Data Generation

Yuchen Li (li215), section 1UG

April 14, 2019

Contents

Team Powers

Standard Probability Distributions

Normal

```
genNormalPowers <- function(n, mean=0, sd=1) {
  # INPUT:
  # n is the number of teams

  # OUTPUT:
  # returns a vector of team powers, sorted in decreasing order
  powers <- rnorm(n, mean, sd)
  return(sort(abs(powers), decreasing=TRUE))
}
```

Accept-Reject

```
# Reference: adapted from Yuchen Li (li215), HW2, Exercise 4

acceptReject <- function(nsim, f, min, max, M) {
  # INPUT:
  # nsim is the number of simulations
  # f is the target distribution
  # min is the min value in the domain of f
```

```
  # max is the max value in the domain of f
  # max
  # M >= sup{f(x)}

  # OUTPUT:
  # returns a vector of random variates sampled from f, using the
  #   Accept-Reject method with Unif(min, max) as the reference distribution
  k1 = 0              # counter for accepted samples
  j1 = 0              # number of iterations required to get desired sample size
  y1 = numeric(nsim)    # storing the sample
  while(k1 < nsim){
      u = runif(1)
      x = runif(1, min, max)  # random variate from reference distribution
      g1 = 1
      if (u < f(x) / M / g1) {
        # condition of accepting x in our sample
        k1 = k1 + 1
        y1[k1] = x
      }
      j1 = j1 + 1
  }
  return(sort(y1, decreasing=TRUE))
}
```

Inverse CDF

```
inverseCDF <- function(n, inv_cdf) {
  # INPUT:
  # n is the number of simulations
  # inv_cdf is the inverse CDF function for f

  # OUTPUT:
  # returns a vector of random variates sampled from PDF f,
  #   using the Inverse CDF method
  u = runif(n)
  y = numeric(n)
  for (i in 1:n) {
    y[i] = inv_cdf(u[i])
  }
  return(sort(y, decreasing=TRUE))
}
```

Cross-Team Winning Probabilities

```
genCrossTeamWinningProbabilities <- function(powers) {
  # INPUT:
  # powers is the teams powers

  # OUTPUT:
  # returns an n x n matrix M where M_{ij} is the probability of team-i beating team-j
```

```
  n = length(powers)
  probs = matrix(nrow=n, ncol=n)
  for (i in 1:n) {
    for (j in 1:n) {
      probs[i,j] = powers[i] / (powers[i] + powers[j])
    }
  }
  return(probs)
}
```

Seeding

What are the other good methods than random selection? (In the data generation part, we do not have actual competition data yet.)

```
# Example
sample(1:8, size=2)
```

```
## [1] 4 2
```

Examples

Normal(0, 1) team powers

```
genCrossTeamWinningProbabilities(
 genNormalPowers(4)
)
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,] 0.50000000 0.9010777 0.9212617 0.9679326
## [2,] 0.09892230 0.5000000 0.5622644 0.7681805
## [3,] 0.07873834 0.4377356 0.5000000 0.7206542
## [4,] 0.03206739 0.2318195 0.2793458 0.5000000
```

Normal(10, 0.01) team powers

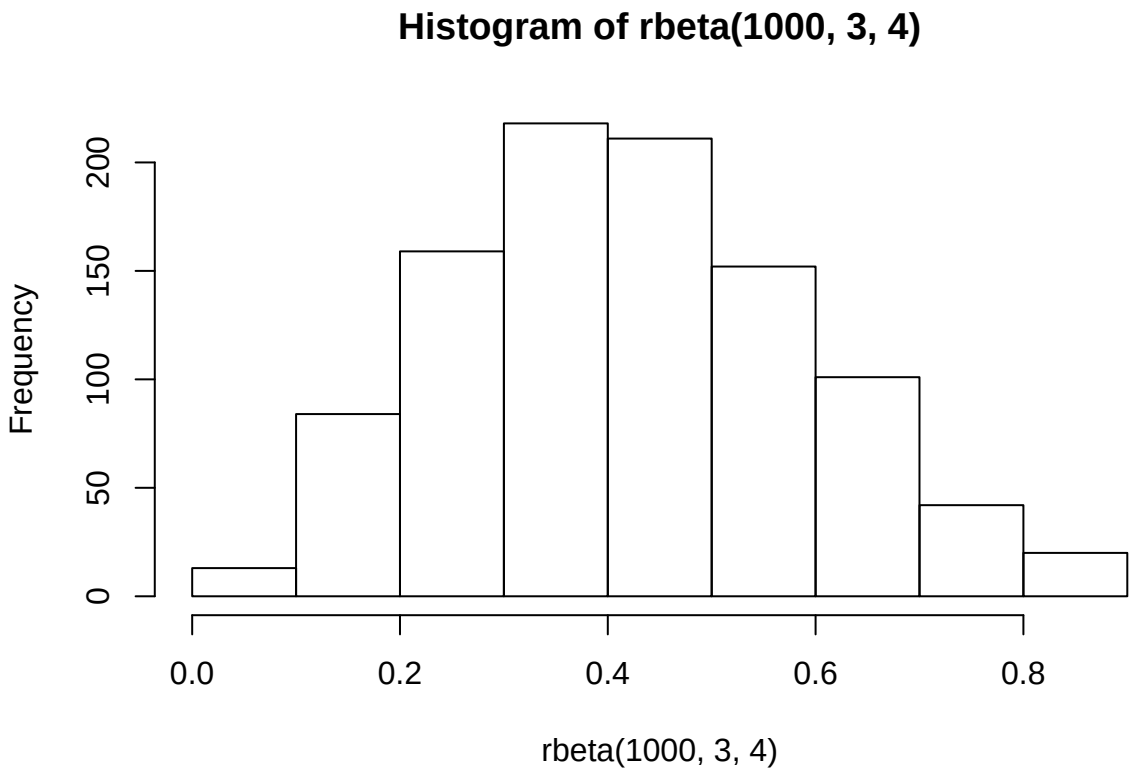
Note the probabilities are closer to 0.5

```
genCrossTeamWinningProbabilities(
 genNormalPowers(4, mean=10, sd=0.1)
)
```

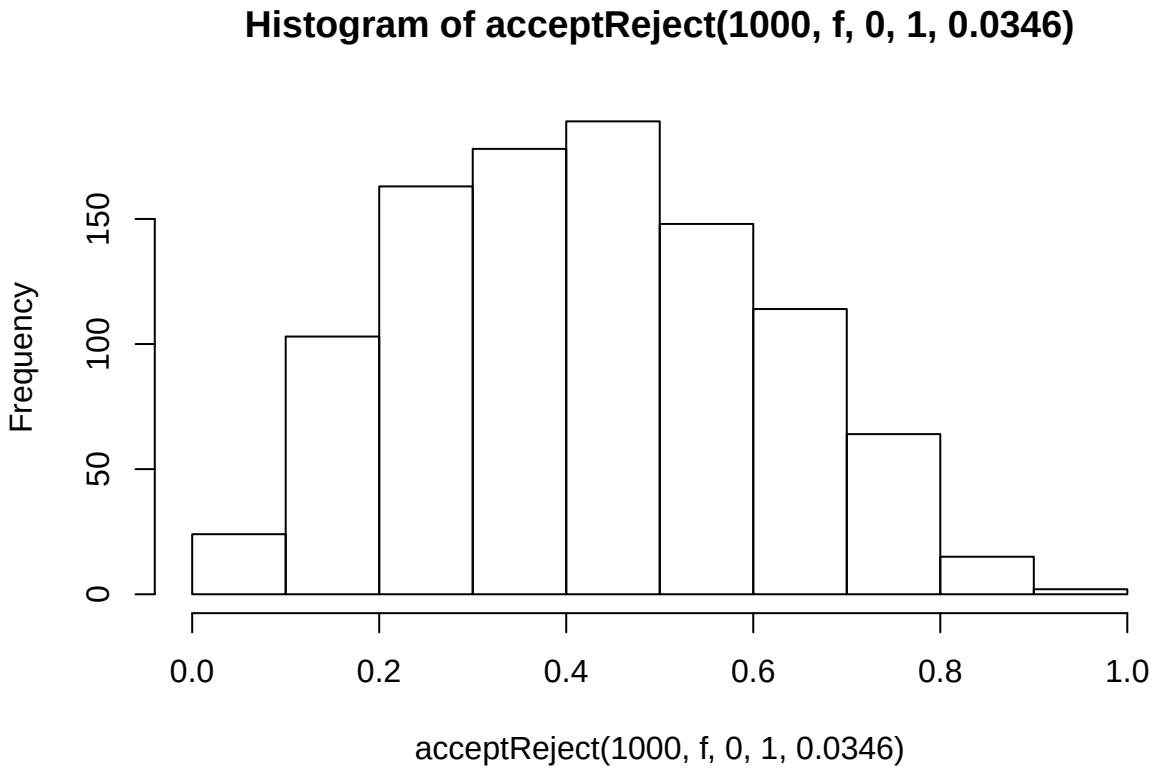
```
##           [,1]      [,2]      [,3]      [,4]
## [1,] 0.5000000 0.5002865 0.5029684 0.5048360
## [2,] 0.4997135 0.5000000 0.5026819 0.5045495
## [3,] 0.4970316 0.4973181 0.5000000 0.5018677
## [4,] 0.4951640 0.4954505 0.4981323 0.5000000
```

Beta(3, 4) team powers using Accept-Reject

```
# Test `acceptReject`
f <- function(x) {return(x^2 * (1-x)^3)}
hist(rbeta(1000, 3, 4))
```




```
hist(acceptReject(1000, f, 0, 1, 0.0346))
```

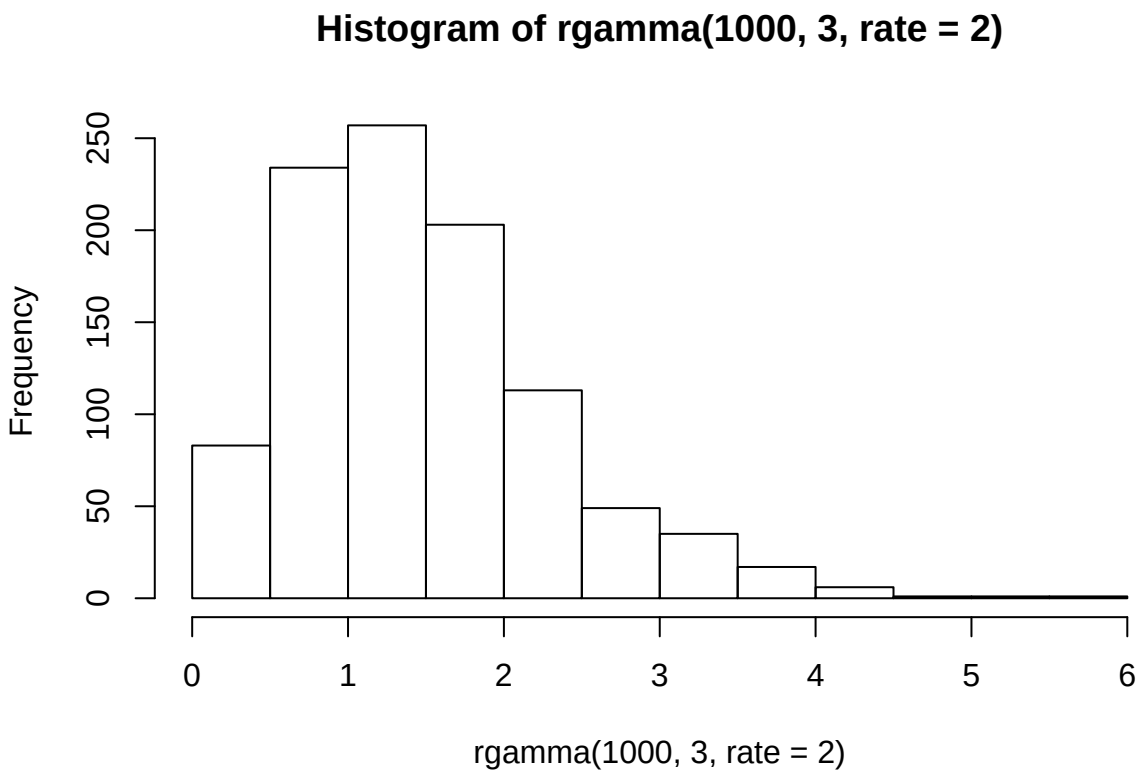



```
# Actual
genCrossTeamWinningProbabilities(
 acceptReject(4, f, 0, 1, 0.0346)
)
```

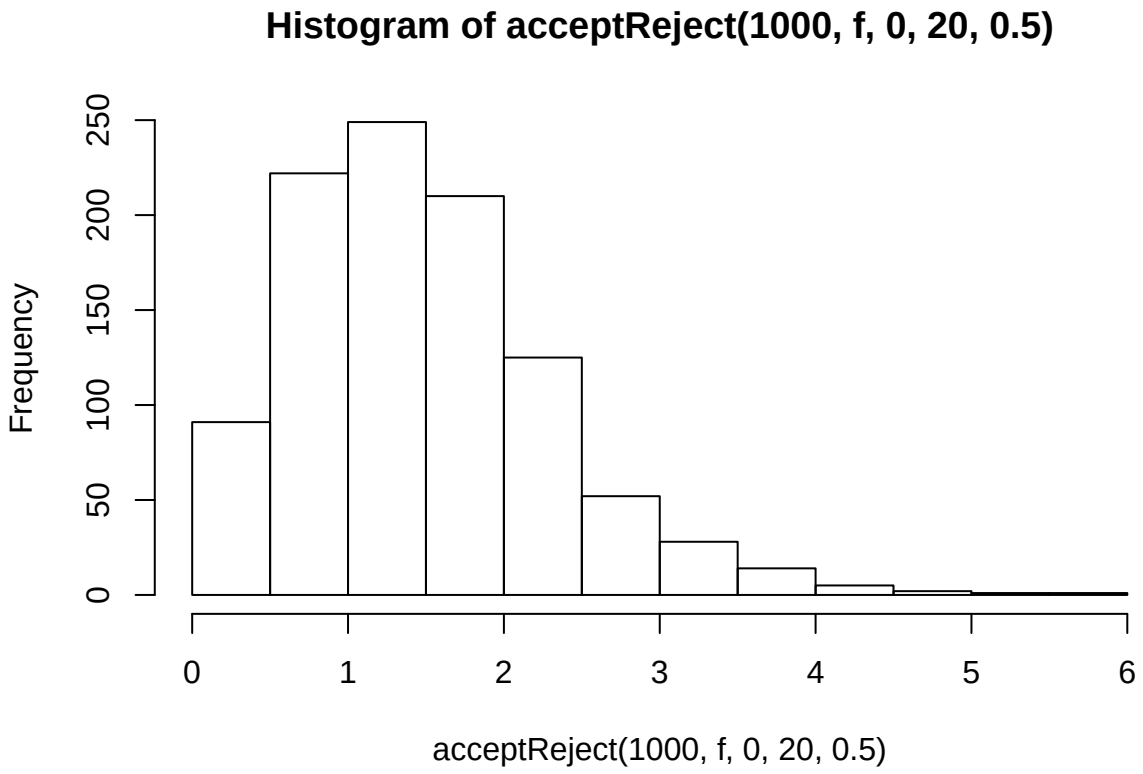
```
##           [,1]      [,2]      [,3]      [,4]
## [1,] 0.5000000 0.5429589 0.6381450 0.6911963
## [2,] 0.4570411 0.5000000 0.5975006 0.6532730
## [3,] 0.3618550 0.4024994 0.5000000 0.5593183
## [4,] 0.3088037 0.3467270 0.4406817 0.5000000
```

Gamma(3, 2) team powers using Accept-Reject

```
# Test `acceptReject`
f <- function(x) {
    return(2^3 / gamma(3) * x^2 * exp(-2*x))
}
hist(rgamma(1000, 3, rate=2))
```




```
hist(acceptReject(1000, f, 0, 20, 0.5))
```



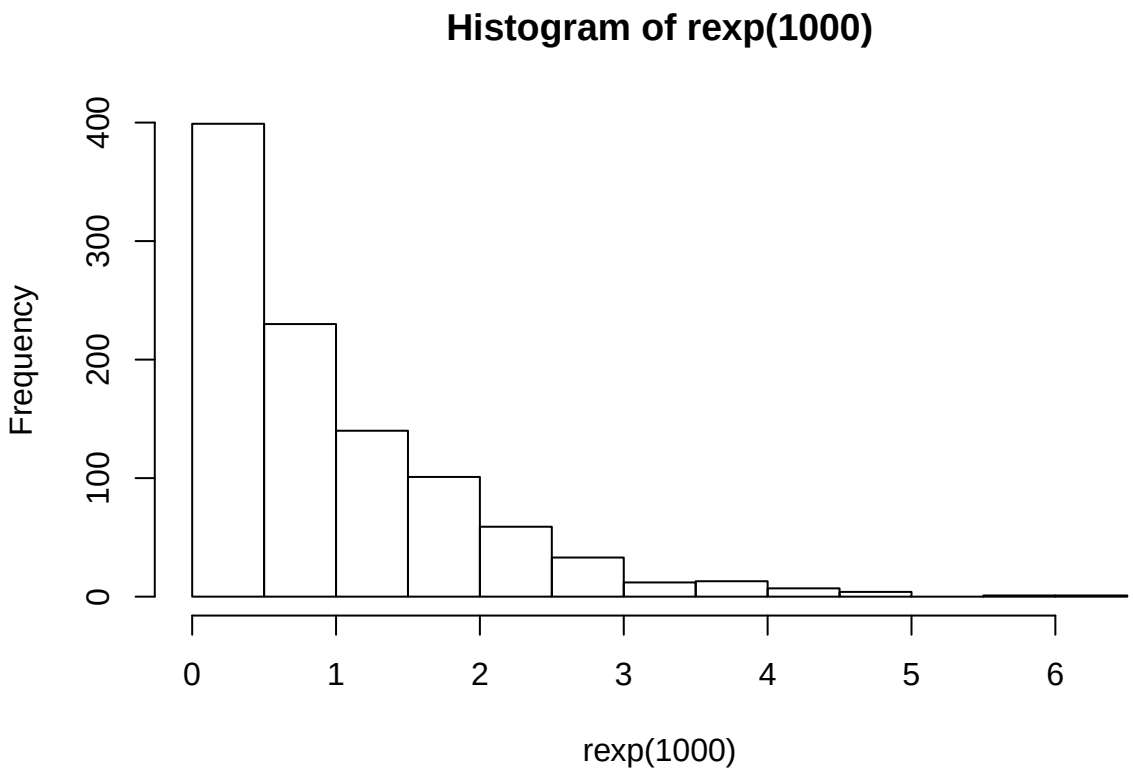

```
# Actual
genCrossTeamWinningProbabilities(
 acceptReject(4, f, 0, 20, 0.5)
```

```
)
```

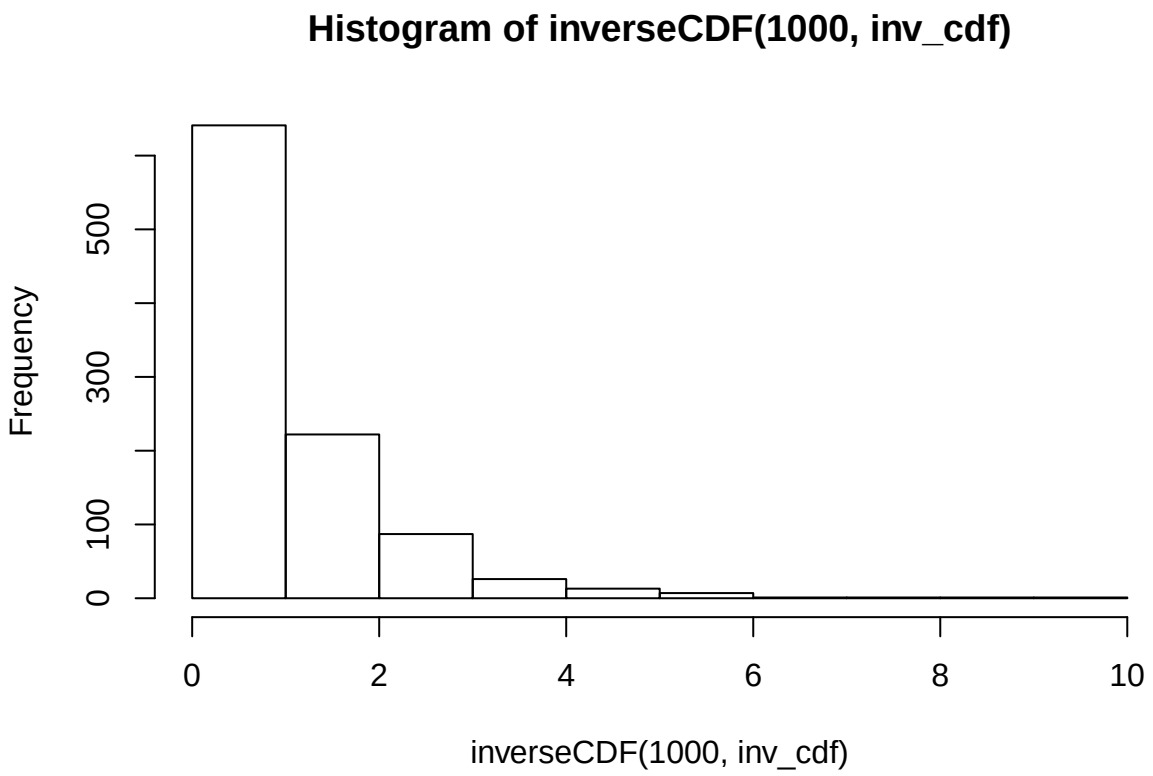
```
##            [,1]      [,2]      [,3]      [,4]
## [1,] 0.50000000 0.5633657 0.6796006 0.9157035
## [2,] 0.43663434 0.5000000 0.6217786 0.8938343
## [3,] 0.32039939 0.3782214 0.5000000 0.8366369
## [4,] 0.08429651 0.1061657 0.1633631 0.5000000
```

Exp(1) team powers using Inverse CDF

```
# Test `inverseCDF`
inv_cdf <- function(x) {return(-log(x))}
hist(rexp(1000))
```




```
hist(inverseCDF(
    1000,
    inv_cdf
  )
)
```




```
# Actual
genCrossTeamWinningProbabilities(
 inverseCDF(
    4,
    inv_cdf
  )
)
```

```
##            [,1]      [,2]      [,3]      [,4]
## [1,] 0.50000000 0.5639417 0.8295896 0.9077138
## [2,] 0.43605831 0.5000000 0.7901031 0.8837940
## [3,] 0.17041037 0.2098969 0.5000000 0.6689218
## [4,] 0.09228623 0.1162060 0.3310782 0.5000000
```