

L2 Ergodicity

kmg

1/22/2019

Read in Data

```
BorkLoc <- "~/Dropbox/Classes/IAV/Data/Borkenau/Data/"
file.names <- list.files(BorkLoc)
BorkData <- list()
append_data <- matrix(,1,30) # Create a 1x30 matrix of NAs
for (p in 1:length(file.names)){
  BorkData[[p]] <- read.csv(paste0(BorkLoc, file.names[p]), header = FALSE)
  append_data <- rbind(append_data, BorkData[[p]]) # one matrix with all
  # individuals concatenated
}
append_data <- append_data[-1,] # remove row of NAs
```

The grand mean

First, let's look at the average obtained across all the days across all the individuals

```
mean(append_data[,18],na.rm = TRUE)
## [1] 1.679293
sd(append_data[,18], na.rm = TRUE)
## [1] 1.575031
```

Each individual provided the same number of observations, thus they are all equally weighted (i.e., provide equal influence on the overall average).

Cross-sectional means

One might argue that by taking the average this way we are obtaining better measurements than what is seen in cross-sectional research. In cross-sectional research we only have one measurement per person and thus may not capture each individual well. (The signal-to-noise ratio is lower.)

We can randomly select one observation per individual and arrive at a distribution of new mean estimates. Each estimate obtained emulates cross-sectional analyses a bit better. This wasn't done in the reading but we include here for completeness.

If we just do this once it directly corresponds to what we might see in cross-sectional data.

```
temp <- matrix(, length(BorkData), 1)
for (p in 1:length(BorkData)) # select one observation from each person
```

```

    {temp[p] <- sample(BorkData[[p]][,18], 1) # Randomly sample one of the
observations from variable 18
}
mean(temp)
## [1] 1.590909

```

We can do it across numerous iterations, say 1000, to get a sense of what the distribution of means on item 18 might be like for this sample to be studied in a cross-section manner.

```

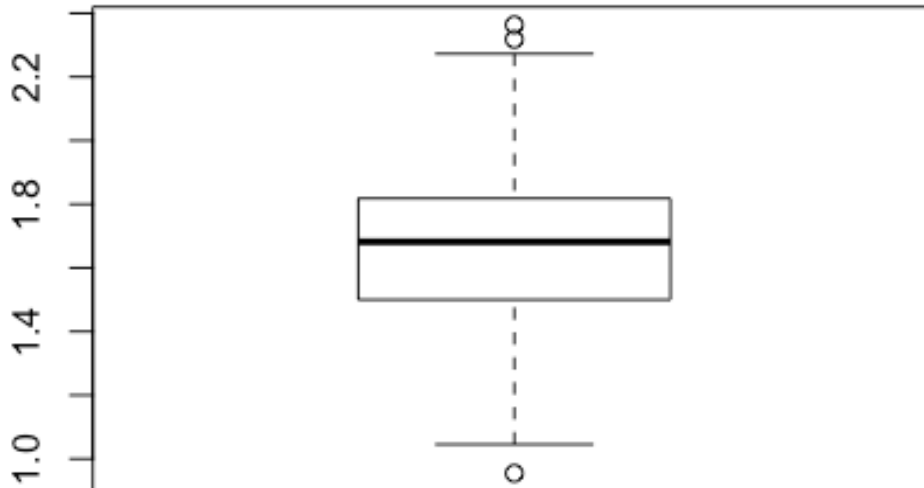
mean_18 <- matrix(,1000,1)
for (iter in 1:1000){
  temp <- matrix(, length(BorkData), 1) # create a new vector with 22 NA
# observations
  for (p in 1:length(BorkData)) # select one observation from each person
  {
    temp[p] <- sample(BorkData[[p]][,18], 1) # Randomly sample one of the
observations
#from variable 18
  }
  mean_18[iter] <- mean(temp)
  ## Note: Your values may differ due to sampling.
}
max(mean_18)
## [1] 2.363636

# The maximum mean found was:
min(mean_18)
## [1] 0.9545455

# The minimum mean found was:
mean(mean_18)
## [1] 1.677682

# Here is the distribution across our iterations:
boxplot(mean_18)

```



We can see that estimates obtained from random draws will vary. However, the average is approximately the same as the original.

Distribution within individuals

It's helpful to see how the mean and variance of values differ for individuals when assess across time.

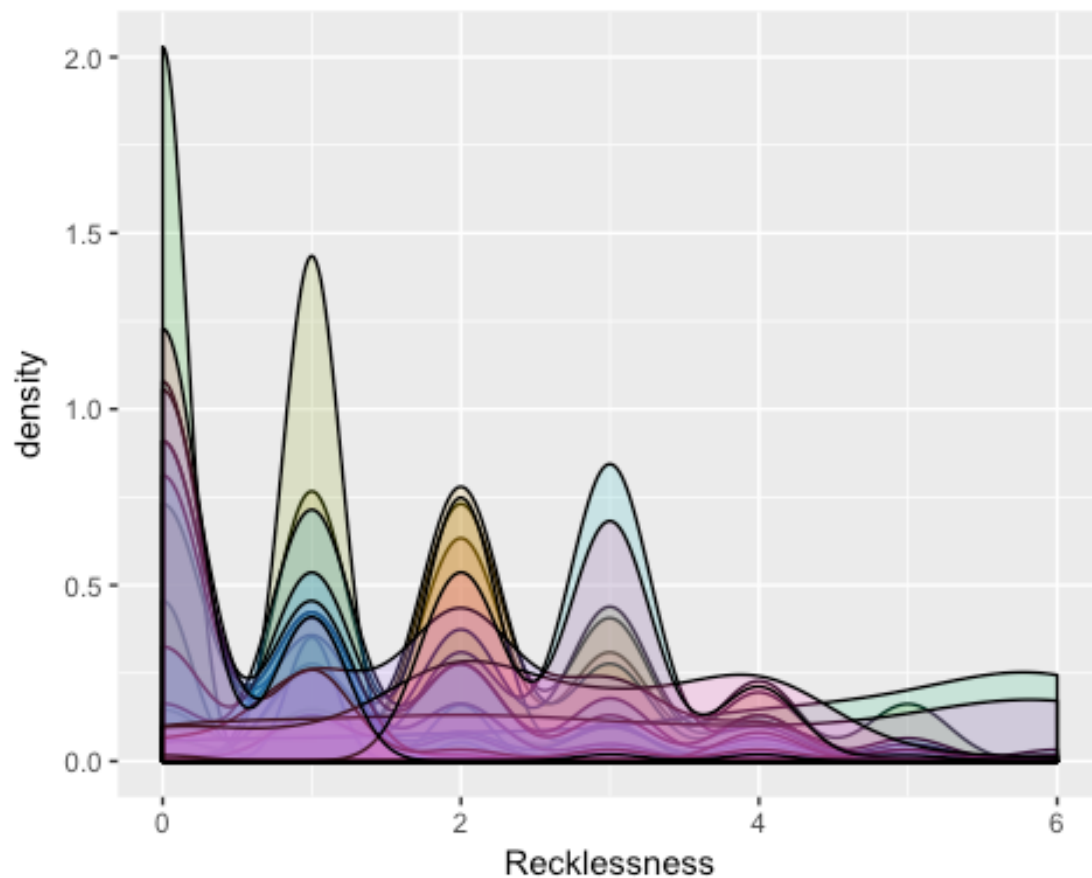
Let's start by looking at the distributions:

```
library(ggplot2)
# Prepare data by adding category variable
for (p in 1:22){
  categories <- rep(paste0("Person", p), 90)
  if (p == 1)
    categories_bind <- t(categories)
  else
    categories_bind <- c(categories_bind, categories)
}

append_data_bind <- cbind(append_data, categories_bind)

# plot histograms of all individuals
```

```
ggplot(append_data_bind, aes(append_data_bind[,18], fill = categories_bind))
+ geom_density(alpha = 0.2) + theme(legend.position="none") +
xlab("Recklessness")
```



Each color represents a different individual

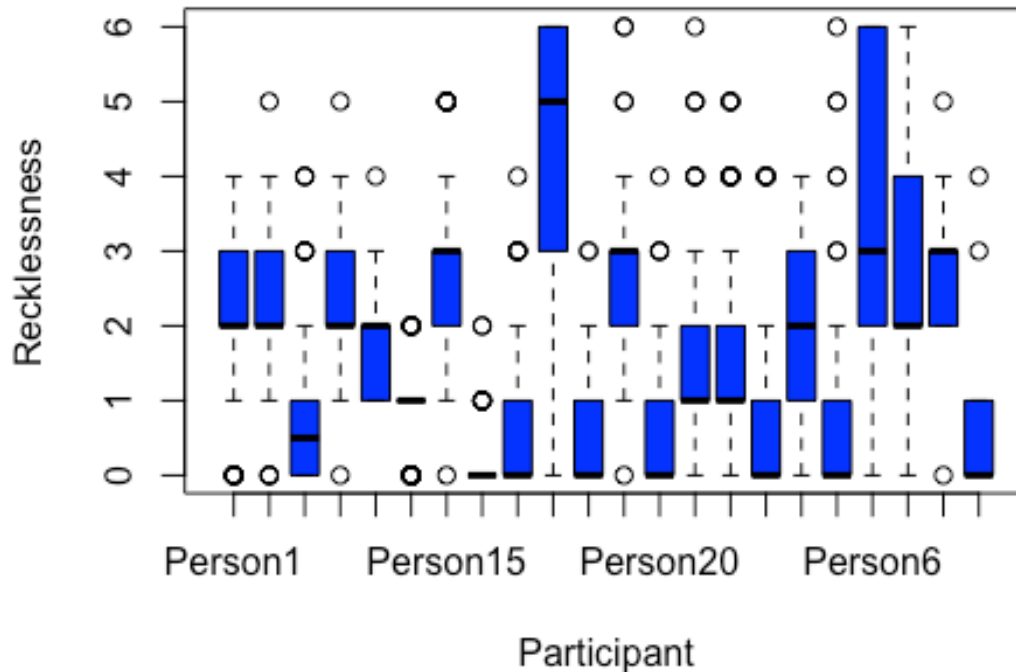
We can see that individuals have different averages, with peaks occurring at reckless = 0, 1, 2, and 4. It appears that different colors (i.e., people) peak at these different values. This suggests that there are subsets of individuals who have these respective averages or modes. Some of the individuals have high kurtosis with means at these values.

We also see that some individuals have a flattened histograms. This suggests a uniform distribution whereby their values on reckless varies greatly.

Another way to investigate these results is to look at boxplots.

```
boxplot(V18~categories_bind,data=append_data_bind, main="Distribution of
Recklessness Values Across Time", xlab="Participant", ylab="Recklessness",
col = "blue")
```

Distribution of Recklessness Values Across Time



Autocovariance function

We have seen that individuals here differ in their averages and standard deviations in at least one variable. We can also look at how the individuals may vary in the temporal relations of the variables.

The autocovariance function was introduced in Chapter 2:

$$c(u) = T^{-1}\Sigma[y(t)y(t - u)]$$

where Σ here is the summation operator. This indicates that we add up each value of y at t times the value of y at $t - u$.

Let's set $u = 1$ for an autocovariance estimate at a lag of one. We can calculate that estimate for the first individual with the following code:

```
# Create the lagged time series  
embedded <- embed(scale(BorkData[[1]][,18], center = TRUE, scale = FALSE), 2)  
# This 'embedded' matrix has two variables: the first is Reckless at t and  
the second is reckless at a lag of t-1.  
  
# We can add the products of each pair of at each row by:
```

```
t(embedded[,1])%*%embedded[,2]/90
```

```
##           [,1]  
## [1,] 0.4327846
```

```
# the t() above transposes the first vector so we can multiply them.  
# T = 90 is the denominator
```

Fortunately there are functions already available that will do this for us at each lag.

Let's verify that we get the same value.

```
library(forecast)  
Acf(BorkData[[1]][,18], lag.max = 1, type = c("covariance"), na.action =  
na.contiguous, demean = TRUE, plot = FALSE)
```

```
##  
## Autocovariances of series 'BorkData[[1]][, 18]', by lag  
##  
##      0      1  
## 1.284 0.433
```

Activities

Try to do the following:

1. Find out the lag-0 autocovariance for variable 18 for Person 1 using the matrix multiplication approach. What does this value mean? It should match the first value provided by the Acf function.
2. Pick a variable that you think might be more consistent in average value across individuals (see the labels provided with the data). Replicate what we did above for 'Reckless'. What is the same, and what is different?
3. In Chapter 2 there was a boxplot for all the autocovariance estimates at a lag of 1 across the individuals. Can you replicate this figure?