Homework 3 is posted
due in class next Tuesday 4/9

Midterm is graded

<table>
<thead>
<tr>
<th></th>
<th>P1 (15)</th>
<th>P2 (10)</th>
<th>P3 (15)</th>
<th>P4 (10)</th>
<th>P5 (10)</th>
<th>P6 (10)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>13.6</td>
<td>9.33</td>
<td>14</td>
<td>9.3</td>
<td>9.7</td>
<td><strong>8.03</strong></td>
<td>63.9</td>
</tr>
<tr>
<td>St.Dev</td>
<td><strong>2.17</strong></td>
<td>1.58</td>
<td>2.32</td>
<td>2.1</td>
<td>1.2</td>
<td><strong>2.36</strong></td>
<td>7.79</td>
</tr>
<tr>
<td></td>
<td><strong>1.38</strong></td>
<td>0.67</td>
<td>1</td>
<td>0.7</td>
<td>0.3</td>
<td><strong>1.97</strong></td>
<td>6.08</td>
</tr>
</tbody>
</table>
Easy solution:
3 letters I have to go to places of L, L, N. For remaining letters, there are 3 spots for N, the rest is fixed. Number of permutations is $6!/(3!2!1!)=60$. Prob=3/60=1/20=0.05
6. **(10 points)** In a data communication system, several messages that arrive at a node are bundled into a packet before they are transmitted over the network. Assume the messages arrive according to a Poisson process with the mean rate equal to one message per two minutes. Five messages are required to form a packet and the packet is formed immediately after the last message has arrived.

**(a) (5 points)** What is the probability that a time interval between two consecutive messages is longer than 4 minutes?

**Answer:** \( \lambda = 1 \text{ message/2 minutes} = 0.5 \text{ messages/minute.} \)

Exponential distribution

\[
P(X > 4) = \exp(-0.5 \times 4) = \exp(-2) = 0.1353
\]

**(b) (5 points)** What is the mean time until a packet is formed, that is, until exactly five messages have arrived at the node?

**Answer:** Using Erlang distribution with \( r = 5 \), \( \lambda = 0.5 \) one gets \( (5/0.5) \text{ minutes} = 10 \text{ minutes} \)
Descriptive statistics: Populations, Samples, Histograms, Quartiles, Sample mean and variance
Two types of reasoning

Logical reasoning

Physical laws → Types of reasoning → Product designs

Population

Sample

Statistical reasoning: Inference of population properties from a finite sample
Numerical Summaries of Data

• Data are the numerical observations of a phenomenon of interest.

• The totality of all observations is a population.
  – Population can be infinite (e.g. random variable)
  – It can be very large (e.g. 7 billion humans)

• A (usually small) portion of the population collected for analysis is a random sample.

• We want to use sample to infer facts about populations

• The inference is not perfect but gets better and better as sample size increases.
Some Definitions

• The random variables $X_1, X_2, \ldots, X_n$ are a random sample of size $n$ if:
  a) The $X_i$ are independent random variables.
  b) Every $X_i$ has the same probability distribution.

• Such $X_1, X_2, \ldots, X_n$ are also called independent and identically distributed (or i. i. d.) random variables.
Ways to describe a sample:

Histogram
approximates PDF
(or PMF)
load PINT_binding_energy;
dfittool(binding_energy)
Histograms with Unequal Bin Widths

• If the data is tightly clustered in some regions and scattered in others, it is visually helpful to use narrow bin widths in the clustered region and wide bin widths in the scattered areas.

• To approximate the PDF, the rectangle area, not the height, must be proportional to the bin relative frequency.

\[
\text{Rectangle height} = \frac{\text{bin relative frequency}}{\text{bin width}}
\]
Cumulative Frequency Plot

- Cumulative probability
- Protein-protein binding energy (units of kT)
- Blue line: binding energy data
- Red line: normal fit with $\mu = -14.6$ and $\sigma = 2.92$
Median, Quartiles, Percentiles

• The **median** $q_2$ divides the sample into two equal parts: 50% ($n/2$) of sample points below $q_2$ and 50% ($n/2$) points above $q_2$

• The **three quartiles** partition the data into four equally sized counts or segments.
  – 25% of the data is less than $q_1$.
  – 50% of the data is less than $q_2$, the median.
  – 75% of the data is less than $q_3$.

• There are **100 percentiles**. $n$-th percentile $p_n$ is defined so that $n\%$ of the data is less than $p_n$
Why do whales jump from cliffs?

Why do I feel dizzy?

Why are there slaves in the Bible?

Why are there different fingerprints?

Why is there a line through the word “red” in my address?

Why do dogs attack people?

Why are there ants in my laptop?

Why do I say “uh”?

Why is sea salt better than table salt?

Why do I get cold sores?

Why do spiders bite itch?

Why are there so many spiders in my house?

Why are there so many ants in my room?

Why are there so many spiders in my house?

Why are there so many birds in my backyard?

Why are there so many bees in my house?

Why do I get hiccups?

Why are there so many bees in my house?

Why do I get cold sores?

Why are there so many ants in my room?

Why are there so many spiders in my house?

Why do I get hiccups?

Why are there so many birds in my backyard?

Why are there so many bees in my house?

Why do I get cold sores?

Why are there so many ants in my room?

Why are there so many spiders in my house?

Why do I get hiccups?

Why are there so many birds in my backyard?

Why are there so many bees in my house?

Why do I get cold sores?

Why are there so many ants in my room?

Why are there so many spiders in my house?

Why do I get hiccups?

Why are there so many birds in my backyard?

Why are there so many bees in my house?

Why do I get cold sores?

Why are there so many ants in my room?

Why are there so many spiders in my house?

Why do I get hiccups?

Why are there so many birds in my backyard?

Why are there so many bees in my house?

Why do I get cold sores?

Why are there so many ants in my room?

Why are there so many spiders in my house?

Why do I get hiccups?

Why are there so many birds in my backyard?

Why are there so many bees in my house?

Why do I get cold sores?

Why are there so many ants in my room?

Why are there so many spiders in my house?

Why do I get hiccups?

Why are there so many birds in my backyard?

Why are there so many bees in my house?

Why do I get cold sores?

Why are there so many ants in my room?

Why are there so many spiders in my house?

Why do I get hiccups?
Matlab exercise

• Find the median and lower & upper quartiles of n=100 sample drawn from a continuous uniform distribution in [0,1]

• Do not use built-in Matlab functions for this exercise!

• Hint: use [a,b]=sort(r1); to rank order your sample. The variable a returns r1 sorted in the increasing order.

• How to find quartiles from a?
How to find median & quartiles

- % Example: find median and lower quartile of
- % a sample with n=100 drawn from uniform
- r1=rand(100,1);
- [a,b]=sort(r1);
- med=(a(50)+a(51))./2
- sum(r1<med) % verify
- q1=(a(25)+a(26))./2
- sum(r1<q1) % verify
Box-and-Whisker Plot

- A box plot is a graphical display showing **Spread**, **Outliers**, **Center**, and **Shape** (SOCS).
- It displays the **5-number summary**: \( min \), \( q_1 \), **median**, \( q_3 \), and **max**.

**Figure 6-13** Description of a box plot.
Matlab exercise:

• Generate a sample with $n = 1000$ following standard normal distribution
• Calculate median, first, and third quartiles
• Calculate IQR and find ranges shown below
• Find and count left and right outliers
• Do not use built-in Matlab functions for this!
• Make box and whisker plot: use boxplot