CSC-591: Foundations of Data Science
T/Th. 12:50-2:05pm. EBI-1005.

Ranga Raju Vatsavai
Chancellors Faculty Excellence Associate Professor in Geospatial Analytics
Department of Computer Science, North Carolina State University (NCSU)
Associate Director, Center for Geospatial Analytics, NCSU
&
Joint Faculty, Oak Ridge National Laboratory (ORNL)

W14: 11/19/15

Today

- Bayesian Inference
- Conclusion of statistical inference

NC STATE UNIVERSITY

Bayesian Data Analysis

- Bayesian analysis is a statistical procedure which endeavors to estimate parameters of an underlying distribution based on the observed distribution. BDA can be idealized into 3 steps:
  - Setting up a full probability model – a joint probability distribution for all observable and unobservable quantities in a problem
  - Conditioning on observed data – calculating and interpreting appropriate posterior distribution
  - Evaluating the fit of the model

11/19/15 CSC-591. 3

NC STATE UNIVERSITY

Bayesian Inference

- Bayes' rule plays central role
- P(A|B) = P(B|A)P(A) / P(B)
  - A (class); B(variables)
- P(A|B) – posterior probability
- P(A) – prior probability
- P(B|A) – conditional (or class conditional) probability (likelihood)

11/19/15 CSC-591. 4

NC STATE UNIVERSITY

Posterior Distribution

- Posterior distribution is the most important quantity in Bayesian inference.

$$f(\theta \mid x) = \frac{f(x \mid \theta) f(\theta)}{\int f(x \mid \theta) f(\theta) \, d\theta}$$

- Let X=x denote the observed realization of a uni- or multivariate r.v. X with density function f(x|θ). Specifying a prior distribution f(θ) allows us to compute the density function f(θ|x) of the posterior distribution using Bayes' theorem.

NC STATE UNIVERSITY

Prior Distribution

- Bayesian inference allows the probabilistic specification of prior beliefs through a prior distribution.
- It is often useful and justified to restrict the range of possible prior distributions to a specific family with one or two parameters, say. The choice of this family can be based on the type of likelihood function encountered.

NC STATE UNIVERSITY

Conjugate Prior Distributions

- A pragmatic approach to choosing a prior distribution is to select a member of a specific family of distributions such that the posterior distribution belongs to the same family. This is called a *conjugate prior distribution*.
- Let L(θ) = f (x | θ) denote a likelihood function based on the observation X = x. A class ζ of distributions is called *conjugate with respect to* L(θ) if the posterior distribution f (θ | x) is in ζ for all x whenever the prior distribution f (θ) is in ζ .

NC STATE UNIVERSITY

Conjugate Prior Distributions

Summary of conjugate prior distributions for different likelihood functions

Likelihood	Conjugate prior distribution	Posterior distribution
$X \mid \pi \sim \mathrm{Bin}(n, \pi)$	$\pi \sim \mathrm{Be}(\alpha, \beta)$	$\pi \mid x \sim \mathrm{Be}(\alpha + x, \beta + n - x)$
$X \mid \pi \sim \mathrm{Geom}(\pi)$	$\pi \sim \mathrm{Be}(\alpha, \beta)$	$\pi \mid x \sim \mathrm{Be}(\alpha + 1, \beta + x - 1)$
$X \mid \lambda \sim \mathrm{Po}(e \cdot \lambda)$	$\lambda \sim \mathrm{G}(\alpha, \beta)$	$\lambda \mid x \sim \mathrm{G}(\alpha + x, \beta + e)$
$X \mid \lambda \sim \mathrm{Exp}(\lambda)$	$\lambda \sim \mathrm{G}(\alpha, \beta)$	$\lambda \mid x \sim \mathrm{G}(\alpha + 1, \beta + x)$
$X \mid \mu \sim \mathrm{N}(\mu, \sigma^2 \text{ known})$	$\mu \sim \mathrm{N}(\nu, \tau^2)$	$\mu \mid x \sim \mathrm{N}\left(\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1} \cdot \left(\frac{x}{\sigma^2} + \frac{\nu}{\tau^2}\right), \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1}\right)$
$X \mid \sigma^2 \sim \mathrm{N}(\mu \text{ known}, \sigma^2)$	$\sigma^2 \sim \mathrm{IG}(\alpha, \beta)$	$\sigma^2 \mid x \sim \mathrm{IG}(\alpha + \frac{1}{2}, \beta + \frac{1}{2}(x - \mu)^2)$

Example

- Assume that the prior distribution for the proportion of defectives produced by a machine is

p	0.1	0.2
f(p)	0.6	0.4

- Let x be the number of defectives among a random sample of size 2.
- Find the posterior probability distribution of p, given that x is observed

Solution

p	0.1	0.2
f(p)	0.6	0.4

- Q: X follows which distribution?
- The random variable X follows a binomial distribution

- Compute marginal distribution
  - $f(x|0.1)f(0.1)+f(x|0.2)f(0.2)$

Solution

p	0.1	0.2
f(p)	0.6	0.4

- Q: X follows which distribution?
- The random variable X follows a binomial distribution

$$f(x \mid p) = b(x;2;p) = \binom{2}{x} p^x q^{2-x}, x = 0,1,2.$$

- Compute marginal distribution
  - f(x|0.1)f(0.1)+f(x|0.2)f(0.2)

$$\binom{2}{x}\left[(0.1)^x(0.9)^{2-x}(0.6)+(0.2)^x(0.8)^{2-x}(0.4)\right]$$

Solution

p	0.1	0.2
f(p)	0.6	0.4

- Hence, for x = 0, 1, 2, we obtain marginal probabilities as
- The posterior probability of p = 0.1 and 0.2, given x, is

NC STATE UNIVERSITY

Solution

p	0.1	0.2
f(p)	0.6	0.4

- Hence, for x = 0, 1, 2, we obtain marginal probabilities as

x	0	1	2
F(x)	0.742	0.236	0.022

- The posterior probability of p = 0.1 and 0.2, given x, is

$$f(0.1 \mid x) = \frac{f(x \mid 0.1) f(0.1)}{(0.1)^x (0.9)^{2-x} (0.6) + (0.2)^x (0.8)^{2-x} (0.4)}$$

$$f(0.2 \mid x) = 1 - f(0.1 \mid x)$$

NC STATE UNIVERSITY

Solution

y	0.1	0.2
p(y)	0.6	0.4

x	0	1	2
F(x)	0.742	0.236	0.022

- Suppose that x = 0 is observed
- Then, posterior is computed as
- Compute for other values of x

NC STATE UNIVERSITY

Solution

p	0.1	0.2
f(p)	0.6	0.4

x	0	1	2
F(x)	0.742	0.236	0.022

- Suppose that x = 0 is observed
- Then

$$f(0.1|0) = \frac{f(0|0.1)f(0.1)}{(0.1)^0(0.9)^{2-0}(0.6)+(0.2)^0(0.8)^{2-0}(0.4)}$$

$$= \frac{(0.1)^0(0.9)^{2-0}(0.6)}{(0.1)^0(0.9)^{2-0}(0.6)+(0.2)^0(0.8)^{2-0}(0.4)} = 0.655$$

$$f(0.2|0) = 1 - f(0.1|0) = 1 - 0.655 = 0.345$$

- Compute for other values of x

NC STATE UNIVERSITY

Practice

- Previous example, prior distribution of p is discrete. Now assume that p is a uniform distribution and compute posterior distribution of p.
  - From the conjugate prior table, check if the posterior distribution is a beta distribution
  - What are the parameters of this beta distribution

NC STATE UNIVERSITY

Solution

- The random variable X follows a binomial distribution $f(x \mid p) = b(x;2;p) = \binom{2}{x} p^x q^{2-x}, x = 0,1,2.$

- Compute marginal distribution

NC STATE UNIVERSITY

Solution

- The random variable X follows a binomial distribution $f(x \mid p) = b(x;2;p) = \binom{2}{x} p^x q^{2-x}, x = 0,1,2.$

- Compute marginal distribution

$$\int_0^1 f(x \mid p) f(p)\, dp = \binom{2}{x} \int_0^1 p^x (1-p)^{2-x}\, dp$$

Solution

$$g(x) = \binom{2}{x} \int_0^1 p^x (1-p)^{2-x} \, dp$$

- We can directly evaluate g(x) for each value of x
  x=0, g(0)=
- The posterior distribution of p, given x, is

$$f(p \mid x) = \frac{\binom{2}{x} \int_0^1 p^x (1-p)^{2-x}}{1/3} = 3\binom{2}{x} p^x (1-p)^{2-x}, 0 < p < 1.$$

- Posterior is a beta distribution, with parameters α=x+1 and β=3-x.
- So, if x=0 is observed, the posterior distribution of p is a beta distribution with parameters (1, 3)

Interval Estimation

- Similar to the classical CI, in Bayesian analysis we can calculate a 100(1-α)% Bayesian interval using the posterior distribution.

The interval $a < \theta < b$ will be called a $100(1-\alpha)\%$ **Bayesian interval** for θ if

$$\int_{-\infty}^{a} \pi(\theta|x) \, d\theta = \int_{b}^{\infty} \pi(\theta|x) \, d\theta = \frac{\alpha}{2}.$$

NC STATE UNIVERSITY

Interval Estimation

- In classical approach, 95% CI means that if an experiment is repeated again and again, the probability that the intervals calculated according to the rule will cover the true parameter is 95%.
- However, in Bayesian interval interpretation, say for a 95% interval, we can state that the probability of the unknown parameter falling into the calculated interval is 95%.

NC STATE UNIVERSITY

Example

- Suppose that X ~ Be(x;n,p), with known n=2, and the prior distribution of p is uniform f(p) = 1, for 0 < p < 1, find 95% Bayesian interval for p.

NC STATE UNIVERSITY

Solution

- From our previous example, x = 0, the posterior distribution is a beta distribution with parameters 1 and 3), i.e., f(p|0) = 3(1-p)2, for 0 < p < 1. Thus we need to solve for a and b using Bayesian interval formulae, which yields the following

$$0.025 = \int_0^a 3(1-p)^2\,dp = 1-(1-a)^3$$
$$and$$
$$0.025 = \int_b^1 3(1-p)^2\,dp = (1-b)^3$$

NC STATE UNIVERSITY

Solution

- Solving these equations result in a = 0.0084 and b = 0.7076. Therefore, the probability that p falls in (0.0084, 0.7076) is 95%.

Contingency Tables

- Data about two variables (bivariate) can be represented as contingency table which is useful estimating joint and marginal probabilities

Age (yr)	Rank: Full professor R_1	Associate professor R_2	Assistant professor R_3	Instructor R_4	Total
Under 30 A_1	2	3	57	6	68
30–39 A_2	52	170	163	17	402
40–49 A_3	156	125	61	6	348
50–59 A_4	145	68	36	4	253
60 & over A_5	75	15	3	0	93
Total	430	381	320	33	1164

Venn Diagram for Contingency Tables

Age (yr)	Rank: Full professor R_1	Associate professor R_2	Assistant professor R_3	Instructor R_4	Total
Under 30 A_1	2	3	57	6	68
30–39 A_2	52	170	163	17	402
40–49 A_3	156	125	61	6	348
50–59 A_4	145	68	36	4	253
60 & over A_5	75	15	3	0	93
Total	430	381	320	33	1164

	R_1	R_2	R_3	R_4
A_1	$(A_1 \& R_1)$	$(A_1 \& R_2)$	$(A_1 \& R_3)$	$(A_1 \& R_4)$
A_2	$(A_2 \& R_1)$	$(A_2 \& R_2)$	$(A_2 \& R_3)$	$(A_2 \& R_4)$
A_3	$(A_3 \& R_1)$	$(A_3 \& R_2)$	$(A_3 \& R_3)$	$(A_3 \& R_4)$
A_4	$(A_4 \& R_1)$	$(A_4 \& R_2)$	$(A_4 \& R_3)$	$(A_4 \& R_4)$
A_5	$(A_5 \& R_1)$	$(A_5 \& R_2)$	$(A_5 \& R_3)$	$(A_5 \& R_4)$

P(A1) = ?
P(R2) = ?
P(A1 & R2) = ?

Joint Probabilities from CT

		Rank				
		Full professor R_1	Associate professor R_2	Assistant professor R_3	Instructor R_4	$P(A_i)$
Age (yr)	Under 30 A_1	0.002	0.003	0.049	0.005	0.058
	30–39 A_2	0.045	0.146	0.140	0.015	0.345
	40–49 A_3	0.134	0.107	0.052	0.005	0.299
	50–59 A_4	0.125	0.058	0.031	0.003	0.217
	60 & over A_5	0.064	0.013	0.003	0.000	0.080
	$P(R_j)$	0.369	0.327	0.275	0.028	1.000

Joint probabilities are displayed in the cell and marginal distributions in the margin.

Acknowledgements

- Weiss, et. al. Walpole, et. al.