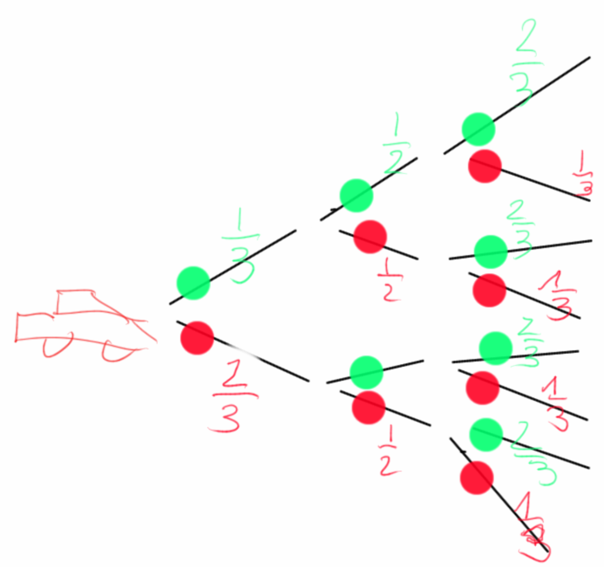
Module 1 - Basic Probability and Statistics

Pawel Chilinski

October 8, 2013

Part 1 - Probability

1 Find a distribution function $F_x(x) = P(X \leq x)$, $x \in R$ of random variable X.



X - random variable equal to a number of times the driver had to stop on his way due to red light in crossroads.

$P(X = 0) = \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{9}$

$P(X = 1) = \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{7}{18}$

$P(X = 2) = \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{7}{18}$

$P(X = 3) = \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{9}$

$$F_x(x) = \begin{cases} 0 & x < 0 \\ 1/9 & 0 \leq x < 1 \\ 1/2 & 1 \leq x < 2 \\ 8/9 & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

2 Find the probability that a person randomly chosen from the population has IQ

a) above 130

```
> 1-pnorm(130,mean=100,sd=15)
[1] 0.02275013
```

b) between 100 and 120

```
> pnorm(120,mean=100,sd=15)-pnorm(100,mean=100,sd=15)
[1] 0.4087888
```

3 A pair of random variables (X, Y) has a joint discrete distribution

a) marginal distributions

X\Y	-1	0	1	p_x
0	0.1	0.1	0	0.2
1	0.2	0.2	0.1	0.5
2	0.1	0.1	0.1	0.3
p_y	0.4	0.4	0.2	

b) Calculate P(X > 2Y)

P(X=x, Y=y) := P(x,y)

P(X > 2Y)=P(0,-1)+P(1,-1)+P(1,0)+P(2,-1)+P(2,0)=0.1+0.2+0.2+0.1+0.1=0.7

c) Are random variables X and Y independent? Calculate Cov(X,Y) and Cor(X,Y)

They are not independent because $P(0,-1) = 0.1 \neq P(X = 0) \cdot P(Y = -1) = 0.2 \cdot 0.4 = 0.08$

```
> EX <- 0.2 * 0 + 0.5 * 1 + 0.3 * 2
> EX

[1] 1.1

> EY <- 0.4 * -1 + 0.4 * 0 + 0.2 * 1
> EY

[1] -0.2

> E_XY <- 0.2 * 1 * -1 + 0.1 * 1 * 1 + 0.1 * 2 * -1 + 0.1 * 2 * 1
> E_XY

[1] -0.1

> var_X <- 0.2*(0-EX)^2 + 0.5*(1-EX)^2 + 0.3*(2-EX)^2
> var_X

[1] 0.49

> var_Y <- 0.4*(-1-EY)^2 + 0.4*(0-EY)^2 + 0.2*(1-EY)^2
> var_Y

[1] 0.56

> cov_XY <- E_XY-EX*EY
> cov_XY

[1] 0.12

> cor_XY <- cov_XY/sqrt(var_X*var_Y)
> cor_XY

[1] 0.2290811
```

d) Find conditional distribution

$P(X|Y=-1) = \frac{P(X,Y=-1)}{P(Y=-1)}$

X|Y=-1	0	1	2
	$\frac{0.1}{0.4}$	$\frac{0.2}{0.4}$	$\frac{0.1}{0.4}$

=

X|Y=-1	0	1	2
	0.25	0.5	0.25

$P(Y|X=0) = \frac{P(Y,X=0)}{P(X=0)}$

$$\begin{array}{c|ccc} \text{Y|X=0} & \text{-1} & 0 & 1 \\ \hline & \frac{0.1}{0.2} & \frac{0.1}{0.2} & \frac{0}{0.2} \end{array} = \begin{array}{c|ccc} \text{Y|X=0} & \text{-1} & 0 & 1 \\ \hline & 0.5 & 0.5 & 0 \end{array}$$

4 Using central limit theorem find an approximate distribution of the time in which a cyclist covers 50km of the route

The expected value of the time that cyclist needs to complete 1km is: $EX = \frac{1.4+1.8}{2} = 1.6$ min
The variance of the time that cyclist needs to complete 1km is: $\sigma_x^2 = \frac{(1.8-1.4)^2}{12} = 0.0133$

So according to CLT the we can model the time needed to complete 50 km route by normal distribution $N(50 \cdot 1.6, 50 \cdot 0.0133) = N(80, 0.665)$.

Part 2 - Statistics

5 Find 95% confidence interval for the expected value of a textbook price

Because $\frac{\bar{P}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ then

```
> mu_p <- 28.4
> sd_p <- 4.75
> n <- 50
> conf_interval_0_95 <- c(mu_p-qnorm(0.975)*sd_p/sqrt(n) , mu_p + qnorm(0.975)*sd_p/sqrt(n))
> conf_interval_0_95

[1] 27.08339 29.71661
```

6 Perform a statistical test to validate their hypothesis, with significance level equal to 0.05.

$H_0 : fraction = 0.25$
$H_1 : fraction \neq 0.25$

Here we deal with example of binomial distribution. And we standardize fraction assuming H_0 holds:

```
> n <- 400
> frac_est <- 79/400
> frac_0 <- 0.25
> frac_standardized <- (p_est-p_0)/sqrt(p_est*(1-p_est)/n)
> frac_standardized

[1] -2.637444
```

and critical region for standardized fraction is:

```
> left_critical_region <- c(-Inf, qnorm(0.025))
> left_critical_region

[1]      -Inf -1.959964
```

$\cup$

```
> right_critical_region <- c(qnorm(0.975),Inf)
> right_critical_region

[1] 1.959964      Inf
```

So we have to reject H_0 because standardized value falls into critical region:

```
> frac_standardized > left_critical_region[1] & frac_standardized < left_critical_region[2]

[1] TRUE
```

7 Test the hypothesis that the factual mean value of the lenses does not meet the requirements, with significance level equal to 0.05.

$H_0 : fraction = 3.2$
$H_1 : fraction \neq 3.2$

Here we deal with example with unknon standard deviation:

```
> n <- 50
> w_est <- 3.05
> w_0 <- 3.2
> SE <- 0.34
> w_standardized <- (w_est-w_0)/SE
> w_standardized
```

```
[1] -0.4411765
```

so to compute critical region we are using t-distribution:

```
> left_critical_region <- c(-Inf, qt(0.025,n-1))
> left_critical_region
```

```
[1]      -Inf -2.009575
```

$\bigcup$

```
> right_critical_region <- c(qt(0.975,n-1),Inf)
> right_critical_region
```

```
[1] 2.009575      Inf
```

so the standardized value is not in the critical region

```
> (w_standardized > left_critical_region[1] & w_standardized < left_critical_region[2]) |
+                  (w_standardized > right_critical_region[1] & w_standardized < right_critical_region[2])
```

```
[1] FALSE
```

so we do not reject the H_0 that the factual mean equals required value.