Performance of Stochastic Programming Solutions
Operations Research

Anthony Papavasiliou
Performance of Stochastic Programming Solutions

1. The Expected Value of Perfect Information

2. The Value of the Stochastic Solution

3. Basic Inequalities

4. Estimating Performance
Two-Stage Stochastic Linear Programs

\[
\begin{align*}
\min z &= c^T x + \mathbb{E}_\omega [\min q(\omega)^T y(\omega)] \\
\text{s.t. } Ax &= b \\
T(\omega)x + W(\omega)y(\omega) &= h(\omega) \\
x \geq 0, y(\omega) &\geq 0
\end{align*}
\]

- First stage decisions \( x \in \mathbb{R}^{m_1}, c \in \mathbb{R}^{n_1}, b \in \mathbb{R}^{m_1}, A \in \mathbb{R}^{m_1 \times n_1} \)
- For a given realization \( \omega \), second-stage data are \( q(\omega) \in \mathbb{R}^{n_2}, h(\omega) \in \mathbb{R}^{m_2}, T(\omega) \in \mathbb{R}^{m_2 \times n_1}, W(\omega) \in \mathbb{R}^{m_2 \times n_2} \)
- All random variables of the problem are assembled in a single random vector
  \( \xi^T(\omega) = (q(\omega)^T, h(\omega)^T, T_1.(\omega), \ldots, T_{m_2}.(\omega), W_1.(\omega), \ldots, W_{m_2}.(\omega)) \)
Motivation

Is it worth solving a stochastic program?

- How well could we do if we knew the future?
- How well could we do with a simpler model (e.g. expected value problem)?
Table of Contents

1. The Expected Value of Perfect Information
2. The Value of the Stochastic Solution
3. Basic Inequalities
4. Estimating Performance
Notation

\[ z(x, \xi) = c^T x + Q(x, \omega) + \delta(x|K_1) \]

\[ Q(x, \xi) = \min_y \left\{ q(\omega)^T y | W(\omega)y = h(\omega) - T(\omega)x \right\} \]

- **What is the interpretation of** \( z(x, \xi) \)?
- **Define** \( K_1 = \{ x | Ax = b, x \geq 0 \} \) as the set of feasible first-stage decisions
- **Define** \( K_2(\omega) = \{ x | \exists y : W(\omega)y = h(\omega) - T(\omega)x \} \) as the set of first-stage decisions that have a feasible reaction in the second stage for \( \omega \in \Omega \)
- It can be that \( z(x, \xi) = +\infty \) (if \( x \notin K_1 \cap K_2(\omega) \))
- It can be that \( z(x, \xi) = -\infty \) (unbounded below)
Wait-and-See, Here-and-Now

- The **wait-and-see** value is the expected value of reacting with perfect foresight $x^*(\xi)$ to $\xi$:

  $$WS = \mathbb{E}[\min_x z(x, \xi)]$$
  $$\mathbb{E}[z(x^*(\xi), \xi)]$$

- The **here-and-now** value is the expected value of the recourse problem:

  $$SP = \min_x \mathbb{E}[z(x, \xi)]$$

- We have swapped $\min$ and $\mathbb{E}$. What’s the difference?
  - Which one is more difficult to compute?
The **expected value of perfect information** is the difference between the two solutions:

$$EVPI = SP - WS$$

Interpretation: value of a perfect forecast for the future
Example: Capacity Expansion Planning

<table>
<thead>
<tr>
<th>Technology</th>
<th>Fuel cost ($/MWh)</th>
<th>Inv cost ($/MWh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coal</td>
<td>25</td>
<td>16</td>
</tr>
<tr>
<td>Gas</td>
<td>80</td>
<td>5</td>
</tr>
<tr>
<td>Nuclear</td>
<td>6.5</td>
<td>32</td>
</tr>
<tr>
<td>Oil</td>
<td>160</td>
<td>2</td>
</tr>
<tr>
<td>DR</td>
<td>1000</td>
<td>0</td>
</tr>
</tbody>
</table>

Table: Probability of (i) reference load duration curve: 10%, (ii) 10x wind scenario: 90%.

<table>
<thead>
<tr>
<th></th>
<th>Duration (hours)</th>
<th>Level (MW)</th>
<th>Level (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reference scenario</td>
<td>10x wind scenario</td>
<td></td>
</tr>
<tr>
<td>Base load</td>
<td>8760</td>
<td>0-7086</td>
<td>0-3919</td>
</tr>
<tr>
<td>Medium load</td>
<td>7000</td>
<td>7086-9004</td>
<td>3919-7329</td>
</tr>
<tr>
<td>Peak load</td>
<td>1500</td>
<td>9004-11169</td>
<td>7329-10315</td>
</tr>
<tr>
<td>Technology</td>
<td>SP solution</td>
<td>Reference</td>
<td>10x wind</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td>Coal</td>
<td>5085</td>
<td>1918</td>
<td>3410</td>
</tr>
<tr>
<td>Gas</td>
<td>1311</td>
<td>2165</td>
<td>2986</td>
</tr>
<tr>
<td>Nuclear</td>
<td>3919</td>
<td>7086</td>
<td>3919</td>
</tr>
<tr>
<td>Oil</td>
<td>854</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ SP = 340316 \ $/h \]
\[ z(x^*(''Ref''), ''Ref'') = 382288 \ $/h \]
\[ z(x^*(''10x''), ''10x'') = 329383 \ $/h \]
\[ WS = 334673 \ $/h \]
\[ EVPI = 5643 \ $/h = 1.7\% \cdot SP \]

Note: wait-and-see model never chooses oil
Expected (or mean) value problem:

\[
EV = \min_x z(x, \bar{\xi}), \bar{\xi} = \mathbb{E}[\xi]
\]

Expected value solution \(x^*(\bar{\xi})\): optimal solution of expected value problem
The expected value of using the EV solution measures the performance of $x^*(\bar{\xi})$ (optimal second-stage reactions given $x^*(\bar{\xi})$):

$$EEV = \mathbb{E}[z(x^*(\bar{\xi}), \xi)]$$

The value of the stochastic solution is

$$VSS = EEV - SP$$

- Which one is easier to compute: WS, SP, or EEV? Which one is harder?
- What can we say about VSS if $x^*(\xi)$ is independent of $\xi$?
Example: Capacity Expansion Planning

Table: Optimal investment and fixed cost for the stochastic program and the expected value problem.

<table>
<thead>
<tr>
<th></th>
<th>SP investment (MW)</th>
<th>EV investment (MW)</th>
<th>SP fixed cost ($/h)</th>
<th>EV fixed cost ($/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coal</td>
<td>5085</td>
<td>3261</td>
<td>81360</td>
<td>52176</td>
</tr>
<tr>
<td>Gas</td>
<td>1311</td>
<td>2905</td>
<td>6,555</td>
<td>14525</td>
</tr>
<tr>
<td>Nuclear</td>
<td>3919</td>
<td>4235</td>
<td>125408</td>
<td>135520</td>
</tr>
<tr>
<td>Oil</td>
<td>854</td>
<td>0</td>
<td>1708</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>11169</td>
<td>10401</td>
<td>215031</td>
<td>202221</td>
</tr>
</tbody>
</table>
Example: Capacity Expansion Planning

**Table:** Variable cost for the SP and EV models.

<table>
<thead>
<tr>
<th></th>
<th>SP var cost ($)</th>
<th>EV var cost ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>25473</td>
<td>25473</td>
</tr>
<tr>
<td>Block 2</td>
<td>64858</td>
<td>60070</td>
</tr>
<tr>
<td>Block 3</td>
<td>4854</td>
<td>4854</td>
</tr>
<tr>
<td>Block 4</td>
<td>9799</td>
<td>29209</td>
</tr>
<tr>
<td>Block 5</td>
<td>17960</td>
<td>17959</td>
</tr>
<tr>
<td>Block 6</td>
<td>2340</td>
<td>13268</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>125285</strong></td>
<td><strong>150834</strong></td>
</tr>
</tbody>
</table>

- $EEV = 12739$ $$/h$
- Investment cost of $EV$ solution is lower than $SP$ solution
- $EV$ investment cannot serve peak demand in "Ref" scenario
# Table of Contents

1. The Expected Value of Perfect Information
2. The Value of the Stochastic Solution
3. Basic Inequalities
4. Estimating Performance
For every $\xi$, we have $z(x^*(\xi), \xi) \leq z(x^*, \xi)$ where $x^*$ is the optimal solution to the stochastic program.

Taking expectations on both sides, $WS \leq SP$

Interpretation: we can do better if we have a crystal ball (i.e. we know the future in advance)
Lazy Solution

- $x^*$ is the optimal solution of
  
  $$
  \min_x \mathbb{E}[z(x, \xi)]
  $$

- $x^*(\bar{\xi})$ is a solution (not necessarily optimal), therefore
  
  $$
  \min_x \mathbb{E}[z(x, \xi)] = SP \leq EEV = \mathbb{E}[z(x^*(\bar{\xi}), \xi)]
  $$

Interpretation: we do worse when we are lazy (i.e. when we do not account for uncertainty explicitly)

Would anything change if some of the $x, y$ were integer?
Jensen’s Inequality

Suppose $f$ is convex and $\xi$ is a random variable, then $f(\mathbb{E}[\xi]) \leq \mathbb{E}[f(\xi)]$
Suppose $c, W, T$ are independent of $\omega$ (i.e., $\xi = h$): then $EV \leq WS$

- We will show that $z(x, h)$ is jointly convex in $(x, h)$
- We know that $f(\xi) = \min_x z(x, \xi)$ is convex in $\xi$
- From Jensen’s inequality, we have $\mathbb{E}[f(\xi)] \geq f(\mathbb{E}[\xi])$

Interpretation: EV (the lazy solution) is a biased estimate of expected cost. Is it optimistic, or pessimistic?
Proof that $z(x, h)$ is convex in $(x, h)$

- Consider $x_1, x_2$ and $\lambda \in (0, 1)$. Without loss of generality, assume $Ax_1 = b, Ax_2 = b, x_1, x_2 \geq 0$.
- $z(x_i, h_i) = c^T x_i + q^T y_i$, where $y_i = \min\{q^T y | Wy = h_i - T x_i, y \geq 0\}, i = \{1, 2\}$
- $z(\lambda x_1 + (1 - \lambda) x_2, \lambda h_1 + (1 - \lambda) h_2) = c^T (\lambda x_1 + (1 - \lambda) x_2) + q^T y_\lambda$, where $y_\lambda = \min\{q^T y | Wy = \lambda h_1 + (1 - \lambda) h_2 - T(\lambda x_1 + (1 - \lambda) x_2), y \geq 0\}$
- $\lambda y_1 + (1 - \lambda) y_2$ is a feasible solution for \(\min\{q^T y | Wy = \lambda h_1 + (1 - \lambda) h_2 - T(\lambda x_1 + (1 - \lambda) x_2), y \geq 0\}\). Therefore, we have $q^T y_\lambda \leq \lambda q^T y_1 + (1 - \lambda) q^T y_2$.
- It follows that $z(\lambda x_1 + (1 - \lambda) x_2, \lambda h_1 + (1 - \lambda) h_2) \leq \lambda z(x_1, h_1) + (1 - \lambda) z(x_2, h_2)$
Example: Capacity Expansion Planning

Does the cap ex problem satisfy the assumptions of slide 20?

For the capacity expansion problem:

\[ WS = EV = 334674 \, \$/h \]

Exercise: show that \( EV = WS \) holds in general for the two-stage stochastic capacity expansion problem with demand uncertainty.
Consider the following problem:

\[
\begin{align*}
\min_{x \geq 0} & \quad 2x + \mathbb{E}_\xi [\xi \cdot y] \\
\text{s.t.} & \quad y \geq 1 - x \\
& \quad y \geq 0
\end{align*}
\]

where \( P[\xi = 1] = 3/4, \ P[\xi = 3] = 1/4 \)

Does this problem satisfy the assumptions of slide 20?
Optimal second-stage decision: \( y = 1 - x \) if \( 1 - x \geq 0 \), \( y = 0 \) otherwise

Trade-off: by increasing \( x \) we can push \( y \) to lower values, but incur certain cost \( 2x \)

For \( \bar{\xi} = \frac{3}{4} + \frac{3}{4} = \frac{3}{2} \) we have \( \{ \min 2x + \frac{3}{2}y | y \geq 1 - x, x \geq 0, y \geq 0 \} \)

Optimal solution: \( x^*(\bar{\xi}) = 0, \ y = 1 \) with \( EV = \frac{3}{2} \)

To compute \( WS \), note that for \( \xi = 1 \) the optimal first-stage decision is \( x = 0 \), with cost of 1, while for \( \xi = 3 \) the optimal first-stage decision is \( x = 1 \), with cost of 2:

\[
WS = \frac{3}{4} + \frac{1}{4} \cdot 2 = \frac{5}{4} < EV
\]
We have established that

- \( VSS \geq 0, \ EVPI \geq 0 \)
- \( VSS \leq EEV - EV, \ EVPI \leq EEV - EV \)
- If \( EEV - EV = 0 \) then \( VSS = 0, \ EVPI = 0 \) (for example, if \( x^*(\xi) \) independent of \( \xi \) - this is rare)
1. The Expected Value of Perfect Information
2. The Value of the Stochastic Solution
3. Basic Inequalities
4. Estimating Performance
Computing $EV$, $SP$, $WS$, $EEV$

- Computing $EV$: single linear program
- Computing two-stage $SP$: (multi-cut) L-shaped method
- Computing multi-stage $SP$: nested decomposition, SDDP
- $EEV$ and $WS$: simulation

Notes:

- Generalization of $WS$ to multiple stages is fairly obvious
- Generalization of $EEV$ to multiple stages is not obvious
- Consider discretization of $n$ random variables at $d$ values each, exact computation of $EEV$ and $WS$ requires solving $d^n$ linear programs
Estimating $WS$ and $EEV$

Estimation of $WS$ and $EEV$ through sample mean approximation:

- For $i = 1, \ldots, K$
  - Sample $\xi_i = \xi(\omega_i)$
  - Compute $x^*(\bar{\xi})$
  - Compute $WS_i = z(x^*(\xi_i), \xi_i)$ and $EEV_i = c^T x^*(\bar{\xi}) + Q(x^*(\bar{\xi}), \xi_i)$

- Estimate $\bar{WS} = \frac{1}{K} \sum_{i=1}^{K} WS_i$ and $\bar{EEV} = \frac{1}{K} \sum_{i=1}^{K} EEV_i$
Suppose $\xi(\omega)$ is continuous, does this complicate the computation of EV, SP, EEV and WS?

Central limit theorem: Suppose $\{X_1, X_2, \ldots\}$ is a sequence of i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2 < \infty$. Then as $n$ approaches infinity, $\sqrt{n}(S_n - \mu)$ converge in distribution to a normal $N(0, \sigma^2)$:

$$\sqrt{n}\left(\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) - \mu\right) \xrightarrow{d} N(0, \sigma^2).$$

Can we use the CLT? What would the $X_i$ be in our case?
Example: Slow Convergence of Sample Average Approximation

The cost $C$ of operating a facility is

- $C(N) = 1$ under normal operations, $f(N) = 0.9$
- $C(E) = 100$ under emergency operations, $f(E) = 0.1$

$$
\mu = 0.1 \cdot 100 + 0.9 \cdot 1 = 10.9
$$

$$
\sigma = \sqrt{0.9 \cdot (1 - 10.9)^2 + 0.1 \cdot (100 - 10.9)^2} = 29.7
$$
Rare outcome (1 out of 10 samples) influences expected value calculation since it contributes by \( \frac{0.1 \cdot 100}{10.9} = 91.7\% \) to expected value.

From central limit theorem, in order to get estimate of \( \mathbb{E}[C] \) to be within 5% with 95.4% confidence, we need \( 2 \frac{\sigma}{\sqrt{n}} = 0.05 \mu \), from which \( n = 11879! \).
Figure: A sample of the evolution of the moving average $\frac{1}{n} \sum_{i=1}^{n} C(\omega_i)$ where $\omega_i$ denotes the outcome of sample $i$.

Note sensitivity of sample average to emergency outcome.
Suppose we wish to estimate $\mathbb{E}[C(\omega)]$, where $\omega$ is distributed according to $f(\omega)$

- Sample average pulls samples $\omega_i$ according to distribution $f(\omega)$ and estimates $\mathbb{E}[C(\omega)]$ with $\sum_{i=1}^{N} \frac{1}{N} C(\omega_i)$

- **Importance sampling** pulls samples $\omega_i$ according to distribution $g(\omega) = \frac{f(\omega) \cdot C(\omega)}{\mathbb{E}[C]}$ and estimates $\mathbb{E}[C(\omega)]$ with $\sum_{i=1}^{N} \frac{1}{N} \frac{f(\omega_i) \cdot C(\omega_i)}{g(\omega_i)}$
Motivation of Importance Sampling

Note that  
\[ E[C(\omega)] = \int_{\Omega} C(\omega) \cdot f(\omega) d\omega = \int_{\Omega} \frac{C(\omega) \cdot f(\omega)}{g(\omega)} g(\omega) d\omega \]

- The random variable \( \frac{C(\omega) \cdot f(\omega)}{g(\omega)} \), which is distributed according to \( g(\omega) \), also has expectation \( E[C] \)
- Which \( g(\omega) \) *minimizes* the variance of this new random variable?

\[ g(\omega) = \frac{C(\omega) \cdot f(\omega)}{E[C]} \]

*Any* sample evaluates to \( E[C] \)!

- We cheated: \( g(\omega) \) requires knowledge of \( E[C] \), which is what we are estimating
- But we learned something: pull samples according to contribution to expected value, \( \frac{C(\omega) \cdot f(\omega)}{E[C]} \). Even if we do not know \( E[C] \), we can *approximate* it.
Problem: rare ‘bad’ outcome had the greatest influence on expected value

Remedy: redefine distribution so that we observe ‘bad’ outcome earlier, then adjust our expected value calculations in order to unbias result

\[ g(\omega_1) = \frac{f(\omega_1) \cdot C(\omega_1)}{\mathbb{E}[C]} = \frac{0.9 \cdot 1}{10.9} = \frac{0.9}{10.9} \]

\[ g(\omega_2) = \frac{f(\omega_2) \cdot C(\omega_2)}{\mathbb{E}[C]} = \frac{0.1 \cdot 100}{10.9} = \frac{10}{10.9} \]

Estimates from sampling \( \omega_1, \omega_2 \) are constant and equal to \( \mathbb{E}[C] \):

\[ C(\omega_1) \cdot \frac{f(\omega_1)}{g(\omega_1)} = 1 \cdot \frac{0.9}{\frac{0.9}{10.9}} = 10.9 \]

\[ C(\omega_2) \cdot \frac{f(\omega_2)}{g(\omega_2)} = 100 \cdot \frac{0.1}{\frac{10}{10.9}} = 10.9 \]