Probabilistic reconciliation via conditioning

Giorgio Corani (giorgio.corani@idsia.ch)

IDSIA, Switzerland

Research group



Dario Azzimonti, Nicolò Rubattu e Lorenzo Zambon

- Probabilistic forecasting and reconciliation.
- Bayesian data analysis, Bayesian networks, Gaussian Processes.

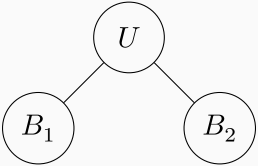
SUPSI (Lugano, Switzerland)

- We are open to collaboration, exchange of students, visiting.

Outline

- Gaussian variables
  - analytical reconciliation
  - relation with Kalman Filter and minT
  - properties of reconciled mean and variance
- Count variables
  - reconciliation via virtual evidence
  - properties of reconciled mean and variance
  - sampling
- Software
- Open problems

Notation



$\mathbf{B}_{t+h}$: actual values of bottom time series at time $t + h$

$\mathbf{U}_{t+h}$: actual values of upper time series at time $t + h$

$\mathbf{Y}_{t+h}$: actual values of all time series at time $t + h$

Summing matrix $\mathbf{S}$

$$\mathbf{Y}_{t+h} = \mathbf{SB}_{t+h}$$

$$\mathbf{S} = \begin{bmatrix} \mathsf{A} \\ \mathsf{I} \end{bmatrix}$$

Forecast notation

- $\hat{\mathbf{b}}_{t+h}$, $\hat{\mathbf{u}}_{t+h}$: base *point forecasts* for bottom and upper ts
- $\tilde{\mathbf{b}}_{t+h}$, $\tilde{\mathbf{u}}_{t+h}$: reconciled *point forecasts* for bottom and upper ts
- $\widehat{\mathbf{\Sigma}}_b$, $\widehat{\mathbf{\Sigma}}_u$: covariance matrices (1-step ahead) of the residuals of the base forecasts for bottom and upper ts
- $\widetilde{\mathbf{\Sigma}}_b$: reconciled covariance matrixes for bottom time series.

k_h

The covariance matrix of the base forecast for time $t+h$ are assumed to be $k_h\widehat{\mathbf{\Sigma}}_b$ and $k_h\widehat{\mathbf{\Sigma}}_u$.

Gaussian base forecasts

- We start from a Gaussian predictive distribution on each time series.
- We then assume a jointly Gaussian predictive distribution over the bottom time series:

$$\hat{p}(\mathbf{B}_{t+h}) = N\left(\hat{\mathbf{b}}_{t+h}, k_h \widehat{\boldsymbol{\Sigma}}_b\right)$$

Probabilistic bottom-up

- We get the joint density as:

$$p_{BU}(\mathbf{Y}_{t+h}) = p_{BU}(\mathbf{S}\mathbf{B}_{t+h}) = N\left(\mathbf{S}\hat{\mathbf{b}}_{t+h}, k_h\mathbf{S}\widehat{\mathbf{\Sigma}}_b\mathbf{S}'\right).$$

Conditioning on the base forecast of the upper variables

We model the point forecasts of the upper time series as noisy observations of sums of the bottom time series:

$$\widehat{\mathbf{U}}_{t+h} \sim N\left(A\mathbf{B}_{t+h}, k_h \widehat{\boldsymbol{\Sigma}}_u\right)$$

- $\widehat{\mathbf{U}}_{t+h}$ denotes the random variable, while $\hat{\mathbf{u}}_{t+h}$ denotes its specific value.

Conditioning to obtain the reconciled density of the bottom time series

$$\tilde{p}(\mathbf{B}_{t+h}) = p(\mathbf{B}_{t+h} \mid \hat{\mathbf{u}}_{t+h}) \propto \hat{p}(\mathbf{B}_{t+h})\hat{p}\left(\hat{\mathbf{u}}_{t+h} \mid \mathbf{B}_{t+h}\right)$$

- The conditioning is solved using the linear-Gaussian model, obtaining the normal reconciled density $\tilde{p}(\mathbf{B}_{t+h})$ (Corani et al, ECML 2020).
- We get the reconciled joint $\tilde{p}(\mathbf{Y}_{t+h})$ by “pre-multiplying” $\tilde{p}(\mathbf{B}_{t+h})$ by $\mathbf{S}$.

Relation with the update step of the Kalman filter

- Reconciliation requires the same type of conditioning of the state-update of the Kalman filter, which is also solved by the linear-Gaussian model.

$$\begin{cases} \mathbf{B}_{t+h} & \sim N(\hat{b}, k_h \widehat{\boldsymbol{\Sigma}}_b) & \text{prior on the state vector} \\ \\ \widehat{\mathbf{U}}_{t+h} & \sim N(\mathbf{A}\mathbf{B}_{t+h}, k_h \widehat{\boldsymbol{\Sigma}}_u) & \text{observation equation} \end{cases}$$

Reconciled mean and variance (Corani et al., 2020)

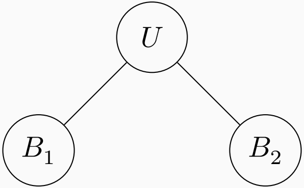
$$\mathbf{G} = \left(\widehat{\mathbf{\Sigma}}_b \mathbf{A}^T\right)\left(\mathbf{A}\widehat{\mathbf{\Sigma}}_b \mathbf{A}^T + \widehat{\mathbf{\Sigma}}_u + \widehat{\mathbf{\Sigma}}_b \mathbf{A}^T\right)^{-1}$$

$$\tilde{\mathbf{b}}_{t+h} = \hat{\mathbf{b}}_{t+h} + \mathbf{G}(\underbrace{\hat{\mathbf{u}}_{t+h} - \mathbf{A}\hat{\mathbf{b}}_{t+h}}_{\text{incoherence}})$$

$$\widetilde{\mathbf{\Sigma}}_b = \left(\widehat{\mathbf{\Sigma}}_b - \mathbf{G}(\mathbf{A}\widehat{\mathbf{\Sigma}}_b)\right) k_h$$

- Equivalent to minT with a block-diagonal covariance matrix.
- We obtain the same reconciled mean and variance of minT by adding the covariance Cov$(\mathbf{B}_{t+1}, \widehat{\mathbf{U}}_{t+h})$, which adds some terms in the above equations.

Reconciled upper mean in 1-level hierarchies



- Assuming no correlation between bottom and upper forecasting errors, $\tilde{u}$ is the combination of $\hat{u}$ and $(\hat{b}_1 + \hat{b}_2)$ (Corani et al., 2020; Hollyman et al., 2022):

$$\tilde{u} = \underbrace{\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \sigma_2^2 + \sigma_u^2}}_{w} \hat{\mathbf{u}} + \underbrace{\frac{\sigma_u^2}{\sigma_1^2 + \sigma_2^2 + \sigma_u^2}}_{1-w} \mathbf{A}\hat{\mathbf{b}},$$

where w depends on the variances of the base forecasts.

Reconciled variance

$$\widetilde{\mathbf{\Sigma}}_b = k_h \left(\widehat{\mathbf{\Sigma}}_b - \mathbf{G}(\mathbf{A}\widehat{\mathbf{\Sigma}}_b) \right)$$

- The variance of each variable decreases after reconciliation: see Zambon et al., (2023); Wickramasuriya et. al. (2022).
- k_h is unknown!

The k_h problem

- Heuristic choices for k_h (Corani et al, ECML20):
  - $k_h = 1$
  - $k_h = h$

Open problem

How to better estimate $\widehat{\mathbf{\Sigma}}_b, \widehat{\mathbf{\Sigma}}_u$ for h-steps ahead forecasts?

Reconciling count variables

Count variables

- A count variable takes non-negative integer values $\{0, 1, 2, 3, ...\}$.

- We have "vectors" of counts $\mathbf{y} \in \mathbb{N}^n$, $\mathbf{b} \in \mathbb{N}^{n_b}$ and $\mathbf{u} \in \mathbb{N}^{n_u}$:
  - n : number of total time series
  - n_b : number of bottom time series
  - n_u : number of upper time series

- The set of coherent vectors in $\mathbb{N}^n$ is:

$$\mathbf{y} \in \mathbb{N}^n : \exists \mathbf{b} \in \mathbb{N}^{n_b} \text{ such that } \mathbf{y} = \mathbf{S}\mathbf{b}.$$

Notation for probability mass functions

- $\hat{p}$: pmf of base forecasts
- p_{BU}: pmf of the probabilistic bottom-up
- $\tilde{p}$: reconciled pmf
- In the following for simplicity we drop the time subscript.

Independence assumptions

Dealing with counts we assume independence of the base forecasts:

- $\widehat{p}(\mathbf{b}) = \prod_i^{n_b} \widehat{p}_i(b_i)$
- $\widehat{p}(\mathbf{y}) = \widehat{p}(\mathbf{b})\widehat{p}(\mathbf{u}) = \prod_i^{n_b} \widehat{p}_i(b_i) \prod_j^{n_u} \widehat{p}_j(u_j)$

Open problem

Model correlations between base forecasts of count variables.

How to obtain a predictive joint from predictive marginals?

Probabilistic bottom-up

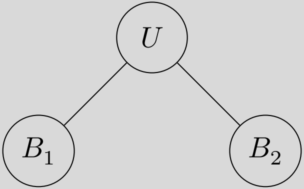
- The pmf $p_{BU}(\mathbf{y})$ uses $\hat{p}(\mathbf{b})$ as the only source of information.
- Consider the bottom vector $\mathbf{b}^*$, with $\mathbf{A}\mathbf{b}^* = u^*$

$$p_{BU}(\mathbf{y}) = p(\mathbf{u}, \mathbf{b}^*) = \begin{cases} \hat{p}(\mathbf{b}) & \text{if } \mathbf{u} = \mathbf{u}^* \\ 0 & \text{otherwise} \end{cases}.$$

- See also Olivares et al., IJF 2023.

Probability of coherence of the base forecasts

Example



- B_1 and B_2 take values in {0, 1}
- U takes values in {0, 1, 2}.

Probability of coherence, p_c

Assuming independence:

$$p_c = \hat{p}_{B1}(B_1 = 0) \cdot \hat{p}_{B2}(B_2 = 0) \cdot \hat{p}_U(U = 0)$$

$$+ \hat{p}_{B1}(B_1 = 0) \cdot \hat{p}_{B2}(B_2 = 1) \cdot \hat{p}_U(U = 1)$$

$$+ \hat{p}_{B1}(B_1 = 1) \cdot \hat{p}_{B2}(B_2 = 0) \cdot \hat{p}_U(U = 1)$$

$$+ \hat{p}_{B1}(B_1 = 1) \cdot \hat{p}_{B2}(B_2 = 1) \cdot \hat{p}(U = 2)$$

In general:

$$p_c = \sum_{\mathbf{b},\mathbf{u}} \hat{p}(\mathbf{b}) \; \mathbb{1}_{\mathbf{Ab}=\mathbf{u}} \; \hat{p}(\mathbf{u})$$

Conditioning on the base forecasts of the upper time series (Corani et al., 2023).

- We use virtual evidence (Pearl, 2009) as a counterpart of the linear-Gaussian model for discrete variables .

Conditioning on the base forecasts of the upper time series (Corani et al., 2023).

- Given a 1-level hierarchy, the vector $\mathbf{b}^*$ and $u^* = \mathbf{A}\mathbf{b}^*$:

$$\tilde{p}(\mathbf{b}^*) \propto \sum_u p_{BU}(u, \mathbf{b}^*) \;\; \hat{p}(u)$$

$$= \;\; \hat{p}(\mathbf{b}^*) \;\; \mathbb{1}_{\mathbf{A}\mathbf{b}=u} \;\; \hat{p}(u)$$

$$= \;\; \hat{p}(\mathbf{b}^*) \;\; \hat{p}(u^*)$$

- The normalizing constant is p_c. In general the updating has to be computed via sampling.

Effect of reconciliation on the relative probabilities

Consider two bottom vectors $\mathbf{b}^*$ and $\mathbf{b}^{**}$:

$$\frac{\tilde{p}(\mathbf{b}^{**})}{\tilde{p}(\mathbf{b}^*)} = \frac{\hat{p}(\mathbf{b}^{**})}{\hat{p}(\mathbf{b}^*)} \cdot \frac{\hat{p}(u^{**})}{\hat{p}(u^*)}$$

- $\dfrac{\tilde{p}(\mathbf{b}^{**})}{\tilde{p}(\mathbf{b}^*)}$ does not change if $\hat{p}(u)$ is uniform.
- $\dfrac{\tilde{p}(\mathbf{b}^{**})}{\tilde{p}(\mathbf{b}^*)}$ increases if $\hat{p}(u^{**}) > \hat{p}(u^*)$.
- If $\hat{p}(u^*) = 1$, reconciliation assigns positive probability only to the bottom vectors which sum up to u^*.

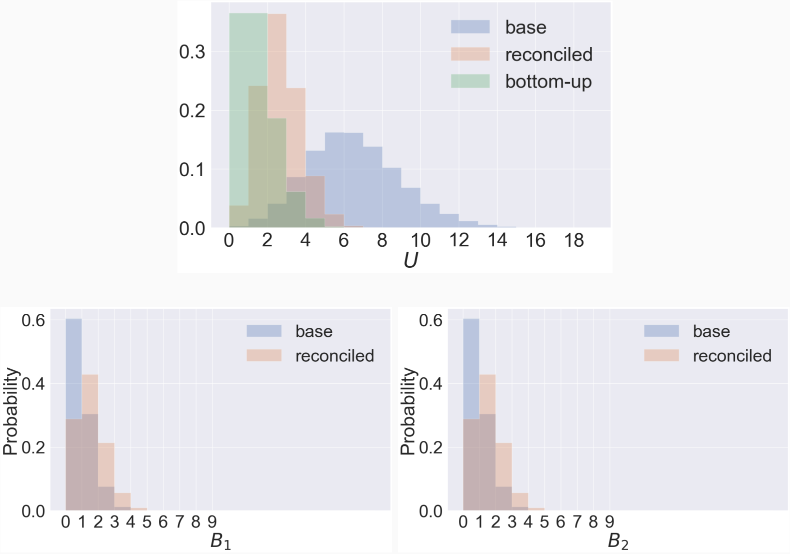
Multiple updates with virtual evidence

- An update is performed for each upper time series of the hierarchy.
  - The posterior (reconciled distribution) of the previous step becomes the prior (base forecast) of the next update.
  - The updates are commutative.
- We assume the conditional independence of the base forecast of the upper time series.

Properties of the reconciled distribution (Zambon et al, 2023)

- While in the Gaussian case reconciliation decreases the variance of every variable…
- .. in the count case reconciliation can decrease or *increase* the variance of the bottom variables.
- The increase of variance is more likely with a small p_c, i.e., when the base forecasts are largely incoherent.

Poisson example ($\lambda_{B1} = \lambda_{B2} = 0.5$, $\lambda_U = 6$)



Reconciled Mean of discrete variables

- The reconciled mean of the upper time series can be lower than both the bottom-up and the base mean (*concordant-shift*).
- This happens for instance if most base forecasts are skewed towards low counts.
- The reconciled means of all the time series are shifted towards zero.
- See Zambon et al. (2023) for a case study about the reconciliation of count time series related to rare events on financial markets.

Temporal reconciliation on *carparts* (Corani et al., 2023)

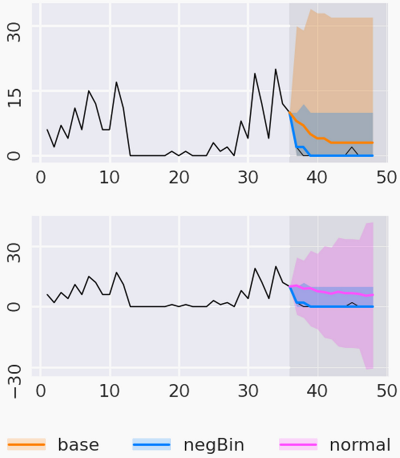
- Temporal hierarchy, whose bottom level is constituted by monthly time series.
- Base forecast in form of samples, generated by tscount.
- We fit a negative binomial on the samples of each base forecasts and then we reconcile using virtual evidence.
- As a term of comparison, we fit a normal on the samples of each base forecasts and then we reconcile using Gaussian reconciliation.

Temporal reconciliation on *carparts* (Corani et al., 2023)

The negbin reconciled pmf is asymmetric, with median 0.

It revises the mean downwards (*concordant shift*) and reduces the variance.

Gaussian reconciliation is a poor model for low-counts time series.



Sampling the reconciled distribution

- Outside the Gaussian case, we need sampling the reconciled distribution.
- Corani et al., (2023) use probabilistic programming.
  - reconciles any parametric base forecast with minimal coding.
  - can become time consuming
- Zambon et al (2022): bottom-up importance sampling (BUIS)
  - immediate
  - reconciles also base forecasts constituted by samples

The BayesRecon package

- Published on CRAN in July 2023.
- Analytical reconciliation of Gaussian base forecasts.
- BUIS reconciliation of non-Gaussian base forecasts.

```r
library(bayesRecon)

#lambdas of Poisson base forecasts for U, B1, B2
base_forecasts <- list(6, 0.5, 0.5)

p_tilde <- reconc_BUIS(S, base_forecasts, in_type="params",
                         distr="poisson")
```

https://CRAN.R-project.org/package=bayesRecon

Open problems

- reconciling hierarchies with mixed types of base forecasts
- estimating Σ_h in Gaussian reconciliation
- representing correlated base forecasts over counts
- cross-temporal probabilistic reconciliation

Papers on reconciliation via conditioning

- Corani, Azzimonti and Rubattu, "Probabilistic Reconciliation of Count Time Series." *International Journal of Forecasting*. 2023
- Zambon, Azzimonti and Corani, Efficient probabilistic reconciliation of forecasts for real-valued and count time series. *Arxiv: 2210.02286v2* (2022)
- Zambon, Agosto, Giudici and Corani, Properties of the reconciled distributions for Gaussian and count forecasts. *ArXiv:2303.15135* (2023)
- Corani, Azzimonti et al., "Probabilistic reconciliation of hierarchical forecast via Bayes' rule.", *Proc. ECML 2020*