MIT 2.853/2.854

Introduction to Manufacturing Systems

Probability

Stanley B. Gershwin
Laboratory for Manufacturing and Productivity
Massachusetts Institute of Technology
gershwin@mit.edu

I flip a coin 100 times, and it shows heads every time.

Question: What is the probability that it will show heads on the next flip?

Another Trick Question

I flip a coin 100 times, and it shows heads every time.

Question: *How much would you bet* that it will show heads on the next flip?

Probability and Statistics

Still Another Trick Question

I flip a coin 100 times, and it shows heads every time.

Question: *What odds would you demand before you bet* that it will show heads on the next flip?

Probability and Statistics

Probability $\neq$ Statistics

Probability: mathematical theory that describes uncertainty.

Statistics: set of techniques for extracting useful information from data.

Interpretations of probability

Frequency

The probability that the outcome of an experiment is A is P...

if the experiment is performed a large number of times and the fraction of times that the observed outcome is A is P.

Interpretations of probability

State of belief

The probability that the outcome of an experiment is A is P...

if that is the opinion (ie, belief or state of mind) of an observer *before* the experiment is performed.

Interpretations of probability

Example of State of Belief: Betting odds

The probability that the outcome of an experiment is A is P...

if *before the experiment is performed* a risk-neutral observer would be willing to bet \$1 against more than $\$\frac{1-P}{P}$.

The expected value (slide **??**) of the bet is greater than

$$(1 - P) \times (-1) + (P) \times \left(\frac{1 - P}{P}\right) = 0$$

Interpretations of probability

Abstract measure

The probability that the outcome of an experiment is A is $P(A)$...

if $P()$ satisfies a certain set of conditions: *the axioms of probability.*

Interpretations of probability

Axioms of probability

Let U be a set of *samples* . Let $\mathcal{E}_1$, $\mathcal{E}_2$, ... be subsets of U.

Let $\emptyset$ be the *null* (or *empty*) *set* , the set that has no elements.

- $0 \leq P(\mathcal{E}_i) \leq 1$
- $P(U) = 1$
- $P(\emptyset) = 0$
- If $\mathcal{E}_i \cap \mathcal{E}_j = \emptyset$, then $P(\mathcal{E}_i \cup \mathcal{E}_j) = P(\mathcal{E}_i) + P(\mathcal{E}_j)$

Probability Basics
Discrete Sample Space

Notation, terminology:

- ω is often used as the symbol for a generic sample.
- Subsets of U are called *events.*
- $P(\mathcal{E})$ is the *probability* of $\mathcal{E}$.

Probability Basics

Discrete Sample Space

- *Example:* Throw a single die. The possible outcomes are $\{1, 2, 3, 4, 5, 6\}$. ω can be any one of those values.

- *Example:* Consider $n(t)$, the number of parts in inventory at time t. Then

$$\omega = \{n(1), n(2), ..., n(t), ...\}$$

is a *sample path.*

Probability Basics

Discrete Sample Space

- An event can often be defined by a statement. For example,

$$\mathcal{E} = \{\text{There are 6 parts in the buffer at time } t = 12\}$$

Formally, this can be written

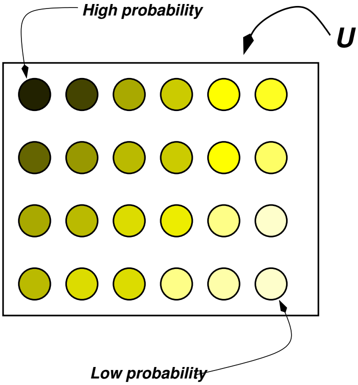
$$\mathcal{E} = \text{the set of all } \omega \text{ such that } n(12) = 6$$

or,

$$\mathcal{E} = \{\omega | n(12) = 6\}$$

Probability Basics

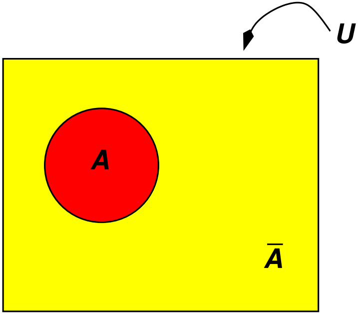
Discrete Sample Space



Probability Basics

Set Theory

Venn diagrams

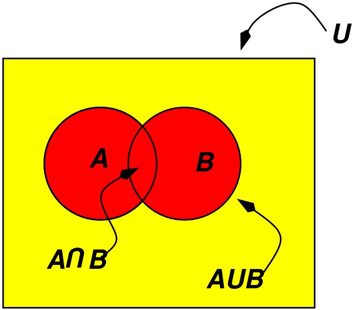


$$P(\bar{A}) = 1 - P(A)$$

Probability Basics

Set Theory

Venn diagrams



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Probability Basics

Independence

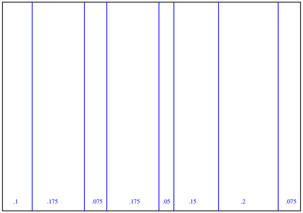
A and B are *independent* if

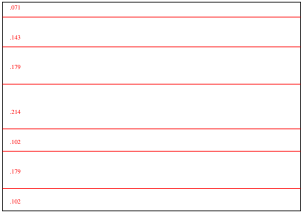
$$P(A \cap B) = P(A)P(B).$$

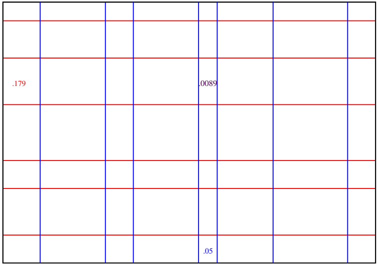
grid figure to illustrate independence

Probability Basics

Independence





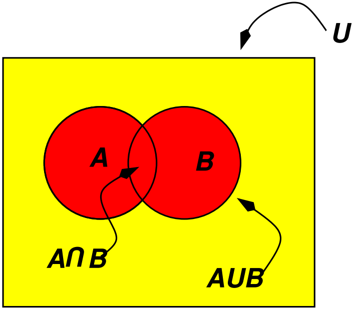


Probability Basics

Conditional Probability

If $P(B) \neq 0$,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



We can also write $P(A \cap B) = P(A|B)P(B)$.

Probability Basics

Conditional Probability

$$P(A|B) = P(A \cap B)/P(B)$$

Example: Throw a die.Let

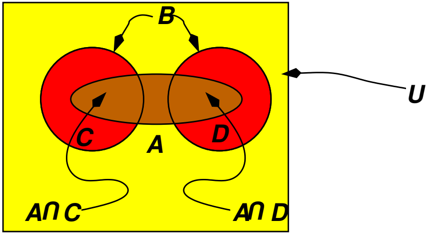
- A is the event of getting an odd number (1, 3, 5).
- B is the event of getting a number less than or equal to 3 (1, 2, 3).

Then $P(A) = P(B) = 1/2, P(A \cap B) = P(1, 3) = 1/3.$

Also, $P(A|B) = P(A \cap B)/P(B) = 2/3.$

Probability Basics

Law of Total Probability



- Let $B = C \cup D$ and assume $C \cap D = \emptyset$. Then

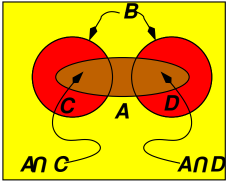
  $P(A|C) = \dfrac{P(A \cap C)}{P(C)}$ and $P(A|D) = \dfrac{P(A \cap D)}{P(D)}$.

Probability Basics

Also,

- $P(C|B) = \dfrac{P(C \cap B)}{P(B)} = \dfrac{P(C)}{P(B)}$ because $C \cap B = C$.

  Similarly, $P(D|B) = \dfrac{P(D)}{P(B)}$



- $A \cap B = A \cap (C \cup D) = (A \cap C) \cup (A \cap D)$

- Therefore

  $P(A \cap B) = P(A \cap (C \cup D))$

  $\qquad = P(A \cap C) + P(A \cap D)$ because $(A \cap C)$ and $(A \cap D)$ are disjoint.

Probability Basics

Law of Total Probability

- Or, from the definition of conditional probability,

  $P(A|B)P(B) = P(A|C)P(C) + P(A|D)P(D)$

  or,

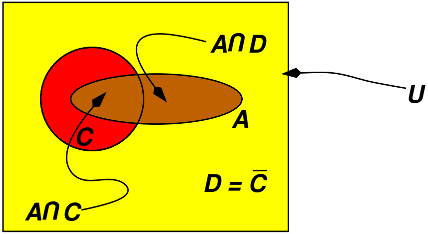
  $$\frac{P(A|B)P(B)}{P(B)} = \frac{P(A|C)P(C)}{P(B)} + \frac{P(A|D)P(D)}{P(B)}$$

  or,

  $$P(A|B) = P(A|C)P(C|B) + P(A|D)P(D|B)$$

Probability Basics

Law of Total Probability

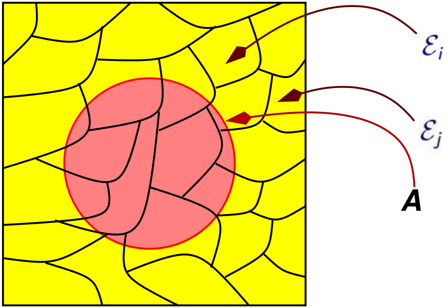


An important case is when $C \cup D = B = U$, so that $A \cap B = A$. Then $P(A) = P(A \cap C) + P(A \cap D)$ or

$$\boxed{P(A) = P(A|C)P(C) + P(A|D)P(D)}$$

Probability Basics

Law of Total Probability



More generally, if A and $\mathcal{E}_1, \ldots \mathcal{E}_k$ are events and

$$\mathcal{E}_i \text{ and } \mathcal{E}_j = \emptyset, \text{ for all } i \neq j$$

and

$$\bigcup_j \mathcal{E}_j = \text{ the universal set}$$

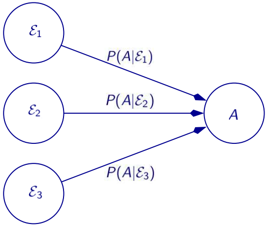
(ie, the set of $\mathcal{E}_j$ sets is *mutually exclusive* and *collectively exhaustive*) then ...

Probability Basics
Law of Total Probability

$$\sum_j P(\mathcal{E}_j) = 1$$

and

$$\boxed{P(A) = \textstyle\sum_j P(A|\mathcal{E}_j) P(\mathcal{E}_j).}$$



Probability Basics

Law of Total Probability

Example

$A = \{$I will have a cold tomorrow.$\}$
$\mathcal{E}_1 = \{$It is raining today.$\}$
$\mathcal{E}_2 = \{$It is snowing today.$\}$
$\mathcal{E}_3 = \{$It is sunny today.$\}$

(Assume $\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3 = U$ and $\mathcal{E}_1 \cap \mathcal{E}_2 = \mathcal{E}_1 \cap \mathcal{E}_3 = \mathcal{E}_2 \cap \mathcal{E}_3 = \emptyset$.)

Then $A \cap \mathcal{E}_1 = \{$I will have a cold tomorrow *and* it is raining today$\}$.
And $P(A|\mathcal{E}_1)$ is the probability I will have a cold tomorrow *given* that it is raining today.

etc.

Probability Basics

Law of Total Probability

Then

{I will have a cold tomorrow.}=
{I will have a cold tomorrow *and* it is raining today} $\cup$
{I will have a cold tomorrow *and* it is snowing today} $\cup$
{I will have a cold tomorrow *and* it is sunny today}

so

P({I will have a cold tomorrow.})=
P({I will have a cold tomorrow *and* it is raining today}) $+$
P({I will have a cold tomorrow *and* it is snowing today}) $+$
P({I will have a cold tomorrow *and* it is sunny today})

Probability Basics

Law of Total Probability

$P(\{\text{I will have a cold tomorrow.}\})=$

$P(\{\text{I will have a cold tomorrow} \mid \text{it is raining today}\})P(\{\text{it is raining today}\}) +$

$P(\{\text{I will have a cold tomorrow} \mid \text{it is snowing today}\})P(\{\text{it is snowing today}\}) +$

$P(\{\text{I will have a cold tomorrow} \mid \text{it is sunny today}\})\ P(\{\text{it is sunny today}\})$

or

$$P(A) = P(A|\mathcal{E}_1)P(\mathcal{E}_1) + P(A|\mathcal{E}_2)P(\mathcal{E}_2) + P(A|\mathcal{E}_3)P(\mathcal{E}_3)$$

Probability Basics

Random Variables

Let V be a vector space. Then a *random variable* X is a mapping (a function) from U to V.

If $\omega \in U$ and $x = X(\omega) \in V$, then X is a random variable.

Example: V could be the real number line.

Typical notation :

- Upper case letters (X) are usually used for random variables and corresponding lower case letters (x) are usually used for possible values of random variables.
- Random variables ($X(\omega)$) are usually not written as functions; the argument (ω) of the random variable is usually not written. *This sometimes causes confusion.*

Probability Basics

Random Variables

Flip of a Coin

Let U={H,T}. Let $\omega =$ H if we flip a coin and get heads; $\omega =$ T if we flip a coin and get tails.

Let V be the real number line.

$X(T) = 0$
$X(H) = 1$

Assume the coin is fair. *(No tricks this time!)* Then
$P(\omega = \text{ T }) = P(X = 0) = 1/2$
$P(\omega = \text{ H }) = P(X = 1) = 1/2$

Flip of Three Coins

Let $U = \{$HHH, HHT, HTH, HTT, THH, THT, TTH, TTT$\}$.

Let $\omega =$ HHH if we flip 3 coins and get 3 heads; $\omega =$HHT if we flip 3 coins and get 2 heads and *then* one tail, etc. *The order matters!* There are 8 samples.

- $P(\omega) = 1/8$ for all ω.

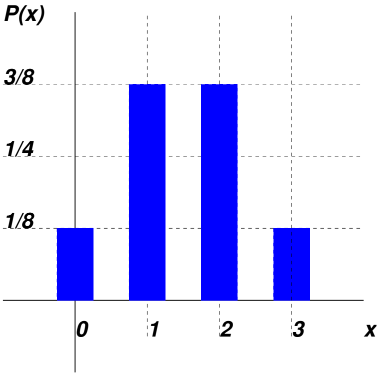
Let X be the *number* of heads. Then $X = 0$, 1, 2, or 3.

- $P(X = 0)$=1/8; $P(X = 1)$=3/8; $P(X = 2)$=3/8; $P(X = 3)$=1/8.

There are 4 distinct values of X.

Probability Basics
Probability Distributions

Let $X(\omega)$ be a random variable. Then $P(X(\omega) = x)$ is the *probability distribution* of X (usually written $P(x)$). For three coin flips:



Probability Basics
Probability Distributions

Shorthand:

- Instead of writing $P(X(\omega) = x)$, people often write $P(x)$ if the meaning is unambiguous.

Mean and Variance:

- *Mean (average):* $\bar{x} = \mu_x = E(X) = \sum_x xP(x)$
- *Variance:* $V_x = \sigma_x^2 = E(X - \mu_x)^2 = \sum_x (x - \mu_x)^2 P(x)$
- *Standard deviation (sd):* $\sigma_x = \sqrt{V_x}$
- *Coefficient of variation (cv):* σ_x / μ_x

Probability Basics

Probability Distributions

For three coin flips:

$$\bar{x} = 1.5$$

$$V_x = 0.75$$

$$\sigma_x = 0.866$$

$$\text{cv} = 0.577$$

Probability Basics

Functions of a Random Variable

- A function of a random variable is a random variable.

- *Special case: linear function*

  For every ω, let $Y(\omega) = aX(\omega) + b$. Then

  - $\bar{Y} = a\bar{X} + b$.
  - $V_Y = a^2 V_X$; $\qquad \sigma_Y = |a|\sigma_X$.

Discrete Random Variables

1. Bernoulli

Flip a biased coin.

X^B is 1 if outcome is heads; 0 if tails.

Let p be a real number, $0 \le p \le 1$.

$P(X^B = 1) = p$.

$P(X^B = 0) = 1 - p$.

X^B is a *Bernoulli random variable.*

Discrete Random Variables

2. Binomial

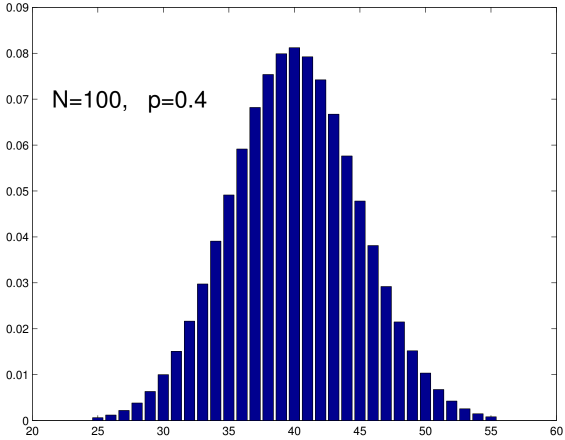
The sum of N independent Bernoulli random variables X_i^B with the same parameter p is a *binomial* random variable X^b.

$$X^b = \sum_{i=0}^{N} X_i^B$$

$$P(X^b = x) = \frac{N!}{x!(N-x)!} p^x (1-p)^{(N-x)}$$

Discrete Random Variables

2. Binomial probability distribution



$P(n)$

n

Discrete Random Variables

3. Geometric

The number of independent Bernoulli random variables X_i^B with the same parameter p tested *until the first 1 appears* is a *geometrically distributed* random variable X^g.

$$\begin{array}{cccccccccc} 1 & 2 & 3 & 4 & \ldots & k-4 & k-3 & k-2 & k-1 & k \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \mathbf{1} \end{array}$$

$$\longleftarrow \quad k \quad \longrightarrow$$

$$\boxed{X^g = k \text{ if } X_1^B = 0,\ X_2^B = 0,\ \ldots,\ X_{k-1}^B = 0,\ X_k^B = 1}$$

Discrete Random Variables

3. Geometric

To calculate $P(X^g = k)$, observe that $P(X^g = 1) = p$, so $P(X^g > 1) = 1 - p$. Also, observe that $\{X^g > k\}$ is a subset of $\{X^g > k - 1\}$.

Then

$$\begin{aligned} P(X^g > k) &= P(X^g > k | X^g > k-1) P(X^g > k-1) \\ &= (1-p) P(X^g > k-1), \end{aligned}$$

because

$$P(X^g > k | X^g > k-1) = P(X_1^B = 0, \ldots, X_k^B = 0 \,| X_1^B = 0, \ldots, X_{k-1}^B = 0)$$
$$= 1 - p$$

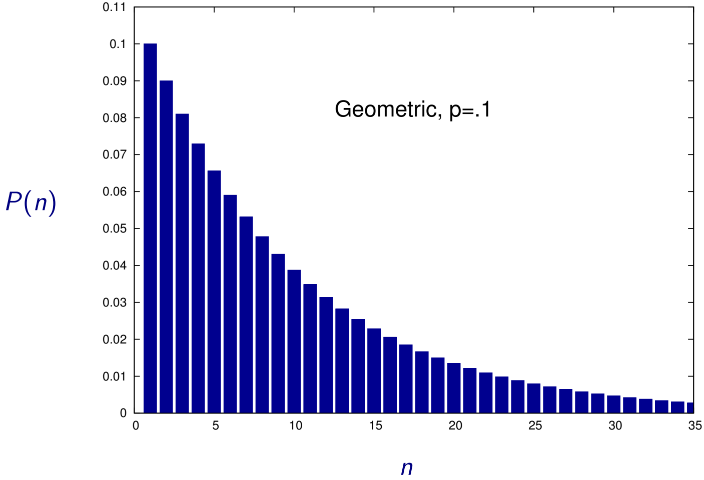
so

$$P(X^g > 1) = 1 - p, \; P(X^g > 2) = (1-p)^2, \; \ldots \; P(X^g > k-1) = (1-p)^{k-1}$$

and $P(X^g = k) = P(\{X^g > k-1\} \text{ and } \{X_k^B = 1\}) = (1-p)^{k-1} p$.

Discrete Random Variables

3. Geometric probability distribution



Discrete Random Variables

4. Poisson Distribution

$$P(X^P = n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

Discussion later.

Continuous Random Variables
Philosophical Issues

1. *Mathematically* , continuous and discrete random variables are very different.
2. *Quantitatively* , however, some continuous models are very close to some discrete models.
3. Therefore, which kind of model to use for a given system is a matter of *convenience* .

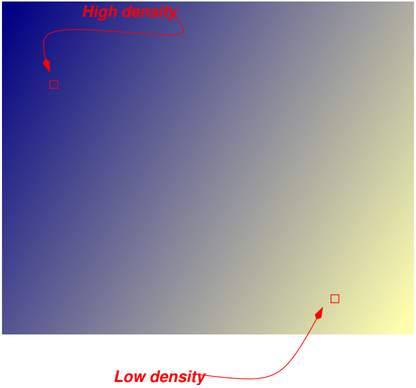
Continuous Random Variables

Philosophical Issues

Example: The production process for small metal parts (nuts, bolts, washers, etc.) might better be modeled as a continuous flow than as a large number of discrete parts.

Continuous Random Variables

Philosophical Issues

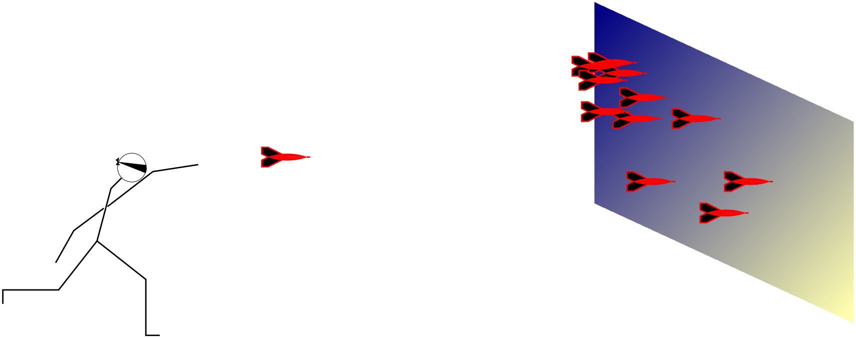


The probability of a two-dimensional random variable being in a small square is the *probability density* times the area of the square. (The definition is similar in higher-dimensional spaces.)

Compare with slide **??**.

Continuous Random Variables

Philosophical Issues



Continuous Random Variables

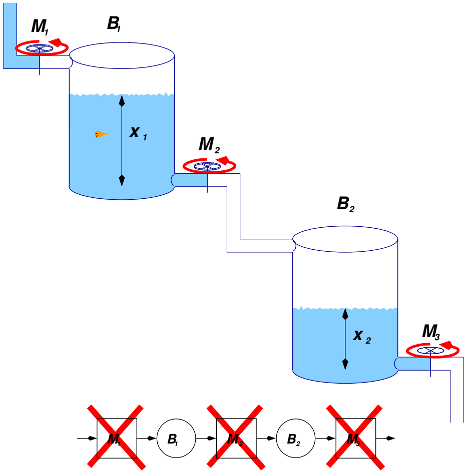
Spaces

Dimensionality

- Continuous random variables can be defined
  - ⋆ in one, two, three, ..., infinite dimensional spaces;
  - ⋆ in finite or infinite regions of the spaces.
- Continuous random variables can have
  - ⋆ probability measures with the same dimensionality as the space;
  - ⋆ lower dimensionality than the space;
  - ⋆ a mix of dimensions.

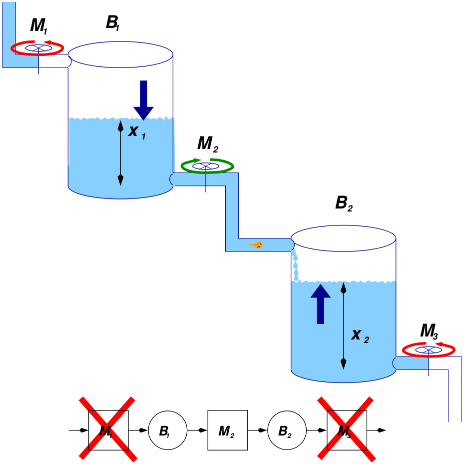
Continuous Random Variables

No change in water levels



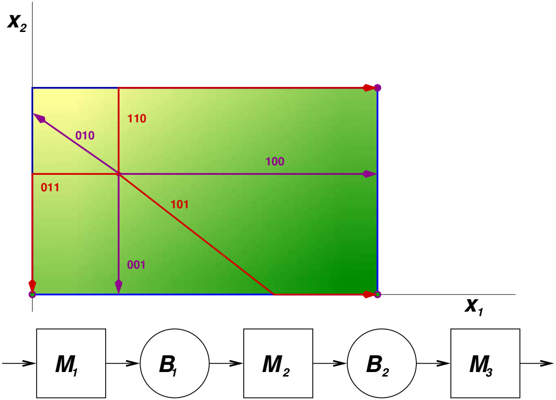
Continuous Random Variables

One kind of change in water levels



Continuous Random Variables

Trajectories

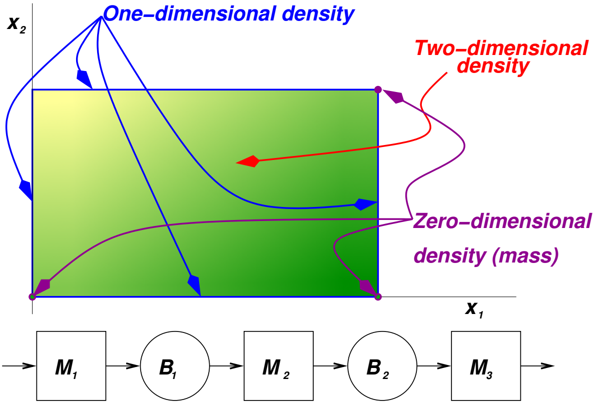


Trajectories of buffer levels in the three-machine line if the machine states stay constant for a long enough time period.

Notation: 110 means M_1 and M_2 are operational and M_3 is down, 100 means M_1 is operational, M_2 and M_3 are down, etc.

Continuous Random Variables

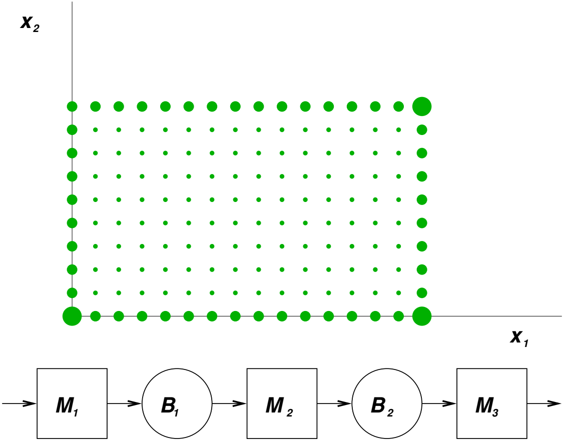
Two-dimensional probability distribution



Probability distribution of the amount of material in each of the two buffers.

Continuous Random Variables

Discrete approximation of the probability distribution



Probability distribution of the amount of material in each of the two buffers.

Continuous Random Variables

Densities and Distributions

In one dimension, $F()$ is the *cumulative probability distribution of* X if

$$F(x) = P(X \leq x)$$

$f()$ is the *density function of* X if

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Therefore,

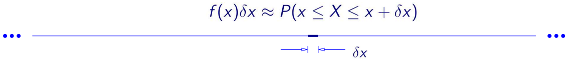
$$f(x) = \frac{dF}{dx}$$

wherever F is differentiable.

Continuous Random Variables

Densities and Distributions

Fact: $f(x)\delta x \approx P(x \leq X \leq x + \delta x)$ for sufficiently small δx.



Fact: $F(b) - F(a) = \int_a^b f(t)dt$

Definition: Expected value of $x = \bar{x} = \int_{-\infty}^{\infty} tf(t)dt$

Continuous Random Variables

Standard Normal Distribution

The density function of the *normal* (or *gaussian*) distribution with mean 0 and variance 1 (the *standard normal*) is given by

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

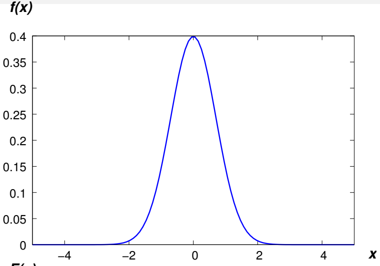
The *normal distribution function* is

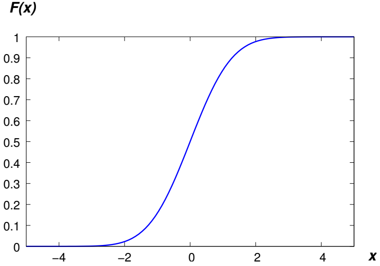
$$F(x) = \int_{-\infty}^{x} f(t)dt$$

(There is no closed form expression for $F(x)$.)

Continuous Random Variables

Standard Normal Distribution





Continuous Random Variables

Normal Distribution

Notation: $N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2.

Note: Some people write $N(\mu, \sigma)$ for the normal distribution with mean μ and variance σ^2.

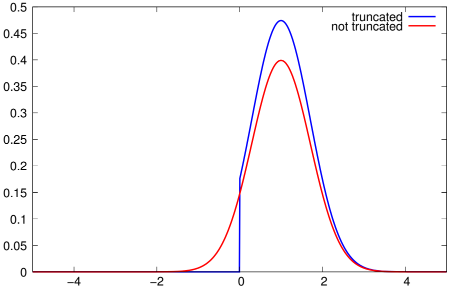
Fact: If X and Y are normal, then $aX + bY + c$ is normal.

Fact: If X is $N(\mu, \sigma)$, then $\frac{X-\mu}{\sigma}$ is $N(0, 1)$, the standard normal.

Consequently, $N(\mu, \sigma)$ easy to compute from $N(0, 1)$. This is why $N(0, 1)$ is tabulated in books.

Continuous Random Variables

Truncated Normal Density (1)

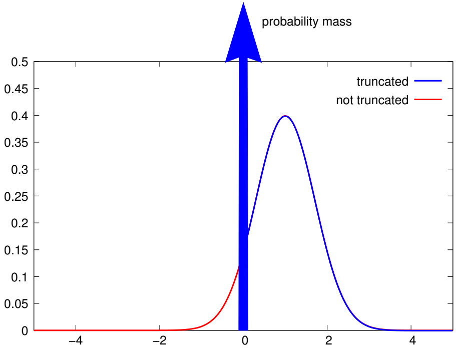


$f_T(x)\delta x = P(x \leq X \leq x + \delta x) = \dfrac{f(x)}{1 - F(0)}\delta x$ where $F()$ and $f()$ are the normal distribution and density functions with parameters μ and σ.

Note: μ and σ are the parameters of $f(x)$, *not* $f_T(x)$.

Continuous Random Variables

Truncated Normal Density (2)



$f_{T'}(x)\delta x = P(x \leq X \leq x + \delta x) = f(x)\delta x$ for $x > 0$ and $P(X = 0) = F(0)$ where $F()$ and $f()$ are the normal distribution and density functions with parameters μ and σ.

Here again, μ and σ are the parameters of $f(x)$, *not* $f_{T'}(x)$.

For *both* kinds of truncation, $f_T(x)$ and $f_{T'}(x)$ are close to $f(x)$ when $\mu \gg \sigma$, and not otherwise.

Continuous Random Variables

Law of Large Numbers

Let $\{X_k\}$ be a sequence of independent identically distributed *(i.i.d.)* random variables that have finite mean μ. Let S_n be the sum of the first n X_ks, so

$$S_n = X_1 + \ldots + X_n$$

Then for every $\epsilon > 0$,

$$\lim_{n \to \infty} P\left(\left| \frac{S_n}{n} - \mu \right| > \epsilon \right) = 0$$

That is, *the average approaches the mean.*

Continuous Random Variables

Central Limit Theorem

Let $\{X_k\}$ be a sequence of i.i.d. random variables with finite mean μ and finite variance σ^2.

Then as $n \to \infty$, $P(\frac{S_n - n\mu}{\sqrt{n}\sigma}) \to N(0, 1)$.

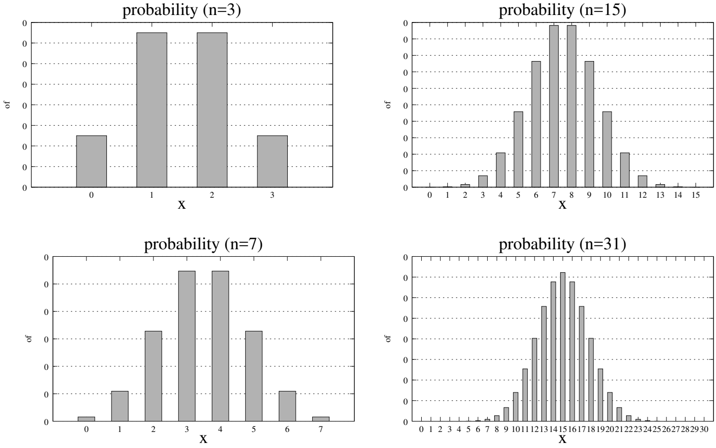
If we define A_n as S_n/n, the average of the first n X_ks, then this is equivalent to:

As $n \to \infty$, $P(A_n) \to N(\mu, \sigma/\sqrt{n})$.

Continuous Random Variables

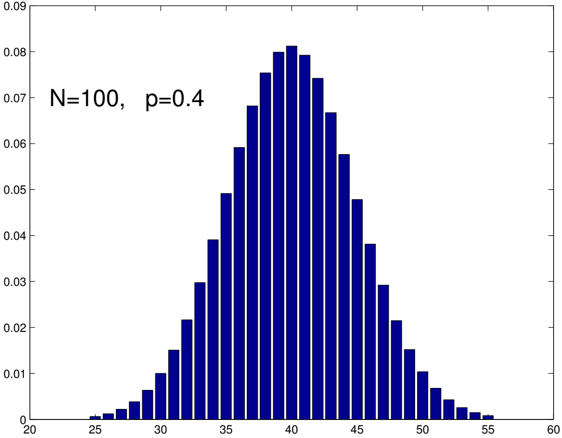
Coin flip examples

Probability of x heads in n flips of a fair coin



Continuous Random Variables

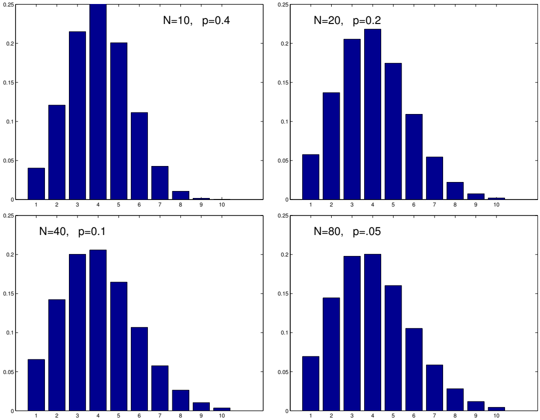
Binomial probability distribution approaches normal for large N.



Continuous Random Variables

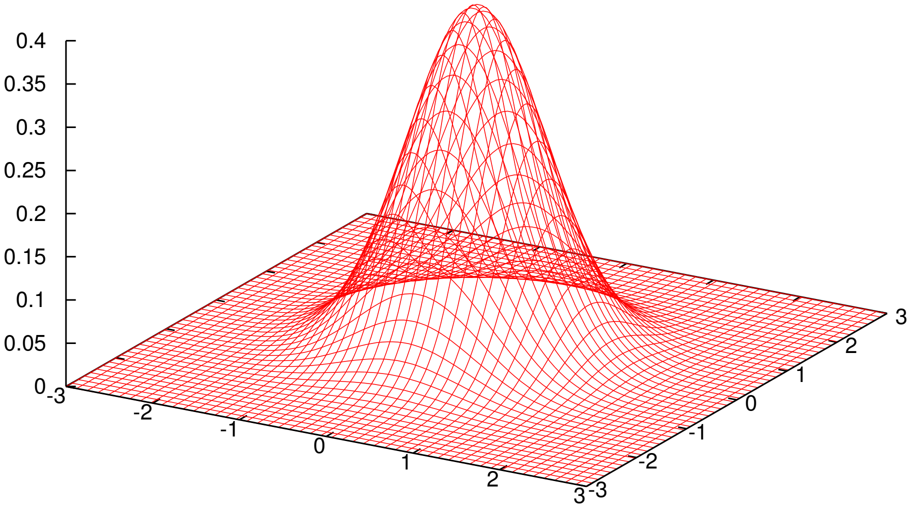
Binomial distributions

Note the resemblance to a *truncated* normal in these examples.



Normal Density Function

... in Two Dimensions



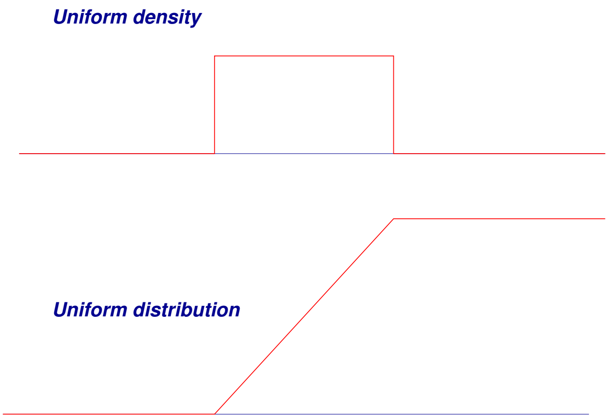
More Continuous Distributions

Uniform

$$f(x) = \frac{1}{b-a} \quad \text{for } a \leq x \leq b$$

$$f(x) = 0 \qquad \text{otherwise}$$

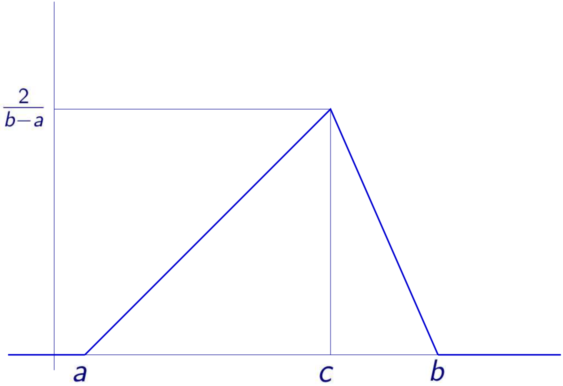
More Continuous Distributions
Uniform



More Continuous Distributions

Triangular

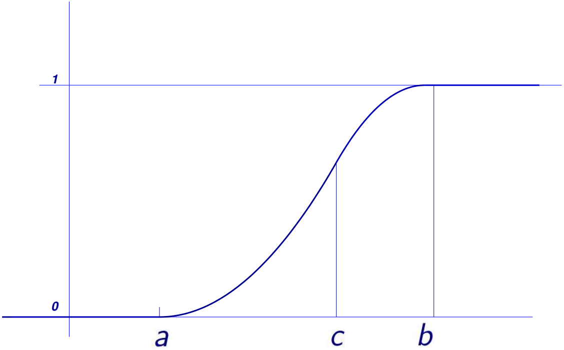
Probability density function



More Continuous Distributions

Triangular

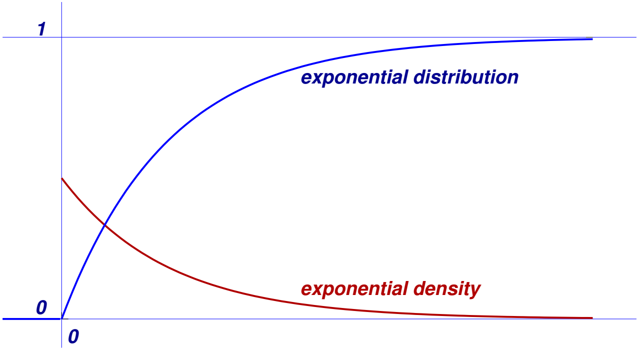
Cumulative distribution function



More Continuous Distributions

Exponential

- Very often used for the time until a specified event occurs.
- **Density:** $f(t) = \lambda e^{-\lambda t}$ **for** $t \geq 0$; $f(t) = 0$ **otherwise;**
- Distribution: $F(t) = P(T \leq t) = 1 - e^{-\lambda t}$ for $t \geq 0$; $F(t) = 0$ otherwise.



More Continuous Distributions

Exponential

- Close to the geometric distribution but for continuous time.
- *Very* mathematically convenient.
- Memorylessness:

  $$P(T > t + x | T > x) = P(T > t)$$

  Suppose an exponentially distributed process is started at time 0 and the event of interest has not occurred yet at time x. Then the probability distribution of the time after x at which it occurs is the same as the original exponential distribution. The process has no "memory" of when it was actually started.

Another Discrete Random Variable

Poisson Distribution

$$P(X^P = x) = e^{-\lambda t}\frac{(\lambda t)^x}{x!}$$

is the probability that x events happen in $[0, t]$ if the events are independent and the times between them are exponentially distributed with parameter λ.

Typical examples: arrivals and services at queues. *(Next lecture!)*

NOT Random

...but almost

- A *pseudo-random number generator* is a set of numbers $X_0, X_1, ...$ where there is a function F such that

$$X_{n+1} = F(X_n)$$

and F is such that the sequence of X_n satisfies certain conditions.

- For example,
  - there is a known finite maximum $X^{\max}$,
  - $0 \leq X_n \leq X^{\max}$,
  - and the sequence $U_0, U_1, ...$ (where $U_i = X_i / X^{\max}$) *looks like* a set of uniformly distributed, independent random variables.
    - That is, statistical tests say that the probability of the sequence *not* being independent uniform random variables is very small.

NOT Random

...but almost

- The sequence is deterministic: it is determined by X_0, the *seed* of the random number generator.
- If you use the same seed twice, you get the same sequence both times. This can be convenient, especially in development of software.
- If you use different seeds, you get completely different sequences, even if the seeds are close to one another.
- Pseudo-random number generators are used extensively in *simulation*.