Exercise 10-1  Conditional Probability

If screening for a disease, there are several possible outcomes. Let $T^+$, $\neg T$ denote the events that the test is positive and negative, respectively, and $D$, $\neg D$ denote the events of having and not having the disease, respectively. There are two major criteria to evaluate tests by:

- Sensitivity: Probability (in practice more likely: ratio) of positively tested people having the disease, i.e., $P(T \mid D)$.
- Specificity: Probability (or ratio) of negatively tested people not having the disease, i.e., $P(\neg T \mid \neg D)$.

Now, assume a (realistic) test for HIV with a sensitivity and specificity of 99.9%. Suppose that a person is randomly selected from a population where 1% are infected with HIV and tested with the aforementioned test. Compute the probability that the person has HIV if:

(a) The test is positive.
(b) The test is negative.

Exercise 10-2  Maximum Likelihood Estimator I

Suppose that $X$ is a discrete random variable with the following probability mass function, where $0 < \theta < 1$ is a parameter.

<table>
<thead>
<tr>
<th>$X$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X)$</td>
<td>$2\theta/3$</td>
<td>$\theta/3$</td>
<td>$2(1-\theta)/3$</td>
<td>$(1-\theta)/3$</td>
</tr>
</tbody>
</table>

The following 10 independent observations were taken from such a distribution: $(3, 0, 2, 1, 3, 2, 1, 0, 2, 1)$. What is the maximum likelihood estimate of $\theta$?
Exercise 10-3  Human Height

Assume that the height of a human from a finite population is a Gaussian random variable:

\[ P_w(x_i) = \mathcal{N}(x_i; \mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{(x_i - \mu)^2}{2\sigma^2} \right) \]

For independent \( x_i \in \mathbb{R} \) from such a population \( w = (\mu, \sigma)^T \in \mathbb{R}^2 \) holds

\[ P_w(x_1, \ldots, x_N) = \prod_{i=1}^{N} P_w(x_i) = \prod_{i=1}^{N} \mathcal{N}(x_i; \mu, \sigma^2) = \]

\[ = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)^2 \right) \]

a) Determine the maximum likelihood estimator of \( P_w(x_1, \ldots, x_N) \).

b) Compute the corresponding estimators for the four height datasets in the file body_sizes.txt and visualize the respective distributions. How does the estimator reflect the understanding of the underlying data?