Recitation: inference of Gaussian distribution

Recitation problem statement

Consider a sample of n i.i.d. Gaussian random variables $X_1, \ldots, X_n$, with unknown parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.

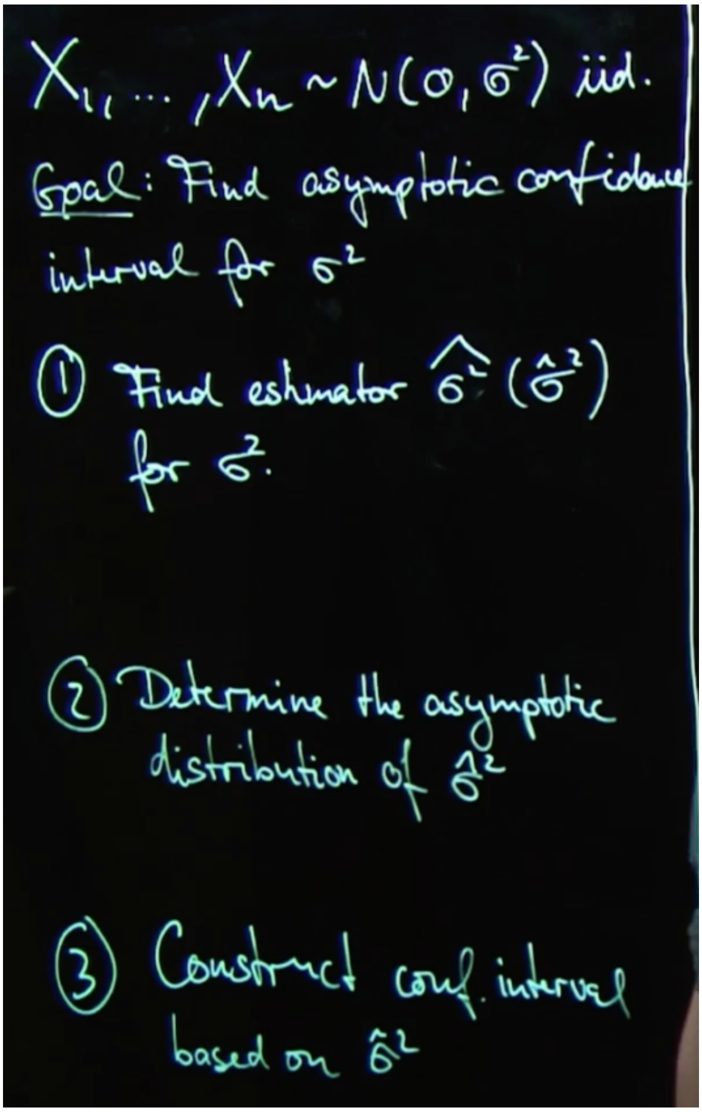
1. Let $\hat{\sigma}^2$ be the sample variance of $X_1, \ldots, X_n$. Recall an expression of $\hat{\sigma}^2$.

2. Prove that $\hat{\sigma}^2$ is a consistent estimator of σ^2.

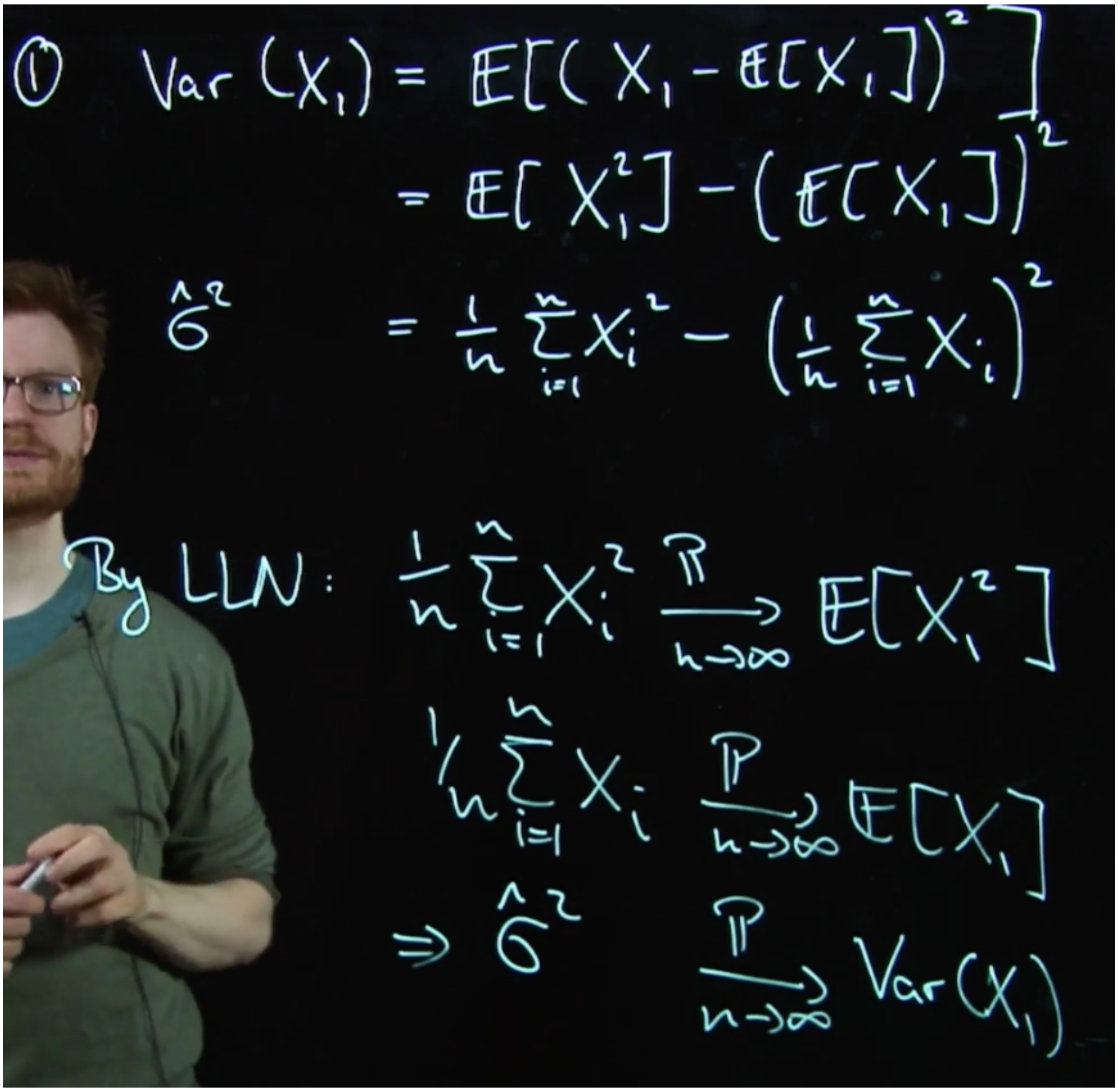
3. Using the central limit theorem, prove that $\left(\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{i=1}^{n} X_i^2\right)'$ is asymptotically normal.

4. Conclude that $\hat{\sigma}^2$ is asymptotically normal and compute its asymptotic variance.

5. Using the previous questions, find a confidence interval for σ^2 with asymptotic level 95% (use two methods).

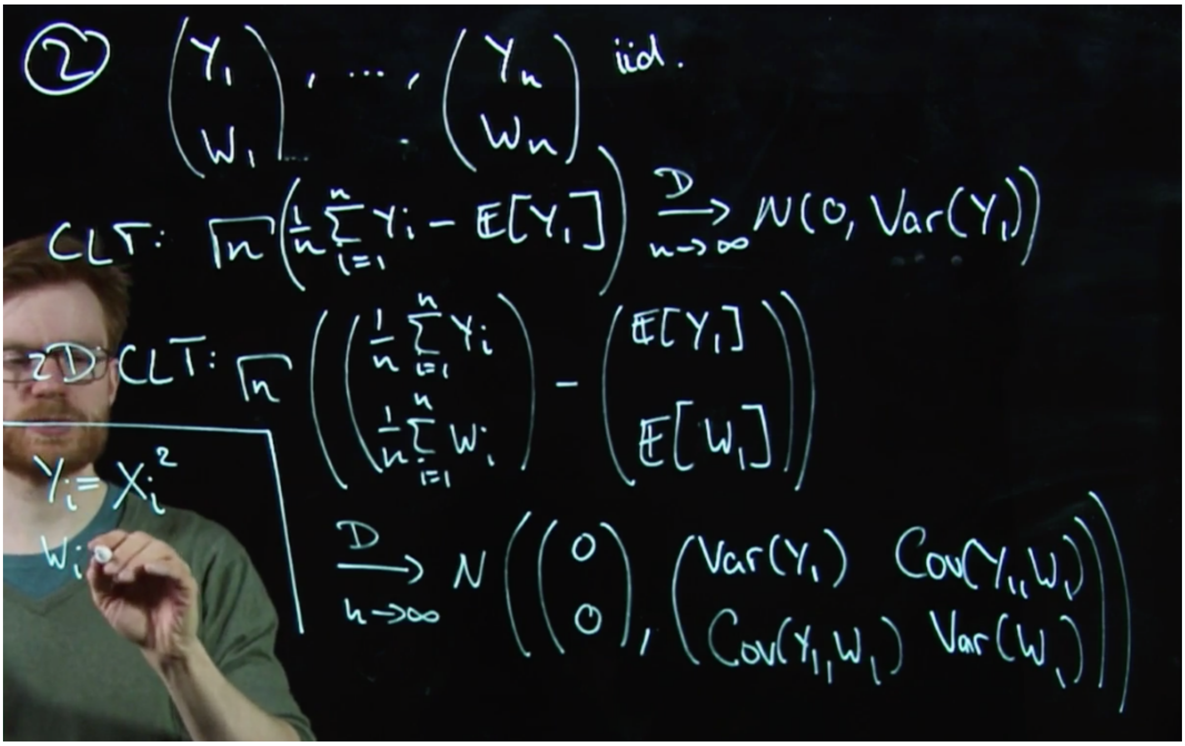


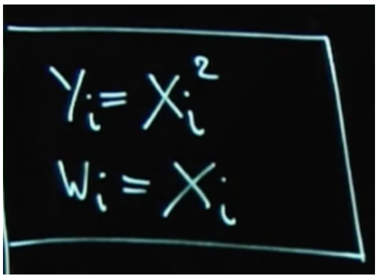
Getting a consistent estimator of the variance

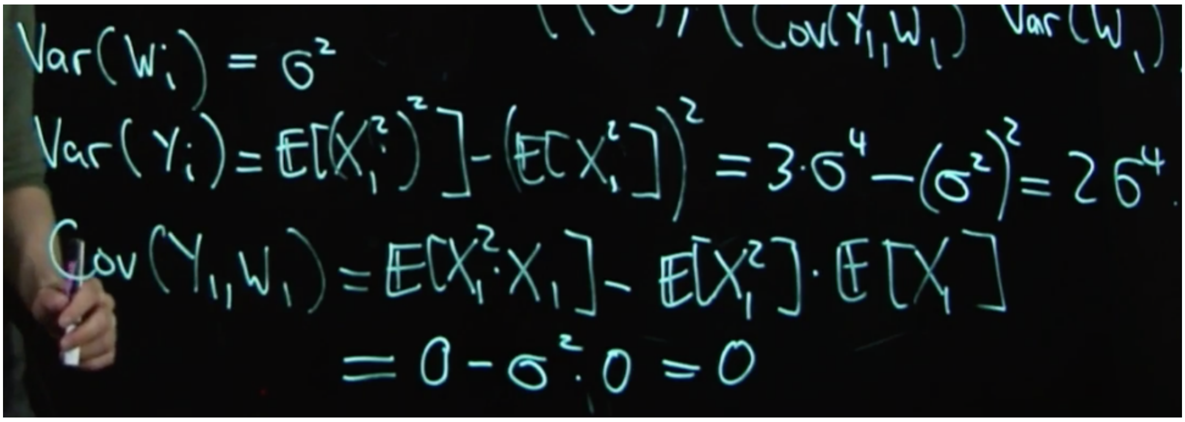


- this estimator should go to 0 since the mean of the distribution is 0
-

Applying the multivariate central limit theorem







- easily goes to 0 here because we are dealing with a centric distribution around 0

Applying the multivariate delta method

$$② \text{ Delta Method: } \sqrt{n}\big(\underset{\mathbb{R}^d}{T_n} - \theta\big) \xrightarrow[n\to\infty]{D} N(0, \underset{\mathbb{R}^{d\times d}}{\Sigma})$$

$$g: \mathbb{R}^d \to \mathbb{R}, \text{ continuously differentiable at } \theta$$

$$\Rightarrow \sqrt{n}\big(g(T_n) - g(\theta)\big) \longrightarrow N\big(0, \nabla g(\theta)^T \Sigma \nabla g(\theta)\big)$$

$$Y_i = X_i^2, \quad W_i = X_i, \quad T_n = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} X_i^2 \\ X_i \end{pmatrix}, \quad \hat{\sigma}^2 = g(T_n),$$

$$g(y,w) = y - w^2, \quad \nabla g(y,w) = \begin{pmatrix} 1 \\ -2w \end{pmatrix}, \quad \theta = \begin{pmatrix} E[Y_1] \\ E[W_1] \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \end{pmatrix}$$

$$\sqrt{n}\cdot\big(\hat{\sigma}^2 - \sigma^2\big) \xrightarrow[n\to\infty]{D} N\left(0, \begin{pmatrix} 1 \\ 0 \end{pmatrix}^T \begin{pmatrix} 2\sigma^4 & 0 \\ 0 & \sigma^2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right)$$

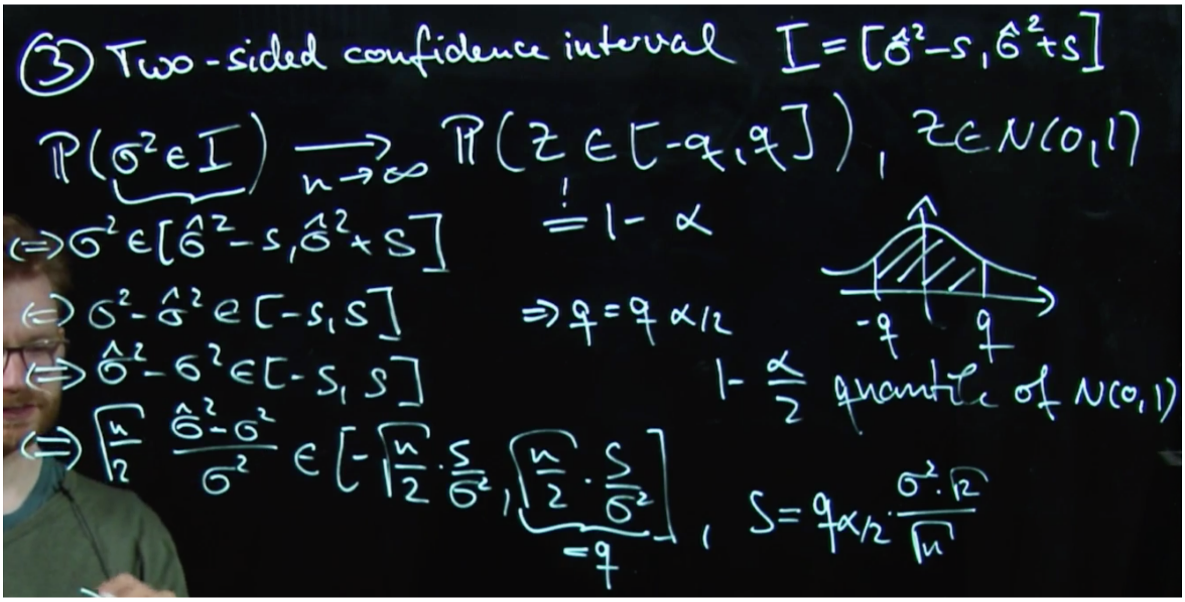
$$= N(0, 2\sigma^4)$$

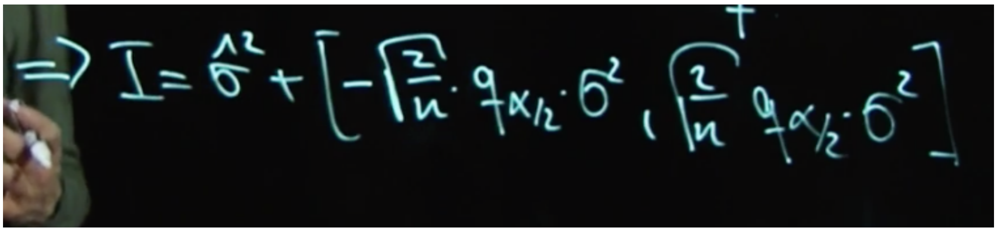
- delta method gives us the asymptotic distribution, because we know that it is continuously differentiable
- given a sequence of random variables T_n, such that they are centred by theta and re-scaled by sqrt(n) - if it goes to a normal distribution with covariance matrix, and ew have a continuously differentiable function g => (then) we know that the limiting distribution is as above (3rd line), and the covariance matrix is transformed by the gradient
- intuitively this means how sensitive is g to each of the variables
- delta g is the partial derivatives with respect to y and w
- inserting covariance matrix, terms were calculated in previous part
-

Confidence interval

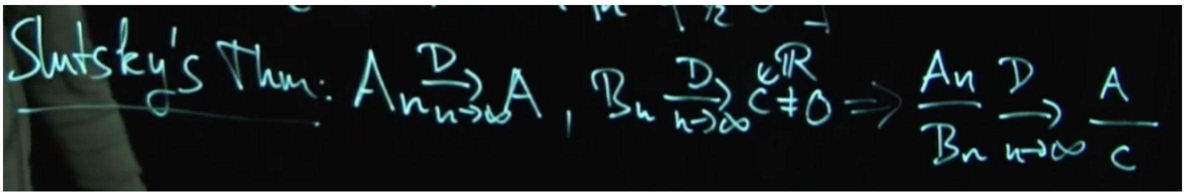
$$\sqrt{n}\,\frac{\hat{\sigma}^2 - \sigma^2}{\sqrt{2}\,\sigma^2} \xrightarrow[n\to\infty]{D} N(0,1)$$

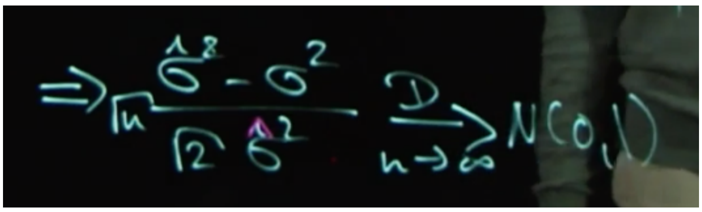
using this formula





- this confidence interval isn't great because it's dependent on the parameter we are trying to estimate (sigma^2)
- we are interested in an asymptotic value so we can instead use sigma_hat^2
- this is done using Slutsky's theorem





this allows us to turn all sigma^2 into sigma_hat^2