Statistical and Mathematical Methods for Data Analysis

Dr. Syed Faisal Bukhari
Associate Professor
Department of Data Science
Faculty of Computing and Information Technology
University of the Punjab

Textbooks

- **Probability & Statistics for Engineers & Scientists**, Ninth Edition, Ronald E. Walpole, Raymond H. Myer

- **Elementary Statistics: Picturing the World,** 6th Edition, Ron Larson and Betsy Farber

- **Elementary Statistics**, 13th Edition, Mario F. Triola

Reference books

- **Probability and Statistical Inference, Ninth Edition,** Robert V. Hogg, Elliot A. Tanis, Dale L. Zimmerman
- **Probability Demystified**, Allan G. Bluman
- **Schaum's Outline of Probability,** Second Edition, Seymour Lipschutz, Marc Lipson
- **Python for Probability, Statistics, and Machine Learning,** José Unpingco
- **Practical Statistics for Data Scientists: 50 Essential Concepts,** Peter Bruce and Andrew Bruce
- **Think Stats: Probability and Statistics for Programmers,** Allen Downey

References

Readings for these lecture notes:

- **Probability & Statistics for Engineers & Scientists**, Ninth edition, Ronald E. Walpole, Raymond H. Myer

- **Elementary Statistics**, 10th Edition, Mario F. Triola

- **Probability Demystified**, Allan G. Bluman

These notes contain material from the above three books.

“Failure will never overtake me if my determination to succeed is strong enough.”

Og Mandino

Binomial Distribution [1]

A **binomial distribution** is obtained from a probability experiment called a binomial experiment. The experiment must satisfy these conditions:

1. Each trial can have only **two outcomes** or **outcomes that can be reduced to two outcomes**. The outcomes are usually considered as a success or a failure.
2. There is a **fixed number** of trials.
3. The outcomes of each trial are **independent** of each other.
4. The probability of a **success** must remain the **same** for each trial.

Binomial Distribution [3]

Example: Explain why the probability experiment of tossing three coins is a binomial experiment.

Binomial Distribution [2]

1. There are only **two outcomes** for each trial, head and tail. Depending on the situation, either heads or tails can be defined as a success and the other as a failure.
2. There is a **fixed number** of trials. In this case, there are three trials since three coins are tossed or one coin is tossed three times.
3. The outcomes are **independent** since tossing one coin does not effect the outcome of the other two tosses.
4. The probability of a success (say heads) is $\frac{1}{2}$ and it **does not change**. Hence the experiment meets the conditions of a binomial experiment.

Binomial Distribution [3]

The binomial probability formula is used to compute **probabilities for binomial random** variables. The binomial probability formula is given as:

$$b(x; n, p) = c_x^n p^x q^{n-x}, \; x = 0, 1, 2, \ldots$$

where $c_x^n = \dfrac{n!}{x!(n-x)!}$

OR

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \; x = 0, 1, 2, \ldots$$

where n = the total number of trials

x = the number of successes (0, 1, 2, 3, . . . , n)

p = the probability of a success

q = the probability of a failure

Binomial Distribution [4]

Example: A coin is tossed three times. Find the probability of getting **two heads** and **a tail** in any given order.

Binomial Distribution [5]

Solution:

Here n = 3

$p = \frac{1}{2}$ **(probability of a head)**

$\because p + q = 1$

$q = 1 – p = 1 - \frac{1}{2} = \frac{1}{2}$ **(probability of a tail)**

Let x denotes **number of heads**

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \; x = 0, 1, 2, \ldots$$

$$b(2; 3, 0.5) = \binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{3-2} = 3\left(\frac{1}{4}\right)\left(\frac{1}{2}\right) = \frac{3}{8}$$

Alternative approach

Solution:

Here n = 3

$p = \frac{1}{2}$ **(probability of a tail)**

$\because p + q = 1$

$q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$ **(probability of a head)**

Let x denotes **number of tails**

$b(x; n, p) = \binom{n}{x} p^x q^{n-x}$, x = 0, 1, 2, ...

$b(1; 3, 0.5) = \binom{3}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{3-1} = 3\left(\frac{1}{2}\right)\left(\frac{1}{4}\right) = \frac{3}{8}$

Binomial Distribution [6]

- It is the case of sampling with **replacement.**
- It has two parameters i.e., **n** and **p**.
- **Mean = np**, **Variance =npq** and Standard deviation = sqrt(npq).
- If p = q then the distribution is **symmetrical** irrespective of the value of n.
  - If **p < 0.5** the distribution is **negatively skewed**.
  - If **p > 0.5** then the distribution is **positively skewed**.

Binomial Distribution [7]

- The **normal distribution (Gaussian distribution)** can be regarded as the **limiting form** of the **binomial distribution** when **neither p nor q is very small** where n, the of trials is **very large** and.

Note: Because the normal approximation is not accurate for small values of n, a good rule of thumb is to use the normal approximation only if $np \geq 10$ and $np(1-p) \geq 10$.

Expected Value

The **mean of a random variable** represents what you would expect to happen over **thousands of trials**. It is also called the **expected value.**

Definition: The **expected value** of a **discrete random variable** is equal to the **mean of the random variable**.

Expected Value = E(x) = μ = $\sum xP(x)$

Note: Although **probabilities** can never be **negative**, the **expected value of a random variable** can be **negative**

Example At a raffle, **1500** tickets are sold at **$ 2** each for four prizes of **$ 500, $250, $150, and $75**. You buy one ticket. What is the expected value of your gain?

Solution

$E(x) = \sum xP(x)$

Gain (\$), *x*	P(x)	xP(x)
500 – 2 = 498	$\frac{1}{1500}$	$\frac{498}{1500}$
250 – 2 = 248	$\frac{1}{1500}$	$\frac{248}{1500}$
150 – 2 = 148	$\frac{1}{1500}$	$\frac{148}{1500}$
75 - 2 = 73	$\frac{1}{1500}$	$\frac{73}{1500}$
-2	$\frac{1496}{1500}$	$\frac{-2992}{1500}$
	$\sum P(x) = \frac{1500}{1500} = 1$	$\sum xP(x) = -1.35$

Interpretation Because the **expected value** is **negative**, you can expect to **lose an average of $1.35** for each ticket you buy.

The Mean and Standard Deviation for a Binomial Distribution [1]

Example: A die is tossed **180 times** and the number of threes obtained is recorded. Find the mean or expected number of threes.

The Mean and Standard Deviation for a Binomial Distribution [2]

Solution:

n = 180 and p = $\frac{1}{6}$

Mean = np = (180)($\frac{1}{6}$)

Mean = 30

Hence, one would expect on average 30 threes.

The Mean and Standard Deviation for a Binomial Distribution [3]

Example: Twelve cards are selected from a deck and each card is replaced before the next one is drawn. Find the average number of diamonds.

The Mean and Standard Deviation for a Binomial Distribution [4]

Solution:

In this case, n =12 and p = $\frac{13}{52}$

Mean = np = $(12)(\frac{13}{52})$

Mean = 3

The Mean and Standard Deviation for a Binomial Distribution [3]

Example: A die is rolled 180 times. Find the standard deviation of the number of threes.

The Mean and Standard Deviation for a Binomial Distribution [4]

Solution:

Standard deviation = $\sqrt{\mathrm{npq}}$

Standard deviation = $\sqrt{180\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)}$

The standard deviation is 5

The Mean and Standard Deviation for a Binomial Distribution [5]

Roughly speaking, most of the values fall within two standard deviations of the mean.

$\mu - 2\sigma <$ most values $< \mu + 2\sigma$

In the die example, we can expect most values will fall between

30 – 2*5 < most values < 30 + 2*5

20 < most values < 40

The Mean and Standard Deviation for a Binomial Distribution [6]

Example: An archer hits the bull's eye **80%** of the time. If he shoots **100** arrows, find the mean and standard deviation of the number of bull's eyes. If he travels to many tournaments, find the approximate range of values.

The Mean and Standard Deviation for a Binomial Distribution [6]

Solution:

$n = 100$, $p = 0.80$, $q = 1 - p = 1 - 0.80 = 0.20$

Mean $= \mu = np = 100 \times 0.80 = 80$

Standard Deviation $= \sigma = \sqrt{npq}$

$= \sqrt{100 \times 0.80 \times 0.20} = 4$

$\mu - 2\sigma <$ most values $< \mu + 2\sigma$

$80 - 8 <$ most values $< 80 + 8$

$72 <$ most values < 88

Example: The probability that a certain kind of component will survive a shock test is $\frac{3}{4} \cdot$ Find the probability that exactly **2** of the **next 4** components tested survive.

Solution:

Here n = 4

$p = \frac{3}{4}$ **(probability of survive a shock test)**

$\because p + q = 1$

$q = 1 - p = 1 - \frac{3}{4} = \frac{1}{4}$

Let x denotes number of components that survive a shock test

$b(x; n, p) = \binom{n}{x} p^x q^{n-x}$, x = 0, 1, 2, 3…,n

$b\left(2; 3, \frac{3}{4}\right) = \binom{4}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^{4-2} = \frac{27}{128} = 0.2109$

Table A.1 Binomial Probability Sums $\sum_{x=0}^{r} b(x;n,p)$

		p									
n	r	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
1	0	0.9000	0.8000	0.7500	0.7000	0.6000	0.5000	0.4000	0.3000	0.2000	0.1000
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500	0.1600	0.0900	0.0400	0.0100
	1	0.9900	0.9600	0.9375	0.9100	0.8400	0.7500	0.6400	0.5100	0.3600	0.1900
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250	0.0640	0.0270	0.0080	0.0010
	1	0.9720	0.8960	0.8438	0.7840	0.6480	0.5000	0.3520	0.2160	0.1040	0.0280
	2	0.9990	0.9920	0.9844	0.9730	0.9360	0.8750	0.7840	0.6570	0.4880	0.2710
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.6561	0.4096	0.3164	0.2401	0.1296	0.0625	0.0256	0.0081	0.0016	0.0001
	1	0.9477	0.8192	0.7383	0.6517	0.4752	0.3125	0.1792	0.0837	0.0272	0.0037
	2	0.9963	0.9728	0.9492	0.9163	0.8208	0.6875	0.5248	0.3483	0.1808	0.0523
	3	0.9999	0.9984	0.9961	0.9919	0.9744	0.9375	0.8704	0.7599	0.5904	0.3439
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0313	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.5314	0.2621	0.1780	0.1176	0.0467	0.0156	0.0041	0.0007	0.0001	0.0000
	1	0.8857	0.6554	0.5339	0.4202	0.2333	0.1094	0.0410	0.0109	0.0016	0.0001
	2	0.9842	0.9011	0.8306	0.7443	0.5443	0.3438	0.1792	0.0705	0.0170	0.0013
	3	0.9987	0.9830	0.9624	0.9295	0.8208	0.6563	0.4557	0.2557	0.0989	0.0159
	4	0.9999	0.9984	0.9954	0.9891	0.9590	0.8906	0.7667	0.5798	0.3446	0.1143
	5	1.0000	0.9999	0.9998	0.9993	0.9959	0.9844	0.9533	0.8824	0.7379	0.4686
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0	0.4783	0.2097	0.1335	0.0824	0.0280	0.0078	0.0016	0.0002	0.0000	
	1	0.8503	0.5767	0.4449	0.3294	0.1586	0.0625	0.0188	0.0038	0.0004	0.0000
	2	0.9743	0.8520	0.7564	0.6471	0.4199	0.2266	0.0963	0.0288	0.0047	0.0002
	3	0.9973	0.9667	0.9294	0.8740	0.7102	0.5000	0.2898	0.1260	0.0333	0.0027
	4	0.9998	0.9953	0.9871	0.9712	0.9037	0.7734	0.5801	0.3529	0.1480	0.0257
	5	1.0000	0.9996	0.9987	0.9962	0.9812	0.9375	0.8414	0.6706	0.4233	0.1497
	6		1.0000	0.9999	0.9998	0.9984	0.9922	0.9720	0.9176	0.7903	0.5217
	7			1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^{r} b(x;n,p)$

		p									
n	r	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
8	0	0.4305	0.1678	0.1001	0.0576	0.0168	0.0039	0.0007	0.0001	0.0000	
	1	0.8131	0.5033	0.3671	0.2553	0.1064	0.0352	0.0085	0.0013	0.0001	
	2	0.9619	0.7969	0.6785	0.5518	0.3154	0.1445	0.0498	0.0113	0.0012	0.0000
	3	0.9950	0.9437	0.8862	0.8059	0.5941	0.3633	0.1737	0.0580	0.0104	0.0004
	4	0.9996	0.9896	0.9727	0.9420	0.8263	0.6367	0.4059	0.1941	0.0563	0.0050
	5	1.0000	0.9988	0.9958	0.9887	0.9502	0.8555	0.6846	0.4482	0.2031	0.0381
	6		0.9999	0.9996	0.9987	0.9915	0.9648	0.8936	0.7447	0.4967	0.1869
	7		1.0000	1.0000	0.9999	0.9993	0.9961	0.9832	0.9424	0.8322	0.5695
	8				1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0	0.3874	0.1342	0.0751	0.0404	0.0101	0.0020	0.0003	0.0000		
	1	0.7748	0.4362	0.3003	0.1960	0.0705	0.0195	0.0038	0.0004	0.0000	
	2	0.9470	0.7382	0.6007	0.4628	0.2318	0.0898	0.0250	0.0043	0.0003	0.0000
	3	0.9917	0.9144	0.8343	0.7297	0.4826	0.2539	0.0994	0.0253	0.0031	0.0001
	4	0.9991	0.9804	0.9511	0.9012	0.7334	0.5000	0.2666	0.0988	0.0196	0.0009
	5	0.9999	0.9969	0.9900	0.9747	0.9006	0.7461	0.5174	0.2703	0.0856	0.0083
	6	1.0000	0.9997	0.9987	0.9957	0.9750	0.9102	0.7682	0.5372	0.2618	0.0530
	7		1.0000	0.9999	0.9996	0.9962	0.9805	0.9295	0.8040	0.5638	0.2252
	8			1.0000	1.0000	0.9997	0.9980	0.9899	0.9596	0.8658	0.6126
	9					1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000		
	1	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	
	2	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	
	3	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0064	0.0001
	5	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7		0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8		1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9				1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10					1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	0	0.3138	0.0859	0.0422	0.0198	0.0036	0.0005	0.0000			
	1	0.6974	0.3221	0.1971	0.1130	0.0302	0.0059	0.0007	0.0000		
	2	0.9104	0.6174	0.4552	0.3127	0.1189	0.0327	0.0059	0.0006	0.0000	
	3	0.9815	0.8389	0.7133	0.5696	0.2963	0.1133	0.0293	0.0043	0.0002	
	4	0.9972	0.9496	0.8854	0.7897	0.5328	0.2744	0.0994	0.0216	0.0020	0.0000
	5	0.9997	0.9883	0.9657	0.9218	0.7535	0.5000	0.2465	0.0782	0.0117	0.0003
	6	1.0000	0.9980	0.9924	0.9784	0.9006	0.7256	0.4672	0.2103	0.0504	0.0028
	7		0.9998	0.9988	0.9957	0.9707	0.8867	0.7037	0.4304	0.1611	0.0185
	8		1.0000	0.9999	0.9994	0.9941	0.9673	0.8811	0.6873	0.3826	0.0896
	9			1.0000	1.0000	0.9993	0.9941	0.9698	0.8870	0.6779	0.3026
	10					1.0000	0.9995	0.9964	0.9802	0.9141	0.6862
	11						1.0000	1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^{r} b(x; n, p)$

		p									
n	**r**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
12	**0**	0.2824	0.0687	0.0317	0.0138	0.0022	0.0002	0.0000			
	1	0.6590	0.2749	0.1584	0.0850	0.0196	0.0032	0.0003	0.0000		
	2	0.8891	0.5583	0.3907	0.2528	0.0834	0.0193	0.0028	0.0002	0.0000	
	3	0.9744	0.7946	0.6488	0.4925	0.2253	0.0730	0.0153	0.0017	0.0001	
	4	0.9957	0.9274	0.8424	0.7237	0.4382	0.1938	0.0573	0.0095	0.0006	0.0000
	5	0.9995	0.9806	0.9456	0.8822	0.6652	0.3872	0.1582	0.0386	0.0039	0.0001
	6	0.9999	0.9961	0.9857	0.9614	0.8418	0.6128	0.3348	0.1178	0.0194	0.0005
	7	1.0000	0.9994	0.9972	0.9905	0.9427	0.8062	0.5618	0.2763	0.0726	0.0043
	8		0.9999	0.9996	0.9983	0.9847	0.9270	0.7747	0.5075	0.2054	0.0256
	9		1.0000	1.0000	0.9998	0.9972	0.9807	0.9166	0.7472	0.4417	0.1109
	10				1.0000	0.9997	0.9968	0.9804	0.9150	0.7251	0.3410
	11					1.0000	0.9998	0.9978	0.9862	0.9313	0.7176
	12						1.0000	1.0000	1.0000	1.0000	1.0000
13	**0**	0.2542	0.0550	0.0238	0.0097	0.0013	0.0001	0.0000			
	1	0.6213	0.2336	0.1267	0.0637	0.0126	0.0017	0.0001	0.0000		
	2	0.8661	0.5017	0.3326	0.2025	0.0579	0.0112	0.0013	0.0001		
	3	0.9658	0.7473	0.5843	0.4206	0.1686	0.0461	0.0078	0.0007	0.0000	
	4	0.9935	0.9009	0.7940	0.6543	0.3530	0.1334	0.0321	0.0040	0.0002	
	5	0.9991	0.9700	0.9198	0.8346	0.5744	0.2905	0.0977	0.0182	0.0012	0.0000
	6	0.9999	0.9930	0.9757	0.9376	0.7712	0.5000	0.2288	0.0624	0.0070	0.0001
	7	1.0000	0.9988	0.9944	0.9818	0.9023	0.7095	0.4256	0.1654	0.0300	0.0009
	8		0.9998	0.9990	0.9960	0.9679	0.8666	0.6470	0.3457	0.0991	0.0065
	9		1.0000	0.9999	0.9993	0.9922	0.9539	0.8314	0.5794	0.2527	0.0342
	10			1.0000	0.9999	0.9987	0.9888	0.9421	0.7975	0.4983	0.1339
	11				1.0000	0.9999	0.9983	0.9874	0.9363	0.7664	0.3787
	12					1.0000	0.9999	0.9987	0.9903	0.9450	0.7458
	13						1.0000	1.0000	1.0000	1.0000	1.0000
14	**0**	0.2288	0.0440	0.0178	0.0068	0.0008	0.0001	0.0000			
	1	0.5846	0.1979	0.1010	0.0475	0.0081	0.0009	0.0001			
	2	0.8416	0.4481	0.2811	0.1608	0.0398	0.0065	0.0006	0.0000		
	3	0.9559	0.6982	0.5213	0.3552	0.1243	0.0287	0.0039	0.0002		
	4	0.9908	0.8702	0.7415	0.5842	0.2793	0.0898	0.0175	0.0017	0.0000	
	5	0.9985	0.9561	0.8883	0.7805	0.4859	0.2120	0.0583	0.0083	0.0004	
	6	0.9998	0.9884	0.9617	0.9067	0.6925	0.3953	0.1501	0.0315	0.0024	0.0000
	7	1.0000	0.9976	0.9897	0.9685	0.8499	0.6047	0.3075	0.0933	0.0116	0.0002
	8		0.9996	0.9978	0.9917	0.9417	0.7880	0.5141	0.2195	0.0439	0.0015
	9		1.0000	0.9997	0.9983	0.9825	0.9102	0.7207	0.4158	0.1298	0.0092
	10			1.0000	0.9998	0.9961	0.9713	0.8757	0.6448	0.3018	0.0441
	11				1.0000	0.9994	0.9935	0.9602	0.8392	0.5519	0.1584
	12					0.9999	0.9991	0.9919	0.9525	0.8021	0.4154
	13					1.0000	0.9999	0.9992	0.9932	0.9560	0.7712
	14						1.0000	1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^{r} b(x;n,p)$

		p									
n	*r*	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
15	**0**	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000				
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000			
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000		
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001		
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0093	0.0007	0.0000	
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0950	0.0152	0.0008	
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000
	8		0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9		0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0022
	10		1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11			1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12				1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13					1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14						1.0000	0.9995	0.9953	0.9648	0.7941
	15							1.0000	1.0000	1.0000	1.0000
16	**0**	0.1853	0.0281	0.0100	0.0033	0.0003	0.0000				
	1	0.5147	0.1407	0.0635	0.0261	0.0033	0.0003	0.0000			
	2	0.7892	0.3518	0.1971	0.0994	0.0183	0.0021	0.0001			
	3	0.9316	0.5981	0.4050	0.2459	0.0651	0.0106	0.0009	0.0000		
	4	0.9830	0.7982	0.6302	0.4499	0.1666	0.0384	0.0049	0.0003		
	5	0.9967	0.9183	0.8103	0.6598	0.3288	0.1051	0.0191	0.0016	0.0000	
	6	0.9995	0.9733	0.9204	0.8247	0.5272	0.2272	0.0583	0.0071	0.0002	
	7	0.9999	0.9930	0.9729	0.9256	0.7161	0.4018	0.1423	0.0257	0.0015	0.0000
	8	1.0000	0.9985	0.9925	0.9743	0.8577	0.5982	0.2839	0.0744	0.0070	0.0001
	9		0.9998	0.9984	0.9929	0.9417	0.7728	0.4728	0.1753	0.0267	0.0005
	10		1.0000	0.9997	0.9984	0.9809	0.8949	0.6712	0.3402	0.0817	0.0033
	11			1.0000	0.9997	0.9951	0.9616	0.8334	0.5501	0.2018	0.0170
	12				1.0000	0.9991	0.9894	0.9349	0.7541	0.4019	0.0684
	13					0.9999	0.9979	0.9817	0.9006	0.6482	0.2108
	14					1.0000	0.9997	0.9967	0.9739	0.8593	0.4853
	15						1.0000	0.9997	0.9967	0.9719	0.8147
	16							1.0000	1.0000	1.0000	1.0000

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^{r} b(x; n, p)$

		p									
n	**r**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
17	**0**	0.1668	0.0225	0.0075	0.0023	0.0002	0.0000				
	1	0.4818	0.1182	0.0501	0.0193	0.0021	0.0001	0.0000			
	2	0.7618	0.3096	0.1637	0.0774	0.0123	0.0012	0.0001			
	3	0.9174	0.5489	0.3530	0.2019	0.0464	0.0064	0.0005	0.0000		
	4	0.9779	0.7582	0.5739	0.3887	0.1260	0.0245	0.0025	0.0001		
	5	0.9953	0.8943	0.7653	0.5968	0.2639	0.0717	0.0106	0.0007	0.0000	
	6	0.9992	0.9623	0.8929	0.7752	0.4478	0.1662	0.0348	0.0032	0.0001	
	7	0.9999	0.9891	0.9598	0.8954	0.6405	0.3145	0.0919	0.0127	0.0005	
	8	1.0000	0.9974	0.9876	0.9597	0.8011	0.5000	0.1989	0.0403	0.0026	0.0000
	9		0.9995	0.9969	0.9873	0.9081	0.6855	0.3595	0.1046	0.0109	0.0001
	10		0.9999	0.9994	0.9968	0.9652	0.8338	0.5522	0.2248	0.0377	0.0008
	11		1.0000	0.9999	0.9993	0.9894	0.9283	0.7361	0.4032	0.1057	0.0047
	12			1.0000	0.9999	0.9975	0.9755	0.8740	0.6113	0.2418	0.0221
	13				1.0000	0.9995	0.9936	0.9536	0.7981	0.4511	0.0826
	14					0.9999	0.9988	0.9877	0.9226	0.6904	0.2382
	15					1.0000	0.9999	0.9979	0.9807	0.8818	0.5182
	16						1.0000	0.9998	0.9977	0.9775	0.8332
	17							1.0000	1.0000	1.0000	1.0000
18	**0**	0.1501	0.0180	0.0056	0.0016	0.0001	0.0000				
	1	0.4503	0.0991	0.0395	0.0142	0.0013	0.0001				
	2	0.7338	0.2713	0.1353	0.0600	0.0082	0.0007	0.0000			
	3	0.9018	0.5010	0.3057	0.1646	0.0328	0.0038	0.0002			
	4	0.9718	0.7164	0.5187	0.3327	0.0942	0.0154	0.0013	0.0000		
	5	0.9936	0.8671	0.7175	0.5344	0.2088	0.0481	0.0058	0.0003		
	6	0.9988	0.9487	0.8610	0.7217	0.3743	0.1189	0.0203	0.0014	0.0000	
	7	0.9998	0.9837	0.9431	0.8593	0.5634	0.2403	0.0576	0.0061	0.0002	
	8	1.0000	0.9957	0.9807	0.9404	0.7368	0.4073	0.1347	0.0210	0.0009	
	9		0.9991	0.9946	0.9790	0.8653	0.5927	0.2632	0.0596	0.0043	0.0000
	10		0.9998	0.9988	0.9939	0.9424	0.7597	0.4366	0.1407	0.0163	0.0002
	11		1.0000	0.9998	0.9986	0.9797	0.8811	0.6257	0.2783	0.0513	0.0012
	12			1.0000	0.9997	0.9942	0.9519	0.7912	0.4656	0.1329	0.0064
	13				1.0000	0.9987	0.9846	0.9058	0.6673	0.2836	0.0282
	14					0.9998	0.9962	0.9672	0.8354	0.4990	0.0982
	15					1.0000	0.9993	0.9918	0.9400	0.7287	0.2662
	16						0.9999	0.9987	0.9858	0.9009	0.5497
	17						1.0000	0.9999	0.9984	0.9820	0.8499
	18							1.0000	1.0000	1.0000	1.0000

n	r	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
						p					
19	0	0.1351	0.0144	0.0042	0.0011	0.0001					
	1	0.4203	0.0829	0.0310	0.0104	0.0008	0.0000				
	2	0.7054	0.2369	0.1113	0.0462	0.0055	0.0004	0.0000			
	3	0.8850	0.4551	0.2631	0.1332	0.0230	0.0022	0.0001			
	4	0.9648	0.6733	0.4654	0.2822	0.0696	0.0096	0.0006	0.0000		
	5	0.9914	0.8369	0.6678	0.4739	0.1629	0.0318	0.0031	0.0001		
	6	0.9983	0.9324	0.8251	0.6655	0.3081	0.0835	0.0116	0.0006		
	7	0.9997	0.9767	0.9225	0.8180	0.4878	0.1796	0.0352	0.0028	0.0000	
	8	1.0000	0.9933	0.9713	0.9161	0.6675	0.3238	0.0885	0.0105	0.0003	
	9		0.9984	0.9911	0.9674	0.8139	0.5000	0.1861	0.0326	0.0016	
	10		0.9997	0.9977	0.9895	0.9115	0.6762	0.3325	0.0839	0.0067	0.0000
	11		1.0000	0.9995	0.9972	0.9648	0.8204	0.5122	0.1820	0.0233	0.0003
	12			0.9999	0.9994	0.9884	0.9165	0.6919	0.3345	0.0676	0.0017
	13			1.0000	0.9999	0.9969	0.9682	0.8371	0.5261	0.1631	0.0086
	14				1.0000	0.9994	0.9904	0.9304	0.7178	0.3267	0.0352
	15					0.9999	0.9978	0.9770	0.8668	0.5449	0.1150
	16					1.0000	0.9996	0.9945	0.9538	0.7631	0.2946
	17						1.0000	0.9992	0.9896	0.9171	0.5797
	18							0.9999	0.9989	0.9856	0.8649
	19							1.0000	1.0000	1.0000	1.0000
20	0	0.1216	0.0115	0.0032	0.0008	0.0000					
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000				
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002				
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0000			
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003			
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000		
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003		
	7	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	
	8	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	
	9	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	
	10		0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000
	11		0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001
	12		1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004
	13			1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024
	14				1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113
	15					0.9997	0.9941	0.9490	0.7625	0.3704	0.0432
	16					1.0000	0.9987	0.9840	0.8929	0.5886	0.1330
	17						0.9998	0.9964	0.9645	0.7939	0.3231
	18						1.0000	0.9995	0.9924	0.9308	0.6083
	19							1.0000	0.9992	0.9885	0.8784
	20								1.0000	1.0000	1.0000

Binomial Distribution [1]

Example 1: The probability that a patient recovers from a rare blood disease is **0.4**. If **15** people are known to have contracted this disease, what is the probability that

(a) at least 10 survive

(b) from 3 to 8 survive

(c) exactly 5 survive

Solution

(a) P(X ≥10) = 1 - P(X < 10) = 1- $\sum_{x=0}^{x=9} b(x; 15,0.4)$ =1- .9662=0.033

(b) P(3 ≤ X ≤ 8) = $\sum_{x=3}^{x=8} b(x; 15,0.4)$

= $\sum_{x=0}^{x=8} b(x; 15,0.4)$ - $\sum_{x=0}^{x=2} b(x; 15,0.4)$

(By using Table A.1of the Appendix for n = 1,2,..., 20, and selected values of p; from 0.1 to 0.9)

= .9050 - 0.0271 = 0.8779

Binomial Distribution [2]

(c) $P(X = 5) = \sum_{x=0}^{x=5} b(x; 15, 0.4) - \sum_{x=0}^{x=4} b(x; 15, 0.4)$

$= 0.4032 - 0.2173 = 0.1859$

(By using Table A.1of the Appendix for n = 1,2,..., 20, and selected values of p; from 0.1 to 0.9)

Conjectured: Opinion or judgment based on inconclusive or incomplete evidence; guesswork

Or

An opinion or conclusion based on guesswork.

The commentators made various conjectures about the outcome of the next election.

Binomial Distribution [3]

It is conjectured that an impurity exists in **30%** of all **drinking wells** in a certain **rural community**. In order to gain some insight into the true extent of the problem, it is determined that **some testing is necessary.** It is too expensive to test all of the wells in the area, so **10** are randomly selected for testing.

(a) Using the binomial distribution, what is the probability that exactly **3 wells** have the impurity, assuming that the **conjecture is correct**?

(b) What is the probability that **more than 3 wells** are impure?

Solution :

(a) $b(3; 10, 0.3) = \sum_{x=0}^{x=3} b(x; 10, 0.3) - \sum_{x=0}^{x=2} b(x; 10, 0.3)$

$= 0.6496 - 0.3828 =$ **0.2668**

(b) $P(X > 3) = 1 - P(X \leq 3)$

$= 1 - \sum_{x=0}^{x=3} b(x; 10, 0.3)$

$= 1 - 0.6496 =$ **0.3504**

Binomial Distribution [3]

Example : Consider the situation of the previous example. The **"30% are impure"** is merely a conjecture put forth by the area water board. Suppose **10 wells** are randomly selected and **6 are** found to contain the impurity. What does this imply about the conjecture? Use a probability statement.

Solution: We must first ask: "If the conjecture is correct, is it likely that we could have found 6 or more impure wells?"

P(X ≥6) = 1 - P(X < 6) = $\sum_{x=0}^{x=10} b(\text{x}; 10,0.3)$ - $\sum_{x=0}^{x=5} b(\text{x}; 10,0.3)$

= 1 - 0.9 5 2 7 = 0.0473 or **4.7%**

- As a result, it is very unlikely **(4.7% chance)** that **6 or more wells** would be found impure if only **30%** of all are impure.
- This casts considerable doubt on the conjecture and suggests that the **impurity problem** is much **more severe**.

or

$$P(X = 6) = \sum_{x=0}^{x=6} b(x; 10, 0.3) - \sum_{x=0}^{x=5} b(x; 10, 0.3)$$

$$= 0.9894 - 0.9524 = 0.0370$$

Binomial Distribution [4]

Example : A family has 6 children. Find the probability P that there are

(i) 3 boys and 3 girls.

(ii) fewer boys than girls.

Assume that the probability of any particular child being a boy is 1/2.

Solution:

Here n = 6 and p = q = 1/2.

Let X denotes number of boys

(i) $P(X=3) = {}_6C_3(1/2)^3(1/2)^3 = 20/64 =$ **5/16**

(ii) **There are fewer boys than girls if there are 0, 1 or 2 boys.**

$P(X \leq 2) = {}_6C_0(1/2)^0(1/2)^6 + {}_6C_1(1/2)^1(1/2)^5 + {}_6C_2(1/2)^2(1/2)^4$

$=$ **11/32**

Relationship to the Binomial Distribution [1]

- There is an interesting relationship between the: **hypergeometric** and the **binomial distribution**. As one might expect, if **n is small compared to N**, the nature of the **N** items changes **very little** in each draw.

- So a **binomial distribution** can be used to approximate the **hypergeometric distribution** when **n is small, compared** to **N**.

- In fact, as a **rule of thumb** the **approximation** is good when $\frac{n}{N} \leq 0.05$ **or 5 %.**

Relationship to the Binomial Distribution [2]

- As a result, the **binomial distribution** may be viewed as a **large population** edition of the **hypergeometric distributions**

The mean and variance then come from the formulas

Mean = **np** = $\frac{nk}{N}$

Variance = **npq** = $\frac{nk}{N} * (\frac{N-k}{N})$

$\frac{\mathbf{N-n}}{\mathbf{N-1}}$ is **negligible** when **n is small** relative to **N**

Example Jury Selection In the case of *Castaneda v. Partida* it was noted that although **80%** of the population in a **Texas county** is **Mexican-American**, only **39%** of those summoned for grand juries were **Mexican-American**. Let’s assume that we need to select **12 jurors** from a population that is **80% Mexican-American**, and we want to find the probability that among **12 randomly** selected jurors, exactly **7** are **Mexican-Americans**.

a. Does this procedure result in a **binomial distribution**?

b. If this procedure does result in a binomial distribution, identify the values of ***n, x, p, and q***.

Solution a.

1. The number of **trials (12)** is **fixed**.

2. The **12 trials** are **independent**. (Technically, the 12 trials involve selection without replacement and are not independent, but we can assume independence because we are randomly selecting only 12 members from a very large population.)

3. Each of the **12 trials** has **two categories of outcomes**: The juror selected is either **Mexican-American** or is **not**.

4. For each **juror selected**, the probability that he or she is **Mexican-American** is **0.8** (because 80% of this population is Mexican-American). That probability of **0.8 remains** the same for each of the ***12 jurors***.

b. Having concluded that the given procedure does result in a binomial distribution, we now proceed to identify the values of ***n*, *x*, *p*, and *q***.

1. With **12 jurors** selected, we have $n = 12$.

2. We want the probability of exactly **7 Mexican-Americans**, so $x = 7$.

3. The probability of success (getting a Mexican-American) for one selection is 0.8, so $p = 0.8$.

4. The probability of failure (not getting a Mexican-American) is **0.2**, so $q = 0.2$.