Bayesian Inference Resistant to Outliers, using Super Heavy-tailed Distributions, for the Calculation of Premiums

Speaker: Alain Desgagné
Coauthor: Jean-François Angers

UQAM

August 12, 2006
Introduction.

- **Context:** A sample of $n$ claims is collected for a specified product of an insurance company.
- **Objective:** Determine a distribution for the next claim which is robust to outliers.
- **Method:** Robust combination of the $n$ claims with the prior information, using the Bayesian model and super heavy-tailed densities.
Introduction.

- **Context:** A sample of $n$ claims is collected for a specified product of an insurance company.
- **Objective:** Determine a distribution for the next claim which is robust to outliers.
- **Method:** Robust combination of the $n$ claims with the prior information, using the Bayesian model and super heavy-tailed densities.
Context: A sample of $n$ claims is collected for a specified product of an insurance company.

Objective: Determine a distribution for the next claim which is robust to outliers.

Method: Robust combination of the $n$ claims with the prior information, using the Bayesian model and super heavy-tailed densities.
Bayesian context.

- Let $X_1, \ldots, X_n$ be $n$ random variables conditionally independent given the scale parameter $\sigma$, corresponding to the amount of claims.
- Let the conditional densities of $X_i|\sigma$ be given by $\frac{1}{\sigma} f_i\left(\frac{x_i}{\sigma}\right)$, where $X_i \in \mathbb{R}^+, \sigma \in \mathbb{R}^+, i = 1, \ldots, n$.
- The prior density of $\sigma$ is $\frac{1}{x_0} \pi_{\sigma}\left(\frac{\sigma}{x_0}\right)$, where $x_0 \in \mathbb{R}^+$ is a known scale parameter.
- The posterior density of the scale parameter $\sigma$ is given by
  $$
  \pi(\sigma|x_1, \ldots, x_n) = \frac{\frac{1}{x_0} \pi\left(\frac{\sigma}{x_0}\right) \prod_{i=1}^n \frac{1}{\sigma} f_i\left(\frac{x_i}{\sigma}\right)}{\int_0^\infty \frac{1}{x_0} \pi\left(\frac{\sigma}{x_0}\right) \prod_{i=1}^n \frac{1}{\sigma} f_i\left(\frac{x_i}{\sigma}\right) d\sigma}.
  $$
- The predictive distribution of a next claim $X_{n+1}$ is given by
  $$
  f(y|x_1, \ldots, x_n) = \int_0^\infty \frac{1}{\sigma} f_{n+1}\left(\frac{y}{\sigma}\right) \pi(\sigma|x_1, \ldots, x_n) d\sigma.
  $$
Bayesian context.

- Let $X_1, \ldots, X_n$ be $n$ random variables conditionally independent given the scale parameter $\sigma$, corresponding to the amount of claims.
- Let the conditional densities of $X_i|\sigma$ be given by $\frac{1}{\sigma} f_i\left(\frac{x_i}{\sigma}\right)$, where $X_i \in \mathbb{R}^+, \sigma \in \mathbb{R}^+, i = 1, \ldots, n$.
- The prior density of $\sigma$ is $\frac{1}{x_0} \pi_{\sigma}\left(\frac{\sigma}{x_0}\right)$, where $x_0 \in \mathbb{R}^+$ is a known scale parameter.
- The posterior density of the scale parameter $\sigma$ is given by
  \[
  \pi(\sigma|x_1, \ldots, x_n) = \frac{\frac{1}{x_0} \pi\left(\frac{\sigma}{x_0}\right) \prod_{i=1}^n \frac{1}{\sigma} f_i\left(\frac{x_i}{\sigma}\right)}{\int_{0}^{\infty} \frac{1}{x_0} \pi\left(\frac{\sigma}{x_0}\right) \prod_{i=1}^n \frac{1}{\sigma} f_i\left(\frac{x_i}{\sigma}\right) d\sigma}.
  \]
- The predictive distribution of a next claim $X_{n+1}$ is given by
  \[
  f(y|x_1, \ldots, x_n) = \int_{0}^{\infty} \frac{1}{\sigma} f_{n+1}\left(\frac{y}{\sigma}\right) \pi(\sigma|x_1, \ldots, x_n) d\sigma.
  \]
Bayesian context.

- Let $X_1, ..., X_n$ be $n$ random variables conditionally independent given the scale parameter $\sigma$, corresponding to the amount of claims.
- Let the conditional densities of $X_i|\sigma$ be given by $\frac{1}{\sigma}f_i\left(\frac{x_i}{\sigma}\right)$, where $X_i \in \mathbb{R}^+, \sigma \in \mathbb{R}^+, i = 1, ..., n$.
- The prior density of $\sigma$ is $\frac{1}{x_0}\pi_\sigma\left(\frac{\sigma}{x_0}\right)$, where $x_0 \in \mathbb{R}^+$ is a known scale parameter.
- The posterior density of the scale parameter $\sigma$ is given by
  \[
  \pi(\sigma|x_1, \ldots, x_n) = \frac{\frac{1}{x_0}\pi_\sigma\left(\frac{\sigma}{x_0}\right)\prod_{i=1}^{n}\frac{1}{\sigma}f_i\left(\frac{x_i}{\sigma}\right)}{\int_0^\infty \frac{1}{x_0}\pi_\sigma\left(\frac{\sigma}{x_0}\right)\prod_{i=1}^{n}\frac{1}{\sigma}f_i\left(\frac{x_i}{\sigma}\right)d\sigma}.
  \]
- The predictive distribution of a next claim $X_{n+1}$ is given by
  \[
  f(y|x_1, \ldots, x_n) = \int_0^\infty \frac{1}{\sigma}f_{n+1}\left(\frac{y}{\sigma}\right)\pi(\sigma|x_1, \ldots, x_n)d\sigma.
  \]
Let $X_1, ..., X_n$ be $n$ random variables conditionally independent given the scale parameter $\sigma$, corresponding to the amount of claims.

Let the conditional densities of $X_i|\sigma$ be given by $\frac{1}{\sigma} f_i\left(\frac{x_i}{\sigma}\right)$, where $X_i \in \mathbb{R}^+, \sigma \in \mathbb{R}^+, i = 1, ..., n$.

The prior density of $\sigma$ is $\frac{1}{x_0} \pi_{\sigma}(\frac{\sigma}{x_0})$, where $x_0 \in \mathbb{R}^+$ is a known scale parameter.

The posterior density of the scale parameter $\sigma$ is given by

$$\pi(\sigma|x_1, \ldots, x_n) = \frac{\frac{1}{x_0} \pi_{\sigma}(\frac{\sigma}{x_0}) \prod_{i=1}^{n} \frac{1}{\sigma} f_i\left(\frac{x_i}{\sigma}\right)}{\int_0^{\infty} \frac{1}{x_0} \pi_{\sigma}(\frac{\sigma}{x_0}) \prod_{i=1}^{n} \frac{1}{\sigma} f_i\left(\frac{x_i}{\sigma}\right) d\sigma}.$$ 

The predictive distribution of a next claim $X_{n+1}$ is given by

$$f(y|x_1, \ldots, x_n) = \int_0^{\infty} \frac{1}{\sigma} f_{n+1}\left(\frac{y}{\sigma}\right) \pi_{\sigma}(\sigma|x_1, \ldots, x_n) d\sigma.$$
Bayesian context.

- Let $X_1, ..., X_n$ be $n$ random variables conditionally independent given the scale parameter $\sigma$, corresponding to the amount of claims.
- Let the conditional densities of $X_i|\sigma$ be given by $\frac{1}{\sigma} f_i\left(\frac{x_i}{\sigma}\right)$, where $X_i \in \mathbb{R}^+, \sigma \in \mathbb{R}^+, i = 1, ..., n$.
- The prior density of $\sigma$ is $\frac{1}{x_0} \pi_\sigma\left(\frac{\sigma}{x_0}\right)$, where $x_0 \in \mathbb{R}^+$ is a known scale parameter.
- The posterior density of the scale parameter $\sigma$ is given by
  \[
  \pi(\sigma|x_1, \ldots, x_n) = \frac{\frac{1}{x_0} \pi\left(\frac{\sigma}{x_0}\right) \prod_{i=1}^n \frac{1}{\sigma} f_i\left(\frac{x_i}{\sigma}\right)}{\int_0^\infty \frac{1}{x_0} \pi\left(\frac{\sigma}{x_0}\right) \prod_{i=1}^n \frac{1}{\sigma} f_i\left(\frac{x_i}{\sigma}\right) d\sigma}.
  \]
- The predictive distribution of a next claim $X_{n+1}$ is given by
  \[
  f(y|x_1, \ldots, x_n) = \int_0^\infty \frac{1}{\sigma} f_{n+1}\left(\frac{y}{\sigma}\right) \pi(\sigma|x_1, \ldots, x_n) d\sigma.
  \]
Robustness to outliers depends on the choice of the prior and the likelihood.

For example, log-normal distributions produce sensitive inference to outliers.

Outliers in this context is conflicting information, which can be an extreme observation as well as a misspecification of the scale parameter of the prior density.

The tails of the prior and the likelihood determine if the posterior density of $\sigma$ and the predictive distribution of a next claim are robust to outliers.
Robustness to outliers depends on the choice of the prior and the likelihood.

For example, log-normal distributions produce sensitive inference to outliers.

Outliers in this context is conflicting information, which can be an extreme observation as well as a misspecification of the scale parameter of the prior density.

The tails of the prior and the likelihood determine if the posterior density of $\sigma$ and the predictive distribution of a next claim are robust to outliers.
Conditions of robustness.

- Robustness to outliers depends on the choice of the prior and the likelihood.
- For example, log-normal distributions produce sensitive inference to outliers.
- Outliers in this context is conflicting information, which can be an extreme observation as well as a misspecification of the scale parameter of the prior density.
- The tails of the prior and the likelihood determine if the posterior density of $\sigma$ and the predictive distribution of a next claim are robust to outliers.
Robustness to outliers depends on the choice of the prior and the likelihood.

For example, log-normal distributions produce sensitive inference to outliers.

Outliers in this context is conflicting information, which can be an extreme observation as well as a misspecification of the scale parameter of the prior density.

The tails of the prior and the likelihood determine if the posterior density of $\sigma$ and the predictive distribution of a next claim are robust to outliers.
Conditions of robustness.

- Our paper established the conditions of robustness. Simply stated, the theoretical results say that:
  - 1) if the tails of the prior and the likelihood are sufficiently heavy,
  - 2) if the number of conflicting information is less or equal to half of the observations,
  then
- \( \sigma|x_n \xrightarrow{L} \sigma|x_k \) as the outliers tend to 0 or infinity, where \( x_k \) is the vector of non-outliers, and the density of the random variables \( \sigma|x_n \) and \( \sigma|x_k \) evaluated at the point \( y \) are given by \( \pi(y|x_n) \) and \( \pi(y|x_k) \).
Conditions of robustness.

- Our paper established the conditions of robustness. Simply stated, the theoretical results say that:
- 1) if the tails of the prior and the likelihood are sufficiently heavy,
- 2) if the number of conflicting information is less or equal to half of the observations, then
- \( \sigma |_{\hat{x}_n} \xrightarrow{L} \sigma |_{\hat{x}_k} \) as the outliers tend to 0 or infinity, where \( \hat{x}_k \) is the vector of non-outliers, and the density of the random variables \( \sigma |_{\hat{x}_n} \) and \( \sigma |_{\hat{x}_k} \) evaluated at the point \( y \) are given by \( \pi(y |_{\hat{x}_n}) \) and \( \pi(y |_{\hat{x}_k}) \).
Our paper established the conditions of robustness. Simply stated, the theoretical results say that:

1) if the tails of the prior and the likelihood are sufficiently heavy,

2) if the number of conflicting information is less or equal to half of the observations,

\[ \sigma|_{\tilde{x}_n} \xrightarrow{L} \sigma|_{\tilde{x}_k} \] as the outliers tend to 0 or infinity, where \( \tilde{x}_k \) is the vector of non-outliers, and the density of the random variables \( \sigma|_{\tilde{x}_n} \) and \( \sigma|_{\tilde{x}_k} \) evaluated at the point \( y \) are given by \( \pi(y|_{\tilde{x}_n}) \) and \( \pi(y|_{\tilde{x}_k}) \).
Our paper established the conditions of robustness. Simply stated, the theoretical results say that:

1) if the tails of the prior and the likelihood are sufficiently heavy,
2) if the number of conflicting information is less or equal to half of the observations,

\[ \sigma|x_n \xrightarrow{L} \sigma|x_k \] as the outliers tend to 0 or infinity, where \( x_k \) is the vector of non-outliers, and the density of the random variables \( \sigma|x_n \) and \( \sigma|x_k \) evaluated at the point \( y \) are given by \( \pi(y|x_n) \) and \( \pi(y|x_k) \).
Conditions of robustness.

- What “sufficiently heavy” means?
  - Densities with exponential tails such as Normal and gamma densities are not sufficiently enough to produce robust inference.
  - Heavy-tailed densities such as Student and Pareto are not sufficiently enough to produce complete robust inference.
  - However, they will produce “partial” robustness, in the sense that an outlier will have an impact on the inference, but this impact will be limited.
  - Super heavy-tailed densities, such as log-Student or log-Pareto densities are sufficiently heavy and satisfy the condition of complete robustness.
  - The impact of conflicting information will disappear gradually as the conflict increase.
Conditions of robustness.

- What “sufficiently heavy” means?

- Densities with exponential tails such as Normal and gamma densities are not sufficiently enough to produce robust inference.

- Heavy-tailed densities such as Student and Pareto are not sufficiently enough to produce complete robust inference.

- However, they will produce “partial” robustness, in the sense that an outlier will have an impact on the inference, but this impact will be limited.

- Super heavy-tailed densities, such as log-Student or log-Pareto densities are sufficiently heavy and satisfy the condition of complete robustness.

- The impact of conflicting information will disappear gradually as the conflict increase.
Conditions of robustness.

- What “sufficiently heavy” means?
- Densities with exponential tails such as Normal and gamma densities are not sufficiently enough to produce robust inference.
- Heavy-tailed densities such as Student and Pareto are not sufficiently enough to produce complete robust inference.
- However, they will produce “partial” robustness, in the sense that an outlier will have an impact on the inference, but this impact will be limited.
- Super heavy-tailed densities, such as log-Student or log-Pareto densities are sufficiently heavy and satisfy the condition of complete robustness.
- The impact of conflicting information will disappear gradually as the conflict increase.
Conditions of robustness.

- What “sufficiently heavy” means?
- Densities with exponential tails such as Normal and gamma densities are not sufficiently enough to produce robust inference.
- Heavy-tailed densities such as Student and Pareto are not sufficiently enough to produce complete robust inference.
- However, they will produce “partial” robustness, in the sense that an outlier will have an impact on the inference, but this impact will be limited.
- Super heavy-tailed densities, such as log-Student or log-Pareto densities are sufficiently heavy and satisfy the condition of complete robustness.
- The impact of conflicting information will disappear gradually as the conflict increase.
Conditions of robustness.

- What “sufficiently heavy” means?

- Densities with exponential tails such as Normal and gamma densities are not sufficiently enough to produce robust inference.

- Heavy-tailed densities such as Student and Pareto are not sufficiently enough to produce complete robust inference.

- However, they will produce “partial” robustness, in the sense that an outlier will have an impact on the inference, but this impact will be limited.

- Super heavy-tailed densities, such as log-Student or log-Pareto densities are sufficiently heavy and satisfy the condition of complete robustness.

- The impact of conflicting information will disappear gradually as the conflict increase.
Robustness

Conditions of robustness.

- What “sufficiently heavy” means?
- Densities with exponential tails such as Normal and gamma densities are not sufficiently enough to produce robust inference.
- Heavy-tailed densities such as Student and Pareto are not sufficiently enough to produce complete robust inference.
- However, they will produce “partial” robustness, in the sense that an outlier will have an impact on the inference, but this impact will be limited.
- Super heavy-tailed densities, such as log-Student or log-Pareto densities are sufficiently heavy and satisfy the condition of complete robustness.
- The impact of conflicting information will disappear gradually as the conflict increase.
Example.

- We observe 5 claims $X_1, \ldots, X_5 = 380, 420, 600, 650, 760$.
- We choose a non-informative distribution for the prior:
  \[
  \frac{1}{x_0} \pi_{\sigma} \left( \frac{\sigma}{x_0} \right) \propto \frac{1}{\sigma}
  \]
- We compare two models for the likelihood: the log Normal and the log Student.
- Log Normal:
  \[
  \frac{1}{\sigma} f_i \left( \frac{x_i}{\sigma} \right) = \frac{1}{sx_i} N \left( \frac{\log x_i - \log \sigma}{s} \right),
  \]
  where $N(\cdot)$ is the density of a $N(0,1)$
Example.

- We observe 5 claims $X_1, \ldots, X_5 = 380, 420, 600, 650, 760$.
- We choose a non-informative distribution for the prior:
  $$\frac{1}{x_0} \pi_\sigma \left( \frac{\sigma}{x_0} \right) \propto \frac{1}{\sigma}$$
- We compare two models for the likelihood: the log Normal and the log Student.
- Log Normal:
  $$\frac{1}{\sigma} f_i \left( \frac{x_i}{\sigma} \right) = \frac{1}{sx_i} N \left( \frac{\log x_i - \log \sigma}{s} \right),$$
  where $N(\cdot)$ is the density of a $N(0,1)$
We observe 5 claims $X_1, \ldots, X_5 = 380, 420, 600, 650, 760$.

We choose a non-informative distribution for the prior:

$$
\frac{1}{x_0} \pi_\sigma \left( \frac{\sigma}{x_0} \right) \propto \frac{1}{\sigma}
$$

We compare two models for the likelihood: the log Normal and the log Student.

Log Normal:

$$
\frac{1}{\sigma} f_i \left( \frac{x_i}{\sigma} \right) = \frac{1}{sx_i} N \left( \frac{\log x_i - \log \sigma}{s} \right),
$$

where $N(\cdot)$ is the density of a $N(0,1)$
Example.

- We observe 5 claims $X_1, \ldots, X_5 = 380, 420, 600, 650, 760$.
- We choose a non-informative distribution for the prior:
  
  $$
  \frac{1}{x_0} \pi_\sigma \left( \frac{\sigma}{x_0} \right) \propto \frac{1}{\sigma}
  $$

- We compare two models for the likelihood: the log Normal and the log Student.
  - Log Normal:
    
    $$
    \frac{1}{\sigma} f_i \left( \frac{x_i}{\sigma} \right) = \frac{1}{sx_i} N \left( \frac{\log x_i - \log \sigma}{s} \right),
    $$

    where $N(\cdot)$ is the density of a $N(0,1)$
We observe 5 claims $X_1, \ldots, X_5 = 380, 420, 600, 650, 760$.

We choose a non-informative distribution for the prior:

$$\frac{1}{x_0} \pi_{\sigma}(\frac{\sigma}{x_0}) \propto \frac{1}{\sigma}$$

We compare two models for the likelihood: the log Normal and the log Student.

Log Normal:

$$\frac{1}{\sigma} f_i(\frac{x_i}{\sigma}) = \frac{1}{sx_i} N \left( \log x_i - \log \sigma, \frac{1}{s} \right),$$

where $N(\cdot)$ is the density of a $N(0,1)$
Example.

- **Log Student:**

  \[
  \frac{1}{\sigma} f_i \left( \frac{x_i}{\sigma} \right) = \frac{1}{s x_i} T \left( \frac{\log x_i - \log \sigma}{s} \right),
  \]

  where \( T(\cdot) \) is the density of a Student with 5 degrees of freedom.

- The scale parameter \( \sigma \) behave more like a location parameter while the parameter \( s \) behave like a scale parameter.

- When the densities are expressed in term of \( \sigma \) as it is the case in the posterior density, \( x_i \) becomes the scale parameter (and behave as a location parameter).
Example.

- Log Student:

\[
\frac{1}{\sigma} f_i \left( \frac{x_i}{\sigma} \right) = \frac{1}{s x_i} T \left( \frac{\log x_i - \log \sigma}{s} \right),
\]

where \( T(\cdot) \) is the density of a Student with 5 degrees of freedom.

- The scale parameter \( \sigma \) behave more like a location parameter while the parameter \( s \) behave like a scale parameter.

- When the densities are expressed in term of \( \sigma \) as it is the case in the posterior density, \( x_i \) becomes the scale parameter (and behave as a location parameter).
Example.

- **Log Student:**

  \[
  \frac{1}{\sigma} f_i \left( \frac{x_i}{\sigma} \right) = \frac{1}{sx_i} T \left( \frac{\log x_i - \log \sigma}{s} \right),
  \]

  where \( T(\cdot) \) is the density of a Student with 5 degrees of freedom.

- The scale parameter \( \sigma \) behave more like a location parameter while the parameter \( s \) behave like a scale parameter.

- When the densities are expressed in term of \( \sigma \) as it is the case in the posterior density, \( x_i \) becomes the scale parameter (and behave as a location parameter).
Log–Normal and Log–Student densities with parameters set as scale=5 and shape=0.5

Speaker: Alain Desgagné
Coauthor: Jean-François Angers (UQAM)

Bayesian Inference Resistant to Outliers, using Super Heavy-tailed Distributions, for the Calculation of Premiums

August 12, 2006 9 / 16
Log–Normal and Log–Student densities
with parameters set as scale=5 and shape=0.5
Since the tails are too heavy for the posterior mean to exists, we estimate $\sigma$ with the posterior median.

We look at the posterior median of $\sigma$ for different values of $x_5$ for both models.

If the observation $x_5$ is removed from the analysis, we find that:
- the posterior median of $\sigma$ for the log Normal model is 4.8 (all numbers are expressed in hundreds)
- the posterior median of $\sigma$ for the log Student model is 4.6
Since the tails are too heavy for the posterior mean to exists, we estimate $\sigma$ with the posterior median.

We look at the posterior median of $\sigma$ for different values of $x_5$ for both models.

If the observation $x_5$ is removed from the analysis, we find that:

- the posterior median of $\sigma$ for the log Normal model is 4.8 (all numbers are expressed in hundreds)
- the posterior median of $\sigma$ for the log Student model is 4.6
Since the tails are too heavy for the posterior mean to exists, we estimate $\sigma$ with the posterior median.

We look at the posterior median of $\sigma$ for different values of $x_5$ for both models.

If the observation $x_5$ is removed from the analysis, we find that:

- the posterior median of $\sigma$ for the log Normal model is 4.8 (all numbers are expressed in hundreds)
- the posterior median of $\sigma$ for the log Student model is 4.6
Since the tails are too heavy for the posterior mean to exist, we estimate $\sigma$ with the posterior median.

We look at the posterior median of $\sigma$ for different values of $x_5$ for both models.

If the observation $x_5$ is removed from the analysis, we find that:

- the posterior median of $\sigma$ for the log Normal model is 4.8 (all numbers are expressed in hundreds)
- the posterior median of $\sigma$ for the log Student model is 4.6
Since the tails are too heavy for the posterior mean to exist, we estimate $\sigma$ with the posterior median.

We look at the posterior median of $\sigma$ for different values of $x_5$ for both models.

If the observation $x_5$ is removed from the analysis, we find that:

- the posterior median of $\sigma$ for the log Normal model is 4.8 (all numbers are expressed in hundreds)
- the posterior median of $\sigma$ for the log Student model is 4.6
Posterior Median of Sigma for Different Values of X5

Estimation of sigma

- Log Normal Model
- Log Student Model

Speaker: Alain Desgagné
Coauthor: Jean-François Angers (UQAM)

Bayesian Inference Resistant to Outliers, using Super Heavy-tailed Distributions, for the Calculation of Premiums

August 12, 2006 12 / 16
Posterior Median of Sigma for Different Values of X5

Estimation of sigma

Log Student Model

Speaker: Alain Desgagné
Coauthor: Jean-François Angers (UQAM)

Bayesian Inference Resistant to Outliers, using Super Heavy-tailed Distributions, for the Calculation of Premiums

August 12, 2006 13 / 16
Prior, Likelihood and Posterior with the Student Model, when $X_5=7.6$
Prior, Likelihood and Posterior with the Student Model, when $X_5=15$
Conclusion.

- Robust Bayesian Inference for scale parameter is possible.
- Modelling the prior and the likelihood using super heavy-tailed distributions satisfy the conditions of robustness.
- Calculation of premiums resistant to outliers is then possible using the predictive distribution.
Conclusion.

- Robust Bayesian Inference for scale parameter is possible.
- Modelling the prior and the likelihood using super heavy-tailed distributions satisfy the conditions of robustness.
- Calculation of premiums resistant to outliers is then possible using the predictive distribution.
Conclusion.

- Robust Bayesian Inference for scale parameter is possible.
- Modelling the prior and the likelihood using super heavy-tailed distributions satisfy the conditions of robustness.
- Calculation of premiums resistant to outliers is then possible using the predictive distribution.