MINIMAL VARIATION PROBLEM OF PRODUCTION

ALEKSANDER, MARIANNE, FRED

1. A global problem

Given n locations (countries in Europe, say), with *annual* capacity factors for wind and PV given by the random variables $W := (W_1, \ldots, W_n)^\top$ and $PV = (PV_1, \ldots, PV_n)^\top$. Define the $2n$-dimensional vector of wind and PV capacity factors by $R = (W, PV)^\top$ (with R indicating *renewable* capacities. Denote by c the $2n$-dimensional vector of capacities installed in each location, given by $c = (c^w, c^{pv})^\top$, $c^i = (c_1^i, \ldots, c_n^i)^\top$, $i = W, PV$. The production becomes

$$P = c^\top R$$

Notice that all vectors are *column* vectors, and that we use $x^\top$ to denote the transpose of a vector x.

Our first global problem is to try to allocate capacities in the locations meeting two aims: first, the production should have as little variance as possible, and secondly, it should ensure that (random) annual demand D is covered with high probability. This can be formulated as

$$\inf_c c^\top \mathrm{Cov}(R) c \tag{1}$$

under the constraint that

$$\mathbb{P}\left(c^\top R \leq D\right) \leq \epsilon \tag{2}$$

for a given (small) ϵ.

The installation of capacities should be so that demand is met with probability at least $1 - \epsilon$, ad at the same time having as little variance as possible. Additional constraints are

(1) All capacities are non-negative and upper bounded: $0 \leq c \leq c_{\max}$, where $c_{\max}$ is a vector of maximal possible capacity of the given technology (wind/PV) at given location. The upper limit reflects land area, for example, and maximum possible costs that can be used in that country due to decisions.
(2) Total capacity is limited (by costs, say): $c^\top \mathbf{1} = \bar{c}$, where $\bar{c}$ is some positive number.

Notice that by dividing all capacities in the vector c by $\bar{c}$, the second constraint becomes $(c/\bar{c})^\top \mathbf{1} = 1$, and we can interpret $c/\bar{c}$ as the *fraction* (or percentage) of the available total capacity that is installed in a given location with a given technology. One must re-scale constraint 1 as well if one choose this framework.

Date: Outline for internship.

1.1. **The random variables.** We can find probability distribution of R as follows: From capacity factor time series data, aggregate up to an annual basis over the years one has available. Estimate mean and covariance matrix from these data. Same goes for the demand, that should possibly also be covariated to R. I.e., a big covariance matrix should be built for (R, D) based on annual data. It is not unnatural to assume that we have a multivariate Gaussian random variable.

An alternative could be to model the time series of daily or hourly capacity factors by an AR-process (same for demand), and integrate these up to annual figures. Based on the annual variables, one could *derive* the covariances for (R, D).

1.2. **The probability constraint.** Let us look at a reformulation of the probability constraint

$$\mathbb{P}\left(c^\top R \leq D\right) \leq \epsilon$$

Introduce $r := \mathbb{E}[R], d := \mathbb{E}[D]$ and $\sigma_c^2 := \mathrm{Var}(c^\top R - D)$. Define the random variable X_c by

$$X := \frac{c^\top R - D - (c^\top r - d)}{\sigma_c}$$

Notice that if (R, D) is assumed to be a multivariate Gaussian random variable, then X is standard normal, and we denote its distribution function by Φ. Notice that X depends on c in the definition, however, as we have just argued, it is always standard normal and we can ignore this dependency. We have that

$$c^\top R - D = c^\top r - d + \sigma_c X$$

and hence,

$$\begin{aligned}
\mathbb{P}\left(c^\top R \leq D\right) &= \mathbb{P}\left(c^\top r - d + \sigma_c X \leq 0\right) \\
&= \mathbb{P}\left(X \leq \frac{d - c^\top r}{\sigma_c}\right) \\
&= \Phi\left(\frac{d - c^\top r}{\sigma_c}\right) \\
&\leq \epsilon
\end{aligned}$$

Φ is an increasing function, so we find that

$$\frac{d - c^\top r}{\sigma_c} \leq \Phi^{-1}(\epsilon)$$

In conclusion, the probabilistic constraint can be reformulated as

$$c^\top r \geq d - \sigma \Phi^{-1}(\epsilon) \tag{3}$$

This is a standard linear inequality constraint.

It is to be remarked that capacity factors R and demand D are positive variables, and thus a Gaussian assumption is a-priori not valid. Still the above argumentation could be used as an approximation, in the case we appeal to different distributions for capacities and demand. Further, notice also that σ_c^2 can be computed from the covariances of R and D. $\Phi^{-1}(\epsilon)$ is the ϵ-quantile of the standard normal distribution, and readily available.

1.3. **The task.** The task has two parts:

(1) Find mean and covariance for capacity factors and demand on annual basis.
(2) Implement the quadratic program with linear constraints

Given these, one can start solving for different choices of ϵ as well as maximum capacities in each location as well as total allowed installation.

Interesting questions for validation are: how good is the normal assumption for X? For the optimal c, what is the probability of "black-out" (i.e. is the probabilistic constraint binding or not)? How many locations (if any) will have zero allocated capacity (which mean they are superfluous).

2. A non-global problem

If we model the R and D as a stochastic time series, denoted R_t and D_t, we could ask for still a maximally smooth production over a year by minimizing the annualized production variance. But instead of have a yearly probability constraint on the production meeting demand, we could have a timewise constraint. This could take the form

$$\mathbb{P}\left(\min_t c^\top R_t - D_t \geq 0\right) \geq 1 - \epsilon$$

This is the same as

$$\mathbb{P}\left(c^\top R_1 - D_1 \geq 0 \,\&\, \ldots \,\&\, c^\top R_{365} - D_{365} \geq 0\right) \geq 1 - \epsilon$$

if we want to make sure that the *daily* production meets demand with $1-\epsilon$ certainty.

If we have linear (AR) models for R_t and D_t, we will have the following (roughly). $c^{top} R_1 - D_1$ will depend linearly on $X_1 \sim N(0,1)$. Next, $c^\top R_2 - D_2$ will depend linearly on X_1 and well as an independent $X_2 \sim N(0,1)$. Moving forward we can express the 365-variate random variable $c^\top R_1 - D_1, \ldots, c^\top R_{365} - D_{365}$ as $\mu + VX$, where X is a multivariate Gaussian of independent mean zero variables with variance one, and V is a lower triangular matrix, and finally μ is a vector of constants (means). Thus, we get the probability constraint

$$\mathbb{P}(VX \geq -\mu) \geq 1 - \epsilon$$

Maybe this could be massaged further in some efficient way?