COMP 7180: Quantitative Methods for Data Analytics and Artificial Intelligence

Lecture 10 – Basic Probability Distributions and Bayesian Inference

Lecturer:

Dr. Yang Liu, Dr. Jun Qi

New Announcement

- Homework 2 has been released, and the deadline is 30 November.
- Quiz 2 will be on 26 November (1.5 Hours, 18:45 – 20:15). The test covers the convex optimization and probability topics.
- The final exam will be on 12 December (19:00 -- 22:00).
- Our last class will be on 3 December.

Updates on Homework 2 (1/2)

- For Problem 1 (c), we provide the hint that

  The affine transformation preserves the convexity.

  $f = g\big(h(x)\big)$ is a convex function:

  - if *h is convex and nondecreasing, and g is convex.*
  - *if h is convex and nonincreasing, and g is concave.*

Proof: $f''(x) = h''\big(g(x)\big)g'(x)^2 + h'\big(g(x)\big)g''(x)$

Please check the conditions.

Updates on Homework 2 (2/2)

- For the function $f(x) = \frac{1}{2}x^2$, we compute its conjugate function based on the definition:

$$f^*(y) = \sup_{x \in \mathbb{R}} \{x\,y - \frac{1}{2}x^2\}$$

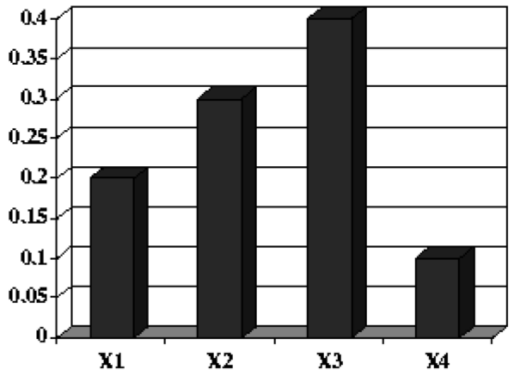
which is an unconstrained convex optimization problem. So, you could find the optimal x by taking the first-order gradient.

Discrete Variable – Recall

$$X \in \{x_1, x_2, \ldots, x_n\}$$

$$P(x_i) \geq 0$$

$$\sum_i P(x_i) = 1$$

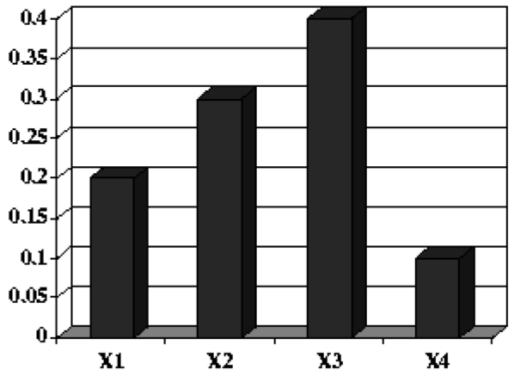


From Discrete to Continuous

$X \in \{x_1, x_2, \ldots, x_n\}$

$P(x_i) \geq 0$

$\sum_i P(x_i) = 1$

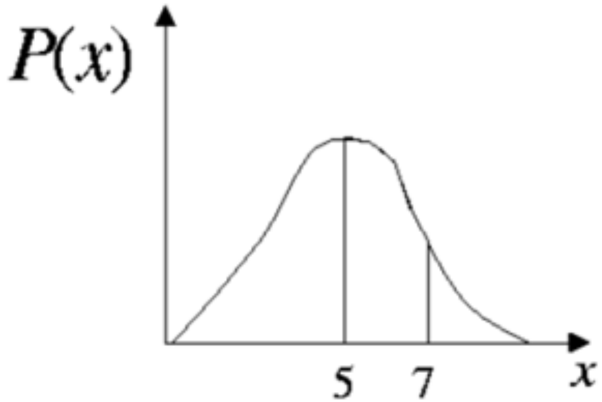


Probability Distribution Function (PDF)

$X \in [0,10]$

$P(x) \geq 0$

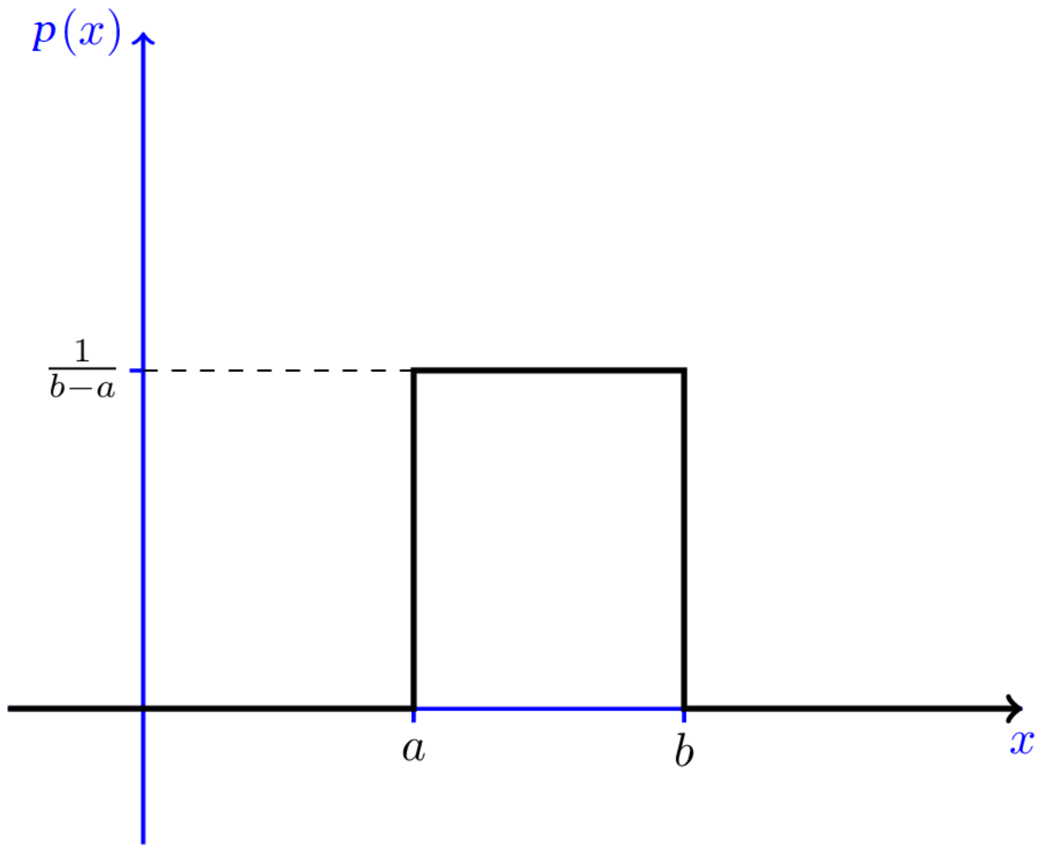
$$\int_0^{10} P(x)\,dx = 1$$



An Example of Continuous Variable

- Uniform Distribution:

If x is *uniformly distributed* over the interval from a to b, i.e. $[a, b]$, then the value of the Probability Distribution Function (PDF) of x is a constant $1/(b-a)$ in the interval from a to b.



An Exercise

- Let X be a continuous random variable with the following PDF:

$$p(x) = \begin{cases} ce^{-x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Find c.

An Exercise

- Let X be a continuous random variable with the following PDF:

$$p(x) = \begin{cases} ce^{-x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Find c.

Solution: To find c, we can use the property in page 3:

$$1 = \int_0^\infty p(x)dx = \int_0^\infty ce^{-x}dx = c\left[-e^{-x}\right]_0^\infty = c[(0)-(-1)] = c$$

Expectation (or Mean)

- The *expectation* or *mean* of a random variable X is denoted by $E[X]$ and defined as:

$$E[X] = \sum_{x \in X} xp(x) \quad \text{for discrete } x \text{ ;}$$

and

$$E[X] = \int_{-\infty}^{+\infty} xp(x)dx \quad \text{for continuous } x \text{ .}$$

- In other n words, we are taking a weighted sum of the values that x can take on, where the weights are the probabilities of those values. The expected value has a physical interpretation as the “center of mass” of the distribution.

Expectation (or Mean)

- A simple example:

  Assume that X is a discrete random variable with 4 possible values: x_1, x_2, x_3, x_4, and with equal probability on these 4 values. What is the expectation of x?

  Solution: According to the definition, we have

$$E[X] = \sum_{x \in X} x p(x)$$

$$= \frac{1}{4}x_1 + \frac{1}{4}x_2 + \frac{1}{4}x_3 + \frac{1}{4}x_4$$

$$= \frac{x_1+x_2+x_3+x_4}{4}$$

Expectation (or Mean)

- An exercise:

Let us consider an European roulette game that players can bet on any single number from 1, 2, …, 36. The number 0 is considered as winning for the casino. The probability of the appearance of any number from 0-36 is equal, i.e., 1/37.

Bob places 1 dollar on a specific number. If he wins, he can take the 1 dollar back and gains the extra 35 dollars from the casino; if he loses, he will lose the 1 dollar he bet on that number.

Now, the question is: what is the expectation of the gain of Bob?

Expectation (or Mean)

- An exercise:

Let us consider a European roulette game that players can bet on any single number from 1, 2, …, 36. The number 0 is considered as winning for the casino. The probability of the appearance of any number from 0-36 is equal, i.e., 1/37.

Bob places 1 dollar on a specific number. If he wins, he can take the 1 dollar back and gains the extra 35 dollars from the casino; if he loses, he will lose the 1 dollar he bet on that number.

Now, the question is: what is the expectation of the gain of Bob?

Solution: $E[X] = x_1 p(x_1) + x_2 p(x_2) =$
$(35)\times(1/37) + (-1)\times(36/37) = -1/37$

Expectation of Functions

- The *expectation* or *mean* of $f(X)$ (a function of random variable x) is denoted by $E[f(X)]$ and defined as:

$$E[f(X)] = \sum_{x \in X} f(x)p(x) \quad \text{for discrete } x\text{ ;}$$

and

$$E[f(X)] = \int_{-\infty}^{+\infty} f(x)p(x)dx \quad \text{for continuous } x\text{ .}$$

Properties of Expectation

- For any two random variables x and y and any constants a, $b \in R$, the following equations hold:

$$E[a] = a \tag{1}$$

$$E[aX] = aE[X] \tag{2}$$

$$E[X + Y] = E[X] + E[Y] \tag{3}$$

$$E[aX + bY] = aE[X] + bE[Y] \tag{4}$$

Properties of Expectation

- For any two random variables x and y and any constants a, $b \in R$, the following equations hold:

$$E[a] = a \quad (1)$$

Why? From the definition of expectation, we have

$E[a] = \sum_{a \in A} ap(a) = a \times 1 = a.$

So, we have: the expectation of a constant is the constant itself.

Properties of Expectation

- For any two random variables x and y and any constants a, $b \in R$, the following equations hold:

$$E[aX] = aE[X] \tag{2}$$

Why? From the definition of expectation, we have

$E[aX] = \sum_{x \in X} axp(x) = a \sum_{x \in X} xp(x) = aE[X].$

Properties of Expectation

- For any two random variables x and y and any constants $a, b \in R$, the following equations hold:

$$E[X + Y] = E[X] + E[Y] \quad (3)$$

Why? From the definition of expectation, we have

$$E[X + Y] = \sum_{x \in X} \sum_{y \in Y} (x + y) p(x, y)$$

$$= \sum_{x \in X} \sum_{y \in Y} x p(x, y) + \sum_{x \in X} \sum_{y \in Y} y p(x, y)$$

$$= \sum_{x \in X} x (\sum_{y \in Y} p(x, y)) + \sum_{y \in Y} y (\sum_{x \in X} p(x, y)$$

$$= \sum_{x \in X} x p(x) + \sum_{y \in Y} y p(y) = E[X] + E[Y]$$

Properties of Expectation

- For any two random variables X and Y and any constants $a, b \in \mathbb{R}$, the following equations hold:

$$E[aX + bY] = aE[X] + bE[Y] \tag{4}$$

Why? From the proof of (1) and (2), we have

$E[aX + bY] = E[aX] + E[bY] = aE[X] + bE[Y]$.

Properties of Expectation

- If two random variables x and y are independent, then the following identity holds:

$$E[XY] = E[X]E[Y]$$

Why? From the definition of expectation, we have

$E[XY] = \sum_{x \in X} \sum_{y \in Y} (xy) p(x, y)$.

Since x and y are independent, we know that $p(x, y) = p(x)p(y)$. Therefore, we have

$$E[XY] = \sum_{x \in X} \sum_{y \in Y} (xy) p(x) p(y)$$

$$= \left(\sum_{x \in X} x p(x)\right) \left(\sum_{y \in Y} y p(y)\right) = E[X]E[Y].$$

Variance

- Expectation provides a measure of the "center" of a distribution, but sometimes, we are also interested in what the "spread" is about that center. Therefore, we define the variance $Var(X)$ of a random variable x as follows:

$$Var(X) = E[(X - E(X))^2]$$

- In other words, this is the average squared deviation of the values of X from the mean of X.

Properties of Variance

- For any random variable X and any a, $b \in \mathbb{R}$, the following equations hold:

$$Var(X) = E[X^2] - E[X]^2 \qquad (1)$$

$$Var(aX+b) = a^2 Var(X) \qquad (2)$$

Properties of Variance

- For any random variable X and any $a, b \in \mathbb{R}$, the following equations hold:

$$Var(X) = E[X^2] - E[X]^2 \qquad (1)$$

Why? From the definition of variance, we have

$$Var(X) = E[(X - E(X))^2] = E[X^2 - 2XE(X) + E(X)^2]$$

$$= E[X^2] - E[2XE(X)] + E[E(X)^2].$$

Note that $E(X)$ and $E(X)^2$ are constants, so that

$$Var(X) = E[X^2] - 2E(X)^2 + E(X)^2 = E[X^2] - E[X]^2.$$

Properties of Variance

- For any random variable x and any $a, b \in R$, the following equations hold:

$$Var(aX+b) = a^2 Var(X) \qquad (2)$$

Why? From (1), we have

$$Var(aX+b) = E[(aX+b)^2] - E[aX+b]^2$$

$$= E[a^2X^2 + 2abX + b^2] - (aE[X]+b)^2$$

$$= (a^2E[X^2] + 2abE[X] + b^2) - (a^2E[X]^2 + 2abE[X] + b^2)$$

$$= a^2E[X^2] - a^2E[X]^2 = a^2 (E[X^2] - E[X]^2) = a^2 Var(X)$$

Properties of Variance

- If two random variables X and Y are independent, then

$$Var(X + Y) = Var(X) + Var(Y)$$

Why? Using the linearity of the expectation and the identity $E(XY) = E(X)E(Y)$, which holds by the independence of x and y, we can write

$$Var(X+Y) = E[(X+Y)^2] - (E[X+Y])^2$$

$$= E[X^2+2XY+Y^2] - (E[X]^2+2E[X]E[Y]+E[Y]^2)$$

$$= (E[X^2] - E[X]^2) + (E[2XY] - 2E[X]E[Y]) + (E[Y^2] - E[Y]^2)$$

$$= \quad Var(x) \quad + \quad 0 \quad + \quad Var(y)$$

Covariance

- The covariance of two random variables X and Y is denoted by $Cov(X, Y)$ and defined as

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] \; .$$

- We have

$$Cov(X, X) = Var(X)$$

Properties of Covariance

- If two random variables X and Y are independent, then

$$Cov(X, Y) = 0 .$$

Why? From the definition of covariance, we have

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$$

$$= E[XY - E(X)Y - XE(Y) + E(X)E(Y)]$$

$$= E[XY] - E[E(X)Y] - E[XE(Y)] + E[E(X)E(Y)]$$

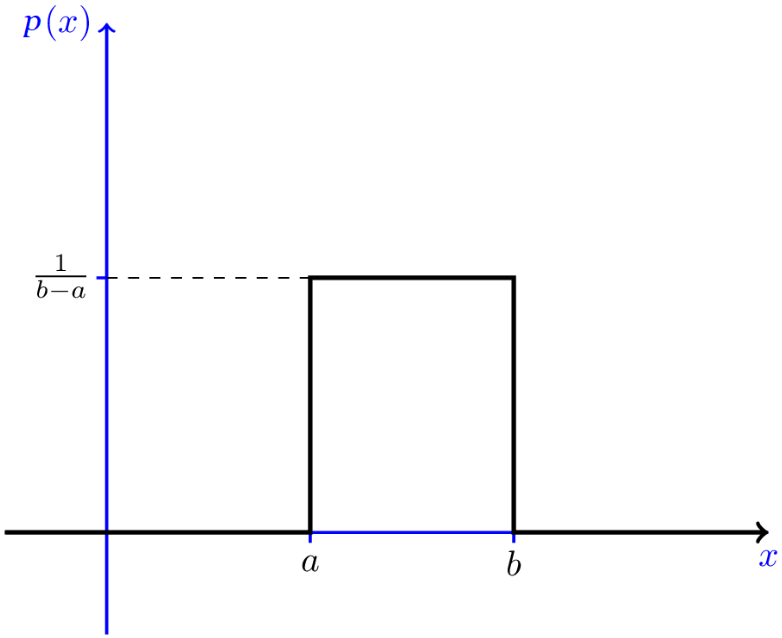
$$= E[XY] - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$$

$$= E[XY] - E(X)E(Y) = 0.$$

Basic Probability Distributions

- **Uniform Distribution**: Given a random variable X, we call its distribution a uniform distribution if it assigns equal probability mass in a region. For any real numbers a, b , and $a < b$, the uniform distribution is defined as:

$$p(x|a,b) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & otherwise \end{cases}$$



- The expectation and variance of the uniform distribution are

$$E(X) = \frac{a+b}{2}, \quad Var(X) = \frac{(b-a)^2}{12}$$

Basic Probability Distributions

- Exercise: Show that the expectation and variance of the uniform distribution are

$$E(X) = \frac{a+b}{2}, \quad Var(X) = \frac{(b-a)^2}{12}$$

From the definition, we have

$$E(X) = \int_a^b xp(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\int_a^b xdx$$

$$= \frac{1}{b-a} \times \frac{1}{2}x^2|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

Basic Probability Distributions

- Exercise: Show that the expectation and variance of the uniform distribution are

$$E(X) = \frac{a+b}{2}, \quad Var(X) = \frac{(b-a)^2}{12}$$

From the property of variance and expectation, we have

$$Var(X) = E[X^2] - E[X]^2 = \int_a^b x^2 p(x)dx - (\int_a^b xp(x)dx)^2$$

$$= \int_a^b x^2 \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \frac{1}{b-a} \int_a^b x^2 dx - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{1}{b-a} \times \frac{1}{3} x^3 |_a^b - \left(\frac{a+b}{2}\right)^2 = \frac{b^3 - a^3}{3(b-a)} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

Basic Probability Distributions

- **Bernoulli Distribution**: It is a distribution for a single binary variable X∈ {0,1}, and it is governed by a single continuous parameter $\mu \in [0,1]$ that represents the probability of X= 1. The Bernoulli distribution is defined as

$$p(x|\mu) = \mu^{x}(1-\mu)^{1-x}, x \in \{0,1\}$$

- The expectation and variance of the Bernoulli distribution are

$$E(X) = \mu, \qquad Var(X) = \mu(1-\mu)$$

Binary outcome probability of a coin flip experiment

Basic Probability Distributions

- **Binomial Distribution**: It is a generalization of the Bernoulli distribution to a distribution over integers. It can be used to describe the probability of observing m occurrences of X= 1 in a set of N samples from a Bernoulli distribution where $\mathrm{p}(X = 1) = \mu \in [0,1]$. The Binomial distribution is defined as

$$p(m|N, \mu) = \binom{N}{m} \mu^m (1 - \mu)^{N-m}$$

- The expectation and variance of the binomial distribution are

$$E(X) = N\mu, \qquad Var(X) = N\mu(1 - \mu)$$

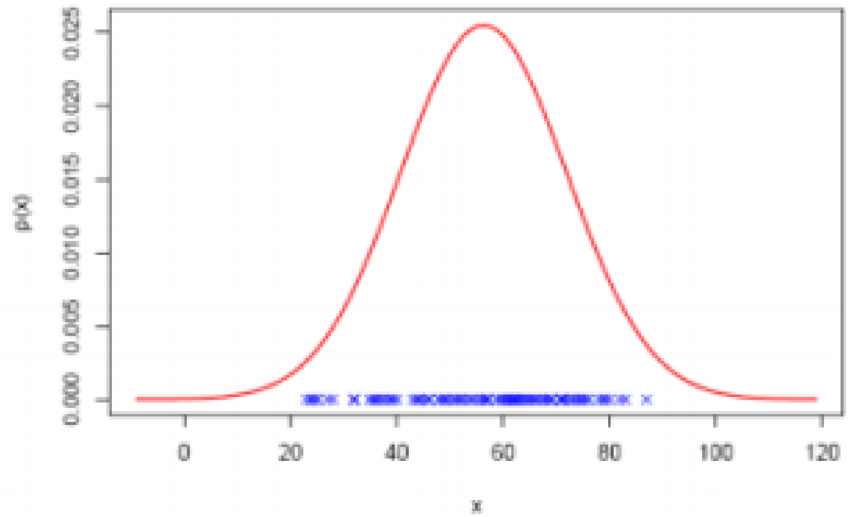
Basic Probability Distributions

- **Gaussian Distribution (Normal Distribution)**: It is the most widely used model for the distribution of continuous variables. For a single variable x, the Gaussian distribution can be written as

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

- The expectation and variance of the Gaussian distribution are

$$E(X) = \mu, \quad Var(X) = \sigma^2$$



Bayesian Learning

Bayesian Theorem

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

$$P(A|B)\,P(B) = P(A \cap B) = P(B|A)\,P(A).$$

$$\boxed{P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}}.$$

Bayes' Rule

$$\overbrace{P(C \mid \boldsymbol{x})}^{\textit{posterior}} = \frac{\overbrace{P(C)}^{\textit{prior}}\,\overbrace{p(\boldsymbol{x} \mid C)}^{\textit{likelihood}}}{\underbrace{p(\boldsymbol{x})}_{\textit{evidence}}}$$

$$P(C=0)+P(C=1)=1$$
$$p(\boldsymbol{x})=p(\boldsymbol{x} \mid C=1)P(C=1)+p(\boldsymbol{x} \mid C=0)P(C=0)$$
$$p(C=0 \mid \boldsymbol{x})+P(C=1 \mid \boldsymbol{x})=1$$

An Example

Does patient have cancer or not?

A patient takes a lab test and the result comes back positive. The test returns a correct positive result in only 98% of the cases in which the disease is actually present, and a correct negative result in only 97% of the cases in which the disease is not present. Furthermore, .008 of the entire population have this cancer.

$$P(cancer) = .008, P(\neg cancer) = .992$$

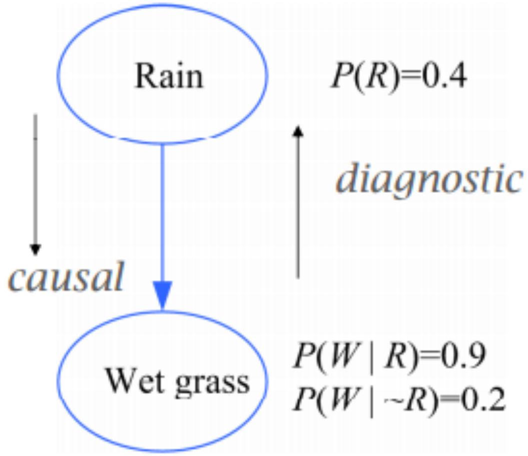
$$P(+ \mid cancer) = .98, P(- \mid cancer) = .02$$

$$P(+ \mid \neg cancer) = .03, P(- \mid \neg cancer) = .97$$

$$P(cancer \mid +) = \frac{P(+ \mid cancer) P(cancer)}{P(+)}$$

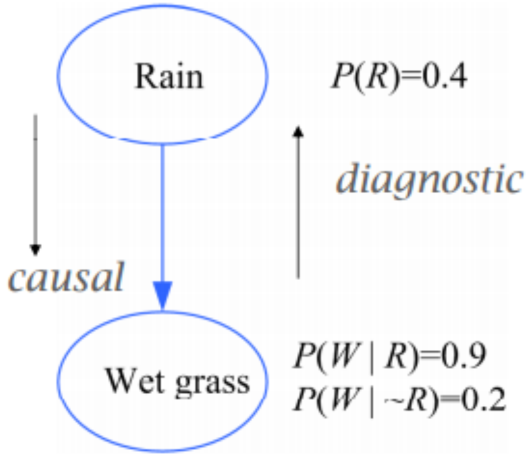
$$P(\neg cancer \mid +) = \frac{P(+ \mid \neg cancer) P(\neg cancer)}{P(+)}$$

Exercise 1: Diagnostic inference



Diagnostic inference:
Knowing that the grass is wet, what is the probability that rain is the cause?

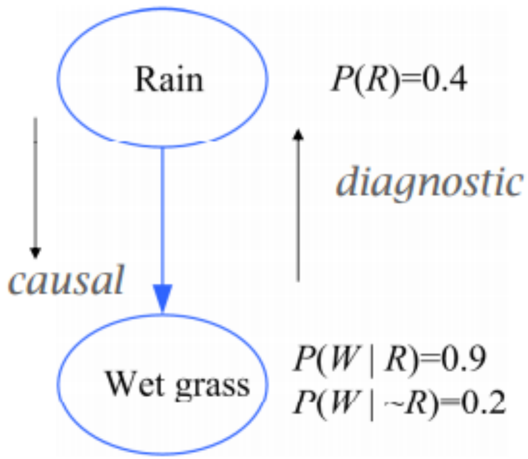
Exercise 1: Diagnostic inference



Diagnostic inference:
Knowing that the grass is wet, what is the probability that rain is the cause?

$$P(R \mid W)$$

Exercise 1: Diagnostic inference



Diagnostic inference:
Knowing that the grass is wet, what is the probability that rain is the cause?

$$P(R \mid W) = \frac{P(W \mid R)P(R)}{P(W)}$$

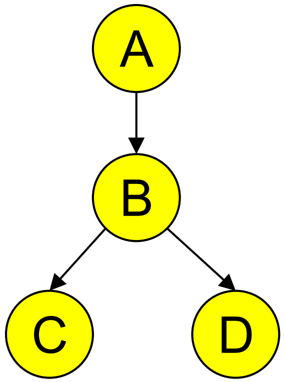
$$= \frac{P(W \mid R)P(R)}{P(W \mid R)P(R) + P(W \mid \sim R)P(\sim R)}$$

$$= \frac{0.9 \times 0.4}{0.9 \times 0.4 + 0.2 \times 0.6} = 0.75$$

A Bayesian Network

A Bayesian network is made up of:

1. A Directed Acyclic Graph



2. A set of tables for each node in the graph

A	P(A)
false	0.6
true	0.4

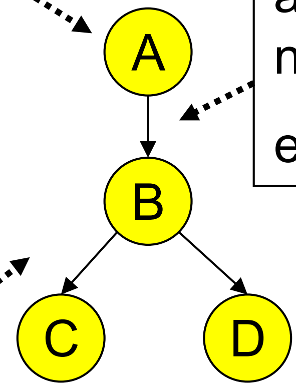
A	B	P(B|A)
false	false	0.01
false	true	0.99
true	false	0.7
true	true	0.3

B	D	P(D|B)
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

A Directed Acyclic Graph

Each node in the graph is a random variable



A node *X* is a parent of another node *Y* if there is an arrow from node *X* to node *Y*

e.g. *A* is a parent of *B*

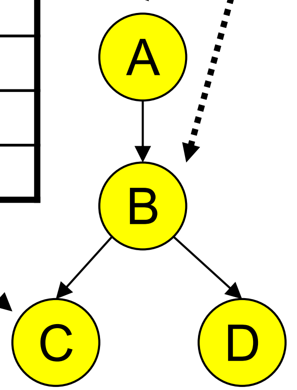
Informally, an arrow from node *X* to node *Y* means *X* has a direct influence on *Y*

A Set of Tables for Each Node

A	P(A)
false	0.6
true	0.4

A	B	P(B|A)
false	false	0.01
false	true	0.99
true	false	0.7
true	true	0.3

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1



B	D	P(D|B)
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95

Each node X_i has a conditional probability distribution P(X_i | Parents(X_i)) that quantifies the effect of the parents on the node

The parameters are the probabilities in these conditional probability tables (CPTs)

A Set of Tables for Each Node

Conditional Probability Distribution for C given B

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

For a given combination of values of the parents (B in this example), the entries for P(C=true | B) and P(C=false | B) must add up to 1
e.g. P(C=true | B=false) + P(C=false |B=false)=1

If you have a Boolean variable with k Boolean parents, this table has 2^{k+1} probabilities (but only 2^k need to be stored)

The Joint Probability Distribution

The joint probability distribution over all the variables X_1, …, X_n in the Bayesian network can be calculated using the formula:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i \mid Parents(X_i))$$

Where Parents(X_i) means the values of the Parents of the node X_i with respect to the graph

Using a Bayesian Network Example

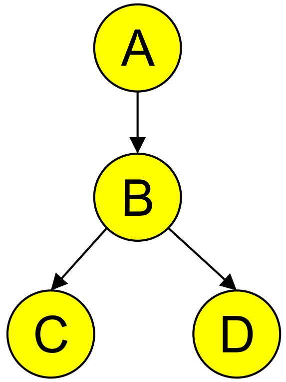
Using the network in the example, suppose you want to calculate:

$P(A = true, B = true, C = true, D = true)$

$= P(A = true) * P(B = true \mid A = true) *$

$P(C = true \mid B = true) * P(D = true \mid B = true)$

$= (0.4)*(0.3)*(0.1)*(0.95)$



A	P(A)
false	0.6
true	0.4

A	B	P(B|A)
false	false	0.01
false	true	0.99
true	false	0.7
true	true	0.3

B	D	P(D|B)
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

Using a Bayesian Network Example

Using the network in the example, suppose you want to calculate:

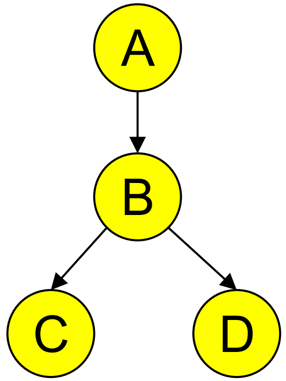
$$P(A = true, B = true, C = true, D = true)$$

$$= P(A = true) * P(B = true \mid A = true) * P(C = true \mid B = true) * P(D = true \mid B = true)$$

$$= (0.4)*(0.3)*(0.1)*(0.95)$$

This is from the graph structure

These numbers are from the conditional probability tables

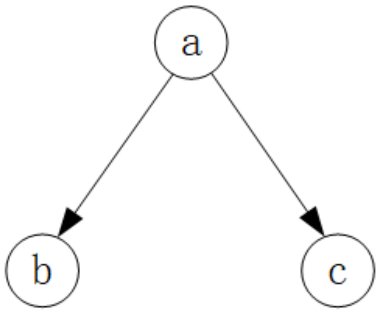


Three Important Structures of BN

Chain Rule: $$P(A_1, \dots, A_n) = \Pi_{i=1}^{n} P(A_i | A_1, \dots, A_{i-1})$$

Joint Probability: $$p(x_1, x_2, x_3, \dots, x_p) = \prod_{i=1}^{p} p(x_i | x_{pa(i)})$$

- Structure 1: Tail to Tail



Chain: $p(a,b,c) = p(a)*p(b|a)*p(c|a,b)$

Joint: $p(a,b,c) = p(a)*p(b|a)*p(c|a)$

$p(c|a) = p(c|a,b)$

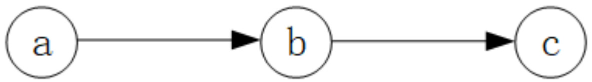
b and c are **conditional independent** given a

Three Important Structures of BN

Chain Rule: $P(A_1, \ldots, A_n) = \Pi_{i=1}^{n} P(A_i | A_1, \ldots, A_{i-1})$

Joint Probability: $p(x_1, x_2, x_3, \ldots, x_p) = \prod_{i=1}^{p} p(x_i | x_{pa(i)})$

- Structure 2: Head to Tail



Chain: *p(a,b,c) = p(a)*p(b|a)*p(c|a,b)*
Joint: *p(a,b,c)=p(a)*p(b|a)*p(c|b)*

p(c|b) = p(c|a,b)

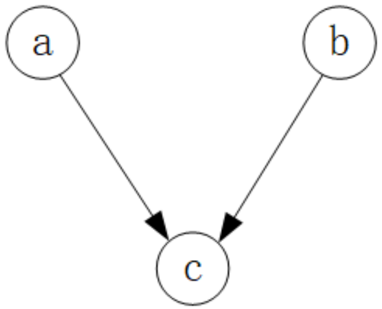
a and c are **conditional independent** given b

Three Important Structures of BN

Chain Rule: $P(A_1, \ldots, A_n) = \Pi_{i=1}^{n} P(A_i | A_1, \ldots, A_{i-1})$

Joint Probability: $p(x_1, x_2, x_3, \ldots, x_p) = \prod_{i=1}^{p} p(x_i | x_{pa(i)})$

- Structure 3: Head to Head



Chain: $p(a,b,c) = p(a)*p(b|a)*p(c|a,b)$
Joint: $p(a,b,c) = p(a)*p(b)*p(c|a,b)$

$p(b|a) = p(b)$

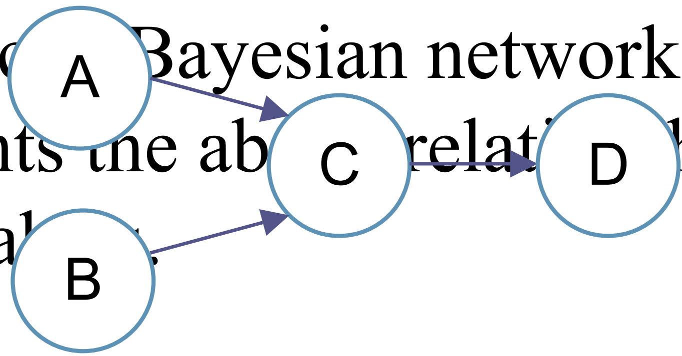
a and b are **independent**

Exercises

- For the probability distribution P(A, B, C, D), if the following equation holds

$$P(A, B, C, D) = P(A) * P(B) * P(C|A, B) * P(D|C),$$

Construc Bayesian network which represents the ab relati hip among the varia .



Exercises

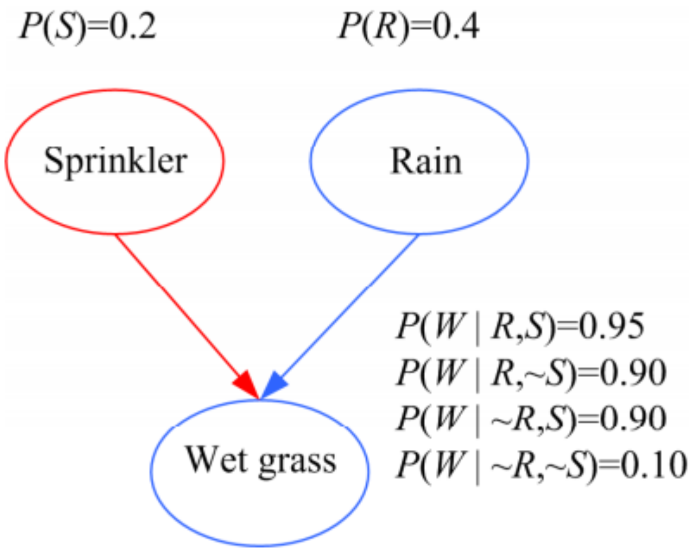
- For the probability distribution P(A, B, C, D), if the following equation holds

$$P(A, B, C, D) = P(A) * P(B) * P(C|A, B) * P(D|C),$$

which of the following statements is(are) true?

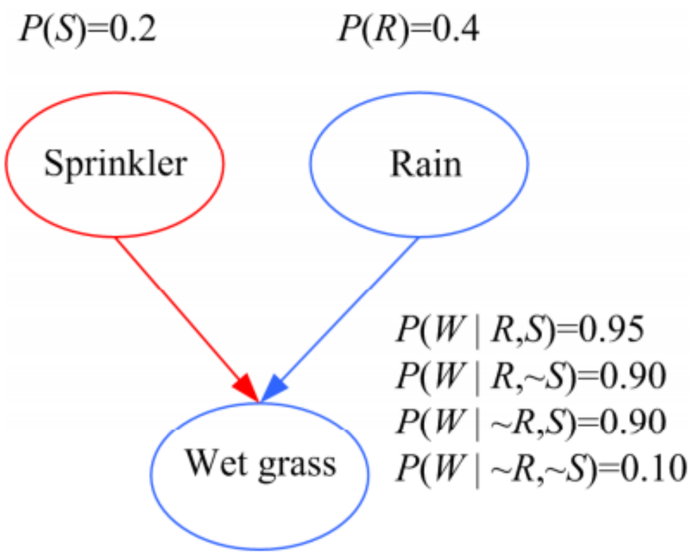
(i) A and B are independent
(ii) A and B are conditionally independent given C
(iii) A and D are independent
(iv) A and D are conditionally independent given C

Exercise 2: **Causal** vs Diagnostic Inference



Causal inference: ***If the***

Exercise 2: **Causal** vs Diagnostic Inference



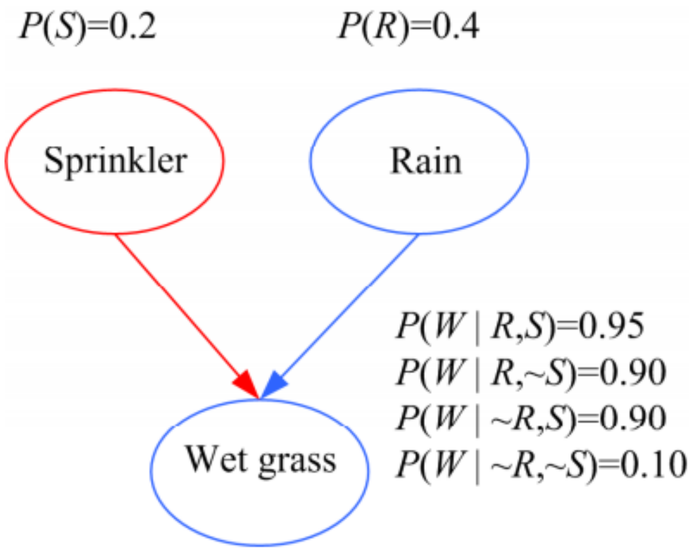
*Causal inference: **If the sprinkler is on, what is the probability that the grass is wet?***

$$P(W|S) = P(W|R,S)\,P(R|S) + P(W|\sim R,S)\,P(\sim R|S)$$

$$= P(W|R,S)\,P(R) + P(W|\sim R,S)\,P(\sim R)$$

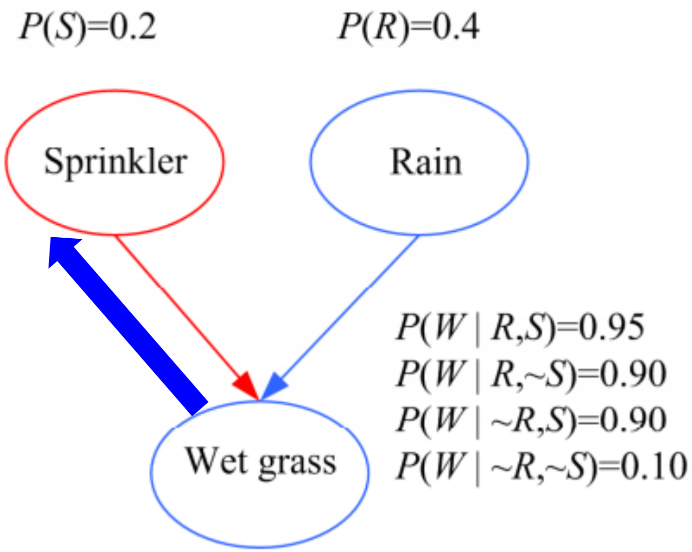
$$= 0.95*0.4 + 0.9*0.6 = 0.92$$

Exercise 2: Causal vs **Diagnostic** Inference



Diagnostic inference: If the grass is wet, what is the probability that the sprinkler is on?

Exercise 2: Causal vs **Diagnostic** Inference



Diagnostic inference: If the grass is wet, what is the probability that the sprinkler is on?

$P(S|W) = P(W,S) / P(W)$
$= P(W,S) / [P(W,S) + P(W, \sim S)]$
$= P(W|S)*P(S) / [P(W|S)*P(S) + P(W|\sim S)*P(\sim S)]$
$= 0.92*0.2/[0.92*0.2 + P(W|\sim S)*P(\sim S)]$
$= 0.92*0.2/[0.92*0.2 + 0.42*0.8]$
$= 0.3538$

$P(W|\sim S) = P(W,R|\sim S) + P(W,\sim R|\sim S)$
$= P(W|R,\sim S)*P(R|\sim S) + P(W|\sim R,\sim S)*P(\sim R|\sim S)$
$= P(W|R,\sim S)*P(R) + P(W|\sim R,\sim S)*P(\sim R) = 0.9*0.4+0.1*0.6 = 0.42$

The Bad News

- Exact inference is feasible in small to medium-sized networks
- Exact inference in large networks takes a very long time

We resort to approximate inference techniques which are much faster and give pretty good results

One Last Issue…

We still haven’t said where we get the Bayesian network from. There are two options:

- Get an expert to design it
- Learn it from data

Thank You!