# Expected Shortfall

Fabian Baier

Technische Universität München

8. September 2025

# Structure

1. Basic definitions and intuition
2. Translation invariance and positive homogeneity
3. Prerequisites for the proofs
4. Important proofs for subadditivity

# Assumptions throughout the presentation

- An **atomless** probability space $(\Omega, \mathcal{F}, \mathbb{P})$, i.e. a $U[0,1]$ random variable exists on it.
- All random variables are defined on $(\Omega, \mathcal{F}, \mathbb{P})$.
- $X$ is a *real-valued* r.v. representing **loss**: $X > 0$ loss, $X < 0$ profit.
- Spaces: $L^0$ (all r.v.'s), $L^1$ (integrable), $L^\infty$ (bounded).
- Notation: $F_X$ is the distribution of $X$ (under $\mathbb{P}$).

# Definitions: Value at Risk (VaR) and Expected Shortfall (ES)

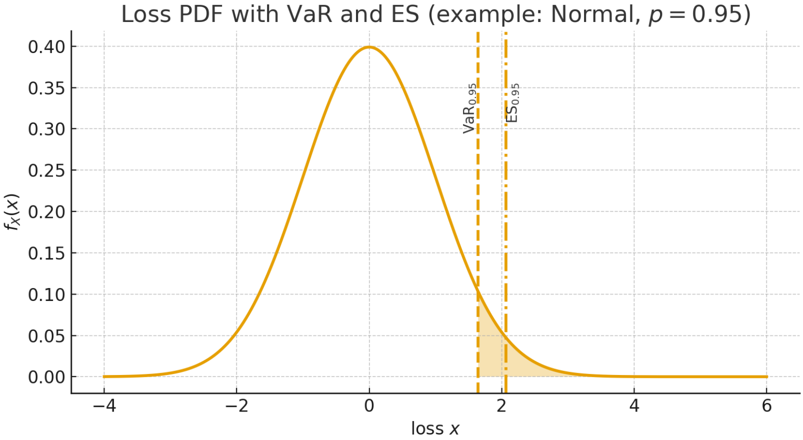
For $p \in (0,1)$, the risk measures $\mathrm{VaR}_p : L^0 \to \mathbb{R}$ and $\mathrm{ES}_p : L^0 \to \mathbb{R} \cup \{+\infty\}$ are defined as

$$\mathrm{VaR}_p(X) = \inf\{x \in \mathbb{R} : \ \mathbb{P}(X \leq x) \geq p\}, \quad X \in L^0,$$

$$\mathrm{ES}_p(X) = \frac{1}{1-p} \int_p^1 \mathrm{VaR}_q(X)\, dq, \quad X \in L^0.$$

ES is also known as Conditional Value at Risk (CVaR) and Average Value at Risk (AVaR).

# Illustration: VaR and ES on a loss PDF




Example with $X \sim \mathcal{N}(0,1)$ and $p = 0.95$. The shaded region is $\{x \geq \mathrm{VaR}_p\}$. Dashed line: $\mathrm{VaR}_p$. Dash–dot line: $\mathrm{ES}_p = \mathbb{E}[X \mid X \geq \mathrm{VaR}_p]$.

# Coherent risk measure axioms

Let $\rho : \mathcal{X} \subseteq L^0 \to \mathbb{R} \cup \{+\infty\}$ be a risk measure that assigns capital to a loss r.v. $X$. For any $X, Y \in \mathcal{X}$, $c \in \mathbb{R}$, $\lambda \geq 0$, and $\theta \in [0, 1]$:

- **Monotonicity:** $X \leq Y$ a.s. $\Rightarrow \rho(X) \leq \rho(Y)$.
- **Translation invariance (cash–invariance):** $\rho(X + c) = \rho(X) + c$.
- **Positive homogeneity:** $\rho(\lambda X) = \lambda \, \rho(X)$.
- **Subadditivity:** $\rho(X + Y) \leq \rho(X) + \rho(Y)$.
- **Convexity:** $\rho(\theta X + (1 - \theta) Y) \leq \theta \rho(X) + (1 - \theta) \rho(Y)$.

## Monetary risk measure with positive homogenity: convexity $\iff$ subadditivity

Assume $\rho$ is positively homogeneous.

**(Subadditivity $\Rightarrow$ Convexity)** For $\theta \in [0,1]$,

$$\rho(\theta X + (1-\theta)Y) \leq \rho(\theta X) + \rho((1-\theta)Y) = \theta\rho(X) + (1-\theta)\rho(Y).$$

**(Convexity $\Rightarrow$ Subadditivity)** Using $\theta = \frac{1}{2}$,

$$\rho(X+Y) = 2\,\rho\left(\frac{X+Y}{2}\right) \leq 2\left(\frac{1}{2}\rho(X) + \frac{1}{2}\rho(Y)\right) = \rho(X) + \rho(Y).$$

Thus, it is sufficient for our purposes to prove that ES satisfies subadditivity.

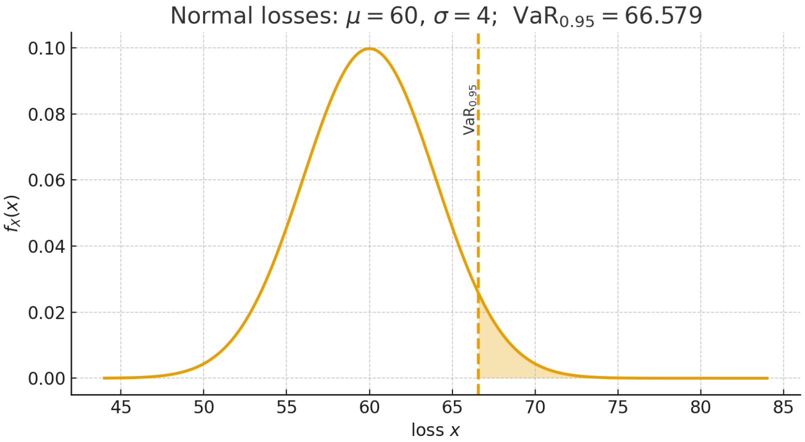
# Example: normal distributed losses — VaR

Assume $X \sim \mathcal{N}(\mu, \sigma^2)$ is a loss r.v. Then

$$F_X(x) = \mathrm{P}(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

$$\begin{aligned}
\mathrm{VaR}_p(X) &= \inf\{\, x : \ \mathrm{P}(X \leq x) \geq p \,\} \\
&= \inf\left\{\, x : \ \Phi\left(\frac{x-\mu}{\sigma}\right) \geq p \,\right\} \\
&= \mu + \sigma\, \Phi^{-1}(p).
\end{aligned}$$

# Example: continued




Parameters: $\mu = 60$, $\sigma = 4$. Here $\mathrm{VaR}_{0.95} = \mu + \sigma\,\Phi^{-1}(0.95) \approx 66.58$.
So there is a 5% chance that the loss exceeds $\mathrm{VaR}_{0.95}(X)$, i.e.

$$\mathrm{P}(X > \mathrm{VaR}_{0.95}(X)) = 0.05.$$

# Example: normal distributed losses — ES

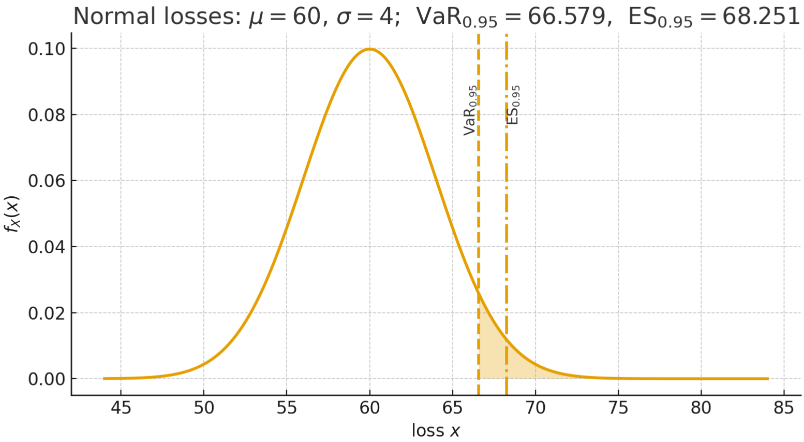
Assume $X \sim \mathcal{N}(\mu, \sigma^2)$ and $p \in (0,1)$. Using

$$\mathrm{ES}_p(X) = \frac{1}{1-p} \int_p^1 \mathrm{VaR}_q(X)\, dq \quad \text{and} \quad \mathrm{VaR}_q(X) = \mu + \sigma\, \Phi^{-1}(q),$$

set $q = \Phi(t)$ so $dq = \varphi(t)\, dt$ and $t$ runs from $z_p = \Phi^{-1}(p)$ to $\infty$:

$$\begin{aligned} \mathrm{ES}_p(X) &= \frac{1}{1-p} \int_p^1 (\mu + \sigma\, \Phi^{-1}(q))\, dq \\ &= \mu + \frac{\sigma}{1-p} \int_{z_p}^{\infty} t\, \varphi(t)\, dt = \mu + \sigma\, \frac{\varphi(z_p)}{1-p}, \qquad z_p = \Phi^{-1}(p). \end{aligned}$$

# Example: continued




Parameters: $\mu = 60$, $\sigma = 4$. Here $\mathrm{VaR}_{0.95} \approx 66.58$ and $\mathrm{ES}_{0.95} \approx 68.25$.

Interpretation: $\mathrm{ES}_{0.95}(X) = \mathrm{E}[X \mid X \geq \mathrm{VaR}_{0.95}(X)]$ is the *average loss* in the worst 5% of cases.

# VaR: translation invariance & positive homogeneity

For $p \in (0,1)$ and any loss r.v. $X$, $c \in \mathbb{R}$, $\lambda \geq 0$:

$$\mathrm{VaR}_p(X + c) = \mathrm{VaR}_p(X) + c, \qquad \mathrm{VaR}_p(\lambda X) = \lambda\,\mathrm{VaR}_p(X).$$

Definition used: $\mathrm{VaR}_p(X) = \inf\{\, x : \; \mathrm{P}(X \leq x) \geq p \,\}$.

# Proof: translation invariance of VaR

**Claim.** $\mathrm{VaR}_p(X + c) = \mathrm{VaR}_p(X) + c$ for any $c \in \mathbb{R}$.
**Proof.** Let $F_X(x) = \mathrm{P}(X \leq x)$.

$$F_{X+c}(x) = \mathrm{P}(X + c \leq x) = \mathrm{P}(X \leq x - c) = F_X(x - c).$$

Hence

$$\begin{aligned}
\mathrm{VaR}_p(X + c) &= \inf\{\, x : \; F_{X+c}(x) \geq p \,\} \\
&= \inf\{\, x : \; F_X(x - c) \geq p \,\} \\
&= c + \inf\{\, y : \; F_X(y) \geq p \,\} \\
&= c + \mathrm{VaR}_p(X).
\end{aligned}$$

$\square$

# Proof: positive homogeneity of VaR

**Claim.** $\mathrm{VaR}_p(\lambda X) = \lambda\,\mathrm{VaR}_p(X)$ for $\lambda \geq 0$.
**Proof.** If $\lambda = 0$, then $\lambda X = 0$ a.s., so $\mathrm{VaR}_p(0) = 0 = \lambda\,\mathrm{VaR}_p(X)$.
For $\lambda > 0$, using $F_X(x) = \mathrm{P}(X \leq x)$,

$$F_{\lambda X}(x) = \mathrm{P}(\lambda X \leq x) = \mathrm{P}\left(X \leq \tfrac{x}{\lambda}\right) = F_X\left(\tfrac{x}{\lambda}\right).$$

Therefore

$$\begin{aligned}
\mathrm{VaR}_p(\lambda X) &= \inf\{\, x : \; F_{\lambda X}(x) \geq p \,\} \\
&= \inf\{\, x : \; F_X(x/\lambda) \geq p \,\} \\
&= \lambda \inf\{\, y : \; F_X(y) \geq p \,\} \\
&= \lambda\,\mathrm{VaR}_p(X).
\end{aligned}$$

$\square$

# ES inherits translation invariance & positive homogeneity

Recall $\mathrm{ES}_p(X) = \dfrac{1}{1-p}\displaystyle\int_p^1 \mathrm{VaR}_q(X)\,dq$. Using the identities for VaR:

$$\mathrm{ES}_p(X+c) = \frac{1}{1-p}\int_p^1 \left(\mathrm{VaR}_q(X)+c\right)dq = \mathrm{ES}_p(X)+c,$$

$$\mathrm{ES}_p(\lambda X) = \frac{1}{1-p}\int_p^1 \lambda\,\mathrm{VaR}_q(X)\,dq = \lambda\,\mathrm{ES}_p(X) \qquad (\lambda \geq 0).$$

These equalities hold in the extended real sense (if $\mathrm{ES}_p(X) = +\infty$, both sides are $+\infty$).

# Why VaR can fail subadditivity (counterexample)

Let $X, Y$ be independent losses with

$$\mathrm{P}(X=0)=0.96, \quad \mathrm{P}(X=10)=0.04, \qquad Y \text{ i.i.d. to } X.$$

At level $p=0.95$:

$$\mathrm{VaR}_{0.95}(X)=0, \qquad \mathrm{VaR}_{0.95}(Y)=0.$$

But for $Z=X+Y$,

$$\mathrm{P}(Z=0)=0.9216, \quad \mathrm{P}(Z=10)=0.0768, \quad \mathrm{P}(Z=20)=0.0016,$$

hence

$$\mathrm{VaR}_{0.95}(Z)=10.$$

Therefore

$$\mathrm{VaR}_{0.95}(X+Y)=10 \; > \; 0+0=\mathrm{VaR}_{0.95}(X)+\mathrm{VaR}_{0.95}(Y),$$

so VaR is *not* subadditive.

# Prerequisites for the proofs

# Generalized inverse (quantile)

Let $F_X(x) = \mathrm{P}(X \leq x)$ be a cdf (nondecreasing, right–continuous). For $p \in (0,1)$ the *generalized inverse / left–continuous quantile* is

$$F_X^{-1}(p) \;=\; \inf\{\, x : \; F_X(x) \geq p \,\}.$$

This quantity is the basis of $\mathrm{VaR}_p(X)$: $\mathrm{VaR}_p(X) = F_X^{-1}(p)$.

# Generalized inverse: key properties

For $p \in (0,1)$ and any cdf $F_X$:

- **Monotone in $p$:** if $p_1 \leq p_2$, then $F_X^{-1}(p_1) \leq F_X^{-1}(p_2)$.
- **Left–continuous in $p$.**
- **Inequalities:**

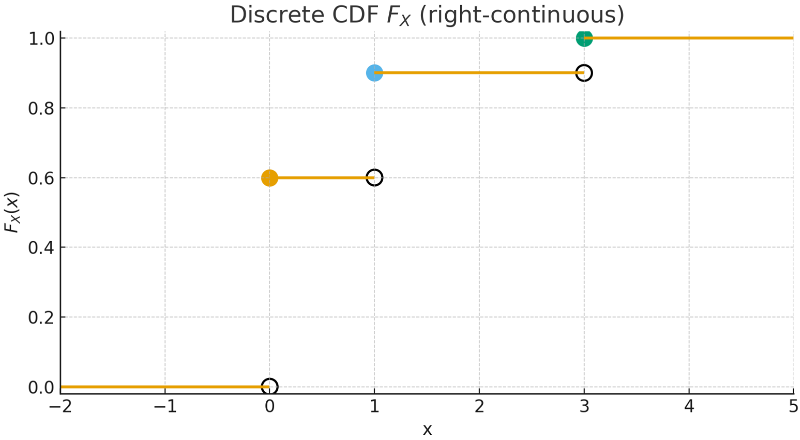
$$F_X(F_X^{-1}(p)) \;\geq\; p, \qquad F_X^{-1}(F_X(x)) \;\leq\; x.$$

- If $F_X$ is *continuous and strictly increasing*, then it is a true inverse:

$$F_X(F_X^{-1}(p)) = p, \qquad F_X^{-1}(F_X(x)) = x.$$

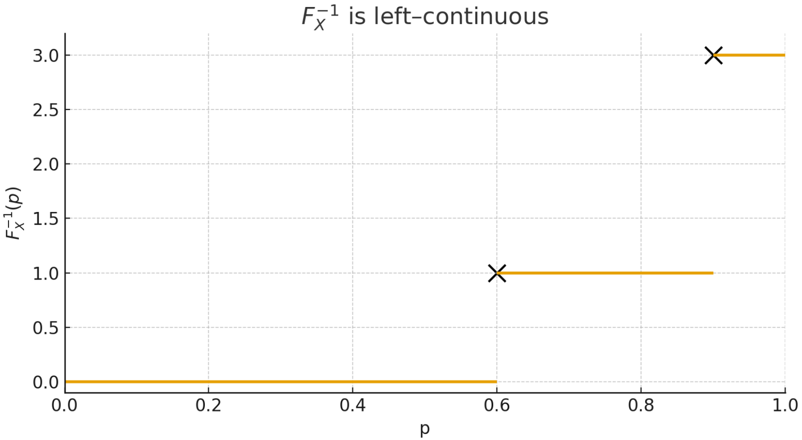
# Generalized inverse example: CDF view

Consider a discrete loss $X$ with $\mathrm{P}(X = 0) = 0.6$, $\mathrm{P}(X = 1) = 0.3$, $\mathrm{P}(X = 3) = 0.1$. Then $F_X$ jumps to 0.6 at $x = 0$, to 0.9 at $x = 1$, and to 1 at $x = 3$.




Reading off at $p = 0.75$ gives $F_X^{-1}(0.75) = 1$ as the first $x$ with $F_X(x) \geq 0.75$.

# Generalized inverse example: quantile view




If the cdf is flat, then the generalized inverse has a jump — and *vice versa*.

# Lemma 1 (Quantile transform)

For any random variable $X$ with cdf $F_X(x) = \mathrm{P}(X \leq x)$, there exists $U_X \sim U[0,1]$ such that

$$X \;=\; F_X^{-1}(U_X) \quad \text{almost surely.}$$

**Remark:**
$F_X(X) \sim U[0,1] \quad \Longleftrightarrow \quad F_X$ is continuous.
In particular, if $F_X$ is continuous, one may take $U_X = F_X(X)$.

# Distributional equality $F_X^{-1}(U) \stackrel{d}{=} X$

We are proving a slightly weaker statement.
Let $F_X(x) = \mathrm{P}(X \leq x)$ be a cdf and define the generalized inverse $F_X^{-1}(p) = \inf\{x : F_X(x) \geq p\}$ for $p \in (0,1)$. Let $U \sim U[0,1]$.
For any real $t$,

$$\{F_X^{-1}(U) \leq t\} \iff \{U \leq F_X(t)\} \quad (\text{since } F_X^{-1}(p) \leq t \iff p \leq F_X(t))$$

Hence

$$\mathrm{P}\big(F_X^{-1}(U) \leq t\big) = \mathrm{P}\big(U \leq F_X(t)\big) = F_X(t),$$

because $U$ is uniform on $[0,1]$ and $F_X(t) \in [0,1]$. Thus the cdf of $F_X^{-1}(U)$ equals $F_X$, so

$$F_X^{-1}(U) \stackrel{d}{=} X.$$

# Lemma 2 (representation of ES)

For $p \in (0,1)$ and an integrable loss r.v. $X \in L^1$,

$$\mathrm{ES}_p(X) = \mathrm{VaR}_p(X) + \frac{1}{1-p}\,\mathrm{E}[(X - \mathrm{VaR}_p(X))_+]\,.$$

Here $(x)_+ := \max\{x, 0\}$.

*Interpretation.*

- $\mathrm{VaR}_p(X)$ is the loss threshold leaving tail probability $1-p$.
- $(X - \mathrm{VaR}_p(X))_+$ is the *exceedance* above that threshold.
- $\frac{1}{1-p}\,\mathrm{E}[(X - \mathrm{VaR}_p(X))_+] = \mathrm{E}[X - \mathrm{VaR}_p(X) \mid X > \mathrm{VaR}_p(X)]$ for continuous $F_X$ (in general: the average tail exceedance).

Thus, $\mathrm{ES}_p$ = "threshold" + "average exceedance" = the mean loss in the worst $1-p$ fraction of cases.

# Lemma 2 — Proof

$$\begin{aligned}\mathrm{ES}_p(X) &= \frac{1}{1-p}\int_p^1 \mathrm{VaR}_q(X)\,dq \\ &= \mathrm{VaR}_p(X) + \frac{1}{1-p}\int_p^1 \left(\mathrm{VaR}_q(X) - \mathrm{VaR}_p(X)\right) dq.\end{aligned}$$

Write $\mathrm{VaR}_q(X) = F_X^{-1}(q)$ and let $U \sim U[0,1]$. Since $F_X^{-1}$ is nondecreasing,

$$\begin{aligned}\int_p^1 \left(F_X^{-1}(q) - F_X^{-1}(p)\right) dq &= \int_0^1 \left(F_X^{-1}(q) - F_X^{-1}(p)\right)_+ dq \\ &= \mathrm{E}\left[\left(F_X^{-1}(U) - F_X^{-1}(p)\right)_+\right].\end{aligned}$$

By Lemma 1, $F_X^{-1}(U) \overset{d}{=} X$, hence

$$\mathrm{ES}_p(X) = \mathrm{VaR}_p(X) + \frac{1}{1-p}\mathrm{E}[(X - \mathrm{VaR}_p(X))_+].$$

# Lemma 3 (tail-split identity for ES)

For $p \in (0,1)$ and any integrable loss r.v. $X \in L^1$,

$$(1-p)\,\mathrm{ES}_p(X) = \mathrm{E}\Big[X\,\mathbf{1}_{\{X>\mathrm{VaR}_p(X)\}}\Big] + \mathrm{VaR}_p(X)\,\big(\mathrm{P}(X \le \mathrm{VaR}_p(X)) - p\big).$$

- $\mathrm{E}\Big[X\,\mathbf{1}_{\{X>\mathrm{VaR}_p(X)\}}\Big]$ is the contribution of all exceedances above the VaR threshold (the strict tail).
- $\mathrm{VaR}_p(X)(\mathrm{P}(X \le \mathrm{VaR}_p(X)) - p)$ accounts for an atom at the VaR level and vanishes when $F_X$ is continuous.
- Hence, if $F_X$ is continuous,

$$\begin{aligned}(1-p)\,\mathrm{ES}_p(X) &= \mathrm{E}\big[X\,\mathbf{1}_{\{X>\mathrm{VaR}_p(X)\}}\big] \\ &\Rightarrow\ \mathrm{ES}_p(X) = \mathrm{E}[X \,|\, X > \mathrm{VaR}_p(X)]\,.\end{aligned}$$

i.e., ES is the average loss in the worst $1-p$ fraction of outcomes.

# Lemma 3 — Proof

Let $v := \mathrm{VaR}_p(X)$. From Lemma 2,

$$\mathrm{ES}_p(X) = v + \frac{1}{1-p}\,\mathrm{E}[(X-v)_+]\,.$$

Multiplying by $(1-p)$ and expanding $(X-v)_+ = (X-v)\mathbf{1}_{\{X>v\}}$ gives

$$\begin{aligned}(1-p)\mathrm{ES}_p(X) &= (1-p)v + \mathrm{E}\Big[(X-v)\mathbf{1}_{\{X>v\}}\Big] \\ &= \mathrm{E}\Big[X\mathbf{1}_{\{X>v\}}\Big] - v\,\mathrm{P}(X>v) + (1-p)v \\ &= \mathrm{E}\Big[X\mathbf{1}_{\{X>v\}}\Big] + v\big(\mathrm{P}(X \le v) - p\big),\end{aligned}$$

which is the desired identity. $\square$

# Important proofs for subadditivity

# Theorem: ES is subadditive

For $p \in (0,1)$, $\mathrm{ES}_p$ is subadditive on $L^0$:

$$\mathrm{ES}_p(X+Y) \;\leq\; \mathrm{ES}_p(X) \;+\; \mathrm{ES}_p(Y) \qquad \text{for all } X, Y \in L^0.$$

**Remark.** In practice $\mathrm{ES}_p$ is used on $L^1$.

**Our plan.** We prove subadditivity of $\mathrm{ES}_p$ for *bounded* losses $X, Y \in L^\infty$ as this implies the theorem for $X, Y \in L^0$.

# Lemma 4 (Law of Large Numbers for order statistics)

Let $(X_i)_{i\geq 1}$ be i.i.d. losses with $X_1 \in L^\infty$. For $n \in \mathbb{N}$, let $X_{[i,n]}$ denote the $i$-th *largest* value among $X_1, \ldots, X_n$. Then for any $p \in (0,1)$,

$$\lim_{n\to\infty} \frac{1}{n(1-p)} \sum_{i=1}^{\lfloor n(1-p)\rfloor} X_{[i,n]} \;=\; \mathrm{ES}_p(X_1) \quad \text{almost surely.}$$

Intuition: the average of the worst $\lfloor n(1-p)\rfloor$ sample losses converges to the population tail mean at level $p$.

# Proof based on LLN for order statistics — setup

Fix integers $m \leq n$ and write

$$A_m^n := \{(i_1, \ldots, i_m) : \; 1 \leq i_1 < \cdots < i_m \leq n\}.$$

For a sequence $X_1, \ldots, X_n$ and the descending order statistics $X_{[1,n]} \geq \cdots \geq X_{[n,n]}$,

$$\sum_{i=1}^{m} X_{[i,n]} \; = \; \max\left\{ X_{i_1} + \cdots + X_{i_m} : \; (i_1, \ldots, i_m) \in A_m^n \right\}.$$

# Proof based on LLN for order statistics — key inequality

Let $X, Y \in L^\infty$. Take i.i.d. copies $(X_1, Y_1), (X_2, Y_2), \ldots$ distributed as $(X, Y)$, and set $Z_i := X_i + Y_i$. Then, for any $m \le n$,

$$\begin{aligned}
\sum_{i=1}^{m} Z_{[i,n]} &= \max\left\{ Z_{i_1} + \cdots + Z_{i_m} : \ (i_1, \ldots, i_m) \in A^n_m \right\} \\
&= \max\left\{ (X_{i_1} + \cdots + X_{i_m}) + (Y_{i_1} + \cdots + Y_{i_m}) \right\} \\
&\le \max\left\{ X_{i_1} + \cdots + X_{i_m} \right\} + \max\left\{ Y_{i_1} + \cdots + Y_{i_m} \right\} \\
&= \sum_{i=1}^{m} X_{[i,n]} + \sum_{i=1}^{m} Y_{[i,n]}.
\end{aligned}$$

# Proof based on LLN for order statistics — conclusion

Choose $m = \lfloor n(1-p) \rfloor$ and divide the inequality by $n(1-p)$:

$$\frac{1}{n(1-p)} \sum_{i=1}^{\lfloor n(1-p) \rfloor} Z_{[i,n]} \leq \frac{1}{n(1-p)} \sum_{i=1}^{\lfloor n(1-p) \rfloor} X_{[i,n]} + \frac{1}{n(1-p)} \sum_{i=1}^{\lfloor n(1-p) \rfloor} Y_{[i,n]}$$

By Lemma 4 (LLN for order statistics), as $n \to \infty$,

$$\frac{1}{n(1-p)} \sum_{i=1}^{\lfloor n(1-p) \rfloor} X_{[i,n]} \to \mathrm{ES}_p(X), \quad \frac{1}{n(1-p)} \sum_{i=1}^{\lfloor n(1-p) \rfloor} Y_{[i,n]} \to \mathrm{ES}_p(Y),$$

$$\frac{1}{n(1-p)} \sum_{i=1}^{\lfloor n(1-p) \rfloor} Z_{[i,n]} \to \mathrm{ES}_p(X+Y) \quad \text{a.s.}$$

Hence $\mathrm{ES}_p(X+Y) \leq \mathrm{ES}_p(X) + \mathrm{ES}_p(Y)$ for $X, Y \in L^\infty$. $\square$

# Optimization view: define the objective

Fix $p \in (0,1)$ and $X \in L^\infty$. For $t \in \mathbb{R}$ set

$$f_X(t) \;:=\; t \;+\; \frac{1}{1-p}\,\mathrm{E}[(X-t)_+]\,.$$

We will use that **VaR is a minimizer of** $f_X$ and **ES is the minimum value of** $f_X$.

# Lemma 5 (VaR is an optimizer)

For $p \in (0,1)$ and $X \in L^\infty$,

$$\mathrm{VaR}_p(X) \;\in\; \arg\min_{t \in \mathbb{R}} f_X(t).$$

*Proof (sketch).* Write $t_0 = \mathrm{VaR}_p(X)$ and note

$$f_X(t) = t + \frac{1}{1-p}\int_t^\infty (1 - F_X(x))\,dx$$

(by integration by parts). For $t_1 > t_0$, $F_X(x) \geq p$ on $(t_0, t_1)$, so

$$f_X(t_1) - f_X(t_0) = (t_1 - t_0) - \frac{1}{1-p}\int_{t_0}^{t_1}(1 - F_X(x))dx \;\geq\; 0.$$

For $t_2 < t_0$, $F_X(x) \leq p$ on $(t_2, t_0)$, giving $f_X(t_2) - f_X(t_0) \geq 0$. Hence $t_0$ minimizes $f_X$. $\square$

# Lemma 6 (ES is the minimum value)

For $p \in (0,1)$ and $X \in L^\infty$,

$$\mathrm{ES}_p(X) \;=\; \min_{t\in\mathbb{R}} f_X(t).$$

*Proof.* Lemma 2 gives

$$\mathrm{ES}_p(X) = \mathrm{VaR}_p(X) + \frac{1}{1-p}\,\mathrm{E}[(X - \mathrm{VaR}_p(X))_+] = f_X\big(\mathrm{VaR}_p(X)\big).$$

By Lemma 5, $\mathrm{VaR}_p(X)$ is a minimizer of $f_X$, so
$f_X(\mathrm{VaR}_p(X)) = \min_t f_X(t)$. $\square$

# Proof based on the optimization property (subadditivity)

Let $f_Z(t) := t + \frac{1}{1-p}\,\mathrm{E}[(Z-t)_+]$. For $X, Y \in L^\infty$, choose $t_1 = \mathrm{VaR}_p(X)$, $t_2 = \mathrm{VaR}_p(Y)$. Then

$$\begin{aligned}
\mathrm{ES}_p(X) + \mathrm{ES}_p(Y) &= f_X(t_1) + f_Y(t_2) \\
&\geq t_1 + t_2 + \frac{1}{1-p}\,\mathrm{E}[(X-t_1)_+ + (Y-t_2)_+] \\
&\geq t_1 + t_2 + \frac{1}{1-p}\,\mathrm{E}[(X+Y-(t_1+t_2))_+]\,.
\end{aligned}$$

Therefore

$$\mathrm{ES}_p(X) + \mathrm{ES}_p(Y) \;\geq\; \min_{t\in\mathbb{R}} f_{X+Y}(t) \;=\; \mathrm{ES}_p(X+Y) \quad \text{(by Lemma 6).}$$

$\square$

# Estimator of $\mathrm{ES}_p$ from order statistics

Given losses $X_1, \ldots, X_n$ and level $p \in (0,1)$, let $X_{[1:n]} \geq \cdots \geq X_{[n:n]}$ be the descending order statistics. Set

$$m = \lfloor n(1-p) \rfloor, \qquad \alpha = n(1-p) - m \in [0,1).$$

Define the empirical Expected Shortfall by

$$\widehat{\mathrm{ES}}_p^{(n)} = \frac{1}{n(1-p)} \left( \sum_{i=1}^{m} X_{[i:n]} \; + \; \alpha \, X_{[m+1:n]} \right).$$

If $n(1-p) \in \mathbb{N}$ (i.e., $\alpha = 0$), this reduces to the average of the largest $m$ observations: $\widehat{\mathrm{ES}}_p^{(n)} = \frac{1}{m} \sum_{i=1}^{m} X_{[i:n]}$.

By the LLN for order statistics ( Lemma 4), $\widehat{\mathrm{ES}}_p^{(n)} \to \mathrm{ES}_p(X_1)$ a.s. under mild conditions (e.g., $X_1 \in L^1$ and continuity at $\mathrm{VaR}_p$).