Binomial Probability Distribution

Animal Health Risk Analysis
Canadian Food Inspection Agency
3851 Fallowfield Road
Ottawa, Ontario, Canada K2H 8P9

Analyse des risques de la santé des animaux
Agence canadienne d'inspection des aliments
3851, chemin Fallowfield
Ottawa (Ontario) Canada K2H 8P9

Canadian Food Inspection Agency (CFIA)

Agence canadienne d'inspection des aliments (ACIA)

Properties of the Binomial Distribution

One of the most elementary discrete random variables is the coin-tossing experiment. Each toss is called a trial. In any single trial their will be a probality associated with a particular event such as a head on the coin. The probability will not change from one trial to the next. Such trials are said to be independent and are often called Bernoulli trials.

<u>Bernoulli trials</u>
A Bernoulli trial has only two outcomes: success or failure, denoted as 1 or 0. If the probability of success is p the Bernoulli distribution is $f(x) = p^x q^{1-x}$.

<u>Example</u>: Coin toss with a balanced coin, where 0 denotes tail and 1 denotes head. When x = 0, $f(x) = 0.5^0\ 0.5^{1-0} = 1 \times ½ = ½$ and when x = 1, $f(x) = 0.5^1\ 0.5^{1-1} = ½ \times 1 = ½$.

<u>Binomial Distribution</u>
A binomial probability distribution determines the probabilities for the number of wins (successes) in n identical trials where these conditions are met:

n identical trials

each trial consists of two outcomes, either a success, S, or a failure, F

the probability of success on a single trial is equal to p and remains the same from trial to trial, the probability of failure is equal to (1-*p*) = *q*

Properties of the Binomial Distribution

Numerous coin tossing experiments are conducted almost daily in regulatory and academic veterinary medicine. A survey of swine producers to determine if they feed household food waste to their pigs is one such experiment. Each swine producer selected is analagous to the toss of an unbalanced coin, since the probability of a 'yes' is not one-half but very likely a small probability. Other experiments include: studies to determine the effectiveness of a La Sota Newcastle disease virus vaccine in chickens to protect against experimental challenge with velogenic virus; a survey of cattle ranchers to estimate the compliance with FMD vaccination of all cattle over 4 months of age; and a study to estimate the proportion of longissimus dorsi muscles which have a pH value of less than 6.2 at 24 hours of maturation. These experiments approximate for all practical purposes binomial experiments. Each trial will result in one of two outcomes (protection or not; compliance or not; and pH above or below 6.2). The probability for success will remain the same for trial to trial and will be unaffected by the outcome on any of the other trials. In the survey of the cattle ranchers, the probability for success will remain approximately constant as long as the population of cattle ranchers is large in comparison with the survey sample *n*.

n = a fixed number of identical Bernoulli trials
p = the probability of success in each trial
X = the number of successes in n trials

Properties of the Binomial Distribution

The random variable X is called a binomial random variable. Its distribution is called a binomial distribution. The probability of *x* successes in *n* trials is:

$$P(X = x) = C_x^n p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x} \quad \text{Where} \quad C_x^n = \frac{n!}{x!(n-x)!}$$

For values of x = 1, 2, ..., n and where *n*! =*n*(*n*-1)(*n*-2) ... (2)(1) and 0! = 1.

<u>Rule of Thumb</u>

When the sample (the *n* identical trials) comes from a large population, the probability of success *p* is the same from trial to trial. When the population size N is small, the probability of success *p* can change dramatically from trial to trial and the experiment is not binomial.

If the sample size is large relative to the population size, that is, $n/N \geq 0.05$, then the resulting experiment is not binomial.

Properties of the Binomial Distribution

Probability density function:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Cumulative distribution function:

$$F(x) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i}$$

Parameter: n>0, n is an integer, $0 \le p \le 1$

Domain: {0, 1, 2, 3, ..., n}

Mean (μ): np

Variance (σ^2): npq

Mode: p (n+1)-1 and p(n+1) if p(n+1) is an integer, $\lfloor p(n+1) \rfloor$ otherwise

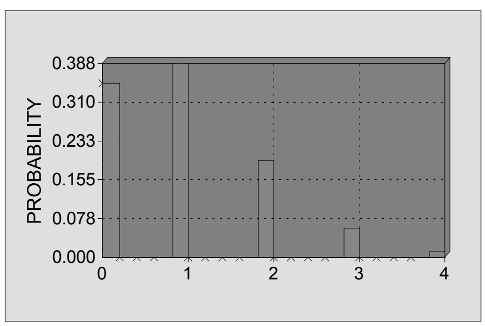
Coefficient of skewness (α_3): (q-p)/$\sqrt{npq}$

Coefficient of kurtosis (α_4): 3 - 6/n + 1/npq

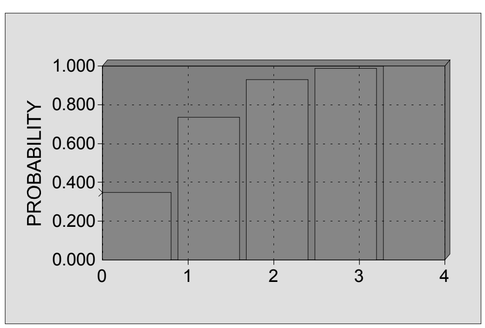
$\lfloor\ \rfloor$ = Greatest integer function

Properties of the Binomial Distribution

Probability mass function for BIN(n = 10, p = 0.1)



Cumulative distribution function for BIN(n = 10, p = 0.1)



Binomial Cumulative Probabilities Table Examples

What is the probability that X≤2 for a binomial random variable with n = 5 and p = 0.2?

This can be written P(BIN(n =5, p = 0.2) ≤ 2) = F(2)

$$= \sum_{x=0}^{2} \binom{5}{x} 0.2^{x} \times 0.8^{5-x} = 0.9421$$

For P(2≤BIN(n =5, p = 0.2) ≤ 4) = F(4) - F(2) =P(X≤4) - P(X≤2) = 0.9997- 0.9421 = 0.0576 from the binomial cumulative probabilities tables on the next slide. The cumulative probabilities are obtained from the following expression:

$$P(X \le x) = \sum_{i=0}^{x} \binom{n}{i} p^{i} (1-p)^{n-i}$$

Binomial Cumulative Probabilities Table Examples

Table of Binomial Distribution Cumulative Probabilities

						p					
n	***x***	**0.05**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.35**	**0.4**	**0.45**	**0.5**
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.9928	0.9720	0.9393	0.8960	0.8438	0.7840	0.7183	0.6480	0.5748	0.5000
	2	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
	2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000
	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125
	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9688
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Limousine Survival Example

Ten years ago 5 purebred Limousine heifer calves were imported for 5 different owners from a country subsequently reporting cases of bovine spongiform encephalopathy. An effort to find these imported cattle was about to be initiated after first answering the question of the probability that all 5 would be alive and the probability that at least one would be alive. Based on national statistics of survival of purebred beef cattle at 10 years of age, the probability of survival to 10 years of age was 0.25.

These questions involve the Binomial distribution because 1) the experiment consists of 5 identical trials, 2) each trial results in one of two outcomes, 3) the probability of success on single trial is equal to p = 0.25, and 4) the trials are independent. The probability that all 5 cattle would be alive represents the case in which 5 successes out of 5 trials is entertained.

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} = \frac{5!}{5!(0)!} 0.25^5 \times 0.75^0 = 0.001$$

The probability that at least one cow would be living is estimated as follows:

$$P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$$

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \frac{5!}{0!(5-0)!} \times 0.25^0 \times 0.75^5 = 1 - 0.24 = 0.76$$

Limousine Survival Example

Another approach is to generate probability density function and cumulative distribution function tables for the binomial distribution where n = 5 and p = 0.25. Such tables can be created in Microsoft Excel using the following BINOMDIST function for this example, employing a 0 for the last parameter to represent the probability density function and 1 for the cumulative distribution function.

Probability density function giving the individual binomial probabilities for BIN(5,0.25).

x	P(X=x)
0	0.237305
1	0.395508
2	0.263672
3	0.087891
4	0.014648
5	0.000977

=BINOMDIST(0,5,0.25,0)

Cumulative distribution function giving the cumulative binomial probabilities for BIN(5,0.25).

x	P(X<=x)
0	0.237305
1	0.632813
2	0.896484
3	0.984375
4	0.999023
5	1

=BINOMDIST(0,5,0.25,1)

Both questions P(X=5) and P(X≥1) = 1 - P(X=0) can be answered from the above tables.

Relationship Between Binomial and Normal Probability Distributions

Probabilities associated with binomial experiments are readily obtainable for the formula BIN(n,p) of the binomial distribution or from binomial tables when n is small. The normal distribution is often a good approximation to a discrete distribution when the latter takes on a symmetric bell shape. The binomial distribution is nicely approximated by the normal in practical problems when one works with the cumulative distribution function.

If X is a binomial random variable as the number of successes in *n* independent trials with *p* as the probability of success and 1-*p* as the probability of failure on any single trial and with mean (μ) =*np* and variance σ^2 = *npq*, then the limiting form of the distribution of:

$$Z = \frac{X - np}{\sqrt{npq}}$$

as n $\rightarrow\infty$, is the standard normal distribution N(0, 1) or NORMALZ(0,1).

It turns out that the normal distribution with μ = np and σ^2 = *np*(1-*p*) not only provides a very accurate approximation to the binomial distribution when *n* is large and *p* is not extremely close to 0 or 1, but also provides a fairly good approximation when *n* is small and *p* is reasonably close to ½.

Relationship Between Binomial and Normal Probability Distributions

Note: Since we are approximating a discrete random variable with a continuous one, we need to apply a continuity correction. That is, if a binomial random variable X takes on the value 'a' exactly, then the corresponding normal X must take on the interval of values (a - 0.5, a + 0.5) since the probability of an exact value for a continuous randomvariable is always 0.

Example - The rate of operative complications in a certain complex vascular reconstructive procedure is 30%. This includes all complications, ranging in severity from wound separation or infection to death, and is the proportion of patients showing some complication. Assuming independence of the occurrence of complication in different patients, in a series of 100 such procedures, what is the probability that there will be:

(a) at most 18 patients with operative complication?
(b) exactly 16 patients with operative complication?

Let X = # of complications occurring.

(a) $$P(X \leq 18) = \sum_{x=0}^{18} \binom{100}{x} (.3)^x (.7)^{100-x}$$

Relationship Between Binomial and Normal Probability Distributions

We need to use the normal approximation since the probability cannot be found in the binomial tables. If X' is now the corresponding r.v. we want P(X′≤ 18.5).

$$\because P(X' \le 18.5) = P(Z \le \frac{18.5 - n\pi}{\sqrt{n\pi(1-\pi}}$$

$$= P(Z \le \frac{18.5 - 30}{\sqrt{21}}) \cong P(Z \le -2.51) = 0.006$$

(b)

$$P(X = 16) = P(15.5 \le X' \le 16.5) = P(\frac{15.5 - 30}{\sqrt{21}} \le Z \le \frac{16.5 - 30}{\sqrt{21}})$$

$$\cong P(-3.16 \le Z \le -2.95) = P(Z \le -2.95) - P(Z \le -3.16)$$
$$= 0.0016 - 0.0008 = 0.0008$$

Relationship Between Binomial and Normal Probability Distributions

<u>Example</u> - Find the probability that a student can guess from 110 to 120 correct answers out of 200 questions on a true-false examination. (No subtraction of wrong from right is done.)

Let X = no. of correct answers out of 200 questions on an exam.

∴ X is binomial with n = 200 and π = ½.
∴ P(110≤X≤120) =P(109.5≤X′≤120.5) if X is N(nπ, nπ (1-π))

$$= P(\frac{109.5 - 100}{\sqrt{50}} \le Z \le \frac{120.5 - 100}{\sqrt{50}}) \cong P(1.34 \le Z \le 2.90)$$

$$= P(Z \le 2.90) - P(Z \le 1.34) = .9981 - .9099 = .0882$$

∴ a student will guess 110 to 120 correct answers out of the 200 T-F exam questions with an 8.8% chance.

<u>Note</u>: If in a binomial distribution n is very large while π is very small, let m = nπ and use the Poisson probability $\frac{e^{-m}m^x}{x!}$ **to approximate the binomial probability** ${}_n c_x \pi^x (1-\pi)^{n-x}$ **for any x value.**