MULTIVARIATE PROBABILITY DISTRIBUTIONS IN R

Multivariate t-distributions

Surajit Ray
Reader, University of Glasgow

Parameters for multivariate distributions

Distribution	Location Parameter	Scale Parameter
Normal	`mean`	`sigma`
t	`delta`	`sigma`
Skew-normal	`xi`	`Omega`
Skew-t	`xi`	`Omega`

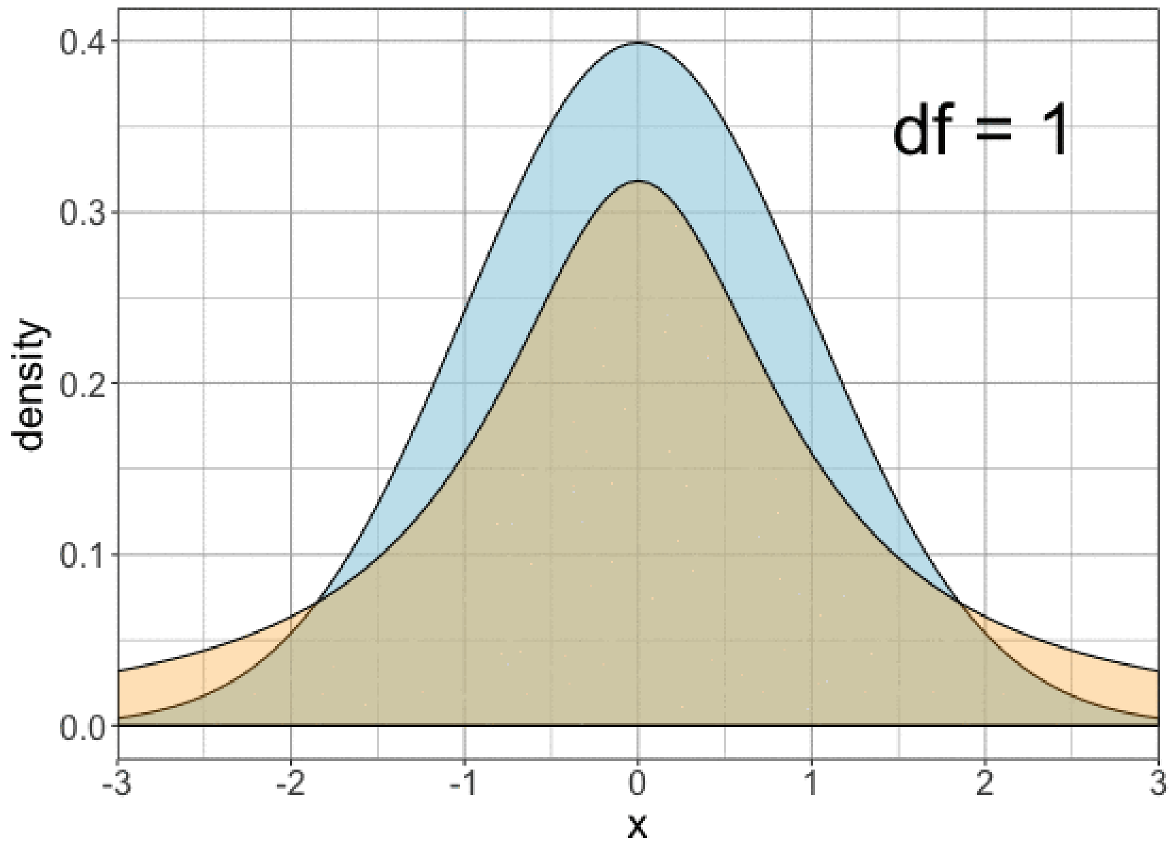
Parameters for multivariate distributions

Distribution	Location Parameter	Scale Parameter	Degrees of freedom
Normal	`mean`	`sigma`	No
t	`delta`	`sigma`	Yes
Skew-normal	`xi`	`Omega`	No
Skew-t	`xi`	`Omega`	Yes

Comparing univariate normal with univariate t-distributions

Comparision

- Standard normal
- t with different `df`'s

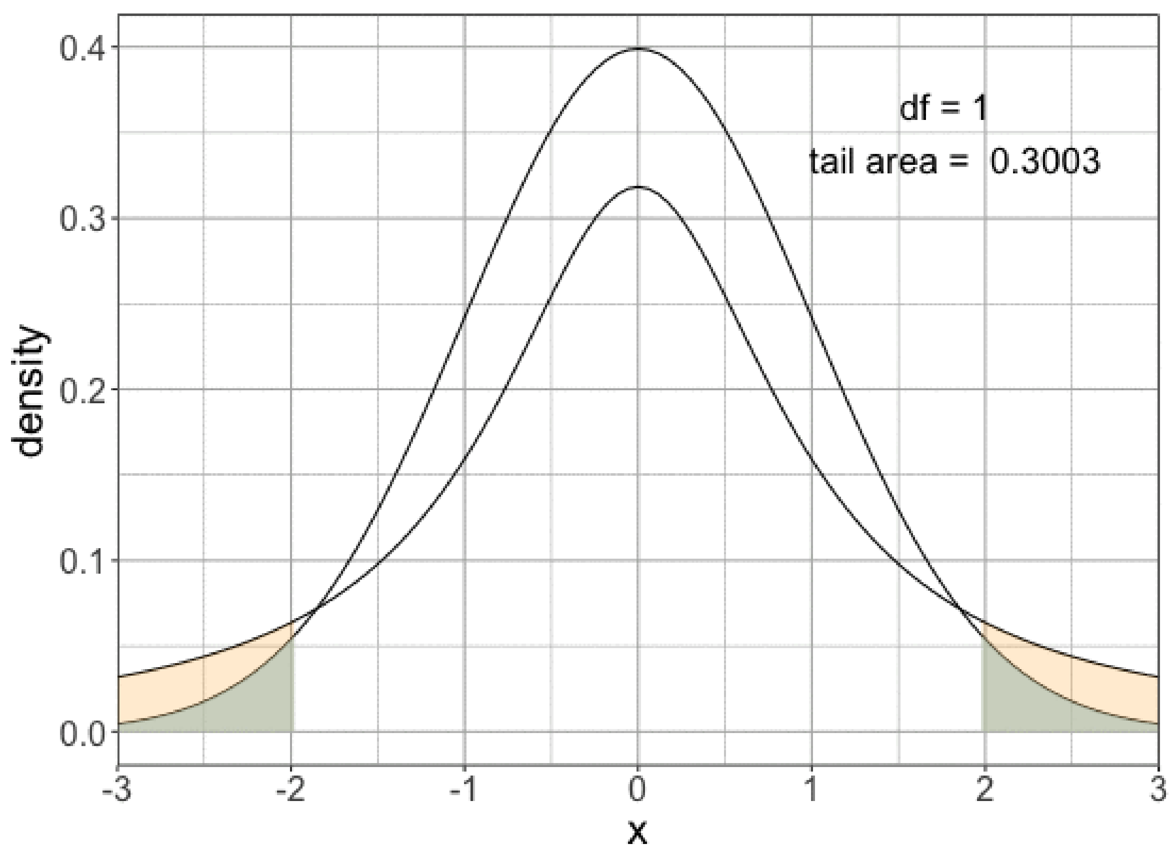


Comparing normal and t-distribution tails

Tails are fatter for the same cutoff

$$P(X < -1.96 \text{ or } X > 1.96)$$

Distribution	Probability
Normal	0.05
t(df=1)	0.3
t(df=8)	0.0857
t(df=20)	0.0641
t(df=30)	0.0593



Multivariate t-distribution notation

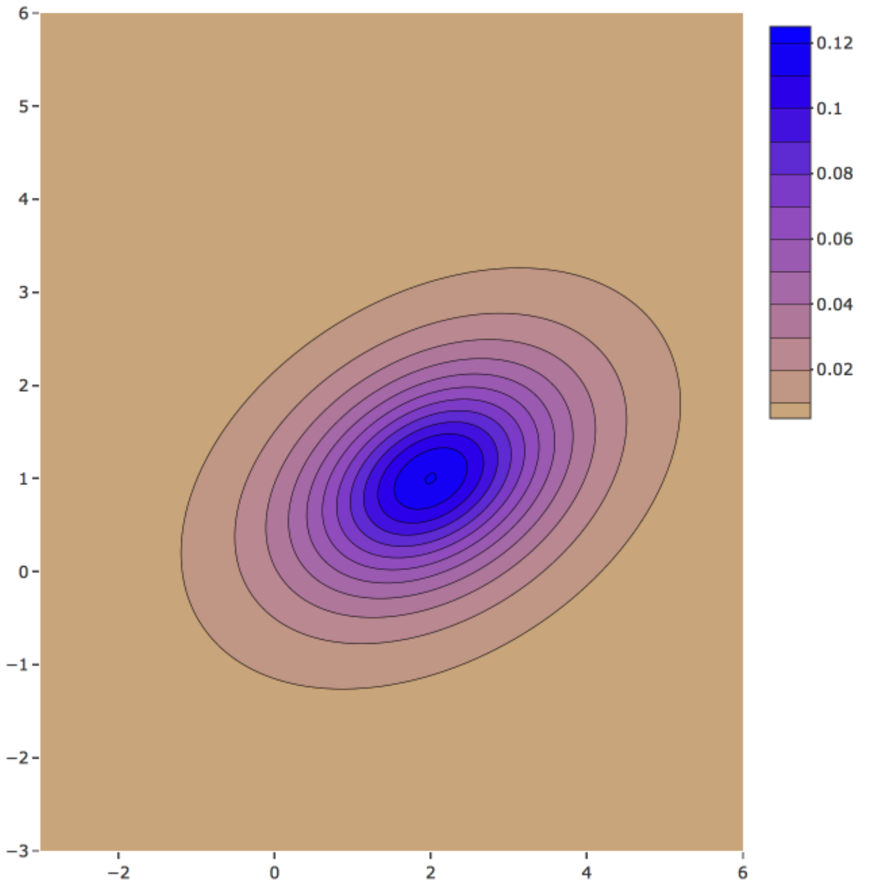
- Generalization of the univariate Student's t-distribution
  - Widely used version has only one degree of freedom for all dimensions and is denoted by

$$t_{df}(\delta, \Sigma)$$

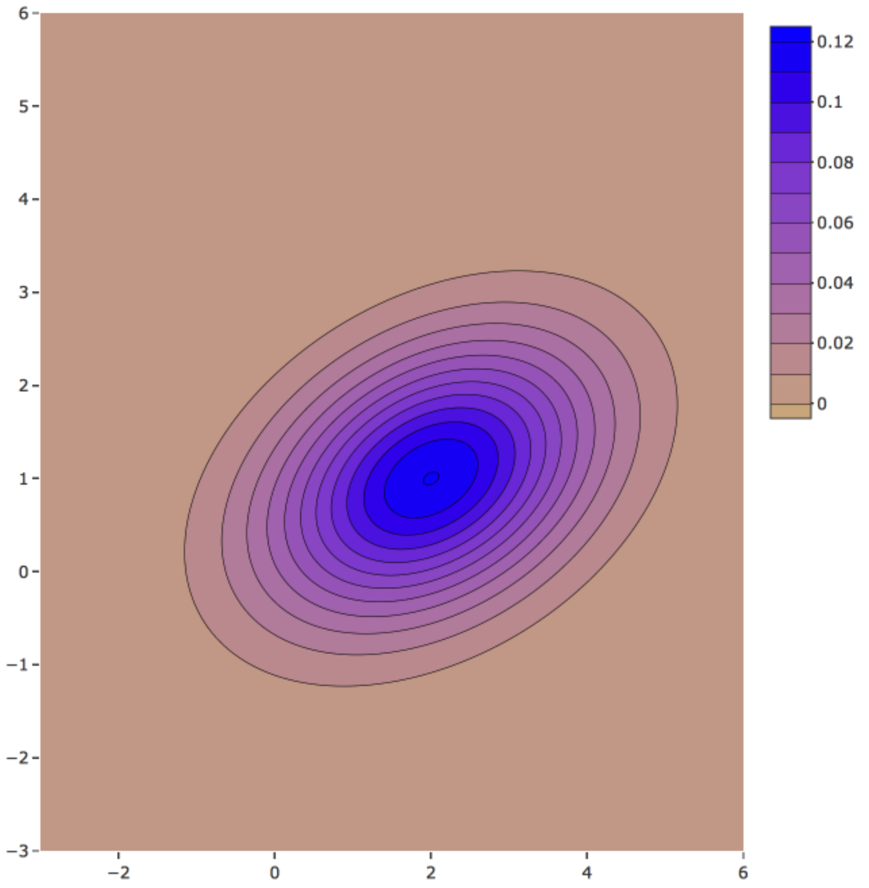
Contours of bivariate normal and t-distributions

$$\mu = \delta = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix}$$

Contours of a t with `df = 3`



Contours of a bivariate normal



Functions for multivariate t-distributions

- Functions include:
  - `rmvt(n, delta, sigma, df)`
  - `dmvt(x, delta, sigma, df)`
  - `qmvt(p, delta, sigma, df)`
  - `pmvt(upper, lower, delta, sigma, df)`

Generating random samples

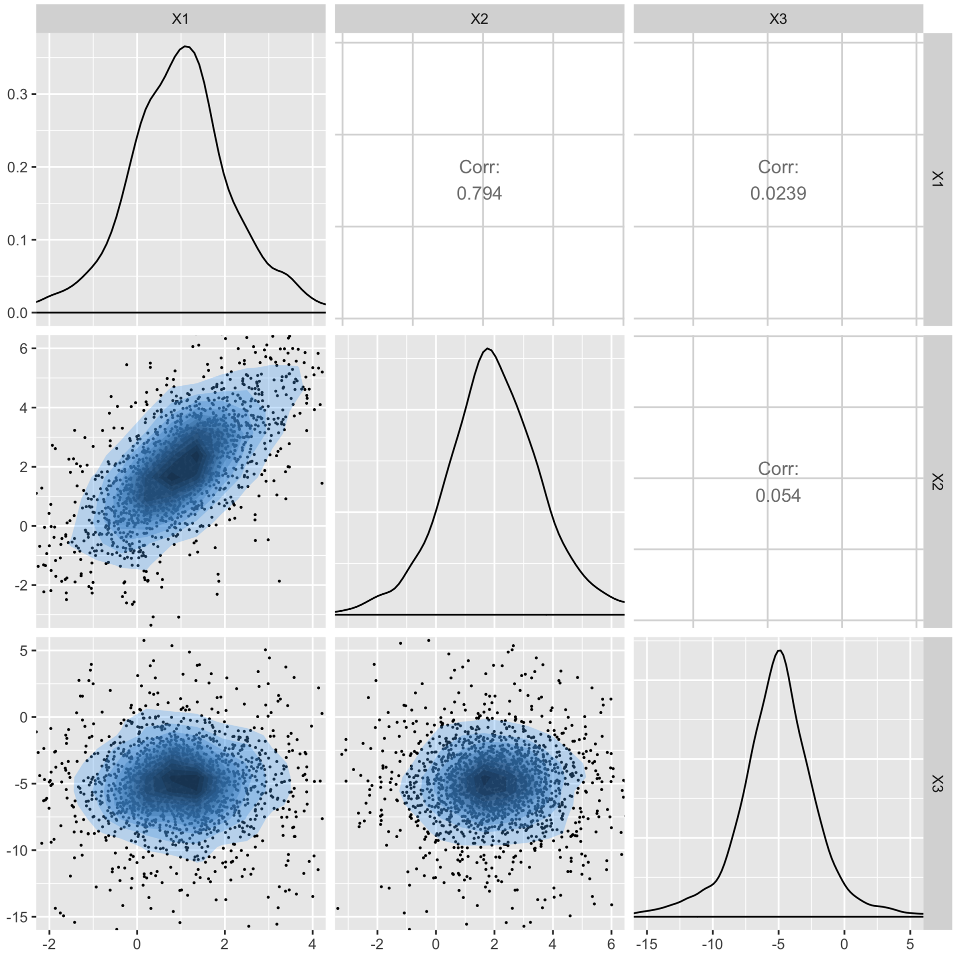
Generate samples from 3 dimensional t with $\delta = \begin{pmatrix} 1 \\ 2 \\ -5 \end{pmatrix}, \ \Sigma = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 5 \end{pmatrix}, \ df = 4.$

```
# Specify delta and sigma
delta <- c(1, 2, -5)
sigma <- matrix(c(1, 1, 0,
                  1, 2, 0,
                  0, 0, 5), 3, 3)

# Generate samples
t.sample <- rmvt(n = 2000, delta = delta, sigma = sigma, df = 4)
```

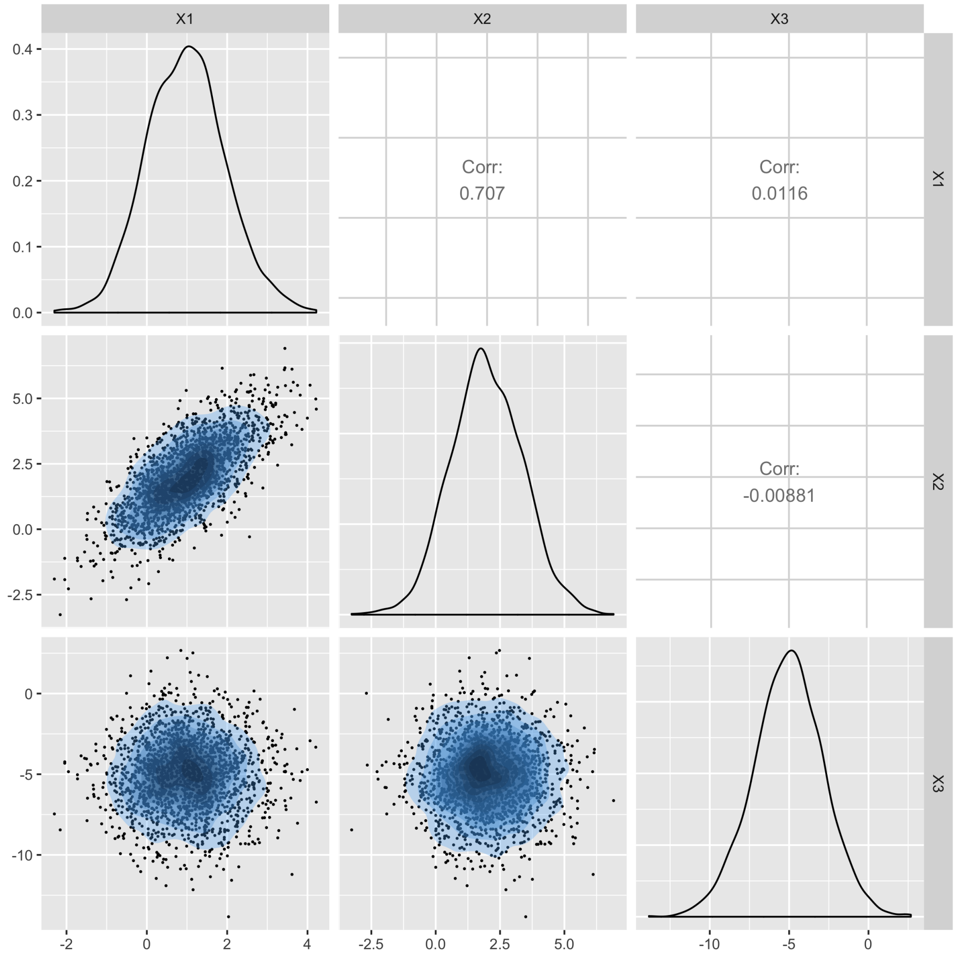
```
head(t.sample)

       [,1]   [,2]    [,3]
[1,] -1.256 -1.518 -12.340
[2,]  1.479  1.908  -7.647
[3,] -0.152  1.357  -9.011
[4,]  1.938  2.531  -4.534
[5,] -1.019 -2.371  -0.794
[6,]  0.832  0.336  -7.625
```

Comparing with normal samples

t-distribution with 4 degrees of freedom

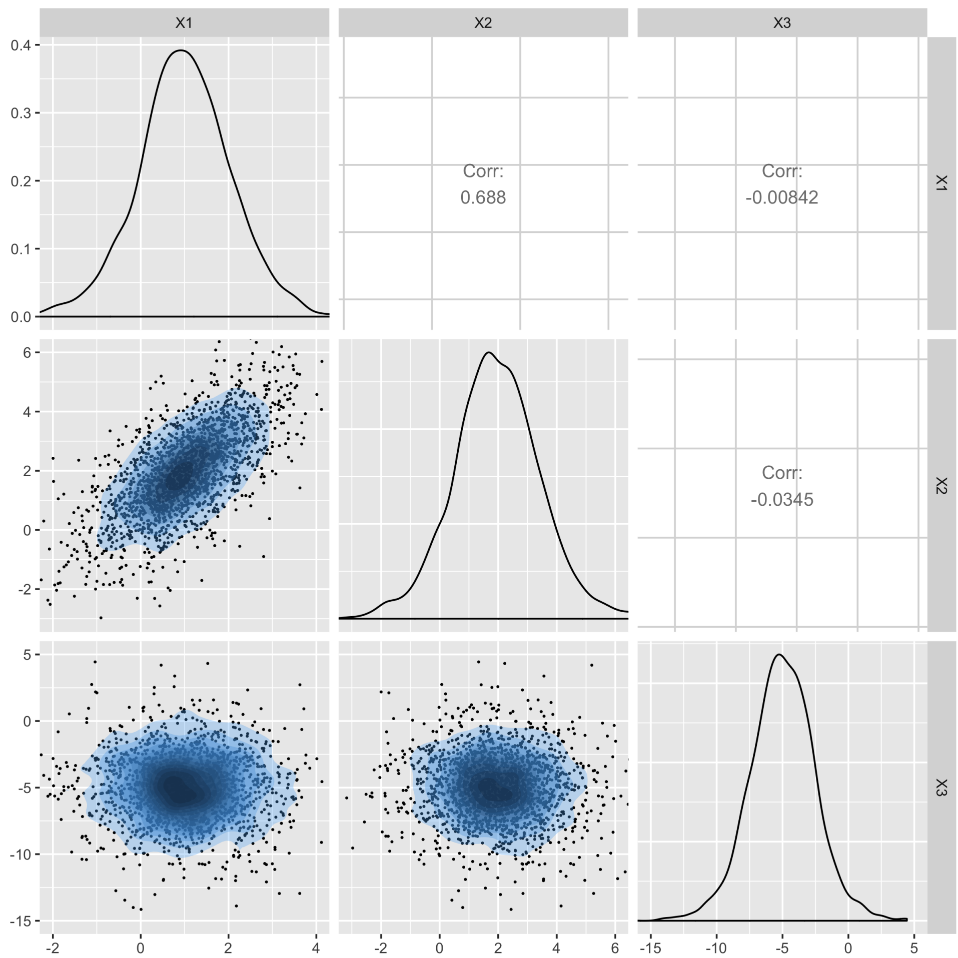


Normal distribution

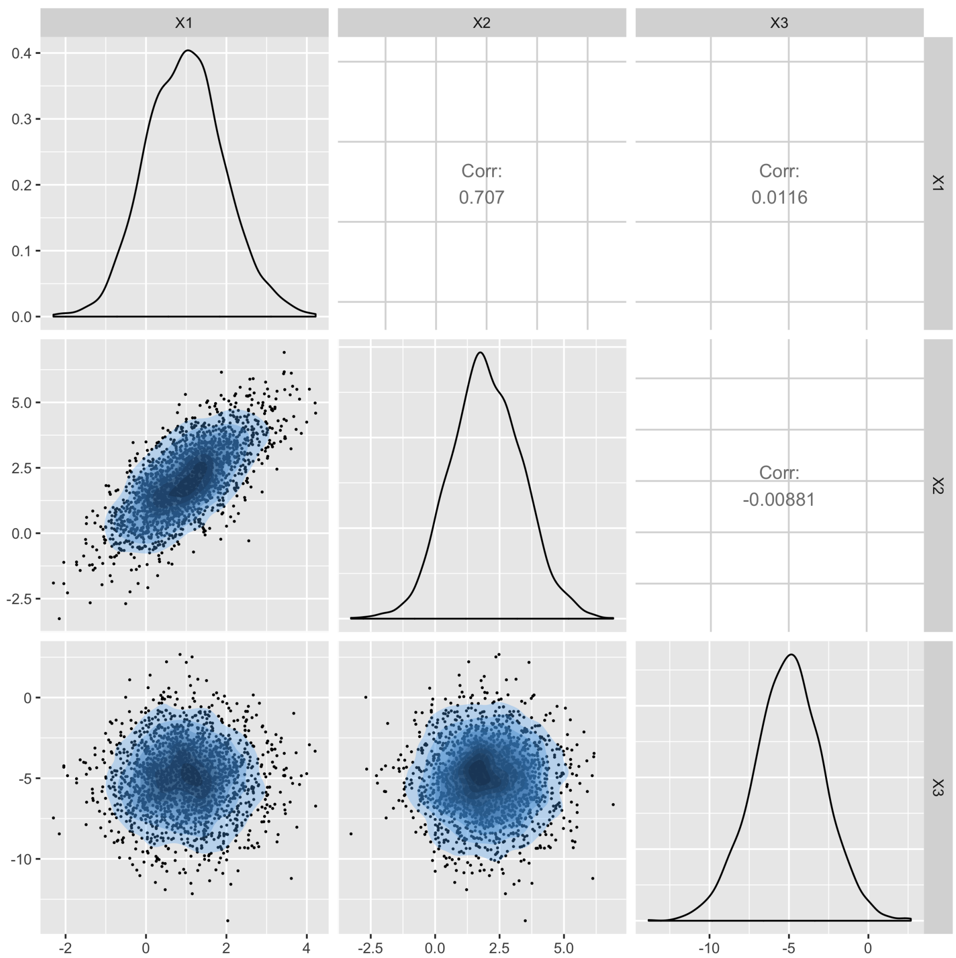


Comparing with normal samples

t-distribution with 10 degrees of freedom



Normal distribution



MULTIVARIATE PROBABILITY DISTRIBUTIONS IN R

Let's generate samples from a multivariate t-distribution!

MULTIVARIATE PROBABILITY DISTRIBUTIONS IN R

Density and cumulative density for multivariate-t

Surajit Ray
Reader, University of Glasgow

Example of multivariate t-distribution

- Individual stocks
  - Univariate t
- Portfolio (3 stocks)
  - Multivariate t
- Probability that all three stocks between $100-150
  - `pmvt()`
- Range of values that the stocks fluctuate 95% of the time
  - `qmvt()`

Density using dmvt

```
dmvt(x, delta = rep(0, p), sigma = diag(p), log = TRUE)
```

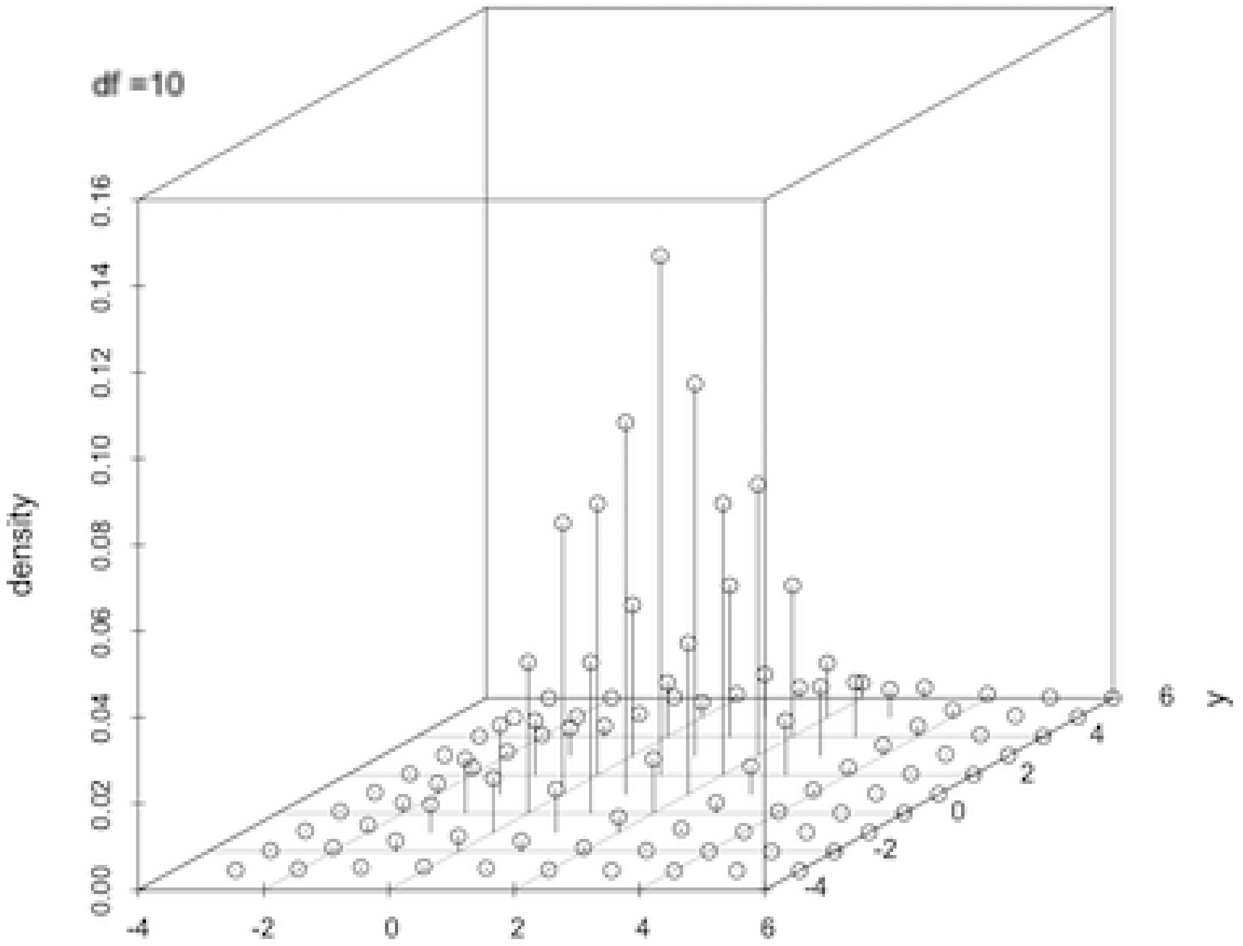
- `x` can be a vector or a matrix
- Unlike `dmvnorm` the default calculation is in **log** scale

To get the densities in natural scale use

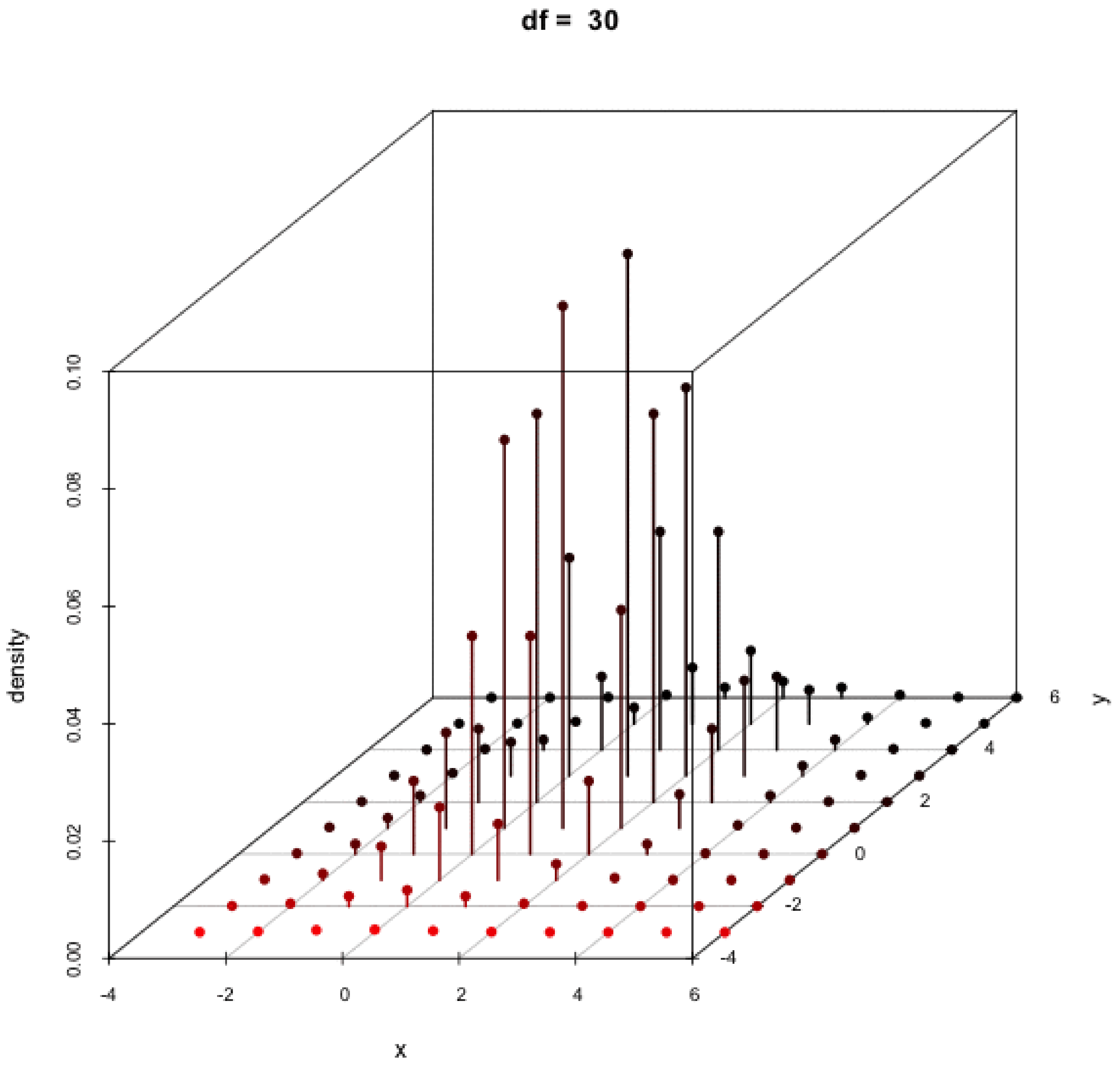
```
dmvt(x, delta = rep(0, p), sigma = diag(p), log = FALSE)
```

Calculating the density of a multivariate t-distribution on a grid

```
x <- seq(-3, 6, by = 1); y <- seq(-3, 6, by = 1)
d <- expand.grid(x = x, y = x)
del1 <- c(1, 2); sig1 <- matrix(c(1, .5, .5, 2), 2)
dens <- dmvt(as.matrix(d), delta = del1, sigma = sig1, df = 10, log = FALSE)

scatterplot3d(cbind(d, dens), type = "h", zlab = "density")
```



Effect of changing the degees of freedom



Cumulative density using pmvt

```
pmvt(lower = -Inf, upper = Inf, delta, sigma, df, ...)
```

- Calculates the cdf or volume similar to normal `pmvnorm()` function

```
pmvt(lower = c(-1, -2), upper = c(2, 2), delta = c(1, 2), sigma = diag(2), df = 6)
```

```
[1] 0.3857
attr(,"error")
[1] 0.0002542
attr(,"msg")
[1] "Normal Completion"
```

Inverse cdf of t-distribution

```
qmvt(p, interval, tail, delta, sigma, df)
```

- Computes the quantile of the multivariate t-distribution
- Computation techniques similar to `qmvnorm()` function

Calculate the 0.95 quantile for 3 degrees of freedom

```
qmvt( p = 0.95, sigma = diag(2), tail = "both", df = 3)

$quantile
[1] 3.96

$f.quantile
[1] -1.05e-06

attr(,"message")
[1] "Normal Completion"
```

MULTIVARIATE PROBABILITY DISTRIBUTIONS IN R

Let's put these functions into practice!

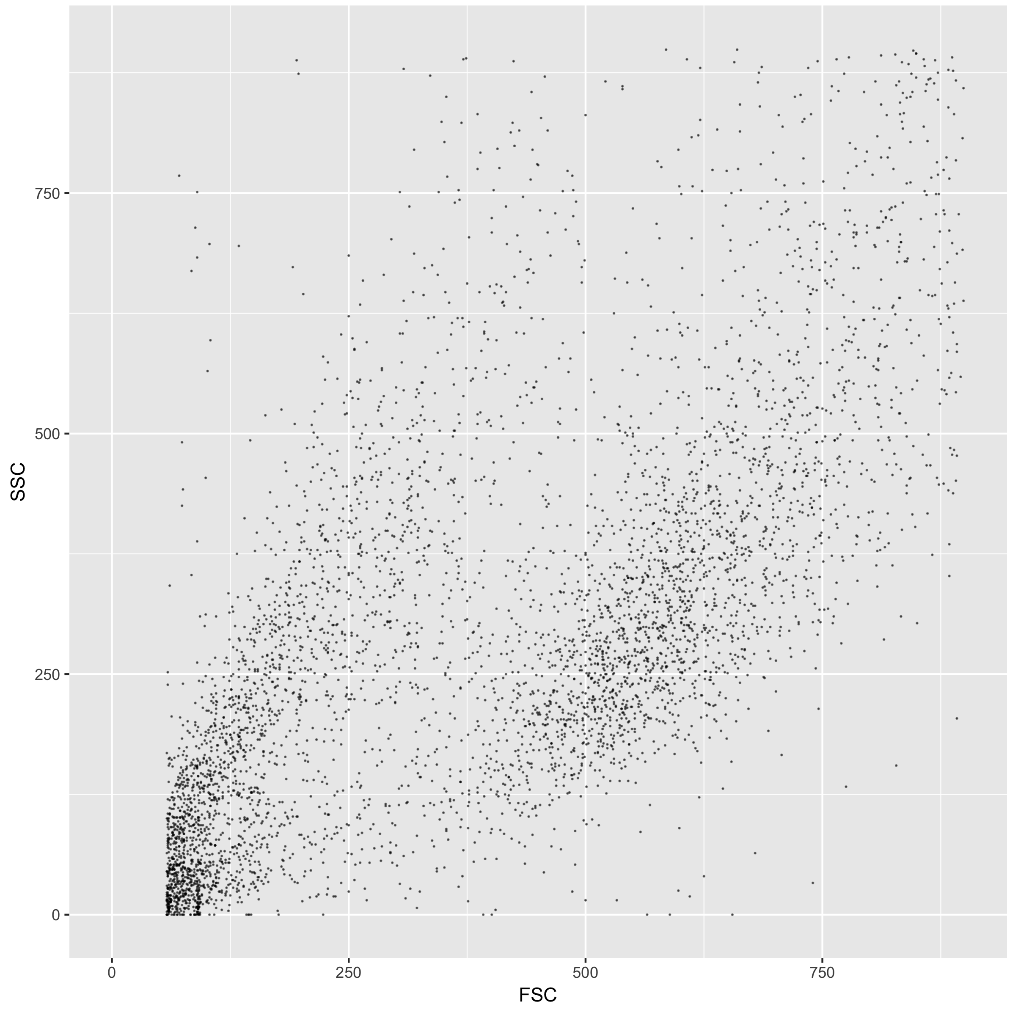
MULTIVARIATE PROBABILITY DISTRIBUTIONS IN R

Multivariate skew distributions

Surajit Ray
Reader, University of Glasgow

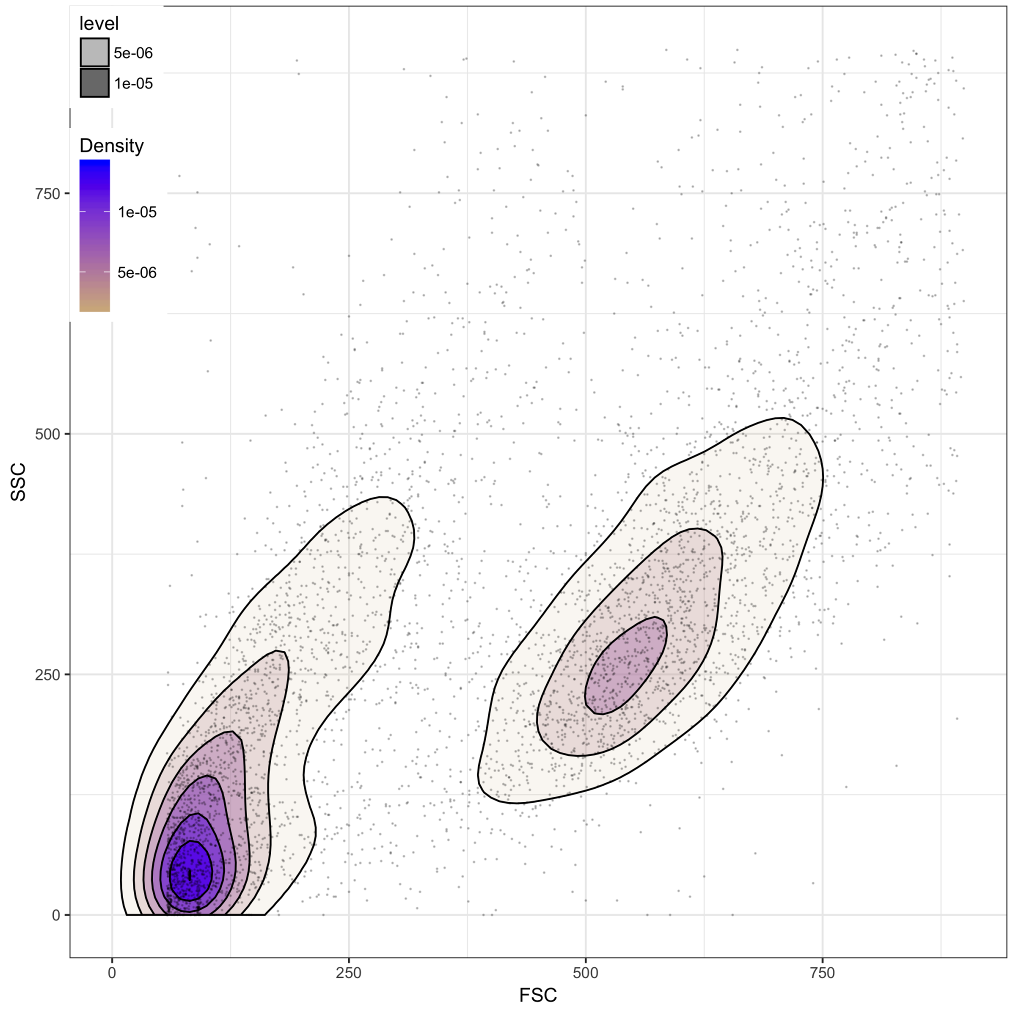
Skew multivariate distribution: scatterplot

Flow cytometry data -- side scatter (SSC) and forward scatter (FSC)



Skew multivariate distribution: contour plot

Flow cytometry data -- side scatter (SSC) and forward scatter (FSC)



Univariate skew-normal distribution

General skew-normal is denoted by $SN(\xi, \omega, \alpha)$

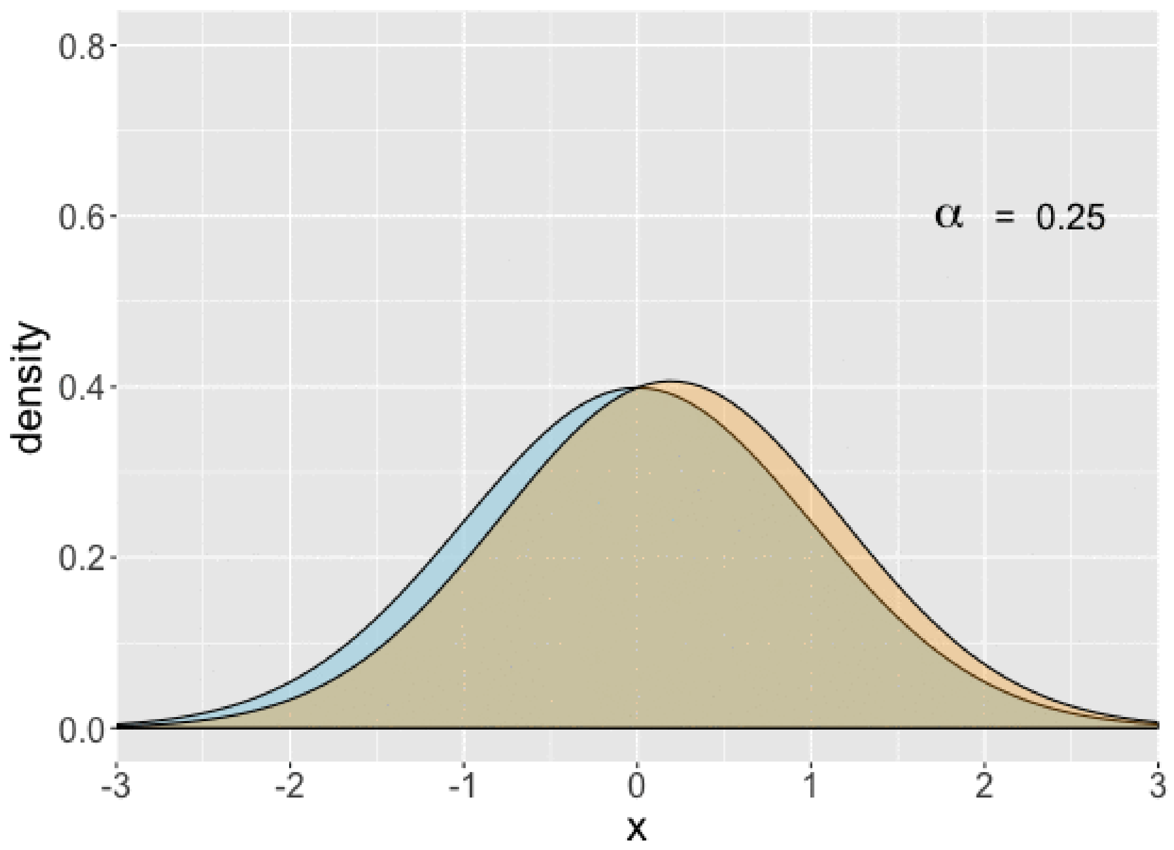
- ξ and ω are the location and scale parameters

Simplest form: $z \sim SN(\alpha)$

- α is the skewness parameter

Range of univariate skew-normal distributions

Comparing $SN(\alpha)$ to a standard Normal



- For $\alpha > 0$ skewed to the right
- For $\alpha < 0$ skewed to the left
- $SN(0)$ is the same as a standard Normal

Multivariate skew-normal distribution

Notations: three-dimensional multivariate skew-normal distribution

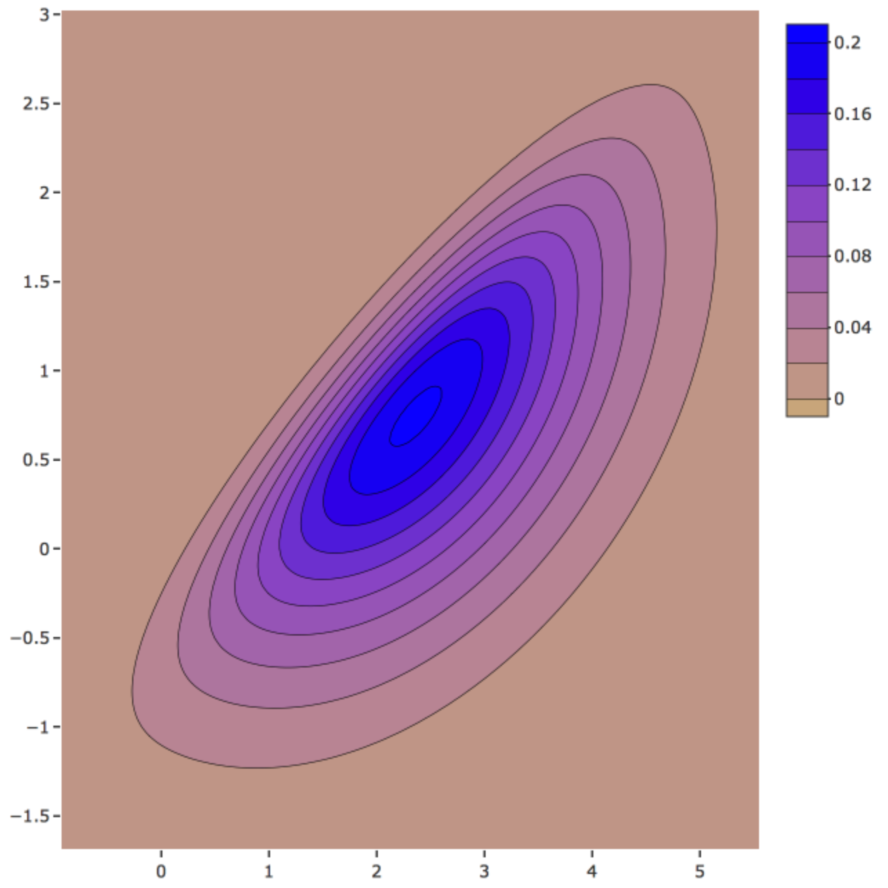
$SN(\xi, \Omega, \alpha)$

- ξ location parameter (vector of length 3)
- Ω variance-covariance parameter (3×3 matrix)
- α skewness parameter (vector of length 3)

Bivariate skew-normal distribution contour plot

Bivariate skew-normal

$$\xi = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix}, \alpha = \begin{pmatrix} -3 \\ 3 \end{pmatrix}.$$



Functions for skew-normal distribution

From `sn` library:

- `dmsn(x, xi, Omega, alpha)`
- `pmsn(x, xi, Omega, alpha)`
- `rmsn(n, xi, Omega, alpha)`
    - Need to specify `xi`, `Omega`, `alpha`

Functions for skew-t distribution

From `sn` library:

- `dmst(x, xi, Omega, alpha, nu)`
- `pmst(x, xi, Omega, alpha, nu)`
- `rmst(n, xi, Omega, alpha, nu )`
    - Need to specify `xi`, `Omega`, `alpha`, `nu` (degrees of freedom)

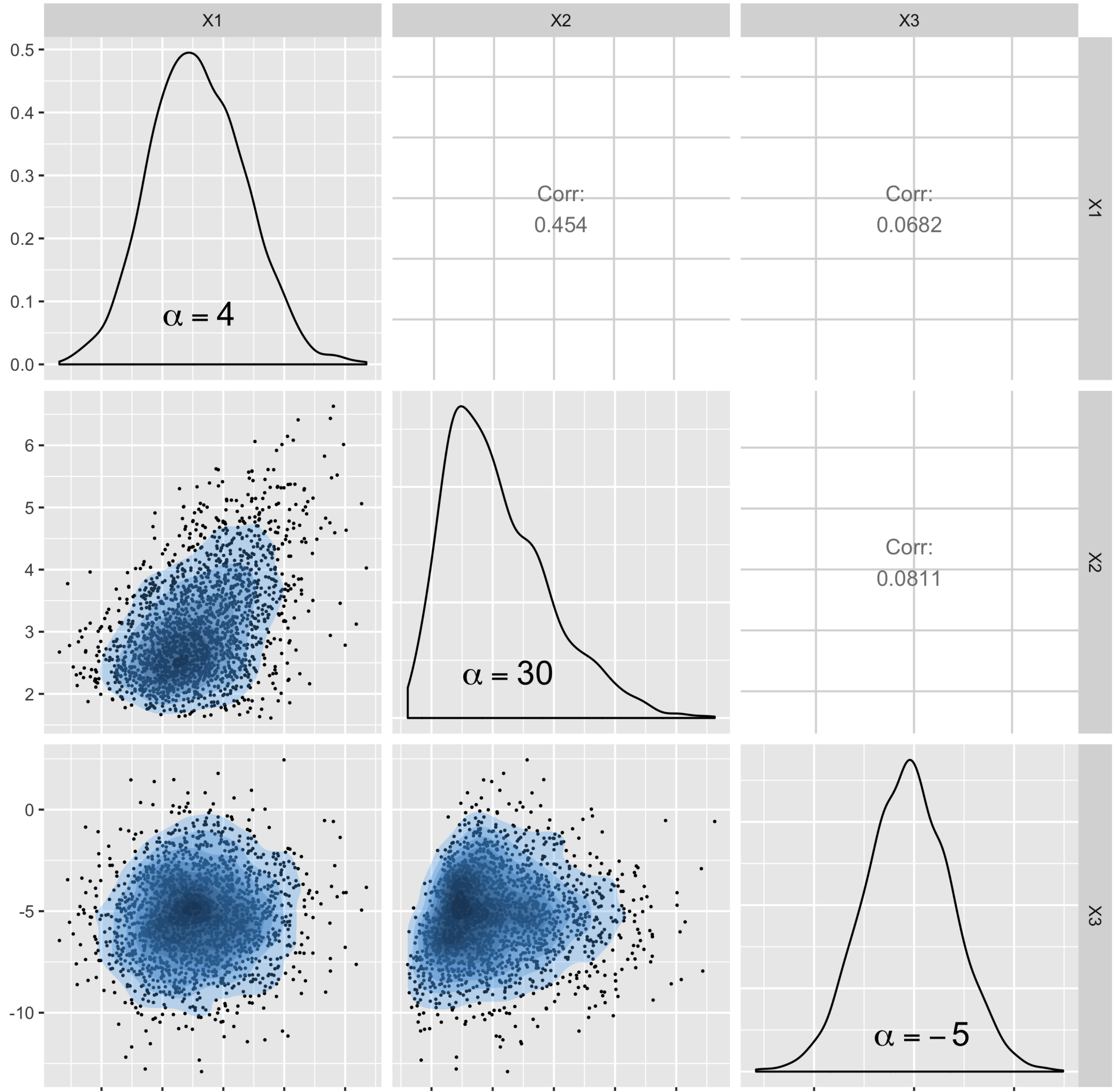
Generating skew-normal samples

Generate 2000 samples from 3 dimensional skew-normal

$$SN\left(\xi = \begin{pmatrix} 1 \\ 2 \\ -5 \end{pmatrix}, \Omega = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 5 \end{pmatrix}, \alpha = \begin{pmatrix} 4 \\ 30 \\ -5 \end{pmatrix}\right)$$

```
# Specify xi, Omega and alpha
xi1 <- c(1, 2, -5)
Omega1 <- matrix(c(1, 1, 0,
                   1, 2, 0,
                   0, 0, 5), 3, 3)
alpha1 <- c(4, 30, -5)

# Generate samples
skew.sample <- rmsn(n = 2000, xi = xi1, Omega = Omega1, alpha = alpha1)
```

Sample from skew-normal distribution



Generating skew-t samples

Generate 2000 samples from 3 dimensional skew-t with

$$\xi = \begin{pmatrix} 1 \\ 2 \\ -5 \end{pmatrix}, \Omega = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 5 \end{pmatrix}, \alpha = \begin{pmatrix} 4 \\ 30 \\ -5 \end{pmatrix}, df = 4$$

```
# Generate samples
skewt.sample <- rmst(n = 2000, xi = xi1, Omega = Omega1, alpha = alpha1, nu = 4)
```

Estimation of parameters from data

- Need iterative algorithm to estimate the parameters of a skew-normal distribution
  - No explicit equation to calculate parameters
- Several functions in `sn` package, including `msn.mle()` function

Estimation of parameters from data

```
msn.mle(y = skew.sample,
        opt.method = "BFGS")
```

```
# Parameter estimation output
$dp
$dp$beta
           X1     X2     X3
[1,] 1.024 2.021 -4.81

$dp$Omega
            X1        X2        X3
X1  0.9154  0.8865 -0.1507
X2  0.8865  1.8276 -0.3560
X3 -0.1507 -0.3560  5.0352

$dp$alpha
     X1     X2      X3
 3.670 28.465 -5.029
```

Samples were generated using:

$$\xi = \begin{pmatrix} 1 \\ 2 \\ -5 \end{pmatrix}, \Omega = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 5 \end{pmatrix}, \alpha = \begin{pmatrix} 4 \\ 30 \\ -5 \end{pmatrix}$$

MULTIVARIATE PROBABILITY DISTRIBUTIONS IN R

Now let's do some exercises with skew-normal distributions!