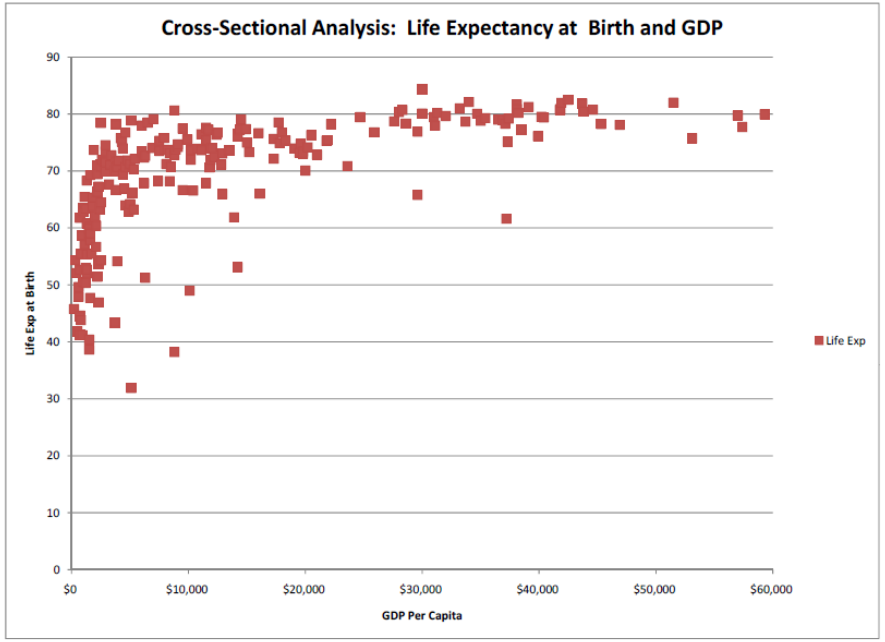
13. Probability and Statistics

Mwaso Mnensa

The Structure of Economic Data

Cross sectional

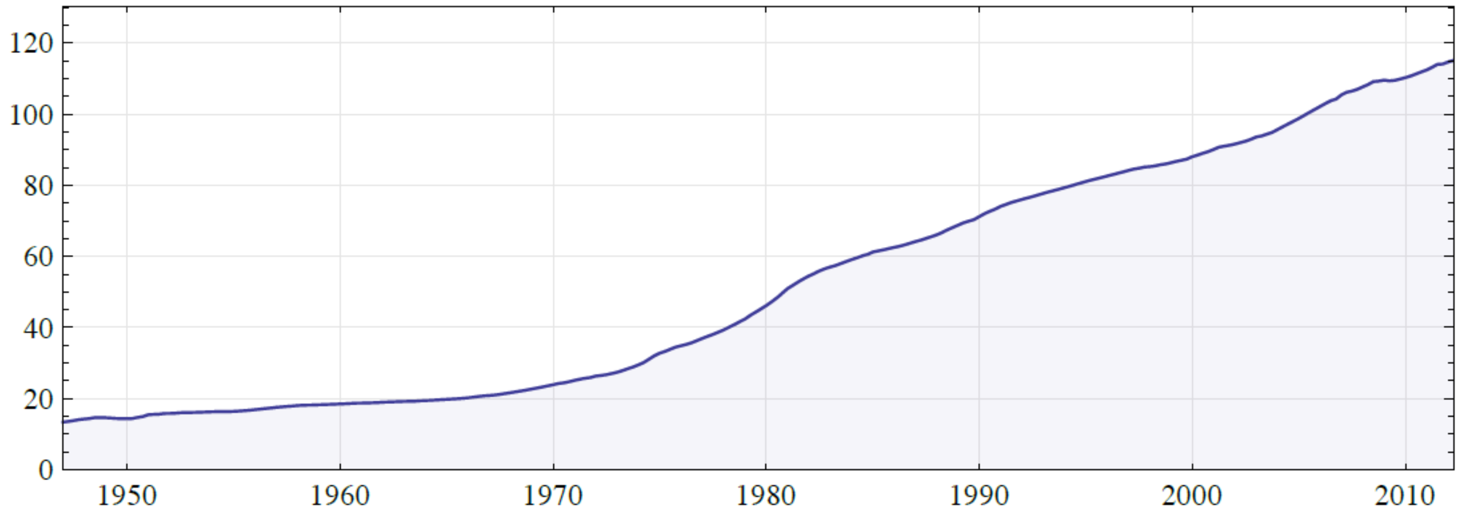
A sample of individuals (or other units) at a given point in time. Often can be assumed to be independent and identically distributed so the order of observations doesn't matter.



Time Series

- Observations on one individual (or other unit) over time, typically not independent across time so chronological ordering of the data is important. Often used for macroeconomic variables.

US Consumer Price Indices



(from Q1 1947 to Q2 2012)

(index reference base: 2005 = 100)

- **Pooled cross section**

Repeated cross-sections conducted at more than one point intime. Ordering doesn’t matter, but typically important to include a variable indicating when sample was taken.

- **Panel data**

A sample of the same individuals/units followed over time.

Population, parameters and random sampling

- **Population**

Any well-defined group of subjects of interest e.g. Math Review students. As a theoretical idea, you can think of a population as a group of subjects of interest of infinite size.

- **Parameter**

A constant (usually unknown) that describes the relationship among variables in the population

e.g. mean hours of daily studying in the population

e.g. a relationship between two variables in the population.

If we have a model:

$\mathit{Midterm\ Score} = \beta_0 + \beta_1\, \mathit{Study\ Hours} + \text{error}$, then β_0 and β_1 are parameters

- **Random sample**

A sample consisting of independent observations drawn from a common population (i.e. the variables come from a common distribution). Independent means that the fact that one individual is in the sample does not affect the likelihood of any other individual being in the sample.

We say variables from a random sample are independently identically distributed (i.i.d.); the sample should be “representative” of the population; every member of the population should have an equal chance of being selected.

Example

- Suppose Prof. Terry randomly chooses one of his former students. Suppose she earned an 83 on the final exam and studied 8 hours/week. That is one individual in the sample. Now suppose Prof. Terry asks her the scores of her friends from APEC8002, would this help me generate a random sample of the scores?

Describing a sample

- Histogram – a bar graph showing the frequency with which data takes on certain values
- Sample mean = average = $\bar{X} = \left(\frac{1}{\mathrm{n}}\right) \sum X_i$
- Mode – most frequent value
- From the histogram we can also see the maximum and minimum values.
- Percentile – value x for which a certain percentage of the sample lies below x (e.g. if 10% of students study less than 1.5 hours, then 1.5 is the 10th percentile.)
- Median – 50th percentile or middle observation if the observations are ranked by size. (If there are an even number, take the average of the two middle observations.)

Basic Probability

- Experiment - any procedure that could theoretically be infinitely repeated and has a well-defined set of outcomes

  (e.g. 3-coin flip or growing wheat on a parcel of land)

- Random trial – an experiment run one time (e.g. one 3-coin flip)

- Sample space – the set or list of all possible outcomes from a random trial

  (e.g. : HHH, HHT, HTH, HTT, THH, THT, TTH, TTT)

- Event – a subset of the outcomes in the sample space (e.g. exactly one head)
- Relative frequency – fraction of random trials for which an event occurs --- fraction of observations for which a variable X takes on some value or some range of values (e.g. Suppose we generated the following data doing the experiment 62 times.

HHH 10

HHT 9

HTH 8

HTT 9 *

THH 4

THT 5 *

TTH 12 *

TTT 5)

- (We got exactly one head—the outcomes marked with * — 26 times.)

- Probability –the relative frequency approached in the limit as a sample becomes larger and larger or as an experiment is repeated more and more times, denoted by $\Pr(A)$, $P(A)$, $or\ p_j$ - a measure of how likely it is that an outcome will occur when we conduct the next random trial.
- *Pr(exactly one head) = 3/8*

- Probability tree – a diagram of potential outcomes used to calculate probabilities

- “A or B” = union of A and B = the set of outcomes in event A or event B or both

Conditional probability $\Pr(A|\, B)$

- The probability of A occurring given that B occurs $\Pr(\mathrm{A}|\mathrm{B})$ = “probability of A given B” = “probability of A conditional on B” = the probability of an event A given that event B has occurred:
- $P(A \mid B) = P(A \text{ and } B) P(B)$
- Prob (exactly one head|first flip is a tail) = 0.25/0.5 = 0.5

Complement

- An event that includes everything else but the outcomes in event A (designated $A^c, \sim A,$ “$not\ A$”).

Independent events

- Events for which the probability of one event is unrelated to the outcome of the other event
- Example 1. A standard example of independence is flipping a coin again and again. Because the outcome on any particular flip has nothing to do with the outcomes on other flips, independence is an appropriate assumption.
- Example 2. If 2 cards are drawn with replacement (i.e., I draw a card and I immediately put it again into the deck), the event of drawing a red card on the first trial and that of drawing a red card on the second trial are independent.
- If A and B are independent, then $\Pr(A|B) = \Pr(A)$
- Therefore $\Pr(A \; and \; B) = \Pr(A|B) * \Pr(B) = \Pr(A) * \Pr(B)$ if events are independent

Properties of the probability function

1. $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$.
2. $\mathbb{P}[\varnothing] = 0$.
3. $\mathbb{P}[A] \leq 1$.
4. **Monotone Probability Inequality**: If $A \subset B$, then $\mathbb{P}[A] \leq \mathbb{P}[B]$.
5. **Inclusion-Exclusion Principle**: $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$.
6. **Boole's Inequality**: $\mathbb{P}[A \cup B] \leq \mathbb{P}[A] + \mathbb{P}[B]$.
7. **Bonferroni's Inequality**: $\mathbb{P}[A \cap B] \geq \mathbb{P}[A] + \mathbb{P}[B] - 1$.

Example

- In a three coin toss with a fair coin, are the events “exactly one head” and “first flip is a tail” independent?

Example

- Suppose we roll a fair die once.

a) Write the sample space for this experiment

b) Compute the probabilities of the following events:

A = {2,4,6} B = {1,2,3,4}

c) What is the probability of {A) B}

d) Are these events independent? Why?

Now suppose that die is “fixed” such that P(6) = 3/8 whereas P(1)=P(2)=…=P(5)=1/8 and we repeat the experiment.

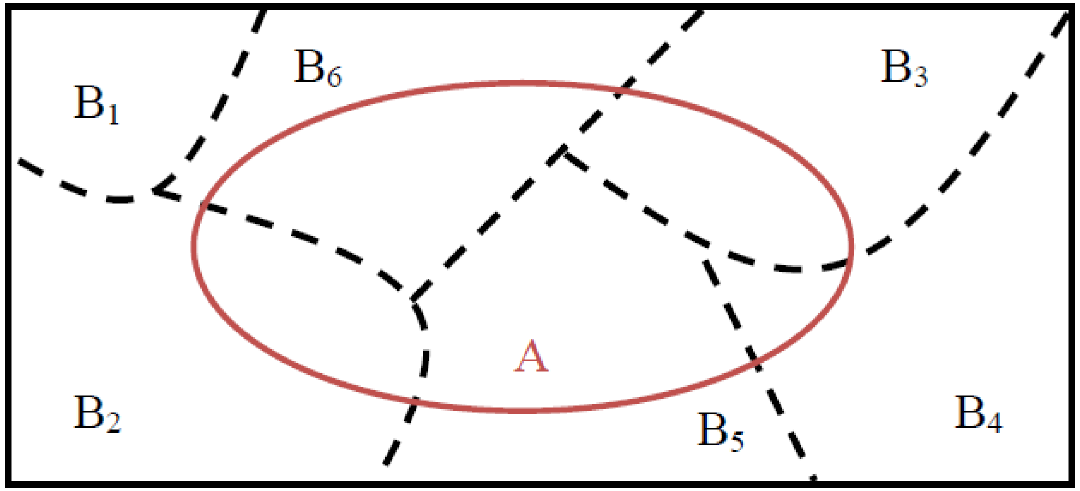
e) Are events A and B independent in this case?

Example

- Using the definition of independence and the rules of probability show that if A and B are independent events then $A\ and\ B^c$ are also independent.

Law of total probability

- Suppose we can partition the sample space in n---disjoint events: $B_1, \ldots, B_n$. Then we have:
- $P(A) = \sum P(A|B_i) * P(B_i)$
- We can graphically see this:



Examples

There are 3 different urns containing colored balls. Urn 1 has 5 reds and 3 greens. Urn 2 has 6 reds, 20 greens, 1 blue. Urn 3 has 10 red balls.

- a) If a ball is chosen at random and the probabilities of choosing any urn are the same, what is the probability of drawing one green ball?
- b) Suppose that the probability of choosing urn 1 is P(Urn 1) = 1/3 and the probability of choosing urn 2 is P(Urn 2) = 1/6. What is the probability of drawing a blue ball?

Random variable

A variable that takes on numerical values and has an outcome determined by an experiment (e.g. 0, 1, 2 or 3 where variable=# of heads, 0-50 bushels of wheat, number of hours of TV watched.)

- **Binary random variable**

Bernoulli random variable or dummy variable – a random variable that takes the value 0 or 1 (can define qualitative variables this way – female=1, male=0, or did you watch TV yesterday yes/no.)

- The probability X=1 is written as $\Pr(X = 1) = \theta, where\ 0 \leq \theta \leq 1.$
- **Discrete random variable**

A random variable that takes on a finite (or countably infinite) number of values (e.g. the number of days it snows in Twin Cities in a winter or a dummy variable.)

- **Continuous random variable**

A random variable that takes an infinite number of values, and takes on any real value with zero probability (e.g. the exact temperature outside, exact arm span.)

F. Probability Distributions for Discrete Random Variables

F. Probability Distributions for Discrete Random Variables

- The probability distribution function (p.d.f.) of a discrete random variable X describes the possible outcomes of X and their probabilities. Denoted f(x), it is the probability that the discrete random variable X takes on some particular value x
- Note that, since f(x) is a probability, $0 \leq f(x) \leq 1$

Probability distributions of discrete variables

The **probability distribution function (p.d.f.)** of a discrete random variable X describes the possible outcomes of X and their probabilities. Denoted $f(x)$, it is the probability that the discrete random variable X takes on some particular value x

$$f(x_j) = \begin{cases} p_j & \text{where } j = 1, \ldots, k \\ 0 & \text{otherwise} \end{cases}$$

Note that, since f(x) is a probability, $0 \leq f(x) \leq 1$

Example: Suppose we flip a coin three times

j	X_j = # heads	$f(x_j) = p_j = Pr(X=x_j)$
1	0	0.125
2	1	0.375
3	2	0.375
4	3	0.125

Cumulative distribution function

The **cumulative distribution function (c.d.f.)** describes the probability that the random variable X is less than or equal to a particular value x.

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

Note the c.d.f. is a non-decreasing function of x and $0 \leq F(x) \leq 1$.

j	X_j=#heads	$f(x_j) = p_j = Pr(X=x_j)$	$F(x_j) = Pr(X<=x_j)$
1	0	0.125	0.125
2	1	0.375	0.5
3	2	0.375	0.875
4	3	0.125	1

Example

If X is the total number of heads in two tosses of a coin, what is the probability $X=1$? What is the probability $X \leq 1$?

Example

Consider the experiment of tossing 5 fair coins independently from one another, and define the random variable X = *"observed number of heads"*.

a) Define all possible outcomes from this experiment

b) Compute the probabilities of each outcome you defined in a)

c) Draw the p.d.f of the random variable X

d) Draw the c.d.f. of the random variable X

e) Does every outcome of the experiment correspond to exactly one value of the random variable?

Joint distributions and marginal distributions

The joint probability of A and B is Pr(A and B).

If X and Y are two discrete random variables, then they have a joint distribution with a joint probability density function. The joint pdf is the probability that $X = x_j$ and $Y = y_j$ for every particular value of x and y.

$f_{X,Y}(x,y) = P(X = x \text{ and } Y = y)$

When we are considering more than one variable, we refer to the probability distribution function of one variable as the **marginal distribution** $f(x)$ or $f(y)$.

Note $f(x) = \sum_y f(x,y)$ and $f(y) = \sum_x f(x,y)$

H. Conditional Probability Distributions (discrete)

H. Conditional Probability Distributions (discrete)

Recall the rules for conditional probability:

Pr(A|B) = "probability of A given B" = "probability of A conditional on B"
= the probability of an event A given that event B has occurred

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

It follows that:

$$P(A \text{ and } B) = P(A \mid B) \times P(B)$$

Similarly with distributions: the conditional p.d.f describes the p.d.f of Y given that $X = x_j$.

$$f_{Y|X}(y \mid x) = f(y \mid x) = P(Y = y \mid X = x)$$

$$f(y \mid x) = \frac{f(x,y)}{f(x)}$$

Also,

$$f(x \mid y) = \frac{f(x,y)}{f(y)}$$

Independence

Recall independence of events: A and B are independent events if knowing the outcome of A does not change the probability of B (and vice-versa).
$P(A \mid B) = P(A)$ if A and B are independent

For random variables, if X and Y are independent, then the p.d.f. of Y conditional on X is the same as the overall p.d.f. of Y since knowing information about X does not provide information about Y.

That is:

If X and Y are independent → $f(y \mid x) = f(y)$

Note that this has to be true for all values of X and Y

That is, information about the value of one variable gives no information about the probability of the value of another variable.

For independent events: $P(A \text{ and } B) = P(A) \times P(B)$

Independence

Random variables X and Y are independent if and only if:

$$f_{Y|X}(y \mid x) = f(y) \text{ for all } x \text{ and } y$$

$$f_{X|Y}(x \mid y) = f(x) \text{ for all } x \text{ and } y$$

Therefore, X and Y are independent if and only if:

$$f_{X,Y}(x,y) = f_X(x) \times f_Y(y) \text{ for all } x \text{ and } y$$

If X is independent of Y, then Y is independent of X.

X and Y are independent if knowing the outcome of X does not change the probabilities associated with different outcomes for Y and vice-versa.
In other words, no matter where we slice the distribution, we get the same p.d.f.

e.g. coin flips of a fair coin are independent

J. Continuous variables

Probability Functions For Continuous Random Variables

For a continuous variable, there are an infinite number of outcomes and the probability of any particular outcome is zero. Instead we consider the probability that X is between two points a and b. We can construct a probability density function, which is analogous to the p.d.f described above. Graphically, the probability is the area under the curve between points a and b.

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Note: Over the full range of values over which X is defined we expect the function to integrate to one; otherwise it is not a proper p.d.f.

Examples

Example: Uniform distribution from 0 to 12 (time on a clock)

$$f(x) = \begin{cases} \frac{1}{12} & \text{if } 0 \le x \le 12 \\ 0 & \text{otherwise} \end{cases}$$

What is the probability of time being between 3 and 5?

Another example of a continuous distribution: Standard normal p.d.f.

$$f(x) = \frac{1}{\sqrt{2\Pi}} e^{\frac{-x^2}{2}}$$

Cumulative distribution function

The **cumulative distribution function (c.d.f.)** describes the probability that the random variable X is less than or equal to a particular value x_j.

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

Note the c.d.f. is a non-decreasing function of x and $0 \le F(x) \le 1$. Also note that for a continuous variable $F(x_j) = P(X < x_j)$ as well. (Why?)

Example

If Z is distributed as a standard normal:

a) What is the probability that $Z<=1.96$?

b) What is the probability that $Z<=1$?

c) What is the probability that $Z<=0$?

Example

Suppose the random variable *X* has the following p.d.f.

$$f(x) = \begin{cases} ax^2 & \text{if } 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

a) What value of the constant *a* would make *f(x)* a “proper” p.d.f?

b) What is the Pr(1<*X*<2)?

c) What is the Pr(1<*X*)?

Joint distributions

In the case of continuous variables, we can represent the probability that x is between a and b and y is between c and d as the double integral:

$$P(a \leq x \leq b \text{ and } c \leq y \leq d) = \int_c^d \int_a^b f(x,y)\,dx\,dy$$

Example

Suppose X is a continuous random variable having the following p.d.f:

$$f(x) = \begin{cases} x & \text{if } 0 \leq x \leq 1 \\ 2 - x & \text{if } 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

a. Draw the p.d.f.

b. Calculate Pr (1<=X<=2)

K. Bayes rule

For two events A and B,

$$P(B \mid A) = \frac{P(B) \times P(A \mid B)}{P(A)}$$

$$P(B \mid A) = \frac{P(B) \times P(A \mid B)}{P(A \mid B) \times P(B) + P(A \mid \overline{B}) \times P(\overline{B})}$$

This is particularly useful for example if we are interested in updating our beliefs about the likelihood of an event after we observe additional information. For example, suppose a friend tells you she had a nice conversation with a stranger on the train. Without any other information, you would guess that there is 50% chance that the stranger is a woman. Now suppose you are told that this person had long hair. You can use Bayes' theorem to calculate the probability that the stranger is a women given that she has long hair.

Example

$$f(y \mid x) = \frac{f(y) \times f(x \mid y)}{f(x)}$$

In the previous example let P(Woman) = 0.5, and suppose you knew that 75% of women have long hair, that is P(Long hair | Woman) = 0.75, while 30% of men do: P(Long hair |Man) = 0.3. We are interested in finding P(Woman|Long Hair). Applying Bayes' rule we have:

$$P(\text{Woman} \mid \text{Long Hair}) = \frac{P(\text{Woman})P(\text{Long Hair} \mid \text{Woman})}{P(\text{Long Hair})}$$

$$= \frac{P(\text{Woman})P(\text{Long Hair} \mid \text{Woman})}{[P(\text{Long Hair} \mid \text{Woman})P(\text{Woman}) + P(\text{Long Hair} \mid \text{Man})P(\text{Man})]}$$

$$= \frac{.5(.75)}{[(.75)(.5) + (.3)(.5)]} = 0.714$$

Example

The probability that a beginning golfer makes a good shot if she selects the correct club is 1/3. The probability the shot is good with the wrong club is 1/5. In her bag are 4 clubs, of which one is the correct club. The golfer chooses a club at random and makes a shot.

a. What is the probability she makes a good shot?

b. Given that she makes a good shot, what is the probability that she chose the correct club?

L. Expected value and sample mean

Expected value (or population mean) of a discrete variable is a weighted average of the possible values of X, where the weights are the probability of each outcome.

$$
\begin{aligned}
&E(X) = \mu_X \\
&= X_1 f(X_1) + X_2 f(X_2) + X_3 f(X_3) + \ldots + X_k f(X_k) \\
&= \sum_{j=1}^{k} x_j f(x_j)
\end{aligned}
$$

For a continuous r.v. E(X) is an integral:

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

Example

Recall that a 12 hour clock has the following p.d.f.:

$$f(x) = \begin{cases} \frac{1}{12} & \text{if } 0 \leq x \leq 12 \\ 0 & \text{otherwise} \end{cases}$$

What is the expected value of the time on a 12-hour clock?

Properties of expected values

a) For constant *c*, *E(c)=c*.
b) For constants *a* and *b*, $E(aX+b)=aE(X)+b$.
c) $E\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} E(X_i)$ (The expected value of the sum is the sum of the expected value.)

As we noted above, the sample mean of a variable is the sum of the observed values of X divided by the number of observations, often called "X-bar".

$$\bar{X}=\tfrac{1}{n}\sum_{i=1}^{n} X_i$$

$\bar{X}$ is called the first moment of the sample.

Variance

If $E(X) = \mu_X$, then the **variance** is:

$$Var(X) = \sigma_X^2 = E[(X - \mu_X)^2]$$

The variance tells us how far X typically is from its mean. A large variance suggests that X often takes values far from its mean and the p.d.f. is wide.

Properties of variance

If a and b are constants,

$Var(a) = 0$

$Var(aX) = a^2\ Var(X)$

If X and Y are **independent** random variables, then $Var(X+Y) = Var(X)+Var(Y)$. It is also true that $Var(X\text{-}Y) = Var(X)+Var(Y)$ as long as they are independent.

$Var(aX + b) = a^2\ Var(X)$. (This follows from above rules)

$Var(X) = E(X\text{-}\mu)^2 = E(X^2) - [E(X)]^2$

For a sample, the variance is called s^2:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Example

1. The random variable Y has a mean of 1 and a variance of 4. Let Z=12(Y−1). Find the expected value and the variance of Z.

2. Let X and Y be random variables with p.d.f. given by:

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \qquad f(y) = \begin{cases} 1 & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

a. Draw f(y) and f(x)
b. Find E(Y) and E(X)
c. Find Var(Y) and Var(X)
d. Let Z=X+Y. Find E(Z) and Var(Z). Intuitively do you think that Z will have a uniform distribution?

Standard deviation

The **standard deviation of** a random variable X is the positive square root of the variance or the sample variance.

$$sd(X) = \sigma_X = +\sqrt{Var(X)}$$

$$sd(X) = s_X = +\sqrt{s^2}$$

It is a measure of the typical deviation from the mean.

Properties of standard deviation:

For constants a and b, $sd(aX + b) = |a|\, sd(X)$

For constant c, $sd(c) = 0$

Standardizing a random variable

Sometimes we will want to transform a r.v. into another r.v. with standard units

We can call Z the standardized random variable:

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

Example: What is the expected value and variance of Z?

$$E[Z] = E\left[\frac{X - \mu_X}{\sigma_X}\right] = \frac{1}{\sigma_X} E[X] - \frac{1}{\sigma_X}\mu_X = 0$$

$$Var[Z] = Var\left[\frac{X - \mu_X}{\sigma_X}\right] = E\left(\frac{X - \mu_X}{\sigma_X}\right)^2 - 0$$

$$= \frac{1}{\sigma_X^2}\left(E\left[X^2\right] - 2\mu_X E[X] + \mu_X^2\right)$$

$$= \frac{1}{\sigma_X^2}\sigma_X^2 = 1$$

Covariance

How do two random variables vary with one another?
For example, how does rainfall vary with grain output?
How does education vary with health?

The covariance indicates the degree to which two variables move together:

$$\operatorname{cov}(X,Y) = \sigma_{XY} = E\left[(X-\mu_X)(Y-\mu_Y)\right]$$