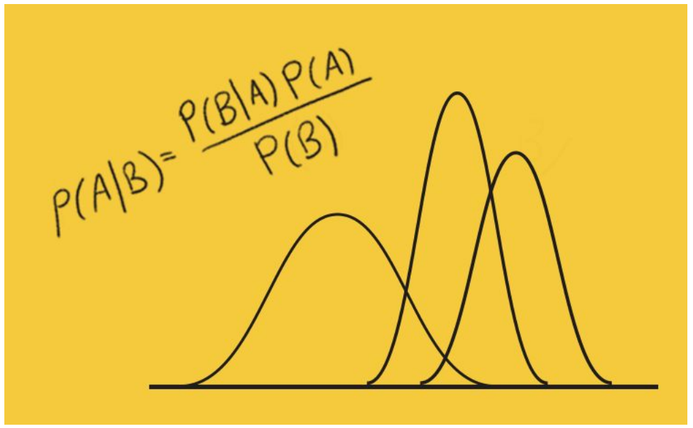


Bayesian Statistics (Basic) Cheat Sheet

V2020.10.14

(Dr Yan Xu)

Probability

➔ joint probability

$$\mathbb{P}(A = a_1, B = b_1)$$

➔ conditional probability

$$\mathbb{P}(A = a_1 | B = b_1) = \frac{\mathbb{P}(A = a_1, B = b_1)}{\mathbb{P}(B = b_1)}$$

Independence

➔ (absolutely) independent

$$\mathbb{P}(A = a | B = b) = \mathbb{P}(A = a)$$

➔ conditionally independent

$$\mathbb{P}(A = a | B = b, C = c) = \mathbb{P}(A = a | C = c)$$

➔ absolutely independent ***does not imply*** conditionally independent.

➔ also, conditionally independent ***does not imply*** absolutely independent.

Bayes Theorem

➔ Formula for 1 dependent variable:

$$\mathbb{P}(A = a_1 | B = b_1) = \frac{\mathbb{P}(A = a_1, B = b_1)}{\mathbb{P}(B = b_1)}$$

$$= \frac{\mathbb{P}(A = a_1) \times \mathbb{P}(B = b_1 | A = a_1)}{\mathbb{P}(B = b_1)}$$

- $\mathbb{P}(A = a_1)$ is called the **prior** probability
- $\mathbb{P}(B = b_1 | A = a_1)$ is called the **likelihood**
- $\mathbb{P}(B = b_1)$ is called the **evidence**
- $\mathbb{P}(A = a_1 | B = b_1)$ is called the **posterior** probability

➔ The case for two dependent variables:

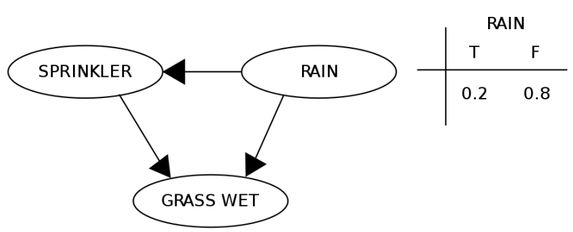
$$\mathbb{P}(A = a_1 | B = b_1, C = c_1) = \frac{\mathbb{P}(A = a_1 | C = c_1) \times \mathbb{P}(B = b_1 | A = a_1, C = c_1)}{\mathbb{P}(B = b_1 | C = c_1)}$$

Bayes Network (or Graphical Model)

➔ A directed acyclic graph (DAG)

i. nodes: variables
ii. edges: dependencies

RAIN	SPRINKLER T	SPRINKLER F
F	0.4	0.6
T	0.01	0.99



RAIN T	RAIN F
0.2	0.8

SPRINKLER	RAIN	GRASS WET T	GRASS WET F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

(source: wikipedia)

➔ determining independence: d-separation method

i. draw the ancestral graph
ii. connect parents of joint child
iii. remove directions
iv. remove given variables
v. path -> dependence

Posterior Distrubution Inference

➔ Enumeration
➔ Elimination
➔ Sampling (*conditional prob -> prob*)
- Rejection Sampling
- Gibbs Sampling
- Metropolis-Hasting Sampling

Conjugate Prior

➔ An Example:
likelihood function: ***Binomial(N, p)***
prior ***p***: **Beta(*α*, β)**
posterior ***p|X*** : **Beta(*α*+x, β+(n-x))**

➔ If likelihood function is discrete:

Likelihood distribution	Conjugate prior
Binomial	Beta
Poisson	Gamma
Geometric	Gamma
Multinomial	Dirichlet

➔ If likelihood function is continuous:

Likelihood distribution	Conjugate prior
Uniform	Pareto
Exponential	Gamma
Normal distribution, known σ^2	Normal distribution
Normal distribution, known μ	Inverse Gamma

Common Bayesian ML Models

➔ Naive Bayes
➔ Bayesian Linear Regression
➔ Bayesian Logistic Regression
➔ Gaussian Mixture Model (or Bayesian K-mean)
➔ Latent Dirichlet Allocation (topic model)