Introduction to Bayesian networks for gravitational wave scientists

Velimir M. Ilić

Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia

28 March, 2023
4th Training School on Gravitational Waves, Geophysics and Machine Learning
Aristotle University of Thessaloniki, Greece

Outline

- Bayesian networks were independently invented and widely used in different scientific disciplines such as artificial intelligence, statistics and signal processing.
- Using Bayesian networks, the structure of multivariate probabilistic distributions can be represented in elegant and efficient way which enables analysis of probabilistic dependences in the distribution.
- This makes them particularly usefull for the models with large number of parameters such as appear in gravitiaional wave modeling.
- In this lecture we discuss their applications for Bayesian inference in gravitational wave detection and point out some open questions in this research direction.

Directed acyclic graphs

- A directed graph is defined as a pair $G = (V, E)$ with V a finite set of vertices and E is a set of directed edges, $E \subseteq V \times V$ for which it holds that for each $(u, v) \in E : (v, u) \notin E$.
- Vertices are denoted by an lowercase letter such as v, u and w.
- A directed edge $(u, v) \in A$ is also denoted by $u \rightarrow v$.
- A path in a directed graph G is a sequence $v_1, v_2, \ldots, v_k$ of vertices in V, where $v_i \rightarrow v_{i+1}$.
- Directed cycle if is path which begins and ends at the same vertex, i.e. $v_1 = v_k$.
- A directed graph $G = (V, E)$ is called directed acyclic graph (DAG) if it contains no directed cycles.
- If $u \rightarrow v \in A$, u is called a parent of v and v is called a chilled of u.
- The set of parents of v is denoted by $\mathrm{pa}(v)$; the set of children of vertex u is denoted by $\mathrm{ch}(u)$.
- The set of descendants of u, denoted by $\delta(u)$, is the set of vertices $\delta(u) \subset V$, where there exists a path from u to each $v \in \delta(u)$, but no path from v to u.

[Lucas, Gámez and Salmerón 2007]

- We consider finite sets of indices, which are denoted by capital letters: U, V, W...
- For singleton sets $\{v\}$, we will write the element v instead of the set $\{v\}$.
- Let X_v be a discrete random variable which takes values x_v corresponding to $v \in V$. We define $X_W = (X_v)_{v \in W}$ and $x_W = (x_v)_{v \in W}$ for any subset $W \subseteq V$, where a natural order is assumed.
- Examples: If $U = \{a, b, c, f, l, s\}$, then $x_U = (x_a, x_b, x_c, x_f, x_l, x_s)$. If $W = \{a, b, f\}$, then $X_W = (X_a, X_b, X_f)$.
- Bayesian network is a pair (G, P), where $G = (V, E)$ is a directed acyclic graph and P is a joint probability distribution defined on a set of random variables X_V, which admits a recursive factorisation in terms of local (conditional) joint probability distributions $P\left(x_v \mid x_{\pi(v)}\right)$ as follows

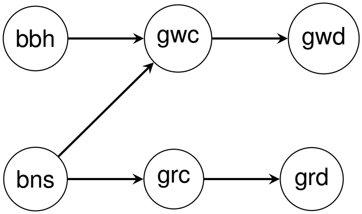
$$P(x_V) = \prod_{v \in V} P\left(x_v \mid x_{\mathrm{pa}(v)}\right)$$

[Lucas, Gámez and Salmerón 2007]

Simple Bayesian network

BBH	Binary black hole merger happens	$bbh \in \{\perp, \top\}$
GWC	"Strong enough"GW is created	$gwc \in \{\perp, \top\}$
GWD	GW is detected at the Earth	$gwd \in \{\perp, \top\}$
BNS	Binary neutron star merger happens	$bns \in \{\perp, \top\}$
GRC	"Strong enough"short gamma-ray burst is created	$grc \in \{\perp, \top\}$
GRD	Short gamma-ray burst is detected at the Earth	$grd \in \{\perp, \top\}$

$$P(bbh, gwc, gwd, bns, grc, grd) =$$
$$P(bbh)P(gwc \mid bbh, bns)P(gwd \mid gwc)$$
$$P(bns)P(grc \mid bns)P(grd \mid grc)$$



see also [Williams 2019]

Conditional independence

- Let X_V be a set of random variables with $U, W, Z \subseteq V$ disjoint sets of vertices, and let P be a joint probability distribution defined on X
- X_U is said to be conditionally independent of X_W given X_Z, denoted by

$$X_U \perp\!\!\!\perp_P X_W \mid X_Z$$

  if

$$P\left(X_U \mid X_W, X_Z\right) = P\left(X_U \mid X_Z\right).$$

  or equivalently

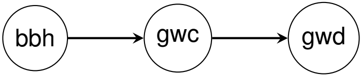
$$P\left(X_U, X_W \mid X_Z\right) = P\left(X_U \mid X_Z\right) P\left(X_W \mid X_Z\right).$$

- Conditional independence can be also interpreted as follows: hearning about X_W has no effect on our knowledge concerning X_U given our beliefs concerning X_Z, and vice versa.
- If $X_U \perp\!\!\!\perp_P X_W \mid X_Z$ does not hold, then X_U and X_W are said to be conditionally dependent given X_Z, which is written as follows:

$$X_U \not\perp\!\!\!\perp_P X_W \mid X_Z$$

Chain pattern

$$P(bbh, gwc, gwd) = P(bbh)P(gwc \mid bbh)P(gwd \mid gwc)$$



- In general

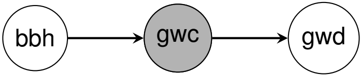
$$P(gwd, bbh) = \sum_{gwc} P(bbh)P(gwc \mid bbh)P(gwd \mid gwc) \neq P(bbh)P(gwd)$$

so that

$$BBH \not\perp GWD \mid \varnothing$$

Chain pattern

$$P(bbh, gwc, gwd) = P(bbh)P(gwc \mid bbh)P(gwd \mid gwc)$$



- In general

$$P(gwd, bbh) = \sum_{gwc} P(bbh)P(gwc \mid bbh)P(gwd \mid gwc) \neq P(bbh)P(gwd)$$

so that

$$BBH \not\perp\!\!\!\perp GWD \mid \varnothing$$

- Bayes' theorem

$$P(bbh \mid gwc) = \frac{P(bbh)P(gwc \mid bbh)}{P(gwc)}$$

- Conditional independence

$$P(bbh, gwd \mid gwc) = \frac{P(bbh, gwc, gwd)}{P(gwc)} = P(bbh \mid gwc)P(gwd \mid gwc)$$

so that

$$BBH \perp\!\!\!\perp GWD \mid GWC$$

Fork pattern

$$P(gwc, bns, grc) = P(gwc \mid bns)P(bns)P(grc \mid bns)$$

$$gwc \leftarrow bns \rightarrow grc$$

- In general

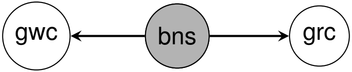
$$P(gwc, grc) = \sum_{gwc} P(gwc \mid bns)P(bns)P(grc \mid bns) \neq P(gwc)P(grc)$$

so that

$$GWC \not\perp GRC \mid \emptyset$$

Fork pattern

$$P(gwc, bns, grc) = P(gwc \mid bns) P(bns) P(grc \mid bns)$$



- In general

$$P(gwc, grc) = \sum_{gwc} P(gwc \mid bns) P(bns) P(grc \mid bns) \neq P(gwc) P(grc)$$

so that

$$GWC \not\perp\!\!\!\perp GRC \mid \varnothing$$

- Conditional independence

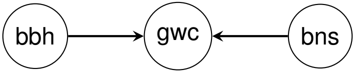
$$P(gwc, grc \mid bns) = \frac{P(gwc, bns, grc)}{P(bns)} = P(gwc \mid bns) P(grc \mid bns)$$

so that

$$GWC \perp\!\!\!\perp GRC \mid BNS$$

Collider pattern

$$P(bbh, gwc, bns) = P(gwc \mid bbh, bns)P(bbh)P(bns)$$



- In general

$$P(bbh, bns) = \sum_{gwc} P(gwc \mid bbh, bns)P(bbh)P(bns) = P(bbh)P(bns)$$

so that

$$BBH \perp\!\!\!\perp BNS \mid \varnothing$$

Collider pattern

$$P(bbh, gwc, bns) = P(gwc \mid bbh, bns) P(bbh) P(bns)$$

- In general

$$P(bbh, bns) = \sum_{gwc} P(gwc \mid bbh, bns) P(bbh) P(bns) = P(bbh) P(bns)$$

so that

$$BBH \perp\!\!\!\perp BNS \mid \varnothing$$

- Conditional independence does not hold in general

$$P(bbh, bns \mid gwc) = \frac{P(bns, gwc, bbh)}{P(gwc)} \neq P(bbh \mid gwc) P(bns \mid gwc)$$

so that

$$BBH \perp\!\!\!\perp BNS \mid GWC$$

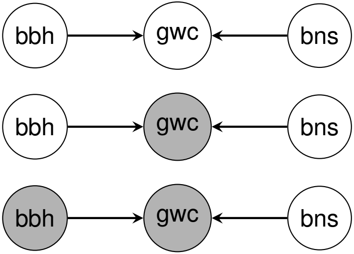
Conditioning in collider pattern

$$P(BBH = bbh, GWC = gwc, BNS = bns) =$$
$$P(GWC = gwc \mid BBH = bbh, BNS = bns)P(BBH = bbh)P(BNS = bns)$$

$P(BNS = bns)$	
$bns = \perp$	$bns = \top$
0,9	0,1

$P(BBH = bbh)$	
$bbh = \perp$	$bbh = \top$
0,9	0,1

$P(GWC = gwc \mid BBH = bbh, BNS = bns)$				
	$bbh = \perp$, $bns = \perp$	$bbh = \perp$, $bns = \top$	$bbh = \top$, $bns = \perp$	$bbh = \top$, $bns = \top$
$gwc = \perp$	0,8	0,2	0,2	0,1
$gwc = \top$	0,2	0,8	0,8	0,9



- No observation:
$$P(BNS = \top) = 0{,}1$$
- $GWC = \top$ observed:
$$P(BNS = \top \mid GWC = \top) = ?$$
- $GWC = \top$ and $BBH = \top$ observed:
$$P(BNS = \top \mid GWC = \top, BBH = \top) = ?$$

Conditioning in collider pattern

- Marginal probabilities:

$$P(BNS = bns, GWC = gwc) = \sum_{bbh \in \{\perp, \top\}} P(GWC = gwc, BBH = bbh, BNS = bns)$$

$$P(GWC = gwc) = \sum_{bns \in \{\perp, \top\}} P(BNS = bns, GWC = gwc)$$

$$P(GWC = gwc, BBH = bbh) = \sum_{bns \in \{\perp, \top\}} P(GWC = gwc, BBH = bbh, BNS = bns)$$

- Conditional probabilities:

$$P(BNS = bns \mid GWC = gwc) = \frac{P(BNS = bns, GWC = gwc)}{P(GWC = gwc)}$$

$$P(BNS = bns \mid GWC = gwc, BBH = bbh) = \frac{P(BNS = bns, GWC = gwc, BBH = bbh)}{P(GWC = gwc, BBH = bbh)}$$

Conditioning in collider pattern

- Marginal probabilities:

$$P(BNS = bns, GWC = \top) = \sum_{bbh \in \{\perp, \top\}} P(BNS = bns, GWC = \top, BBH = bbh)$$

$$P(GWC = \top) = \sum_{bns \in \{\perp, \top\}} P(BNS = bns, GWC = \top)$$

$$P(GWC = \top, BBH = \top) = \sum_{bns \in \{\perp, \top\}} P(BNS = bns, GWC = \top, BBH = \top)$$

- Conditional probabilities:

$$P(BNS = \top \mid GWC = \top) = \frac{P(BNS = \top, GWC = \top)}{P(GWC = \top)}$$

$$P(BNS = \top \mid GWC = \top, BBH = \top) = \frac{P(BNS = \top, GWC = \top, BBH = \top)}{P(GWC = \top, BBH = \top)}$$

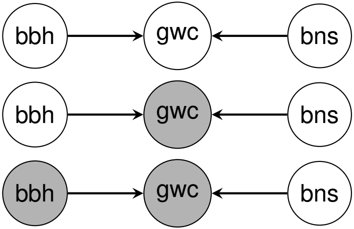
Conditioning in collider pattern

$$P(BBH = bbh, GWC = gwc, BNS = bns) =$$
$$P(GWC = gwc \mid BBH = bbh, BNS = bns)P(BBH = bbh)P(BNS = bns)$$

$P(BNS = bns)$	
$bns = \perp$	$bns = \top$
0,9	0,1

$P(BBH = bbh)$	
$bbh = \perp$	$bbh = \top$
0,9	0,1

$P(GWC = gwc \mid BBH = bbh, BNS = bns)$				
	$bbh = \perp$, $bns = \perp$	$bbh = \perp$, $bns = \top$	$bbh = \top$, $bns = \perp$	$bbh = \top$, $bns = \top$
$gwc = \perp$	0,8	0,2	0,2	0,1
$gwc = \top$	0,2	0,8	0,8	0,9



$$P(BNS = \top) = 0{,}1$$

$$P(BNS = \top \mid GWC = \top) \approx 0{,}257$$

$$P(BNS = \top \mid GWC = \top, BBH = \top) \approx 0{,}111$$

$$P(BNS = \top \mid GWC = \top) > P(BNS = \top \mid GWC = \top, BBH = \top) > P(BNS = \top)$$

Monte hall problem

1. The car is uniformly distributed behind one of three closed doors. The other two are empty. The guest makes a uniform random choice at one door, trying to guess and win the car.
2. The host selects a door which is not selected by the guest and which does not hide the car. If the car is behind the guest's door, the host makes a random uniform selection between the empty doors.
3. The host opens the selected door. Will the guest increase the chances of winning by switching the choice to another closed door?

Monte hall problem

1. The car is uniformly distributed behind one of three closed doors. The other two are empty. The guest makes a uniform random choice at one door, trying to guess and win the car.
2. The host selects a door which is not selected by the guest and which does not hide the car. If the car is behind the guest's door, the host makes a random uniform selection between the empty doors.
3. The host opens the selected door. Will the guest increase the chances of winning by switching the choice to another closed door?

$$P(G=g)=\frac{1}{3} \quad P(C=c)=\frac{1}{3}$$

$$P(H=h \mid C=c, G=g)=$$

$$[h \neq g][h \neq c]\left([g \neq c]+\frac{1}{2}\cdot[g=c]\right)$$

$$g,h,c \in \{1,2,3\}$$

$$\begin{tikzcd} g \arrow[r] & h & c \arrow[l] \end{tikzcd}$$

Monte Hall conditioning

- Joint probability distribution:

$$P(G=g,H=3,C=c)=\frac{1}{9}\cdot[g\neq 3][c\neq 3]\Big([g\neq c]+\frac{1}{2}\cdot[g=c]\Big)$$

- Marginal probabilities:

$$P(H=3,C=c)=\sum_{g=1}^{3}P(G=g,H=3,C=c)=\frac{1}{6}\cdot[c\neq 3]$$

$$P(H=3)=\sum_{c=1}^{3}P(H=3,C=c)=\frac{1}{3}$$

$$P(G=1,H=3)=\sum_{c=1}^{3}P(G=1,H=3,C=c)=\frac{1}{6}$$

- Conditional probabilities:

$$P(C=1\mid H=3)=\frac{P(H=3,C=1)}{P(H=3)}=\frac{1}{2}$$

$$P(C=1\mid H=3,G=1)=\frac{P(G=1,H=3,C=1)}{P(G=1,H=3)}=\frac{1}{3}$$

Modified Monte Hall problem

1. The car is uniformly distributed behind one of three closed doors. The other two are empty. The guest makes a uniform random choice at one door, trying to guess and win the car.
2. The host selects a door which is not selected by the guest uniformly randomly. If the car is behind the host's door, the car driver switches the car to another door uniformly randomly, otherwise nothing changes.
3. The host opens the selected door. Will the guest increase the chances of winning by switching the choice to another closed door?

Modified Monte Hall problem

1. The car is uniformly distributed behind one of three closed doors. The other two are empty. The guest makes a uniform random choice at one door, trying to guess and win the car.
2. The host selects a door which is not selected by the guest uniformly randomly. If the car is behind the host's door, the car driver switches the car to another door uniformly randomly, otherwise nothing changes.
3. The host opens the selected door. Will the guest increase the chances of winning by switching the choice to another closed door?

$$P(G=g)=\frac{1}{3} \quad P(H=h\mid G=g)=\frac{1}{2}\cdot[g\neq h]$$

$$P(C=c\mid H=h)=\frac{1}{2}\cdot[c\neq h]$$

$$g,h,c\in\{1,2,3\}$$

$$g \longrightarrow h \longrightarrow c$$

Modified Monte Hall conditioning

- Joint probability distribution:

$$P(G=g, H=h, C=c) = \frac{1}{12} \cdot [g \neq h][c \neq h]$$

- Marginal probabilities $(G=1, H=3)$:

$$P(H=3, C=c) = \sum_{g=1}^{3} P(G=g, H=3, C=c) = \frac{1}{6} \cdot [c \neq 3]$$

$$P(H=3) = \sum_{c=1}^{3} P(H=3, C=c) = \frac{1}{3}$$

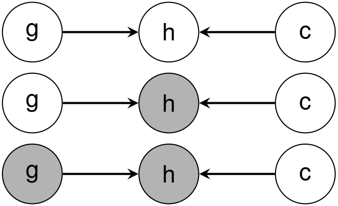
$$P(G=1, H=3) = \sum_{c=1}^{3} P(G=1, H=3, C=c) = \frac{1}{6}$$

- Conditional probabilities:

$$P(C=1 \mid H=3) = \frac{P(H=3, C=1)}{P(H=3)} = \frac{1}{2}$$

$$P(C=1 \mid H=3, G=1) = \frac{P(G=1, H=3, C=1)}{P(G=1, H=3)} = \frac{1}{2}$$

Monte Hall problem



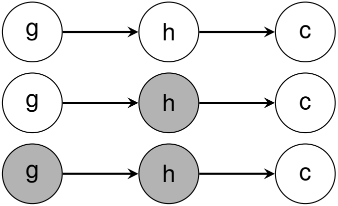
$$P(C=1) = \frac{1}{3}$$

$$P(C=1 \mid H=3) = \frac{1}{2}$$

$$P(C=1 \mid H=3, G=1) = \frac{1}{3}$$

$$P(C=1 \mid H=3) > P(C=1 \mid H=3, G=1) = P(C=1)$$

Modified Monte Hall problem



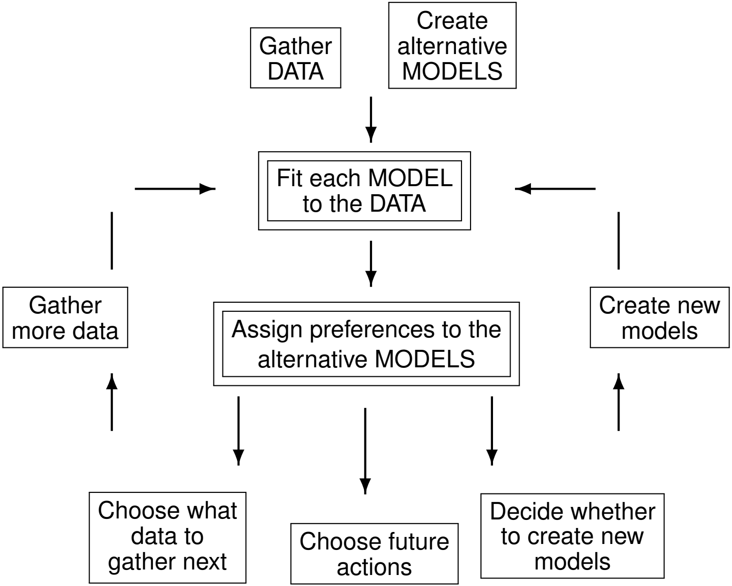
$$P(C=1) = \frac{1}{3}$$

$$P(C=1 \mid H=3) = \frac{1}{2}$$

$$P(C=1 \mid H=3, G=1) = \frac{1}{2}$$

$$P(C=1 \mid H=3) = P(C=1 \mid H=3, G=1) > P(C=1)$$

Bayesian inference and model selection



[MacKay 2003]

The d-separation criterion

Path blocking

A set W of vertices is said to block a path p if either of the following two condition is satisfied:

1. p contains at least one arrow-emitting node (middle node in a chain or a fork pattern) that is in W.
2. p contains at least one collision node (middle node in a collider pattern) that is outside W and has no descendant in W

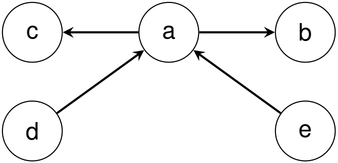
The d-separation criterion

Let U, V, W be disjoint sets of vertices. If W blocks all paths from $u \in U$ to $v \in V$, it is said to "d-separate u and v". Two set U and V are said to be d-separated by W if any pair u and v from the two sets are d-separated by W. If so, then we have

$$X_U \perp\!\!\!\perp X_V \mid X_W$$

[Bishop 2006, Lucas, Gámez and Salmerón 2007, Deng]

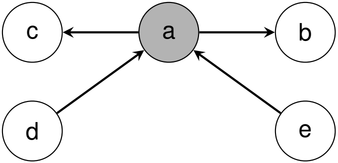
The d-separation criterion



$X_b \not\perp\!\!\!\perp X_c$ ($b \to a \to c$ is **open**: fork)

$X_c \not\perp\!\!\!\perp X_e$ ($e \to a \to c$ is **open**: chain)

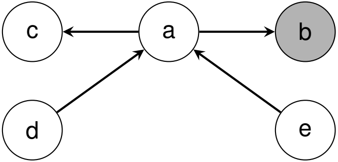
$X_d \perp\!\!\!\perp X_e$ ($d \to a \to e$ is **blocked**: collider)



$X_b \not\perp\!\!\!\perp X_c \mid X_a$ ($b \to a \to c$ is **blocked**: fork)

$X_c \not\perp\!\!\!\perp X_e \mid X_a$ ($e \to a \to c$ is **blocked**: chain)

$X_d \perp\!\!\!\perp X_e \mid X_a$ ($d \to a \to e$ is **open**: collider)

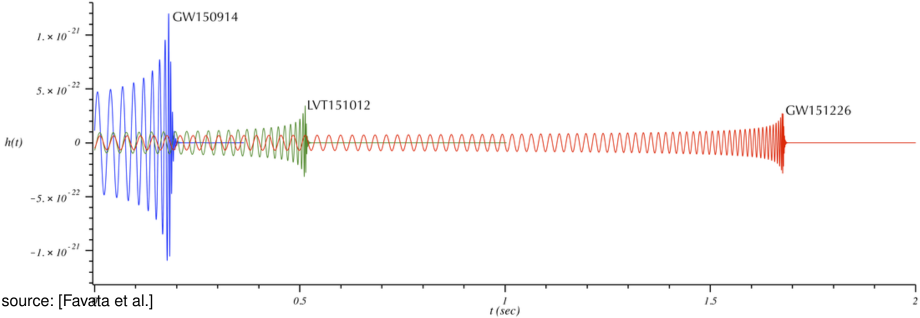


$X_b \not\perp\!\!\!\perp X_c \mid X_a$ ($b \to a \to c$ is **open**: fork)

$X_c \not\perp\!\!\!\perp X_e \mid X_a$ ($e \to a \to c$ is **open**: chain)

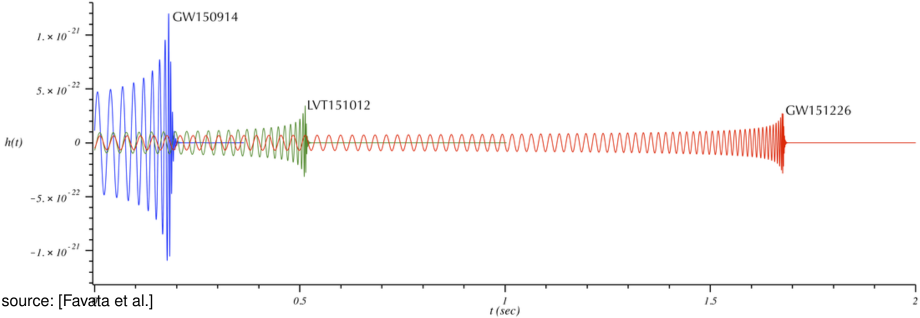
$X_d \perp\!\!\!\perp X_e \mid X_a$ ($d \to a \to e$ is **open**: collider)

Back to the gravitational waves

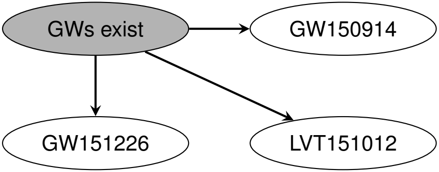


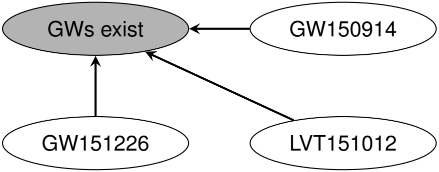
source: [Favata et al.]

Back to the gravitational waves



source: [Favata et al.]



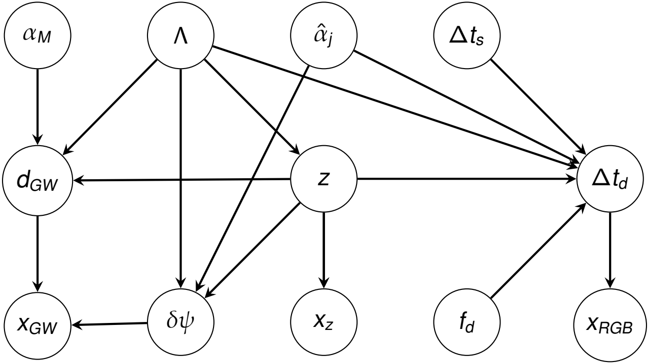


$\text{LVT151012} \perp\!\!\!\perp \text{GW150914}, \text{GW151226} \mid \text{GWs exist}$

$\text{LVT151012} \not\perp\!\!\!\perp \text{GW150914}, \text{GW151226} \mid \text{GWs exist}$

see also [D'Agostini 2002]

A cosmological Bayesian network model

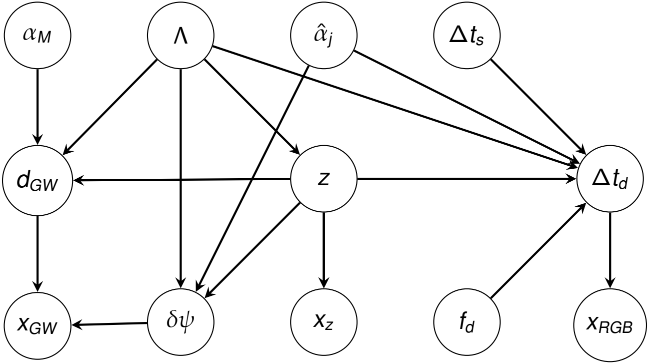


α_M GW friction	d_{GW} GW luminosity distance	Δt_s Initial GW-EM delay	x_{GW} GW data
Λ Huble constant	z Source red shift	Δt_d GW-EM delay	x_{RGB} RGB data
$\hat{\alpha}_j$ GW dispersion	$\delta\psi$ GW phase shift	f_d Merger frequency	x_z HG data

[Mastrogiovanni et al. 2020]

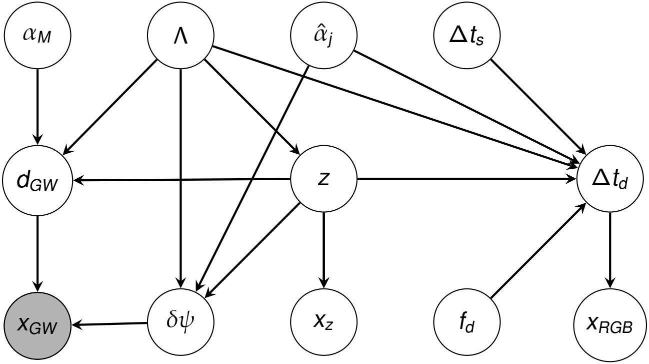
Quiz time...

Are α_M and $\hat{\alpha}_j$ correlated?

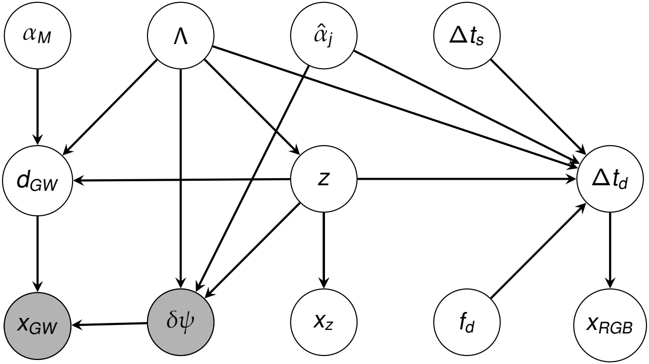


A cosmological Bayesian network model: quiz time

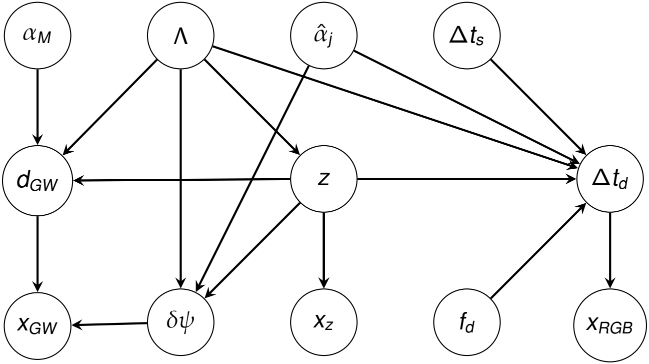
Are α_M and $\hat{\alpha}_j$ correlated if x_{GW} is given?



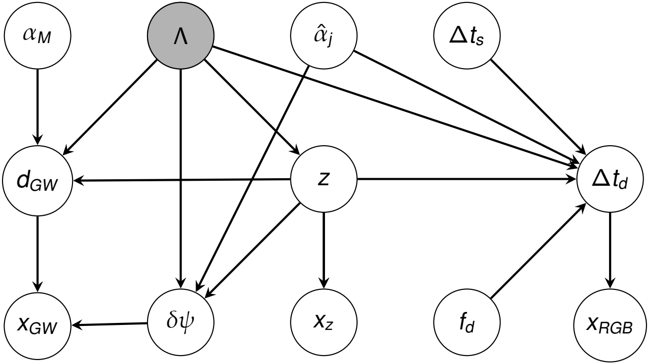
Are α_M and $\hat{\alpha}_j$ correlated if x_{GW} and $\delta\psi$ are given?



Are α_M and Δt_d correlated?

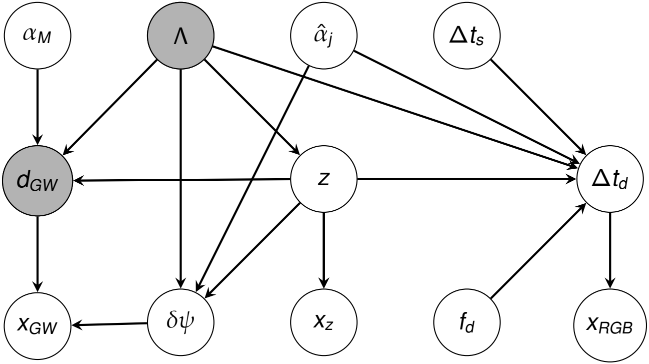


Are α_M and Δt_d correlated if Λ is given?



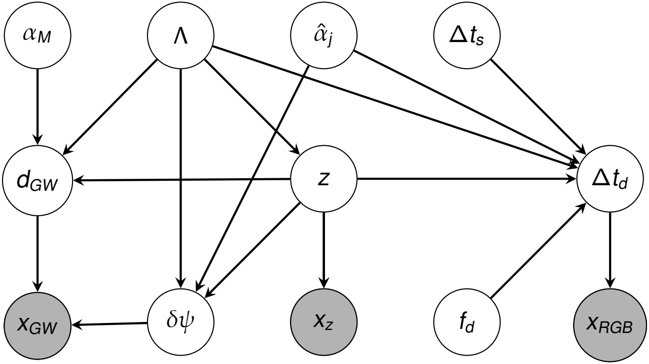
A cosmological Bayesian network model: quiz time

Are α_M and Δt_d correlated if Λ and d_{GW} are given?



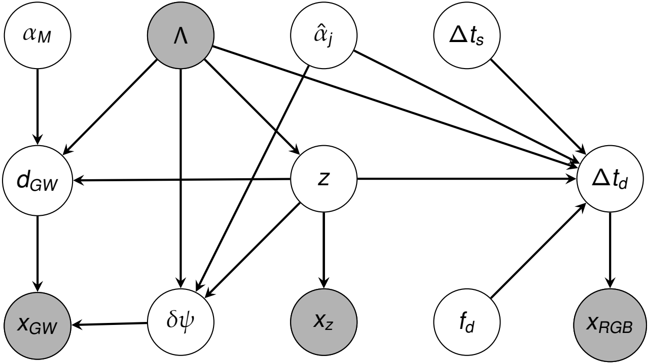
A cosmological Bayesian network model: quiz time

Are α_M and $\hat{\alpha}_j$ correlated if x_{GW}, x_z and x_{RGB} are given?



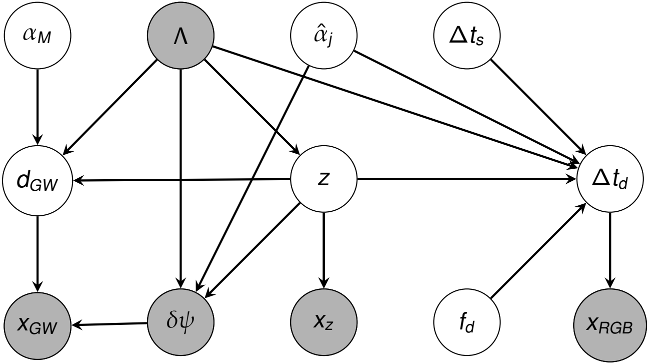
A cosmological Bayesian network model: quiz time

Are α_M and $\hat{\alpha}_j$ correlated if x_{GW}, x_z, x_{RGB} and Λ are given?



A cosmological Bayesian network model: quiz time

Are α_M and $\hat{\alpha}_j$ correlated if x_{GW}, x_z, x_{RGB}, Λ and $\delta\psi$ are given?



Used sources and recommended readings

- MacKay, David JC, and David JC Mac Kay. Information theory, inference and learning algorithms. Cambridge university press, 2003.
- Bishop, Christopher M., and Nasser M. Nasrabadi. Pattern recognition and machine learning. Vol. 4. No. 4. New York: springer, 2006.
- Lucas, Peter, José A. Gámez, and Antonio Salmerón Cerdan, eds. Advances in Probabilistic Graphical Models. Vol. 213. Springer, 2007.
- Williams, Daniel. Inference methods for gravitational wave data analysis. Diss. University of Glasgow, 2019.
- Mastrogiovanni, S., D. A. Steer, and M. Barsuglia. ”Probing modified gravity theories and cosmology using gravitational-waves and associated electromagnetic counterparts.”Physical Review D 102.4 (2020): 044009.
- D’Agostini, Giulio. ”The waves and the sigmas (to say nothing of the 750 gev mirage).” arXiv preprint arXiv:1609.01668 (2016).
- Favata, Marc et al. Sounds of spacetime: https://www.soundsofspacetime.org/second-detection-gw151226–lvt151012.html
- Berry, Christopher. An introduction to probability: Inference and learning from data: https://cplberry.com/tag/bayes-theorem/
- Deng, Alex. Causal Inference and Its Applications in Online Industry: https://alexdeng.github.io/causal/

Thank You for your attention!