Proposition 3.2.1. Suppose that $Z_1, Z_2, \ldots, Z_n$ are i.i.d. random variables each with a $\mathrm{N}(0,1)$ distribution, and write $\mathbf{Z} = (Z_1, \ldots, Z_n)^T$. Suppose that $\mathbf{A}$ is an orthogonal $n \times n$ matrix, and define $\mathbf{Y} = \mathbf{AZ}$, with $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$. Then the $Y_1, Y_2, \ldots, Y_n$ are also i.i.d. random variables each with a $\mathrm{N}(0,1)$ distribution, and furthermore $\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n Z_i^2$. ♦

A is orthogonal: $AA^T = I = A^T A$

$$\sum_{i=1}^n Y_i^2 = Y^T Y = (Z^T A^T)(AZ) = Z^T (A^T A) Z = Z^T (I) Z = Z^T Z = \sum_{i=1}^n Z_i^2$$

$Y = AZ$

$Y^T = Z^T A^T$

$\det(A) = ?$

$\det(I) = 1$

$$1 = \det(I) = \det(AA^T) = \det(A)\det(A^T) = \det(A)\det(A)$$

$$\Rightarrow 1 = (\det(A))^2$$

$\Rightarrow \det(A) = \pm 1$ or $|\det(A)| = 1$

Recall $Z_1, Z_2, \ldots, Z_n$ are iid standard normal r.v.s

Joint p.d.f. for vector z.

$$f(z) = \prod_{i=1}^{n} f(z_i)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} z_i^2\right]$$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} z_i^2\right]$$

A is orthogonal and invertible

$$Y = Az$$

$$\Rightarrow z = A^{-1} Y$$

$$g(y) = \frac{1}{|\det(A)|} f(A^{-1} y)$$

$$= \frac{1}{1} f(z)$$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} z_i^2\right)$$

$$g(y) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} y_i^2\right)$$

$\Rightarrow Y_1, Y_2, \ldots, Y_n$ are $N(0,1)$

Theorem 3.2.2. Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. random variables distributed according to $\mathrm{N}(\mu, \sigma^2)$, with $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$. Then $\overline{X}$ and S^2 are independent random variables and

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}. \tag{3.4}$$

♦

Define $Z_i = \dfrac{X_i - \mu}{\sigma}$ for $i = 1, 2, \ldots, n$

$\Rightarrow Z_i \sim N(0,1)$

Choose orthogonal A with first row equal to $u = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)$

Construction: Start with I_n
replace first row with
(not orthogonal)
Then use Gram-Schmidt orthogonalisation procedure.

Have $\underbrace{Z_1, \ldots, Z_n}_{\overset{iid}{\sim} N(0,1)}$, A - orthogonal

Define $Y = AZ$ $\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{pmatrix} \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$

$\Rightarrow \quad Y_1, \ldots, Y_n$ are iid $N(0,1)$

AND $\quad \sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} Z_i^2$

AND $\quad Y_1 = u Z = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} Z_i$

$$\left(\bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i \right)$$

$$\Rightarrow Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i = \underbrace{\frac{n}{\sqrt{n}}}_{\sqrt{n}} \underbrace{\left(\frac{1}{n} \sum_{i=1}^{n} Z_i \right)}_{\bar{Z}}$$

$$\Rightarrow Y_1 = \sqrt{n}\, \bar{Z}$$

Let's look at sample variance

$$\sum_{i=1}^{n} (Z_i - \bar{Z})^2 \underset{\text{Exercise}}{=} \sum_{i=1}^{n} Z_i^2 - n(\bar{Z})^2$$

$$= \sum_{i=1}^{n} Y_i^2 - (Y_1)^2$$

$$= \sum_{i=2}^{n} Y_i^2$$

All Y_i are independent

All Y_i^2 are independent

AND so $\sum_{i=2}^{n} Y_i^2$ is independent of Y_1^2

(and Y_1)

$\Rightarrow \sum_{i=1}^{n} (Z_i - \bar{Z})^2$ is independent

of $\sqrt{n}\,\bar{Z}$

(and so of $\bar{Z}$)

$\Rightarrow \sum_{i=1}^{n} (X_i - \bar{X})^2$ is independent

of $\bar{X}$

$\Rightarrow$ S^2 is independent of $\bar{X}$

The distribution of S^2:

$$\sum_{i=2}^{n} Y_i^2 = \sum_{i=1}^{n} (Z_i - \bar{Z})^2$$

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

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

$$\Rightarrow \sum_{i=2}^{n} Y_i^2 \sim \chi^2_{n-1}$$

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

Random variable

X

Statistic, eg. $\bar{X}$

EstimatOR $\bar{X}$

confidence interval

$[L(X), U(X)]$

Probability

$$P(\theta \in [L(X), U(X)]) = 1-\alpha$$

realisation

x

Statistic, eg. $\bar{x}$

estimatE $\bar{x}$

confidence interval

$[L(x), U(x)]$

Confidence

$\theta \overset{?}{\in} (167, 178)$

with ~~probability~~

with confidence

Day 1 : $(167, 178)$ $n=10$ 95%

Day 2 : $(165, 181)$ 95% $n=10$

Day 3 $(164, 177)$ 95% $n=10$

θ has a fixed, true, but unknown value

(FREQUENTIST)