Maximum-likelihood estimation
Given samples from a parameterized family of distributions $\mathcal{D}(\theta)$, determine $\theta$.

For example, consider the uniform distribution with parameters $\theta=(a,b)$: $\text{Unif}(a,b)$
paramater estimation

- Number of promoters
- Normalized CpG
Samples from $\text{Unif}(a,b)$

Compute samples $X_1, X_2, \ldots, X_n$

Compute $k$th sample moment: $\hat{m}_k = \frac{1}{n} \sum_{i=1}^{k} (X_i)^k$

Solve equations:

$$\hat{m}_1 = \frac{\hat{a} + \hat{b}}{2} \quad \hat{m}_2 - (\hat{m}_1)^2 = \frac{(\hat{a} - \hat{b})^2}{12}$$
Samples from $\text{Unif}(a,b)$

$$\hat{a} = \hat{m}_1 - \sqrt{3}\hat{\sigma} \quad \hat{b} = \hat{m}_1 + \sqrt{3}\hat{\sigma}$$

$$\hat{\sigma} = \sqrt{\hat{m}_2 - (\hat{m}_1)^2}$$

Is there a more principled way to think about the “best” estimator?
P(E | θ): Probability of event E given model θ

Viewed as a function of E (fixed θ), it's a probability.

Given sample HHTTHH of possibly biased coin flips, estimate:

θ = probability of Heads

P(HHTTHH | 0.6) > P(HHTTHH | 0.5)

i.e. event HHTTHH is more likely if θ = 0.6

Which value of θ makes HHTTHH most likely?
\[ P(\text{HHTHH} \mid \theta) \]:
probability of HHTHH, given \( P(H) = \theta \)

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>( \theta^4(1-\theta) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.0013</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0313</td>
</tr>
<tr>
<td>0.8</td>
<td>0.0819</td>
</tr>
<tr>
<td>0.95</td>
<td>0.0407</td>
</tr>
</tbody>
</table>

Likelihood
maximum-likelihood estimation (MLE)

**MLE:** One of many approaches to parameter est.

Likelihood of independent observations $x_1, x_2, \ldots$

$$L(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} P(x_i \mid \theta)$$

Now choose $\theta$ which maximizes the **likelihood**.

**Typical approach:**

$$\frac{\partial}{\partial \theta} L(\vec{x} \mid \theta) = 0 \quad \text{OR} \quad \frac{\partial}{\partial \theta} \log L(\vec{x} \mid \theta) = 0$$
\[
P(\text{HHTHH} \mid \theta) = \theta^4(1-\theta)
\]

<table>
<thead>
<tr>
<th>(\theta)</th>
<th>(\theta^4(1-\theta))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.0013</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0313</td>
</tr>
<tr>
<td>0.8</td>
<td>0.0819</td>
</tr>
<tr>
<td>0.95</td>
<td>0.0407</td>
</tr>
</tbody>
</table>

Probability of HHTHH, given \(P(H) = \theta\).
Example 1

\(N\) coins flips \(x_1, x_2, \ldots, x_N\)

\(N_0\) tails, \(N_1\) heads, \(N_0 + N_1 = N\)

\(\theta = \) probability of heads

\[
L(x_1, x_2, \ldots, x_n \mid \theta) = (1 - \theta)^{n_0} \theta^{n_1}
\]

\[
\log L(x_1, x_2, \ldots, x_n \mid \theta) = n_0 \log(1 - \theta) + n_1 \log \theta
\]

\[
\frac{\partial}{\partial \theta} \log L(x_1, x_2, \ldots, x_n \mid \theta) = \frac{-n_0}{1 - \theta} + \frac{n_1}{\theta}
\]

Setting to zero and solving:

\[
\hat{\theta} = \frac{n_1}{n}
\]
Given $N$ normal samples, estimate mean and variance.

\[
f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

\[
\theta = (\mu, \sigma^2)
\]
Suppose we know that $\sigma^2 = 1$, but $\mu$ is unknown.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\theta = (\mu, \sigma^2)$$
Suppose we know that $\sigma^2 = 1$, but $\mu$ is unknown.
Suppose we know that $\sigma^2 = 1$, but $\mu$ is unknown.
Suppose we know that $\sigma^2 = 1$, but $\mu$ is unknown.
Suppose we know that $\sigma^2 = 1$, but $\mu$ is unknown. Which is more likely?
\[ x_i \sim N(\mu, \sigma^2), \quad \sigma^2 = 1, \quad \mu \text{ unknown} \]

\[
L(x_1, x_2, \ldots, x_n | \theta) = \prod_{1 \leq i \leq n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \theta)^2}{2}}
\]

\[
\ln L(x_1, x_2, \ldots, x_n | \theta) = \sum_{1 \leq i \leq n} -\frac{1}{2} \ln 2\pi - \frac{(x_i - \theta)^2}{2}
\]

\[
\frac{d}{d\theta} \ln L(x_1, x_2, \ldots, x_n | \theta) = \sum_{1 \leq i \leq n} (x_i - \theta)
\]

\[
\frac{dL}{d\theta} = \left( \sum_{1 \leq i \leq n} x_i \right) - n\theta = 0
\]

\[
\hat{\theta} = \left( \sum_{1 \leq i \leq n} x_i \right) / n = \bar{x}
\]

Sample mean is MLE of actual mean.
Suppose we know that data is normal, but both $\sigma^2$ and $\mu$ are unknown.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$$

$\theta = (\mu, \sigma^2)$
Suppose we know that data is normal, but both $\sigma^2$ and $\mu$ are unknown.
Suppose we know that data is normal, but both $\sigma^2$ and $\mu$ are unknown.
Suppose we know that data is normal, but both $\sigma^2$ and $\mu$ are unknown.
Suppose we know that data is normal, but both $\sigma^2$ and $\mu$ are unknown.
Suppose we know that data is normal, but both $\sigma^2$ and $\mu$ are unknown.
example 3

\[ x_i \sim N(\mu, \sigma^2), \; \mu, \sigma^2 \text{ both unknown} \]

\[
\ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) = \sum_{1 \leq i \leq n} -\frac{1}{2} \ln 2\pi \theta_2 - \frac{(x_i - \theta_1)^2}{2\theta_2}
\]

\[
\frac{\partial}{\partial \theta_1} \ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) = \sum_{1 \leq i \leq n} \frac{(x_i - \theta_1)}{\theta_2} = 0
\]

\[
\hat{\theta}_1 = \left( \sum_{1 \leq i \leq n} x_i \right) / n = \bar{x}
\]

Likelihood surface

Sample mean is MLE of actual mean (again).
\[ x_i \sim N(\mu, \sigma^2), \ \mu, \sigma^2 \text{ both unknown} \]

\[
\ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) = \sum_{1 \leq i \leq n} -\frac{1}{2} \ln 2\pi \theta_2 - \frac{(x_i - \theta_1)^2}{2\theta_2}
\]

\[
\frac{\partial}{\partial \theta_2} \ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) = \sum_{1 \leq i \leq n} -\frac{1}{2} \frac{2\pi}{2\pi \theta_2} + \frac{(x_i - \theta_1)^2}{2\theta_2^2} = 0
\]

\[
\hat{\theta}_2 = \left( \sum_{1 \leq i \leq n} (x_i - \hat{\theta}_1)^2 \right) / n = \bar{s}^2
\]

A consistent but **biased** estimate of population variance.

**Unbiased:**

\[
\hat{\theta}_2' = \sum_{1 \leq i \leq n} \frac{(x_i - \hat{\theta}_1)^2}{n-1}
\]
Samples from \textbf{Unif}(a,b)

Determine the MLE estimates for \textbf{a} and \textbf{b}.

\[
L(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} P(x_i \mid \theta)
\]