University of Southampton

SESA6085 – Advanced Aerospace Engineering Management

Tutorial 2

2024-2025

Dr David Toal

University of Southampton

Q3.1

Q3.1

- Use maximum likelihood estimation to fit a log-normal PDF to the failure data provided
- Calculate the 95% confidence bounds in the MLE estimates
- Calculate the probability that the component will fail between 30 and 55 years.

49.5742	39.0231	28.241	15.4782	16.6129	56.0018	39.1695	23.3995	9.34463
25.07	16.3242	7.1327	24.5379	95.74	13.3444	45.9901	22.8084	45.0592
49.9091	10.9584	22.0764	20.182	19.9721	14.9876	12.6724	17.3804	5.13221
21.0974	13.3389	12.4727	14.2646	17.2633	14.3847	17.5576	25.6905	7.18939
33.0662	19.2535	13.5511	20.9923	16.0762	24.9305	12.2074	5.23867	18.3745
16.6716	40.6632	18.8055	12.6727	11.24	21.6383	14.7064	17.601	15.0596
32.7493	18.8793	15.8212	6.69836	11.7989	27.7205	18.9279	72.2692	8.4983
18.6928	13.4246	38.2967	32.851	13.5905	30.0529	15.1524	60.4526	18.3166
5.44084	6.10256	20.3197	8.4967	35.7099	11.3333	31.722	18.1207	11.6619
15.0325	25.5701	20.1378	26.1686	12.1678	38.3752	13.6005	6.39596	10.8213

A3.1

- Recall the general expression for the likelihood equation

$$L(\theta) = \prod_{i=1}^{n} f(t_i;\theta)$$

- For our log-normal distribution the PDF is

$$f(t;\mu,\sigma) = \frac{1}{t\sigma(2\pi)^{1/2}} \exp\left[-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2\right]$$

- The natural log of the PDF is

$$\ln(f) = -\ln(t) - \ln(\sigma) - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2$$

A3.1

- The log-likelihood equation is therefore

$$l(\mu, \sigma) = \sum_{i=1}^{n} \ln(f(t; \mu, \sigma))$$

$$l(\mu, \sigma) = -n\ln(\sigma) - \frac{n}{2}\ln(2\pi) - \sum_{i=1}^{n}\ln(t_i) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln(t_i) - \mu)^2$$

- With derivatives of

$$\frac{\partial l}{\partial \mu} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(\mu - \ln(t_i))$$

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(\ln(t_i) - \mu)^2$$

A3.1

- Setting the derivatives equal to zero, the MLEs become

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \ln(t_i)$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\ln(t_i) - \hat{\mu})^2}$$

- What do you notice about these?
  - They are the mean and standard deviation of the natural log of the data
  - The log normal is therefore directly linked to the normal distribution

A3.1

- Using these equations and the given data we can fit a log-normal PDF

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \ln(t_i) = 2.916$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\ln(t_i) - \hat{\mu})^2} = 0.593$$

- N.B. be careful using standard deviation functions in Matlab, Excel, Python etc. often these use ${}^{1}/_{n-1}$ by default instead of ${}^{1}/_{n}$

A3.1

- To calculate the confidence bounds we need to determine an expression for the variance in the MLE parameters
- We do this by determining the Fisher information matrix

$$I_{ij} = E\left[-\frac{\partial^2 l(t;\theta)}{\partial\theta_i \partial\theta_j}\right]$$

- Whose inverse defines the covariance matrix

$$I^{-1} = \begin{bmatrix} \mathrm{Var}(\theta_1) & \mathrm{Cov}(\theta_1,\theta_2) & \cdots & \mathrm{Cov}(\theta_1,\theta_k) \\ \mathrm{Cov}(\theta_2,\theta_1) & \mathrm{Var}(\theta_2) & \cdots & \mathrm{Cov}(\theta_2,\theta_k) \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}(\theta_k,\theta_1) & \mathrm{Cov}(\theta_k,\theta_2) & \cdots & \mathrm{Var}(\theta_k) \end{bmatrix}$$

A3.1

- We know our first derivatives, from these we can calculate the second derivatives

$$\frac{\partial^2 l}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2 l}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}(\ln(t_i) - \mu)^2$$

$$\frac{\partial^2 l}{\partial \mu \partial \sigma} = \frac{2}{\sigma^3}\sum_{i=1}^{n}(\mu - \ln(t_i))$$

- These are the same as the normal distribution derivatives but with $\ln(t_i)$ instead of t_i

A3.1

- Defining the negative of the expected values

$$E\left[-\frac{\partial^2 l}{\partial \mu^2}\right] = \frac{n}{\hat{\sigma}^2}$$

$$E\left[-\frac{\partial^2 l}{\partial \sigma^2}\right] = \frac{2n}{\hat{\sigma}^2}$$

$$E\left[-\frac{\partial^2 l}{\partial \mu \partial \sigma}\right] = 0$$

- The information and correlation matrices are therefore

$$I = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & 2n/\sigma^2 \end{bmatrix} \qquad \& \qquad I^{-1} = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & \sigma^2/2n \end{bmatrix}$$

- With $\mathrm{Var}(\hat{\mu}) = \hat{\sigma}^2/n \qquad \mathrm{Var}(\hat{\sigma}) = \hat{\sigma}^2/2n$

A3.1

- Using the data and MLE the variances are

$$\mathrm{Var}(\hat{\mu}) = 0.00390$$

$$\mathrm{Var}(\hat{\sigma}) = 0.00195$$

- With the 95% bounds ($Z_{\alpha/2} = 1.96$) then

$$2.79 \leq \hat{\mu} \leq 3.04$$

$$0.51 \leq \hat{\sigma} \leq 0.68$$

A3.1

- Given the MLEs the probability of failure is simple to calculate

$$P_{\text{Failure}} = F(55) - F(30)$$

$$P_{\text{Failure}} = 0.174$$

- With the CDF easily calculated used Excel, Matlab or Python

Q3.2

Q3.2

- 15 components are tested, 12 of which have failure times recorded while 3 did not fail during the 24 hours in which the experiment was run
- Failures are assumed to follow an exponential distribution
- The recorded failure times are at:

t	5.9653	2.3421	0.7956	0.1604	0.0151	3.0238
	4.2605	1.2375	8.6255	6.8242	0.7645	5.8619

A3.2

- Recall our PDF and reliability functions for an exponential distribution

$$f(t) = \lambda e^{-\lambda t} \qquad\qquad R(t) = e^{-\lambda t}$$

- Our general expression for the likelihood function is:

$$L(\lambda) = \left\{\prod_{i=1}^{n_U} \lambda e^{-\lambda t_i}\right\}\left\{\prod_{i=1}^{n_{C_R}} e^{-\lambda T_R}\right\}$$

- Where n_U and n_{C_R} are the number of uncensored and right censored data points respectively

A3.2

- Taking natural logs of this function:

$$l(\lambda) = n_U \ln(\lambda) - \lambda \sum_{i=1}^{n_U} t_i - n_{C_R} \lambda T_R$$

- With the derivative with respect to λ given by:

$$\frac{\partial l(\lambda)}{\partial \lambda} = \frac{n_U}{\lambda} - \sum_{i=1}^{n_U} t_i - n_{C_R} T_R$$

- Where $n_U + n_{C_R} = n$

A3.2

- Setting this derivative equal to zero and rearranging gives us an expression for the right censored MLE for an exponential distribution

$$\frac{n_U}{\lambda} - \sum_{i=1}^{n_U} t_i - n_{C_R} T_R = 0$$

$$\lambda = \frac{n_U}{\sum_{i=1}^{n_U} t_i + n_{C_R} T_R}$$

- Which when used with our tabulated data gives:

$$\lambda = 0.1073$$

Q3.3

Q3.3

- A technician performs a series of 25 failure tests on a component. The tests are run for a total of 7 hours. At the end of the test 1 component has not yet failed. The technician also did not record an exact failure time for 4 components which failed during the first 0.25 hours of the test.

0.2967	0.9752	1.6013	3.1594	4.6646
0.3594	1.3873	1.6849	3.4113	5.0373
0.4060	1.4796	2.2832	3.6384	5.0554
0.5223	1.5322	2.7337	3.7169	6.4226

- Calculate the maximum likelihood estimate of λ assuming that the distribution is exponential in nature.

A3.3

- Now we have both left and right-censored data

$$f(t) = \lambda e^{-\lambda t} \qquad F(t) = 1 - e^{-\lambda t} \qquad R(t) = e^{-\lambda t}$$

- Our expression for the likelihood function becomes…

$$L(\lambda) = \left\{\prod_{i=1}^{n_U} \lambda e^{-\lambda t_i}\right\}\left\{\prod_{i=1}^{n_{C_R}} e^{-\lambda T_R}\right\}\left\{\prod_{i=1}^{n_{C_L}} 1 - e^{-\lambda T_L}\right\}$$

- Which after taking natural logs becomes…

$$l(\lambda) = n_U \ln(\lambda) - \lambda \sum_{i=1}^{n_U} t_i + n_{C_L} \ln\left(1 - e^{-\lambda T_L}\right) - n_{C_R} \lambda T_R$$

A3.3

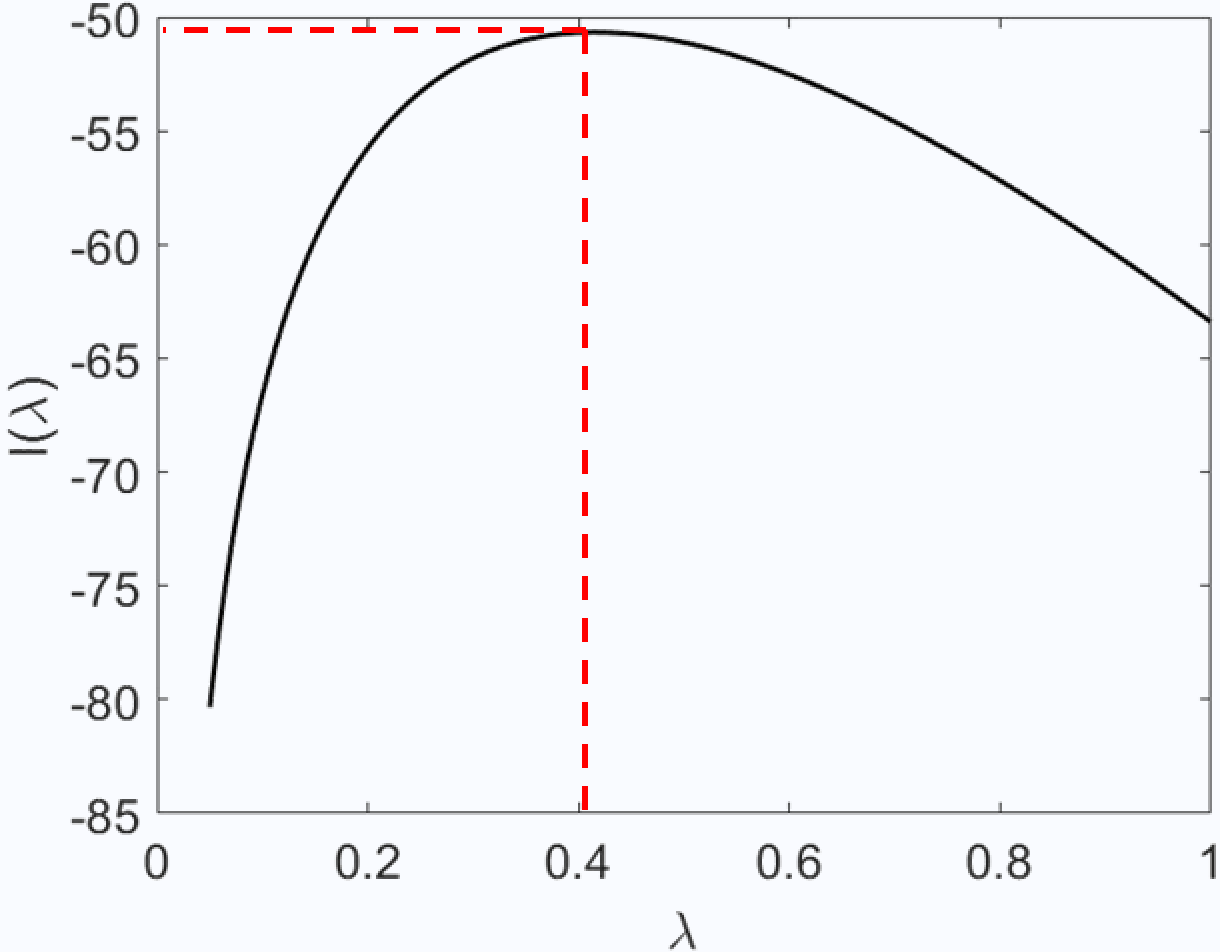
- We can attempt to take partial derivatives with respect to λ

$$\frac{\partial l(\lambda)}{\partial \lambda} = \frac{n_U}{\lambda} - \sum_{i=1}^{n_U} t_i - n_{C_R} T_R + \frac{n_{C_L} T_L \exp(-\lambda T_L)}{1 - \exp(-\lambda T_L)}$$

- But what do we notice?
- There is no closed form solution for $\hat{\lambda}$ when this is set to zero
- How do we solve for this?

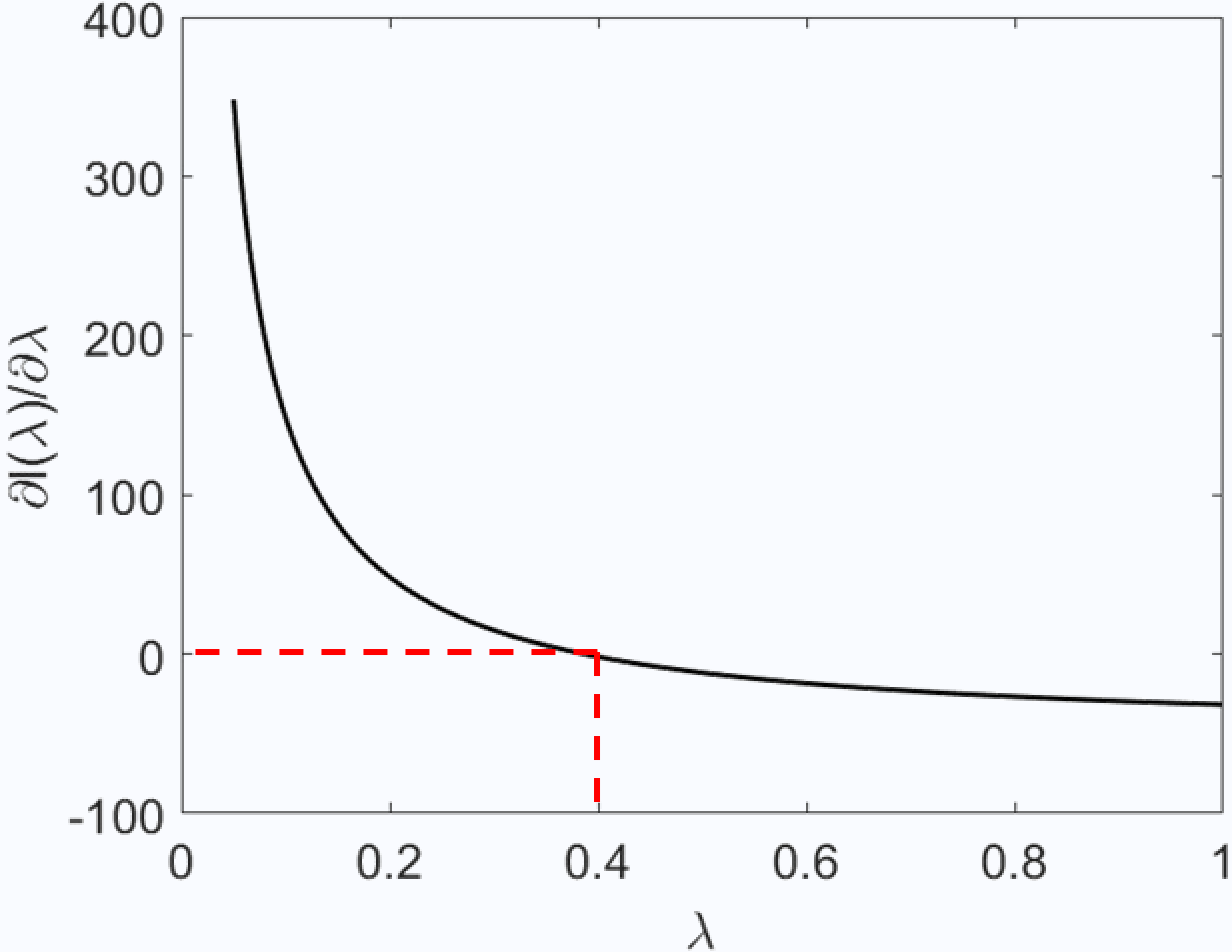
A3.3

- Graphically…



A3.3

- Alternative graphical approach…



A3.3

- A more general approach is to employ an optimisation algorithm
- The log-likelihood equation can be defined in Excel
  - One cell defining λ
  - Another cell defining $l(\lambda)$
  - Excel's "solver" can be used to maximise $l(\lambda)$ by changing λ

$$\hat{\lambda} = 0.4148$$

- N.B. sometimes the function $l(\lambda)$ can be highly multi-modal so the choice of optimisation algorithm can be important

Q4.1

Q4.1

- Consider the following set of observations
- Fit a 2-dimensional normal distribution to the data set
- What are the steps?

1. Calculate the mean for x_1 and x_2
2. Calculate the covariance matrix using this mean

x_1	x_2
0.8265	1.7644
0.8576	1.6610
1.7552	2.3293
1.0233	2.3513
1.3421	1.5131
0.8821	2.2614
0.7454	1.7379
1.0525	1.1175
1.2380	1.3376
1.0518	2.3298
1.5904	1.8017
1.0710	1.8711
0.9222	1.1830
0.6637	2.0190
1.1401	1.8232

A4.1

- Let's use Matlab for this calculation

1. Import the data into the variables `x1` and `x2` in Matlab

```
x1=[0.8265;0.8576;1.7552;1.0233…];
x2=[1.7644;1.6610;2.3293;2.3513…];
```

2. Calculate and store the mean of each dataset

```
Mu(1) = mean(x1);
Mu(2) = mean(x2);
```

A4.1

3. With these mean values calculate the covariance matrix

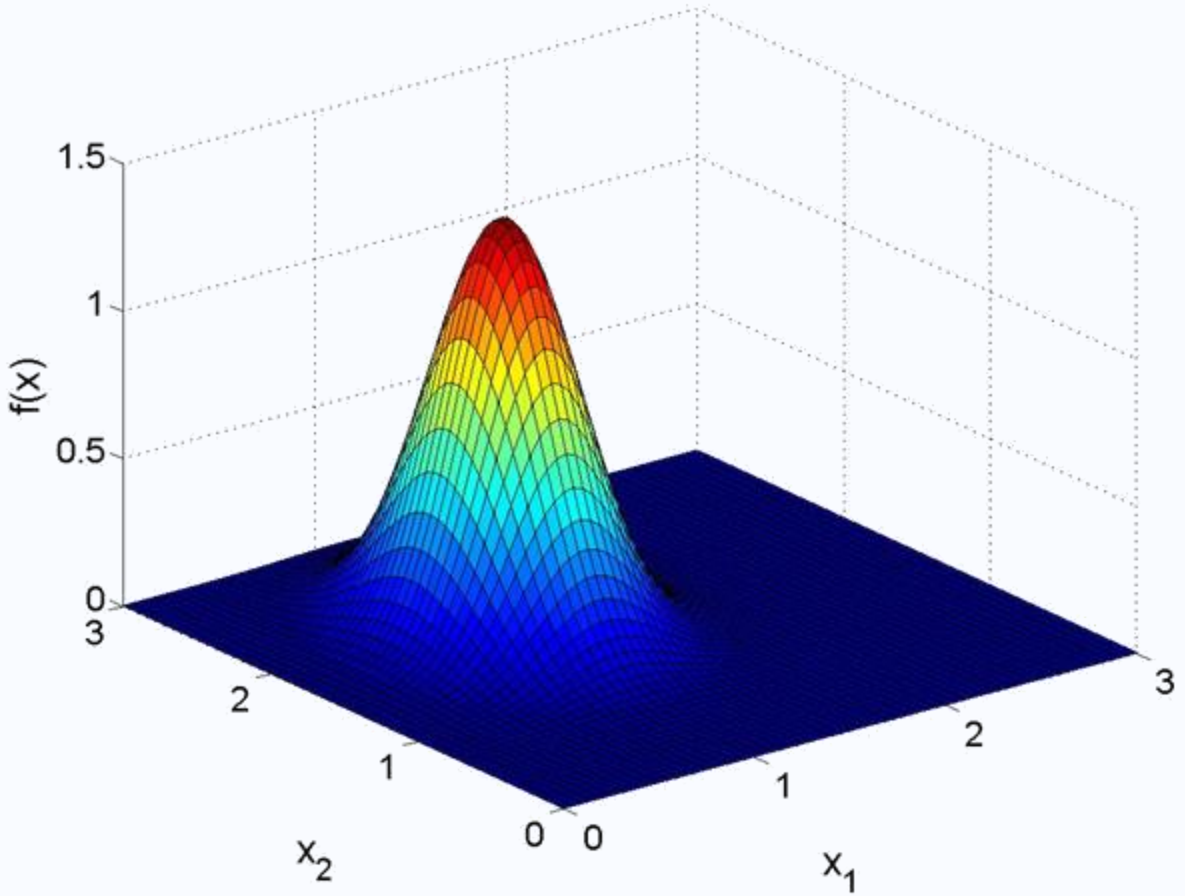
```
Sigma(1,1)=((x1-Mu(1))'*(x1-Mu(1)))/15
Sigma(2,2)=((x2-Mu(2))'*(x2-Mu(2)))/15
Sigma(1,2)=((x1-Mu(1))'*(x2-Mu(2)))/15
Sigma(2,1)= Sigma(1,2)
```

4. This gives us the following distribution

$$\boldsymbol{\mu} = \begin{bmatrix} 1.0775 \\ 1.8067 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 0.0854 & 0.0106 \\ 0.0106 & 0.1532 \end{bmatrix}$$

A4.1

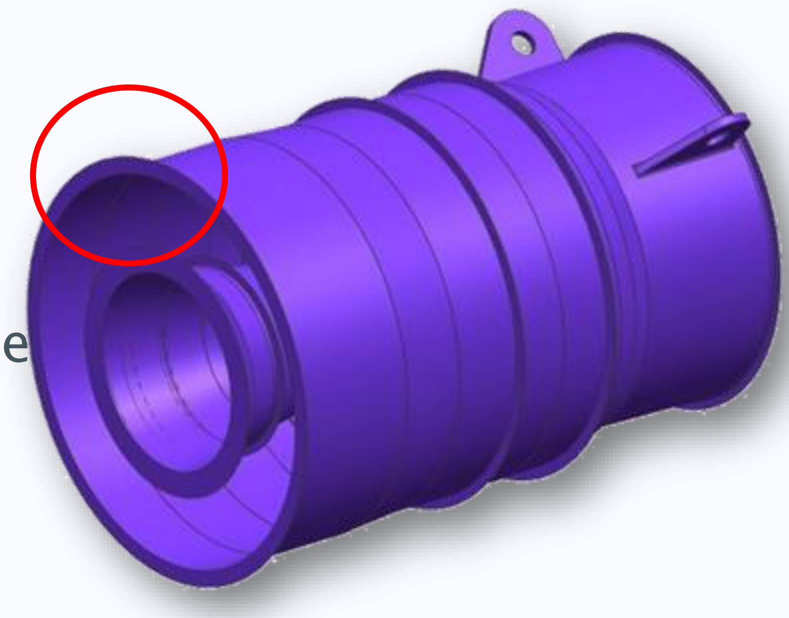
5. We can then use Matlab to plot the resulting PDF



Q4.2

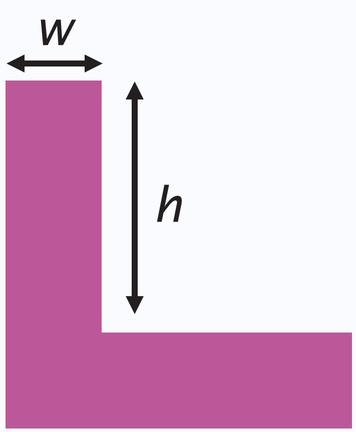
Q4.2

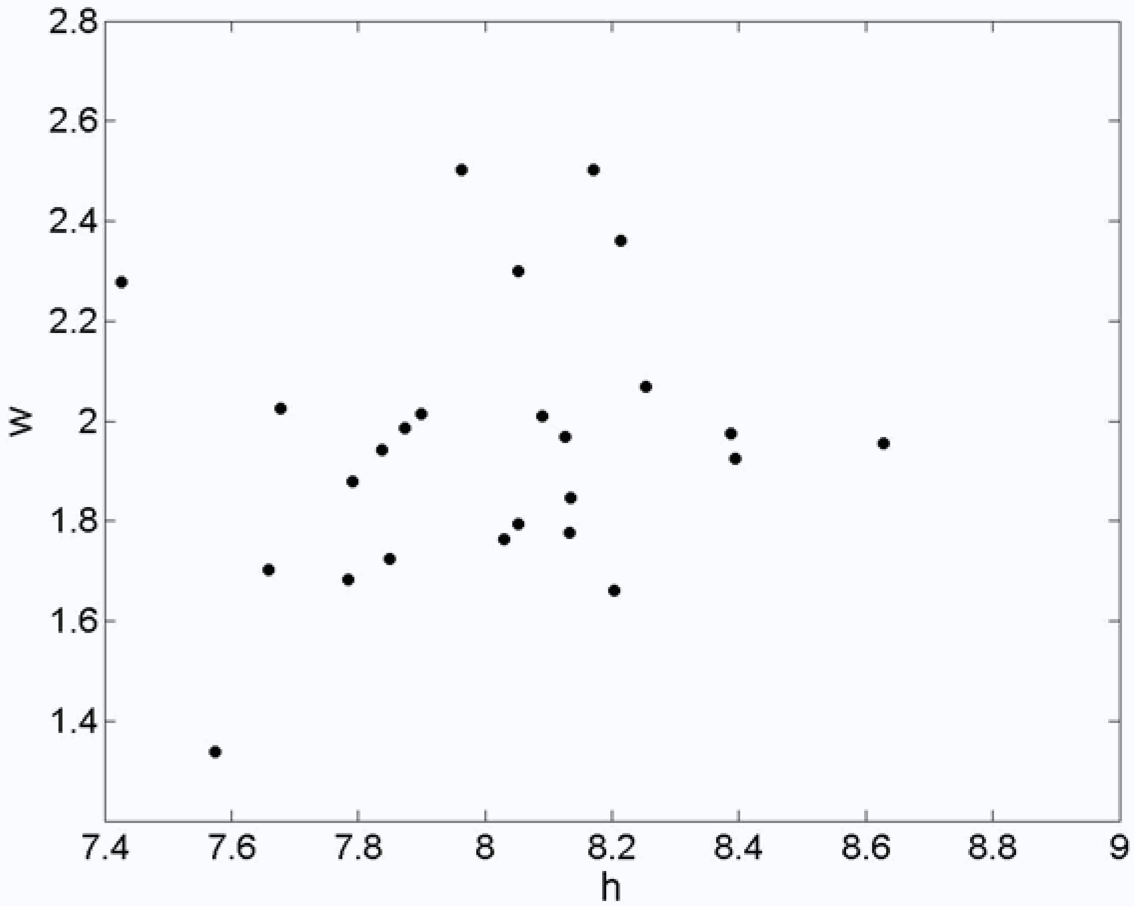
- Let’s consider the manufacture of a casing for a gas turbine
- Specifically, the milling of the highlighted flange
- Upon manufacture measurements are taken of the width and height of a series of casings



A4.2

- The measurement data is found to be distributed as follows when plotted





A4.2

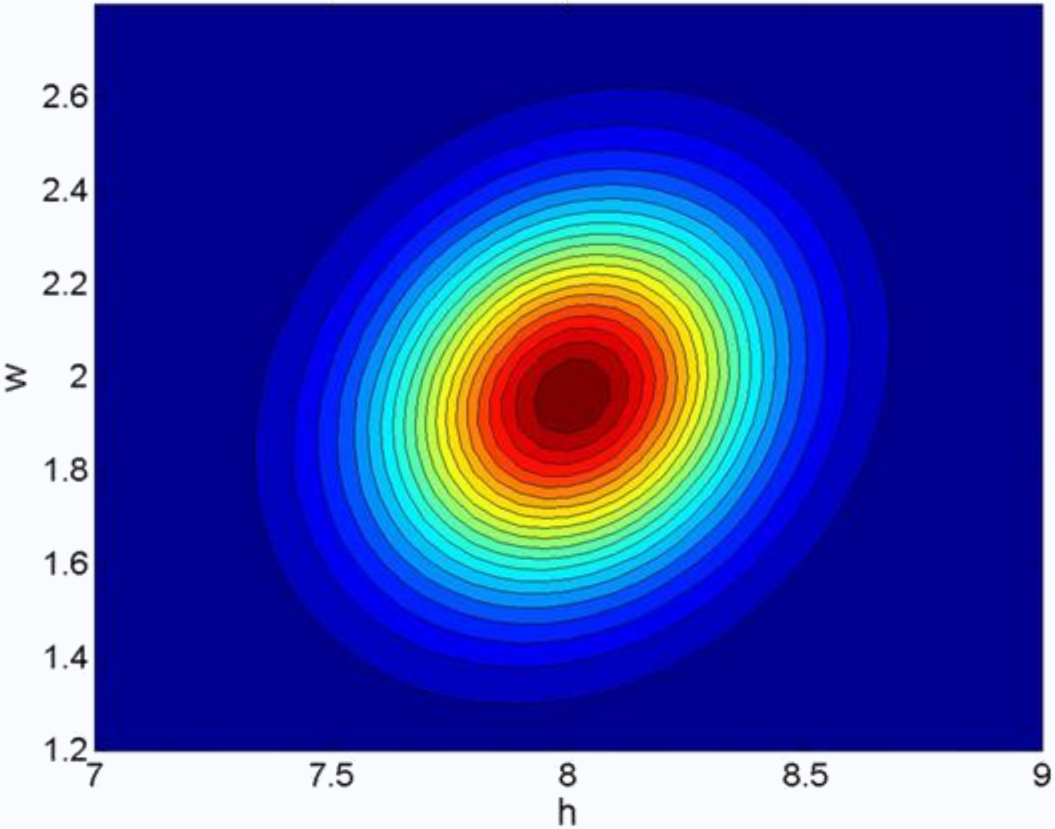
- If the flange width falls below 1.4mm or the flange height falls below 7.5mm then the casing has to be scrapped
- If the flange width is above 2.4mm or the flange height is above 8.5mm then the casing has to be machined again
- Calculate the probability of a casing being scrapped
- Calculate the probability of a casing being reworked
- How would you proceed?

A4.2

- Assuming a two dimensional normal distribution fitting our measurement data gives us the following:

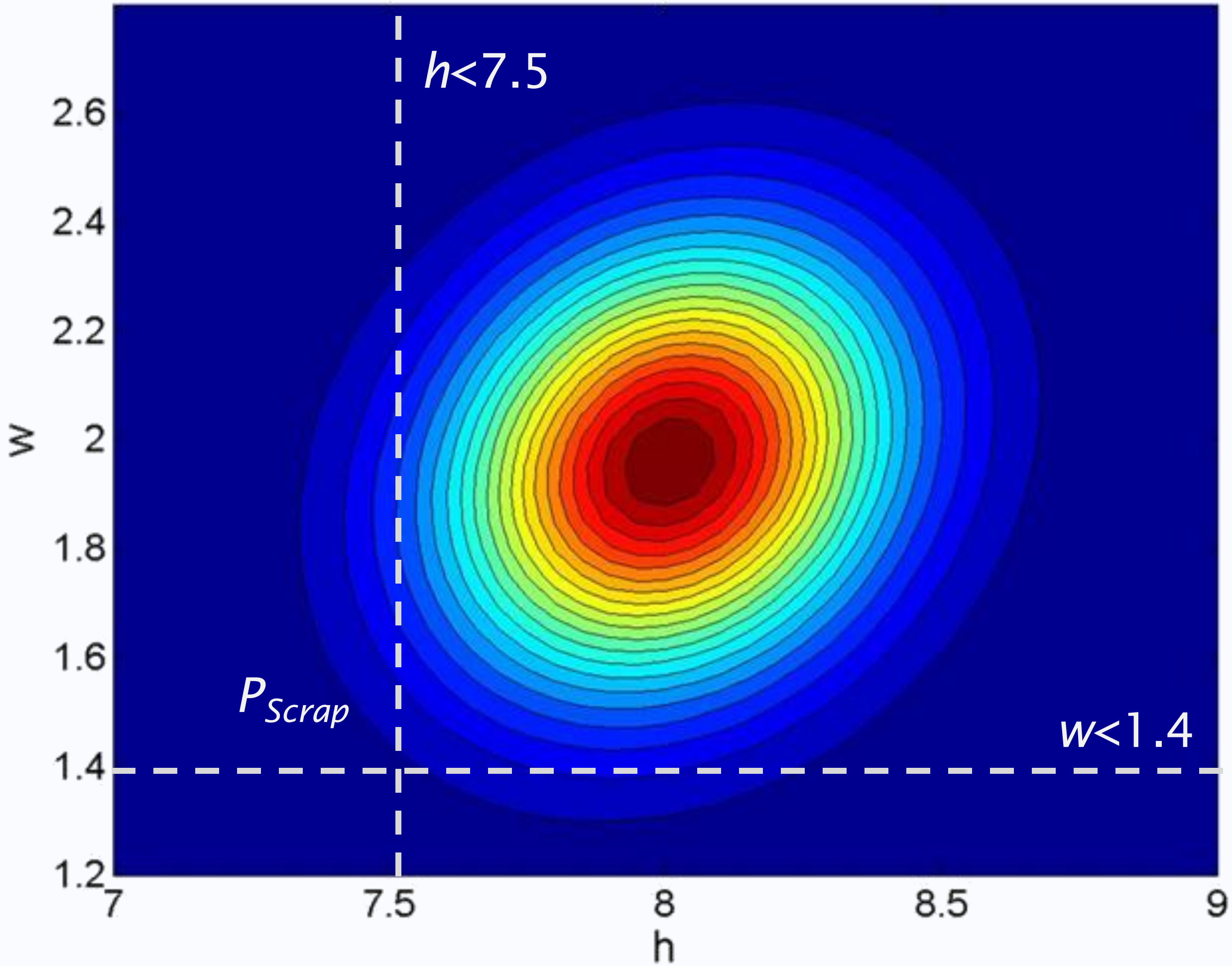
$$\boldsymbol{\mu} = \begin{bmatrix} 8.0081 \\ 1.9591 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.0737 & 0.0136 \\ 0.0136 & 0.0714 \end{bmatrix}$$



A4.2

- What is the probability of scrap?



A4.2

- The probability of scrap is therefore:

$$P_{scrap} = F(7.5, \infty) + F(\infty, 1.4) - \mathrm{F}(7.5,1.4)$$

- Using Matlab's `mvncdf()` function‡

```
P_scrap = mvncdf([7.5,inf],Mu,Sigma)+mvncdf([inf,1.4],Mu,Sigma)-
mvncdf([7.5,1.4],Mu,Sigma)
```

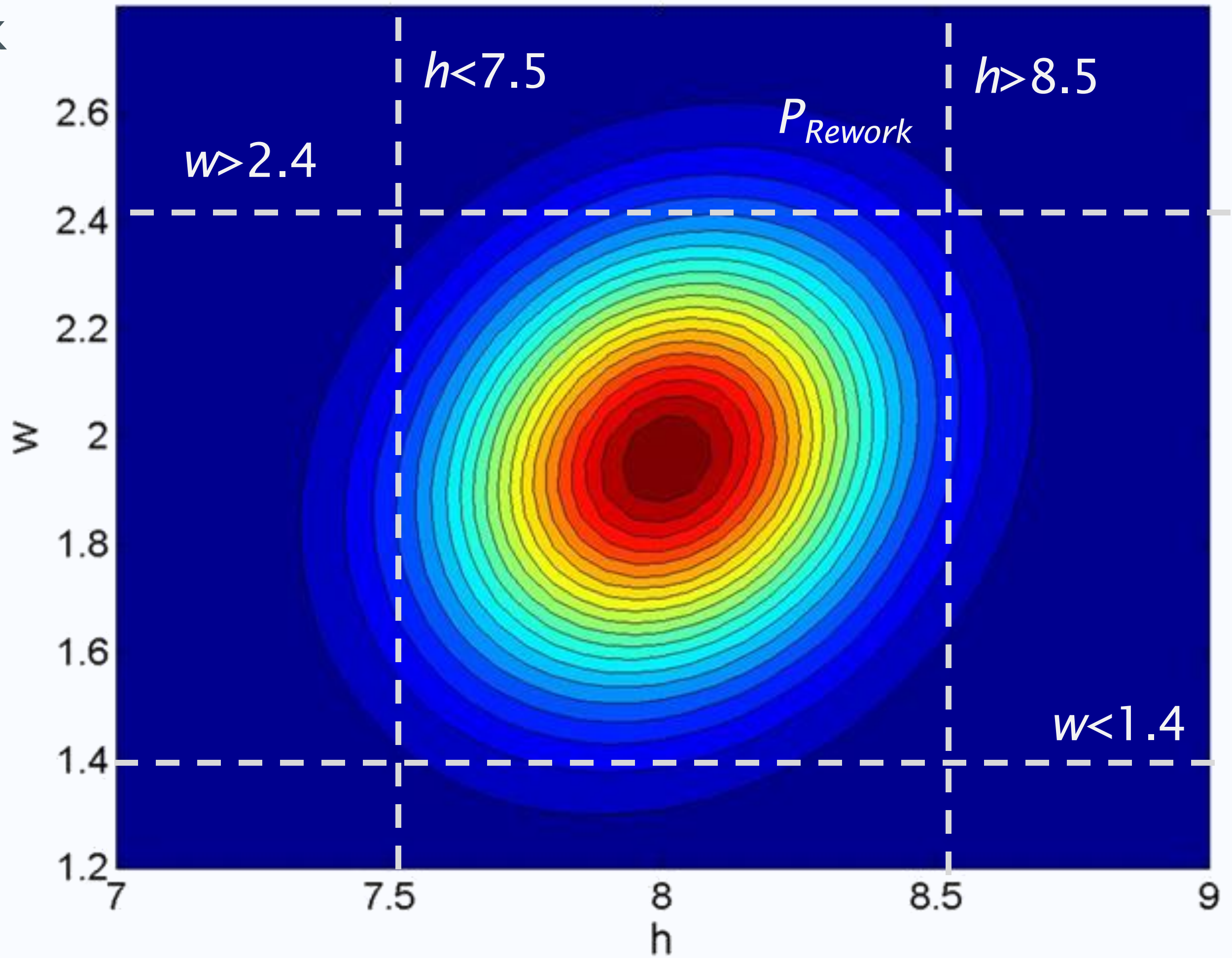
- The probability of scrappage is

$$P_{scrap} = 0.0474$$

‡Recall from lecture 4 that in Python `scipy.stats.multivariate_normal` could be used

A4.2

- We can use a similar technique to calculate the probability of rework



A4.2

- The probability of rework is therefore:

$$P_{Rework} = 1 - F(8.5,2.4) - \big(F(7.5,\infty) - F(7.5,2.4)\big) - \big(F(\infty,1.4) - F(8.5,1.4)\big)$$

- Using Matlab's `mvncdf()` function
- The probability of rework is

$$P_{Rework} = 0.0801$$

University of Southampton