

## Case study:

### Analysis of the authorship of the *Tirant lo Blanc*

## Case study: Change Point Analysis

---

### Analysis of the authorship of the *Tirant lo Blanc*

There is the controversial debate about if the *Tirant lo blanc* has been written by one author or two authors:

- Joanot Martorell (died in 1465)
- Joanot Martorell & Martí Joan de Galba (printed the book in 1490)

## Case study: Change Point Analysis

### *Tirant lo Blanc*

	Llargada paraula										$N$
	1	2	3	4	5	6	7	8	9	10+	
Capítol 1	21	59	44	19	33	20	16	17	9	17	255
Capítol 2	53	113	80	49	52	33	28	36	16	16	476
Capítol 3	109	274	239	128	112	110	76	51	43	32	1.174
Capítol 4	69	150	126	71	60	71	47	32	23	21	670
...	...	...	...	...	...	...	...	...	...	...	...
Capítol 486	45	88	91	46	40	28	13	30	11	10	402
Capítol 487	48	49	62	53	41	36	21	9	16	13	348

Part of the table of counts of words of each length in each chapter.  $N_i$  is the total number of words in that chapter

## Case study: Change Point Analysis

### Objective:

To identify a change in style that might indicate a change of author

Study case inspired by a  
real case of stylometry

We want to detect if a coin has  
been changed during the  
experiments

## Write the Statistical Model

**Objective:**

Identify when the coin is changed

	# faces	# tosses
sample 1	4	6
sample 2	5	8
sample 3	9	12
...	...	...
sample 49	10	13
sample 50	7	10

## Change Point

---

### Objective:

Identify when the coin has been changed

#### Parameters of interest

- When the coin is changed
- Head probability of the first coin
- Head probability of the second coin

## Change Point

---

A sequence of ordered binomial variables

$$p(y_i | N_i, \theta_1, \theta_2, r) = \begin{cases} \text{Bin}(N_i, \theta_1) & \text{if } i < r \\ \text{Bin}(N_i, \theta_2) & \text{if } i \geq r \end{cases}$$

$$p(\underline{y} = (y_1, \dots, y_n) | \underline{N}, \theta_1, \theta_2, r) = \prod_{i=1}^{r-1} \text{Bin}(y_i | N_i, \theta_1) \prod_{i=r}^n \text{Bin}(y_i | N_i, \theta_2)$$

#### Parameters

- $r$  change point location
- $\theta_1$  head probability before the change point
- $\theta_2$  head probability after the change point

## Change Point

---

### Bayesian Model

$$p(y_i | N_i, \theta_1, \theta_2, r) = \begin{cases} \text{Bin}(N_i, \theta_1) & \text{si } i < r \\ \text{Bin}(N_i, \theta_2) & \text{si } i \geq r \end{cases} \quad \text{for } i = 1 \dots 50$$

$$\pi(r = i) = \frac{1}{50}$$

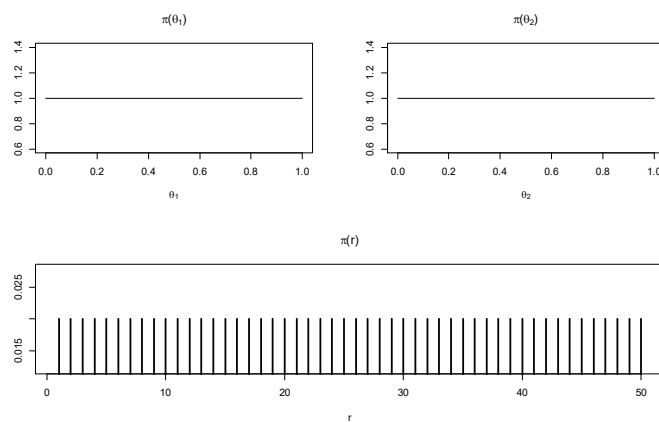
$$\pi(\theta_1) = \text{Beta}(1,1)$$

$$\pi(\theta_2) = \text{Beta}(1,1)$$

## Change Point

---

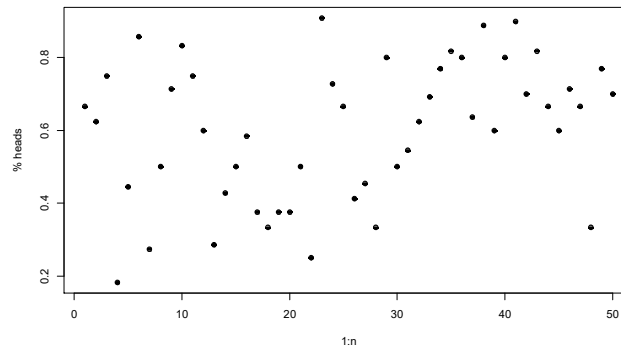
### The prior distribution



## Change Point

### The Data

	# faces	# tosses
sample 1	4	6
sample 2	5	8
sample 3	9	12
...	...	...
sample 49	10	13
sample 50	7	10



## Change Point

JAGS'code

```
model
{
  for (i in 1:50) {
    Y[i] ~ dbin(theta[S[i]], N[i])
    S[i] <- 1+step(i-r)
  }
  for (j in 1:2) {
    theta[j] ~ dbeta(1,1)
  }
  r ~ dcat(p[])
}
```

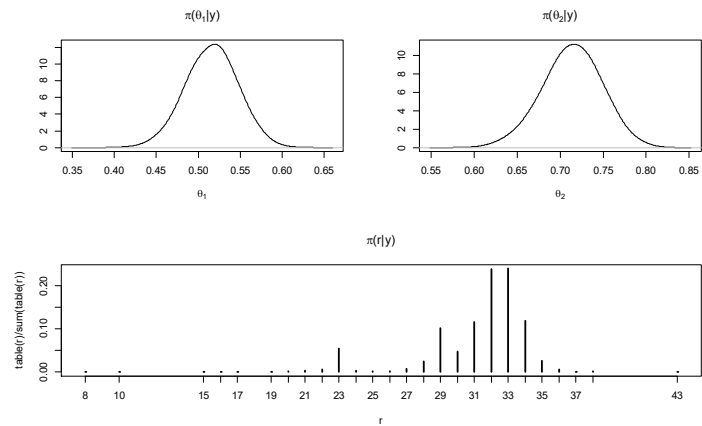
$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

$p[] = c(1/50, \dots, 1/50)$

$\text{step}(x)$  1 if  $x \geq 0$ , otherwise 0

# Change Point

## The posterior distribution



# Styolometry: Change Point Analysis

## Analysis of the authorship of the *Tirant lo Blanc*

### Objective:

To identify a change in style that might indicate a change of author

	Llargada paraula										N
	1	2	3	4	5	6	7	8	9	10+	
Capítol 1	21	59	44	19	33	20	16	17	9	17	255
Capítol 2	53	113	80	49	52	33	28	36	16	16	476
Capítol 3	109	274	239	128	112	110	76	51	43	32	1.174
Capítol 4	69	150	126	71	60	71	47	32	23	21	670
...	...	...	...	...	...	...	...	...	...	...	...
Capítol 486	45	88	91	46	40	28	13	30	11	10	402
Capítol 487	48	49	62	53	41	36	21	9	16	13	348

## Styolometry: Change Point Analysis

---

I encourage you to read:

Giron, J., Ginebra, J. and Riba, A. (2005). Bayesian analysis of a multinomial sequence and homogeneity of literary style. *The American Statistician*, 59, 19-30.