# Statistics for Chemical Engineers:
## From Data to Models to Decisions

Victor M. Zavala

Department of Chemical and Biological Engineering
University of Wisconsin-Madison

*victor.zavala@wisc.edu*

**Chapter 7: Decision-Making Under Uncertainty**

## Motivation

- We have used *statistical techniques* to model known/unknown aspects of a system.
- We now want to use these capabilities to *make decisions*.
- Accounting for uncertainty is *essential* when making decisions (ensure robustness).
- Making decisions under uncertainty boils down to *comparing* RVs and to *shaping* the probability distribution of RVs.

# Motivation

- Discussion will address following specific questions:
  - What can happen if we ignore uncertainty?
  - How do we measure risk/uncertainty?
  - How can we obtain a decision that anticipates and responds to uncertainty?
  - How can we compare random variables?
  - How can we shape probability distribution of random variables?
- These questions are studied in area of *stochastic optimization*.
- This is a broad area that has applications in engineering, science, and finance.

# Motivation

- Consider system:

$$Y = \varphi(X, u)$$

- $X = (X_1, ..., X_n)$ is multivariate input RV and $Y$ is univariate output RV.
- System model is $\varphi$ and we interpret this as performance function (e.g., cost).
- Realizations of $X$ propagate through system to generate realizations of $Y$:

$$y_\omega = \varphi(x_\omega, u),\ \omega \in \mathcal{S}$$

- Uncertainty in $X$ propagates through system to generate variability in $Y$.
- e.g., input can be prices of energy that affect profit of a process.
- Can use decision $u$ to protect system against variability of $X$.

# Deterministic Decision-Making

- Decision-makers (DMs) often make decisions by *ignoring uncertainty*.
- This *deterministic* approach uses following logic.
- Assume $X$ takes *representative* value, such as *historical average*:

$$x = \frac{1}{S} \sum_{\omega \in \Omega} x_\omega.$$

- Input $X$ is thus assumed to be deterministic with value $x$.
- This implies that $Y$ is deterministic with value $y = \varphi(x, u)$.
- Want to compare decisions $u$ & $u'$ with outputs $y(u) = \varphi(x, u)$ & $y(u') = \varphi(x, u')$.
- If we want to minimize output, we select $u$ over $u'$ if $y(u) \leq y(u')$.
- Comparing decisions *is trivial* in deterministic setting, because we compare *numbers*.

# Deterministic Decision-Making

- Consider now that we want to find a decision $u$ that minimizes $\varphi(x, u)$.
- This is done by solving optimization problem:

$$u_D^* = \underset{u \in \mathcal{U}}{\operatorname{argmin}} \, \varphi(x, u)$$

- Here, $\mathcal{U}$ is set of possible decisions.
- Deterministic DM ignores inherent variability of $X$ seen in real-life.
- For instance, “optimal” decision $u_D^*$ is only optimal for assumed value $X = x$.
- As such, decision might be vulnerable when exposed to other outcomes of $X$.
- It is important to evaluate how $u_D^*$ behaves when exposed to other outcomes of $X$:

$$y_\omega = \varphi(x_\omega, u_D^*), \; \omega \in \mathcal{S}.$$

- This provides a complete picture of how *vulnerable (or robust)* is decision $u_D^*$.

## Monte Carlo Simulations

- An approach to account for uncertainty in decision-making is to compute decisions for different realizations of $X$.

- Consider solution of set of optimization problems:

$$u_\omega^* = \underset{u \in \mathcal{U}}{\operatorname{argmin}}\, \varphi(x_\omega, u).$$

  for realizations $x_\omega$, $\omega \in \mathcal{S}$.

- This computes a decision $u_\omega^*$ that is optimal for realization $x_\omega$ of $X$.

- This approach is known as *sensitivity analysis* or *Monte Carlo analysis* and provides useful insight into how decision changes under possible outcomes of $X$.

- However, this approach does not deliver a single (implementable) decision $u$ that is optimal *for all outcomes* of $X$.

# Example: Deterministic Decision-Making `ch7_generator_det.m`

- Facility wants to purchase on-site generator to mitigate rising electricity costs from the local power grid.
- Power demand/load of facility is random variable $L$.
- Capacity of generator is $u$ and residual demand is $R = \max(L - u, 0)$.
- Cost of residual demand purchased from grid is $\alpha_R \cdot R^2$.
- Investment cost for generator is $\alpha_u \cdot u^2$ and total cost is thus:

$$Y(u) = \alpha_R \cdot R^2 + \alpha_u \cdot u^2.$$

- To simplify decision, DM uses representative value of load $r = \max_{\omega \in \mathcal{S}} r_\omega$ (peak).
- Cost in this case is deterministic: $y(u) = \alpha_R \cdot r^2 + \alpha_u \cdot u^2$.

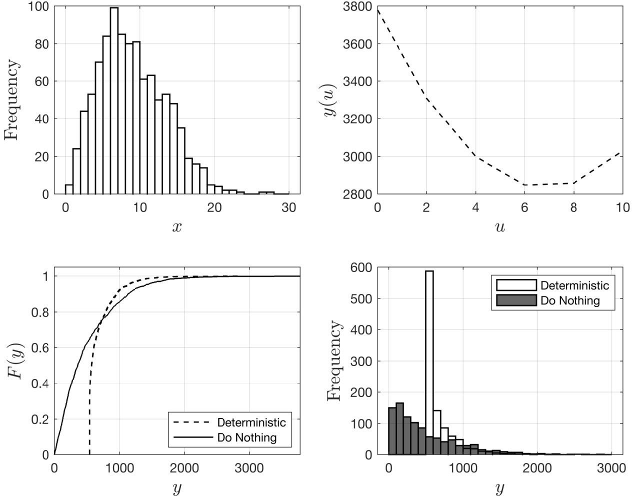
# Example: Deterministic Decision-Making `ch7_generator_det.m`

- DM has possibility of selecting generators of capacities $\mathcal{U} = \{0, 1, 2, ...., 10\}$.
- Case for $u = 0$ is "do-nothing" case (current system) and $u = 10$ is max capacity.
- Deterministic generator design problem is:

$$\min_{u \in \mathcal{U}} \; \alpha_R \cdot r^2 + \alpha_u \cdot u^2.$$

- We can see that cost is minimized with a generator of capacity $u_D^* = 6$.
- We conduct MC simulations to determine how generator performs under different load outcomes.
- Simulations generate costs $y_\omega(u) = \alpha_R \cdot r_\omega^2 + \alpha_u \cdot (u_D^*)^2$ for $\omega \in \mathcal{S}$.
- Average cost of new system $67.4$ and average cost of current system is $50.1$.
- However, new system avoids extreme/peak costs (helps with budgeting).

# Example: Deterministic Decision-Making `ch7_generator_det`



# Attitudes Towards Risk

- Given shortcomings of deterministic approach, we want to make decisions that *explicitly* take uncertainty into account.

- We denote output as $Y(u) = \varphi(X, u)$ with pdf $f_{Y(u)}$ and cdf $F_{Y(u)}$.

- Decision $u$ can be used to manipulate pdf/cdf $Y(u)$.

- The ability to manipulate pdf/cdf enables mitigation of *risk*.

- Quantifying and mitigating risk is essential.

# Attitudes Towards Risk

- Risk is often associated with a decision (e.g., “this is a risky investment”).
- Risk is often associated with extreme events (e.g., bankruptcy or system collapsing).
- There is no *unique* definition of risk.
- Because humans have different perceptions of risk.
- Humans also differ in their risk tolerance (e.g., some DMs love to take risks).
- Perceptions of risk are analogous to perceptions of happiness and fairness.
- Want to investigate what *definitions* of risk exist and what makes a definition *proper*.

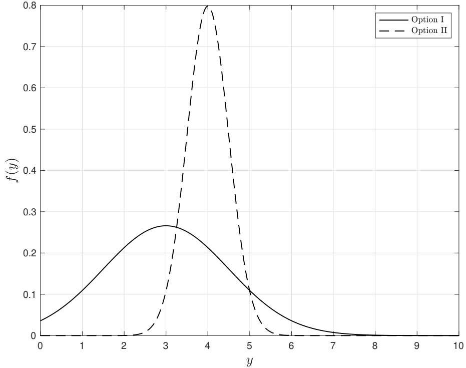
# Example: Comparing Random Variables `ch7_mean_var_comparison`

- Example illustrates how disagreements naturally arise due to different perceptions of risk.

- Have options $u$ (Option I) and $u'$ (Option II) with random costs $Y(u)$ and $Y(u')$.

- Option $u$ has low mean cost but high var; option $u'$ has high mean cost but low var.

Which option is better?



# Example: Comparing Random Variables `ch7_mean_var_comparison`

- There is no unique answer, “better” depends on how DM attributes value to different aspects of the cost distributions.

- If DM values average cost, $u$ is clearly better than $u'$, but this choice fails to capture the fact that $u$ has a a higher uncertainty and has a higher probability of incurring a large cost. In other words, option $u$ is more “risky”.

- If DM values certainty (less “risk”) they will protect themselves against outcomes of high cost. As such, they would prefer $u'$ (even if this leads to higher mean cost).

- To formalize comparisons, we need to model the preferences of decision-makers.

- This is done using *risk measures*.

## Attitudes Towards Risk

The following are typical attitudes towards risk and associated risk measures:

- *Risk-Neutral*: DM prefers $Y(u)$ over $Y(u')$ if

$$\mathbb{E}_{Y(u)} \leq \mathbb{E}_{Y(u')}$$

  DM focuses on average costs and is not concerned about extreme outcomes.

- *Risk-Conscious*: DM prefers $Y(u)$ over $Y(u')$ if

$$\mathbb{P}(Y(u) \leq \bar{y}) \geq \mathbb{P}(Y(u') \leq \bar{y})$$

  DM wants a decision that is *more likely* to achieve low outcomes.

# Attitudes Towards Risk

- *Risk-Averse*: DM prefers $Y(u)$ over $Y(u')$ if

$$\max Y(u) \leq \max Y(u').$$

  DM has pessimistic view of uncertainty (because it might lead to large losses).

- *Risk-Taker*: DM prefers $Y(u)$ over $Y(u')$ if

$$\min Y(u) \leq \min Y(u').$$

  DM has optimistic view of uncertainty (because it might lead to gains).

# Example: Attitudes Towards Risk `ch7_mean_var_comparison`

We revisit cost pdf/cdf of previous example and analyze risk attitudes.

- *Risk-neutral* DM: you select option $u$ because this has lower expected cost.

- *Risk-conscious* DM: you select threshold for cost that you are willing to tolerate (e.g., 5 MUSD). You select option $u'$ because this has a higher probability of achieving a cost that is below a threshold.

- *Risk-averse* DM: you focus on outcomes of high cost. You select $u'$ because this has no outcomes with high cost (e.g., above 6 MUSD). DM focuses on right tail of pdf.

- *Risk-taking* DM: you focus on outcomes of low cost. You select $u$ there this has outcomes of low cost (e.g., below 1 MUSD). DM focuses on left tail of pdf. Note that $u$ yields outcomes with negative cost (revenue). DM thus exploit risk.

## Example: Comparing Investment Options `ch7_npv_comparison`

- Have a couple of investment opportunities:
  - Capital $c_I = \$25,000$ with annual rate of return $R_I \sim \mathcal{N}(0.5, 0.1)$.
  - Capital $c_{II} = \$50,000$ with annual rate of return $R_{II} \sim \mathcal{N}(0.4, 0.2)$.
- Net present value for options (10 yr, interest rate of 5%) are:

$$NPV_I = c_I - R_I \cdot c_I \cdot \frac{(1-(1+0.05)^{-10})}{i}$$
$$NPV_{II} = c_{II} - R_{II} \cdot c_{II} \cdot \frac{(1-(1+0.05)^{-10})}{i}.$$

- Want NPV to be as small as possible (negative values indicate revenue).

# Example: Comparing Investment Options `ch7_npv_comparison`

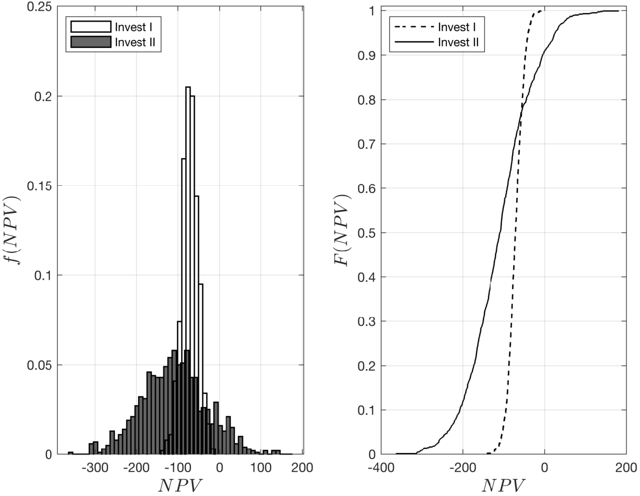



Figure: Density functions for NPVs for competing investment options.

# Example: Comparing Investment Options `ch7_npv_comparison`

- Option I has high mean but is always negative (always makes money and is “safer”).
- Moreover, Option I has a low variance (low uncertainty).
- Option II has high variance and loses money 10% of time; in some instances it can lose more than \$150,000.
- However, with Option II, we make money 90% of the time and it is possible to make a lot of money (at levels that Option I cannot reach).
- If you are risk-averse (do not want to lose money), I seems better.
- If you are a risk-taker (you might be willing to take a chance of losing money some times in exchange of large gains), II is better.

# Risk Measures

- Risk measure is a scalar function $\rho(Y(u))$ that maps RV $Y(u)$ to a number.
- Risk measure is a *summarizing statistic*.
- Risk measure aims to extract key aspects/features of RV.
- We have mentioned some common risk measures that capture risk attitudes.
- We now present more sophisticated measures and analyze them in detail.

# Risk Measures

- Mean-Variance:

$$\mathbb{MV}_\kappa[Y(u)] = \mathbb{E}[Y(u)] + \kappa \cdot \mathbb{V}[Y(u)]$$

for some weight $\kappa \in \mathbb{R}_+$; for $\kappa \to 0$ we prefer mean, for $\kappa \gg 0$ we prefer variance.

- Probability of Loss:

$$\mathbb{P}(Y(u) > \bar{y})$$

for some threshold $\bar{y} \in \mathbb{R}$. This is also often expressed as $\mathbb{P}(Y(u) \leq \bar{y})$.

- Event probability:

$$\mathbb{P}(Y(u) \in \mathcal{A})$$

where the event $\mathcal{A}$ capture conditions/constraints that $Y(u)$ should satisfy.

- Value-at-Risk:

$$\mathbb{Q}_{Y(u)}(\alpha)$$

for probability $\alpha \in [0, 1]$. This is the $\alpha$-quantile of $Y(u)$.

# Risk Measures

- Conditional Value-at-Risk (CVaR):

$$\mathbb{E}[Y(u)|Y(u) \geq \mathbb{Q}_{Y(u)}(\alpha)]$$

for some probability $\alpha \in [0, 1]$. This is also known as "expected loss".

- Mean Absolute Deviation:

$$\mathbb{E}\left[|Y(u) - \mathbb{E}[Y(u)]|\right]$$

- Worst-Case and Best-Case Value:

$$\max Y(u) \quad \text{and} \quad \min Y(u)$$

# Risk Measures

- Probability of loss resembles our colloquial definition of risk. But this fails to capture important aspects (magnitude of losses).
- There is a broad selection of risk measures that one can select.
- In fact, any statistic can be used as risk measure (e.g., mode, moments, median.
- It is also important to highlight that the selection of a risk measure should be based on nature of RV to be compared.
- e.g., comparing Gaussian RVs (symmetric RVs with left and right tails) is not the same as comparing exponential RVs (asymmetric RVs with only right tails).

## Example: Measuring Risk for Investment Optimization `ch7_npv_comparison`

We compute risk measures for investment example.

- Expected Value:
  $\mathbb{E}[Y(u)] = -\$70,891$ and $\mathbb{E}[Y(u')] = -\$109,048$
- Standard Deviation:
  $\mathbb{SD}[Y(u)] = \$19,284$ and $\mathbb{V}[Y(u')] = \$79,288$
- Probability of Loss:
  $\mathbb{P}(Y(u) > 0) = 0\%$ and $\mathbb{P}(Y(u') > 0) = 9\%$
- Probability of Gain:
  $\mathbb{P}(Y(u) \leq \bar{y}) = 6.4\%$ and $\mathbb{P}(Y(u') \leq \bar{y}) = 55.90\%$ for $\bar{y} = -\$100,000$.
- Value-at-Risk:
  $\mathbb{Q}_{Y(u)}(\alpha) = -\$46,699$ and $\mathbb{Q}_{Y(u')}(\alpha) = -\$4,935$ for $\alpha = 0.9$.
- Conditional Value-at-Risk:
  $\mathrm{CVaR}_\alpha[Y(u)] = -\$37,530$ and $\mathrm{CVaR}_\alpha[Y(u')] = \$34,810$ for $\alpha = 0.9$
- Worst-Case:
  $\max Y(u) = -\$9,129$ and $\max Y(u') = \$181,413$.
- Best-Case:
  $\min Y(u) = -\$140,609$ and $\min Y(u') = -\$363,708$.

# Stochastic Dominance

- Given that there is a wide range of possible risk measures, we can ask ourselves:

What are desirable properties of a risk measure?
What is a proper risk measure?

- An important concept that arises here is that of *stochastic dominance* (SD).

- SD provides a formal framework to compare RVs in a consistent manner.

- We say that $Y$ stochastically dominates $Y'$ (written as $Y \preceq Y'$) if:

$$\mathbb{P}(Y \leq \bar{y}) \geq \mathbb{P}(Y' \leq \bar{y}) \text{ for all } \bar{y} \in \mathcal{D}_Y$$

## Stochastic Dominance

- Notation $Y \preceq Y'$ is generalization of $y \leq y'$ (used for comparing deterministic vars).
- Sense of better defined by SD is *probabilistic* in nature.
- SD says that $Y$ is better $Y'$ if prob of $Y$ taking a value less than $\bar{y}$ is higher than prob of $Y'$ being below same threshold (for all the possible threshold values $\bar{y}$).
- We can also express SD using the cdf as

$$F_Y(\bar{y}) \geq F_{Y'}(\bar{y}) \text{ for all } \bar{y} \in \mathcal{D}_Y.$$

- In simple terms, $Y \preceq Y'$ indicates that cdf of $Y$ is above that of $Y'$ for all $\bar{y}$.

# Stochastic Dominance

- It is important to note how SD differs from other ways to compare RVs.
- We often think that a decision dominates another if this *always* better; this implies:

$$y_\omega \leq y'_\omega, \quad \omega \in \Omega$$

- This requirement is quite strict and uncommon in practice.
- SD tells us is that outcomes of $Y$ should be lower than threshold $\bar{y}$ more often than outcomes of $Y$. This is more common.
- A risk-conscious DM would require that $\mathbb{P}(Y \leq \bar{y}) \geq \mathbb{P}(Y' \leq \bar{y})$ *for a single* threshold value $\bar{y}$. SD requires that this holds for all thresholds.
- There are *relaxed* versions of SD that are also often used; e.g., might require that $\mathbb{P}(Y \leq \bar{y}) \geq \mathbb{P}(Y' \leq \bar{y})$ holds *for some* $\bar{y}$.
- SD makes it clear that, to *properly* compare RVs, it is necessary to compare their *density functions*; moreover, comparison is based on cdfs (not pdfs).
- Comparing cdfs is more intuitive.

# Coherent Risk Measures

- There are a number of properties that a proper (a.k.a. coherent) risk measure should satisfy.
- These properties have been proposed based on usefulness in *practical applications* and based on mathematical consistency.
- Interestingly and, as we we will see shortly, these properties are analogous to properties of vector norms.
- One can think of the following properties as a set of rules (a.k.a. mathematical axioms) that give us guidance into how to select a suitable risk measure.
- Similar types of axioms exist for defining proper measures of fairness.

## Coherent Risk Measures

We say that the risk measure $\rho$ is *proper* if it satisfies the properties:

- Monotonicity:

  $$\text{If } Y \preceq Y' \text{ then } \rho(Y) \leq \rho(Y').$$

  If $Y$ stochastically dominates $Y'$, then its risk should also be lower.

- Translation Invariance:

  $$\rho(Y + c) = \rho(Y) + c \text{ for } c \in \mathbb{R}.$$

  Adding a constant should not affect risk.

- Positive Homogeneity:

  $$\rho(c \cdot Y) = c \cdot \rho(Y) \text{ for } c \in \mathbb{R}_+.$$

  Scaling an RV by a constant should not affect risk (e.g., risk of a cost expressed in USD or Euros should be the same up to a factor).

- Subadditivity:

  $$\rho(Y + Y') \leq \rho(Y) + \rho(Y').$$

  The risk of a combined pair of RVs should not exceed sum of their individual risks.

# Properties of Risk Measures: Expected Value

We can now formally analyze if risk measures are proper or not.

- $\mathbb{E}[Y]$ captures the magnitude of $Y$ and has an intuitive interpretation.
- $\mathbb{E}[Y]$ is a proper risk measure; as such, it is consistent with SD.
- $\mathbb{E}[Y]$ ignores outcomes with extreme values.

## Properties of Risk Measures: Variance

- $\mathbb{V}[Y]$ is a measure of variability/uncertainty and is thus intuitive.
- Variance is often used in conjunction with expected value: $\mathbb{E}[Y] + \kappa \mathbb{V}[Y]$.
- Mean-variance was proposed by Markowitz (1950s) and became standard in finance.
- Reducing variance reduces small/large outcomes (reduces left/right tails).
- Symmetric penalization is often undesirable (only interested in right tail).
- Variance is *not proper*; it does not satisfy monotonicity (not consistent with SD).

# Properties of Risk Measures: Expected Loss

- Expected loss (CVaR) overcomes deficiencies of variance and has become widely used in finance.
- Expected loss of $Y$ at a given probability $\alpha$ is:

$$\mathbb{E}[Y|Y \geq \mathbb{Q}_Y(\alpha)]$$

- Recall that $\alpha$-quantile of $Y$ is threshold value $\bar{y}$ satisfying

$$\mathbb{P}(Y \leq \bar{y}) = \alpha$$

- Expected loss is expected value of costs above $\bar{y} = \mathbb{Q}_Y(\alpha)$.
- Expected loss can also be written as:

$$\mathbb{E}[(Y - \mathbb{Q}_Y(\alpha))_+]$$

# Properties of Risk Measures: Expected Loss

- CVaR captures magnitude of outcomes of high value, while ignoring those of small magnitude (asymmetric measure).
- This provides important benefit over the variance, as it only focuses on right tail.
- If $\alpha = 0$, quantile $\mathbb{Q}_Y(\alpha)$ is min value of $Y \implies$ CVaR is $\mathbb{E}[Y]$.
- If $\alpha = 1$, quantile $\mathbb{Q}_Y(\alpha)$ is max value of $Y \implies$ CVaR is $\max Y$.
- CVaR is a proper risk measure.
- CVaR captures size of losses above quantile.
- A caveat of CVaR is that it does not offer direct control on probability of loss.

# Properties of Risk Measures: Probability of Loss

- The probability of loss is one of most widely used measures of risk.
- The meaning of probability of loss is intuitive in many applications and is a widely used of notion of risk.
- This measure does not take into consideration the actual magnitude of the losses.
- Probability of loss is *not coherent*; it does not satisfy translation invariance.
- However, probability of loss is directly compatible with SD (compares probabilities).

## Risk Measures as Norms

- Coherency of risk measures resemble coherency of *vector norms* (e.g., triangle inequality is subadditivity).

- Norms allow us to measure and compare vectors.

  How can we tell if vector $\mathbf{y} = (y_1, y_2, ..., y_S)$ is better than $\mathbf{y}' = (y_1', y_2', ..., y_S')$?

- We can say that $\mathbf{y}$ is better than $\mathbf{y}'$ if total magnitude of its entries is lower:

$$\sum_{i=1}^{S} y_i \leq \sum_{i=1}^{S} y_i'.$$

  Can be written as $\frac{1}{S}\sum_{i=1}^{S} y_i \leq \frac{1}{S}\sum_{i=1}^{S} y_i'$ (analogous to mean and $\ell_1$ norm $\|\mathbf{y}\|_1$).

- We can say that $\mathbf{y}$ is better than $\mathbf{y}'$ if:

$$\max_i \; y_i \leq \max_i \; y_i'.$$

  Analogous to worst-case measure and to an $\ell_\infty$ norm $\|\mathbf{y}\|_\infty$.

# Risk Measures as Norms

- We can say that $\mathbf{y}$ is better than $\mathbf{y}'$ if:

$$y_i \leq y_i' \text{ for all } i = 1, ..., S.$$

  Analogous to say that RV $Y$ is *always better* than the RV $Y'$.

- This comparison is *not consistent* because te vector entries of $\mathbf{y}$ and $\mathbf{y}'$ are not arranged in order.

- More consistent approach would first arrange entries in decreasing order and then compare entries.

- This is *precisely analogous* to SD; we establish a threshold on entries and count how many entries are below that threshold value.

- Another possibility is to arrange entries in increasing order and sum the largest $k$ entries. This is *directly analogous to CVaR*.

# Example: Risk Measures as Norms

Consider the vector $y = (100, 70, 50, 20, 10)$

- Mean is:
  $(1/5) \cdot (100 + 70 + 50 + 20 + 10) = 50$
- Worst-Case is: $100$

The CVaR values for vector is:

- If $\alpha = 5/5$ then CVaR is:
  $(1/5) \cdot (100 + 70 + 50 + 20 + 10) = 50$
- If $\alpha = 4/5$ then CVaR is:
  $(1/4) \cdot (100 + 70 + 50 + 20) = 60$
- If $\alpha = 3/5$ then CVaR is:
  $(1/3) \cdot (100 + 70 + 50) = 73$
- If $\alpha = 1/5$ then CVaR is:
  $(1/1) \cdot (100) = 100$

Note how the CVaR has mean and worst-case values as extremes.

# Stochastic Optimization

- Consider now the situation in which we do not want to compare decisions but we would like to *find best decision* possible.

- We formulate optimization problems that explicitly take uncertainty into account.

- These will use risk measures to optimally shape the distribution of $Y(u)$.

- We discuss important notions of *conflict resolution* (trading-off risk measures) and *recourse* (corrective decisions).

# Stochastic Optimization: Basic Formulation

- Ultimate goal is to find decision $u$ that optimizes $Y(u) = \varphi(X, u)$.

- Complication arises because $\varphi(X, u)$ is an RV; as such, problem:

$$u^* = \underset{u \in \mathcal{U}}{\text{argmin}} \ \varphi(X, u)$$

  does not make mathematical sense.

- To formulate problem properly, we optimize a risk measure:
  - Mean-Variance:

$$u^* = \underset{u \in \mathcal{U}}{\text{argmin}} \ \mathbb{E}[\varphi(X, u)] + \kappa \mathbb{V}[\varphi(X, u)]$$

  - Probability of loss:

$$u^* = \underset{u \in \mathcal{U}}{\text{argmin}} \ \mathbb{P}(\varphi(X, u) > y)$$

  - Expected shortfall:

$$u^* = \underset{u \in \mathcal{U}}{\text{argmin}} \ \mathbb{E}[\varphi(X, u) \,|\, \varphi(X, u) \geq \mathbb{Q}(\alpha)]$$

# Stochastic Optimization: Basic Formulation

- Stochastic formulations obtain decisions $u^*$ that protect us (hedge) against uncertainty (or exploit uncertainty).

- Type of protection/exploitation depends on risk measure that we use.

- *Ultimate goal* is to manipulate shape of pdf/cdf of $Y(u)$.

- Observation is key, because it is quite difficult (if not impossible) to have full control over shape of pdf/cdf.

- For instance, by selecting to minimize a specific aspect of RV (e.g., average), we might need to sacrifice another aspect (e.g., probability of loss or variance).

- This will introduce *conflicting trade-offs*.

# Example: Deterministic vs. Stochastic Decision-Making `ch7_generator_stoch`

- Consider generator sizing problem and assume now that DM would like to size system in a way that it captures uncertainty of demand.

- DM formulates optimization problem that minimizes CVaR of cost (with $\alpha = 0.9$):

$$\min_{u \in \mathcal{U}} \; \mathrm{CVaR}_{\alpha}[Y(u)].$$

- Stochastic problem selects a generator with capacity of $u^* = 4$, which is smaller than deterministic $u_D^* = 6$.

# Example: Deterministic vs. Stochastic Decision-Making `ch7_generator_stoch`

- We compare the pdf/cdf for the deterministic and stochastic design.
- It is clear that stochastic solution reduces overall costs.
- Stochastic solution offers same level of protection to extreme costs as deterministic.
- This indicates that deterministic solution is *unnecessarily* conservative.

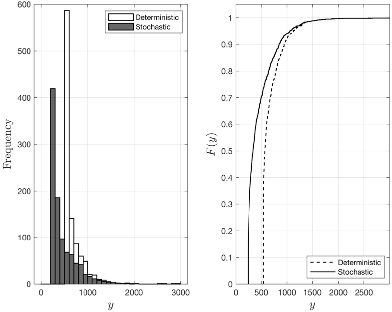



Figure: Pdf and Cdf for generator sizing problem using deterministic and stochastic decision-making.

## Conflict Resolution

- To deal with conflicting measures, we can minimize a risk measure, while constraining another.
- This can be formulated as problem:

$$\begin{aligned} \min_{u \in \mathcal{U}} \;& \rho_o(Y(u)) \\ \text{s.t.} \;& \rho_c(Y(u)) \leq \bar{\rho} \end{aligned}$$

- Solution of problem for different values of $\bar{\rho}$ gives family of optimal solutions that is known as *trade-off (Pareto) frontier.*
- From trade-off frontier, a DM can select a *compromise solution*.
- Can also construct frontier by minimizing weighted sum of risk measures:

$$\min_{u \in \mathcal{U}} \; \rho_o(Y(u)) + \kappa \cdot \rho_c(Y(u)),$$

where $\kappa \in \mathbb{R}_+$ is weighting factor.

# Example: Conflict Resolution in Sizing `ch7_generator_conflict`

- DM is now interested in exploring trade-off between expected cost and variance:

$$\min_{u \in \mathcal{U}} \mathbb{E}[Y(u)] + \kappa \cdot \mathbb{SD}[Y(u)].$$

- DM explores trade-off by trying different values of $\kappa \in [0, 20]$.
- When $\kappa = 0$, generator is sized to minimize mean cost (while ignoring large costs).
- When $\kappa$ is large, generator is sized to try to reduce large costs.
- DM is aware that $\mathrm{CVaR}_\alpha[Y(u)]$ can capture similar behavior (by tuning $\alpha$).
- Specifically, when $\alpha = 0$, CVaR is $\mathbb{E}[Y(u)]$ and, when $\alpha = 1$, CVaR is $\max Y(u)$.
- DM thus solves problem:

$$\min_{u \in \mathcal{U}} \mathrm{CVaR}_\alpha[Y(u)]$$

for different values of $\alpha = [0, 1]$.

## Example: Conflict Resolution in Sizing `ch7_generator_conflict`

- In top-left, can see that there is conflict between mean cost and its variance as we span $\kappa$.
- In top-right, we see how optimal capacity increases as we put more emphasis on SD.
- In bottom row we se trade-off curve of mean and SD as we span $\alpha$. This has a similar effect as spanning $\kappa$.
- This highlights how CVaR can be used to explore trade-offs. However, we notice differences in size of generators.
- Difference is due to asymmetric penalization of CVaR vs. symmetric of SD.

# Example Conflict Resolution in Sizing `ch7_generator_conflict`

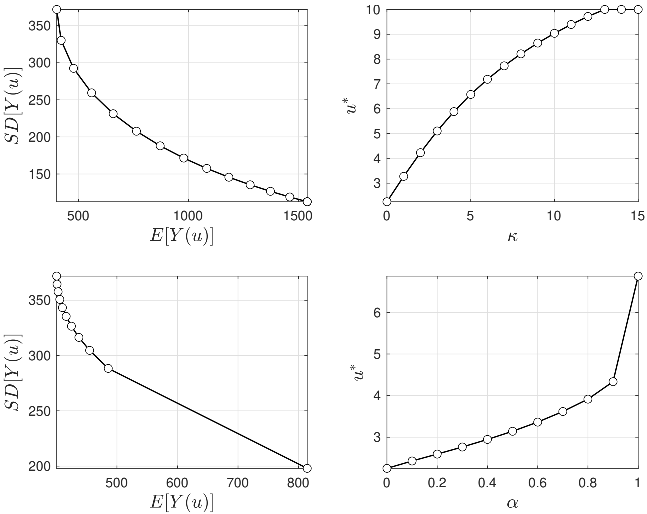



Figure: Trade-off curve obtained as we span $\kappa$ in mean-variance risk measure (top). Trade-off curve obtained as we span $\alpha$ in CVaR risk measure (bottom).

# Stochastic Optimization: Formulations with Recourse

- Stochastic formulation finds a decision $u$ that protect us against uncertainty.
- This decision is taken in *anticipation* of uncertainty.
- However, in practical situations, we also have ability to protect ourselves by using *corrective actions* that are taken once we face a particular outcome of uncertainty.
- Such corrective actions are known as *recourse of feedback*.
- Consider, *design of a chemical process* in the face of uncertain market conditions.
- The design sets the size and layout of equipment.
- Once process is installed, we can correct how we run the process to respond to market conditions (e.g., by correcting operating conditions).
- *Sequential* decisions capture how humans anticipate & respond to uncertainty.

## Stochastic Optimization: Formulations with Recourse

- Stochastic optimization can capture sequential decisions; consider the system:

$$Y = \varphi(X, u, W(X))$$

- Decision $u$ is made in anticipation of uncertainty (preemptive).
- Decision $W(X)$ is made in response to outcomes of $X$ (corrective).
- Note that $W(X)$ is an RV (a.k.a. recourse/feedback policy).
- Formulate problem that gives best preemptive $u$ and corrective $W(X)$ decisions:

$$\min_{u \in \mathcal{U}, W(X) \in \mathcal{W}(u)} \; \rho_o(\varphi(X, u, W(X))).$$

  where $\mathcal{W}(u)$ is the set of possible corrective decisions.
- Analogous formulations can be considered that capture risk constraints.
- Optimization formulation is known as two-stage formulation.

## Example: Two-Stage Formulation for Sizing `ch7_generator_twostage.m`

- Want to size generator to control costs for demands $X = (X_1, X_2)$.
- Generator can be operated flexibly; power output $W(X))$ satisfies $W(X) \leq u$.
- Generator power can be adjusted depending on load conditions $X$ observed.
- Amount of residual load requested from power grid is:

$$R = X_1 + X_2 - W(X)$$

The cost function of system is:

$$Y(u, W(X)) = \alpha_R \cdot R^2 + \alpha_u \cdot u^2.$$

- Assume we can only select between a couple of capacities $\mathcal{U} = \{1, 10\}$.
- To prevent power waste, we assume $0 \leq W(X) \leq X_1 + X_2$ and thus:

$$\mathcal{W}(u) = \{0 \leq W(X) \leq X_1 + X_2,\ W(X) \leq u\}.$$

- This set depends on decision $u$ and it dictates how flexible the system can be.

# Example: Two-Stage Formulation for Sizing `ch7_generator_twostage.m`

- We find optimal sizing by formulating two-stage stochastic problem:

$$\min_{u \in \mathcal{U}, W(X) \in \mathcal{W}(u)} \; \mathrm{CVaR}_\alpha \left[ Y(u, W(X)) \right].$$

- Solution of problem yields generator size of $u^* = 1$.

- Compare against deterministic solution (assume worst-case values of $X_1, X_2$):

$$\begin{aligned} \min_{u \in \mathcal{U}} \;& \alpha_R \cdot r^2 + \alpha_u \cdot u^2 \\ \text{s.t. } & r = x_1 + x_2 - w \\ & 0 \leq w \leq x_1 + x_2 \\ & w \leq u. \end{aligned}$$

- Size that results from this problem is $u_D^* = 10$ (more conservative).

## Stochastic Optimization: Shaping Formulations

- Goal of stochastic optimization is to *shape* distribution of $Y(u)$.
- Instead of using risk measures, consider a *target* $Y^*$ that captures *ideal behavior*.
- We aim to minimize distance between $Y(u)$ and target $Y^*$.
- Distance between RVs can be measured as:

$$\mathbb{D}(Y(u), Y^*) = \int_{y \in \mathcal{D}_Y} |F_{Y(u)}(y) - F_{Y^*}(y)| dy.$$

- This is known as *Wasserstein distance*.
- Formulate a stochastic optimization problem that minimizes the distance:

$$\min_{u \in \mathcal{U}} \ \mathbb{D}(Y(u), Y^*).$$

- Formulation is analogous to how optimal trajectories are obtained (e.g., for rockets).
- e.g., compute a trajectory that is as close as possible to target trajectory.

# Quantifying Flexibility and Robustness

- We are often interested in exploring how *flexible/robust* a system is.
- Consider performance $\varphi(X,u)$; operation is *feasible* if event $\varphi(X,u) \leq \bar{\varphi}$ holds.
- Conversely, operation is deemed *infeasible* if event $\varphi(X,u) > \bar{\varphi}$ holds.
- Want to determine prob that system has feasible operation under decision $u$:

$$\delta(u) = \mathbb{P}(\varphi(X,u) \leq \bar{\varphi}).$$

- $\delta(u) \in [0,1]$ is a *measure of flexibility/robustness*.
- We can find decision $u$ that maximizes flexibility $\delta(u)$ by solving:

$$u^* = \underset{u \in \mathcal{U}}{\text{argmax}} \ \delta(u).$$

- Consider now that system is described by *performance functions* $\varphi_k(X,u)$, $k \in \mathcal{K}$.
- We can measure the flexibility of this system by using probability:

$$\delta(u) = \mathbb{P}(\varphi_k(X,u) \leq \bar{\varphi}_k, \ k \in \mathcal{K}).$$

## Bayesian Optimization

- Bayesian Optimization (BO) is a special type of stochastic optimization algorithm.
- BO provides "closed-loop" paradigm that integrates experimental design/data collection, modeling/learning, uncertainty quantification, and decision-making.
- Consider system:

$$Y(u) = \varphi(u) + \epsilon$$

- $\varphi(u)$ is performance, $\epsilon$ is random noise, $u$ is decision, and $Y$ is observed output.
- Decision $u$ represents conditions under which we can run the system (e.g., experimental design).
- Goal of BO is to find conditions $u$ that maximize system performance:

$$\max_{u \in \mathcal{U}} \;\; \varphi(u)$$

# Bayesian Optimization

- Challenge is that we *do not know* system function $\varphi(u)$ (system is *black box*).
- We need to build a *data-driven* model using data $(u_\omega, y_\omega)$, $\omega \in \mathcal{S}$.
- This model is known as a *surrogate* model and is used to predict performance.
- Can use diverse models (e.g., linear regression, physics, neural nets, basis).
- Popular modeling approach used in BO is *kriging*, as it captures nonlinear behavior and quantifies uncertainty.

# Bayesian Optimization

- Imagine we have a kriging model built using data $(u_\omega, y_\omega)$, $\omega \in \mathcal{S}$.
- Kriging provides mean prediction model $m_\ell(x)$ and uncertainty model $\sigma_\ell(u)$.
- Define initial dataset as $\mathcal{D}_\ell$ and initial model as $(m_\ell(u), \sigma_\ell(u))$.
- Here, $\ell$ is an iteration counter.
- We can use prediction/uncertainty model for a couple of tasks:
  - Compute decision $u_{\ell+1}$ that maximizes the expected performance:
  $$\max_{u \in \mathcal{U}} \; m_\ell(u).$$
  - Compute decision $u_{\ell+1}$ that maximizes information:
  $$\max_{u \in \mathcal{U}} \; \sigma_\ell(u).$$
- 1st task is *exploitation*, as it exploits knowledge $(m_\ell(u))$ to max performance.
- 2nd task is *exploration*, as it explores input space to max info $(\sigma_\ell(u))$.

# Bayesian Optimization

- Select new decision $u'$ that balances exploration/exploitation by solving:

$$\max_{u \in \mathcal{U}} \; m_\ell(u) + \kappa \sigma_\ell(u).$$

- Objective is known as the *acquisition function* $AF_\ell(u) = m_\ell(u) + \kappa \cdot \sigma_\ell(x)$.
- $\kappa \in \mathbb{R}_+$ is parameter that balances exploration/exploitation.
- Run experiment $u'$ in the system to obtain new data point $(u', y')$ and augment dataset $\mathcal{D}_{\ell+1} \leftarrow \mathcal{D}_\ell \cup (u', y')$.
- Use new dataset to re-train kriging model $\mu_{\ell+1}(u), \sigma_{\ell+1}(u)$ and use this to obtain new input that maximizes $AF_{\ell+1}(u)$.
- Loop is repeated over iterations $\ell, \ell+1, \dots$ .
- BO selects experiments to progressively improve model and maximize performance.
- BO captures how humans navigate world by collecting data, extract knowledge from data (learn models), and leverage knowledge to make decisions.

# Example: BO for Experimental Design `ch7_bayesopt_yield.m`

- A team of scientists want to determine optimal conditions to extract pectin from sunflower heads using oxalic acid as a solvent.
- Need to identify how how pectin yield ($Y$) is influenced by $u_1$ (extraction temperature), $u_2$ (extraction time), and $u_3$ (oxalic acid concentration).
- Unfortunately, there is *no physical understanding* of system and yield thus needs to optimized directly using data.
- Apply BO to automatically identify experimental conditions that maximize yield.

## Example: BO for Experimental Design `ch7_bayesopt_yield.m`

- Construct an initial kriging model with 3 experiments.
- Subset of experiments proposed by BO and evolution of yield are shown below.
- BO first runs the 3 inputs at their highest values to reduce uncertainty of model.
- BO then progressively lowers the values of the inputs but eventually realizes that lowering temperature $(u_1)$ is not a good strategy.
- As such, it decides to increase temperature while, continuing to lowering time $(u_2)$ and oxalic acid conc $(u3)$.
- Yield is improved quickly and it eventually saturates, converging to the conditions $u_1 = 1$, $u_2 = 0.426$, and $u_3 = 0.24$.
- BO identifies optimal conditions in less than 30 experiments.

# Example: BO for Experimental Design `ch7_bayesopt_yield.m`

| | | |
|---|---|---|
| 1.0000 | 1.0000 | 1.0000 |
| 1.0000 | 1.0000 | 1.0000 |
| 0.9946 | 0.9946 | 0.9692 |
| 0.8355 | 0.5647 | 0.5133 |
| 0.8697 | 0.7806 | 0.3191 |
| 1.0000 | 0.6770 | 0.2967 |
| ⋮ | ⋮ | ⋮ |
| 1.0000 | 0.4505 | 0.2219 |
| 1.0000 | 0.4426 | 0.2283 |
| 1.0000 | 0.4356 | 0.2338 |
| 1.0000 | 0.4300 | 0.2380 |
| 1.0000 | 0.4259 | 0.2411 |
| 1.0000 | 0.4259 | 0.2411 |

Table: Subset of experiment history suggested by BO.

# Example: BO for Experimental Design `ch7_bayesopt_yield.m`

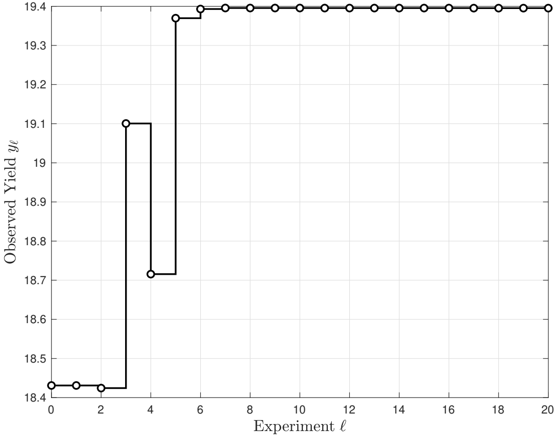



Figure: Evolution of yield using experiments proposed by BO.