Sample variance

$N(\mu, \sigma^2) = N(\theta_1, \theta_2)$

$x_1, x_2, \ldots, x_n$ are independent r.v.'s for $N(\theta_1, \theta_2)$.

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \hat{\theta}_2 = \underline{\frac{1}{n}}\sum_{i=1}^{n}(x_i - \hat{\theta}_1)^2$$

$\mathrm{Var}(Y) = E[(Y-\mu)^2]$, where $\mu = E[Y]$.

$\Omega = \{x_1, x_2, \ldots, x_n\}$, uniform

$\hat{\theta}_2$ is the variance in this case.

Sample variance

Bias.

__Defn:__ An estimator $\hat{\theta}$ of θ is <u>unbiased</u> iff

$$E[\hat{\theta}] = \theta.$$

Back to the MLEs $\hat{\theta}_1$ and $\hat{\theta}_2$ for $N(\theta_1, \theta_2)$.

$$E[\hat{\theta}_1] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n}\sum_{i=1}^{n}\theta_1$$

$$= \frac{1}{n}\cdot n\theta_1 = \theta_1, \text{ so } \hat{\theta}_1 \text{ is unbiased.}$$

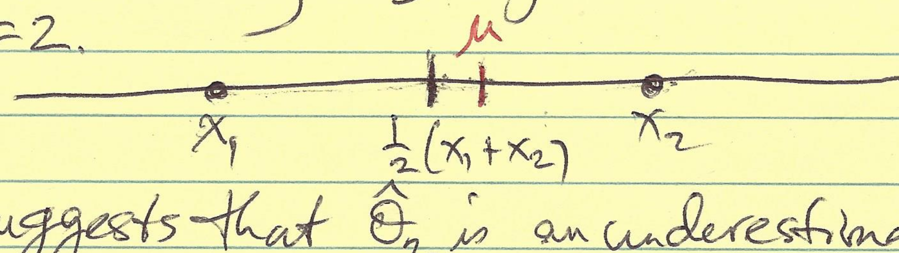
But $\hat{\theta}_2$ is not an unbiased estimator for θ_2.

$\hat{\theta}_2' = \underline{\frac{1}{n-1}}\sum_{i=1}^{n}(x_i - \hat{\theta}_1)^2$ is an unbiased estimator for θ_2.

$\hat{\theta}_2$ and $\hat{\theta}_2'$ 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 θ_2

Confidence intervals

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

Could we find a Δ such that

$\theta \in [\hat{\theta} - \Delta, \hat{\theta} + \Delta]$ with probability 95%, say.

This is called the 95% confidence interval.

Ex: MLE $\hat{\theta}_1$ of mean μ in $N(\mu, \sigma^2)$.

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

$$\mathrm{Var}(\hat{\theta}_1) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} x_i\right)$$

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

$$\hat{\theta}_1 \sim N(\mu, \sigma^2/n)$$

$$\frac{\hat{\theta}_1 - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$P\left(-z < \frac{\hat{\theta}_1 - \mu}{\sigma/\sqrt{n}} < +z\right) = \Phi(z) - \Phi(-z) = 2\Phi(z) - 1$$

$$P\left(-z < \frac{\mu - \hat{\theta}_1}{\sigma/\sqrt{n}} < +z\right) = 2\Phi(z) - 1$$

$$P\left(\hat{\theta}_1 - \cancel{z/\sqrt{n}}\ \frac{z\sigma}{\sqrt{n}} < \mu < \hat{\theta}_1 + \frac{z\sigma}{\sqrt{n}}\right) = 2\Phi(z) - 1 = 0.95$$

$$2\Phi(z) = 1.95$$

$$\Phi(z) = 0.975$$

$$z \approx \cancel{2.26}\ 1.96$$

$$\Delta = \frac{\overset{1.96}{\cancel{2.26}}\,\sigma}{\sqrt{n}}$$

$$P(\mu \in [\hat{\theta}_1 - \Delta, \hat{\theta}_1 + \Delta]) \approx 0.95$$