POLITECNICO DI MILANO

Estimation of reliability parameters from Experimental data

(Part 2)

This lecture

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

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

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

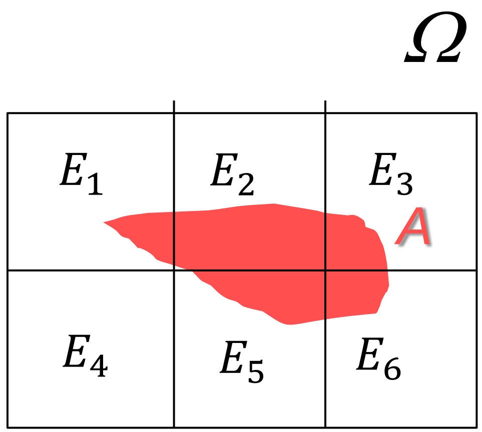
- **Bayesian Approach (subjective probability definition)**
  - ϑ is a random quantity (epistemic uncertainty)

Subjective Definition of Probability

P(A|K) is the **degree of belief** of the **assigner** about the occurrence of A (numerical encoding of the **state of knowledge** – K - of the assessor)

Theorem of Total Probability

- Let us consider a partition of the sample space Ω into n mutually exclusive and exhaustive events. In terms of Boolean events:



- Given any event A in Ω,

$$A = \bigcup_{j=1}^{n} A \cap E_j$$

$$P(A) = \sum_{j=1}^{n} P\left(A \cap E_j\right) = \sum_{j=1}^{n} P\left(A|E_j\right)P(E_j)$$

The Bayesian Subjective Probability Framework

P(E|K) is the **degree of belief** of the **assigner** with regard to the occurrence of E (numerical encoding of the **state of knowledge** – K - of the assessor)

When P(E) can be considered 'objective' from the scientific point of view?

- De Finetti: objective=coherent:
  - It uses total body of knowledge
  - It complies with theory of probability

Bayes Theorem to update the probability assignment in light of new data

Updated

$$P(E_i \mid A) = \frac{P(A \mid E_i)P(E_i)}{P(A)} = \frac{P(A \mid E_i)P(E_i)}{\sum_{j=1}^{n} P(A \mid E_j)P(E_j)}$$

Old

new data

The Bayesian Subjective Probability Framework

Rare events

- Frequentist school: we cannnot associate a probabilty to them
  - E.g. a pump of a NPP, constant failure rate, λ, 10000 hours of operation, 0 failures

$$\hat{\lambda}_{MLE}=\frac{k}{TTT}=0\ h^{-1}$$

- Bayesian School: we can associate a probability to them based on expert judgment, and then, as evidence is collected, we can update the probability using the Bayes Theorem

Repeatibility:

- Bayesian school: two assessors can be coherent and … still disagree

The Bayesian approach to parameter estimation

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

$P(\vartheta)$ = Prior distribution (subjective probability)

experimental evidence

- When a sample of failure times $\mathrm{E} = \{t_1, t_2, \ldots, t_n\}$ becomes available, the estimate of ϑ 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*
- $P(E) = \begin{cases} \sum_i P(E|\vartheta_i)P(\vartheta_i) \\ \int_\vartheta P(E|\vartheta)P(\vartheta)d\vartheta \end{cases}$ Theorem of Total Probability

➡ $P(\vartheta|E) = k \cdot P(\vartheta) \cdot L(\vartheta)$

Observations: The Two Types of Uncertainty

- 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)

	Frequentist	**Bayesian**
Parameter, θ	fixed, unknown number, $\hat{\theta}$	random variable $\Theta, P(\theta\|E)$
inference	ad hoc estimation methods (e.g., MLE)	Bayesian updating, logical extension of the theory of probability
Source of Information	experimental data	expert judgment + experimental data

Exercise 1: Bayes Theorem

- You feel that the frequency of heads, ϑ, on tossing a particular coin is either 0.4, 0.5 or 0.6. Your prior probabilities are:

$$P(\vartheta_1 = 0.4) = 0.1$$

$$P(\vartheta_2 = 0.5) = 0.7$$

$$P(\vartheta_3 = 0.6) = 0.2$$

- You toss the coin just once and the toss results is tail: $E = \{'tail'\}$
- Questions:
  1. Update the probability of ϑ
  2. Consider the denominator of Bayes' theorem and interpret it.

Exercise 2

- **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 ϑ , 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 ϑ. This information can be summarized in terms of the following probabilities that constitute the production manager's prior distribution of ϑ :**

$$P(\vartheta = 0.01) = 0.60$$

$$P(\vartheta = 0.05) = 0.30$$

$$P(\vartheta = 0.10) = 0.08$$

$$P(\vartheta = 0.25) = 0.02$$

- **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?**

Bayesian approach: continuous updating of the parameter distribution

We know:

$$P(R|S) = \frac{P(R,S)}{P(S)} = \frac{P(S|R)P(R)}{P(S)}$$

Therefore,

$$P(R|S,H) = \frac{P(R,S,H)}{P(S,H)}$$

By implementing conditional probability law $P(A,B) = P(A|B)\ P(B)$

$$= \frac{P(S,R|H)P(H)}{P(S|H)\ P(H)} = \frac{P(S|R,H)P(R|H)\ P(H)}{P(S|H)\ P(H)}$$

$$\rightarrow P(R|S,H) = \frac{P(S|R,H)P(R|H)}{P(S|H)}$$

Bayesian approach: continuous updating of the parameter distribution

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

$$P(\vartheta | E_{n-1}, t_n) = \frac{P(\vartheta, E_{n-1}, t_n)}{P(E_{n-1}, t_n)} = P(\vartheta, t_n | E_{n-1}) \frac{P(E_{n-1})}{P(E_{n-1}, t_n)} =$$

$$= P(t_n | \vartheta, E_{n-1}) P(\vartheta | E_{n-1}) \frac{P(E_{n-1})}{P(E_{n-1}, t_n)}$$

$$= P(t_n | \vartheta, \cancel{E_{n-1}}) P(\vartheta | E_{n-1}) \frac{P(E_{n-1})}{P(t_n | E_{n-1}) P(E_{n-1})} = P(\vartheta | E_{n-1}) \frac{P(t_n | \vartheta)}{P(t_n | E_{n-1})}$$

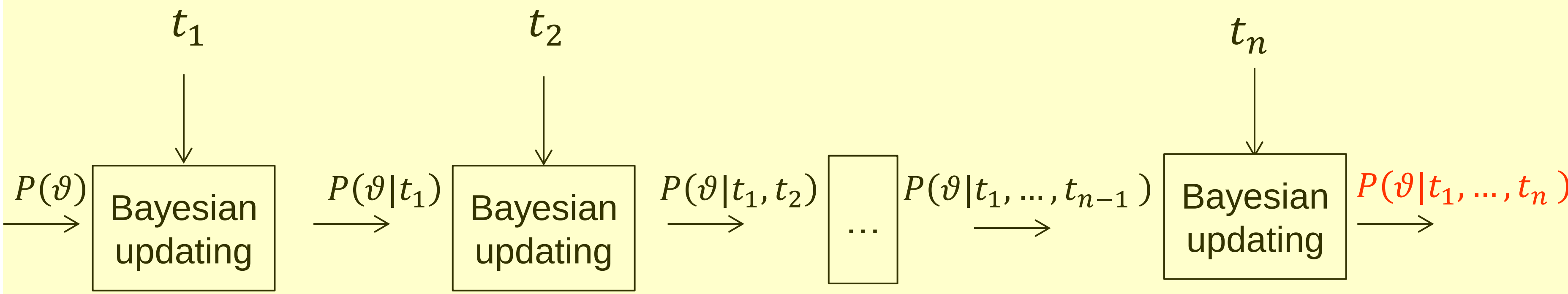
Note: when the parameter value is known, ϑ no longer depends on T, e.g. $T | \vartheta \sim N(\theta, \sigma^2)$

$$P(t_n | E_{n-1}) = \int P(\vartheta | E_{n-1}) \cdot \underbrace{P(t_n | \vartheta)}_{\text{Independent from the previous } E_{n-1}} d\vartheta$$

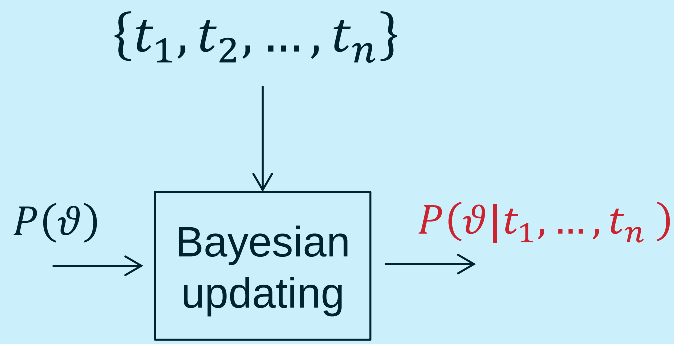
Independent from the previous E_{n-1}

Bayesian approach: coherence

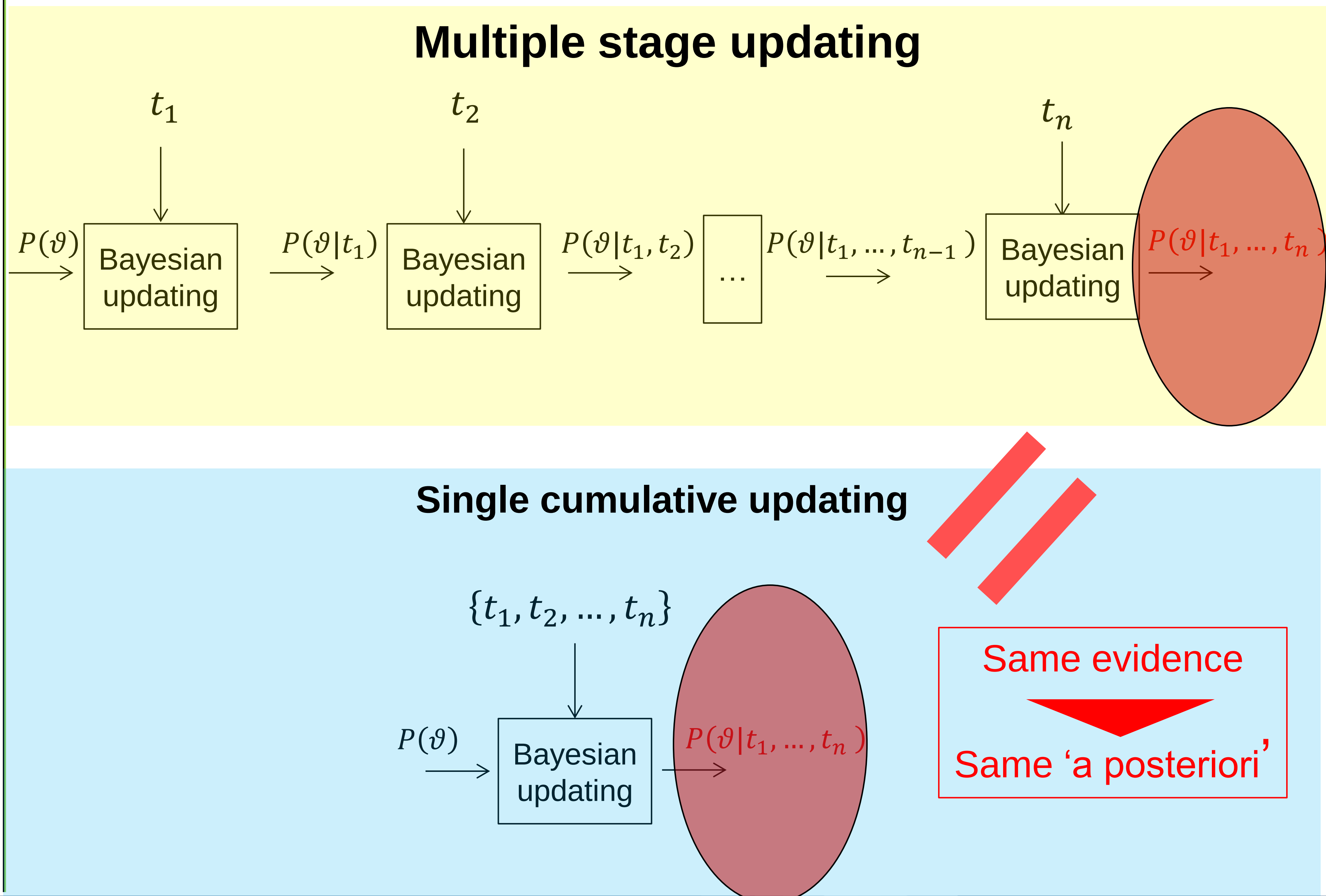
Multiple stage updating



Single stage updating



Bayesian approach: coherence



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

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

with:

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

Conditional probability

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

First updating

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

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

same as a single cumulative updating!

Bayesian approach: some observations on the updating process

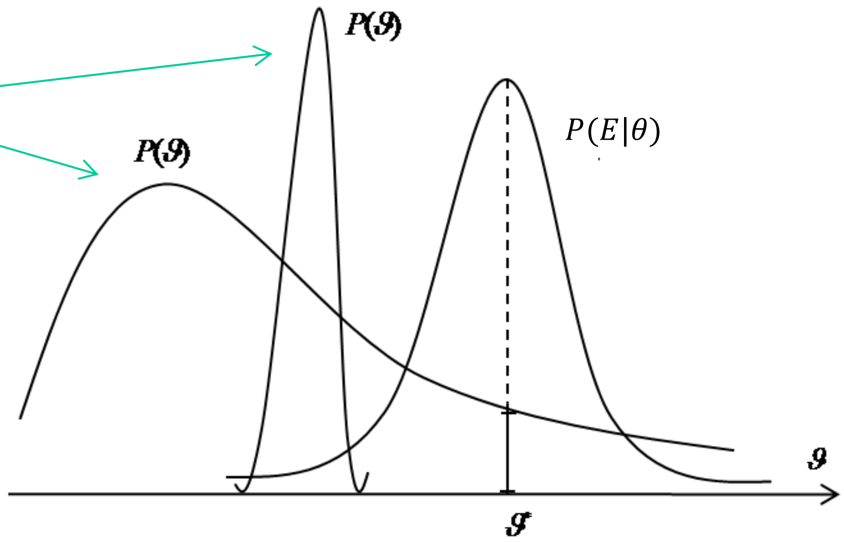
$$P(\vartheta|E) \propto P(\vartheta) \cdot P(E|\vartheta)$$

Posterior ∝ Prior ·Likelihood

- In correspondence of values of ϑ for which both prior and likelihood are small → the posterior will be small
- bulk of the posterior where 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 *E?*

Posterior depends on the relative strength of prior and likelihood



BAYESIAN APPROACH:

LARGE EVIDENCE

Bayesian approach to parameter estimation: Large evidence - Example

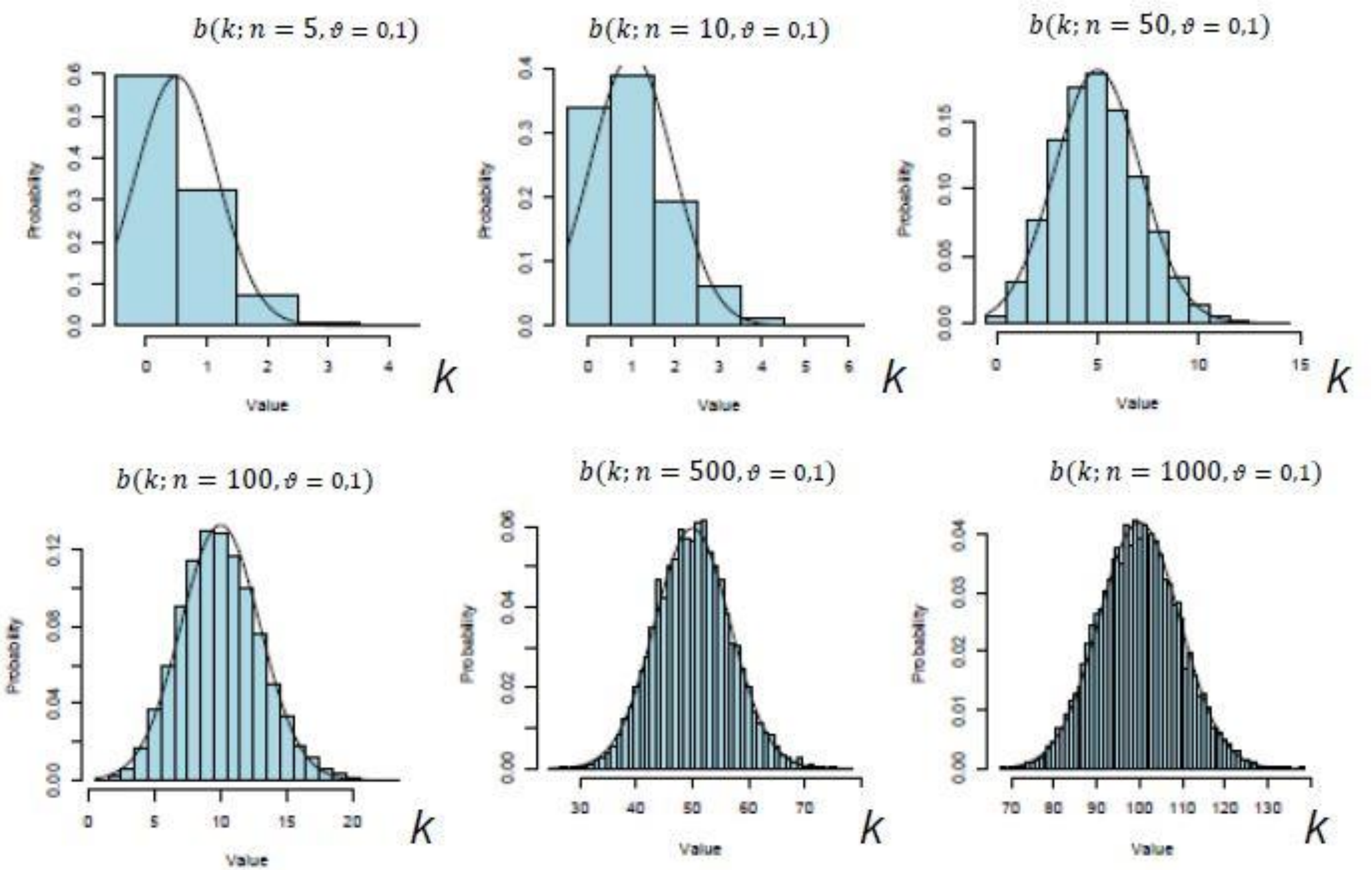
- Parameter $\vartheta = P\{'\text{success}'\}$ = P('success')
- Evidence: $E = \{k$ successes on n trials$\}$

- $P(\vartheta|E) \propto P(\vartheta) \cdot P(E|\vartheta)$ 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:

- $\lim_{n \to \infty} b(k; n, \vartheta) = N(n\vartheta, n\vartheta(1-\vartheta))$

Normal distribution as a limit of the binomial distribution ($\vartheta = 0.1$)



Bayesian approach to parameter estimation: Large evidence - Example

- Parameter $\vartheta = P\{'\text{success}'\}$ = P('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:

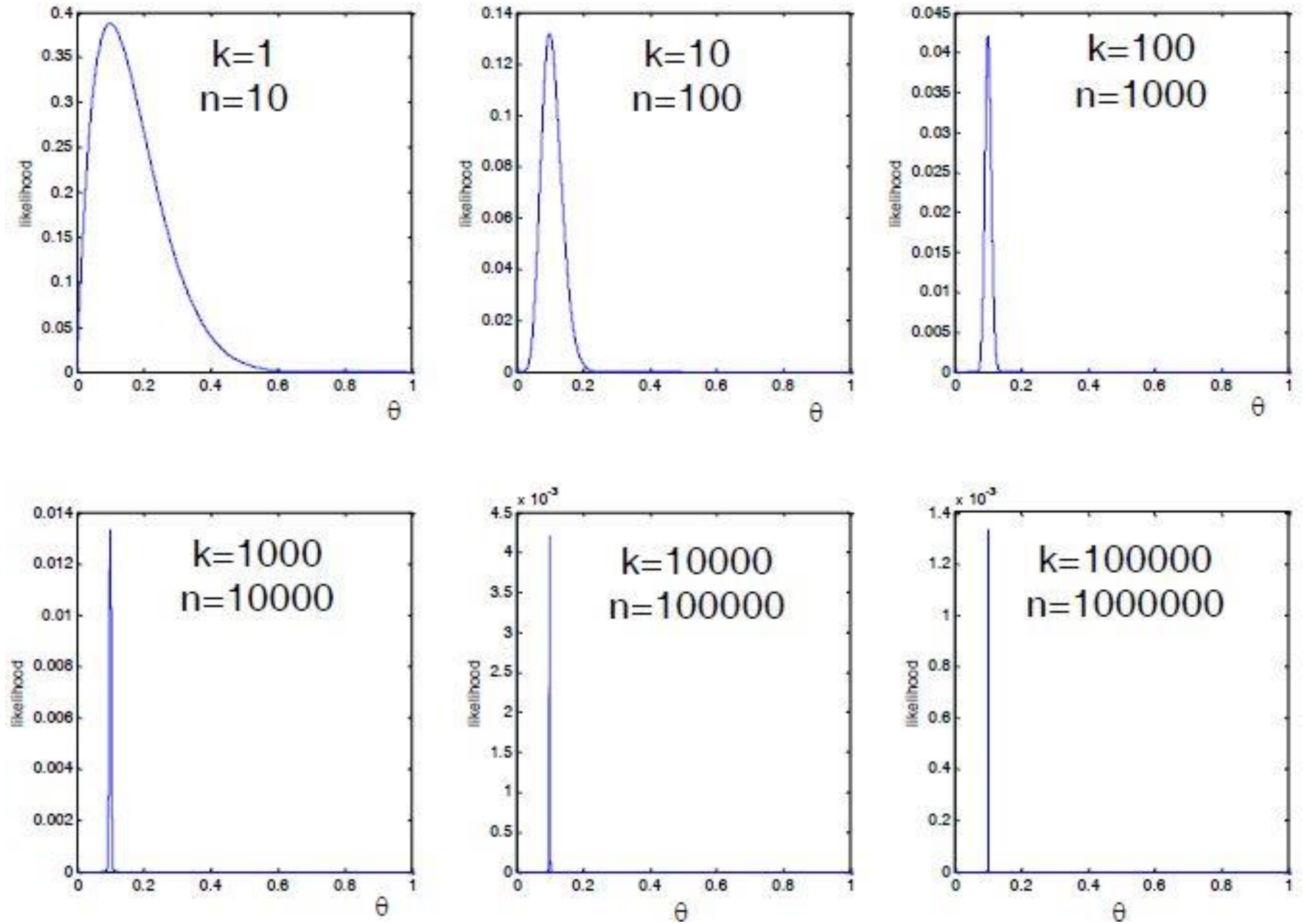
- $\lim_{n\to\infty} b(k; n, \vartheta) = N(n\vartheta, n\vartheta(1-\vartheta)) = \frac{1}{(\sqrt{2\pi})(n\vartheta(1-\vartheta)} e^{\frac{(k-n\vartheta)^2}{2n\vartheta(1-\vartheta)}}$

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

Notice that:
- k and n are known
- ϑ is unknown

- $L(\vartheta)$ is a continuous function of ϑ
- Maximum of the likelihood for:
  $\frac{\partial L(\vartheta)}{\partial \vartheta} = 0 \rightarrow \vartheta_{max} = \frac{k}{n}$

Likelihood behavior when $n \to \infty$, assuming $n/k = 0.1$



Bayesian approach to parameter estimation: Large evidence - Example

$$n \to \infty$$

$$\lim_{n\to\infty} L(\vartheta) = \delta(\vartheta - \vartheta_{max}) = \delta\left(\vartheta - \frac{k}{n}\right)$$ Likelihood

$$P(\vartheta|E) = const \cdot \delta(\vartheta - \vartheta_{max}) P(\vartheta) = \delta(\vartheta - \vartheta_{max}) = \delta\left(\vartheta - \frac{k}{n}\right)$$ Posterior

Bayesian statistics ≡ frequentist statistics ($\hat{\vartheta}^{MLE} = \frac{k}{n}$)

(the prior has no effects on the posterior)

(Bayesian ≠ 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

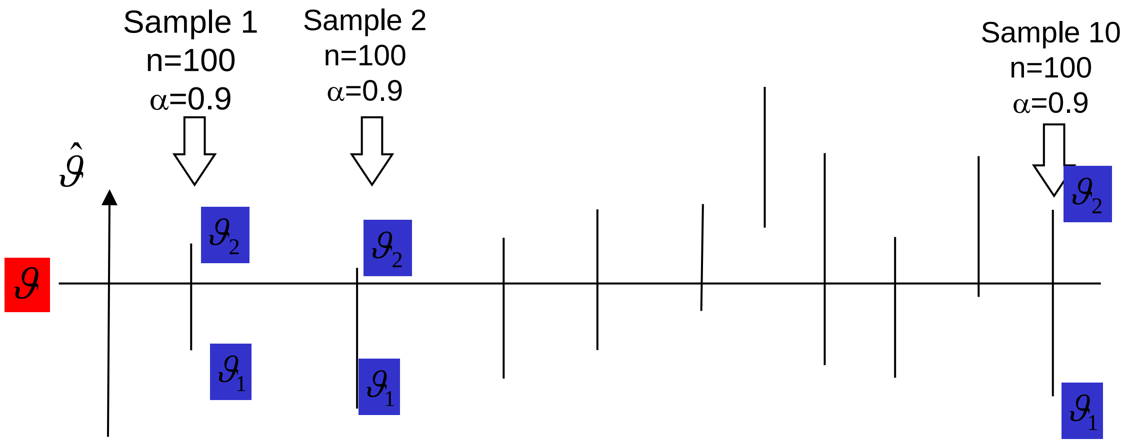
The posterior tends to a delta as well, centered around the only value which is now the true value (perfect knowledge)

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

Confidence intervals vs Credible intervals

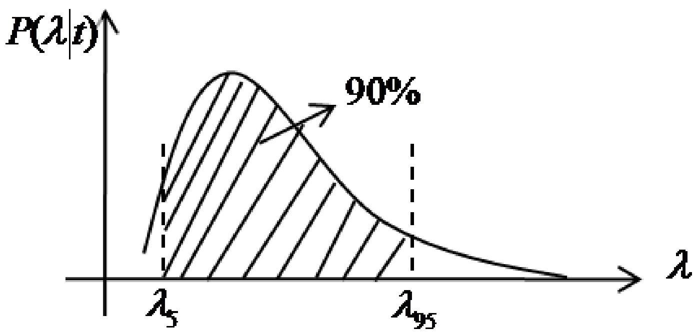
- **Classical statistics:**

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% credible 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.



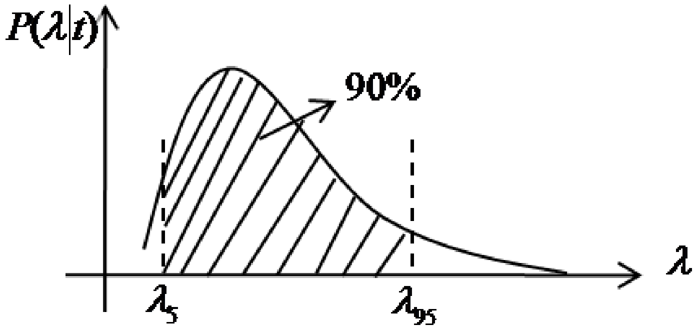
Confidence intervals vs Credible intervals

- **Classical statistics:**

Confidence intervals capture the uncertainty about the interval we have obtained (i.e., whether it contains the true value or not). Thus, they cannot be interpreted as a probabilistic statement about the true parameter values.

- **Bayesian statistics:**

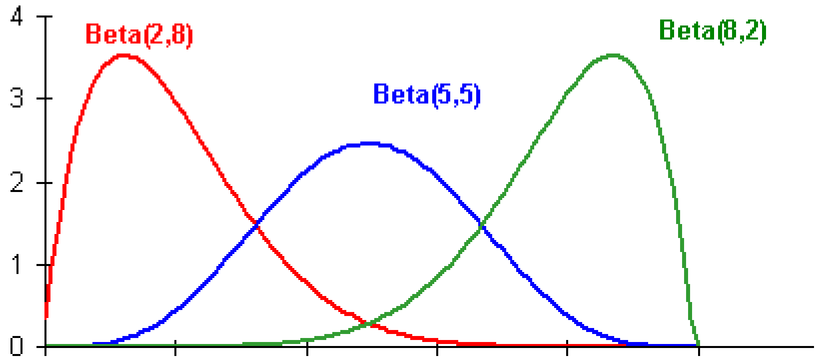
Credible intervals capture our current uncertainty in the location of the parameter values and thus can be interpreted as probabilistic statement about the parameter



Conjugate distributions

- The likelihood $L(\vartheta)$ and the prior $f(\vartheta)$ are called conjugate distributions if the posterior $\pi(\vartheta|E)$ is in the same family of the prior distribution
- Example:
  - Likelihood: binomial distribution
  - Prior = Beta Distribution(*q*,*r*)

Posterior = Beta distribution (different parameters)



Conjugate distributions

- The likelihood $L(\vartheta)$ and the prior $f(\vartheta)$ are called conjugate distributions is the posterior $\pi(\vartheta|E)$ is in the same family of the prior distribution
- Example:
  - Likelihood: binomial distribution
  - Prior = Beta Distribution

Posterior = Beta distribution (different parameters)

Basic random variable	Parameter	Prior and posterior distributions of parameter
Binomial		Beta
$p_X(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$	θ	$f_\Theta(\theta) = \frac{\Gamma(q+r)}{\Gamma(q)\Gamma(r)} \theta^{q-1}(1-\theta)^{r-1}$

Mean and Variance of Parameter	Posterior Statistics
$E(\Theta) = \frac{q}{q+r}$	$q'' = q' + x$
$\mathrm{Var}(\Theta) = \frac{qr}{(q+r)^2(q+r+1)}$	$r'' = r' + n - x$

Bayesian approach to parameter estimation: Families of conjugate distributions

- **Conjugate distributions characteristics:**
  - **posterior ≡ prior with updated parameters**
  - **estimates ≡ simple analytical (mean and variance)**

Basic random variable	Parameter	Prior and posterior distributions of parameter	Mean and Variance of Parameter	Posterior Statistics
Binomial $p_X(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$	θ	Beta $f_\Theta(\theta) = \frac{\Gamma(q+r)}{\Gamma(q)\Gamma(r)} \theta^{q-1}(1-\theta)^{r-1}$	$E(\Theta) = \frac{q}{q+r}$ $\text{Var}(\Theta) = \frac{qr}{(q+r)^2(q+r+1)}$	$q'' = q' + x$ $r'' = r' + n - x$
Exponential $f_X(x) = \lambda e^{-\lambda x}$	λ	Gamma $f_\Lambda(\lambda) = \frac{\nu(\nu\lambda)^{k-1} e^{-\nu\lambda}}{\Gamma(k)}$	$E(\lambda) = \frac{k}{\nu}$ $\text{Var}(\lambda) = \frac{k}{\nu^2}$	$\nu'' = \nu' + \sum_i x_i$ $k'' = k' + n$
Normal $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$ (with known σ)	μ	Normal $f_M(\mu) = \frac{1}{\sqrt{2\pi}\sigma_\mu} \exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_\mu}{\sigma_\mu}\right)^2\right]$	$E(\mu) = \mu_\mu$ $\text{Var}(\mu) = \sigma_\mu^2$	$\mu_\mu'' = \frac{\mu_\mu'(\sigma^2/n) + \bar{x}\sigma_\mu'^2}{\sigma^2/n + (\sigma_\mu')^2}$ $\sigma_\mu'' = \sqrt{\frac{(\sigma_\mu')^2(\sigma^2/n)}{(\sigma_\mu')^2 + \sigma^2/n}}$
Normal $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$	μ, σ	Gamma-Normal $f(\mu,\sigma) = \left\{\frac{1}{\sqrt{2\pi}\sigma/n} \exp\left[-\frac{1}{2}\left(\frac{\mu-\bar{x}}{\sigma/\sqrt{n}}\right)^2\right]\right\} \cdot \left\{\frac{[(n-1)/2]^{(n+1)/2}}{\Gamma[(n+1)/2]} \left(\frac{s^2}{\sigma^2}\right)^{(n-1)/2} \cdot \exp\left(-\frac{n-1}{2}\frac{s^2}{\sigma^2}\right)\right\}$	$E(\mu) = \bar{x}$ $\text{Var}(\mu) = s^2\left[\frac{n-1}{n(n-3)}\right]$ $E(\sigma) = s\sqrt{\frac{n-1}{2}} \frac{\Gamma[(n-2)/2]}{\Gamma[(n-1)/2]}$ $\text{Var}(\sigma) = s^2\left(\frac{n-1}{n-3}\right) - E^2(\sigma)$	$n'' = n' + n$ $n''\bar{x}'' = n'\bar{x}' + n\bar{x}$ $(n''-1)s''^2 + n''\bar{x}''^2 = [(n'-1)s'^2 + n'\bar{x}'^2] + [(n-1)s^2 + n\bar{x}^2]$
Poisson $p_X(x) = \frac{(\mu t)^x}{x!} e^{-\mu t}$	μ	Gamma $f_M(\mu) = \frac{\nu(\nu\mu)^{k-1} e^{-\nu\mu}}{\Gamma(k)}$	$E(\mu) = \frac{k}{\nu}$ $\text{Var}(\mu) = \frac{k}{\nu^2}$	$\nu'' = \nu' + t$ $k'' = k' + x$
Lognormal $f_X(x) = \frac{1}{\sqrt{2\pi}\zeta x} \cdot \exp\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\zeta}\right)^2\right]$ (with known ζ)	λ	Normal $f_\Lambda(\lambda) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\lambda-\mu}{\sigma}\right)^2\right]$	$E(\lambda) = \mu$ $\text{Var}(\lambda) = \sigma^2$	$\mu'' = \frac{\mu'(\zeta^2/n) + \sigma^2\overline{\ln x}}{\zeta^2/n + \sigma^2}$ $\sigma'' = \sqrt{\frac{\sigma^2(\zeta^2/n)}{\sigma^2 + \zeta^2/n}}$

Exercise 3: failure rate of a motor-driven pump of a NPP

- Assume that the failure rate of the pump is constant, λ
- Assume the following three types of relevant information on this machine:
  - E1: engineering knowledge (description of the design and construction of the pump)
  - E2: past performance of similar pumps in similar plants

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

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

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

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 λ and its 95 percentile

Bayesian Vs Frequentist for large amount of data

- **As already pointed out, the results of the Bayesian and classical analyses converge with large amounts of data. The influence of the prior parameters *α'*, *β'* decreases.**

$$\bar{\lambda}'' = \frac{\alpha''}{\beta''} = \frac{\alpha' + k}{\beta' + t} \to \frac{k}{t} = \hat{\lambda}_{\mathrm{MLE}} \qquad \text{for } k, t \to \infty$$

$$\bar{\sigma}''_{\lambda} = \frac{\sqrt{\alpha''}}{\beta''} = \frac{\sqrt{\alpha' + k}}{\beta' + t} \to 0$$

- **Thus for large amounts of data, the posterior distribution will be highly peaked around the MLE estimate** $\hat{\lambda}_{\mathrm{MLE}} = \frac{k}{t}$

- **It can also be shown that the Bayesian and frequentist 95 percentiles will converge; one should, however, keep in mind the differences:**
  - **BAYESIAN = analyst's subjective uncertainty concerning the value of the random variable λ**
  - **FREQUENTIST = variability in the estimation of λ (true value)**
- **Finally, note that as the evidence increases our state of knowledge on the parameter increases and at 'perfect knowledge' (∞ evidence) the uncertainty on its value is zero: however, the failure process remains inherently aleatory.**