlo part and compute $\sqrt{\hat{Var}\left[\hat{\beta}^{(s)}\right]}$,plot histogram for $\hat{\beta}^{(s)}$. What would happen if you drew Y and D independently from the original sample, instead of as a pair? (hint: think about the value of coefficient on D)

Solution. We implement the system in Python and compute the relevant numbers to obtain:

$$\begin{aligned}\frac{1}{S}\sum_{s=1}^{S}\hat{\beta}_{OLS}^{(s)} &= \left[\begin{array}{cc} 1.994699 & 3.016365 \end{array}\right]^{\top} \\ \sqrt{\hat{Var}\left[\hat{\beta}_k^{(s)}|X_1, ..., X_N\right]} &= \left[\begin{array}{cc} 0.014328 & 0.02001449 \end{array}\right]^{\top}\end{aligned}$$

If we drew Y and D independently from the original sample, then we will be running a regression where there is nodependence around Y on D . Thus the β_0 estimate will simply be the mean of Y, i.e.

$$\mathbb{E}\left[Y\right] = 0.5\left(5 + 2\right) = 3.5$$

and the β_1 estimate will be very close to 1. We plot the histogram below. The first graph is for the first element of $\hat{\beta}^{(s)}$ and the second graph is for the second element of $\hat{\beta}^{(s)}$.