Can We Improve the Learning Part of the Online Fulfillment Algorithm?

D. Gueorguiev 6/20/23

Notation and quantities

Notation

$A, B, \dots, Z$ – with capital Latin letters we will denote *scalar quantities* which are either essential algorithm parameters or constants which will not change during the algorithm execution; for example, *number of feasible nodes for the current bundle* (scalar constant) will be denoted with N and *inventory for given SKU on given node* (algorithm parameter) will be denoted with I. Graphs will also be denoted with capital Latin letters for historical reasons.

$a, b, \dots, z$ – with small Latin letters we will denote *variable/unknown (integral or not) quantities*, not necessarily scalar. For example, with x we can denote the number of order-lines fulfilled at a given node.

$\alpha, \beta, \dots, \omega$ – with small Greek letters we will denote *variable/unknown (integral or not) quantities*, not necessarily scalar.

$\mathcal{A}, \mathcal{B}, \dots, \mathcal{Z}$ – with capital Script letters we will denote a *set* (ordered or unordered) of quantities of the same type; for example, with $\mathcal{S}$ we will denote the set of SKUs in some bundle of some order

$\mathscr{a}, \mathscr{b}, \dots, \mathscr{z}$ – with small Script letters we will denote attributes, concepts which will be in the form of subscript for parameters and quantities. For example, UCP_j^{daily} denotes daily unit capacity at node j.

$\mathbb{A}, \mathbb{B}, \dots, \mathbb{Z}$ – with double struck Latin capital letters we will denote standard number sets. For example

$\mathbb{C}$ - the set of complex numbers

$\mathbb{N}$ - the set of natural numbers

$\mathbb{R}$ - the set of the real numbers

$\mathbb{Z}$ - the set of integer numbers

Quantities and parameters

$\mathcal{T}$ – ordered set of order and bundle arrival indices

$\mathcal{S}$ – the set of all products

$\mathcal{J}$ – the set of all nodes

$\mathcal{V}$ – the set of all objectives

$\mathcal{K}$ – the set of all *node kinds*

$\mathcal{M}_t$– optimal fulfillment options (matching) set

O_t – order received at time t

$INV_{s,j}^t$ – inventory for product s at node j for order bundle at time t

CAP_j^t – capacity at node j for order bundle at time t

OV_v^t – objective value for objective v of the fulfillment decision for order bundle at time t

OPT_v^t – objective value at optimum total for objective v for order bundle at time t

$R_{m^*}^t$ – reward for the optimal fulfillment option of the order bundle at time t

Q_s^t – quantity for product s in the order bundle at time t

$OC_{s,j}^t$ – opportunity cost for product s at node j for the order bundle at time t

$\delta OC_{s,j,\kappa}^t$ – opportunity cost delta for product s at node j and node kind κ for the order bundle at time t

f_u – the opportunity cost update function

AS^t– the algorithm state for order bundle at time t

CRT – cost-reward transformation

SC – shipping cost value

$\overline{SC}$ – average shipping cost

SQ_s– quantity for product s

$SQ_p^{tot} = \sum_{s\in p} SQ_s$ – the total units for the products in the current bundle p

UCP_j^{daily} – daily unit capacity at node j

R – reward for fulfillment option
DV_j^{cap} – capacity dual variable at node j
$DV_{s,j}^{inv}$ – inventory dual variable at node j
$\hat{R}$ – adjusted reward for fulfillment option
$E[R_o]$ – expected reward for fulfillment option o
$E[\hat{R}_o]$ – expected reward for fulfillment option o
$E[R_{m^*}]$ – expected reward for optimal fulfillment option
$E[\hat{R}_{m^*}]$ – expected adjusted reward for optimal fulfillment option

Problem Statement

A fundamental assumption made in the design of the online fulfillment algorithm is that the fulfillment decision of the current order bundle depends only on aggregated state of the environment taken right after the fulfillment decision about the previous order bundle was made.

Everywhere throughout the discussion we assume that all received orders can be ordered in time as a sequence O_1, $O_2, \dots, O_{t-1}, O_t, \dots$ and if we know the time index t of the current order O_t we can always tell the sequence of orders $O_1, O_2, \dots, O_{t-1}$ received before O_t. Further, for multi-bundle orders we will assume that there is a strict order of the bundles within the current order at time t. Thus, for the online fulfillment algorithm the time is represented with the indexes of the received order bundles ordered as a time sequence: $t_1, t_2, \dots, t_{k-1}, t_k, \dots$. We will denote the ordered set of those indices with $\mathcal{T}$.

The algorithmic state of the online fulfillment algorithm is given with:

The current inventory at the nodes at the time t which will be denoted with $INV_{s,j}^t$. The subscript s denotes the SKU (or product and the subscript j denotes the node index for which the inventory quantity of this product is given.

The current capacity at the nodes at the time t which will be denoted with CAP_j^t. The subscript j denotes the node index for which the capacity quantity is specified.

The current values of the objectives at optimization total at time t which will be denoted with OPT_v^t. Here the subscript v denotes the v-th objective. Let us we denote with OV_v^t the objective value at time t for the v-th objective. Then we have the following recurrent relation:

$$OPT_v^t = OPT_v^{t-1} + OV_v^t \qquad (1)$$

The current values of the opportunity cost total at time t which will be denoted with $OC_{s,j}^t$. As before the subscript s denotes the SKU (or product) and the subscript j denotes the node index for the current opportunity cost.
The following recurrent relation is valid for the opportunity cost total:

$$OC_{s,j}^t = OC_{s,j}^{t-1} + f_u\left(R_{m^*}^t, CAP_j^t, INV_{s,j}^t, SQ_s^t\right) \qquad (2)$$

Here $R_{m^*}^t$ denotes the reward for the optimal fulfillment option of the order bundle at time t. Here $m^* \in \mathcal{M}_t$ represents an optimal fulfillment option (matching) from the optimal fulfillment options (matching) set $\mathcal{M}_t$.
SQ_s^t denotes the quantity for product s in the order bundle at time t.
f_u is the *opportunity cost update function* which is linearly proportional with respect to $R_{m^*}^t$ and SQ_s^t, and inversely proportional with respect to CAP_j^t and $INV_{s,j}^t$.

Let us denote the current algorithmic state at time t with AS^t defined as:

$$AS^t = \left\{INV_{s,j}^t, CAP_j^t, OPT_v^t, OC_{s,j}^t; s \in \mathcal{S}, j \in \mathcal{J}, v \in \mathcal{V}\right\} \qquad (3)$$

Here $\mathcal{S}$ is the set of all products, $\mathcal{J}$ is the set of all nodes and $\mathcal{V}$ is the set of all objectives.

The new fulfillment decision is based on finding a fulfillment option which maximizes the adjusted multi-objective reward.

The opportunity cost total at time t in general will have dependency on the *kind of* node represented by the function $\kappa(j) \in \mathcal{K}$ where $\mathcal{K} \equiv \{1, \dots, K\}$: $OC_{s,j,\kappa}^{t}$. If the node does not change its kind from the beginning of time until t then this dependency is only implicit - $OC_{s,j,\kappa}^{t} \equiv OC_{s,j}^{t}$. As stated earlier the total opportunity cost for the order bundle at moment t will be updated using the update function f_u:

$$OC_{s,j,\kappa}^{t} = f_u\left(OC_{s,j,\kappa}^{t-1}, R_{m^*}^{t}, CAP_j^t, INV_{s,j}^t, SQ_s^t\right) \qquad (4)$$

The opportunity cost delta $OC_{s,j,\kappa}^{t}$ should be updated in such way that the total reward will increase in an optimal way from this manipulation. The question which will be discussed here is how to update the opportunity costs such that the update will bring benefits in collecting future rewards.

Example Scenarios

Let us consider few example scenarios starting with the simplest scenarios and progressing gradually to more elaborate scenarios.

Single objective Shipping Cost, two nodes and single product with unit quantity

In the first example we will assume that we have only a single objective – *shipping cost*. Let us assume we have a region $\mathcal{R}$ with $\dim(\mathcal{R}) = 1$ partitioned into $J = 2$ subregions. Each subregion contains a single node $N_i, i = 1{,}2$ as shown on the figure below.

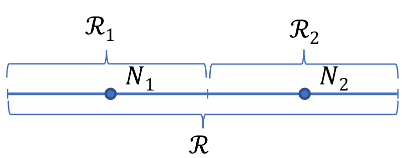


Figure 1: region layout for the scenario with single objective and two nodes

For simplicity of the analysis let us assume that each point on the region $\mathcal{R}$ represent a distinct zip code i.e. there is a continuum of zip codes in $\mathcal{R}$. Another simplifying assumption which we will make is that the shipping cost is proportional to the distance from the order zip code to the fulfilling node.
Here is how we will calculate the reward in this single objective scenario:
$\overline{SC}$ – mean shipping cost for the considered set of nodes given with $\overline{SC} = \frac{\sum_j SC_j}{N}$
CRT – cost-reward transformation:

$$CRT(SC, \overline{SC}, c_1) = \begin{cases} e^{\frac{-(SC-\overline{SC})}{\overline{SC}}}, & \overline{SC} \neq 0 \\ e^{-c_1 SC}, & \overline{SC} = 0 \end{cases} \qquad (5)$$

The reward R for specific fulfillment option in this case is given with

$$R = CRT(SC, \overline{SC}, c_1) \qquad (6)$$

The current update rule for the opportunity cost is shown in the equations below.

We split the total opportunity cost into capacity-related opportunity cost DV_j^{cap} and inventory-related opportunity cost $DV_{s,j}^{inv}$:

$$OC_{s,j} = DV_{s,j}^{inv} + DV_j^{cap} \qquad (7)$$

Each of these two kinds of opportunity costs will be updated with a separate update function as:

Inventory Dual Variable update function f_u^{inv} is given with:

$$DV_{s,j}^{inv}(t) = f_u^{inv}\left(DV_{s,j}^{inv}(t-1), R_{m^*}^t, SQ_s^t, INV_{s,j}^{t0}, \alpha, \beta\right) \qquad (8)$$

$$f_u^{inv}(\cdot) = DV_{s,j}^{inv}(t-1) \times \left(1 + \alpha \frac{SQ_s^t}{INV_{s,j}^{t0}}\right) + \beta R_{m^*}^t \frac{SQ_s^t}{INV_{s,j}^{t0}} \frac{SQ_s^t}{SQ_p^{tot}} \qquad (9)$$

Using simplified notation:

$$DV_{s,j}^{inv} := DV_{s,j}^{inv}\left(1 + \alpha \frac{SQ_s}{INV_{s,j}^{t0}}\right) + \beta R_{m^*} \frac{SQ_s}{INV_{s,j}^{t0}} \frac{SQ_s}{SQ_p^{tot}} \qquad (10)$$

α, β – algorithm parameters; Usually, $\alpha = 1.0$ and $\beta = 0.4$.

Capacity Dual Variable update function is given with:

$$DV_j^{cap}(t) = f_u^{cap}\left(DV_j^{cap}(t-1), R_{m^*}^t, SQ_s^t, UCP_j^{daily}, \alpha, \beta\right) \qquad (11)$$

$$f_u^{cap}(\cdot) = DV_j^{cap}(t-1) \times \left(1 + \alpha \frac{SQ_p^{tot}}{UCP_j^{daily}}\right) + \beta R_{m^*}^t \frac{SQ_p^{tot}}{UCP_j^{daily}} \frac{SQ_p^{tot}}{UCP_j^{daily}} \qquad (12)$$

Using simplified notation:

$$DV_j^{cap} := DV_j^{cap}\left(1 + \alpha \frac{SQ_p^{tot}}{UCP_j^{daily}}\right) + \beta R_{m^*} \frac{SQ_p^{tot}}{UCP_j^{daily}} \frac{SQ_p^{tot}}{UCP_j^{daily}} \qquad (13)$$

Here $SQ_p^{tot} = \sum_{s \in p} SQ_s$ is the total fulfillment units of all products in the current bundle p.

The adjusted reward for the two possible fulfillment options for $j = 1{,}2$ is given with

$$\hat{R}_j = R_j - OC_{s,j} \qquad (14)$$

Since we are dealing with single product, we can drop the subscript s from all parameters.
Additionally, if we assume that all orders are with single bundle and unit of quantity (9) and (12) simplify to:

$$f_u^{inv}(\cdot) = DV_{s,j}^{inv}(t-1) \times \left(1 + \alpha \frac{1}{INV_j^{t0}}\right) + \beta R_{m^*}^t \frac{1}{INV_j^{t0}} \qquad (15)$$

$$f_u^{cap}(\cdot) = DV_j^{cap}(t-1) \times \left(1 + \alpha \frac{1}{UCP_j^{daily}}\right) + \beta R_{m^*}^t \frac{1}{UCP_j^{daily}} \frac{1}{UCP_j^{daily}} \qquad (16)$$

The Expected reward with uniform random distribution of the incoming order along $\mathcal{R}$ is given with

$$E(R_j) = \int_0^1 CRT(SC_j, \overline{SC})dx \qquad (17)$$

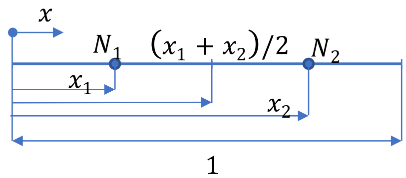


Figure 2: node layout for the scenario with single objective and two nodes

Thus, we have:

$E(R_j) = \int_0^1 e^{\frac{-(|x_j - x| - \overline{SC})}{\overline{SC}}} dx$ where $\overline{SC} = \frac{\sum_{j=1}^{N} |x_j - x|}{N}$ with $N = 2$ (18)

We denote $\bar{x}_{12} = \frac{x_1 + x_2}{2}$, $r_{12} = \frac{x_2 - x_1}{2}$, $d_{12} = x_2 - x_1$ (19)

In order to solve the integral in (18) we need to consider the following four cases:

Case 1) $0 < x < x_1$

In this case we have $\overline{SC} = \bar{x}_{12} - x$. Then $E(R_1) = \int_0^{x_1} e^{\frac{-(x_1 \cancel{-x} - \bar{x}_{12} \cancel{+x})}{\bar{x}_{12} - x}} dx$ which becomes

$E(R_1) = \int_0^{x_1} e^{\frac{r_{12}}{\bar{x}_{12} - x}} dx.$ (20)

Similarly, $E(R_2) = \int_0^{x_1} e^{\frac{-(x_2 \cancel{-x} - \bar{x}_{12} \cancel{+x})}{\bar{x}_{12} - x}} dx$ becomes $E(R_2) = \int_0^{x_1} e^{-\frac{r_{12}}{\bar{x}_{12} - x}} dx.$ (21)

Obviously in this case $CRT(SC_1, \overline{SC}) = e^{\frac{r_{12}}{\bar{x}_{12} - x}}$ and $CRT(SC_2, \overline{SC}) = e^{-\frac{r_{12}}{\bar{x}_{12} - x}}$ (22)

$\min\left(CRT(SC_1, \overline{SC})\right)\Big|_0^{x_1} = e^{\frac{r_{12}}{\bar{x}_{12}}}$ at $x = 0$, $\max\left(CRT(SC_1, \overline{SC})\right)\Big|_0^{x_1} = e$ at $x = x_1$

$\min\left(CRT(SC_2, \overline{SC})\right)\Big|_0^{x_1} = {}^1\!/_e$ at $x = x_1$, $\max\left(CRT(SC_1, \overline{SC})\right)\Big|_0^{x_1} = e^{-\frac{r_{12}}{\bar{x}_{12}}}$ at $x = 0$

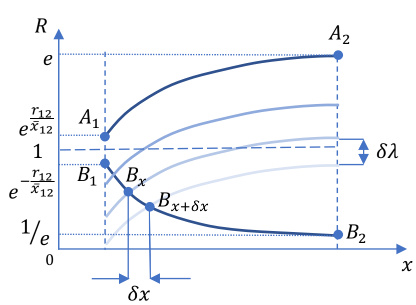


Figure 3: The rewards for case 1): $0 < x < x_1$

Case 2) $x_1 < x < x_1 + r_{12}$

In this case we have $\overline{SC} = r_{12}$. Then $E(R_1) = \int_{x_1}^{x_1 + r_{12}} e^{\frac{-(x - x_1 - r_{12})}{r_{12}}} dx$ which becomes

$E(R_1) = \int_{x_1}^{x_1 + r_{12}} e^{\frac{\bar{x}_{12} - x}{r_{12}}} dx.$ (23)

Similarly, $E(R_2) = \int_{x_1}^{x_1 + r_{12}} e^{\frac{-(x_2 - x - r_{12})}{r_{12}}} dx$ becomes $E(R_2) = \int_{x_1}^{x_1 + r_{12}} e^{-\frac{\bar{x}_{12} - x}{r_{12}}} dx.$ (24)

Obviously in this case $CRT(SC_1, \overline{SC}) = e^{\frac{\bar{x}_{12}-x}{r_{12}}}$ and $CRT(SC_2, \overline{SC}) = e^{-\frac{\bar{x}_{12}-x}{r_{12}}}$ (25)

$\min\left(CRT(SC_1, \overline{SC})\right)\Big|_{x_1}^{x_1+r_{12}} = 1$ at $x = x_1 + r_{12}$, $\max\left(CRT(SC_1, \overline{SC})\right)\Big|_{x_1}^{x_1+r_{12}} = e$ at $x = x_1$

$\min\left(CRT(SC_2, \overline{SC})\right)\Big|_{x_1}^{x_1+r_{12}} = 1/e$ at $x = x_1$, $\max\left(CRT(SC_2, \overline{SC})\right)\Big|_{x_1}^{x_1+r_{12}} = 1$ at $x = x_1 + r_{12}$

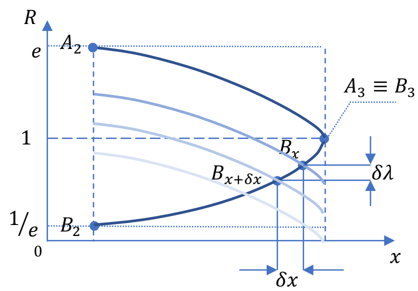


Figure 4: The rewards for case 2): $x_1 < x < x_1 + r_{12}$

Case 3) $x_1 + r_{12} < x < x_2$

In this case we have $\overline{SC} = r_{12}$. Then $E(R_1) = \int_{x_1+r_{12}}^{x_2} e^{\frac{-(x-x_1-r_{12})}{r_{12}}} dx$ which becomes

$E(R_1) = \int_{x_1+r_{12}}^{x_2} e^{-\frac{x-\bar{x}_{12}}{r_{12}}} dx.$ (26)

Similarly, $E(R_2) = \int_{x_1+r_{12}}^{x_2} e^{\frac{-(x_2-x-r_{12})}{r_{12}}} dx$ becomes $E(R_2) = \int_{x_1+r_{12}}^{x_2} e^{\frac{x-\bar{x}_{12}}{r_{12}}} dx$ (27)

Obviously in this case $CRT(SC_1, \overline{SC}) = e^{-\frac{x-\bar{x}_{12}}{r_{12}}}$ and $CRT(SC_2, \overline{SC}) = e^{\frac{x-\bar{x}_{12}}{r_{12}}}$ (28)

$\min\left(CRT(SC_1, \overline{SC})\right)\Big|_{x_1+r_{12}}^{x_2} = 1/e$ at $x = x_2$, $\max\left(CRT(SC_1, \overline{SC})\right)\Big|_{x_1+r_{12}}^{x_2} = 1$ at $x = x_1 + r_{12}$

$\min\left(CRT(SC_2, \overline{SC})\right)\Big|_{x_1+r_{12}}^{x_2} = 1$ at $x = x_1 + r_{12}$, $\max\left(CRT(SC_2, \overline{SC})\right)\Big|_{x_1+r_{12}}^{x_2} = e$ at $x = x_2$

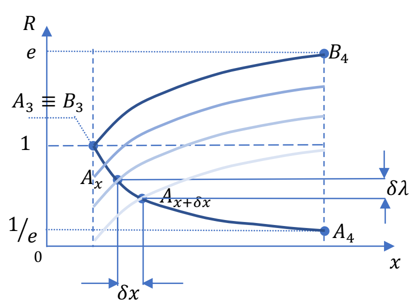


Figure 5: The rewards for case 3): $x_1 + r_{12} < x < x_2$

Case 4) $x_2 < x < 1$

In this case we have $\overline{SC} = x - \bar{x}_{12}$. Then $E(R_1) = \int_{x_2}^{1} e^{\frac{-(\cancel{x}-x_1+\bar{x}_{12}-\cancel{x})}{x-\bar{x}_{12}}} dx$ which becomes

$E(R_1) = \int_{x_2}^{1} e^{-\frac{r_{12}}{x-\bar{x}_{12}}} dx.$ (29)

Similarly, $E(R_2) = \int_{x_2}^{1} e^{\frac{-(\cancel{x}-x_2+\bar{x}_{12}-\cancel{x})}{x-\bar{x}_{12}}} dx$ becomes $E(R_2) = \int_{x_2}^{1} e^{\frac{r_{12}}{x-\bar{x}_{12}}} dx.$ (30)

Obviously in this case $CRT(SC_1, \overline{SC}) = e^{-\frac{r_{12}}{x-\bar{x}_{12}}}$ and $CRT(SC_2, \overline{SC}) = e^{\frac{r_{12}}{x-\bar{x}_{12}}}$ (31)

$\min\left(CRT(SC_1, \overline{SC})\right)\Big|_{x_2}^{1} = 1/e$ at $x = x_2$, $\max\left(CRT(SC_1, \overline{SC})\right)\Big|_{x_2}^{1} = e^{-\frac{r_{12}}{1-\bar{x}_{12}}}$ at $x = 1$
$\min\left(CRT(SC_2, \overline{SC})\right)\Big|_{x_2}^{1} = e$ at $x = x_2$, $\max\left(CRT(SC_2, \overline{SC})\right)\Big|_{x_2}^{1} = e^{\frac{r_{12}}{1-\bar{x}_{12}}}$ at $x = 1$

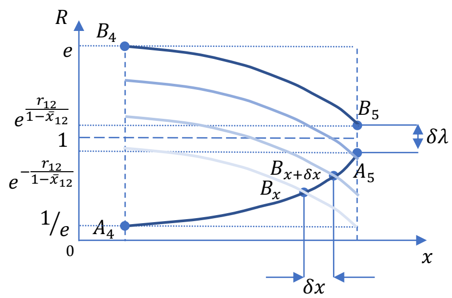


Figure 6: The rewards for case 4): $x_2 < x < 1$

In order to obtain the expected reward for the optimal fulfillment option m^* we want to evaluate $E\left(\max(R_1, R_2)\right) = \int_0^1 \max\left(CRT(SC_1, \overline{SC}), CRT(SC_2, \overline{SC})\right) dx$

Case 1) $0 < x < x_1$

We have $CRT(SC_1, \overline{SC}) > CRT(SC_2, \overline{SC})$ so $E\left(\max(R_1, R_2)\right)\Big|_0^{x_1} = E(R_1)\big|_0^{x_1} = \int_0^{x_1} e^{\frac{r_{12}}{\bar{x}_{12}-x}} dx$ (32)

Case 2) $x_1 < x < \bar{x}_{12}$

We have $CRT(SC_1, \overline{SC}) > CRT(SC_2, \overline{SC})$ so $E\left(\max(R_1, R_2)\right)\Big|_{x_1}^{\bar{x}_{12}} = E(R_1) = \int_{x_1}^{x_1+r_{12}} e^{\frac{\bar{x}_{12}-x}{r_{12}}} dx$ (33)

Case 3) $\bar{x}_{12} < x < x_2$

We have $CRT(SC_1, \overline{SC}) < CRT(SC_2, \overline{SC})$ so $E\left(\max(R_1, R_2)\right)\Big|_{\bar{x}_{12}}^{x_2} = E(R_2) = \int_{x_1+r_{12}}^{x_2} e^{\frac{x-\bar{x}_{12}}{r_{12}}} dx$ (34)

Case 4) $x_2 < x < 1$

We have $CRT(SC_1, \overline{SC}) < CRT(SC_2, \overline{SC})$ so $E\left(\max(R_1, R_2)\right)\Big|_{x_2}^{1} = E(R_2) = \int_{x_2}^{1} e^{\frac{r_{12}}{x-\bar{x}_{12}}} dx$ (35)

Thus, for the total expected reward of the optimal fulfillment option m^* one can write:

$E\left(R_{m^*}^t\right) = x_1 E\left(\max(R_1, R_2)\right)\Big|_0^{x_1} + r_{12} E\left(\max(R_1, R_2)\right)\Big|_{x_1}^{\bar{x}_{12}} + r_{12} E\left(\max(R_1, R_2)\right)\Big|_{\bar{x}_{12}}^{x_2} + (1 - x_2) E\left(\max(R_1, R_2)\right)\Big|_{x_2}^{1}$ (36)

Looking into (22), (25), (28), (31) one may want to conclude that:

$$E\left(R_{m^*}^t\right) = x_1 E(R_1)|_0^{x_1} + r_{12}E(R_1)|_{x_1}^{\bar{x}_{12}} + r_{12}E(R_2)|_{\bar{x}_{12}}^{x_2} + (1-x_2)E(R_2)|_{x_2}^{1} \qquad (37)$$

There is a problem with the equation (37) – it claims that it computes the expected reward for the optimal fulfillment option, but it does not account for the opportunity cost. Opportunity cost will need to be accounted for each fulfillment option before the reward of the optimal fulfillment option is computed. We need to factor eq. (14) in (36) before the reward of the optimal fulfillment option is computed.

Thus, in general, we will have the following expression for the marginal expected adjusted by the opportunity cost reward:

$$E\left(\hat{R}_{m^*}^t\right) = E\left(\max\left(\hat{R}_1, \hat{R}_2\right)\right) \quad (38)$$

$$E\left(\max\left(\hat{R}_1, \hat{R}_2\right)\right) = x_1\, E\left(\max\left(\hat{R}_1, \hat{R}_2\right)\right)\Big|_0^{x_1} + r_{12}\, E\left(\max\left(\hat{R}_1, \hat{R}_2\right)\right)\Big|_{x_1}^{\bar{x}_{12}} + r_{12}\, E\left(\max\left(\hat{R}_1, \hat{R}_2\right)\right)\Big|_{\bar{x}_{12}}^{x_2} +$$
$$(1-x_2)\, E\left(\max\left(\hat{R}_1, \hat{R}_2\right)\right)\Big|_{x_2}^{1} \qquad (39)$$

where $\hat{R}_j$ is given with (14). In our scenario of a single product, we have only a node dependency in the opportunity cost which will be denoted with λ_1 and λ_2 accordingly:

$$\hat{R}_j = R_j - \lambda_j \qquad (40)$$

The Case of Zero Opportunity Cost

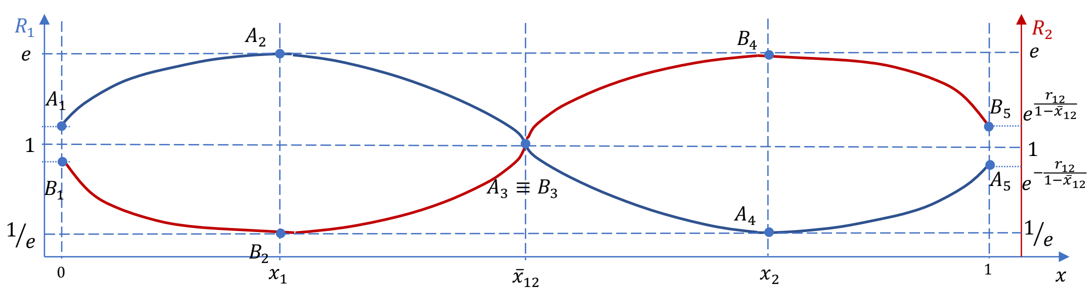


Figure 7: Rewards with zero opportunity costs for the scenario with single objective and two nodes

The opportunity costs are zero in case this is *the first order* ($t = 0$) to be processed by this set of nodes. If the opportunity costs are zero in (40), that is $\lambda_j = 0, j = 1,2$, then (39) transforms to (37). In turn, (37) can be written explicitly as a sum of integrals as:

$$E\left(R_{m^*}^0\right) = x_1 \int_0^{x_1} e^{\frac{r_{12}}{\bar{x}_{12}-x}} dx + r_{12} \int_{x_1}^{x_1+r_{12}} e^{\frac{\bar{x}_{12}-x}{r_{12}}} dx + r_{12} \int_{x_1+r_{12}}^{x_2} e^{\frac{x-\bar{x}_{12}}{r_{12}}} dx + (1-x_2) \int_{x_2}^{1} e^{\frac{r_{12}}{x-\bar{x}_{12}}} dx \quad (41)$$

Using the solutions of these integrals provided in the Appendix (41) transforms to

$$E\left(R_{m^*}^0\right) = x_1\left(-r_{12}e + \bar{x}_{12}e^{\frac{r_{12}}{\bar{x}_{12}}} + r_{12}\left(\mathrm{Ei}(1) - \mathrm{Ei}\left(\frac{r_{12}}{\bar{x}_{12}}\right)\right)\right) +$$

$$(1-x_2)\left(-r_{12}e+(1-\bar{x}_{12})e^{\frac{r_{12}}{1-\bar{x}_{12}}}+r_{12}\left(\mathrm{Ei}(1)-\mathrm{Ei}\left(\frac{r_{12}}{1-\bar{x}_{12}}\right)\right)\right)-2{r_{12}}^2+{r_{12}}^2e+{r_{12}}^2e^{-1} \qquad (42)$$

Using (15) and (16) we can obtain the opportunity cost at $t = 0$ as:

$\lambda_{m^*}^0 = \gamma_{m^*}R_{m^*}^0$, $\lambda_j^0 = 0$ for $j \neq m^*$ (43)

Here γ_{m^*} is a parameter which depends on $UCP_{m^*}^{daily}$ and INV_j^{t0}.

Note that for the first order $E(m^* = 1) = \bar{x}_{12}$ and $E(m^* = 2) = 1-\bar{x}_{12}$ (44)

For the conditional expectations $E\left(R_{m^*}^0|m^* = 1\right)$ and $E\left(R_{m^*}^0|m^* = 2\right)$ we have accordingly:

$$E\left(R_{m^*}^0|m^* = 1\right) = \frac{1}{\bar{x}_{12}}\left[x_1\left(-r_{12}e+\bar{x}_{12}e^{\frac{r_{12}}{\bar{x}_{12}}}+r_{12}\left(\mathrm{Ei}(1)-\mathrm{Ei}\left(\frac{r_{12}}{\bar{x}_{12}}\right)\right)\right)-{r_{12}}^2+{r_{12}}^2e\right] \qquad (45)$$

$$E\left(R_{m^*}^0|m^* = 2\right) = \frac{1}{(1-\bar{x}_{12})}\left[(1-x_2)\left(-r_{12}e+(1-\bar{x}_{12})e^{\frac{r_{12}}{1-\bar{x}_{12}}}+r_{12}\left(\mathrm{Ei}(1)-\mathrm{Ei}\left(\frac{r_{12}}{1-\bar{x}_{12}}\right)\right)\right)-{r_{12}}^2+{r_{12}}^2e^{-1}\right]$$

(46)

Then the conditional expectations for the opportunity costs can be found from (43)-(46):

$$E\left(\lambda_{m^*}^0|m^* = 1\right) = \frac{\gamma_{m^*}}{\bar{x}_{12}}\left[x_1\left(-r_{12}e+\bar{x}_{12}e^{\frac{r_{12}}{\bar{x}_{12}}}+r_{12}\left(\mathrm{Ei}(1)-\mathrm{Ei}\left(\frac{r_{12}}{\bar{x}_{12}}\right)\right)\right)-{r_{12}}^2+{r_{12}}^2e\right] \qquad (47)$$

$$E\left(\lambda_{m^*}^0|m^* = 2\right) = \frac{\gamma_{m^*}}{(1-\bar{x}_{12})}\left[(1-x_2)\left(-r_{12}e+(1-\bar{x}_{12})e^{\frac{r_{12}}{1-\bar{x}_{12}}}+r_{12}\left(\mathrm{Ei}(1)-\mathrm{Ei}\left(\frac{r_{12}}{1-\bar{x}_{12}}\right)\right)\right)-{r_{12}}^2+{r_{12}}^2e^{-1}\right]$$

(48)

Notice that when $\bar{x}_{12} = \frac{1}{2}$ then $E\left(\lambda_{m^*}^0|m^* = 1\right) = E\left(\lambda_{m^*}^0|m^* = 2\right)$ and $E\left(R_{m^*}^0|m^* = 1\right) = E\left(R_{m^*}^0|m^* = 2\right)$.

The Case of Non-Zero Opportunity Cost

Indexing (39) with the order index t yields:

$\hat{R}_j^t = R_j^t - \lambda_j^{t-1}$ where $\lambda_j^{t-1} = 0$ when $t = 0$.

Thus, non-zero opportunity cost will be present when we calculate rewards for $t > 0$.

For simplicity of the analysis for now we will assume that in the update expressions (15) and (16) α is equal to 0. With this assumption we will ignore the linear scaling component of the new opportunity cost with the accumulated opportunity cost to this moment.
Also let us assume that node 1 was chosen n_1 times and node 2 was chosen n_2 times until the previous moment $t-1$. Then the conditional expectation of the opportunity cost for the node 1 and node 2 assignments at moment $t-1$ will be given with:

$$E\left(\lambda_{m^*}^{t-1}|m^* = 1, n_1 \text{ times}\right) = n_1\frac{\gamma_{m^*}}{\bar{x}_{12}}\left[x_1\left(-r_{12}e+\bar{x}_{12}e^{\frac{r_{12}}{\bar{x}_{12}}}+r_{12}\left(\mathrm{Ei}(1)-\mathrm{Ei}\left(\frac{r_{12}}{\bar{x}_{12}}\right)\right)\right)-{r_{12}}^2+{r_{12}}^2e\right] \qquad (49)$$

$$E\left(\lambda_{m^*}^{t-1}|m^* = 2, n_2 \text{ times}\right) = n_2 \frac{\gamma_{m^*}}{(1-\bar{x}_{12})}\left[(1-x_2)\left(-r_{12}e + (1-\bar{x}_{12})e^{\frac{r_{12}}{1-\bar{x}_{12}}} + r_{12}\left(\mathrm{Ei}(1) - \mathrm{Ei}\left(\frac{r_{12}}{1-\bar{x}_{12}}\right)\right)\right) - {r_{12}}^2 + {r_{12}}^2 e^{-1}\right] \quad (50)$$

$$n_1 + n_2 = t - 1 \quad (51)$$

For brevity we will denote the conditional expectation $E\left(\lambda_{m^*}^{t-1}|m^* = 1, n_1 \text{ times}\right)$ with $\bar{\lambda}_1^{t-1}(n_1)$ and $E\left(\lambda_{m^*}^{t-1}|m^* = 2, n_2 \text{ times}\right)$ with $\bar{\lambda}_2^{t-1}(n_2)$. Then (38) can be expanded as:

$$E\left(\hat{R}_{m^*}^t\right) = E\left(\max\left(R_1^t - \bar{\lambda}_1^{t-1}(n_1), R_1^t - \bar{\lambda}_2^{t-1}(n_2)\right)\right) \quad (52)$$

Let us denote the difference $\delta\lambda = \left|\bar{\lambda}_1^{t-1} - \bar{\lambda}_2^{t-1}\right| > 0$. We will show a graphical solution from which one can obtain the conditional expectation of the adjusted by the opportunity cost reward. When the opportunity cost is zero the curves of the two node rewards is shown on Figure 7. On Figure 7 and Figure 8 the bold blue curve represents the unadjusted reward for node 1 and the bold red curve represents the unadjusted reward of node 2. With non-zero opportunity cost accumulated from past decisions until the moment $t-1$ we will need to translate down the bold blue curve by the amount $\delta\lambda$ in case $\bar{\lambda}_1^{t-1} < \bar{\lambda}_2^{t-1}$ **or** we will need to translate down the bold red curve by the amount $\delta\lambda$ in case $\bar{\lambda}_1^{t-1} > \bar{\lambda}_2^{t-1}$. These translations are visualized on Figure 8 with lighter hues of blue and red accordingly.
For instance, when the unadjusted reward curve for node 1 is translated by a total of $\delta\lambda_1'' = \bar{\lambda}_2^{t-1} - \bar{\lambda}_1^{t-1} > 0$ the bold blue curve will be translated down so that it will intersect with the bold red curve at point B_1'' in region $[0, x_1]$ and at point B_3'' in region $[x_1, \bar{x}_{12}]$. Thus, the translated blue curve, corresponding to the adjusted reward of node 1 will still be higher than the bold red curve corresponding to the adjusted reward of node 2, in the sector $[B_1'', B_3''] \subset [0, \bar{x}_{12}]$. For comparison, when there is zero opportunity cost, and therefore zero translation, the reward of node 1 would dominate the reward of node 2 in the whole region $[0, \bar{x}_{12}]$.
Similar observations can be made for the case when $\delta\lambda_2'' = \bar{\lambda}_1^{t-1} - \bar{\lambda}_2^{t-1} > 0$.
In this case the reward of node 2 would dominate the reward of node 1 in the sector $[A_1'', A_3''] \subset [\bar{x}_{12}, 1]$. For comparison, with zero opportunity cost, and therefore zero translation, the reward of node 2 would dominate the reward of node 1 in the whole region $[\bar{x}_{12}, 1]$.

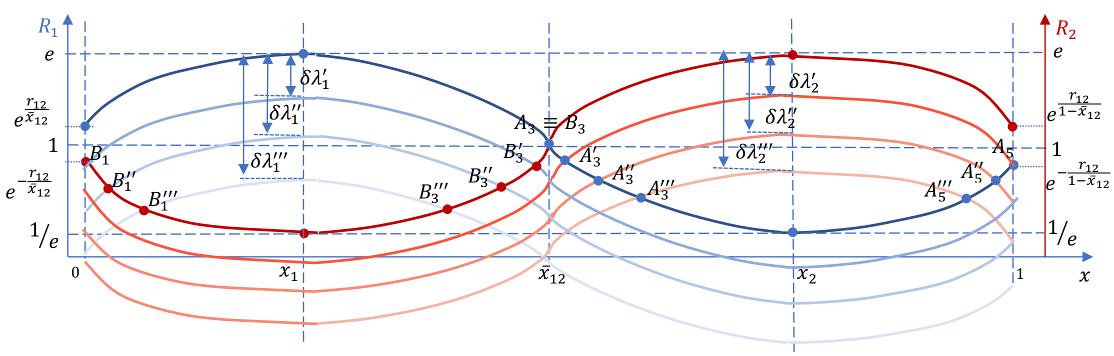


Figure 8: Rewards and non-zero opportunity costs for the scenario with single objective and two nodes

When $\bar{\lambda}_2^{t-1} - \bar{\lambda}_1^{t-1} > 0$ the locations along the x axis of the points B_1'' and B_3'' can be computed by the set of equations:

$$e^{\frac{r_{12}}{\bar{x}_{12}-x}} - \bar{\lambda}_1^{t-1} = e^{-\frac{r_{12}}{\bar{x}_{12}-x}} - \bar{\lambda}_2^{t-1}, \; x \in [0, x_1] \quad (53)$$

If B_1'' exists it will be a root of (53) in the interval $[0, x_1]$.

$e^{\frac{\bar{x}_{12}-x}{r_{12}}} - \bar{\lambda}_1^{t-1} = e^{-\frac{\bar{x}_{12}-x}{r_{12}}} - \bar{\lambda}_2^{t-1}, \; x \in [x_1, \bar{x}_{12}]$ (54)
If B_3'' exists it will be a root of (54) in the interval $[x_1, \bar{x}_{12}]$.

When $\bar{\lambda}_1^{t-1} - \bar{\lambda}_2^{t-1} > 0$ the locations along the x axis of the points A_1'' and A_3'' can be computed by the set of equations:

$e^{-\frac{x-\bar{x}_{12}}{r_{12}}} - \bar{\lambda}_1^{t-1} = e^{\frac{x-\bar{x}_{12}}{r_{12}}} - \bar{\lambda}_2^{t-1}, \; x \in [\bar{x}_{12}, x_2]$ (55)
If A_3'' exists it will be a root of (55) in the interval $[\bar{x}_{12}, x_2]$.

$e^{-\frac{r_{12}}{x-\bar{x}_{12}}} - \bar{\lambda}_1^{t-1} = e^{\frac{r_{12}}{x-\bar{x}_{12}}} - \bar{\lambda}_2^{t-1}, \; x \in [x_2, 1]$ (56)
If A_5'' exists it will be a root of (56) in the interval $[x_2, 1]$.

Notice that each of the equations (53)-(56) can be transformed into a quadratic equation with appropriate variable substitution and expressions for the x coordinates of B_1'', B_3'', A_3'', A_5'' can be derived in closed form.

Thus, for the case when $\delta\lambda_1'' = \bar{\lambda}_1^{t-1} - \bar{\lambda}_2^{t-1} > 0$ we will compute the marginal expected adjusted reward $E(\hat{R}_{m^*}^t)$ as well as the conditional expected adjusted rewards $E(\hat{R}_{m^*}^t | m^*(t) = 1)$ and $E(\hat{R}_{m^*}^t | m^*(t) = 2)$ as:

Case 1) $\bar{\lambda}_2^{t-1} - \bar{\lambda}_1^{t-1} > 0$, both B_1'' and B_3'' exist

In this case there are two distinct real roots of the system (53), (54), resulting in $B_1'' \in [0, x_1]$ and $B_3'' \in [x_1, \bar{x}_{12}]$.

$$
\begin{aligned}
E(\hat{R}_{m^*}^t) = E\left(\max\left(R_1^t - \bar{\lambda}_1^{t-1}(n_1), R_1^t - \bar{\lambda}_2^{t-1}(n_2)\right)\right) = \\
OB_1'' \int_0^{OB_1''} \left[e^{-\frac{r_{12}}{\bar{x}_{12}-x}} - \bar{\lambda}_2^{t-1}(n_2)\right] dx + \\
(x_1 - OB_1'') \int_{OB_1''}^{x_1} \left[e^{\frac{r_{12}}{\bar{x}_{12}-x}} - \bar{\lambda}_1^{t-1}(n_1)\right] dx + \\
[r_{12} - (\bar{x}_{12} - OB_3'')] \int_{x_1}^{OB_3''} \left[e^{\frac{\bar{x}_{12}-x}{r_{12}}} - \bar{\lambda}_1^{t-1}(n_1)\right] dx + \\
(\bar{x}_{12} - OB_3'') \int_{OB_3''}^{x_1+r_{12}} \left[e^{-\frac{\bar{x}_{12}-x}{r_{12}}} - \bar{\lambda}_2^{t-1}(n_2)\right] dx + \\
r_{12} \int_{x_1+r_{12}}^{x_2} \left[e^{\frac{x-\bar{x}_{12}}{r_{12}}} - \bar{\lambda}_2^{t-1}(n_2)\right] dx + \\
(1 - x_2) \int_{x_2}^{1} \left[e^{\frac{r_{12}}{x-\bar{x}_{12}}} - \bar{\lambda}_2^{t-1}(n_2)\right] dx \qquad (57)
\end{aligned}
$$

$$
\begin{aligned}
E(\hat{R}_{m^*}^t | m^*(t) = 1) = E\left(R_1^t - \bar{\lambda}_1^{t-1}(n_1) | m^*(t) = 1\right) = \\
\frac{x_1 - OB_1''}{OB_3'' - OB_1''} \int_{OB_1''}^{x_1} \left[e^{\frac{r_{12}}{\bar{x}_{12}-x}} - \bar{\lambda}_1^{t-1}(n_1)\right] dx + \\
\frac{r_{12} - (\bar{x}_{12} - OB_3'')}{OB_3'' - OB_1''} \int_{x_1}^{OB_3''} \left[e^{\frac{\bar{x}_{12}-x}{r_{12}}} - \bar{\lambda}_1^{t-1}(n_1)\right] dx \qquad (58)
\end{aligned}
$$

$$E\left(\hat{R}_{m^*}^t|m^*(t)=2\right)=E\left(R_2^t-\bar{\lambda}_2^{t-1}(n_2)|m^*(t)=2\right)=$$
$$\frac{OB_1''}{OB_1''+2\bar{x}_{12}-OB_3''}\int_0^{OB_1''}\left[e^{-\frac{r_{12}}{\bar{x}_{12}-x}}-\bar{\lambda}_2^{t-1}(n_2)\right]dx+$$
$$\frac{\bar{x}_{12}-OB_3''}{OB_1''+2\bar{x}_{12}-OB_3''}\int_{OB_3''}^{x_1+r_{12}}\left[e^{-\frac{\bar{x}_{12}-x}{r_{12}}}-\bar{\lambda}_2^{t-1}(n_2)\right]dx+$$
$$\frac{r_{12}}{OB_1''+2\bar{x}_{12}-OB_3''}\int_{x_1+r_{12}}^{x_2}\left[e^{\frac{x-\bar{x}_{12}}{r_{12}}}-\bar{\lambda}_2^{t-1}(n_2)\right]dx+$$
$$\frac{1-x_2}{OB_1''+2\bar{x}_{12}-OB_3''}\int_{x_2}^{1}\left[e^{\frac{r_{12}}{x-\bar{x}_{12}}}-\bar{\lambda}_2^{t-1}(n_2)\right]dx \qquad (59)$$

The integrals in (57)-(59) are not difficult to be evaluated explicitly and expression similar to (42) can be obtained.

Case 2) $\bar{\lambda}_1^{t-1}-\bar{\lambda}_2^{t-1}>0$, both A_1'' and A_3'' exist

In this case there are two distinct real roots of the system (55), (56), resulting in $A_3''\in[\bar{x}_{12},x_2]$ and $A_5''\in[x_2,1]$.

This case can be handled in similar fashion as Case 1) $\bar{\lambda}_1^{t-1}-\bar{\lambda}_2^{t-1}>0$ and the derivation of $E\left(\hat{R}_{m^*}^t\right)$, $E\left(\hat{R}_{m^*}^t|m^*(t)=1\right)$ and $E\left(\hat{R}_{m^*}^t|m^*(t)=2\right)$ is left as an exercise to the reader.

Case 3) $\bar{\lambda}_2^{t-1}-\bar{\lambda}_1^{t-1}>0$, neither B_1'' nor B_3'' exist

In this case there are no real roots of the system (53), (54), resulting in a blue curve well below the local minimum of the bold red curve in the region $[0,\bar{x}_{12}]$. Then we have:

$$E\left(\hat{R}_{m^*}^t\right)=E\left(\max\left(R_1^t-\bar{\lambda}_1^{t-1}(n_1),R_1^t-\bar{\lambda}_2^{t-1}(n_2)\right)\right)=E\left(\hat{R}_{m^*}^t|m^*(t)=2\right)=$$
$$x_1\int_0^{x_1}\left[e^{-\frac{r_{12}}{\bar{x}_{12}-x}}-\bar{\lambda}_2^{t-1}(n_2)\right]dx+$$
$$r_{12}\int_{x_1}^{x_1+r_{12}}\left[e^{-\frac{\bar{x}_{12}-x}{r_{12}}}-\bar{\lambda}_2^{t-1}(n_2)\right]dx+$$
$$r_{12}\int_{x_1+r_{12}}^{x_2}\left[e^{\frac{x-\bar{x}_{12}}{r_{12}}}-\bar{\lambda}_2^{t-1}(n_2)\right]dx+$$
$$(1-x_2)\int_{x_2}^{1}\left[e^{\frac{r_{12}}{x-\bar{x}_{12}}}-\bar{\lambda}_2^{t-1}(n_2)\right]dx \qquad (60)$$

The integrals in (60) are not difficult to be evaluated explicitly and expression similar to (42) can be obtained.

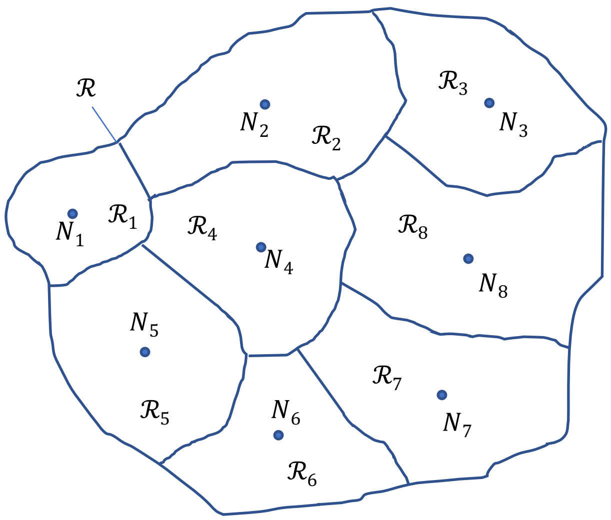
Obviously, $E\left(\hat{R}_{m^*}^t|m^*(t)=1\right)=0$ in this case as node 1 has been chosen too many times in the past so the new decision definitely will be node 2 no matter where the order zip code is.

Case 4) $\bar{\lambda}_1^{t-1}-\bar{\lambda}_2^{t-1}>0$, neither A_1'' nor A_3'' exist

In this case there are no real roots of the system (55), (56), resulting in a red curve well below the local minimum of the bold blue curve in the region $[\bar{x}_{12},1]$.

This case mirrors Case 3) and expressions for the marginal expected reward and conditional expected rewards can be obtained in closed form similarly.

Single objective Shipping Cost, N nodes and single product with unit quantity



//TODO: finish the analysis of the scenario with N nodes

Appendix

The solutions of certain definite integrals:

a. $\int_{x_1}^{x_2} e^{\frac{a}{b-x}} dx = -be^{\frac{a}{b-x_2}} + x_2 e^{\frac{a}{b-x_2}} + be^{\frac{a}{b-x_1}} - x_1 e^{\frac{a}{b-x_1}} + a\mathrm{Ei}\left(\frac{a}{b-x_2}\right) - a\mathrm{Ei}\left(\frac{a}{b-x_1}\right)$

Here $\mathrm{Ei}(x)$ is the exponential integral and it is given with $\mathrm{Ei}(x) = \int_{-x}^{\infty} \frac{e^{\xi}}{\xi} d\xi$.
$\mathrm{Ei}(x)$ converges slowly for arguments with larger moduli.
Converging series for $\mathrm{Ei}(x)$ (found by Srinivasa Ramanujan) is given with:

$$\mathrm{Ei}(x) = \gamma + \ln(x) + e^{x/2} \sum_{n=1}^{\infty} \frac{(-1)^{n-1} x^n}{n! 2^{n-1}} \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \frac{1}{2k+1}$$

Using the alternating series above one can establish the following asymptotic bounds

$$1 - \frac{3x}{4} \leq \mathrm{Ei}(x) - \gamma - \ln(x) \leq 1 - \frac{3x}{4} + \frac{11x^2}{36}$$

b. $\int_{x_1}^{x_2} e^{\frac{x-a}{b}} dx = b\left(e^{\frac{x_2-a}{b}} - e^{\frac{x_1-a}{b}}\right)$