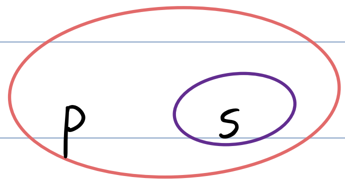
Lesson 1 The big picture

1.2 Probability and Random Sample



Randomly select individuals to form a sample, to make inference about the characteristics of a population!

1.3 Sample space

Sample space or outcome space is the collection of all possible outcomes of a study

1.4 Types of Data

Data
- Numeric
  - Discrete
  - Continuous
- Categorical

1.5 Histogram and density plot

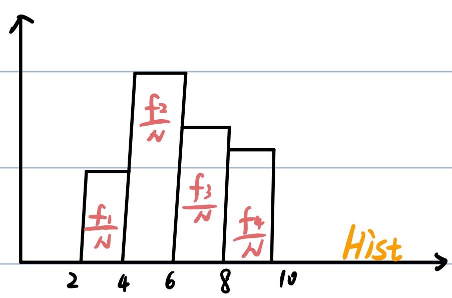
How to make a histogram and a density plot

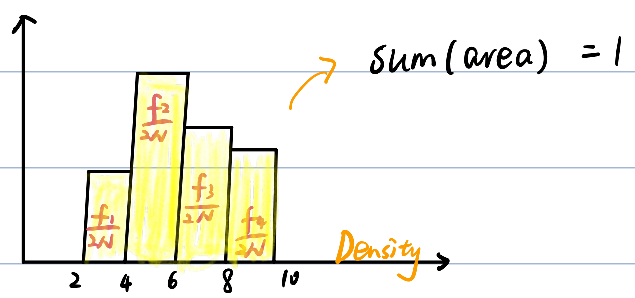
- denote f as frequency; N as total number of occurrences

1) calculate $\frac{f_i}{N}$ for a histogram

   Calculate $\frac{f_i}{N\,(\text{width of a bar})}$ for a density plot

2) draw the plot





Lesson 2 Properties of probability

2.1 Why probability?

understand how likely an event happens

2.2 Event

Event is a subset of sample space (recall - sample space or outcome space is all possible outcomes of a random experiment)

- Review of set theory

1) ϕ is null set

2) $A \cup B$: union

3) $A \cap B$: intersection

4) $A' = A^c$: complement , the elements not in A

2.5 what is probability?

Formally speaking, probability is a set function , P, that assigns a real number to an event

P(A=event)=a —> real number

- Axioms of probability

1) P(A) >0

2) P(S) = 1

3) if A1, A2, ..., An are mutually exclusive, then $P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A1)+P(A2)+....+P(An)$

2.6 Five Theorems

1) $P(A) = 1 - P(A')$

2) $P(\phi) = 0$

3) $P(A) \le 1$

4) if $A \subseteq B$, $P(A) \le P(B)$

5) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Lesson 3 Counting Techniques

3.1 The multiplication principle

If there are

n_1 outcomes of a random experiment for E_1,

n_2 outcomes of a random experiment for E_2

......

n_k outcomes of a random experiment for E_k,

then, there are $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ outcomes of the composite experiment, $E_1E_2\ldots E_m$

*To determine n_i, one should pay attention that if the experiment allows replication.

3.2 Permutation

(A generalization of multiplication rule)

If there are n positions we want to fill with n objects, then you have

n choices for the 1st position,

$n-1$ choices for the 2nd position,

$n-2$ choices for the 3rd position,

....

1 choice for the last position

then, the total outcomes are $n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1 = n!$

This is denoted as $_nP_n = n!$

Similarly, if there are r positions and n objects, then you will have

n choices for the 1st position,

n-1 choices for the 2nd position,

n-2 choices for the 3rd position,

....

n-r+1 choices for the rth position,

then, the total number of choices are n•(n-1)•(n-2)•...•(n-r+1), which can be written as

$$nPr = \frac{n!}{(n-r)!}$$

3.3 Combination

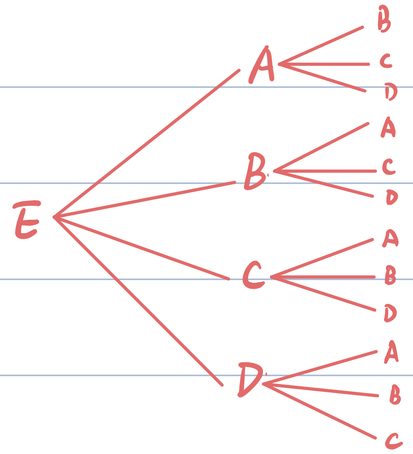
Unlike permutation cares about who is the first, the second, ..., the last, combination does not consider order when choosing objects to fill in each position. Then, combination has fewer total outcomes than permutation.

$$nCr = C^n_r = \binom{n}{r} = \frac{n!}{r!\,(n-r)!} = \frac{{}_nP_r}{r!}$$

An example to show the relation

Suppose we have 5 objects and 3 positions,

if we consider order when filling the positions, it would have 12•5=60 outcomes , ${}_5P_3 = \frac{5!}{(5-3)!} = 5\times4\times3$



permutation : EAB is different from EBA

Combination : EAB is the same as EBA , AEB, ABE, BAE , BEA

If it is combination it has (60/6)=10 outcomes , $C^5_3 = \frac{5\times4\times3}{3\times2} = \frac{5!}{(5-3)!\cdot3!} = \frac{n!}{r!\,(n-r)!}$

Lesson 4 Conditional Probability

The probability that A happens given B has already happened.

4.2 Conditional Probability

- The conditional probability of an event A given B has already happened is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{given } P(B) > 0$$

- Properties of conditional probability

1) $P(A|B) \geq 0$

2) $P(B|B) = 1$

3) If $A_1, A_2, \cdots, A_n$ are mutually independent, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n | B) = P(A_1|B) + P(A_2|B) + \cdots + P(A_n|B)$$

4.3 Multiplication Rule

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

$$P(A \cap B \cap C) = P(C|A \cap B) \cdot P(A \cap B) = P(C|A \cap B) \cdot P(A|B) \cdot P(B)$$

Lesson 5 Independent Events

5.1 Independent events

Two events are independent if one happens does not affect the probability of the occurrence of the other, which means $P(A|B)=P(A)$ and $P(B|A)=P(B)$.

Then, we have

$$P(A|B)=P(A)=\frac{P(A\cap B)}{P(B)} \Rightarrow P(A\cap B)=P(A)\cdot P(B)$$

So, if $P(A\cap B)=P(A)\cdot P(B)$, then A and B are independent, O.W. they are dependent.

5.3 Three Theorems

1) If A and B are independent, then A and B' are also independent.

$$P(B'|A)=\frac{P(B'\cap A)}{P(A)}$$

$$\begin{aligned}
\Rightarrow P(B'\cap A) &= P(B'|A)\cdot P(A)\\
&= P(A)\cdot(1-P(B|A))\\
&= P(A)-P(A)\cdot P(B|A)\\
&= P(A)-P(A)\cdot P(B) \quad \text{A, B are independent}\\
&= P(A)\cdot(1-P(B))\\
&= P(A)\cdot P(B')
\end{aligned}$$

2) If A and B are independent, then A' and B are independent, too.

3) If A and B are independent, then A' and B' are independent.

$$\begin{aligned}
P(A'\cap B') &= P(A'|B')\cdot P(B')\\
&= (1-P(A|B'))\cdot(1-P(B))\\
&\qquad \text{Theorem 1)}\\
&= (1-P(A))\cdot(1-P(B))\\
&= P(A')\cdot P(B')
\end{aligned}$$

5.3 Mutually independent events

Three events are mutually independent if and only if the following conditions hold:

1) the events are pairwise independent. That is

$$P(A \cap B) = P(A) \cdot P(B) \quad \& \quad P(B \cap C) = P(B) \cdot P(C) \quad \& \quad P(A \cap C) = P(A) \cdot P(C)$$

2) $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$

*pairwise independence does not ensure mutual independence

Lesson 6 Bayes' Theorem

When you have the conditional probability of $P(A|B)$ and you want $P(B|A)$, then you need Bayes' theorem .

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A)}$$

Where $P(A)$ and $P(B)$ are called prior probability, $P(B|A)$ is called posterior probability.

Prior probability is what we know about the event A, B before we obtain any additional information.

Posterior probability is the a conditional probability after we know additional information about the relations among events.

6.2 A generalization

Given events B1, B2, B3,...,Bm are mutually exclusive and exhaustive, which means

$$B_i \cap B_j = \phi \quad (i \neq j) \quad \& \quad B_1 \cup B_2 \cup \cdots \cup B_m = S \quad \text{(sample space)} \quad \& \quad P(B_i) > 0$$

Now, if $A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_m)$

then $P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_m) = \sum_{i=1}^{m} P(A \cap B_i)$ } Law of Total probability

$$= P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + \cdots + P(A|B_m) \cdot P(B_m)$$

$$= \sum_{i=1}^{m} P(A|B_i) \cdot P(B_i)$$

so, as long as $P(A)$ is greater than 0, the posterior probability of $P(B_k|A)$ is

$$P(B_k|A) = \frac{P(B_k \cap A)}{P(A)}$$

$$= \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i=1}^{m} P(A|B_i) \cdot P(B_i)}$$

Lesson 7 Discrete Random Variables

Random Variable X

Given a random experiment with a sample space S, a random variable is a set function to assign one and the only one real number to each element in the sample space.

e.g. randomly select 5 puppies from a puppy party group (10), X is the number of puppies which is less than 10 lbs. X could be 0,1,2,3,4,5.

X is the random variable, {0,1,2,3,4,5} are all the possible values of X, which is called support or space.

7.2 Discrete Random Variable

A random variable X is a discrete random variable if:

- the space/support is finite
- the possible values of X is countable , even though it is infinite

7.2 Probability Mass Function

The probability that a random variable X takes on x is denoted as $P(X=x)$ or $f(x)$. It is called probability mass function (p.m.f).

The probability mass function satisfies:

- $P(X=x)=f(x)>0$, if x is in the support S (if x does not belong to S, then $P(X=x) = 0$)
- $\sum_{x\in S} f(x)=1$
- $P(X \in A)= \sum_{x\in A} f(x)$

7.3 The cumulative discrete function (CDF)

The CDF of a random variable is $F_X(t)=P(X<=t)$.

A CDF has the following properties:

1) $F_X(t)$ is non-decreasing

2) $F_X(t)$ is between 0 and 1

3) $F_X(min) = P(X<=min)= P(X=min)$

4) $F_X(max) = P(X<= max) = P(X<=max)=1$

5) $P(X>t) = 1-P(X<=t) = 1-F_X(t)$

7.4 Hypergeometric distribution

Supports there are two groups of objects in the set N, and m out of N belong to the first group, and N-m belong to the second group. We want to select n in total from the N objects without replacement, and x out of n are from Group 1, and n-x are from Group 2.

Then, the p.m.f of X is called the hypergeometric distribution with a function

$$P(X=x) = f(x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$

where $x<=n$, $x<=m$, $(n-x)<=(N-m)$

Lesson 8 Mathematical Expectation

Mean is the expected value of a random variable X, denoted E(X) or u. It is a weighted average, an average of the values weighted by their respective individual probabilities.

8.1 A definition

If f(x) is a p.m.f of a random variable X with a support S, and the the summation exists $\sum_{x \in S} X \cdot f(X)$, then, it is called the mathematical expectation.

$$E[X] = \sum_{x \in S} X \cdot f(x)$$

A generalization: $E[u(x)] = \sum_{x \in S} u(x) \cdot f(x)$, where $u(X)$ is a function of X.

8.2 Properties of expectation

- E[c] = c (c is a constant)
- E[cx] = cE[x] (c is a constant, X is a random variable)
- E[cx+dY] = cE[X]+dE[Y] (c,d are constant, X,Y are random variables)
- E[XY]=E[X]E[Y] if X and Y are independent

8.3 Mean of X

Taking the generalization form of expectation, the mean of X is simply

$$E[u(x)] = E[x] = \sum_{x \in S} x \cdot f(x)$$

This is also called the first moment of X (about the origin) @not sure what the origin mean 😢

8.4 Variance of X

Taking the generalization form of expectation, the variance is when $u(x) = (x-u)^2$

$$E[u(x)] = E[(x-u)^2] = \sum_{x \in S} (x-u)^2 f(x) = \sigma^2 = Var(X)$$

The variance of X can be also called the second moment of X about the mean. The squared root of variance, σ , is called the standard deviation of X.

- Theorem

$$\sigma^2 = E[(x-\mu)^2] = E[x^2 - 2\mu x + \mu^2]$$
$$= E[x^2] - 2\mu \cdot E[x] + E[\mu^2]$$
$$= E[x^2] - 2\mu \cdot \mu + \mu^2$$
$$= E[x^2] - \mu^2$$

- Theorem

Given random variable X has mean and variance μ_x, σ_x^2, then the mean and variance of the random variable $Y = ax + b$ are

$$E[Y] = E[ax+b] = E[ax] + E[b] = a \cdot E[x] + b = a \cdot \mu_x + b = \mu_Y$$

$$Var(Y) = E[Y^2] - \mu_Y^2$$
$$= E[(ax+b)^2] - (a \cdot \mu_x + b)^2$$
$$= E[a^2x^2 + 2abx + b^2] - (a^2 \mu_x^2 + 2ab\mu_x + b^2)$$
$$= a^2 \cdot E[x^2] + \cancel{2abE[x]} + \cancel{b^2} - a^2 \cdot \mu_x^2 - \cancel{2ab\mu_x} - \cancel{b^2}$$
$$= a^2 \underbrace{(E[x^2] - \mu_x^2)}_{\sigma_x^2}$$
$$= a^2 \sigma_x^2$$

8.5 Sample Means and Variances

Sample mean is the average of a random sample from the population, denoted $\bar{x}$.

$$\bar{x} = \frac{1}{n} \cdot (x_1 + x_2 + \cdots + x_n) = \sum_{i=1}^{n} \frac{1}{n} \cdot x_i$$

The sample mean summarizes the "location" or "center" of the data.

Sample variance measures the spread of the sample data from the population, denoted s.

$$s = \frac{1}{n-1} \cdot \left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\right]$$
$$= \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \cdot \left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right]$$

Lesson 9 Moment Generating Function

9.1 Moment generating function

Let X be a discrete random variable with a p.m.f, f(x), and a support S. Then the moment generating function of X is

$$M(t) = E[e^{tx}] = \sum_{x \in S} e^{tx} \cdot f(x)$$

as long as the summation is finite for some interval of t around 0.

Personally, I use MGF very little. I will not continue the review of this section. Penn State has more open source about MGF on its website.

Lesson 10 The Binomial Distribution

10.1 The probability mass function

A binomial random variable X follows a binomial distribution denoted $X \sim b(n,p)$ with a P.m.f.

$$f(x) = \binom{n}{x} \cdot p^{x} \cdot (1-p)^{n-x} , \quad E[x] = np , \quad Var(x) = np(1-p)$$

10.2 Is X Binomial?

Binomial random variable satisfies the following conditions:

- An experiment or trials is repeated exactly the same way
- Each of the n trials has only two outcomes. One is called "success ", the other is called "failure ".

  Each single trial is a Bernoulli trial.
- The n trails are independent
- The probability of a success is p, and that of a failure is 1-p
- The random variable X, is the number of success in the n trials

10.3 Cumulative binomial probabilities

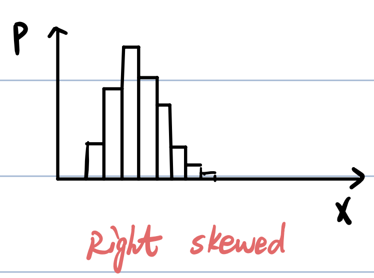
Cumulative probability function is $F(x) = P(X \le x)$. For discrete random variable, it is simply a summation:

$$F(x) = \sum_{m=0}^{x} f(m) = P(X=0) + P(X=1) + \cdots + P(X=m)$$

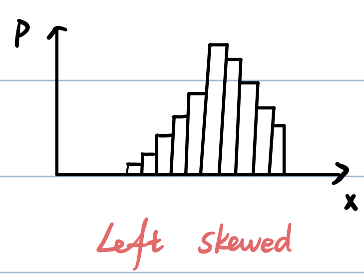
10.4 Effect of n and p on Shape

n and p in binomial distribution are called parameter. The size of n and p affect the shape of the distribution.

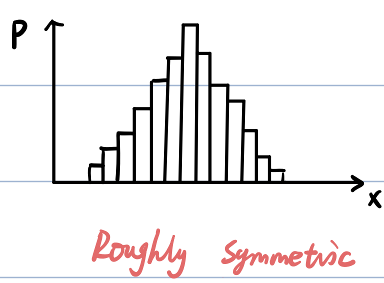
- Small p and small n.



- large p and small n.



- p=0.5



Lesson 11 Geometric and Negative Binomial Distribution

11.1 Geometric Distribution (Bernoulli until the first success)

Geometric distribution is consisted of a series of Bernoulli trials. There are two outcomes:

1. There two possible outcomes

2. The trials are independent

3. The probability of a success, p, is the same from trial to trial.

Let X denote the number of trials until the first success. Then. The probability function is

$$f(x) = P(X=x) = (1-p)^{x-1} \cdot p$$

11.2 Key properties of a Geometric Random Variable

Theorem 1 $F(x) = P(X <= x) = 1-(1-p)^x$

Theorem 2 $E[x] = 1/p$

Theorem 3 $Var(x) = (1-p)/p^2$

11.4 Negative binomial distribution (Bernoulli until the rth success)

Negative binomial distribution is a series of Bernoulli trials. That is

1. There are two possible outcomes

2. Each trial is independent from the other

3. The probability of success, p, is the same for all the trials

Let X be the total number of trails until the rth success. The p.m.f of a Negative binomial distribution is

$$f(x) = P(X=x) = \binom{x-1}{r-1} \cdot (1-p)^{x-r} \cdot p^{r-1} \cdot p$$

$\binom{x-1}{r-1}$ means before reaching the rth success, in the previous x-1 trials, you have r-1 success out of them. It is like choosing (r-1) success from the (x-1) trials.

11.5 Key properties of negative binomial distribution

Theorem 1 $E[X]=r/p$

Theorem 2 $Var[X] = r(1-p)^2/p$

Lesson 12 The Poisson Distribution

12.1 Poisson Distribution

Let the discrete random variable X denote the number of times an event occurs in an interval of a time, or space. Then X may be a Poisson random variable with $x = 0,1,2,3,\dots$

Then x follows an approximate Poisson process with parameter $\lambda > 0$ if:

1. The number of events occurring in non-overlapping intervals are independent
2. The probability of exactly one event in a short interval of length is approximately $\frac{\lambda}{n}$
3. The probability of two or more events happening in a short interval is zero.

With these conditions, the p.m.f of a Poisson Distribution is

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

where λ is the mean and variance of X.

12.3 Poisson Properties

Theorem 1 Moment Generating Function $M(t) = e^{\lambda(e^{t}-1)}$ for $-\infty < t < \infty$

Theorem 2 $E[x] = \lambda$

Theorem 3 $var(x) = \lambda$

Theorem 4 Poisson distribution could be an approximate of Binomial distribution when n is large enough. Then, $p = \frac{\lambda}{n}$

Note, Poisson approximation to the binomial distribution works well only when n is large and p is small. In general, the approximation works well if $n >= 20$ and $p <= 0.05$ or if $n >= 100$ and $p <= 0.1$.

Lesson 13 Exploring Continuous Data

13.1 Histogram

1. Get the total number of objects, n, in the sample
2. Define k class intervals $(c_0, c_1]$, $(c_2, c_3]$...$(c_{k-1}, c_k]$
3. Count the objects to get the frequency, f_i, of each class i
4. Calculate the relative frequency (proportion) of each class by dividing the class frequency by the total number of objects, n, $\frac{f_i}{n}$

5. For a frequency histogram, draw rectangle for each class with the class interval as the base, and the height equal to the frequency of the class

6. For a relative frequency histogram, draw a rectangle for each class with the class interval as the base, and the height equal to the relative frequency of the class

7. For a density histogram, draw a rectangle for each class with the class interval as the base, and the height equal to $\frac{f_i}{n(C_i - C_{i-1})}$

13.3 Order Statistics and Sample Percentiles

If we have a sample of n observations represented as $X_1, X_2, X_3, \ldots X_n$, then, when the observations are ordered from the smallest to the largest, the resulting ordered data are called the order statistics of the sample, and represented as $y_1 <= y_2 <= y_3 <= \ldots <= y_n$. That is, y_1 is the smallest data point and the first order statistic. The second smallest data point is y_2, the second order statistic. From the order statistics, it is easy to find the sample percentile.

- Definition of sample percentile

if $0<p<1$, then the $(100p)$th sample percentile has approximately np sample observations less than it, and the rest $n(1-p)$ are greater than it.

Some sample percentiles have special names:

- Q1: first quartile, the 25th percentile
- Q2: second quartile or median, 50th percentile
- Q3: third quartile, 75th percentile
- IQR=Q3-Q1 (IQR: interquartile range)

Typical method to find a particular sample percentile:

1. If (n+1)p is an integer, then the (100p)th sample percentile is the (n+1)pth order statistic
2. If (n+1)p is not an integer, but rather equals r plus some proper fraction, e.g. (n+1)p =r+a/b.

   Then, the (100p)th sample percentile is a value between the rth and (r+1)th order statistic. The exact form is $y_r + \left(\frac{a}{b}\right)\cdot\left(y_{r+1} - y_r\right)$

note, the definition of percentile is the sample value should be exactly less than and greater than the percentile, that is why (n+1) is used here

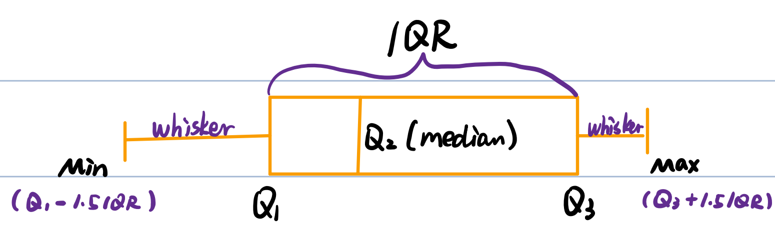
Example If n = 64, find its 25th percentile

n=64, p=0.25, (n+1)p =65•0.25=16.25=16+0.25

Then the 25th percentile is $y_{16} + 0.25\,(y_{17} - y_{16})$

13.4 Box Plots

Five-Number Summary: Q1, Q2, Q3, Min, Max. The summary statistics are the components of a box plot.

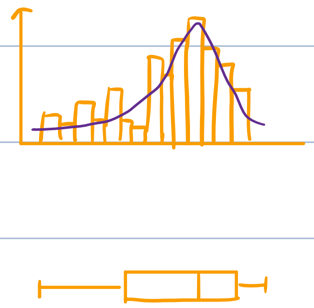


Different resources may define the whiskers differently. Some uses min and max directly, others may use lower limit (Q1-1.5IQR) and upper limit (Q3+1.5IQR). Points beyond the two limits are outliers.

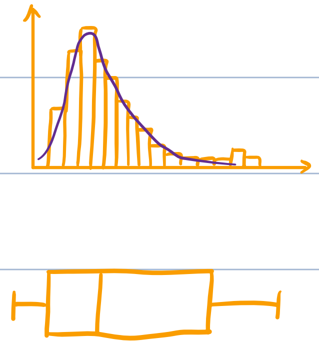
13.5 Shapes of Distributions

Histogram and box plots can be helpful in suggesting the shape of probability distributions.

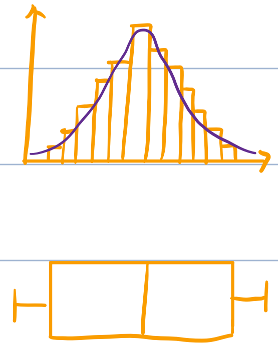
- Skewed Left.



- Skewed Right.



- Symmetric



Lesson 14 Continuous Random Variables

A continuous random variable is different from a discrete random variable because it takes an uncountably infinite number of possible numbers.

14.1 Probability Density Functions

- Probability density function (p.d.f)

the probability density function of a continuous random variable X with support S is an integrable function f(x) satisfying the following:

1. $f(x) > 0$ for all x in the support
2. $\int_S f(x)\,dx = 1$, The area under the pdf curve in the support S is 1
3. $P(X \in A) = \int_A f(x)\,dx$, where A is some interval in the support S

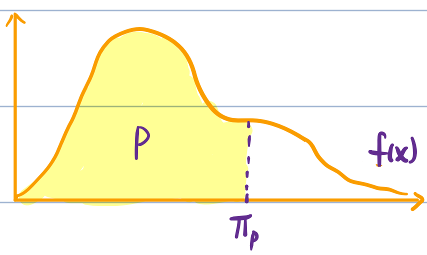
14.2 Cumulative Distribution Function (c.d.f)

The cumulative distribution function of a random variable X is

$$P(X \le x) = P(X < x) = \int_{-\infty}^{x} f(x)\,dx$$

14.3 Finding Percentiles

If X is a random variable, the (100p)th percentile is a number π_p such that the area under the curve of f(x) to the left of π_p is p.



$$P = \int_{-\infty}^{\pi_p} f(x)\,dx = P(X \le \pi_p) = F(\pi_p)$$

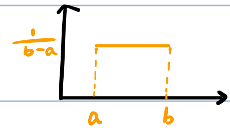
14.4 Special Expectations

- Mean $\mu = E[x] = \int x \cdot f(x) \cdot dx$
- variance $\sigma^2 = Var(x) = \int (x-\mu)^2 \cdot f(x)\,dx$
- Moment Generating Function $M(t) = \int e^{tx} \cdot f(x)\,dx$

14.6 Uniform Distributions

- Uniform Distribution

1. PDF : $X \sim U(a,b)$ $\quad f(x) = \frac{1}{b-a}$. for $a < x < b$,



2. CDF: $F(x) = P(X<x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{x} \frac{1}{b-a}\,dx = \frac{x-a}{b-a}$

14.7 Uniform Properties

- Mean $\quad \mu = E[x] = \frac{a+b}{2}$

$$= \int_a^b x \cdot f(x)\,dx = \int_a^b \frac{1}{b-a} \cdot x\,dx = \frac{1}{b-a} \cdot \frac{x^2}{2}\Big|_a^b = \frac{1}{b-a} \cdot \frac{b^2-a^2}{2}$$

$$= \frac{1}{b-a} \cdot \frac{(a+b)(b-a)}{2}$$

$$= \frac{a+b}{2}$$

- Variance

$$\sigma^2 = Var(x) = \frac{(b-a)^2}{12}$$

- Moment Generating Function

$$M(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

Lesson 15 Exponential, Gamma, and Chi-Square Distributions

15.1 Exponential Distribution

Recall that a Poisson random variable is the number of occurrences within a certain time interval or space. While, an exponential distribution is the usually used when is the first occurrence.

For example, suppose the mean number of customers to arrive at a bank in a 1-hour time interval is 10. If this is a Poisson distribution, then 10 is the mean or the parameter in the Poisson distribution, $\lambda = 10$.

While, on the other hand, if the study interest is the average time or the waiting time when the first customer comes in, it is an exponential distribution, with a parameter $\theta = \frac{1}{10}$ hr.

15.2 Exponential Properties

- p.d.f.

$$f(x) = \frac{1}{\theta} \cdot e^{-\frac{x}{\theta}}$$

- mean

$$\mu = E[x] = \theta$$

- variance

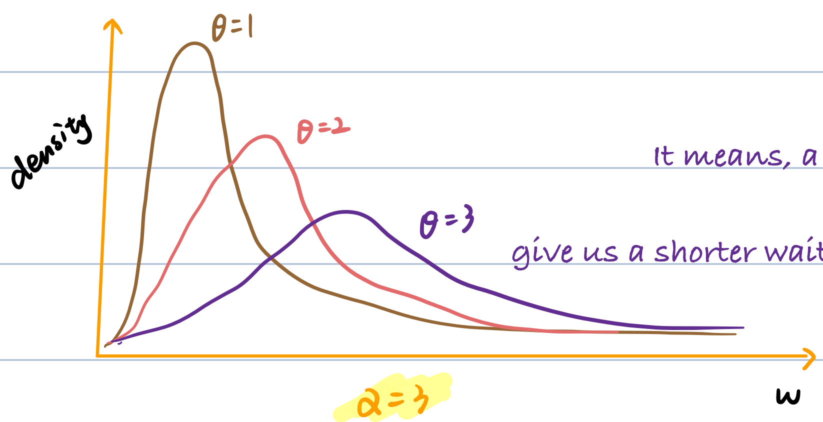
$$\sigma^2 = Var(x) = \theta^2$$

15.4 Gamma Distributions

In an Poisson process with mean lambda, the waiting time X until the first event occurs follows an exponential distribution, with mean $\theta = \frac{1}{\lambda}$. We now let W denote the waiting time until the αth occurrence. Then W follows a Gamma Distribution with a p.d.f.

$$f(w) = \frac{1}{(\alpha-1)!\,\theta^{\alpha}} \cdot e^{-\frac{w}{\theta}} \cdot w^{\alpha-1}$$

- Effect of θ and α on the distribution



It means, a shorter average waiting time, theta, tends to give us a shorter waiting time until the 3rd occurrence.

Similar situation for alpha when theta is fixed

15.5 Gamma Properties

- Gamma function

$$\Gamma(t) = \int_0^\infty y^{t-1} \cdot e^{-y}\, dy$$
$$= (t-1)\Gamma(t-1)$$

If t is an integer, then $\Gamma(t) = (t-1)!$

- Mean $\mu = E[X] = \alpha\theta$
- Variance $\sigma^2 = Var(X) = \alpha\theta^2$
- Moment Generating Function $M(t) = \dfrac{1}{(1-\theta t)^\alpha}$
- Exponential distribution is a special case of Gamma distribution when alpha=1

15.8 Chi-Squared Distribution

Chi-Squared distribution is a special case of Gamma distribution when theta=2 and alpha=r/2,

where r is a positive integer.

- P.d.f.

$$f(x) = \frac{1}{\Gamma(\frac{r}{2})2^{\frac{r}{2}}} \cdot x^{\frac{r}{2}-1} \cdot e^{-\frac{x}{2}} \quad , \quad x>0 \quad , \quad r>0$$

We say X follows a Chi-Squared distribution with a degree of freedom of r. Denoted $\chi^2(r)$

- Mean $\mu = E[X] = r$
- Variance $\sigma^2 = Var(X) = 2r$
- Moment Generating Function $M(t) = \dfrac{1}{(1-2t)^{\frac{r}{2}}}$
- Degree of Freedom

1. D.F. : the number of values that are free to vary as you estimate the parameters, N-P
2. It indicates how much independent information goes into a parameter estimation.

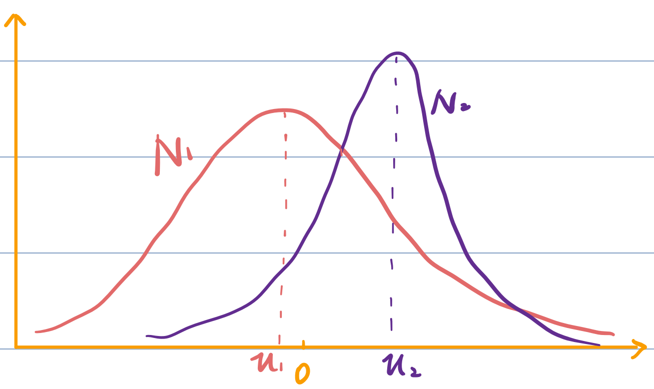
Lesson 16 Normal Distribution

16.1 The Distribution and Its Characteristics

- p.d.f.

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

- All normal curves are bell-shaped
- All normal curves are symmetric about the mean mu
- The area under an entire normal curve is 1
- $f(x) > 0$
- $\lim_{x \to \infty} f(x) = 0$
- The height of any normal curve is maximized at $x = \mu$
- The shape of any normal curves depends on its mean and standard deviation



$X_1 \sim N_1(\mu_1, \sigma_1^2)$

$X_2 \sim N_2(\mu_2, \sigma_2^2)$

$\mu_1 < \mu_2, \quad \sigma_1^2 < \sigma_2^2$

16.2 Finding Normal Probabilities

- Theorem

if $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X-\mu}{\sigma^2}$ follows $N(0,1)$, which is called standard normal distribution.

- c.d.f

$$P(X \le x) = F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

16.4 Normal Properties

- mean μ
- variance σ^2
- MGF $M(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$

16.5 The Standard Normal and The Chi-Squared Distributions

- Theorem

If X is normally distributed with mean and sigma, then $V = \left(\frac{X-u}{\sigma}\right)^2 = Z^2 \sim \chi^2(1)$

16.6 Some Application

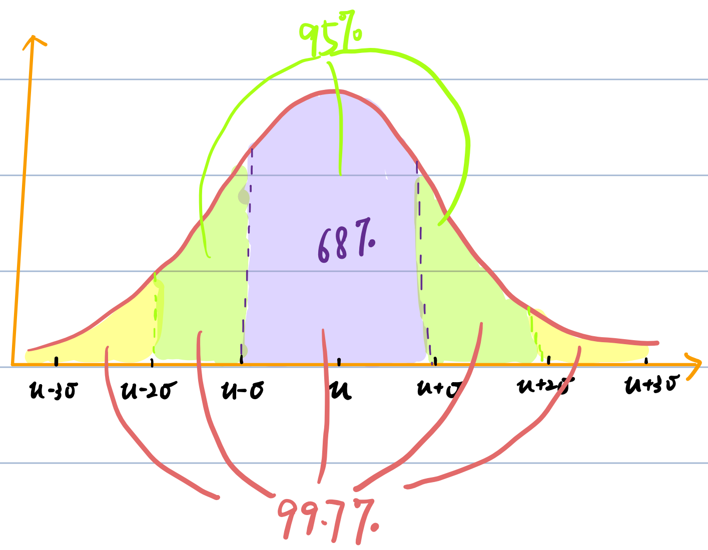
- Interpretation of Z

$$Z = \frac{X-u}{\sigma}$$

Z tells us how many standard deviations above or below the mean that X falls. E.g., if Z=2, it means X is 2 SD above the mean, mu. If Z=-2, it means X is 2 SD below the mean.

- The Empirical Rule

1. approximately 68% of data fall with in one SD from the mean
2. approximately 95% of data fall with in two SD from the mean
3. approximately 99.7% of data fall with in three SD from the mean



Lesson 17 Distribution of Two Discrete Random Variables

Suppose we toss two fair dices, what is the probability that the first dice is 6 and the second dice is 1?

This is a particular case of two discrete random variables, that $P(X=6, Y=1)=?$

- Joint Probability Mass Function

Let X and Y be two discrete random variables. Let S denote the two-dimentional support of X and Y.

Then, the function $f(x,y)=P(X=x, Y=y)$ is a joint probability mass function if

1. $0 \le f(x,y) \le 1$
2. $\sum\sum_{(x,y)\in S} f(x,y) = 1$
3. $P[(X,Y)\in A] = \sum\sum_{(x,y)\in A} f(x,y)$, where A is a subset of the support S.

- Marginal Probability Mass Function of X

Let X be a discrete random variable with support S_1, and let Y be a discrete random variable with support S_2,. Let X and Y have the joint probability mass function $f(x,y)$ with support S. Then, the probability mass function of X alone, which is called marginal probability mass function of X, is defined by

$$f_X(x) = \sum_y f(x,y) = P(X=x) \quad , \quad x \in S_1$$

where, for each x in the support S_1, the summation is taken over all possible values of y.

Similarly, the probability mass function of Y alone, which is called the marginal probability mass function of Y, is defined by:

$$f_Y(y) = \sum_x f(x,y) = P(Y=y) \quad , \quad y \in S_2$$

Where, for each y in the support S_2, the summation s taken over all possible values of x.

- Independent and Dependent Random Variables

The random variables X and Y are independent if $P(X=x, Y=y) = P(X=x)\cdot P(Y=y)$

- Expected Values

Let X be a discrete random variable with support S1, and let Y be a discrete random variable with support S2. Let X and Y be discrete random variables with joint p.m.f. $f(x,y)$ on the support S. If $u(X,Y)$ is a function of these two random variables, then

$$E[u(X,Y)] = \sum_{(x,y)\in S}\sum u(X,Y) f(X,Y)$$

if it exists, it is called expected values of $u(X,Y)$.

If $u(X,Y) = u(X)$, $u_X = E[X] = \sum_{x\in S_1}\sum_{y\in S_2} x \cdot f(x,y)$, if it exists, it is called the expected value of X.

If $u(X,Y) = u(Y)$, $u_Y = E[Y] = \sum_{x\in S_1}\sum_{y\in S_2} y \cdot f(x,y)$, if it exists, it is called the expected value of Y.

- Variance

similarly, $\sigma_X^2 = Var(X) = \sum_{x\in S_1}\sum_{y\in S_2} (x - u_X)^2 \cdot f(x,y)$

$\sigma_Y^2 = Var(Y) = \sum_{x\in S_1}\sum_{y\in S_2} (y - u_y)^2 \cdot f(x,y)$

Lesson 18 The Correlation Coefficient

18.1 Covariance of X and Y

Covariance is used to quantify the dependence between two random variables X and Y.

- Covariance

Let X and Y be random variables with means μ_X and μ_Y. The covariance of X and Y, denoted COV(X,Y) or σ_{XY}, is defined as:

$$COV(X,Y) = \sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)]$$

That is, if X and Y are discrete random variables with joint support S, then the covariance is

$$COV(X,Y) = \sigma_{XY} = \sum\sum_{(x,y)\in S} (x-\mu_X)(Y-\mu_Y)\cdot f(x,y)$$

If X and Y are continuous random variable with support S1 and S2, respectively, the covariance is

$$cov(X,Y) = \sigma_{XY} = \int_{S_1}\int_{S_2} (x-\mu_X)(y-\mu_Y)\cdot f(x,y)\, dx\, dy$$

- Theorem

For any random variables X and Y with means μ_X and μ_Y, the covariance of X and Y can be calculated as: $COV(X,Y) = E[XY] - \mu_X\mu_Y$

18.2 Correlation Coefficient of X and Y

Let X and Y be any two random variables with standard deviations σ_X and σ_Y, respectively. The correlation coefficient of X and Y, denoted Corr(X,Y) or ρ_{XY}

$$\rho_{XY} = Corr(X,Y) = \frac{COV(X,Y)}{\sigma_X\sigma_Y} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

- Interpretation of Correlation $(0 \le \rho_{XY} \le 1)$

1. $\rho_{XY} = 1$ Perfectly positively linearly related.
2. $\rho_{XY} = -1$ Perfectly negatively linearly related.
3. $\rho_{XY} = 0$ No linear relationship, may have other nonlinear relations
4. $\rho_{XY} > 0$ Positively linearly related
5. $\rho_{XY} < 0$ Negatively linearly related

18.3 Understanding Pho

- Theorem

If X and Y are independent, then $Corr(X,Y)=Cov(X,Y)=0$.

But, the converse of the theorem is not necessarily true. $corr(X,Y)=0$ does not mean X and Y are independent. It means X and Y are not linearly dependent.