(1) Discrete Random Variables

1. Discrete Uniform

    - Probability Mass Function

    - Mean

    - Variance

2. Binomial

    - Probability Mass Function

    - Mean

    - Variance

(2) Continuous Random Variables

Distribution function: The distribution function of a random variable X of the continuous type, defined in terms of the p.d.f. of X, is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \quad -\infty < x < \infty$$

Popular continuous distributions

1. Uniform

    - Probability Density Function

    - Distribution function

    - Mean

    - Variance

2. Normal

   - Probability density Function

   - Distribution function

   - Mean

   - Variance

<u>Theorem</u> If $X_i, i = 1, ..., n$, are independent random variables that are normally distributed with respective parameters $\mu_i, \sigma_i^2, i = 1, ..., n$, then $\sum_{i=1}^n X_i$ is normally distributed with parameters $\sum_{i=1}^n \mu_i$ and $\sum_{i=1}^n \sigma_i^2$

3. Chi-square

- Probability density Function

- Mean

- Variance

Theorem Let $W = X_1 + X_2 + \cdots + X_n$, a sum of n mutually independent and identically distributed chi-square random variables with $r_1, r_2, ..., r_n$ degrees of freedom, respectively. Then, W has a Chi-square distribution with $r_1 + r_2 + \cdots + r_n$ degrees of freedom.

Theorem Let $Z_1, Z_2, ..., Z_n$ have standard normal distributions, $N(0, 1)$. If these random variables are independent, then $W = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ has a distribution that is $\chi^2_{(n)}$.

4. Student's t distribution

Let

$$T = \frac{Z}{\sqrt{U/r}},$$

where Z is a random variable that is $N(0,1)$, U is a random variable that is $\chi^2_{(r)}$, and Z and U are independent. Then T has a t distribution with p.d.f.

$$f(t) = \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\Gamma(r/2)} \frac{1}{(1+t^2/r)^{(r+1)/2}}, \qquad -\infty < t < \infty.$$

5. F distribution

Let

$$W = \frac{U/r_1}{V/r_2},$$

where U and V are independent chi-square variables with r_1 and r_2 degrees of freedom, respectively. Then W has the F distribution with r_1 and r_2 degrees of freedom.

(3) Sampling distribution and Estimation

Independence

Definition: Let (X, Y) be a bivariate random vector with pdf or pmf $f(x, y)$, and marginal pdfs or pmfs $f_X(x)$ and $f_Y(y)$. Then X and Y are called independent random variables if, for every $x \in \mathcal{R}$ and $y \in \mathcal{R}$,

$$f(x, y) = f_X(x)\ f_Y(y).$$

Random sample

Definition: Let the random variables $X_1, X_2, ..., X_n$ have a joint density $f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$, then

$$f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = f(x_1)f(x_2)\cdots f(x_n),$$

where $f(\cdot)$ is the (common) density of each X_i. Furthermore, $X_1, X_2, ..., X_n$ is defined to be a random sample of size n from a population with density $f(\cdot)$.

Sampling Distributions

Theorem: Let $X_1, ..., X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Then

(a) $E(\bar{X}) = \mu$,
(b) $Var(\bar{X}) = \sigma^2/n$,
(c) $E(S^2) = \sigma^2$.

Theorem: Let $X_1, ..., X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution, then

(a) $\bar{X}$ and S^2 are independent random variables,
(b) $\bar{X}$ has a $N(\mu, \sigma^2/n)$ distribution,
(c) $(n-1)S^2/\sigma^2$ has a chi squared distribution with $n-1$ degrees of freedom.

Theorem: (*Central Limit Theorem*) Let $X_1, X_2, ...$ be a sequence of iid random variables whose mgfs exist in a neighborhood of 0. Let $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 > 0$. (Both μ and σ^2 are finite since the mgf exists.) Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then for any $x, -\infty < x < \infty$,

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy,$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

Maximum Likelihood Estimators

Theorem: (Invariance Property of Maximum Likelihood Estimators) If $\hat{\theta}$ is the MLE of θ, then for any function $h(\theta)$, the MLE of $h(\theta)$ is $h(\hat{\theta})$.

(4) Test of Hypothesis

1. A statistical hypothesis is a conjecture about a population parameter. This conjecture may or may not be true.

2. The null hypothesis, symbolized by H_0, is a statistical hypothesis that states that there is no difference between a parameter and a specific value, or that there is no difference between two parameters.

3. The alternative hypothesis, symbolized by H_1, is a statistical hypothesis that states the existence of a difference between a parameter and a specific value, or states that there is a difference between two parameters.

4. Two-sided test Vs one-sided test

   (a) Two-sided test

   A medical researcher is interested in finding out whether a new medication will have any undesirable side effects. The researcher is particularly concerned with the pulse rate of the patients who take the medication. Will the pulse rate increase, decrease, or remain unchanged after a patient takes the medication? The researcher knows that the mean pulse rate for the population under study is 82 beats per minute.

   (b) One-sided test

   A chemist invents an additive to increase the life of an automobile battery. The mean lifetime of the automobile battery without the additive is 36 months. The chemist want to test whether the additive is useful.

5. Error in testing, power of a test

6. p-value

   The p-value (or probability value) is the probability of getting a sample statistic (such as the mean) or a more extreme sample statistic in the direction of the alternative hypothesis when the null hypothesis is true.