Problem 1:

1)

1. **[4p]** Fill in the remaining part of the function `problem1_rejection` in order to produce samples from the density, using rejection sampling:

$$f(x) = C(\sin(x))^{10}, \quad 0 \le x \le \pi$$

where (C) is a value such that (f) above is a density (i.e. integrates to one).
Hint: you do not need to know the value of (C) to perform rejection sampling.

So: We want to sample from the density

$f(x) = C \cdot (\sin(x))^{10}$, $0 \le x \le \pi$, C unknown

we use Rejection sampling.

Step 1: Identify unnormalized target.

This means we simply remove the constant from the density function, meaning we have:

$\tilde{f}(x) = (\sin(x))^{10}$

Step 2: Choose a proposal distribution $g(x)$:

We choose a Uniform distribution on the interval $[0, \pi]$ (given from exercise)

$X \sim \text{Uniform}(0, \pi)$

For a uniform distribution on an interval of length π, the probability density is constant everywhere on that interval and zero outside it.

A probability density must integrate to 1, so we require: $\int_0^\pi g(x)\,dx = 1$

If $g(x)$ is constant, call it k:

$$\int_0^{\pi} k\,dx = k\pi = 1$$

$$\Rightarrow k\pi = 1 \Rightarrow k = \frac{1}{\pi} \Rightarrow g(x) = \frac{1}{\pi}$$

$$\Rightarrow g(x) = \frac{1}{\pi}, \quad (0 \le x \le \pi)$$

Step 3:

Finding M:

we need to bound the ratio

$$\frac{\tilde{f}(x)}{g(x)} = \frac{(\sin(x))^{10}}{\frac{1}{\pi}} = \pi \cdot (\sin(x))^{10}$$

M is always the maximum of $\frac{\tilde{f}(x)}{g(x)}$

and $(\sin(x))^{10} \le 1$, meaning maximum (M) for $\pi \cdot (\sin(x))^{10} \le \pi$, this means $M = \pi$

$$M = \pi$$

Step 4:

Acceptance probability:

The acceptance probability is

$$\alpha(x) = \frac{\tilde{f}(x)}{M \cdot g(x)} = \frac{(\sin(x))^{10}}{\pi \cdot \frac{1}{\pi}} = (\sin(x))^{10}$$

So we accept if $U \le (\sin(x))^{10}$.

Step 5 : Final algorithm:

- To generate one sample:

1) Draw $x \sim \text{Uniform}(0, \pi)$

This is proposed distribution $g(x)$.

2) Compute $\tilde{f} = (\sin(x))^{10}$

This is unnormalized targed density

3) Draw $U \sim \text{Uniform}(0, 1)$

This U is used for the acceptance test.

It's just a uniform random number for the rejection decision.

4) Accept x if $U \le \tilde{f}$

This is the acceptance probability:

$$\alpha(x) = \frac{\tilde{f}(x)}{M \cdot g(x)}$$

Problem 1 :

3)

3. **[2p]** Define (X) as a random variable with the density given in part 1. Denote

$$Y = \left(X - \frac{\pi}{2}\right)^2$$

and use the 10 000 samples to estimate $E[Y]$. Store the result in `problem1_expectation`.

We want to estimate $E[Y] = E\left[\left(X - \frac{\pi}{2}\right)^2\right]$

Using Monte Carlo, if we have samples $X_1, \dots, X_n$ drawn from the distribution of X, then a natural estimator of $E[Y]$ is:

$$E[Y] \approx \frac{1}{n} \cdot \sum_{i=1}^{n} \left(X_i - \frac{\pi}{2}\right)^2$$

Step-by-step procedure:

1) Use the rejection sampling function from Part 1 to generate $n = 10000$ samples from the distribution X: $X_1, \dots, X_n$

2) For each sample X_i, compute:

$$Y_i = \left(X_i - \frac{\pi}{2}\right)^2$$

3) Approximate $E[Y]$ by taking the average of these Y_i values.

4) In code, do this efficiently by doing:

- Store all X_i in an array
- Compute Y = (problem1_samples - np.pi/2)**2.
- Then taking np.mean(Y)

Problem 1:

4

4. **[2p]** Use Hoeffding's inequality to produce a 95% confidence interval of the expectation above and store the result as a tuple in the variable `problem1_interval`.

From the previous part:

we have i.i.d samples $X_1, \dots, X_n$ from density $f(x)$ on $[0, \pi]$.

- We define $Y_i = (X_i - \frac{\pi}{2})^2$

we are interested in the expectation

$$E[Y] = E[(X - \tfrac{\pi}{2})^2].$$

Our estimator is the sample mean:

$$\bar{Y} = \frac{1}{n} \cdot \sum_{i=1}^{n} Y_i = \frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \tfrac{\pi}{2})^2$$

Step 4: Find bounds $[a, b]$ for Y_i.

we need constants a and b such that

$$a \leq Y_i \leq b$$

we know:

- $X \in [0, \pi]$
- The midpoint is $\frac{\pi}{2}$

We need to find the minimum value of Y which is a, then the maximum value of Y which is b, and try different combinations of X inside Y to achieve this.

So: $Y = (X - \frac{\pi}{2})^2$

a = minimum value: $X = \frac{\pi}{2} \Rightarrow Y = 0$

So: $a = 0$

b = maximum value: $X = 0$ or $\pi \Rightarrow Y = (\frac{\pi}{2})^2$

So: $b = (\frac{\pi}{2})^2$

Step 2: Hoeffding's inequality:

For i.i.d random variablem $Y_1, \dots, Y_n$ with each $Y_i \in [0, \pi]$, Hoeffding's inequality says:

$$\mathbb{P}(|\bar{Y} - E[Y]| \geq \varepsilon) \leq 2\exp\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right)$$

We want a 95% confidence interval. This means that we want:

$$\mathbb{P}(|\bar{Y} - E[Y]| < \varepsilon) \geq 0.95$$

Equivalently:

$$\mathbb{P}(|\bar{Y} - E[Y]| \geq \varepsilon) \leq 0.05$$

So: set $\delta = 0.05$

and choose ε so that:

$$2\exp\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right) = \delta$$

$$= 2 \cdot e^{\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right)}$$

<u>Now we solve:</u>

$$2 \cdot e^{\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right)} = \delta \iff e^{\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right)} = \frac{\delta}{2}$$

$$\Rightarrow \frac{-2n\varepsilon^2}{(b-a)^2} = \ln\left(\frac{\delta}{2}\right)$$

$\Rightarrow$ (Multiply both sides with -1)

$$\Rightarrow \frac{2n\varepsilon^2}{(b-a)^2} = -\ln\left(\frac{\delta}{2}\right)$$

$$\Rightarrow \frac{2n\varepsilon^2}{(b-a)^2} = \ln\left(\frac{2}{\delta}\right)$$

$$\Rightarrow \varepsilon^2 = \frac{(b-a)^2}{2n} \cdot \ln\left(\frac{2}{\delta}\right)$$

$$\Rightarrow \varepsilon = (b-a) \cdot \sqrt{\frac{\ln\left(\frac{2}{\delta}\right)}{2n}}$$

$$\Rightarrow \underline{\underline{\varepsilon \approx 0.0335.}}$$

Last step: The 95% confidence interval:

Hoeffding gives:

$$\mathbb{P}(|\bar{Y} - E[Y]| \leq \varepsilon) \geq 1 - \delta = 0.95$$

so a valid 95% confidence interval is

$$[\bar{Y} - \varepsilon, \bar{Y} + \varepsilon].$$

Problem 1:

(5)

5. **[4p]** Can you calculate an approximation of the value of (C) from part 1 using random samples?
Provide a plot of the histogram from part 2 together with the true density as a curve (this requires the value of (C)).
Explain what method you used and what answer you got.

I'm using Monte Carlo integration with a Uniform$(0, \pi)$ distribution to approximate $\int_0^\pi (\sin(x))^{10} dx$, and then taking $C \approx 1/\hat{I}$.

Step 1: what is C?

We have the density $f(x) = C \cdot (\sin(x))^{10}$, $0 \leq x \leq \pi$

Because f is a probability density, it must integrate to 1:

$$\int_0^\pi f(x) dx = \int_0^\pi C \cdot (\sin(x))^{10} dx = 1$$

<u>So</u>: $C \cdot \int_0^\pi (\sin(x))^{10} dx = 1 \iff C = \dfrac{1}{\int_0^\pi (\sin(x))^{10} dx}$

Call $I = \int_0^\pi (\sin(x))^{10} dx$

Then $C = \dfrac{1}{I}$

<u>Problem</u>: We can't easily compute I exactly, so we approximate it from random samples.

<u>Step 2</u>: Monte carlo integration with $U \sim \text{Uniform}(0, \pi)$

- Let U be a random variable with $U \sim \text{Uniform}(0, \pi)$,

it's density is same as exercise ①,

$g(x) = \frac{1}{\pi}$, $0 \le x \le \pi$

For any function $h(x)$,

$$E[h(U)] = \int_0^\pi h(x) g(x) dx = \int_0^\pi h(x) \cdot \frac{1}{\pi} dx = \frac{1}{\pi} \cdot \int_0^\pi h(x) dx$$

Rearrange: $\int_0^\pi h(x) dx = \pi \cdot E[h(U)]$

now $h(x) = \tilde{f}(x) = (\sin(x))^{10}$

$$\Rightarrow I = \int_0^\pi ((\sin(x))^{10} dx = \pi E[\sin(U)^{10}]$$

0

<u>So</u>: If we can approximate $\pi \cdot E[\sin(U)^{10}]$, we get an approximation to I.

<u>Step 3</u>: Monte Carlo estimator for I:

Take N i.i.d samples:

$$U_1, U_2, \ldots, U_N \overset{iid}{\sim} \text{Uniform}(0, \pi)$$

Compute:

$$h(U_i) = (\sin(U_i))^{10}, \quad i = 1, \ldots, N$$

The sample mean of these values approximates the expectation:

$$\frac{1}{N} \cdot \sum_{i=1}^{N} h(U_i) \approx E[h(U)] = E[(\sin(U))^{10}]$$

<u>So</u>: A monte carlo estimator of the integral is:

$$\hat{I} = \frac{\pi}{N} \cdot \sum_{i=1}^{N} h(U_i)$$

As N gets large, $\hat{I} \to I$ (Law of large numbers).

<u>Step 4</u>: Estimator for C

Recall $C = \frac{1}{I}$

We plug in $\hat{I}$ instead of the unknown true I:

$$\hat{c} = \frac{1}{\hat{I}} = \frac{1}{\frac{\pi}{N} \cdot \sum_{i=1}^{N} (\sin(U_i))^{10}} = \frac{N}{\pi \cdot \sum_{i=1}^{N} (\sin(U_i))^{10}}$$

This is the answer:

$$\hat{c} = \frac{1}{\frac{\pi}{N} \cdot \sum_{i=1}^{N} (\sin(U_i))^{10}}$$

Problem 2:

① 1. [3p]

Use **Hoeffding's inequality** and compute the **95% confidence intervals** for **precision and recall** (etc.) on the **test set**. Store your intervals for each class in the variables:

- `problem2_precision0`
- `problem2_recall0`
- `problem2_precision1`
- `problem2_recall1`

Each of these should be a **tuple** `(lower, upper)`.

We want a 95% Hoeffding confidence interval for:

- Precision and Recall for Class 1, and
- Precision and Recall for Class 0

based on the test set.

So:

Each of these metrics is just an average of 0 and 1 outcomes (success/failure), so it's Bernoulli sample mean, and Hoeffding applies with

$a = 0$

$b = 1$

$$\Rightarrow e^{\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right)} = e^{\left(\frac{-2n\varepsilon^2}{(1-0)^2}\right)} = e^{-2n\varepsilon^2}$$

Step 1: Hoeffding inequality and solving ε.

General Hoeffding for bounded variables in $[a, b]$:

$$\mathbb{P}(|\hat{p} - p| \geq \varepsilon) \leq 2e^{\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right)}$$

For Bernoulli, values are as stated above in $[0, 1]$, so $a = 0$, $b = 1$:

$$\Rightarrow \mathbb{P}(|\hat{p} - p| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}$$

We want a 95% confidence interval, so:

- Set right hand side equal to $\alpha = 0.05$

$$\Rightarrow 2e^{-2n\varepsilon^2} = 0.05$$

$$\Rightarrow -2n\varepsilon^2 = \ln(0.025)$$

$$\Rightarrow \varepsilon^2 = -\frac{\ln(0.025)}{2n}$$

$$\Rightarrow \varepsilon = \sqrt{\frac{-\ln(0.025)}{2n}}$$

Step 2 :

- From the test set predictions we need to make a confusion matrix and use ravel() to convert multi-dimension to 1d-array with the FN, TP.... values to calculate Recall and Precision.

- **TP**: Predicted 1, true 1
- **FP** : Predicted 1, true 0
- **FN** : Predicted 0, true 1
- **TN** : Predicted 0, true 0

Step 3 :

Class 1 precision:

$$\hat{p} = \widehat{\text{Precision}}_1 = \frac{TP}{TP+FP}$$

$\hat{p}$ in Bernoulli

$n_{prec^{(1)}} = TP + FP$ (use this n in the Bernoulli)

$$\hat{p} = \underbrace{\frac{1}{n_{prec^{(1)}}} \cdot \sum_{j=1}^{n_{prec^{(1)}}} z_j^{(prec^{(1)})}}_{\text{sample mean}} = \frac{TP}{TP+FP}$$

This means we only consider the test points where the model predicted 1.

Class 1 Recall :

$$\hat{P} = \widehat{recall}_{(1)} = \frac{TP}{TP+FN}$$

$$n_{rec^{(1)}} = TP+FN$$

Treating class 0 as the positive class (Class 0 precision)

$$\hat{P} = \widehat{Precision}_{(0)} = \frac{TN}{TN+FN}$$

$$n_{prec^{(0)}} = TN+FN$$

Class 0 Recall :

$$\hat{P} = \widehat{recall}_{(0)} = \frac{TN}{TN+FP}$$

$$n_{rec^{(0)}} = TN+FP$$

Step 4 :

The intervals are :

- Before clipping : $\hat{p} \pm \epsilon$

• Clip to $[0,1]$:

$\Rightarrow$ lower $= \max\{0, \hat{p} - \varepsilon\}$

upper $= \min\{1, \hat{p} + \varepsilon\}$

7. Recap in one sentence per metric

- **Precision1**: average of "is this predicted-1 point actually 1?" over TP+FP trials.
- **Recall1**: average of "for this actual-1 point, did we predict 1?" over TP+FN trials.
- **Precision0**: average of "is this predicted-0 point actually 0?" over TN+FN trials.
- **Recall0**: average of "for this actual-0 point, did we predict 0?" over TN+FP trials.

Each is a Bernoulli mean → apply Hoeffding with that specific n, and form $\hat{p} \pm \sqrt{-\ln(0.025)/(2n)}$ clipped to $[0,1]$.

Problem 2:

2

2. [3p]

You are interested in minimizing the **average cost** of your classifier.
The hospital will use the model as a screening tool:

- If the model predicts **CHD = 1**, the patient is sent for further investigation.
- If the model predicts **CHD = 0**, nothing is done.

You decide to use the following costs:

- True positive (CHD = 1, predicted 1): cost = 0
- True negative (CHD = 0, predicted 0): cost = 0
- False positive (CHD = 0, predicted 1): cost = 10
- False negative (CHD = 1, predicted 0): cost = 300 *(worst case)*

Complete the function `problem2_cost(model, threshold, X, Y)` to compute the **average cost per person** for a given prediction threshold, using `model.predict_proba`.

1: What "average cost per person" mean:

Imagine we freeze a threshold t (say $t = 0.5$) and look only at the test set (or any dataset X, Y).

step 1:

For each person i, the model gives you

$$p_i = \mathbb{P}(CHD = 1 \mid X_i)$$

Ex:

patient	p_i (model prob for CHD=1)	true y_i
1	0.83	1 (by $0.83 \geq t = 0.5$)
2	0.12	0
⋮	⋮	⋮

These probabilities are:

```
Probas = model.predict_proba(X)[:,1]
```

"array of p_i"

Step 2:

Turn probabilities into predictions via threshold.

- Pick threshold t. (0.5 this case)
- If $p_i \geq t \rightarrow$ Predict $\hat{y}_i = 1$
- If $p_i < t \rightarrow$ Predict $\hat{y}_i = 0$

This is:

```
y_pred = (probas >= threshold).astype(int)
```

Step 3: Classify each prediction as TP, TN, FP, FN

For each patient, compare y_i and $\hat{y}_i$:

- TP: $y=1$, $\hat{y}=1$
- TN: $y=0$, $\hat{y}=0$
- FP: $y=0$, $\hat{y}=1$
- FN: $y=1$, $\hat{y}=0$

Step 4: Translate to costs

The exercise gives the cost table:

- TP: cost = 0
- TN: cost = 0
- FP: cost = 10
- FN: cost = 300

- So, total cost over all patients are:

  $$10 \cdot FP + 300 \cdot FN$$

  because TP and TN contribute 0.

- If $FP = 8$ and $FN = 3$, then total cost would be:

  $$10 \cdot 8 + 300 \cdot 3 = 980$$

Step 5: Average cost per person

If we have N persons (rows in dataset), the average cost per person is:

$$\text{Average Cost} = \frac{\text{Total Cost}}{N} = \frac{10 \cdot FP + 300 \cdot FN}{N}$$

OBS:
" X is just the dataset we are evaluating".

Problem 2:

③

3. [4p]

Select the **threshold** between 0 and 1 that minimizes the **average cost** on the **test set**.
Check, for example, **100 evenly spaced thresholds** between 0 and 1.

Store:

- the optimal threshold in `problem2_optimal_threshold`
- the cost at this threshold (on the test set) in `problem2_cost_at_optimal_threshold`

Only done in code!

Problem 2:

④

4. [3p]

With your newly computed threshold, compute the **cost of putting the model in production** by evaluating the cost on the **validation set**.

Also compute a **99% confidence interval** for this cost using **Hoeffding's inequality**, and store it as:

- `problem2_cost_at_optimal_threshold_validation`
- `problem2_cost_interval = (lower, upper)`

Step 2.1 — Define the random variable

For each patient i in the validation set:

- We apply the model with the chosen threshold.
- The patient gets a **cost** depending on the outcome:
  - True positive (Y=1, pred=1): cost = 0
  - True negative (Y=0, pred=0): cost = 0
  - False positive (Y=0, pred=1): cost = 10
  - False negative (Y=1, pred=0): cost = 300

Call this **per-person random cost** C_i.

So for each patient:

$$C_i \in \{0, 10, 300\}$$

Now:

Hoeffding range $[a, b]$:

Hoeffding needs that each variable is bounded:

$$a \leq C_i \leq b$$

Here:

- Minimum possible cost per person: $a = 0$
- Maximum possible cost per person: $b = 300$

On the validation set of size n, we compute

$$\hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} C_i$$

This is exactly what problem2_cost() returns:

"average cost per person on that dataset".

So: Hoeffding's inequality (99 % confidence):

Hoeffding for bounded variables says:

$$\mathbb{P}(|\hat{\mu}-\mu| \geq \varepsilon) \leq 2 \cdot e^{\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right)}$$

we want a 99% confidence interval, so:

confidence interval: $1-\alpha = 0.99$

So $\alpha = 0.01$

we choose ε such that:

$$2e^{\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right)} = \alpha$$

Solve for ε:

$$e^{\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right)} = \frac{0.01}{2}$$

$$\Rightarrow \frac{-2n\varepsilon^2}{(b-a)^2} = \ln(0.005)$$

$$\Rightarrow \varepsilon^2 = -\frac{(b-a)^2 \cdot \ln(0.005)}{2n}$$

$$\Rightarrow \varepsilon = (b-a) \cdot \sqrt{\frac{-\ln(0.005)}{2n}}$$

$$a = 0 \ , \ b = 300$$

$$\Rightarrow \varepsilon = 300 \cdot \sqrt{\frac{-\ln(0.005)}{2n}}$$

Confidence interval:

From before:

- $\hat{\mu}$ = problem2_cost_at_optimal_threshold_validation().
- ε as above

Then the 99% Hoeffding confidence interval is:

$$[\hat{\mu} - \varepsilon \ , \ \hat{\mu} + \varepsilon]$$

If we want to be super strict about bounds, we clip at $[0, 300]$:

- $\text{lower} = \max[0, \hat{\mu} - \varepsilon]$
- $\text{upper} = \min[300, \hat{\mu} + \varepsilon]$

Problem 3:

1) What is the transition matrix?

Matrix A:

	A	B	C	D
A	0	0.2	0	0.8
B	0	0	1	0
C	0	1	0	0
D	0.5	0	0.5	0

Matrix B:

	A	B	C	D	E	F
A	0	1	0	0	0	0
B	0	0	1	0	0	0
C	0	0.5	0	0.5	0	0
D	0	0	0.5	0	0.5	0
E	0	0	0	0	0	1
F	0.5	0	0	0	0.5	0

2) Are the Markov chains irreducible?

- A Markov chain is irreducible if:
  - You can get from every state to every other state
  - As long as there is some path with positive probability.
  - * If matrix has no zeros at all, it is trivially irreducible, because all transitions can occur in 1 step

* This does not need to necessarily happen in one step

So: No zeros in matrix -> irreducible !

So:

- **Matrix A**: Not irreducible because states B and C can never reach A or D.

- **Matrix B**: This is irreducible because all 6 states are connected in a way that you can move from any state to any other via some path with positive probability.

③ (i) Is the Markov chains aperiodic?
(ii) What is the period of each state?

- (ii): The period tells us:
  "What is the greatest common diviser (GCD) of all times t where you can return to that same state exactly at time t.

Matrix A: A -> A in 2 steps
B -> B in 2 steps
C -> C in 2 steps
D -> D in 2 steps

$\Rightarrow GCD = (2,2,2,2) = 2$

Matrix B: $A \to A$ in 6 steps
$B \to B$ in 2 steps
$C \to C$ in 2 steps
$D \to D$ in 2 steps
$E \to E$ in 2 steps
$F \to F$ in 2 steps

$\Rightarrow GCD(6,2,2,2,2,2) \Rightarrow \underline{\underline{2}}$

$6 = 2 \cdot 3$

- (i): Matrix is aperiodic if every state have period 1.

Answer: All have period 2, meaning no matrix is aperiodic

OBS: If Matrix A would look like [2, 2, 2, 3], then GCD would be 1. Ant then it would be aperiodic.

---

5. T = The first hitting time of state D, starting from state A.

- Compute $P(T=1), \dots, P(T=5), P(T=\infty)$

* General method for first hitting times

To compute $P(T=t)$, $T = \min\{k : X_k = D\}$

1: look at the transition matrix to find all possible paths starting at A.

2: For $P(T=t)$:

- list all paths of length t that
  - do not enter D at steps $1, 2, \dots, t-1$
  - enter D exactly at step t.
- Multiply the transition probabilities along each valid path.
- Sum the probabilities if more than one valid path exists.

3: For $P(T=\infty)$:

$$P(T=\infty) = 1 - \sum_{t=1}^{\infty} P(T=t)$$

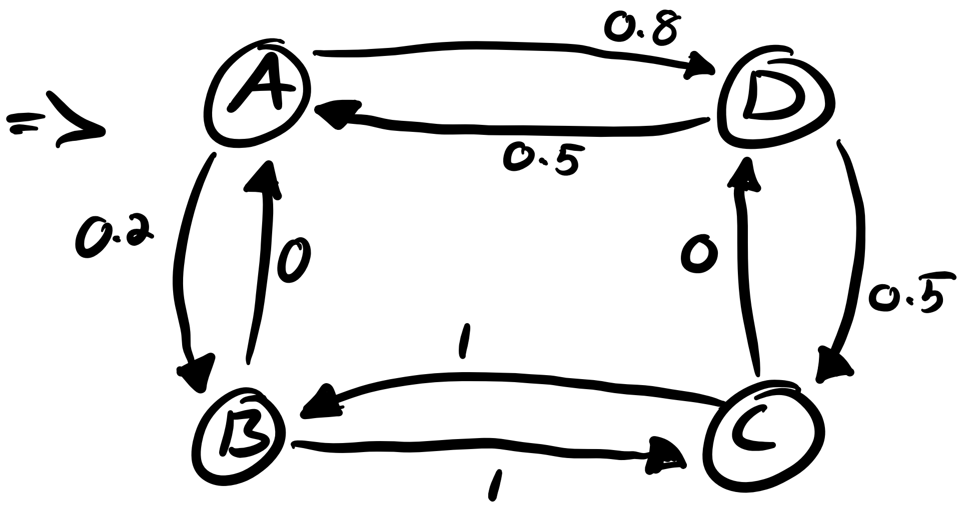
or identify cycles that never reach D.

So: (solution)

Chain A:

$$\begin{array}{c|cccc} & A & B & C & D \\ \hline A & 0 & 0.2 & 0 & 0.8 \\ B & 0 & 0 & 1 & 0 \\ C & 0 & 1 & 0 & 0 \end{array} \Rightarrow$$

$D \begin{pmatrix} 0.5 & 0 & 0.5 & 0 \end{pmatrix}$



Ex: From A: to B: 0.2
to D: 0.8

- $P(T=1)$: This means we hit D immediately:

$$P(A \to D) = 0.8, \text{ so } \boxed{P(T=1) = 0.8}$$

- $P(T=2)$: We must avoid D at first step, then enter D at second step:

- Only way to avoid D from A is to go to B:

$$A \to B = 0.2$$

- Then from B to D is not possible ($B \to D = 0$).

So: $\boxed{P(T=2) = 0}$

- $P(T=3)$:

  - Paths: A -> (non - D) -> (non - D) -> D

  - The only way to avoid D from A is again:

    A -> B -> C -> D ?

    check transitions:

    - A -> B = 0.2
    - B -> C = 1
    - C -> D = 0

    Thus no path reaches D at step 3.

    $$\boxed{P(T=3) = 0}$$

- $P(T=4)$:

  Try: A -> B -> C -> B -> D ?

  Transitions:

  A -> B = 0.2
  B -> C = 1
  C -> B = 1
  B -> D = 0

  $$\Rightarrow \boxed{P(T=4) = 0}$$

- $P(T=5)$:

Try: $A \to B \to C \to B \to C \to D$

but $C \to D = 0$

$$\Rightarrow \boxed{P(T=5) = 0}$$

- $P(T=\infty)$:
  - we hit D only if we take $A \to D$ at time 1.

  otherwise we go:

  $A \to B \to C \to B \to C \to B \ldots$

  This is a 2-cycle and D will never be reached.

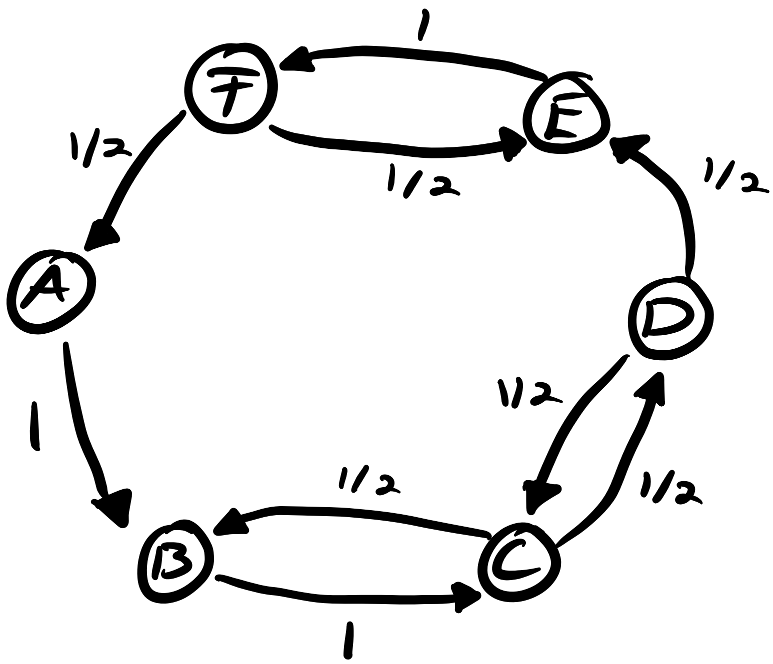
  Thus: $P(T=\infty) = 1 - \sum_{i=1}^{5} \underbrace{P(T_i = t)}_{1-0.8-0-0-0-0=0.2} = 1 - 0.8 = 0.2$

$$\Rightarrow \boxed{P(T=\infty) = 0.2}$$

- **Matrix B**:

	A	B	C	D	E	F
A	0	1	0	0	0	0
B	0	0	1	0	0	0
C	0	0.5	0	0.5	0	0
D	0	0	0.5	0	0.5	0
E	0	0	0	0	0	1
F	0.5	0	0	0	0.5	0

=>



$\boxed{P(T=1) = 0}$

- can't go from A to D in 1-time step

$\boxed{P(T=2) = 0}$, same here

- $\underline{P(T=3)}$:

  A -> B -> C -> D

  => $1 \cdot 1 \cdot 0.5 = 0.5$

  => $\boxed{P(T=3) = 0.5}$

- $\underline{P(T=4)}$:

  A -> B -> C -> B -> C

  This must be D

Therefore not possible!

$\Rightarrow \boxed{P(T=4) = 0}$

- $P(T=5)$:

$A \to B \to C \to B \to C \to D$

$\Rightarrow 1 \cdot 1 \cdot 0.5 \cdot 1 \cdot 0.5 = 0.25$

$\Rightarrow \boxed{P(T=5) = 0.25}$

- $P(T=\infty)$

-This chain will hit D with probability 1 because whenever you are in C, you have probability 0.5 of jumping into D, and you return to C infinetely often (from B).

-This is a classic recurrent non-absorbing but eventually absorbing chance 0.5 each visit

$\to$ absorption probability $= 1$

therefore $\boxed{P(T=\infty) = 0}$