r.s : $X_1, X_2, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$\Rightarrow E(X_i) = \mu, \quad Var(X_i) = \sigma^2 \ \forall i$.

$$Y = \underbrace{\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2}_{\text{constructed r.v.}} \sim \chi^2_{(n)}$$

$(\bar{X})$

[n = degrees of freedom

= No. of independent variables.

= No. of variables - No. of restrictions]

$(n) \quad - (0) = n.$

Result

Let $y_1, y_2, \cdots, y_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Then

$$\frac{(n-1)s^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

sample mean

has a χ^2 Distribution with $(n-1)$ degrees of freedom (df).

s^2 = sample variance

$s^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$.

$$\chi^2 = \sum \left(\frac{y_i - \bar{y}}{\sigma} \right)^2 = \frac{1}{\sigma^2} \sum (y_i - \bar{y})^2$$

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

F- distribution

Consider 2 independent normal distributions:

$X_1 \sim N(\mu_1, \sigma_1^2)$ $\quad X_2 \sim N(\mu_2, \sigma_2^2)$ ... X_1, X_2 are independent

draw r.s of size n_1

construct $\chi^2_{(1)}$

$[df = n_1]$

draw a r.s of size n_2

construct $\chi^2_{(2)}$ .... independent

$[df = n_2]$.

$\chi^2_{(1)}$ and $\chi^2_{(2)}$ are independent chi-sq variates with n_1 & n_2 d.f respectively.

constructed r.v

$$F = \frac{\chi^2_{(1)}/n_1}{\chi^2_{(2)}/n_2} \sim F_{n_1, n_2}$$

e.g 9

In e.g 6, the ounces of fill from the bottling machine are assumed to have a normal distribution with $\sigma^2 = 1$. Suppose that we plan to select a random sample ten bottles and measure the amount of fill in each bottle. If these ten observations are used to calculate s^2, it might be useful to specify an interval of values that will include s^2 with a high probability. Find numbers b_1 and b_2 such that

$$P(b_1 \le s^2 \le b_2) = 0.90.$$

$n = 10$, $\quad X_1, X_2, \ldots, X_{10} \overset{iid}{\sim} N(\mu, 1)$.

$s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$ --- sample variance.

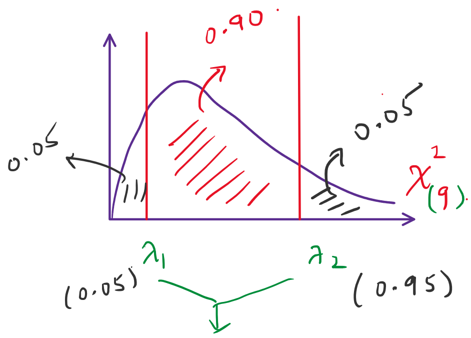
90% C.I for s^2 : $P[b_1 \le s^2 \le b_2] = 0.90$.

$$\Rightarrow P\left[b_1 \le \frac{1}{n-1}\sum(x_i-\bar{x})^2 \le b_2\right] = 0.90.$$

$$\Rightarrow P\left[(n-1)b_1 \le \sum(x_i-\bar{x})^2 \le (n-1)b_2\right] = 0.90.$$

$\chi^2_{(n-1)}$

Note: $\frac{\sum(x_i-\bar{x})^2}{\sigma^2} \sim \chi^2_{(n-1)}$, $n=10$



In this question: $\sum(x_i-\bar{x})^2 \sim \chi^2_{(9)}$

$\lambda_1 = (9)\,b_1 \Rightarrow b_1 = \lambda_1/9$

$\lambda_2 = (9)\,b_2 \Rightarrow b_2 = \lambda_2/9$.

obtained from R / χ^2 tables.

e.g 10

If we take independent samples of size $n_1 = 6$ and $n_2 = 10$ from two normal populations with equal variances, find the number b such that

$$P\left(\frac{s_1^2}{s_2^2} \le b\right) = 0.95.$$

2 independent normal poplns:

$X_1 \sim N(\mu_1, \sigma^2)$ $\qquad X_2 \sim N(\mu_2, \sigma^2)$.

$\left[\begin{array}{l} X_1 \sim N(\mu_1, \sigma^2) \\ \hookrightarrow \text{r.s of size } n_1 = 6 \end{array}\right.$ $\left[\begin{array}{l} X_2 \sim N(\mu_2, \sigma^2) \\ \hookrightarrow \text{r.s of size } n_2 = 10 \end{array}\right.$

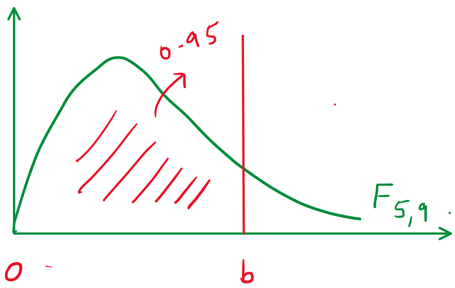
s_1^2 = sample var. for popln 1 $= \frac{1}{5}\sum_{i=1}^{6}(x_{1i}-\bar{x}_1)^2$.

s_2^2 = sample var for popln 2 $= \frac{1}{9}\sum_{j=1}^{10}(x_{2j}-\bar{x}_2)^2$.

& $\dfrac{\sum_{i=1}^{6}(x_{1i}-\bar{x}_1)^2}{\sigma^2} \sim \chi^2_{(5)}$ & $\dfrac{\sum_{j=1}^{10}(x_{2j}-\bar{x}_2)^2}{\sigma^2} \sim \chi^2_{(9)}$.

$\therefore F = \dfrac{\sum_{i=1}^{6}(x_{1i}-\bar{x}_1)^2/5}{\sum_{j=1}^{10}(x_{2j}-\bar{x}_2)^2/9} \sim F_{5,9}$.

$= \dfrac{s_1^2}{s_2^2} \sim F_{5,9}$.



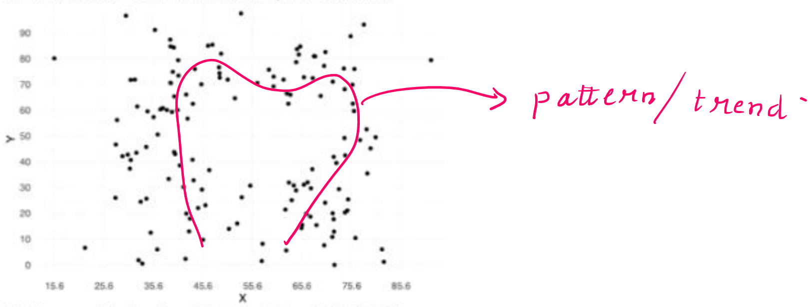
↓ computed from R/ F-Tables.

F-variate

To find: $P\left[\dfrac{s_1^2}{s_2^2} \leq b\right] = 0.95$.

Part c)

Below you can see a scatter plot of 142 realizations of (X, Y). The correlation coefficient of X and Y is 0. In other words, they are not correlated.



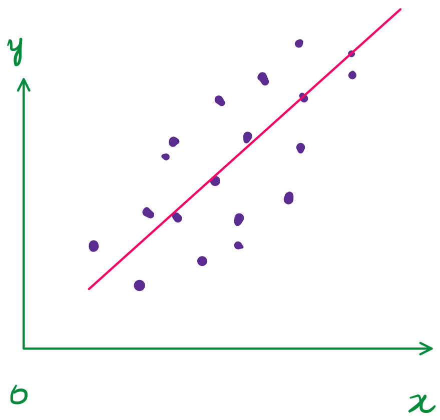
Would you consider these 2 variables to be independent? Explain in one or two sentences.

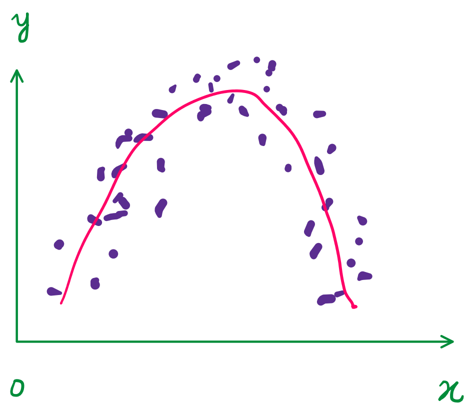
Solution

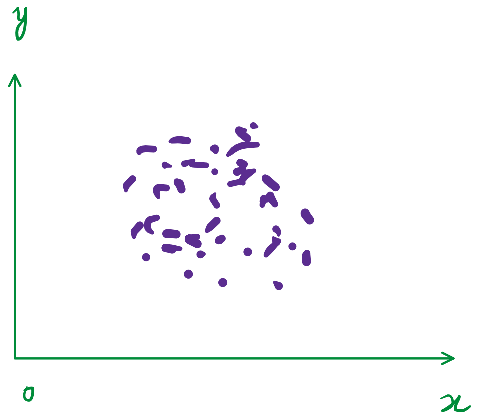
No. While they are uncorrelated, there is a large gap in the centre of the plot. It suggests that X is unlikely to fall between 45 and 65 when Y is between 40 and 60.

Uncorrelated vs Independent

y y y







Linear- Relationship b/w X & Y.
(r is high)

Non-linear-Relationship b/w X & Y.
(r is lower).

No visible pattern by X & Y.
($r \approx 0$) ⇒ uncorrelated
seems to be independent

Correlation: measure of association
⇒ [measure of linear relationships].

In General: Independence ⇒ uncorrelated.

i.e if X, Y are independent, they are obviously uncorrelated.

But uncorrelated $\not\Rightarrow$ Independence [Not necessarily]

Suppose that $Y_1, Y_2, \cdots, Y_n$ constitute a random sample from the density function

$$f(y|\theta) = \begin{cases} e^{-(y-\theta)}, & y \geq \theta, \\ 0, & \text{elsewhere}, \end{cases}$$

where θ is an unknown, positive constant.

(a) Find an estimator $\hat{\theta}_1$ for θ by the method of moments. $m_1' = \mu_1'$

(b) Find an estimator $\hat{\theta}_2$ for θ by the method of maximum likelihood.

(c) Adjust $\hat{\theta}_1$ and $\hat{\theta}_2$ so that they are unbiased. Find the efficiency of the adjusted $\hat{\theta}_1$ relative to the adjusted $\hat{\theta}_2$.

r.s: $Y_1, Y_2, \ldots, Y_n$.

$$m_1' = \frac{1}{n}\sum Y_i = \bar{Y}$$

$$\mu_1' = \int^{\infty} y \cdot f(y)\, dy = \int^{\infty} y\, e^{-(y-\theta)}\, dy$$ Let $y - \theta = t$

$$\mu_1' = \int_\theta^\infty y \cdot f(y)\, dy = \int_\theta^\infty y\, e^{-(y-\theta)}\, dy$$

Let $y - \theta = t$

$dy = dt$

$\Rightarrow y = \theta,\ t = 0$

$\Rightarrow y \to \infty,\ t \to \infty$.

$$= \int_0^\infty (t+\theta)\, e^{-t}\, dt$$

$$= \int_0^\infty t\, e^{-t}\, dt + \theta \underbrace{\int_0^\infty e^{-t}\, dt}_{=1}$$

$$= \underbrace{\int_0^\infty t\, e^{-t}\, dt}_{=1} + \theta = (1+\theta)$$

$$\int_0^\infty e^{-t}\, dt = -\left[e^{-t}\right]_0^\infty$$

$$= -\left[\underbrace{e^{-\infty}}_{0} - \underbrace{e^{-0}}_{1}\right]$$

$$= -(-1) = 1.$$

$$\int t e^{-t}\, dt = t \int e^{-t} dt - \int -e^{-t} dt$$

$$= -t e^{-t} + \int e^{-t}\, dt$$

$$= -t e^{-t} - e^{-t}$$

$$\int_0^\infty t e^{-t}\, dt = -\left[e^{-t}(t+1)\right]_0^\infty.$$

$$= -\left[e^{-\infty}(\cdot) - e^{-0}(0+1)\right]$$

$$= -(-1) = 1$$

$\therefore$ According to MOM: $\mu_1' = m_1'$

$$\Rightarrow 1 + \theta = \bar{Y}$$

$$\Rightarrow \hat{\theta}_{MOM} = (\bar{Y} - 1) = \hat{\theta}_1.$$

(b) $f(y|\theta) = e^{-(y-\theta)}$, $y \geq \theta$.

r.s: $Y_1, Y_2, \ldots, Y_n$.

$$f(y_i|\theta) = e^{-(y_i - \theta)},\ \forall i,\ y_i \geq \theta.$$

Likelihood fn: $L(\theta | y_1, \ldots, y_n) = \prod_{i=1}^n f(y_i|\theta)$

$$= \prod_{i=1}^n e^{-y_i} \cdot e^{\theta} = e^{n\theta} e^{-\sum y_i}$$

$$= e^{n\theta - \sum y_i}$$

Log-likelihood fn: $\ell(\theta) = n\theta - \sum y_i$ .... [linear in θ].

For MLE: $\frac{\partial \ell}{\partial \theta} = 0 \Rightarrow n - 0 = 0 \Rightarrow \boxed{n = 0}$ Incorrect!

$\ell(\theta) = n\theta - \sum y_i$ .. [Obj: To find that value of θ that max $\ell(\theta)$]

$f(y|\theta) = e^{-(y-\theta)}$, $y \geq \theta$.

$f(y|\theta) = e^{-(y-\theta)}, \ y \geq \theta.$

r.s: $Y_1, Y_2, \cdots, Y_n$.

$\Rightarrow Y_1 \geq \theta, \ Y_2 \geq \theta, \cdots, Y_n \geq \theta$.

ordered r.s: $\theta \leq Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$.

relationship b/w parameter θ & sample obs.

Put $\theta = Y_{(1)}$: $\ell_{(1)}(\theta) = nY_{(1)} - \sum y_i$

$\theta = Y_{(2)}$: $\ell_{(2)}(\theta) = nY_{(2)} - \sum y_i$

$\vdots$

$\boxed{\theta = Y_{(n)}}$: $\ell_{(n)}(\theta) = nY_{(n)} - \sum y_i$ .... max possible value of likelihood fn.

$\therefore \hat{\theta}_{MLE} = Y_{(n)} = \hat{\theta}_2$.

e.g 4

Let $Y_1, Y_2, \cdots, Y_n$ be a random sample of observations from a uniform distribution with probability density function $f(y_i|\theta) = 1|\theta$, for $0 \leq y_i \leq \theta$ and $i = 1, 2, \cdots, n$. Find the MLE of θ.

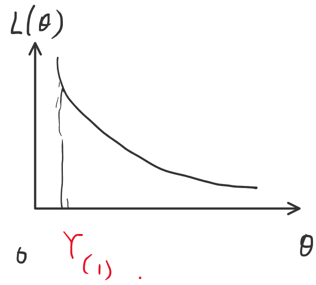
$f(y) = \frac{1}{\theta}, \ 0 \leq y \leq \theta$.

r.s: $Y_1, Y_2, \cdots, Y_n$.

$f(y_i|\theta) = \frac{1}{\theta}, \ 0 \leq y_i \leq \theta, \ \forall i$

Likelihood fn: $L(\theta) = \prod_{i=1}^{n} f(y_i|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} = \frac{1}{\theta^n}$.

Log-likelihood fn: $\ell(\theta) = -n \ln \theta$.



For MLE: $\frac{\partial \ell}{\partial \theta} = 0 \Rightarrow -\frac{n}{\theta} = 0 \Rightarrow \boxed{n = 0}$ ✗.

Now, $0 \leq y_i \leq \theta$.

$0 \quad Y_{(1)} \quad\quad \theta$

Now, $0 \leq y_i \leq \theta$.

Ordered r.s: $\boxed{0 \leq Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)} \leq \theta}$

$\theta = Y_{(1)} \Rightarrow L_{(1)}(\theta) = \frac{1}{Y_{(1)}^n} \Rightarrow \max. \Rightarrow \hat{\theta}_{MLE} = Y_{(1)}$.

$\theta = Y_{(2)} \Rightarrow L_{(2)}(\theta) = \frac{1}{Y_{(2)}^n}$.

$\vdots$

$\theta = Y_{(n)} \Rightarrow L_{(n)}(\theta) = \frac{1}{Y_{(n)}^n}$.

$\hat{\theta}_1 = \bar{Y} - 1$

$f(y) = e^{-(y-\theta)}, \; y \geq \theta$.

$E(\hat{\theta}_1) = \frac{1}{n} \sum E(Y_i) - 1$.

$E(Y) = (\theta + 1)$.

$= \frac{1}{n} \sum (\theta+1) - 1$.

$= \frac{\cancel{n}(\theta+1)}{\cancel{n}} - 1 = \theta + \cancel{1} - \cancel{1} = \theta$.

$\hat{\theta}_2 = Y_{(n)} = \max\{Y_1, Y_2, \ldots, Y_n\}$.

$E(\hat{\theta}_2) = \int_\theta^\infty y_{(n)} \cdot \boxed{f(y_{(n)})} \, dy$. $\rightarrow$ pdf of $y_{(n)}$

$f(y_n) = n[F(y)]^{n-1} \cdot f(y)$.

$F(y) = P[Y \leq y] = \int_\theta^y e^{-(y-\theta)} \, d\theta$.

$y - \theta = t$

$dy = d\theta$.

$y = \theta, \; t = 0$

$y = y, \; t = y$

$= \int_0^y e^{-t} \, dt$

$= -[e^{-t}]_0^y$

$= -(e^{-y} - e^0) = -(e^{-y} - 1)$

$$= -\left(e^{-y} - e^{0}\right) = -\left(e^{-y} - 1\right)$$
$$= \left(1 - e^{-y}\right)$$

$$f(y_n) = n\left[1 - e^{-y}\right]^{(n-1)} e^{-(y-\theta)}, \quad y \geq \theta.$$

$$E(\hat{\theta}_2) = \int_{\theta}^{\infty} y \cdot n\left[1 - e^{-y}\right]^{(n-1)} e^{-(y-\theta)}\, dy.$$

$$= n \cdot e^{\theta} \int_{\theta}^{\infty} y\left[1 - e^{-y}\right]^{(n-1)} e^{-y}\, dy.$$