Quantile-Filtered Bayesian Learning for the Correlation Class

Hermann Held
Potsdam Institut for Climate Impact Research (PIK), 14412 Potsdam, Germany
hermann.held@pik-potsdam.de

Abstract

We introduce a new rule for Bayesian updating of classes of precise priors. The rule combines Walley's generalized Bayes rule with a filter based on prior quantiles of the observational evidence. We introduce this new "quantile-filtered Bayesian update rule" because in many situations, Walley's generalized Bayes rule reveals counter-intuitively non-informative, dilation-type results while an alternative rule, the maximum likelihood update rule after Gilboa and Schmeidler, is not robust against imprecise priors that are contaminated with spurious information. Our new quantile-based update rule addresses the former issue and fully resolves the latter. By the new rule we update an imprecise prior that was recently motivated by expert interviews with climate, ecosystem and economic modelers: a "correlation class" of precise priors with arbitrary correlation structure, however, prescribed precise marginals. For an insurance situation we demonstrate that under our new rule a set of clients would be insured that is disregarded under standard generalized Bayesian updating.

Keywords. Bayesian updating, Generalized Bayes rule, imprecise probability, robust Bayesian approach, modeling expert opinions, prescribed marginals, unknown correlation structure.

1 Introduction

Complex numerical models provide key working horses within climate, ecosystem and economic research and hence their output strongly influences the discussion on ecologically and economically sustainable climate policies. In turn, model output strongly depends on various tuning parameters which cannot fully be determined through objective data in general. For that reason Bayesian methods become increasingly popular in these fields as they would allow to incorporate subjective prior knowledge on model parameters, often aggregated from scattered sources of information in the brains of modelers, in a statistical analysis. A recent semi-formalized expert elicitation aimed at generic patterns of knowledge vs. ignorance in modelers' prior information on multivariate model parameters [8, 9]. As a key result modelers across disciplines stated to hold fundamentally more precise information on marginals than on the (higher order) correlation structure among parameters.

This key finding from above elicitation fueled our interest in Tchen's imprecise model [14] (further investigated in [2, 5, 6, 11, 12]) consisting of a class $\mathcal{P}$ of precise measures P the marginals of which would all equal certain prescribed marginals. We call this class "correlation class". When updating a correlation class along the lines of global Bayesian robustness [1], i.e. element-wise updating according to standard Bayes rule, then observing the extremes of ensuing answers as the prior varies over the class, we found non-informative imprecise posteriors over a wide range of potential observations y [8, 9].

This is in line with prior results on a similar $\mathcal{P}$ by [11] (see also Seidenfeld and Wasserman [13] for a discussion of such a dilation phenomenon where posterior bounds are dilated even for all possible measurements y). In case the set $\mathcal{P}$ is convex this updating procedure is equivalent to Walley's generalized Bayes rule [15]. We will call this element-wise updating and subsequent extremizing "GBR" throughout this article regardless of whether $\mathcal{P}$ is convex or not. (An alternative class displaying imprecise correlations is introduced in [10] characterized by a radially symmetric possibility measure. However as no results on Bayesian updating have been published for that class so far, we disregard it in the context of this article.)

Gilboa's and Schmeidler's maximum likelihood update rule [7] delivers much more informative results. Their rule is equivalent to applying GBR – not to $\mathcal{P}$ but – to the subset of those precise priors that would maximize the prior expectation of the evidence y. In [8, 9] that rule is generalized by not completely disregarding those priors that would not maximize prior expectation of y but by giving any element of $\mathcal{P}$ an influence, weighted by its prior expectation of y. (y may either represent a single sample or a number of samples that can be combined to the multi-variate observation y.) However as against GBR, both likelihood update rules face the problem that spurious information may enter the final result: in case $\mathcal{P}$ contains an unjustified element that accidentally displays high prior expectation of y, this may result in a posterior that is more precise than for the uncontaminated version of $\mathcal{P}$.

For that reason here we present a new updating rule that combines important advantages of GBR and the latter two likelihood updating rules: (i) it is more informative than GBR and (ii) in case $\mathcal{P}$ is contaminated this contamination would not add spurious information to the posterior.

We are aware that there exists the further method of reducing the class of priors in view of evidence as described in [3, 4]. However the relation to our work appears intricate and its elucidation shall be outlined elsewhere.

This article is organized as follows. In Section 2 we introduce the new updating rule. In Section 3 we apply that rule to the briefly recapitulated imprecise prior in [8, 9] motivated by above expert elicitation. In Section 4 we regularize our prior by bounding the gradients of densities making up the imprecise prior. In Section 5 we offer an interpretation of our new updating rule that involves also concepts from classical statistics and therefore might be controversial. In Section 6 we compare the results of various updating methods from the point of view of an idealized insurance company. Finally, in Section 7 we summarize our findings and outline the most pressing issues from the point of view of a modeler.

2 The Quantile-Filtered Bayesian Update Rule

The crucial element of our new updating rule is the filter that acts on $\mathcal{P}$, before GBR is applied.

Definition 1 *Let $\mathcal{P}$ represent an imprecise prior made up by a non-empty set of precise priors. Let $Q \in]0,1[$. Let P_L denote the probability measure induced by a precise prior $P' \in \mathcal{P}$ and the precise likelihood L on the space of all potential observations Y. Then $\mathcal{Y}_{\mathcal{P}LYQ}$ is a* **generator of a Q-filtered Bayesian update rule (QFB)** *iff $\mathcal{Y}_{\mathcal{P}LYQ} : \mathcal{P} \to 2^Y$ with $\forall_{y \in Y} \forall_{P' \in \mathcal{P}} \;\; P_L(y \in \mathcal{Y}_{\mathcal{P}LYQ}(P')) \geq Q$.*

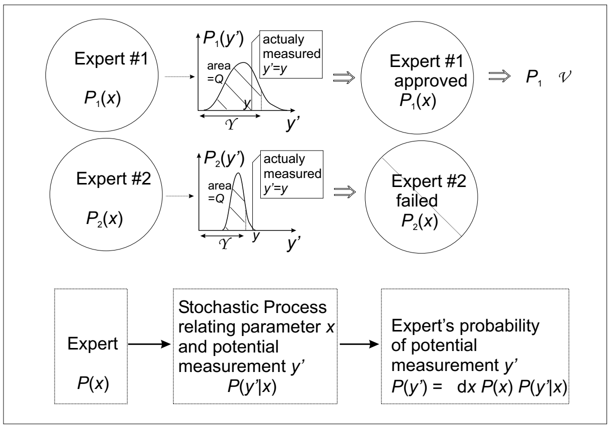

Figure 1: Scheme for the construction of the subset $\mathcal{V}$ in the class of priors. Any prior (here identified with a different "expert") induces – through a given likelihood – a probability measure on the space of potential measurements y' (bottom). Once the measurement has been realized, i.e. $y' := y$, one can disregard priors that display y outside of a quantile, characterized by a pre-set probability Q.

Hence $\mathcal{Y}$ maps $P' \in \mathcal{P}$ onto a prior ($\geq Q$)-quantile in observation space. As an illustrative example, in Figure 1, the two elements of $\mathcal{P}$, P_1, P_2, are mapped onto an interval $]-\infty, y'_{\max}]$ within the respective abscissa (the latter denoting the space of potential observations y').

Definition 2 *Let $\mathcal{P}, Q, L, P_L, Y$ as above and $\mathcal{Y}_{\mathcal{P}LYQ}$ the accompanying generator of a Q-filtered Bayesian update rule. Then $\mathcal{V}_{\mathcal{P}LYQ}$ is a* **Q-GBR-filter** *iff $\mathcal{V}_{\mathcal{P}LYQ} : Y \to 2^{\mathcal{P}}, \quad y \mapsto \{P' \in \mathcal{P} \mid y \in \mathcal{Y}_{\mathcal{P}LYQ}(P')\}$.*

Hence for given observation y, $\mathcal{V}(y)$ represents those priors for which y is not too "far-fetched" (see Figure 1).

Definition 3 *Let $\mathcal{V}$ be according to previous Defs. Then we call the operation GBR $\circ\, \mathcal{V}$ a* **quantile filtered Bayesian learning rule** *(QFB).*

Before we discuss a desirable property of QFB w.r.t. contaminations, we would like to recall that GBR shares this property:

Theorem 1 *Let $\overline{\mathcal{U}}_{\text{GBR}} : Y \otimes \mathcal{P} \to \mathbb{R}$ the "updating operator" maximizing the ensuing answer of Bayesian learning over the class of priors along GBR, and $\underline{\mathcal{U}}_{\text{GBR}}$ the analogue minimization operator. Then $\forall_{y \in Y} \;\; \underline{\mathcal{U}}_{\text{GBR}}(y, \mathcal{P} \cup \mathcal{P}_c) \leq \underline{\mathcal{U}}_{\text{GBR}}(y, \mathcal{P}) \leq \overline{\mathcal{U}}_{\text{GBR}}(y, \mathcal{P}) \leq \overline{\mathcal{U}}_{\text{GBR}}(y, \mathcal{P} \cup \mathcal{P}_c)$.*

This relation simply follows from the fact that the sup(inf)-operator is monotonous w.r.t. set-extension.

It implies that a contamination $\mathcal{P}_c$ would not add spurious information to the posterior result. In general, such a relation is violated by the two likelihood updating rules mentioned before, but importantly it holds for QFB:

Theorem 2 *Let* $\overline{\mathcal{U}}_{\mathrm{QFB}} : Y \otimes \mathcal{P} \rightarrow \mathbb{R}$ *the "updating operator" maximizing the ensuing answer of Bayesian learning over the class of priors along QFB, and* $\underline{\mathcal{U}}_{\mathrm{QFB}}$ *the analogue minimization operator. Then* $\forall_{y \in Y}\ \ \underline{\mathcal{U}}_{\mathrm{QFB}}(y, \mathcal{P} \cup \mathcal{P}_c) \leq \underline{\mathcal{U}}_{\mathrm{QFB}}(y, \mathcal{P}) \leq \overline{\mathcal{U}}_{\mathrm{QFB}}(y, \mathcal{P}) \leq \overline{\mathcal{U}}_{\mathrm{QFB}}(y, \mathcal{P} \cup \mathcal{P}_c)$.

This Theorem readily follows from the fact that the way the operator GBR $\circ\ \mathcal{V}$ acts on $P' \in \mathcal{P}$ does *not* depend on the other elements of $\mathcal{P}$. This is in contrast to the other two likelihood update rules for which the relative weight (the prior expectation of y) of P', compared to the other priors matters. We regard the fact that those Theorems hold as a key advantage of QFB and GBR. It now remains to show that QFB is significantly more informative than GBR in relevant cases.

3 Specification and updating of the correlation class

3.1 The imprecise prior and the likelihood

In order to keep the discussion as transparent as possible we decide on the simplest non-trivial $\mathcal{P}$ and likelihood possible. We consider the uncertain parameter $(x_1, x_2)^t \in \mathbb{R}^2$, the "observation" or "evidence" $y \in \mathbb{R}$. Furthermore for any element of $\mathcal{P}$, any of its two marginals should equal $N(\mu, \sigma^2)$, a Gaussian with mean μ and σ^2 variance[1]. A likelihood employed shall write $L(x_1, x_2) \equiv P(y|x_1, x_2) := N(\kappa x_1 + x_2, \sigma_\eta^2)(y)$, κ known to the modeler. From now on whenever results are not displayed in analytic form, we choose the specific parameter values $\mu = 1/2, \sigma = 1/4$, κ:=1.05 (as κ=1 would lead to a degenerate and $|\kappa| \gg 1$ to a trivial case [8, 9]), $\sigma_\eta := \sigma/10$.

So far we have specified only marginals, hence we do not rule out multi-modal densities. However we find the subset of unimodal prior densities more convincing a model for generic prior expert knowledge. This is conveniently implemented by requiring that any prior shall be a 2D Gaussian, although admittedly hereby we potentially disregard too many priors. For pragmatic reasons, however, we stick to this computationally convenient case for the remainder of the article. It is shown in [8, 9] that then

[1]For a multivariate application, the first entry would represent a vector of means, the second the symmetric covariance matrix.

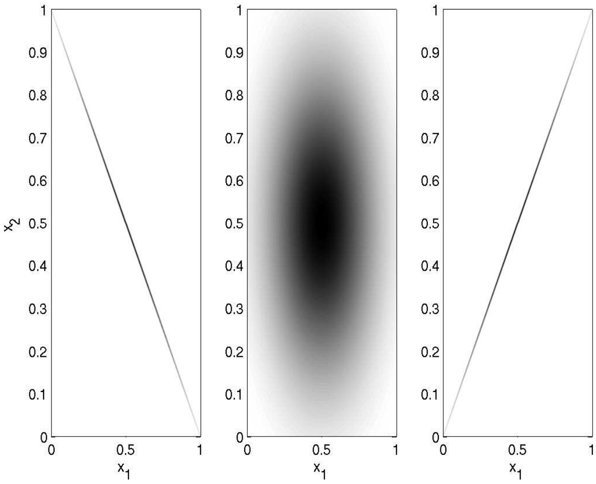

Figure 2: Three extreme representatives of the class of Gaussian priors with prescribed marginals. From left to right: maximally anticorrelated case ($f = -1$), uncorrelated case ($f = 0$), and maximally correlated case ($f = 1$) – for a definition of the parameter f see Eq. 1).

$$\mathcal{P} = \{P \mid \exists_{f \in [-1,1]}\ P\ \sim N((\mu, \mu)^t, \Sigma(f))\} \text{ with } \quad (1)$$
$$\forall_{f \in [-1,1]} \quad \Sigma(f) := \sigma^2 \begin{pmatrix} 1 & f \\ f & 1 \end{pmatrix}.$$

$f = 0$ represents standard Bayesian updating with an uncorrelated prior, $f = 1$ ($f = -1$) the fully (anti)correlated prior. Accompanying densities are displayed in Figure 2.

Finally we select the functional we are interested in – the *probability of ruin*:

Definition 4 *Let* $P \in \mathcal{P}$. *Let* $x_1^* \in \mathbb{R}$. *Then we define the probability of ruin as*

$$P^* := \int_{x_1^*}^{\infty} dx_1 \int_{-\infty}^{+\infty} dx_2\ P(x_1, x_2).$$

In the context of climate modeling, κx_1^* could represent a well-known critical value of global mean temperature beyond which "catastrophic" global warming impacts may occur, and x_1, x_2 two uncertain climate model parameters.

3.2 Bayesian learning

In order to generate the posterior probability of ruin per precise prior, the posterior marginal for x_1 is key. In [8, 9] it is shown that

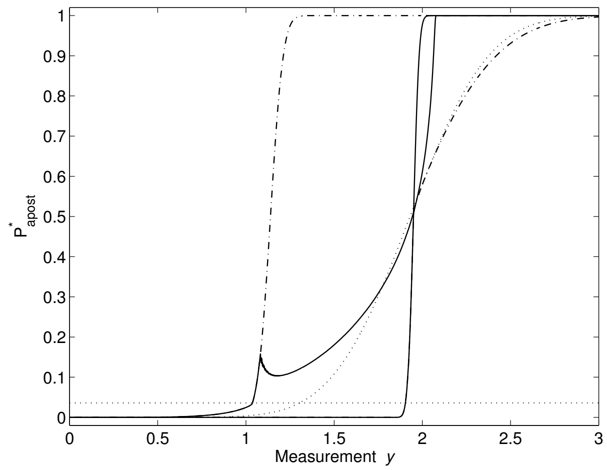

Figure 3: Probability of ruin (upper and lower) for the correlation class parameterized by the correlation coefficient $f \in [-1, 1]$ after Eq. 1 for $\kappa = 1.05, x_1^* = 0.95$, $\sigma_\eta = \sigma/10$. Horizontal dotted line: apriori value, curved dotted: (standard) uncorrelated case, dashed-dotted: GBR, solid: QFB for $Q = 98\%$, the lower probabilities of ruin for GBR and QFB coalescing. GBR reveals quasi non-informative posterior results for $y \in [1.3, 1.8]$. Quite the contrary the new QFB is informative for any $y \in \mathbb{R}$. For an expanded representation of the "avoided crossing" region around (2,1/2), see the following Figure.

$$\begin{aligned} P_{\rm post}(x_1|y) &\sim N(\mu'(f,y), \sigma'^2(f,y)) \quad \text{with} \qquad (2)\\ \mu'(f,y) &= (\mu(1-(1-f)(\kappa-1)\ \sigma^2/\sigma_\eta^2) \\ &\quad +(f+\kappa)\ y\ \sigma^2/\sigma_\eta^2) \\ &\quad /(1+(1+2f\kappa+\kappa^2)\ \sigma^2/\sigma_\eta^2), \\ \sigma'(f) &= \sigma \sqrt{\frac{1+(1-f^2)\ \sigma^2/\sigma_\eta^2}{1+(1+2f\kappa+\kappa^2)\ \sigma^2/\sigma_\eta^2}}. \end{aligned}$$

We utilize this expression to calculate the posterior probability of ruin

$$P^*_{\rm apost}(f,y) = \int_{x_1^*}^{\infty} N\left(\mu'(f,y), (\sigma'(f))^2\right)(x_1)\ dx_1. \quad (3)$$

From this we obtain the upper probability of ruin in the case of GBR by

$$\overline{P}^*_{\rm apost.GBR}(y) = \sup_{f\in[-1,1]} P^*_{\rm apost}(f,y). \quad (4)$$

For QFB we need to define generator of a Q-filtered Bayesian update rule $\mathcal{Y}$. As larger y will imply higher

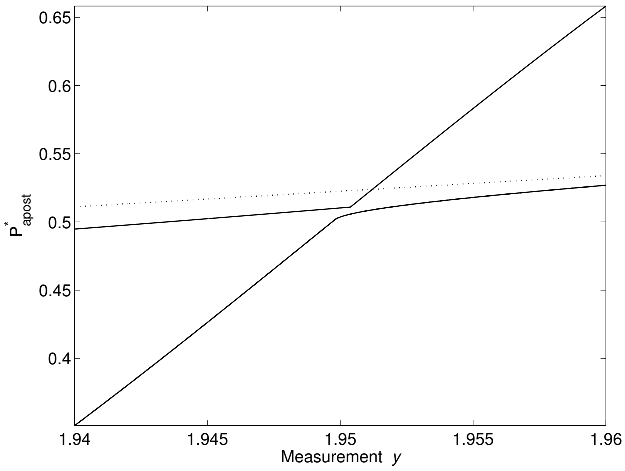

Figure 4: Expansion of the previous Figure's center that shows an "avoided crossing" of the solid lines. We explain this feature of almost precise posterior probability in Subsection 3.3 by an approximate symmetry in the transfer function $(x_1, x_2) \to y$ in combination with Gaussian symmetry.

probabilities of ruin in general, we expect that the following prescription will lead to informative posteriors $(0 < Q < 1)$:

$$\begin{aligned} \forall_{f\in[-1,1[}\quad \mathcal{Y}(f) &:= \ ]-\infty, y_{\max}(f)] \quad \text{with} \qquad (5)\\ Q &=: \int_{-\infty}^{y_{\max}(f)} dy\ P_{y;f.\rm prior}(y) \quad \text{and} \\ \mathcal{Y}(f=1) &:= \ ]-\infty,\infty[\qquad (6) \end{aligned}$$

Let $\mathcal{V}$ be the filter generated by $\mathcal{Y}$. Then latter equation ensures that for all y, $\mathcal{V}(y) \neq \emptyset$ (for a more extended discussion the reader is put off to the more "philosophical" Subsection 5.3 – here we just point to Theorem 1 which ensures that no spurious information is added when making a class of priors, subject to GBR, larger). In order to operationalize Eq. 5 we need $P_{y;f.\rm prior}$. In [8, 9] we show

$$P_{y;f.\rm prior} \sim N(\mu(1+\kappa), \sigma^2(1+2\kappa f+\kappa^2)+\sigma_\eta^2). \quad (7)$$

Then

$$\overline{P}^*_{\rm apost.QFB}(y) = \sup_{f\in f(\mathcal{V}(y))} P^*_{\rm apost}(f,y) \quad (8)$$

if $f(\mathcal{V}(y))$ denotes the set of f-values needed to parameterize $\mathcal{V}(y)$. (For $\underline{P}_{\rm apost}$, "sup" is to be replaced by "inf" in above equations.)

We do not claim that our choice of $\mathcal{Y}$ generates the most informative QFB. Here we just would like to demonstrate that even a rather unsophisticated choice leads to much more informative results than GBR does.

The dependency of the probability of ruin on y is depicted in Figure 3 for GBR (dashed-dotted curves for upper and lower probability of ruin), the new QFB (solid curves) under a choice of $Q = 98\%$, for comparison also the assumption of independent parameters (uncorrelated case $f = 0$).

We observe that in general QFB is much more informative than GBR – i.e. the difference of upper and lower probability of ruin is smaller for QFB than for GBR. A bizarre feature can be observed for QFB however: the upper probability of ruin is not a monotonous function of y, a feature occurring in an even more pronounced way for the maximum likelihood update rule [8, 9]. There we attribute this to a certain degenerate feature within $\mathcal{P}$, related to $f = -1$ and becoming virulent at $y = (\kappa - 1)x_1^* + 2\mu \approx 1.05$. We propose that such effects would vanish if a non-parametric class of priors were considered. A skeptic of our new method may now argue that also the superiority of QFB over GBR as displayed in Figure 3 may be a result of degenerate priors and would vanish under more regular imprecise priors. In the following Section we show that this is not the case, but QFB is robustly more informative even when we "regularize" $\mathcal{P}$. Before that, however, we would like to interpret the striking convergence of QFB-upper and lower probability of ruin to the value 1/2 as displayed in Figure 3.

3.3 An "avoided crossing" for QFB

The fact that we use a Gaussian class of priors leads to a series of peculiar phenomena of which the "avoided crossing" of upper vs. lower solid curve at $\sim (2, 1/2)$ in Figure 3 may be of special interest. For readers that would like to focus more on the general statements of this article we suggest that they skip this Subsection and proceed directly with Section 4.

The key reason for the almost precise QFB posterior at $y \approx 2$ is easiest accessed in considering the following double limit of Eq. 3 on the posterior mean

$$\forall_{f \in]-1,1]} \; \forall_{y \in \mathbb{R}} \quad \lim_{\kappa \to 1} \lim_{\sigma_\eta \to 0} \mu'(f, y; \kappa, \sigma_\eta) = \frac{y}{2}, \tag{9}$$

i.e. for the whole class, the posteriors will be centered at $y/2$ (with differing variances).

This implies that

$$\left\{ \frac{y}{2} = x_1^* \right\} \Rightarrow \left\{ \forall_{f \in]-1,1]} \; \lim_{\kappa \to 1} \lim_{\sigma_\eta \to 0} P^*_{\text{apost}}(f, y) = \frac{1}{2} \right\}. \tag{10}$$

As $f = -1$ is not element of the volume of confidence at $y/2 = x_1^*$, from this Eq. we conclude a precise posterior at that $y \approx 2$.

We now investigate how this exact prosterior dilutes into an avoided crossing for $\kappa = 1.05, \sigma_\eta = \sigma/10 = 1/40$. For this, it is important to note that Eq. 3 can be rewritten as

$$P^*_{\text{apost}}(f, y) = \int_{-\infty}^{y} dy' \; N\left(\frac{x_1^* - \mu_0(f)}{\mu_1(f)}, \frac{\sigma'(f)}{\mu_1(f)} \right)(y'), \tag{11}$$

whereby the two new functions $\mu_0(f) + y\mu_1(f) := \mu'(f, y)$ are determined by the (in y) linear relation Eq. 3. From Eq. 11 we learn that for any f, $P^*_{\text{apost}}(f, y)$ is an error function in y. Now we deduce the analytic form of the lower solid line before the crossing. After verifying $\partial\mu'/\partial f < 0$ (for $y > (1 + \kappa)\mu$) and $d\sigma'/df < 0$, we conclude that for given y, $P^*_{\text{apost}}(f, y)$ decreases with f. Hence the QFB lower bound is generated by the single posterior $P^*_{\text{apost}}(f = 1, y)$ for $y < y_c$. We define y_c as the "crossing value" $P^*_{\text{apost}}(f = 1, y_c) := 1/2 \Rightarrow y_c \approx 1.9497$, also compare to Figure 4.

While the lower QFB bound before the crossing is made up by a single f (i.e. a single prior) in terms of one single error function, the QFB upper bound is the envelope of error functions generated according to Eq. 11 from different f's. This is related to the fact that the upper bound per y is generated from the lower bound f_- of the interval of confidence $[f_-(y), 1]$ and $df_-/dy > 0$. However, locally in y, the upper bound can be related to one single f. We find numerically $f_-(y_c) \approx 1/2$ (in accordance with Figure 4, QFB excludes the uncorrelated case (dotted line $\Leftrightarrow f = 0 \notin [f_-(y_c), 1]$)). We can now address the following question: what parameters determine the width of the avoided crossing

$$\Delta P^*_{\text{apost.QFB.ac}} := P^*_{\text{apost}}(f_-(y_c), y_c) - \frac{1}{2}. \tag{12}$$

Let $\Delta f := 1 - f_-(y_c)$, i.e. the difference of the QFB upper bound f to the prior's f that generates the QFB lower bound. Utilizing Eq. 11 we then derive in first order perturbation theory

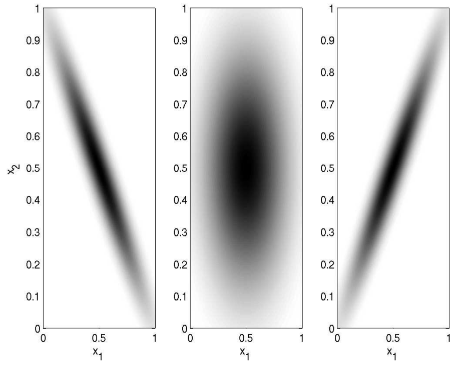

Figure 5: Extreme cases of priors after bounding the gradient. Left: $f = -f^*$, center: $f = 0$, right: $f = f^*$, $f^* \approx 0.95434$. The bound f^* was chosen such that the expert "can resolve not more than 5 items per typical marginal parameter scale" (in our case $[0, 1]$, spanning 4σ – for details see [8, 9]). For that reason, the densities displayed in the left and the right panel are smoother than their counterparts in Figure 2.

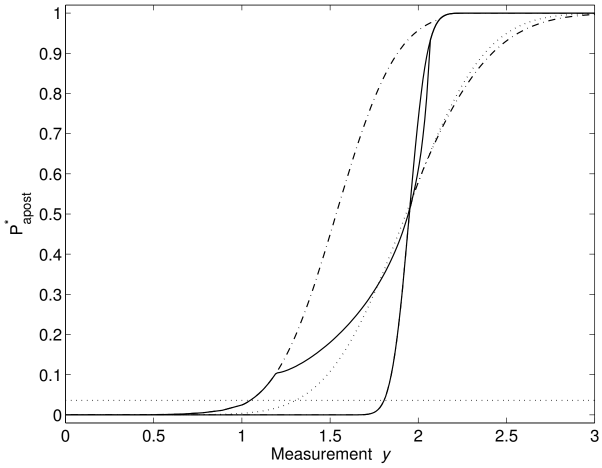

Figure 6: Upper and lower probabilties of ruin as in Figure 3, yet for bounded gradients of prior densities (dashed-dotted lines: GBR, solid lines: QFB, dotted curved line: updated uncorrelated precise prior, horizontal). Note that even for this class of priors "regularized" by gradient bounding (equivalent to $|f| \leq f^* \approx 0.95434$), QFB is more informative than GBR.

$$\lim_{\sigma_\eta \to 0} \Delta P^*_{\text{apost.QFB.ac}} \approx \frac{1}{4\sqrt{\pi}} (\kappa - 1) \frac{x_1^* - \mu}{\sigma} \sqrt{1 - \Delta f}. \tag{13}$$

Inserting the values of our example, we obtain $\Delta P^*_{\text{apost.QFB.ac}} \approx 0.01$, in accordance with the distance within the crossing displayed in Figure 4. The last equation also reveals that the avoided crossing becomes an exact crossing if x_1, x_2 influence y symmetrically, i.e. $\kappa \to 1$, in accordance with Eq. 10.

4 Introducing a gradient filter

Following Walley [16] we regard it as meaningful to bound the gradient of densities within a class of priors. It is very questionable that in general an expert will hold such a sophisticated prior knowledge that bizarre density structures of arbitrary gradient could be distinguished in her or his brain. For our class this would imply to disregard priors with too large $|f|$.

Working with such a "regularized" class of priors comes with the additional advantage that effects like those at $y = (\kappa - 1)x_1^* + 2\mu \approx 1.05$ may vanish as our class becomes more similar to a non-parametric, however, gradient-bounded class which the "imprecise community" may find more adequate for generic expert knowledge in the future.

The question now is how to restrict $|f|$. Following [8, 9] we argue that in general, an expert will not be able to distinguish more than 5 "major blocks" per parameter dimension. This idea is formalized in [8, 9] and leads to the prescription $|f| \leq 0.95434$. Figure 5 then represents the bounded-gradient counterpart of Figure 2. In the long run, this issue must ultimately be addressed by suitable expert elicitations and social experiments that would reveal the expert's "prior resolution."

In fact Figure 6 reveals that even after bounding the density gradient over $\mathcal{P}$ QFB stays qualitatively more informative than GBR. In addition, for this "regularized imprecise prior" now also QFB responds monotonously w.r.t. observation y what is more in line with intuition.

Hence QFB seems to combine both desirable features discussed in the Introduction: it is informative and it does not absorb spurious information (Theorem 2). For that reason we regard it as worthwhile to look for an interpretation of QFB. (The reader may start using QFB for pragmatic reasons even if she or he does not want to follow the assumptions in the interpretation given below.)

5 Interpretation and nesting of quantile-filtered Bayesian learning

5.1 Interpretation of QFB

We present one possible interpretation QFB that is based on the following two assumption:

(1) Any prior class of precise measures specified by an

expert contains "the adequate, yet un-identified" precise measure for that actual assessment;
(2) when considering the sequence of the expert's assessments over her or his life and transforming each "adequate precise prior" to a uniform prior by a suitable coordinate transformation, then the sequence of accordingly transformed "true states of the world" (the sequence of true parameter values) would behave as drawn from a uniform distribution.

Assumption 1 reminds of a situation in which a king needs to listen to a series of agents, knowing that only one agent has really been sent by the king's friend whereas the others are from "false friends".

Assumption 2 shall be illustrated by a special case first: suppose an expert performed a series of assessments $a_1, ..., a_n, ..., a_N$, whereby at each assessment a_n she or he would be asked for the probability of whether a certain "true" state of the world s_n belonged to a certain set S_n. We denote this probability as $P(s_n \in S_n)$ and we assume further that for any n, the expert would claim $P(s_n \in S_n) = p$. We now imagine that some time will have passed by and in the corse of history the true nature of $s_1, ..., s_N$ will have become public, i.e. the expert's customers will then be able to objectively determine the index function $\text{ind}(s_n \in S_n)$ – that is 1 in case the statement is true and 0 otherwise. Then Assumption 2 requires that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \text{ind}(s_n \in S_n) \to p \tag{14}$$

in the sense of the law of large numbers. Hence we require that in a frequentistic sense the expert will have been neither over- nor under-confident, i.e. the life-averaged assessment prior p was "adequate."

Therefore in more general terms, Assumption 2 implies that in a world in which an expert would specify prior knowledge always as uniform distribution on [0,1] per assessment, later generations would find the histogram of true states of the world, the expert had assessed, converge to that uniform distribution over the life-span of the expert.

That way, we choose an interpretation of subjective probability that allows us to treat it not only as epistemic uncertainty, but also as aleatoric uncertainty, i.e., as a stochastic process that governs the relation of the expert to reality during her or his life. Those users that could accept such an interpretation of experts' knowledge have the chance to interpret the combination of "choose the parameter" and "predict, given that parameter, the measurement y" as a joint stochastic process. If the former is described by $P(x)$ and the latter by $P(y|x)$, then, given the expert's P:

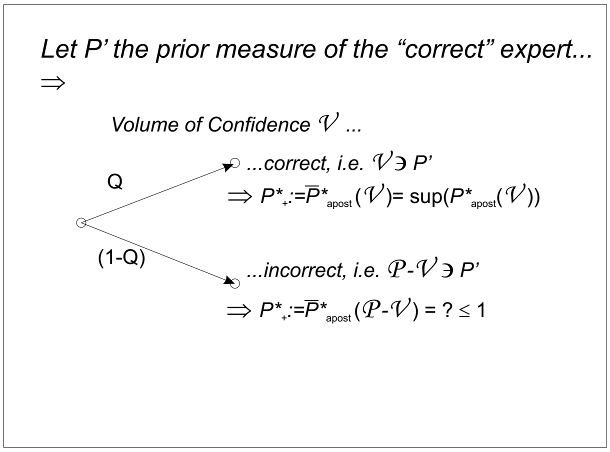

Figure 7: Nesting the classical volume of confidence $\mathcal{V}$ in a decision situation. In our frequentist's interpretation we can explicitly take care of the possibility that $\mathcal{V}$ may *not* contain the adequate prior. For that we utilize a probability tree, resulting in Eqs. 15 and 16.

$P(y) = \int dx\ P(x)\ P(y|x)$.

As in our interpretation, for any prior, $P(y)$ is generated by a stochastic process, it must be possible to evaluate the elements within the set of priors on the basis of the measurement utilizing frequentistic statistics. In particular we are interested in defining a classical volume of confidence within the set of priors as a filter, conditioned on y.

By construction $\mathcal{V}$ of Definition 2 is such a classical volume of confidence with the confidence value Q. Q can then be interpreted as follows: it represents a (lifetime averaged) lower bound for the relative frequency that an expert does include the "adequate" precise prior in $\mathcal{V} \subset \mathcal{P}$ in case for any inference situation such an "adequate prior" exists.

Given this interpretation we can make use of the fact that upper (lower) posterior probabilities of the event we are interested in are bounded functionals over $\mathcal{P}$ and ask whether we can somehow also account for those cases in which $\mathcal{V}$ fails, i.e. does not contain the "adequate prior".

5.2 Proposing a nesting formula

One may now ask how a decision-maker may deal with the fact that the volume of confidence does not hold with certainty but only with probability Q. If $Q \approx 1$, in many applications of classical tests, this aspect is simply ignored and the volume of confidence is dealt with as if it were certain.

However, here we would like to suggest an exact approach that explicitly takes care of those cases for which the volume of confidence fails, appearing with probability $(1-Q)$. We "nest" the classical uncertainty $(1-Q)$ into the Bayesian scheme by a probability-tree argument (see Figure 7).

Let P_+^* and P_-^* the upper and lower probabilities of ruin derived, after the quantile filter has been applied before GBR. In case 1, the classical volume was correct, and $\overline{P}^*_{\text{apost}} = P_+^*$, being true with probability Q. In case 2, the classical volume was wrong, and we set $\overline{P}^*_{\text{apost}} = 1$ as a conservative estimate of that quantity, with probability $(1-Q)$. (Analogously we can proceed with the *lower* probability of ruin.)

According to the thereby induced tree diagram,

$$\underline{P}^*_{\text{apost.QFB.nest}} = Q \cdot P_-^* + (1-Q)\cdot 0, \quad (15)$$
$$\overline{P}^*_{\text{apost.QFB.nest}} = Q \cdot P_+^* + (1-Q)\cdot 1. \quad (16)$$

In the following, we will call the upper and lower probabilities of ruin "nested".

In case one subscribed to the two assumptions given in the beginning of this Section, one could interpret $\underline{P}^*_{\text{apost.QFB.nest}}$ and $\overline{P}^*_{\text{apost.QFB.nest}}$ as upper and lower bounds for relative frequencies of "ruins" over a sequence of equivalent assessments, in the limit of large numbers (of assessments). To the best of our knowledge, this is the first time the incompleteness of interval estimates is addressed.

5.3 Treatment of an empty $\mathcal{V}(y)$

How to proceed if y is such an "outlier" that $\mathcal{V}(y) = \emptyset$? One could proceed in saying that no expert were available, hence there were no information on P^*_{apost}. However, that lack of posterior information is counterintuitive. If the quantile filter is used together with GBR, we know that adding a prior to the class does not result in spurious information. Hence if $\mathcal{V}(y) = \emptyset$ we could add a prior P_a from the original class that is most informative, e.g. the maximum likelihood prior. No spurious information is added by re-introducing P_a due to Theorem 1. This is exactly the argument that was used when setting up Eq. 6.

We would like to illustrate what updating of the imprecise prior with our new rule QFB may mean in a decisions situation. Hence, before presenting the implementation of above combinations of learning rules and filters, we now introduce a stylized potential user of our ideas.

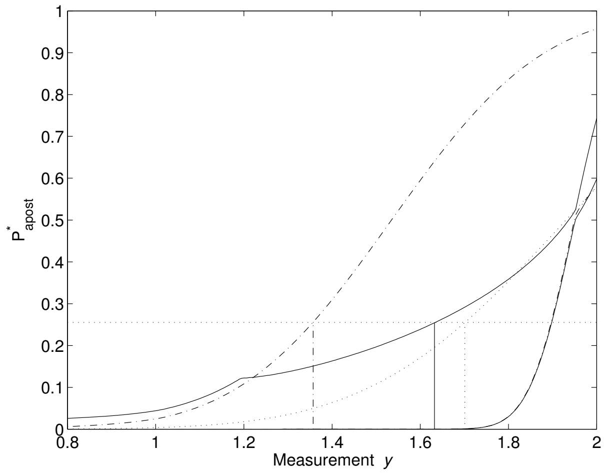

Figure 8: Inverting Figure 6: reading from a prescribed maximum probability of ruin (horizontal line) the accompanying maximum y^*, as needed for the stylized insurance problem. As against Figure 6, here we have involved the nesting correction for QFB, that amounts, however, only to an upwards shift of 0.02. We observe that according to QFB clients with characteristic $y \in [1.34, 1.66]$ could be insured in addition to GBR.

6 Various updating rules for a stylized insurance situation

Following [8, 9] we consider an admittedly rather stylized insurance company that plans to insure a fixed number of clients J each of which comes with a potential standard loss of 1, the behaviorally identical clients' willingness to pay (for a premium) of $2^{-1+1/\alpha}p^{1/\alpha}, \alpha := 3$, for the upper probability of ruin per client p. If the company asks for a residual upper probability for bankrupt, i.e. net loss, of 0.1%, then in a Gaussian approximation we obtain as upper probabilities of ruin allowed per client: 0.12927417 or 0.27004601 for $J = 30$ or $J = 100$ respectively.

With these numbers we enter the ordinate in Figure 6 and read the maximum characteristic y^* per client with which that client would still be insured. Within that Figure the concept of a maximum allowed y makes sense as all curves monotonously increase. In Figure 8 we further illustrate this inversion for the case of 30 clients, i.e. $\overline{P}^*_{\text{apost}} \approx 27\%$. The only difference is that for QFB we show the nesting-corrected results according to Eqs. 15 and 16 (for the upper probability of ruin, this amounts approximately to an addition of $1-Q = 0.02$ that is almost negligible). Interestingly, clients with much higher y could be insured according to QFB than according to GBR.

We summarize threshold values y^* that denote the

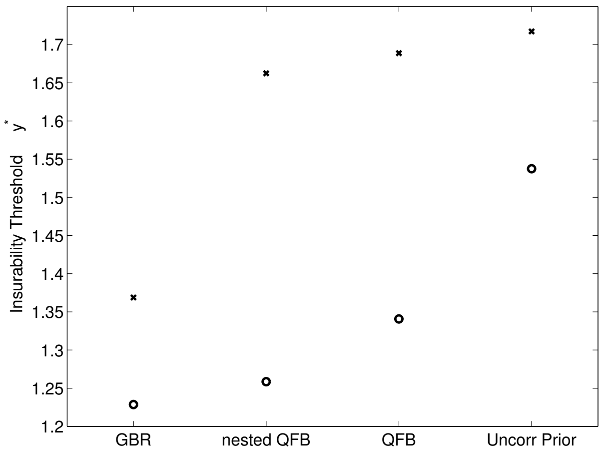

Figure 9: y^* as upper limit of y's with which clients would be insured: Circles: pooling with 30 clients; crosses: pooling with 100 clients. The abscissa indicates the four learning rules according to the tabular of this Section. (Any entry for $\kappa = 1.05, x_1^* = 0.95, \sigma_\eta = \sigma/10, Q = 98\%$.) According to QFB significantly more risky clients could be insured than for GBR.

maximum y with which a client would get insured in the following tabular:

	updating rule $\quad J$	30	100
1	GBR	1.23	1.37
2	QFB after nesting	1.26	1.66
3	QFB before nesting	1.34	1.69
4	uncorrelated prior	1.54	1.72

As expected, the standard Bayesian updating (uncorrelated prior) is found on the optimistic (upper) end of y^*. Otherwise QFB significantly out-competes GBR in that it would allow the insurance company to tap a new class of clients. This tabular is visualized in Figure 9.

7 Summary and Conclusions

This article introduces a new rule for Bayesian updating of imprecise priors that can be represented by classes of precise priors. Our quantile-filted Bayesian learning rule (QFB) disregards those priors that would see the evidence y outside a certain Q-quantile before updating along (a modified version of) generalized Bayes' rule (GBR). The aspect of disregarding priors in view of the evidence before applying GBR is along the idea of Gilboa and Schmeidler to consider only those priors that would maximize prior probability of y. However, in contrast to their rule, QFB has the advantage that it does not add spurious information in case the imprecise prior is contaminated by a "wrong" precise prior. QFB and GBR share the latter advantage.

We demonstrate QFB for a (special version of a) class of precise priors with prescribed marginals and arbitrary correlations. Such a class has been motivated by a recent expert elicitation among modelers working along issues of climate policy advice in the broadest sense. We find that QFB is considerably more informative than GBR for this class.

This suggests an interpretation of QFB. (The reader may use QFB on pragmatic grounds even if she or he would not like to follow the interpretation given in this paragraph.) One possible interpretation assumes (1) that within the class of priors, one prior is the – un-identified – "adequate" and (2) that prior measure can be given a frequentistic interpretation: the life-averaged successes and failures of an expert. Then QFB would imply that with probability $\geq Q$, QFB would acknowledge this adequate prior within the GBR-step. A nesting correction would probabilistically capture the cases if which the adequate prior would be lost. This is possible as upper and lower posterior probabilities are bounded functionals over the set of priors. Hence a nesting-corrected QFB would reveal upper and lower bounds on frequencies of events when averaged over the life of an expert. Remarkably, even after nesting-correcting QFB, QFB remains much more informative than GBR. Hereby we would like to stress that our implementation of QFB is by no means optimized w.r.t. being as informative as possible. One could further optimize Q together with the quantile functional.

Finally we illustrate the effects of various updating rules for the example of a stylized insurance situation. Under QFB much more risky clients could be insured compared to GBR.

A skeptic may argue that any updating rule which disregards precise priors in view of the evidence before applying GBR would be logically inconsistent, as the evidence were used twice: firstly, the evidence is used to disregard priors from then applying GBR. Secondly, those priors that "have made it," are again treated in view of the evidence, namely by standard Bayes' rule.

This counter-argument would apply for Gilboa's and Schmeidler's rule as well as for QFB. Such type of discussion is beyond the scope of this paper, however, we observe that society very often just behaves like that: it would listen more carefully to experts (i.e. precise priors) that have stated the evidence stronger in advance.

With this article we would like to fuel a discussion on the adequate update rule when updating classes of priors: is it allowed to disregard priors in view of the evidence before Bayesian updating? If yes, what is a meaningful filter? In addition, subsequent algorithms are needed to update imprecise priors that are much more precise on marginals than on (higher order) correlations. Those items seem to be crucial when modeling Bayesian updating of state-of-the-art models in politically influential modeling areas.

In any case it appears as stimulating and satisfying to see experts' relief when not being forced to specify precise measures but instead much less informative measures. We regard this observation as a key motivation for further investments in adequate imprecise models of prior knowledge and generalized Bayesian updating. This also implies the use of social data based choices of non-parametric priors and subsequent numerics.

Finally, we ultimately understand this constribution as an invitation to the "imprecise community" to develop a sound axiom system (as suggested by one of our reviewers) about updating and imprecision (in relation to information).

Acknowledgements

We would like to thank H. Küchenhoff for drawing our attention to the issue of "double-counting" the observational information in non-GBR updating rules. Furthermore we are grateful for the very valuable comments of three anonymous referees. The author has been supported by the Volkswagen Foundation under grant number II/78470.

References

[1] J. O. Berger, D. R. Insua, and F. Ruggeri. Bayesian robustness. In D. R. Insua and F. Ruggeri, editors, *Robust Bayesian analysis*, volume 125, pages 1–32. Lecture Notes in Statististics, Springer, New York, 2000.

[2] D. Berleant and Jianzhong Zhang. Using Pearson correlations to improve envelopes around the distribution of functions. *Reliable Computing*, 10(2):139–161, 2004.

[3] F. Coolen. Imprecise conjugate prior densities for the one-parameter exponential family of distributions. *Statistics & Probability Letters*, 16:337–342, 1993.

[4] F. Coolen. *PhD Thesis.* Eindhoven University of Technology, 1994.

[5] J. Dhaene, M. Denuit, M. J. Goovaerts, R. Kaas, and D. Vyncke. The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics*, 31:3–33, 2002.

[6] M. J. Frank, R. B. Nelsen, and B. Schweizer. Best-possible bounds for the distribution of a sum – a problem of Kolmogorov. *Probability Theory and Related Fields*, 74:199–211, 1987.

[7] I. Gilboa and D. Schmeidler. Updating ambiguous beliefs. *Journal of Economic Theory*, 59:33–49, 1993.

[8] H. Held, E. Kriegler, and T. Augustin. Bayesian learning for a class of priors with prescribed marginals. *preprint server http://www.stat.uni-muenchen.de/sfb386/*, 488, 2006.

[9] H. Held, E. Kriegler, and T. Augustin. Bayesian learning for a class of priors with prescribed marginals. *International Journal of Approximate Reasoning*, submitted.

[10] H. Held and T. Schneider von Deimling. Transformation of possibility functions in a climate model of intermediate complexity. *Advances in Soft Computing*, 6:337–345, 2006.

[11] M. Lavine, L. Wasserman, and R. L. Wolpert. Bayesian inference with specified prior marginals. *Journal of the American Statistical Association*, 86(416):964–971, 1991.

[12] M. Martel-Escobar, F. J. Vázquez-Polo, and A. Hernández-Bastida. Analysing the independence hypothesis in models for rare errors: an application to auditing. *Appl. Statist.*, 54(4):795–804, 2005.

[13] T. Seidenfeld and L. Wasserman. Dilation for convex sets of probabilities. *Annals of Statistics*, 21:1139–1154, 1993.

[14] A. H. Tchen. Inequalities for distributions with given marginals. *The Annals of Probability*, 8(4):814–827, 1980.

[15] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.

[16] P. Walley. A bounded derivative model for prior ignorance about a real-valued parameter. *Scandinavian Journal of Statistics*, 24:463–483, 1997.