

Convex Duality

Stephen J. Mildenhall

2026-03-15



1 The Fenchel Conjugate and Fenchel-Young Inequality

Convex duality (or the Legendre-Fenchel transform) allows us to describe a convex function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ not by its points $(x, f(x))$, but by its supporting affine functions. The **Fenchel conjugate** f^* is defined as:

$$f^*(y) = \sup_x \langle y, x \rangle - f(x).$$

The sup can be taken over the **effective domain** of a f , defined as $\text{dom } f = \{x \in \mathbb{R}^n \mid f(x) < \infty\}$. The function f is **proper** if there is at least one point where $f(x)$ is finite. Then we can write (abbreviating $\langle y, x \rangle$ to xy)

$$f^*(y) = \sup_{x \in \text{dom } f} xy - f(x).$$

For a fixed vector (slope) y , the conjugate $f^*(y)$ represents the maximum vertical distance from $f(x)$ to the linear function yx . Geometrically, this supremum occurs where the derivative $f'(x) = y$, assuming f is differentiable. If f is twice differentiable, convexity implies $f''(x) \geq 0$, but the sup involves $-f$, the critical point is a maximum. At the critical point, the line $L_y(x) = yx - f^*(y)$ is the tangent line to f with slope y . If f is not differentiable there it is a supporting affine function (line). In either case $f(x) \geq L_y(x)$ for all x .

The definition of the dual function f^* yields one of the most useful bounds in analysis. Since $f^*(y)$ is the least upper bound of $yx - f(x)$ over all x , it follows that for any specific pair (x, y) :

$$f^*(y) \geq yx - f(x).$$

Rearranging gives the **Fenchel-Young Inequality**:

$$f(x) + f^*(y) \geq xy. \tag{1}$$

The Fenchel-Young inequality has a nice economic interpretation:

- $f(x)$ = cost of producing x ,
- $\langle x, y \rangle$ = revenue from selling x at prices y , and
- $f^*(y)$ = maximal profit available at prices y .

Figure 1: Visualizing the Fenchel Conjugate for a quadratic primal.

Figure 2: Visualizing the Fenchel Conjugate for the absolute value primal. For slopes $s < -1$ or $s > 1$, f^* takes the value infinity (not shown) because $yx - |x|$ diverges.

Equation 1 says that the revenue from any chosen output cannot exceed cost plus the maximum profit achievable at those prices.

Equality in Fenchel-Young holds exactly when x attains the sup (is an optimizer) in the definition of $f^*(y)$: then $f^*(y) = \langle x, y \rangle - f(x)$, i.e.,

$$f(x) + f^*(y) = \langle x, y \rangle.$$

If pricing y is optimal for output x there is just enough revenue to cover costs and pay profits. Before moving on to subgradients and subdifferentials, here are some examples of conjugate pairs.

2 Examples and Illustrations of Conjugate Pairs

2.1 The Quadratic: $f(x) = \frac{1}{2}x^2$

The supremum of $yx - \frac{1}{2}x^2$ occurs where the derivative w.r.t x is zero: $y - x = 0 \Rightarrow x = y$. Substituting gives

$$f^*(y) = y(y) - \frac{1}{2}y^2 = \frac{1}{2}y^2.$$

The squared Euclidean norm is its own conjugate.

2.2 The Absolute Value function

Here $f(x) = |x|$ and we seek $\sup_x \{yx - |x|\}$.

- If $|y| > 1$, then as $x \rightarrow \infty$ (if $y > 1$) or $x \rightarrow -\infty$ (if $y < -1$), the term $yx - |x|$ grows without bound. Thus, $f^*(y) = \infty$.
- If $|y| \leq 1$, the maximum value is 0 (at $x = 0$).

Therefore,

$$f^*(y) = \mathbb{1}_{[-1,1]}(y) = \begin{cases} 0 & \text{if } |y| \leq 1 \\ +\infty & \text{otherwise.} \end{cases}$$

2.3 The Convex Indicator Function: $f(x) = \mathbb{1}_{[0,1]}(x)$

Here $f(x) = 0$ if $x \in [0, 1]$ and ∞ otherwise. $f^*(y) = \sup_{x \in [0,1]} \{yx - 0\}$.

- If $y > 0$, the max is at $x = 1$, so $f^*(y) = y$.
- If $y \leq 0$, the max is at $x = 0$, so $f^*(y) = 0$.

Thus, $f^*(y) = \max(0, y)$, which is the “hinge” function.

2.4 The Linear Function: $f(x) = ax$

$$f^*(y) = \sup_x \{yx - ax\} = \sup_x \{(y - a)x\}.$$

- If $y \neq a$, the supremum is ∞ .
- If $y = a$, the supremum is 0.

Thus, $f^*(y) = \mathbb{1}_{\{a\}}(y)$, which does not make a good graphic! However, by symmetry, the dual of $\mathbb{1}_{\{a\}}$ is the line. Thus, lines (and affine functions generally) are their own biconjugates.

Figure 3: Visualizing the Fenchel Conjugate for a convex indicator function primal.

Figure 4: Visualizing the Fenchel Conjugate for the convex indicator of a point is a line.

2.5 Other Examples

Here are examples for a quartic, log, exp and a W-shaped non-convex function.

3 Tables: Conjugate Pairs and the Action of Transformations on Conjugates

For a set $C \subseteq \mathbb{R}^n$, the (convex analysis) **indicator function** $I_C : \mathbb{R}^n \rightarrow \{0, \infty\}$ is defined as:

$$I_C(x) = \begin{cases} 0 & x \in C \\ +\infty & x \notin C \end{cases}.$$

This function differs from the standard probabilistic indicator (which outputs 1 or 0) to ensure that the function remains convex when C is a convex set. If C is closed and convex, I_C is a closed, proper, and convex function.

For a set $C \subseteq \mathbb{R}^n$, the convex **support function** $\sigma_C : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is defined as:

$$\sigma_C(y) = \sup_{x \in C} \langle y, x \rangle.$$

Geometrically, for a unit vector y , $\sigma_C(y)$ represents the maximum displacement of the set C in the direction y . It describes the distance from the origin to the supporting hyperplane of C orthogonal to y .

The support function is the Fenchel conjugate of the indicator function:

$$\sigma_C(y) = I_C^*(y)$$

If C is a closed convex set, the relationship is symmetric by the Fenchel-Moreau theorem, meaning $I_C(x) = \sigma_C^*(x)$.

Table 1: Conjugate pairs.

Function $f(x)$	dom f	Conjugate $f^*(y)$	dom f^*
$\langle a, x \rangle + b$	\mathbb{R}^n	$I_{\{a\}}(y) - b$	$\{a\}$
$\frac{1}{2}x^\top Qx, Q \succ 0$	\mathbb{R}^n	$\frac{1}{2}y^\top Q^{-1}y$	\mathbb{R}^n
$\frac{1}{2}\ x\ _2^2$	\mathbb{R}^n	$\frac{1}{2}\ y\ _2^2$	\mathbb{R}^n
$\ x\ $ (any norm)	\mathbb{R}^n	$I_{\{y: \ y\ _* \leq 1\}}(y)$	$\{y : \ y\ _* \leq 1\}$
$I_C(x)$	C	$\sigma_C(y) := \sup_{x \in C} \langle y, x \rangle$	\mathbb{R}^n
$\sigma_C(x)$, with C closed and convex	\mathbb{R}^n	$I_C(y)$	C
e^x	\mathbb{R}	$y \log y - y$	$[0, \infty)$
$-\log x$	$(0, \infty)$	$-1 - \log(-y)$	$(-\infty, 0)$
$x \log x$, with $0 \log 0 := 0$	$[0, \infty)$	e^{y-1}	\mathbb{R}
$\frac{1}{p}\ x\ _p^p, p > 1$	\mathbb{R}^n	$\frac{1}{q}\ y\ _q^q, \frac{1}{p} + \frac{1}{q} = 1$	\mathbb{R}^n
$\log(\sum_i e^{x_i})$	\mathbb{R}^n	$\sum_i y_i \log y_i$	$\{y \geq 0 : \sum_i y_i = 1\}$
$-\sum_i \log x_i$	\mathbb{R}_{++}^n	$-n - \sum_i \log(-y_i)$	\mathbb{R}_{--}^n
$\max(0, x)$	\mathbb{R}	$I_{[0,1]}(y)$	$[0, 1]$
$\frac{1}{x}$	$(0, \infty)$	$-2\sqrt{-y}$	$(-\infty, 0]$
$\text{dist}(x, C)$	\mathbb{R}^n	$\sigma_C(y) + I_{\{y: \ y\ _* \leq 1\}}(y)$	$\{y : \ y\ _* \leq 1\}$

Figure 5: Visualizing the Fenchel Conjugate for a quartic primal.

Figure 6: Visualizing the Fenchel Conjugate for a log primal.

Table 2: Action of transformations on conjugates.

Operation	Function $g(x)$	Conjugate $g^*(y)$	Notes on domain / qualifications
Scalar multiplication	$af(x), a > 0$	$a f^*(y/a)$	$\text{dom } g^* = a \text{ dom } f^*$
Scaling	$f(ax), a \neq 0$	$f^*(y/a)$	$\text{dom } g^* = a \text{ dom } f^*$
Translation	$f(x - b)$	$f^*(y) + \langle b, y \rangle$	same domain as f^*
Affine perturbation	$f(x) + \langle a, x \rangle + b$	$f^*(y - a) - b$	$\text{dom } g^* = a + \text{dom } f^*$
Linear map	$f(Ax + b)$	$\inf_{z: A^\top z = y} \{f^*(z) - \langle z, b \rangle\}$	finite only if $y \in A^\top(\text{dom } f^*)$
Pointwise sum	$f_1(x) + f_2(x)$	$\text{cl}(f_1^* \square f_2^*)(y)$	closure may be needed; without it, equality holds under standard qualification conditions
Infimal convolution	$(f_1 \square f_2)(x)$	$f_1^*(y) + f_2^*(y)$	domain contained in $\text{dom } f_1^* \cap \text{dom } f_2^*$
Perspective	$t f(x/t), t > 0$	$I_{\{(y,s): f^*(y)+s \leq 0\}}(y, s)$	epigraph-style indicator form

4 The Fenchel Conjugate of a General Function

Let $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be **any** function. We define its conjugate f^* as:

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \{\langle y, x \rangle - f(x)\}.$$

Even if f is a jagged, discontinuous, or highly non-convex function, the resulting f^* will always be well-behaved in one specific way: it is always convex. The proof relies on the fact that the supremum of a family of convex functions is convex.

1. For each fixed x , define the function $g_x(y) = \langle x, y \rangle - f(x)$.
2. Notice that $g_x(y)$ is an **affine function** of y (essentially a line/hyperplane). All affine functions are convex.
3. The conjugate $f^*(y)$ is the pointwise supremum of these affine functions: $f^*(y) = \sup_x g_x(y)$.
4. By the property of convex functions, the supremum of any collection of convex functions is itself convex.

5 Subgradients and Subdifferentials

Define $\partial f(x_0)$ to be the set of prices y consistent with x_0 being the profit maximizing output level, that is consistent with

$$\forall x: \langle x_0, y \rangle - f(x_0) \geq \langle x, y \rangle - f(x),$$

Rearranging,

$$\partial f(x_0) := \{y: f(x) \geq f(x_0) + \langle x - x_0, y \rangle, \forall x\}.$$

Thus:

$$y_0 \in \partial f(x_0) \iff x_0 \in \arg \max_x \langle x, y_0 \rangle - f(x).$$

A subgradient is a price vector supporting x_0 as profit-maximizing output.

For f^* , the subdifferential is simply

$$\partial f^*(y_0) = \{x: f^*(y) \geq f^*(y_0) + \langle x, y - y_0 \rangle, \forall y\}.$$

Figure 7: Visualizing the Fenchel Conjugate for an exponential primal.

Figure 8: Visualizing the Fenchel Conjugate for a non-convex primal function. The conjugate is remains convex.

Thus, $x_0 \in \partial f^*(y_0)$ iff the affine function

$$y \mapsto f^*(y_0) + \langle x_0, y - y_0 \rangle$$

supports f^* from below at y_0 . Equivalently, $x_0 \in \partial f^*(y_0)$ exactly when x_0 attains the maximum profit achievable at price y_0 , that is, $f^*(y_0) = \langle x_0, y_0 \rangle - f(x_0)$. Put another way, they are outputs for which y_0 is the optimal price.

Economically, since $f^*(y)$ is maximal profit as a function of prices, this says x_0 is the marginal response of optimal profit to changes in price. In the differentiable case,

$$\nabla f^*(y_0) = x_0,$$

so the gradient of the profit function is the optimal production plan.

6 The Biconjugate and The Fenchel-Moreau Theorem

When we take the conjugate of the conjugate, we get the **biconjugate** f^{**} . This is where the lack of convexity in the original f creates an interesting result.

If f is already convex and lower semi-continuous (lsc)¹, then $f^{**} = f$. However, if f is **non-convex**, then f^{**} is the **convex envelope** (the tightest possible convex lower bound) of f .

$$f^{**}(x) = \text{cl}(\text{conv } f)(x).$$

Imagine a non-convex function as a landscape with a deep, narrow valley. The biconjugate f^{**} is what you get if you stretched a rubber sheet below the landscape and pull it up—it bridges the non-convex gaps. $\text{conv } f$ pulls the sheet up, and cl fills in any holes where the function fails to be lower semicontinuous.

In optimization and risk theory, the difference between $f(x)$ and $f^{**}(x)$ is the gap between f and its closed convex envelope. If $f(x) = f^{**}(x)$, we have Strong Duality. This is guaranteed for convex, proper, lsc functions by the Fenchel-Moreau Theorem (see below). If $f(x) > f^{**}(x)$, there is a gap. This usually occurs in discrete optimization or non-convex problems (like those involving binary choices).

Here are formal statements of what we have been discussing.

Proposition 6.1. *The following are equivalent:*

1. $y_0 \in \partial f(x_0)$,
2. $x_0 \in \partial f^*(y_0)$,
3. $f(x_0) + f^*(y_0) = \langle x_0, y_0 \rangle$.

Proof. 1 \implies 3:

$$y_0 \in \partial f(x_0) \iff f(x) \geq f(x_0) + \langle x - x_0, y_0 \rangle \quad \forall x.$$

¹A function $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous at a point x_0 if

$$f(x_0) \leq \liminf_{n \rightarrow \infty} f(x_n)$$

whenever $x_n \rightarrow x_0$. Heuristically, f may jump down at x_0 , but not up: the value at the limit point cannot lie above the limiting lower envelope of nearby values. Equivalently, the epigraph

$$\text{epi}(f) = \{(x, t) : f(x) \leq t\}$$

is closed in \mathbb{R}^{n+1} . A proper convex function is continuous on the interior of its effective domain, so any failure of lower semicontinuity can occur only at the boundary. Duality is built on supporting hyperplanes, and these see the closed epigraph. Thus, if a convex function is not lower semicontinuous, the biconjugate replaces it by its lower-semicontinuous closure. If f is already convex and lower semicontinuous, then $f^{**} = f$.

Rearrange:

$$\langle x_0, y_0 \rangle - f(x_0) \geq \langle x, y_0 \rangle - f(x) \quad \forall x.$$

Taking the supremum over x , remembering x_0 is part of the sup, gives

$$f^*(y_0) = \langle x_0, y_0 \rangle - f(x_0),$$

which is exactly

$$f(x_0) + f^*(y_0) = \langle x_0, y_0 \rangle.$$

3 \implies 1: If equality holds, then

$$f^*(y_0) = \langle x_0, y_0 \rangle - f(x_0),$$

and since for all x ,

$$f^*(y_0) \geq \langle x, y_0 \rangle - f(x),$$

we get

$$\langle x_0, y_0 \rangle - f(x_0) \geq \langle x, y_0 \rangle - f(x),$$

which gives

$$f(x) \geq f(x_0) + \langle x - x_0, y_0 \rangle,$$

showing $y_0 \in \partial f(x_0)$.

The equivalences with $x_0 \in \partial f^*(y_0)$ follows by the same argument with the roles reversed. \square

Proposition 6.1 describes algebraic equivalences based on the definitions, independent of the properties of f . However, if f is not convex then $\partial f(x_0)$ can be empty almost everywhere, because a nonconvex function generally has no global supporting affine function at a given point. Even though the equivalences hold formally, they are much less useful without convexity.

Remark 6.1 (Economic interpretations). At fixed y :

- $f^*(y)$ is maximal profit;
- $x_0 \in \partial f^*(y)$ means x_0 is an optimal production plan at prices y .

At fixed x_0 :

- $y_0 \in \partial f(x_0)$ means y_0 is a supporting price system making x_0 optimal.

When either membership holds, so does the other, and we have equilibrium between quantity and prices:

$$f(x_0) + f^*(y_0) = \langle x_0, y_0 \rangle.$$

Revenue splits exactly into cost plus optimal profit.

Example 6.1 (The Differentiable Case). If f is differentiable at x_0 , then

$$\partial f(x_0) = \{ \nabla f(x_0) \},$$

and the unique supporting price is marginal cost:

$$y_0 = \nabla f(x_0).$$

If f^* is differentiable at y_0 , then

$$\nabla f^*(y_0) = x_0.$$

Whenever both gradients exist at corresponding points, the gradient maps are inverse in the sense that

$$y_0 = \nabla f(x_0) \iff x_0 = \nabla f^*(y_0).$$

7 The Biconjugate f^{**}

The biconjugate function is the conjugate of the conjugate, defined as

$$f^{**}(x) := \sup_y \langle x, y \rangle - f^*(y).$$

It searches for price systems that support output x . More precisely:

- for every y , $\langle x, y \rangle - f^*(y)$ is the profit-compatible lower bound (Fenchel-Young) on the cost of producing x ;
- $f^{**}(x)$ is the best such lower bound obtainable from linear pricing systems.

Thus, f^{**} is the supremum of all affine minorants of f generated by dual prices.

Recall, a function is **proper** if $f(x) < \infty$ for some x , and $f(x) > -\infty$ for all x .

Theorem 7.1. *Let $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ be proper. Then, the biconjugate f^{**} is always convex and lsc, and*

$$f^{**} \leq f.$$

In addition, if f is convex and lower semicontinuous then

$$f^{**} = f.$$

Proof. For each fixed y , the function

$$x \mapsto \langle x, y \rangle - f^*(y)$$

is affine and hence continuous, convex, and lower semicontinuous. Since f^{**} is the supremum of these affine functions it follows that f^{**} is convex and lower semicontinuous.

Next we show $f^{**} \leq f$. For any x, y ,

$$f^*(y) = \sup_u \langle u, y \rangle - f(u) \geq \langle x, y \rangle - f(x),$$

so

$$\langle x, y \rangle - f^*(y) \leq f(x).$$

Taking the supremum over y gives

$$f^{**}(x) \leq f(x).$$

Now assume f is convex and lower semicontinuous, and fix $x_0 \in \mathbb{R}^n$. To prove the reverse inequality $f(x_0) \leq f^{**}(x_0)$, let $r < f(x_0)$. Then

$$(x_0, r) \notin \text{epi}(f), \quad \text{epi}(f) := \{ (x, t) : t \geq f(x) \}.$$

Since f is convex, $\text{epi}(f)$ is convex; since f is lower semicontinuous, $\text{epi}(f)$ is closed. Hence, by the separating hyperplane theorem, there exist $(a, b) \neq (0, 0)$ and $\alpha \in \mathbb{R}$ such that

$$\langle a, x \rangle + bt \geq \alpha \quad \text{for all } (x, t) \in \text{epi}(f),$$

while

$$\langle a, x_0 \rangle + br < \alpha.$$

One must have $b > 0$: $b < 0$ is impossible because the epigraph is upward closed in the t -direction, and $b = 0$ contradicts the fact that $(x_0, f(x_0)) \in \text{epi}(f)$.

Set

$$y = -a/b, \quad c = \alpha/b.$$

Then from the separation inequality,

$$f(x) \geq \langle x, y \rangle + c \quad \text{for all } x,$$

and from the strict inequality at (x_0, r) ,

$$r < \langle x_0, y \rangle + c.$$

The first display implies

$$\langle x, y \rangle - f(x) \leq -c \quad \text{for all } x,$$

hence

$$f^*(y) \leq -c.$$

Therefore

$$\langle x_0, y \rangle - f^*(y) \geq \langle x_0, y \rangle + c > r,$$

so by definition of f^{**} ,

$$f^{**}(x_0) \geq r.$$

Since this holds for every $r < f(x_0)$, we obtain

$$f^{**}(x_0) \geq f(x_0).$$

Combined with $f^{**}(x_0) \leq f(x_0)$, this gives

$$f^{**}(x_0) = f(x_0).$$

As x_0 is arbitrary, $f^{**} = f$. □

Convexity is used in the proof to ensure $\text{epi}(f)$ is convex and lsc to ensure it is closed; together these allow application of the hyperplane separation theorem.

Remark 7.1 (Constrained optimization). The conjugate f^* first appears here as a profit function, with y interpreted as a price vector for outputs. In constrained optimization, a closely related but distinct construction appears: one introduces Lagrange multipliers y that price constraints rather than goods. These multipliers are not part of the original economic model; they are auxiliary variables used to build lower bounds on the primal problem. Under suitable conditions, optimal multipliers acquire an economic meaning as shadow prices: they measure the marginal value of relaxing the constraints. Thus Fenchel conjugacy and Lagrangian duality are different ideas, but they are tightly linked, because conjugates appear naturally when one minimizes a Lagrangian over the primal variables.

8 Inf-Convolution

The **inf-convolution** (sometimes called epi-sum) of two functions f and g , denoted by $(f \square g)$, is defined as:

$$(f \square g)(x) = \inf_{u+v=x} \{f(u) + g(v)\} = \inf_u \{f(u) + g(x-u)\}.$$

If you think of functions in terms of their epigraphs (the set of points on and above the graph), inf-convolution corresponds exactly to the Minkowski sum $\widehat{\text{minkowski}}$ of those epigraphs:

$$\text{epi}(f \square g) = \text{epi}(f) + \text{epi}(g).$$

A “Magic Formula” of convex analysis is that the conjugate of an inf-convolution is the simple sum of the conjugates:

$$(f \square g)^* = f^* + g^*.$$

This formula is always true. The converse

$$(f + g)^* = (f^* \square g^*)$$

is much more sensitive. For it to hold, you generally need the functions to be proper, convex, and lsc, and you usually need a qualification condition (like Slater’s Condition) to ensure the infimum in the convolution is actually attained and the duality gap is zero.

9 The Conjugate of a Risk Measure

We have seen that for an extended real-valued convex function f on \mathbb{R}^n , the conjugate is

$$f^*(y) = \sup_x \langle x, y \rangle - f(x).$$

Here $\langle x, y \rangle$ is the linear part, and $f(x)$ is the original convex function.

Let ρ be a convex risk measure defined on a linear space of losses X , for example on L^∞ . We use the loss convention throughout, so larger X means worse outcomes.

Since ρ is a convex function, it has an abstract conjugate on the dual space. Under the Fatou property, this yields an exact dual representation in terms of countably additive measures, or equivalently L^1 densities. Measures \mathbb{Q} replace pricing vectors y and define a linear functional on the loss space. $\mathbb{Q}(X) = \langle X, \mathbb{Q} \rangle$ plays the role of the pairing $\langle x, y \rangle$. The conjugate of ρ is

$$\rho^*(\mathbb{Q}) := \sup_X \mathbb{Q}(X) - \rho(X).$$

So the parallel is exact:

- $\mathbb{Q}(X)$ is the linear part;
- $\rho(X)$ is the primal convex functional;
- $\rho^*(\mathbb{Q})$ measures how far the linear functional \mathbb{Q} sits above ρ .

If \mathbb{Q} is represented by a density $Z = d\mathbb{Q}/dP$, then

$$\mathbb{Q}(X) = P(XZ),$$

and the same conjugate becomes

$$\rho^*(Z) = \sup_X P(XZ) - \rho(X).$$

10 The Dual Representation

With that notation, the dual representation of ρ takes the form

$$\rho(X) = \sup_{\mathbb{Q} \ll P} \mathbb{Q}(X) - \rho^*(\mathbb{Q}),$$

or equivalently

$$\rho(X) = \sup_Z P(XZ) - \rho^*(Z).$$

The dual representation of a convex risk measure is the infinite-dimensional analogue of the usual bi-conjugation theorem for convex functions. For an ordinary convex function f , equality $f^{**} = f$ holds when f is lower semicontinuous in the usual topology on \mathbb{R} . For a convex risk measure ρ on L^∞ , the corresponding topology is the weak-* topology $\sigma(L^\infty, L^1)$. In the risk-measure literature this lower-semicontinuity condition is usually expressed as the Fatou property. Under the standard assumptions, convexity plus Fatou is therefore exactly the condition needed for the conjugate dual representation to hold with countably additive dual measures or, equivalently, L^1 densities².

The literature often writes the dual term as $\alpha(\mathbb{Q})$ rather than $\rho^*(\mathbb{Q})$, to emphasize its interpretation as a penalty attached to the scenario \mathbb{Q} . But from the convex-analysis point of view, it is simply the conjugate, and it is exactly analogous to f^* .

²Strictly speaking, if one starts from abstract convex duality on L^∞ , the conjugate ρ^* naturally lives on the full Banach dual $(L^\infty)^*$. This dual is not L^1 , but ba , the space of bounded finitely additive signed measures absolutely continuous with respect to P . Thus, without additional assumptions, the abstract dual representation involves finitely additive set functions rather than countably additive probability measures. The Fatou property, or equivalently $\sigma(L^\infty, L^1)$ lower semicontinuity under the standard assumptions, is what permits one to restrict attention to the countably additive part and write the dual representation in terms of $Q \ll P$, or densities $Z \in L^1$. In this sense, finitely additive measures arise at the level of general Banach-space duality, while the Fatou property rules out the need to leave the countably additive world.

11 The Coherent Case

Now suppose ρ is coherent. In particular, ρ is positively homogeneous:

$$\rho(\lambda X) = \lambda \rho(X), \quad \lambda \geq 0.$$

This forces the conjugate ρ^* to take only the values 0 and ∞ . In other words, ρ^* must be an indicator function of a convex set of dual measures. To see why, fix a dual measure \mathbf{Q} and recall that

$$\rho^*(\mathbf{Q}) = \sup_X \mathbf{Q}(X) - \rho(X).$$

There are now two possibilities.

First, suppose there exists some X such that

$$\mathbf{Q}(X) - \rho(X) > 0.$$

Then for every $\lambda > 0$,

$$\mathbf{Q}(\lambda X) - \rho(\lambda X) = \lambda \mathbf{Q}(X) - \rho(\lambda X).$$

Using positive homogeneity of ρ ,

$$\begin{aligned} \mathbf{Q}(\lambda X) - \rho(\lambda X) &= \lambda \mathbf{Q}(X) - \lambda \rho(X) \\ &= \lambda (\mathbf{Q}(X) - \rho(X)). \end{aligned}$$

Since we assumed $\mathbf{Q}(X) - \rho(X) > 0$, this tends to ∞ as $\lambda \rightarrow \infty$. Hence

$$\rho^*(\mathbf{Q}) = \infty.$$

Second, suppose instead that

$$\mathbf{Q}(X) - \rho(X) \leq 0 \quad \text{for all } X.$$

Then automatically

$$\rho^*(\mathbf{Q}) \leq 0.$$

But taking $X = 0$ gives

$$\rho^*(\mathbf{Q}) \geq \mathbf{Q}(0) - \rho(0) = 0,$$

so in fact

$$\rho^*(\mathbf{Q}) = 0.$$

Thus there are only two possibilities:

$$\rho^*(\mathbf{Q}) = \begin{cases} 0, & \text{if } \mathbf{Q}(X) \leq \rho(X) \text{ for all } X, \\ \infty, & \text{otherwise.} \end{cases}$$

If we define

$$\mathcal{Q}_\rho := \{ \mathbf{Q} : \mathbf{Q}(X) \leq \rho(X) \text{ for all } X \},$$

then ρ^* is exactly the convex $0/\infty$ indicator function of \mathcal{Q}_ρ :

$$\rho^*(\mathbf{Q}) = I_{\mathcal{Q}_\rho}(\mathbf{Q}).$$

Substituting this into the dual representation gives

$$\rho(X) = \sup_{\mathbf{Q} \in \mathcal{Q}_\rho} \mathbf{Q}(X).$$

In the coherent case the general linear functional minus penalty representation collapses to a pure worst-case expectation over an admissible convex set of dual scenarios.

The behavior of coherent risk measures is exactly the same convex-analytic behavior of ordinary functions on \mathbb{R} . For example, let

$$f(x) = I_{[a,b]}(x) = \begin{cases} 0, & x \in [a, b], \\ \infty, & x \notin [a, b]. \end{cases}$$

Then f is convex, and its conjugate is

$$f^*(y) = \sup_{x \in [a,b]} xy.$$

So the indicator of a convex set becomes the support function of that set.

Here the same mechanism is running in the opposite direction:

- a coherent risk measure ρ is sublinear;
- the conjugate of a sublinear functional is an indicator;
- therefore ρ is the support function of a convex dual set.

So coherent risk measures stand to their dual sets exactly as support functions stand to convex sets in elementary convex analysis.

12 Interpretation

A coherent risk measure is a monotone cash-additive sublinear functional on losses. The sublinearity is the decisive point here: for a general convex risk measure, the conjugate ρ^* can assign different finite penalties to different dual scenarios; for a coherent risk measure, positive homogeneity rules that out. A scenario is either admissible, in which case the penalty is 0, or inadmissible, in which case the penalty is ∞ .

Thus:

- for a general convex risk measure, one has

$$\rho(X) = \sup_{\mathbb{Q}} \mathbb{Q}(X) - \rho^*(\mathbb{Q});$$

- for a coherent risk measure, one has

$$\rho(X) = \sup_{\mathbb{Q} \in \mathcal{Q}_\rho} \mathbb{Q}(X).$$

So a coherent risk measure is simply the worst expected loss over a convex admissible set of scenarios.

13 Lower Semicontinuity and the Fatou Property

To obtain an exact dual representation, one needs the appropriate lower-semicontinuity property. In the risk-measure literature this is usually expressed as the Fatou property. On L^∞ , the Fatou property says that if X_n are uniformly bounded and $X_n \rightarrow X$ almost surely, then

$$\rho(X) \leq \liminf_n \rho(X_n).$$

This is the probabilistically natural form of lower semicontinuity: the risk at the limit cannot exceed the liminf of the risks along the approximating sequence. In this sense, risk may jump down at the limit, but not up. This is exactly what one wants for risk management: if a uniformly bounded sequence X_n converges almost surely to X , and if each X_n satisfies the risk bound $\rho(X_n) \leq c$, then the Fatou property implies $\rho(X) \leq c$: risk control is preserved under passage to the limit.

From the convex-analysis point of view, the corresponding topology is the weak-* topology $\sigma(L^\infty, L^1)$. Under the standard assumptions for convex monetary risk measures on L^∞ , the Fatou property is equivalent to weak-* lower semicontinuity. Thus Fatou plays, in this setting, the same role that lower semicontinuity plays in ordinary convex duality: it ensures that the biconjugate representation is exact and that the dual variables may be taken to be countably additive measures, or equivalently L^1 densities.

In finite dimensions, proper convex functions are automatically continuous on the interior of their effective domains, so lower semicontinuity issues arise only at the boundary. On L^∞ , the analogous subtlety is topological rather than geometric: the key issue is whether the functional is closed in the relevant weak-* topology. For convex risk measures this closure is expressed by the Fatou property. A striking theorem of Elyès Jouini et al. (2006) shows that law invariance implies Fatou for monetary convex risk measures under the usual assumptions on Ω . Thus, within the law-invariant setting, the lower-semicontinuity needed for exact dual representation is automatic. In general, the Fatou property is

precisely the extra closure condition needed to rule out those boundary-type pathologies and recover an exact dual representation on countably additive measures.

14 Summary Tables

Table 3: Terms from convex analysis.

Concept	Primal Requirement	Dual Result
Fenchel Conjugate f^*	Any function f	Always Convex
Biconjugate f^{**}	Any function f	Convex Envelope of f
Strong Duality	Convex, proper, lsc	$f = f^{**}$
Duality Gap	Non-convexity	$f - f^{**} > 0$

15 Appendix: Dual representations on ba and L^1 , and the question of attainment

A convex risk measure on L^∞ admits two closely related dual stories. The first is the raw convex-analytic story: since the Banach dual of L^∞ is ba, the space of bounded finitely additive signed measures absolutely continuous with respect to \mathbb{P} , Fenchel-Moreau duality naturally produces a representation over ba. The second is the probabilistic story: under an additional regularity condition, usually the Fatou property, the same risk measure admits an equivalent representation using only countably additive measures, or equivalently L^1 densities. A separate issue is whether the supremum in the dual representation is actually achieved by some dual variable. These two questions are logically distinct. One asks what the correct dual space is; the other asks whether an optimizer exists in that space. Both matter in applications, especially for coherent and spectral risk measures.

Let $\rho : L^\infty \rightarrow \mathbb{R} \cup \infty$ be proper and convex. The most basic conjugate is

$$\alpha(\mu) = \sup_{X \in L^\infty} \mu(X) - \rho(X), \quad \mu \in \text{ba},$$

and the corresponding biconjugate formula is

$$\rho(X) = \sup_{\mu \in \text{ba}} \mu(X) - \alpha(\mu),$$

provided ρ is lower semicontinuous for the relevant duality. This is the abstract dual representation. It is called abstract because it is the representation guaranteed by general convex analysis before one imposes additional probabilistic structure. The dual variables are continuous linear functionals on L^∞ ; as a matter of functional analysis those are elements of ba, not just of L^1 . Thus “abstract attainment” means attainment of the supremum at this most general level: for a fixed X , there exists some $\mu_X \in \text{ba}$ such that

$$\rho(X) = \mu_X(X) - \alpha(\mu_X).$$

Nothing more is meant. One is not yet asserting that μ_X is countably additive, or that it has a density, or even that it has a direct probabilistic interpretation. One is only saying that the supporting functional exists in the full topological dual.

This distinction matters because existence of a supporting functional is often easier in ba than in L^1 . The reason is compactness. The weak-* topology coming from the dual pair (L^∞, ba) is designed precisely so that closed bounded pieces of the dual are compact enough for convex duality to work. If the penalty α has weak-* compact lower level sets, then for each fixed X the map

$$\mu \mapsto \mu(X) - \alpha(\mu)$$

is weak-* upper semicontinuous, so the supremum is attained. This is the convex-analytic mechanism behind abstract attainment. One should think of it as the dual analogue of saying that a continuous function on a compact set attains its maximum. The set is not compact in norm, but it is compact in the weak-* topology, and that is enough.

There is, however, an important difference between saying that a dual optimizer exists in ba and saying that it is countably additive. Countably additive measures are the honest probability measures of probability theory. Finitely additive measures are more general linear functionals. They obey finite additivity but need not preserve limits of monotone sequences, so they need not interact well with almost sure convergence. This is why they are often described as pathological from the probabilistic point of view. But that label needs care. They are not pathological in convex analysis; they are the correct dual objects. Nor are they rare in important examples. They arise naturally whenever the risk measure probes extreme tail behavior in a way that is not fully captured by L^1 densities.

The canonical example is the essential supremum. The coherent risk measure

$$\rho(X) = \text{ess sup}(X)$$

is perfectly natural and extremely important. It is the limiting case of many distortion and spectral constructions, and it appears whenever one places nonzero weight on the very worst outcomes. Yet its supporting functionals are not generally represented by countably additive probabilities. Intuitively, to support the essential supremum at a point X , the dual variable must concentrate entirely on the set where X is as large as possible. If that set is not an atom of positive probability, no countably additive probability can put full mass there while remaining absolutely continuous with respect to \mathbb{P} . A finitely additive measure can. So one should not say that ba only encodes pathologies. It also encodes genuine worst-case support for perfectly natural risk measures. Distortion risk measures with a jump at 0, equivalently a mass at the extreme tail in the associated spectral description, are exactly of this kind. They contain an essential-supremum component, and that component is what forces finitely additive support in the most general dual description.

This point is worth making carefully. In the spectral language, a risk measure with a mass at 0 in the weighting measure assigns positive weight to the very highest quantiles, in the limit to the worst loss itself. That is not an exotic feature; it is often economically meaningful. The difficulty is only that the worst point may not be attained on a set of positive probability. When it is not, the supporting dual object that “looks only at the worst tail” cannot be countably additive. Thus finitely additive measures are best thought of not as strange intruders, but as the closure of countably additive tail selectors under the dual topology. They are the objects needed to represent limiting support at the edge of the loss distribution.

The passage from ba to countably additive measures is governed by regularity. For convex monetary risk measures on L^∞ , the standard condition is the Fatou property:

$$X_n \rightarrow X \text{ a.s.}, \sup_n |X_n|_\infty < \infty \implies \rho(X) \leq \liminf_n \rho(X_n).$$

This is the natural lower semicontinuity condition for bounded almost sure convergence. It says that risk cannot jump upward at the limit of a bounded a.s. convergent sequence. Under the usual assumptions, this is equivalent to weak-* lower semicontinuity for the pairing between L^∞ and L^1 , and it allows one to replace the ba representation by an L^1 representation:

$$\rho(X) = \sup_{Z \in L^1} \mathbb{P}[XZ] - \alpha(Z),$$

or, in the monetary-risk-measure convention,

$$\rho(X) = \sup_{Z \in L^1_+, \mathbb{P}Z=1} \mathbb{P}[XZ] - \alpha(Z).$$

Equivalently, one may write the representation over countably additive probabilities $Q \ll P$ with Radon-Nikodym density $Z = dQ/dP$. This is the probabilistically meaningful representation. The role of the Fatou property is precisely to rule out the need for purely finitely additive support.

In that sense the ba representation is always the background theorem, and the L^1 representation is the sharpened theorem obtained once one adds regularity. The former comes from the geometry of the

Banach space; the latter comes from compatibility with probabilistic limits. This is why, in the literature on convex risk measures, one so often sees a progression from a general Fenchel-Moreau statement on the full dual to a more concrete theorem under the Fatou property. It is not that the first theorem is wrong or useless. Rather, it is more general than what probability theory by itself would suggest, because L^∞ has a larger dual than L^1 .

One should also separate very clearly the issue of representation from the issue of attainment. Even if a risk measure admits a representation over countably additive measures, the supremum need not be achieved by any single countably additive optimizer. There may only exist maximizing sequences. Conversely, the supremum may be attained in ba even when no countably additive optimizer exists. Thus the move from sup to max is an additional compactness question, not a mere consequence of having an L^1 representation.

In the coherent case this is especially transparent. If ρ is coherent, then the penalty collapses to an indicator of a convex dual set, and the representation takes the form

$$\rho(X) = \sup_{Q \in \mathcal{Q}} Q(X)$$

for some admissible family \mathcal{Q} of dual measures. The distinction between sup and max is now simply the distinction between taking the supremum of a continuous linear functional over a set and asking whether the set is compact enough for the maximum to be attained. If \mathcal{Q} is weakly compact in the relevant topology, then for each fixed X there exists a worst-case measure $Q_X \in \mathcal{Q}$ such that

$$\rho(X) = Q_X(X).$$

If \mathcal{Q} is merely closed and convex but not compact, then one may have sequences Q_n with $Q_n(X) \uparrow \rho(X)$ and no optimizer.

In the countably additive setting the standard compactness criterion is uniform integrability. If \mathcal{Q} is represented by a set of densities

$$\mathcal{Z} = \left\{ Z = \frac{dQ}{dP} : Q \in \mathcal{Q} \right\} \subset L_+^1,$$

then weak compactness in L^1 is typically obtained from boundedness together with uniform integrability, by the Dunford-Pettis theorem. This is the natural condition ensuring that the supremum over countably additive measures is actually attained. Without uniform integrability, mass may drift further and further into thinner and thinner tails, producing no limit in L^1 . One then gets approximate worst-case scenarios but not a genuine worst-case density.

This phenomenon has a clear financial meaning. A maximizing sequence of densities may concentrate on ever more adverse portions of the loss distribution, but the limiting object may no longer be an honest density. In the countably additive category the limit escapes. In the larger space ba it may reappear as a finitely additive measure. This is another way to understand the role of ba : it is the natural compactification of the probabilistic dual domain. What does not converge in L^1 may still converge in the weak-* topology of the larger dual space. Then the abstract supremum becomes an abstract maximum.

From this point of view, finitely additive measures arise for two rather different reasons. The first is geometric: they are required because $(L^\infty)^* = \text{ba}$. The second is limit-theoretic: they often represent limits of increasingly concentrated countably additive tail measures. The essential supremum illustrates both at once. For each $\varepsilon > 0$, one may choose a countably additive measure that puts all its mass on a set where X lies within ε of its essential supremum. Such measures form a maximizing family, but unless the supremum is actually attained on a positive-probability atom, no countably additive maximizer exists. In the weak-* closure, however, there is a finitely additive selector that supports the exact essential supremum. This is precisely the sense in which ba closes the gap between sup and max .

The law-invariant case simplifies the picture. One of the central results of E. Jouini et al. (2008) is that, on L^∞ , law invariance already forces enough regularity to imply the Fatou property for convex monetary risk measures. Hence law-invariant convex risk measures admit dual representations by countably additive measures. This includes the usual coherent and spectral examples, apart from the subtleties just noted at the extreme worst-case edge. Thus for most economically meaningful risk measures used in actuarial and

financial applications, the L^1 representation is the right working representation. The ba representation remains the correct general background, and it becomes indispensable when discussing limiting worst-case components such as essential-supremum terms.

There is therefore no contradiction between saying that finitely additive measures are often probabilistically undesirable and saying that they are indispensable in some natural examples. They are undesirable if one wants dual variables to behave like ordinary probabilities under limits. They are indispensable if one wants complete convex-analytic closure, or exact support for worst-case functionals that live on the edge of the distribution. A good way to say this is that L^1 captures ordinary probabilistic pricing, while ba captures the weak-* closure of such pricing rules, including limiting tail selectors.

For applications it helps to keep four statements distinct. First, every proper convex lower semicontinuous functional on L^∞ has an abstract dual representation on ba. Second, if the risk measure has the Fatou property, then this representation can be written using only countably additive measures, or L^1 densities. Third, even in the countably additive representation the dual supremum need not be attained unless the admissible density set has additional compactness, typically via uniform integrability. Fourth, if countably additive attainment fails, abstract attainment may still hold in ba, with the optimizer interpreted as a finitely additive supporting functional. These four points together are the full story behind the recurrent questions “why ba?” and “when do we have max instead of sup?”

For the purposes of this monograph, the practical lesson is simple. When discussing duality at the level of general convex analysis on L^∞ , the natural dual space is ba, and abstract maximizers may live there. When discussing economically meaningful and law-invariant risk measures with good continuity properties, one usually works with countably additive measures and L^1 densities. When discussing worst-case components, such as essential-supremum terms or spectral weights with mass at the extreme tail, one should expect finitely additive support to re-enter the picture, at least as a limiting or closure device. Finally, whenever a dual representation is written with a supremum, the question of whether the supremum is achieved is a separate compactness question and should be treated as such, not taken for granted from the representation itself.

16 Appendix: Distinguishing Examples

Here a set of canonical examples. Each one isolates one phenomenon. An important distinction is between functions and equivalence classes of functions mod null sets. On bounded measurable functions, you still see individual points of ω , so Dirac masses are available and many worst-case functionals are countably additive. On L^∞ , you quotient out null sets, so pointwise information disappears. Then the same-looking functional may lose its countably additive maximizer, and finitely additive charges appear as the missing closure. We start with a handy reference table.

We use the loss sign convention, positive values are bad. To convert to the payout convention simply substitute $-X$ for X . Föllmer and Schied (2016) use the payout convention.

16.1 Table of Distinguishing Examples

Table 4 lays out several risk measures that distinguish between the various properties we have introduced. In the table and remainder of this appendix, M_1 means countably additive probability measures. $M_{1,f}$ means finitely additive probability charges. Fatou means lower semicontinuity under bounded a.s. convergence.

Table 4: Important distinguishing examples of risk measures.

Example ρ	Dual space / representation	Limit behavior	What it teaches
P[X]	Single countably additive model; density $Z = 1$	Continuous from above and below; Fatou; weak-* lsc	The fully regular benchmark; everything works and the dual max is attained trivially

Example ρ	Dual space / representation	Limit behavior	What it teaches
$\rho_\gamma(X) = \gamma^{-1} \log \mathbb{P} e^{-\gamma X}$	Robust rep on M_1 or L^1 : $\sup_Q \{Q(X) - \gamma^{-1} H(Q P)\}$	Good limit behavior; Fatou; optimizer exists in M_1	Prototype convex, non-coherent risk measure with genuine penalty
$\max\{Q_1(X), Q_2(X)\}$	Coherent rep on convex hull of $\{Q_1, Q_2\} \subset M_1$	Good limit behavior; max attained	Prototype coherent robust representation over countably additive models
$\mu(X)$ with μ a free-ultrafilter charge on \mathbb{N}	Representation only on $M_{1,f}$; not on M_1	Fails continuity from above; e.g. $1_{A_m} \downarrow 0$ but $\mu(A_m) = 1$	Canonical example showing finitely additive \neq countably additive; failure is about tail limits
$\text{ess sup}(X)$ on nonatomic L^∞	Sup over M_1 exists, but max may fail there; max restored on $M_{1,f}$	Fatou holds; countably additive representation exists; countably additive optimizer may not exist	Canonical example for sup over M_1 , max over $M_{1,f}$; finitely additive support is natural, not merely pathological
$\sup_\omega(X(\omega))$ on bounded measurable functions	Often represented by Dirac masses $\delta_\omega \in M_1$ if supremum is attained pointwise	Pointwise worst-case evaluation on the raw function space	Distinguishes measurable functions from L^∞ classes: before quotienting by null sets, point masses can be honest maximizers

We now discuss these examples in more detail.

16.2 Plain Vanilla: Ordinary Expectation

Take

$$\rho(X) = \mathbb{P}[X].$$

This is coherent, linear, continuous from above, continuous from below, Fatou, weak-* lower semicontinuous, and its dual set is the singleton \mathbb{P} , or in density form the singleton 1.

This example teaches nothing subtle, but it is the reference point. Every limit property one could reasonably ask for holds, and the dual representation is on a countably additive measure with an attained maximum.

When you want to remember what the good case looks like, this is it.

16.3 Chocolate: Convex, Well-Behaved, But Not Coherent

Take

$$\rho_\gamma(X) = \frac{1}{\gamma} \log \mathbb{P}(e^{-\gamma X}), \quad \gamma > 0.$$

Then

$$\rho_\gamma(X) = \sup_{Q \ll P} \left\{ Q(X) - \frac{1}{\gamma} H(Q | P) \right\},$$

where $H(Q | P)$ is relative entropy.

This is the standard robust representation: many countably additive models, with a genuine penalty rather than an indicator. It has Fatou, weak-* lower semicontinuity, and the supremum is attained by a countably additive optimizer with density proportional to $e^{-\gamma X}$.

This is the prototype for convex but not coherent, and everything still works in L^1 .

16.4 Strawberry: Coherent Robust Example

The term **robust** reflects the interpretation that risk is assessed not under a single reference model, but against model uncertainty represented by a family of alternative measures and a penalty for departing from the reference law.

Take two probability measures Q_1, Q_2 and set

$$\rho(X) = \max \{ Q_1(X), Q_2(X) \}.$$

This is coherent, and its dual set is the convex hull of Q_1, Q_2 . Again everything is countably additive, and the maximum is attained.

This is the prototype for Follmer-Schied Proposition 4.15: coherent means worst case over a convex set of models, with no genuine penalty beyond the indicator of that set.

16.5 Hazelnut: $M_{1,f}$ versus M_1 : a Free Ultrafilter On \mathbb{N}

Let $\Omega = \mathbb{N}$, let $\mathcal{F} = 2^{\mathbb{N}}$, and let μ be a finitely additive probability extending the Frechet filter, so that every cofinite set has μ -mass 1 and every finite set has mass 0. Define

$$\rho(X) = \mu(X), \quad X \in \ell^\infty.$$

This is coherent and perfectly legitimate from the convex-analytic point of view. It has a dual representation with a maximum on $M_{1,f}$, indeed by the single measure μ itself. But it is not representable by countably additive probabilities.

Why not? Because countable additivity is exactly what fails on tail limits. If

$$A_m = \{ m, m+1, m+2, \dots \},$$

then $1_{A_m} \downarrow 0$ pointwise, but

$$\mu(A_m) = 1 \quad \text{for all } m.$$

So the associated functional is not continuous from above, and Fatou-type regularity fails.

This is the canonical example showing that finitely additive and countably additive really are different, and that the difference is exactly about monotone limits.

16.6 Pistachio: Sup Over M_1 Vs Max Over $M_{1,f}$

Now take a nonatomic probability space, say $([0, 1], \mathcal{B}, \lambda)$, and define

$$\rho(X) = \text{ess sup}(X).$$

This is coherent and law invariant. It has the Fatou property, hence a representation over countably additive probabilities absolutely continuous with respect to λ :

$$\rho(X) = \sup_{Q \ll \lambda} Q(X).$$

But in general the supremum is not attained by any countably additive Q .

For a concrete example, take

$$X(\omega) = \omega.$$

Then

$$\rho(X) = \text{ess sup}(\omega) = 1.$$

For every countably additive $Q \ll \lambda$,

$$Q(\omega) < 1,$$

because the only way to get expectation exactly 1 would be to put all mass at $\omega = 1$, and that is impossible under $Q \ll \lambda$ since $\lambda(1) = 0$.

But the supremum is still 1, because one can choose Q_n concentrated on $[1 - 1/n, 1]$ and get expectations tending to 1.

So here is the phenomenon:

- over countably additive measures, one has a supremum but no maximizer;
- over finitely additive measures, one gets an actual maximizer.

This is the best example of sup versus max. It also shows why finitely additive duals are not merely pathological. They are exactly what closes up the missing worst-case selector at the boundary.

16.7 Pistachio With a Twist

Now forget the quotient space L^∞ and work on bounded measurable functions on $[0, 1]$. Define

$$\rho(X) = \sup_{\omega \in [0,1]} (X(\omega)).$$

Then the maximum is attained by a Dirac mass:

$$\rho(X) = \max_{\omega \in [0,1]} \delta_\omega(X)$$

whenever the supremum is attained by the function.

So on the raw function space, worst-case evaluation is often represented by honest countably additive point masses. On the quotient L^∞ , those point masses are not absolutely continuous with respect to λ , and pointwise location no longer makes sense modulo null sets. That is exactly why the same intuition becomes ess sup, and the maximizer may disappear from M_1 and reappear only in $M_{1,f}$.

This is the canonical example distinguishing bounded measurable functions from equivalence classes in L^∞ .

16.8 What Each Flavor Teaches

Expectation teaches the fully regular countably additive case.

Entropic risk teaches the convex robust case with a genuine penalty and an attained countably additive optimizer.

Finite worst-case mixtures teach the coherent countably additive case.

The ultrafilter example teaches that finitely additive functionals appear exactly when monotone-limit regularity fails.

Essential supremum on L^∞ teaches that even with Fatou and countably additive representation, the dual supremum need not be attained in M_1 ; the missing maximizer lives in the finitely additive closure.

Worst-case evaluation on bounded measurable functions teaches that some of this subtlety is created by passing to equivalence classes modulo null sets.

16.9 How the Limit Properties Line Up

Continuity from above is the strongest monotone-limit condition in this story. It rules out ultrafilter-type tail selectors and forces countable additivity.

The Fatou property is the lower-semicontinuity version appropriate to bounded almost sure convergence. It is weaker in spirit than full order continuity, but still strong enough on L^∞ to recover countably additive dual representations.

Weak-* lower semicontinuity is the Banach-space reformulation of the same idea on L^∞ .

Failure of these properties means that the functional can react to limiting tail structure that ordinary probabilities do not see. That is when finitely additive measures enter.

References

- Föllmer, H., & Schied, A. (2016). *Stochastic Finance: An Introduction in Discrete Time* (Fourth.). Berlin, Boston: Walter de Gruyter. <https://doi.org/10.1017/CBO9781107415324.004>
- Jouini, Elyès, Schachermayer, W., & Touzi, N. (2006). Law invariant risk measures have the Fatou property. In S. Kusuoka & A. Yamazaki (Eds.), *Advances in Mathematical Economics* (Vol. 9, pp. 49–71). Springer Japan. https://doi.org/10.1007/4-431-34342-3/_4

Jouini, E., Schachermayer, W., & Touzi, N. (2008). Optimal risk sharing for law invariant monetary utility functions. *Mathematical Finance*, 18(2), 269–292. <https://doi.org/10.1111/j.1467-9965.2007.00332.x>