Section 2.5

Random Variables
Random variables
Random variables

Random variable assigns a numerical value to each outcome of a random experiment.

**Notation**: Random variable: r.v.

**Example**: Select 13 cards from a deck of 52. Random variable $X =$ number of Aces.
Possible values of $X$: 0, 1, 2, 3, 4. $X$ is a discrete r.v.

**Example**: Select a student at random from a class of 200. Random variable $X =$ height of the selected student.
Possible values of $X$: [5ft, 7ft] is a continuous r.v.
DISCRETE RANDOM VARIABLES

Def. A random variable is discrete if its values form a discrete set. That is, a discrete r.v. takes on finite (or countable) number of values.

What do we need to specify/describe a r.v? Need its values and their probabilities!

We describe a discrete r.v. using probability mass function or probability distribution: list of all values of the random variable with the corresponding probabilities: \( p(x) = P(X = x) \).

X random variable. Values of X: \( x_1, x_2, \ldots, x_n \)

Probabilities: \( P(X = x_1) = p_1, P(X = x_2) = p_2, \ldots, P(X = x_n) = p_n \).

Every \( p_i \geq 0 \) and \( p_1 + p_2 + \ldots + p_n = 1 \).

Typically, probability distribution of a random variable is given as a table:

<table>
<thead>
<tr>
<th>Values of the r.v X</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( \ldots )</th>
<th>( x_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>probabilities</td>
<td>( p_1 )</td>
<td>( p_2 )</td>
<td>( \ldots )</td>
<td>( p_n )</td>
</tr>
</tbody>
</table>
Discrete random variables. Example.

Toss a fair coin 3 times. Let $X$ be the number of Heads in the 3 tosses. Find the probability distribution of $X$.

**Sln.** Sample space $S=\{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$. 8 equally likely outcomes.

$X =$ # H. Possible values of $X$: 0, 1, 2, 3. Need probabilities.

$P(X=0)=P(0 \text{ Heads})=P(TTT)=1/8$, (X assigned number 0 to experiment outcome TTT)

$P(X=1)=P(1 \text{ Head})=P(TTH, THT, HTT)=3/8$,

$P(X=2)=P(2 \text{ Heads})=P(HHT, HTH, THH)=3/8$,

$P(X=3)=P(3 \text{ Heads})=P(HHH)=1/8$.

**Probability distribution of $X$:**

<table>
<thead>
<tr>
<th>$X$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>probs</td>
<td>1/8</td>
<td>3/8</td>
<td>3/8</td>
<td>1/8</td>
</tr>
</tbody>
</table>
Discrete random variables. Example.

The number of flaws in 1 inch length of a wire varies from wire to wire. Overall, 48% of the wires have no flaws, 39% have 1 flaw, 12% have 2 flaws, and 1% have three flaws. Let $X$ be the number of flaws in a randomly selected piece of wire. Construct the probability distribution of $X$.

Sln.

$X = \# \text{ flaws}$. Possible values of $X$: 0, 1, 2, 3. Need probabilities.

$P(X=0) = 0.48$, $P(X=1) = 0.39$, $P(X=2) = 0.12$, and $P(X=3) = 0.01$.

Probability distribution of $X$:

<table>
<thead>
<tr>
<th>$X$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>probs</td>
<td>0.48</td>
<td>0.39</td>
<td>0.12</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Cumulative Distribution Function of a Discrete Random Variable

**Definition:** For any random variable, the cumulative distribution function (cdf) is: \( F(x) = P(X \leq x) \).

Let \( X \) a discrete random variable, let its cdf be: \( F(x) = P(X \leq x) \).

Let \( p(x) = P(X = x) \) be the probability mass function.

Then: \( F(x) = \sum_{t \leq x} P(X = t) = \sum_{t \leq x} p(t) \), and \( \sum_x P(X = x) = \sum_x p(x) = 1 \).
Example: Flaws in wires. Recall the pmf: $P(X = 0) = 0.48$, $P(X = 1) = 0.39$, $P(X = 2) = 0.12$, and $P(X = 3) = 0.01$. Find $F(1)$, $F(-2)$, $f(2.1)$, $F(5)$, $F(3)$.

\[
F(1) = P(X \leq 1) = P(X=1)+P(X=0)= 0.48 + 0.39 = 0.87
\]
\[
F(-2) =P(X \leq -2) = 0
\]
\[
F(2.1) = P(X \leq 2.1) = P(X =2)+ P(X=1)+P(X=0)= 0.12+ 0.48 + 0.39 =0.99
\]
\[
F(5)= P(X \leq 5) =1.
\]
\[
F(3) = P(X \leq 3) = 0.48 + 0.39 + 0.12 + 0.01 = 1
\]

Plot the cdf of the number of flaws in a piece of wire.

Start with the formula for the cdf:

\[
F(x) = 0 \text{ for } x < 0
\]
\[
F(x)=0.48 \text{ for } 0 \leq x < 1
\]
\[
F(x)=0.87 \text{ for } 1 \leq x < 2
\]
\[
F(x)=0.99 \text{ for } 2 \leq x < 3
\]
\[
F(x)= 1 \text{ for } x \geq 3.
\]
MEAN AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Measures of center and spread of a r.v.

X – discrete r.v. X has probability distribution:

Values of the r.v X | $x_1$ | $x_2$ | … | $x_n$
---|---|---|---|---
probabilities | $p_1$ | $p_2$ | … | $p_n$

Center of the distribution: mean, or Expected value of X, $EX$ or $E(X)$ or $\mu$:

$$\mu = EX = x_1 p_1 + x_2 p_2 + \cdots + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

NOTES: The mean represents the “long-run-average” value of X. If we average many values of X, we expect to get a number close to $EX$.

$EX$ is a weighted average of the values of X, weights are the probabilities of the values.
Variance and Standard Deviation of a discrete random variable

Measure of spread around the mean \( \mu \) of a random variable. 
\( X \) discrete rv

**Variance of X**

Take squared deviations from the mean and add them up with the same weights as for the mean:

\[
\text{Var} X = \sigma_X^2 = \sum_x (x - \mu_X)^2 P(X = x) \\
= \sum_x x^2 P(X = x) - \mu_X^2.
\]

Standard deviation of \( X \): 
\[ \sigma = \sqrt{\sigma^2} = \sqrt{\text{Var} X}. \]

**NOTES:** Both variance and standard deviation of any random variable are nonnegative.
Your winnings on a lottery are a random variable with the following probability distribution:

<table>
<thead>
<tr>
<th>Winnings in $</th>
<th>100</th>
<th>200</th>
<th>1,000</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilities</td>
<td>0.4</td>
<td>0.3</td>
<td>?</td>
<td>0.2</td>
</tr>
</tbody>
</table>

a. Find the probability that you win $1,000.
b. What is the probability that you will win at least $150?
c. What is your expected winning?
d. What is the standard deviation of the winnings?

**Solution.** \( X = \) winnings in $.

a. Since sum of all probs=1, then \( P(X=1,000)=1-(0.4+0.3+0.2)=0.1. \)

b. \( P(X \geq 150)=P(X=200 \text{ or } X=1,000)=0.3 +0.1=0.4. \)

c. \( EX = 0 \times 0.2 + 100 \times 0.4 + 200 \times 0.3 +1,000 \times 0.1= 200. \) You should expect to win about $200 if you play for a long time.

d. \( VarX= (100-200)^2 \times 0.4+(200-200)^2 \times 0.3+(1000-200)^2 \times 0.1+(0-200)^2 \times 0.4=69,600 \)

The standard deviation of winnings is \( \sqrt{VarX} = $263.82 \)
Summary for Discrete Random Variables

Let $X$ be a discrete random variable. Then

- The probability mass function of $X$ is the function $p(x) = P(X = x)$.

- The cumulative distribution function of $X$ is the function $F(x) = P(X \leq x)$.

$$F(x) = \sum_{t \leq x} p(t) = \sum_{t \leq x} P(X = t)$$

- $\sum_{x} p(x) = \sum_{x} P(X = x) = 1$, where the sum is over all the possible values of $X$. 
The Probability Histogram

- When the possible values of a discrete random variable are evenly spaced, the probability mass function can be represented by a histogram, with rectangles centered at the possible values of the random variable.

- The area of the rectangle centered at a value $x$ is equal to $P(X = x)$.

- Such a histogram is called a probability histogram, because the areas represent probabilities.

Example: Probability Histogram for the Number of Flaws in a Wire

The pmf is: $P(X = 0) = 0.48$, $P(X = 1) = 0.39$, $P(X = 2) = 0.12$, and $P(X = 3) = 0.01$. 
Continuous Random Variables

A random variable $X$ is **continuous** if its probabilities are given by areas under a curve given by function $f(x)$. Let $a < b$ be any two numbers.

$$ P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = \int_{a}^{b} f(x)dx. $$

The function $f(x)$ is called a **probability density function (pdf)** for the r.v.

Let $X$ be a continuous random variable with pdf $f(x)$. Then

$$ \int_{-\infty}^{\infty} f(x)dx = 1. $$

Also:

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

$$ P(X \geq a) = P(X > a) = \int_{a}^{\infty} f(x)dx. $$
More on Continuous Random Variables

- Let $X$ be a continuous random variable with pdf $f(x)$. The **cumulative distribution function** of $X$ is the function

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

- The mean of $X$ is given by

$$\mu_X = \int_{-\infty}^{\infty} xf(x)dx.$$  

- The variance of $X$ is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)dx$$

$$= \int_{-\infty}^{\infty} x^2 f(x)dx - \mu_X^2.$$
Median and Percentiles for a Continuous Random Variable

Let $X$ be a continuous random variable with pdf $f(x)$ and cdf $F(x)$.

- The median of $X$ is the point $x_m$ that solves the equation

$$F(x_m) = P(X \leq x_m) = \int_{-\infty}^{x_m} f(x)dx = 0.5.$$ 

- If $p$ is any number between 0 and 100, the $p$th percentile is the point $x_p$ that solves the equation

$$F(x_p) = P(X \leq x_p) = \int_{-\infty}^{x_p} f(x)dx = p / 100.$$ 

- Note: the median is the 50th percentile.
A certain radioactive mass emits alpha particles from time to time. The time between emissions, in seconds, is random, with probability density function

\[
f(x) = \begin{cases} 
0.1e^{-0.1x} & x > 0 \\
0 & x \leq 0.
\end{cases}
\]

Find the median and 60th percentile of the emission times.
Example

A hole is drilled in a sheet-metal component, and then a shaft is inserted through the hole. The shaft clearance is equal to the difference between the radius of the hole and the radius of the shaft. Let the random variable $X$ denote the clearance, in millimeters. The probability density function of $X$ is

$$f(x) = \begin{cases} 1.25(1 - x^4), & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

1. Components with clearances larger than 0.8 mm must be scrapped. What proportion of components are scrapped?

2. Find the cumulative distribution function $F(x)$, mean clearance and variance of clearance.
Chebyshev’s Inequality

- Let $X$ be a random variable with mean $\mu_X$ and standard deviation $\sigma_X$. Then
  
  $$P\left(\left|X - \mu_X\right| \geq k\sigma_X\right) \leq \frac{1}{k^2}$$

- Chebyshev’s inequality is valid for any random variable and does not require knowledge of the distribution. The bound tends to overestimate the desired probability.

- Example: For the problem on shaft clearance, find the probability of the clearance to be further from the mean clearance by more than 2 standard deviations by (A) Chebyshev Rule, and (B) exact integration of the pdf.