STAT 511

Lecture 8: Continuous Random Variables: an Introduction

Devore: Section 4.1-4.3

Prof. Michael Levine

October 10, 2018

Continuous Random Variables: a motivating example

- What probability distribution formalizes the notion of "equally likely" outcomes in the unit interval $[0,1]$?
- If we assign $P(X=0.5)=\varepsilon$ for any real $\varepsilon>0$, we have a serious problem.
- Consider the event $E=\left\{\frac{1}{2},\frac{1}{3},\frac{1}{4},\ldots,\right\}$
- Then,

$$P(E)=P\left(\cup_{j=2}^{\infty}\left\{\frac{1}{j}\right\}\right)=\sum_{j=2}^{\infty}\varepsilon=\infty$$

- We *must* assign a probability of zero to *every outcome* x in $[0,1]$

Interpretation

- There is nothing shocking about it : an empty set (an impossible event) *must* have probability zero but nobody ever said that an event that has probability zero is always impossible...
- We also conclude that any countable event also has probability zero as well
- Moreover, if we think of "equally likely" outcomes as meaning that an outcome is equally likely to be in two subintervals of equal length, we have

$$1 = P(X \in [0,1]) = P(X \in [0,0.5]) + P(X \in [0.5,1]) - P(X = 0.5)$$
$$= P(X \in [0,0.5]) + P(X \in [0.5,1])$$

and, therefore $P(X \in [0,0.5]) = P(X \in [0.5,1]) = \frac{1}{2}$

Continuous uniform distribution

- Let S be the sample space, $X(S) = [0,1]$ and each $x \in [0,1]$ is equally likely. Then, for any $0 \leq a \leq b \leq 1$ $P(X \in [a,b]) = b - a$
- This is called **continuous uniform distribution**. Its cdf is easy to compute:
  1. If $y < 0$, $F(y) = P(X \leq y) = 0$
  2. If $y \in [0,1]$, $F(y) = P(X \leq y) = P(X \in [0,y]) = y$
  3. If $y > 1$,$F(y) = P(X \leq y) = P(X \in [0,1]) = 1$

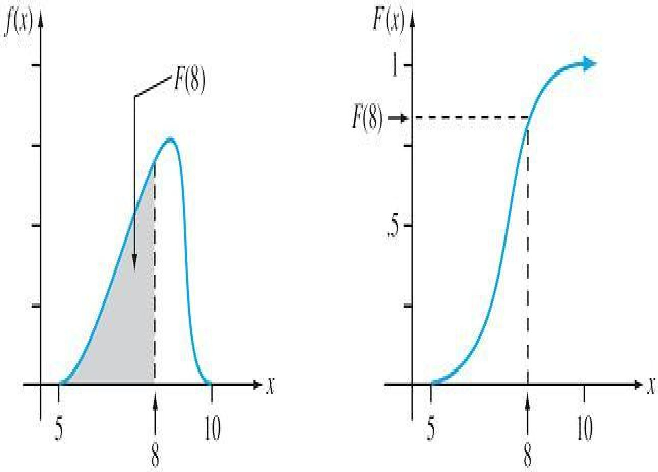
Continuous random variable: definition

- A random variable X is **continuous** if its set of possible values is an entire interval of numbers
- The function f is called a **probability density function** (pdf; compare to pmf) if $f(x) \geq 0$ for any $x \in R$ and $\int_{-\infty}^{\infty} f(x)\,dx = 1$.
- A random variable is continuous if there exists a pdf f such that for any two numbers a and b,

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

- For any two numbers a and b with $a < b$ $P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b)$.

Figure :

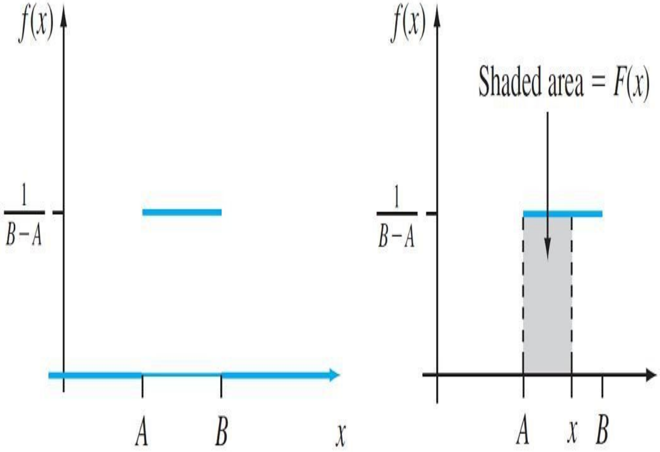


Uniform distribution

- Clearly, a RV has a uniform distribution on the interval $[A, B]$ if the pdf of X is

$$f(x; A, B) = \begin{cases} \frac{1}{B-A} \text{ if } A \leq x \leq B \\ 0 \text{ otherwise} \end{cases}$$

Figure :



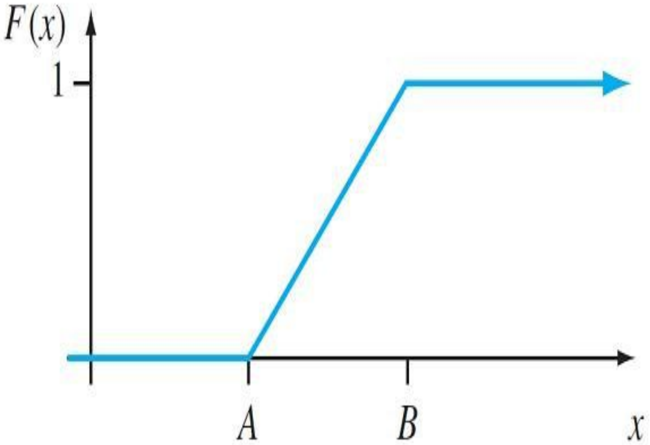
Cumulative Distribution Function (cdf)

- The cumulative distribution function $F(x)$ of a continuous RV X is defined for every number x as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\,dy$$

- For each x $F(x)$ is the area under the density curve to the left of x.
- **Ex**. Let X be a thickness of a metal sheet that has a uniform distribution on $[A, B]$. For $A \leq x \leq B$
$F(x) = \int_{-\infty}^{x} f(y)\,dy = \frac{x-A}{B-A}$.

Figure :



Using $F(x)$ to Compute Probabilities

- Let X be a continuous RV with pdf $f(x)$ and cdf $F(x)$. Then for any number a,

$$P(X > a) = 1 - F(a)$$

- For any numbers a and b such that $a < b$,

$$P(a \leq X \leq b) = F(b) - F(a)$$

Example

- Suppose the pdf of the magnitude X of a dynamic load on a bridge (in newtons) is

$$f(x; A, B) = \begin{cases} \frac{1}{8} + \frac{3}{8}x \text{ if } 0 \leq x \leq 2 \\ 0 \text{ otherwise} \end{cases}$$

- Then, for any $0 \leq x \leq 2$,

$$F(x) = \int_{-\infty}^{x} f(y)\, dy = \int_{0}^{x} \left(\frac{1}{8} + \frac{3}{8}y \right) dy = \frac{x}{8} + \frac{3}{16}x^2$$

- Based on the above, we have

$$P(1 \leq X \leq 1.5) = F(1.5) - F(1) = \frac{19}{64}$$

Obtaining $f(x)$ from $F(x)$

- If X is a continuous RV with pdf $f(x)$ and cdf $F(x)$, then at every number x for which the derivative $F^{'}(x)$ exists,

$$f(x) = F^{'}(x)$$

- **Ex.** Consider the uniform cdf

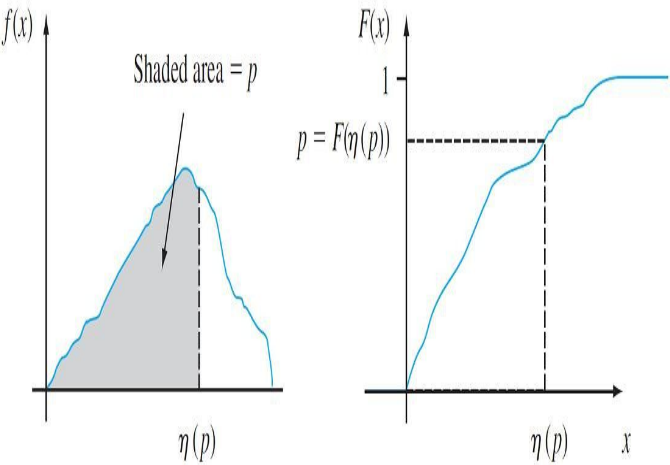
$$f(x; A, B) = \begin{cases} 0 \text{ if } x < A \\ \frac{x-A}{B-A} \text{ if } A \leq x \leq B \\ 1 \text{ if } x > B \end{cases}$$

- The pdf is then equal $F^{'}(x) = \frac{1}{B-A}$ for $A \leq x \leq B$ and 0 otherwise

Percentiles

- Let $0 < p < 1$. The $(100p)$ th percentile of the distribution of a continuous RV X is denoted by $\eta(p)$ and is defined from the equation
$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)\,dy.$$
- The median of a continuous distribution, denoted by $\tilde{\mu}$ is the 50th percentile. The defining equation is
$$0.5 = F(\tilde{\mu})$$
- That is, half the area under the density curve is to the left of $\tilde{\mu}$.

Figure :



Expected Value of a RV

- The expected or mean value of a continuous RV X with pdf $f(x)$ is
$$E(X) = \mu = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$
- If X is a continuous RV with pdf $f(x)$, then for any function $h(x)$
$$E(h(X)) = \mu_{h(X)} = \int_{-\infty}^{\infty} h(x) \cdot f(x)\, dx$$

Example

- In the "broken stick" ecological model, the proportion of the resource controlled by species 1 has the uniform distribution on $[0, 1]$
- The species that controls the majority of this resource controls the amount

$$h(X) = \max(X, 1 - X) = \begin{cases} 1 - X & 0 \leq X \leq \frac{1}{2} \\ X & \frac{1}{2} \leq X \leq 1 \end{cases}$$

- The expected amount controlled by the species having majority control is

$$E\, h(X) = \int_0^1 \max(x, 1 - x) * 1 d\, x = \frac{3}{4}$$

Variance and Standard Deviation

- The variance of continuous RV X with pdf $f(x)$ and mean μ is

$$V(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\, dx = E(X-\mu)^2$$

- The standard deviation is $\sigma_X = \sqrt{V(X)}$
- The shortcut formula is

$$V(X) = E(X^2) - [E(X)]^2$$

- For any constants a and b,

$$V(aX+b) = a^2 \cdot V(X)$$

and $\sigma_{aX+b} = |a| \cdot \sigma_X$

Example

- For X = weekly gravel sales, we established $E(X) = \frac{3}{8}$
- By definition,

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 x^2 \frac{3}{2}(1 - x^2)\,dx = \frac{1}{5}$$

- Finally, $V(X) = 0.059$ and $\sigma_X = .244$

Definition

- A continuous RV X is said to have a *normal distribution* with parameters μ and σ^2, $-\infty < \mu < \infty$ and $0 < \sigma^2$, if the pdf of X is
$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$
for all $-\infty < x < \infty$.
- The normal distribution is very important as it describes a very wide variety of data. Heights, weights and other physical characteristics of different populations, measurement errors in scientific experiments and many other types of data are readily described by the normal distribution

Normal Distribution

- Moreover, sums and averages of a large number of non-normal variables can be described as normal under some suitable conditions.
- It is easy to see that $f(x;\mu,\sigma^2) > 0$; a little more difficult to confirm that
$$\int_{-\infty}^{\infty} f(x;\mu,\sigma^2)\,dx = 1$$
- μ is the mean:
$$E(X) = \mu$$
and σ^2 is the variance:
$$V(X) = \sigma^2$$

Standard Normal Distribution

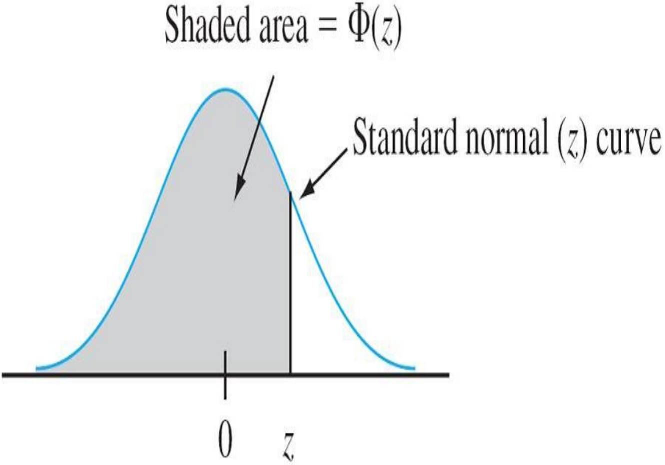
- The normal distribution with parameter values $\mu = 0$ and $\sigma^2 = 1$ is called a *standard normal distribution*.
- A random variable that has a standard normal distribution is called a *standard normal random variable* and is denoted by Z.
- Its pdf is

$$f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

- Its cdf is

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} f(y; 0, 1)\, dy$$

Figure :



Examples

- Let Z be the standard normal random variable. Find
  1. $P(Z \leq 0.85) = 0.8023$ (area under the curve to the left of 0.85)
  2. $P(Z > 1.32) = 1 - P(Z \leq 1.32) = 0.0934$
  3. $P(-2.1 \leq Z \leq 1.78) = P(Z \leq 1.78) - P(Z \leq -2.1) = 0.9625 - 0.0179 = 0.9446$ - the area to the left of 1.78 minus the area to the left of -2.1

- z_α is the value on the measurement axis for which the area under the z curve that lies to the right of it is equal to α

Example

- **Ex.** Let Z be the standard normal variable. Find z if $P(Z < z) = 0.9278$
  - Look at the table and find an entry $= 0.9278$ then read back to find $z = 1.46$
- Find z such that $P(z < Z < z) = 0.8132$
  - The standard normal distribution is symmetric so $P(-z < Z < z) = 2P(0 < Z < z)$
  - $P(0 < Z < z) = P(Z < z) - \frac{1}{2}$
  - Thus,$2P(Z < z) - 1 = 0.8132$ or $P(Z < z) = 0.9066$
  - From the table, $z = 1.32$

Nonstandard normal distribution

- If X has a normal distribution with mean μ and standard deviation σ, then
$$Z = \frac{X - \mu}{\sigma}$$
has the standard normal distribution

Example

- Let X be a normal random variable with $\mu = 80$ and $\sigma = 20$
- Find $P(X \leq 65)$

$$P(X \leq 65) = P\left(Z \leq \frac{65-80}{20}\right) = P(Z \leq -.75) = .2266$$

Example

- The breakdown voltage of a randomly chosen diode of a particular type is normally distributed. What is the probability that a diode's breakdown voltage is within 1 standard deviation of its mean value?

$$P(\mu-\sigma\le X\le\mu+\sigma)=P\left(-1.00\le Z\le 1.00\right)$$
$$=\Phi(1.00)-\Phi(-1.00)=0.6826$$

Normal Approximation to the Binomial Distribution

- Let X be a binomial RV based on n trials, each with probability of success p.
- If the binomial probability histogram is not too skewed, X may be approximated by a normal distribution with $\mu = np$ and $\sigma = \sqrt{np(1-p)}$ as long as $np \geq 10$ and $n(1-p) \geq 10$.
- More specifically,

$$P(X \leq x) = B(x; n, p) \approx \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

Example

- At a particular small college the pass rate of Intermediate Algebra is 72%. If 500 students enroll in a semester determine the probability that at most 375 students pass.
- First, $\mu = np = 500 \cdot (.72) = 360$
- Next, $\sigma = \sqrt{npq} = \sqrt{500 \cdot (.72) \cdot (.28)} \approx 10$
- Finally,

$$P(X \leq 375) \approx \Phi\left(\frac{375.5 - 360}{10}\right) = \Phi(1.55) = 0.9394$$