Definition. ***Confidence Interval and Confidence Level*** *Let* $X_1, \cdots, X_n \sim f_\theta$. *A* $100(1-\alpha)\%$ *confidence interval for* θ *is a pair of statistics* $L = L(X_1, \cdots, X_n)$ *and* $U = U(X_1, \cdots, X_n)$ *such that*

$$P(L \leq \theta \leq U) = 1 - \alpha$$

We call $100(1-\alpha)\%$ *the confidence level*

Remark. L and U are random variables while the population parameter θ is fixed and unknown. Therefore in construction of confidence interval, it is the interval itself that is random. It is incorrect to claim that θ has $100(1-\theta)\%$ chance of lying in the confidence interval. Or we expect $(1-\alpha)\%$ of the times that the interval covers the true population parameter. However, it is correct to say that if we had infinitely many random samples and calculated the confidence limits for each, $100(1-\alpha)\%$ of the resultant intervals would include the true parameter

Definition. ***The Pivotal Method*** *A pivotal quantity (or simply pivot) is a function* $g(X_1, \cdots, X_n; \theta)$ *of the data and parameter of interest, whose distribution does not depend on any unknown parameter*

1. *Find a pivot* $g(X_1, \cdots, X_n; \theta)$ *and identify its distribution*

2. *Find a and b such that* $P(a \leq g(X_1, \cdots, X_n) \leq b) = 1 - \alpha$ *using the distributinon*

3. *Find L and U such that* $P(L \leq \theta \leq U) = 1 - \alpha$

Definition. ***The large number theory method*** *When it comes difficult to find pivot and its accopaning distribution, we invoke large number theory, namly* $\hat{\theta}_{MLE} \sim AN(\theta, \mathcal{I}^{-1}(\hat{\theta}_{MLE}))$ *We can construct a* $(1-\alpha)\%$ *confidence interval of the form*

$$\hat{\theta}_{MLE} \pm \frac{z_{1-\alpha/2}}{\sqrt{\mathcal{I}^{-1}(\hat{\theta}_{MLE})}}$$

Remark. Here is some example $(1-\alpha)\%$ confidence interval calculated using the large number theory method

1. **Exponential** $\hat{\lambda}_{MLE} = 1/\overline{X}$ and $\mathcal{I}(\lambda) = n/\lambda^2$ so then

$$\frac{1}{\overline{X}} \pm \frac{z_{1-\alpha/2}}{\overline{X}\sqrt{n}}$$

2. **Poisson** $\hat{\lambda}_{MLE} = \overline{X}$ and $\mathcal{I}^{-1}(\lambda) = n/\lambda$ so then we ahve

$$\overline{X} \pm z_{1-\alpha/2}\sqrt{\frac{\overline{X}}{n}}$$

Definition. ***Standar Error*** *The standard deviation of the sampling distribution is called standard error of θ*

$$\sigma_{\hat{\theta}} = \sqrt{\frac{\theta}{n}}$$

We can derive an approximation by substituting θ with $\hat{\theta}$

$$s_{\hat{\theta}} = \sqrt{\frac{\hat{\theta}}{n}}$$

Definition. *The **Squared error loss** is given by*

$$l(\hat{\theta}, \theta) = (\theta - \hat{\theta})^2$$

which inflicts harsh penulties on large deviatios from true parameter value and forgiving for small deviations.

Definition. *The **Mean Squared Error** of an estimator $\hat{\theta}$ of a parameter θ is*

$$MSE(\hat{\theta}, \theta) = \mathbb{E}[(\hat{\theta} - \theta)^2]$$

which is a measure of estimation accuracy

Theorem. ***The Bias-Variance Decomposition*** *Let $\hat{\theta}$ be an estimator of a parameter θ. The bias of $\hat{\theta}$ is given by*

$$b(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}] - \theta$$

Then

$$MSE(\hat{\theta}, \theta) = b^2(\hat{\theta}, \theta) + Var[\hat{\theta}]$$

Definition. ***Comparison of estimator of normal variance*** *For $X_1, \cdots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$. Compure following estimators of σ^2*

1. $S^2 \dfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ *(sample variance)*

2. $\hat{\sigma}^2 = \dfrac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$ *(MME and MLE)*

Definition. ***Unbiased Estimator*** *We say that $\hat{\theta}$ is an unbiased estimator of θ if*

$$\mathbb{E}[\hat{\theta}] = \theta \quad or \quad b(\hat{\theta}, \theta) = 0$$

Remark.

1. $\overline{X}$ is an unbiased estimator of $\mu = \mathbb{E}[X]$ by law of large numbers.

2. S^2 is an unbiased estimator of $\sigma^2 = Var[X]$ (using χ_n^2)

3. Unbiased estimator has property $MSE(\hat{\theta}, \theta) = Var[\hat{\theta}]$

Theorem. ***Cramer-Rao Lower Bound*** *Let $X_1, \cdots, X_n \sim f_\theta$ and let $\hat{\theta}$ be an unbiased estimator of θ. Under some regularity conditions,*

$$Var[\hat{\theta}] \geq \mathcal{I}^{-1}(\theta)$$

where $\mathcal{I}(\theta)$ is the Fisher Information. Intuitively, the variance of any unbiased estimator is at least as high as the inverse of the Fisher information.

Remark. This theorem provides that the maximum likelihood estimator (MLE) is asymptotically optimal, as in such condition $\hat{\theta}_{MLE}$ is unbiased and achieves the lower bound. This means that no consistent estimator has lower asymptotic mean squared error than the MLE (or other estimators attaining this bound). For a finite sample size however, the MLE estimate may not be efficient.

Definition. ***Efficiency***

1. *We say that an **unbiased estimator** $\hat{\theta}$ of a parameter θ is **finite sample efficient** (or simply efficient) if*
$$Var[\hat{\theta}] = \mathcal{I}^{-1}(\theta)$$

   In other word, an unbiased estimator which achieves the Cremer-Rao lower bound is said to be efficient. This implies that such a solution achieves the lowest possible MSE among all unbiased methods

2. *We say that $\hat{\theta}$ is **asymptotically efficient** if*

$$\lim_{n \to \infty} \frac{Var[\hat{\theta}]}{\mathcal{I}^{-1}(\theta)} = 1$$

3. *The **relative efficiency** of an unbiased estimator $\hat{\theta}_1$ of θ with respect to another unbiased estiamtor $\hat{\theta}_2$ is*

$$eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{Var[\hat{\theta}_2]}{Var[\hat{\theta}_1]}$$

   If efficiency is smaller than 1, $\hat{\theta}_1$ has a larger variance. This comparison is useful when two estimators have different bias.

Remark.
Some examples of unbiased estimators and their efficiency

1. $\overline{X}$ of normal distribution is efficient

2. S^2 is asymptotically efficient

3. MLE is asymptotically unbiased and asymptotically efficient. Although MLE is not necessarily unbiased and efficient for finite sample size.