Homework 1

Gabriel Lahman

1/18/2019

Question 1

```
poisson_pop = function(N, theta) {
  return (rpois(N, theta))
}



N = 1000
theta = 2.718
num_pops = 100

# Each row is one population
populations = matrix(nrow=num_pops,ncol=N)

# Generate 100 random samples
for (pop in 1:100){
  populations[pop,] = poisson_pop(N, theta)
}

pop_means = rowMeans(populations)

num_samples = 250
sample_size = 49

sample_avgs = matrix(nrow=num_pops, ncol=num_samples)

# For each population, take a 250 samples of 49 values and calculate their means
for (pop in 1:100){
  for (sample in 1:num_samples){
    s = sample.int(N, sample_size)
    sample_avgs[pop,sample] = mean(populations[pop,s])
  }
}
```

```
dif_sq = c(rep(0,num_pops))
for (pop in 1:num_pops) {
  dif_sq[pop] = mean((populations[pop] - sample_avgs[pop,])^2)
}
```

A

```
dif = mean((theta - pop_means)^2)
```

```
## [1] "The average difference between the population means and theta is 0.00257011"
```

B

```
popSampleDif = c(rep(0,num_pops))
for (i in 1:100){
  popSampleDif[i] = mean((pop_means[i] - sample_avgs[i,])^2)
}
avgPopSampleDif = mean(popSampleDif)
```

```
## [1] "The avg. of the squared difference between a population's mean and its sample
means is 0.0527159278259059"
```

C

I would say that the variation between the population means and theta is more meaningful. Since the populations are of size 1000, while the samples are only of size 49, the population mean will have a smaller variation and its small value shows the the true value of the population means is tending towards θ.

D

```
difSampleTheta = mean((theta - sample_avgs)^2)
```

```
## [1] "The overal squared difference between the sample means and theta is 0.0552970
937442732"
```

E

Yes, I would be comfortable using the sample average to estimate θ. Since the sample average is unbiased, that means $E[sampleaverage] = \theta$, so as n grows to infinity, the sample average tends to θ.

Question 2

A

Expected Value is defined as $\sum_{i=1}^{k} x_i p_i$ where x are the events and p is the probability of any given event. For a Bernoulli distribution, there are two possible outcomes, 1 or 0, and their respective probabilities are θ and $1-\theta$. So, $E[X] = P(1) * (1) + P(0) * (0) = \theta * (1) + (1-\theta) * (0) = \theta$

Variance is defined as $E[X^2] - E[X]^2$.

$E[X^2] = P(1) * (1^2) + P(0) * (0^2) = \theta * (1) + (1-\theta) * (0) = \theta$

So, $E[X^2] - E[X]^2 = \theta - \theta^2 = \theta(1-\theta)$

B

$E[\hat{\theta}] = E[\bar{y}] = E[(\sum_{i=1}^{n} y_i)/n] = (\sum_{i=1}^{n} E(y_i))/n = (\sum_{i=1}^{n} \theta)/n = (n * \theta)/n = \theta$

Since $E[\hat{\theta}] = \theta$, $\hat{\theta}$ is unbiased.

$Var(\hat{\theta}) = Var((\sum_{i=1}^{n} y_i)/n) = Var(\sum_{i=1}^{n} \frac{y_i}{n}) = (\sum_{i=1}^{n} Var(y_i))/n^2 = (\sum_{i=1}^{n} \sigma^2)/n^2 = (n\sigma^2)/n^2 = \sigma^2/n$

C

$E[\tilde{\theta}] = E[\frac{\sum_{i=1}^{n} y_i + 1}{n+2}] = \frac{\sum_{i=1}^{n} E[y_i] + 1}{n+2} = \frac{n\theta+1}{n+2}$

Since $\frac{n\theta+1}{n+2} \neq \theta$, $\tilde{\theta}$ is biased

$Var(\tilde{\theta}) = Var(\frac{(\sum_{i=1}^{n} y_i)+1}{n+2}) = \frac{Var((\sum_{i=1}^{n} y_i)+1)}{(n+2)^2} = \frac{Var(\sum_{i=1}^{n} y_i)}{(n+2)^2} = \frac{\sum_{i=1}^{n} Var(y_i)}{(n+2)^2} = \frac{\sum_{i=1}^{n} \sigma^2}{(n+2)^2} = \frac{n\sigma^2}{(n+2)^2}$

The variance of $\hat{\theta} = \frac{\sigma^2}{n} = \frac{n\sigma^2}{n^2}$. Since $(n+2)^2 > n^2$, $\frac{n\sigma^2}{(n+2)^2} < \frac{n\sigma^2}{n^2}$ and therefore $\tilde{\theta} < \hat{\theta}$.

D

$Bias(\hat{\theta}) = \theta - \theta = 0$

$Risk(\hat{\theta}) = 0^2 + \frac{\sigma^2}{n} = \frac{\sigma^2}{n}$

$Bias(\tilde{\theta}) = \frac{n\theta+1}{n+2} - \theta = \frac{1-2\theta}{n+2}$

$Risk(\tilde{\theta}) = (\frac{1-2\theta}{n+2})^2 + \frac{n\sigma^2}{(n+2)^2} = \frac{(1-2\theta)^2}{(n+2)^2} + \frac{n\sigma^2}{(n+2)^2} = \frac{n\sigma^2+(1-2\theta)^2}{(n+2)^2}$

The risk for $\hat{\theta}$ is only equal to its variance, which seems to say there is not much risk associated with that estimator. However, since $\tilde{\theta}$ is biased, its risk appears to be much more, as the bias outweighs the smaller variance.