Problem: Let $\{x_1 \dots x_N\}$ be iidrv with $x_i \sim N(m, \sigma^2)$ and m, σ^2 *unknown*.
Want: To compute $\hat{m}_{MLE}$ and $\hat{\sigma}^2_{MLE}$ based on observations $\{X_1 \dots X_N\}$.
Solution: $f_{x_1 \dots x_N}(X_1 \dots X_N) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(X_i - m)^2/\sigma^2}$ since x_i indpt rvs.

Set: $0 = \frac{\partial}{\partial m} \log f_{x_1 \dots x_N} = \frac{\partial}{\partial m}[-\frac{N}{2}\log(2\pi) - \frac{N}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^N (X_i - m)^2/\sigma^2]$
$= \frac{1}{\sigma^2}\sum_{i=1}^N (X_i - m) = 0 \rightarrow \hat{m}_{MLE} = \frac{1}{N}\sum_{i=1}^N X_i = \textit{sample mean}.$

Set: $0 = \frac{\partial}{\partial \sigma^2} \log f_{x_1 \dots x_N} = \frac{\partial}{\partial \sigma^2}[-\frac{N}{2}\log(2\pi) - \frac{N}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^N (X_i - m)^2/\sigma^2]$
$= -\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{2}\sum_{i=1}^N (X_i - m)^2/(\sigma^2)^2 = 0 \rightarrow \hat{\sigma}^2_{MLE} = \frac{1}{N}\sum_{i=1}^N (X_i - m)^2.$
Replace m in $\hat{\sigma}^2_{MLE}$ with $\hat{m}_{MLE} \rightarrow \hat{\sigma}^2_{MLE} = \textit{sample variance}.$
Note: $\hat{\sigma}_{MLE} = \sqrt{\hat{\sigma}^2_{MLE}}$: MLE *commutes* with nonlinear functions $g(a)$.
Why? $\underset{A}{\text{argmax}} f_{r|a}(R|A) = \underset{g(A)}{\text{argmax}} f_{r|g(a)}(R|g(A))$. No Jacobian for $a \rightarrow g(a)$.

Q: What are some desirable properties for estimators to have?
DEF: *Unbiased* estimator has $E[\hat{a}(x_1 \dots x_N)] = A$ (x_i now treated as rvs).
DEF: *Asymptotically unbiased* estimator has $\lim_{N\rightarrow\infty} E[\hat{a}(x_1 \dots x_N)] = A$.

- Sample mean is unbiased: $E[\hat{m}] = E[\frac{1}{N}\sum_{i=1}^N x_i] = \frac{1}{N}\sum_{i=1}^N E[x_i] = m.$
- Sample variance is *biased*: $E[\hat{\sigma}^2] = E[\frac{1}{N}\sum_{i=1}^N (X_i - \hat{m})^2] = \frac{N-1}{N}\sigma^2.$
- Sample variance is *asymp.* unbiased: $\lim_{N\rightarrow\infty} E[\hat{\sigma}^2] = \lim_{N\rightarrow\infty} \frac{N-1}{N}\sigma^2 = \sigma^2.$

Note: $\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^N (X_i - \frac{1}{N}\sum_{j=1}^N X_j)^2$ is an unbiased estimator of σ^2.
Note: If we *know* mean m, then the sample variance is *unbiased* estimator.
Note: Algebra for sample variance biased and consistent is on pp. 274-5.

DEF: Seq. of rvs $\{a_1, a_2 \dots\} \rightarrow a$ *in probability* if $\lim_{N\rightarrow\infty} Pr[|a_N - a| > \epsilon] = 0.$
DEF: *Consistent* estimator has $\lim_{N\rightarrow\infty} \hat{a}(x_1 \dots x_N) = a$ *in probability*.
Means: *More data helps:* The distribution of $\hat{a}$ becomes tighter around a.

- Sample mean is consistent: Use the Chebyschev inequality:
$Pr[|\hat{m} - m| > \epsilon] = Pr[|\hat{m} - E[\hat{m}]| > \epsilon] \leq \frac{\sigma^2_{\hat{m}}}{\epsilon^2} = \frac{\sigma^2}{N\epsilon^2} \rightarrow 0$ as $N \rightarrow \infty$.

Using: $\hat{m}$ unbiased and $\sigma^2_{\hat{m}} = \frac{1}{N^2}\sum_{i=1}^N \sigma^2 = \frac{\sigma^2}{N}$. We have just proved the:
Thm: *Weak Law of Large Numbers:* Let $\{x_i\}$ be iidrv with $E[x_i], \sigma_i^2 < \infty$.
$\rightarrow$ The sample mean is a *consistent* estimator of the expectation $E[x_i]$.
Means: Mean $\hat{m}$ of *data* approaches mean $m = E[x_i]$ of *random variables.*