Summarizing the Posterior

Patrick Lam

Outline

Outline

Marginal Distributions

Marginal Distributions

Suppose you have a joint distribution of two parameters, $p(\theta_1, \theta_2)$.

Marginal Distributions

Suppose you have a joint distribution of two parameters, $p(\theta_1, \theta_2)$.

How do you get the marginal distribution of $p(\theta_1)$?

Marginal Distributions

Suppose you have a joint distribution of two parameters, $p(\theta_1, \theta_2)$.

How do you get the marginal distribution of $p(\theta_1)$? Integrate

Marginal Distributions

Suppose you have a joint distribution of two parameters, $p(\theta_1, \theta_2)$.

How do you get the marginal distribution of $p(\theta_1)$? Integrate

$$p(\theta_1) = \int p(\theta_1, \theta_2) d\theta_2$$

Marginal Distributions

Suppose you have a joint distribution of two parameters, $p(\theta_1, \theta_2)$.

How do you get the marginal distribution of $p(\theta_1)$? Integrate

$$p(\theta_1) = \int p(\theta_1, \theta_2) d\theta_2$$

This works for more than two parameters as well:

Marginal Distributions

Suppose you have a joint distribution of two parameters, $p(\theta_1, \theta_2)$.

How do you get the marginal distribution of $p(\theta_1)$? Integrate

$$p(\theta_1) = \int p(\theta_1, \theta_2) d\theta_2$$

This works for more than two parameters as well:

$$p(\theta_1) = \int \int p(\theta_1, \theta_2, \theta_3) d\theta_2 d\theta_3$$

Marginal Distributions

Suppose you have a joint distribution of two parameters, $p(\theta_1, \theta_2)$.

How do you get the marginal distribution of $p(\theta_1)$? Integrate

$$p(\theta_1) = \int p(\theta_1, \theta_2) d\theta_2$$

This works for more than two parameters as well:

$$p(\theta_1) = \int \int p(\theta_1, \theta_2, \theta_3) d\theta_2 d\theta_3$$

We can sometimes do the integrals analytically, but more often than not we want to simulate from the marginal distribution.

Integrating a Grid

Integrating a Grid

Suppose we were doing a grid approximation of a bivariate distribution of θ_1 and θ_2.

Integrating a Grid

Suppose we were doing a grid approximation of a bivariate distribution of θ_1 and θ_2.

Table: Joint Grid

θ_1	θ_2	$p(\theta_1, \theta_2)$
0	0	0.4
0	0.5	0.7
0	1	0.5
0.5	0	0.6
0.5	0.5	0.8
0.5	1	0.8
1	0	0.7
1	0.5	0.6
1	1	0.5

Integrating a Grid

Suppose we were doing a grid approximation of a bivariate distribution of θ_1 and θ_2.

Table: Joint Grid

θ_1	θ_2	$p(\theta_1, \theta_2)$
0	0	0.4
0	0.5	0.7
0	1	0.5
0.5	0	0.6
0.5	0.5	0.8
0.5	1	0.8
1	0	0.7
1	0.5	0.6
1	1	0.5

$$\begin{aligned} p(\theta_1 = 0) &= 0.4 + 0.7 + 0.5 = 1.6 \\ p(\theta_1 = 0.5) &= 0.6 + 0.8 + 0.8 = 2.2 \\ p(\theta_1 = 1) &= 0.7 + 0.6 + 0.5 = 1.8 \end{aligned}$$

Integrating a Grid

Suppose we were doing a grid approximation of a bivariate distribution of θ_1 and θ_2.

Table: Joint Grid

θ_1	θ_2	$p(\theta_1, \theta_2)$
0	0	0.4
0	0.5	0.7
0	1	0.5
0.5	0	0.6
0.5	0.5	0.8
0.5	1	0.8
1	0	0.7
1	0.5	0.6
1	1	0.5

$$
\begin{aligned}
p(\theta_1 = 0) &= 0.4 + 0.7 + 0.5 = 1.6 \\
p(\theta_1 = 0.5) &= 0.6 + 0.8 + 0.8 = 2.2 \\
p(\theta_1 = 1) &= 0.7 + 0.6 + 0.5 = 1.8
\end{aligned}
$$

Table: Marginal Grid

θ_1	$p(\theta_1)$
0	1.6
0.5	2.2
1	1.8

Integrating a Grid

Suppose we were doing a grid approximation of a bivariate distribution of θ_1 and θ_2.

Table: Joint Grid

θ_1	θ_2	$p(\theta_1, \theta_2)$
0	0	0.4
0	0.5	0.7
0	1	0.5
0.5	0	0.6
0.5	0.5	0.8
0.5	1	0.8
1	0	0.7
1	0.5	0.6
1	1	0.5

$$
\begin{aligned}
p(\theta_1 = 0) &= 0.4 + 0.7 + 0.5 = 1.6 \\
p(\theta_1 = 0.5) &= 0.6 + 0.8 + 0.8 = 2.2 \\
p(\theta_1 = 1) &= 0.7 + 0.6 + 0.5 = 1.8
\end{aligned}
$$

Table: Marginal Grid

θ_1	$p(\theta_1)$
0	1.6
0.5	2.2
1	1.8

We can sample from the marginal grid to get simulations from the marginal distribution $p(\theta_1)$.

Integrating Simulations

Integrating Simulations

Now suppose that we have a bunch of simulated draws from the joint distribution (obtained either by sampling from the joint grid or other simulation methods).

Table: Joint Simulations

Draw #	θ_1	θ_2
1	0.5	1
2	1	0
3	1	0
4	0	1
5	0	0.5
6	0.5	0.5
7	1	0.5
8	0.5	0
9	0.5	1
10	0.5	0.5

Integrating Simulations

Now suppose that we have a bunch of simulated draws from the joint distribution (obtained either by sampling from the joint grid or other simulation methods).

Table: Joint Simulations

Draw #	θ_1	θ_2
1	0.5	1
2	1	0
3	1	0
4	0	1
5	0	0.5
6	0.5	0.5
7	1	0.5
8	0.5	0
9	0.5	1
10	0.5	0.5

Table: Marginal Simulations

Draw #	θ_1
1	0.5
2	1
3	1
4	0
5	0
6	0.5
7	1
8	0.5
9	0.5
10	0.5

Integrating Simulations

Now suppose that we have a bunch of simulated draws from the joint distribution (obtained either by sampling from the joint grid or other simulation methods).

Table: Joint Simulations

Draw #	θ_1	θ_2
1	0.5	1
2	1	0
3	1	0
4	0	1
5	0	0.5
6	0.5	0.5
7	1	0.5
8	0.5	0
9	0.5	1
10	0.5	0.5

Table: Marginal Simulations

Draw #	θ_1
1	0.5
2	1
3	1
4	0
5	0
6	0.5
7	1
8	0.5
9	0.5
10	0.5

To obtain simulations from the marginal distribution, we simply drop the simulated values from the other parameter(s).

Outline

Summarizing the Posterior

Summarizing the Posterior

Suppose we now have either analytical or simulated posterior.

Summarizing the Posterior

Suppose we now have either analytical or simulated posterior.

What are some of the ways we can summarize the posterior?

Summarizing the Posterior

Suppose we now have either analytical or simulated posterior.

What are some of the ways we can summarize the posterior?

Consider the previous example of the beta-binomial model with a Beta(1,1) prior (uniform) and a Beta$(y + \alpha, n - y + \beta)$ posterior, where $n = 500$, $y = 285$, $\alpha = 1$, and $\beta = 1$.

Posterior Mean

Posterior Mean

Analytically:

$$\frac{y+\alpha}{(y+\alpha)+(n-y+\beta)} = \frac{286}{286+216} \approx 0.57$$

Posterior Mean

Analytically:

$$\frac{y+\alpha}{(y+\alpha)+(n-y+\beta)} = \frac{286}{286+216} \approx 0.57$$

Simulation:

```
> a <- 1
> b <- 1
> posterior.unif.prior <- rbeta(10000, shape1 = 285 + a, shape2 = 500 -
+     285 + b)
> mean(posterior.unif.prior)

[1] 0.5699
```

Posterior Variance

Posterior Variance

Analytically:

$$\frac{(y+\alpha)(n-y+\beta)}{[(y+\alpha)+(n-y+\beta)]^2[(y+\alpha)+(n-y+\beta)+1]} \approx 0.00049$$

Posterior Variance

Analytically:

$$\frac{(y+\alpha)(n-y+\beta)}{[(y+\alpha)+(n-y+\beta)]^2[(y+\alpha)+(n-y+\beta)+1]} \approx 0.00049$$

Simulation:

```
> var(posterior.unif.prior)
[1] 0.0004832
```

Posterior Variance

Analytically:

$$\frac{(y+\alpha)(n-y+\beta)}{[(y+\alpha)+(n-y+\beta)]^2[(y+\alpha)+(n-y+\beta)+1]} \approx 0.00049$$

Simulation:

```
> var(posterior.unif.prior)
[1] 0.0004832
```

Take the square root for standard deviation.

Posterior Probability $0.5 < \pi < 0.6$

Posterior Probability $0.5 < \pi < 0.6$

Analytically:

$$\int_{0.5}^{0.6} \frac{\Gamma((y+\alpha)+(n-y+\beta))}{\Gamma(y+\alpha)\Gamma(n-y+\beta)} \pi^{(y+\alpha-1)}(1-\pi)^{(n-y+\beta-1)} d\pi \approx 0.91$$

Posterior Probability $0.5 < \pi < 0.6$

Analytically:

$$\int_{0.5}^{0.6} \frac{\Gamma((y+\alpha)+(n-y+\beta))}{\Gamma(y+\alpha)\Gamma(n-y+\beta)} \pi^{(y+\alpha-1)}(1-\pi)^{(n-y+\beta-1)} d\pi \approx 0.91$$

Simulation:

```
> mean(posterior.unif.prior > 0.5 & posterior.unif.prior < 0.6)
[1] 0.9114
```

Central 95% Credible Interval

Central 95% Credible Interval

In Bayesian statistics, we use the terms *credible sets* and *credible intervals* rather than confidence intervals.

Central 95% Credible Interval

In Bayesian statistics, we use the terms *credible sets* and *credible intervals* rather than confidence intervals.

Find the central interval that contains 95% of the area of the posterior.

Central 95% Credible Interval

In Bayesian statistics, we use the terms *credible sets* and *credible intervals* rather than confidence intervals.

Find the central interval that contains 95% of the area of the posterior.

Simulation:

```
> quantile(posterior.unif.prior, probs = c(0.025, 0.975))

  2.5%  97.5%
0.5269 0.6125
```

95% Highest Posterior Density Region

95% Highest Posterior Density Region

Find the smallest interval(s) that contains 95% of the area of the posterior.

95% Highest Posterior Density Region

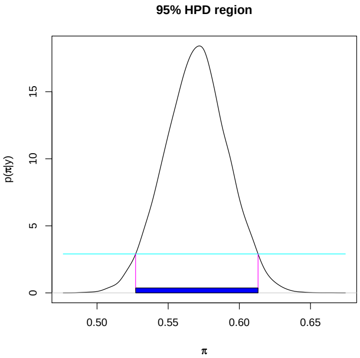
Find the smallest interval(s) that contains 95% of the area of the posterior.

Simulation:

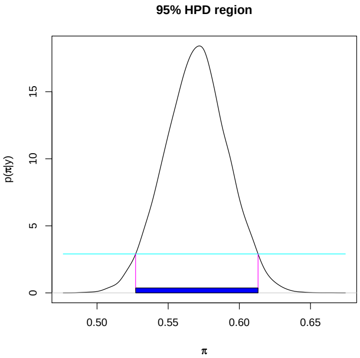
```
> library(hdrcde)
> hdr(posterior.unif.prior, prob = 95)$hdr

      [,1]   [,2]
95% 0.5271 0.6132
```

```
> hdr.den(posterior.unif.prior, prob = 95, main = "95% HPD region",
+     xlab = expression(pi), ylab = expression(paste("p(", pi,
+         "|y)")))
```



```
> hdr.den(posterior.unif.prior, prob = 95, main = "95% HPD region",
+     xlab = expression(pi), ylab = expression(paste("p(", pi,
+         "|y)")))
```



Very similar to central 95% credible interval for Beta posterior.

What about for this posterior?