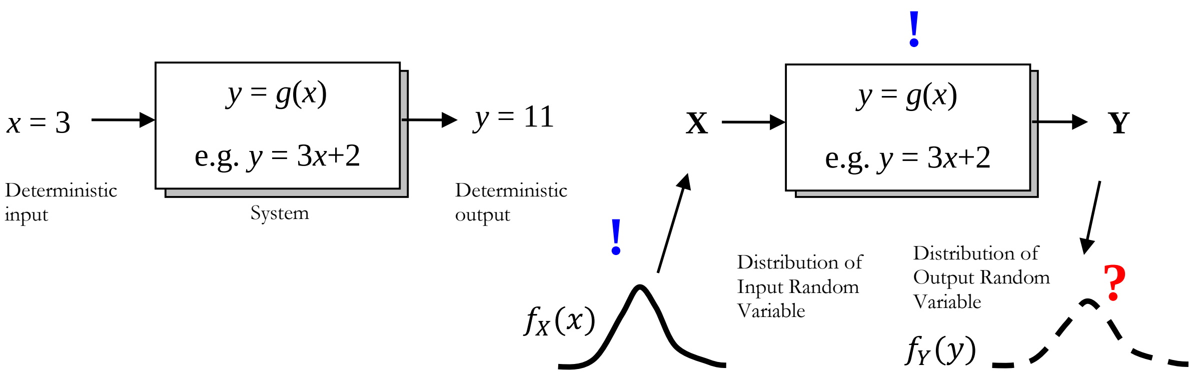
Seoul National University
Dept. of Civil and Environmental Engineering

Instructor: Junho Song
junhosong@snu.ac.kr

457.212 Statistics for Civil & Environmental Engineers
In-Class Material: Class 16
Function of Random Variables (A&T 4.2)



1. **Derived Distribution** of the function of a r.v., $Y = g(X)$

   (a) **P**robability **M**ass **F**unction $p_Y(y)$

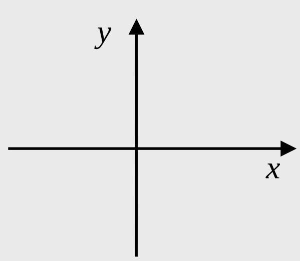
   - Sum up the probability mass function values of X that satisfy $y = g(X)$

$$p_Y(y) = P(Y = y) = \sum_{all\ x_i : y = g(x_i)} p_X(x_i)$$

Example 1: Consider a discrete random variable X that follows the PMF given on the right.

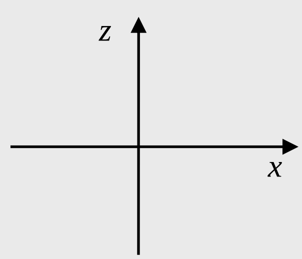
Find the PMF of the following functions of X :

x	$p_X(x)$
–1	0.25
0	0.40
1	0.25
2	0.10



(a) $y = -x + 1$

y	$p_Y(y)$



(b) $z = x^2 + 1$

z	$p_Z(z)$

(b) **C**umulative **D**istribution **F**unction (Discrete), $F_Y(y)$

- Sum up the probability mass function values of X that satisfy $g(X)$ y

$$F_Y(y) = P(Y \le y) = \sum_{all\ x_i : g(x_i) \le y} p_X(x_i)$$

Example 1 (Contd.): CDF of Y and Z ?

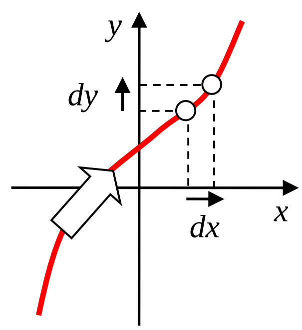
(c) **C**umulative **D**istribution **F**unction (Continuous), $F_Y(y)$

- (   ) the probability (   ) function of X for the range(s) satisfying $g(X)$ y

$$F_Y(y) = P(Y \le y) = \int_{g(x) \le y} f_X(x)dx$$

(d) **P**robability **D**ensity **F**unction, $f_Y(y)$

(i) One-to-one mapping and $\dfrac{dy}{dx} > 0$



Due to the one-to-one mapping,

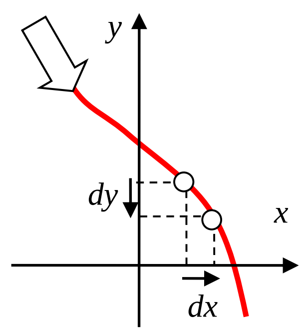
$$P(x < X \le x + dx) = P(y < Y \le y + dy)$$

$$f_X(x)dx = f_Y(y)dy$$

Therefore,

$$f_Y(y) =$$

(ii) One-to-one mapping and $\dfrac{dy}{dx} < 0$



Due to the one-to-one mapping,

$$P(x < X \le x + dx) = P(y < Y \le y + dy)$$

$$f_X(x)dx = f_Y(y)(-dy)$$

Therefore,

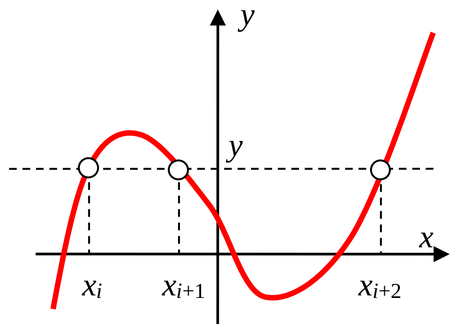
$$f_Y(y) =$$

(i) & (ii) One-to-one mapping: $f_Y(y) =$

Instructor: Junho Song
junhosong@snu.ac.kr

(iii) Non one-to-one mapping

$$P(y < Y \le y + dy) = \sum P(x_i < X \le x_i + dx)$$



Therefore,

$$f_Y(y) = \sum_{\text{all } x_i : g(x_i) = y} f_X(x_i) \left| \frac{dx}{dy} \right|_{x = x_i}$$

Example 2: $X \sim N(\mu, \sigma^2)$ and consider the linear function $U = g(X) = (X - \mu)/\sigma$

(a) The probability density function of U, $f_U(u)$?

(b) What type of the distribution does U follow? Mean? Standard deviation?

* Recall

$$P(X \le a) = P\left(\frac{X - \mu}{\sigma} \le \frac{a - \mu}{\sigma}\right) = P\left(U \le \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right)$$

?

(c) Draw PDF of U when $\mu = 3, \sigma = 2$.

```
par(mfrow=c(2,1)) # create subplots (2 by 1)
samp2X = rnorm(10000, mean=3, sd=2)
samp2U = (samp2X-3)/2
density(samp2U) # kernel-based estimation of PDF
plot(density(samp2U))

real2U = seq(-4, 4, by=0.01)
real2_density = dnorm(real2U, mean=0, sd=1) # standard normal PDF
plot(real2U, real2_density, pch=".")
```

Instructor: Junho Song
junhosong@snu.ac.kr

Example 3: $X \sim LN(\lambda, \zeta^2)$. The PDF of $Y = \ln X$?
(Distribution of the natural logarithm of a lognormal random variable)

Draw PDF of Y when $\lambda = 3, \zeta = 2$.

```
par(mfrow=c(2,1))
samp3X = rlnorm(10000, meanlog=3, sdlog=1)
samp3Y = log(samp3X)
plot(density(samp3Y))

real3Y = seq(-1, 7, by=0.01)
real3_density = dnorm(real3Y, mean=3, sd=1)
plot(real3Y, real3_density, pch=".")
```

Example 4: The strain energy (E) accumulated during a linearly elastic behavior is proportional to the square of the applied force, S^2, i.e. $E = aS^2$ where a is a positive constant. When S follows the standard normal distribution, what is the PDF of the strain energy?

Draw PDF of E when $a = 2$.

Instructor: Junho Song
junhosong@snu.ac.kr

```
a=2
par(mfrow=c(2,1))
samp4S = rnorm(100000, mean=0, sd=1)
samp4E = a*samp4S^2
plot(density(samp4E, from=0), xlim=c(0,5), ylim=c(0,0.8))
# 'from' option in 'density' function: starting point of PDF estimation

real4E = seq(0,5,by=0.001)
Ex4ftn = function(x) {1/sqrt(2*pi*a*x)*exp(-x/2/a)}
real4_density = Ex4ftn(real4E)
plot(real4E, real4_density, pch=".", xlim=c(0,5), ylim=c(0,0.8))
```

2. Important Examples of **Derived Distributions**

(a) The **sum** of s.i. **Poisson** random variables follows a __________ distribution

$X_i,\ i=1,..,n$ ~ Poisson r.v.'s with ν_i (mean occurrence rate)

$Z=\sum_{i=1}^{n} X_i$ ~ ________ r.v. with $\nu=\sum_{i=1}^{n}\nu_i$

→ See Example 4.5 in A&T

(b) **Linear functions** (including sum) of **normal** r.v.'s follow __________ distribution

$X_i,\ i=1,...,n$ ~ Normal r.v.'s with μ_i and σ_i (mean and standard deviation)

$Z=\sum_{i=1}^{n} a_i X_i + a_0$ ~ ________ r.v. with $\mu_Z=?$ and $\sigma_Z=?$ (Next class)

(c) **Products** or **quotients** of **lognormal** r.v.s follow ______________

$X_i,\ i=1,..,n$ ~ lognormal with λ_i and ζ_i

$Z=a_0\prod_{i=1}^{n} X_i^{a_i}$ ~ _________ r.v.

→ Why? Take the natural logarithm: $\ln Z=\ln a_0+\sum a_i \ln X_i$ ~ Normal