STATISTICAL INFERENCE - Peer Assignment

Version: V00

Date: 21-dec-2014

GitHub Repository: <https://github.com/A6111E/datasciencecoursera/tree/master/Statistical_Inference/Assignment_Part1>

#### Introduction:

**Exponential Distribution:** is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate.

The exponential distribution occurs naturally when describing the lengths of the inter-arrival times in a homogeneous Poisson process. Describes the time for a continuous process to change state.

In real scenarios, the assumption of a constant rate (or probability per unit time) is rarely satisfied.

For example, the rate of incoming phone calls differs according to the time of day. But if we focus on a time interval during which the rate is roughly constant, such as from 2 to 4 p.m. during work days, the exponential distribution can be used as a good approximate model for the time until the next phone call arrives.

Examples which yield approximately exponentially distributed variables:

* Time until a radioactive particle decays, or the time between clicks of a geiger counter.
* Time it takes before your next telephone call

Exponential variables can also be used to model situations where certain events occur with a constant probability per unit length, such as the distance between mutations on a DNA strand, or between roadkills on a given road.

**Expected Value:** the mean or expected value of an exponentially distributed random variable with rate parameter is given by:

**Variance:**

**Standard Deviation (is equal to the mean):**

**Mean Standard Error:**

**Sample Mean:**

**t Confidence Interval:**

*Partial Source:* <http://en.wikipedia.org/wiki/Exponential_distribution>

### Data Processing:

#### Data Set

#### Table: Summary Simulation Random Exponential Distribution

##### Simulations = 1000 - Sample Size = 40 - lambda = 0.2

colNames <- c("Trial\_1", "Trial\_100", "Trial\_500", "Trial\_750", "Trial\_1000")  
pr1 <- DT[ , colNames, with = FALSE]  
  
xt <- xtable(summary(pr1))  
print(xt, type = "html", floating = FALSE)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Trial\_1** | **Trial\_100** | **Trial\_500** | **Trial\_750** | **Trial\_1000** |
| Min. : 0 | Min. : 0 | Min. : 0.2 | Min. : 0.1 | Min. : 0.0 |
| 1st Qu.: 2 | 1st Qu.: 2 | 1st Qu.: 1.0 | 1st Qu.: 0.4 | 1st Qu.: 1.5 |
| Median : 4 | Median : 3 | Median : 2.3 | Median : 1.7 | Median : 3.7 |
| Mean : 6 | Mean : 5 | Mean : 4.0 | Mean : 2.7 | Mean : 4.9 |
| 3rd Qu.: 6 | 3rd Qu.: 6 | 3rd Qu.: 5.5 | 3rd Qu.: 3.3 | 3rd Qu.: 7.2 |
| Max. :32 | Max. :37 | Max. :16.9 | Max. :16.1 | Max. :21.0 |

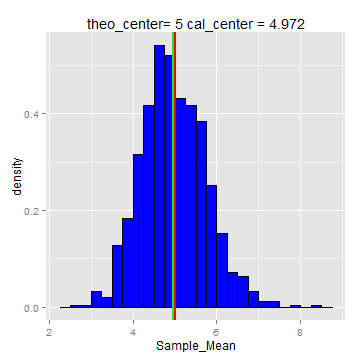
#### Table: Mean Value for each Simulation

xt <- xtable(DT1[1:10, ], digits = 2)  
print(xt, type = "html", include.rownames = FALSE, hline.after=c(1), floating = FALSE)

|  |  |
| --- | --- |
| **Trial** | **Sample\_Mean** |
| 1 | 5.55 |
| 2 | 4.67 |
| 3 | 4.26 |
| 4 | 4.61 |
| 5 | 5.30 |
| 6 | 3.63 |
| 7 | 6.23 |
| 8 | 3.68 |
| 9 | 4.84 |
| 10 | 4.98 |

#### Graphic: Random Exponential Distribution

theo\_center <- mu # distribution theoretical center  
cal\_center <- mean(p) # distribution calculated center  
  
p1 <- ggplot(DT1, aes(x = Sample\_Mean))  
 p1 <- p1 + geom\_histogram(fill = "blue",  
 binwidth = 0.25, aes(y = ..density..), colour = "black")  
 p1 <- p1 + geom\_vline(xintercept = theo\_center, size = 0.75, colour = "red")  
 p1 <- p1 + geom\_vline(xintercept = cal\_center, size = 0.75, colour = "green")  
 mse <- round(cal\_center, 3)  
 p1 <- p1 + labs(title = paste('theo\_center = ', theo\_center, ' cal\_center = ', mse))  
grid.arrange(p1, ncol = 1)



Remarks: Red Vertical Line: distribution theoretical center / Green Vertical Line: distribution calculated center

##### Q1 - Results:

The calculated center from 1000 simulations for exponential distributions with , is approximately , which corresponds to theoretical center of distribution ().

Increasing the simulations, the calculated gets closer to theoretical center.

**A. Theoretical Center:**

**B. Calculated Center:**

##### Q2 - Results:

The calculated variance from simulations for exponential distributions samples with and theoretical variance are:

**A. Theoretical Variance:**

**B. Calculated Variance:**

Increasing the simulations, the calculated gets near to theoretical variance, and the exponential distributions trend to be a standard normal distribution.

For example:

A. The calculated variance from simulation for exponential distributions random variables is (for this simulation).

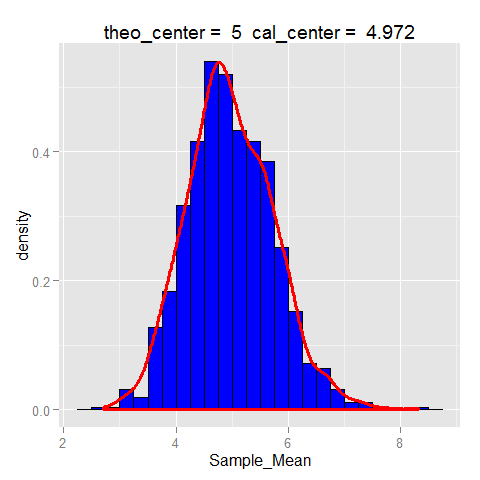
B. The theoretical variance , is

Decreasing the simulations, the calculated variance gets far away to theoretical variance, and the exponential distributions trend NOT to be a standard normal distribution.

##### Q3 - Results:

#### Graphic: Random Exponential Distribution ~ Standard Normal Distribution

p3 <- ggplot(DT1, aes(x = Sample\_Mean))  
 p3 <- p3 + geom\_histogram(fill = "blue",  
 binwidth = 0.25, aes(y = ..density..), colour = "black")  
 p3 <- p3 + geom\_density(size = 1.25, colour = "red")  
 mse <- round(cal\_center, 3)  
 p3 <- p3 + labs(title = paste('theo\_center = ', mu, ' cal\_center = ', mse))  
grid.arrange(p3, ncol = 1)



For a standard normal distribution, , and of the normal density, lies within , and standard deviations from the mean, respectively

The calculted percentage from the sample means of the exponential distribution () are:

**A. Percentage of Values between +/-1 Standard Deviation:**

**B. Percentage of Values between +/-2 Standard Deviation:**

**C. Percentage of Values between +/-3 Standard Deviation:**

Both, the graphic "Random Exponential Distribution ~ Standard Normal Distribution", and the calculated percentage values, indicates that if the sample trend to be infinite, the random exponential distribution will convert to a standard normal distribution.

The confidence interval for the quantification of uncertainty in the exponential dsitribution is:

**A. Confidence Interval 95% - Up:**

**B. Confidence Interval 95% - Down:**

### Session Information

## R version 3.1.2 (2014-10-31)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
##   
## locale:  
## [1] LC\_COLLATE=Spanish\_Colombia.1252 LC\_CTYPE=Spanish\_Colombia.1252   
## [3] LC\_MONETARY=Spanish\_Colombia.1252 LC\_NUMERIC=C   
## [5] LC\_TIME=Spanish\_Colombia.1252   
##   
## attached base packages:  
## [1] grid stats graphics grDevices utils datasets methods   
## [8] base   
##   
## other attached packages:  
## [1] gridExtra\_0.9.1 knitr\_1.8 xtable\_1.7-4 ggplot2\_1.0.0   
## [5] data.table\_1.9.4  
##   
## loaded via a namespace (and not attached):  
## [1] chron\_2.3-45 colorspace\_1.2-4 digest\_0.6.4 evaluate\_0.5.5   
## [5] formatR\_1.0 gtable\_0.1.2 htmltools\_0.2.6 labeling\_0.3   
## [9] MASS\_7.3-35 munsell\_0.4.2 plyr\_1.8.1 proto\_0.3-10   
## [13] Rcpp\_0.11.3 reshape2\_1.4 rmarkdown\_0.3.3 scales\_0.2.4   
## [17] stringr\_0.6.2 tools\_3.1.2 yaml\_2.1.13