Chapter 4: Random Variables And Expectation

ST2334 Probability and Statistics[1]
(Academic Year 2014/15, Semester 1)

Department of Statistics & Applied Probability (DSAP)
National University of Singapore (NUS)[1]

Outline

Introduction I

Learning Outcomes

Questions to Address: ♦ What a random variable (r.v.) is ♦ Difference between/characterizations of discrete & continuous r.v.'s ♦ How to define jointly distributed r.v.'s ♦ What an expectation & its interpretation are ♦ Theoretical properties & results of expectations ♦ What a variance/covariance is ♦ What a moment generating function (mgf) is ♦ Determination of distribution/kth moment by mgf ♦ How to obtain mgf of a r.v./a linear transformation of a r.v./a sum of independent r.v.'s

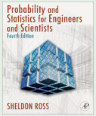
Introduction II

Learning Outcomes (continued)

`Concept & Terminology:` ♦ numerical events ♦ discrete/continuous r.v. ♦ probability mass/density function ♦ cumulative distribution function ♦ Bernoulli/uniform r.v. ♦ expected value/expectation/mean ♦ long–run average ♦ expectation of a function of a r.v./a function of ≥ 2 r.v.'s/a linear combination of r.v.'s/≥ 2 independent r.v.'s ♦ variance & standard deviation ♦ variability/spread of a r.v. ♦ variance of a linear transformation of a r.v./a sum of independent r.v.'s ♦ covariance of 2 r.v.'s ♦ covariance of 2 sums of r.v.'s ♦ moment generating function (mgf) ♦ kth moment ♦ mgf of a linear transformation of a r.v./a sum of independent r.v.'s

Introduction III

Mandatory Reading



✏ Section 4.1 – Section 4.8 (except Section 4.3.2)

Why do we study Random Variables?

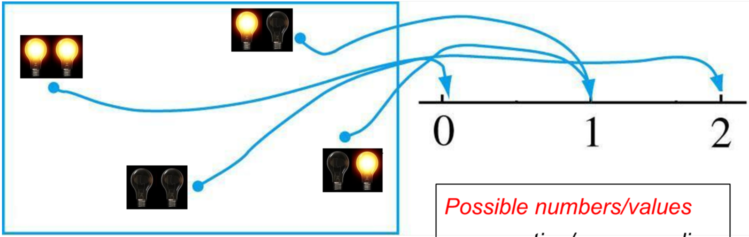
Why Study Random Variables? I

- *Usually*, scientists, businessmen, or engineers, are interested in *characteristics* in different experiments defined by events that are in terms of numbers, called *numerical events*, though the sample space S may consist of *qualitative* or *quantitative* outcomes
  1. Manufacturers: proportion of defective light bulbs from a lot (*qualitative* outcomes from {, } for each light bulb)
  2. Market researchers: preference in a scale of 1–10 (*quantitative*) of consumers about a proposed product
  3. Research physicians: changes in certain reading (*quantitative*) from patients who are prescribed a new drug
  4. Singapore Pool: how much revenue is generated (depending on different experiments including Singapore Sweep, Toto, 4D, football games & so on, with *qualitative* outcomes in terms of different combinations of numbers, football match results & so on)

Random Variable: A unified framework to handle them all!

Why Study Random Variables? II

- A rule of association is needed to associate *all outcomes in S* with the *possible numbers/values representing/corresponding to the events of interest*
- **Examine 2 light bulbs**: We are interested in the *# of defective bulbs* $(= 0, 1, 2)$, but not whether a specific bulb is defective or not. Such a rule is illustrated below:



Sample space S

Possible numbers/values representing/corresponding to the events of interest

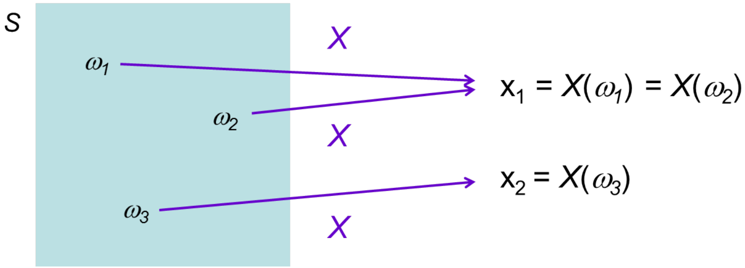
Why Study Random Variables? III

- A *random variable* serves as such a *rule of association*
  - *a variable*: takes on different possible numerical values
  - *random*: different values depending on which outcome occurs
- When an experiment is conducted, an outcome (subject to uncertainty) is *realized* $\Rrightarrow$ a corresponding *"random" value* is *realized* according to the *rule*
  1. Manufacturers: the # of defective light bulbs from a lot of certain fixed size N is any integer from 0 to N
  2. Market researchers: preference score of consumers about the proposed product is any integer from 1 to 10
  3. Research physicians: change in certain reading from a patient who is prescribed the new drug is any *real #* from $-\infty$ to ∞
  4. Singapore Pool: the revenue is some real # (but probably can never be a negative #, right?)

Definition

For a given sample space S of an experiment, a *random variable (r.v.)* is a function X whose *domain* is S & whose *range* is the set of real #'s

$$X : S \to \mathbb{R}$$



Random Variables II

- *Uppercase* letters $X, Y, Z, \ldots$ to denote r.v.'s
- *Lowercase* letters $x, y, z, \ldots$ to denote their possible values
- *e.g.,* in the example of **Examine 2 light bulbs**:
  - Let X denote the # of defective bulbs
  - *Domain*: $S = \{$, , , $\}$
  - X is a r.v. defined by
    - $X($$) = 0$
    - $X($$) = X($$) = 1$
    - $X($$) = 2$

    $\Rrightarrow$ *X takes on values $x = 0, 1, 2$* $\Leftrightarrow$ *Range*: $\{0, 1, 2\}$
  - **Note**: When this experiment is conducted, a value x from $\{0, 1, 2\}$ would be *realized*, say, $X = 0$, governed by the magnitudes of $P(X = 0)$, $P(X = 1)$ & $P(X = 2)$ where P is the *same function* defined at Page 19 (Ch.3)
- *Mainly 2 types of r.v.'s*: Discrete versus Continuous

Definition

A *discrete r.v.* is a r.v. that can take on only a finite or at most a countably infinite # of values

- *Range* is
  - *finite* (*e.g.,* $\{0, 1, 2\}$ in the example of **Examine 2 light bulbs**), or
  - *countably infinite* (*e.g.,* Range $= \{0, 1, 2, \ldots\}$ in concerning # of wrong phone calls received in a day)
- To *characterize* (*i.e.,* fully understand) X, one has to specify the *probs attached to each value in its range*

What is $P(X = x)$ for all possible values x?

- **Note**: *All the discussion & properties related to P* (in terms of $P(A)$) in `Ch.3` *apply to probs of any r.v.*

Probability Mass Function

- Let $A = \{\omega \in S : X(\omega) = x\}$ be the set/event containing all the outcomes $\omega \in S$ which are mapped to any value x by X
- Following from equation (1) at Page 23 (Ch.3), for any $x \in \mathbb{R}$,

$$P(X = x) \equiv P(A) = \sum_{\{\omega \in S : X(\omega) = x\}} P(\{\omega\})$$

$\Rrightarrow$ **Simply add probs of all $\omega \in A$**

Definition

The *probability mass function (pmf)* of a *discrete r.v.* with range $\{x_1, x_2, \ldots\}$ is a function p s.t. $p(x_i) = P(X = x_i)$ & $\sum_i p(x_i) = 1$

Example 1: Discrete r.v.'s, & pmf I

Toss 3 fair coins:

- $S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$
- Let X be the total # of heads observed
- X is a *discrete* r.v. taking on $\{0, 1, 2, 3\}$ defined by

$$X(HHH) = 3; X(HHT) = X(HTH) = X(HHT) = 2;$$
$$X(HTT) = X(THT) = X(TTH) = 1; X(TTT) = 0$$

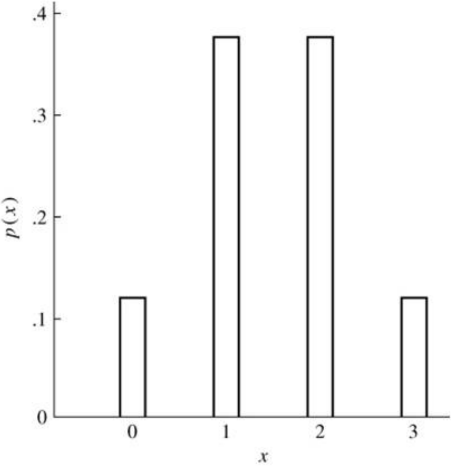
- $P(X = 3) = P(\{HHH\}) = 1/8$
  $P(X = 2) = P(\{HHT, HTH, THH\}) = 3/8$
  $P(X = 1) = P(\{HTT, THT, TTH\} = 3/8$
  $P(X = 0) = P(\{TTT\}) = 1/8$

Example 1: Discrete r.v.'s, & pmf II

- The pmf of X is given by

$$p(x) = \begin{cases} .125, & x = 0, 3 \\ .375, & x = 1, 2 \\ 0, & \text{otherwise} \end{cases}$$

- The length/height of the vertical bar at any point x_i is $p(x_i)$, with the sum of all the bar's height as 1
- Refer to the set $\{0, 1, 2, 3\}$ as the *support of the pmf*
  - the values *at which $p(x)$ is non–zero*
  - *identical to the range of X*



Example 2: The Bernoulli r.v. I

- *Bernoulli trial*: an experiment whose outcomes can be classified as, generically "*success (S)*" or "*failure (F)*"
- A function X which maps S to 1 & F to 0 defines the **simplest discrete r.v.** which takes on only 2 possible values

Definition

A *Bernoulli r.v.* with *parameter or probability of success* $0 < p < 1$ takes on only 2 values, 0 & 1, with $p(1) = p = 1 - q = 1 - p(0)$

- Write $X \sim Ber(p)$
- *pmf* of X:

$$p(x) = \begin{cases} p^x q^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

Example 2: *The Bernoulli r.v. II*

- Different examples of Bernoulli trials include: ❶ **Toss a coin**; ❷ **Examine a light bulb**; ❸ **Whether it rains in NUS today**; ❹ **Clear ST2334 by the end of the semester or not**
- **Win the 1st prize in the next draw of 4D**: Make a single bet on the # 2222. Define a r.v. X by

$$X = \begin{cases} 1, & \text{if the 4–digit \# drawn for the 1st prize is 2222} \\ 0, & \text{otherwise} \end{cases}$$

Then, *e.g.,* if the 4–digit # drawn for the 1st prize is 2345, we have $X(2345) = 0$, whereas $X(2222) = 1$
$\because$ $P(\text{the 4–digit \# drawn for the 1st prize is 2222}) = 10^{-4}$
$\therefore$ X is a *Bernoulli r.v. with parameter/prob of success* 10^{-4}, or simply, write $X \sim Ber(10^{-4})$

Continuous Random Variables I

- Among real world "experiments", many "natural" variables/quantities of interest whose set of possible values is *uncountable*, *e.g.*,
  1. temperature range on any day
  2. amount of annual income via export in a country
  3. lifetime of a light bulb
  4. height of a newborn baby
  5. weight loss after exercise
  6. amount of distance traveled by a car before breakdown
- Based on the graph at Page 9, define a *one–to–one* function X which maps these uncountably infinite # of outcomes in S onto exactly themselves, respectively
- Viewing X as a r.v., the *range of X* is an *interval* (possibly unbounded) or *a union of intervals*, & this kind of r.v.'s are called

continuous (cont.) random variables

Continuous Random Variables II

In case we follow what we did in discussing discrete r.v.'s, we would have to look at $P(X = x)$ for every possible value of x in $\mathbb{R}$ of the cont. r.v. X

Impractical & Illegitimate to talk about $P(X = x)$ for every x!

- *Uncountably infinite* # of x or $P(X = x)$ to be dealt with!
- *For any cont r.v.'s*, we must have, for all possible value of x,

$$P(X = x) = 0$$

due to $P(S) = \sum_{\text{all } x} P(X = x) = 1$ (**Axiom ❶ of prob**, Ch. 3)

Probability Density Function I

The idea: Replace the pmf (or the graph similar to the one at `Page 14`) by *another function* called the *probability density function (pdf)* (or, equivalently, another graph)

	pmf for Discrete r.v.'s	*pdf* for Cont. r.v.'s
Defined on (*i.e.*, support)	Finite/countably infinite # of points	A continuum of values
prob	Height/length of the bar at each possible value	Area under the function between 2 points
The Total prob $= 1$	Sum of height of all the bars	Area under the function over all possible values
Computation	Addition/Subtraction	Integration

Remark: In fact, for brevity, we sometimes refer to both pmf & pdf as *density* in a unified manner

Probability Density Function II

Definition

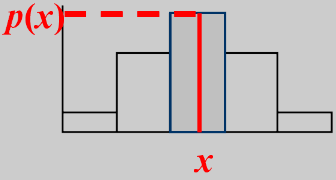
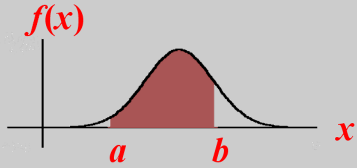
The *probability density function (pdf)* of a cont. r.v. X is an integrable function $f : \mathbb{R} \to [0, \infty)$ satisfying

1. $f(x) \geq 0$ for any $x \in \mathbb{R}$
2. f is *piecewise cont.*
3. $\int_{-\infty}^{\infty} f(x)\,dx = 1$
4. The prob that X takes on a value in the interval $(a, b) \subset \mathbb{R}$ equals the *area under the curve f between a & b*:

$$P(X \in (a, b)) = P(a < X < b) = \int_a^b f(x)\,dx$$

- *The value of $f(x) \neq$ prob $P(X = x)$*
- **Note**: For brevity, use *density* to *refer to both pmf/pdf of any r.v.*

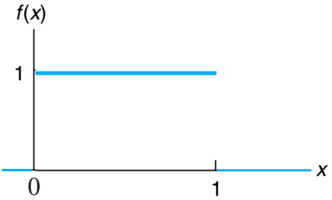
Comparison Between pmf & pdf

	RANDOM VARIABLE, X	
Type	**Discrete**	**Continuous**
Values	A finite/countable set of numbers $x_1, x_2, x_3, \ldots$	All numbers in an interval
Probability	Probability Mass Function, p ***pmf*** $P(X = x) = p(x)$ 	Probability Density Function, f ***pdf*** $P(a < X < b) = \left[\begin{array}{c}\text{area} \\ \text{under the} \\ \text{graph of } f \\ \text{over } (a,b)\end{array}\right]$ 

Example 3: *Uniform r.v.*

The **simplest cont. r.v.** is a *uniform/standard uniform r.v.*, which has a *range as the interval* $[0, 1]$, & is defined by a *uniform/rectangular pdf*

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



or, in terms of the *support of f*,

$$f(x) = 1, \qquad 0 \le x \le 1$$

- Write $X \sim U(0, 1)$
- The *prob that X is in any interval of length h* is equal to h
- A *generalization*: a *uniform r.v. on any interval* $[\alpha, \beta] \subset \mathbb{R}$, $X \sim U(\alpha, \beta)$, is defined by

$$f(x) = \frac{1}{\beta - \alpha}, \qquad \alpha \le x \le \beta$$

Cumulative Distribution Function

Definition

The *cumulative distribution function (cdf)* (or simply distribution function) of *any* r.v. X is defined by

$$F(x) = P(X \leq x) = P(X \in (-\infty, x]), \qquad -\infty < x < \infty,$$

which must

1. be always *non–decreasing*
2. satisfy $\lim_{x \to -\infty} F(x) = 0$ & $\lim_{x \to \infty} F(x) = 1$

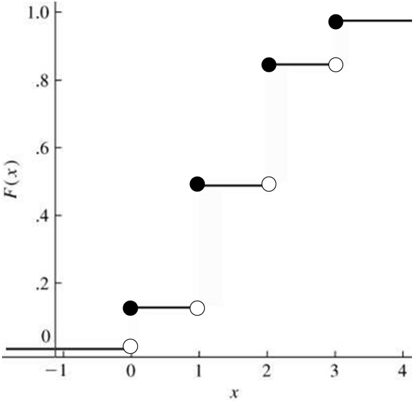
- Both pmf/pdf & cdf are *necessary & sufficient* ways to *uniquely characterize a r.v.*, but it is more common to use pmf/pdf in practice
- $$F(x) = \begin{cases} \displaystyle\sum_{\{x_i : x_i \leq x\}} p(x_i), & \text{when } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{x} f(t)\, dt, & \text{when } X \text{ is cont.} \end{cases}$$

Example 4: cdf of a Discrete r.v.

- The cdf of the discrete r.v. X in Example 1 is given by

$$F(x) = \begin{cases} 0, & x < 0 \\ .125, & 0 \le x < 1 \\ .5, & 1 \le x < 2 \\ .875, & 2 \le x < 3 \\ 1, & x \ge 3 \end{cases}$$

  - stays at 0 from $x = -\infty$ until 0 (the smallest possible value of X)
  - at 0, jumps by $P(X = 0) = .125$
  - stays at .125 from $x = 0$ until 1 (the next possible value of X)
  - at 1, jumps by $P(X = 1) = .375$
  - ...
  - stays at 1 from $x = 3$ (the largest possible value of X) until ∞



Some More Properties of Cont. r.v.'s

More Properties of Continuous Random Variables

1. $P(X = c) = \int_c^c f(x)dx = 0$ for any $c \in \mathbb{R}$
2. $P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$
   $= P(X \leq b) - P(X < a) = F(b) - F(a)$
3. For small $\delta > 0$, if f is cont. at c,
   $$P\left(c - \frac{\delta}{2} \leq X \leq c + \frac{\delta}{2}\right) = \int_{c-\frac{\delta}{2}}^{c+\frac{\delta}{2}} f(x)\,dx \approx \delta f(c)$$
4. $F'(x) = \dfrac{d}{dx}F(x) = f(x)$

- From **Property ❸**, the prob that X is in a small interval around c is *proportional to $f(c)$*
  $\Rrightarrow$ X more likely takes on a value around c rather than d if $f(c) > f(d)$

Example 5: Cont. r.v.'s, pdf & cdf

Suppose that a r.v. X has a pdf given by $f(x) = \begin{cases} Cx^{-2}, & x > 10 \\ 0, & x \leq 10 \end{cases}$. Find

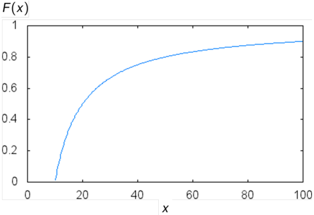
❶ the value of C, ❷ the cdf of X, & ❸ $P(X > 15)$

Solution: First of all, the *support of the pdf* is $S = [10, \infty)$

❶ Since f is a pdf, we must have $\int_{-\infty}^{\infty} f(x)\, dx = 1$, *i.e.*,

$$\int_{-\infty}^{10} 0\, dx + \int_{10}^{\infty} Cx^{-2}\, dx = C\left[\frac{x^{-2+1}}{-2+1}\right]_{10}^{\infty} = 1 \quad \Rightarrow \quad C = 10$$

❷ $$F(x) = \begin{cases} 10\int_{10}^{x} t^{-2}\, dt = 10\left[-t^{-1}\right]_{10}^{x} \\ \qquad\qquad = 1 - \frac{10}{x}, & x \geq 10 \\ 0, & x < 10 \end{cases}$$



❸ $P(X > 15) = 1 - P(X \leq 15) = 1 - F(15)$
$$= 1 - \frac{1}{3} = \frac{2}{3}$$

Jointly Distributed Random Variables

- Obviously, it is possible to define *as many r.v.'s* as we want from any *single experiment* through *different rules of association*
- *1 common situation*: For every experimental unit in the experiment, measurements of ≥ 2 attributes are recorded, *e.g.,*
  1. any investor is interested in the return of his/her investments in a portfolio that contains several stocks & funds in both local & overseas markets
     - *Random returns* of the individual derivatives: $X_1, X_2, \ldots, X_N$
     - For risk management: How $X_1, X_2, \ldots, X_N$ vary simultaneously?
     - What is the total return $Y = v(X_1, \ldots, X_N) = X_1 + \ldots + X_N$ like?
  2. observe height X & weight Y of new–born babies; we may be interested in determining *a relationship between the 2 characteristics*
     - There may be some pattern between height & weight that can be described by an "appropriately chosen" function u s.t. $Y = u(X)$
- To characterize these r.v.'s *simultaneously* (*i.e.,* understand the relationship between these r.v.'s), we look at the *joint probability distribution* of the many r.v.'s

Joint Cumulative Distribution Function I

Definition

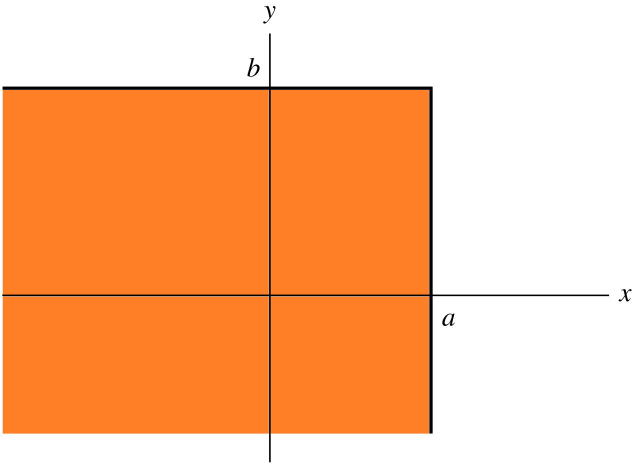
The *relationship between any 2 r.v.'s*, X & Y, is determined by the *joint cumulative distribution function (joint cdf)*:

$$F(a, b) = P(X \leq a, Y \leq b), \qquad -\infty < a, b < \infty$$

- $F(\cdot, \cdot)$ is a function from $\mathbb{R}^2 \equiv \mathbb{R} \times \mathbb{R}$ to $[0, 1]$
- $\{X \leq a, Y \leq b\} \equiv \{X \leq a\} \cap \{Y \leq b\}$: The "," inside the former set stands for *intersection*
- *To get the RHS of F*:
  1. Write $\{X \leq a\} = \{\omega \in S : X(\omega) \leq a\}$ & $\{Y \leq b\} = \{\omega \in S : Y(\omega) \leq b\}$
  2. Take the intersection of the 2 sets of ω's
  3. Add probs of all the common ω's from the intersection

Joint Cumulative Distribution Function II

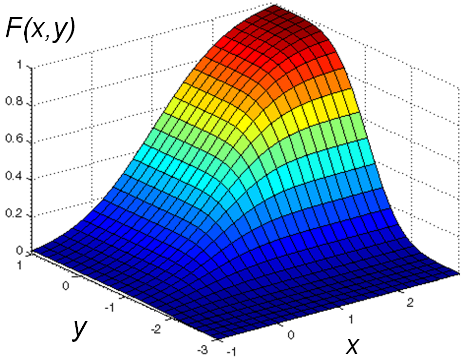
- $F(a, b)$ gives the prob of *all the common outcomes ω in the intersection* that are mapped by X & Y onto the region in orange



Joint Cumulative Distribution Function III

- Instead of a 2–d graph (at Page 24 or 26) for plotting the cdf of 1 r.v., the joint cdf F is plotted visually as a *3–d graph* with
  1. X–axis: values of X
  2. Y–axis: values of Y
  3. Z–axis (*i.e.,* the vertical axis): values of $F(x, y)$ for different x & y

  An example of a plot of *a joint cdf of 2 cont. r.v.'s*:



Definition

Let X & Y be *2 discrete r.v.'s* defined on the same sample space S. The *joint probability mass function*, $p(x, y)$, is defined for each pair $(x, y) \subset \mathbb{R}^2$ by

$$p(x, y) = P(X = x, Y = y) \tag{1}$$

- $0 \leq p(x, y) \leq 1$ & $\sum_x \sum_y p(x, y) = 1$
- *To get the RHS of $p(x, y)$*:
  1. Write $\{X = x\} = \{\omega \in S : X(\omega) = x\}$ & $\{Y = y\} = \{\omega \in S : Y(\omega) = y\}$
  2. Take the intersection of the 2 sets of ω's
  3. Add probs of all the common ω's from the intersection

Joint Probability & Marginal pmf

- In (1), only 1 pair (x, y) is of interest. However, usually of interest could be *a collection of (x, y)-pairs* represented by *some sets $C \subset \mathbb{R}^2$ in the xy-plane*
  e.g., $C = \{(x, y) \in \mathbb{R}^2 : x + y = 5\}$ or $C = \{(x, y) \in \mathbb{R}^2 : x/y \leq 3\}$

Probability of Any Event

Let $C \subset \mathbb{R}^2$ be any set consisting of pairs of (x, y) values. Then,

$$\underline{P((X, Y) \in C)} = \sum_{(x,y)\in C} p(x, y)$$

Marginal pmf

The *marginal probability mass functions* of X & of Y, denoted by p_X & p_Y, can be recovered from the joint pmf as

$$p_X(x) = P(X = x) = \sum_y p(x, y) \quad \& \quad p_Y(y) = P(Y = y) = \sum_x p(x, y)$$

Example 6: Joint Distribution of 2 Discrete r.v.'s I

Consider Example 1 about tossing 3 fair coins again

- Define the r.v.'s $\begin{cases} X &= \text{\# of tails in the first 2 tosses} \\ Y &= \text{\# of heads in all the 3 tosses} \end{cases}$

  $\Rrightarrow$ X takes on $0, 1, 2$, & Y takes on $0, 1, 2, 3$ (*i.e.*, Y is exactly the same r.v. X in Example 1)
- Mappings from all outcomes $\omega \in S$ to the values of X & of Y

ω	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$X(\omega)$	0	0	1	1	1	1	2	2
$Y(\omega)$	3	2	2	2	1	1	1	0

- Hence, we have

$$\begin{aligned} p(0,2) &= P(X=0, Y=2) = P(\{HHT\}) = 1/8 \\ p(0,3) &= P(X=0, Y=3) = P(\{HHH\}) = 1/8 \\ p(1,1) &= P(X=1, Y=1) = P(\{HTT, THT\}) = 2/8 \\ p(1,2) &= P(X=1, Y=2) = P(\{HTH, THH\}) = 2/8 \end{aligned}$$

Example 6: *Joint Distribution of 2 Discrete r.v.'s II*

$$
\begin{aligned}
p(2,0) &= P(X=2, Y=0) = P(\{TTT\}) = 1/8 \\
p(2,1) &= P(X=2, Y=1) = P(\{TTH\}) = 1/8 \\
p(x,y) &= 0, \qquad \text{otherwise}
\end{aligned}
$$

- The joint pmf $p(x,y)$ of X & Y in a tabular form:

	y			
x	0	1	2	3
0	0	0	1/8	1/8
1	0	2/8	2/8	0
2	1/8	1/8	0	0

- From the above table, *e.g.*, $P(X < 2, Y \geq 1) = 3/4$ & $P(X = Y) = 1/4$
- *Summing up the columns/rows* in the above table gives p_X/p_Y

Definition

The *joint probability density function* of *2 cont. r.v.'s* X & Y is an *integrable* function $f : \mathbb{R}^2 \to [0, \infty)$ satisfying

1. $f(x, y) \geq 0$ for any $(x, y) \in \mathbb{R}^2$
2. $f(x, y)$ is a *piecewise cont.* function of x & of y
3. $P((X, Y) \in \mathbb{R}^2) = 1$
4. The prob that (X, Y) takes on a pair (x, y) in any set $C \subset \mathbb{R}^2$ equals the *volume of the object/space over the region C & bounded above by the surface $z = f(x, y)$*, which is expressible as a *double integral*:

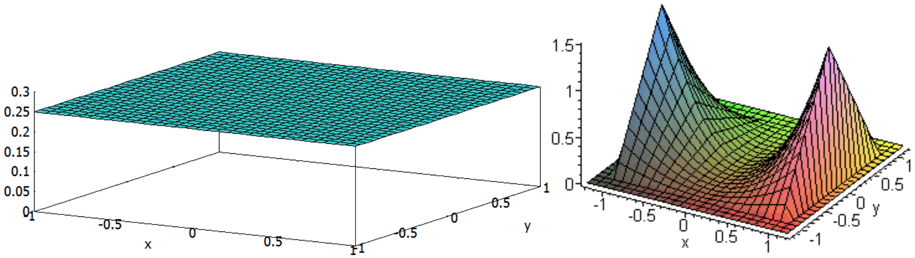
$$P((X, Y) \in C) = \int_C \int f(x, y)\, dy\, dx = \int_C \int f(x, y)\, dx\, dy \qquad (2)$$

- The *simplest example* of a joint pdf, f, is a *constant function*, which is equivalent to

  A *flat surface* on a bounded set $S \subset \mathbb{R}^2$ (plotted as below on the left)

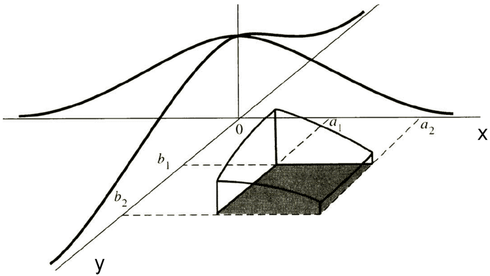
  where S is called *the support of f* (same as that in Examples 1 & 5 for a pmf/pdf)

- Plots of 2 *joint pdf's $f(x, y)$ of 2 cont. r.v.'s* (at the right: f is not a constant function):



Computing Joint Probability $P((X,Y) \in C)$ *I*

- **Ideally** when
  1. $S = \mathbb{R}^2$
  2. $C = \{(x,y) \in \mathbb{R}^2 : a_1 \le x \le a_2, b_1 \le y \le b_2\}$, where a_1, a_2, b_1, b_2 are some *fixed constants*, is the rectangle in black in the xy-plane below



Following **property** ❹ at Page 35,

$$P((X,Y) \in C) = P(a_1 \le X \le a_2, b_1 \le Y \le b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f(x,y)\, dy\, dx$$

Computing Joint Probability $P((X,Y)\in C)$ II

- But, **commonly**, $S \subset \mathbb{R}^2 \Leftrightarrow f(x,y) = \begin{cases} g(x,y) \neq 0, & (x,y) \in S \\ 0, & (x,y) \in \mathbb{R}^2 \backslash S \end{cases}$

  Following **property ❹** at Page 35, $P((X,Y)\in C)$ equals

$$\begin{aligned}
&\int_{C\cap\mathbb{R}^2}\int f(x,y)\,dy\,dx = \int_{C\cap(S\cup(\mathbb{R}^2\backslash S))}\int f(x,y)\,dy\,dx \\
&= \int_{C\cap S}\int f(x,y)\,dy\,dx + \int_{C\cap(\mathbb{R}^2\backslash S)}\int f(x,y)\,dy\,dx \\
&= \underline{\int_{C\cap S}\int g(x,y)\,dy\,dx} + \int_{C\cap(\mathbb{R}^2\backslash S)} 0\,dy\,dx
\end{aligned}$$

$$\boxed{P((X,Y)\in C) = \int_{C\cap S}\int g(x,y)\,dy\,dx} \tag{3}$$

Computing Joint Probability $P((X, Y) \in C)$ *III*

Evaluating a Double Integral

For any set $D \subset \mathbb{R}^2$ & a known function $g(x, y) \neq 0$, a *double integral*

$$\int_D \int g(x, y)\, dy\, dx \tag{4}$$

is evaluated by

1. integrating over *all possible values of* y in the inner integral holding x (the variable in the outer integral) *as a constant at each of its possible value*
2. integrating over all possible values of x in the outer integral

Computing Joint Probability $P((X, Y) \in C)$ *IV*

- **Our goal**: Find *some expressions, a_1, a_2, b_1, b_2, as the limits* such that (4) becomes

$$\int_D \int g(x, y)\, dy\, dx = \int_{a_1}^{a_2} \int_{b_1}^{b_2} g(x, y)\, dy\, dx \tag{5}$$

so that we can *evaluate* ❶ the inner integral in y by treating x as a constant & then ❷ the outer integral in x

- **In practice**, *plot the domain of integration D* on the xy-plane to
  - ❶ locate all *possible values of x as $a_1 \leq x \leq a_2$* from D
  - ❷ locate all *possible values of y as $b_1 \leq y \leq b_2$* from D with *x fixed at each of its possible value on $[a_1, a_2]$*

If ❷ fails, *break the interval (a_1, a_2) for x in ❶ into 2 or more pieces* so that the RHS of (5) becomes a sum of 2 or more double integrals

Marginal cdf's & Marginal pdf's

Marginal Cumulative Distribution Functions

The *marginal cdf*'s of X & of Y, denoted by F_X & F_Y, equal

$$F_X(x) = F(x, \infty) \qquad \& \qquad F_Y(y) = F(\infty, y)$$

Marginal Probability Density Functions

The *marginal pdf*'s of X & of Y, denoted by f_X & f_Y, are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \qquad \& \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

- **Note**: When $S \subset \mathbb{R}^2$, the latter 2 integrals become
  1. $\int_{b_1}^{b_2} g(x, y)\, dy$: $[b_1, b_2]$ gives all possible values of y when x is a fixed value in S
  2. $\int_{a_1}^{a_2} g(x, y)\, dx$: $[a_1, a_2]$ gives all possible values of x when y is a fixed value in S

Example 7: Joint pdf & Computing Joint Probability I

Consider

$$f(x,y) = \begin{cases} 24xy, & 0 \le x \le 1,\ 0 \le y \le 1,\ x+y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

1. $S = \{(x,y) \in \mathbb{R}^2 : 0 \le x \le 1,\ 0 \le y \le 1,\ x+y \le 1\}$ is the support of f
2. We can *show that f is a proper joint pdf* as follows:
   - Clearly, $f(x,y) \ge 0$ for $-\infty < x, y < \infty$, & f is piecewise cont.
   - To verify that *the double integral of f over $\mathbb{R}^2$ is 1*, it involves computing the *double integral*

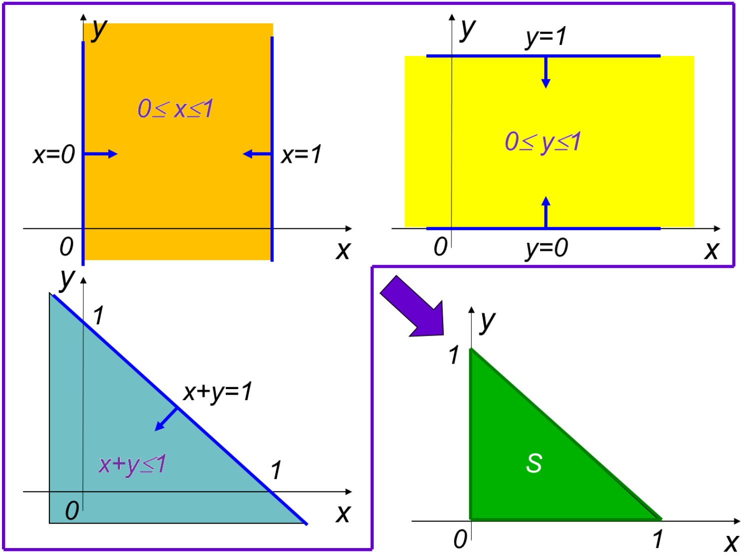
$$\int_{\mathbb{R}^2}\int f(x,y)\,dy\,dx = \int_S\int f(x,y)\,dy\,dx + \int_{\mathbb{R}^2\setminus S}\int f(x,y)\,dy\,dx$$

$$= \underline{\int_S\int 24xy\,dy\,dx} + \int_{\mathbb{R}^2\setminus S} 0\,dy\,dx$$

Refer to (4), we have $g(x,y) = 24xy$ & $D = S$

$\Rrightarrow$ Next, draw the *domain of integration S* (at the next page)

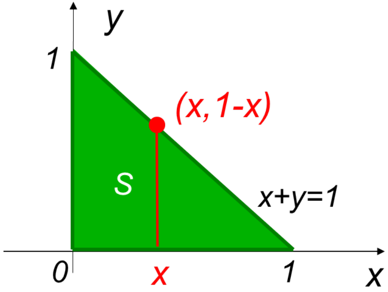
Example 7: Joint pdf & Computing Joint Probability II

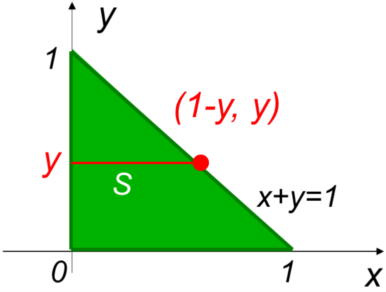


Example 7: *Joint pdf & Computing Joint Probability III*

$$\int_S\int 24xy\,dy\,dx = 24\int_0^1\left[\int_0^{1-x} xy\,dy\right]dx = 24\int_0^1 x\left[\frac{y^2}{2}\right]_0^{1-x} dx$$

$$= 12\int_0^1 x(1-x)^2\,dx = 12\int_0^1 (x-2x^2+x^3)\,dx = 1$$

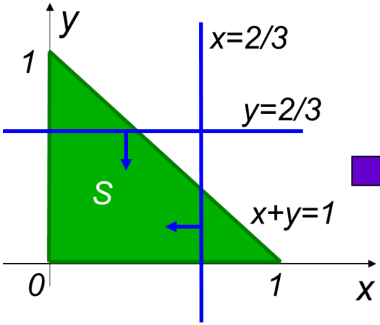


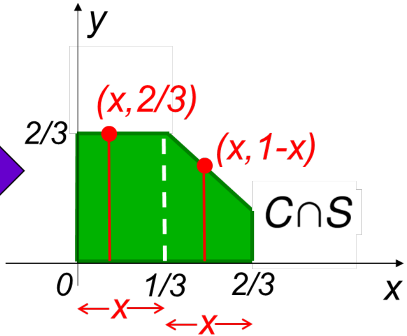


Alternatively, $\int_S\int 24xy\,dx\,dy = 24\int_0^1\left[\int_0^{1-y} xy\,dx\right]dy = 1$

Example 7: *Joint pdf & Computing Joint Probability IV*

3. Now, suppose that we are interested in computing the prob that *both X & Y take on values less than* $2/3$, *i.e.*, $P((X, Y) \in C)$ where $C = \left\{(x, y) \in \mathbb{R}^2 : x < 2/3, y < 2/3\right\}$
   - From (3), it suffices to compute the *double integral (4)* with $g(x, y) = 24xy$ & $D = C \cap S$ represented by the **green pentagon** below at the right





Example 7: *Joint pdf & Computing Joint Probability V*

- Unfortunately, it is *impossible to assign unique limits for the inner integral when we consider all x values from 0 to 2/3*
  $\Rrightarrow$ *Split $C \cap S$* (*i.e.,* the **green pentagon**) *into 2 pieces wrt x* from 0 to 1/3 & from 1/3 to 2/3 (*i.e.,* the **green rectangle** & the **green trapezium**) in order to *obtain unique (& different) limits for the inner integrals* over the 2 resulting regions

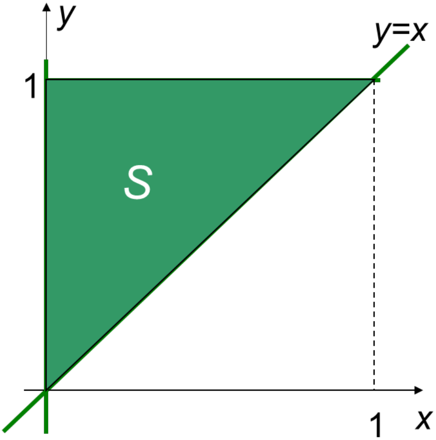
$$
\begin{aligned}
P\left(X < \frac{2}{3}, Y < \frac{2}{3}\right) &= \int_{C \cap S}\int 24xy\, dy\, dx = \int\int 24xy\, dy\, dx \\
&= \int\int 24xy\, dy\, dx + \int\int 24xy\, dy\, dx \\
&= \int_0^{1/3}\left[\int_0^{2/3} 24xy\, dy\right] dx + \int_{1/3}^{2/3}\left[\int_0^{1-x} 24xy\, dy\right] dx \\
&= \int_0^{1/3} \frac{16}{3} x\, dx + \int_{1/3}^{2/3} 12x(1-x)^2\, dx \\
&= \frac{8}{27} + \left[6x^2 - 8x^3 + 3x^4\right]_{1/3}^{2/3} = \frac{7}{9}
\end{aligned}
$$

Example 8: *Joint Probability & Marginal pdf's I*

Suppose that the joint pdf of X & Y is defined by

$$f(x,y) = \begin{cases} 6(1-y), & 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

1. What is $P\left(X \le \frac{3}{4}, Y \ge \frac{1}{2}\right)$?
2. Find the marginal pdf's



Example 8: Joint Probability & Marginal pdf's II

To compute the required prob, draw the *domain of integration* $C \cap S \equiv A \cup B$ of the double integral $\int_{C\cap S} 6(1-y)\,dy\,dx$ as in the graph below (where $C = \{(x,y) \in \mathbb{R} : x < 3/4, y < 1/2\}$):

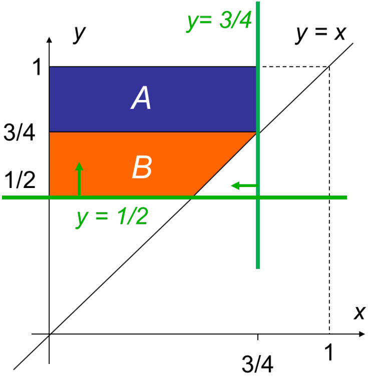
$$\int_A \int 6(1-y)\,dx\,dy + \int_B \int 6(1-y)\,dx\,dy$$

$$= \int_{3/4}^{1} \int_0^{3/4} 6(1-y)\,dx\,dy$$

$$+ \int_{1/2}^{3/4} \int_0^{y} 6(1-y)\,dx\,dy$$

$$= \int_{3/4}^{1} \frac{9}{2}(1-y)dy + \int_{1/2}^{3/4} 6(y-y^2)dy$$

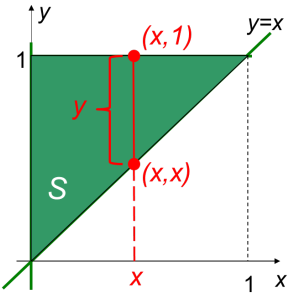
$$= \frac{9}{2}\left[y - \frac{y^2}{2}\right]_{3/4}^{1} + \left[3y^2 - 2y^3\right]_{1/2}^{3/4} = \frac{31}{64}$$



Example 8: Joint Probability & Marginal pdf's III

For $0 < x < 1$ (*i.e.,* all possible values of X from the support S),

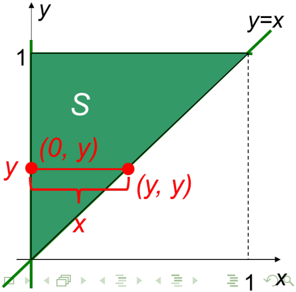
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{x}^{1} 6(1-y)\,dy$$
$$= \left[6y - 3y^2\right]_1^x = 3(1-x)^2$$



Note: Fix x at a certain value beforehand

For $0 < y < 1$ (*i.e.,* all possible values of Y from the support S),

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{0}^{y} 6(1-y)\,dx$$
$$= 6y(1-y)$$



Generalization to ≥ 3 *Cont. r.v.'s*

For $m \geq 3$ cont. r.v.'s, $X_1, X_2, \ldots, X_n$, defined on the same S:

- *Joint pdf*: $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \dfrac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{X_1,\ldots,X_m}(x_1,\ldots,x_n)$
- *1–dim marginal pdf*'s expressible as $(n-1)$–folded integrals:
$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\, dx_1 \cdots dx_{i-1}\, dx_{i+1} \cdots dx_n$$
for all i
- *2–dim joint pdf*'s expressible $(n-2)$–folded integrals: *e.g.*,
$$f_{X_1,X_2}(x_1,x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\, dx_3 \cdots dx_n$$
- *3–dim*, …, *$(n-1)$–dim joint pdf*'s similarly defined
- *Joint probability* for any $C \subset \mathbb{R}^n$ expressible as an m–folded integral:
$$P((X_1,\ldots,X_n) \in C) = \int_C \cdots \int f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\, dx_1 \cdots dx_n$$

Independent Random Variables I

Definition

Random variables $X_1, \ldots, X_n$ are said to be *independent (indept)* if their *joint cdf* factors into the product of their marginal cdf's:

$$F(x_1, \ldots, x_n) = F_{X_1}(x_1) \times F_{X_2}(x_2) \times \cdots \times F_{X_n}(x_n)$$

for all values $x_1, \ldots, x_n \in \mathbb{R}$; otherwise $X_1, \ldots, X_n$ are said to be *dependent*

- *Alternative definitions* of *independence (indep)*: Replace cdf's by pmf's & pdf's for discrete & cont. r.v.'s, respectively

Independent Random Variables II

Definition

Discrete r.v.'s $X_1, \ldots, X_n$ are said to be *indept* if their *joint pmf* factors into the product of their marginal pmf's:

$$p(x_1, \ldots, x_n) = p_{X_1}(x_1) \times p_{X_2}(x_2) \times \cdots \times p_{X_n}(x_n)$$

for all values $x_1, \ldots, x_n \in \mathbb{R}$

Definition

Cont. r.v.'s $X_1, \ldots, X_n$ are said to be *indept* if their *joint pdf* factors into the product of their marginal pdf's:

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \times f_{X_2}(x_2) \times \cdots \times f_{X_n}(x_n)$$

for all values $x_1, \ldots, x_n \in \mathbb{R}$

Checking For Independence

- All the 3 previous definitions about indep of r.v.'s *require all the $n + 1$ expressions/functions* at both the LHS & the RHS of the displayed identities (in theory, given the joint cdf/pmf/pdf at the LHS, one can obtain the n marginal expressions at the RHS by *marginalizations*, but the computations could be very tedious & costly in time)

A Shortcut for Checking Independence

Random variables $X_1, \ldots, X_n$ are *indept* if and only if (iff) there exist functions $g_1, \ldots, g_n : \mathbb{R} \to \mathbb{R}$ s.t. for all $x_1, \ldots, x_n \in \mathbb{R}$, we have the joint pmf/pdf

$$f(x_1, \ldots, x_n) = g_1(x_1) \times g_2(x_2) \times \cdots \times g_n(x_n)$$

- *To check for indep*: Re-arrange the expression of the joint density to see if it is a *product* of some functions, each of which involves at most a single x_i

Example 9: *Falsify Independence*

1. Consider Example 6 at Page 33. X & Y are *dependent* as

$$0 = P(X = 0, Y = 0) \neq P(X = 0)P(Y = 0) = \frac{1}{8} \times \frac{2}{8} = \frac{1}{32}$$

2. In Example 7, X & Y are jointly distributed with pdf

$$f(x, y) = \begin{cases} 24xy, & 0 \leq x \leq 1,\ 0 \leq y \leq 1,\ x + y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

   One can verify that $f_X(.7), f_Y(.7) > 0$ (Find them as an exercise), while $f(.7, .7) = 24(.7)(.7) > 0 \Rrightarrow X$ & Y are *dependent*
   Alternatively, they are *dependent* as values of X & of Y are related through $x + y \leq 1$ in the support S

Example 10: *Verify Independence By Shortcut*

Suppose that

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y}, & 0 < x < \infty, 0 < y < \infty \\ 0, & \text{otherwise} \end{cases}$$

Let

$$g(x) = \begin{cases} 2e^{-x}, & 0 < x < \infty \\ 0, & \text{otherwise} \end{cases} \quad \& \quad h(y) = \begin{cases} e^{-2y}, & 0 < y < \infty \\ 0, & \text{otherwise} \end{cases}$$

Then, it is easily verified that

$$f(x,y) = g(x)h(y), \qquad x, y \in \mathbb{R}$$

$\therefore$ X & Y are *indept*

Example 11: Obtain Joint pmf/pdf of Independent r.v.'s

1. Suppose that X & Y are both discrete, & that they are indept. Then, one can *obtain the joint pmf*

$$p(x,y) = p_X(x)p_Y(y), \qquad \text{for all } x, y$$

   by multiplying the marginal pmf's when they are known

2. Suppose that X & Y are indept standard normal r.v.'s with pdf's $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2},\ x \in \mathbb{R}$, & $f_Y(y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2},\ y \in \mathbb{R}$. Then, the *joint pdf of X & Y is given by*

$$\begin{aligned} f(x,y) = f_X(x)f_Y(y) &= \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}}e^{-y^2/2} \\ &= \frac{1}{2\pi}e^{-(x^2+y^2)/2}, \qquad x, y \in \mathbb{R} \end{aligned}$$

A Representative of a r.v./distribution

- In general: >1 possible values for any r.v. X

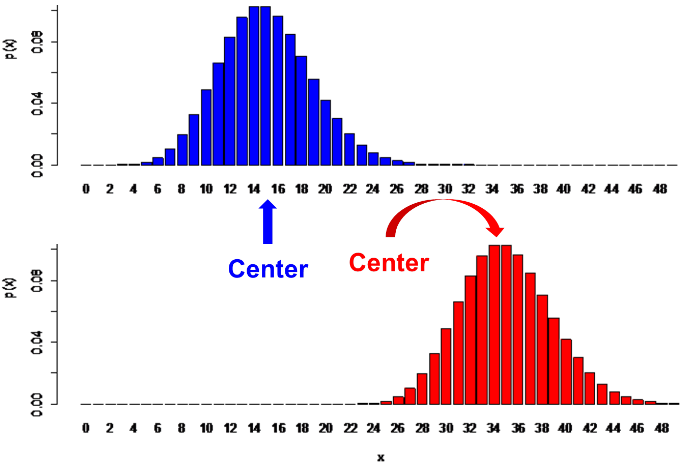
  Different values are observed in different occasions

  Which value is observed/realized is *governed by the pdf/pmf* (according to probs associated with each value), which is a *function*

- **"1 number" versus "A function"** : More handy to have 1 particular *informative value/number* to summarize/represent the density (a mathematical function)
  - roughly understand the r.v. in some sense
  - compare between different r.v.'s

"Center" of a r.v.

1 common way to obtain such an informative number in regards to the *"center" of a distribution* is based on the idea of an **average**



Expected Value of a Discrete r.v.

Definition

For a *discrete* r.v. X with density $p(x)$, the *expected value/expectation/mean of X*, denoted by $\mu/\mu_X/E(X)$, is

$$E(X) = \sum_i x_i \, p(x_i)$$

provided that $\sum_i |x_i| p(x_i) < \infty$. If the sum diverges, the expectation is undefined

- Regarded as the center of mass of $p(x)$
- A *weighted average* of all possible values that X can take on, each value being weighted by the prob that X assumes it

Expected Value of a Cont. r.v.

Definition

For a *cont.* r.v. X with density $f(x)$, the *expected value/expectation/mean of X*, denoted by $\mu/\mu_X/E(X)$, is

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

provided that $\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty$. If the integral diverges, the expectation is undefined

Example 12: *Expected Value = Center of Mass*

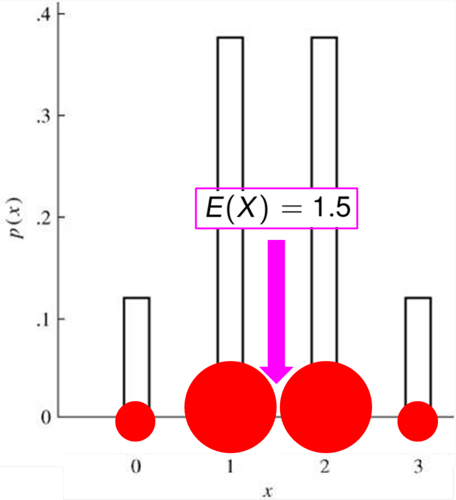
Consider the **Toss 3 fair coins** example in Example 1

The pmf of X is given by

$$p(x) = \begin{cases} .125, & x = 0, 3 \\ .375, & x = 1, 2 \\ 0, & \text{otherwise} \end{cases}$$

Hence,

$$\begin{aligned} E(X) &= (0+3)(.125) + (1+2)(.375) \\ &= 1.5 \end{aligned}$$



Example 13: Expected Value = Probability

Consider I_A, the *indicator r.v. of A*, *for any event of interest $A \subset S$*, defined by

$$I_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{if } A^c \text{ occurs} \end{cases}$$

Clearly, I_A is a discrete r.v., taking on 2 values, 1 & 0, with pmf summarized by $p(1) = P(I_A = 1) = P(A)$ & $p(0) = P(I_A = 0) = P(A^c) = 1 - P(A)$ $\Rightarrow$ $I_A \sim Ber(P(A))$, a *Bernoulli r.v. with parameter $P(A)$* defined in Example 2

It is straightforward that

$$E(I_A) = (1)[P(A)] + (0)[1 - P(A)] = P(A)$$

Theoretically, one can always

view any prob as an expectation or vice versa

Note: In general, for $X \sim Ber(p)$, $E(X) = p$

Example 14: Expected Value of a Discrete .v.

Consider a discrete r.v. X with pmf $p(x) = \binom{n}{x} p^x q^{n-x},\ x = 0, 1, \ldots, n$

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{n} xp(x) = \sum_{x=1}^{n} (x) \frac{n!}{x!(n-x)!} p^x q^{n-x} \\
&= (np) \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} \qquad [\text{let } y = x-1] \\
&= (np) \left[\sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} p^y q^{n-1-y} \right] = np
\end{aligned}
$$

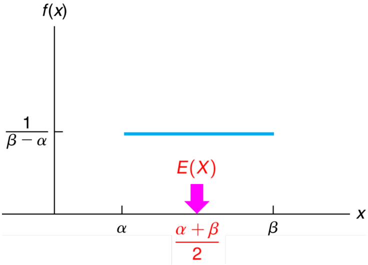
since the latter sum is 1 as it sums up all probs of another r.v. defined similarly as X with n replaced by $n-1$

Remark: This r.v. is said to follow a *binomial distribution*, which will be discussed in more details in Ch.5

Example 15: Expected Value of a Uniform r.v.

Consider $X \sim U(\alpha, \beta)$ in Example 3 with $f(x) = \dfrac{1}{\beta - \alpha}$, $\alpha \le x \le \beta$, which is a flat curve with constant value $\dfrac{1}{\beta - \alpha}$ on the interval $[\alpha, \beta]$

$$
\begin{aligned}
E(X) &= \left[\int_{-\infty}^{\alpha} + \int_{\alpha}^{\beta} + \int_{\beta}^{\infty}\right] x f(x)\, dx \\
&= 0 + \int_{\alpha}^{\beta} (x) \frac{1}{\beta - \alpha}\, dx + 0 \\
&= \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} x\, dx = \frac{1}{\beta - \alpha} \left[\frac{x^2}{2}\right]_{\alpha}^{\beta} \\
&= \frac{1}{\beta - \alpha} \left[\frac{\beta^2 - \alpha^2}{2}\right] = \frac{\alpha + \beta}{2}
\end{aligned}
$$

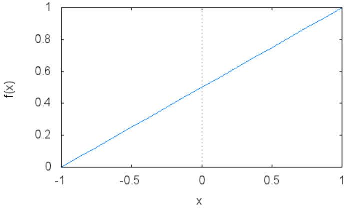


It is intuitive that the “center” of such a uniform distribution on $[\alpha, \beta]$ is at the mid–point of the boundaries α & β

Example 16: Expectation of a Cont. r.v.

Find $E(X)$ when the pdf of X is

$$f(x) = \begin{cases} (x+1)/2, & -1 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$



Solution:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)\,dx \\ &= \int_{-\infty}^{-1} xf(x)\,dx + \int_{-1}^{1} xf(x)\,dx + \int_{1}^{\infty} xf(x)\,dx \\ &= \int_{-\infty}^{-1} x(0)\,dx + \int_{-1}^{1} x(x+1)/2\,dx + \int_{1}^{\infty} x(0)\,dx \\ &= \int_{-1}^{1} \left[\frac{x^2}{2} + \frac{x}{2}\right] dx = \left[\frac{x^3}{6} + \frac{x^2}{4}\right]_{-1}^{1} = \frac{1}{3} \end{aligned}$$

Long–run Average Interpretation of an Expectation

An expected value $E(X)$ can be interpreted as **a long–run average**:
When *the same experiment is repeated/replicated for a lot of times independently*

What is "expected" to be the value of the r.v. of interest?

- **Toss 3 fair coins**: When we repeatedly toss 3 fair coins for a lot of times, the average of all "# of heads" would be close to $E(X) = 1.5$
- **Indicator r.v.**: When we repeatedly perform an experiment for a lot of times independently, the proportion of times of occurrence of A would be close to $E(I_A) = P(A)$; This supports

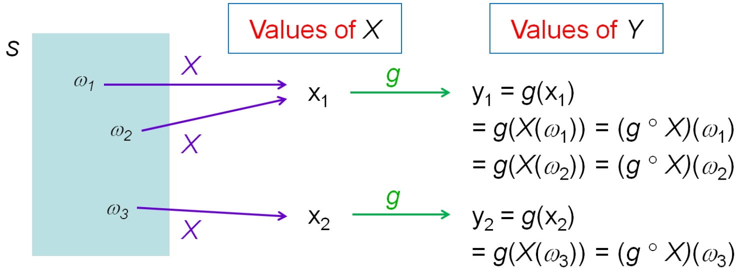
  empirical determination of any prob

- The **uniform r.v.** $X \sim U(\alpha, \beta)$: When we repeatedly *select a value on* $[\alpha, \beta]$ *randomly* for a lot of times, the average of all the selected #'s would be close to $E(X) = (\alpha + \beta)/2$

Expectation of a Function of a r.v. I

Given a r.v. X with known density, it is often that we may be interested in *another r.v.* $Y = g(X)$ which is defined as a *known function* g (either *one–to–one* or *many–to–one*) of X

- *e.g.,* interested in the revenue of a shop (Y) which depends on the sales (X) (assuming that the r.v., sales, is fully understood)



- The *composite function* $g \circ X$ defines a *new r.v.* Y *from* S *to* $\mathbb{R}$

Expectation of a Function of a r.v. II

Expectation of a Function of a r.v.

Suppose that $Y = g(X)$ for a known function $g : \mathbb{R} \to \mathbb{R}$

1. If X is a *discrete* r.v. with pmf $p_X(x)$, then $E(Y) = \sum_x g(x)p_X(x)$ provided that $\sum_x |g(x)|p_X(x) < \infty$
2. If X is a *cont.* r.v. with pdf $f_X(x)$, then $E(Y) = \int_{-\infty}^{\infty} g(x)f_X(x)\,dx$ provided that $\int_{-\infty}^{\infty} |g(x)|f_X(x)\,dx < \infty$

- In general, $E(g(X)) \neq g(E(X))$
- **Note** : $p_X(x)/f_X(x)$ is available $\Rrightarrow$ No need to derive $p_Y(y)/f_Y(y)$ in computing $E(Y)$ (with $Y = g(X)$) based on definition at Page 59 or 60

Example 17: Expectation of a Function of a r.v.

For a r.v. X with pmf given by $p(x) = \begin{cases} .125, & x = 0,3 \\ .375, & x = 1,2 \\ 0, & \text{otherwise} \end{cases}$. Find $E(X^2)$ & $E(2X+3)$

Solution:

1. $E(X^2) = \sum_{x=0}^{3} x^2 p(x) = 0^2(.125) + 1^2(.375) + 2^2(.375) + 3^2(.125)$
   $= 3 \; (\neq (E(X))^2 = 1.5^2 = 2.25)$
2. $E(2X+3) = \sum_{x=0}^{3} [2x+3]p(x)$
   $= [2(0)+3](.125)+[2(1)+3](.375)+[2(2)+3](.375)+[2(3)+3](.125)$
   $= 3(.125) + 5(.375) + 7(.375) + 9(.125) = 6$

Example 18: *Expectation of a Function of a r.v.*

Consider the cont. r.v. X with pdf given by $f(x) = 10x^{-2},\ x > 10$, in Example 5. Find $E(1/X)$ & $E(X^3)$

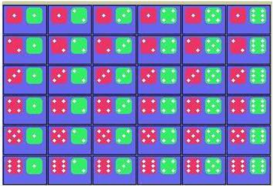
Solution:

1. $E(1/X) = \int_{10}^{\infty} (1/x)10x^{-2}\,dx = \int_{10}^{\infty} 10x^{-3}\,dx = \frac{10}{-3+1}[x^{-3+1}]_{10}^{\infty}$
$= \frac{10}{-2}[0 - 10^{-2}] = \frac{1}{20}$
2. $E(X^3) = \int_{10}^{\infty} (x^3)10x^{-2}\,dx = \int_{10}^{\infty} 10x\,dx = \frac{10}{2}[x^2]_{10}^{\infty} = \infty$

Expectation of a Function of ≥ 2 Random Variables I

Consider a toy example about *tossing 2 fair dice*:

- Sample space is represented by $S = \{(i, j) : i, j = 1, \dots, 6\}$ of cardinality as 36
- Define X_i be the # observed in the ith toss, for $i = 1, 2$
- Our interest is the sum of 2 #'s obtained, which is equivalent to $Z = X_1 + X_2$



We have in the above example:

1. Both X_1 & X_2 are r.v.'s defined from the same sample space S
2. *$Z = g(X_1, X_2) = X_1 + X_2$ is another function from $S \to \mathbb{R}$* defined by $Z(\omega) = X_1(\omega) + X_2(\omega)$ for any outcome/pair $\omega = (i, j) \in S,\ i, j = 1, \dots, 6$

$$\boxed{Z = X_1 + X_2 \textbf{ is a r.v.: } S \to \mathbb{R}}$$

Expectation of a Function of ≥ 2 *Random Variables II*

In general, the g function can be any function from $\mathbb{R}^2 \to \mathbb{R}$. Set $Z(\omega) = g(X(\omega), Y(\omega))$ for any $\omega \in S$. Then,

$$Z = g(X, Y) \text{ is a r.v.: } S \to \mathbb{R}$$

Expectation of a Function of 2 r.v.'s

1. If X & Y have a joint pmf $p(x, y)$, then

$$E(Z) = E[g(X, Y)] = \sum_y \sum_x g(x, y)p(x, y)$$

2. If X & Y have a joint pdf $f(x, y)$, then

$$E(Z) = E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy$$

Expectation of a Function of ≥ 2 Random Variables III

In general, the g function can be any function from $\mathbb{R}^n \to \mathbb{R}$. Set $Z(\omega) = g(X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$ for any $\omega \in S$. Then,

$$Z = g(X_1, \ldots, X_n) \text{ is a r.v.: } S \to \mathbb{R}$$

Expectation of a Function of $n \geq 2$ r.v.'s

1. If $X_1, \ldots, X_n$ have a joint pmf $p(x_1, \ldots, x_n)$, then

$$E[g(X_1, \ldots, X_n)] = \sum_{x_1} \cdots \sum_{x_n} g(x_1, \ldots, x_n) p(x_1, \ldots, x_n)$$

2. If $X_1, \ldots, X_n$ have a joint pdf $f(x_1, \ldots, x_n)$, then

$$E[g(X_1, \ldots, X_n)] = \int \cdots \int g(x_1, \ldots, x_n) f(x_1, \ldots, x_n) \, dx_1 \cdots dx_n$$

Expectation of a Function of ≥ 2 Random Variables IV

Expectation of a Sum of r.v.'s

Suppose that $X_1, \ldots, X_n$ are $n > 1$ r.v.'s with expectations $E(X_i)$. Then,

$$E(Z) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

Expectation of a Linear Combination of r.v.'s

Suppose that $X_1, \ldots, X_n$ are $n \geq 1$ r.v.'s with expectations $E(X_i)$. For fixed constants, $a, b_1, \ldots, b_n \in \mathbb{R}$, the *expected value of $Z = a + \sum_{i=1}^{n} b_i X_i$* is

$$E(Z) = a + \sum_{i=1}^{n} b_i E(X_i)$$

- *e.g.,* in Example 17, $E(2X + 3)$ can be alternatively computed as $E(2X + 3) = 2E(X) + 3 = 2(1.5) + 3 = 6$

Example 19: The Sample Mean I

Definition

1. Let $X_1, \ldots, X_n$ be *independent and identically distributed (i.i.d.) r.v.'s* having cdf F & mean μ. Such a sequence of r.v.'s is said to constitute a *sample or random sample* from the distribution characterized by F, & n is called the *sample size*
2. The *constant μ* is called the *population mean*
3. The *sample mean*, denoted by $\overline{X}$, is defined by

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- F is viewed as the *population of interest* (introduced in Ch.0): The *relative frequency histogram* of all measurements/values from such population is *identical to* the corresponding *density f/p of F*

Example 19: The Sample Mean II

- $\overline{X}$ is a r.v. defined as the *mean* of all n measurements in the sample

What is $E(\overline{X})$?

- $$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}(n\mu) = \mu$$

Mean of the Sample Mean Equals the Population Mean

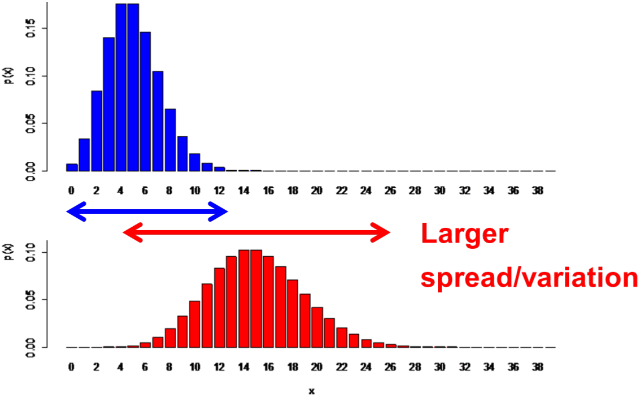
The expected value/mean of the sample mean always equals μ:

$$E(\overline{X}) = \mu$$

Note: In Statistics, our interest is usually about the unknown value of a population mean μ. The *sample mean* is often used to *estimate/provide a guess of* μ provided that there is a sample available from the population

Another common way to obtain an informative # in describing/summarizing a r.v. concerns about the *variation, or spread*, *i.e.,*

how dispersed the distribution is about its center



Variance I

Definition

For any r.v. X with mean $\mu < \infty$, the *variance of X* is defined by

$$\text{Var}(X) = E[(X - \mu)^2] \geq 0$$

provided that the expectation exists. The *standard deviation (sd) of X*, denoted by $SD(X)$, is defined by

$$SD(X) = +\sqrt{\text{Var}(X)}$$

- The mean/average of $Y = (X - \mu)^2$, the *squared deviation of X from the expected value of X*
- $\text{Var}(X) = 0 \Leftrightarrow X$ is not a r.v. & is a fixed constant equal to μ
- *$\text{Var}(X)$ has "strange" or usually meaningless units*: *e.g.,* when X is in unit of "S\$", $\text{Var}(X)$ would be in unit of "S\$^{2}", while the *sd* of X is in the *same* unit of "S\$" as X

Variance II

Computational Formula for Var(X)

For any r.v. X with mean $\mu < \infty$, the *variance of X* can be *equivalently computed* as

$$\mathrm{Var}(X) = E(X^2) - \mu^2$$

provided that $E(X^2)$ exists, where

$$E(X^2) = \begin{cases} \sum_x x^2 p(x), & X \text{ is discrete with pmf } p(x) \\ \int x^2 f(x)\, dx, & X \text{ is cont. with pdf } f(x) \end{cases}$$

- Proved by the linearity property of the expectation
- It *suffices to compute $E(X^2)$*

Variance III

Variance of a Linear Transformation of a r.v.

If $\text{Var}(X)$ exists & $Y = a + bX$ for some fixed constants $a, b \in \mathbb{R}$, then

$$\text{Var}(Y) = b^2\text{Var}(X)$$

- It is *intuitive* that adding any constant a to a r.v. X does *NOT* affect the spread of a r.v./distribution, as it merely shifts both the whole density p/f & the center of the distribution, $E(X)$, by a units on the horizontal axis
- The *sd of Y* follows as $SD(Y) = |b|SD(X)$

Example 20: *Variance & Standard Deviation*

Consider the r.v. X with pmf $p(x) = \begin{cases} .125, & x = 0, 3 \\ .375, & x = 1, 2 \\ 0, & \text{otherwise} \end{cases}$. Find Var($X$)

Solution:

- With $E(X) = 1.5$, by definition,

$$\begin{aligned} \mathrm{Var}(X) &= \sum_{x:p(x)>0} (x - E(X))^2 p(x) \\ &= (0 - 1.5)^2(.125) + (1 - 1.5)^2(.375) + (2 - 1.5)^2(.375) \\ &\quad + (3 - 1.5)^2(.125) \\ &= 2.25(.125) + .25(.375) + .25(.375) + 2.25(.125) = .75 \end{aligned}$$

- Alternatively, by its computational formula & $E(X^2) = 3$,

$$\mathrm{Var}(X) = 3 - E(X)^2 = 3 - 1.5^2 = .75$$

- The sd of X is given by $+\sqrt{.75} = .866$

Example 21: Variance of a Discrete r.v.

Let's compute the variance of the *binomial r.v.* considered in Example 14.

We have X with pmf $p(x) = \binom{n}{x} p^x q^{n-x}$, for $x = 0, 1, \ldots, n$, & $E(X) = np$.

As $E(X^2) = E[X(X-1)] + E(X)$, it suffices to compute

$$
\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{n} x(x-1)p(x) = \sum_{x=2}^{n} [x(x-1)] \frac{n!}{x!(n-x)!} p^x q^{n-x} \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x} \quad [\text{let } y = x-2] \\
&= n(n-1)p^2 \left[\sum_{y=0}^{n-2} \frac{(n-2)!}{y!(n-2-y)!} p^y q^{n-2-y} \right] = n(n-1)p^2
\end{aligned}
$$

Then, $Var(X) = n(n-1)p^2 + np - n^2p^2 = npq$

Example 22: Variance of a Cont. r.v.

Find the variance of X with pdf $f(x) = \begin{cases} (x+1)/2, & -1 < x < 1 \\ 0, & \text{otherwise} \end{cases}$

Solution: $E(X) = 1/3$ (obtained in Example 16)

1. By definition, $\text{Var}(X) = E\left[\left(X - \frac{1}{3}\right)^2\right] = \int_{-1}^{1} \left(x - \frac{1}{3}\right)^2 \frac{x+1}{2}\, dx$

$$= \frac{1}{18} \int_{-1}^{1} (x+1)(3x-1)^2\, dx = \frac{1}{18} \int_{-1}^{1} (9x^3 + 3x^2 - 5x + 1)\, dx = \frac{2}{9}$$

2. Alternatively, $E(X^2) = \int_{-1}^{1} (x^2) \frac{x+1}{2}\, dx = \int_{-1}^{1} \left[\frac{x^3}{2} + \frac{x^2}{2}\right] dx$

$$= \left[\frac{x^4}{8} + \frac{x^3}{6}\right]_{-1}^{1} = \frac{1}{3}$$

$$\Rightarrow \text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{1}{3} - \frac{1}{9} = \frac{2}{9}$$

Covariance I

- Either mean or variance provides a number which describes certain characteristics of a r.v./distribution. When it comes to ≥ 2 r.v.'s, they can only be used as a criterion of comparison, but not of describing any *relationship between the r.v.'s*
- Define the *covariance* of 2 r.v.'s as a measure of their degree of *linear association* – *degree to which X & Y "go together"*

How strong is the relationship/association between 2 r.v.'s?

Definition

If X & Y are jointly distributed r.v.'s with finite marginal expectations μ_X & μ_Y, respectively, the *covariance of X & Y* is

$$\text{Cov}(X, Y) = \text{Cov}(Y, X) = E\left[(X - \mu_X)(Y - \mu_Y)\right]$$

provided that the expectation exists

Covariance II

- Defined as an average of all the *product of the deviations of X from its mean & the deviation of Y from its mean*
- Computed by the result of $E[g(X, Y)]$
- **Note** : $Var(X) = Cov(X, X)$

Computational Formula For Cov(X, Y)

The covariance of X & Y can be equivalently computed by

$$\text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y$$

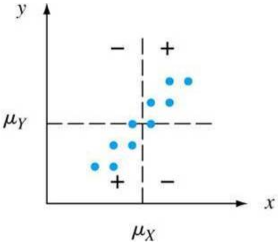
Independence $\Rrightarrow$ Zero Covariance

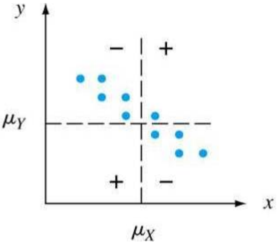
Indep of X & Y implies zero covariance of X & Y, but the converse is NOT true

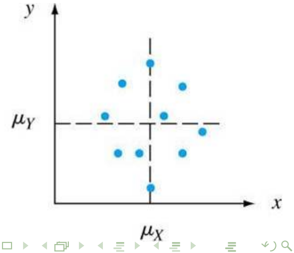
$$E(XY) = \mu_X \mu_Y \Rrightarrow \text{Cov}(X, Y) = 0$$

Covariance III

- Ranges from $-\infty$ to ∞:
  - +ve (X & Y are positively correlated): In general, both X & Y are larger/smaller than their respective means
  - -ve (X & Y are negatively correlated): In general, 1 of X & Y is larger (*resp.* smaller) than its mean while the other is smaller (*resp.* larger) than its mean
  - $= 0$ (X & Y are uncorrelated): there are +ve & -ve products of deviations which offset each other
- Observe pairs of observations (x, y) from the joint density: $Cov(X, Y)$ is +ve, -ve, ≈ 0







Example 23: Zero Covariance $\not\Rightarrow$ Independence

Here is a simple example illustrating the *converse of Indep implies zero covariance* is *NOT TRUE*. Two dependent r.v.'s X & Y having zero covariance can be obtained by letting X be a r.v. s.t.

$$P(X=0)=P(X=1)=P(X=-1)=\frac{1}{3},$$

& define

$$Y=\begin{cases}0, & \text{if } X \neq 0 \\ 1, & \text{if } X=0\end{cases}$$

Now, $XY=0$ with prob 1, so $E(XY)=0$. Also, $E(X)=0$ & thus

$$\text{Cov}(X,Y)=E(XY)-E(X)E(Y)=0$$

We have *zero covariance* of X & Y; however, X & Y are clearly *not indept* from the definition of Y

Example 24: Covariance of 2 Discrete r.v.'s

Suppose that the joint & the marginal pmf's for X = "automobile policy deductible amount" & Y = "homeowner policy deductible amount" are

$p(x,y)$	y		
x	0	100	200
100	.20	.10	.20
250	.05	.15	.30

x	100	250
$p_X(x)$	.5	.5

y	0	100	200
$p_Y(y)$	.25	.25	.5

from which $\mu_X = 175$ & $\mu_Y = 125$. Then,

$$\begin{aligned}
\text{Cov}(X, Y) &= E(XY) - (175)(125) \\
&= (100)(100)(.1) + (100)(200)(.2) + (250)(100)(.15) \\
&\quad + (250)(200)(.3) - 21,875 \\
&= 1,875
\end{aligned}$$

Note: Compute Cov(X, Y) in Example 6

Example 25: Covariance of 2 Cont. r.v.'s

Consider X & Y jointly distributed as the joint pdf $f(x, y)$ in Example 7

$$\begin{aligned}
E(XY) &= \int_{\mathbb{R}^2}\int (xy)f(x,y)\,dx\,dy = \int_S\int (xy)24xy\,dx\,dy \\
&= 24\int_0^1\int_0^{1-y} x^2y^2\,dx\,dy = 8\int_0^1 y^2[x^3]_0^{1-y}\,dy \\
&= 8\int_0^1 y^2(1-y)^3\,dy = 8\int_0^1 (y^2 - 3y^3 + 3y^4 - y^5)\,dy = \frac{2}{15}
\end{aligned}$$

$$E(X) = \int_{\mathbb{R}^2}\int (x)f(x,y)\,dx\,dy = \int_0^1\int_0^{1-y} 24x^2y\,dx\,dy = \cdots = -\frac{2}{5}$$

$$E(Y) = \int_{\mathbb{R}^2}\int (y)f(x,y)\,dx\,dy = \int_0^1\int_0^{1-y} 24xy^2\,dx\,dy = \cdots = \frac{2}{5}$$

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{2}{15} - \left(-\frac{2}{5}\right)\left(\frac{2}{5}\right) = \frac{22}{75}$$

Covariance IV

Covariance of 2 Sums of r.v.'s

Suppose that $U = a + \sum_{i=1}^{n} b_i X_i$ & $V = c + \sum_{j=1}^{m} d_j Y_j$ for constants $a, b_1, \ldots, b_n, c, d_1, \ldots, d_m \in \mathbb{R}$. Then,

$$\text{Cov}(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_i \, d_j \, \text{Cov}(X_i, Y_j)$$

Some special cases of the above result about covariances:

- $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$ for any $a, b \in \mathbb{R}$
- $\text{Cov}(X+Y, Z+W) = \text{Cov}(X, Z) + \text{Cov}(X, W) + \text{Cov}(Y, Z) + \text{Cov}(Y, W)$

Covariance V

Variance of a Linear Combination of r.v.'s

For any n r.v.'s, $X_1, \ldots, X_n$, & fixed constants $a_1, \ldots, a_n, b_1, \ldots, b_n$,

$$\text{Var}\left(a + \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n} b_i^2 \text{Var}(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} b_i\, b_j\, \text{Cov}(X_i, X_j)$$

A special case of the above result about covariances:

- $\text{Var}(X \pm Y) = \text{Cov}(X \pm Y, X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X, Y)$

Variance of a Sum of Independent r.v.'s

Suppose that $X_1, \ldots, X_n$ are n indept r.v.'s. Then,

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i)$$

Example 26: *Variance of the Sample Mean*

Consider the same set–up as in Example 19. Let $\sigma^2 > 0$ be the common variance of each X_i, called the *population variance*, which can be computed from F by definition. What is $\mathrm{Var}(\overline{X})$?

Solution: $\mathrm{Var}(\overline{X}) = \left(\frac{1}{n}\right)^2 \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{\sigma^2}{n}$

where the second last equality follows from indep of all the X_i's

Variance of the Sample Mean Equals the Population Variance Divided by the Sample Size

The variance of the sample mean $\overline{X}$ always equals the ratio of σ^2 & n:

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

Note: Results of both $E(\overline{X}) = \mu$ in Example 19 & $\mathrm{Var}(\overline{X}) = \sigma^2/n$ here are *valid for all F (i.e., any population)*

Example 27: Zero Covariance of $\overline{X}$ and Any Deviation

Consider the same set–up as in Example 19. Show that $\mathrm{Cov}(\overline{X}, X_i - \overline{X}) = 0$, *i.e.*,

the sample mean & a deviation from it are uncorrelated

Solution:

$$\begin{aligned}\mathrm{Cov}(X_i - \overline{X}, \overline{X}) &= \mathrm{Cov}(X_i, \overline{X}) - \mathrm{Cov}(\overline{X}, \overline{X}) \\ &= \mathrm{Cov}\left(X_i, \frac{1}{n}\sum_{j=1}^{n} X_j\right) - \mathrm{Var}(\overline{X}) \\ &= \frac{1}{n}\sum_{j=1}^{n}\mathrm{Cov}(X_i, X_j) - \frac{\sigma^2}{n} = 0\end{aligned}$$

where the final equality follows since

$$\mathrm{Cov}(X_i, X_j) = \begin{cases} 0, & \text{if } j \neq i \qquad \text{[by indep]} \\ \sigma^2, & \text{if } j = i \qquad \text{[since } \mathrm{Var}(X_i) = \sigma^2\text{]} \end{cases}$$

Moment Generating Function of a r.v. I

Definition

The *moment generating function (mgf)* of a r.v. X is defined by, for a constant $t \in \mathbb{R}$,

$$\phi(t) = E(e^{tX}) = \begin{cases} \displaystyle\sum_x e^{tx} p(x), & \text{if } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} e^{tx} f(x)\, dx, & \text{if } X \text{ is cont.} \end{cases}$$

when the expectation exists

- $\phi(t)$ is a function of t (*i.e.,* does not contain x anymore)
- *Generates* all the *moments* $E(X^k)$, for $k = 1, 2, \ldots,$ of the r.v. X
- Due to the usage of $\phi(t)$, we only care about t to be in an open interval containing zero $\Leftrightarrow t \in (-a, b)$ for some $a, b \in \mathbb{R}$

Moment Generating Function of a r.v. II

Characterization of a r.v.

If the mgf of X exists for t in an open interval containing zero, it *uniquely determines the probability distribution of X*

Determination of the kth moment of a r.v.

If the mgf of X exists for t in an open interval containing zero, then the *kth moment of X* is given by

$$E(X^k) = \phi^k(0) = \left.\frac{d^k}{dt^k}\phi(t)\right|_{t=0}$$

for any $k = 1, 2, \ldots$

Example 28: mgf & Moments of a Binomial r.v.

Consider the *binomial r.v.* in Example 14 again. Its *mgf* $\phi(t)$ equals

$$E(e^{tX}) = \sum_x e^{tx} p(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} = (pe^t + q)^n$$

Now, we can obtain *some moments & variance* of this r.v.:

1. Differentiating $M(t)$ once yields

$$\phi^1(t) = n(pe^t + q)^{n-1} pe^t \qquad \& \qquad E(X) = \phi^1(0) = np$$

2. Differentiating $\phi^1(t)$ once yields

$$\phi^2(t) = n(n-1)(pe^t + q)^{n-2}(pe^t)^2 + n(pe^t + q)^{n-1} pe^t$$

& $E(X^2) = \phi^2(0) = n(n-1)p^2 + np$
$\therefore \mathrm{Var}(X) = E(X^2) - [E(X)]^2 = npq$

Example 29: mgf & Moments of a $U(\alpha, \beta)$ r.v.

For $X \sim U(\alpha, \beta)$, its *mgf* $\phi(t)$ equals

$$E(e^{tX}) = \int_\alpha^\beta \frac{e^{tx}}{\beta - \alpha} dx = \frac{1}{\beta - \alpha} \int_\alpha^\beta e^{tx} dx = \frac{1}{\beta - \alpha} \left[\frac{e^{tx}}{t} \right]_\alpha^\beta = \frac{e^{\beta t} - e^{\alpha t}}{(\beta - \alpha)t}$$

Now, we can obtain *some moments & variance*:

1. Differentiating $\phi(t)$ once yields

$$\phi^1(t) = \frac{\beta e^{\beta t} - \alpha e^{\alpha t}}{(\beta - \alpha)t} - \frac{e^{\beta t} - e^{\alpha t}}{(\beta - \alpha)t^2} \quad \& \quad E(X) = \phi^1(0) = \frac{\alpha + \beta}{2}$$

2. Differentiating $\phi^1(t)$ once yields

$$\phi^2(t) = \frac{\beta^2 e^{\beta t} - \alpha^2 e^{\alpha t}}{(\beta - \alpha)t} - \frac{2(\beta e^{\beta t} - \alpha e^{\alpha t})}{(\beta - \alpha)t^2} + \frac{2(e^{\beta t} - e^{\alpha t})}{(\beta - \alpha)t^3}$$

$$\& \; E(X^2) = \phi^2(0) = \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} = \frac{\beta^2 + \alpha\beta + \alpha^2}{3}$$

$$\therefore \mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{(\beta - \alpha)^2}{12}$$

Moment Generating Function of a Function of a r.v./r.v.'s

mgf of a Linear Transformation of a r.v.

If X has the mgf $\phi_X(t)$, for fixed constants $a, b \in \mathbb{R}$, the *mgf of $Y = a + bX$* equals

$$\phi_Y(t) = e^{at}\phi_X(bt) \tag{6}$$

mgf of a Sum of Independent r.v.'s

If $X_1, \ldots, X_n$ are *indept* r.v.'s with mgf's ϕ_{X_i}, then the *mgf of $Z = X_1 + X_2 + \cdots + X_n$*

$$\phi_Z(t) = \phi_{X_1}(t) \times \phi_{X_2}(t) \times \cdots \times \phi_{X_n}(t) \tag{7}$$

on the common interval where all the n mgf's at the RHS exist

- Both of the above results are *extremely useful* as they require only mgf's of individual r.v.'s but not manipulation of any joint densities

Example 30: Relationship Between $U(0,1)$ & $U(\alpha,\beta)$ r.v.'s

Example 29 gives the mgf of $X \sim U(\alpha,\beta)$ as

$$\phi_X(t) = \frac{e^{\beta t} - e^{\alpha t}}{(\beta - \alpha)t}$$

Obviously, setting $\alpha = 0$ & $\beta = 1$ yields the mgf of $U \sim U(0,1)$ as

$$\phi_U(t) = \frac{e^t - 1}{t}$$

Re-arrange the above expression of $\phi_X(t)$ as in the form of (6):

$$\phi_X(t) = e^{\alpha t}\frac{e^{(\beta-\alpha)t} - 1}{(\beta - \alpha)t} = e^{\alpha t}\phi_U((\beta - \alpha)t)$$

$\Rightarrow$ *$X \sim U(\alpha,\beta)$ is a linear transformation of $U \sim U(0,1)$* through $X = \alpha + (\beta - \alpha)U$

Example 31: mgf of The Sample Mean

Let us find the mgf of *the sample mean* $\overline{X}$ defined in Example 19

- First, note that
$$\overline{X} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n$$
- $\because X_1, \ldots, X_n$ are i.i.d. r.v.'s. Suppose that the mgf of each X_i is $\phi_{X_i}(t) = E(e^{tX_i}) = g(t)$ for some function g
- By (6) with $a = 0$ & $b = 1/n$, the mgf of $\dfrac{1}{n}X_i$ for all i is

$$\phi_{X_i/n}(t) = \phi_{X_i}(t/n) = g(t/n)$$

By (7), the *mgf of* $\overline{X}$ is given by

$$\phi_{\overline{X}}(t) = \phi_{X_1/n}(t) \times \phi_{X_2/n}(t) \times \cdots \times \phi_{X_n/n}(t) = [g(t/n)]^n$$

($\because$ indep of $X_1, \ldots, X_n \Rrightarrow$ indep of $X_1/n, \ldots, X_n/n$)