Why We Don’t Know What We Talk About When We Talk About Probability

Nassim N. Taleb

Distinguished Lecture Series,
Humbold University, Berlin, Department of Applied Statistics,
January 28, 2012

January 2012 - DRAFT- cannot be cited yet without permission

Explains why naive discussions of probability are extremely harmul. It explains the problems in discussing probability and expected payoffs in the presence of convexity effects, and shows how 1) the dynamics of exposure are divorced from those of probability, and 2) how probability needs to be used as a *kernel* not an end result. Shows the difference between *naive* probability in the philosophical and other discussions and *kernel* probability. Explains the difference between “knowing” and “doing”, the “Aristotelian” v/s the “Thalesian”. Following the logic leads us to the central error made in Fukushima. Why, for the planet’s sake, people should stop talking about probability and focus on robustness.

I- Standard Probability Confusions

1- Probability is too primitive a term for discussion of the most basic issues

In *Fooled by Randomness* (Taleb, 2001/2005), the character is asked which was *more probable* that a given market would go higher or lower by the end of the month. Higher, he said, much more probable. But then it was revealed that he was making trades that benefit if that particular market *goes down*. This of course, appears to be paradoxical for the nonprobabilist but very ordinary for traders (yes, the market is more likely to go up, but should it go down it will fall much much more) and obvious when written down in probabilistic form, so for S_t the maket price at period t there is nothing incompatible between $P[S_{t+1}>S_t]$ and $E[S_{t+1}]<S_t$ (where E is the expectation), when one takes into account full distribution for S_{t+1}, which comes from having the mean much much lower than the median (under negative skewness of the distribution). Plus this example illustrates the common confusion between a *bet* and an exposure (a bet is a binary outcome, an exposure has more nuanced results and depends on full distribution). This example shows one of the extremely elementary mistakes of talking about *probability presented* as single numbers not distributions of outcomes, but when we go deeper into the subject, many less obvious, or less known paradox-style problems occur. Simply, it is of the opinion of the author, that it is not rigorous to talk about “probability” as a final product, or even as a “foundation” of decisions.

Another anecdote concerns a ski resort in the Lebanon. The owner of the ski resort, deploring lack of snow, deposited at a shrine the Virgin Mary a $100 wishing for snow. Snow came, with such abundance, and avalanches, with people stuck in the cars, and the resort was forced to close, prompting the owner to quip “I should have only given $25”. What the owner did is discover the notion of nonlinear exposure and extreme events. Under so-called “fat tails”, there is no such thing as a “typical event”, and nonlinearity causes even more severe problems.

Most of the problems discussed in this paper (except the last one involving metaprobabilities, in Section II) might be trivial mathematically and well known (on paper), but they are only retrospectively trivial as researchers somewhere, under pressures to simplify, engage in Procrustean bed types of simplifications that lead to chains of severe errors, the entire thing buried in the derivations.

For instance, making a mistake between a bet and a payoff (binary) that depends on the full distribution is extremely common among sophisticated people, which led high ranking social scientists to start the so-called "prediction markets" grounded in the idea that binary bets on markets going higher or lower are hedges (they are not, as was asserted in Taleb (1997) yet the incidence of such confusion has increased since).

Trivial perhaps, but serious. Very serious. Otherwise we would not have had the crisis that started in 2008 in spite of thousands of PhDs in quantitative fields missing the point and similar ones staring at them. Indeed the probabilistic mistakes that led the the collapse of the banking system were trivial and remain uncured: banks focusing on "probability" of profit and loss, causing them to engage in negative Black Swan exposures, a high probability of "mildly profitable", low probability of disaster, and a negative overall expected return while being convinced that “being profitable most of the time” is their mission.

2- Exposures to X Should Not Be Confused With Knowledge About X --and They Are

Exposure, not knowledge. Next we go into a deeper set of problems less obvious in the discourse, that have a severe philosophical consequence. The confusion is as follows. Take *X* a random or nonrandom variable, and *F(X)* the exposure, payoff, the effect of *X* on you, the end bottom line. (To be technical, *X* is higher dimensions, in $\mathbb{R}^N$ but less assume for the sake of the examples in the introduction that it is a simple one-dimensional variable).

The disconnect. As a practitioner and risk taker I realize the disconnect: people (nonpractitioners) talking to me about X (with the implication that we practitioners should care about X in running our affairs) while I had been thinking about F(X), nothing but F(X). And the straight confusion since Aristotle between X and F(X) has been chronic. Sometimes people mention F(X) as utility but miss the full payoff. And the confusion is at two level: one, simple confusion; second, in the decision-science literature, seeing the difference and not realizing that action on F(X) is easier than action on X.

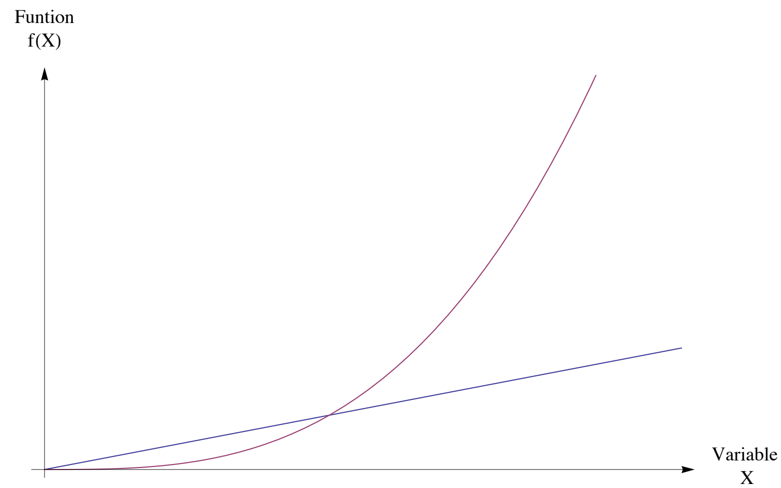
Examples:

X is arsenic, coffee, or aspirin, *F(X)* is the response of your body to the ingestion of arsenic, coffee, or aspirin (obviously the dose response is extremely nonlinear making F complicated).
X can be a shock hitting a coffee cup, *F(X)* is the response in change in *value* of the coffee cup and its *utility* for you.

X is a "prediction", *F(X)* is how it affects me.

X can be rain, *F(X)* is the effect of rain on you.

X is unemployment in Senegal, $F_1(X)$ is the effect on the bottom line of the IMF, and $F_2(X)$ is the effect on your grandmother (which I assume is minimal).
X can be a stock price, but you own an option on it, so F(X) is your exposure an option value for X, or, even more complicated the utility of the exposure to the option value.
X can be changes in wealth, F(X) the convex-concave value function of Kahneman-Tversky, how these "affect" you. One can see that F(X) is vastly more stable or robust than X.



A convex and linear function of a variable X. Confusing f(X) (on the vertical) and X (the horizontal) is more and more significant when f(X) is nonlinear. The more convex f(X), the more the statistical and other properties of f(X) will be divorced from those of X. For instance, the mean of f(X) will be different from f(Mean of X), by Jensen's ineqality. But beyond Jensen's inequality, the difference in risks between the two will be more and more considerable. When it comes to probability, the more nonlinear f, the less the probabilities of X matter compared to the nonlinearity of f. Moral of the story: focus on f, which we can alter, rather than the measurement of the elise properties of X.

There are infinite numbers of functions *F* depending on a unique variable *X*.

All utilities need to be embedded in *F*.

Limitations of knowledge. What is crucial, our limitations of knowledge apply to X not necessarily to *F(X)*. We have no control over X, some control over *F(X)*. In some cases a very, very large control over *F(X)*.

My point is that we should not confuse "knowledge" of *X* with that of F(X), the effect of the payoff, an error often made, actually, almost always made. And "knowledge of X" does not crudely translate in knowledge of *F(X)*; *F(X)* is the full package, the begining and end package. F(X) needs to be computed in a completely different manner, as we will see, using all possible values of X. This seems naive, but people do, as something is lost in the translation.

- My irritation with the treatment of the Black Swan problem is as follows: people focus on X ("predicting X"). My point although we do not understand X, we can deal with it by working on F which we can understand, while others work on predicting *X* which we can't because small probabilities are incomputable, particularly in "fat tailed" domains. F(X) is how the end result affects you.
- The probability distribution of *F(X)* is markedly different from that of X, particularly when *F(X)* is nonlinear. We need a nonlinear transformation of the distribution of *X* to get F(X). We had to wait until 1964 to get a paper on "convex transformations of random variables", Van Zwet (1964).

Bad news: F is almost always nonlinear.

The central point about what to understand: When *F(X)* is convex, say as in trial and error, or with an option, we do not need to understand *X* as much as our exposure to H. Simply the statistical properties of *X* are swamped by those of *H*. That's the point of *Antifragility* in which exposure is more important than the naive notion of "knowledge", that is, understanding *X*.

Fragility and Antifragility:

- When *F(X)* is concave (fragile), errors about X can translate into extreme negative values for F. When *F(X)* is convex, one is immune from negative variations.
- The more nonlinear F the less the probabilities of X matter in the probability distribution of the final package *F*.
- Most people confuse the probabilites of X with those of F. I am serious: the *entire* literature reposes largely on this mistake.

So, for now ignore discussions of X that do not have *F*. And, for Baal's sake, focus on *F*, not *X*.

3- Probability and Decisions Cannot be Naively Extracted from Each Other

I will rephrase the previous two points, as they are nonseparable. Luckily, we do not use probabilities in daily (and less daily) decisions, at least not in the raw form presented in the literature -doing so would have made us exit the gene pool. Probabilities appear to be only present in textbooks (even then, often in mistaken and naive forms). And for a reason: as we saw so far, probabilities are a crude, and distorted input that cannot naively isolated into a single number, and mathematically extracted from its context and *F*, the payoff package represented.

So decisions should not be based on these crude single probabilities, which I call "raw probabilities" - technically one needs to use a double kernel across values of payoff, and *errors of probability* (or metaprobabilities, which we leave for Section II), regardless of the statement made -except when the probabilities are deemed both single and *a priori*, not subjected to change. (Nor should we discuss single probabilities as "starting points" because these don't generalize. Because of nonlinearities of the effects, these statements are usually Procrustean-bed reductions.)

So the first severe error is to talk about "raw" probabilities with the implication that they lead to decisions outside "ludic" contexts (such as casinos, lotery tickets and matters that are constructed and tend to be limited to textbooks and quite absent from real life). A raw probability is any point estimate reduction of the sort "5% chance of rain" or "5% chance of war", etc., which discretizes the event and expressed it using a single number rather than a full distribution (discrete across many numbers or continuous using densities). To repeat the confusion, these raw probabilities are nice, but once the "effect" of different levels of rain or war are taken into account and the variations within the quantity of rain, severity of war, etc., then the effects vary as these need to be computed from within the integral. (The condition in which transforming these events of multiple layers and effects into a single event would be allowed is: full linearity of payoff across all values of the random variable).

This error, which we saw is commonly made by economists discussing "betting" markets and setups (or "Value at Risk"), is also severely made by philosophers discussing "philosophy of probability" (see Hacking, 1984, Hajek, 2010). The irony is that while many such persons take millions of decisions without explicit probability in their natural ecology; they fail to realize that when it comes to the literature the tendency is to focus on examples and concepts related to "lotteries" and similar matters that do not resemble the ecology of the real world, as it is one of the rare (constructed) cases of truly single-event probabilities. It looks like people are vastly more intelligent with probability when they don't talk about it.

Example: ask any philosopher dealing with probability how many times he has used probability as input into decision over his life, which patently should include billions of small decisions. The only answers I usually get from them is lottery examples. People may embed the probability kernel into their decisions but F is dominant (Taleb, 2012) --this is similar to an "emerging property" resuting from nonlinearities.

The second and third errors are much more severe and have been largely missed in the literarure, and are squarely behind the errors causing Black Swan effects: they are the results of not integrating (literally) the variations, model errors, and difficulties in the derivation of the probabilities themselves --probabilities don't fall from the sky and we need to embed uncertainties about them. These uncertainties are not trivial and affect small probabilities and, correspondingly, their payoffs in severe ways.

4- Some Technical Discussions

Raw Probability: So where x is the random variable, the first error would be to take (as in textbooks) what we have called earlier the "raw probability" (with the usual "triple" where p is the measure, etc., over a set interval Dx that can be selected in any possible way on the σ-Algebra):

$$P_{raw} = \int_{Dx} p(x)\, dx \tag{1}$$

when in fact we need to take into account the "payoff", which may be nonlinear (actually, should be except rare exceptions nonlinear). The payoff can be a function F(X) (which includes contractual effects, preferences and utilities), which is standard in the literature since Luce and Raiffa (1958).

Raw Expected Payoff: So the first naive kerneling is what we call the "raw expectation", which makes a difference in situations of discontinu-ities and nonlinearities for H:

$$EP_{raw} = \int_{Dx} F(x)\, p(x)\, dx \tag{2}$$

So, naively, the linear $F(x)$= x in situation in which EP_{raw} would be the standard expectation over Dx. $F(x) = x^n$ would be the n^{th}noncentral moment. And something like $H(x) = (x - k)^+$ (where k is a "strike" value), is a situation of explicit optionality, or introducting standard utility would give U(x).

First Fallacy: Making a decision based on P_{raw} or discussing P_{raw} as something that can lead to decision or "building block" of probabilities other than entertainment. We see that P_{raw}multiplied by some scalar (that corresponds to some type of payoff) does not directly translate into EP_{raw}. EP_{raw} under convexity of F cannot be trivially inferred from P_{raw}. This is the Thales/Aristotle Problem in *Antifragility*, Taleb (2012a) in which Thales' performance comes from the type of bet he made, (here the H) , not from his knowledge of the odds of the event (the P), yet is misattributed by Aristotle to the understanding of the odds.

Irrelevance of likelihood. For instance, reversal in likelihood do not necessarity lead to reversal in decision (or expected payoff); often the opposite. For instance, one can easily have $P_1 > P_2$, with $EP_1 < EP_2$, in spite of having the same first order exposure, $\frac{\partial F_1}{\partial x} = \frac{\partial F_2}{\partial x}$ say in the case were H_1more convex than H_2 ($\frac{\partial^2 F_1}{\partial x^2} > \frac{\partial^2 F_2}{\partial x^2}$ for all values of x). In that case, EP_2 is more impacted by the "tails", and is more concerned with larger values of x, the second order effects.

Importance of the point. Aside from the Aristotle-Thales problem, the mistake is seen all across naive uses of probability, and permeates the texts in the philosophy of probability. Probability is not an end product, and is not raw material that can be naively used as "building block" or "foundation".

The foundation or "building block" is F(X) p(x), nothing else, period.

Example of severe such mistake in finance: Value-at-Risk is defined as finding s such as $P(\Delta W(s) > VaR) = p$, where p the probability (typically 99%), ΔW the changes in wealth, s is the state variable or vector, “VaR” is the value one does not want to lose given the 99% confidence level. the “value at risk”. The problem of course is that the expected loss conditional on $\Delta W(s) < VaR$ can be (and usually is) monstrously high.

Experience of the author. There is a difference between packages of the style "digital" (more like a De Finetti bet, P[X>k]) purely probabilistic and packages that require "full distribution". Often the rise of the value of the “digital” is accompanied with a drop in the value of $E[X \mid_{x>k}]$. See *Dynamic Hedging*, Taleb (1997)

Error in De Finetti. De Finetti, in the magisterial *La Prévision*, focuses on prediction off an expected value, not a probability. But this remains a point estimate, which does not capture the additional convexity of H, the effect of the prediction, and, of course, as we see in the third error below, the error around the prediction.

Betting pathologies. Observing someone’s bet does not reveal his probability, rather his utility or his H. Say I fear the election of candidate x. I would make a bet (over its “statistical” measure, what we call the $\mathcal{P}$ measure in finance) that x would win -should he win, I would be compensated as a hedge. So betting may show a lot more preferences and desire to hedge (the utility embedded in H) than the probability.

Assymetries of F(X) **in life:** almost all life situations (with include biological systems) are marred with nonlinear effect: survival, death, concave utility of gains, optionality, loss of job, starvation, dose-response, etc.

5- Philosophical Implications of the Standard Confusions

This shows the difference between “I know that” and “I do”, and why “knowledge” and “doing” do not naively map into each other. “I know” is a very, very lower form of statement. It can only map directly into action in situations where either a)there is absolutelty no probabilistic dimension , b) F is completely linear across all values of x. In other words, total absence of second order effects.

II- Error and Consequence of Missing Metaprobabilities

Correcting for Errors in Probabilities With Metaprobabilities

Now the central mistake. If probabilities have errors, then Both $P_{\text{raw}}$and E_{raw} themselves have problems and should not be used without another element, the kernel embedding uncertainty about the probability (whatever its source, wrong distribution, wrong parameter, etc.).

A probability that has measurement error that depends on a stochastic term λ needs to be integrated across q(λ), weighted values of λ using the kernel $p_\lambda(x)$ in place of p(x)

So taking Eq. (1) and correcting it, the raw probability becomes, taking one level of error terms λ, the convolution (where, to avoid confusion, the kernel $p_\lambda(x)$ means "probability of x using λ as a parameter in the computation", with q(λ) the first order "metaprobability", or the probability distribution for λ; $\lambda \in \mathcal{R}^N$). Assuming independance between x and λ:

$$p^*(x) = \int_{D\lambda} p_\lambda(x)\, q(\lambda)\, d\lambda \tag{3}$$

and P_{adjusted} the integral over $p^*(x)$ as in Eq. 1

Note that the uncertainty over the probability distribution can be accomodated using *q(λ)*.

Error on Probability: $\frac{P_{\text{adjusted}}}{P_{\text{raw}}} \neq 1$. This is compounded by small probabilities. We saw in Taleb (2012b), Taleb and Douady (2012) that, in the presence of errors $\frac{P_{\text{adjusted}}}{P_{\text{raw}}}$ can exceed 10^7 for rare events of the "six sigma" variety and can explain fat tails. The problem of Fukushima was ignoring (by criminally incompetent idiots) of the necessary convexity effect from the metadistribution q, with the ratio $\frac{P_{\text{adjusted}}}{P_{\text{raw}}} > 10^5$

But we still can't talk naively about $P_{adjusted}$.

(Note: This λ is in higher dimenstions and includes the case where error terms may have error terms. One can use λ as, say the standard deviation, which itself has a standard deviation, which causes the ratio to swell even more. Assuming $\lambda_n$is the error in estimating λ_{n-1} and, further, assuming independence of the errors, we can keep chaining.)

Second the probabilities themselves are subjected to errors of missing the nonlinear effect of metaprobabilities.

$$EP_{adjusted} = \int_{D\lambda} \int_{Dx} F(x)\, p_\lambda(x)\, q(\lambda)\, dx\, d\lambda \tag{4}$$

assuming independence, to simplify

$$\text{EP}_{\text{adjusted}} = \int_{Dx} F(x)\, p^*(x)\, dx \tag{5}$$

The Most Severe Error. The ratio $\frac{EP_{adjusted}}{EP_{raw}}$ is where people tend to get in severe trouble --it is completely missing in economics. This is what makes portfolio theory in particular, and other forms of optimization fragile to model error and prone to blowups.

Conclusion

This paper is meant to show the disease, not the cure. Should we ban the use of probability? Most certainly, we humans tend to build the most risks (particularly risks of tail events, of "blowups") when we discuss probability, so we need to prevent systems from being based on probabilistic computations. In other words, we need to work on changing F(X) such that the understanding of the properties of X matter little, an operation called "robustification" by the author. *Antifragility* (Taleb 2012) builds such maps.

Acknowledgments

Raphael Douady, Emanuel Derman, Jean-Pierre Aubin, Bruno Dupire, Gur Huberman ...

References

De Finetti, B. (1937). La prévision : ses lois logiques, ses sources subjectives, Institut Henri Poincaré

Hacking, I. (1984). The emergence of probability : a philosophical study of early ideas about probability, induction and statistical inference, Cambridge Univ Pr.

Hacking, I. (1990). The taming of chance, Cambridge Univ Pr.

Hajek, A. (2003). Interpretations of probability, Stanford Encyclopedia of Philosophy

Taleb, N.N., and Douady, R. (2012), A Map and Simple Heuristic to Detect Fragility, Antifragility, and Model Error, under revision, *Quantitative Finance*

Taleb, N.N. (1997), *Dynamic Hedging*, Wiley

Taleb, N.N. (2001, 2005), *Fooled by Randomness*, Penguin (UK) and Random House (US)

Taleb, N.N. (2012a), *Antifragility*, Penguin (UK) and Random House (US), in press

Taleb, N.N. (2012b), The Future Has Thicker Tails than the Past : Model Error as Branching Counterfactuals, preprint

Van Zwet,W.R. (1964). *Convex Transformations of Random Variables*, Mathematical Center Amsterdam, 7