MultiVariate Normal Distribution

Jing Qin

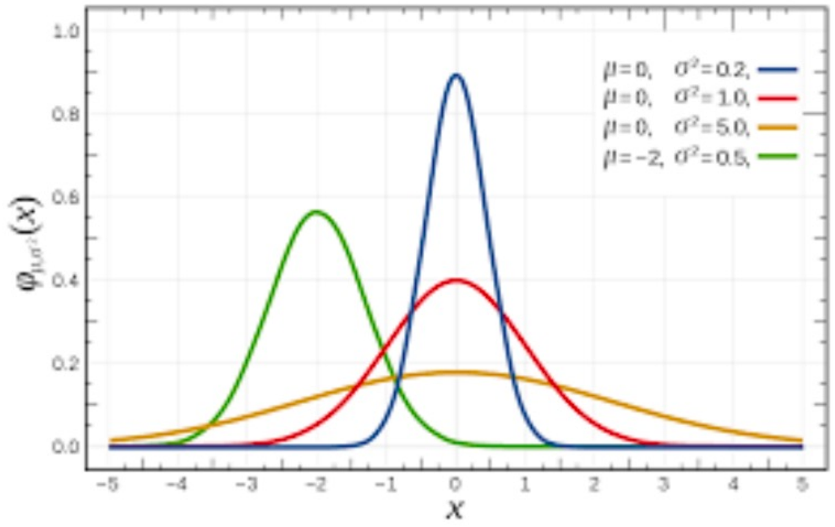
22/03/2022

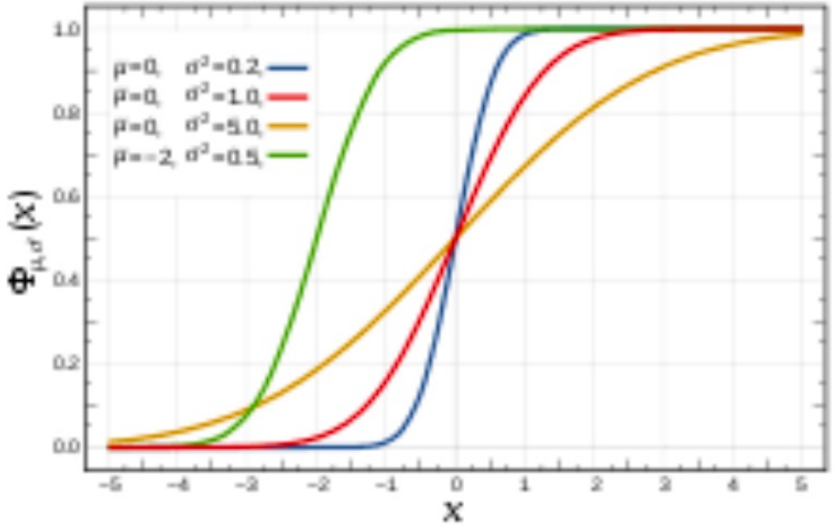
Re-cap: univariate normal distribution

Assume a r.v. X satisfies a normal/Gaussian distribution $N(\mu, \sigma^2)$, i.e.

$$X \sim N(\mu, \sigma^2)$$

- Probability density function $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\}$





- $E(X) = \mu$ and $Var(X) = \sigma^2$

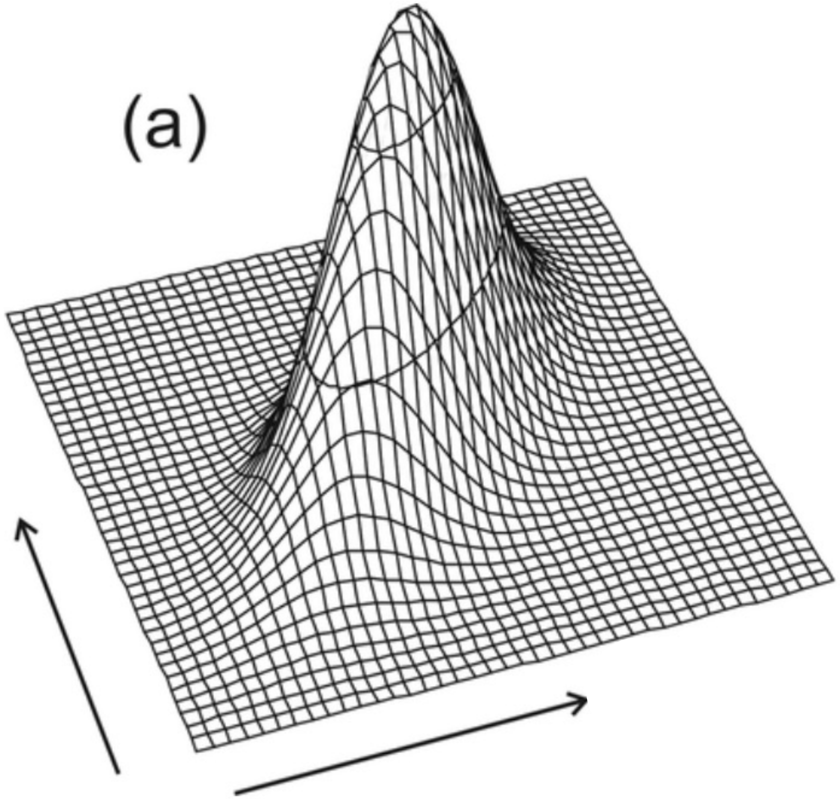
- Estimation about μ and σ with some given data: confidence interval and hypothesis test.

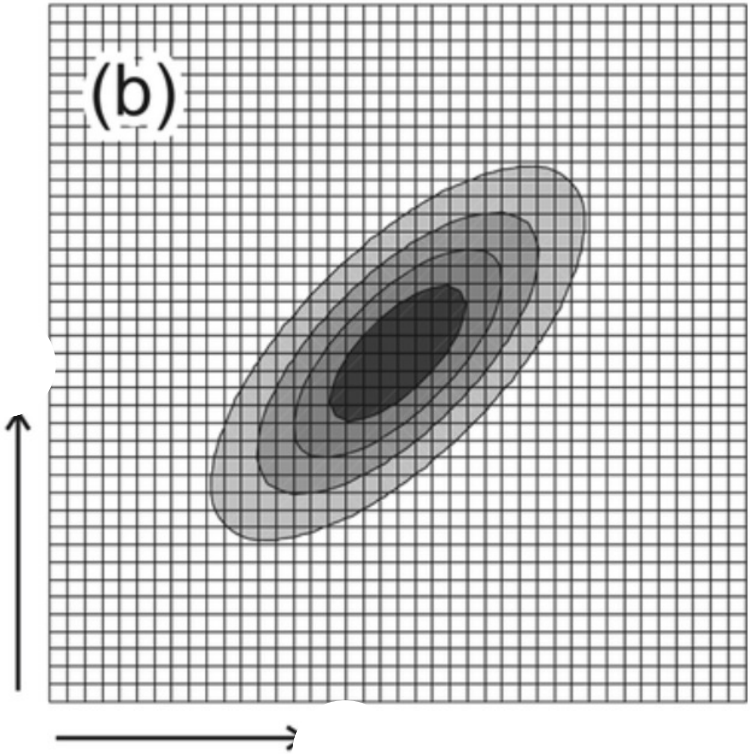
Bivariate normal distribution $\boldsymbol{p} = \mathbf{2}$

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$





$$\frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}} \quad \times$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right\}$$

Towards general: vector and matrix form

$$\boxed{\frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right\}}$$

$$\boldsymbol{\mu} = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix} \qquad |\Sigma| = \sigma_1^2\sigma_2^2(1-\rho^2)$$

$$\Sigma = \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix} \qquad \Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{pmatrix}\sigma_2^2 & -\rho\sigma_1\sigma_2\\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{pmatrix}$$

$$\boxed{\frac{1}{(2\pi)^{2/2}|\Sigma|^{1/2}} \exp\{-(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2\}}$$

?

For general $\boldsymbol{p} \geq \mathbf{2}$, joint PDF (4-4) $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$

$$\frac{1}{(2\pi)^{2/2}|\Sigma|^{1/2}} \exp\{-(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2\}$$

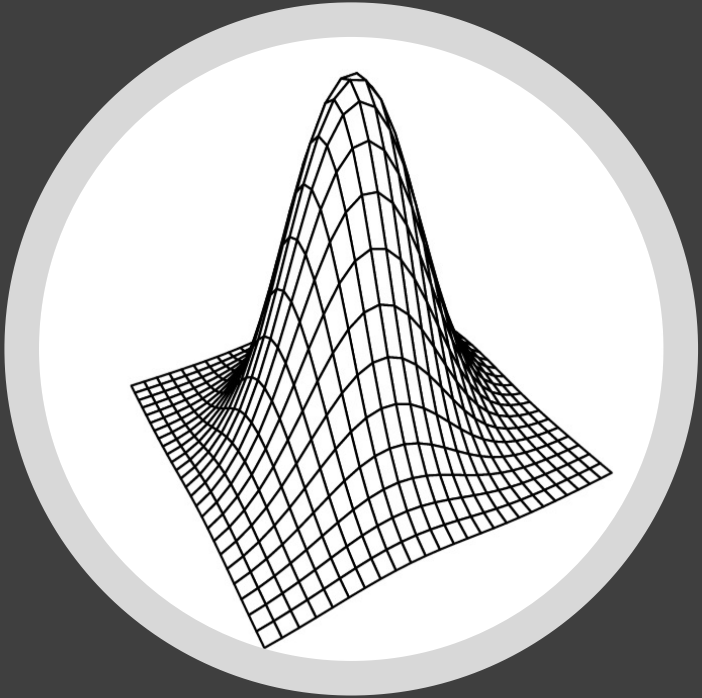
$$\Big\downarrow$$

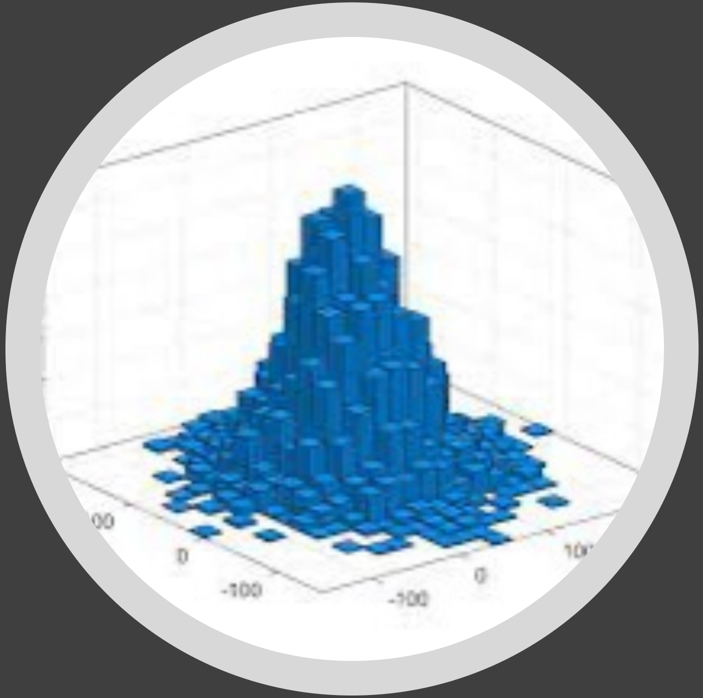
$$\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\{-(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2\}$$

This is why we need the vectors!

Consider $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)'$ and $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)'$

$E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $Cov(\boldsymbol{X}) = \Sigma$.





<u>In practice:</u> *Is my data normally distributed*?

Example: radiation data with door open+closed (t4-1.dat; t4-5.dat)

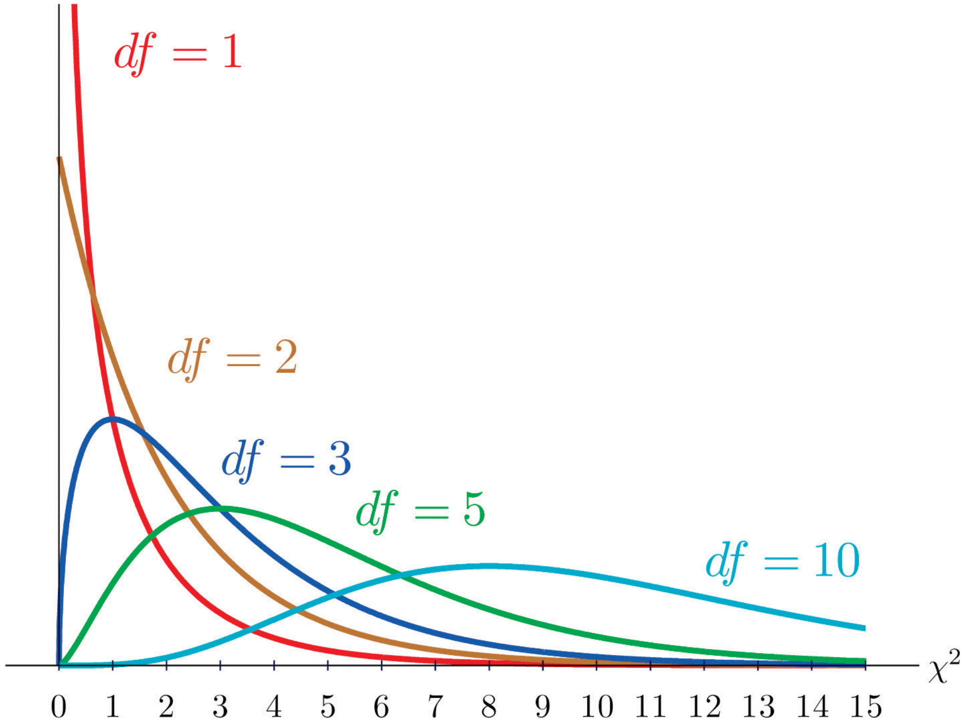
Quadratic form (4-8)

Quadratic form

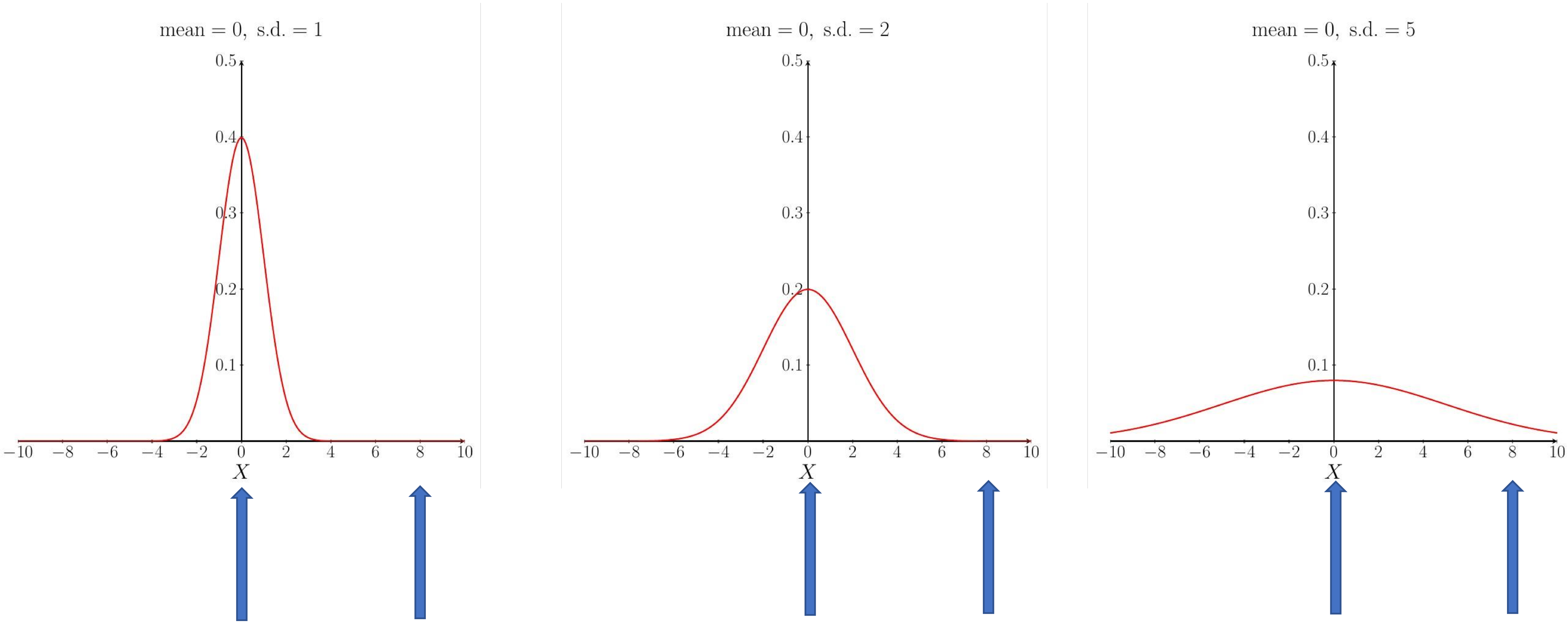
$$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma) \Rightarrow \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\{-\boxed{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}/2\}$$

p-dimension 1-dimension

- (4-8) Assume $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, we have $(\boldsymbol{X}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \sim \chi^2_p$

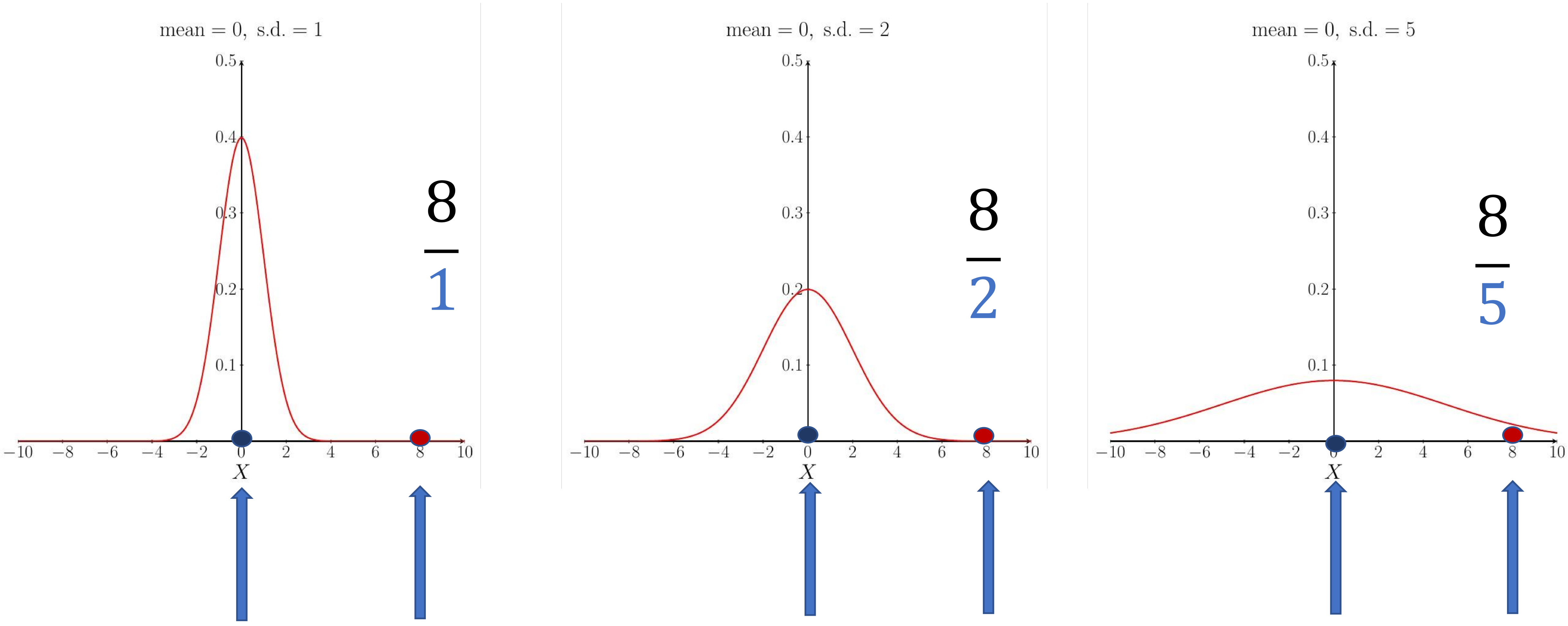


Statistical/Mahalanobis distance



Statistical/Mahalanobis distance

Bigger variability, smaller difference



$$\frac{8}{1} \qquad \frac{8}{2} \qquad \frac{8}{5}$$

Statistical/Mahalanobis distance

- (2-17, 4-3) The quadratic form $(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ is referred to as squared statistical/Mahalanobis distance. R cmd `mahalanobis()`

From $\boldsymbol{x}$ *to* $\boldsymbol{\mu}$

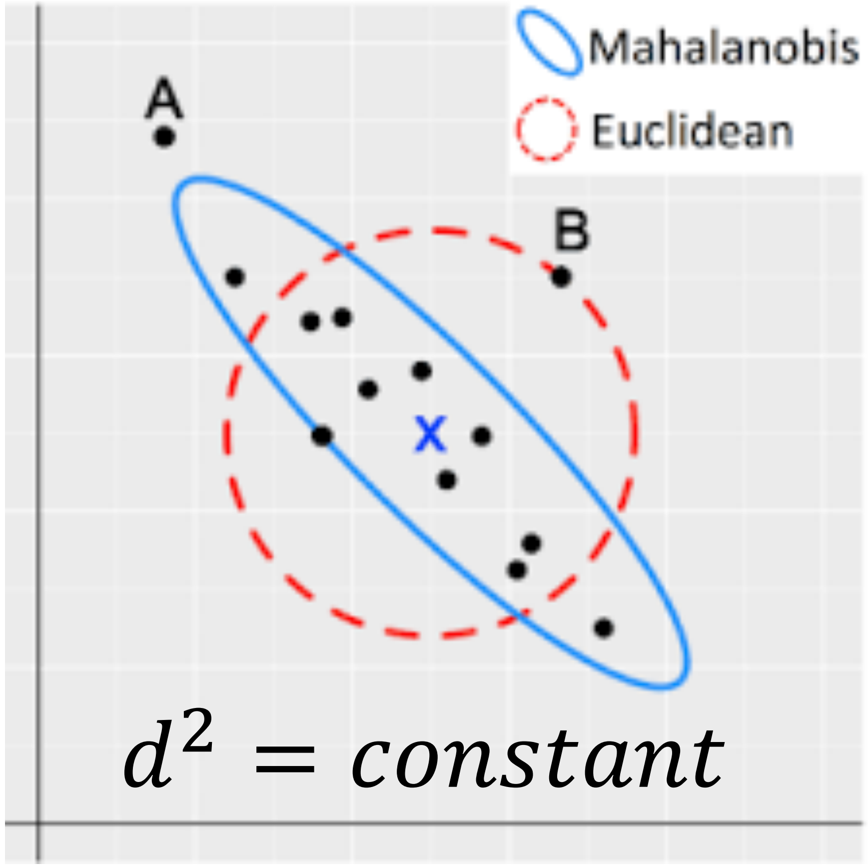
$$d_E(\boldsymbol{x},\boldsymbol{y}) = \sqrt{(\boldsymbol{x}-\boldsymbol{y})^T \cdot (\boldsymbol{x}-\boldsymbol{y})}$$

$$d_M(\boldsymbol{x},\boldsymbol{y}) = \sqrt{(\boldsymbol{x}-\boldsymbol{y})^T \cdot S^{-1} \cdot (\boldsymbol{x}-\boldsymbol{y})}$$

$$= \sqrt{\begin{bmatrix} x_1-y_1 & x_2-y_2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix} \begin{bmatrix} x_1-y_1 \\ x_2-y_2 \end{bmatrix}}$$

$$= \sqrt{\begin{bmatrix} \frac{x_1-y_1}{\sigma_1^2} & \frac{x_2-y_2}{\sigma_2^2} \end{bmatrix} \begin{bmatrix} x_1-y_1 \\ x_2-y_2 \end{bmatrix}}$$

$$= \sqrt{\frac{(x_1-y_1)^2}{\sigma_1^2} + \frac{(x_2-y_2)^2}{\sigma_2^2}}$$



- Result (4.7) The solid ellipsoid of $\boldsymbol{x}$ values satisfying

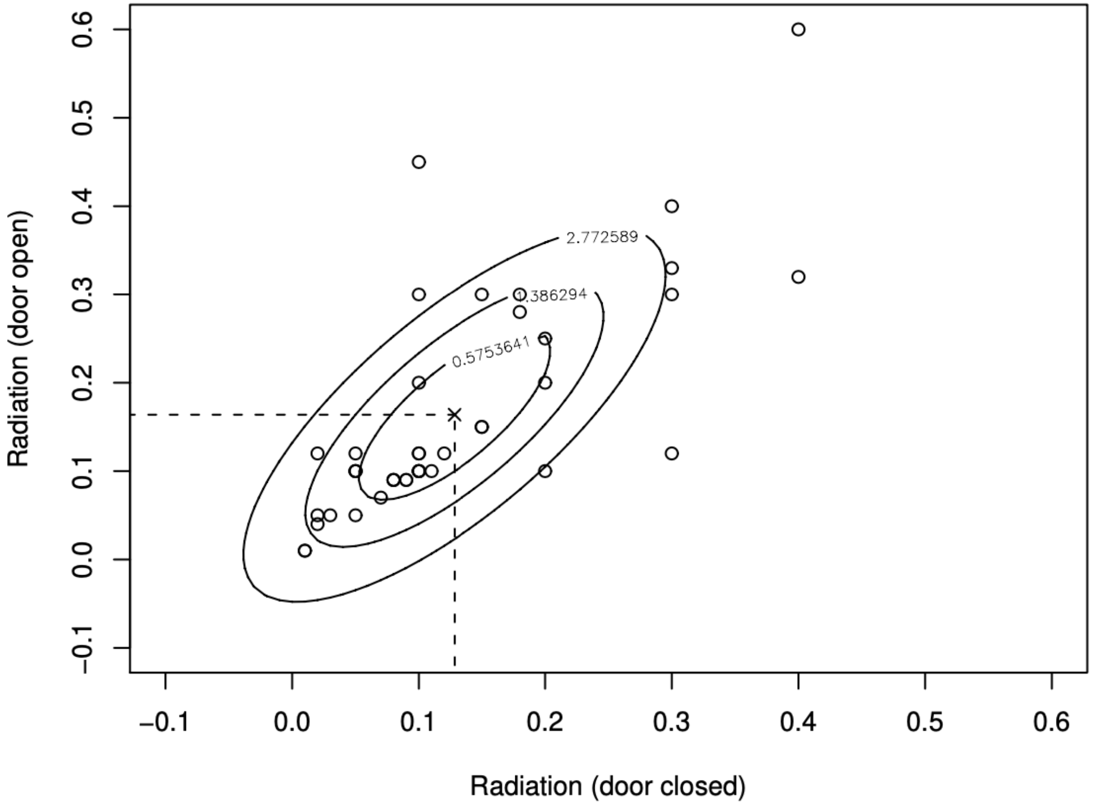
$$(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \le c^2 = \chi_p^2(\alpha)$$

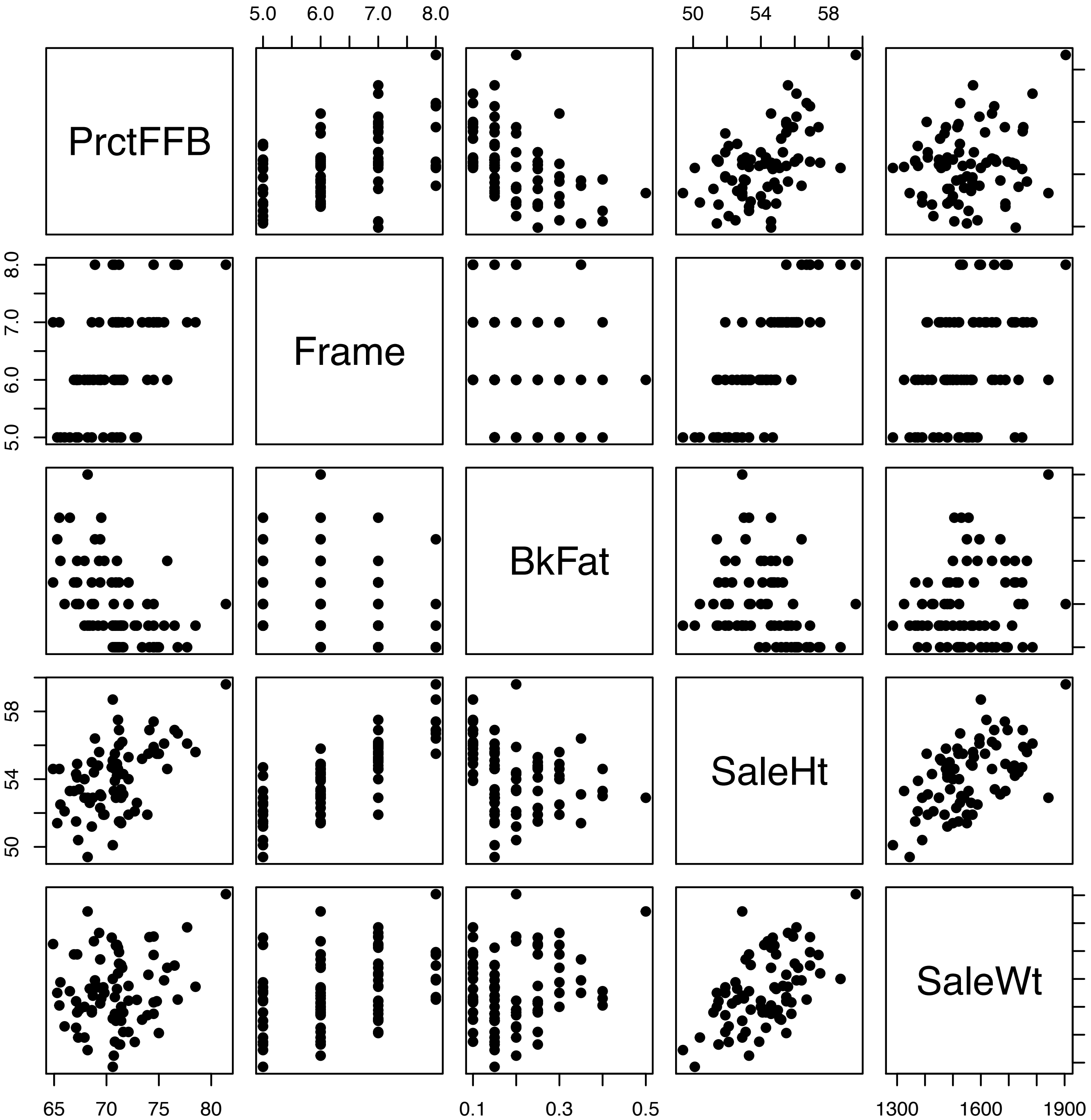
has probability $1-\alpha$.

$qchisq(\alpha, p)$

$1-\alpha$	Observed count	Expected count
0.25	17	10.5
0.50	29	21
0.75	33	31.5

Expected number of observations versus data. Note n = 42




PrctFFB
Frame
BkFat
SaleHt
SaleWt
5.0 6.0 7.0 8.0
50 54 58
80 75 70 65
0.5 0.3 0.1
1900 1600 1300
65 70 75 80
0.1 0.3 0.5
1300 1600 1900

Q-Q plot, again

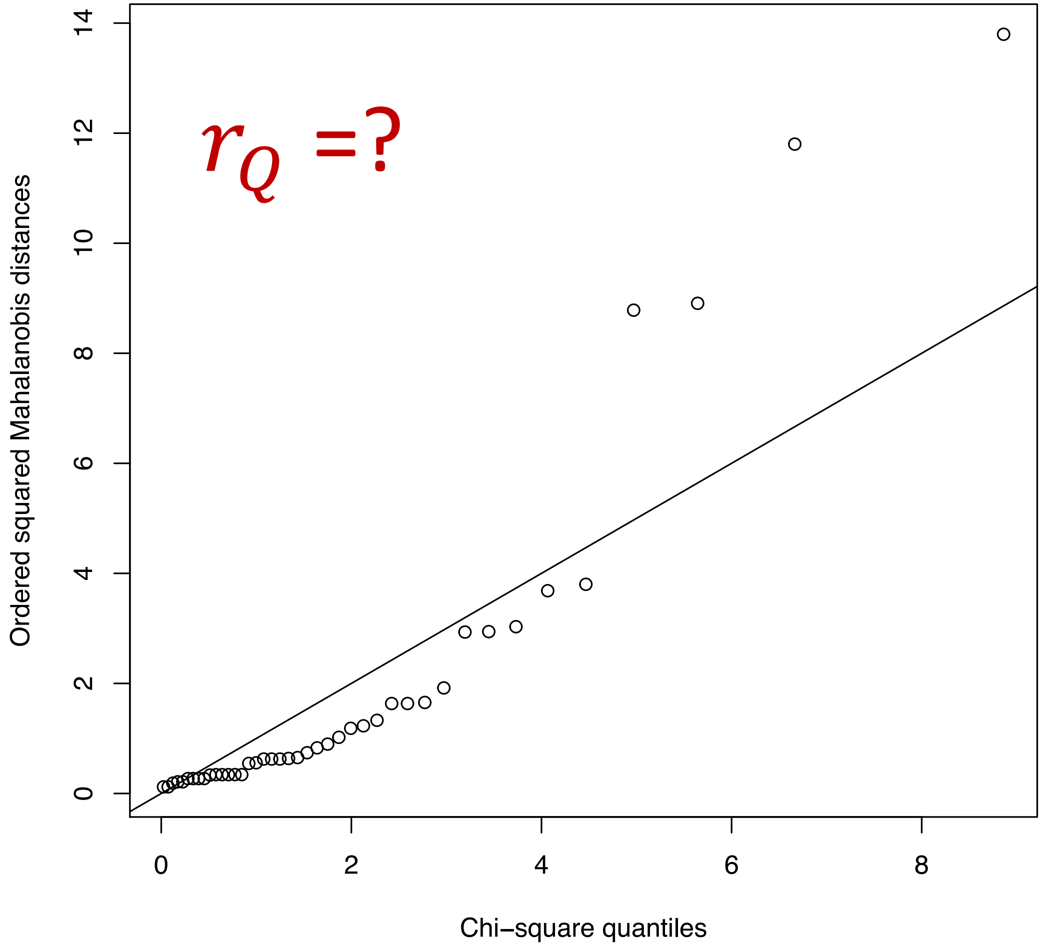
Dimension?

- Result (4.7) Assume $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, we have $\boxed{(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu})} \sim \chi^2_p$

Q-Q plot, again

Dimension?

- Result (4.7) Assume $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, we have $\boxed{(\boldsymbol{X}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu})} \sim \chi_p^2$



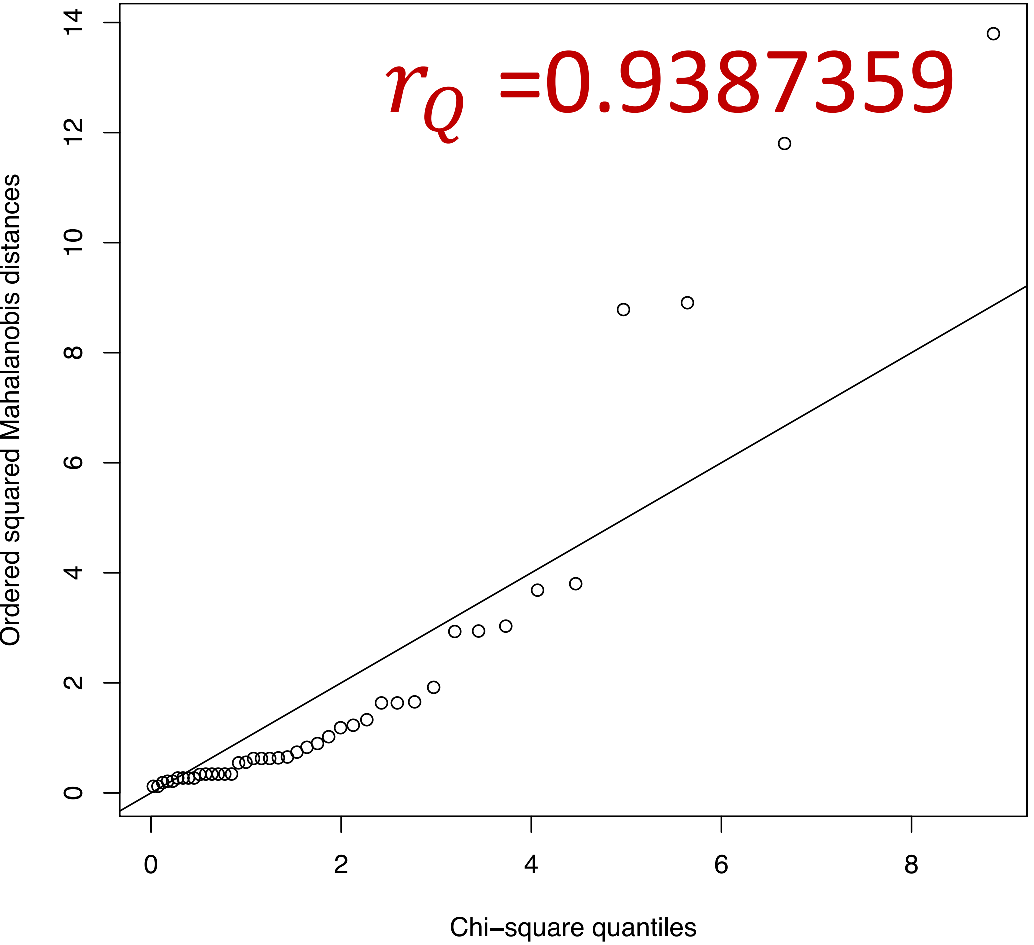
Can we use Table 4.2 again?

Table 4.2 Critical Points for the Q–Q Plot Correlation Coefficient Test for Normality

Sample size n	Significance levels α .01	.05	.10
5	.8299	.8788	.9032
10	.8801	.9198	.9351
15	.9126	.9389	.9503
20	.9269	.9508	.9604
25	.9410	.9591	.9665
30	.9479	.9652	.9715
35	.9538	.9682	.9740
40	.9599	.9726	.9771
45	.9632	.9749	.9792
50	.9671	.9768	.9809
55	.9695	.9787	.9822
60	.9720	.9801	.9836
75	.9771	.9838	.9866
100	.9822	.9873	.9895
150	.9879	.9913	.9928
200	.9905	.9931	.9942
300	.9935	.9953	.9960

Q-Q plot, again

- Result (4.7) Assume $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, we have $\boxed{(\boldsymbol{X}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu})} \sim \chi_p^2$



1. Generate multiple datasets with χ_p^2 for n=42
2. Make Q-Q plots for each of the dataset and derive r_Q respectively
3. Collect all the r_Q and find the critical value for some given significant level.

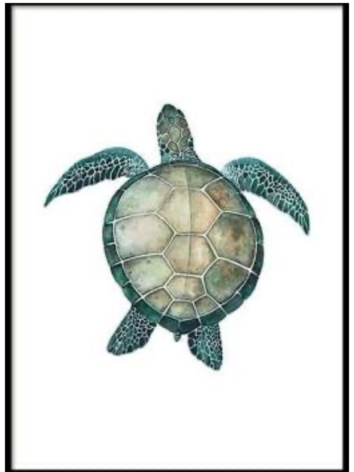
```
> source("FindCrikChi2.R")
> result1[[2]]
[1] 0.9948543
```

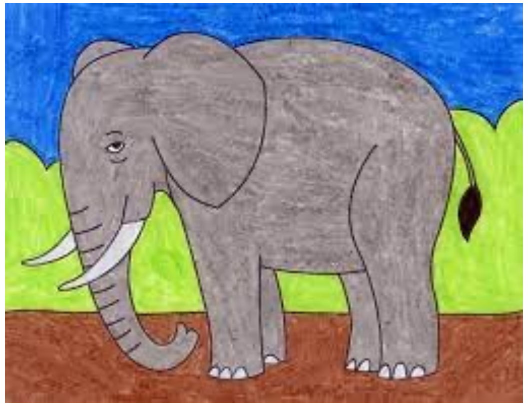
Conclusion ?

Is *quadratic form* enough for assessing normality?

- Result (4.7) Assume $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, we have $\boxed{(\boldsymbol{X}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu})} \sim \chi_p^2$

This animal is a turtle. **This is an animal with 4 legs.**





Check MVN, continued

- (Result 4.4) Any subset of a MVN distributed random vector is normally distributed.

Example 4.5 (The distribution of a subset of a normal random vector)

If $\mathbf{X}$ is distributed as $N_5(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, find the distribution of $\begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$. We set

$$\mathbf{X}_1 = \begin{bmatrix} X_2 \\ X_4 \end{bmatrix}, \qquad \boldsymbol{\mu}_1 = \begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \quad , \quad \boldsymbol{\Sigma}_{11} = \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}$$

In summary, basic track of checking MVN

- Test univariate normality for each marginal distribution with QQ-plot.
- Test bivariate normality for each pair of attributes. For example, a matrix of scatterplots and QQ-plot based on $(\boldsymbol{X}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \sim \chi^2_2$
- Test over all MVN using QQ-plot based on $(\boldsymbol{X}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \sim \chi^2_p$
- Linear pattern in QQ-plot can be evaluated through hypothesis test.

Advanced track: well, it is by taking care of all the possible subsets of the attributes…or PCA

Useful properties of MVN

- (Result 4.3) Let $\boldsymbol{A}$ be a $(q \times p)$ numeric matrix, then

$$\boldsymbol{AX} \sim N_q(\boldsymbol{A\mu}, \boldsymbol{A}\Sigma\boldsymbol{A}')$$

- Exercise 2 Find the mean vector and the total variance of $\boldsymbol{AX}$.
  Given that $\boldsymbol{A} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$ and $\boldsymbol{X} = (X_1, X_2, X_3)'$.

Exp 4.4

Further we know $\boldsymbol{X} \sim N_3(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = (1, 2, 1)'$ and

$$\Sigma = \begin{pmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

Check MVN, continued

- (Result 4.4) Any subset of a MVN distributed random vector is normally distributed.

Example 4.5 (The distribution of a subset of a normal random vector)

If $\mathbf{X}$ is distributed as $N_5(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, find the distribution of $\begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$. We set

$$\mathbf{X}_1 = \begin{bmatrix} X_2 \\ X_4 \end{bmatrix}, \qquad \boldsymbol{\mu}_1 = \begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{11} = \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}$$

Useful properties of MVN

- (Result 4.4) Any subset of a MVN distributed random vector is normally distributed.

Example 4.6 (The equivalence of zero covariance and independence for normal variables) Let $\underset{(3\times 1)}{\mathbf{X}}$ be $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Question 1: What is the distribution of (X_1, X_2)?

$Cov(\boldsymbol{X}, \boldsymbol{Y})$ as a notation

The matrix of all covariances between elements in $\boldsymbol{X}$ and $\boldsymbol{Y}$

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \qquad \boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

$$Cov(\boldsymbol{X}, \boldsymbol{Y}) = \begin{bmatrix} Cov(X_1, Y_1) & Cov(X_1, Y_2) & Cov(X_1, Y_3) \\ Cov(X_2, Y_1) & Cov(X_2, Y_2) & Cov(X_2, Y_3) \end{bmatrix}$$

It is sometimes convenient to use the Cov $(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$ notation where

$$\text{Cov}\,(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \mathbf{\Sigma}_{12}$$

is a matrix containing all of the covariances between a component of $\mathbf{X}^{(1)}$ and a component of $\mathbf{X}^{(2)}$.

(2-38)+(2-40)

$$\mathbf{X} = \left[\begin{array}{c} X_1 \\ \vdots \\ X_q \\ \hdashline X_{q+1} \\ \vdots \\ X_p \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c} X_1 \\ \vdots \\ X_q \end{array}}\right\} q \\ \\ \left.\vphantom{\begin{array}{c} X_1 \\ \vdots \\ X_p \end{array}}\right\} p - q \end{array} = \left[\begin{array}{c} \mathbf{X}^{(1)} \\ \hdashline \mathbf{X}^{(2)} \end{array}\right] \quad \text{and} \quad \boldsymbol{\mu} = E(\mathbf{X}) = \left[\begin{array}{c} \mu_1 \\ \vdots \\ \mu_q \\ \hdashline \mu_{q+1} \\ \vdots \\ \mu_p \end{array}\right] = \left[\begin{array}{c} \boldsymbol{\mu}^{(1)} \\ \hdashline \boldsymbol{\mu}^{(2)} \end{array}\right] \tag{2-38}$$

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \begin{array}{c} \\ q \\ p-q \end{array} \begin{array}{c} \begin{array}{cc} q & p-q \end{array} \\ \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right] \\ (p\times p) \end{array}$$

$$= \left[\begin{array}{ccc:ccc} \sigma_{11} & \cdots & \sigma_{1q} & \sigma_{1,q+1} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{q1} & \cdots & \sigma_{qq} & \sigma_{q,q+1} & \cdots & \sigma_{qp} \\ \hdashline \sigma_{q+1,1} & \cdots & \sigma_{q+1,q} & \sigma_{q+1,q+1} & \cdots & \sigma_{q+1,p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pq} & \sigma_{p,q+1} & \cdots & \sigma_{pp} \end{array}\right] \tag{2-40}$$

Useful properties of MVN

Matrix of 0's

Result 4.5.

(a) If $\underset{(q_1\times 1)}{\mathbf{X}_1}$ and $\underset{(q_2\times 1)}{\mathbf{X}_2}$ are independent, then $\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{0}$, a $q_1 \times q_2$ matrix of zeros.

(b) If $\begin{bmatrix} \mathbf{X}_1 \\ \hline \mathbf{X}_2 \end{bmatrix}$ is $N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \hline \boldsymbol{\mu}_2 \end{bmatrix}, \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right]\right)$, then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

(c) If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent and are distributed as $N_{q_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $N_{q_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$, respectively, then $\begin{bmatrix} \mathbf{X}_1 \\ \hline \mathbf{X}_2 \end{bmatrix}$ has the multivariate normal distribution

$$N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \hline \boldsymbol{\mu}_2 \end{bmatrix}, \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \hline \mathbf{0}' & \boldsymbol{\Sigma}_{22} \end{array}\right]\right)$$

Useful properties of MVN

- (Result 4.4) Any subset of a MVN distributed random vector is normally distributed.

Example 4.6 (The equivalence of zero covariance and independence for normal variables) Let $\underset{(3\times 1)}{\mathbf{X}}$ be $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Are X_1 and X_2 independent? What about (X_1, X_2) and X_3?

Question 2: are X_1, X_2 independent?
Question 3: are (X_1, X_2) and X_3 independent?

Linear combinations of **random vectors**

Assume $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ is are independent and identically distributed and $\boldsymbol{X}_j \sim N_p(\boldsymbol{\mu}, \Sigma)$.

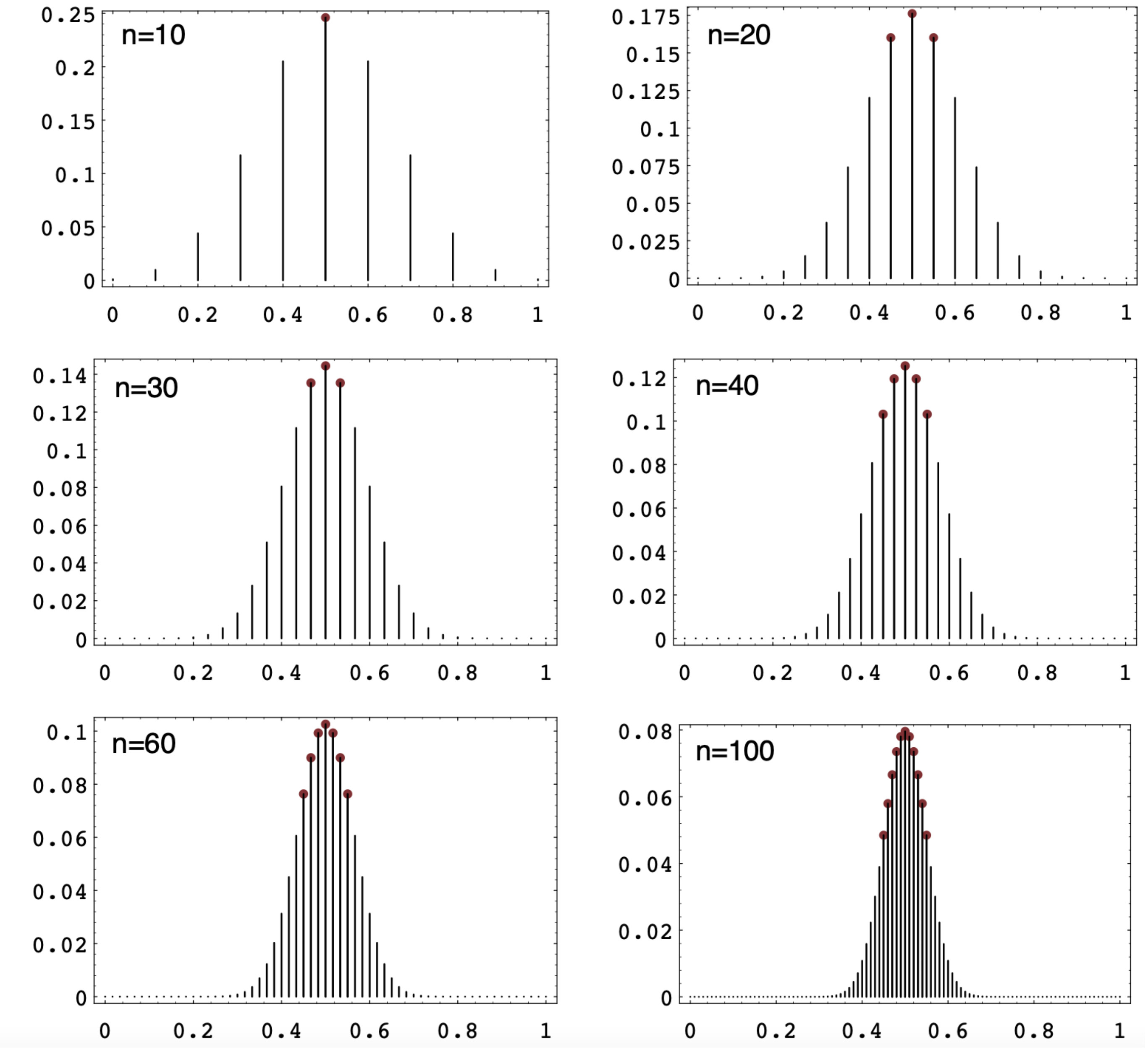
- (Result 4.8) If $\boldsymbol{V}_1 = c_1\boldsymbol{X}_1 + c_2\boldsymbol{X}_2 + \cdots + c_n\boldsymbol{X}_n$, then

$$V_1 \sim N_p\left(\sum c_j\boldsymbol{\mu}, \left(\sum c_j^2\right)\Sigma\right)$$

What is the distribution of the sample mean?

$$\overline{\boldsymbol{X}} = \frac{1}{n}\sum \boldsymbol{X}_j$$

See the shape and how the variance decreases while n increases



CLT: we know sample mean vector really well

$$\bar{X}\ approx \sim N_p(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{n})$$

Result 4.13 (The central limit theorem). Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent observations from any population with mean $\boldsymbol{\mu}$ and finite covariance $\boldsymbol{\Sigma}$. Then

$$\sqrt{n}\,(\overline{\mathbf{X}} - \boldsymbol{\mu}) \text{ has an approximate } N_p(\mathbf{0}, \boldsymbol{\Sigma}) \text{ distribution}$$

for large sample sizes. Here n should also be large relative to p.

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent observations from a population with mean $\boldsymbol{\mu}$ and finite (nonsingular) covariance $\boldsymbol{\Sigma}$. Then

$$\sqrt{n}\,(\overline{\mathbf{X}} - \boldsymbol{\mu}) \text{ is approximately } N_p(\mathbf{0}, \boldsymbol{\Sigma})$$

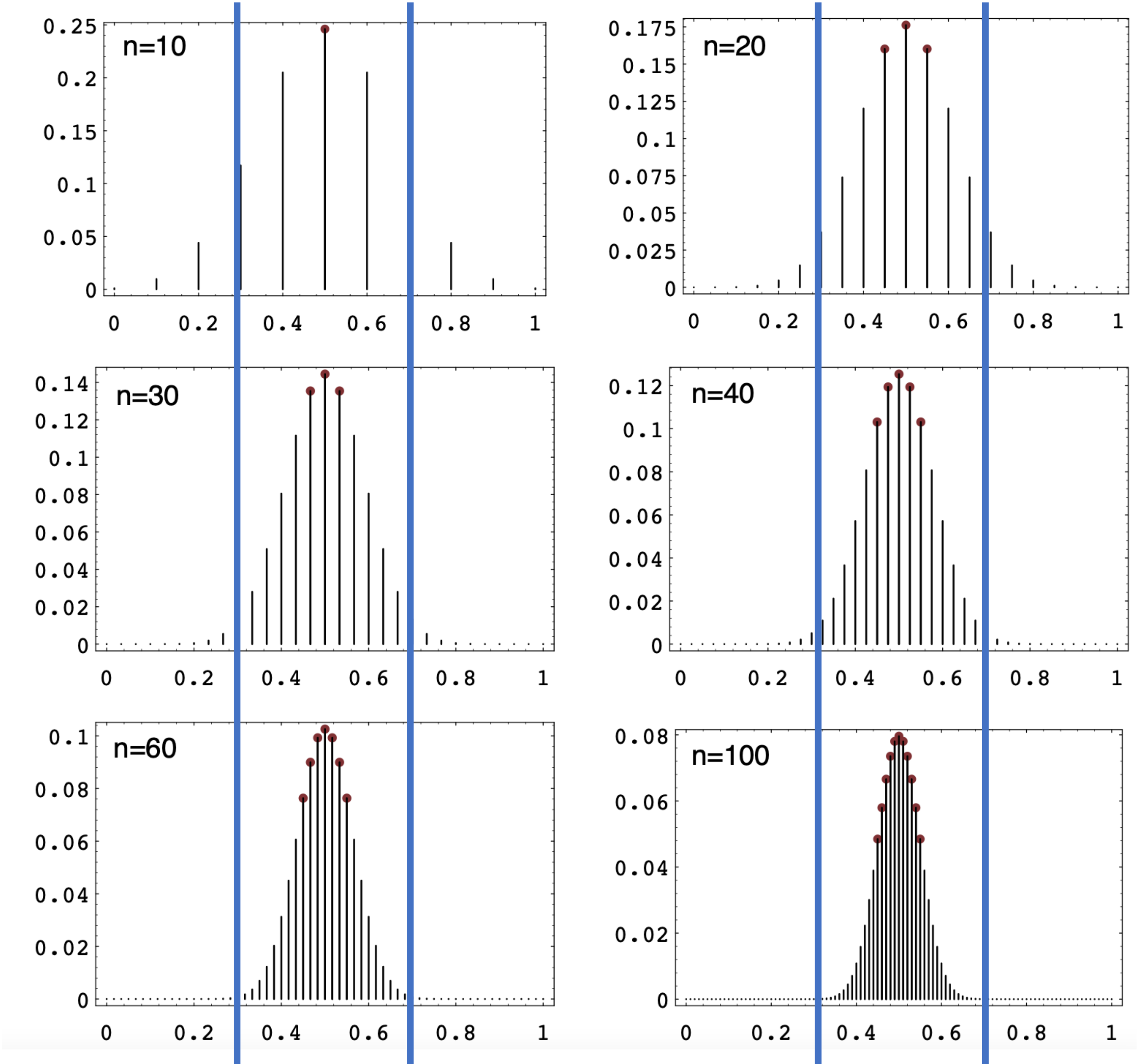
and (4-28)

$$n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) \text{ is approximately } \chi^2_p$$

for $n - p$ large.

Assume $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, we have $(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2_p$

$\mathit{Now\ on\ the\ tails}$: $\mathrm{P}[-0.2 < \bar{Y} - 0{,}5 < 0.2] \longrightarrow 0$ as n grows big



Result (4.12) LLN

Result 4.12 (Law of large numbers). Let $Y_1, Y_2, \ldots, Y_n$ be independent observations from a population with mean $E(Y_i) = \mu$. Then

$$\bar{Y} = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$$

converges in probability to μ as n increases without bound. That is, for any prescribed accuracy $\varepsilon > 0$, $P[-\varepsilon < \bar{Y} - \mu < \varepsilon]$ approaches unity as $n \to \infty$.

0,2 0,5

Proof. See [9]. ■

As a direct consequence of the law of large numbers, which says that each $\bar{X}_i$ converges in probability to μ_i, $i = 1, 2, \ldots, p$,

$$\bar{\mathbf{X}} \text{ converges in probability to } \boldsymbol{\mu} \tag{4-26}$$

Also, each sample covariance s_{ik} converges in probability to σ_{ik}, $i, k = 1, 2, \ldots, p$, and

$$\mathbf{S} \ (\text{or } \hat{\boldsymbol{\Sigma}} = \mathbf{S}_n) \text{ converges in probability to } \boldsymbol{\Sigma} \tag{4-27}$$