Cows, risk, and SDDP.jl

Oscar Dowson

Department of Industrial Engineering and Management Sciences
Northwestern

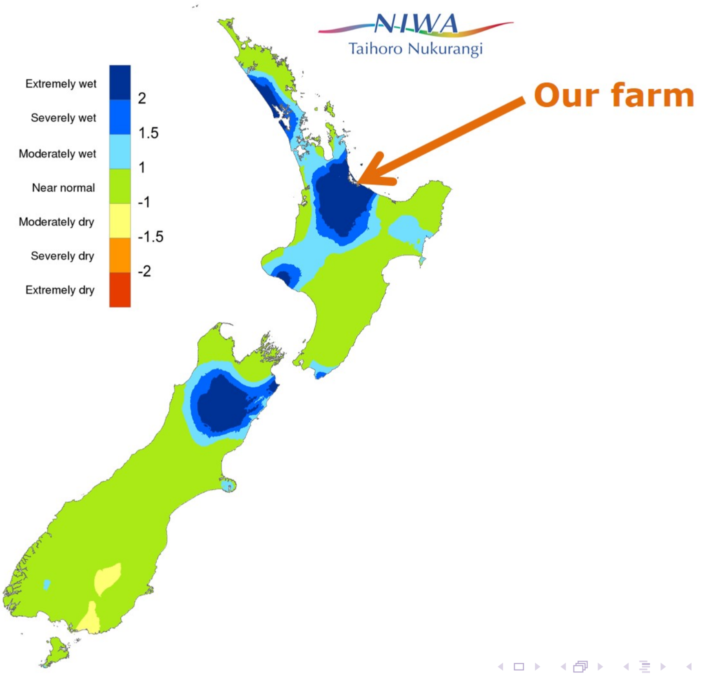
October 29, 2018

The dairy farmer problem

Northwestern ENGINEERING

maximise:	revenue from milk production less operating costs
by deciding:	the number of cows to farm
	the quantity of grass to feed
	the quantity of supplement to feed
	when to dry-off the herd
subject to:	obtaining a high Body Condition Score at the end of the season
	uncertainty in grass growth
	uncertainty in the milk price

SPI Drought Index for 9am 27/08/2017 to 9am 26/09/2017




In my opinion,
all palm oil
should be banned.

What role can stochastic programming play?

Northwestern ENGINEERING

- The farmer made a sequence of "bad" decisions (in hindsight)
  - They had too many cows to begin with
  - They didn't buy more feed when it was cheap
  - They didn't sell their cows while the price was high
- But they were also unlucky. It was the wettest spring in recent memory.
- Given the information available at the time, did they make the right decisions?

Is the last 30 years of experience a good heuristic for the next 30 years?

To learn more about this, come to my talk

Wednesday, October 31, 2018 @ 2:00 p.m. Room 274 Animal Sciences Bldg.

Why care about risk?

If the tail matters more than the average.

- For a farmer, bad years mean cows starve or you go bankrupt and lose your farm
- As for the question of whether the farmer made the right decision last year, it depends on how the value risk.

Static Risk Measures

Northwestern ENGINEERING

Definition

A *risk measure* $\mathbb{F}$ is a function that maps a random variable to a real number.

Definition

A *risk measure* $\mathbb{F}$ is a function that maps a random variable to a real number.

Math

We restrict our attention to random variables with a finite sample space $\Omega := \{z_1, z_2, \ldots, z_K\}$ equipped with a sigma algebra of all subsets of Ω and respective (strictly positive) probabilities $\{p_1, p_2, \ldots, p_K\}$.

We denote the random variable with the uppercase Z.

Definition

The *Average Value-at-Risk* at the β quantile ($\mathsf{AV@R}_{1-\beta}$) is:

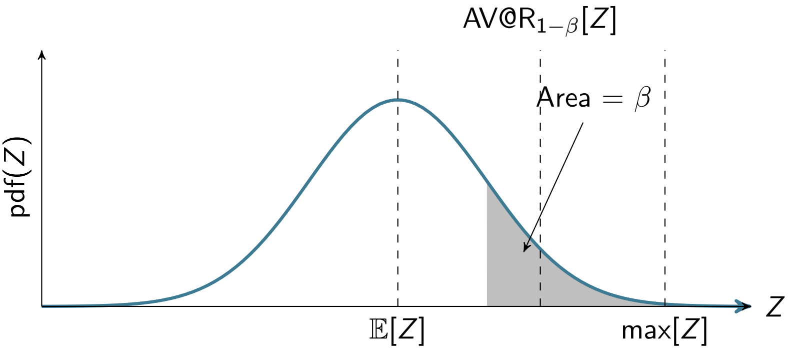
$$\mathsf{AV@R}_{1-\beta}[Z] = \inf_{\zeta} \left\{ \zeta + \frac{1}{\beta} \sum_{k=1}^{K} p_k (z_k - \zeta)_+ \right\},$$

where $(x)_+ = \max\{0, x\}$. (Rockafellar and Uryasev 2002)

Note that when $\beta = 1$, $\mathsf{AV@R}_{1-\beta}[Z] = \mathbb{E}[Z]$, and $\lim_{\beta \to 0} \mathsf{AV@R}_{1-\beta}[Z] = \max[Z]$.

Static Risk Measures

Northwestern ENGINEERING



Definition

A *coherent* risk measure is a risk measure $\mathbb{F}$ that satisfies the axioms of Artzner et al. 1999. For two discrete random variables Z_1 and Z_2, each with drawn from a sample space with K elements, the axioms are:

- **Monotonicity**: If $Z_1 \leq Z_2$, then $\mathbb{F}[Z_1] \leq \mathbb{F}[Z_2]$.
- **Sub-additivity**: For Z_1, Z_2, then $\mathbb{F}[Z_1 + Z_2] \leq \mathbb{F}[Z_1] + \mathbb{F}[Z_2]$.
- **Positive homogeneity**: If $\lambda \geq 0$ then $\mathbb{F}[\lambda Z] = \lambda \mathbb{F}[Z]$.
- **Translation equivariance**: If $a \in \mathbb{R}$ then $\mathbb{F}[Z + a] = \mathbb{F}[Z] + a$.

Northwestern ENGINEERING

We can also define coherent risk measures in terms of *risk sets*. That is, a coherent risk measure $\mathbb{F}$ has a dual representation that can be viewed as taking the expectation of the random variable with respect to the worst probability distribution within some set $\mathfrak{A}$ of possible distributions:

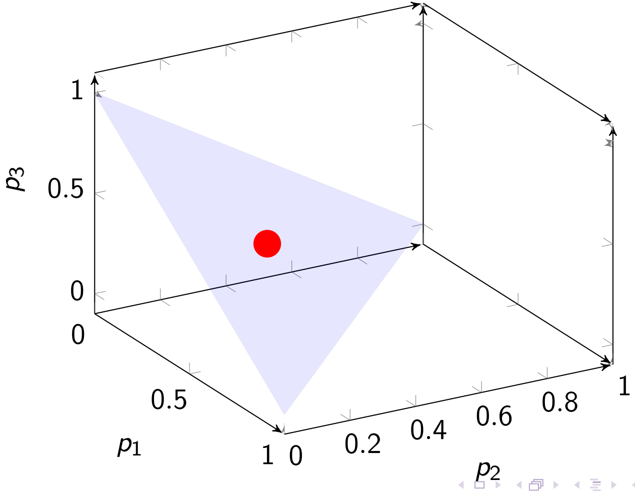
$$\mathbb{F}[Z] = \sup_{\xi \in \mathfrak{A}} \mathbb{E}_{\xi}[Z] = \sup_{\xi \in \mathfrak{A}} \sum_{k=1}^{K} \xi_k z_k, \tag{1}$$

where $\mathfrak{A}$ is a convex subset of:

$$\mathfrak{P} = \left\{ \xi \in \mathbb{R}^K \ : \ \sum_{k=1}^{K} \xi_k = 1, \ \xi \geq 0 \right\}.$$

Northwestern ENGINEERING

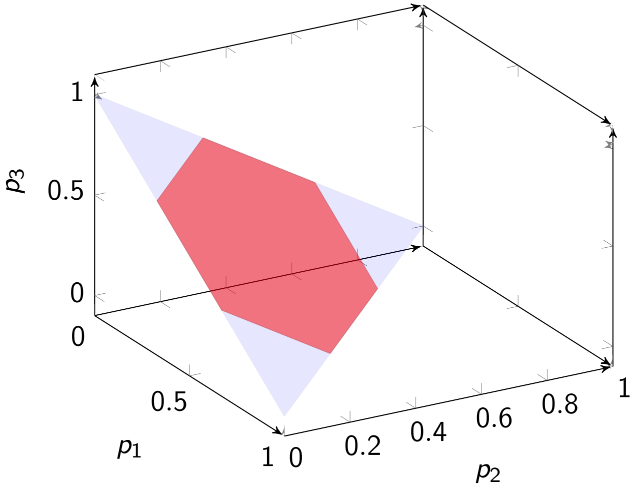
If $\mathfrak{A}$ is a singleton, containing only the original probability distribution, then the risk measure $\mathbb{F}$ is equivalent to the expectation operator.



Static Risk Measures

Northwestern ENGINEERING

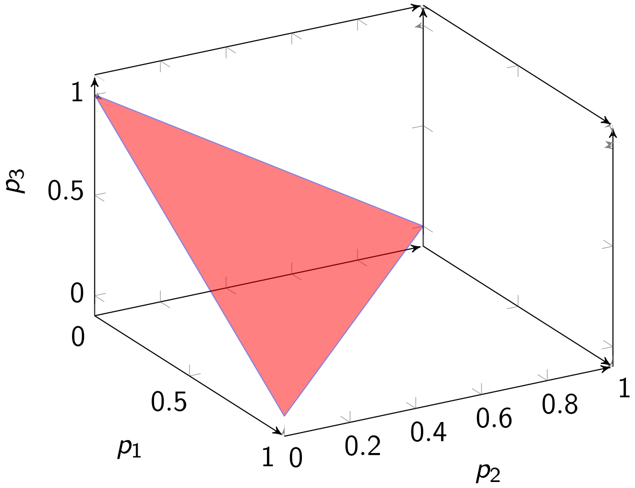
If $\mathfrak{A} = \left\{ \xi \in \mathfrak{P} \mid \xi_k \leq \frac{p_k}{\beta},\ k = 1, 2, \ldots, K \right\}$, then the risk measure $\mathbb{F}$ is equivalent to $\mathsf{AV@R}_{1-\beta}$.



Static Risk Measures

Northwestern ENGINEERING

If $\mathfrak{A} = \mathfrak{P}$, then $\mathbb{F}$ is the Worst-case risk measure.



Okay, so those are static risk measures. How do these translate to the multistage case?

Let's assume we have three stages, $t = 1, 2, 3$.

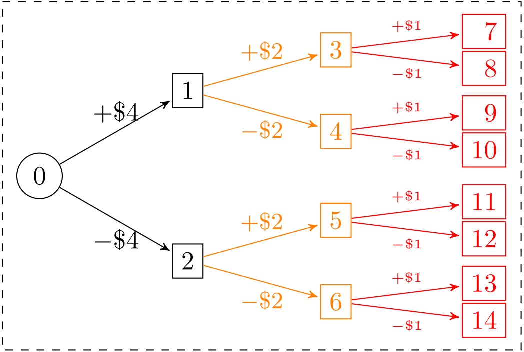
In stage t, the cost incurred is a random variable Z_t that depends on the realization of the noise terms ω_1, ω_2, ..., ω_t.

How do we take the risk of $\mathbb{F}[Z_1, Z_2, Z_3]$?

End-of-horizon risk measure

See, e.g., Pflug and Pichler 2016; Baucke, Downward, and Zakeri 2018

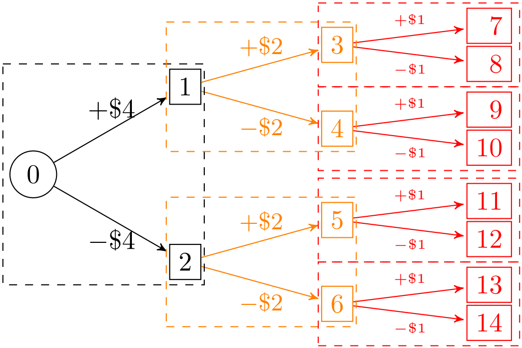
$$\mathbb{F}[Z_1, Z_2, Z_3] = \mathbb{F}_{\omega_1,\omega_2,\omega_3}[Z_1 + Z_2 + Z_3]$$



Nested risk measure

See, e.g., Ruszczyński 2010; Philpott, de Matos, and Finardi 2013

$$\mathbb{F}[Z_1, Z_2, Z_3] = \mathbb{F}_{\omega_1}[Z_1 + \mathbb{F}_{\omega_2|\omega_1}[Z_2 + \mathbb{F}_{\omega_3|\omega_1,\omega_2}[Z_3]\,]\,]$$



Risk in SDDP

Northwestern ENGINEERING

Recall our favourite dynamic programming recursion:

$$\begin{aligned} V_t(x_t, \omega_t) = \min_{u_t} \quad & C_t(x_t, u_t, \omega_t) + \mathop{\mathbb{E}}_{\omega_{t+1} \in \Omega_{t+1}} [V_{t+1}(x_{t+1}, \omega_{t+1})] \\ \text{s.t.} \quad & x_{t+1} = T_t(x_t, u_t, \omega_t) \\ & u_t \in U_t(x_t, \omega_t), \end{aligned}$$

where the decision-rule $\pi_t(x_t, \omega_t)$ takes the value of u_t in the optimal solution.

Risk in SDDP

Northwestern ENGINEERING

Recall our favourite dynamic programming recursion:

$$\begin{aligned} V_t(x_t, \omega_t) = \min_{u_t} \quad & C_t(x_t, u_t, \omega_t) + \underset{\omega_{t+1} \in \Omega_{t+1}}{\mathbb{F}} \left[V_{t+1}(x_{t+1}, \omega_{t+1}) \right] \\ \text{s.t.} \quad & x_{t+1} = T_t(x_t, u_t, \omega_t) \\ & u_t \in U_t(x_t, \omega_t), \end{aligned}$$

where the decision-rule $\pi_t(x_t, \omega_t)$ takes the value of u_t in the optimal solution.

Risk in SDDP

Northwestern ENGINEERING

Recall our favourite dynamic programming recursion:

$$
\begin{aligned}
V_t(\bar{x}_t, \omega_t) = \min_{u_t} \quad & C_t(x_t, u_t, \omega_t) + \theta_{t+1} \\
\text{s.t.} \quad & x_t = \bar{x}_t, \quad [\lambda_t] \\
& x_{t+1} = T_t(x_t, u_t, \omega_t) \\
& u_t \in U_t(x_t, \omega_t) \\
& \theta_{t+1} \geq \alpha_{t+1}^k + {\beta_{t+1}^k}^\top (x_{t+1} - \bar{x}_{t+1}^k),
\end{aligned}
$$

where the decision-rule $\pi_t(x_t, \omega_t)$ takes the value of u_t in the optimal solution.

Recall our favourite dynamic programming recursion:

$$
\begin{aligned}
V_t(\bar{x}_t, \omega_t) = \min_{u_t} \quad & C_t(x_t, u_t, \omega_t) + \theta_{t+1} \\
\text{s.t.} \quad & x_t = \bar{x}_t, \quad [\lambda_t] \\
& x_{t+1} = T_t(x_t, u_t, \omega_t) \\
& u_t \in U_t(x_t, \omega_t) \\
& \theta_{t+1} \geq \alpha_{t+1}^k + {\beta_{t+1}^k}^\top (x_{t+1} - \bar{x}_{t+1}^k),
\end{aligned}
$$

where the decision-rule $\pi_t(x_t, \omega_t)$ takes the value of u_t in the optimal solution.

Northwestern ENGINEERING

Recall

Given an original probability distribution $\{p_1, p_2, \dots, p_K\}$ and a coherent risk measure $\mathbb{F}$, there exists a *changed* probability distribution $\{\xi_1, \xi_2, \dots, \xi_K\}$ such that $\mathbb{F}[Z] = \mathbb{E}_\xi[Z]$.

The meat of the matter. Consider the following (re-stated) proposition from Philpott, de Matos, and Finardi 2013:

Proposition

Suppose for each $\omega \in \Omega$, that $\lambda(\bar{x}, \omega)$ is a subgradient of $V(x, \omega)$ at $\bar{x}$. Then, given $\mathbb{F}[V(\bar{x}, \omega)] = \mathbb{E}_{\xi}[V(\bar{x}, \omega)]$, $\mathbb{E}_{\xi}[\lambda(\bar{x}, \omega)]$ is a subgradient of $\mathbb{F}[V(x, \omega)]$ at $\bar{x}$.

The meat of the matter. Consider the following (re-stated) proposition from Philpott, de Matos, and Finardi 2013:

Proposition

Suppose for each $\omega \in \Omega$, that $\lambda(\bar{x}, \omega)$ is a subgradient of $V(x, \omega)$ at $\bar{x}$. Then, given $\mathbb{F}[V(\bar{x}, \omega)] = \mathbb{E}_{\xi}[V(\bar{x}, \omega)]$, $\mathbb{E}_{\xi}[\lambda(\bar{x}, \omega)]$ is a subgradient of $\mathbb{F}[V(x, \omega)]$ at $\bar{x}$.

The meat of the matter. Consider the following (re-stated) proposition from Philpott, de Matos, and Finardi 2013:

Proposition

Suppose for each $\omega \in \Omega$, *that* $\lambda(\bar{x}, \omega)$ *is a subgradient of* $V(x, \omega)$ *at* $\bar{x}$. *Then, given* $\mathbb{F}[V(\bar{x}, \omega)] = \mathbb{E}_{\xi}[V(\bar{x}, \omega)]$, $\mathbb{E}_{\xi}[\lambda(\bar{x}, \omega)]$ *is a subgradient of* $\mathbb{F}[V(x, \omega)]$ *at* $\bar{x}$.

The meat of the matter. Consider the following (re-stated) proposition from Philpott, de Matos, and Finardi 2013:

Proposition

Suppose for each $\omega \in \Omega$, that $\lambda(\bar{x}, \omega)$ is a subgradient of $V(x, \omega)$ at $\bar{x}$. Then, given $\mathbb{F}[V(\bar{x}, \omega)] = \mathbb{E}_{\xi}[V(\bar{x}, \omega)]$, $\mathbb{E}_{\xi}[\lambda(\bar{x}, \omega)]$ is a subgradient of $\mathbb{F}[V(x, \omega)]$ at $\bar{x}$.

Northwestern ENGINEERING

So what is this saying?

To obtain a cut for $\mathbb{F}_{\omega_{t+1} \in \Omega_{t+1}} [V_{t+1}(x_{t+1}, \omega_{t+1})]$

- We can go and solve the $t+1$ stage problems to obtain an objective value $\bar{\theta}_{\omega_{t+1}}$ and a dual vector $\lambda_{\omega_{t+1}}$ for each realization of ω_{t+1}.
- Normally, we take the expectation of these to get the cut

$$\theta_{t+1} \geq \mathbb{E}[\bar{\theta}_{\omega_{t+1}}] + \mathbb{E}[\lambda_{\omega_{t+1}}]^\top (x_{t+1} - \bar{x}_{t+1})$$

- Instead, we compute ξ such that $\mathbb{F}[\bar{\theta}_{\omega_{t+1}}] = \mathbb{E}_\xi[\bar{\theta}_{\omega_{t+1}}]$ and then take the risk-adjusted expectation to get the cut

$$\theta_{t+1} \geq \mathbb{E}_\xi[\bar{\theta}_{\omega_{t+1}}] + \mathbb{E}_\xi[\lambda_{\omega_{t+1}}]^\top (x_{t+1} - \bar{x}_{t+1})$$

Northwestern ENGINEERING

Link

https://github.com/odow/talks/blob/master/2018/uw_luedtke.ipynb

But Wait! Danger ahead!

Northwestern ENGINEERING

Do these nested risk measures make sense?

Remember how the *end-of-horizon* risk measure made the most sense:

$$\mathbb{F}[X_1, X_2, X_3] = \mathbb{F}[X_1 + X_2 + X_3]$$

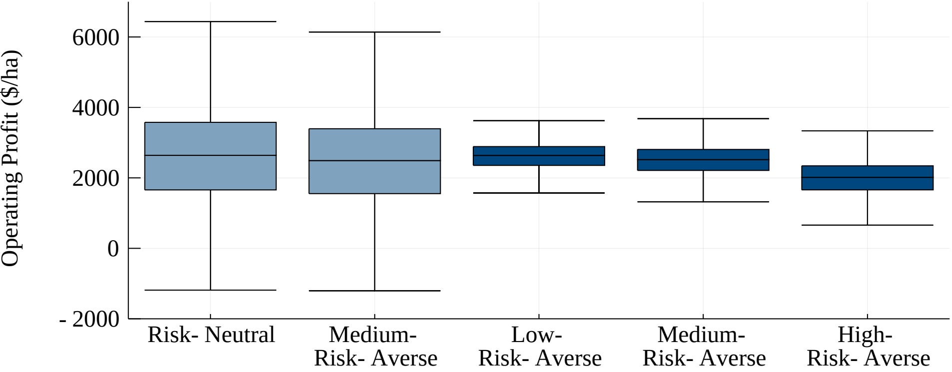
But we actually used the *nested* risk measure:

$$\mathbb{F}[X_1, X_2, X_3] = \mathbb{F}[X_1 + \mathbb{F}[X_2 + \mathbb{F}[X_3 \mid X_1, X_2] \mid X_1]]$$

What is the interpretation of a nested risk measure? This can lead to perverse, counter-intuitive results!

POWDER

Northwestern ENGINEERING



Northwestern ENGINEERING

Philippe Artzner et al. “Coherent Measures of Risk”. In: *Mathematical Finance* 9.3 (1999), pp. 203–228.

Regan Baucke, Anthony Downward, and Golbon Zakeri. “A Deterministic Algorithm for Solving Multistage Stochastic Minimax Dynamic Programmes”. In: *Optimization Online* (2018). URL: http://www.optimization-online.org/DB_FILE/2018/02/6449.pdf.

Georg Ch. Pflug and Alois Pichler. “Time-Inconsistent Multistage Stochastic Programs: Martingale Bounds”. In: *European Journal of Operational Research* 249.1 (2016), pp. 155–163.

Andy Philpott, Vitor de Matos, and Erlon Finardi. “On Solving Multistage Stochastic Programs with Coherent Risk Measures”. In: *Operations Research* 61.4 (2013), pp. 957–970.

References II

Northwestern ENGINEERING

Tyrrell R. Rockafellar and Stanislav P. Uryasev. "Conditional Value-at-Risk for General Loss Distributions". In: *Journal of Banking and Finance* 26 (2002), pp. 1443–1471.

Andrzej Ruszczyński. "Risk-Averse Dynamic Programming for Markov Decision Processes". In: *Mathematical Programming* 125.2 (2010), pp. 235–261.