Primer on Probability/Random Variables

The 0th Module

Real World Experiments and Mathematical Abstraction

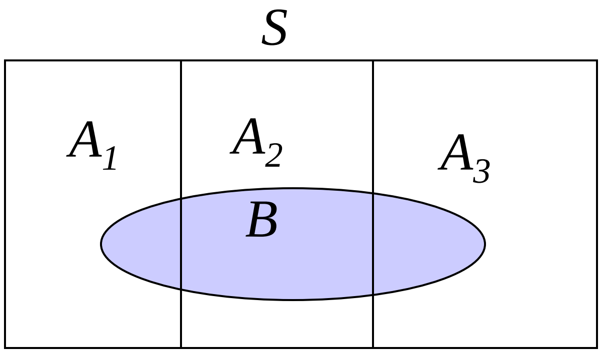
- Experiments
  - Measurement of voltage across a resistance
  - Roll a die
- Three entities in the real world experiments
  - The set of all possible *outcomes*
  - Grouping of the *outcomes* into classes, called *results*
  - The *relative frequencies* of occurrences of the *results*
- The corresponding mathematical abstractions
  - *The sample space*
  - *The set of events*
  - *The probability measure* assigned on each of these *events*

Fundamental Definitions in Set Theory

- A set is a collection of *objects* (*elements*).
  - $A = \{v: 0 \le v \le 5 \text{ volts}\}$
  - $B_1 = \{1, 2, 3, 4\}$, $B_2 = \{\text{head, tail}\}$
- A subset C of A is another set whose elements are also elements of A.
  - $C = \{1, 2\} \subset B_1$
  - We say C belongs to B_1
- Set operations: Union and Intersection
  - $B_1 \cup B_2 = \{1,2,3,4, \text{head, tail}\}$
  - $B_1 \cap C = \{1, 2\}$ (Sometimes, a shorthand notation, B_1C, is used)
- The *empty set* or *null set* $\{\varnothing\}$ (or simply $\varnothing$) is the set having no elements.

Fundamental Definitions in Set Theory

- Two sets A and B are *mutually exclusive* or *disjoint* if they have no common elements.
  - $A \cap B = AB = \varnothing$
- A partition U of a set S is a collection of mutually exclusive subsets A_i of S whose union equals S.
  - $S = A_1 \cup A_2 \cup A_3$ and $A_i A_j = \varnothing$ for any i, $j \neq i$
- In the figure below, $U=[A_1, A_2, A_3]$, and the subset $B = (A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B)$

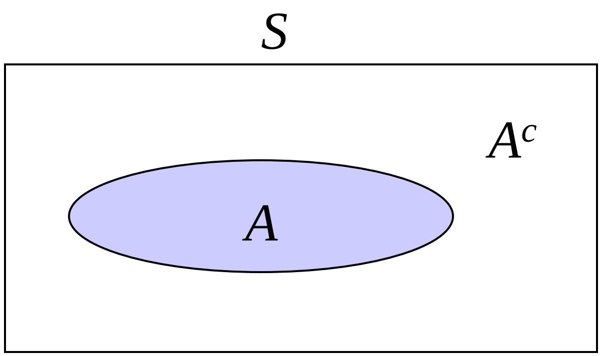


Sample Spaces and *Events*

- A sample space Ω, which is called *the certain event of a particular experiment*, is the collection of all *experimental outcomes* (objects).
  - An object in Ω is called *a sample point*; is usually denoted by ω.
- Subsets of a sample space is called *events*.
  - Grouping of the *outcomes* into the subsets
  - A set of sample points
  - A = {ω: some condition(s) on ω is provided here}, the event A is the set of all ω satisfying the condition(s) on ω.
  - An event consisting of a single element is called an *elementary event.*

Complement of an Event

- We define a complement of an event A as the set of all outcomes of S which are not included in A.
- We denote $A^c = S \backslash A$.



Examples of *Sample Spaces* and *Events* (*Results*)

- Die experiment: $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - $A = \{\omega: \text{odd}\} = \{1, 3, 5\}$
  - $B = \{\omega: \text{even}\} = \{2, 4, 6\}$
- The closed interval of the real line:

  $\Omega = [0, 1] = \{\omega: 0 \leq \omega \leq 1\}$
  - $A = \{\omega: 0.2 \leq \omega \leq 0.7\}$
- All time functions $f(t)$, $-\infty < t < \infty$
  - An event may be a set of all time functions whose energy is less than 1.
- A finite sample space of N elements → There are 2^N possible subsets.

Trial

- A single performance of an experiment is called a *trial*.
- In each trial we observe a single outcome $a_i \in S$.
- We say an event A occurs during this trial when A contains a_i .
- From a single trial, multiple events can occur.
- Roll a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
  - Now, suppose after a trial, an outcome "1" is observed.
  - Then, the events {1}, {1, 3, 5}, {1, 3}, and all the rest $2^5 - 3$ events that contain "1" as an element, it can be said, have occurred.

On the Occurrence of Events In a Trial

- We say an event $A=\{a_1, a_2, a_3\}$ has occurred in a trial, if any one element of the set, namely, a_1, a_2, or a_3, was the outcome of the trial.

- The event Ω occurs in every trial.

Probability Measure

- *An assignment* of a real number from the interval [0, 1] to the *events* defined on Ω.

  - Ex) Fair die: All faces occur equally likely with probability 1/6.

  - Ex-2) Unfair die: face-1 event occurs with probability 1/3, the rest 5 faces with 2/15.

  - Ex-3)You can create and use your own rule which suits your needs the most (your betting rule in Gambling for example).

- Probability measure P(A) is assigned to a *field E* of subsets (events) of the sample space Ω.

  $P: E \rightarrow [0, 1]$

Relative Frequency vs. Probability Measure

- The assignment of probability measure to an event A, P(A), may be done in terms of relative frequency of occurrences in N independent trials

  $$P(A) = \lim_{n\to\infty} n_A/N$$

  where n_A is the number of occurrence of event A in N trials

- Ex-1) a coin is tossed 100 times.
  - The event of head occurred 51 times.
  - Then, P(A) = 51/100
- Ex-2) An experienced gambler watches the cards played, and updates his table of probability measures assigned only on the events of his interests and makes bets accordingly

Axiomatic Definition of Probability

- The assignment of probability to events should follow the three fundamental rules (Kolmogoroff's axioms)
- 1. $0 \leq P(A) \leq 1$ (The frequency of an event)
- 2. $P(\Omega) = 1$ (In every trial there is an outcome)
- 3. If $A \cap B = \varnothing$, then $P(A \cup B) = P(A) + P(B)$
  - Die: frequency(1 or 2) = frequency(1) + frequency(2), $\{1\} \cap \{2\} = \varnothing$
- In the theory of probability, all conclusions are direct or indirect consequences of these three axioms.
- These conclusions allow us to predict -- by calculation – the probability of occurrence of observable(or wanting-to-observe) events in real world experiments.

- Reference: Web-site: http://www.kolmogorov.com/Kolmogorov.html

Examples

- A coin toss experiment: S = {h, t}
- Events are the four subsets of S, {∅}, {t}, {h}, {h, t}.
- It forms a sigma field.
- A sigma field is a collection of sets which is closed under the *union* and the *complement* operations.
- A complement of {t} is {h} in this example.
- We will use superscript c to denote complement, i.e., $\{t\}^c = \{h\}$ and $S^c = \{\varnothing\}$.
- We may assign P{t} = *p* and P{h} = *q*, i.e., $p + q = 1$.

Coin Toss Three Times Experiment (1)

- S ={hhh, hht, hth, htt, ttt, tth, tht, thh}.
- Assume a fair coin; then head/tail occurs with equal prob.
- First, consider the naive case such that all 2^8 possible events are of interest, and then the probability assignment is trivial.
- Ex) The probability of an event {hht, hhh} is
  – P{hht, hhh} = P{hht} + P{hhh} = 2/8 .

Coin Toss Three Times Experiment (2)

- Now, consider a non trivial case:
- Suppose we are interested in the occurrence of an event A = {hth, tht} only.
- Then, we assign probability to only those events in the sigma field formed by A, i.e., $\{A, A^c, \{\varnothing\}, S\}$.
- Thus, assign P(A) = 1/4 (The coin is a fair coin).
- We note that the probability measure satisfies all the conditions of the Kolmogoroff's axioms.

Conditional Probability

- Given any two events A and B, the conditional probability $P(A|B)$ of an event A is defined as

  $$P(A \mid B) := P(AB)/P(B)$$

  whenever $P(B) \neq 0$.

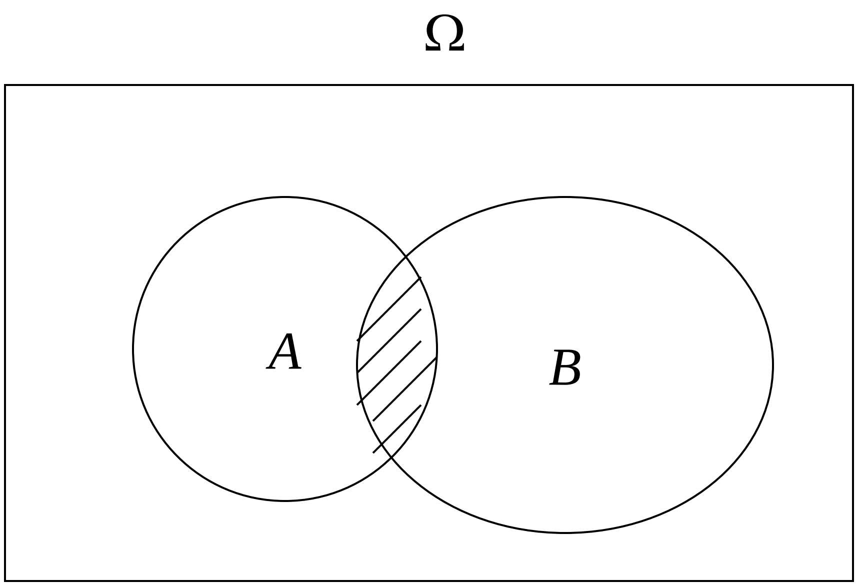
- $P(A \mid A) = 1$

- In the Coin-Toss Three Times experiment, let A={hhh} and B = {a head in the first toss} = {hhh, hht, hth, htt}

  $P(A \mid B) = (1/8)/(1/2) = 1/4$.

Probability of Joint Event

- Notation: P(A, B) = P(AB) = P(A ⋂ B)
- We refer P(A, B) as the probability of a "joint event A and B."



Probability of Joint Event

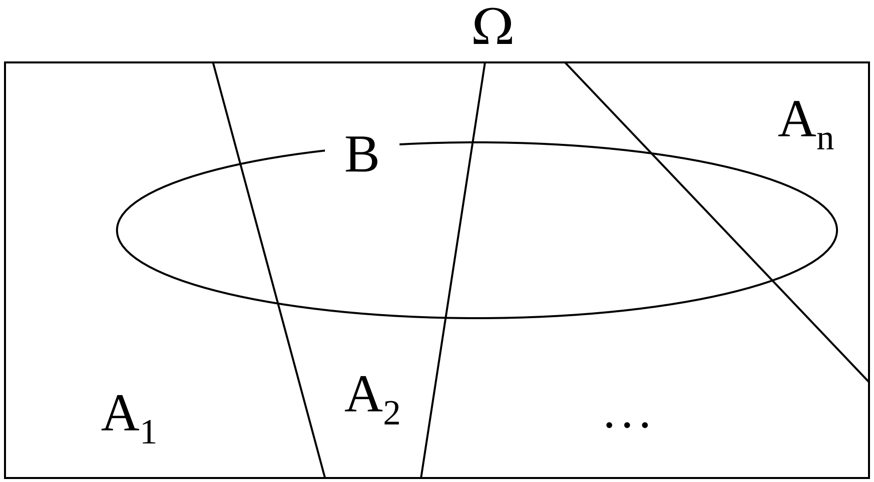
- $P(A, B) = P(A \mid B)\, P(B)$
  $= P(B \mid A)\, P(A)$

- A box contains three white balls, w_1, w_2, and w_3 and two red balls r_1 and r_2. We remove two balls in succession. What is the probability that the first removed is white and the second is red?

Independence

- If P(A|B) = P(A) or P(B|A) = P(B), the two events A and B are, said to be, (statistically) independent with each other.
- Coin Toss Twice:
  - Ω = {hh, ht, th, tt}
  - Suppose we use numbers a and b in [0, 1] with $a + b = 1$ in the following manner:
  - $P\{hh\}=a^2$, $P\{ht\}=P\{th\}=ab$, $P\{tt\}=b^2$
  - Note that the assignment satisfies the axioms: $a^2+2ab+b^2 = (a+b)^2 = 1$
  - Now, define two events A={head at the first toss} and B={head at the second toss}
  - Note $P(A)=aa + ab = a$ and $P(B)= ba + aa = a$
  - $P(A, B) = P\{hh\} = a^2 = P(A)\,P(B)$
  - Then, we note A and B are mutually independent.

Theorem of Total Probability (Very Important)



- If $U=[A_1, A_2, \ldots, A_n]$ is a partition of Ω and B is an arbitrary event, then

  $P(B) = P(B, A_1) + P(B, A_2) + P(B, A_3) + P(B, A_4)$

  $= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) + P(B|A_4)P(A_4)$

Bayes' Theorem [Very Important]

- From the results of the conditional probability and the total probability theorem, we could easily get the following,

$$
\begin{aligned}
P(A_i|B) &= \frac{P(A_i, B)}{P(B)} \\
&= \frac{P(B|A_i)\,P(A_i)}{\sum_{k=1}^{n} P(B|A_k)\,P(A_k)}
\end{aligned}
$$

Examples of Bayes' Theorem

- Box-1 contains *a* white balls and *b* black balls. Box-2 contains *c* white balls and *d* black balls. One ball is drawn from Box-1 and inserted into Box-2. Then, a ball is drawn from Box-2.
- What is the probability that a ball drawn from Box-2 is white?

- What is the probability that the first draw from Box-1 was black, given that a white ball was obtained at the second draw from Box-2

Permutation/Combination

- Consider a set of N distinct objects
- *Permutation*: The total number of distinctive arrangements (each in an ordered sequence) of N distinct objects is

  $N!$

- The total number of distinctive arrangements when taking K objects out of N distinct objects is

  $N(N-1)(N-2) \ldots (N-K+1) = N!/(N-K)!$

- *Combination*: The total number of ways to select K objects out of N distinct objects is

$$\binom{N}{K} = \frac{N!}{(N-K)!K!}$$

Bernoulli Trials

- Observe the occurrence of an event A in each trial
- The event A occurs with P(A) = p and P(A^c) = 1- p = q
- Find the probability of a compound event that there are k occurrences of event A in N trials
- None in *N* … $(1-p)^N$
- One in *N* … $N\,p(1-p)^{N-1}$
- Two in *N* … $\binom{N}{2} p^2 (1-p)^{N-2}$

…

- In general,

$$P\{\text{A occurs k times in N trials}\} = \binom{N}{k} p^k (1-p)^{N-k}$$

Random Variable and Processes

- A signal is a function of time

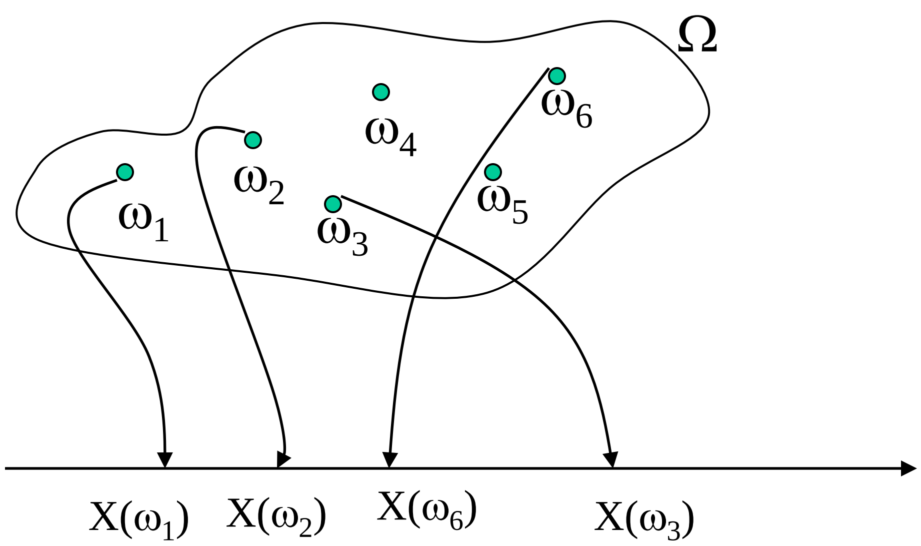
  ex) $y(t) = \sin(2\pi f_c t)$, this is a deterministic signal
- A random signal: the value of the signal at a fixed time *t* is a random variable

  ex) $y(t) = \sin(2\pi f_c t + \theta)$, $0 \le t \le T$

  where θ is +180 degree with probability 1/2 or -180 degree with prob. 1/2
- A random process y(*t*) is a collection of different random variables at each time *t*
  - Stochastic processes

Random Variable

- A function $X: \Omega \rightarrow R$ (Domain is Ω, range is R)
  - Given any ω, the function specifies a finite real number $X(\omega)$



A random variable is a function whose domain is Ω, the range of this function is usually a real line (Real-valued random variable). Also, it has a probability distribution $Pr\{X <= x\}$ associated with it.

Motivation for RV

- It may be easier to deal with numbers, instead of abstract objects.

Events Described by Random Variables

- We know now that we assign probability to the field of subsets of Ω .
- Note that with the use of a random variable, the subsets of *range* space are associated with the subsets of Ω.
- Thus, events defined on the outcomes of experiments can be described by the subsets of the range space of the function.

Examples of Random Variables

- Roll a die experiment
  - 6 outcomes, $\Omega = \{f_1, f_2, f_3, \ldots, f_6\}$
  - We may define a random variable X_1 which has the following rule

    $X_1(f_1) = 10$, $X_1(f_2) = 20$, $X_1(f_3)=30$, $X_1(f_4)=40$, $X_1(f_5)=50$, and $X_1(f_6)=60$
  - We may also define a random variable X_2 which uses the following rule

    $X_2(f_1) = -1$, $X_2(f_2) = -2$, $X_2(f_3) = -3$, $X_2(f_4)=+3$, $X_2(f_5)=+2$,…, and $X_2(f_6) = +1$
  - It's up to the designer to choose a map for convenience

Examples of Random Variables (2)

- According to the r.v.s X_1 and X_2, we can say the following:
- A subset $\{\omega: X_1(\omega) = 10, 30, 50\}$ is equivalent to the event $\{f_1, f_3, f_5\} = \{\omega: \text{odd}\}$.
- Similarly, a subset $\{\omega: X_2(\omega) = -1, -2\}$ is equivalent to the event $\{f_1, f_2\}$.
- Thus, we can talk about assigning a probability measure on the events described by random variables, in exactly the same way we do with the events of Ω.

Distribution Function

- Suppose the probability measure defined on the die experiment was

  $P(f_k) = 1/6$ for all $k=1, 2, .., 6$

- Then, correspondingly we could have the probability measure defined on the random variables X_1 and X_2
- For X_1, we have

  $P(X_1 = 10) = 1/6$, $P(X_1=20) = 1/6$, …

- For X_2, we have

  $P(X_1 = -1) = 1/6$, $P(X_2=-2) = 1/6$, $P(X_2=-3) = 1/6$…

- Probability assignment is easy with finite and countable sample space.

Distribution Function (2)

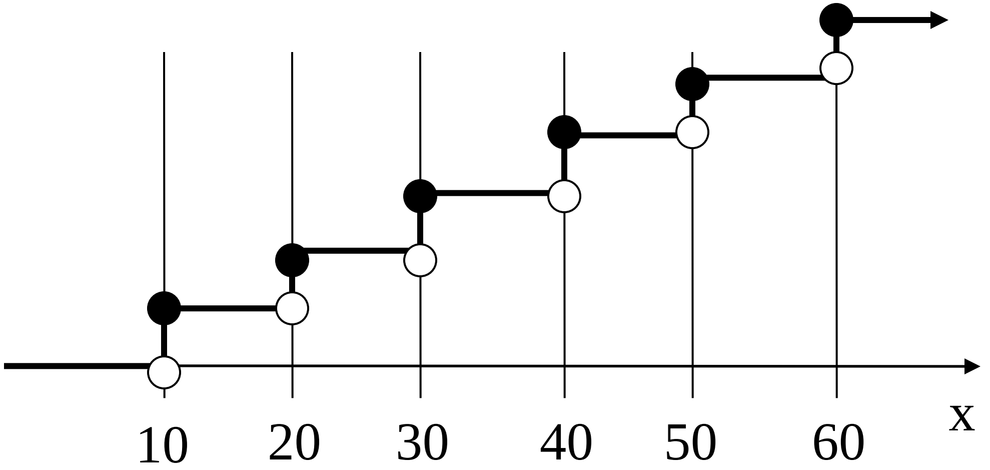
- We use a cumulative distribution function to deal with an infinite uncountable sample space.
  - For example, S = [0, 1].
- The probability is assigned on the intervals of interest.
  - A collection of intervals, say events, is of interest.
  - A sigma field can be formed for the collection of intervals.
  - Distribution function $F_X(x)$ of a random variable X is defined as

$$F_X(x) := P(\omega: X(\omega) \leq x)$$

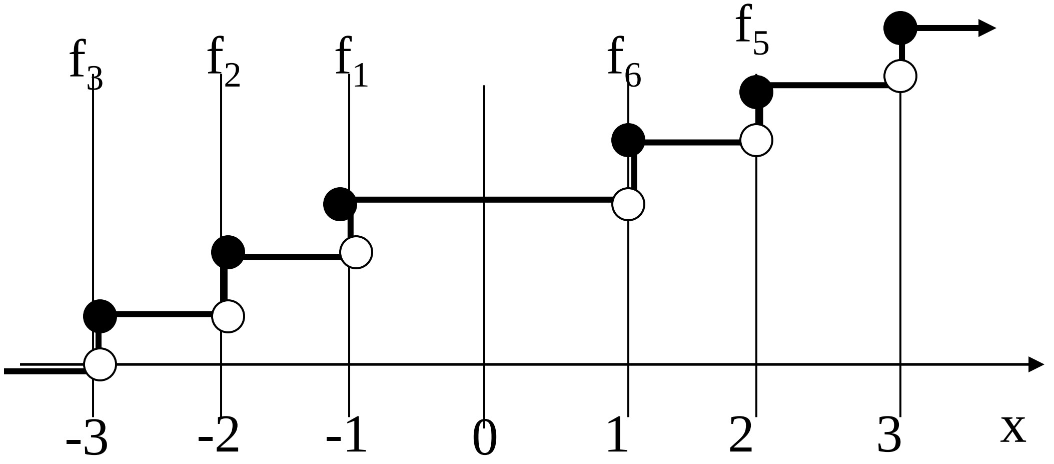
  - It is called the cumulative distribution function (CDF) of X.
- Examples) Find the distribution functions for random variable X_1 and X_2 that were defined in the roll-a-die experiment.

Distribution Function (3)

- $F_{X1}(x) = P(X_1 \leq x)$
- Note that the function is right continuous



- $F_{X2}(x) = P(X_2 \leq x)$



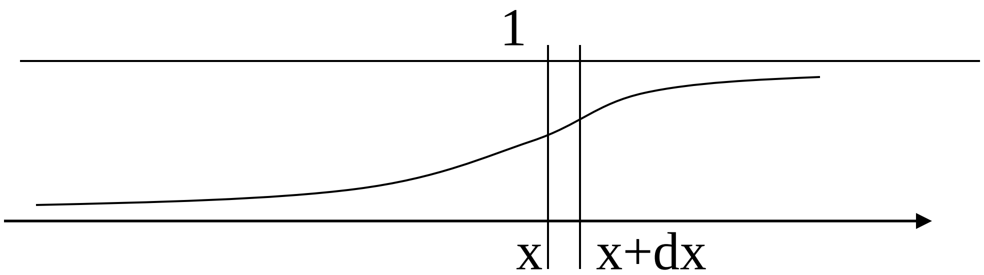
Properties of Distribution Function F(x)

- Non-decreasing function of x: For $x_2 > x_1$, $F(x_2) \geq F(x_1)$
- Continuous from the right.

  $\lim_{\varepsilon \Downarrow 0} F(x+\varepsilon) = F(x)$,
- $F(-\infty) = P(X \leq -\infty) = 0$
- $F(+\infty) = P(X \leq \infty) = 1$
- $0 \leq F(x) \leq 1$

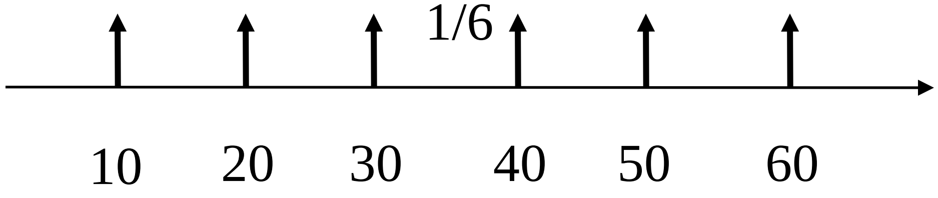
Probability Density Function f(x)

- $f(a) := dF(x)/dx \,|_{x=a}$
- $P\{x < X \leq X+dx\} = P\{X \leq x+dx\} - P\{X \leq x\} = f(x)\,dx$



- Example of pdf of X_1:

  $f_1(x) = (1/6) \sum_{k=1}^{6} \delta(x - 10\,k)$



Ensemble Averages (Expected Value)

- Ensemble Average E(X)
- 1^{st} moment: $m_1 = E\{X\} := \int_{=\infty}^{\infty} x\, f(x)\, dx$
- Note that this operator is a linear operator

- 2nd moment: $m_2 = E\{X^2\} = \int_{=\infty}^{\infty} x^2 f(x)\, dx$
- $Var(X) = E\{(X - E\{X\})^2\} = E\{(X - m_1)^2\}$

  $= E(X^2) - E\{X\}m_1 - m_1\, E\{X\} + E(m_1^{\,2}) = m_2 - m_1^{\,2}$

Ensemble Average of Product XY

- X and Y are two random variables with PDF $f_x(x)$ and $f_y(y)$
- Then, $f_{XY}(x, y)$ is the joint density function
- $E\{XY\} = \int \int x\, y\, f_{xy}(x, y)\, dx\, dy$
  - This is called the **Correlation** of the two random variables X and Y
  - Note, what happens when X and Y are independent
  - When E{XY} = E{X}E{Y}, X and Y are said to be **mutually uncorrelated**
  - Note, if you have two indep. r.v.s, then they are uncorrelated, but not vice versa
- E{(X-E(X))(Y-E(Y))} is called the **Covariance**
  - Note what happens when two are uncorrelated

Binomial Distribution

- Binomial PDF: prob. of obtaining k "1"s in N Bernoulli trials

  $P(k) = \binom{N}{k} p^k (1-p)^{N-k}$

  By letting x = k, where k = 0, 1, 2, …, N

  $f(x) = \sum_{k=0}^{N} P(k)\, \delta(x - k)$

  – Binomial expansion: $(p + q)^N = \sum_{k=0}^{N} P(k)$

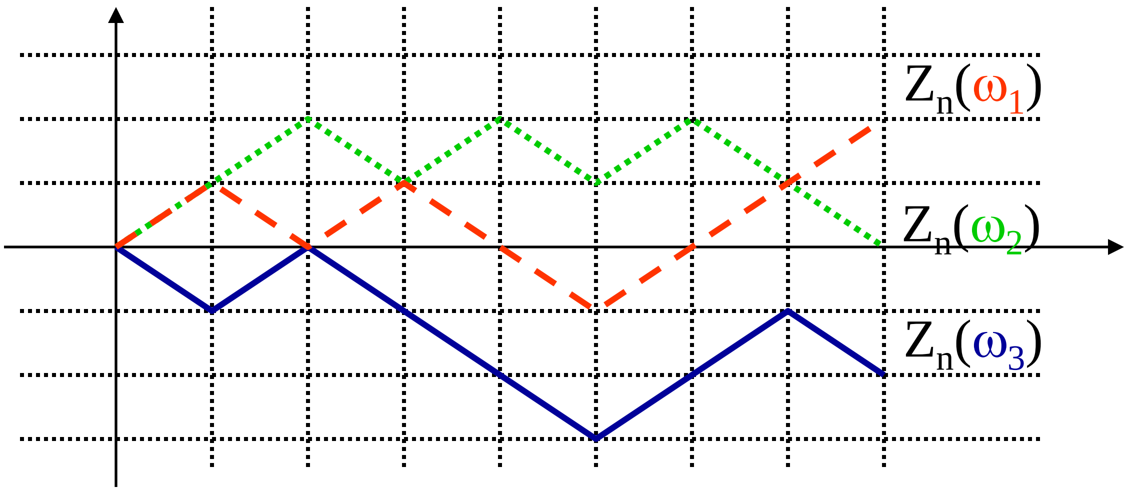
  $= \sum_{k=0}^{N} \binom{N}{k} p^k (1-p)^{N-k}$

Random Processes (Stochastic Processes)

- A random process can be described as a *collection of random variables* parameterized by time index *t*.
- Continuous random process $\{x_t, t \in [0, \infty]\}$
  - For a fixed *t*, x_t is a random variable.
- Discrete-time random process $\{x_k\}$, such that $x_1, x_2, \ldots, x_k, \ldots$
  - Again, each x_k is a random variable.
- Ex) Flipping a coin repeatedly x_k = 1 with prob. p

  or -1 with prob. 1-p
- Ex2) $Z_n := \sum_{k=1}^{n} X_k$

Random Processes (Stochastic Processes)



Each is called a ***sample path***.

Ensemble: Collection of every possible sample paths

- Suppose we observe a path being taken by the random process Z_n in 8 steps.
- There are 2^8 possible paths. This collection is called *ensemble*.
- In an observation, Z_n takes a particular path. It is called a *sample path* taken by the random process in an experiment.
- We may interpret it as an outcome of a random experiment: choosing one object out of 2^8 objects.
- We use $Z_n(\omega)$ to denote a particular sample path.

Stationary Processes (Strict Sense)

- A random process $x(t)$ is said to be stationary to the order N, if for any $t_1, t_2, \ldots, t_N$,
  $f_x(x(t_1), x(t_2), \ldots, x(t_N)) = f_x(x(t_1+t_0), x(t_2+t_0), \ldots, x(t_N+t_0))$
  where t_0 is any arbitrary real constant.
- That is, the joint distribution function is shift-invariant in time.
- If this holds for any N, then we say the process is *strictly stationary*.

Ergodic Random Process (Important)

- If time average ´ ensemble average, then ergodic.
- A random process is said to be *ergodic* if the time average of any sample path is equal to the ensemble average (expectation).
  - $E(x(t)) = \lim_{T\to\infty} (1/T) \int_T x(t)\, dt$ (ergodic in mean)
  - $E(x^2(t)) = \lim_{T\to\infty} (1/T) \int_T x^2(t)\, dt$ (ergodic in 2^{nd} moment)
- An ergodic process must be a stationary process (but not vice versa).
  - If a process is non-stationary, then the ensemble average of the process changes over time.
  - Not all stationary processes are ergodic.
    - Select a coin from a box containing two coins with different weight in a box and throw them repeatedly.

Example of Ergodic Processes

- Show that $x(t) = \cos(2\pi f_0 t + \theta)$ is ergodic in mean and 2nd moment, where θ is uniformly distributed over $[0, 2\pi]$.
- $E(x(t)) = (1/2\pi) \int_0^{2\pi} \cos(2 \pi f_0 t + \theta)\, d\theta$

  $= (1/2\pi) \sin(2 \pi f_0 t + \theta) |_0^{2\pi}$

  $= 0$
- $E(x^2(t)) = (1/2\pi) \int_0^{2\pi} \cos^2(2 \pi f_0 t + \theta)\, d\theta$

  $= (1/2\pi) (1/2) \int_0^{2\pi} 1 + \cos(2 \pi 2f_0 t + 2\theta)\, d\theta$

  $= 1/2$
- $T_0 = 1/f_0$
- $<x(t)> = (1/T_0) \int_0^{T0} \cos(2\pi f_0 t + \theta)\, dt = 0$
- $<x^2(t)> = (1/T_0) \int_0^{T0} \cos^2(2\pi f_0 t + \theta)\, dt$

  $= (1/2T_0) \int_0^{T0} 1+\cos(2\pi 2f_0 t + 2\theta)\, dt = 1/2$

HW#0

- Complete the following problems and submit by the next lecture.
- Will be checked, but not graded.

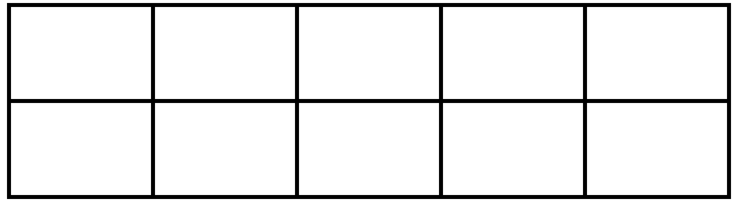
Mutually Exclusive vs. Independence

- The events A and B are mutually exclusive. Can they be independent?

Probability/Random Variable/Distribution

- A coin with Pr{tail} = p is tossed n times.
  - (a). Find the probability of the event that shows k heads in n trials.
  - (b). What is the conditional probability that the first toss is head given that there are 2 heads in n tosses?
  - (c). Let X be the random variable denoting the number of heads. Specify the domain and the range of this random variable.
  - (d). Sketch the cumulative distribution function of X for n = 6. Assume p = 0.1.

Probability



- Consider a box shown above. It has 10 pockets. Two balls are thrown into the box in sequence. A ball can be placed in any pocket with equal probability. No pocket can hold two balls. No balls can be placed outside the box.
  - (a) What is the probability that both balls are placed into the same column?
  - (b) What is the probability that both balls are placed into the same row?
  - (c) What is the probability that the two balls are separated into both different row and different column?
  - (d) Is there any other case? Justify your answer.

Joint distribution/conditional probability

- Two i.i.d. (indep. identically distr.) binary random variables, X_1 and $X_2 \in \{1,-1\}$ with p and (1-p). What's the conditional probability $\Pr(X_1=1|X_2=1)$?
- Now consider a series of binary random variables, X_1, X_2, X_3, ….. X_1 produces equally likely outcomes, the second and the rest are i.i.d. random variables producing the outcome 1 with probability *p* and outcome -1 with probability (1-*p*) where *p* is a number between zero and 1. The number *p* is determined at the first experiment. p is 1/2 if X_1= 1 or 1/4 if X_1=-1.
  - What is $\Pr\{X_4 = 1\}$?
  - What about $\Pr\{X_1 + X_2 = 2\}$?

Joint distribution/conditional probability

- In this problem, θ, U, V, e_1 and e_2 are all binary {0, 1} random variables. Let's use notation $P_\theta = \Pr(\theta=1)$, and thus $\Pr(\theta = 0) = 1 - P_\theta$. The same goes for the other random variables. For example, $P_{e1} = \Pr(e_1 = 1)$, and $P_{e2} = \Pr(e_1 = 1)$.
- Suppose U and V are binary random variable, i.e.

  $$U = \theta + e_1 \text{ modulo } 2$$

  $$V = \theta + e_2 \text{ modulo } 2$$

  where $P_\theta = p$, $P_{e1} = p_1$ and $P_{e2} = p_2$, and θ, e_1 and e_2 are mutually independent.

1. For $P_\theta = 0.6$, $P_{e1} = 0.1$ and $P_{e2} = 0.2$, find the joint distribution $\Pr(U = x, V=y)$.
2. Repeat 1 with $P_\theta = 0.9$, $P_{e1} = 0.01$ and $P_{e2} = 0.02$.

Urn Problem

- A box contains m white balls and n black balls. Balls are drawn at random one at a time without replacement. Find the probability of encountering a white ball by the k-th draw.

Total Probability/Bayes' Theorem

- Suppose there is a test for a prostate cancer which is known to be 95% accurate. A person took the test and the result came out positive. Suppose that the person comes from a population of a million, where 20,000 people suffer from that disease. What can we conclude about the probability that the person under test has that particular cancer.