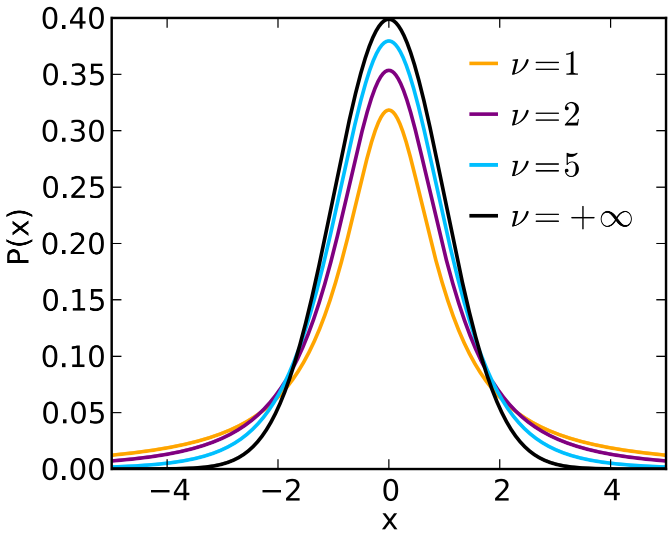
Student-t distribution

Yuanye Liu
2022.11

Introducion

Student-t distribution is a heavy-tailed distribution alternative to normal distribution, which has more patience for the outliers, with more robustness. Not like Gaussian distribution that much sensitve.



When the degree of freedom goes to infinity, the student-t distribution is a Gaussian. The parameter ν reflects the robustness of the distribution.

Owing to the features of student-t, we can use T-test to calculate the confidence interval of μ with unknown standard variance in Gaussian distribution.

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
$$\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

Generation of Student-t distribution

For the unitary situation

Suppose we have a Chi-square random variable, with degree of freedom ν. $Y \sim \chi_\nu^2$,

$$f_Y(y) = \frac{y^{\frac{\nu}{2}-1}e^{-\frac{y}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}$$

where $\Gamma(z)$ is the Gamma function, which is,

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$$

And we have a standard distribution, $Z \sim N(0,1)$, then we can construct a student-t distribution,

$$X = \frac{Z}{\sqrt{Y/\nu}} \sim t_\nu$$

For the multivariate student-t

Similarily, we need a Chi-square random variable, with degree of freedom ν, $Y \sim \chi_\nu^2$, and we have a multivariate Gaussian distribution, $Z \sim N(0, \Sigma)$. where Σ is a $p \times p$ positive matrix.

So the multivariate student-t is,

$$X = \frac{Z}{\sqrt{Y/\nu}} + \mu \sim t_p(\mu, \Sigma, \nu)$$

And here I want to introduce a perspective to understand t distribution. From the eqution above, we can find out that, $\frac{Z}{\sqrt{Y/\nu}} \sim t_p(0, \Sigma, \nu)$. Meanwhile, if we regard the $u = Y/\nu$ as a constant or say fixed number, it's just a modification of covariance matrix in a $N(0, \Sigma)$, which is to say, if we fix u, we have,

$$X|u \sim N(\mu, \Sigma/u)$$

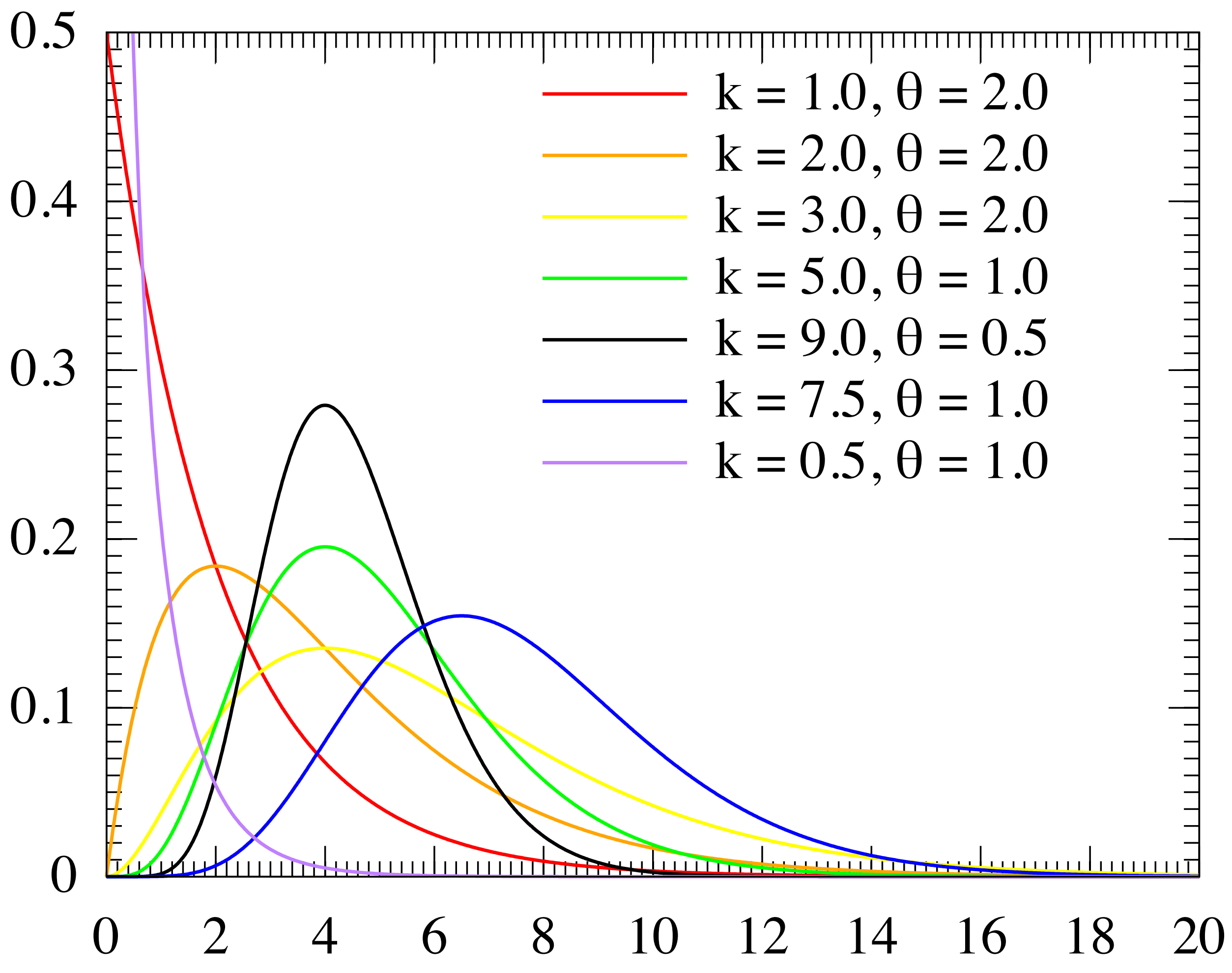
Student t distribution is a gaussian distribution with a **stochastic covariance** matrix.

And, the $u = Y/\nu$ is a Gamma distribution, satisfying $u \sim \text{Gamma}(\frac{\nu}{2}, \frac{\nu}{2})$. also called scaled chi-square distribution.

For $x \sim \text{Gamma}(\alpha, \beta)$,

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}e^{-\beta x}\beta^\alpha}{\Gamma(\alpha)} \quad \text{for } x > 0 \quad \alpha, \beta > 0$$

where $\Gamma(\alpha)$ is the Gamma function.



So when ν goes to infinity, the u goes to 1, so the t distribution goes to a Gaussion $N(\mu, \Sigma)$.

Sampling

$$u_i|\nu \overset{iid}{\sim} \text{Gamma}(\nu/2, \nu/2)$$
$$X_i|\mu, \Sigma, u_i \overset{ind}{\sim} N(\mu, \Sigma/u_i)$$

Probability Density Function(PDF)

Unitary

For the unitary situation, I use the differential method to calculate the density,

- $P(x < X < x + \mathrm{d}x) = f(x)\mathrm{d}x$
- $P(X = x) = f(x)\mathrm{d}x$
- $\mathrm{d}x = \sqrt{\frac{y}{\nu}}\mathrm{d}t$

And we have,

$$
\begin{aligned}
f_T(t)\mathrm{d}t &= P(T=t) \\
&= P(X = t\sqrt{Y/n}) \\
&= f_X(t\sqrt{Y/n})\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}}\exp\{-\frac{t^2Y}{2n}\}\mathrm{d}x \\
f_T(t) &= \frac{1}{\sqrt{2\pi}}\exp\{-\frac{t^2Y}{2n}\}\sqrt{Y/n} \\
&= E\{E\{\frac{1}{\sqrt{2\pi}}\exp\{-\frac{t^2Y}{2n}\}\sqrt{Y/n}|Y=y\}\} \\
&= E\{E\{\frac{1}{\sqrt{2\pi}}\exp\{-\frac{t^2y}{2n}\}\sqrt{y/n}\}\} \\
&= E\{\int_0^{+\infty} f_Y(y)\frac{1}{\sqrt{2\pi}}\exp\{-\frac{t^2y}{2n}\}\sqrt{y/n}\mathrm{d}y\} \\
&= \int_0^{+\infty} f_Y(y)\frac{1}{\sqrt{2\pi}}\exp\{-\frac{t^2y}{2n}\}\sqrt{y/n}\,\mathrm{d}y \\
&= \int_0^{+\infty} \frac{y^{\frac{n}{2}-1}e^{-\frac{y}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\frac{1}{\sqrt{2\pi}}\exp\{-\frac{t^2y}{2n}\}\sqrt{y/n}\,\mathrm{d}y \\
&= \frac{1}{\sqrt{2n\pi}}\frac{1}{2^{n/2}\Gamma(\frac{n}{2})}\int_0^{+\infty} y^{\frac{n-1}{2}}\exp\{-(\frac{1}{2}+\frac{t^2}{2n})y\}\mathrm{d}y \\
\overset{u=(\frac{1}{2}+\frac{t^2}{2n})y}{\rightarrow} &= \frac{1}{\sqrt{2n\pi}}\frac{1}{2^{n/2}\Gamma(\frac{n}{2})}(\frac{1}{2}(1+\frac{t^2}{n}))^{-\frac{n+1}{2}}\int_0^{+\infty} u^{\frac{n+1}{2}-1}\exp\{-u\}\mathrm{d}u \\
&= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}
\end{aligned}
$$

Which is,

$$
f_T(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}(1+\frac{t^2}{n})^{-\frac{\nu+1}{2}}a
$$

Mulitivariate case

For the multivariate case, it's similar,

$$
f_T(t) = \frac{\Gamma\{(\nu+p)/2\}}{\Gamma(\nu/2)(\nu\pi)^{p/2}|\Sigma|^{1/2}}\{1+\frac{1}{\nu}(t-\mu)^\top\Sigma^{-1}(t-\mu)\}^{-(\nu+p)/2}
$$

where $\delta_T(\mu,\Sigma) = (t-\mu)^\top\Sigma^{-1}(t-\mu)$ is the Mahalanobis distance from t to the center μ with respect to Σ.

Likelihood of μ, Σ with observed ν

The likelihood of (μ, Σ, ν) is

$$
\begin{aligned}
L(\mu,\Sigma,\nu|Y,\tau) &= f(Y|\mu,\Sigma,\tau)f(\tau|\nu) \\
&= \prod_{i=1}^{n}\left((\frac{\nu}{2})^{\nu/2}\tau_i^{\nu/2-1}\frac{\exp\{-\frac{\nu}{2}\tau_i\}}{\Gamma(\frac{\nu}{2})}\right)\left(\frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma/\tau_i|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(y_i-\mu)^\top(\Sigma/\tau_i)^{-1}(y_i-\mu)\}\right)
\end{aligned}
$$

the log-likihood is,

$$
\begin{aligned}
l(\mu,\Sigma,\nu|Y,\tau) &= \sum_{i=1}^{n}\frac{\nu}{2}\log\frac{\nu}{2}+(\frac{\nu}{2}-1)\log\tau_i-\frac{\nu}{2}\tau_i-\log\Gamma(\frac{\nu}{2}) \\
&\quad -\frac{p}{2}\log 2\pi-\frac{1}{2}\log|\Sigma/\tau_i|-\frac{1}{2}(y_i-\mu)^\top(\Sigma/\tau_i)^{-1}(y_i-\mu) \\
&= n[\frac{\nu}{2}\log\frac{\nu}{2}-\log\Gamma(\frac{\nu}{2})-\frac{p}{2}\log 2\pi-\frac{1}{2}\log|\Sigma|] \\
&\quad +\sum_{i=1}^{n}(\frac{\nu+p}{2}-1)\log\tau_i-\frac{\nu}{2}\tau_i-\frac{1}{2}\tau_i(y_i^\top\Sigma^{-1}y_i-2\mu^\top\Sigma^{-1}y_i+\mu^\top\Sigma^{-1}\mu) \\
&= -\frac{np}{2}\log(2\pi)+\sum_{i=1}^{n}(\frac{p}{2}-1)\log\tau_i \qquad (Const) \\
&\quad +\frac{n\nu}{2}\log\frac{\nu}{2}-n\log\Gamma(\frac{\nu}{2})+\frac{\nu}{2}\sum_{i=1}^{n}(\log\tau_i-\tau_i) \qquad (\nu) \\
&\quad -\frac{n}{2}\log|\Sigma|-\frac{1}{2}tr\{\Sigma^{-1}\sum_{i=1}^{n}\tau_i y_i y_i^\top\}+\mu^\top\Sigma^{-1}\sum_{i=1}^{n}\tau_i y_i-\frac{1}{2}\mu^\top\Sigma^{-1}\mu\sum_{i=1}^{n}\tau_i \qquad (\mu,\Sigma)
\end{aligned}
$$

Student-t Mixture Model

Notation

index of observations: $i = 1, \cdots, K$

index of components $j = 1, \cdots, N$

$\{x_j\}$: the observations.

$\{z_j\}$: the component-label vector, $z_{ij} = 1$ means the j_{th} observation belongs to the i_{th} componet.

$\{u_j\}$: missing data, is the parameter for the j_{th} observation.

$\{\mu_i\}$: the center of the i_{th} component.

$\{\Sigma_i\}$: the covariance of the i_{th} component.

$\{\nu_i\}$: the degree of freedom of the i_{th} component.

Complete log-likelihood

Complete data vector

$$x_c = [x_1, \cdots, x_N, z_1, \cdots, z_N, u_1, \cdots, u_N]$$

Complete likelihood:

$$\begin{aligned}
L_c(\pi, \nu, \mu, \Sigma; x, z, u) &= P(x, z, u; \pi, \nu, \mu, \Sigma) \\
&= P(z; \pi)P(u|z; \nu)P(x|z, u; \nu, \mu, \Sigma) \\
&= \prod_{j=1}^{N}\prod_{k=1}^{K} P(z_{kj}; \pi_k)P(u_j|z_{kj}; \nu_k)P(x_j|z_{kj}, u_j; \nu_k, \mu_k, \Sigma_k) \\
&= \prod_{j=1}^{N}\prod_{k=1}^{K} \pi_k^{z_{kj}} \,\mathrm{Gamma}(u_j; \frac{\nu_k}{2}, \frac{\nu_k}{2})^{z_{kj}} N(x_j; \mu_k, \frac{\Sigma_k}{u_j})^{z_{kj}} \\
&= \prod_{j=1}^{N}\prod_{k=1}^{K} \left\{\pi_k \,\mathrm{Gamma}(u_j; \frac{\nu_k}{2}, \frac{\nu_k}{2}) N(x_j; \mu_k, \frac{\Sigma_k}{u_j})\right\}^{z_{kj}}
\end{aligned}$$

Log-likelihood:

$$\begin{aligned}
l_c(\pi, \nu, \mu, \Sigma; x, z, u) &= \sum_{j=1}^{N}\sum_{k=1}^{K} z_{kj}(\log \pi_k + \log Gamma(u_j; \frac{\nu_k}{2}, \frac{\nu_k}{2}) + \log N(x_j; \mu_k, \frac{\Sigma_k}{u_j})) \\
&= \sum_{j=1}^{N}\sum_{k=1}^{K} z_{kj}[\log \pi_k + \frac{\nu_k}{2}\log(\frac{\nu_k}{2}) + (\frac{\nu_k}{2} - 1)\log u_j - \frac{\nu_k}{2}u_j - \log\Gamma(\frac{\nu_k}{2}) \\
&\quad - \frac{p}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_k/u_j| - \frac{1}{2}(y_j - \mu_k)^\top(\Sigma_k/u_j)^{-1}(y_j - \mu_k)]
\end{aligned}$$

$$\begin{aligned}
l_1(\pi) &= \sum_{j=1}^{N}\sum_{k=1}^{K} z_{kj}\log \pi_k \\
l_2(\nu) &= \sum_{j=1}^{N}\sum_{k=1}^{K} z_{kj}(\frac{\nu_k}{2}\log(\frac{\nu_k}{2}) + (\frac{\nu_k}{2} - 1)\log u_j - \frac{\nu_k}{2}u_j - \log\Gamma(\frac{\nu_k}{2})) \\
&= \sum_{j=1}^{N}\sum_{k=1}^{K} z_{kj}(\frac{\nu_k}{2}\log(\frac{\nu_k}{2}) + \frac{\nu_k}{2}(\log u_j - u_j) - \log u_j - \log\Gamma(\frac{\nu_k}{2})) \\
l_3(\mu, \Sigma) &= \sum_{j=1}^{N}\sum_{k=1}^{K} z_{kj}(-\frac{p}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_k| - \frac{1}{2}u_j(x_j - \mu_k)^\top\Sigma_k^{-1}(x_j - \mu_k))
\end{aligned}$$

Parameter: $\Psi = \{\pi, \theta, \nu\} = \{\pi_j, \theta_j, \nu_j\}_{j=1}^{K}$.

EM

E-step.

Compute the $Q(\Psi;\Psi^{(t)}) = E\{l_c(\pi,\nu,\mu,\Sigma;x,z,u)\}$. We should compute the $E(Z_{ij})$ and $E(U_j)$.

At the t_{th} stage, for $E(Z_{ij})$,

$$\begin{aligned}
E(Z_{ij}|x_j;\Psi^{(t)}) &= P(z_{ij}=1|x_j;\Psi^{(t)}) \\
&= \frac{P(z_{ij}=1,x_j;\Psi^{(t)})}{P(x_j;\Psi^{(t)})} \\
&= \frac{\pi_i^t f(x_j;\mu_i^t,\Sigma_i^t,\nu_i^t)}{\sum_l \pi_l^t f(x_j;\mu_l^t,\Sigma_l^t,\nu_l^t)} \\
&:= \tau_{ij}^t
\end{aligned}$$

for the $E(U_j|X_j)$, consider the U_j, we already have,

$$\begin{aligned}
P(U_j|X_j,z_{ij}) &= \frac{P(U_j|z_{ij})P(X_j|U_j,z_{ij})}{P(X_j|z_{ij})} \\
&= \frac{\frac{(\frac{\nu_i}{2})^{\frac{\nu_i}{2}}U_j^{\frac{\nu_i}{2}-1}\exp\{-\frac{\nu_i}{2}U_j\}}{\Gamma(\frac{\nu_i}{2})}\frac{\exp\{-\frac{1}{2}(X_j-\mu_i)^\top(\Sigma_i/U_j)^{-1}(X_j-\mu_i)\}}{(2\pi)^{p/2}|\Sigma_i/U_j|^{1/2}}}{\frac{\Gamma\{(\nu_i+p)/2\}\{1+\frac{1}{\nu_i}(X_j-\mu_i)^\top(\Sigma_i)^{-1}(X_j-\mu_i)\}^{-(\nu_i+p)/2}}{\Gamma(\frac{\nu_i}{2})(\nu_i\pi)^{p/2}|\Sigma_i|^{1/2}}} \\
&= \frac{\frac{(\frac{\nu_i}{2})^{\frac{\nu_i}{2}}U_j^{\frac{\nu_i}{2}-1}\exp\{-\frac{\nu_i}{2}U_j\}}{\Gamma(\frac{\nu_i}{2})}\frac{\exp\{-\frac{U_j}{2}\delta_{X_j}(\mu_i,\Sigma_i)\}}{(2\pi)^{p/2}|\Sigma_i/U_j|^{1/2}}}{\frac{\Gamma\{(\nu_i+p)/2\}\{1+\frac{1}{\nu_i}\delta_{X_j}(\mu_i,\Sigma_i)\}^{-(\nu_i+p)/2}}{\Gamma(\frac{\nu_i}{2})(\nu_i\pi)^{p/2}|\Sigma_i|^{1/2}}} \\
&= \frac{(\frac{\nu_i}{2})^{\frac{\nu_i}{2}}U_j^{\frac{\nu_i}{2}-1}\exp\{-\frac{\nu_i}{2}U_j\}\frac{\exp\{-\frac{U_j}{2}\delta_{X_j}(\mu_i,\Sigma_i)\}}{(2\pi)^{p/2}|\Sigma_i/U_j|^{1/2}}}{\frac{\Gamma\{(\nu_i+p)/2\}\{1+\frac{1}{\nu_i}\delta_{X_j}(\mu_i,\Sigma_i)\}^{-(\nu_i+p)/2}}{(\nu_i\pi)^{p/2}|\Sigma_i|^{1/2}}} \\
&= \frac{(\frac{\nu_i}{2})^{\frac{\nu_i+p}{2}}U_j^{\frac{\nu_i+p}{2}-1}\exp\{-\frac{\nu_i+\delta_{X_j}(\mu_i,\Sigma_i)}{2}U_j\}}{\Gamma\{(\nu_i+p)/2\}\{1+\frac{1}{\nu_i}\delta_{X_j}(\mu_i,\Sigma_i)\}^{-(\nu_i+p)/2}} \\
&= \frac{(\frac{\nu_i}{2})^{\frac{\nu_i+p}{2}}\{1+\frac{1}{\nu_i}\delta_{X_j}(\mu_i,\Sigma_i)\}^{(\nu_i+p)/2}U_j^{\frac{\nu_i+p}{2}-1}\exp\{-\frac{\nu_i+\delta_{X_j}(\mu_i,\Sigma_i)}{2}U_j\}}{\Gamma\{(\nu_i+p)/2\}} \\
&= \frac{(\frac{\nu_i+\delta_{X_j}(\mu_i,\Sigma_i)}{2})^{(\nu_i+p)/2}U_j^{(\nu_i+p)/2-1}\exp\{-\frac{\nu_i+\delta_{X_j}(\mu_i,\Sigma_i)}{2}U_j\}}{\Gamma\{(\nu_i+p)/2\}} \\
&= Gamma(U_j;\frac{\nu_i+p}{2},\frac{\nu_i+\delta_{X_j}(\mu_i,\Sigma_i)}{2})
\end{aligned}$$

So, the expectation of U_j,

$$E(U_j|X_j, z_{ij}; \Psi^{(t)}) = \frac{\nu_i^{(t)} + p}{\nu_i^{(t)} + \delta_{X_j}(\mu_i^{(t)}, \Sigma_i^{(t)})} := u_{ij}^{(t)}$$

for the $E(\log U)$, we have a theorem, if a random variable $X \sim \text{Gamma}(\alpha, \beta)$, the expectation of $\log X$ is,

$$E(\log X) = \psi(\alpha) - \log(\beta)$$

where the ψ is the DiGamma function.

So,

$$\begin{aligned}
E(\log U_j|X_j, z_{ij=1}; \Psi^{(t)}) &= \psi(\frac{\nu_i^{(t)} + p}{2}) - \log(\frac{\nu_i^{(t)} + \delta_{X_j}(\mu_i^{(t)}, \Sigma_i^{(t)})}{2}) \\
&= \psi(\frac{\nu_i^{(t)} + p}{2}) + \log(\frac{\nu_i^{(t)} + p}{\nu_i^{(t)} + \delta_{X_j}(\mu_i^{(t)}, \Sigma_i^{(t)})}) - \log(\frac{\nu_i^{(t)} + p}{2}) \\
&= \log u_{ij}^{(t)} + \{\psi(\frac{\nu_i^{(t)} + p}{2}) - \log(\frac{\nu_i^{(t)} + p}{2})\}
\end{aligned}$$

The last term can be interpreted as the correction for just inputing the conditional mean value $u_{ij}^{(k)}$ for U_j in $\log U_j$.

So the $Q(\Psi; \Psi^{(k)})$ can be given by,

$$Q(\Psi; \Psi^{(t)}) = Q_1(\pi; \Psi^{(t)}) + Q_2(\nu; \Psi^{(t)}) + Q_3(\theta; \Psi^{(t)})$$

Where,

$$\begin{aligned}
Q_1(\pi; \Psi^{(t)}) &= \sum_{j=1}^{N} \sum_{i=1}^{K} \tau_{ij}^{(t)} \log \pi_i \\
Q_2(\nu; \Psi^{(t)}) &= \sum_{j=1}^{N} \sum_{i=1}^{K} \tau_{ij}^{(t)} Q_{2j}(\nu_i; \Psi^{(t)}) \\
Q_3(\theta; \Psi^{(t)}) &= \sum_{j=1}^{N} \sum_{i=1}^{K} \tau_{ij}^{(t)} Q_{3j}(\theta_i; \Psi^{(t)})
\end{aligned}$$

where, on ignoring terms not involving the parameters we are concerned.

$$\begin{aligned}
Q_{2j}(\nu_i; \Psi^{(t)}) &= -\log \Gamma(\frac{\nu_k}{2}) + \frac{\nu_k}{2} \log(\frac{\nu_k}{2}) + \frac{\nu_k}{2}\{(\log u_{ij}^{(t)} - u_j^{(t)}) + \psi(\frac{\nu_i^{(t)} + p}{2}) - \log(\frac{\nu_i^{(t)} + p}{2})\} \\
Q_{3j}(\theta_i; \Psi^{(t)}) &= -\frac{p}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma_i| - \frac{1}{2} u_{ij}^{(t)} (x_j - \mu_k)^\top \Sigma_i^{-1} (x_j - \mu_k)
\end{aligned}$$

M-step

We compute the π, μ, Σ in a closed form,

$$\pi_i^{(t+1)} = \sum_{j=1}^{N} \frac{\tau_{ij}^{(t)}}{N}$$

$$\mu_i^{(t+1)} = \frac{\sum_{j=1}^{N} \tau_{ij}^{(t)} u_{ij}^{(t)} x_j}{\sum_{j=1}^{N} \tau_{ij}^{(t)} u_{ij}^{(t)}}$$

$$\Sigma_i^{(t+1)} = \frac{\sum_{j=1}^{N} \tau_{ij}^{(t)} u_{ij}^{(t)} (y_j - \mu_i^{(k+1)})(y_j - \mu_i^{(k+1)})^\top}{\sum_{j=1}^{N} \tau_{ij}^{(t)}}$$

for the $\nu_i^{(t)}$, it's the solution of the equation below,

$$-\psi(\frac{1}{2}\nu_i) + \log(\frac{1}{2}\nu_i) + 1 + \frac{1}{n_i^{(t)}} \sum_{j}^{N} (\log u_{ij}^{(t)} - u_{ij}^{(t)}) + \psi(\frac{\nu_i^{(t)} + p}{2}) - \log \frac{\nu_i^{(t)} + p}{2} = 0$$

it can compute by **one-dimensional search**, such as the half-interval method.