

Statistics 406 Lab 7
 GSI: Adam Rothman
 Tuesday, October 31

1. If $T \sim \text{Binomial}(n, p)$, then T has probability mass function:

$$P(T = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \text{ for } k = 0 \dots n$$

The binomial coefficient may create an overflow problem, thus we can take advantage of the log function:

$$\begin{aligned} P(T = k) &= \exp[\log P(T = k)] \\ \log P(T = k) &= \log \left(\binom{n}{k} p^k (1 - p)^{n-k} \right) \\ &= \log \binom{n}{k} + \log p^k + \log(1 - p)^{n-k} \\ &= \log \frac{n!}{k!(n-k)!} + k \log(p) + (n - k) \log(1 - p) \\ &= \log \frac{n(n-1) \cdots (n-k+1)}{k!} + k \log(p) + (n - k) \log(1 - p) \\ &= \sum_{j=n-k+1}^n \log j - \sum_{j=1}^k \log j + k \log(p) + (n - k) \log(1 - p) \\ P(T = k) &= \exp \left[\sum_{j=n-k+1}^n \log j - \sum_{j=1}^k \log j + k \log(p) + (n - k) \log(1 - p) \right] \end{aligned}$$

By using the above expression for calculating $P(T = k)$ we can overcome the numerical instability that the binomial coefficient may cause; however, keep in mind that p cannot be too small otherwise the $\log(1 - p)$ term will have very large magnitude.

Here is a numerically stable R implementation for calculating $P(T = k)$:

```
P <- 0
if ((k>0) & (k<n))
{
  logBinomCoeff <- sum(log(seq(n-k+1, n))) - sum(log(seq(1, k)))
  logP <- logBinomCoeff + k*log(p) + (n-k)*log(1-p)
  P <- exp(logP)
}else if(k == 0)
{
```

```

P <- (1-p)^n
}else if(k == n)
{
P <- p^n
}

```

This can also be done using the built in R binomial cdf:

```
P <- pbinom(k,n,p)-pbinom(k-1,n,p)
```

2. Consider a random variable $X \sim \text{Poisson}(\lambda)$. The probability mass function is:

$$P(X = k) = \exp(-\lambda)\lambda^k/k!$$

We can use the same procedure as we used in problem 1 in order to devise a numerically stable calculation of $P(X = k)$.

$$\begin{aligned}
P(X = k) &= \exp[\log P(X = k)] \\
\log P(X = k) &= \log(\exp(-\lambda)\lambda^k/k!) \\
&= -\lambda + k \log(\lambda) + \log(1/k!) \\
&= -\lambda + k \log(\lambda) - \sum_{j=1}^k \log(j) \\
P(X = k) &= \exp \left[-\lambda + k \log(\lambda) - \sum_{j=1}^k \log(j) \right]
\end{aligned}$$

You will use the above expression for writing your R code for question 1a. of the homework.

3. In the second question of this week's homework, you're asked to consider using the normal approximation to the binomial distribution in order to conduct a hypothesis test: $H_0 : p = 0.5$ and $H_1 : p > 0.5$. You're asked to calculate the exact level of the test. This means you must first find the cut point T_0 such that assuming H_0 is true, $P(T > T_0) \approx 0.05$. This value can be found by looking at the 0.95 quantile of the standard normal distribution. Then use R to find the exact value of $P(T > T_0)$ using the binomial distribution.

Keep in mind that the expression for T involves \bar{X} , so if you can isolate \bar{X} on the left side of the inequality inside the probability, then $n\bar{X}$ has the binomial distribution.