22.10.19

Sampling Distributions - 2016 Question Paper

1. (d) If $X \sim \chi^2_{(n)}$, then prove that $\frac{X-n}{\sqrt{2n}}$ is a $N(0,1)$ variate for large n.

Let $Z = \frac{X-n}{\sqrt{2n}} \Rightarrow M_Z(t) = E(e^{tZ}) = E\left[e^{t\left(\frac{X-n}{\sqrt{2n}}\right)}\right] = E\left[e^{\left(\frac{tX}{\sqrt{2n}} - \frac{tn}{\sqrt{2n}}\right)}\right]$

$\Rightarrow M_Z(t) = e^{-tn/\sqrt{2n}} \cdot E\left[e^{\frac{t}{\sqrt{2n}}X}\right] = e^{-tn/\sqrt{2n}} \cdot M_X\left[\frac{t}{\sqrt{2n}}\right]$ $\quad (\because M_X(t) = (1-2t)^{-n/2})$

$\Rightarrow M_Z(t) = e^{-tn/\sqrt{2n}} \left(1 - \frac{2t}{\sqrt{2n}}\right)^{-n/2}$

$\Rightarrow \log M_Z(t) = -\sqrt{\frac{n}{2}}\,t - \frac{n}{2}\log\left(1 - \frac{2t}{\sqrt{2n}}\right) = -\sqrt{\frac{n}{2}}\,t - \frac{n}{2}\log\left(1 - \sqrt{\frac{2}{n}}\,t\right)$

$= -t\sqrt{\frac{n}{2}} + \frac{n}{2}\left[\sqrt{\frac{2}{n}}\,t + \frac{1}{2}\left(\frac{2}{n}\right)t^2 + O(n^{-3/2})\right]$

$= -t\sqrt{\frac{n}{2}} + t\sqrt{\frac{n}{2}} + \frac{t^2}{2} + O(n^{-3/2})$

$\Rightarrow \lim_{n\to\infty} \log M_Z(t) = \frac{t^2}{2} \Rightarrow \lim_{n\to\infty} M_Z(t) = e^{t^2/2}$ = M.G.F. of $N(0,1)$.

$\therefore Z = \frac{X-n}{\sqrt{2n}} \sim N(0,1)$ as $n \to \infty$.

(c) If variable t has Student's distribution with 2 degrees of freedom, then find $P(-\sqrt{2} \le t \le \sqrt{2})$.

Given $t \sim t_{(2)} \Rightarrow f(t) = \dfrac{1}{\sqrt{n}\,B\left(\frac{n}{2},\frac{1}{2}\right)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}\,; -\infty < t < \infty \quad (n=2)$

$\Rightarrow f(t) = \dfrac{1}{\sqrt{2}\,B\left(1,\frac{1}{2}\right)\left(1+\frac{t^2}{2}\right)^{3/2}}\,; -\infty < t < \infty.$

$P(-\sqrt{2} \le t \le \sqrt{2}) = \displaystyle\int_{-\sqrt{2}}^{\sqrt{2}} \frac{1}{\sqrt{2}\,B\left(1,\frac{1}{2}\right)\left(1+\frac{t^2}{2}\right)^{3/2}}\,dt$

$= \displaystyle\int_{-\pi/4}^{\pi/4} \frac{\sqrt{2}\sec^2\theta\,d\theta}{\sqrt{2}\cdot 2\cdot(1+\tan^2\theta)^{3/2}}$

$= \displaystyle\int_{-\pi/4}^{\pi/4} \frac{1}{2}\cdot\cos\theta\,d\theta = \frac{1}{2}\left[\sin\theta\right]_{-\pi/4}^{\pi/4} = \frac{1}{\sqrt{2}}$

Put $\frac{t^2}{2} = \tan^2\theta \Rightarrow \frac{t}{\sqrt{2}} = \tan\theta$

$t = \sqrt{2} \Rightarrow \theta = \tan^{-1}\frac{\sqrt{2}}{\sqrt{2}} = \pi/4$

$t = -\sqrt{2} \Rightarrow \theta = -\pi/4$

$\left(B\left(1,\frac{1}{2}\right) = \frac{\Gamma(1)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{3}{2}\right)} = \frac{\sqrt{\pi}}{\frac{1}{2}\sqrt{\pi}} = 2\right)$

$\therefore P(-\sqrt{2} \le t \le \sqrt{2}) = \frac{1}{\sqrt{2}}$

(f) Let X_1 and X_2 be independent random variables with density law $f(x) = e^{-x}; x \geq 0$, then show that $Z = \frac{X_1}{X_2}$ has F-distribution.

Given $f(x) = e^{-x}; x \geq 0$

$\therefore X_1 \sim Exp(1) \Rightarrow X_1 \sim G(1,1) \Rightarrow \frac{X_1}{2} \sim G(\frac{1}{2}, 1) = G(\frac{1}{2}, \frac{2}{2})$ (But $G(\frac{1}{2}, \frac{n}{2}) = \chi^2_{(n)}$)

$\therefore \frac{X_1}{2} \sim \chi^2_{(2)} \Rightarrow X_1 \sim 2 \cdot \chi^2_{(2)}$.

Similarly, $X_2 \sim 2 \cdot \chi^2_{(2)}$.

Now, $Z = \frac{X_1}{X_2} = \frac{2\chi^2_{(2)}}{2\chi^2_{(2)}} = \frac{\chi^2_{(2)}/2}{\chi^2_{(2)}/2}$

$\Rightarrow Z \sim F(2,2)$ ($\because X_1$ and X_2 are independent)

But, if $X \sim \chi^2_{(n)}$ & $Y \sim \chi^2_{(m)}$ & X & Y are independent, then $\frac{X/n}{Y/m} \sim F(n,m)$

$\therefore Z = \frac{X_1}{X_2}$ has F-distribution.

(g) If $X \sim U(0,1)$, then show that $-2\log X \sim \chi^2_{(2)}$.

Given $X \sim U(0,1)$. Let $Z = -2\log X$.

$M_Z(t) = E(e^{tZ}) = E[e^{(-2\log x)t}] = E[e^{\log x^{-2t}}] = E(x^{-2t})$

$= \int_0^1 x^{-2t}(1)\,dx = \left[\frac{x^{1-2t}}{1-2t}\right]_0^1 = \frac{1}{1-2t} = (1-2t)^{-1} = (1-2t)^{-2/2}$

$=$ M.G.F. of $\chi^2_{(2)}$ ($\because$ M.G.F. of $\chi^2_{(n)} = (1-2t)^{-n/2}$)

$\therefore Z \sim \chi^2_{(2)}$, i.e, $-2\log X \sim \chi^2_{(2)}$.

Section-B

5a.

(a) If $X_1, X_2, \ldots, X_n$ are independent random variables with continuous distribution functions $F_1, F_2, \ldots, F_n$ respectively, then show that:

$-2\log[F_1(x_1) \cdot F_2(x_2) \ldots F_n(x_n)] \sim \chi^2_{(2n)}$.

We know that, distribution function of any random variable follows standard uniform distribution, i.e.,

$F_i(x_i) \sim U(0,1) \quad \forall\, i = 1, 2, \ldots, n$.

Let $Z = -2\log(F_1 F_2 \ldots F_n)$

Now, $M_Z(t) = E(e^{tZ}) = E\left[e^{t(-2\log(F_1F_2\ldots F_n))}\right]$

$= E\left[e^{\log(F_1 \cdot F_2 \ldots F_n)^{-2t}}\right]$

$= E\left[(F_1F_2\ldots F_n)^{-2t}\right] = \left[E(F_1^{-2t})\right]^n$ ($\because F_i$'s are i.i.d.'s)

Now, $E(F_1^{-2t}) = \int_0^1 F_1^{-2t}(1)\,dF_1 = \left[\frac{F_1^{1-2t}}{1-2t}\right]_0^1 = (1-2t)^{-1}.$

$\therefore M_Z(t) = \left[(1-2t)^{-1}\right]^n = (1-2t)^{-n} = (1-2t)^{-2n/2}$ which is the M.G.F. of $\chi^2_{(2n)}$ distribution.

$\therefore Z = -2\log(F_1(x_1).F_2(x_2)\ldots.F_n(x_n)) \sim \chi^2_{(2n)}$

(b) Let $P_x = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})}\int_0^x w^{\frac{n-2}{2}} e^{-w}\,dw,\ x>0$. Show that $x < \frac{n}{1-P_x}$.

Let $W \sim \chi^2_{(n)} \Rightarrow P_x = P(W<x)$, we have to show that $x < \frac{n}{1-P_x}$.

$\therefore E(W) = n$ and $var(W) = 2n$.

$$\Rightarrow n = \int_0^\infty w.\frac{e^{-w/2}w^{n/2-1}}{2^{n/2}\Gamma(\frac{n}{2})}dw = \int_0^x \frac{e^{-w/2}w^{n/2}}{2^{n/2}\Gamma(n/2)}dw + \int_x^\infty \frac{w.e^{-w/2}w^{n/2-1}}{2^{n/2}\Gamma(n/2)}dw$$

$$\Rightarrow n > \int_x^\infty w.\frac{e^{-w/2}w^{n/2-1}}{2^{n/2}\Gamma(\frac{n}{2})}dw > \int_x^\infty x.\frac{e^{-w/2}w^{\frac{n-2}{2}}}{2^{n/2}\Gamma(n/2)}dw = -x(P_x-1)$$

$$\Rightarrow n > x(1-P_x) \Rightarrow x < \frac{n}{1-P_x}$$

(c) For the t-distribution with n d.f., prove that $\mu_{2r} = \frac{n(2r-1)}{(n-2r)}\mu_{2r-2};\ n>2r.$

Let $f(t)$ be the density function of t-distribution. Since $f(t)$ is symmetrical about the line $t=0$, we have μ_1' (mean) $=0$.

Now, moments of even order are given by:

$$\mu_{2r} = \mu_{2r}'(\text{about origin}) = \int_{-\infty}^{\infty} t^{2r} f(t)\,dt = 2\int_0^\infty t^{2r} f(t)\,dt$$

$$= \frac{2}{B(\frac{1}{2},\frac{n}{2})\sqrt{n}}\int_0^\infty \frac{t^{2r}}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}dt.$$

The integral is convergent if $2r<n$.

Put $1+\frac{t^2}{n} = \frac{1}{y} \Rightarrow t^2 = \frac{n(1-y)}{y} \Rightarrow 2t\,dt = \frac{-n}{y^2}dy$

$t=0 \Rightarrow y=1$ & $t\to\infty \Rightarrow y\to 0$.

$$\therefore \mu_{2r} = \frac{2}{\sqrt{n}\,B\left(\frac{1}{2},\frac{n}{2}\right)} \int_1^0 \frac{t^{2r}(-n)}{\left(\frac{1}{y}\right)^{\frac{n+1}{2}}(2ty^2)}\cdot dy = \frac{n}{\sqrt{n}\,B\left(\frac{1}{2},\frac{n}{2}\right)} \int_0^1 (t^2)^{(2r-1)/2}\cdot y^{\frac{n+1}{2}-2}\cdot dy$$

$$= \frac{\sqrt{n}}{B\left(\frac{1}{2},\frac{n}{2}\right)} \int_0^1 \left[n\left(\frac{1-y}{y}\right)\right]^{r-\frac{1}{2}}\cdot y^{((n+1)/2)-2}\,dy$$

$$= \frac{n^r}{B\left(\frac{1}{2},\frac{n}{2}\right)} \int_0^1 y^{\frac{n}{2}-r-1}(1-y)^{r-\frac{1}{2}}\,dy = \frac{n^r}{B\left(\frac{1}{2},\frac{n}{2}\right)}\cdot B\left(\frac{n}{2}-r,\, r+\frac{1}{2}\right), \quad n>2r$$

$$= n^r.\ \frac{\Gamma[(n/2)-r].\ \Gamma(r+\frac{1}{2})}{\Gamma(\frac{1}{2}).\ \Gamma(n/2)}, \quad n>2r$$

$$= n^r\ \frac{(r-\frac{1}{2})\cdot(r-\frac{3}{2})\cdots(\frac{3}{2})\cdot(\frac{1}{2})\cdot\Gamma(\frac{1}{2})\cdot\Gamma(\frac{n}{2}-r)}{\Gamma(\frac{1}{2})(\frac{n}{2}-1)(\frac{n}{2}-2)\cdots(\frac{n}{2}-r)\cdot\Gamma(\frac{n}{2}-r)}, \quad n>2r$$

$$\Rightarrow \mu_{2r} = n^r\ \frac{(2r-1)(2r-3)\cdots 3.1}{(n-2)(n-4)\cdots(n-2r)}, \quad n>2r \quad \to (1)$$

Replace $2r$ by $2r-2$ in (1);

$$\mu_{2r-2} = n^{r-1}.\ \frac{(2r-3)(2r-5)\cdots 3.1}{(n-2)(n-4)\cdots(n-2r-2)}, \quad n>2r \quad \to (2)$$

$$\frac{(1)}{(2)} \Rightarrow \frac{\mu_{2r}}{\mu_{2r-2}} = \frac{n(2r-1)}{(n-2r)}, \quad n>2r \Rightarrow \mu_{2r} = \frac{n(2r-1)}{(n-2r)}\mu_{2r-2}, \quad n>2r$$

6.
(a) Prove that $\frac{ns^2}{\sigma^2}$ is distributed as Chi-square with $(n-1)$ d.f. where s^2 and σ^2 are the variances of sample of size n and popⁿ respectively.

7.
(b) If $\bar{x}$ and S^2 be the usual sample mean and sample variance based on a random sample of n observations from $N(\mu,\sigma^2)$ and if $T = \frac{\bar{x}-\mu}{S/\sqrt{n}}$, then prove that $Cov(\bar{x},T) = \frac{\sigma\sqrt{n-1}\cdot\Gamma\left(\frac{n-2}{2}\right)}{\sqrt{2n}\cdot\Gamma\left(\frac{n-1}{2}\right)}$, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

x_i's are sample of n observations from $N(\mu,\sigma^2) \Rightarrow \bar{x}$ and $x_i-\bar{x}$ are independent.

$\Rightarrow \bar{x}$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ are independent.

$\therefore E(T) = E\left(\frac{\bar{x}-\mu}{S/\sqrt{n}}\right) = E(\bar{x}-\mu).E\left(\frac{1}{S/\sqrt{n}}\right) = 0.E\left(\frac{1}{S/\sqrt{n}}\right) = \underline{0}$ $\quad (\because E(\bar{x})=\mu)$

$\therefore E(T^2) = E\left[\frac{(\bar{x}-\mu)^2}{S^2/n}\right] = n.E(\bar{x}-\mu)^2.E\left(\frac{1}{S^2}\right)$

Now, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{\sigma^2}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2 \Rightarrow \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \rightarrow ①$.

Now, $\text{cov}(\bar{x},T) = E(\bar{x}T) - E(\bar{x})E(T)$ $\quad (\because E(T)=0)$

$= E[((\bar{x}-\mu)+\mu)T] = E[(\bar{x}-\mu)T] + \mu E(T)$

$= E[(\bar{x}-\mu)T]$

$= E\left[(\bar{x}-\mu)\left(\frac{\bar{x}-\mu}{S/\sqrt{n}}\right)\right] = E(\bar{x}-\mu)^2.E\left(\frac{1}{S/\sqrt{n}}\right)$ $\quad (\because \bar{x}$ & S are ind.)

$= \sqrt{n}.\text{Var}(\bar{x}).E\left(\frac{1}{S}\right) = \sqrt{n}.\frac{\sigma^2}{n}.E\left(\frac{1}{S}\right) = \frac{\sigma^2}{\sqrt{n}}.E\left(\frac{1}{S}\right)$.

From ①, $\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2 \sim \chi^2_{(n-1)}$

$$\Rightarrow E\left[\frac{1}{\sqrt{\frac{(n-1)S^2}{\sigma^2}}}\right] = \int_0^\infty \frac{1}{\theta}.\frac{e^{-\theta/2}\,\theta^{\frac{n-1}{2}-1}}{2^{\frac{n-1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)}d\theta$$

$$= \frac{1}{2^{(n-1)/2}\,\Gamma\left(\frac{n-1}{2}\right)}.\int_0^\infty e^{-\theta/2}.\theta^{\frac{n-2}{2}-1}\,d\theta$$

$$= \frac{1}{2^{(n-1)/2}\,\Gamma\left(\frac{n-1}{2}\right)}.\frac{\Gamma\left(\frac{n-2}{2}\right)}{\left(\frac{1}{2}\right)^{\frac{n-2}{2}}} = \frac{1}{\sqrt{2}}.\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}.$$

$$\Rightarrow E\left(\frac{1}{S}\right) = \frac{\sqrt{n-1}}{\sigma}.\frac{1}{\sqrt{2}}.\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

$$\therefore \text{Cov}(\bar{x},T) = \frac{\sigma^2}{\sqrt{n}}.\frac{\sqrt{n-1}}{\sigma}.\frac{1}{\sqrt{2}}.\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} = \frac{\sigma\sqrt{n-1}.\Gamma\left(\frac{n-2}{2}\right)}{\sqrt{2n}.\Gamma\left(\frac{n-1}{2}\right)}.$$

Hence proved.

8.(b) Let $x_1, x_2, \dots, x_n$ be independent observations from $N(\mu,\sigma^2)$ and let $\bar{x}$ and s^2 be the sample mean and sum of the squares of deviations from mean respectively. Let x' be one more observation independent of previous ones.

Obtain the distribution of $U = \frac{x'-\bar{x}}{s}\sqrt{\frac{n(n-1)}{n+1}}$.

We know that, $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ and $x' \sim N(\mu, \sigma^2)$

$$\Rightarrow x'-\bar{x} \sim N\left(\mu-\mu, \sigma^2+\frac{\sigma^2}{n}\right) = N\left(0, \sigma^2\left(\frac{n+1}{n}\right)\right)$$

$$\therefore \frac{x'-\bar{x}}{\sigma\sqrt{\frac{n+1}{n}}} = Z \text{ (say)} \sim N(0,1).$$

Again, $\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2 \sim \chi^2_{(n-1)} \Rightarrow \frac{s^2}{\sigma^2} \sim \chi^2_{(n-1)}$.

Also, $s^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2$ is independent of $\bar{x}$ $\Rightarrow Y = \frac{s^2}{\sigma^2}$ is independent of Z.

$$\therefore \frac{Z}{\sqrt{Y/(n-1)}} \sim t_{(n-1)} \Rightarrow \left[\frac{x'-\bar{x}}{\sigma\sqrt{\frac{n+1}{n}}}\right]\left[\frac{\sigma\sqrt{n-1}}{s}\right] \sim t_{(n-1)}$$

If $A \sim N(0,1)$ & $B \sim \chi^2_{(n)}$ & A, B are independent, then $\frac{A}{\sqrt{B/n}} \sim t_{(n)}$.

$$\Rightarrow \frac{x'-\bar{x}}{s}\sqrt{\frac{n(n-1)}{n+1}} \sim t_{(n-1)}, \text{ i.e., } \underline{\underline{U \sim t_{(n-1)}}}$$

(c) Prove that if $X \sim F_{m,n}$ and $Y \sim F_{n,m}$, then for every $a>0$, $P(X \le a) + P(Y \le \frac{1}{a}) = 1$.

Given $X \sim F_{m,n}$ and $Y \sim F_{n,m}$.

$$\therefore f_X(x) = \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}}{B\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} ; \quad 0<x<\infty$$

$$f_Y(y) = \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}} y^{\frac{n}{2}-1}}{B\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{n}{m}y\right)^{\frac{m+n}{2}}} ; \quad 0<y<\infty.$$

$$\therefore \text{L.H.S.} = P(X\le a) + P(Y \le 1/a) = \int_0^a \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}\,dx}{B\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} + \int_0^{1/a} \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}} y^{\frac{n}{2}-1}\,dy}{B\left(\frac{n}{2},\frac{m}{2}\right)\left(1+\frac{n}{m}y\right)^{\frac{n+m}{2}}} \quad \text{--- (1)}$$

In second integral, put $y = \frac{1}{x}$; $dy = -\frac{1}{x^2}dx$; $y=0 \Rightarrow x\to\infty$ & $y=\frac{1}{a} \Rightarrow x=a$.

$$\text{i.e., } P(Y \le 1/a) = \int_\infty^a \frac{\left(\frac{n}{m}\right)^{n/2}\left(\frac{1}{x}\right)^{\frac{n}{2}-1}\left(\frac{-1}{x^2}\right)}{B\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{n}{mx}\right)^{\frac{m+n}{2}}}dx$$

$$= \left(\frac{n}{m}\right)^{n/2}\left(\frac{n}{m}\right)^{m/2} \int_a^\infty \frac{\left(\frac{m}{n}\right)^{m/2}\left(\frac{mx}{n}\right)^{\frac{m+n}{2}} x^{\frac{n}{2}-1}}{B\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}} x^{\frac{n}{2}+1}}dx$$

$$\Rightarrow P\left(Y \le \frac{1}{a}\right) = \left(\frac{n}{m}\right)^{\frac{m+n}{2}} \left(\frac{m}{n}\right)^{\frac{m+n}{2}} \int_a^\infty \frac{\left(\frac{m}{n}\right)^{m/2} x^{\frac{m}{2}-1}}{B\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} \cdot dx$$

$$= \int_a^\infty \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}}{B\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} \cdot dx = P(X \ge a)$$

$\therefore$ From ①, L.H.S. $= P(X \le a) + P(Y \le a) = \displaystyle\int_0^a \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}\, dx}{B\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} + \int_a^\infty \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}}{B\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} dx$

$$= \int_0^\infty \frac{x^{\frac{m}{2}-1}\left(\frac{m}{n}\right)^{\frac{m}{2}}}{B\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} dx = P(0 \le X < \infty) = \underline{\underline{1}}$$

$$\therefore \ P(X \le a) + P\left(Y \le \frac{1}{a}\right) = 1$$

(a) What is a contingency table? Describe how the χ^2 distribution may be used to test whether the two attributes are independent.

Contingency table is a table which shows various cell frequencies of a bivariate classification of attributes.

Test for independence :-

Let A and B be two attributes divided into classes $A_1, A_2, \dots, A_m$ and $B_1, B_2, \dots, B_n$ respectively.

B / A	B_1	B_2	...	B_n	Total
A_1	O_{11}	O_{12}	...	O_{1n}	$O_{1\cdot}$
A_2	O_{21}	O_{22}	...	O_{2n}	$O_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_m	O_{m1}	O_{m2}	...	O_{mn}	$O_{m\cdot}$
Total	$O_{\cdot 1}$	$O_{\cdot 2}$	...	$O_{\cdot n}$	N

(CONTINGENCY TABLE)

O_{ij} = No. of observations posessing attributes A_i and B_j.

$P(A_i) = \frac{O_{i\cdot}}{N}$; $P(B_j) = \frac{O_{\cdot j}}{N}$

$P(A_iB_j) = P(A_i)\,P(B_j)$ if A and B are two independent attributes.

e_{ij} = Expected no. of observations posessing attributes A_i and B_j.

$= NP(A_iB_j) = \frac{O_{i\cdot}\, O_{\cdot j}}{N}$ if A_i & B_j are independent.

No. of independent classes $= (m-1)(n-1)$.

$$\therefore \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{(O_i - e_i)^2}{e_i} \sim \chi^2_{((m-1)(n-1))}$$

$\therefore$ If Cal. χ^2 < Tab. $\chi^2_{(m-1)(n-1)}(\alpha)$, then A & B are independent at α% level of sig.

6.(b) Define F-distribution. For F-dist^n with n_1, n_2 d.f. show that mean is independent of n_1 and mode lies b/w 0 and 1.

If $X \sim \chi^2_{(n_1)}$, $Y \sim \chi^2_{(n_2)}$ and X, Y are independent, then $\frac{X/n_1}{Y/n_2} \sim F_{n_1, n_2}$.

p.d.f. of F_{n_1,n_2} is $f(x) = \dfrac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(1+\frac{n_1}{n_2}x\right)^{\frac{n_1+n_2}{2}}}$; $0 \le x < \infty$. $= x'$ (say)

$$E(x') = \int_0^\infty \frac{x\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)\left(1+\frac{n_1}{n_2}x\right)^{\frac{n_1+n_2}{2}}}dx = \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)}\int_0^\infty \frac{x^{\frac{n_1}{2}}}{\left(1+\frac{n_1}{n_2}x\right)^{\frac{n_1+n_2}{2}}}dx$$

Put $\frac{n_1}{n_2}x = \theta$

$dx = \frac{n_2}{n_1}d\theta$

$$= \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \frac{1}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)} \int_0^\infty \frac{\left(\frac{n_2}{n_1}\right)^{\frac{n_1}{2}}\theta^{\frac{n_1}{2}}}{(1+\theta)^{\frac{n_1+n_2}{2}}}\frac{n_2}{n_1}d\theta$$

$$= \frac{n_2}{n_1}\cdot\frac{1}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)}\int_0^\infty \frac{\theta^{\frac{n_1}{2}+1-1}}{(1+\theta)^{\frac{n_1+n_2}{2}}}d\theta = \frac{n_2}{n_1}\cdot\frac{1}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)}\cdot B\left(\frac{n_1}{2}+1, \frac{n_2}{2}-1\right)$$

$$= \frac{n_2}{n_1}\frac{\Gamma\left(\frac{n_1}{2}+1\right)\cdot\Gamma\left(\frac{n_2}{2}-1\right)}{\Gamma\left(\cancel{\frac{n_1+n_2}{2}}\right)}\cdot\frac{\Gamma\left(\cancel{\frac{n_1+n_2}{2}}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} = \frac{n_2}{\cancel{n_1}}\frac{\cancel{n_1}}{2}\frac{1}{\left(\frac{n_2}{2}-1\right)}$$

$$= \frac{n_2}{\cancel{2}}\frac{\cancel{2}}{n_2-2} = \underline{\underline{\frac{n_2}{n_2-2}}}$$

$\therefore$ Mean of $F_{n_1,n_2} = \frac{n_2}{n_2-2}$ (independent of n_1)

To find mode of F_{n_1,n_2}:

$$\log(f(x)) = \log\left[\frac{(n_1/n_2)^{n_1/2}}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)}\right] + \left(\frac{n_1}{2}-1\right)\log x - \left(\frac{n_1+n_2}{2}\right)\log\left(1+\frac{n_1}{n_2}x\right)$$

$$\Rightarrow \frac{d(\log(f(x)))}{dx} = 0 + \left(\frac{n_1}{2}-1\right)\frac{1}{x} - \left(\frac{n_1+n_2}{2}\right)\frac{1}{\left(1+\frac{n_1}{n_2}x\right)}\cdot\frac{n_1}{n_2} = 0$$

$$\Rightarrow \left(\frac{n_1-2}{\cancel{2}}\right)\frac{1}{x} = \frac{n_1}{\cancel{n_2}}\left(\frac{n_1+n_2}{\cancel{2}}\right)\left(\frac{\cancel{n_2}}{n_2+n_1x}\right)$$

$$\Rightarrow (n_1-2)(n_2+n_1x) = n_1x(n_1+n_2)$$

$$\Rightarrow [n_1(n_1-2) - n_1(n_1+n_2)]x = n_2(2-n_1)$$

$$\Rightarrow [\cancel{n_1^2} - 2n_1 - \cancel{n_1^2} - n_1n_2]x = n_2(2-n_1)$$

$$\Rightarrow x = \frac{n_2(n_1-2)}{n_1(n_2+2)} = \left(\frac{n_1-2}{n_1}\right)\left(\frac{n_2}{n_2+2}\right)$$

$\therefore$ Mode of $F_{n_1,n_2} = \left(\frac{n_1-2}{n_1}\right)\left(\frac{n_2}{n_2+2}\right)$ but $\frac{n_1-2}{n_1} < 1$ and $\frac{n_2}{n_2+2} < 1$ & $x > 0$ for f-distn.

$\therefore$ Mode of F_{n_1,n_2} lies b/w 0 and 1.

7.(a) If $X \sim F_{m,n}$, then show that $U = \frac{mX}{n+mX} \sim \beta_I\left(\frac{m}{2}, \frac{n}{2}\right)$.

Let $V \sim \chi^2_{(m)}$, $W \sim \chi^2_{(n)}$ and V, W are independent.

$\therefore X = \frac{V/m}{W/n} \sim F_{m,n}. \Rightarrow X = \frac{n}{m} \cdot \frac{V}{W} \sim F_{m,n}.$

$$U = \frac{mX}{n+mX} = \frac{m\left(\frac{n}{m}\frac{V}{W}\right)}{n + m\left(\frac{n}{m}\frac{V}{W}\right)} = \frac{n\frac{V}{W}}{n\left(1+\frac{V}{W}\right)} = \frac{V}{V+W} \sim \beta_I\left(\frac{m}{2}, \frac{n}{2}\right)$$

$\therefore U = \frac{mX}{n+mX} \sim \beta_I\left(\frac{m}{2}, \frac{n}{2}\right)$.

$\left(\because V \sim \chi^2_{(m)}, W \sim \chi^2_{(n)} \text{ \& } V, W \text{ are ind. Then } \frac{V}{V+W} \sim \beta_I\left(\frac{m}{2}, \frac{n}{2}\right)\right)$

1.(d) Show that the sum of independent chi-square variates is also a chi-square variate.

Let $X \sim \chi^2_{(m)}$ and $Y \sim \chi^2_{(n)}$ and X, Y are independent.

Now, $M_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX}).E(e^{tY}) = (1-2t)^{-n/2}(1-2t)^{-m/2}$

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

$=$ M.G.F. of $\chi^2_{(m+n)}$.

$\therefore X+Y \sim \chi^2_{(m+n)}$, by uniqueness theorem.

$\therefore$ Sum of independent chi-square variates is also chi-square variate.

(e) If $X \sim F_{2,4}$, then show that $P(X \geq 2) = 1/4$.

$X \sim F_{2,4} \Rightarrow f_X(x) = \dfrac{\left(\frac{2}{4}\right)^{2/2} x^{2/2-1}}{B\left(\frac{2}{2}, \frac{4}{2}\right)\left(1+\frac{2}{4}x\right)^{\frac{2+4}{2}}}$; $0<x<\infty$ $= \dfrac{1}{B(1,2)\left(1+\frac{x}{2}\right)^3 . 2}$; $0<x<\infty$.

$P(X \geq 2) = \displaystyle\int_2^\infty \frac{1}{2B(1,2)\left(1+\frac{x}{2}\right)^3}dx = \frac{\Gamma(3)}{2\Gamma(1)\Gamma(2)}\int_2^\infty \left(1+\frac{x}{2}\right)^{-3}dx$

$= \dfrac{\frac{1}{2}\left[\left(1+\frac{x}{2}\right)^{-2}\right]_2^\infty}{-\frac{1}{2}} = -\left[0-\frac{1}{4}\right] = \frac{1}{4}$

$\therefore P(X \geq 2) = \frac{1}{4}$

(f) If $X \sim F_{m,n}$ and $Y \sim F_{n,m}$, then show that $P(X \leq a) + P\left(X \leq \frac{1}{a}\right) = 1 \; \forall a$.

Already solved in Q.P. 2016 (8(a)).

Section-B

5.(a) Obtain mean deviation about mean of t-distribution with n d.f.

As $f(t)$, the density function of t-distn is symmetrical about the line $t=0$, $E(t) =$ mean of t-distn $= 0$.

M.D. (mean) $= \displaystyle\int_{-\infty}^{\infty} |t-0| . \frac{1}{\sqrt{n}\,B\left(\frac{n}{2},\frac{1}{2}\right)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} .dt = \frac{2}{\sqrt{n}\,B\left(\frac{n}{2},\frac{1}{2}\right)}\int_0^\infty \frac{t\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$

Put $t^2/n = y \Rightarrow t = \sqrt{ny}$

$dt = \frac{\sqrt{n}}{2\sqrt{y}}dy$

$$= \frac{2}{\sqrt{n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_0^\infty \frac{\sqrt{ny}\cdot\sqrt{n}}{2\sqrt{y}\,(1+y)^{\frac{n+1}{2}}}\,dy = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)} B\left(1,\frac{n-1}{2}\right)\sqrt{n}$$

$$= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)} \cdot \frac{\Gamma(1)\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}\sqrt{n} = \frac{\sqrt{n}\,\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)} \qquad \left(\because \Gamma\left(\tfrac{1}{2}\right)=\sqrt{\pi}\right)$$

(b) If X is a chi-square variate with n d.f., then prove that for large 'n', $\sqrt{2X} \sim N(\sqrt{2n}, 1)$.

We know that $Z = \frac{X-n}{\sqrt{2n}} \sim N(0,1)$ as $n\to\infty$ if $X \sim \chi^2_{(n)}$.

$$F_Z(z) = P(Z \le z) = P\left(\frac{X-n}{\sqrt{2n}} \le z\right) = P(X \le n+\sqrt{2n}\,z) = P(2X \le 2n + 2z\sqrt{2n})$$

$$= P\left(\sqrt{2X} \le \sqrt{2n+2z\sqrt{2n}}\right) = P\left[\sqrt{2X} \le \sqrt{2n}\left(1+\sqrt{\frac{2}{n}}\,z\right)^{1/2}\right]$$

$$= P\left[\sqrt{2X} \le \sqrt{2n}\left(1+\frac{1}{2}\sqrt{\frac{2}{n}}\,z + \frac{\frac{1}{2}\left(\frac{1}{2}-1\right)}{2!}\left(\sqrt{\frac{2}{n}}\,z\right)^2 + O(n^{-3/2})\right)\right]$$

$$= P\left[\sqrt{2X} \le \sqrt{2n} + z + O(n^{-3/2})\right] = P(\sqrt{2X}-\sqrt{2n} \le z) \quad (\text{for large } n)$$

$\Rightarrow$ $\sqrt{2X}-\sqrt{2n} \sim N(0,1)$ for large values of n.

$\Rightarrow$ $\sqrt{2X} \sim N(\sqrt{2n},1)$ for large values of n.

Hence proved.

(c) Show that t-distribution tends to normal distribution for large n.

$$f(t) = \frac{1}{\sqrt{n}\,B\left(\frac{1}{2},\frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} \;;\; -\infty < t < \infty.$$

Now, $\lim_{n\to\infty} \frac{1}{\sqrt{n}\,B\left(\frac{1}{2},\frac{n}{2}\right)} = \lim_{n\to\infty} \frac{1}{\sqrt{n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}$ $\left(\because \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \;\&\; \lim_{n\to\infty} \frac{\Gamma(n+k)}{\Gamma(n)} \approx n^k\right)$

$$\approx \frac{1}{\sqrt{n}}\cdot\frac{1}{\sqrt{\pi}}\cdot\left(\frac{n}{2}\right)^{\frac{1}{2}} = \frac{1}{\sqrt{2\pi}}$$

$$\therefore \lim_{n\to\infty} f(t) = \lim_{n\to\infty} \frac{1}{\sqrt{n}\,B\left(\frac{1}{2},\frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} = \lim_{n\to\infty} \frac{1}{\sqrt{n}\,B\left(\frac{1}{2},\frac{n}{2}\right)} \times \lim_{n\to\infty} \frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

$$= \frac{1}{\sqrt{2\pi}} \times \frac{1}{e^{\lim_{n\to\infty}\left(\frac{n+1}{2}\right)\left(\frac{t^2}{n}\right)}} = \frac{1}{\sqrt{2\pi}}\,e^{-t^2/2} \;;\; -\infty < t < \infty. = \text{p.d.f. of } N(0,1).$$

$\therefore$ For large d.f., t-distribution tends to standard normal distⁿ.

2017-2

6.(b) If $X \sim F_{m,n}$ distribution, then obtain the distribution of mX as $n \to \infty$. Also obtain the mode of the F-distribution.

$$X \sim F_{m,n} \Rightarrow f(x) = \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}}{B\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} \; ; \; 0<x<\infty$$

$$\Rightarrow f(x) = \frac{(mx)^{m/2-1} \cdot m}{n^{m/2} B\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}}$$

Now, $$\lim_{n\to\infty} n^{m/2} B\left(\frac{m}{2},\frac{n}{2}\right) = \lim_{n\to\infty} \frac{n^{m/2}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{m+n}{2}\right)} = \lim_{n\to\infty} \frac{n^{m/2}\Gamma(m/2)}{\left(\frac{\Gamma\left(\frac{n}{2}+\frac{m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\right)}$$

$$= \lim_{n\to\infty} \frac{n^{m/2}\Gamma\left(\frac{m}{2}\right)}{(n/2)^{m/2}} \approx 2^{m/2}\Gamma\left(\frac{m}{2}\right). \quad \left(\because \lim_{n\to\infty}\frac{\Gamma(n+l)}{\Gamma(n)} \approx n^{l}\right)$$

Also $$\lim_{n\to\infty}\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}} = \lim_{n\to\infty}\left[\left(1+\frac{mx}{n}\right)^{\frac{n}{mx}}\right]^{\frac{mx}{n}\left(\frac{m+n}{2}\right)} = e^{mx/2}$$

$$\therefore \lim_{n\to\infty} d(F_{m,n}) = \lim_{n\to\infty} \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}}{B\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} \cdot dx = \frac{(mx)^{\frac{m}{2}-1}\, d(mx)}{2^{m/2}\Gamma\left(\frac{m}{2}\right) e^{mx/2}}$$

$$= \frac{e^{-mx/2}(mx)^{m/2-1}}{2^{m/2}\Gamma(m/2)}\, d(mx)$$

$$\Rightarrow \quad \underline{mX \sim \chi^2_{(m)}}$$

Mode of $F_{m,n}$ is obtained as $\underline{\left(\frac{m-2}{m}\right)\left(\frac{n}{n+2}\right)}$ in Q.No. 6(b) of 2016 Q.P.

8.(a). Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ and $\bar{x}$ and S^2 respectively be the sample mean and sample variance. Let $X_{n+1} \sim N(\mu,\sigma^2)$, and assume that $X_1, X_2, \ldots, X_n, X_{n+1}$ are independent. Obtain the distribution of:

$$U = \frac{X_{n+1}-\bar{X}}{S}\sqrt{\frac{n}{n+1}}.$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ and } X_{n+1} \sim N(\mu,\sigma^2) \Rightarrow X_{n+1}-\bar{X} \sim N\left(0, \sigma^2\left(\frac{n+1}{n}\right)\right)$$

$$\therefore \frac{X_{n+1}-\bar{X}}{\sigma\sqrt{\frac{n+1}{n}}} = Z \text{ (say)} \sim N(0,1).$$

Again, $$\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 \sim \chi^2_{(n-1)} \Rightarrow \frac{(n-1)S}{\sigma^2} \sim \chi^2_{(n-1)}.$$

Also, $S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ is independent of $\bar{X} \Rightarrow V = \frac{(n-1)S}{\sigma^2}$ is independent of Z.

$$\Rightarrow \frac{Z}{\sqrt{V/(n-1)}} \sim t_{(n-1)} \Rightarrow \left(\frac{X_{n+1}-\bar{x}}{\sigma\sqrt{\frac{n+1}{n}}}\right)\cdot\left(\frac{\sigma\sqrt{n-1}}{\sqrt{(n-1)}S}\right) \sim t_{(n-1)} \Rightarrow \frac{X_{n+1}-\bar{X}}{S}\cdot\sqrt{\frac{n}{n+1}} \sim t_{(n-1)}$$

$\therefore U \sim t_{(n-1)}$

8(b) If $X \sim F_{n_1,n_2}$, then show that its mean is independent of n_1.

Already solved in Q. 6(b) in 2016 Q.P.

6(a) For a chi-square distribution with n d.f., prove that: $\mu_{r+1} = 2r(\mu_r + n\mu_{r-1})$, $r \geq 1$.

$X \sim \chi^2_{(n)} \Rightarrow E(X) = n$.

M.G.F. of x about mean $= M(t) = E\left(e^{t(x-n)}\right) = e^{-nt}E\left(e^{tx}\right) = e^{-nt}(1-2t)^{-n/2}$

Take logarithm on both sides $\Rightarrow \log(M(t)) = -nt - \frac{n}{2}\log(1-2t)$

Differentiate both sides w.r.t. t:

$$\frac{M'(t)}{M(t)} = -n + \frac{n}{2}\left(\frac{2}{1-2t}\right) = \frac{2nt}{1-2t} \Rightarrow (1-2t)M'(t) = M(t)\, 2nt$$

Differentiating r times w.r.t. t by Liebnitz theorem, we get:

$$(1-2t)M^{r+1}(t) + r(-2)M^r(t) = 2nt\,M^r(t) + 2nr\,M^{r-1}(t)$$

Putting $t=0$ and using the relation, $\mu_r = \left[\frac{d^r}{dt^r}M(t)\right]_{t=0} = M^r(0)$, we get:

$$\mu_{r+1} - 2r\mu_r = 2nr\mu_{r-1} \Rightarrow \mu_{r+1} = 2r(\mu_r + n\mu_{r-1}),\ r \geq 1. \rightarrow (1)$$

Hence proved.

Hence find β_1 and β_2. Also discuss the limiting form of χ^2 distn.

Taking $r = 1,2,3$ in (1), we get:

$\mu_2 = 2n\mu_0 = 2n$; $\mu_3 = 4(\mu_2 + n\mu_1) = 8n$ $\quad (\because \mu_1 = 0\ \&\ \mu_0 = 1)$

$\mu_4 = 6(\mu_3 + n\mu_2) = 48n + 12n^2$.

$$\therefore \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{8}{n} \text{ and } \beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{12}{n}.$$

$M_x(t) = (1-2t)^{-n/2}$, $|t| < \frac{1}{2}$. Let $Z = \frac{X-\mu}{\sigma}$

$\therefore$ M.G.F. of standard chi-square variate $= M_{\frac{X-\mu}{\sigma}}(t) = e^{-\mu t/\sigma}\, M_X\left(\frac{t}{\sigma}\right)$

$$\Rightarrow M_Z(t) = e^{-nt/\sqrt{2n}}\left(1 - \frac{2t}{\sqrt{2n}}\right)^{-n/2} \quad (\because \mu = n \text{ and } \sigma^2 = 2n)$$

$$\therefore K_Z(t) = \log M_Z(t) = -t\sqrt{\frac{n}{2}} - \frac{n}{2}\log\left(1-\sqrt{\frac{2}{n}}\,t\right)$$

$$= -t\sqrt{\frac{n}{2}} + \frac{n}{2}\left[t\sqrt{\frac{2}{n}} + \frac{t^2}{2}\cdot\frac{2}{n} + \frac{t^3}{3}\left(\frac{2}{n}\right)^{3/2} + \cdots\right]$$

$$= -t\sqrt{\frac{n}{2}} + t\sqrt{\frac{n}{2}} + \frac{t^2}{2} + O(n^{-1/2}) = \frac{t^2}{2} + O(n^{-1/2})$$

$$\therefore \lim_{n\to\infty} K_Z(t) = \frac{t^2}{2} \Rightarrow \lim_{n\to\infty} M_Z(t) = e^{t^2/2} = \text{M.G.F. of } N(0,1).$$

$\therefore$ As $n\to\infty$, standard chi-square variate tends to standard normal variate.

7.(a) Prove that if $n_1=n_2$, the median of F distribution is at $F=1$, and the quartiles Q_1 and Q_3 satisfy the condition $Q_1 Q_3 = 1$.

Since $n_1=n_2=n$ (say), the median (M) of $F_{n_1,n_2}=F_{n,n}$ distn is given by: $P[F(n,n) < M] = 0.5 \rightarrow (1)$ $\Rightarrow P\left[\frac{1}{F(n,n)} \geq \frac{1}{M}\right] = 0.5$

$$\Rightarrow P\left[F(n,n) \geq \frac{1}{M}\right] = 0.5 \qquad \left(\because F(n,n) = \frac{1}{F(n,n)}\right)$$

$$\Rightarrow P\left[F(n,n) < \frac{1}{M}\right] = 1 - P\left[F(n,n) \geq \frac{1}{M}\right] = 0.5 \rightarrow (2)$$

From (1) and (2), $M = \frac{1}{M} \Rightarrow M^2 = 1$ (but $F>0$)

$$\Rightarrow \underline{\underline{M=1}}$$

$\therefore$ Median of $F(n,n)$ is at $F=1$.

Similarly, for Q_1 and Q_3, we have:

$P[F(n,n) < Q_1] = 0.25 \rightarrow (3)$ and $P[F(n,n) > Q_3] = 0.25$

$\Rightarrow P\left[\frac{1}{F(n,n)} < \frac{1}{Q_3}\right] = 0.25 \Rightarrow P\left[F(n,n) < \frac{1}{Q_3}\right] = 0.25 \rightarrow (4)$

From (3) and (4), $Q_1 = \frac{1}{Q_3} \Rightarrow \underline{\underline{Q_1 Q_3 = 1}}$

7(b) Discuss the t-test for testing the significance for the difference of two population means.

Suppose we want to test if two samples (independent) $x_i\ (i=1,2,\ldots,n_1)$ and $y_j\ (j=1,2,\ldots,n_2)$ of sizes n_1 and n_2 have been drawn from two normal populations with equal means μ_X and μ_Y respectively. (i.e., test if $\mu_X=\mu_Y$)

Under the null hypothesis (H_0), samples have been drawn from the populations with same means, i.e., $\mu_x = \mu_Y$, against the alternative hypothesis (H_1) that they are either (i) $\mu_x \neq \mu_Y$, (ii) $\mu_x < \mu_Y$ or (iii) $\mu_x > \mu_Y$.

Test statistic is given by:

$$t = \frac{\bar{x} - \bar{y} - E(\bar{x} - \bar{y})}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}} = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$E(\bar{x} - \bar{y}) = E(\bar{x}) - E(\bar{y}) = \mu_x - \mu_Y$

But under H_0, $\mu_x = \mu_Y$

$\therefore$ Under H_0, $E(\bar{x} - \bar{y}) = 0$

Compare this calculated t with tabulated t-distn with $(n_1 + n_2 - 2)$ degrees of freedom to take a decision i.e., make a conclusion.

Here, we make three fundamental assumptions:

1. Parent popn's from which samples have been drawn are normally distributed.
2. The popn variances are equal and unknown, i.e., $\sigma_x^2 = \sigma_Y^2 = \sigma^2$, where σ^2 is unknown.
3. The two samples are random and independent of each other.

Before applying t-test for testing equality of means, it is desirable to test the equality of popn variances by applying F-test. If variances do not come out to be equal, t-test becomes invalid and in that case Behren's 'd'-test based on fiducial intervals is used.

Scanned by CamScanner

1. (ii) If x is F-variate with 2 and n ($n \geq 2$) d.f., then show that $P(F \geq K) = \left(1+\frac{2K}{n}\right)^{-\frac{n}{2}}$.

Given $x \sim F_{2,n} \Rightarrow f(x) = \dfrac{\left(\frac{2}{n}\right)^{\frac{2}{2}} x^{\frac{2}{2}-1}}{B\left(\frac{2}{2},\frac{n}{2}\right)\left(1+\frac{2}{n}x\right)^{\frac{n+2}{2}}}$; $0 < x < \infty$

$$\therefore P(F \geq K) = \int_K^\infty \frac{\left(\frac{2}{n}\right) dx}{B\left(1,\frac{n}{2}\right)\left(1+\frac{2}{n}x\right)^{\frac{n}{2}+1}} = \left(\frac{2}{n}\right) \frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma(1)\,\Gamma\left(\frac{n}{2}\right)} \int_K^\infty \left(1+\frac{2}{n}x\right)^{-\frac{n}{2}-1} dx$$

$$= \left(\frac{2}{n}\right)\left(\frac{n}{2}\right)^1 \left[\frac{\left(1+\frac{2}{n}x\right)^{-\frac{n}{2}}}{\left(\frac{-n}{2}\right)\left(\frac{2}{n}\right)}\right]_K^\infty = -\left[0-\left(1+\frac{2}{n}K\right)^{-n/2}\right] \quad \left(\because \frac{\Gamma(n+1)}{\Gamma(n)} \approx n^1\right)$$

$$\Rightarrow P(F \geq K) = \left(1+\frac{2}{n}K\right)^{-n/2}$$

7.(a) Discuss the chi-square test of goodness of fit of a theoretical distn to an observed distn by stating conditions of validity.

If O_i $(i=1,2,3,\dots,n)$ is a set of observed (experimental) frequencies and e_i $(i=1,2,3,\dots,n)$ is the corresponding set of expected (theoretical or hypothetical) frequencies, then Karl Pearson's chi-square, is given by:

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(O_i-e_i)^2}{e}\right], \left(\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} e_i\right)$$ follows chi-square distn with $(n-1)$ degrees of freedom.

CONDITIONS FOR THE VALIDITY OF χ^2-test:

1. Sample observations should be independent.
2. Constraints on the cell frequencies, if any, should be linear.
3. N, the total frequency should be reasonably large, say greater than 50.
4. No theoretical cell frequency should be less than 5.

Accept H_0 if $\chi^2 \leq \chi^2_{(n-1)}(\alpha)$ and reject H_0 if $\chi^2 > \chi^2_{(n-1)}(\alpha)$, where $\chi^2_{(n-1)}(\alpha)$ is tabulated value of χ^2 chi-square for $(n-1)$ d.f. and level of significance α.

1.(iii) Write a note on paired-t-test.

In this case, we have the conditions:

(i) the sample sizes are equal, i.e., $n_1 = n_2 = n$ (say)

(ii) the sample observations are paired together, i.e., the pair of observations (x_i, y_i) $(i = 1, 2, 3, \ldots, n)$ corresponds to the same (i^{th}) sample unit. The problem is to ~~test~~ test if the sample means differ significantly or not.

For example, we want to test the efficiency of a particular drug, for inducing sleep. We note the readings of hours of sleep on i^{th} individual before and after the drug respectively. So, here, the same set of sample with differing observations are seen.