

Introduction to Bayesian Statistics

Shin-ichi Mayekawa

mayekawa@hum.titech.ac.jp

www.ms.hum.titech.ac.jp

Outline

1. Introduction to Bayes Theorem
2. Binomial Model and Poisson Model
3. Normal Model
4. Normal Regression Model
5. Bayesian Numerical Calculations (Markov Chain Monte Carlo)
6. Application of MCMC
7. Model Selection
8. miscellaneous topics

2

Multiple Independent Observations

- If two events are mutually independent the joint probability of two events can be written as

$$\begin{aligned} \Pr(A_{i1} \& A_{i2}|B_j) &= \Pr(A_{i1}, A_{i2}|B_j) \\ &= \Pr(A_{i1}|B_j) \times \Pr(A_{i2}|B_j) \end{aligned}$$

3

Multiple Independent Observations

$\Pr(A_{i1}, A_{i2} | B_j) = \Pr(A_{i1} | B_j) \Pr(A_{i2} | B_j)$

CPT or Model Probability

	B_1	...	B_j	...	B_b
A_1, A_1	$\Pr(A_1 B_1)\Pr(A_1 B_1)$...	$\Pr(A_1 B_j)\Pr(A_1 B_j)$...	$\Pr(A_1 B_b)\Pr(A_1 B_b)$
A_1, A_2	$\Pr(A_1 B_1)\Pr(A_2 B_1)$...	$\Pr(A_1 B_j)\Pr(A_2 B_j)$...	$\Pr(A_1 B_b)\Pr(A_2 B_b)$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
A_{i1}, A_{i2}	$\Pr(A_{i1} B_1)\Pr(A_{i2} B_1)$...	$\Pr(A_{i1} B_j)\Pr(A_{i2} B_j)$...	$\Pr(A_{i1} B_b)\Pr(A_{i2} B_b)$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
A_a, A_a	$\Pr(A_a B_1)\Pr(A_a B_1)$...	$\Pr(A_a B_j)\Pr(A_a B_j)$...	$\Pr(A_a B_b)\Pr(A_a B_b)$
$\Pr(B)$	$\Pr(B_1)$...	$\Pr(B_j)$...	$\Pr(B_b)$

If $A_i, i = 1, 2, \dots, a$ are mutually exclusive and exhaustive

$$\sum_{i1=1}^a \sum_{i2=1}^a \Pr(A_{i1}, A_{i2}|B_j) = 1$$

Joint Probability

$$\Pr(A_{i1}, A_{i2}, B_j) = \Pr(A_{i1}, A_{i2}|B_j) \times \Pr(B_j)$$

Posterior Probability

$$\Pr(B_j|A_{i1}, A_{i2}) = \frac{1}{\sum_{j=1}^b \Pr(A_{i1}, A_{i2}, B_j)} \Pr(A_{i1}, A_{i2}, B_j)$$

4

Multiple Independent Observations

Joint Probability Table

$$\Pr(A_{i1}, A_{i2}, B_j) = \Pr(A_{i1}, A_{i2} | B_j) \times \Pr(B_j)$$

	B_1	...	B_j	...	B_b	$\Pr(A, A)$
A_1, A_1	$\Pr(A_1 B_1)\Pr(A_1 B_1)\Pr(B_1)$...	$\Pr(A_1 B_j)\Pr(A_1 B_j)\Pr(B_j)$...	$\Pr(A_1 B_b)\Pr(A_1 B_b)\Pr(B_b)$	$\Pr(A_1, A_1)$
A_1, A_2	$\Pr(A_1 B_1)\Pr(A_2 B_1)\Pr(B_1)$...	$\Pr(A_1 B_j)\Pr(A_2 B_j)\Pr(B_j)$...	$\Pr(A_1 B_b)\Pr(A_2 B_b)\Pr(B_b)$	$\Pr(A_1, A_2)$
...
A_{i1}, A_{i2}	$\Pr(A_{i1} B_1)\Pr(A_{i2} B_1)\Pr(B_1)$...	$\Pr(A_{i1} B_j)\Pr(A_{i2} B_j)\Pr(B_j)$...	$\Pr(A_{i1} B_b)\Pr(A_{i2} B_b)\Pr(B_b)$	$\Pr(A_{i1}, A_{i2})$
...
A_a, A_a	$\Pr(A_a B_1)\Pr(A_a B_1)\Pr(B_1)$...	$\Pr(A_a B_j)\Pr(A_a B_j)\Pr(B_j)$...	$\Pr(A_a B_b)\Pr(A_a B_b)\Pr(B_b)$	$\Pr(A_a, A_a)$
$\Pr(B)$	$\Pr(B_1)$...	$\Pr(B_j)$...	$\Pr(B_b)$	

Posterior Probability

$$\Pr(B_j | A_{i1}, A_{i2}) = \frac{1}{\sum_{j=1}^b \Pr(A_{i1}, A_{i2}, B_j)} \Pr(A_{i1}, A_{i2}, B_j)$$

Multiple Independent Observations

Posterior Probability

$$\Pr(B_j | A_{i1}, A_{i2}) = \frac{1}{\sum_{j=1}^b \Pr(A_{i1}, A_{i2}, B_j)} \Pr(A_{i1}, A_{i2}, B_j)$$

	B_1	...	B_j	...	B_b	$\Pr(A_{i1}, A_{i2})$
A_1, A_1	$\frac{\Pr(A_1 B_1)\Pr(A_1 B_1)\Pr(B_1)}{\Pr(A_1, A_1)}$...	$\frac{\Pr(A_1 B_j)\Pr(A_1 B_j)\Pr(B_j)}{\Pr(A_1, A_1)}$...	$\frac{\Pr(A_1 B_b)\Pr(A_1 B_b)\Pr(B_b)}{\Pr(A_1, A_1)}$	$\Pr(A_1, A_1)$
A_1, A_2	$\frac{\Pr(A_1 B_1)\Pr(A_2 B_1)\Pr(B_1)}{\Pr(A_1, A_2)}$...	$\frac{\Pr(A_1 B_j)\Pr(A_2 B_j)\Pr(B_j)}{\Pr(A_1, A_2)}$...	$\frac{\Pr(A_1 B_b)\Pr(A_2 B_b)\Pr(B_b)}{\Pr(A_1, A_2)}$	$\Pr(A_1, A_2)$
...
A_{i1}, A_{i2}	$\frac{\Pr(A_{i1} B_1)\Pr(A_{i2} B_1)\Pr(B_1)}{\Pr(A_{i1}, A_{i2})}$...	$\frac{\Pr(A_{i1} B_j)\Pr(A_{i2} B_j)\Pr(B_j)}{\Pr(A_{i1}, A_{i2})}$...	$\frac{\Pr(A_{i1} B_b)\Pr(A_{i2} B_b)\Pr(B_b)}{\Pr(A_{i1}, A_{i2})}$	$\Pr(A_{i1}, A_{i2})$
...
A_a, A_a	$\frac{\Pr(A_a B_1)\Pr(A_a B_1)\Pr(B_1)}{\Pr(A_a, A_a)}$...	$\frac{\Pr(A_a B_j)\Pr(A_a B_j)\Pr(B_j)}{\Pr(A_a, A_a)}$...	$\frac{\Pr(A_a B_b)\Pr(A_a B_b)\Pr(B_b)}{\Pr(A_a, A_a)}$	$\Pr(A_a, A_a)$
$\Pr(B)$	$\Pr(B_1)$...	$\Pr(B_j)$...	$\Pr(B_b)$	

Coin Tossed Twice

CPT of A given B
and Marginal of B
B1:1/2 B2:2/3

A1:HH	.25000	.44444
A2:HT	.25000	.22222
A3:TH	.25000	.22222
A4:TT	.25000	.11111

Bm .50000 .50000

Joint Probability of A and B
and marginal of A
B1:1/2 B2:2/3 Am

A1:HH	.12500	.22222	.34722
A2:HT	.12500	.11111	.23611
A3:TH	.12500	.11111	.23611
A4:TT	.12500	.05556	.18056

Bm .50000 .50000 1.0000

Posterior Probability

B1:1/2 B2:2/3 sum

A1:HH	.36000	.64000	1.
A2:HT	.52941	.47059	1.
A3:TH	.52941	.47059	1.
A4:TT	.69231	.30769	1.

Bm .50000 .50000

Note that the posterior odds of not Fair to Fair, given Head, is larger with this example.

$$\frac{\Pr(\text{notFair}|\text{Head})}{\Pr(\text{Fair}|\text{Head})} = \frac{.571}{.429} = 1.333 < 1.778 = \frac{.640}{.360} = \frac{\Pr(\text{notFair}|\text{Head, Head})}{\Pr(\text{Fair}|\text{Head, Head})}$$

Multiple observations provide more information on B.

Coin Tossed Twice

CPT of A given B
and Marginal of B
B1:1/2 B2:2/3

A1:HH	.25000	.44444
A2:HT	.25000	.22222
A3:TH	.25000	.22222
A4:TT	.25000	.11111

Bm .50000 .50000

Joint Probability of A and B
and marginal of A
B1:1/2 B2:2/3 Am

B1:1/2 B2:2/3 Am

A1:HH	.12500	.22222	.34722
A2:HT	.12500	.11111	.23611
A3:TH	.12500	.11111	.23611
A4:TT	.12500	.05556	.18056

Bm .50000 .50000 1.0000

Posterior Probability

B1:1/2 B2:2/3 sum

A1:HH	.36000	.64000	1.
A2:HT	.52941	.47059	1.
A3:TH	.52941	.47059	1.
A4:TT	.69231	.30769	1.

Bm .50000 .50000

Note also that, $\Pr(\text{Fair}|\text{Head, Tail}) = \Pr(\text{Fair}|\text{Tail, Head})$.

Number of Heads is sufficient for the inference.

Two Test Scores

Test Score (two scores)				Joint Probability of A and B and the Marginal					Posterior Probability			
Conditional Probability of A given B (Model)				PROB2	B1:L	B2:M	B3:H	Am	PROB3	B1:L	B2:M	B3:H
PROB1	B1:L	B2:M	B3:H									
A1:00	.01000	.01000	.00000	A1:00	.00250	.00500	.00000	.00750	A1:00	.33333	.66667	.00000
A2:01	.03000	.01000	.00000	A2:01	.00750	.00500	.00000	.01250	A2:01	.60000	.40000	.00000
A3:02	.04000	.02000	.00000	A3:02	.01000	.01000	.00000	.02000	A3:02	.50000	.50000	.00000
A4:03	.02000	.03000	.00000	A4:03	.00500	.01500	.00000	.02000	A4:03	.25000	.75000	.00000
A5:04	.00000	.02000	.00000	A5:04	.00000	.01000	.00000	.01000	A5:04	.00000	1.0000	.00000
A6:05	.00000	.01000	.00000	A6:05	.00000	.00500	.00000	.00500	A6:05	.00000	1.0000	.00000
A7:10	.03000	.01000	.00000	A7:10	.00750	.00500	.00000	.01250	A7:10	.60000	.40000	.00000
A8:11	.09000	.01000	.00000	A8:11	.02250	.00500	.00000	.02750	A8:11	.81818	.18182	.00000
A9:12	.12000	.02000	.00000	A9:12	.03000	.01000	.00000	.04000	A9:12	.75000	.25000	.00000
A10:13	.06000	.03000	.00000	A10:13	.01500	.01500	.00000	.03000	A10:13	.50000	.50000	.00000
A11:14	.00000	.02000	.00000	A11:14	.00000	.01000	.00000	.01000	A11:14	.00000	1.0000	.00000
A12:15	.00000	.01000	.00000	A12:15	.00000	.00500	.00000	.00500	A12:15	.00000	1.0000	.00000
A13:20	.04000	.02000	.00000	A13:20	.01000	.01000	.00000	.02000	A13:20	.50000	.50000	.00000
A14:21	.12000	.02000	.00000	A14:21	.03000	.01000	.00000	.04000	A14:21	.75000	.25000	.00000
A15:22	.16000	.04000	.01000	A15:22	.04000	.02000	.00250	.06250	A15:22	.64000	.32000	.04000
A16:23	.08000	.06000	.02000	A16:23	.02000	.03000	.00500	.05500	A16:23	.36364	.54545	.09091
A17:24	.00000	.04000	.04000	A17:24	.00000	.02000	.01000	.03000	A17:24	.00000	.66667	.33333
A18:25	.00000	.02000	.03000	A18:25	.00000	.01000	.00750	.01750	A18:25	.00000	.57143	.42857
A19:30	.02000	.03000	.00000	A19:30	.00500	.01500	.00000	.02000	A19:30	.25000	.75000	.00000
A20:31	.06000	.03000	.00000	A20:31	.01500	.01500	.00000	.03000	A20:31	.50000	.50000	.00000
A21:32	.08000	.06000	.02000	A21:32	.02000	.03000	.00500	.05500	A21:32	.36364	.54545	.09091
A22:33	.04000	.09000	.04000	A22:33	.01000	.04500	.01000	.06500	A22:33	.15385	.69231	.15385
A23:34	.00000	.08000	.08000	A23:34	.00000	.03000	.02000	.05000	A23:34	.00000	.60000	.40000
A24:35	.00000	.03000	.06000	A24:35	.00000	.01500	.03000	.04500	A24:35	.00000	.50000	.50000
A25:40	.00000	.02000	.00000	A25:40	.00000	.01000	.00000	.01000	A25:40	.00000	1.0000	.00000
A26:41	.00000	.02000	.00000	A26:41	.00000	.01000	.00000	.01000	A26:41	.00000	1.0000	.00000
A27:42	.00000	.04000	.04000	A27:42	.00000	.02000	.01000	.03000	A27:42	.00000	.66667	.33333
A28:43	.00000	.06000	.08000	A28:43	.00000	.03000	.02000	.05000	A28:43	.00000	.60000	.40000
A29:44	.00000	.04000	.16000	A29:44	.00000	.02000	.04000	.06000	A29:44	.00000	.33333	.66667
A30:45	.00000	.02000	.12000	A30:45	.00000	.01000	.03000	.04000	A30:45	.00000	.25000	.75000
A31:50	.00000	.01000	.00000	A31:50	.00000	.00500	.00000	.00500	A31:50	.00000	1.0000	.00000
A32:51	.00000	.01000	.00000	A32:51	.00000	.00500	.00000	.00500	A32:51	.00000	1.0000	.00000
A33:52	.00000	.02000	.03000	A33:52	.00000	.01000	.00750	.01750	A33:52	.00000	.57143	.42857
A34:53	.00000	.03000	.06000	A34:53	.00000	.01500	.01500	.03000	A34:53	.00000	.50000	.50000
A35:54	.00000	.02000	.12000	A35:54	.00000	.01000	.03000	.04000	A35:54	.00000	.25000	.75000
A36:55	.00000	.01000	.09000	A36:55	.00000	.00500	.02250	.02750	A36:55	.00000	.18182	.81818
Bm	.25000	.50000	.25000	Bm	.25000	.50000	.25000	1.0000	Bm	.25000	.50000	.25000

9

Two Test Scores

Note that, with this example (CPT) the sum of two test scores is not sufficient. For example the posterior probability of B given sum = 6 are shown below.

Conditional Probability of A given B (Model)				Joint Probability of A and B and the Marginal					Posterior Probability			
PROB1	B1:L	B2:M	B3:H	PROB2	B1:L	B2:M	B3:H	Am	PROB3	B1:L	B2:M	B3:H
A12:15	.00000	.01000	.00000	A12:15	.00000	.00500	.00000	.00500	A12:15	.00000	1.0000	.00000
A32:51	.00000	.01000	.00000	A32:51	.00000	.00500	.00000	.00500	A32:51	.00000	1.0000	.00000
A17:24	.00000	.04000	.04000	A17:24	.00000	.02000	.01000	.03000	A17:24	.00000	.66667	.33333
A27:42	.00000	.04000	.04000	A27:42	.00000	.02000	.01000	.03000	A27:42	.00000	.66667	.33333
A22:33	.04000	.09000	.04000	A22:33	.01000	.04500	.01000	.06500	A22:33	.15385	.69231	.15385

If CPT is proportional to the PDF of the Normal distribution, sum is the sufficient statistics.

10

Multiple Observations in General

Conditional Probability

$$\Pr(A_{i1}, A_{i2}, \dots, A_{in} | B_j) = \prod_{k=1}^n \Pr(A_{ik} | B_j)$$

Posterior Probability

$$\Pr(B_j | A_{i1}, A_{i2}, \dots, A_{in}) = \frac{\Pr(A_{i1}, A_{i2}, \dots, A_{in} | B_j) \Pr(B_j)}{\sum_{j=1}^b \Pr(A_{i1}, A_{i2}, \dots, A_{in} | B_j) \Pr(B_j)}$$

$$\Pr(B_j | A_{i1}, A_{i2}, \dots, A_{in}) = \frac{(\prod_{k=1}^n \Pr(A_{ik} | B_j)) \Pr(B_j)}{\sum_{j=1}^b ((\prod_{k=1}^n \Pr(A_{ik} | B_j)) \Pr(B_j))} \propto (\prod_{k=1}^n \Pr(A_{ik} | B_j)) \Pr(B_j)$$

11

Numerical Category Values

- When the categories have numerical value, CPT can be expressed by a formula $f(\cdot|\cdot)$.

Let A and B be the random variables whose values are α_i and β_j .

CPT

$$\Pr(A_i | B_j) = \Pr(A = \alpha_i | B = \beta_j) = f(\alpha_i | \beta_j)$$

12

Coin Toss: Bernoulli Distribution

A: Indicator Variable

$$A = \begin{cases} 1 & \text{Head} \\ 0 & \text{Tail} \end{cases} \quad \alpha = \begin{cases} 1 & \text{Head} \\ 0 & \text{Tail} \end{cases}$$

B: Success Probability

$$0 \leq \beta_j \leq 1$$

$$\Pr(A = 1|B = \beta_j) = \beta_j \quad \& \quad \Pr(A = 0|B = \beta_j) = 1 - \beta_j$$

CPT: Bernoulli Distribution

$$\Pr(A = \alpha|B = \beta_j) = f(\alpha|\beta_j) = \beta_j^\alpha (1 - \beta_j)^{1-\alpha}$$

13

Coin Toss: Binomial Distribution

Two observations: $A_{(1)}$ and $A_{(2)}$

$$\begin{aligned} \Pr(A_{(1)} = \alpha_{(1)} \& A_{(2)} = \alpha_{(2)} | B_j = \beta_j) &= \Pr(A_{(1)} = \alpha_{(1)} | B_j = \beta_j) \times \Pr(A_{(2)} = \alpha_{(2)} | B_j = \beta_j) \\ &= f(\alpha_{(1)} | \beta_j) \times f(\alpha_{(2)} | \beta_j) \\ &= \beta_j^{\alpha_{(1)}} (1 - \beta_j)^{1-\alpha_{(1)}} \times \beta_j^{\alpha_{(2)}} (1 - \beta_j)^{1-\alpha_{(2)}} \\ &= \beta_j^{\alpha_{(1)} + \alpha_{(2)}} (1 - \beta_j)^{2 - (\alpha_{(1)} + \alpha_{(2)})} \end{aligned}$$

N observations: $A_{(k)}, k = 1, 2, \dots, N$

$$\begin{aligned} \Pr(A_{(1)} = \alpha_{(1)} \& A_{(2)} = \alpha_{(2)}, \dots, A_{(N)} = \alpha_{(N)} | B_j = \beta_j) \\ &= \prod_{k=1}^N f(A_{(k)} = \alpha_{(k)} | \beta_j) \\ &= \prod_{k=1}^N \beta_j^{\alpha_{(k)}} (1 - \beta_j)^{1-\alpha_{(k)}} \\ &= \beta_j^{\sum_{k=1}^N \alpha_{(k)}} (1 - \beta_j)^{N - \sum_{k=1}^N \alpha_{(k)}} \\ &\propto \Pr(\text{number of Heads out of } N = \sum_{k=1}^N \alpha_{(k)} | B_j = \beta_j) \end{aligned}$$

14

Coin Toss: Binomial Distribution

Let Y be the random number representing the number of Heads out of N tosses.

There are $\binom{N}{y}$ sequences of $A_{(k)}, k = 1, 2, \dots, N$, which result in y Heads, i.e.

$$\sum_{k=1}^N \alpha_{(k)} = y$$

CPT: Binomial Distribution

$$\begin{aligned} \Pr(Y = y | B_j = \beta_j) &= \text{number of combinations} \times \\ &\Pr\left(\sum_{k=1}^N \alpha_{(k)} = y | B_j = \beta_j\right) \\ &= f(y | \beta_j) \\ &= \binom{N}{y} \times \beta_j^y (1 - \beta_j)^{N-y} \end{aligned}$$

Note that when $N=1$, Binomial = Bernoulli.

15

Binomial Distribution

CPT:

$$\Pr(Y = y | B = \beta_j) = \binom{N}{y} \times \beta_j^y (1 - \beta_j)^{N-y}$$

$Y = \{0, 1, 2, \dots, N\}$

Joint Probability

$$\begin{aligned} \Pr(Y = y, B = \beta_j) &= \Pr(Y = y | B = \beta_j) \times \Pr(B = \beta_j) \\ &= \binom{N}{y} \times \beta_j^y (1 - \beta_j)^{N-y} \Pr(B = \beta_j) \end{aligned}$$

Posterior Probability

$$\Pr(B = \beta_j | Y = y) \propto \Pr(Y = y, B = \beta_j) \propto \beta_j^y (1 - \beta_j)^{N-y} \Pr(B = \beta_j)$$

$$\begin{aligned} \Pr(B = \beta_j | Y = y) &= \frac{\Pr(Y = y, B = \beta_j)}{\sum_{j=1}^b \Pr(Y = y, B = \beta_j)} \\ &= \frac{\beta_j^y (1 - \beta_j)^{N-y} \Pr(B = \beta_j)}{\sum_{j=1}^b \beta_j^y (1 - \beta_j)^{N-y} \Pr(B = \beta_j)} \end{aligned}$$

16

Extending B to continuous RV

- About continuous random variables

17

Discrete Random Variables

- probability function

$$f(x) = \Pr(X = x) \quad x_0 \leq X \leq x_1$$

$$\sum_{x=x_0}^{x_1} f(x) = 1$$

- distribution function

$$\Pr(X \leq x) = F(x) = \sum_{z=x_0}^x f(z)$$

$$\Pr(a < X \leq b) = \sum_{x=x_0}^b f(x) - \sum_{x=x_0}^a f(x) = F(b) - F(a)$$

$$F^{-1}(\alpha) = \min x \in R : p \leq F(x)$$

18

Continuous Random Variables

- probability distribution function and density (pdf)

$$\Pr(X \leq x) = F(x) = \int_{-\infty}^x f(z) dz \quad \int_X f(x) dx = 1$$

- probabilities

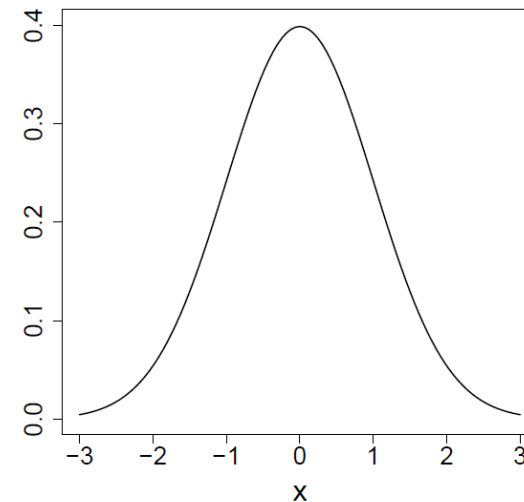
$$\Pr(a \leq X \leq b) = \int_a^b f(x) dx$$

$$\Pr(X = x) = 0$$

$$\Pr(x - \varepsilon < X \leq x + \varepsilon) \propto f(x)$$

19

pdf



20

Joint PDF, Conditional PDF

- joint pdf

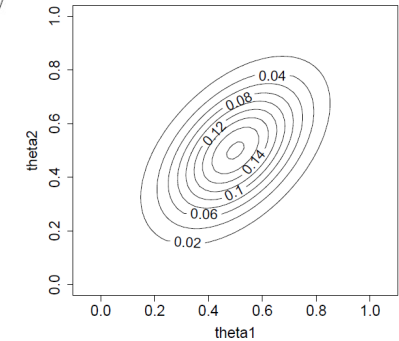
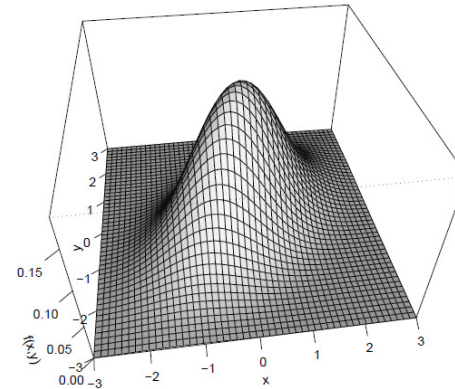
$$\int_Y \int_X f(x, y) dx dy = 1$$

- conditional pdf

$$f(x|y) = \frac{f(x, y)}{f(y)} \quad f(y|x) = \frac{f(x, y)}{f(x)}$$

$$\int_X f(x|y) dx = 1, \quad \int_Y f(y|x) dy = 1$$

Joint PDF



		B is discrete		B is continuous
		categorical	numerical	numerical
A is discrete	categorical	A	D	G
	numerical	B	E	H
A is continuous	numerical	C	F	I
			cpt=pdf, h=pf	cpt=pdf, h=pdf

- A: simplest case
- B: test score
- C: body height M/F
- D:
- E: coin toss F/MF
- F: battery test
- G:
- H: Bernoulli/Binomial
- I: Normal

Bayes Theorem (continuous)

- Bayes theorem

$$h(\theta|y) = \frac{f(\theta, y)}{f(y)} = \frac{f(y|\theta)h(\theta)}{\int_{\Theta} f(y|\theta)h(\theta)d\theta}$$

- proportionality

$$h(\theta|y) = \frac{1}{\int_{\Theta} f(y|\theta)h(\theta)d\theta} f(y|\theta)h(\theta) \propto f(y|\theta)h(\theta)$$

- Posterior pdf is proportional to Likelihood x Prior pdf

Bayesian Inference

- Specify the model distribution (likelihood.)

$$f(y|\theta) \quad \theta \text{ is called the parameter of the model.}$$

- Express your prior belief on the model parameter as the prior distribution. (prior pdf)

$$h(\theta)$$

- Update your prior belief on the model parameter by combining the prior pdf and the likelihood.

$$h(\theta|y) = \frac{1}{\int_{\Theta} f(y|\theta)h(\theta)d\theta} f(y|\theta)h(\theta)$$

$$\propto f(y|\theta)h(\theta)$$

25

- Going back to Binomial Likelihood

Based on the # of successes out of N trials, we must make inference on the parameter q which is the probability of success.

$$\beta_j, j = 1, 2, \dots, b, \quad b \rightarrow +\infty$$

$$\beta_j \equiv \theta$$

- How do we specify the prior distribution?

26

Non-informative Prior

- If we do not have any knowledge on θ

$$h(\theta) = 1$$

- Posterior distribution of θ

$$h(\theta|y) \propto f(y|\theta)h(\theta) = f(y|\theta)$$

$$= \theta^y(1 - \theta)^{N-y}$$

$$\int_0^1 \theta^y(1 - \theta)^{N-y}d\theta = B(y + 1, N - y + 1)$$

$$h(\theta|y) = \frac{1}{B(y + 1, N - y + 1)} \theta^y(1 - \theta)^{N-y}$$

pdf of Beta distribution

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}$$

27

Beta Distribution

- pdf: $X \sim \text{Beta}(\alpha, \beta)$

$$f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} \times x^{\alpha-1}(1 - x)^{\beta-1}$$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

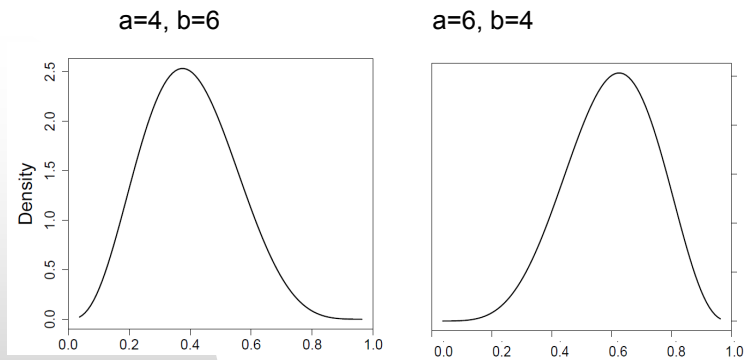
- moments

$$\mathcal{E}(X) = \frac{\alpha}{\alpha + \beta}, \quad \text{var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

$$\text{mode}(X) = \frac{\alpha - 1}{\alpha + \beta - 2} \quad (\alpha > 1 \& \beta > 1)$$

28

Beta Distribution



29

Identifying Posterior Distribution

If we know

$$h(\theta|y) \propto \theta^{y+\alpha_0-1}(1-\theta)^{N-y+\beta_0-1}$$

then we know

$$\theta | \sim \text{Beta}(y + \alpha_0, N - y + \beta_0)$$

30

Summary of Posterior Distribution

Posterior Summary

$$\begin{aligned} \mathcal{E}(\theta|y) &= \frac{y+1}{N+2} \\ \text{mode}(\theta|y) &= \frac{y}{N} \\ \text{var}(\theta|y) &= \frac{(y+1)(N-y+1)}{(N+2)^2(N+3)} \end{aligned}$$

31

How to Choose Point Estimate

- Expected loss

Choose d which minimizes:

$$\mathcal{E}L(d|y) = \mathcal{E}_{\Theta}(L(d, \theta)|y) = \int_{\Theta} L(d, \theta)h(\theta|y)d\theta$$

- Various types of expected loss

Squared Error Loss = $(d - \theta)^2$

minimized by:

mean

Bilinear Error Loss = $\begin{cases} a(d - \theta) & \text{if } d \leq \theta \\ b(\theta - d) & \text{if } d > \theta \end{cases}$

median

0-1 Error Loss = $\begin{cases} 0 & \text{if } |d - \theta| \leq b \\ 1 & \text{if } |d - \theta| > b \end{cases}$

mode

32

Point Estimate

- Squared error loss

$$\begin{aligned}\mathcal{E}L(d|\mathbf{y}) &= \int_{\Theta} (d - \theta)^2 h(\theta|\mathbf{y}) d\theta = \int_{\Theta} (d^2 - 2d\theta + \theta^2) h(\theta|\mathbf{y}) d\theta \\ &= d^2 - 2d \int_{\Theta} \theta h(\theta|\mathbf{y}) d\theta + \int_{\Theta} \theta^2 h(\theta|\mathbf{y}) d\theta\end{aligned}$$

- Minimum by differentiation

$$d = \int_{\Theta} \theta h(\theta|\mathbf{y}) d\theta = \mathcal{E}(\theta|\mathbf{y})$$

33

Natural Conjugate Prior

- If you can express your prior knowledge of θ by a Beta distribution:

$$\begin{aligned}h(\theta) &= \frac{1}{B(\alpha_0, \beta_0)} \theta^{\alpha_0-1} (1-\theta)^{\beta_0-1} \\ &\propto \theta^{\alpha_0-1} (1-\theta)^{\beta_0-1}\end{aligned}$$

- Posterior is also a Beta distribution.

$$\begin{aligned}h(\theta|\mathbf{y}) &\propto \theta^y (1-\theta)^{N-y} \times \theta^{\alpha_0-1} (1-\theta)^{\beta_0-1} \\ &= \theta^{y+\alpha_0-1} (1-\theta)^{N-y+\beta_0-1} \\ &= \text{pdf of } \text{Beta}(y + \alpha_0, N - y + \beta_0)\end{aligned}$$

34

Summary of Posterior Distribution

- Posterior summary in general

$$\mathcal{E}(\theta|\mathbf{y}) = \frac{\alpha_0 + y}{\alpha_0 + \beta_0 + N},$$

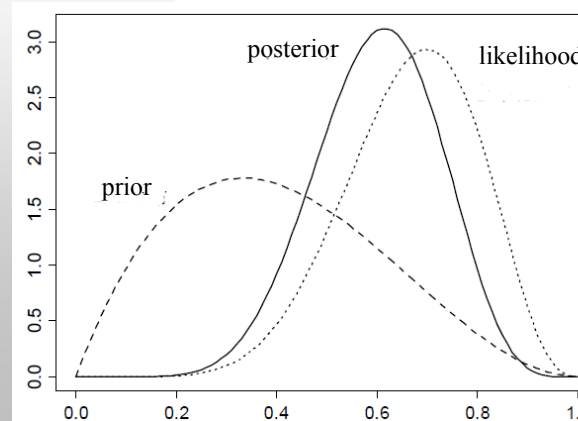
$$\text{mode}(\theta|\mathbf{y}) = \frac{\alpha_0 + y - 1}{\alpha_0 + \beta_0 + N - 2}$$

$$\text{var}(\theta|\mathbf{y}) = \frac{(\alpha_0 + y)\beta_0 + N - y}{(\alpha_0 + \beta_0 + N)^2(\alpha_0 + \beta_0 + N + 1)}$$

35

Example

- Ten tosses of a coin:
H T H H T H H H H T $N=10, y=7$
- Prior: $B(2,3)$
- Posterior: $B(9,6) = B(2+7, 3+10-7)$



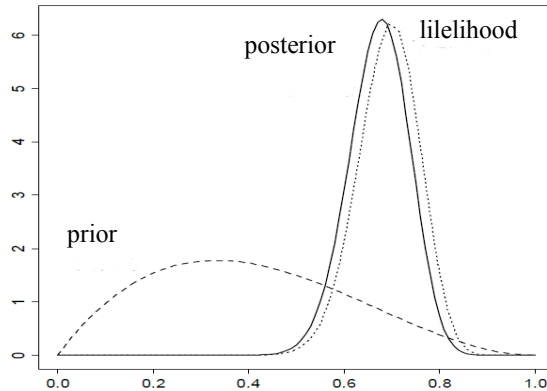
	prior	posterior
mean	0.40	0.60
mode	0.33	0.62
std	0.20	0.12

Posterior var(std) is smaller than prior var.

36

Example

- 50 tosses of a coin:
N=50, y=35
- Prior: B(2,3)
- Posterior: B(37,18)



	prior	posterior
mean	0.40	0.67
mode	0.33	0.68
std	0.20	0.06

Posterior var(std) is smaller than when N=10.

37

Posterior Variance

- Expected to be small

$$\mathcal{E}(\theta|y) = \text{var}(\theta) - \text{var}(\mathcal{E}(\theta|y))$$

- Some exceptions

prior:	Beta(10,2)	Beta(10,2)
data:	n=3, x=1	n=10, x=1
posterior:	Beta(11,4)	Beta(11,11)

38

Predictive Distribution

- posterior predictive
What do we know about the next observation after observing N obs.

$$\begin{aligned}
 h(\tilde{y}|\mathbf{y}) &= \int f(\tilde{y}|\pi)h(\pi|\mathbf{y})d\pi \\
 &= \int \pi^{\tilde{y}}(1-\pi)^{1-\tilde{y}} \times \frac{1}{B(\alpha_1, \beta_1)} \pi^{\alpha_1-1}(1-\pi)^{\beta_1-1} \\
 &= \frac{1}{B(\alpha_1, \beta_1)} \int \pi^{\alpha_1+\tilde{y}-1}(1-\pi)^{\beta_1-\tilde{y}} \\
 &= \frac{B(\alpha_1 + \tilde{y}, \beta_1 + 1 - \tilde{y})}{B(\alpha_1, \beta_1)} \\
 &= \frac{\Gamma(\alpha_1 + \tilde{y})\Gamma(\beta_1 + 1 - \tilde{y})}{\Gamma(\alpha_1 + \beta_1 + 1)} \times \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} \\
 &= \frac{1}{\alpha_1 + \beta_1} \times \frac{(\alpha_1 + \tilde{y} - 1)!(\beta_1 - \tilde{y})!}{(\alpha_1 - 1)!(\beta_1 - 1)!}
 \end{aligned}$$

39

Normal Model

- voltage of battery

α	$\beta = 1.2$	$\beta = 1.5$
0.95	0.001	0.000
1.00	0.005	0.000
1.05	0.030	0.000
1.10	0.104	0.000
1.15	0.220	0.000
1.20	0.282	0.000
1.25	0.220	0.001
1.30	0.104	0.005
1.35	0.030	0.030
1.40	0.005	0.104
1.45	0.001	0.220
1.50	0.000	0.282
1.55	0.000	0.220
1.60	0.000	0.104
1.65	0.000	0.030
1.70	0.000	0.005
1.75	0.000	0.001

A digital voltage meter is used to measure the voltage of a battery whose true voltage is either 1.2v or 1.5v.

The digital voltage meter display the voltage with the step of 0.05 volt as shown in the CPT.

$$\Pr(B = \beta_j | A = \alpha_i) = \frac{\frac{1}{c_j} \exp(-\frac{1}{2}(\alpha_i - \beta_j)^2/0.05) \Pr(B = \beta_j)}{\sum_k^a \frac{1}{c_k} \exp(-\frac{1}{2}(\alpha_i - \beta_k)^2/0.05) \Pr(B = \beta_k)}$$

40