# *Prior-Posterior Analysis and Conjugacy*

Econ 690

Purdue University

# Outline

# Review of Basic Framework

- Quantities to become known under sampling are denoted by the $T$-dimensional vector $y$,
- The remaining unknown quantities are denoted by the $k$-dimensional vector $\theta \in \Theta \subseteq \mathcal{R}^k$.
- Standard manipulations show:

$$p(y, \theta) = p(\theta)p(y|\theta) = p(y)p(\theta|y),$$

where $p(\theta)$ is the prior density, $p(\theta|y)$ is the posterior density and $p(y|\theta)$ is the likelihood function

- We also note

$$p(y) = \int_{\Theta} p(\theta)L(\theta)d\theta$$

is the marginal density of the observed data, also known as the marginal likelihood)

# Bayes Theorem

*Bayes' theorem for densities* follows immediately:

$$p(\theta|y) = \frac{p(\theta)L(\theta)}{p(y)} \propto p(\theta)L(\theta).$$

- The shape of the posterior can be learned by plotting the right hand side of this expression when $k = 1$ or $k = 2$.
- Obtaining moments or quantiles, however, requires the integrating constant, i.e., the marginal likelihood $p(y)$.
- In most situations, the required integration cannot be performed analytically.
- In simple examples, however, this integration can be carried out. Many of these cases arise in conjugate situations. By "conjugacy," we mean that the functional forms of the prior and posterior are the same.

# Conjugate Bernoulli Trials

Given a parameter $\theta$ where $0 < \theta < 1$, consider $T$ iid Bernoulli random variables $Y_t$ $(t = 1, 2, \cdots, T)$, each with probability mass function (p.m.f.):

$$p(y_t|\theta) = \begin{cases} \theta & \text{if } y_t = 1 \\ 1-\theta & \text{if } y_t = 0 \end{cases} = \theta^{y_t}(1-\theta)^{1-y_t}.$$

The likelihood function associated with this data is

- 

where $m = T\overline{y}$ is the number of successes (i.e., $y_t = 1$) in $T$ trials.

# Conjugate Bernoulli Trials

Suppose prior beliefs concerning $\theta$ are represented by a Beta distribution with p.d.f.

- 

where $\underline{\alpha} > 0$ and $\underline{\delta} > 0$ are known, and $\mathcal{B}(\underline{\alpha}, \underline{\delta}) = \Gamma(\underline{\alpha})\Gamma(\underline{\delta})/\Gamma(\underline{\alpha} + \underline{\delta})$ is the Beta function defined in terms of the Gamma function $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \exp(-t) dt$.

# Conjugate Bernoulli Trials

- Note that the Beta is a reasonable choice of prior, since it incorporates the necessary constraint that $\theta \in (0, 1)$.
- Also note that $\underline{\alpha}$ and $\underline{\delta}$ are chosen by you!
- Some guidance in this regard can be obtained my noting:

# Conjugate Bernoulli Trials

By Bayes' Theorem:

$$p(\theta|y) \propto p(\theta)p(y|\theta).$$

Putting the previous parts together, we obtain

- 

where

- 

Thus, the posterior distribution for $\theta$ is also of the Beta form, $\theta|y \sim B(\overline{\alpha}, \overline{\delta})$ so that the beta density is a conjugate prior for the Bernoulli sampling model.

# Conjugate Bernoulli Trials

From our handout on "special" distributions, we know that

$$E(\theta|y) = \frac{\overline{\alpha}}{\overline{\alpha} + \overline{\delta}} = \frac{\underline{\alpha} + T\overline{y}}{\underline{\alpha} + \underline{\delta} + T}.$$

Similarly, the prior mean is

$$E(\theta) \equiv \mu = \frac{\underline{\alpha}}{\underline{\alpha} + \underline{\delta}}.$$

Expanding the posterior mean a bit further, we find:

- 

# Conjugate Bernoulli Trials

$$E(\theta|y) = w_T \overline{y}_T + (1 - w_T)\mu,$$

a weighted average of the sample mean $\overline{y}$ and the prior mean $\mu$.
What happens as $T \rightarrow \infty$?
Note that

$$w_T = \frac{T}{\underline{\alpha} + \underline{\delta} + T}$$

and thus as $T \rightarrow \infty$, $w_T \rightarrow 1$, and thus the posterior mean $E(\theta|y)$ approaches the sample mean $\overline{y}_T$.
This is sensible, and illustrates that, in large samples, information from the data dominates information in the prior (provided the prior is not dogmatic).

# Conjugate Bernoulli Analysis: Example

Consider the 2011 record for the Purdue football team: $(T = 12, \overline{y} = .5)$:

$$y = [1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 1]'.$$

As a "neutral" fan, before the season started, you had little prior information about Purdue's success probability $\theta$.
You summarized this lack of information by choosing

- 

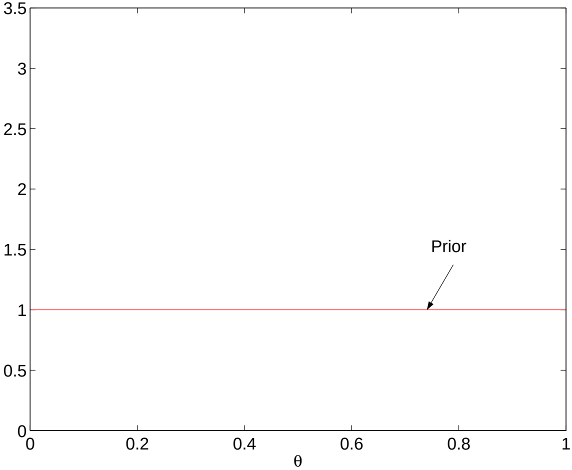
and thus

$$p(\theta) = I(0 < \theta < 1),$$

i.e., a uniform prior over the unit interval.

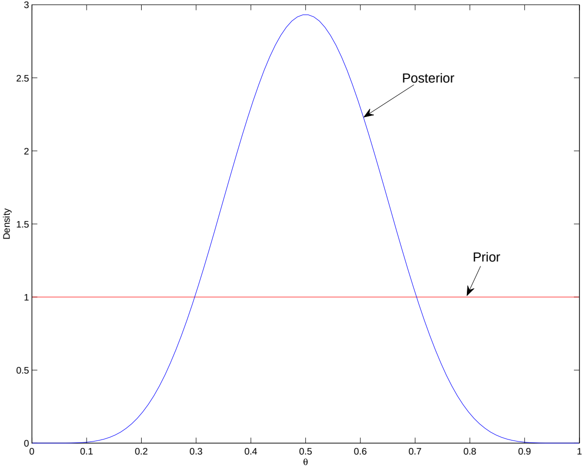
# Conjugate Bernoulli Analysis: Example

Your prior over $\theta$ can be graphed as follows:



# Conjugate Bernoulli Analysis: Example

Your posterior beliefs, after observing all 12 games, is as follows:

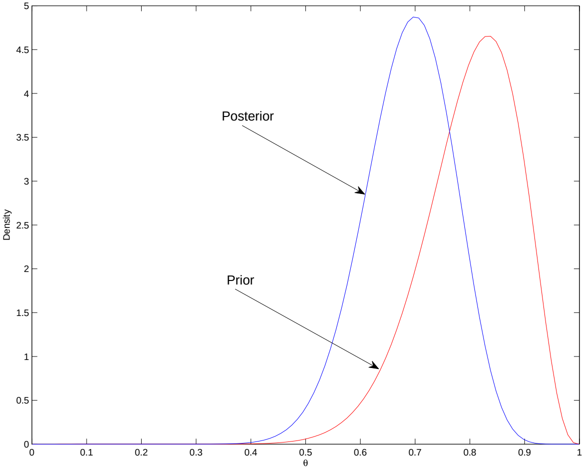


# Conjugate Bernoulli Analysis: Example

- Now suppose, instead of having "no" prior information, you expected that Purdue would win 80 percent of its games this season.
- You incorporate these beliefs by choosing the following prior hyperparameters:
  - 

  Note that this implies
  - 
- The prior and posterior under this scenario are as follows:

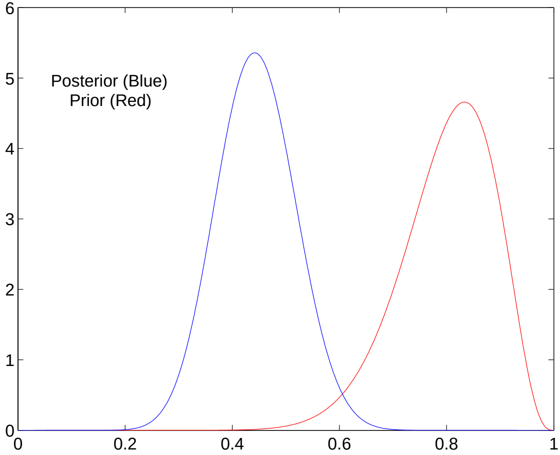
# Conjugate Bernoulli Analysis, Example



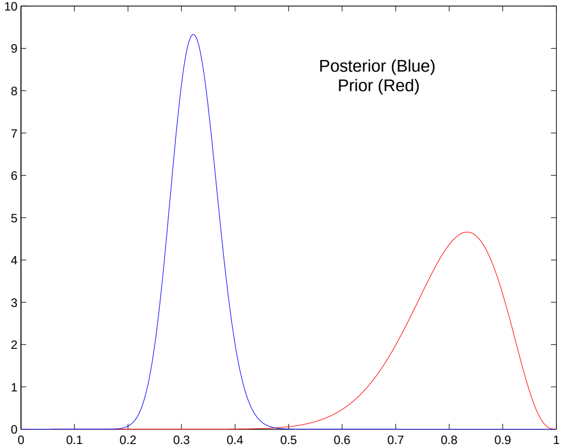
# Conjugate Bernoulli Analysis, Example

- To illustrate the impact of the sample size on the posterior, let us conduct an experiment.
- Using $\theta = .25$ as the "true" probability of the data generating process, let's generate $y$ vectors of length $N = 25, 100, 1,000$, where $y_i = 1$ with probability .25 and 0 otherwise, for all $i$.
- Keep the same "optimistic" prior.
- Examine how the posterior changes as the sample size increases.

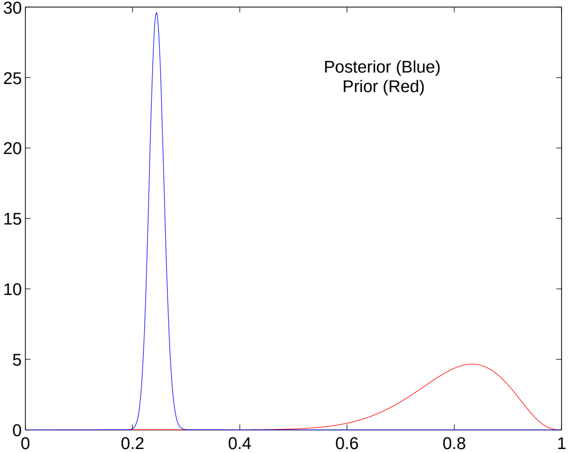
# Conjugate Bernoulli Analysis, Example: $N = 25$



# Conjugate Bernoulli Analysis, Example: $N = 100$



# Conjugate Bernoulli Analysis, Example: $N = 1,000$



# Conjugate Bernoulli Analysis: Marginal Likelihood

Consider, for this problem, determining the marginal likelihood $p(y)$:

$$p(y) = \int_{\Theta} p(\theta)p(y|\theta)d\theta.$$

Here the integration is reasonably straightforward:

-

where the last integral equals unity because the integrand is a Beta p.d.f. for $\theta$.

# Conjugate Exponential Analysis

Suppose $Y_t$ $(t = 1, 2, \cdots, T)$ is a random sample form an Exponential distribution $f_{EXP}(y_t|\theta) = \theta \exp(-\theta y_t)$, which has mean $\theta^{-1}$.

In addition, suppose that the prior distribution of $\theta$ is the Gamma distribution $G(\underline{\alpha}, \underline{\beta})$ where $\underline{\alpha} > 0$ and $\underline{\beta} > 0$:

$$p(\theta) \propto \theta^{\underline{\alpha}-1} \exp(-\theta/\underline{\beta}).$$

What is the posterior distribution of $\theta$?

# Conjugate Exponential Analysis

The likelihood function is

- 

Define $\overline{\alpha} = \underline{\alpha} + T$ and $\overline{\beta} = (\underline{\beta}^{-1} + T\overline{y})^{-1}$.
Using Bayes Theorem, the posterior density is

- 

Therefore, $\theta | y \sim G(\overline{\alpha}, \overline{\beta})$. Thus the Gamma prior is a conjugate prior for the exponential sampling model.

# Conjugate Exponential Analysis

Using properties of the Gamma distribution, we know:

- 

Note, in this parameterization of the exponential,

$$E(y|\theta) = \frac{1}{\theta}$$

and the MLE is

$$\hat{\theta}_{MLE} = \frac{1}{\overline{y}_T}.$$

# Conjugate Exponential Analysis: Example

Assume that the duration of the life of a lightbulb is described by an exponential density,

$$p(y_i|\theta) = \theta^{-1} \exp(-\theta^{-1} y_i).$$

We parameterize the exponential in this way to work in terms of the mean of $y$.

You obtain data on 10 continuously running light bulbs and find that they last 25, 20, 40, 75, 15, 30, 30, 10, 20 and 40 days, respectively. Using an inverse gamma prior for $\theta$ of the form

- 

derive the posterior distribution of $\theta$ and plot it alongside the prior.

# Conjugate Exponential Analysis: Example

Note

- 

Combining this with our prior, we obtain

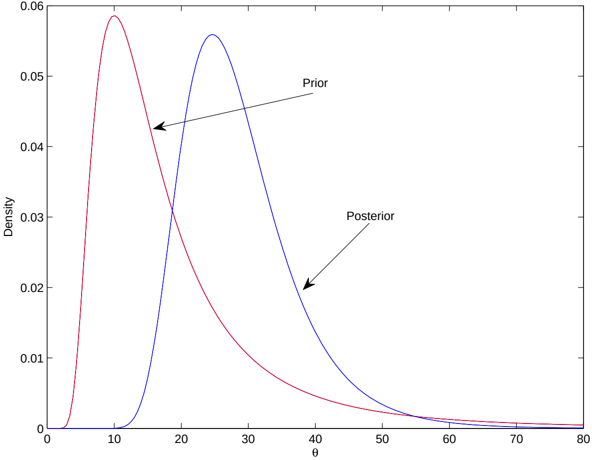
$$p(\theta|y) \propto \theta^{-(\underline{\alpha}+T+1)} \exp\left(-\theta^{-1}[T\overline{y} + \underline{\beta}^{-1}]\right).$$

This is in the form of an

- 

density.

# Conjugate Exponential Analysis, Example: $\underline{\alpha} = 3$, $\underline{\beta} = 1/40$.



# Conjugate Exponential Analysis: Example

In this example, our choice of prior hyperparameters produced a prior that had a mean and standard deviation equal to 20. To see this, note (from the distributional catalog notes):

- 

and

- 

# Conjugate Exponential Analysis: Example

Think about what the output represents and what kinds of questions you can answer:

*What is the (posterior) probability that a light bulb has an average life span of more than 30 days?*

-

# Conjugate Exponential Analysis: Example

*Suppose I intend to purchase a light bulb tomorrow. Based on the data that I have observed (as well as my own prior beliefs), what is the probability that the light bulb I purchase will last at least 30 days?*

Let $y_f$ denote the future, as yet unobserved duration of our light bulb. We would first seek to recover

- 

the *posterior predictive density*. We can do this (see future notes on prediction) and obtain:

- 

(Note that the posterior predictive density and the $\theta$ posterior distribution are *not the same thing!*)

# Conjugate Poisson Analysis

Suppose $Y_t(\ t = 1, 2, \cdots, T)$ is a random sample from a Poisson distribution with mean $\theta$, i.e.,

$$p(y_t|\theta) = \frac{\theta^{y_t} \exp(-\theta)}{y_t!}, \quad y_t = 0, 1, 2, \ldots$$

and that the prior distribution of $\theta$ is the Gamma distribution $G(\underline{\alpha}, \underline{\beta})$ :

$$p(\theta) \propto \theta^{\underline{\alpha}-1} \exp(-\theta/\underline{\beta}).$$

Find the posterior distribution of $\theta$.

# Conjugate Poisson Analysis

The likelihood function is

- 

Define $\overline{\alpha} = \underline{\alpha} + T\overline{y}$ and $\overline{\beta} = (\underline{\beta}^{-1} + T)^{-1}$. Using Bayes Theorem, the posterior density is proportional to:

- 

Therefore, $\theta | y \sim G(\overline{\alpha}, \overline{\beta})$. Thus, the Gamma prior is a conjugate prior for the Poisson sampling model.

# Conjugate Poisson Analysis

As before, note

- 

Therefore, the posterior mean converges to $\overline{y}_T$ as $T \rightarrow \infty$. Similarly,

$$\text{Var}(\theta|y) = \overline{\alpha}\overline{\beta}^2 \rightarrow 0 \quad \text{as } T \rightarrow \infty.$$