Probability

Probability Distributions

Given a set $S = \{s_1, \ldots, s_n\}$ a probability distribution is a way to assign probabilities to outcomes ie $P(s_i)$.

To be valid it must satisfy:

(1) $0 \le P(s_i) \le 1$

(2) $P(s_1) + P(s_2) + \cdots + P(s_n) = 1$

(3) If $E \subseteq S$

$$P(E) = \sum_i P(E_i)$$

eg $E = \{s_1, s_2, s_5\}$ $\quad P(E) = P(s_1) + P(s_2) + P(s_5)$

Some immediate results of these properties are:

- $P(S) = 1$
- $P(\{\}) = 0$
- $P(E_1 \cap E_2) = P(E_1) + P(E_2) - P(E_1 \cup E_2)$
- $P(E') = 1 - P(E)$

<u>Using these properties</u>

<u>ex</u> $S = \{s_1, s_2, s_3, s_4\}$

	s_1	s_2	s_3	s_4
$P(s_i)$	.2	.3		.4

- $P(s_3) =$?

$$.2 + .3 + P(s_3) + .4 = 1$$
$$P(s_3) + .9 = 1$$
$$P(s_3) = .1$$

- Let $E = \{s_2, s_3\}$ $E' = \{s_1, s_4\}$

$P(E') =$?

$P(E') = P(s_1) + P(s_4)$
$= .2 + .4$
$= .6$

$P(E) = P(s_2) + P(s_3)$
$= .3 + .1$
$= .4$

$P(E') = 1 - P(E)$
$= 1 - .4$

Two Different Distributions

We'll reference 2 different probability distributions:

1) Estimated Probability aka Relative Frequency
2) Theoritical Probability aka Modeled Probability

Estimated Probability

Suppose you go to grocery stores to examine apples, noting if they are bruised.

The frequency of bruised apples would be the number of bruised apples observed.

The comparison of frequencies can be misleading.

At store 1: frequency of bruised apples was 20
At store 2: frequency of bruised apples was 4

that at store 1 you checked 100 apples and at store 2 you only checked 5.

<u>Relative frequency</u> is the frequency of an event divided by the total sample size.

$$P(E) = \frac{f_r(E)}{\#\text{ of trials}}$$

Relative frequency of bruised apples...

at store 1 is $20/100 = 0.2$

at store 2 is $4/5 = 0.8$

As a probability distribution relative frequency aka estimated probability has limitations. The field of statistics studies ways to measure the 'trustworthiness' of estimated probabilities.

Theoretical / Modeled Probability

Our second probability distribution, modeled probability, defines the probability of an outcome as the predicted relative frequency of that outcome over a large number of trials.

The fact that the relative frequency of an event stabilizes over large trials comes from the Central Limit Theorem.

For experiments where every outcome can be assumed to be equally likely we can directly calculate its modeled probability.

Calculating Modeled Probability

Given a set of outcomes $S = \{s_1, \ldots, s_n\}$ if every outcome is equally likely

$$P(s_i) = \frac{1}{n(S)}$$

$$P(E) = \frac{n(E)}{n(S)}$$

ex Experiment: flipping a coin 3 times

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

$$P(HHH) = \frac{1}{n(S)} = \frac{1}{8}$$

$$P(\text{at least 1 heads}) = \frac{n(\text{at least 1 heads})}{n(S)} = \frac{7}{8}$$

$$P(\text{at most 1 heads}) = \frac{3}{8}$$

ex Experiment: rolling 2 colored dice

S = {

11,	12,	13,	14,	15,	16,
21,	22,	23,	24,	25,	26,
31,	32,	33,	34,	35,	36,
41,	42,	43,	44,	45,	46,
51,	52,	53,	54,	55,	56,
61,	62,	63,	64,	65,	66 }

$$P(\text{dice sum to } 7) = \frac{n(\text{sum to } 7)}{n(S)}$$

$$= \frac{6}{36} = 1/6$$

$$P(\text{dice both even})$$

$$= \frac{9}{36} = 1/4$$