

# Today • Bayesian Inference • Conclusion of statistical inference 11/19/15 • Raju Vatsavai CSC-591.2

# **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

🛭 Raju Vatsava

CSC-591.

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

© Raju Vatsava

#### **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  $f(x|\theta)$ . Specifying a prior distribution  $f(\theta)$  allows us to compute the density function  $f(\theta|x)$  of the posterior distribution using Bayes' theorem.

11/19/15

🗅 Raju Vatsava

CSC-591. 5

NC STATI UNIVERSIT

#### **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.

11/19/15

© Raju Vatsavai

# **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.

11/19/15

🛭 Raju Vatsava

CSC-591. 7

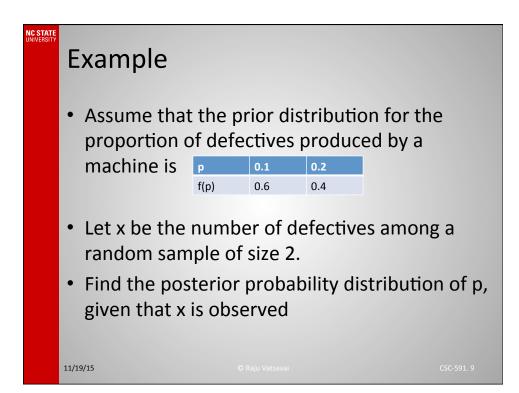
#### NC STATE UNIVERSITY

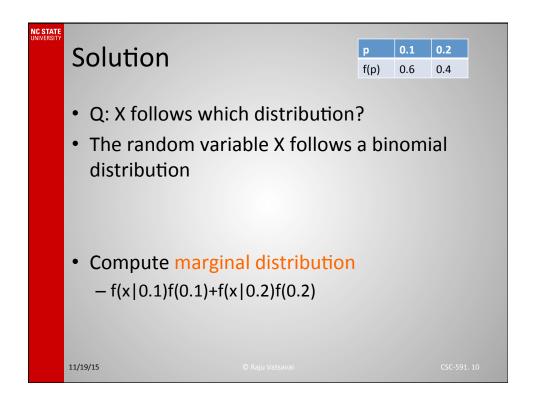
# **Conjugate Prior Distributions**

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

11/19/15

Raju Vatsavai





# Solution

р	0.1	0.2
f(p)	0.6	0.4

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

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

• Compute marginal distribution

$$-f(x|0.1)f(0.1)+f(x|0.2)f(0.2)$$

$$\begin{pmatrix} 2 \\ x \end{pmatrix} [(0.1)^{x}(0.9)^{2-x}(0.6) + (0.2)^{x}(0.8)^{2-x}(0.4)]$$

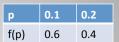
11/19/15

Raju Vatsavai

CSC E01 11

NC STATE UNIVERSITY

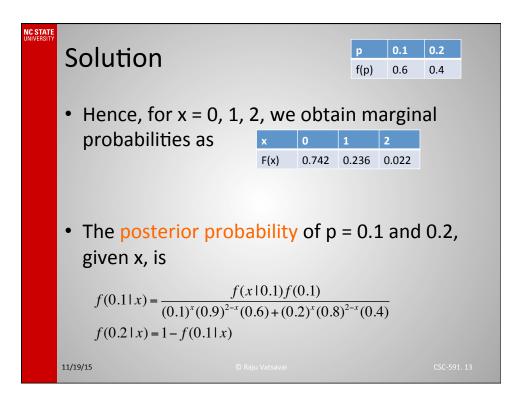
# Solution

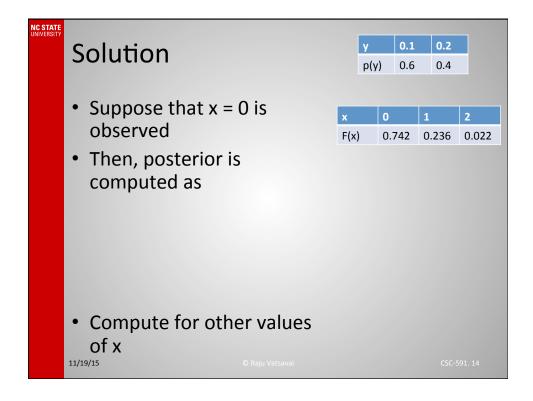


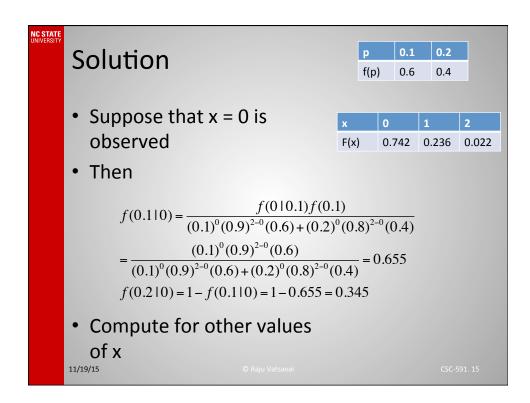
- Hence, for x = 0, 1, 2, we obtain marginal probabilities as
- The posterior probability of p = 0.1 and 0.2, given x, is

11/19/15

Raju Vatsava







# 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

11/19/15 © Raju Vatsavai

# Solution

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

11/19/15

🛭 Raju Vatsava

CSC-591, 17

NC STAT UNIVERSIT

# Solution

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

$$\int_0^1 f(x \mid p) f(p) dp = \begin{pmatrix} 2 \\ x \end{pmatrix} \int_0^1 p^x (1-p)^{2-x} dp$$

11/19/15

© Raju Vatsava

# Solution

 $g(x) = \begin{pmatrix} 2 \\ x \end{pmatrix} \int_0^1 p^x (1-p)^{2-x} dp$ 

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

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

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

11/19/15

© Raju Vatsav:

CSC-591. 1

NC STATI

# **Interval Estimation**

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

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

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

11/19/15

Raju Vatsavai

#### **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%.

11/19/15

🕽 Raju Vatsava

CSC-591. 2:

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

11/19/15

Raju Vatsavai

# 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(1p)2, for 0

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

11/19/15

🕽 Raju Vatsava

CSC-591, 23

NC STATE

# 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%.

11/19/15

Raju Vatsavai

