Econometrics – Problem Set #2

Junwoo Yang

March 24, 2021

#2.16 Y is distributed $N(10, 100)$, and you want to calculate $P(Y \leq 5.8)$. Unfortunately, you do not have your textbook, and do not have access to a normal probability table like Appendix Table 1. However, you do have your computer and a computer program that can generate i.i.d. draws from the $N(10, 100)$ distribution. Explain how you can use your computer to compute an accurate approximation for $P(Y \leq 5.8)$.

Proof. We can generate i.i.d. random variables $Y_1, \cdots, Y_n$. Consider Bernoulli random variable X_i such that $X_i = 1$ if $Y_i \leq 5.8$, and $X_i = 0$ otherwise. Note that expected values of both X_i and $\overline{X}$ are $P(Y \leq 5.8)$, i.e., $E(X_i) = E(\overline{X}) = P(Y \leq 5.8)$. The low of large number says that $\overline{X}$ converges in probability to $P(Y \leq 5.8)$ as n goes to infinity. Therefore, we can get an accurate approximation for $P(Y \leq 5.8)$ by generating a myriad of samples and calculating $\overline{X}$. □

#2.18 In any year, the weather can inflict storm damage to a home. From year to year, the damage is random. Let Y denote the dollar value of damage in any given year. Suppose that in 95% of the years $Y = \$0$ but in 5% of the years $Y = \$30,000$.

(a) What are the mean and standard deviation of the damage in any year?

Proof. Since $P(Y = 0) = 0.95$, $P(Y = 30000) = 0.05$, the mean of Y is

$$\mu_Y = 0 \times P(Y = 0) + 30000 \times P(Y = 30000) = 1500.$$

The standard deviation of Y is

$$\begin{aligned}\sigma_Y &= \sqrt{E[(Y - \mu_Y)^2]} = \sqrt{(-1500)^2 \times 0.95 + 28500^2 \times 0.05} \\ &= \sqrt{42750000} = 6538.\end{aligned}$$

□

(b) Consider an insurance pool of 120 people whose homes are sufficiently dispersed so that, in any year, the damage to different homes can be viewed as independently distributed random variables. Let $\overline{Y}$ denote the average damage to these 120 homes in a year.

(i) What is the expected value of the average damage $\overline{Y}$?

Proof. $Y_1, \cdots, Y_{120} \sim (\mu_Y, \sigma_Y^2)$ i.i.d., $\overline{Y} \sim (\mu_Y, \sigma_Y^2/n)$. $\therefore E(\overline{Y}) = \mu_Y = 1500$. □

(ii) What is the probability that $\overline{Y}$ exceeds \$3,000?

Proof. By CLT, we can approximate that probability:

$$\begin{aligned}P(\overline{Y} > 3000) = 1 - P(\overline{Y} \leq 3000) &= 1 - P\left(\frac{\overline{Y} - \mu_Y}{\sigma_Y/\sqrt{n}} \leq \frac{3000 - \mu_Y}{\sigma_Y/\sqrt{n}}\right) \\ &\approx 1 - \Phi\left(\frac{1500}{6538/\sqrt{120}}\right) = 0.005983.\end{aligned}$$

□

#2.20 Consider three random variables, X, Y, and Z. Suppose that Y takes on k values $y_1, \cdots, y_k$; that X takes on l values $x_1, \cdots, x_l$; and that Z takes on m values $z_1, \cdots, z_m$. The joint probability distribution of X, Y, Z is $P(X = x, Y = y, Z = z)$, and the conditional probability distribution of Y given X and Z is $P(Y = y|X = x, Z = z) = \frac{P(Y=y,X=x,Z=z)}{P(X=x,Z=z)}$.

(a) Explain how the marginal probability that $Y = y$ can be calculated from the joint probability distribution.

Proof. $P(Y = y) = \sum_{i=1}^{l} \sum_{j=1}^{m} P(X = x_i, Y = y, Z = z_j)$. □

(b) Show that $E(Y) = E[E(Y|X, Z)]$.

Proof. Note that $E[E(Y|X, Z)]$ is function of X and Z, while $E(Y|X, Z)$ is function of Y given X and Z.

$$\begin{aligned}
E[E(Y|X,Z)] &= E\left[\sum_{h=1}^{k} y_h P(Y = y_h|X = x_i, Z = z_j)\right] \\
&= \sum_{i=1}^{l}\sum_{j=1}^{m}\left[\sum_{h=1}^{k} y_h P(Y = y_h|X = x_i, Z = z_j)\right] P(X = x_i, Z = z_j) \\
&= \sum_{h=1}^{k} y_h \sum_{i=1}^{l}\sum_{j=1}^{m} P(Y = y_h|X = x_i, Z = z_j)P(X = x_i, Z = z_j) \\
&= \sum_{h=1}^{k} y_h \sum_{i=1}^{l}\sum_{j=1}^{m} P(X = x_i, Y = y_h, Z = z_j) \\
&= \sum_{h=1}^{k} y_h P(Y = y_h) \\
&= E(Y)
\end{aligned}$$

□

#2.22 Suppose you have some money to invest, for simplicity \$1, and you are planning to put a fraction w into a stock market mutual fund and the rest, $1 - w$, into a mutual fund. Suppose that \$1 invested in a stock fund yields R_s after one year and that \$1 invested in mutual fund yields R_b. Suppose that R_s is random with mean 0.06 and standard deviation 0.09, and suppose that R_b is random with mean 0.04 and standard deviation 0.05. The correlation between R_s and R_b is 0.3. If you place a fraction w of your money in the stock fund and the rest, $1 - w$, in the mutual fund, then the return on your investment is $R = wR_s + (1 - w)R_b$.

(a) Suppose that $w = 0.2$. Compute the mean and standard deviation of R.

Proof.

$$\begin{aligned}
\mu_R &= E[wR_s + (1 - w)R_b] = 0.2\mu_{R_s} + 0.8\mu_{R_b} = 0.2 \times 0.06 + 0.8 \times 0.04 = 0.044 \\
\sigma_R^2 &= \mathrm{Var}(wR_s + (1 - w)R_b) \\
&= w^2\sigma_{R_s}^2 + (1 - w)^2\sigma_{R_b}^2 + 2w(1 - w)\underbrace{\mathrm{Corr}(R_s, R_b)\sigma_{R_s}\sigma_{R_b}}_{\mathrm{Cov}(R_s,R_b)} \\
&= 0.2^2 \times 0.09^2 + 0.8^2 \times 0.05^2 + 2 \times 0.2 \times 0.8 \times 0.3 \times 0.09 \times 0.05 = 0.002356 \\
\sigma_R &= \sqrt{0.002356} = 0.0485
\end{aligned}$$

□

(b) Suppose that $w = 0.8$. Compute the mean and standard deviation of R.

Proof. Similarly, $\mu_R = 0.056$, $\sigma_R = 0.0756$ □

(c) What value of w makes the mean of R as large as possible? What is the standard deviation of R for this value of w?

Proof. Since $\mu_{R_s} > \mu_{R_b}$, μ_R has a maximum of $\mu_{R_s} = 0.06$ when $w = 1$. At this time, σ_R is just $\sigma_{R_s} = 0.09$. □

(d) What is the value of w that minimizes the standard deviation of R?

Proof. When σ_R^2 is minimum, σ_R is minimum. The derivative of σ_R^2 w.r.t. w is

$$\begin{aligned}\frac{d\sigma_R^2}{dw} &= \frac{d}{dw}(w^2\sigma_{R_s}^2 + (1-w)^2\sigma_{R_b}^2 + 2w(1-w)\operatorname{Corr}(R_s, R_b)\sigma_{R_s}\sigma_{R_b}) \\ &= 2w\sigma_{R_s}^2 - 2(1-w)\sigma_{R_b}^2 + (2-4w)\operatorname{Corr}(R_s, R_b)\sigma_{R_s}\sigma_{R_b} \\ &= 2w \times 0.09^2 - 2(1-w) \times 0.05^2 + (2-4w) \times 0.3 \times 0.09 \times 0.05 \\ &= 0.0158w - 0.0023.\end{aligned}$$

Thus, σ_R is minimized when $w = 0.14557$. □

#2.23 This exercise provides an example of a pair of random variables, X and Y, for which the conditional mean of Y given X depends on X but $\operatorname{Corr}(X, Y) = 0$. Let X and Z be two independently distributed standard normal random variables, and let $Y = X^2 + Z$.

(a) Show that $E(Y|X) = X^2$.

Proof. $E(Y|X) = E(X^2 + Z|X) = E(X^2|X) + E(Z|X) = X^2 + E(Z) = X^2 + 0 = X^2$. □

(b) Show that $\mu_Y = 1$.

Proof. $E(Y) = E(X^2 + Z) = E(X^2) + E(Z) = E(X^2) = \operatorname{Var}(X) + E(X)^2 = 1$. □

(c) Show that $E(XY) = 0$.

Proof. In #2.13 (b), Problem Set #1, I proved that the third moment of normal random variable centered at zero is zero.

$$\begin{aligned}E(XY) &= E(X^3 + XZ) = E(X^3) + E(XZ) \\ &= 0 + E[(X - \mu_X)(Z - \mu_Z)] = \operatorname{Cov}(X, Z) = 0.\end{aligned}$$ □

(d) Show that $\operatorname{Cov}(X, Y) = 0$ and thus $\operatorname{Corr}(X, Y) = 0$.

Proof. $\operatorname{Cov}(X, Y) = 0 \Rightarrow \operatorname{Corr}(X, Y) = 0$ is trivial by definition of correlation.

$\operatorname{Cov}(X, Y) = E[X(X^2 + Z - 1)] = E(X^3 + XZ - X) = E(X^3) + E(XZ) - E(X) = 0.$ □

#2.26 Suppose that $Y_1, Y_2, \cdots, Y_n$ are random variables with a common mean μ_Y; a common variance σ_Y^2; and the same correlation ρ (so that the correlation between Y_i and Y_j is equal to ρ for all pairs i and j, where $i \neq j$).

(a) Show that $\mathrm{Cov}(Y_i, Y_j) = \rho\sigma_Y^2$ for $i \neq j$.

Proof. $\mathrm{Cov}(Y_i, Y_j) = \mathrm{Corr}(Y_i, Y_j)\sigma_{Y_i}\sigma_{Y_j} = \rho\sigma_Y^2$. □

(b) Suppose that $n = 2$. Show that $E(\overline{Y}) = \mu_y$ and $\mathrm{Var}(\overline{Y}) = \frac{1}{2}\sigma_Y^2 + \frac{1}{2}\rho\sigma_Y^2$.

Proof.

$$
\begin{aligned}
E(\overline{Y}) &= E\left(\frac{Y_1 + Y_2}{2}\right) = \frac{E(Y_1) + E(Y_2)}{2} = \frac{2\mu_Y}{2} = \mu_Y \\
\mathrm{Var}(\overline{Y}) &= \frac{\mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) + 2\,\mathrm{Cov}(Y_1, Y_2)}{2^2} = \frac{2\sigma_Y^2 + 2\rho\sigma_Y^2}{4} = \frac{\sigma_Y^2 + \rho\sigma_Y^2}{2}.
\end{aligned}
$$

□

(c) For $n \geq 2$, show that $E(\overline{Y}) = \mu_Y$ and $\mathrm{Var}(\overline{Y}) = \frac{1}{n}\sigma_Y^2 + \frac{n-1}{n}\rho\sigma_Y^2$.

Proof. Note that $E(Y_i^2) = \sigma_Y^2 + \mu_Y^2$ and $E(Y_iY_j) = \mathrm{Cov}(Y_i, Y_j) + \mu_Y^2 = \rho\sigma_Y^2 + \mu_Y^2$.

$$
\begin{aligned}
E(\overline{Y}) &= E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{n\mu_Y}{n} = \mu_Y \\
E(\overline{Y}^2) &= E\left(\frac{1}{n^2}\sum_{i,j} Y_iY_j\right) = \frac{1}{n^2}\sum_{i,j} E(Y_iY_j) \\
&= \frac{1}{n^2}\left\{\sum_{i=j} E(Y_iY_j) + \sum_{i\neq j} E(Y_iY_j)\right\} \\
&= \frac{1}{n^2}\left\{\sum_{i=1}^{n} E(Y_i^2) + (n^2 - n)(\rho\sigma_Y^2 + \mu_Y^2)\right\} \\
&= \frac{1}{n^2}\left\{n(\sigma_Y^2 + \mu_Y^2) + n(n-1)(\rho\sigma_Y^2 + \mu_Y^2)\right\} \\
&= \frac{1}{n}\left\{\sigma_Y^2 + (n-1)\rho\sigma_Y^2 + n\mu_Y^2\right\} \\
&= \frac{1}{n}\sigma_Y^2 + \frac{n-1}{n}\rho\sigma_Y^2 + \mu_Y^2 \\
\mathrm{Var}(\overline{Y}) &= E(\overline{Y}^2) - E(\overline{Y})^2 = \frac{1}{n}\sigma_Y^2 + \frac{n-1}{n}\rho\sigma_Y^2
\end{aligned}
$$

where $(n^2 - n)$ is the number of all pairs except cases with $i = j$. □

(d) When n is very large, show that $\mathrm{Var}(\overline{Y}) \simeq \rho\sigma_Y^2$.

Proof. By above result, $\mathrm{Var}(\overline{Y})$ converges to $\rho\sigma_Y^2$ as n goes to infinity, i.e.,

$$
\lim_{n\to\infty} \mathrm{Var}(\overline{Y}) = \lim_{n\to\infty}\left(\frac{1}{n}\sigma_Y^2 + \frac{n-1}{n}\rho\sigma_Y^2\right) = \rho\sigma_Y^2.
$$

□