Seoul National University
Dept. of Civil and Environmental Engineering

Instructor: Junho Song
junhosong@snu.ac.kr

457.212 Statistics for Civil & Environmental Engineers

In-Class Material: Class 10

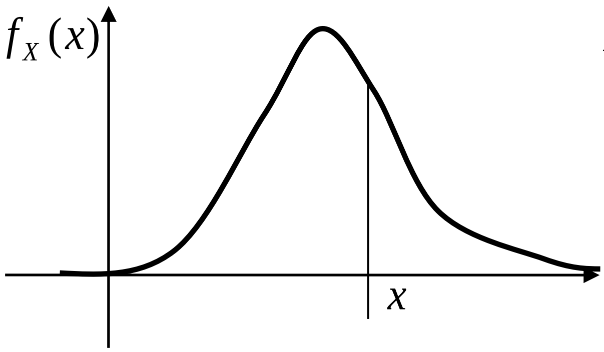
Probability Distribution Functions and Partial Descriptors (A&T: 3.1)

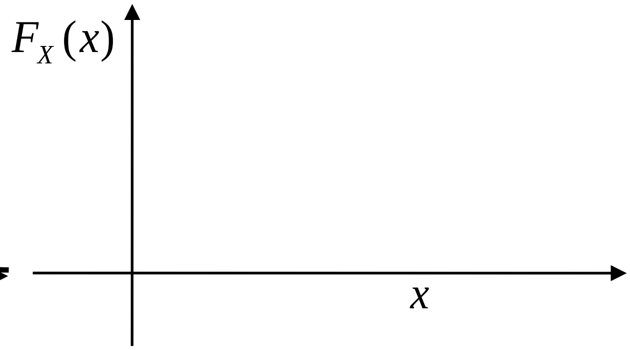
1. **Cumulative Distribution Function** (CDF) of a **continuous** random variable X, $F_X(x)$

(a) Definition

$$F_X(x) = P(\qquad\qquad)$$

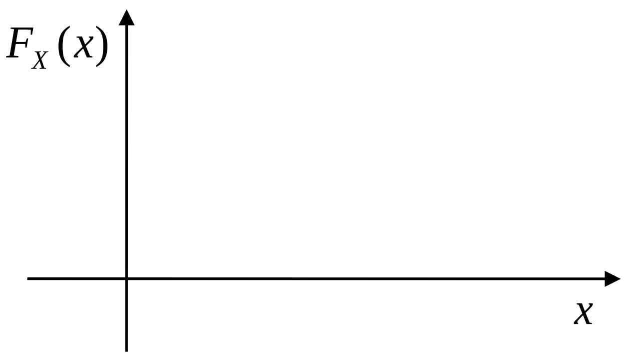
$$= \int f_X(x)dx$$





(b) Properties

- $\dfrac{dF_X(x)}{dx} =$. In summary, [PDF] [CDF]
- Non-decreasing (because we are integrating a non-negative function)
- $F_X(-\infty) = P(\qquad\qquad) =$
- $F_X(\infty) = P(\qquad\qquad) =$

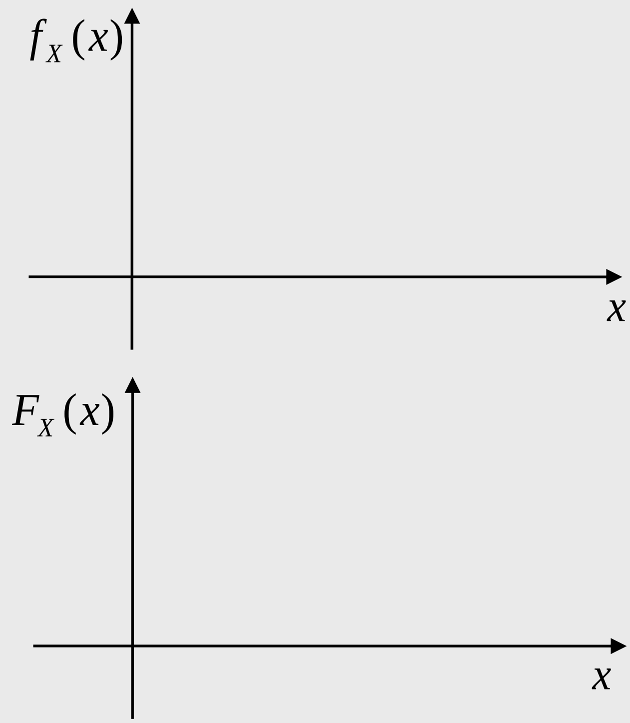


Example 1: Suppose the likelihood of accidents is uniform along the 100 km highway. Let X denote the distance between the starting point and an accident location. Determine

(a) Probability density function (PDF) of X and plot

(b) Cumulative density function (CDF) of X and plot:

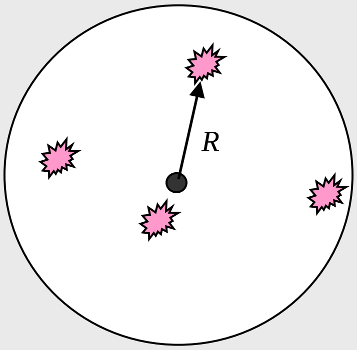
(c) $P(20 \le X \le 50)$ by use of PDF and CDF




```
ex01_sample = runif(10000, min=0, max=100)
# generate random numbers following uniform distribution U(0,100)

hist(ex01_sample, freq=FALSE, breaks=seq(0,100,5), main="PDF")
# plot empirical PDF from samples: use hist with freq=FALSE

plot(ecdf(ex01_sample), verticals=TRUE, pch="", main="CDF")
# plot empirical CDF using samples

ex01_c = subset(ex01_sample, (ex01_sample>=20 & ex01_sample<=50))
# subset of elements satisfying the condition
c_ans = length(ex01_c)/length(ex01_sample) # P(20≤X≤50)
```

Instructor: Junho Song
junhosong@snu.ac.kr

Example 2: Suppose an object can fall anywhere within a 10-km radius circle at random (i.e. uniform likelihood over points inside the circle). PDF and CDF of the distance between the location and the center of the circle, R ?




```
# creating a user-defined function
ex02 = function(num) { # create a function
samp = c()
repeat{ # repeated infinitely without 'break'
  x = runif(1, min=-10, max=10)
  y = runif(1, min=-10, max=10)
  if(x^2+y^2 < 100){ # random sample is located inside of the circle
    samp = rbind(samp, c(x,y))
  }
  if(dim(samp)[1]==num) break # break when number of sample is 'num'
}
return(samp) # result of function 'ex02' is 'samp'
}

# scatter plot of samples
ex02_sample = ex02(1000)
plot(ex02_sample, xlab="x", ylab="y", asp=1)

# draw CDF
distance = sqrt(ex02_sample[,1]^2+ex02_sample[,2]^2)
ex02_sample = cbind(ex02_sample, distance)
plot(ecdf(ex02_sample[,3]), verticals=TRUE, pch="", main="CDF")
```

Instructor: Junho Song
junhosong@snu.ac.kr

2. **Partial Descriptors** of a random variable

(a) "Complete" description by probability functions:

(b) "Partial" descriptors: measures of key characteristics; can derive from ( )

Note:

- Expectation: $E[\cdot] = \int_{-\infty}^{\infty} (\cdot) f_X(x) dx$ (continuous) or $\sum_{\text{all } x} (\cdot) p_X(x)$ (discrete)
- Moment: $E[\cdot X^n] = \int_{-\infty}^{\infty} x^n f_X(x) dx$ or $\sum_{\text{all } x} x^n p_X(x)$
- Central Moment, $E[(X - \mu_X)^n] = \int_{-\infty}^{\infty} (x - \mu_X)^n f_X(x) dx$ or $\sum_{\text{all } x} (x - \mu_X)^n p_X(x)$

	Name	Definition	Meaning (PDF/CDF)
Measure of Central Location	Mean, μ_X	First moment, $\mathrm{E}[X]$	Location of the ( ) of an area underneath ( )
	Median, $x_{0.5}$	$F_X(x_{0.5}) = 0.5$ $F_X^{-1}(0.5)$	The value of a r.v. at which values above and below it are ________lly probable. If symmetric?
	Mode, $\tilde{x}$	$\arg\max_x f_X(x)$	The outcome that has the ________est probability mass or density
Measure of Dispersion	Variance, σ_X^2	Second-order central moment $\mathrm{E}[(X - \mu_X)^2]$ $= \mathrm{E}[X^2] - \mathrm{E}[X]^2$	Average of squared deviations
	Standard Deviation, σ_X	$\sqrt{\sigma_X^2}$	Radius of ( )

Instructor: Junho Song
junhosong@snu.ac.kr

	Coefficient of Variation (C.O.V.), δ_X	$\frac{\sigma_X}{\mid\mu_X\mid}$	__________ed radius of ( )
Asymmetry	Coefficient of Skewness, γ_X	Third-order central moment normalized by σ_X^3, $\frac{E[(X-\mu_X)^3]}{\sigma_X^3}$	Behavior of two tails > 0 = 0 < 0
Flatness	Coefficient of Kurtosis, κ_X	Fourth-order central moment normalized by σ_X^4, $\frac{E[(X-\mu_X)^4]}{\sigma_X^4}$	"Peakedness" - more of the variance is due to infrequent extreme deviations, as opposed to frequent modestly-sized deviations.

Example 3: Compute the mean, median, mode, variance, standard deviation and coefficient of variation for a discrete random variable X whose PMF is given as

x	$P_X(x)$
0	0.20
1	0.50
2	0.30

Instructor: Junho Song
junhosong@snu.ac.kr

Example 4: Compute the mean, median, mode, variance, standard deviation and coefficient of variation for a continuous random variable X whose PDF is given by $f_X(x) = (3/1000)x^2, \quad 0 \le x \le 10$ and 0 elsewhere. What is your guess on the sign of the coefficient of the skewness? Confirm your guess by computing it.

```
# mean
ex04_pdf = function(x) {x*(3/1000)*x^2}
mean_data = integrate(ex04_mean, lower=0, upper=10) # integrate function
mean = mean_data$value # only 'value' of the integrate result

# variance
ex04_var = function(x) {(x-mean)^2*(3/1000)*x^2}
var_data = integrate(ex04_var, lower=0, upper=10)
var = var_data$value
std = sqrt(var)

# skewness
ex04_skew = function(x) {(x-mean)^3*(3/1000)*x^2}
skew_data = integrate(ex04_skew, lower=0, upper=10)
skew = skew_data$value/std^3

# kurtosis
ex04_kurt = function(x) {(x-mean)^4*(3/1000)*x^2}
kurt_data = integrate(ex04_kurt, lower=0, upper=10)
kurt = kurt_data$value/std^4
```