Lecture 3 Solutions

Isabel Fulcher

8/7/2018

Normal distribution exercises

1. Generate 10 draws from a normal random variable with mean 5 and variance 4. What is the sample mean of these variables? What is the true mean?

```
set.seed(12)
x <- rnorm(10,5,2)
mean(x)
```

```
## [1] 4.065572
```

2. How would you expect the previous answer to change if we increased the number of draws to 10,000? Check your intuition.

```
set.seed(12)
x <- rnorm(10000,5,2)
mean(x)
```

```
## [1] 4.997451
```

3. The probability density function (pdf) for a normal random variable X with mean μ and variance σ^2 is as follows,

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} exp(\frac{1}{2\sigma^2}(x-\mu)^2)$$

What is the value of $f_X(0)$ when $\mu = 0$ and $\sigma^2 = 2$? Calculate using the above expression and built-in R function.

```
#evaluate expression
(1/sqrt(2*pi*2))
```

```
## [1] 0.2820948
```

```
#built-in R function
dnorm(0,0,sqrt(2))
```

```
## [1] 0.2820948
```

4. If $X \sim N(5, 4)$, what is $P(X < 2)$? What is $P(X \leq 2)$? What is $P(X > 2)$?

```
# P(X < 2) = P(X <= 2)
pnorm(2,5,4)
```

```
## [1] 0.2266274
```

```
# P(X > 2)
1-pnorm(2,5,4)
```

```
## [1] 0.7733726
```

```
pnorm(2,5,4,lower.tail=FALSE)
```

```
## [1] 0.7733726
```

Poisson distribution exercises

1. Generate a random sample of size 100,000 from a Poisson distribution with rate parameter 5

```r
set.seed(80417)
xsamp <- rpois(100000,5)
```

2. Compute the sample mean and variance. What do you notice?

```r
mean(xsamp)
```

```
## [1] 4.99062
```

```r
var(xsamp)
```

```
## [1] 5.014122
```

3. Compute the empirical estimates of $P(X = 0)$, $P(X = 1)$, and $P(X = 2)$ for $X \sim Pois(5)$ from your sample.
4. Compare the estimates to the true value of the probability mass function (pmf) at $x = 0, 1, 2$. Use the pmf and the built-in R function. Note the pmf is given by:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \text{ where } x \in 0, 1, 2, ... \text{ and } \lambda > 0$$

5. Compute an estimate of $P(0 < X < 3)$ from the sample and compare to true value.

```r
#estimate
mean(xsamp > 0 & xsamp < 3)
```

```
## [1] 0.11977
```

```r
#true value
dpois(1,5)+dpois(2,5) #P(X=1) + P(X=2)
```

```
## [1] 0.1179141
```

```r
ppois(2,5)-ppois(0,5) #P(X <= 2) - P(X = 0)
```

```
## [1] 0.1179141
```

6. Find the smallest k such that $0.30 < P(X < k)$ for $X \sim Pois(5)$ using a while loop. Then, do this using a built-in R function.

```r
# Thank you, Beau!
k = 0
pr = 0
while (pr < .3){
  pr = pr + ppois(k,5)
  k = k + 1
}
k
```

```
## [1] 4
```

Group exercises

To display your results, create a table in Rmarkdown using the kable() function. Try to make it as clean as possible (i.e. column headers, title, digits, etc.).

Fun note: you will prove these results formally in the probability course!

Group 1

Generate 100,000 samples from a geometric distribution with $p = 0.3$. Estimate $P(X \geq s + t|X \geq t)$ and $P(X \geq s)$ for $s = 4$ and $t = 1, 2, 3, 4, 5, 6$. Compare to the true values. What do you notice? Google 'memoryless property distribution' and take a look at the wiki page on memorylessness. What does this suggest about the geometric distribution?

```
xsamp <- rgeom(100000,.3)
T <- 1:6

pr_func <- function(x,s,t) {
  pr1 <- round(mean(x[x >= t] >= s + t),3)
  pr2 <- round(mean(x >= s),3)
  return(data.frame("t" = t, "cond" = pr1, "marg" = pr2))
}

results <- sapply(T,pr_func,x=xsamp,s=4)

# Create a table
kable(t(results), align=rep('c',times=3),
      col.names = c("t","P(X >= 4+t | X >= t)","P(X >= 4)"),
      caption="Exercise 1 results")
```

Table 1: Exercise 1 results

t	P(X >= 4+t | X >= t)	P(X >= 4)
1	0.24	0.24
2	0.242	0.24
3	0.241	0.24
4	0.242	0.24
5	0.243	0.24
6	0.24	0.24

Group 2

Generate 10,000 samples from a $Bin(3, 0.5)$ and another 10,000 samples from $Bin(5, 0.5)$. Compute the empirical cdf of the sum of the two samples and compare to the distribution function of a $Z \sim Bin(8, 0.5)$ random variable. What does this suggest about the distribution of $X + Y$ where $X \sim Bin(n_1, p)$ and $Y \sim Bin(n_2, p)$?

```
# Two sample distribution
x <- rbinom(10000,3,.5)
y <- rbinom(10000,5,.5)
xysum <- x+y
```

```r
emp.cdf <- function(samp,val) {
  p <- round(mean(samp <= val),3)
  return(data.frame("k" = val, "prob" = p))
}

values <- 0:8
results.sum <- sapply(values,emp.cdf,samp=xysum)

# Empirical CDF Bin(8,0.5)
results.z <- round(sapply(values,pbinom,size=8,prob=0.5),3)

# Create table
results <- cbind(t(results.sum),results.z)
kable(results,align=rep('c',times=3),
      col.names = c("k","Empirical Pr(X+Y <= k)","Pr(Z <= k)"),
      caption="Exercise 2 results")
```

Table 2: Exercise 2 results

k	Empirical Pr(X+Y <= k)	Pr(Z <= k)
0	0.003	0.004
1	0.034	0.035
2	0.141	0.145
3	0.356	0.363
4	0.634	0.637
5	0.861	0.855
6	0.967	0.965
7	0.996	0.996
8	1	1