Empirical Distribution of Caller Wait Times

Khalid Elsadig, Krish Mallina, and Joshua Parr

August 1, 2024

1 Introduction

1.1 Problem

In order to answer questions about a random variable with regards to probability, expected value, etc., it is important to know the distribution of the variable. In this project, we used simulation to create an experimental distribution of random variable $\boldsymbol{W}$, the total wait time spent by a caller getting tickets from a ticketing agency.

1.2 Outline

We are given the following assumptions:

- callers need 3 seconds to dial
- callers need 2 seconds to hang up
- callers will dial 3 times before giving up on getting tickets
- callers will spend a maximum of 1.5 minutes waiting for the switchboard to connect them to an agent before hanging up
- the caller can be connected to any one of 4 agents A, B, C, or D, whom will take 72, 96, 81, or 114 seconds, respectively, to get the callers their ticket

This report breaks down into three major sections:

1. Theory: Analyzing the problem using theoretical probability
2. Experimental Testing: Combining theory with programming to simulate 500 callers
3. Results: Values and descriptions are given to describe the distribution of wait times

2 Theory

We are given that the amount of time the switchboard takes to connect a caller to an agent in minutes, X, is defined by the probability density function (PDF) in Equation 1:

$$f(x) = \frac{3}{16}\sqrt{x}, \quad 0 < x < 4 \tag{1}$$

Given this PDF, we integrate over its range to yield the cumulative distribution function (CDF) defined in Equation 2:

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{8}x^{\frac{3}{2}} & 0 \leq x \leq 4 \\ 1 & 4 \leq x \end{cases} \tag{2}$$

The caller will hang up on the switchboard after 1.5 minutes. With this CDF, we approximate that the caller will not hang up on the switchboard with probability 0.2296. We calculated this using Equation 3:

$$F(1.5) \approx 0.2296 \tag{3}$$

With this CDF, we can derive the inverse CDF by solving for x in terms of u, the probability (see Equation 4 for reference). This will be specifically useful later for calculating the wait time the caller spends at the switchboard. The inverse CDF is given in Equation 5

$$F(x) = u = \frac{1}{8}x^{\frac{3}{2}}, \quad 0 < x < 4 \tag{4}$$

$$F^{-1}(u) = 4u^{\frac{3}{2}} = x \tag{5}$$

Using our given parameters, we have devised a flowchart to outline what needs to be tested and calculated during our simulation. First, we test to see if the the caller waits or hangs up on the switchboard. The former has a probability of 0.2296 whereas the latter has a probability of 0.7704. If they hang up, the flowchart has indicated that the caller will redial, given that the number of total dials is less than 3. If the number of total dials is equal to 3, the simulation terminates. This is represented with an arrow looping back to the start. If the caller is successfully connected to an agent, they are assigned to Agent A, B, C, or D at probabilities of 0.2, 0.3, 0.1, or 0.4, respectively. This is also indicated with arrows pointing to each possible outcome. The added wait time is indicated by hexagons with text stating "w+=" the wait time for each section. The flowchart can be seen in Figure 1.

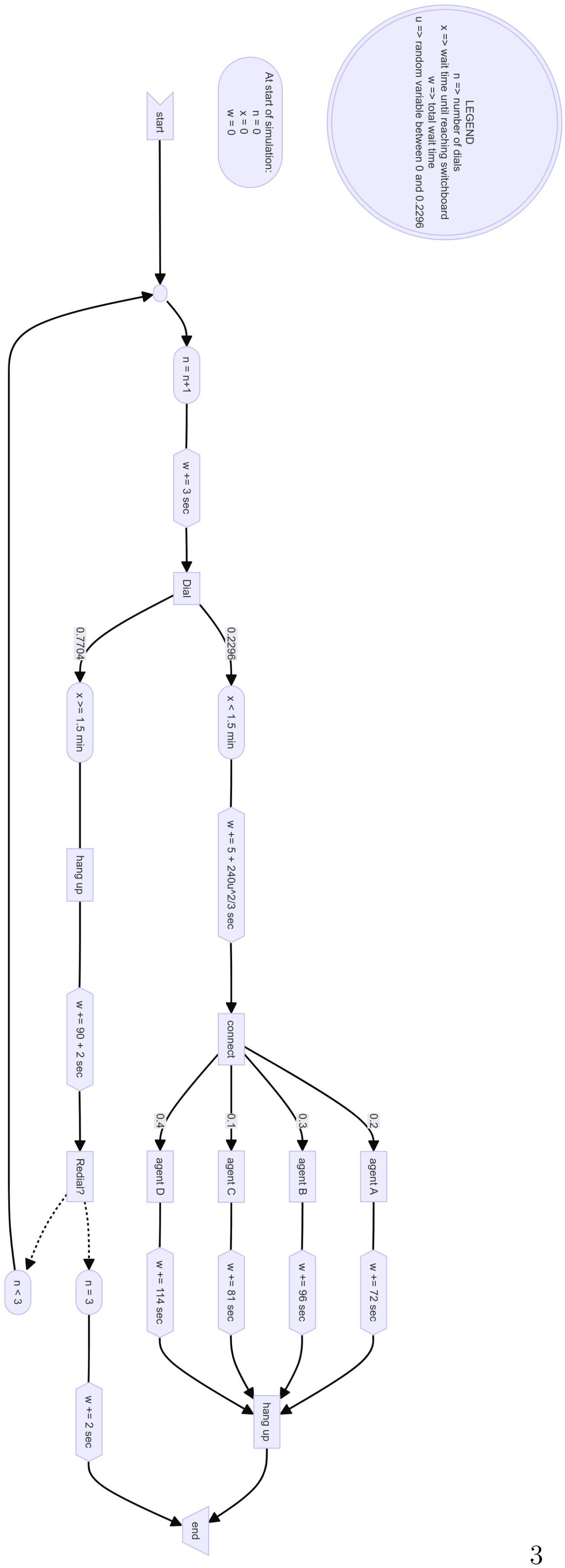

Figure 1: Diagram for Total Wait Time Calculation

3 Algorithm

The algorithm for calculating W was first diagrammed, as shown in Figure 1. Our algorithm begins by defining a variable W, representing wait time in seconds, and n, representing the number of times the caller has dialed the ticketing agency. W is set to 0 by default since the caller has not yet called the agency, and n is set to 1 by default since the simulation starts with the caller dialing the agency. Since dialing requires 3 seconds, W is increased by 3. We then simulate the caller's decision to either wait for the switchboard or hang up. To this end, we created a random variable "u", which is a random number between 0 and 1. "U" is generated using the linear congruential random number generator (LCRNG). The LCRNG is designed to reliably yield random values between 0 to 1 given four parameters: a starting value, multiplier, increment, and modulus. For this simulation, we used 1000 for the starting value, 24693 for the multiplier, 3517 for the increment, and 2^{17} for the modulus. Now that we can reliably generate random values between 0 and 1, we must determine which range of values corresponds to the caller waiting for the switchboard, and which range corresponds to the caller hanging up. For this, we are given that the Probability Density Function (PDF) of the wait time to connect to the switchboard in minutes, X, is defined by the formula 1 and that the caller will hang up after 1.5 minutes. So, the caller will hang up if the switchboard takes longer than 1.5 minutes to connect them to an agent. Therefore, if we want to find the range of values that define whether the caller waits for the switchboard or not, then we must find the probability that the switchboard takes less than 1.5 minutes to connect the caller to an agent.

To do this, we find the Cumulative Distribution Function (CDF) as seen in Equation 2.

The probability that the switchboard takes less than 1.5 minutes to connect the caller to an agent is given by Equation 3. So, if "u" is less than 0.2296, then the simulated caller will be connected to an agent. Otherwise, the caller will hang up. If the caller hangs up, W is increased by 92 seconds because the caller spent 90 seconds waiting for the switchboard and another 2 seconds hanging up. If n is equal to 3, the simulation terminates for that caller, and their wait time, W, is stored to a list of wait times. Otherwise, n is increased by 1 and the caller is put through the switchboard simulation again.

If the caller is connected to an agent, then the time the caller spent waiting for the switchboard must be calculated. For this purpose, we derive the inverse CDF from the CDF in Equation 4 by solving for "x", yielding the inverse CDF in Equation 5.
Note that because this is the inverse CDF we are inputting a probability to get a wait time as the result, which is why the variable is "u" instead of "x". Inputting our value of "u" generated from the LCRNG to make the switchboard decision gives us the time the caller spent waiting for the switchboard in minutes. We multiply this by 60 to put it in units of seconds and increment our variable W by the result.

After the switchboard, the caller can be connected to any one of four agents:

- Agent A who will cause the caller to wait for 72 seconds, with a probability of 0.2;
- Agent B who will cause the caller to wait for 96 seconds, with a probability of 0.3;
- Agent C who will cause the caller to wait for 81 seconds, with a probability of 0.1; or
- Agent D who will cause the caller to wait for 114 seconds, with a probability of 0.4.

To determine which agent the caller gets connected to, we set up a discrete random variable generator (DRVG). The DRVG begins by generating a random number using the LCRNG. Since there is a 0.2 probability of connecting to agent A, we checked if u <= 0.2. Next, we checked if u <= 0.5 for agent B. Then, we checked if u <= 0.6 for agent C. Finally, if all other checks returned false, then the caller was assigned to agent D. It's important to note that in this sequence of checks, the next check happens only if all previous checks returned false (i.e. u <= 0.5 won't be checked if u <= 0.2 is already true). W is then incremented by 72 sec, 96 sec, 81 sec, or 114 sec for agents A, B, C, or D, respectively since that is the time the caller will spend waiting for each agent. The simulation is then terminated and W is stored to a list of wait times.

4 Results

4.1 u-Values

The following equation shows the 51st, 52nd, and 53rd random numbers generated by our linear congruential random number generator:

$$(u_{51}, u_{52}, u_{53}) = (0.5157, 0.4273, 0.7682) \tag{6}$$

4.2 W-Values

The table below shows the total wait times(in numerical order) of the first 500 callers stored by our simulation.

Total Caller Wait Time, $\boldsymbol{W}$ (seconds)									
	W		W		W		W		W
1	80.3071	37	138.7785	73	169.5722	109	205.5162	145	252.5748
2	94.1915	38	140.5359	74	169.9606	110	205.8428	146	254.086
3	95.3157	39	141.1892	75	170.3122	111	206.1945	147	260.2726
4	97.3515	40	141.2223	76	171.5853	112	206.5641	148	260.9217
5	98.4907	41	141.5188	77	173.1257	113	208.0975	149	261.1297
6	100.6623	42	142.1368	78	174.667	114	210.023	150	261.2642
7	103.2418	43	142.6206	79	174.9412	115	211.6621	151	262.7708
8	111.0151	44	142.8792	80	175.6144	116	212.1993	152	263.2627
9	111.5355	45	144.372	81	177.1878	117	216.1923	153	264.5197
10	111.8546	46	145.0255	82	177.9691	118	216.4159	154	264.7319
11	111.9284	47	145.4658	83	178.755	119	217.0977	155	264.8431
12	112.1221	48	147.2786	84	179.5655	120	222.0242	156	265.1108
13	113.8462	49	147.8268	85	181.3524	121	223.1316	157	265.867
14	114.3173	50	148.5819	86	183.1141	122	223.3693	158	267.5842
15	116.0086	51	148.7654	87	183.846	123	225.3265	159	267.6548
16	116.5726	52	149.549	88	185.138	124	226.262	160	267.8556
17	117.0033	53	151.0193	89	185.5159	125	228.6134	161	268.7979
18	117.2202	54	151.3067	90	186.3637	126	228.8354	162	268.9724
19	117.9866	55	153.3622	91	188.2121	127	228.9047	163	268.9896
20	119.0878	56	155.179	92	188.4414	128	230.8098	164	269.2824
21	119.6808	57	155.2182	93	188.5782	129	230.9967	165	269.7165
22	119.9502	58	155.5691	94	189.8017	130	233.58	166	269.7277
23	122.5215	59	157.1133	95	189.9846	131	234.8241	167	270.3213
24	123.2499	60	157.361	96	190.4448	132	238.1241	168	271.8592
25	123.663	61	158.9056	97	191.225	133	238.8074	169	272.4474
26	124.5231	62	159.2791	98	192.1103	134	242.4745	170	272.715
27	126.375	63	160.3096	99	194.0527	135	243.4974	171	274.2085
28	126.474	64	161.115	100	197.6703	136	243.7896	172	275.3449
29	127.7412	65	161.373	101	199.6102	137	245.6746	173	276.2874
30	130.89	66	161.3856	102	200.2785	138	246.6203	174	277.2058
31	131.3597	67	161.9931	103	201.9623	139	246.7337	175	279.0577
32	133.3938	68	164.2842	104	202.4861	140	248.7061	176	280.3901
33	133.8456	69	167.8973	105	203.616	141	249.9359	177	280.4988
34	134.1934	70	168.0693	106	204.0363	142	251.6927	178	281.6192
35	135.5157	71	168.7182	107	204.2229	143	251.7917	179	281.7948
36	137.4657	72	168.8095	108	204.2404	144	251.8383	180	281.92

	W		W		W		W		W
181	283.8098	217	285	253	285	289	285	325	285
182	284.3616	218	285	254	285	290	285	326	285
183	285	219	285	255	285	291	285	327	285
184	285	220	285	256	285	292	285	328	285
185	285	221	285	257	285	293	285	329	285
186	285	222	285	258	285	294	285	330	285
187	285	223	285	259	285	295	285	331	285
188	285	224	285	260	285	296	285	332	285
189	285	225	285	261	285	297	285	333	285
190	285	226	285	262	285	298	285	334	285
191	285	227	285	263	285	299	285	335	285
192	285	228	285	264	285	300	285	336	285
193	285	229	285	265	285	301	285	337	285
194	285	230	285	266	285	302	285	338	285
195	285	231	285	267	285	303	285	339	285
196	285	232	285	268	285	304	285	340	285
197	285	233	285	269	285	305	285	341	285
198	285	234	285	270	285	306	285	342	285
199	285	235	285	271	285	307	285	343	285
200	285	236	285	272	285	308	285	344	285
201	285	237	285	273	285	309	285	345	285
202	285	238	285	274	285	310	285	346	285
203	285	239	285	275	285	311	285	347	285
204	285	240	285	276	285	312	285	348	285
205	285	241	285	277	285	313	285	349	285
206	285	242	285	278	285	314	285	350	285
207	285	243	285	279	285	315	285	351	285
208	285	244	285	280	285	316	285	352	285
209	285	245	285	281	285	317	285	353	285
210	285	246	285	282	285	318	285	354	285
211	285	247	285	283	285	319	285	355	285
212	285	248	285	284	285	320	285	356	285
213	285	249	285	285	285	321	285	357	285
214	285	250	285	286	285	322	285	358	285
215	285	251	285	287	285	323	285	359	285
216	285	252	285	288	285	324	285	360	285

	W		W		W		W		W
361	285	389	285	417	293.3632	445	331.0744	473	354.2739
362	285	390	285	418	294.4835	446	331.9605	474	354.416
363	285	391	285	419	297.3299	447	332.3286	475	355.3074
364	285	392	285	420	298.9599	448	334.2107	476	356.0144
365	285	393	285	421	299.0722	449	334.4166	477	356.998
366	285	394	285	422	299.3522	450	335.589	478	357.8778
367	285	395	285	423	299.7238	451	335.6519	479	359.6681
368	285	396	285	424	300.5991	452	339.3097	480	360.0043
369	285	397	285	425	301.2471	453	339.5582	481	360.4234
370	285	398	285	426	301.4074	454	340.3136	482	361.98
371	285	399	285	427	301.578	455	340.3652	483	369.5725
372	285	400	285	428	301.623	456	341.2485	484	370.9778
373	285	401	285	429	304.0791	457	341.6779	485	371.545
374	285	402	285	430	304.5648	458	341.9322	486	373.4285
375	285	403	285	431	305.3452	459	342.0571	487	373.4858
376	285	404	285	432	305.8555	460	344.6113	488	374.2855
377	285	405	285	433	305.8575	461	344.7547	489	377.6651
378	285	406	285	434	306.5702	462	346.8255	490	380.8398
379	285	407	285	435	307.4492	463	346.9279	491	381.9
380	285	408	285	436	308.6389	464	347.4985	492	383.2242
381	285	409	285.9764	437	315.5264	465	348.2687	493	387.3722
382	285	410	286.5354	438	317.8071	466	349.1675	494	388.0763
383	285	411	287.4136	439	318.7817	467	349.6406	495	390.0089
384	285	412	287.8985	440	321.3445	468	349.8958	496	395.3299
385	285	413	288.1364	441	322.3189	469	350.0553	497	398.1403
386	285	414	290.756	442	324.1895	470	351.7905	498	398.6936
387	285	415	291.0886	443	327.8902	471	351.9297	499	401.5143
388	285	416	291.7107	444	328.1919	472	354.1094	500	401.9495

4.3 Analysis

The sample space of the first 500 W-values is:

$$S_W = [80.3071, 401.9495] \tag{7}$$

It's important to note that the above sample space is just for the discrete experimental values of W that we calculated. Theoretically the sample space of W is given below.

$$S_W = [80, 402] \tag{8}$$

The mean total waiting time for the first 500 callers is approximately 260.744 seconds. The median wait time is 285 seconds. The first quartile and third quartile are 228.724 and 285, respectively. The median wait time comprises of approximately 45% of the data and occurs when the caller dials 3 times and fails to connect to the switch board each time. The probability of this occurrence is given by the equation below:

$$(1 - 0.2296)^3 \approx 0.46 \tag{9}$$

This probability is consistent with our gathered data. The following figure shows the graph of the cumulative distribution function (CDF) of $\boldsymbol{W}$.

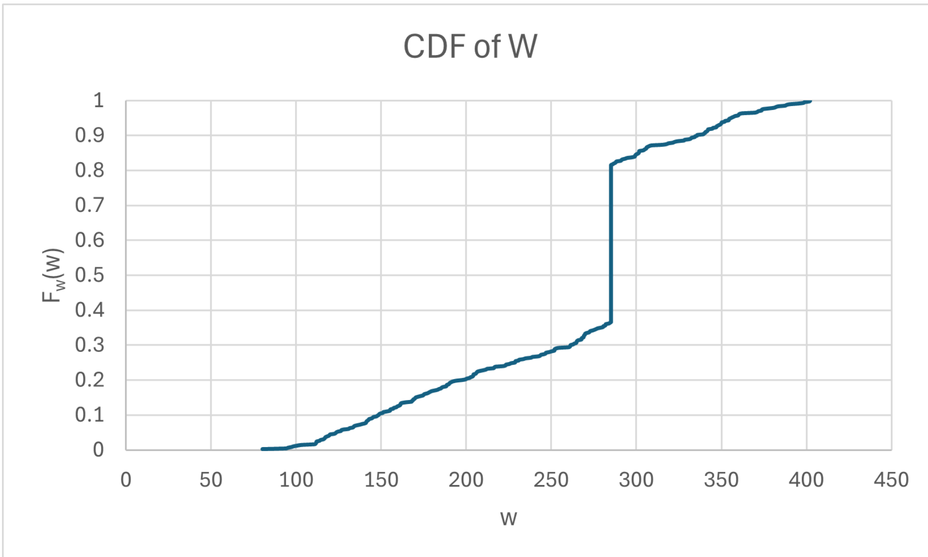

Figure 2: Cumulative Distribution Function of Total Wait Time, $\boldsymbol{W}$ (s)

The jump in the CDF occurs at the median wait time described above. The mean being less than the median means that the data is skewed left and that it is more probable for W to be greater than the mean.