

# STAT4DS / HOMEWORK 3 - PART A

*Maria Luisa Croci, Valerio Antonini*

---

## Index

1. Exercise I: Ultra-fast exercise
  1. DAG 1
  2. DAG 2
2. Exercise II: Sampling a DAG
  1. Joint distribution
  2. Simulation on R
  3. Plot each componenet of  $X$

## Exercise I: Ultra-fast exercise

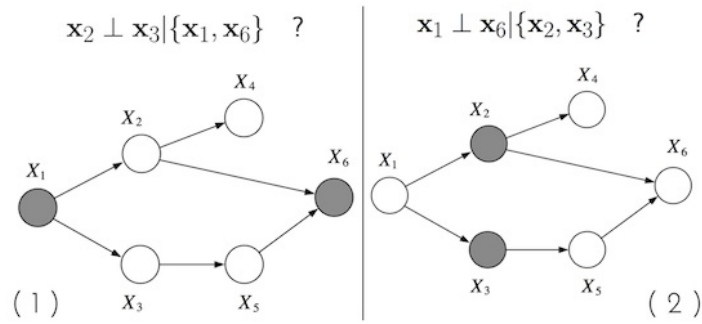
In the following exercise we will work on **Graphical Models**, that are probabilistic models for which a graph expresses the conditional independence structure between random variables.

The conditional independence helps to represent large joint distributions. In general independence is an useful property, but rarely we have two random variables that are independent. Infact it's more common that two r.v. are independent given an additional one. In formula:

$$(X \perp Y \mid Z) \Leftrightarrow \begin{cases} P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z) \\ P(X \mid Y, Z) = P(X \mid Z) \\ P(Y \mid X, Z) = P(Y \mid Z) \end{cases} \quad \text{since for symmetry } (X \perp Y \mid Z) = (Y \perp X \mid Z)$$

**So, given the two following DAGs, how can we check if the conditional independence relationships below are satisfied?**

In order to verify these relationships we use *Markov Condition* and *d-separation*.



In general a DAG (directed acyclical graph) is used to represent hierarchical models, where nodes are ordered such that parents come always before children (following a topological order).

## Markov Condition

Given a graph  $G$  with  $V = \{X_1, \dots, X_d\}$  the set of vertex, Markov property define that each node depends only on its immediate parents. In formula:

$$(X_d \perp X_{pred(d)pa(d)} | X_{pa(d)})$$

where  $pred(d)$  is the predecessors of  $d$  and  $pa(d)$  the parents of  $d$ .

## d-separation

Given a DAG, we say that if two nodes  $X$  and  $Y$  are d-separated by  $Z$ , the conditional independence relationships  $(X \perp Y | Z)$  is satisfied, otherwise  $X$  and  $Y$  are d-connected.

But how to check the d-separation?

1. Consider a DAG in which  $X_A, X_B, X_C$  are three disjoint sets of nodes.
2. Consider a set  $U$  that includes all the possible undirected paths from any node in  $X_A$  to any node in  $X_B$ .
3. We define a path as *blocked* if and only if at least one of the following conditions is satisfied:
  - $U$  contains a **chain** that is  $\rightarrow \bullet \rightarrow$  with  $\bullet \in X_C$
  - $U$  contains a **common cause** that is  $\leftarrow \bullet \rightarrow$  with  $\bullet \in X_C$
  - $U$  contains a **collider** that is  $\rightarrow \circ \leftarrow$  with  $\circ \notin X_C$  and nor is any of its decedents
4. If all paths in  $U$  are blocked, then  $X_A$  is d-separated from  $X_B$  by  $X_C$  and the joint distribution over all the variables in the graph is satisfied. Briefly  $X_A \perp X_B | X_C$

Now we can answer to our questions:

### 1.1 DAG 1

Is  $(X_2 \perp X_3 | \{X_1, X_6\})$  true?

1.  $X_A = \{X_2\}, X_B = \{X_3\}, X_C = \{X_1, X_6\}$ .
2.  $U = \{X_2 \rightarrow X_6 \leftarrow X_5 \leftarrow X_3, X_2 \leftarrow X_1 \rightarrow X_3\}$  is the set of undirected paths from  $X_2$  to  $X_3$ .
3. The first path  $\{X_2 \rightarrow X_6 \leftarrow X_5 \leftarrow X_3\}$  is not blocked (infact it doesn't have neither chain and common cause, moreover the collider case is not satisfied since  $X_6 \in X_C$ ). The second path  $X_2 \leftarrow X_1 \rightarrow X_3$  is blocked because it has a common cause with  $X_1 \in X_C$ .
4. Since only one path is blocked,  $X_2$  is d-connected (not d-separated) from  $X_3$  by  $\{X_1, X_6\}$  and the relationship  $X_2 \perp X_3 | \{X_1, X_6\}$  is not satisfied.

### 1.2 DAG 2

Is  $(X_1 \perp X_6 | \{X_2, X_3\})$  true?

1.  $X_A = \{X_1\}, X_B = \{X_6\}, X_C = \{X_2, X_3\}$ .
2.  $U = \{X_1 \rightarrow X_3 \rightarrow X_5 \rightarrow X_6, X_1 \rightarrow X_2 \rightarrow X_6\}$  is the set of undirected paths from  $X_1$  to  $X_6$ .

3. The two paths are blocked because they both have a chain with  $X_3 \in X_C$  and  $X_2 \in X_C$ .
4. Since both paths are blocked,  $X_1$  is d-separated from  $X_6$  by  $\{X_2, X_3\}$  and the relationship  $X_1 \perp X_6 | \{X_2, X_3\}$  is satisfied.

Moreover we can prove it directly checking if  $p(x_6 | x_1, x_2, x_3) = p(x_6 | x_2, x_3)$ .

From Markov condition, we define the joint probability of our graph as:

$$\begin{aligned} p(x_1, x_2, \dots, x_6) &= p(x_1)p(x_2 | x_1)p(x_3 | x_1, x_2)p(x_4 | x_1, x_2, x_3)p(x_5 | x_1, x_2, x_3, x_4)p(x_6 | x_1, x_2, x_3, x_4, x_5) \\ &= p(x_1)p(x_2 | x_1)p(x_3 | x_1)p(x_4 | x_2)p(x_5 | x_3)p(x_6 | x_2, x_5) \end{aligned}$$

Using the chain case we have that the graph induced factorization as:

$$p(x_6 | x_1, x_3, x_3) = \frac{p(x_1, x_2, x_3, x_6)}{p(x_1, x_2, x_3)} = \frac{p(x_1)p(x_2 | x_1)p(x_3 | x_1)p(x_6 | x_2, x_5)}{p(x_1)p(x_2 | x_1)p(x_3 | x_1)} = p(x_6 | x_2, x_5)$$

and

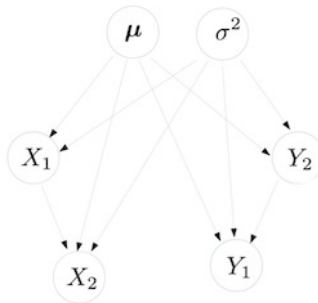
$$p(x_6 | x_2, x_3) = \frac{p(x_2, x_3, x_6)}{p(x_2, x_3)} = \frac{p(x_2 | x_1)p(x_3 | x_1)p(x_6 | x_2, x_5)}{(x_2 | x_1)p(x_3 | x_1)} = p(x_6 | x_2, x_5)$$

.

We can check it also using the definition on conditional independence, infact:

$$(X_1 \perp X_6 | \{X_2, X_3\}) \Leftrightarrow \begin{cases} P(X_1, X_6 | \{X_2, X_3\}) = P(X_1 | \{X_2, X_3\})P(X_6 | \{X_2, X_3\}) \\ P(X_1 | \{X_2, X_3, X_6\}) = P(X_1 | \{X_2, X_3\}) \\ P(X_6 | \{X_1, X_2, X_3\}) = P(X_6 | \{X_2, X_3\}) \end{cases} \quad \text{since for symmetry } (X_1 \perp X_6 | \{X_2, X_3\}) = (X_6 \perp X_1 | \{X_2, X_3\})$$

## Exercise II: Sampling a DAG



In the DAG above, we define:

- The random vector  $X = [X_1, X_2] \sim N(\mu, \sigma^2 I_2)$  where  $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
- The vector  $\mu = [\mu_1, \mu_2]$
- The matrix  $\sigma^2 = \Sigma$  of variance and covariance:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

where  $\rho$  is the correlation between  $X_1$  and  $X_2$ .

- The conditional distribution of  $X_2$  given  $X_1$  is:

$$X_2 | X_1 \sim N(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho(x_1 - \mu_1), (1 - \rho)\sigma_2^2)$$

- So since  $\sigma^2 I_2 = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$ , for this reason  $X_2 | X_1 \sim N(\mu_2, \sigma_2^2)$

The process to define the vector  $Y = [Y_1, Y_2]$  is the same.

### 2.1 Joint distribution of the DAG above

Using conditional distributions of multivariate Normal vectors we can define:

$$\begin{aligned}
f(x_1, x_2, y_1, y_2, \mu, \sigma^2) &= f(\mu) f(\sigma^2) f(x_1 | \mu, \sigma^2) f(x_2 | \mu, \sigma^2, x_1) f(y_2 | \mu, \sigma^2) f(y_1 | \mu, \sigma^2, y_2) \\
&= f(\mu) f(\sigma^2) f(x_1 | \mu, \sigma^2) f(x_2 | \mu, \sigma^2) f(y_2 | \mu, \sigma^2) f(y_1 | \mu, \sigma^2) \\
&= f(\mu) f(\sigma^2) f(x_1) f(x_2) f(y_2) f(y_1)
\end{aligned}$$

## 2.2 Simulation on R

From this article (<http://www.stat.columbia.edu/~gelman/research/published/taumain.pdf>) we can assume as prior distributions the normal and gamma distributions for respectively  $\mu$  and  $\sigma^2$ .

–  $\mu_1 \sim N(0, 10)$  and  $\mu_2 \sim N(0, 10)$

–  $\sigma^2 \sim \text{Gamma}(1, 1)$

Given the joint probability defined in the previous exercise 2.1, through the forward/ancestral sampling we sample the variables in topological order. In other words, we start by sampling the variables with no parents; then we sample from the next generation by conditioning these variables' conditional probability distributions to values sampled at the first step. We proceed like this until all the variables have been sampled.

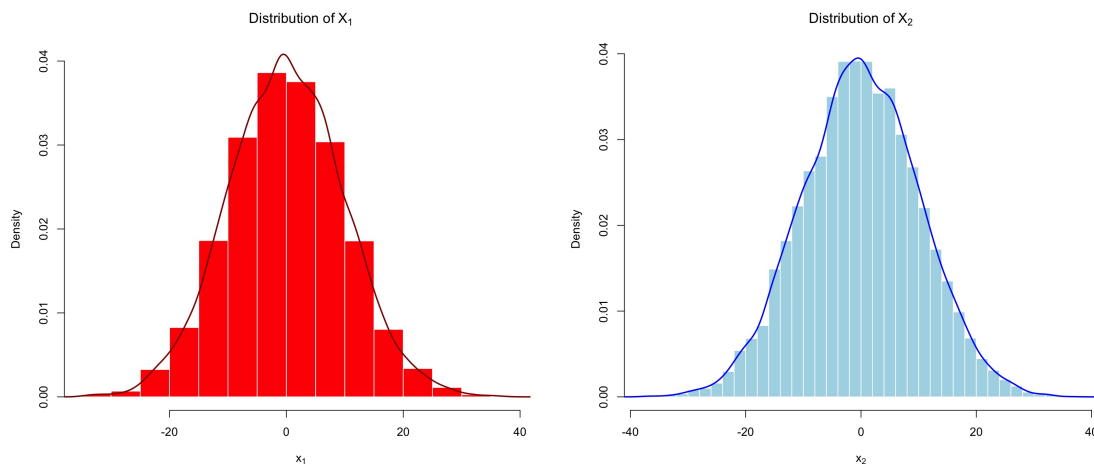
```
#initialize empty vectors
x1 <- c()
x2 <- c()
y1 <- c()
y2 <- c()

M <- 10000 #num of replication of ancestral sampling

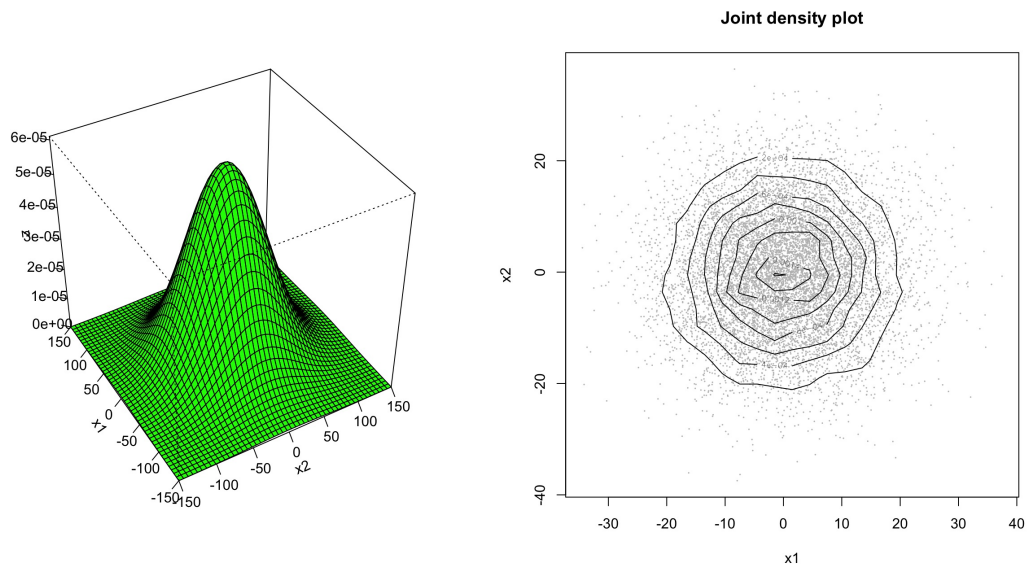
for (i in 1:M) {
  mu <- c(rnorm(2, 0, 10)) #vector of mu1 and mu2
  standard_deviation = sqrt(rgamma(1,1))
  x1[i] <- rnorm(1, mean = mu[1], sd = standard_deviation)
  x2[i] <- rnorm(1, mean = mu[2], sd = standard_deviation)
  y1[i] <- rnorm(1, mean = mu[1], sd = standard_deviation)
  y2[i] <- rnorm(1, mean = mu[2], sd = standard_deviation)
}
```

## 2.3 Plot each component of the vector $X$

Looking at the following plot we could expect that  $X_1$  and  $X_2$  follow a normal distribution



## 3D Plot



We decide to use the One-Sample Kolmogorov-Smirnov test to compare  $X_1$  and  $X_2$  with a reference probability distribution (in our case a normal distribution):

$H_0$  :  $X_1$  has normal distribution

$H_1$  :  $X_1$  doesn't have normal distribution

How does the test works on R? In input we have the parameter *alternative* referred to the alternative hypothesis  $H_1$  : in our case it is the default *two.sided*, that is in the null hypothesis  $H_0$  the true distribution function of  $x$  is equal to the hypothesized distribution function (one-sample case).

```
ks.test(x1, "pnorm", mean=mean(x1), sd=sd(x1))
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: x1
## D = 0.0062456, p-value = 0.8302
## alternative hypothesis: two-sided
```

```
ks.test(x2, "pnorm", mean=mean(x2), sd=sd(x2))
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: x2
## D = 0.006112, p-value = 0.8491
## alternative hypothesis: two-sided
```

The p-values of the tests are grathen than  $\alpha = 0.05$ , it means that we accept the null hypothesis and the two variables  $X_1$  and  $X_2$  are distributed as normal.

---

*Collaborations: Alice Schirina' and Eleonora Barocco & Daniele Sanna and Giorgio Zannini Quirini*

---