Aalto University
School of Science

Decision making and problem solving – probability calculus revision material

2019

Why probabilities?

- Decisions are often made under uncertainty
  - "How many metro drivers should be recruited = trained, when future traffic is uncertain?"
- Probability theory dominates the modeling of uncertainty in decision analysis
  - Well established rules for computations
  - Understandable
  - Other models (e.g., evidence theory, fuzzy sets) are not covered here
- Learning objective: refresh memory about probability theory and calculations

The sample space

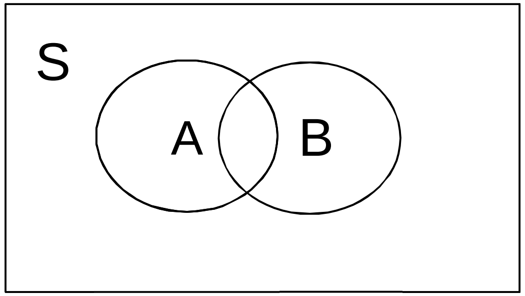
- Sample space S = set of all possible outcomes
- Examples:
  - A coin toss: S = {H,T}
  - Two coin tosses: S = {HH, TT, TH, HT}
  - Number of rainy days in Helsinki in 2018: S={1,...,366}
  - Grades from four courses: $S=G \times G \times G \times G=G^4$, where G={0,...,5}
  - Average m^2-price for apartments in Helsinki area next year $S = [0,\infty)$ euros

Simple events and events

- Simple event: an individual outcome from *S*
  - A coin toss: T
  - Two coin tosses: TT
  - Number of rainy days in Helsinki in 2018: 180
  - Grades from four courses: (4, 5, 3, 4)
  - Average m^2-price for apartments in Helsinki in 2019: 4000 €
- Event: a collection of one or more outcomes (i.e., a subset of the sample space: $E \subseteq S$)
  - Two coin tosses: First toss tails, E={TT, TH}
  - Number of rainy days in Helsinki in 2018: Less than 100, E={0,...,99}
  - Grades from four courses: Average at least 4.0, $E = \left\{ z \in G^4 \middle| \frac{1}{4} \sum_{i=1}^{4} z_i \geq 4.0 \right\}$
  - Average m^2-price for apartments in Helsinki in 2019: Above 4000€, E=(4000, ∞)

Events derived from events: Complement, union, and intersection

- **Complement** A^c of A = all outcomes in S that are not in A
- **Union** $A \cup B$ of two events A and B = all outcomes that are in A **or** B (or both)
- **Intersection** $A \cap B$ = all outcomes that are in both events
- A and B with no common outcomes are **mutually exclusive**
- A and B are **collectively exhaustive** if $A \cup B = S$



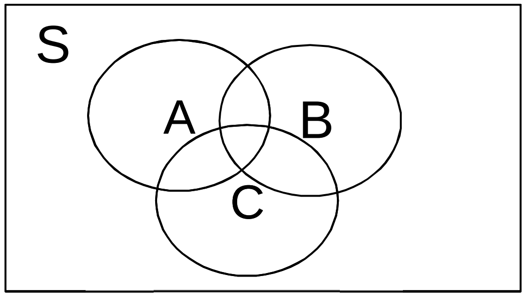
Events derived from events: Laws of set algebra

Commutative laws: $A \cup B = B \cup A$, $A \cap B = B \cap A$

Associative laws: $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$,

Distributive laws: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

DeMorgan's laws: $(A \cup B)^C = A^C \cap B^C$, $(A \cap B)^C = A^C \cup B^C$

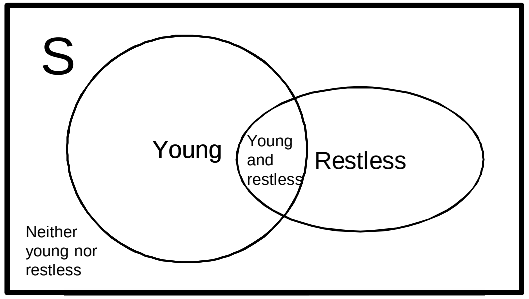


Probability measure

- **Definition:** Probability *P* is a function that maps all events *A* onto real numbers and satisfies the following three axioms:
  1. P(S)=1
  2. $0 \leq P(A) \leq 1$
  3. If A and B are mutually exclusive (i.e., $A \cap B = \emptyset$) then $P(A \cup B) = P(A) + P(B)$

Properties of probability (measures)

- From the three axioms it follows that
  I. $P(\emptyset)=0$
  II. If $A \subseteq B$, then $P(A) \leq P(B)$
  III. $P(A^C) = 1 - P(A)$
  IV. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



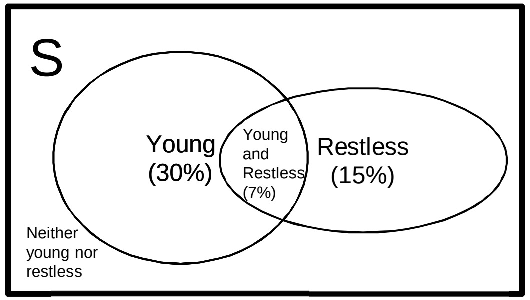
- In a given population, 30% of people are young, 15% are restless, and 7% are both young and restless. A person is randomly selected from this population. What is the chance that this person is
  - Not young? 1. 30% 2. 55% 3. 70%
  - Young but not restless? 1. 7% 2. 15% 3. 23%
  - Young, restless or both? 1. 38% 2. 45% 3. 62%

Independence

Definition: Two events A and B are independent if

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

- A person is randomly selected from the population on the right.
- Are events ”the person is young” and ”the person is restless” independent?
  - No: $0.07 \neq 0.3 \times 0.15$



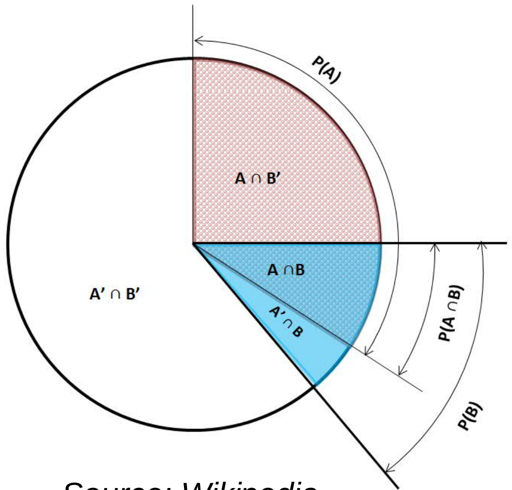
Conditional probability

Definition: Conditional probability P(*A*|*B*) of *A* given that *B* has occurred is

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

Note: If A and B are independent, the probability of *A* does not depend on whether *B* has occurred or not:

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



Source: Wikipedia

Joint probability vs. conditional probability

Example:

- A farmer is trying to decide on a farming strategy for next year. Experts have made the following forecasts about the demand for the farmer's products.
- Questions:
  - What is the probability of high wheat demand?
    1. 40% 2. 65% 3. 134%
  - What is the probability of low rye demand?
    1. 11% 2. 35% 3. 45%
  - What is the (conditional) probability of high wheat demand, if rye demand is low?
    1. 40% 2. 55% 3. 89%
  - Are the demands independent?
    1. Yes 2. No

Joint probability

	Wheat demand		
Rye demand	Low	High	**Sum**
Low	0.05	0.4	0.45
High	0.3	0.25	0.55
Sum	0.35	0.65	1

Conditional probability

	Wheat demand		
Rye demand	Low	High	**Sum**
Low	0.11	0.89	1
High	0.55	0.45	1
Sum	0.66	1.34	

Law of total probability

- If $E_1,\ldots,E_n$ are mutually exclusive and $A = \bigcup_i E_i$, then

$$P(A)=P(A|E_1)P(E_1)+\ldots+P(A|E_n)P(E_n)$$

- Most frequent use of this law:
  - Probabilities $P(A|B)$, $P(A|B^c)$, and $P(B)$ are known
  - These can be used to compute $P(A)=P(A|B)P(B)+P(A|B^c)P(B^c)$

Bayes' rule

- **Bayes' rule:** $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
- Follows from
  1. The definition of conditional probabilty: $P(A|B) = \frac{P(A \cap B)}{P(B)}$, $P(B|A) = \frac{P(B \cap A)}{P(A)}$,
  2. Commutative laws: $P(B \cap A) = P(A \cap B)$.

Bayes' rule

Example:

- The probability of a fire in a certain building is 1/10000 any given day.
- An alarm goes off whenever there is an actual fire, but also once in every 200 days for no reason.
- Suppose the alarm goes off. **What is the probability that there is a fire?**

Solution:

- F=Fire, F^c=No fire, A=Alarm, A^c=No alarm
- $P(F)=0.0001$ $P(F^c)=0.9999$, $P(A|F)=1$, $P(A|F^c)=0.005$

Law of total probability: $\mathbf{P(A)}=P(A|F)P(F)+P(A|F^c)\,P(F^c)=\mathbf{0.0051}$

Bayes: $P(F|A) = \frac{P(A|F)P(F)}{P(A)} = \frac{1 \cdot 0.0001}{0.0051} \approx 2\%$

Random variables

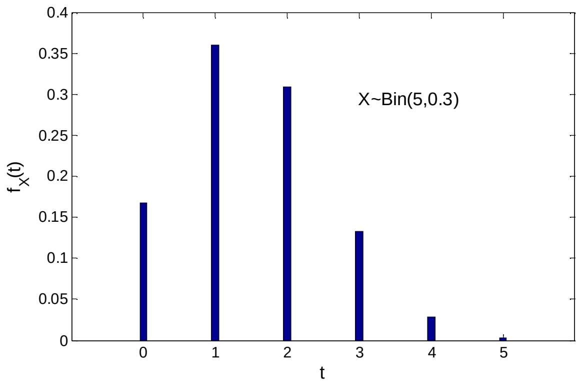
- A random variable is a mapping from sample space *S* to real numbers (discrete or continuous scale)

- The probability measure *P* on the sample space defines a **probability distribution** for these real numbers

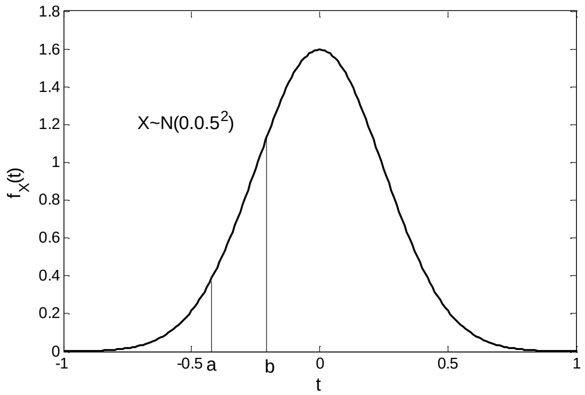
- Probability distribution can be represented by
  - Probability mass (discrete) / density (continuous) function
  - Cumulative distribution function

Probability mass/density function (PMF & PDF)

- PMF of a discrete random variable is $f_X(t)$ such that
  – $f_X(t)=P(\{s \in S | X(s)=t\})$ = probability
  – $\sum_{t\in(a,b]} f_X(t) = P(X \in (a,b])$ = probability



- PDF of a continuous random variable is $f_X(t)$ such that
  – $f_X(t)$ is NOT a probability
  – $\int_a^b f_X(t)dt = P(X \in (a,b])$ is a probability



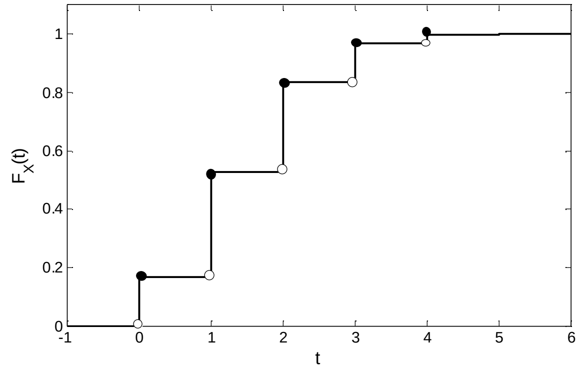
Cumulative distribution function (CDF)

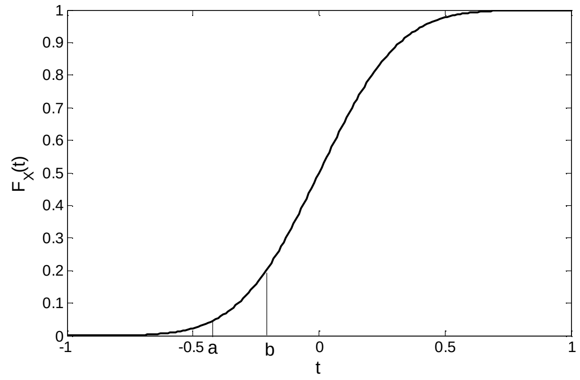
- The CDF of random variable X is

$$F_X(t) = P(\{s \in S | X(s) \le t\})$$

(often $F(t) = P(X \le t)$)

- Properties
  - F_X is non-decreasing
  - $F_X(t)$ approaches 0 (1) when t decreases (increases)
  - $P(X>t)=1-F_X(t)$
  - $P(a<X\le b)= F_X(b)- F_X(a)$





Expected value

- The expected value of a random variable is the weighted average of all possible values, where the weights represent probability mass / density at these values

Discrete X	Continuous X
$E[X] = \sum_t t f_X(t)$	$E[X] = \int_{-\infty}^{\infty} t f_X(t) dt$

- A function g(X) of random varibale X is itself a random variable, whereby

$$E[g(X)] = \sum_t g(t) f_X(t)$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(t) f_X(t) dt$$

Expected value: Properties

- If $X_1, \dots, X_n$ and $Y = \sum_{i=1}^{n} X_i$ are random variables, then
$$E[Y] = \sum_{i=1}^{n} E[X_i]$$
- If random variable $Y=aX+b$ where a and b are constants, then
$$E[Y] = aE[X] + b$$

- **NB!** In general, E[g(X)]=g(E[X]) does NOT hold:
  - Let $X \in \{0,1\}$ with P(X=1)=0.7. Then,

    $E[X] = 0.3 \cdot 0 + 0.7 \cdot 1 = 0.7$,

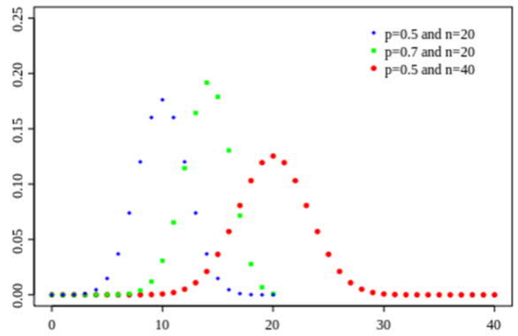
    $E[X^2] = 0.3 \cdot 0^2 + 0.7 \cdot 1^2 = 0.7 \neq 0.49 = (E[X])^2$.

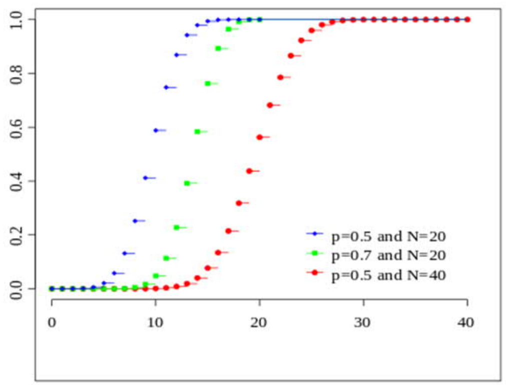
Random variables vs. sample space

- Models are often built by directly defining distributions (PDF/PMF or CDF) rather than starting with the sample space
  - Cf. alternative models for coin toss:
    1. Sample space is S={H,T} and its probability measure P(s)=0.5 for all $s \in S$
    2. PMF is given by $f_X(t)=0.5$, $t \in \{0,1\}$ and $f_X(t)=0$ elsewhere
- Computational rules that apply to event probabilities also apply when these probabilities are represented by distributions
- Detailed descriptions about the properties and common uses of different kinds of discrete and continuous distributions are widely documented
  - Elementary statistics books
  - Wikipedia

Binomial distribution

- n independent binary (0/1, no/yes) trials, each with success probability $p=P(X=1)$
- The number X ~ Bin(n,p) of successful trials is a random variable that follows the binomial distribution with parameters n and p
- PMF: $P(X=t) = f_X(t) = \binom{n}{t} p^t (1-p)^{n-t}$
- Expected value E[X]=np
- Variance Var[X]=np(1-p)





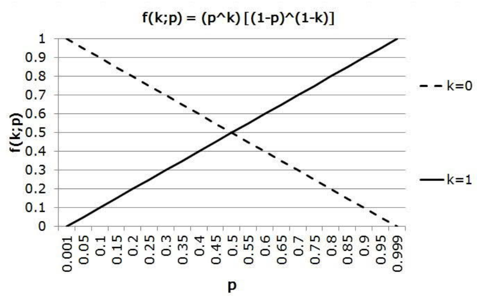
Source: Wikipedia

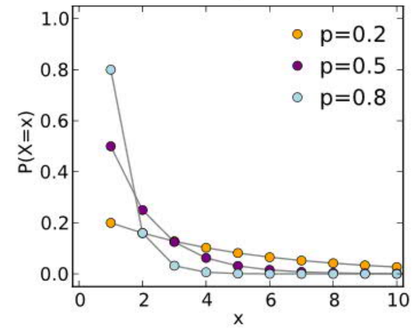
Other common discrete distributions

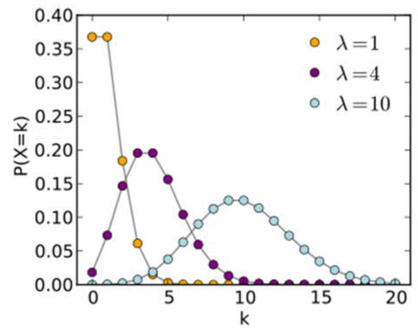
- **Bernoulli distribution**
  - If X ∈{0,1} is the result of a single binary trial with success probability p, then X~Bernoulli(p).
  - $f_X(t) = p^t(1-p)^{1-t}$

- **Geometric distribution**
  - If X ∈{1,2,3,...} is the number of Bernoulli trials needed to get the first success, then X~Geom(p).
  - $f_X(t) = p(1-p)^{t-1}$

- **Poisson distribution**
  - Let X ∈{1,2,3,...} be the number of times that an event occurs during a fixed time interval such that (i) the average occurence rate λ is known and (ii) events occur independently of the last event time. Then, X~Poisson(λ).
  - $f_X(t) = \frac{\lambda^k e^{-\lambda}}{k!}$







Source: Wikipedia

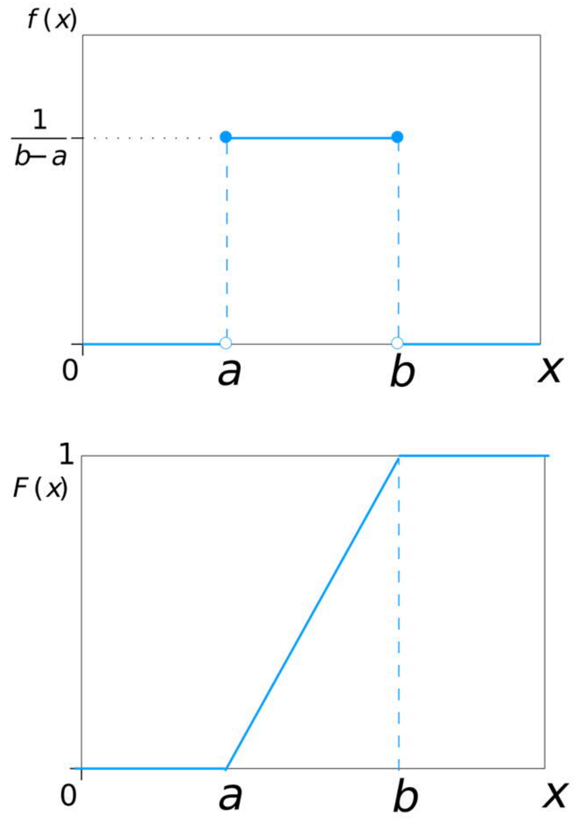
Uniform distribution

- Let $X \in [a,b]$ such that each real value within the interval has equal probability. Then, X~Uni(a,b)

- $f_X(t) = \begin{cases} \frac{1}{b-a}, & \text{for } a \leq t \leq b \\ 0, & \text{otherwise} \end{cases}$

- $E[X] = \frac{a+b}{2}$

- $Var[X] = \frac{1}{12}(b-a)^2$



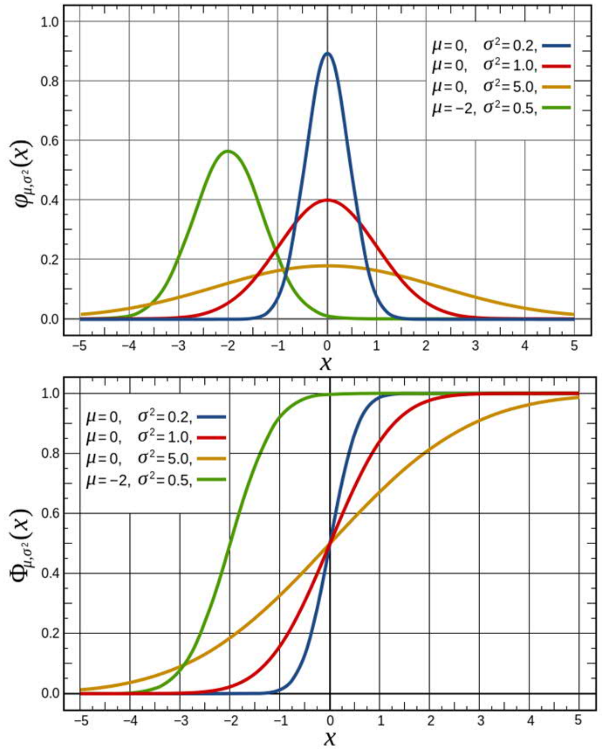
Source: Wikipedia

Normal distribution N(μ, σ^2)

- $f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$
- $E[X] = \mu, Var[X] = \sigma^2$
- The most common distribution for continuous random variables

- **Central limit theorem:** Let $X_1,\ldots,X_n$ be independent and identically distributed random variables with E[X_i]= μ and Var[X_i]=σ^2. Then,

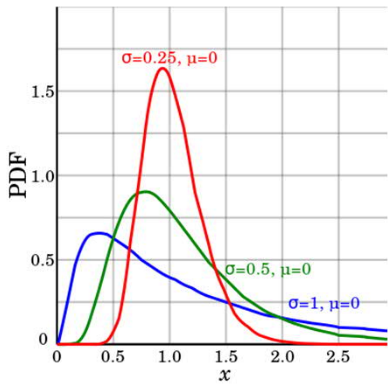
$$\frac{\sum_{i=1}^{n} X_i}{n} \sim_a N\left(\mu, \frac{\sigma^2}{n}\right).$$



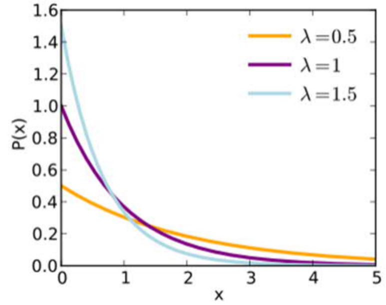
Source: Wikipedia

Other common continuous distributions

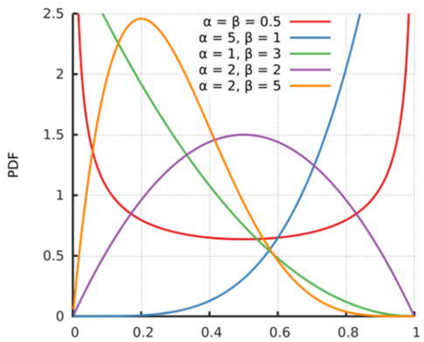
- Log-normal distribution: if $X \sim N(\mu, \sigma^2)$, then $e^X \sim \mathrm{LogN}(\mu, \sigma^2)$



- Exponential distribution Exp(λ): describes the time between events in a Poisson process with event occurrence rate λ



- Beta distribution Beta(α,β): distribution for X∈[0,1] that can take various forms



Why Monte Carlo simulation?

- When probabilitistic models are used to support decision making, alternative decisions often need to be described by 'performance indices' such as
  - Expected values – e.g., expected revenue from launching a new product to the market
  - Probabilities of events – e.g., the probability that the revenue is below 100k€

- It may be difficult, time-consuming or impossible to calculate such measures analytically

- Monte Carlo simulation:
  - Use of a computer program to generate samples from the probability model
  - Estimation of expected values and event probabilites from these samples

Monte Carlo simulation of a probability model

Probability model

- Random variable $X \sim f_X$

$$E[X]$$

$$E[g(X)]$$

$$P(a < X \leq b)$$

Monte Carlo simulation

- Sample $(x_1, \ldots, x_n)$ from f_X

$$\frac{\sum_{i=1}^{n} x_i}{n}$$

$$\frac{\sum_{i=1}^{n} g(x_i)}{n}$$

$$\frac{|\{i \in \{1, \ldots, n\} | x_i \in (a, b)\}|}{n}$$

Uni(0,1) distribution in MC – discrete random variables

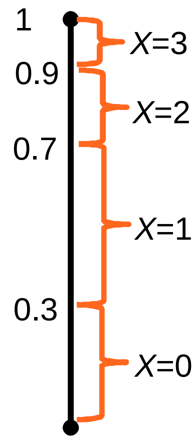
- ❑ Some softwares only generate random numbers from Uni(0,1)-distribution
- ❑ Samples from Uni(0,1) can, however, be transformed into samples from many other distributions

❑ <u>Discrete distribution:</u>

- Let $X \in \{x_1, \ldots, x_n\}$ such that $f_X = P(X = x_i) = p_i$.
- Divide interval [0,1] into n segments of lengths $p_1, \ldots, p_n$.
- Sample values u_j from Uni(0,1).
- Transform the sample: If $u_j \in [\sum_{k=0}^{i-1} p_k, \sum_{k=0}^{i} p_k)$ where $p_0 = 0$, then $X_j = x_i$.

U~Uni(0,1)

$u_1 = 0.4565$
$u_2 = 0.8910$
$u_3 = 0.3254$
$\vdots$



X~f_X

$x_1 = 1$
$x_2 = 2$
$x_3 = 1$
$\vdots$

Demand x / week	Prob. f_X of demand
0	0.3
1	0.4
2	0.2
3	0.1

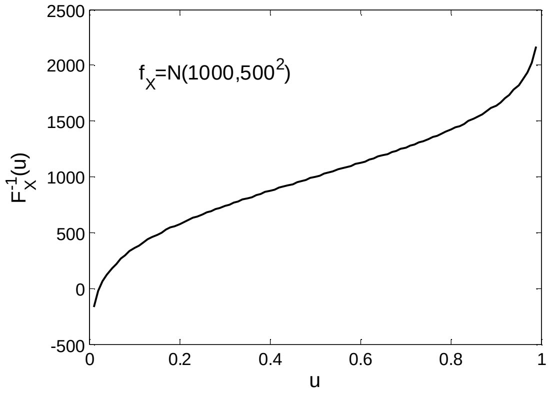
Uni(0,1) distribution in MC – continuous random variables

- Assume that the CDF of random variable X has an inverse function F_X^{-1}. Then, the random variable $Y = F_X^{-1}(U)$ where U~Uni(0,1) follows the same distribution as X:

$$F_Y(t) = P(Y \le t) = P(F_X^{-1}(U) \le t) = P\big(U \le F_X(t)\big) = F_X(t)$$

- Continuous distribution:
  - Let $X \sim F_X$ (CDF)
  - Sample values u_j from Uni(0,1).
  - Transform the sample: $X_j = F_X^{-1}(u_j)$

U~Uni(0,1)	X~f_X
u_1 = 0.4565	x_1 = 945.4
u_2 = 0.8910	x_2 = 1615.9
u_3 = 0.3254	x_3 = 773.7
⋮	⋮



Monte Carlo simulation in Excel

VLOOKUP looks for the cell value in the 1st column of the table. The value in the 3rd column of the table is returned to the current cell.

=VLOOKUP(G7;$B$7:$D$10;3;TRUE)

AVERAGE(H7:H206)

			C	D	E	F	G	H
2								
3						True mean	0.5	1.1
4						Sample mean	0.498714	1.085
5								
6	i	Sum p0:p(i-1)	Probability pi	Demand xi		Sample	u	x
7	1	0	0.3	0		1	0.009979	3;TRUE)
8	2	0.3	0.4	1		2	0.423969	1
9	3	0.7	0.2	2		3	0.931674	3
10	4	0.9	0.1	3		4	0.963706	3
11		1				5	0.500698	1
12						6	0.628946	1
13						7	0.056035	0
14						8	0.762916	2
15						9	0.401607	1
16						10	0.937021	3
17						11	0.862141	2
18						12	0.895572	2

RAND() generates a random number from Uni(0,1)

STDEV.S(E8:E207)

=NORM.INV(E8;1000;500)

D	E	F	G
True mean	0.5	1000	
Sample mean	0.518524	1020.184	
True stdev	0.288675	500	
Sampe stdev	0.296019	503.2426	
Sample	u	x	
1	0.049976	177.4551	
2	0.205365	588.695	
3	0.874753	1574.575	
4	0.970594	1944.799	
5	0.968038	1926.357	
6	0.643137	1183.428	
7	0.26185	681.174	
8	0.404865	879.6124	
9	0.642356	1182.382	
10	0.200953	580.889	
11	0.297499	734.1966	
12	0.858584	1536.989	

Monte Carlo simulation in Matlab

```
S=200; %Number of simulation rounds
p=[0.3 0.4 0.2 0.1]; %PMF for x
P=[0.3 0.7 0.9 1]; %CDF for x
X=[0 1 2 3]; %Possible values of x
Sample=zeros(S,1); %Initialize the sample vector
for k=1:S;
    r=rand; %Random number from Uni(0,1)
    counter=1; %Start looking from the first value of X
    while(r>P(counter)) %While r is greater than the CDF at current value of X...
        counter=counter+1; %We go to the next value of X.
    end %When r is lower than the CDF at the current value of X...
    Sample(k)=X(counter); %We have found the value of X corresponding to r
end
TrueMean=p*X'
SampleMean=mean(Sample)
```

Monte Carlo simulation in Matlab

- Statistics and Machine Learning Toolbox makes it easy to generate numbers from various distributions
- E.g.,
  - `Y=normrnd(mu,sigma,m,n)`: $m \times n$-array of X~N(mu,sigma)
  - `Y=betarnd(A,B,m,n)`: $m \times n$-array of X~Beta(A,B)
  - `Y=lognrnd(mu,sigma,m,n)`: $m \times n$-array of X~LogN(mu,sigma)
  - `Y=binornd(N,P,m,n)`: $m \times n$-array of X~Bin(N,P)
  - ...

Summary

- Probability is the dominant way of capturing uncertainty in decision models

- Well-established computational rules provide means to derive probabilities of events from those of other events
  - Conditional probability, law of total probability, Bayes' rule

- To support decision making, probabilistic models are often used to compute performance indices (expected values, probabilities of events, etc.)

- Such indices can easily be computed through Monte Carlo simulation