Design and Inference of a Call Center with an Answering Machine (IVR)

Polina Khudyakov

Joint work with Prof. Avishai Mandelbaum, Prof. Malka Gorfine and Prof. Paul Feigin

June 29, 2010

Outline

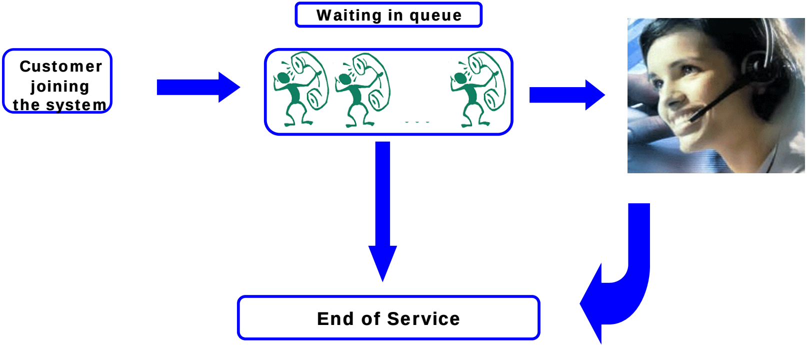
- Design of a Call Center with an IVR
- Customer Patience Analysis
- Service Time Analysis

Background

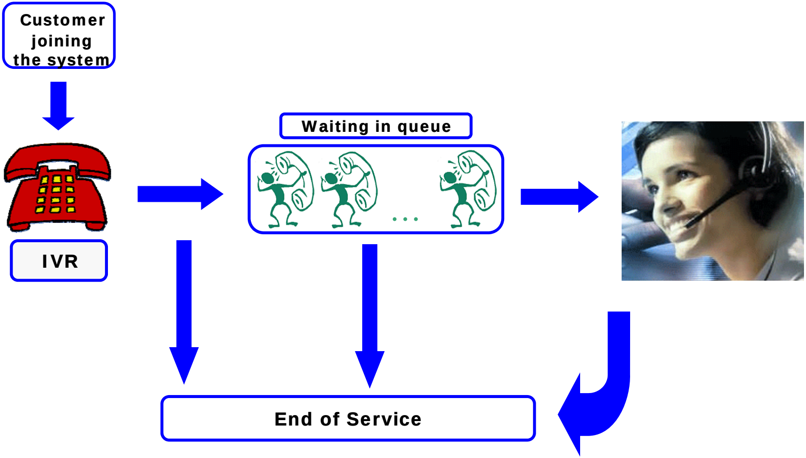
- Halfin and Whitt (1981) (M/M/S)
- Massey and Wallace (2004) (M/M/S/N)
- Garnet, Mandelbaum and Reiman (2002) (M/M/S+M)
- Srinivasan, Talim and Wang (2002) (Call center with an IVR)

Customer interaction with a call center (Model I)

M/M/S/N+M



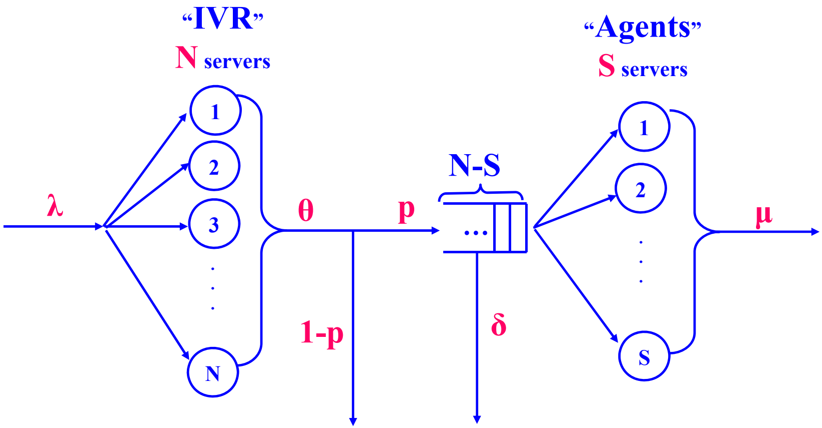
Customer interaction with a call center (Model II)



Model description

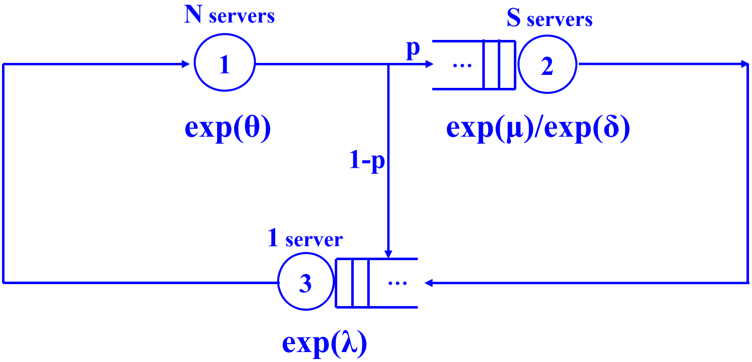
- N - number of trunk lines
- $Poisson(\lambda)$ - arrival process
- p - probability to request agent's service
- S - number of agents
- $exp(\theta)$ - IVR service time
- $exp(\mu)$ - agent's service time
- $exp(\delta)$ - customer patience time

Schematic model



Closed Jackson Network

- Brandt, Brandt, Spahl and Weber (1997)
- Srinivasan, Talim and Wang (2004)



Stationary Probabilities

$$\pi(i,j) = \begin{cases} \pi_0 \frac{1}{i!} \left(\frac{\lambda}{\theta}\right)^i \frac{1}{j!} \left(\frac{\lambda p}{\mu}\right)^j, & j \le S,\ 0 \le i+j \le N; \\ \pi_0 \frac{1}{i!} \left(\frac{\lambda}{\theta}\right)^i \frac{1}{S!} \left(\frac{\lambda p}{\mu}\right)^S \frac{(\lambda p)^{j-S}}{\prod_{k=1}^{j}(S\mu + k\delta)} & j \ge S,\ 0 \le i+j \le N; \\ 0 & \text{otherwise,} \end{cases}$$

where

$$\pi_0 = \Big[\sum_{i=0}^{N-S} \sum_{j=S}^{N-i} \frac{1}{i!} \Big(\frac{\lambda}{\theta}\Big)^i \frac{1}{S!} \Big(\frac{\lambda p}{\mu}\Big)^S \frac{\Big(\lambda p\Big)^{j-S}}{\prod_{k=1}^{j}(S\mu + k\delta)} + \sum_{i+j \le N, j \le S-1} \frac{1}{i!} \Big(\frac{\lambda}{\theta}\Big)^i \frac{1}{j!} \Big(\frac{\lambda p}{\mu}\Big)^j \Big]^{-1}.$$

Exact Performance Measures

$$P(W>0)=\sum_{i=0}^{N-S}\sum_{j=S}^{N-i}\chi(i+j,j),$$

where $\chi(k,j)$, $0\leq j<k\leq N$, is the probability that the system is in state (k,j), given that a call (among the $k-j$ customers) is about to complete its IVR service:

$$\chi(k,j)=\frac{(k-j)\,\pi(k-j,j)}{\sum\limits_{l=0}^{N}\sum\limits_{m=0}^{l}(l-m)\,\pi(l-m,m)}.$$

$$P(Ab|W>0)=\frac{\sum_{i=0}^{N-S}\sum_{j=S+1}^{N}\pi(i,j)(j-S)\delta}{\sum_{i=0}^{N-S}\sum_{j=S+1}^{N}\pi(i,j)\,(S\mu+(j-S)\delta)},$$

$$E[W|W>0]=\frac{1}{\delta}P(Ab|W>0).$$

The domain for asymptotic analysis: QED

- Model I (M/M/S/N+M)

$$(i) \quad \lim_{\lambda\to\infty} N - S = \eta\sqrt{\frac{\lambda}{\mu}}, \quad 0 < \eta < \infty;$$

$$(ii) \quad \lim_{\lambda\to\infty} S = \frac{\lambda}{\mu} + \beta\sqrt{\frac{\lambda}{\mu}}, \quad -\infty < \beta < \infty.$$

- Model II (Call Center with an IVR)

$$(i) \quad \lim_{\lambda\to\infty} N - S = \frac{\lambda}{\theta} + \eta\sqrt{\frac{\lambda}{\theta}} \quad -\infty < \eta < \infty;$$

$$(ii) \quad \lim_{\lambda\to\infty} S = \frac{\lambda p}{\mu} + \beta\sqrt{\frac{\lambda p}{\mu}}, \quad -\infty < \beta < \infty \, .$$

QED Approximations (Model I)

Theorem:

Let the variables λ, S and N tend to ∞ simultaneously and satisfy QED conditions, where μ is fixed. Then the asymptotic behavior of the system is described in terms of the following performance measures:

- the probability $P(W>0)$ that a served customer waits after the IVR:

$$\lim_{\lambda\to\infty} P(W>0) = \left(1 + \frac{\sqrt{\frac{\delta}{\mu}}\Phi(\beta)\varphi(\beta\sqrt{\frac{\mu}{\delta}})}{\varphi(\beta)\left[\Phi(\eta\sqrt{\frac{\delta}{\mu}}+\beta\sqrt{\frac{\mu}{\delta}})-\Phi(\beta\sqrt{\frac{\mu}{\delta}})\right]}\right)^{-1},$$

- the probability of abandonment, given waiting:

$$\lim_{\lambda\to\infty}\sqrt{S}P(Ab|W>0) = \frac{\sqrt{\frac{\delta}{\mu}}\varphi(\beta\sqrt{\frac{\mu}{\delta}})-\beta\left[\Phi(\eta\sqrt{\frac{\delta}{\mu}}+\beta\sqrt{\frac{\mu}{\delta}})-\Phi(\beta\sqrt{\frac{\mu}{\delta}})\right]}{\Phi(\eta\sqrt{\frac{\delta}{\mu}}+\beta\sqrt{\frac{\mu}{\delta}})-\Phi(\beta\sqrt{\frac{\mu}{\delta}})}.$$

QED Approximations (Model II)

Theorem:

Let the variables λ, S and N tend to ∞ simultaneously and satisfy the QED conditions, where μ, p, θ, δ are fixed. Then the asymptotic behavior of our system, is captured by the following performance measures:

- the probability $P(W>0)$ that a served customer waits after the IVR:

$$\lim_{\lambda\to\infty} P(W>0) = \left(1+\frac{\gamma}{\xi_1-\xi_2}\right)^{-1};$$

- the probability of abandonment, given waiting:

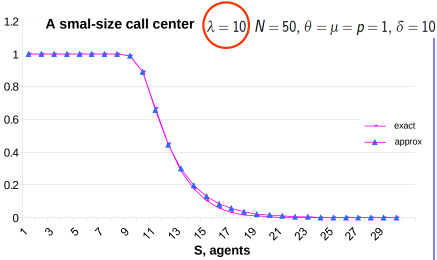
$$\lim_{\lambda\to\infty}\sqrt{S}P(Ab|W>0) = \frac{\sqrt{\frac{\delta}{\mu}}\varphi(\beta\sqrt{\frac{\mu}{\delta}})\Phi(\eta) - \beta\int\limits_{\beta\sqrt{\frac{\mu}{\delta}}}^{\infty}\Phi(\eta+(\beta\sqrt{\frac{\mu}{\delta}}-t)\sqrt{\frac{p\theta}{\mu}})\varphi(t)dt}{\int\limits_{\beta\sqrt{\frac{\mu}{\delta}}}^{\infty}\Phi(\eta+(\beta\sqrt{\frac{\mu}{\delta}}-t)\sqrt{\frac{p\theta}{\mu}})\varphi(t)dt}.$$

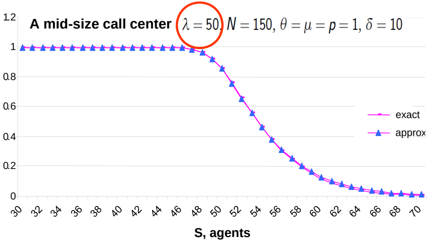
where

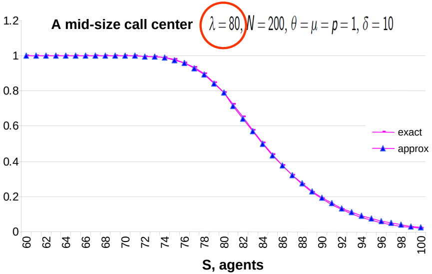
$$\xi_1 = \sqrt{\frac{\mu}{\delta}}\frac{\varphi(\beta)}{\varphi(\beta\sqrt{\frac{\mu}{\delta}})}\int\limits_{-\infty}^{\eta}\Phi\left((\eta-t)\sqrt{\frac{\delta}{p\theta}}+\beta\sqrt{\frac{\mu}{\delta}}\right)\varphi(t)dt,$$

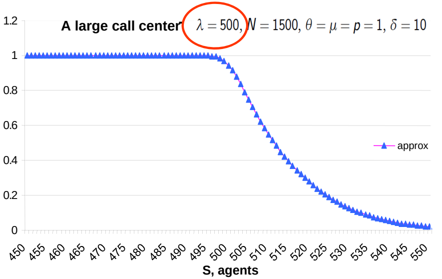
$$\xi_2 = \sqrt{\frac{\mu}{\delta}}\frac{\varphi(\beta)}{\varphi(\beta\sqrt{\frac{\mu}{\delta}})}\Phi(\beta\sqrt{\frac{\mu}{\delta}})\Phi(\eta), \qquad \gamma = \int\limits_{-\infty}^{\beta}\Phi\left(\eta+(\beta-t)\sqrt{\frac{p\theta}{\mu}}\right)\varphi(t)dt.$$

Illustration of the $P(W > 0)$ approximation



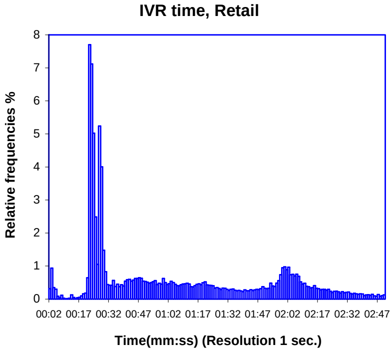




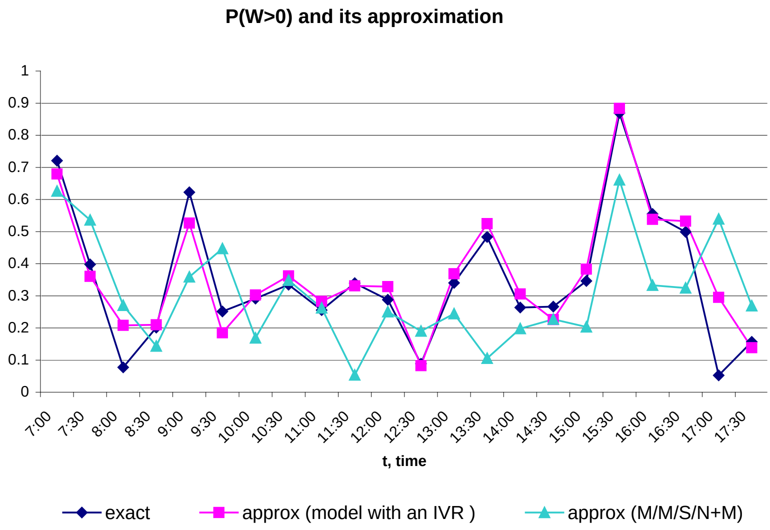


Data Description

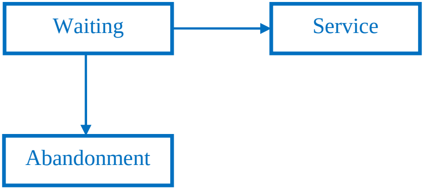
- Service type - Retail (80% of total customers)
- April 12, 2001 (an ordinary week day)
- Observed time period is from 07:00 till 18:00
- About 600 agents per shift



True vs. Approximated Performance Measures



Customer Patience Analysis



- *Customer patience* is the willingness of a customer to wait before being served
- Customers receive the service before they lose their patience
- Censored data
- Survival analysis

Customer Patience Analysis (cont.)

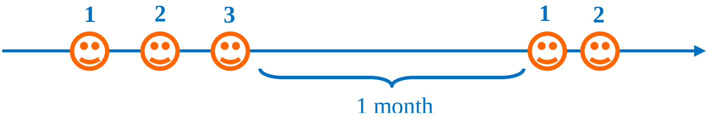
- Does customers' individual features influence on their waiting behavior?
  - observed ⇒ covariates
  - unobserved ⇒ frailty

Customer Patience Analysis (cont.)

- Does customers' individual features influence on their waiting behavior?
  - observed $\Rightarrow$ covariates
  - unobserved $\Rightarrow$ frailty
- How customers' experience with the system affect of their waiting behavior?

Customer Patience Analysis (cont.)

- Does customers' individual features influence on their waiting behavior?
  - observed ⇒ covariates
  - unobserved ⇒ frailty
- How customers' experience with the system affect of their waiting behavior?
- Data selection



The Model

- Conditional hazard function

$$\lambda_{ij}(t|w_i) = \lambda_{0j}(t) w_i e^{\beta^T Z_{ij}} \qquad i = 1, ..., n \quad j = 1, ..., m_i$$

- $\lambda_{0j}(t)$ is an unspecified baseline hazard function of call j
- β is a p-dimensional vector of unknown regression coefficients
- w_i are i.i.d. random variables with known density $f(w) \equiv f(w; \theta)$
- θ is an unknown vector of parameters

Estimation Procedure

- Gorfine, Zucker and Hsu (*Biometrika*, 2006)

$$L = \prod_{i=1}^{n}\prod_{j=1}^{m_i}\{\lambda_{0j}\left(T_{ij}\right)e^{\beta^T Z_{ij}}\}^{\delta_{ij}}\prod_{i=1}^{n}\int_0^{\infty} w^{N_{i\cdot}(\tau)}e^{-wH_{i\cdot}(\tau)}f(w)dw$$

Step 1.

Given the value of $\gamma = (\theta, \beta^T)$ estimate $\left\{\Lambda_{0j}(\tau)\right\}_{j=1}^{m}$ by using

$$\Delta\hat{\Lambda}_{0j}(\tau_{jk}) = \left(\sum_{i=1}^{n} dN_{ij}(\tau_{jk})\right) / \left(\sum_{i=1}^{n} h_{ij}(\tau_{jk})Y_{ij}(\tau_{jk})\right)$$

Step 2.

Given value of $\left\{\hat{\Lambda}_{0j}\right\}_{j=1}^{m}$, estimate γ by using the score equations $U(\gamma, \{\hat{\Lambda}_{0j}\}_{j=1}^{m}) = 0$

Step 3.

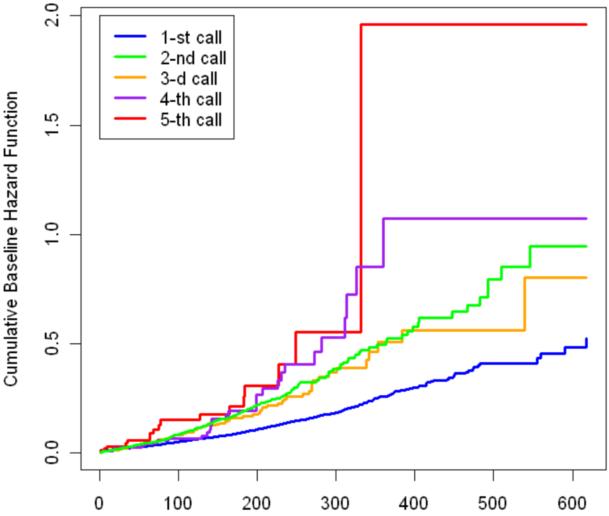
Repeat Steps 1 and 2 until convergence is reached with respect to $\left\{\hat{\Lambda}_{0j}\right\}_{j=1}^{m}$ and $\hat{\gamma}$.

Data Analysis

- The sample consists of $n = 50,000$
- Each customer did at most 5 calls
- Standard errors are based on 150 bootstrap samples
- Estimates of the model parameters:

	$\hat{\theta}$	$\hat{\beta}_1$	$\hat{\beta}_2$
Mean	0.9973	-0.3006	-0.1211
SD	0.1767	0.1046	0.0567

Data Analysis (cont.)



Test for Equality of Baseline Hazard Functions

- $H_0: \qquad \Lambda_{01}(t) = \Lambda_{02}(t) = \Lambda_0(t) \qquad \textit{for all} \quad 0 < t \leq \tau$

- Our test statistic:

$$S_n(\tau) = \frac{1}{\sqrt{n}} \int_0^\tau \hat{W}_n(s) \frac{\bar{Y}_1(s)\bar{Y}_2(s)}{\bar{Y}_1(s) + \bar{Y}_2(s)} \left\{ d\hat{\Lambda}_{01}(s) - d\hat{\Lambda}_{02}(s) \right\}$$

- $\bar{Y}_j(s) = \sum_{i=1}^n \hat{h}_{ij}(s) Y_{ij}(s)$

- $\hat{W}_n(s)$ a nonnegative weight function

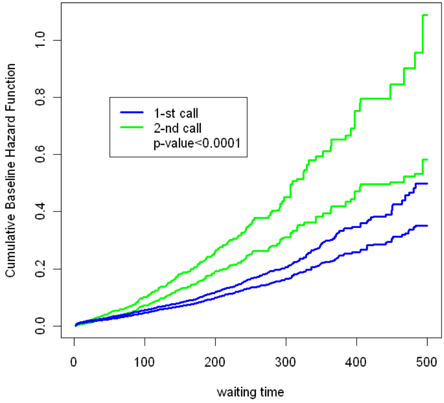
Data analysis

- Results of the 10 tests
- The sample consists from $n = 50,000$ customers:

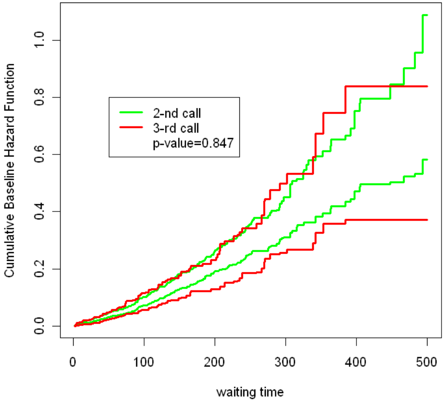
	1 vs. 2	1 vs. 3	1 vs. 4	1 vs. 5	2 vs. 3
$S_n(250)$	-0.464	-0.771	-0.051	-0.048	0.027
$\hat{\sigma}_I(250)$	0.039	0.199	0.018	0.016	0.027
$S_n(250)/\hat{\sigma}_I(250)$	-11.915	-3.871	-2.884	-3.019	1.024
$p-value$	<0.001	<0.001	0.002	0.001	0.847
$FDR \quad p-value$	<0.001	0.003	0.042	0.003	1.000
	2 vs. 4	**2 vs. 5**	**3 vs. 4**	**3 vs. 5**	**4 vs. 5**
$S_n(250)$	0.058	-0.029	-0.014	-0.030	-0.019
$\hat{\sigma}_I(250)$	0.163	0.016	0.016	0.015	0.012
$S_n(250)/\hat{\sigma}_I(250)$	0.355	-1.841	-0.907	-2.051	-1.493
$p-value$	0.639	0.033	0.182	0.020	0.068
$FDR \quad p-value$	1.000	0.096	1.000	0.042	0.422

95% Pointwise Confidence Intervals

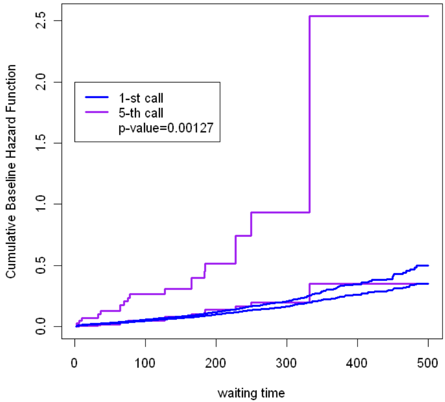
- 1-st call vs. 2-nd call

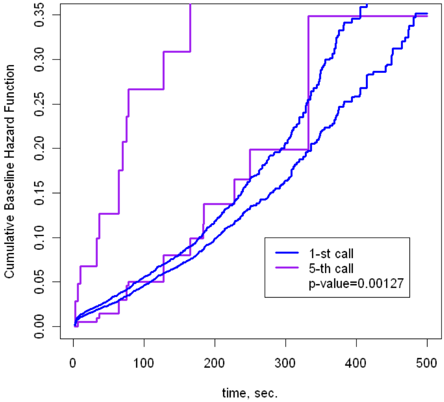


- 2-nd call vs. 3-rd call

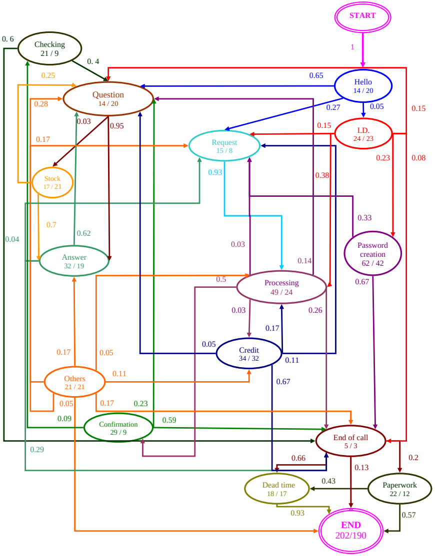


1-st call vs. 5-th call





Phase Type Distribution for Agents' Service Time



Fitting the Agents' Service Time