Empirical Cumulative Distribution

Song Liu (song.liu@bristol.ac.uk)

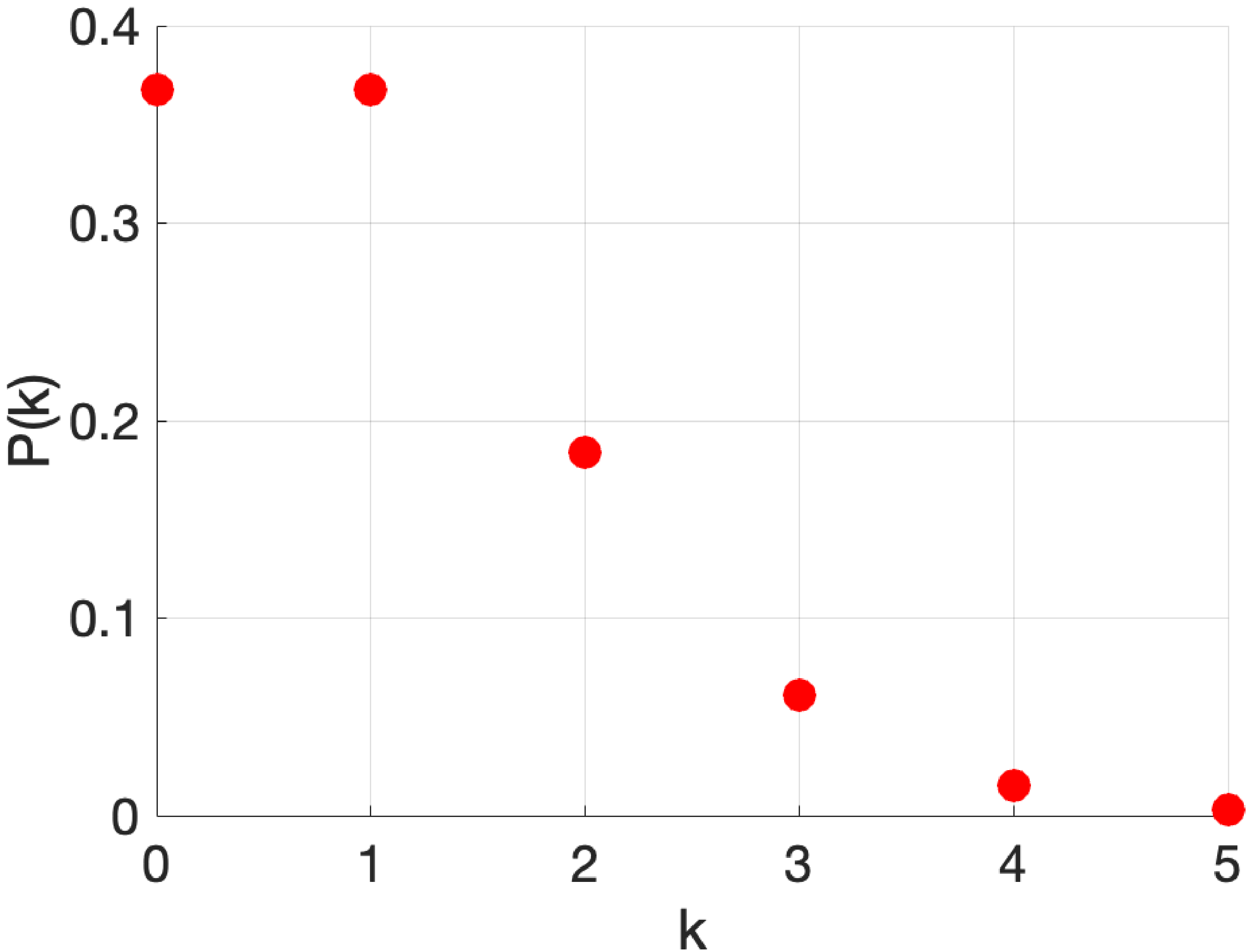
GA 18, Fry Building,

Microsoft Teams (search "song liu").

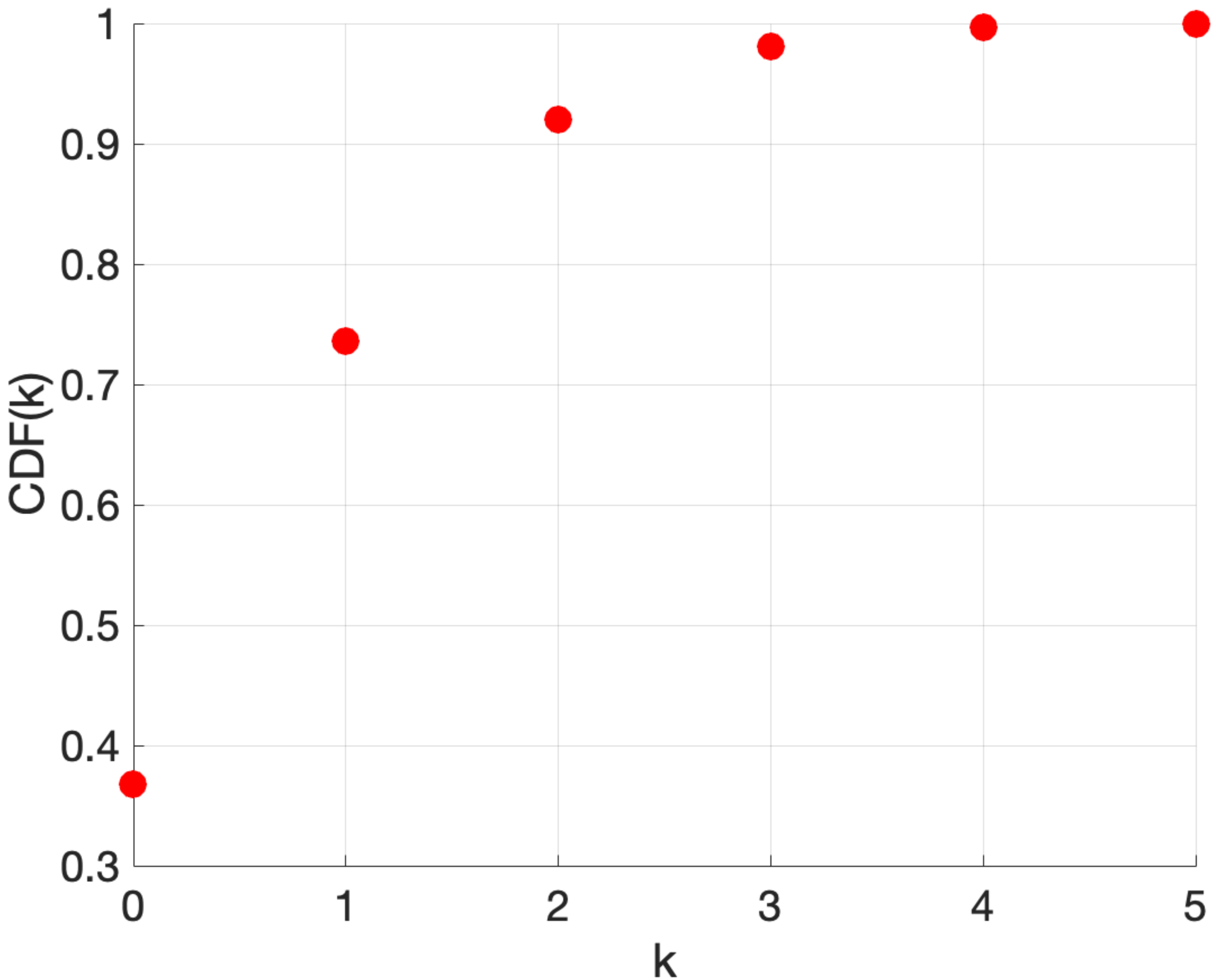
Poisson Distribution

- Poisson distribution is the probability of the number of events occurring during a fixed interval.
  - given a constant mean rate.
- On average, a flood happens once a hundred years. What is the probability of having k floods in the next a hundred years?
  - $P(k) = \frac{\lambda^k \exp(-\lambda)}{k!} = \frac{\exp(-1)}{k!}$
  - where λ is the mean rate.
  - Assuming the flood rate is a constant.

Probability Mass Function



Cumulative Distribution Function



Empirical Cumulative Distribution Function

- In reality, we do not know the distribution our data comes from.
- In this case, we can still visualize the distribution of our data using a concept called Empirical Cumulative Distribution Function (ECDF).
- Assume you have n random samples from an unknown distribution, $X_1, X_2, ...X_n$:
  - $\mathrm{ECDF}(k) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq k)$
  - $1(X \leq k) = 1$ if $X \leq k$, and 0 otherwise.

Computing ECDF

- Using a simple program, we can compute ECDF easily.
- Notice $\sum_{i=1}^{n} 1(X_i \leq k)$ is just the count of my samples which are $\leq k$.

Pseudo Code

- input: an `int` array `X` with length `len`.
- output: a `double` array `ECDF` with values of $\mathrm{ECDF}(k=0),\ldots\mathrm{ECDF}(k=5)$.
- For loop: k = 0...5
  - ECDF[i] = (double) count(k)/n;

Pseudo Code `count`

- input: an `int` array `X` with length `len`.
- input: `k`
- output: number of elements in `X` smaller or equal to `k`.
- c = 0
- For loop: i = 1...n
    - if x_i <= k
        - c++;

Practice

- Write your code using provided skeleton code and data.
- What distribution do you think the data come from?