# Statistics 251: Lab 4 Pre-reading
# Minimum and Maximum of Independent Random Variables

**Objectives:**

- Draw samples from certain distributions in R.
- Visualize distribution of minimum or maximum of independent random variables by simulations.
- Calculate probabilities using R and compare with results calculated by hand.

## 1 Calculating pdf's and probabilities by hand

### 1.1 Maximum

Suppose a system consists of two components, A and B, connected in parallel (see Figure 1). The lifetime of components A and B are independent exponential random variables with mean 40 and 50 respectively (in hours).

(i) What is the density function of the lifetime of the system? Does it also follow an exponential distribution?

(ii) What is the probability that the system will fail before 45 hours?

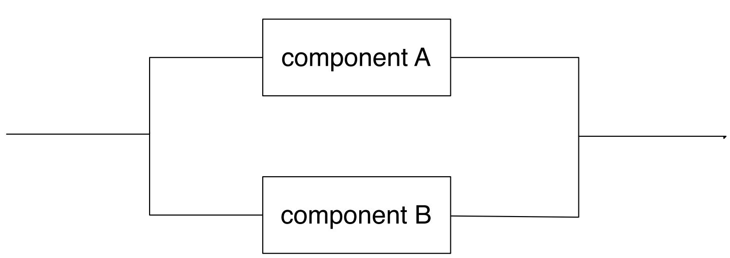



Figure 1: A system of two components connected in parallel.

Let $X_A$ be the lifetime of component A, $X_B$ be the lifetime of component B, $Y$ be the lifetime of the system. From the question, we know that $X_A \sim \exp(1/40)$, $X_B \sim \exp(1/50)$, and $Y = \max(X_A, X_B)$.

(i) To find the density function of $Y$, you need to calculate the cdf first, that is,

$$\begin{aligned}
F(y) &= P(Y \leq y) \\
&= P(X_A \leq y \text{ and } X_B \leq y) \\
&= P(X_A \leq y) \times P(X_B \leq y) \qquad \text{because } X_A \text{ and } X_B \text{ are independent} \\
&= (1 - e^{-y/40}) \times (1 - e^{-y/50}) \\
&= 1 - e^{-y/40} - e^{-y/50} + e^{-9y/200}
\end{aligned}$$

Then, the pdf of $Y$ is

$$f(y) = F'(y) = \frac{1}{40}e^{-y/40} + \frac{1}{50}e^{-y/50} - \frac{9}{200}e^{-9y/200},$$

which means $Y$ doesn't follow an exponential distribution.

(ii) By (i), we have

$$\begin{aligned}
P(Y < 45) &= P(Y \leq 45) \qquad \text{because Y is a continuous random variable} \\
&= 1 - e^{-45/40} - e^{-45/50} + e^{-9\times 45/200} \\
&= 0.4
\end{aligned}$$

## 1.2 Minimum

Consider the case when the two components are connected in series (see Figure 2).

(i) What are the pdf and distribution of $Y$ now?

(ii) What is the probability that the system will fail before 45 hours?

Note that in this case, the lifetime of the system is defined as $Y = \min(X_A, X_B)$.

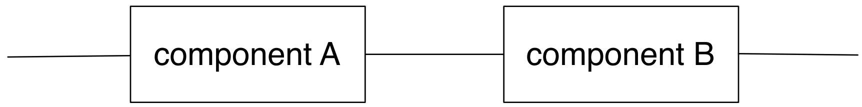



Figure 2: A system of two components connected in series.

(i) Again, we calculate the cdf of $Y$ first,

$$\begin{aligned}
F(y) &= P(Y \leq y) \\
&= 1 - P(Y > y) \\
&= 1 - P(X_A > y \text{ and } X_B > y) \\
&= 1 - P(X_A > y) \times P(X_B > y) \qquad \text{because } X_A \text{ and } X_B \text{ are independent} \\
&= 1 - e^{-y/40} \times e^{-y/50} \\
&= 1 - e^{-9y/200}
\end{aligned}$$

Then the pdf of $Y$ is

$$f(y) = F'(y) = \frac{9}{200} e^{-9y/200}$$

and $Y$ follows an exponential distribution with rate=9/200.

(ii) By (i), we have

$$\begin{aligned}
P(Y < 45) &= 1 - e^{-9 \times 45/200} \\
&= 0.87
\end{aligned}$$

# 2 Simulations by R

In Section 1, we calculated the pdf's and probabilities by hand. Sometimes, the calculation may be very challenging and finding the right answer won't be easy. A good way to double-check if your results are correct is to conduct simulations using R. Here, we run some simulations to verify our answers in Section 1.

```
n <- 1000                    # sample size
Xa <- rexp(n, rate=1/40)     # draw samples of lifetime of component A
Xb <- rexp(n, rate=1/50)     # draw samples of lifetime of component B

X <- cbind(Xa, Xb)         # generate a matrix
head(X)                    # check the first parts of the matrix
```

We first consider the maximum case where $Y = \max(X_A, X_B)$.

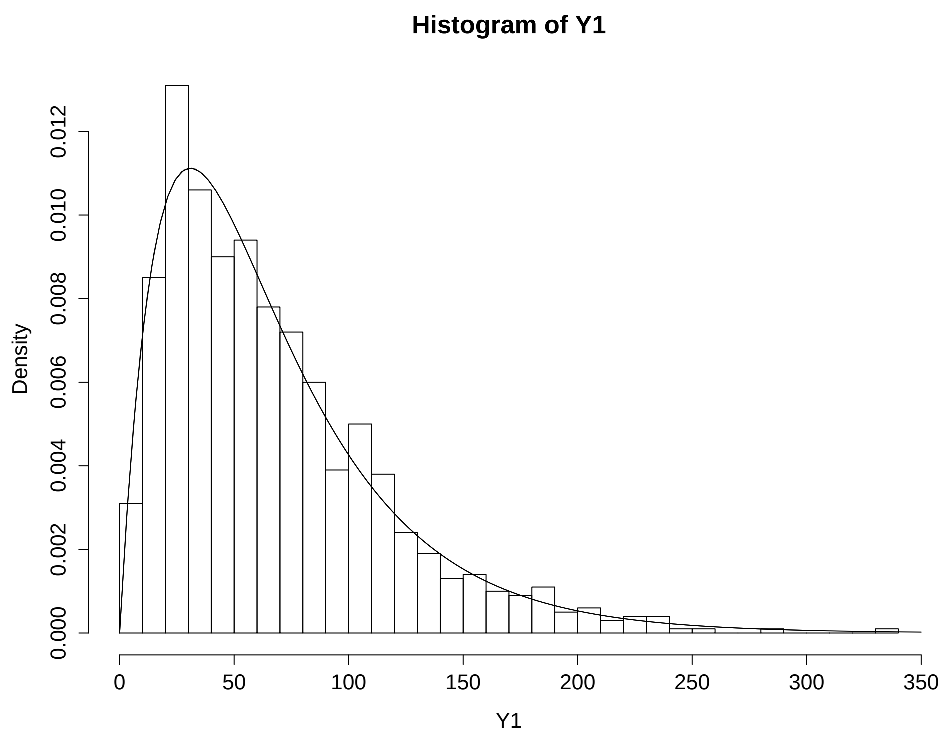
```
Y1 <- apply(X, 1, max)   # get the max value in each row, which is
                         # the lifetime of the system

hist(Y1, prob=T, breaks=40)   # draw a histogram

# define a function of the pdf in Section 1.1
f1 <- function(y){
       1/40*exp(-y/40)+1/50*exp(-y/50)-9/200*exp(-9*y/200)
      }

# draw a curve based on the above function and add it to the histogram
curve(f1, from = 0, to = 350, add=T)
```

In the histogram of $Y_1$, the pattern of the probability density of the sample is very close to the curve of the pdf function. Therefore, our calculation in Section 1.1 was correct! :D




Now we estimate the probability $P(Y < 45)$ based on the sample.

```
p1 <- sum(Y1 < 45)/n    # return 0.406
```

The estimated value is very close to the true value 0.4. Note that as the sample size $n$ increases, the estimate is more accurate.

Similarly, we consider the minimum case below where $Y = \min(X_A, X_B)$.

```
Y2 <- apply(X, 1, min)    # get the min value in each row
```

```
hist(Y2, prob=T, breaks=40) # draw a histogram

# add a curve based on the pdf in Section 1.2, which is an
# exponential distribution with rate 9/200.
curve(dexp(x, rate=9/200), from=0, to=150, add=T)

# estimated P(Y<45) based on the sampled data
p2 <- sum(Y2 < 45)/n   # return 0.899
```