BUDS Training: Expectations Lab

Raven Ico

Problem

The mean of a random variable can be thought of as a long run average of repeated samples (the weak law of large numbers). In this problem we will see that sample means from a normal distribution tend to a limit as the sample size increases but sample means from a Cauchy distribution do not. This is a visual illustration of why the Cauchy distribution doesn't have a mean.

The following code generates 1000 random samples of a Cauchy distribution and calculates the cumulative mean. The first element of `sim_df$cum_mean` is equal to the first element of `sim_df$cauchy_rv`, the second element of `sim_df$cum_cauchy` is equal to `mean(sim_df$cauchy_rv[1:2])`, `sim_df$cum_cauchy[10] = mean(sim_df$cauchy_rv[1:10])`, etc.

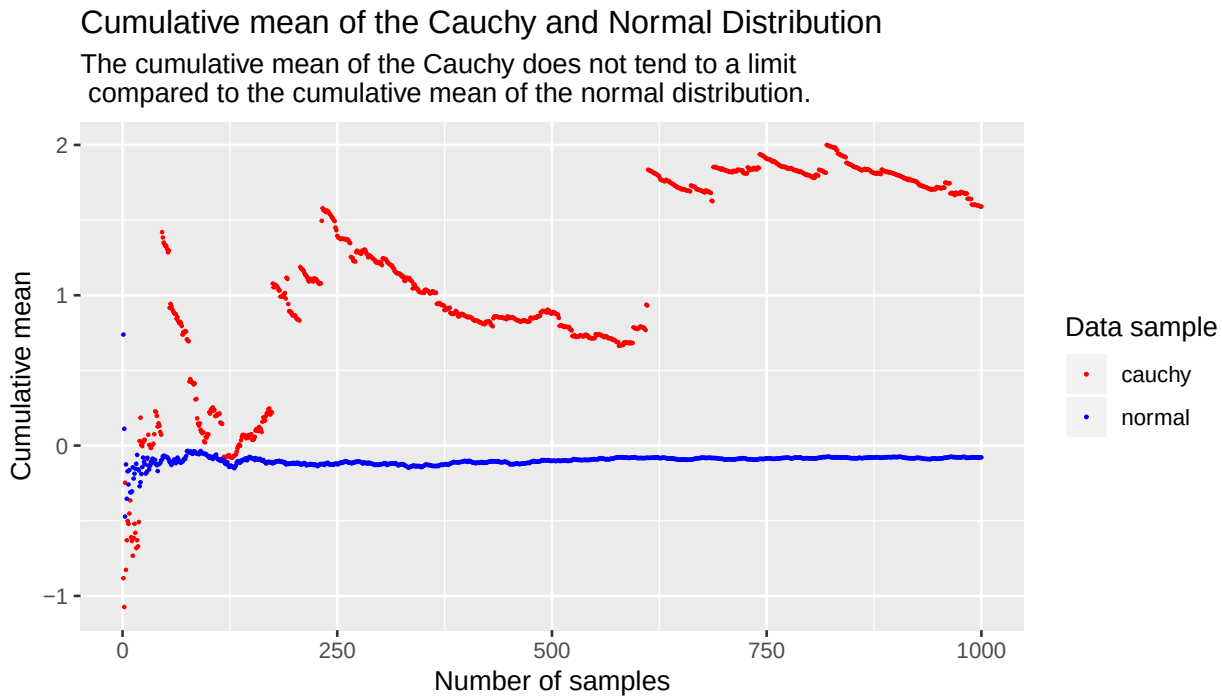
```
set.seed(2019) # Makes our results reproducible
sim_df <- data.frame(n = 1:1000,
                     cauchy_rv = rcauchy(1000, location = 0, scale = 1))

sim_df$cum_cauchy = cumsum(sim_df$cauchy_rv)/sim_df$n
```

Add two more columns to the dataset, the first should be 1000 random samples from a standard normal distribution and the second should be the cumulative mean of these samples. Create a pretty plot showing the cumulative mean from each of the distributions by the number of samples used in the mean. Make sure to add informative labels for the axes.

```
set.seed(2019) # Makes our results reproducible
sim_df$norm_rv <- rnorm(1000, mean = 0, sd = 1)
sim_df$cum_norm <- cumsum(sim_df$norm_rv)/sim_df$n

ggplot(sim_df)+
  geom_point(mapping=aes(n,cum_cauchy, col='cauchy'), size=0.2)+
  geom_point(mapping=aes(n,cum_norm,col='normal'), size=0.2)+
  labs(x = 'Number of samples', y='Cumulative mean',
       title = 'Cumulative mean of the Cauchy and Normal Distribution',
       subtitle='The cumulative mean of the Cauchy does not tend to a limit \n compared to the cumulati
  scale_colour_manual(name="Data sample",values=c(cauchy="red", normal="blue"))
```


Cumulative mean of the Cauchy and Normal Distribution
The cumulative mean of the Cauchy does not tend to a limit
compared to the cumulative mean of the normal distribution.
Cumulative mean
2
1
0
−1
0
250
500
750
1000
Number of samples
Data sample
cauchy
normal