Probability and Random Signals

MATH

$$\sum_{k=0}^{n} k = \frac{n(n+1)}{2} \; ; \; \sum_{k=0}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\sum_{k=0}^{\infty} b^k = \frac{1}{1-b} \quad ; \text{for } |b|<1$$

$$\left(\sum_{i=1}^{n} a_i\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j$$

$$\underline{x}^T A \underline{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} [A]_{i,j} x_i x_j$$

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \approx 1 + x$$

$$e^{ix} = \cos(x) + i\sin(x)$$

$$\text{n!} = \int_0^{\infty} e^{-t} t^n dt$$

Leibniz's Rule

$$\frac{d}{dz}\left(\int_{a(z)}^{b(z)} f(x,z)dx\right) = b'(z)\cdot f(b(z),z) - a'(z)\cdot f(a(z),z) + \int_{a(z)}^{b(z)} \frac{\partial}{\partial z} f(x,z)dx$$

Function g(x) is **nonnegative definite** if

$$(\forall n)(\forall x_1 \ldots x_n)(\forall a_1 \ldots a_n)\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^* g(x_i - x_j) \ge 0$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,,\; \overline{g(x)} = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$$

Unit-impulse function δ or Dirac δ function

- δ(x) = $\frac{d}{dx}U(x)$
- U(x) = $\int_{-\infty}^{x} d(t)\,dt$
- $\int_{-\infty}^{\infty} g(t)\,d(t-a)\,dt$ = g(a)

$$\frac{d}{dx}\int_a^x f(t)dt = f(x)$$

$$\frac{d}{dx}\int_a^{v(x)} f(t)dt = \frac{dv}{dx}\cdot\frac{d}{dv}\int_a^{v(x)} f(t)dt = v'(x)\,f(v(x))$$

$$\frac{d}{dx}\int_{u(x)}^{v(x)} f(t)dt = \frac{d}{dx}\int_a^{v(x)} f(t)dt - \frac{d}{dx}\int_a^{u(x)} f(t)dt$$

$$= v'(x)\,f(v(x)) - u'(x)\,f(u(x))$$

SET

(with the major exception of the sample space)
an unordered collection of distinct elements/members

- ∈ ⇒ **set membership** or being an **element of the set**
  - undefined
- ∉ ⇒ **not an element of**

∅ = {} = empty/null set = (a,a)

cardinality = #elements in a set → ‖ ‖

- $\mathbf{N}_\mathbf{n}$ = {0, 1, 2, …, n-1}
- **Z** = {0, ±1, ±2, …}
- $\mathbf{Z}^+$ = {1, 2, 3, …}
- **N** = {0, 1, 2, 3, …}
- *R* = {x: -∞ < x < ∞}

- countably infinite set ⇒ has exactly as many members as there

are integers
- o Z, Z^+, N

- inclusion: A is a subset of B
  $A \subset B \leftrightarrow \{\forall w, w \in A \to w \in B\}$
  - $B \supset A$: B is a superset of A:
- Identity/Equality: A, B are equal
  $A = B \leftrightarrow \{A \subset B \wedge B \subset A\}$
- Reflexivity: $A \subset A$
- Transitivity: $A \subset B \wedge B \subset C \to A \subset C$

Boolean set operation
- complementation $A^C, A', \overline{A} = \{w : w \notin A\}$
- union $A \cup B = \{w : w \in A \vee w \in B\}$
- intersection $A \cap B = \{w : w \in A \wedge w \in B\}$
- difference $A - B = A \cap B^c = \{w : w \in A \wedge w \notin B\}$

event language
- A $\to$ A occurs
- $A^c \to$ A does not occur
- $A \cup B \to$ either A or B occur
- $A \cap B \to$ both A and B occur

- disjoint: $A \perp B \Leftrightarrow A \cap B = \varnothing$
- pair-wise disjoint, mutually exclusive:
  for $\{A_k\}$, $A_i \cap A_j = \varnothing$ when $i \neq j$

- indempotence: $(A^c)^c = A$
- commutativity (symmetry): $A \cup B = B \cup A$, $A \cap B = B \cap A$
- associativity:
  - o $A \cap (B \cap C) = (A \cap B) \cap C$
  - o $A \cup (B \cup C) = (A \cup B) \cup C$
- distributivity
  - o $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
  - o $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

EE310 By Prapun Suksompong

- **de Morgan laws**
  - o $(A \cup B)^c = A^c \cap B^c$
  - o $(A \cap B)^c = A^c \cup B^c$

Partition

$\Pi = \{B_i\}$ is a partition of Ω
- if and only if
  - $\Omega = \bigcup B_i$ and
  - $B_i \perp B_j$ when $i \neq j$
- $\to B_i \perp B_j \to A \cap B_i \perp A \cap B_j$
- $\to A = \bigcup_i (A \cap B_i)$
- $P(B_i) = 0 \to P(A|B_i)P(B_i) = 0$

The sequence of events $\{A_1, A_2, A_3, \ldots\}$ is monotone-increasing sequence of events if and only if
$A_1 \subset A_2 \subset A_3 \subset \ldots$
- $\bigcup_{i=1}^{n} A_i = A_n$
- $\lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i$

The sequence of events $\{B_1, B_2, B_3, \ldots\}$ is monotone-decreasing sequence of events if and only if
$B_1 \supset B_2 \supset B_3 \supset \ldots$
- $\bigcap_{i=1}^{n} B_i = B_n$
- $\lim_{n\to\infty} B_n = \bigcap_{i=1}^{\infty} B_i$

(event-) indicator function

$I_A : \Omega \to \{0,1\}$

$$I_A(w) = \begin{cases} 1, & \text{if } w \in A \\ 0, & \text{otherwise} \end{cases}$$

- $A = \{w : I_A(w) = 1\}$
- $A = B \Leftrightarrow I_A = I_B$
- $I_{A^c}(w) = 1 - I_A(w)$
- $A \subset B \Leftrightarrow \{\forall w, I_A(w) \le I_B(w)\} \Leftrightarrow \{\forall w, I_A(w) = 1 \to I_B(w) = 1\}$
- $I_{A\cap B}(w) = \min(I_A(w), I_B(w)) = I_A(w) \cdot I_B(w)$
- $I_{A\cup B}(w) = \max(I_A(w), I_B(w)) = I_A(w) + I_B(w) - I_A(w) \cdot I_B(w)$

Enumeration / Combinatorics / Counting

Given a set of n distinct items, select a distinct ordered sequence (word) of length r drawn from this set

- Sampling with replacement

  $\mu_{n,r} = n^r$
  - $\mu_{n,1} = n$
  - $\mu_{1,r} = 1$
  - $\mu_{n,r} = \mu_{n,r-1}$ for $r > 1$
  - $\|\Omega\| = r \to \|\text{power set}\| = \left\|2^\Omega\right\| = 2^{\|\Omega\|}$
  - Ex. #binary string of length r = 2^r
- Sampling without replacement

  $$(n)_r = \prod_{i=0}^{r-1}(n-i) = \frac{n!}{(n-r)!} = \underbrace{n\cdot(n-1)\cdots(n-(r-1))}_{\text{r terms}} \;;\; r \le n$$
  - $= 0$ if $r > n$
  - $(n)_1 = n$
  - $(n)_r = (n-(r-1))(n)_{r-1}$
    - Ex. $(7)_5 = (7-4)(7)_4$
  - $(1)_r = \begin{cases} 1 & \text{if } r = 1 \\ 0 & \text{if } r > 1 \end{cases}$

$$\frac{(n)_r}{n^r} = \frac{\prod_{i=0}^{r-1}(n-i)}{\prod_{i=0}^{r-1}(n)} = \prod_{i=0}^{r-1}\left(1-\frac{i}{n}\right) \approx \prod_{i=0}^{r-1}\left(e^{-\frac{i}{n}}\right) = e^{-\frac{1}{n}\sum_{i=0}^{r-1} i} = e^{-\frac{r(r-1)}{2n}}$$

$$\approx e^{-\frac{r^2}{2n}}$$

The number of arrangements (permutations) of n ≥ 0 distinct items $(n)_n = n!$

$0! = 1! = 1$

$n! = n(n-1)!$

$n! = \int_0^\infty e^{-t} t^n dt$

Stirling's Formula: $n! \approx \sqrt{2p\, n}\, n^n e^{-n} = \left(\sqrt{2p\, e}\right) e^{\left(n+\frac{1}{2}\right)\ln\left(\frac{n}{e}\right)}$

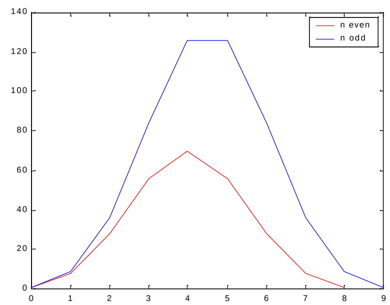
binomial coefficient

$$\binom{n}{r} = \frac{(n)_r}{r!} = \frac{n!}{(n-r)!\,r!}$$

- the number of unordered sets of size r drawn from an alphabet of size n without replacement
- the number of subsets of size r that can be formed from a set of n elements

- reflection property: $\binom{n}{r} = \binom{n}{n-r}$
- $\binom{n}{n} = \binom{n}{0} = 1$
- $\binom{n}{1} = \binom{n}{n-1} = n$
- $\binom{n}{r} = 0$ if $n < r$

- $\max_r \binom{n}{r} = \binom{n}{\left\lfloor \frac{n+1}{2} \right\rfloor}$



Binomial theorem

$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}$

- let x = y = 1 $\rightarrow \sum_{r=0}^{n} \binom{n}{r} = 2^n$

Entropy function

H(p) = $-p\log_b(p) - (1-p)\log_b(1-p)$

- binary: b = 2

  $H_2(p) = -p\log_2(p) - (1-p)\log_2(1-p)$

- $\binom{n}{r} \approx 2^{nH_2\left(\frac{r}{n}\right)}$
- $\frac{1}{n+1} 2^{nH\left(\frac{r}{n}\right)} \le \binom{n}{r} \le 2^{nH\left(\frac{r}{n}\right)}$

Multinomial Counting

$$\binom{n}{n_1 n_2 \ldots n_r} = \prod_{i=1}^{r} \binom{n - \sum_{k=0}^{i-1} n_k}{n_i}$$

$$= \binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \binom{n-n_1-n_2}{n_3} \cdots \binom{n_r}{n_r} = \frac{n!}{\prod_{i=1}^{r} n!}$$

- Arrange n = $\sum_{i=1}^{r} n_i$ tokens when having r types of symbols and n_i indistinguishable copies/tokens of a type i symbol
- multinomial coefficient

Multinomial Theorem

$$(x_1 + \ldots + x_r)^n = \sum_{i_1=0}^{n} \sum_{i_2=0}^{n-i_1} \cdots \sum_{i_{r-1}=0}^{n-\sum_{j<r-1} i_j} \frac{n!}{\left(n - \sum_{k<n} i_k\right)! \prod_{k<n} i_k!} x_r^{\,n-\sum_{j<r} i_j} \prod_{k=1}^{r-1} x_k^{\,i_k}$$

r-ary entropy function

- $p_i \ge 0$
- $\sum_{i=1}^{r} p_i = 1$

$$H(\underline{p}) = -\sum_{i=1}^{r} p_i \log_b p_i$$

Let $p_i = \frac{n_i}{n}$

- $\sum_{i=1}^{r} p_i = \sum_{i=1}^{r} \frac{n_i}{n} = \frac{1}{n} \sum_{i=1}^{r} n_i = 1$

$$\frac{n!}{\prod_{i=1}^{r} n_i!} \approx 2^{nH_2(\underline{p})}$$

bars and stars

- Ex. distribution of r = 10 balls into n = 5 cells
- ****|***||**|* $\Rightarrow$ (4,3,0,2,1)
- n-1 bars ; r stars
- #{distinguishable arrangements} = $\binom{n-1+r}{r} = \binom{n-1+r}{n-1}$

CLASSICAL PROBABILITY

random experiment (Ω , $\mathcal{A}$,P)

Equipossibility

- The bases for indentifying equipossibility were
  - physical symmetry
  - balance of information
- meaningful only for finite sample space

selected at random $\Rightarrow$ equipossible cases

The classical probability of an event A

P(A) = $\dfrac{\|A\|}{\|\Omega\|}$

$= \dfrac{\text{The number of cases favorable to the outcome of the event}}{\text{The total number of possible case}}$

- $P(A) \geq 0$
- $P(\Omega) = 1$
- $P(\varnothing) = 0$
- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - $\|A \cup B\| = \|A\| + \|B\| - \|A \cap B\|$
- $A \perp B \rightarrow P(A \cup B) = P(A) + P(B)$
- $P(A^c) = 1 - P(A)$
- $\Omega = \{\omega_1, \ldots, \omega_n\}$, $P(\{\omega_i\}) = \dfrac{1}{n} \rightarrow P(A) = \sum_{w \in A} p(w)$
  - The probability of an event is equal to the sum of the probabilities of its component outcomes because

outcomes are mutually exclusive

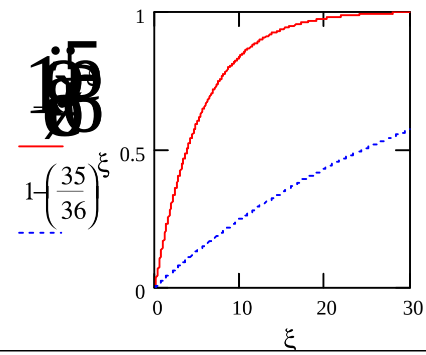
Chevalier de Mere's Scandal of Arithmetic

A = obtaining at least one six in four tosses of a fair die

$P(A) = 1-\left(\frac{5}{6}\right)^4 = .518$

B = obtaining at least double-six in 24 tosses of a pair dice

$P(B) = 1-\left(\frac{35}{36}\right)^{24} = .491$

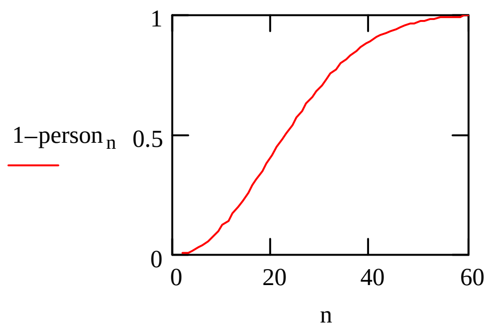


Probability of coincidence birthday

Probability that two people in your class have the same birthday

= 1 if n $\geq$ 365

$= 1-\left(\underbrace{\frac{365}{365}\cdot\frac{364}{365}\cdot\cdots\cdot\frac{365-(n-1)}{365}}_{\text{n terms}}\right)$ if $0 \leq n \leq 365$



Conditional Probability

Classical Conditional Probability

- Given that event B ≠ ∅ occurred
- Conditional Probability of A given B:

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

= the updated probability of the event A given that we now know that B occurred.

- $P(A|B) = P(A \cap B|B) \geq 0$
- $P(\Omega|B) = P(B|B) = P(A_{\supset B}|B) = 1$
- If $A \perp C$,

EE310 By Prapun Suksompong

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

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

$$;\{A \perp C \to (A \cap B) \perp (A \cap C)\}$$

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

- $P(A \cap B) = P(B)P(A|B)$
- $P(A \cap B) \leq P(A|B)$

Multiplication/Sequence Theorem :

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)\prod_{i=2}^{n} P\left(A_i \middle| \bigcap_{j<i} A_j\right)$$

n =2 ⟶ definition of conditional probability

Let B = $\bigcap_{i=1}^{n} A_i$

$$P\left(\bigcap_{i=1}^{n+1} A_i\right) = P(B \cap A_{n+1}) = P(A_{n+1}|B)P(B)$$

$$= P(A_{n+1}|B)P(A_1)\prod_{i=2}^{n} P\left(A_i \middle| \bigcap_{j<i} A_j\right)$$

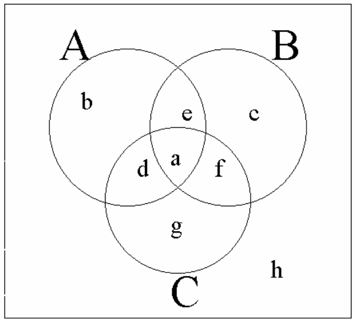
$$= P(A_1)\left\{P\left(A_{n+1} \middle| \bigcap_{i=1}^{n} A_i\right)\prod_{i=2}^{n} P\left(A_i \middle| \bigcap_{j<i} A_j\right)\right\}$$

$$= P(A_1)\prod_{i=2}^{n+1} P\left(A_i \middle| \bigcap_{j<i} A_j\right)$$

- $P(A \cap B \cap C)$
  $= P((A \cap B) \cap C) = P(A \cap B) \times P(C|A \cap B)$
  $= P(A) \times P(B|A) \times P(C|A \cap B)$

PM Page

- $P(A\cap B) = P(A)\times P(B|A) = \frac{a+b+d+e}{n}\cdot\frac{e+a}{a+b+d+e} = \frac{e+a}{n}$
- $P(A\cap B\cap C) = P(A\cap B)\times P(C|A\cap B) = \frac{a+e}{n}\cdot\frac{a}{a+e} = \frac{a}{n}$



Markov process

- a causal chain in which event A_i is produced solely by its temporal predecessor A_{i-1}
- $P\left(A_i \middle| \bigcap_{j<i} A_j\right) = P(A_i|A_{i-1})$; $\forall i > 1$

Total Probability Theorem

Assumption:

- $\Pi = \{B_i\}$ is a partition of Ω
- $P(B_i) = 0 \rightarrow P(A|B_i)P(B_i) = 0$

Result:

- $P(A) = \sum_i P(A|B_i)P(B_i)$
- Urn models: n urns $\Rightarrow$ partition of Ω
  each urn U_i has $a_{1i} + a_{2i} + a_{3i} + \ldots + a_{pi} = n_i$
  - $P(U_i) = \frac{1}{n}$ if equally possible
  - $P(a_{ji}|U_i) = \frac{a_{ji}}{n_i}$

EE310 By Prapun Suksompong

- $P(a_j) = \sum_{i=1}^{p} \frac{a_{ji}}{n_i}\cdot P(U_i)$

Bayes Theorem

Assumption:

- $P(E_j) > 0$
- C_i is a partition of Ω
- $P(C_i) = 0 \rightarrow P(E_j|C_i)P(C_i) = P(E_j\cap C_i) = 0$

Results:

- $P(C_i|E_j) = \frac{P(E_j\cap C_i)}{P(E_j)} = \frac{P(E_j|C_i)P(C_i)}{\sum_k \{P(E_j|C_k)P(C_k)\}}$

- In this form, know $P(\text{output}_j|\text{input}_i)$ and $P(\text{input}_i)$.
  - find $P(\text{output}_j)$ by total probability theorem
  - find $P(\text{input}_i|\text{output}_j)$ by Bayes theorem

Consider the following table

$C_i \xrightarrow{S} E_j$

$\|E\| = \|C\| = \|E_1\| + \|E_2\| = \|C_1\| + \|C_2\|$

	E_1	E_2	
C_1	$\|C_1\cap E_1\|$	$\|C_1\cap E_2\|$	$\|C_1\|$
C_2	$\|C_2\cap E_1\|$	$\|C_2\cap E_2\|$	$\|C_2\|$
	$\|E_1\|$	$\|E_2\|$	$\|E\| = \|C\|$

Notice that

$$P(E_j) = \frac{\|E_j\|}{\|E\|}$$

$$P(E_j|C_i) = \frac{P(E_j\cap C_i)}{P(C_i)} = \frac{\|E_j\cap C_i\|}{\|C_i\|}$$

$$P(C_j) = \frac{\|C_j\|}{\|C\|}$$

$$P(E_j)=\sum_k \{P(E_j\cap C_k)\}=\sum_k \{P(E_j\mid C_k)P(C_k)\}$$

$$\sum_k \{P(E_j\mid C_k)P(C_k)\}=\sum_k\left(\frac{\|E_j\cap C_k\|}{\|C_k\|}\frac{\|C_k\|}{\|C\|}\right)$$

$$=\sum_k\left(\frac{\|E_j\cap C_k\|}{\|C\|}\right)$$

$$=\frac{1}{\|C\|}\sum_k\left(\|E_j\cap C_k\|\right)$$

$$=\frac{\|E_j\|}{\|C\|}=\frac{\|E_j\|}{\|E\|}$$

$$P(C_i\mid E_j)=\frac{P(C_i\cap E_j)}{P(E_j)}=\frac{\|C_i\cap E_j\|}{\|E_j\|}$$

$$=\frac{P(E_j\mid C_i)P(C_i)}{P(E_j)}=\frac{P(E_j\mid C_i)P(C_i)}{\sum_k\{P(E_j\mid C_k)P(C_k)\}}$$

Monte Hall's Game

- Started with showing a contestant 3 closed doors behind of which was a prize
- The contestant selected a door
- but before the door was opened, Monte Hall, who knew which door hid the prize, opened a remaining door.
- The contestant was then allowed to either stay with his original guess or change to the other closed door.
- Question: better to stay or to switch

R = right door ; S = switch door

We will find P(R|S) ;

Case 1: C_1: P(R) = $\frac{1}{3}$, P(R|S) = 0

Case 2: C_2: P(R^c) = $\frac{2}{3}$, P(R|S) = 1

$\therefore$ P(R|S) = $\frac{1}{3}(0) + \frac{2}{3}(1) = \frac{2}{3}$

False positives on Diagnostic Tests

- D = Have a disease
- + = positive test
- p_D = probability of having disease
- P(+|D) = 1 $\Rightarrow$
  - P($+^c$|D) = 0
  - have disease $\rightarrow$ always positive result
- P(+|D^C) = p_+ = not have disease $\rightarrow$ positive result

	D	D^c	
+	$P(+\cap D)$ $= P(+\mid D)P(D)$ $= 1(p_D) = p_D$	$P(+\cap D^c)$ $= P(+\mid D^c)P(D^c)$ $= p_+(1-p_D)$	$P(+)$ $= P(+\cap D) + P(+\cap D^c)$ $= p_D + p_+(1-p_D)$
$+^c$	$P(+^c\cap D)$ $= P(+^c\mid D)P(D)$ $= 0(p_D) = 0$	$P(+^c\cap D^c)$ $= P(+^c\mid D^c)P(D^c)$ $= (1-p_+)(1-p_D)$	$P(+^c)$ $= P(+^c\cap D) + P(+^c\cap D^c)$ $= (1-p_+)(1-p_D)$
	$P(D)$ $= P(+\cap D) + P(+^c\cap D)$ $= p_D$	$P(D^c)$ $= P(+\cap D^c) + P(+^c\cap D^c)$ $= 1 - p_D$	$P(\Omega) = 1$

$$P(D\mid +)=\frac{P(D\cap +)}{P(+)}=\frac{p_D}{p_D+p_+(1-p_D)}=\frac{1}{1+\frac{p_+}{p_D}(1-p_D)}$$

$$\approx\frac{P_D}{P_+}\quad ;\text{for rare disease } (p_D \ll 1)$$

Probability Foundations

propensity view of probability

The probability P(A) of the event A is a numerical measure of the propensity or tendency of the event A to occur in a (not necessarily repeatable) performance of a random experiment $\mathcal{E}$

- For indefinitely repeatable experiments,
  probabilistic propensity is displayed
  in the long-run relative frequency of the occurrence of A
  in n repeated, unlinked performances $\mathcal{E}_{1,\ldots,}\mathcal{E}_n$ of the random experiment $\mathcal{E}$

Sample point
A representation of a possible outcome of an experiment.

Sample space : Ω
- set of all possible outcomes of the experiment
- do so without duplication
- at a level of detail sufficient for our interests
- this list is complete in a practical sense, albeit usually not complete either regarding all logically or physically possible outcomes

Event $\Rightarrow$ an outcome or a collection of outcomes.

(Event) Algebra / Field $\mathcal{A}$

① a collection/class of subsets of Ω
② closed under
  - 1) complementation
  - 2) finite union

- **σ-field/algebra** if closed under complementation and countable union.

- $\Omega, \varnothing \in \mathcal{A}$
- closed under
  - finite intersection
  - difference

- The smallest $\mathcal{A} = \{\Omega, \varnothing\}$
- The largest $\mathcal{A} = 2^{\Omega}$
- Why not all subsets?
  - If Ω is finite, too large
  - If Ω is not finite, may not be able to assign a probability to all possible subsets.

EE310 By Prapun Suksompong

- Intersections of Algebra is an algebra

- **Relative frequency** of an event A

$$r_n(A) = \frac{1}{n}\sum_{1}^{n} I_A(w_i) = \frac{N_n(A)}{n}$$

RF0 r_n: $\mathcal{A} \to$

RF1 $r_n(A) \ge 0$

RF2 $r_n(\Omega) = 1$

RF3 $A \perp B \Rightarrow r_n(A \cup B) = r_n(A) + r_n(B)$

Kolmogorov's Axioms for probability

K0 Setup
P: $\mathcal{A} \to \Re$

K1 Nonnegativity
$P(A) \ge 0$

K2 Unit normalization
$P(\Omega) = 1$

K3 Finite additivity
$A \perp B \Rightarrow P(A \cup B) = P(A) + P(B)$

- The above axioms suffice when sample space has finite number of sample points

K4 Monotone continuity
$A_{i+1} \subset A_i$ and $\bigcap_i A_i = f \Rightarrow \lim_{i\to\infty} P(A_i) = 0$

K4′ Countable or σ-additivity
$\{i \ne j \Rightarrow A_i \perp A_j\} \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

- K4, K4′ Equivalence:
  If P satisfies K0-K3, then
  it satisfies K4' if and only if it satisfies K4

probability measure $\Leftrightarrow$ satisfies K0 – K4

- $P(A^c) = 1 - P(A)$
- $P(\varnothing) = 0$
- $0 \le P(A) \le 1$
- If $P(A) = 1$, A is not necessary Ω.
- $A \supset B \Rightarrow P(A) \ge P(B)$
- $P(A \cup B) \ge \max(P(A),P(B)) \ge \min(P(A), P(B)) \ge P(A \cap B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Finite Disjoint Unions:

  $$i \ne j,\ A_i \perp A_j \Rightarrow P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$
- Countable Disjoint Unions:

  $$i \ne j,\ A_i \perp A_j \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$
- If $\Pi = \{A_i\}$ = countable partition of Ω, then

  $P(B) = \sum_i P(B \cap A_i)$
- Boole's Inequality:

  $$P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$$
- Inclusion-Exclusion Principle

  $$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{f \ne I \subset \{1,\ldots,n\}} (-1)^{\|I\|+1} P\left(\bigcap_{i \in I} A_i\right)$$

  $$\begin{aligned} P(A_1 \cup A_2 \cup A_3) = {} & P(A_1) + P(A_2) + P(A_3) \\ & - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\ & + P(A_1 \cap A_2 \cap A_3) \end{aligned}$$

- Given a common event algebra $\mathcal{A}$, probability measures $P_1,\ldots,P_m$, and the numbers $\lambda_1,\ldots,\lambda_m$, $\lambda_i \ge 0$, $\sum_1^m l_i = 1$

  a convex combination $P(A) = \sum_1^m l_i P_i(A)$ of probability measures $\{P_i\}$ is a probability measure

A formal definition of probability involves the specification of

- Ω
- $\mathcal{A}$
- A set function P which is defined on the elements of $\mathcal{A}$ and which has the properties specified by Kolmogorov's Axioms

- A probability assignment which is consistent with Kolmogorov's Axioms can always be made to the elements of a σ-algebra.

Probability space $\Rightarrow (\Omega, \mathcal{A}, P)$

PMF: Probability Mass Function

pdf for finite or countably infinite $\Omega = \{\omega_I\}$

pmf

① $p:\Omega \to [0,1]$

② $\sum_{w \in \Omega} p(w) = 1$

when enumerated, $p(\omega_I) = p_i$

- $p(\omega) = P(\{\omega\})$
- $P(A) = \sum_{w \in A} p(w)$

- convex combination of pmfs $\sum_i l_i p^{(i)}(w)$ is a pmf

- **Random** $\mathcal{R}_n$: $p_i = \frac{1}{n}$ for $\Omega = N_n$
  - classical game of chance / classical probability drawing at random
  - fair gaming devices (well-balanced coins and dice, well shuffled decks of cards)
  - high-rate coded digital data
  - experiment where
    - there are only n possible outcomes and they are all equally probable
    - there is a balance of information about outcomes

- **Bernoulli**
  - $\Omega = \{0,1\}$
  - $p_0 = 1-p$
  - $p_1 = p$
  - $\equiv \mathcal{B}(1,p)$

- **Binomial** $\mathcal{B}(n,p)$: $p_i = \binom{n}{i} p^i (1-p)^{n-i}$; for $\Omega = N_{n+1}$

  $0 \le p \le 1 \Rightarrow$probability of single occurrence of A = $\mathcal{B}(1,1)$
  - If have $\mathcal{E}_1, \ldots, \mathcal{E}_n$ n unlinked repetition of $\mathcal{E}$ and event A for E

    $\mathcal{B}(n,p)$= the probability that A occurs k times in $\mathcal{E}_1, \ldots, \mathcal{E}_n$
  - maximum probability value k = $\lfloor (n+1)p \rfloor \approx np$

    (Average #errors)
  - $\beta(1, \frac{1}{2}) \Rightarrow$ binomial that also random
  - #heads in n toss of a coin (p = 0.5)
  - #errors in n symbols of text

    (p = the probability of an error in a single symbol of text)

EE310 By Prapun Suksompong

- **Geometric** $\mathcal{G}(\beta)$: $p_i = (1-\beta)\beta^i$; $\Omega = N$, $0 \le \beta < 1$
  - $\beta = \frac{m}{m+1}$, m = mean/average waiting time/ lifetime
  - P(X=k) = P{ k failures followed by a success }

    =P^k\{failure\} P\{success\}
  - lifetimes of components, measured in discrete time units, when the fail catastrophically (without degradation due to aging)
  - waiting times
    - for next customer in a queue
    - between radioactive disintegrations
    - between photon emission
  - number of repeated, unlinked random experiments that must be performed prior to the first occurrence of a given event A
    - number of coin tosses prior to the first appearance of a 'head'
    - number of trials required to observe the first success

- **Poisson** $\mathcal{P}(\lambda)$: $p_i = e^{-l}\frac{l^i}{i!}$; $\Omega = N$, $0 \le \lambda$
  - λ= mean/average #counts = IT
  - T = observation time
  - I = an event intensity / rate of occurrence / current
  - most probable value i = $\lfloor l \rfloor$, peak value $p_{\lfloor l \rfloor} \approx \frac{1}{\sqrt{2pl}}$
  - i.i.d. $N_k \sim \mathcal{P}(\lambda_k) \to N = \sum N_k \sim \mathcal{P}(\Sigma\lambda_k)$
  - rare events limit of the binomial (large n, small b_n)

    let b_n depend on n and $\lim_{n\to\infty} nb_n = l > 0$ then

    $b_n \to 0$

    $$\lim_{n\to\infty} \binom{n}{i} b^i (1-b)^{n-i} = e^{-l}\frac{l^i}{i!}$$

- o #photons emitted by a light source of intensity I [photons/second] in time T ($\lambda = IT$)
- o #atoms of radioactive material of mass m undergoing decay in time T ($\lambda \propto mT$)
- o #clicks in a Geiger counter in T seconds when the average number of click in 1 second is I
- o #dopant atoms deposited to make a small device such as an FET
- o #customers arriving in a queue or workstations requesting service from a file server in time T ($\lambda \propto T$)
- o number of occurrences of rare events in time T
- o #soldiers kicked to death by horses

cdf: Univariate cumulative distribution function

cdf $F_X(x) = P(\{X: X \le x\})$

Lebesgue-Stieltjes integral

$$P(A) = \int_{-\infty}^{\infty} I_A(x)\,dF_X(x)$$

Riemann integral

$$P(A) = \int_{-\infty}^{\infty} I_A(x) f_X(x)\,dx = \int_A f(x)\,dx$$

- $1 \ge F(x) \ge 0$
- $P((a,b]) = F(b) - F(a)$

1. nondecreasing: $x_2 > x_1 \Rightarrow F_X(x_2) > F_X(x_1)$
2. Right Continuity: $F(x^+) = F(x)$

- $P(X = a) = P([a,a]) = F(a) - F(a^-)$ = jump height @ X = a
  - If F(x) is continuous @ a, then P(X = a) = 0

3. The left-hand limit: $\lim_{x \to -\infty} F(x) = 0$
4. The right-hand limit: $\lim_{x \to \infty} F(x) = 1$

- convex combination of cdfs $\sum_{i=1}^{n} l_i F_i(x)$ is a cdf

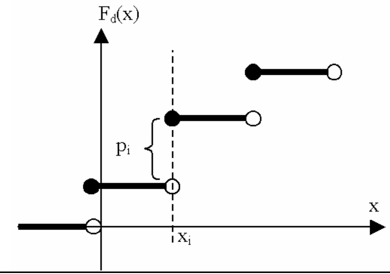
- F_X is continuous on the left at point *a* if and only if P(X=a) = 0.

3 types of cdfs

① Discrete CDF: $F_d(x) = \sum_i p_i U(x - x_i)$

$\{p_i\} \rightarrow$ pmf

- $P(X=x_i) = p_i$
- piecewise constant



② Absolutely Continuous CDF: $F_{ac}(x) = \int_{-\infty}^{x} f(z)\,dz$

$f(z) \rightarrow$ pdf

- P(X = x) = 0

③ Singular CDF $F_s(x)$

- the only discontinuities that a pdf can have are jump discontinuities
- pdf can have at most a countable number of jump discontinuities.

Decomposition:

Any F(x)

$= \lambda_d F_d(x) + \lambda_{ac} F_{ac}(x) \{+ \lambda_s F_s(x)\}$; $\sum l = 1$

$$= l \underbrace{\sum_i p_i U(x - x_i)}_{\text{step}} + (1 - l) \underbrace{\int_{-\infty}^{x} f(y)\,dy}_{\text{continuous}}$$

$$\frac{d}{dx} F(x) = l \sum_i p_i d(x - x_i) + (1 - l) f(x)$$

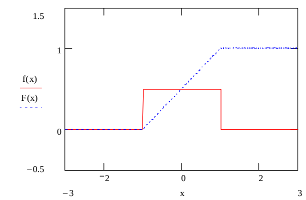
CDF representation for pmf p(ω) $\Omega = \{\omega_i\} \to F(x) = \sum_{i: w_i \le x} p(w_i) = \sum_i p(w_i) U(x - w_i)$
For a random experiment $\mathcal{E}$, can estimate the cdf from data on the actual outcomes $x_1,\ldots,x_n$ of n unlinked repetitions of $\mathcal{E}$ through empirical distribution function $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} U(x - x_i)$
the number of observations lying in the interval (a,b] = the height of the histogram $= n\left(\hat{F}_n(b) - \hat{F}_n(a)\right)$
$F(z^-)$ = the value of the cdf F as z is approached from the left (the lower cdf value if there is a jump discontinuity at z)
pdf: Probability Density Function
$x \in \Re$
pdf ⇔ ① $f(x) \ge 0$ ② $\int_{\Omega} f(x)dx = P(\Omega) = 1$
• f(x) can > 1 • convex combination of pdfs $\sum_i l_i f_i(x)$ yields a pdf
1:1 correspondence between P, cdf, pdf $P(A) = \int_A f(x)dx = \int_{-\infty}^{\infty} I_A(x) f(x) dx$ $F_X(x) = P(\{X : X \le x\}) = \int_{-\infty}^{x} f_X(x)dx$ $f_X(x) = \frac{d}{dx} F_X(x)$

pmf p is a special case of pdf f $p(x_i) = p_i = P(\{\omega = x_i\})$ $f(\omega) = \sum_i p_i d(w - x_i)$

- **Uniform** U(a,b)

$$f(x) = \frac{1}{b-a} U(x-a) U(b-x) = \begin{cases} 0 & x < a, x > b \\ \frac{1}{b-a} & a \le x \le b \end{cases}$$

$$F(x) = \begin{cases} 0 & x < a, x > b \\ \frac{x-a}{b-a} & a \le x \le b \end{cases}$$



continuous generalization of R(n)

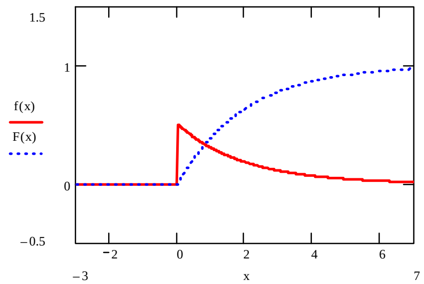
- o use with caution to represent ignorance about a parameter taking value in [a,b]
- o phase of oscillators ⇒ [-π,π] or [0,2π]
- o phase of received signals in incoherent communications → usual broadcast carrier phase φ ~ U(-π,π)
- o mobile cellular communication: multipath → path phases ϕ_c ~ U(-π,π)

- **Exponential** $\mathcal{E}(\alpha)$

  $f(x) = ae^{-ax}U(x)$; $\alpha > 0$

  $$F(x) = \left(1 - e^{-ax}\right)U(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-ax} & x \geq 0 \end{cases}$$

  $P(X>x) = e^{-\alpha x}U(x)$



continuous version of $\mathcal{G}(\beta)$

- Lack of memory property

  $$P\{X>k+c \mid X>k\} = \frac{P\{X > k + c\}}{P\{X > k\}} = P\{X>c\}$$

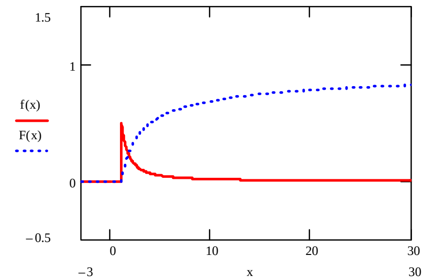
  - lifetimes (continuous time) of components of systems that fail without aging (memorylessness - eg wine glass) : mean life = $\frac{1}{a}$
  - waiting times between successive
    - photon arrivals
    - electron emissions from a cathode
    - radioactive decays
    - customer/packet arrivals
    - dopant atoms arrival in an implant process
  - duration of telephone or wireless call

EE310 By Prapun Suksompong

- **Pareto** $\mathcal{P}ar(\alpha)$: heavy-tailed model/density

  $f(x) = ax^{-a-1}U(x-1)$; $\alpha > 0$

  $$F(x) = \left(1 - \frac{1}{x^a}\right)U(x-1) = \begin{cases} 0 & x < 1 \\ 1 - \frac{1}{x^a} & x \geq 1 \end{cases}$$

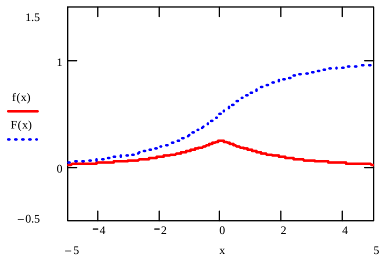


  - distribution of wealth
  - flood heights of the Nile river
  - designing dam height
  - (discrete) sizes of files requested by web users
  - waiting times between successive keystrokes at computer terminals
  - (discrete) sizes of files stored on Unix system file servers
  - running times for NP-hard problems as a function of certain parameters

- **Laplacian** $\mathcal{L}(\alpha)$

$$f(x)=\frac{a}{2}e^{-a|x|}\ ;\ \alpha>0$$

$$F(x)=\begin{cases}\frac{1}{2}e^{ax} & x<0\\ 1-\frac{1}{2}e^{-ax} & x\geq 0\end{cases}$$



  - o amplitudes of speech signals
  - o amplitudes of differences of intensities between adjacent pixels in an image

EE310 By Prapun Suksompong

- **Normal/Gaussian** $\mathcal{N}(m,\sigma^2)$

$$f(x)=\frac{1}{s\sqrt{2p}}e^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2}$$

m $\rightarrow$ mean
σ^2 $\rightarrow$ variance
σ $\rightarrow$ standard deviation ≥ 0
error function:

$$erf(z)=\Phi(z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2p}}e^{-\frac{x^2}{2}}dx$$

$$=1-\bar{\Phi}(z)$$

complementary error function:

$$cerf(z)=\bar{\Phi}(z)=1-\Phi(z)=\int_{z}^{\infty}\frac{1}{\sqrt{2p}}e^{-\frac{x^2}{2}}dx$$

$$=\frac{1}{\sqrt{2p}}\left(\frac{e^{-\frac{x^2}{2}}}{x}-\int_{x}^{\infty}\frac{e^{-\frac{x^2}{2}}}{x^2}dx\right)$$

$$\approx\frac{e^{-\frac{x^2}{2}}}{x\sqrt{2p}}\ \text{ for large x}$$

F(x) = $\Phi\left(\frac{x-m}{s}\right)$

- thermal noise
  - in resisters
  - produced by all dissipative physical systems operating above 0 K
  - Voltage across a resistor R [ohm] @ T [K]
    $V \sim N(0, \sigma^2)$; $\sigma^2 = 4kTRB$
    B : single-sided (only positive, physical frequencies) bandwidth [Hz]
    k = Boltzmann's constant= $1.38\times10^{-23}\left[\frac{Watt}{Hz\cdot K}\right]$
  - a device have a noise temperature of T K $\to$ average power P = kTB would be delivered to R_{load} under fictitious assumption that the source is also a resistance $R_s = R_{load}$ @ T
    - 3 K universal background radiation
    - 290 K radiation from the Earth as seen from space
- shot noise produced by the random arrivals of individual photons or electrons
- low-frequency noise ($\frac{1}{f}$, flicker, semiconductor, excess noises) as found in
  - low-frequency amplifiers
  - variation in quartz crystal oscillator frequency
- many kinds of measurement errors
- variabilities in parameters of
  - manufactured components
  - biological organisms (height, weight, intelligence)
- certain characteristics oflarge-scale systems formed out of many loosely interacting components

Reliability assessment
components that do not degrade significantly due to aging

- due to manufacturing errors, they may fail on the first use with small probability λ
- if they do not fail immediately, lifetime L ~ E(α)

$$f_L(x) = \lambda\delta(x) + (1-\lambda)\alpha e^{-\alpha x}U(x)$$

Rayleigh

$$F(x) = \left(1 - e^{-\alpha x^2}\right)U(x)$$

$$f(x) = 2\alpha x e^{-\alpha x^2}U(x)$$

$$P(\{X > t\}) = 1 - F(t) = \begin{cases} e^{-\alpha t^2} & t \ge 0 \\ 1 & t < 0 \end{cases}$$

- noise X at the output of AM envelope detector when no signal is present
- If X and Y are independent, identically distributed normal random variables, then R ≡ $\sqrt{X^2+Y^2}$ has a Rayleigh density

Cauchy

f(z) = $\frac{\alpha}{\pi}\frac{1}{\alpha^2+z^2}$ where a > 0

- If X and Y are independent, identically distributed normal random variables, then Z ≡ $\frac{Y}{X}$ has a Rayleigh density

Multivariate

uncountably infinite sample spaces

$\Omega \subset \Re^n$

orthant / semi-infinite corner with northeast vertex specified by the point $\underline{x}$

$C_{\underline{x}} = \{\underline{X} : (\forall i \le n)\ X_i \le x_i\}$

multivariate / joint cdf

$F_{\underline{X}}(\underline{x}) = P(C_{\underline{x}})$

- o non-decreasing in each variable :
  $\underline{x}_a < \underline{x}_b$ component-wise $\rightarrow F_{\underline{X}}(\underline{x}_a) < F_{\underline{X}}(\underline{x}_b)$
- o $\lim_{x_i \to -\infty} F_{\underline{X}}(\underline{x}) = 0$
- o $\lim_{x_n \to \infty} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = F_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1})$
- o $\lim_{x_1,\ldots,x_n \to \infty} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = 1$
- o for n =2
  $P(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2)$
  $= F_{\underline{X}}(b_1,b_2) - F_{\underline{X}}(a_1,b_2) - F_{\underline{X}}(b_1,a_2) + F_{\underline{X}}(a_1,a_2)$

multivariate / joint pdf

PDF1	$f(\underline{x}) \ge 0$; $\forall \underline{x} \in \Re^n$
PDF2	$\int_{\Re^n} f_{\underline{X}}(\underline{x}) d\underline{x} = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f_{\underline{X}}(\underline{x}) dx_1 \ldots dx_n = 1$

$F_{\underline{X}}(\underline{x}) = P(C_{\underline{x}}) = \int_{\Re^n} f_{\underline{X}}(\underline{x}) d\underline{x} = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_{\underline{X}}(\underline{x}) dx_1 \ldots dx_n$

$P(A) = \int_A f_{\underline{X}}(\underline{x}) d\underline{x} = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f_{\underline{X}}(\underline{x}) I_A(\underline{x}) d\underline{x}$

$f_{\underline{X}}(\underline{x}) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F_{\underline{X}}(\underline{x})$

$f_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1}) = \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) dx_n$

product or independence construction
when have unlinked or unrelated $\{E_i\}$,
X_i: outcome of E_i (univariate)

$f_{\underline{X}}(\underline{x}) = \prod_1^n f_i(x_i)$, $F_{\underline{X}}(\underline{x}) = \prod_1^n F_i(x_i)$

- $F_{X_a^b}(x_a^b) = F_{X_a,X_{a+1},\ldots,X_b}(x_a, x_{a+1},\ldots,x_b)$; $b \ge a$
- $F_{X_a^a}(x_a^a) = F_{X_a}(x_a)$

Functions of Random Variables

Random variable

- a real-valued X that somewhat unpredictably takes on a specific numerical value x when the appropriate random experiment is performed.
- vector-valued random quantities: $\underline{X}$
- a function, mapping, or transformation from a given initial probability space $\Omega, \mathcal{A}, P$ to a final probability space $\Omega_X, \mathcal{B}, P_X$
  $X: \Omega \rightarrow \Omega_X$, $X(\omega) \in \Omega_X$

Measurability

The function X is measurable wrt. the algebras $\mathcal{A}, \mathcal{B} \Leftrightarrow$

$(\forall B \in \mathcal{B})\ X^{-1}(B) \in \mathcal{A}$

Random variables are assumed to be measurable wrt. the relevant σ-algebras

an initial probability space $\Omega, \mathcal{A}, P$

and

a final probability space $\Omega_X, \mathcal{B}, P_X$

are linked by a random variable X
whenever
$X: \Omega \rightarrow \Omega_X$
is measurable wrt. the algebras $\mathcal{A},\mathcal{B}$
and
$(\forall B \in \mathcal{B})\ P_X(B) = P(X^{-1}(B))$

$Y = g(X)$

- $\mathcal{E}_X = (\Omega_X, \mathcal{A}_X, P_X)$
  $\mathcal{E}_Y = (\Omega_Y, \mathcal{A}_Y, P_Y)$
- $\underline{\Omega_X} \xrightarrow{g} \underline{\Omega_Y} = \{Y: (\exists X \in \Omega_X)\ Y = g(X)\}$
- $\Omega_Y \supset g(\Omega_X)$
- for $A_Y \in \Omega_Y$, if $g^{-1}(A_Y) = \varnothing$, then $P_Y(A_Y) = 0$
- g is measurable wrt. the algebras $\mathcal{A}_X, \mathcal{A}_Y \Leftrightarrow$
  $(\forall A_Y \in \mathcal{A}_Y)\ g^{-1}(A_Y) \in \mathcal{A}_X$

Given P_X, g, need P_Y

- $P_Y(A_Y) = P_X(g^{-1}(A)) = \int\limits_{\{x: x \in g^{-1}(A_Y)\}} f_X(x)dx = \int\limits_{g^{-1}(A_Y)} f_X(x)dx$

To determine $f_{\underline{Y}}$ from $f_{\underline{X}}$

- $F_Y(y) = P_Y(Y \le y) = P_X(g(X) \le y) = \int\limits_{\{x: g(x) \le y\}} f_X(x)dx$
- $f_Y(y) = \frac{d}{dy} F_Y(y)$

SISO

Linear: **$Y = aX+b$**
continuous F_X (for $a < 0$ part)

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

- $X \sim N(m,\sigma^2)$, $Y = aX+b \to Y \sim N(am+b, a^2\sigma^2)$

Power Law **$Y = X^n$**

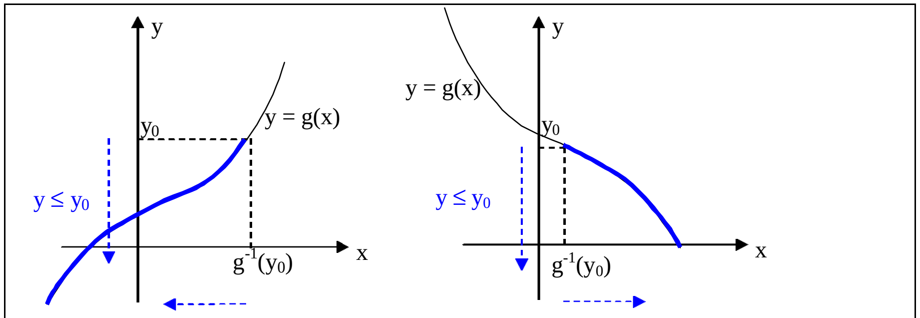
- n odd: $f_Y(y) = \frac{1}{n} y^{\frac{1}{n}-1} f_X\left(y^{\frac{1}{n}}\right)$
- n even & continuous F_X:
  $$f_Y(y) = \frac{1}{n} y^{\frac{1}{n}-1}\left[f_X\left(y^{\frac{1}{n}}\right) + f_X\left(-y^{\frac{1}{n}}\right)\right]U(y)$$

$g^{-1}(y) = \min\{x: g(x) \ge y\}$

Monotone g (strictly increasing or strictly decreasing; one-to-on single variable transformation)

$$f_Y(y) = \left|\frac{d}{dy} g^{-1}(y)\right| f_X\left(g^{-1}(y)\right)$$

- $F_Y(y) = P(Y \le y)$
  - For monotone-increasing function,
    $F_Y(y) = P\left(X \le g^{-1}(y)\right) = F_X(g^{-1}(y))$
    $$f_Y(y) = f_X\left(g^{-1}(y)\right)\underbrace{\left(\frac{d}{dy} g^{-1}(y)\right)}_{\ge 0}$$
  - For monotone-decreasing function,
    $F_Y(y) = P\left(X \ge g^{-1}(y)\right)$
    $= 1 - P\left(X \le g^{-1}(y)\right) + P\left(X = g^{-1}(y)\right) = 1 - F_X(g^{-1}(y))$
    $$f_Y(y) = -f_X\left(g^{-1}(y)\right)\underbrace{\left(\frac{d}{dy} g^{-1}(y)\right)}_{\le 0}$$



$X,Y \sim \mathcal{E}(\alpha) \Rightarrow Z = X-Y \sim \mathcal{L}(\alpha)$

Ex. $Y = X^2$

note that $Y \geq 0$; so $f_Y(y) = F_Y(y) = 0$ for $y < 0$

$F_Y(y) = P(Y \leq y) = P(0 \leq X^2 \leq y)$

For $y \geq 0$, $F_Y(y) = P(0 \leq X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y})$

$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$

$$f_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{2\sqrt{y}} f_X\left(\sqrt{y}\right) + \frac{1}{2\sqrt{y}} f_X\left(-\sqrt{y}\right) & y \geq 0 \end{cases}$$

Ex. $Y = X^2 U(x)$

note that $Y \geq 0$; so $f_Y(y) = F_Y(y) = 0$ for $y < 0$

$F_Y(y) = P(Y \leq y) = P(X^2 U(X) \leq y)$

note that $X^2 U(X) \geq 0$

$F_Y(y) = P(Y \leq y) = P(0 \leq X^2 U(X) \leq y)$

$F_Y(0) = P(0 \leq X^2 U(X) \leq 0) = P(X^2 U(X) = 0) = P(X=0) + P(U(X)=0)$

$= 0 + P(X<0) = F_X(0)$

For $y > 0$, $F_Y(y) = P(0 \leq X^2 U(X) \leq y) = P(X<0) + P(0 \leq X \leq \sqrt{y})$

$= P(X \leq \sqrt{y}) = F_X(\sqrt{y})$

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ F_X(0) & y = 0 \\ F_X\left(\sqrt{y}\right) & y > 0 \end{cases}$$

$$f_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{2\sqrt{y}} f_X\left(\sqrt{y}\right) & y > 0 \end{cases} + F_X(0)\delta(y)$$

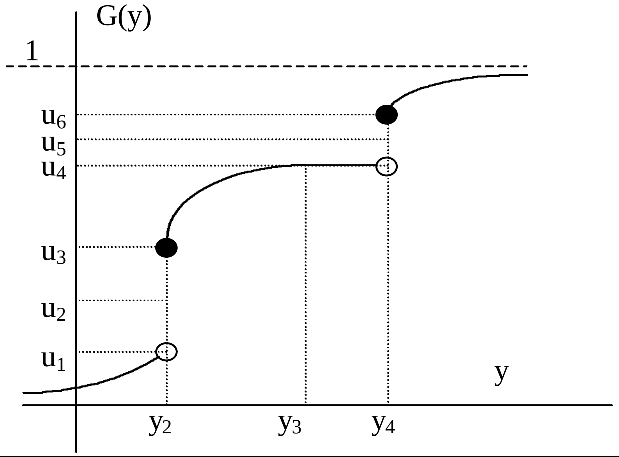
To generate $Y \sim F_Y(y) = G(y)$ from $U \sim \mathcal{U}(0,1)$

Inverse cdf

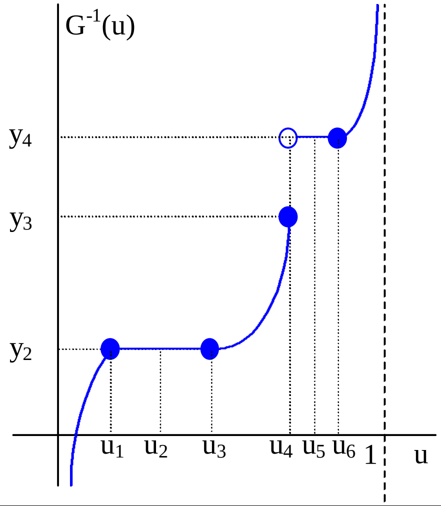
The inverse G^{-1} to a cdf G is

$G^{-1}(u) = \min\{y: G(y) \geq u\}$

- Both G and G^{-1} are nondecreasing functions
- $G(G^{-1}(u)) \geq u$
- $G^{-1}(G(y)) \leq y$
- ? G^{-1} of confusing u is @ jump



- $G(y_2) = u_3$
- $G^{-1}(u_3) = \min\{y: G(y) \geq u_3\} = y_2$
- $G^{-1}(u_1<u_2<u_3) = \min\{y: G(y) \geq u_2\} = y_2$
  G(y) jump from u_1^- to u_3 at y_2 so, min $G(y) \geq u_2$ is u_3
  $G(y) = u_3 \rightarrow y = y_2$
- $G^{-1}(u_1) = \min\{y: G(y) \geq u_1\} = y_2$
  no $G(y) = u_1$
  G(y) jump from u_1^- to u_3 at y_2 so, min $G(y) \geq u_1$ is u_3
- $G^{-1}(u_6) = y_4$
- $G^{-1}(u_5) = \min\{y: G(y) \geq u_5\} = \{y: G(y) = u_6\} = y_4$
- $G^{-1}(u_4) = \min\{y: G(y) \geq u_4\} = \min\{y: G(y) = u_4\} = y_3$



- Note the difference between choosing the value at u_1 and u_4
  incoming increasing $\Rightarrow$ choose y at jump
  incoming constant $\Rightarrow$ choose y at left end of constant
  problem if have constant piece $G((-\infty,a)) = 0$ ans.: $G^{-1}(0) = 0$

$U \sim \mathcal{U}(0,1) \rightarrow Y = G^{-1}(U) \sim$ cdf G

strictly increasing continuous cdf F

- F^{-1} is continuous and strictly increasing

$Y \sim F \rightarrow X = F(Y) \sim \mathcal{U}(0,1)$

: probability integral transformation

MIMO: Multiple Input – Multiple Output

$\dim(\underline{X}) = \dim(\underline{Y}) \Rightarrow n = m$

For a 1:1 nonlinear transfrmation g from $\underline{X}$ to $\underline{Y}$, $\underline{Y} = g(\underline{X})$,
with differentiable inverse g^{-1}, $\underline{X} = g^{-1}(\underline{Y})$

$f_{\underline{Y}}(\underline{y}) = f_{\underline{X}}(\, g^{-1}(\underline{y})\,)\, |\det J|$

$J = [J_{i,j}]$, $J_{i,j} = \dfrac{\partial x_i}{\partial y_j} = \dfrac{\partial}{\partial y_j} g_i^{-1}(\underline{y})$

$$\det J = \begin{vmatrix} \frac{\partial}{\partial y_1} g_1^{-1}(y_1,\dots y_n) & \cdots & \frac{\partial}{\partial y_n} g_1^{-1}(y_1,\dots y_n) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial y_1} g_n^{-1}(y_1,\dots y_n) & \vdots & \frac{\partial}{\partial y_n} g_n^{-1}(y_1,\dots y_n) \end{vmatrix}$$

$$= \frac{1}{\begin{vmatrix} \frac{\partial}{\partial x_1} g_1(x_1,\dots x_n) & \cdots & \frac{\partial}{\partial x_n} g_1(x_1,\dots x_n) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} g_n(x_1,\dots x_n) & \vdots & \frac{\partial}{\partial x_n} g_n(x_1,\dots x_n) \end{vmatrix}}.$$

Ex $\underline{Y} = A\underline{X} + \underline{b}$

$f_{\underline{Y}}(\underline{y}) = f_{\underline{X}}(A^{-1}\underline{y} - A^{-1}\underline{b}) \cdot \frac{1}{|A|}$

Ex Transformation from Cartesian coordinates (x,y) to polar coordinates (r,θ) $\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} r \\ q \end{pmatrix}$

x = r cos θ , y = r sin θ , r = $\sqrt{x^2 + y^2}$, θ = $\tan^{-1}\left(\frac{y}{x}\right)$

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial q} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial q} \end{vmatrix} = \begin{vmatrix} \cos q & -r\sin q \\ \sin q & r\cos q \end{vmatrix} = r \;:\; dx\,dy = r\,dr\,d\theta$$

$f_{R,\Theta}(r, \theta) = f_{X,Y}(x = r\cos\theta ,\ y = r\sin\theta)\, r$

EE310 By Prapun Suksompong

Special case:
Transforming Uniforms into Normals

$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{U}(0,1)$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \sqrt{-2s^2 \log x_1}\cos(2px_2) \\ \sqrt{-2s^2 \log x_1}\sin(2px_2) \end{pmatrix} \sim N(0,\sigma^2)$$

For $\underline{X}$ be i.i.d. pdf type C
means $f_{\underline{X}}(\underline{x}) = f_{X_1}(x_1)\cdot f_{X_2}(x_2)\cdots f_{X_n}(x_n)$, $X_i \sim C$

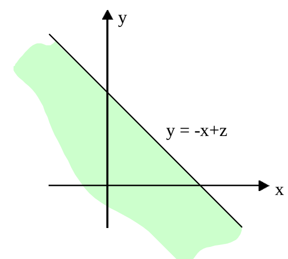
MIMO transformation: dim($\underline{Y}$) = m < n = dim($\underline{X}$)

solution: augment $\underline{Y}$ to $\underline{Y}'$, dim($\underline{Y}'$) = dim($\underline{X}$)
(addition of n-m variablesto $\underline{Y}$ so that the new overall transformation has a unique differentiable inverse)

Sums of random variables

Z = X + Y

P(Z ≤ z) = P(X+Y≤z)



$F_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x'} f_{X,Y}(x', y')\,dy'dx'$

$f_Z(z) = \frac{d}{dz}F_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x', z-x')\,dx' \Rightarrow$ superposition integral

If X and Y are independent random variables, then

$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx \Rightarrow$ convolution integral

Product of random variable

If random variable X and Y are statistically independent and if their pdf f_X and f_Y respectively, exist almost everywhere, the the probability density of their product

Z = XY

is given by the formula

$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y\left(\frac{z}{x}\right)dx$

Expectation and Moments

Expectation

$EX = \int_{-\infty}^{0} x f_X(x)\,dx + \int_{0}^{\infty} x f_X(x)\,dx$

provided that at least one of the two integrals is finite (can't have (-∞)+(+∞))

For discrete-valued X

$EX = \sum_{i: x_i \le 0} x_i P(X = x_i) + \sum_{i: x_i > 0} x_i P(X = x_i)$

provided that at least one of these sums is finite

EE310 By Prapun Suksompong

- P(X=c) = 1 → EX = E(c) = c
- P(X≥0) = 1 → EX ≥ 0
- E(aX) = aEX
- E(X+Y) = EX+EY
  $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} EX_i$
- E(aX+bY+c) = aEX + bEY + c ; for finite EX, EY
- P(X≥Y) = 1 → EX ≥ EY
- $EX = -\int_{-\infty}^{0} F_X(x)\,dx + \int_{0}^{\infty} (1 - F_X(x))\,dx$ *
  if $\lim_{x\to-\infty} xF(x)dx = \lim_{x\to\infty} x(1-F(x))dx = 0$
- $E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$ *

Ex. g(x) = U(t-x)

$E(g(X)) = \int_{-\infty}^{\infty} U(t-x) f_X(x)\,dx = \int_{-\infty}^{t} f_X(x)\,dx = F_X(t)$

- $E\left(\left(\sum_{i=1}^{n} X_i\right)^2\right) = E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} E(X_i X_j)$

n^{th} Moment: $E(X^n) = \int_{-\infty}^{\infty} x^n f_X(x)\,dx$; if exists

- Finiteness of Moments: j<k → { $|E(X^k)|<\infty \to |E(X^j)|<\infty$ }
  - the existence of finite higher-order moments implies the existence of finite lower-order moments

n^{th} Central Moment: $E\left((X - EX)^n\right) = \int_{-\infty}^{\infty} (x - EX)^n f_X(x)\,dx$; if exists

- n = 1: E(X-EX) = 0
- $E\left(X^n\right)=\sum_{k=0}^{n}\binom{n}{k}(EX)^{n-k}E\left((X-EX)^k\right)$
- the k[th] moment exists ↔ the k[th] central moment exists finite

Variance: VAR(X)

= σ_X^2

= second central moment = $E\left((X-EX)^2\right)=\int_{-\infty}^{\infty}(x-EX)^2 f_X(x)dx$

= $E(X^2)-(EX)^2$

- VAR(X) ≥ 0
- VAR(c) = 0
- VAR(X+c) = VAR(X)
- VAR(aX) = a^2VAR(X)
- 0 ≤ VAR(X) ≤ $E(X^2)$
- EX = 0 →
  - VAR(X) = $E(X^2)$
  - VAR(X+Y) = VAR(X) + VAR(Y) + 2COV(X,Y)

standard deviation = $+\sqrt{VAR(X)}$

Correlation: E(XY) = $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{X,Y}(x,y)dxdy$

Covariance:

COV(X,Y) = E(X-EX)(Y-EY) = E(XY) – (EX)(EY)

- COV(X,Y) = COV(Y,X)
- COV(X+c,Y+c) = COV(X,Y)
- COV(aX,bY) = abCOV(X,Y)
- COV(aX+b,cY+d) = acCOV(X,Y)
- { E(XY) = (EX)(EY) ↔ COV(X,Y) = 0 } → X,Y are uncorrelated
- E(XY) = 0 → X,Y are orthogonal

EE310 By Prapun Suksompong

- EX = 0 or EY = 0

  → COV(X,Y) = E(XY)

  → orthogonality is equivalent to uncorrelatedness
- $VAR\left(\sum_{1}^{n}X_i\right)=\sum_{i=1}^{n}\sum_{j=1}^{n}COV\left(X_i,X_j\right)$
- there can be a high degree of nonlinear dependence that is not seen by the covariance

autocorrelation / normalized covariance:

$\rho_{X,Y}=\dfrac{COV(X,Y)}{\sqrt{VAR(X)\cdot VAR(Y)}}=\dfrac{COV(X,Y)}{s_X s_Y}$

- $0\le\rho_{X,Y}\le 1$
- large $\rho_{X,Y}$ →

  high degree of linear dependence between X and Y

Estimator

- Observe X
- Inference: make a linear approximation $\hat{Y}(X)=$ aX+b to Y
- mean squared error = $E(Y-\hat{Y})^2 = E(Y-aX-b)^2$

$E(Y-aX-b)^2$

$= E\left((Y-EY)-a(X-EX)+(\underbrace{-b+EY-aEX}_{c})\right)^2$

$= E\big((Y-EY)^2+a^2(X-EX)^2+c^2$
$\quad -2a(Y-EY)(X-EX)+2c(Y-EY)-ac(X-EX)\big)$

$= VAR(Y)+a^2VAR(X)+c^2-2aCOV(X,Y)$

- select a, c that yield a MMSE (minimum mean squared error) choose c = 0 ⇒ b = EY – a EX

∴ $\hat{Y}(X)$ = aX + EY – a EX = a(X-EX) + EY

or can use

$$\frac{d}{db}E\left(Y-\hat{Y}\right)^2 = E\frac{d}{db}\left(Y-\hat{Y}\right)^2 = E\left(-2\left(Y-\hat{Y}\right)\underbrace{\left(\frac{d\hat{Y}}{db}\right)}_{1}\right) = 0$$

- E$\hat{Y}$ = EY , b = EY – a EX

now we have $\hat{Y}(X)$ = a(X-EX) + EY

Y - $\hat{Y}(X)$ = (Y-EY) – a(X-EX)

E(Y – aX – b)^2 = VAR(Y) + a^{2}VAR(X) – 2aCOV(X,Y)

$$\frac{d}{da}E\left(Y-\hat{Y}\right)^2 = 0 + 2aVAR(X) - 2COV(X,Y)$$

To minimize E(Y-$\hat{Y}$)^2 , need a = $\frac{COV(X,Y)}{VAR(X)}$

$$\frac{d}{da}E\left(Y-\hat{Y}\right)^2 = 2E\left(Y-\hat{Y}\right)\frac{d\hat{Y}}{da} = 2E\left(Y-\hat{Y}\right)(X-EX)$$

$E\left(Y-\hat{Y}\right)(X-EX)$ = 0 : orthogonality condition

$$E\left(Y-\hat{Y}\right)(X-EX) = E\left((Y-EY) - a(X-EX)\right)(X-EX)$$
$$= COV(Y,X) - aVAR(X) = 0$$

E(Y – aX – b)^2 = E((Y-EY) – a(X-EX))^2
(= VAR(Y) + a^{2}VAR(X) – 2aCOV(X,Y))

- $\hat{Y}(X) = \frac{COV(X,Y)}{VAR(X)}(X-EX) + EY$ *

E(Y – aX – b)^2

$$= VAR(Y) + \left(\frac{COV(X,Y)}{VAR(X)}\right)^2 VAR(X)$$
$$-2\left(\frac{COV(X,Y)}{VAR(X)}\right)COV(X,Y)$$
$$= VAR(Y) - \frac{\left(COV(X,Y)\right)^2}{VAR(X)}$$
$$= VAR(Y)\cdot\left(1 - \frac{\left(COV(X,Y)\right)^2}{VAR(Y)\cdot VAR(X)}\right)$$
$$= VAR(Y)\cdot\left(1 - r_{X,Y}^2\right)$$

- performance = E(Y-$\hat{Y}$)^2 = VAR(Y) (1 - $\rho_{X,Y}^2$)

worse case

if $\rho_{X,Y}$ = 0 ↔ COV(X,Y) = 0 ↔ X and Y are uncorrelated , then

- o have biggest mean squared error
- o $\hat{Y}(X)$ = EY (not using X)
- o performance = VAR(Y)

Ex Observe X = Y + N = noisy reading on Y

Infer Y using $\hat{Y}(X)$ = aX+b

Assume EN = 0, COV(N,Y) = 0 (uncorrelated)

EN = 0 → EX = E(Y+N) = EY + EN = EY + 0 = EY b: $E\hat{Y} = EY$

E(aX+b) = aEX + b = aEY + b = EY

b = EY(1-a)

COV(X,Y)

= E(X-EX)(Y-EY) = E(Y+N – EY)(Y-EY)

= E $(Y-EY)^2$ – E(N)(Y-EY)

= VAR(Y) - E(N-EN)(Y-EY) ; EN = 0

= VAR(Y) – COV(N,Y) = VAR(Y)

VAR(X)

= $E(X-EX)^2 = E(Y+N-EY)^2$

= $E(Y-EY)^2 + E(N^2) + 2E(N)(Y-EY)$

= $VAR(Y) + E(N-EN)^2 + 2E(N-EN)(Y-EY)$

= VAR(Y) + VAR(N) + 2COV(N,Y)

= VAR(Y) + VAR(N)

$$a = \frac{COV(X,Y)}{VAR(X)} = \frac{VAR(Y)}{VAR(Y)+VAR(N)}$$

$$\hat{Y}(X) = a(X-EX) + EY = \frac{\frac{VAR(Y)}{VAR(N)}}{\frac{VAR(Y)}{VAR(N)}+1}(X-EY) + EY$$

- $\frac{VAR(Y)}{VAR(N)}$ = SNR: signal-to-noise ratio
- If SNR $\to \infty$, then
  X is a good measurement
  $\hat{Y}(X)$ = X
- If SNR << 1
  $\hat{Y}(X)$ = EY

$\rho_{X,Y}$

$$= \frac{(COV(X,Y))^2}{VAR(Y)\cdot VAR(X)} = \frac{(VAR(Y))^2}{VAR(Y)\cdot(VAR(Y)+VAR(N))}$$

$$= \frac{1}{1+\frac{VAR(N)}{VAR(Y)}}$$

performance

$$= VAR(Y)(1 - \rho_{X,Y}^2) = \frac{VAR(N)}{1+\frac{VAR(N)}{VAR(Y)}}$$

Extensions of Expectation to Vector, Matrix, and Complex-Valued Variables
$E\mathbf{M} = [EM_{i,j}]$ $E\mathbf{X} = [EX_i]$
$Z = X + iY \to EZ = EX + iEY$
• $E(\mathbf{A}\underline{X}+\mathbf{B}\underline{Y}+\underline{c}) = \mathbf{A}E\underline{X}+\mathbf{B}E\underline{Y}+\underline{c}$
A is nonnegative definite matrix $\leftrightarrow \underline{a}^T\mathbf{A}\underline{a} = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j}a_ia_j \ge 0$ $\leftrightarrow$ symmetrix and all its eigenvalues are nonnegative
Correlation Matrix: $\mathbf{R}_{\underline{X}} = E(\underline{X}\underline{X}^T) = [E(X_iX_j)]$ • symmetric
Cross-correlation Matrix: $\mathbf{R}_{X,Y} = E(\underline{X}\underline{Y}^T) = [E(X_iY_j)]$
• not symmetric in general • $\mathbf{R}_{X,X} = \mathbf{R}_X$
Covariance Matrix: $\mathbf{C}_X = E((\underline{X}-E\underline{X})(\underline{X}-E\underline{X})^T) = \mathbf{R}_{\underline{X}} - (E\underline{X})(E\underline{X})^T = [COV(X_i,X_j)]$ $= COV(\underline{X},\underline{X})$
• symmetric • have $VAR(X_i)$ on its diagonal line $\leftarrow COV(X_i,X_i) = VAR(X_i)$

Cross-covariance Matrix:
COV($\underline{X}$,$\underline{Y}$) = E(($\underline{X}$-E$\underline{X}$)($\underline{Y}$-E$\underline{Y}$)^T) = [COV(X_i,Y_j)]

- E$\underline{X}$ = 0 $\rightarrow$ $\mathbf{R}_X = \mathbf{C}_X$
- **A** is a $\mathbf{R}_X$ or $\mathbf{C}_X$ $\leftrightarrow$ A is nnd

Linear transformation

$\underline{Y}$ = %$\underline{X}$ + $\underline{b}$

- E$\underline{Y}$ = $ E$\underline{X}$ + $\underline{b}$
- 5 $_Y$ = $ 5 $_X$$ ^{T} + $\underline{b}\underline{b}^T$ + $\underline{b}$(E$\underline{X}$)^T$ ^{T} + $ (E$\underline{X}$)$\underline{b}^T$
- &$_Y$ = $ &$_X$$ ^{T}

X ~ N($\underline{m}$, C) $\Rightarrow$ Y ~ N(A$\underline{m}$+b,ACA^T)

Wiener Filtering

Observe $\underline{X}$ infer Y by $\hat{Y}(\underline{X}) = \underline{a}^T\underline{X}$+b

To minimize MSE

- E$\hat{Y}$ = EY , b = EY – $\underline{a}^T$E$\underline{X}$
- orthogonality condition:
  E(Y-$\hat{Y}$)(X_i-EX_i) = 0
  E($\underline{X}$-E$\underline{X}$)(Y-$\hat{Y}$) = 0
- $\underline{a}$ = $\mathbf{C}_X^{-1}$COV($\underline{X}$,Y)

Conditional Probability

(revised probability)

P(A|B) : the conditional probability of A given B

- P(A|B) $\geq$ 0
- P(Ω|B) = 1
- P(B|B) = 1
- P(A|B) = P(A$\cap$B|B)
- **A $\supset$ B $\rightarrow$ P(A|B) = 1**
- $A_1 \perp A_2 \rightarrow P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$
- For fixed B, P(.|B) is a probability measure
- P(A|.)
- A $\perp$ B $\rightarrow$ P(B|A) = 0 , and if P(B) $\neq$ 0, P(A|B) = 0

For P(B) > 0

F_X(x|B) = P({X:X≤x}|B) = $\dfrac{P(\{X \leq x\} \cap B)}{P(B)} = \dfrac{\int_{-\infty}^{x} f(x') I_B(x') dx'}{P(B)}$

$$f(x|B) = \frac{f(x) I_B(x)}{P(B)} = \begin{cases} \dfrac{f(x)}{P(B)} & \text{if } x \in B \\ 0 & \text{if otherwise} \end{cases}$$

Markov Dependence

$\forall$i > 1 ; $P\left(A_i \middle| \bigcap_{j=1}^{i-1} A_j\right) = P(A_i \mid A_{i-1})$

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)\prod_{k=2}^{n} P(A_k|A_{k-1})$$

Ex	• the successive orderings $\{A_i\}$ of a deck of cards in repeat shuffles

Discrete X, Discrete Y

P(Y = y_i|X=x_j) = $\dfrac{P(Y = y_i, X = x_j)}{P(X = x_j)}$

P(Y=y_i) = $\sum_j P(Y = y_i | X = x_j) P(X = x_j)$

P(X = x_j|Y=y_i) = $\dfrac{P(Y = y_i | X = x_j) P(X = x_j)}{P(Y = y_i)}$

P(Y_1=y_1,…,Y_n=y_n) = $P(Y_1 = y_1) \times \prod_{j=2}^{n} P(Y_j = y_j | Y_1 = y_1, \ldots, Y_{j-1} = y_{j-1})$

- P(A|B) + P(A^c|B) = 1
- P(A|B) + P(A|B^c) = ?
- P(A|B)P(B) + P(A|B^c)P(B^c) = P(A)

Ex	S: **BSC: binary symmetric channel** X,Y $\in$ {0,1} $\underset{\{0,1\}}{X} \to S \to \underset{\{0,1\}}{Y}$ P(Y=1-x|X=x) = p
	• P(Y=1|X=0) = P(Y=0|X=1) = p = P(error) (Given) • P(Y=x|X=x) = 1-p P(Y=0|X=0) = P(Y=1|X=1) = 1-p • P(Y=y) = P(Y=y|X=0) $\times$P(X=0)+ P(Y=y|X=1) $\times$P(X=1) • P(X=0,Y=0) = P(Y=0|X=0)$\times$ P(X=0) = (1-p)$\times$P(X=0) • P(X=0|Y=0) = $\frac{P(Y=0 \mid X=0)P(X=0)}{P(Y=0)}$ $= \frac{(1-p)\times P(X=0)}{(1-p)\times P(X=0)+p\times P(X=1)}$ • P(X=1|Y=0) = 1 – P(X=0|Y=0) • P(X=1|Y=1) = $\frac{P(Y=1 \mid X=1)P(X=1)}{P(Y=1)}$ $= \frac{(1-p)\times P(X=1)}{(1-p)\times P(X=1)+p\times P(X=0)}$ $= \frac{1}{1+\frac{p\times P(X=0)}{(1-p)\times P(X=1)}}$
	Estimator $\hat{X}$ (Y) of X
	E($\hat{X}$ -X)^2 = (1)P($\hat{X}$ $\neq$x) + (0)P($\hat{X}$ =x) = P($\hat{X}$ $\neq$x) = P(error)
	Observe Y = 1

	• **Decide X = 1** if **P(X=1|Y=1) $\geq$ P(X=0|Y=1)** = 1 - P(X=1|Y=1) $\Rightarrow$ **P(X=1|Y=1) $\geq$ 0.5** $\Rightarrow \frac{1}{1+\frac{p\times P(X=0)}{(1-p)\times P(X=1)}} \geq \frac{1}{2}$ $\Rightarrow \frac{p\times P(X=0)}{(1-p)\times P(X=1)} \leq 1$ $\Rightarrow \frac{p}{(1-p)} \leq \frac{P(X=1)}{P(X=0)}$ $\Rightarrow$ not only depend on p

Ex	Queue Buffer
	▪ Buffer stores upto b bits
	▪ State of buffer: $B_t = k_t$; $0 \leq k_t \leq b \Rightarrow k_t$ bits in buffer
	▪ $B_t \to B_{t+1} \Rightarrow$ change k
	  ▪ remove r_{t+1} = r bits/cycle from buffer ; if $k_t < r$, empty the buffer
	  ▪ add L_t = length of t^{th} cycle packet

	$P(B_{t+1}=k_{t+1} \mid B_t=k_t)$ = ? Want $k_{t+1} = k_t+L_{t+1}-r_{t+1}$ $P(B_{t+1}=k_{t+1} \mid B_t=k_t) = P(L_{t+1}=k_{t+1}-k_t+r_{t+1})$ ▪ = 0 if $k_t - r_{t+1} < 0$ or $k_t - r_{t+1} > k_{t+1}$ ▪ = 0 if $k_{t+1} > b$ or $k_{t+1} < 0$
	Buffer overflow P(overflow) = $P(k_{t+1}>b) = \sum_{i=0}^{b} P(k_{t+1} > b \mid k_t = i)P(k_t = i)$ $= \sum_{i=0}^{b} P(L_{t+1} > b - i + r_{t+1})P(k_t = i)$
Case of P(B) = 0	
$P(X_2 \in A \mid X_1 = x_1) = \int_A f_{X_2\mid X_1}(x_2 \mid x_1)dx_2$	

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)} = \frac{\partial}{\partial x_2} F_{X_2|X_1}(x_2 \mid x_1)$$

$$F_{X_2|X_1}(x_2 \mid x_1) = \int_{-\infty}^{x_2} f_{X_2|X_1}(x_2|x_1)\,dx_2$$

Continuous X , Continuous Y

Given $F_{X,Y}(x,y)$. Find $P(Y \in B | X \in A)$

- $F_X(x) = F_{X,Y}(x,+\infty)$
- Find $P(X \in A)$ from $F_X(x)$.
- Find $f_{X,Y}(x,y)$ from $F_{X,Y}(x,y)$.
- $P(X \in A, Y \in B) = \int_B \int_A f_{X,Y}(x', y')\,dx'dy'$
- $P(Y \in B | X \in A) = \dfrac{P(X \in A, Y \in B)}{P(X \in A)}$

- $P(Y \in B | X=x) = \int_B f_{Y|X}(y' \mid x)\,dy'$

$$= \lim_{\Delta x \to 0} P(Y \in B \mid x \le X \le x + \Delta x)$$

$$= \frac{\left(\int_B f_{X,Y}(x, y')\,dy'\right)\Delta x}{f_X(x)\Delta x} = \frac{\int_B f_{X,Y}(x, y')\,dy'}{f_X(x)}$$

- $F_{Y|X}(y|x) = \dfrac{\int_{-\infty}^{y} f_{X,Y}(x, y')\,dy'}{f_X(x)}$
- Conditional-probability density of Y given X=x:

$f_{Y|X}(y|x) = \dfrac{d}{dy} F_{Y|X}(y \mid x)$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

- $f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x) = f_{X|Y}(x|y)\, f_Y(y)$

$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x)\,dx$

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)}$$

$$f_Y = \frac{1}{\int_{-\infty}^{\infty} \frac{f_{X|Y}}{f_{Y|X}}\,dx}$$

- $\int_{-\infty}^{\infty} f_{Y|X}(y' \mid x)\,dy' = 1$

Binary Decision-Making

State of the world: Hypothesis

- **H_0 = null hypothesis** → status quo → **target absent**
- **H_1 = alternate hypothesis** → unique alternative to the status quo → **target present**

Prior probability

- $\pi_1 = P(H=H_1) = P(X=H_1)$
- $\pi_0 = P(H=H_0) = P(X=H_0)$

$\pi_1 + \pi_0 = 1$

Make a measurement Y, $f_{Y|H}(y|H=H_i) = f_i(y)$ ⇒ likelihood func.

Likelihood ratio : $\Lambda(y)$ = $\dfrac{f_{Y|H}(y \mid H = H_1)}{f_{Y|H}(y \mid H = H_0)} = \dfrac{f_1(y)}{f_0(y)}$

Threshold : τ = $\dfrac{P(H = H_0)}{P(H = H_1)} = \dfrac{p_0}{p_1}$

Discrete X → Continuous Y

$f_Y(y) = \sum_{i=0}^{1} f_{Y|H}(y \mid H = H_i) P(H = H_i) = f_1(y)p_1 + f_0(y)p_0$

posterior probability:

$$P(H=H_i|Y=y) = \frac{f_{Y|H}(y \mid H = H_i) P(H = H_i)}{\sum_{i=0}^{1} f_{Y|H}(y \mid H = H_i) P(H = H_i)} = \frac{f_i(y)p_i}{f_1(y)p_1 + f_0(y)p_0}$$

- $P(H=H_1|Y=y) + P(H=H_0|Y=y) = 1$

Decision rule: $\hat{X}(Y) = \begin{cases} H_1 & ; y \in S_1 \\ H_0 & ; y \in S_0 = S_1^c \end{cases}$

Minimum error probability design

$\Rightarrow$ choose S_0 to minimize $P(\hat{X}(Y) \neq H)$

$$\begin{aligned} P(\hat{X}(Y) \neq H) &= p_0 P(\hat{X}(Y) = H_1 | H = H_0) + p_1 P(\hat{X}(Y) = H_0 | H = H_1) \\ &= p_0 P(Y \in S_1 | H = H_0) + p_1 P(Y \in S_0 | H = H_1) \\ &= p_0 (1 - P(Y \in S_0 | H = H_0)) + p_1 P(Y \in S_0 | H = H_1) \\ &= p_0 - p_0 \int_{S_0} f_0(y)dy + p_1 \int_{S_0} f_1(y)dy \\ &= p_0 + \int_{S_0} (p_1 f_1(y) - p_0 f_0(y))dy \\ &= p_0 + p_1 \int_{S_0} (\Lambda(y) - t) f_0(y)dy \end{aligned}$$

Want $S_0 = \{y : \Lambda(y) < \tau\}$

S_1's approach

$$\begin{aligned} P(\hat{X}(Y) \neq H) &= p_0 P(Y \in S_1 | H = H_0) + p_1 P(Y \in S_0 | H = H_1) \\ &= p_0 P(Y \in S_1 | H = H_0) + p_1 (1 - P(Y \in S_1 | H = H_1)) \\ &= p_0 \int_{S_1} f_0(y)dy + p_1 - p_1 \int_{S_1} f_1(y)dy \\ &= p_1 - \int_{S_1} (p_1 f_1(y) - p_0 f_0(y))dy \\ &= p_1 - p_1 \int_{S_1} (\Lambda(y) - t) f_0(y)dy \end{aligned}$$

Want $S_1 = \{y : \Lambda(y) \geq \tau\}$

$$\frac{\Lambda(y)}{t} = \frac{p_1 f_{Y|H}(y \mid H = H_1)}{p_0 f_{Y|H}(y \mid H = H_0)} = \frac{\dfrac{p_1 f_{Y|H}(y \mid H = H_1)}{f_Y(y)}}{\dfrac{p_0 f_{Y|H}(y \mid H = H_0)}{f_Y(y)}}$$

$$= \frac{P(H = H_1 | Y = y)}{P(H = H_0 | Y = y)}$$

MAP (maximum a posteriori) rule

$\hat{X}(Y)$ = H_1 if

- $P(H=H_1|Y=y) \geq P(H=H_0|Y=y)$
- **$P(H=H_1|Y=y) \geq 0.5$**

and = H_0 otherwise

performance = $P(\hat{X}(Y) \neq H)$

Hypothesis Testing

Have no information about the prior probabilities $P(H=H_i)$

(Does not know τ)

Know $\Lambda(y)$

Error

- P_{FA} : False alarm : decide $\hat{X}(Y)$ = H_1 when X = H_0
  $= \int_{S_1} f_0(y)\,dy$
- Missed detection : decide $\hat{X}(Y)$ = H_0 when X = H_1

Neyman-pearson Rule :

maximize P_D subject to the upper bound α on P_{FA}

EE310 By Prapun Suksompong

- α = size of the statistical test = fixed P_{FA}
  $= P((\hat{X}(Y)=H_1|X=H_0)$
- detection probability = P_D = β = power of the test
  $= P(\hat{X}(Y)=H_1|X=H_1) = \int_{S_1} f_1(y)\,dy$
- Generally,
  - o $\alpha = 1 \Leftrightarrow \beta = 1$
  - o $\alpha = 0 \Leftrightarrow \beta = 0$ (>0 possible)

NP Solution

When $(\forall i \forall c)\ P(\Lambda(y)=c|H_i) = 0$,

the decision rule: $\hat{X}(Y) = \begin{cases} H_1 & if\ \Lambda(y) \ge t \\ H_0 & if\ \Lambda(y) < t \end{cases}$

has the highest P_D among all decision rules having $P_{FA} \le \alpha$

$\Rightarrow$ likelihood-ratio-threshold rule

- $P(\Lambda(y)\ge\tau|H_0)$
  - is a continuous nonincreasing function of τ
  - $\tau = 0 \to P_D = 1$
  - $\tau = \infty \to P_D = 0$
  - For any $0 < \alpha < 1$,
    there is $\tau = \tau_\alpha$ such that $P(\Lambda(y)\ge\tau|H_0) = \alpha$
- $P_D = P(\Lambda(y)\ge\tau|H_1)$
- $P_D \ge P_{FA}$

$\hat{X}(Y) = H_1$ if $\Lambda(y) \ge \tau_\alpha$

Solve for τ_α from $P_{FA} = \alpha = \displaystyle\int_{\{\Lambda(y)\ge t_a\}} f_0(y)dy$

ROC : Receiver Operating Characteristic

$\Rightarrow P_D = \rho(P_{FA})$

- $\rho(P_{FA})$ is a concave function
- $\rho'(P_{FA}) = \tau_\alpha$

Additive measurement noise

$Y = S + N$

Assumption: S,N are independent $\Rightarrow$ independent additive noise

$f_{S,N}(s,n) = f_S(s)f_N(n)$

$$\begin{pmatrix} S \\ N \end{pmatrix} \to \begin{pmatrix} S \\ Y \end{pmatrix} = \begin{pmatrix} S \\ S+N \end{pmatrix};\ |J| = \begin{vmatrix} \frac{\partial}{\partial s}s & \frac{\partial}{\partial y}s \\ \frac{\partial}{\partial s}y-s & \frac{\partial}{\partial y}y-s \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$$

$f_{S,Y}(s,y) = f_{S,N}(s,y\text{-}s) = f_S(s)f_N(y\text{-}s)$

$$f_{Y|S}(y|s) = \frac{f_{S,Y}(s,y)}{f_S(s)} = \frac{f_S(s)f_N(y-s)}{f_S(s)} = f_N(y-s)$$

Can also get this from

$F_{Y|S}(y|s)$

$= P(Y\le y|S=s) = P(S+N\le y|S=s) = P(s+N\le y|S=s) = P(N\le y\text{-}s|S=s)$

By independent assumption

$P(N\le y\text{-}s|S=s) = P(N\le y\text{-}s) = F_{Y|S}(y|s) = P(N\le y\text{-}s) = F_N(y\text{-}s)$

$f_{Y|S}(y|s) = f_N(y\text{-}s)$

$f_Y(y) = \displaystyle\int_{-\infty}^{\infty} f_{S,Y}(s,y)ds = \int_{-\infty}^{\infty} f_N(y-s)f_S(s)ds = f_N(y) * f_S(y)$

$$f_{S|Y}(s|y) = \frac{f_{Y|S}(y|s)f_S(s)}{\int_{-\infty}^{\infty} f_{Y|S}(y|s')f_S(s')ds'} = \frac{f_N(y-s)f_S(s)}{\int_{-\infty}^{\infty} f_N(y-s')f_S(s')ds'}$$

Infer S by value $\hat{S}(Y) = \arg\max_s\left[f_{S|Y}(s|y)\right]$

$$= \arg\max_s\left[\frac{f_N(y-s)f_S(s)}{f_Y(y)}\right] = \arg\max_s\left[f_N(y-s)f_S(s)\right]$$

Assume $N \sim \mathcal{N}(0,\sigma_n^2)$ and $X \sim \mathcal{N}(m_x,\sigma_x^2)$

$EY = ES + EN = ES$

$$\hat{S}(Y) = \arg\max_s\left[\frac{1}{s_n\sqrt{2p}}e^{-\frac{1}{2}\left(\frac{(y-s)-(m_y-m_s)}{s_n}\right)^2}\frac{1}{s_s\sqrt{2p}}e^{-\frac{1}{2}\left(\frac{S-m_s}{s_s}\right)^2}\right]$$

$$= \arg\min_s\left[\left(\frac{(y-s)}{s_n}\right)^2 + \left(\frac{S-m_s}{s_s}\right)^2\right]$$

$$= \frac{s_n^2 m_s + s_s^2 y}{s_n^2 + s_s^2} = \frac{m_s + (SNR)y}{1+(SNR)}$$

- SNR $\to 0 \Rightarrow \hat{S}(Y)$ = $m_s \Rightarrow$ ignore the measurement
- SNR $\to \infty \Rightarrow \hat{S}(Y)$ = y $\Rightarrow$ ignore noise (good measurements)

If
- $f_N(n)$ is a sharply peaked function
- $f_S(s)$ is slowly varying

then

$\hat{S}(Y) \approx \arg\max_s \left[f_N(y-s) \right] \to$ not require $f_S(s)$

maximum likelihood (ML) rule

$$\tilde{S}(Y) = \arg\max_s \left[f_N(y-s) \right]$$

Multivariable

$$P\left(Y_1^n \in A \mid X_1^m = x_1^m\right) = \int_A f_{Y_1^n|X_1^m}\left(y_1^n \mid x_1^m\right) dy_1 \ldots dy_n$$

$$\frac{\partial^n}{\partial y_1 \ldots \partial y_n} F_{Y_1^n|X_1^m}\left(y_1^n \middle| x_1^m\right) = f_{Y_1^n|X_1^m}\left(y_1^n \middle| x_1^m\right)$$

Conditional CDF

$$F_{Y_1^n|X_1^m}\left(y_1^n \middle| x_1^m\right) = \int_{-\infty}^{y_1} dy_1' \cdots \int_{-\infty}^{y_n} dy_n' \, f_{Y_1^n|X_1^m}\left(y_1'^n \middle| x_1^m\right)$$

$$F_{Y_1^n|X_1^m}\left(y_1^n \middle| x_1^m\right) = \int_{-\infty}^{y_1} dy_1' \cdots \int_{-\infty}^{y_n} dy_n' \, \frac{f_{X_1^m, Y_1^n}\left(x_1^m, y_1'^n\right)}{f_{X_1^m}\left(x_1^m\right)}$$

Product/independent model:

$$f_{X_1^m, Y_1^n}\left(x_1^m, y_1^n\right) = \left(\prod_{i=1}^m f_{X_i}(x_i)\right) \cdot \left(\prod_{i=1}^n f_{Y_i}(y_i)\right)$$

$$f_{Y_1^n|X_1^m}\left(y_1^n \middle| x_1^m\right) = \frac{f_{X_1^m, Y_1^n}\left(x_1^m, y_1^n\right)}{f_{X_1^m}\left(x_1^m\right)} = \frac{\left(\prod_{i=1}^m f_{X_i}(x_i)\right) \cdot \left(\prod_{i=1}^n f_{Y_i}(y_i)\right)}{\left(\prod_{i=1}^m f_{X_i}(x_i)\right)}$$

$$= \prod_{i=1}^n f_{Y_i}(y_i)$$

Markov Dependence: ($\forall$i > 1) $f_{X_i|X_1^{i-1}} = f_{X_i|X_{i-1}}$

- m > 0 $\to$ $f_{X_{m+1}^n|X_1^m} = \prod_{i=m+1}^n f_{X_i|X_{i-1}}$

Independence

unlinked :
- without a causal connection in a physical setting
- without one outcome being informative about another outcome in an information-theoretic or belief-based setting

P[X$\in$A,Y$\in$B] = P[X$\in$A] P[Y$\in$B]

P[X$\le$x, Y$\le$y] = P[X$\le$x] P[Y$\le$y]

$F_{X,Y}(x,y) = F_X(x) F_Y(y)$

P[X=x_i, Y=y_i] = P[X=x_i] P[Y=y_i]

two variables

A and B are **independent** : A$\perp\!\!\!\perp$B

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

- $A \perp\!\!\!\perp B$ if and only if either
  - $P(A) = 0$ or
  - $P(B) = 0$ or
  - $P(B|A) = P(B)$ and $P(A|B) = P(A)$
- $A \perp\!\!\!\perp B$ & $P(B) > 0 \to P(A|B) = P(A)$
  $B \perp\!\!\!\perp A$ & $P(A) > 0 \to P(B|A) = P(B)$
- $B \perp\!\!\!\perp A \Leftrightarrow A \perp\!\!\!\perp B$
- ($A \perp\!\!\!\perp A \Leftrightarrow P(A) = 0$ or 1: trivial) $\Rightarrow \forall B, A \perp\!\!\!\perp B$
  - $P(A) = P(A)^2$
  - Ex. $\varnothing$, Ω is independent of all other events
  - $P(A) = 0$ or $1 \Leftrightarrow A \perp\!\!\!\perp B$ and $A \perp B$
- $A \perp\!\!\!\perp B \Leftrightarrow A \perp\!\!\!\perp B^c \Leftrightarrow A^c \perp\!\!\!\perp B \Leftrightarrow A^c \perp\!\!\!\perp B^c$
  - $P(B) = P(A \cap B) + P(A^c \cap B)$
    $P(A^c \cap B) = P(B) - P(A)P(B) = P(A^c)P(B)$
- two events are nontrivially independent only if they overlap properly
- $A \perp B$ & $A \perp\!\!\!\perp B \Rightarrow P(A) = 0$ or $P(B) = 0$
- If $P(A) \neq 0$ and $P(B) \neq 0$, then A and B cannot be both mutually exclusive and statistically independent.
- $A \subset B$ & $A \perp\!\!\!\perp B \Rightarrow P(A) = 0$ or $P(B) = 1$
- if $\|\Omega\| < 4$, then there are no nontrivially independent events
  - For $\|\Omega\| = 3$, if $A = \{\omega_1\}$, then
    - $B = \{\omega_2\}$ or $\{\omega_3\}$ or $\{\omega_2, \omega_3\} \to A \perp B$
    - $B = \{\omega_1, \omega_2\}$ or $\{\omega_1, \omega_3\} \to A \subset B$

- Independence implies a lack of covariance
  - It is not the case that a lack of covariance always implies independence because covariance only measures linear association.

- $\perp\!\!\!\perp$ and $\perp$
  - A and A^c are mutually exclusive. However, they are not independent (since if one occurs, the other cannot).

- The null event is statistically independent of any other event.
  - $P(A) = 0 \Rightarrow$ A is statistically independent of any other event B in that sample space.

Multivariable

Boolean function $f(A_1,\ldots,A_n) = F$ is constructed through iterated use of Boolean set operations of complementation, union, or intersection

events $\{A_1,\ldots,A_n\}$ are mutually independent: $\coprod_1^n A_i \Leftrightarrow$

$$(\forall I \subset \{1,\ldots,n\}) \quad P\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} P(A_i)$$

The event $A_1, A_2, \ldots, A_n$ are said to be mutually independent if and only if the relations

$P(A_i \cap A_j) = P(A_i)P(A_j)$

$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)\, P(A_k)$

……………

$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2)\cdots P(A_n)$

Hold for all combinations of the indices such that

$1 \le i < j < k < \cdots \le n$

- trivially true for $\|I\| < 2$
- $(\forall I \subset \{1,\ldots,n\})(\forall j \notin I) \quad P\left(A_j \,\middle|\, \bigcap_{i\in I} A_i\right) = P(A_j)$
- symmetrical $\Rightarrow$ ordering of events is irrelevant
- Let $B_i = A_i$ or A_i^c, $\coprod_1^n A_i \Leftrightarrow \coprod_1^n B_i$
- Given two nonoverlaping collections of event drawn from a larger collection of mutually independent events,
  the two new set, formed by choosing arbitrary Boolean functions defined on the two collections,
  will themselves be independent

Assume $\coprod_{1}^{n} A_i$
• occurrence of all of these events $P\left(\bigcap_{1}^{n} A_i\right) = \prod_{1}^{n} P(A_i)$
• non-occurrence of all of them $P\left(\bigcap_{1}^{n} A_i^c\right) = \prod_{1}^{n} P(A_i^c) = \prod_{1}^{n} (1 - P(A_i))$
• occurrence of at least one of them $P\left(\bigcup_{1}^{n} A_i\right) = 1 - P\left(\bigcap_{1}^{n} A_i^c\right) = 1 - \prod_{1}^{n} P(A_i^c) = 1 - \prod_{1}^{n} (1 - P(A_i))$
• occurrence of exactly one event, no matter which one $P\left(\bigcup_{i=1}^{n}\left(A_i \cap \bigcap_{j \neq i} A_j^c\right)\right) = \sum_{i=1}^{n} P\left(\left(A_i \cap \bigcap_{j \neq i} A_j^c\right)\right)$ $= \sum_{i=1}^{n} P(A_i) \prod_{j \neq j} (1 - P(A_j))$ $= \left(\prod_{i=1}^{n} (1 - P(A_i))\right)\left(\sum_{k=1}^{n} \frac{P(A_k)}{1 - P(A_k)}\right)$

Independent experiments Suppose that we are concerned with the outcomes of n different experiment $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$. Suppose further that the sample space Ω_k of the k[th] of these n experiments is partitioned by the m_k events A_{ki} , $i_k = 1, 2, \ldots, m_k$. The n given experiments are then said to be statistically independent if and only if the equation $P\left(A_{1i_1} \bigcap A_{2i_2} \bigcap \ldots \bigcap A_{ni_n}\right) = P\left(A_{1i_1}\right) P\left(A_{2i_2}\right) \cdots P\left(A_{ni_n}\right)$ holds for every possible set of n integers {i , i_2 n}, where the k ranges from 1 to m .
• $F_{X_a^b}\left(x_a^b\right) = F_{X_a, X_{a+1}, \ldots, X_b}\left(x_a, x_{a+1}, \ldots, x_b\right)$; b a $F_{X_a^a}\left(x_a^a\right) = F_{X_a}\left(x_a\right)$
$_1,\ldots,$X } to be independent if the individual random experiments E ,…,E_n independent
$\coprod_{i=1}^{n} X_i$ $\Leftrightarrow F_{X_1^n}\left(x_1^n\right) = \prod_{i=1}^{n} F_{X_i}\left(x_i\right)$ $\Leftrightarrow f_{X_1^n}\left(x_1^n\right) = \prod_{i=1}^{n} f_{X_i}\left(x_i\right)$ $\Leftrightarrow f_{X_{m+1}^n \mid X_1^m}\left(x_{m+1}^n \mid x_1^m\right) = f_{X_{m+1}^n}\left(x_{m+1}^n\right)$; $\forall$(m,n) $\Leftrightarrow$ P(X_1=x_1,…,X_n=x_n) = $\prod_{i=1}^{n} P\left(X_i = x_i\right)$

- $E\left(\prod_1^n X_i\right) = \prod_1^n EX_i$
- $E\left(\prod_1^n h_i(X_i)\right) = \prod_1^n Eh_i(X_i)$
- $E(X_i X_j) = \begin{cases} EX_i EX_j & i \neq j \\ EX_i^2 & i = j \end{cases}$
- independence $\rightarrow$ uncorrelatedness (the converse fails)
    - o two random variables can be uncorrelated even though the one determines the other

i.i.d : independent and identically distributed

$\Rightarrow$ mutually independent and have a common F_X

$\Rightarrow (\forall i)\ F_{X_i}(x) = F_X(x)$

- $f_{X_1^n}(x_1^n) = \prod_{i=1}^n f_X(x_i)$
- $P(X_1=x_1,\ldots,X_n=x_n) = \prod_{i=1}^n P(X = x_i)$

Let

X be a two-dimentional random vector (X_1, X_2) whose components X_1 and X_2 are independent random variables.

Let

Y be the two-dimensional random vector (Y_1,Y_2) whose components are the random variables

$Y_1 = g_1(X_1)$ and

$Y_2 = g_2(X_2)$

respectively;

where g_1 and g_2 are Borel functions defined on Ω_{X1} and Ω_{X2} respectively.

$\rightarrow$

random variables Y_1 and Y_2 are statistically independent.

Maxima and Minima of Random Variables

$Y = \min X_i$

$Z = \max X_i$

$F_Z(z) = P(Z \le z) = P(\max X_i \le z) = P(\ (\forall i)\ X_i \le z\ ) = P\left(\bigcap_i A_i\right)$

A_i is the event that $X_i \le z \rightarrow$ i.i.d

$P(A_i) = F_{X_i}(z) = F_X(z)$

$F_Z(z) = P\left(\bigcap_i A_i\right) = \prod_i P(A_i) = (F_X(z))^n$

$f_Z(z) = n(F_X(z))^{n-1} f_X(z)$

For large n,

assume median μ_Z to be large

$\rightarrow F_X(\mu_Z) \approx 1$

$F_Z(\mu_Z) = \frac{1}{2} = e^{-\ln 2}$

$= (\ F_X(\mu_Z)\ )^n = (\ 1-(\ 1-F_X(\mu_Z)\ )\ )^n = (\ 1-x\ )^n$

$\approx e^{-x} = e^{-n(1-F_X(m_Z))}$

$F_X(\mu_Z) = 1-\frac{\ln 2}{n}$

$F_Y(y) = P(Y \le y) = P(\min X_i \le y) = P(\ (\exists i)\ X_i \le z\ ) = P\left(\bigcup_i A_i\right)$

$= 1 - P\left(\bigcap_i A_i^c\right) = 1 - \prod_i P(A_i^c) = 1 - (1 - F_X(y))^n$

$f_Y(y) = n(1 - F_X(y))^{n-1} f_X(y)$

For large n, $F_X(\mu_Y)$ will be small

$F_Y(\mu_Y) = \frac{1}{2} = e^{-\ln 2}$

$=1-(\ 1-F_X(\mu_Y)\ )^n = 1-(1-x)^n \approx 1-e^{-x} = 1-e^{-nF_X(m_Y)}$

$F_X(\mu_Y) = \frac{\ln 2}{n}$