MATH 460: Probability & Statistics January_2019

Ch 4 Special Probability Distribution Functions

Class 4 Notes **Dr. Basilio**

Wed Jan_16 $\cup$ Tues Jan_22

* * *

Guiding Question(s)

(1) What do the binomial probability distributions look like?

(2) What do probability density functions look like?

(3) How can we calculate probabilities for continuous random variables?

Chapter 4: Probability Distribution Functions

Binomial Distribution

Definition 1: Binomial-Probability-Distribution

For a **Binomial probability distribution** it is important that we have a random process or experiment and must satisfy:

1. The experiment has a fixed number of trails
2. The trials must be **independent** (that is, the outcome of any individual trial doesn't effect the probabilities of other trials)
3. Each trail must have all outcomes classificed into exactly two categories: *success* and *failure*
4. The probability of success remains the same in all trails

Let $P(X = x)$ denote the **probability of exactly x successful trails out n in a binomial probability distribution**, then

$$\boxed{P(X = x) = \binom{n}{x} p^x q^{n-x} =_n C_x p^x q^{n-x}} \tag{1}$$

where

- n be the total number of trails run in the experiment
- X be a random variable of a single "successful" trail
- p be the probability of the successful trail X
- q be the probability of trail X failing. (NOTE: $p + q = 1$, or $q = 1 - p$)
- x be the number of successful trials of X. So notice that x can take values from 0 up to n, i.e. $x = 0, 1, 2, 3, \ldots, n$.

Recall that: $\binom{n}{x} =_n C_x = \frac{n!}{x!(n-x)!}$.

Example 1: Binomial-Distribution-Probability

A die is tossed 3 times. What is the probability of

(a) No fives turning up?

Solution: Let $n = 3$. Let $X = \{$ rolling a 5 $\}$. The probability of X being successful is $p = 1/6$ and $q = 5/6$ is probability of X failing. We want X to be successful zero times! so $x = 0$:

$$\begin{aligned} P(X = 0) &= \binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{3-0} \\ &= \left(\frac{3!}{0!(3-0)!}\right)\cdot 1\cdot\left(\frac{5}{6}\right)^3 \\ &= \left(\frac{5}{6}\right)^3 = \frac{125}{216} \approx 0.5787037 \approx 0.579 \end{aligned}$$

So the probability of no five turning up if a die is tossed 3 times is approximately 57.9%. □

(b) 1 five turning up?

Solution: Let $n = 3$. Let $X = \{$ rolling a 5 $\}$. The probability of X being successful is $p = 1/6$ and $q = 5/6$ is probability of X failing. We want X to be successful once so $x = 1$:

$$\begin{aligned} P(X = 1) &= \binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{3-1} \\ &= \left(\frac{3!}{1!(3-1)!}\right)\cdot\frac{1}{6}\cdot\left(\frac{5}{6}\right)^2 \\ &= 3\cdot\frac{1}{6}\cdot\frac{25}{36} \\ &= \frac{25}{72} \approx 0.347\bar{2} \approx 0.3472 \end{aligned}$$

So the probability of one five turning up if a die is tossed 3 times is approximately 34.7%. □

(c) 3 fives turning up?

Solution: n, X, p, and q are exactly as in part (a) and (b), but this time we want $x = 3$:

$$\begin{aligned} P(X = 3) &= \binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^{3-3} \\ &= \frac{1}{216} \approx 0.00462963 \approx 0.0046 \end{aligned}$$

So the probability of three fives turning up if a die is tossed 3 times is approximately 0.5%. □

Definition 2: Binompdf-vs-Binomcdf

USING CALCULATOR TI83: binompdf(n,p,x)

$\boxed{DIST}$ key in yellow ($\boxed{2nd}$ > $\boxed{VARS}$) > Scroll to 10 “binompdf” or scroll to A “binomcdf”

- Binompdf is when we want exactly x trails to be successful so this is binomial distribution pdf. Thus this is 1-valued random variables.
- Binomcdf is when we want multiple values of x to be true. It is defined as

$$\text{binomcdf}(n, p, x) = P(X \leq x) \tag{2}$$

Notice the sneaky "≤" less than or equal to sign in the binomcdf. This means:

$$\begin{aligned}\text{binomcdf}(n,p,x) &= P(X \leq x) = P(X = 0,1,2,\ldots,x)\\ &= P(X=0)+P(X=1)+P(X=2)+\cdots+P(X=x)\end{aligned}$$

This can help us with "at most" and "at least" type of problems.

Example 2: Binomcdf-probability

What is the probability of at least four successful trials out of a total of 6 trails in a random experiment, with probability of success of a single trial being 25%?

Solution: Here $n = 6$ and $p = 0.25$. Notice we must convert the percentage to a decimal. We don't need to know what X is. We want at least four successful trials, so $x = 4$ is when exactly 4 trials are successful. When $x = 5$ is when exactly 5 trials are successful, and $x = 6$ is when exactly 6 trials are successful.

One way to find the answer is: $P(X = 4) + P(X = 5) + P(X = 6)$. This is alot to type into the calculator.

ANOTHER WAY (FASTER): we can use binomcdf! Because binomcdf(6,0.25,3) calculates $P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$ quickly we can use this to find the remaining probability with the "1 minus trick:"

$$\begin{aligned}\underbrace{P(X \geq 4)}_{\text{at least 4 successful trials}} &= 1 - P(X < 4) \qquad \text{(because } P(X<4)+P(X\geq 4)=1\text{)}\\ &= 1 - P(X \leq 3) \qquad \text{(because } P(X<4)=P(X\geq 3)\text{)}\\ &= 1 - \text{binomcdf}(6, 0.25, 3) = 1 - 0.96240234375 = 0.03759765625\end{aligned}$$

So the probability of at least 4 successful trials is approximately 3.76%. □

Activity 1: Binomial-Distribution-Probability

For each part, please label the n, X, and p in addition to your work and answer. Leave answers as decimals and round to three decimal places.

Find the probability that in tossing a fair coin three times, there will appear

(a) three heads

(b) two tails and a head

(c) at least one head

(d) not more than one tail

Activity 2: Binomial-Distribution-Probability

For each part, please label the n, X, and p in addition to your work and answer. Give answers as percentages and round to one decimal place.

Find the probability that in five tosses of a fair die, a 3 will appear

(a) twice

(b) at most once

(c) at least two times

Activity 3: Probability

A manufacturer of metal pistons finds that on the average, 12% of his pistons are rejected because they are either oversized or undersized. What is the probability that a batch of 10 pistons will contain

(a) no more than 2 rejects?

(b) at least 2 rejects?

Visualizing a Binomial Distribution

Activity 4: Visualizing-Binomial-Distribution

Let X be the number of heads that turn up after flipping a coin five times. Then $n = 5$ and x can be $0, 1, 2, 3, 4, 5$. We can calculate the probability of zero heads turning up with $P(X = 0)$, one head turning up with $P(X = 1)$, etc. Using our calculator check that:

$$P(X=0) = \frac{1}{32}, P(X=1) = \frac{5}{32}, P(X=2) = \frac{10}{32}, P(X=3) = \frac{10}{32}, P(X=4) = \frac{5}{32}, P(X=5) = \frac{1}{32}$$

(a) Plot a histogram for the random variable X probability distribution.

To do this, on the horizontal axis scale from $x = 0, 1, 2, 3, 4, 5$ and the vertical axis scale from 0 to $10/32$ with $1/32$ intervals.

(b) Describe any interesting features from your histogram.

(c) Sketch what you think the histogram would look like for the same random variable X but the number of trails is $n = 100$.

Definition 3: Binompdf-Histogram

USING CALCULATOR TI83:

- $\boxed{DIST}$ key in yellow ($\boxed{2nd}$ > $\boxed{VARS}$) > Scroll to 10 "binompdf" > Type binompdf(n,p) (NOTICE WITHOUT THE x!!) > $\boxed{STO->}$ > $\boxed{L2}$ (this saves to List L2). Create x=0,1,. . . ,n into L1.
- $\boxed{StatPlot}$ ($\boxed{2nd}$ $\boxed{Y=}$) > Turn "On" Plot 1 > Select Bar graph image and set "Xlist:L1" and "Freq:L2" > $\boxed{Graph}$. (Change viewing window manually in $\boxed{Window}$)

Activity 5: Visualizing-Binomial-Distribution

Let our experiment be shooting free-throws. Assume that the probability of making a freethrow is 70% and that these are independent events. Let X be the number of made in taking six shots.

(a) Use your calculator to find $P(X = x)$ for $x = 0, 1, 2, 3, 4, 5, 6$.

(b) Plot a bar graph for the random variable X probability distribution.

(c) Describe any interesting features from your histogram. Is it positively skewed, negatively skewed, or symmetric?

Activity 6: Probability

Hospital records show that of the patients suffering from a certain disease, 75% die of it.

(a) What is the probability that of 6 randomly selected patients, 4 will recover? Give your answer as a percent rounded to the nearest tenth.

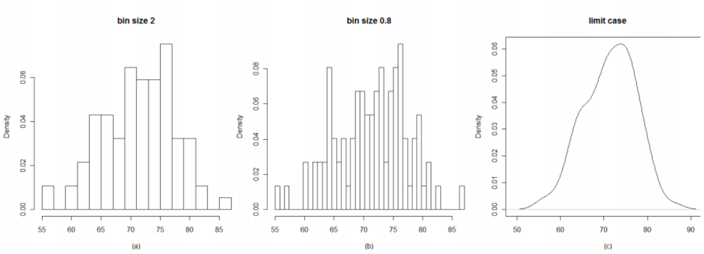
(b) Plot the probability distribution function.

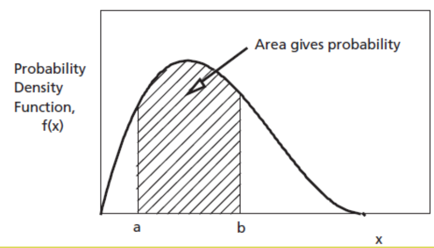
(c) Is the histogram positively skewed, negatively skewed, or symmetric?

Probability Density Functions

Let's consider the random variable X to be the temperature in a room. We want to know the probability that the temperature is in any given interval. For example, what's the probability for the temperature between 70° and 80°?

Ultimately, we want to know the probability distribution for X. One way to do that is to record the temperature from time to time and then plot the histogram. However, when you plot the histogram, it's up to you to choose the bin size.
But if we make the bin size finer and finer (meanwhile we need more and more data), the histogram will become a smooth curve which will represent the probability distribution for X.





Definition 4: Probability Density Functions

Recall a random variable can be either discrete or continuous. If we want to know the probability of a random variable in certain range, then we

- **Probability Density Functions:** Let X be a continuous random variable. Then a **probability density (or probability distribution) function (pdf) of** X is a function $f(x)$ provided that:
  1. f is a continuous function on $\mathbb{R}$
  2. f is nonnegative, that is, $f(x) \geq 0$
  3. The probability that X takes on a value in the interval $[a, b]$ is the area above this interval and under the graph of the density function:

     $$\boxed{P(a \leq X \leq b) = \text{Area under the curve between } x = a \text{ and } x = b} \tag{3}$$

     The graph of $f(x)$ is often referred to as the density curve.
  4. The total area under the graph of $f(x)$ is 1. This corresponds to $P(S) = 1$.

Example 3: Probability Density Function

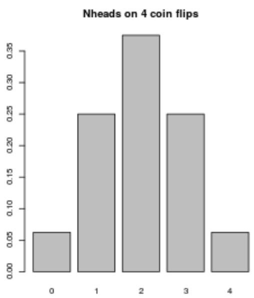
(a) NOTE: for continuous pdf, $P(X = x) = 0$ (why?)

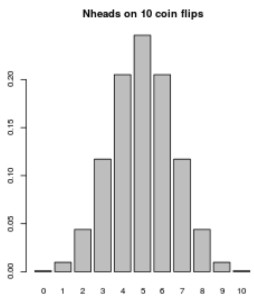
(b) $X =$ temperature in the room. What is $P(X = 70)$, that is, the temperature in the room is *exactly* 70°? Well, logically, $P(X = 70) = 0$ right? Notice it is NOT the y-coordinate of the point $(70, 0.056)$. The key to using a PDF is that you must know a RANGE of values for X, not insist on exact values. This is because we are dealing with infinite possibilties unlike the discrete case.

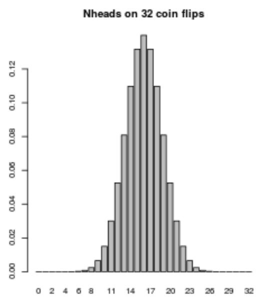
Normal Distribution

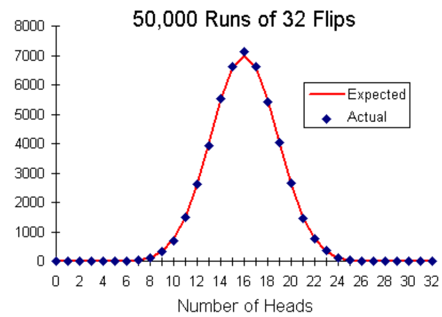
The normal distribution is the most widely known and used of all distributions. Because the normal distribution approximates many natural phenomena so well, it has developed into a standard of reference for many probability problems.

If you recall Activity 4, I asked you to visualize the histogram of the probability of flipping a coin 100 times. What if we did it for 1,000,000 times? What is we imagine $n \to \infty$. We go from a discrete random variable X to a continuous random variable X in the limit. That means we can talk about it's PDF $f(x)$. What does the graph of $f(x)$ look like for flipping a coin infinitely many times?







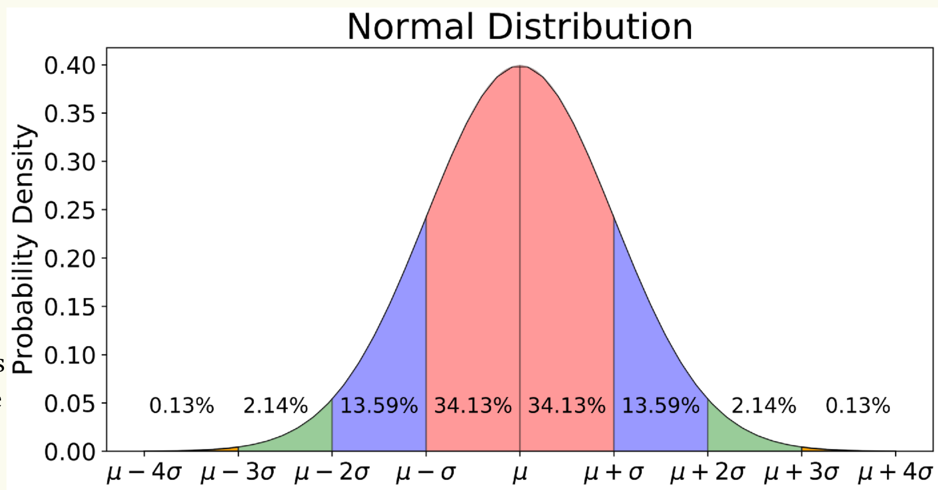


Definition 5: Normal-Distribution

- **Normal Distribution:** Is the PDF for a very special function, and it's graph is given by:

  Properties:

  - Symmetric, bell shaped
  - Continuous on all of $\mathbb{R}$
  - Mean: $\bar{x} = \mu$
  - Standard deviation: σ
  - The associated PDF $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$
  - Two parameters, μ and σ. Note that the normal distribution is actually a family of distributions since μ and σ determine the shape of the distribution.



Definition 6: Normal Distributions in Ti83

- Calculating probabilities outside of empirical rule: use table or calculator!

  USING CALCULATOR TI83: normalcdf(min, max, μ, σ)

  $\boxed{DIST}$ key in yellow ($\boxed{2nd}$ > $\boxed{VARS}$) > Scroll to 2. "normalcdf"

Definition 7: Empirical Rule

- **Empirical Rule (68-95-99.7 Rule):** is a shorthand used to remember the percentage of values that lie within a band around the mean in a normal distribution with a width of two, four and six standard deviations, respectively. More accurately, 68.27%, 95.45% and 99.73% of the values lie within one, two and three standard deviations of the mean, respectively.

  In mathematical notation, these facts can be expressed as follows, where X is an observation from a normally distributed random variable, $\mu (= \bar{x})$ is the mean of the distribution, and σ is its standard deviation:

$$\text{Within 1 standard deviation of mean: } P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 68\%$$
$$\text{Within 2 standard deviations of mean: } P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95\%$$
$$\text{Within 3 standard deviations of mean: } P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 99.7\%$$

Why is the standard normal distribution useful?

- Many things actually are normally distributed, or very close to it. For example, height and intelligence are approximately normally distributed; measurement errors also often have a normal distribution
- The normal distribution is easy to work with mathematically. In many practical cases, the methods developed using normal theory work quite well even when the distribution is not normal.
- There is a very strong connection between the size of a sample N and the extent to which a sampling distribution approaches the normal form. Many sampling distributions based on large N can be approximated by the normal distribution even though the population distribution itself is definitely not normal.

Activity 7: Normal-Distribution

Weight (in grams) of bags of sugar from a factory are normally distributed, with a mean of 1000g, and standard deviation of 13g. Find the following.

(a) The probability that a randomly selected bad of sugar weighs in between 974g and 1000g. Do this with a calculator.

(b) The percentage of bags whose weight is above 1026g. Do this with a calculator.

Activity 8: Normal-Distribution

The time it takes employees to get to work from home (in minutes) is normally distributed with a mean of 30 minutes, and a standard deviation of 5 minutes. Find the following.

(a) The percentage of employees that take between 20 and 40 minutes to get to work. Do this without a calculator.

(b) The percentage of employees that take between 28 and 37 minutes to get to work. Do this with a calculator.

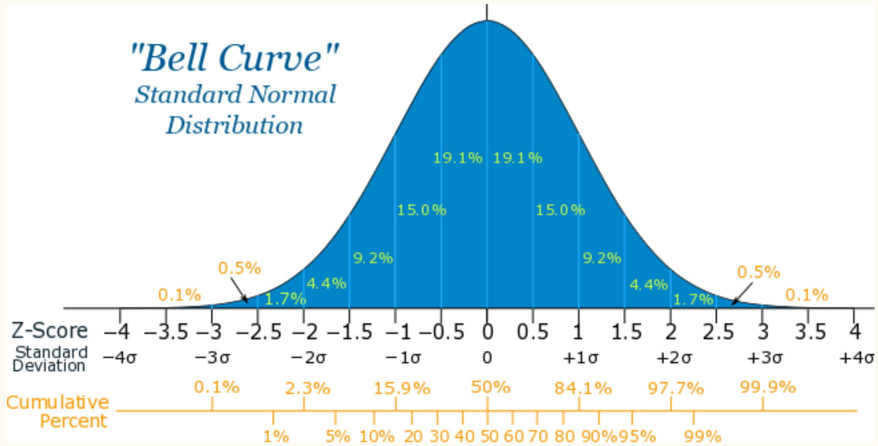
Standard Normal Distribution

Definition 8: Standard-Normal-Distribution

- **Standard Normal Distribution:** Is the PDF for a very special function, and it's graph is given by:

  Properties:

  - Symmetric, bell shaped
  - Continuous on all of $\mathbb{R}$
  - Mean: $\bar{x} = \mu = 0$
  - Standard deviation: $\sigma = 1$
  - The associated PDF $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$



- Calculating probabilities outside of empirical rule: use table or calculator!

  USING CALCULATOR TI83: normalcdf(min, max) – NOTICE: when using Standard Normal we do not need to put μ and σ since the calculator already has is programed for $\mu = 0$ and $\sigma = 1$

  $\boxed{DIST}$ key in yellow ($\boxed{2nd}$ > $\boxed{VARS}$) > Scroll to 2. "normalcdf"

Definition 9: Using-Normal-Distribution-Z-Scores

- General Procedure: As you might suspect from the formula for the normal density function, it would be difficult and tedious to do the calculus every time we had a new set of parameters for μ and σ. So instead, we usually work with the standardized normal distribution, where $\mu = 0$ and $\sigma = 1$. That is, rather than directly solve a problem involving a normally distributed variable X with mean μ and standard deviation σ, an indirect approach is used.

- We first convert the problem into an equivalent one dealing with a normal variable measured in standardized deviation units, called a **standardized normal variable**. To do this,

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

- A table of standardized normal values (Appendix C) can then be used to obtain an answer in terms of the converted problem.

- If necessary, we can then convert back to the original units of measurement. To do this, simply note that, if we take the formula for z, multiply both sides by σ, and then add μ to both sides, we get

$$\boxed{x = z \cdot \sigma + \mu}$$

- The interpretation of z values is straightforward. Since $\sigma = 1$, if $z = 2$, the corresponding x value is exactly 2 standard deviations above the mean. If $z = -1$, the corresponding x value is one standard deviation below the mean. If $z = 0$, x = the mean, i.e. μ.

KEY: suffices to know the standard normal since we can go back and forth between normal and standard normal using the formulas

Activity 9: Convert-z-values

Convert each of the following between x and z values.

(a) $x = 35$ where $\mu = 40$, $\sigma = 2$

(b) $x = 130$ where $\mu = 100$, $\sigma = 12$

(c) $z = -0.57$ where $\mu = 14$, $\sigma = 1.5$

Activity 10: Standard-Normal-Distribution

Find the area under the standard normal curve between

(a) $z = 0$ and $z = 1.2$

(b) $z = -0.68$ and $z = 0$

(c) $z = -0.46$ and $z = 2.21$

(d) to the right of $z = -1.28$

Activity 11: Standard-Normal-Distribution

The mean weight of 500 male students at a certain college is 151 lb and the standard deviation is 15 lb. Assuming that the weights are normally distributed, find without using a calculator how many students weigh

(a) between 120 and 155 lb

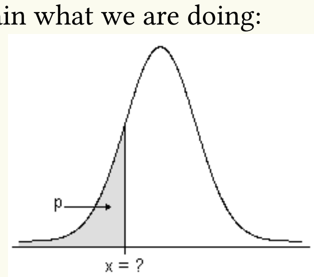
(b) more than 185 lb.

Inverse Normal Distribution

Definition 10: Inverse-Normal-Distribution

This is an informal name to the process of working backwards from a known (or given) probability to find an x-value.

Given probability p of the area under the curve of the PDF to the left of the value x_p, we want to find what the value x_p is. It really helps to draw a picture to help explain what we are doing:



Essentially, we solving for x in the equation: $P(X \leq x) = p$.

USING CALCULATOR TI83: invNorm(p,μ,σ)

$\boxed{DIST}$ key in yellow ($\boxed{2nd}$ > $\boxed{VARS}$) > Scroll to 3. "invNorm("

- Given probability p of the area under the curve of the PDF to the left of the value x_p, we can find what the value x_p is by using the "inverse normal distribution:"

$$\text{invNorm}(p, \mu, \sigma) = x_p \tag{4}$$

Activity 12: Inverse-Normal-Distribution

Find the 90th percentile for a normal distribution with a mean of 70 and a standard deviation of 4.5.

Activity 13: Inverse-Normal-Distribution

The time it takes employees to get to work from home (in minutes) is normally distributed with a mean of 30 minutes, and a standard deviation of 5 minutes. Find:

(a) the percentage of employees that take between 28 and 37 minutes to get to work (Hint: this is not an inverse problem)

(b) The number of minutes the longest it would take the bottom employee in the bottom 5% of the data to get to work. (Hint: this is an inverse problem)

Activity 14: Inverse-Normal-Distribution

An average light bulb manufactured in a factory lasts 280 days with a standard deviation of 45 days. Assume that bulb life is normally distributed.

(a) What is the probability that an Acme light bulb will last at most 360 days? (Hint: this is not an inverse problem)

(b) What bulb life separates the bottom 12%? (Hint: this is an inverse problem)