Sample average approximation (SAA)

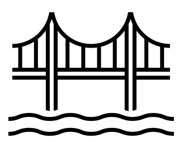
Iiro Tynnilä

Outline

1. **Motivation and background**
2. **Sample average approximation**
   - Sample average & Central Limit Theorem
3. **Convergence**
4. **Optimality gap**
5. **Performance and examples**
6. **Conclusion**

Background

- **Stochastic programming solves optimization problems involving uncertainty**
- **Often useful data is collected to help decision making**
- **We want solution that performs well on average**
- **Important fact is, that even if uncertainty is involved, it follows known distribution or it can be estimated**



Background

- **Previously we learnt that size of the models increases fast**
- **Stochastic programming is computationally intense**
- **"linear two-stage stochastic programs with fixed recourse are #P-hard even if the random problem data is governed by independent uniform distributions"**

(Hanasusanto et al., 2016)

What is SAA?

- **Objective is to reduce scenario set to manageable size**
- **This is done with sampling process**
- **“True” value is then estimated from the sample**
- **Common method is to use random style Monte Carlo sampling**
- **Technique can be viewed as a part of scenario generation**

Reminder, two-stage problem

Familiar formulation for two-stage problem

$$\min_{x \in X} \{ g(x) := c^\top x + \mathbb{E}[Q(x,\xi)] \}$$

(Shapiro,Pilpott,2007)

where $[Q(x,\xi)]$ is the optimal value for second-stage and ξ is vector containing the data

SAA

Expected function of the random vector

$$q(x) = \mathbb{E}\left[Q(x, \xi)\right]$$

Usually vector ξ is very large, thus we generate a N-sized sample

$$\xi^1, \xi^2, \xi^3, \dots, \xi^N$$

Conveniently we know that each $\xi^j, j = 1,,, N$ has the same probability distribution as ξ and that the sample is iid.

(Shapiro,Pilpott,2007)

SAA

Then we can approximate the expected function $q(x)$,

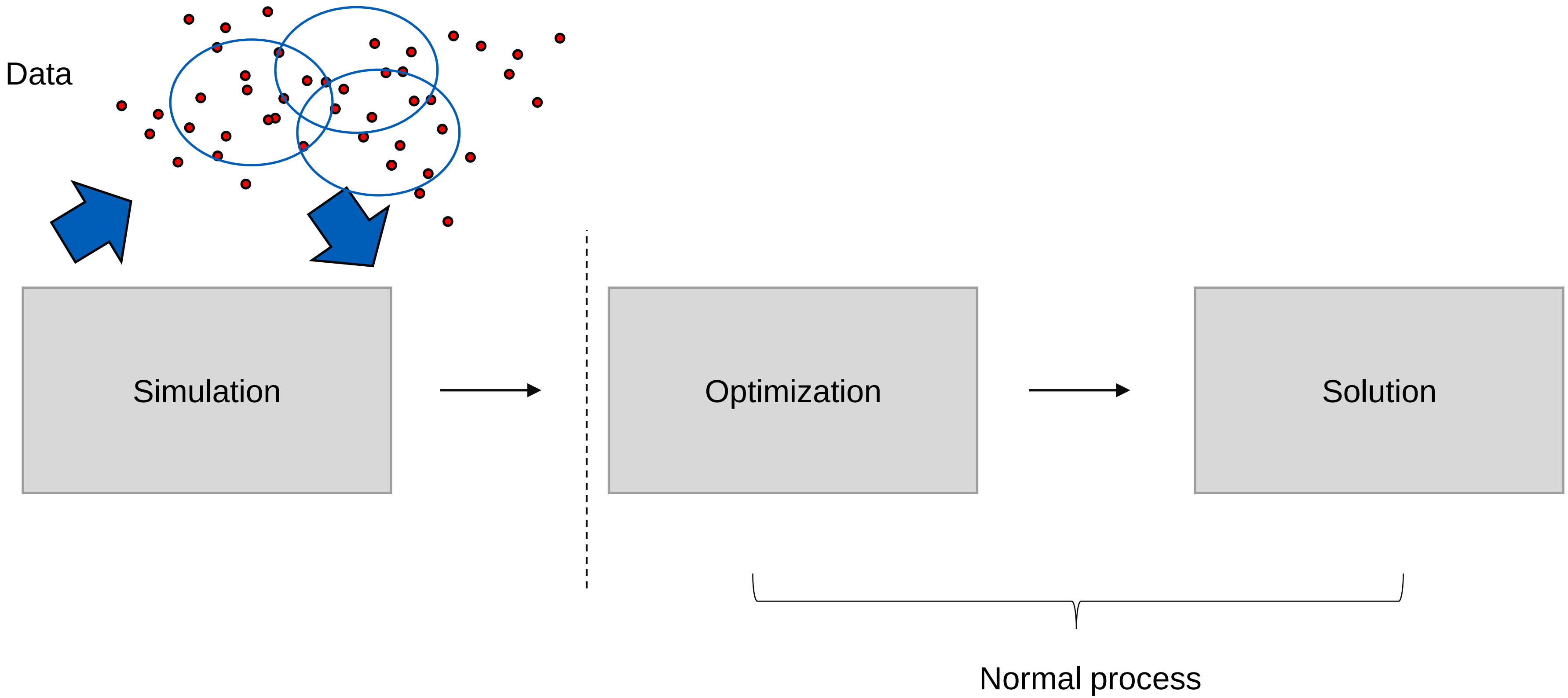
$$\hat{q}_N(x) = \frac{1}{N}\sum_{j=1}^{N} Q(x, \xi^j)$$

And the optimization problem by

$$\min_{x \in X}\left\{\hat{g}_N(x) := c^{\mathsf{T}}x + \frac{1}{N}\sum_{j=1}^{N} Q(x, \xi^j)\right\}$$ (Shapiro,Pilpott,2007)

Note that SAA is just the same problem with each simplified scenario having the same probability $\frac{1}{N}$

SAA



SAA

Notes:

- **SAA is not a solution algorithm or technique**
- **One has to still solve the problem using appropriate method**
- **Obtained SAA problem is just easier to solve**

Monte Carlo-method

- **The obtained $\hat{q}_N(x)$ is an unbiased estimator for $q(x)$**
- **In Monte Carlo sampling each ξ^j is independent of each other and then by the Central Limit Theorem we know that**

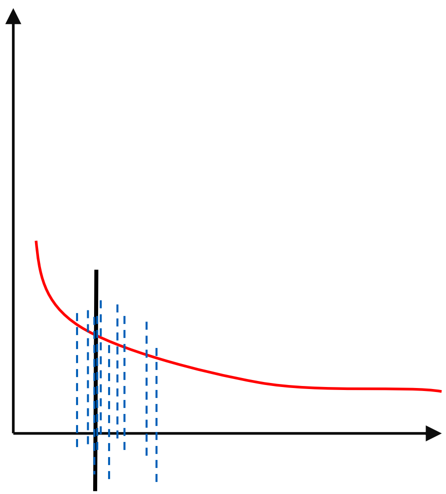
$$VAR[\hat{q}_N(x)] = \frac{1}{N}\sigma^2(x)$$
$$\sigma^2\,(x) := VAR[Q(x,\xi)]$$

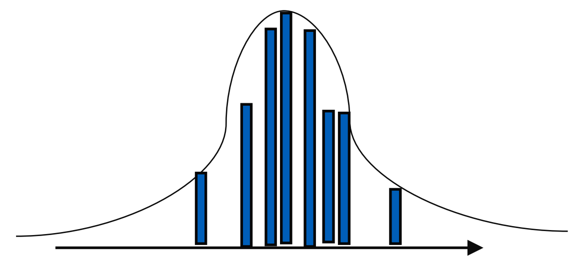
(Linderoth, Shapiro, Wright, 2006)

Central limit theorem

- **Reminder:**
- **The central limit theorem states that the sampling distribution of the sample mean approaches a normal distribution as the size of the sample grows**

Central limit theorem

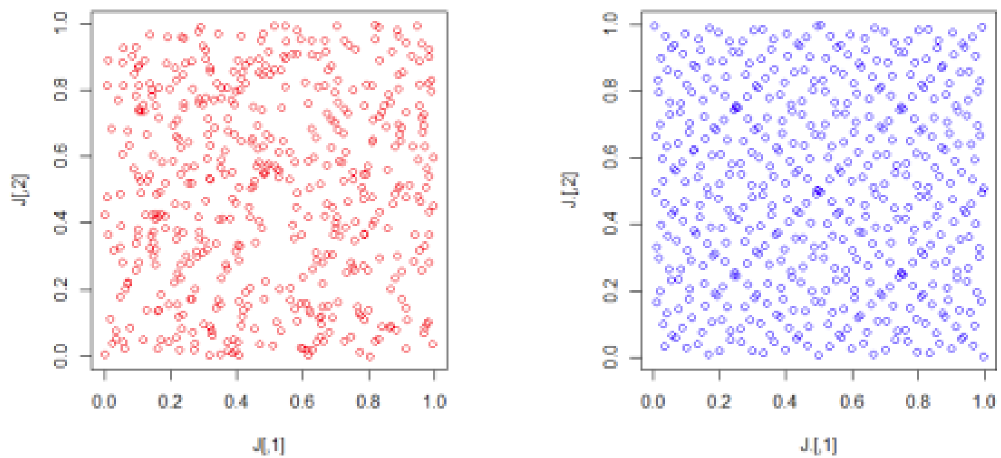




Starting distribution doesn't matter

Sampling

- **Different sampling techniques, e.g. LHS** (Linderoth, Shapiro, Wright, 2006)



Random sample and Sobol's sequence

SAA

Remarks:

- **The complexity of the of the problem depends on the variance of $Q(x,\xi)$, not the number of scenarios**
- **In SAA, following conditions ensure good approximation**

(i) it is possible to generate a sample of realizations of the random vector ξ,

(ii) for moderate values of the sample size it is possible to efficiently solve the obtained SAA problem,

(iii) the true problem has relatively complete recourse,

(iv) variability of the second-stage (optimal value) function is not "too large".

(Shapiro,Pilpott,2007)

Convergence of the SAA

- **Let us denote**

$$\boldsymbol{v}^* = \min_{x \epsilon \boldsymbol{S}} \boldsymbol{g}(\boldsymbol{x}) \qquad \widehat{\boldsymbol{v}}_{\boldsymbol{N}} = \min_{x \epsilon \boldsymbol{S}} \widehat{\boldsymbol{g}}_{\boldsymbol{N}}(\boldsymbol{x})$$

- **By law of large numbers we can show that**

$$\max_{x \epsilon \boldsymbol{S}} |\boldsymbol{g}(\boldsymbol{x}) - \widehat{\boldsymbol{g}}_{\boldsymbol{N}}(\boldsymbol{x})| \rightarrow \mathbf{0} \textbf{ as } \mathrm{N} \rightarrow \infty$$

and thus $\widehat{\boldsymbol{v}}_{\boldsymbol{N}} \rightarrow \boldsymbol{v}^*$ as $\mathrm{N} \rightarrow \infty$ with probability of 1.

(Kleywegt, Shapiro, Homem-de-Mello, 2002)

Convergence of SAA

- **Previously has been shown that SAA works, but how fast?**
- **It happens exponentially fast**

$$\lim_{N\to\infty} \frac{1}{N} \log\left[1 - P(\hat{x}_N = x^*)\right] \;=\; -\gamma^*$$

, where γ^* is constant and > 0

- **Proof in** (Kleywegt, Shapiro, Homem-de-Mello, 2002, p.482-486) (Linderoth, Shapiro, Wright, 2006, p.219-220)

Condition number

- **How sensitive function is subject to its input values**
- **Ill or well conditioned problem**
- **Singleton optimal correspond to well conditioning**
- **Describes sample size needed to obtain $\hat{x}_N = x^*$**
- **Impractical, because you need to obtain x^***

(Kleywegt, Shapiro, Homem-de-Mello, 2002, p.486)

Identifying optimal solutions

Optimality gap

- **To evaluate SAA performance we need new metrics**
- **Given feasible point $\hat{x}$ we clearly heave that $g(\hat{x}) \geq v^*$**
- **To study further qualities we define true optimality gap as**

$$gap(\hat{x}) := g(\hat{x}) - v^*$$

(Shapiro,Pilpott,2007)

Lower Bound

- **It is easily shown that $\mathbb{E}\left[\hat{v}_N\right] \leq v^*$** (Wai-Kei, Morton, Wood, 1999)
- **$\mathbb{E}\left[\hat{v}_N\right]$ can be estimated by solving SAA problems M times, obtaining estimator**

$$\bar{v}_{N,M} := \frac{1}{M}\sum_{j=1}^{M} \hat{v}_N^j$$

(Shapiro,Pilpott,2007)

since $\hat{v}_N{}^1, \ldots, \hat{v}_N{}^M$ are independent, variance of $\hat{v}_{N,M}$

$$\hat{\sigma}^2_{N,M} := \frac{1}{M(M-1)}\sum_{j=1}^{M}\left(\hat{v}_N^j - \bar{v}_{N,M}\right)^2$$

Lower Bound

- **And then we get the lower bound estimate**

$$L_{N,M} := \hat{v}_{N,M} \pm t_{\alpha,\nu}\, \hat{\sigma}_{N,M}$$

where $t_{\alpha,\nu}$ is Student's t-distribution with α-critical value and $\nu = M - 1$

- **Usually it is sufficient to substitute t-distribution with normal distribution with preferable confidence interval**

(Shapiro,Pilpott,2007)

Upper bound

- **Opposite of lower bound, it is again easily shown that for any feasible $\hat{x}$,**

$$\boldsymbol{g(\hat{x}) \geq v^*}$$

- **By choosing $\hat{x}$ near the optimal, we can estimate v^***
- **Starting with N'-sized SAA problem, $g(\hat{x})$ is estimated by**

$$\boldsymbol{\hat{g}_{N'}(\hat{x}) = c^T\hat{x} + \hat{q}_{N'}(\hat{x})}$$

(Shapiro,Pilpott,2007)

Upper bound

- Sample variance of $\widehat{q}_{N'}(\widehat{x})$ is calculated similar to $L_{N,M}$

$$\hat{\sigma}^2_{N'}(\hat{x}) := \frac{1}{N'(N'-1)} \sum_{j=1}^{N'} \left[Q(\hat{x}, \xi^j) - \hat{q}_{N'}(\hat{x}) \right]^2$$

and then upper bound is obtained

$$U_{N'}(\widehat{x}) \coloneqq \widehat{g}_{N'}(\widehat{x}) \pm z_\alpha \, \widehat{\sigma}_{N'}(\widehat{x})$$

- z_α is critical value of normal distribution with α confidence level
- This is and earlier result are proven by CLT (Shapiro,Pilpott,2007)

Optimality gap

- **With lower and upper bound estimates we can construct estimator for the true optimality gap**

$$\widehat{gap}\,(\hat{x}) := U_{N'}(\hat{x}) - L_{N,M}$$

- **By the law of large numbers this tend to the true optimality gap with probability 1**

(Shapiro,Pilpott,2007)

How to choose N and M?

- **Straight forward answer is the bigger the better**
- **“If the computational complexity of solving the SAA problem increases faster than linearly in the sample size N, it may be more efficient to choose a smaller sample size N and to generate and solve several SAA problems”**
- **Increasing only M, leads to Bernoulli style problem, where one needs to asses probability for answer being “true” optimal, which is problem specific**

(Kleywegt, Shapiro, Homem-de-Mello, 2002)

Increase M or N?

- **Suppose that M replications with sample size N have been performed so far.**

- **When $\frac{1}{M+1}$ becomes sufficiently small, additional SAA replications with the same sample size are not likely to be worth the effort, and either the sample size N should be increased or the procedure should be stopped**

(Kleywegt, Shapiro, Homem-de-Mello, 2002)

Performance and examples

Example

Table 2 Test Problem Dimensions

Name	Application	Scenarios	First-Stage Size	Second-Stage Size
20term	Vehicle Assignment	1.1×10^{12}	(3, 64)	(124, 764)
gbd	Aircraft Allocation	6.5×10^{5}	(4, 17)	(5, 10)
LandS	Electricity Planning	10^{6}	(2, 4)	(7, 12)
ssn	Telecom Network Design	10^{70}	(1, 89)	(175, 706)
storm	Cargo Flight Scheduling	6×10^{81}	(185, 121)	(528, 1259)

(Linderoth, Shapiro, Wright, 2006)

Table 3 Lower and Upper Bound Estimates for v^*, Monte Carlo Sampling

Problem	N	$E\hat{v}_N$ (95% conf. int.)	Best $f(\hat{x}_N^j)$ (95% conf. int.)
20term	50	253361.33 ± 944.06	254317.96 ± 19.89
20term	100	254024.89 ± 800.88	254304.54 ± 21.20
20term	500	254324.33 ± 194.51	254320.39 ± 27.36
20term	1000	254307.22 ± 234.04	254333.83 ± 18.85
20term	5000	254340.78 ± 85.99	254341.36 ± 20.32
gbd	50	1678.62 ± 66.73	1655.86 ± 1.34
gbd	100	1595.24 ± 42.41	1656.35 ± 1.19
gbd	500	1649.66 ± 13.60	1654.90 ± 1.46
gbd	1000	1653.50 ± 12.32	1655.70 ± 1.49
gbd	5000	1653.13 ± 4.37	1656.40 ± 1.31
LandS	50	227.19 ± 4.03	225.71 ± 0.12
LandS	100	226.39 ± 3.99	225.55 ± 0.12
LandS	500	226.02 ± 1.43	225.61 ± 0.12
LandS	1000	225.96 ± 0.76	225.70 ± 0.13
LandS	5000	225.72 ± 0.52	225.70 ± 0.12
ssn	50	4.11 ± 1.23	12.68 ± 0.05
ssn	100	7.66 ± 1.31	11.20 ± 0.05
ssn	500	8.54 ± 0.34	10.28 ± 0.04
ssn	1000	9.31 ± 0.23	10.09 ± 0.03
ssn	5000	9.98 ± 0.21	9.86 ± 0.05
storm	50	15506271.7 ± 22043.4	15499092.17 ± 845.26
storm	100	15482549.9 ± 19213.8	15499056.00 ± 623.30
storm	500	15498139.8 ± 4152.8	15498468.02 ± 684.65
storm	1000	15500879.4 ± 4847.1	15498893.02 ± 695.90
storm	5000	15498121.3 ± 1879.0	15498646.89 ± 696.08

Table 4 Lower and Upper Bound Estimates for v^*, Latin Hypercube Sampling

Problem	N	$E\hat{v}_N$ (95% conf. int.)	Best $f(\hat{x}_N^j)$ (95% conf. int.)
20term	50	254307.57 ± 371.80	254328.69 ± 4.73
20term	100	254387.00 ± 252.13	254312.50 ± 5.77
20term	500	254296.43 ± 117.95	254315.82 ± 4.59
20term	1000	254294.00 ± 95.22	254310.33 ± 5.23
20term	5000	254298.57 ± 38.74	254311.55 ± 5.56
gbd	50	1644.21 ± 10.71	1655.628 ± 0.00
gbd	100	1655.62 ± 0.00	1655.628 ± 0.00
gbd	500	1655.62 ± 0.00	1655.628 ± 0.00
gbd	1000	1655.62 ± 0.00	1655.628 ± 0.00
gbd	5000	1655.62 ± 0.00	1655.628 ± 0.00
LandS	50	222.59 ± 2.75	225.647 ± 0.004
LandS	100	225.57 ± 0.16	225.630 ± 0.004
LandS	500	225.65 ± 0.05	225.628 ± 0.004
LandS	1000	225.64 ± 0.03	225.633 ± 0.005
LandS	5000	225.62 ± 0.02	225.624 ± 0.005
ssn	50	10.10 ± 0.81	11.380 ± 0.023
ssn	100	8.90 ± 0.36	10.542 ± 0.021
ssn	500	9.87 ± 0.22	10.069 ± 0.026
ssn	1000	9.83 ± 0.29	9.996 ± 0.025
ssn	5000	9.84 ± 0.10	9.913 ± 0.022
storm	50	15497683.7 ± 1078.8	15498722.14 ± 17.97
storm	100	15499255.3 ± 1011.7	15498739.40 ± 18.34
storm	500	15498661.4 ± 280.8	15498733.67 ± 17.71
storm	1000	15498598.1 ± 148.5	15498735.73 ± 19.93
storm	5000	15498657.8 ± 73.9	15498739.41 ± 19.11

(Linderoth, Shapiro, Wright, 2006)

Resource allocation problem

TABLE 4.1

Condition numbers κ, optimal values v^*, and values $g(\bar{x})$ of optimal solutions $\bar{x}$ of expected value problems $\max_x G(x, E[W])$, for instances presented.

Instance	Condition number κ	Optimal value v^*	Expected value $g(\bar{x})$
10D	107000	42.7	26.2
10R	410	46.3	28.2
20D	954000	96.5	75.9
20R	233	130.3	109.0

Table 4.1 shows the condition numbers, the optimal values v^*, and the values $g(\bar{x})$ of the optimal solutions $\bar{x}$ of the associated expected value problems $\max_x G(x, E[W])$ for the four instances.

(Kleywegt, Shapiro, Homem-de-Mello, 2002)

Resource allocation problem

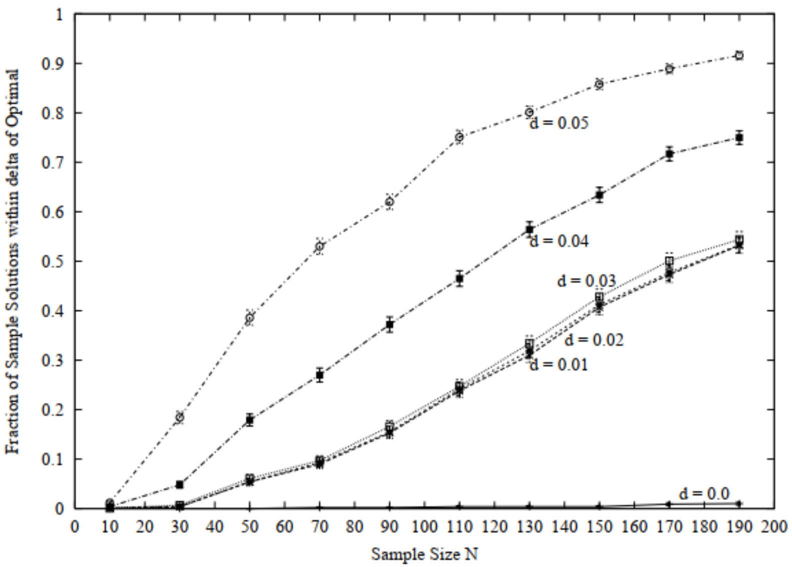


FIG. 4.1. *Probability of* SAA *optimal solution* $\hat{x}_N$ *having objective value* $g(\hat{x}_N)$ *within relative tolerance* d *of the optimal value* v^*, $\hat{P}[v^* - g(\hat{x}_N) \leq d\,v^*]$, *as a function of sample size* N *for different values of* d, *for instance* 20D.

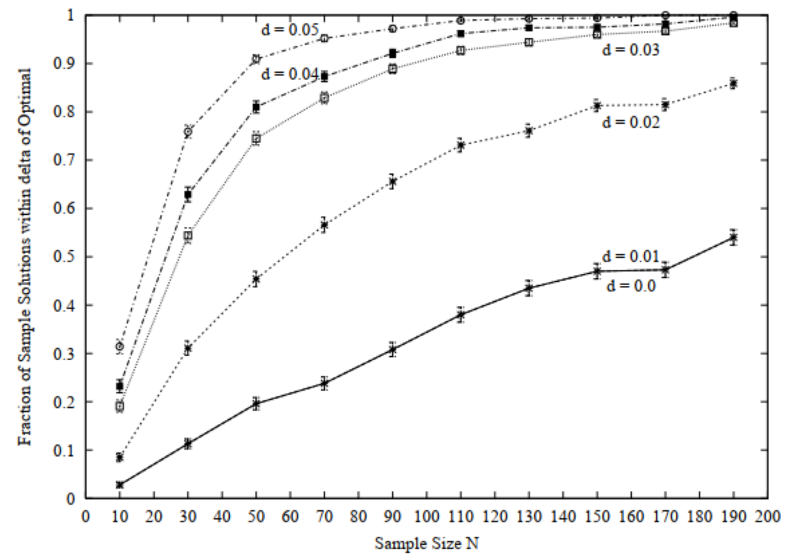


FIG. 4.2. *Probability of* SAA *optimal solution* $\hat{x}_N$ *having objective value* $g(\hat{x}_N)$ *within relative tolerance* d *of the optimal value* v^*, $\hat{P}[v^* - g(\hat{x}_N) \leq d\,v^*]$, *as a function of sample size* N *for different values of* d, *for instance* 20R.

(Kleywegt, Shapiro, Homem-de-Mello, 2002)

Conclusion

- **SAA is simple and proven way to simplify stochastic programming**
- **Sample sizes needed for good solution can be very low depending on the problem**
- **Ill conditioned problems can be hard to solve with SAA**
- **Methods for easier solution assessment would be greatly welcome**

References

1. Hanasusanto, Grani A., Daniel Kuhn, and Wolfram Wiesemann. "A comment on “computational complexity of stochastic programming problems”." *Mathematical Programming* 159.1 (2016): 557-569.

2. Kleywegt, Anton J., Alexander Shapiro, and Tito Homem-de-Mello. "The sample average approximation method for stochastic discrete optimization." *SIAM Journal on Optimization* 12.2 (2002): 479-502.

3. Linderoth, Jeff, Alexander Shapiro, and Stephen Wright. "The empirical behavior of sampling methods for stochastic programming." Annals of Operations Research 142.1 (2006): 215-241.

4. Mak, Wai-Kei, David P. Morton, and R. Kevin Wood. "Monte Carlo bounding techniques for determining solution quality in stochastic programs." *Operations research letters* 24.1-2 (1999): 47-56.

5. Shapiro, Alexander, and Andy Philpott. "A tutorial on stochastic programming." Manuscript. Available at www2. isye. gatech. edu/ashapiro/publications. html 17 (2007).