Central Limit Theorem

Gabriele Degola

25/11/2021

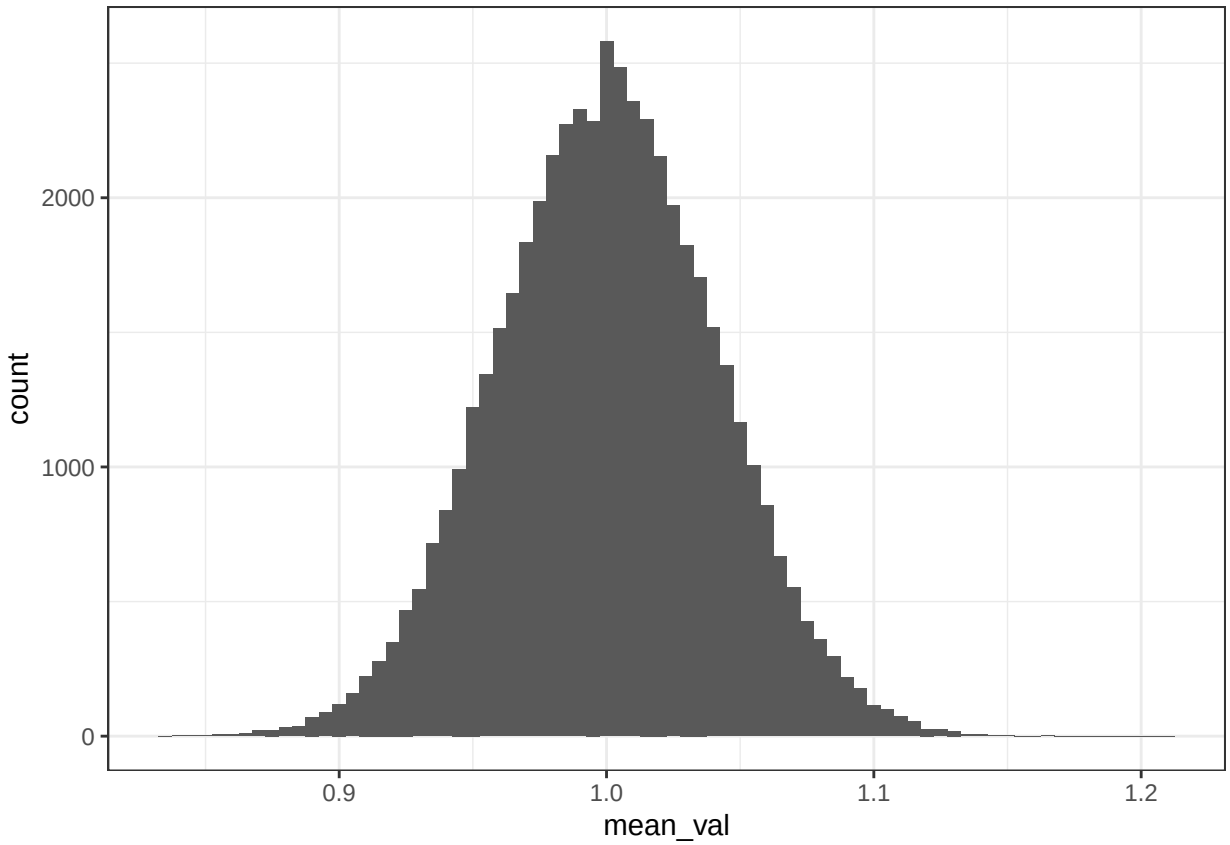
```
library(tidyverse)
```

```
P = 50000
N = 200
df = data.frame(val = runif(n=N*P, min = 0, max = 2), group = 1:P)
df_agg <- df %>% group_by(group) %>% summarise(mean_val = mean(val))
# df_agg$mean_val
# probability of having mean value in the interval
mean(df_agg$mean_val < 1.01 & df_agg$mean_val > 0.99)
```

```
## [1] 0.1946
```

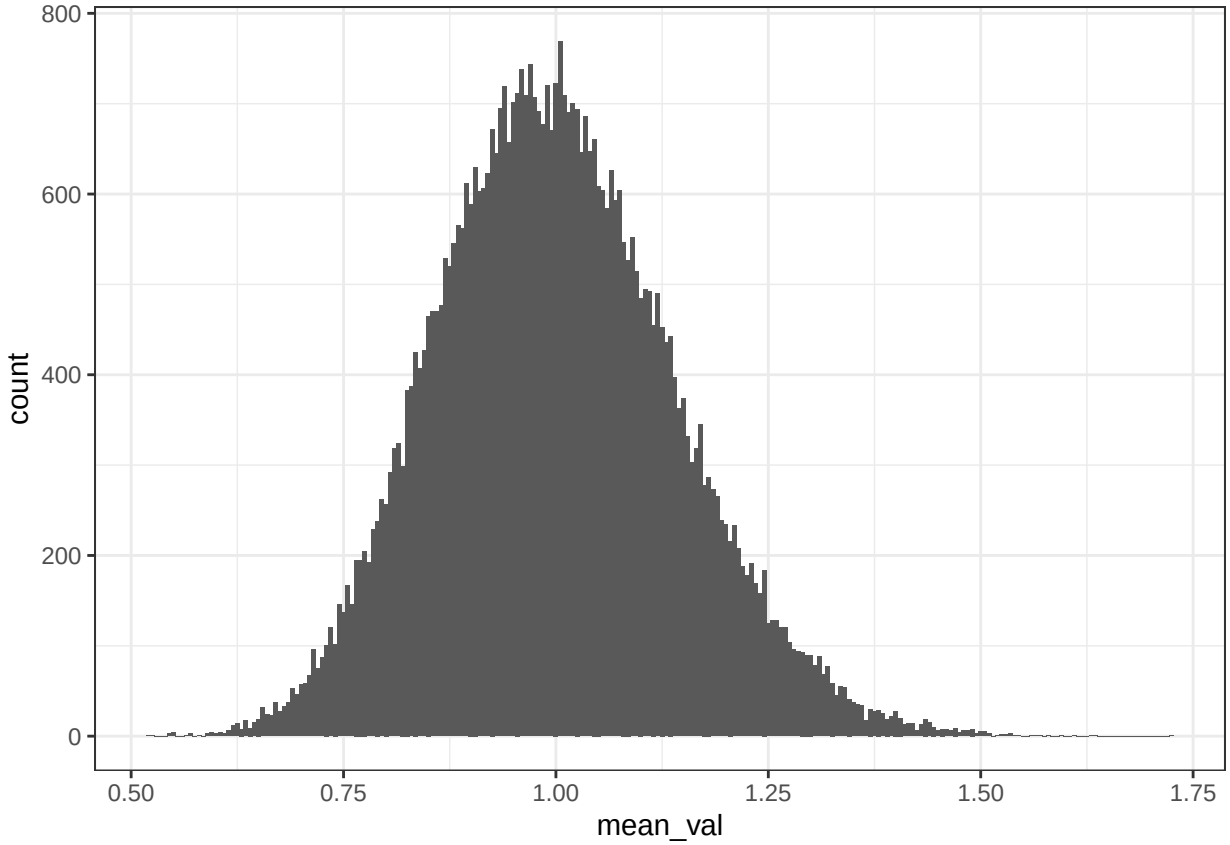
Experimentally, to be 10 times more precise we need 100 more observations.

```
df_agg %>% ggplot() + geom_histogram(aes(x=mean_val), binwidth = 0.005) + theme_bw()# +
↪  xlim(0, 2)
```




```
P = 50000
N = 50
df = data.frame(val = rexp(n=N*P, rate=1), group = 1:P)
df_agg <- df %>% group_by(group) %>% summarise(mean_val = mean(val))

df_agg %>% ggplot() + geom_histogram(aes(x=mean_val), binwidth = 0.005) + theme_bw()
```



Whatever the original distribution from which data are drawn, distribution converges to normal distribution for $\lim_{N \to \inf} N$. The more experiments are made, the more confident we are.