Chapter 8: SUMMARY

Two principal methods of point estimation: (1) Method of Moments and (2) Method of Maximum Likelihood.

Method of Moments:
The kth moment of a random variable X is defined by $\mu_k = E(X^k)$. If $X_1, \cdots, X_n$ are the sample observations from the distribution of the random variable X, then the kth sample moment is defined by $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^n X_i^k$. For example μ_1 is the population mean, and $\bar{X}$ is the sample mean.
The method of moments estimates parameters by finding expressions for them in terms of the lowest possible order moments, and then equating these moments to the correponding sample moments, for example equate μ_1 to $\bar{X}$, equate μ_2 to $\frac{1}{n}\sum_{i=1}^n X_i^2$ etc.

Maximum Likelihood:
The likelihood function is the joint pdf or pf of the sample observations. The maximum likelihood estimate (MLE) of a parameter θ (real or vector valued) is obtained by maximizing the likelihood with respect to θ. For example, if $X_1, \cdots, X_n$ are iid sample observations from a common pdf or pf $f(x|\theta)$, then the likelihood function is given by $L(\theta) = \prod_{i=1}^n f(x_i|\theta)$. Instead of maximizing $L(\theta)$ with respect to θ, often we maximize $l(\theta) = \sum_{i=1}^n \log f(x_i|\theta)$ with respect to θ.
An Important Result: If $\hat{\theta}_{\text{MLE}}$ is the MLE of θ, then $\hat{\theta}_{\text{MLE}}$ is asymptotically distributed as $\text{N}(\theta, \frac{1}{nI(\theta)})$, where $I(\theta) = E[\{\frac{\partial \log f(X|\theta)}{\partial\theta}\}^2] = E[-\frac{\partial^2 \log f(X|\theta)}{\partial\theta^2}]$.

An estimator $\hat{\theta}$ of θ is said to be unbiased if $E(\hat{\theta}) = \theta$.
An estimator $\hat{\theta}$ of θ based on a sample of size n is said to be consistent if $\hat{\theta}$ converges in probability to θ, i.e. for every $\epsilon > 0$, $\text{P}(|\hat{\theta} - \theta| > \epsilon) \to 0$ as $n \to \infty$.
A sufficient condition to prove consistency of an estimator $\hat{\theta}$ of θ is to show that $E(\hat{\theta}) \to \theta$ as $n \to \infty$ and $V(\hat{\theta}) \to 0$ as $n \to \infty$.

Standard error of an estimator $\hat{\theta}$ of θ, written as $\sigma_{\hat{\theta}} = V^{1/2}(\hat{\theta})$. The asymptotic standard error of $\hat{\theta}_{\text{MLE}}$ based on iid samples of size n is $(\frac{1}{nI(\theta)})^{1/2}$.

Often $\bar{X}$ based on iid samples of size n turns out to be the natural estimator of θ. The standard error of $\bar{X}$ denoted by $\sigma_{\bar{X}} = \sigma/n^{1/2}$, where σ is the population standard deviation. The calculation of σ varies from one distribution to another.

A confidence interval for a population parameter θ is a random interval calculated from the sample which contains θ with some specified probability.
If $X_1, \cdots, X_n$ are iid samples from a $\text{N}(\mu, \sigma^2)$ distribution, then a $100(1-\alpha)\%$ two-sided confidence interval for μ is $\bar{X} \pm t_{n-1;\alpha/2} S/n^{1/2}$, where $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ and $t_{n-1;\alpha/2}$ is the upper $100\alpha/2\%$ point of Student's t-distribution with $n-1$ degrees of freedom. Similarly, a chisquare distribution can be used to construct a confidence interval for σ^2.

A $100(1-\alpha)\%$ large sample confidence interval for θ based on $\hat{\theta}_{\text{MLE}}$ is $\hat{\theta}_{\text{MLE}} \pm z_{\alpha/2}(\frac{1}{nI(\hat{\theta})})^{1/2}$.

If $\bar{X}$ is an estimator for θ, then the large sample $100(1-\alpha)\%$ confidence interval for θ is $\bar{X} \pm z_{\alpha/2}\hat{\sigma}_{\bar{X}}$ or $\bar{X} \pm z_{\alpha/2}\hat{\sigma}/n^{1/2}$.