Bill Foote

Acting on Bayes

Contents

Beyond regression

Suppose we say that a regression is terrific for description, perhaps for some explanation. But we say further that description and explanation is simply not enough to make any kind of decision. Our transcendent realism will not abide it. Sure we have a way to predict outcomes. But we have a decision to make. Do not our views on outcomes matter? They certainly do!

We use regression modeling as a descriptive tool for studying how an outcome can be predicted given some input variables. A completely different approach is to model a decision outcome as a balancing of goals or utilities or, better yet, beliefs. Here a belief is a justifiable, perhaps through our extensive regression analysis, truth. The belief itself relates to something believed to be. We cross a threshold of rationality, of intelligibility, of something beyond mere observation and sensibility. We act on what we believe to be true.

The values we hold to be true guide us. Following Hildebrand (1953), there are two distinct and subsisting notions of value. One is the merely satisfying, even in an ecstasy of the beauty of a painting, poem, dance, even a deal. Still the experience is one of satisfaction. Consequentialism falls into this notional pit. Economic rational choice theory would have us order our actions according to, essentially, subjective preferences, and actions leading to consequences. We weigh the consequences against our subjectively determined preferences, beliefs. And each of us has these. Others may agree or disagree with us and our preferences. If they agree they might transact with us. We both act on mutually acceptable consequences whatever our beliefs separately. The act might be to transact, exchange, or the act might be to retain the status quo, business as usual, as we walk away from one another, perhaps to transact with others and their subjectively wrought preferences.

Zellner has it!

We can sometimes express our analytical life in the simple terms of doing one of two things D_1, drink from well #1 , or D_2, switch to well #2 . Relative to these two decisions why would we even countenance these alternatives? The simple reason might be one well is contaminated, the other is not. We might not know for sure. If we choose and the water is bad, we, or at least some of us, will be sick, perhaps die eventually. If we choose and the water is good, we should all be just fine and live out our day.

This table replicates Zellner (1971), with some modifications and changes in nomenclature.

	S_1	S_2
Do D_1	0	$L(do\, D_1, S_2)$
Do D_2	$L(do\, D_2, S_1)$	0

In this simple, primitive, table, D_x means decision alternative D_x, and doD_x means act on D_x. The losses incurred when doing, acting on any decision will be given.

How do we choose? At one level we can rationally, reasonably choose one expected level of loss over another, given the data, and the states of the world which might erupt. In our example of water contamination below, the community will be involved in promoting the greater common good. An aspect of this involvement is the principled discussion of decision alternatives in the face of good and/or bad news, uncertainty, and the plausibility of otherwise uncertain outcomes. Also the community will expend resources or not. What the community can review, having become aware of the issues and accepting we, as a community, must act, is to follow a reasonable formation of decision alternatives.

Here is one way to provide a comparison between the two alternatives, where E is our expectation of loss (cost) L. The expected loss of doing $do\, D_x$, we observe data $\mathbf{y}$. We work along the columns of our decision matrix.[1]

$$E(L \mid do\, D_1) = p(S_1 \mid \mathbf{y})L(S_1, do\, D_1) + p(S_2 \mid \mathbf{y})L(S_2, do\, D_1)$$

Because in our simplistic decision matrix, $L(S_1, do\, D_1) = 0$, we have this simplification.

$$E(L \mid do\, D_1) = p(S_2 \mid \mathbf{y})L(S_2, do\, D_1)$$

[1] This does look a lot like Pearl (2016) and his *do*− operator)

Similarly, without any further ado (someone should check this!), we also have for the doing, this expression (no further puns!).

$$E(L \mid do\, D_2) = p(S_1 \mid \mathbf{y})L(S_1, do\, D_2)$$

If we, as a community, believe that, with Polya (1954), Zellner (1971) and Gelman et al. (2004), a reasonable choice of one decision over another is when the community's discernment of expected loss is different between the decisions. Let's compare the results of our Bayesian (we did use posterior probabilities after all!) decision analysis.

We would choose $do\, D_1$ over $do\, D_2$ if this condition holds and remembering we want to **minimize the (expected) loss**, that is, choose the decision with the smallest expected loss.

$$E(L \mid do\, D_1) < E(L \mid do\, D_2)$$

The intuition makes sense if we want to minimize loss when acting, especially on spotty information. The agove relation is the same as saying that this condition then holds.

$$p(S_2 \mid \mathbf{y})L(S_2, do\, D_1) < p(S_1 \mid \mathbf{y})L(S_1, do\, D_2)$$

And these expressions, decision rules, are the same as saying something about the odds.

$$\frac{p(S_1 \mid \mathbf{y})L(S_1, do\, D_2)}{p(S_2 \mid \mathbf{y})L(S_2, do\, D_1)} > 1$$

Let's not let the subscripts fool us. In the numerator we have the expected loss if we were to act on, $do\, D_1$. In the denominator we have the expected loss if we were to act on, $do\, D_2$.

Intuitively reasonable? Mostly, yes, this seems reasonable in the absence of any other consideration. Here is a motivating example.

Example 1. ______________________________

Arsenic in Bangladesh

Gelman et al. (2004) examine the problem of making policy for avoiding arsenic contamination in water wells in Bangladesh. How can we understand the relation between distance between wells, arsenic level, and the decision to switch? It makes some sense that people with higher arsenic levels would be more likely to switch. The policy decision they model is a recommendation to switch to another well and / or drill new wells.

The actual health risk is believed to be related to arsenic concentration.

To set up a discrete choice model, we specify a value function, which represents the strength of preference for one decision over the other — in this case, the preference for switching, D_1, as compared to not switching, D_2. The value function can be scaled so that zero no loss from making a decision, while a positive value represents the loss from making the decision.[2]

What is new about this analysis is the deliberate sidestepping of a purely statistical hypothesis of inference. Instead we directly model decisions instead of hypotheses, confront the decisions with data mediated by a model which the decision maker believes in, all to yield the plausibility of a range of decision alternatives. This approach follows and builds on a decades-old multi-action-multi-state loss model reminiscent of contingency tables, strategy, and games.

Example 2.

Based on Zellner (1971), p. 317 and informed by Gelman et al. (2004)

It costs resources, financial, human, materials, to build, maintain, test wells drilled for potable water. We have 10 observations of contamination levels from a Gaussian population of wells with unknown mean contamination level μ and standard deviation $\sigma = 1$. The mean contamination levels (states of the world) are $S_1 = 2.0$ and $S_2 = 1.0$, and the prior probabilities of $D_1 = drink$ the water from the wells and $D_2 = switch$ to another set of wells we know are probably not contaminated, are 1/2 each. If the wells are not contaminated, and we decided to drink from the wells, there will be no extra cost of drinking water. If the wells are contaminated, and we decided to switch to other (perhaps more reliable) wells, we will assume again there will no extra cost of resources. Otherwise if the wells are not contaminated, and we decided to switch to other the wells, there will at least be a cost to switch to other sources of potable water equal to 2 (pick your currency units and scale; millions of Bolivars, for example). Also, if the wells are contaminated, and we decided not to switch to other the wells and drink the water, there will be a cost to care for the sick and dying, as well as search for other sources of potable water equal to 4 (again pick your currency units and scale; millions of Bolivars, for example).

Quite a problem to solve, a puzzle to figure out, but there may be a bit of a mystery to wrangle over when we exercise our will collectively, or even individually, to actual drink or switch. Our actions will necessarily precede the potentials of outcomes.

What is the common good, the greater good? The goods here are the provision by a community of potable, safe, reliably so, water. Water is needed for life, so

[2] This model is similar to the **latent-data interpretation** of logistic regression.

the principle we might apply is a version of the primacy of life. Life envelopes the community. The community exists to promote life. We have become aware as a community of a problem of contamination. We now accept we must do something, as a community, and we will suffer costs and anxiety. So be it. Doing so will promote life, and at the very least we will learn from inferring the consequences of our actions.[3]

One solution should at least proceed from these considerations.

```
#library( tidyverse )
#library( rethinking )
n <- 10
y <- rnorm( n, 2.5, 1)
d <- tibble(
  y = y
)
mu <- seq( from=0, to=3, by=0.1 )
sigma <- 1
prior <- rep( 1, length(mu) ) #equally likely
# we sample to get an idea
likelihood <- dnorm( mu, sigma  )
likelihood_prior <- likelihood * prior
posterior <- likelihood_prior / sum( likelihood_prior )
precis( posterior, hist=FALSE )
```

```
##                 mean         sd        5.5%      94.5%
## posterior 0.03225806 0.01356694 0.008916462 0.04764759
```

```
Pr_1_y <- posterior[ mu==1 ] #under the low contaminant S_1
Pr_2_y <- posterior[ mu==2 ] #under the high contaminant S_2
EL_1 <- Pr_2_y * 4
EL_2 <- Pr_1_y * 2
ifelse( EL_1 < EL_2, "drink from wells", "switch to other wells")
```

```
## [1] "switch to other wells"
```

The answer seems to be to switch to other wells, given the 10 observations of well water.

Or we can perform this analysis.

```
m_wells <- quap( alist(
  y ~ dnorm( mu, 1),
  mu <- a,
  a ~ dnorm( 5, 1 ),
  sigma <- 1
```

[3]Nussbaum (2011) on Harsanyi (1977) for a Kantian-Aristotelian analogy and Hirschfeld (2018) for a Thomistic-Transcendental Realist approach.

```
), data = d )
precis( m_wells, hist=FALSE )
```

```
##      mean        sd     5.5%    94.5%
## a 2.90351 0.3015113 2.421637 3.385383
```

```
Pr_1_2 <- dnorm(2.58, 2, 1)
```

Rules to think by

Zellner (1971) espouses a Bayesian analysis of inferential deductive logic for hypotheses as well as for decisions. He follows Jeffreys (1966) and his nine rules of inductive inference. We can restate them here using the epistemological language of Lonergan (1957) and his canons of general empirical method.[4]

Rule 1. "All hypotheses used may be explicitly stated, and the conclusions must follow from the hypotheses." (Jeffreys (1966), p. 8) A complete set of hypotheses, decision alternatives, must be explicit to the analyst and the consumer of the decision maker. But further, these hypotheses must deductively result in a "complete explanation" (Lonergan (1957), p. 107-109) in the sense that for a universe of discourse carved out by the analyst and decision maker, the hypotheses must be able to result in a comprehensive set of conclusions. One imagines the wrangle over the logically inconsistent use of *modus tollens* to base inferences with the Null Hypothesis Significance Test framework. If *modus tollens* is IF A, THEN B, AND NOT B, THEN NOT A then our deduction that NOT A is valid. However if we deny the antecedent and argue IF A, THEN B, AND NOT A, THEN NOT B we deduce inconsistently by denying the antecedent.[5]

Example 3.

If burglars, who do not have a key, entered by the front door (A the antecedent), then they must have forced the lock, which requires a key (B the consequent).

Now suppose we have data which indicates that **the lock was not forced, $\sim B$, denying the consequent**. We can deduce from this data, namely, that the burglars did not force the lock, $\sim B$, that thus they also did not enter by the front door, $\sim A$. This is ***modus tollens*** or **denying the consequent**.

If the data we have is that the burglars did not enter by the front door, $\sim A$ denying the antecedent, then it is illogical to deduce that they did not force the lock. $\sim B$. This is the fallacy of **denying**

[4] The canons are listed, related to one another, and explicated in *Insight*, chapter 3. Therein is a thematic approach to the inference rules set out by Jeffreys, Jaynes and Zellner.

[5] See this note in the *Oxford Reference* site: https://www.oxfordreference.com/display/10.1093/oi/authority.20110803095711627

the antecedent. They might have forced the lock, was startled by the owners' attack cat, and decided to enter the house by another route.

Rule 2. "The theory must be self-consistent; that is, it must not be possible to derive contradictory conclusions from the postulates and any given set of observational data." (Jeffreys (1966), p. 5) This rule works in tandem with the next rule.

Rule 3. "Any rule must be applicable in practice. A definition is useless unless the thing defined can be recognized in terms of the definition when it occurs. The existence of a thing or the estimate of a quantity must not involve an impossible experiment." (Jeffreys (1966), p. 5) Here are Lonergan (1957)'s Canons of Selection, Relevance and Operations at work. We can unpack this rule by a three-way division of labor. But we will only stop by the Canon of Relevance here. We defer to Rule 5 for Canons of Selection and Operations, not that they do not apply here! There is strong connection between Rules 3 and 5 bridged by Operations and Selection within the context of Operations. Both of these point to Relevance, here. In fact it seems that Relevance and Operations are flip-sides of the same coin.

Relevant data, propositions, criteria for choosing one intelligible pattern rather than another, one decision alternative than another, all presuppose that data can be applied to different sets of hypotheses, inform different species of decisions, each require different sets of criteria. Data are raw and that flexible. Using statistics as derived data can be even more problematic. Averages of transactions can wash out very high and very low levels of measures, both informative in their extremes. Perhaps they might be the only relevant part of a data set, whereas the averages tell us nothing much interesting about the pattern of the data we call intelligibility. And thus when applied to a decision the analysis is irrelevant, plausibly. A most important part of the Canon of Relevance is that it views data as emanating from beings as they relate to one another, not to our measurements, the instruments by which we measure and record, even to the observers. This canon is a first step to a principled notion of objectivity. There is are immanent patterns latent in all data which has nothing to do with how the analysis is run, why we are analyzing something in the first case, even what are the materials and resources involved in the analysis. These immanent patterns in sensible data we are collecting are part and parcel of the formal causality of the analysis. Nascent, inchoate somewhat, yes, but immanently and thus relevantly critical to our analysis mission.

Rule 4. "The theory must provide explicitly for the possibility that inference made by it may turn out to be wrong." (Jeffreys (1966), p. 6) Jeffreys goes on to state, categorically, we might be wrong do to error, incomplete and evolving data, and an attitude of never allowing for revision. On the same page he goes on to say "... [W]e have a certain amount of confidence that it will be right in any particular case, though this confidence does not amount to logical certainty." (*ibid.*) Lonergan (1957) agrees and uses this notion to buttress an uncertainty principle in our analytical work. "[While i]t is true enough that data are hazy, that measurements are not perfectly accurate, that measuring can distort the measured object ... [, yet one] can affirm them [and] continue to misconceive classical laws[, such falling bodies.]" (p. 125) How do we characterize the ways of affirming the indeterminant nature of laws in the concrete? Not imaginatively, but as the "indeterminancy of the abstract." (p. 125)

Rule 5. "The theory must not deny any empirical proposition *a priori*; any precisely stated empirical proposition must be formally capable of being accepted, in the sense of the last rule [4], given a moderate amount of relevant evidence." (Jeffreys (1966), p. 6) Yes, only relevant propositions are allowed. The only relevant propositions are those that have some sort of chance of surviving the rigors of the Canon of Selection (Lonergan (1957), p. 94-97) and the Canon of Operations (p. 97-99) implicit in our formal approach to finding patterns, intelligibility in otherwise unintelligible data. Selection would restrict us to sensible data, data we can observe as existing outside of our mental images of the data in our minds. This canon and here Rule 5 restrict us to propositions against which we can apply the restricted sensible data. If we cannot do this, we cannot possibly search for intelligible anything in objective reality, namely the pattern in the data susceptible to description, explanation, and perhaps prediction. Operations is a many-headed hydra.

Operations would help us expand our consciousness about data using the cumulative font of previous work. This means we use prior research to inform answers to current questions, but it must be relevant. Operations also helps us construct mentally, for all analysis is a figment of the imagination and the intellect. Constructions are cumulative and so are they verified. We keep what works and throw the rest into a bin for reuse. In this way they are also systematic in that different pieces of constructions provide integrity, harmony, and even clarity to any analysis. And if they do not, then they go into the bin for reuse, repurposing. With operational constructions we understand and have a history of previous endeavors, how well they did or did not work, what we can use or discard from previous analyses. In a word or two, operations allow us to transcend our current state of operations and envision a different, hopefully better by some rubric, way to proceed. All of this is the answer to a question about the shape or form of the analysis.

These five rules are considered by many, including Jeffreys and Zellner, to be *essential*. Jeffreys also indicates three more useful rules.

Rule 6. "The number of postulates should be reduced to a minimum." (Jeffreys (1966), p. 6) This is often stated as a variant of William of Occam's Razor. What William actually stated was apparently "Numquam pluralitas non est ponenda sine necessitate." (Any plurality is not to be posited without necessity.) This principle is a component of the Canon of Parsimony for Lonergan (1957), p.102, wherein any scientist must eliminate any statement which is not verifiable and include only those statements which, currently, are. On the other hand, Lonergan's Canon of Relevance (p. 103) also requires the albeit parsimonious inclusion of any insights which add to the raw data. While laws of nature might in a particular experiment seem to fail, repeated experiments should be able to come to a statistical law which can be affirmed as verified. The key is to understand the difference between a law and an event. Events can deviate from laws (trends, hypotheses, models), but to be verifable they cannot deviate in any systematic way. If they do the other side of parsimony directs us to consider including the perhaps newly discovered deviatoin into a new formulation of the law, trend, or model. [6]

Rule 7. "While we do not regard the human mind as a perfect reasoner, we must accept it as the only one available. The theory need not represent actual thought processes in detail, but should agree with them in outline."" (Jeffreys (1966), p. 9) As with Lonergan (1957), p. 124-125, our brilliant ideas are abstractions from practical reality. Thus they can only be further explained and used to predict when yoked with the concrete circumstances of new events as they unfold. Thus also more work by Jeffreys' imperfect human reasoner is always needed to determine further insights mashed together with concrete situations. As if this is not enough, Lonergan goes on to imply if a new situation arises, then unsystematic deviations might need either to be incorporated into new models or else discarded as residues, at least relative to the abstraction called a model. But one last point is that if the abstract understanding called a model begins to degenerate, there is the "inverse insight" (p. 125) that this degeneration is also intelligible and must be reported as a claim. Thus this rule seems to be a species of Lonergans Canon of Statistical Residues. But it also seems to require another Canon of Relevance to "fix our attention on what insight adds to data." (p. 125)

Something must be said about the imperfect human reasoner. And this will not be a foray as much as into psychology as it will be into epistemology, the study of how it is we know anything at all. The object is all being, asking the question *what is it?* presupposing perhaps but at the least alongside the more basic question *is it?* and all of this incompletely (we do not know all the rules of this road) and imperfectly (we will at times erroneous apply whatever rules we seem to know). The first question is *formal causality*, what is the nature of, blueprint for, scaffolding, framework, approach we will use for our analysis. The second question is about the existence of the object of our various desires to understand. But it raises a question of *final causality* called *why bother?* or *what's the purpose?* the end we would like to achieve on this mission, should we accept it. Both of these raise two more questions of

[6] This is an often used principle of logic and selection of the minimal set of hypotheses needed to answer a question. St. Thomas Aquinas uses this principle to set up an objection to the existence of God, which he roundly refutes by the superiority of causal over simple explanations of anything. *Summa Theologiae*, q.3, a.3, ad 2.

material causality, the what's in the process, the people, technology platforms, resources, materials, inputs, and *efficient causality*, the how the process of inputs, activities, technology, know-how, conventional wisdom, yielding desired outputs. Let's apply this causal heuristic to get a thumbnail sketch of the imperfect human reasoner.

- The reasoner, or community of reasoners for that matter, are *aware* (not not, remember the adjective imperfect) of raw data, use operations to observe and record the data, or at least somehow remember the data.
- The reasoner then *understands* the data by throwing provisional hypotheses like darts onto a known dart board. There is a center on this dart board, the goal, the finality of the understanding. Some darts stick (they are consistent with the data), others eventually will fall off to the floor. Whatever is left is the pattern of data we might call the beginnings of an understanding.
- While there are several pointed projectiles on the dartboard only those close enough to the center are very plausible, others not so. We can even measure the distances of darts to the center to quantify our analysis. We might say at this point that we can make a first judgment based on the *is it?* question of *yes it is.*

There is a pattern in the data. This is what we will mean by *systematic*. It might also answer something about the nature of the data, the *what is it?*. Attributes like long-lived, very complicated, and it is blue by the way can be verified on the dartboard. We have verified in data, using a reasonable approach to belief in plausibility, called a distance. We can even add up all of the distances to normalize the distance in a metric we call a probability. Quite a leap and all part of the answer to the question of *what is it?* and the blueprint we are using to answer questions at all.

Rule 8. "In view of the great complexity of induction, we cannot help to develop it more thoroughly than deduction." (Jeffreys (1966), p. 10) And thus the role of the use of the logic of pure mathematics. But if we use pure mathematics in this logical way, then the notion of probability is not that of empirical frequencies as we usually view them in elementary statistics courses, rather the ordered degrees of reasonable, rational, belief about a claim. This means also that any inductively constructed claim should emanate from a deductively valid logic. One builds on the other. One verifies the other in concrete events. One provides a deduced ordering of hypotheses when hypotheses are confronted by data. There is thus some sort of *uncertainty principle* lurking here as Lonergan points out: "For the concrete includes a nonsystematic component, and so the concrete cannot be deduced in its full determinancy from any set of systematic premises." (Lonergan (1957), p. 123)

Upshots

What is the upshot, if any, from all of this dense verbiage?

- Statistical thinking and deciding is not for the faint-hearted. A leap into a heuristic is required. But such a heuristic is based on how we know and decide at all. Above all we need a measure of belief we can use in a deduction about the consistency of data with the abstraction called a model.
- Every analysis should begin with a four cause charter. Why are we doing this analysis? (The purpose must be embedded in all aspects of our analysis especially the criteria we apply.) What is the nature of this analysis? (We recall the impassioned plea of the Canon of Relevance and Rule 3.) What materials, resources, platforms do we need? How do we go from point A to point B in the process of constructing a higher viewpoint for the consumer of the analysis? (Operations, Relevance)
- Indeterminacy and uncertainty: All models are abstractions from reality. When the rubber hits the road, are all plans are thrown out? No just relegated to the recycle bin. Thus prepare to minimize our maximum analytical grief. With indeterminacy, this grief will evolve differently for different people.
- Getting to a complete explanation, but updating when new data, new constructions, new criteria inevitably occur.

Here is the beginning of a way to incorporate many of the currents throughout the Rules and Canons as we construct and analyze not only hypotheses, but also their analogical correlate decision alternatives.

Loss, and hopefully found

We look for clean water W, also find land L in the process. Without water, people will die. Some land is also needed to sustain life, absorb water, and so on. Suppose we know that the amount of water relative to land to keep 100 (multiply by the correct scaling factor to render reality) people alive is 70% of the overall mass on which people are located, the rest, 30%, land to live on.

We have data which suggests that we found in 10 experimental probes of our joint water/land mass, that 6 probes found water and 4 found land. We want to safeguard life so we propose that given the risk (probability) of not finding *enough* water it is possible people will perish. If we find too much water, there will be not enough room for people to live, as apparently they do not have the technology (yet) to live extensively on water. Although some very few folks seem to, at times. An we have no idea at all before we send water and land sensitive probes what the proportion of water is.

Given all of this we first determine the probabilities of various proportions of water, p. A grid will help us specify these various proportions. A binomially distributed observation model will help us compute the likelihood of a proposed proportion given the probe data, namely 6 water and 4 land for a total of 10 tries.

```
W <- 6
n <- 10
L <- n - W
grid_out <- 1000
p_grid <- seq( from=0, to=1, length.out=grid_out)
```

We just generated 1000 equally spaced proposed proportions p from zero water and all land, $p = 0.00$, to no water at all, all land $p = 1.00$, all with the `seq()` function. With this specification we **completely covered** the space of possible proportions, realizing we can zoom into each equally spaced interval, should we ever want to.

Next we combine the data W and n with the proposed proportions p using a **relevant** plausibility generator known as the **binomial distribution.** It is relevant since it is the distribution which can analyze binary data, such as water and land, in multiple sampling probes. We assume we know nothing at all about the plausibility of water on this mass of land and water, so that our prior expectation of a proportion is quite vague, just equally likely, even before we apply data, nay, before we sample data.

```
# we have no idea
prior <- rep( 1, grid_out )
# we sample to get an idea
likelihood <- dbinom( W, n, prob=p_grid )
likelihood_prior <- likelihood * prior
posterior <- likelihood_prior / sum( likelihood_prior )
```

A little exploration of the posterior's impact on proportions needs this sampling of the `p_grid`, with replacement, of course.

```
n_samples <- 1e6
samples <- sample( p_grid, prob=posterior, n_samples, replace=TRUE )
lower <- 0.05; upper <- 0.95
sum( samples >= lower & samples < upper ) / n_samples
```

```
## [1] 0.999899
```

```
quantile( samples, c( 0.10, 0.90 ))
```

```
##       10%       90%
## 0.4004004 0.7597598
```

And the action is mostly between the lower and upper bound coded here. A little bit of zoom-in shows that their is a 80% (90% - 10%) probability that the proportions are between a low of about 0.40 and a high almost of 0.76. We remember that our goal is 70%; more than this target we do not have enough land, less than this target we do not have enough water. It does not look too promising for some people?

We have a decision to make

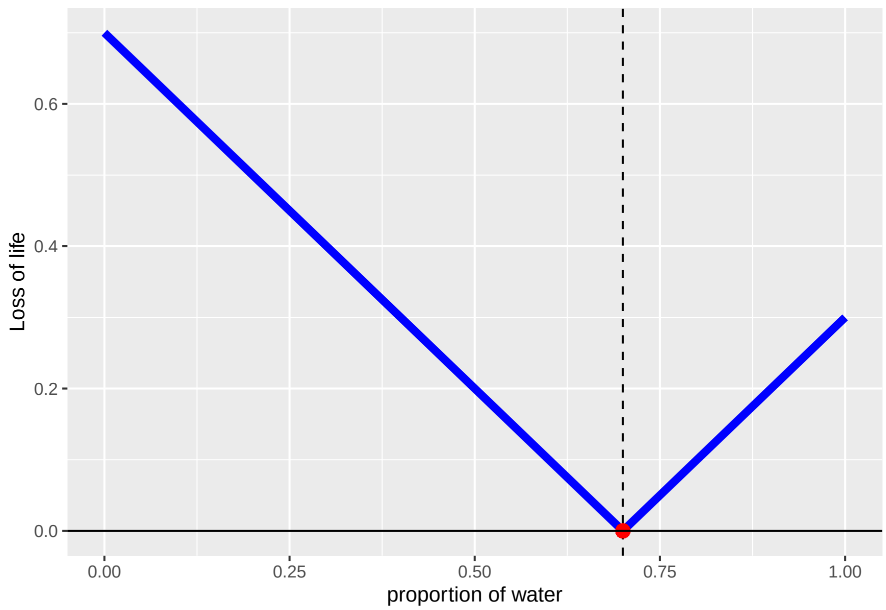
We suppose that we will lose people at a constant rate proportional to the distance our actual proportion of water is from the target of 70%, in both directions! One side is too much land and too little water ($p < 0.70$), the other is too much water and too little land ($p > 0.70$). A plot tells the story.

To illustrate these ideas suppose our data is all positive (ratio data in fact). If $p^* = 0.70$ then the loss function is the aabasolute value of the distance between actual p and target p^*. In our very simplistic example, the loss is both symmetric, and that the rate is unitary.

$$L(p, p^*) = |p - p^*| \tag{1}$$

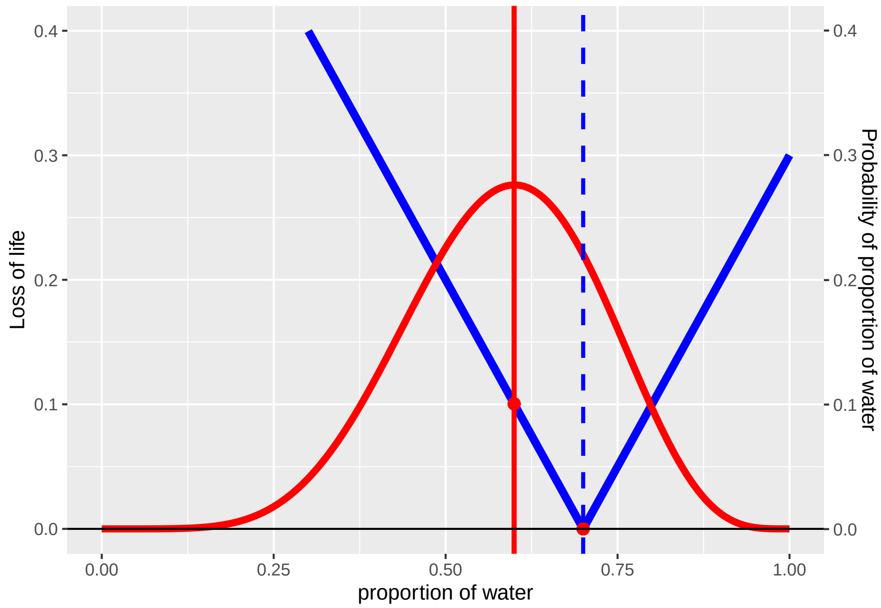
Our loss function has this appearance, the so-called check function.

```
p_star <- 0.70
alpha <- 1.0
#X <- seq(0, 1, length.out = 100)
loss <- alpha * abs(p_grid - p_star)
Lp <- tibble(
  p = p_grid,
  Loss = loss,
  posterior = posterior
  )
p <- Lp |>
  ggplot(aes( x = p, y = Loss )) +
  geom_line(color = "blue", size = 2.0) +
  geom_vline(xintercept = p_star, linetype = "dashed") +
  geom_point(x = p_star, y = 0, color = "red",  size = 3) +
  geom_hline(yintercept = 0.0) +
  xlab("proportion of water") + ylab("Loss of life")
p #plotly::ggplotly(p)
```



Now let's see what happens when we superimpose the posterior distribution of proposed proportions onto the loss plot.

```
max_post_p <- p_grid[ which.max( posterior ) ]
p <- Lp |>
  ggplot(aes( x=p, y=Loss )) +
  geom_line( color="blue", size = 2.0 ) +
  geom_line( aes( x=p, y=posterior*100 ), color="red", size=1.75) +
  geom_vline( xintercept=p_star, linetype="dashed", color="blue", size=1.0 ) +
  geom_point( x=p_star, y=0, color="red",  size=2.5 ) +
  geom_point( x=max_post_p, y=p_star-max_post_p, color="red", size=2.5 ) +
  geom_vline( xintercept=max_post_p, color="red", size=1.25 ) +
  geom_hline( yintercept=0.0 ) +
  xlab( "proportion of water" ) + ylab( "Loss of life" ) +
  scale_y_continuous(
    # Features of the first axis
    name = "Loss of life",
    limits = c( 0, 0.4 ),
    # Add a second axis and specify its features
    sec.axis = sec_axis(~.*1, name="Probability of proportion of water")
  )
p
```



Well that loss function certainly does not line up with the median, the most likely *a posteriori* probability! We will have a loss! So the question becomes can we intervene somehow to minimize the loss? We must have a lever to pull. This lever might be something that the proportion of water could depend on. The search continues!

Ground control, we have a problem

Now to a completely made-up model. Let's suppose we have the following generative model of independent (and controllable) x_t somehow influencing dependent variable y_t. The t subscript is just a sequential index.

$$y_t \sim \text{Normal}(\mu_y, \sigma_y^2), \tag{2}$$
$$\mu_y = \alpha + \beta x_t, \tag{3}$$
$$x_t \sim \text{Normal}(0, 1), \tag{4}$$
$$\alpha \sim \text{Normal}(1, 1), \tag{5}$$
$$\beta \sim \text{Normal}(2, 1), \tag{6}$$
$$\sigma_y \sim \text{Exponential}(1) \tag{7}$$

So far, just another day of simple linear regression.

```
n <- 1000
alpha <- rnorm( n, 1, 1)
beta <- rnorm( n, 2, 1)
x <- rnorm( n, 0, 1 )
y <- alpha + beta*x
d <- tibble(
  x = x,
  y = y
)
d |>
  summary()
```

```
##        x                  y
##  Min.   :-3.35971   Min.   :-9.5856
##  1st Qu.:-0.71286   1st Qu.:-0.4737
##  Median :-0.02985   Median : 0.9040
##  Mean   :-0.04074   Mean   : 0.9057
##  3rd Qu.: 0.59728   3rd Qu.: 2.2835
##  Max.   : 3.39166   Max.   : 9.5554
```

We then run the quadratic approximation estimation of the model, just reversing what we already know.

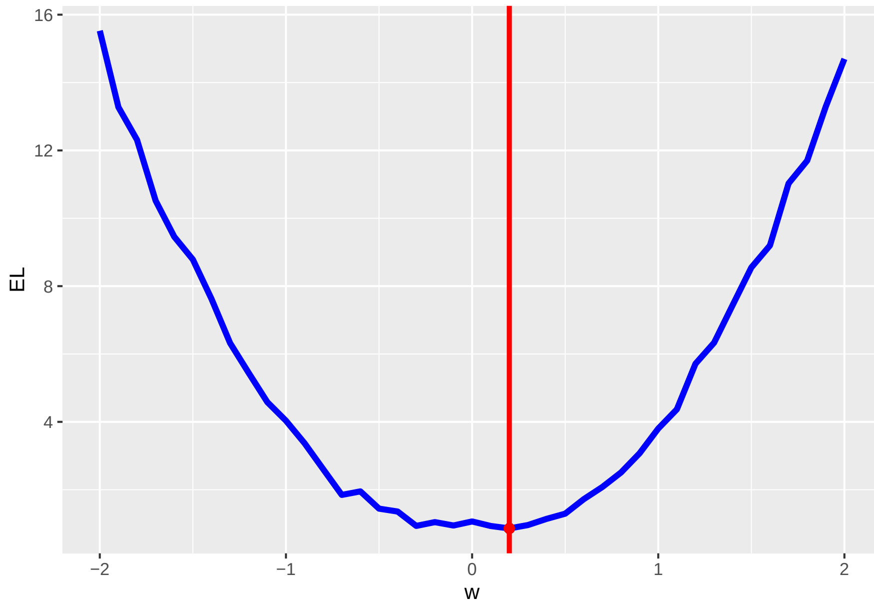
```
#library(rethinking)
m_0 <- quap(
  alist(
    y ~ dnorm( mu_y, sigma_y ),
    mu_y <-  alpha + beta*x,
    alpha ~ dnorm( 0, 1),
    beta ~ dnorm( 0, 1),
    sigma_y ~ dexp( 1 )
  ), data = d
)
precis( m_0, hist=FALSE )
```

```
##              mean         sd      5.5%    94.5%
## alpha   0.9827609 0.04500962 0.9108269 1.054695
## beta    1.9401743 0.04592874 1.8667713 2.013577
## sigma_y 1.4235410 0.03179769 1.3727222 1.474360
```

Nicht Neues here. The model estimate parameter values very close to the one's we generated our data with.

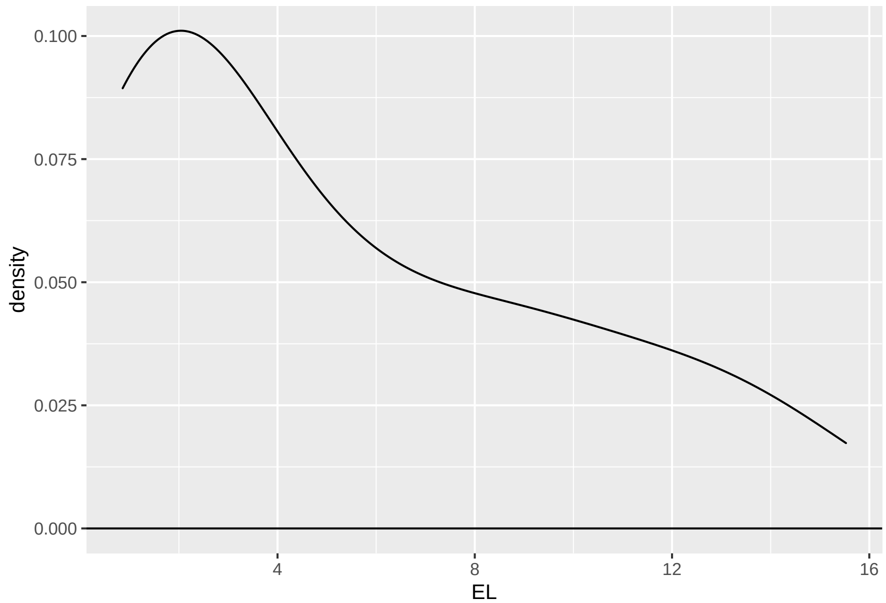
With all this talk of generated data, let's generate **predicted values** of y calling these z using the same model m_0 but with a set of projected x values we will call w. The `rethinking::sim()` function will produce (trying not to over-use generate!) as many columns as there are elements of w we project, perhaps 41. For each column `sim()` simulates values of z, perhaps 1000.

```
w <- seq( from=-2, to=2, by=0.1)
z <- sim( m_0, data=list( x=w ))
a <- 1 #target z
L <- (z-a)^2 #Loss
L_median <- apply( L, 2, median ) #expected Loss
d_L <- tibble(
  w = w,
  EL = L_median
)
L_min <- min( L_median )
w_star <- w[ which.min( L_median ) ] #min_w EL
p <- d_L |>
  ggplot( aes( x=w, y=EL) ) +
  geom_line( color = "blue", size = 1.5 ) +
  geom_point( aes( x=w_star, y=L_min ), color="red", size=2.00 ) +
  geom_vline( xintercept=w_star, color="red", size=1.25 ) #+
  #geom_histogram( aes( x=EL )) +
  #scale_y_continuous(
    # Features of the first axis
    #name = "Loss of life",
    #limits = c( 0, 0.4 ),
    # Add a second axis and specify its features
    #sec.axis = sec_axis(~.*0.001, name="Probability of of clean water")
  #)
p
```



And now the distribution of the expected loss, $E(L(z \mid y, w))$, just for comparison.

```
#under consruction
d_L |>
  ggplot( aes( x=EL ) ) +
  geom_density() +
  geom_hline( yintercept=0 )
```




```
library(MKmisc)
summ_EL <- d_L |>
  summarize(
    mean = mean( EL ),
    mad = sum( abs(EL - mean(EL)))/length(EL),
    Q25 = quantile( EL, 0.25),
    Q50 = quantile( EL, 0.50),
    Q75 = quantile( EL, 0.75),
    IQR = quantile( EL, 0.75) - quantile( EL, 0.25),
    tail = max( EL ) - quantile( EL, 0.75),
  )
summ_EL
```

This is definitely not a Gaussian distribution, more like an exponential mixed with a Gamma. We know that this is what a Generalized Pareto Distribution tends to look like. Extreme values seem important. The median of the medians is xxx. The IQR is xxx. The difference between the tail and the IQR is xxx.

We now have something of an answer to the question of what value of the independent and controllable $w = x$ will minimize the loss which results from failing to achieve the target a.[7] We also now have a principled way to begin to challenge an intervention.

As if this is not enough let's superimpose $\Pr(z \mid y, w^*)$ onto the simulated loss

[7]This is known as a stochastic optimal **Linear-Quadratic-Predictor (LQP)** problem.

function. We repeat the entire workflow here to embed it in our analytical memories.

```
w <- seq( from=-2, to=2, by=0.1)
z <- sim( m_0, data=list( x=w ))
a <- 1 #target z
L <- (z-a)^2 #Loss
L_median <- apply( L, 2, median ) #expected Loss
d_L <- tibble(
  w = w,
  EL = L_median
)
L_min <- min( L_median )
w_star <- w[ which.min( L_median ) ] #min_w EL
p <- d_L |>
  ggplot( aes( x=w, y=EL) ) +
  geom_line( color = "blue", size = 1.5 ) +
  geom_point( aes( x=w_star, y=L_min ), color="red", size=2.00 ) +
  geom_vline( xintercept=w_star, color="red", size=1.25 )
```

xxx]

What could possibly be next?

Was that enough? For now at least. We accomplished a lot. But perhaps the biggest takeaway is our new found ability to deploy simulation with probability with compatibility of hypotheses with data to arrive at learning about the data. We learned because we inferred. But then we extended this capability to the realm of decision making itself.

References

Gelman, Andrew, Matilde Trevisani, Hao Lu, and Alexander Van Geen. 2004. "Direct Data Manipulation for Local Decision Analysis as Applied to the Problem of Arsenic in Drinking Water from Tube Wells in Bangladesh." *Risk Analysis* 24 (6): 1597–1612. https://doi.org/https://doi.org/10.1111/j.0272-4332.2004.00553.x.

Harsanyi, John C. 1977. "Morality and the Theory of Rational Behaviour." *Social Research* 44 (4).

Hildebrand, Dietrich Von. 1953. *Christian Ethics*. McKay.

Hirschfeld, Mary L. 2018. *Aquinas and the Market: Toward a Humane Economy*. Harvard University Press.

Jeffreys, Harold. 1966. *Theory of Probability*. 3rd ed. Oxford: The Clarendon Press.

Lonergan, Bernard J. F. 1957. *Insight: A Study of Human Understanding*. Philosophical Library.

Nussbaum, Martha C. 2011. *Creating Capbilities: A Human Development Approach*. Belknap, Harvard.

Pearl, Judea. 2016. *Causal Inference: A Primer*. Wiley.

Polya, George. 1954. *Mathemtics and Plausible Reasoning*. Princeton, N.J.: Princeton University Press.

Zellner, Arnold. 1971. *An Introduction to Bayesian Inference in Econometrics*. John Wiley; Sons, Inc.