Distributions of Functions of Random Variables

Sean Mulherin

March 1, 2024

1 Introduction

In this paper, you will learn some of the most notable distributions of functions of random variables. Distributions are determined by the random variable(s) they describe. There are many different types of random variables and so there are many different types of distributions. Functions of random variables enable us to apply transformations such that particular behaviors of unique phenomena can be modeled.

I will begin by looking at the most common methods used to derive functions of random variables. This is done to elucidate the origins of distributions introduced thereafter. Using such methods, I will derive many notable distributions beginning with the simplest case: discrete univariate distributions. Then, I will introduce continuous univariate distributions to help transition to the more versatile case of multivariate models. Finally, I will discuss the case of conditional distributions chained together through interrelated random variables called hierarchical models.

2 Methods of Deriving Functions of Random Variables

1. Method of Cumulative Density Function (cdf)
2. Method of Transformation
3. Method of Jacobian (aka Change-of-Variables)

2.1 Method of cdf

The cumulative density function is a non-decreasing function that models the probability that a random variable X is less than or equal to a specified value x. It can be applied to either discrete or continuous random variables.

Definition 1. Let Y be a continuous random variable with cdf $F_Y(y)$. Assume injective transformations.

Case 1. Let $Y = g(x)$ be a monotonically increasing function. The cdf of Y can be transformed to the cdf of X in the following way

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

The pdf is given by

$$f_Y(y) = F'_Y(y) = \frac{d}{dy}F_x(g^{-1}(y)) = f_x(g^{-1}(y))\frac{d}{dy}g^{-1}(y)$$

Case 2. Let $Y = g(x)$ be a monotonically decreasing function. The cdf of Y can be transformed to the cdf of X in the following way

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$

The pdf is given by

$$f_Y(y) = F'_Y(y) = -\frac{d}{dy}F_x(g^{-1}(y)) = -f_x(g^{-1}(y))\frac{d}{dy}g^{-1}(y)$$

Example. Let $X \sim \Gamma(\alpha, \beta)$. Find the distribution of $Y = aX + b$, $a > 0$.
First, we obtain $g^{-1}(y)$

$$Y = aX + b \Rightarrow g(X) = aX + b \Rightarrow g^{-1}(y) = \frac{y-b}{a}$$

Apply the method of cdf

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(X \leq \frac{y-b}{a}) = F_X(\frac{y-b}{a})$$

Differentiate both sides with respect to y to obtain the pdf

$$f_Y(y) = f_X(\frac{y-b}{a}) \cdot \frac{1}{a} = \frac{(\frac{y-b}{a})^{\alpha-1}e^{-(\frac{y-b}{a\beta})}}{\Gamma(\alpha)\beta^\alpha} \cdot \frac{1}{a} = \frac{(y-b)^{\alpha-1}e^{\frac{b-y}{a\beta}}}{\Gamma(\alpha)(\alpha\beta)^\alpha}$$

2.2 Method of Transformation

The method of transformation follows directly from the method of cdf.

Definition 2. Let X be a continuous random variable with probability density function $f(x)$. Assume injective transformations. Let $Y = g(x)$ be a monotonic function; thus,

$$\text{If } u > v \Rightarrow g(u) > g(v)$$

$$\text{If } u < v \Rightarrow g(u) < g(v)$$

The pdf of Y is given by

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

Theorem 1. Let X be a continuous random variable and let $Y = g(x)$.

Case 1. If $g(x)$ is a monotonically increasing function, then $F_Y(y) = F_X(g^{-1}(y))$
Case 2. If $g(x)$ is a monotonically decreasing function, then $F_Y(y) = 1 - F_X(g^{-1}(y))$

Example. Let $X \sim \Gamma(\alpha, \beta)$. Find the distribution of $Y = aX + b$, $a > 0$. Recall from the previous example, $g^{-1}(y) = \frac{y-b}{a}$ Apply the method of transformation to obtain the pdf

$$\begin{aligned} f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| &= f_X(\frac{y-b}{a})\left|\frac{d}{dy}\frac{y-b}{a}\right| \\ &= f_X\left(\frac{y-b}{a}\right)\frac{1}{a} \\ &= \frac{(\frac{y-b}{a})^{\alpha-1}e^{-(\frac{y-b}{a\beta})}}{\Gamma(\alpha)\beta^\alpha}\frac{1}{a} \\ &= \frac{(y-b)^{\alpha-1}e^{\frac{b-y}{a\beta}}}{\Gamma(\alpha)(\alpha\beta)^\alpha} \end{aligned}$$

2.3 Method of Jacobian (aka Change-of-Variables)

The method of Jacobian enables us to derive joint distributions of multivariate random vectors.

Definition 3. Let X, Y be continuous random variables with joint pdf $f_{XY}(x, y)$. Suppose $U = g_1(X, Y)$ and $V = g_1(X, Y)$. Assume injective transformations.
We first solve U, V for x, y in terms of u, v to obtain

$$x = h_1(u, v) \text{ and } y = h_2(u, v)$$

Case 1. We compute the Jacobian J_1, defined as the determinant of the matrix of partial derivatives of g_1 and g_2 w.r.t x and y.

$$J_1 = \begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial x} \end{bmatrix} = \frac{\partial g_1}{\partial x}\frac{\partial g_2}{\partial y} - \frac{\partial g_1}{\partial y}\frac{\partial g_2}{\partial x}$$

Thus, the pdf is given by

$$f_{U,V}(u, v) = f_{X,Y}(x = h_1, y = h_2)|J_1|^{-1}$$

where $|J_1|^{-1}$ is the inverse of the absolute value of the Jacobian matrix.

Case 2. Compute the Jacobian J_2, defined as the determinant of the matrix of partial derivatives of h_1 and h_2 w.r.t x and y.

$$J_2 = \begin{bmatrix} \frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y} \\ \frac{\partial h_2}{\partial x} & \frac{\partial h_2}{\partial x} \end{bmatrix} = \frac{\partial h_1}{\partial x}\frac{\partial h_2}{\partial y} - \frac{\partial h_1}{\partial y}\frac{\partial h_2}{\partial x}$$

Thus, the pdf is given by

$$f_{U,V}(u,v) = f_{X,Y}(x = h_1, y = h_2)|J_2|$$

where $|J_1|^{-1}$ is the inverse of the absolute value of the Jacobian matrix.

Note, both **Case 1** and **Case 2** are effective strategies, but one should use mathematical judgement to determine the less strenuous of the two.

Example. Let X_1 and X_2 be independent exponential random variables with parameters λ_1 and λ_2, respectively. Find the joint pdf of U and V, where $U = X_1 + X_2$ and $V = X_1 - X_2$. The joint pdf is given by

$$f_{X_1,X_2}(x_1, x_2) = f_{x_1}(x_1) \cdot f_{X_2}(x_2) = \lambda_1 e^{-\lambda_1 x_1} \cdot \lambda_1 e^{-\lambda_1 x_1}$$

Solve for X_1 and X_2 in terms of U and V

$$X_1 = \frac{U+V}{2} \qquad X_2 = \frac{U-V}{2}$$

Compute the Jacobian

$$J_1 = \begin{bmatrix} \frac{\partial U}{\partial x_1} & \frac{\partial U}{\partial x_2} \\ \frac{\partial V}{\partial x_1} & \frac{\partial V}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = -2$$

Finally, the joint pdf of U and V is given by

$$\begin{aligned} f_{U,V}(u,v) &= f_{X_1,X_2}(x_1 = \frac{u+v}{2}, x_2 = \frac{u-v}{2})| - 2|^{-1} \\ &= \lambda_1 \lambda_2 \cdot \frac{1}{2} \, e^{-\lambda_1 \frac{u+v}{2}} e^{-\lambda_1 \frac{u-v}{2}} \end{aligned}$$

3 Discrete Univariate Models

Discrete univariate models describe models with exactly one discrete random variable. A variable is said to be discrete if there exists a countable number of values within its mapped set (range). We use a probability mass function (pmf) to model distributions with discrete random variables. I will cover the following models:

1. Uniform
2. Bernoulli
3. Binomial
4. Poisson

3.1 Uniform

Definition 4. X is said to follow a *Uniform distribution* if

$$f(x|a,b) = \frac{1}{b-a+1} \qquad \text{for some } a,b \in \mathbb{Z} \text{ s.t. } b \geq a$$

Uniform distributions model events that have a fixed likelihood of occurring within a specified range. Note, the uniform distribution has a continuous case as well, depending on the range of values. Some notable applications include: rolling a fair die, flipping a fair coin, or picking from a deck of cards.

Example. Consider a fair die. Find the probability of rolling a 3.

$$f(3|1,6) = \frac{1}{6-1+1} = \frac{1}{6} = 0.16\bar{6}$$

3.2 Bernoulli

Definition 5. X is said to follow a *Bernoulli distribution* if

$$f(x|p) = P(X=x|p) = \begin{cases} p & 1 \\ 1-p & 0 \end{cases} \qquad 0 \leq p \leq 1$$

Bernoulli distributions model singular events that have exactly two possible outcomes. Some notable applications include: a weighted coin toss, the success or failure of a drug, or a free throw in basketball.

Example. Consider a weighted coin. The probability of heads is 65% (p = 0.65). Let X = 1 represent heads and X = 0 represent tails.

$$f(x|p=0.65) = P(X=x|p=0.65) = \begin{cases} 0.65 & 1 \\ 0.35 & 0 \end{cases}$$

Therefore, the probability of flipping heads $P(X=1|p=0.65) = 0.65$.

3.3 Binomial

Definition 6. X is said to follow a *binomial distribution* if

$$f(x|n,p) = P(X=x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0,1,...,n \quad \text{for some } n \in \mathbb{Z} \text{ and } 0 \leq p \leq 1$$

Binomial distributions model successive independent Bernoulli events (trials). Some notable applications include: flipping a coin multiple times, winning a sports tournament, or taking a multiple choice test.

Example. Consider a 10 question test. Each question has 4 choices with only 1 correct answer. Each choice is equally likely to occur. If you randomly guess on each question, find the probability of answering 8 questions correctly.

$$f(8|10, 0.25) = \binom{10}{8} 0.25^8 (1 - 0.25)^{10-8} = 0.024$$

Therefore, the likelihood of answering 8 questions correctly at random is 2.4%.

3.4 Poisson

Definition 7. X is said to follow a *Poisson distribution* if

$$f(x|\lambda) = P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, ... ; \quad 0 \leq \lambda < \infty$$

Poisson distributions model the frequency of independent events that occur within a specified time and/or space interval given a constant frequency rate. Some notable applications include: the number of calls arriving at a call center in an hour, the number of cars driving on a highway in a day, the number of customers entering a restaurant in a week, and the number of goals scored in a game.

Example. On average, 8 cars enter a highway every 60 minutes. Find the probability that 5 cars enter the highway in the next 60 minutes.

$$f(5|8) = \frac{e^{-8}8^5}{5!} = 0.0916$$

Therefore, the likelihood that 8 cars will enter the highway in the next 60 minutes is 9.16%.

4 Continuous Univariate Models

Continuous univariate models describe models with exactly one continuous random variable. A variable is said to be continuous if there exists an uncountably infinite number of values within its mapped set (range). We use a probability density function (pdf) to model distributions with continuous random variables. I will cover the following models:

1. Exponential
2. Gamma
3. Normal (aka Gaussian)
4. Beta
5. Chi squared
6. Student's t
7. F (aka Variance-Ratio)

4.1 Exponential

Definition 8. X is said to follow an *Exponential distribution* if

$$f(x|\lambda) = \frac{1}{\lambda}e^{-\frac{x}{\lambda}}, \quad 0 \leq x < \infty, \quad \lambda > 0$$

Exponential distributions follow directly from Poisson distributions, modeling the time it takes for a Poisson event to occur or, equivalently, the time between independent successive Poisson events. The exponential distribution is able to model such phenomena given only the average rate at which the event occurs (λ). It is a monotonically decreasing function and so it is most suitable for phenomena that decay at a decreasing rate. Some notable applications include: radioactive decay, time between bus stop arrivals, time between calls at a call center, or time between customer arrivals at a store.

Example. Consider a bus stop where, on average, a bus arrives every 15 minutes. Find the probability a bus will take longer than 15 minutes once a previous bus departs.

First, we obtain the pdf $\lambda = 15$.

$$f(x|\lambda = 15) = \frac{1}{15}e^{-\frac{x}{15}}$$

Then, we integrate with respect to x to obtain the cdf.

$$\begin{aligned} F(X > 15) &= 1 - F(X < 15) \\ &= 1 - \int_0^{15} \frac{1}{15}e^{-\frac{x}{15}}dx \\ &= 1 - \left(-e^{-\frac{x}{15}}\right)\Big|_0^{15} \\ &= 1 - (-e^{-1} + e^0) \\ &= 0.3678 \end{aligned}$$

Therefore, the likelihood of waiting longer than 15 minutes for the next bus is 36.78%.

Theorem 2. Let $X \sim exp(\lambda)$. Then, $\sum_{i=1}^{n} X_i \sim \Gamma(n, \lambda)$
The sum of n exponential distributions with rate λ follows a gamma distribution with shape parameter n and rate λ

4.2 Gamma

Definition 9. X is said to follow a *Gamma distribution* if

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}, \quad 0 \leq x < \infty, \quad \alpha > 0, \quad \beta > 0$$

The gamma distribution follows from the exponential distribution, modeling the time it takes for n successive independent Poisson events to occur. The gamma distribution is for the exponential what the binomial distribution is for the Bernoulli. Gamma distributions model events that hold strictly non-negative values given a shape parameter (α) and scale parameter (β). Because of these parameters, the gamma distribution provides a flexible fit to model positive, right-skewed phenomena. In particular, gamma is most suitable in modeling events that have extreme initial excitement and then calm down as time increases. Some notable applications include: social media posts, wait times in a queue, survival analysis, rainfall, and investment returns.

Example. A battery factory manufactures batteries that follow a gamma distribution with shape parameter $\alpha = 3$ and $\beta = 25$ hours. Find the probability that a battery lasts less than 20 hours.

First, we obtain the pdf.

$$f(X < 20|\alpha = 3, \beta = 25) = \frac{x^{3-1}e^{-\frac{x}{25}}}{\Gamma(3)25^3}$$

Then, we integrate with respect to x to obtain the cumulative density function (cdf).

$$\begin{aligned} F(X < 20) &= \int_0^{20} \frac{x^{3-1}e^{-\frac{x}{25}}}{\Gamma(3)25^3} dx \\ &= \frac{1}{2 \cdot 25^3} \int_0^{20} x^2 e^{-\frac{x}{25}} dx = 0.0474 \qquad\qquad \Gamma(1) = \Gamma(2) = 1 \end{aligned}$$

Therefore, the likelihood of a battery lasting less than 20 hours is 4.74%.

Theorem 3. $\Gamma(\alpha, \beta) \sim N(\alpha\beta, \alpha\beta^2)$ as $\alpha \to \infty$

Theorem 4. $\chi^2_p \sim \Gamma(\frac{p}{2}, 2)$
The χ^2 distribution with p degrees of freedom is a special case of the Gamma distribution with $\alpha = \frac{p}{2}$ and $\beta = 2$.

$$f(x|\alpha = \frac{p}{2}, \beta = 2) = \frac{x^{\frac{p}{2}-1}e^{-\frac{x}{2}}}{\Gamma(\frac{p}{2})2^{\frac{p}{2}}}$$

Theorem 5. $\Gamma(1, \beta) = exp(\beta)$
The exponential distribution is a special case of the Gamma distribution with $\alpha = 1$.

$$f(x|\alpha = 1, \beta) = \frac{x^{1-1}e^{-\frac{x}{\beta}}}{\Gamma(1)\beta^1} = \frac{1}{\beta}e^{-\frac{x}{\beta}}$$

Theorem 6. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Proof: Let $Z \sim N(0,1)$. Then, the pdf is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

We take the integral with respect to z to obtain the cdf. Rewrite using the property of symmetry.

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 2\int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 1$$

Let $y = \frac{1}{2}z^2 \Rightarrow z = (2y)^{-\frac{1}{2}}$ and $\frac{dy}{dz} = z \Rightarrow dz = z^{-1}dy$
Substituting,

$$\begin{aligned}
2\int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y} z^{-1} dy &= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-y} z^{-1} dy \\
&= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-y} (2y)^{-\frac{1}{2}} dy \\
&= \int_{0}^{\infty} e^{-y} y^{-\frac{1}{2}} dy \\
&= \int_{0}^{\infty} e^{-y} y^{\frac{1}{2}-1} dy = \sqrt{\pi} \\
&= \Gamma(\frac{1}{2}) = \sqrt{\pi}
\end{aligned}$$

□

4.3 Normal (aka Gaussian)

Definition 10. X is said to follow a *Normal distribution* if

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$$

Note, if $X \sim N(0,1)$, then the distribution is referred to as the *standard normal distribution.*

Normal distributions model the probability of an event occurring given the mean and variance of the population to which the event belongs. Notably, the *Central Limit Theorem* states that the distribution of the means of large samples repeatedly taken (independently and with replacement) from any population with any underlying distribution will approximate a normal distribution. Because of this fundamental principal in statistics, it is one of the most commonly used distributions and has many useful applications. Among them include: human height/weight/age/blood pressure, test scores, and seasonal temperatures.

Example. The average grade in a math class is 81% with a standard deviation of 5 percentage points. What is the probability a student earned 84% or above in the class?

We are asked to find $P(X \geq 84|\mu = 81, \sigma = 5)$. First, we standardize the value of interest to calculate the Z-score.

$$Z = \frac{X-\mu}{\sigma} = \frac{84-81}{5} = 0.6$$

Now that we have standardized, we need to find $P(z \geq 0.6) = 1 - P(z < 0.6)$. We consult a Z-score table or use statistical software to obtain the following value: $P(z \geq 0.6) = 1 - 0.7257 = 0.2742$. Therefore, the likelihood of a student earning 84% or above in the math class is 27.42%.

4.4 Beta

Definition 11. X is said to follow a *Beta distribution* if

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\beta(\alpha, \beta)}, \quad 0 \leq x \leq 1, \quad \alpha > 0, \quad \beta > 0$$

where $\beta(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

Beta distributions model the *probability* of an event's success (x), specifying the number of desired successes (α) and number of failures (β). This is similar to the binomial distribution, which models the *number* of an event's successes. Mathematically, one can see the similarities between the beta and binomial distributions through inspection of the numerator of their respective pdfs. Notice, the beta pdf is scaled by the term $\frac{1}{\beta(\alpha,\beta)}$ to bound the range such that values are between 0 and 1; this is necessary because of the fundamental principals of probability. The beta distribution offers a flexible fit for modeling probabilities and so provides versatile suitability. Some notable applications include: advertisement click rates, social media interaction rates, defective product likelihood, and medical treatment efficacy.

Example. Consider a medical treatment clinical trial. The treatment is a new drug and we want to find the probability that it has a success rate above 50%. A total of 500 people have already received the drug treatment, 255 of which experienced success. If 800 new patients are prescribed this treatment, find the probability of at least half of them experiencing success.

Note, prior knowledge informs us that the treatment has a $\frac{255}{500} = 0.51$ success rate. We want to find the likelihood of having at least a $\frac{400}{800} = 0.5$ success rate for the next round of trails.

First, we obtain the pdf.

$$f(x > 0.5|\alpha = 255, \beta = 245) = \frac{x^{255-1}(1-x)^{245-1}}{\beta(255, 245)}$$

Then, we integrate to obtain the cdf.

$$\begin{aligned}
F(x > 0.5) = 1 - F(x < 0.5) &= 1 - \int_0^{0.5} \frac{x^{254}(1-x)^{244}}{\beta(255,245)} dx \\
&= 1 - \frac{1}{\beta(255,245)} \int_0^{0.5} x^{254}(1-x)^{244} dx \\
&= 1 - \frac{\Gamma(500)}{\Gamma(255)\Gamma(245)} \int_0^{0.5} x^{254}(1-x)^{244} dx \\
&= 1 - \frac{499!}{254!244!} \int_0^{0.5} x^{254}(1-x)^{244} dx \\
&= 1 - 0.3272 \\
&= 0.6727
\end{aligned}$$

Therefore, the likelihood of the drug treatment succeeding for at least half of the patients to which it is prescribed is 67.27%.

Theorem 7. $f(x|\alpha = 1, \beta = 1) \sim Unif(0,1)$

Proof:

$$\begin{aligned}
f(x|\alpha = 1, \beta = 1) = \frac{x^{1-1}(1-x)^{1-1}}{\beta(1,1)} &= \frac{1}{\beta(1,1)} \\
&= \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)} = 1
\end{aligned}$$

4.5 Chi-square

Definition 12. X is said to follow an *Chi-square distribution* if

$$f(x|p) = \frac{x^{\frac{p}{2}-1} e^{-\frac{x}{2}}}{\Gamma(\frac{p}{2}) 2^{\frac{p}{2}}}, \quad 0 \leq x < \infty, \quad p = 1, 2, ...$$

Chi-square distributions model events that follow the sum of squared normal distributions. Because it models squared values, the chi-square distribution is suitable for modeling non-negative phenomena. Furthermore, since it inherits from the normal distribution, chi-square is particularly useful as a stress test to determine underlying properties of observed data such as independence and goodness-of-fit. Some notable applications include: testing relationships between categorical variables (test of independence), testing how well the observed data fits the expected data (goodness-of-fit test), and testing distributions across different populations (homogeneity test).

Example. We want to determine if a relationship exists between education level and political party association. We survey 500 people and collect data in a frequency table shown below.

	High School	College	Graduate	Total
Republican	54	32	12	98
Democrat	87	46	20	153
Independent	20	23	19	62
Total	161	101	51	313

First, we specify our *null hypothesis* H_o and *alternative hypothesis* H_a:
H_o : there exists no relationship between education level and political party association.
H_a : there exists an relationship between education level and political party association.
We also need to decide on a confidence level: $\alpha = 0.05$.

Then, we calculate the expected frequencies to compare our observed data to what the data "should" be, calculated as the ratio between elements and their respective row/column totals. We place this in a frequency table shown below.

	High School	College	Graduate
Republican	50.41	31.62	6.26
Democrat	42.5	22.5	9.77
Independent	3.96	4.56	3.76

Next, we compute the χ^2 test statistic defined and calculated as

$$\chi^2 = \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 16.05$$

where O_{ij} represents the observed values and E_{ij} represents the expected values.

Now, we need to calculate the *critical value*, defined as the degrees of freedom and calculated with the row number R and column number C as follows:

$$df = (R - 1)(C - 1) = (3 - 1)(3 - 1) = 4$$

Finally, because $\chi^2 = 16.05 > 4$, we reject H_o with a 95% confidence level. Therefore, there is a statistically significant amount of evidence supporting the claim that there is a relationship between education level and political party.

Theorem 8. Chi-square with 1 degree of freedom (df) is equivalent to the square of a standard normal distribution. Mathematically,

$$\exists Z \sim N(0,1) \text{ s.t } Z^2 \sim X_1^2$$

Theorem 9. Chi-square approximates normal if df > 90.

$$\chi_n^2 \approx N(\mu, \sigma) \text{ as } n \to \infty$$

4.6 Student's t

Definition 13. T is said to follow a *t distribution* with p degrees of freedom if

$$f(T|p) = \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} \frac{1}{\sqrt{p\pi}} \frac{1}{(1+(\frac{t^2}{p}))^{\frac{p+1}{2}}} \qquad -\infty < t < \infty; \quad p = 1, 2, ...$$

Student's t distributions model the comparison of two population means. It is a popular tool to approximate normal distributions when sample size is small and/or a population's variance is unknown, as it is often the case that a population's variance is unknown. The t distribution is similar in shape to a Gaussian bell curve, but with larger tails. Some notable applications include: comparing the efficacy of a treatment group to a placebo group in a medical trial, comparing one production line to another in a manufacturing factory, and comparing one class to another in a school.

Example. Consider 30 plants of the same species. Half of the plants are kept inside while the other half are kept outside. After one month, the outside plants have grown an average of 2 inches with a standard deviation of 0.5 inches, the inside plants have grown an average of 1.8 inch with a standard deviation of 0.6 inches. Determine if there is a statistically significant difference in plant growth between the two groups. This exercise is left for the reader.

Theorem 10. $T_p \sim N(0, 1)$ as $p \to \infty$

4.7 F (aka Variance-Ratio)

Definition 14. Consider two independent chi-square distributions, χ_p^2 and χ_q^2, where p, q are the degrees of freedom, respectively. X is said to follow a *F distribution* if

$$f(x|p,q) = \frac{(\frac{\chi_p^2}{p})}{(\frac{\chi_q^2}{q})}$$

The pdf is expressed as,

$$f(x|p,q) = \frac{\Gamma(\frac{p+q}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})} \left(\frac{p}{q}\right)^{p/2} \frac{x^{\frac{p-2}{2}}}{(1+x(\frac{p}{q}))^{\frac{p+q}{2}}}; \quad 0 \leq x < \infty; \quad p, q = 1, 2, ...$$

F distributions model the ratio of variances between two independent and normally distributed populations. Because of this, it inherits many properties from the chi-square distribution such as right skew and non-negativity. The suitability of the F distribution is similar to that of the T distribution with subtle differences. The T distribution tests the difference in *means* between *exactly two* populations whereas the F distribution tests the difference in *variances* between *two or more* populations. Some notable applications include: testing equality of variances between two or more populations in an ANOVA test, and testing the *power* of a model in regression analysis which gauges it's significance.

Example. A chef wants to perfect her apple pie recipe. She baked 10 pies each for recipes A, B, and C, making a total of 30 pies. She cuts them into eighths and hands 1 slice each to 240 customers (at random) for them taste and rate their pie on a scale from 1-5 (5 being the most satisfied). On average, customers rated recipe A with a score of 3.2, recipe B with a score of 4.6, and recipe C with a score of 4.1. Determine if there exists statistically significant differences in customer satisfaction between the 3 different recipes. This exercise is left for the reader.

Theorem 11. $F_{1,q} = T_p^2$

Proof: The T test statistic is defined as

$$T = \frac{Z}{\sqrt{\frac{W}{q}}} \qquad \text{for some } Z \sim N(0,1) \text{ and } W \sim \chi_q^2$$

Substituting, we obtain

$$T = \frac{Z}{\sqrt{\frac{\chi_q^2}{q}}}$$

Thus it follows

$$T^2 = \frac{Z^2}{\left(\frac{\chi_q^2}{q}\right)} = \frac{\left(\frac{\chi_1^2}{1}\right)}{\left(\frac{\chi_q^2}{q}\right)} \sim F_{1,q} \qquad \text{(Theorem 7)}$$

5 Multivariate Models

Multivariate models describe models with more than one random variable. From this, it is necessary to express the response variables as a random vector. Such models are helpful in modeling the more complex world. Realistically, most phenomena have more than one variable so multivariate models are more suitable for our natural environment compared to their univariate counterpart.

5.1 Joint Distributions

Joint distributions model the combination of two or more variables. They are used to describe the probability of multiple independent events as they occur simultaneously. Joint distributions completely describe multivariate models. Similar to univariate models, joint distributions can be expressed as a joint pmf or joint pdf depending on the continuity of the random variables they describe. We consider both cases below.

5.1.1 Discrete

Definition 15. $(X_1, ..., X_n)$ have a joint probability mass function expressed as

$$f(X_1, ..., X_n) = P(X_1 = x_1, ..., X_n = x_n) = \sum_{X_1} ... \sum_{X_n} f(x_1, ..., x_n) = 1$$

Example. Consider a fair coin and die. We want to model the joint distribution of flipping the coin and rolling the die simultaneously. Note, the coin flip is independent of the die roll. Because both variables have a discrete range, we will describe their joint distribution with a probability mass function. Write the joint pmf and use it to find the likelihood of simultaneously rolling a 4 and flipping a heads after one trial.

Let random variables D and C describe the probability of the desired outcome of the die and coin, respectively. Because of independence, we know the joint pmf, denoted $f(D, C)$, is equal to the product of their marginal distributions. Mathematically,

$$f(D, C) = f(D) \cdot f(C) \qquad \text{for some, } f(D) \sim Bernoulli(p = 1/6)$$
$$f(C) \sim Bernoulli(p = 1/2)$$

It follows,

$$\begin{aligned} f(D = d, C = c) &= f(D = d) \cdot f(C = c) \\ &= f(D = 4) \cdot f(C = Heads) \\ &= (1/6)(1/2) \\ &= 2/3 = 0.333 \end{aligned}$$

Therefore, the likelihood of simultaneously flipping a heads and rolling a 4 is 33.33%.

5.1.2 Continuous

Definition 16. $(X_1, ..., X_n)$ have a joint probability density function expressed as

$$f(X_1, ..., X_n) = P(X_1 = x_1, ..., X_n = x_n) = \int_{X_1} ... \int_{X_n} f(x_1, ..., x_n) dx_1 ... dx_n = 1$$

Example. We want to model the likelihood of someone's height H (in inches) and weight W (in pounds). Let H and W be jointly continuous random variables with joint pdf:

$$f(H, W) = P(H = h | W = w) = \frac{w}{990h^2}$$

$$\text{where,} \quad 30 < w < 300, \quad 30 < h < 90$$

Find $f(H < 60, W < 130)$

$$\begin{aligned}
f(H < 60, W < 130) &= \int_H \int_W f(H < 60, W < 130)\ dwdh \\
&= \frac{1}{990} \int_{30}^{60} \int_{30}^{130} \frac{w}{h^2}\ dwdh \\
&= \frac{1}{990} \int_{30}^{60} \left[\frac{w^2}{2h^2} \right]_{30}^{130} dh \\
&= \frac{1}{990} \int_{30}^{60} \frac{130^2 - 30^2}{2h^2}\ dh \\
&= \frac{1}{990} \left[\frac{-8000}{h} \right]_{30}^{60} \\
&= \frac{1}{990} \left[\frac{-8000}{60} + \frac{8000}{30} \right] \\
&= 0.1346
\end{aligned}$$

Therefore, the likelihood of someone weighing between 30 and 130 pounds while also having a height between 30 and 60 inches is 13.46%.

5.2 Marginal Distributions

Marginal distributions allow us to focus on one particular variable taken from a joint distribution where each individual random variable has a univariate distribution. A note of caution, marginal distributions do not necessarily completely describe multivariate models; thus, we are not necessarily able to determine a joint distribution from the sum of marginals. Conversely, we are and, in fact, must determine marginal distributions from their associated joint distributions.

5.2.1 Discrete

Definition 17. Let $(X_1, ..., X_n)$ be a discrete random vector with joint pmf expressed as $f(X_1, ..., X_n)$. X_i has a marginal pmf denoted

$$f(X_i) = P(X_i = x_i) = \sum ... \sum f(x_1, ..., x_{i-1}, x_{i+1}, ..., x_n)$$

5.2.2 Continuous

Definition 18. Let $(X_1, ..., X_n)$ be a continuous random vector with joint pdf expressed as $f(X_1, ..., X_n)$. X_i has a marginal pdf denoted

$$f(X_i) = P(X_i = x_i) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) dx_1 ... dx_{i-1} dx_{i+1} dx_n$$

Example. We will use the previous example introduced in subsection **4.2.1**. The marginal distributions, denoted $f_H(h)$ and $f_W(w)$, are as follows

$$f_H(h) = \int_{30}^{130} f(h,w)\ dw = \int_{30}^{130} (w/h^2)\ dw = \frac{8000}{h^2}$$

$$f_W(w) = \int_{30}^{60} f(h,w)\ dh = \int_{30}^{60} (w/h^2)\ dh = \frac{w}{60}$$

6 Mixture Distributions

Mixture distributions engender from multivariate models that deal with *different types* of distributions. Until now, we have dealt solely with random vectors wherein all the variables follow the same distribution. Realistically, this is not always the case.

Example. Consider the random vector $[X, Y, Z]$, where $X|Y$ follows a *binomial distribution*, $Y|Z$ follows an *Poisson distribution*, and Z follows a *gamma distribution*. This is a model involving a mixture of different distributions chained together through conditional relationships, called a *hierarchical model*. We say hierarchical models lead to mixture distributions. This can be written mathematically as follows,

$$\begin{aligned} X|Y &\sim binom(n,p) \\ Y|Z &\sim Poisson(Z) \\ Z &\sim gamma(\alpha,\beta) \end{aligned}$$

7 Afterword

This is not a complete guide to distributions of functions of random variables. Notable distributions left out include: negative binomial, cauchy, double exponential, logistic, lognormal, pareto, and weibull. Furthermore, some proofs and exercises were proposed and left unsolved. The diligent reader will engage in distributions, proofs, and exercises of personal significance.

Lecture notes from Prof. Nicolas Christou's Mathematical Statistics course (UCLA STATS 404) were referenced throughout the writing of this paper along with "Statistical Inference" by Casella and Berger.