Review of Probability

Dmitri Loguinov
Texas A&M University

January 19, 2017
Agenda

- Probability space
- Probability measure
- Random variables
- Distributions
  - Discrete
  - Continuous
- Memoryless property
- Wrap-up
**Probability Space**

- **Definition:** probability space $\Omega$ is the set of all possible outcomes of a random experiment
  - We use $\omega$ to denote the random outcome of a particular experiment
- **Definition:** event $A$ is a subset of $\Omega$: $A \subset \Omega$
  - 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
  - $\Omega = \{\text{rainy, cloudy, sunny}\}$, $A = \{\text{rainy, cloudy}\}$
- Q: how many events in this probability space?
• 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

\[
A \cap A = AA = A \\
AB = BA \\
A(B \cup C) = AB \cup AC \\
(AB)^c = A^c \cup B^c \\
A \cup (BC) = (A \cup B)(A \cup C) \\
(A \cup B)^c = A^cB^c
\]
### Probability Measure

**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 $AB = \emptyset$, 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 $A_1, A_2, \ldots \subseteq \mathcal{F}, A_iA_j = \emptyset, i \neq j$
- we have $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$
Probability Measure 2

• **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 \( C \subseteq A \Rightarrow P(A \setminus C) = P(A) - P(C) \)
  
  - Proof (same reasoning, axiom 3):
    \[
    P(A \setminus C) + P(C) = P(A \setminus C \cup C) = P(A)
    \]

• **Exercise 3**: show \( P(A \cup B) = P(A) + P(B) - P(AB) \)
  
  - Proof:
    \[
    P(A \cup B) = P(A \cup B \setminus AB) = P(A) + P(B \setminus AB)
    = P(A) + P(B) - P(AB)
    \]


**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 $i \neq j$: $A_iA_j = \emptyset$

- **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?

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th></th>
<th>X</th>
<th>X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>
• **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 \( \Omega \)

• **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

• Furthermore, if set $\{A_i\}$ is a partition:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_i P(B|A_i)P(A_i)}$$
**Independence**

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

- From the definition:

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

- 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: $AB^c = A \setminus AB$

  $A^cB^c = (A \cup B)^c$
**Random Variables**

- For convenience, we would rather work with numbers than actual outcomes $\omega$.

**Definition**

- A random variable $X(\omega)$ is a real-valued function that maps $\Omega$ to $\mathbb{R}$.

- 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 $\omega$ 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 $\omega$ 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 \( x_1, x_2, \ldots \)
  - 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:
  \[
  X = \begin{cases} 
  1 & \text{w.p. } p \\ 
  0 & \text{w.p. } 1 - p 
  \end{cases}
  \]
• Example
  - Coin toss where probability of a head is $p$
  - Outcome $\omega$ of each toss is either heads or tails
  - Define $X(\text{head}) = 1$, $X(\text{tail}) = 0$
  - Then $X$ is a Bernoulli variable (if $p = \frac{1}{2}$, 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 \( p = \frac{1}{2} \) and \( n = 15 \)
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 \( F(x) \):

\[
F(t) = \int_{-\infty}^{t} f(x) \, dx
\]

\[
P(a \leq X \leq b) = F(b) - F(a) = \int_{a}^{b} f(x) \, dx
\]
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 \([a, b] \):
    \[ f(x) = \frac{1}{b - a}, \quad F(x) = \frac{x - a}{b - a} \]
  - Exponential:
    \[ f(x) = \lambda e^{-\lambda x}, \quad F(x) = 1 - e^{-\lambda x} \]
Memoryless Distributions

• We can now demonstrate how exponential distributions “forget” memory

• Suppose inter-bus delays are exponential and we know that the last bus left $t$ time units ago
  - What is the distribution of the remaining waiting time?

• Define $X$ to be the inter-bus delay and $W = X - t$ to be the wait time
  - Then we have:

$$P(W > x|X > t) = P(X - t > x|X > t)$$
$$= P(X > x + t|X > t)$$
Memoryless Distributions 2

• Rewriting:

\[
P(W > x | X > t) = \frac{P(X > x + t, X > t)}{P(X > t)} = \frac{F^c(t + x)}{F^c(t)}
\]

• Now substituting the tail of the exponential distribution, we have complete independence of \( t \):

\[
P(W > x | X > t) = \frac{e^{-\lambda (t+x)}}{e^{-\lambda t}} = e^{-\lambda x}
\]
Memoryless Distributions 3

• This means that regardless of how long ago the last bus left, the delay to the next bus is only determined by $x$
  
  Thus, $W$ is another exponential distribution with the same mean, which explains why the wait was always 20 minutes in the previous lecture.

• Definition
  
  A distribution is called memoryless if and only if

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

• This means that the current “age” of $X$ does not affect its remaining life.
Wrap-up

Using the last equation, observe:

\[ P(X > t + x) = P(X - t > x | X > t)P(X > t) = F^c(x)F^c(t) \]

Since the left side is \( F^c(t+x) \), we have an equivalent definition of memoryless distributions:

\[ F^c(t + x) = F^c(x)F^c(t) \]

Exercise: show that the exponential distribution is the only memoryless distribution.
Practice

• Driver decisions
  – Your gas tank holds 20 gallons, but fuel gauge is broken
  – You drive around and all of a sudden see a good deal
  – If filling up \( X \) gallons and \( X < 10 \), you pay \( X^2 \) dollars, otherwise \( 1.5X \) dollars; but have to fill the tank to the max
  – You only have $20 on you
  – What is the likelihood of successful fill-up?

• Suppose \( Y \) is the cash due if you go for it

\[
P(Y < 20) = P(Y < 20|X < 10)P(X < 10) + P(Y < 20|X \geq 10)P(X \geq 10)
\]
Practice 2

• Assuming a uniform distribution for $X$

$$P(Y < 20) = P(0 \leq X < \sqrt{20}|X < 10)\frac{1}{2}$$

$$+ P(10 \leq X < 20/1.5|X \geq 10)\frac{1}{2}$$

• This can be easily computed as

$$P(Y < 20) = \frac{\sqrt{5} + 5/3}{10} \approx 0.3903 \cdots$$

• Graphical explanation only valid when $\omega$ is uniform