Assignment 10

Hema Sri Cheekatla, CS21BTECH11013

June 17, 2022

Outline

Questioin

On observing 100 samples of x, find the 0.95 confidence interval of the median $x_{0.5}$ of x

Concept

u Percentile:
The u percentile of a random variable x is by definition a number x_u such that $F(x_u) = u$. Thus x_u is the inverse function $F^{-1}(u)$ of the distribution F(x) of x.

Let y_k be the k^{th} number assuming that all the numbers are arranged in ascending order

$y_k < x_u < x_{k+r-1}$, iff atleast k and atmost $k + r - 1$ of the samples x_i are less than x_u

Concept

The event $\{y_k < x_u < y_{k+r-1}\}$ occurs iff number of success of event $\{x \leq x_u\}$ in n repetitions of the experiment is atleast k and atmost $k+r-1$.

And since $\Pr(x \leq x_u) = u$ which means success probability $p = u$, we can obtain the following

$$\Pr(y_k < x_u < y_{k+r-1}) = \sum_{m=k}^{k+r-1} \binom{n}{m} u^m (1-u)^{n-m} \tag{2.0.1}$$

Concept

On approximation of above equation and assuming that X is a Normal or Gaussian Random Variable and n is large, for specific value of u, we obtain the following

$$\Pr\left(y_k < x_u < y_{k+r-1}\right) = G\left(\frac{k+r-0.5-nu}{\sqrt{nu(1-u)}}\right) - G\left(\frac{k-0.5-nu}{\sqrt{nu(1-u)}}\right) = \gamma \tag{2.0.2}$$

Where, G is a Gaussian distribution function

$$G(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy \tag{2.0.3}$$

Concept

From these equations, for a specific γ, r is minimum if nu is near the center of the interval $(k, k+r)$. This yeilds,

$$k \simeq nu - z_{1-\delta/2}\sqrt{nu(1-u)} \tag{2.0.4}$$

$$k + r \simeq nu - z_{1-\delta/2}\sqrt{nu(1-u)} \tag{2.0.5}$$

where z_u is the standard normal percentile with ,

$$u = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_u} e^{-z^2/2}\, dz \tag{2.0.6}$$

Solution:

We have ,

$$n = 100 \tag{3.0.1}$$
$$u = 0.5 \tag{3.0.2}$$
$$\gamma = 1 - \delta = 0.95 \tag{3.0.3}$$
$$\implies \delta/2 = 0.025 \tag{3.0.4}$$
$$\implies z_{1-\delta/2} \simeq 2 \tag{3.0.5}$$

Solution

Hence k and k+r are given by,

$$k \simeq (100)(0.5) - 2(\sqrt{(100)(0.5)(1-0.5)}) \tag{3.0.6}$$

$$k = 40 \tag{3.0.7}$$

$$\text{Similarly, } k + r \simeq (100)(0.5) + 2(\sqrt{(100)(0.5)(1-0.5)}) \tag{3.0.8}$$

$$\implies k + r = 60 \tag{3.0.9}$$

Therefore the median of $\mathbf{x}$ is inbetween y_{40} and y_{60}

That is the median is inbetween 40th and 60th number.