Random Variables and Discrete Probability Functions

Blake Pappas

09/28/2021

Review of Discrete Families of Distributions

Binomial Probabilities in R

Suppose a random variable X has a binomial distribution with $n = 15$ and $\pi = 0.89$; that is, X is the number of successes in a binomial experiment with 15 trials and probability of success equal to 0.89.

To find the probability that X = x for some value of x, use the `dbinom()` function. Set the "size" option equal to n and the "prob" option equal to π.

```
# P(X = 11)
dbinom(11, size = 15, prob = 0.89)

## [1] 0.05546174

# P(X = 14)
dbinom(14, size = 15, prob = 0.89)

## [1] 0.3228078
```

To find the probability that $X \leq x$, use the `pbinom()` function.

```
# P(X <= 11)
pbinom(11, size = 15, prob = 0.89)

## [1] 0.07420982
```

To find the probability that $X > x$, there are two options. First, set the lower.tail equal to FALSE.

```
# P(X > 11)
pbinom(11, size = 15, prob = 0.89, lower.tail = FALSE)

## [1] 0.9257902
```

The second option is to use the complement rule:

```
# P(X > 11) = 1 - P(X <= 10)
1 - pbinom(11, 15, 0.89)

## [1] 0.9257902
```

Poisson Probabilities in R

Assume that "Y" is a random variable with a Poisson distribution with the rate parameter, λ, equal to 3.5. To find $P(Y = y)$, plug y and λ into the `dpois` function.

```
# Set the Poisson Rate Parameter
lambda.value <- 3.5

# Find P(Y = 2)
dpois(2, lambda = lambda.value)

## [1] 0.184959

# Find P(Y=0)
dpois(0, lambda = lambda.value)

## [1] 0.03019738
```

Use the `ppois()` function to find the probability that Y is less than or equal to some value.

```
# Find P(Y <= 4)
ppois(4, lambda = lambda.value)

## [1] 0.725445
```

Remember that for a discrete random variable, $P(Y \leq 4)$ is not equal to $P(Y < 4)$. To find $P(Y < 4)$, make use of the fact that $P(Y < 4) = P(Y \leq 3)$.

```
ppois(3, lambda = lambda.value)

## [1] 0.5366327
```

There is also a `lower.tail` option to obtain $P(Y > y)$ for some value y. Again, keep track of the strict and non-strict inequalities.

```
# P(Y > 4)
ppois(4, lambda = lambda.value, lower.tail = FALSE)

## [1] 0.274555

# P(Y >= 4)
ppois(3, lambda = lambda.value, lower.tail = FALSE)

## [1] 0.4633673
```

Generating a Random Variable from a Distribution

Sometimes it is helpful to generate random variables from a certain distribution in order to approximate probabilities.

To generate random variables from the Bernoulli distribution, use the `dbinom()` function with the "size" option set to 1.

```r
# Generate 10 Random Bernoulli Random Variables with pi = 0.65
rbinom(10, size = 1, prob = 0.65)

##  [1] 1 0 1 1 1 1 0 0 0 1
```

Binomial random variables are generated in a similar way, but set the “size” option equal to n.

```r
# Generate 10 Random Binomial Random Variables with pi = 0.65 and n = 25
rbinom(10, size = 25, prob = 0.65)

##  [1] 19 14 13 18 19 10 19 15 12 16
```

Simulation to Approximate Probabilities

Large samples from probability distributions can be used to approximate probabilities. For example, let’s consider a random variable X whose distribution is the binomial distribution with n = 12 and π = 0.10. To approximate the probability that X < 2, we first generate a large number (10000) of values from the distribution.

```r
no.samples <- 10000
binomial.samples <- rbinom(no.samples, size = 12, prob = 0.10)
```

The proportion of samples with values less than 2 is an approximation of $P(X < 2)$. See the code below for the approximation.

```r
sum(binomial.samples <= 2) / no.samples

## [1] 0.893
```

We can compare this to the true value given by the `pbinom()` function. The two numbers should be close. See the code below for the exact answer.

```r
pbinom(2, size = 12, prob = 0.1)

## [1] 0.88913
```

Try decreasing and increasing the `no.samples` above to see how the approximation improves with larger samples.

Simulation to Approximate Expected Values

We sometimes call $E(X)$ the “mean” of X. One reason that this makes sense is that $E(X)$ is the number that would be the sample average of an infinitely large sample from the probability distribution of X.

This means that $E(X)$ can be approximated using simulation. To do this, take a very large random sample from the probability distribution of X. Then, calculate the mean of these samples. When the number of samples is large enough, the mean of the samples will be very close to the true $E(X)$ value.

Example:

We know that $E(X)$ for the binomial distribution with n = 12 and π = 0.1 is equal to $n\pi =$ 1.2.

We can also approximate this with simulation by taking the sample average of the binomial samples from the previous example.

```
mean(binomial.samples)

## [1] 1.1827
```

Example: Poisson Calculations

Suppose the random variable Y has a Poisson distribution. Compute the following probabilities:

a. $P(Y = 4)$ given $\lambda = 2$.

```
dpois(4, lambda = 2)

## [1] 0.09022352
```

b. $P(Y = 4$ given $\lambda = 3.5$.

```
dpois(4, lambda = 3.5)

## [1] 0.1888123
```

c. $P(Y > 4)$ given $\lambda = 2$.

```
ppois(4, lambda = 2, lower.tail = FALSE)

## [1] 0.05265302
```

d. $P(1 \leq Y < 4)$ given $\lambda = 2$.

```
ppois(3, lambda = 2) - ppois(1, lambda = 2)

## [1] 0.4511176
```

Example: Quality Control

The quality control department examines all the products returned to a store by customers. An examination of the returned products yields the following assessment: 5% are defective and not repairable, 45% are defective but repairable, 35% have small surface scratches but are functioning properly, and 15% have no problems.

Compute the following probabilities for a random sample of 20 returned products.

a. All of the 20 returned products have some type of problem.

```r
no.samples <- 10000
binomial.samples <- rbinom(no.samples, size = 20, prob = 0.85)
sum(binomial.samples == 20) / no.samples

## [1] 0.0369
```

b. Exactly 6 of the 20 returned products are defective and not repairable.

```r
no.samples <- 10000
binomial.samples <- rbinom(no.samples, size = 20, prob = 0.05)
sum(binomial.samples == 6) / no.samples

## [1] 5e-04
```

c. Of the 20 returned products, 6 or more are defective and not functioning properly.

```r
no.samples <- 10000
binomial.samples <- rbinom(no.samples, size = 20, prob = 0.5)
sum(binomial.samples >= 6) / no.samples

## [1] 0.9789
```

d. None of the 20 returned products has any sort of defect.

```r
no.samples <- 10000
binomial.samples <- rbinom(no.samples, size = 20, prob = 0.50)
sum(binomial.samples == 0) / no.samples

## [1] 0
```

Example: Expected Value of a Binomial Random Variable

For parts a, b, and c, assume that the random variable X has a binomial distribution with n = 8 and π = 0.3.

a. What is the expected value (mean) of X?

Answer: The expected value (mean) of x is 2.4.

b. The answer will be the same if you use the general formula for expected value,

$$E(X) = \sum_{x \in \mathcal{X}} x \, P(X = x).$$

Write code below that will calculate E(X) according to this formula. Find the $P(X = x)$ values using either the formula for the binomial pmf or the `dbinom()` function. Your answer should match part a.

Answer: The expected value (mean) of x is 2.4. E(X) = nπ, so E(X) = 8 x 0.3.

c. Use the `rbinom()` function to approximate the E(X) through simulation. Try taking samples of size of 100, 1000, 5000, and 10,000. How many samples does it take to obtain what appears to be a reasonably close approximation?

```r
mean(rbinom(100, size = 8, prob = 0.3))
```

```r
## [1] 2.35
mean(rbinom(1000, size = 8, prob = 0.3))
## [1] 2.428
mean(rbinom(5000, size = 8, prob = 0.3))
## [1] 2.3714
mean(rbinom(10000, size = 8, prob = 0.3))
## [1] 2.4218
```

Answer: It takes about 10,000 samples to obtain a reasonably close approximation.

Example: Expected Value of a Poisson Random Variable

Use the `rpois()` function to approximate E(Y) if Y is a Poisson random variable with $\lambda = 8.2$.

```r
Y <- 10000

s <- sum(rpois(Y, lambda = 8.2))

s / Y
## [1] 8.2318
```

Answer: The expected value of a Poisson random variable with $\lambda = 8.2$ is approximately 8.2.

Example: Unknown pi

We generated three random variables from a binomial distribution with $n = 30$ and a fixed value of π. The values are 6, 8, and 5.

The value of π is unknown, but is one of the following values: 0.17, 0.24, 0.31, 0.39. Use the `dbinom()` function to find $P(X = 6)$, $P(X = 8)$, and $P(X = 5)$ for each of the possible values of π:

a. $= $ = 0.17

```r
dbinom(6, size = 30, prob = 0.17)
## [1] 0.1637531
dbinom(8, size = 30, prob = 0.17)
## [1] 0.06771461
dbinom(5, size = 30, prob = 0.17)
## [1] 0.1918802
```

b. $\pi = 0.24$

```r
dbinom(6, size = 30, prob = 0.24)

## [1] 0.1564644

dbinom(8, size = 30, prob = 0.24)

## [1] 0.153802

dbinom(5, size = 30, prob = 0.24)

## [1] 0.118913
```

c. $\pi = 0.31$

```r
dbinom(6, size = 30, prob = 0.31)

## [1] 0.07147765

dbinom(8, size = 30, prob = 0.31)

## [1] 0.1422154

dbinom(5, size = 30, prob = 0.31)

## [1] 0.0381829
```

d. $\pi = 0.39$

```r
dbinom(6, size = 30, prob = 0.39)

## [1] 0.0147206

dbinom(8, size = 30, prob = 0.39)

## [1] 0.05931249

dbinom(5, size = 30, prob = 0.39)

## [1] 0.005525888
```

Do the calculations provide any insight into which π values are most plausible? Give a guess as to which π value was the one used to generate the random sample.

Answer: The calculations do not give me any insight into which π values are most plausible. If I had to guess, I would say π has a value of 0.24.