Statistics 406 Problem Set 7

Due in lab, Tuesday November 7th

1. Let X be a random variable following a Poisson distribution with parameter λ. The probability that X has value k is

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

where the sample space is $k = 0, 1, 2, \ldots$.

(a) Devise a numerically stable R implementation of the Poisson probability function. To be "numerically stable," you should not directly calculate the factorial. Then use the ppois function in R to get the exact probability and compare results for a few points in the sample space.

Since ppois gives the left tail probability, to get the exact value $P(X = k)$, you will need to use

$$P(X = k) = P(X \le k) - P(X \le k - 1) = \texttt{ppois(k,}\lambda) - \texttt{ppois(k} - 1\texttt{,}\lambda)$$

for $k > 0$, or just ppois(0,λ) for $k = 0$.

Solution:

```
## The Poisson parameter.
for (L in c(1,3))
{
  ## Check the first few points in the sample space.
  for (k in 0:5)
  {
    ## A numerically stable calculation of the Poisson probability.
    p <- -L + k*log(L)
    if (k>0) { p <- p - sum(log(1:k)) }
    p <- exp(p)

    ## Get the exact probability.
    p1 <- ppois(k, L)
    if (k>0) { p1 <- p1 - ppois(k-1, L) }

    print(c(p, p1))
  }
}
```

(b) The expected value and the variance of the Poisson distribution are both equal to λ. Write an R program that directly approximates the expected value and variance, by truncating as follows

$$EX = \sum_{k=0}^{\infty} kP(X = k) \approx \sum_{k=0}^{N} kP(X = k)$$

$$\text{var}(X) = \sum_{k=0}^{\infty} (k - EX)^2 P(X = k) \approx \sum_{k=0}^{N} (k - EX)^2 P(X = k),$$

where N is a large positive constant (say $N = 200$).

Solution:

```
## The expected value.
E <- 0

## The variance.
V <- 0

## The Poisson distribution parameter.
L <- 1

## Approximate the mean and variance by summing over the first 200
## points in the sample space.
for (k in (0:200))
{
    ## A numerically stable calculation of the Poisson probability.
    p <- -L + k*log(L)
    if (k>0) { p <- p - sum(log(1:k)) }
    p <- exp(p)

    ## Update the expected value.
    E <- E + k*p

    ## Update the variance.
    V <- V + (k-L)^2*p
}
```

(c) Write an R program to calculate the 90^{th}, 95^{th} and 99^{th} percentiles of the Poisson distribution with $\lambda = 10$. Since the sample space is discrete, we will define the Q^{th} percentile to be the greatest point k in the sample space such that $P(X \leq k) < Q$.

Solution:

```
## The Poisson parameter.
L <- 10
K <- NULL

## Calculate three quantiles.
for (Q in c(0.9, 0.95, 0.99))
{
  ## The cumulative probability.
  p <- 0

  ## The current point in the sample space.
  k <- 0

  ## Keep adding on to the cumulative probability until the quantile
  ## is reached.
  while (1)
  {
    ## Calculate the Poisson probability and add it on to the cumulative
    ## probability.
    p1 <- -L + k*log(L)
    if (k>0) { p1 <- p1 - sum(log(1:k)) }
    p1 <- exp(p1)
    p <- p + p1

    ## Stop if we've passed the quantile.
    if (p > Q) { break }

    ## Otherwise move to the next point in the sample space.
    k <- k + 1
  }

  ## The quantile, as we have defined it, is the point before the
  ## cumulative probability passes Q.
  K <- c(K, k-1)
}
```

(d) Use a normal approximation to approximate the 90^{th}, 95^{th} and 99^{th} of the Poisson distribution with $\lambda = 10$.

Hint: to do this, set up the equation

$$P(X \leq Q) = 0.95$$

where X is the Poisson random variable, and Q is the quantile we are looking for. Then standardize both sides of the inequality so that the standard normal probability can be used. Finally, solve for Q.

Solution: Let θ be either 0.9, 0.95, or 0.99. Then

$$\begin{aligned}\theta &= P(X \le Q) \\ &= P(\frac{X-\lambda}{\sqrt{\lambda}} \le \frac{Q-\lambda}{\sqrt{\lambda}}) \\ &\approx P(Z \le \frac{Q-\lambda}{\sqrt{\lambda}})\end{aligned}$$

Therefore $(Q-\lambda)/\sqrt{\lambda}$ is the θ quantile of the standard normal distribution. Let F_θ be this quantile. Then solving for Q yields

$$Q = \lambda + \sqrt{\lambda}F_\theta.$$

In R:

```
Q <- NULL
L <- 10
for (T in c(0.9, 0.95, 0.99))
{
  Q <- c(Q, L + sqrt(L)*qnorm(T))
}
```

2. Suppose we collect the results from n iid Bernoulli trials, and are interested in testing the null hypothesis that the success probability is $1/2$ against the alternative hypothesis that the success probability is greater than $1/2$. We plan to use the test statistic

$$T = 2\sqrt{n}(\bar{X} - 1/2)$$

to do this.

(a) Using a normal approximation for the null sampling distribution of T, what is the critical value for a nominal 0.05 level, one-sided test?

Solution: Since T is standardized, the critical value for a one-sided test is 1.64.

(b) Using the pbinom function in R, calculate the exact level of the test for sample sizes $n = 10, 20$, and 30.

Solution: We reject the null hypothesis if $T > 1.64$. This is equivalent to rejecting if

$$n\bar{X} > n/2 + 1.64\sqrt{n}/2,$$

where $n\bar{X}$ is the number of heads.

```
## Storage for the level results.
P <- NULL

## Loop over the sample sizes.
for (n in c(10,20,30))
{
  ## The number of heads needed to reject.
  T <- n/2 + 1.64*sqrt(n)/2

  ## The exact level.
  level <- 1-pbinom(T, n, 1/2)

  ## Save the level.
  P <- c(P, level)
}
```

(c) Use pbinom to calculate the exact power of the test when the alternative is that the success probability is 0.6, for sample sizes 10, 20, and 30.

Solution:

```
## Storage for the power results.
P <- NULL

## Loop over the sample sizes.
for (n in c(10,20,30))
{
  ## The number of heads needed to reject.
  T <- n/2 + 1.64*sqrt(n)/2

  ## The exact power.
  power <- 1-pbinom(T, n, 0.6)
```

```
  ## Save the power.
  P <- c(P, power)
}
```

(d) Using exact calculation with pbinom, determine the sample size that would give exact power 0.8 for the alternative success probability 0.6.

Solution:

```
## The initial sample size.
n <- 10

## Increment the sample size until the power becomes 0.8.
while (1)
{
  ## The number of heads needed to reject.
  T <- n/2 + 1.64*sqrt(n)/2

  ## The exact power.
  power <- 1-pbinom(T, n, 0.6)

  ## Check if the power is big enough, if so, we're done.
  if (power > 0.8) { break }

  ## Otherwise increase the sample size by one and try again.
  n <- n+1
}
```

(e) Using exact calculation with pbinom, determine the alternative success probability (greater than 1/2) that would give exact power at least 0.8 for sample size 30. It is sufficient to consider alternative success probabilities in the sequence $0.51, 0.52, \ldots$.

Solution:

```
## The sample size.
n <- 30

## The starting alternative success probability.
p_alt <- 0.51

while (1)
{
  ## The number of heads needed to reject.
  T <- n/2 + 1.64*sqrt(n)/2
```

```
  ## The power for the current p_alt value.
  power <- 1-pbinom(T, n, p_alt)

  ## If the power is at least 0.8, we're done.
  if (power > 0.8) { break }

  ## Otherwise, try an easier alternative.
  p_alt <- p_alt + 0.01
}
```