Lecture 3

Random Variables and Their Properties

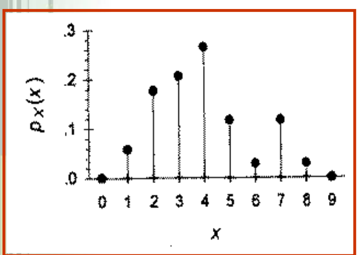
Statistics for
Civil & Environmental Engineers

Discrete Random Variables

❖ **Distribution Functions**

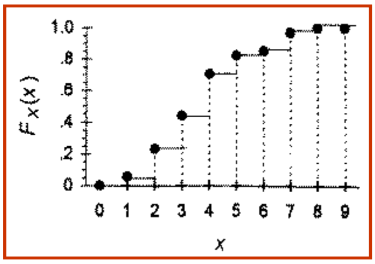
➢ **Probability Mass Function**

$$p_X(x) = \Pr[X = x]$$



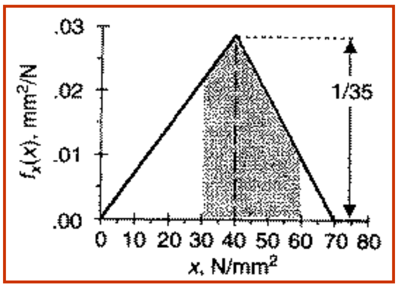
➢ **Cumulative Distribution Function**

$$F_X(x) = \sum_{X_k \le x} p_X(x_k)$$



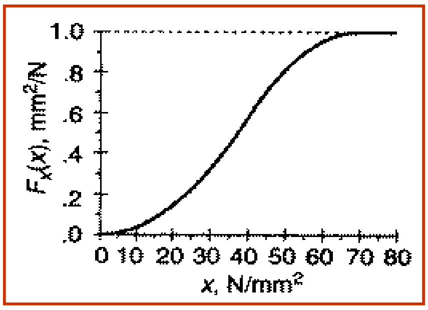
Continuous Random Variables

- **Probability Density Function**
  - $f_X(x) \geq 0$
  - $\Pr[x_1 \leq X \leq x_2] =$



Continuous Random Variables

- **Cumulative Distribution Function of a Continuous RV**
  - $F_X(x) = \int_{-\infty}^{x} f_X(x)dz$
  - $\frac{dF_X(x)}{dx} = f_x(X)$



Example 3.6: Earthquake Intensity

Example 3.6. Earthquake intensity. For another application of the pdf and cdf, consider the occurrence of earthquakes in a region for which the cdf can be simplified to the exponential distribution

$$F_X(x) = 1 - e^{-\lambda x},$$

where λ is a parameter and the random variable X is the magnitude of an earthquake in the region in the range $0 \le x \le +\infty$. From Eq. (3.1.6*b*) the pdf is given by

Example 3.6: Earthquake Intensity

$$f_X(x) = \lambda e^{-\lambda x}.$$

It is estimated that λ is 0.2. Equation (3.1.5*a*) is clearly applicable, and it is easy to show by integration the validity of Eq. (3.1.5*c*) in this case. The probability of an earthquake exceeding 10 units, for example, is given by

$$\Pr[X \ge 10] = 1 - F_X(10) = e^{-2} = .135.$$

The pdf and cdf are shown in Fig. 3.1.5.

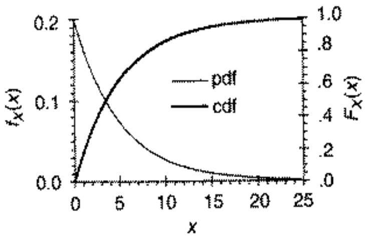


FIGURE 3.1.5
Probability density function and cumulative distribution function exponentially distributed earthquake intensities X.

Description of Random Variables

- **Expectation and Other Population Measures**
  - **Mean or Expected Value**

In the discrete case: $\mu_X = E[X] =$

In the continuous case: $\mu_X = E[X] =$

건설환경공학부

Description of Random Variables

- **Expectation and Other Population Measures**
  - **Definition of Expectation Operator**

Let X be a random variable and g(x) a function of X. The expectation of the function g(x) is given by

$$E[g(x)] = \sum_{all\ x_i} g(x_i) p_X(x_i)$$

if X is a discrete variable with mass points x_i and

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

if X is a continuous variable with pdf $f_X(x)$, provided that the series and the integral are absolutely convergent.

건설환경공학부

Description of Random Variables

- **Expectation and Other Population Measures**
  - **Properties of Expectation Operator**

E[a]= a, for a constant a;

E[ah(X)]= aE[h(X)], for a constant a;

$E[ah_1(X)+bh_2(X)]=$

for two constant a and b;

$E[h_1(X)] \geq E[h_2(X)]$ if $h_1(X) \geq h_2(X)$,

for two functions $h_1(x)$ and $h_2(x)$

Description of Random Variables

- **Expectation and Other Population Measures**
  - **Chebyshev inequality**

$Pr[h(X) \geq m] \leq m^{-1}E[h(X)]$, for every m > 0 and h(x) ≥ 0

Example 3.11. Reliability bounds using Chebyshev inequality. Let the squared deviation of a random variable X from its mean μ_X be represented by $h(X) = (X - \mu_X)^2$,

$$E[h(X)] = E[(X - \mu_X)^2] = \sigma_X^2.$$

Let $m = k^2\sigma_X^2$. Then by using Chebyshev inequality,

$$Pr[(X - \mu_X)^2 \geq k^2\sigma_X^2] \leq 1/k^2, \qquad \text{for every } k.$$

It follows that

$$Pr[|x - \mu_X| < k\sigma_X] = Pr[-k\sigma_X < |X - \mu_X| < k\sigma_X] \geq 1 - 1/k^2,$$

Description of Random Variables

- **Chebyshev inequality (cont.)**

which can be written as

$$\Pr[\mu_X - k\sigma_X < X < \mu_X + k\sigma_X] \geq 1 - 1/k^2.$$

This expression states that the probability that X falls within $k\sigma_x$ units of μ_X is greater than or equal to $1 - 1/k^2$. For $k = 2$, the lower bound to probability is $\frac{3}{4}$, and for $k = 3$ it is $\frac{8}{9}$. Using this rule, one can establish operational bounds without specifying the probability law of the investigated system. For example, consider a supply system subject to a random load or demand with known mean and standard deviation. An engineer who designs the capacity of this system to satisfy any demand ranging within two standard deviations of the mean ($k = 2$) does so with the knowledge that the reliability of this system will not be less than 75 percent. However, as we shall see in subsequent chapters, a higher reliability is obtained by making further assumptions—for example, that several moments are known.

Multiple Random Variables

- **Joint Probability Distribution(1)**
  - **Conditional Probability Density Function**

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$f_{X,Y}(x,y) = f_{Y|X}(y|x) f_X(x) = f_{X|Y}(x|y) f_Y(y)$$

  - **Independent Continuous Random Variables**

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

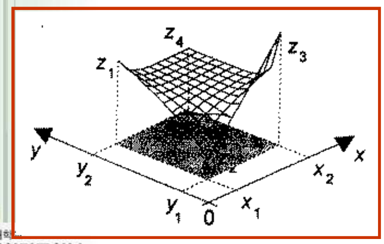
Multiple Random Variables

❖ Joint Probability Distribution(2)

➢ Marginal Probability Density Function

$$f_X(x)=\int_{-\infty}^{+\infty} f_{X|Y}(x,y)f_Y(y)dy=\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$$

$$f_Y(y)=$$



Schematic diagram of bivariate pdf represented by heights of a curved surface in the given ranges of the variables

Multiple Random Variables

❖ Properties of Multiple Variables

➢ Covariance and Correlation

$$Cov[X_1,X_2]=E[(X_1-E[X_1])]=E[X_1,X_2]-E[X_1]E[X_2]$$

$$E[X_1X_2]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1x_2 f_{X_1X_2}(x_1,x_2)dx_1dx_2$$

➢ Conditional Expectation

$$E[X|Y=y_j]=\sum_{all\ i} x_i\, p_{X|Y}(x_i|y_j)$$

$$E[Y|X=x_i]=\sum_{all\ j} y_j\, p_{Y|X}(y_j|x_i)$$

Assignment 2

- **Chapter 2**

  Problems

  2.1 / 2.4 / 2.8 / 2.18 / 2.23

- **Chapter 3**

  Problems

  3.3 / 3.6 / 3.16 / 3.20

Due date: September 29th, 2009