2.1 Sampling Distribution of The Sample Mean

Steven Cheung
2 February 2022

If a family of probability models is indexed by two or more unknown parameters, $\theta_1, \ldots, \theta_k$. finding maximum likelihood estimates require the solution of k simultaneous equaions

e.g for k=2, $\frac{\partial L(\theta_1,\theta_2)}{\partial\theta_1} = 0$ and $\frac{\partial L(\theta_1,\theta_2)}{\partial\theta_2} = 0$

Theorem (Central Limit Theory)

Let $X_1, \ldots, X_n$ be a random sample n from a population with mean μ and variance σ^2. Then $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \frac{\sigma^2}{n}$

Proof of the Central Limit Theorem)

$E(\bar{x}) = E(\frac{\sum_{i=1}^{n} x_i}{n}) = \sum_{i=1}^{n} E(x_i) = \sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu$
$Var(\bar{x}) = Var(\frac{\sum_{i=1}^{n} x_i}{n}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$

$\mu_{\bar{x}} = E(\bar{X}) = \mu$
$var(\bar{x}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$
$\sigma_{\bar{X}}$ - standard error

From Chebyshev's Inequality

$$P(|\bar{X} - \mu_{\bar{X}}| < k\sigma_{\bar{X}}) \geq 1 - \frac{1}{k^2}$$

let $\epsilon = \frac{(k\sigma)}{\sqrt{n}} > 0$

$$P(|\bar{X} - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2}$$

Let $\{c_1, \ldots, c_n\}$ be a finite population. Then the population mean $\mu = (\frac{1}{N})\sum_{i=1}^{n} c_i$, and the population variance $\sigma^2 = (\frac{1}{N})\sum_{i=1}^{n}(c_i - \mu)^2$

Theorem:

Let $X_1, \ldots, X_n$ be a random sample of size n (chosen without replacement) from a finite population $\{c_1, \ldots, c_n\}$, then

$$E(\bar{X}) = \mu$$
$$var(\bar{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

The factor $\frac{(N-n)}{(N-1)}$ is often called the finite population correction factor