PIVOT TABLES

(continued...)

Laboratory: comparing two automated systems

THE DATASET

`pivot_system_comparison.xlsx`

700 case records: 350 evaluated by System A and 350 by System B. The systems were tested on **different cases**.

Each row corresponds to a single evaluated case.

Field	Meaning
`DecisionID`	Unique identifier for the case
`System`	A or B
`Complexity`	Standard or Complex
`Correct`	Yes or No

We want to compare the two systems without confusing performance with the difficulty of the cases assigned to them.

What exactly does accuracy measure?

For a system s, its **overall accuracy** is

$$\widehat{P}(\text{Correct} \mid \text{System} = s) = \frac{\#\{\text{correct cases evaluated by } s\}}{\#\{\text{all cases evaluated by } s\}}.$$

For a complexity category c, its **accuracy within that category** is

$$\widehat{P}(\text{Correct} \mid s, c) = \frac{\#\{\text{correct cases for } s \text{ in } c\}}{\#\{\text{cases for } s \text{ in } c\}}.$$

THE DENOMINATOR CHANGES

The overall rate uses every case assigned to the system. The conditional rate uses only cases in one complexity category.

These are different statistical questions. A PivotTable lets us calculate both while keeping the relevant denominator visible.

PivotTable setup for overall accuracy

1. Select a cell, then choose **Insert > PivotTable** and use a new worksheet.
2. Place **System** in **Rows**.
3. Place **Correct** in **Columns**.
4. Place **DecisionID** in **Values**; confirm that Excel uses **Count of DecisionID**.
5. Open **Value Field Settings > Show Values As** and choose **% of Row Total**. Format the values with one decimal place.

HOW TO READ THE TABLE

Within a system's row,

$$\text{Yes percentage} = \frac{\text{Yes count}}{\text{row total}} = \text{accuracy}.$$

Each system row must sum to 100%.

Exercise: system accuracy

QUESTIONS

1. Use a PivotTable to calculate the overall accuracy of each system.
2. Calculate the accuracy of each system separately for Standard and Complex cases.
3. Which system appears better overall?
4. Which system is better within each complexity category?
5. Explain the apparent contradiction.
6. If both systems were applied to a population containing 50% Standard and 50% Complex cases, which would have the higher expected accuracy?

Solution: compare the systems overall

Questions 1 and 3. First ignore complexity and use all 350 cases assigned to each system.

System	Yes	No	Total	Accuracy
A	273	77	350	$273/350 = 78.0\%$
B	289	61	350	$289/350 \approx 82.6\%$

IF WE STOP HERE

System B appears better overall: **82.6% versus 78.0%**.

Its observed advantage is about **4.6 percentage points**.

This answer describes the two observed samples. It does not yet ask whether the systems received cases of comparable difficulty.

Now compare cases of the same complexity

Copy the first PivotTable and change only its row structure:

1. Keep **System** in **Rows**.
2. Add **Complexity** *above* System in **Rows**.
3. Keep **Correct** in **Columns**.
4. Keep **Count of DecisionID** in **Values** and **Show Values As: % of Row Total**.

WHAT CHANGES?

For the row **Standard − A**, the denominator is no longer 350. It is the 87 Standard cases evaluated by A:

$$\widehat{P}(\text{Correct} \mid A, \text{Standard}) = \frac{81}{87}.$$

Each system is now compared with the other system **inside the same complexity category**.

Solution: accuracy within each category

Questions 2 and 4. Read the Yes percentage for each System–Complexity row.

Complexity	System A	System B	Winner
Standard	$81/87 \approx 93.1\%$	$234/270 \approx 86.7\%$	**A**
Complex	$192/263 \approx 73.0\%$	$55/80 \approx 68.8\%$	**A**

WITHIN COMPARABLE CASES

System A is more accurate for **Standard** cases.

System A is also more accurate for **Complex** cases.

The advantage of A is about 6.4 percentage points among Standard cases and 4.3 points among Complex cases.

The ranking reverses

Put the two PivotTable readings next to each other.

Comparison	A	B	Winner
Overall observed samples	78.0%	82.6%	**B**
Within Standard cases	93.1%	86.7%	**A**
Within Complex cases	73.0%	68.8%	**A**

THE APPARENT CONTRADICTION

B is ahead when all cases are pooled, yet A is ahead after we compare like with like in **both** complexity categories.

The percentages are not inconsistent. Before naming the phenomenon, we need to see how the cases were distributed between the systems.

The systems received different case mixes

In a new Pivot Table, show the values as ordinary counts as displayed below. Then compute (a part) each category's share of the system total.

	System A		System B	
Complexity	**Cases**	**Share**	**Cases**	**Share**
Standard	87	24.9%	270	77.1%
Complex	263	75.1%	80	22.9%
Total	350	100.0%	350	100.0%

THE KEY DIFFERENCE

A was tested mainly on **Complex** cases.

B was tested mainly on **Standard** cases.

Both systems have higher accuracy on Standard cases. Therefore, a sample containing more Standard cases tends to have a higher overall rate.

Why there is no mathematical contradiction

The two comparisons answer different questions.

OVERALL COMPARISON

How accurate was each system on the particular mix of cases it received?

$$\widehat{P}(\text{Correct} \mid \text{System})$$

CONDITIONAL COMPARISON

How accurate was each system among cases of the same complexity?

$$\widehat{P}(\text{Correct} \mid \text{System}, \text{Complexity})$$

The overall rates combine **performance** and **case mix**.

The conditional rates hold complexity fixed.

Simpson's paradox

For system s and category c, write

$$\widehat{p}_{sc} = \text{accuracy of } s \text{ within } c, \qquad w_{sc} = \frac{\text{cases for } s \text{ in } c}{\text{all cases for } s}.$$

Then the observed overall accuracy is

WEIGHTED-AVERAGE RULE

$$\widehat{p}_s = \sum_c w_{sc}\widehat{p}_{sc}, \qquad \sum_c w_{sc} = 1.$$

Simpson's paradox occurs when an aggregated comparison reverses the comparison found within the relevant groups.

Here A wins within every complexity group, but B wins overall because $w_{Ac} \neq w_{Bc}$: the two overall averages use different weights.

A fair deployment question uses common weights

Question 6 specifies a target population containing 50% Standard and 50% Complex cases.

SAME TARGET POPULATION

$$q_{\text{Standard}} = 0.50, \qquad q_{\text{Complex}} = 0.50.$$

For both systems, calculate

$$\widehat{p}_s^{\,\text{target}} = 0.50\,\widehat{p}_{s,\text{Standard}} + 0.50\,\widehat{p}_{s,\text{Complex}}.$$

The category accuracies come from the stratified PivotTable. We replace each system's observed case mix with the **same 50/50 mix**.

This procedure is called **standardization**: it makes the population used for the comparison explicit.

Solution: expected accuracy for a 50/50 population

Question 6. Use the exact category fractions and round only the final displayed percentages.

$$\widehat{p}_A^{\,\text{target}} = \tfrac{1}{2}\left(\frac{81}{87}\right) + \tfrac{1}{2}\left(\frac{192}{263}\right) \approx \boxed{83.1\%},$$

$$\widehat{p}_B^{\,\text{target}} = \tfrac{1}{2}\left(\frac{234}{270}\right) + \tfrac{1}{2}\left(\frac{55}{80}\right) \approx \boxed{77.7\%}.$$

RESULT

For the common 50/50 population, System A has the higher expected accuracy by about **5.3 percentage points**.

The ranking now agrees with the within-category comparisons because the two systems are evaluated with identical category weights.

The 50/50 calculation in Excel

Copy the category accuracies from the PivotTable into a small helper table. Keep their full underlying precision; display them with one decimal place.

Complexity	A accuracy	B accuracy	Target weight
Standard	81/87	234/270	50%
Complex	192/263	55/80	50%

If the table occupies **A1:D3**, calculate the weighted averages:

A: `=SUMPRODUCT(B2:B3,$D$2:$D$3)`

B: `=SUMPRODUCT(C2:C3,$D$2:$D$3)`

Format the results as percentages: **83.1%** and **77.7%**. Check that the target weights sum to 100%.

Do not calculate from percentages that were manually rounded first.

What the comparison tells us

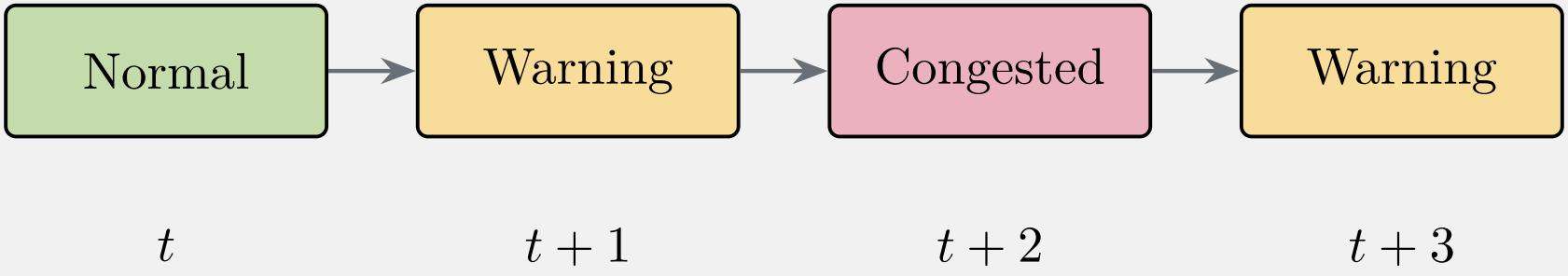
THREE CORRECT STATEMENTS

1. On the two **observed samples**, B has the higher overall accuracy: 82.6% versus 78.0%.
2. Within both **Standard** and **Complex** cases, A has the higher accuracy.
3. On a common **50/50 target population**, A has the higher expected accuracy: 83.1% versus 77.7%.

Markov Chains with Pivot Tables

Why Markov chains?

Many operational systems move repeatedly among a small number of states.



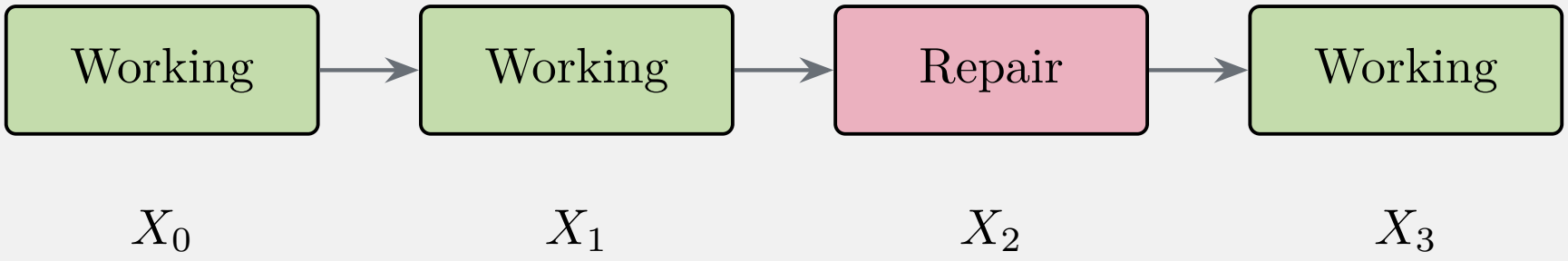
THE QUESTION

If we know the current state, what is the probability of transitioning to a specific new state?

A Markov chain answers this question with a transition model.

State and observation time

Let X_t denote the state of a system at observation time t.



Here the system is inspected once per hour. The sequence is

$$X_0 = W, \qquad X_1 = W, \qquad X_2 = R, \qquad X_3 = W.$$

The index t counts observations. It does not have to mean one hour.

The state space

The **state space** is the set of possible modeled states. For the machine example,

$$\mathcal{S} = \{W, R\},$$

where W means Working and R means Repair.

A USEFUL STATE DEFINITION

At each observation time, exactly one state applies. Together, the states cover every situation that the model intends to represent.

The observation interval matters

Transition probabilities always refer to a specified time step.

Observation interval	**Question represented by a transition**
1 minute	What state appears one minute later?
15 minutes	What state appears 15 minutes later?
1 day	What state appears on the next day?

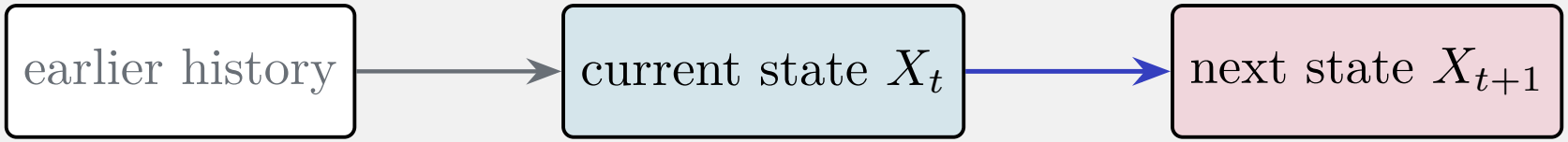
SAME SYSTEM, DIFFERENT MATRIX

A 15-minute transition matrix and a one-day transition matrix usually contain different probabilities.

The Markov property

A process has the Markov property when the current state contains all the information from the past that is needed for the next-step distribution.

$$\Pr(X_{t+1} = j \mid X_t = i, X_{t-1}, \ldots, X_0) = \Pr(X_{t+1} = j \mid X_t = i).$$



Once X_t is known, earlier states add no further predictive information for X_{t+1}.

Markov does not mean independent

Consecutive states can be strongly dependent.

Suppose

$$\Pr(X_{t+1} = R \mid X_t = W) = 0.10, \qquad \Pr(X_{t+1} = R \mid X_t = R) = 0.60.$$

The probability of Repair next hour changes with the current state.

Therefore X_t and X_{t+1} are not independent.

WHAT THE ASSUMPTION REMOVES

The Markov property removes additional dependence on states before X_t. It does not remove dependence between the current and next state.

Constant transition probabilities

A **time-homogeneous** Markov chain uses the same one-step probabilities at every time:

$$p_{ij} = \Pr(X_{t+1} = j \mid X_t = i).$$

The value of p_{ij} does not depend on t.

A SEPARATE ASSUMPTION

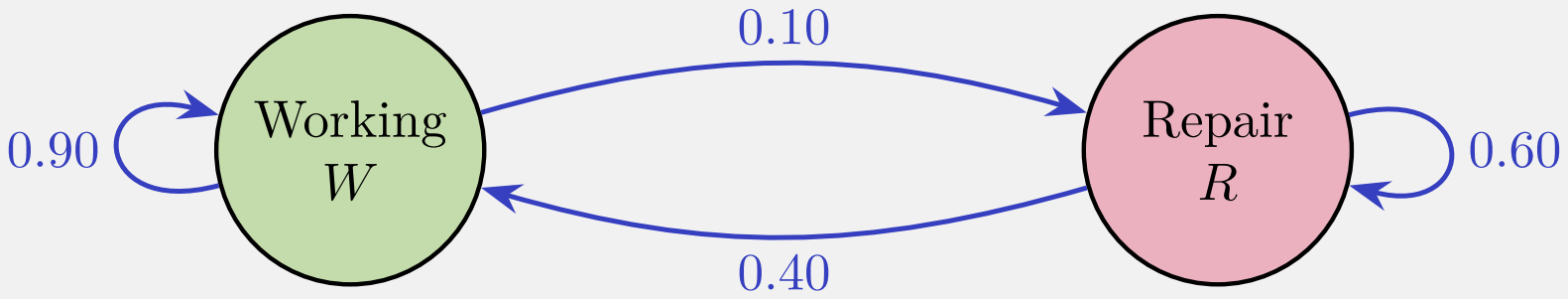
The Markov property concerns which past information matters. Time homogeneity concerns whether the probabilities remain constant over time.

If day and night operations behave differently, estimate separate matrices or include the shift in the state description.

The transition graph

Each node is a state.

Each arrow carries a one-step conditional probability.



The self-loops represent staying in the same state for another observation interval.

The transition matrix

Place transition probabilities in a matrix.

Rows represent the current state, and columns represent the next state.

	W_{t+1}	R_{t+1}
W_t	0.90	0.10
R_t	0.40	0.60

In symbols,

$$p_{ij} = \Pr(X_{t+1} = j \mid X_t = i).$$

READ FROM ROW TO COLUMN

Start in the row for the current state. Move to the column for the next state.

Every row is a probability distribution

For each current state i,

$$p_{ij} \geq 0, \qquad \sum_j p_{ij} = 1.$$

For the Working row,

$$0.90 + 0.10 = 1.$$

For the Repair row,

$$0.40 + 0.60 = 1.$$

WHY ROWS SUM TO ONE

Starting from one current state, the next observation must fall in exactly one of the modeled next states.

Exercise: the next observation

QUESTION

The machine is currently in Repair. Use

$$P = \begin{pmatrix} 0.90 & 0.10 \\ 0.40 & 0.60 \end{pmatrix}$$

with the state order (W, R).

What is the probability that the machine is Working one hour later?
What is the probability that it remains in Repair?

Solution: read the Repair row

The current state is R, so use the second row of P:

	W_{t+1}	R_{t+1}
R_t	**0.40**	**0.60**

Therefore,

$$\Pr(X_{t+1} = W \mid X_t = R) = 0.40,$$

$$\Pr(X_{t+1} = R \mid X_t = R) = 0.60.$$

These probabilities describe uncertainty. They do not specify which outcome must occur.

A distribution over states

The current state is uncertain: we know only the current probability of being in a given state (the deterministic case being a sub-case).

Write the row vector

$$\boldsymbol{\mu}_t = \big(\Pr(X_t = W), \Pr(X_t = R)\big).$$

Examples:

$$\boldsymbol{\mu}_t = (1, 0) \quad \text{means Working is known with certainty,}$$
$$\boldsymbol{\mu}_t = (0.60, 0.40) \quad \text{means 60\% Working and 40\% Repair.}$$

The entries are nonnegative and sum to one.

Updating a distribution

For row-vector notation, the next distribution is

$$\boxed{\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t P.}$$

Each next-state probability combines all ways of arriving at that state. For Working,

$$\Pr(X_{t+1} = W) = \Pr(X_t = W)p_{WW} + \Pr(X_t = R)p_{RW}.$$

WEIGHTED AVERAGE

The current state probabilities supply the weights. The relevant transition probabilities supply the conditional rates.

A small distribution example

Suppose

$$\boldsymbol{\mu}_t = (0.60, 0.40), \qquad P = \begin{pmatrix} 0.90 & 0.10 \\ 0.40 & 0.60 \end{pmatrix}.$$

Then

$$\boldsymbol{\mu}_{t+1} = (0.60, 0.40)\, P = (0.60, 0.40) \begin{pmatrix} 0.90 & 0.10 \\ 0.40 & 0.60 \end{pmatrix} = (0.70, 0.30).$$

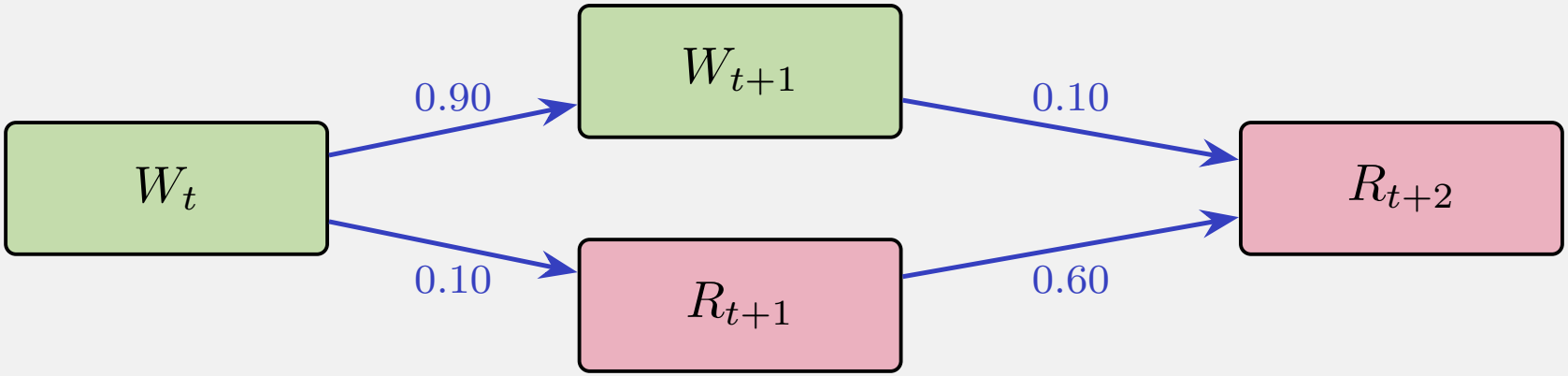
For the Working component,

$$0.60(0.90) + 0.40(0.40) = 0.70.$$

One hour later, the model assigns 70% probability to Working and 30% to Repair.

Two steps require all intermediate paths

Starting from Working, the machine can be in Repair two hours later through two paths.



$$\Pr(W_t \to R_{t+2}) = 0.90(0.10) + 0.10(0.60) = 0.15.$$

Multiply along each path, then add the mutually exclusive paths.

The two-step transition matrix

Matrix multiplication performs the same path calculation for every pair of states:

$$P^2 = P\,P = \begin{pmatrix} 0.90 & 0.10 \\ 0.40 & 0.60 \end{pmatrix} \begin{pmatrix} 0.90 & 0.10 \\ 0.40 & 0.60 \end{pmatrix} = \begin{pmatrix} 0.85 & 0.15 \\ 0.60 & 0.40 \end{pmatrix}.$$

The entry in row i, column j is

$$(P^2)_{ij} = \sum_k p_{ik} p_{kj} = \Pr(X_{t+2} = j \mid X_t = i).$$

Thus $(P^2)_{WR} = 0.15$ agrees with the path calculation.

Several steps into the future

For any positive integer n,

$$(P^n)_{ij} = \Pr(X_{t+n} = j \mid X_t = i),$$

and a starting distribution evolves according to

$$\boxed{\boldsymbol{\mu}_{t+n} = \boldsymbol{\mu}_t P^n.}$$

MATRIX POWERS

P^n means ordinary matrix multiplication repeated n times. It does not mean raising each entry of P to the power n.

The time horizon is n times the observation interval.

A stationary distribution

A distribution $\boldsymbol{\pi}$ is stationary when one transition leaves it unchanged:

$$\boxed{\boldsymbol{\pi} = \boldsymbol{\pi} P,} \qquad \sum_i \pi_i = 1.$$

Individual systems may continue to change state. The population proportions remain constant.

LONG-RUN INTERPRETATION

When the chain converges, π_i is the long-run fraction of observation times spent in state i.

A stationary distribution describes behavior under the transition probabilities encoded in P.

Stationary distribution of the machine

Write $\boldsymbol{\pi} = (a, b)$ for Working and Repair. Solve

$$(a, b) \begin{pmatrix} 0.90 & 0.10 \\ 0.40 & 0.60 \end{pmatrix} = (a, b), \qquad a + b = 1.$$

The Working equation gives

$$a = 0.90a + 0.40b \quad \Longrightarrow \quad 0.10a = 0.40b \quad \Longrightarrow \quad a = 4b.$$

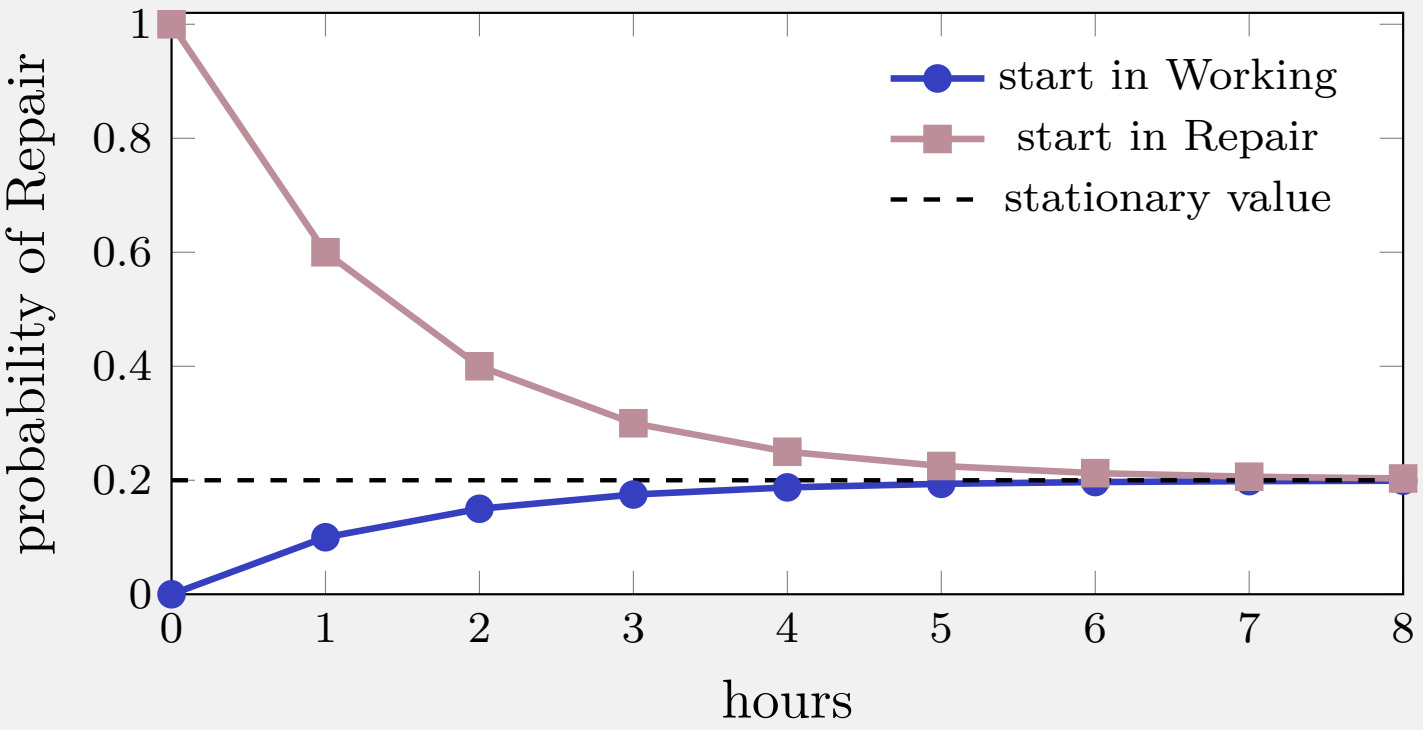
Therefore,

$$\boxed{\boldsymbol{\pi} = (0.80, 0.20).}$$

Under this model, Repair occupies 20% of observation times in the long run.

Convergence toward the stationary distribution

The probability of Repair approaches 0.20 from either starting state.



When the simple model is credible

The model requires more than a matrix with rows that sum to one.

- The chosen state should contain the information needed for the next-step distribution.
- The observation interval should remain fixed.
- A time-homogeneous model assumes stable transition probabilities.
- The data should contain enough transitions from every state.

A MODEL, NOT A CAUSAL CLAIM

An estimated transition matrix describes observed movements under the conditions in the data. It does not by itself prove what caused those movements.

Estimating transition probabilities from data

Suppose a dataset records consecutive states

$$X_0, X_1, \ldots, X_T.$$

For every pair of states i and j, count

$$N_{ij} = \#\{t : X_t = i,\ X_{t+1} = j\}.$$

The number of observed departures from state i is

$$N_i = \sum_j N_{ij}.$$

These counts form the empirical transition-count matrix.

From counts to estimated probabilities

Normalize each row of the count matrix:

$$\boxed{\widehat{p}_{ij} = \frac{N_{ij}}{N_i} = \frac{N_{ij}}{\sum_j N_{ij}}.}$$

THE DENOMINATOR

For row i, the denominator includes every observed transition that starts from state i.

The resulting matrix $\widehat{P}$ estimates the one-step conditional probabilities. A PivotTable can produce both N_{ij} and the row percentages.

APPLIED LABORATORY

congestion in an automated warehouse

`warehouse_congestion_markov.xlsx`

The warehouse congestion problem

An automated fulfillment center routes totes to packing stations. Every 15 minutes, the control room classifies congestion from the number of waiting totes:

Low	0–39 waiting totes	normal flow
Medium	40–79 waiting totes	queues are building
High	80 or more waiting totes	intervention may be needed

OPERATIONS QUESTION

Given the congestion state now, what is the probability distribution 15 or 30 minutes later?

The synthetic observation period

The workbook contains a teaching dataset generated from a stable three-state process.

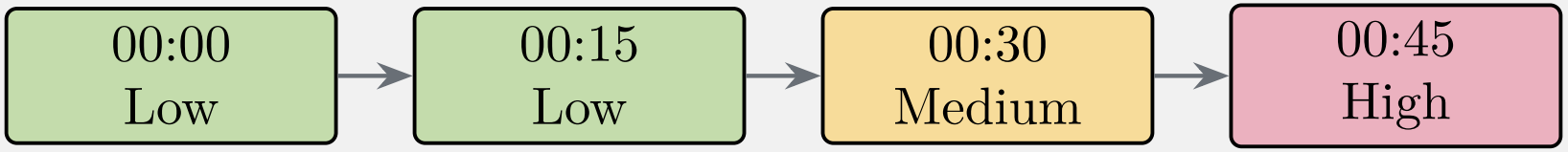
`warehouse_congestion_markov.xlsx`

- 673 congestion observations from 5 January 2026 at 00:00 to 12 January 2026 at 00:00.
- One observation every 15 minutes.
- 672 consecutive transitions.
- one continuously operating hypothetical warehouse.

The data are synthetic. The exercise focuses on the estimation workflow and its interpretation.

From observations to transition rows

Two consecutive observations create one transition.



Interval start	**CurrentState**	**NextState**
00:00	Low	Low
00:15	Low	Medium
00:30	Medium	High

With 673 observations, there are 672 transitions. The final observation has no later state inside the sample.

The workbook fields

Use the Excel Table `TransitionsTable`. One row represents one 15-minute transition.

Field	Meaning
`TransitionID`	Unique identifier for the transition
`IntervalStart`	Date and time of the current observation
`CurrentState`	Congestion state at the start of the interval
`NextState`	Congestion state 15 minutes later

The current state supplies the row of the transition matrix. The next state supplies the column.

PivotTable setup for transition counts

1. Select a cell in `TransitionsTable`. Choose **Insert** > **PivotTable** and use a new worksheet.
2. Place **CurrentState** in **Rows**.
3. Place **NextState** in **Columns**.
4. Place **TransitionID** in **Values**. Confirm **Count of TransitionID**.
5. Manually order both state dimensions as Low, Medium, High.

WHAT THIS TABLE ESTIMATES FIRST

The first PivotTable contains counts N_{ij}, not probabilities.

Solution: the transition-count matrix

Excel may initially show **High, Low, Medium** (alphabetical order). We want **Low, Medium, High** for both rows and columns.

1. Right-click the **High row label**, then choose:
   Move → Move "High" to End.
2. Repeat on the **High column label** using the same command.

Excel moves the corresponding counts automatically:

	NextState			
CurrentState	**Low**	**Medium**	**High**	**Row total**
Low	189	55	14	258
Medium	48	126	51	225
High	20	44	125	189
Grand total	257	225	190	672

For example, 125 observed transitions started in High and also ended in High.

PivotTable setup for transition probabilities

Keep the same PivotTable field arrangement. Then:

1. Open **Value Field Settings** for Count of TransitionID.
2. Choose **Show Values As**.
3. Select **% of Row Total**.
4. Format the values as percentages with one decimal place.

WHY ROW TOTAL?

Within the Low row, the denominator must be all 258 transitions that start from Low. The same rule applies separately to Medium and High.

Solution: the estimated transition matrix

The row percentages give

$$\widehat{P} = \begin{array}{c|ccc} & \text{Low} & \text{Medium} & \text{High} \\ \hline \text{Low} & 73.3\% & 21.3\% & 5.4\% \\ \text{Medium} & 21.3\% & 56.0\% & 22.7\% \\ \text{High} & 10.6\% & 23.3\% & 66.1\% \end{array}$$

For example,

$$\widehat{p}_{HH} = \frac{125}{189} \approx 0.661.$$

Each displayed row sums to 100.0%, apart from possible rounding in the visible percentages.

Reading the warehouse estimates

The diagonal entries measure one-step persistence.

Current state	Probability of staying	Interpretation
Low	$189/258 = 73.3\%$	largest diagonal
Medium	$126/225 = 56.0\%$	lower persistence
High	$125/189 = 66.1\%$	congestion often persists

CONDITIONAL READING

If the warehouse is High now, the estimated probability that it remains High 15 minutes later is 66.1%.

A 30-minute forecast from Low

Thirty minutes equals two 15-minute transitions. Therefore use $\widehat{P}^2$.

$$(\widehat{P}^2)_{LH} = \widehat{p}_{LL}\widehat{p}_{LH} + \widehat{p}_{LM}\widehat{p}_{MH} + \widehat{p}_{LH}\widehat{p}_{HH}.$$

Using the exact fractions behind the PivotTable percentages,

$$(\widehat{P}^2)_{LH} = \frac{189}{258}\frac{14}{258} + \frac{55}{258}\frac{51}{225} + \frac{14}{258}\frac{125}{189} \approx 0.124.$$

Starting from Low, the estimated probability of High congestion 30 minutes later is **12.4%**.

The estimated long-run distribution

Find $\widehat{\boldsymbol{\pi}} = \widehat{\boldsymbol{\pi}}\widehat{P}$, with $\sum_i \widehat{\pi}_i = 1$, by iterating in Excel.

1. Put **unrounded** $\widehat{P}$ in `B3:D5`. Rows = current state; columns = next state. Order both as Low, Medium, High.
2. Enter `1, 0, 0` in `B9:D9` (start in Low).
3. In `B10`: `=$B9*B$3+$C9*B$4+$D9*B$5`

   Copy right to `D10`; then fill `B10:D10` down to row 59.
4. Each row multiplies the previous distribution by $\widehat{P}$. At row 59 (50 steps), the values stabilize to six decimals:

$$\widehat{\boldsymbol{\pi}} \approx (0.379975,\ 0.334937,\ 0.285088).$$

Check: the values sum to 1 and stay stable in subsequent rows.

INTERPRETATION

At stationarity, about 28.5% of observations are in High: approximately 2.28 hours per eight hours, on average.