Tutorial 2: Properties of Random Variables

Anh Le

August 25, 2015

Agenda:

1. Implement formulas for Expected Values, Variance, etc. in R (and learn vectorized operation along the way)
2. Download data automatically from the web (no more point and click!)
3. Draw the plots you saw from lectures in R

1. Implement expected value and variance formula

Use `sum()` (to get the sum) and `length()` (to get the number of elements in a vector). Calculate:

$$E(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Use vectorized operation!

$$Var(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - E(X))^2$$

Let's break down this formula. Mathematically, the formula mean that for each element `X_i` in the vector `X`: - subtract `E(X)` from `X_i`, square the result - then we add up all the results and divide by `n - 1`

So we can naively translate that into code as follows:

```
myVec <- rnorm(1000, mean = 2, sd = 5)

myVar1 <- function(X) {
  n <- length(X)

  sum = 0
  # For each element X_i
  for (i in 1:n) {
    # Subtract E(X), square the result, then add the results together
    sum = sum + (X[i] - mean(X)) ** 2
  }

  return(sum / (n - 1))
}

myVar1(myVec)
```

```
## [1] 22.63753
```

```
var(myVec)
```

```
## [1] 22.63753
```

But loops in R are notoriously slow! We should use vectorized operation instead. For example,

```
X <- 1:5

# To subtract E(X) from each element
X - mean(X)
```

```
## [1] -2 -1  0  1  2
```

```
# To square all elements
X ** 2
```

```
## [1]  1  4  9 16 25
```

```
# To calculate the sum of squares
sum(X ** 2)
```

```
## [1] 55
```

Let's use this to rewrite `myVar1` so that it's faster:

```
myVar2 <- function(X) {
  return(sum((X - mean(X)) ** 2) / (length(X) - 1))
}

myVar2(myVec)
```

```
## [1] 22.63753
```

```
myVar1(myVec)
```

```
## [1] 22.63753
```

```
var(myVec)
```

```
## [1] 22.63753
```

Let's compare the speed:

```
library(rbenchmark) # install.packages if you don't have the package
benchmark(myVar1(myVec), myVar2(myVec))
```

```
##            test replications elapsed relative user.self sys.self
## 1 myVar1(myVec)          100   0.587    293.5     0.588        0
## 2 myVar2(myVec)          100   0.002      1.0     0.002        0
##   user.child sys.child
## 1          0         0
## 2          0         0
```