MATH 7241: Problem Set #3

Due date: Friday October 7

Reading: relevant background material for these problems can be found in the class notes, and in Ross (Chapters 2,3,5) and in Grinstead and Snell (Chapters 1,2 3, 6).

Exercise 1 Let $N, X_1, X_2, \dots$ be independent random variables, where the X_k are identically distributed with $\mathbb{E}[X_k] = \mu$ and $\mathrm{VAR}[X_k] = \sigma^2$. Also $\mathrm{RAN}(N) = \{1, 2, 3, \dots\}$, and both mean and variance of N are finite. Show that

$$\mathrm{VAR}\left(\sum_{k=1}^{N} X_k\right) = \sigma^2\, \mathbb{E}[N] + \mu^2\, \mathrm{VAR}[N]$$

[Hint: let $Y = \sum_{k=1}^{N} X_k$, and note that $\mathrm{VAR}[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$. Use conditioning on N to compute $\mathbb{E}[Y^2]$, and the use the result from class about $\mathbb{E}[Y]$].

$\mathbb{E}[Y^2 \mid N = n]$

$\Rightarrow \mathbb{E}[Y^2] = \mathbb{E}[\mathbb{E}[Y^2 \mid N]]$

$\mathbb{E}[W^2] = \mathrm{VAR}[W] + \mathbb{E}[W]^2$

Exercise 2 Suppose X has a uniform distribution on $[0, 2]$, and Y is uniformly distributed on $[0, X^2]$. Find the expected value of XY.

$$\mathbb{E}[XY \mid X = x]$$

$$\mathbb{E}[XY] = \int_0^2 \mathbb{E}[XY \mid X = x] \, f_X(x) \, dx$$

Exercise 3 Let X be an exponential random variable with rate λ, and let $t > 0$ be a fixed number.

a) Use the memoryless property to compute

$$\mathbb{E}[X \mid X > t]$$

b) By combining the result of part (a) with the total probability probability formula for $\mathbb{E}[X]$, compute

$$\mathbb{E}[X \mid X \leq t]$$

$X > t \Rightarrow X = t + X'$

$\mathbb{E}[X] = \mathbb{E}[X \mid A]\,\mathbb{P}(A) + \mathbb{E}[X \mid A^c]\,\mathbb{P}(A^c)$

$\mathbb{P}(X = x) = \mathbb{P}(X = x \mid A)\,\mathbb{P}(A) + \mathbb{P}(X = x \mid A^c)\,\mathbb{P}(A^c)$

Exercise 4 Let X_n be the symmetric random walk starting at 0. In class we derived the formula

$$P(X_n = n - 2k) = \frac{n!}{(n-k)!\,k!}\,2^{-n}, \quad k = 0, 1, 2, \ldots, n.$$

Define $x = k/n$ and use Stirling's formula $n! \sim n^n\sqrt{2\pi n}e^{-n}$ to show that when n and k are both large, the following asymptotic formula holds:

$$P(X_n = n - 2k) \simeq \frac{1}{\sqrt{2\pi n x(1-x)}}\,e^{-n[\log 2 - h(x)]}$$

where

$$h(x) = -x\,\log x - (1-x)\,\log(1-x)$$

$$n! = n^n\sqrt{2\pi n}\,e^{-n}$$

Exercise 5 Suppose that $\{X_i\}$ are IID uniform random variables on the interval $[-1, 1]$. Let Z be a standard normal random variable. Using the CLT, find the number a so that

$$\lim_{n\to\infty} P\Big(\sum_{i=1}^{n} X_i \geq \sqrt{n}\Big) = P(Z \geq a)$$

[Hint: you will need to find the mean and variance of X, which is uniform on $[-1, 1]$].

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = Z.$$

Exercise 6 A random variable can take values $\{1, 2, 3, 4\}$. The observed frequencies from 200 measurements are $\{85, 70, 25, 20\}$ respectively. The null hypothesis is

$$H_0 : p_1 = 0.4,\ p_2 = 0.3,\ p_3 = 0.2,\ p_4 = 0.1$$

and the alternative hypothesis is that these are not the probabilities. Use goodness of fit and the chi-square distribution to test H_0 at the 1% significance level: find the expected frequencies, find the number of degrees of freedom, find the critical value from the tables, state the decision rule, find the test statistic of the data, and state your conclusion.

Simple Random Walk

The simple random walk starts at zero, and at each step moves one unit either forwards (up) or backwards (down). We write X_n for the position of the random walk after n steps. So the possible values of X_n are

$$Ran(X) = \{-n, -n+2, \ldots, n-2, n\}$$

The pdf of X_n is the list of probabilities

$$\mathbb{P}(X = n-2k), \quad k = 0, 1, \ldots, n$$

These can be calculated using the binomial distribution as

$$\mathbb{P}(X = n-2k) = \mathbb{P}(k \text{ down}, n-k \text{ up}) = \binom{n}{k} 2^{-n}$$

$\mathbb{E}[X_n] = 0$

$\mathbb{E}[|X_n|]$ – average distance moved.

$\sqrt{\mathbb{E}[X_n^2]}$ – root mean square distance after n steps.

$$S_k = \begin{cases} +1 & \text{if } k^{\text{th}} \text{ step to right} \\ -1 & \text{if } k^{\text{th}} \text{ step to left.} \end{cases}$$

$$X_n = S_1 + S_2 + S_3 + \cdots + S_n.$$

↑ independent →

IID

S_k	1	-1
prob.	$\frac{1}{2}$	$\frac{1}{2}$

$$\mathbb{E}[S_k] = 1\left(\tfrac{1}{2}\right) - 1\left(\tfrac{1}{2}\right) = 0$$

$$VAR[S_k] = \mathbb{E}[S_k^2] - \left(\mathbb{E}[S_k]\right)^2$$

$$= (1)^2\left(\tfrac{1}{2}\right) + (-1)^2\tfrac{1}{2}$$

$$= 1$$

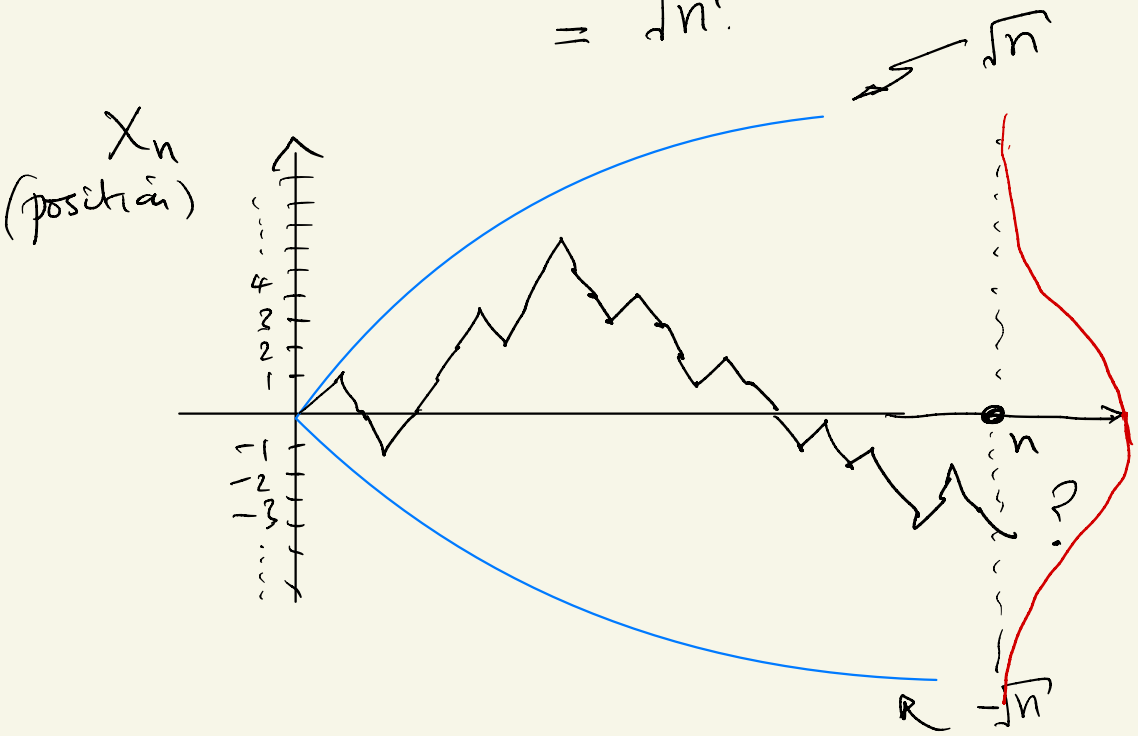
$$\Rightarrow \mathbb{E}[X_n] = 0$$

$$VAR[X_n] = n$$

$$\Rightarrow \text{r.m.s. distance} = \sqrt{\mathbb{E}[X_n^2]}$$

$$= \sqrt{\text{VAR}[X_n]}$$

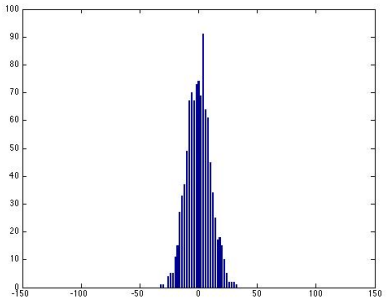
$$= \sqrt{n}.$$

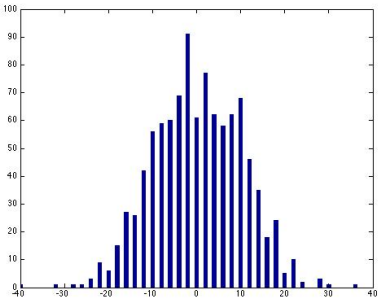


Random walk increases as $\sqrt{n}$, so velocity $\to 0$ as $n \to \infty$.

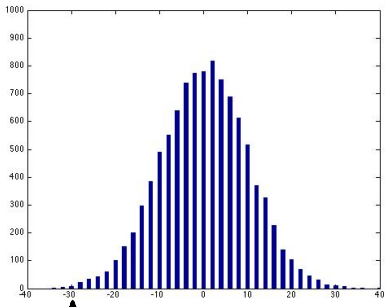
This is called diffusion.

But first let's do an experiment. Repeat the random walk many times and compute the long-run fraction of occurrences of each possible value of X_n. This is called the empirical pdf of X_n. Here $n = 100$ steps and sample 1000 times.





Note that the re-scaled pdf has a definite shape. Here is a longer run, where $n = 100$ and we sample $10,000$ times.



how likely? expect about 10 trials to end up below -30.

X_{100} is very well approximated as a normal r.v.

The pdf looks like a normal curve. We can explain this observation using the CLT. The value $X_n = n - 2k$ is achieved by k down steps and $n - k$ up steps. So we can write

$$X_n = S_1 + S_2 + \cdots + S_n$$

where

$$S_k = \begin{cases} 1 & \text{if } k^{th} \text{ step goes up} \\ -1 & \text{if } k^{th} \text{ step goes down} \end{cases}$$

Each step S_k is an independent random variable with a very simple pmf:

$$\mathbb{P}(S_k = 1) = \mathbb{P}(S_k = -1) = \frac{1}{2}$$

Easily calculate that

$$\mathbb{E}[S_k] = 0, \quad \text{VAR}[S_k] = 1$$

Therefore since $X_n = S_1 + \cdots + S_n$ we get

$$\mathbb{E}[X_n] = 0, \quad \text{VAR}[X_n] = n$$

Note that the S_k are IID, so we can apply the CLT. In this case

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i = n^{-1/2} X_n$$

$$\frac{\sum_{i=1}^{n} S_i - n(0)}{\sqrt{n}} = Z$$

and therefore the CLT says

$$Z_n \to Z, \quad X_n \simeq n^{1/2} Z$$

So for large n, the position X_n is a rescaled standard normal. This explains the graphs shown above.

$n = 100.$

$$\mathbb{P}(X_{100} < -30)$$

$$\text{(CLT)} \quad = \mathbb{P}(\sqrt{100}\, Z < -30)$$

$$= \mathbb{P}(Z < -3)$$

$$= \mathbb{P}(Z > 3) = 1 - \mathbb{P}(Z \leq 3)$$

$$= 1 - 0.9987 = 0.0013$$

Tables of the Normal Distribution

Probability Content from -oo to Z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Far Right Tail Probabilities

Z	P{Z to oo}	Z	P{Z to oo}	Z	P{Z to oo}	Z	P{Z to oo}
2.0	0.02275	3.0	0.001350	4.0	0.00003167	5.0	2.867 E-7
2.1	0.01786	3.1	0.0009676	4.1	0.00002066	5.5	1.899 E-8
2.2	0.01390	3.2	0.0006871	4.2	0.00001335	6.0	9.866 E-10
2.3	0.01072	3.3	0.0004834	4.3	0.00000854	6.5	4.016 E-11
2.4	0.00820	3.4	0.0003369	4.4	0.000005413	7.0	1.280 E-12
2.5	0.00621	3.5	0.0002326	4.5	0.000003398	7.5	3.191 E-14
2.6	0.004661	3.6	0.0001591	4.6	0.000002112	8.0	6.221 E-16
2.7	0.003467	3.7	0.0001078	4.7	0.000001300	8.5	9.480 E-18
2.8	0.002555	3.8	0.00007235	4.8	7.933 E-7	9.0	1.129 E-19
2.9	0.001866	3.9	0.00004810	4.9	4.792 E-7	9.5	1.049 E-21

These tables are public domain.
They are produced by APL programs written by the author,
William Knight

χ^2 distribution (Chi-Square)

Normal random variables are important because combinations of them arise in many statistical tests. Suppose that $Z_1, \ldots, Z_m$ are IID standard normal random variables, and suppose that

$$X = Z_1^2 + \cdots + Z_m^2$$

Then the random variable X is said to have a *chi-square* χ^2 *distribution with* m *degrees of freedom.* We indicate this by

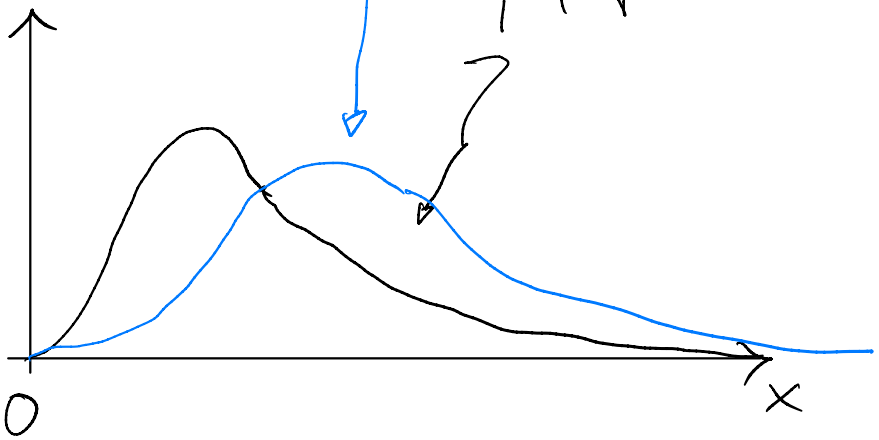
$$X \sim \chi^2(m), \quad m = df = \text{degrees of freedom}$$

χ^2 chi square

larger $m \Rightarrow$ further to the right

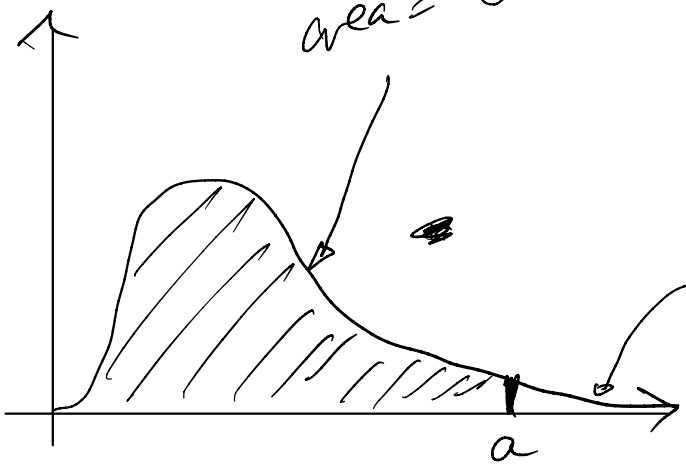
$X > 0$

pdf of X



m = integer = degrees of freedom

Suppose $X \sim \chi^2(5)$ ← degrees of freedom



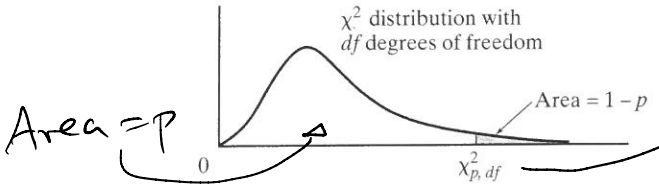
Find the number a so that

$$P(X \leq a) = 0.95$$

Tables. $a = \chi^2_{0.95,5}$

$= 11.07$

Table A.3 Upper and Lower Percentiles of χ^2 Distributions



Area = p

$\chi^2_{p,df}$

p = 0.95

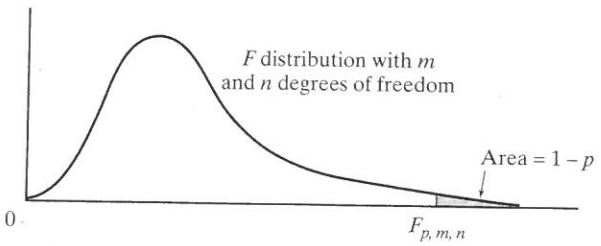
df = 5

	p							
df	0.010	0.025	0.050	0.10	0.90	0.95	0.975	0.99
1	0.000157	0.000982	0.00393	0.0158	2.706	3.841	5.024	6.635
2	0.0201	0.0506	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.336	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289
23	10.196	11.688	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892
31	15.655	17.539	19.281	21.434	41.422	44.985	48.232	52.191
32	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486
33	17.073	19.047	20.867	23.110	43.745	47.400	50.725	54.776
34	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061

Table A.3 Upper and Lower Percentiles of χ^2 Distributions (*cont.*)

	p							
df	0.010	0.025	0.050	0.10	0.90	0.95	0.975	0.99
35	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342
36	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619
37	19.960	22.106	24.075	26.492	48.363	52.192	55.668	59.892
38	20.691	22.878	24.884	27.343	49.513	53.384	56.895	61.162
39	21.426	23.654	25.695	28.196	50.660	54.572	58.120	62.428
40	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691
41	22.906	25.215	27.326	29.907	52.949	56.942	60.561	64.950
42	23.650	25.999	28.144	30.765	54.090	58.124	61.777	66.206
43	24.398	26.785	28.965	31.625	55.230	59.304	62.990	67.459
44	25.148	27.575	29.787	32.487	56.369	60.481	64.201	68.709
45	25.901	28.366	30.612	33.350	57.505	61.656	65.410	69.957
46	26.657	29.160	31.439	34.215	58.641	62.830	66.617	71.201
47	27.416	29.956	32.268	35.081	59.774	64.001	67.821	72.443
48	28.177	30.755	33.098	35.949	60.907	65.171	69.023	73.683
49	28.941	31.555	33.930	36.818	62.038	66.339	70.222	74.919
50	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154

Source: *Scientific Tables*, 6th ed. (Basel, Switzerland: J.R. Geigy, 1962), p. 36.



The figure above illustrates the percentiles of the F distributions shown in Table A.4. Table A.4 is used with permission from Wilfrid J. Dixon and Frank J. Massey, Jr., *Introduction to Statistical Analysis* 2nd ed. (New York: McGraw-Hill, 1957), pp. 389–404.

Properties of chi-square: it is a sum if independent r.v.'s, so

$$\begin{aligned} \mathbb{E}[X] &= m\,\mathbb{E}[Z^2] = m \\ \mathrm{VAR}[X] &= m\,\mathrm{VAR}[Z^2] = 2m \\ M_X(t) &= (M_{Z^2}(t))^m = (1-2t)^{-m/2} \quad \text{for } t < 1/2 \end{aligned}$$

It is also clear that a sum of independent chi-squares is again chi-square:

$$X \sim \chi^2(m),\ Y \sim \chi^2(n) \Rightarrow X + Y \sim \chi^2(m+n)$$

Use tables (or calculator) to compute probabilities for chi-square variables.

$X = Z_1^2 + Z_2^2 + \cdots + Z_m^2$

$Y = W_1^2 + W_2^2 + \cdots + W_n^2$

$X + Y = Z_1^2 + Z_2^2 + \cdots + Z_m^2 + W_1^2 + \cdots + W_n^2 \sim \chi^2(m+n)$

The multinomial distribution is the joint pdf for the frequencies of different results in a sequence of random trials. For example, suppose that a die is rolled ten times, and the following outcomes recorded:

Outcome	1	2	3	4	5	6
Probability	1/6	1/6	1/6	1/6	1/6	1/6
Observed frequency	2	1	0	3	1	3

What is the probability of this result? The answer is

$$\frac{10!}{2!\,1!\,0!\,3!\,1!\,3!}\,(1/6)^{10} = 8.34 \times 10^{-4}$$

0! = 1

Here is the general result: suppose X takes values $x_1, \ldots, x_m$ with probabilities $p_1, \ldots, p_m$. Suppose X is measured N times, and let $R_1, \ldots, R_m$ be the numbers of times each value is measured, so $R_1 + \cdots + R_m = N$. Then the probability to get the frequencies $n_1, \ldots, n_m$ is

$$\mathbb{P}\left(R_1 = n_1, \ldots, R_m = n_m\right) = \frac{N!}{n_1! \cdots n_m!}\, p_1^{n_1} \ \ldots \ p_m^{n_m}$$

N = number of trials.
p_i = prob. to get $X = i$.
$\Rightarrow \mathbb{E}[\text{number of trials with } X = i] = Np_i$.

Goodness of fit distribution

The expected frequency for the ith value is $\mathbb{E}[R_i] = Np_i$. The goodness-of-fit random variable is defined as

$$D = \sum_{i=1}^{m} \frac{(R_i - N p_i)^2}{N p_i} = \sum_{i=1}^{m} \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i}$$

The main result is that D has a χ^2 distribution with $m-1$ degrees of freedom. That is,

$$D \sim \chi^2(m-1)$$

(The precise statement says that this holds true in the limit where $N \to \infty$.)

Die roll.

$N = 10$ rolls.

X	1	2	3	4	5	6
prob	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
Expected number	$\frac{5}{3}$	$\frac{5}{3}$	$\frac{5}{3}$	$\frac{5}{3}$	$\frac{5}{3}$	$\frac{5}{3}$
Observed number	2	1	0	3	1	3

H_0 : null hypothesis

H_0 : These are the probabilities for X

$$D = \frac{\left(2 - \frac{5}{3}\right)^2}{5/3} + \frac{\left(1 - \frac{5}{3}\right)^2}{5/3} + \frac{\left(0 - \frac{5}{3}\right)^2}{5/3} + \cdots$$

$$D = 4.4$$

We know that

$$D \sim \chi^2(6-1) = \chi^2(5)$$

Ask the question:

how likely that $\chi^2(5)$

would be equal to or

lager than 4.4 ?

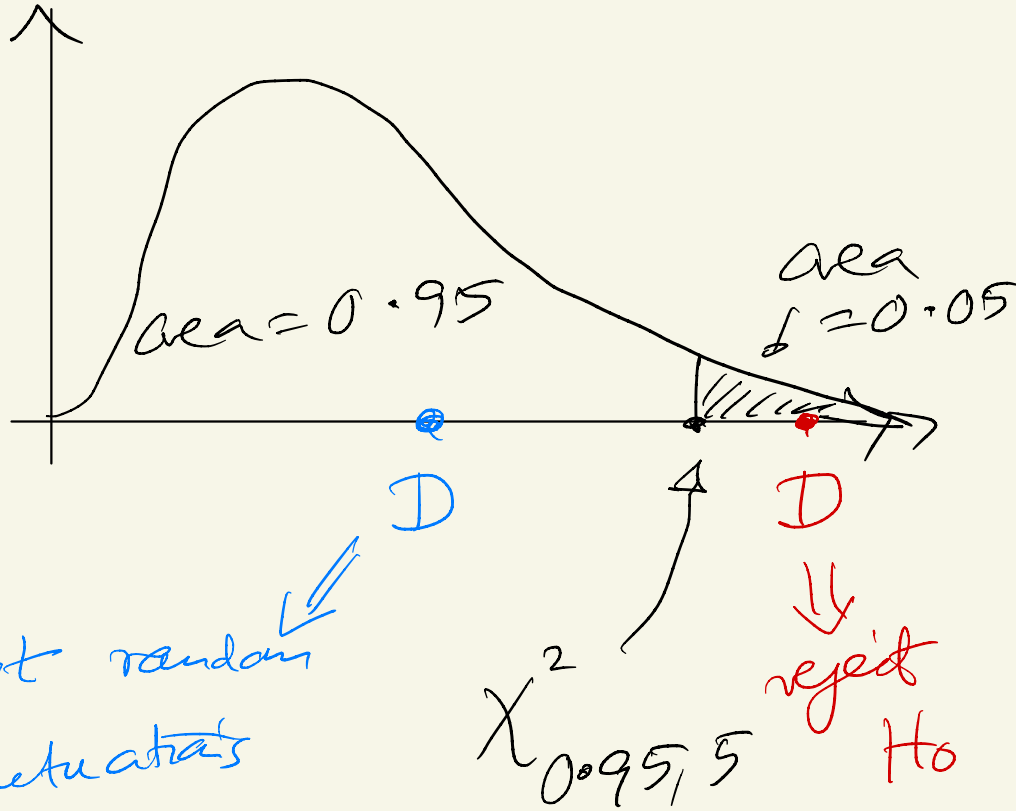
Chose a significance level

$$\alpha = 0.05$$

Decision rule:

reject H_0 if

$$D > \chi^2_{0.95,5}$$



do not reject H_0

What is our decision?

$$D = 4.4$$

$$\chi^2_{0.95,5} = 11.07$$

$\Rightarrow$ do not reject H_0

i.e. data is consistent with our model as a fair die.

Pearson's Goodness of fit Test

Suppose we have the sequence of observed frequencies $n_1, \ldots, n_m$ from N measurements, and we want to test the hypothesis that the measurements arise from the multinomial distribution with probabilities $p_1, \ldots, p_m$. We use D to design a statistical test based on the observed data.

$H_0 \longrightarrow$ **Null hypothesis** H_0: $R_1, \ldots, R_m$ are multinomial $p_1, \ldots, p_m$

Alternative hypothesis H_1: $R_1, \ldots, R_m$ are not multinomial $p_1, \ldots, p_m$.

We compute the test statistic using the observed data and the null hypothesis parameters:

$$d = \sum_{i=1}^{m} \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i} = \sum_{i=1}^{m} \frac{(n_i - N\,p_i)^2}{N\,p_i}$$

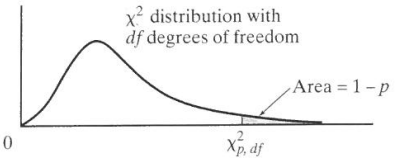
If H_0 is true then we know that d comes from a χ^2 distribution. So we can calculate the probability of finding a value as extreme as d, and use this to decide whether or not to reject the null hypothesis. The result is: we reject H_0 at significance level α if

$$d > \chi^2_{1-\alpha,m-1}$$

where the critical value $\chi^2_{1-\alpha,m-1}$ can be found from the chi-square tables. For example if $\alpha = 0.05$ (typical value) and $m = 10$ the value is

$$\chi^2_{0.95,9} = 16.919$$

Table A.3 Upper and Lower Percentiles of χ^2 Distributions



	p							
df	0.010	0.025	0.050	0.10	0.90	0.95	0.975	0.99
1	0.000157	0.000982	0.00393	0.0158	2.706	3.841	5.024	6.635
2	0.0201	0.0506	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.336	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289
23	10.196	11.688	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892
31	15.655	17.539	19.281	21.434	41.422	44.985	48.232	52.191
32	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486
33	17.073	19.047	20.867	23.110	43.745	47.400	50.725	54.776
34	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061

Example 4 *A die is rolled 30 times and the frequency of each outcome is recorded in the following table. Test at significance level* $\alpha = 0.05$ *to decide if the die is fair. So the null hypothesis* H_0 *is that each outcome has probability* $1/6$.

Outcome	*1*	*2*	*3*	*4*	*5*	*6*
Probability under H_0	1/6	1/6	1/6	1/6	1/6	1/6
Observed frequency	4	11	5	3	5	2
Expected frequency under H_0	5	5	5	5	5	5

$N = 30$.

$N p_i$

The test statistic is

$$d = \sum_{i=1}^{6} \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i} = 10$$

We have $\alpha = 0.05$ *and* $df = m - 1 = 5$ *so the critical value is*

$$\chi^2_{1-\alpha,df} = \chi^2_{0.95,5} = 11.07$$

Since $d < \chi^2_{1-\alpha,df}$ *we do* not *reject* H_0*, and so we accept that the die may be fair; that is we do* not *reject the null hypothesis.*

MTH 7241 Fall 2022: Prof. C. King

Notes 4: Finite Markov chains

A. A. Markov

Andrei Andreyevich Markov (1856 - 1922) founded the modern theory of stochastic processes, and gave his name to the special class we will consider here. He was an early activist for human rights in Imperial Russia, and in 1912 he protested Leo Tolstoy's excommunication from the Russian Orthodox Church. Markov was undaunted by extensive calculations and was very good at them. In what has now become the famous first application of Markov chains, A. A. Markov studied the sequence of 20,000 letters in A. S. Pushkins poem "Eugeny Onegin", discovering that the stationary vowel probability is $p = 0.432$, that the probability of a vowel following a vowel is $p_1 = 0.128$, and that the probability of a vowel following a consonant is $p_2 = 0.663$.

Markov chains

Markov's great contribution was the systematic analysis of a class of sequences of random variables $X_1, X_2, \ldots$ which are dependent, but only in the simplest possible way. In a sense they are the nearest to independent chains. Thus for example the sequence of random steps $S_1, S_2, \ldots$ is independent, while the random walk $\{X_k = S_1 + \cdots + S_k\}$ is not independent. However they are both Markov chains.

The theory has an enormous range of applications, including:

- statistical physics
- queueing theory
- communication networks
- voice recognition
- bioinformatics
- Google's pagerank algorithm
- computer learning and inference
- economics
- gambling
- data compression

Definition of the chain

Let Ω be a finite or countably infinite sample space. A collection of Ω-valued random variables $\{X_0, X_1, X_2, \dots\}$ is called a discrete-time Markov chain on Ω if it satisfies the *Markov condition*:

$$P(X_{n+1} = j \,|\, X_0 = i_0, \dots, X_n = i_n) = P(X_{n+1} = j \,|\, X_n = i_n) \tag{1}$$

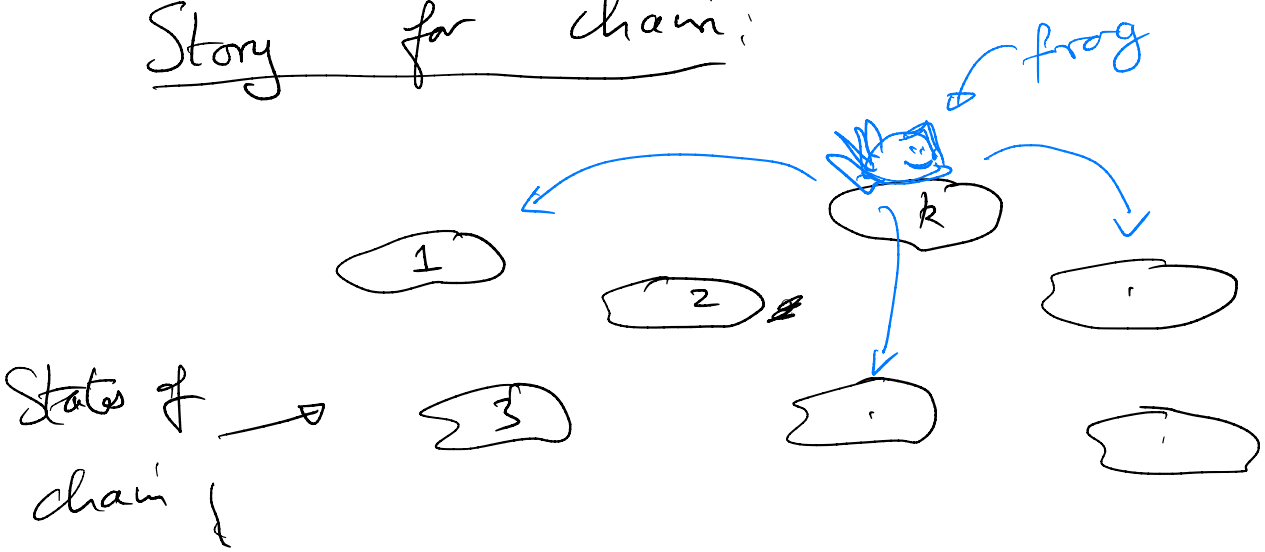
for all $n \geq 0$ and all states $j, i_0, \dots, i_n \in \Omega$.

Regarding the index of X_n as a discrete time the Markov condition can be summarized by saying that the conditional distribution of the future state X_{n+1} conditioned on the present and past states $X_0, \dots, X_n$ is equal to the conditional distribution of X_{n+1} conditioned on the present state X_n. In other words, the future (random) behavior of the chain only depends on where the chain sits right now, and not on how it got to its present position.

$X_0 \to X_1 \to X_2 \to X_3 \to \cdots$

- chain of random variables
- discrete time (one step at each tick of the clock).
- each X_i is a random variable taking values in the same state space Ω "omega"

Story for chain:



e.g. lilypads in a pad.

X_n = position of random jumps after n steps.

Transition mechanism: frog jumps to another lilypad at each time step. The new lilypad is chosen randomly.

Markov chain condition:

the frog has no memory!

next step = future

present position at i

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = k_{n-1}, X_{n-2} = k_{n-2}, \dots, X_0 = k_0)$$

past positions

$$= \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

Markov condition

$$= \mathbb{P}(X_1 = j \mid X_0 = i)$$

homogeneous chain

"clock starts over" at every jump.

We say that

$$P(X_{n+1} = j \,|\, X_n = i)$$

is the *transition probability* from state i to state j after n steps. We will mostly consider homogeneous chains, meaning that the transition probabilities do not depend on n. In this case for all n and $i, j \in \Omega$

$$P(X_{n+1} = j \,|\, X_n = i) = P(X_1 = j \,|\, X_0 = i) = p_{ij} \tag{2}$$

This defines the transition matrix P with entries p_{ij}. Any transition matrix P must satisfy these properties:

(P1) $p_{ij} \geq 0$ for all $i, j \in \Omega$

(P2) $\sum_{j \in \Omega} p_{ij} = 1$ for all $i \in \Omega$

Such a matrix is also called row-stochastic. So a square matrix is a transition matrix if and only if it is non-negative and row-stochastic.

An equivalent way to state the Markov condition is for any sequence of states $i, j, k, l, \ldots,$

$$P(X_0 = i, \ldots, X_1 = j, X_2 = k, X_3 = l, \ldots) = P(X_0 = i)\, p_{ij}\, p_{jk}\, p_{kl}, \ldots$$