Random variables, Expectation, Mean and Variance

Slides are adapted from STAT414 course at PennState

https://onlinecourses.science.psu.edu/stat414/

Random variable

- **Definition.** Given a random experiment with sample space *S,* a **random variable *X*** is a set function that assigns one and only one real number to each element *s* that belongs in the sample space *S*.
- The set of all possible values of the random variable *X,* denoted *x,* is called the **support**, or **space**, of *X.*

Example

- A rat is selected at random from a cage of male (M) and female rats (F). Once selected, the gender of the selected rat is noted. The sample space is thus:
- $S = \{M, F\}$
- Define the random variable X as follows:
- Let $X = 0$ if the rat is male.
- Let $X = 1$ if the rat is female.
- Note that the random variable X assigns one and only one real number (0 and 1) to each element of the sample space (M and F). The support, or space, of X is $\{0, 1\}$.
- Note that we don't necessarily need to use the numbers 0 and 1 as the support. For example, we could have alternatively (and perhaps arbitrarily?!) used the numbers 5 and 15, respectively. In that case, our random variable would be defined as $X = 5$ of the rat is male, and $X = 15$ if the rat is female.

Discrete random variable

- **Definition.** A random variable X is a **discrete random variable** if:
- there are a finite number of possible outcomes of X, or
- there are a countably infinite number of possible outcomes of X.

- Examples of discrete data include the number of siblings a randomly selected person has, the total on the faces of a pair of six-sided dice

Probability mass function

Definition. The **probability mass function**, $P(X = x) = f(x)$, of a discrete random variable X is a function that satisfies the following properties:

(1) $P(X = x) = f(x) > 0$ if $x \in$ the support S

(2) $\sum_{x \in S} f(x) = 1$

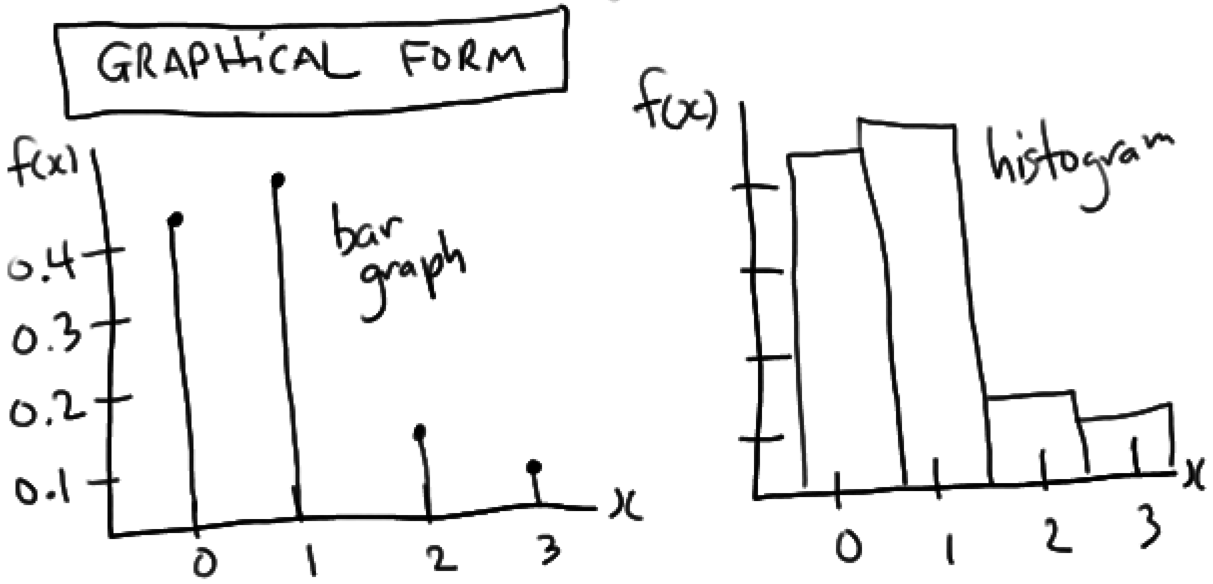
(3) $P(X \in A) = \sum_{x \in A} f(x)$

Example

- Let X equal the number of siblings of students in a class. The support of X is, of course, 0, 1, 2, 3, ... Because the support contains a countably infinite number of possible values, X is a discrete random variable with a probability mass function. Find $f(x) = P(X = x)$, the probability mass function of X, for all x in the support.

TABULAR FORM

x	0	1	2	3
$f(x) = P(X=x)$	0.41	0.45	0.11	0.03



Cumulative Distribution Function

$F_X(t)=P(X\leq t)$

The cdf of random variable X has the following properties:

1.$F_X(t)$ is a nondecreasing function of t, for $-\infty<t<\infty$.
2.The cdf, $F_X(t)$, ranges from 0 to 1. This makes sense since $F_X(t)$ is a probability.
3.If X is a discrete random variable whose minimum value is a, then $F_X(a)=P(X\leq a)=P(X=a)=f_X(a)$. If c is less than a, then $F_X(c)=0$.
4.If the maximum value of X is b, then $F_X(b)=1$.
5.Also called the *distribution function*.
6.All probabilities concerning X can be stated in terms of F.

Hypergeometric Distribution

- A crate contains 50 light bulbs of which 5 are defective and 45 are not. A Quality Control Inspector randomly samples 4 bulbs without replacement. Let *X* = the number of defective bulbs selected. Find the probability mass function, *f*(*x*), of the discrete random variable *X*.

Support: $x = 0, 1, 2, 3$ or 4

PMF: $P(X=0), P(X=1), \cdots, P(X=4)$

$$P(X=0) = \frac{\binom{5}{0}\binom{45}{4}}{\binom{50}{4}} \qquad P(X=1) = \frac{\binom{5}{1}\binom{45}{3}}{\binom{50}{4}}$$

$$P(X=2) = \frac{\binom{5}{2}\binom{45}{2}}{\binom{50}{4}}$$

In general:

$$f(x) = P(X=x) = \begin{cases} \dfrac{\binom{5}{x}\binom{45}{4-x}}{\binom{50}{4}} & x = 0,1,2,3,4 \\ 0, & \text{otherwise} \end{cases}$$

Hypergeometric Distribution

If we randomly select n items without replacement from a set of N items of which:

- m of the items are of one type
- and $N - m$ of the items are of a second type

then the probability mass function of the discrete random variable X is called the **hypergeometric distribution** and is of the form:

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

where the support S is the collection of nonnegative integers x that satisfies the inequalities:

- $x \leq n$
- $x \leq m$
- $n - x \leq N - m$

Mean

- Toss a fair, six-sided die many times. ***In the long run*** , what would the average (or "**mean**") of the tosses be?

$$\text{Mean} = \frac{1+3+2+\cdots+6+3+1}{36}$$

$$= \frac{(1+1+\cdots+1)+(2+2+\cdots+2)+\cdots+(6+6+\cdots+6)}{36}$$

$$= 1\left(\frac{6}{36}\right)+2\left(\frac{6}{36}\right)+\cdots+6\left(\frac{6}{36}\right)$$

$$= 1\left(\frac{1}{6}\right)+2\left(\frac{1}{6}\right)+\cdots+6\left(\frac{1}{6}\right)$$

$$= 1\cdot P(X=1)+2\cdot P(X=2)+\cdots+6(P(X=6))$$

- The mean is a weighted average, that is, an average of the values weighted by their respective individual probabilities.
- The mean is called the expected value of X, denoted $E(X)$ or by μ,

Expected value

Definition. If $f(x)$ is the p.m.f. of the discrete random variable X with support S, and if the summation:

$$\sum_{x \in S} u(x) f(x)$$

exists (that is, it is less than ∞), then the resulting sum is called the **mathematical expectation**, or the **expected value** of the function $u(X)$. The expectation is denoted $E[u(X)]$. That is:

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

Expectation

Theorem. When it exists, the mathematical expectation E satisfies the following properties:

(a) If c is a constant, then $E(c) = c$.

(b) If c is a constant and u is a function, then:

$E[cu(X)]=cE[u(X)]$

(a) $E(c) = \sum_{x \in S} c f(x) = c \sum_{x \in S} f(x) = c(1) = c$ ✓

(b) $E[cu(X)] = \sum_{x \in S} cu(x)f(x) = c \sum_{x \in S} u(x) f(x)$

$= cE[u(X)]$ ✓

Expectation

Theorem. Let c_1 and c_2 be constants and u_1 and u_2 be functions. Then, when the mathematical expectation E exists, it satisfies the following property:

$$E[c_1u_1(X) + c_2u_2(X)] = c_1E[u_1(X)] + c_2E[u_2(X)]$$

$$E\left[\sum_{i=1}^{k} c_i u_i(X)\right] = \sum_{i=1}^{k} c_i E[u_i(X)]$$

Mean

Definition. When the function $u(X) = X$, the expectation of $u(X)$, when it exists:

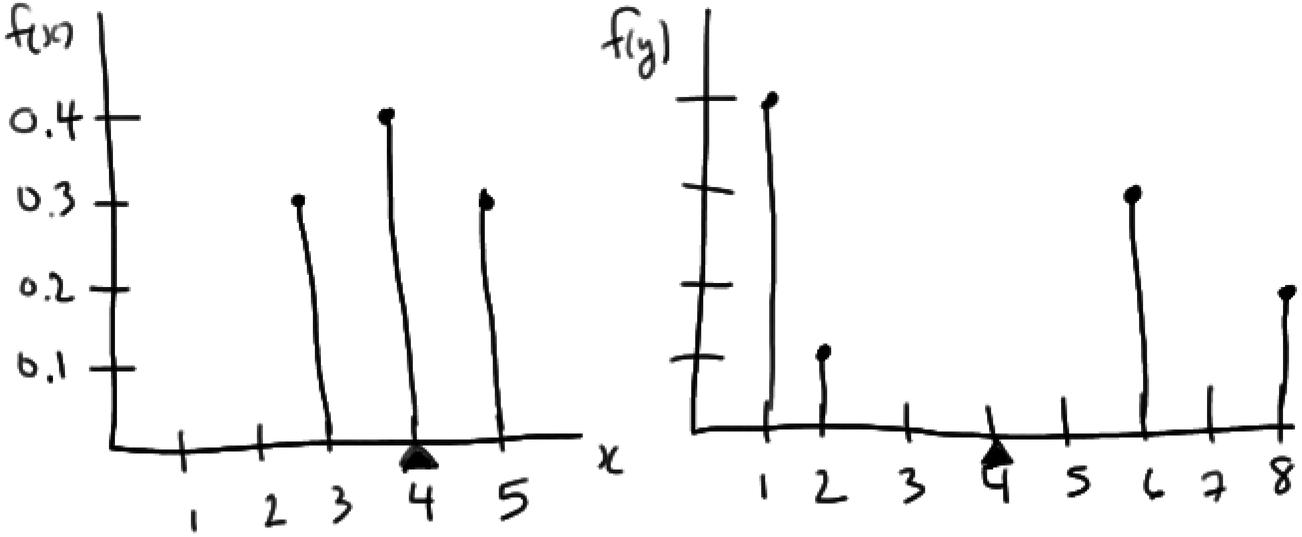
$$E[u(X)] = E(X) = \sum_{x \in S} x f(x)$$

is called the **expected value of** X, and is denoted $E(X)$. Or, it is called the **mean of** X, and is denoted as μ (the greek letter mu, read "mew"). That is, $\mu = E(X)$. The expected value of X can also be called the **first moment about the origin**.

Example

x	3	4	5
$f(x)$	0.3	0.4	0.3

y	1	2	6	8
$f(y)$	0.4	0.1	0.3	0.2



- $\mu_X = E(X) = 3(0.3) + 4(0.4) + 5(0.3) = 4$
- $\mu_Y = E(Y) = 1(0.4) + 2(0.1) + 6(0.3) + 8(0.2) = 4$

Variance

Definition. When $u(X) = (X - \mu)^2$, the expectation of $u(X)$:

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

is called the **variance of** X, and is denoted as $Var(X)$ or σ^2 ("sigma-squared"). The variance of X can also be called the **second moment of** X **about the mean** μ.

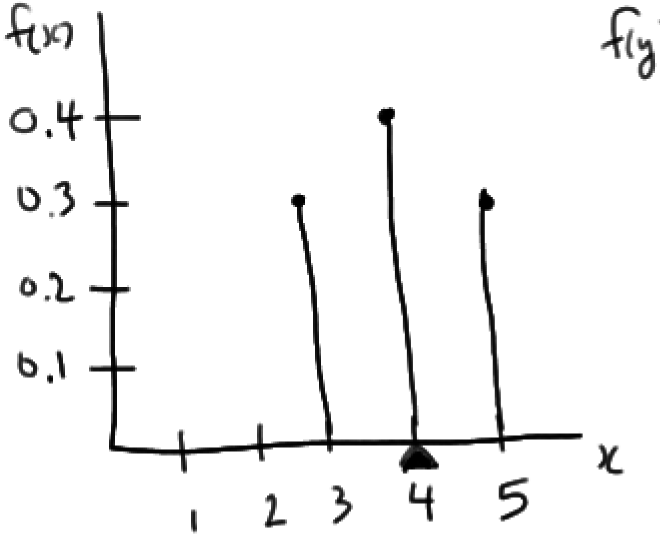
The positive square root of the variance is called the **standard deviation of** X, and is denoted σ ("sigma"). That is:

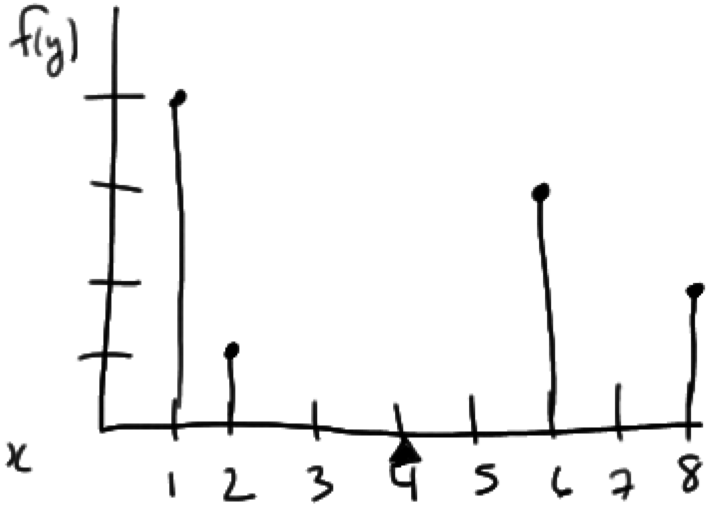
$$\sigma = \sqrt{Var(X)} = \sqrt{\sigma^2}$$

Example

x	3	4	5
f(*x*)	0.3	0.4	0.3

y	1	2	6	8
f(*y*)	0.4	0.1	0.3	0.2





Example

x	3	4	5
$f(x)$	0.3	0.4	0.3

y	1	2	6	8
$f(y)$	0.4	0.1	0.3	0.2

$$\sigma_X^2 = E[(X-\mu)^2] = (3-4)^2(0.3) + (4-4)^2(0.4) + (5-4)^2(0.3) = 0.6$$

And, therefore, the standard deviation of X is:

$$\sigma_X = \sqrt{0.6} = 0.77$$

Now, the variance of Y is calculated as:

$$\sigma_Y^2 = E[(Y-\mu)^2] = (1-4)^2(0.4) + (2-4)^2(0.1) + (6-4)^2(0.3) + (8-4)^2(0.2) = 8.4$$

And, therefore, the standard deviation of Y is:

$$\sigma_Y = \sqrt{8.4} = 2.9$$

Variance

Theorem: An easier way to calculate the **variance** of a random variable X is:

$$\sigma^2 = Var(X) = E(X^2) - \mu^2$$

$$\sigma^2 = E\left[(X-\mu)^2\right] = E\left[X^2 - 2X\mu + \mu^2\right]$$

$$= E(X^2) - 2\mu E(X) + \mu^2$$

$$= E(X^2) - 2\mu^2 + \mu^2$$

$$\sigma^2 = E(X^2) - \mu^2$$

Sample Mean and Variance

Definition. The **sample mean**, denoted $\bar{x}$ and read "x-bar," is simply the average of the n data points x_1, x_2, ..., x_n:

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

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

Definition. The **sample variance**, denoted s^2 and read "s-squared," summarizes the "spread" or "variation" of the data:

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

The **sample standard deviation**, denoted s is simply the positive square root of the sample variance. That is:

$$s = \sqrt{s^2}$$

$X \quad Y$ independent random variables

$E(X) = \mu_X \qquad E(Y) = \mu_Y$

$Var(X) = E((X-\mu_X)^2) = \sigma_X^2 \qquad Var(Y) = E((Y-\mu_Y)^2) = \sigma_Y^2$

$Z = X + Y \qquad E(-Y) = -E(Y) \qquad A = X - Y$

$E(Z) = E(X+Y) = E(X) + E(Y) \qquad E(A) = E(X-Y) = E(X) - E(Y)$

$\mu_Z = \mu_X + \mu_Y \qquad \mu_{X-Y} = \mu_A = \boxed{\mu_X - \mu_Y}$

$Var(Z) = Var(X) + Var(Y) \qquad \sigma_A^2 = \sigma_{X-Y}^2 = \sigma_{X+(-Y)}^2 = \sigma_X^2 + \boxed{\sigma_{-Y}^2}$

$\sigma_Z^2 = \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 \qquad \sigma_{-Y}^2 = Var(-Y) = E((-Y - E(-Y))^2)$

$E((-1)^2 (Y + E(-Y))^2)$

$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 \qquad E((Y - E(Y))^2) = \sigma_Y^2$

KhanAcademy, Random Variables

Moments

- The expected values $E(X)$, $E(X^2)$, $E(X^3)$, ..., and $E(X^r)$ are called **moments**
- *Recall that:*
- $\mu = E(X)$
- $\sigma^2 = Var(X) = E(X^2) - \mu^2$

Moment Generating Function

Let X be a discrete random variable with probability mass function $f(x)$ and support S. Then:

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

is the **moment generating function of *X*** as long as the summation is finite for some interval of t around 0.

(1) The mean of X can be found by evaluating the first derivative of the moment-generating function at $t = 0$. That is:

$$\mu = E(X) = M'(0)$$

(2) The variance of X can be found by evaluating the first and second derivatives of the moment-generating function at $t = 0$. That is:

$$\sigma^2 = E(X^2) - [E(X)]^2 = M''(0) - [M'(0)]^2$$

- The moment generation function for a binomial variable is

$$M(t) = [(1-p) + pe^t]^n$$

$$M'(t) = n[1-p+pe^t]^{n-1}(pe^t)$$

And, setting $t = 0$, we get the binomial mean $\mu = np$:

To find the variance, we first need to take the second derivative of $M(t)$ with respect to t. Doing so, we get:

$$M''(t) = n[1-p+pe^t]^{n-1}(pe^t) + (pe^t)n(n-1)[1-p+pe^t]^{n-2}(pe^t)$$

And, setting $t = 0$, and using the formula for the variance, we get the binomial variance $\sigma^2 = np(1-p)$:

Binomial Random variable

Definition. A discrete random variable X is a **binomial random variable** if:

1. An experiment, or trial, is performed in exactly the same way n times.
2. Each of the n trials has only two possible outcomes. One of the outcomes is called a "success," while the other is called a "failure." Such a trial is called a **Bernoulli trial**.
3. The n trials are independent.
4. The probability of success, denoted p, is the same for each trial. The probability of failure is $q = 1 - p$.
5. The random variable X = the number of successes in the n trials.

Binomial Distribution

Definition. The **probability mass function of a binomial random variable** X is:

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

We denote the **binomial distribution** as $b(n, p)$. That is, we say:

$$X \sim b(n, p)$$

where the tilde (~) is read "as distributed as," and n and p are called **parameters** of the distribution.

Example

$P(\text{P}) =$ 0.8 and $P(\text{N}) = 0.2$.

$$P(X=0) = P(\text{NNN}) = 0.2 \times 0.2 \times 0.2 = 1 \times (0.8)^0 \times (0.2)^3$$

And, by independence and mutual exclusivity of NNP, NPN, and PNN:

$$P(X=1) = P(\text{NNP}) + P(\text{NPN}) + P(\text{PNN}) = 3 \times 0.8 \times 0.2 \times 0.2 = 3 \times (0.8)^1 \times (0.2)^2$$

Likewise, by independence and mutual exclusivity of PPN, PNP, and NPP:

$$P(X=2) = P(\text{PPN}) + P(\text{PNP}) + P(\text{NPP}) = 3 \times 0.8 \times 0.8 \times 0.2 = 3 \times (0.8)^2 \times (0.2)^1$$

Finally, by independence:

$$P(X=3) = P(\text{PPP}) = 0.8 \times 0.8 \times 0.8 = 1 \times (0.8)^3 \times (0.2)^0$$

Cumulative Probability Distribution

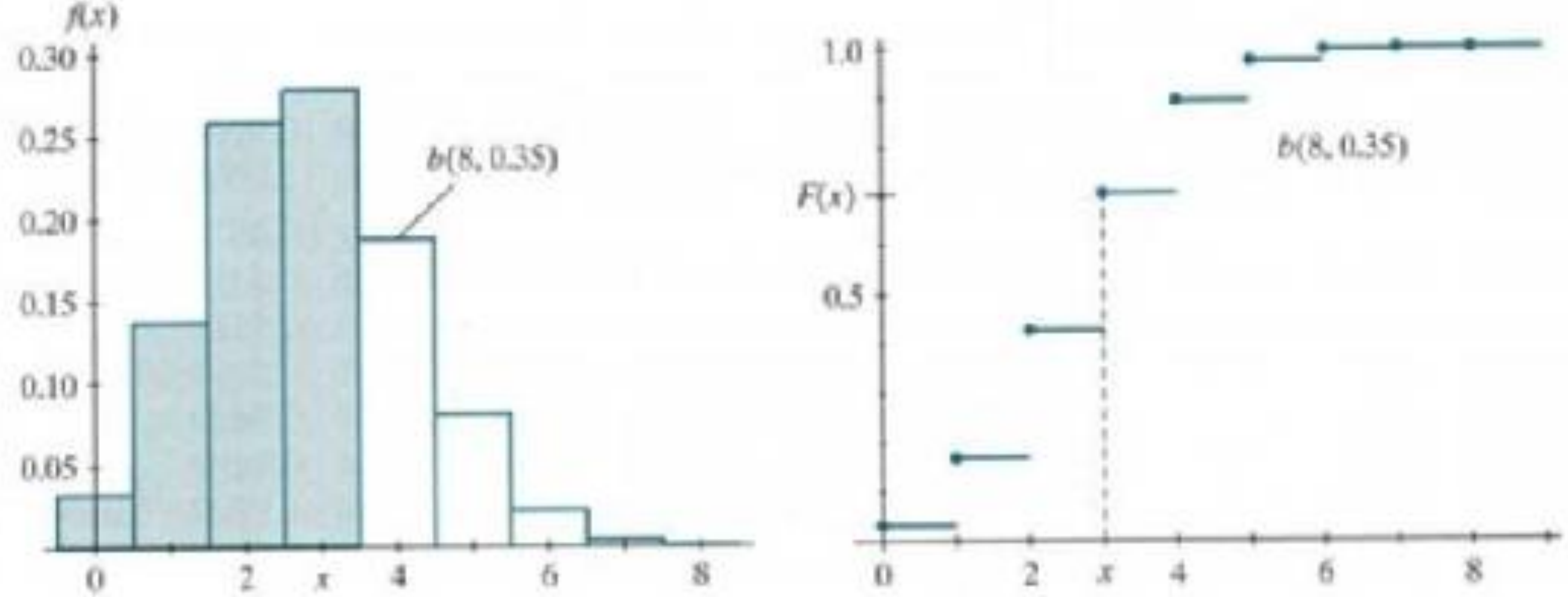
Definition. The function:

$$F(x) = P(X \leq x)$$

is called a **cumulative probability distribution**. For a discrete random variable X, the cumulative probability distribution $F(x)$ is determined by:

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

Cumulative Probability Function



$$F(x) = P(X \le x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Mean and Variance of a Binomial R.V.

Theorem. If X is a binomial random variable, then the **mean of X** is:

$$\mu = np$$

Theorem. If X is a binomial random variable, then the **variance of X** is:

$$\sigma^2 = np(1-p)$$

and the **standard deviation of X** is:

$$\sigma = \sqrt{np(1-p)}$$

$$\mu = E(x) = \sum_{x=0}^{n} x \cdot \frac{n!}{x!\,(n-x)!}\, p^{x} (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!\,(n-x)!}\, p^{x} (1-p)^{n-x}$$

Let $k = x-1$, so $x = k+1$

$$= \sum_{k=0}^{n-1} \frac{n!}{k!\,(n-1-k)!}\, p^{k+1} (1-p)^{n-1-k}$$

$$= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!\,(n-1-k)!}\, p^{k} (1-p)^{n-1-k}$$

$\rightarrow b(n-1, p)$

$\rightarrow 1$

$= np$

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

TRICK!

$$= E(x^2) \underbrace{- E(x) + E(x)}_{\text{add } 0} - [E(x)]^2$$

$$= \underset{①}{E[X(X-1)]} + \underset{②}{E(X)} - \underset{③}{[E(X)]^2}$$

Proof. The definition of the expected value of a function gives us:

$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1) \times f(x) = \sum_{x=0}^{n} x(x-1) \times \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

The first two terms of the summation equal zero when $x = 0$ and $x = 1$. Therefore, the bottom index on the summation can be changed from $x = 0$ to $x = 2$, as it is here:

$$E[X(X-1)] = \sum_{x=2}^{n} x(x-1) \times \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$E(x(x-1)) = \sum_{x=2}^{n} \frac{n!}{(x-2)!\,(n-x)!} p^x (1-p)^{n-x}$$

Let $k = x-2$, so $x = k+2$

$$= \sum_{k=0}^{n-2} \frac{n!}{k!\,(n-2-k)!} p^{k+2} (1-p)^{n-2-k}$$

$$= n(n-1)p^2 \sum_{k=0}^{n-2} \frac{(n-2)!}{k!\,(n-2-k)!} p^k (1-p)^{n-2-k}$$

$\to b(n-2, p)$ $\to 1$

$$= n(n-1)p^2$$

$$\sigma^2 = E[X(X-1)] + E(X) - (E(X))^2$$

$$= n(n-1)p^2 + np - n^2p^2$$

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$= np - np^2 = np(1-p) \checkmark$$

Continuous Random Variable

- the amount of rain, in inches, that falls in a randomly selected storm
- the weight, in pounds, of a randomly selected student
- the square footage of a randomly selected three-bedroom house

- 13 mm rain drop may be 13.456789 or 13.00000000012

Density histogram
Histogram of continuous data

- Measured nose lengths

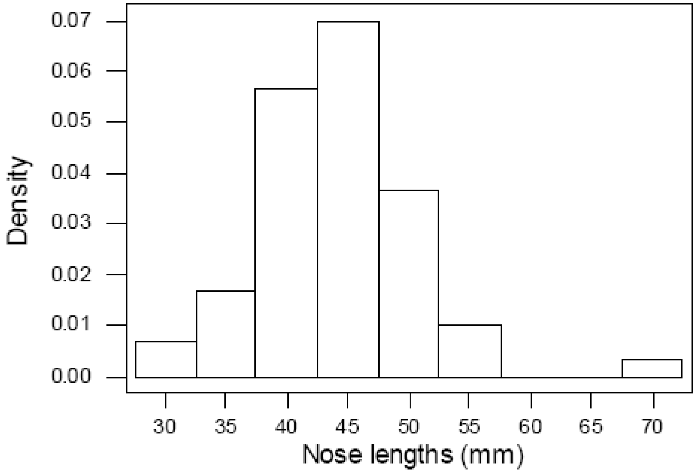
(error in measurement)

38 50 38 40 35 52 45 50 40 32 40 47 70 55 51
43 40 45 45 55 37 50 45 45 55 50 45 35 52 32
45 50 40 40 50 41 41 40 40 46 45 40 43 45 42
45 45 48 45 45 35 45 45 40 45 40 40 45 35 52

1. Determine the number, n, in the sample.
2. Define k class intervals $(c_0, c_1]$, $(c_1, c_2]$, ..., $(c_{k-1}, c_k]$.
3. Determine 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 in the sample — that is, $f_i \div n$.
5. For a **frequency histogram**: draw a 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

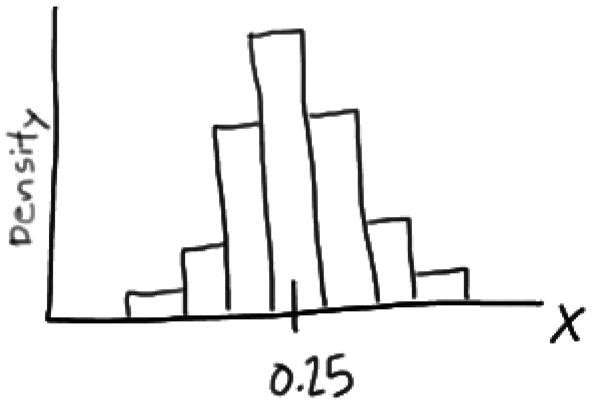
$h(x)=f_i / n(c_i-c_{i-1})$ for $c_{i-1}<x\le c_i$ $i = 1, 2,..., k$.

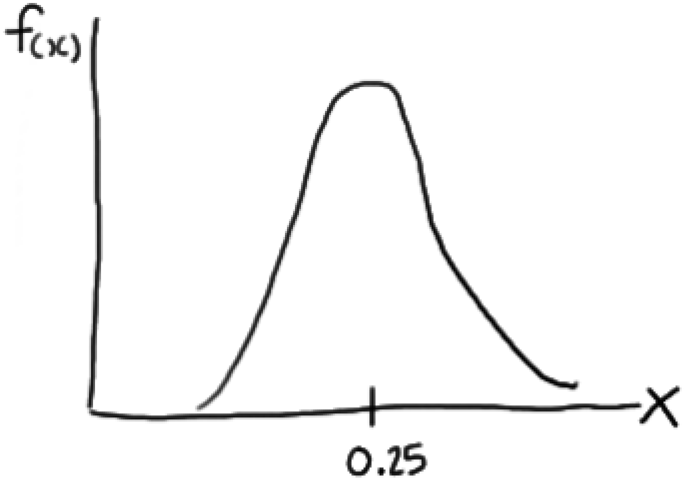
Class interval	Tally	Frequency	Relative Frequency	Density height
27.5-32.5	||	2	0.033	0.0066
32.5-37.5	|||||	5	0.083	0.0166
37.5-42.5	||||| ||||| ||||| ||	17	0.283	0.0566
42.5-47.5	||||| ||||| ||||| ||||| |	21	0.350	0.0700
47.5-52.5	||||| ||||| |	11	0.183	0.0366
52.5-57.5	|||	3	0.050	0.010
57.5-62.5		0	0	0
62.5-67.5		0	0	0
67.5-72.5	|	1	0.017	0.0034
		60	0.999 (rounding)	

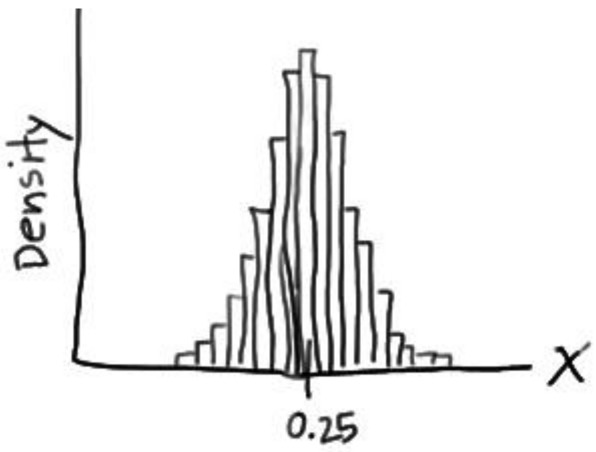


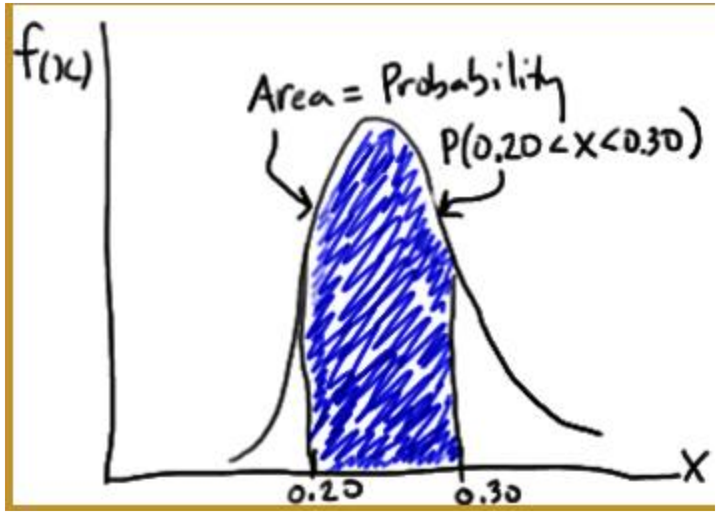
the area of the entire histogram equals 1.

Probability Density Function









Probability Density Function (p.d.f)

Definition. The **probability density function** ("**p.d.f.**") of a continuous random variable X with support S is an integrable function $f(x)$ satisfying the following:

(1) $f(x)$ is positive everywhere in the support S, that is, $f(x) > 0$, for all x in S

(2) The area under the curve $f(x)$ in the support S is 1, that is:

$$\int_S f(x)dx = 1$$

(3) If $f(x)$ is the p.d.f. of x, then the probability that x belongs to A, where A is some interval, is given by the integral of $f(x)$ over that interval, that is:

$$P(X \in A) = \int_A f(x)dx$$

Notes

- when X is continuous, $P(X = x) = 0$ for all x in the support.

- $P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$ for any constants a and b.

Cumulative Distribution Function

- Recall that c.d.f for discrete R.V. :

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

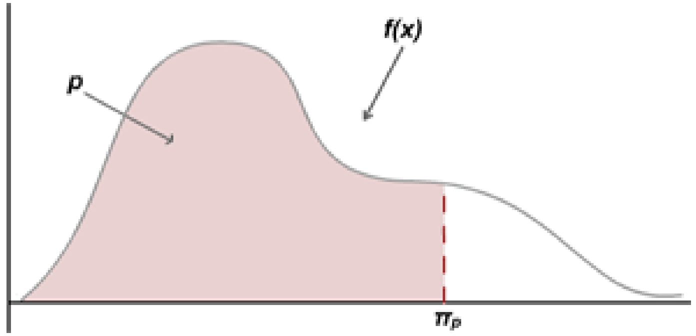
Definition. The **cumulative distribution function** ("**c.d.f.**") of a continuous random variable X is defined as:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

for $-\infty < x < \infty$.

Percentiles

Definition. If X is a continuous random variable, then the **$(100p)^{th}$ percentile** is a number π_p such that the area under $f(x)$ and to the left of π_p is p.



That is, p is the integral of $f(x)$ from $-\infty$ to π_p:

$$p = \int_{-\infty}^{\pi_p} f(x)dx = F(\pi_p)$$

Some percentiles are given special names:

- The 25th percentile, $\pi_{0.25}$, is called the **first quartile** (denoted q_1).
- The 50th percentile, $\pi_{0.50}$, is called the **median** (denoted m) or the **second quartile** (denoted q_2).
- The 75th percentile, $\pi_{0.75}$, is called the **third quartile** (denoted q_3).

Expectations

Definition. The **expected value** or **mean** of a continuous random variable X is:

$$\mu = E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

The **variance** of a continuous random variable X is:

$$\sigma^2 = Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx$$

Alternatively, you can still use the shortcut formula for the variance, $\sigma^2 = E(X^2) - \mu^2$, with:

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx$$

The **standard deviation** of a continuous random variable X is:

$$\sigma = \sqrt{Var(X)}$$

The **moment generating function** of a continuous random variable X, if it exists, is:

$$M(t) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx$$

for $-h < t < h$.

$$E(X) = M'(0)$$

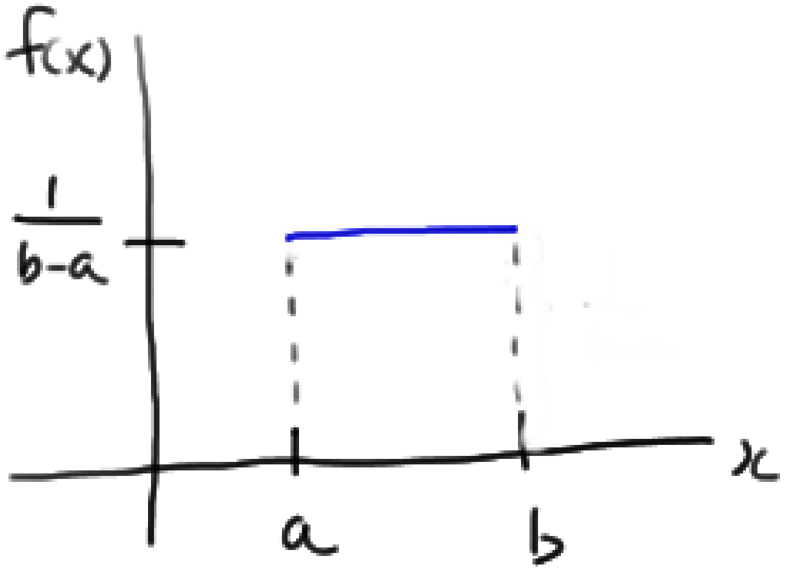
$$Var(X) = M''(0) - (M'(0))^2$$

Uniform Distributions

Definition. A continuous random variable X has a **uniform distribution**, denoted $\boldsymbol{U(a, b)}$, if its probability density function is:

$$f(x) = \frac{1}{b-a}$$

for two constants a and b, such that $a < x < b$. A graph of the p.d.f. looks like this:

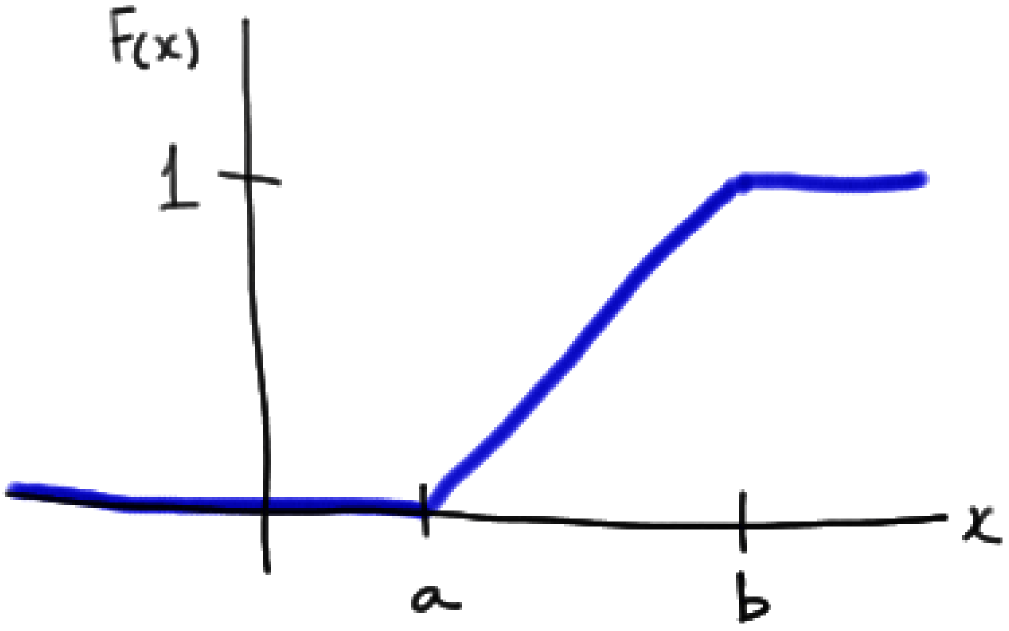


CDF of a uniform R.V.

Definition. The **cumulative distribution function of a uniform random variable** X is:

$$F(x) = \frac{x-a}{b-a}$$

for two constants a and b such that $a < x < b$. A graph of the c.d.f. looks like this:



Properties of Uniform variables

Theorem. The mean of a continuous uniform random variable defined over the support $a < x < b$ is:

$$\mu = E(X) = \frac{a+b}{2}$$

$$\mu = E(X) = \int_a^b x\left(\frac{1}{b-a}\right)dx$$

$$= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_{x=a}^{x=b} = \frac{1}{2(b-a)}(b^2-a^2)$$

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

$$E(x^2) = \int_a^b x^2\left(\frac{1}{b-a}\right)dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_{x=a}^{x=b}$$

$$= \frac{1}{b-a}\frac{(b^3-a^3)}{3} = \frac{1}{b-a}\cdot\frac{(b-a)(b^2+ab+a^2)}{3}$$

$$= \frac{b^2+ab+a^2}{3}$$

Theorem. The variance of a continuous uniform random variable defined over the support $a < x < b$ is:

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

$$\sigma^2 = E(X^2) - \mu^2 = \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2$$

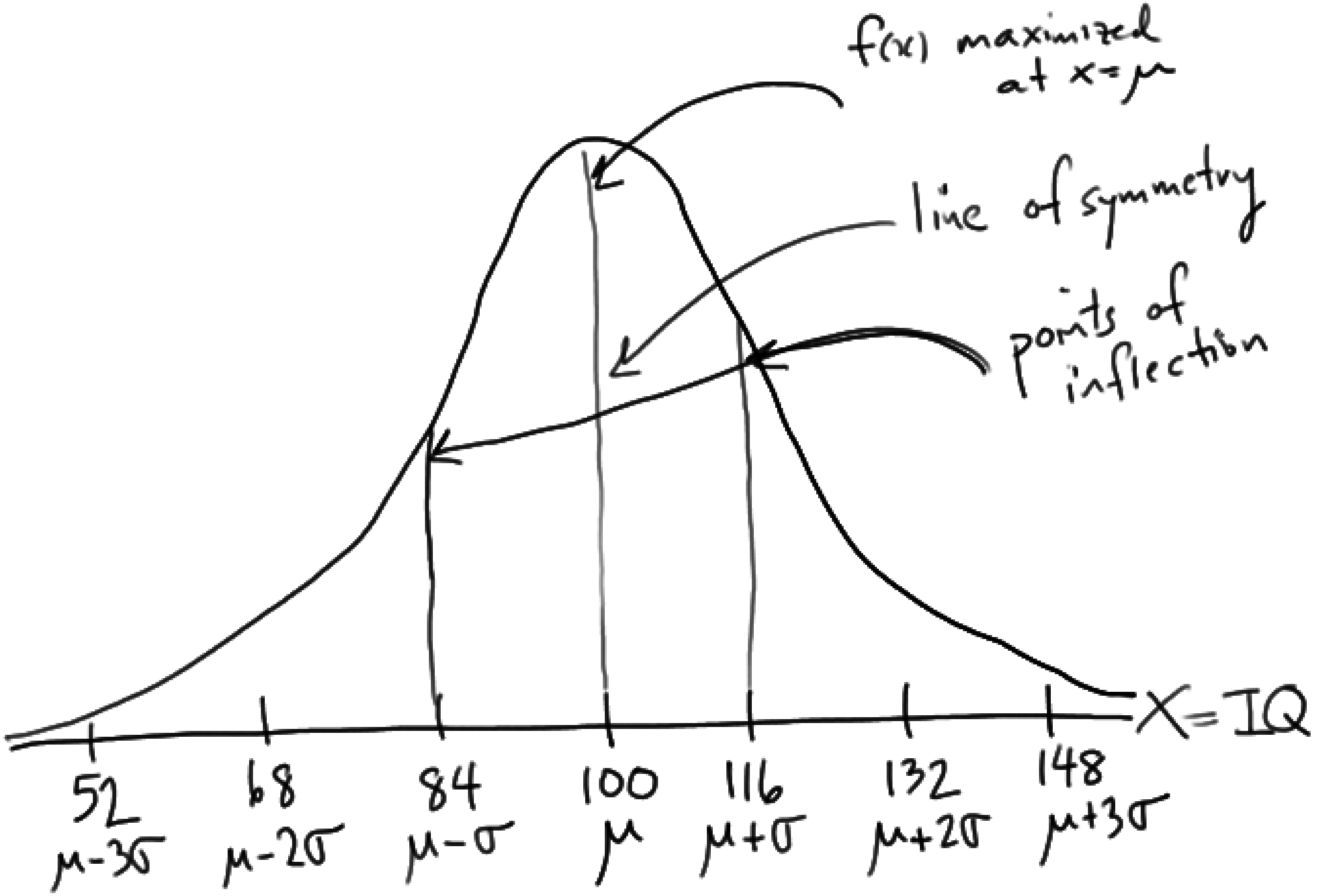
Normal distributions

Definition. The continuous random variable X follows a **normal distribution** if its probability density function is defined as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

for $-\infty < x < \infty$, $-\infty < \mu < \infty$, and $0 < \sigma < \infty$. The **mean** of X is μ and the **variance** of X *is* σ^2. We say $\boldsymbol{X \sim N(\mu, \sigma^2)}$.

Example



f(x) maximized at x=μ
line of symmetry
points of inflection
X=IQ
52
μ-3σ
68
μ-2σ
84
μ-σ
100
μ
116
μ+σ
132
μ+2σ
148
μ+3σ

Characteristics

1) All normal curves are **bell-shaped** with points of inflection at $\mu \pm \sigma$.

Proof. The proof is left for you as an exercise.

2) All normal curves are **symmetric about the mean** μ.

Proof. All normal curves are symmetric about the mean μ, because $f(\mu+x) = f(\mu-x)$ for all x. That is:

$$f(\mu+x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x+\mu-\mu}{\sigma}\right)^2\right\} = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right\}$$

equals:

$$f(\mu-x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\mu-x-\mu}{\sigma}\right)^2\right\} = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{-x}{\sigma}\right)^2\right\} = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right\}$$

Therefore, by the definition of symmetry, the normal curve is symmetric about the mean μ.

3) The area under an entire normal curve is 1.

Characteristics

4) All normal curves are positive for all x. That is, $f(x) > 0$ for all x.

Proof. The standard deviation σ is defined to be positive. The square root of 2π is positive. And, the natural exponential function is positive. When you multiply positive terms together, you, of course, get a positive number.

5) The limit of $f(x)$ as x goes to infinity is 0, and the limit of $f(x)$ as x goes to negative infinity is 0. That is:

$$\lim_{x \to \infty} f(x) = 0 \text{ and } \lim_{x \to -\infty} f(x) = 0$$

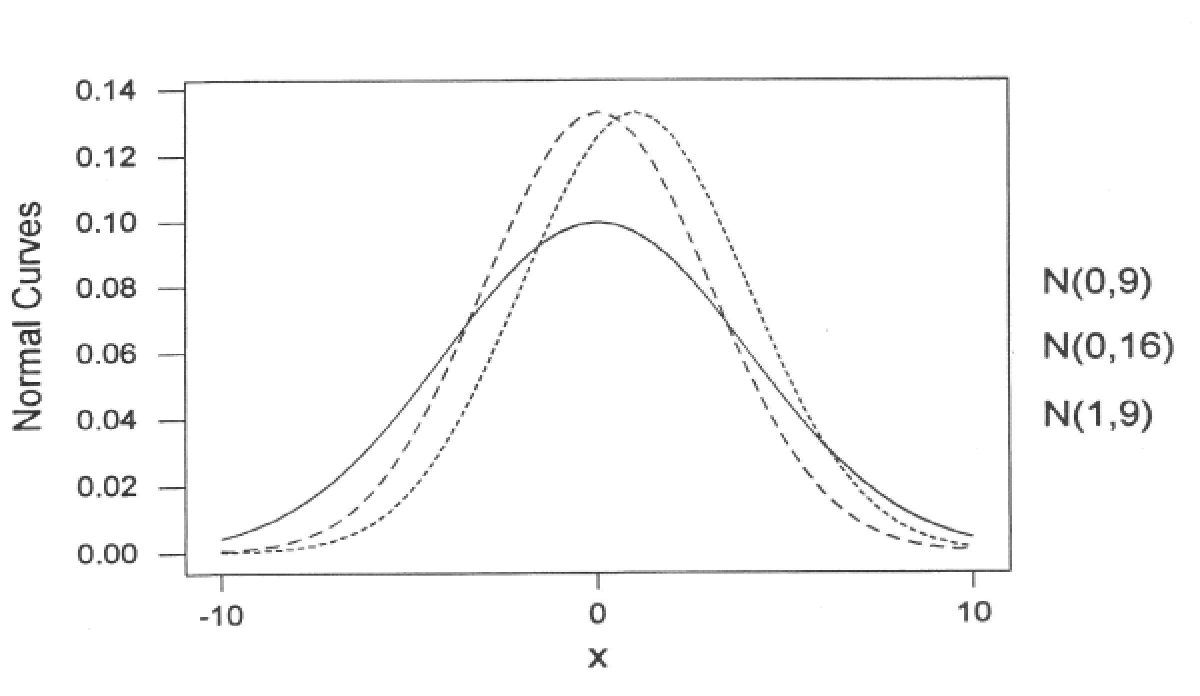
Proof. The function $f(x)$ depends on x only through the natural exponential function $\exp[-x^2]$, which is known to approach 0 as x approaches infinity or negative infinity.

6) The height of any normal curve is maximized at $x = \mu$.

Characteristics

7) The shape of any normal curve depends on its mean μ and standard deviation σ.

Proof. Given that the curve $f(x)$ depends only on x and the two parameters μ and σ, the claimed characteristic is quite obvious. An example is perhaps more interesting than the proof. Here is a picture of three superimposed normal curves —one of a $N(0, 9)$ curve, one of a $N(0, 16)$ curve, and one of a $N(1, 9)$ curve:

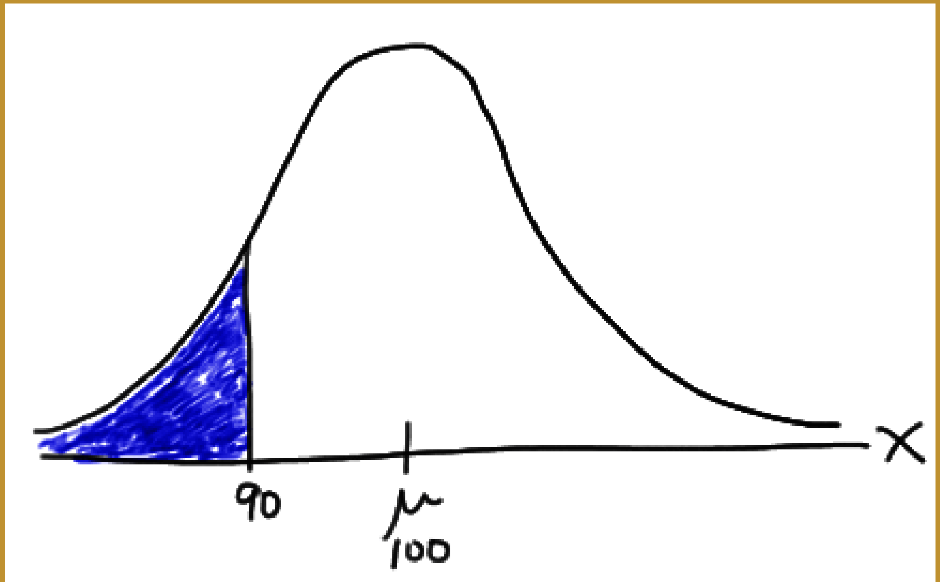


As claimed, the shapes of the three curves differ, as the means μ and standard deviations σ differ.

Example

Let X equal the IQ of a randomly selected American. Assume $X \sim N(100, 16^2)$. What is the prob
randomly selected American has an IQ below 90?

Solution. As is the case with all continuous distributions, finding the probability involves fi
under the curve and to the left of the line $x = 90$:



That is:

$$P(X \leq 90) = F(90) = \int_{-\infty}^{90} \frac{1}{16\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-100}{16}\right)^2\right\} dx$$

Standard Normal distribution

Theorem. If $X \sim N(\mu, \sigma^2)$, then:

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

follows the $N(0,1)$ distribution, which is called the **standardized** (or **standard**) **normal distribution**.

The theorem leads us to the following strategy for finding probabilities $P(a < X < b)$ when a and b are constants, and X is a normal random variable with mean μ and standard deviation σ:

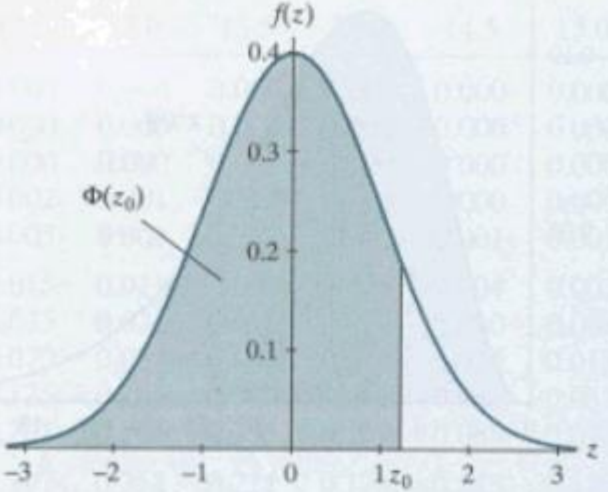
1) Specify the desired probability in terms of X.

2) Transform X, a, and b, by:

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

3) Use the standard normal $N(0,1)$ table, typically referred to as the ***Z*-table**, to find the desired probability.

z-table



$$P(Z \le z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\, dw$$

$$\Phi(-z) = 1 - \Phi(z)$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879

Properties of normal distributions

Recall that the probability density function of a normal random variable is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

Theorem. The mean and variance of a normal random variable X are, respectively, μ and σ^2.