MA-250 Probability and Statistics

Nazar Khan

PUCIT

Lecture 18

So far we have covered …

1. Random Experiments – processes with uncertain outcomes
2. Sample Space – outcomes of experiments
3. Events
4. Probability – assigns numbers between 0 and 1 to events
5. Independence – P(ABC…)=P(A)P(B)P(C)…

So far we have covered …

6. Random Variables – assign labels to each outcome
   – X(HHH)=3 if random variable X is the number of heads
   – X(HHH)=0 if random variable X is the number of tails
7. Probability **Density** of a random variable

Values of X	0	1	2	3	4
Probabilities	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

← **Labels**

← **Probabilities**

8. Cumulative Probability **Distribution** of a random variable – P(*X*<=*t*)

So far we have covered …

- **Discrete** Random Variables – set of outputs is real valued, countable set
- Now we study **continuous** random variables
  - set of outputs is real valued, uncountable set
  - we can't count, but we can still **measure!**

CONTINUOUS RANDOM VARIABLES

Discrete vs. Continuous

Discrete R.V.	Continuous R.V.
Number of heads in n coin tosses	A number from the interval [a,b] where $a,b \in R$
Year of birth of all students in this class	Exact weight of all students in this class
Number of phone calls per minute at a telephone exchange	Time between successive phone calls at a telephone exchange
Winning time of Olympic 100m races rounded to the nearest 100^{th} of a second.	Exact winning time of Olympic 100m races

Continuous R.V Properties

- Range of continuous R.V is an uncountable set.
- <u>Distribution</u> must obey the **fundamental theorem of calculus**

$$P(s < X \leq t) = F(t) - F(s) = \int_s^t f(x)dx$$

- For any real number *a*,

$$P(X = a) = P(a < X \leq a) = F(a) - F(a) = 0$$

  – Let X be a real number chosen randomly between 5 and 10. Find (i) P(6<X<7)? (ii) P(X=7)?

Deriving a Density

- Suppose we pick a point at random from the interval [3, 14].
  - Sample space is S = [3, 14].
  - X(s)=s i.e., random variable X is the selected point from [3,14]
- Any subinterval of the form X<=t will have length t-3
- For any subinterval A, P(A) = length(A)/length(S) = length(A)/(14-3)
- Therefore, **distribution function F(t)** of random variable X is

$$F(t) = \mathbb{P}(X \leq t) = \begin{cases} 0 & \text{if } t < 3, \\ \frac{t-3}{14-3} & \text{if } t \in [3, 14], \\ 1 & \text{if } t > 14. \end{cases}$$

- By differentiating the distribution function F(t), we get the **density function** of X

$$f(x) = \begin{cases} \frac{1}{14-3} & \text{if } x \in (3, 14), \\ 0 & \text{otherwise.} \end{cases}$$

Example

- Recall that a density function
  - is non-negative
  - with total area 1
- Area from 0 to infinity under $e^{-.01x}$ is given by $\int_0^\infty e^{-.01x}dx = \frac{1}{.01} = 100$
- We can define a **density function** *f(x)* using this result

$$f(x) = \begin{Bmatrix} .01e^{-.01x} & , x \geq 0 \\ 0 & , x < 0 \end{Bmatrix}$$

- The **distribution function** F(t)=

$$F(t) = \begin{Bmatrix} \int_0^t .01e^{-.01x}dx = \left.\frac{e^{-.01x}}{-.01}\right|_0^t = 1 - e^{-.01t} & , t \geq 0 \\ 0 & , t < 0 \end{Bmatrix}$$

Lifespan of Car Windshields

- It has been empirically observed that the time X it takes for a windshield to develop a crack has density $f(x) = \begin{cases} .01e^{-.01x} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$ where X is measured in years.
- The probability that a new windshield will crack within *t* years is $P(X \leq t) = F(t) = 1 - e^{-.01t}$
- Therefore

$$P(X \leq 6\,\text{months}) = F(.5) = 1 - e^{-.01(.5)} = 0.00499$$
$$P(X \leq 100\,\text{years}) = F(100) = 1 - e^{-.01(.5)} = 0.63212$$

- **H.W. Find the constant c > 0 so that the given function is a density of some continuous random variable X. If no such constant exists, explain why.**

(i) $f(x) = cx$, on $[0, 1]$,	(ii) $f(x) = ce^x$, on $[0, \ln 2]$.
(iii) $f(x) = ce^{-3x}$, on $[0, \infty)$,	(iv) $f(x) = ce^{3x}$, on $(-\infty, 0]$.
(v) $f(x) = c/x^5$, on $[1, \infty)$,	(vi) $f(x) = c/x$, on $[1, \infty)$.
(vii) $f(x) = cxe^{-x^2}$, on $[0, \infty)$,	(viii) $f(x) = c/x$, on $[1, e]$.
(ix) $f(x) = c \sin x$, on $(0, \pi)$,	(x) $f(x) = \frac{c}{(1-x)^2}$, on $[0, 1)$.

For functions that are densities of some random variable, compute their distribution function F(t).

- **H.W. Classify the random variable to be discrete, continuous or neither and give your reasoning. In each case provide two possible values of the random variable.**
  1. **Waiting time until a specific bank goes bankrupt.**
  2. **Out of 100 specific banks the percentage of those that go bankrupt in one year.**
  3. **Number of eggs laid by a female turtle.**
  4. **Weight of a trout caught from a river.**
  5. **Difference of actual versus the advertized arrival time of a flight.**
  6. **Time to decay of a Uranium-235 atom.**
  7. **Amount of soda in a randomly selected can which is supposed to have 250 milliliters.**
  8. **Air pressure in a randomly selected inflated football.**
  9. **Number of wing flaps per minute of an eagle.**

Continuous Random Variable

- Density of a continuous random variable is a non-negative function, *f*, defined over R so that

$$\int_{\mathbb{R}} f(x)\,dx = 1$$

- Distribution function of a continuous random variable *X* is F(t) = P(X ≤ t), for all t∈R.
- Random variable *X* is continuous if there exists a density *f* so that,

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

for all t∈R.

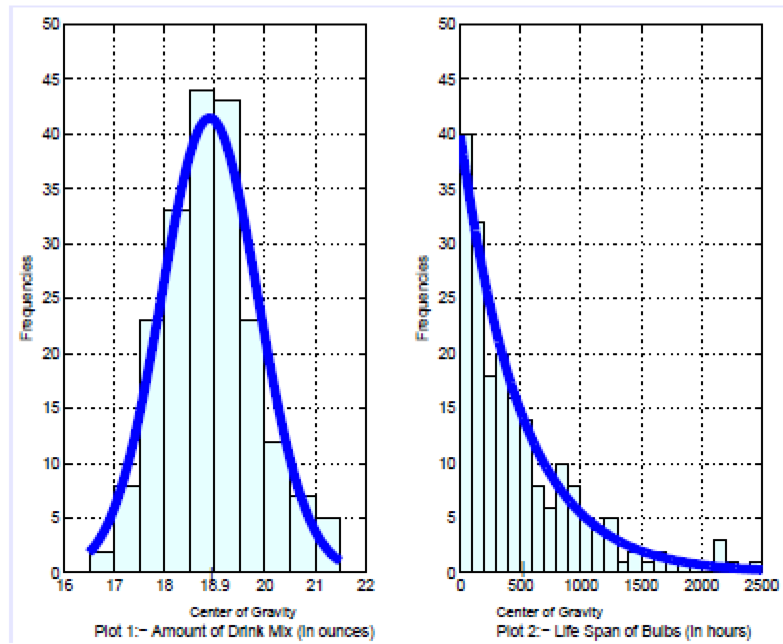
Properties of a Distribution Function

- A distribution function, F, always has the following properties
  1. F(t) is a non-decreasing function of t,
  2. $F(-\infty) = 0$, $F(\infty) = 1$,
  3. F(t) is right continuous for all $t \in R$.
- F(t) is the distribution of a continuous random variable if, **in addition**, there exists a density f, so that $\frac{d}{dt}F(t) = f(t)$ for $t \in R$.

SOME SPECIAL CONTINUOUS RANDOM VARIABLES

Motivation

- Data from real world experiments can often be approximated using densities of some known RVs.
- Allows us to approximate the data using concise mathematical functions.



Data following a normal density

Data following an exponential density

Uniform Random Variable

- Such a random variable takes values in a bounded interval, say (a, b), with density

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{for } x \in (a,b) \\ 0, & \text{otherwise} \end{cases}$$

- Denoted by X ∼ Uniform(a, b).
- Whenever we say "pick a point randomly …", then the picked point X is a uniform random variable.

Exponential Random Variable

- Takes values in the interval [0, ∞), with density

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{for } x \in (0, \infty) \\ 0, & \text{otherwise} \end{cases}$$

- The constant λ > 0 is a parameter of the density.
- Denoted by X ~ Exp(λ).

Exponential Random Variable

- Time it takes for a particular window glass to crack (due to some accident).
- Time it takes for a bulb to stop working.
- Time it takes for an electrical circuit to malfunction.
- Time it takes for a radioactive atom to decay.

Standard Normal Random Variable

- For modeling measurement errors.
- Takes values in R, with density

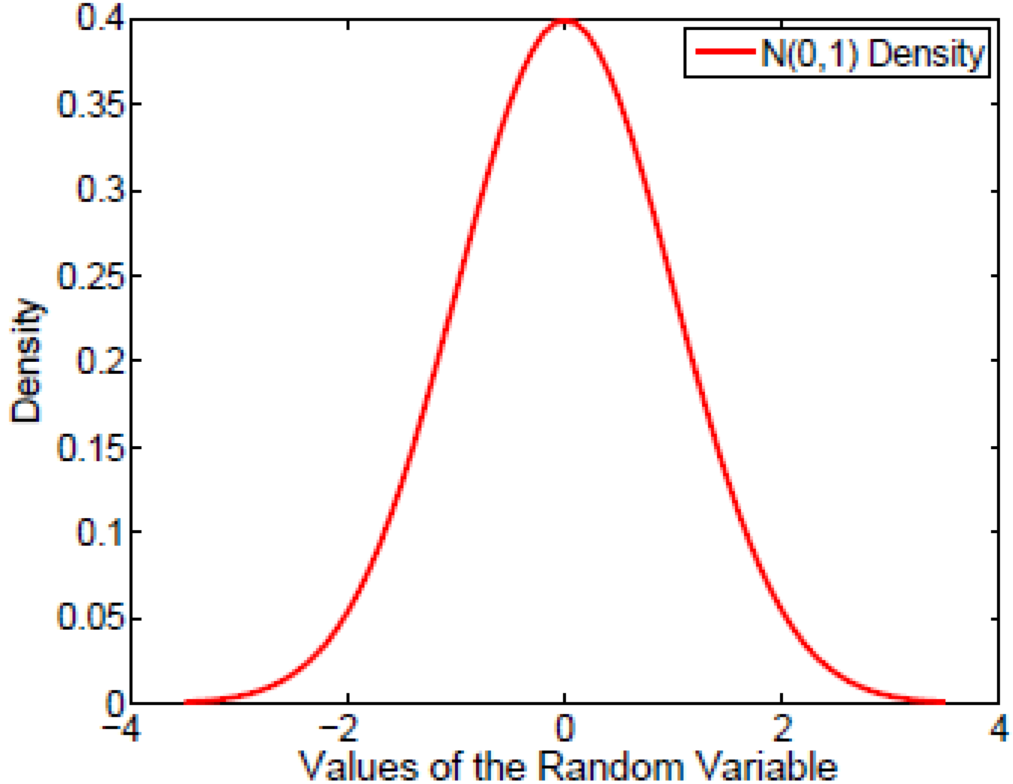
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \text{ for } x \in R$$

- Denoted by X ~ N(0, 1).
- Distribution function

$$F(t) = P(X \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

  – Cannot be integrated into a closed form.
  – That's why we have the Normal Table.

Standard Normal Random Variable



The Standard Normal Density

Normal Random Variable

- Takes values in R, with density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ for } x \in R$$

- Denoted by X ~ N(μ,σ^2) where constants, $\mu \in$R, $\sigma > 0$, are parameters of the density.
  - μ corresponds to the average value of *X*.
  - σ corresponds to the standard deviation of *X*.

Other Continuous Random Variables

- Gamma
- Chi-square
- Beta
- Cauchy
- Lognormal
- Logistic