

Percentiles and linear recursion

Wigner Summer Camp

Data and Compute Intensive Sciences Research Group

Balázs, Paszkál, Vince, Levente,
Antal
Éva, Hajni

7-11 July 2025



Table of Contents

Percentiles

Linear Recursions

Percentiles

What are Percentiles?

- ▶ A percentile indicates the value below which a given percentage of observations fall.
- ▶ For example, the 25th percentile (Q1) is the value below which 25% of the data fall.
- ▶ Common percentiles: 25th (Q1), 50th (median), 75th (Q3).

Numerical Example

Given the data: 1, 3, 4, 7, 8, 10, 12.

- ▶ 25th percentile \Rightarrow 3.5 (using linear interpolation, see next slide).
- ▶ 50th percentile (median) \Rightarrow 7.
- ▶ 75th percentile \Rightarrow 9.

Calculating Percentiles in PyTorch

```
1 torch.quantile(tensor, q=0.25,  
2                 interpolation='linear')
```

- ▶ To compute the quantile, we map $q \in [0, 1]$ to a quantile index $i_q = q \cdot (n - 1)$, where n is the number of data points.
- ▶ Let $i = \lfloor i_q \rfloor$, $j = \lceil i_q \rceil$, with sorted data values $a = x_i$ and $b = x_j$.
- ▶ Define $f = i_q - i$, the fractional part.
- ▶ Result is then computed as:
 - ▶ Linear: $a + (b - a) \cdot f$.
 - ▶ Lower: a .
 - ▶ Higher: b .
 - ▶ Nearest: a or b , whichever index is closer to i_q (rounding down at 0.5).
 - ▶ Midpoint: $(a + b)/2$.

Interpolation Methods: Example

Data: 10, 20, 30, 40. Then

- ▶ 25th percentile lies between 10 and 20.
 - ▶ Linear: $10 + (20 - 10) \cdot 0.25 = 12.5$.
 - ▶ Lower: 10.
 - ▶ Higher: 20.
 - ▶ Nearest: 10.
 - ▶ Midpoint: $(10 + 20)/2 = 15$.

Exercise 1: Skewed Toward Low Values

Given the data: 1, 1, 1, 2, 3, 4, 10

- ▶ Calculate the 25th, 50th, and 75th percentiles.

Solution to Exercise 1

Sorted data: 1, 1, 1, 2, 3, 4, 10

- ▶ 25th percentile: Between 1 and 1 $\Rightarrow 1$
- ▶ 50th percentile: Middle value = 2
- ▶ 75th percentile: Between 4 and 10
 $\Rightarrow 4 + 0.25 \cdot (10 - 4) = 5.5$

Exercise 2: Skewed Toward High Values

Given the data: 1, 6, 7, 8, 9, 9, 9

- ▶ Calculate the 25th, 50th, and 75th percentiles.

Solution to Exercise 2

Sorted data: 1, 6, 7, 8, 9, 9, 9

- ▶ 25th percentile: Between 1 and 6 $\Rightarrow 1 + 0.5 \cdot (6 - 1) = 3.5$
- ▶ 50th percentile: Middle value = 8
- ▶ 75th percentile: Between 9 and 9 $\Rightarrow 9$

Exercise 3: Values Near Zero

Given the data: 0, 0, 0, 1, 1, 2, 3

- ▶ Calculate the 25th, 50th, and 75th percentiles.

Solution to Exercise 3

Sorted data: 0, 0, 0, 1, 1, 2, 3

- ▶ 25th percentile: Between 0 and 0 $\Rightarrow 0$
- ▶ 50th percentile: Middle value = 1
- ▶ 75th percentile: Between 1 and 2 $\Rightarrow 1 + 0.25 \cdot (2 - 1) = 1.25$

Linear Recursions

What is a Linear Recursion?

- ▶ A sequence where each element is a linear combination of previous elements.
- ▶ General form:

$$x_n = w_1x_{n-1} + w_2x_{n-2} + \cdots + w_kx_{n-k}. \quad (1)$$

- ▶ Requires initial values x_0, \dots, x_{k-1} .
- ▶ Typically, $x_n, w_i \in \mathbb{R}$ (real numbers), but $x_n \in \mathbb{Z}$ or complex values are also common in specific applications.

Famous Examples

- ▶ Fibonacci sequence:

$$x_n = x_{n-1} + x_{n-2}, \quad x_0 = 0, \quad x_1 = 1. \quad (2)$$

- ▶ Exponential growth: $x_n = ax_{n-1}, a > 1$.
- ▶ Weighted average: $x_n = 0.8x_{n-1} + 0.2x_{n-2}$.

Effect of Weights and Initial Values

- ▶ Weights determine how past values influence the future.
- ▶ Initial values can lead to different growth or oscillation patterns.
- ▶ Some sequences stabilize, others diverge.

Recursions and Discretizing Functions

- ▶ Recursions can approximate continuous functions.

Summary

- ▶ **Percentiles** help understand the distribution of data, identifying values below which a given percentage of data falls.
 - ▶ Different interpolation methods influence the computed percentile value.
 - ▶ PyTorch supports flexible percentile computations.
- ▶ **Linear recursions** generate sequences based on weighted combinations of previous values.
 - ▶ Used to model growth, oscillations, or smooth approximations.
 - ▶ Provide a bridge between discrete sequences and continuous functions.