# 7.1 Approximating a Distribution with a Large Sample
# 7.2 A Simple Case of the Metropolis Algorithm

110191019 鍾侑廷

# Markov chain Monte Carlo (MCMC)

- A large number of $\theta$ values sampled from the posterior distribution are used to approximate the posterior distribution $p(\theta|D)$.

- The method assumes that the prior distribution is easily evaluated, and the value of the likelihood function can be computed for any specified values of D and θ.

$$\underbrace{P(\text{Claim}|\text{Data})}_{\text{Posterior}} = \frac{\overbrace{P(\text{Data}|\text{Claim})}^{\text{Likelihood}} \times \overbrace{P(\text{Claim})}^{\text{Prior}}}{\underbrace{P(\text{Data})}_{\text{Marginal Probability}}}$$

# 7.1 Approximating a distribution with a large sample

an example of approximating an exact mathematical distribution
by a large random sample of representative values.

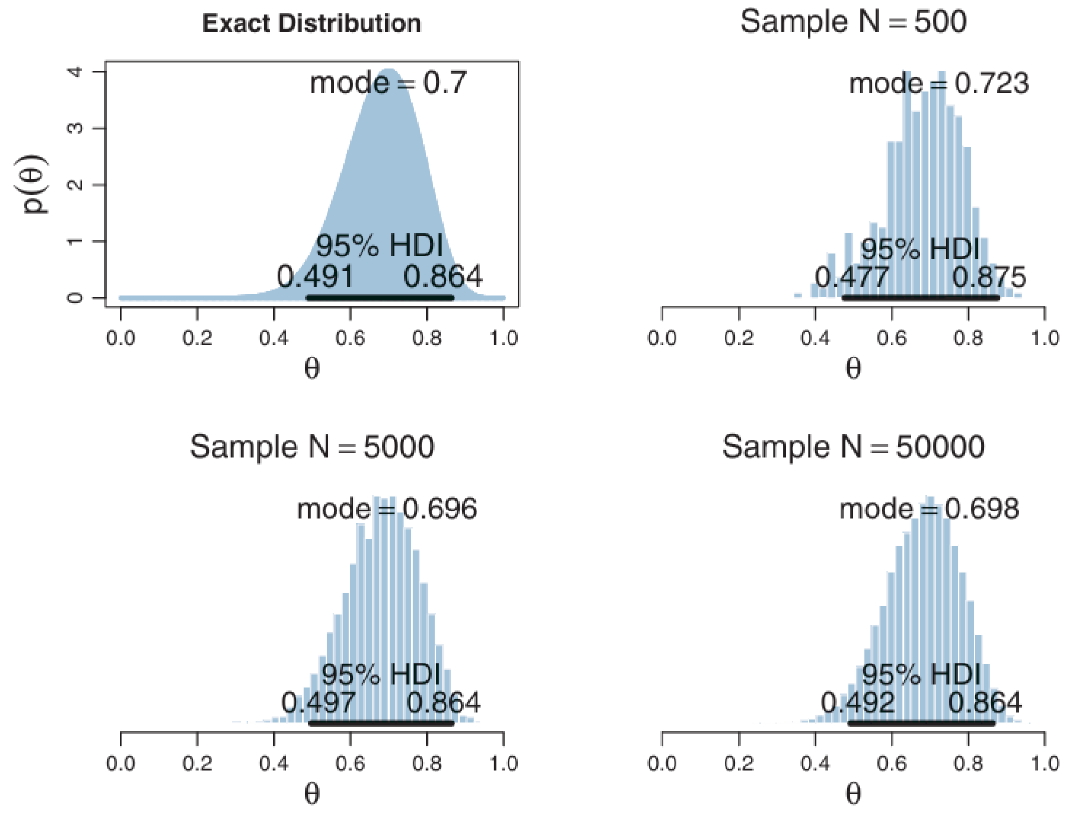


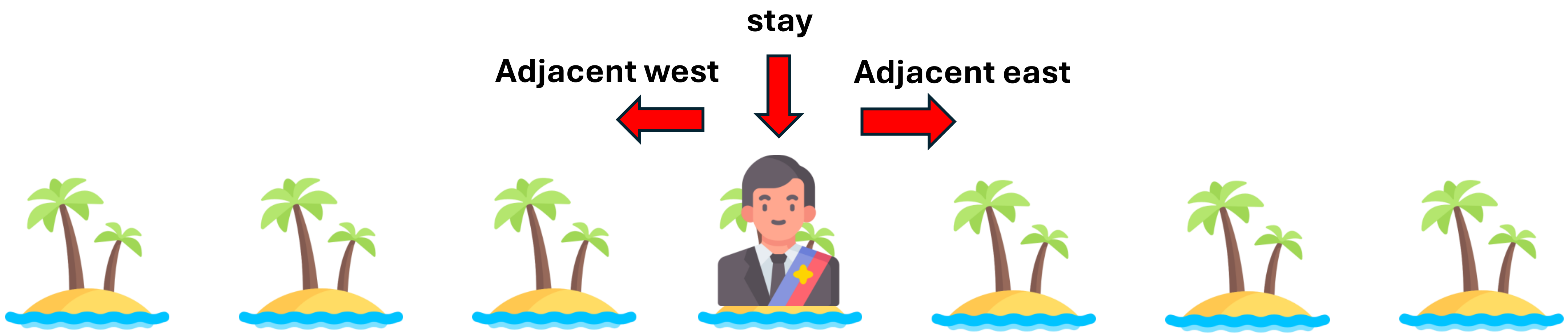

**Figure 7.1** Large representative samples approximate the continuous distribution in the upper-left panel. The larger the sample, the more accurate the approximation. (This happens to be a beta($\theta$|15, 7) distribution.)

# How can we sample a large number of representative values from a distribution?

Goal ： visit all the islands proportionally to their relative population

？： what the total population of the island chain is and how many islands there are

！： The number of people on the current island and the number of people on the destination island.



# Politician's Move rules

flips a fair coin to decide whether go to east or west

the proposed island > the current island

**Go to the proposed island !**

the proposed island < the current island

the ratio of the proposed island population to the current island population:

pmove = Pproposed / Pcurrent

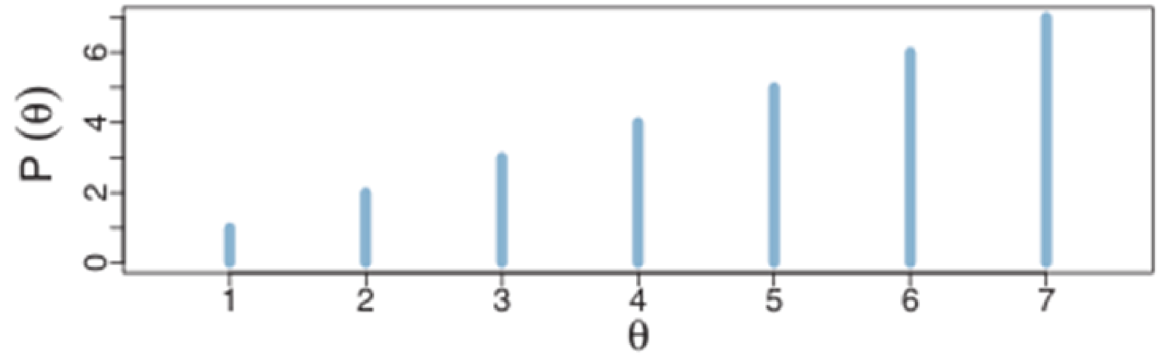
spinning a fair spinner marked on its circumference with uniform values from 0 to 1

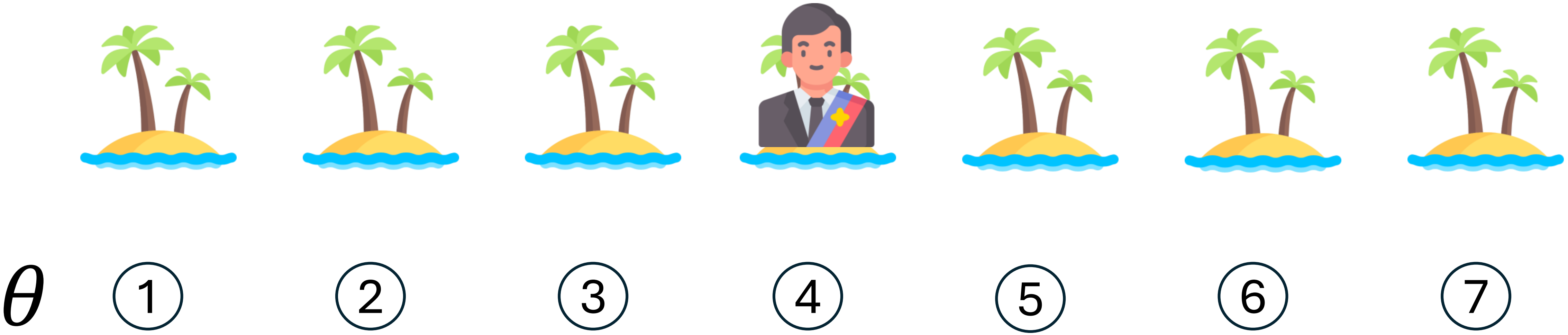
the pointed-to value is between 0 and pmove,
**Go to the proposed island !**

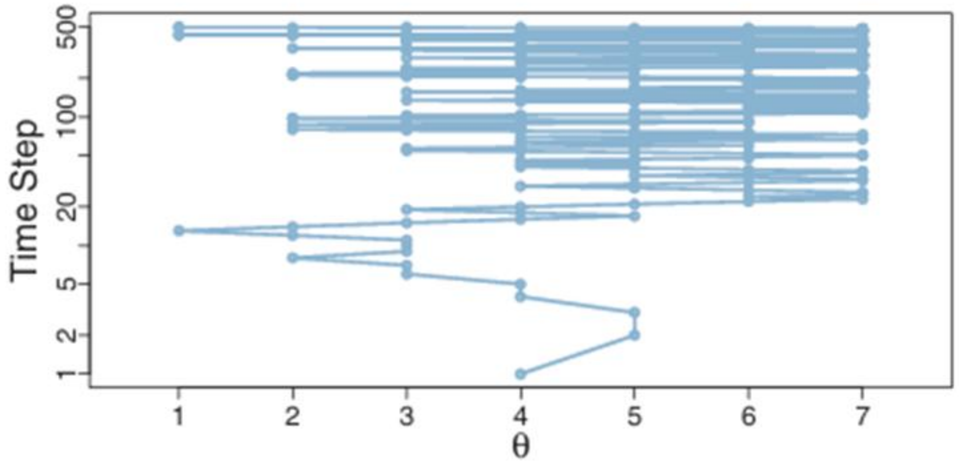
**Otherwise,**
Stay on the current island

In the long run, the probability that the politician is on any one of the islands exactly matches the relative population of the island!

# 7.2.2. A random walk






DAY 1

DAY 2

DAY 3


 right

P(5) > P(4)

 left

P(4) < P(5)

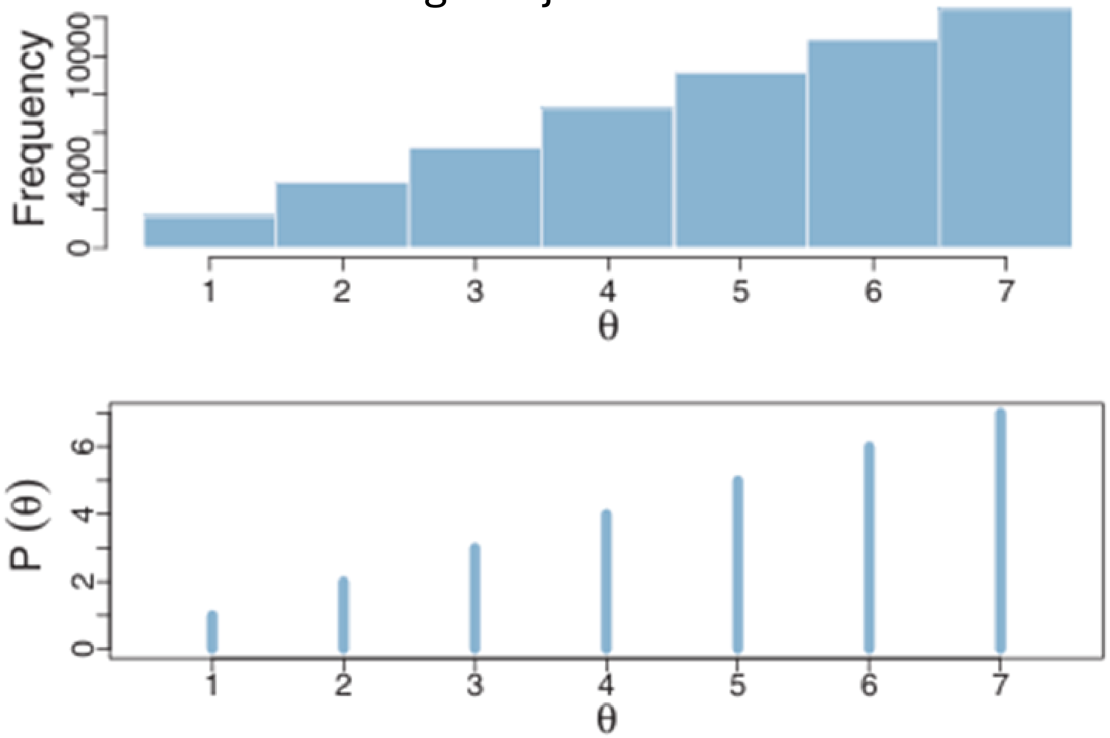
pmove = P(4) /P(5) = 4/5 = 0.8

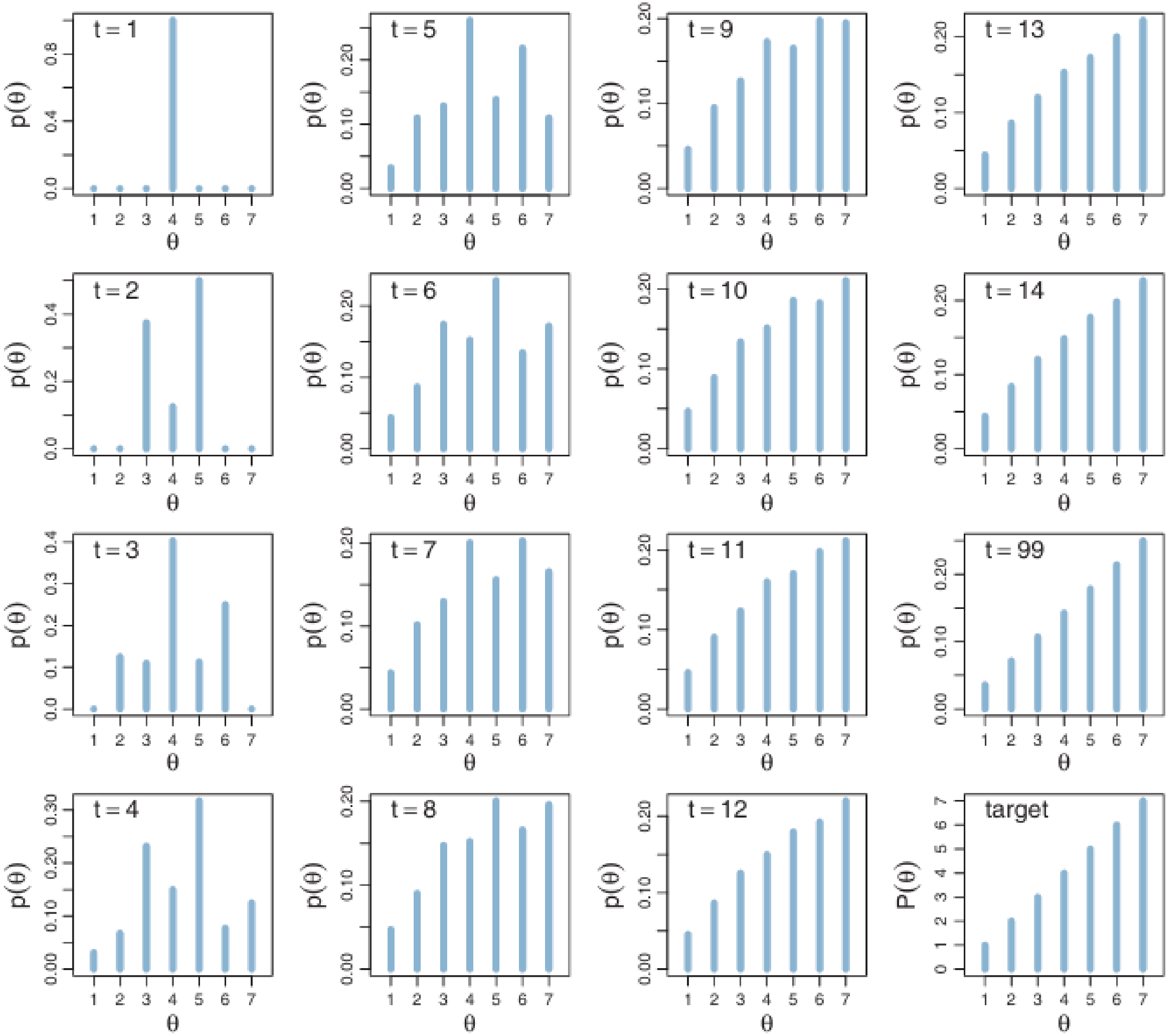
 > 0.8

# 7.2.2. A random walk

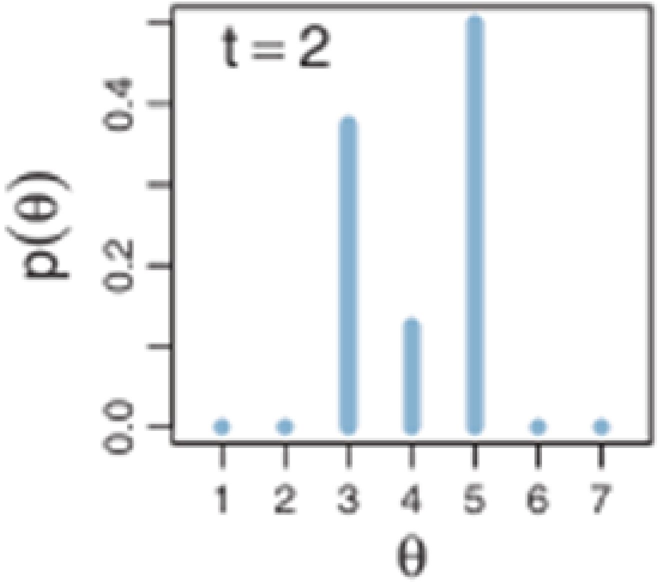
a sequence generated this way will converge, as the sequence gets longer, to an arbitrarily close approximation of the actual relative probabilities.

the frequencies with which each position is visited during this junket.




t = 1
t = 2
t = 3
t = 4
t = 5
t = 6
t = 7
t = 8
t = 9
t = 10
t = 11
t = 12
t = 13
t = 14
t = 99
target
p(θ)
P(θ)
θ

# 7.2.3. General properties of a random walk




<u>Move to 5 :</u>

50 percent

<u>Move to 3 :</u>

50 percent

$P(\theta=3) < P(\theta=4)$

$P(\theta=3) / P(\theta=4) = 3/4 = 0.75$

$0.5 \times 0.75 = 0.375$

<u>Stay on 4 :</u>

$1-0.5-0.375 = 0.125$

# 7.2.3. General properties of a random walk

- **Further Time Steps (t = 3 and beyond):**
- The process continues similarly, expanding the range of possible positions.
- At $t$=3, the politician can be at $\theta$=2 to $\theta$=6, but not beyond.
- Over time, , we can approximate the target probability at each value of θ by simply counting the relative number times that the traveler visited that value.

flips a fair coin to decide whether go to east or west

50%

the proposed island > the current island

**Go to the proposed island !**

50%

the proposed island < the current island

the ratio of the proposed island population to the current island population:

pmove = P($\theta$ proposed) / P($\theta$current)

spinning a fair spinner marked on its circumference with uniform values from 0 to 1

the pointed-to value is between 0 and pmove,
**Go to the proposed island !**

**Otherwise,**
Stay on the current island

The range of possible proposed moves, and the probability of proposing each, is called the proposal distribution.

flips a fair coin to decide whether go to east or west

$P(\theta \text{proposed}) > P(\theta \text{current})$

**Go to the proposed island !**

pmove = 1

$P(\theta \text{proposed}) < P(\theta \text{current})$

the ratio of the proposed island population to the current island population:

pmove = $P(\theta \text{ proposed}) / P(\theta \text{current})$

$$p_{\text{move}} = \min\left(\frac{P(\theta_{\text{proposed}})}{P(\theta_{\text{current}})}, 1\right)$$

A random value u is drawn from a uniform distribution over [0,1].

u ≤ pmove , the move is accepted, and the position is updated to $\theta$proposed .

**Otherwise,**

Stay on the current island

# 7.2.4. Why We Care

In the random walk process, we must be able to perform the following three tasks:

1. Generate a random value from the proposal distribution to create $\theta_{proposed}$.
2. Evaluate the target distribution at any proposed position to compute $P(\theta_{proposed})/P(\theta_{current})$
3. Generate a random value from a uniform distribution to decide whether to accept or reject the proposed move based on $p_{move}$.

# 7.2.4. Why We Care

- By performing these three steps, we can **indirectly** achieve something that we could not necessarily do directly:
  We can generate random samples from the target distribution.

- Even when the target distribution is not normalized, we can still generate random samples from it.

$$\underbrace{P(\text{Claim}|\text{Data})}_{\text{Posterior}} = \frac{\overbrace{P(\text{Data}|\text{Claim})}^{\text{Likelihood}} \text{ x } \overbrace{P(\text{Claim})}^{\text{Prior}}}{\underbrace{P(\text{Data})}_{\text{Marginal Probability}}}$$

# 7.2.5. Why It Works

Probability of Moving from $\theta$ to $\theta$+1

$$\frac{\boxed{p(\theta \to \theta+1)}}{\boxed{p(\theta+1 \to \theta)}} = \frac{0.5 \min\left(P(\theta+1)/P(\theta), 1\right)}{0.5 \min\left(P(\theta)/P(\theta+1), 1\right)}$$

Probability of Moving from $\theta$+1 to $\theta$

$$= \begin{cases} \frac{1}{P(\theta)/P(\theta+1)} & \text{if } P(\theta+1) > P(\theta) \\ \frac{P(\theta+1)/P(\theta)}{1} & \text{if } P(\theta+1) < P(\theta) \end{cases}$$

$$= \frac{P(\theta+1)}{P(\theta)} \tag{7.2}$$

The relative transition probabilities between these positions match the relative values of the target distribution.

# Transition Probability Matrix

$$
\begin{bmatrix}
\ddots & p(\theta-2 \rightarrow \theta-1) & 0 & 0 & 0 \\
\ddots & p(\theta-1 \rightarrow \theta-1) & p(\theta-1 \rightarrow \theta) & 0 & 0 \\
0 & p(\theta \rightarrow \theta-1) & p(\theta \rightarrow \theta) & p(\theta \rightarrow \theta+1) & 0 \\
0 & 0 & p(\theta+1 \rightarrow \theta) & p(\theta+1 \rightarrow \theta+1) & \ddots \\
0 & 0 & 0 & p(\theta+2 \rightarrow \theta+1) & \ddots
\end{bmatrix}
\begin{bmatrix}
\ddots & 0.5\min\left(\frac{P(\theta-1)}{P(\theta-2)}, 1\right) & 0 & 0 & 0 \\
\ddots & \begin{array}{c} 0.5\left[1-\min\left(\frac{P(\theta-2)}{P(\theta-1)}, 1\right)\right] \\ +\, 0.5\left[1-\min\left(\frac{P(\theta)}{P(\theta-1)}, 1\right)\right] \end{array} & 0.5\min\left(\frac{P(\theta)}{P(\theta-1)}, 1\right) & 0 & 0 \\
0 & 0.5\min\left(\frac{P(\theta-1)}{P(\theta)}, 1\right) & \begin{array}{c} 0.5\left[1-\min\left(\frac{P(\theta-1)}{P(\theta)}, 1\right)\right] \\ +\, 0.5\left[1-\min\left(\frac{P(\theta+1)}{P(\theta)}, 1\right)\right] \end{array} & 0.5\min\left(\frac{P(\theta+1)}{P(\theta)}, 1\right) & 0 \\
0 & 0 & 0.5\min\left(\frac{P(\theta)}{P(\theta+1)}, 1\right) & \begin{array}{c} 0.5\left[1-\min\left(\frac{P(\theta)}{P(\theta+1)}, 1\right)\right] \\ +\, 0.5\left[1-\min\left(\frac{P(\theta+2)}{P(\theta+1)}, 1\right)\right] \end{array} & \ddots \\
0 & 0 & 0 & 0.5\min\left(\frac{P(\theta+1)}{P(\theta+2)}, 1\right) & \ddots
\end{bmatrix}
\tag{7.3}
$$

# how matrix multiplication operates:

Consider a matrix T.

- The value in its rth row and cth column is denoted as Trc.
- We can multiply the matrix on its left side by a row vector w, which yields another row vector.
- The cth component of the product wT is $\sum_r w_r T_{rc}$

# To use the transition matrix in Equation 7.3:

- We put the current location probabilities into a row vector w.
- This vector w represents where we are.
- If at the current time we are definitely in location $\theta=4$, then w has 1.0 in its $\theta=4$ component and zeros everywhere else.
- To determine the probability of locations at the next time step, we simply multiply w by T.

- At every time step, we multiply the current position probability vector w by the transition probability matrix T.
- This gives the position probabilities for the next time step.
- By repeatedly multiplying by T, we derive the long-run position probabilities.
- This process is exactly what generated the graphs in Figure 7.3.

# Stable Target Distribution

- If the position probability vector matches the target distribution, it remains unchanged in the next step.

$$w = [\ldots, P(\theta-1), P(\theta), P(\theta+1), \ldots]/Z, \text{ where } Z = \sum_\theta P(\theta)$$

- The target distribution is stable and does not change under the transition matrix $T$.

# Mathematical Proof

- Computing the $\theta$ Component $\sum_r w_r T_{r\theta}$

$$\sum_r w_r T_{r\theta} = P(\theta-1)/Z \cdot 0.5 \min\left(\frac{P(\theta)}{P(\theta-1)}, 1\right)$$
$$+ P(\theta)/Z \cdot \left(0.5\left[1-\min\left(\frac{P(\theta-1)}{P(\theta)}, 1\right)\right] + 0.5\left[1-\min\left(\frac{P(\theta+1)}{P(\theta)}, 1\right)\right]\right)$$
$$+ P(\theta+1)/Z \cdot 0.5 \min\left(\frac{P(\theta)}{P(\theta+1)}, 1\right) \quad (7.4)$$

Case 1: $P(\theta) > P(\theta - 1)$ and $P(\theta) > P(\theta + 1)$;
Case 2: $P(\theta) > P(\theta - 1)$ and $P(\theta) < P(\theta + 1)$;
Case 3: $P(\theta) < P(\theta - 1)$ and $P(\theta) > P(\theta + 1)$;
Case 4: $P(\theta) < P(\theta - 1)$ and $P(\theta) < P(\theta + 1)$

- consider Case 1, when P(θ) > P(θ−1) and P(θ) > P(θ+1)

$$
\begin{aligned}
\sum_r w_r T_{r\theta} &= P(\theta-1)/Z \cdot 0.5 \\
&\quad + P(\theta)/Z \cdot \left(0.5\left[1-\left(\frac{P(\theta-1)}{P(\theta)}\right)\right] + 0.5\left[1-\left(\frac{P(\theta+1)}{P(\theta)}\right)\right]\right) \\
&\quad + P(\theta+1)/Z \cdot 0.5 \\
&= 0.5\, P(\theta-1)/Z \\
&\quad + 0.5P(\theta)/Z - 0.5P(\theta)/Z\frac{P(\theta-1)}{P(\theta)} + 0.5P(\theta)/Z - 0.5P(\theta)/Z\frac{P(\theta+1)}{P(\theta)} \\
&\quad + 0.5\, P(\theta+1)/Z \\
&= P(\theta)/Z
\end{aligned}
$$

Conclusion: In all cases, the result reduces to $P(\theta)/Z$, proving stability.