Non-parametric confidence-based cost estimation

Citation for published version:
Rossi, R 2016, Non-parametric confidence-based cost estimation.

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Other version

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Consider a stochastic constraint optimisation problem $P$ [3], which without loss of generality in what follows will be formulated as a profit maximisation problem. Let $P_{lb}$ and $P_{ub}$ be two sampled stochastic constraint optimisation problems [2] obtained from $P$ such that the optimal solution to $P_{lb}$ underestimates the optimal solution to $P$ with probability $\alpha$ and the optimal solution to $P_{ub}$ overestimates the optimal solution to $P$ with probability $\alpha$. These sampled stochastic constraint optimisation problems can be obtained by using the notion of $(\alpha, \vartheta)$-solution [2].

Let $\omega_{i, lb}$ for $i = 1, \ldots, M$ be the finite-time discrete stochastic process representing the objective values obtained by repeatedly solving $P_{lb}$ $M$ times.

Let $\omega_{i, ub}$ for $i = 1, \ldots, M$ be the finite-time discrete stochastic process representing the objective values obtained by repeatedly solving $P_{ub}$ $M$ times.

Although we do not know the exact distribution of $\omega_{i, lb}$ and $\omega_{i, ub}$, we know that these stochastic processes are stationary. In addition we know that $\omega_{i, lb}$ will underestimate the optimal solution of $P$ with probability $\alpha$ and that $\omega_{i, lb}$ will overestimate the optimal solution of $P$ with probability $\alpha$.

We run the stochastic process $\omega_{i, lb}$ for $i = 1, \ldots, M$ and store the optimal profit obtained for each of these instances into an array $K_{lb}$ sorted in ascending order; we also run the stochastic process $\omega_{i, ub}$ for $i = 1, \ldots, M$ and store the optimal profit obtained for each of these instances into an array $K_{ub}$ sorted in ascending order.

Let $\text{bin}^{-1}(M, \alpha)$ be the inverse cumulative distribution of a binomial distribution with $M$ trials and a success probability $\alpha$; let $k_{lb}$ be the $(1 - \alpha)/2$-quantile of this distribution; finally, let $k_{ub}$ be the $1 - (1 - \alpha)/2$-quantile of $\text{bin}^{-1}(M, 1 - \alpha)$. With confidence $\alpha$ element at position $k_{lb}$ of $K_{lb}$ is a lower bound and element at position $k_{ub} + 1$ of $K_{ub}$ is an upper bound to the true optimal cost.\footnote{Elements of $K_{i}$ are indexed as follows: $1, \ldots, |K_{i}|$. Note that in statistics the $k^{th}$-smallest value of a statistical sample is known as $k^{th}$ order statistic [1].}

1 Elements of $K_{i}$ are indexed as follows: $1, \ldots, |K_{i}|$. Note that in statistics the $k^{th}$-smallest value of a statistical sample is known as $k^{th}$ order statistic [1].
Example

Assume that the value of the optimal solution to $\mathcal{P}$ is $\mu = 30$; $\sigma = 5$; $G^{-1}$ denotes the inverse cumulative distribution function of a standard normally distributed random variable; $\omega_{\text{lb}}^{i}$ is normally distributed with mean $\mu_{\text{lb}} = \mu + \sigma G^{-1}(1 - \alpha)$; $\omega_{\text{ub}}^{i}$ is normally distributed with mean $\mu_{\text{ub}} = \mu + \sigma G^{-1}(\alpha)$; $M = 20$. If we fix $M = 20$ and $\alpha = 0.9$, it follows that $k_{\text{lb}} = 16$ and $k_{\text{ub}} = 4$. Therefore element 16 of $K_{\text{lb}}$ is a lower bound for $\mu$ and element 5 of $K_{\text{ub}}$ is an upper bound for $\mu$ with probability $\alpha$. We replicated the process 10000 times and obtained the distributions shown in Fig. 1 and Fig. 2 for the $k_{\text{lb}}$ order statistics of $K_{\text{lb}}$ and the $k_{\text{ub}}$ order statistics of $K_{\text{ub}}$, respectively. The confidence interval obtained for $\mu$, defined by the lower and the upper bound obtained in each run as illustrated, covers the true value of $\mu$ (i.e. 30) with frequency $0.9154 \geq \alpha$. In Fig. 3 and Fig. 4 we demonstrate how the distribution of the optimality gap varies when $M$ takes value 20 or 100.
References


Appendix: Mathematica code

```mathematica
kLBArray={}; kUBArray={};
M=20; \[Mu]=30; \[Sigma]=5; \[Alpha]=0.9;
\[Mu]LB=InverseCDF[NormalDistribution[\[Mu],\[Sigma]],(1-\[Alpha])];
\[Mu]UB=InverseCDF[NormalDistribution[\[Mu],\[Sigma]],\[Alpha]];
counter=0; R=10000;
For[x=1,x<=R,x++,
dLB=NormalDistribution[\[Mu]LB,\[Sigma]];
dUB=NormalDistribution[\[Mu]UB,\[Sigma]];
sLB=RandomReal[dLB,M];
sUB=RandomReal[dUB,M];
sLBSorted=Sort[sLB];
sUBSorted=Sort[sUB];
lb=InverseCDF[BinomialDistribution[M,0.9],(1-\[Alpha])/2];
ub=InverseCDF[BinomialDistribution[M,0.1],1-(1-\[Alpha])/2];
kLB=sLBSorted[[lb]];
kUB=sUBSorted[[ub+1]];
kLBArray=Append[kLBArray,kLB];
kUBArray=Append[kUBArray,kUB];
If[kLB<=\[Mu] && kUB>=\[Mu],counter++];
];
N[counter/R]
Histogram[kLBArray]
Histogram[kUBArray]
Histogram[kUBArray-kLBArray]
Mean[kUBArray-kLBArray]
```