10:20 7/29/2011

Chapter 7

Discrete Probability

7.1 INTRO TO PROBABILITY

DEF: The study of probability uses special jargon.

- A ***sample space*** is a nonempty set.
- An ***experiment*** is a process that produces a point in a sample space.
- An ***event*** is a subset of a sample space.
- The ***event space*** is the power set of the sample space.

Experiment 7.1.1: Toss a coin.
Sample space: $U = \{H, T\}$
Event space: $P(U) = \big\{\emptyset, \{H\}, \{T\}, \{H, T\}\big\}$.

Remark: Often, some of the more interesting events have special names.

Experiment 7.1.2: Toss two coins.
Sample space: $U = \{HH, HT, TH, TT\}$
Event space: subsets of U
Named events:

- ***match*** $= \{HH, TT\}$
- ***at least one head*** $= \{HT, TH, HH\}$

UNIFORM PROBABILITY MEASURE

DEF: The ***uniform probability measure*** on a finite sample space S assigns to each event E the probability

$$p(E) = \frac{|E|}{|S|}$$

Experiment 7.1.1, continued: coin toss.

Sample space: $U = \{H, T\}$

Event	$\emptyset$	$\{H\}$	$\{T\}$	$\{H, T\}$
Probability	0	$\frac{1}{2}$	$\frac{1}{2}$	1

This example is readily generalized.

Experiment 7.1.2, continued: two coins

Sample space: $U = \{HH, HT, TH, TT\}$

Event Name	at least one head	one of each
Event	$\{HT, TH, HH\}$	$\{HT, TH\}$
Probability	$\frac{3}{4}$	$\frac{2}{4}$

Experiment 7.1.3: roll a die

Sample space: $U = \{1,2,3,4,5,6\}$

Event Name	odd	even	2 mod 3
Event	$\{1,3,5\}$	$\{2,4,6\}$	$\{2,5\}$
Probability	$\frac{3}{6}=\frac{1}{2}$	$\frac{3}{6}=\frac{1}{2}$	$\frac{2}{6}=\frac{1}{3}$

Experiment 7.1.4: roll two dice

Sample space: $U = \{\overbrace{11,\ldots,16}, \ \ldots, \ \overbrace{61,\ldots,66}\}$

Event Name	doubles	sum is nine
Event	$\{11,22,\ldots,66\}$	$\{36,45,54,63\}$
Probability	$\frac{6}{36}=\frac{1}{6}$	$\frac{4}{36}=\frac{1}{9}$

Experiment 7.1.5: roll three dice

Sample space: $U = \{\overbrace{111,\ldots,116}, \ \ldots, \ \overbrace{661,\ldots,666}\}$

Event Name	triples	doubles	singles
Event	$\{111, \text{ etc.}\}$	$\{112, \text{ etc.}\}$	$\{123, \text{ etc.}\}$
Probability	$\frac{6}{216}=\frac{1}{36}$	$\frac{90}{216}=\frac{5}{12}$	$\frac{120}{216}=\frac{5}{9}$

Experiment 7.1.6: A base-10 numeral is randomly chosen from the range

$$000 \ldots 999$$

Q1: What is the probability that the numeral contains no 3's or 5's?

Ans. There are 8^3 base-10 numerals containing no 3's or 5's and 10^3 three-digit numerals altogether. Thus, the probability is

$$\frac{8^3}{10^3} = \frac{512}{1000} = 0.512$$

Q2: What is the probability that the numeral contains one 3 and no 5's?

Ans. The 3 could occur as each of the three digits. There would be 8^2 possibilities for the other two digits. Thus, the probability is

$$\frac{3 \cdot 8^2}{10^3} = \frac{192}{1000} = 0.192$$

DEF: ***(standard) deck of cards***

$$\begin{array}{cccccccc} 2\clubsuit & 3\clubsuit & \ldots & 10\clubsuit & J\clubsuit & Q\clubsuit & K\clubsuit & A\clubsuit \\ 2\diamondsuit & 3\diamondsuit & \ldots & 10\diamondsuit & J\diamondsuit & Q\diamondsuit & K\diamondsuit & A\diamondsuit \\ 2\heartsuit & 3\heartsuit & \ldots & 10\heartsuit & J\heartsuit & Q\heartsuit & K\heartsuit & A\heartsuit \\ 2\spadesuit & 3\spadesuit & \ldots & 10\spadesuit & J\spadesuit & Q\spadesuit & K\spadesuit & A\spadesuit \end{array}$$

Experiment 7.1.7: deal one card from deck

Sample space: standard 52-card deck

Event Name	heart	seven
Event	$\{2\heartsuit, \ldots, A\heartsuit\}$	$\{7\clubsuit, 7\diamondsuit, 7\heartsuit, 7\spadesuit\}$
Probability	$\frac{13}{52} = \frac{1}{4}$	$\frac{4}{52} = \frac{1}{13}$

$$\begin{array}{cccccccc}
2\clubsuit & 3\clubsuit & \ldots & 10\clubsuit & J\clubsuit & Q\clubsuit & K\clubsuit & A\clubsuit \\
2\diamondsuit & 3\diamondsuit & \ldots & 10\diamondsuit & J\diamondsuit & Q\diamondsuit & K\diamondsuit & A\diamondsuit \\
2\heartsuit & 3\heartsuit & \ldots & 10\heartsuit & J\heartsuit & Q\heartsuit & K\heartsuit & A\heartsuit \\
2\spadesuit & 3\spadesuit & \ldots & 10\spadesuit & J\spadesuit & Q\spadesuit & K\spadesuit & A\spadesuit
\end{array}$$

Experiment 7.1.8: deal five cards from deck

Sample space: all possible 5-card hands.

Event Name	Probability
4 of a kind	$\dfrac{\binom{13}{1}\cdot\binom{48}{1}}{\binom{52}{5}}$
full house	$\dfrac{\binom{13}{1}\cdot\binom{4}{3}\cdot\binom{12}{1}\cdot\binom{4}{2}}{\binom{52}{5}}$
3 of a kind	$\dfrac{\binom{13}{1}\cdot\binom{4}{3}\cdot\binom{48}{1}\cdot\binom{44}{1}}{\binom{52}{5}}$

INFINITE SAMPLE SPACES (optional)

If a sample space is finite, it is sufficient to specify the probabilities of singletons, and to infer the probabilities of compound events by addition. However, this does not work for infinite sample spaces.

Experiment 7.1.9: pick a natural number

Probabilities of singletons

$$0 = p(0) = p(1) = p(2) = \cdots$$

do not yield probabilities of infinite subsets:

$$p(\text{even}), \quad p(\text{power of two})$$

Uniform probability for an arbitrary set E of numbers is defined by the rule

$$p(E) = \lim_{n\to\infty} \frac{|\{E \cap \{0, 1, \ldots, n-1\}|}{n}$$

This illustrates why the most general kind of probability measures (discussed elsewhere, but not here) are assigned to the event space, rather than to the sample space.

PROBABILITY of the COMPLEMENTARY EVENT

Proposition 7.1.1. *Let E be an event in a finite sample space S, under the uniform probability distribution. Then*

$$p(\overline{E}) = 1 - p(E)$$

Pf:

$$
\begin{array}{lrcl}
(1) & |S| & = & |E| + |\overline{E}| \quad \text{by Rule of Sum} \\
(2) & \dfrac{|S|}{|S|} & = & \dfrac{|E|}{|S|} + \dfrac{|\overline{E}|}{|S|} \\
(3) & 1 & = & p(|E|) + p(\overline{E}) \\
(4) & p(\overline{E}) & = & 1 - p(E)
\end{array}
\qquad \diamondsuit
$$

Experiment 7.1.10: A (putatively fair) coin is tossed 10 times. Find the probability of at least one tail.

$$
\begin{aligned}
\text{Answer} : p\,(\#\text{tails} \geq 1) &= 1 - p\,(10 \text{ heads}) \\
&= 1 - \frac{1}{1024} \\
&= \frac{1023}{1024}
\end{aligned}
$$

PROBABILITY of a UNION of EVENTS

Proposition 7.1.2. *Let E_1 and E_2 be events in a finite sample space S, under the uniform probability distribution. Then*

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

Pf:

$$|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2| \quad \text{by Incl-Excl}$$

$$\frac{|E_1 \cup E_2|}{|S|} = \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|}$$

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2) \qquad \diamondsuit$$

Experiment 7.1.11: An integer is chosen from the interval $[1, \ldots, 100]$. Find the probability that it is divisible either by 6 or by 15.

Solution:

$$\begin{aligned} p(6\backslash n \vee 15\backslash n) &= p\,(6\backslash n) + p\,(15\backslash n) - p\,(30\backslash n) \\ &= \frac{16}{100} + \frac{6}{100} - \frac{3}{100} \\ &= \frac{19}{100} \end{aligned}$$

7.2 PROBABILITY THEORY

In §6.1 we defined the *uniform probability measure* for an event E in a *finite sample space* S:

$$pr(E) = \frac{|E|}{|S|}$$

In general, a ***discrete probability measure*** assigns probabilities to all events of a *finite or countably infinite sample space.*

DEF: A ***discrete probability measure*** assigns to every subset E of a countable sample space S a real number $pr(E)$ satisfying these three axioms:

A1. $0 \leq pr(E) \leq 1$, for every event $E \subseteq S$.

A2. $pr(S) = 1$.

A3. For any finite or countably infinite collection of mutually exclusive subsets $\{A_j : j \in J\}$ of the sample space S,

$$pr\left(\bigcup_{j \in J} A_j\right) = \sum_{j \in J} pr(A_j)$$

Example 7.2.1: Uniform probability measure (§6.1) on a finite sample space is a probability measure. The three axioms are easily verified.

Example 7.2.2: For any finite sample space S, assign to each possible outcome $s \in S$, a number $p(s)$, such that

(1) $$0 \leq p(s) \leq 1$$

(2) $$\sum_{s \in S} p(s) = 1$$

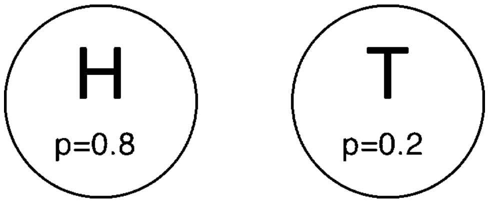
Then to any event E, we assign the probability

$$pr(E) = \sum_{s \in E} p(s)$$

TWO STANDARD EXAMPLES

Here are two non-uniform probability measures for some familiar sample spaces.

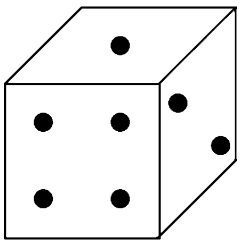
DEF: The ***standard loaded coin*** has probability $p(H) = 0.8$ and $p(T) = 0.2$.



DEF: The ***standard loaded die*** has sample space

$$U = \{1, 2, 3, 4, 5, 6\}$$

Then assign the singleton $\{j\}$ the probability $\frac{j}{21}$.



PROPERTIES of PROBABILITY MEASURES

Prop 7.2.1. *Let p be a probability measure on a sample space S. Then*

$$p(\emptyset) \;=\; 0$$

Pf:

$$\begin{aligned} p(S) &= p(S \cup \emptyset) \quad \text{since } S = S \cup \emptyset \\ &= p(S) + p(\emptyset) \quad \text{by Axiom A3} \end{aligned} \qquad \diamondsuit$$

Prop 7.2.2. *Let p be a probability measure on a sample space S, and let E be an event. Then*

$$p(\overline{E}) \;=\; 1 - p(E)$$

Pf:

$$\begin{aligned} 1 &= p(S) \quad \text{by Axiom A2} \\ &= p(E \cup \overline{E}) \quad \text{since } S = E \cup \overline{E} \\ &= p(E) + p(\overline{E}) \quad \text{by Axiom A3} \end{aligned} \qquad \diamondsuit$$

BERNOULLI DISTRIBUTIONS

DEF: A ***Bernoulli distribution*** is a probability measure on a sample space with exactly two points.

Example 7.2.3: Flip standard loaded coin.

Sample space $= \{H, T\}$

Event space $= \{\emptyset, \{H\}, \{T\}, \{H,T\}\}$

$$pr(\{H\}) = \frac{4}{5} \qquad pr(\{T\}) = \frac{1}{5}$$

Example 7.2.4: Roll a fair die.

Sample space $= \{1, 2, 3, 4, 5, 6\}$

$$pr(3) = \frac{1}{6} \quad pr(\neg 3) = pr(\{1,2,4,5,6\}) = \frac{5}{6}$$

Example 7.2.5: The standard loaded die induces a Bernoulli distribution with

$$pr(even) = \frac{4}{7} \quad \text{and} \quad pr(odd) = \frac{3}{7}$$

NOTATION: Let $U = \{x, y\}$ be a binary sample space. Then the set

$$\{\, x\cdots xxx, \quad x\cdots xxy, \quad ..., \quad y\cdots yyy \,\}$$

of length-n sequences in U is denoted U^n.

REVIEW : The number of n-strings in $\{x, y\}$ with k occurrences of x and $n-k$ occurrences of y is

$$\binom{n}{k}$$

Prop 7.2.3. *A Bernoulli distribution on* $\{x, y\}$

$$p(x) = p \qquad p(y) = 1 - p$$

induces a probability measure on the sample space $\{0,\ 1,\ ...,\ n\}$ *in which, for* $k = 0, 1, \ldots, n$

$$pr(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

is the probability of obtaining k *occurrences of* x *and* $n - k$ *occurrences of* y. ♢

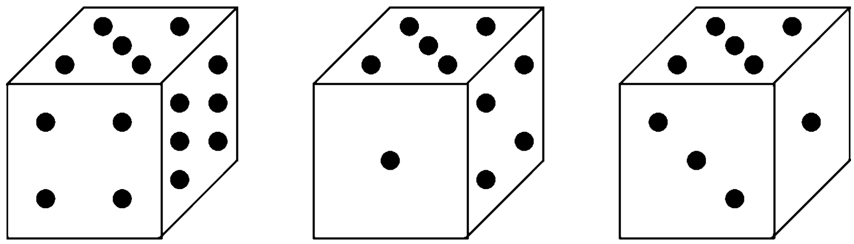
DEF: A ***binomial probability distribution*** on the sample space $\{0, 1, \ldots, n\}$ is a probability induced by a Bernoulli distribution, as in Prop 7.2.3.

Example 7.2.6: Flip standard loaded coin 10 times. Then for $k = 0, \ldots, 10$

$$pr(k \text{ heads}) = \binom{10}{k} 0.8^k \, 0.2^{10-k}$$

CLASSROOM EXERCISE

Seven standard loaded dice are rolled.
What is the probability of three 5's ??



CONDITIONAL PROBABILITY

Example 7.2.7: A fair dime and a fair nickel are flipped. An oracle tells you that at least one of them is heads. What is the probability that the other also is heads?

Analysis: If the oracle says the dime is heads, then there are two equally likely cases for the nickel. Similarly, if the oracle says that the nickel is heads, then there are two equally likely cases for the dime.

Question: How many cases are there if the oracle doesn't say which coin?

Experiment: Program a computer to initialize the variables n and $h2$ at 0. Run 10,000 trials. On each trial, generate two random real numbers between 0 and 1. If at least one is less than 0.5, then increment the variable n by 1. If both are less than 0.5, then also increment $h2$ by 1.

Calculate the fraction $\frac{h2}{n}$.

DEF: Let p be a probability distribution on a sample space U, and let Y be an event. Then the ***conditional probability*** of event E given that event Y has occurred is

$$pr(E \mid Y) = \frac{p(E \cap Y)}{p(Y)}$$

Example 7.2.7, continued:

$$p(HH \mid \neg TT) = \frac{p(HH \wedge \neg TT)}{p(\neg TT)} = \frac{p(HH)}{p(\neg TT)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Example 7.2.8: Two fair dice are rolled. At least one is a four. What is the probability that both are fours?

$$p(44 \mid \text{1-or-2 4's}) = \frac{p(44 \wedge \text{1-or-2 4's})}{p(\text{1-or-2 4's})} = \frac{1/36}{11/36} = \frac{1}{11}$$

INDEPENDENCE

DEF: Let pr be a probability distribution on a sample space U. Event E is ***probabilistically independent*** of event Y if $pr(E \mid Y) = pr(E)$.

Prop 7.2.4. *Let pr be a probability distribution on a sample space U, and let event E be probabilistically independent of event Y. Then event Y is probabilistically independent of event E.*

Pf: $pr(E) = pr(E \mid Y) = pr(E \cap Y)/p(Y)$. Therefore,

$$\begin{aligned} pr(Y) &= \frac{pr(E \cap Y)}{pr(E)} \\ &= \frac{pr(Y \cap E)}{pr(E)} = pr(Y \mid E) \qquad \diamondsuit \end{aligned}$$

Example 7.2.9: Roll two standard loaded dice. Let Y be the event that the first die is a one, and E the event that the sum is odd. Then

$$pr(E) = 2 \cdot \frac{3}{7} \cdot \frac{4}{7} = \frac{24}{49} \quad \text{and}$$

$$pr(E \mid Y) = \frac{pr(E \cap Y)}{p(Y)} = \frac{pr(12, 14, 16)}{1/21}$$

$$= \frac{2/441 + 4/441 + 6/441}{1/21} = \frac{12}{21} = \frac{4}{7}$$

Example 7.2.10: Roll two fair dice. Let Y be the event that the first die is a one, and E the event that the sum is odd. Then

$$pr(E) = 2 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2} \quad \text{and}$$

$$pr(E \mid Y) = \frac{p(E \cap Y)}{p(Y)} = \frac{p(12, 14, 16)}{1/6}$$

$$= \frac{3/36}{1/6} = \frac{3}{6} = \frac{1}{2}$$

Remark: Disjoint events are usually NOT probabilistically independent.

RANDOM VARIABLES

DEF: A ***probability space*** is a pair consisting of a sample space U, called its ***domain*** and a probability measure on U.

DEF: A ***random variable*** is a real-valued function on the domain of a probability space.

Remark: Since a random variable is a function, it is *not* a variable, and it is not random.

Example 7.2.11: Flip a coin three times. Let $X(t)$ be the number of heads that occurs. Then

$$\begin{aligned} X(TTT) &= 0 \\ X(TTH) = X(THT) = X(HTT) &= 1 \\ X(THH) = X(HHT) = X(HTH) &= 2 \\ X(HHH) &= 3 \end{aligned}$$

7.3 BAYES'S THEOREM

Bayes's Theorem is a powerful tool for inference. The theorem itself is motivated by examples of the following type.

Example 7.3.1: An internet service provider estimates that two-thirds of all email messages are spam, and it decides to filter out some of the most frequent spam varieties by identifying characteristic keywords.

In particular, some 12600 of 100000 recent spam messages contained the word "Rolex". However, in another 20000 messages that were not spam, the word "Rolex" appeared 100 times. A message arrives containing the word "Rolex". Estimate the probability that the message is spam.

Modeling. Let E be the event that an arriving message is spam, and let R be the event that the arriving message contains the word "Rolex". You are asked to estimate

$$pr(E \mid R)$$

from the information above.

continued

Example 7.3.1, continued: By the definition of conditional probability, we have

$$pr(E \mid R) = \frac{pr(E \cap R)}{pr(R)}$$

Neither the numerator nor the denominator occurs directly in the information we are given,

$$\begin{aligned} pr(E) &\approx \frac{2}{3} \\ pr(R \mid E) &\approx \frac{12600}{100000} = 0.126 \\ pr(R \mid \overline{E}) &\approx \frac{120}{20000} = 0.006 \end{aligned}$$

Nonetheless, it is sufficient to calculate them both:

$$\begin{aligned} pr(E \cap R) &= pr(R \cap E) && (1) \\ &= pr(R \mid E)\, pr(E) \\ &\approx 0.126 \cdot \frac{2}{3} = 0.084 \\ pr(R) &= pr(R \cap E) + pr(R \cap \overline{E}) && (2) \\ &= pr(R \mid E)\, pr(E) + pr(R \mid \overline{E})\, pr(\overline{E}) \\ &\approx 0.126 \cdot \frac{2}{3} + 0.006 \cdot \frac{1}{3} \\ &= 0.084 + 0.002 = 0.086 \end{aligned}$$

Example 7.3.1, continued: In summary, we have

$$pr(E \mid R) = \frac{pr(E \cap R)}{pr(R)} \approx \frac{0.084}{0.086} \approx 0.9767$$

In general, the formula we have used to calculate $pr(E \mid R)$ is called ***Bayes's Theorem***.

Bayes's Theorem.

$$pr(E \mid R) = \frac{pr(R \mid E)\, pr(E)}{pr(R \mid E)\, pr(E) + pr(R \mid \overline{E})\, pr(\overline{E})}$$

Pf: The formulas for the numerator and denominator are derived in Example 7.3.1. ◊

In Example 7.3.1, we could have estimated

$$pr(E \mid R)$$

directly by partitioning a sample of messages containing the word "Rolex" into spam and non-spam. In various other circumstances, direct measurement may be impossible.

7.4 EXPECTED VALUE AND VARIANCE

DEF: The **e**xpected value of a random variable X on a discrete probability space (S, p) is the sum

$$E(X) = \sum_{s \in S} X(s)p(s)$$

Example 7.4.1: The expected outcome of a fair die is

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$$
$$= \frac{21}{6} = \frac{7}{2}$$

Example 7.4.2: The expected outcome of the standard loaded die is

$$1 \cdot \frac{1}{21} + 2 \cdot \frac{2}{21} + \cdots + 6 \cdot \frac{6}{21} = \frac{91}{21} = \frac{13}{3}$$

Over a non-uniform probability space the expected value of a random variable is the weighted mean.

Example 7.4.3: Flip a fair coin three times. The expected number of heads is

$$0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} \;=\; \frac{12}{8} \;=\; \frac{3}{2}$$

This calculation is based on the binomial distribution.

Example 7.4.4: Flip a standard loaded coin three times. The expected number of heads is

$$\begin{aligned} & 0 \cdot \frac{1}{125} + 1 \cdot \frac{12}{125} + 2 \cdot \frac{48}{125} + 3 \cdot \frac{64}{125} \\ = \;& \frac{0 + 12 + 96 + 192}{125} \;=\; \frac{300}{125} \;=\; \frac{12}{5} \end{aligned}$$

This calculation too is based on the binomial distribution.

SUMMING RANDOM VARIABLES

Theorem 7.4.1. *Let X_1 and X_2 be random variables on a probability space (S, p). Then*

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

Pf:

$$\begin{aligned} E(X_1 + X_2) &= \sum_{s \in S} (X_1(s) + X_2(s))p(s) \\ &= \sum_{s \in S} X_1(s)p(s) + X_2(s)p(s) \\ &= \sum_{s \in S} X_1(s)p(s) + \sum_{s \in S} X_2(s)p(s) \\ &= E(X_1) + E(X_2) \end{aligned}$$

Example 7.4.5: When two fair dice are rolled, here are both calculations:

$$E(X_1) + E(X_2) = \frac{7}{2} + \frac{7}{2} = 7 \quad \text{and}$$

$$E(X_1 + X_2) = \frac{1}{36} \sum_{j=1}^{6} \sum_{k=1}^{6} (j + k) = \frac{252}{36} = 7$$

Theorem 7.4.2. *Let $X_1, \ldots, X_n$ be random variables on a probability space (S, p). Then*

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$$

Pf: By induction on n, using Thm 7.4.1. ♢

Example 7.4.6: When 100 fair coins are tossed, the expected number of heads is

$$\frac{1}{2} \cdot 100 = 50$$

Example 7.4.7: When 100 standard loaded coins are tossed, the expected number of heads is

$$0.8 \cdot 100 = 80$$

GEOMETRIC DISTRIBUTION

DEF: The geometric distribution on the positive integers is

$$pr(k) = (1-p)^{k-1}p$$

Example 7.4.8: A coin with $p(H) = p$ is tossed until the first occurrence of heads. Then the probability of requiring exactly k tosses is $(1-p)^{k-1}p$. We observe that

$$\begin{aligned}\sum_{k=1}^{\infty}(1-p)^{k-1}p &= p\sum_{k=1}^{\infty}(1-p)^{k-1} \\ &= \frac{p}{1-(1-p)} = 1\end{aligned}$$

It is proved in the text that

$$\begin{aligned}E(X) &= \sum_{k=1}^{\infty}(1-p)^{k-1}pk \\ &= p\,\frac{d}{dx}(1-x)^{-1}\Big|_{x=1-p} = \frac{1}{p}\end{aligned}$$

INDEPENDENT RANDOM VARIABLES

DEF: The random variables X and Y on the probability space (S, p) are ***i***ndependent if for all real numbers r_1 and r_2

$$p(X = r_1 \wedge Y = r_2) \;=\; p(X = r_1) \cdot p(Y = r_2)$$

Example 7.4.9: Suppose that X is the sum of two fair dice and that Y is the product. Then

$$p(X = 2) = \frac{1}{36} \quad \text{and} \quad p(Y = 5) = \frac{1}{18}$$

However,

$$p(X = 2 \wedge Y = 5) \;=\; 0 \;\neq\; \frac{1}{36} \cdot \frac{1}{18}$$

Thus X and Y are not independent.

VARIANCE and STANDARD DEVIATION

DEF: The **v**ariance of a random variable X on a probability space (S,p) is the sum

$$\sigma^2(X) \;=\; V(X) \;=\; \sum_{s\in U} \big(X(s) - E(X)\big)^2 p(s)$$

DEF: The **s**tandard deviation of a random variable X on a probability space (U,p) is

$$\sigma(X) \;=\; \sqrt{V(X)}$$

Example 7.4.10: Flip a fair coin three times. The variance of the number of heads is

$$\left(0-\frac{3}{2}\right)^2\cdot\frac{1}{8}+\left(1-\frac{3}{2}\right)^2\cdot\frac{3}{8}+\left(2-\frac{3}{2}\right)^2\cdot\frac{3}{8}+\left(3-\frac{3}{2}\right)^2\cdot\frac{1}{8}$$
$$= \frac{9}{4}\cdot\frac{1}{8} + \frac{1}{4}\cdot\frac{3}{8} + \frac{1}{4}\cdot\frac{3}{8} + \frac{9}{4}\cdot\frac{1}{8} = \frac{24}{32} = \frac{3}{4}$$

The standard deviation is

$$\sqrt{\frac{3}{4}} \;=\; \frac{\sqrt{3}}{2}$$

Example 7.4.11: Flip a standard loaded coin three times. The variance of the number of heads is

$$\left(0 - \frac{12}{5}\right)^2 \cdot \frac{1}{125} \;+\; \left(1 - \frac{12}{5}\right)^2 \cdot \frac{12}{125}$$
$$+\left(2 - \frac{12}{5}\right)^2 \cdot \frac{48}{125} \;+\; \left(3 - \frac{12}{5}\right)^2 \cdot \frac{64}{125}$$
$$= \frac{144 + 49 \cdot 12 + 4 \cdot 48 + 9 \cdot 64}{5^5} = \frac{1500}{5^5} = \frac{12}{25}$$

The standard deviation is

$$\sqrt{\frac{12}{25}} = \frac{2\sqrt{3}}{5}$$

CHEBYSHEV INEQUALITY

Chebyshev Inequality. *Let X be a random variable on any probability space that has a mean and a variance. Then*

$$pr\big(\,|X(s) - E(X)| \geq k\sigma(X)\big) \;\leq\; \frac{1}{k^2}$$