NHH

INSTITUTT FOR FORETAKSØKONOMI

DEPARTMENT OF BUSINESS AND MANAGEMENT SCIENCE

FOR 7 2015

ISSN: 1500-4066

January 2015

Discussion paper

Best estimate reporting with asymmetric loss

BY

Jostein Lillestøl AND **Richard Sinding-Larsen**

Norges Handelshøyskole

NORWEGIAN SCHOOL OF ECONOMICS

Best estimate reporting with asymmetric loss

Jostein Lillestøl[1]

Norwegian School of Economics

Richard Sinding-Larsen[2]

Norwegian University of Science and Technology

Jan. 21, 2015

Abstract

This paper considers the problem of point prediction based on a predictive distribution, representing the uncertainty about the outcome. The issue explored is the reporting of a single characteristic, typically the mean, the median or the mode, in the context of a skewed distribution and asymmetric loss. Special attention is given to the two-piece normal distribution and asymmetric piecewise linear and quadratic loss. The practical context for the issue is the yearly reporting of remaining petroleum resources given by the authorities to stakeholders that may ask for just a single number.[3]

[1] Department of Business and Management Science, Norwegian School of Economics, Helleveien 30, N-5045 Bergen, Norway; e-mail: jostein.lillestol@nhh.no

[2] Department of Geology and Mineral Resources Engineering, Norwegian University of Science and Technology, Sem Sælands veg 1, N-7491 Trondheim; e-mail: richard.sinding-larsen@ntnu.no

[3] We owe thanks to Per Blystad and Benvenutta Henriksen at the Norwegian Petroleum Directorate for arranging a seminar on this issue at NPD in the Fall 2014 with fruitful discussions.

Best estimate reporting with asymmetric loss

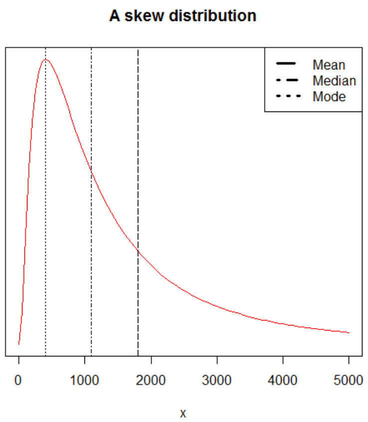
Consider the problem of point prediction based on a predictive distribution, representing the uncertainty about the outcome. The context may be

- individual: gamble on outcome with loss/gain
- business: budget a future uncertain outcome
- national: report an anticipated economic state

The predictive distribution may originate from individual beliefs, a single expert or a panel of experts. The tools for derivation may range from none at all to sophisticated analytic models and methods. Examples of all three contexts may be prediction of the oil price at a certain future point or period in time. Often such predictions are reported periodically based on updated knowledge, as a time series of predictions. In each epoch the predictions problem is the same, based on the current predictive distribution, in some cases just updates from new data. Another example of national importance is the periodic reporting of remaining oil reserves in a continental shelf, discovered and undiscovered, as basis for the political process of planning for the future. Since we have both depletion and new discoveries, the situation is sequentially slightly different, but at a given point in time it is essentially the same.

The question of remaining oil deposits relates to many fields of knowledge. To establish the predictive distribution experts from |the relevant disciplines are given the opportunity to express the expectations and uncertainties related to their field of knowledge. This is then combined by simulations, giving a single predictive distribution, representing the current best knowledge of the uncertainties of the remaining deposits. This distribution may be useful in various contexts, by planners and politicians. Some may comprehend this distribution as is, but many requests an estimate given as a single number, to be used as input to planning calculation or just as a political argument. The common practice of reporting the mean (expected value) of the distribution has some implications that are not always taken serious.

The predictive distribution may typically look like this, i.e. skewed with a long right tail.



The Mean (M_2), Median (M_1) and Mode (M_0) will then be located as indicated $M_0 \leq M_1 \leq M_2$. This means that the reporting of the mean could be far from the most likely outcome, the Mode.

The art of guessing the outcome may be judged in terms of a loss function L where L(x,y) denotes the loss of guessing x when the outcome turns out to be y. With a given loss function we may then guess/choose the action x that minimizes the expected loss calculated from the predictive distribution F. We will tacitly assume that this distribution is of continuous type, unimodal and that the expectation exists. It is a well-known fact that the optimum x is

$$x = M_2 \quad \text{for } L(x,y) = (y-x)^2 \qquad \text{(quadratic error loss)}$$
$$x = M_1 \quad \text{for } L(x,y) = |y-x| \qquad \text{(linear error loss)}$$
$$x = M_0 \quad \text{for } L(x,y) = \delta(x,y) \qquad \text{(0-1 error loss)}$$

This is so whatever the distribution of y. If the predictive distribution is derived in a Bayesian context, the corresponding prediction is commonly referred to as a Bayes estimate.

What is the possible impact of this for practice? An individual decision maker may possibly have an idea of his/her possible losses and choose action accordingly. The same may be true for a business. For a provider of knowledge to other parties the problem is what to report. The opportunities are:

i. Report the whole distribution, from which different characteristics can be derived (and reported as well).
ii. Report some key characteristics (Mean, Median, Mode, Quantiles)
iii. Report just one key characteristic

With extensive reporting it may be a risk that the users just look at the quantity they are mostly used to, typically the mean, regardless of this being the most relevant or not. The provider should therefore imagine the context the report is likely to be used and emphasize the relevant characteristics, and even refrain from reporting some less relevant ones. Clearly there are contexts where the mean is of minor relevance, but the median and/or mode are.

It is interesting to note that a many central banks nowadays use so-called fan-charts when reporting projections of inflation and other macroeconomic variables for several periods ahead, a practice initiated by the Bank of England, see Britton, Fisher &Whitley (1997). Fan chart is a generic term, and may be displayed using the mean, median or mode as point of departure and displayed as the central line. Bank of England reports a modal fan chart while other central banks, e.g. Sweden and Poland use median/quantile based fan charts, see Kowalczyk (2013). There seem to be a consensus that charts based on the mean do not provide the appropriate message to the users of these reports. If we are just considering the choice of one-point reporting, we should nevertheless take into account the preferred type of more extensive reporting. If we offer a modal fan chart for the more sophisticated reader the one-point prediction for the novice should be the mode as well.

The loss functions above are symmetric. In many cases the loss of over-prediction is larger than the loss of under-prediction. A recent detailed and deep going account of issues related to point forecast and their relationship to loss functions is given in a series of papers by Gneiting (2008, 2011,a,b) with many references to the literature, which goes back at least to the 1960's. We will here limit ourselves to asymmetric losses that generalize the first two above, namely piecewise linear loss and piecewise quadratic loss. A natural generalization of the one-point 0-1 loss is to define no loss in an interval around the true value. A symmetric interval gives rise to an optimal prediction being the midpoint in the so-called modal interval. This may be generalized to an asymmetric interval. We will not pursue this, but make comparisons directly with the mode itself.

The asymmetric piecewise linear loss function is defined as follows: For $0 < \alpha < 1$ let

$$\begin{aligned} L(x,y) &= \alpha\ |y-x| && \text{if } x \leq y \\ &= (1-\alpha)\ |y-x| && \text{if } x \geq y \end{aligned}$$

With this loss function the optimal point prediction is the α-quantile of the predictive distribution, i.e. q_α so that $F(q_\alpha) = \alpha$, where F denotes the cumulative distribution.

Here the ratio $\rho = (1-\alpha)/\alpha$ represents the relative size of losses from over- and under-prediction of the same size respectively. For α=½ we are back to the median, and for α < ½ we predict less than the median. In fact the α-quantile is optimal for a wider class of loss distributions named the generalized piecewise linear loss function (GPL). Moreover, if the α-quantile is optimal whatever F, the loss function L must be of GPL type, provided some reasonable conditions on L.

One may ask the fundamental question: Is the mean optimal under a wider class of loss functions than the quadratic, whatever the predictive distribution? Insights to this are given by Granger (1969), Savage (1971) and Banerjee, Guo & Wang (2005). It turns out that the mean is optimal if and only if the loss function is of so-called Bregman type (under reasonable regularity conditions). It should be noted that the only one in this class which depends on the prediction error only is the quadratic loss function.

We are mainly interested in the consequences of using the mean in case of asymmetric loss when the mean is not optimal, but optimal in the symmetric limit. This way we may challenge the indiscriminate use of the mean by some rhetoric questions. Imagine the following dialogue:

- "You report the mean only?"
- "Yes! I have heard that this is optimal!"
- "Well, that's true in a sense, but do your stakeholders regard over-prediction and under-prediction equally worse?"
- "No! Over-prediction would be more regrettable!"
- "Then the main reason for reporting the mean vanishes, but points to good alternatives!"

The idea of the challenger is to argue from the imagined premises of the reporter, by modifying the loss function for which the mean is optimal. To avoid confusion the main rationale (linear loss) for quantiles is kept out of the discussion.

The asymmetric piecewise quadratic loss function is defined by

$$
\begin{aligned}
L(x,y) &= \alpha\,(y-x)^2 && \text{if } x \leq y \\
&= (1-\alpha)\,(y-x)^2 && \text{if } x \geq y
\end{aligned}
$$

Again $\alpha < ½$ corresponds to over-prediction being the most serious. The point prediction that minimizes the expected loss is now given implicitly by the solution of the following equation

$$x = \mu + \frac{2\alpha-1}{\alpha}(xF(x) - F_1(x))$$

where $F_1(x) = \int_{-\infty}^{x} y dF(y)$ and $\mu = EY = F_1(+\infty)$. This formula appears maybe first in the Econometrica paper of Newey and Powell (1987), where the solution is given the name expectile and properties that resembles that of quantiles are pointed out.

Since $xF(x) > F_1(x)$ it follows, as expected, that $x < \mu$ for $\alpha < 1/2$, when over-prediction is the most serious. If we use $\rho = \frac{1-\alpha}{\alpha}$ as parameter we have $\alpha = \frac{1}{1+\rho}$ and the coefficient $\frac{2\alpha-1}{\alpha} = 1 - \rho$.

In the case of F being a location-scale family of distributions, i.e. of form

$$F(x) = G(\frac{x-\mu}{\tau})$$

where G is its standardized distribution with τ as scale parameter, we may write the equation on parameter-free form with $z = \frac{x-\mu}{\tau}$ as

$$z = \frac{2\alpha-1}{\alpha}(zG(z) - G_1(z))$$

Examples of this are the Gaussian distribution and the Logistic distribution. With symmetric distribution it follows, by $G(-z) = 1 - G(z)$ and $G_1(-z) = G_1(z)$, that when z is the solution for $(\alpha, 1-\alpha)$ the –z is the solution for $(1-\alpha, \alpha)$.

The equation may be easily solved by a simple univariate solver, like *uniroot* in the package R. We may also compute the solution iteratively by

$$z_{n+1} = \frac{2\alpha-1}{\alpha}(z_n G(z_n) - G_1(z_n)) \quad \text{with} \quad z_0 = 0$$

This is easily accomplished in spreadsheet, but some care must be taken to assure convergence in specific applications. Sometimes non-convergence is solved by choosing alternative starting values.

In the case of F being a scale family of distributions parameterized by its expectation , i.e. of form

$$F(x) = H(\frac{x}{\mu})$$

where G is its standardized distribution with expectation one, we may write the equation on parameter-free form with $z = x/\mu$ as

$$z = 1 + \frac{2\alpha - 1}{\alpha}(zH(z) - H_1(z))$$

An example of this is the exponential distribution. The corresponding iterative scheme will be

$$z_{n+1} = 1 + \tfrac{2\alpha-1}{\alpha}(z_n H(z_n) - H_1(z_n)) \quad \text{with} \quad z_0 = 1$$

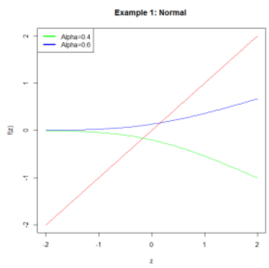
Note that the move away from the mean now is multiplicative instead of additive as above.

Example 1: Normal distribution N(μ, σ^2)

Take $z = \frac{x-\mu}{\sigma}$ and note that $G(z) = \Phi(z)$ and $G_1(z) = -\phi(z)$, where $\Phi(z)$ and $\phi(z)$ are the standard Gaussian cumulative distribution and density respectively. We therefore have

$$z = \frac{2\alpha - 1}{\alpha}(z\Phi(z) + \phi(z))$$

with recursion accordingly. The solution may be illustrated graphically by plotting the right hand side of the equation f(z) as function of z and see where it crosses the 45 degree line z. In Figure 1 this is illustrated for α=0.4 and α=0.6. We also give the optimal z for various α=0.1, 0.2,…, 0.9.



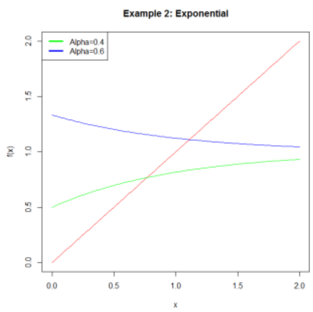
alpha	rho	znorm
0.1	9.00	-0.8616
0.2	4.00	-0.5492
0.3	2.33	-0.3371
0.4	1.50	-0.1617
0.5	1.00	0.0000
0.6	0.67	0.1617
0.7	0.43	0.3371
0.8	0.25	0.5492
0.9	0.11	0.8616

The computations also show that by doubling the losses on one side we have to move the optimal guess about 26% of the standard deviation from the expectation to the other side. With four times the size of loss to one side we have to move the optimal, prediction about 55% of the standard deviation to the other side.

Example 2: Exponential distribution(μ)

In this case take $z = x/\mu$. We now get $H(z) = 1 - e^{-z}$ and $H_1(z) = 1 - e^{-z} - ze^{-z}$, so that $z = 1 + \frac{2\alpha-1}{\alpha}(z - 1 - e^{-z})$, which can be rearranged to

$$z = 1 + \frac{2\alpha - 1}{1 - \alpha}e^{-z}.$$



alpha	rho	zexpo
0.1	9.00	0.4102
0.2	4.00	0.5801
0.3	2.33	0.7226
0.4	1.50	0.8588
0.5	1.00	1.0000
0.6	0.67	1.1572
0.7	0.43	1.3468
0.8	0.25	1.6035
0.9	0.11	2.0401

Note the obvious special cases z=0 for α=0 and x tending to infinity when α tends to one.

Example 3: Lognormal distribution

With logY distributed Normal(θ, τ^2) we have for Y itself

$$\text{Mean}=\mu = e^{\theta+1/2\tau^2}, \text{ Median}= e^{\theta} \text{ and Mode} = e^{\theta-\tau^2}$$

$$F(x) = \Phi(\tfrac{logx-\theta}{\tau}) \quad \text{and} \quad F_1(x) = \mu\cdot\Phi(\tfrac{logx-\theta}{\tau} - \tau)$$

so that

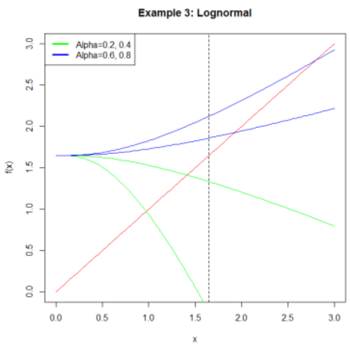
$$x = \mu + \frac{2\alpha-1}{\alpha}(\Phi\left(\frac{logx-\theta}{\tau}\right) - \mu\cdot\Phi\left(\frac{logx-\theta}{\tau} - \tau\right))$$

with recursion

$$x_{n+1} = \mu + \frac{2\alpha-1}{\alpha}(\Phi\left(\frac{logx_n-\theta}{\tau}\right) - \mu\cdot\Phi\left(\frac{logx_n-\theta}{\tau} - \tau\right))$$

As an illustration take $\theta = 0$ and $\tau = 1$ so that
Mean= $\mu = e^{0.5} = 1.6487$, Median= $e^0 = 1$ and Mode= $e^{-1} = 0.3679$.

$$x_{n+1} = \mu + \frac{2\alpha-1}{\alpha}(\Phi(logx_n) - e^{0.5}\cdot\Phi(logx_n - 1))$$



alpha	rho	xlogn
0.1	9.00	0.7209
0.2	4.00	0.9732
0.3	2.33	1.1929
0.4	1.50	1.4118
0.5	1.00	1.6487
0.6	0.67	1.9255
0.7	0.43	2.2787
0.8	0.25	2.7931
0.9	0.11	3.7704

It is of interest to see to what extent the optimal prediction is shifted from the expectation towards the median and the mode for increasing loss attached to over-prediction over under-prediction. We may phrase this conversely by asking how large the difference has to be to give the median or the mode as optimal. In the example above we have

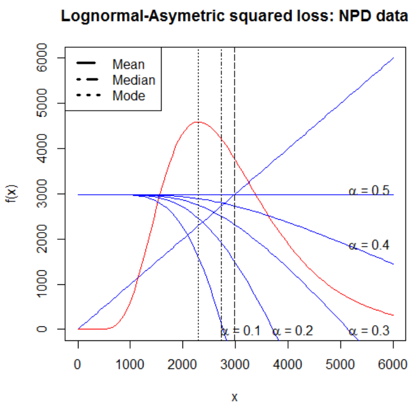
ρ	α	X
1	0.50	1.6787
2	0.33	1.2650
3	0.25	1.0846
4	0.20	0.9332

We get exactly x=1 (Median) for ρ=3.72 with α=0.2118 and exactly x=0.3679 (Mode) for ρ=62.37 with α=0.01578. Consequently we see that the Median is superior when the loss of over-prediction is 3.7 times larger than for under-prediction of a given size. From this it looks like an extreme unbalance in losses is required to make the Mode optimal. We shall see that this is not so in general for any choice of parameters in the lognormal distributions.

In connection with the examples above we could also ask the questions: For which ρ will the Mode be superior to the Mean, and for which ρ will the Mode be superior to the Mean?

Example 4: A Lognormal fit to NPD data

The Norwegian Petroleum Directorate NPD provides periodically estimates of the remaining deposits, discovered and undiscovered, on the Norwegian Shelf. Based on an established predictive distribution they have reported (in some units): Mean=2980 and (5%, 95%)–Quantiles, her named (Lower, Upper)= (935, 5420). This is all we know, but experience tells that such distributions are skewed to the right, maybe similar to the lognormal. However, with this data it is apparent a lognormal with the given mean cannot be found that gives good fit to both tails, even if we choose the three-parameter (translated) lognormal. Here the context of reporting an over-prediction error is the most serious, and a good fit to the right tail of the distribution is the most important. We therefore assume a lognormal distribution and use the reported Mean=2980 and Upper=5420 to fix the parameters (θ,τ) in the corresponding normal distribution. We obtained (θ,τ) = (7.913, 0.416). This gives a lognormal distribution with Median=2738 and Mode=2298. Alternatively we could fit a three-parameter lognormal by stipulating the minimum possible value. With the given data a reasonable choice is min=60, and then we get the excess fitted to a lognormal distribution with (θ,τ) = (7.890, 0.424). However, this does not make any difference for the points made here. In the graph below we displayed the determination of the optimal x for the cases α=0.1, 0.2, 0.3, 0.4, 0.5 with an overlaid lognormal distribution.



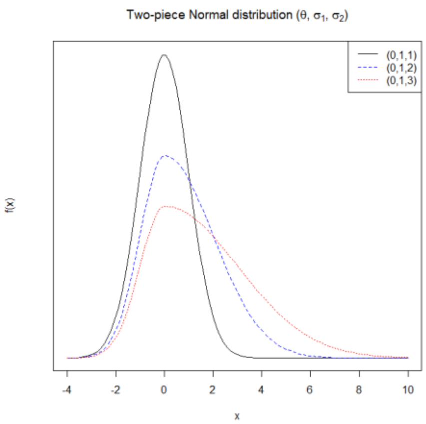
The computations show that α=0.371 with ρ=2.155 makes the Median optimal, and α=0.169 with ρ=5.917 makes the Mode optimal. So, if we the loss by over-prediction of a given size is twice that of under-prediction of the same size, the Median is optimal. Similarly it takes about five times more to make the Mode optimal. In the given context this is not an unreasonable unbalance between the two.

Asymmetric squared loss and the Two-piece Normal distribution

In situations where both tails of the distribution are relevant, and we know they are different, we may want to model them separately, in conjunction with some centrality measure, say mean, median or mode. An opportunity of this kind is offered by so-called two-piece normal distribution, sometimes referred to as the Fechner distribution. This is defined by the density

$$\begin{aligned} f(y) &= A \cdot e^{-\frac{1}{2\sigma_1^2}(y-\theta)^2} && y \leq \theta \\ &= A \cdot e^{-\frac{1}{2\sigma_2^2}(y-\theta)^2} && y \geq \theta \end{aligned}$$

where $= \frac{1}{\sqrt{2\pi}(\sigma_1+\sigma_2)/2}$. We see that this is two pieces of possible different normal distributions, scaled to give a common value $f(\theta) = A$ at the mode θ. The probabilities of the left and right pieces will be $\frac{\sigma_1}{\sigma_1+\sigma_2}$ and $\frac{\sigma_2}{\sigma_1+\sigma_2}$ respectively. For $\sigma_1 < \sigma_2$ we have positive skewness, i.e. longer right tail.



The expressions for the expectation and the variance are

$$EY = \theta + \sqrt{\frac{2}{\pi}}(\sigma_2 - \sigma_1) \qquad\qquad var(Y) = \left(1 - \frac{2}{\pi}\right)(\sigma_2 - \sigma_1)^2 + \sigma_1\sigma_2$$

The cumulative distribution may be expressed in terms of the cumulative standard normal as

$$\begin{aligned} F(y) &= \frac{2\sigma_1}{\sigma_1 + \sigma_2}\,\Phi(\frac{y-\theta}{\sigma_1}) && y \leq \theta \\ &= \frac{2\sigma_2}{\sigma_1 + \sigma_2}\,\Phi\left(\frac{y-\theta}{\sigma_2}\right) + \frac{\sigma_1 - \sigma_2}{\sigma_1 + \sigma_2} && y \geq \theta \end{aligned}$$

From this follows expressions for the quantiles in terms of standard normal quantiles: The left quantile $y_p = \theta + \sigma_1 \cdot z_q$ with $q = p \cdot \frac{\sigma_1+\sigma_2}{2\sigma_1}$ and the right quantile $y_{1-p} = \theta + \sigma_2 \cdot z_{1-q}$ with $q = p \cdot \frac{\sigma_1+\sigma_2}{2\sigma_2}$. The median may be expressed as

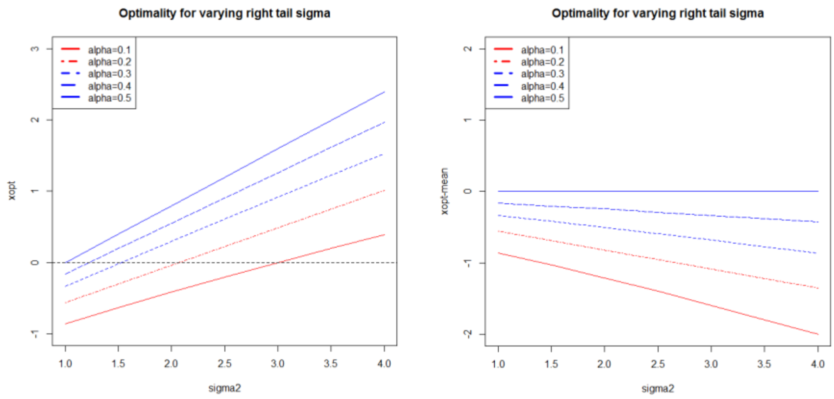
$$
\begin{aligned}
med(Y) &= \theta + \sigma_1 \Phi^{-1}\left(1 - \frac{\sigma_1 + \sigma_2}{4\sigma_1}\right) \qquad \sigma_1 > \sigma_2 \\
&= \theta + \sigma_2 \Phi^{-1}\left(1 - \frac{\sigma_1 + \sigma_2}{4\sigma_2}\right) \qquad \sigma_1 < \sigma_2
\end{aligned}
$$

The quantile functions may now be used for the analysis of optimal one-point prediction with asymmetric linear loss. For the asymmetric squared loss we need the function F_1(x) as well. However, in practice we do not need the explicit analytic expression. We may just do the numeric integration based on the density.

The two-piece normal distribution has some added advantages over the lognormal distribution for modelling this kind of data, besides the opportunity to fit both tails individually. We know that the predictive distribution is established by some kind of aggregation, and then we expect that there will be central limit effect that points towards normality. However, we see that such distributions may nevertheless be fairly skew, mainly due to more extreme underlying skewness and dependencies of the aggregates. Still it is attractive to at least having a skew distribution with symmetry and normality as a limit, i.e. when the defining standard deviations become equal.

Example 5: Two-piece normal distribution

Let us see how the optimal prediction is affected by the combination of asymmetric loss and skewness of the distribution by taking $\sigma_1 = 1$ and vary σ_2 for the case of mode=θ=0. In the following plot we have the optimal x for each of the cases α=0.1, 0.2, 0.3, 0.4, 0.5 plotted for σ_2 in the range from 1.0 to 4.0. For $\alpha = 0.5$ the optimal prediction is the expectation, which comes out as a straight line in σ_2, as expected. We see that the other cases come out as approximately straight lines as well. In the graph we have drawn the zero-line corresponding to the mode. The median curve (not drawn) will also be approximately a straight line, starting at (1,0) with slope 2/3 ending at (4,2). The calculations of the optimal predictions are performed by the iterative scheme of the kind above. Convergence was reached in the cases of not too large asymmetry of losses (cases in blue), while special attention had to given to the cases of large asymmetry (cases in red). With a slight modification of the graph we see more directly the size of the push away from the mean.



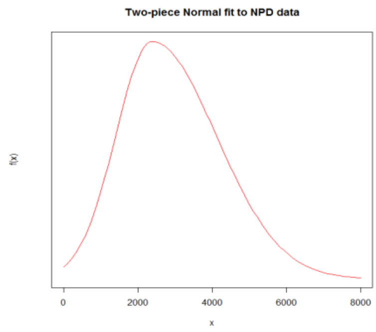
From the graph we see the combinations of asymmtry of both loss and distribution that lead to predictions below, around and above the mode respectively. For instance we are at the mode for combinations (α, σ_2) = (0.3, 1.5), (0.2, 2.0), (0.1, 3.0). Here it is worthwhile to be remained that I the former case of little asymmetry the mean and the mode are not much different, but in the latter case of large asymmetric distribution they are very different, and it takes more asymmetry of the loss to be pushed all the way to the mode. Consequently we have the following important conclusion in this context: The mode may be a good replacement for the mean both in case of small joint asymmetry and large joint asymmtry (but for slightly different reasons).

Example 6: A two-piece normal fit to NPD data

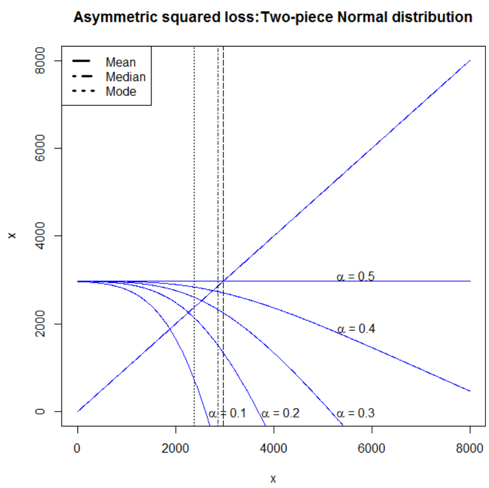
Consider again the data from The Norwegian Petroleum Directorate where they reported a predictive distribution with (5%, 95%)–quantiles (Lower, Upper)= (935, 5420) and Mean=2980. Fitting a two-piece normal distribution to this gave the following:

$$\theta = 2383, \quad \sigma_1 = 977\,, \ \sigma_2 = 1725$$

From this we obtain the Median=2858. The fitted distribution looks like this:



The corresponding analysis of optimal one point prediction gave the following plot:



We see that Median is optimal for about α=0.45 corresponding to ρ=1.22, and Mode is optimal for about α=0.25 corresponding to ρ=3.00. A more detailed analysis computing the actual losses shows that: Median is better than Mean for α<0.472 (corresponding to ρ>1.12), Mode is better than Mean for α<0.365 (corresponding to ρ>1.74), and Mode is better than Median for α<0.339 (corresponding to ρ>1.95). This means that Mean is outperformed by the Median even for the minor increase of 12% in the losses of over-prediction in comparison with under-prediction of the same size and that the Mean is outperformed by the Mode for a 74% increase. At 95% increase the Mode also outperforms the Median.

Some questions of interest now are: What is a reasonable value of α? Is it possible to elicit α for a well-defined group of stakeholders? What should guide the reporter when there are several groups of stakeholders with differing interests and views? What should guide us when the stakeholders are not clearly defined? Should we give priority to some (more important) stakeholders?

In some cases it may be possible to imagine that all (important) stakeholders incur a loss corresponding to an α beyond a certain size that makes the mean less valuable than the mode. In the case of reporting remaining oil reserves the main stakeholders are the politicians and planners, although it is the population at large, at the end, that suffer the losses by decisions based on mistaken reporting. If weimagine a doubled loss for over-prediction over under-prediction of the same size, then our analysis provides strong arguments for reporting the mode instead of the mean, but the mode is reasonable also for situations with asymmetry far less than that. This is so both when given alone as a one-point prediction and in conjunction with some measure of uncertainty. preferably a modal interval, just as the (5%, 95%) quantile range goes with the median. The difference between the two is that the modal interval explicitly takes into account the skewness of the distribution. In case of reporting an economic quantity several stages ahead this may be extended to a modal fan chart.

References

Banerjee, A., Guo, X., & Wang, H. (2005) On the optimality of conditional expectation as a Bregman predictor. *IEEE Transactions on Information Theory*, 51(7): 2664-2669.

Basu, S., & Markov, S. (2004). Loss function assumptions in rational expectations tests on financial analysts' earnings forecasts. *Journal of Accounting and Economics*, 38, 171–203

Britton E., Fisher P. & Whitley J. (1998) The Inflation Report projections: understanding the fan chart, Bank of England Quarterly Bulletin, February 1998.

Gneiting, T. (2008) Probabilistic forecasting. *Journal of the Royal Statistical Society, Series A; Statistics in Society*, 171, 319-321.

Gneiting, T. (2011a) Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106, 746–762.

Gneiting, T. (2011b) Quantiles as point forecasts. *International Journal of Forecasting*, 27,197-207.

Granger, C. W. J. (1969) Prediction with a generalized cost of error function. *Operational Research Quarterly*, 20, 199–207.

McCullough, B. D. (2000) Optimal prediction with a general loss function. *Journal of Combinatorics, Information and System Science*, 25, 1-4, 2007-221.

Newey, W. K. & Powell, J.L. (1987) Asymmetric Least Squares Estimation and Testing. *Econometrica*, Vol. 55, No. 4, 819-847.

Patton, A. J., & Timmermann, A. (2007) Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics*, 140, 884–918.

Savage. L.J. (1971) Elicitation of personal probabilities and expectations**.** *Journal of the American Statistical Association,* 66, 783–801.

Whiteman, (1996) Bayesian prediction under asymmetric linear loss forecasting state tax revenues in Iowa. In *Bayesian Inference in Statistics and Econometrics, Essays in honor of Seymour Geisser*, *(Eds. Lee, Johnson & Zellner)*, pp. 149-166

Zellner, A. (1981) Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, 81, 446-451.