MAT370
Applied Probability
Sample Student Project

The Extreme Value of a Sample from an Exponentially Distributed Population

Overview

An extreme value from a sample (a maximum or a minimum) is often of interest. For example, when designing a sewer system or a dam, an engineer would like to make predictions about the magnitudes of future precipitation events based upon available data. A dam, for instance, should be able to withstand future water levels without being absurdly immense.

This project considers the relatively easy case of exponentially distributed random variables. A theoretical model for the distribution of the maximum value from a sample of size n is developed, and checked against data from samples of size 2.

Theory

Consider a random variable, X, which is exponentially distributed. That is, one with the following probability density function

$$f(x) = \begin{cases} \alpha e^{-\alpha x} & \text{if } 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

In this case X has the cumulative distribution function

$$F(x) = \begin{cases} 1 - e^{-\alpha x} & \text{if } 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

where $F(x) = Prob[X \leq x]$.

Suppose now that we have a family of *iid* random variables X_i. That is, the random variables are *independent, identically distributed* with common cumulative distribution function $F(x)$. We calculate the cumulative distribution for $max(X_1, X_2, \ldots, X_n)$.

Consider first $Z = max(X, Y)$ for independent random variables X and Y with common cumulative distribution function $F(x)$. The cumulative distribution function for Z is defined to be the function

$$F_Z(x) = Prob[Z \leq x].$$

But $Z \leq x$ if and only if $X \leq x$ and $Y \leq x$. That is,

$$Prob[Z \leq x] = Prob[X \ and \ Y \leq x] = Prob[X \leq x]Prob[Y \leq x]$$

since X and Y are independent. Thus, since X and Y have common cumulative distribution function $F(x)$

$$Prob[Z \leq x] = F_Z(x) = (F(x))^2.$$

And finally, since the probability density function

$$f_Z(x) = F_Z'(x),$$

we have

$$f_Z(x) = 2(F(x))^1 F'(x).$$

For the general case, where $Z = max(X_1, X_2, \ldots, X_n)$, we have

$$f_Z(x) = n(F(x))^{(n-1)} F'(x).$$

For example, suppose X and Y have exponential distribution with $\alpha = 1$. Then their common density is given by

$$f(x) = \begin{cases} e^{-x} & \text{if } 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

and their cumulative distribution function

$$F(x) = \begin{cases} 1 - e^{-x} & \text{if } 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

Then, according to the previous discussion, $Z = max(X, Y)$ will have cumulative distribution

$$Prob[Z \leq x] = F_Z(x) = (F(x))^2 = \begin{cases} (1 - e^{-x})^2 & \text{if } 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

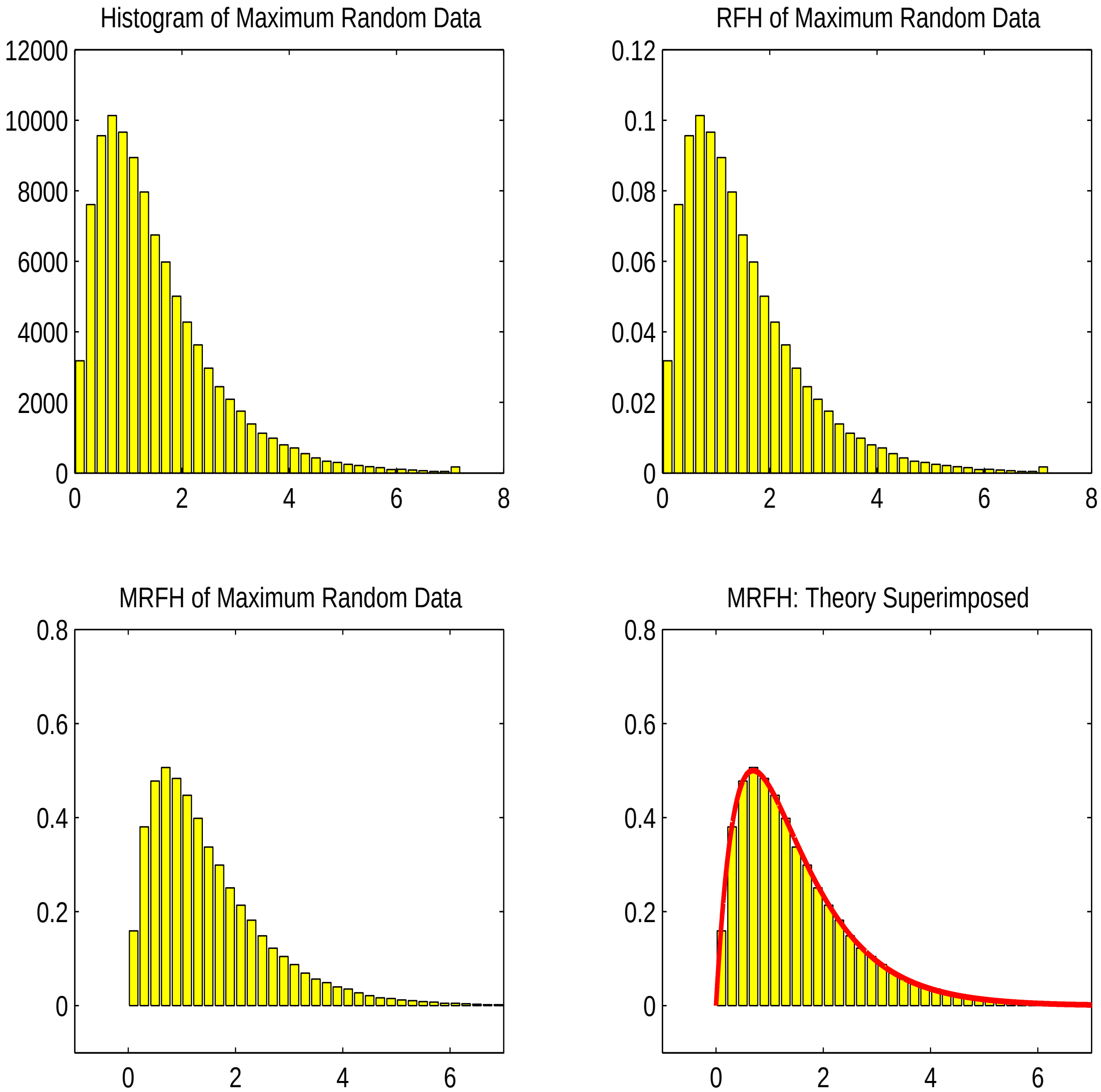
and probability density function $f_Z(x) = F'_Z(x)$

$$f_Z(x) = 2(F(x))^1 F'(x) f(x) = \begin{cases} 2(1 - e^{-x})(e^{-x}) & \text{if } 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

Computation

The following *MATLAB* code produced the accompanying graphs.

```
alpha = 1;
dataX = exprnd(alpha,100000,1);
dataY = exprnd(alpha,100000,1);
dataZ = max(dataX,dataY);

intervalWidth = 0.2;

subplot(2,2,1)
h = hist(dataZ,0.1:intervalWidth:7.1);
bar(0.1:intervalWidth:7.1,h,'y')
title('Histogram of Maximum Random Data')

subplot(2,2,2)
h = h/sum(h);
bar(0.1:intervalWidth:7.1,h,'y')
title('RFH of Maximum Random Data')

subplot(2,2,3)
h = h/intervalWidth;
bar(0.1:intervalWidth:7.1,h,'y')
title('MRFH of Maximum Random Data')
axis([-1 7 -0.1 0.8])

subplot(2,2,4)
x = 0:0.001:12;
theoryZ = 2*(1-exp(-x)).*(exp(-x));
bar(0.1:intervalWidth:7.1,h,'y')
title('MRFH: Theory Superimposed')
hold on
plot(x,theoryZ)
axis([-1 7 -0.1 0.8])
```


Histogram of Maximum Random Data
RFH of Maximum Random Data
MRFH of Maximum Random Data
MRFH: Theory Superimposed
12000
10000
8000
6000
4000
2000
0
0.12
0.1
0.08
0.06
0.04
0.02
0.8
0.6
0.4
0.2
0
2
4
6
8