Inference with the multivariate normal distribution

Today's handout: 1) an excellent multivariate graphical summary 2) two examples where unbiasedness is a silly inferential criterion
3) A code simulating from multivariate distributions. Notice in this simulation that:
(a) the entries in $\frac{1}{n} \mathbb{M}$ (where $m \sim W_p(\mathbb{F}, \nu)$) are unbiased estimates of the entries of $\mathbb{F}$ for any $\nu > 0$
(b) the variability of the entries of $\frac{1}{n} \mathbb{M}$ around those of $\mathbb{F}$ goes to 0 at a $\sqrt{n}$ rate on the scale
standard Neyman (unbiased) inference

story \( Y = (x_1, \ldots, x_n) \sim N(\mu, \sigma^2) \)

\[ x \sim N(\mu, \sigma^2) \]

unbiased

inference about \( \mu \):

\[ (\bar{x} | \mu, \sigma^2) \sim N(\mu, \frac{\sigma^2}{n}) \]

point estimate \( \hat{\mu} \) of \( \mu \) is \( \hat{\mu} = \bar{x} \);

what about internal estimate? Need good estimate of \( \sigma^2 \) to assess uncertainty of \( \bar{x} \) as estimate of \( \mu \);

\[ S_n^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \]

is unbiased.

repeated sampling for \( \sigma^2 \) and \( \frac{(n-1)S_n^2}{\sigma^2} \sim X^2_{n-1} \)

From above, \( (\bar{x} - \mu | \sigma^2) \sim N(0, \frac{\sigma^2}{n}) \)

\[ \frac{\bar{x} - \mu}{\sigma/S_n} \sim N(0, 1) \] so if he knew

Neyman could use his confidence trick.
\[ \Pr \left[ -t^{*}(1-\frac{\alpha}{2}) < \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} < t^{*}(1-\frac{\alpha}{2}) \right] = 1 - \alpha \]

looks like a probability statement about \( \mu \) but isn't

\[ \Pr \left[ \mu \in \bar{x} \pm t^{*}(1-\frac{\alpha}{2}) \frac{\sigma}{\sqrt{n}} \right] \]

where \( t^{*}(\cdot) \) is the standard normal (CDF)

\[ \bar{x} \pm \frac{t^{*}(1-\frac{\alpha}{2}) \sigma}{\sqrt{n}} \]

is a 100(1-\( \alpha \))% confidence interval for \( \mu \) (Neyman, 1937)

Since we don't know \( \sigma \), obviously try to try is to use \( \hat{\sigma} = s_n \) and work out \( s_n \)-distribution of

\[ \frac{\bar{x} - \mu}{s_n / \sqrt{n}} \sim t_{n-1} \] (Student (1908), Fisher (1915))

or equivalently

\[ \left( \frac{\bar{x} - \mu}{s_n / \sqrt{n}} \right)^2 \sim F_{1, n-1} \]
(note that \( \frac{(\bar{x} - \mu)}{\sigma / \sqrt{n}} \) can be written \( (\bar{x} - \mu) (\sigma^2/n) \). From this \( x \pm t_{n-1} (1-\alpha/2) \frac{\sigma}{\sqrt{n}} \) 100(1-\alpha)% CI for \( \mu \).

Story (Fisher, 1925): first max likelihood

First \((x_1, \ldots, x_n | \mu, \sigma^2) \sim \mathcal{N}(\mu, \sigma^2) \) \((i = 1, \ldots, n)\)

Joint sampling distribution is

\[
p(x_1, \ldots, x_n | \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2} \right]
\]

\[
= (\sigma^2)^{-\frac{n}{2}} (2\pi)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right]
\]
so likelihood function is
\[ L(\mu, \sigma^2 | x) = c \sigma^{-n} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right) \]
and log likelihood function is
\[ \ell(\mu, \sigma^2 | x) = c - \frac{n}{2} \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \]
\[ \frac{\partial}{\partial \mu} \ell(\mu, \sigma^2 | x) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0 \]
\[ \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} \]
\[ \frac{\partial}{\partial \sigma} \ell(\mu, \sigma^2 | x) = - \frac{n}{2\sigma} + \frac{1}{2\sigma^3} \sum_{i=1}^{n} (x_i - \mu)^2 = 0 \]
\[ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = s^2 \text{ (biased)} \]
\[ \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = s^2 \text{ (unbiased)} \]
\[
\frac{1}{\sigma^2} \mu \frac{d^2}{d\mu^2} \mathcal{L}(\mu, \sigma^2 | x) = -\frac{5}{\sigma^2} \tag{Hessian}
\]

\[
\frac{1}{\sigma^2} \mu \frac{d^2}{d\mu^2} \mathcal{L}(\mu, \sigma^2 | x) = \frac{2}{\sigma^3} \sum_{i=1}^{n} (x_i - \mu) \sum_{i=1}^{n} (x_i - \mu)
\]

\[
\frac{1}{\sigma^2} \mu \frac{d^2}{d\sigma^2} \mathcal{L}(\mu, \sigma^2 | x) = \frac{5}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2
\]

\[\text{Fisher Information matrix is} \]

\[\mathbb{I} = \begin{pmatrix} \frac{5}{\sigma^2} & 0 \\ 0 & \frac{5}{2\sigma^4} \end{pmatrix} \]

This is \( \mathbb{I} = \begin{pmatrix} \frac{5}{\sigma^2} & 0 \\ 0 & \frac{5}{2\sigma^4} \end{pmatrix} \). Fisher showed that
If the model has parameter \( \theta = (\mu, \sigma^2) \) here, of least \( k \), then under mild regularity conditions, for large \( n \):

(a) \( \mathbb{E}_{\theta} (\hat{\theta}_{\text{MLE}}) = \theta + O(\frac{1}{n}) \) vs. \( \theta \) vs.

so \( \hat{\theta}_{\text{MLE}} \xrightarrow{\text{vs.}} \theta \) (consistency),

(b) \( \sqrt{n} (\hat{\theta}_{\text{MLE}} - \theta) \xrightarrow{\text{d}} N_k (0, \hat{I}^{-1}) \)

Here \( \hat{I}^{-1} = \begin{pmatrix} \hat{\sigma}^2 & 0 \\ 0 & \hat{\sigma}^2 \end{pmatrix} \) so \( \hat{\sigma}^2 \) large

sample vs. \( \mathbb{V}_{\theta} (\hat{\theta}_{\text{MLE}}) = \mathbb{V}_{\theta} (\hat{x}) = \frac{\hat{\sigma}^2}{n} \)

and \( \bar{x} \pm z^* \left( 1 - \frac{x}{2} \right) \frac{\sqrt{\hat{\sigma}^2}}{\sqrt{n}} \) is an approximate internal estimate for \( \mu \).
Bayesian inferential story (Laplace, 1780): if first

\[(\mu, \sigma^2) \sim p(\mu, \sigma^2)\]

same likelihood

\[(X_i | \mu, \sigma^2) \sim N(\mu, \sigma^2)\]

as with Fisher (of course);

one flexible class of priors for this model is the conjugate priors specified hierarchically (as mixture):

\[p(\mu, \sigma^2) = p(\sigma^2) p(\mu | \sigma^2)\]

\[p(\sigma^2)\] is an inverse Gamma, most easily worked

with as the scaled inverse \(\chi^2\) distribution:

\[\sigma^2 \sim \chi^2(\nu_0, \sigma_0^2)\]

\[= \Gamma^{-1}(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2})\]; this acts like a
prior dataset with \( n \) observations

and prior estimate \( \sigma^2 \) for \( \sigma^2 \)

conditional conjugate \( p(\mu | \sigma^2) \)

is normal: \( (\mu | \sigma^2) \sim N(\mu_0, \sigma^2 \sigma^{-2}/\kappa_0) \)

\( \mu_0 \) acts like a prior dataset for \( \mu \) given \( \sigma^2 \) with \( \kappa_0 \) observations,

scale \( \sigma \) (on the \( \sigma \) scale), and prior

estimate \( \mu_0 \)

the resulting joint prior

is \( p(\mu, \sigma^2) \sim \chi^2(\mu, \sigma^2, \kappa_0, \sigma_0^2) \)

with

density \( \frac{1}{\sigma_0^{-2}} \exp\left\{ -\frac{1}{2\sigma_0^2} \left[ \kappa_0 \sigma_0^{-2} + \kappa_0 (\mu_0 - \mu)^2 \right] \right\} \)

by conjugacy joint posterior also has to be \( N - \chi^2 \): write out
Bayes's Theorem \( p(\mu, \sigma^2 | x) \propto p(\mu | \sigma^2) p(\sigma^2) p(x | \mu, \sigma^2) \) & simplify to get

\[
p(\mu, \sigma^2 | x) \sim N - \chi^2 \left( \mu_n, \frac{\sigma^2}{\kappa_n} ; \kappa_n, \sigma^2 \right)
\]

with nicely intuitive updating rules:

\[
\begin{align*}
\kappa_n &= \kappa_0 + n \\
\mu_n &= \frac{\kappa_0 \mu_0 + n \bar{x}}{\kappa_0 + n} \\
\sigma_n^2 &= \sigma_0^2 + (n - 1) \bar{s}^2 + \frac{1}{\kappa_0 + \frac{1}{n}} \left( \bar{x} - \mu_0 \right)^2
\end{align*}
\]

approximate (weighted average of 3 sources of variability information: \( \sigma_0^2, \sigma^2, \) and \( (\bar{x} - \mu_0)^2 \)) & precise (weight)}
and now we can marginalize:

$$p(\mu | x) = \int_0^\infty p(\mu, \sigma^2 | x) \, d\sigma^2$$

to get the marginal posterior for \( \mu \), which turns out to be a scaled \( t \)-distribution:

$$p(\mu | x) \sim t_{\nu_n}(\mu | \mu_n, \frac{\sigma_n^2}{\nu_n})$$

$$= \frac{c}{\left[ 1 + \frac{\nu_n (\mu - \mu_n)^2}{\nu_n \sigma_n^2} \right]^{\nu_n + 1}}$$

so the posterior mean for \( \mu \) (a good point estimate) is \( \mu_n \) and a \( 100(1-\alpha) \% \) interval can be constructed from quantiles of the \( t_{\nu_n} \) distribution.
generalizing to $p > 1$}

\[ (X_1, \ldots, X_p) \sim N_p (\mu, \Sigma) \] for $i = 1, \ldots, n$

we already know that $(\bar{X} | \mu, \Sigma) \sim N_p (\mu, \frac{1}{n} \Sigma)$ so $\bar{X}$ is unbiased for $\mu$; how about uncertainty assessment? (p = 1) confidence set for $\mu$ was defined by

\[
Pr \left[ \left| \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \right| \leq t_{n-1} (1 - \frac{\alpha}{2}) \right] = 1 - \alpha
\]

or equivalently

\[
Pr \left[ \left( \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \right)^2 \leq F_{1, n-1} (1 - \frac{\alpha}{2}) \right] = 1 - \alpha
\]
The suggested re-write of \[
\left( \frac{\bar{x} - \mu}{s/\sqrt{n}} \right)^2
\] is \[n (\bar{x} - \mu) (s^2)^{-1} (\bar{x} - \mu)\], with \(n > 1\).

One obvious thing to do is to base confidence set on \[
\frac{1}{n} n (\bar{x} - \mu)^2 S_n^{-1} (\bar{x} - \mu)
\]

I hope that its distribution has something to do with the F dist.

---

Fact: in this model

\[
\left( \frac{n-p}{p(n-1)} \right)^{\frac{1}{2}} (\bar{x} - \mu)^{\frac{1}{2}} S_n^{-1} (\bar{x} - \mu) \sim F_{p, n-p}
\]
so a $100(1 - \alpha)\%$ confidence set for $\mu$ is defined by

$$P_{\alpha} \left[ \frac{1}{n} \left( \bar{x} - \mu \right) \left( \bar{x} - \mu \right)^T \leq \frac{(n - 1)}{n} \sigma^2 F_{\alpha}^{-1} (x) \right] = 1 - \alpha$$

What does this set look like?

density of $N_p(\mu, \Sigma)$ is

$$p(x) = c \exp \left[ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right]$$

level sets of this distribution (curves of constant density) are

(by inspection) of the form

$$\left( \frac{x - \mu}{\sigma} \right)^T \Sigma^{-1} \left( \frac{x - \mu}{\sigma} \right) = \kappa \geq 0$$
Given $\mu \approx \bar{x}$, for $p \geq 2$, the contours are ellipses: $N(k \theta)$ have plotted $p = 2$

for us (p. 39) with $\mu = (3, 3)$

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

for general $k \geq 2$ the level sets $(x - \bar{x})^T A^{-1} (x - \bar{x}) = c$, $A > 0$, are called ellipsoids (hard to visualize for $p > 2$).

Data example

from Johnson & Wichern, chapters 4-5.
microwave oven manufacturers have to test their products for radiation emitted both with open and closed doors. A random sample of 15 ovens from manufacturer X is:}

---

**Table 4.5 Radiation Data (Door Open)**

<table>
<thead>
<tr>
<th>Oven no</th>
<th>Radiation</th>
<th>Oven no</th>
<th>Radiation</th>
<th>Oven no</th>
<th>Radiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.30</td>
<td>16</td>
<td>.20</td>
<td>31</td>
<td>.10</td>
</tr>
<tr>
<td>2</td>
<td>.09</td>
<td>17</td>
<td>.04</td>
<td>32</td>
<td>.10</td>
</tr>
<tr>
<td>3</td>
<td>.30</td>
<td>18</td>
<td>.10</td>
<td>33</td>
<td>.10</td>
</tr>
<tr>
<td>4</td>
<td>.10</td>
<td>19</td>
<td>.01</td>
<td>34</td>
<td>.30</td>
</tr>
<tr>
<td>5</td>
<td>.10</td>
<td>20</td>
<td>.60</td>
<td>35</td>
<td>.12</td>
</tr>
<tr>
<td>6</td>
<td>.12</td>
<td>21</td>
<td>.12</td>
<td>36</td>
<td>.25</td>
</tr>
<tr>
<td>7</td>
<td>.09</td>
<td>22</td>
<td>.10</td>
<td>37</td>
<td>.20</td>
</tr>
<tr>
<td>8</td>
<td>.10</td>
<td>23</td>
<td>.05</td>
<td>38</td>
<td>.40</td>
</tr>
<tr>
<td>9</td>
<td>.09</td>
<td>24</td>
<td>.05</td>
<td>39</td>
<td>.33</td>
</tr>
<tr>
<td>10</td>
<td>.10</td>
<td>25</td>
<td>.15</td>
<td>40</td>
<td>.32</td>
</tr>
<tr>
<td>11</td>
<td>.07</td>
<td>26</td>
<td>.30</td>
<td>41</td>
<td>.12</td>
</tr>
<tr>
<td>12</td>
<td>.05</td>
<td>27</td>
<td>.15</td>
<td>42</td>
<td>.12</td>
</tr>
<tr>
<td>13</td>
<td>.01</td>
<td>28</td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>.45</td>
<td>29</td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>.12</td>
<td>30</td>
<td>.28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Data courtesy of J. D. Cryer.

---

**Table 4.1 Radiation Data (Door Closed)**

<table>
<thead>
<tr>
<th>Oven no</th>
<th>Radiation</th>
<th>Oven no</th>
<th>Radiation</th>
<th>Oven no</th>
<th>Radiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.15</td>
<td>16</td>
<td>.10</td>
<td>31</td>
<td>.10</td>
</tr>
<tr>
<td>2</td>
<td>.09</td>
<td>17</td>
<td>.02</td>
<td>32</td>
<td>.20</td>
</tr>
<tr>
<td>3</td>
<td>.18</td>
<td>18</td>
<td>.10</td>
<td>33</td>
<td>.11</td>
</tr>
<tr>
<td>4</td>
<td>.10</td>
<td>19</td>
<td>.01</td>
<td>34</td>
<td>.30</td>
</tr>
<tr>
<td>5</td>
<td>.05</td>
<td>20</td>
<td>.40</td>
<td>35</td>
<td>.02</td>
</tr>
<tr>
<td>6</td>
<td>.12</td>
<td>21</td>
<td>.10</td>
<td>36</td>
<td>.20</td>
</tr>
<tr>
<td>7</td>
<td>.08</td>
<td>22</td>
<td>.05</td>
<td>37</td>
<td>.20</td>
</tr>
<tr>
<td>8</td>
<td>.05</td>
<td>23</td>
<td>.03</td>
<td>38</td>
<td>.30</td>
</tr>
<tr>
<td>9</td>
<td>.08</td>
<td>24</td>
<td>.05</td>
<td>39</td>
<td>.30</td>
</tr>
<tr>
<td>10</td>
<td>.10</td>
<td>25</td>
<td>.15</td>
<td>40</td>
<td>.40</td>
</tr>
<tr>
<td>11</td>
<td>.07</td>
<td>26</td>
<td>.10</td>
<td>41</td>
<td>.30</td>
</tr>
<tr>
<td>12</td>
<td>.02</td>
<td>27</td>
<td>.15</td>
<td>42</td>
<td>.05</td>
</tr>
<tr>
<td>13</td>
<td>.01</td>
<td>28</td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>.10</td>
<td>29</td>
<td>.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>.10</td>
<td>30</td>
<td>.18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Data courtesy of J. D. Cryer.

---

they don't tell us the units of measurement, but presumably higher # = more radiation.

Sanity check:
- radiation - open
- radiation - closed

(Not true)
Can we use sampling model \( p = 4 \) \( \chi_{1 \mu, \sigma} \sim N_2 (\mu, \sigma) \) 

(Conditionally) IID

OK but 93 %

(Open) shows long right-

hand tail

Shape:

Standard fix: transform to approximate normality: Box-Cox power transform family \( x + \frac{x^p - 1}{p} \) for some power \( p \), with \( x + \ln(x) \) for \( p = 0 \); can experiment & find that with this data set \( p = 0.25 \) works well:
Two natural forms government regulation could take: ① Ensure that a high percentage (say 99%) of the microwave oven produced emit no more than \( c_1 \) of radiation (door closed) & \( c_2 \) (door open) with high probability (say 95%); ② Ensure that the (population) mean amount of radiation (door) \( \mu_x \leq c_1 \) and \( \mu_y \leq c_2 \) with high (\( 95\% \)) prob.
Example 5.3 (Constructing a confidence ellipse for \( \mu \)) Data for radiation from microwave ovens were introduced in Examples 4.10 and 4.17. Let

\[ x_1 = \sqrt{\text{measured radiation with door closed}} \]

and

\[ x_2 = \sqrt{\text{measured radiation with door open}} \]

For the \( n = 42 \) pairs of transformed observations, we find that

\[ \bar{x} = \begin{bmatrix} .564 \\ .603 \end{bmatrix}, \quad S = \begin{bmatrix} .0144 & .0117 \\ .0117 & .0146 \end{bmatrix}, \]

\[ S^{-1} = \begin{bmatrix} 203.018 & -163.391 \\ -163.391 & 200.228 \end{bmatrix} \]

The eigenvalue and eigenvector pairs for \( S \) are

\[ \lambda_1 = .026, \quad e_1 = [ .704, .710 ] \]

\[ \lambda_2 = .002, \quad e_2 = [ -.710, .704 ] \]

The 95% confidence ellipse for \( \mu \) consists of all values \((\mu_1, \mu_2)\) satisfying

\[ 42[.564 - \mu_1, .603 - \mu_2] \begin{bmatrix} 203.018 & -163.391 \\ -163.391 & 200.228 \end{bmatrix} \begin{bmatrix} .564 - \mu_1 \\ .603 - \mu_2 \end{bmatrix} \]

\[ \leq \frac{2(41)}{40} F_{2,40}(.05), \]

or, since \( F_{2,40}(.05) = 3.23, \)

\[ 42(203.018)(.564 - \mu_1)^2 + 42(200.228)(.603 - \mu_2)^2 \]

\[ - 84(163.391)(.564 - \mu_1)(.603 - \mu_2) \leq 6.62 \]

To see whether \( \mu' = [.562, .589] \) is in the confidence region, we compute

\[ 42(203.018)(.564 - .562)^2 + 42(200.228)(.603 - .589)^2 \]

\[ - 84(163.391)(.564 - .562)(.603 - .589) = 1.30 \leq 6.62 \]

We conclude that \( \mu' = [.562, .589] \) is in the region. Equivalently, a test of \( H_0: \mu = [.562, .589] \) would not be rejected in favor of \( H_1: \mu \neq [.562, .589] \) at the \( \alpha = .05 \) level of significance.
The joint confidence ellipsoid is plotted in Figure 5.1. The center is at \( \bar{x}' = [0.564, 0.603] \), and the half-lengths of the major and minor axes are given by

\[
\sqrt{\lambda_1} \frac{p(n-1) F_{p,n-p}(\alpha)}{n(n-p)} = \sqrt{0.026} \frac{2(41)}{42(40)} (3.23) = 0.064
\]

and

\[
\sqrt{\lambda_2} \frac{p(n-1) F_{p,n-p}(\alpha)}{n(n-p)} = \sqrt{0.002} \frac{2(41)}{42(40)} (3.23) = 0.018
\]

respectively. The axes lie along \( e'_1 = [0.704, 0.710] \) and \( e'_2 = [-0.710, 0.704] \) when these vectors are plotted with \( \bar{x} \) as the origin. An indication of the elongation of the confidence ellipse is provided by the ratio of the lengths of the major and minor axes. This ratio is

\[
\frac{2\sqrt{\lambda_1} \sqrt{p(n-1) F_{p,n-p}(\alpha)}}{2\sqrt{\lambda_2} \sqrt{p(n-1) F_{p,n-p}(\alpha)}} = \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} = \frac{161}{1045} = 3.6
\]

**Figure 5.1** A 95% confidence ellipse for \( \mu \) based on microwave-radiation data.

The length of the major axis is 3.6 times the length of the minor axis.

So what they have shown is that anybody who thought the population variance of the variable validation 0.25 (closed) and validation 0.25 (open) were \((0.562, 0.589)\) might well be right; however there's a problem with interpreting results from a data-transformed analysis.
it may well be that \( x \sim N \left( \mu, \sigma^2 \right) \), but unfortunately

\[ E(y) = E(x^{1/4}) \neq \left[ E(x) \right]^{1/4} \]

in fact, by Jensen's inequality:

\[
\begin{cases}
\text{convex} & \Rightarrow \text{bowl shaped up} \\
\text{linear} & \Rightarrow \text{flat}
\end{cases}
\]

\[
\begin{cases}
\text{concave} & \Rightarrow \text{bowl shaped down} \\
\quad \quad \quad \quad \quad \quad \quad \quad \text{if low}
\end{cases}
\]

we will have that \( \left[ E(x) \right]^{1/4} \geq E(x^{1/4}) \)

since \( f(x) = x^{1/4} \) is concave; this means that more work is needed to get the confidence-ellipsoid result back onto the radiation source (later)
The likelihood story ($p > 1$) we know already that if $(x_1, \ldots, x_i, \ldots)$ has

$$p(x_i) = \frac{1}{(2\pi) \frac{1}{2}} \exp \left\{ -\frac{1}{2} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \right\}$$

so the joint sampling distribution is

$$p(x_1, \ldots, x_i, \ldots) \propto \prod_{i=1}^{n} \left\{ (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \right\} \right\}$$

and the likelihood function is thus

$$L(\mu, \Sigma | \mathbf{x}) \propto \prod_{i=1}^{n} \exp \left\{ -\frac{1}{2} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \right\}$$

we write this as

$$L(\mathbf{x}; \mu, \Sigma) = \left| 2\pi \Sigma \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right\}$$
(hence revealing themselves as frequentist at heart), the log-likelihood function is then
\[ L(\mu, \Sigma) = -\frac{n}{2} \ln |2\pi \Sigma| \]
\[ -\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)' \Sigma^{-1} (x_i - \mu) \]

We show that this can be simplified to yield
\[ L(\mu, \Sigma | x) = -\frac{n}{2} \ln |2\pi \Sigma| \]
\[ -\frac{n}{2} \ln (\Sigma') - \frac{n}{2} \ln (\Sigma) - \frac{1}{2} (\Sigma^{-1} - \Sigma') (\Sigma^{-1} - \Sigma) \]
from which it is immediately clear that \((\bar{x}, \Sigma')\) forms a sufficient statistic for \((\mu, \Sigma)\) (surprise).
extra notes

\( \text{rvish (newpack)} \)

\text{library (in v-t norm)}

\text{rvish}

\[ x \sim N(\mu, \sigma^2) \]

\[ x - \mu \sim N(0, \sigma^2/n) \]

\[ \frac{x - \mu}{\sigma / \sqrt{n}} \sim N(0, 1) \]
Bayesian ($\rho = 1$)

\[ \Theta = (\mu, \sigma^2) \]

\[ \left\{ \left( x_i \mid \mu, \sigma^2 \right) \overset{i.i.d.}{\sim} N(\mu, \sigma^2) \right\}_{i=1, \ldots, n} \]

\[ (\mu, \sigma^2) \sim \pi(\mu, \sigma^2) \]

\[ \pi(\mu, \sigma^2) = \pi(\sigma^2) \pi(\mu \mid \sigma^2) \]

\[ \overset{\chi^2}{\sim} N \text{- mixture} \]

\[ (\mu, \sigma^2) \leftrightarrow \left\{ \left( x_i \mid \sigma^2 \right) \right\} \]

\[ h(\mu \mid \sigma^2 | X) = N - \chi^2 \]

\[ p(\mu | X) = t - \]

\[ p(\mu | X) = t - \]
\[ E(x^{1/4}) = \left[ E(x) \right]^{1/4} \]