4. Bayesian Inference, Part II: Hypothesis Testing

ECE 302 Fall 2009 TR 3-4:15pm
Purdue University, School of ECE
Prof. Ilya Pollak
Hypothesis Testing

• Suppose the unobserved random variable $X$ is discrete and can assume a finite number of values, say, 1,2,...,$m$. 
Hypothesis Testing

• Suppose the unobserved random variable $X$ is discrete and can assume a finite number of values, say, 1,2,...,$m$.

• Then the problem of estimating $X$ from some observation $Y$ is called an $m$-ary hypothesis testing problem.
Hypothesis Testing

• Suppose the unobserved random variable $X$ is discrete and can assume a finite number of values, say, $1,2,...,m$.

• Then the problem of estimating $X$ from some observation $Y$ is called an $m$-ary hypothesis testing problem.

• The $m$ hypotheses are $\{X=1\}$, $\{X=2\}$, ..., $\{X=m\}$.
Hypothesis Testing

• Suppose the unobserved random variable $X$ is discrete and can assume a finite number of values, say, $1,2,...,m$.

• Then the problem of estimating $X$ from some observation $Y$ is called an $m$-ary hypothesis testing problem.

• The $m$ hypotheses are \{X=1\}, \{X=2\}, ..., \{X=m\}.

• If $m=2$, then this is a binary hypothesis testing problem.
Example 3.20: Signal Detection

Transmitter \[\rightarrow\text{signal } S = +1 \text{ or } -1, \text{ with prob } p \text{ and } 1-p\] \[\rightarrow\text{Noisy channel} \quad Y = S + W\] \[\rightarrow\text{Detector} \quad \hat{S}(Y) = \pm 1\]

Noise \(W \sim N(0,1)\) independent of \(S\)
Example 3.20: Signal Detection

<table>
<thead>
<tr>
<th>Transmitter</th>
<th>signal $S = +1$ or $-1$, with prob $p$ and $1-p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noisy channel</td>
<td>$Y = S + W$</td>
</tr>
<tr>
<td>Detector</td>
<td>$\hat{S}(Y) = \pm 1$</td>
</tr>
</tbody>
</table>

Noise $W \sim N(0,1)$ independent of $S$

$$
P(S = 1 \mid Y = y) = \frac{f_{Y|S}(y \mid 1)p_S(1)}{f_Y(y)} = \frac{f_{Y|S}(y \mid 1)p}{f_{Y|S}(y \mid 1)p + f_{Y|S}(y \mid -1)(1 - p)}
= \frac{1}{\sqrt{2\pi}} e^{-(y-1)^2 / 2} p
= \frac{1}{\sqrt{2\pi}} e^{-(y-1)^2 / 2} p + \frac{1}{\sqrt{2\pi}} e^{-(y+1)^2 / 2} (1 - p)
= \frac{pe^y}{pe^y + (1 - p)e^{-y}}$$
Example 3.20: Signal Detection

Transmission of a signal $S = +1$ or $-1$, with probability $p$ and $1-p$.

**Transmitter**

Signal $S = +1$ or $-1$, with prob $p$ and $1-p$

**Noisy channel**

$Y = S + W$

$W \sim N(0,1)$ independent of $S$

**Detector**

$\hat{S}(Y) = \pm 1$

$P(S = 1 | Y = y) = \frac{pe^y}{pe^y + (1-p)e^{-y}}$

$P(S = -1 | Y = y) = 1 - P(S = 1 | Y = y) = \frac{(1-p)e^{-y}}{pe^y + (1-p)e^{-y}}$
Example 3.20: Signal Detection

Transmitter \(\text{signal } S = +1 \text{ or } -1, \text{ with prob } p \text{ and } 1-p\)  \rightarrow \text{Noisy channel} \(Y = S + W\)  \rightarrow \text{Detector} \(\hat{S}(Y) = \pm 1\)

Noise \(W \sim N(0,1)\) independent of \(S\)

\[
P(S = 1 \mid Y = y) = \frac{pe^y}{pe^y + (1-p)e^{-y}}
\]

\[
P(S = -1 \mid Y = y) = 1 - P(S = 1 \mid Y = y) = \frac{(1-p)e^{-y}}{pe^y + (1-p)e^{-y}}
\]

MAP detector is:

say \(\hat{S} = 1\)

\[
P(S = 1 \mid Y = y) \quad > \quad P(S = -1 \mid Y = y)
\]

say \(\hat{S} = -1\)
Example 3.20: Signal Detection

Transmitter

\[ S = \pm 1 \text{ or } -1, \text{ with prob p and } 1-p \]

Noisy channel

\[ Y = S + W \]

Detector

\[ \hat{S}(Y) = \pm 1 \]

Noise \( W \sim N(0,1) \)

independent of \( S \)

\[
P(S = 1 \mid Y = y) = \frac{pe^y}{pe^y + (1-p)e^{-y}}
\]

\[
P(S = -1 \mid Y = y) = 1 - P(S = 1 \mid Y = y) = \frac{(1-p)e^{-y}}{pe^y + (1-p)e^{-y}}
\]

MAP detector is:

say \( \hat{S} = 1 \)

\[
P(S = 1 \mid Y = y) > P(S = -1 \mid Y = y)
\]

say \( \hat{S} = -1 \)

\[
\frac{pe^y}{pe^y + (1-p)e^{-y}} > \frac{(1-p)e^{-y}}{pe^y + (1-p)e^{-y}}
\]

say \( \hat{S} = 1 \)

\[
\frac{pe^y}{pe^y + (1-p)e^{-y}} < \frac{(1-p)e^{-y}}{pe^y + (1-p)e^{-y}}
\]

say \( \hat{S} = -1 \)
Example 3.20: Signal Detection

Transmitter

signal $S = +1$ or $-1$, with prob $p$ and $1−p$

Noisy channel

$Y = S + W$

Detector

$\hat{S}(Y) = \pm 1$

Noise $W \sim N(0,1)$ independent of $S$

$$P(S = 1 \mid Y = y) = \frac{pe^y}{pe^y + (1−p)e^{-y}}$$

$$P(S = -1 \mid Y = y) = 1 - P(S = 1 \mid Y = y) = \frac{(1−p)e^{-y}}{pe^y + (1−p)e^{-y}}$$

MAP detector is:

say $\hat{S} = 1$

$$P(S = 1 \mid Y = y) > P(S = -1 \mid Y = y)$$

say $\hat{S} = -1$

$$\frac{pe^y}{pe^y + (1−p)e^{-y}} > \frac{(1−p)e^{-y}}{pe^y + (1−p)e^{-y}}$$

say $\hat{S} = 1$

$$e^{2y} > \frac{1−p}{p}$$

say $\hat{S} = -1$
Example 3.20: Signal Detection

Transmitter \[\rightarrow\text{signal } S = +1 \text{ or } -1, \text{ with prob } p \text{ and } 1-p\] \[\rightarrow\text{noisy channel}\] \[Y = S + W\] \[\rightarrow\text{detector}\] \[\hat{S}(Y) = \pm 1\]

Noise \(W \sim N(0,1)\) independent of \(S\)

**Probability Calculation**

\[
P(S = 1 \mid Y = y) = \frac{pe^y}{pe^y + (1-p)e^{-y}}
\]

\[
P(S = -1 \mid Y = y) = 1 - P(S = 1 \mid Y = y) = \frac{(1-p)e^{-y}}{pe^y + (1-p)e^{-y}}
\]

**MAP Detector**

\[
P(S = 1 \mid Y = y) > P(S = -1 \mid Y = y)
\]

- Say \(\hat{S} = 1\)

\[
P(S = -1 \mid Y = y) > P(S = 1 \mid Y = y)
\]

- Say \(\hat{S} = -1\)

**Decision Rule**

\[
e^{2y} > \frac{1-p}{p}
\]

- Say \(\hat{S} = 1\)

\[
y > \frac{1}{2} \ln \left( \frac{1-p}{p} \right)
\]

- Say \(\hat{S} = -1\)
Example 3.20: MAP Detector

Transmitter \(\xrightarrow{\text{signal } S = +1 \text{ or } -1, \text{ with prob } p \text{ and } 1-p} \) Noisy channel \(\xrightarrow{Y = S + W} \) Detector \(\xrightarrow{\hat{S}(Y) = \pm 1} \)

Denoting \(\eta = \frac{1}{2} \ln \left(\frac{1-p}{p}\right)\), we have:

\[
\hat{S}(Y) = \begin{cases} 
1, & \text{if } Y > \eta \\
-1, & \text{if } Y < \eta 
\end{cases}
\]
Example 3.20: MAP Detector

Detector

Transmitter

signal $S = \pm 1$, with prob $p$ and $1-p$

Noisy channel

$Y = S + W$

Detector

$\hat{S}(Y) = \pm 1$

Denoting $\eta = \frac{1}{2} \ln \left( \frac{1 - p}{p} \right)$, we have:

$$\hat{S}(Y) = \begin{cases} 
1, & \text{if } Y > \eta \\
-1, & \text{if } Y < \eta 
\end{cases}$$

Several comments:

1. What if $Y = \eta$?
Example 3.20: MAP Detector

Denoting $\eta = \frac{1}{2}\ln\left(\frac{1-p}{p}\right)$, we have:

$$\hat{S}(Y) = \begin{cases} 
1, & \text{if } Y > \eta \\
-1, & \text{if } Y < \eta
\end{cases}$$

Several comments:

1. What if $Y = \eta$? Then it does not matter what value to choose, as both have 1/2 probability of a correct decision.
Example 3.20: MAP Detector

Transmitter \(\rightarrow\) Noisy channel \(\rightarrow\) Detector

signal \(S = +1\) or \(-1\), with prob \(p\) and \(1-p\)

\[ Y = S + W \]

\[ \hat{S}(Y) = \pm 1 \]

Denoting \(\eta = \frac{1}{2} \ln \left( \frac{1-p}{p} \right)\), we have:

\[ \hat{S}(Y) = \begin{cases} 
1, & \text{if } Y > \eta \\
-1, & \text{if } Y < \eta 
\end{cases} \]

Several comments:

1. What if \(Y = \eta\)? Then it does not matter what value to choose, as both have 1/2 probability of a correct decision.
2. If \(p = 1/2\), then \(\eta = 0\).
Example 3.20: MAP Detector

![Diagram of signal transmission and detection process]

Denoting $\eta = \frac{1}{2} \ln \left( \frac{1-p}{p} \right)$, we have:

$$\hat{S}(Y) = \begin{cases} 
1, & \text{if } Y > \eta \\
-1, & \text{if } Y < \eta 
\end{cases}$$

Several comments:

1. What if $Y = \eta$? Then it does not matter what value to choose, as both have $1/2$ probability of a correct decision.
2. If $p = 1/2$, then $\eta = 0$.
3. If $p = 1$, then $\eta = -\infty$. Always say $\hat{S} = 1$, since $P(S = -1) = 0$. 

Signal $S = +1$ or -1, with prob $p$ and $1-p$.

Noise $W \sim N(0,1)$ independent of $S$. 

Y = S + W
Example 3.20: MAP Detector

![Diagram showing transmitter, noisy channel, and detector with signal $S = +1$ or $-1$, noise $W \sim N(0,1)$, and $\hat{S}(Y) = \pm 1$.]

Denoting $\eta = \frac{1}{2} \ln \left( \frac{1-p}{p} \right)$, we have:

$$\hat{S}(Y) = \begin{cases} 
1, & \text{if } Y > \eta \\
-1, & \text{if } Y < \eta 
\end{cases}$$

Several comments:

1. What if $Y = \eta$? Then it does not matter what value to choose, as both have $1/2$ probability of a correct decision.
2. If $p = 1/2$, then $\eta = 0$.
3. If $p = 1$, then $\eta = -\infty$. Always say $\hat{S} = 1$, since $P(S = -1) = 0$.
4. If $p = 0$, then $\eta = \infty$. Always say $\hat{S} = -1$, since $P(S = 1) = 0$. 
HW 8 Prob 5: MAP Detector

Transmitter

signal $X = 0$ or $1$, with prob $p_0$ and $p_1$

Noisy channel

$Y = X + W$

Noise $W \sim N(0, \sigma^2)$ independent of $X$

Detector

$\hat{X}(Y)$
HW 8 Prob 5: MAP Detector

Transmitter \( \rightarrow \) Noisy channel \( \rightarrow \) Detector

signal \( X = 0 \) or \( 1 \), with prob \( p_0 \) and \( p_1 \)

\( Y = X + W \)

Noise \( W \sim N(0, \sigma^2) \)

independent of \( X \)

\[
P(X = 1 \mid Y = y) = \frac{f_{Y \mid X}(y \mid 1)p_X(1)}{f_Y(y)} = \frac{f_{Y \mid X}(y \mid 1)p_1}{f_{Y \mid X}(y \mid 0)p_0 + f_{Y \mid X}(y \mid 1)p_1}
\]
HW 8 Prob 5: MAP Detector

Transmitter \[\xrightarrow{\text{signal } X = 0 \text{ or } 1, \text{ with prob } p_0 \text{ and } p_1}\] Noisy channel \[\xrightarrow{Y = X + W}\] Noise \(W \sim N(0, \sigma^2)\) independent of \(X\) Detector \[\hat{X}(Y)\]

\[P(X = 1 \mid Y = y) = \frac{f_{Y|X}(y \mid 1)p_X(1)}{f_Y(y)} = \frac{f_{Y|X}(y \mid 1)p_1}{f_{Y|X}(y \mid 0)p_0 + f_{Y|X}(y \mid 1)p_1}\]

\[= \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y-1)^2}{2\sigma^2}} p_1\]

\[= \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}} p_0 + \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y-1)^2}{2\sigma^2}} p_1\]

\[= \frac{p_1}{p_0 e^{\frac{y^2}{2\sigma^2}} + p_1}\]
HW 8 Prob 5: MAP Detector

Transmitter \[ \begin{align*}
\text{signal } X = 0 \text{ or } 1, \\
\text{with prob } p_0 \text{ and } p_1
\end{align*} \]

Noisy channel \[ Y = X + W \]

Detector \[ \hat{X}(Y) \]

Noise \( W \sim \mathcal{N}(0, \sigma^2) \) independent of \( X \)

\( P(X = 1 \mid Y = y) \quad > \quad P(X = 0 \mid Y = y) \)

say \( \hat{X} = 1 \)

\( P(X = 1 \mid Y = y) \quad < \quad P(X = 0 \mid Y = y) \)

say \( \hat{X} = 0 \)
HW 8 Prob 5: MAP Detector

Transmitter

\( \text{signal } X = 0 \text{ or } 1, \text{ with prob } p_0 \text{ and } p_1 \)

Noisy channel

\( Y = X + W \)

Detector

\( \hat{X}(Y) \)

Noise \( W \sim N(0, \sigma^2) \) independent of \( X \)

say \( \hat{X} = 1 \)

\[
P(X = 1 \mid Y = y) > \frac{p_1}{p_0 e^{2\sigma^2} + p_1}
\]

say \( \hat{X} = 0 \)

\[
P(X = 0 \mid Y = y) < \frac{-2y + 1}{p_0 e^{2\sigma^2} + p_1}
\]

say \( \hat{X} = 1 \)

\[
\frac{p_1}{p_0 e^{2\sigma^2} + p_1} > \frac{-2y + 1}{p_0 e^{2\sigma^2} + p_1}
\]

say \( \hat{X} = 0 \)
HW 8 Prob 5: MAP Detector

Transmitter \[\xrightarrow{\text{signal } X = 0 \text{ or } 1, \text{ with prob } p_0 \text{ and } p_1}\] Noisy channel \[\xrightarrow{\text{Y} = X + W}\] Detector \[\hat{X}(Y)\]

Noise \(W \sim N(0, \sigma^2)\) independent of \(X\)

\[
P(X = 1 | Y = y) > \begin{cases} p_1 e^{\frac{-2y+1}{2\sigma^2}} & \text{say } \hat{X} = 1 \\ p_0 e^{\frac{-2y+1}{2\sigma^2}} + p_1 & \text{say } \hat{X} = 0 \end{cases}
\]

\[
P(X = 0 | Y = y) < \begin{cases} p_0 e^{\frac{-2y+1}{2\sigma^2}} & \text{say } \hat{X} = 1 \\ p_0 e^{\frac{-2y+1}{2\sigma^2}} + p_1 & \text{say } \hat{X} = 0 \end{cases}
\]
HW 8 Prob 5: MAP Detector

Transmitter \( \xrightarrow{\text{signal } X = 0 \text{ or } 1, \text{ with prob } p_0 \text{ and } p_1} \) Noisy channel \( \xrightarrow{\text{Y} = X + W} \) Detector

\( \hat{X}(Y) \)

Noise \( W \sim N(0, \sigma^2) \) independent of \( X \)

\[
P(X = 1 \mid Y = y) > P(X = 0 \mid Y = y)
\]

say \( \hat{X} = 1 \)

\[
P(X = 1 \mid Y = y) > \frac{p_1 e^{-2y+1}}{p_0 e^{-2\sigma^2} + p_1}
\]

say \( \hat{X} = 1 \)

\[
\frac{p_1 e^{-2y+1}}{p_0 e^{-2\sigma^2} + p_1} > \frac{p_0 e^{-2y+1}}{p_0 e^{-2\sigma^2} + p_1}
\]

say \( \hat{X} = 0 \)

\[
y > \sigma^2 \ln \left( \frac{p_0}{p_1} \right) + \frac{1}{2}
\]

say \( \hat{X} = 0 \)

\[
y < \sigma^2 \ln \left( \frac{p_0}{p_1} \right) + \frac{1}{2}
\]

say \( \hat{X} = 0 \)
HW 8 Prob 5: MAP Detector

Transmitter \( \rightarrow \) Noisy channel \( \rightarrow \) Detector

Signal \( X = 0 \) or \( 1 \), with prob \( p_0 \) and \( p_1 \)

\( Y = X + W \)

Noise \( W \sim N(0, \sigma^2) \) independent of \( X \)

Denoting \( \eta = \sigma^2 \ln \left( \frac{p_0}{p_1} \right) + \frac{1}{2} \), we have:

\[
\hat{X}(Y) = \begin{cases} 
1, & \text{if } Y > \eta \\
0, & \text{if } Y < \eta 
\end{cases}
\]
HW 8 Prob 5: MAP Detector

Transmitter \[\xrightarrow{\text{signal } X = 0 \text{ or } 1, \ \text{with prob } p_0 \text{ and } p_1}\] Noisy channel \[\xrightarrow{Y = X + W}\] Detector \[\xrightarrow{\hat{X}(Y)}\]

Noise \( W \sim N(0, \sigma^2) \) independent of \( X \)

\[
P(X = 1 \mid Y = y) = \frac{f_{Y\mid X}(y \mid 1)p_1}{f_Y(y)} \quad \text{and} \quad P(X = 0 \mid Y = y) = \frac{f_{Y\mid X}(y \mid 0)p_0}{f_Y(y)}
\]

Therefore, MAP decision is equivalent to:

\[
say \hat{X} = 1 \quad \text{if} \quad f_{Y\mid X}(y \mid 1)p_1 > f_{Y\mid X}(y \mid 0)p_0
\]

\[
say \hat{X} = 0 \quad \text{if} \quad f_{Y\mid X}(y \mid 1)p_1 < f_{Y\mid X}(y \mid 0)p_0
\]
HW 8 Prob 5: Weighted PDFs of Y conditioned on X=0 and on X=1
HW 8 Prob 5: $P(\text{correct detection})$

$P(\text{correct detection}) = P(\hat{X}(Y) = X)$
HW 8 Prob 5: $P$(correct detection)

\[
P(\text{correct detection}) = P(\hat{X}(Y) = X) \\
= P(\hat{X}(Y) = X | X = 0)P(X = 0) + P(\hat{X}(Y) = X | X = 1)P(X = 1)
\]
(by total probability theorem)
HW 8 Prob 5: $P$(correct detection)

$$P(\text{correct detection}) = P(\hat{X}(Y) = X)$$

$$= P(\hat{X}(Y) = X \mid X = 0)P(X = 0) + P(\hat{X}(Y) = X \mid X = 1)P(X = 1)$$

$$= P(Y < \eta \mid X = 0)p_0 + P(Y > \eta \mid X = 1)p_1$$
HW 8 Prob 5: $P(\text{correct detection})$

$P(\text{correct detection}) = P(\hat{X}(Y) = X)$

$$= P(\hat{X}(Y) = X | X = 0)P(X = 0) + P(\hat{X}(Y) = X | X = 1)P(X = 1)$$

$$= P(Y < \eta | X = 0)p_0 + P(Y > \eta | X = 1)p_1$$

$$= p_0 \int_{-\infty}^{\eta} f_{Y|X}(y | 0) dy + p_1 \int_{\eta}^{\infty} f_{Y|X}(y | 1) dy$$
HW 8 Prob 5: $P$(correct detection)

$P$(correct detection) = $P(\hat{X}(Y) = X)$

$$= P(\hat{X}(Y) = X \mid X = 0)P(X = 0) + P(\hat{X}(Y) = X \mid X = 1)P(X = 1)$$

$$= P(Y < \eta \mid X = 0)p_0 + P(Y > \eta \mid X = 1)p_1$$

$$= p_0 \int_{-\infty}^{\eta} f_{Y \mid X}(y \mid 0) \, dy + p_1 \int_{\eta}^{\infty} f_{Y \mid X}(y \mid 1) \, dy$$

$$= p_0 \Phi\left(\frac{\eta}{\sigma}\right) + p_1 \left(1 - \Phi\left(\frac{\eta - 1}{\sigma}\right)\right)$$

(Because $f_{Y \mid X}(y \mid 0)$ is Gaussian with mean 0 and standard deviation $\sigma$, and $f_{Y \mid X}(y \mid 1)$ is Gaussian with mean 1 and standard deviation $\sigma$.)
HW 8 Prob 5: $p_0 = p_1 = \frac{1}{2}$, $\sigma = 0.1$

Noise small compared to the separation between 0 and 1, therefore high probability of correct detection, 0.9999997.
HW 8 Prob 5: $p_0 = p_1 = \frac{1}{2}, \sigma = 0.5$

Probability of correct detection = 0.84.
HW 8 Prob 5: $p_0 = p_1 = \frac{1}{2}$, $\sigma = 1$

Significant noise, therefore lower probability of correct detection, $0.69$. As $\sigma$ goes to $\infty$, $P($correct decision$)$ approaches $1/2$. 
HW 8 Prob 5: $p_0 \approx 1$, $p_1 \approx 0$

Suppose $\sigma = 1$, $p_0 = 0.99$, $p_1 = 0.01$
HW 8 Prob 5: \( p_0 \approx 1, \ p_1 \approx 0 \)

Suppose \( \sigma = 1, \ p_0 = 0.99, \ p_1 = 0.01 \)

Then \( \eta = \sigma^2 \ln \left( \frac{p_0}{p_1} \right) + \frac{1}{2} = \ln 99 + 0.5 \approx 4.6 \)
HW 8 Prob 5: $p_0 \approx 1$, $p_1 \approx 0$

Suppose $\sigma = 1$, $p_0 = 0.99$, $p_1 = 0.01$

Then $\eta = \sigma^2 \ln \left( \frac{p_0}{p_1} \right) + \frac{1}{2} = \ln 99 + 0.5 \approx 4.6$

$P(\text{correct detection}) = p_0 \Phi \left( \frac{\eta}{\sigma} \right) + p_1 \left( 1 - \Phi \left( \frac{\eta - 1}{\sigma} \right) \right)$
HW 8 Prob 5: \( p_0 \approx 1, \ p_1 \approx 0 \)

Suppose \( \sigma = 1, \ p_0 = 0.99, \ p_1 = 0.01 \)

Then \( \eta = \sigma^2 \ln \left( \frac{p_0}{p_1} \right) + \frac{1}{2} = \ln 99 + 0.5 \approx 4.6 \)

\[
P(\text{correct detection}) = p_0 \Phi \left( \frac{\eta}{\sigma} \right) + p_1 \left( 1 - \Phi \left( \frac{\eta - 1}{\sigma} \right) \right)
\]

\[
= 0.99 \Phi(4.6) + 0.01(1 - \Phi(3.6))
\]

\[
\approx 0.99 \cdot 0.999998 + 0.01 \cdot 0.000159
\]

\[
\approx 0.989999
\]
HW 8 Prob 5: $p_0 \approx 1$, $p_1 \approx 0$

Suppose $\sigma = 1$, $p_0 = 0.99$, $p_1 = 0.01$

Then $\eta = \sigma^2 \ln \left( \frac{p_0}{p_1} \right) + \frac{1}{2} = \ln 99 + 0.5 \approx 4.6$

$$P(\text{correct detection}) = p_0 \Phi \left( \frac{\eta}{\sigma} \right) + p_1 \left( 1 - \Phi \left( \frac{\eta - 1}{\sigma} \right) \right)$$

$$= 0.99\Phi(4.6) + 0.01(1 - \Phi(3.6))$$

$$\approx 0.99 \cdot 0.999998 + 0.01 \cdot 0.000159$$

$$\approx 0.989999$$

Is this a great detector?
HW 8 Prob 5: $p_0 \approx 1, \ p_1 \approx 0$

Suppose $\sigma = 1, \ p_0 = 0.99, \ p_1 = 0.01$

Then $\eta = \sigma^2 \ln\left(\frac{p_0}{p_1}\right) + \frac{1}{2} = \ln 99 + 0.5 \approx 4.6$

$$P(\text{correct detection}) = p_0 \Phi\left(\frac{\eta}{\sigma}\right) + p_1 \left(1 - \Phi\left(\frac{\eta - 1}{\sigma}\right)\right)$$

$$= 0.99 \Phi(4.6) + 0.01(1 - \Phi(3.6))$$

$$\approx 0.99 \cdot 0.999998 + 0.01 \cdot 0.000159$$

$$\approx 0.989999$$

Is this a great detector?

Say, $X = 0$ means "no burglar" and $X = 1$ means "burglar."

$P(\text{correct detection} \mid \text{no burglar}) = \Phi(4.6) \approx 0.999998$

$P(\text{correct detection} \mid \text{burglar}) = 1 - \Phi(3.6) = 0.000159$
Suppose $\sigma = 1$, $p_0 = 0.99$, $p_1 = 0.01$

Then $\eta = \sigma^2 \ln \left( \frac{p_0}{p_1} \right) + \frac{1}{2} = \ln 99 + 0.5 \approx 4.6$

$$P(\text{correct detection}) = p_0 \Phi \left( \frac{\eta}{\sigma} \right) + p_1 \left( 1 - \Phi \left( \frac{\eta-1}{\sigma} \right) \right)$$

$$= 0.99 \Phi(4.6) + 0.01 (1 - \Phi(3.6))$$

$$\approx 0.99 \cdot 0.999998 + 0.01 \cdot 0.000159$$

$$\approx 0.989999$$

Is this a great detector?

Say, $X = 0$ means "no burglar" and $X = 1$ means "burglar."

$$P(\text{correct detection} \mid \text{no burglar}) = \Phi(4.6) \approx 0.999998$$

$$P(\text{correct detection} \mid \text{burglar}) = 1 - \Phi(3.6) = 0.000159$$

Actually, this detector is horrendous: precisely when it matters most, it is very likely to make a mistake!
HW 8 Prob 5: all errors are not necessarily created equal!

Suppose a false alarm is 100 times less costly than making an error when there is a burglar:

\[
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 1) = 100 \\
\text{cost}(\hat{X}(Y) = 1 \text{ and } X = 0) = 1
\]
HW 8 Prob 5: all errors are not necessarily created equal!

Suppose a false alarm is 100 times less costly than making an error when there is a burglar:
\[
\begin{align*}
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 1) &= 100 \\
\text{cost}(\hat{X}(Y) = 1 \text{ and } X = 0) &= 1 \\
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 0) &= \text{cost}(\hat{X}(Y) = 1 \text{ and } X = 1) = 0
\end{align*}
\]
HW 8 Prob 5: all errors are not necessarily created equal!

Suppose a false alarm is 100 times less costly than making an error when there is a burglar:

\[
\begin{align*}
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 1) &= 100 \\
\text{cost}(\hat{X}(Y) = 1 \text{ and } X = 0) &= 1 \\
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 0) &= \text{cost}(\hat{X}(Y) = 1 \text{ and } X = 1) = 0
\end{align*}
\]

Instead of minimizing \( P(\text{error}) \), let us minimize \( E[\text{cost}] \).
HW 8 Prob 5: all errors are not necessarily created equal!

Suppose a false alarm is 100 times less costly than making an error when there is a burglar:

\[
\begin{align*}
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 1) &= 100 \\
\text{cost}(\hat{X}(Y) = 1 \text{ and } X = 0) &= 1 \\
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 0) &= \text{cost}(\hat{X}(Y) = 1 \text{ and } X = 1) = 0
\end{align*}
\]

Instead of minimizing \( P(\text{error}) \), let us minimize \( E[\text{cost}] \).

\[
P(\text{error}) = P(\hat{X}(Y) \neq X) = P(\hat{X}(Y) = 0 \text{ and } X = 1) + P(\hat{X}(Y) = 1 \text{ and } X = 0) \\
= p_1 P(\hat{X}(Y) = 0 \mid X = 1) + p_0 P(\hat{X}(Y) = 1 \mid X = 0),
\]
HW 8 Prob 5: all errors are not necessarily created equal!

Suppose a false alarm is 100 times less costly than making an error when there is a burglar:

\[
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 1) = 100 \\
\text{cost}(\hat{X}(Y) = 1 \text{ and } X = 0) = 1 \\
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 0) = \text{cost}(\hat{X}(Y) = 1 \text{ and } X = 1) = 0
\]

Instead of minimizing \( P(\text{error}) \), let us minimize \( E[\text{cost}] \).

\[
P(\text{error}) = P(\hat{X}(Y) \neq X) = P(\hat{X}(Y) = 0 \text{ and } X = 1) + P(\hat{X}(Y) = 1 \text{ and } X = 0) \\
= p_1 P(\hat{X}(Y) = 0 \mid X = 1) + p_0 P(\hat{X}(Y) = 1 \mid X = 0),
\]

\[
E[\text{cost}] = 100 P(\hat{X}(Y) = 0 \text{ and } X = 1) + P(\hat{X}(Y) = 1 \text{ and } X = 0)
\]
HW 8 Prob 5: all errors are not necessarily created equal!

Suppose a false alarm is 100 times less costly than making an error when there is a burglar:
\[
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 1) = 100
\]
\[
\text{cost}(\hat{X}(Y) = 1 \text{ and } X = 0) = 1
\]
\[
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 0) = \text{cost}(\hat{X}(Y) = 1 \text{ and } X = 1) = 0
\]

Instead of minimizing \( \text{P(error)} \), let us minimize \( \text{E}[\text{cost}] \).

\[
\text{P(error)} = \text{P}(\hat{X}(Y) \neq X) = \text{P}(\hat{X}(Y) = 0 \text{ and } X = 1) + \text{P}(\hat{X}(Y) = 1 \text{ and } X = 0)
\]
\[
= p_1 \text{P}(\hat{X}(Y) = 0 \mid X = 1) + p_0 \text{P}(\hat{X}(Y) = 1 \mid X = 0),
\]
\[
\text{E}[\text{cost}] = 100 \text{P}(\hat{X}(Y) = 0 \text{ and } X = 1) + \text{P}(\hat{X}(Y) = 1 \text{ and } X = 0)
\]
\[
= 100 p_1 \text{P}(\hat{X}(Y) = 0 \mid X = 1) + p_0 \text{P}(\hat{X}(Y) = 1 \mid X = 0)
\]
HW 8 Prob 5: all errors are not necessarily created equal!

Suppose a false alarm is 100 times less costly than making an error when there is a burglar:
\[ \text{cost}(\hat{X}(Y) = 0 \text{ and } X = 1) = 100 \]
\[ \text{cost}(\hat{X}(Y) = 1 \text{ and } X = 0) = 1 \]
\[ \text{cost}(\hat{X}(Y) = 0 \text{ and } X = 0) = \text{cost}(\hat{X}(Y) = 1 \text{ and } X = 1) = 0 \]

Instead of minimizing \( P(\text{error}) \), let us minimize \( E[\text{cost}] \).
\[
P(\text{error}) = P(\hat{X}(Y) \neq X) = P(\hat{X}(Y) = 0 \text{ and } X = 1) + P(\hat{X}(Y) = 1 \text{ and } X = 0) \\
= p_1 P(\hat{X}(Y) = 0 \mid X = 1) + p_0 P(\hat{X}(Y) = 1 \mid X = 0),
\]
\[
E[\text{cost}] = 100 P(\hat{X}(Y) = 0 \text{ and } X = 1) + P(\hat{X}(Y) = 1 \text{ and } X = 0) \\
= 100 p_1 P(\hat{X}(Y) = 0 \mid X = 1) + p_0 P(\hat{X}(Y) = 1 \mid X = 0) \\
= P(\hat{X}(Y) = 0 \mid X = 1) + 0.99 P(\hat{X}(Y) = 1 \mid X = 0) 
\]
HW 8 Prob 5: all errors are not necessarily created equal!

Suppose a false alarm is 100 times less costly than making an error when there is a burglar:
\[
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 1) = 100
\]
\[
\text{cost}(\hat{X}(Y) = 1 \text{ and } X = 0) = 1
\]
\[
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 0) = \text{cost}(\hat{X}(Y) = 1 \text{ and } X = 1) = 0
\]

Instead of minimizing \( P(\text{error}) \), let us minimize \( E[\text{cost}] \).
\[
P(\text{error}) = P(\hat{X}(Y) \neq X) = P(\hat{X}(Y) = 0 \text{ and } X = 1) + P(\hat{X}(Y) = 1 \text{ and } X = 0)
\]
\[
= p_1 P(\hat{X}(Y) = 0 \mid X = 1) + p_0 P(\hat{X}(Y) = 1 \mid X = 0),
\]
\[
E[\text{cost}] = 100 P(\hat{X}(Y) = 0 \text{ and } X = 1) + P(\hat{X}(Y) = 1 \text{ and } X = 0)
\]
\[
= 100 p_1 P(\hat{X}(Y) = 0 \mid X = 1) + p_0 P(\hat{X}(Y) = 1 \mid X = 0)
\]
\[
= P(\hat{X}(Y) = 0 \mid X = 1) + 0.99 P(\hat{X}(Y) = 1 \mid X = 0)
\]

Actually, let's minimize \( E[\text{cost}/1.99] \) which is the same as minimizing \( E[\text{cost}] \):
HW 8 Prob 5: all errors are not necessarily created equal!

Suppose a false alarm is 100 times less costly than making an error when there is a burglar:

\[
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 1) = 100 \\
\text{cost}(\hat{X}(Y) = 1 \text{ and } X = 0) = 1 \\
\text{cost}(\hat{X}(Y) = 0 \text{ and } X = 0) = \text{cost}(\hat{X}(Y) = 1 \text{ and } X = 1) = 0
\]

Instead of minimizing \( P(\text{error}) \), let us minimize \( E[\text{cost}] \).

\[
P(\text{error}) = P(\hat{X}(Y) \neq X) = P(\hat{X}(Y) = 0 \text{ and } X = 1) + P(\hat{X}(Y) = 1 \text{ and } X = 0) \\
= p_1 P(\hat{X}(Y) = 0 \mid X = 1) + p_0 P(\hat{X}(Y) = 1 \mid X = 0),
\]

\[
E[\text{cost}] = 100 P(\hat{X}(Y) = 0 \text{ and } X = 1) + P(\hat{X}(Y) = 1 \text{ and } X = 0) \\
= 100 p_1 P(\hat{X}(Y) = 0 \mid X = 1) + p_0 P(\hat{X}(Y) = 1 \mid X = 0) \\
= P(\hat{X}(Y) = 0 \mid X = 1) + 0.99 P(\hat{X}(Y) = 1 \mid X = 0)
\]

Actually, let's minimize \( E[\text{cost}/1.99] \) which is the same as minimizing \( E[\text{cost}] \):

\[
E \left[ \frac{\text{cost}}{1.99} \right] = \frac{1}{1.99} P(\hat{X}(Y) = 0 \mid X = 1) + \frac{0.99}{1.99} P(\hat{X}(Y) = 1 \mid X = 0)
\]
HW 8 Prob 5: all errors are not necessarily created equal!

\[
P(\text{error}) = P(\hat{X}(Y) \neq X) = P(\hat{X}(Y) = 0 \text{ and } X = 1) + P(\hat{X}(Y) = 1 \text{ and } X = 0) \\
= p_1P(\hat{X}(Y) = 0 | X = 1) + p_0P(\hat{X}(Y) = 1 | X = 0),
\]

\[
E\left[\frac{\text{cost}}{1.99}\right] = \frac{1}{1.99} P(\hat{X}(Y) = 0 | X = 1) + \frac{0.99}{1.99} P(\hat{X}(Y) = 1 | X = 0)
\]
HW 8 Prob 5: all errors are not necessarily created equal!

\[ P(\text{error}) = P(\hat{X}(Y) \neq X) = P(\hat{X}(Y) = 0 \text{ and } X = 1) + P(\hat{X}(Y) = 1 \text{ and } X = 0) \]

\[ = p_1 P(\hat{X}(Y) = 0 \mid X = 1) + p_0 P(\hat{X}(Y) = 1 \mid X = 0), \]

\[ E\left[ \frac{\text{cost}}{1.99} \right] = \frac{1}{1.99} P(\hat{X}(Y) = 0 \mid X = 1) + \frac{0.99}{1.99} P(\hat{X}(Y) = 1 \mid X = 0) \]

But this is a MAP detection problem with new prior probabilities

\[ p_{0,\text{new}} = \frac{0.99}{1.99}, \quad p_{1,\text{new}} = \frac{1}{1.99} \]
HW 8 Prob 5: all errors are not necessarily created equal!

\[ P(\text{error}) = P(\hat{X}(Y) \neq X) = P(\hat{X}(Y) = 0 \text{ and } X = 1) + P(\hat{X}(Y) = 1 \text{ and } X = 0) \]

\[ = p_1 P(\hat{X}(Y) = 0 \mid X = 1) + p_0 P(\hat{X}(Y) = 1 \mid X = 0), \]

\[ E \left[ \frac{\text{cost}}{1.99} \right] = \frac{1}{1.99} P(\hat{X}(Y) = 0 \mid X = 1) + \frac{0.99}{1.99} P(\hat{X}(Y) = 1 \mid X = 0) \]

But this is a MAP detection problem with new prior probabilities

\[ p_{0,\text{new}} = \frac{0.99}{1.99}, \quad p_{1,\text{new}} = \frac{1}{1.99} \]

The solution is our MAP detector with threshold \( \eta_{\text{new}} = \sigma^2 \ln \left( \frac{p_{0,\text{new}}}{p_{1,\text{new}}} \right) + \frac{1}{2} = \ln 0.99 + 0.5 \approx 0.49 \)
HW 8 Prob 5: all errors are not necessarily created equal!

\[ P(\text{error}) = P(\hat{X}(Y) \neq X) = P(\hat{X}(Y) = 0 \text{ and } X = 1) + P(\hat{X}(Y) = 1 \text{ and } X = 0) \]

\[ = p_1 P(\hat{X}(Y) = 0 \mid X = 1) + p_0 P(\hat{X}(Y) = 1 \mid X = 0), \]

\[ E\left[ \frac{\text{cost}}{1.99} \right] = \frac{1}{1.99} P(\hat{X}(Y) = 0 \mid X = 1) + \frac{0.99}{1.99} P(\hat{X}(Y) = 1 \mid X = 0) \]

But this is a MAP detection problem with new prior probabilities

\[ p_{0,\text{new}} = \frac{0.99}{1.99}, \quad p_{1,\text{new}} = \frac{1}{1.99} \]

The solution is our MAP detector with threshold \( \eta_{\text{new}} = \sigma^2 \ln \left( \frac{p_{0,\text{new}}}{p_{1,\text{new}}} \right) + \frac{1}{2} = \ln 0.99 + 0.5 \approx 0.49 \)

\[ P(\text{correct detection}) = p_0 \Phi\left( \frac{\eta_{\text{new}}}{\sigma} \right) + p_1 \left( 1 - \Phi\left( \frac{\eta_{\text{new}} - 1}{\sigma} \right) \right) \]

\[ = 0.99 \Phi(0.49) + 0.01(1 - \Phi(-0.51)) \]

\[ \approx 0.99 \cdot 0.6879 + 0.01 \cdot 0.6950 \approx 0.6880 \]
HW 8 Prob 5: all errors are not necessarily created equal!

\[ \mathbf{P}({\text{error}}) = \mathbf{P}(\hat{X}(Y) \neq X) = \mathbf{P}(\hat{X}(Y) = 0 \text{ and } X = 1) + \mathbf{P}(\hat{X}(Y) = 1 \text{ and } X = 0) \]
\[ = p_1 \mathbf{P}(\hat{X}(Y) = 0 \mid X = 1) + p_0 \mathbf{P}(\hat{X}(Y) = 1 \mid X = 0), \]

\[ \mathbf{E}\left[ \frac{\text{cost}}{1.99} \right] = \frac{1}{1.99} \mathbf{P}(\hat{X}(Y) = 0 \mid X = 1) + \frac{0.99}{1.99} \mathbf{P}(\hat{X}(Y) = 1 \mid X = 0) \]

But this is a MAP detection problem with new prior probabilities

\[ p_{0,\text{new}} = \frac{0.99}{1.99}, \quad p_{1,\text{new}} = \frac{1}{1.99} \]

The solution is our MAP detector with threshold \( \eta_{\text{new}} = \sigma^2 \ln\left( \frac{p_{0,\text{new}}}{p_{1,\text{new}}} \right) + \frac{1}{2} = \ln 0.99 + 0.5 \approx 0.49 \)

\[ \mathbf{P}(\text{correct detection}) = p_0 \Phi\left( \frac{\eta_{\text{new}}}{\sigma} \right) + p_1 \left( 1 - \Phi\left( \frac{\eta_{\text{new}} - 1}{\sigma} \right) \right) \]
\[ = 0.99\Phi(0.49) + 0.01(1 - \Phi(-0.51)) \]
\[ \approx 0.99 \cdot 0.6879 + 0.01 \cdot 0.6950 \approx 0.6880 \]

\[ \mathbf{P}(\text{correct detection} \mid \text{no burglar}) = \Phi(0.49) \approx 0.6879 \]
\[ \mathbf{P}(\text{correct detection} \mid \text{burglar}) = 1 - \Phi(-0.51) \approx 0.6950 \]
Receiver operating characteristic: tradeoff between false alarms and correct detections

- $P(\text{alarm} \mid \text{no burglar}) = \text{(conditional) false alarm probability, or probability of Type I error}$
- $P(\text{alarm} \mid \text{burglar}) = \text{(conditional) correct detection probability, or } (1 - \text{probability of Type II error})$
Receiver operating characteristic: tradeoff between false alarms and correct detections

\[ \eta = \infty \]
Always detect no burglar
Zero correct detection probability
No false alarms
Receiver operating characteristic: tradeoff between false alarms and correct detections

- $\eta = \infty$: Always detect no burglar
  - Zero correct detection probability
  - No false alarms

- $\eta = -\infty$: Always detect burglar
  - Correct detection probability 1
  - False alarm probability 1
Receiver operating characteristic: tradeoff between false alarms and correct detections

- \( \eta = \infty \) Always detect no burglar
  - Zero correct detection probability
  - No false alarms
- \( \eta = 0.49 \)
  - In the example,
    - \( P(\text{correct detection}) \approx 0.6950 \)
    - \( P(\text{false alarm}) \approx 0.3121 \)
- \( \eta = -\infty \)
  - Always detect burglar
  - Correct detection probability 1
  - False alarm probability 1
Receiver operating characteristic: tradeoff between false alarms and correct detections

\[ \eta = \infty \]
Always detect no burglar
Zero correct detection probability
No false alarms

\[ \eta = 0.49 \]
In the example,
P(correct detection) \approx 0.6950
P(false alarm) \approx 0.3121
Ex. 8.10: Signal Detection and the Matched Filter

To improve noise resilience,
when we want to send a "0" we send a sequence $a \equiv (a_1, \ldots, a_n)$
when we want to send a "1" we send a sequence $b \equiv (b_1, \ldots, b_n)$
Ex. 8.10: Signal Detection and the Matched Filter

To improve noise resilience,
   when we want to send a "0" we send a sequence \( a \equiv \{a_1, \ldots, a_n\} \)
   when we want to send a "1" we send a sequence \( b \equiv \{b_1, \ldots, b_n\} \)

(Both \( a \) and \( b \) are known to the receiver.)
Ex. 8.10: Signal Detection and the Matched Filter

To improve noise resilience,
when we want to send a "0" we send a sequence \( a \equiv \{a_1, \ldots, a_n\} \)
when we want to send a "1" we send a sequence \( b \equiv \{b_1, \ldots, b_n\} \)

(Both \( a \) and \( b \) are known to the receiver.)

Thus, the transmitted sequence is \( X \equiv (X_1, \ldots, X_n) \), where
either \( X_i = a_i \) for all \( i \), or \( X_i = b_i \) for all \( i \).
Ex. 8.10: Signal Detection and the Matched Filter

To improve noise resilience,

- when we want to send a "0" we send a sequence \(a \equiv \{a_1, \ldots, a_n\}\)
- when we want to send a "1" we send a sequence \(b \equiv \{b_1, \ldots, b_n\}\)

(Both \(a\) and \(b\) are known to the receiver.)

Thus, the transmitted sequence is \(X \equiv \{X_1, \ldots, X_n\}\), where either \(X_i = a_i\) for all \(i\), or \(X_i = b_i\) for all \(i\).

Suppose we have a fixed "energy budget" \(\mathcal{E}\), i.e.,

\[
a_1^2 + \ldots + a_n^2 = b_1^2 + \ldots + b_n^2 = \mathcal{E}
\]
Ex. 8.10: Signal Detection and the Matched Filter

To improve noise resilience,
when we want to send a "0" we send a sequence \( a \equiv (a_1, \ldots, a_n) \)
when we want to send a "1" we send a sequence \( b \equiv (b_1, \ldots, b_n) \)
(Both \( a \) and \( b \) are known to the receiver.)

Thus, the transmitted sequence is \( X \equiv (X_1, \ldots, X_n) \), where either \( X_i = a_i \) for all \( i \), or \( X_i = b_i \) for all \( i \).

Suppose we have a fixed "energy budget" \( \mathcal{E} \), i.e,
\[
a_1^2 + \ldots + a_n^2 = b_1^2 + \ldots + b_n^2 = \mathcal{E}
\]

Assume that the prior probabilities for \( X = a \) and \( X = b \) are both 1/2.
Ex. 8.10: Signal Detection and the Matched Filter

To improve noise resilience,

when we want to send a "0" we send a sequence \( \mathbf{a} \equiv (a_1, \ldots, a_n) \)
when we want to send a "1" we send a sequence \( \mathbf{b} \equiv (b_1, \ldots, b_n) \)

(Both \( \mathbf{a} \) and \( \mathbf{b} \) are known to the receiver.)

Thus, the transmitted sequence is \( \mathbf{X} \equiv (X_1, \ldots, X_n) \), where
either \( X_i = a_i \) for all \( i \), or \( X_i = b_i \) for all \( i \).

Suppose we have a fixed "energy budget" \( \mathcal{E} \), i.e,
\[
a_1^2 + \ldots + a_n^2 = b_1^2 + \ldots + b_n^2 = \mathcal{E}
\]

Assume that the prior probabilities for \( \mathbf{X} = \mathbf{a} \) and \( \mathbf{X} = \mathbf{b} \) are both 1/2.

Assume that the received sequence is \( \mathbf{Y} \equiv (Y_1, \ldots, Y_n) \), with \( Y_i = X_i + W_i \).
Ex. 8.10: Signal Detection and the Matched Filter

To improve noise resilience,
when we want to send a "0" we send a sequence \(a \equiv (a_1, \ldots, a_n)\)
when we want to send a "1" we send a sequence \(b \equiv (b_1, \ldots, b_n)\)
(Both \(a\) and \(b\) are known to the receiver.)

Thus, the transmitted sequence is \(X \equiv (X_1, \ldots, X_n)\), where
either \(X_i = a_i\) for all \(i\), or \(X_i = b_i\) for all \(i\).

Suppose we have a fixed "energy budget" \(\mathcal{E}\), i.e,
\[
a_1^2 + \ldots + a_n^2 = b_1^2 + \ldots + b_n^2 = \mathcal{E}
\]

Assume that the prior probabilities for \(X = a\) and \(X = b\) are both \(1/2\).

Assume that the received sequence is \(Y \equiv (Y_1, \ldots, Y_n)\), with \(Y_i = X_i + W_i\),
where the noise random variables \(W_i\) are standard Gaussian
Ex. 8.10: Signal Detection and the Matched Filter

To improve noise resilience,
  when we want to send a "0" we send a sequence $a \equiv \{a_1, \ldots, a_n\}$
  when we want to send a "1" we send a sequence $b \equiv \{b_1, \ldots, b_n\}$
(Both $a$ and $b$ are known to the receiver.)

Thus, the transmitted sequence is $X \equiv (X_1, \ldots, X_n)$, where either $X_i = a_i$ for all $i$, or $X_i = b_i$ for all $i$.

Suppose we have a fixed "energy budget" $\mathcal{E}$, i.e,
$$a_1^2 + \ldots + a_n^2 = b_1^2 + \ldots + b_n^2 = \mathcal{E}$$

Assume that the prior probabilities for $X = a$ and $X = b$ are both 1/2.

Assume that the received sequence is $Y \equiv (Y_1, \ldots, Y_n)$, with $Y_i = X_i + W_i$,
where the noise random variables $W_i$ are standard Gaussian, independent of each other and independent of $X$. 

Ex. 8.10: Signal Detection and the Matched Filter

Find the MAP detector
Ex. 8.10: What is the conditional density of $Y$ given $X=a$?

$$f_{Y|X}(y \mid x = a) = \frac{\partial^n}{\partial y_1 \ldots \partial y_n} F_{Y|X}(y \mid x = a)$$
Ex. 8.10: What is the conditional density of $Y$ given $X=a$?

$$f_{Y|X}(y \mid x = a) = \frac{\partial^n}{\partial y_1 \ldots \partial y_n} F_{Y|X}(y \mid x = a) = \frac{\partial^n}{\partial y_1 \ldots \partial y_n} P(Y_1 \leq y_1, \ldots, Y_n \leq y_n \mid X = a)$$
Ex. 8.10: What is the conditional density of $Y$ given $X=a$?

$$f_{Y|X}(y \mid x = a) = \frac{\partial^n}{\partial y_1 \ldots \partial y_n} F_{Y|X}(y \mid x = a) = \frac{\partial^n}{\partial y_1 \ldots \partial y_n} P(Y_1 \leq y_1, \ldots, Y_n \leq y_n \mid X = a)$$

$$= \frac{\partial^n}{\partial y_1 \ldots \partial y_n} P(a_1 + W_1 \leq y_1, \ldots, a_n + W_n \leq y_n \mid X = a)$$
Ex. 8.10: What is the conditional density of $Y$ given $X=a$?

$$f_{Y|X}(y \mid x = a) = \frac{\partial^n}{\partial y_1 \cdots \partial y_n} F_{Y|X}(y \mid x = a) = \frac{\partial^n}{\partial y_1 \cdots \partial y_n} P(Y_1 \leq y_1, \ldots, Y_n \leq y_n \mid X = a)$$

$$= \frac{\partial^n}{\partial y_1 \cdots \partial y_n} P(a_1 + W_1 \leq y_1, \ldots, a_n + W_n \leq y_n \mid X = a)$$

$$= \frac{\partial^n}{\partial y_1 \cdots \partial y_n} [P(a_1 + W_1 \leq y_1) \cdots P(a_n + W_n \leq y_n)]$$

because $W_1, \ldots, W_n, X$ are independent.
Ex. 8.10: What is the conditional density of $Y$ given $X=\mathbf{a}$?

\[
f_{Y \mid X}(y \mid x = \mathbf{a}) = \frac{\partial^n}{\partial y_1 \ldots \partial y_n} F_{Y \mid X}(y \mid x = \mathbf{a}) = \frac{\partial^n}{\partial y_1 \ldots \partial y_n} P(Y_1 \leq y_1, \ldots, Y_n \leq y_n \mid X = \mathbf{a})
\]

\[
= \frac{\partial^n}{\partial y_1 \ldots \partial y_n} P(a_1 + W_1 \leq y_1, \ldots, a_n + W_n \leq y_n \mid X = \mathbf{a})
\]

\[
= \frac{\partial^n}{\partial y_1 \ldots \partial y_n} \left[ P(a_1 + W_1 \leq y_1) \cdots P(a_n + W_n \leq y_n) \right]
\]

\[
= \frac{\partial^n}{\partial y_1 \ldots \partial y_n} \left[ P(W_1 \leq y_1 - a_1) \cdots P(W_n \leq y_n - a_n) \right]
\]
Ex. 8.10: What is the conditional density of $Y$ given $X=a$?

$$f_{Y|X}(y \mid x = a) = \frac{\partial^n}{\partial y_1 \ldots \partial y_n} F_{Y|X}(y \mid x = a) = \frac{\partial^n}{\partial y_1 \ldots \partial y_n} P(Y_1 \leq y_1, \ldots, Y_n \leq y_n \mid X = a)$$

$$= \frac{\partial^n}{\partial y_1 \ldots \partial y_n} P(a_1 + W_1 \leq y_1, \ldots, a_n + W_n \leq y_n \mid X = a)$$

$$= \frac{\partial^n}{\partial y_1 \ldots \partial y_n} \left[ P(a_1 + W_1 \leq y_1) \cdot \ldots \cdot P(a_n + W_n \leq y_n) \right]$$

$$= \frac{\partial^n}{\partial y_1 \ldots \partial y_n} \left[ P(W_1 \leq y_1 - a_1) \cdot \ldots \cdot P(W_n \leq y_n - a_n) \right]$$

$$= \frac{\partial^n}{\partial y_1 \ldots \partial y_n} \left[ F_{W_1}(y_1 - a_1) \cdot \ldots \cdot F_{W_n}(y_n - a_n) \right]$$
Ex. 8.10: What is the conditional density of $Y$ given $X = a$?

\[
f_{Y|X}(y \mid x = a) = \frac{\partial^n}{\partial y_1 \cdots \partial y_n} F_{Y|X}(y \mid x = a) = \frac{\partial^n}{\partial y_1 \cdots \partial y_n} P(Y_1 \leq y_1, \ldots, Y_n \leq y_n \mid X = a)
\]

\[
= \frac{\partial^n}{\partial y_1 \cdots \partial y_n} P(a_1 + W_1 \leq y_1, \ldots, a_n + W_n \leq y_n \mid X = a)
\]

\[
= \frac{\partial^n}{\partial y_1 \cdots \partial y_n} \left[ P(a_1 + W_1 \leq y_1) \cdots P(a_n + W_n \leq y_n) \right]
\]

\[
= \frac{\partial^n}{\partial y_1 \cdots \partial y_n} \left[ P(W_1 \leq y_1 - a_1) \cdots P(W_n \leq y_n - a_n) \right]
\]

\[
= \frac{\partial^n}{\partial y_1 \cdots \partial y_n} \left[ F_{W_1}(y_1 - a_1) \cdots F_{W_n}(y_n - a_n) \right]
\]

\[
= f_{W_1}(y_1 - a_1) \cdots f_{W_n}(y_n - a_n)
\]
Ex. 8.10: What is the conditional density of $Y$ given $X=a$?

$$f_{Y|X}(y \mid x = a) = \frac{\partial^n}{\partial y_1 \ldots \partial y_n} F_{Y|X}(y \mid x = a) = \frac{\partial^n}{\partial y_1 \ldots \partial y_n} P(Y_1 \leq y_1, \ldots, Y_n \leq y_n \mid X = a)$$

$$= \frac{\partial^n}{\partial y_1 \ldots \partial y_n} P(a_1 + W_1 \leq y_1, \ldots, a_n + W_n \leq y_n \mid X = a)$$

$$= \frac{\partial^n}{\partial y_1 \ldots \partial y_n} \left[ P(a_1 + W_1 \leq y_1) \cdot \ldots \cdot P(a_n + W_n \leq y_n) \right]$$

$$= \frac{\partial^n}{\partial y_1 \ldots \partial y_n} \left[ P(W_1 \leq y_1 - a_1) \cdot \ldots \cdot P(W_n \leq y_n - a_n) \right]$$

$$= \frac{\partial^n}{\partial y_1 \ldots \partial y_n} \left[ F_{W_1}(y_1 - a_1) \cdot \ldots \cdot F_{W_n}(y_n - a_n) \right]$$

$$= f_{W_1}(y_1 - a_1) \cdot \ldots \cdot f_{W_n}(y_n - a_n) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_1-a_1)^2}{2}} \cdot \ldots \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_n-a_n)^2}{2}}$$
Ex. 8.10: What is the conditional density of $Y$ given $X=a$?

$$f_{Y|X}(y \mid x = a) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_1-a_1)^2}{2}} \cdot \ldots \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_n-a_n)^2}{2}}$$

$$= (2\pi)^{-n/2} \exp \left( -\frac{1}{2} \sum_{i=1}^{n} (y_i - a_i)^2 \right)$$
Ex. 8.10: Similarly, the conditional density of $Y$ given $X=b$ is:

$$f_{Y|X}(y \mid x = b) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_1-b_1)^2}{2}} \cdot \ldots \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_n-b_n)^2}{2}}$$

$$= (2\pi)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (y_i - b_i)^2\right)$$
Ex. 8.10: MAP decision rule

\[
P(X = a \mid Y = y) = \frac{f_{y|x}(y \mid x = a)P(X = a)}{f_Y(y)} \quad \text{and} \quad P(X = b \mid Y = y) = \frac{f_{y|x}(y \mid x = b)P(X = b)}{f_Y(y)}
\]
Ex. 8.10: MAP decision rule

\[
P(X = a \mid Y = y) = \frac{f_{Y \mid X}(y \mid x = a)P(X = a)}{f_{Y}(y)} \quad \text{and} \quad P(X = b \mid Y = y) = \frac{f_{Y \mid X}(y \mid x = b)P(X = b)}{f_{Y}(y)}
\]

Since \( P(X = a) = P(X = b) = 1/2 \), the MAP decision is equivalent to:

\[
f_{Y \mid X}(y \mid x = a) > f_{Y \mid X}(y \mid x = b)
\]

say \( \hat{x} = a \)

\[
f_{Y \mid X}(y \mid x = a) < f_{Y \mid X}(y \mid x = b)
\]

say \( \hat{x} = b \)
**Ex. 8.10: MAP decision rule**

\[
P(X = a \mid Y = y) = \frac{f_{Y \mid X}(y \mid x = a)P(X = a)}{f_Y(y)} \quad \text{and} \quad P(X = b \mid Y = y) = \frac{f_{Y \mid X}(y \mid x = b)P(X = b)}{f_Y(y)}
\]

Since \( P(X = a) = P(X = b) = 1/2 \), the MAP decision is equivalent to:

\[\begin{align*}
\text{say } \hat{X} &= a \\
& \quad \left( f_{Y \mid X}(y \mid x = a) > f_{Y \mid X}(y \mid x = b) \right) \\
& \quad \left( \text{say } \hat{X} = b \right)
\end{align*}\]

\[
(2\pi)^{-n/2} \exp\left( -\frac{1}{2} \sum_{i=1}^{n} (y_i - a_i)^2 \right) > (2\pi)^{-n/2} \exp\left( -\frac{1}{2} \sum_{i=1}^{n} (y_i - b_i)^2 \right)
\]

\[\begin{align*}
& \quad \left( \text{say } \hat{X} = a \right) \\
& \quad \left( \text{say } \hat{X} = b \right)
\end{align*}\]
Ex. 8.10: MAP decision rule

\[
P(X = a \mid Y = y) = \frac{f_{Y \mid X}(y \mid x = a)P(X = a)}{f_Y(y)} \text{ and } P(X = b \mid Y = y) = \frac{f_{Y \mid X}(y \mid x = b)P(X = b)}{f_Y(y)}
\]

Since \( P(X = a) = P(X = b) = 1/2 \), the MAP decision is equivalent to:

say \( \hat{x} = a \)

\[
f_{Y \mid X}(y \mid x = a) > \quad f_{Y \mid X}(y \mid x = b)
\]

say \( \hat{x} = b \)

\[
(2\pi)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (y_i - a_i)^2 \right) > \quad (2\pi)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (y_i - b_i)^2 \right)
\]

say \( \hat{x} = a \)

\[
\sum_{i=1}^{n} (y_i - a_i)^2 < \quad \sum_{i=1}^{n} (y_i - b_i)^2
\]

say \( \hat{x} = b \)
Ex. 8.10: MAP decision rule

\[\sum_{i=1}^{n} (y_i - a_i)^2 < \sum_{i=1}^{n} (y_i - b_i)^2\]

say \(\hat{x} = a\)

say \(\hat{x} = b\)
Ex. 8.10: MAP decision rule

\begin{align*}
\text{say } \hat{x} &= a \\
\sum_{i=1}^{n} (y_i - a_i)^2 &< \sum_{i=1}^{n} (y_i - b_i)^2 \\
\text{say } \hat{x} &= b \\
\sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} y_i a_i + \sum_{i=1}^{n} a_i^2 &< \sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} y_i b_i + \sum_{i=1}^{n} b_i^2 \\
\text{say } \hat{x} &= b \\
\sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} y_i a_i + \sum_{i=1}^{n} a_i^2 &> \sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} y_i b_i + \sum_{i=1}^{n} b_i^2
\end{align*}
Ex. 8.10: MAP decision rule

\[ \sum_{i=1}^{n} (y_i - a_i)^2 < \sum_{i=1}^{n} (y_i - b_i)^2 \]

say \( \hat{x} = a \)

\[ \sum_{i=1}^{n} y_i^2 - 2\sum_{i=1}^{n} y_i a_i + \sum_{i=1}^{n} a_i^2 < \sum_{i=1}^{n} y_i^2 - 2\sum_{i=1}^{n} y_i b_i + \sum_{i=1}^{n} b_i^2 \]

say \( \hat{x} = a \)

\[ \sum_{i=1}^{n} y_i a_i > \sum_{i=1}^{n} y_i b_i \]

say \( \hat{x} = b \)
Ex. 8.10: MAP decision rule

\[
\sum_{i=1}^{n} (y_i - a_i)^2 < \sum_{i=1}^{n} (y_i - b_i)^2
\]

say \( \hat{x} = a \)

\[
\sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} y_i a_i + \sum_{i=1}^{n} a_i^2 < \sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} y_i b_i + \sum_{i=1}^{n} b_i^2
\]

say \( \hat{x} = a \)

\[
\sum_{i=1}^{n} y_i a_i > \sum_{i=1}^{n} y_i b_i
\]

say \( \hat{x} = b \)

\[
\sum_{i=1}^{n} (y_i - a_i)^2 > \sum_{i=1}^{n} (y_i - b_i)^2
\]

say \( \hat{x} = b \)

\[
\sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} y_i b_i + \sum_{i=1}^{n} b_i^2 < \sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} y_i a_i + \sum_{i=1}^{n} a_i^2
\]

say \( \hat{x} = b \)

**Matched filter:** we match the received signal with each of the two candidate signals by forming the inner products.
Ex. 8.10: MAP decision rule

\[
\sum_{i=1}^{n} (y_i - a_i)^2 < \sum_{i=1}^{n} (y_i - b_i)^2 \quad \text{say } \hat{x} = a
\]

\[
\sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} y_i a_i + \sum_{i=1}^{n} a_i^2 < \sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} y_i b_i + \sum_{i=1}^{n} b_i^2 \quad \text{say } \hat{x} = a
\]

\[
\sum_{i=1}^{n} y_i a_i < \sum_{i=1}^{n} y_i b_i \quad \text{say } \hat{x} = a
\]

\[
\langle y, a \rangle > \langle y, b \rangle \quad \text{say } \hat{x} = a
\]

\[
\langle y, b \rangle
\]

\[
\text{Matched filter: we match the received signal with each of the two candidate signals by forming the inner products.}
\]
Ex. 8.10: MAP decision rule

\[ \sum_{i=1}^{n} (y_i - a_i)^2 < \sum_{i=1}^{n} (y_i - b_i)^2 \]

say \( \hat{x} = a \)

\[ \sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} y_i a_i + \sum_{i=1}^{n} a_i^2 \]
\[ > \sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} y_i b_i + \sum_{i=1}^{n} b_i^2 \]

say \( \hat{x} = a \)

\[ \sum_{i=1}^{n} y_i a_i \]
\[ < \sum_{i=1}^{n} y_i b_i \]

say \( \hat{x} = a \)

\[ \langle y, a \rangle \]
\[ < \langle y, b \rangle \]

say \( \hat{x} = a \)

\[ \langle y, a \rangle \]
\[ < \langle y, b \rangle \]

say \( \hat{x} = b \)

**Matched filter**: we match the received signal with each of the two candidate signals by forming the inner products. We then select the hypothesis that corresponds to the higher value ("the best match").
Ex 8.5: Spam Filtering

• An email system receives a message, and needs to automatically determine whether it is spam or legitimate.
Ex 8.5: Spam Filtering

- An email system receives a message, and needs to automatically determine whether it is spam or legitimate.

Model: Bernoulli r.v. $X$,

$$X = \begin{cases} 
0, & \text{if the message is legitimate} \\
1, & \text{if the message is spam}
\end{cases}$$
Ex 8.5: Spam Filtering

- An email system receives a message, and needs to automatically determine whether it is spam or legitimate.

Model: Bernoulli r.v. $X$,

$$X = \begin{cases} 
0, & \text{if the message is legitimate} \\
1, & \text{if the message is spam} 
\end{cases}$$

$p_X(0) \equiv P(X = 0)$ and $p_X(1) \equiv P(X = 1)$ are assumed to be known.
Ex 8.5: Spam Filtering

• An email system receives a message, and needs to automatically determine whether it is spam or legitimate.

Model: Bernoulli r.v. $X$,

$$X = \begin{cases} 
0, & \text{if the message is legitimate} \\
1, & \text{if the message is spam} 
\end{cases}$$

Suppose $w_1, \ldots, w_n$ are special words or phrases whose appearance suggests spam.

Let $Y_i = \begin{cases} 
0, & \text{if } w_i \text{ does not appear in the message} \\
1, & \text{if } w_i \text{ appears in the message} 
\end{cases}$
Ex 8.5: Spam Filtering

• An email system receives a message, and needs to automatically determine whether it is spam or legitimate.

Model: Bernoulli r.v. $X$,

$$X = \begin{cases} 
0, & \text{if the message is legitimate} \\
1, & \text{if the message is spam} \end{cases}$$

Suppose $w_1, \ldots, w_n$ are special words or phrases whose appearance suggests spam.

Let $Y_i = \begin{cases} 
0, & \text{if } w_i \text{ does not appear in the message} \\
1, & \text{if } w_i \text{ appears in the message} \end{cases}$

Assume $p_{Y_i|X}(y_i | 0)$ and $p_{Y_i|X}(y_i | 1)$ are known.
Ex 8.5: Spam Filtering

• An email system receives a message, and needs to automatically determine whether it is spam or legitimate.

Model: Bernoulli r.v. $X$,

$$X = \begin{cases} 
0, & \text{if the message is legitimate} \\
1, & \text{if the message is spam} 
\end{cases}$$

Suppose $w_1, \ldots, w_n$ are special words or phrases whose appearance suggests spam.

Let $Y_i = \begin{cases} 
0, & \text{if } w_i \text{ does not appear in the message} \\
1, & \text{if } w_i \text{ appears in the message} 
\end{cases}$

Assume $p_{Y_i|X}(y_i | 0)$ and $p_{Y_i|X}(y_i | 1)$ are known.

Assume $Y_1, \ldots, Y_n$ are conditionally independent given $X$. 
Ex 8.5: Spam Filtering

• An email system receives a message, and needs to automatically determine whether it is spam or legitimate.

Model: Bernoulli r.v. $X$,

$$X = \begin{cases} 
0, & \text{if the message is legitimate} \\
1, & \text{if the message is spam} 
\end{cases}$$

Suppose $w_1, \ldots, w_n$ are special words or phrases whose appearance suggests spam.

Let $Y_i = \begin{cases} 
0, & \text{if } w_i \text{ does not appear in the message} \\
1, & \text{if } w_i \text{ appears in the message} 
\end{cases}$

Assume $p_{Y_i|X}(y_i \mid 0)$ and $p_{Y_i|X}(y_i \mid 1)$ are known.

Assume $Y_1, \ldots, Y_n$ are conditionally independent given $X$.

$$p_{X|Y}(x \mid y) = \frac{p_X(x)p_{Y|X}(y \mid x)}{p_Y(y)}$$
Ex 8.5: Spam Filtering

• An email system receives a message, and needs to automatically determine whether it is spam or legitimate.

Model: Bernoulli r.v. $X$,

$$X = \begin{cases} 
0, & \text{if the message is legitimate} \\
1, & \text{if the message is spam} 
\end{cases}$$

Suppose $w_1, \ldots, w_n$ are special words or phrases whose appearance suggests spam.

Let $Y_i = \begin{cases} 
0, & \text{if } w_i \text{ does not appear in the message} \\
1, & \text{if } w_i \text{ appears in the message} 
\end{cases}$

Assume $p_{Y_i|X}(y_i | 0)$ and $p_{Y_i|X}(y_i | 1)$ are known.

Assume $Y_1, \ldots, Y_n$ are conditionally independent given $X$.

$$p_{X|Y}(x | y) = \frac{p_X(x) p_{Y|X}(y | x)}{p_Y(y)} = \frac{p_X(x) \prod_{i=1}^{n} p_{Y_i|X}(y_i | x)}{p_Y(y)}$$
Ex 8.5: MAP Spam Filtering

\[
p_{X|Y}(x | y) = \frac{p_X(x)p_{Y|x}(y | x)}{p_Y(y)} = \frac{p_X(x)\prod_{i=1}^{n} p_{Y_i|x}(y_i | x)}{p_Y(y)}
\]

MAP decision rule:

\[
p_X(1)\prod_{i=1}^{n} p_{Y_i|x}(y_i | 1) > p_X(0)\prod_{i=1}^{n} p_{Y_i|x}(y_i | 0)
\]

say \( \hat{x} = 1 \)

\[
p_X(1)\prod_{i=1}^{n} p_{Y_i|x}(y_i | 1) < p_X(0)\prod_{i=1}^{n} p_{Y_i|x}(y_i | 0)
\]

say \( \hat{x} = 0 \)