UNCERTAINTY—THE DESCRIPTION OF RANDOM EVENTS

A. Some introductory definitions

1. Event/realization: the rolling of a pair of dice, the taking of a measurement, the performing of an experiment.
2. Outcome: the result of rolling the dice, taking the measurement, etc.
3. Deterministic event: an event whose outcome can be predicted realization after realization, e.g., the measured length of a table to the nearest cm.
4. Random event/process/variable: an event/process that is not and cannot be made exact and, consequently, whose outcome cannot be predicted, e.g., the sum of the numbers on two rolled dice.
5. Probability: an estimate of the likelihood that a random event will produce a certain outcome.

B. What's deterministic and what's random depends on the degree to which you take into account all the relevant parameters.

1. Mostly deterministic—only a small fraction of an outcome cannot be accounted for
   a. The measurement of the length of a table—temperature/humidity variation, measurement resolution, instrument/observer error; and, at the quantum level, intrinsic uncertainty.
   b. The measurement of a volume of water—evaporation, meniscus, temperature dependency.

2. Half and half—a significant fraction of an outcome cannot be accounted for
   a. The measurement of tidal height in an ocean basin—measurement is contaminated by effects from wind and boats.

3. Mostly random—most of an outcome cannot be accounted for
   a. The sum of the numbers on two rolled dice.
   b. The trajectory of a given molecule in a solution
   c. The stock market
   d. Fluctuations in water pressure in a municipal water supply

C. The "sample space"—the set of all possible outcomes

1. A rolled die: 1, 2, 3, 4, 5, 6	(discrete outcome)
2. A flipped coin: H, T	(discrete outcome)
3. The measured width of a room: ?	(continuous outcome)
4. The measured period of a pendulum: ?	(continuous outcome)

D. Statistical regularity—if a random event is repeated many times, it will produce a distribution of outcomes. This distribution will approach an asymptotic form as the number of events increases.

1. If the distribution is represented as the number of occurrences of each outcome, the distribution is called the **frequency distribution function**.

2. If the distribution is represented as the percentage of occurrences of each outcome, the distribution is called the **probability distribution function.**

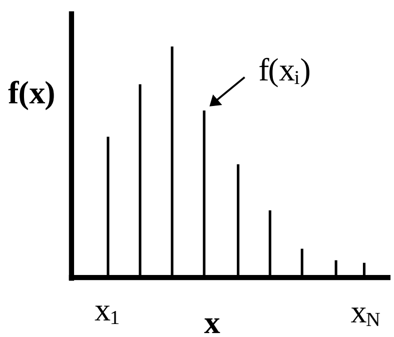
3. We attribute probability to the asymptotic distribution of outcome occurrences.
   a) A "fair" die will produce each of its six outcomes with equal likelihood; therefore we say that the probability of getting a "1" is 1/6.

E. Description of a random variable X

1. Distribution of discrete outcomes

   a) $\Pr(X = x_i) = f(x_i) \Rightarrow$ the probability that the variable X takes the value x_i is a function of the value x_i.

   b) $f(x_i)$ is called the probability distribution function



Probability distribution function

   c) Properties of discrete probabilities

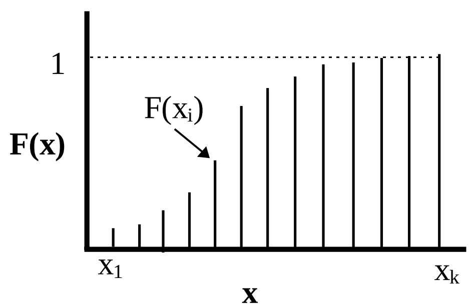
      i) $\Pr(X = x_i) = f(x_i) \geq 0$ for all i

      ii) $\sum_{i=1}^{k} \Pr(X = x_i) = \sum_{i=1}^{k} f(x_i) = 1$ for k possible discrete outcomes

      iii) $\Pr(a < X \leq b) = F(b) - F(a) = \sum_{a < x_i \leq b} f(x_i)$

d) Cumulative discrete probability distribution function

$\Pr(X \le x') = F(x') = \sum_{i=1}^{j} f(x_i)$, where x_j is the largest discrete value of X less than or equal to x'. $\Pr(X \le x_k) = 1$.



Cumulative distribution function

e) Examples of discrete probability distribution functions

i) Distribution function for throwing a die:

$f(x_i) = 1/6 \qquad$ for i = 1, 6

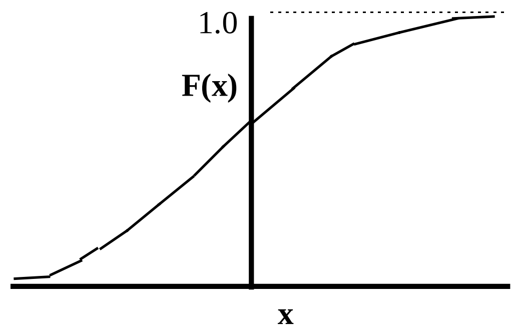
ii) Distribution function for the sum of two thrown dice

$f(x_i) =$	1/36	for $x_1 = 2$
	2/36	for $x_2 = 3$
	3/36	for $x_3 = 4$
	4/36	for $x_3 = 5$
	5/36	for $x_5 = 6$
	6/36	for $x_5 = 7$
	5/36	for $x_7 = 8$
	4/36	for $x_8 = 9$
	3/36	for $x_9 = 10$
	2/36	for $x_{10} = 11$
	1/36	for $x_{11} = 12$

2. Distribution of continuous outcomes

a) Cumulative distribution function

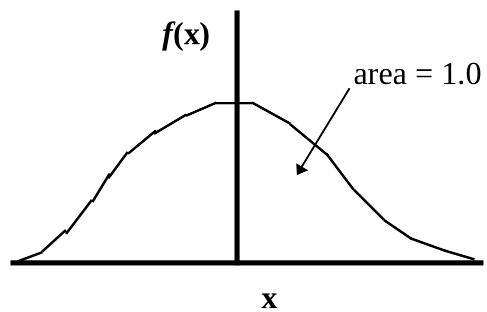
$$\Pr(X \le x) = F(x) = \int_{-\infty}^{x} f(x)\,dx$$



Cumulative distribution function

b) Probability density (distribution) function (p.d.f.)

$f(x) = dF(x)/dx$



Probability density function

c) Properties of F(x) and f(x)

i) $F(-\infty) = 0$, $0 \le F(x) \le 1$, $F(\infty) = 1$

ii) $\Pr(a < X \le b) = F(b) - F(a) = \int_a^b f(x)dx$

d) Examples of continuous p.d.f.s

i) "top hat" or uniform distribution:
$f(x) = 0$ for $-a < x < a$
$f(x) = 1/(2a)$ for $|x| \le a$

ii) Gaussian distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$, where μ and σ are given constants

iii) Poisson, Binomial are other important named distributions

3. Another way of representing a distribution function: moments

   a) the r^th^ moment about the origin: $\nu_r = \sum_{i=1}^{k} x_i^{\,r} f(x_i)$

      i) the 1^st^ moment about the origin is the mean $\mu = \nu_1 = \sum_{i=1}^{k} x_i f(x_i)$

   b) the r^th^ moment about the mean $\mu_r = \sum_{i=1}^{k} \left(x_i - \mu\right)^r f(x_i)$

      i) the 2^nd^ moment about the mean μ_2 is the variance $\sigma^2 = \sum_{i=1}^{k} \left(x_i - \mu\right)^2 f(x_i)$

   c) for many distribution functions, knowing all the moments of f(x) is equivalent to knowing f(x) itself.

4. Important moments

   a) the mean μ, the "center of gravity"
   b) the variance σ^2: a measure of spread.
      i) the standard deviation: $\sigma = \sqrt{\sigma^2}$
   c) the skewness $\dfrac{\mu^3}{\left[\sigma^2\right]^{\frac{3}{2}}}$: a measure of asymmetry
   d) the kurtosis $\dfrac{\mu^4}{\left(\sigma^2\right)^2}$: a measure of "peakedness"

F. Estimation of random variables (RVs)

1. Assumptions/procedures
   a) There exists a stable underlying p.d.f. for the RV
   b) Investigating the characteristics of the RV consists of obtaining sample outcomes and making inferences about the underlying distribution.

2. Sample statistics on a random variable X (or from a large sample "population")
   a) The i^{th} sample outcome of the variable X is denoted X_i .

   b) The sample mean: $\overline{X} = \dfrac{1}{N}\sum_{i=1}^{N} X_i$, where *N* is the sample size.

   c) The sample variance : $s^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2$

   d) $\overline{X}$ and s^2 are only estimates of the mean μ and variance σ^2 of the underlying p.d.f. ($\overline{X}$ and s^2 are estimates for the sample. μ and σ^2 characteristics of the population from which the sample was taken.)

e) $\overline{X}$ and s^2 are themselves random variables. As such, they have their own means and variances (which can be calculated).

3. Expected value
   a) The expected value of a random variable X is written E(X). It is the value that one would obtain if a very large number of samples were averaged together.
   b) As defined above:
      i) $E[\overline{X}] = \mu$, i.e., the expected value of the sample mean is the population mean.
      ii) $E[s^2] = \sigma^2$, i.e., the expected value of the sample variance is the population variance.
   c) Expectation allows us to use sample statistics to infer population statistics.
   d) Properties of expectation:
      i) E[aX + bY] = aE[X] + bE[Y], where a, b are constants
      ii) If Z = g(X). then E[Z] = E[g(X)] = $\sum_{\substack{all\,values \\ x\,of\,X}} g(x)\Pr(X = x)$

         Example: Throw a die. If the die shows a "6" you win $5; else, you lose a $1. What's the expected value Z of this "game"?

         Pr(X = 1) = 1/6 g(1) = -1
         Pr(X = 2) = 1/6 g(2) = -1
         Pr(X = 3) = 1/6 g(3) = -1
         Pr(X = 4) = 1/6 g(4) = -1
         Pr(X = 5) = 1/6 g(5) = -1
         Pr(X = 6) = 1/6 g(6) = 5

         $\therefore$ E[Z] = (-1) * 5 * 1/6 + 5 * 1/6 = 0 , i.e., you would expect to neither win nor lose

      iii) E[XY] = E[X] E[Y] provided X and Y are "independent", i.e., samples of X cannot be used to predict anything out sample of Y (and vice versa).

         Example: You have error-prone measurements for the height X and width Y of a picture. What is the expected value of the area of the picture XY? Answer: E[X] E[Y].

G. Selected engineering uses of statistics

1. Measurement and errors
   a) "Best" estimate of value based on error-prone measurements, e.g., three different measurements give three different answers
   b) Compound measurements each of which is error-prone, e.g., the volume of a box whose sides are measured with error
   c) Standard error (standard deviation) of "best" estimate, i.e., error bars
   d) Techniques to reduce the standard error

i) repeated measurements
ii) different measurement strategy

2. Characterizing random populations
    a) Distribution of traffic accidents at an intersection
    b) Quantity of lumber in a forest
    c) Distribution of “seconds” on an assembly line
    d) Voltage fluctuation characteristics on a transmission line