

# Central Limit Theorem

Arnaud Legrand

11/25/2021

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##     filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
N = 20  
mean(runif(n = N, min = 0, max = 2))
```

```
## [1] 0.6805527
```

```
N = 10  
P = 50  
X = runif(n = N*P, min = 0, max = 2)  
df = data.frame(val = X, group = 1:P)  
df %>% group_by(group) %>% summarize(val_mean = mean(val), val_sd = sd(val)) -> df_agg  
df_agg$val_mean
```

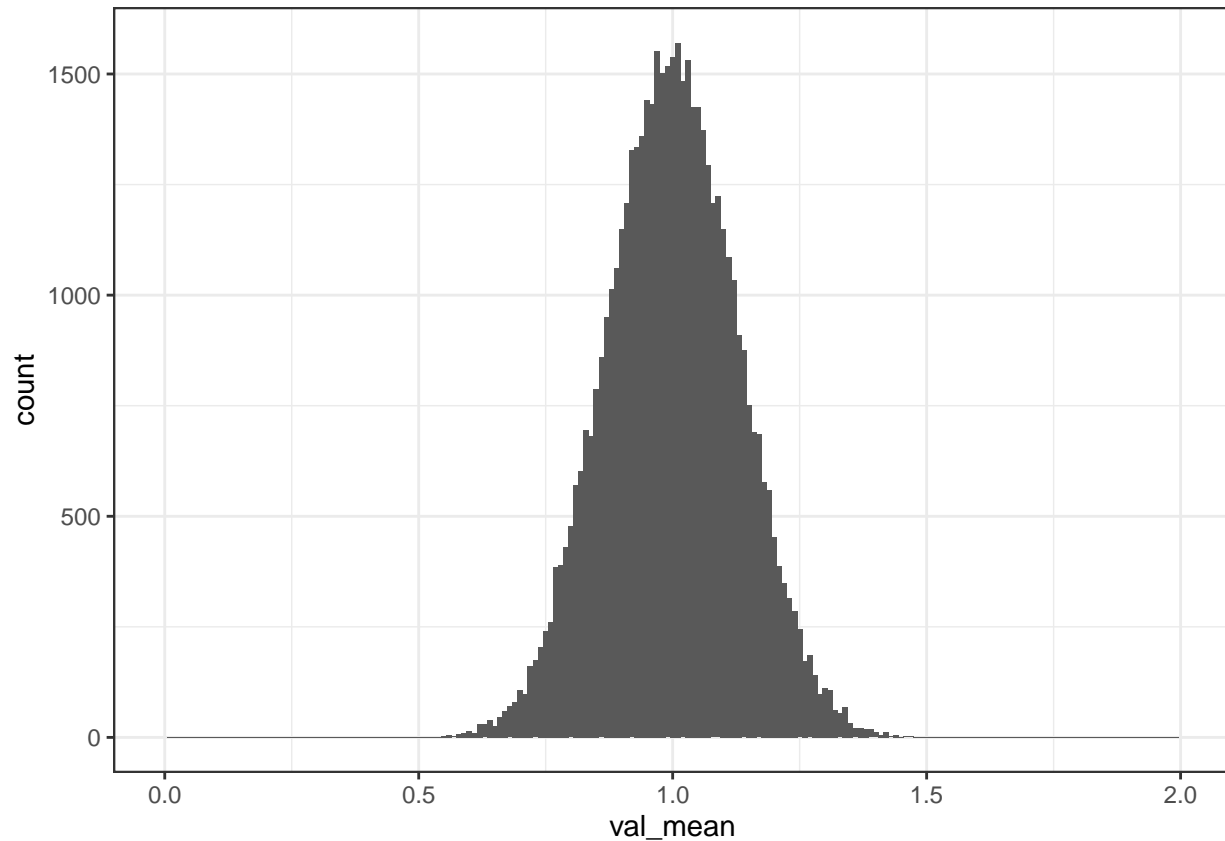
```
## [1] 1.0902048 0.8834930 1.2262204 1.0558558 0.9310699 1.0202701 0.9131170  
## [8] 0.7659628 1.0966566 1.0777541 0.7778909 0.7089121 1.0370745 0.9294479  
## [15] 1.0774324 0.9635767 0.8905227 0.9591402 1.0617740 1.0765316 0.9893997  
## [22] 1.2605089 1.1066899 1.2740734 0.7271290 0.9227281 1.0759741 0.8339250  
## [29] 0.9545622 1.1725405 1.0799072 0.6856388 0.8789199 1.2260159 1.0594542  
## [36] 1.2421002 0.9736293 1.4731255 0.9361124 0.8512528 0.9978027 1.0338051  
## [43] 0.8561914 1.0058075 1.0822441 0.7399158 0.9090966 1.0149396 0.8661574  
## [50] 0.9838605
```

```

N = 20
P = 50000
X = runif(n = N*P, min = 0, max = 2)
df = data.frame(val = X, group = 1:P)
df %>% group_by(group) %>% summarize(val_mean = mean(val), val_sd = sd(val)) -> df_agg
df_agg %>% ggplot() + geom_histogram(aes(x = val_mean), binwidth = 0.01) + theme_bw() + xlim(0,2)

```

```
## Warning: Removed 2 rows containing missing values (geom_bar).
```



```

N = 1000
P = 5000
X = runif(n = N*P, min = 0, max = 2)
df = data.frame(val = X, group = 1:P)
df %>% group_by(group) %>% summarize(val_mean = mean(val), val_sd = sd(val),
                                   val_se = val_sd/sqrt(n()), ok = 1<val_mean+2*val_se & 1>val_mean-2*val_se)
mean(df_agg$ok)

```

```
## [1] 0.9526
```