Convexity and Conflation Biases as Bregman Divergences: A note

Eric Briys
Ceregmia - U. Antilles - Guyane
eric.b@cyberlibris.com

Brice Magdalou
LAMETA - U. Montpellier 1
brice.magdalou@univ-montp1.fr

Richard Nock
Ceregmia - U. Antilles - Guyane
rnock@martinique.univ-ag.fr

November 23, 2012

Abstract

In "Antifragile" [1], Taleb provides a fresh perspective on how one may gain from disorder. In this short note, we formalize and unify in a single premium (a schematic view of) the concavity/convexity and conflation effects described by Taleb. We show that this premium relies on a generalization of a well-known class of distortion measures of information geometry, namely Bregman divergences. We exemplify some properties of this premium, and discuss them in the light of "Antifragile" [1].

1 Introduction

Everything gains or loses from volatility. Things that lose from volatility are fragile. Things that gain from uncertainty are the opposite. In his latest opus (2012), Nassim Nicholas Taleb call these things that gain from disorder antifragile [1]. Taleb advocates a quest for antifragility, a quest that nature has long embarked into. Sadly enough, mankind is more often than not exposed to fragility without knowing it. The reasons for this lethal blindness are many. According to Taleb, we suffer from what he calls the conflation error and we overlook the convexity bias.

The conflation error is due to the fact that we spend too much time and resources trying to predict (unpredictable) outcomes: In the course of doing so, we take the tree for the forest. We are obsessed by prediction while we should focus on what our exposure to uncertain outcomes is all about. What matters is not, say, the uncertain outcome x but how it affects us. What matters to us is our response function $f(x)$. Confusing x for $f(x)$ or the expectation $\mathrm{E}(x)$ for $f(\mathrm{E}(x))$, for that matter, is the conflation error. This error compounds with the convexity bias. Indeed, even if we were avoiding the conflation error on focusing on $f(\mathrm{E}(x))$, this would have the unfortunate consequence that we miss whether we are in a fragile or antifragile position. The right move is to use $\mathrm{E}(f(x))$ instead of $f(\mathrm{E}(x))$. If f is concave we lose from disorder, if f is convex we gain from it.

In this note, we show that Taleb's concepts can be summarized into a single premium which we call Taleb's premium. Borrowing from recent developments in information theory we show that the conflation error and the convexity bias can be measured by Bregman divergences.

2 The almighty mean

The notion of expected value, of average, is very familiar to us. It arises spontaneously from the natural tendency of the human mind to unify in a more or less comprehensive synthesis quantitative impressions it receives from objects[1].

Averages are addictive. Whatever the quantity of interest (sizes, weights, financial returns, interest rates, rainfall levels, temperatures, growth rates, budget deficits, sales, costs...), it is first summarized by its average. When it comes to predicting the value of a random outcome from observations, averages come again into play.

But, as we all know: "*The expected value is not to be expected*". This is a casual way of saying that relying on the average inevitably loses information about the data it is supposed to synthesize. In other words, averaging yields a trade-off between data compression and information loss. This trade-off is well-known in econometrics where the loss in information is usually measured by the mean squared error: Under this assumption, the conditional average is the unique (best) predictor that minimizes the mean squared error.

One may wonder, given the sheer popularity of the (conditional) average if there are other (more general) information loss functions for which the average is the unique best predictor of future outcome. This would give another strong rationale for its ubiquity. Recent work in information theory [2] shows that under so-called Bregman loss functions, aka Bregman divergences the (conditional) average is the unique best predictor: It minimizes the expectation of the Bregman divergence[2].

Definition 1 *[3] Let f be some convex differentiable function. The Bregman divergence between x and y is: $D_f(x\|y) \doteq f(x) - f(y) - (x-y)f'(y)$, where f' is the derivative of f.*

Bregman divergences generalize well-known distortion measures used in geometry (Mahalanobis divergence), statistics (Kullback-Leibler divergence), signal processing (Itakura-Saito divergence), etc. . In layman's terms, the Bregman loss function measures the rest of the Taylor expansion: Under a (convex) Bregman loss function, this expected rest is minimized when the average is taken as the best predictor.

But there is more to this result. Indeed, the minimum of the expected Bregman loss function has a nice and useful interpretation. It is equal to the Jensen's gap.

Definition 2 *For any set $\mathcal{S} = \{x_1, x_2, ...\}$ equipped with discrete probabilities $\mathcal{P} = \{p_1, p_2, ...\}$, we let $\mu_\mathcal{S} \doteq \mathrm{E}_\mathcal{P}\mathcal{S}$.*

Jensen's gap over f is then the difference between the expected value of f and the f-value for the average; it is a special case of Bregman divergence as we have:

$$\mathrm{E}_\mathcal{P} f(x) - f(\mu_\mathcal{S}) \quad = \quad \mathrm{E}_\mathcal{P} D_f(x\|\mu_\mathcal{S}) \ . \tag{1}$$

Banerjee et alli call this gap the Bregman information. To get the intuition of this result, assume that the random variable is Gaussian. In that case, the Bregman information is equal to the variance. Indeed, compressing the Gaussian variable to its mean is tantamount to losing the variance information, which is exactly what Jensen's inequality boils down to. Under the normality assumption, the loss function measured at the average of the random variable diverges from the expected loss function by the variance factor.

[1] Lottin, Joseph: "La théorie des moyennes et son emploi dans les sciences d'observation", Revue néo-scolastique, 16ème année, n°64, 1909, p 537

[2] Note that this result works both ways: If the conditional average is the best predictor, the loss function is a Bregman divergence. If the loss function is a Bregman divergence, the unique best predictor is the conditional average.

3 The mean can be mean

Despite all its advantages, focusing on the mean can truly be misleading. Indeed, as forcefully reminded in Nassim Taleb's latest book [1], we spend too much time and resources on (predicting unpredictable) random outcomes. According to Taleb, we should rather uncover what our true exposure to uncertain outcomes is all about. After all, what affects us is not the uncertain prospect per se but how our exposure to it transforms its impact into either a favorable or damaging outcome. In other words, by spending too much energy on the random prospect x, we confuse x for $f(x)$ where f describes our exposure to x. What matters is our response function, $f(x)$, to x. Confusing x for $f(x)$ is what Taleb calls the conflation error.

This error hides a more pervasive consequence. And, this is where Jensen's come into play again. Assume that you try to predict x and come up with the expectation $\mathrm{E}(x) = \mu_{\mathcal{S}}$ as its best predictor. Assume that you are not subject to the conflation error and do measure $f(\mu_{\mathcal{S}})$. There you go says Taleb: You miss the shape of f and what this shape implies to your "health".

Let's assume that f is concave. This is the typical exposure of an investor who for instance has shorted a call option on a stock. The investor is said to be negative gamma. In that case, Jensen's inequality tells the investor that $f(\mu_{\mathcal{S}})$ overestimates the outcome of his trading position. Indeed, for a concave f:

$$\mathrm{E}_{\mathcal{P}} f(x) - f(\mu_{\mathcal{S}}) \quad < \quad 0 \ .$$

In other words, the investor is (negatively) exposed to volatility. Any major swings in the underlying stock price or its volatility translate into losses. This is what Taleb calls missing the convexity (concavity for that matter) bias. The investor is experiencing fragility.

If the same trader were long the same call option, his exposure would be convex and his position would be antifragile: He would gain from noise, from volatility. Indeed:

$$\mathrm{E}_{\mathcal{P}} f(x) - f(\mu_{\mathcal{S}}) \quad > \quad 0 \ .$$

The expected value $\mu_{\mathcal{S}}$ underestimates what the convex exposure of the investor will translate into. Again, this is what Taleb calls the convexity bias.

4 The Taleb (antifragility / conflation) premium

The Taleb premium integrates both the non-linearity effects of f and the conflation of f for x. Under Taleb's setting, it is defined as: $\mathrm{E}_{\mathcal{P}} f(x) - \mu_{\mathcal{S}}$. Let us split this premium into two components:

$$\mathrm{E}_{\mathcal{P}} f(x) - \mu_{\mathcal{S}} \quad = \quad \underbrace{\mathrm{E}_{\mathcal{P}} f(x) - f(\mu_{\mathcal{S}})}_{(a)} + \underbrace{f(\mu_{\mathcal{S}}) - \mu_{\mathcal{S}}}_{(b)} \ .$$

(a) is Jensen's gap, which quantifies the non-linearity effects, and (b) is the conflation effect. We now formally define Taleb's premium in a more extended setting, in which the conflation effect is more general.

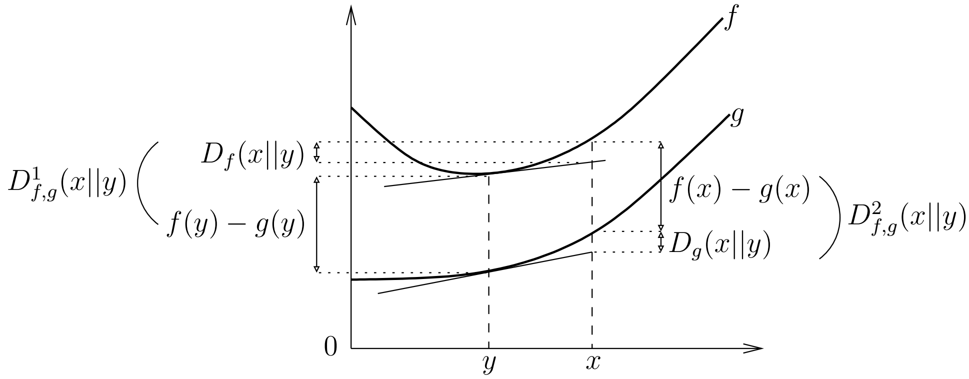

Figure 1: A schematic view of type-1 and type-2 divergences, $D^1_{f,g}(x\|y)$ and $D^2_{f,g}(x\|y)$.

4.1 Definition

Let f, g, h be three real-valued, differentiable functions. The (f, g, h)-divergence between reals $x \in \mathrm{dom}(f)$ and $y \in \mathrm{dom}(g) \cap \mathrm{dom}(h)$ is:

$$D_{f,g,h}(x\|y) \doteq f(x) - g(y) - (x-y)h'(y) \ . \quad (2)$$

When $f = g = h$ is convex, one obtains Bregman divergences (Definition 1). We specialize (2) in two skewed divergences that generalize Bregman divergences as well:

1. Type-1 divergences, in which $h = f$, noted $D^1_{f,g}(x\|y)$;
2. Type-2 divergences, in which $h = g$, noted $D^2_{f,g}(x\|y)$.

Figure 1 presents an example view of type-1 and -2 divergences when $f > g$.

Definition 3 *For any set $\mathcal{S} = \{x_1, x_2, ...\}$ equipped with discrete probabilities $\mathcal{P} = \{p_1, p_2, ...\}$, we let*

$$\mathrm{P}_{f,g}(\mathcal{S}) \doteq \mathrm{E}_{\mathcal{P}} f(x) - g(\mu_{\mathcal{S}}) \quad (3)$$

denote the T-premium, where f and g are convex differentiable, and $f \geq g$.

If one takes $g(x) = x$, we recover Taleb's original T-premium.

4.2 Some properties

Figure 2 displays the T-premium in Taleb's framework ($g(x) = x$). One sees that the T-premium integrates two parts relevant to antifragility [1]: the first one caused by the Jensen's gap, and the second one due to conflation (misunderstanding of f for g), which imposes $f(x) \geq x (= g(x))$. The following Lemma makes the observation more precise.

Lemma 1 *Letting* $\mathrm{Id}(x) \doteq x$*, we have:*

$$\begin{aligned} D^1_{f,g}(x\|y) &= D_f(x\|y) + (f(y) - g(y)) \ , & (4)\\ D^2_{f,g}(x\|y) &= D_g(x\|y) + (f(x) - g(x)) \ , & (5)\\ \mathrm{P}_{f,g}(\mathcal{S}) &= \mathrm{E}_{\mathcal{P}} D^1_{f,g}(x\|\mu_{\mathcal{S}}) \ , & (6)\\ &= \mathrm{E}_{\mathcal{P}} D^2_{f,g}(x\|\mu_{\mathcal{S}}) \ . & (7) \end{aligned}$$

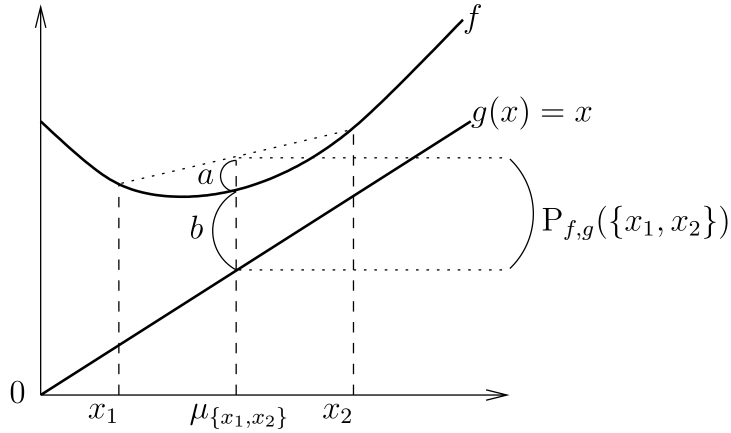

Figure 2: Decomposition of $\mathrm{P}_f(\mathcal{S})$ in two effects, the first caused by nonlinearity (a), and the second due to the conflation effect of $f(x)$ for x (b).

Eqs. (6) and (6) show that the T-premium can be written as an expected type-1 or -2 divergence with respect to the sample's average. This is a generalization of other "premia" famed in geometry, statistics, signal processing, that can be written as the average of a Bregman divergence with respect to a sample's average. Each of them quantifies the loss in information that one pays by replacing a sample by its average, thus somehow reducing his/her "field of vision" over the data.

Bregman divergence play a fundamental role in the quantification of these losses. They are remarkable because they allow to analyze the ubiquitous and polymorphous nature of information using a *single* class of measures, whatever this information be: sounds, colors, bits, probabilities, stock values, etc. [4, 5]. Bregman divergences satisfy a number of useful properties. Type-1 and -2 divergences (and so, the T-premium) turn out to satisfy interesting generalizations of these properties. Assume that f, g are strictly convex in $\mathrm{P}_{f,g}(\mathcal{S})$. The following Lemma is an example, which relies on *convex conjugates*: informally, the convex conjugate $f^\star$ of f and f have the property that their derivatives are inverse of each other.

Lemma 2 *Assume f, g strictly convex differentiable. The following holds true:*

$$D^1_{f,g}(x\|y) = D^1_{f^\star,g^\star}(f'(y)\|f'(x)) + y(f'(y) - g'(y)) \ , \quad (8)$$
$$D^2_{f,g}(x\|y) = D^2_{f^\star,g^\star}(f'(y)\|f'(x)) + x(f'(x) - g'(x)) \ . \quad (9)$$

Hence, whenever $f' = g'$,

$$D^1_{f,g}(x\|y) = D^1_{f^\star,g^\star}(f'(y)\|f'(x)) \ , \quad (10)$$
$$D^2_{f,g}(x\|y) = D^2_{f^\star,g^\star}(f'(y)\|f'(x)) \ . \quad (11)$$

The important thing about this result is that the distortion between x and y may be the same as between $f'(y)$ and $f'(x)$ (remark that parameters are switched in (10) and (11): type-1 and -2 divergences are in general not symmetric). Hence, there is a *dual representation* for x and y which may provide the same information for distortions in data. However, because of the assymetry of (10 — 11) and the differences between f and $f^\star$, confusing x for the dual representation $f'(x)$ may result in large uncontrolled errors. This is another view of the conflation error.

This dual representation is very helpful to improve the qualitative understanding of data in particular domains, and very helpful to improve models and decisions. Consider finance: type-1 and -2 divergences are a generalization of Bregman divergences. In this more specialized Bregman case for which the duality relationships (10) and (11) always hold, we can model risk premia as a generalization of Markowitz famed approach [5]. The dual representations turn out to be the spaces of returns, and the space of allocations [5]. In *one* case only, *i.e.* for a *single* choice of f, these different representations collapse as *one*. Thus, in Taleb's vocabulary [1], there is an implicit conflation made between observations of potentially different nature. This, obviously, may impair considerably the "perception" of data, increase the risk of errors, and ultimately make behaviors more fragile.
— Bullseye for Taleb[3]: this (so) particular case coincides with Markowitz setting [5].

References

[1] N.-N. Taleb, *Antifragile: things that gain from disorder*, Random House, November 2012.

[2] A. Banerjee, X. Guo, and H. Wang, "On the optimality of conditional expectation as a bregman predictor," *IEEE Trans. on Information Theory*, vol. 51, pp. 2664–2669, 2005.

[3] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Comp. Math. and Math. Phys.*, vol. 7, pp. 200–217, 1967.

[4] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*, Oxford University Press, 2000.

[5] R. Nock, B. Magdalou, E. Briys, and F. Nielsen, "On tracking portfolios with certainty equivalents on a generalization of Markowitz model: the fool, the wise and the adaptive," in *Proc. of the 28 th International Conference on Machine Learning*. 2011, pp. 73–80, Omnipress.

[3]Cf the "Central Triad" table [1].