60615A : Decision Analysis
Monte Carlo Simulation Part I

Prof. Carolina Osorio

Slides of Prof. Erick Delage, in collaboration with Prof. Geneviève Gauthier
Department of Decision Sciences
HEC Montréal

Outline

Outline

1. Statistical analysis
2. Evaluate the estimation error

Empirical distribution

- Definition : the empirical distribution is obtained by giving an equal probability to each values in a scenario set, called a sample.
- The empirical distribution based on the sample $Z_1, Z_2, ..., Z_M$ is

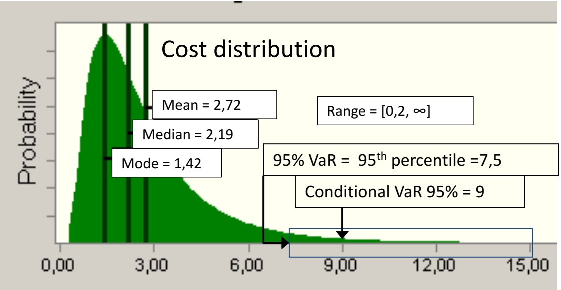
$$F_Z^M(z) = \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}\{Z_i \leq z\}$$

  In fact, $F_Z^M(z)$ is the proportion of time we observe that Z_i is smaller or equal to z in the sample.
- The analysis of a Monte-Carlo simulation is done assuming that the empirical distribution captures the likelihood of all possible realizations.
- Example : see nuclear plant case study (Excel file)

Relevant statistics

There exists several measures useful to quantify risk :

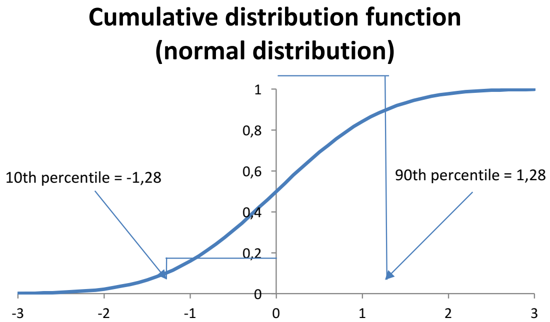
- Expected value
- Standard deviation
- Range
- Probability of reaching a target
- Percentile
- Value-at-risk
- Conditional value-at-risk
- Expected shortfall



Percentile / Value-at-risk / Conditional value-at-risk

- Definition : the n-th percentile is the smallest value below which we are assured that the random variable has more than $n\%$ chances of occurring.
  - If $F_Z(\cdot)$ is continuous (and strictly increasing), the n-th percentile is the value z for which $F_Z(z) = n/100$
- Definition : the value-at-risk for a level of confidence of α is the value of losses for which we believe we have a $\alpha \times 100\%$ chances of doing better.
  - If $F_Z(\cdot)$ is continuous and Z is a cost, $VaR_\alpha(Z)$ is the value z for which $F_Z(z) = \alpha$ (for a profit it is $F_Z(-z) = 1 - \alpha$)
- Definition : the conditional value-at-risk for a level of confidence of α is the expected value of losses conditional to obtaining a worst value than $VaR_\alpha(Z)$.
  - If Z is a cost, we can compute it with : $E[Z|Z \geq VaR_\alpha(Z)] = E[Z \cdot \mathbb{1}\{Z \geq VaR_\alpha(Z)\}]/(1-\alpha)$

Example



- 10-th percentile = -1,28 ; 90-th percentile = 1,28
- VaR-90% = 1,28 (wheter Z is a profit or an expense)
- CVaR-90% $= E[Z|Z \geq 1,28] \approx 1.75$ (if $Z =$ expense), $-E[Z|Z \leq -1,28] \approx 1.75$ (if Z is a profit)

Stochastic dominance

- Definition : the alternative x_A dominates stochastically the alternative x_B if for every level of performance, the alternative x_A has a higher chance of surpassing this level than the alternative x_B.
- Mathematically, if $g(x, Z) = \text{profit}$:

$$P(g(x_A, Z) \geq \alpha) \geq P(g(x_B, Z) \geq \alpha)\,,\ \forall\, \alpha \in \mathbb{R}$$
$$P(g(x_A, Z) \leq \alpha) \leq P(g(x_B, Z) \leq \alpha)\,,\ \forall\, \alpha \in \mathbb{R}$$

- Consequence of interest : every person that maximizes expected utility will agree on the fact that x_A dominates x_B.
- Can be verified by comparing the distribution functions.

Deterministic dominance

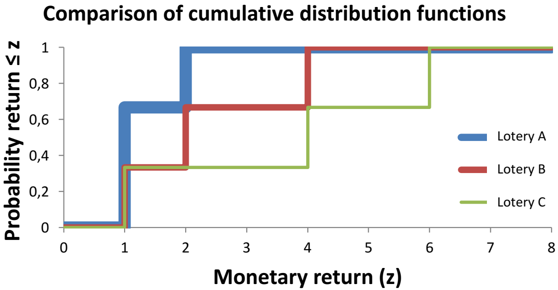
- Definition : the alternative x_A dominates deterministically the alternative x_B if we are guaranteed that the outcomes of x_A will be preferred to the outcomes of x_B under all possible realizations of the uncertain parameters.
- Mathematically, if $g(x, Z) = \text{profit}$:

$$P(g(x_A, Z) \geq g(x_B, Z)) = 1$$

- Sufficient condition : x_A dominates x_B if the most negative outcomes of the alternative x_A are preferred to the most positive outcomes of the alternative x_B.
- Can't always be verified by comparing the distribution functions.
- Determ. dominance implies stoch. dominance but not the reverse.

Example of Dominance

- Consider the 3 lotteries related to the roll of a dice :
  - Lottery A : (1-2) win 1$, (3-4) win 1$, (5-6) win 2$
  - Lottery B : (1-2) win 1$, (3-4) win 2$, (5-6) win 4$
  - Lottery C : (1-2) win 6$, (3-4) win 4$, (5-6) win 1$



- «A is determ. dominated by B but not by C» , «A & B are stoch. dominated by C»

Outline

Estimation of the expected value

- Let $Z_1, Z_2, ..., Z_M$ be the simulated values of the uncertain parameters of the problem.
- In fact, the values $g(x, Z_1), g(x, Z_2), ..., g(x, Z_M)$ are independent and identically distributed random variables since each is the result of a Monte-Carlo simulation.
- The Monte-Carlo estimator approximates $E[g(x, Z)]$ with

$$\hat{\theta}_M(x, Z) = \frac{1}{M}\sum_{i=1}^{M} g(x, Z_i)$$

- Note that the estimated value is a random variable : i.e., the values estimated by two analysts will usually not be the same.

Estimation of the expected value

The Monte-Carlo estimator is :

- Without bias : $E[\hat{\theta}_M(x,Z)] = E[g(x,Z)]$
- Convergent : $Var[\hat{\theta}_M(x,Z)] = Var[g(x,Z)]/M \rightarrow 0$ when $M \rightarrow \infty$
  I.e., the strong law of large numbers guarantees that $\hat{\theta}_M$ converges almost surely to $E[g(x,Z)]$.

$$P(\lim_{M\rightarrow\infty} |\hat{\theta}_M - E[g(x,Z)]| = 0) = 1$$

- Asymptotically normal : by the central limit theorem :

$$\frac{\hat{\theta}_M - E[g(x,Z)]}{\sigma/\sqrt{M}} \rightarrow \mathcal{N}(0,1) \text{ when } M \rightarrow \infty$$

where $\sigma^2 = Var[g(x,Z)]$

Confidence interval for the expected value (Excel file)

- Let $\hat{\theta}_M$ be a Monte-Carlo estimator of $E[g(x, Z)]$.
- Let $\hat{\sigma}^2_M$ be the unbiased MC estimator of $Var[g(x, Z)]$.

$$\hat{\sigma}^2_M = \frac{1}{M-1}\sum_{i=1}^{M}(g(x, Z_i) - \hat{\theta}_M)^2$$

- When the size of M is sufficiently large,

$$\hat{\theta}_M \approx \mathcal{N}(E[g(x, Z)], \hat{\sigma}^2_M/M)$$

- We can therefore estimate a confidence interval of level $1 - \alpha$

$$E[g(x, Z)] \in \left[\hat{\theta}_M - \Phi^{-1}(1 - \frac{\alpha}{2})\frac{\hat{\sigma}_M}{\sqrt{M}}, \hat{\theta}_M + \Phi^{-1}(1 - \frac{\alpha}{2})\frac{\hat{\sigma}_M}{\sqrt{M}}\right]$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of the standard normal distribution $\mathcal{N}(0, 1)$.

MC Quantile Estimator

- Definition : for every $\alpha \in (0,1)$, the MC estimator $\hat{z}_\alpha^M$ of the α-quantile of the distribution F_X is the α-quantile of the empirical distribution $F_Z^M(z)$.
- In practice :
  - Considering that $Z_{(1)}, Z_{(2)}, \ldots, Z_{(M)}$ is the ordered version of the sample $Z_1, Z_2, \ldots, Z_M$
  - We can measure
$$\hat{z}_\alpha^M = Z_{(\lceil M\alpha \rceil)}$$
where $\lceil x \rceil$ is the smallest superior integer to x.
- Confidence interval of level $1-\beta$:

$$Z_{(\lfloor M(\alpha-\Delta_\beta)\rfloor)} \le z_\alpha \le Z_{(\lceil M(\alpha+\Delta_\beta)\rceil+1)}$$

where $\Delta_\beta = \Phi^{-1}(1-\frac{\beta}{2})\sqrt{\alpha(1-\alpha)}/\sqrt{M}$

Confidence interval - Example

- Suppose the size of the sample is 10000, what are the indices of the ordered sample list that allow us to determine a 95% confidence interval for the 10th percentile ?
  Hint : $\Phi^{-1}(97.5\%) \approxeq 1.96$

- The margin of error on the index of this percentile is $\Delta = 1.96 \times \sqrt{0.1(1-0.1)/10000} = 0.0059$
- The minimum index is therefore $\lfloor 10000 \times (0.1 - 0.0059) \rfloor = 941$
- The maximum index is therefore $\lceil 10000 \times (0.1 + 0.0059) \rceil + 1 = 1060$