# Probability and Random Variables

Sanford Gordon

February 28, 2003

## 1 The Sample Space

*Random experiment:* repeatable procedure with uncertain outcome.

*Sample outcome:* A possible observation of the experiment, labeled $s$.

*Sample space:* Set of all possible outcomes of an experiment, $S$.

For example, consider the experiment of two coins flipped at once:

$$S = \{(h, h), (h, t), (t, h), (t, t)\}$$

Or consider the experiment of a dart thrown at a (very large) target:

$$S = \mathbb{R}^2 = \{x, y : -\infty < x < \infty, -\infty < y < \infty\}$$

An event $E$ is a subset of the sample space $E \subseteq S$. If we observe outcome $s$ and $s \in E$, we say “the event $E$ occurred.”

## 2 Probability

Two events, $E$ and $F$, are said to be disjoint (mutually exclusive) if $E \cap F = \emptyset$.

The complement of $E$, $E^c$, is the set of points in $S$ but not in $E$. $E^c = \{x | x \in S \text{ and } x \notin E\}$. Accordingly, $E \cup E^c = S$.

Let $\mathcal{F}$ be the collection of all events. A function $P : \mathcal{F} \rightarrow [0, 1]$ is called a *probability measure* provided:

1) $P(S) = 1$.

2) If $E_1, E_2, \ldots, E_n$ are disjoint events, then $P(\cup_i E_i) = \sum_i P(E_i)$.
Corollary: $P(E) + P(E^c) = 1.$

Proof:

\[
E \cup E^c = S \\
E \cap E^c = \emptyset \\
P(E) + P(E^c) = P(E \cup E^c) = P(S) = 1
\]

If events are not disjoint, $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

3 Conditional Probability

(A simple example: red and white balls in an urn)

3.1 Definition

The conditional probability of an event $F$ given event $E$ has occurred is a function, $P(F|E)$ such that:

\[
P(E \cap F) = P(F)P(E|F) = P(E)P(F|E).
\]

Rearranging gives

\[
P(E|F) = \frac{P(E \cap F)}{P(F)}
\]

3.2 Total Probability

$N$ disjoint events $F_i | i = 1^N$ constitute a partition of $S$ if $\cup_i F_i = S$. Given such a partition and an event $E$,

\[
P(E) = P(E \cap F_1) + P(E \cap F_2) + \cdots + P(E \cap F_N)
\]

\[
= \sum_{i=1}^{N} P(E \cap F_i)
\]

Using (1), we have the law of total probability:
\[ P(E) = \sum_{i=1}^{N} P(E|F_i)P(F_i). \quad (4) \]

### 3.3 Independence

Events \( E \) and \( F \) are independent if \( P(E \cap F) = P(E)P(F) \).

Given equation (1), this is equivalent to stating \( P(F|E) = P(F) \).

### 3.4 Bayes Rule

Let \( F_i \mid i=1 \) constitute a partition of \( S \). By the definition of conditional probability, \( P(F_j|E) = P(F_j \cap E)/P(E) \). Again using the definition, we can express the numerator as \( P(E|F_j)P(F_j) \). Using the law of total probability, we can express the denominator as \( \sum_{i=1}^{N} P(E|F_i)P(F_i) \). So

\[ P(F_j|E) = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{N} P(E|F_i)P(F_i)} \quad (5) \]

Equation (5) is known as Bayes Rule.

(A famous example: the blood test)

### 4 Random Variables

#### 4.1 Preliminaries

A random variable \( X \) is a function that assigns a real value to each possible outcome \( s \) in the sample space \( S \). Formally, \( X : S \rightarrow \mathbb{R} \).

Consider a role of two dice. One possible random variable is a measure of whether the sum exceeds six.

\[ X(s) = \begin{cases} 1 & \text{if the sum exceeds six} \\ 0 & \text{otherwise.} \end{cases} \]
4.2 Discrete Random Variables

4.2.1 Definitions

A random variable \( X \) is discrete if it takes a finite number of values or a countably infinite number of values.

The \textit{probability mass function} or pmf of \( X \) is \( p(a) = P\{X = a\} \). For admissible values of \( X, x_i, p(x_i) > 0 \); for all other values, \( p(x) = 0 \). A proper pmf has the property \( \sum_{i=1}^{\infty} p(x_i) = 1 \).

The \textit{cumulative distribution function} (cdf) \( F(a) \) is

\[
F(a) = \sum_{x_i \leq a} p(x_i)
\]

4.2.2 Expectation and Variance of Discrete Random Variables

Definition of Expectation: \( E[x] = \sum_{i=1}^{\infty} x_i p(x_i) \),

Rules for expectations:

\[
E[g(X)] = \sum_{i=1}^{\infty} g(x_i)p(x_i) \\
E[aX + b] = aE[X] + b \quad \text{for constants } a, b \\
E[X + Y] = E[X] + E[Y] \quad \text{if } X, Y \text{ are both random variables}
\]

Definition of Variance: \( Var(X) = E[(X - E(X))^2] \)

4.2.3 Bernoulli Random Variables

\( X \) is distributed \textit{Bernoulli} if it takes on values of 0 or 1 with a given probability. The pmf is

\[
p(0) = P\{X = 0\} = 1 - p \\
p(1) = P\{X = 1\} = p
\]

Let’s calculate the mean and variance.
4.2.4 Geometric Random Variables

Conduct independent Bernoulli trials with success probability $p$ until a success occurred. The number of trials until the first success is a random variable said to follow a geometric distribution with parameter $p$. The pmf is:

$$p(k) = P\{X = k\} = p(1 - p)^{k-1}$$

4.2.5 Binomial Random Variables

Preliminaries.

Generically, for $n$ trials, the number of possible sequences of $k$ successes is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

The number of successes in $n$ Bern($p$) trials, $X$, is a binomial random variable with parameters $p$ and $n$. Note that $X$ can take on values from zero to $n$. The pmf is

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

4.3 Continuous Random Variables: Preliminaries

A random variable $X$ is continuous if it can take on a continuum of values.

Define a function $f(x)$ for all $x \in \mathbb{R}$ with the property that for any two constants $a$ and $b$, with $a \leq b$,

$$P\{X \in [a, b]\} = \int_a^b f(x)dx.$$ 

The function $f(x)$ is called the probability density function (pdf) of $X$. Note that

$$P\{X \in [-\infty, +\infty]\} = \int_{-\infty}^{+\infty} f(x)dx = 1.$$ 

Also,

$$P\{X = a\} = \int_a^a f(x)dx = 0.$$
Next, define the cumulative distribution function of continuous random variable $X$, $F(a)$ as

$$P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)\,dx = 0.$$ 

### 4.3.1 Expectation of Continuous Random Variables

The intuition here follows exactly from discrete random variables, except we replace summations with integrals:

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

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\,f(x)\,dx$$

### 4.3.2 Uniform Random Variables

$X$ is said to be distributed Uniform over the interval $(0,1)$, i.e. $U(0, 1)$ if

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

A random variable is said to be distributed Uniform over the interval $(a,b)$, i.e. $U(a, b)$, for $a < b$, if

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise.} \end{cases}$$

### 4.3.3 Normal (Gaussian) Random Variables

A random variable is said to be normally distributed with mean $\mu$ and variance $\sigma^2$, i.e. $N(\mu, \sigma^2)$ if

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.$$ 

If $\mu = 0$ and $\sigma^2 = 1$, $X$ is said to be a standard normal random variable. The standard normal pdf is often represented as $\phi(x)$.
\[
\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}
\]

Likewise, the cdf of a standard normal random variable is represented as \( \Phi(\cdot) \). There is no known closed-form expression for \( \Phi \).

Let \( X \sim N(\mu, \sigma^2) \). Define a new variable, called the “z-transform” of \( X \): \( Z = (X - \mu)/\sigma \). Then \( Z \sim N(0, 1) \).