A population is called multinomial if its data is categorical and belongs to a collection of discrete non-overlapping classes. The null hypothesis for goodness of fit test for multinomial distribution is that the observed frequency fi is equal to an expected count ei in each category. It is to be rejected if the p-value of the following Chi-squared test statistics is less than a given significance level α. In the built-in data set survey, the Smoke column records the survey response about the student’s smoking habit. As there are exactly four proper response in the survey: "Heavy", "Regul" (regularly), "Occas" (occasionally) and "Never", the Smoke data is multinomial. It can be confirmed with the levels function in R. > library(MASS)   # load the MASS package > levels(survey$Smoke) [1] "Heavy" "Never" "Occas" "Regul" As discussed in the tutorial Frequency Distribution of Qualitative Data, we can find the frequency distribution with the table function. > smoke.freq = table(survey$Smoke) > smoke.freq Heavy Never Occas Regul   11   189   19   17 Suppose the campus smoking statistics is as below. Determine whether the sample data in survey supports it at .05 significance level.   Heavy   Never   Occas   Regul   4.5%   79.5%   8.5%   7.5% We save the campus smoking statistics in a variable named smoke.prob. Then we apply the chisq.test function and perform the Chi-Squared test. > smoke.prob = c(.045, .795, .085, .075) > chisq.test(smoke.freq, p=smoke.prob)   Chi-squared test for given probabilities data: smoke.freq X-squared = 0.1074, df = 3, p-value = 0.991 As the p-value 0.991 is greater than the .05 significance level, we do not reject the null hypothesis that the sample data in survey supports the campus-wide smoking statistics. Conduct the Chi-squared goodness of fit test for the smoking data by computing the p-value with the textbook formula. All BlogsStatistics ResourcesMultinomial DistributionProbabilityRandom ProcessesMarkov ChainMarkov ModelPoisson ProcessProbabilityGET: WSM ALL COURSES ACCESSTable Of ContentsMultinomial distribution is a multivariate version of the binomial distribution. It is the probability distribution of the outcomes from a multinomial experiment. It is used in the case of an experiment that has a possibility of resulting in more than two possible outcomes.You are free to use this image on your website, templates, etc.. Please provide us with an attribution link.The simplest technique to construct a multinomial random variable is to replicate an experiment (by drawing n uniform random numbers and assigning them to certain bins based on the cumulative value of the p vector) to produce a multinomial random variable. In instructional statistics, this distribution is put to various uses.The multinomial distribution represents the likelihood of receiving a certain set of counts where each trial has a discrete number of possible outcomes.The most direct goodness-of-fit test is based on the multinomial distribution of response patterns.Its applications and use cases frequently involve the evaluation of the likelihood of a set of outcomes that are usually more than two or at least two.It is a type of probability distribution of which binomial distribution is a subtype.The multinomial distribution is a multivariate discrete distribution. Multinomial experiments include the following characteristics: The experiment comprises of n repeated trials. Its results have statistical significance. There are a finite number of possible outcomes for each trial, and the likelihood of any event occurring is constant throughout the experiment. The results of one experiment do not influence the results of the others. Assuming the model is valid, the most straightforward way to determine a model's fit is to use the multinomial distribution of response patterns. Accordingly, there are 'nm' potential response patterns ranging from (l,...,l) to (m,...,m) for n items with m response categories for each item. Hence, all response patterns have the same probability of occurring if u is a typical pattern.Thus, to estimate the frequencies of the response patterns, use multinomial distribution with parameters n and actual probability for all the response patterns. According to the multivariate central limit theorem, the multivariate normal distribution can approximate the distribution for large sample sizes.While considering the entire data, the distribution of the observations has a multinomial shape for observations from different Poisson distributions. Probabilities in the multinomial distribution are based on the Poisson mean for each cell multiplied by all Poisson mean values.A multinomial experiment has a subtype known as a binomial one. In this regard, there is one major distinction. Accordingly, in a binomial experiment, there are only two possibilities for each trial. At the same time, each experiment in a multinomial trial has the potential difference for two or more different results.For instance, you perform n times an experiment with K outcomes. Then, you denote by Xi the number of times you obtain the i-th outcome. In that case, the random vector X is defined as X = is a multinomial random vector. The multinomial distribution is the generalization of the binomial distribution to the case of n repeated trials where there are more than two possible outcomes for each. If an event may occur with k possible outcomes, each with a probability, pi(i = 1,1,…,k), with ∑k(i=1)pi = 1, and if ri is the number of the outcome associated with 'pi' occurs, then the random variables ri (i = 1,2,…,k-1) have a multinomial probability defined asf (r1,r2,….,rk-1) = n! ∏ki=1 piri/∏ki=1 ri!, ri = 0,1,2,….,n.Note that each of the ri ranges from 0 to n inclusive with only (k-1) variables because of the linear constraint:∑ki=1 ri = nJust as the binomial distribution tends to the univariate normal, so does the multinomial distribution tend to limit to the multivariate normal distribution.Suppose a random variable X has a multinomial distribution. In that case, the following multinomial distribution calculator calculates the likelihood that event 1 occurs exactly x1 times, event 2 occurs exactly x2 times, event 3 occurs exactly x3 times, and so on. Hence following is the multinomial distribution formula:Probability = n!*(p1x1 * p2x2 * … * pkxk)/(x1!*x2!*…*xk!)Where: n: the total number of events x1, x2, xk: the number of occurrences of event 1, event 2, and event k, respectively. p1, p2, pk: the likelihood that results in 1, 2, and k happen, respectively, in a trail. Let us have a look at the multinomial distribution example to understand the concept better:Rebecca, a portfolio manager, utilizes it to assess the probability of her client's investment. For 60% of the time, she chooses a small-cap index to outperform a large-cap index. Whereas for 40% of the time, Rebecca opts for a large-cap index to outperform a small-cap index. For 10% of the time, the indexes may have the same or approximate return.Since the trial may last a full year of trading days in such cases, Rebecca uses actual market data to validate the outcomes. In scenarios where the likelihood of this set of occurrences is high, her clients may tempt to overinvest in the small-cap index.Likewise, Neil, a financial analyst, uses this method to evaluate the likelihood of events, like potential quarterly sales for a business when its competitors post lower-than-expected profits. All BlogsStatistics ResourcesMultinomial DistributionProbabilityRandom ProcessesMarkov ChainMarkov ModelPoisson ProcessProbabilityGET: WSM ALL COURSES ACCESSTable Of ContentsMultinomial distribution is a multivariate version of the binomial distribution. It is the probability distribution of the outcomes from a multinomial experiment. It is used in the case of an experiment that has a possibility of resulting in more than two possible outcomes.You are free to use this image on your website, templates, etc.. Please provide us with an attribution link.The simplest technique to construct a multinomial random variable is to replicate an experiment (by drawing n uniform random numbers and assigning them to certain bins based on the cumulative value of the p vector) to produce a multinomial random variable. In instructional statistics, this distribution is put to various uses.The multinomial distribution represents the likelihood of receiving a certain set of counts where each trial has a discrete number of possible outcomes.The most direct goodness-of-fit test is based on the multinomial distribution of response patterns.Its applications and use cases frequently involve the evaluation of the likelihood of a set of outcomes that are usually more than two or at least two.It is a type of probability distribution of which binomial distribution is a subtype.The multinomial distribution is a multivariate discrete distribution. Multinomial experiments include the following characteristics: The experiment comprises of n repeated trials. Its results have statistical significance. There are a finite number of possible outcomes for each trial, and the likelihood of any event occurring is constant throughout the experiment. The results of one experiment do not influence the results of the others. Assuming the model is valid, the most straightforward way to determine a model's fit is to use the multinomial distribution of response patterns. Accordingly, there are 'nm' potential response patterns ranging from (l,...,l) to (m,...,m) for n items with m response categories for each item. Hence, all response patterns have the same probability of occurring if u is a typical pattern.Thus, to estimate the frequencies of the response patterns, use multinomial distribution with parameters n and actual probability for all the response patterns. According to the multivariate central limit theorem, the multivariate normal distribution can approximate the distribution for large sample sizes.While considering the entire data, the distribution of the observations has a multinomial shape for observations from different Poisson distributions. Probabilities in the multinomial distribution are based on the Poisson mean for each cell multiplied by all Poisson mean values.A multinomial experiment has a subtype known as a binomial one. In this regard, there is one major distinction. Accordingly, in a binomial experiment, there are only two possibilities for each trial. At the same time, each experiment in a multinomial trial has the potential difference for two or more different results.For instance, you perform n times an experiment with K outcomes. Then, you denote by Xi the number of times you obtain the i-th outcome. In that case, the random vector X is defined as X = is a multinomial random vector. The multinomial distribution is the generalization of the binomial distribution to the case of n repeated trials where there are more than two possible outcomes for each. If an event may occur with k possible outcomes, each with a probability, pi(i = 1,1,…,k), with ∑k(i=1)pi = 1, and if ri is the number of the outcome associated with 'pi' occurs, then the random variables ri (i = 1,2,…,k-1) have a multinomial probability defined asf (r1,r2,….,rk-1) = n! ∏ki=1 piri/∏ki=1 ri!, ri = 0,1,2,….,n.Note that each of the ri ranges from 0 to n inclusive with only (k-1) variables because of the linear constraint:∑ki=1 ri = nJust as the binomial distribution tends to the univariate normal, so does the multinomial distribution tend to limit to the multivariate normal distribution.Suppose a random variable X has a multinomial distribution. In that case, the following multinomial distribution calculator calculates the likelihood that event 1 occurs exactly x1 times, event 2 occurs exactly x2 times, event 3 occurs exactly x3 times, and so on. Hence following is the multinomial distribution formula:Probability = n!*(p1x1 * p2x2 * … * pkxk)/(x1!*x2!*…*xk!)Where: n: the total number of events x1, x2, xk: the number of occurrences of event 1, event 2, and event k, respectively. p1, p2, pk: the likelihood that results in 1, 2, and k happen, respectively, in a trail. Let us have a look at the multinomial distribution example to understand the concept better:Rebecca, a portfolio manager, utilizes it to assess the probability of her client's investment. For 60% of the time, she chooses a small-cap index to outperform a large-cap index. Whereas for 40% of the time, Rebecca opts for a large-cap index to outperform a small-cap index. For 10% of the time, the indexes may have the same or approximate return.Since the trial may last a full year of trading days in such cases, Rebecca uses actual market data to validate the outcomes. In scenarios where the likelihood of this set of occurrences is high, her clients may tempt to overinvest in the small-cap index.Likewise, Neil, a financial analyst, uses this method to evaluate the likelihood of events, like potential quarterly sales for a business when its competitors post lower-than-expected profits. The multinomial distribution is a generalization of the binomial distribution to k categories instead of just binary (success/fail). For n independent trials each of which leads to a success for exactly one of k categories, the multinomial distribution gives the probability of any particular combination of numbers of successes for the various categories Example: Rolling a die N times In Data mining, When we discuss everything in terms of text classification, i.e. Topic Modeling: Each document has its own distribution over topics. Each topic has its own distribution over the words. The (multinomial) distribution over words for a particular topic The (multinomial) distribution over topics for a particular document Two chess players have the probability Player A would win is 0.40, Player B would win is 0.35, game would end in a draw is 0.25. The multinomial distribution can be used to answer questions such as: “If these two chess players played 12 games, what is the probability that Player A would win 7 games, Player B would win 2 games, the remaining 3 games would be drawn?” dmultinom(x=c(7,2,3), prob = c(0.4,0.35,0.25)) ## [1] 0.02483712 In a little town, 40% of the eligible voters prefer candidate A, 10% prefer candidate B, 50% have no preference. You randomly sample 10 eligible voters. What is the probability that 4 will prefer candidate A, 1 will prefer candidate B, 5 will have no preference ? dmultinom(x=c(4,1,5), prob = c(0.4,0.1,0.5)) ## [1] 0.1008 A distribution that shows the likelihood of the possible results of a experiment with repeated trials in which each trial can result in a specified number of outcomes that is greater than two. A multinomial distribution could show the results of tossing a dice, because a dice can land on one of six possible values. By contrast, the results of a coin toss would be shown using a binomial distribution because there are only two possible results of each toss, heads or tails. Two additional key characteristics of a multinomial distribution are that the trials it illustrates must be independent (e.g., in the dice experiment, rolling a five does not have any impact on the number that will be rolled next) and the probability of each possible result must be constant (e.g., on each roll, there is a one in six chance of any number on the die coming up). one.dice