Lisbon School of Economics & Management
Universidade de Lisboa

STATISTICS I

Economics / Finance/ Management

2nd Year/2nd Semester

2024/2025

LESSON 11

Professor: Elisabete Fernandes
E-mail: efernandes@iseg.ulisboa.pt

https://doity.com.br/estatistica-aplicada-a-nutricao

https://basiccode.com.br/produto/informatica-basica/

Roadmap:

- Probability
- Random variable and two dimensional random variables:
  - Distribution
  - Joint distribution
  - Marginal distribution
  - Conditional distribution functions
- Expectations and parameters for a random variable and two dimensional random variables
- Discrete Distributions
- Continuous Distributions

Bibliography: Miller & Miller, John E. (2014) Freund's Mathematical Statistics with applications, 8th Edition, Pearson Education, [MM]

1

Continuous Distributions: Exercises

Exponential Distribution, Gamma Distribution and Chi-Square Distribution

Chapter 6

8. (Lack of memory of the exponential random variable) Let $X \sim Exp(\lambda)$, prove that $P(X > x + s|X > x) = P(X > s)$ for any $x \geq 0$ and $s \geq 0$.

Gamma distribution: The *gamma cumulative distribution* function is defined for $x > 0$, $a > 0$, $b > 0$, by the integral

$$F_X(x) = \frac{1}{b^a \Gamma(a)} \int_0^x u^{a-1} e^{-\frac{u}{b}} du$$

2025

where $\Gamma(t) = \int_0^\infty e^{-u} u^{t-1} du$ is the Gamma function. The parameters a and b are called the shape parameter and scale parameter, respectively.

The probability density function for the gamma distribution is

$$f_X(x) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-\frac{x}{b}}$$

4. $\boxed{\text{Let } X_i \sim Exp(\lambda_i),\ i = 1, 2, ..., k, \text{ be independent random variables, then } Y = \min\{X_1, X_2, ..., X_k\} \sim Exp(\textstyle\sum_{i=1}^k \lambda_i)}$

Exercise 8

$$X \sim \mathcal{E}x(\lambda), \quad \lambda > 0$$

$$f_X(x|\lambda) = \lambda e^{-\lambda x} \quad (x > 0)$$

- $X > s$
- $X > x + s$
- $X > s \wedge X > x + s$

$$F_X(x|\lambda) = \int_0^x f_X(u|\lambda)\,du = \int_0^x \lambda e^{-\lambda u}\,du = -\left[e^{-\lambda u}\right]_{u=0}^{u=x} = -\left(e^{-\lambda x} - e^0\right) =$$

$$= 1 - e^{-\lambda x} \quad (x > 0)$$

$$P(X > x + s \mid X > s) = \frac{P(X > x + s)}{P(X > s)} = \frac{1 - F_X(x+s)}{1 - F_X(s)}$$

$$= \frac{1 - (1 - e^{-\lambda(s+x)})}{1 - (1 - e^{-s\lambda})} = \frac{e^{-\lambda(s+x)}}{e^{-s\lambda}} = \frac{e^{-\lambda s} \cdot e^{-\lambda x}}{e^{-s\lambda}} = e^{-\lambda x} =$$

$$= 1 - (\underbrace{1 - e^{-\lambda x}}_{F_X(x)}) = 1 - F_X(x) = 1 - P(X \leq x) = P(X > x), \quad \text{QED}$$

9. Let $X_i \sim Exp(\lambda_i)$, $i = 1, 2$, be independent random variables. Prove that $Y = \min\{X_1, X_2\} \sim Exp(\lambda_1 + \lambda_2)$.

(**Hint:** Note that $P(Y > y) = P(X_1 > y, X_2 > y)$).

Exercise 9

$X \sim \mathcal{E}x(\lambda_i)$, $\lambda_i > 0$ $(i = 1, 2)$ $\qquad X_1 \perp X_2$

$$f_{X_i}(x_i | \lambda_i) = \lambda_i e^{-\lambda_i x_i}, \quad x_i > 0 \quad (i = 1, 2)$$

$$F_{X_i}(x_i | \lambda_i) = 1 - e^{-\lambda_i x_i} \quad (x_i > 0)$$

$$Y = \min(X_1, X_2)$$

$$P(Y > y) = P(X_1 > y, X_2 > y) = P(X_1 > y)\, P(X_2 > y) =$$

$$= (1 - F_{X_1}(y))(1 - F_{X_2}(y)) =$$

$$= (1 - (1 - e^{-\lambda_1 y}))(1 - (1 - e^{-\lambda_2 y})) = e^{-\lambda_1 y} e^{-\lambda_2 y} =$$

$$= e^{-(\lambda_1 + \lambda_2) y}$$

Exercise 9

$$F_Y(y) = P(Y \leqslant y) = 1 - P(Y > y) = 1 - e^{-(\lambda_1 + \lambda_2)y}$$

This is the C.D.F. of a $\mathcal{E}x(\lambda_1 + \lambda_2)$ random variable.

Therefore $Y \sim \mathcal{E}x(\lambda_1 + \lambda_2)$, Q.E.D.

10. Let $X_i \sim Exp(\lambda)$, $i = 1, 2$, be independent random variables.

(a) Find the distribution of $Y_1 = \min\{X_1, X_2\}$ and $Y_2 = \max\{X_1, X_2\}$.

(b) Find the expected value of $Z = \lambda^2 Y_1 - \frac{2}{3} Y_2$.

Exercise 10 a)

$X_i \sim Ex(\lambda)$; $i = 1,2$ $\quad X_1 \perp X_2$

$f_{X_i}(x \mid \lambda) = \lambda e^{-\lambda x} \quad (x > 0)$; $\quad i = 1,2$

$F_{X_i}(x \mid \lambda) = 1 - e^{-\lambda x} \quad (x > 0)$; $\quad i = 1,2$

a)

$Y_1 = \min\{X_1, X_2\}$

From ex 9 we know that $Y_1 \sim ex(2\lambda)$

$Y_2 = \max\{X_1, X_2\}$

$$F_{Y_2}(y) = P(\overset{\max\{X_1, X_2\}}{Y_2} \leq y) = P(X_1 \leq y, X_2 \leq y) =$$

$$= P(X_1 \leq y)\, P(X_2 \leq y) = F_{X_1}(y)\, F_{X_2}(y) =$$

$$= (1 - e^{-\lambda y})(1 - e^{-\lambda y}) = \underbrace{(1 - e^{-\lambda y})^2}_{[F_{X_i}(x)]^2} =$$

$$= 1 + e^{-2\lambda y} - 2e^{-\lambda y} \quad (y > 0)$$

$$f_{Y_2} = \begin{cases} F'_{Y_2}(y) = 2\lambda e^{-\lambda y} - 2\lambda e^{-2\lambda y} = 2\lambda(e^{-\lambda y} - e^{-2\lambda y}) & (y \geq 0) \\ 0 & (\text{elsewhere}) \end{cases}$$

Exercise 10 b)

$Y_1 \sim \mathcal{E}x(2\lambda) \Rightarrow E(Y_1) = \frac{1}{2\lambda}$

$E(Y_2) = \int_{-\infty}^{+\infty} y \, f_{Y_2}(y)\,dy = \int_0^{+\infty} 2\lambda y\,(e^{-\lambda y} - e^{-2\lambda y})\,dy =$

$= \int_0^{+\infty} 2\lambda y e^{-\lambda y} dy - \int_0^{+\infty} 2\lambda y e^{-2\lambda y} dy =$

$= 2 \underbrace{\int_0^{+\infty} y \lambda e^{-\lambda y} dy}_{(*)} - \underbrace{\int_0^{+\infty} y (2\lambda) e^{-(2\lambda) y} dy}_{(*)} =$

$2 \times \frac{1}{\lambda} - \frac{1}{2\lambda} = \frac{4}{2\lambda} - \frac{1}{2\lambda} = \frac{3}{2\lambda}$ (*)

(*) If $X \sim \mathcal{E}x(\lambda)$ then $E(X) = \int_0^{+\infty} x\, f_X(x)\,dx = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$

Therefore, $\int_0^{+\infty} y e^{-\lambda y} dy = \frac{1}{\lambda}$

(*) If $X \sim \mathcal{E}x(2\lambda)$ then $E(X) = \int_0^{+\infty} x\, f_X(x)\,dx = \int_0^{+\infty} x (2\lambda) e^{-(2\lambda) x} dx = \frac{1}{2\lambda}$

Therefore $\int_0^{+\infty} y(2\lambda) e^{-(2\lambda) y} dy = \frac{1}{2\lambda}$

$Z = \lambda^2 Y_1 - \frac{2}{3} Y_2$

$E(Z) = E(\lambda^2 Y_1 - \frac{2}{3} Y_2) = \lambda^2 E(Y_1) - \frac{2}{3} E(Y_2) = \frac{\lambda^2}{2\lambda} - \frac{2}{3} \times \frac{3}{2\lambda} =$

$= \frac{\lambda}{2} - \frac{1}{\lambda}$

Exercise 10 b)

(*) Alternatively, we could have used integration by parts:

$$\boxed{\int (u v') = uv - \int (v u')}$$

$$\int y e^{-2\lambda y} = -\frac{y e^{-2\lambda y}}{2\lambda} - \int -\frac{e^{-2\lambda y}}{2\lambda}\, dy = -\frac{y e^{-2\lambda y}}{2\lambda} + \frac{1}{2\lambda}\int e^{-2\lambda y} =$$

$$= -\frac{y e^{-2\lambda y}}{2\lambda} + \frac{1}{2\lambda}\left(-\frac{e^{-2\lambda y}}{2\lambda}\right) = -\frac{y e^{-2\lambda y}}{2\lambda} - \frac{e^{-2\lambda y}}{4\lambda^2}$$

$$\boxed{\begin{array}{ll} u = y & u' = 1 \\ v' = e^{-2\lambda y} & v = -\frac{e^{-2\lambda y}}{2\lambda} \end{array}}$$

$$uv = -y\frac{e^{-2\lambda y}}{2\lambda}$$

$$vu' = v = -\frac{e^{-2\lambda y}}{2\lambda}$$

$$\int y e^{-\lambda y} = -\frac{y e^{-\lambda y}}{\lambda} - \int -\frac{e^{-\lambda y}}{\lambda}\, dy = -\frac{y e^{-\lambda y}}{\lambda} + \frac{1}{\lambda}\int e^{-\lambda y} =$$

$$= -\frac{y e^{-\lambda y}}{\lambda} + \frac{1}{\lambda}\left(-\frac{e^{-\lambda y}}{\lambda}\right) = -\frac{y e^{-\lambda y}}{\lambda} - \frac{e^{-\lambda y}}{\lambda^2}$$

$$\boxed{\begin{array}{ll} u = y & u' = 1 \\ v' = e^{-\lambda y} & v = -\frac{e^{-\lambda y}}{\lambda} \end{array}}$$

$$uv = -y\frac{e^{-\lambda y}}{\lambda}$$

$$vu' = v = -\frac{e^{-\lambda y}}{\lambda}$$

Exercise 10 b)

Auxiliary calculations:

$$\lim_{a \to +\infty} a e^{-a\lambda} = \lim_{a \to +\infty} \frac{a}{e^{a\lambda}} = 0$$

$$\lim_{a \to +\infty} e^{-a\lambda} = \lim_{a \to +\infty} \frac{1}{e^{a\lambda}} = 0$$

$$\lim_{b \to +\infty} b e^{-2\lambda b} = \lim_{b \to +\infty} \frac{b}{e^{2\lambda b}} = 0$$

$$\lim_{b \to +\infty} e^{-2\lambda b} = \lim_{b \to +\infty} \frac{1}{e^{2\lambda b}} = 0$$

$$E(Y_2) = \int_{-\infty}^{+\infty} y \, f_{Y_2}(y)\, dy = \int_0^{+\infty} 2\lambda y \left(e^{-\lambda y} - e^{-2\lambda y}\right) dy =$$

$$= \int_0^{+\infty} 2\lambda y e^{-\lambda y} dy - \int_0^{+\infty} 2\lambda y e^{-2\lambda y} dy =$$

$$= 2\lambda \int_0^{+\infty} y e^{-\lambda y} dy - 2\lambda \int_0^{+\infty} y e^{-2\lambda y} dy =$$

$$= 2\lambda \lim_{b \to +\infty} \left[-\frac{y e^{-\lambda y}}{\lambda} - \frac{e^{-\lambda y}}{\lambda^2} \right]_0^a - 2\lambda \lim_{b \to +\infty} \left[-\frac{y e^{-2\lambda y}}{2\lambda} - \frac{e^{-2\lambda y}}{4\lambda^2} \right]_0^b =$$

$$= 2\lambda \lim_{a \to +\infty} \left(-\frac{a e^{-a\lambda}}{\lambda} - \frac{e^{-a\lambda}}{\lambda^2} - \left(-\frac{0}{\lambda} - \frac{e^0}{\lambda^2} \right) \right) - 2\lambda \lim_{b \to +\infty} \left(-\frac{b e^{-2\lambda b}}{2\lambda} - \frac{e^{-2\lambda b}}{4\lambda^2} - \left(-\frac{0}{2\lambda} - \frac{e^0}{4\lambda^2} \right) \right) =$$

$$= 2\lambda \left(0 - 0 + 0 + \frac{1}{\lambda^2} \right) - 2\lambda \left(0 - 0 + 0 + \frac{1}{4\lambda^2} \right) = \frac{2\lambda}{\lambda^2} - \frac{2\lambda}{4\lambda^2} = \frac{2}{\lambda} - \frac{1}{2\lambda} = \frac{4}{2\lambda} - \frac{1}{2\lambda} = \frac{3}{2\lambda}$$

11. The lifetime in years of an electronic component is a continuous random variable X that follows

$$X \sim Exp(1)$$

(a) Find the lifetime L which a typical component is 60% certain to exceed.

(b) If five components are sold to a manufacturer, find the probability that at least one of them will have a lifetime less than L years.

Exercise 11 a)

$X \equiv$ lifetime of a component (in years)

$X \sim \mathcal{E}x(1) \qquad f_X(x) = e^{-x} \; (x > 0)$

a)

$$P(X > L) = 0.6 \Leftrightarrow 1 - F_X(L) = 0.6 \Leftrightarrow$$

$$\Leftrightarrow 1 - \int_0^L e^{-x}\,dx = 0.6 \Leftrightarrow$$

$$\Leftrightarrow 1 - \left[-e^{-x}\right]_0^L = 0.6 \Leftrightarrow$$

$$\Leftrightarrow 1 - (-e^{-L} + e^{0}) = 0.6 \Leftrightarrow$$

$$\Leftrightarrow 1 - (1 - e^{-L}) = 0.6 \Leftrightarrow$$

$$\Leftrightarrow 1 - 1 + e^{-L} = 0.6 \Leftrightarrow$$

$$\Leftrightarrow e^{-L} = 0.6 \Leftrightarrow L = -\ln(0.6) \simeq 0.51$$

Answer: Aproximately 0.51 hours.

Exercise 11 b)

$n = 5$ components

$p = P(X < L) = 1 - 0.6 = 0.4$

$Y \equiv$ # of components (in 5) with lifetime less than L.

$Y \sim Bin(5, 0.4)$ $\quad f_Y(y) = \binom{5}{y} 0.4^y \times 0.6^{5-y} \quad (y = 0,1,2,3,4,5)$

$P(Y \geqslant 1) = 1 - P(Y < 1) = 1 - f_Y(0) = 1 - \binom{5}{0} 0.4^0 \times 0.6^5 \simeq 0.9222$

12. The time intervals between successive trains stopping in a certain rail station have an exponential distribution with mean 6 minutes.

(a) Find the probability that the time interval between two consecutive trains is less than 5 minutes.

(b) Find a time interval t such that we can be 95% sure that the time interval between two successive trains will be greater than t.

(c) Assume that the number of trains arriving in one hour is modeled by a Poisson random variable. Compute the probability that in a random hour 5 trains stop at the train station.

(d) If we have counted 8 trains in the first hour, what is the probability that two of them arrived in the first 30 minutes?

Exercise 12 a)

$X \equiv$ Time (in minutes) between sucessive trains

$X \sim \mathcal{E}x(1/6) \quad E(X) = 6 \qquad f_X(x) = \frac{1}{6} e^{-\frac{x}{6}} \quad (x > 0)$

$F_X(x) = 1 - e^{-\frac{x}{6}} \quad (x > 0)$

a)

$P(X < 5) = F_X(5) = 1 - e^{-\frac{5}{6}} \simeq 0.5654$

Exercise 12 b)

$$P(x > t) = 0.95 \Leftrightarrow 1 - F_x(t) = 0.95 \Leftrightarrow$$

$$\Leftrightarrow 1 - \left(1 - e^{-\frac{t}{6}}\right) = 0.95 \Leftrightarrow$$

$$\Leftrightarrow e^{-\frac{t}{6}} = 0.95 \Leftrightarrow$$

$$\Leftrightarrow -\frac{t}{6} = \ln(0.95) \Leftrightarrow$$

$$\Leftrightarrow t = -6 \ln(0.95) \simeq 0.308$$

Exercise 12 c)

Time (in minutes) between sucessive trains $\sim Ex(1/6)$

trains per minute $\sim Poi\left(\frac{1}{6}\right)$

trais per hour $\sim Poi\left(\frac{1}{6} \times 60\right) = Poi(10)$

$Y \equiv$ # of trains arriving in one hour $\sim Poi(10)$

$$f_Y(y) = \frac{e^{-10}\, 10^y}{y!} \quad (y = 0, 1, 2, \dots)$$

$$P(Y = 5) = f_Y(5) = \frac{e^{-10}\, 10^5}{5!} = 0.0378$$

Exercise 12 d)

$m = 8$ trains (in 1 hour)

$W \equiv$ # of trains arriving in the first 30 min $\sim Poi\left(\frac{10}{2}\right) = Poi(5)$

$Z \equiv$ # of trains arriving in the last 30 min $\sim Poi(5)$

$$P(W=2 \mid Y=8) = \frac{P(W=2 \wedge Y=8)}{P(Y=8)} = \frac{P(W=2 \wedge Z=6)}{P(Y=8)} =$$

$$= \frac{f_W(2)\, f_Z(6)}{f_Y(8)} = \frac{\frac{e^{-5}5^2}{2!} \times \frac{e^{-5}5^6}{6!}}{\frac{e^{-10}10^8}{8!}} \simeq 0.1094$$

13. Compute the following probabilities:

(a) If Y is distributed $\chi^2(4)$ find $P(Y \leq 7.78)$.

(b) If Y is distributed $\chi^2(10)$ find $P(Y > 18.31)$.

(c) If Y is $\chi^2(1)$ find $P(Y \leq 3.8416)$.

Chi-squared Distribution

Adriana, 2025

In the case of the chi-squared random variables we have:

1. $E(X) = v$.
2. $Var(X) = 2v$.
3. Let X_1, X_2,...,X_k be independent random variables with Chi-squared distribution $X_1 \sim \chi^2(v_1)$ and $X_2 \sim \chi^2(v_2)$, ..., $X_k \sim \chi^2(v_k)$, then $\sum_{i=1}^{k} X_i \sim \chi^2\left(\sum_{i=1}^{k} v_i\right)$.

Exercise 13 a)

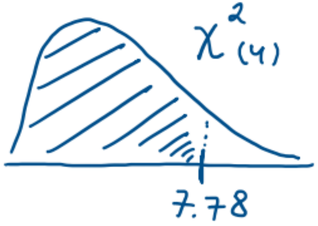
<u>Table 5:</u> $\chi^2_{\alpha,\nu}$: $P(\chi^2_{\nu} > \chi^2_{\alpha,\nu}) = \alpha$

R.V.

$\chi^2_{\alpha,\nu} \in \mathbb{R}^+$

a)

$Y \sim \chi_{(4)}$ $P(Y \leq 7.78) = 0.9$



using excel, because 7.78 is not on table V

fx =DIST.CHIQ(7,78;4;VERDADEIRO)

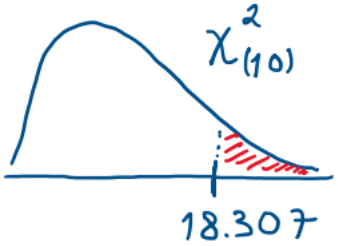
	A	B	C	D	E
1					
2		0,90002			
3					

b)

Exercise 13 b)

To use table 5

$Y \sim \chi^2_{(10)}$ $\qquad P(Y > 18.31) \simeq P(Y > \underbrace{18.307}_{\chi^2_{0.05,10}}) = 0.05$



▨ = 0.05

Exercise 13 c)

e)

To use table 5

$Y \sim \chi^2_{(1)}$

$$P(Y \leq 3.8416) \simeq P(Y \leq 3.841) = 1 - P(Y > 3.841)$$

$$= 1 - 0.05 = 0.95$$

= 0.05

= 0.95

14. Using the moment generating function, show that if $X \sim Gamma(a, b)$ and $Y = 2X/b$, then $Y \sim \chi^2(2a)$.

Gamma Distribution

Gamma distribution: The *gamma cumulative distribution* function is defined for $x > 0$, $a > 0$, $b > 0$, by the integral

$$\boxed{F_X(x) = \frac{1}{b^a \Gamma(a)} \int_0^x u^{a-1} e^{-\frac{u}{b}} \, du}$$

where $\Gamma(t) = \int_0^\infty e^{-u} u^{t-1} du$ is the Gamma function. The parameters a and b are called the shape parameter and scale parameter, respectively.

The probability density function for the gamma distribution is

$$\boxed{f_X(x) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-\frac{x}{b}}}$$

Adriana, 2025

Exercise 14

$X \sim \text{Gamma}(a, b)$ $\qquad Y = \frac{2}{b} X$

$M_X(t) = (1 - bt)^{-a} \quad (t < \frac{1}{b})$

$Q_1 \sim \chi^2(K)$

$M_{Q_1}(t) = (1 - 2t)^{-\frac{K}{2}} \quad (t < \frac{1}{2})$

Consequently:

$Q_2 \sim \chi^2(2a)$

$M_{Q_2}(t) = (1 - 2t)^{-\frac{2a}{2}} = (1 - 2t)^{-a} \quad (t < \frac{1}{2})$

To demonstrate that $Y = \frac{2}{b} X \sim \chi^2(2a)$ we need to show that $M_Y(t) = M_{Q_2}(t) = (1 - 2t)^{-a} \quad (t < \frac{1}{2})$.

Exercise 14

$$M_Y(t) = E(e^{tY}) = E(e^{\frac{2}{b}tX}) = M_X\left(\frac{2}{b}t\right) =$$

$$= \left(1 - b\frac{2}{b}t\right)^{-a} = (1-2t)^{-a} \quad \left(t < \frac{1}{2}\right)$$

Conclusion: $Y \sim \chi^2(2a)$, QED.

15. Prove that if X_1 and X_2 are independent random variables with Gamma distribution $X_1 \sim Gamma(a_1, b)$ and $X_2 \sim Gamma(a_2, b)$, then $X_1 + X_2 \sim Gamma(a_1 + a_2, b)$. (**Hint:** Recall that if $X \sim Gamma(a, b)$, then $M_X(t) = (1 - bt)^{-a}$ for $t < 1/b$).

Exercise 15

$X_1 \perp X_2$

$X_1 \sim \text{Gamma}(a_1, b)$ $\qquad M_{X_1}(t) = (1 - bt)^{-a_1} \quad (t < \frac{1}{b})$

$X_2 \sim G(a_2, b)$ $\qquad M_{X_2}(t) = (1 - bt)^{-a_2} \quad (t < \frac{1}{b})$

$$M_{X_1+X_2}(t) = E(e^{t(X_1+X_2)}) = E(e^{tX_1} \cdot e^{tX_2}) \overset{\text{because } X_1 \perp X_2}{=}$$

$$= E(e^{tX_1})E(e^{tX_2}) = M_{X_1}(t) M_{X_2}(t) =$$

$$= (1 - bt)^{-a_1}(1 - bt)^{-a_2} = \underbrace{(1 - bt)^{-(a_1+a_2)} \qquad (t < \tfrac{1}{b})}_{\text{M.G.F. of the Gamma}(a_1+a_2,\, b)}$$

Therefore $X_1 + X_2 \sim G(a_1 + a_2, b)$, QED.

16. Suppose customers arrive at a store according to a Poisson process, where the expected number of customers per hour is 0.5.

   (a) Knowing that 4 customers arrived at the store during the morning (4 hours) what is the probability that in this day (8 hours) the store receives more than 15 customers?

   (b) Compute the probability that the first customer does not arrive during the first hour (since the opening hour of the store).

   (c) What is the distribution of the time until the second customer arrives?

   (d) Find the probability that one has to wait at least half an hour until the second customer arrives.

   (e) Find the probability that one has to wait at least five hours until the fourth customer arrives.

Exercise 16 a)

$X \equiv$ # of costumers in 1 hour $\sim Poi(0.5)$

$E(X) = 0.5$

a)

$Y_M \equiv$ # of costumers in one morning (4 hours) $\sim Poi(4 \times 0.5) = Poi(2)$

$Y_A \equiv$ " " " " " afternoon (" ") $\sim Poi(2)$ $\quad f_{Y_A}(y) = \dfrac{e^{-2}\,2^y}{y!} \quad (y = 0, 1, 2, \dots$

$Y_D \equiv$ " " 8 hours $\sim Poi(8 \times 0.5) = Poi(4)$

$$P(Y_D > 15 \mid Y_M = 4) = \frac{P(Y_A > 11, Y_M = 4)}{f_{Y_M}(4)} = \frac{(1 - F_{Y_A}(11))\,\cancel{f_{Y_M}(4)}}{\cancel{f_{Y_M}}(4)} =$$

$$= 1 - F_A(11) = 1 - \sum_{y=0}^{11} \frac{e^{-2}\,2^y}{y!} \simeq 0$$

Cálculo no R:

```
> 1 - sum((exp(-2) * 2 ^ (0:11)) / factorial(0:11))
[1] 1.364615e-06
> 1 - ppois(11, 2)
[1] 1.364615e-06
>
```

Cálculo no excel

fx =1 - DIST.POISSON(11;2;VERDADEIRO)

	A	B	C	D
1	1,3646E-06			
2				

Exercise 16 b)

$T \equiv$ time until the first customer (in hours) $\sim Ex(0.5$

$F_T(t) = 1 - e^{-\frac{t}{2}} \quad (t > 0)$

$P(T > 1) = 1 - F_T(1) = 1 - (1 - e^{-\frac{1}{2}}) \simeq 0.61$

Exercise 16 c) and d)

c)

$T_2 \equiv$ time until the second customer arrives (in hours) $\sim \text{Gamma}(\underset{m}{2}, \underset{\lambda}{0.5}) = \chi^2(\underset{2m}{4})$

d)

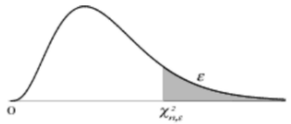
$P(T_2 \geqslant 0.5) \overset{\text{To use table V}}{\simeq} P(T_2 \geqslant 0.484) \overset{\text{Table V}}{=} 0.975$

Exercise 16 e)

$T_4 \equiv$ time until the fourth customer arrives (in hours) $\sim$ Gamma $(4, 0.5) = \chi\ (8)$

$P(T_4 \geqslant 5) \simeq P(T_4 \geqslant 5.071) = 0.75$

$\chi^2_{n.\varepsilon} : P(X > \chi^2_{n.\varepsilon}) = \varepsilon$



n \ ε	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005	.001
1	.000	.000	.001	.004	.016	.102	.455	1.323	2.706	3.841	5.024	6.635	7.879	10.827
2	.010	.020	.051	.103	.211	.575	1.386	2.773	4.605	5.991	7.378	9.210	10.597	13.815
3	.072	.115	.216	.352	.584	1.213	2.366	4.108	6.251	7.815	9.348	11.345	12.838	16.266
4	.207	.297	.484	.711	1.064	1.923	3.357	5.385	7.779	9.488	11.143	13.277	14.860	18.466
5	.412	.554	.831	1.145	1.610	2.675	4.351	6.626	9.236	11.070	12.832	15.086	16.750	20.515
6	.676	.872	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	14.449	16.812	18.548	22.457
7	.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	16.013	18.475	20.278	24.321
8	1.344	1.647	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	17.535	20.090	21.955	26.124
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	19.023	21.666	23.589	27.877
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	20.483	23.209	25.188	29.588

or :

fx =1-DIST.CHIQ(5;8;VERDADEIRO())

	A	B	C	D
1	0,75757613			
2				

$\rightarrow P(T_4 \geqslant 5)$

2

Continuous Distributions: Exercises

Normal Distribution, Uniforme Distribution and Central Limite Theorem

Chapter 6

Normal Distribution

Adriana, 2025

The most famous continuous distribution is the *normal distribution* (introduced by Abraham de Moivre, 1667-1754). The normal probability density function is given by

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.}$$

The cumulative distribution function does not have a close form solution:

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

When a random variable X follows a normal distribution with parameters μ and σ^2 we write $X \sim N(\mu, \sigma^2)$.

$$X \sim N(\mu, \sigma^2) \Rightarrow \frac{X-\mu}{\sigma} \sim N(0,1)$$

Properties:

1. Moment generating function $M_X(t) = e^{(\mu t + 0.5\sigma^2 t^2)}$
2. $E(X) = \mu$.
3. $Var(X) = \sigma^2$

17. Compute the following probabilities:

(a) If Y is distributed $N(1,4)$, find $P(Y \leq 3)$.

(b) If Y is distributed $N(3,9)$, find $P(Y > 0)$.

(c) If Y is distributed $N(50,25)$, find $P(40 \leq Y \leq 52)$.

(d) If Y is distributed $N(0,1)$, find $P(|Y| > 1.96)$.

Exercise 17 a), b), c) and d)

(a) 0.841

(b) 0.841

(c) 0.633

(d) 0.05

18. Prove that if the random variables $X_i, i = 1, 2$ have a normal distribution, $X_i \sim N(\mu_i, \sigma_i^2)$, and are independent and if $Y = aX_1 + bX_2 + c$, then $Y \sim N(\mu_Y, \sigma_Y^2)$, where $\mu_Y = a\mu_1 + b\mu_2 + c$ and $\sigma_Y^2 = a^2\sigma_1^2 + b^2\sigma_2^2$.

(**Hint:** Recall that if $X \sim N(\mu, \sigma^2)$, then $M_X(t) = e^{(\mu t + 0.5\sigma^2 t^2)}$ and note that functions of independent random variables are also independent).

Exercise 18

$X_i \sim N(\mu_i, \sigma_i^2)$, $\quad M_{X_i}(t) = E(e^{tX_i}) = e^{\mu_i t + 0.5\sigma_i^2 t^2} \quad (i = 1, 2)$

$X_i \perp X_j \quad (i \neq j)$

$Y = aX_1 + bX_2 + c$

because of independence $\quad = E(e^{ct}) = e^{ct}$, because $c\,t \in \mathbb{R}$

$$M_Y(t) = M_{aX_1 + bX_2 + c}(t) = M_{aX_1}(t)\, M_{bX_2}(t)\, M_c(t) =$$

$$= M_{X_1}(at)\, M_{X_2}(bt)\, e^{ct} = e^{\mu_1 a t + 0.5\sigma_1^2 (at)^2}\, e^{\mu_2 b t + 0.5\sigma_2^2 (bt)^2}\, e^{ct}$$

$$= e^{(\mu_1 a + \mu_2 b + c)t + 0.5(a^2\sigma_1^2 + b^2\sigma_2^2)t^2} = e^{\mu_Y t + 0.5\sigma_Y^2 t^2}$$

This is the M.G.F. of a $N(\mu_Y, \sigma_Y^2)$, where $\mu_Y = a\mu_1 + b\mu_2 + c$ and $\sigma_Y^2 = a^2\sigma_1^2 + b^2\sigma_2^2$. It is therefore demonstrated that $Y \sim N(\mu_Y, \sigma_Y^2) = N(a\mu_1 + b\mu_2 + c,\ a^2\sigma_1^2 + b^2\sigma_2^2)$.

Exercise 18

Note:

functions of independent variables are also independent

$$M \quad t) = E\left(e^{(aX_1 + bX_2 + c)t}\right) = E\left(e^{aX_1 t + bX_2 t + ct}\right) =$$

$$= E\left(e^{aX_1 t} e^{bX_2 t} e^{ct}\right) = E\left(e^{aX_1 t}\right) E\left(e^{bX_2 t}\right) E\left(e^{ct}\right) =$$

$$= M_{X_1}(at)\, M_{X_2}(bt)\, e^{ct} = \cdots$$

19. Suppose that diameter of a certain component produced in a factory can be modeled by a normal distribution with mean $10cm$ and standard deviation $3cm$.

   (a) Find the probability that the diameter of a random component is larger than $13cm$.

   (b) Find the probability that the diameter of a random component is less than $7cm$.

   (c) Selecting randomly 10 components, what is the probability that 2 of them have a diameter less than $7cm$?

   (d) What is the expected number of components that we have to inspect to find 1 with a diameter less than $7cm$?

Exercise 19 a) and b)

(a) Let X be the diameter of a certain component produced in a factory.

$$X \sim N(\mu = 10, \sigma^2 = 9).$$

Then,

$$P(X > 13) = P\left(\frac{X-10}{3} > \frac{13-10}{3}\right) = P(Z > 1) = 1 - \Phi(1) = 0.159$$

where $Z \sim N(0,1)$.

(b)

$$P(X < 7) = P\left(\frac{X-10}{3} < \frac{7-10}{3}\right) = P(Z < -1) = 1 - \Phi(1) = 0.159$$

the last equality following in light of the symmetry of the normal distribution.

Exercise 19 c) and d)

(c) Let Y be the random variable that represents the number of components that have a diameter less than $7cm$, in a set of 10 components.

$$Y \sim Bin(n = 10, p = 0.159)$$

because $p = P(X < 7)$. The requested probability is

$$P(Y = 2) = \binom{10}{2} 0.159^2 (1 - 0.159)^8 = 0.285.$$

(d) Let Z be the random variable that we have to inspect to find 1 with a diameter less than $7cm$.

$$Z \sim Geo(p), \quad \text{where } p = P(X < 7) = 0.159$$

Then,

$$E(Z) = 1/p \approx 6.3,$$

which means that, in average, one has to inspect 7 components.

20. A baker knows that the daily demand for a specific type of bread is a random variable X such that $X \sim N(\mu = 50, \sigma^2 = 25)$. Find the demand which has probability 1% of being exceeded.

Exercise 20

61.65

21. Assume that X_i, with $i = 1, 2, 3$ represent the profit, in million of euros, of 3 different companies located in 3 different countries. If

$$X_1 \sim N(1, 0.01), \quad X_2 \sim N(1.5, 0.03), \quad X_3 \sim N(2, 0.06)$$

(a) Which company is more likely to have a profit greater than 1.5 millions?

(b) What is the probability of the profit of these 3 companies does not exceed 4 millions of euros? (Assume independence.)

Exercise 21 a)

(a) Due to the symmetry of the normal distributions, we know that

$$P(X_1 > 1.5) < P(X > 1) = 0.5, \quad P(X_2 > 1.5) = 0.5$$
$$P(X_3 > 1.5) > P(X_3 > 2) = 0.5.$$

Therefore, company 3 is more likely to exceed a profit of 1.5 million.

Exercise 21 b)

(b) From the properties of independent normal random variables, we know that

$$X_1 + X_2 + X_3 \sim N(\mu = 1 + 1.5 + 2, \sigma^2 = 0.01 + 0.02 + 0.06).$$

Therefore, if $Z \sim N(0, 1)$, then

$$\begin{aligned} P(X_1 + X_2 + X_3 < 4) &= P\left(\frac{X_1 + X_2 + X_3 - 4.5}{0.3} < \frac{4 - 4.5}{0.3}\right) \\ &= P(Z < -5/3) \approx 0.05. \end{aligned}$$

22. The time elapsed since failure until repair (designated as repair time) of a certain type of machines is a random variable with exponential distribution with mean of 2 hours.

(a) What is the probability that a broken machine has a repair time of 1 hour or less?

(b) If 10 broken machines were randomly selected, what is the probability of the fastest repair be performed in less than 15 minutes? (assume independence)

(c) What is the probability that the total repair time of 50 broken machines does not exceed 90 hours? (assume independence)

Exercise 22 a)

Let X be the random variable that represents the time (in hours) elapsed since failure until repair.

$$X \sim Exp(\lambda), \quad \lambda = 1/E(X) = 0.5.$$

Then,

$$P(X \leq 1) = F_X(1) = 1 - e^{-1 \times 0.5} = 0.393.$$

Exercise 22 b)

Let X_i, with $i = 1 \cdots, 10$, be the random variable that represents the time (in hours) elapsed since failure until repair of the i^{th} machine.

$$X_i \sim Exp(\lambda), \quad \lambda = 1/E(X) = 0.5.$$

and the the random variables $X_1, \cdots, X_{10}$ are independent. From what we already saw $\min\{X_1, \cdots, X_{10}\} \sim Exp(5)$. Then,

$$P(\min\{X_1, \cdots, X_{10}\} < 1/4) = 1 - e^{-5/4} \approx 0.713$$

The Central Limit Theorem

Adriana, 2025

Theorem: (*The Central Limit Theorem* - Lindberg-Levy)

Assume that X_i, $i = 1, \ldots, n$ are independent, $E(X_i) = \mu_X$, and $Var(X_i) = \sigma_X < +\infty$, then the distribution of

$$\boxed{Z = \frac{\sum_{i=1}^{n} X_i - n\mu_X}{\sigma_X \sqrt{n}} = \frac{\sqrt{n}\left(\overline{X} - \mu_X\right)}{\sigma_X}}$$

converges to a standard normal distribution as n tends to infinity. We write $Z \overset{a}{\sim} N(0, 1)$ where the symbol $\overset{a}{\sim}$ reads "distributed asymptotically"

Normal Distribution: Results

If $X_1, X_2, \cdots X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$ then the following holds true:

i)

$$\boxed{\sum_{i=1}^{n} X_i \sim N(\mu n, \sigma^2 n) \quad \text{or equivalently} \quad \overline{X} \sim N(\mu, \sigma^2/n)}$$

ii)

$$\boxed{\frac{\sum_{i=1}^{n} X_i - \mu n}{\sigma\sqrt{n}} \sim N(0,1) \quad \text{or equivalently} \quad \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)}$$

Adriana, 2025

Exercise 22 c)

Let X_i, with $i = 1 \cdots, 50$, be the random variable that represents the time (in hours) elapsed since failure until repair of the i^{th} machine.

$$X_i \sim Exp(\lambda), \quad \lambda = 1/E(X) = 0.5.$$

and the the random variables $X_1, \cdots, X_{50}$ are independent. Then the total repair time of 50 broken machines is given by

$$T_{50} = \sum_{i=1}^{50} X_i.$$

Due to the Limit Central Theorem, we get that

$$T_{50} = \sum_{i=1}^{50} X_i \overset{a}{\sim} N(\mu, \sigma^2)$$

Exercise 22 c)

where

$$\mu = \sum_{i=1}^{50} E(X_i) = \sum_{i=1}^{50} \frac{1}{\lambda} = 100$$

and

$$\sigma^2 = \sum_{i=1}^{50} Var(X_i) = \sum_{i=1}^{50} \frac{1}{\lambda^2} = 200.$$

Then,

$$P(T_{50} < 90) = P\left(\frac{T_{50} - 100}{10\sqrt{2}} < \frac{90 - 100}{10\sqrt{2}}\right) \approx P(Z < -0.71) = 0.239.$$

23. Suppose that you roll a balanced die 36 times. Let Y denote the sum of the outcomes in each of the 36 rolls. Estimate the probability that $108 \leq Y \leq 144$.

Exercise 23

$$0.858$$

24. Suppose that a book with 300 pages contains on average 1 misprint per page. Assume that the number of misprints per page is a Poisson random variable.

(a) What is the probability that a random page has 2 or more misprints?

(b) What is the probability that there will be at least 100 pages which contain 2 or more misprints? (assume independence)

(c) What is the probability that there will be no more than 200 misprints in the book?

Exercise 24 a), b) and c)

(a) 0.264

(b) 0.0033

(c) ≈ 0

Uniforme Distribution

The probability density function of the *uniform random variable* on an interval (a, b), where $a < b$, is the function

$$\boxed{f_X(x) = \begin{cases} 0 & \text{if} \quad x \leq a \\ \frac{1}{b-a} & \text{if} \quad a < x < b \\ 0 & \text{if} \quad b \leq x \end{cases}}$$

Adriana, 2025

The cumulative distribution function is the function

$$\boxed{F_X(x) = \begin{cases} 0 & \text{if} \quad x \leq a \\ \frac{x-a}{b-a} & \text{if} \quad a \leq x \leq b \\ 1 & \text{if} \quad b \leq x \end{cases}}$$

3. $E(X) = (a+b)/2.$
4. $Var(X) = (b-a)^2/12.$
5. $Skewness = \gamma_1 = 0.$

Remark: If X is a *uniform random variable* in the interval (a, b) we write $X \sim U(a, b)$.

25. Assume that the number of hours per week that a student spends studying for the course of Statistics 1 follows a continuous uniform distribution in the interval $(0, 5)$.

    (a) What is the probability that a random student spends more than 3 hours studying for the course of Statistics 1?

    (b) In a group of 300 students, what is the probability that more than 100 spend more than 3 hours studying for the course of Statistics 1?

    (c) In a group of 300 students, what is the probability that, on average, students spend more than 4 hours studying for the course of Statistics 1?

Exercise 25 a)

(a) Let X be the random variable that represents the number of hours that students spend studying for the course of Statistics 1.

$$X \sim U(0,5).$$

Then,

$$P(X > 3) = \int_3^5 \frac{3}{5} dx = \frac{2}{5}.$$

Exercise 25 b)

Let Y be the random variable that represents the number of students, in 300, that spend more than 3 hours studying for the course of Statistics 1.

$$Y \sim Bin(n = 300, p), \quad \text{with } p = P(X \geq 3) = \frac{2}{5}.$$

As the number of trials is large enough, the central limit theorem allows us to say that

$$Y \overset{a}{\sim} N(\mu, \sigma^2)$$

where,

$$\mu = n \times p = 120 \quad \text{and} \quad \sigma^{=} n \times p \times (1 - p) = 72.$$

Therefore,

$$P(Y > 100) = P\left(\frac{Y - 120}{\sqrt{72}} > \frac{100 - 120}{\sqrt{72}}\right) \approx P(Z > -2.36) = 0.99$$

Exercise 25 c)

Let X_i, with $i = 1, \cdots, 300$, be the random variable that represents the number of hours that student i spends studying for the course of Statistics 1. Then

$$X_i \sim U(0,5), \quad \text{for } i = 1, \cdots, 300$$

and X_i, with $i = 1, \cdots, 300$ are independent random variables. Therefore, the average number of hours spent by students studying for the course of statistics one is modeled by

$$\overline{X} = \frac{1}{300}\sum_{i=1}^{300} X_i.$$

From the properties of expected value and variance, we get

$$E(\overline{X}) = \frac{1}{300}\sum_{i=1}^{300} E(X_i) = E(X_i) = 2,5$$

and

$$Var(\overline{X}) = \underbrace{\left(\frac{1}{300}\right)^2 \sum_{i=1}^{300} Var(X_i)}_{\text{due to independence}} = \frac{1}{300} Var(X_i) = \frac{1}{300} \times \frac{25}{12}.$$

Therefore, from the central limit theorem, we get that

$$Z = \frac{\overline{X} - 2,5}{\sqrt{\frac{1}{300} \times \frac{25}{12}}} \overset{a}{\sim} N(0,1).$$

Exercise 25 c)

The intended probability follows

$$P(\overline{X} > 4) \approx P\left(Z > \frac{4 - 2,5}{\sqrt{\frac{1}{300} \times \frac{25}{12}}}\right) = P(Z > 18) \approx 0.$$

Thanks!

Questions?