Estimation of components failure rates from statistical data: Bayesian approach
Frequentist approach: a limitation

- Rare events (Frequentist school)
  - a new type of pump in an energy production plant, constant failure rate, $\lambda$, 1000 hours of operation, 0 failures

\[
\hat{\lambda}_{MLE} = \frac{k}{T} = 0 \, h^{-1}
\]

It is difficult to associate failure occurrence probabilities to this type of pump
This lecture

Life test → \((t_1, t_2, ..., t_n)\) → Estimate \(\vartheta\) of \(f_T(t|\vartheta)\)

For example: \(\lambda\) of \(f_T(t) = \lambda e^{-\lambda t}\)

- Frequentist approach (frequentist probability definition)
  - \(\vartheta\) is a fixed unknown parameter
  - From \((t_1, t_2, ..., t_n)\) find an estimator \(\hat{\vartheta}\) of \(\vartheta\)

- Bayesian Approach (subjective probability definition)

From Lecture 2: Definition of subjective probability:
\(P(A|K)\) is the degree of belief of the assigner with regard to the occurrence of \(A\) (numerical encoding of the state of knowledge – \(K\) - of the assessor)
The Bayesian approach to parameter estimation

Life test → \( (t_1, t_2, \ldots, t_n) \) → Estimate \( \vartheta \) of \( f_T(t|\vartheta) \)

For example: \( \lambda \) of \( f_T(t) = \lambda e^{-\lambda t} \)

- Frequentist approach (frequentist probability definition)
  - \( \vartheta \) is a fixed unknown parameter
  - From \( (t_1, t_2, \ldots, t_n) \) find an estimator \( \hat{\vartheta} \) of \( \vartheta \)

- Bayesian Approach (subjective probability definition)
  - \( \vartheta \) is a random quantity (epistemic uncertainty)
The Bayesian approach to parameter estimation

- \( \theta = \) parameter of the failure time distribution, \( f_T(t; \theta) \)
- \( \theta \) is a random quantity (epistemic uncertainty)
- Assessor provides a probability distribution of \( \theta \) based on its knowledge, experience,…:

\[
P(\theta) = \text{Prior distribution} \quad (\text{subjective probability})
\]

- When a sample of failure times \( E = \{t_1, t_2, \ldots, t_n\} \) becomes available, the estimate of \( \theta \) is updated by using the Bayesian theorem:

\[
P(\theta|E)
\]

Posterior
distribution
The Bayesian approach to parameter estimation

• \( \vartheta \) = parameter of the failure time distribution, \( f_T(t; \vartheta) \)
• \( \vartheta \) is a random quantity (epistemic uncertainty)
• Assessor provides a probability distribution of \( \vartheta \) based on its knowledge, experience, …:

\[
P(\vartheta) = \text{Prior distribution (subjective probability)}
\]

• When a sample of failure times \( E = \{t_1, t_2, \ldots, t_n\} \) becomes available, the estimate of \( \vartheta \) is updated by using the Bayes' theorem:

\[
P(\vartheta|E) = P(\vartheta) \frac{P(E|\vartheta)}{P(E)}
\]

Posterior distribution
Bayes Formula

\[ P(\vartheta|E) = P(\vartheta) \frac{P(E|\vartheta)}{P(E)} \]

- \( P(E|\vartheta) = L(\vartheta) \) ← likelihood of the evidence \( E \)
- \( P(E) = \sum_i P(E|\vartheta_i)P(\vartheta_i) \)
- \( P(E) = \int_\vartheta P(E|\vartheta)P(\vartheta)d\vartheta \)  

Theorem of Total Probability

\[ P(\vartheta|E) = k \cdot P(\vartheta) \cdot L(\vartheta) \]
• **Aleatory uncertainty** on the failure time: \( t \rightarrow P(t|\vartheta) \)

   Example: the failure time distribution is an exponential distribution
   \[
   t \sim f_T(t|\lambda) = \lambda e^{-\lambda t}
   \]

• **Epistemic uncertainty** on the parameter value, conditional on the background knowledge \( K \) (expert judgment, experimental data, …):
   \( \vartheta \rightarrow P(\vartheta|K) \)
   - the epistemic uncertainty can be updated through Bayes theorem
   - as the evidence increases, the background knowledge \( K \) improves and the epistemic uncertainty reduces
Comparing Bayesian and frequentist approaches (parameter estimation)

<table>
<thead>
<tr>
<th></th>
<th>Frequentist</th>
<th>Bayesian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter, $\theta$</td>
<td>fixed, unknown number, $\hat{\theta}$</td>
<td>random variable $\Theta$, $P(\theta</td>
</tr>
<tr>
<td>inference</td>
<td>ad hoc estimation methods (e.g. MLE)</td>
<td>Bayesian updating, logical extension of the theory of probability</td>
</tr>
<tr>
<td>Source of Information</td>
<td>Experimental data</td>
<td>Expert judgment + Experimental data</td>
</tr>
</tbody>
</table>
Frequentist vs. Bayesian statistics: Confidence intervals

- **Classical statistics:**
a 90% confidence interval means that there is a 0.9 probability that the interval contains the parameter which is a fixed value, although unknown.

- **Bayesian statistics:**
the parameter is a random variable with a given distribution and the 90% confidence interval tells me that right now, with my current knowledge, I am 90% confident that the true value (which I will discover when I gain perfect knowledge) will fall within these bounds.
Exercise 1: Bayes Theorem

• You feel that the frequency of heads, $\theta$, on tossing a particular coin is either 0.4, 0.5 or 0.6. Your prior probabilities are:
  
  \[
  P(\theta_1 = 0.4) = 0.1 \\
  P(\theta_2 = 0.5) = 0.7 \\
  P(\theta_3 = 0.6) = 0.2 \\
  \]

• You toss the coin just once and the toss results is tail: $E = \{ \text{tail} \}$

• Questions:
  
  1. Update the probability of $\theta$
  2. Consider the denominator of Bayes’ theorem and interpret it.
Exercise 1: Solution (I)

- We apply the Bayes’ theorem to revise the prior probabilities after the evidence $E$ that the toss results in tail.

\[
P(\vartheta_i|E) = P(\vartheta_i) \frac{P(E|\vartheta_i)}{P(E)}
\]

- Likelihood: $P(E|\vartheta_i) = P('tail'|\vartheta_i) = 1 - \vartheta_i$:

\[
\begin{align*}
P(E|\vartheta_1 &= 0.4) = 1 - 0.4 = 0.6 \\
P(E|\vartheta_2 &= 0.5) = 1 - 0.5 = 0.5 \\
P(E|\vartheta_3 &= 0.6) = 1 - 0.6 = 0.4 
\end{align*}
\]

- $P(E) = \sum_{i=1}^{3} P(\vartheta_i) \cdot P(E|\vartheta_i) = 0.1 \cdot 0.6 + 0.7 \cdot 0.55 + 0.2 \cdot 0.4 = 0.49 = 0.49$

- Posterior:

\[
\begin{align*}
P(\vartheta_1 &= 0.4) = P(\vartheta_1) \cdot \frac{P(E|\vartheta_1)}{P(E)} = 0.1 \cdot \frac{0.6}{0.49} = 0.1224 \\
P(\vartheta_2 &= 0.5) = P(\vartheta_2) \cdot \frac{P(E|\vartheta_2)}{P(E)} = 0.7 \cdot \frac{0.5}{0.49} = 0.7143 \\
P(\vartheta_3 &= 0.6) = P(\vartheta_3) \cdot \frac{P(E|\vartheta_3)}{P(E)} = 0.2 \cdot \frac{0.4}{0.49} = 0.1633 
\end{align*}
\]
Exercise 1: Solution (II)

The denominator: \( P(E) = \sum_{i=1}^{3} P(\theta_i) \cdot P(E|\theta_i) \) can be interpreted as the probability of a result in tails in a single trial based on our prior knowledge.
Suppose that a production manager is concerned about the items produced by a certain manufacturing process. More specifically, he is concerned about the proportion of these items that are defective. From past experience with the process, he feels that $\vartheta$, the proportion of defectives, can take only four possible values: 0.01, 0.05, 0.10 and 0.25. Moreover, he has observed the process and he has some information concerning $\vartheta$. This information can be summarized in terms of the following probabilities that constitute the production manager’s prior distribution of $\vartheta$:

$$P(\vartheta = 0.01) = 0.60$$
$$P(\vartheta = 0.05) = 0.30$$
$$P(\vartheta = 0.10) = 0.08$$
$$P(\vartheta = 0.25) = 0.02$$

The production manager assumes that the process can be thought of as a Bernoulli process, with the assumption of stationarity and independence appearing reasonable. That is, the probability that only one item is defective remains constant for all items produced and is independent of the past history of defectives from the process.

A sample of $n = 5$ items is taken from the production process, and $k = 1$ of the 5 is found to be defective. How can this information be combined with the prior information?
Solution

- The sampling distribution of the number of defectives in 5 trials, given any particular value of $\vartheta$, is a binomial distribution. The likelihoods are thus:

\[
P(k = 1|n = 5, \vartheta = 0.01) = \binom{5}{1} (0.01) \cdot (0.99)^4 = 0.0480
\]

\[
P(k = 1|n = 5, \vartheta = 0.05) = \binom{5}{1} (0.05) \cdot (0.95)^4 = 0.2036
\]

\[
P(k = 1|n = 5, \vartheta = 0.10) = \binom{5}{1} (0.10) \cdot (0.90)^4 = 0.3280
\]

\[
P(k = 1|n = 5, \vartheta = 0.25) = \binom{5}{1} (0.25) \cdot (0.75)^4 = 0.3955
\]

- Bayes’ theorem can be written in the form:

\[
P(\vartheta_i|E) = P(\vartheta_i) \frac{P(E|\vartheta_i)}{P(E)}
\]

where $t_1$ represents the sample result, 1 defective in 5 trials.
The posterior probabilities are summarized in the following Table:

<table>
<thead>
<tr>
<th>(1) Prior probability</th>
<th>(2) Likelihood</th>
<th>(3) (Prior probability)</th>
<th>(4) Likelihood</th>
<th>(5) Posterior probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>.60</td>
<td>.0480</td>
<td>.02880</td>
<td>.02880/.12403 = .232</td>
</tr>
<tr>
<td>.05</td>
<td>.30</td>
<td>.2036</td>
<td>.06108</td>
<td>.06108/.12403 = .492</td>
</tr>
<tr>
<td>.10</td>
<td>.08</td>
<td>.3280</td>
<td>.02624</td>
<td>.02624/.12403 = .212</td>
</tr>
<tr>
<td>.25</td>
<td>.02</td>
<td>.3955</td>
<td>.00791</td>
<td>.00791/.12403 = .064</td>
</tr>
<tr>
<td>1.00</td>
<td></td>
<td>.12403</td>
<td></td>
<td>1.000</td>
</tr>
</tbody>
</table>
Exercise 2: (continue)

- The production manager decides to take another sample of size $n = 5$, in order to obtain more information about the production process, and he found $k = 2$ defectives. Update the probability of $\theta$ using the new experimental evidence.
Exercise 2: (Solution)

- The production manager decides to take another sample of size $n = 5$, in order to obtain more information about the production process, and he found $k = 2$ defectives. Update the probability of $\theta$ using the new experimental evidence.

<table>
<thead>
<tr>
<th>$p$</th>
<th>Prior probability</th>
<th>Likelihood</th>
<th>$(\text{Prior probability}) \times (\text{likelihood})$</th>
<th>Posterior probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>.232</td>
<td>.0010</td>
<td>.00023</td>
<td>.005</td>
</tr>
<tr>
<td>.05</td>
<td>.492</td>
<td>.0214</td>
<td>.01053</td>
<td>.244</td>
</tr>
<tr>
<td>.10</td>
<td>.212</td>
<td>.0729</td>
<td>.01545</td>
<td>.359</td>
</tr>
<tr>
<td>.25</td>
<td>.064</td>
<td>.2637</td>
<td>.01688</td>
<td>.392</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td></td>
<td>.04309</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Bayesian approach: continuous updating of the parameter distribution

- $E_n = \{t_1, t_2, ..., t_n\} \rightarrow P(\vartheta|E_n)$ already updated (it will be our prior for next updating)
- $t_{n+1} =$ new evidence $\rightarrow E_{n+1} = \{t_1, t_2, ..., t_n, t_{n+1}\}$

\[ P(\vartheta|E_n, t_{n+1}) = P(\vartheta|E_{n+1}) = P(\vartheta|E_n) \cdot \frac{P(t_{n+1}|\vartheta)}{P(t_{n+1}|E_n)} \]

\[ P(t_{n+1}|E_n) = \int P(\vartheta|E_n) \cdot P(t_{n+1}|\vartheta)d\vartheta \]
Exercise 3: (continue)

What would you obtain in Exercise 1 if you start from the prior knowledge:

\[
\begin{align*}
P(\vartheta = 0.01) &= 0.60 \\
P(\vartheta = 0.05) &= 0.30 \\
P(\vartheta = 0.10) &= 0.08 \\
P(\vartheta = 0.25) &= 0.02
\end{align*}
\]

and you perform a single updating with \( n=10 \) and \( k=3 \)?
Bayesian approach

Multiple stage updating

\[ P(\theta) \rightarrow P(\theta | t_1) \rightarrow P(\theta | t_1, t_2) \rightarrow \ldots \rightarrow P(\theta | t_1, \ldots, t_{n-1}) \rightarrow P(\theta | t_1, \ldots, t_n) \]

Single stage updating

\[ \{t_1, t_2, \ldots, t_n\} \]

\[ P(\theta) \rightarrow P(\theta | t_1, \ldots, t_n) \]
Bayesian approach: coherence

Multiple stage updating

\[ P(\theta) \xrightarrow{t_1} P(\theta|t_1) \xrightarrow{t_2} P(\theta|t_1, t_2) \xrightarrow{\ldots} P(\theta|t_1, \ldots, t_{n-1}) \xrightarrow{t_n} P(\theta|t_1, \ldots, t_n) \]

Single cumulative updating

\{t_1, t_2, \ldots, t_n\}

\[ P(\theta) \xrightarrow{t_1, \ldots, t_n} P(\theta|t_1, \ldots, t_n) \]

Same evidence

Same ‘a posteriori’
Updating on $t_{n+2}$ after $t_{n+1}$ (multiple stage updating)

$$P(\vartheta|E_{n+1}, t_{n+2}) = P(\vartheta|E_{n+2}) = \underbrace{P(\vartheta|E_{n+1}) \cdot \frac{P(t_{n+2}|\vartheta)}{P(t_{n+2}|E_{n+1})}}_{\text{Conditional probability}}$$

with:

$$P(t_{n+2}|E_{n+1}) = P(t_{n+2}|E_n, t_{n+1}) = \frac{P(t_{n+2}, t_{n+1}|E_n)}{P(t_{n+1}|E_n)}$$

$$P(\vartheta|E_{n+1}) = P(\vartheta|E_n) \cdot \frac{P(t_{n+1}|\vartheta)}{P(t_{n+1}|E_n)}$$

First updating
Updating on $t_{n+2}$ after $t_{n+1}$ (multiple stage updating)

$$P(\theta|E_{n+2}) = P(\theta|E_{n+1}) \cdot \frac{P(t_{n+2}|\theta)}{P(t_{n+2}|E_{n+1})}$$

with:

$$P(t_{n+2}|E_{n+1}) = P(t_{n+2}|E_n, t_{n+1}) = \frac{P(t_{n+2}, t_{n+1}|E_n)}{P(t_{n+1}|E_n)}$$

First updating

$$P(\theta|E_{n+1}) = P(\theta|E_n) \cdot \frac{P(t_{n+1}|\theta)}{P(t_{n+1}|E_n)}$$

Conditional probability

$$P(\theta|E_{n+2}) = P(\theta|E_{n+1}) \cdot \frac{P(t_{n+2}|\theta)}{P(t_{n+2}|E_{n+1})} = P(\theta|E_n) \cdot \frac{P(t_{n+1}|\theta)}{P(t_{n+1}|E_n)} \cdot \frac{P(t_{n+2}|\theta)}{P(t_{n+2}, t_{n+1}|E_n)}$$

same as a single cumulative updating!
Bayesian approach: some observations on the updating process

\[ P(\theta|E) \propto P(\theta) \cdot P(E|\theta) \]

Posterior \( \propto \) Prior \( \cdot \) Likelihood

- In correspondence of values of \( \theta \) for which both prior and likelihood are small \( \rightarrow \) the posterior will be small
- bulk of the posterior where both prior and likelihood are not negligible
Bayesian approach: some observations on the updating process

\[ P(\theta|E) \propto P(\theta) \cdot P(E|\theta) \]

Posterior \( \propto \) Prior \( \cdot \) Likelihood

- In correspondence of values of \( \theta \) for which both prior and likelihood are small → the posterior will be small
- bulk of the posterior where both prior and likelihood are not negligible
- If the prior is very sharp (strong prior evidence), it will not change much unless the evidence is very strong

Which of the two prior will be more influenced by the evidence \( t \)?

Posterior depends on the relative strength of prior and likelihood
Bayesian approach to parameter estimation: Large evidence - Example

- Parameter $\vartheta = P\{\text{success}'\} = P(\text{success}')$
- Evidence: $E = \{k \text{ successes on } n \text{ trials}\}$

$P(\vartheta|E) \propto P(\vartheta) \cdot P(E|\vartheta)$ with:
- Prior: $P(\vartheta)$
- Likelihood: $L(\vartheta) = P(E|\vartheta) = b(k; n, \vartheta) = \binom{n}{k} \vartheta^k (1 - \vartheta)^{n-k}$

It is possible to show that:
- $\lim_{n \to \infty} b(k; n, \vartheta) = N(n\vartheta, n\vartheta(1 - \vartheta))$
Normal distribution as a limit of the binomial distribution

\[
b(k; n = 5, \varphi = 0.1) \quad \quad \quad \quad \quad \quad b(k; n = 10, \varphi = 0.1) \quad \quad \quad \quad \quad \quad b(k; n = 50, \varphi = 0.1)
\]

\[
b(k; n = 100, \varphi = 0.1) \quad \quad \quad \quad \quad \quad b(k; n = 500, \varphi = 0.1) \quad \quad \quad \quad \quad \quad b(k; n = 1000, \varphi = 0.1)
\]
Bayesian approach to parameter estimation: Large evidence - Example

- Parameter $\vartheta = P\{\text{success}'\}$
- Evidence: $E = \{k$ successes on $n$ trials$\}$

$$P(\vartheta|E) \propto P(\vartheta) \cdot P(E|\vartheta)$$ with:
  - Prior: $P(\vartheta)$
  - Likelihood: $L(\vartheta) = P(E|\vartheta) = b(k; n, \vartheta) = \binom{n}{k} \vartheta^k (1 - \vartheta)^{n-k}$

It is possible to show that:

$$\lim_{n \to \infty} b(k; n, \vartheta) = N(n\vartheta, n\vartheta(1 - \vartheta)) = \frac{1}{(\sqrt{2\pi}) (n\vartheta(1-\vartheta))} e^{-\frac{(k-n\vartheta)^2}{2n\vartheta(1-\vartheta)}}$$

$$L(\vartheta) = \frac{1}{(\sqrt{2\pi}) (n\vartheta(1-\vartheta))} e^{-\frac{(k-n\vartheta)^2}{2n\vartheta(1-\vartheta)}}$$

Notice that:
- $k$ and $n$ are known
- $\vartheta$ is unknown

- $L(\vartheta)$ is a continuous function of $\vartheta$
- Maximum of the likelihood for:
  $$\frac{\partial L(\vartheta)}{\partial \vartheta} = 0 \Rightarrow \vartheta_{max} = \frac{k}{n}$$
Behaviour of the likelihood for $n \to \infty$, assuming $k/n=0.1$.

- $k=1$, $n=10$
- $k=10$, $n=100$
- $k=100$, $n=1000$
- $k=1000$, $n=10000$
- $k=10000$, $n=100000$
Bayesian approach to parameter estimation: Large evidence - Example

\[ n \to \infty \]

Likelihood: \[ \lim_{n \to \infty} L(\theta) = \delta(\theta - \theta_{max}) = \delta \left( \theta - \frac{k}{n} \right) \]

Posterior: \[ P(\theta | E) = \text{const} \cdot \delta(\theta - \theta_{max})P(\theta) = \text{const} \cdot \delta(\theta - \theta_{max}) \cdot P(\theta_{max}) = \delta(\theta - \theta_{max}) = \delta \left( \theta - \frac{k}{n} \right) \]

Bayesian statistics \equiv\ frequentist statistics \left( \hat{\theta}_{MLE} = \frac{k}{n} \right) 

(the prior has no effects on the posterior)

(Bayesian \neq\ frequentist only for scarce evidence when prior beliefs count)
Bayesian approach to parameter estimation: Large evidence

Evidence becomes stronger and stronger

The likelihood tends to a delta function

The posterior tends to a delta as well, centered around the only value which is now the true value (perfect knowledge)

The classical and bayesian statistics become identical in the results (not conceptually)
Conjugate distributions

- The likelihood $L(\vartheta)$ and the prior $f(\vartheta)$ are called conjugate distributions if the posterior $f(\vartheta | E)$ is in the same family of the prior distribution.

- Example:
  - Likelihood: binomial distribution
  - Prior = Beta Distribution($q, r$)

$$f_{\vartheta}(\vartheta) = \frac{\Gamma(q + r)}{\Gamma(q)\Gamma(r)} \vartheta^{q-1}(1 - \vartheta)^{r-1}$$

Posterior = Beta distribution (different parameters)

![Graph showing Beta distributions](image-url)
Conjugate distributions

- The likelihood $L(\theta)$ and the prior $f(\theta)$ are called conjugate distributions if the posterior $f(\theta|E)$ is in the same family of the prior distribution.

- Example:
  - Likelihood: binomial distribution
  - Prior = Beta Distribution($q,r$)
  - Posterior = Beta distribution (different parameters)

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Prior and Posterior distributions</th>
<th>Mean and Variance of the parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L(\theta) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}$</td>
<td>$f_\theta(\theta) = \frac{\Gamma(q+r)}{\Gamma(q)\Gamma(r)} \theta^{q-1} (1 - \theta)^{r-1}$</td>
<td>$E[\theta] = \frac{q}{q + r}$</td>
</tr>
<tr>
<td>$\text{Var}[\theta] = \frac{qr}{(q + r)^2(q + r + 1)}$</td>
<td>$q'' = q' + k$</td>
<td></td>
</tr>
<tr>
<td>$r'' = r' + n - k$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameters of the prior:
- $q'$
- $r'$

Parameters of the posterior:
- $q''$
- $r''$
Bayesian approach to parameter estimation: Families of conjugate distributions

- Conjugate distributions characteristics:
  - posterior $\equiv$ prior with updated parameters
  - estimates $\equiv$ simple analytical (mean and variance)

<table>
<thead>
<tr>
<th>Basic random variable</th>
<th>Parameter</th>
<th>Prior and posterior distributions of parameter</th>
<th>Mean and Variance of Parameter</th>
<th>Posterior Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binomial</td>
<td>Beta</td>
<td>$p(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$</td>
<td>$E(\theta) = \frac{q}{q + r}$</td>
<td>$q^* = q' + x$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\text{Var}(\theta) = \frac{qr}{(q + r)^2(q + r + 1)}$</td>
<td>$r^* = r' + n - x$</td>
</tr>
<tr>
<td>Exponential</td>
<td>Gamma</td>
<td>$f_X(x) = \lambda e^{-\lambda x}$</td>
<td>$E(\lambda) = \frac{k}{\nu}$</td>
<td>$\nu^* = \nu' + \sum x_i$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\text{Var}(\lambda) = \frac{k}{\nu^2}$</td>
<td>$k^* = k' + n$</td>
</tr>
<tr>
<td>Normal</td>
<td>Normal</td>
<td>$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right]$</td>
<td>$E(\mu) = \mu$</td>
<td>$\mu^* = \frac{\mu' (\sigma^2/n) + 2\sigma^* \sigma^3}{\sigma^2 n + (\sigma^3)^2}$</td>
</tr>
<tr>
<td>(with known $\sigma$)</td>
<td></td>
<td></td>
<td>$\text{Var}(\mu) = \sigma^3$</td>
<td>$\sigma^* = \sqrt{(\sigma^3)^2 + \sigma^2/n}$</td>
</tr>
<tr>
<td>Normal</td>
<td>Gamma-Normal</td>
<td>$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right]$</td>
<td>$E(\mu) = \bar{x}$</td>
<td>$n^* = n' + n$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$E(\sigma) = \sqrt{\frac{n - 1 - \Gamma[n / 2, 1/2]}{2 \Gamma[n / 2]}} \left[\frac{(n - 1) \Gamma[n/2]}{\Gamma[n/2]}</td>
<td></td>
<td></td>
</tr>
</tbody>
</table><p>ight]^{\sigma^2/2 - \frac{n}{2}} \exp\left[-\frac{1}{2}\left(\frac{\sigma^2}{\sigma^2}\right)^2\right]$ | $\text{Var}(\sigma) = \frac{n - 1}{n - 3} - \bar{x}^2$ | $(n^* - 1) \sigma^2 + n \sigma^2$ |
| Poisson               | Gamma     | $p(x) = \frac{\mu^x e^{-\mu}}{x!}$             | $E(\mu) = \frac{k}{\nu}$     | $\nu^* = \nu' + t$ |
|                       |           |                                               | $\text{Var}(\mu) = \frac{k}{\nu^2}$ | $k^* = k' + \pi$ |
| Lognormal             | Normal    | $f_X(x) = \frac{1}{2\sigma \nu} \exp\left[-\frac{1}{2}\left(\ln x - \frac{\lambda}{\sigma}\right)^2\right]$ | $E(\lambda) = \frac{\mu}{\nu}$ | $\mu^* = \mu' (\sigma^2/n) + \frac{\sigma \ln x}{\sigma^2/n + \sigma^2}$ |
| (with known $\nu$)    |           |                                               | $\text{Var}(\lambda) = \sigma^3$ | $\sigma^* = \sqrt{\sigma^3 + \frac{\sigma^2}{\nu^2}}$ |</p>
Exercise 3: failure rate of a motor-driven pump of a NPP

- The failure rate of the pump is constant, $\lambda$
- We have available the following three sources of information:
  - E1: engineering knowledge (description of the design and construction of the pump)
  - E2: past performance of similar pumps in similar plants

$$\lambda' = 3 \cdot 10^{-5} \text{ h}^{-1}$$

$$\sigma_{\lambda}' = 7.4 \cdot 10^{-5} \text{ h}^{-1}$$

- E3: performance of the specific machine = 0 failures in $t = 1000$ h

Questions:
1) Use E1 and E2 to build the prior of the failure rate distribution, $P(\lambda)$
2) Update the prior using the information in E3. Determine the point estimator of $\lambda$, its variance and its 95 percentile
Exercise 3: failure rate of a motor-driven pump of a NPP - Solution

- To choose the family of the prior \( P(\lambda) \), I observe that E3 can be seen as the result of a poisson experiment: 0 occurrence in 1000h:

\[
P[k \text{ failures in } [0,t]] = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad \text{Poisson process}
\]

- The conjugate distribution to the poisson is the Gamma with two parameters \( \alpha', \beta' \):

\[
\text{Prior}: P(\lambda|E_1) = \Gamma(\alpha',\beta') = \frac{\beta'^{\alpha'} \lambda^{\alpha'-1}}{\Gamma(\alpha')} e^{-\beta' \lambda'}
\]

\[
\bar{\lambda'} = \frac{\alpha'}{\beta'} \quad \sigma_{\lambda'} = \frac{\sqrt{\alpha'}}{\beta'}
\]

- I model the prior as a Gamma function with parameters \( \alpha', \beta' \) obtained from:

\[
\frac{\alpha'}{\beta'} = \bar{\lambda'} = 3 \cdot 10^{-5}
\]

\[
\frac{\sqrt{\alpha'}}{\beta'} = \sigma_{\lambda'} = 7.4 \cdot 10^{-5}
\]

\[
\alpha' = \frac{\bar{\lambda'}^2}{\sigma_{\lambda'}^2} = 0.1644
\]

\[
\beta' = \frac{\bar{\lambda'}}{\sigma_{\lambda'}^2} = 5478
\]
Exercise 3: failure rate of a motor-driven pump of a NPP - Solution

- Gamma and Poisson are conjugate

- Posterior is still a Gamma distribution with parameters

\[ \alpha'' = \alpha' + k = 0.1644 \]
\[ \beta'' = \beta' + t = 6478 \]

\[ \bar{\lambda}'' = \frac{\alpha}{\beta} = 2.5 \cdot 10^{-5} \]
\[ \sigma_{\lambda''} = \frac{\sqrt{\alpha}}{\beta} = 6.2 \cdot 10^{-5} \]
\[ \lambda''_{95} = 1.4 \cdot 10^{-4} \]
Exercise 3: failure rate of a motor-driven pump of a NPP - Solution

- Frequentist statistics:

\[
\hat{\lambda}_{\text{MLE}} = \frac{k}{t} = \frac{0}{1000} = 0 \text{h}^{-1} \quad \lambda_{95}^F = \frac{\chi_{95}^2 (2k + 2)}{2t} = \frac{\chi_{95}^2 (2)}{2000h} = \frac{5.99}{2000h} = 3 \cdot 10^{-3} \text{h}^{-1}
\]
As already pointed out, the results of the Bayesian and frequentist analyses converge with large amounts of data. The influence of the prior parameters $\alpha'$, $\beta'$ decreases.

\[
\bar{\lambda}'' = \frac{\alpha''}{\beta''} = \frac{\alpha' + k}{\beta' + t} \rightarrow \frac{k}{t} = \hat{\lambda}_{\text{MLE}} \quad \text{for } k, t \rightarrow \infty
\]

\[
\bar{\sigma}_{\hat{\lambda}}'' = \frac{\sqrt{\alpha''}}{\beta''} = \frac{\sqrt{\alpha' + k}}{\beta' + t} \rightarrow 0
\]

Large amounts of data $\rightarrow$ the posterior distribution is highly peaked around the MLE estimate

\[
\hat{\lambda}_{\text{MLE}} = \frac{k}{t}
\]

It can also be shown that the Bayesian and frequentist 95 percentiles will converge;

The failure process remains inherently aleatory
Reliability of a component in a Bayesian framework

• Posterior distribution of the failure rate: $p''(\lambda|E)$
• How to compute the component reliability? $R(t) = P(T > t) = ?$
Reliability of a component in a Bayesian framework

- Posterior distribution of the failure rate: $p''(\lambda|E)$
- How to compute the component reliability?

Total Probability Theorem

$$R(t) = P(T > t) = \int_0^{+\infty} P(T > t|\lambda)p''(\lambda|E)d\lambda = \int_0^{+\infty} R(t|\lambda)p''(\lambda|E)d\lambda$$