Expected values

The *average* or *mean* or *expected value* of \( x_1, x_2, x_3, \ldots, x_n \) is

\[
\frac{x_1 + x_2 + \ldots + x_n}{n} = x_1 \frac{1}{n} + x_2 \frac{1}{n} + \ldots + x_n \frac{1}{n} = \sum_1^n x_ip(x_i)
\]

where \( p(x_i) = \frac{1}{n} \) is the probability of \( x_i \) assuming the \( n \) numbers are equally likely. If the \( x_i \)'s are not equally likely, then \( \frac{x_1 + x_2 + \ldots + x_n}{n} \) is no longer correct but \( \sum_1^n x_ip(x_i) \) remains correct. For a continuous random variable \( X \) the expected value of \( X \) is \( E(X) = \int xf(x)dx \) where \( f(x) \) is the probability density function for the values of \( X \).
Besides computing the average $E(X)$ of a random variable, we often wish to compute the expected value $E(g(X))$ of a function $g(x)$ which acts on the values $x$ of a random variable $X$.

Suppose:
- $X$ is a random variable,
- $g(x)$ is a function on the values $x$ of $X$,
- $S$ is the state space (set of possible values $x$ of $X$),
- $p(x) = P[X = x]$ is the probability mass function (discrete case) or (continuous case) that
- $f(x)$ is the probability density function for $X$.

**Definition.** The *expected value* of $g(x)$ is

\[
E[g(X)] = \sum_{x \in S} g(x)p(x) \text{ (discrete case)}
\]

or

\[
E[g(X)] = \int_S g(x)f(x)dx \text{ (continuous case)}.
\]
A fair coin is tossed. Suppose

\[ X = \begin{cases} 
1 & \text{if heads} \\
0 & \text{if tails} 
\end{cases} \]

The states of \( X \) are its values 0; \( \{0, 1\} \) is its state space. You win $2.00 if you get heads; lose $1.00 if you get tails. How much can you expect to win?
A fair coin is tossed. Suppose

\[ X = \begin{cases} 
1 & \text{if heads} \\
0 & \text{if tails} 
\end{cases} \]

The states of \( X \) are its values 0; \( \{0, 1\} \) is its state space. You win $2.00 if you get heads; lose $1.00 if you get tails. How much can you expect to win?

Let \( g(1) = \) the amount won with heads = $2.00 and \( g(0) = \) the amount won with tails = -$1.00. The expected value of \( g \) is

\[
E[g(X)] = g(1)p(1) + g(0)p(0)
\]

\[
= ($2.00)(\frac{1}{2}) + (-$1.00)(\frac{1}{2}) = $0.50
\]
Special cases

**Mean.** When \( g(x) = x \), the expected value of \( g(x) \) is \( E[X] \), the mean of \( X \). In the continuous case \( E[X] = \int xf(x)dx \).

**Nth moment.** When \( g(x) = x^n \) the expected value of \( g(x) \) is the \( n \)th moment of \( x \) written \( E[X^n] \). In the continuous case it is \( E[X^n] = \int x^n f(x)dx \).

**Moment generating function.** \( M(t) = E[e^{tx}] = \int e^{tx} f(x)dx \). The partial derivatives of \( M(t) \) generate the mean and the other moments. \( \partial M/\partial t |_{t=0} = E[X] \). \( \partial^2 M/\partial t^2 |_{t=0} = E[X^2] \).

**Variance, Standard Deviation.** For a random variable \( X \) with mean \( \mu = E[X] \), the expected value of \( (x - \mu)^2 \) is the variance of \( X \), written \( \sigma^2 \) or \( \sigma^2_X \). In the continuous case,

\[
\sigma^2_X = E[(X - E[X])^2] = \int (x - \mu)^2 f(x)dx
\]

The standard deviation \( \sigma \) is the square root of the variance \( \sigma = \sqrt{\sigma^2} \).
Basic expectation properties

For functions $g(X)$, $h(Y)$ of random variables $X$ and $Y$, and constants $a$ and $b$:

$$E[a] = a, \ E[aX] = aE[X], \ E[X + Y] = E[X] + E[Y].$$

Thus

$$E[ag(X) + bh(Y)] = aE[g(X)] + bE[h(Y)]$$

$$\sigma_{aX}^2 = a^2 \sigma_X^2 \text{ and } \sigma_{aX} = |a| \sigma_X.$$

If $X$ and $Y$ are independent then

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2,$$

$$E[XY] = E[X]E[Y], \ E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$

Visualize independent random variables as perpendicular vectors whose lengths are their variances.

\[ \begin{align*}
\sigma_X^2 &\quad \text{\(\sigma_X^2\)} \\
\sigma_{X+Y}^2 &\quad \text{\(\sigma_{X+Y}^2\)} \\
\sigma_Y^2 &\quad \text{\(\sigma_Y^2\)}
\end{align*} \]
A conditional expectation, is the expected value of a random variable given that some condition is known to be true. Typically the random variable is a function $g(X, Y)$ of two variables and the condition is that the value of one of the random variables is known, i.e., $X = x$ for some $x$.

**Definition.** The conditional expectation of $g(X, Y)$ given that $X = x$ is

$$E[g(X, Y)|X = x] = \int g(x, y)f(y|x)dy$$

where

$$f(y|x) = f(x, y)/f_X(x),$$

$f(x, y)$ is the joint density function and $f_X(x) = \int f(x, y)dy$ is the marginal density function for $x$. 
In a matrix, the average of all matrix entries 
= the average of the column averages.
= the average of the row averages

Here is the continuous analog.

**Theorem.**

\[
E[g(X, Y)] = E[E[g(X, Y)|X = x]]
\]

\[
= E[E[g(X, Y)|Y = y]].
\]

**Proof.**

\[
E\left[ E[g(X, Y) | X = x] \right]
\]

\[
= \int E[g(X, Y)] | X = x \right| f_X(x) \, dx
\]

\[
= \int [ E[g(x, Y)] ] f_X(x) \, dx
\]

\[
= \int \left[ \int g(x, y) f(y|x) \, dy \right] f_X(x) \, dx
\]

\[
= \int \left[ \int g(x, y) f(x, y)/f_X(x) \, dy \right] f_X(x) \, dx
\]

\[
= \int \int g(x, y) f(x, y) \, dy \, dx
\]

\[
= E[g(X, Y)].
\]
**Monotone continuity theorem.**

For any monotone (either increasing or decreasing) sequence of sets (events):

The limit of $A_1 \subseteq A_2 \subseteq A_3 \subseteq ...$ is $\bigcup_{i=1}^{\infty} A_i$ and

$$\lim_{i \to \infty} P[A_i] = P[\bigcup A_i].$$

The limit of $A_1 \supseteq A_2 \supseteq A_3 \supseteq ...$ is $\bigcap_{i=1}^{\infty} A_i$ and

$$\lim_{i \to \infty} P[A_i] = P[\bigcap A_i].$$

- $P(X \leq x)$ is the cumulative distribution of $X$.

The event $X \leq x$ is the intersection of the decreasing sequence $X \leq x + \frac{1}{n}$. Hence $P[X \leq x] = \lim_{n \to \infty} P[X \leq x + \frac{1}{n}]$.

The event $X \leq x$ is the union of the increasing sequence $X \leq x - \frac{1}{n}$. Hence $P[X < x] = \lim_{n \to \infty} P[X \leq x - \frac{1}{n}]$. 
Convergence Theorems

**Definition.** A sequence \( X_1, X_2, X_3, \ldots \) of random variables is *bounded* if there are upper and lower bound random variables \( U \) and \( L \) such that \( U \leq X_i \leq L \) for each \( i \).

**Convergence theorem.** Suppose \( X_1, X_2, X_3, \ldots \) is a sequence of random variables with expectations \( E[X_n] \). Suppose the sequence is either increasing, decreasing or bounded. If \( \lim_{n \to \infty} X_n \) exists, then

\[
E[\lim_{n \to \infty} X_n] = \lim_{n \to \infty} E[X_n].
\]
Strong law of large numbers

"An average of averages is an average". The average of a matrix of numbers is the average of its row averages.

Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of independent, identically distributed random variables with common mean $\mu$. What is the expected value of the average of the first $n$ values?

$$E[(X_1 + X_2 + \ldots + X_n)/n]$$

$$= (E[X_1] + E[X_2] + \ldots + E[X_n])/n$$

$$= (\mu + \mu + \ldots + \mu)/n = (n\mu)/n = \mu$$

What is the expected average of the infinitely many values? To make this precise, we take the limit.

$$E[\lim_{n \to \infty} (X_1 + X_2 + \ldots + X_n)/n]$$

$$= \lim_{n \to \infty} E[(X_1 + X_2 + \ldots + X_n)/n]$$

$$= \lim_{n \to \infty} \mu = \mu.$$ 

Thus the expected average of infinitely many random variables with common mean $\mu$ is just $\mu$.

The Strong Law says much more. Not only is this the average, it is (with probability 1) the only possible value!
Notation: Suppose $X$ is a random variable. Thus the value of $X$ depends on the outcome of $\omega$ some random experiment. Let $X(\omega)$ be the value of $X$ on outcome $\omega$.

Suppose the experiment is tossing a coin infinitely many times. Suppose $X_i = 1$ if the $i$th toss is H (heads) and $X_i = 0$ if it is T (tails). An outcome $\omega$ of this experiment is an infinite sequence of heads and tails.

Suppose $\omega = \langle H, H, T, H, T, T, .... \rangle$. Then 

$\langle X_1(\omega), X_2(\omega), X_3(\omega), X_4(\omega), ... \rangle = \langle 1, 1, 0, 1, 0, 0, ... \rangle$.

**Strong law of large numbers.** For almost all outcomes $\omega$:

$$\lim_{n \to \infty} \sum_{k=1}^{n} X_k(\omega)/n = \mu$$

“For almost all” means that the set of outcomes $\omega$ for which this is false has probability 0.
Let’s begin by considering the experiment which randomly chooses an infinite decimal \( r = .d_1d_2d_3d_4 \ldots \) by first choosing the first digit \( d_1 \in \{0,1,2,3,4,5,6,7,8,9\} \), then the second \( d_2 \), the third, \( \ldots \). Suppose the digits are all equally likely.

Given an outcome \( r \) of this experiment, let \( D_i = \) the \( i^{\text{th}} \) decimal place of \( r \). Thus if \( r = .d_1d_2d_3d_4 \ldots \), then \( D_1 = d_1 \), \( D_2 = d_2 \), \( \ldots \). The \( D_1, D_2, D_3, \ldots \) are independent identically distributed random variables.

What is the average value of a randomly chosen digit?

What is the mean \( \mu = E[D_i] \) of the \( i^{\text{th}} \) digit?

What is the average of the first \( n \) digits?

\[
E[(D_1 + D_2 + \ldots + D_n)/n] = ?.
\]

To find the average of all infinitely many digits in \( r \), we take the limit \( \lim_{n \to \infty} (D_1 + D_2 + \ldots + D_n)/n \).

Since the sequence is bounded we can calculate the expected value of the average of the digits as above with the \( X_i \)’s:

\[
E[\lim_{n \to \infty} \sum_{i=1}^{n} D_i/n] = \lim_{n \to \infty} E[\sum_{i=1}^{n} D_i/n] = \lim_{n \to \infty} \mu = \mu.
\]

Thus the expected average of the digits is 9/2 as claimed by the Strong Law of Large Numbers.
Not every $r$ has this average. For example, if $r = .020202020202...$

the average of the first two digits is 1 which is $< 9/2$.

What is the average of the digits of $r$?

What is average for all the digits of $r = .8989898989...$ ?

It is reasonable to expect that the average of the digits for a randomly chosen $r$ is roughly a bell curve with minimum value 0 for $r = .00000...$ and maximal value 9 for $r = .999999...$ and with the center of the distribution being $9/2$.

The Strong Law says no. The distribution has a single point mass $9/2$ which has probability 1 and all else has probability 0.