Notes and Proof Sketches
Risk Optimization 2025/2026

These notes are intended as a personal study aid and include some preliminary proof sketches, example ideas, and potential clarifications to the Risk Optimization 2025/2026 lecture notes.

Indhold

1 Proof sketches

1.1 proof of proposition 4.14

Let us recall Proposition 4.12. Since we assume $\alpha^- < \alpha < \alpha^+ < 1$, we obtain the following ordering for the tail distributions:

$$F_L^{\alpha-}(t) > F_L^{\alpha}(t) > F_L^{\alpha+}(t)$$

for all $t \geq VaR_\alpha(L)$ where $F_L(t) < 1$. This implies that for the survival functions, the inequality is reversed:

$$1 - F_L^{\alpha-}(t) < 1 - F_L^{\alpha}(t) < 1 - F_L^{\alpha+}(t)$$

For F_L^β, $\beta \in \{\alpha^-, \alpha, \alpha^+\}$, by definition, $L \geq \mathrm{VaR}_\alpha(L)$ almost surely. Using Proposition 4.1 with $z = \mathrm{VaR}_\alpha(L)$:

$$\mathbb{E}_{F^\beta}[L] = \mathrm{VaR}_\alpha(L) + \mathbb{E}_{F^\beta}[(L - \mathrm{VaR}_\alpha(L))_+] = \mathrm{VaR}_\alpha(L) + \int_{\mathrm{VaR}_\alpha(L)}^{\infty} \big(1 - F_L^\beta(t)\big)\, dt.$$

Observe that since $\alpha^+ < 1$, the distributions are not identical everywhere on the interval $[VaR_\alpha(L), \infty)$. Therefore, when we integrate the strict inequality of the survival functions, the strict inequalities are preserved:

$$\int_{VaR_\alpha(L)}^{\infty} (1 - F_L^{\alpha-}(t))\, dt < \int_{VaR_\alpha(L)}^{\infty} (1 - F_L^{\alpha}(t))\, dt < \int_{VaR_\alpha(L)}^{\infty} (1 - F_L^{\alpha+}(t))\, dt$$

Adding $\mathrm{VaR}_\alpha(L)$ to these terms yields $\mathbb{E}_{F^{\alpha-}}[L] < \mathbb{E}_{F^{\alpha}}[L] < \mathbb{E}_{F^{\alpha+}}[L]$. By Definitions 4.10 and 4.11, this implies:

$$\mathrm{CVaR}_\alpha^-(L) < \mathrm{CVaR}_\alpha(L) < \mathrm{CVaR}_\alpha^+(L).$$

1.2 proof of proposition 4.21

First, we establish the value of VaR. According to the assumption in Proposition 4.21, the probabilities accumulate such that $\sum_{s=1}^{s_\alpha} \pi_s \geq \alpha > \sum_{s=1}^{s_\alpha - 1} \pi_s$. Applying Definition 1.2 it follows that $VaR_\alpha(L) = L_{s_\alpha}$.

Next, we express CVaR using Corollary 4.20:

$$CVaR_\alpha(L) = VaR_\alpha(L) + \frac{1}{1-\alpha}\mathbb{E}[(L - VaR_\alpha(L))_+]$$

Substituting $VaR_\alpha(L) = L_{s_\alpha}$ and expanding the expectation for the discrete support yields

$$CVaR_\alpha(L) = L_{s_\alpha} + \frac{1}{1-\alpha}\sum_{s=1}^{S} \pi_s (L_s - L_{s_\alpha})_+$$

By the sorted assumption $L_1 < \cdots < L_S$, we have $(L_s - L_{s_\alpha})_+ = 0$ for all $s \leq s_\alpha$. Consequently, the summation reduces to the tail indices:

$$CVaR_\alpha(L) = L_{s_\alpha} + \frac{1}{1-\alpha} \sum_{s=s_\alpha+1}^{S} \pi_s (L_s - L_{s_\alpha})$$

We factor out L_{s_α} and split the sum:

$$CVaR_\alpha(L) = \frac{1}{1-\alpha} \left[(1-\alpha) L_{s_\alpha} + \sum_{s=s_\alpha+1}^{S} \pi_s L_s - L_{s_\alpha} \sum_{s=s_\alpha+1}^{S} \pi_s \right]$$

Using the property of probability measures that $\sum_{s=s_\alpha+1}^{S} \pi_s = 1 - \sum_{s=1}^{s_\alpha} \pi_s$, we substitute into the expression:

$$CVaR_\alpha(L) = \frac{1}{1-\alpha} \left[\left((1-\alpha) - \left(1 - \sum_{s=1}^{s_\alpha} \pi_s \right) \right) L_{s_\alpha} + \sum_{s=s_\alpha+1}^{S} \pi_s L_s \right]$$

Simplifying the terms:

$$CVaR_\alpha(L) = \frac{1}{1-\alpha} \left[\left(\sum_{s=1}^{s_\alpha} \pi_s - \alpha \right) L_{s_\alpha} + \sum_{s=s_\alpha+1}^{S} \pi_s L_s \right]$$

2 ideas of examples

2.1 Example 1.4 - The Newsvendor Problem

(reference?)

A newspaper vendor must decide in advance how many copies $x \in \mathbb{R}^+$ to order for a particular day, before the actual demand is known. Let the daily demand for the newspaper be a non-negative random variable ξ.

The vendor faces a trade-off. Ordering too many copies leads to a surplus cost of c_o per unsold newspaper, while ordering too few results in a shortage cost c_u for each potential sale lost. The goal is to choose the optimal order quantity x that balances these two costs given the uncertain demand.

For a fixed decision x, the realized loss depends on the outcome of ξ. Specifically, the loss function is given by

$$f(x, \xi) = c_o(x - \xi)_+ + c_u(\xi - x)_+$$

Maybe add the definition of $(x)_+ = \max(0, x)$*, since it hasn't been introduced yet??*

Since ξ is random, the loss $f(x, \xi)$ is itself a random variable for any chosen x. Consequently, the problem of finding an optimal x is not well-defined until we specify exactly what aspect of this random loss we wish to minimize. Should the retailer focus on the loss expected on an average day? Or should they focus on avoiding a catastrophic loss in a worst-case scenario? The answer depends entirely on how one chooses to model the uncertainty and the decision maker's specific attitude toward risk.

2.2 Example 1.5 - Example 1.4 continued

(this should go on page 8-9?)

The Newsvendor problem fits the framework of two-stage stochastic programming. We treat the order quantity of newspapers x as the first-stage decision. We introduce second-stage variables $w^+(\omega)$ to represent shortage and $w^-(\omega)$ to represent surplus of newspapers. The balance constraint matches the form of (1.9):

$$x + w^+(\omega) - w^-(\omega) = \boldsymbol{\xi}(\omega) \quad \text{a.s.}$$

Identifying the costs as $q^+ = c_u$ and $q^- = c_o$, and noting that the first-stage cost is implicitly zero in this formulation as the cost is fully captured by the deviation from demand, the problem becomes:

$$\begin{aligned} \min_{x,w^+,w^-} \quad & \mathbb{E}\left[c_u w^+(\omega) + c_o w^-(\omega)\right] \\ \text{s.t.} \quad & x + w^+(\omega) - w^-(\omega) = \boldsymbol{\xi}(\omega) \quad \text{a.s.} \\ & x \geq 0, \quad w^+(\omega) \geq 0, \quad w^-(\omega) \geq 0 \end{aligned}$$

following proof of thm 2.3 - Tying stochastic dominance to VAR

It is interesting to observe that the quantile-function technique used in the proof of Proposition 2.3 provides a direct link between stochastic dominance and VaR.

Recall that for a confidence level α from chapter 1, the Value-at-Risk is defined as the α-quantile of the distribution:

$$\mathrm{VaR}_\alpha(L) = \min\{z \mid F_L(z) \geq \alpha\} = F_L^{-1}(\alpha)$$

If X and Y represent losses and X dominates Y, FSD implies (as seen in proof of thm 2.3):

$$F_X^{-1}(\alpha) \geq F_Y^{-1}(\alpha) \quad \forall \alpha \in (0,1)$$

Substituting the definition of VaR:

$$\mathrm{VaR}_\alpha(X) \geq \mathrm{VaR}_\alpha(Y)$$

Thus, if X dominates Y by First-Order Stochastic Dominance, then for every possible confidence level α, the Value-at-Risk of X is higher to that of Y.

2.3 example (2.1)+(2.2)+(2.3)

Perhaps you could collect Examples 2.1, 2.2, and 2.3 and present a generalization. This would require stating Jensen's inequality in the appendix, possibly with a reference to Le Gall, Measure Theory, Probability, and Stochastic Processes, which is the primary textbook for the Probability Theory 1 course at KU and where Jensen's inequality is stated and proved on page 66.

proposition 2.xx?

Let $\boldsymbol{X}$ be a random variable representing a lottery with finite expectation. The classification of a decision maker's risk attitude is determined by the application of Jensen's inequality (A.XX) to their utility function u. Specifically, a decision maker is risk-averse if and only if u is concave, which implies that $u(\mathbb{E}[\boldsymbol{X}]) \geq \mathbb{E}[u(\boldsymbol{X})]$. That is, the utility of the sure expected value exceeds the expected utility of the gamble. A decision maker is risk-seeking if and only if u is convex, implying that $u(\mathbb{E}[\boldsymbol{X}]) \leq \mathbb{E}[u(\boldsymbol{X})]$. Finally, risk neutrality corresponds to an linear utility function where the decision maker is indifferent between the two quantities, satisfying $u(\mathbb{E}[\boldsymbol{X}]) = \mathbb{E}[u(\boldsymbol{X})]$.

2.4 Table for chapter 2

It would be helpful to include a table that clearly distinguishes between the conditions for gains and for losses. Since we frequently switch between the two frameworks, a concise "translation table" would reduce the risk of sign errors. Specifically, Section 2.5 could be expanded or integrated with such a side-by-side comparison, showing the corresponding definitions for gains versus losses (e.g., FSD inequalities, SSD integral inequalities). This would provide a quick reference for exercises and improve clarity throughout the material.

Concept	**Gains / Returns (X)**	**Losses / Costs (L)**
Objective	Maximize	Minimize
Better Outcome	Higher values	Lower values
First-Order (FSD)	$X \succeq_{(1)} Y$	$L_A \succeq_{(1)} L_B$
Condition:	$F_X(t) \leq F_Y(t), \quad \forall t$	$F_{L_A}(t) \geq F_{L_B}(t), \quad \forall t$
Visual:	CDF of X is below CDF of Y	CDF of L_A is above CDF of L_B
Second-Order (SSD)	$X \succeq_{(2)} Y$	$L_A \succeq_{(2)} L_B$
Integral Condition:	$\int_{-\infty}^{\eta} F_X(t)dt \leq \int_{-\infty}^{\eta} F_Y(t)dt$	$\int_{-\infty}^{\eta} F_{L_A}(t)dt \geq \int_{-\infty}^{\eta} F_{L_B}(t)dt$
Expectation Condition:	$\mathbb{E}[(\eta - X)_+] \leq \mathbb{E}[(\eta - Y)_+]$	$\mathbb{E}[(L_A - \eta)_+] \leq \mathbb{E}[(L_B - \eta)_+]$

Note for gains, we use $(\eta - X)_+$ which measures deviation below a target. For Losses, we use $(L - \eta)_+$ which measures deviation above a target.

2.5 section 3.2 - Choosing M

I thought more carefully about choosing M, as I was skeptical of simply taking the supremum of $f_i(x, \xi)$ for practical cases.

I think it would be really valuable to add a remark in Section 2.3 discussing the practical difficulty of selecting the constant M_{ik}. You could explain that just picking an arbitrarily large number creates a weak LP relaxation. Basically, if M is too big, the solver "cheats"by setting the binary variable to a tiny fraction, 10^{-6} to satisfy the math without "paying"the cost, which destroys the lower bounds and kills the solver's speed. But you could also point out the theoretical solution is also not possible always. Calculating the perfectly tight constant M is often difficult, as it may require maximizing a convex loss function. This is typical, since the technique above is usually applied in problems where the underlying geometry is complicated. Consequently, this maximization problem is likely to be challenging as well. It would be a great practical insight i guess.

my inspiration and where i got the above from:

https://orinanobworld.blogspot.com/2018/09/choosing-big-m-values.html

2.6 section 3.2 - Choosing K

As mentioned in the lectures, a short discussion on the choice of the sample size K could be nice. While I consulted Lectures on Stochastic Programming: Modeling and Theory by Alexander Shapiro, my current level of experience does not allow me to fully assess how much background theory is required to make these notes completely self-contained. Nevertheless, the material under could provide a sufficient theoretical foundation for understanding the main ideas in K?

It is also important to point out, as emphasized in the lectures, that increasing K leads to improved precision only at a very slow rate. However, i could not find a good way to incorporate this.

3.2 SAA method and choice of K

...(Right before section 3.2 maybe)

Returning to the minimization problem of the expected value,

$$\min_{x \in \mathcal{X}} \{f(x) := \mathbb{E}[F(x, \boldsymbol{\xi})]\} \tag{1}$$

we apply the SAA method by generating a sample $\xi^1, \ldots, \xi^K$ of K independent and identically distributed (i.i.d.) realizations of the random vector $\boldsymbol{\xi}$. We then replace the expected value in the objective function (or the probability in a constraint) with the empirical mean computed from this sample. Thus, the SAA is given by

$$\min_{x \in \mathcal{X}} \left\{ \hat{f}_K(x) := \frac{1}{K} \sum_{j=1}^{K} F(x, \xi^j) \right\}. \tag{2}$$

The crucial question for a decision maker is to determine the sample size K required to ensure that the solution to the approximate problem is a reliable estimator of the solution to the true problem.

Let us first consider the estimation of the objective function value for a fixed decision $x \in \mathcal{X}$. Since the sample is i.i.d., the Law of Large Numbers guarantees that the sample average $\hat{f}_K(x)$ converges pointwise to the true mean $f(x)$ with probability one as $K \to \infty$. If the random variable $F(x, \boldsymbol{\xi})$ has a finite variance, then the error of the approximation converges in distribution to a normal variable and by CLT:

$$\sqrt{K}\left(\hat{f}_K(x) - f(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(x)). \tag{3}$$

This asymptotic normality allows us to construct confidence intervals for the true value $f(x)$. For a large sample size K, the error $|\hat{f}_K(x) - f(x)|$ is bounded by $z_{\alpha/2}\sigma(x)/\sqrt{K}$ with probability $1-\alpha$, where $z_{\alpha/2}$ is the critical value of the standard normal distribution. To see this:

$$\begin{aligned} P\left(-z_{\alpha/2} \leq \frac{\hat{f}_K(x) - f(x)}{\sigma(x)/\sqrt{K}} \leq z_{\alpha/2}\right) &= 1-\alpha \\ \iff P\left(\left|\frac{\hat{f}_K(x) - f(x)}{\sigma(x)/\sqrt{K}}\right| \leq z_{\alpha/2}\right) &= 1-\alpha \\ \iff P\left(|\hat{f}_K(x) - f(x)| \leq z_{\alpha/2}\frac{\sigma(x)}{\sqrt{K}}\right) &= 1-\alpha \end{aligned}$$

If a decision maker requires the estimation error to be smaller than a tolerance ϵ with confidence $1-\alpha$, we can derive the requisite sample size algebraically by setting the half-width of the confidence interval equal to ϵ. This yields a sample size requirement of

$$K \geq \frac{z_{\alpha/2}^2\sigma^2(x)}{\epsilon^2}. \tag{4}$$

Note in the above, that since σ^2 is usually unknown, we replace it with the sample variance $\hat{\sigma}^2$ calculated from the data. However, the result only applies to the estimation of a single, fixed point x. In optimization, we search through the entire feasible set $\mathcal{X}$ to find an optimal solution. Ensuring that the SAA function $\hat{f}_K(x)$ is close to $f(x)$ for all x simultaneously requires a stronger condition (uniform convergence?).

continue (theorem 5.17, improve the precision of the solution by one order of magnitude, the sample size must increase by a factor of one hundred)

Ltterature: Shapiro, Alexander, Darinka Dentcheva, and Andrzej Ruszczyński. Lectures on Stochastic Programming: Modeling and Theory. Primarily chapter 5.

3 small corrections

3.1 page 33 error

for (2.8) it says under:

"If w is convex in x, then the SSD conditions are convex for a given η"

Isn't this false? Since w appears as $-w$ in the expression, shouldn't w be concave for $-w$ to be convex, and then the whole expression is convex as desired?

also, (2.8), $f(x)$ should be concave to have a tractable optimization problem, due to the "max- we minimize convex functions and maximize concave.

That is, the feasiable set is convex as required, with w being concave and therefore the expression convex, but the $f(x)$ is concave due to the max.

3.2 page 38 error

A small error on page 38. It should say:

$$\int_{-\infty}^{\eta} F_{\boldsymbol{W}}(\alpha)\, d\alpha \leq \int_{-\infty}^{\eta} F_{\boldsymbol{Y}}(\alpha)\, d\alpha \quad \forall \eta \in \mathbb{R}$$

holds if and only if

$$\mathbb{E}\big[(\eta - \boldsymbol{W})_+\big] \leq \mathbb{E}\big[(\eta - \boldsymbol{Y})_+\big] \quad \forall \eta \in \mathbb{R}.$$

Notice the exchange of η with $\boldsymbol{W}$ and $\boldsymbol{Y}$ in the expectation. This also aligns with the proof of Corollary 2.10.

3.3 Proposition 4.22

I do not see why $G_\alpha(L, z)$ should be bounded (i.e $|G_\alpha(L, z)| \leq M \;\; \forall z$), since as $z \to \infty$, the expression appears to diverge to ∞. Instead, one can rely on the finiteness of $G_\alpha(L, z)$ together with its convexity on the open set $\mathbb{R}$ to conclude that it is continuous. It seems that this is indeed the argument the proof is using: one employs the fact that

$$G_\alpha(L, z) < \infty,$$

which expresses finiteness (a norm bound), rather than bounded.

This does not change the proof in 4.24, however i would prefer the wording finiteness rather than bounded.