Unifying Statistical Distributions:

Exploring Connections between Discrete and Continuous Models

Spencer Sewell

Advisor: Dr. Ping Ye

University of North Georgia

January 23rd, 2023

Abstract

In statistics there appear to be four major sources for probability distributions, although the list is not limited to this: "the uniform distributions, either discrete Uniform(n), or continuous Uniform(a, b), those that come from the Bernoulli process or sampling with or without replacement: Bernoulli(p), Binomial(n, p), Geometric(p), NegativeBinomial(p, r), and Hypergeometric(N, M, n). those that come from the Poisson process: Poisson(λ, t), Exponential(λ), Gamma(λ, r), and Beta(α, β), [and] those related to the Central Limit Theorem: Normal(μ, σ2), ChiSquared(ν), T(ν), and F(ν1, ν2)." (Joyce, 2006). If you explore the distributions further, connections between these various distributions can be made, both defined on a discrete and continuous basis. First, we will define what a probability distribution is, and then explore the various types of distributions in both a discrete and continuous case. Then, through a thorough examination of both types of models, we will uncover the similarities and differences between them.

1 Introduction

Probability distributions are a fundamental concept in statistics and probability theory, representing the functions that describe the likelihood of each possible outcome of a random event. Any time we look at data sets composed of variables with more than one viable answer - rolling dice, flipping coins, measuring atmospheric pressure fluctuations - we are dealing with some form of probability distribution. These probabilities are assigned to a random variable, which are typically expressed as numerical quantities from the results of a random experiment or observation. A defining feature is that these probabilities will sum to 1.

Let's introduce two main probability functions: the probability mass function (PMF) for discrete probability distributions, and the probability density function (PDF) for continuous probability distributions. These functions are used to find various properties of a probability distribution, such as its mean, variance, and standard deviation. They serve well to describe the central tendency and spread of a data set. As a secondary source of information, we can use the moment generating function. The moment generating function (MGF) generates the "moments" of the distribution, which are the expected values of the exponential function of a random variable. By taking the derivatives, we can find the mean, variance, skewness, and kurtosis. Finally, we can introduce the two main types of probability distributions: discrete and continuous. When a random variable can only take on a finite or countably infinite number of values, we define it as a discrete distribution. Examples of discrete random variables include the number of heads obtained in a series of coin flips (either a yes/no outcome), the number of defective items in a batch of products, and the number of customers who arrive at a store in a given hour. Since the possible values are countable, the PMF can be represented as a probability distribution table or graph, and the probabilities assigned must sum to 1. If a random variable can take on any value within a defined range or interval, we define it as a continuous distribution. Examples of continuous random variables include the weight or height of individuals in a population, the time taken to complete a task, and the amount of rainfall in a given area. Since the possible values are uncountable, the PDF cannot necessarily be represented as a probability distribution table or graph. Instead, we use the area under the curve generated by the distribution. Like the discrete data, the area under the curve will be equal to 1.

At the fundamental level of highschool and undergraduate statistics, the connections between these various distributions are rarely covered. Exploring these connections and highlighting the significance of how they work and interact with each other might help students achieve a stronger foundation of statistical distributions. To achieve this, several techniques can be used; One of the techniques for connecting distributions is to identify special cases where one distribution can be derived from another distribution by setting some of the parameters equal to specific values. Another technique for connecting distributions is to use transformations, which allow a mapping of one probability distribution to another. Some common types of transformations include linear transformations, cumulative distribution function transformations, and moment generating transformations. Additionally, we can employ the use of limits to show that the random variables of a moment generating function converge to a specific distribution as the number of observations increases.

2 Connections

2.1 Discrete Connections

For a great introductory set of distributions which easily show both a special case and a moment generating transformation through their connections, let's take the Bernoulli Distribution and the Binomial Distribution.

The Bernoulli Distribution is written as:

$X \sim$ Bernoulli(p) = $f(x) = p^x(1-p)^{(1-x)}$

With a mean and variance of $\mu = p$, $\sigma^2 = p(1-p)$. As we will be using a moment generating transformation, let's also introduce the moment generating function of the Bernoulli Distribution as $M(t) = E[e^{tX}] = (1-p) + pe^t$.

The Bernoulli distribution describes the outcome of a single binary (yes/no) experiment, where the outcome can be either success or failure. The most common example is the flip of a coin.

The Binomial Distribution is written as:

$X \sim$ binomial(n,p) = $f(x) = \binom{n}{x} p^x(1-p)^{(n-x)}$

With a mean and variance of $\mu = np$, $\sigma^2 = np(1-p)$. Again, let's show the moment generating function of the Binomial Distribution as $M(t) = E[e^{tX}] = (1-p+pe^t)^n$.

The Binomial Distribution is used to model the number of successes in a fixed number of independent trials, where each trial can either result in success or failure. It is particularly useful when the probability of success in each trial is the same, and the trials are independent of each other.

There is an obvious similarity in the two distribution functions. The Bernoulli Distribution is a special case of the Binomial distribution where n=1. If we have a Binomial Distribution where $n = 1$, we have $\binom{1}{x}p^x(1-p)^{((1)-x)}$, which simplifies to $p^x(1-p)^{1-x}$, the Bernoulli Distribution. It is a very simple and fundamental example of the special case. Now, let us look at the other way around, moving from a Bernoulli Distribution to a Binomial Distribution. It is here where we can take the moment generating functions and use a moment generating transformation, in this case multiplying n Bernoulli moment generating functions together, to reach the Binomial moment generating function. Let $X_1, X_2, \dots, X_n$ be mutually independent Bernoulli(p) random variables, then $Y = \Sigma_{i=1}^{n} X_i$ is binomial(n,p).

Proof: Through the uniqueness property of moment generating functions, we can identify which probability function a random variable follows. The moment generating function of X_i is

$M(t) = E[e^{tX}] = (1-p) + pe^t$ for $i = 1, 2, ...n$, $-\infty < t < \infty$

We can see that if we multiply each individual moment generating function together, we get

$M_Y(t) = \Pi_{i=1}^{n}(1 - p + pe^t) = (1 - p + pe^t)^n$

Which, by the uniqueness property of the moment generating function, is the mgf for a Binomial random variable. □

One can also take the Negative Binomial Distribution and the Geometric Distribution and just as easily unify them using the exact same techniques. It is almost the exact same case, where the geometric distribution is the special case of the negative binomial distribution where n=1, and the moment generating function of the negative binomial distribution is a transformation of the moment generating function of the geometric distribution. Now, let's look at an example of a limiting technique, which is a bit more rigorous than our previous examples. We will introduce the Poisson distribution.

The Poisson Distribution is written as:

$X \sim \text{Poisson}(\lambda) = f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$

The Poisson distribution is a probability distribution that describes the likelihood of a certain number of events occurring within a fixed interval of time or space, given a known average rate of occurrence. It is often used to model rare events, such as the number of car accidents in a day or the number of customers who arrive at a store in an hour.

The connection to be made here is to the binomial distribution. The Poisson distribution is related to the binomial distribution in that it can be used as an approximation of the binomial distribution in situations where the probability of success is small and the number of trials is large. Specifically, if n is large and p is small, such that np is moderate or large, then the binomial distribution can be approximated by the Poisson distribution with a parameter λ = np. In terms of limits, The Poisson Distribution is the limit of the Binomial Distribution with λ = np as n approaches infinity and p approaches 0.

Proof: Let X be a random variable that follows the Binomial Distribution. We will begin by substituting p with $\frac{\lambda}{n}$:

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

$= \binom{n}{x} (\frac{\lambda}{n})^x (1 - \frac{\lambda}{n})^{(n-x)}$

$$=\frac{n!}{x!(n-x)!}(\frac{\lambda}{n})^x(1-\frac{\lambda}{n})^{(n-x)}$$

$$=\frac{\lambda^x}{x!}\frac{n!}{(n-x)!}(\frac{1}{n^x})(1-\frac{\lambda}{n})^{(n-x)}$$

Now, let us look at $\frac{n!}{(n-x)!}(\frac{1}{n^x})$ and expand it further

$$=\frac{n(n-1)(n-2)...(n-x+1)(n-x)!}{(n-x)!n^x}$$

$$=\frac{n(n-1)(n-2)...(n-x+1)}{n^x}$$

$$=\frac{n}{n}(\frac{n-1}{n})(\frac{n-2}{n})...(\frac{(n-x+1)}{n})$$

$$=1(1-\frac{1}{n})(1-\frac{2}{n})...(1-\frac{x+1}{n})$$

Now, using the product rule of limits, we can take the limit of our expanded distribution

First, note that $e^x = \lim_{n\to\infty}(1+\frac{x}{n})^n$

$$f(x) = \frac{\lambda^x}{x!}(1-\frac{1}{n})(1-\frac{2}{n})...(1-\frac{x+1}{n})(1-\frac{\lambda}{n})^{n-x}$$

$$f(x) = \frac{\lambda^x}{x!}\lim_{n\to\infty}(1-\frac{1}{n})(1-\frac{2}{n})...(1-\frac{x+1}{n})\lim_{n\to\infty}(1-\frac{\lambda}{n})^n\lim_{n\to\infty}(1-\frac{\lambda}{n})^{-x}$$

$$=\frac{\lambda^x}{x!}(1)(e^{-\lambda})(1)$$

$$=\frac{\lambda^x e^{-\lambda}}{x!} \quad \square$$

2.2 Continuous Connections

The use of special cases, transformations, and limiting techniques are not just unique to connecting discrete distributions. We can also use it to connect various continuous distributions. As before, we will look at two different examples. First, let’s look at the connection between the Gamma Distribution and the Chi Square distribution.

The Gamma Distribution is written as:

X ~ gamma(α,β) = $f(x) = \frac{x^{\beta-1}e^{-x/\alpha}}{\alpha^{\beta}\Gamma(\beta)}$

With a mean and variance of E[X] = αβ, V[X] = $\alpha^2\beta$. The moment generating function is defined as $M(t) = E[e^{tX}] = (1-\alpha t)^{-\beta}$.

The gamma distribution is a useful tool that models a wait time until a kth event. It can be used to model service times, lifetimes of objects, and repair times.

The Chi-Square Distribution is written as:

X ~ χ^2(n) = $f(x) = \frac{x^{(n/2)-1}e^{-x/2}}{2^{n/2}\Gamma(n/2)}$

The Chi-Square distribution is most commonly used in inferential statistics, and is used to test the goodness of fit of a data distribution.

The connection between the two distributions is as simple as the connection between the Bernoulli and Binomial distribution. It is easy to see that the Chi-Square distribution is a special case of the Gamma distribution when α = 2 and β = n/2. However, we can also make a second connection using a transformation. If X ~ gamma(α, β), then 2X/α ~ χ^2(n), where β = n/2.

Proof:

Let X be a random variable such that it follows a gamma distribution with the probability density function $f_x(x) = \frac{x^{\beta-1}e^{-x/\alpha}}{\alpha^{\beta}\Gamma(\beta)}$

We will define a one to one transformation Y = g(x) = $\frac{2X}{\alpha}$ such that the inverse transform is X = g^{-1}(Y) = $\frac{\alpha Y}{2}$ and $\frac{\partial X}{\partial Y} = \frac{\alpha}{2}$

The probability density function of Y is

$$f_Y(y) = f_X(g^{-1}(y))|\frac{dx}{dy}|$$

$$= f_x(x) = \frac{(\frac{\alpha y}{2})^{\beta-1}e^{-(\frac{\alpha y}{2})/\alpha}}{\alpha^\beta\Gamma(\beta)}|\frac{\alpha}{2}|$$

$$= \frac{y^{\beta-1}e^{-y/2}}{2^\beta\Gamma(\beta)}$$

$$= \frac{y^{(n/2)-1}e^{-y/2}}{2^{n/2}\Gamma(n/2)}$$

Which is the Chi-Square probability density function with n degrees of freedom. □

Now, let’s look at a continuous connection that uses a moment generating transformation and limiting techniques.

The Normal Distribution is written as:

$X \sim N(\mu,\sigma^2) = f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

With a moment generating function $M(t) = E[e^{tX}] = e^{\frac{t(t\sigma^2+2\mu)}{2}}$

The Normal distribution is a versatile distribution; not merely for height and weight variables. It is commonly used as an approximation of other distributions, such as the binomial or the Poisson. For instance, when n is sufficiently large and p=0.5, the Normal distribution approximates the outcome of a binomial distribution. When n is sufficiently large and p is sufficiently small, we can use the Normal distribution as a fit to a Poisson distribution.

We can connect the Gamma distribution to the Normal distribution; we claim that the limiting distribution of the Gamma(α, β) distribution is the N(μ, σ^2) distribution where $\mu = \alpha\beta$ and $\sigma^2 = \alpha^2\beta$.

Proof:

Let X be a random variable such that it follows the gamma(α, β) distribution with pdf:

X ~ gamma(α,β) = $f(x) = \frac{x^{\beta-1}e^{-x/\alpha}}{\alpha^{\beta}\Gamma(\beta)}$ x > 0

The Moment Generating Function of X is

$M(t) = E[e^{tX}] = (1 - \alpha t)^{-\beta}$ t < 1/α

The mean and the variance of X is E[X] = αβ and V [X] = $\alpha^2\beta$. We will do a one to one transformation by the standard normal variate corresponding to X, which is

Z = $\frac{X-E[X]}{\sqrt{V[X]}}$

Y = g(X) = $\frac{X-\alpha\beta}{\alpha\sqrt{\beta}}$ = $\frac{X}{\alpha\sqrt{\beta}} - \sqrt{\beta}$

Now, the Moment Generating Function of Y is

$M_Y(t) = E[e^{tY}]$

= $E[e^{t(\frac{X}{\alpha\sqrt{\beta}} - \sqrt{\beta})}]$

= $e^{-t\sqrt{\beta}}E[e^{t(\frac{X}{\alpha\sqrt{\beta}})}]$

= $e^{-t\sqrt{\beta}}M_X(\frac{t}{\alpha\sqrt{\beta}})$

= $e^{-t\sqrt{\beta}}(1 - (\frac{t}{\sqrt{\beta}}))^{-\beta}$ $t < \sqrt{\beta}$

The limit of the moment generating function Y as $\beta \to \infty$ is

$$\lim_{\beta\to\infty} M_Y(t) = \lim_{\beta\to\infty} e^{-t\sqrt{\beta}}(1 - (\frac{t}{\sqrt{\beta}}))^{-\beta} = e^{\frac{t^2}{2}}$$

Which, by the uniqueness property of the moment generating function, is the mgf of a standard normal random variable. □

3 Future Analysis

3.1 The Additional "Features" of Distributions

While one might use some of the techniques we have shown (special cases, transformations, limiting in terms of the moment generating function), there are several other characteristics of distributions we have not covered. Let's take the uniqueness property of the moment generating function for instance. If you can show that a random variable follows a moment generating function that is the same as some known moment generating function, then it follows the same distribution. However, there are some distributions that have no moment generating function, one of these being the Cauchy Distribution. In these cases, there are other defining features of distributions that might be used to make connections.

The survivor and hazard functions are common features for a lot of distributions. These two characteristics find use in describing the time-to-event variable T with many different distribution models, including but not limited to exponential distribution, Weibull distribution, and log-normal distribution. Known as the complementary cumulative distribution function (CCDF), the survival function provides the probability that T is no larger than a given value. In simple terms, if $S(t) = P(T > t)$, then $S(t)$ gives us the surviving proportion beyond time t. The survival function always starts at 1 and decreases over time. In contrast, the hazard function describes the level of risk T poses for an event occurring at any given instance. As a simple example: if you've already lasted three years into a five-year term with any particular manufacturer's product without experiencing any problems, it suggests that you'll likely make it through a total term safely if you persist further.

Another feature to be utilized is a distribution's characteristic function, which is defined as $\varphi(t) = E[e^{itX}]$, where 'i' denotes the imaginary unit and 't' represents a real number. With the characteristic function being a complex-valued function of 't', it can be viewed as the Fourier transform of the probability density function of X. What makes it a powerful tool in terms of connecting distributions is that, like the moment generating function, there exists a uniqueness property.

Typically these features are not covered in a fundamental statistics or probability class, so further research, classes, etc. on their definitions and interactions might be needed in order to apply the connection techniques.

3.2 Compound/Mixture Distribution

In the real world, we find that statistics is not as simple as just using one distribution. There are an infinite number of scenarios that might require parameters that fall outside of simple univariate distributions. The use of compound (or mixture) distributions might be beneficial.

A compound probability distribution arises when a random variable X follows a specific distribution F, which depends on an uncertain parameter Θ. This unknown parameter Θ, in turn, follows another distribution G. The resulting distribution H is described as the outcome of combining distributions F and G. Simply, a compound distribution is a distribution defined in terms of a second distribution depending on a parameter that is itself a random variable of another distribution. We find methods of this commonly used in academia for Bayesian Statistics and in the workforce for actuarial science. With further study on the properties of distributions, one might be able to apply them to make connections between compound distributions. For example, one might look into a connection on the compound poisson distribution; one of the properties is that the sum of independent compound poisson distributions also follows a compound poisson distribution. Working on a proof of this can be left for future analysis.

4 Conclusion

As we have shown, there are several methods to connect one distribution to another. Using special cases, transformations, and taking the limits of distributions allow for a beautiful flow map between each distribution. As mentioned before, it can be very helpful for students to make these connections so they have a deeper understanding of how distributions work. Sometimes it might be obvious and have a reason, such as realizing the connection between the poisson and binomial distribution. When the sample size is large enough, the poisson distribution can approximate the binomial distribution. Hence, the connection we made through limiting techniques as n approaches infinity. Sometimes, the connection may not be as clear or have a reason, such as the connection between the gamma distribution and the normal distribution. We proved that there was a connection between the two distributions, but what is the significance of the connection? While it might not have given us a clear definition like the poisson and the binomial, one might realize that the connection simply comes from the fact that the gamma and the normal distribution come from the same exponential distribution family.

On an additional note, when taking into account the almost infinite possibilities of compound distributions, whether it be uni-, bi-, or multivariate, this might imply that there are undiscovered distributions. As described before, while the compounding of any distributions might be possible, they might serve little to no purpose. Especially given the fact that there are so many unique statistical scenarios in the real world. With practice and understanding of the distributions, one might find it easier to discover their own unique distributions to apply for their observation or experiment. This is left for future research.

References:

Crooks, G. E. (2019). *Field guide to continuous probability distributions*. Three Plus One. Retrieved April 21, 2023, from http://threeplusone.com/pubs/fieldguide/

Evans, M., & Rosenthal, J. S. (2010). *Probability and statistics: The Science of Uncertainty*. W.H. Freeman and Company.

Joyce, D. (2006). *Common probability distributions - Clark University*. Department of Mathematics | Clark University. Retrieved January 23, 2023, from https://mathcs.clarku.edu/~djoyce/ma217/distributions2.pdf

SMU. (n.d.) *Nonlife Actuarial Models Chapter 1 | Claim Frequency Distribution*. Southern Methodist University. Retrieved from http://www.mysmu.edu/faculty/yktse/NAM/NAM_S1.pdf

Stanford. (2017). *Survival distributions, hazard functions, cumulative hazards.* Department of Mathematics | Stanford University. Retrieved April 22, 2023, from https://web.stanford.edu/~lutian/coursepdf/unit1.pdf