Sample Variance
\[ N(x_1, \sigma^2) = N(\theta_1, \theta_2) \]
\[ x_1, x_2, \ldots, x_n \] are independent r.v.'s for \( N(\theta_1, \theta_2) \).
\[ \hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2 \]
\[ \text{Var}(Y) = E[(Y - \mu)^2] \] where \( \mu = E[Y] \).
\[ \theta_2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 \text{, uniform} \]
\( \hat{\theta}_2 \) is the variance in this case.

Sample Variance

Bias.

Definition: An estimator \( \hat{\theta} \) of \( \theta \) is unbiased if
\[ E[\hat{\theta}] = \theta. \]

Back to the MLEs \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) for \( N(\theta_1, \theta_2) \).
\[ E[\hat{\theta}_1] = E\left[ \frac{1}{n} \sum_{i=1}^{n} x_i \right] = \frac{1}{n} \sum_{i=1}^{n} E[x_i] = \frac{1}{n} \sum_{i=1}^{n} \theta_1 = \frac{1}{n} \cdot n \theta_1 = \theta_1, \text{ so } \hat{\theta}_1 \text{ is unbiased.} \]

But \( \hat{\theta}_2 \) is not an unbiased estimator for \( \theta_2 \).
\[ \hat{\theta}_2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2 \] is an unbiased estimator for \( \theta_2 \).

\( \hat{\theta}_2 \) and \( \hat{\theta}_1 \) differ by a factor \( \frac{n-1}{n} \).

Intuition why \( \hat{\theta}_2 \) might be biased.

\[ n = 2. \]

suggests that \( \hat{\theta}_2 \) is an underestimate of \( \theta_2 \).
Confidence intervals

Problem: MLE $\hat{\theta}$ of $\theta$ is wrong with probability $\frac{1}{2}$.

Could we find a $\Delta$ such that $\theta \in [\hat{\theta} - \Delta, \hat{\theta} + \Delta]$ with probability $95\%$, say.

This is called the $95\%$ confidence interval.

Ex: MLE $\hat{\theta}$ of mean $\mu$ in $N(\mu, \sigma^2)$.

Ind. samples $x_1, x_2, \ldots, x_n \sim N(\mu, \sigma^2)$.

$\text{Var}(\hat{\theta}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} x_i \right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^{n} x_i \right)$

$= \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(x_i) = \frac{1}{n^2} \cdot n \sigma^2 = \frac{\sigma^2}{n}$.

$\hat{\theta} \sim N(\mu, \sigma^2/n)$

$\frac{\hat{\theta} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

$P(-z < \frac{\hat{\theta} - \mu}{\sigma/\sqrt{n}} < +z) = \Phi(z) - \Phi(-z) = 2 \Phi(z) - 1$

$P(-z < \frac{\mu - \hat{\theta}}{\sigma/\sqrt{n}} < +z) = 2 \Phi(z) - 1$

$P(\hat{\theta} - \frac{3.845}{\sqrt{n}} < \mu < \hat{\theta} + \frac{3.845}{\sqrt{n}}) = 2 \Phi(z) - 1 = 0.95$

$2 \Phi(z) = 1.95$

$\Phi(z) = 0.975$

$z \approx 2.58$ 1.96

$\Delta = \frac{3.845}{\sqrt{n}}$

$P(\mu \in [\hat{\theta} - \Delta, \hat{\theta} + \Delta]) \approx 0.95$