DIS 14A

1 Working with the Law of Large Numbers

(a) A fair coin is tossed and you win a prize if there are more than 60% heads. Which is better: 10 tosses or 100 tosses? Explain.

(b) A fair coin is tossed and you win a prize if there are more than 40% heads. Which is better: 10 tosses or 100 tosses? Explain.

(c) A coin is tossed and you win a prize if there are between 40% and 60% heads. Which is better: 10 tosses or 100 tosses? Explain.

(d) A coin is tossed and you win a prize if there are exactly 50% heads. Which is better: 10 tosses or 100 tosses? Explain.

Solution:

(a) 10 tosses. By LLN, the sample mean should have higher probability to be close to the population mean as n increases. Therefore the average proportion of coins that are heads should be closer to 0.50, and has a lower chance of being greater than 0.60 if there are 100 tosses compared with 10 tosses.

(b) 100 tosses. Based on the first part, consider the inverse of the event "more than 60% heads" and the symmetry of heads and tails.

(c) 100 tosses. Based on the first part, consider the union of the events "more than 60% heads" and "more than 60% tails" ("less than 40% heads").

(d) 10 tosses. Compare the probability of getting equal number of heads and tails between $2n$ and

$2n+2$ tosses.

$$\begin{aligned}
\mathbb{P}[n \text{ heads in } 2n \text{ tosses}] &= \binom{2n}{n}\frac{1}{2^{2n}} \\
\mathbb{P}[n+1 \text{ heads in } 2n+2 \text{ tosses}] &= \binom{2n+2}{n+1}\frac{1}{2^{2n+2}} = \frac{(2n+2)!}{(n+1)!(n+1)!}\cdot\frac{1}{2^{2n+2}} \\
&= \frac{(2n+2)(2n+1)2n!}{(n+1)(n+1)n!n!}\cdot\frac{1}{2^{2n+2}} \\
&= \frac{2n+2}{n+1}\cdot\frac{2n+1}{n+1}\binom{2n}{n}\cdot\frac{1}{2^{2n+2}} < \left(\frac{2n+2}{n+1}\right)^2\binom{2n}{n}\cdot\frac{1}{2^{2n+2}} \\
&= 4\binom{2n}{n}\cdot\frac{1}{2^{2n+2}} = \binom{2n}{n}\frac{1}{2^{2n}} = \mathbb{P}[n \text{ heads in } 2n \text{ tosses}]
\end{aligned}$$

The larger n is, the less probability we'll get 50% heads. □

Note: By Stirling's approximation, $\binom{2n}{n}2^{-2n}$ is roughly $(\pi n)^{-1/2}$ for large n.

2 Uniform Probability Space

Let $\Omega = \{1,2,3,4,5,6\}$ be a uniform probability space. Let also $X(\omega)$ and $Y(\omega)$, for $\omega \in \Omega$, be the random variables defined in the table:

Table 1: All the rows in the table correspond to random variables.

ω	1	2	3	4	5	6	$\mathbb{E}[\cdot]$
$X(\omega)$	0	0	1	1	2	2	
$Y(\omega)$	0	2	3	5	2	0	
$X^2(\omega)$							
$Y^2(\omega)$							
$XY(\omega)$							
$L[Y \mid X](\omega)$							

(a) Fill in the blank entries of the table. In the column to the far right, fill in the expected value of the random variable.

(b) Are the variables correlated or uncorrelated? Are the variables independent or dependent?

(c) Calculate $\mathbb{E}[(Y - L[Y \mid X])^2]$.

Solution:

(a) See the following table:

ω	1	2	3	4	5	6	$\mathbb{E}[\cdot]$
$X(\omega)$	0	0	1	1	2	2	1
$Y(\omega)$	0	2	3	5	2	0	2
$X^2(\omega)$	0	0	1	1	4	4	$5/3$
$Y^2(\omega)$	0	4	9	25	4	0	7
$XY(\omega)$	0	0	3	5	4	0	2
$L[Y \mid X](\omega)$	2	2	2	2	2	2	2

The third, fourth, and fifth rows can be calculated directly from the corresponding X and Y values. Recall that

$$L[Y \mid X] = \frac{\text{cov}(X,Y)}{\text{var}(X)}\big(X - \mathbb{E}(X)\big) + \mathbb{E}(Y).$$

But $\text{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 2 - (1)(2) = 0$, so $L[Y \mid X] = \mathbb{E}(Y) = 2$ for all ω.

(b) Since $\text{cov}(X,Y) = 0$, the variables are uncorrelated. But, we see that $\mathbb{P}(Y = 0) = 1/3$ and $\mathbb{P}(Y = 0 \mid X = 3) = 0$, so the two variables are not independent. Recall that independence implies uncorrelation, but the converse is not true.

(c)

$$\mathbb{E}[(Y - L[Y \mid X])^2] = \mathbb{E}[(Y - 2)^2] = \frac{4+0+1+9+0+4}{6} = 3$$

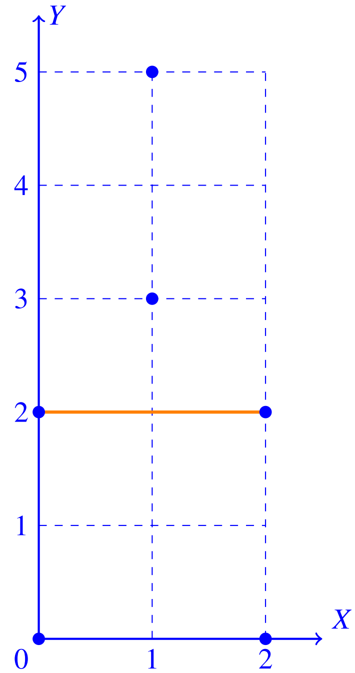

Figure 1: Visualization of regression. The circles are the (X,Y) points. The orange line is the LLSE.

3 LLSE

We have two bags of balls. The fractions of red balls and blue balls in bag A are $2/3$ and $1/3$ respectively. The fractions of red balls and blue balls in bag B are $1/2$ and $1/2$ respectively. Someone gives you one of the bags (unmarked) uniformly at random. You then draw 6 balls from that same bag with replacement. Let X_i be the indicator random variable that ball i is red. Now, let us define $X = \sum_{1\le i\le 3} X_i$ and $Y = \sum_{4\le i\le 6} X_i$. Find $L(Y \mid X)$. *Hint*: Recall that

$$L(Y \mid X) = \mathbb{E}(Y) + \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X)}\big(X - \mathbb{E}(X)\big).$$

Also remember that covariance is bilinear.

Solution:

Note that although the indicator random variables are not independent, we can still apply linearity of expectation. By symmetry, we also know that each indicator follows the same distribution.

Therefore:

$$\begin{aligned}
\mathbb{E}[X] = \mathbb{E}[Y] = 3 \cdot \mathbb{E}(X_1) = 3 \cdot \mathbb{P}(X_1 = 1) = 3 \cdot \left(\frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{2}\right) = \frac{7}{4}. \\
\operatorname{cov}(X, Y) &= \operatorname{cov}\Big(\sum_{1 \le i \le 3} X_i, \sum_{4 \le j \le 6} X_j\Big) = 9 \cdot \operatorname{cov}(X_1, X_4) \\
&= 9 \cdot \big(\mathbb{E}(X_1 X_4) - \mathbb{E}(X_1) \cdot \mathbb{E}(X_4)\big). \\
\mathbb{E}(X_1 X_4) - \mathbb{E}(X_1)\,\mathbb{E}(X_4) &= \mathbb{P}(X_1 = 1, X_4 = 1) - \mathbb{P}(X_1 = 1)^2 \\
&= \left[\frac{1}{2} \cdot \left(\frac{2}{3}\right)^2 + \frac{1}{2} \cdot \left(\frac{1}{2}\right)^2\right] - \left[\frac{1}{2} \cdot \left(\frac{2}{3}\right) + \frac{1}{2} \cdot \left(\frac{1}{2}\right)\right]^2 = \frac{1}{144}. \\
\operatorname{var}(X) &= \operatorname{cov}\Big(\sum_{1 \le i \le 3} X_i, \sum_{1 \le j \le 3} X_j\Big) \\
&= 3 \cdot \operatorname{var}(X_1) + 6 \cdot \operatorname{cov}(X_1, X_2) = 3\big(\mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2\big) + 6 \cdot \frac{1}{144} \\
&= 3\left[\frac{7}{12} - \left(\frac{7}{12}\right)^2\right] + 6 \cdot \frac{1}{144} = \frac{111}{144}.
\end{aligned}$$

So,

$$L(Y \mid X) = \frac{7}{4} + \frac{9}{111}\left(X - \frac{7}{4}\right) = \frac{3}{37} X + \frac{119}{74}.$$