1/27/2025

Monday, January 27, 2025 12:18 PM

If $\vec{y} \sim MVN(\vec{\mu}, \Sigma)$, then

$$\vec{y} = \vec{\mu} + \underset{\sim}{R}^T \vec{z}, \quad \vec{z} \sim N(0, \underset{\sim}{I})$$

- For univariate: Let $X_i \sim N(\mu, \sigma^2)$

$$\Rightarrow \quad X_i = \mu + \sigma \cdot Z_i, \quad Z_i \sim N(0, 1)$$

Estimators

- Let $\vec{\Theta} \in \mathbb{R}^p$ be a parameter (vector) of a model. We want an estimator that is associated with the distribution we are considering.
  - The estimator is a R.V., but the estimate is a realization from the estimator.

Ex) Suppose you have $x_1, x_2, \ldots, x_N$.

- We can define the estimators

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i, \quad S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$$

- Given realizations, x_i, $i = 1, \ldots, N$

$$\Rightarrow \quad \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i, \quad s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

- If we assume $X_i \sim N(\mu, \sigma^2)$, then the <u>Sampling Distribution</u> of $\overline{X}$ + S^2 are

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{N}\right), \quad S^2 \sim \frac{\sigma^2}{N-1} \chi^2(N-1)$$

- Note: what happens as $N \to \infty$
  - $\overline{X} \to \delta(x-\mu)$
  - $S^2 \to \delta(x-\sigma^2)$

<u>Confidence Intervals</u>

- Based on realizations of X, $\vec{x} = [x_1, \ldots, x_N]$, we want a interval estimator, $L(\vec{x})$ + $U(\vec{x})$, where for any parameter $\vec{\theta}$ based on $\vec{x}$

$$L(\vec{x}) \leq \theta \leq U(\vec{x})$$

- If we have $L(\vec{x})$ + $U(\vec{x})$, then a $1-\alpha$ Confidence interval is defined as

$$P[L(\vec{x}) \leq \theta \leq U(\vec{x})] = 1-\alpha$$

- Note: θ is unknown, but fixed.

<u>Inference</u>

- Let y_i denote observations from some process, + let $f(\underline{x}_i; \vec{\theta})$, where $\underline{x}_i$ is independent variable + $\vec{\theta}$ parameters.
  - We assume f approximates the true process. Then

- We assume f approximates the true process. Then

$$Y_i = f(x_i ; \vec{\theta}) + \varepsilon_i , \quad i = 1, \ldots, N$$

where ε_i accounts for difference between Y_i + f.

- Y_i: random observations (real. y_i)
- ε_i: random measurement errors (real ε_i)
- $\vec{\theta}$: Parameters
  - ☐ Frequentist: fixed, unknown
  - ☐ Bayesian: random variables