3.3 Random Variables

Introduction

A random variable assigns a number to each outcome in a sample space.

The possible values of a discrete random variable can be arranged in a (finite or infinite) sequence. The values of a continuous random variable are not between any two there’s a third.

e.g.

- \( W \) = the height of a randomly-chosen woman. \( W \) is a discrete random variable whose possible values are ...

- \( X \) = the number of heads in two tosses of a fair coin. \( X \) is a discrete random variable whose possible values are ...

- \( Y \) = the sum of two fair six-sided dice. \( Y \) is a discrete random variable whose possible values are ...

- \( Z \) = the number of times a coin is flipped until heads appears. \( Z \) is a discrete random variable whose possible values are ...

Discrete Random Variables

The probability mass function (or probability density function) \( p \) of a discrete random variable \( X \) is

\[
p(x) = P(\{\cdots\})
\]

e.g. Referring to the previous example, the probability mass function of \( X \) is

\[
x \begin{array}{c|ccc} p(x) & 0 & 1 & 2 \end{array}
\]

e.g. The mass function of \( Y \) is

\[
y \begin{array}{ccccccccccc} p(y) & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{array}
\]

e.g. The mass function of \( Z \) is \( p(z) = \left\{ \begin{array}{ll} \left(\frac{1}{2}\right)^z, & \text{for } z = 1, 2, 3, \ldots \\
\text{ otherwise} & \end{array} \right. \)

The Cumulative Distribution Function of a Discrete Random Variable

The cumulative distribution function (CDF) \( F \) of a discrete random variable \( X \) is

\[
F(x) = P(\{\cdots\}) = \sum_{t \leq x} p(t) = \sum_{t \leq x} P(X = t)
\]
e.g. The cumulative distribution function of $X$ is

<table>
<thead>
<tr>
<th>$x$</th>
<th>$\ldots$</th>
<th>$-1$</th>
<th>$0$</th>
<th>$1$</th>
<th>$2$</th>
<th>$\ldots$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F(x)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

e.g. The CDF of $Y$ is

| $y$ | $\ldots$ | $0$ | $1$ | $2$ | $3$ | $4$ | $5$ | $6$ | $7$ | $8$ | $9$ | $10$ | $11$ | $12$ | $13$ | $\ldots$ |
|-----|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----------|
| $F(y)$ |        |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |          |

Note that $\sum_x p(x) = \sum_x P(X = x) = 1$.

Mean and Variance for Discrete Random Variables

The expected value (or mean or population mean) of $X$ is the __________________________ of the possible values $X$ can take:

$$\mu_X = E(X) = \sum_x x \cdot p(x)$$

(It’s the __________ of the distribution of $X$. For an $X$ that takes on $n$ equally-likely values, the formula is familiar: $E(X) = \text{___________________________}$.)

e.g. $\mu_X = E(X) = \text{___________________________}$

e.g. $\mu_Y = E(Y) = \text{___________________________}$

e.g. $\mu_Z = E(Z) = \sum_{z=1}^{\infty} z \cdot p(z) = \sum_{z=1}^{\infty} z \cdot \left(\frac{1}{2}\right)^z = 2$ (not demonstrated)

The (population) variance of $X$ is the expected value of its ________________ from its mean:

$$\sigma_X^2 = E[(X - \mu_X)^2] = \sum_x (x - \mu_X)^2 \cdot p(x)$$

(An alternate form is $\sigma_X^2 = \sum_x x^2 \cdot p(x) - \mu_X^2$.)

(For an $X$ that takes on $n$ equally-likely values, the formula is almost familiar: $\sigma_X^2 = \text{___________________________}$.)

e.g. $\sigma_X^2 = E[(X - 1)^2] = \sum_{x=0}^{2} (x - 1)^2 \cdot p(x) = \text{___________________________}$

e.g. $\sigma_Y^2 = E[(Y - 7)^2] = \sum_{y=2}^{12} (y - 7)^2 \cdot p(y) = \text{___________________________}$
The (population) standard deviation of $X$ is $\sigma_X = \sqrt{\sigma_X^2}$.

e.g. $\sigma_X =$

The Probability Histogram

When the possible values of a discrete random variable are evenly spaced, the distribution can be represented by a probability histogram, which, for each value of the variable, has a rectangle centered at that value whose ______ (or ______) indicates the probability of that value.

e.g. The probability histogram for $X$ is

Refining a Histogram

e.g. Here’s a histograms of heights of a sample of women.

Suppose the mean and median are each 64” and the standard deviation is 2.7”. What is the proportion of heights •

- less than 64”?
- less than 61”?
- between 61” and 70”?
Refining the bins (and to avoid empty bins) gives:

Repeated refining leads to a to get under the curve. Rescale its vertical axis under the curve.

Now the under the curve above a range of values is the proportion of heights in that range.

**Continuous Random Variables**

The probabilities of a continuous random variable \( X \) are specified by under a probability density function \( f \) such that

\[
P(a \leq X \leq b) = \int_a^b f(x) \, dx
\]

= area under \( f \) between \( a \) and \( b \) (draw)

Notes:

- \( P(X \leq b) = \int_{-\infty}^b f(x) \, dx \)
- \( P(X \geq a) = \int_a^\infty f(x) \, dx \)
- \( \int_{-\infty}^\infty f(x) \, dx = \) 
- \( P(X = a) = P(a \leq X \leq a) = \int_a^a f(x) \, dx = \), so whether endpoints are included for continuous probability calculations

**Computing Probabilities with the Probability Density Function**

e.g. For the random variable \( X \) with density function \( f(x) = \begin{cases} 2x, & \text{for } 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases} \) (draw),

\[
P(X > \frac{1}{2}) =
\]
The Cumulative Distribution Function of a Continuous Random Variable

The cumulative distribution function $F$ of a continuous random variable $X$ is

$$F(x) = P(\text{__________}) = \int_{-\infty}^{x} f(t) \, dt \text{ (draw)}$$

e.g. Referring to the previous example, find $F(x)$ (draw twice, as area under $f(t)$ and as function).

Mean and Variance for Continuous Random Variables

The expected value (or mean or population mean) of $X$ is the ________________ of the possible values $X$ can take:

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx$$

(It’s the ________________ of the distribution of $X$.)

e.g. Referring to the previous example, $\mu_X = E(X) =$

The (population) variance of $X$ is the expected value of its ________________ from its mean:

$$\sigma_X^2 = E\left[(X - \mu_X)^2\right] = \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f(x) \, dx$$

(An alternate form is $\sigma_X^2 = \int_{-\infty}^{\infty} x^2 \cdot f(x) \, dx - \mu_X^2$.)

e.g. Referring to the previous example, $\sigma_X^2 =$

The (population) standard deviation of $X$ is $\sigma_X = \sqrt{\sigma_X^2}$. 