Development statistics 06

S06 Sample Distribution

Fujikawa, Kiyoshi
Nagoya University, GSID

Distribution of Sample Mean

- Population =Normal Distribution

$$N(\mu, \sigma^2)$$

- Distribution of sample mean
- Variance becomes smaller then that of the population.

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Mean of sample mean

$$E(\overline{X}) = E(\frac{1}{n}\sum x_i) = \frac{1}{n}E(\sum x_i)$$
$$= \frac{1}{n}E(\sum \mu) = \frac{n\mu}{n} = \mu$$

Variance of sample mean

$$Var(\overline{X}) = Var(\frac{1}{n}\sum x_i)$$
$$= \frac{1}{n^2}Var(\sum x_i)$$
$$= \frac{1}{n}Var(x_i) = \frac{1}{n}\sigma^2$$

Standardization of Normal distribution

- Population =Normal Distribution

$$N(\mu,\sigma^2)$$

- Distribution of sample mean

$$\overline{X} \sim N(\mu,\frac{\sigma^2}{n}) \Rightarrow z=\frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

Sample Variance

- Maximum likelihood variance(最尤分散)
- Large sample theory
- Consistency (一致性)

$$s^2=\frac{\sum_i (x_i-\overline{x})^2}{n}$$

- Unbiased variance
- small sample theory
- Unbiased (不偏性)

$$s^2=\frac{\sum_i (x_i-\overline{x})^2}{n-1}$$

Basic knowledge

$$Var(x_i) = \sigma^2$$
$$= E(x_i - \mu)^2$$
$$= E(x_i^2) - 2\mu E(x_i) + \mu^2$$
$$= E(x_i^2) - \mu^2$$

$$E(x_i^2) = \sigma^2 + \mu^2$$

mean of sample variance

$$E\sum(x_i - \overline{X})^2$$
$$= E\sum(x_i^2) - 2E(\overline{X}\sum x_i) + nE(\overline{X}^2)$$
$$= E\sum(x_i^2) - 2nE(\overline{X}^2) + nE(\overline{X}^2)$$
$$= E\sum(x_i^2) - nE(\overline{X}^2)$$
$$= n(\sigma^2 + \mu^2) - n(\frac{1}{n}\sigma^2 + \mu^2)$$
$$= (n-1)\sigma^2$$

Unbiased variance

$$E\frac{1}{n-1}\sum(x_i - \overline{X})^2 = \sigma^2$$

Chi-square distribution

- The population is $N(0,1)$
- Take sample whose size is n

$$x_i \sim N(0,1)$$

- The squared sum of n samples follows Chi-square distribution with degree of freedom n : $\chi^2\ (n)$

$$C = \sum_{i=1}^{n} x_i^2 \sim \chi^2(n)$$

Distribution of sample variance

- The population is $N(\mu,\sigma^2)$
- Take samples whose size is n

Unbiased sample Variance

$$s^2 = \frac{\sum_i (x_i - \overline{x})^2}{n-1}$$

Chi square distribution

$$C = (n-1)\frac{s^2}{\sigma^2} \sim \chi^2(n-1)$$

T distribution

$$z \sim N(0,1) \qquad C \sim \chi^2(n)$$

$$t = \frac{Z}{\sqrt{\frac{C}{n}}} \sim t(n)$$

Practically speaking

$$z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1) \qquad C = (n-1)\frac{s^2}{\sigma^2} \sim \chi^2(n-1)$$

$$t = \frac{\overline{X} - \mu}{\sqrt{\frac{s^2}{n}}} \sim t(n-1)$$

Binomial distribution

- Probability of "success or failure"
  - 1 (success) with probability p
  - 0 (failure) with probability 1-p
- Trial times : n
- Probability of total points "x"

$$P(x) = {}_n C_x \, p^x (1-p)^{n-x}$$

Mean & variance of Binomial Distribution

- Binomial distribution
  = total of Bernoulli distribution
- Mean

$$\mu = 1 * p + 1 * p + \cdots = np$$

- Variance

$$\sigma^2 = 1^2 p + 1^2 p + \cdots = np(1-p)$$

Distribution of proportion

- Distribution of ratio

$$\frac{x}{n} = \hat{p} \sim N(p, p(1-p)/n)$$

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim N(0,1)$$

- If the sample size is large enough

$$z = \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \sim N(0,1)$$