EE 214 Machine Problem 1.3

Paul Vincent S. Nonat
Electrical and Electronics Engineering Institute
University of the Philippines Diliman
paul.vincent.nonat@eee.upd.edu.ph

1. The Problem

The Pratt-Woodruff experiment where some 32 students were made to guess five cards with different ESP symbols. The subject picks up one card, concentrates, and guesses the symbol on the card. Out of 60,000 guesses, 12,489 were correct. The probability of this happening modeled as a binomial distribution, given by

$$P(X \geq 12,489) = \sum_{k=12,489}^{60,000} \binom{60,000}{k} (\frac{1}{5})^k (\frac{4}{5})^{60,000-k} \tag{1}$$

Instead of using the De Moivre-Laplace Limit Theorem (since direct calculation can cause overflow problems), this probability can be calculated using a different method. Take the log of the right hand side, i.e. from,

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \tag{2}$$

Then use the relationship between the factorial and the gamma function $x! = \Gamma(x+1)$. In Python 3, the *math function gamma* computes the gamma function.

Write a Python program *prat woodruff.py* containing the function *binpmf(n,p)* to compute the above probability using this method. Plot the probability curve.

2. The Solution

The problem can be addressed using *Binomial(n,p) Random Variable*, in which

$$P_X(x) = \binom{n}{k} p^k (1-p)^{n-k} \tag{3}$$

The problem asks us to find $P_X(X \geq k)$. To solve this, we take the $\sum_k^n P(X = k)$ in order to get its probability. However, taking this method is compute-intensive, which can result in computation overflow. Hence, we will represent the $P(X = k)$ probability function using natural logarithm given by

$$\ln[P(X = k)] = \ln(n!) - \ln(k!) - \ln(n-k)! + k\ln(p) + (n-k)\ln(1-p) \tag{4}$$

Since $n = 60,000$, therefore taking $n!$ and $(n-k)!$ will also cause computation overflow due to high number of recurrent multiplication. To prevent this from occurring, we introduce to use the $\Gamma()$ function to solve this factorial overflow calculation issue.

$$\ln(n!) = \ln[\Gamma(n+1)] \tag{5}$$

Substituting this relationship to eqn. 4, we get the following

$$\ln[P(k)] = \ln[\Gamma(n+1)] - \ln[\Gamma(k+1)] - \ln[\Gamma(n+k+1)] + k\ln(p) + (n-k)\ln(1-p) \tag{6}$$

The function must be transformed back to its original form after computing for their natural logarithm counterpart, to calculate for its PMF using

$$P(X \geq k) = \sum_{k}^{n} e^{l \ln[P(k)]} \tag{7}$$

3. Results and Discussion

To compute for the probability of Pratt-Woodruff experiment, eqn. 5-7 were used as the basis of the python program to crunch the numbers from k to n.

```
def ln_factorial(n):
    ln_fac=0
    ln_fac=math.lgamma(n+1)
    return ln_fac

def binpmf(n,p,k):
    p_X_k=np.zeros(((n-k+1),2))
    print("Calculating ln("+str(n)+")")
    ln_n_fac=ln_factorial(n)
    z=0
    for j in range(k,n+1,1):
        ln_k_fac=ln_factorial(j)
        ln_nk_fac=ln_factorial(n-j)
        ln_X_p_k= ln_n_fac - ln_k_fac - ln_nk_fac
        + (j*math.log(p)) + ((n-j)*math.log(1-p))
        if j==0:
            p_X_k[z][0]= math.exp(ln_X_p_k)
        else:
            p_X_k[z][0]=  p_X_k[z-1][0]+math.exp(ln_X_p_k)
        print(p_X_k[z])
        p_X_k[z][1]=k+z
        z=z+1
    return p_X_k;
```

The *ln factorial* function was used to simulate eqn. 5, while the *binpmf(n,p,k)* function were used for calculating the pmf of the problem. This function works by calculating the $\ln[P(k)$ until $k = n$, which is then transformed to the probability notation using the *math.exp()* function as presented in eqn. 7.

The graph as shown in Fig. 1 represents the probability curve of the Pratt-woodruff experiment. k for $1 \leq n \leq 1000$.

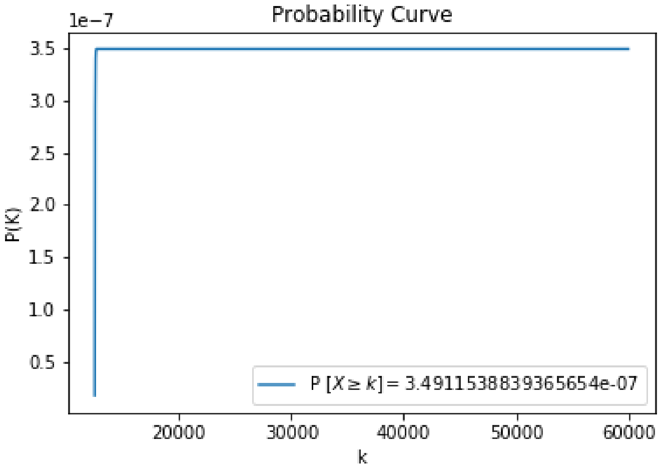

Fig. 1. Pratt-woodruff Experiment

To Validate the result, I used the *DeMoivre-Laplace Limit theorem*

$$\lim_{n\to} P\left(a \leq \frac{X - np}{\sqrt{np(1-p)}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{\frac{-z^2}{2}} dz \tag{8}$$

This equation returns the same output for all n, k, and p as the natural logarithm based solution used in this problem.

4. Conclusion

A solution to model the probability distribution of the Pratt-Woodruff experiment with $p = \frac{1}{5}$, $n = 60,000$, and $k = 12,489$ was developed using the natural logarithm $\ln[P(X = k]$ of the binomial probability distribution. Also, the relationship for $n!$ to the $\Gamma()$ function was used to eradicate the occurrence of computation overflow through the simulation.

The result was validated by comparing the output of this method to the one using

DeMoivre-Laplace Limit theorem, which produces the same output. Hence, the binomial probability distribution for problems with very large n can also be solved using $\ln[P(X = k]$ and the relationship between $n!$ and $\Gamma()$.