PIVOT TABLES

(continued...)

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.

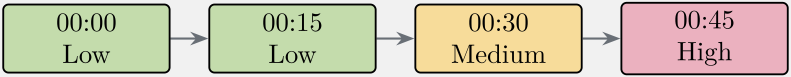
`warehouse_congestion_markov.xlsx`

- 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.

Why one equation is redundant

For any candidate row vector $\boldsymbol{u} = (u_L, u_M, u_H)$, define the **stationarity residual** for state j by

$$r_j(\boldsymbol{u}) = (\boldsymbol{u}\widehat{P})_j - u_j = \sum_{i \in \{L,M,H\}} u_i \widehat{p}_{ij} - u_j.$$

Stationarity means (L) $r_L(\boldsymbol{u}) = 0$, (M) $r_M(\boldsymbol{u}) = 0$, and (H) $r_H(\boldsymbol{u}) = 0$.

Since every row of $\widehat{P}$ sums to 1,

$$\begin{aligned}\sum_j r_j(\boldsymbol{u}) &= \sum_i u_i \left(\sum_j \widehat{p}_{ij}\right) - \sum_j u_j \\ &= \sum_i u_i - \sum_i u_i = 0.\end{aligned}$$

This identity holds for **every** candidate vector $\boldsymbol{u}$; neither stationarity nor normalization has been assumed.

THE PRECISE CONCLUSION

Since $r_H = -(r_L + r_M)$, $r_L = r_M = 0$ forces $r_H = 0$. The three equations are linearly dependent: any two imply the third. Keep two and impose $u_L + u_M + u_H = 1$ as the normalization equation.

Why we need the sum-to-one constraint

Multiplying a stationary vector by any number c preserves stationarity:

$$\widehat{\boldsymbol{\pi}} = \widehat{\boldsymbol{\pi}}\widehat{P} \quad \Longrightarrow \quad c\widehat{\boldsymbol{\pi}} = (c\widehat{\boldsymbol{\pi}})\widehat{P}.$$

A probability distribution must have **nonnegative entries summing to 1**.

Keep (L) and (M), and replace (H) with the sum-to-one condition:

$$\boxed{\begin{aligned} (\widehat{p}_{LL} - 1)\widehat{\pi}_L + \widehat{p}_{ML}\widehat{\pi}_M + \widehat{p}_{HL}\widehat{\pi}_H &= 0, \\ \widehat{p}_{LM}\widehat{\pi}_L + (\widehat{p}_{MM} - 1)\widehat{\pi}_M + \widehat{p}_{HM}\widehat{\pi}_H &= 0, \\ \widehat{\pi}_L + \widehat{\pi}_M + \widehat{\pi}_H &= 1. \end{aligned}}$$

Keep $\widehat{P}$ in `B3:D5`, then solve $A\,x = b$, with

	A: `H3:J5`			b: `L3:L5`
Row	**H**	**I**	**J**	**L**
3	`=B3-1`	`=B4`	`=B5`	`0`
4	`=C3`	`=C4-1`	`=C5`	`0`
5	**1**	**1**	**1**	**1**

Row 5 enforces the sum-to-one constraint.

The stationary eigenvector in Excel

Leave `N3:N5` empty, enter in `N3` and press Enter:

```
=MMULT(MINVERSE(H3:J5),L3:L5)
```

`MINVERSE` computes A^{-1}; `MMULT` multiplies it by b.

`N3:N5` returns **0.379975, 0.334937, 0.285088**, vertically in Low, Medium, High order.

Check: nonnegative values, sum $= 1$.

Exercise: start from Medium

QUESTIONS

Use the estimated matrix $\widehat{P}$.

1. If the current state is Medium, what is the complete state distribution after 15 minutes?
2. What is the complete state distribution after 30 minutes?
3. What is the probability of High congestion after 30 minutes?

Keep the full transition fractions during the calculation. Round only the final results.

Solution: two steps from Medium

After 15 minutes, read the Medium row:

$$\boldsymbol{\mu}_{t+1} = (0, 1, 0)\widehat{P} \approx (0.213,\ 0.560,\ 0.227).$$

After 30 minutes,

$$\boldsymbol{\mu}_{t+2} = (0, 1, 0)\widehat{P}^2 \approx (0.300,\ 0.412,\ 0.288).$$

Therefore,

$$\Pr(X_{t+2} = \text{High} \mid X_t = \text{Medium}) \approx \boxed{28.8\%}.$$

This is the probability of being High at the end of 30 minutes.

How operations can use the model

The transition matrix can support short-horizon planning.

- Forecast the next congestion distribution from the current state.
- Estimate how often a High state tends to persist.
- Compare transition patterns across shifts, staffing levels, or operating policies.

DECISIONS MAY CHANGE THE MATRIX

If managers add staff or change routing rules, the historical matrix may no longer describe future transitions. Re-estimate and validate the model after a material operating change.

Checks before using the estimates

1. Confirm that every row represents one consecutive 15-minute transition.
2. Exclude transitions across missing observations or system shutdowns.
3. Keep the state order Low, Medium, High in every table and matrix.
4. Verify that each probability row sums to 100%.
5. Check the number of departures from each state before trusting its row.
6. Test the fitted matrix on a later period and compare important operating segments.

The PivotTable makes the denominator visible. That visibility is a valuable error check.

What to remember

FOUR IDEAS

1. The current state and P determine the next-step distribution.
2. The formula $\boldsymbol{\mu}_{t+n} = \boldsymbol{\mu}_t P^n$ gives the distribution after n steps.
3. A stationary distribution satisfies $\boldsymbol{\pi} = \boldsymbol{\pi} P$.
4. A PivotTable estimates P by counting CurrentState-to-NextState transitions and converting each row to percentages.

YOUR TURN

the fraud-review queue

`fraud_review_queue_markov.xlsx`

The fraud-review queue

A digital bank's automated monitoring system routes flagged payments to an analyst queue. Every 30 minutes, the operations team records the number of open alerts and assigns one state.

Controlled	0–19 open alerts
Busy	20–49 open alerts
Backlog	50 or more open alerts

OPERATIONS QUESTION

Given the queue state now, what does the observed transition pattern imply about the next 30 or 60 minutes?

The fraud-review dataset

The Excel Table `ReviewTransitions` contains 630 valid 30-minute transitions from 14 monitoring days. Each row belongs to a Day, Evening, or Overnight session. Rows never join two different sessions.

Field	Meaning
`TransitionID`	unique transition identifier
`SessionID`, `SessionDate`	session identifier and date
`Shift`	Day, Evening, or Overnight
`IntervalStart`	start of the 30-minute interval
`CurrentState`, `NextState`	queue state now and 30 minutes later
`OpenAlertsStart`, `OpenAlertsNext`	open-alert counts behind the state labels
`AnalystsOnDuty`	analysts assigned to the session

UNIT OF ANALYSIS

One row represents one consecutive transition inside a monitoring session.

Your questions: the transition matrix

WORK IN THE STATE ORDER: CONTROLLED, BUSY, BACKLOG

1. Use a PivotTable to construct the transition-count matrix.
2. Convert the counts into the estimated transition-probability matrix.
3. Which state is most persistent? Use the diagonal probabilities to justify the answer.

Keep row totals visible. They are the denominators of the conditional probabilities.

Your questions: operational risk

USE THE SAME PIVOTTABLE ESTIMATE

4. If the queue is in Backlog now, what is the probability that it will be Controlled after 30 minutes?
5. If the queue is Busy now, what is the complete state distribution after 60 minutes?
6. Compare $\Pr(\text{Backlog}_{t+1} \mid \text{Busy}_t)$ across the three shifts. Which shift has the highest observed escalation probability?

Round displayed probabilities to one decimal place. Use the full-precision row percentages from the PivotTable, or the exact count ratios, for the two-step calculation.

Solution: transition counts

PivotTable fields

- Rows: `CurrentState`
- Columns: `NextState`
- Values: `TransitionID`
- Summarize values by **Count**

Order both state dimensions as Controlled, Busy, Backlog.

	NextState			
Current	**C**	**B**	**K**	**Total**
C	196	60	14	270
B	60	122	43	225
K	14	43	78	135
Grand total	270	225	135	630

C = Controlled, B = Busy, K = Backlog.

The row totals show 270 departures from Controlled, 225 from Busy, and 135 from Backlog.

Solution: transition probabilities

Open **Value Field Settings**. Under **Show Values As**, choose **% of Row Total**.

$$\widehat{P} = \begin{array}{c|ccc} & \text{Controlled} & \text{Busy} & \text{Backlog} \\ \hline \text{Controlled} & 72.6\% & 22.2\% & 5.2\% \\ \text{Busy} & 26.7\% & 54.2\% & 19.1\% \\ \text{Backlog} & 10.4\% & 31.9\% & 57.8\% \end{array}$$

The displayed Backlog row sums to 100.1% because of rounding. The underlying probabilities sum to 100%.

ANSWERS TO QUESTIONS 2–4

Controlled is the most persistent state because $196/270 = 72.6\%$ is the largest diagonal entry. From Backlog, the 30-minute probability of Controlled is $14/135 = 10.4\%$.

Solution: 60 minutes from Busy

Compute $\widehat{P}^2$ explicitly in Excel. The square means matrix multiplication: $\widehat{P}^2 = \widehat{P}\,\widehat{P}$. If $\widehat{P}$ occupies `M2:O4`, click `Q2` (the top-left cell of an empty 3×3 block) and enter

```
=MMULT($M$2:$O$4,$M$2:$O$4).
```

Microsoft 365 spills the full matrix $\widehat{P}^2$ into `Q2:S4`. Because Busy is the second state, read its 60-minute distribution in the second row, `Q3:S3`.

$$(0,1,0)\widehat{P}^2 \approx (0.358,\ 0.414,\ 0.228).$$

State after 60 minutes	Controlled	Busy	Backlog
Probability	35.8%	41.4%	22.8%

Solution: comparison across shifts

Add `Shift` to the PivotTable **Filters** area. Keep **% of Row Total**, select one shift at a time, and read the Busy row and Backlog column.

Shift	Busy to Backlog	Busy departures	Probability
Day	22	90	24.4%
Evening	14	79	17.7%
Overnight	7	56	12.5%

ANSWER TO QUESTION 6

The Day shift has the highest observed Busy-to-Backlog escalation probability: $22/90 = 24.4\%$.

YOUR TURN

protecting a live exam platform

`exam_platform_reliability_markov.xlsx`

The live-exam reliability problem

A university monitors its online exam platform every five minutes. The current median latency determines one of three states.

Stable	below 400 milliseconds
Degraded	400–1,999 milliseconds
Interrupted	2,000 milliseconds or more

The backup connection costs money, so it is activated only when the estimated probability of **at least one interruption during the next 15 minutes** is 15% or more.

OPERATIONS QUESTION

The platform is Stable at the beginning. Should operations activate the backup connection?

The exam-platform dataset

The Excel Table `PlatformTransitions` contains 1,960 valid five-minute transitions from 40 exam sessions. Each session contains 49 consecutive transitions; rows never join two different sessions.

Field	**Meaning**
`TransitionID`	unique transition identifier
`SessionID`, `SessionDate`	exam session and calendar date
`Window`, `IntervalStart`	session window and start of the five-minute interval
`CurrentState`, `NextState`	platform state now and five minutes later
`CandidatesOnline`	candidates connected during the interval
`MedianLatencyStartMs`, `MedianLatencyNextMs`	median platform response time (ms), measured at interval start and end

UNIT OF ANALYSIS

One row represents one consecutive five-minute transition inside an exam session.

Your questions: estimate the Markov model

WORK IN THE STATE ORDER: STABLE, DEGRADED, INTERRUPTED

1. Use a PivotTable to construct the transition-count matrix and the row-normalized transition-probability matrix $\widehat{P}$.
2. Is Interrupted an absorbing state? Quantify the one-step probability of recovering directly from Interrupted to Stable.
3. Starting from Stable, find the complete state distribution after 15 minutes, or three five-minute transitions.

Use the unrounded PivotTable percentages when calculating matrix powers.

Your questions: read the event carefully

THE WORDING OF THE RISK MEASURE MATTERS

4. Assume that the platform is Stable at time 0. What is the probability that it will be Interrupted *exactly* 15 minutes later?
5. Does this endpoint probability also equal the probability of at least one interruption between time 0 and time 15, given the same Stable initial state? Explain.
6. Starting from Stable at time 0, estimate the probability of at least one interruption during the following 15 minutes.
7. Apply the 15% backup rule. What decision follows? What decision would follow from incorrectly using the answer to Question 4?

Solution: estimate the transition matrix

In the PivotTable use `CurrentState` in Rows, `NextState` in Columns, and a Count of `TransitionID` in Values. If Excel sorts the labels alphabetically, move them manually into the order Stable, Degraded, Interrupted.

Transition counts

	S	D	I	Total
S	976	183	61	1,220
D	156	260	104	520
I	88	77	55	220
Total	1,220	520	220	1,960

Show Values As: % of Row Total

$$\widehat{P} = \begin{pmatrix} .80 & .15 & .05 \\ .30 & .50 & .20 \\ .40 & .35 & .25 \end{pmatrix}.$$

ANSWER TO QUESTION 2

Interrupted is not absorbing: $\widehat{p}_{II} = 25\%$, and the estimated probability of moving directly from Interrupted to Stable is $88/220 = 40\%$.

Solution: the 15-minute endpoint distribution

1. Put the unrounded matrix $\widehat{P}$ in `M3:O5`.
2. In `Q3`, enter
   ```
   =MMULT($M$3:$O$5,$M$3:$O$5)
   ```
   to spill $\widehat{P}^2$ into `Q3:S5`.
3. In `U3`, enter
   ```
   =MMULT($Q$3:$S$5,$M$3:$O$5)
   ```
   to spill $\widehat{P}^3$ into `U3:W5`.
4. Stable is the first state, so read the first row, `U3:W3`.

ANSWERS TO QUESTIONS 3–4

$$(1,0,0)\widehat{P}^3 = (0.660750,\ 0.240875,\ 0.098375).$$

The probability of being Interrupted exactly 15 minutes later is 9.8%.

Solution: endpoint probability versus path probability

Answer to Question 5: No.

Let X_k denote the platform state after k five-minute transitions. The value calculated in Question 4 answers an **endpoint** question:

$$(\widehat{P}^3)_{SI} = \Pr(X_3 = \text{Interrupted} \mid X_0 = \text{Stable}).$$

Question 6 asks about the whole path:

$$\Pr(X_1 = I \text{ or } X_2 = I \text{ or } X_3 = I \mid X_0 = S).$$

For example, the path

$$\text{Stable} \longrightarrow \text{Interrupted} \longrightarrow \text{Stable} \longrightarrow \text{Stable}$$

contains an interruption but is not counted by $(\widehat{P}^3)_{SI}$. This matters because Interrupted can be left in the original chain.

HOW TO ANSWER QUESTION 6

Track the probability of remaining in Stable or Degraded throughout all three transitions, then take the complement.

Solution: use a restricted transition matrix R

Compute the requested probability from its complement:

$$\Pr(\text{at least one } I \mid X_0 = S) = 1 - \Pr(\text{no } I \text{ in steps } 1, 2, 3 \mid X_0 = S).$$

What is R? For "no interruption", every transition must stay within Stable and Degraded. Take the S, D rows and columns of $\widehat{P}$:

$$R = \begin{array}{c|cc} & S & D \\ \hline S & .80 & .15 \\ D & .30 & .50 \end{array}.$$

Remove column and row I: entering in or starting from Interrupted disqualifies the path.

Do not renormalize R! The row deficits are $p_{SI} = .05$ and $p_{DI} = .20$: exactly the probabilities of entering Interrupted from Stable and Degraded.

How is R used? Start from $u_0 = (1, 0)$. Multiplying by R extends each interruption-free path by one step, so set $u_{k+1} = u_k R$. Thus u_k gives the probabilities of no interruption through step k, ending in S and D. After three steps,

$$u_3 = (.60650, .20025), \qquad .60650 + .20025 = .80675.$$

Answer: $1 - .80675 = .19325 \approx 19.3\%.$

Solution: track the surviving probability mass

Keep only transitions among Stable and Degraded:

$$R = \begin{pmatrix} .80 & .15 \\ .30 & .50 \end{pmatrix}.$$

The rows of R need not sum to 1: the missing probability is the mass that has already moved into Interrupted.

1. Put R in `M10:N11` and enter $1, 0$ in `M14:N14`.
2. In `M15`, enter

   `=MMULT(M14:N14,$M$10:$N$11).`

3. Select `M15` and fill it down through `M17`; each formula spills into the cell to its right. Each row gives the surviving Stable–Degraded mass after another step.
4. In `P17`, enter

   `=1-SUM(M17:N17).`

Alternative solution with an absorbing hit state

Reinterpret the state space so that it records whether Interrupted has ever been visited.

- S: no hit yet, currently Stable
- D: no hit yet, currently Degraded
- I_{hit}: at least one hit has occurred

Once the event occurs, I_{hit} is absorbing:

$$P_{\text{hit}} = \begin{array}{c|ccc} & S & D & I_{\text{hit}} \\ \hline S & .80 & .15 & .05 \\ D & .30 & .50 & .20 \\ I_{\text{hit}} & 0 & 0 & 1 \end{array}.$$

The upper-left block is R. The third column collects the mass that reaches Interrupted.

INTERPRETATION

I_{hit} means “an interruption has occurred”, not “currently Interrupted”. The original platform chain can still recover.

Three-step propagation with P_{hit}

$$h_0 = (1,0,0), \qquad h_k = h_0 P_{\text{hit}}^k.$$

The first two coordinates track paths that are still interruption-free. The third tracks paths that have hit Interrupted by step k.

Step k	**No-hit Stable**	**No-hit Degraded**	**Hit by step** k
0	1.00000	0.00000	0.00000
1	0.80000	0.15000	0.05000
2	0.68500	0.19500	0.12000
3	0.60650	0.20025	0.19325

Thus

$$(1,0,0)P_{\text{hit}}^3 = (.60650, .20025, .19325),$$

and

$$\Pr(X_1 = I \text{ or } X_2 = I \text{ or } X_3 = I \mid X_0 = S) = .19325 \approx 19.3\%.$$

Excel calculation of P_{hit}^3

1. Put P_{hit} in `M20:O22`, with rows and columns ordered as S, D, I_{hit}.
2. In `Q20`, enter

   ```
   =MMULT($M$20:$O$22,$M$20:$O$22)
   ```

   to spill P_{hit}^2 into `Q20:S22`.
3. In `U20`, enter

   ```
   =MMULT($Q$20:$S$22,$M$20:$O$22)
   ```

   to spill P_{hit}^3 into `U20:W22`.
4. Read the first-row, third-column entry in `W20`: 0.19325.

REPLACE THE ORIGINAL THIRD ROW

The row $(.40, .35, .25)$ allows recovery after an interruption. Replace it by $(0, 0, 1)$ so the bookkeeping chain remembers that the event occurred.

Decision from the absorbing hit state

Starting from Stable,

$$h_3 = (1,0,0)P_{\text{hit}}^3 = \big(0.60650,\ 0.20025,\ \boxed{0.19325}\big),$$

with coordinates ordered as (S, D, I_{hit}).

Since I_{hit} is absorbing and means that an interruption has occurred, its probability gives the answer directly:

$$\Pr(\exists\, k \in \{1,2,3\} : X_k = I \mid X_0 = S) = (h_3)_{I_{\text{hit}}} = 0.19325 \approx 19.3\%.$$

OPERATIONS DECISION

Since $19.3\% > 15\%$, activate the backup connection.

Manual calculation: organize all paths

Starting from $X_0 = S$, three transitions generate $3^3 = 27$ possible state sequences (X_1, X_2, X_3). These sequences are not equally likely. The $2^3 = 8$ sequences containing only S and D never visit I, so $27 - 8 = 19$ sequences contain at least one interruption. For any particular path,

$$\Pr(S \to x_1 \to x_2 \to x_3) = p_{Sx_1} p_{x_1x_2} p_{x_2x_3}.$$

Let $\tau_I = \min\{k \geq 1 : X_k = I\}$. Every path containing an interruption has one and only one first-hit time:

$$\{\tau_I \leq 3\} = \{\tau_I = 1\} \,\dot\cup\, \{\tau_I = 2\} \,\dot\cup\, \{\tau_I = 3\}.$$

First hit	**Required path prefix**	**Full paths represented**
Step 1	$S \to I$	$3^2 = 9$
Step 2	$S \to \{S, D\} \to I$	$2 \times 3 = 6$
Step 3	$S \to \{S, D\} \to \{S, D\} \to I$	$2^2 = 4$
Total		$9 + 6 + 4 = 19$

First hit at step 1: nine paths

Here $X_1 = I$, while X_2 and X_3 may be any state. The nine full paths are:

Path	**Product and probability**
$S \to I \to S \to S$	$.05(.40)(.80) = .016000$
$S \to I \to S \to D$	$.05(.40)(.15) = .003000$
$S \to I \to S \to I$	$.05(.40)(.05) = .001000$
$S \to I \to D \to S$	$.05(.35)(.30) = .005250$
$S \to I \to D \to D$	$.05(.35)(.50) = .008750$
$S \to I \to D \to I$	$.05(.35)(.20) = .003500$
$S \to I \to I \to S$	$.05(.25)(.40) = .005000$
$S \to I \to I \to D$	$.05(.25)(.35) = .004375$
$S \to I \to I \to I$	$.05(.25)(.25) = .003125$

Adding the nine mutually exclusive paths gives

$$\Pr(\tau_I = 1 \mid X_0 = S) = \boxed{.050000}.$$

This equals p_{SI} because the probabilities of all later continuations from I sum to one.

First hit at step 2: six paths

Now $X_1 \in \{S, D\}$ and $X_2 = I$. The final state X_3 may be any state, giving six full paths:

Path	**Product and probability**
$S \to S \to I \to S$	$.80(.05)(.40) = .016000$
$S \to S \to I \to D$	$.80(.05)(.35) = .014000$
$S \to S \to I \to I$	$.80(.05)(.25) = .010000$
$S \to D \to I \to S$	$.15(.20)(.40) = .012000$
$S \to D \to I \to D$	$.15(.20)(.35) = .010500$
$S \to D \to I \to I$	$.15(.20)(.25) = .007500$

The six path events are mutually exclusive, so their probabilities add. Separate the first three rows from the last three:

$$\begin{aligned}\Pr(\tau_I = 2 \mid X_0 = S) &= .80(.05)(.40 + .35 + .25) + .15(.20)(.40 + .35 + .25)\\ &= 0.04 + 0.03 = \boxed{0.07}\end{aligned}$$

First hit at step 3: four paths

For $\tau_I = 3$, $X_1, X_2 \in \{S, D\}$ and $X_3 = I$:

Path	**Product and probability**
$S \to S \to S \to I$	$.80(.80)(.05) = .032000$
$S \to S \to D \to I$	$.80(.15)(.20) = .024000$
$S \to D \to S \to I$	$.15(.30)(.05) = .002250$
$S \to D \to D \to I$	$.15(.50)(.20) = .015000$

The four path events listed on the previous slide are mutually exclusive, so their probabilities add:

$$\begin{aligned}\Pr(\tau_I = 3 \mid X_0 = S) &= .80(.80)(.05) + .80(.15)(.20) + \\ &\quad + .15(.30)(.05) + .15(.50)(.20) \\ &= \boxed{.073250}.\end{aligned}$$

Adding the three disjoint first-hit groups gives

$$\begin{aligned}\Pr(\tau_I \le 3 \mid X_0 = S) &= .050000 + .070000 + .073250 \\ &= .193250 = 19.325\%.\end{aligned}$$