Combining probabilities/information from different sources

Lets say I have three independent sources and each of them make predictions for the weather tomorrow. The first one says that the probability of rain tomorrow is 0, then the second one says that the probability is 1, and finally the last one says that the probability is 50%. I would like to know the total probability given that information.

If apply the multiplication theorem for independent events I get 0, which doesn't seem correct. Why is not possible to multiply all three if all sources are independent? Is there some Bayesian way to update the prior as I get new information?

Note: This is not homework, is something that I was thinking about.

probability bayesian pooling model-averaging forecast-combination

edited May 11 at 17:27
Richard Hardy
21.8k 4 30 94

asked Jun 6 '15 at 22:25
Biela Diela
61 5

Do you know how reliable the independent sources are – Dilip Sarwate Jun 6 '15 at 22:44

No, a priori I would assume that all sources are equally reliable. – Biela Diela Jun 8 '15 at 8:08

1 This is a good question I am thinking about, too. I would add second question: If all predictions were 0.75, what would the combined probability be? Higher than 0.75? What would a formal framework be for analyzing this kind of questions? – Karsten W. Dec 29 '15 at 7:48

2 There's not really enough information; we need some model of how the predictions are expected to relate to reality. – Glen_b ♦ Dec 29 '15 at 9:16

I am not quite sure what is meant by "all sources are equally reliable" when the sources provide statements regarding probabilities or confidence/trust levels. If we are talking about the probability-that-a-certain-probability-has-a-given-value that seems to bring up conceptual problems. BTW, if sources 1 and 2 are equally reliable, they must both be right with probability 0.50... (and the probability of rain is 1/2). – A.G. Dec 30 '15 at 9:53

6 Answers

You ask about three things: (a) how to combine several forecasts to get single forecast, (b) if Bayesian approach can be used in here, and (c) how to deal with zero-probabilities.

Combining forecasts, is a common practice. If you have several forecasts than if you take average of those forecasts the resulting combined forecast should be better in terms of accuracy than any of the individual forecasts. To average them you could use weighted average where weights are based on inverse errors (i.e. precision), or information content. If you had knowledge on reliability of each source you could assign weights that are proportional to reliability of each source, so more reliable sources have greater impact on the final combined forecast. In your case you do not have any knowledge about their reliability so each of the forecasts have the same weight and so you can use simple arithmetic mean of the three forecasts

$$0\% \times .33 + 50\% \times .33 + 100\% \times .33 = (0\% + 50\% + 100\%)/3 = 50\%$$

As was suggested in comments by *@AndyW* and *@ArthurB.*, other methods besides simple weighted mean are available. Many such methods are described in literature about averaging expert forecasts, that I was not familiar with before, so thanks guys. In averaging expert forecasts sometimes we want to correct for the fact that experts tend to regress to the mean (Baron et al, 2013), or make their forecasts more extreme (Ariely et al, 2000; Erev et al, 1994). To achieve this one could use transformations of individual forecasts p_i, e.g. logit function

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) \tag{1}$$

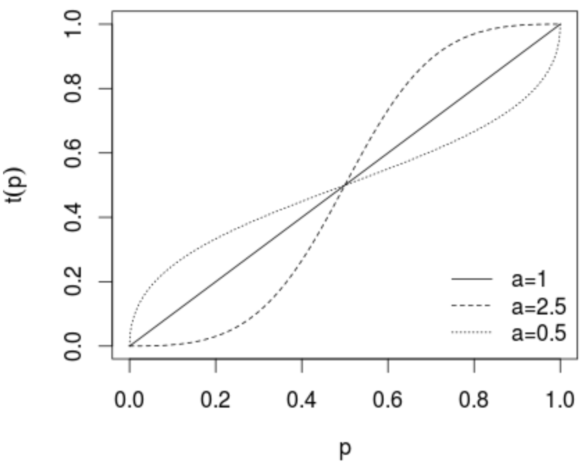
odds to the a-th power

$$g(p_i) = \left(\frac{p_i}{1-p_i}\right)^a \tag{2}$$

where $0 < a < 1$, or more general transformation of form

$$t(p_i) = \frac{p_i^a}{p_i^a + (1-p_i)^a} \tag{3}$$

where if $a = 1$ no transformation is applied, if $a > 1$ individual forecasts are made more extreme, if $0 < a < 1$ forecasts are made less extreme, what is shown on picture below (see Karmarkar, 1978; Baron et al, 2013).



After such transformation forecasts are averaged (using arithmetic mean, median, weighted mean, or other method). If equations (1) or (2) were used results need to be back-transformed using inverse logit for (1) and inverse odds for (2). Alternatively, geometric mean can be used (see Genest and Zidek, 1986; cf. Dietrich and List, 2014)

$$\hat{p} = \frac{\prod_{i=1}^N p_i^{w_i}}{\prod_{i=1}^N p_i^{w_i} + \prod_{i=1}^N (1-p_i)^{w_i}} \tag{4}$$

or approach proposed by Satopää et al (2014)

$$\hat{p} = \frac{\left[\prod_{i=1}^N \left(\frac{p_i}{1-p_i}\right)^{w_i}\right]^a}{1 + \left[\prod_{i=1}^N \left(\frac{p_i}{1-p_i}\right)^{w_i}\right]^a} \tag{5}$$

where w_i are weights. In most cases equal weights $w_i = 1/N$ are used unless *a priori* information that suggests other choice exists. Such methods are used in averaging expert forecasts so to correct for under- or overconfidence. In other cases you should consider if transforming forecasts to more, or less extreme is justified since it can make resulting aggregate estimate fall out of the boundaries marked by the lowest and the greatest individual forecast.

If you have *a priori* knowledge about rain probability you can apply Bayes theorem to update the forecasts given the *a priori* probability of rain in similar fashion as described in here. There is also a simple approach that could be applied, i.e. calculate weighted average of your p_i forecasts (as described above) where prior probability π is treated as additional data point with some prespecified weight w_π as in this IMDB example (see also source, or here and here for discussion; cf. Genest and Schervish, 1985), i.e.

$$\hat{p} = \frac{\left(\sum_{i=1}^N p_i w_i\right) + \pi w_\pi}{\left(\sum_{i=1}^N w_i\right) + w_\pi} \tag{6}$$

From your question however it does not follow that you have any *a priori* knowledge about your problem so you would probably use uniform prior, i.e. assume *a priori* 50% chance of rain and this does not really change much in case of example that you provided.

For dealing with zeros, there are several different approaches possible. First you should notice that 0% chance of rain is not really reliable value, since it says that it is *impossible* that it will rain. Similar problems often occur in natural language processing when in your data you do not observe some values that possibly can occur (e.g. you count frequencies of letters and in your data some uncommon letter does not occur at all). In this case the classical estimator for probability, i.e.

$$p_i = \frac{n_i}{\sum_i n_i}$$

where n_i is a number of occurrences of ith value (out of d categories), gives you $p_i = 0$ if $n_i = 0$. This is called *zero-frequency problem*. For such values you *know* that their probability is nonzero (they exist!), so this estimate is obviously incorrect. There is also a practical concern: multiplying and dividing by zeros leads to zeros or undefined results, so zeros are problematic in dealing with.

The easy and commonly applied fix is, to add some constant β to your counts, so that

$$p_i = \frac{n_i + \beta}{(\sum_i n_i) + d\beta}$$

The common choice for β is 1, i.e. applying uniform prior based on Laplace's rule of succession, $1/2$ for Krichevsky-Trofimov estimate, or $1/d$ for Schurmann-Grassberger (1996) estimator. Notice however that what you do here is you apply out-of-data (prior) information in your model, so it gets subjective, Bayesian flavor. With using this approach you have to remember of assumptions you made and take them into consideration. The fact that we have strong *a priori* knowledge that there should not be any zero probabilities in our data directly justifies the Bayesian approach in here. In your case you do not have frequencies but probabilities, so you would be adding some *very small* value so to correct for zeros. Notice however that in some cases this approach may have bad consequences (e.g. when dealing with logs) so it should be used with caution.

Schurmann, T., and P. Grassberger. (1996). Entropy estimation of symbol sequences. *Chaos, 6,* 41-427.

Ariely, D., Tung Au, W., Bender, R.H., Budescu, D.V., Dietz, C.B., Gu, H., Wallsten, T.S. and Zauberman, G. (2000). The effects of averaging subjective probability estimates between and within judges. *Journal of Experimental Psychology: Applied, 6*(2), 130.

Baron, J., Mellers, B.A., Tetlock, P.E., Stone, E. and Ungar, L.H. (2014). Two reasons to make aggregated probability forecasts more extreme. Decision Analysis, 11(2), 133-145.

Erev, I., Wallsten, T.S., and Budescu, D.V. (1994). Simultaneous over-and underconfidence: The role of error in judgment processes. *Psychological review, 101*(3), 519.

Karmarkar, U.S. (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational behavior and human performance, 21*(1), 61-72.

Turner, B.M., Steyvers, M., Merkle, E.C., Budescu, D.V., and Wallsten, T.S. (2014). Forecast aggregation via recalibration. *Machine learning, 95*(3), 261-289.

Genest, C., and Zidek, J. V. (1986). Combining probability distributions: a critique and an annotated bibliography. *Statistical Science, 1*, 114–135.

Satopää, V.A., Baron, J., Foster, D.P., Mellers, B.A., Tetlock, P.E., and Ungar, L.H. (2014). Combining multiple probability predictions using a simple logit model. *International Journal of Forecasting, 30*(2), 344-356.

Genest, C., and Schervish, M. J. (1985). Modeling expert judgments for Bayesian updating. *The Annals of Statistics*, 1198-1212.

Dietrich, F., and List, C. (2014). Probabilistic Opinion Pooling. (Unpublished)

edited Dec 30 '15 at 9:39

answered Dec 29 '15 at 9:20

Tim ♦
42.5k 7 91 163

1 I wanted to add to this rather than start a new answer. Another well known method is to combine the three (or N) probabilities by taking their geometric mean (rather than their arithmetic mean). Hinton points out that this gives a model with a very high or low probability, the 'veto' power amongst others, rather than averaging everything which may at times work against you. – Berkmeister Dec 29 '15 at 9:30

So, if the three forecasts were all 75%, and no information on their reliability is available, the final forecast would be 75%? – Karsten W. Dec 29 '15 at 9:41

@KarstenW. yes, why would you expect something different? If you have no a priori information, than this is the only information that you have, so you have no reason to consider the final result to be different... – Tim ♦ Dec 29 '15 at 9:53

1 Haven't read any of Tetlock's academic papers, but I would start there. Such as *Two Reasons to Make Aggregated Probability Forecasts More Extreme*. I will look up Phil's exact wording, I may be mis-remembering the word *extremify*. – Andy W Dec 29 '15 at 14:22

1 I was close with *extremified*, but not quite. I should have used *extremized*, see here. Besides the Baron et al. paper mentioned, I see Ville Satopää has some work on the topic arxiv.org/abs/1506.06405. – Andy W Dec 29 '15 at 16:29

There are two way to think of the problem. One is to say that the sources observe a noisy version of the latent variable "it will rain / it will not rain".

For instance, we could say that each source draws its estimates from a $Beta(a+b, a)$ distribution if it will rain, and a $Beta(a, a+b)$ distribution if it will not.

In this case, the a parameter drops out and the three forecast, x, y, and z would be combined as

$$p = \frac{1}{1 + \left(\frac{1}{x} - 1\right)^b \left(\frac{1}{y} - 1\right)^b \left(\frac{1}{z} - 1\right)^b}$$

b is a parameter controlling how under ($b > 1$) or over ($b < 1$) confident the sources are. If we assume that the sources estimates are unbiased, then $b = 1$ and the estimate simplifies as

$$\frac{p}{1-p} = \frac{x}{1-x}\frac{y}{1-y}\frac{z}{1-z}$$

Which is just saying: the odds of rain is the product of the odds given by each source. Note that it is not well defined if a source gives an estimate of exactly 1 and another gives an estimate of exactly 0, but under our model, this never happens, the sources are never *that* confident. Of course we could patch the model to allow for this to happen.

This model works better if you're thinking of three people telling you whether or not it rained yesterday. In practice, we know that there is an irreducible random component in the weather, and so it might be better to assume that nature first picks a probability of rain, which is noisily observed by the sources, and then flips a biased coin to decide whether or not it is going to rain.

In that case, the combined estimate would look much more like an average between the different estimates.

edited Dec 29 '15 at 22:23

answered Dec 29 '15 at 13:35

Arthur B.
1,736 6 11

What would x, y, z be in this model? – Karsten W. Dec 29 '15 at 17:31

It would be the three different predictions. – Arthur B. Dec 29 '15 at 17:32

The example you were wondering about would be $x = y = z = \frac{3}{4}$. In the framework I suggested as a reasonable choice, you would have $p = \frac{27}{28}$. This is because $\frac{3}{4}$ represents 3 to 1 odds, so the product represents 27 to 1 odds, or a $\frac{27}{28}$ probability. – Arthur B. Dec 29 '15 at 18:04

Going from 3/4 to 27/28 is a bit extreme, it is like three people were telling you that the sky is dark blue and you concluded it is black... – Tim ♦ Dec 29 '15 at 21:58

It depends on the model. Here I'm assuming each source has a noisy view on a latent binary variable, rain or no rain. It's more like three different people tell you it rained yesterday. You can also model the system as there being a latent probability of rain and the forecast sources as getting a noisy version of that forecast. – Arthur B. Dec 29 '15 at 22:12

Their numbers for rain likelihood is only half the story, as we'd have to temper their predictions with the probability that they are accurate when making guesses.

Because something like rain is mutually exclusive(it's either raining or isn't, in this setup), they cannot all simultaneously be correct with 75% probability as Karsten suggested (I think, hard to tell with the confusion I hear about what it means to find "combined probability").

Taking into consideration their individual abilities to predict the weather, we could take a stab (a la Thomas Bayes, as in a generally blind shot in the dark) at what the chance of rain is tomorrow.

Station 1 is correct in their predictions 60% of the time, the second 30% of the time, and the last station a poor 10% of the time.

E[rain]=PxX+PyY+Pz*Z is the form we're looking at here:

(.6)(0)+(.3)(1)+(.1)(.5) = E[rain] = 35% chance of rain with made up prediction accuracies.

answered Dec 29 '15 at 8:55

Havok
21 1

This algorithm can produce values above 1. – Andy W Dec 29 '15 at 14:36

In the framework of Transferable Belief Model (TBM), it is possible to combine different predictions using for instance the "conjunctive rule of combination". In order to apply this rule, you need to transform the probabilities of the predictions into basic belief assignments. This can be achieved with the so-called Least-Committed-Principle. In R:

```
library(ibelief)
#probabilities
p1 <- c(0.99, 0.01) # bad results for 0 and 1
p2 <- c(0.01, 0.99)
p3 <- c(0.5, 0.5)

# basic belief assignment,
# each row represents a subset of (rain, not rain)
# each column represents one prediction
Mat <- LCPrincple(rbind(p1,p2,p3))

# combine beliefs
m <- DST(Mat, 1)

# resulting probability distribution (pignistic probability)
mtobetp(m)
# returns 0.5 and 0.5
```

For the second example of three independent predictions of 0.75, this approach returns a higher value:

```
p4 <- c(0.75, 0.25)
Mat <- LCPrincple(rbind(p4,p4,p4))
m <- DST(Mat, 1)
mtobetp(m)
#returns 0.9375 0.0625
```

This is not very far from the Bayesian approach shown in Arthur B's answer.

edited Dec 30 '15 at 12:16

answered Dec 29 '15 at 18:27

Karsten W.
497 4 22

I think it's worthwhile to look at the weighting scheme based on inverse errors mentioned in one of the answers. If the sources are truly independent and we constrain the weights to sum to one, the weights are given by

$$w_1 = \frac{\sigma_2^2\sigma_3^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2},\ w_2 = \frac{\sigma_1^2\sigma_3^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2},\ w_3 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2}.$$

If, as the OP states, the forecasts are equally reliable, then all weights will simplify to $\frac{1}{3}$ and the combined forecast for the given example will be 50%.

Note that the values of σ_i do not need to be known if their relative proportions are known. So if $\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = 1 : 2 : 4$, then the forecast in the example would be

$$f = \frac{8}{14} * (0) + \frac{4}{14} * (1) + \frac{2}{14} * (0.5) = 0.3571$$

edited Jan 7 '16 at 23:21

answered Jan 7 '16 at 23:15

soakley
3,056 3 10 20

For combining reliability, my go-to formula is r1xr2xr3÷(r1xr2xr3+(1-r1)x(1-r2)x(1-r3). So for the 3 sources of reliability 75% all saying the same thing, i would have .75^3 ÷ (.75^3 + .25^3) =>

96% reliability of the combined response

answered Sep 20 at 3:27

user3902302
99

1 This doesn't seem to be a proper answer to the question. – Michael Chernick Sep 20 at 3:58

Admittedly, it was more of a response to KarstenW comments than a direct response to the question. – user3902302 Sep 22 at 0:25