Uncertainty and statistics
Part II

Håkon Tjelmeland

CGF and Department of Mathematical Sciences
Norwegian University of Science and Technology

Friday October 29th 2021

Recall from last time

- Stochastic variable, X
  - X represents an unknown quantity
  - $p(x)$ describes our knowledge about X

- The effect of conditioning
  - two stochastic variables, $p(x_1, x_2)$
  - $p(x_1) = \int_{-\infty}^{\infty} p(x_1, x_2) dx_2$ describes our knowledge about x_1
  - after we have observed a value for x_2,

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

    describes our updated knowledge about x_1

- Three strategies to specify $p(x_1, x_2)$
  - specify a formula for $p(x_1, x_2)$ directly (often Gaussian)
  - conditional probability, $p(x_1, x_2) = p(x_1)p(x_2|x_1)$
  - hierarchical model, $x_1, x_2|\theta \sim p(x|\theta)$ indep., and $\theta \sim p(\theta)$

- Use graphical model to visualise how we specified $p(x_1, x_2)$

Plan for today

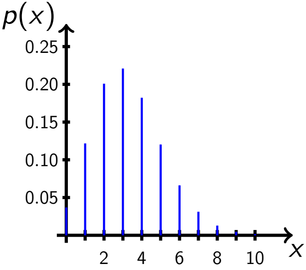
- Interpretation of a distribution, $p(x)$
- Modelling of noisy observations
- Independence and conditional independence
- Use of conditional independence for modelling of n variables
  - Markov chain, CO_2 leakage example
  - a larger network example
- Hierarchical models
  - the effect of conditioning
- How to look at n-dimensional distributions
  - what quantities are we interested in?

Plan for today

- ⋆ Interpretation of a distribution, $p(x)$
- ⋆ Modelling of noisy observations
- ⋆ Independence and conditional independence
- ⋆ Use of conditional independence for modelling of n variables
  - Markov chain, CO_2 leakage example
  - a larger network example
- ⋆ Hierarchical models
  - the effect of conditioning
- ⋆ How to look at n-dimensional distributions
  - what quantities are we interested in?

Interpretation of $p(x)$

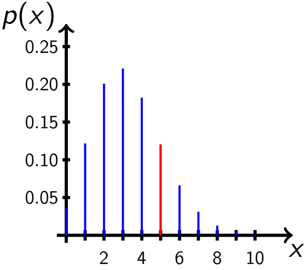
- ⋆ Discrete stochastic variable



- ⋆ Interpretation: $p(x) = P(X = x)$

Interpretation of $p(x)$

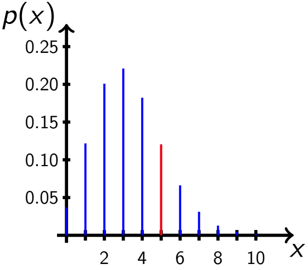
- ⋆ Discrete stochastic variable



- ⋆ Interpretation: $p(x) = P(X = x)$
  - for example: $p(5) = P(X = 5)$

Interpretation of $p(x)$

- Discrete stochastic variable



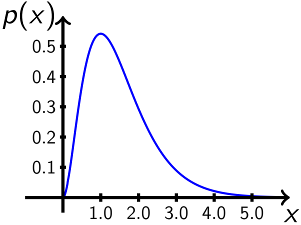
- Interpretation: $p(x) = P(X = x)$
  - for example: $p(5) = P(X = 5)$

- With two stochastic variables

$$p(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$$

Interpretation of $p(x)$

- Continuous stochastic variable

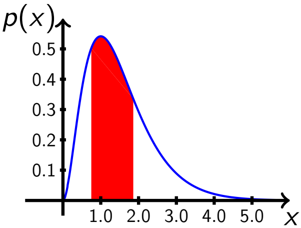


- Interpretation:

$$P(a \leq X \leq b) = \int_a^b p(x)dx$$

Interpretation of $p(x)$

- Continuous stochastic variable



- Interpretation:

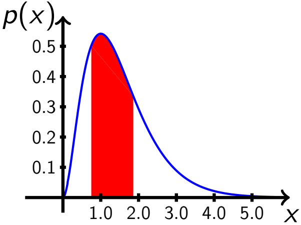
$$P(a \leq X \leq b) = \int_a^b p(x)dx$$

  - for example:

$$P(0.75 \leq X \leq 1.86) = \int_{0.45}^{1.86} p(x)dx$$

Interpretation of $p(x)$

- ⋆ Continuous stochastic variable



- ⋆ Interpretation:

$$P(a \leq X \leq b) = \int_a^b p(x)dx$$

  - for example:

$$P(0.75 \leq X \leq 1.86) = \int_{0.45}^{1.86} p(x)dx$$

  - two continuous stochastic variables

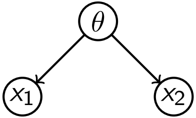
$$P((X_1, X_2) \in A) = \iint_A p(x_1, x_2)dx_1 dx_2$$

Plan for today

- Interpretation of a distribution, $p(x)$
- Modelling of noisy observations
- Independence and conditional independence
- Use of conditional independence for modelling of n variables
  - Markov chain, CO_2 leakage example
  - a larger network example
- Hierarchical models
  - the effect of conditioning
- How to look at n-dimensional distributions
  - what quantities are we interested in?

Recall: Hierarchical model example

⋆ Assume: $x_1, x_2|\theta \sim p(x|\theta)$ independently, $\theta \sim p(\theta)$



⋆ Simplistic example:

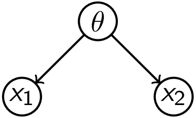
- $\theta \in \{0, 1\}$: source rock has produced hydrocarbons
- $x_1 \in \{0, 1\}$: hydrocarbons present in prospect 1
- $x_2 \in \{0, 1\}$: hydrocarbons present in prospect 2
- assume probabilities: $p(\theta = 1) = 0.3$

$$p(x_1 = 1|\theta = 0) = 0, \; p(x_1 = 1|\theta = 1) = 0.6$$
$$p(x_2 = 1|\theta = 0) = 0, \; p(x_2 = 1|\theta = 1) = 0.6$$

⋆ We found and compared: $p(x_1 = 1)$ and $p(x_1 = 1|x_2 = 0)$

Recall: Hierarchical model example

- Assume: $x_1, x_2|\theta \sim p(x|\theta)$ independently, $\theta \sim p(\theta)$



- Simplistic example:
  - $\theta \in \{0, 1\}$: source rock has produced hydrocarbons
  - $x_1 \in \{0, 1\}$: hydrocarbons present in prospect 1
  - $x_2 \in \{0, 1\}$: hydrocarbons present in prospect 2
  - assume probabilities: $p(\theta = 1) = 0.3$

$$p(x_1 = 1|\theta = 0) = 0, \;\; p(x_1 = 1|\theta = 1) = 0.6$$
$$p(x_2 = 1|\theta = 0) = 0, \;\; p(x_2 = 1|\theta = 1) = 0.6$$

  - We found and compared: $p(x_1 = 1)$ and $p(x_1 = 1|x_2 = 0)$

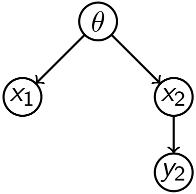
- Now assume we observe x_2 with noise
  - how can we model this?

Modelling of observation noise

- New stochastic variable for observation: $y_2 \in \{0, 1, 2\}$
  - assume

$$
\begin{aligned}
p(y_2 = 0|x_2 = 0) &= 0.6 \qquad & p(y_2 = 0|x_2 = 1) &= 0.1 \\
p(y_2 = 1|x_2 = 0) &= 0.3 \qquad & p(y_2 = 1|x_2 = 1) &= 0.2 \\
p(y_2 = 2|x_2 = 0) &= 0.1 \qquad & p(y_2 = 2|x_2 = 1) &= 0.7
\end{aligned}
$$

- New graphical model:



- Joint distribution:

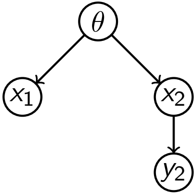
$$p(\theta, x_1, x_2, y_2) = p(\theta)p(x_1|\theta)p(x_2|\theta)p(y_2|x_2)$$

Modelling of observation noise

- New stochastic variable for observation: $y_2 \in \{0, 1, 2\}$
  - assume

$$p(y_2 = 0|x_2 = 0) = 0.6 \qquad p(y_2 = 0|x_2 = 1) = 0.1$$
$$p(y_2 = 1|x_2 = 0) = 0.3 \qquad p(y_2 = 1|x_2 = 1) = 0.2$$
$$p(y_2 = 2|x_2 = 0) = 0.1 \qquad p(y_2 = 2|x_2 = 1) = 0.7$$

- New graphical model:



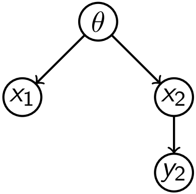
- Joint distribution:

$$p(\theta, x_1, x_2, y_2) = p(\theta)p(x_1|\theta)p(x_2|\theta)p(y_2|x_2)$$

- Probabilities of interest:

$$p(x_1 = 1|y_2 = 0), \quad p(x_2 = 1|y_2 = 0)$$
$$p(x_1 = 1|y_2 = 1), \quad p(x_2 = 1|y_2 = 1)$$
$$p(x_1 = 1|y_2 = 2), \quad p(x_2 = 1|y_2 = 2)$$

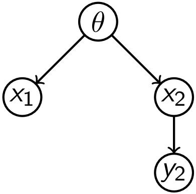
Modelling of observation noise



- "Detailed" calculation of one probability

$$
\begin{aligned}
p(x_1 = 1 | y_2 = 0) &= \frac{p(x_1 = 1, y_2 = 0)}{p(y_2 = 0)} \\
&= \frac{\sum_{\theta=0}^{1} \sum_{x_2=0}^{1} p(\theta, x_1 = 1, x_2, y_2 = 0)}{\sum_{\theta=0}^{1} \sum_{x_1=0}^{1} \sum_{x_2=0}^{1} p(\theta, x_1, x_2, y_2 = 0)} \\
&= \frac{0.054}{0.51} = 0.1059
\end{aligned}
$$

Modelling of observation noise



- "Detailed" calculation of one probability

$$
\begin{aligned}
p(x_1 = 1 | y_2 = 0) &= \frac{p(x_1 = 1, y_2 = 0)}{p(y_2 = 0)} \\
&= \frac{\sum_{\theta=0}^{1} \sum_{x_2=0}^{1} p(\theta, x_1 = 1, x_2, y_2 = 0)}{\sum_{\theta=0}^{1} \sum_{x_1=0}^{1} \sum_{x_2=0}^{1} p(\theta, x_1, x_2, y_2 = 0)} \\
&= \frac{0.054}{0.51} = 0.1059
\end{aligned}
$$

- Resulting probabilities:

$$
\begin{array}{ll}
p(x_1 = 1 | y_2 = 0) = 0.1059 & p(x_2 = 1 | y_2 = 0) = 0.0353 \\
p(x_1 = 1 | y_2 = 1) = 0.1532 & p(x_2 = 1 | y_2 = 1) = 0.1277 \\
p(x_1 = 1 | y_2 = 2) = 0.3981 & p(x_2 = 1 | y_2 = 2) = 0.6058
\end{array}
$$

$$
p(x_1 = 1) = 0.18 \qquad\qquad p(x_2 = 1) = 0.18
$$

Plan for today

- ⋆ Interpretation of a distribution, $p(x)$
- ⋆ Modelling of noisy observations
- ⋆ Independence and conditional independence
- ⋆ Use of conditional independence for modelling of n variables
  - Markov chain, CO_2 leakage example
  - a larger network example
- ⋆ Hierarchical models
  - the effect of conditioning
- ⋆ How to look at n-dimensional distributions
  - what quantities are we interested in?

Independence and conditional independence

- Independence: x_1 and x_2 are independent if

$$p(x_1|x_2) = p(x_1)$$

  - from this it follows

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)} = p(x_1) \quad \Leftrightarrow \quad p(x_1, x_2) = p(x_1) \cdot p(x_2)$$

  - in turn this implies

$$p(x_2|x_1) = \frac{p(x_1, x_2)}{p(x_1)} = \frac{p(x_1) \cdot p(x_2)}{p(x_1)} = p(x_2)$$

- Graphical model showing that x_1 and x_2 are independent:

Independence and conditional independence

- Conditional independence: x_1 and x_2 are conditionally independent given θ if

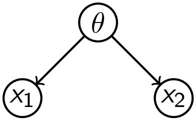
$$p(x_1|x_2, \theta) = p(x_1|\theta)$$

  - from this it follows

$$p(x_1, x_2|\theta) = p(x_1|\theta) \cdot p(x_2|\theta)$$

  - note that x_1 and x_2 are then not independent!

- Graphical model showing that x_1 and x_2 are conditionally independent given θ



How conditional independence helps specifying a model

- Simplistic example:
  - $\theta \in \{0,1\}$: source rock has produced hydrocarbons
  - $x_1 \in \{0,1\}$: hydrocarbons present in prospect 1
  - $x_2 \in \{0,1\}$: hydrocarbons present in prospect 2
  - $y_2 \in \{0,1,2\}$: noisy observation of x_2
- Want a model for θ, x_1, x_2 and y_2, i.e. $p(\theta, x_1, x_2, y_2)$
- Always true:

$$\begin{aligned} p(\theta, x_1, x_2, y_2) &= p(\theta) \cdot p(x_1|\theta) \cdot p(x_2|\theta, x_1) \cdot p(y_2|\theta, x_1, x_2) \\ &= p(\theta) \cdot p(x_1|\theta) \cdot p(y_2|\theta, x_1) \cdot p(x_2|\theta, x_1, y_2) \\ &\vdots \\ &= p(y_2) \cdot p(x_2|y_2) \cdot p(x_1|y_2, x_2) \cdot p(\theta|y_2, x_2, x_1) \end{aligned}$$

How conditional independence helps specifying a model

- Simplistic example:
  - $\theta \in \{0,1\}$: source rock has produced hydrocarbons
  - $x_1 \in \{0,1\}$: hydrocarbons present in prospect 1
  - $x_2 \in \{0,1\}$: hydrocarbons present in prospect 2
  - $y_2 \in \{0,1,2\}$: noisy observation of x_2
- Want a model for θ, x_1, x_2 and y_2, i.e. $p(\theta, x_1, x_2, y_2)$
- Always true:

$$\begin{aligned} p(\theta, x_1, x_2, y_2) &= p(\theta) \cdot p(x_1|\theta) \cdot p(x_2|\theta, x_1) \cdot p(y_2|\theta, x_1, x_2) \quad \Leftarrow \\ &= p(\theta) \cdot p(x_1|\theta) \cdot p(y_2|\theta, x_1) \cdot p(x_2|\theta, x_1, y_2) \\ &\vdots \\ &= p(y_2) \cdot p(x_2|y_2) \cdot p(x_1|y_2, x_2) \cdot p(\theta|y_2, x_2, x_1) \end{aligned}$$

How conditional independence helps specifying a model

- Simplistic example:
  - $\theta \in \{0,1\}$: source rock has produced hydrocarbons
  - $x_1 \in \{0,1\}$: hydrocarbons present in prospect 1
  - $x_2 \in \{0,1\}$: hydrocarbons present in prospect 2
  - $y_2 \in \{0,1,2\}$: noisy observation of x_2
- Want a model for θ, x_1, x_2 and y_2, i.e. $p(\theta, x_1, x_2, y_2)$
- Always true:

$$\begin{aligned}
p(\theta, x_1, x_2, y_2) &= p(\theta) \cdot p(x_1|\theta) \cdot p(x_2|\theta, x_1) \cdot p(y_2|\theta, x_1, x_2) \quad \Leftarrow \\
&= p(\theta) \cdot p(x_1|\theta) \cdot p(y_2|\theta, x_1) \cdot p(x_2|\theta, x_1, y_2) \\
&\;\vdots \\
&= p(y_2) \cdot p(x_2|y_2) \cdot p(x_1|y_2, x_2) \cdot p(\theta|y_2, x_2, x_1)
\end{aligned}$$

- Assume conditional independence

$$p(x_2|\theta, x_1) = p(x_2|\theta)$$
$$p(y_2|\theta, x_1, x_2) = p(y_2|x_2)$$

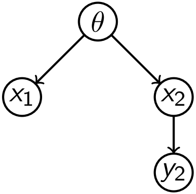
How conditional independence helps specifying a model

- Simplistic example:
  - $\theta \in \{0, 1\}$: source rock has produced hydrocarbons
  - $x_1 \in \{0, 1\}$: hydrocarbons present in prospect 1
  - $x_2 \in \{0, 1\}$: hydrocarbons present in prospect 2
  - $y_2 \in \{0, 1, 2\}$: noisy observation of x_2
- Want a model for θ, x_1, x_2 and y_2, i.e. $p(\theta, x_1, x_2, y_2)$
- Always true:

$$
\begin{aligned}
p(\theta, x_1, x_2, y_2) &= p(\theta) \cdot p(x_1|\theta) \cdot p(x_2|\theta, x_1) \cdot p(y_2|\theta, x_1, x_2) \quad \Leftarrow \\
&= p(\theta) \cdot p(x_1|\theta) \cdot p(y_2|\theta, x_1) \cdot p(x_2|\theta, x_1, y_2) \\
&\vdots \\
&= p(y_2) \cdot p(x_2|y_2) \cdot p(x_1|y_2, x_2) \cdot p(\theta|y_2, x_2, x_1)
\end{aligned}
$$

- Assume conditional independence

$$
\begin{aligned}
p(x_2|\theta, x_1) &= p(x_2|\theta) \\
p(y_2|\theta, x_1, x_2) &= p(y_2|x_2)
\end{aligned}
$$

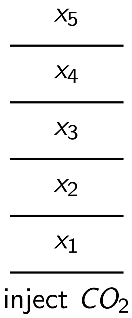


Plan for today

- ⋆ Interpretation of a distribution, $p(x)$
- ⋆ Modelling of noisy observations
- ⋆ Independence and conditional independence
- ⋆ Use of conditional independence for modelling of n variables
  - Markov chain, CO_2 leakage example
  - a larger network example
- ⋆ Hierarchical models
  - the effect of conditioning
- ⋆ How to look at n-dimensional distributions
  - what quantities are we interested in?

Conditional independence example: CO_2 leakage

- ⋆ Inject CO_2 in layered reservoir
- ⋆ $x_i \in \{0, 1\}$: injected CO_2 present in layer i



- ⋆ Modelling

$$p(x_1, x_2, x_3, x_4, x_5)$$
$$= p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_3)p(x_5|x_4)$$

where

$$p(x_1 = 1) = p_1$$
$$p(x_i = 1|x_{i-1} = 0) = 0, \quad p(x_i = 1|x_{i-1} = 1) = p_i$$

Conditional independence example: CO_2 leakage

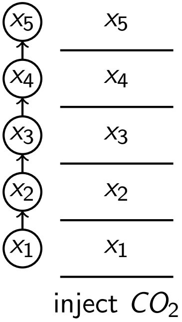
- Inject CO_2 in layered reservoir
- $x_i \in \{0, 1\}$: injected CO_2 present in layer i

- Modelling

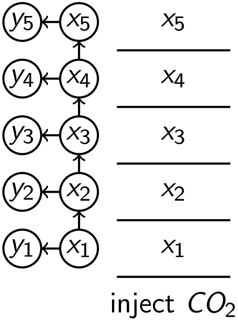
$$p(x_1, x_2, x_3, x_4, x_5)$$
$$= p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_3)p(x_5|x_4)$$

where

$$p(x_1 = 1) = p_1$$
$$p(x_i = 1|x_{i-1} = 0) = 0, \quad p(x_i = 1|x_{i-1} = 1) = p_i$$



Conditional independence example: CO_2 leakage



- Inject CO_2 in layered reservoir
- $x_i \in \{0, 1\}$: injected CO_2 present in layer i
- Modelling

  $$p(x_1, x_2, x_3, x_4, x_5)$$
  $$= p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_3)p(x_5|x_4)$$

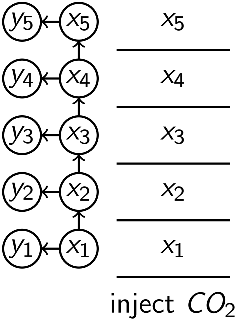
  where

  $$p(x_1 = 1) = p_1$$
  $$p(x_i = 1|x_{i-1} = 0) = 0, \quad p(x_i = 1|x_{i-1} = 1) = p_i$$

- Observation related to each layer, y_i

  $$p(y_i|x_i)$$

Conditional independence example: CO_2 leakage



- Inject CO_2 in layered reservoir
- $x_i \in \{0, 1\}$: injected CO_2 present in layer i

- Modelling

  $$p(x_1, x_2, x_3, x_4, x_5)$$
  $$= p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_3)p(x_5|x_4)$$

  where

  $$p(x_1 = 1) = p_1$$
  $$p(x_i = 1|x_{i-1} = 0) = 0, \quad p(x_i = 1|x_{i-1} = 1) = p_i$$

- Observation related to each layer, y_i

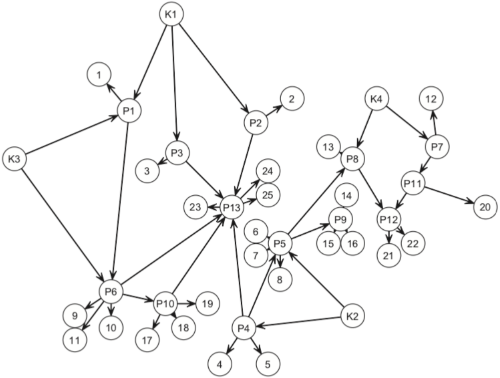
  $$p(y_i|x_i)$$

- Distributions of interest:

  $$p(x_i|y_1, y_2, y_3, y_4, y_5), i = 1, 2, 3, 4, 5$$

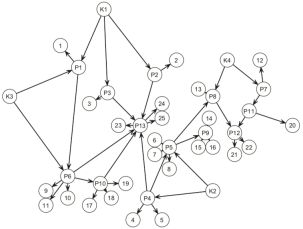
A larger graphical model example

- ⋆ Reference: Martinelli, G., Eidsvik, J. and Hauge, R. (2013). Dynamic decision making for graphical models applied to oil exploration, European Journal of Operational Research.



- ⋆ Three types of nodes: Areas which may have produced HC (K), macro-regions able to store HC (P), prospect nodes
- ⋆ A stochastic variable for each node, $x_k \in \{\text{dry}, \text{oil}, \text{gas}\}$

A larger graphical model example



- ⋆ Joint distribution

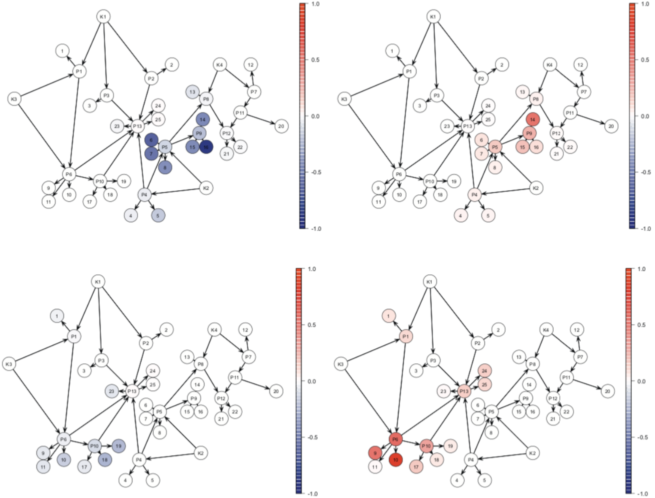
$$p(x) = \prod_k p(x_k | x_{\text{pa}(k)})$$

- ⋆ Number of possible values of x: $3^{42} \approx 10^{20}$
- ⋆ Cost of drilling, profit if finding oil/gas
- ⋆ In which prospects should we drill exploration wells?
  - sequential decisions
  - finding oil/gas gives profit
  - observations are important for later decision
  - assume we want to maximize profit
  - include discounting
- ⋆ Maximisation problem

$$\max_{i \in N} \left\{ \sum_{j=1}^{3} p(x_i = j) \left[r_i^j + \delta \max_{s \in N \setminus \{i\}} \left\{ \sum_{l=1}^{3} p(x_s = l | x_i = j) \left(r_s^l + \ldots \right), 0 \right\} \right], 0 \right\}$$

The effect of conditioing

- ⋆ The effect of drilling in prospect 14 (top) and 10 (bottom)
  - – observe dry (left) or oil (right)
- ⋆ Colours indicate $p(x_k = \text{oil}|x_{\text{obs}}) - p(x_k = \text{oil})$

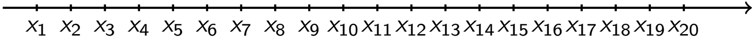


Plan for today

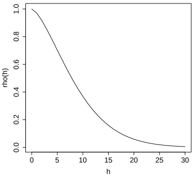
- ⋆ Interpretation of a distribution, $p(x)$
- ⋆ Modelling of noisy observations
- ⋆ Independence and conditional independence
- ⋆ Use of conditional independence for modelling of n variables
  - Markov chain, CO_2 leakage example
  - a larger network example
- ⋆ Hierarchical models
  - the effect of conditioning
- ⋆ How to look at n-dimensional distributions
  - what quantities are we interested in?

Hierarchical models and the effect of conditioning

- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable



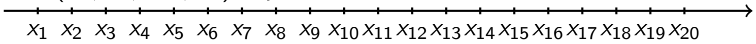
- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i-j|)$



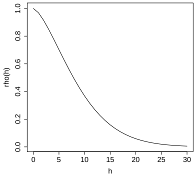
- ⋆ Assume $\mu = 0$, $\sigma^2 = 1$ and $\rho(h) = \exp\left\{-(h/10)^{1.5}\right\}$

Hierarchical models and the effect of conditioning

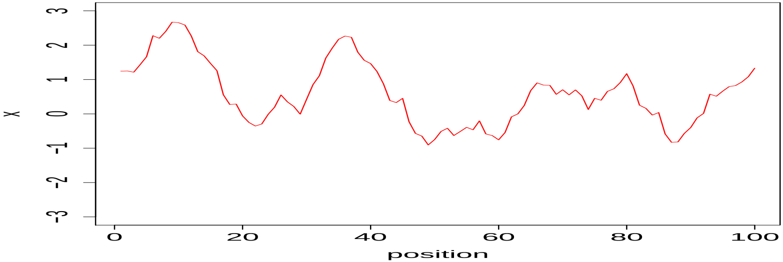
- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable



- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i - j|)$

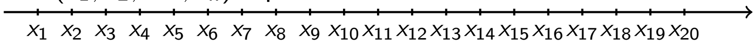


- ⋆ Assume $\mu = 0$, $\sigma^2 = 1$ and $\rho(h) = \exp\left\{-(h/10)^{1.5}\right\}$

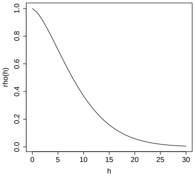


Hierarchical models and the effect of conditioning

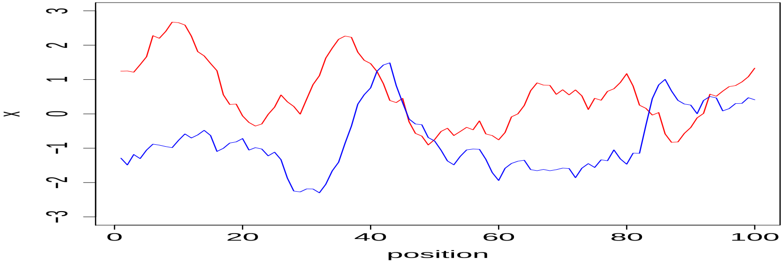
- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable



- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i - j|)$

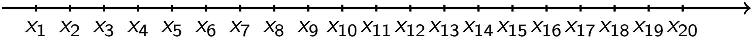


- ⋆ Assume $\mu = 0$, $\sigma^2 = 1$ and $\rho(h) = \exp\left\{-(h/10)^{1.5}\right\}$

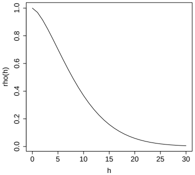


Hierarchical models and the effect of conditioning

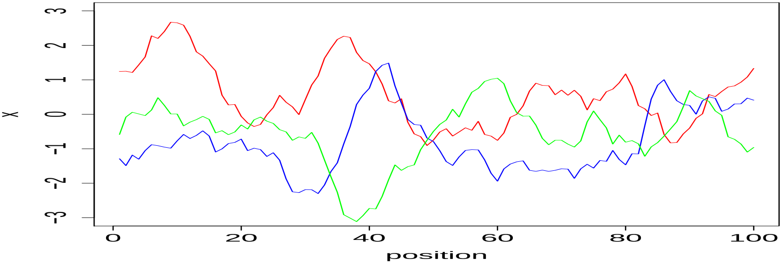
- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable



- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i-j|)$

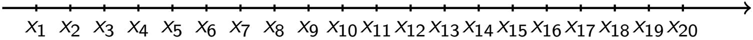


- ⋆ Assume $\mu = 0$, $\sigma^2 = 1$ and $\rho(h) = \exp\left\{-(h/10)^{1.5}\right\}$

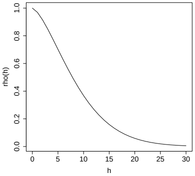


Hierarchical models and the effect of conditioning

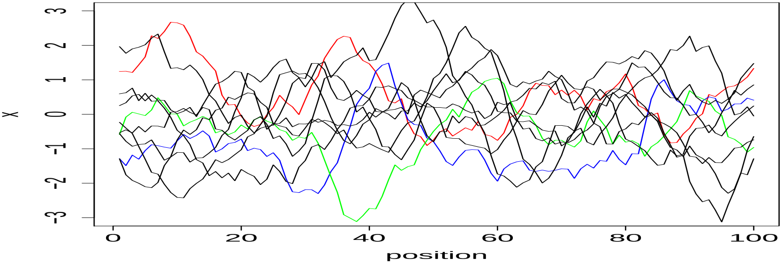
- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable



- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i - j|)$

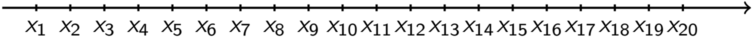


- ⋆ Assume $\mu = 0$, $\sigma^2 = 1$ and $\rho(h) = \exp\{-(h/10)^{1.5}\}$

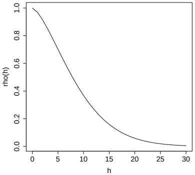


Graphical models and the effect of conditioning

- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable



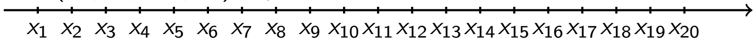
- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i-j|)$



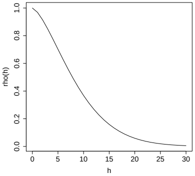
- ⋆ Assume $\mu = 0$, $\sigma^2 = 1$ and $\rho(h) = \exp\left\{-(h/10)^{1.5}\right\}$

Graphical models and the effect of conditioning

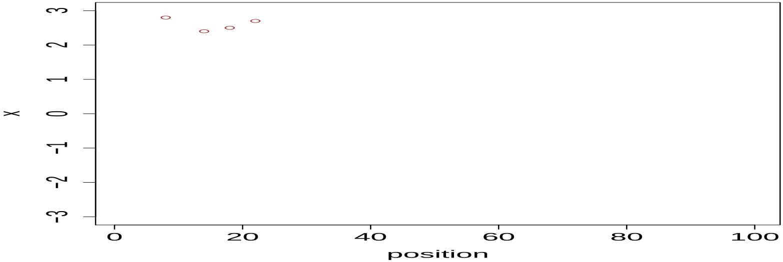
- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable



- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i - j|)$

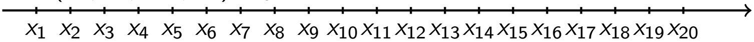


- ⋆ Assume $\mu = 0$, $\sigma^2 = 1$ and $\rho(h) = \exp\left\{-(h/10)^{1.5}\right\}$
- ⋆ Assume observed: $x_8 = 2.8$, $x_{14} = 2.4$, $x_{18} = 2.5$, $x_{22} = 2.7$

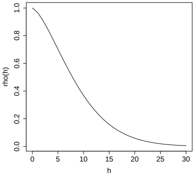


Graphical models and the effect of conditioning

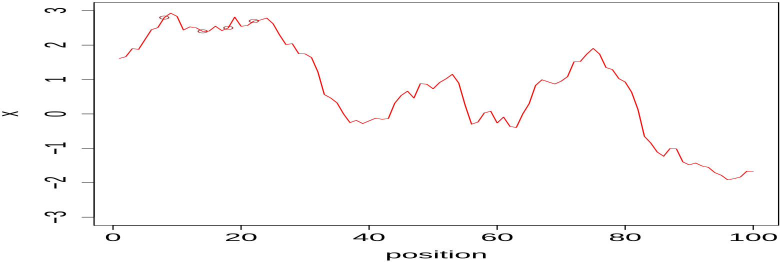
- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable



- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i - j|)$

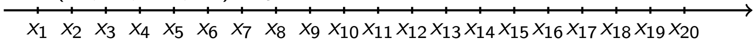


- ⋆ Assume $\mu = 0$, $\sigma^2 = 1$ and $\rho(h) = \exp\{-(h/10)^{1.5}\}$
- ⋆ Assume observed: $x_8 = 2.8$, $x_{14} = 2.4$, $x_{18} = 2.5$, $x_{22} = 2.7$

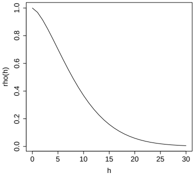


Graphical models and the effect of conditioning

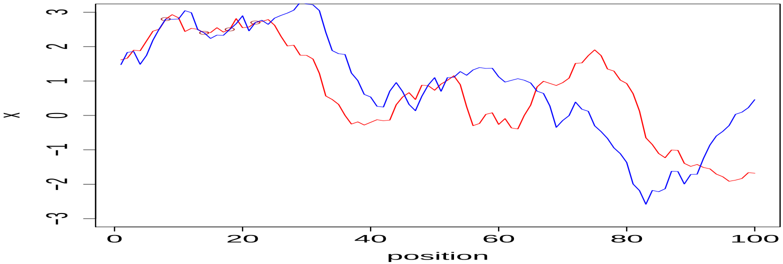
- ⋆ $x = (x_1, x_2, \dots, x_n)$: spatial variable



- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i-j|)$

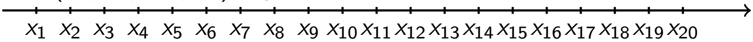


- ⋆ Assume $\mu = 0$, $\sigma^2 = 1$ and $\rho(h) = \exp\left\{-(h/10)^{1.5}\right\}$
- ⋆ Assume observed: $x_8 = 2.8$, $x_{14} = 2.4$, $x_{18} = 2.5$, $x_{22} = 2.7$

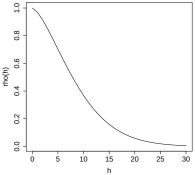


Graphical models and the effect of conditioning

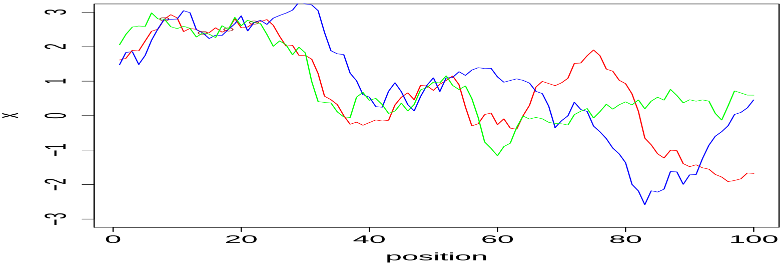
- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable



- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i - j|)$

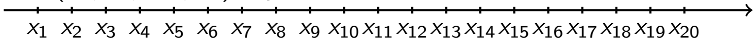


- ⋆ Assume $\mu = 0$, $\sigma^2 = 1$ and $\rho(h) = \exp\left\{-(h/10)^{1.5}\right\}$
- ⋆ Assume observed: $x_8 = 2.8$, $x_{14} = 2.4$, $x_{18} = 2.5$, $x_{22} = 2.7$

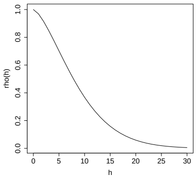


Graphical models and the effect of conditioning

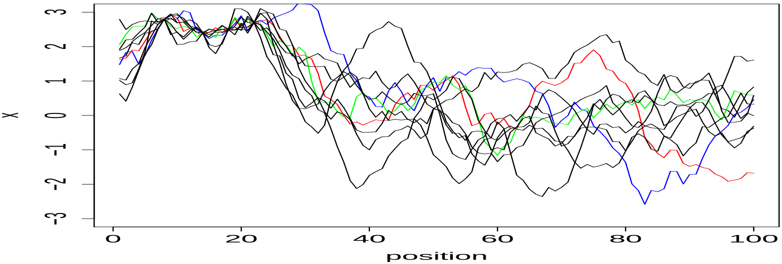
- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable



- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i - j|)$

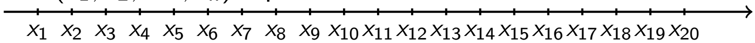


- ⋆ Assume $\mu = 0$, $\sigma^2 = 1$ and $\rho(h) = \exp\left\{-(h/10)^{1.5}\right\}$
- ⋆ Assume observed: $x_8 = 2.8$, $x_{14} = 2.4$, $x_{18} = 2.5$, $x_{22} = 2.7$

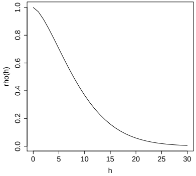


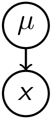
Graphical models and the effect of conditioning

- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable

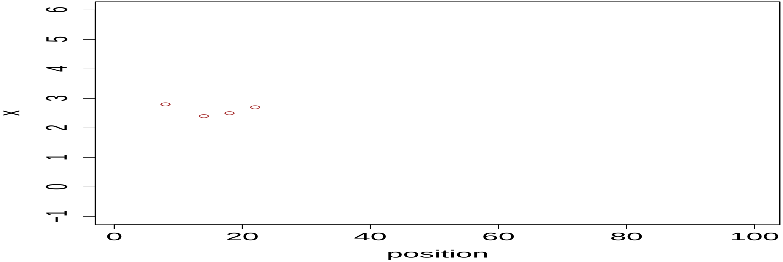


- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i - j|)$



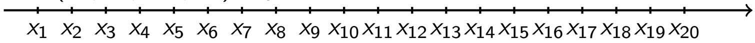


- ⋆ Assume $\mu \sim N(0, 5^2)$, $\sigma^2 = 1$ and $\rho(h) = \exp\{-(h/10)^{1.5}\}$
- ⋆ Assume observed: $x_8 = 2.8$, $x_{14} = 2.4$, $x_{18} = 2.5$, $x_{22} = 2.7$

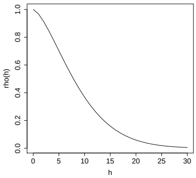


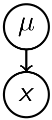
Graphical models and the effect of conditioning

- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable

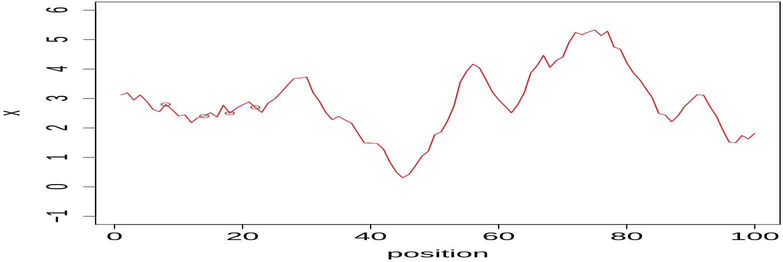


- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i - j|)$



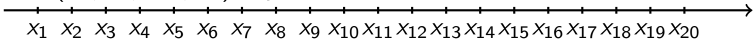


- ⋆ Assume $\mu \sim N(0, 5^2)$, $\sigma^2 = 1$ and $\rho(h) = \exp\left\{-(h/10)^{1.5}\right\}$
- ⋆ Assume observed: $x_8 = 2.8$, $x_{14} = 2.4$, $x_{18} = 2.5$, $x_{22} = 2.7$

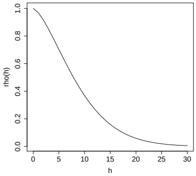


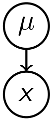
Graphical models and the effect of conditioning

- ⋆ $x = (x_1, x_2, \dots, x_n)$: spatial variable

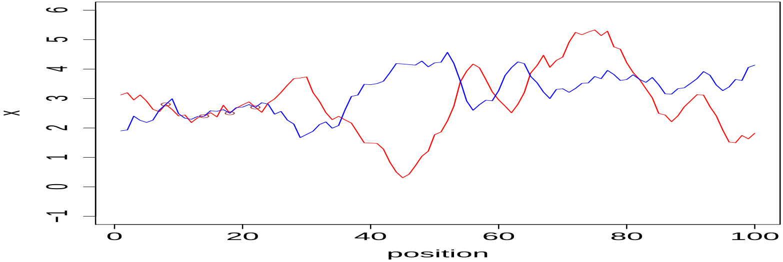


- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i - j|)$



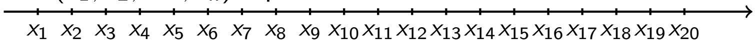


- ⋆ Assume $\mu \sim N(0, 5^2)$, $\sigma^2 = 1$ and $\rho(h) = \exp\left\{-(h/10)^{1.5}\right\}$
- ⋆ Assume observed: $x_8 = 2.8$, $x_{14} = 2.4$, $x_{18} = 2.5$, $x_{22} = 2.7$

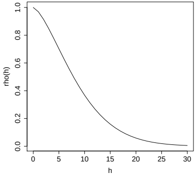


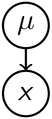
Graphical models and the effect of conditioning

- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable

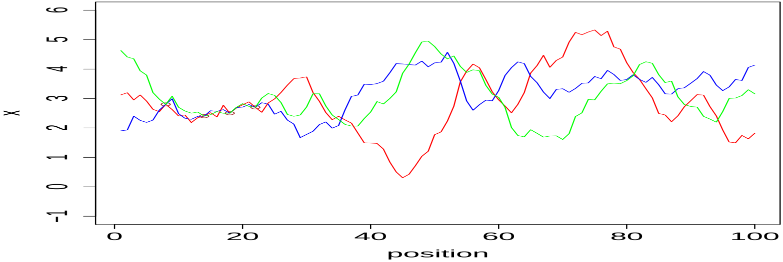


- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i - j|)$



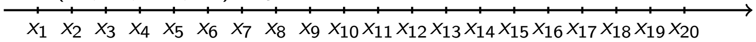


- ⋆ Assume $\mu \sim N(0, 5^2)$, $\sigma^2 = 1$ and $\rho(h) = \exp\{-(h/10)^{1.5}\}$
- ⋆ Assume observed: $x_8 = 2.8$, $x_{14} = 2.4$, $x_{18} = 2.5$, $x_{22} = 2.7$

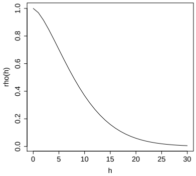


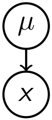
Graphical models and the effect of conditioning

- ⋆ $x = (x_1, x_2, \ldots, x_n)$: spatial variable

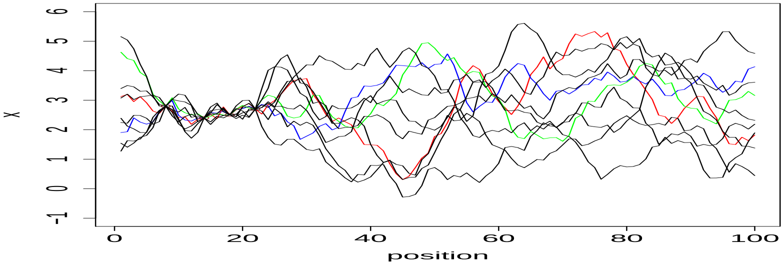


- ⋆ Assume x to be Gaussian
- ⋆ $\mathrm{E}[x_i] = \mu$
- ⋆ $\mathrm{Var}[x_i] = \sigma^2$
- ⋆ $\mathrm{Corr}[x_i, x_j] = \rho(|i - j|)$





- ⋆ Assume $\mu \sim N(0, 5^2)$, $\sigma^2 = 1$ and $\rho(h) = \exp\{-(h/10)^{1.5}\}$
- ⋆ Assume observed: $x_8 = 2.8$, $x_{14} = 2.4$, $x_{18} = 2.5$, $x_{22} = 2.7$

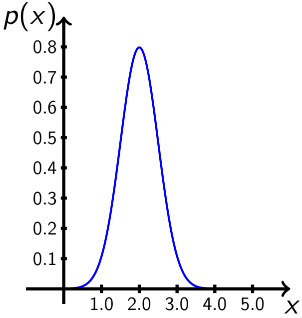


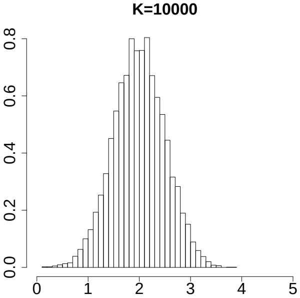
Plan for today

- ⋆ Interpretation of a distribution, $p(x)$
- ⋆ Modelling of noisy observations
- ⋆ Independence and conditional independence
- ⋆ Use of conditional independence for modelling of n variables
  - Markov chain, CO_2 leakage example
  - a larger network example
- ⋆ Hierarchical models
  - the effect of conditioning
- ⋆ How to look at n-dimensional distributions
  - what quantities are we interested in?

How to look at *n*-dimensional distributions

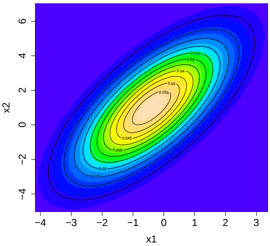
⋆ How to look at 1-dimensional distributions?

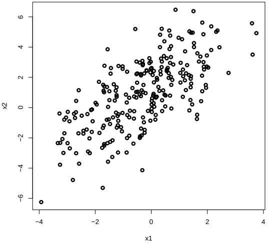




How to look at *n*-dimensional distributions

⋆ How to look at 2-dimensional distributions?





What quantities are we interested in?

- Assume stochastic variables, $x_1, \ldots, x_n, y_1, \ldots, y_m$
  - we have a formula for $p(x_1, \ldots, x_n, y_1, \ldots, y_m)$
  - we have observed $y_1, \ldots, y_m$
  - distribution of interest

$$p(x_1, \ldots, x_n | y_1, \ldots, y_m) = \frac{p(x_1, \ldots, x_n, y_1, \ldots, y_m)}{p(y_1, \ldots, y_m)}$$

where

$$p(y_1, \ldots, y_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1, \ldots, x_n, y_1, \ldots, y_m) dx_1 \cdots dx_n$$

What quantities are we interested in?

- Assume stochastic variables, $x_1, \ldots, x_n, y_1, \ldots, y_m$
  - we have a formula for $p(x_1, \ldots, x_n, y_1, \ldots, y_m)$
  - we have observed $y_1, \ldots, y_m$
  - distribution of interest

$$p(x_1, \ldots, x_n | y_1, \ldots, y_m) = \frac{p(x_1, \ldots, x_n, y_1, \ldots, y_m)}{p(y_1, \ldots, y_m)}$$

where

$$p(y_1, \ldots, y_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1, \ldots, x_n, y_1, \ldots, y_m) dx_1 \cdots dx_n$$

- We may be interested in
  - each x_i; $i = 1, \ldots, n$, i.e.

$$p(x_i | y_1, \ldots, y_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1, \ldots, x_n | y_1, \ldots, y_m) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n$$

  - each pair (x_i, x_j); $i, j = 1, \ldots, n$, i.e. $p(x_i, x_j | y_1, \ldots, y_m)$
  - some function of $x_1, \ldots, x_n$

$$z = g(x_1, \ldots, x_n)$$

What quantities are we interested in?

- If we cannot evaluate the integrals, we can do Monte Carlo sampling
- Assume we can do Monte Carlo sampling from

$$p(x_1, \ldots, x_n | y_1, \ldots, y_m) = \frac{p(x_1, \ldots, x_n, y_1, \ldots, y_m)}{p(y_1, \ldots, y_m)}$$

What quantities are we interested in?

- ⋆ If we cannot evaluate the integrals, we can do Monte Carlo sampling
- ⋆ Assume we can do Monte Carlo sampling from

$$p(x_1, \ldots, x_n | y_1, \ldots, y_m) = \frac{p(x_1, \ldots, x_n, y_1, \ldots, y_m)}{p(y_1, \ldots, y_m)}$$

- ⋆ If $(x_1, \ldots, x_n) \sim p(x_1, \ldots, x_n | y_1, \ldots, y_m)$ we have

What quantities are we interested in?

- ⋆ If we cannot evaluate the integrals, we can do Monte Carlo sampling
- ⋆ Assume we can do Monte Carlo sampling from

$$p(x_1, \ldots, x_n | y_1, \ldots, y_m) = \frac{p(x_1, \ldots, x_n, y_1, \ldots, y_m)}{p(y_1, \ldots, y_m)}$$

- ⋆ If $(x_1, \ldots, x_n) \sim p(x_1, \ldots, x_n | y_1, \ldots, y_m)$ we have
  - $x_i \sim p(x_i | y_1, \ldots, y_m)$

What quantities are we interested in?

- If we cannot evaluate the integrals, we can do Monte Carlo sampling
- Assume we can do Monte Carlo sampling from

$$p(x_1, \ldots, x_n | y_1, \ldots, y_m) = \frac{p(x_1, \ldots, x_n, y_1, \ldots, y_m)}{p(y_1, \ldots, y_m)}$$

- If $(x_1, \ldots, x_n) \sim p(x_1, \ldots, x_n | y_1, \ldots, y_m)$ we have
  - $x_i \sim p(x_i | y_1, \ldots, y_m)$
  - $(x_i, x_j) \sim p(x_i, x_j | y_1, \ldots, y_m)$

What quantities are we interested in?

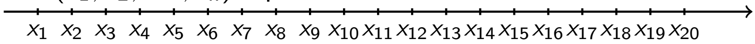
- ⋆ If we cannot evaluate the integrals, we can do Monte Carlo sampling
- ⋆ Assume we can do Monte Carlo sampling from

$$p(x_1, \ldots, x_n | y_1, \ldots, y_m) = \frac{p(x_1, \ldots, x_n, y_1, \ldots, y_m)}{p(y_1, \ldots, y_m)}$$

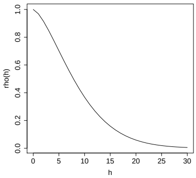
- ⋆ If $(x_1, \ldots, x_n) \sim p(x_1, \ldots, x_n | y_1, \ldots, y_m)$ we have
  - $x_i \sim p(x_i | y_1, \ldots, y_m)$
  - $(x_i, x_j) \sim p(x_i, x_j | y_1, \ldots, y_m)$
  - $z = g(x_1, \ldots, x_n) \sim p(z | y_1, \ldots, y_m)$

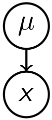
Gaussian example revisited

- $x = (x_1, x_2, \ldots, x_n)$: spatial variable

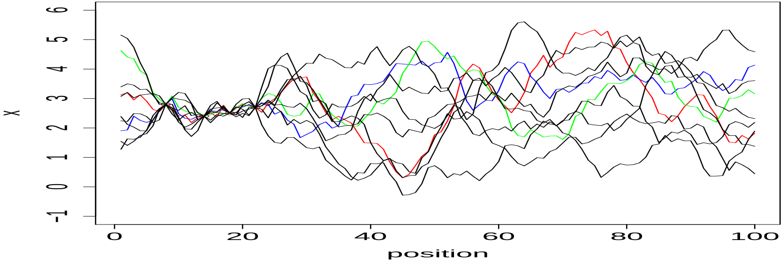


- Assume x to be Gaussian
- $\mathrm{E}[x_i] = \mu$
- $\mathrm{Var}[x_i] = \sigma^2$
- $\mathrm{Corr}[x_i, x_j] = \rho(|i - j|)$



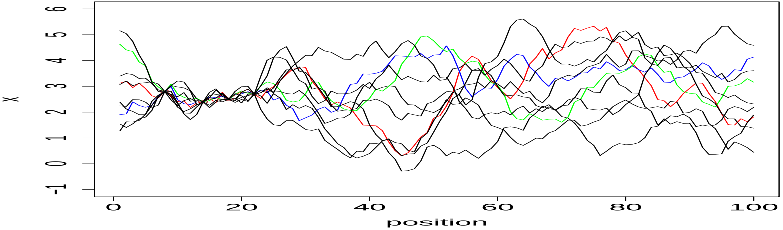


- Assume $\mu \sim N(0, 5^2)$, $\sigma^2 = 1$ and $\rho(h) = \exp\{-(h/10)^{1.5}\}$
- Assume observed: $x_8 = 2.8$, $x_{14} = 2.4$, $x_{18} = 2.5$, $x_{22} = 2.7$



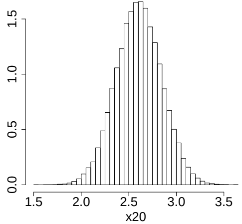
Marginal distributions

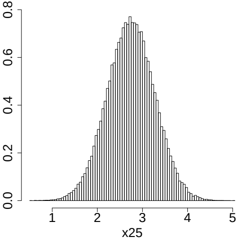
- ⋆ Ten realisations

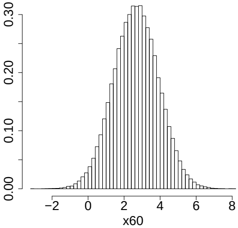


- ⋆ Marginal distributions for x_{20}, x_{25} and x_{60}

$p(x_{20}|x_8, x_{14}, x_{18}, x_{22})$ $\quad$ $p(x_{25}|x_8, x_{14}, x_{18}, x_{22})$ $\quad$ $p(x_{60}|x_8, x_{14}, x_{18}, x_{22})$

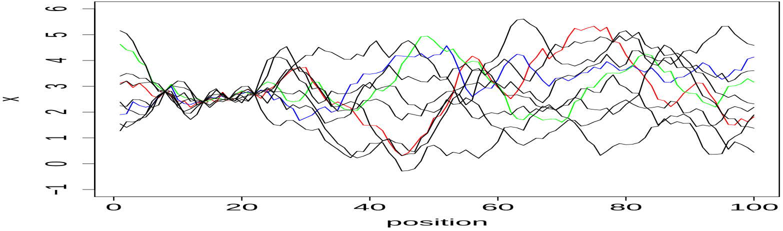




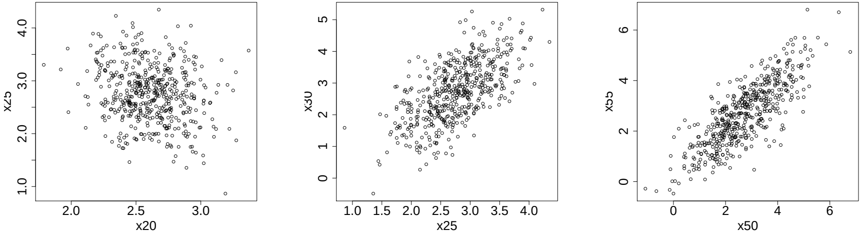


Bivariate distributions

⋆ Ten realisations

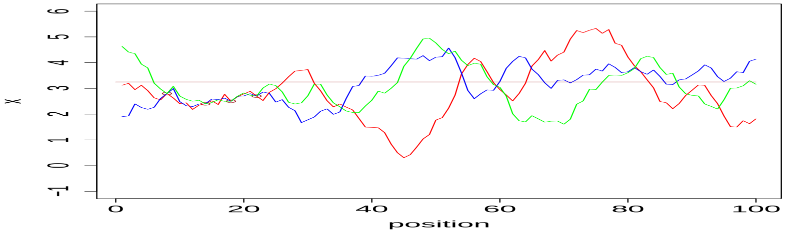


⋆ Some bivariate distributions



A non-linear function of $x_1, \ldots, x_n$

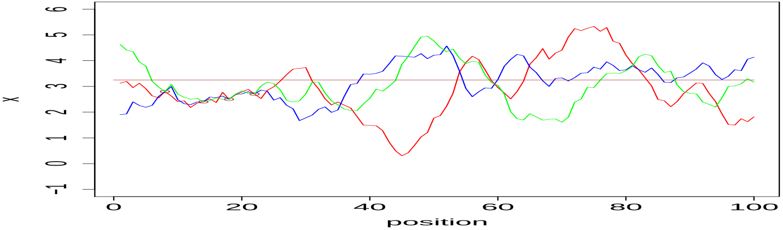
- ⋆ Three realisations



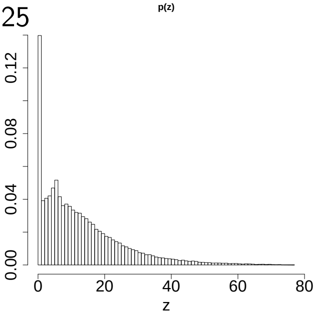
- ⋆ Let $z = g(x_1, \ldots, x_n)$ be the length of the longest continuous interval where $x_i \geq 3.25$
- ⋆ Distribution of z:

A non-linear function of $x_1, \ldots, x_n$

- ⋆ Three realisations



- ⋆ Let $z = g(x_1, \ldots, x_n)$ be the length of the longest continuous interval where $x_i \geq 3.25$
- ⋆ Distribution of z:



Plan for next time

- ⋆ Part I: Introduction, stochastic variables and the effect of conditioning
- ⋆ Part II: Modelling of dependence, conditional independence
- ⋆ Part III: Bayesian inversion, prior and posterior distribution
- ⋆ Part IV: Spatial model for categorical variables, Markov chain Monte Carlo
- ⋆ Part V: Dynamic state space models, Kalman and ensemble Kalman filters