Lecture 3 Exercises

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

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

Poisson distribution exercises

1. Generate a random sample of size 100,000 from a Poisson distribution with rate parameter 5
2. Compute the sample mean and variance. What do you notice?
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.
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.

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?

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 $Y \sim Bin(n_2, p)$?