9.4 Random Numbers from Various Distributions

Distributions

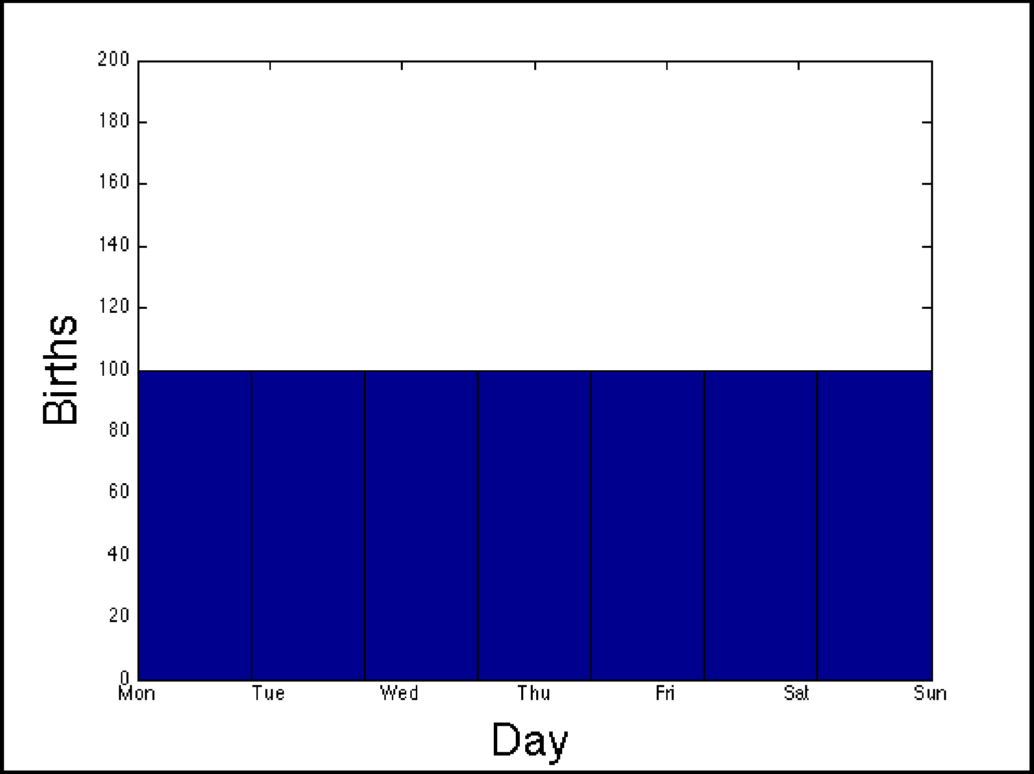
- A distribution of numbers is a description of the portion of times each possible outcome or range of outcomes occurs on average.

- A histogram is a display of a given distribution.

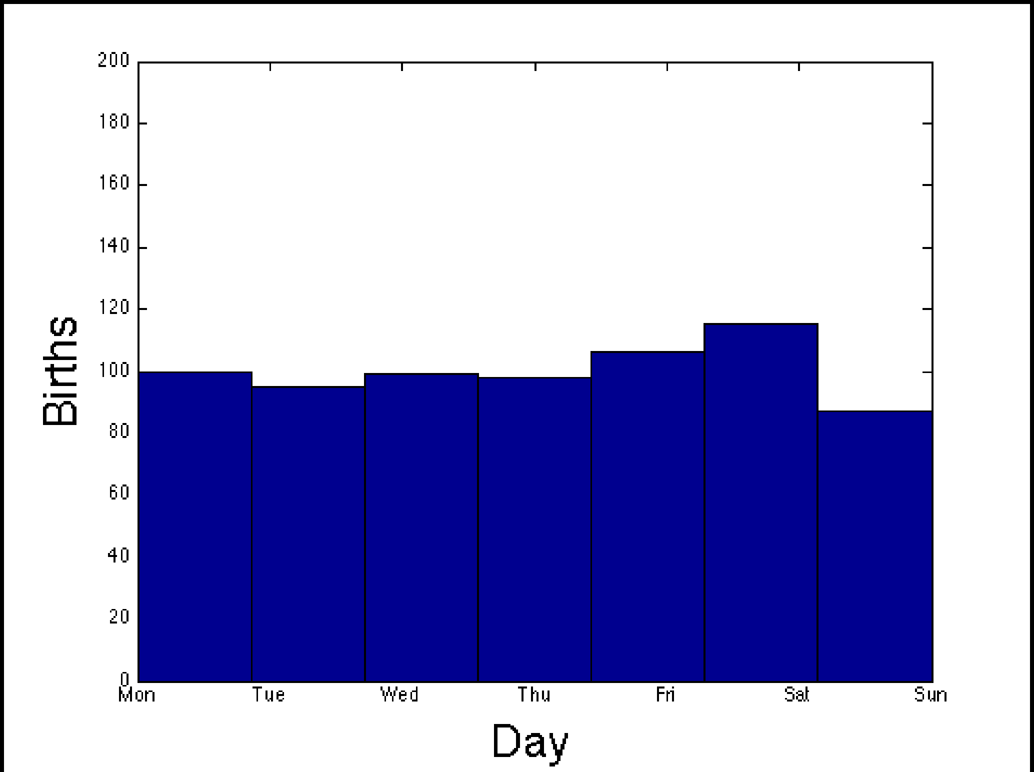
- In a uniform distribution all outcomes are equally likely.

Uniform Distribution: Hypothetical

Consider a hospital at which there are an average of 100 births per day:



Uniform Distribution: Random-Number Simulation

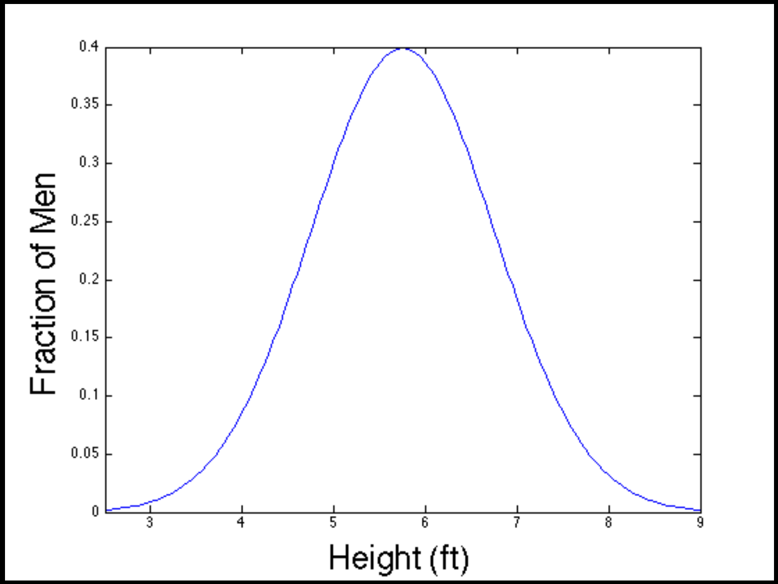


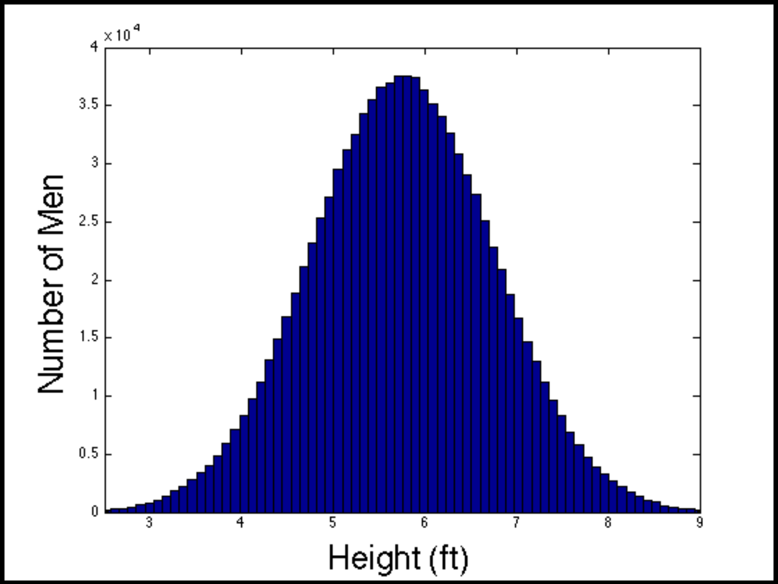
Discrete vs. Continuous Distributions

- A discrete distribution is one in which the values (*x*, *y* axis values) are discrete (countable, finite in number).
-
- A continuous distribution is one in which the values (*x*, *y* axis values) are continuous (not countable).

- In practice, we can model continuous distributions discretely by binning….

Binning


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
3
4
5
6
7
8
9
Fraction of Men
Height (ft)


× 10^4
4
3.5
3
2.5
2
1.5
1
0.5
0
3
4
5
6
7
8
9
Number of Men
Height (ft)

Probability Density Function

- For a discrete distribution, a probability density function (or density function or probability function) tells us the probability of occurrence of its input.

- For a continuous distribution, the PDF indicates the probability that a given outcome falls inside a specific range of values.

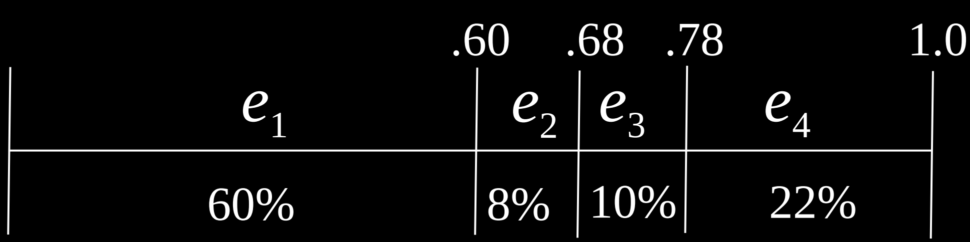
Probability Density Function

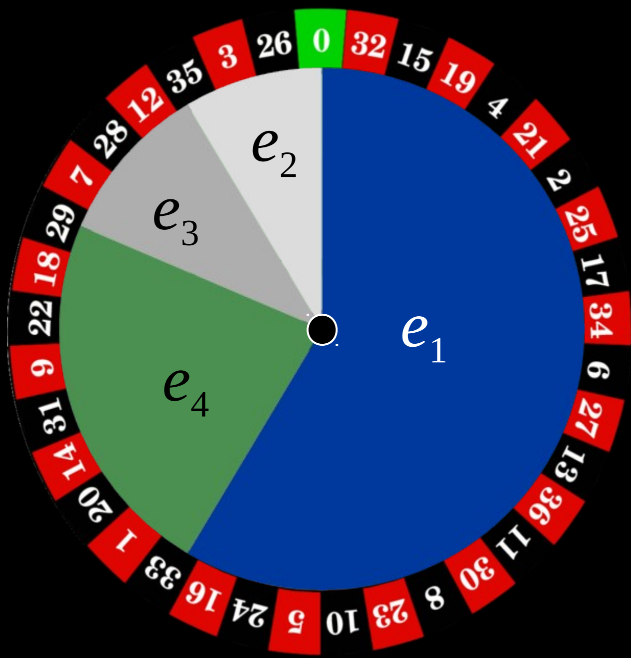
- For a discrete distribution, we just report the value at that bin.

- For a continuous distribution, we integrate between the ends of the interval (Fundamental Theorem of Calculus), or approximate that using numerical methods (Euler, RK4)

Generating Random Numbers in Non-Uniform Distributions

Imagine a biased roulette wheel for picking events (outcomes) e_1, e_2, ….





Generating Random Numbers in Non-Uniform Distributions

Given probabilities $p_1, p_2, \ldots.$ for events $e_1, e_2, \ldots.$:

Generate *rand*, a uniform random floating-point number in [0,1); that is, from zero up to but excluding 1.

If $rand < p_1$ then use e_1

Else if $rand < p_1+p_2$ then use e_2

…

Else if $rand < p_1+p_2+\ldots+p_{n\text{-}1}+$then use $e_{n\text{-}1}$

Else use e_n

Normal (Gaussian) Distributions

Carl Friedrich Gauss
(1777-1855)

Gauss
Gerling
Plücker
Klein
Story
Lefschetz
Tucker
Minsky
Winston
Waltz
Pollack
Levy
You

Normal (Gaussian) Distributions

- The standard deviation σ of a set of values is their average difference from their mean μ.
- In a normal distribution (so-called because it is so common) 68.3% of the values are within $\pm\sigma$ (one standard deviation) of μ; 95.5% are within $\pm 2\sigma$; and 99.7% are within $\pm 3\sigma$. *I.e.*, *extreme values are rare.*
- Where do these strange percentages come from?

Normal (Gaussian) Distributions

- Normal distribution has probability density function

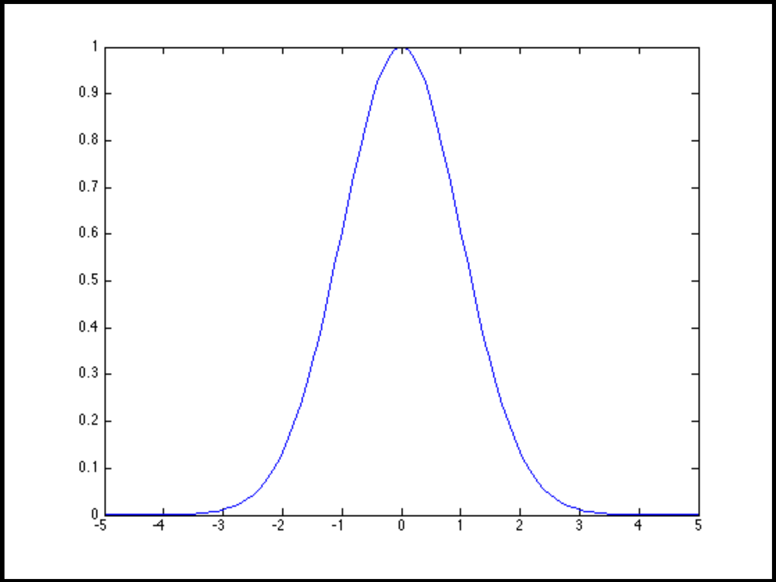
$$\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$$

- Random number generators typically use $\mu = 0$, $\sigma = 1$, so this simplifies to

$$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

Normal (Gaussian) Distributions

- is a constant, so the shape is given by ; *i.e.*, something that reaches a peak at $x = 0$ and tapers off rapidly as x grows positive or negative:



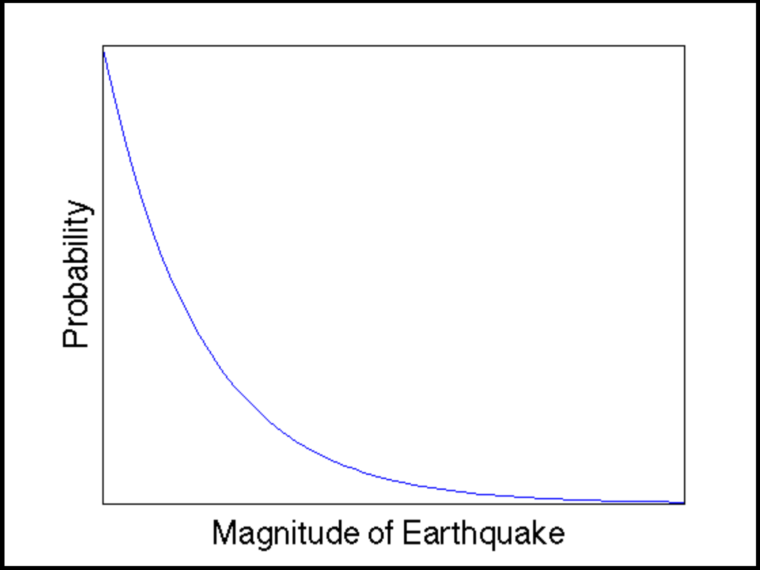
- How can we build such a distribution from our uniformly-distributed random numbers?

Box-Muller-Gauss Method for Normal Distributions with Mean μ And Standard Deviation σ

- Start with two uniform random numbers:
  - a in [0,2π)
  - $rand$ in [0,1)
- Then compute $b = \sigma \sqrt{-\ln(rand)}$
- Obtain two normally distributed numbers
  - $b \sin(a) + \mu$
  - $b \cos(a) + \mu$

Exponential Distributions

- Common pattern is exponentially decaying PDF, also called $1/f$ noise (a.k.a. pink noise)
  - *noise* = random
  - f = *frequency*; *i.e.*, larger events are less common
  - *pink* because uniform distribution is "white" (white light = all frequencies)
- "Universality" is a current topic of controversy (Shalizi 2006)



Exponential Method for PDF $|r|e^{rt}$ where $t>0$, $r<0$

- Start with uniform random $rand$ in [0,1)
- Compute $\ln(rand)/r$
- *E.g.*, $\ln(rand)$ / (-2) gives $1/f$ noise

Rejection Method

- To get random numbers in interval $[a, b)$ for distribution $f(x)$:
  - Generate *randInterval*, a uniform random number in $[a, b)$
  - Generate *randUpperBound*, a uniform random number in [0, upper bound for f)
  - If $f(randInterval) > randUpperBound$ then use *randInterval*
- *E.g.* for normal distribution with μ = 0, σ = 1,

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

upperBound = 1

randInterval = approx. [-3,3)