2024 Spring –SPCO HW#2

1. Show that $E[(X-a)^2]=Var(X)+(\mu-a)^2$.

2. Let X be a discrete random variable with the following probability mass function (pmf):

$$P(X=x;\lambda)=\frac{\lambda^{2x}}{(2x)!\cosh(\lambda)} \text{ for } x=0, 1, 2, 3,\ldots$$

   (1) Verify $\Sigma P(X)=1$ (Hint: Taylor series of $\cosh(\lambda)$)

   (2) For $\lambda=30$, calculate the weighted average of X (weighted by its associated probability) and plot the pmf and cdf for $X=0, 1, 2, \ldots, 60$ using Excel. Do the same for $\lambda=50$. Could you guess the expected value for this random variable?

3. Show that the variance of a binomial random variable is $np(1-p)$.

4. Show the variance of a Poisson random variable is λ based on its relationship with the binomial distribution. (Note: the Poisson distribution model is the only probability distribution with its mean equal to its variance)

5. Using excel, plot the Binomial distributions with $(n, p)=$ (10, 0.4), (40, 0. 1), (100, 0.04), (500, 0.008), (1000, 0.004) and a Poisson distribution with $\lambda=4$ on the same chart. What do you observe from the differences among these distributions?

6. Under the identical and independent (iid) assumption, what would be the probability for a number to appear in a draw (期) of Power Lotto 638's first-set winning numbers? Assuming a Binomial distribution, calculate the expected value and variance of a number's appearances in $n=50$, 100, and 500 draws of Power Lotto first-set winning numbers. Compare the "average" and "sample variance" calculated for 50, 100, and 500 runs in Problem 2 of HW#1 to the expected value and variance of the Binomial probability model.

7. Find from your real-life experience a random phenomenon that can be likely modeled with Poisson distribution. Why?

8. 24Spring-HW2.xls is the actual statistics for the number of deaths per month caused by traffic accidents in different regions of Taiwan. Calculate the average and sample variance for the number of deaths for each region, respectively. Describe what you observed from these averages and sample variances. What would be the probability model that can best describe the number of deaths per month for each region based on the average and the sample variance?