Problems on random vectors (2/2)

Ignacio Cascos, Department of Statistics, Universidad Carlos III de Madrid

2018/19

1. Assume heights for women follow a normal distribution with mean of 64 inches and standard deviation 3 inches, while men's heights follow a normal distribution with mean of 70 inches and standard deviation 3 inches. Assume further that 51% of the individuals in a population are women and the joint distribution of the heights of a couple of siblings (one sister and one brother) is normal with correlation 0.6.
   1. What is the probability that a woman selected at random is taller than 68 inches? Denote the height of a random woman by $X_w \sim \mathrm{N}(64, 3)$, $P(X_w > 68) = 0.0912112$.
   2. What height is exceeded by 65% of women? The solution (on h) to the equation $P(X_w \geq h) = 0.65$ is $h = 62.8440386$.
   3. Compute the mean and variance of the height of an individual selected at random. Denote by X the height of an individual selected at radom. With probabilty 0.51 it will be a woman and with probability 0.49, a man. It is thus a mixture of two normal distributions.

$$\begin{aligned}\mathbb{E}[X] &= 0.51 \times 64 + 0.49 \times 70 = 66.94.\\ \mathbb{E}[X^2] &= 0.51 \times (9 + 64^2) + 0.49 \times (9 + 70^2) = 4498.96.\\ \mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 17.9964.\end{aligned}$$

   4. What is the probability that an individual selected at random is taller than 68 inches? $P(X > 68) = 0.4127964$.
   5. What height is exceeded by 65% of individuals? The solution (on h) to the equation $P(X \geq h) = 0.65$ is $h = 65.0621154$.
   6. Consider now a couple of siblings (one sister and one brother). What is the distribution of their average height? Denote the average height of a couple of siblings by $Y = (X_w + X_m)/2 \sim \mathrm{N}(\mu = 67, \sigma = \sqrt{9/4 + 9/4 + 2 \times 3 \times 3 \times 0.6/4} = \sqrt{7.2})$.
   7. What is the probability that the midheight from part 7. is greater than 68 inches? $P(Y > 68) = 0.3546941$.
   8. What midheight is exceeded by 65% of couples of siblings (one sister and one brother)? The solution (on h) to the equation $P(Y \geq h) = 0.65$ is $h = 65.9660767$.

```r
library(EnvStats)
1-pnorm(68,mean=64,sd=3)
```

```
## [1] 0.09121122
```

```r
qnorm(0.35,mean=64,sd=3)
```

```
## [1] 62.84404
```

```r
qnormMix(0.35,mean1=64,sd1=3,mean2=70,sd2=3,p.mix=0.49)
```

```
## [1] 65.06212
```

```r
1-.51*pnorm(68,mean=64,sd=3)-.49*pnorm(68,mean=70,sd=3)
```

```
## [1] 0.4127964
```

```r
1-pnorm(68,mean=67,sd=sqrt(7.2))
```

```
## [1] 0.3546941
```

```r
qnorm(0.35,mean=67,sd=sqrt(7.2))
```

```
## [1] 65.96608
```

2. A firm has workers of 5 types with five different salaries

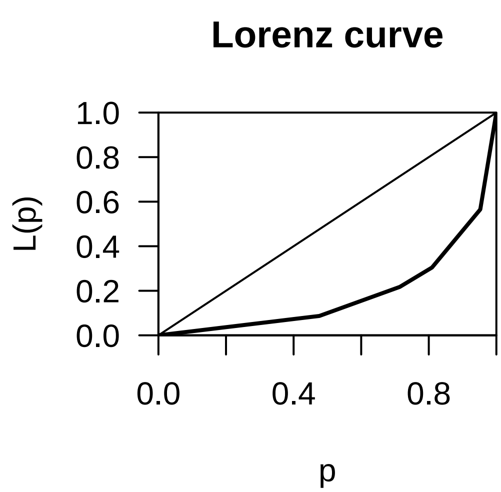
number of workers	10	5	2	3	1
salary	1000	3000	5000	10000	50000

1. Represent the Lorenz curve of the salaries. What is the Gini index?
2. The firm will spend 10000 extra monetary units in salaries, how should they spend that amount of money in order that no worker will earn less salary than now and the distribution of salaries is as egalitarian as possible. What will the new Gini index be? They will spend those 10000 extra m.u. in the workers with the smallest salaries, so the salary of each of them will be increased by 1000 m.u.
3. Answer the previous question if instead of 10000, the firm can dispose of 35000 extra monetary units. The first 20000 m.u. will be used to increment the salary of the 10 workers that now earn 1000 m.u. until they earn 3000 m.u. each. The remaining 15000 m.u. will be used to increment the salary of each of the 15 workers that now earn 3000 m.u., until they earn 4000 m.u.

```r
library(ineq)
salary=rep(c(1000,3000,5000,10000,50000),c(10,5,2,3,1))
sum(salary)
```

```
## [1] 115000
```

```r
plot(Lc(salary))
```



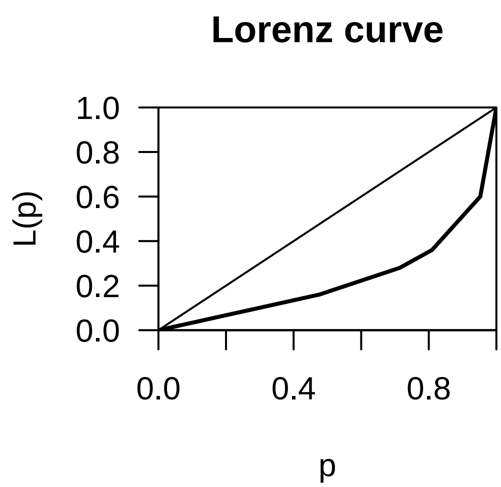

```r
ineq(salary,type="Gini")
```

```
## [1] 0.6376812
```

```r
salary=rep(c(2000,3000,5000,10000,50000),c(10,5,2,3,1))
sum(salary)
```

```
## [1] 125000
```

```r
plot(Lc(salary))
```



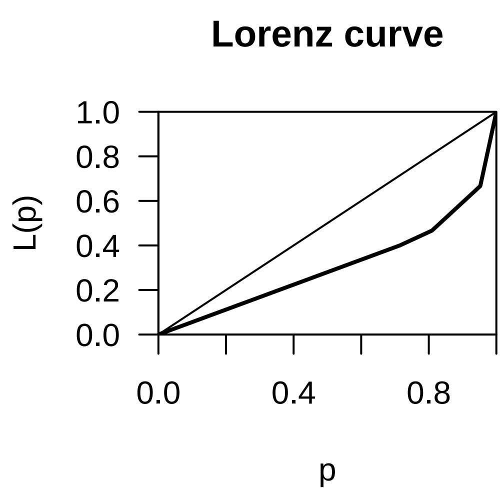

```r
ineq(salary,type="Gini")
```

```
## [1] 0.5447619
```

```r
salary=rep(c(4000,5000,10000,50000),c(15,2,3,1))
sum(salary)
```

```
## [1] 150000
```

```r
plot(Lc(salary))
```




```r
ineq(salary,type="Gini")
```

```
## [1] 0.3904762
```

3. When events occur in a time period $(0,a)$ with regard to a Poisson process, it is well known that, conditioned on the total number of events k, the joint distribution of the times at with the events occur follows a uniform distribution in $(0,a)^k$. That is, if X_i represents the arrival time of one of them, then $X_i \sim \mathrm{U}(0,a)$ and it is independent of the other X's.
    1. If ten patients visit the Emergency Room (ER) of a hospital between 9 and 10, what is the probability that fifth of them reaches the ER before 9:30? Denote by X the number of patients that reach the ER before 9.30. Since the arrival time (in minutes after 9:00) of each of the 10 patients is uniformly distributed in the interval $(0,60)$, and they are independent, then $X \sim \mathrm{B}(n=10,p=0.5)$, and we have to compute $P(X \geq 5) = 0.6230469$.
    2. If ten patients visit the ER between 9 and 10, what is the probability that the first 3 of them arrive before $9:20$, 4 of them between $9:20$ and $9:40$, and the last 3 of them after $9:40$? The distribution of the random vector (X_1,X_2,X_3) where X_1 is the number of patients that reach the hospital between 9 and 9:20, X_2 is the number of patients that reach the hospital between 9:20 and 9:40, and X_3 is the number of patients that reach the hospital between 9:40 and 10 is Multinomial with parameters $\mathrm{M}(n=10,p=(1/3,1/3,1/3))$, and $P(X_1=3,X_2=4,X_3=3)=0.0711274$.
    3. If two patients visit the ER of a hospital between 9 and 10, what is the conditional distribution of the time at which the second patient arrived at the ER given that the first patient arrived at time t_1? The joint density mass function of $X_{1:2}, X_{2:2}$ (given in minutes after 9:00) is

    $$f_{X_{1:2},X_{2:2}}(t_1,t_2) = \begin{cases} 1/1800 & \text{if } 0 < t_1 < t_2 < 60 \\ 0 & \text{otherwise} \end{cases} .$$

    The marginal density mass function of $X_{1:2}$ is

    $$f_{X_{1:2}}(t_1) = 2[1-F_X(t_1)]f_X(t_1) = \begin{cases} (60-t_1)/1800 & \text{if } 0 < t_1 < 60 \\ 0 & \text{otherwise} \end{cases} .$$

    Finally the conditional density mass function of $X_{2:2}$ given $X_{1:2}=t_1$ when $0 < t_1 < 60$ is

    $$f_{X_{2:2}|X_{1:2}}(t_2|t_1) = \frac{f_{X_{1:2},X_{2:2}}(t_1,t_2)}{f_{X_{1:2}}(t_1)} = \begin{cases} 1/(60-t_1) & \text{if } t_1 < t_2 < 60 \\ 0 & \text{otherwise} \end{cases} .$$

    In conclusion $X_{2:2}|(X_{1:2}=t_1) \sim \mathrm{U}(t_1,60)$.

```r
1-pbinom(4,size=10,prob=0.5)
```

```
## [1] 0.6230469
```

```r
dmultinom(c(3,4,3),size=10,prob=c(1/3,1/3,1/3))
```

```
## [1] 0.07112737
```