

# LAW OF LARGE NUMBERS FOR RISK MEASURES

FREDDY DELBAEN

ABSTRACT. Under appropriate integrability conditions the risk measure of the sample measures for a law invariant risk measure converge almost surely to the risk measure of the sampled random variable. The results follow from general convergence theorems based on the theory of Orlicz spaces.

## 1. INTRODUCTION AND NOTATION

The notation we use is the standard notation of probability theory.  $(\Omega, \mathcal{F}, \mathbb{P})$  will denote a probability space without atoms. If  $\mathbb{Q}$  is a probability defined on  $\mathcal{F}$  and absolutely continuous with respect to  $\mathbb{P}$ , we will identify  $\mathbb{Q}$  with its Radon-Nikodým derivative  $h = d\mathbb{Q}/d\mathbb{P}$ . A basic example of such an atomless space is  $([0, 1], \mathcal{B}, m)$  where  $\mathcal{B}$  is the Borel sigma algebra and  $m$  is Lebesgue measure. This space will also play a fundamental role in the sequel.

Because of neganumerophobia we will use a setup closer to premium calculation. For a convex closed set of probability densities,  $\mathcal{S} \subset L^1$  we denote by  $\rho$  the functional (whenever defined)

$$\rho(\xi) = \sup\{\mathbb{E}[\xi h \mid h \in \mathcal{S}]\}.$$

This functional is up to sign changes the same as a coherent monetary utility function. The interpretation of  $\rho$  is the same as for the utility functions. Among the many uses we mention risk measures and insurance premium principles defined on random variables. A basic problem is the following. Suppose we see a large sample of a random variable. Can we define the functional  $\rho$  on such samples and does this procedure converge to the risk measure of the random variable? The best is to put the problem in the framework of risk measures that are law invariant (law determined or rearrangement invariant). That means that when  $\eta$  and  $\eta'$  have the same distribution, then  $\rho(\eta) = \rho(\eta')$ . Since we will be working on atomless probability spaces, this property is also equivalent to the property that  $\mathcal{S}$  is rearrangement invariant. Law determined risk measures can be defined for probability distributions on  $\mathbb{R}$  with compact support (the latter to avoid integrability problems). It therefore does not matter on which probability space we work. In later sections we will start with a risk measure defined for random variables defined on  $\Omega$  but then we will “transport” it to random variables defined on  $[0, 1]$ . For more information on rearrangement invariant measures see [9] where the general form of these risk measures is given.

The risk measures we will use are sometimes defined on bigger spaces than  $L^\infty$ , the space of bounded random variables. The best way to study this is to use Orlicz spaces. For the theory of these spaces we refer to [10]. In this book the reader will find a very general approach. We will restrict ourselves to Young functions that are more regular than the ones in [10]. For our presentation this restriction is of no importance but it avoids a lot of – in

---

*Date:* First version April 26, 2020, this version July 14, 2020.

*AMS-CLASSIFICATION* 90B50, 91B06, 91B16, 91G99

our opinion cosmetic –problems. When we say  $\Phi$  is a Young function, we mean that  $\Phi$  is an even real valued function defined on  $\mathbb{R}$  with the help of its derivative  $\phi$ . This derivative function  $\phi$  will be nondecreasing on  $\mathbb{R}_+$ , continuous at 0,  $\phi(x) = 0$  if and only if  $x = 0$  and  $\lim_{x \rightarrow +\infty} \phi(x) = +\infty$ . We do not require more regularity for  $\phi$ . The conjugate Young function  $\Psi$  is then defined using the right continuous inverse function,  $\psi$ , of  $\phi$ . More precisely for  $x \geq 0$ :  $\psi(x) = \inf\{u \mid \phi(u) > x\}$ . Hence  $\Psi(x) = \int_0^{|x|} \psi(u) du$ . We realise that the class of Orlicz spaces so defined, does not contain  $L^1$  or  $L^\infty$ , so when appropriate we will make additional remarks. We follow [10] and define

$$\begin{aligned} L^\Phi &= \{\xi \mid \text{there is } k > 0 \text{ with } \mathbb{E}[\Phi(k\xi)] < \infty\} \\ M^\Phi &= \{\xi \mid \text{for all } k > 0 \text{ we have } \mathbb{E}[\Phi(k\xi)] < \infty\} = L^{(\Phi)} \subset L^\Phi. \end{aligned}$$

These spaces are usually equipped with the Luxemburg norm

$$\|\xi\|_\Phi = \inf \left\{ \alpha > 0 \mid \mathbb{E} \left[ \Phi \left( \frac{\xi}{\alpha} \right) \right] \leq 1 \right\}.$$

The dual space of  $M^\Phi$  is the space  $L^\Psi$  and we have for  $\xi \in L^\Phi, \eta \in L^\Psi$  that  $\xi\eta \in L^1$  and  $|\mathbb{E}[\xi\eta]| \leq 2\|\xi\|_\Phi\|\eta\|_\Psi$ . The dual space of  $L^\Phi$  is not necessarily  $L^\Psi$ . Indeed we have that  $(L^\Phi)^* = L^\Psi$  if and only if  $L^\Phi = M^\Phi$  if and only if  $\Phi$  satisfies the  $\Delta_2$  condition: there is  $x_0$  and a constant  $C$  such that for all  $x \geq x_0$  the inequality  $\Phi(2x) \leq C\Phi(x)$  is valid. If this is the case one can modify the function  $\Phi$  without changing the spaces and such that in the inequality we may suppose that  $x_0 = 0$ .

## 2. MAIN RESULT

In this section we suppose that the scenario set  $\mathcal{S}$ , for the risk measure  $\rho$ , is supposed to be a subset of  $L^\Phi$ . The set  $\mathcal{S}$  is a convex closed subset of  $L^1$  and hence it must be bounded in  $L^\Phi$ . Indeed, if  $h_n \in \mathcal{S}$  is such that  $\|h_n\|_\Phi$  is an unbounded sequence, say  $\|h_n\|_\Phi \geq 4^n$ , then the element  $h = \sum_n \frac{h_n}{2^n}$  is still in  $\mathcal{S}$  but we have for each  $n$ :  $\|h\|_\Phi \geq \|h_n 2^{-n}\|_\Phi \geq 2^{-n} 4^n = 2^n$ . This would mean that  $\mathcal{S}$  is not a subset of  $L^\Phi$ . The set  $\mathcal{S}$  is also weak\*, meaning  $\sigma(L^\Phi, M^\Psi)$ , compact. To see this we must only show that it is weak\* closed. This follows from the (obvious) continuity of the injection  $L^\Phi \rightarrow L^1$  for the topologies  $\sigma(L^\Phi, M^\Psi)$  and  $\sigma(L^1, L^\infty)$ , just observe that  $L^\infty \subset M^\Phi$ . This continuity also implies that  $\mathcal{S}$  is closed in  $L^\Phi$ . We remark that when  $\mathcal{S} \subset L^1$  is a convex set of probabilities that is closed in  $L^\Phi$ , this does not necessarily imply that  $\mathcal{S}$  must be closed in  $L^1$ .

Because of this boundedness, the functional  $\rho$  with scenario set  $\mathcal{S}$  can be defined on  $L^\Psi$ . Indeed for  $\eta \in L^\Psi$  we have

$$|\rho(\eta)| = |\sup\{\mathbb{E}[\eta h] \mid h \in \mathcal{S}\}| \leq 2\|\eta\|_\Psi \sup_{h \in \mathcal{S}} \|h\|_\Phi < \infty.$$

As expected this boundedness also implies that  $\rho$  is continuous on  $L^\Psi$ . For completeness we include the obvious proof.

**Lemma 1.** *If  $\eta_n$  is a sequence in  $L^\Psi$  tending to  $\eta$  for  $\|\cdot\|_\Psi$ , then  $\rho(\eta_n) \rightarrow \rho(\eta)$ . More precisely  $|\rho(\eta) - \rho(\eta')| \leq 2 \sup_{h \in \mathcal{S}} \|h\|_\Phi \|\eta - \eta'\|_\Psi$ .*

**Proof.** We only have to prove the inequality. For given  $\varepsilon > 0$  and  $\eta$  we choose  $h \in \mathcal{S}$  with  $\rho(\eta) \leq \mathbb{E}[h\eta] + \varepsilon$ . This gives

$$\rho(\eta) \leq \mathbb{E}[h\eta] + \varepsilon = \mathbb{E}[h\eta'] + \mathbb{E}[h(\eta - \eta')] + \varepsilon \leq \rho(\eta') + 2\|h\|_\Phi \|\eta - \eta'\|_\Psi + \varepsilon.$$

Because  $\varepsilon > 0$  was arbitrary we get  $\rho(\eta) \leq \rho(\eta') + 2 \sup_{h \in \mathcal{S}} \|h\|_{\Phi} \|\eta - \eta'\|_{\Psi}$ . We can interchange  $\eta$  and  $\eta'$  and then get the desired result.  $\square$

*From now on we suppose that  $\rho$  is rearrangement invariant.*

For a probability measure  $\mu$  on  $\mathbb{R}$ , we use the notation  $\rho(\mu)$  for the outcome  $\rho(\eta)$ , where  $\mu$  is the distribution of  $\eta$ . When  $\mathcal{S} \subset L^{\Phi}$ , this is defined as soon as  $\int_{\mathbb{R}} \Psi(x) d\mu < \infty$ , i.e.  $\eta \in L^{\Psi}$ .

For probability laws on  $\mathbb{R}$  there is an easy way to transform convergence of probability distributions into convergence of random variables (a special case of the Skorohod lemma). Suppose that  $\mu_n$  converges weak\* to  $\mu$ , i.e. for each bounded continuous function  $g$  defined on  $\mathbb{R}$  we have  $\int_{\mathbb{R}} g(x) d\mu_n \rightarrow \int_{\mathbb{R}} g(x) d\mu$ . The distribution functions are defined as  $F_n(x) = \mu_n(-\infty, x]$ ,  $F(x) = \mu(-\infty, x]$ . Weak\* convergence is then equivalent to  $F_n(x) \rightarrow F(x)$  for each continuity point  $x$  of  $F$ . The quantile functions are defined as  $q_n(x) = \inf\{u \mid F_n(u) > x\}$ ,  $q(x) = \inf\{u \mid F(u) > x\}$ . The quantile functions,  $q_n$ , are defined on the unit interval  $(0, 1)$  and under the Lebesgue measure they have the distribution functions  $\mu_n$ . The convergence of  $\mu_n$  is then transformed into almost everywhere convergence of the quantile functions. Convergence  $q_n \rightarrow q$  in the Orlicz space space  $L^{\Psi}$  is not easily defined in terms of the measures  $\mu_n$ . The following is a partial answer to this question

**Proposition 1.** *Suppose that  $\mu_n$  converges weak\* to  $\mu$ . If for each  $k > 0$ ,  $\int_{\mathbb{R}} \Psi(kx) \mu_n(dx) \rightarrow \int_{\mathbb{R}} \Psi(kx) \mu(dx) < \infty$ , we have  $\|q_n - q\|_{\Psi} \rightarrow 0$ . Consequently  $\rho(\mu_n) \rightarrow \rho(\mu)$ .*

**Proof.** We have  $\rho(\mu_n) = \rho(q_n)$ ,  $\rho(\mu) = \rho(q)$  so the last line follows from the lemma. To prove the norm convergence we have to estimate  $\int_{[0,1]} \Psi(k|q_n(u) - q(u)|) du$ . We use the inequalities

$$\int \Psi(k|q_n - q|) \leq \int \Psi(2k \max(q_n, q)) \leq \int \Psi(2kq_n) + \int \Psi(2kq).$$

Because  $q_n \rightarrow q$  a.s. and  $\int \Psi(2kq_n) \rightarrow \int \Psi(2kq)$ , Scheffé lemma implies that there is  $n_0$  such that  $\Psi(2kq_n)$ ,  $n \geq n_0$  forms a uniformly integrable family. Consequently the elements  $\int \Psi(k|q_n - q|)$  (for  $n \geq n_0$ ) form a uniformly integrable family. Therefore for every  $k$ :  $\int_{[0,1]} \Psi(k|q_n(u) - q(u)|) du \rightarrow 0$ . This shows that  $\lim_n \|q_n - q\|_{\Psi} \rightarrow 0$ .  $\square$

*Remark 1.* The hypothesis implies that  $q \in M^{\Psi}$  but there is no need to suppose that  $q_n \in M^{\Psi}$ .

We are now ready to state the main result of this section.

**Theorem 1.** *We use the notation and assumptions from the beginning of this section. Let  $\xi \in M^{\Psi}$ . Let  $\xi_n: (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$  be an iid sequence with the same distribution  $\mu$  as  $\xi$ . For each  $N$  we define the sample distribution as  $\mu_n = \frac{1}{N} \sum_{n=1}^N \delta_{\xi_n}$  where  $\delta_y$  denotes the Dirac measure concentrated in  $y$ . Then almost surely  $\rho(\mu_n) \rightarrow \rho(\mu)$ .*

**Proof.** The usual law of large numbers allows to conclude that almost everywhere the random measures  $\frac{1}{N} \sum_{n=1}^N \delta_{\xi_n}$  converge weak\* to  $\mu$ . The variant of Skorohod's theorem then permits to find for almost every  $\omega \in \Omega$ , increasing functions  $q_N$  on  $[0, 1]$  such that  $q_N$  has as its distribution  $\mu_n = \frac{1}{N} \sum_{n=1}^N \delta_{\xi_n}$ . They converge almost everywhere (on  $[0, 1]$ ) to  $q$ , the increasing rearrangement of  $\xi$ . Again the law of large numbers shows that almost surely for every  $0 < K \in \mathbb{N}$ ,  $\frac{1}{N} \sum_{n=1}^N \Psi(K\xi_n)$  converges to  $\mathbb{E}[\Psi(K\xi)]$  (where the set of measure 1 can

be chosen to be independent of  $K$ ). On  $[0, 1]$  this translates as  $\int_{[0,1]} \Psi(Kq_N) \rightarrow \int_{[0,1]} \Psi(Kq)$ . We now apply the previous proposition.  $\square$

*Remark 2.* With minor notational modifications, the above proof also works for the case  $\Psi(x) = x$ , i.e. the  $L^1$  space. In particular it also proves the theorem for the tail expectation. In this case the result was proved by Van Zwet, [?].

### 3. A MORE GENERAL RESULT

In the section we generalise the result of the preceding section to the case of random variables in  $L^\Psi$  (and not just  $M^\Psi$ ). We need an additional hypothesis on the scenarioset. We recall the result of Ando, [3], which characterises the  $\sigma(L^\Phi, L^\Psi)$  relative sequentially compact sets.

**Theorem 2.** (Ando) *A set  $A \subset L^\Phi$  is  $\sigma(L^\Phi, L^\Psi)$  relative sequentially compact if and only if*

$$\lim_{\lambda \rightarrow \infty} \sup_{h \in A} \lambda \mathbb{E} \left[ \Phi \left( \frac{h}{\lambda} \right) \right] = 0.$$

*In particular if  $\Psi$  is  $\Delta_2$  then bounded sets  $A \subset L^\Phi$  automatically satisfy the above limit condition.*

We use the same notation as in the previous sections. In particular  $\rho$  is defined on  $L^\Psi$  and is rearrangement invariant.

**Proposition 2.** *Suppose that  $\mathcal{S}$  satisfies the conditions of Ando's theorem. Suppose that  $\mu_n$  converges weak\* to  $\mu$ . If for some  $k > 0$ ,  $\int_{\mathbb{R}} \Psi(kx) \mu_n(dx) \rightarrow \int_{\mathbb{R}} \Psi(kx) \mu(dx) < \infty$ , we have  $\sup_{h \in \mathcal{S}} \mathbb{E}[h|q_n - q|] \rightarrow 0$ . Consequently  $\rho(\mu_n) \rightarrow \rho(\mu)$ .*

**Proof.** By homogeneity we may suppose that  $k = 2$  and that  $\int_{\mathbb{R}} \Psi(2x) \mu_n(dx) \rightarrow \int_{\mathbb{R}} \Psi(2x) \mu(dx)$ . We may also suppose that for all  $n$ :  $\int_{\mathbb{R}} \Psi(2x) \mu_n(dx) < \infty$ . We replace the integrals on  $\mathbb{R}$  by integrals of the quantile functions and get that  $\int_{[0,1]} \Psi(2q_n) \rightarrow \int_{[0,1]} \Psi(2q)$ . Scheffé's lemma then implies that the sequence  $\Psi(2q_n)$  is uniformly integrable. The obvious inequality  $\Psi(|q_n - q|) \leq \Psi(2q_n) + \Psi(2q)$  then implies that the sequence  $\Psi(|q_n - q|)$  is also uniformly integrable. Hence  $\int_{[0,1]} \Psi(|q_n - q|) \rightarrow 0$ , since  $q_n \rightarrow q$  a.s. . We now have

$$\sup_{h \in \mathcal{S}} \int_{[0,1]} h|q_N - q| \rightarrow 0.$$

For given  $\epsilon > 0$  we take  $\lambda$  so that  $\lambda \int_{[0,1]} \Phi(\frac{h}{\lambda}) \leq \epsilon$ . Then we take  $N_0$  so that for  $N \geq N_0$ :  $\lambda \int_{[0,1]} \Psi(|q_N - q|) \leq \epsilon$ . Young's inequality now yields  $h|q_N - q| \leq \lambda \Phi(\frac{h}{\lambda}) + \lambda \Psi(|q_N - q|)$  and after integration we get for  $N \geq N_0$ :

$$\sup_{h \in \mathcal{S}} \int_{[0,1]} h|q_N - \eta| \leq 2\epsilon.$$

$\square$

**Theorem 3.** *Suppose that the scenarioset  $\mathcal{S}$  satisfies the condition of Ando's theorem. Let  $\xi \in L^\Psi$ . Let  $\xi_n: (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$  be an iid sequence with the same distribution  $\mu$  as  $\xi$ . For each  $N$  we define the sample distribution as  $\mu_n = \frac{1}{N} \sum_{n=1}^N \delta_{\xi_n}$  where  $\delta_y$  denotes the Dirac measure concentrated in  $y$ . Then almost surely  $\rho(\mu_N) \rightarrow \rho(\mu)$ .*

**Proof.** We already know that almost surely  $\mu_N$  converges weak\* to  $\mu$ . Without loss of generality we may replace  $\xi$  by a multiple of it and hence we may suppose that  $\mathbb{E}[\Psi(2\xi)] \leq 1$ . The law of large numbers implies that:

$$\frac{1}{N} \sum_{k=1}^N \Psi(2\xi_k) \rightarrow \mathbb{E}[\Psi(2\xi)],$$

almost surely and in  $L^1$ . We can now apply the previous proposition.  $\square$

#### 4. THE CASE OF CHOQUET INTEGRATION

Choquet integration provides us with one of the most used examples of rearrangement invariant risk measures. As proved by David Schmeidler, [12], [8], a rearrangement invariant risk measure that is commonotone is given by the core of a convex game and hence uses a scenarioset of the form

$$\mathcal{S} = \{h \in L^1 \mid \mathbb{E}[h] = 1 \text{ and for each } A \in \mathcal{F} : \mathbb{E}[h\mathbf{1}_A] \geq f(\mathbb{P}[A])\},$$

where  $f$  is a convex function  $f: [0, 1] \rightarrow [0, 1]$ ,  $f(0) = 0$ ,  $f(1) = 1$ . In our context, where we want the risk measure to be defined on a larger space than  $L^\infty$  we must use a weakly ( $\sigma(L^1, L^\infty)$ ) compact set  $\mathcal{S}$ . This is equivalent to the left continuity of  $f$  at the point 1. Ryff's theorem, [11], [8], then says that  $\mathcal{S}$  is the  $L^1$ -closed convex hull of random variables having the same law as the derivative  $f'$  of  $f$ . The statement  $\mathcal{S} \subset L^\Phi$  is equivalent to  $f' \in L^\Phi$ . One can show that when  $f' \in L^\Phi[0, 1]$ , the closed convex hull taken in  $L^\Phi$  is the same as when the closedness is taken in  $L^1$ . The set  $\mathcal{S}$  is also the set of those nonnegative random variables  $h$  such that for each convex function  $\beta: \mathbb{R}_+ \rightarrow \mathbb{R}$  we have  $\mathbb{E}[\beta(h)] \leq \int_{[0,1]} \beta(f')$  (convex dominance). Because of this dominance the set  $\mathcal{S}$  automatically satisfies Ando's criterium for  $\sigma(L^\Phi, L^\Psi)$  sequential compactness. As a result the topologies  $\sigma(L^\Phi, L^\Psi)$  and  $\sigma(L^1, L^\infty)$  coincide on the set  $\mathcal{S}$ . We get that the statistical results of the previous section apply to this case. For iid sequences we find the almost sure convergence

$$\xi_{[n:1]} f\left(\frac{1}{n}\right) + \sum_{k=1}^{k=n-1} \xi_{[n:k+1]} \left( f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right) \right) = \int_0^1 q_n f' \rightarrow \int_0^1 q f' = \rho(\xi)$$

where  $\xi_{[n:1]}, \dots, \xi_{[n:n]}$  are the observations  $\xi_1, \dots, \xi_n$  in increasing order. This result can also be obtained using integration theory.

#### REFERENCES

- [1] Amir Ahmadi-Javid: Entropic Value-at-Risk: A New Coherent Risk Measure, *Journ. of Optim. Theory and Appl.*, **155**, 1105–1123, (2012)
- [2] Amir Ahmadi-Javid: Addendum to Entropic Value-at-Risk: A New Coherent Risk Measure, *Journ. of Optim. Theory and Appl.*, **155**, 1124–1128, (2012)
- [3] Ando, Tsuyoshi: Weakly Compact Sets in Orlicz Spaces, *Can. J. Math.* **14**, 170–176, (1962)
- [4] Artzner, Ph., F. Delbaen, J.-M. Eber, and D. Heath: Thinking Coherently, *RISK*, **November 97**, 68–71, (1997)
- [5] Artzner, Ph., F. Delbaen, J.-M. Eber, and D. Heath: Characterisation of Coherent Risk Measures, *Mathematical Finance* **9**, 145–175, (1999)
- [6] Delbaen, F.: Coherent Risk Measures on General Probability Spaces *in Advances in Finance and Stochastics*, pp. 137, Springer, Berlin (2002)

- [7] Delbaen, F: Coherent Risk Measures, *Lectures given at the Cattedra Galileiana at the Scuola Normale Superiore di Pisa, March 2000*, Published by the *Scuola Normale Superiore di Pisa*, (2002)
- [8] Delbaen, F: Monetary Utility Functions, *Lectures held in 2008 and published in the series "Lectures Notes of the University of Osaka"*, (2011)
- [9] Kusuoka, S.: On Law Invariant Coherent Risk Measures, *Advances in Mathematical Economics*, **3**, 83–95, (2001)
- [10] Rao, M.,M., Ren, Z.D.:Theory of Orlicz Spaces Pure and Applied Mathematics, **146**, Marcel Dekker, (1991), New York, Basel, Hong Kong
- [11] Ryff, J. V.: Extreme points of some convex subsets of  $L^1(0, 1)$ . *Proc. Amer. Math. Soc.*, **18**, 10261034, (1967)
- [12] Schmeidler, D.: Integral Representation without Additivity, *Proc. Amer. Math. Soc.*,**97**, pp. 255–261, (1986)

DEPARTEMENT FÜR MATHEMATIK, ETH ZÜRICH, RÄMISTRASSE 101, 8092 ZÜRICH, SWITZERLAND

INSTITUT FÜR MATHEMATIK, UNIVERSITÄT ZÜRICH, WINTERTHURERSTRASSE 190, 8057 ZÜRICH, SWITZERLAND