CHAPTER 1

SAMPLING DISTRIBUTIONS

§1.1 SAMPLING

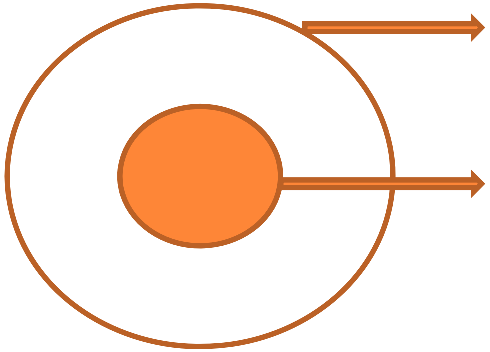
Def 1.1 : The totality of elements which are under discussion and about which information is desired is called the *target population*.

Def 1.2 : Let $X_1,\ldots,X_n$ be random variables having a joint density $f_{X_1,\ldots,X_n}(\cdot,\ldots,\cdot)$ that factors as follows : $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f(x_1)f(x_2)\cdots f(x_n)$, where $f(\cdot)$ is the common density of each X_i. Then $X_1,\ldots,X_n$ is defined to be a *random sample* of size n from a population with density $f(\cdot)$.

Note : $x_1, \ldots, x_n$ are observed values (or observations) of $X_1, \ldots, X_n$.

Def 1.3 : Let $X_1, \ldots, X_n$ be a random sample from a population with a density $f(\cdot)$, then this population is called the *sample population*.

Ex1.1 :



20-year-old males in the US

$\Rightarrow$ target population

some large city we sampled

$\Rightarrow$ sample population

Goal : Study the religious habits of 20-year-old males in the US.

Def 1.4 : Let $X_1,\ldots,X_n$ be a random sample from a population with a density $f(\cdot)$. The distribution of $X_1,\ldots,X_n$ is defined to be the joint distribution of $X_1,\ldots,X_n$; that is ,

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

Ex 1.2 : Suppose X is a random variable having a Bernoulli distribution $f(x) = p^x q^{1-x}, x = 0,1$.Let X_1 and X_2 be a random sample from $f(\cdot)$.Then

$$f_{X_1,X_2}(x_1,x_2) = f(x_1)f(x_2) = p^{x_1+x_2}q^{2-x_1-x_2}, x_1,x_2 = 0,1 .$$

<u>Remark 1.1</u> : The definition of random sampling here has ruled out sampling from a finite population without replacement, since then, the results of drawings are not independent.

<u>Note</u> : Suppose the form of a density $f(\cdot\,;\theta)$ is known, but the parameter θ is unknown. Let $X_1,\ldots,X_n$ be a random sample from $f(\cdot\,;\theta)$ and $T = t(x_1,\ldots,x_n)$ represent the value of some function which estimates the unknown parameter θ. We want to determine which function will be the best one to estimate θ.

Def 1.5 : Suppose $X_1,\ldots,X_n$ is a random sample from the distribution of a random variable X. Then any function $T = \ell(X_1,\ldots,X_n)$ of the sample is called a *statistic*.

Ex1.3 : Suppose $X_1,\ldots,X_n$ is a random sample from a population. Then

1. $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ is a statistic.
2. $\min\left[X_1,\ldots,X_n\right] + \max\left[X_1,\ldots,X_n\right]$ is also a statistic.

Def 1.6 : Let X be a random variable. The *rth (population) moment* of X is defined as $\mu_r' = E\left(X^r\right)$.In particular,if $r=1$, then $\mu_1' = E\left(X\right) = \mu$, the mean of X .

Def 1.7 : Let X be a random variable. The *rth central moment* of X about μ is defined by $\mu_r = E\left[\left(X-\mu\right)^r\right]$. In particular,if $r=2$, then $\mu_2 = E\left[\left(X-\mu\right)^2\right] = \sigma^2$, the variance of X .

Def 1.8 : Let $X_1,\ldots,X_n$ be a random sample from a population. The *rth sample moment about 0* is defined to be $M_r^{'}=\frac{1}{n}\sum_{i=1}^{n}X_i^r$. If $r=1$, we get the sample mean $\bar{X}=\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$.

Also,the *rth sample moment about $\bar{X}$* is defined by $M_r=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^r$.

Thm 1.1 : Let $X_1, \ldots, X_n$ be a random sample from a population. Then

(a) $E\left(M_r^{'}\right) = \mu_r^{'}$ $\left(\text{if } \mu_r^{'} \text{ exists}\right)$;

(b) $Var\left(M_r^{'}\right) = \dfrac{1}{n}\left\{E\left(X^{2r}\right) - \left[E\left(X^{r}\right)\right]^2\right\}$

$$= \frac{1}{n}\left[\mu_{2r}^{'} - \left(\mu_r^{'}\right)^2\right] \quad \left(\text{if } \mu_{2r}^{'} \text{ exists}\right).$$

In particular, if $r = 1$,then $E\left(\bar{X}\right) = \mu$ and

$Var\left(\bar{X}\right) = \dfrac{\sigma^2}{n}$, where μ and σ^2 are the mean and variance of the population, respectively.

Def 1.9 : Let $X_1,\dots,X_n$ be a random sample from a population. The *sample variance* is defined by

$$S^2 = S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \;, n > 1.$$

Thm1.2 : Let $X_1,\dots,X_n$ be a random sample from a population with mean μ and variance σ^2. Then

$$E\left(S^2\right) = \sigma^2 \text{ and } Var\left(S^2\right) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right),\; n > 1.$$

where μ and σ^2 are the mean and variance of the population, respectively.

§1.2 SAMPLE MEAN

Thm1.3 : *(Weak law of large number)*

Let $X_1,\ldots,X_n$ be a random sample from a distribution with mean $E(X_i)=\mu$ and variance $Var(X_i)=\sigma^2<\infty$. Then , for every $\varepsilon>0$, we have $\lim\limits_{n\to\infty} P\left(\left|\bar{X}_n-\mu\right|<\varepsilon\right)=1$.

Thm1.4 : *(Strong law of large number)*

Let $X_1,\ldots,X_n$ be a random sample from a distribution with mean $E(X_i)=\mu$ and variance $Var(X_i)=\sigma^2<\infty$. Then , for every $\varepsilon>0$, we have $P\left(\lim\limits_{n\to\infty}\left|\bar{X}_n-\mu\right|<\varepsilon\right)=1$.

Thm1.5 : *(Central limit theorem; CLT)*

Let $X_1,\ldots,X_n$ be a random sample from a distribution with finite mean $E\left(X_i\right)=\mu$ and variance $Var\left(X_i\right)=\sigma^2<\infty$. Then, the random variable $Z_n=\dfrac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$ approaches the standard normal distribution $N\left(0,1\right)$ as $n\to\infty$. That is,

$$F_{Z_n}\left(z\right)=P\left(Z_n\le z\right)\to\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz=\Phi\left(z\right).$$

§1.3 SAMPLING FROM THE NORMAL DISTRIBUTION

Thm1.6 : Let $X_1,\ldots,X_n$ be a random sample from a $N(\mu,\sigma^2)$.Then $\bar{X} \sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$.

Thm1.7 : Let $X_1,\ldots,X_n$ be independent random variables and $X_i \sim N(\mu_i,\sigma_i^{\,2})$, $i=1,\ldots,n$. Than

$$U=\sum_{i=1}^{n}\left(\frac{X_i-\mu_i}{\sigma_i}\right)^2 \sim \chi^2(n).$$

Cor1.1 : If $X \sim N(\mu,\sigma^2)$,then $\left(\dfrac{X-\mu}{\sigma}\right) \sim \chi^2(1)$.

Cor1.2 : Let $X_1,\ldots,X_n$ be a random sample from a $N\left(\mu,\sigma^2\right)$.

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

Thm1.8 : Let $X_1,\ldots,X_n$ be a random sample from a $N\left(\mu,\sigma^2\right)$. Then (a) $\bar{X}$ and S^2 are independent random variables. (b) $\frac{\left(n-1\right)S^2}{\sigma^2}\sim\chi^2\left(n-1\right)$.

Thm1.9 : Suppose U and V are independent variables and $U\sim\chi^2\left(m\right)$ and $V\sim\chi^2\left(n\right)$. Then $X=\frac{U/m}{V/n}\sim F\left(m,n\right)$.

Cor1.3 : Let $X_1,\ldots,X_m$ be a random sample from a $N\left(\mu_X,\sigma_X^2\right)$, and let $Y_1,\ldots,Y_n$ be a random sample from a $N\left(\mu_Y,\sigma_Y^2\right)$. Suppose the two samples are independent. Then

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F\left(m-1,n-1\right).$$

In particular,if $\sigma_X^2=\sigma_Y^2$, then $\dfrac{S_X^2}{S_Y^2} \sim F\left(m-1,n-1\right)$.

Remark1.2 : If $X \sim F(m,n)$, then

1. $\mathrm{E}(X) = \dfrac{n}{n-2}$ for $n > 2$ and

$$Var(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \text{ for } n > 4.$$

2. $\dfrac{1}{X} \sim F(n,m)$ and $F_\alpha(m,n) = \dfrac{1}{F_{1-\alpha}(n,m)}$.

Thm1.10 : If $\mathrm{Z} \sim N(0,1)$ and $U \sim \chi^2(r)$.Then

$$X = \frac{\mathrm{Z}}{\sqrt{U/r}} \sim t(r).$$

Remark1.3 :

(1) If $r=1$, then the t-distribution reduces to a Cauchy distribution.

(2) As r increases, the t-distribution approaches the $N(0,1)$.

(3) $X^2 = \dfrac{Z^2}{U/r} \sim F(1,r)$.

Cor1.4 : Suppose that $X_1,\ldots,X_n$ is a random sample from a $N(\mu,\sigma^2)$. Then $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ and $\dfrac{(\bar{X}-\mu)^2}{S^2/n} \sim F(1,n-1)$.

§1.4 ORDER STATISTICS

1.4.1 DEFINITIONS AND DISTRIBUTIONS

Def 1.10 : Let $X_1,\ldots,X_n$ be a random sample from a population. Then $Y_1 \le Y_2 \le \cdots \le Y_n$ $\left(X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}\right)$, where $Y_i\left(X_{(i)}\right), i = 1,\ldots n,$ are the X_i arranged in order of increasing magnitudes and are defined to be the *order statistics* corresponding to $X_1,\ldots,X_n$.

Note :

(i) Y_i are statistics $\left(\text{functions of } X_1,\ldots,X_n\right)$ and are in order.

(ii) Order statistics are NOT independent $\left(Y_j \geq y \Rightarrow Y_{j+1} \geq y\right)$.

(iii) $Y_1 = \min\left[X_1,\ldots,X_n\right]$ and $Y_n = \max\left[X_1,\ldots,X_n\right]$.

Thm1.11 : Let $Y_1,\ldots,Y_n$ denote the order statistics of a random sample, $X_1,\ldots,X_n$ from a c.d.f $F(\cdot)$.

Then $F_{Y_\alpha}(y)=\sum_{j=\alpha}^{n}\binom{n}{j}\left[F(y)\right]^j\left[1-F(y)\right]^{n-j}$.

Cor1.5 : From Theorem 1.11,we have

(a) $F_{Y_n}(y)=\binom{n}{n}\left[F(y)\right]^n\left[1-F(y)\right]^0=\left[F(y)\right]^n$;

(b) $F_{Y_1}(y)=\sum_{j=1}^{n}\binom{n}{j}\left[F(y)\right]^j\left[1-F(y)\right]^{n-j}$

$$=1-\left[1-F(y)\right]^n$$

Thm1.12 :Let $Y_1,\ldots,Y_n$ denote the order statistics of a random sample, $X_1,\ldots,X_n$ from a continuous population with c.d.f $F(\cdot)$ and p.d.f $f(\cdot)$.

(a)The p.d.f of Y_α is

$$f_{Y_\alpha}(y)=\frac{n!}{(\alpha-1)!(n-\alpha)!}\left[F(y)\right]^{\alpha-1}\left[1-F(y)\right]^{n-\alpha}f(y).$$

(b) The joint p.d.f of Y_α and Y_β, $1 \le \alpha < \beta \le n$, is

$$f_{Y_\alpha, Y_\beta}\left(y_\alpha, y_\beta\right) = \frac{n!}{(\alpha-1)!(\beta-\alpha-1)!(n-\beta)!}\left[F\left(y_\alpha\right)\right]^{\alpha-1}$$

$$\times\left[F\left(y_\beta\right)-F\left(y_\alpha\right)\right]^{\beta-\alpha-1}\left[1-F\left(y_\beta\right)\right]^{n-\beta} f\left(y_\alpha\right) f\left(y_\beta\right),$$ for

$-\infty < y_\alpha < y_\beta < \infty$.

(c) The joint p.d.f of $Y_1, \ldots, Y_n$ is

$$f_{Y_1 \cdots Y_n}\left(y_1, \cdots, y_n\right) = \begin{cases} n! f\left(y_1\right) \cdots f\left(y_n\right) & y_1 < y_2 < \cdots < y_n \\ 0 & \text{otherwise.} \end{cases}$$

Def1.11 : Let $Y_1,\ldots,Y_n$ be the order statistics of a random sample, $X_1,\ldots,X_n$, from a population.

(i) The *sample median* is defined by

$$M=\begin{cases} Y_{\frac{n+1}{2}} & \text{if } n \text{ is odd} \\ \dfrac{Y_{\frac{n}{2}}+Y_{\frac{n}{2}+1}}{2} & \text{if } n \text{ is even.} \end{cases}$$

(ii) The *sample range* is defined to be $R=Y_n-Y_1$.

(iii) The *sample midrange* is defined to be $T=\dfrac{Y_1+Y_n}{2}$.

Remark1.4 :

(i) If $n = 2k+1$, then Y_{k+1} is the sample median and $f_{k+1}(y)$ is given by Theorem1.12(a).

(ii) If $n = 2k$, the sample median is $\dfrac{(Y_k + Y_{k+1})}{2}$. We can obtain the distribution by a transformation starting with the joint density of Y_k and Y_{k+1} given by Theorem1.12(b).

Thm1.13 : If R is the sample range and T is the sample midrange from a continuous population with cdf $F(\cdot)$ and pdf $f(\cdot)$. Then the joint pdf of R and T is given by $f_{R,T}(r,t) = n(n-1)$

$$\times\left[F(t+r/2) - F(t-r/2)\right]^{n-2} f(t-r/2) f(t+r/2),$$

$r > 0$, and the marginal distributions are given by

$f_R(r) = \int_{-\infty}^{\infty} f_{R,T}(r,t)\,dt$ and $f_T(t) = \int_0^{\infty} f_{R,T}(r,t)\,dr.$

Ex1.5 : $X_1,\ldots,X_n \sim iid\ U(0,1)$

$$f_{R,T}(r,t) = n(n-1)\left[\left(t+\frac{r}{2}\right)-\left(t-\frac{r}{2}\right)\right]\cdot 1\cdot 1$$

$$= n(n-1)r^{n-2},\ 0<r<1,\ \frac{r}{2}<t<1-\frac{r}{2}$$

$$f_R(r) = \int_{-\infty}^{\infty} f_{R,T}(r,t)\,dt$$

$$= \int_{\frac{r}{2}}^{1-\frac{r}{2}} n(n-1)r^{n-2}dt = \left[n(n-1)r^{n-2}\cdot t\right]_{\frac{r}{2}}^{1-\frac{r}{2}}$$

$$= n(n-1)(1-r)r^{n-2},\ 0<r<1.$$

$$f_T(t) = \int_{-\infty}^{\infty} f_{R,T}(r,t)\,dr$$

$$= \begin{cases} \int_0^{2(1-t)} n(n-1) r^{n-2} dr, & \frac{1}{2} < t < 1 \\ \int_0^{2t} n(n-1) r^{n-2} dr, & 0 < t \le \frac{1}{2} \end{cases}$$

$$= \begin{cases} n(2t)^{n-1}, & 0 < t \le \frac{1}{2} \\ n\left[2(1-t)\right]^{n-1}, & \frac{1}{2} < t < 1 \end{cases}$$

$$\therefore R \sim Beta(n-1,2),\ E(R)=\frac{n-1}{n+1}$$

$$E(T)=\int_t t\cdot f_T(t)\,dt$$ or

$$E(T)=E\left(\frac{Y_1+Y_n}{2}\right)=\frac{1}{2}\left[E(Y_1)+E(Y_n)\right]$$

$$=\frac{1}{2}\left(\frac{1}{n+1}+\frac{n}{n+1}\right)=\frac{1}{2}$$

1.4.2 ASYMPTOTIC DISTRIBUTIONS

Ex1.6 : Let $Y_1,\ldots,Y_n$ denote the order statistics of a sample of size n from $U(0,\theta)$.

Then $f(x)=\frac{1}{\theta},\ 0<x<\theta.\quad F(x)=\frac{x}{\theta}$

So $F_{Y_n}(y)=\left[F(y)\right]^n=\begin{cases}\left(\frac{y}{\theta}\right)^n, & 0\le y<\theta\\ 1, & \theta\le y<\infty\end{cases}$

Thus, $\lim_{n\to\infty}F_{Y_n}(y)=\begin{cases}0, & 0\le y<\theta\\ 1, & \theta\le y<\infty\end{cases}$

Ex1.7 : Y_n: nth order statistics of a random sample from $U(0,\theta)$. Let $Z_n = n(\theta - Y_n)$. Then the p.d.f of is

$$h_{Z_n}(z) = f_{Y_n}\left(\theta - \frac{z}{n}\right)\cdot\left|\frac{dy}{dz}\right|$$

$$= n\cdot\left(\frac{\theta - \frac{z}{n}}{\theta}\right)^{n-1}\cdot\frac{1}{\theta}\cdot\left|-\frac{1}{n}\right|$$

$$= \frac{\left(\theta - \frac{z}{n}\right)^{n-1}}{\theta^n},\ 0 < z < n\theta$$

So the p.d.f of Z_n is

$$H_{Z_n}(z)=\begin{cases}0, & z<0\\ \int_0^z h_{Z_n}(u)\,du, & 0\le z<n\theta\\ 1, & n\theta\le z\end{cases}$$

$$=\begin{cases}0, & z<0\\ \int_0^z \frac{\left(\theta-\frac{u}{n}\right)^{n-1}}{\theta^n}\,du=1-\left(1-\frac{z}{n\theta}\right)^n, & 0\le z<n\theta\\ 1, & n\theta\le z<\infty\end{cases}$$

Hence the limiting distribution of Z_n is

$$H_Z = \lim_{n\to\infty} H_{Z_n}\left(z\right) = \begin{cases} 0, & z < 0 \\ \lim_{n\to\infty}\left[1-\left(1-\dfrac{z}{n\theta}\right)^n\right], & z \geq 0 \end{cases}$$

$$= \begin{cases} 0, & z < 0 \\ 1-e^{-\frac{z}{\theta}}, & z \geq 0 \end{cases}$$