Normal samples: Distribution Theory

Theorem: Suppose $X_1, \ldots, X_n$ are independent $N(\mu, \sigma^2)$ random variables. Then

1. $\bar{X}$ (sample mean)and s^2 (sample variance) independent.

2. $n^{1/2}(\bar{X} - \mu)/\sigma \sim N(0, 1)$.

3. $(n - 1)s^2/\sigma^2 \sim \chi^2_{n-1}$.

4. $n^{1/2}(\bar{X} - \mu)/s \sim t_{n-1}$.

Proof: Let $Z_i = (X_i - \mu)/\sigma$.

Then $Z_1, \ldots, Z_p$ are independent $N(0,1)$.

So $Z = (Z_1, \ldots, Z_p)^t$ is multivariate standard normal.

Note that $\bar{X} = \sigma\bar{Z}+\mu$ and $s^2 = \sum(X_i-\bar{X})^2/(n-1) = \sigma^2\sum(Z_i - \bar{Z})^2/(n-1)$ Thus

$$\frac{n^{1/2}(\bar{X} - \mu)}{\sigma} = n^{1/2}\bar{Z}$$

$$\frac{(n-1)s^2}{\sigma^2} = \sum(Z_i - \bar{Z})^2$$

and

$$T = \frac{n^{1/2}(\bar{X} - \mu)}{s} = \frac{n^{1/2}\bar{Z}}{s_Z}$$

where $(n-1)s_Z^2 = \sum(Z_i - \bar{Z})^2$.

So: reduced to $\mu = 0$ and $\sigma = 1$.

Step 1: Define

$$Y = (\sqrt{n}\bar{Z}, Z_1 - \bar{Z}, \ldots, Z_{n-1} - \bar{Z})^t \,.$$

(So Y has same dimension as Z.) Now

$$Y = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}$$

or letting M denote the matrix

$$Y = MZ \,.$$

It follows that $Y \sim MVN(0, MM^t)$ so we need to compute MM^t:

$$MM^t = \left[\begin{array}{c|cccc} 1 & 0 & 0 & \cdots & 0 \\ \hline 0 & 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & -\frac{1}{n} & \ddots & \cdots & -\frac{1}{n} \\ 0 & \vdots & \cdots & & 1-\frac{1}{n} \end{array}\right]$$

$$= \left[\begin{array}{c|c} 1 & 0 \\ \hline & \\ 0 & Q \end{array}\right] .$$

Solve for Z from Y: $Z_i = n^{-1/2}Y_1 + Y_{i+1}$ for $1 \le i \le n-1$. Use the identity

$$\sum_{i=1}^{n}(Z_i - \bar{Z}) = 0$$

to get $Z_n = -\sum_{i=2}^{n} Y_i + n^{-1/2}Y_1$. So M invertible:

$$\Sigma^{-1} \equiv (MM^t)^{-1} = \left[\begin{array}{c|c} 1 & 0 \\ \hline & \\ 0 & Q^{-1} \end{array}\right].$$

Use change of variables to find f_Y. Let $\mathbf{y}_2$ denote vector whose entries are $y_2, \ldots, y_n$. Note that

$$y^t\Sigma^{-1}y = y_1^2 + \mathbf{y}_2^t Q^{-1}\mathbf{y}_2\,.$$

Then

$$\begin{aligned} f_Y(y) =& (2\pi)^{-n/2}\exp[-y^t\Sigma^{-1}y/2]/|\det M| \\ =& \frac{1}{\sqrt{2\pi}}e^{-y_1^2/2}\times \\ & \frac{(2\pi)^{-(n-1)/2}\exp[-\mathbf{y}_2^t Q^{-1}\mathbf{y}_2/2]}{|\det M|}. \end{aligned}$$

Note: f_Y is ftn of y_1 times a ftn of $y_2, \ldots, y_n$.

Thus $\sqrt{n}\bar{Z}$ is independent of $Z_1 - \bar{Z}, \ldots, Z_{n-1} - \bar{Z}$.

Since s_Z^2 is a function of $Z_1 - \bar{Z}, \ldots, Z_{n-1} - \bar{Z}$ we see that $\sqrt{n}\bar{Z}$ and s_Z^2 are independent.

Also, density of Y_1 is a multiple of the function of y_1 in the factorization above. But factor is standard normal density so $\sqrt{n}\bar{Z} \sim N(0,1)$.

First 2 parts done. Third part is a homework exercise.

Derivation of the χ^2 density:

Suppose $Z_1, \ldots, Z_n$ are independent $N(0,1)$. Define χ^2_n distribution to be that of $U = Z_1^2 + \cdots + Z_n^2$. Define angles $\theta_1, \ldots, \theta_{n-1}$ by

$$\begin{aligned}
Z_1 &= U^{1/2} \cos\theta_1 \\
Z_2 &= U^{1/2} \sin\theta_1 \cos\theta_2 \\
\vdots &= \vdots \\
Z_{n-1} &= U^{1/2} \sin\theta_1 \cdots \sin\theta_{n-2} \cos\theta_{n-1} \\
Z_n &= U^{1/2} \sin\theta_1 \cdots \sin\theta_{n-1} \,.
\end{aligned}$$

(Spherical co-ordinates in n dimensions. The θ values run from 0 to π except last θ from 0 to 2π.) Derivative formulas:

$$\frac{\partial Z_i}{\partial U} = \frac{1}{2U} Z_i$$

and

$$\frac{\partial Z_i}{\partial \theta_j} = \begin{cases} 0 & j > i \\ -Z_i \tan\theta_i & j = i \\ Z_i \cot\theta_j & j < i \,. \end{cases}$$

Fix $n = 3$ to clarify the formulas. Use shorthand $R = \sqrt{U}$

Matrix of partial derivatives is

$$\begin{bmatrix} \frac{\cos\theta_1}{2R} & -R\sin\theta_1 & 0 \\ \frac{\sin\theta_1\cos\theta_2}{2R} & R\cos\theta_1\cos\theta_2 & -R\sin\theta_1\sin\theta_2 \\ \frac{\sin\theta_1\sin\theta_2}{2R} & R\cos\theta_1\sin\theta_2 & R\sin\theta_1\cos\theta_2 \end{bmatrix}.$$

Find determinant by adding $2U^{1/2}\cos\theta_j/\sin\theta_j$ times col 1 to col $j+1$ (no change in determinant).

Resulting matrix lower triangular; diagonal entries:

$$\frac{\cos\theta_1}{R}, \frac{R\cos\theta_2}{\cos\theta_1}, \frac{R\sin\theta_1}{\cos\theta_2}$$

Multiply these together to get

$$U^{1/2}\sin(\theta_1)/2$$

(non-negative for all U and θ_1).

General n: every term in the first column contains a factor $U^{-1/2}/2$ while every other entry has a factor $U^{1/2}$.

FACT: multiplying a column in a matrix by c multiplies the determinant by c.

SO: Jacobian of transformation is

$$u^{(n-1)/2}u^{-1/2}/2 \times h(\theta_1, \theta_{n-1})$$

for some function, h, which depends only on the angles.

Thus joint density of $U, \theta_1, \ldots \theta_{n-1}$ is

$$(2\pi)^{-n/2} \exp(-u/2) u^{(n-2)/2} h(\theta_1, \cdots, \theta_{n-1})/2\,.$$

To compute the density of U we must do an $n-1$ dimensional multiple integral $d\theta_{n-1} \cdots d\theta_1$.

Answer has the form

$$cu^{(n-2)/2} \exp(-u/2)$$

for some c.

Evaluate c by making

$$\begin{aligned}\int f_U(u)du &= c\int_0^\infty u^{(n-2)/2}\exp(-u/2)du \\ &= 1.\end{aligned}$$

Substitute $y = u/2$, $du = 2dy$ to see that

$$\begin{aligned}c2^{n/2}\int_0^\infty y^{(n-2)/2}e^{-y}dy &= c2^{n/2}\Gamma(n/2) \\ &= 1.\end{aligned}$$

CONCLUSION: the χ_n^2 density is

$$\frac{1}{2\Gamma(n/2)}\left(\frac{u}{2}\right)^{(n-2)/2} e^{-u/2}1(u>0)\,.$$

Fourth part: consequence of first 3 parts and def'n of t_ν distribution.

Defn: $T \sim t_\nu$ if T has same distribution as

$$Z/\sqrt{U/\nu}$$

for $Z \sim N(0,1)$, $U \sim \chi^2_\nu$ and Z, U independent.

Derive density of T in this definition:

$$\begin{aligned}P(T \le t) &= P(Z \le t\sqrt{U/\nu}) \\ &= \int_0^\infty \int_{-\infty}^{t\sqrt{u/\nu}} f_Z(z) f_U(u)\, dz du\end{aligned}$$

Differentiate wrt t by differentiating inner integral:

$$\frac{\partial}{\partial t}\int_{at}^{bt} f(x)dx = bf(bt) - af(at)$$

by fundamental thm of calculus. Hence

$$\frac{d}{dt}P(T \le t) = \int_0^\infty \frac{f_U(u)}{\sqrt{2\pi}} \left(\frac{u}{\nu}\right)^{1/2} \exp\left(-\frac{t^2 u}{2\nu}\right) du\,.$$

Plug in

$$f_U(u) = \frac{1}{2\Gamma(\nu/2)}(u/2)^{(\nu-2)/2}e^{-u/2}$$

to get

$$f_T(t) = \frac{\int_0^\infty (u/2)^{(\nu-1)/2}e^{-u(1+t^2/\nu)/2}du}{2\sqrt{\pi\nu}\Gamma(\nu/2)}.$$

Substitute $y = u(1+t^2/\nu)/2$, to get

$$dy = (1+t^2/\nu)du/2$$

$$(u/2)^{(\nu-1)/2} = [y/(1+t^2/\nu)]^{(\nu-1)/2}$$

leading to

$$f_T(t) = \frac{(1+t^2/\nu)^{-(\nu+1)/2}}{\sqrt{\pi\nu}\Gamma(\nu/2)}\int_0^\infty y^{(\nu-1)/2}e^{-y}dy$$

or

$$f_T(t) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\Gamma(\nu/2)}\frac{1}{(1+t^2/\nu)^{(\nu+1)/2}}.$$