Chapter 4. Probability: The Study of Randomness

Random Variables

- A random variable is a variable whose value is a numerical outcome of a random phenomenon.

- Typically use capital letters such as \(X\), \(Y\), or \(Z\) to denote random variables. For example, \(\bar{X}\) will be used to denote the sample mean of an SRS of size \(n\).

(EXAMPLE: Evolution or Creationism?)

Pew Research Center conducted opinion polls on the American public’s beliefs about the origins of life on earth. Let \(X\) be the number of people who accept the idea that humans evolved over time from a simple random sample of 100.

(EXAMPLE: Customer Service Calls)

Many businesses operate call centers to serve customers. Let \(X\) be the length (in seconds) of a customer service call made to a bank call center.

Discrete random variable

- It has a finite number of possible values.

- The probability distribution of the variable \(X\) lists the values and their probabilities:

<table>
<thead>
<tr>
<th>Value of (X)</th>
<th>(x_1)</th>
<th>(x_2)</th>
<th>(\ldots)</th>
<th>(x_k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>(p_1)</td>
<td>(p_2)</td>
<td>(\ldots)</td>
<td>(p_k)</td>
</tr>
</tbody>
</table>

where \(0 \leq p_i \leq 1\) and \(p_1 + p_2 + \cdots + p_k = 1\).

(EXAMPLE: Tossing a Coin)

Let \(X\) be the number of heads in four tosses of a fair coin.

\[
\begin{array}{cccccccc}
\text{TTTT} & \text{HTTT} & \text{THTT} & \text{TTHT} & \text{THHT} & \text{HHTT} & \text{HTHH} & \text{HHHT} & \text{HHHH} \\
\hline \\
X = 0 & X = 1 & X = 2 & X = 3 & X = 4
\end{array}
\]

1
The probability distribution of $X$ is

<table>
<thead>
<tr>
<th>Value of $X$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.0625</td>
<td>0.25</td>
<td>0.375</td>
<td>0.25</td>
<td>0.0625</td>
</tr>
</tbody>
</table>

and below is the probability histogram of the variable

Describe the event of getting at least one head in terms of $X$, and find the probability of the event.

**Continuous random variable**

- It takes all values in an interval of numbers.
- The probability distribution of the variable $X$ is described by a density curve.
- The probability of any event is the area under the density curve over the region described by the event.

- All continuous probability distributions assign probability 0 to every individual outcome. That is, $P(X = x) = 0$. Only intervals of values have positive probability. Consequently, two events $\{X > x\}$ and $\{X \geq x\}$ have the same probability.
• A density curve can be equivalently represented by the equation of the height of the curve $p(x)$ at any given value $x$.

• Example of the probability distribution of a continuous random variable

  i) A **uniform distribution** is the probability distribution of a randomly chosen number from 0 to 1.

  ![Uniform Distribution Example](image)

  ![Uniform Distribution Example](image)

  ii) **Normal distributions** are described by Normal curves.

  ![Normal Distribution Example](image)

  If $X$ has $N(\mu, \sigma)$ distribution, then the standardized variable $Z = \frac{X - \mu}{\sigma}$ has $N(0, 1)$.

  \[
P(X < x) = P\left(Z < \frac{x - \mu}{\sigma}\right).
  \]

  (EXAMPLE: Sample Proportion)

  When the population proportion $p$ is 60%, it can be shown that the sample proportion $\hat{p}$ from an SRS of size 100 has approximately $N(0.6, 0.049)$. What is the probability that the sample proportion $\hat{p}$ is different from the population proportion $p$ by more than 5%?  

3
Means and Variances of Random Variables

Consider a discrete random variable $X$ with the probability distribution

<table>
<thead>
<tr>
<th>Value of $X$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$\cdots$</th>
<th>$x_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>$p_1$</td>
<td>$p_2$</td>
<td>$\cdots$</td>
<td>$p_k$</td>
</tr>
</tbody>
</table>

The mean of $X$

- A measure of the center of the distribution.
- $\mu_X$: the mean of a random variable $X$.
- An average of outcomes weighted by their probabilities.

$$
\mu_X = x_1p_1 + x_2p_2 + \cdots + x_kp_k = \sum x_ip_i.
$$

The variance of $X$

- A measure of the variability of the distribution.
- $\sigma^2_X$: the variance of a random variable $X$.
- An average of squared deviation of the variable $X$ from its mean $\mu_X$ weighted by their probabilities.

$$
\sigma^2_X = (x_1 - \mu_X)^2p_1 + (x_2 - \mu_X)^2p_2 + \cdots + (x_k - \mu_X)^2p_k
= \sum(x_i - \mu_X)^2p_i
$$

- The standard deviation $\sigma_X$ of $X$: $\sqrt{\sigma^2_X}$.

EXAMPLE: Pick-3 Game

You choose a three-digit number from 000 to 999. A three-digit winning number is chosen at random. $500 is paid if you win. Let $X$ be the amount of your pay-off in each game.

<table>
<thead>
<tr>
<th>Value of $X$</th>
<th>0</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>.999</td>
<td>.001</td>
</tr>
</tbody>
</table>

(a) What is the mean of $X$?

(b) What is the variance of $X$?

(c) What is the standard deviation of $X$?
(Question) How do we define the mean and variance of a continuous random variable $X$?

The mean $\mu_X$ versus $\bar{x}$

- Denote $n$ observations chosen at random from the probability distribution of a random variable $X$ by $x_1, x_2, \cdots, x_n$.

- What is the connection between a sample mean $\bar{x}$ and the mean $\mu_X$ of the random variable $X$?

- (EXAMPLE: Pick-3 Game)
  - In this example, $x_i$ takes values either $0$ or $500$. The sample mean is given by
    \[
    \bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)
    \]
    \[
    = \frac{1}{n}(0 \times n_{\text{lose}} + 500 \times n_{\text{win}})
    \]
    \[
    = 0 \times \frac{n_{\text{lose}}}{n} + 500 \times \frac{n_{\text{win}}}{n}.
    \]
  - What values would $(n_{\text{lose}}/n)$ and $(n_{\text{win}}/n)$ approach as the sample size $n$ increases?
    - Consequently, as the sample size increases, $\bar{x}$ would get close to $0(0.999) + 500(0.001) = 0.5 = \mu_X$.

- The mean of a random variable, often called the expected value, is interpreted as the long-run average outcome.

Law of Large Numbers

- The law of large numbers is a central theorem in probability, which says that the average of the values of a random variable $X$ observed in many trials must approach $\mu_X$.

- Draw $n$ independent observations at random from a population. Let $X$ be a random variable that takes values of the observations.

  - The sample mean $\bar{x}$ of observations eventually approaches the mean $\mu_X$ of the population as the number of observations increases.

  - The sample variance $s^2$ gets close to the variance $\sigma^2_X$ of the population as the sample size increases.

  - In particular, if $X$ takes 1 or 0 (yes or no), then the law of large numbers says that the sample proportion $\hat{p}$ ($= \bar{x}$) of 1 (yes) approaches the population proportion $p$ ($= \mu_X$).
Rules for Means

- For a random variable $X$ and fixed numbers $a$ and $b$, $\mu_{a+bX} = a + b\mu_X$.
- For random variables $X$ and $Y$, $\mu_{X+Y} = \mu_X + \mu_Y$.

Rules for Variances

- For a random variable $X$ and fixed numbers $a$ and $b$, $\sigma^2_{a+bX} = b^2\sigma^2_X$.
- The Addition Rule for Variances of Independent Random Variables:
  
  For independent variables $X$ and $Y$,
  
  $$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$
  
  $$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$$.

- Two random variables $X$ and $Y$ are independent if knowing the occurrence of any event involving $X$ alone tells us nothing about the occurrence of any event involving $Y$ alone.

EXAMPLE: Pick-3 Game

You buy a $1 ticket for Pick-3 game on each of two different days. Let $X$ and $Y$ be pay-offs on the two tickets.

(a) What is the mean of the net winnings $(X - 1)$ in each game?

(b) What is the variance of the net winnings $(X - 1)$ in each game?

(c) What is the mean of total pay-off $(X + Y)$?

(d) What is the variance of total pay-off $(X + Y)$?

(e) What is the standard deviation of $(X + Y)$?