Ch 5 : Probability To Statistics

Contents

June 27, 2017

1

Random Sampling

[ToC]

1.1 Probability and Statistics

[ToC]

Random variables $X_1, X_2, \ldots, X_n$ are said to be a **random sample** of size n from distribution F if

1. The X_i's are independent
2. Each X_i has distribution F.

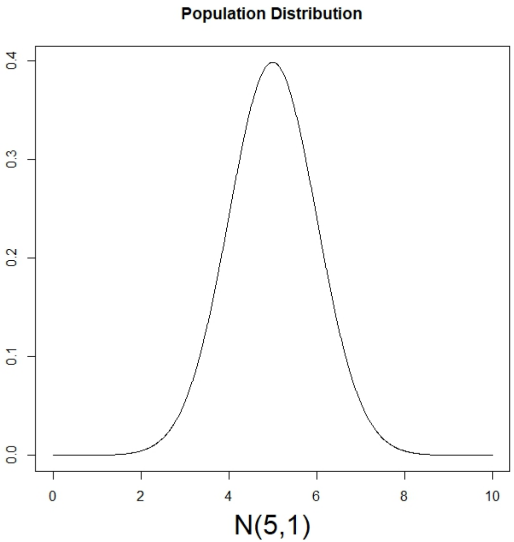
   1st run of the Experiment $->$ realization of X_1
   2nd run of the Experiment $->$ realization of X_2
   3rd run of the Experiment $->$ realization of X_3
   $\vdots$

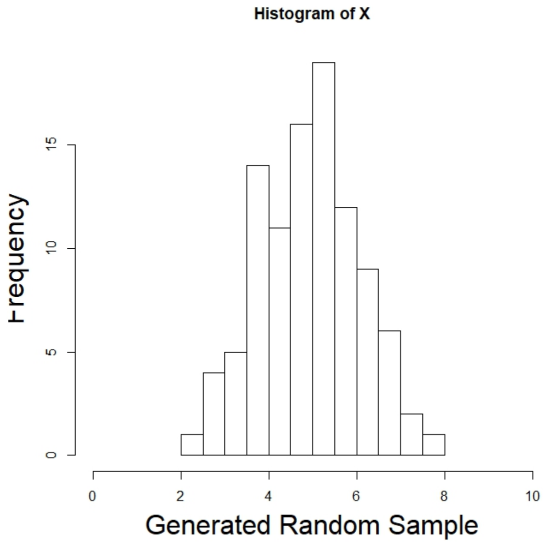
3. $\{X_1\, X_2, \ldots, X_3\}$ is the dataset.
4. F is called the population distribution.

Example: Population Distribution to Data

```
par(mfrow=c(1,2))
x=seq(0,10,.01)
plot(x, dnorm(x, 5,1), type='l', xlab='N(5,1)', ylab='', cex.lab=2, main='Population Distribution')
X <- rnorm(100, 5, 1)
hist(X, 15, xlab='Generated Random Sample', cex.lab=2, xlim=c(0,10))
```

Example: Population Distribution to Data





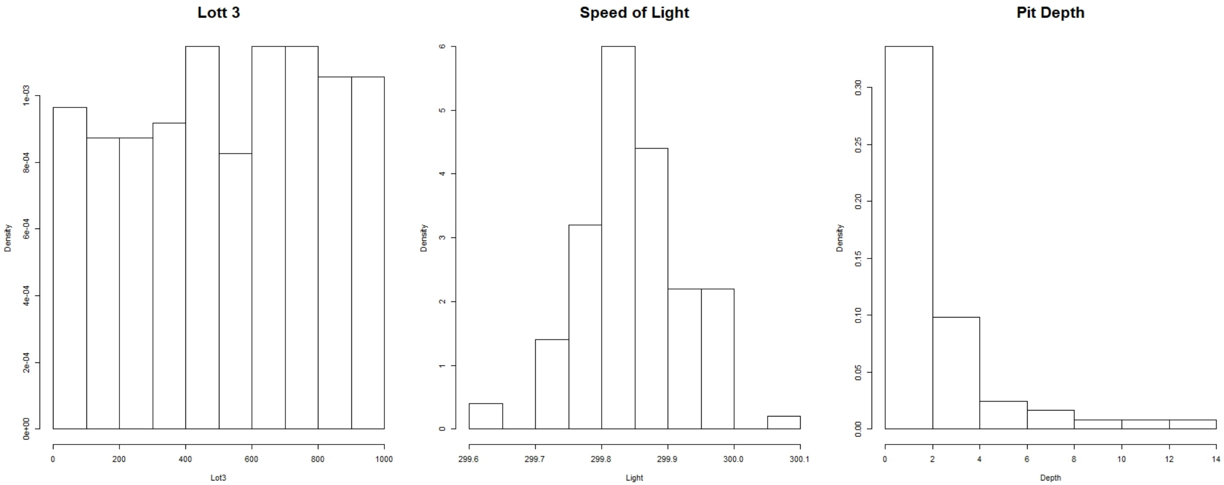
Example: Data to Population Distribution

```
D2 = read.csv("http://gozips.uakron.edu/~nmimoto/pages/datasets/Lottery.txt")
D3 = read.csv("http://gozips.uakron.edu/~nmimoto/pages/datasets/Light.csv")
D4 = read.csv("http://gozips.uakron.edu/~nmimoto/pages/datasets/PitCorrosion.txt")

Lot3=D2$Lot_3    #-$---
Light=D3$Light
Depth=D4$depth


par(mfrow=c(1,3))
hist(Lot3,  freq=F, main='Lott 3', cex.main=2)
hist(Light, freq=F, main='Speed of Light', cex.main=2)
hist(Depth, freq=F, main='Pit Depth', cex.main=2)
```

Example: Data to Population Distribution



Can you guess the shape of f?

1.2 Guessing the population distritution (Distribution Fitting)

[ToC]

---

1. Histogram(data) vs pdf(Theoretical)

2. EDF(data) vs CDF(Theoretical)

3. Probability Plot

$$\text{Sample Percentiles} \quad vs \quad \text{Theoretical Percentiles}$$

Probability Plot (q-q plot)

(From Chapter 4)

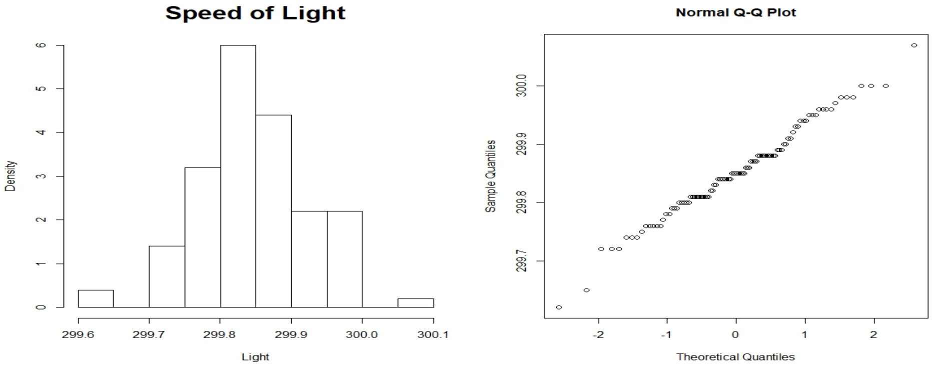
If you guessed the population distribution is A, then you can check your guess by plotting q-q plot. q-q plot is a plot of

$$\left[\ i\text{th ordered obs.}\ \right] \text{ vs } \left[100 \times \frac{(i-.5)}{n}\text{th (theoretical) percentile from your guees of A}\right].$$

If the data are indeed sample from A, then the q-q plot should look like a line.

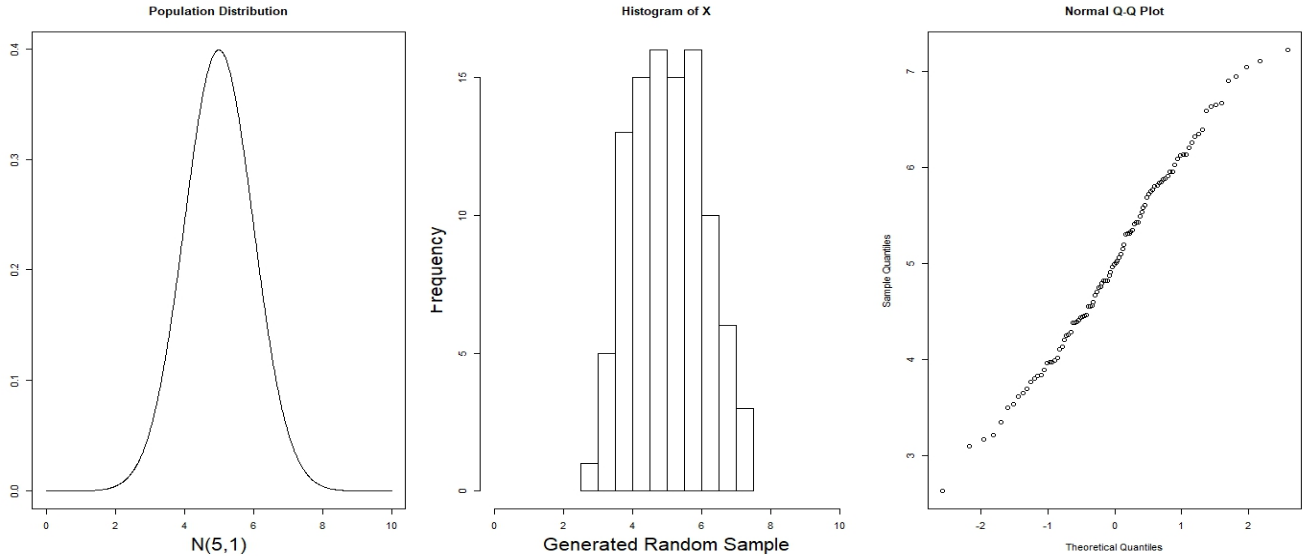
Example: q-q Nomal Plot for Light Data

```
par(mfrow=c(1,2))
hist(Light, freq=F, main='Speed of Light', cex.main=2)
qqnorm(Light)
```



Example: q-q nomal plot

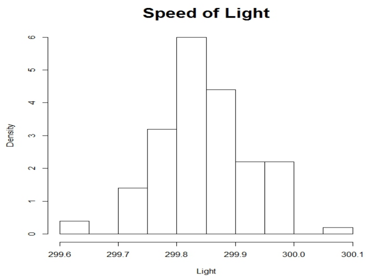
```
n=100
par(mfrow=c(1,3))
x=seq(0,10,.01)
plot(x, dnorm(x, 5,1), type='l', xlab='N(5,1)', ylab='', cex.lab=2, main='Population Distribution')
X <- rnorm(n, 5, 1)
hist(X, 15, xlab='Generated Random Sample', cex.lab=2, xlim=c(0,10))
qqnorm(X)
```

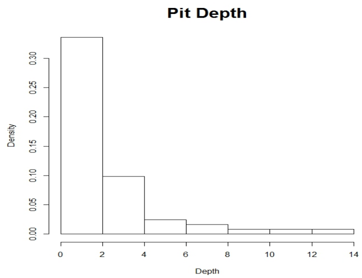

Population Distribution
N(5,1)
Histogram of X
Frequency
Generated Random Sample
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles

1.3 Parameter Estimation

[ToC]

After you Guessed the population distritution, we need to estimate the parameter(s).





$N(\mu, \sigma^2)$?

$\mathrm{Exp}(\lambda)$?

Two Major Methods

for coming up with an estimator in general.

1. Method of Moments

2. Maximum Likelihood Estimation

Method of Moments

- For $X \sim \mathrm{N}(\mu, \sigma^2)$, $E(X) = \mu$. Use $\bar{X}$ to estimate μ.
- For $X \sim \mathrm{Exp}(\lambda)$, $E(X) = 1/\lambda$. Use $1/\bar{X}$ to estimate λ.

Two Important Cases

1. $\{X_1, \ldots, X_n\}$ are R.S. from a population with mean μ and standard deviation σ.

$$\overline{X} \text{ estimates } \mu$$

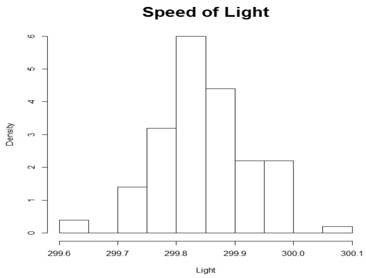
$$S^2 \text{ estimates } \sigma^2$$

2. $\{X_1, \ldots, X_n\}$ are R.S. from a population, which has only 0 or 1 as possible outcomes. Probability for getting 1 for each X_i is p.

$$\overline{X} \text{ estimates } p$$

1.4 Example: Speed of Light Experiment

[ToC]



- Assume that each measurement X_i is a random sample from $\mathrm{N}(c, \sigma^2)$
- That is same thing as to say

$$X_i \sim c + \varepsilon \qquad \epsilon \sim N(0, \sigma^2)$$

- $$\bar{X} \text{ estimates } \; c = [\text{Speed of Light}]$$

- How good is the estimation?

2

Linear Combination of Normal RVs

[ToC]

2.1 Sampling Distribution of the Sample Mean

[ToC]

- What is the distribution of $\bar{X}$ when each X_i is a random sample from Normal distribution?

$$X_i \sim N(\mu, \sigma^2)$$

- applet

http://onlinestatbook.com/stat_sim/sampling_dist/index.html

Theoretical Consideration

- Suppose $X \sim N(2, 3)$ and $Y \sim N(4, 2)$. X and Y are independent.

- What is $P(X + Y < 5) =?$