

Supplemental Notes

This content is protected and copyright reserved. It may not be shared, copied, distributed, uploaded, or otherwise used without the prior express written authorization of Dr. Brandon Franzke.

Rules of Inference

(the logical schema)

modus Ponens

(Modus ponendo ponens)

"mode that affirms by affirming"

Premise 1: $P \rightarrow Q$

Premise 2: P

Conclusion: Q

Theorem: $[(P \rightarrow Q) \wedge P] \rightarrow Q$

Prf: By truth table

P	Q	$(P \rightarrow Q) \wedge P \rightarrow Q$					
1	1	1	1	1	1	1	1
0	1	0	1	0	0	1	1
1	0	1	0	0	1	1	0
0	0	0	1	0	0	1	0

Diagram illustrating the truth table for the theorem $[(P \rightarrow Q) \wedge P] \rightarrow Q$. The table shows the truth values of P, Q, and the logical expression $(P \rightarrow Q) \wedge P \rightarrow Q$. The final column, representing the truth value of the entire expression, is circled in red. Red arrows point from the circled column to the first two columns (P and Q) in the first row, indicating the logical flow or the conditions under which the expression is true.

Modus Tollens

(modus tollendo tollens)

"mode that denies by denying"

Premise 1: $P \longrightarrow Q$

Premise 2: $\sim Q$

Conclusion: $\sim P$

Defn: A fallacy is a logical error
(not factual error)

Theorem: $[(P \longrightarrow Q) \& \sim Q] \longrightarrow \sim P$

truth $\swarrow$ factual truth
 $\searrow$ logical truth

"grass is green"

"green is green" $\leftarrow$ tautology

Fallacy of affirming the consequent

P1: $P \longrightarrow Q$

P2: Q

C: $\therefore P$

Fallacy of denying the antecedent.

P1: $P \longrightarrow Q$

P2: $\sim P$

C: $\therefore \sim Q$

Conditional:

$$P \longrightarrow Q$$

Contra positive:

$$\sim Q \longrightarrow \sim P$$

← same logical truth
as conditional.

Converse:

$$Q \longrightarrow P$$

i.e. $[P \longrightarrow Q] \longleftrightarrow [\sim Q \longrightarrow \sim P]$.

Theorem: "Material implication"

$$P \longrightarrow Q = \sim P \vee Q$$

Defn: An argument is VALID iff the premises
logically imply the conclusion

(can't make if-part TRUE and then-part FALSE)

A proposition is valid by structure, not content

Defn: An argument is SOUND iff it is valid and all
premises are true

Ex: $[(P \vee Q) \rightarrow R] \leftrightarrow [P \rightarrow (Q \rightarrow R)]$.

P	Q	R	$[(P \vee Q) \rightarrow R] \leftrightarrow [P \rightarrow (Q \rightarrow R)]$											
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	1	1	0	1	1	1	1	1	0	1	1	1	1	1
1	0	1	1	1	0	1	1	1	1	1	0	1	1	1
0	0	1	0	0	0	1	1	1	0	1	0	1	1	1
1	1	0	1	1	1	0	0	1	1	0	1	0	0	0
0	1	0	0	1	1	0	0	0	0	1	1	0	0	0
1	0	0	1	1	0	0	0	0	1	1	0	1	0	0
0	0	0	0	0	0	1	0	1	0	1	0	1	0	0

1 2 3 4 1 3 2 1

$\therefore$ False (since the "final" column is not all 1's)

Defn: Suppose $(\Omega, \mathcal{Q}, P)$ is a probability space.
Events A and B are (statistically) independent
w.r.t. P iff $P(A \cap B) = P(A) \cdot P(B)$

"joint factors into marginals"

Defn: Events $A_1, A_2, \dots$ are independent $\longleftrightarrow$ every finite subset is independent.

Q: Can A be independent of itself?

A: iff $P(\underbrace{A \cap A}_A) = P(A) \cdot P(A) \longleftrightarrow P(A) = P(A)^2$
 $\therefore$ only if $P[A] = 0$ or $P[A] = 1$.

Q: Can A be independent of X ?

A: iff $P(\underbrace{A \cap X}_A) = P(A) \cdot \underbrace{P(X)}_{=1 \text{ (cert)}} \longleftrightarrow P(A) = P(A)$
 $\therefore$ every set is independent of X .

Q: Does A indep. $B \longrightarrow A \cap B = \emptyset$? (they are disjoint).

A: iff $P(\underbrace{A \cap B}_{\emptyset \text{ by assump.}}) = P(A) \cdot P(B) \longleftrightarrow P[A] = 0 \text{ or } P[B] = 0$.

no in general

Ex: Coin flip w/ $P[\text{head}] = p$

$$\therefore P[\text{tail}] = 1-p.$$

(independent flips)

Flip twice:

$$P[H_1 \cap H_2] = P[H_1] \cdot P[H_2] = P[H]^2 = p^2$$

$$P[H_1 \cap T_2] = P[H_1] \cdot P[T_2] = P[H] \cdot P[T] = p(1-p)$$

$$P[T_1 \cap H_2] = P[T_1] \cdot P[H_2] = P[T] \cdot P[H] = (1-p)p$$

$$P[T_1 \cap T_2] = P[T_1] \cdot P[T_2] = P[T]^2 = (1-p)^2$$

Note: $P[\{H_1 \cap H_2\} \cup \{H_1 \cap T_2\} \cup \{T_1 \cap H_2\} \cup \{T_1 \cap T_2\}]$

$$\downarrow = p^2 + 2p(1-p) + (1-p)^2$$

$$= P[\Omega] = 1$$

$$\therefore = 1.$$

Q: $P[1 \text{ tail AND } 1 \text{ head (any order)}]$

$$\begin{aligned} &= P[\{H_1 \cap T_2\} \cup \{T_1 \cap H_2\}] \stackrel{\text{disjoint CA}}{=} P[H_1 \cap T_2] + P[T_1 \cap H_2] \\ &= 2p(1-p) \end{aligned}$$

Q: $P[\text{at least 1 tail}] = P[\{H_1 \cap T_2\} \cup \{T_1 \cap H_2\} \cup \{T_1 \cap T_2\}]$

$$= 2p(1-p) + (1-p)^2$$

$$= 1 - P[(\text{at least 1 tail})^c]$$

$$= 1 - P[\text{no head}]$$

$$= 1 - p^2$$

Defn: Conditional Probability $P(\underbrace{B|A}_{\text{conditional}}) = \frac{P(\overbrace{A \cap B}^{\text{joint}})}{\underbrace{P(A)}_{\text{marginal}}}$ if $P(A) > 0$.

$$\begin{aligned} \therefore A \text{ \& B independent } &\longleftrightarrow P(B|A) = P(B) \\ &\longleftrightarrow P(A|B) = P(A) \end{aligned}$$

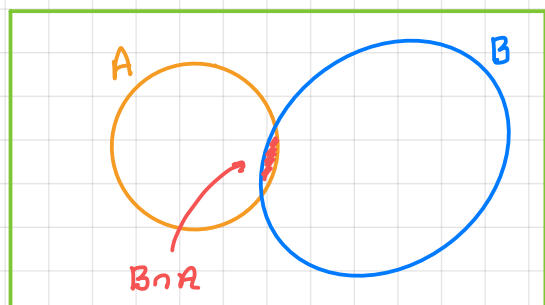
Prf: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot \cancel{P(B)}}{\cancel{P(B)}} = P(A).$

Important: $P[A] = 0 \not\rightarrow A = \emptyset$ ($A = \emptyset \rightarrow P[A] = 0$)

Note: Conditional probability is not transitive
 $P[A|B] \neq P[B|A]$

Intuition: If you know A occurred then you constrain the size of sample space of B

$B \cap A$
 $B^c \cap A$



the value $P[A]$ normalizes area to 1.

Ex: Urn experiment w/ 4-numbered balls: #1 and #2 blue
#3 and #4 red

Events
A: blue ball
B: even ball
C: # ball > 2

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2} \quad (\text{note} = P[A])$$

"blue given even"

$$P[A|C] = \frac{P[A \cap C]}{P[C]} = \frac{0}{\frac{1}{2}} = 0 \quad (\text{note} \neq P[A])$$

"blue given # > 2"

$\therefore$ conditioning can (but may not) affect computed probability.

★ Note: $P[A \cap B] = P[A|B] \cdot P[B]$
 $= P[B|A] \cdot P[A]$

Ex: Urn experiment w/ 5-balls, 2 blue, 3 red

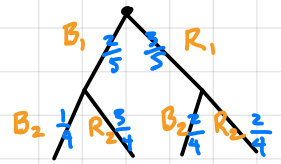
Draw 2 balls: $P[2 \text{ blue balls}]$

Consider as 2 "sub" experiments

$$P[B_1 \cap B_2] = P[B_1] \cdot P[B_2 | B_1]$$

$$= \frac{2}{5} \cdot \frac{1}{4}$$

$$= \frac{1}{10}$$



Thrm: Suppose A and B are independent then: $P[A|B] = P[A]$

$$P[B|A] = P[B]$$

Prf: $P[B|A] = \frac{P[A \cap B]}{P[A]} = \frac{P[B] \cdot \cancel{P[A]}}{\cancel{P[A]}} = P[B]$

and similarly for $P[A|B]$.

★ Conditioning does not affect the probability of independent events.

Thrm: Probabilistic Modus Ponens

$$\text{Premise 1: } P(B|A) \geq c$$

$$\text{Premise 2: } P(A) \geq a$$

$$\text{Conclusion: } P(B) \geq a \cdot c$$

$$\text{Prf: } P(B) \geq P(A \cap B)$$

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

$$= P(A) \cdot P(B|A)$$

$$\geq a \cdot c$$

monotonicity

$$P(A) > 0$$

defn cond. prob.

premise 1 and 2.

QED

$$\text{Check: } a = c = 1 \quad \therefore P(B) = 1$$

Thrm: Probabilistic Modus Tollens

$$\text{Premise 1: } P(B|A) \geq c > 0$$

$$\text{Premise 2: } P(B) \leq b.$$

$$\text{Conclusion: } P(A) \leq \min\left(1, \frac{b}{c}\right)$$

$$\text{Prf: } P(A) \leq P(A) \left(\frac{P(B)}{P(A \cap B)} \right) \quad \text{since } P(A \cap B) \leq P(B)$$

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

$$\leq \frac{b}{P(B|A)}$$

premise 1.

$$\leq \frac{b}{c}$$

premise 2

$$\text{Always } P(A) \leq 1 \quad \therefore P(A) \leq \min\left(1, \frac{b}{c}\right)$$

QED.

Proof by induction. Prove $A(n)$ holds $\forall n$.

(1) Prove $A(1)$

(2) Show $A(n) \rightarrow A(n+1)$

Ex: $n^2 + n$ even $\forall n$

• Basis, $n=1$ $1^2 + 1 = 2 = 2 \cdot k$ for $k=1$ $\therefore$ even QED basis

• Induction, Suppose $n^2 + n = 2 \cdot k'$ for some $k' \in \mathbb{Z}^+$.

$$(n+1)^2 + (n+1) = n^2 + 2n + 1 + n + 1 = \underbrace{n^2 + n}_{2k'} + 2 \underbrace{(n+1)}_{k''} = 2(k' + k'') \quad \text{QED.}$$

Theorem: ("Booles inequality") $P\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n P(A_k)$

Prf: (by induction on $n=2, 3, 4, \dots$)

Basis Step: $n=2$.

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(A_1 \cap A_2) && \text{Add. theorem} \\ &\leq P(A_1) + P(A_2) = \sum_{k=1}^2 P(A_k) && P[A] \geq 0 \end{aligned}$$

Induction Step: - Assume holds for n . QED-basis
- Derive that it holds for $n+1$

Induction Hypothesis (IH).

$$P\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n P(A_k)$$

$$\begin{aligned} \Rightarrow P\left(\bigcup_{k=1}^{n+1} A_k\right) &= P\left(\left(\bigcup_{k=1}^n A_k\right) \cup A_{n+1}\right) \\ &\leq P\left(\bigcup_{k=1}^n A_k\right) + P(A_{n+1}) \\ &\leq \sum_{k=1}^n P(A_k) + P(A_{n+1}) \\ &= \sum_{k=1}^{n+1} P(A_k) \end{aligned}$$

QED

Thm: ("Inclusion-Exclusion")

Note: generalizes CA.

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n)$$

Technique: Unions are usually hard to work with, $\therefore$ so either

1) De Morgan's

2) Additivity (find disjoint)

Theorem: ("Multiplication theorem")

$$P\left(\bigcap_{k=1}^n A_k\right) = P(A_1) \cdot P(A_2 | A_1) \cdots P(A_n | A_1 \cap \dots \cap A_{n-1})$$

no dependence.



Compare to: Independence: $= P(A_1) \cdot P(A_2) \cdots P(A_n) = \prod_{k=1}^n P(A_k)$

Markov: $= P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_2) \cdots P(A_n | A_{n-1})$

↑
depend only on
previous 1.

$$= P(A_1) \cdot \prod_{k=2}^n P(A_k | A_{k-1})$$

Ex: Drawing 3 aces

w/ replacement (independent samples)

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$$

w/o replacement

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 | A_1)$$

$$P[\text{draw 3 aces}] = P(A_1 \cap A_2 \cap A_3)$$

w/o replacement

$$= P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2)$$

$$= \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50}$$

$$= \frac{1}{5525} \quad (\approx 0.00018)$$

Prf: (by induction on $n = 2, 3, 4, \dots$)

Basis Step: $n = 2$.

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 | A_1) \quad \text{by defn cond. prob.}$$

QED-basis

Induction Step:

- Assume holds for n .
- Derive that it holds for $n+1$

Induction Hypothesis (IH).

$$P\left(\bigcap_{k=1}^n A_k\right) = P(A_1) \cdot P(A_2 | A_1) \cdots P(A_n | A_1 \cap \dots \cap A_{n-1})$$

$$\begin{aligned} \Rightarrow P\left(\bigcap_{k=1}^{n+1} A_k\right) &= P\left(\bigcap_{k=1}^n A_k \cap A_{n+1}\right) \\ &\stackrel{\text{assoc.}}{=} P\left(\left(\bigcap_{k=1}^n A_k\right) \cap A_{n+1}\right) \\ &= P\left(\bigcap_{k=1}^n A_k\right) \cdot P(A_{n+1} | \bigcap_{k=1}^n A_k) \\ &= P(A_1) \cdot P(A_2 | A_1) \cdots P(A_n | A_1 \cap \dots \cap A_{n-1}) \\ &\quad \cdot P(A_{n+1} | \bigcap_{k=1}^n A_k) \end{aligned}$$

QED.

Defn: Suppose $(\Omega, \mathcal{Q})$ is a measurable space. Then

$\{H_k\} \subset \mathcal{Q}$ partition Ω ("is a partition") iff

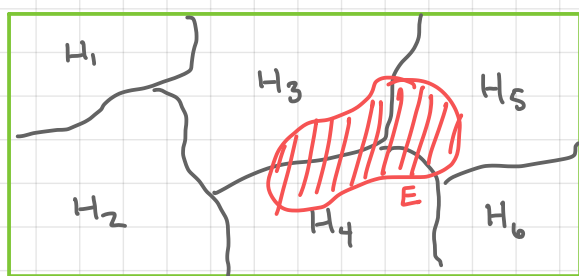
union that
does not overlap
 $\therefore CA$

$$(1) \bigcup_k H_k = \Omega$$

$$(2) \text{ pairwise-disjoint } H_i \cap H_j = \emptyset \text{ if } i \neq j$$

Theorem: (of Total Probability)

$$P(E) = \sum_k P(H_k) \cdot P(E|H_k) \text{ if } \{H_k\} \text{ partition } \Omega$$



x

Hypotheses: H_k

Evidence: E

Prf: $P(E) = P[E \cap \Omega]$

$$= P\left[E \cap \left(\bigcup_k H_k\right)\right]$$

since $\{H_k\}$ partition Ω

$$= P\left(\bigcup_k (E \cap H_k)\right)$$

$$= \sum_k P(E \cap H_k)$$

since $\{H_k\}$ partition Ω

$$= \sum_k P(H_k) \cdot P(E|H_k)$$

defn of cond. prob.

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

QED

Ex: Urn w/ 5-balls, 2 blue, 3 red

$P[\text{second ball is red}]$

events: $B_1 = \{b_1, n r_2, b_1, n b_2\}$
 $\sim B_1 = R_1 = \{r_1, n r_2, r_1, n r_2\}$ } partition: B_1 vs. $\sim B_1$

$$\begin{aligned}\therefore P[r_2] &= P[r_2 | b_1] \cdot P[b_1] + P[r_2 | r_1] \cdot P[r_1] \\ &= \frac{3}{4} \cdot \frac{2}{5} + \frac{1}{2} \cdot \frac{3}{5} \\ &= \frac{3}{5} \quad (60\%) \end{aligned}$$

Note: (another representation)

Data = Evidence

Hypotheses $H_1, H_2, \dots, H_n$ partition sample space

$$P[E] = \sum_{k=1}^n \underbrace{P[E | H_k]}_{\text{"likelihood"}} \cdot \underbrace{P[H_k]}_{\text{"prior"}}$$

Later in this course:

$$\left\{ \begin{array}{ll} \text{Total Expectation} & E_x[X] = E_Y[E[X|Y]] \\ \text{Total Variance} & V_x[X] = E_Y[V[X|Y]] + V_Y[E[X|Y]]. \end{array} \right.$$

extremely important for problem solving

Special case: $\Omega = A \cup A^c$

$$\therefore P(B) \stackrel{\text{total prob.}}{=} P(A) \cdot P(B|A) + P(A^c) \cdot P(B|A^c)$$

Note: for any B.

Thm: $P[B|A] = \frac{P[A|B] \cdot P[B]}{P[A]}$ for any event B. (if $P[A] > 0$)

Prf: $P[B|A] = \frac{P[A \cap B]}{P[A]} \cdot \frac{P[B]}{P[B]} = \frac{P[A|B] \cdot P[B]}{P[A]}$ QED.

Theorem: (Bayes' Theorem)

$$P(H_j | E) = \frac{P(E|H_j) \cdot P(H_j)}{\sum_k P(E|H_k) \cdot P(H_k)} \quad \text{if } \{H_k\} \text{ partitions } \Omega.$$

Prf: $P(H_j | E) = \frac{P(E \cap H_j)}{P(E)}$ defn of cond. prob

$$= \frac{P(H_j) \cdot P(E|H_j)}{P(E)} \quad \text{" " "}$$
$$= \frac{P(H_j) \cdot P(E|H_j)}{\sum_k P(H_k) \cdot P(E|H_k)} \quad \text{total prob.}$$

QED

$P(H_j)$: Prior

$P(H_j | E)$: Posterior

$P(E|H_j)$: Likelihood

$$\therefore P(H|E) = \frac{a}{a+b}$$

$$P(H^c|E) = \frac{b}{a+b}$$

IMAC format

I: Issue

- associative memory ("spot the issue")

M: Math rule

- rote memorization

A: Apply math to facts (Analysis)

- pattern matching

- element because fact.

C: Conclusion

- deductive logic

Mnemonic:

"Party Unconditionally To Conquer Bayes"

P: Partition $\{H_k\}$ or $A \cup A^c$

U: Unconditional probability $P(B)$

T: Total Probability

C: Conditional Probability $P(B|A)$

B: Bayes Theorem

Issue Spotting Sequence

- Is there a partition?

- Unconditional probability $P(B)$?

$\therefore$ Use total probability

- Conditional probability $P(B|A)$?

$\therefore$ Use Bayes Theorem

Ex: Special case H vs H^c

$$H \cup H^c = X$$

$$H \cap H^c = \emptyset$$

$\therefore H$ and H^c partition X

$$P[H|E] = \frac{P[E|H] \cdot P[H]}{P[E|H] \cdot P[H] + P[E|H^c] \cdot P[H^c]}$$

hypothesis
evidence

Ex: Cancer (H) vs. not Cancer (H^c)

Evidence = Test w/ diagnostic:

prior $\rightarrow$ 1% of women have breast cancer

test $\left\{ \begin{array}{l} 80\% \text{ detect cancer when it is there} \\ 9.6\% \text{ detect cancer when it is not there} \end{array} \right.$

$\therefore$ 99% do not

20% false neg.

90.4% true neg.

	(1%) Cancer H		(99%) ~Cancer H^c	
Test	+	TP 0.80	FP 0.096	
	-	FN 0.20	TN 0.904	
		$\downarrow$ $\Sigma = 1$	$\downarrow$ $\Sigma = 1$	

$$P[\text{Cancer} | +] = \frac{P[+ | \text{Cancer}] \cdot P[\text{Cancer}]}{P[+ | \text{Cancer}] \cdot P[\text{Cancer}] + P[+ | \sim \text{Cancer}] \cdot P[\sim \text{Cancer}]}$$

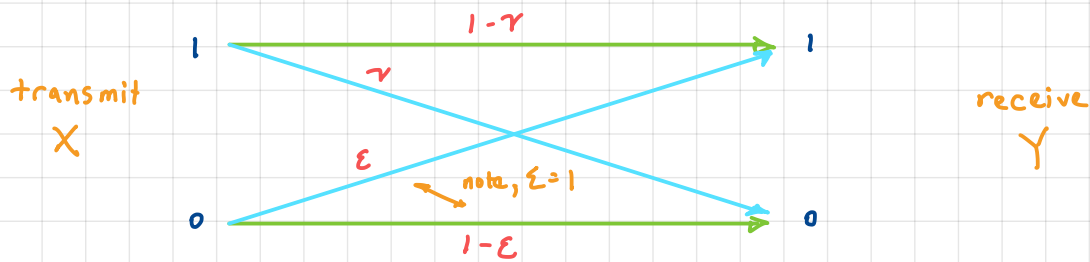
$$= \frac{(0.80)(0.01)}{(0.80)(0.01) + (0.96)(0.99)}$$

$$= 0.0776 \quad (\sim 7.76\%)$$

Note: $\therefore$ get second opinion.

Ex: Binary Channel

Problem: Transmit a single noisy bit. Guess transmitted given received.



$$P[Y=1 | X=1] = 1-r$$

$$P[Y=1 | X=0] = \epsilon$$

$$P[Y=0 | X=1] = r$$

$$P[Y=0 | X=0] = 1-\epsilon$$

Defn: A binary symmetric channel has $r = \epsilon$.

Q: Suppose $Y=1$ (received 1). Guess $X=0$ or $X=1$?

A: It depends on r and ϵ .

Choose hypothesis ($X=0$ or 1) that has higher probability, i.e.

$$\begin{aligned} \text{estimate } \hat{X} &= 1 && \text{if } P[X=1 | Y=1] > P[X=0 | Y=1] \\ \hat{X} &= 0 && \text{else} \end{aligned}$$

"pick biggest"

Defn: The MAP (maximum a priori) estimate is the hypothesis that maximizes the probability given the observation: $\hat{H}^{\text{MAP}} = \arg \max_k P(H_k | E)$

In our problem: $P[X=1 | Y=1] \geq P[X=0 | Y=1]$

Bayes:
$$\frac{P(Y=1 | X=1) \cdot P(X=1)}{P(Y=1 | X=1) \cdot P(X=1) + P(Y=1 | X=0) \cdot P(X=0)} \stackrel{\text{iff}}{\geq} \frac{P(Y=1 | X=0) \cdot P(X=0)}{P(Y=1 | X=1) \cdot P(X=1) + P(Y=1 | X=0) \cdot P(X=0)}$$

$$\overbrace{P(Y=1|X=1)}^{1-\gamma} \cdot \overbrace{P(X=1)}^p \geq \overbrace{P(Y=1|X=0)}^{\varepsilon} \cdot \overbrace{P(X=0)}^{1-p}$$

odds

iff $\frac{P(X=1)}{P(X=0)} \geq \frac{\varepsilon}{1-\gamma}$

Suppose $P(X=1) = P(X=0)$ $\Rightarrow$ $\frac{0.50}{0.50} \geq \frac{\varepsilon}{1-\gamma}$

transmit 50% 1 and 50% 0

$\therefore$ if $Y=1$, choose $\hat{X}=1$ if $\frac{\varepsilon}{1-\gamma} < 1$.

$\hat{X}=0$ if $\frac{\varepsilon}{1-\gamma} > 1$

MAP hypothesis test.

$$P[\text{flip}] < P[\text{not flip}]$$

$$\downarrow$$

$$\varepsilon < 1-\gamma$$

$$\varepsilon > 1-\gamma$$

$$\uparrow$$

$$P[\text{flip}] > P[\text{not flip}]$$