1 The Simple Roy Model

We consider a selection problem. Assume for any of the students in NTU, student i, the wage of going to a small company is w_{0i}, and the wage of going to a big company is w_{1i}.

The wage is modeled to be

$$w_{0i} = \mu_0 + \epsilon_{0i} \tag{1}$$
$$w_{1i} = \mu_1 + \epsilon_{1i} \tag{2}$$

We assume that $\epsilon_{0i} \sim \mathcal{N}(0, \sigma_0^2), \epsilon_{1i} \sim \mathcal{N}(0, \sigma_1^2), Cov(\epsilon_{0i}, \epsilon_{1i}) = \sigma_{01}$.

Let $D = 1$ be the student choosing to go to a big company. The student chooses to go to a big company if it offers a higher wage, so

$$D = 1\{w_{1i} > w_{0i}\} = 1\{\mu_1 + \epsilon_{1i} > \mu_0 + \epsilon_{0i}\}$$

We now observe the average wage of a student that goes to a big company, that is, $E[w_{1i} \mid D = 1]$. As a economist, we are interested in knowing *What will be the expected wage if this student didn't go to a big company, but went to a small company instead?* , in other words, what is $E[w_{0i} \mid D = 1]$?

This is never observed (a.k.a counterfactual), because we can't reverse time and tell the student to change its mind and go to a small company. However, we can somehow derive this unobserved potential outcome by using our knowledge in statistics.

1. Given $X \sim \mathcal{N}(0, 1)$, proof that

$$f(x \mid x > a) = \frac{\phi(x)}{1 - \Phi(a)}$$

   where $\phi(\cdot)$ is the PDF of a normal distribution, and $\Phi(\cdot)$ is the CDF. This is the PDF of a *truncated normal distribution*

2. Find the expectation value $E[X \mid X > a]$.

3. Following 2. Now $Y \sim \mathcal{N}(0, \sigma_y^2)$, and $Corr(X, Y) = \rho_{xy}$.
   Use the result

$$E[Y \mid X > a] = \frac{1}{\Pr(X > a)} \int_a^\infty E[Y \mid X = x] f_X(x) dx$$

   , prove that

$$E[Y \mid X > a] = \rho_{xy} \sigma_y E[X \mid X > a]$$

4. Now, consider the estimation we are interested in, $E[w_{0i} \mid D = 1]$. We rewrite it as

$$\begin{aligned} E[w_{0i} \mid D = 1] &= E[w_{0i} \mid w_{1i} > w_{0i}] \\ &= E[w_{0i} \mid \mu_1 + \epsilon_{1i} > \mu_0 + \epsilon_{0i}] \\ &= \mu_0 + E[\epsilon_{0i} \mid \epsilon_{1i} - \epsilon_{0i} > \mu_0 - \mu_1] \end{aligned}$$

   Define $\nu_i = \epsilon_{1i} - \epsilon_{0i}$, find $\sigma_v^2 = Var(\nu_i)$.

5. Prove that

$$E[w_{0i} \mid D = 1] = \mu_0 + \rho_{o,\nu}\sigma_0 E\left[\frac{\nu_i}{\sigma_\nu} \mid \frac{\nu_i}{\sigma_\nu} > \frac{\mu_0 - \mu_1}{\sigma_\nu}\right]$$

   Where $\rho_{0,\nu}$ is the correlation coefficient of ϵ_{0i} and ν_i.

   Hint. What is $\epsilon_{0i} \mid \nu_i \sim$?

6. Finally, prove that

$$E[w_{0i} \mid D = 1] = \mu_0 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\rho_{01} - \frac{\sigma_0}{\sigma_1}\right)\frac{\phi(z)}{1 - \Phi(z)}$$

   , where $z \equiv \frac{\mu_0 - \mu_1}{\sigma_v}$