ECON 21020 PSET 1

Alex Urquhart

I discussed ideas with Nicholas Limon on this assignment.

Problem #1:

We are given that X has Bernoulli distribution, so $P(x = 1) = p$ and $P(x = 0) = 1 - p$. We also are given that $Z = 3^x - 1$.

(a)

Because of the Bernoulli distribution, X can only take 0 and 1 as values. We then calculate the expected value of Z:

$$E[Z] = \sum_{i=0}^{1} z(x_i)p(z_i)$$

$$E[Z] = \sum_{i=0}^{1} (3^x - 1)p(z_i)$$

$$E[Z] = (3^0 - 1)(1 - p) + (3^1 - 1)(p)$$

$$E[Z] = 0 + 2p = 2p$$

(b)

$$E[Z^2] = \sum_{i=0}^{1} z(x_i)^2 p(z_i)$$

$$E[Z^2] = (3^0 - 1)^2(1 - p) + (3^1 - 1)^2(p)$$

$$E[Z^2] = 0(1 - p) + (2)^2(p)$$

$$E[Z^2] = 0 + 4p = 4p$$

(c)

$$Var(Z) \approx E[Z^2] - (E[Z])^2$$

$$Var(Z) \approx 4p - (2p)^2$$

$$Var(Z) \approx 4p - 4p^2$$

$$Var(Z) \approx 4p(1 - p)$$

While the shortcut formula is an approximation, we can take it as is.

Problem #2:

(a)

$$E[Y] = \sum_{i=0}^{1} y_i p(y_i | X = x_i)$$

$$E[Y] = 0(0.035) + 1(0.965)$$

$$E[Y] = 0.965$$

(b)

$$1 - E[Y] = 1 - 0.965 = 0.035$$

On the probability table, this value represents the proportion of the total population that is unemployed (Y = 0), regardless of degree status (X = 0 or X = 1). Thus, the unemployment rate can be represented by $1 - E[Y]$.

(c)

$$E[Y|X = 1] = \sum_{i=0}^{1} y_i p(y_i | x = 1)$$

$$E[Y|X = 1] = 0(\frac{0.009}{0.398}) + 1(\frac{0.389}{0.398})$$

$$E[Y|X = 1] = \frac{0.389}{0.398} \approx 0.977$$

$$E[Y|X=0] = \sum_{i=0}^{1} y_i p(y_i|x=0)$$

$$E[Y|X=0] = 0(\frac{0.026}{0.602}) + 1(\frac{0.576}{0.602})$$

$$E[Y|X=0] = \frac{0.576}{0.602} \approx 0.957$$

(d)

The degree-specific unemployment rate can be calculated by $P(Y=0|X=a)$, where $a=1$ for college grads and $a=0$ for non-graduates.

For college grads:

$$P(Y=0|X=1) = \frac{0.009}{0.398} \approx 0.023$$

For non-graduates:

$$P(Y=0|X=0) = \frac{0.026}{0.602} \approx 0.043$$

(e)

To calculate both probabilities, we need to calculate $P(X|Y=0)$, where $X=1$ for college grads and $X=0$ for non-graduates.

For college grads:

$$P(X=1|Y=0) = \frac{0.009}{0.035} \approx 0.257$$

For non-graduates:

$$P(X=0|Y=0) = 1 - P(X=1|Y=0)$$

$$P(X=0|Y=0) = 1 - 0.257$$

$$P(X=0|Y=0) = 0.743$$

(f)

For X and Y to be independent, they must fulfill the following condition:

$$P(X \cap Y) = P(X)P(Y)$$

We'll test this in the case where $X = Y = 1$.

$$P(X = 1 \cap Y = 1) \stackrel{?}{=} P(X = 1)P(Y = 1)$$

$$0.389 \stackrel{?}{=} 0.398 * 0.965$$

$$0.389 \neq 0.384$$

Thus, X and Y are not independent.

Problem #3:

We are given that $E[X] = 48.8, \sigma_x = 12.1$, which tells us that $Var(X) = 12.1^2 = 146.41$. We also know that $Y(X) = 1000X$, so we can use the scalar properties of expected value and variance:

$$E[1000X] = 1000E[X] = 1000 * 48.8 = 48800$$

$$Var(1000X) = 1000^2 Var(X) = 10^6 * 146.41 = 1.4641 * 10^8$$

We can also check the standard deviation:

$$\sigma_x = \sqrt{1.4641 * 10^8} = 12100$$

Problem #4:

We are given that $E[GPA|SAT] = 0.70 + 0.002(SAT)$.

(a)

$$E[GPA|SAT = 750] = 0.70 + 0.002(750)$$

$$E[GPA|SAT = 750] = 2.2$$

$$E[GPA|SAT = 1500] = 0.70 + 0.002(1500)$$

$$E[GPA|SAT = 1500] = 3.7$$

(b)

By the Law of Iterated Expectations, we can establish the following relationship:

$$E[E[GPA|SAT] = E[GPA]$$

Thus, we can use the given formula with $SAT = 1000$ to find $E[GPA]$.

$$E[GPA] = E[0.07 + 0.002(1000)]$$

$$E[GPA] = E[2.7] = 2.7$$

Problem #5:

$$MSE = E[(Y - g(x))^2]$$

(a)

$$U = Y - E[Y|X]$$

$$E[U|X] = E[(Y - E[Y|X])|X]$$

By the properties of conditional expectation, we know that $E[g(x) + h(x)Y|X] = g(x) + h(x)E[Y|X]$. Thus, taking $g(x) = -E[Y|X]$ and $h(x) = 1$,

$$E[U|X] = -E[Y|X] + E[Y|X]$$

$$E[U|X] = 0$$

(b)

By the Law of Iterated Expectation, $E[U * h(X)] = E[E[U * h(X)|X]]$.

We know that $E[U|X] = 0$, so we can use the scalar property of expected value, $E[bx] = bE[X]$, to establish that $E[U * h(X)|X] = h(X)E[U|X]$. Thus, $E[U * h(X)|X] = h(X) * 0 = 0$.

We are then left with $E[U * h(X)] = E[0] = 0$.

(c)

$$E[(Y - g(X))^2] \stackrel{?}{\geq} E[(Y - E[Y|X])^2]$$

Let $V = Y - g(X)$, $h(X) = g(X) - E[Y|X]$. We then plug those values in, along with the known equation $U = Y - E[Y|X]$:

$$V = Y - E[Y|X] - h(x) = U - h(x)$$

$$V^2 = (U - h(x))^2$$

$$V^2 = U^2 - 2Uh(x) + h(x)^2$$

Now, we can take the expected values on both sides of the equation.

$$E[V^2] = E[U^2] + E[-2U * h(X)] + E[h(x)^2]$$

We've already established that $E[U * h(X)] = 0$. With this result and the scalar property, we know that $-2E[U * h(X)] = -2(0) = 0$, and we can take out that term:

$$E[V^2] = E[U^2] + E[h(x)^2]$$

We also know that $h(x)^2 \geq 0$, since a squared value cannot be negative. Thus, we can confirm that

$$E[V^2] = E[U^2] + E[h(x)^2] \geq E[U^2]$$

,

Which confirms these subsequent relationships:

$$E[V^2] \geq E[U^2]$$

$$E[(Y - g(X))^2] \geq E[(Y - E[Y|X])^2]$$

.

(d)

This inequality proves that we can use $E[Y|X]$ as our predictor to minimize the MSE, since the smallest possible value of the MSE is when $g(x) = E[Y|X]$.

Problem #6:

We know that X is distributed uniformly along [-1,1], and that $Y = X^2$. Thus, we can set up the covariance formula and plug in Y:

$$Cov[X,Y] = E[XY] - E[X]E[Y]$$

$$Cov[X,Y] = E[X^3] - E[X]E[X^2]$$

We can then use the pdf of the uniform distribution to solve for the expected values:

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx$$

$$Cov[X,Y] = \int_{-\infty}^{\infty} x^3(\frac{1}{2})dx - (\int_{-\infty}^{\infty} x(\frac{1}{2})dx) * (\int_{-\infty}^{\infty} x^2(\frac{1}{2})dx)$$

We can change the bounds to [-1,1], take out 1/2 from each integral, and combine them to simplify:

$$Cov[X,Y] = \frac{1}{2}\int_{-1}^{1} x^3 - x * x^2 dx$$

$$Cov[X,Y] = \frac{1}{2}\int_{-1}^{1} x^3 - x^3 dx$$

$$Cov[X,Y] = \frac{1}{2}\int_{-1}^{1} 0dx$$

$$Cov[X,Y] = 0$$

To check for independence, we can test whether Y is mean independent of X. This would require that $E[Y|X]$ is constant for all values of X.

$$E[Y|X=1] = E[X^2|X=1] = 1$$

$$E[Y|X=0] = E[X^2|X=0] = 0$$

Since we got two different values, Y is NOT mean independent of X, and X and Y are NOT independent.

Problem #7:

(a)

By the Law of Iterated Expectations, $E[U] = E[E[U|X]]$. We can then plug in $E[U|X] = 1$ to get $E[U] = E[1] = 1$.

(b)

Because $E[U|X] = 1$ (a constant value), we know that U is mean independent of X. Thus, we can use the relationship $E[XU] = E[X]E[U] = 1(2) = 2$.

(c)

$$Cov(X, U) = E[XU] - E[X]E[U]$$

Using our result from (b), where $E[X]E[U] = E[XU]$,

$$Cov(X, U) = 2 - 2 = 0$$

(d)

Since the value $E[U|X]$ is constant, we know that U is mean independent of X.

(e)

$$Corr(X, U) = \frac{Cov(X, U)}{\sqrt{Var(X)Var(U)}}$$

$$Corr(X, U) = \frac{0}{\sqrt{Var(X)Var(U)}}$$

$$Corr(X, U) = 0$$

X and U are uncorrelated.

(f)

We can determine whether X is independent of U by solving $E[X|U]$. To do this, however, we would need information about the pdfs of X and U. Thus, we cannot make a determination like this.