Encoding and updating uncertain judgements using hybrid elicitation and quantile-parametrized distributions

Dmytro Perepolkin[1], Ullrika Sahlin[1]

[1] Centre for Environmental and Climate Science, Lund University, Sweden

Introduction

Elicitation of parameters [1, 2] requires that the expert understands the model and the role a particular parameter plays in it. *Elicitation of predictions* [3] mixes variability and uncertainty and, therefore, can not be used for learning from data without making additional assumptions [4, 5]. We propose a *hybrid elicitation* approach, where elicitation of the observable quantities is accompanied by the elicitation of the uncertainty around them. Hybrid elicitation can be used to define the prior distribution for a model defined by a quantile-parametrized distribution.

Objectives

1. Simplify the Bayesian prior specification by adopting a model parameterized by the quantiles of the observable quantity [6]
2. Elicit the food consumption distribution with uncertainty for exposure assessment
3. Update the expert belief with the observations of actual consumption obtained from the food consumption database

Materials and Methods

We extracted the summary statistics for Level 5 food category from the EFSA food consumption database. The following Figure 1 shows the quantile summary of the chronic consumption of bitter chocolate in Sweden for all age categories from different surveys. Each of the curves represents a predictive distribution for consumption of chocolate.

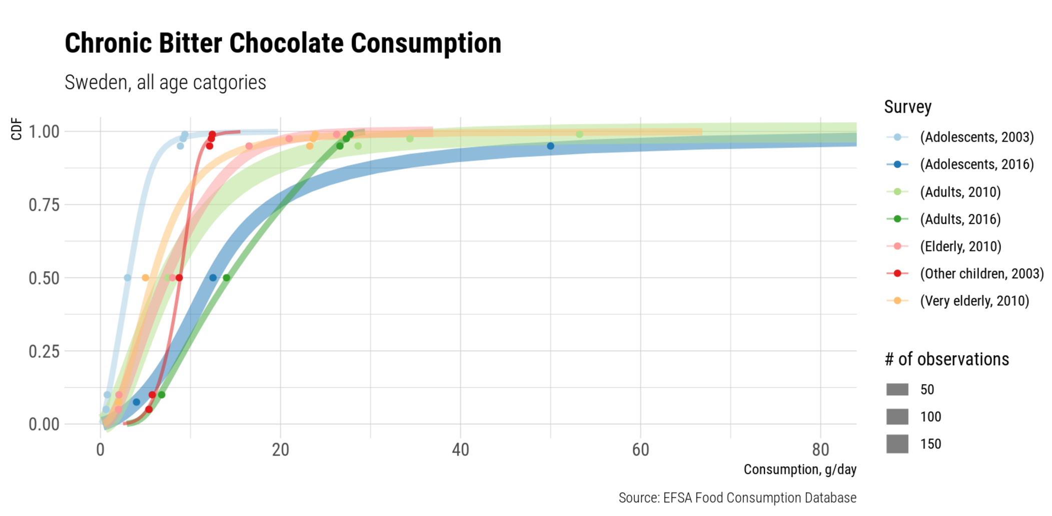

Figure 1: EFSA Food Consumption of bitter chocolate

Categories of consumers based on Swedish National Dietary Survey - Riskmaten 2010-11 are shown in Figure 2. Category cut-off points can be chosen arbitrarily off of the predictive distribution.

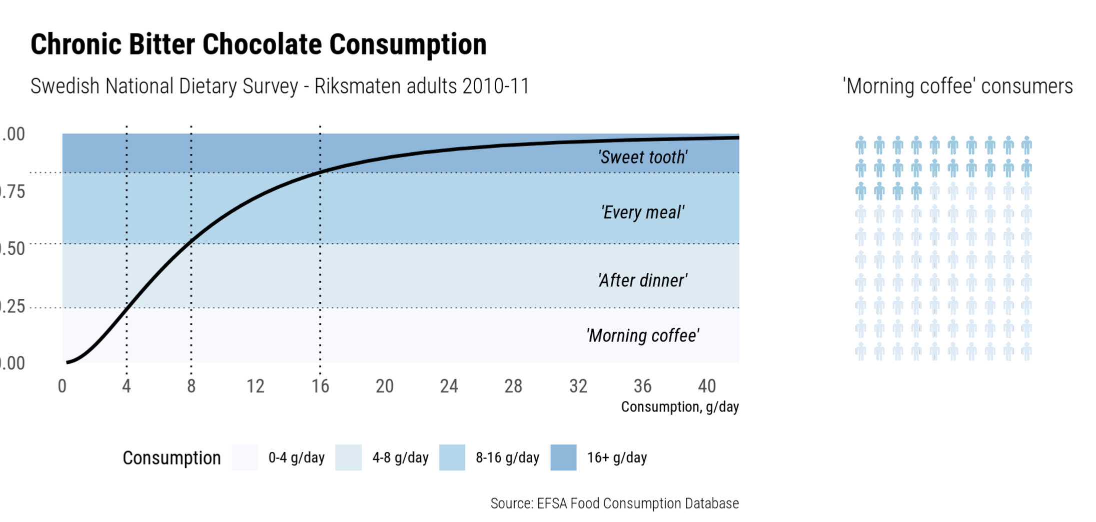

Figure 2: Consumer categories

We elicited the uncertainty in the cumulative probabilities associated with the fixed quantiles [7] using the icon arrays.

> **Interviewer:** Consider a sample of bitter chocolate consumers, say 100 people. According to your assessment there should be only 24 people that consume up to one bit of chocolate a day. We will interpret this assessment as you believing that there's about equal chance that the actual number of *morning coffee* consumers (0-4 g/day) in this sample will be above or below 24, i.e we will interpret it as the median assessment. Would you like to reconsider this value?

We adopt the approach described by [8] for eliciting the uncertainty in the probabilities related to a categorical variable to infer the parameter vector(s) of Dirichlet (or Connor-Mosimann) distribution. In this method the expert assesses the quartiles of the probability for each category using the symmetric percentile triplet elicitation. Elicited uncertainty in the category counts out of hypothetical sample of 100 people were recorded in Table 1.

Table 1: Elicited consumer counts by category

	Category	P25	P50	P75
1	Morning coffee (0-4 g/day)	18	24	30
2	After dinner (4-8 g/day)	25	31	36
4	Sweet tooth (16+ g/day)	8	15	20

* All consumer counts are out of a sample of 100

We encoded the elicited uncertainties in the cumulative probabilities by category into the hyperparameter vector of a Dirichlet distribution using the conditional univariate beta distributions [8]. The resulting Dirichlet distribution can be used as a prior for the model parametrized by quantiles.

Quantile-parameterized distributions (QPDs) are parameterized by a set of quantile-probability pairs (quantile-probability tuple, QPT) describing an observable [9, 10]. Uncertainty about the quantiles was updated using Bayesian inference [11].

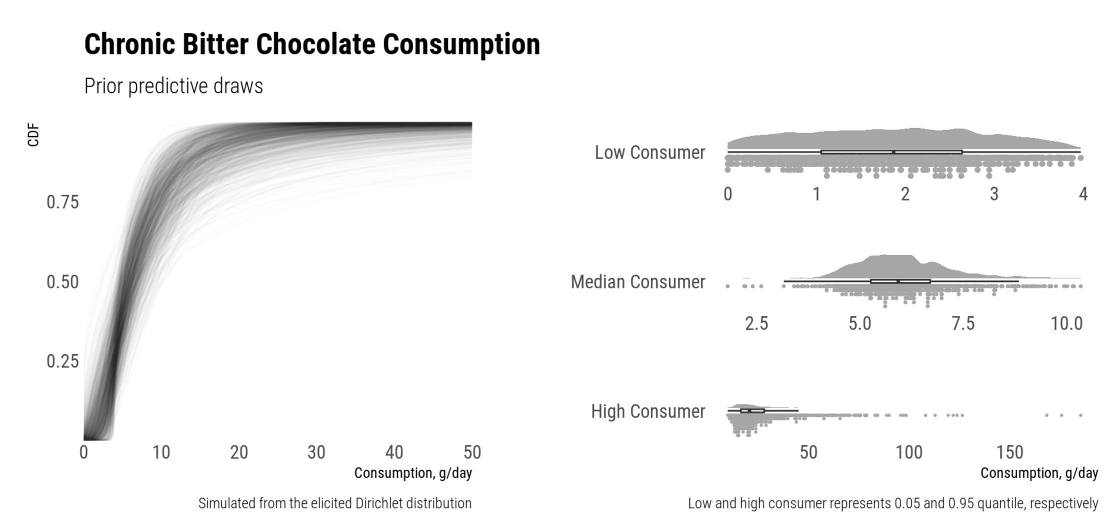

Figure 3: Prior draws from elicited QDirichlet distribution

Estimates of consumption by a median and a high consumer can be extracted from the posterior food consumption distribution together with uncertainty in these estimates, which can be used in exposure assessment (see Figures 3 and 4).

Results and discussion

The hybrid elicitation consists of two phases: elicitation of the quantile values q and elicitation of uncertainty in the cumulative probabilities associated with them (i.e. possible vectors of p which could correspond to the specified vector q). The hyperparameter vector q specifies the location of the QDirichlet prior, while the hyperparameter vector α is responsible for defining its shape.

The parameter vector of Dirichlet distribution combined with the vector of elicited quantiles act as hyper-parameters of the proposed QDirichlet prior, which describes the uncertainty in the parameters of the quantile-parameterized model (1).

$$\begin{aligned}\Delta &\sim Dirichlet(\alpha|q);\\ p &= \Xi_1^n(\Delta);\\ u &= \widehat{F}_{M_n}(x|p,q);\\ u &\sim f(Q_{M_n}(p,q))\end{aligned} \tag{1}$$

where Δ is a simplex of size $n+1$, $\Xi_1^n()$ is the cumulative sum operator, p is a size-n tuple of cumulative probabilities, q is the size-n vector of quantiles, corresponding to the cumulative probabilities p, u is the depth corresponding to the observable x given the parameterizing QPT $\{p,q\}$, $\widehat{F}_{M_n}(x|p,q)$ is the approximated CDF of the metalog distribution, represented by the numerically inverted $\widehat{Q}^{-1}_{M_n}(x|p,q)$), indirectly parameterized by $\{p,q\}_n$ via the vector of metalog coefficients a, and $f(Q_{M_n}(u|p,q)) = [q_{M_n}(u|p,q)]^{-1}$ is the metalog density quantile function (DQF).

Posterior predictive check shows significant reduction of uncertainty about the consumption of bitter chocolate.

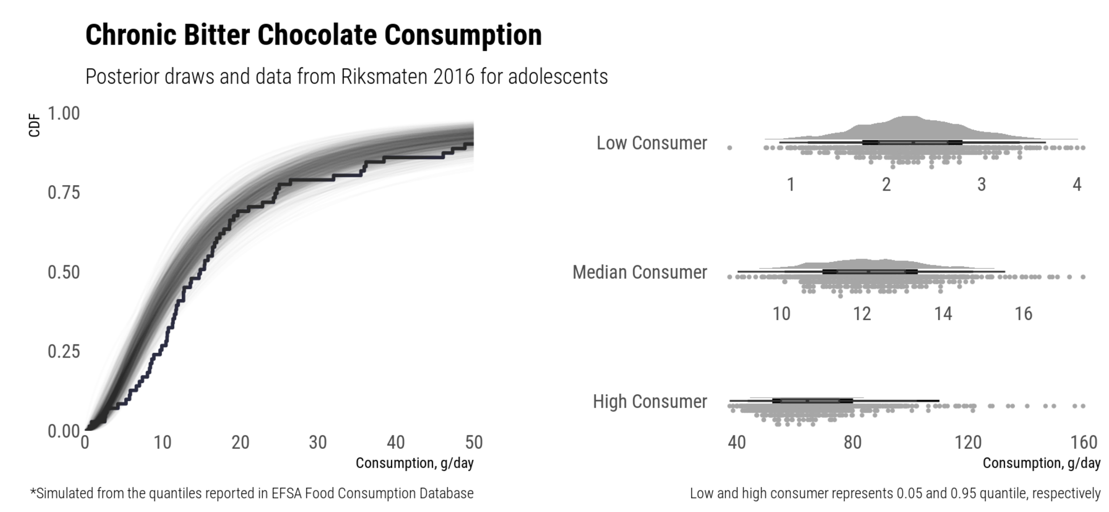

Figure 4: Posterior draws from QDirichlet-Metalog model

Conclusion

Asking experts to provide their uncertainty about the elicited QPT is enough to quantify the uncertainty about the food consumption distribution. This approach is particularly useful when the food consumption data is sparse.

Hybrid elicitation, like *predictive elicitation*, describes only observable quantities. At the same time, like *parametric elicitation*, the hybrid approach results in the characterization of uncertainty in the model parameters. Hybrid elicitation, therefore, can be viewed as observations-level parametric elicitation for quantile-parameterized models [6].

References

1. Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. Journal of the American Statistical Association. 2005;100:680–701.
2. O'Hagan A. Expert knowledge elicitation: Subjective but scientific. The American Statistician. 2019;73:69–81.
3. Kadane J, Wolfson LJ. Experiences in elicitation. Journal of the Royal Statistical Society: Series D (The Statistician). 1998;47:3–19.
4. Akbarov A. Probability elicitation: Predictive approach. PhD thesis. University of Salford; 2009.
5. Hartmann M, Agiashvili G, Bürkner P, Klami A. Flexible prior elicitation via the prior predictive distribution. arXiv:200209868 [stat]. 2020.
6. Perepolkin D, Goodrich B, Sahlin U. Hybrid elicitation and indirect bayesian inference with quantile-parametrized likelihood. 2021.
7. Crowder M. Bayesian priors based on a parameter transformation using the distribution function. Annals of the Institute of Statistical Mathematics. 1992;44:405–16.
8. Elfadaly FG, Garthwaite PH. Eliciting dirichlet and connormosimann prior distributions for multinomial models. TEST. 2013;22:628–46.
9. Keelin TW, Powley BW. Quantile-parameterized distributions. Decision Analysis. 2011;8:206–19.
10. Keelin TW. The metalog distributions. Decision Analysis. 2016;13:243–77.
11. Perepolkin D, Goodrich B, Sahlin U. The tenets of indirect inference in bayesian models. Preprint. https://osf.io/enzgs: Open Science Framework; 2021.