On the Difference between Binary Prediction and True Exposure

With Implications For Forecasting Tournaments and Prediction Markets

N. N. Taleb and P. E. Tetlock

June 2013

Abstract

There are serious differences between predictions, bets, and exposures that have a yes/no type of payoff, the "binaries", and those that have varying payoffs, which we call the "vanilla". Real world exposures tend to belong to the vanilla category, and are poorly captured by binaries. Vanilla exposures are sensitive to Black Swan effects, model errors, and prediction problems, while the binaries are largely immune to them. The binaries are mathematically tractable, while the vanilla are much less so. Hedging vanilla exposures with binary bets can be disastrous--and because of the human tendency to engage in attribute substitution when confronted by difficult questions, decision-makers and researchers often confuse the vanilla for the binary.

Binary vs Vanilla Predictions

Binary: Binary predictions are about well defined discrete events, such as whether a person will win the election, a single individual will die, a team will win a contest. We call them binary because the outcome is either 0 (the event does not take place) or 1 (the event took place). You cannot have five hundred people winning a presidential election.

Vanilla: "Vanilla" predictions, also known as natural exposures, correspond to situations in which the payoff is continuous and can take several values. The "vanilla" designation comes from option exposures that are open-ended as opposed to the binary ones that are called "exotic". The designation "vanilla" is fitting outside option trading because the exposures they designate are naturally occurring continuous variables, as opposed to the binary that which tend to involve abrupt institution-mandated discontinuities. The vanilla add a layer of complication: profits for companies or deaths due to terrorism or war can take many, many potential values. You can predict the company will be "profitable", but the profit could be $1 or 10 billion.

Most of the variables affecting the vanilla are not bounded, or have a remote boundary. So the prediction of the vanilla is marred by Black Swan effects. A few prescient observers saw the potential for war among the Great Power of Europe in the early 20th century but virtually everyone missed the second dimension: that the war would wind up killing an unprecedented twenty million persons, setting the stage for both Soviet communism and German fascism and a war that would claim an additional 60 million, followed by a nuclear arms race from 1945 to the present, which might some day claim 600 million lives.

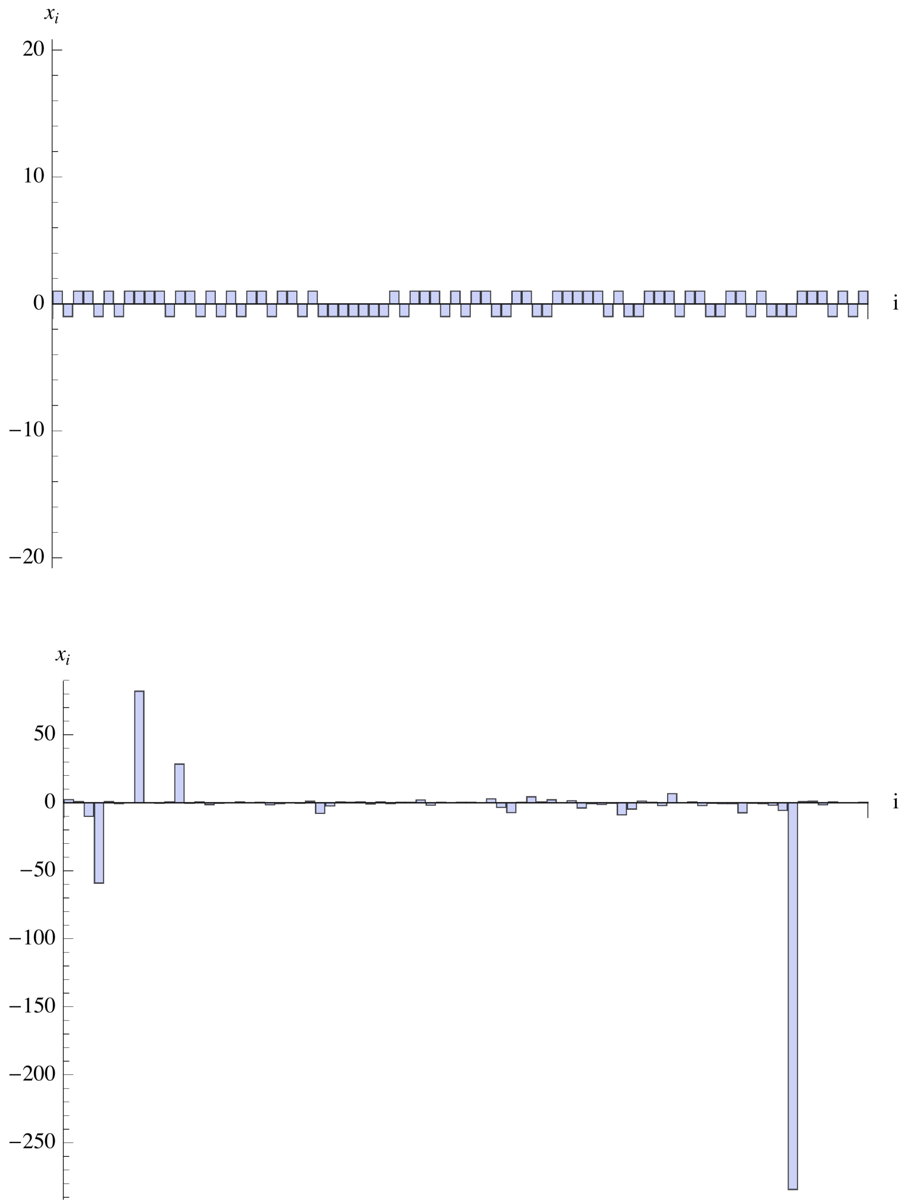

Figure 1 Comparing digital payoff to the vanilla. The vertical payoff shows x_i (x_1, x_2, ...) and the horizontal shows the index *i= (1,2,...)*, as *i* can be time, or any other form of classification. We assume in the first case payoffs of {-1,1}, and open-ended (or with a very remote and unknown bounds) in the second.

The Black Swan is Not About Probability But Payoff

In short, the vanilla has another dimension, the payoff, in addition to the probability, while the binary is limited to the probability. Ignoring this additional dimension is equivalent to living in a 3-D world but discussing it as if it were 2-D, promoting the illusion to all who will listen that such an analysis captures all worth capturing.

Now the Black Swan problem has been misunderstood. We are saying neither that there must be more volatility in our complexified world nor that there must be more outliers. Indeed, we may well have fewer such events but it has been shown that, under the mechanisms of “fat tails”, their “impact” gets larger and larger and more and more unpredictable. The main cause is globalization and the spread of winner-take-all effects across variables (just think of the Google effect), as well as effect of the increased physical and electronic connectivity in the world, causing the weakening of “island effect” a well established fact in ecology by which isolated areas tend to have more varieties of species per square meter than larger ones. In addition, while physical events such as earthquakes and tsunamis may not have changed much in incidence and severity over the last 65 million years (when the dominant species on our planet, the dinosaurs, had a very bad day), their effect is compounded by interconnectivity.

So there are two points here.

Binary predictions are more tractable than exposures

First, binary predictions tend to work; we can learn to be pretty good at making them (at least on short timescales and with rapid accuracy feedback that teaches us how to distinguish signals from noise — all possible in forecasting tournaments as well as in electoral forecasting — see Silver, 2012). Further, these are mathematically tractable: your worst mistake is bounded, since probability is defined on the interval between 0 and 1. But the applications of these binaries tend to be restricted to manmade things, such as the world of games (the “ludic” domain).

It is important to note that, ironically, not only do Black Swan effects not impact the binaries, but they even make them more mathematically tractable, as will see further down.

Binary predictions are often taken as a substitute for vanilla ones

Second, most non-decision makers tend to confuse the binary and the vanilla. And well-intentioned efforts to improve performance in binary prediction tasks can have the unintended consequence of rendering us oblivious to catastrophic vanilla exposure.

The confusion can be traced to attribute substitution and the widespread tendency to replace difficult-to-answer questions with much-easier-to-answer ones. For instance, the extremely-difficult-to-answer question might be whether China and the USA are on an historical trajectory toward a rising-power/hegemon confrontation with the potential to claim far more lives than the most violent war thus far waged (say 10X more the 60M who died in World War II). The much-easier-binary-replacement questions —the sorts of questions likely to pop up in forecasting tournaments or prediction markets — might be whether the Chinese military kills more than 10 Vietnamese in the South China Sea or 10 Japanese in the East China Sea in the next 12 months or whether China publicly announces that it is restricting North Korean banking access to foreign currency in the next 6 months.

The nub of the conceptual confusion is that although predictions and payoffs are completely separate mathematically, both the general public and researchers are under constant attribute-substitution temptation of using answers to binary questions as substitutes for exposure to vanilla risks.

We often observe such attribute substitution in financial hedging strategies. For instance, Morgan Stanley correctly predicted the onset of a subprime crisis, but they had a binary hedge and ended up losing billions as the crisis ended up much deeper than predicted (*Bloomberg Magazine*, March 27, 2008).

Or, consider the performance of the best forecasters in geopolitical forecasting tournaments over the last 25 years (Tetlock, 2005; Tetlock & Mellers, 2011; Mellers et al, 2013). These forecasters may will be right when they say that the risk of a lethal confrontation claiming 10 or more lives in the East China Sea by the end of 2013 is only 0.04. They may be very "well calibrated" in the narrow technical sense that when they attach a 4% likelihood to events, those events occur only about 4% of the time. But framing a vanilla question as a binary question is dangerous because it masks exponentially escalating tail risks: the risks of a confrontation claiming not just 10 lives of 1000 or 1 million. No one has yet figured out how to design a forecasting tournament to assess the accuracy of probability judgments that range between .00000001% and 1%--and if someone ever did, it is unlikely that anyone would have the patience —or lifespan —to run the forecasting tournament for the necessary stretches of time (requiring us to think not just in terms of decades, centuries and millennia).

The deep ambiguity of objective probabilities at the extremes—and the inevitable instability in subjective probability estimates—can also create patterns of systematic mispricing of options. An option or option like payoff is not to be confused with a lottery, and the "lottery effect" or "long shot bias" often discussed in the economics literature that documents that agents overpay for these bets should not apply to the properties of actual options.

In *Fooled by Randomness*, the narrator is asked "do you predict that the market is going up or down?" "Up", he said, with confidence. Then the questioner got angry when he discovered that the narrator was short the market, i.e., would benefit from the market going down. The trader had a difficulty conveying the idea that someone could hold the belief that the market had a higher probability of going up, but that, should it go down, it would go down a lot. So the rational response was to be short.

This divorce between the binary (up is more likely) and the vanilla is very prevalent in real-world variables. Indeed we often see reports on how a certain financial institution "did not have a losing day in the entire quarter", only to see it going near-bust from a monstrously large trading loss. Likewise some predictors have an excellent record, except that following their advice would result in large losses, as they are rarely wrong, but when they miss their forecast, the results are devastating.

Remark*: More technically, for a heavy tailed distribution (defined as part of the subexponential family, see Taleb 2013), with at least one unbounded side to the random variable, the vanilla prediction record over a long series will be of the same order as the best or worst prediction, whichever in largest in absolute value, while no single outcome can change the record of the binary*.

Another way to put the point: to achieve the reputation of “Savior of Western civilization,” a politician such as Winston Churchill needed to be right on only one super-big question (such as the geopolitical intentions of the Nazis)-- and it matters not how many smaller errors that politician made (e.g. Gallipoli, gold standard, autonomy for India). Churchill could have a terrible Brier score (binary accuracy) and a wonderful reputation (albeit one that still pivots on historical counterfactuals).

Finally, one of the authors wrote an entire book (Taleb, 1997) on the hedging and mathematical differences between binary and vanilla. When he was an option trader, he realized that binary options have nothing to do with vanilla options, economically and mathematically. Seventeen years later people are still making the mistake.

Acknowledgments

Bruno Dupire, Raphael Douady, Daniel Kahneman, Barbara Mellers.

References

Mellers, B. et al. (2013). How to win a geopolitical forecasting tournament: The power of teaming and training. Unpublished manuscript, Wharton School, University of Pennsylvania Team Good Judgment Lab.

Silver, Nate, 2012, *The Signal and the Noise*.

Taleb, N.N., 1997, *Dynamic Hedging: Managing Vanilla and Exotic Options*, Wiley

Taleb, N.N., 2001/2004, *Fooled by Randomness,* Random House

Taleb, N.N., 2013, *Fat Tails and Antifragility: Lectures on Probability, Risk and Decisions in the Real World,* Freely Available Web Book, www.fooledbyrandomness.com

Tetlock, P.E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton: Princeton University Press.

Tetlock, P.E., Lebow, R.N., & Parker, G. (Eds.) (2006). *Unmaking the West: What-if scenarios that rewrite world history*. Ann Arbor, MI: University of Michigan Press.

Tetlock, P. E., & Mellers, B.A. (2011). Intelligent management of intelligence agencies: Beyond accountability ping-pong. American Psychologist, 66(6), 542-554.

Appendix: A Semi-Technical Commentary

Fatter tails lower the probability of remote events (the binary) and raise the value of the vanilla. The following intuitive exercise will illustrate what happens when one conserves the variance of a distribution, but "fattens the tails" by increasing the kurtosis. The probability of a certain type of intermediate and large deviation drops, but their impact increases. Counterintuitively, the possibility of staying within a band increases.

Let x be a standard Gaussian random variable with mean 0 (with no loss of generality) and standard deviation σ. Let $P_{>1\sigma}$ be the probability of exceeding one standard deviation. $P_{>1\sigma} = 1 - \frac{1}{2}\operatorname{erfc}\left(-\frac{1}{\sqrt{2}}\right)$, where erfc is the complimentary error function, so $P_{>1\sigma} = P_{<1\sigma} \simeq 15.86\%$ and the probability of staying within the "stability tunnel" between $\pm 1\,\sigma$ is $1 - P_{>1\sigma} - P_{<1\sigma} \simeq 68.3\,\%$.

Let us fatten the tail in a variance-preserving manner, using the "barbell" standard method of linear combination of two Gaussians with two standard deviations separated by $\sigma\sqrt{1+a}$ and $\sigma\sqrt{1-a}$,

$a \in (0,1)$, where a is the "vvol" (which is variance preserving, technically of no big effect here, as a standard deviation-preserving spreading gives the same qualitative result). Such a method leads to the immediate raising of the standard Kurtosis by $(1+a^2)$ since $\frac{\mathbb{E}(x^4)}{\mathbb{E}(x^2)^2} = 3\,(a^2+1)$, where $\mathbb{E}$ is the expectation operator.

$$P_{>1\sigma} = P_{<1\sigma} = 1 - \frac{1}{2}\operatorname{erfc}\left(-\frac{1}{\sqrt{2}\,\sqrt{1-a}}\right) - \frac{1}{2}\operatorname{erfc}\left(-\frac{1}{\sqrt{2}\,\sqrt{a+1}}\right) \tag{1}$$

So then, for different values of a in Eq. 1 as we can see in Figure 2, the probability of staying inside 1 sigma rises, "rare" events become less frequent.

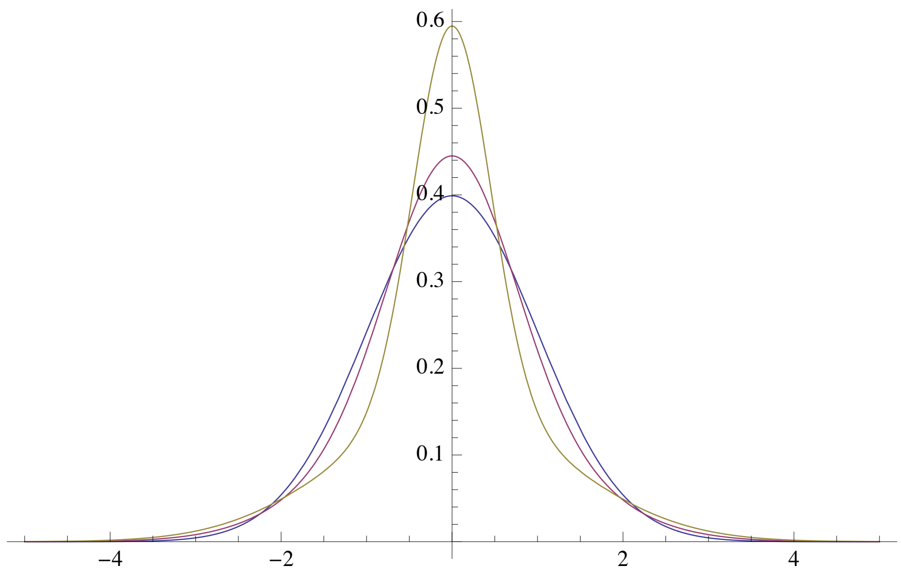

Figure 2 Fatter and fatter tails: different values of *a*. Note that higher peak ⟹lower probability of leaving the ±1 σ tunnel

Note that this example was simplified for ease of argument. In fact the "tunnel" inside of which fat tailedness increases probabilities is between $-\sqrt{\frac{1}{2}\left(5-\sqrt{17}\right)}\,\sigma$ and $\sqrt{\frac{1}{2}\left(5-\sqrt{17}\right)}\,\sigma$ (even narrower than 1 σ in the example, as it numerically corresponds to the area between -.66 and .66), and the outer one is $\pm\sqrt{\frac{1}{2}\left(5+\sqrt{17}\right)}\,\sigma$, that is the area beyond $\pm 2.13\,\sigma$.

The law of large numbers: Getting a bit more technical, the law of large numbers works much faster for the binary than the vanilla (for which it may never work, see Taleb, 2013). The more convex the payoff, the more observations one needs to make a reliable inference. The idea is as follows, as can be illustrated by an extreme example of very tractable binary and intractable vanilla.

Let x_t be the realization of the random variable $X \in (-\infty, \infty)$ at period t, which follows a Cauchy distribution with p.d.f. $f(x_t) \equiv \frac{1}{\pi\left((x_0-1)^2+1\right)}$. Let us set $x_0 = 0$ to simplify and make the exposure symmetric around 0. The Vanilla exposure maps to the variable x_t and has an expectation $\mathbb{E}(x_t) = \int_{-\infty}^{\infty} x_t\, f(x)\, dx$, which is infinite. A bet at x_0 has a payoff mapped by as a Heaviside Theta Function $\theta_{>x_0}(x_t)$ paying 1 if $x_t > x_0$and 0 otherwise. The expectation of the payoff is simply $\mathbb{E}(\theta(x)) = \int_{-\infty}^{\infty} \theta_{>x_0}(x)\, f(x)\, dx = \int_{x_0}^{\infty} f(x)\, dx$, which is simply $P(x > 0)$. So long as a distribution exists, the binary exists and is Bernouilli distributed with probability of success and failure p and $1-p$ respectively.

The irony is that the payoff of a bet on a Cauchy, admittedly the worst possible distribution to work with since it lacks both mean and variance, can be mapped by a Bernouilli distribution, about the most tractable of the distributions. In this case the Vanilla is the hardest thing to estimate, and the binary is the easiest thing to estimate.

Set $S_n = \frac{1}{n} \sum_{i=1}^{n} x_{t_i}$ the average payoff of a variety of vanilla bets x_{t_i}across periods t_i, and $S^\theta{}_n = \frac{1}{n} \sum_{i=1}^{n} \theta_{>x_0}(x_{t_i})$. No matter how large n, $\lim_{n\to\infty} S^\theta{}_n$ has the same properties – the exact same probability distribution –as S_1. On the other hand $\lim_{n\to\infty} S^\theta{}_n = p$; further the presaymptotics of $S^\theta{}_n$ are tractable since it converges to $\frac{1}{2}$ rather quickly, and the standard deviations declines at speed $\sqrt{n}$, since $\sqrt{V(S^\theta{}_n)} = \sqrt{\frac{V(S^\theta{}_1)}{n}} = \sqrt{\frac{(1-p)\, p}{n}}$ (given that the moment generating function for the average is $M(z) = \left(p\, e^{z/n} - p + 1\right)^n$).

The binary has necessarily a thin-tailed distribution, regardless of domain. More, generally, for the class of heavy tailed distributions, in a long time series, the sum is of the same order as the maximum, which cannot be the case for the binary: $\lim_{X\to\infty} \frac{P\left[X > \sum_{i=1}^{n} x_{t_i}\right]}{P\left[X > \text{Max}\left(x_{t_i}\right)_{i\le 2\le n}\right]} = 1$. Compare this to the binary for which $\lim_{X\to\infty} P(X > \text{Max}\, \theta(x_{t_i})_{i\le 2\le n}) = 0$. The binary is necessarily a thin-tailed distribution, regardless of domain.

We can assert the following:

1. The standard deviation of the binary is always smaller or equal to that of the vanilla.
2. The sum of binaries converges at a speed faster or equal to that of the vanilla.

3. The sum of binaries is never dominated by a single event, while that of the vanilla can be.

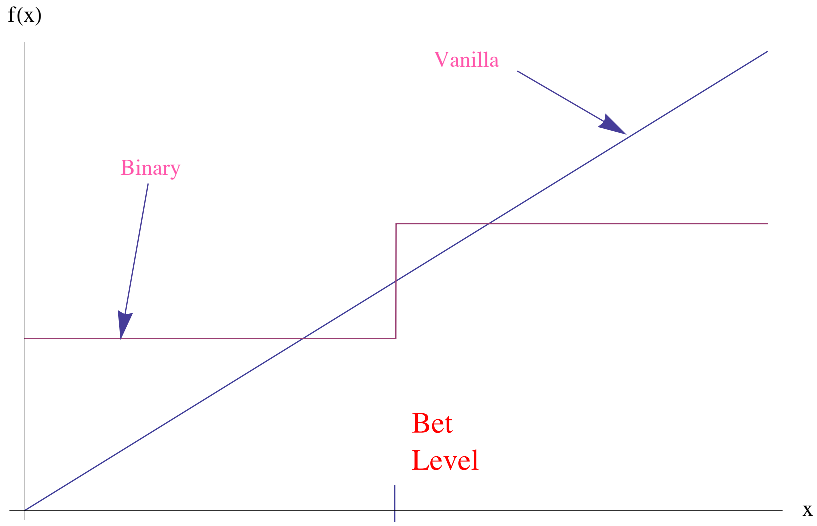

Figure 4 The different classes of payoff f(x) seen in relation to an event x. (When considering options, the vanilla can start at a given bet level, so the payoff would be continuous on one side, not the other).

How is the binary more robust to model error? In the more general case, the expected payoff of the vanilla is expressed as $\int_A x \, dF(x)$ (the unconditional shortfall) while that of the binary= $\int_A dF(x)$, where A is the part of the support of interest for the exposure, typically A≡[K,∞), or (–∞,K]. Consider model error as perturbations in the parameters that determine the calculations of the probabilities. In the case of the vanilla, the perturbation's effect on the probability is multiplied by a larger value of x.

As an example, define a slighly more complicated vanilla than before, with option-like characteristics, $V(\alpha, K) \equiv \int_K^\infty x \, p_\alpha(x) \, dx$ and $B(\alpha, K) \equiv \int_K^\infty p_\alpha(x) \, dx$, where V is the expected payoff of vanilla, B is that of the binary, K is the "strike" equivalent for the bet level, and with x∈[1, ∞) let $p_\alpha(x)$ be the density of the Pareto distribution with minimum value 1 and tail exponent α, so $p_\alpha(x) \equiv \alpha \, x^{-\alpha-1}$.

Set the binary at .02, that is, a 2% probability of exceeding a certain number K, corresponds to an α=1.2275 and a K=24.2, so the binary is expressed as $B(1.2, 24.2)$. Let us perturbate α, the tail exponent, to double the probability from .02 to .04. The result is $\frac{B(1.01,24.2)}{B(1.2,24.2)} = 2$. The corresponding effect on the vanilla is $\frac{V(1.01,24.2)}{V(1.2,24.2)} = 37.4$. In this case the vanilla was ~18 times more sensitive than the binary.