LECTURE 9

SIMULATION AND MODELING

CSE 411

Md. Tanvir Al Amin, Lecturer, Dept. of CSE, BUET

Estimation of mean, var, cor

- Suppose X1, X2, ….. Xn are IID random variables
  - IID = Independent and Identically Distributed
  - With finite population mean = μ
  - Population Variance = σ^2
- Given an observation of X1 … X5
  - Can you find the mean of the underlying distribution ??
  - Or the variance ??

Estimation of mean, var, cor

- Suppose the random variables are taken from a normal distribution
  - With Mean = 2, Variance = 1
  - These are **Population Mean** and **Population Variance**

Estimation of mean, var, cor

- Suppose we make observations of 5 random variables X1, X2, X3, X4, X5
  - Known that X1 … X5 are Identically distributed
  - They are independent
  - What is the mean and variance of their underlying distribution ??
  - Can we estimate it from the observations found ?

Estimation of mean, var, cor

- Some instance

X1	X2	X3	X4	X5	Sample Mean
0.6638	2.7143	3.6236	1.3082	2.8580	2.2358
3.2540	0.4063	0.5590	2.5711	1.6001	1.6781
2.6900	2.8156	2.7119	3.2902	2.6686	2.8353

2.2358, 1.6781, 2.8353 …. But none of these equals 2 !!!

How can we find the population mean ??

Estimation of mean, var, cor

- We can **ESTIMATE** the **population mean** by the **sample mean**
- The sample mean :

$$\overline{X(n)} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- is an unbiased estimator of μ i.e $E\left[\overline{X(n)}\right] = \mu$
  - Which means : If we perform a very large number of independent experiments, each resulting in an $\overline{X(n)}$, the average of the $\overline{X(n)}$ s will be μ
  - Sometimes estimators denoted as $\hat{\mu}$

Estimation of mean, var, cor

- Some instance

	X1	X2	X3	X4	X5	Sample Mean
Exp 1 :	0.6638	2.7143	3.6236	1.3082	2.8580	2.2358
Exp 2 :	3.2540	0.4063	0.5590	2.5711	1.6001	1.6781
Exp 3 :	2.6900	2.8156	2.7119	3.2902	2.6686	2.8353
Exp 4 :	2.2193	1.0781	-0.1707	1.9408	0.9894	1.2114
Exp 5 :	2.6145	2.5077	3.6924	2.5913	1.3564	2.5525
Exp 6 :	2.3803	0.9909	1.9805	1.9518	2.0000	1.8607

2.0623 !!

- **Prove that** The sample mean :

$$\overline{X(n)} = \frac{\sum_{i=1}^{n} X_i}{n}$$

  - is an unbiased estimator of μ
  - Exercise problem 4.16
  - Important

Sample Variance

- **Prove that** Sample Variance

$$S^2(n) = \frac{\sum_{i=1}^{n} [X_i - \overline{X(n)}]^2}{n-1}$$

- is an unbiased estimator of μ
- i.e $E\left[S^2(n)\right] = \sigma^2$
- Exercise problem 4.16
- Important

Estimation precision

- But how much confident we are finding our results ?
- We have no way of assessing how close $\overline{X(n)}$ is to μ
- $\overline{X(n)}$ itself is a random variable with variance

  $$Var\left[\overline{X(n)}\right]$$

- How to asses the precision of the estimation ?
  - Construct a **confidence interval**

Estimation precision

- The bigger the sample size the closer the estimate

$$
\begin{aligned}
&Var\left(\overline{X(n)}\right) \\
&= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var\left(X_i\right) \\
&= \frac{1}{n^2} n\sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

Estimating variance of sample mean

- How good is our estimation ?
  - Find the variance of sample mean
  - $E\left[S^2(n)\right] = \sigma^2$, so

$$E\left[\frac{S^2(n)}{n}\right] = \frac{\sigma^2}{n} = Var(\overline{X(n)})$$

An estimate for $Var(\overline{X(n)})$ is $\frac{S^2(n)}{n} = \frac{\sum_{i=1}^{n}\left[X_i - \overline{X(n)}\right]^2}{n(n-1)}$

Confidence Interval

- Confidence intervals are used to indicate the reliability of an estimate.
  - Perform an experiment
  - Say data : 1.2 1.5 1.68 1.89 0.95 1.49 1.58 1.55 0.50 1.09
  - Suppose they are from an unknown normal distribution of mean μ
  - Suppose the 90% confindence interval **as calculated from the observation** is **[1.10 1.58]**
  - It means : We claim with 90% confidence that μ is in [1.10 1.58]
  - Which means : **If we repeat the experiment several times and calculate confidence intervals, they will contain the true population mean 90% times**

How to construct

- Suppose X_1, X_2, ….. X_n are IID random variables
  - IID = Independent and Identically Distributed
  - With finite population mean = μ
  - Population Variance = σ^2
  - How to construct a confidence interval for μ ??

Central Limit Theorem

- **The most important result in probability theroy**
- Let $Z_n = \frac{\overline{X(n)} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$
- And Let $F_n(z)$ be the distribution function Z_n for a sample size of n; that is $F_n(z) = \Pr\{Z_n \leq z\}$
- Note that : μ and σ^2 are the mean and variance of $\overline{X(n)}$

Central Limit Theorem

- Central Limit Theorem says

$F_n(z) \to \Phi(z)$ as $n \to \infty$, where $\Phi(z)$ is the distribution function of a normal random variable with $\mu = 0$ and $\sigma^2 = 1$ i.e. standard normal random variable

$$\Phi(z) = \frac{1}{2\pi} \int_{-\infty}^{z} e^{-\frac{y^2}{2}} dy \ \text{ for } \ -\infty < z < \infty$$

Central Limit Theorem

- If n is sufficiently large, then the random variable Z_n will be approximately distributed as a standard normal random variable, **regardless of the underlying distribution of the X_i s .**

- It can also be shown that the **sample mean** is approximately distributed as normal random variable with mean μ and variance $\frac{\sigma^2}{n}$

Difficulty

- $F_n(z) \to \Phi(z)$, but the problem is

$$Z_n = \frac{\overline{X(n)} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

- We need to know $\frac{\sigma^2}{n}$ but the population variance is not known actually !!
- But sample variance $S^2(n)$ converges to σ^2 as n gets large, so
  - It can be proved that σ^2 can be replaced by $S^2(n)$ in the expression for Z_n

t-distribution

- If n is sufficiently large, the random variable t_n

$$t_n = \frac{\overline{X}(n) - \mu}{\sqrt{\frac{S^2(n)}{n}}}$$

- Is approximately distributed as a standard random variable. Then $P(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{X}(n) - \mu}{\sqrt{\frac{S^2(n)}{n}}} \leq z_{1-\frac{\alpha}{2}})$

$$= P\left(\overline{X}(n) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n)}{n}} \leq \mu \leq \overline{X}(n) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n)}{n}} \right)$$

$$\approx 1 - \alpha$$

The Confidence Interval

- $z_{1-\frac{\alpha}{2}}$ is the upper $1-\frac{\alpha}{2}$ critical point for a standard normal random variable
- Therefore if n is sufficiently large, an approximate
  - 100(1-α) percent confidence interval is given by

$$\overline{X}(n) \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n)}{n}}$$