CSCE 619-600
Networks and Distributed Processing
Spring 2017

Review of Probability

Dmitri Loguinov
Texas A&M University

January 19, 2017

Agenda

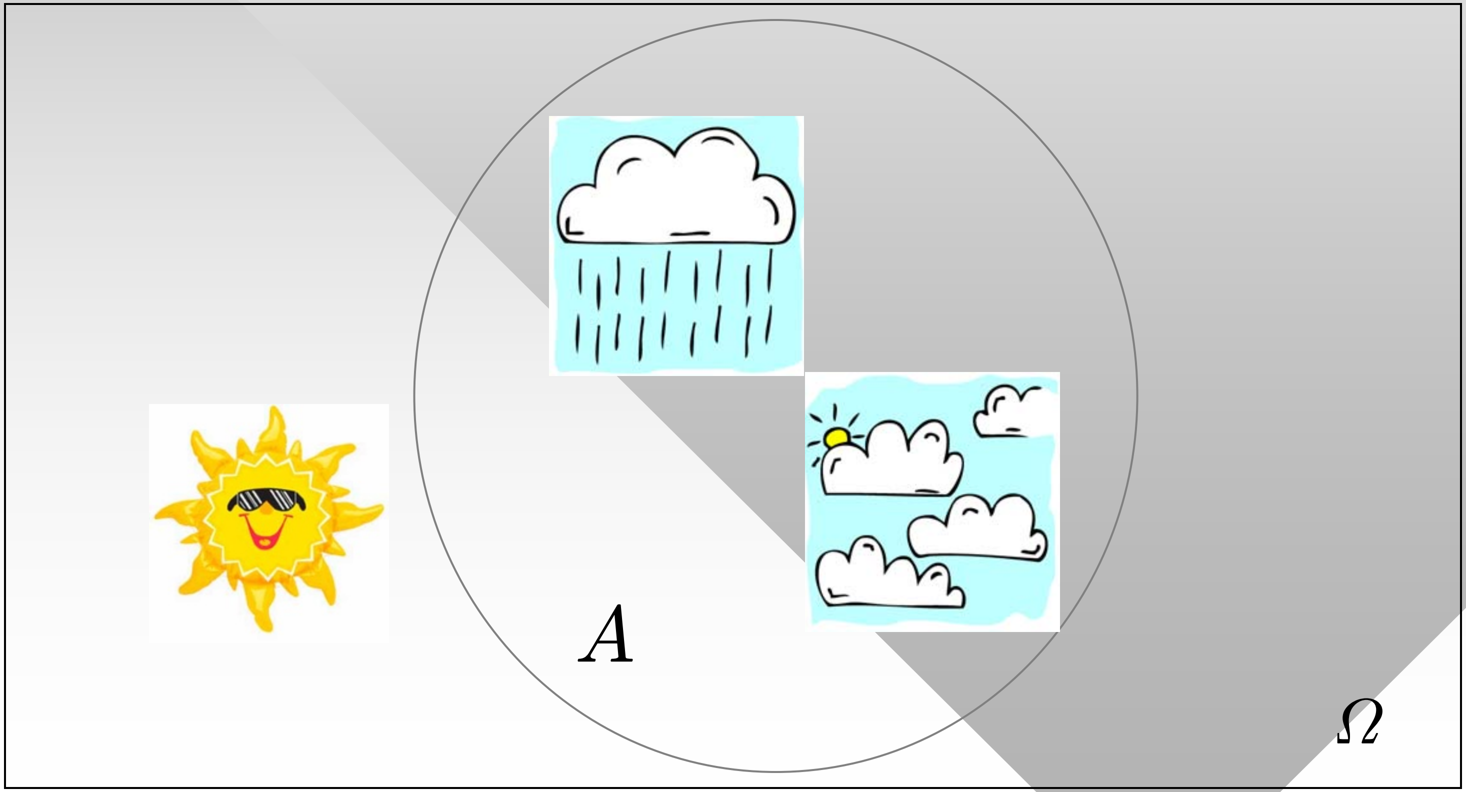
- Probability space
- Probability measure
- Random variables
- Distributions
  - Discrete
  - Continuous
- Wrap-up

Probability Space

- Definition: probability space Ω is the set of all possible outcomes of a random experiment
  - We use ω to denote the random outcome of a particular experiment
- Definition: event A is a subset of Ω: $A \subseteq \Omega$
  - Thus, A may contain multiple outcomes
- Definition: event A occurs if and only if $\omega \in A$
- Probability theory examines the likelihood of events
  - For example, "rainy," "cloudy," and "sunny" are three outcomes of your weather report for today
  - Event A could be "rainy or cloudy"

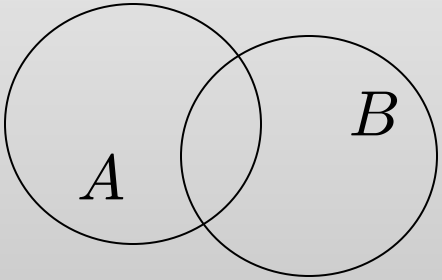
Probability Space 2

- Formalizing the weather example
  - $\Omega = \{$rainy, cloudy, sunny$\}$, $A = \{$rainy, cloudy$\}$
- Q: how many events in this probability space?



Probability Space 3

- It is often convenient to write A^c for the complement of A: $A^c \cup A = \Omega$
- Basic set theory applies to events
  - Use Venn diagrams to show the following



$$A \cap A = AA = A$$

$$AB = BA$$

$$A(B \cup C) = AB \cup AC$$

$$(AB)^c = A^c \cup B^c$$

$$A \cup (BC) = (A \cup B)(A \cup C)$$

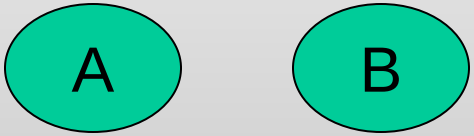
$$(A \cup B)^c = A^c B^c$$

Probability Measure

Think of probability as the area of an event

- Definition: probability measure $P(A)$ is a function that maps events to real numbers and satisfies 3 axioms of probability:

  1) $P(A) \geq 0$
  2) $P(\Omega) = 1$
  3) If $AB = \varnothing$, then $P(A \cup B) = P(A) + P(B)$



- (side note) For infinite sets, axiom 3 is usually strengthened to *countable additivity*:
  - for any set $A_1, A_2, \ldots$ such that $A_i A_j = \emptyset, i \neq j$
  - we have $$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Probability Measure 2

1) $P(A) \geq 0$
2) $P(\Omega) = 1$
3) If $AB = \emptyset$, $P(A \cup B) =$
$P(A)+P(B)$

- Exercise 1: show $P(A^c) = 1 - P(A)$
  - Proof (noticing that A and A^c do not overlap and applying axioms 3 and 2):

$$P(A^c) + P(A) \;=\; P(A \cup A^c) = P(\Omega) = 1$$

- Exercise 2: show $C \subseteq A \Rightarrow P(A \backslash C) = P(A) - P(C)$
  - Proof (same reasoning, axiom 3):

$$P(A \backslash C) + P(C) = P(A \backslash C \cup C) = P(A)$$

- Exercise 3: show $P(A \cup B) = P(A) + P(B) - P(AB)$
  - Proof:

$$\begin{aligned} P(A \cup B) &= P(A \cup B \backslash AB) = P(A) + P(B \backslash AB) \\ &= P(A) + P(B) - P(AB) \end{aligned}$$

Probability Measure 3

- Definition: any two non-intersecting events A and B are called *mutually exclusive*
  - A set of events $\{A_i\}$ is called *pair-wise mutually exclusive* if for all $i \neq j$: $A_i A_j = \varnothing$
- Definition: the probability of observing event A given that B has occurred is called conditional probability $P(A|B)$, which is defined as

$$P(A|B) = \frac{P(AB)}{P(B)}, P(B) > 0$$

- Example: what is the probability that student X will attend the class given that Y showed up?

X				X		X	X		
Y	Y			Y	Y	Y			Y

Probability Measure 4

- Definition: set of events $\{A_i\}$ is called *exhaustive* if

$$\bigcup_{i=1}^{n} A_i = \Omega$$

  - If in addition set $\{A_i\}$ is pair-wise mutually exclusive, then it is called a *partition* of Ω
- Conditional probability is a useful tool in practice
  - Observe that if $\{A_i\}$ is exhaustive, then B can be decomposed as:

$$B = \bigcup_{i=1}^{n} BA_i$$

  - and if additionally $\{A_i\}$ is a partition:

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Probability Measure 7

- Inverting conditional probability
  - Suppose we know $P(B|A)$ and need to find out $P(A|B)$

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(BA)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

  - This is also known as Bayes Theorem
- Task: compute $P(Y|X)$ for the students

X				X		X	X		
Y	Y			Y	Y	Y			Y

Independence

- Definition: events A and B are independent if and only if $P(A \mid B) = P(A)$
- From the definition:

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

- Therefore for independent events:

$$P(AB) = P(A)P(B)$$

- Task: show that if A and B are independent, then so are these pairs of events:
  - (B, A), (A, B^c)
  - (A^c, B), (A^c, B^c)

hints:

$$AB^c = A \backslash AB$$

$$A^cB^c = (A \cup B)^c$$

Random Variables

- For convenience, we would rather work with numbers than actual outcomes ω
- Definition
  - A random variable $X(\omega)$ is a real-valued function that maps Ω to $\mathcal{R}$
- One of the simplest random variables is an indicator of event A

$$X(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

- Often ω is omitted

$$X = \begin{cases} 1 & A \text{ happens} \\ 0 & \text{otherwise} \end{cases}$$

Random Variables 2

- Once we know that the set where X assumes values below x belongs to $\mathcal{F}$, we can define the *cumulative distribution function* (CDF) $F(x)$

$$F(x) = P(\{\omega : X(\omega) \leq x\})$$

- It is the probability that outcome ω is such that the value of $X(\omega)$ is no more than x
  - We usually write:

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

- $F(x)$ is non-decreasing with $F(\infty) = 1$ and $F(-\infty) = 0$
  - Finally define the tail distribution F^c:

$$F^c(x) = P(X > x) = 1 - F(x)$$

Random Variables 3

- Definition: a random variable is discrete if it assumes values from some countable (possibly infinite) set
  - Assume the set consists of x_1, x_2, …
  - Then in addition to the CDF, we often use the *probability mass function* (PMF): $p(i) = P(X = x_i)$
- Clearly, the following holds for all discrete distributions:

$$\sum_{i=1}^{\infty} p(i) = 1$$

- One example of discrete X is the Bernoulli random variable:

$$X = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$$

Random Variables 4

- Example
  - Coin toss where probability of a head is p
  - Outcome ω of each toss is either heads or tails
  - Define $X(\text{head}) = 1$, $X(\text{tail}) = 0$
  - Then X is a Bernoulli variable (if $p =$ ½, the coin is called fair)
- Define Z to be the number of independent tosses before we get the first occurrence of heads
  - What is the PMF of Z?
- To get the first head on toss k, we clearly must sit through $k-1$ tails:

$$P(Z = k) = (1-p)^{k-1}p$$

Random Variables 5

- What we just defined is the *geometric* distribution
- In the next example, we define Y to be the number of heads that come out in n independent tosses
  - Find out $P(Y = k)$
- We need the probability of k heads and $n - k$ tails
  - If we know the position of each head and tail, then:

$$P(Y = k, \text{fixed set of heads}) = p^k(1-p)^{n-k}$$

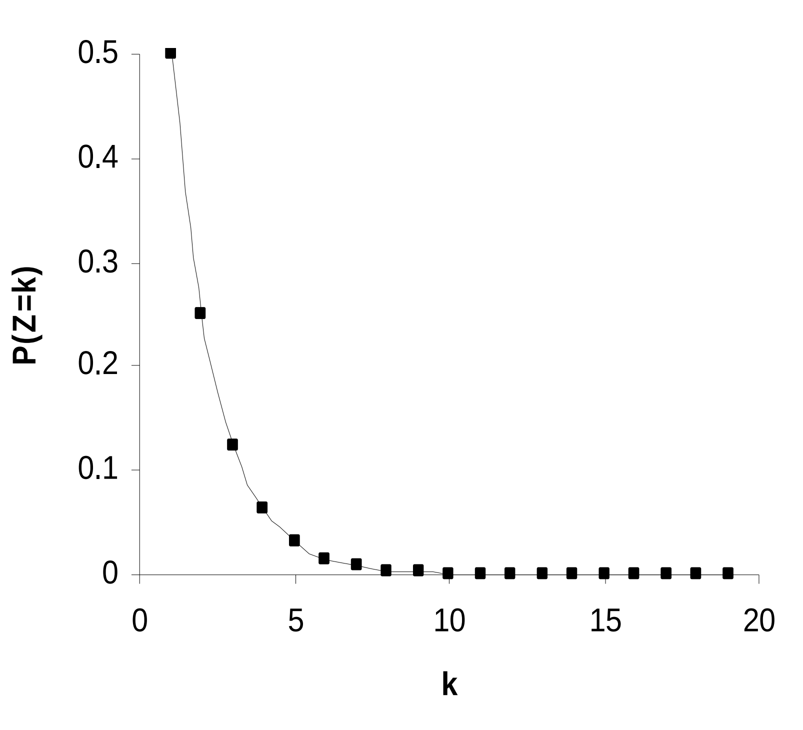
  - Accounting for all possible permutations:

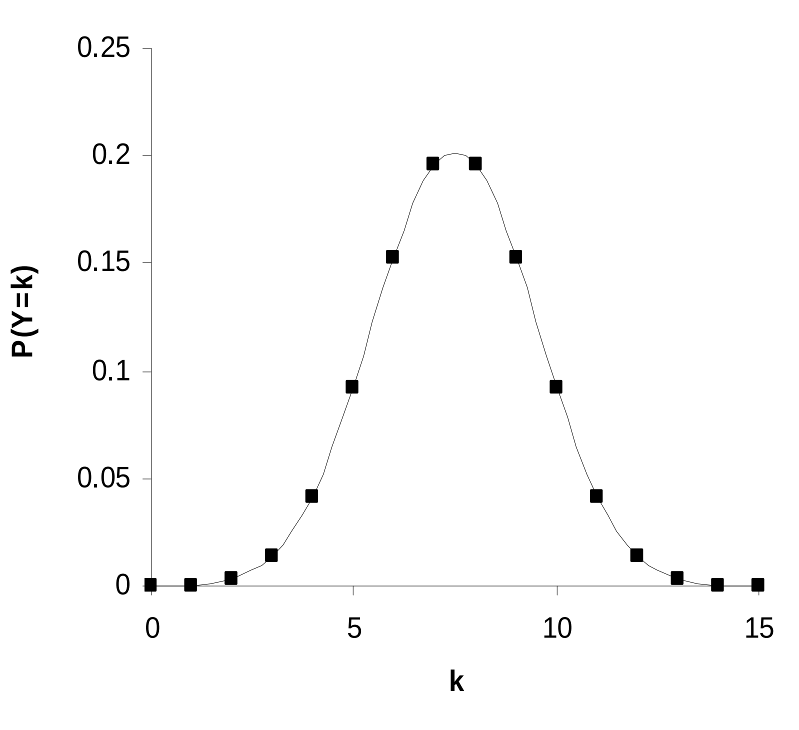
$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

  - we get the binomial distribution

Random Variables 6

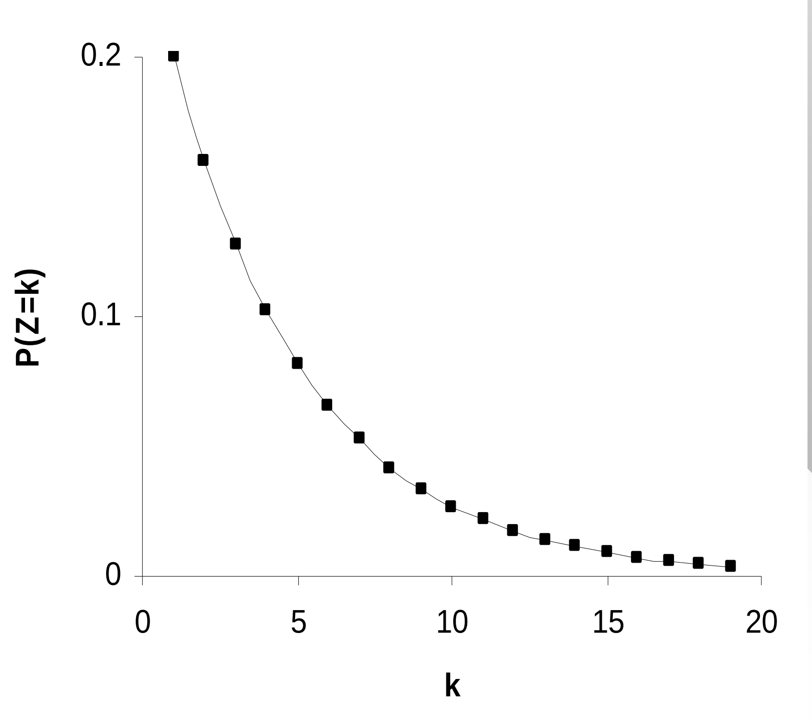
- What do these distributions look like?
  - We can directly plot the value of $P(Y = k)$ for each k
  - Examples below use $p =$ ½ and $n = 15$

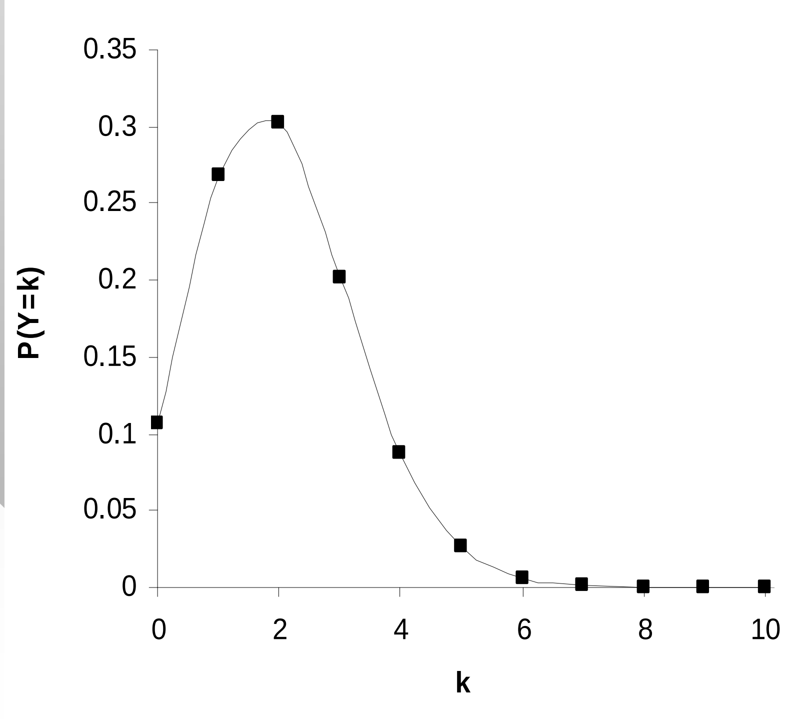




Random Variables 7

- Here is another example for $p = 0.2$ and $n = 10$
  - Notice the change in shape for the binomial distribution
  - For large n, it tends to the Gaussian (previous slide) or Poisson (this slide) distribution





Random Variables 8

- Definition:
  - If $P(X = x) = 0$ for all x (or the cardinality of the set of its values is continuum), then X is said to be *continuous*
- In such cases, we assume that $F(x)$ is differentiable and call its derivative the *density* (PDF) of X

$$f(x) = F'(x)$$

- Note certain properties of $f(x)$ and $F(x)$:

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

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

Random Variables 9

- Also notice that we can differentiate the tail to obtain the density

$$f(x) = \frac{dF(x)}{dx} = \frac{d(1 - F^c(x))}{dx} = -\frac{dF^c(x)}{dx}$$

- Next define some useful distributions
  - Uniform in $[a,b]$:

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

  - Exponential defined in $[0, \infty)$:

$$f(x) = \lambda e^{-\lambda x}, F(x) = 1 - e^{-\lambda x}$$