

Bayesian Non-Conjugate Priors

Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University



Copyright D.B. Rowe 1

Bayesian Statistics



Outline

Prior Information

Deterministic Integration

Stochastic Integration

Non-Conjugate Prior for Binomial RVs (Parabolic prior for p)

Discussion

Homework



Prior Information

When we expect observations from a particular distribution, $f(x/\theta)$, it may be the case that the information that we have about the parameter(s) θ do not fit into the conjugate prior framework. It is too constraining.

In such cases, we have non-conjugate prior distributions which when combined with the likelihood for the forthcoming observations does not necessarily form a "nice" $f(x,\theta)$ or "friendly" $f(\theta/x)$.

We may need to use advanced computational methods.



Prior Information

These advanced computational methods have the goal to compute the (marginal) posterior mean

$$E(\theta \mid x_1, ..., x_n) = \int_{\theta} \theta f(\theta \mid x_1, ..., x_n) d\theta \leftarrow \text{Can't always be found with pencil and paper!}$$

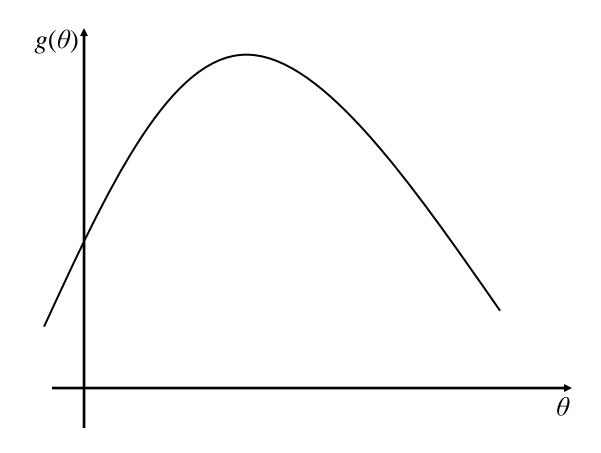
where

$$f(\theta \mid x_1, ..., x_n) = \frac{f(x_1, ..., x_n \mid \theta) f(\theta)}{f(x_1, ..., x_n)}$$

But to do this, don't forget that we also need

$$f(x_1,...,x_n) = \int_{\theta} f(x_1,...,x_n \mid \theta) f(\theta) d\theta \quad \leftarrow \text{ Can't always be found with pencil and paper!}$$

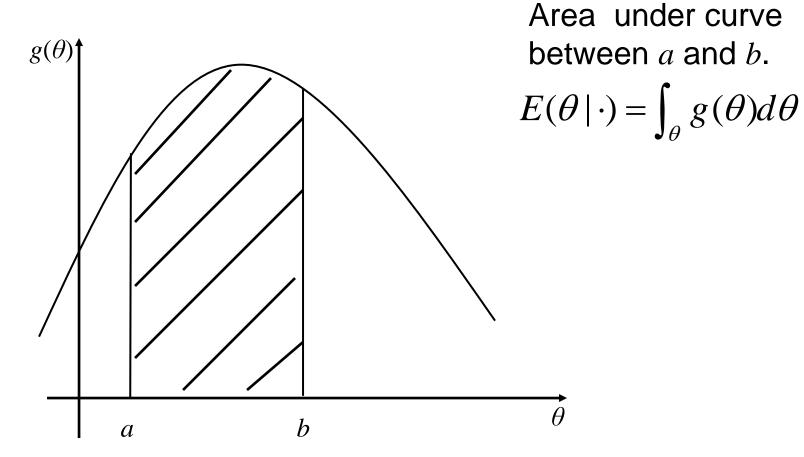




$$E(\theta \mid x_1, ..., x_n) = \int_{\theta} \theta f(\theta \mid x_1, ..., x_n) d\theta$$

$$g(\theta)$$

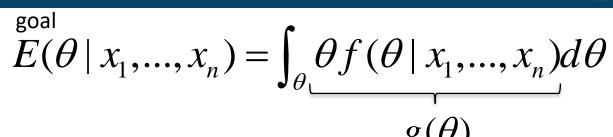




 $E(\theta \mid x_1, ..., x_n) = \int_{\theta} \theta f(\theta \mid x_1, ..., x_n) d\theta$

$$g(\theta) = \theta f(\theta)$$





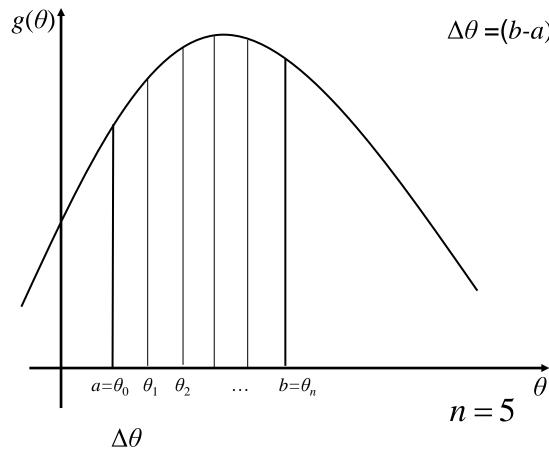
Divide into intervals: $\Delta\theta$ small

$$\Delta \theta = (b - a)/n$$

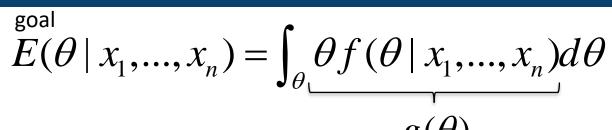
$$\Delta\theta = (b-a)/n$$
 $\Delta\theta = \theta_i - \theta_{i-1}$

think of as

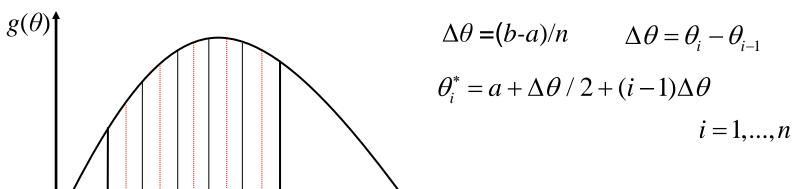
$$g(\theta) = \theta f(\theta)$$





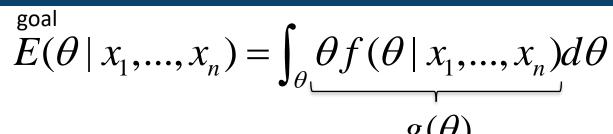


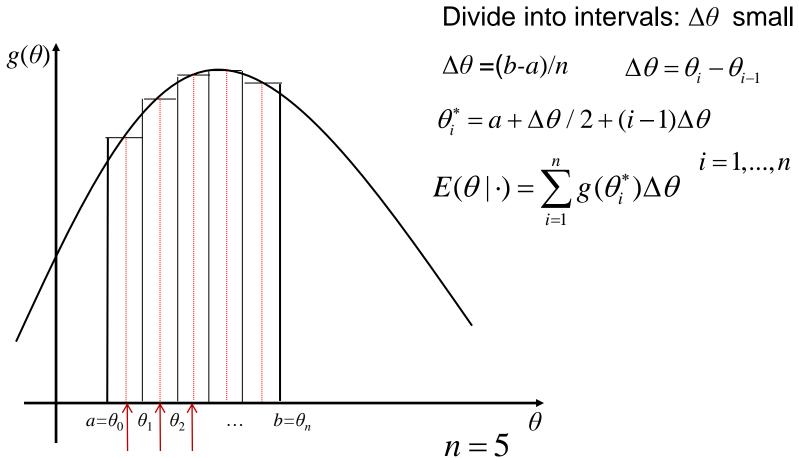
Divide into intervals: $\Delta\theta$ small



 $a = \theta_0 \quad \theta_1 \quad \theta_2 \quad \dots \quad b = \theta_n$ n = 5



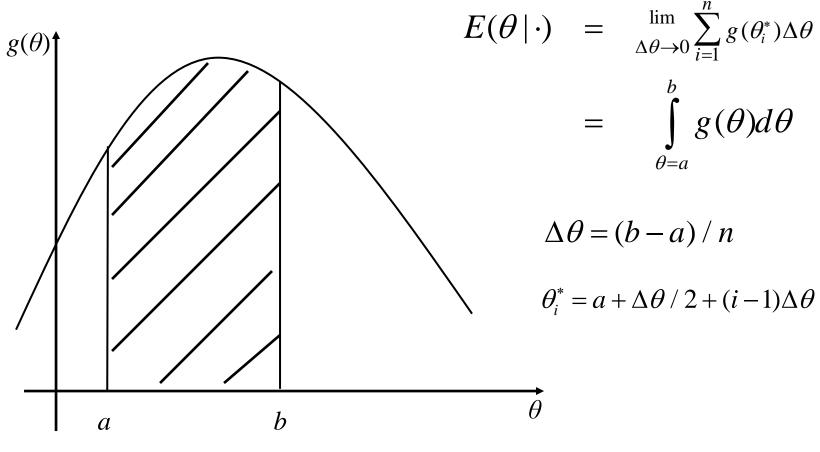




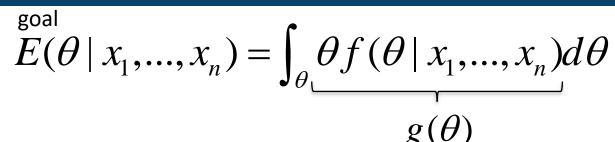


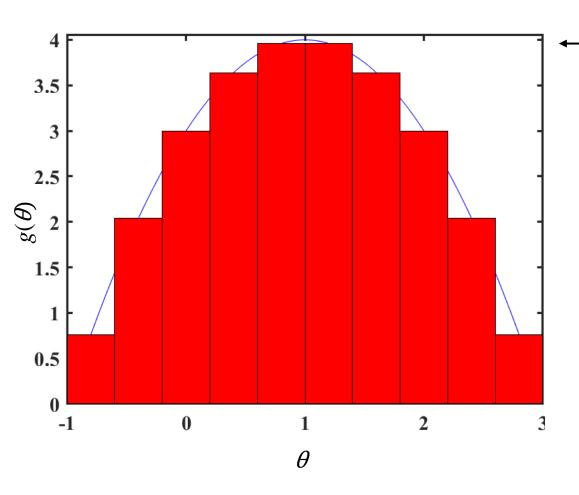
$$E(\theta \mid x_1, ..., x_n) = \int_{\theta} \theta f(\theta \mid x_1, ..., x_n) d\theta$$

$$g(\theta)$$









$$g(\theta) = 4 - (\theta - 1)^2$$

-numerical

$$n=10, \quad \Delta \theta = 0.400$$

$$\hat{E}(\theta \mid \cdot) = \Delta \theta \sum_{i=1}^{10} g(\theta_i^*) = 10.7200$$

analytic

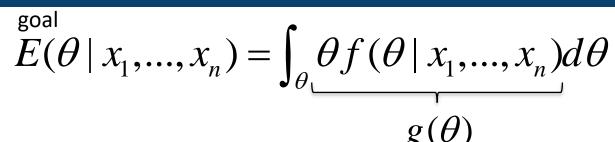
$$\int_{\theta=-1}^{3} [4 - (\theta - 1)^{2}] d\theta = 10.6667$$

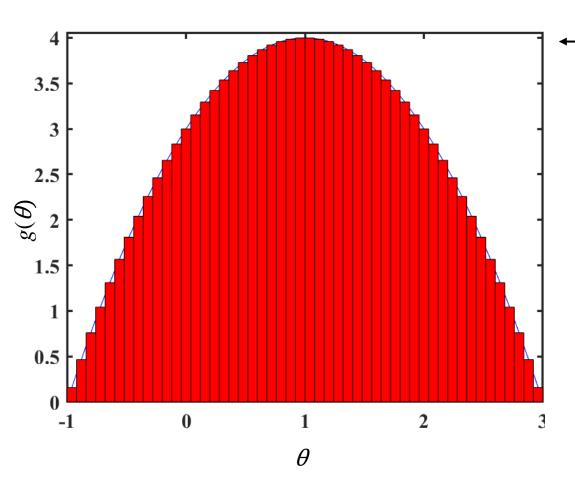
```
% Numerical Integral
a=-1; b=3;
n=10; dt=(b-a)/n;
tpts=(a+dt/2:dt:b)';
gpts=4-(tpts-1).^2;
Ethat=dt*sum(gpts)
```

 $g(\theta) = 4 - (\theta - 1)^2$



Deterministic Integration





numerical

n=50, $\Delta \theta = 0.080$

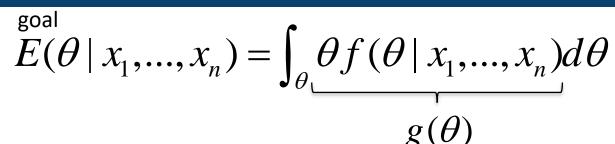
$$\hat{E}(\theta | \cdot) = \Delta \theta \sum_{i=1}^{50} g(\theta_i^*) = 10.6688$$

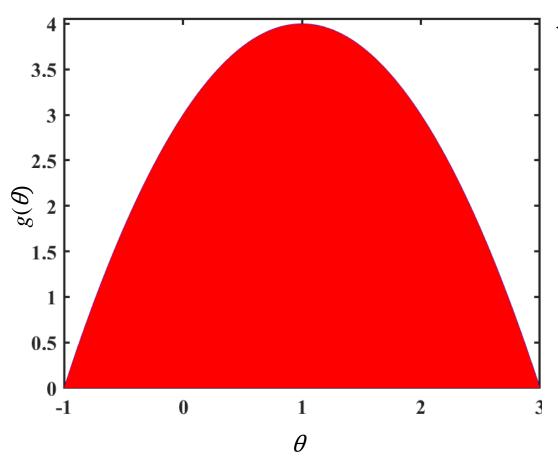
analytic

$$\int_{\theta=-1}^{3} [4 - (\theta - 1)^{2}] d\theta = 10.6667$$

% Numerical Integral
a=-1; b=3;
n=50; dt=(b-a)/n;
tpts=(a+dt/2:dt:b)';
gpts=4-(tpts-1).^2;
Ethat=dt*sum(gpts)







 $\hat{E}(\theta \mid \cdot) = \Delta \theta \sum_{i=1}^{1000} g(\theta_i^*) = 10.6667$ analytic $\int_{\theta=-1}^{3} [4 - (\theta - 1)^2] d\theta = 10.6667$ % Numerical Integral

a=-1; b=3;

n=1000; dt=(b-a)/n;

 $gpts=4-(tpts-1).^2;$

Ethat=dt*sum(gpts)

tpts=(a+dt/2:dt:b)';

numerical

 $n=1000, \Delta\theta=0.004$

D.B. Rowe

 $g(\theta) = 4 - (\theta - 1)^2$



We can also integrate a function $g(\theta)$ via random uniform numbers.

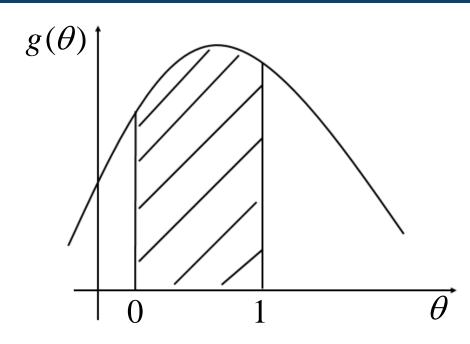
$$E(\theta | \cdot) = \int_0^1 g(u) du \quad \leftarrow \text{ changed symbol of integration}$$

We can view this as there being a PDF f(u) where u is uniformly distributed over (0,1)

$$f(u) = 1$$
 $0 \le u \le 1$

and we are calculating the expected value

$$E(\theta \mid \cdot) = E(g(u)) = \int_0^1 g(u)f(u)du \quad \longleftarrow \text{key idea}$$



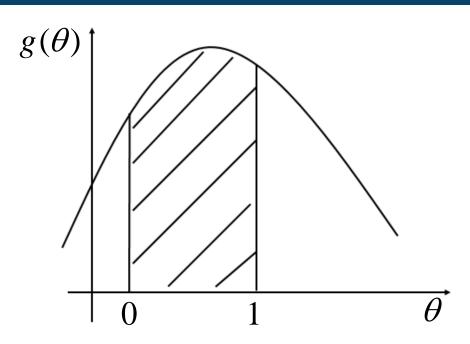


If we have a random sample $u_1, ..., u_n$ from f(u), then we can calculate a sample version of

$$E(\theta \mid \cdot) = E(g(u)) = \int_0^1 g(u) f(u) du$$

with the iid sample $u_1, ..., u_n$, from f(u). The computed values $g(u_1), ..., g(u_n)$ are an iid sample with mean E(g(u)).

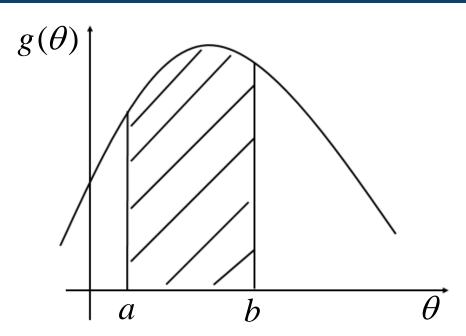
$$\frac{1}{n} \sum_{i=1}^{n} g(u_i) \to E[g(u)] = E(\theta \mid \cdot) \quad \text{as } n \to \infty$$





If we want $E(\theta|\cdot) = \int_a^b g(\theta)d\theta$, then we can transform θ to u as $u = (\theta-a)/(b-a)$, with $du = d\theta/(b-a)$

$$E(\theta \mid \cdot) = \int_{a}^{b} g(\theta) d\theta \qquad d\theta = (b - a) du$$
$$\theta = (b - a) u + a$$



Then our integral becomes

$$E(\theta \mid \cdot) = \int_0^1 g(a + (b - a)u)(b - a)du$$

$$h(u) = g(a + (b-a)u)(b-a)$$

$$f(u) = 1 \qquad 0 \le u \le 1$$

which is the same as

$$E(\theta \mid \cdot) = E(h(u)) = \int_0^1 h(u) f(u) du$$



Let's use this idea to evaluate the same integral.

Deterministic

$$E(\theta \mid \cdot) = \int_{\theta} g(\theta) d\theta$$

$$\hat{E}_n(\theta \mid \cdot) = \Delta \theta \sum_{i=1}^n g(\theta_i^*)$$

$$\hat{E}_{10}(\theta \mid \cdot) = 10.7200$$

Theoretical

$$\int_{-1}^{3} [4 - (\theta - 1)^{2}] d\theta = 10.6667$$

$$(b-a) =$$

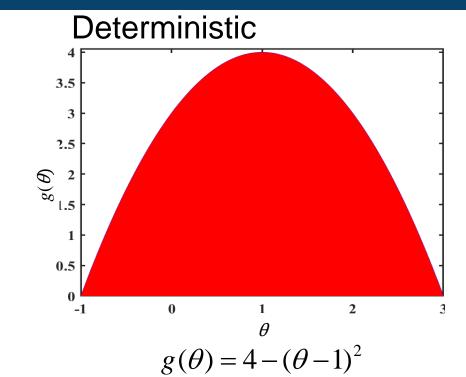
Stochastic

$$h(u) = g(a + (b-a)u)(b-a)$$

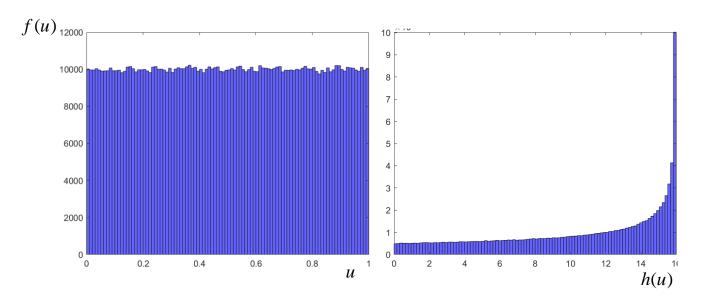
$$\frac{1}{n} \sum_{i=1}^{n} h(u_i) = 10.6718$$

% stochastic integral rnq('default') $a=-1; b=3; n=10^6;$ u=rand(n,1);t = (a + (b-a) * u); $hu = (4 - (t-1) \cdot ^2) * (b-a);$ Ethat=sum(hu)/n

(b-a) = 4



Stochastic





If we want $E(\theta|\cdot) = \int_0^\infty g(\theta)d\theta$, then we can transform θ to u as $u=1/(\theta+1)$, with $du=-1/(\theta+1)^2d\theta=-u^2d\theta$

$$E(\theta \mid \cdot) = \int_0^\infty g(\theta) d\theta \qquad d\theta = -du \mid u^2 \qquad \text{limits} \qquad \theta = 0 \to u = 1$$
$$\theta = 1 \mid u = 1 \qquad \theta = \infty \to u = 0$$

Then our integral becomes

$$E(\theta | \cdot) = \int_0^1 g(1/u - 1) \frac{du}{u^2}$$

$$u = 1/(\theta + 1)$$

$$E(\theta | \cdot) = \int_0^1 h(u) f(u) du$$

$$h(u) = g(1/u - 1)/u^2$$

and we can use the same technique.

$$\overline{E}(\theta \mid \cdot) = \frac{1}{n} \sum_{i=1}^{n} h(u_i)$$



logistic function

If we want $E(\theta|\cdot) = \int_{-\infty}^{\infty} g(\theta)d\theta$, then we can transform θ to u as $u=e^{\theta}/(1+e^{\theta})$, with $du=e^{\theta}/(1+e^{\theta})^2d\theta=ud\theta/(1-u)$

$$E(\theta \mid \cdot) = \int_{-\infty}^{\infty} g(\theta) d\theta \qquad d\theta = u du / (1 - u) \quad \text{limits} \qquad \theta = -\infty \to u = 0$$

$$\theta = \ln\left(\frac{u}{1 - u}\right) \qquad \theta = +\infty \to u = 1$$

Then our integral becomes

$$E(\theta | \cdot) = \int_0^1 g \left(\ln \left(\frac{u}{1 - u} \right) \right) \frac{du}{u(1 - u)}$$

$$E(\theta | \cdot) = \int_0^1 h(u) f(u) du$$

$$u = e^{\theta} / (1 + e^{\theta})$$

$$h(u) = g \left(\ln \left(\frac{u}{1 - u} \right) \right) \frac{1}{u(1 - u)}$$

and we can use the same technique.

$$\overline{E}(\theta \mid \cdot) = \frac{1}{n} \sum_{i=1}^{n} h(u_i)$$



Binomial observation *x*:

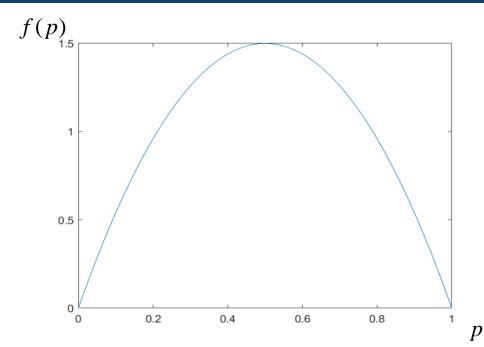
Let's imagine that our prior information for p was consistent with the following parabolic prior for p,

$$f(p) = \frac{3}{2} \left[1 - 4(p - 1/2)^2 \right]$$

$$p \in [0,1]$$

to go along with the binomial likelihood for x

$$f(x \mid p) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x} \qquad p \in [0,1]$$
$$x = 0,1,....,n$$

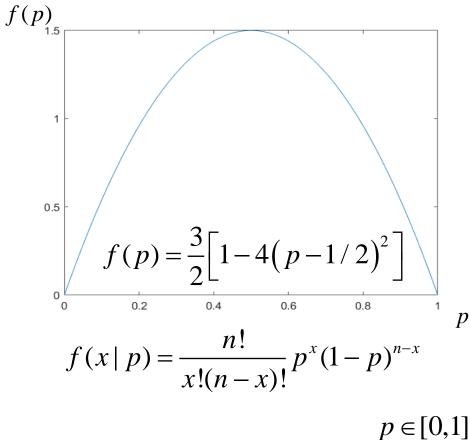




Binomial observation x:

The parabolic prior for p and binomial likelihood for x do not combine to form a "nice" joint distribution f(x,p)

$$f(x,p) = \frac{3}{2} \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x} \left[1 - 4(p-1/2)^{2} \right]$$



or yield a "friendly" posterior distribution f(p/x) for p.

So we need to resort to more advanced methods.

i.e. Try harder for pencil & paper integral, deterministic integration (rectangles), stochastic integration (random u's), MCMC.

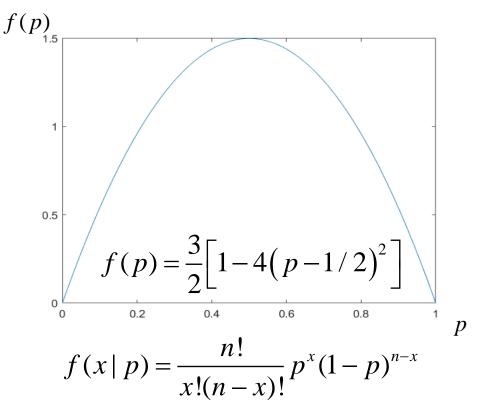


Binomial observation x:

The f(p) and f(x/p) do not combine to form a "nice" f(x,p) or yield a "friendly" posterior f(p/x).

It can be shown (homework problem) that

$$f(x) = 6\frac{(x+1)(n-x+1)}{(n+3)(n+2)(n+1)}$$



$$f(p|x) = \frac{1}{4} \frac{(n+3)!}{(x+1)!(n-x+1)!} p^{x} (1-p)^{n-x} \left[1 - 4(p-1/2)^{2} \right]$$

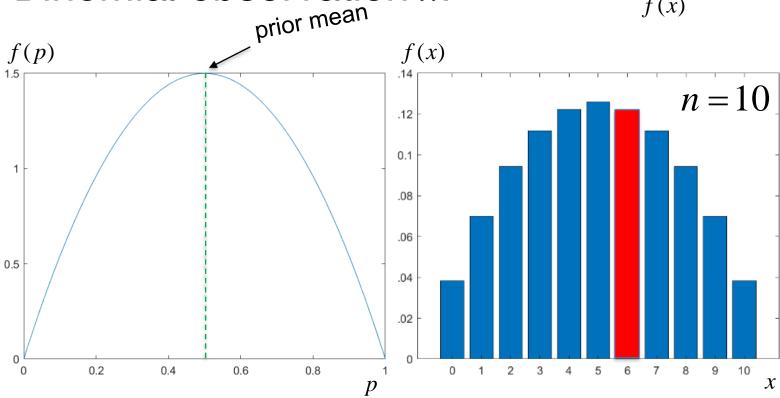
 $p \in [0,1]$



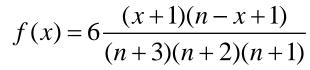
$f(x \mid p) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x}$

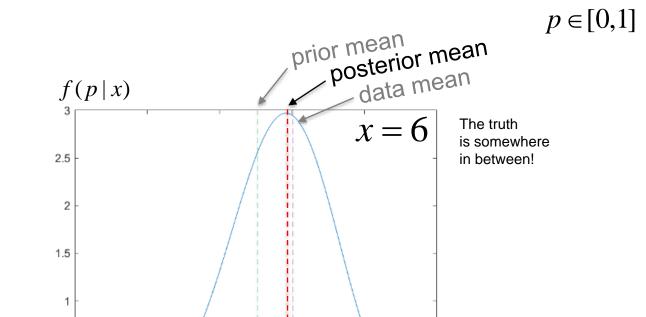
Binomial observation *x*:

$$f(p \mid x) = \frac{f(x \mid p)f(p)}{f(x)}$$



$$f(p) = \frac{3}{2} \left[1 - 4(p - 1/2)^2 \right]$$





$$f(x) = 6 \frac{(x+1)(n-x+1)}{(n+3)(n+2)(n+1)} \qquad f(p \mid x) = \frac{1}{4} \frac{(n+3)!}{(x+1)!(n-x+1)!} p^{x} (1-p)^{n-x} \left[1 - 4(p-1/2)^{2} \right]$$

0.5

0.2



Binomial observation x:

From this posterior PDF for p,

$$f(p|x) = \frac{1}{4} \frac{(n+3)!}{(x+1)!(n-x+1)!} p^{x} (1-p)^{n-x} \left[1 - 4(p-1/2)^{2} \right]$$
$$p \in [0,1]$$

$$f(x \mid p) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x}$$

$$f(p) = \frac{3}{2} \left[1 - 4(p - 1/2)^2 \right]$$

$$n = 10$$
$$x = 6$$

we need to compute summary measures.

i.e. mode, mean, median, variance of p/x.

Similar to what we do when we have conjugate priors.



n = 10

x = 6

Non-Conjugate Prior For Binomial RVs

Binomial observation *x*:

We can differentiate to maximize the posterior

$$f(p|x) = \frac{1}{4} \frac{(n+3)!}{(x+1)!(n-x+1)!} p^{x} (1-p)^{n-x} \left[1 - 4(p-1/2)^{2} \right]$$
$$p \in [0,1]$$

 $f(x \mid p) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x}$

$$f(p) = \frac{3}{2} \left[1 - 4(p - 1/2)^2 \right]$$

Or we can we can try every value of p that is $\Delta p = 0.0001$ apart and select the one that yields the maximum.

$$\operatorname{ArgMax}_{p} f(p \mid x, n) = 0.5833$$



Binomial observation *x*:

To calculate the expected value E(p/x) of

$$f(p|x) = \frac{1}{4} \frac{(n+3)!}{(x+1)!(n-x+1)!} p^{x} (1-p)^{n-x} \left[1 - 4(p-1/2)^{2} \right],$$

$$p \in [0,1]$$

there are several techniques at our disposal exact integration (calculus) deterministic (*numerical analysis*)

and stochastic (statistical simulation).

$$f(x \mid p) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x}$$

$$f(p) = \frac{3}{2} \left[1 - 4(p - 1/2)^2 \right]$$

$$n = 10$$
$$x = 6$$



Binomial observation *x*:

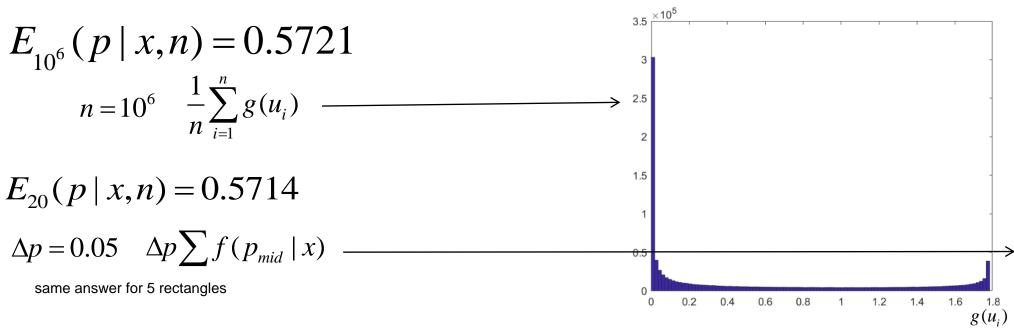
Stochastic and numerical integration.

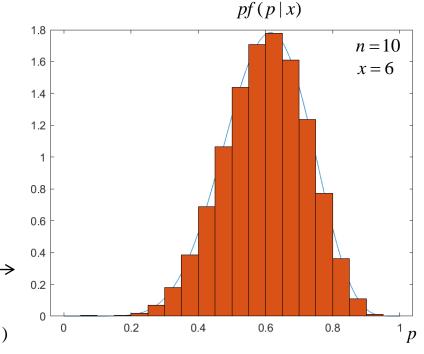
$$f(p|x) = \frac{1}{4} \frac{(n+3)!}{(x+1)!(n-x+1)!} p^{x} (1-p)^{n-x} \left[1 - 4(p-1/2)^{2} \right]$$
$$p \in [0,1]$$

$$f(x \mid p) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x}$$

$$f(p) = \frac{3}{2} \left[1 - 4(p - 1/2)^2 \right]$$

$$n = 10$$
$$x = 6$$







Discussion

Life is more difficult when we have non-conjugate priors!

$$f(x_1,...,x_n) = \int_{\theta} f(x_1,...,x_n \mid \theta) f(\theta) d\theta$$

$$f(\theta \mid x_1,...,x_n) = \frac{f(x_1,...,x_n \mid \theta)f(\theta)}{f(x_1,...,x_n)}$$

$$E(\theta \mid x_1,...,x_n) = \int_{\theta} \theta f(\theta \mid x_1,...,x_n) d\theta$$

Deterministic and Stochastic integration requires $f(x_1,...,x_n)$. It may be extremely difficult to get. May need other methods!

D.B. Rowe

For each n.



Discussion

Questions?

Pencil and Paper Integration

$$E(\theta \mid x_1,...,x_n) = \int_{\theta} \theta f(\theta \mid x_1,...,x_n) d\theta$$

Deterministic Numerical Integration

$$\hat{E}(\theta | \cdot) = \Delta \theta \sum_{i=1}^{n} g(\theta_{i}^{*})$$

$$\Delta \theta = (b-a)/n$$

$$\theta_i^* = a + \Delta\theta / 2 + (i-1)\Delta\theta$$

Stochastic Simulation Integration

$$\overline{E}(\theta \mid \cdot) = \frac{1}{n} \sum_{i=1}^{n} h(u_i)$$

$$a < \theta < b$$

$$0 < \theta < \infty$$

$$-\infty < \theta < \infty$$

$$u = (\theta - a) / (b - a)$$

$$u = 1 / (\theta + 1)$$

$$u = 1/(\theta + 1)$$

$$u = e^{\theta} / (1 + e^{\theta})$$

$$h(u) = g\left(a + (b - a)u\right)(b - a)$$

$$h(u) = g(1/u - 1) / u^2$$

$$h(u) = g(a + (b - a)u)(b - a)$$
 $h(u) = g(1/u - 1)/u^2$ $h(u) = g\left(\ln\left(\frac{u}{1 - u}\right)\right)\frac{1}{u(1 - u)}$



$$f(x \mid p) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x}$$
$$p \in [0,1]$$

1. With the parabolic prior for p (prove integrates to 1)

$$f(p) = \frac{3}{2} \left[1 - 4(p - 1/2)^2 \right] \qquad p \in [0,1]$$

and binomial likelihood for x,

$$f(x \mid p) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x} \qquad x = 0,1,...,n$$

a) prove that the marginal PDF for x is

$$f(x) = 6 \frac{(x+1)(n-x+1)}{(n+3)(n+2)(n+1)}$$
 Want at least a page of calculus. Hint: This is a Beta PDF integral.

b) and that the posterior for p/x is

$$f(p|x) = \frac{1}{4} \frac{(n+3)!}{(x+1)!(n-x+1)!} p^{x} (1-p)^{n-x} \left[1 - 4(p-1/2)^{2} \right]$$



$$f(x | p) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x}$$
$$p \in [0,1]$$

2. With the posterior PDF for *p*

$$f(p|x) = \frac{1}{4} \frac{(n+3)!}{(x+1)!(n-x+1)!} p^{x} (1-p)^{n-x} \left[1 - 4(p-1/2)^{2} \right], \qquad x = 0,1,...,n$$

- a) Calculate the mode and prove or disprove that it is 0.5833. n=10 and x=6 Use deterministic AND stochastic methods.
- b) Calculate the mean and prove or disprove that it is 0.5714. Use both deterministic AND stochastic integration. $\int_{p=0}^{1} pf(p|x)dp?$



$$f(x | p) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x}$$
$$p \in [0,1]$$

 3^* .Use the following non-conjugate prior PDF for p

$$f(p) = \frac{\pi}{2} \cos[\pi(p-1/2)] \qquad p \in [0,1]$$

to combine with the binomial likelihood for x/p.

- a) Try pencil & paper options to prove f(p) integrates to 1.
- b) Try pencil & paper options to calculate f(x) and f(p/x). n=10 and x=6 If you can't do pencil and paper, calculate with deterministic and stochastic methods.
- c) Try pencil & paper options to calculate the mode of f(p/x). n=10 and x=6 If you can't do pencil & paper, calculate with deterministic and stochastic methods.
- d) Try pencil & paper options to calculate the mean of f(p/x). n=10 and x=6 If you can't do pencil & paper, calculate with deterministic and stochastic methods.

* For students in 5790.



$$f(x \mid p) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x}$$

$$p \in [0,1]$$

4**. Make up your own *fun* non-conjugate prior PDF for *p*

$$f(p) = ?????$$
 $p \in [0,1]$

to combine with the binomial likelihood for x/p.

- a) Try pencil & paper options to calculate f(x) and hence f(p/x). If you can't do pencil and paper, calculate computationally. n=10 and x=6
- b) Try pencil & paper options to calculate the mode of f(p/x). If you can't do pencil and paper, calculate computationally. n=10 and x=6
- c) Try pencil & paper options to calculate the mean of f(p/x). If you can't do pencil and paper, calculate computationally. n=10 and x=6

^{**} For students that have had MSSC 6010 and 6020.



$f(x \mid p) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x}$

f(p)

Potential *Fun* Non-Conjugate Priors

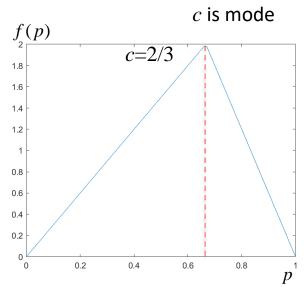
Triangular

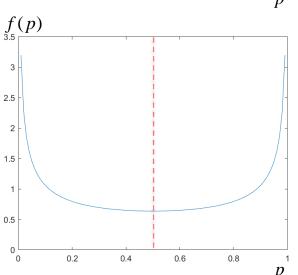
$$f(p) = \begin{cases} \frac{2p}{c}, & 0 \le p \le c \end{cases}$$

$$\frac{2(1-p)}{1-c}, & c \le p \le 1 \end{cases}$$

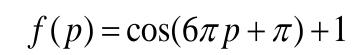
Arcsine U Shaped

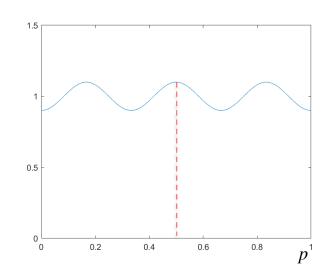
$$f(p) = \frac{1}{\pi \sqrt{p(1-p)}}$$





Roller Coaster





3-Piece Quadratic

$$f(p) = \begin{cases} \frac{27}{2}p^2 & 0 \le p \le \frac{1}{3} \end{cases}$$

$$f(p) = \begin{cases} -27(p - \frac{1}{2})^2 + \frac{9}{4} & \frac{1}{3} \le p \le \frac{2}{3} \end{cases}$$

$$\frac{27}{2}(p-1)^2 & \frac{2}{3} \le p \le 1 \end{cases}$$



$$f(x \mid p) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x}$$

- 5**. Is there a difference between using conjugate vs non-conjugate priors?
 - a) Are the two posterior PDFs and CDFs different? Make plots.
 - b) Are the posterior means and variances different?
 - c) Are the 95th percentiles different?

^{**} For students that have had MSSC 6010 and 6020.