BAGGING ESTIMATION OF AVAILABILITY IN PUBLIC CLOUD STORAGE

MAURIZIO NALDI

UKSIM 2018 – CAMBRIDGE

AVAILABILITY CLAIMS

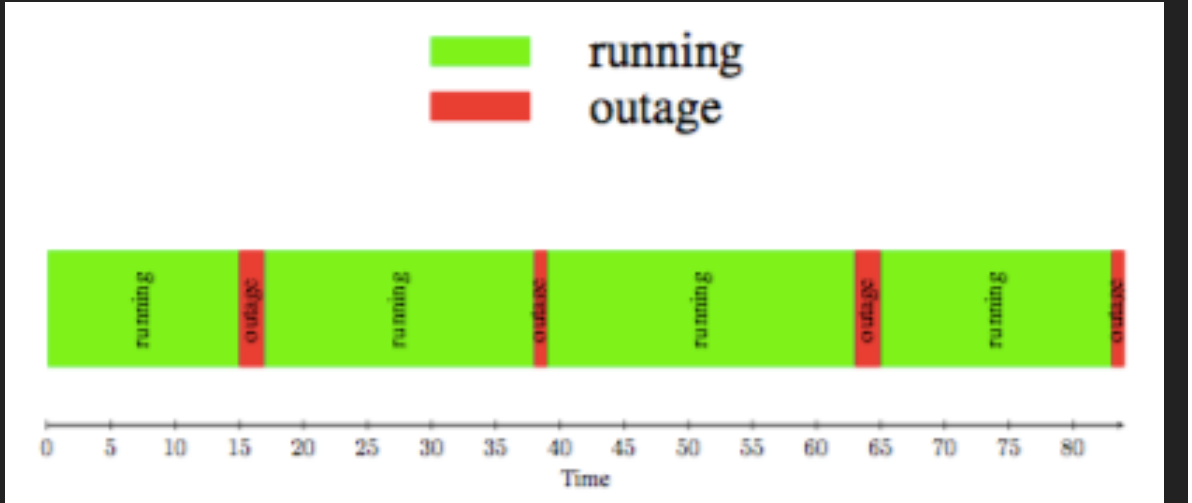
- Availability is typically included in SLAs
- Cloud providers boast of unrealistic availability performances
- Most claim to be able to provide 100% availability

Cloud provider	Availability SLO [%]
Amazon Web Service	99.95
AT&T Synaptic	99.9
CloudSigma	100
ElasticHosts	100
FlexiScale	100
GoGrid	100
JoyentCloud	100
layeredtech	100
Locaweb	99.9
Opsource	100
Rackspace	100
ReliaCloud	100
RSAWEB Cloud servers	ND
SliceHost	ND
Storm On demand	100
Terremark vCloud express	100
VPSNET	100

AVAILABILITY MEASUREMENTS

- Availability is typically measured as the proportion of overall ON time
- This definition provides a single figure for a specific observation interval and does not allow to draw any conclusion as to its statistical accuracy

$$A_0 = \frac{\sum_{i=1}^{n} S_i}{\sum_{i=1}^{n} S_i + \sum_{i=1}^{n} D_i}$$



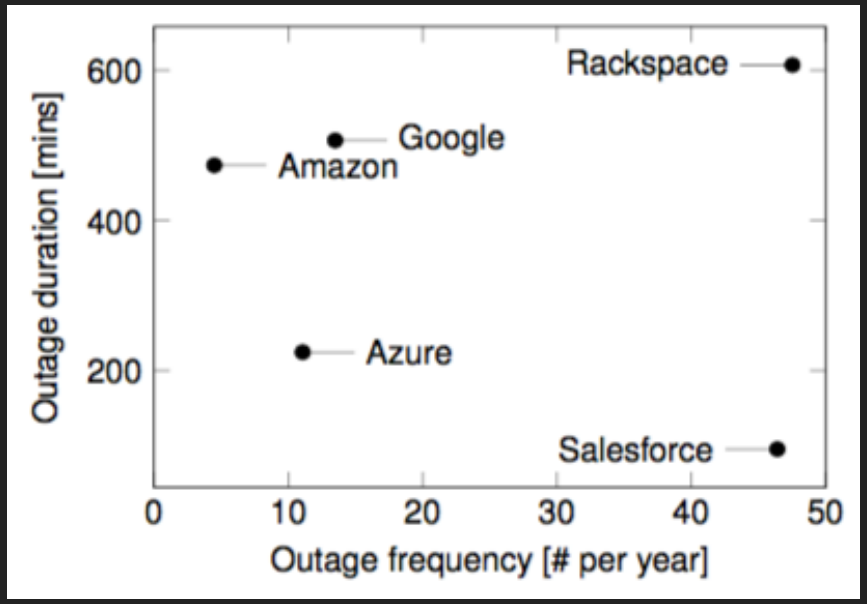
A BOOTSTRAP AVERAGING TECHNIQUE

- We observe a single sequence of operating periods and outages
- From this sequence we generate a set of B sequences by sampling with replacement from the observed durations of operating periods and outages
- Compute the availability (or any other performance metric) for each of these sequences
- Compute the relevant statistics (e.g., average, standard deviation and confidence intervals) from the set of bootstrap-derived availability values

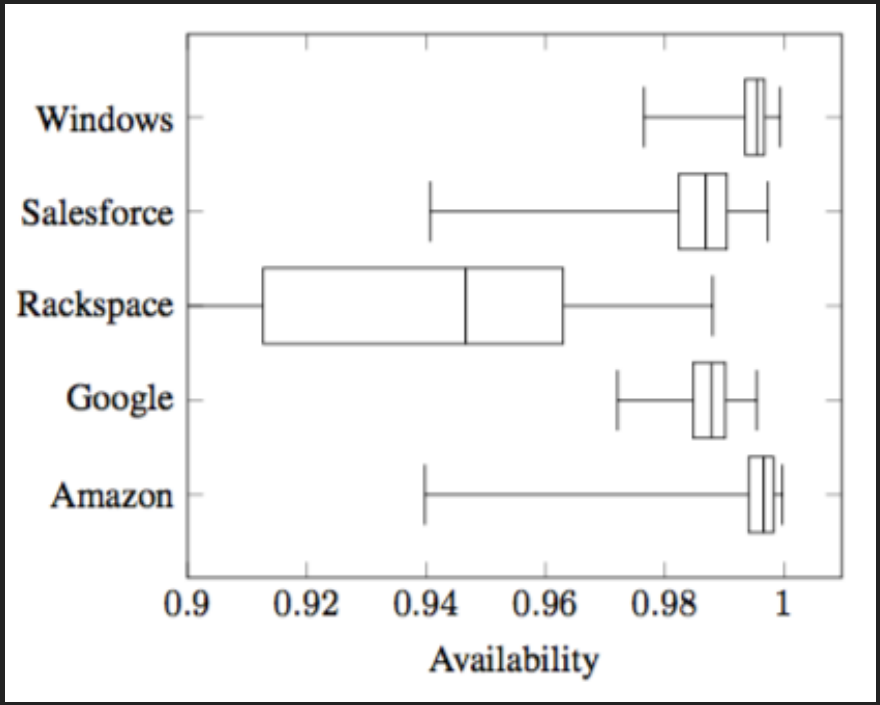
OUR DATASET

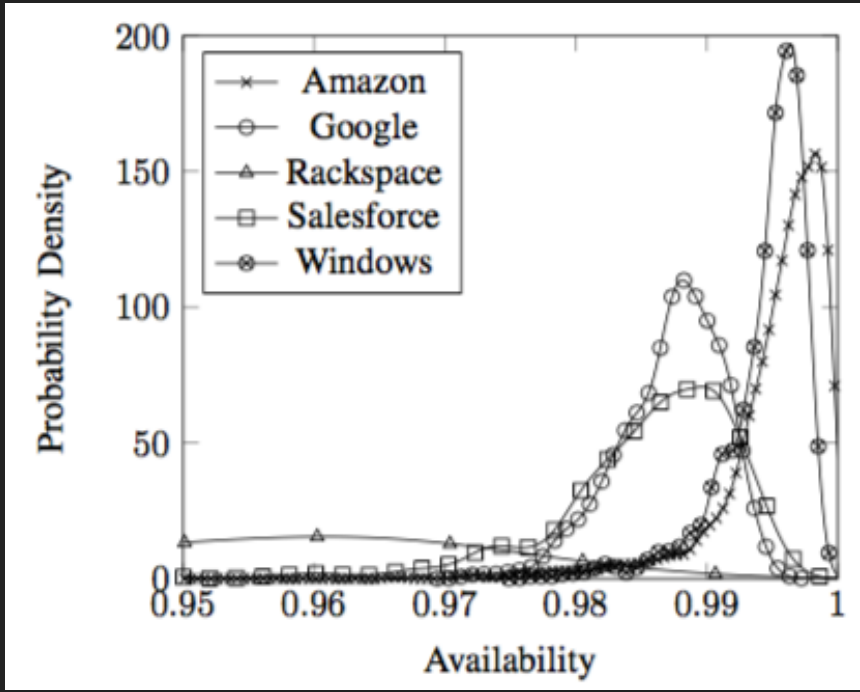
- We used data from Cloutage (cloutage.org) and the International Working Group on Cloud Computing Resiliency, which gather data from customers' observations
- The data collection considered 5 major cloud providers

Provider	Outages per year	Inter-outage times [days]
Google	13.48	27.53
Amazon	4.48	85.6
Rackspace	47.53	7.78
Salesforce	46.4	8.56
Windows Azure	11.06	36.67



AVAILABILITY: THE DISTRIBUTION





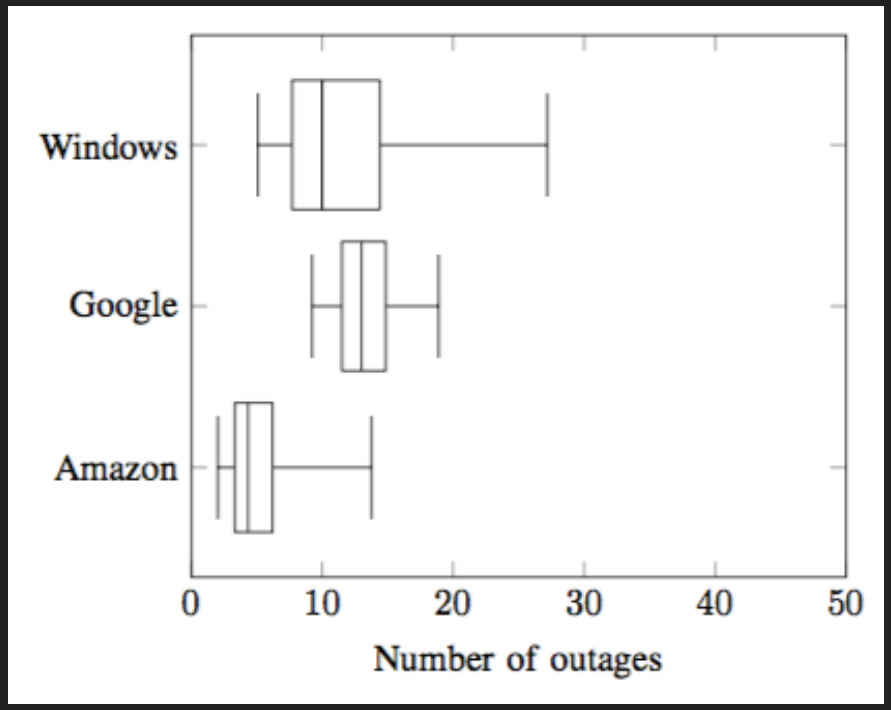
AVAILABILITY: CONFIDENCE INTERVALS

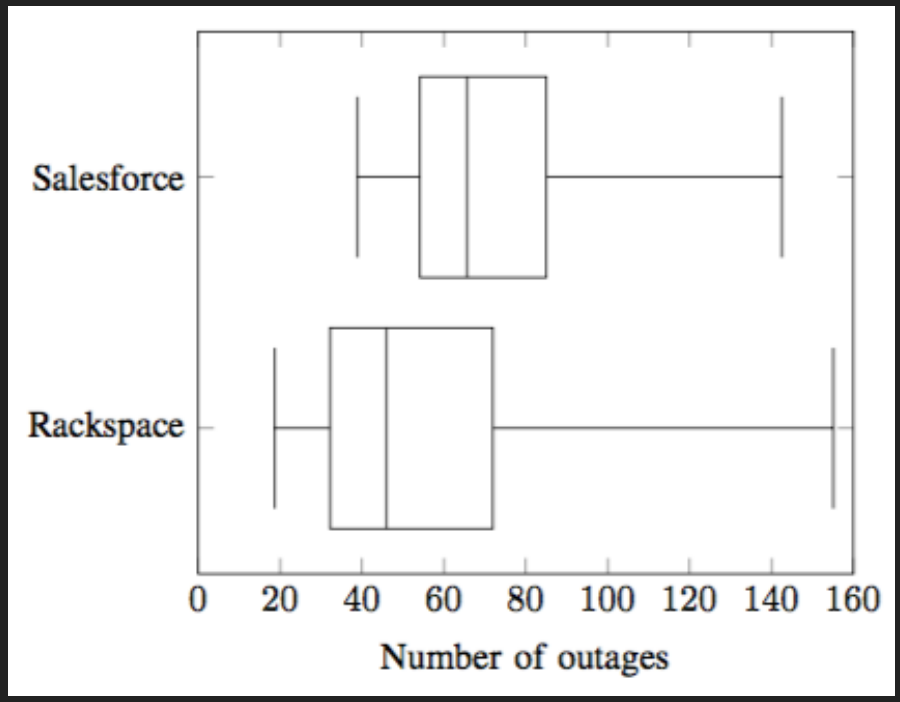
- A simple method to check for compliance with SLAs
- Compute confidence intervals and check if committed values lies inside

Provider	Lower limit	Upper limit
Windows	0.9855082	0.9982058
Salesforce	0.9674701	0.9945194
Rackspace	0.7981062	0.9799475
Google	0.9784518	0.9934143
Amazon	0.9820859	0.9993779

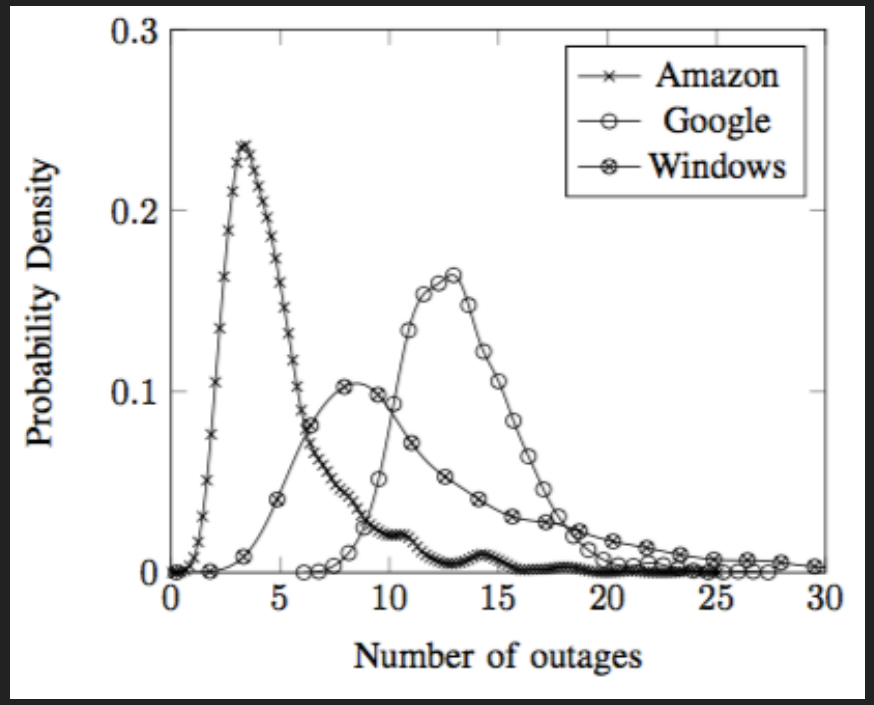
- Confidence intervals include 0.999 for just one operator

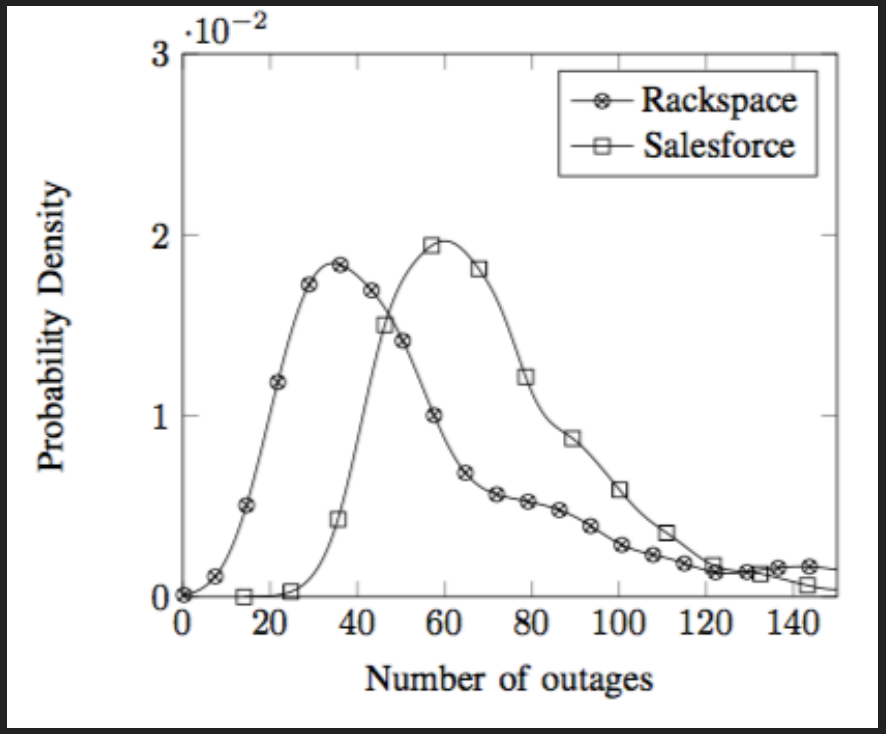
NUMBER OF OUTAGES: THE BOXPLOT





NUMBER OF OUTAGES: THE DISTRIBUTION





NUMBER OF OUTAGES: CONFIDENCE INTERVALS

Provider	Lower limit	Upper limit
Windows	5.075332	27.175377
Salesforce	38.81662	142.48247
Rackspace	18.64925	155.06943
Google	9.242874	18.900919
Amazon	2.024132	13.747629

CONCLUSIONS

- Single sample measurements of the availability of a cloud storage service do not possess statistical accuracy
- A bootstrap-based technique has been proposed to derive statistically sound estimates from a single measurement
- The technique has been tested on a dataset concerning 5 major cloud providers
- Its use allows to compute confidence intervals to be employed in SLA compliance assessment