LECTURE NOTES SIMULATION | CHAPER 7

PROBABILITY REVIEW:

(I) DISCRETE PROBABILITY DISTRIBUTIONS

* Random Variable:

Ex/ Let an unfair die ($P(H)=0{,}6$) is tossed 3 times.

$S = \{HHH, HHT, HTH, THH, THH, THT, TTH, TTT\}$

lex X: # of heads observed. X is a random variable which is a real valued function whose domain is S.:

$X(HHH)=3$

$X(HHT)=X(HTH)=X(THH)=2$

$X(HTT)=X(THT)=X(TTH)=1$

$X(TTT)=0$

* The probability mass function (pmf) of a discrete Random Variable is the function $p(x)$ that gives the probabilities: $P(X=x)=p(x)\geq 0$

Note that; $P(X=3)=0{,}6\cdot 0{,}6\cdot 0{,}6=0{,}064$

$P(X=2)=3\cdot 0{,}6\cdot 0{,}6\cdot 0{,}6=0{,}228$

$P(X=1)=3\cdot 0{,}6\cdot 0{,}6\cdot 0{,}6=0{,}432$

$P(X=0)=0{,}6\cdot 0{,}6\cdot 0{,}6=0{,}216$

X	0	1	2	3
$P(X=x)=p(x)$	0,216	0,432	0,228	0,064

* A random variable X is discrete if it can take only a finite number, or a countably infinitive possible values of X. In general, X takes natural numbers.

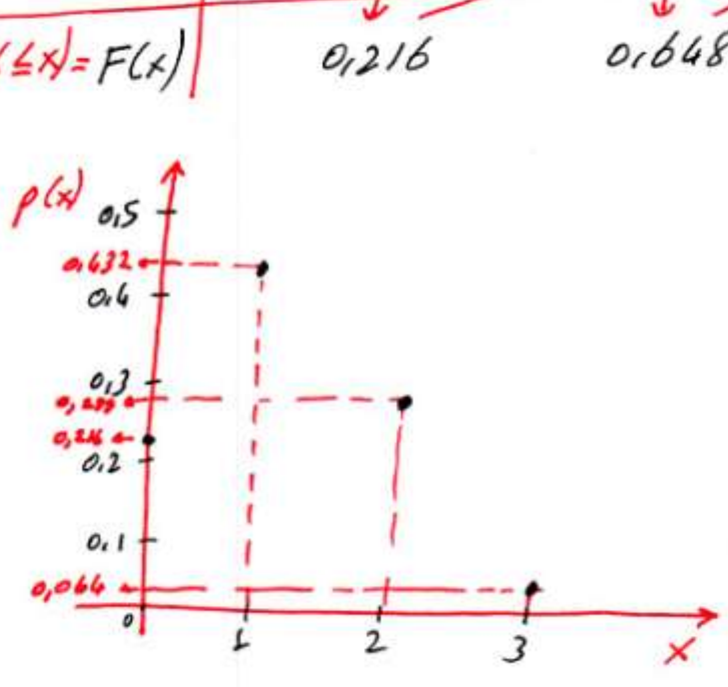
Ex: X: # of accidents in E5 highway in a month
X: # of flirts to find one to marry.

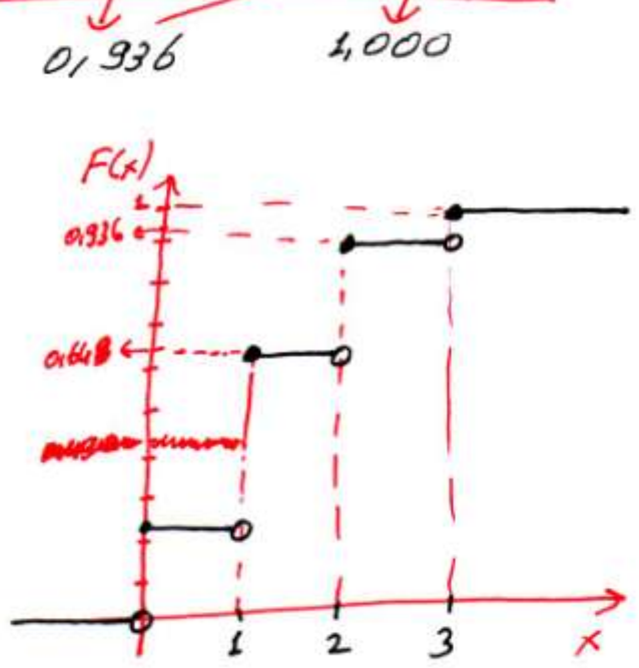
* The (Cummulative) Distribution function (cdf) of a discrete random variable X is defined as;

$$F(x) = P(X \leq x) = \sum_{k=-\infty}^{x} p(k)$$

For X: # of heads obtained example, we have;

X	0	1	2	3
$P(X=x)=p(x)$	0,216	0,432	0,288	0,064
$P(X \leq x)=F(x)$	0,216	0,648	0,936	1,000





Expected Value & Variance

$$\mu = E(X) = \sum_x x \cdot p(x)$$

$$E(g(X)) = \sum_x x \cdot p(x)$$

μ: Expected Value : Mean : Average

Then, $E(X^2) = \sum_x x^2 \cdot p(x)$

$$\sigma^2 = Var(X) = E[(X-\mu)^2] = E(X^2) - \mu^2$$

σ^2: Variance ; σ: Standard Deviation

Ex: The number of children per family is given by the following distribution. Find average number of children in a family and its standard deviation.

X	0	1	2	3
p(x)	0,2	0,4	0,3	0,1

Ans: $\mu = E(X) = 0 \cdot 0{,}2 + 1 \cdot 0{,}4 + 2 \cdot 0{,}3 + 3 \cdot 0{,}1 = \underline{1{,}3}$

$$E(X^2) = 0^2 \cdot 0{,}2 + 1^2 \cdot 0{,}4 + 2^2 \cdot 0{,}3 + 3^2 \cdot 0{,}1 = 2{,}5$$

$$\sigma^2 = 2{,}5 - 1{,}3^2 = 0{,}81$$

$$\sigma = \sqrt{0{,}81} = \underline{0{,}9}$$

* Also remember, for X is random and a,b constants;

$$E(aX+b) = a \cdot E(X) + b$$

$$Var(aX+b) = Var(aX) = a^2 \cdot Var(X)$$

(II) CONTINUOUS PROBABILITY DISTRIBUTIONS

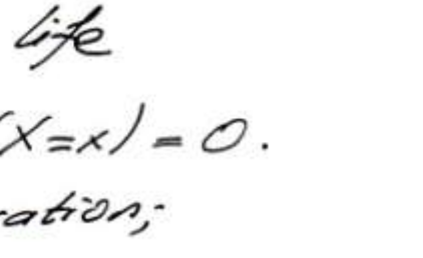
* A random variable X is continuous if it takes values on a real interval $(\alpha; \beta)$. Note that α or β or both can be $-\infty$ or ∞ (respectively :)

Ex: X: Weight of a newborn baby
X: Remaining time of your life

* If X is continuous then $P(X=x)=0$.

We find probabilities by integration;

$$P(a<X<b) = \int_a^b f(x)\,dx = A$$



$f(x)$: probability density function (pdf)

* The (cummulative) Distribution Function (cdf) is;

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(w)\,dw \quad \text{so;} \quad f(x) = \frac{d}{dx} F(x)$$

$$P(a<X<b) = F(b) - F(a)$$

(Equality does NOT matter since $P(X=x)=0$)

Expectation & Variance

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx$$

$$\sigma^2 = Var(X) = E(X^2) - \mu^2$$

Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

Example: The proportion of time X that an industrial robot is in operation during 40-hour work weeks is a random variable with pdf:

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & o.w. \end{cases}$$

a) Find Distribution function

b) Find $P(0{,}2 < X < 0{,}5)$ and $P(X > 0{,}7)$

c) Find $E(X)$ and $Var(X)$

d) If the profit is $Y = 200X - 60$, find $E(Y)$ and $Var(Y)$

Answer: a) $F(x) = \int_0^x 2w\,dw = \frac{2w^2}{2}\Big]_0^x = x^2$ for $0 \leq x \leq 1$

Then, $$F(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

b) $P(0{,}2 < X < 0{,}5) = F(0{,}5) - F(0{,}2) = 0{,}5^2 - 0{,}2^2 = 0{,}21$

$P(X > 0{,}7) = 1 - P(X \leq 0{,}7) = 1 - F(0{,}7) = 1 - 0{,}7^2 = 0{,}51$

c) $E(X) = \int_0^1 x \cdot 2x\,dx = 2 \cdot \frac{x^3}{3}\Big]_0^1 = \frac{2}{3}(1-0) = \frac{2}{3}$

$E(X^2) = \int_0^1 x^2\, 2x\,dx = 2 \cdot \frac{x^4}{4}\Big]_0^1 = \frac{1}{2}(1-0) = \frac{1}{2}$

$Var(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{9-8}{18} = \frac{1}{18}$

d) $E(Y) = 200 \cdot E(X) - 60 = 73{,}33$; $Var(Y) = 200^2 \cdot Var(X) = 2222$

Continuous UNIFORM Distribution

Some pdf's have special properties. Therefore, their characteristics are identified by their parameters and names attributed to them. Hereafter, we'll use the notation like as follows;

$$X \sim \text{Uniform}\ (a;\ b)$$

- X: Random Variable
- $\sim$: is distributed as...
- Uniform: distribution NAME
- $(a;\ b)$: Distribution's parameters.

A distribution is completely given by its pdf: $f(x)$, cdf: $F(x)$, $\mu = E(X)$: mean and $\sigma^2 = Var(X)$: Variance.

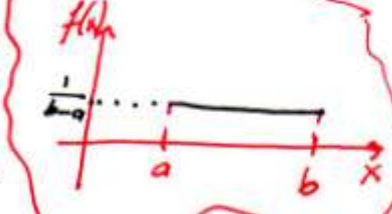
NOTES:

(i) (Sometimes $M_x(t)$: mgf: moment generating function is also defined but it is out of scope of this lecture:)

(ii) (Also note that all of these are NOT always in explicit form)

(iii) (Remember, pdf is used for continuous and pmf is used for discrete Random Variables)

We have: $X \sim \text{Uniform}\ (a;\ b)$



$$f(x) = \frac{1}{b-a} \qquad F(x) = \frac{x-a}{b-a}$$

$$\mu = E(X) = \frac{b+a}{2} \qquad \sigma^2 = Var(X) = \frac{(b-a)^2}{12}$$

$F(x)$; μ and σ^2 formulas can be derived as before. We'll give examples of them.

Random Numbers & Generation (Next:)

Basically, random numbers $R_1, R_2, \ldots$ are assumed to be drawn from the pdf:

$$R \sim \text{Uniform}(0;1)$$

$$f_R(x) = \begin{cases} 1 = \frac{1}{b-a} & 0 \leq x \leq 1 \\ 0 & \text{o.w} \end{cases}$$

So, we have $F_R(x) = x$ and

$$E(R) = \int_0^1 x \cdot 1\, dx = \frac{x^2}{2}\Big]_0^1 = \frac{1}{2} = \frac{b+a}{2}$$

$$Var(R) = E(R^2) - E^2(R) = \int_0^1 x^2 dx - E^2(R)$$

$$= \frac{x^3}{3}\Big]_0^1 - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} = \frac{b-a}{12}$$

We use the term pseudo random numbers to indicate that they're not truly random (since they have a matematical sequence) but they act as if they are random. By Randomness in terms of a simulation process, we mean (i) Uniformity over $(0,1)$
(ii) Independence from the previous values drawn.

We'll see that (at least intuitively) the generated numbers have cycles. To make a simulation statistically acceptable, we need long cycles because when we start to have some random numbers, they're not random ANY MORE!

Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

Linear Congruential Method;

Given the initial random number X_0; which is called the seed, pseudo random numbers are generated as follows;

$$X_{i+1} = (a \cdot X_i + c) \mod m \quad i = 0, 1, 2, \ldots$$

multiplier (a), increment (c), modulus (m)

If $c \neq 0$, the form is called mixed congruential method

If $c = 0$, the form is called multiplicative congruential method

Random Numbers are generated by: $R_i = \frac{X_i}{m}$; $i = 1, 2, \ldots$

6. Use the multiplicative congruential method to generate a sequence of four three-digit random integers. Let $X_0 = 117$, $a = 43$, and $m = 1000$.

10. Use the mixed congruential method to generate a sequence of three two-digit random numbers with $X_0 = 37$, $a = 7$, $c = 29$, and $m = 100$.

11. Use the mixed congruential method to generate a sequence of three two-digit random integers between 0 and 24 with $X_0 = 13$, $a = 9$, and $c = 35$.

6) $X_0 = 117$; $a = 43$; $m = 1000$

$X_1 = 43 \cdot 117 = 5031 = 31 \pmod{1000}$ $\quad R_1 = 0{,}031$

$X_2 = 43 \cdot 31 = 1333 = 333 \pmod{1000}$ $\quad R_2 = 0{,}333$

$X_3 = 43 \cdot 333 = 14319 = 319 \pmod{1000}$ $\quad R_3 = 0{,}319$

$X_4 = 43 \cdot 319 = 13717 = 717 \pmod{1000}$ $\quad R_4 = 0{,}717$

10) $X_0 = 37;\ a = 7;\ c = 29;\ m = 100$

$X_1 = 7.37 + 39 = 298 = 98 \pmod{100}$ $R_1 = 0,98$

$X_2 = 7.98 + 39 = 725 = 25 \pmod{100}$ $R_2 = 0,25$

$X_3 = 7.25 + 39 = 214 = 14 \pmod{100}$ $R_3 = 0,14$

11) $X_0 = 13;\ a = 9;\ c = 35;$ Between 0 and 24 $\Rightarrow$ $m = 25$!

$X_1 = 9.13 + 35 = 152 = 2 \pmod{25}$ $R_1 = \frac{2}{25} = 0,08$

$X_2 = 9.2 + 35 = 53 = 3 \pmod{25}$ $R_2 = \frac{3}{25} = 0,12$

$X_3 = 9.3 + 35 = 62 = 12 \pmod{25}$ $R_3 = \frac{12}{25} = 0,48$

TESTS FOR RANDOM NUMBERS

(i) Frequency Tests (for $R_i \sim Uniform(0;1)$)

* THE Kolmogorov – Smirnov test (KS test)

Compare the continuous cdf $F(x)$ of uniform distribution with empirical cdf $S_N(x)$ of the sample of N observations.

Remember; $F(x) = x \quad 0 \le x \le 1$

Empirical cdf is; $S_N(x) = \dfrac{\#\text{ of } R_1, R_2, \ldots, R_n \le x}{N}$

Test statistics is $D = |F(x) - S_N(x)|$

where $R_{(1)} \le R_{(2)} \le \cdots \le R_{(N)}$ and

$$D^+ = \max_{1 \le i \le N} \left\{ \frac{i}{N} - R_i \right\} \qquad D^- = \max_{1 \le i \le N} \left\{ R_i - \frac{i-1}{N} \right\}$$

$$D = \max (D^+; D^-)$$

Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

Example: let 5 numbers 0,44 ; 0,81 ; 0,14 ; 0,05 ; 0,93 were generated. Test for uniformity using KS test (using 5% sig. level)

Answer: (i) H_0, H_A and α

H_0: $R_i \sim$ Uniform $(0;1)$

H_A: R_i does NOT follow Uniform distribution.

$\alpha = 0{,}05$

(ii) Test Statistics;

$D = \max(D^+; D^-)$; $N = 5$

(iii) Decision Criteria;

Reject H_0 if $D > D_{0{,}05} = 0{,}565$

(iv) Calculation

	i: 1	2	3	4	5
Step 1: $R_{(i)}$	0,05 ≤	0,14 ≤	0,44 ≤	0,81 ≤	0,93
Step 2: i/N (0)	0,20	0,40	0,60	0,80	1,00
$i/N - R_{(i)}$	0,15	(0,26)	0,16	—	0,07
$R_{(i)} - \frac{(i-1)}{N}$	0,05	—	0,04	(0,21)	0,13

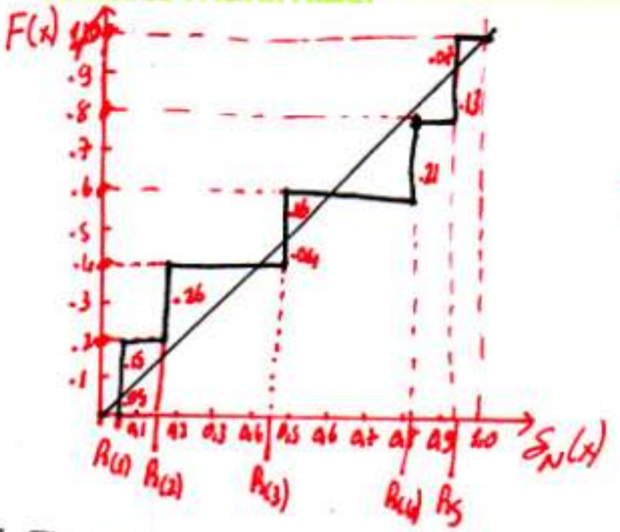
$D^+ = \max\left\{\frac{i}{n} - R_{(i)}\right\} = 0{,}26$ $\quad D^- = \max\left\{R_{(i)} - \frac{(i-1)}{n}\right\} = 0{,}21$

Step 3: $D = \max\{D^+; D^-\} = \underline{0{,}26}$

(v) Decision and Conclusion:

$0{,}26 \leq 0{,}565 \Rightarrow$ Do NOT Reject H_0.

Data does NOT indicate Non-Uniformity at $\alpha = 0{,}05$.



⇒ Graph for $F(x)$ against $S_N(x)$
(Values of $\frac{i}{n} - R_{(i)}$ and $R_{(i)} - \frac{i-1}{n}$

7. The sequence of numbers 0.54, 0.73, 0.98, 0.11, and 0.68 has been generated. Use the Kolmogorov–Smirnov test with $\alpha = 0.05$ to learn whether the hypothesis that the numbers are uniformly distributed on the interval [0, 1] can be rejected.

(iv) Calculation (First 3 steps are the same)

i:	1	2	3	4	$5=N$
$R_{(i)}$	0,11	0,54	0,68	0,73	0,98
i/N	0,20	0,40	0,60	0,80	1,00
$i/N - R_{(i)}$	(0,09)	—	—	0,07	0,02
$R_{(i)} - \frac{(i-1)}{N}$	0,11	0,34	(0,28)	0,13	0,18

$D = Max(0{,}09\ ;\ 0{,}28) = 0{,}28$

(v) $0{,}28 \leq 0{,}565 \Rightarrow$ Do NOT Reject H_0.

· * THE Chi-Square Test

Remember the Goodness of Fit test.
If we divide the interval $(0;1)$ into n equal classes, in each class, we expect $E_i = \frac{N}{n}$ Random Numbers.

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi_{n-1}^2$$

where O_i: Observed # of random #'s in i^{th} interval.

For χ^2 test to be reliable, we need to have $E_i \geq 5$ for each class. For that reason, χ^2 test is usually valid for large samples, say $N \geq 50$.

Example: Consider the following data set divided into $10 = n$ intervals: $[0;0,1), [0,1;0,2), \ldots, [0,9;1,0)$

We have; $N=100$ Random Numbers and $E_i = \frac{100}{10} = 10$.

O_i are given in the table:

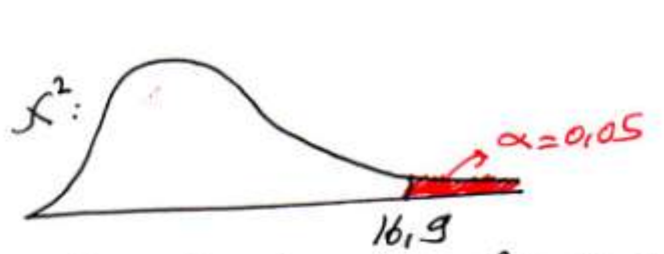
Interval: i	1	2	3	4	5	6	7	8	9	10 (= n)
O_i	8	8	10	9	12	8	10	14	10	11
E_i	10	10	10	10	10	10	10	10	10	10

(i) H_0: $R_i \sim$ Uniform $(0;1)$

H_A: The data set is NOT uniformly distributed

$\alpha = 0,05$

(ii) $\chi^2 = \frac{(O_i - E_i)^2}{E_i}$; $df = n-1 = 10-1 = 9$

(iii) 

Reject H_0 if $\chi^2 > 16,9$

(iv) $\chi^2 = \frac{(8-10)^2}{10} + \frac{(8-10)^2}{10} + \cdots + \frac{(11-10)^2}{10} = 3,4$

(v) $3,4 \leq 16,9$ so we do NOT Reject H_0. The numbers can be assumed to follow a Uniform distribution over 0 and 1.

(ii) Tests for Autocorrelation (for independence)

Remember, ρ: Correlation Coefficient stands for linear dependence of Random variables where $-1<\rho<1$. If $\rho=0$, we say that the random variables are uncorrelated meaning that there's NO linear relationship.

The term autocorrelation is usually used for time series data. We use autocorrelation to detect a systematic pattern of the data, which means that data values are NOT independent. The motivation of making an autocorrelation test to random numbers is: we want to be sure that "periodic numbers are NOT related."

Let, we want to test the autocorrelation between every m numbers (m is also called lag), say every $m=5$ numbers: $5^{th}, 10^{th}, 15^{th}, \ldots$. In general, our interest is on the numbers: $R_i, R_{i+m}, R_{i+2m}, \ldots, R_{i+(M+1)m}$

where M is the largest integer s.t. $i+(M+1)m \leq N$

Our test is:

$$H_0: \rho_{im} = 0$$
$$H_A: \rho_{im} \neq 0$$

with test statistics $Z_0 = \dfrac{\hat{\rho}_m}{\sigma_{\hat{\rho}_{im}}}$

where

$$\hat{\rho}_{im} = \frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+km} \cdot R_{i+(k+1)m}\right] - 0{,}25$$

$$\sigma_{\hat{\rho}_{im}} = \frac{\sqrt{13M+7}}{12(M+1)}$$

Example: Consider the following data set. Test for wheather the 3^{rd}, 8^{th}, 13^{th}, ... numbers in the sequence are autocorrelated:

0,12	0,01	0,23	0,28	0,89	0,31	0,64	0,28	0,83	0,93
0,99	0,15	0,33	0,35	0,91	0,41	0,60	0,27	0,75	0,88
0,68	0,49	0,05	0,43	0,95	0,58	0,19	0,36	0,69	0,87

Answer: $3+(M+1).5 \leq 30$ since we have;

$M = 4$ $\quad$ $i=3$, $m=5$ and $N=30$

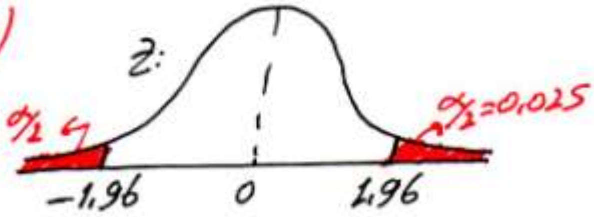
(i) $H_0: \rho_{35} = 0$

$H_A: \rho_{35} \neq 0$

$\alpha = 0,05$

(ii) $z = \dfrac{\hat{\rho}_{35}}{\sigma_{\hat{\rho}_{35}}}$

(iii)



Reject H_0 if $|z| > 1.96$

(iv) $\hat{\rho}_{35} = \frac{1}{4+1} \cdot [0{,}23 \cdot 0{,}28 + 0{,}28 \cdot 0{,}33 + 0{,}33 \cdot 0{,}27 + 0{,}27 \cdot 0{,}05 + 0{,}05 \cdot 0{,}36] - 0{,}25$

$= -1{,}1945$

$$\sigma_{\hat{\rho}_{35}} = \frac{\sqrt{13 \cdot 4 + 7}}{12 \cdot (4+1)} = 0{,}1280$$

$$z = \frac{-1{,}1945}{0{,}1280} = -1{,}516$$

(v) $|-1{,}516| \leq 1{,}96$, do NOT reject H_0. NO evidence for autocorrelation of type ρ_{35} at $\alpha = 0{,}05$.

20. Test the following sequence of numbers for uniformity and independence, using procedures you learned in this chapter: 0.594, 0.928, 0.515, 0.055, 0.507, 0.351, 0.262, 0.797, 0.788, 0.442, 0.097, 0.798, 0.227, 0.127, 0.474, 0.825, 0.007, 0.182, 0.929, 0.852.

20) I'll NOT conduct a KS test.

N=20 is NOT suitable for χ^2 test but I'll apply it to show the counting. let n=4 (To be sure $E_i = \frac{20}{4} = 5 \geq 5$) and then, calculation and counting of χ^2 is as follows:

interval	Tally	O_i	E_i	$(O_i - E_i)^2$
$0 \leq R_i < 0{,}25$	//////	6	5	1
$0{,}25 \leq R_i < 0{,}50$	////	4	5	1
$0{,}50 \leq R_i < 0{,}75$	///	3	5	4
$0{,}75 \leq R_i \leq 1{,}00$	///////	7	5	+ 4
				10

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{1}{5} \cdot 10 = 2$$

(i) $H_0: R_i \sim \text{Uniform}(0;1)$

H_A: NOT H_0

$\alpha = 0{,}05$

(ii) $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$; $df = 4-1=3$

(iii) 

Reject H_0 if $\chi^2 > 7{,}81$

(iv) $\chi^2 = 2$

(v) Do NOT Reject H_0.

To illustrate autocorrelation test, let i=0 and m=4 (N=20)

$(M+1) \cdot 4 \leq 20 \Rightarrow M = 4$ $\quad R_{i+jm}$: 0,055, 0,797, 0,798, 0,825, 0,852

(iv) (First 3 steps are the same)

$$\hat{\rho}_{04} = \frac{1}{4+1}\left[0{,}655 \cdot 0{,}797 + 0{,}797 \cdot 0{,}798 + 0{,}798 \cdot 0{,}825 + 0{,}825 \cdot 0{,}852\right] = 0{,}504$$

$$\hat{\sigma}_{\rho_{04}} = \frac{\sqrt{13 \cdot 4 + 7}}{12(4+1)} = 0{,}1280; \quad Z = \frac{0{,}504}{0{,}1280} = 3{,}94$$

(v) $|3{,}94| > 1{,}96$ so we Reject H_0. Data is NOT independent at $\alpha = 0{,}05$.