BIOS 660/BIOS 672 (3 Credits): Probability and Statistical Inference I

Jianwen Cai

https://sakai.unc.edu/portal/site/bios660-bios672-3-credits

Notes 6

Random Variables **2**
- Random Variables 3
- Random Variables (Formal Definition) 4
- Conventions 5
- Examples 6

Distribution Functions **7**
- Cumulative Distribution Functions 8
- Some Properties of $F(y)$ 9
- Induced Probability Space 10
- Identically distributed rvs 11
- Types of Random Variables 12

Discrete random variables **13**
- Discrete random variables 14
- Probability mass function 15
- Properties of the pmf 16
- Example 17

Continuous Random Variables **18**
- Continuous Random Variabes 19
- Absolute Continuity 20
- cdf — density relation 21
- Properties 22
- Notes 23
- Notes (cont.) 24
- Notes (cont.) 25
- Stochastic ordering 26

Random Variables

Suppose we start with a probability space $(\Omega, \mathcal{A}, P)$. Instead of referring to outcomes and events observed from the sample space Ω, it is often convenient to assign a number to each possible outcome and record that instead.
Example: Flip a coin three times.

- $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$
- $\mathcal{A} = 2^\Omega$
- Define a random variable Y to be the number of heads.
- $Y(HHH) = 3, Y(HHT) = 2, Y(HTT) = 1$, etc.

Random Variables (Formal Definition)

Definition: A random variable Y is a real-valued and measurable function defined on a probability space. That is, $Y : \Omega \to \mathbb{R}$.
Every point ω in Ω maps to a point in $\mathbb{R}$, namely $Y(\omega)$.
Conversely, we define the *inverse image* under Y of a subset B of $\mathbb{R}$ as

$$Y^{-1}(B) = \{\omega : Y(\omega) \in B\}$$

The definition of a random variable requires that the inverse image of every Borel set $B \subset \mathbb{R}$ is an element of $\mathcal{A}$. This property allows us to assign probabilities to random variables. More precisely,

$$P\{Y \in B\} = P\{Y^{-1}(B)\}$$

Conventions

A Random Variable is a set function which takes values on the real line (for now). Often the argument is omitted and one writes Y instead of $Y(\omega)$.

Random variables are usually denoted by capital letters (e.g. Y).
Values which random variable can take on are denoted by lower case letters (e.g. y).

Example: Coin Tosses

$$\Omega : \{H, T\}$$
$$Y(H) = 1, \quad Y(T) = 0$$

If $P(\text{head}) = .5$, then $P(Y = 1) = .5$

Examples

Roll of a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
Define $Y(\omega) = \omega$
$P\{Y(\omega) = \omega\} = P(\omega) = 1/6$

An artificial example:
Suppose $\Omega = \{\omega : 0, \ \pm 1, \ \pm 2, \ \ldots\}$
Define

$$Y(\omega) = a \quad \text{if} \quad \omega \leq 0$$
$$Y(\omega) = b \quad \text{if} \quad \omega > 0$$
$$P\{Y = a\} = \sum_{i=-\infty}^{0} P(i) \quad ,$$
$$P\{Y = b\} = \sum_{i=1}^{\infty} P(i)$$

Cumulative Distribution Functions

Distribution Functions are used to describe the behavior of a rv.

Definition The *cumulative distribution function* (*cdf*) of a random variable Y is a real valued function $F(y)$ defined by

$$F(y) = P\{Y \leq y\} = P\{\omega : Y(\omega) \leq y\}$$

Example: cdf of a die: $F(y) = y/6$

Definition The *survival function* of Y is defined by

$$S(y) = 1 - F(y) = P(Y > y)$$

Some Properties of $F(y)$

1. $0 \leq F(y) \leq 1$
2. $\lim_{y \to -\infty} F(y) = 0$
3. $\lim_{y \to \infty} F(y) = 1$
4. F is nondecreasing: i.e. if $a < b$, then $F(a) \leq F(b)$
5. F is right continuous: that is, for any b and any decreasing sequence $b_n, n \geq 1$ that converges to $b, \lim_{n \to \infty} F(b_n) = F(b)$
6. $P\{a < Y \leq b\} = F(b) - F(a)$

These properties can all be proved using the properties of probability measures.

Induced Probability Space

All probability questions about a random variable can be answered via its *cdf*.
Every random variable defined on a probability space induces a probability space on $\mathbb{R}$:

$$(\Omega, \mathcal{A}, P) \longrightarrow \quad Y(\omega) \longrightarrow \quad (\mathbb{R}, \mathcal{B}, F(\cdot))$$

- Points in Ω are transformed to points on $\mathbb{R}$ (Real line)
- Sets (events in $\mathcal{A}$) are mapped into intervals on real line, i.e., into members of the Borel sets, $\mathcal{B}$.
- P is replaced by $F(\cdot)$.

Because of this, the abstract notion of a sample space recedes, and attention is usually given primarily to random variables and their distributions.
We will sometimes refer to the ‘sample space’ of a random variable, which will be taken to be the values in $\mathbb{R}$ that a random variable takes on.

Identically distributed rvs

The cdf does not contain information about the original sample space.

Example: Toss a coin n times. The number of heads and number of tails have the same distribution.

Definition: Two rvs X and Y are identically distributed if for every Borel set $A \subset \mathbb{R}$,
$P(X \in A) = P(Y \in A)$.

Theorem C&B 1.5.10 The following two statements are equivalent:

a. The rvs X and Y are identically distributed
b. $F_X(x) = F_Y(x)$ for every x.

Note that two rvs can have the same distribution even if they are not equal to one another.

The distinction between two rvs being equal and having the same distribution will become important later in questions of convergence.

Types of Random Variables

A random variable Y can be

- *discrete* - Y takes on a finite or countably infinite number of values
- *continuous* - the range of Y consists of subsets of the real line.
- *mixed* - best to see this with an example

Example of a mixed random variable:
Consider a random variable with *cdf* given by:

$$F(x) = \begin{cases} 0 & x < 0 \\ x/2 & 0 \leq x < 1 \\ 2/3 & 1 \leq x < 2 \\ 11/12 & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

Discrete random variables

Discrete random variables

If a random variable Y takes only a finite or countable number of values, then the *cdf* can be expressed as:

$$F(y) = P\{Y \leq y\} = \sum_{z \leq y} P\{Y = z\}$$

If the sample space of Y is $\Omega = \{y_1, y_2, ...\}$, then

$$F(y) = \sum_{y_i \leq y} P\{Y = y_i\}$$

Probability mass function

Definition The *prob. mass function* (*pmf*) or *frequency function* is a function $f(y)$ defined by

$$f(y) = P\{Y(\omega) = y\}$$

Thus, we can write $F(y) = P\{Y \leq y\} = \sum_{z \leq y} f(z)$.

If the sample space of Y is $\Omega = \{y_1, y_2, ...\}$, then

$$f(y_i) = P(Y = y_i) = P(y_{i-1} < Y \leq y_i) = F(y_i) - F(y_{i-1})$$

Example: Suppose Y is a random variable that takes the values 0, 1 or 2 with probability .5, .3, and .2, respectively. Then

$f(0) = 0.5, f(1) = .3$, and $f(2) = .2$.

Properties of the pmf

Definition: The *domain* of a random variable Y is the set of all values of y for which $f(y) > 0$.

Properties of the pmf:

1. $f(y) > 0$ for at most a countable number of values y. For all other values y, $f(y) = 0$.
2. Let $\{y_1, y_2, \ldots\}$ denote the domain of Y. Then

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

An obvious consequence is that $f(y) \leq 1$ over the domain.

Example: What is the pmf of a deterministic rv (a constant)?

$f(x) = 1$ for $x = k$ and $f(x) = 0$ for $x \neq k$.

Example

In many applications, a formula can be used to represent the *pmf* of a random variable. Suppose Y can take values 1,2 ... with *pmf*

$$f(y) = \begin{cases} \frac{1}{y(y+1)} & y = 1, 2, ... \\ 0 & \text{otherwise} \end{cases}$$

How would we determine if this is an allowable *pmf*?

$$\begin{aligned} \sum_{y=1}^{\infty} \frac{1}{y(y+1)} &= \sum_{y=1}^{\infty} \left(\frac{1}{y} - \frac{1}{y+1} \right) \\ &= \left(1 - \frac{1}{2} \right) + \left(\frac{1}{2} - \frac{1}{3} \right) + \cdots \\ &= 1 - \lim_{y \to \infty} \frac{1}{y} \\ &= 1 \end{aligned}$$

Continuous Random Variables

Continuous Random Variabes

Recall that a random variable is a function that maps a probability space $\{\Omega, \mathcal{A}, P\}$ to the real line, thereby inducing a new probability space:

$$\{\Omega, \mathcal{A}, P\} \longrightarrow \quad Y \longrightarrow \quad (\mathbb{R}, \mathcal{F}, F(\cdot))$$

We have discussed the setting where Y is discrete (i.e. Y can take on a finite or countably infinite number of values).
A random variable Y is called *continuous* if its distribution function $F(y) = P(Y \leq y)$ is a continuous function.

Absolute Continuity

The distribution of a continuous random variable is characterized by the probability of falling in intervals, e.g. $P(Y \in (a, b])$.
We will focus on *absolutely continuous* random variables.

Definition: A function $F(y)$ is *absolutely continuous* if it can be written

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

where for now, you may interpret $\int$ as the usual Riemann integral.

A random variable is said to be absolutely continuous if its distribution function is absolutely continuous.

Note: Absolute continuity is stronger than continuity but weaker than differentiability. An example of an absolutely continuous function is one that is differentiable everywhere except for a countable number of points.

cdf — density relation

If $F(y)$ is absolutely continuous, $f(y)$ is called the *probability density function (pdf)* of Y and

$$F'(y) = \frac{dF(y)}{dy} = f(y).$$

Building on this idea,

$$P(a < Y \leq b) = F(b) - F(a) = \int_{a}^{b} f(x)dx$$

More generally, for a set B,

$$P(Y \in B) = \int_{B} f(x)dx$$

Note that of course B has to be an “allowable” subset of the real line $\mathbb{R}$, that is, a Borel set.

Properties

In general, a function $f(x)$ is a *pdf* iff

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

Examples:

- Suppose $F(x) = 1 - e^{-\lambda x}$ for $x > 0$ and $F(x) = 0$ otherwise. Is $F(x)$ a cdf? What is the associated pdf?
  - $F(x)$ is continuous and nondecreasing, and $lim_{x \to -\infty} F(x) = 0, \;\; lim_{x \to +\infty} F(x) = 1$, so $F(x)$ is a cdf.
  - $f(x) = \frac{d}{dx} F(x) = \lambda e^{-\lambda x}$
- What about $f(x) = 1/x^r$ for $x > 1$ and $f(x) = 0$ otherwise?

$$\int_1^{\infty} \frac{dx}{x^r} = \left[\frac{x^{1-r}}{(1-r)} \right] = \frac{1}{r-1}$$

  Then $f(x)$ is not a pdf but $(r-1)f(x)$ is a pdf.

Notes

- $f(x)$ is not the probability that $Y = x$. In fact, if Y is an absolutely continuous random variable with density function $f(x)$, then $P(Y = x) = 0$. Why?

$$\begin{aligned} P(Y = x) &= \lim_{h \to 0} \int_{x-h}^{x+h} f(u)du \\ &= \lim_{h \to 0} [F(x+h) - F(x-h)] \\ &= F(x+) - F(x-) \\ &= 0 \end{aligned}$$

Notes (cont.)

- More generally, if B is a subset of $\mathbb{R}$ with

$$\int_B dx = 0,$$

  then if Y is an absolutely continuous random variable defined on $\mathbb{R}$, then $P(Y \in B) = 0$ also.
- Because $P(Y = a) = 0$, all the following are equivalent:

$$P(a \le Y \le b), \quad P(a \le Y < b) \quad \text{and} \quad P(a < Y < b)$$

- Also, note that $f(x)$ can exceed one!

Notes (cont.)

- $f(x)$ can be interpreted as the *relative* probability that Y takes the value x. Why? By the mean value theorem, we can say

$$P(x < Y \le x + \Delta) \approx f(x)\Delta$$

  Thus

$$P(Y \in \text{interval of width } \Delta \text{ centered at } a) \approx f(a)\Delta$$

  and

$$P(Y \in \text{interval of width } \Delta \text{ centered at } b) \approx f(b)\Delta$$

  Hence, if $f(b) > f(a)$, we can say that it is more likely for Y to take the values near b rather than near a.

Stochastic ordering

Suppose Y is a rv and define $X = Y + 2$. Then $X > Y$ always.

Now suppose

$$X \sim F_X(t) = (1 - e^{-t})\, 1(t > 0)$$
$$Y \sim F_Y(t) = (1 - e^{-2t})\, 1(t > 0)$$

Then X is not always greater than Y, but it is likely to be.

Definition: X is *stocastically greater* than Y if

$$F_X(t) \leq F_Y(t) \text{ for all } t$$
$$F_X(t) < F_Y(t) \text{ for some } t$$

or equivalently

$$P(X > t) \geq P(Y > t) \text{ for all } t$$
$$P(X > t) > P(Y > t) \text{ for some } t$$