Sampling Distrbutions

Expectations

Let $X = (X_1, X_2, \dots, X_n)^{\mathrm{T}}$ be a set of random variables, the expectation of a function $g(X)$ is defined as

$$E\{g(X)\} = \sum_{x_1 \in X_1} \cdots \sum_{x_n \in X_n} g(X) p(x, \theta)$$

or

$$E\{g(X)\} = \int_{x_1 \in X_1} \cdots \int_{x_n \in X_n} g(X) f(x, \theta) dx_n \cdots dx_1$$

Expected Value and Variance of Linear Functions

Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be random variables with $E(X_i) = \mu_i$ and $E(Y_j) = \tau_j$. Furthermore, let $U = \sum_{i=1}^{n} a_i X_i$ and $V = \sum_{j=1}^{m} b_j Y_j$ where $\{a_i\}_{i=1}^{n}$ and $\{b_j\}_{j=1}^{m}$ are constants. We have the following properties:

$Z = aX + bY$

$E(Z) = a E(X) + b E(Y)$

$Var(Z) = a^2 Var(X) + b^2 Var(Y) + 2ab\, Cov(X,Y)$

- $E(U) = \sum_{i=1}^{n} a_i \mu_i$
- $Var(U) = \sum_{i=1}^{n} a_i^2 Var(X_i) + 2\sum\sum_{i<j} a_i a_j Cov(X_i, X_j)$
- $Cov(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{m} Cov(X_i, Y_j)$

Conditional Expectations

Let X_1 and X_2 be two random variables, the conditional expectation of $g(X_1)$, given $X_2 = x_2$, is defined as

$$E\{g(X_1)|X_2 = x_2\} = \sum_{x_1} g(x_1)p(x_1|x_2)$$

or

$$E\{g(X_1)|X_2 = x_2\} = \int_{x_1} g(x_1)f(x_1|x_2)dx_1.$$

Conditional Expectations

Furthermore,

$$E(X_1) = E_{X_2}\{E_{X_1|X_2}(X_1|X_2)\}$$

and

$$Var(X_1) = E_{X_2}\{Var_{X_1|X_2}(X_1|X_2)\} + Var_{X_2}\{E_{X_1|X_2}(X_1|X_2)\}$$

Sample

When collecting data to construct a sample, the sample is a collection of random variables. Therefore, the sample can be subjected to probability properties.

Sample is a collection of random variables.

iid Random Variables

A sample of random variables are said to be iid if they are identical and independentally distributed.

For example, X and Y are iid, if X and Y has the same distribution $f(\theta)$ and $X \perp Y$

Random Sample

Random Sample is equivalent to iid RV's.

$$\mathbf{X} = (X_1, \ldots, X_n)^T$$

Statistics

$T = g(X)$

T is also a RV

A statistic is a transformation of the the sample data. Before data is calculated, a statistic from a sample can take any value. Therefore, a statistic must be a random variable.

Mean

$g() = \frac{1}{n} \sum x_i$

Let $X_1, X_2, \ldots, X_n$ be a random sample that is *iid*. The mean is defined as:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Distribution

parameter

$X \sim F(\theta)$

Variance

Let $X_1, X_2, \ldots, X_n$ be a random sample that is *iid*. The variance is defined as:

sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$

Sampling Distributions

A sampling distribution is the distribution of a statistic.

Many known statistics have a known distribution.

$\bar{X}$

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, show that $\bar{X} \sim N(\mu, \sigma^2/n)$.

$$\bar{X} = \frac{1}{n}\sum X_i = \sum \frac{1}{n} X_i$$

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

$X_i \sim RV$

MGF Propert

$$\begin{cases} Z = \sum a_i X_i \\ M_Z(t) = \prod_{i=1}^{n} M_{X_i}(a_i t) \end{cases} \quad \text{if } X_i \perp X_j,\ i \neq j$$

$$Z = \bar{X} \qquad a_i = \frac{1}{n}$$

$$M_{\bar{X}}(t) = \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{n}\right) \quad \text{identical}$$

$$M_{\bar{X}}(t) = \prod_{i=1}^{n} e^{\mu t/n + \sigma^2 t^2/2n^2}$$

$$M_{\bar{X}}(t) = e^{\sum_{i=1}^{n} (\mu t/n + \sigma^2 t^2/2n^2)}$$

$$M_{\bar{X}}(t) = e^{n\mu t/n + n\sigma^2 t^2/2n^2}$$

$$M_{\bar{X}}(t) = e^{\mu t + \frac{t^2}{2} \cdot \frac{\sigma^2}{n}}$$

$$M_{\bar{X}}(t) = e^{\mu t + \frac{t^2}{2} \sigma^2}$$

$$\mu' = \mu \qquad \sigma^{2\prime} = \sigma^2/n$$

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

Sum of χ_1^2

$$Z_1 = \left(\frac{X_1 - \mu}{\sigma}\right)^2$$

Let $Z_1^2, \ldots, Z_n^2$ be a iid χ_1^2. Find $Y = \sum_{i=1}^n Z_i^2$

$$Y \sim \chi_n^2$$

$$Y = \sum a_i Z_i$$

$$M_Y(t) = \prod M_{Z_i}(a_i t)$$

$$M_{Z_i^2}(t) = (1-2t)^{-1/2}$$

$A \sim \chi^2_k$

$\rightarrow M_A(t) = (1-2t)^{-k/2}$

$$M_Y(t) = \prod_{i=1}^{n} (1-2t)^{-1/2}$$

$$= (1-2t)^{-n/2}$$

$$= (1-2t)^{-n/2}$$

$Y \sim \chi^2_n$

s^2

$X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$\bar{X} \perp S^2$

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

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$

$\sum Z_i^2 \sim \chi^2_n$

$(a+b)^2$

$\sum_{i=1}^{n} \left(\frac{(X_i - \bar{X}) + (\bar{X} - \mu)}{\sigma} \right)^2$

$a \quad b$

$$\sum \left(\frac{(x_i - \bar{x})^2}{\sigma^2} + \frac{2(x_i - \bar{x})(\bar{x} - \mu)}{\sigma^2} + \frac{(\bar{x} - \mu)^2}{\sigma^2} \right)$$

$$\frac{1}{\sigma^2} \sum \left((x_i - \bar{x})^2 + 2(\bar{x} - \mu)(x_i - \bar{x}) + (\bar{x} - \mu)^2 \right)$$

$$\frac{1}{\sigma^2} \left[\sum (x_i - \bar{x})^2 + 2(\bar{x} - \mu) \sum (x_i - \bar{x}) + n(\bar{x} - \mu)^2 \right]$$

$$\sum (x_i - \bar{x}) = 0$$

$$\sum x_i - \sum \bar{x}$$

$$\sum x_i - n\bar{x}$$

$$n\bar{x} - n\bar{x} = 0$$

$$\bar{x} = \frac{1}{n} \sum x_i$$

$$n\bar{x} = \sum x_i$$

$$\frac{1}{\sigma^2}\left[\sum(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2\right]$$

$$\frac{1}{\sigma^2}\left[(n-1)s^2 + n(\bar{x}-\mu)^2\right]$$

$$\frac{n-1}{n-1}\sum(x_i-\bar{x})^2 \qquad s^2$$

$$\frac{(n-1)s^2}{\sigma^2} + \frac{n(\bar{x}-\mu)^2}{\sigma^2} = W$$

$$W = \sum\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$M_W(t) = (1-2t)^{-n/2}$$

$$M_W(t) = E\left(e^{Wt}\right)$$

$$= E\left(e^{t\left(\frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{X}-\mu)^2}{\sigma^2}\right)}\right)$$

$$= E\left(e^{t\frac{(n-1)S^2}{\sigma^2}} e^{t\frac{n(\bar{X}-\mu)^2}{\sigma^2}}\right)$$

$$E\left(e^{t\frac{(n-1)S^2}{\sigma^2}}\right) E\left(e^{tn\frac{(\bar{X}-\mu)^2}{\sigma^2}}\right)$$

$$M_{\frac{(n-1)S^2}{\sigma^2}}(t)\; M_{\frac{n(\bar{X}-\mu)^2}{\sigma^2}}(t) = M_W(t)$$

↑ Unknown ↑ Unknown? ↑ Known

$$\frac{(\bar{X}-\mu)^2}{\sigma^2/n} = \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

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

$$Z_{\bar{X}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow Z_{\bar{X}} \sim N(0,1)$$

$$Z_{\bar{X}}^2 \sim \chi_1^2$$

MGF of χ^2_1

$$M_{\frac{(n-1)S^2}{\sigma^2}}(t)\; M_{n\frac{(\bar{x}-\mu)^2}{\sigma^2}}(t) = M_{\omega}(t)$$

$$M_{\frac{(n-1)S^2}{\sigma^2}}(t)(1-2t)^{-1/2} = (1-2t)^{-n/2}$$

$$M_{\frac{(n-1)S^2}{\sigma^2}}(t) = (1-2t)^{-n/2+1/2}$$

$$= (1-2t)^{-\frac{1}{2}(n-1)}$$

$$= (1-2t)^{-\frac{(n-1)}{2}}$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

t-distribution

Let $Z \sim N(0,1), W \sim \chi^2_\nu, Z \perp W$; therefore:

$$T = \frac{Z}{\sqrt{W/\nu}} \sim t_\nu$$

F-distribution

Let $W_1 \sim \chi^2_{\nu_1}$ $W_2 \sim \chi^2_{\nu_2}$, and $W_1 \perp W_2$; therefore:

$$F = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F_{\nu_1,\nu_2}$$

Order Statistics

Order statistics are a fundamental concept in statistics and probability, dealing with the properties of sorted random variables. They provide insights into the distribution and behavior of sample data, such as minimum, maximum, and quantiles. Understanding order statistics is crucial in various fields such as risk management, quality control, and data analysis.

Order Statistics

Let $X_1, X_2, \ldots, X_n$ be a sample of n independent and identically distributed (i.i.d.) random variables with a common probability density function $f(x)$. The order statistics are the sorted values of this sample, denoted as:

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

Here, $X_{(1)}$ is the minimum, and $X_{(n)}$ is the maximum of the sample.

Order Statistics

- $X_{(k)}$: The k-th order statistic, representing the k-th smallest value in the sample.
- $X_{(1)}, X_{(n)}$: The minimum and maximum of the sample, respectively.

Distribution of Order Statistic

The distribution of the k-th order statistic $X_{(k)}$ can be derived using combinatorial arguments. Its PDF is given by:

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k}f(x$$

This formula shows how the distribution of $X_{(k)}$ depends on the underlying distribution of the sample and its position k.

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be identical and independent distributed random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. We define

$$Y_n = \sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) \text{ where } \bar{X} = x\frac{1}{n}\sum_{i=1}^{n} X_i.$$

Then, the distribution of the function Y_n converges to a standard normal distribution function as $n \to \infty$.

Central Limit Theorem

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

Central Limit Theorem Proof

Example

Let $X_1, \ldots, X_n \overset{iid}{\sim} \chi^2_p$, the MGF is $M(t) = (1-2t)^{-p/2}$. Find the distribution of $\bar{X}$ as $n \to \infty$.