Stochastic Signal Processing

Lesson 2: Repeated Trials and Random Variables

Weize Sun

Note !!!!!!

- In the text book, a random variable (r.v) is written as $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \dots$, **Lowercase bold letter**
- In some other books, a r.v is written as $X, Y, Z, \dots$, Uppercase letter without bold
- Here we use both, which is, both $\boldsymbol{x}$ and X are used to represent a r.v, in order to be consistent with most books
- However, the $x, y, z, \dots$, lowercase letter without bold, is used to represent a value only and always

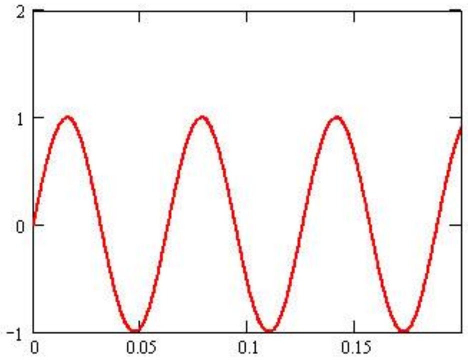
Course reminder

- This is a math class
- The examples in the slides (lecture notes) are very important
- The homework/assignments/experiments are also very important
  - They must be submitted on time
- In some slides, there will be some "review questions (examples)"
  - They will be solved and explained in the beginning of ‘next class’
  - It is recommended to finish it before the ‘next class’ as a review
- Before the final exam, there will be some example questions, but do not review this course fully based on that

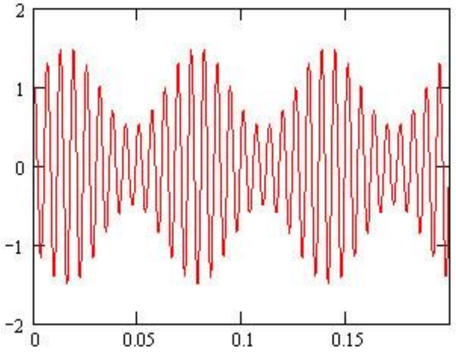
Course reminder

- In the first 4 lessons, I will mainly go directly through the examples, as these lectures are ‘a review of probability that you are supposed learned well’
- But in the remainder lessons, when come to some important examples, I will give some time for you to ‘calculate it’
- Reading:
    - This week: Text book, Chapters 1-2
    - Next week: Text book, Chapters 3-4

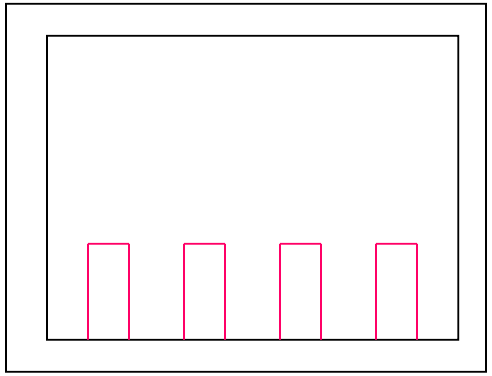
- In real life, we will encounter many kinds of signals every day



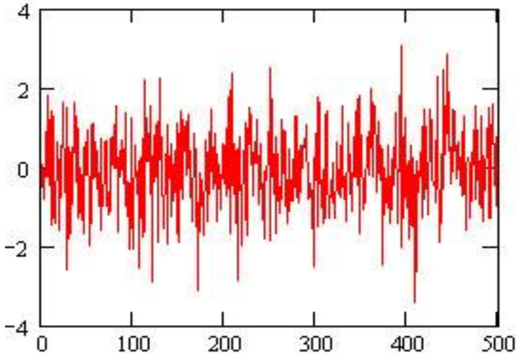
sinusoidal signal



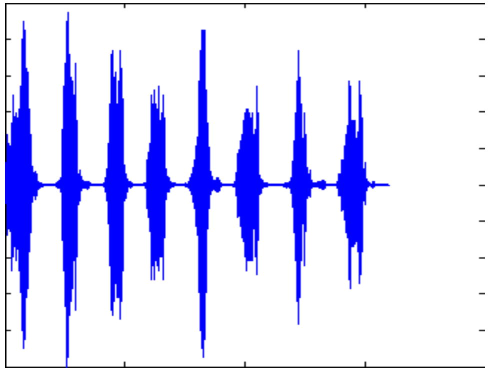
Modulation signal



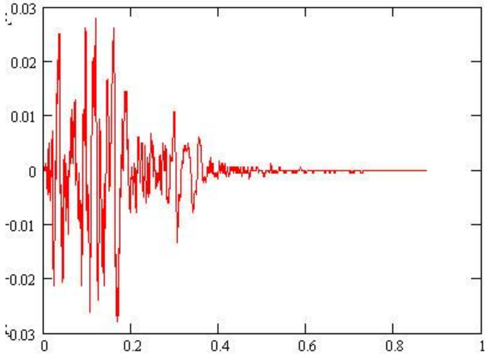
Periodic pulse signal



Noise of radar receiver



Birdsong



blasting signal

The above signals can be simply divided into two categories:

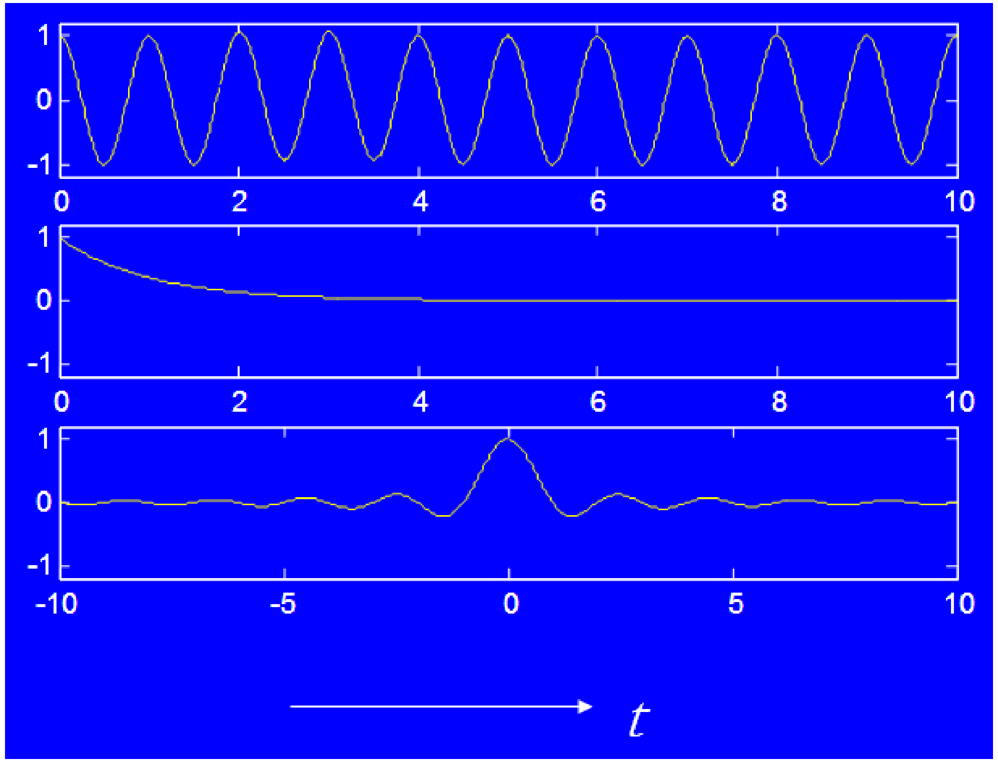
$$\text{Signal}\begin{cases}\text{Deterministic signals}\\ \\ \text{Stochastic signals}\end{cases}$$

Deterministic signal: It follows a fixed rule that can be described by mathematical expressions, rules or tables.

$$\cos 2\pi f_c t$$

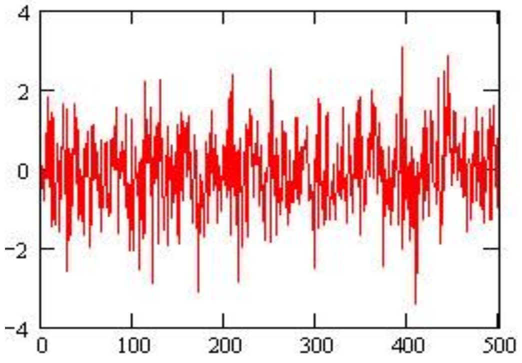
$$e^{-\alpha t}u(t)$$

$$\frac{\sin \pi t}{\pi t}$$

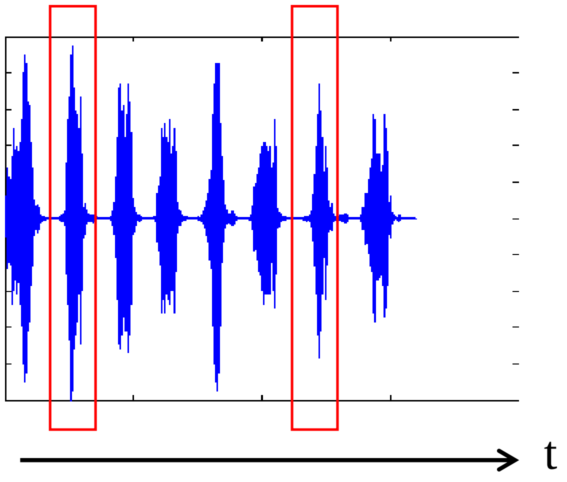


Stochastic signal: Signal with high uncertainty.

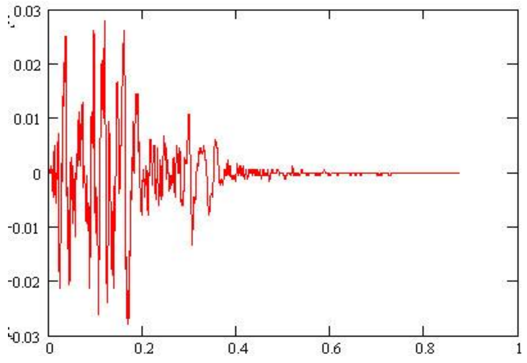
Receiver noise



Birdsong

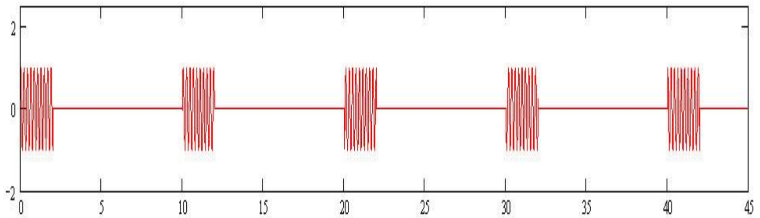


blasting signal

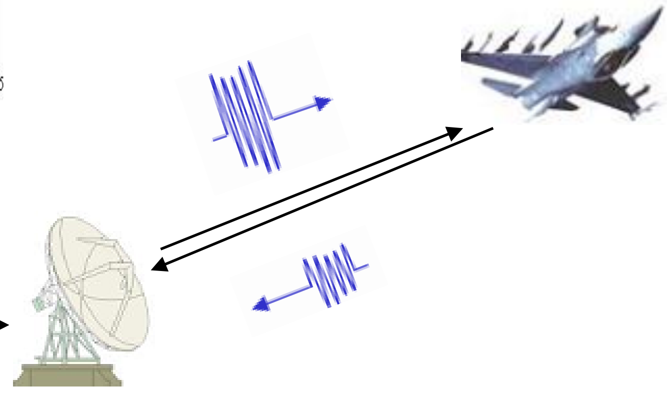


These signals change with time

Basic of Stoc. Sig. Proc.



Radar transmitter



switch

Two basic questions:

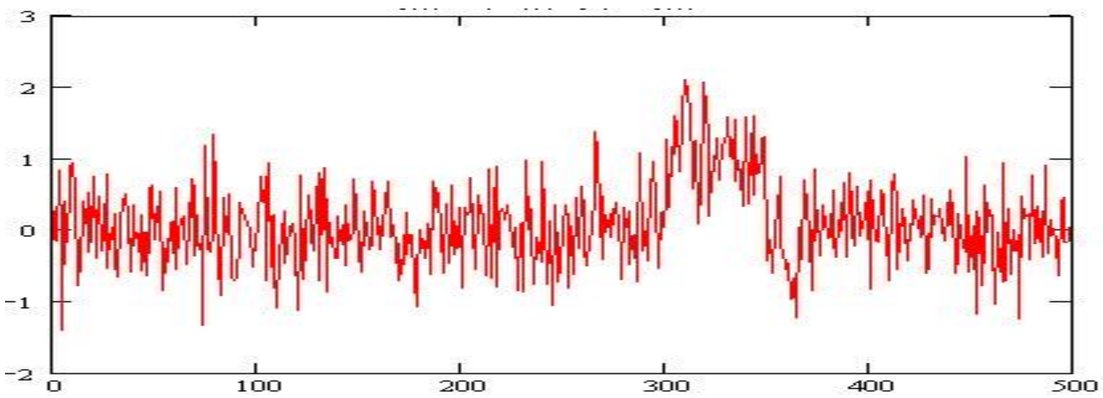
- How to detect echo signals (回波信号)
- How to estimate the parameters of a signal

receiver output

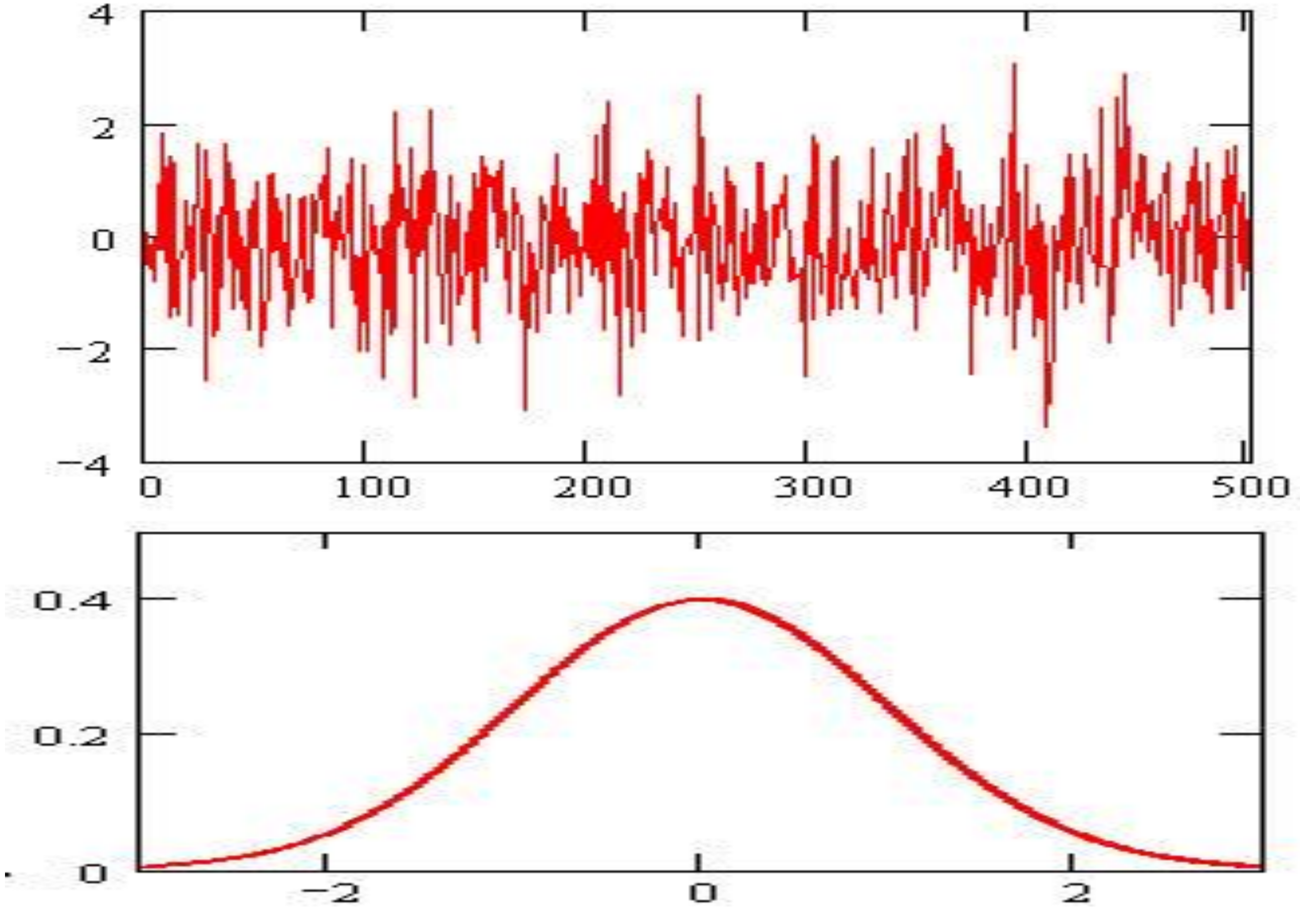
Radar receiver

Receiver output

(including noise)



Example: Signal received without signal (pure noise)



We can use features as:

mean

variance

PDF

…

Probability deals with the signals that do not vary according to time or state

Stochastic signal processing deals with the signals that do vary according to time or state

- **Independence**: Events A and B are independent if

$$P(AB) = P(A)P(B)$$

- It is easy to show that A, B independent implies $\bar{A}, B; A, \bar{B}; \bar{A}, \bar{B}$ are all independent pairs.

- For example, $B = (A \cup \bar{A})B = AB \cup \bar{A}B$ and $AB \cap \bar{A}B = \emptyset$, $\rightarrow$

$$P(B) = P(AB \cup \bar{A}B) = P(AB) + P(\bar{A}B) = P(A)P(B) + P(\bar{A}B)$$

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

- Therefore, $\bar{A}$ and B are independent events.

- Independent events obviously cannot be mutually exclusive:

$P(A) > 0, \quad P(B) > 0$ and A, B independent implies $P(AB) > 0$.Thus if A and B are independent, the event AB cannot be the null set.

- More generally, a family of events $\{A_i\}$ are said to be independent, if for every finite sub collection $A_{i_1}, A_{i_2}, \cdots, A_{i_n}$, we have

$$P\left(\cap_{k=1}^{n} A_{i_k}\right) = \prod_{k=1}^{n} P\left(A_{i_k}\right).$$

- Let $A = A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n$, a union of n independent events. Then by De-Morgan's law:

$$\bar{A} = \bar{A}_1 \bar{A}_2 \cdots \bar{A}_n$$

and using their independence

$$P(\bar{A}) = P(\bar{A}_1 \bar{A}_2 \cdots \bar{A}_n) = \prod_{i=1}^{n} P(\bar{A}_i) = \prod_{i=1}^{n} \big(1 - P(A_i)\big)$$

Thus for any A:

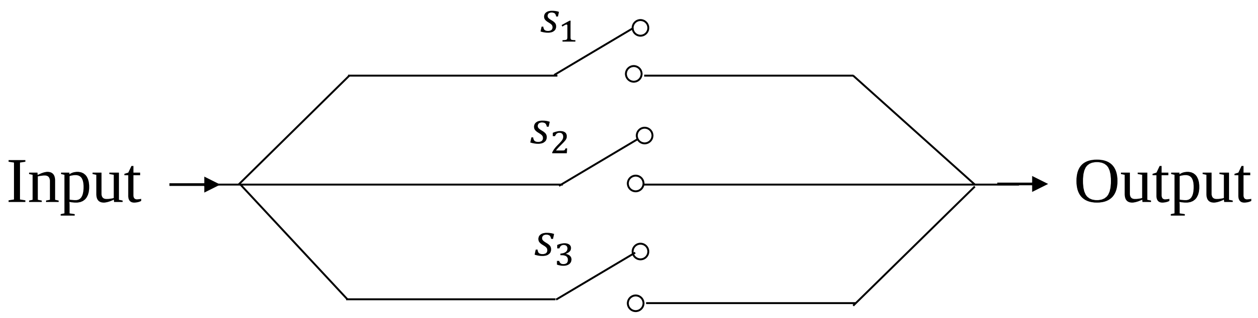
$$P(A) = 1 - P(\bar{A}) = 1 - \prod_{i=1}^{n} \big(1 - P(A_i)\big) \qquad (2\text{-}1)$$

This is a useful result. We can use these results to solve some problems

- Example 1: Three switches connected in parallel operate independently. Each switch remains closed with probability p.

(a) Find the probability of receiving an input signal at the output.

(b) Find the probability that switch S_1 is open given that an input signal is received at the output.



Solution: (a). Let A_i = "Switch S_i is closed". Then $P(A_i) = p,\ i = 1 \rightarrow 3$

Since switches operate independently, we have

$$P(A_iA_j) = P(A_i)P(A_j); P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3)$$

- Let R = "input signal is received at the output". For the event R to occur either switch 1 or switch 2 or switch 3 must remain closed, i.e.,

$$R = A_1 \cup A_2 \cup A_3 \quad P(A) = 1 - P(\bar{A}) = 1 - \prod_{i=1}^{n}\left(1 - P(A_i)\right), (2-1)$$

$$\Downarrow$$

$$P(R) = P(A_1 \cup A_2 \cup A_3) = 1 - (1-p)^3 = 3p - 3p^2 + p^3$$

- Note that the events A_1, A_2, A_3 do not form a partition, since they are not mutually exclusive. Obviously any two or all three switches can be closed (or open) simultaneously. Which is, $P(A_1) + P(A_2) + P(A_3) \neq 1$

- (b). We need $P(\bar{A}_1 \mid R)$. Note that $P(R \mid \bar{A}_1) = P(A_2 \cup A_3) = 2p - p^2$

From Bayes' theorem

$$P(\bar{A}_1 \mid R) = \frac{P(R \mid \bar{A}_1)P(\bar{A}_1)}{P(R)} = \frac{(2p - p^2)(1 - p)}{3p - 3p^2 + p^3} = \frac{2 - 2p + p^2}{3p - 3p^2 + p^3}$$

- Because of the symmetry of the switches, we also have

$$P(\bar{A}_1 \mid R) = P = (\bar{A}_2 \mid R) = P(\bar{A}_3 \mid R)$$

Repeated Trials – combined experiments

- The ‘section 1-1 Combined Experiments’ of the text book introduce the ‘Cartersian Products’(笛卡尔积)
- It is easy understanding. This notion tackles the problem of two independent experiments.

Repeated Trials – combined experiments

- For example
  - Experiment A is running a dice and you will get 1,2,3,4,5,6; there are six outcomes. (S_1 in page 11)
  - Experiment B is to flip a coin and you will get 'head' or 'tail'; there are two outcomes. (S_2 in page 11)
  - The independently running these two experiments, will give 6*2=12 outcomes. ($S = S_1 \times S_2$ in page 11)
  - If it is not a fair dice, i.e., the dice is filled with mercury(水银), and the probability of giving 1 is 95%, giving 2,3,4,5,6 are 1%, respectively.
  - Then, getting a result '3'($P(A_1) = 1\%$) and 'head'($P(A_2) = 50\%$) , which is '3 & head', is $P(A = A_1 \times A_2) = P(A_1)P(A_2) = 0.5\%$
- That's it, read the rest by yourself, knowing such notion is enough

- Assumed that there are repeated independent and identical (i.i.d, 独立同分布) experiments each of which has only two outcomes A or $\bar{A}$ with $P(A)=p$ and $P(\bar{A})=q=1-p$, the probability of exactly k occurrences of A in n such trials is given by:

$$P_n(k) = P(\text{"}A \; occurs\; exactly\, k \;\; times\, in\, n \;\; trials\text{"})$$

$$= \binom{n}{k} p^k q^{n-k}, \; k = 0,1,2,\cdots,n \qquad （2\text{-}2）$$

- Let $\quad X_k = \text{"}exactly\, k \;\; occurrences\; in\, n \;\; trials\text{"}.$

Since the number of occurrences of A in n trials must be an integer $k = 0,1,2,\ldots,n$, either $X_0 \; or \; X_1 \; or \; X_2 \; or \; \ldots or \; X_n$ must occur in such an experiment. Thus $P(X_0 \cup X_1 \cup \cdots \cup X_n) = 1$

- But X_i, X_j are mutually exclusive. Thus

$$P(X_0 \cup X_1 \cup \cdots \cup X_n) = \sum_{k=0}^{n} P(X_k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$$

From the relation

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

It gives $(p+q) = 1$ which is correct.

- For a given *n* and *p* what is the most likely value of *k* ? See Fig. 2.1, the most probable value of *k* is that number which maximizes $P_n(k)$ in (2-2).

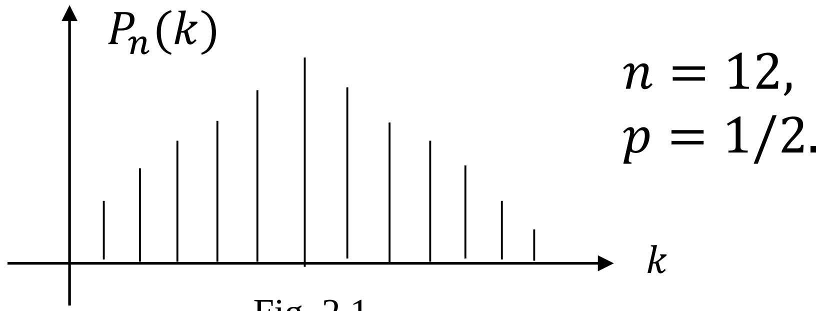


Fig. 2.1

- To obtain this value, consider the ratio

$$\frac{P_n(k-1)}{P_n(k)} = \frac{n!\,p^{k-1}q^{n-k+1}}{(n-k+1)!\,(k-1)!}\frac{(n-k)!\,k!}{n!\,p^k q^{n-k}} = \frac{k}{n-k+1}\frac{q}{p}$$

- Thus $P_n(k) \geq P_n(k-1)$ if $k(1-p) \leq (n-k+1)p$ or $k \leq (n+1)p$. Thus $P_n(k)$ as a function of k increases until

$$k = (n+1)p \tag{2-3}$$

Note that k might not be an integer, in this case, the largest integer k_{max} less than $(n+1)p$ will maximizes the $P_n(k)$

And (2-3) represents the most likely number of successes in n trials.

- From $k = (n+1)p$, k_{max}, the most likely number of successes in n trials, satisfies

$$(n+1)p - 1 \le k_{\max} \le (n+1)p \quad \text{or} \quad p - \frac{q}{n} \le \frac{k_{\max}}{n} \le p + \frac{p}{n}$$

- so that

$$\lim_{n\to\infty} \frac{k_m}{n} = p$$

  - As $n \to \infty$, the ratio of the most probable number of successes (A) to the total number of trials in a Bernoulli experiment tends to p, the probability of occurrence of A in a single trial.

- Example 2: Suppose 5,000 components are ordered. The probability that a part is defective equals to 0.1. What is the probability that the total number of defective parts does not exceed 400 ?

- Example 2: Suppose 5,000 components are ordered. The probability that a part is defective equals to 0.1. What is the probability that the total number of defective parts does not exceed 400 ?

- Solution: Let $Y_k =$ "k parts are defective among 5,000 components".

Then the desired probability is given by

$$P(Y_0 \cup Y_1 \cup \cdots \cup Y_{400}) = \sum_{k=0}^{400} P(Y_k) = \sum_{k=0}^{400} \binom{5000}{k} (0.1)^k (0.9)^{5000-k}$$

- There are too many terms to compute. Clearly, we need a technique to compute the above term in a more efficient manner.

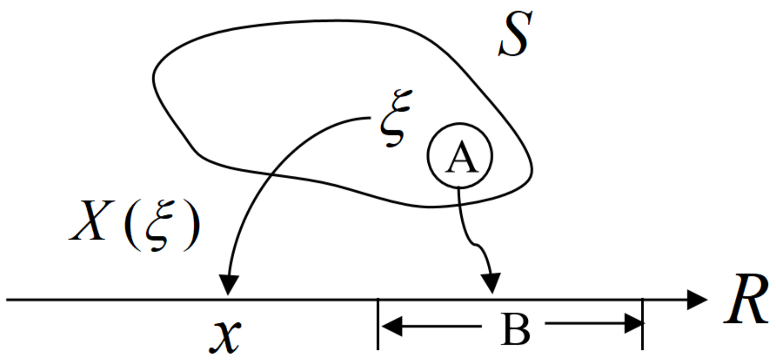
* Read **Bernoulli's theorem** (Page 22-29, text book) yourself, this will not be in the exam.

- Let S is the set, and X is a function that maps every even $\xi \in S$, to a unique point $x \in R$, the set of real numbers.

- Since the outcome ξ is not certain, so is the value $x \in R$.

- If B is some subset of R, we may want to determine the probability of "$X(\xi) \in B$".

- To determine this probability, we can look at the set $A = X^{-1}(B) \in S$ that contains all $\xi \in S$ that maps into B under the function X.



- Then: Probability of the event $X(\xi) \in B = P(X^{-1}(B))$.
- The notion(概念) of random variable (r.v) makes sure that the inverse mapping $X^{-1}(B)$ always results in an event so that we are able to determine the probability for any $B \in$ R:

- **Random Variable (r.v)**: A finite single valued function that maps the set of all experimental outcomes S into the set of real numbers R is said to be a r.v, if the set $\{\xi | X(\xi) \leq x\}$ is an event for every x in R.

- Alternatively, X is said to be a r.v, if $X^{-1}(B) \in F$ where B represents semi-definite intervals of the form $\{-\infty < x \le a\}$ and all other sets that can be constructed from these sets by performing the set operations of union(并), intersection(交) and negation(否定) any number of times. Thus if X is a r.v, then

$$\{\xi | X(\xi) \le x\} \text{ as } \{X \le x\}$$

- is an event for every x. What about $\{ a < X \le b \}$, $\{ X = a\}$? Are they also events? In fact with $b > a$ since $\{X \le a\}$ and $\{X \le b\}$ are events, therefore $\{X > a\} \cap \{X \le b\} = \{a < X \le b\}$ is also an event.

Note that $P\{X > a\} = 1 - P\{X \le a\}$

- Denote

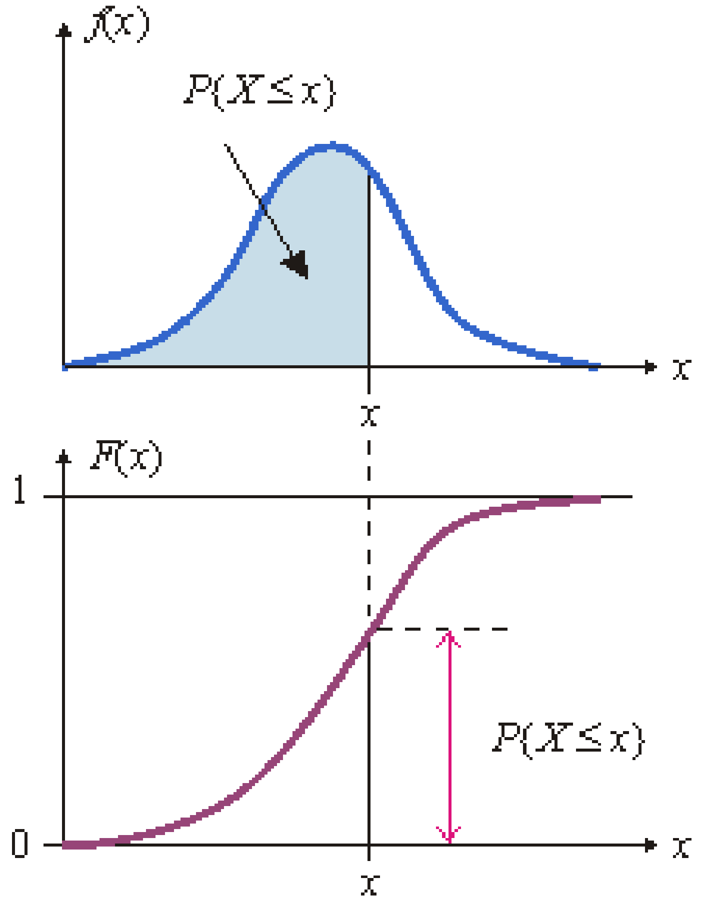
$$P\{\xi \mid X(\xi) \leq x\} = F_X(x) \geq 0$$

The role of the subscript X in $F_X(x)$ is only to identify the actual r.v.

$F_X(x)$ is said to the **Cumulative Distribution Function (CDF)** associated with the r.v X, usually, it is written as $F(x)$.

- The derivative of the $F_X(x)$ is called the **probability density function (pdf)** of the r.v X. Thus

$$f_X(x) = \frac{dF_X(x)}{dx}$$



- Properties of the CDF (!):
  - $F(x)$ is non-decreasing function
  - $F(+\infty) = P(S) = 1$（全集的概率为1）
  - $F(-\infty) = P(\emptyset) = 0$
  - $0 \leq F(x) \leq 1$
  - If $F(x_0) = 0$, then for any $x < x_0$, $F(x) = 0$
  - $P(X > x) = 1 - F(x)$
  - $P(X = x) = F(x) - F(x^-)$
  - $P(x_1 < X \leq x_2) = F(x_2) - F(x^-)$
  - $P(x_1 \leq X \leq x_2) = F(x_2) - F(x^-)$

- Example 3: Toss a coin. $S = \{H, T\}$, suppose the r.v X is such that $X(T) = 0, X(H) = 1$. Find $F_X(x)$

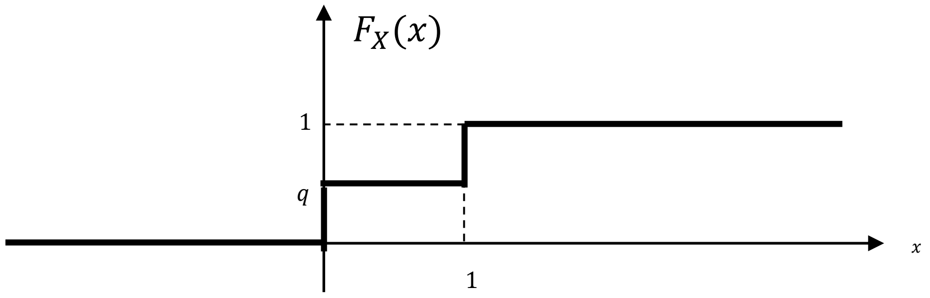
- Example 3: Toss a coin. $S = \{H, T\}$, suppose the r.v X is such that $X(T) = 0, X(H) = 1$. Find $F_X(x)$

Solution:

For $x < 0, \{X(\xi) \leq x\} = \{\phi\}$ so that $F_X(x) = 0$

For $0 \leq x < 1, \{X(\xi) \leq x\} = \{T\}$, so that $F_X(x) = P\{T\} = 1 - p$,

For $x \geq 1$, $\{X(\xi) \leq x\} = \{H, T\} = \mathrm{S}$, so that $F_X(x) = 1$. (see the Figure below)



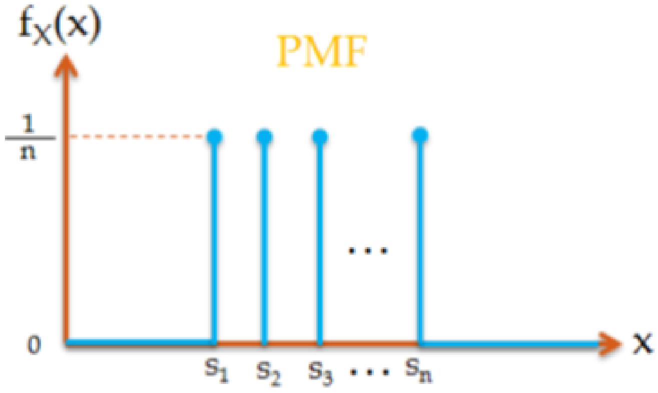
- X is said to be a continuous-type r.v if its distribution function $F_X(x)$ is continuous. In that case $F_X(x) = F_X(x^-)$ for all x, note that

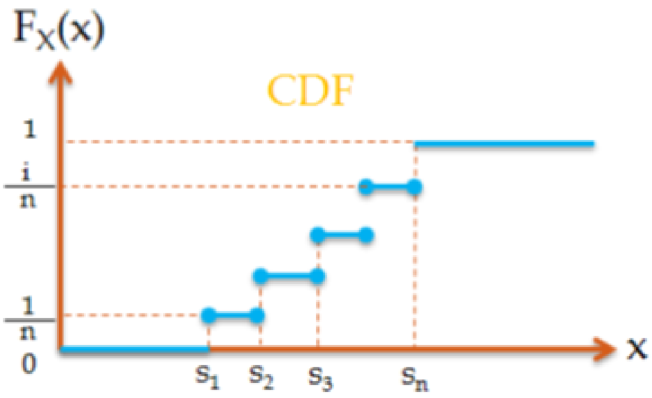
$$P\{X(\xi) = x_0\} = F_X(x_0) - F_X(x_0^-) > 0$$

- Therefore, for continuous-type r.v, $P\{X = x\} = 0$

- If $F_X(x)$ is constant except for a finite number of jump discontinuities (piece-wise constant; step-type), then X is said to be a discrete-type r.v. If x_i is such a discontinuity point, then

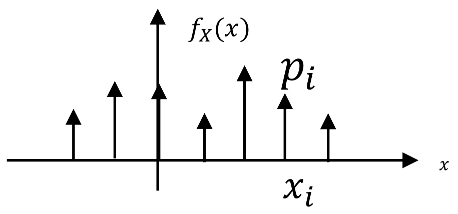
$$p_i = P\{X = x_i\} = F_X(x_i) - F_X(x_i^-)$$



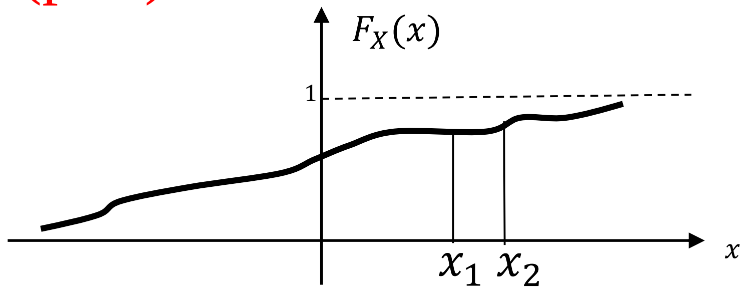


More about pdf:

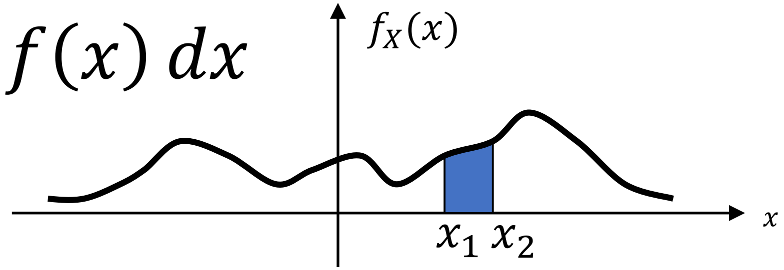
- Continuous type r.v. X: $f_X(x) = \dfrac{dF_X(x)}{dx}$
- discrete type r.v X: $f_X(x) = \sum p_i \delta(x - x_i)$
  - As the figure shows, $f_X(x)$ represents a collection of positive discrete masses, and it is known as the probability mass function (pmf) in the discrete case.



- Properties of pdf (!):
  - $f(x) \geq 0$
  - $\int_{-\infty}^{\infty} f(x)\,dx = 1$
  - $P(x_1 < X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\,dx$

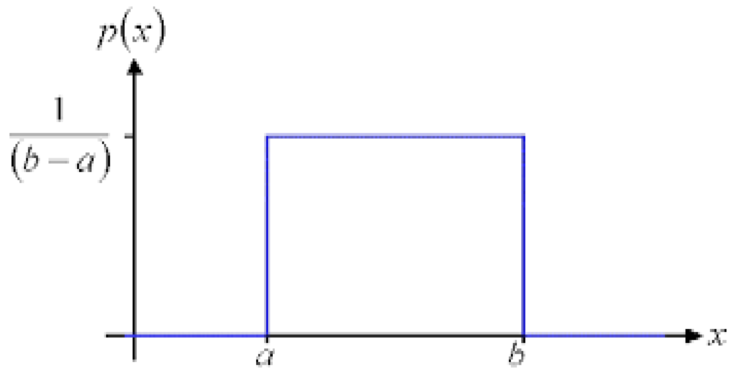


(a) CDF



(b) pdf

- Uniform distribution
  - pdf: $$f(x) = \begin{cases} 1/(b-a), & a < x < b \\ 0, & 其他 \end{cases}$$
  - CDF: $$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases}$$

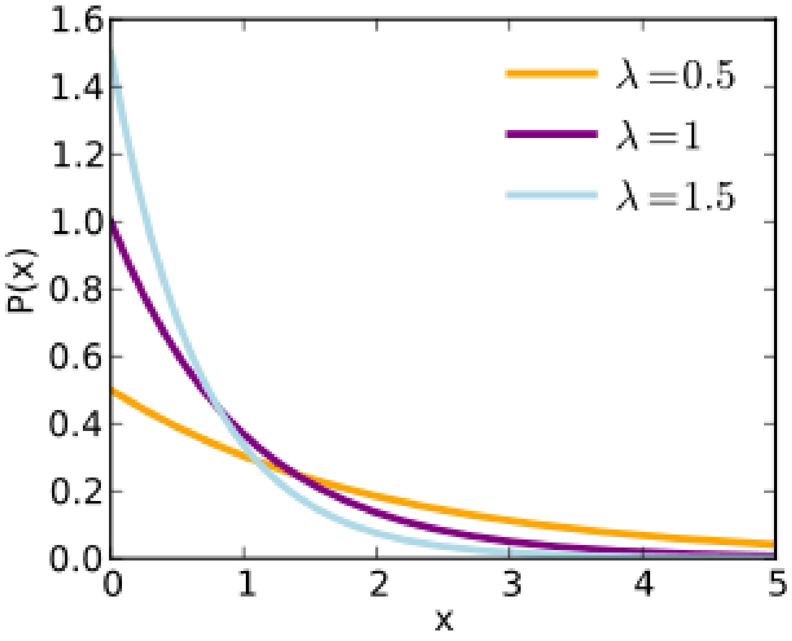


Example: Assume a=0, b=0.1, you can see that between 0 and 0.1, the value of PDF is 10
Which is, the value of PDF can exceed 1
But the value of CDF cannot exceed 1

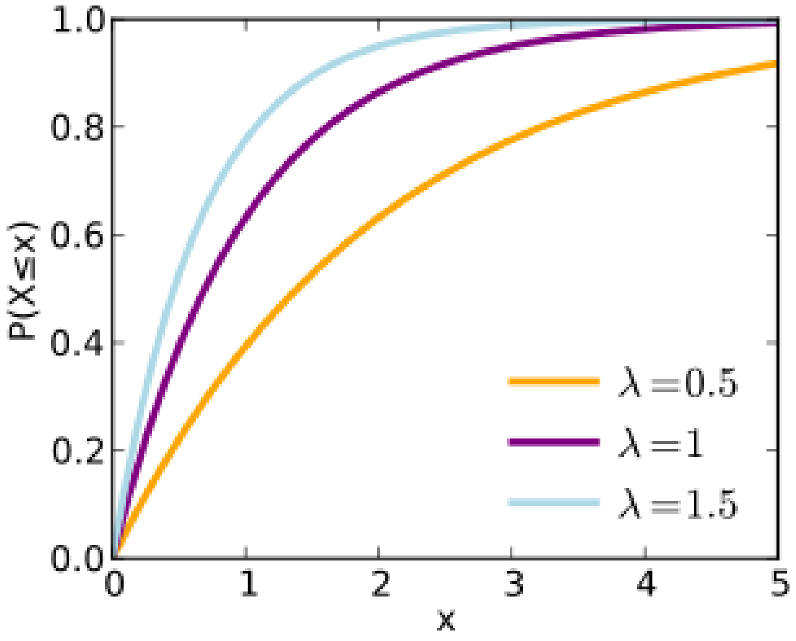
- Exponential Distribution: For a certain $\lambda > 0$, the pdf satisfies

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- Then x is called exponential r.v with parameter λ.



PDF

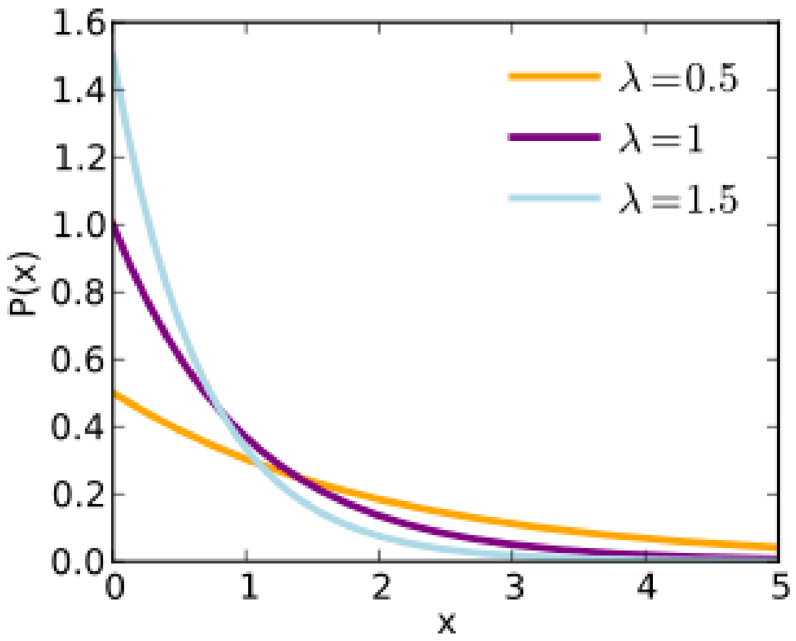


CDF

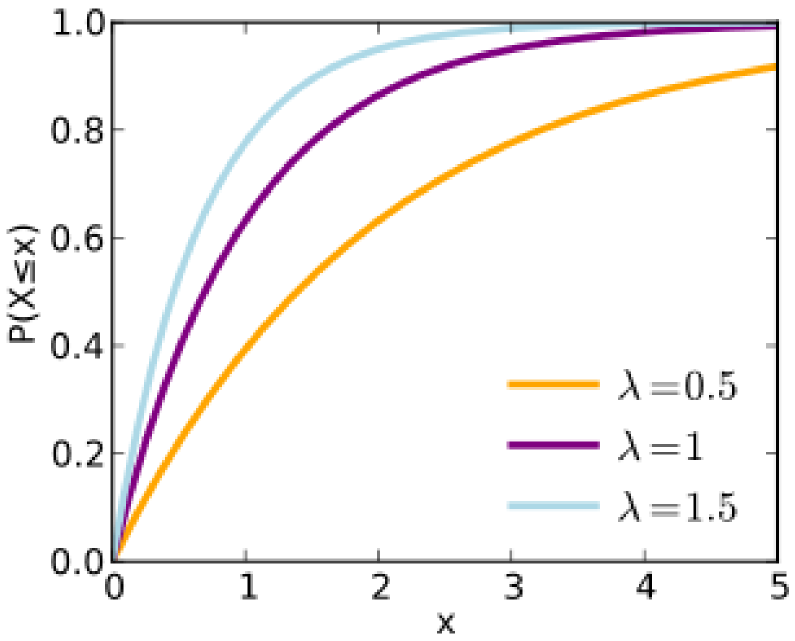
- Often used to model:
  - Arrival time of Poisson random process, component failure time,...

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$F(x) = 1 - e^{-\lambda x},\ x \geqslant 0$$



PDF



CDF

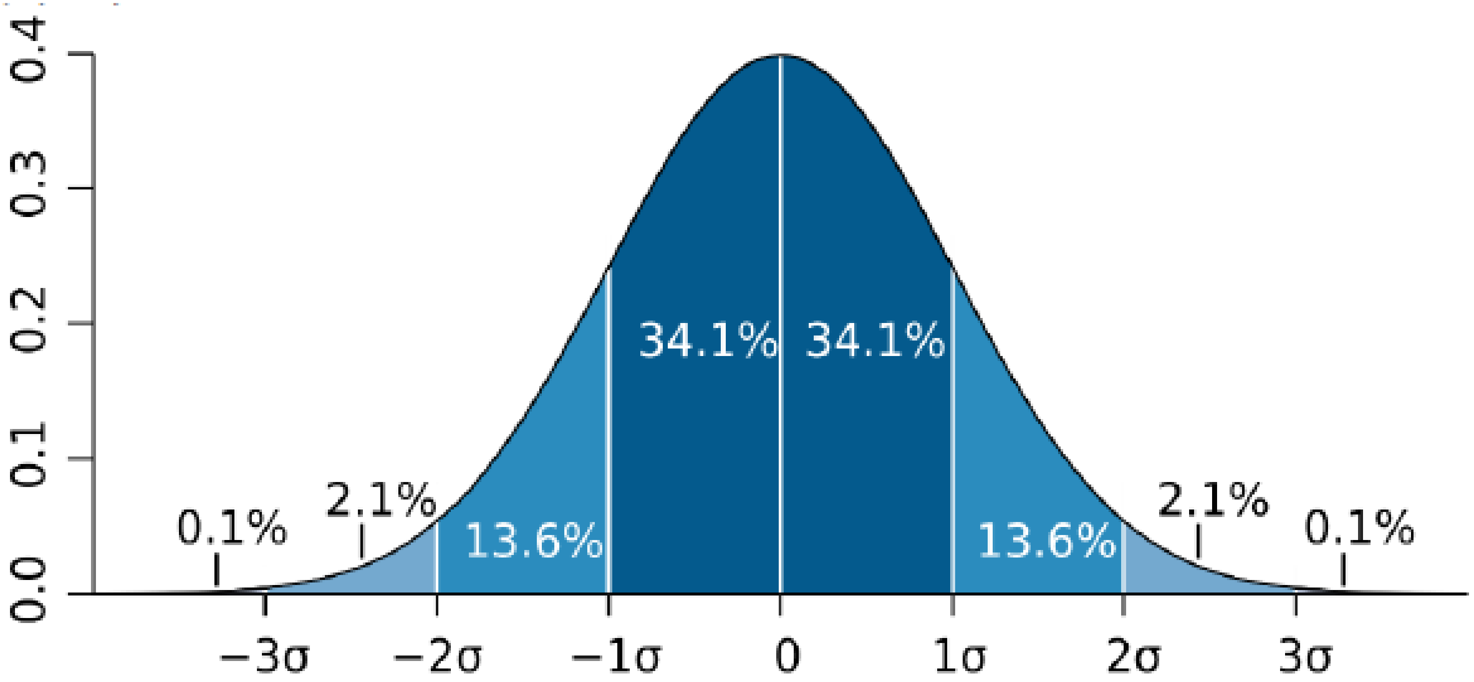
- Normal distribution:
  - A variable X is a normal random variable with parameters $\{\mu, \sigma\}$ if its probability density function satisfies:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

- Usually used to model:
  - electronic system noise
  - the physical characteristics of a large group of people
  - the distribution of student test scores
  - …

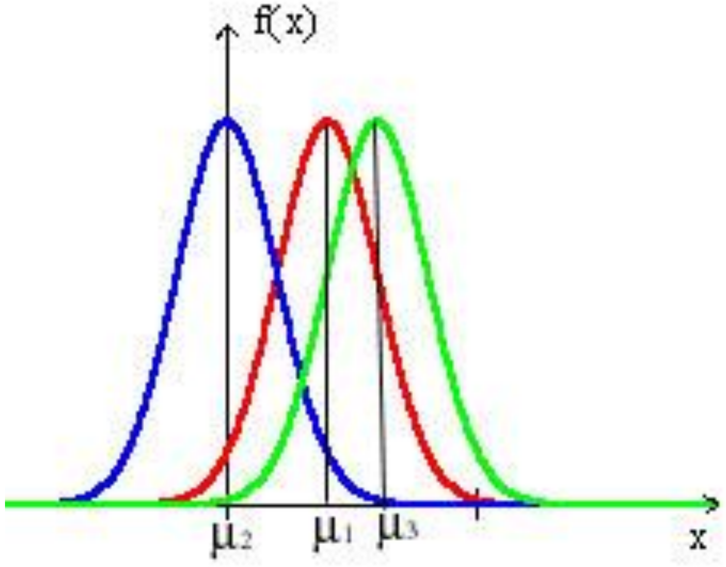
- The pdf:

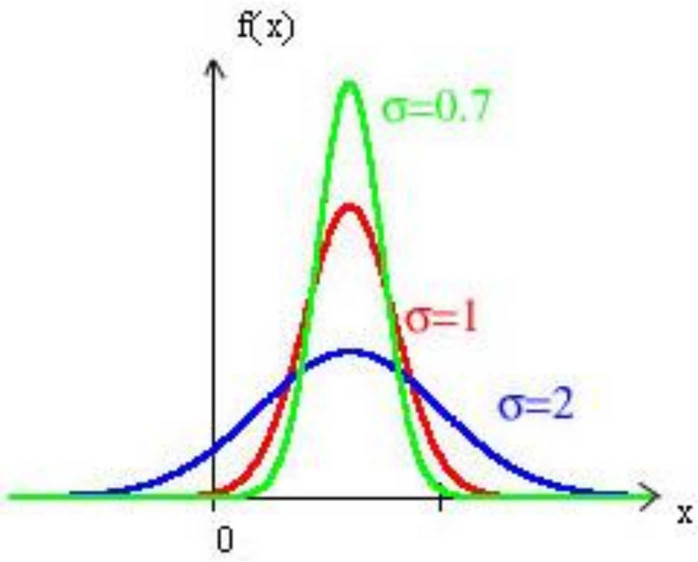
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$



- Results for different means μ and variances σ^2:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

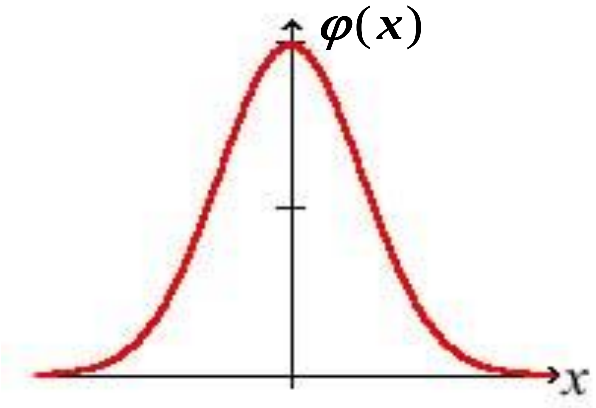


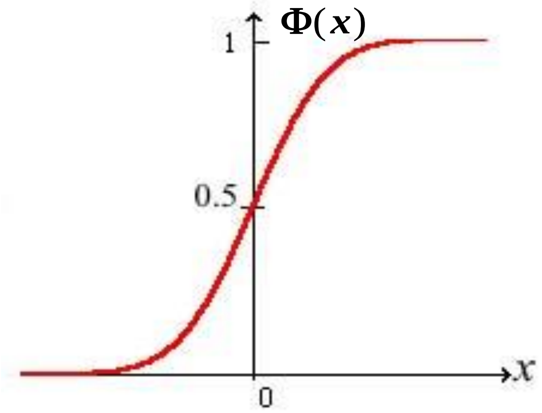


μ Determines the center position of the graph

σ Determines the steepness of the peaks in the graph

- Standard normal distribution: $\mu = 0, \sigma = 1$
  - Probability density function: $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < \infty$
  - Distribution function: $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, -\infty < x < \infty$
  - Theorem (normalization): if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$



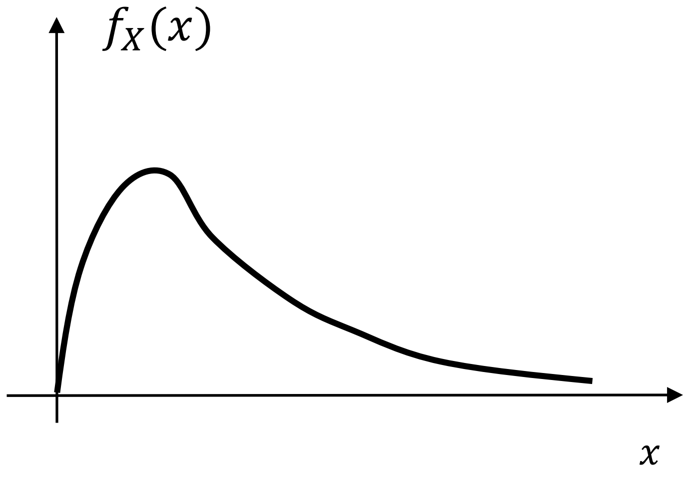


- Standard normal distribution: $\mu = 0, \sigma = 1$
  - Probability density function: $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < \infty$
  - Distribution function: $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, -\infty < x < \infty$
  - Theorem (normalization): if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
    - The importance of the standard normal distribution is that any general normal distribution can be transformed into a standard normal distribution by linear transformation.
    - So we use the standard normal distribution in many systems.
    - We often use $Q(x)=P(X>x)$ to express $P(X>x)$ probability, where X is a random variable that follows a standard normal distribution. For example, for the standard normal distribution definition $Q(2)=P(X>2)$, then, for the more general Gaussian distribution $Q(2)=P(X>2\sigma+\mu)$. And this $Q(2)$can usually be obtained by looking up the table.

Others (了解即可，需要用时会提供pdf和CDF):

- Chi-Square: $X \sim \chi^2(n)$

$$f_X(x) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

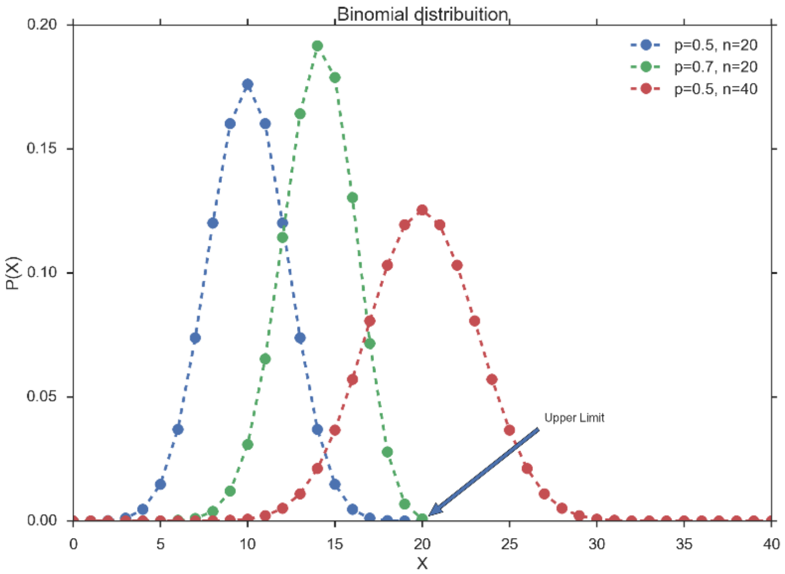


Usually used to model something related to detection theory.

- Bernoulli distribution
  - also known as (0,1) distribution
  - An experiment that results in success or failure:
    - Success → $X = 1, p(1) = P\{X = 1\} = p$
    - Fail → $X = 0, p(0) = P\{X = 0\} = 1 - p$
  - application:
    - Flipping a coin
    - Sending binary data in wireless communication

- Binomial distribution
  - Suppose n independent trials are made, where each successful outcome is p and the failed outcome is $(1-p)$
  - Each experiment is a Bernoulli distribution
  - If X represents the number of successes that appear in n trials, then X is called a binomial random variable with parameters (n,p), which is $X \sim B(n,p)$
  - Probability of success k times:

$$P(X=k)=\binom{n}{k}p^k q^{n-k}, k=0,1,2,\cdots,n$$



- Poisson distribution:
  - $\boldsymbol{X}$ is said to be a Poisson r.v with parameter $\lambda > 0$ if it takes values $0, 1, 2, \ldots, \infty$, with

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

- It is often used to model some application countable problems:
  - Typographical(印刷上的) errors in a book or one page of it
  - The number of users in a cell
  - …………

- Let X represent a Binomial r.v, then:

$$P(k_1 \leq X \leq k_2) = \sum_{k=k_1}^{k_2} P_n(k) = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k}$$

- Since the binomial coefficient $\binom{n}{k} = \frac{n!}{(n-k)!k!}$ grows rapidly with n, it is difficult to compute $P(k_1 \leq X \leq k_2)$ for large n. In this case, we can use Poisson Approximation.

- If, for example, $p \rightarrow 0$ and n$\rightarrow \infty$ such that $np = \lambda$ is a fixed number, we have $\lim_{n\rightarrow\infty, p\rightarrow 0, np=\lambda} P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}$

- Let X represent a Binomial r.v, then:

$$P(k_1 \leq X \leq k_2) = \sum_{k=k_1}^{k_2} P_n(k) = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k}$$

- If, for example, $p \to 0$ and n$\to \infty$ such that $np = \lambda$ is a fixed number, we have

$$\lim_{n\to\infty, p\to 0, np=\lambda} P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

- Poisson distribution:
  - $\boldsymbol{X}$ is said to be a Poisson r.v with parameter $\lambda > 0$ if it takes values $0, 1, 2, \ldots, \infty$, with

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

Random Variables – Poisson Approximation (证明过程，理解)

- Since $\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)$ as well as $\left(1-\frac{\lambda}{n}\right)^k$ tend to 1 as $n \to \infty$,and

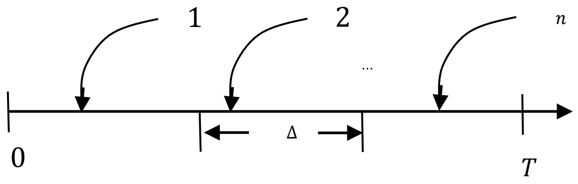
$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

$$P_n(k) = \frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{(np)^k}{k!}(1-np/n)^{n-k} = \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\frac{\lambda^k}{k!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}.$$

→
$$\lim_{n\to\infty, p\to 0, np=\lambda} P_n(k) = \frac{\lambda^k}{k!}e^{-\lambda}$$

- The right side of it represents the Poisson pmf and the Poisson approximation to the binomial r.v is valid in situations where the binomial r.v parameters *n* and *p* diverge to two extremes, such that their product *np* is a constant.

- Many random phenomena in nature in fact follow this pattern. Total number of calls on a telephone line, claims in an insurance company, etc. tend to follow this type of behavior.
- Consider random arrivals such as telephone calls over a line. Let n represent the total number of calls in the interval $(0, T)$. From our experience, as $T \to \infty$ we have $n \to \infty$.
- so that we may assume $n = \mu T$. Consider a small interval of duration Δ as below. If there is only a single call coming in, the probability p of that single call occurring in that interval must depend on its relative size with respect to T.



- Hence we may assume $p = \frac{\Delta}{T}$. Note that $p \to 0$ as $T \to \infty$. However in this case $np = \mu T \cdot \frac{\Delta}{T} = \mu\Delta = \lambda$ is a constant.
- Suppose the interval Δ is of interest to us. A call inside that interval is a "success" (H), whereas one outside is a "failure" (T). This is equivalent to the coin tossing situation, hence the probability $P_n(k)$ of obtaining k calls (in any order) in an interval of duration Δ is given by the binomial pmf. Thus

$$P_n(k) = \frac{n!}{(n-k)!\,k!} p^k (1-p)^{n-k}$$

- and here as $n \to \infty, p \to 0$ such that $np = \lambda$ It is easy to obtain an excellent approximation of $P_n(k)$ in this situation.

- Example 4 - Winning a Lottery: Suppose two million lottery tickets are issued with 100 winning tickets among them. (a) If a person purchases 100 tickets, what is the probability of winning? (b) How many tickets should one buy to be 95% confident of having a winning ticket?

- Example 4 - Winning a Lottery: Suppose two million lottery tickets are issued with 100 winning tickets among them. (a) If a person purchases 100 tickets, what is the probability of winning? (b) How many tickets should one buy to be 95% confident of having a winning ticket?

Solution: The probability of buying a winning ticket

$$p = \frac{\text{No. of winning tickets}}{\text{Total no. of tickets}} = \frac{100}{2 \times 10^6} = 5 \times 10^{-5}$$

Here $n = 100$ and the number of winning tickets X in the n purchased tickets has an approximate Poisson distribution with parameter $\lambda = np = 100 \times 5 \times 10^{-5}$ $= 0.005$. Thus

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

→(a) Probability of winning $= P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda} \approx 0.005$.

- (b) In this case we need a λ that $P(X \geq 1) \geq 0.95$.

$$P(X \geq 1) = 1 - e^{-\lambda} \geq 0.95 \quad \text{implies} \quad \lambda \geq \ln 2\,0 \approx 3$$

Which is, $\lambda = np = n \times 5 \times 10^{-5} \geq \ln 2\,0$, approximately, it is $n \geq 60{,}000$. Thus one needs to buy about 60,000 tickets to be 95% confident of having a winning ticket!

- For any two events A and B, we have defined the conditional probability of A given B as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) \neq 0.$$

- Note that the probability distribution function $F_X(x)$ is given by

$$F_X(x) = P\{\, X(\xi) \leq x \,\},$$

- we may define the conditional distribution of the r.v X given the event B as

$$F_X(x|B) = P\{\, X(\xi) \leq x|B \,\} = \frac{P\{\, (X(\xi) \leq x) \cap B \,\}}{P(B)} \qquad \text{(2-4)}$$

- Thus the definition of the conditional distribution depends on conditional probability, and since it obeys all probability axioms, it follows that the conditional distribution has the same properties as any distribution function. In particular

$$F_X(+\infty|B) = \frac{P\{\,(X(\xi) \le +\infty) \cap B\,\}}{P(B)} = \frac{P(B)}{P(B)} = 1,$$

$$F_X(-\infty|B) = \frac{P\{\,(X(\xi) \le -\infty) \cap B\,\}}{P(B)} = \frac{P(\emptyset)}{P(B)} = 0.$$

$$P(x_1 < X(\xi) \le x_2|B) = \frac{P\{\,(x_1 < X(\xi) \le x_2) \cap B\,\}}{P(B)} \qquad (2\text{-}5)$$

$$= F_X(x_2|B) - F_X(x_1|B),$$

- Since for $x_2 \geq x_1$,

$$(X(\xi) \leq x_2) = (X(\xi) \leq x_1) \cup (x_1 < X(\xi) \leq x_2)\,.$$

- The conditional density function is the derivative of the conditional distribution function. Thus

$$f_X(x|B) = \frac{dF_X(x|B)}{dx},$$

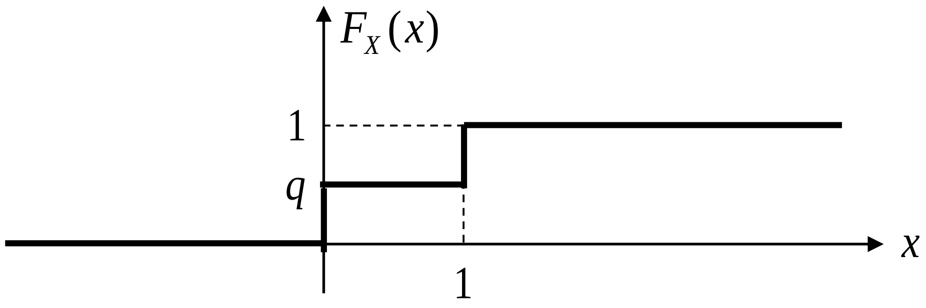
Or $$F_X(x|B) = \int_{-\infty}^{x} f_X(u|B)du.$$

- we can also rewrite (2-5) as

$$P(x_1 < X(\xi) \leq x_2|B) = \int_{x_1}^{x_2} f_X(x|B)dx.$$

- Example 5: Refer to example 3 of this ppt. Toss a coin and $X(T)=0$, $X(H)=1$. Suppose $B = \{H\}$. Determine $F_X(x|B)$.

Solution: From Example 3, $F_X(x)$ is like:

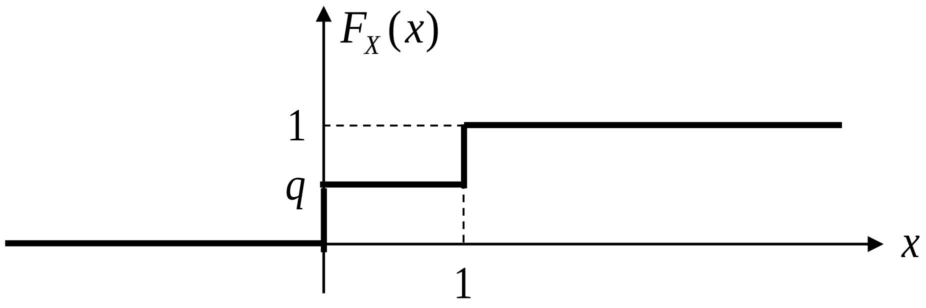


We need $F_X(x|B)$ for all x. (try to write it by yourself first)

Conditional Probability Density Function

- Example 5: Refer to example 3 of this ppt. Toss a coin and $X(T)$=0, $X(H)$=1. Suppose $B = \{H\}$. Determine $F_X(x|B)$.

Solution: From Example 3, $F_X(x)$ is like:



We need $F_X(x|B)$ for all x. (try to write it by yourself first)

- For $x < 0, \{X(\xi) \leq x\} = \emptyset,$ so that $\{\, (X(\xi) \leq x) \cap B\} = \emptyset,$ and $F_X(x|B) = 0$
- For $0 \leq x < 1$, $\{X(\xi) \leq x\} = \{\, T\},$ so that $\{\, (X(\xi) \leq x) \cap B\} = \{\, T\} \cap \{\, H\}$ $= \emptyset$, and $F_X(x|B) = 0$
- For $x \geq 1$, $\{X(\xi) \leq x\} = S$, and $\{\, (X(\xi) \leq x) \cap B\} = \Omega \cap \{\, B\} = \{B\}$, which is, $F_X(x|B) = \frac{P(B)}{P(B)} = 1$

- Example 5: Refer to example 3 of this ppt. Toss a coin and $X(T)=0$, $X(H)=1$. Suppose $B = \{H\}$. Determine $F_X(x|B)$.

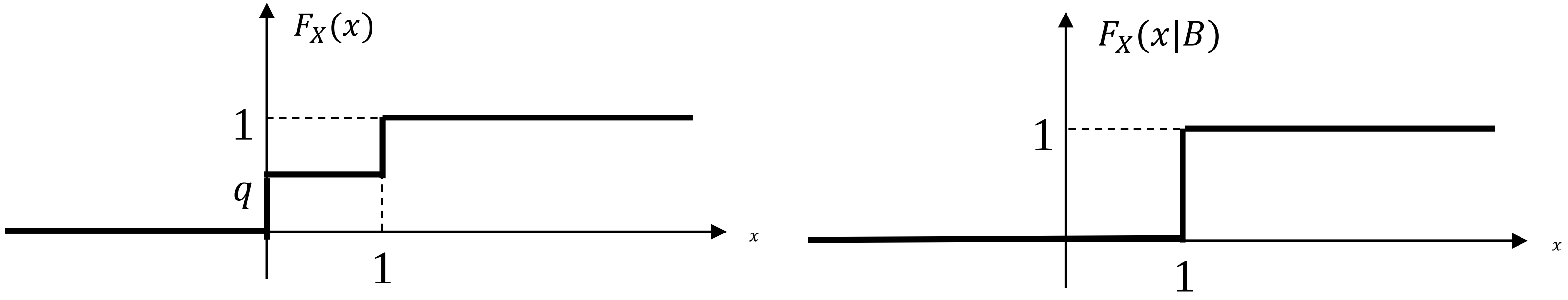
Solution:

$x < 0, \quad \{X(\xi) \leq x\} = \emptyset, \qquad F_X(x|B) = 0.$

$0 \leq x < 1, \quad \{X(\xi) \leq x\} = \{T\}, \qquad F_X(x|B) = 0.$

$x \geq 1, \quad \{X(\xi) \leq x\} = S, \qquad F_X(x|B) = \dfrac{P(B)}{P(B)} = 1$

- The physical meaning of this example is: given a result $B = \{H\}$, the result from the coin toss will always is $X(H)=1$, or says, HEAD.
- It seems meaningless in logic, but this is math



(a) (b)

Conditional Probability Density Function

- Example 6: Given $F_X(x)$, suppose $B = \{X(\xi) \leq a\}$. Find $f_X(x|B)$.

- Example 6: Given $F_X(x)$, suppose $B = \{X(\xi) \le a\}$. Find $f_X(x|B)$.

Solution: We will first determine $F_X(x|B)$. Note that

$$F_X(x \mid B) = P\{X(\xi) \le x \mid B\} = \frac{P\{(X(\xi) \le x) \cap B\}}{P(B)}$$

we have

$$F_X(x|B) = \frac{P\{\,(X \le x) \cap (X \le a)\,\}}{P(X \le a)}$$

- For $x < a$, $(X \le x) \cap (X \le a) = (X \le x)$ so that

$$F_X(x|B) = \frac{P(X \le x)}{P(X \le a)} = \frac{F_X(x)}{F_X(a)}.$$

- For $x \ge a$, $(X \le x) \cap (X \le a) = (X \le a)$ so that

$$F_X(x|B) = 1.$$

- Example 6: Given $F_X(x)$, suppose $B = \{X(\xi) \le a\}$. Find $f_X(x|B)$.

Thus

$$F_X(x|B) = \begin{cases} \dfrac{F_X(x)}{F_X(a)}, & x < a, \\ 1, & x \ge a, \end{cases}$$

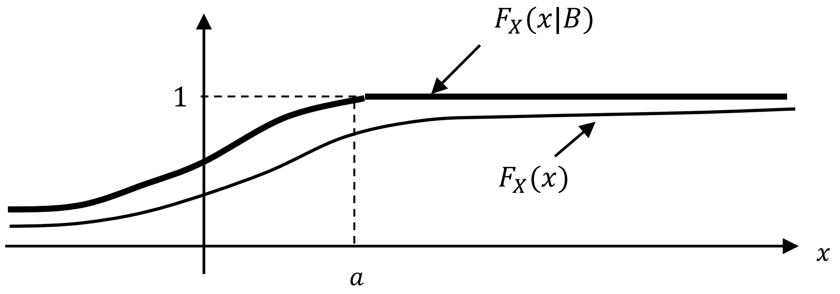
and hence

$$f_X(x|B) = \frac{d}{dx} F_X(x|B) = \begin{cases} \dfrac{f_X(x)}{F_X(a)}, & x < a, \\ 0, & \text{otherwise.} \end{cases}$$

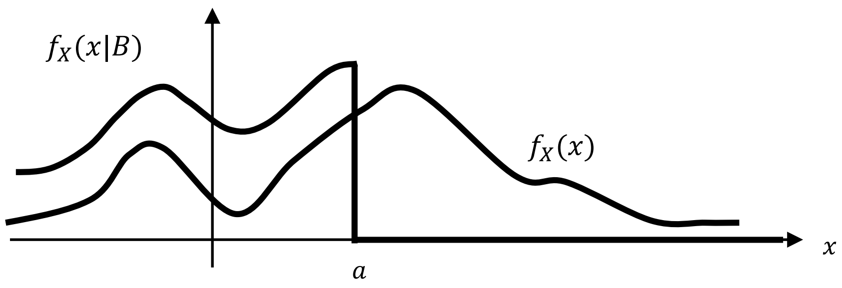
Note: to find a conditional pdf, you can always start with a conditional CDF

- Example 6: Given $F_X(x)$, suppose $B = \{X(\xi) \leq a\}$. Find $f_X(x|B)$.

The conditional CDF and pdf might look like this



(a)



(b)

- the shape of the conditional CDF and pdf are similar to the shape of original CDF and pdf, for$B = \{X(\xi) \leq a\}$.
- for $X(\xi) > a$, the CDF becomes 1 and pdf becomes 0

Conditional Probability Density Function

- We can use the conditional pdf together with the Bayes' theorem to update our a-priori knowledge about the probability of events in presence of new observations.

- Ideally, any new information should be used to update our knowledge. As we see in the next example, conditional pdf together with Bayes' theorem allow systematic updating. For any two events A and B, Bayes' theorem gives

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

- Let $B = \{x_1 < X(\xi) \leq x_2\}$ and it becomes

$$P\{A|(x_1 < X(\xi) \leq x_2)\} = \frac{P((x_1 < X(\xi) \leq x_2)|A)P(A)}{P(x_1 < X(\xi) \leq x_2)}$$

$$= \frac{F_X(x_2|A) - F_X(x_1|A)}{F_X(x_2) - F_X(x_1)} P(A) = \frac{\int_{x_1}^{x_2} f_X(x|A)dx}{\int_{x_1}^{x_2} f_X(x)dx} P(A).$$

Conditional Probability Density Function

- Further, let $x_1 = x, \quad x_2 = x + \varepsilon, \quad \varepsilon > 0$, so that in the limit as $\varepsilon \to 0$,

$$\lim_{\varepsilon \to 0} P\left\{A | (x < X(\xi) \le x + \varepsilon)\right\} = P(A | X(\xi) = x) = \frac{f_X(x|A)}{f_X(x)} P(A).$$

Or $$f_{X|A}(x|A) = \frac{P(A|X=x) f_X(x)}{P(A)}. \qquad \text{(2-6)}$$

- Put $P(A)$ to the left hand side, and take the integral:

$$P(A) \underbrace{\int_{-\infty}^{+\infty} f_X(x|A) dx}_{1} = \int_{-\infty}^{+\infty} P(A|X = x) f_X(x) dx,$$

Or $$P(A) = \int_{-\infty}^{+\infty} P(A|X = x) f_X(x) dx \qquad \text{(2-7)}$$

$$f_{X|A}(x|A) = \frac{P(A|X = x) f_X(x)}{\int_{-\infty}^{+\infty} P(A|X = x) f_X(x) dx}$$

Conditional Probability Density Function

$$f_{X|A}(x|A) = \frac{P(A|X=x)f_X(x)}{\int_{-\infty}^{+\infty} P(A|X=x)f_X(x)dx}$$

- 用简单版的金融量化分析举例：
  - $f_{X|A}(x|A)$：A发生情况下，x发生的概率。例如，GDP增长率为某个范围的情况下，某公司股价变化为x的概率。这个的计算目的在于预测。

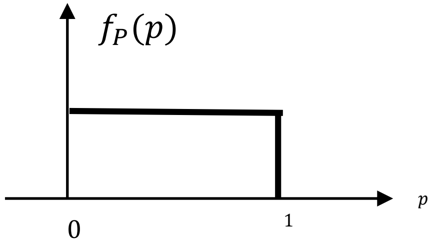
  而以下的所有数据均从历史数据中统计获取：

  - $P(A|X=x)$：股价变化为x时，GDP增长率为某个范围的概率。通过统计从过往N年，看当日股价变化为x时，看GDP增长率然后统计获得
  - $f_X(x)$：股价变化为x的概率；通过统计从过往N年中获得。

- Example 7: Let $p = P(H)$ represent the probability of obtaining a head in a toss. For a given coin (not a fair coin), a-priori(先验) p can possess any value in the interval (0,1). In the absence of any additional information, we may assume the a-priori pdf $f_P(p)$ to be a uniform distribution in interval (0,1). Now suppose we actually perform an experiment of tossing the coin n times, and k heads are observed. This is new information. How can we update $f_P(p)$?

Solution: Let A= "k heads in n specific tosses". Since these tosses result in a specific sequence,

$$P(A|P = p) = p^k q^{n-k},$$



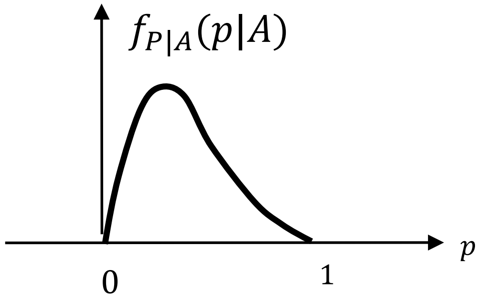
Conditional Probability Density Function

- and using (2-7) we get

$$P(A) = \int_0^1 P(A|P=p) f_P(p) dp = \int_0^1 p^k (1-p)^{n-k} dp = \frac{(n-k)!\,k!}{(n+1)!}.$$

- The a-posteriori(后验) pdf $f_{P|A}(p|A)$ represents the updated information given the event A, and from (2-6)

$$f_{P|A}(p|A) = \frac{P(A|P=p) f_P(p)}{P(A)} = \frac{(n+1)!}{(n-k)!k!} p^k q^{n-k}, 0 < p < 1 \quad (2\text{-}8)$$



Note that the a-posteriori pdf of p in (2-8) is not a uniform distribution, but a beta distribution ($\beta(n,k)$) (do not need to remember this distribution, just understand the basic idea of this example is OK). We can use this a-posteriori pdf to make further predictions

- For example, in the above experiment, what can we say about the probability of a head occurring in the next $(n+1)$th toss?

- Let B= "head occurring in the (n+1)th toss, given that k heads have occurred in n previous tosses".

- Clearly $P(B|P=p)=p,$ and from (2-7):

$$P(B)=\int_0^1 P(B|P=p)f_P(p|A)dp. \qquad (2\text{-}9)$$

- Notice that unlike (2-7), we have used the a-posteriori pdf in (2-9) to reflect our knowledge about the experiment already performed. write (2-8) in (2-9), we get

$$P(B)=\int_0^1 p\cdot\frac{(n+1)!}{(n-k)!\,k!}p^k q^{n-k}dp=\frac{k+1}{n+2}.$$

- Thus, if n =10, and k = 6, then

$$P(B)=\frac{7}{12}=0.58,$$

- which is more realistic compare to p = 0.5.

Conditional Probability Density Function

- To summarize, if the probability of an event X is unknown, one should make noncommittal judgement about its a-priori probability density function $f_X(x)$.
- Usually the uniform distribution is a reasonable assumption in the absence of any other information. Then experimental results (A) are obtained, and out knowledge about X must be updated reflecting this new information.
- Bayes' rule helps to obtain the a-posteriori pdf of X given A. From that point on, this a-posteriori pdf $f_{X|A}(x|A)$ should be used to make further predictions and calculations.
- This example will be in the experiment 1, understand it, and write down the correct program, then the experiment 1 will be very easy!

More Examples – will give solutions next week

1: Assume that the waiting time of a bus follows an $\lambda = 0.1$ exponential distribution. (Time: minutes) What is the probability that you will be able to get on the bus in 10 minutes?

2: Let B represent the event $\{ a < X(\xi) \leq b\}$ with $b > a$. For a given $F_X(x)$, determine $F_X(x|B)$and $f_X(x|B)$

Reading

- This week: Text book, Chapters 1-2
- Next week: Text book, Chapters 3-4