

Overview: Probability Distributions

Michelle Arnetta

Summary

An overview of different types of distributions for both continuous random variables and discrete random variables.

How to use

This overview will contain the following content on each distribution:

- Where to use
- Notation
- Parameter(s)
- Mean and Expected Value (mean = expected value, and the expected value of X will be represented as $E(X)$)
- Variance (the variance of X will be represented as $\text{Var}(X)$)
- Probability Mass Function (PMF) for discrete random variables, or Probability Density Function (PDF) for continuous random variables
- Cumulative Distribution Function (CDF), where possible
- Interactive Figure, where possible
- Example(s)

Although it is not an exhaustive list, this overview can be treated as an introduction to commonly used types of distributions.

Discrete Random Variables

Uniform Distribution (Discrete)

Where to use: The discrete uniform distribution is used when all discrete values in the interval a to b are equally likely.

Notation: $X \sim \text{Uniform}(a, b)$ or $X \sim U(a, b)$

Parameters:

- a = minimum value
- b = maximum value where all values of X for $a \leq x \leq b$ are equally likely

Mean and Expected Value: $E(X) = \frac{a+b}{2}$

Variance: $\text{Var}(X) = \frac{(b-a+1)^2-1}{12}$

PMF: $P(X = x) = \frac{1}{n}$ where $n = b - a + 1$

CDF: $P(X \leq x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{\lfloor x \rfloor - a + 1}{n} & \text{if } a < x < b \\ 1 & \text{if } x \geq b \end{cases}$

where $n = b - a + 1$ and $\lfloor x \rfloor$ is x rounded down to the nearest integer

Figure

PUT FIGURE HERE TOM

Example: You roll a fair six-sided die, where all outcomes (1, 2, 3, 4, 5, and 6) are equally likely. This can be expressed as $X \sim U(1, 6)$. It means 1 is the minimum value and 6 is the maximum value, where all discrete values of X for $1 \leq x \leq 6$ are equally likely.

Bernoulli Distribution

Where to use: The Bernoulli distribution is used for binary data, where one trial is conducted with only two possible outcomes. Examples include success/failure, yes/no, and heads/tails. X indicates whether the trial is a success (when $X = 1$) or failure (when $X = 0$).

Notation: $X \sim \text{Bernoulli}(p)$

Parameter: p = probability of success (where $0 \leq p \leq 1$)

Mean and Expected Value: $E(X) = p$

Variance: $\text{Var}(X) = p(1 - p)$

$$\text{PMF: } P(X = x) = \begin{cases} 1 - p & \text{if } x = 0 \\ p & \text{if } x = 1 \end{cases}$$

$$\text{CDF: } P(X \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - p & \text{if } 0 \leq x < 1 \\ p & \text{if } x \geq 1 \end{cases}$$

Figure

PUT FIGURE HERE TOM

Example: You flip a coin, and the probability of getting 'heads' is 0.5. Taking 'heads' as a success, this can be expressed as $X \sim \text{Bernoulli}(0.5)$, meaning the probability of success in each trial is 0.5.

Binomial Distribution

Where to use: The binomial distribution is used when there are a fixed number of trials (n) and only two possible outcomes for each trial. X represents the number of successes.

Notation: $X \sim \text{Binomial}(n, p)$ or $X \sim B(n, p)$

Parameters:

- n = number of trials
- p = probability of success (where $0 \leq p \leq 1$)

Mean and Expected Value: $E(X) = np$

Variance: $\text{Var}(X) = np(1 - p)$

$$\text{PMF: } P(X = x) = \frac{n!}{(n-x)!x!} p^x q^{(n-x)}$$

CDF: $P(X \leq x) = I_q(n - \lfloor x \rfloor, 1 + \lfloor x \rfloor)$ where $I_x(a, b)$ is the regularized incomplete beta function for a and b with x being the upper bound for integration

Figure

PUT FIGURE HERE TOM

Example: You flip a coin 10 times, and the probability of getting 'heads' is 0.5. Taking 'heads' as a success, this can be expressed as $X \sim B(10, 0.5)$, meaning 10 trials are conducted, where the probability of success in each trial is 0.5.

Multinomial Distribution

Where to use: The multinomial distribution is used when there are a fixed number of trials (n) and more than two possible outcomes for each trial. X_i represents the number of times a specific outcome occurs. Therefore, the mean, variance, and expected value of multinomial distributions are calculated for each X_i , not X .

Notation: $X \sim \text{Multinomial}(n, p)$ or $X \sim M(n, p)$

Parameters:

- n = number of trials
- p = vector with probabilities of each outcome (expressed as p_1, \dots, p_k where k = number of possible mutually exclusive outcomes)

Mean and Expected Value: $E(X_i) = np_i$

Variance: $\text{Var}(X_i) = np_i(1 - p_i)$

PMF: $P(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}$

CDF: COULDNT FIND

Figure

PUT FIGURE HERE TOM

Example: There is a candy jar consisting of 5 red candies, 3 blue candies, and 7 yellow candies.

- The probability of drawing a red candy is $\frac{1}{3}$.
- The probability of drawing a blue candy is $\frac{1}{5}$.
- The probability of drawing a yellow candy is $\frac{7}{15}$

You draw 3 candies from the jar, replacing as you go along. This can be expressed as $X \sim M(3, \frac{1}{3}, \frac{1}{5}, \frac{7}{15})$. It means 3 trials are conducted, where $p_1 = \frac{1}{3}$, $p_2 = \frac{1}{5}$, and $p_3 = \frac{7}{15}$ (and $k = 3$).

Poisson Distribution

Where to use: The Poisson distribution is used when a specific event occurs at some rate λ , and you are counting X , the number of times this event occurs in some interval.

Notation: $X \sim \text{Poisson}(\lambda)$ or $X \sim \text{Pois}(\lambda)$

Parameter: λ = number of times an event occurs within a specific period of time

Mean and Expected Value: $E(X) = \lambda$

Variance: $\text{Var}(X) = \lambda$

PMF: $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$

CDF: $P(X \leq x) = \sum_{i=1}^{\lfloor x \rfloor} \frac{\lambda^x e^{-\lambda}}{x!}$ where $\lfloor x \rfloor$ is x rounded down to the nearest integer

Figure

PUT FIGURE HERE TOM

Example: Customers enter Cantor's Confectionery at an average rate of 20 people per hour. This can be expressed as $X \sim \text{Pois}(20)$, meaning the event of customers entering the store occurs 20 times within an hour.

Negative Binomial Distribution

Where to use: The negative binomial distribution is often used to handle over-dispersed data, which means the variance exceeds the mean. It can serve as an alternative to the Poisson distribution, as the Poisson distribution assumes that the mean is equal to the variance. X represents the number of trials required to reach the targeted number of successes r .

Notation: $X \sim \text{NB}(r, p)$

Parameters:

- r = targeted number of successes
- p = probability of success (where $0 \leq p \leq 1$)

Mean and Expected Value: $E(X) = \frac{r(1-p)}{p}$

Variance: $\text{Var}(X) = \frac{r(1-p)}{p^2}$

PMF: $P(X = x) = \frac{(x+r-1)!}{(r-1)!x!} (1-p)^x p^r$

CDF: $P(X \leq x) = \sum_{i=1}^x \frac{(x+r-1)!}{(r-1)!x!} (1-p)^x p^r$

Figure

PUT FIGURE HERE TOM

Example: You flip a coin multiple times, and the probability of getting 'heads' is 0.5. You decide to stop flipping the coin once you get 3 'heads'; these do not have to be consecutive. Taking 'heads' as a success, this can be expressed as $X \sim \text{NB}(3, 0.5)$. It means the probability of success is 0.5, and you will stop conducting trials after you reach 3 successes.

Geometric Distribution

Where to use: The geometric distribution is used to count X , the number of trials until a successful outcome is reached.

Notation: $X \sim \text{Geometric}(p)$

Parameter: p = probability of success (where $0 \leq p \leq 1$)

Mean and Expected Value: $E(X) = \frac{1}{p}$

Variance: $\text{Var}(X) = \frac{1-p}{p^2}$

PMF: $P(X = x) = (1 - p)^{x-1}p$

CDF: $P(X \leq x) = \begin{cases} 1 - (1 - p)^x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$

Figure

PUT FIGURE HERE TOM

Example: You flip a coin multiple times, and the probability of getting 'heads' is 0.5. You decide to stop flipping the coin once you get a 'heads'. Taking 'heads' as a success, this can be expressed as $X \sim \text{Geometric}(0.5)$. It means the probability of success is 0.5, and you will stop conducting trials after you reach a success.

Continuous Random Variables

Uniform Distribution (Continuous)

Where to use: The continuous uniform distribution is used when all continuous values in the interval a to b are equally likely.

Notation: $X \sim \text{Uniform}(a, b)$ or $X \sim U(a, b)$

Parameters:

- a = minimum value
- b = maximum value where all values of X for $a \leq x \leq b$ are equally likely

Mean and Expected Value: $E(X) = \frac{a+b}{2}$

Variance: $\text{Var}(X) = \frac{(b-a)^2}{12}$

PDF: $P(X = x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$

$$\text{CDF: } P(X \leq x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } x > b \end{cases}$$

Figure

PUT FIGURE HERE TOM

Example:

A machine from Cantor's Confectionery is programmed to chop long candy bars into pieces, each with a length between 30 milimeters to 50 milimeters. Due to variations in the machine, each continuous value between this interval is equally likely. This can be expressed as $X \sim U(30, 50)$. It means 30 is the minimum value and 50 is the maximum value, where all continuous values of X for $30 \leq x \leq 50$ are equally likely.

Normal Distribution

Where to use: The normal distribution is used to model continuous random variables, which can include any positive or negative real values. The use of this distribution is often justified by the Central Limit Theorem: as the sample size increases, the distribution of sample means will resemble a normal distribution more and more closely.

Notation: $X \sim \text{Normal}(\mu, \sigma^2)$ or $X \sim N(\mu, \sigma^2)$

Parameters:

- μ = mean (location parameter)
- σ^2 = variance (squared scale parameter), where σ = standard deviation (scale parameter)

Mean and Expected Value: $E(X) = \mu$

Variance: $\text{Var}(X) = \sigma^2$

PDF: $P(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

CDF: $P(X \leq x) = \frac{1}{2} [1 + \text{erf}(\frac{x-\mu}{\sigma\sqrt{2}})]$ where $\text{erf}(x)$ is the error function of x

Figure

PUT FIGURE HERE TOM

Example: The lengths of chocolate bars produced by Cantor's Confectionery follow a normal distribution with a mean of 5.6 inches and a standard deviation of 1.2. This can be expressed as $X \sim N(5.6, 1.2^2)$, meaning the data is normally distributed, centered at 5.6 (location parameter) and stretched by 1.2 (scale parameter).

Lognormal Distribution

Where to use: The lognormal distribution is used to model continuous random variables with values that are both real and non-negative, wherein the logarithms of these variables follow a normal distribution.

Notation: $X \sim \text{Lognormal}(\mu, \sigma^2)$

Parameters:

- μ = logarithm of location parameter
- σ^2 = logarithm of scale parameter

Mean and Expected Value: $E(X) = \exp(\mu + \frac{\sigma^2}{2})$ where $\exp(x) = e^x$

Variance: $\text{Var}(X) = [\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2)$ where $\exp(x) = e^x$

PDF: $P(X = x) = \frac{1}{x\sigma\sqrt{2\pi}}\exp(-\frac{(\ln x - \mu)^2}{2\sigma^2})$ where $\exp(x) = e^x$ and $\ln x$ is the natural logarithm of x

CDF: $P(X \leq x) = \frac{1}{2}[1 + \text{erf}(\frac{\ln x - \mu}{\sigma\sqrt{2}})]$ where $\text{erf}(x)$ is the error function of x and $\ln x$ is the natural logarithm of x

Figure

PUT FIGURE HERE TOM

Example: The logarithms of Cantor's Confectionery's stock prices follow a normal distribution. The mean of the stock prices' natural logarithms is 8.01, whereas the variance of the stock prices' natural logarithms is 3. This can be expressed as $X \sim \text{Lognormal}(8.01, 3)$, meaning the logarithm of the location parameter is 8.01 and the logarithm of scale parameter is 3.

Exponential Distribution

Where to use: The exponential distribution is used when X is the waiting time before a certain event occurs. It is similar to the geometric distribution, but the exponential distribution uses continuous waiting time instead of the integer number of trials.

Notation: $X \sim \text{Exponential}(\lambda)$ or $X \sim \text{Exp}(\lambda)$

Parameter: λ = number of times an event occurs within a specific period of time

Mean and Expected Value: $E(X) = \frac{1}{\lambda}$

Variance: $\text{Var}(X) = \frac{1}{\lambda^2}$

PDF: $P(X = x) = \lambda e^{-\lambda x}$

CDF: $P(X \leq x) = 1 - e^{-\lambda x}$

Figure

PUT FIGURE HERE TOM

Example: Customers enter Cantor's Confectionery at an average rate of 20 people per hour, and the time distance between each visit can be modeled by an exponential distribution. This can be expressed as $X \sim \text{Exp}(20)$.

Gamma Distribution

Where to use: The gamma distribution generalizes the exponential distribution, allowing for greater or lesser variance. It is used to model positive continuous random variables that have skewed distributions.

Notation: $X \sim \text{Gamma}(\alpha, \theta)$ or $X \sim \text{Gam}(\alpha, \theta)$

Parameters:

- $\alpha = \frac{\mu^2}{\sigma^2}$ (shape parameter)
- $\theta = \frac{\sigma^2}{\mu}$ (scale parameter) where μ = mean and σ^2 = variance

Mean and Expected Value: $E(X) = \alpha\theta$

Variance: $\text{Var}(X) = \alpha\theta^2$

PDF: $P(X = x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta}}$ where $\Gamma(x)$ is the gamma function of x

CDF: $P(X \leq x) = \frac{1}{\Gamma(\alpha)} \text{Gam}(\alpha, \frac{x}{\theta})$ where $\text{Gam}(\alpha, \theta)$ is the PDF of the gamma distribution for parameters α and θ

Figure

PUT FIGURE HERE TOM

Example: You collect historical data on the time to failure of a machine from Cantor's Confectionery. The mean is 83 days and the variance is 50.3. You can then use this to estimate the shape and scale parameters of the Gamma Distribution:

- $\alpha = \frac{83^2}{50.3} = 136.958250497 \approx 137$
- $\theta = \frac{50.3}{83} = 0.60602409638 \approx 0.61$

The distribution can be expressed as $X \sim \text{Gam}(137, 0.61)$, where the shape parameter is 137 and the scale parameter is 0.61.

Beta Distribution

Where to use: The beta distribution is used to model the distribution of probabilities or proportions. Hence, $0 \leq X \leq 1$.

Notation: $X \sim \text{Beta}(\alpha, \beta)$

Parameters:

- $\alpha = k + 1$ (shape parameter)
- $\beta = n - k + 1$ (shape parameter) where k = number of successes and n = number of trials

Mean and Expected Value: $E(X) = \frac{\alpha}{\alpha + \beta}$

Variance: $\text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

PDF: $P(X = x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$ where $B(x, y)$ is the beta function of x and y

CDF: $P(X \leq x) = I_x(\alpha, \beta)$ where $I_x(a, b)$ is the regularized incomplete beta function for a and b with x being the upper bound for integration

Figure

PUT FIGURE HERE TOM

Example: Cantor's Confectionery is visited by 10 customers, and 6 of them purchase something from the store. Taking the buying customers as successes and the total visiting customers as number of trials, there would be 6 successes, allowing you to find the following parameters:

- $\alpha = 6 + 1 = 7$
- $\beta = 10 - 6 + 1 = 5$

Then the distribution of the probabilities of a customer purchasing from Cantor's Confectionery can be expressed as $X \sim \text{Beta}(7, 5)$, meaning the first shape parameter is 7 and the second shape parameter is 5.

χ^2 Distribution

Where to use: The χ^2 distribution is used for hypothesis testing, such as for goodness of fit tests and tests for independence. It is a special case of the gamma distribution, as $\chi^2(k) = \text{Gam}(\frac{k}{2}, 2)$.

Notation: $X \sim \chi^2(k)$

Parameter: k = degrees of freedom in sample

Mean and Expected Value: $E(X) = k$

Variance: $\text{Var}(X) = 2k$

PDF: $P(X = x) = \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$ where $\Gamma(x)$ is the gamma function of x

CDF: $P(X \leq x) = \frac{1}{\Gamma(\frac{k}{2})} \text{Gam}(\frac{k}{2}, \frac{x}{2})$ where $\Gamma(x)$ is the gamma function of x and $\text{Gam}(\alpha, \theta)$ is the PDF of the gamma distribution for parameters α and θ

Figure

PUT FIGURE HERE TOM

Examples:

- **Goodness of Fit Example:** You have a six-sided die with six possible outcomes: 1, 2, 3, 4, 5, and 6. You calculate the expected frequencies of each outcome. Then you roll the die many times and record the observed frequencies of each outcome. Since there are 6 categories,

$$\text{degrees of freedom} = \text{number of categories} - 1 = 6 - 1 = 5$$

This can be expressed as $X \sim \chi^2(5)$, meaning the degrees of freedom is 5.

- **Test for Independence Example:** You are investigating whether there is a correlation between two variables: candy color and flavor. You have 5 categories of colors and 3 categories of flavors. Calculating the degrees of freedom can be done with:

$$(\text{number of rows} - 1)(\text{number of columns} - 1) = (5 - 1)(3 - 1) = (4)(2) = 8$$

Hence, $X \sim \chi^2(8)$, meaning the degrees of freedom is 8.

***F* Distribution**

Where to use: The F distribution is used for the ratio of two independent random χ^2 variables. It is commonly used as a reference distribution in hypothesis testing to compare two variances or more than two means, such as Analysis of Variance (ANOVA) tests.

Notation: $X \sim F(d_1, d_2)$

Parameters:

- d_1 = numerator degrees of freedom
- d_2 = denominator degrees of freedom

Mean and Expected Value: $E(X) = \frac{d_2}{d_2-2}$ for $d_2 > 2$

Variance: $\text{Var}(X) = \frac{2d_2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$

PDF: $P(X = x) = \frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}}}}{x B(\frac{d_1}{2}, \frac{d_2}{2})}$ where $B(x, y)$ is the beta function of x and y

CDF: $P(X \leq x) = I_{\frac{d_1 x}{d_1 x + d_2}}(\frac{d_1}{2}, \frac{d_2}{2})$ where $I_x(a, b)$ is the regularized incomplete beta function for a and b with x being the upper bound for integration

Figure

PUT FIGURE HERE TOM

Example: You have three independent groups of data containing Cantor's Confectionery chocolate bar lengths, and the total sample size is 90. From this, you would like to conduct an ANOVA test investigating if there is a statistically significant difference between each group's means. You find the degrees of freedom using the following methods:

- numerator degrees of freedom = number of groups $- 1 = 3 - 1 = 2$
- denominator degrees of freedom = sample size - number of groups = $90 - 3 = 87$

The F distribution, which will be used as a reference distribution for the ANOVA test, can be expressed as $X \sim F(2, 87)$, meaning the numerator degrees of freedom is 2 and the denominator degrees of freedom is 87.

t Distribution

Where to use: The t distribution is a special case of the F distribution, as $(t(v))^2 = F(1, v)$. This distribution is used for continuous random variables with heavier tails than the normal distribution, and it is often employed in hypothesis testing when the population standard deviation is unknown.

Notation: $X \sim t(v)$

Parameter: v = degrees of freedom (equal to denominator degrees of freedom if related to F distribution)

Mean and Expected Value: $E(X) = 0$

Variance: $\text{Var}(X) = \frac{v}{v-2}$ for $v > 2$

PDF: $P(X = x) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi v} \Gamma(\frac{v}{2})} (1 + \frac{x^2}{v})^{-\frac{v+1}{2}}$ where $\Gamma(x)$ is the gamma function of x

CDF: $P(X \leq x) = \frac{1}{2} + x \Gamma(\frac{v+1}{2}) (\frac{{}_2F_1(\frac{1}{2}, \frac{v+1}{2}; \frac{3}{2}; -\frac{x^2}{v})}{\sqrt{\pi v} \Gamma(\frac{v}{2})})$ where $\Gamma(x)$ is the gamma function of x and ${}_2F_1(a, b; c; z)$ is the hypergeometric function for a, b, c , and z

Figure

PUT FIGURE HERE TOM

Example: You have a sample of 40 measurements of Cantor's Confectionery chocolate bar lengths. From this, you would like to conduct a one sample t-test comparing the sample to a hypothesized mean. You find the degrees of freedom:

$$\text{degrees of freedom} = \text{sample size} - 1 = 40 - 1 = 39$$

The t distribution, which will be used as a reference distribution for the t-test, can be expressed as $X \sim t(39)$, meaning the degrees of freedom is 39.

Further reading

[For more information on hypothesis testing, please see Guide: Introduction to hypothesis testing.]

[For more information on probability, please see Guide: Introduction to Probability.]

[For more information on mean, variance, and standard deviation, please see Guide: Mean, Variance, and Standard Deviation.]

[For more information on PMFs, PDFs, and CDFs, please see Guide: PMFs, PDFs, and CDFs.]

Version history

v1.0: initial version created 4/25 by Michelle Arnetta as part of a University of St Andrews VIP project.

[This work is licensed under CC BY-NC-SA 4.0.](#)