Departament d'Estadística
i Investigació Operativa
UNIVERSITAT POLITÈCNICA DE CATALUNYA

SEMINAR ON MODELING AND SIMULATION

Simulation as a numerical technique to deal with random dynamic systems:

Characterizing data input randomness

Prof. Jaume Barceló

jaume.barcelo@upc.edu

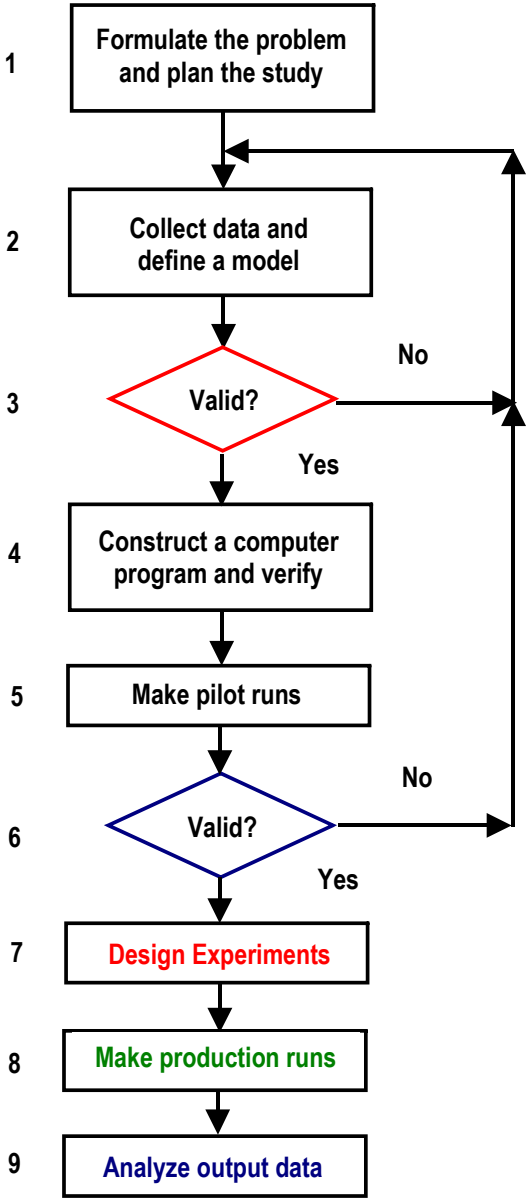
Department of Statistics and Operations Research
Technical University of Catalonia
Barcelona, Spain

Departament d'Estadística
i Investigació Operativa
UNIVERSITAT POLITÈCNICA DE CATALUNYA

REFERENCES

- Stephen Vincent, Input data Analysis, Chapter 3 in Handbook of Simulation, Principles, Methodology, Advances, Applications and Practice, Edited by Jerry Banks, John Wiley, 1998
- J. Banks, J.S. Carson and B.L. Nelson and D.M.Nicol, Discrete-Event System Simulation, Prentice-Hall, 2005. Chapter 10, Input Modeling
- A.M.Law and W.D. Kelton, Simulation Modeling and Analysis, McGraw Hill, 2001. Chapter 6: Selecting Input Probability Distributions

METHODOLOGICAL STEPS IN A SIMULATION STUDY

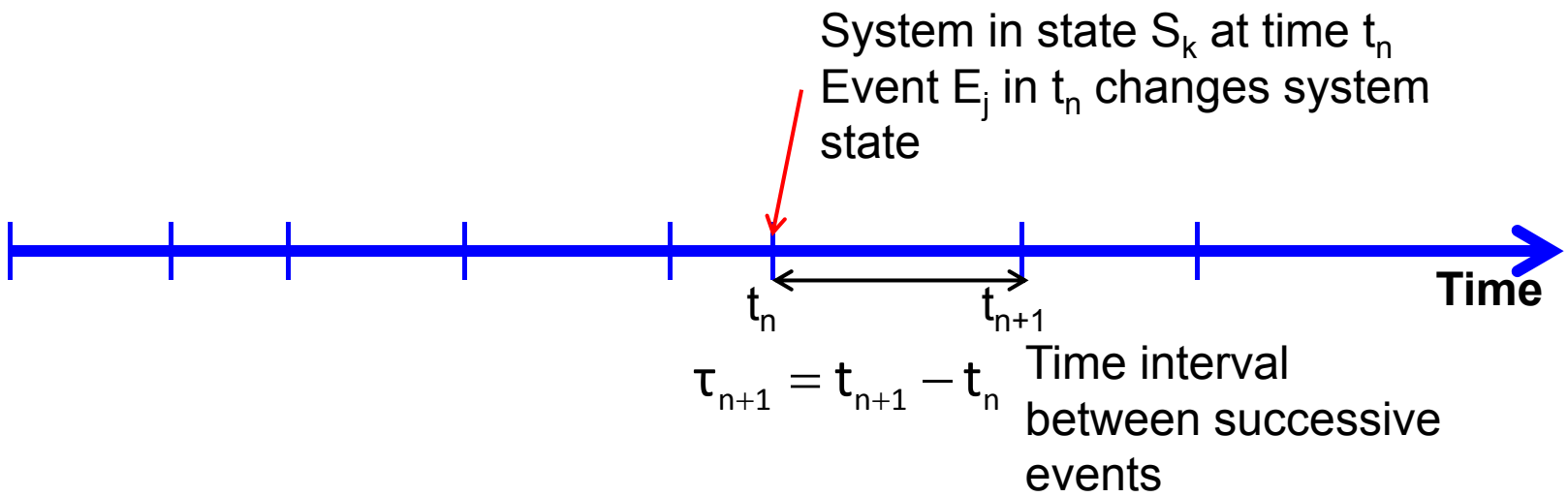


1. Formulate the problem and plan the study: identify the nature of the problem and the requirements to find a solution.
2. Collect data and formulate the model.
3. Check whether the model built is a valid representation of the system for the purposes of the study, that is, verify that the answers provided by the model to the what if questions can be accepted as valid ones.
4. Translate the formal model in terms of a computer program
5. And 6. Check that the computer model performs correctly, that is, is error free and provides acceptable results.
7. Identify the design factors that translate the what if questions in terms of computer experiments.
8. Specify the experimental sampling procedures to gather the data for the statistical analysis that will provide the expected answers.
9. Conduct the simulation experiments on the computer and analyze the outputs.

TIME EVOLUTION IN SIMULATION

ASYNCHRONOUS IN EVENT SCHEDULING

OCCURRENCE OF EVENTS CHANGE SYSTEM STATE



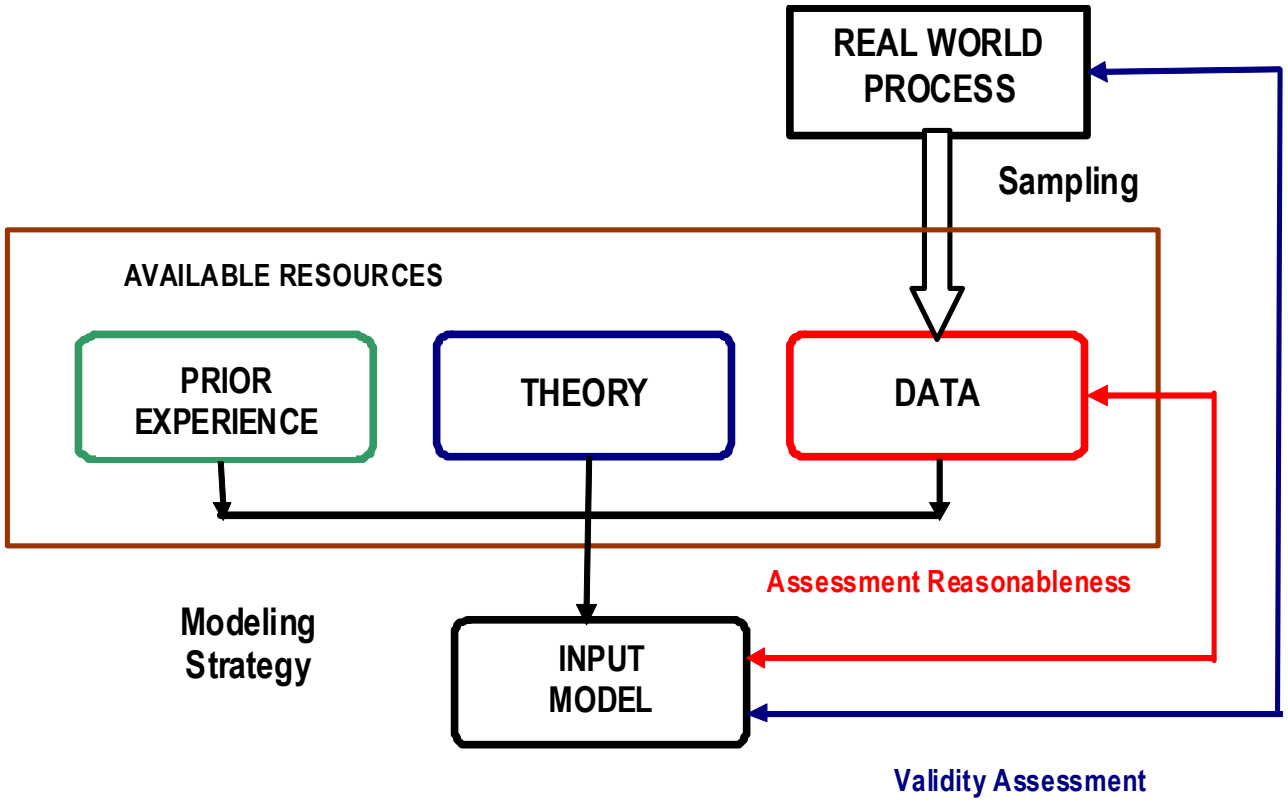
- Events occur randomly $\Rightarrow$ τ is a random variable
- Which random variable?
- How to identify the randomness and the type of randomness (probability distribution that "explains" τ)
- How to generate sequences of events according with the type of randomness identified? $\Rightarrow$ Generating samples of random distributions

MODELING THE INPUT TO SIMULATION MODELS

- The Input Variables to Simulation Models can be viewed as Stochastic Processes (Discrete in our case)
- Stochastic Process: A collection of random variables, X_k, k=1,2,…, (Discrete Time) where subscript k dictates the order of the variables but not the specific time of occurrence, such that:
  - Each X_k is a distinct occurence of the same general random phenomenon X with a probability distribution function $F_k(x) = Pr\{X_k \leq x\}$
  - In the case of the analysis of the random simulation input there is a specific type of stochastic process of special interest: the IID processes (independent, identically distributed)
  - The strongest interrelationship assumptions that we can make are (for instance when, X_k is the service time of the k-th client) :
    - All of the X_k random variables are probabilistically independet of one other
    - All of the X_k random variables follow the same probability distribution and thus are said to be identically distributed: that is $F_k(\bullet)$ has a common form $F(\bullet)$ for all k.
- Alternatives that could be of special interests in simulation:
  - Nonstationary processes: the probability distribution varies as a function of k: $F_k(\bullet)$ is stable but it has a parameterization that depends on time
  - Correlated processes: the subsequent values in the process are not independent of each other (i.e. time series)

INPUT DATA ANALYSIS

- Collect data of sufficient quality, quantity and variety to perform a reasonable analysis
  - Determine wich is the most suitable technology to collect the data and which is the data collection methodology to apply
  - Dessign adequately the sampling process
- Identify the probability distribution most likely representing the input data
- Estimate the parameters of the selected probability distribution
- Evaluate the quality of the fitness of the postulated probability distribution



ASSESING THE INDEPENDENCE OF THE COLLECTED DATA

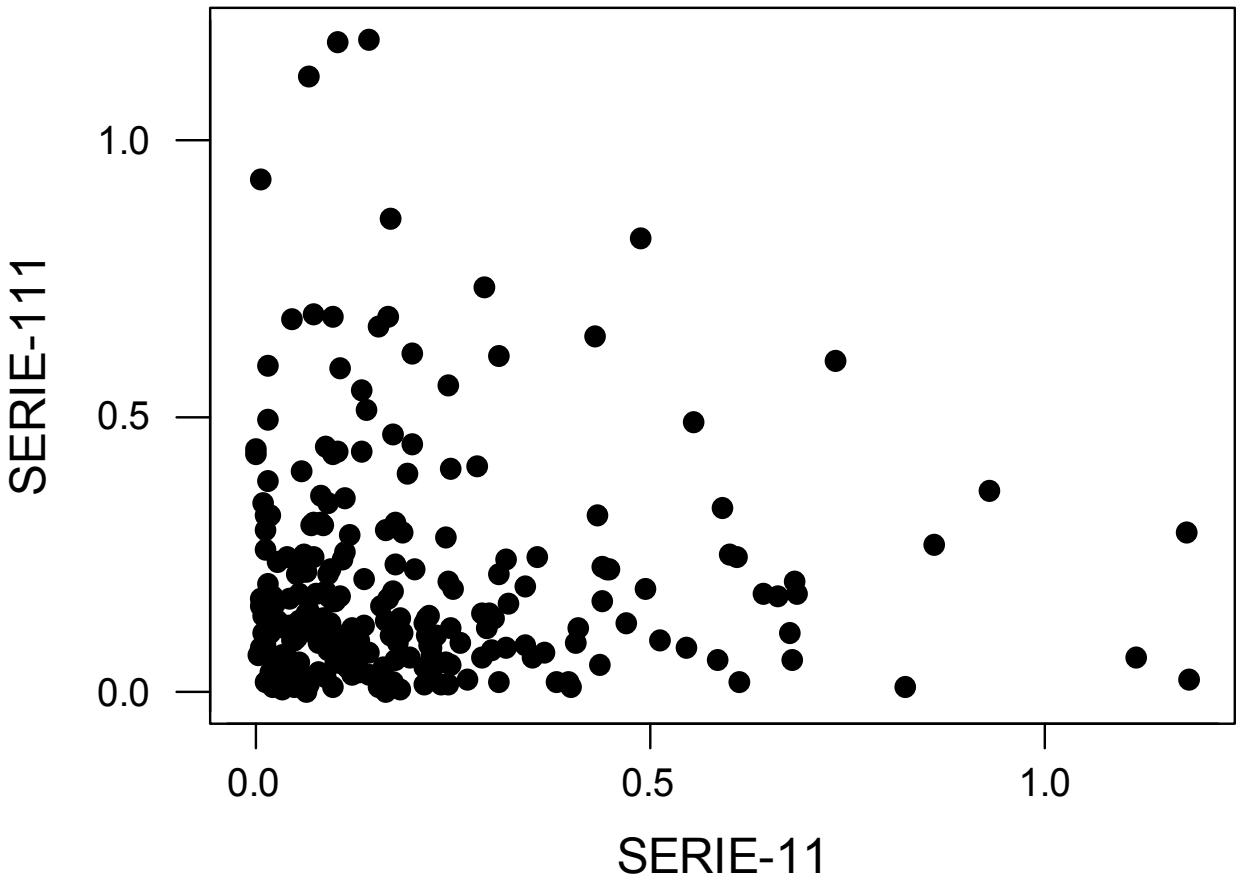
- **Scatter diagram analysis:**
  - **A heuristic procedure that considers the graphical display of sequential pairs of lagged observations**

    (1,2), (2,3), (3,4), ……., (n-1, n)
  - **Correlated samples of observations generate scatter diagrams where points are clustered along the positive or negative diagonals depending on the type of positive or negative correlation.**
  - **Independent samples of observations generate unshaped clusters of points, the type of clouds could depend on the underlying probability distribution..**
- **Spectral analysis: analysis of the autocorrelation function**

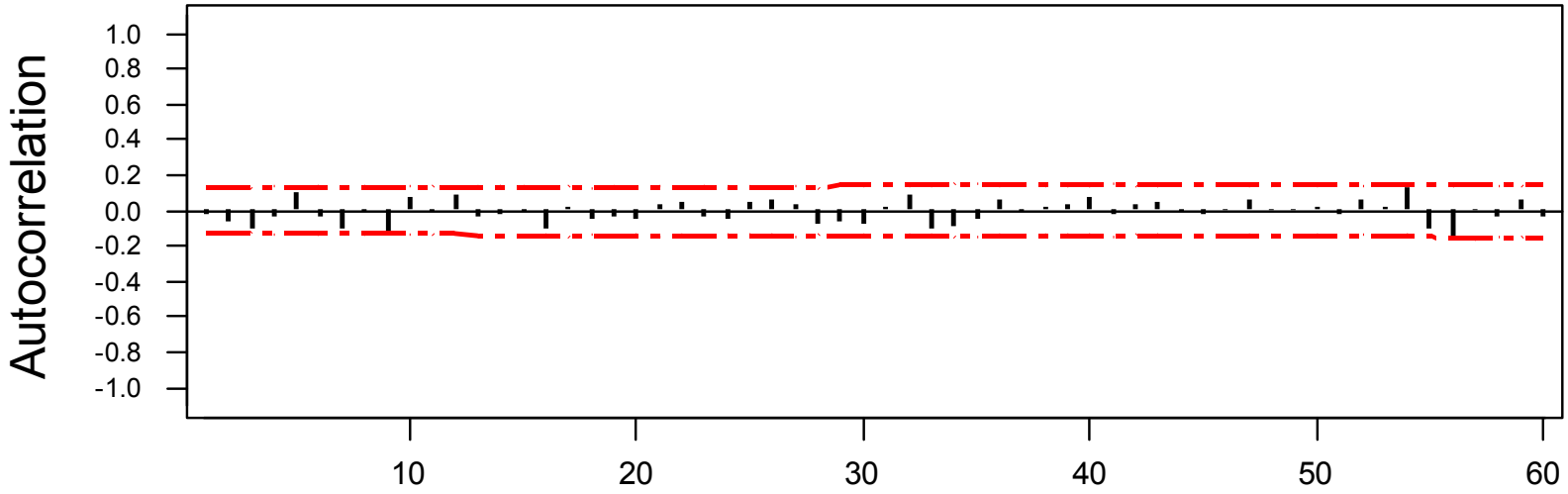
EXAMPLE 1: SERIE-11

			SERIE-11				
0.33615	0.11984	0.34368	0.12491	0.10359	0.10675	0.73578	0.14489
0.59311	0.13784	0.09271	0.09558	0.17286	0.01003	0.29175	0.29555
0.01772	0.2208	0.51194	0.10226	0.10897	0.40066	0.18794	0.16603
0.01791	0.44973	0.14172	0.11456	0.67735	0.05878	0.25245	0.10308
0.01464	0.20095	0.06469	0.29259	0.04589	0.67934	0.11491	0.09119
0.06758	0.24393	0.04048	0.01301	0.11595	0.169	0.03718	0.08214
0.03666	0.60936	0.05222	0.21514	0.12228	0.00659	0.08103	0.00916
0.03523	0.30855	0.10624	0.05491	0.03828	0.0362	0.22367	0.02167
0.0207	0.07357	0.18775	0.10226	0.2253	0.13265	0.09467	1.18317
0.17436	0.10476	0.49406	0.22986	0.44043	0.08778	0.08514	0.14525
0.02388	0.07889	0.0181	0.17913	0.00154	0.12933	0.01448	0.06712
0.26839	0.31957	0.30864	0.64332	0.0648	0.16676	0.24373	0.22522
0.85955	0.02033	0.08337	0.43104	0.28814	0.09917	0.35656	0.06311
0.1716	0.04688	0.22208	0.0007	1.18049	0.08102	0.0848	1.1165
0.16973	0.16184	0.20362	0.16551	0.10461	0.54756	0.34293	0.0687
0.04304	0.02644	0.13824	0.43831	0.10726	0.13514	0.01126	0.0047
0.1171	0.06035	0.01169	0.10438	0.01866	0.18426	0.09885	0.18401
0.40991	0.22409	0.1573	0.21639	0.38199	0.08902	0.05419	0.17591
0.28201	0.44568	0.00893	0.056	0.01635	0.18052	0.12222	0.66337
0.24135	0.0908	0.82254	0.04741	0.61428	0.68739	0.05593	0.15713
0.11251	0.12804	0.48864	0.21575	0.20008	0.07499	0.0572	0.16041
0.12431	0.02981	0.55449	0.30881	0.68305	0.09288	0.17785	0.32255
0.46914	0.12215	0.24523	0.17667	0.09931	0.04882	0.0783	0.43267
0.17541	0.04322	0.03995	0.05506	0.12422	0.43739	0.30097	0.09965
0.09272	0.21958	0.13325	0.24274	0.21592	0.13561	0.0727	0.13036
0.04889	0.06415	0.3024	0.31915	0.09413	0.06409	0.36641	0.05517
0.24667	0.19616	0.08824	0.01325	0.13278	0.35219	0.92859	0.02562
0.07372	0.01823	0.40683	0.23628	0.21855	0.1143	0.00919	0.08949
0.14519	0.39743	0.2487	0.02782	0.05975	0.2483	0.05183	0.2594
0.28668	0.194	0.06105	0.07217	0.5869	0.6019	0.03137	0.01333

Scatter Diagram



Autocorrelation Function for SERIE-11

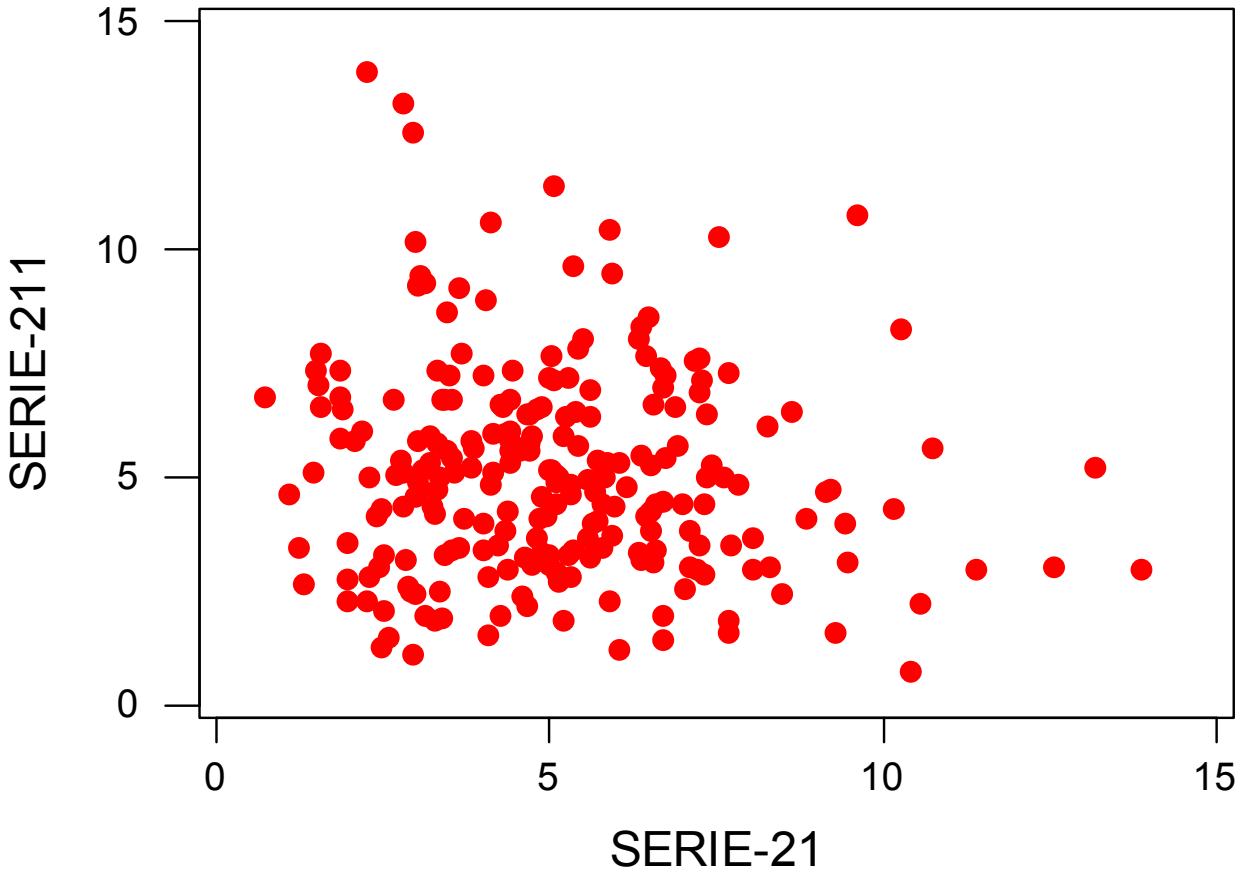


Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	-0.02	-0.29	0.08	13	-0.03	-0.51	16.52	25	0.05	0.76	23.22	37	-0.01	-0.13	37.94	49	0.01	0.16	43.06
2	-0.06	-0.86	0.83	14	-0.02	-0.33	16.65	26	0.07	0.93	24.36	38	0.02	0.31	38.08	50	0.02	0.32	43.23
3	-0.11	-1.68	3.72	15	-0.00	-0.00	16.65	27	0.03	0.43	24.61	39	0.03	0.42	38.35	51	-0.03	-0.35	43.44
4	-0.03	-0.44	3.93	16	-0.10	-1.47	19.30	28	-0.07	-1.02	26.02	40	0.08	1.12	40.32	52	0.07	0.90	44.82
5	0.10	1.51	6.36	17	0.02	0.26	19.39	29	-0.07	-0.93	27.21	41	-0.03	-0.36	40.53	53	0.02	0.24	44.92
6	-0.04	-0.59	6.73	18	-0.04	-0.62	19.87	30	-0.07	-1.04	28.72	42	0.03	0.44	40.84	54	0.13	1.68	49.83
7	-0.11	-1.63	9.65	19	-0.03	-0.44	20.12	31	0.02	0.23	28.79	43	0.05	0.63	41.48	55	-0.11	-1.43	53.51
8	-0.01	-0.13	9.67	20	-0.04	-0.62	20.60	32	0.09	1.31	31.25	44	-0.00	-0.05	41.48	56	-0.14	-1.87	59.89
9	-0.11	-1.71	12.99	21	0.03	0.42	20.83	33	-0.10	-1.41	34.15	45	-0.02	-0.24	41.58	57	0.00	0.05	59.89
10	0.07	1.03	14.21	22	0.05	0.73	21.51	34	-0.09	-1.21	36.32	46	0.01	0.18	41.63	58	-0.04	-0.49	60.35
11	-0.01	-0.19	14.26	23	-0.04	-0.53	21.88	35	-0.04	-0.61	36.88	47	0.07	0.91	43.01	59	0.06	0.82	61.64
12	0.09	1.28	16.21	24	-0.05	-0.66	22.44	36	0.06	0.82	37.91	48	0.01	0.08	43.02	60	-0.04	-0.48	62.09

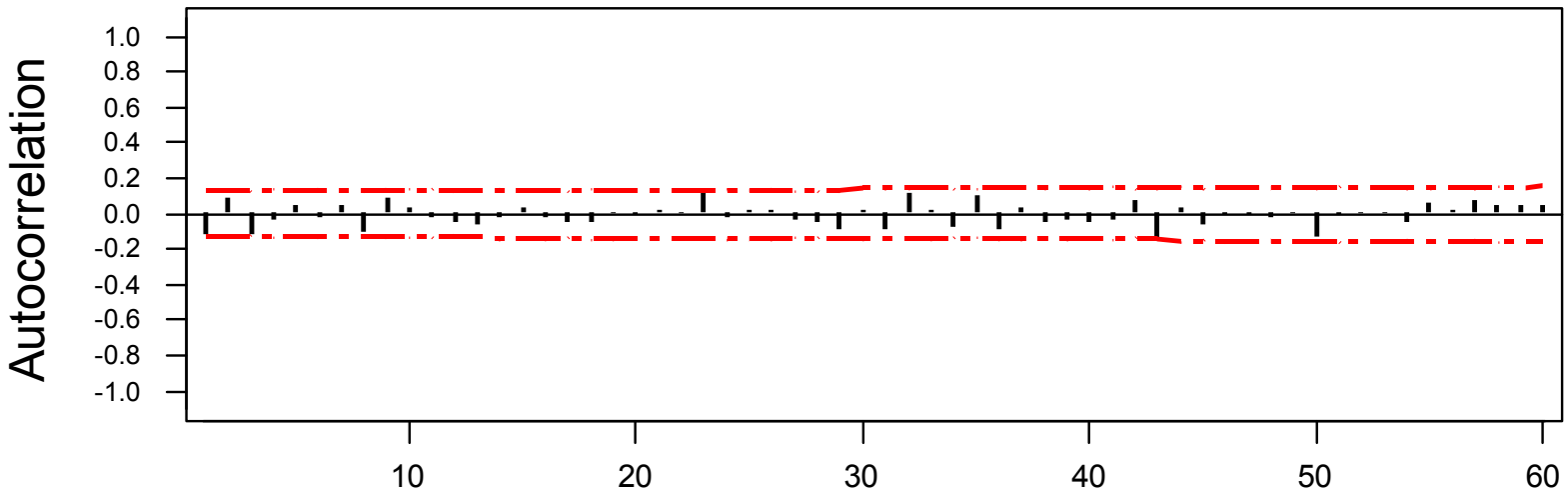
EXAMPLE 2: SERIE-31

			SERIE-21				
5.7361	5.0191	6.6069	5.901	13.1968	8.4939	6.734	12.5522
3.3176	7.361	6.5502	3.1915	2.8022	6.4716	1.8412	2.9525
3.4107	1.8527	4.297	6.3654	2.2988	4.8087	5.2173	11.4037
3.989	3.2645	2.4929	4.6792	1.9818	6.1355	3.8291	5.0474
4.0115	5.5972	2.9106	5.6774	4.242	8.2703	4.3219	5.1078
9.4219	3.4545	5.0961	6.9074	4.3622	10.2652	10.1441	1.4502
3.0595	5.8006	2.7903	5.6149	5.9716	7.5495	3.0024	6.6983
4.7313	4.3986	1.9646	10.7576	4.3459	7.1572	8.0385	4.3956
3.0448	5.0832	3.123	9.6188	2.8155	4.9944	5.4963	7.3253
2.429	3.5763	9.4745	5.3493	4.0832	2.2908	6.3751	4.4423
3.0007	1.9649	5.9328	2.7599	8.8646	2.2423	7.3577	6.6954
7.1995	6.6839	4.1527	5.1347	4.0421	10.5711	1.4953	3.3899
5.29	3.5226	6.4526	4.9969	5.7195	4.1281	2.5926	5.3341
6.528	7.2281	5.3892	3.3664	4.7115	4.947	2.8874	5.8627
4.8937	3.4807	5.7089	4.8745	3.3102	5.5796	7.3171	1.8652
3.0332	1.2211	5.4296	5.1127	2.5292	4.414	3.2987	7.6942
7.1155	6.028	6.7492	4.1474	7.0334	5.7712	5.2836	1.571
5.0654	2.1948	0.7439	2.4136	1.5366	2.0787	5.7805	9.275
2.7094	4.6554	10.4125	4.5931	4.0861	2.4993	3.8286	3.1368
5.1475	9.145	5.9085	4.8656	4.8544	3.3592	6.4999	6.5432
3.1049	3.6576	4.7446	7.8333	5.3092	6.3346	1.8961	1.5726
4.9994	5.5672	9.2035	5.4189	3.2033	5.2448	3.3814	7.6669
7.5988	4.5665	3.0135	3.5154	2.8244	7.414	3.5257	6.45
7.2527	2.9935	8.2895	7.7041	5.3012	6.6798	4.2098	8.6338
3.9954	4.3677	6.3833	3.6804	4.4175	3.4224	3.2878	3.4547
5.6483	3.2517	4.6479	4.8174	6.9775	6.5766	4.9934	3.6475
3.8479	4.6064	1.1035	4.1176	6.6816	4.259	5.138	8.0391
7.1137	5.3006	2.955	3.7168	2.6666	6.5238	5.0206	6.3169
7.2907	6.0291	13.8861	5.9202	1.3012	6.8827	5.8092	5.608
7.6797	4.4036	2.2735	5.2135	2.4592	7.2322	3.0161	4.6935

Scatter Diagram



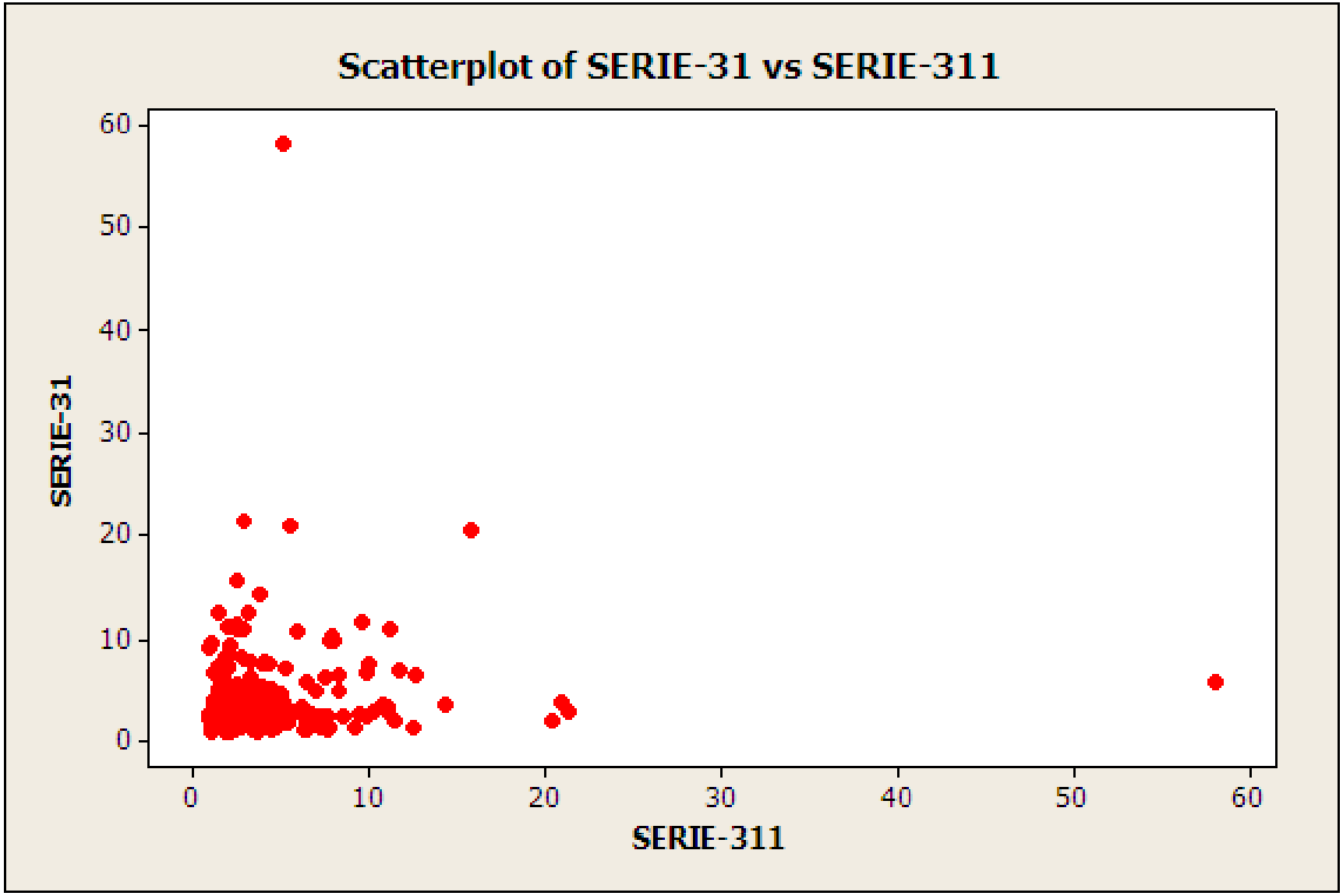
Autocorrelation Function for SERIE-21

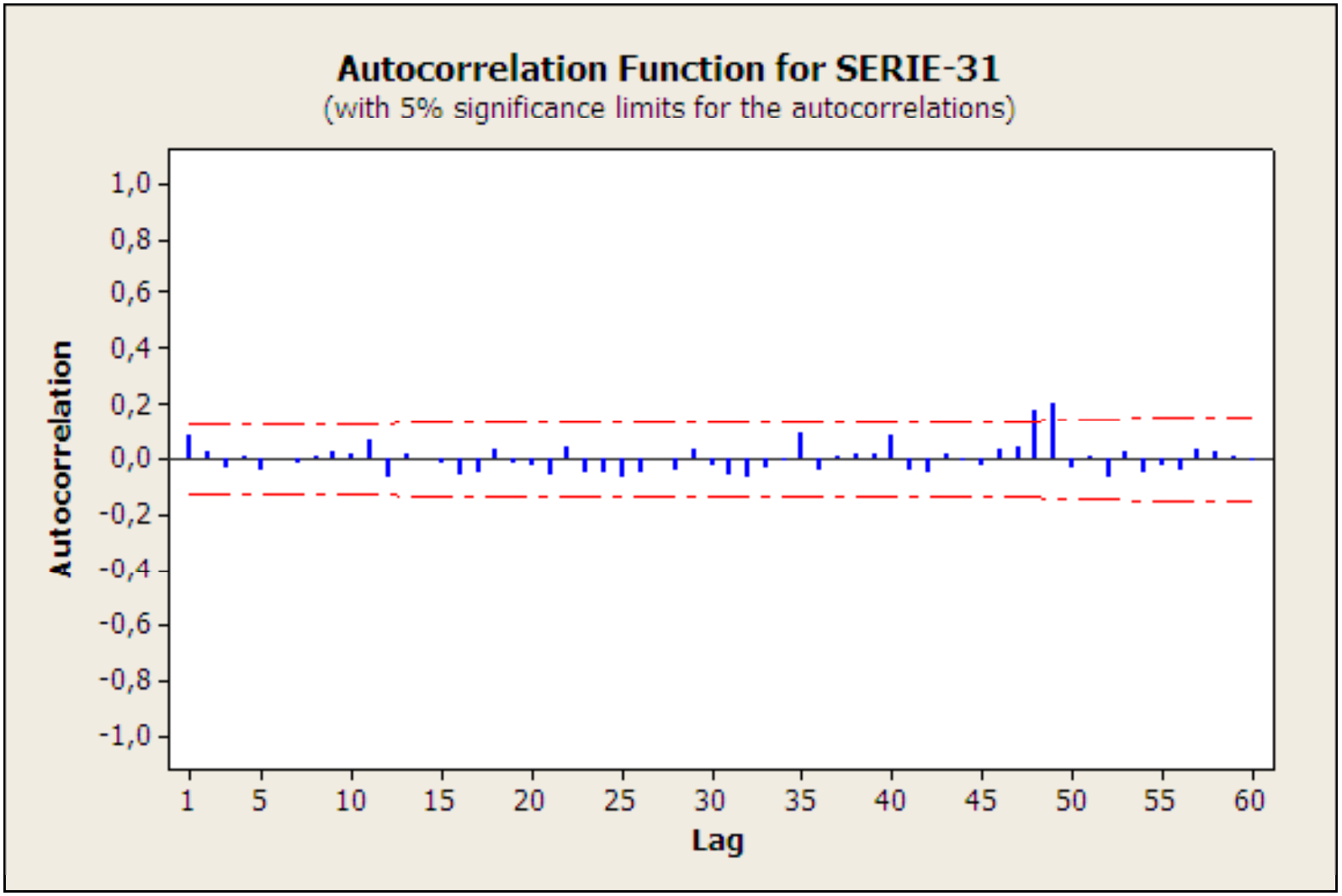


Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	-0.11	-1.72	3.00	13	-0.06	-0.94	16.63	25	0.03	0.38	22.39	37	0.03	0.43	39.53	49	0.00	0.03	51.58
2	0.09	1.37	4.97	14	-0.02	-0.31	16.75	26	0.02	0.30	22.51	38	-0.05	-0.63	40.15	50	-0.13	-1.65	56.39
3	-0.12	-1.81	8.45	15	0.04	0.51	17.06	27	-0.03	-0.46	22.79	39	-0.03	-0.47	40.50	51	-0.01	-0.11	56.42
4	-0.04	-0.60	8.84	16	-0.02	-0.32	17.19	28	-0.05	-0.76	23.57	40	-0.05	-0.64	41.14	52	0.00	0.01	56.42
5	0.04	0.63	9.27	17	-0.05	-0.68	17.77	29	-0.09	-1.32	25.95	41	-0.03	-0.44	41.45	53	0.00	0.04	56.42
6	-0.02	-0.35	9.41	18	-0.04	-0.62	18.25	30	0.02	0.31	26.08	42	0.08	1.03	43.13	54	-0.04	-0.58	57.04
7	0.05	0.75	10.03	19	0.01	0.18	18.29	31	-0.09	-1.21	28.12	43	-0.15	-2.04	49.90	55	0.06	0.77	58.14
8	-0.11	-1.57	12.80	20	-0.01	-0.10	18.31	32	0.12	1.73	32.38	44	0.04	0.50	50.31	56	0.02	0.29	58.29
9	0.08	1.24	14.58	21	0.01	0.21	18.36	33	0.01	0.21	32.44	45	-0.06	-0.82	51.46	57	0.08	1.03	60.29
10	0.04	0.59	14.99	22	0.01	0.13	18.38	34	-0.08	-1.07	34.13	46	-0.01	-0.08	51.47	58	0.05	0.65	61.10
11	-0.02	-0.24	15.06	23	0.12	1.69	22.03	35	0.10	1.42	37.16	47	-0.01	-0.16	51.51	59	0.05	0.64	61.89
12	-0.04	-0.65	15.56	24	-0.02	-0.36	22.20	36	-0.09	-1.17	39.24	48	-0.02	-0.20	51.58	60	0.05	0.63	62.68

EXAMPLE 3: SERIE-31

SERIE 31							
9,79238086	3,39248633	3,85402491	1,90258714	3,06976484	8,46533444	4,31480957	2,48259365
6,68580989	1,11937393	5,34727032	2,32068584	4,38753273	2,50982585	3,40125673	5,4547351
2,69148801	3,53598322	2,84476121	2,01364062	5,13861991	2,09936415	4,68471037	2,9633303
2,83795744	3,25394346	2,2409955	1,63727606	58,0468904	1,28086227	4,50255541	2,15559484
2,10140362	6,33841869	2,64853502	3,93496453	5,86446746	2,42011945	1,81901052	3,01820006
0,89292225	1,22096092	5,08075741	2,01676859	10,7662605	1,74980541	1,9603977	2,84411193
2,70881212	2,78819656	3,79873788	1,61612691	3,64230984	2,14604077	7,4626681	11,0763426
4,33330032	2,36499398	1,14665519	2,41448238	1,70902836	2,44919207	6,30723462	3,39910042
2,02440266	2,64565448	2,55656051	1,19398638	1,4353896	10,984561	1,14236814	3,25566982
9,4503185	3,1337218	3,91616503	3,93034911	7,25269811	3,19941565	1,31331457	1,86610631
2,64312602	1,65727213	7,62439196	2,47126754	1,17716002	7,77550344	3,42084176	11,1952095
2,61528043	1,94334997	2,36726593	11,3757102	2,18887089	9,78287203	3,1717982	10,9769257
1,29081515	7,33347608	1,13005733	1,90404052	2,46025487	2,37891838	12,4602235	2,76331961
4,17440249	1,87874175	2,47107728	3,9736257	3,23641162	3,80285998	1,24497787	1,26386512
1,82689321	3,58996872	2,85779348	2,70500112	5,31722532	1,99165978	3,85923629	1,36883619
1,39426144	1,83819955	7,94811796	8,25933652	1,64643074	1,35325069	1,73625953	1,55807549
1,71633695	1,85730572	9,92515957	6,45343815	2,37905114	6,40810427	1,971637	1,82077954
3,99856237	2,89198853	7,59644796	0,94760782	0,97444619	5,7766491	1,93883983	2,64598078
3,62402545	21,3112177	1,02211793	1,34879682	0,87312188	2,90689608	3,34231849	4,28227596
1,99825507	2,89281644	2,28478915	5,179849	2,13001808	2,91671831	1,63482789	7,69484184
1,37554261	3,79717092	3,6445635	7,08705049	2,11484313	2,4562488	5,50148032	1,16906102
2,51736791	3,81615513	1,97226231	1,84490859	4,82732968	15,7414733	20,9416732	6,7590078
5,1460402	14,3731351	1,38283271	8,21976136	4,71808058	20,390198	3,72579055	1,70939843
1,84756529	3,61184366	12,5906118	4,818646	1,31210639	1,9567882	1,05767598	2,75274251
0,9087296	0,92602806	6,35506688	4,05037532	2,99824938	7,18740676	9,57416884	1,41857286
9,11042665	2,45353161	1,64629926	7,81383314	4,11875192	2,30263886	11,6657984	6,12751461
1,23335251	2,25626652	6,06110448	10,2861503	4,39999267	3,63885231	6,94612822	3,25932895
1,96718497	4,54615784	2,95581249	2,77365266	1,10674577	3,28828788	4,94274274	1,21823087
1,06893998	2,14835343	2,07800633	2,25504004	1,31418536	1,91884905	4,70513733	1,83450638
1,53362506	1,1258206	2,14005427	1,90029756	1,99507052	2,96052798	2,59336284	6,87393724

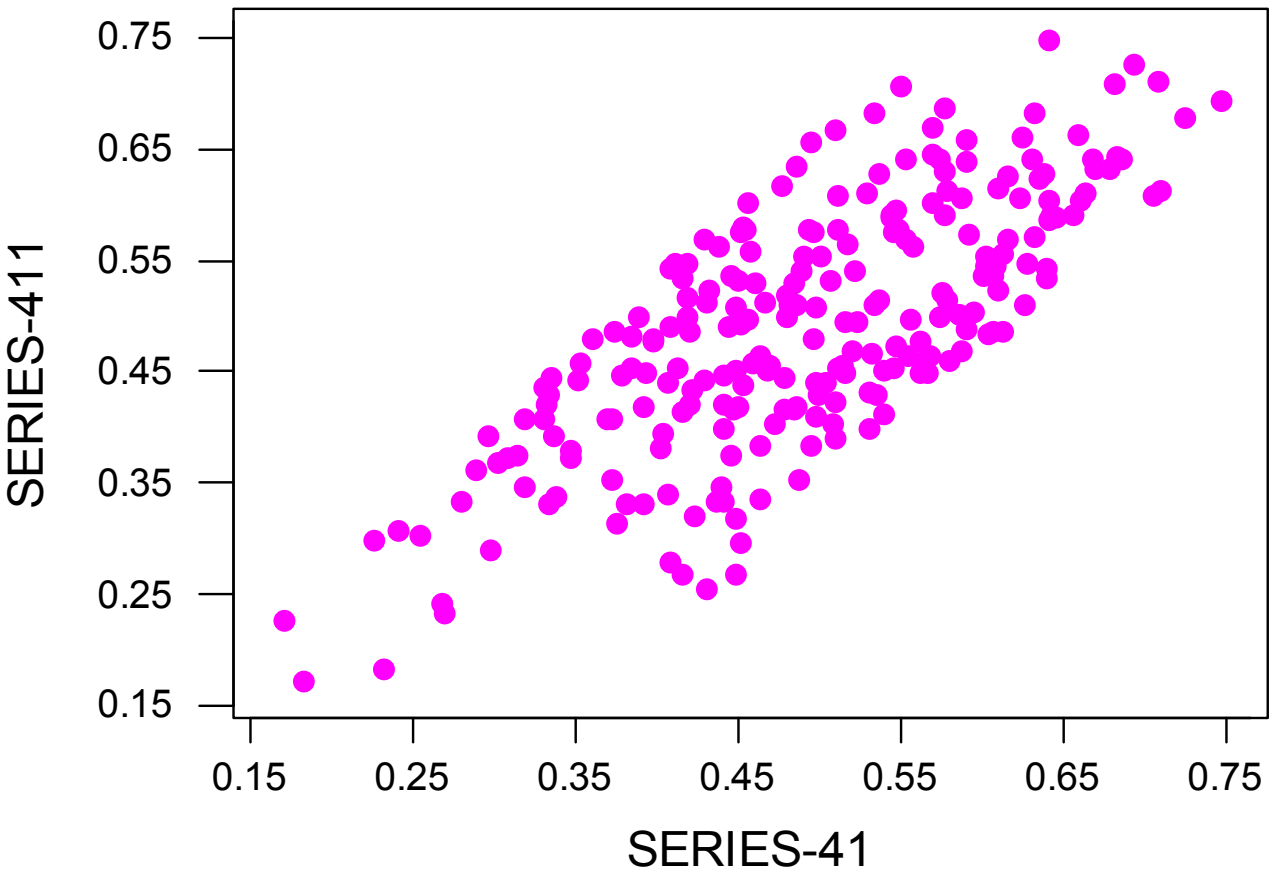

Scatterplot of SERIE-31 vs SERIE-311
SERIE-31
SERIE-311
0
10
20
30
40
50
60


Autocorrelation Function for SERIE-31
(with 5% significance limits for the autocorrelations)
Autocorrelation
1,0
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,8
-1,0
1
5
10
15
20
25
30
35
40
45
50
55
60
Lag

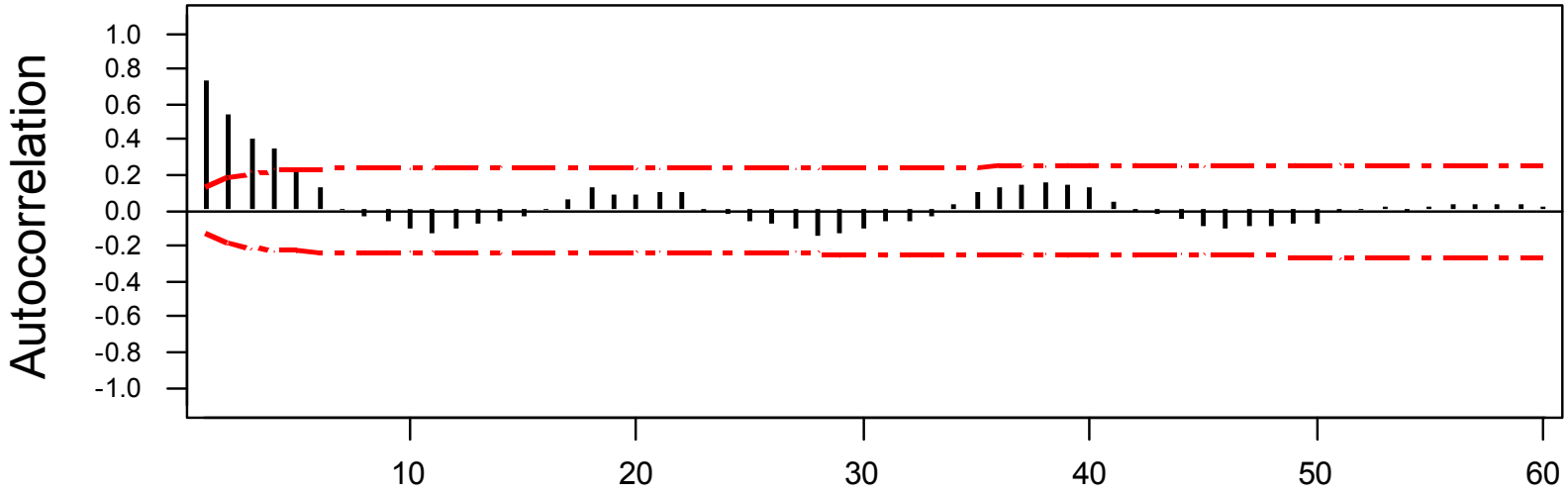
EXAMPLE 4: SERIE-41

			SERIE-41				
0,334212	0,450114	0,577721	0,594755	0,441379	0,638671	0,568783	0,669429
0,43639	0,468004	0,492714	0,547033	0,35176	0,589803	0,552942	0,569291
0,331024	0,520025	0,45176	0,627713	0,37262	0,576383	0,499952	0,42913
0,381603	0,575033	0,539437	0,536883	0,346867	0,454603	0,585526	0,535064
0,402386	0,496259	0,521885	0,445136	0,318639	0,513624	0,640461	0,600902
0,507436	0,456328	0,432649	0,335753	0,449103	0,578011	0,685819	0,455715
0,448208	0,353148	0,421402	0,463451	0,393266	0,511334	0,577392	0,468781
0,562395	0,486963	0,332003	0,567865	0,403727	0,466652	0,547856	0,587942
0,437899	0,590306	0,333288	0,616243	0,47303	0,532253	0,410941	0,645319
0,45251	0,65647	0,440242	0,476857	0,546477	0,506671	0,539781	0,56984
0,384238	0,494462	0,40689	0,397446	0,419187	0,497777	0,489246	0,631601
0,463746	0,523436	0,372177	0,441195	0,391387	0,479453	0,407814	0,677864
0,463045	0,609734	0,30762	0,428444	0,336538	0,398401	0,369002	0,724558
0,553819	0,529745	0,240847	0,334048	0,338685	0,529808	0,302012	0,692981
0,602848	0,459713	0,268165	0,27943	0,406344	0,484572	0,254047	0,747507
0,659932	0,580006	0,415445	0,407411	0,319101	0,604185	0,429837	0,640497
0,625223	0,453518	0,447186	0,330954	0,423473	0,640437	0,498753	0,552428
0,615325	0,412728	0,440303	0,392259	0,509701	0,667708	0,418806	0,490673
0,609658	0,416155	0,497902	0,296674	0,480814	0,50956	0,486165	0,443395
0,6633	0,477736	0,389028	0,452156	0,383957	0,533985	0,613138	0,478842
0,658168	0,562269	0,509438	0,544985	0,494361	0,414908	0,709397	0,360787
0,590229	0,557615	0,485256	0,608307	0,515335	0,48495	0,708074	0,288815
0,543479	0,457891	0,374532	0,705775	0,418385	0,60668	0,681009	0,297553
0,639639	0,458856	0,313412	0,550226	0,450431	0,587672	0,533383	0,225866
0,629955	0,564141	0,375002	0,605137	0,448754	0,543333	0,639648	0,171086
0,576045	0,517568	0,445898	0,623348	0,514896	0,408464	0,573701	0,182909
0,452018	0,479753	0,378776	0,634882	0,536106	0,498401	0,591146	0,232215
0,511164	0,49621	0,346836	0,486374	0,607184	0,574182	0,641769	0,268707
0,431065	0,555504	0,440087	0,419422	0,510562	0,545257	0,682373	0,44779
0,531199	0,612388	0,50399	0,419372	0,626422	0,601665	0,63199	0,566805

Scatter Diagram



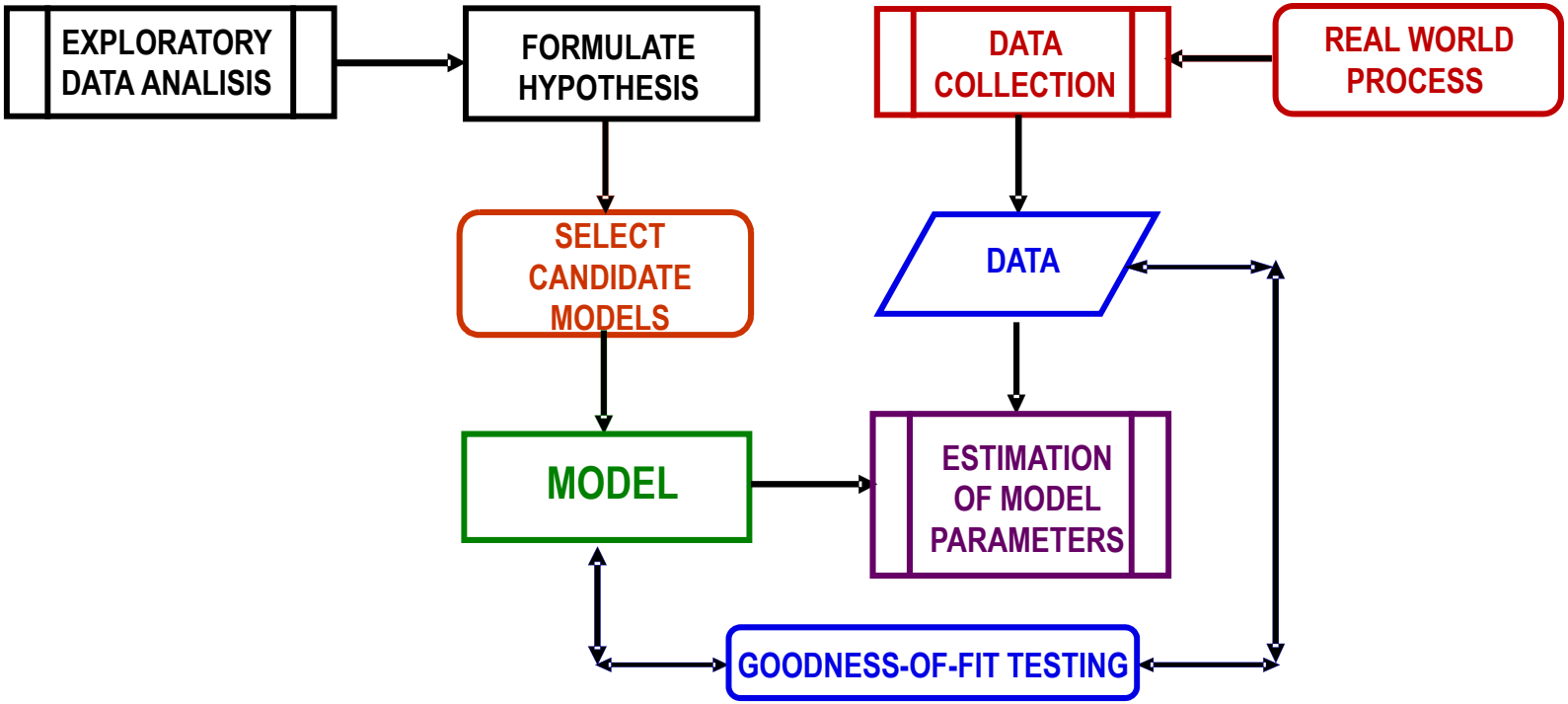
Autocorrelation Function for SERIES-41



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.75	11.55	135.05	13	-0.08	-0.67	310.48	25	-0.06	-0.46	327.68	37	0.14	1.13	363.52	49	-0.07	-0.55	395.91
2	0.55	5.86	208.79	14	-0.06	-0.51	311.47	26	-0.07	-0.59	329.10	38	0.16	1.27	371.09	50	-0.07	-0.55	397.51
3	0.41	3.88	250.54	15	-0.03	-0.28	311.76	27	-0.11	-0.85	332.12	39	0.15	1.16	377.52	51	-0.01	-0.10	397.56
4	0.35	3.10	280.70	16	0.01	0.05	311.77	28	-0.15	-1.17	337.95	40	0.13	1.02	382.55	52	0.00	0.03	397.57
5	0.23	1.95	293.59	17	0.06	0.46	312.58	29	-0.13	-1.08	342.95	41	0.05	0.36	383.19	53	0.02	0.12	397.64
6	0.14	1.13	298.12	18	0.13	1.06	316.93	30	-0.10	-0.81	345.78	42	-0.00	-0.03	383.19	54	-0.01	-0.09	397.69
7	0.01	0.09	298.15	19	0.08	0.69	318.78	31	-0.07	-0.55	347.10	43	-0.02	-0.15	383.31	55	0.02	0.17	397.85
8	-0.04	-0.30	298.48	20	0.09	0.76	321.06	32	-0.06	-0.48	348.14	44	-0.05	-0.37	384.00	56	0.04	0.28	398.29
9	-0.06	-0.47	299.28	21	0.10	0.84	323.85	33	-0.03	-0.26	348.45	45	-0.09	-0.73	386.69	57	0.03	0.26	398.65
10	-0.11	-0.92	302.35	22	0.10	0.83	326.64	34	0.04	0.29	348.83	46	-0.10	-0.80	389.98	58	0.04	0.27	399.04
11	-0.13	-1.04	306.34	23	-0.00	-0.03	326.64	35	0.11	0.86	352.14	47	-0.08	-0.65	392.13	59	0.03	0.24	399.36
12	-0.10	-0.81	308.79	24	-0.02	-0.20	326.80	36	0.14	1.09	357.58	48	-0.08	-0.65	394.30	60	0.02	0.13	399.46

STRATEGIES FOR SELECTING THE PROBABILITY DISTRIBUTION

- Formulate hypothesis based on the descriptive data analysis: Histograms, Box-Plots….
- Select the most suitable candidate models according to the assumptions
- Estimate the parameters of the selected models
- Estimate the quality of the fitness: Q and P Plots, χ^2, K-S,…



PRACTICAL RULES-I

- Coefficient of variation of a random variable X: $C_X = \dfrac{\sigma_X}{\mu_X}$
- Can be interpreted as a measure of the deviation of X with respect to the exponential distribution, $C_X=1$

A. Approximation of distributions whose coefficient of variation is significantly smaller than 1, $C_X<1$.

Family of Gamma probability distributions

$$f_X(x) = \begin{cases} \dfrac{e^{-x/\beta}}{\Gamma(\alpha)} \beta^{-\alpha} x^{\alpha-1} & \text{for } x \geq 0 \\ 0 & \text{for } x = 0 \end{cases} \qquad \Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-\alpha} dy$$

$$\alpha \in R_+, \beta \in R_+ \qquad \alpha = k \in Z_+ \Rightarrow \Gamma(k+1) = k!$$

α shape parameter, β scale parameter. $F_X(x)$ has closed form only if α is integer:

$$F_X(x) = \begin{cases} 1 - e^{-x/\beta} \displaystyle\sum_{j=0}^{\alpha-1} \frac{\left(x/\beta\right)^j}{j!} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

PRACTICAL RULES-II

- **Particular case: when α =1 the distribution $\Gamma(1) \equiv$ exponential with parameter 1/β.**

$$f_X(x) \rightarrow \begin{cases} \text{Mean } \bar{x} = \alpha\beta \\ \text{Variance } \sigma^2 = \alpha\beta^2 \end{cases} \Rightarrow C_X = \frac{\sqrt{\alpha}}{\alpha} < 1$$

- **For α = k integer and β=1/(kμ), Γ(k, 1/(kμ)) is an Erlang distribution with parameter k and mean 1/μ, sum of k exponential distributions I.I.D. with mean 1/(kμ),**

$$\left.\begin{aligned} f_X(x) &= \frac{k\mu(k\mu)^{k-1}x^{k-1}}{(k-1)!}\left(e^{-k\mu}\right)^x \\ F_X(x) &= 1-\left(e^{-k\mu}\right)^x \sum_{j=0}^{k-1}\frac{(k\mu)^j x^j}{j!} \end{aligned}\right\} x \geq 0;\ \text{Mean } \frac{1}{\mu};\ \text{Variance } \frac{1}{k\mu^2}$$

PRACTICAL RULES-III

B. Approximation of distributions with coefficient of variation is significantly greater than 1: C_X>1.

Example: Hiperexponential distribution (Combination of exponential distributions)

$$f_X(x) = \sum_{i=1}^{k} \pi_i \mu_i e^{-\mu_i x} \qquad \left(\pi_i \geq 0, \forall i, \sum_{i=1}^{k} \pi_i = 1 \right)$$

$$F_X(x) = \sum_{i=1}^{k} \pi_i \left(1 - e^{-\mu_i x}\right) = 1 - \sum_{i=1}^{k} \pi_i e^{-\mu_i x} \qquad x \geq 0$$

Mean **Variance :**

$$\bar{x} = \sum_{i=1}^{k} \frac{\pi_i}{\mu_i} = \frac{1}{\mu} \qquad \sigma_X^2 = \frac{1}{\mu^2} + \sum_{i=1}^{k} \sum_{j=1}^{k} \pi_i \pi_j \left(\frac{1}{\mu_i} - \frac{1}{\mu_j} \right)^2$$

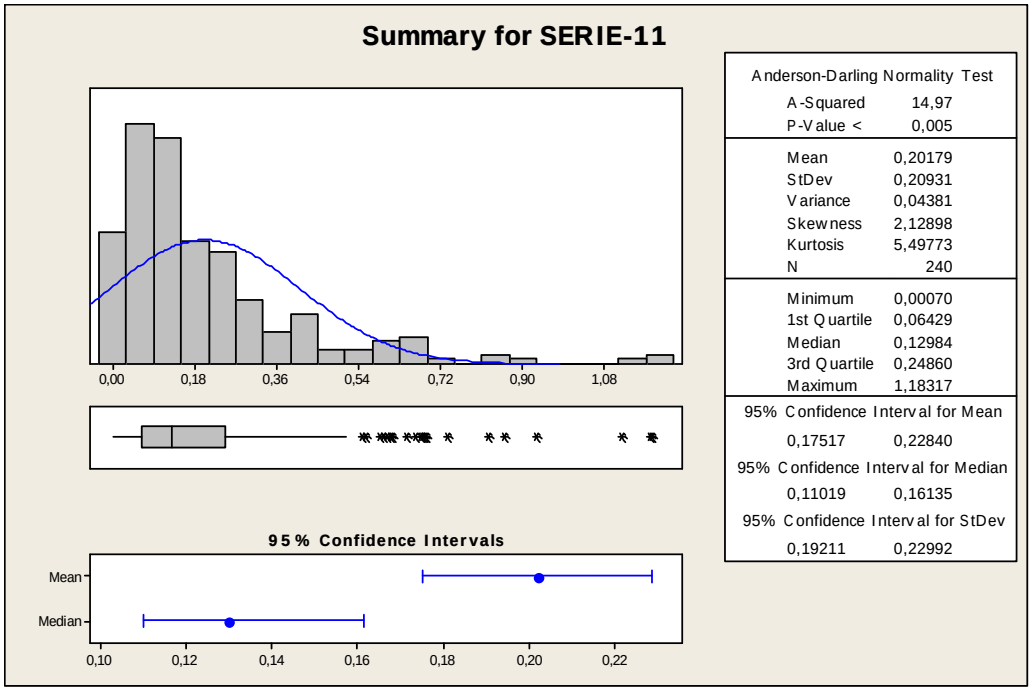
Coefficient of Variation

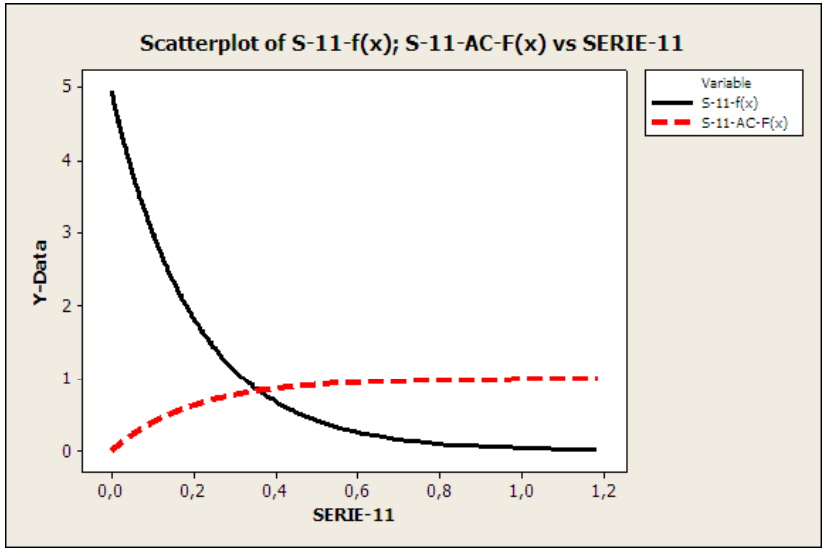
$$C_X^2 = 1 + \mu^2 \sum_{i=1}^{k} \sum_{j=1}^{k} \pi_i \pi_j \left(\frac{1}{\mu_i} - \frac{1}{\mu_j} \right)^2 \geq 1 \quad \left(= 1 \Leftrightarrow \mu_i = \mu_j \forall i, j \right)$$

SUMMARY ANALYSIS OF SERIE 11

- **Serie 11**
  - **Mean** $\mu = 0.20179$
  - **Standard deviation** $\sigma = 0.20931$
  - **Coefficient of Variation** $C_X = 1.03726 \approx 1$

$\Rightarrow$ Significantly close to 1: hypothesis the distribution is exponential

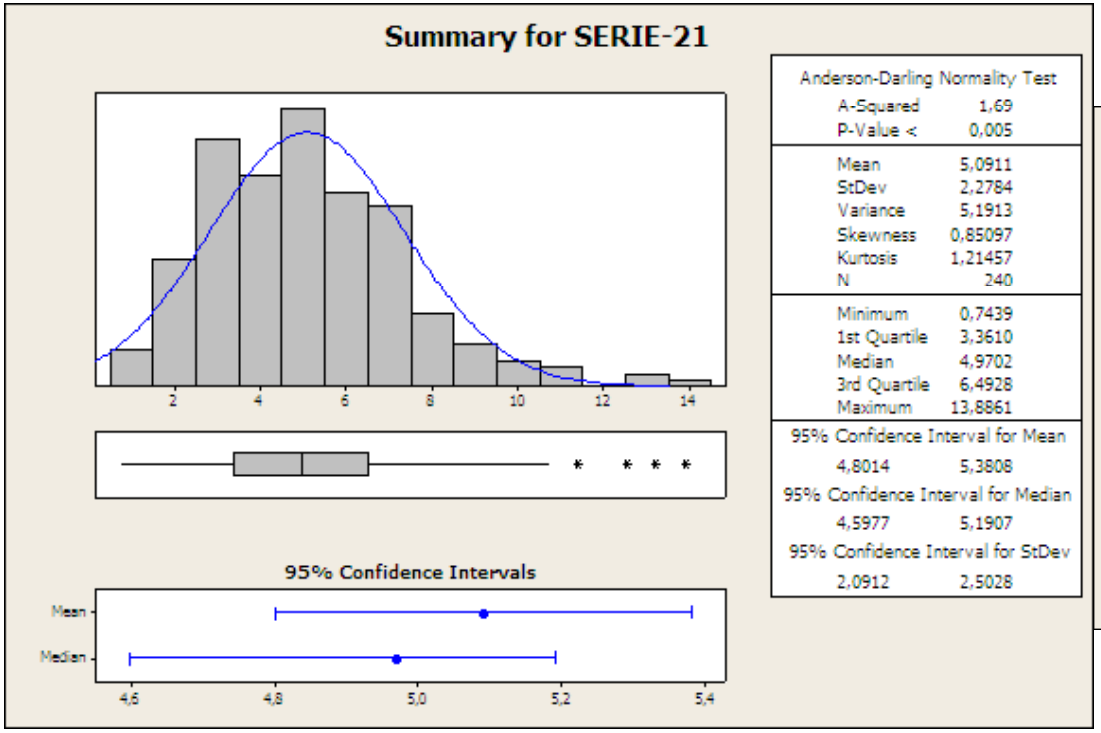


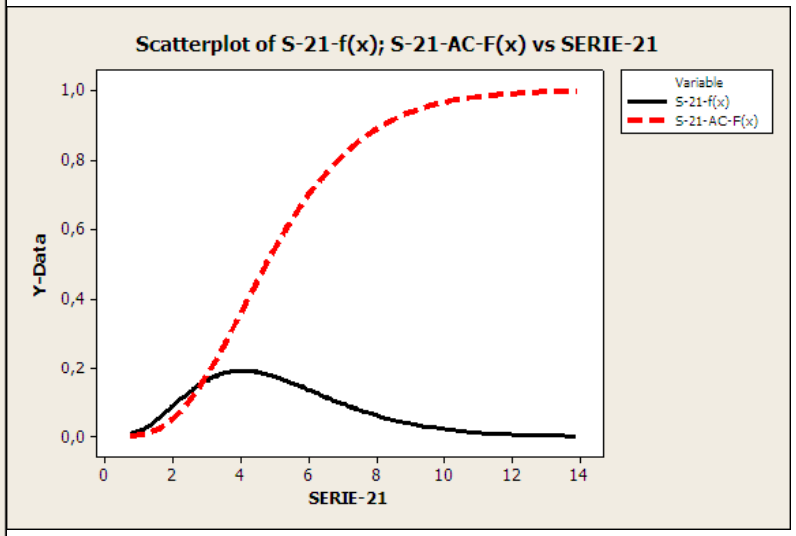


SUMMARY ANALYSIS OF SERIE 21

- **Serie 21**
  - **Mean** $\mu = 5.0911$
  - **Standard deviation** $\sigma = 2.2784$
  - **Coefficient of Variation** $C_X = 0.44752 < 1$

$\Rightarrow$ **Significantly smaller than 1: hypothesis distribution of the Gamma family**

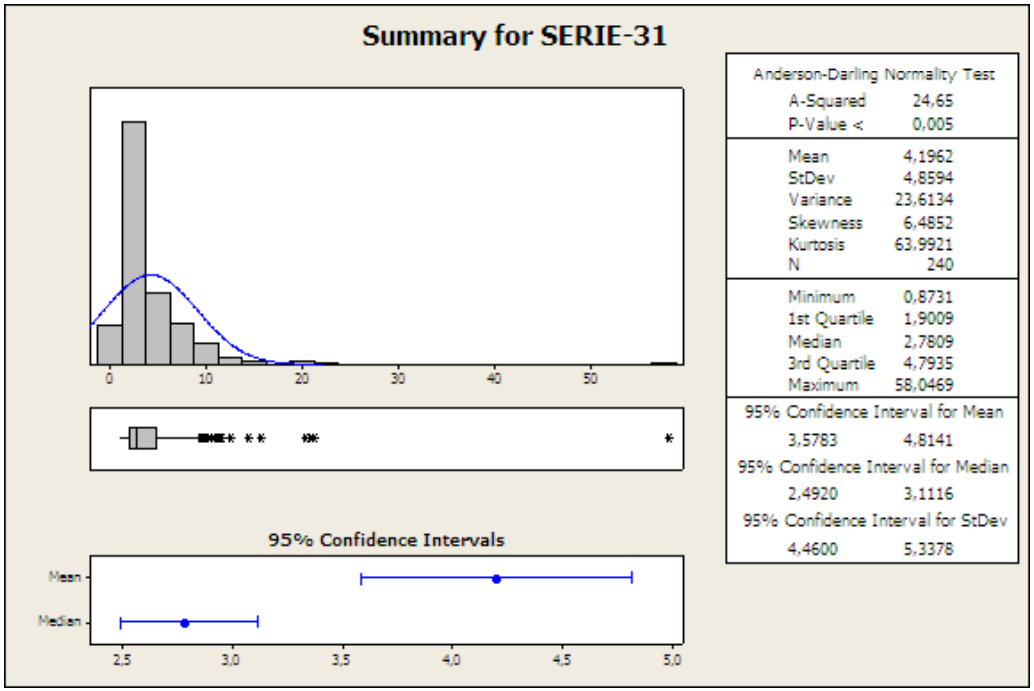




SUMMARY ANALYSIS OF SERIE 31

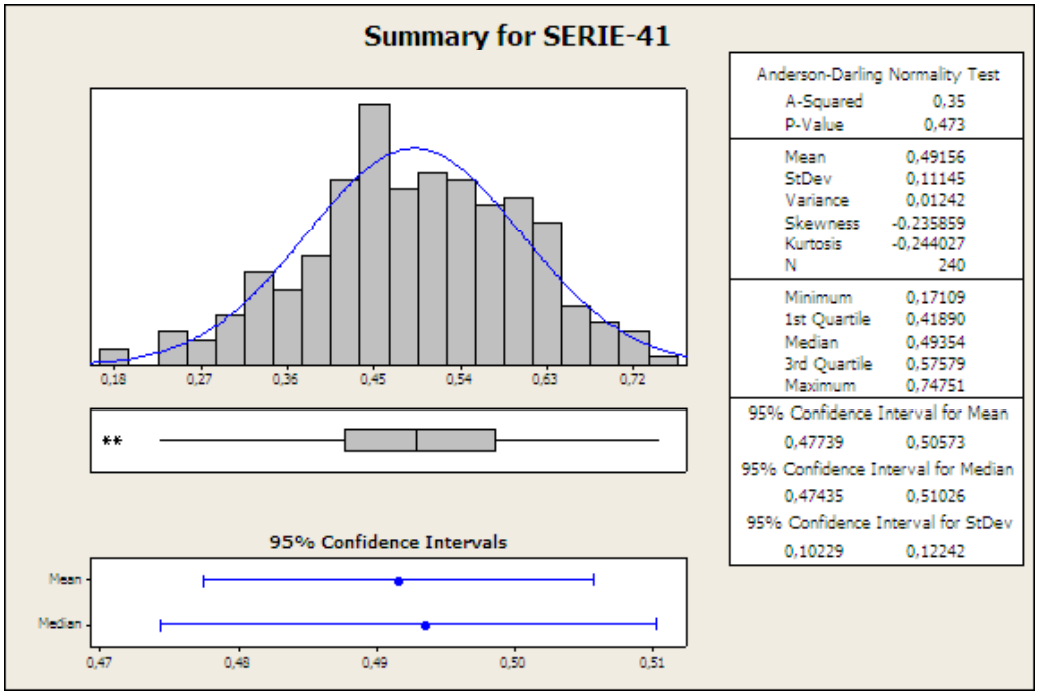
- **Serie 31**
  - **Mean** $\mu = 4.1962$
  - **Standard deviation** $\sigma = 4.8594$
  - **Coefficient of variation** $C_X = 1.15804 > 1$

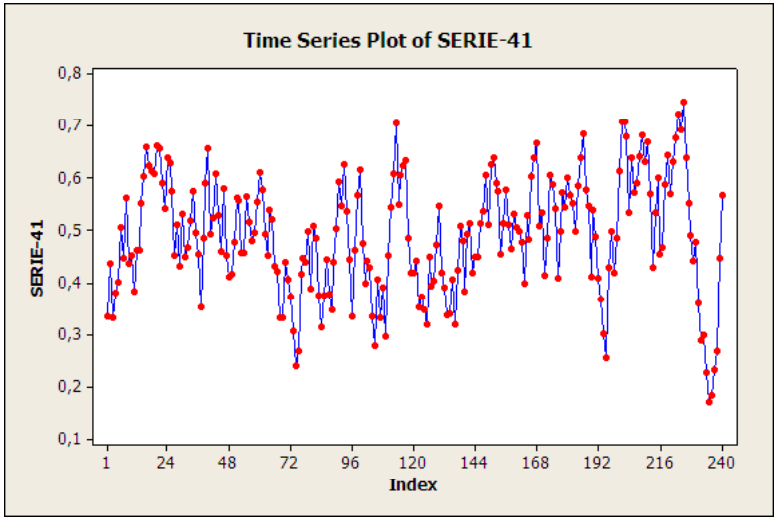
(Pearson of Type 5 with parameters α=2.5 and β=6.666)



SUMMARY ANALYSIS OF SERIE 41

- **Serie 41**
  - **Mean** $\mu = 0.49156$
  - **Standard deviation** $\sigma = 0.11145$
  - **Variation coefficient** $C_X = 0.22672$
  - **Correlated**





ESTIMATION OF PARAMETERS OF THE CANDIDATE DISTRIBUTIONS FROM OBSERVED DATA

- **Hypothesis:** The observed data $X_1, X_2, \ldots, X_n$, are I.I.D.
- **Method:** Maximum-likelihood estimators
  - Let $P_\theta(x)$ be the probability density function which depends on parameter θ that we wish to estimate from the already observed data $X_1, X_2, \ldots, X_n$, which we assume are I.I.D.
  - The Maximum Likelihood Function $L(\theta)$ is defined as:

$$L(\theta) = P_\theta(X_1)\, P_\theta(X_2) \ldots P_\theta(X_n)$$

  - The Maximum Likelihood Estimator $\hat{\theta}$ of parameter θ is defined to be the value that maximizes $L(\theta)$ over all admissible values of θ ". It is the value that "explains better" the observed data.

$$\hat{\theta} \Leftrightarrow \frac{dL(\theta)}{d\theta} = 0$$

- **EXAMPLE 1**: The exponential distribution: $\theta = \mu > 0, f_\mu(x) = \frac{1}{\mu} e^{-x/\mu}, x \geq 0$
- Likelihood Function:

$$L(\mu) = \left(\frac{1}{\mu} e^{-x_1/\mu}\right)\left(\frac{1}{\mu} e^{-x_2/\mu}\right)\ldots\left(\frac{1}{\mu} e^{-x_n/\mu}\right) = \frac{1}{\mu^n} \exp\left(-\frac{1}{\mu}\sum_{i=1}^{n} x_i\right)$$

- To determine the value $\hat{\mu}$ that maximizes $L(\mu)$ for $\mu \geq 0$ it is more convenient to take logarithms and redefine the likelihood function as:

$$l(\mu) = \ln[L(\mu)] = -n\ln\mu - \frac{1}{\mu}\sum_{i=1}^{n} x_i$$

- And as it is strictly increasing: $\underset{\mu}{\text{MAX}}\, L(\mu) \equiv \underset{\mu}{\text{MAX}}\, l(\mu)$

$$\frac{dl(\mu)}{d\mu} = -\frac{n}{\mu} + \frac{1}{\mu^2}\sum_{i=1}^{n} x_i \Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- And furthermore: $\frac{dl^2(\mu)}{d\mu^2} = \frac{n}{\mu^2} - \frac{2}{\mu^3}\sum_{i=1}^{n} x_i < 0$ for $\mu = \hat{\mu}$

- **EXAMPLE 2**: The Gamma $\Gamma(\alpha,\beta)$ distribution:

$$f_X(x) = \begin{cases} \dfrac{e^{-x/\beta}}{\Gamma(\alpha)}\beta^{-\alpha}x^{\alpha-1} & \text{for } x \geq 0 \\ 0 & \text{for } x = 0 \end{cases} \qquad \Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-\alpha}dy$$

$$\alpha \in R_+, \beta \in R_+ \qquad \alpha = k \in Z_+ \Rightarrow \Gamma(k+1) = k!$$

- The Maximum Likelihood Function is:

$$L(\alpha,\beta) = \frac{\beta^{-n\alpha}\left(\prod_{i=1}^{n} X_i\right)^{\alpha-1} \exp\left(-\frac{1}{\beta}\sum_{i=1}^{n} X_i\right)}{[\Gamma(\alpha)]^n}$$

and consequently, the values $\hat{\alpha}$ and $\hat{\beta}$ that maximize the Maximum Likelihood Function, are the solution of the nonlinear system of equations :

$$\frac{\partial L(\alpha,\beta)}{\partial \alpha} = 0, \quad \frac{\partial L(\alpha,\beta)}{\partial \beta} = 0$$

- That means

$$\ln\hat{\beta} + \Psi(\hat{\alpha}) = \frac{\sum_{i=1}^{n} \ln X_i}{n}, \qquad \hat{\alpha}\hat{\beta} = \overline{X}(n) = \frac{\sum_{i}^{n} X_i}{n}$$

where $\Psi(\hat{\alpha}) = \dfrac{\Gamma'(\alpha)}{\Gamma(\alpha)}$ is the digamma function

- Approximate solutions for $\hat{\alpha}$ and $\hat{\beta}$ can be found calculating:

$$T = \left[\ln\left(\overline{X}(n)\right) - \frac{\sum_{i=1}^{n} \ln X_i}{n} \right]^{-1}$$

and estimating $\hat{\alpha}$ as a function of T a from the tables (Law and Kelton, Table 6.19) an then:

$$\hat{\beta} = \frac{\overline{X}(n)}{\hat{\alpha}}$$

- **EXAMPLE 3: Weibull distribution W(α,β)**
- **Functions probability density $f_X(x)$ and distribution $F_X(x)$:**

$$f_X(x) = \begin{cases} \alpha\beta^{-\alpha}x^{\alpha-1}e^{-\left(x/\beta\right)^{\alpha}} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = \begin{cases} 1 - e^{-\left(x/\beta\right)^{\alpha}} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- **Shape parameter $\alpha > 0$, scale parameter $\beta > 0$**

$$\text{Mean } \frac{\beta}{\alpha}\Gamma\left(\frac{1}{\alpha}\right) \qquad \text{Variance } \frac{\beta^2}{\alpha}\left\{2\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha}\left[\Gamma\left(\frac{1}{\alpha}\right)\right]^2\right\}$$

- **The MLE requires to solve the nonlinear system of equations:**

$$\frac{\sum_{i=1}^{n} X_i^{\hat{\alpha}} \ln X_i}{\sum_{i=1}^{n} X_i^{\hat{\alpha}}} - \frac{1}{\hat{\alpha}} = \frac{\sum_{i=1}^{n} \ln X_i}{n}, \quad \text{y} \quad \hat{\beta} = \left(\frac{\sum_{i=1}^{n} X_i^{\hat{\alpha}}}{n}\right)^{\frac{1}{\hat{\alpha}}}$$

- The first equation can be solved numerically to obtain $\hat{\alpha}$ by applying the Method of Newton. The general recursive step at each Newton iteration is:

$$\hat{\alpha}_{k+1} = \hat{\alpha}_k + \frac{A + \left(1/\hat{\alpha}_k\right) - \left(C_k/B_k\right)}{\left(1/\hat{\alpha}_k^2\right) + \left[\left(B_k H_k - C_k^2\right)/B_k^2\right]}$$

- where

$$A = \frac{\sum_{i=1}^{n} \ln X_i}{n}, B_k = \sum_{i=1}^{n} X_i^{\hat{\alpha}_k}, C_k = \sum_{i=1}^{n} X_i^{\hat{\alpha}_k} \ln X_i, H_k = \sum_{i=1}^{n} X_i^{\hat{\alpha}_k} \left(\ln X_i\right)^2$$

- The experience recommends to use as initial value in the iterative process:

$$\hat{\alpha}_0 = \left\{ \frac{\frac{6}{\pi^2}\left[\sum_{i=1}^{n} (\ln X_i)^2 - \left(\sum_{i=1}^{n} \ln X_i\right)^2 \Big/ n\right]}{n-1} \right\}^{-\frac{1}{2}}$$

- And then caluculate the value of $\hat{\beta}$ directly from the second equation

THE LOADING DOCK EXAMPLE REVISITED:
Estimating the arrival and service time from observed data

THE LOADING DOCK EXAMPLE

- Consider a loading dock that has room for one truck and two places for trucks to wait. If a truck is at the dock being unloaded and two trucks are waiting, all other arriving trucks go to other loading docks. When a truck arrives, it begins unloading immediately if the dock is free, otherwise it waits in the queue until the dock is free or is turned away if the queue is full. Arrivals and time services are random. To determine their random behavior samples of interarrival and service times have been carefully collected. Tablex in the two next slides show samples of the data observed

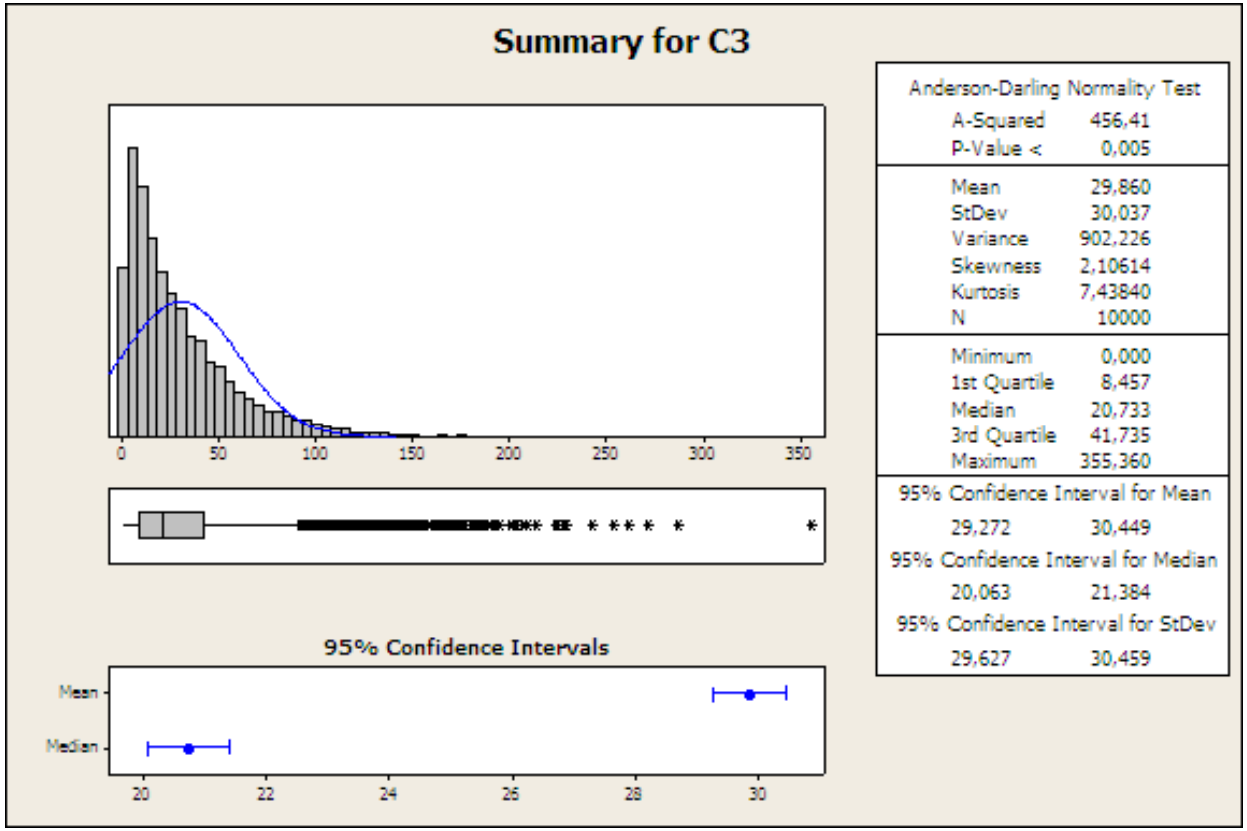
ARRIVALS RANDOMNESS

SAMPLE OF 300 OBSERVED INTERARRIVAL TIMES

84,449	29,369	6,009	0,232	18,162	2,357	16,887	4,136	31,707	22,651
20,798	9,322	19,579	8,234	20,267	23,656	14,643	53,224	3,224	12,603
123,798	5,279	51,162	11,474	98,138	12,052	50,065	1,486	22,167	8,257
28,832	1,364	79,311	69,994	0,867	89,882	24,658	19,221	13,439	11,960
81,400	69,966	0,517	6,897	87,190	17,742	3,924	34,763	2,255	1,161
28,661	90,710	36,390	21,531	30,787	9,270	11,598	106,686	43,787	61,012
58,679	28,087	22,703	13,379	4,266	24,980	49,525	38,362	2,659	9,271
7,936	27,626	23,266	61,353	70,001	11,280	5,080	17,008	8,374	24,529
15,793	106,128	15,722	40,410	89,493	2,410	109,426	3,591	2,258	27,528
62,306	15,075	39,672	3,873	38,693	43,894	4,612	24,001	59,494	16,925
21,640	34,190	19,287	21,327	1,523	11,102	12,455	31,005	8,292	16,364
1,238	12,955	5,952	46,836	75,698	10,202	9,857	102,162	1,391	97,317
40,113	31,746	2,650	17,331	14,693	10,409	18,597	5,149	83,641	36,151
54,941	6,734	23,399	120,366	20,693	40,803	4,757	20,134	5,307	58,116
29,304	14,099	47,489	59,496	14,830	7,135	18,544	23,547	15,321	38,685
19,066	110,143	9,450	24,090	175,779	18,244	6,734	18,676	43,300	0,874
9,088	58,277	36,062	26,495	122,145	14,223	82,478	138,351	8,536	82,226
26,361	12,843	113,100	19,539	9,505	36,290	24,972	17,838	21,782	3,963
26,584	34,633	31,299	2,419	11,248	69,816	44,807	0,793	84,997	20,888
79,500	98,614	0,819	27,899	67,668	73,069	46,924	33,798	20,943	94,802
73,301	19,088	2,279	23,901	44,199	14,385	13,742	25,265	76,738	25,953
69,827	5,487	7,208	27,777	82,324	126,541	21,196	0,072	0,696	1,892
37,781	67,437	33,764	49,645	8,319	9,898	19,139	44,774	40,859	12,577
39,939	80,103	17,242	13,335	14,355	112,727	3,540	19,852	0,288	18,302
31,272	91,915	17,800	68,102	2,618	64,854	18,291	23,067	21,985	34,303
18,656	41,367	27,942	7,577	2,764	14,571	0,794	11,138	17,898	16,604
52,020	3,662	70,446	32,382	4,017	3,454	55,708	96,437	24,355	34,756
4,395	7,266	11,129	44,217	38,611	50,590	46,658	15,014	1,170	6,240
82,497	50,926	6,474	32,033	89,719	39,603	6,380	2,583	60,920	2,995
27,478	0,402	4,842	48,868	60,084	41,741	5,871	22,375	28,743	2,844

ARRIVALS RANDOMNESS (I)

Learning from data: preliminary descriptive analysis



Anderson-Darling Normality Test	
A-Squared	456,41
P-Value <	0,005
Mean	29,860
StDev	30,037
Variance	902,226
Skewness	2,10614
Kurtosis	7,43840
N	10000
Minimum	0,000
1st Quartile	8,457
Median	20,733
3rd Quartile	41,735
Maximum	355,360
95% Confidence Interval for Mean	
29,272	30,449
95% Confidence Interval for Median	
20,063	21,384
95% Confidence Interval for StDev	
29,627	30,459

- Mean: $\mu = 29.860$
- Standard Deviation: $\sigma = 30.037$
- Coefficient of Variation

$$c_V = \frac{\sigma}{\mu} = 1.005927$$

Coefficient of variation $\cong 1 \Rightarrow$ Modeling hypothesis: Interarrival times are exponentially distributed

$$f(t) = 0.03348e^{-0.03348t}$$

ESTIMATION OF THE QUALITY OF THE FITTED DISTRIBUTION

1. Group the empirical observations in a histogram of k classes
2. Let N_j be the number of observations from the sample belonging to the j-th class, the interval of extremes $[a_{j-1}, a_j)$
3. If the sample comes from a theoretical probability function f(x) the theoretical number of obsrtevations expected to fall in this class is Np_j, where N is the sample size (total number of observations) and p_j is the expected proportion, which, according with the probability function f(x), is given by:

$$p_j = \int_{a_{j-1}}^{a_j} f(x)dx$$

4. The emprical value of the corresponding χ^2 distribution is given by

$$\chi^2 = \sum_{j=1}^{k} \frac{(N_j - Np_j)^2}{Np_j}$$

If the fit is accetably good this empirical value of χ^2 will not exceed the theoretical value $\chi^2_{k-1,1-\alpha}$, for k-1 degrees of freedom and an αsignificance level.

5. In our case taking k = 25 classes and costant probabilities p_j=0.04, for each class, the number of cosntant observation per class is Np_j = 300 x 0.04 = 12. Constant probabilities imply variable limits for the classes, given in this example by: $a_j = -29.680\ln(1-\frac{j}{25})$, the results are displayed in the table of the next slide.

χ^2 GOODNESS-OF-FIT TEST FOR THE INTERARRIVALS TIME DISTRIBUTION

j	Interval	N_j	Np_j	$\frac{(N_j - Np_j)^2}{Np_j}$
1	[0, 1.2189)	13	12	0,08333333
2	[1.2189, 2.4897)	12	12	0
3	[2.4897, 3.8171)	12	12	0
4	[3.8171, 5.2062)	12	12	0
5	[5.2062, 6.6630)	9	12	0,75
6	[6.6630, 8.1946)	8	12	1,33333333
7	[8.1946, 9.8091)	12	12	0
8	[9.8091, 11.5158)	10	12	0,33333333
9	[11.5158, 13.3261)	8	12	1,33333333
10	[13.3261, 15.2532)	14	12	0,33333333
11	[15.2532, 17.3133)	9	12	0,75
12	[17.3133, 19.5262)	18	12	3
13	[19.5262, 21.9163)	14	12	0,33333333
14	[21.9163, 24.5144)	14	12	0,33333333
15	[24.5144, 27.3604)	9	12	0,75
16	[27.3604, 30.5065)	12	12	0
17	[30.5065, 34.0235)	10	12	0,33333333
18	[34.0235, 38.0107)	10	12	0,33333333
19	[38.0107, 42.6137)	13	12	0,08333333
20	[42.6137, 48.0578)	11	12	0,08333333
21	[48.0578, 54.7208)	9	12	0,75
22	[54.7208, 63.3110)	12	12	0
23	[63.3110, 75.4182)	12	12	0
24	[75.4182, 96.1156)	20	12	5,33333333
25	[96.1156, ∞)	17	12	2,08333333

Empirical value of χ^2

$$\chi^2 = \sum_{j=1}^{k} \frac{(N_j - Np_j)^2}{Np_j} = 18.33333$$

Theoretical value of χ^2 for α = 0.10 and k-1=24 degrees of freedom

$$\chi^2_{24,0.90} = 33.196$$

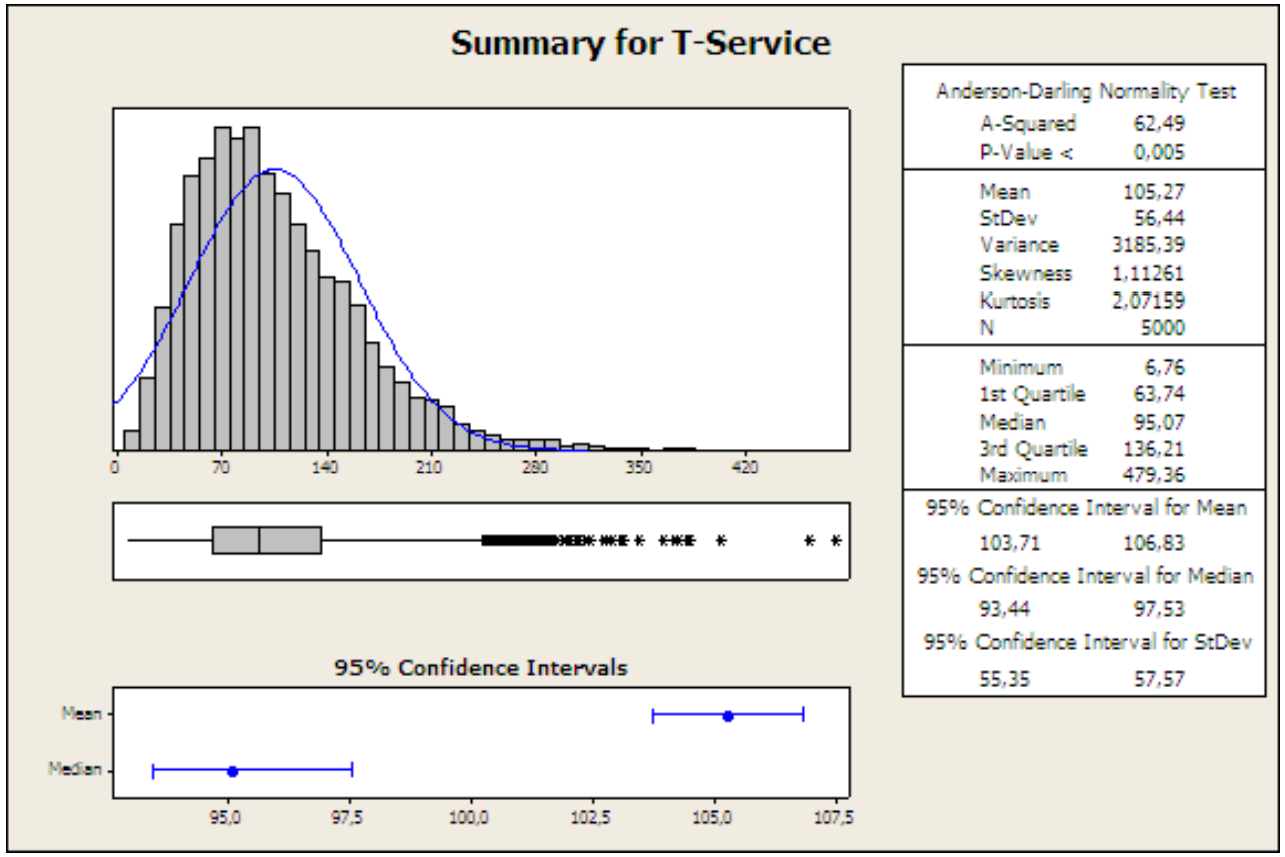
$\chi^2_{24,0.90} > \chi^2 \Rightarrow$ The null hypothesis H_0 "The interarrival times are exponentially distributed according to $f(t) = 0.03348e^{-0.03348t}$" cannot be rejected

SERVICES RANDOMNESS

SAMPLE OF 300 OBSERVED SERVICE TIMES

69,508	164,412	73,916	230,06	211,696	35,368	64,812	95,352	90,376	43,44
110,268	113,412	46,624	97,892	119,748	106,472	63,832	31,416	152,952	54,52
64,288	44,024	63,188	172,156	71,588	56,952	76,78	128,62	97,628	68,736
35,932	42,12	38,544	27,744	52,044	75,48	308,832	42,424	157,012	182,58
110,72	43,272	29,68	60,712	127,904	81,316	223,864	94,48	76,184	130,504
101,332	167,08	66,48	91,356	66,216	69,036	151,54	38,56	187,164	89,852
84,852	78,168	22,896	43	142,468	148,212	118,12	94,632	37,332	37,304
78,072	76,612	59,096	88,804	215,952	65,948	106,5	83,916	148,752	89,024
115,104	68,268	58,584	86,472	89,988	67,936	53,3	141,032	128,924	89,044
54,416	145,144	99,336	222,684	153,324	285,852	185,404	85,748	92,564	112,156
96,04	142,524	45,972	73,064	83,552	270,368	53,548	105,74	94,396	124,26
59,216	87,492	143,624	185,372	31,152	71,408	165,836	65,956	96,776	52,784
75,54	161,068	125,912	93,392	60,056	66,584	168,464	114,804	162,952	110,5
146,648	144,388	69,2	21,74	70,432	146,456	167,524	157,168	113,864	123,552
228,444	93,576	113,004	68,552	81,152	114,612	165,892	136,44	114,1	106,308
139,444	124,416	102,144	41,808	106,632	144,772	123,892	62,8	164,36	195,068
108,648	307,972	152,384	124,724	19,628	59,648	42,812	66,32	72,292	53,368
135,348	93,38	116,432	179,328	95,912	103,244	271,536	91,864	324,512	98,624
150,78	79,636	133,584	62,892	61,908	102,052	22,864	65,156	68,972	85,768
186,332	95,32	160,424	72,736	60,308	41,06	105,58	69,42	38,668	71,84
86,532	85,48	157,6	145,912	84,828	63,696	49,804	61,312	86,656	58,264
146,724	57,7	103,328	218,616	44,464	210,288	200,776	60,964	32,528	114,724
66,752	83,532	111,064	201,72	60,624	58,376	47,084	36,924	110,172	37,404
166,34	55,956	169,536	160,824	78,848	154,956	102,768	80,944	45,268	216,732
34,604	201,692	70,888	165,864	34,632	66,44	71,66	49,652	24,024	210,468
60,956	80,34	107,62	77,176	154,148	113,296	30,816	137,864	50,588	97,832
179,688	118,216	40,06	138,7	52,548	117,152	150,836	151,824	93,92	32,58
70,952	80,968	86,412	97,98	68,7	57,908	161,568	215,708	94,656	134,916
105,324	228,856	83,14	77,776	86,872	94,24	76,476	174,976	58,28	96,664
68,28	123,056	59,784	123,228	149,936	60,288	122,492	136,18	31,972	79,46

ANALYZING SERVICE RANDOMNESS: Descriptive Data Analysis



Coefficient of Variation: $c_x = \frac{\sigma}{\mu} = \frac{56.44}{105.27} = 0.5661 < 1$

$\Rightarrow$ Candidate distributions: Weibull, Gamma, Lognormal, Logistic…

AUXILIARY GRAPHIC TOOLS: PROBABILITY PLOT

A probability plot can be thought as a graphical comparison of an estimate of the true distribution function of the available data X_1, X_2, …., X_n with the distribution function of the fitted distribution.

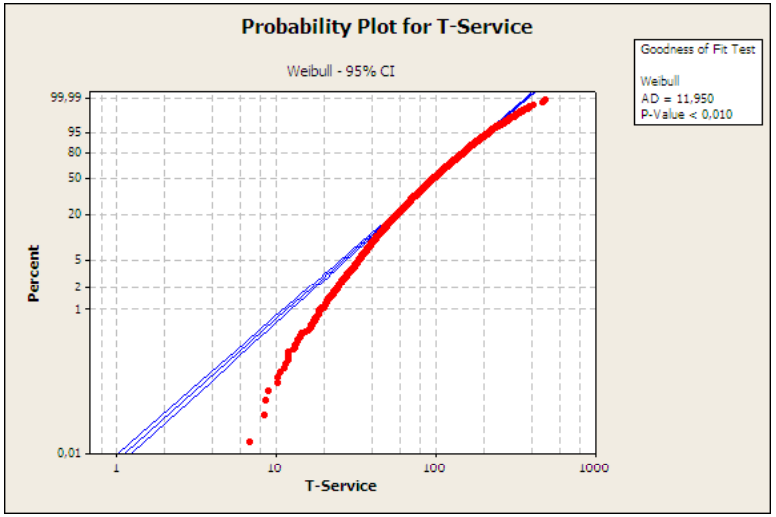
A *probability-probability* plot (P-P) is a graph of the model probability $\hat{F}(X_i)$ versus the sample probability

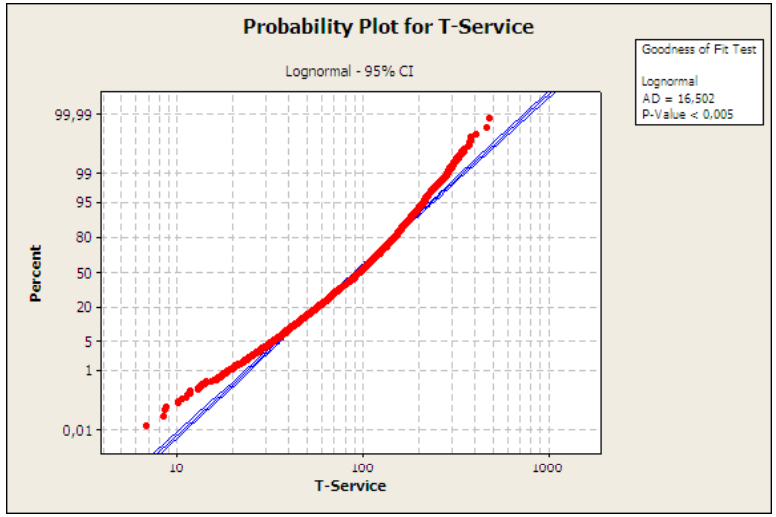
$$\tilde{F}_n(x_i) = q_i = \frac{i - 0.5}{n}, i = 1,2,....,n$$

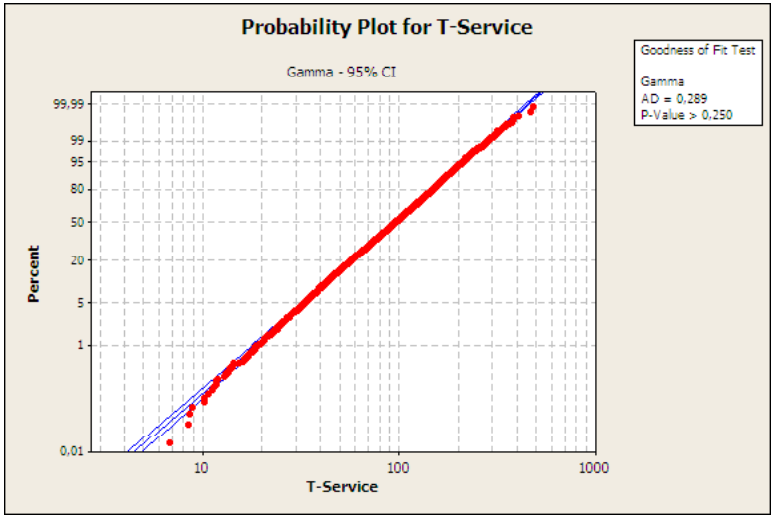
If $\hat{F}(x)$ and $\tilde{F}_n(x)$ are close togeher then the P-P plot will also be approximately linear with an intercept 0 and a slope 1.

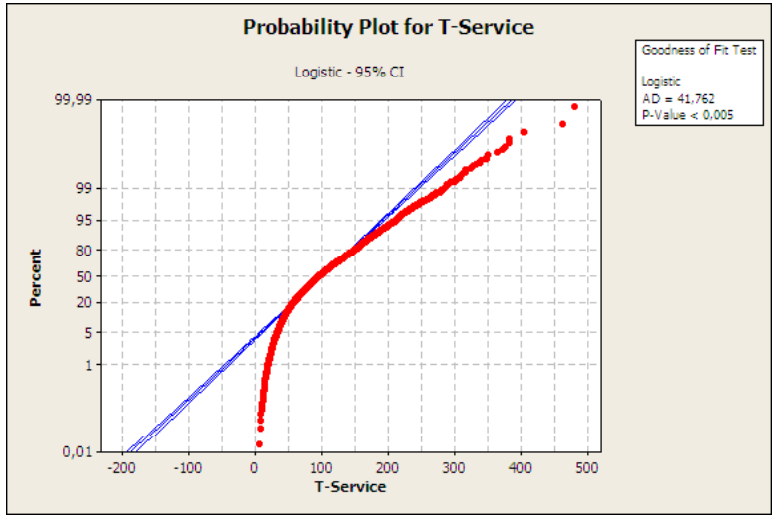
The lineal correlation coefficient of the fit of the P-P points is a measurement of the goodness-of-fit of the proposed distribution.

ANALYZING SERVICE RANDOMNESS: Searching a Model



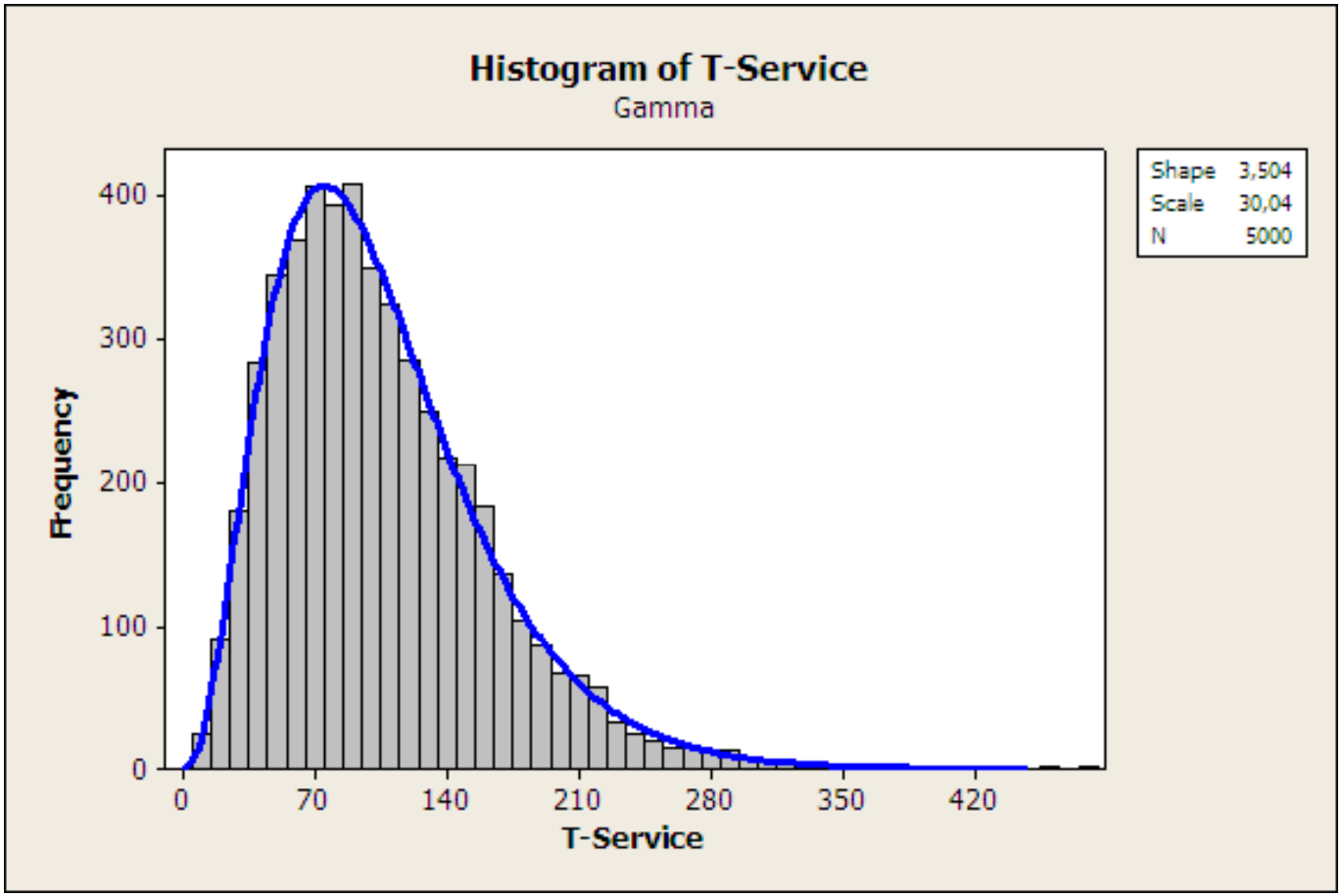






ANALYZING SERVICE RANDOMNESS: Fitting a GAMMA (α,β) Distribution

- Shape parameter α = 3,50416
- Scale parameter β = 30,04149



Departament d'Estadística
i Investigació Operativa
UNIVERSITAT POLITÈCNICA DE CATALUNYA

GENERATING RANDOM DATA INPUTS TO SIMULATION MODELS

RANDOM VARIATE GENERATION

REFERENCES

- J. Banks, J.S. Carson and B.L. Nelson, Discrete-Event System Simulation, Prentice-Hall 1999

  Ch. 9 Random Variate Generation

- S. M. Ross, Simulation, Academic Press 2002

  Ch. 5 Generating Continuous Random Variables

- G.S. Fishman, Discrete-Event Simulation: Modeling, Programming and Analysis, Springer 2001

  Ch. 8 Sampling from Probability Distributions

- Handbook of Simulation: Principles, Methodology, Advances, Applications and Practice, Ed. By J. Banks, John Wiley 1998

  Ch. 5 (by R.C.H. Cheng) Random Variate Generation

GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES: The Inverse Transform Method (I)

- Let X be a continuous random variable with a probability function f(x) and a distribution function F(x)

$$\mathbf{F(x) = P\{X \le x\} = \int_{-\infty}^{x} f(\theta)d\theta}$$

- Let's define the continuous random variable **U=F(x)**
- **PROPOSITION**: *Let* U *be a uniform* (0,1) *random variable. For any continuous distribution function* F *the random variable* X defined by $X=F^{-1}(U)$ *has distribution* F
- $F^{-1}(u)$ is defined to be that value of x such that F(x)=u
- Let F_X denote the distribution function of $X=F^{-1}(U)$. Then:

$$F_X(x) = P\{ X \le x \} = P\{ F^{-1}(U) \le x \}$$

- Since F is a distribution function it follows that F(x) is a monotone increasing function of x and therefore $\{ a \le b \} \Rightarrow \{ F(a) \le F(b) \}$and thus

$$F_X(x) = P\{ F[F^{-1}(U) \le F(x) \} = P\{ U \le F(x) \} = F(x)$$

- Since $F(F^{-1}(U)) = U$, and U is uniform in (0,1)

GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES: The Inverse Transform Method (II)

- **ALGORITHM:**
  - **Generate U from $\mathcal{U}(0,1)$**
  - **Do $X \leftarrow F^{-1}(U)$**
  - **Repeat**
- **Example:** Exponential distribution

$$f(x) = \lambda e^{-\lambda x}; x \geq 0$$

$$F(x) = \int_0^x f(z)dz = \int_0^x \lambda e^{-\lambda z} dz = -e^{-\lambda z}\Big|_0^x = 1 - e^{-\lambda x}$$

- If u is a uniform random number in [0,1] then we get x exponentially distributed by f(x) (exponencial) solving the equation:

$$F(x) = u$$

$$u = 1 - e^{-\lambda x} \Rightarrow x = -\frac{1}{\lambda}\ln(1-u) \quad \left(\text{Also}: x = -\frac{1}{\lambda}\ln(u)\right)$$

GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES: The Inverse Transform Method (III)
THE UNIFORM DISTRIBUTION IN [a,b]

$$f(x) = \frac{1}{b-a}; a \le x \le b$$

$$F(x) = \int_a^x f(z)dz = \int_a^x \frac{dz}{b-a} = \left.\frac{z}{b-a}\right|_a^x = \frac{x-a}{b-a}$$

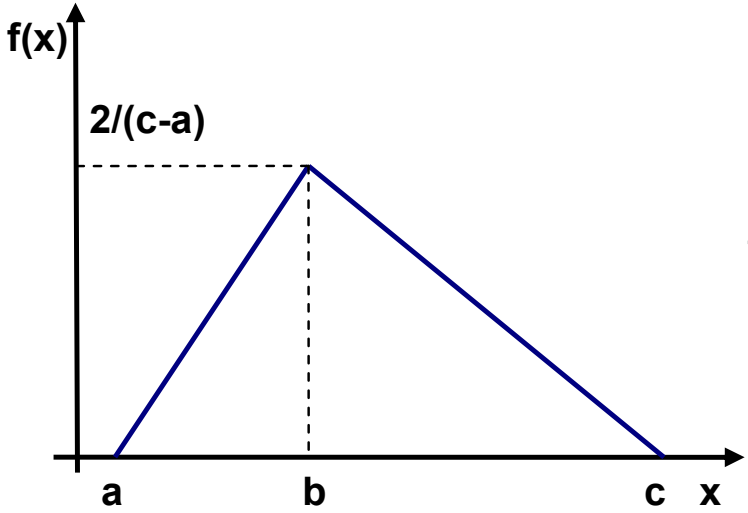
- If u is a number uniformly distributed in [0,1] then one can get x distributed according to f(x) (Uniform in [a,b]) solving the equation :

$$F(x) = u$$

$$u = \frac{x-a}{b-a} \Rightarrow x = a + (b-a)u$$

GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES: The Inverse Transform Method (IV), THE TRIANGULAR DISTRIBUTION

The Triangular distribution (a,b,c), a<b<c has the probability function:



$$f(x)=\begin{cases}\dfrac{2(x-a)}{(b-a)(c-a)} & a\le x<b\\[2ex] \dfrac{2(c-x)}{(c-a)(c-b)} & b\le x<c\\[2ex] 0 & \text{Otherwise}\end{cases}$$

And the distribution function

$$F(x)=\begin{cases}0 & x<a\\[2ex] \dfrac{(x-a)^2}{(b-a)(c-a)} & a\le x\le b\\[2ex] 1-\dfrac{(c-x)^2}{(c-a)(c-b)} & b\le x\le c\\[2ex] 1 & x>c\end{cases}$$

Algorithm:

Do: $\beta=\dfrac{b-a}{c-a}$ **and U=RN(0,1)**

If (U<β) $T=\sqrt{\beta U}$

Otherwise: $T=1-\sqrt{(1-\beta)(1-U)}$

Return X=a+(c-a)T

GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES: The Inverse Transform Method (V) WEIBULL DISTRIBUTION WITH SCALE PARAMETER α>0 AND SHAPE PARAMETER β>0

$$f(x) = \frac{\beta}{\alpha^{\beta}} x^{\beta-1} e^{-\left(x/\alpha\right)^{\beta}}$$

$$F(x) = \int_0^x f(z)dz = 1 - e^{-\left(x/\alpha\right)^{\beta}}$$

- If u is a number unifromly distributed in [0,1] then one can generate x distributed according to f(x) (Weibull (α,β)) solving the equation:

$$F(x) = u$$

$$u = 1 - e^{-\left(x/\alpha\right)^{\beta}} \Rightarrow x = \alpha\left[-\ln(1-u)\right]^{1/\beta}$$

- **ALGORITHM:**
  - **Generate u=RN(0,1)**
  - **Do x = α [-ln(1-u)]$^{1/\beta}$**

GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES:
The Inverse Transform Method (VI)
THE GEOMETRIC DISCRETE DISTRIBUTION

$$f(x) = p(1-p)^x; x = 0,1,2,....;0 < p < 1$$

$$F(x) = \sum_{j=0}^{x} p(1-p)^j = p\frac{1-(1-p)^{x+1}}{1-(1-p)} = 1-(1-p)^{x+1}$$

If u is a number unifromly distributed in [0,1] then one can generate x distributed according to f(x) (Geometric) solving the equation: $F(x-1) = 1-(1-p)^x < u \le 1-(1-p)^{x+1} = F(x)$

$$\Rightarrow \left\{(1-p)^{x+1} \le 1-u < (1-p)^x\right\} \Leftrightarrow \left\{(x+1)\ln(1-p) \le \ln(1-u) \le x\ln(1-p)\right\}$$

And taking into account that 1-p < 1 $\Rightarrow$ ln (1-p) < 0

$$\frac{\ln(1-u)}{\ln(1-p)} - 1 \le x < \frac{\ln(1-u)}{\ln(1-p)}$$

$$\Rightarrow x = \left\lceil \frac{\ln(1-u)}{\ln(1-p)} - 1 \right\rceil$$

GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES: Approximations to the Inverse Transform Method

A. When the distribution function F(x) has an inverse that either is computationally costly to compute or cannot be computed analytically because it has not a closed form then it can be approximated generating a set of pairs of values:

$$(F(x_i),x_i) \text{ such that } x_i < x_{i+1}$$

And applying the following algorithm that inverts the polygonal approximation to F(x) interpolating between these pairs:

1. Find X_i such that $F(X_i) \leq U \leq F(X_{i+1})$
2. Compute:

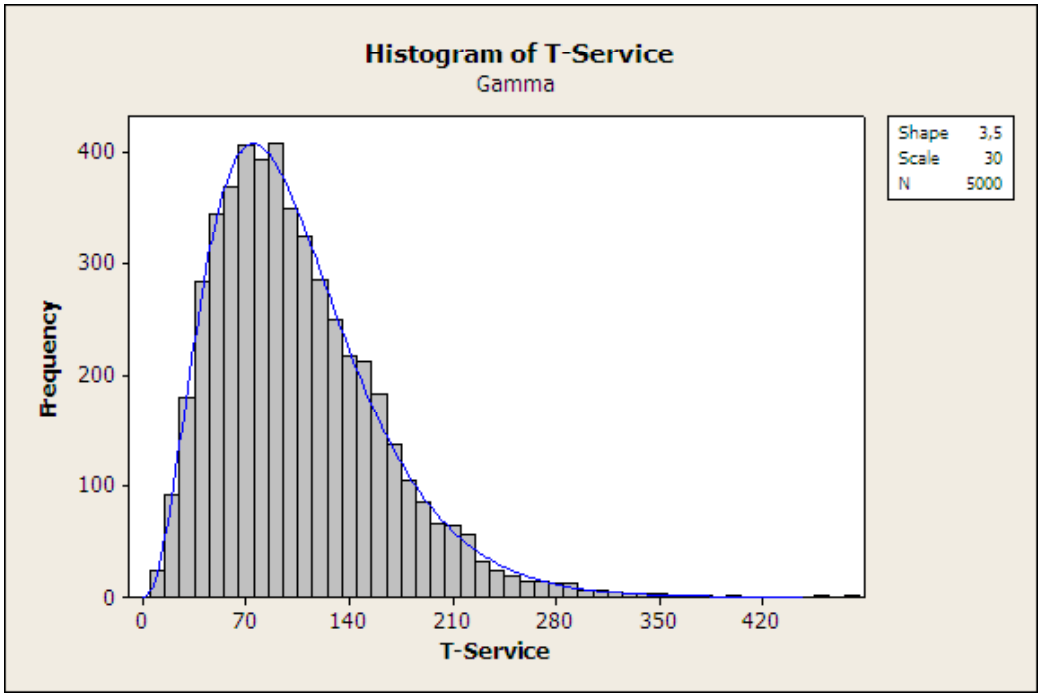
$$X = \frac{[F(X_{i+1}) - U]X_i + [U - F(X_i)]X_{i+1}}{F(X_{i+1}) - F(X_i)}$$

B. Solving numerically the equation U = F(U) by Newton-Raphson, Bisection…..

WORKING WITH EMPIRICAL DISTRIBUTIONS (I)

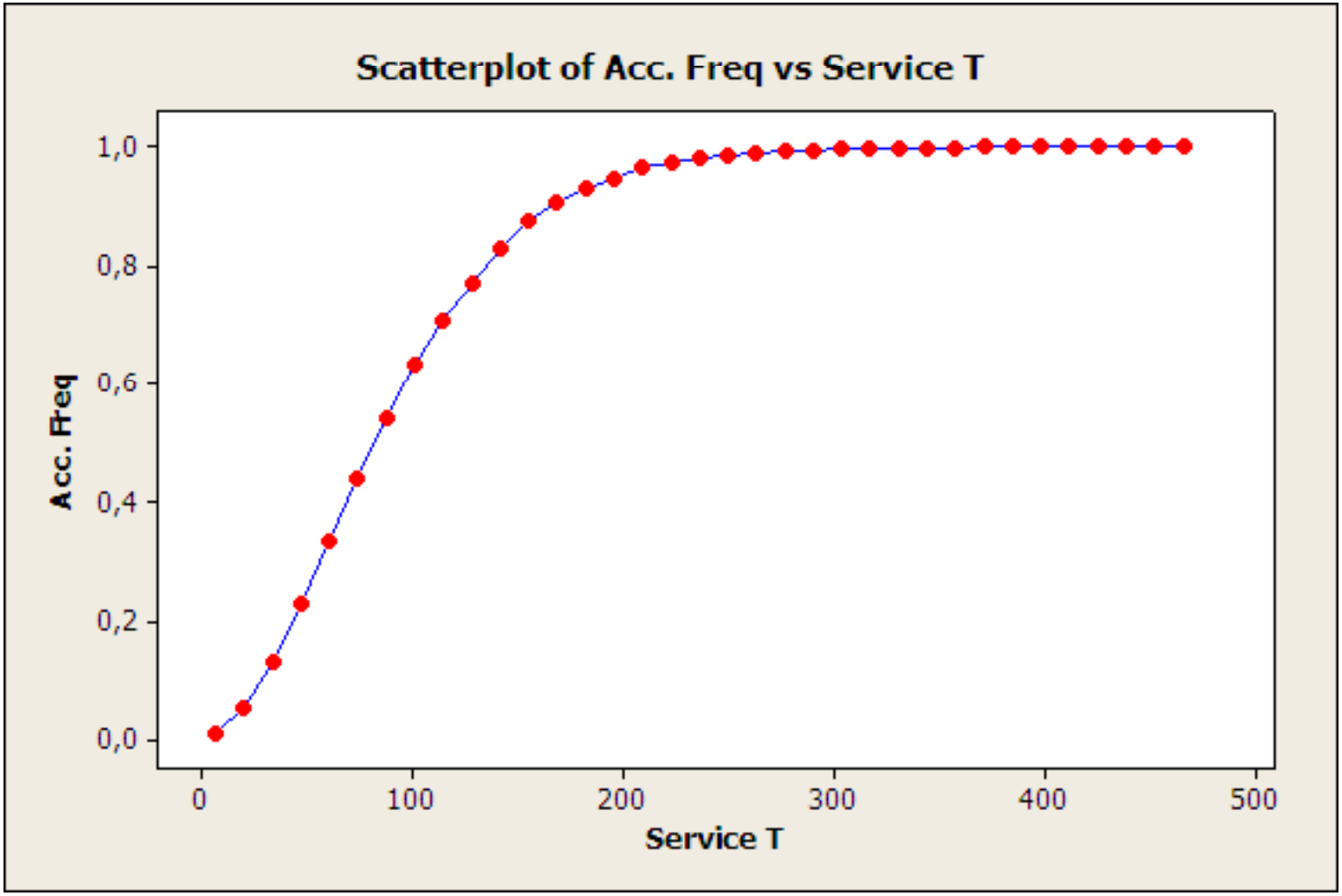
Building the empirical distribution

- Build a histogram whose ends are, respectively, the smallest and the biggest of the observed values [6.764, 479.356]
- Calculate the frequencies and the accumulated frequencies for each class. 35 classes, class width 13.502.



CLASS	SERVICE TIME	FREQ.	ACC. FREQ
1	6,764 ≤ x < 20,266	0,0114	0,0114
2	20,266 ≤ x < 33,768	0,0422	0,0536
3	33,768 ≤ x < 47,270	0,079	0,1326
4	47,270 ≤ x < 60,772	0,096	0,2286
5	60,772 ≤ x < 74,274	0,1064	0,335
6	74,274 ≤ x < 87,776	0,1034	0,4384
7	87,776 ≤ x < 101,278	0,1038	0,5422
8	101,278 ≤ x < 114,780	0,0904	0,6326
9	114,780 ≤ x < 128,282	0,076	0,7086
10	128,282 ≤ x < 141,784	0,0626	0,7712
11	141,784 ≤ x < 155,286	0,0562	0,8274
12	155,286 ≤ x < 168,788	0,0484	0,8758
13	168,788 ≤ x < 182,290	0,0306	0,9064
14	182,290 ≤ x < 195,792	0,0238	0,9302
15	195,792 ≤ x < 209,294	0,0168	0,947
16	209,294 ≤ x < 222,796	0,018	0,965
17	222,796 ≤ x < 236,298	0,0092	0,9742
18	236,298 ≤ x < 249,800	0,006	0,9802
19	249,800 ≤ x < 263,302	0,004	0,9842
20	263,302 ≤ x < 276,804	0,0038	0,988
21	276,804 ≤ x < 290,306	0,0042	0,9922
22	290,306 ≤ x < 303,808	0,002	0,9942
23	303,808 ≤ x < 317,310	0,0022	0,9964
24	317,310 ≤ x < 330,812	0,0006	0,997
25	330,812 ≤ x < 344,314	0,0008	0,9978
26	344,314 ≤ x < 357,816	0,0006	0,9984
27	357,816 ≤ x < 371,318	0,0002	0,9986
28	371,318 ≤ x < 384,820	0,0008	0,9994
29	384,820 ≤ x < 398,322	0	0,9994
30	398,322 ≤ x < 411,824	0,0002	0,9996
31	411,824 ≤ x < 425,326	0	0,9996
32	425,326 ≤ x < 438,828	0	0,9996
33	438,828 ≤ x < 452,330	0	0,9996
34	452,330 ≤ x < 465,832	0,0002	0,9998
35	465,832 ≤ x < ∞	0,0002	1

THE EMPIRICAL DISTRIBUTION



WORKING WITH THE EMPIRICAL DISTRIBUTION

Let $X_1, X_2, \ldots, X_n$ be the observed values, clustered in k classes whose respective ends are son $[a_0,a_1)$, $[a_1,a_2)$, ..., $[a_{j-1}, a_j)$, ...,$[a_{k-1},a_k)$, the intermediate values can be obtained interpolating as follows:

$$F(x) = \begin{cases} 0 \\ F(a_{j-1}) + \dfrac{x - a_{j-1}}{a_j - a_{j-1}}\left[F(a_j) - F(a_{j-1})\right] \\ 1 \end{cases}$$

where F(x) = 0 if $x < a_0$, F(x)=1 if $a_k \leq x$, and F(x) equals the result of evaluating the intepolation formula when al resultado de la evaluación de cuando $a_{j-1} \leq x < a_j$.

A disadvantage of the empirical distribuiton is that they can never generate values smaller than a_0, or grater than a_k, and that can introduce a bias in the samples generated from them.

EXAMPLE: GENERATING A SAMPLE FROM AN EMPIRICAL DISTRIBUTION

1. Generate u uniformly distributed in [0,1]
2. Identify to which class it belongs: $u \in [F(a_{j-1}), F(a_j)]$
3. Calculate

$$x = a_{j-1} + \frac{a_j - a_{j-1}}{F(a_j) - F(a_{j-1})}\left[u - F(a_{j-1})\right]$$

Example: $u=0.6148 \Rightarrow u \in [0.5422, 0.6326]$

$a_{j-1}=101.278$, $a_j= 114.780$

$F(a_{j-1})= 0.5422$, $F(a_j) = 0.6326$

$$x = 101.278 + \frac{114.780 - 101.278}{0.6326 - 0.5422}\left(0.6148 - 0.5422\right) = 112.1214$$

GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES: The Acceptation-Rejection Method (I)

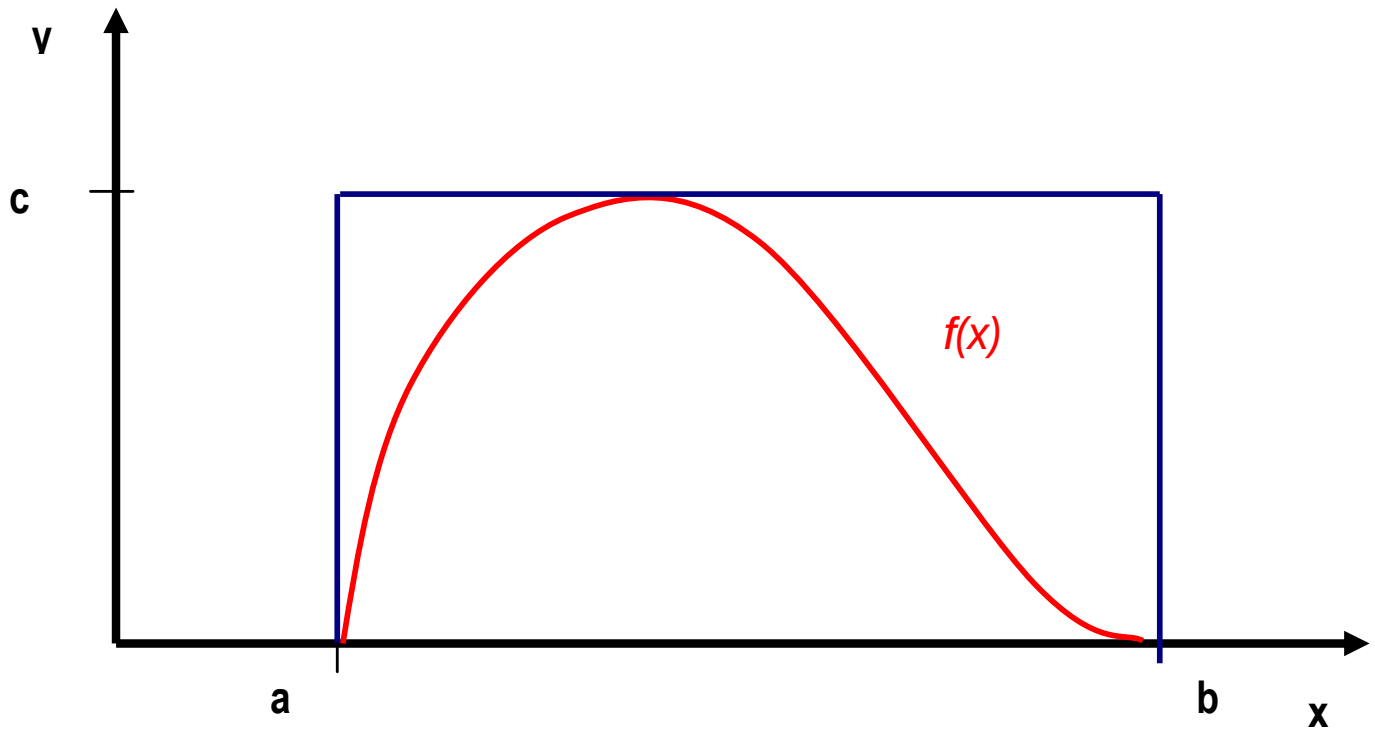
- The Inverse Transform Method requires the explicit (or approximate) knowledge of the distribution function F(x)
- Quite frequently the probability distribution f(x) is known but not the distribution function F(x). A typical case is when the distribution function F(x) has no closed form, as for example in the case of the normal distribution.
- This is the case for using the *rejection method* that only requires to know the probability function f(x), and that it be bounded and not zero in a finite interval [a,b].
- Let's define:

$$c = \max\{f(x) : a \le x \le b\}$$

- Then:
  1. Generate X uniformly distributed in (a,b);
  2. Generate Y uniformly distributed in (0,c);
  3. If $Y \le f(X)$, then accept X,otherwise reject and repeat.

GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES: The Acceptation-Rejection Method(II)

The figure depicts geometrically how the rejection method works. The pair of values X and Y represent the Cartesian coordinates of a point belonging to the rectangle of sides c and b-a. A point below f(X) belongs to the area enclosed by the curve and the interval (a,b), then the value X is accepted, otherwise it is rejected.



GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES: The Acceptation-Rejection Method (III)

- The rejection method is highly inefficient computationally when X is rejected in step 3 with a significant probability, then 2 random numbers are misused..
- $\Rightarrow$ The computational efficiency can be improved if the requirement of X belonging to a finite interval is relaxed, but this requires a probability function g(x) such that a cosntant a exists fro which

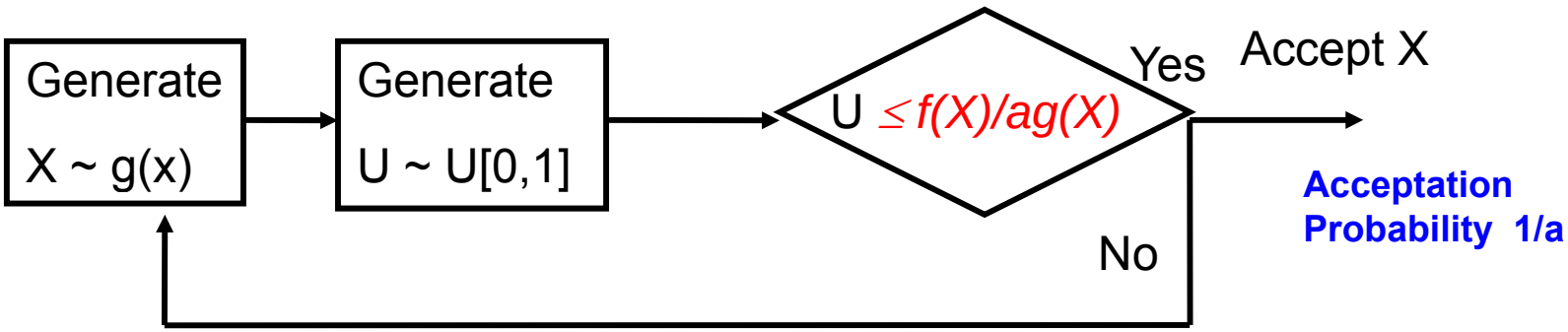
$$f(x) \leq ag(x) \ \forall x$$

- The Generalized Rejection Method is then:
  1. Generate X with probability function g(x);
  2. Generate Y uniformly distributed in (0,ag(X));
  3. If $Y \leq f(X)$, then accept X, otherwise reject and repeat.
- Steps 1 and 2 generate points uniformly distributed under the curve ag(x), the more closer to f(x) the more unlikely steps 1 and 2 will be repeated.
- Step 3 is equivalent to verify if $f(x)/ag(x) \geq U$ when U is uniform in (0,1)

GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES: The Acceptation-Rejection Method (IV)

Generalized Rejection Method:

1. **Generate X with probability function g(x);**
2. **Generate U uniformly distributed in (0,1);**
3. **If U ≤ f(X)/ag(X), then accept *X*, otherwise reject and repeat.**



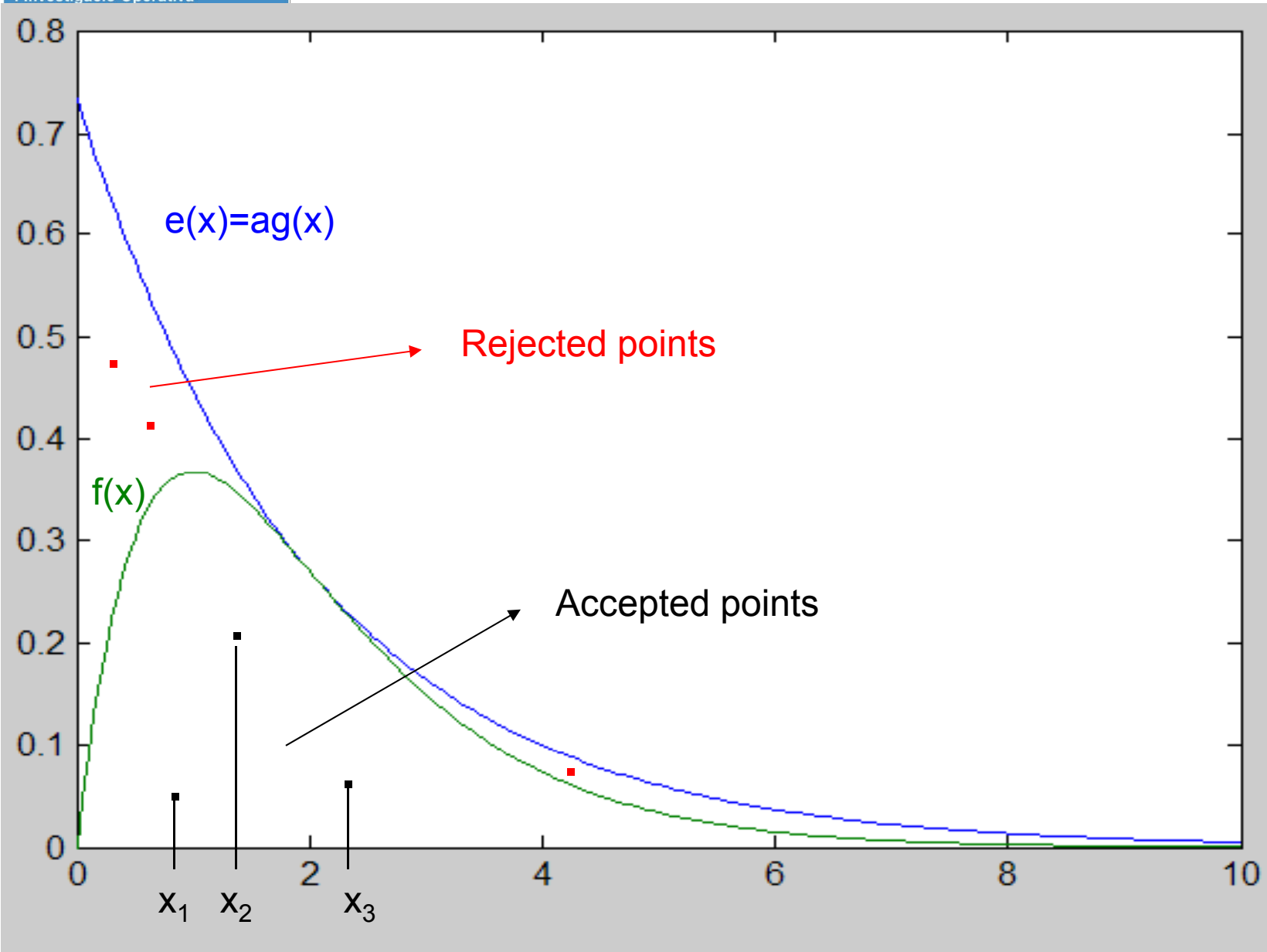
GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES: The Acceptation-Rejection Method (IV)

- The three following hypothesis must hold:
  - It exists a function e(x) that dominates uniformly f(x):

$$e(x) \geq f(x), \ \forall x$$

  - It is possible generate points of coordinates (X,Y) uniformly spread under the graph of e(x) disperse above the x axis
  - Overlapping the graph of f(x) in the same diagram as the graph of e(x) the point (X,Y) will be above or below f(x) depending on whether:

$$Y>f(X) \text{ or } Y\leq f(X)$$


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
e(x)=ag(x)
Rejected points
f(x)
Accepted points
x_1
x_2
x_3

GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES: The Acceptation-Rejection Method (V)

- The method works generating points (X,Y) and returning the X value only if the point is under the graph of (X), that is only if $Y \leq f(X)$.
- Intuitively one can see that the probability of X is proportional to the heigh of f and therefore has the right probalility distribution.
- Two ways computationally efficient of building e(x)
  - That the area between the graphs of e(x) and f(x)be as small as possible
  - That should be easy to gegenrate points under e(x)
- An easy way of building e(x) is by doing e(x)=ag(x) where g(x) is a probability function whose samples are easy to get.
- It can be proof that if X is a value sampled from this probability function then the points of the form:

$$(X,Y) = (X, Uag(x))$$

where U is a variable RN(0,1) generated independently of X, are uniformly distributed under the graph of e(x)

GENERATING SAMPLES OF CONTINUOUS RANDOM VARIATES: The Acceptation-Rejection Method (VII)

- It can be proof that any X accepted by the genearalized rejection method follow a probability distribution F(x).
- It is a consequence of:

$$P\{X \le t | Y \le f(X)\} = \frac{P\{X \le t \text{ e } Y \le f(X)\}}{P\{Y \le f(X)\}} = \frac{F(t)/a}{1/a} = F(t)$$

- Which imples the satisfaction of the second equality

$$P\{X \le t \text{ e } Y \le f(X)\} = \int_{-\infty}^{\infty} P[X \le t \text{ e } Y \le f(X) | X = z] dG(z)$$

$$= \int_{-\infty}^{t} \left[\frac{f(z)}{ag(z)} \right] g(z)dz = \frac{F(t)}{a}$$

- Where the correctness of the term btween the parentheses is a consequence of Y being (conditionally) uniform in [0,ag(z)]; and therefore:

$$P[Y \le f(X)] = 1/a$$

EXAMPLE 1

Utilize the acceptance-rejection method to generate samples from a random variable X whose probability function is:

$$f(x) = 20x(1-x)^3,\ 0 < x < 1$$

(Beta Function with de parameters 2 and 4)

Since it is defined in (0,1) let's consider the rejection method with

$$g(x)=1,\ 0 < x < 1$$

To determine the smallest constant a such that:

$$\frac{f(x)}{g(x)} \le a$$

We must find the maximum of : $\frac{f(x)}{g(x)} = 20x(1-x)^3$

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = 20(1-x)^3 - 60x(1-x)^2 = 0 \ \Rightarrow\ \tilde{x} = \frac{1}{4}$$

$$\frac{f(\tilde{x})}{g(\tilde{x})} = 20\frac{1}{4}\left(1-\frac{1}{4}\right)^3 = \frac{135}{64} \ \Rightarrow\ a = \frac{135}{64}$$

$$\frac{f(x)}{g(x)} = \frac{256}{27}x(1-x)^3$$

ALGORITHM:

STEP 1: Generate the uniform random numbers U_1 and U_2

STEP 2: If $U_2 \le \frac{256}{27}U_1(1-U_1)^3$ Do $X=U_1$

Otherwise return to STEP 1

EXAMPLE 2: Generating samples from the function Gamma (3/2, 1)

$$f(x)=\frac{1}{\Gamma(\alpha)}\beta^{-\alpha}x^{\alpha-1}e^{-x/\beta}=\frac{1}{\Gamma\left(\frac{3}{2}\right)}x^{1/2}e^{x}=Kx^{1/2}e^{-x},x>0,K=\frac{1}{\Gamma\left(\frac{3}{2}\right)}=\frac{2}{\sqrt{\pi}}$$

The fact that the mean of the Gamma function $\Gamma(\alpha,\beta)$ equals to $\alpha\beta$ (=3/2 in this case) suggests to probe as majorant an exponential function with the same mean: $g(x)=\frac{2}{3}e^{-2x/3},x>0$

In which case: $\frac{f(x)}{g(x)}=\frac{Kx^{1/2}e^{-x}}{2/3\,e^{-2x/3}}=\frac{3K}{2}x^{1/2}e^{-x/3}$

An to calculate the constant a we have to find the maximum of $\frac{f(x)}{g(x)}$

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right]=\frac{3K}{2}\left[\frac{1}{2}x^{-1/2}e^{-x/3}-\frac{1}{3}x^{1/2}e^{-x/3}\right]=0\Rightarrow\frac{1}{2}x^{-1/2}-\frac{1}{3}x^{1/2}=0\Rightarrow\bar{x}=\frac{3}{2}$$

And thus: $a=\mathrm{MAX}\left[\frac{f(x)}{g(x)}\right]=\frac{f(\bar{x})}{g(\bar{x})}=\frac{3K}{2}\left(\frac{3}{2}\right)^{1/2}e^{-1/2}=\frac{3^{3/2}}{(2e\pi)^{1/2}}$ And then: $\frac{f(x)}{ag(x)}=\frac{Kx^{1/2}e^{-x}}{\frac{3K}{2}\left(\frac{3}{2}\right)^{1/2}e^{-1/2}\left(\frac{2}{3}\right)e^{-2x/3}}=\left(\frac{2e}{3}\right)^{1/2}x^{1/2}e^{-x/3}$

ALGORITHM:

STEP 1: Generate a uniform random number $U_1\in$RN(0,1); Do $Y=-\frac{3}{2}\ln U_1$

STEP 2: Generate a uniform random number $U_2\in$RN(0,1)

STEP 3: If $U_2\le\left(\frac{2e}{3}\right)^{1/2}Y^{1/2}e^{-Y/3}$ **do X=Y**

Otherwise return to STEP 1

GENERATING SAMPLES FROM A NORMAL STANDARD RANDOM VARIATE Z(0,1) (I)

Using as majoring function the exponential with mean 1: $g(x) = e^{-x}$, $0 < x < \infty$ results:

$$\frac{f(x)}{g(x)} = \sqrt{\frac{2}{\pi}} e^{x-\frac{x^2}{2}}$$ reaching its máximum at:

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{d}{dx}\left[\sqrt{\frac{2}{\pi}} e^{x-\frac{x^2}{2}}\right] = 0 \equiv \text{MAX}\left[x - \frac{x^2}{2}\right] \Rightarrow \bar{x} = 1$$

And thus: $$\frac{f(\bar{x})}{g(\bar{x})} = \sqrt{\frac{2}{\pi}} e^{\frac{1}{2}} = \sqrt{\frac{2e}{\pi}} = a \rightarrow \frac{f(x)}{ag(x)} = \exp\left\{x - \frac{x^2}{2} - \frac{1}{2}\right\} = \exp\left\{-\frac{(x-1)^2}{2}\right\}$$

And therefore the algorithm to generate samples of the absolute value of the random varíate Z normal standard is:

ALGORITHM

STEP 1: Generate an uniform random number U_1; do Y=-ln U_1
STEP 2: Generate an uniform random number U_2
STEP 3: If $U_2 \leq \exp\left\{-\frac{(Y-1)^2}{2}\right\}$ **do X=Y**

Otherwise return to STEP 1

The standard normal Z can be obtained making that it be X *or* –X *with the same probability*

GENERATING SAMPLES FROM A NORMAL STANDARD RANDOM VARIATE Z(0,1) (I)

ALGORITHM:

STEP 1: Generate an uniform random number U_1; do Y_1=-ln U_1
STEP 2: Generate an uniform random number U_2; do Y_2=-ln U_2
STEP 3: If $Y_2 - \frac{(Y_1-1)^2}{2} > 0$ **Do Y=** $Y_2 - \frac{(Y_1-1)^2}{2}$**; GO TO STEP 4**
Otherwise return to STEP 1
STEP 4: Generate an uniform random number U and do:

$$Z = \begin{cases} Y_1 & \text{Si } U \le \frac{1}{2} \\ -Y_1 & \text{Si } U > \frac{1}{2} \end{cases}$$

To generate a normal random variable X~N(μ,σ)do the transform:

$$X = \mu + \sigma Z$$

GAMMA DISTRIBUTION (Rejection Method for $\alpha > 1$, Cheng)

Probability function

$$f(x) = \begin{cases} \dfrac{\beta^{-\alpha}(x)^{\alpha-1}}{\Gamma(\alpha)} \exp\left(-\dfrac{x}{\beta}\right) & x > 0 \\ 0 & \text{Otherwise} \end{cases}$$

where $\Gamma(\alpha)$ is the Gamma function : $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$

Do : $a = (2\alpha - 1)^{-1/2}$, $b = \alpha - \ln 4$, $c = \alpha + a^{-1}$, $d = 1 + \ln 4.5$

While (True){

Do $U_1 = RN(0,1)$ $U_2 = RN(0,1)$

Do $V = a \ln\left[U_1 / (1 - U_1)\right]$, $Y = \alpha e^V$, $Z = U_1 U_2$, $W = b + cV - Y$

If $(W + d - 4.5\, Z \geq 0)$ {

Return $X = \beta Y$

Otherwise

If $(W \geq \ln Z)$ Return $X = \beta Y$

}

}

GAMMA DISTRIBUTION: simple generator $1<\alpha<5$ (Fishman)

While (True) {

Do $U_1 = RN(0,1)$, $U_2 = RN(0,1)$

$V_1 = -\ln U_1$, $V_2 = -\ln U_2$

If $(V_2 > (\alpha - 1)(V_1 - \ln V_1 - 1))$ {

Return $X = \beta V_1$

}

}

GENERATING POISSON SAMPLES BY THE REJECTION METHOD

- A RANDOM POISSON VARIABLE N, DISCRETE WITH MEAN $\lambda>0$ HAS A PROBABILITY FUNCTION

$$p(n) = P(N = n) = \frac{e^{-\lambda}\lambda^n}{n!}, n = 0,1,2,.....$$

- N CAN BE INTERPRETED AS THE NUMBER OF POISSON ARRIVALS IN A UNIT TIME INTERVAL AND THEREFORE THE INTERARRIVAL TIMES A_1, A_2, ... WILL BE EXPONENTIALLY DISTRIBUTED WITH MEAN $1/\lambda$, AND THUS: N = n
- IF AND ONLY IF: $A_1+A_2+.....+A_n \leq 1 < A_1+A_2+.....+A_n+A_{n+1}$.
- AND TAKING INTO ACCOUNT THAT $A_i = -(1/\lambda)\ln(u_i)$ RESULTS

$$\sum_{i=1}^{n} \frac{-1}{\lambda}\ln u_i \leq 1 < \sum_{i=1}^{n+1} \frac{-1}{\lambda}\ln u_i$$

$$\ln\prod_{1}^{n} u_i = \sum_{i=1}^{n}\ln u_i \geq -\lambda > \sum_{i=1}^{n+1}\ln u_i = \ln\prod_{1}^{n+1} u_i$$

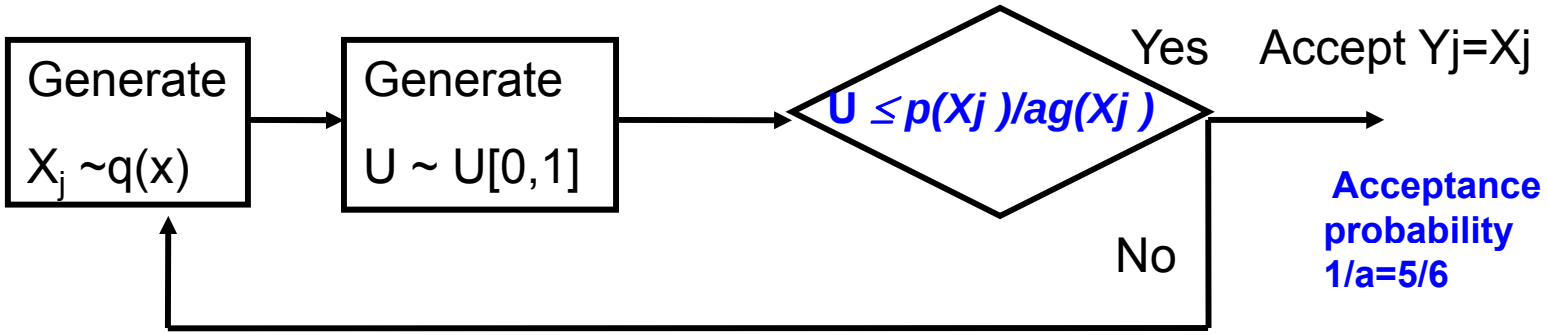
$$\prod_{1}^{n} u_i \geq e^{-\lambda} > \prod_{1}^{n+1} u_i$$

- ALGORITHM:
  - STEP 1: DO n = 0, P = 1
  - STEP 2: GENERATE AN UNIFORM RANDOM NUMBER u_{n+1} AND REPLACE P BY u_{n+1} P
  - STEP 3: IF $P < e^{-\lambda}$ THEN ACCEPT N = n. OTHERWISE REJECT THE CURRENT n, INCREASE n BY ONE UNIT AND REPEAT FROM STEP 2.

ACCEPTANCE REJECTION METHOD FOR DISCRERE RANDOM VARIABLE (DRV)

- Let Y be a DRV with k values and probability function $p(y_j)=p_j$.
- Let X be a DRV with k values and probability function $q(x_j)=q_j$ easy to generate.
- Example: Y_j and X_j values 1 to 10, con funciones de probabilidad: q(x) = {0.1,…,0.1} and p(y)={0.11,0.12,0.09,0.08,0.12,0.1,0.09,0.09,0.1,0.1}
- Let a be such that p(Xj) / q(Xj) ≤a, a=1.2

$$U \leq p(Xj)/ag(Xj) = p(Xj)/0.12$$



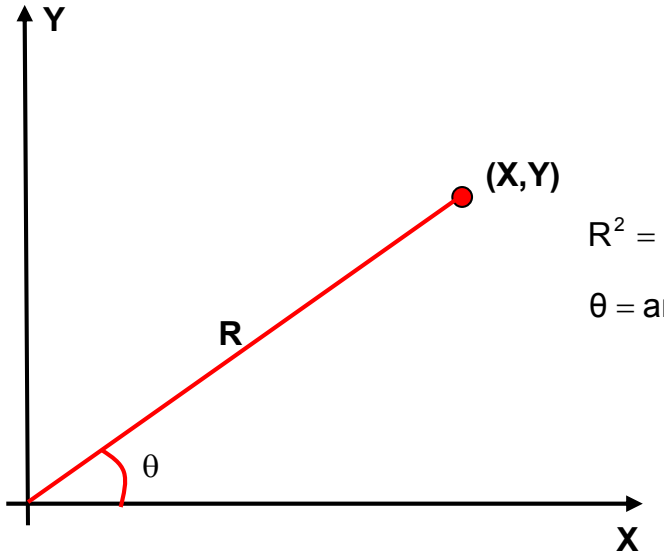
OTHER METHODS

- **GENERATING AN ERLANG E(k,μ) DISTRIBUTION AS SUM OF k EXPONENTIAL, INDEPENDENT RANDOM VARIABLES X_i, IDENTICALLY DISTRIBUTED WITH MEAN 1/kμ**

$$X = \sum_{i=1}^{k} X_i = -\frac{1}{k\mu} \ln\left(\prod_{i=1}^{k} u_i \right)$$

MÉTHOD OF BOX AND MULLER TO GENERATE SAMPLES OF THE STANDARD NORMAL RANDOM VARIABLE (I)

Let X and Y two normal standard independent random variables. Let's denote byr R and θ the polar coordinates of point (X,Y)



$$R^2 = X^2 + Y^2 \qquad (X = R\cos\theta)$$

$$\theta = \text{arc tg}\,\frac{Y}{X} \qquad (Y = R\,\text{sen}\,\theta)$$

Given that X and Y are independent their joint probability function will be the product of the individual probability functions:

$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

To determine the joint probability function of R^2 and θ, $f(R,\theta)$ let's do the variable change:

$$d = x^2 + y^2 \qquad \theta = \tan^{-1}\left(\frac{y}{x}\right) \qquad \text{and then} \quad f(d,\theta) = |J|^{-1} f(x,y)$$

Where J is the Jacobian of the transformation, J=2

METHOD OF BOX AND MULLER TO GENERATE SAMPLES OF THE STANDARD NORMAL RANDOM VARIABLE (II)

$$J=\begin{vmatrix}\frac{\partial d}{\partial x} & \frac{\partial d}{\partial y}\\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y}\end{vmatrix}=\begin{vmatrix}2x & 2y\\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2}\end{vmatrix}=\frac{2x^2}{x^2+y^2}+\frac{2y^2}{x^2+y^2}=2$$

And thus $f(d,\theta)=\frac{1}{2}\frac{1}{2\pi}e^{-\frac{d}{2}},0<d<\infty;0<\theta<2\pi$ which is equal to the product of a random uniform function in $(0,2\pi)$ and an exponential of mean 2, $\left(\frac{1}{2}e^{-\frac{d}{2}}\right)$, what implies that R^2 y θ are independent with R^2 exponential of mean 2 and θ uniformly distributed in $(0,2\pi)$.

$\Rightarrow$ A pair X, Y of normal standard independent random variables can be generated by generating R^2 θ, polar coordinates of point (X,Y) and transforming them to en cartesian coordinates.

METHOD OF BOX AND MULLER TO GENERATE SAMPLES OF THE STANDARD NORMAL RANDOM VARIABLE (III)

BOX AND MULLER ALGORITHM

STEP 1: Generate the independent random numbers U_1 y U_2
PASO 2: Generate R^2, exponential with mean 2: $R^2 = -2 \ln U_1$
Generate θ uniform in $(0,2\pi)$: $\theta = 2\pi U_2$
PASO 3: Change to cartesian coordinates:

$$X = R\cos\theta = \sqrt{-2\ln U_1}\cos(2\pi\pi_2)$$

$$Y = R\,\mathrm{sen}\,\theta = \sqrt{-2\ln U_1}\,\mathrm{sen}(2\pi\pi_2)$$