UNIVERSITÀ DI PISA

Combining Estimates

Thomas Struppeck, FCAS, ASA, CERA

2014

Statistics for data science

A.Y. 2023/2024

Combining estimates

2 scenarios

Combining estimates for the same quantity

We will actually be able to improve on both estimates

Combining estimates for multiple components

We can only hope to not be much worse than the worst one

GOAL

Combining estimates into a single estimate

* In all the frameworks we assume normal distributions as in the paper

1. Combining Two Or More Estimates For The Same Quantity

Framework

When estimating ultimate losses, there may be one estimate based on paid losses and another estimate based on incurred losses

Setting

Two independent estimates, A and B , are available for an unknown parameter θ.

Example
- θ : unpaid claims on a block of business
- θ_A θ_B : two different estimates of θ

Goal

Create a single estimator that allows these two estimates to work in tandem in a way that maximizes the precision of the resulting estimator

Problem: how to combine estimates?

How to combine estimates?

Consider the one-parameter family of estimators obtained by taking **weighted averages of A and B**.
When t is zero, we get the first estimator, and when t is 1, we get the second estimator.

Since each of θ_A and θ_B are normally distributed, and they are independent, the weighted average is normal with mean and variance shown in formula.

We want to **maximize the precision**, which amounts to **minimizing the standard deviation**, which is the same as **minimizing the variance.**

This is easily done by taking the derivative with respect to t and setting it to zero.

Weighted variance

$$(1-t)^2 Var[\hat{\theta}_A] + t^2 Var[\hat{\theta}_B]$$

Combined estimator

$$\hat{\theta}_t = (1-t)\hat{\theta}_A + t\hat{\theta}_B \text{ where } 0 \leq t \leq 1$$

Weighted average

$$(1-t)E[\hat{\theta}_A] + tE[\hat{\theta}_B]$$

Optimization of weights

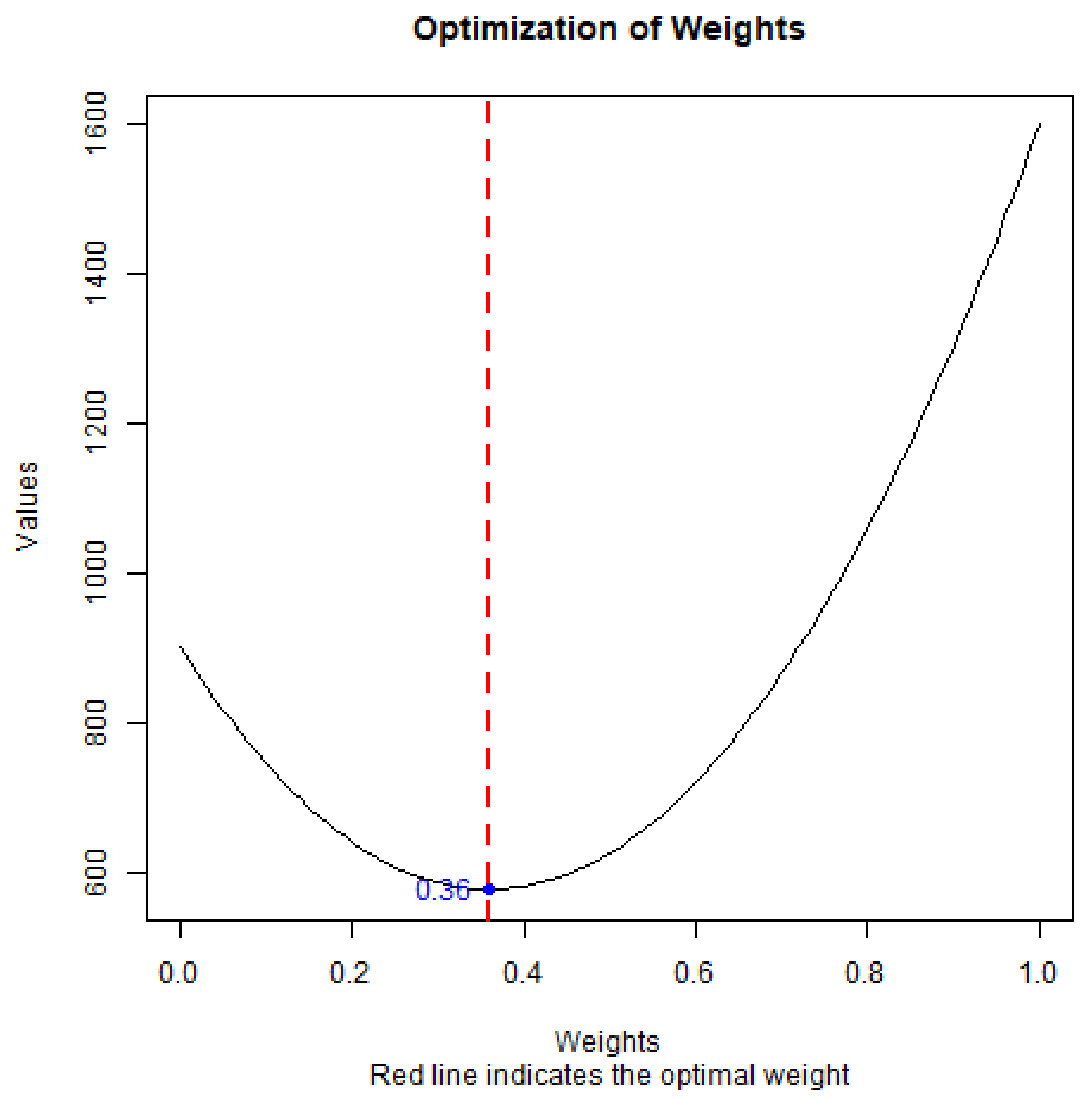
On paper data

Considering:

- θ_A
  - Mean: 250
  - Sd: 30
  - 95% CI: ($190, $310)
- θ_B
  - Mean: 275
  - Sd: 40
  - 95% CI: ($195, $355)

"optimize" R function was applied to find the optimal weight.

Result: 0,36 as in the paper

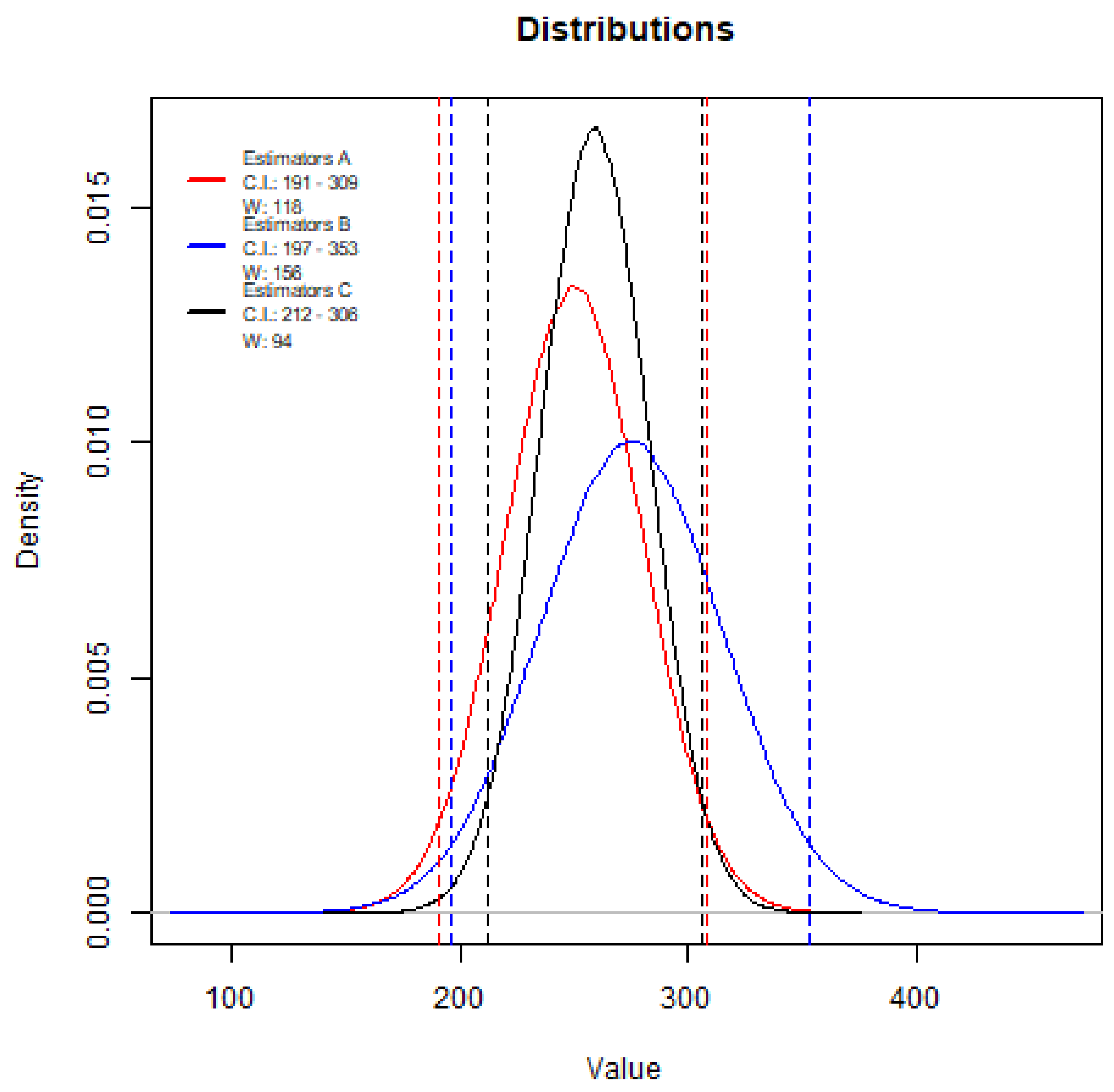


Distributions compared

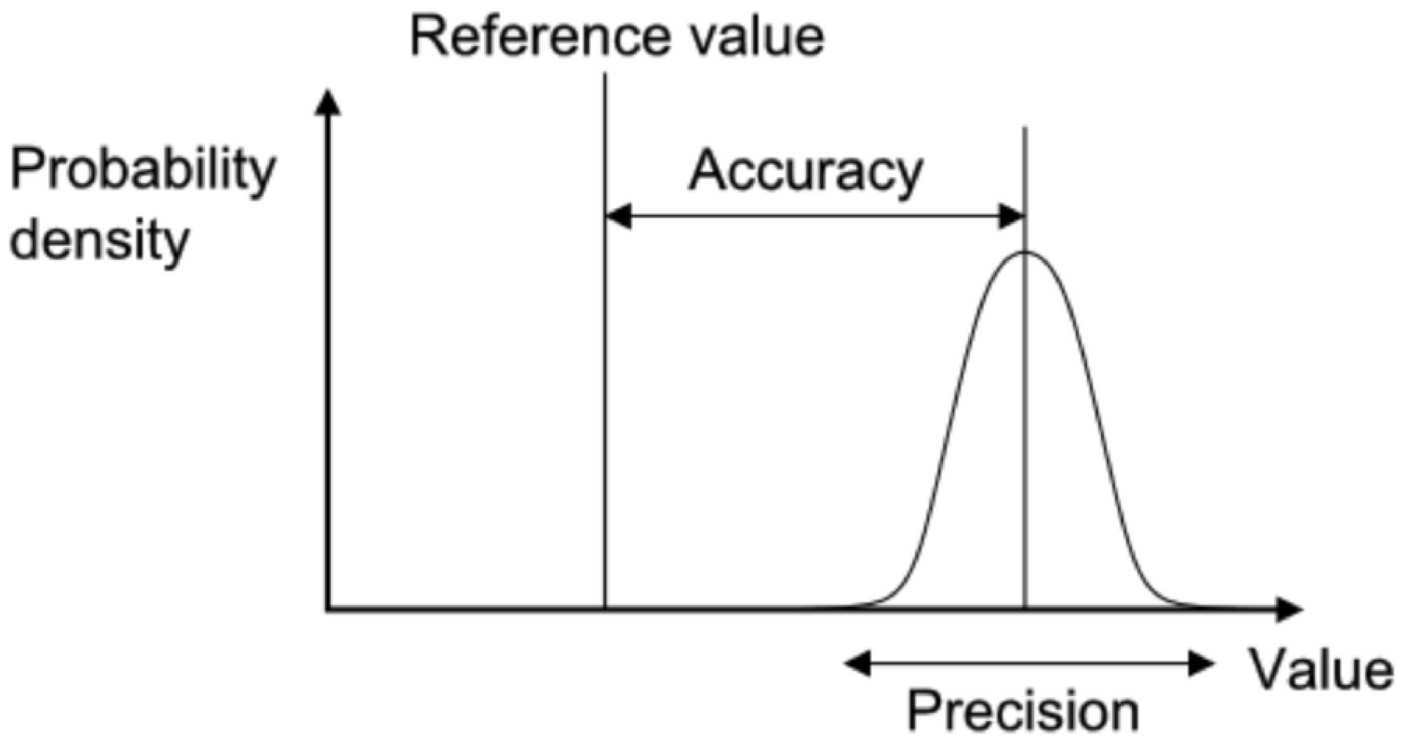
highlighted 95% CI for each estimator

The combined one is **more precise**, since its CI is narrowed.

- θ_C
  - Mean: 259
  - Sd: 24.01
  - 95% CI: ($212, $306)



The weighted average, having a more concentrated density, is more precise, but **is it more accurate?**



Backdrop

Accuracy vs Precision

Accuracy measures how far the true parameter is from our estimate

Precision refers to the spread of estimator values around its expected value

Distance from the best

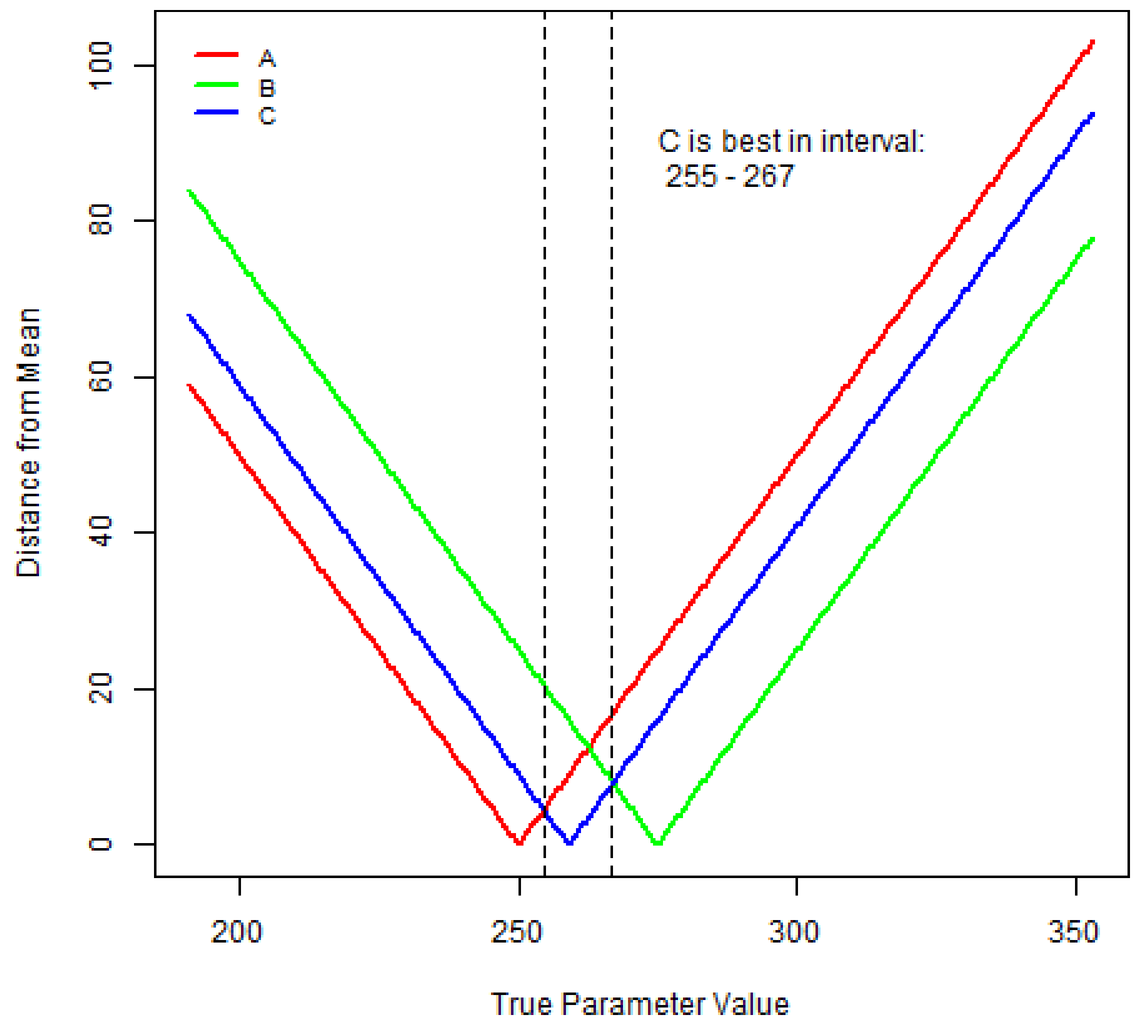
(concept of accuracy)

Since the true parameter is **intrinsically unknowable** we simulate different "true" hypothetical parameters (ranging in our case from lowest/highest 95% CI intervals).

We calculate the distances from this parameters using the estimated means for each estimators.

Result: the combined one (C) is the best in the 255 - 267 interval

in all the other cases it is not the worst
meaning that the blue line is always between the others (the distance is not the highest).



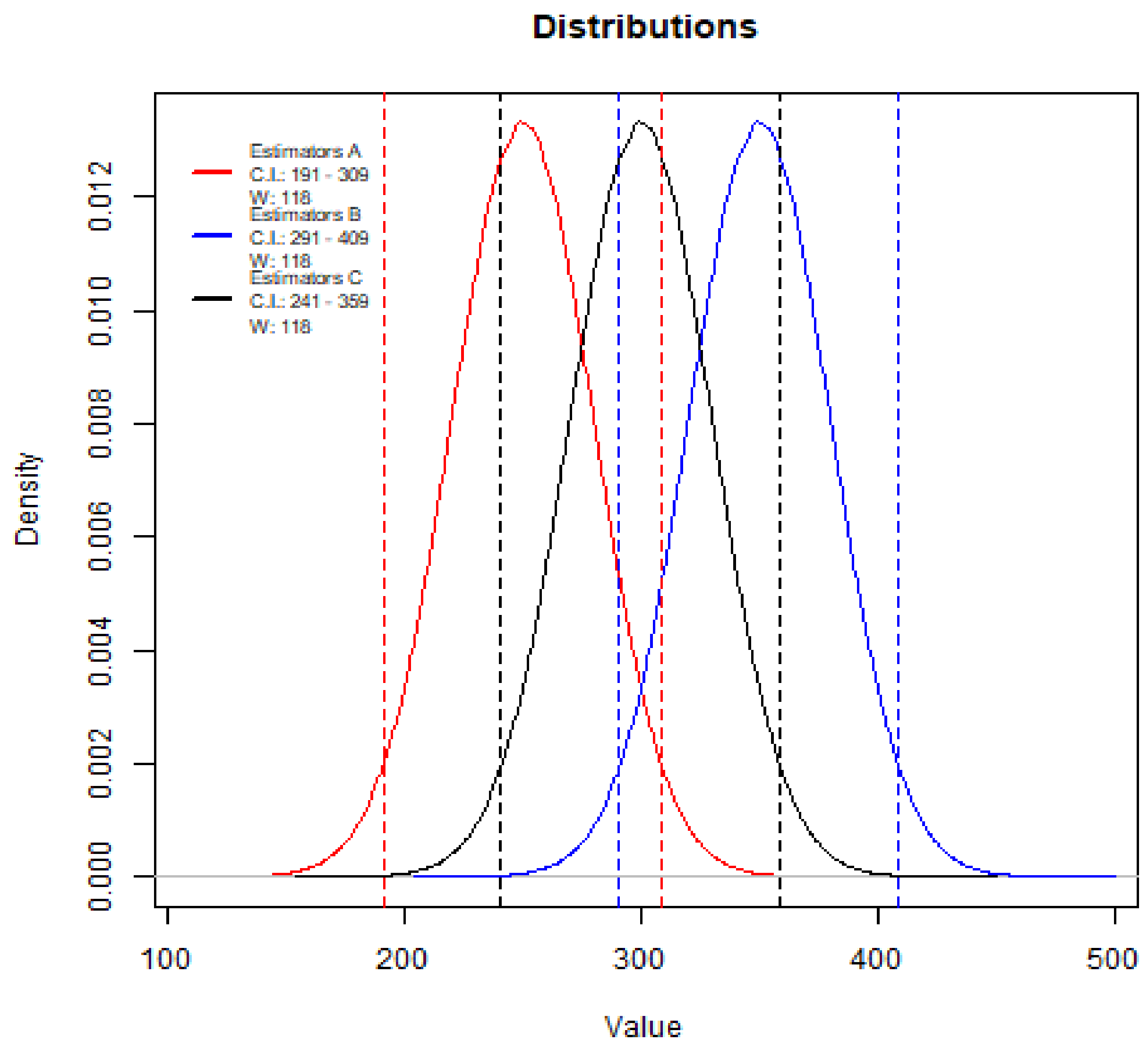
Estimators perfectly correlated

“ Some increase in precision will occur when estimators are combined unless they are 100% correlated ”

Setting estimators B as estimators A + Constant:

- $\boldsymbol{\sigma}_A = \boldsymbol{\sigma}_B$
- $\rho_{AB} = 1$

Result : no increase in precision in the combined Estimator C



Correlation simulation

highlighted 95% CI for each estimator

Considering θ_A and θ_B as in the paper.

Data generated with mvrnorm function to consider the correlation constraints.

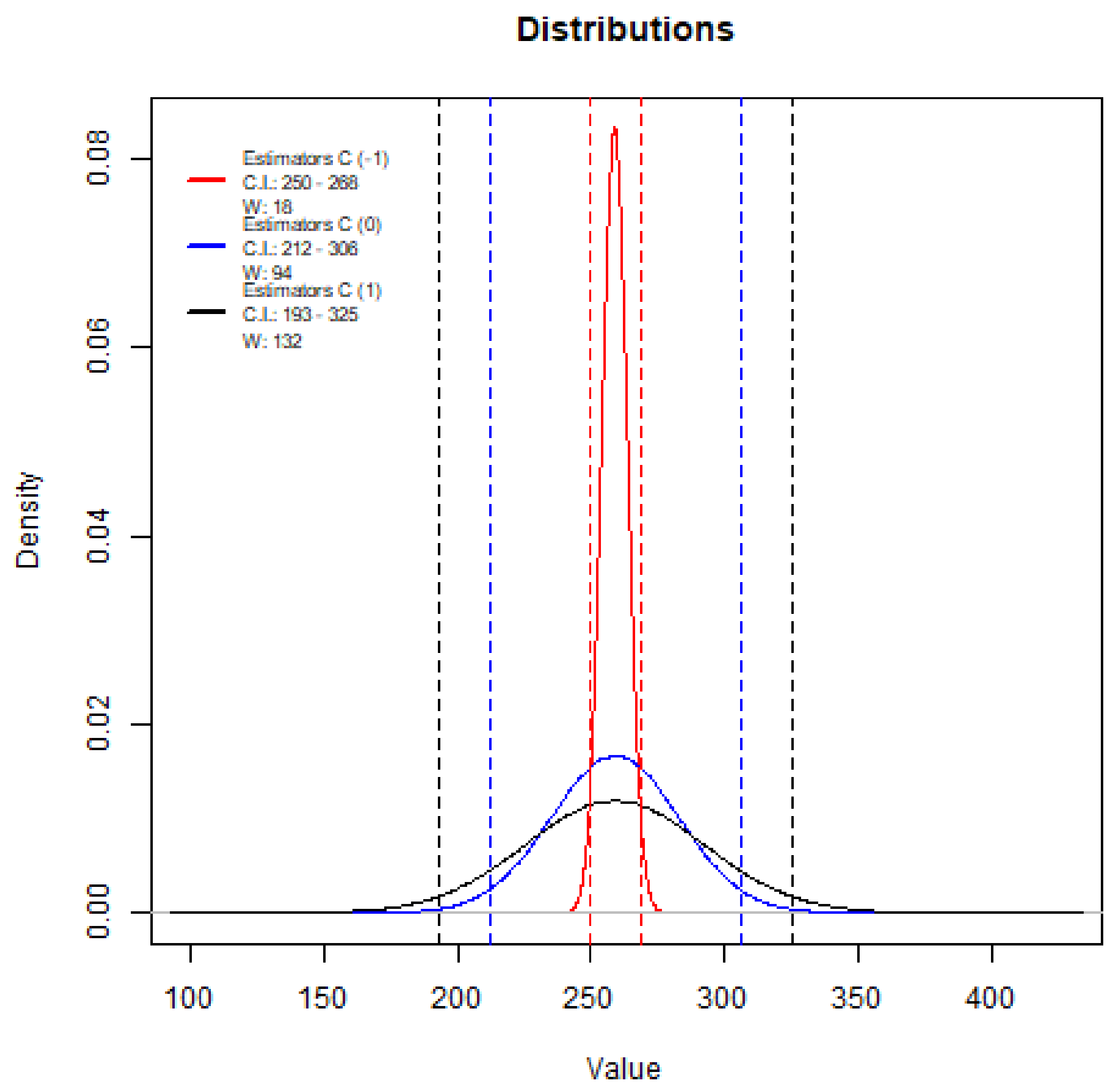
Simulating iteratively increasing correlation coefficient [-1,1] with a 0.1 step.

Result:

$\rho_{AB} = -1 \rightarrow$ the most precise

$\rho_{AB} = 0 \rightarrow$ no correlation

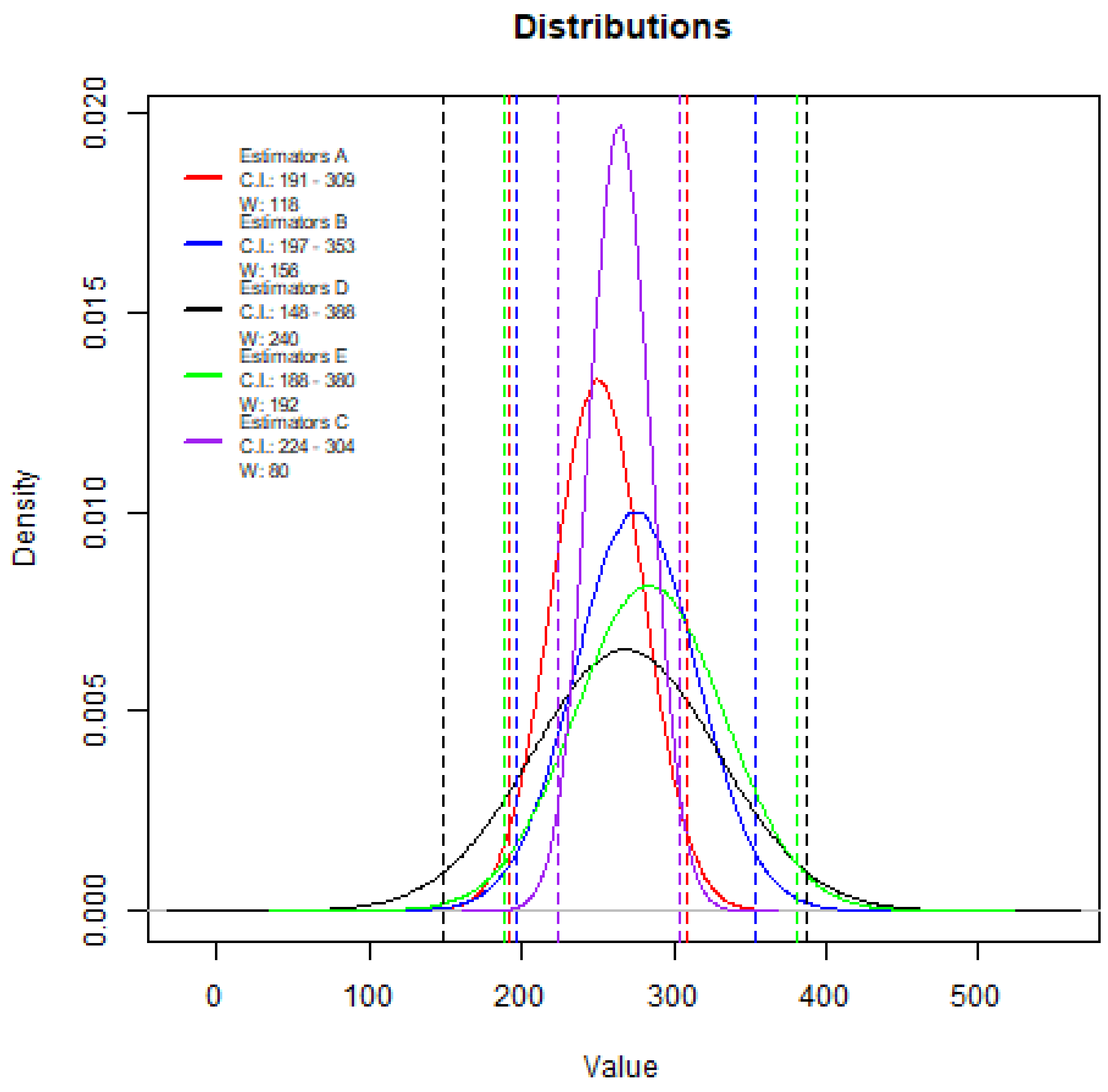
$\rho_{AB} = 1 \rightarrow$ the worst precise



General case

more than two estimates

$$\hat{w}_i = \frac{w_i}{\sum_{j=1}^{n} w_j} = \frac{\frac{1}{\sigma_i^2}}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}} \quad \text{where} \quad w_i = \frac{1}{\sigma_i^2}$$



Consider the **inverse variance weighting** (IVW) formula as the inverse variance of the observation i.

As in the two estimators case we obtain a new combined estimators with the highest precision.

Since the optimization ensure to take the best weights according to the variances, giving more importance to the lower ones.

2. Combining Estimates for Multiple Components

Framework

Estimates for several lines of business that are to be combined into a single estimate

Setting

Three line loss, A, B, C , and we want to obtain an estimate for the sum.
Reasonable to assume that the components are not independent, so assume some degree of correlations among them.
Motivation: crisis related to complementary products

Goal

Create a single estimator that allows to estimate the sum of the total loss

Problem: how to combine estimates?

Example with 3 lines

Starting from the known lower and upper percentile, we extract the expected losses and sd for each line (developing the **percentile_matching** function).

3 type of estimators was developed:

Line of business	Expected Loss	25th percentile	75th percentile
A	100	90	110
B	225	150	300
C	350	200	500

Naïve total

It consider the 100% correlation so it is basically the sum of each component results

With covariance adjustment

Applying the variance covariance matrix to adjust the calculus of the sd with the given correlation

No correlation

Ignoring the correlation among components

$$\Sigma = \begin{pmatrix} \sigma_A & 0 & 0 \\ 0 & \sigma_B & 0 \\ 0 & 0 & \sigma_C \end{pmatrix} \begin{pmatrix} 1 & \rho_{BA} & \rho_{CA} \\ \rho_{AB} & 1 & \rho_{CB} \\ \rho_{AC} & \rho_{BC} & 1 \end{pmatrix} \begin{pmatrix} \sigma_A & 0 & 0 \\ 0 & \sigma_B & 0 \\ 0 & 0 & \sigma_C \end{pmatrix}$$

Results and comparison with generated data

$$CV = \frac{\sigma}{\mu}$$

Line of business	Expected Loss	25th percentile	75th percentile	sd	cv
A	100	90	110	14.8	0.148
B	225	150	300	111.2	0.494
C	350	200	500	222.4	0.653
Naive total	675	440	910	348.4	0.516
With cov.adj.	675	465.3	884.7	310.9	0.461
No correlation	675	507	843	249	0.369

Paper results

Line of business	Expected Loss	25th percentile	75th percentile	sd	cv
A	100.01	90.04	109.98	14.81	0.148
B	224.99	149.99	299.98	111.2	0.494
C	349. 85	199.64	499.99	222.6	0.636
Naive total	674.87	439.7	910.03	348.6	0.516
With cov.adj.	674.87	464.99	884.7	311.7	0.461
No correlation	674.87	506.72	843.02	249.3	0.369

Our results

Example through a User application

Goal

Put all the reasoning about 3 lines done so far into an interactive tool to explore and adjust estimates

Tools

- **Shiny** library for the environment structure and application logic
- **DT** library for interactive table

User interaction

The user can insert multiple lines with their percentiles details and modify the correlation structure

let's see in practice...