(d) The solution to part (b) also applies here.

Section 2.2

2. We are given that the density is of the form $p(x|\omega_i) = ke^{-|x-a_i|/b_i}$.

(a) We seek k so that the function is normalized, as required by a true density. We integrate this function, set it to 1.0,

$$k\left[\int_{-\infty}^{a_i} \exp[(x-a_i)/b_i]dx + \int_{a_i}^{\infty} \exp[-(x-a_i)/b_i]dx\right] = 1,$$

which yields $2b_ik = 1$ or $k = 1/(2b_i)$. Note that the normalization is independent of a_i, which corresponds to a shift along the axis and is hence indeed irrelevant to normalization. The distribution is therefore written

$$p(x|\omega_i) = \frac{1}{2b_i}e^{-|x-a_i|/b_i}.$$

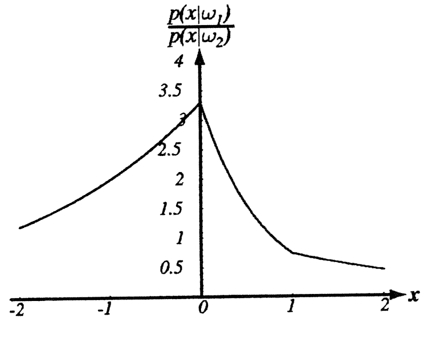
(b) The likelihood ratio can be written directly:

$$\frac{p(x|\omega_1)}{p(x|\omega_2)} = \frac{b_2}{b_1}\exp\left[-\frac{|x-a_1|}{b_1} + \frac{|x-a_2|}{b_2}\right].$$

(c) For the case $a_1 = 0$, $a_2 = 1$, $b_1 = 1$ and $b_2 = 2$, we have the likelihood ratio is

$$\frac{p(x|\omega_2)}{p(x|\omega_1)} = \begin{cases} 2e^{(x+1)/2} & x \leq 0 \\ 2e^{(1-3x)/2} & 0 < x \leq 1 \\ 2e^{(-x-1)/2} & x > 1, \end{cases}$$

as shown in the figure.



Section 2.3

3. We are are to use the standard zero-one classification cost, that is $\lambda_{11} = \lambda_{22} = 0$ and $\lambda_{12} = \lambda_{21} = 1$.

(d) The solution to part (b) also applies here.

Section 2.2

2. We are given that the density is of the form $p(x|\omega_i) = ke^{-|x-a_i|/b_i}$.

(a) We seek k so that the function is normalized, as required by a true density. We integrate this function, set it to 1.0,

$$k\left[\int_{-\infty}^{a_i} \exp[(x-a_i)/b_i]dx + \int_{a_i}^{\infty} \exp[-(x-a_i)/b_i]dx\right] = 1,$$

which yields $2b_i k = 1$ or $k = 1/(2b_i)$. Note that the normalization is independent of a_i, which corresponds to a shift along the axis and is hence indeed irrelevant to normalization. The distribution is therefore written

$$p(x|\omega_i) = \frac{1}{2b_i}e^{-|x-a_i|/b_i}.$$

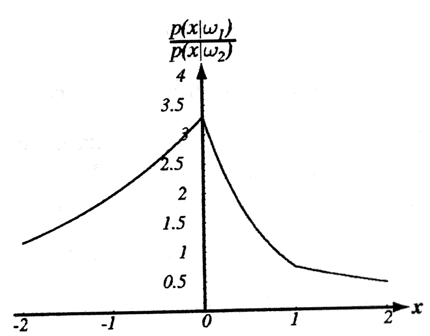
(b) The likelihood ratio can be written directly:

$$\frac{p(x|\omega_1)}{p(x|\omega_2)} = \frac{b_2}{b_1}\exp\left[-\frac{|x-a_1|}{b_1} + \frac{|x-a_2|}{b_2}\right].$$

(c) For the case $a_1 = 0$, $a_2 = 1$, $b_1 = 1$ and $b_2 = 2$, we have the likelihood ratio is

$$\frac{p(x|\omega_2)}{p(x|\omega_1)} = \begin{cases} 2e^{(x+1)/2} & x \le 0 \\ 2e^{(1-3x)/2} & 0 < x \le 1 \\ 2e^{(-x-1)/2} & x > 1, \end{cases}$$

as shown in the figure.



Section 2.3

3. We are are to use the standard zero-one classification cost, that is $\lambda_{11} = \lambda_{22} = 0$ and $\lambda_{12} = \lambda_{21} = 1$.

(a) We have the priors $P(\omega_1)$ and $P(\omega_2) = 1 - P(\omega_1)$. The Bayes risk is given by Eqs. 12 and 13 in the text:

$$R(P(\omega_1)) = P(\omega_1)\int_{\mathcal{R}_2} p(x|\omega_1)dx + (1-P(\omega_1))\int_{\mathcal{R}_1} p(x|\omega_2)dx.$$

To obtain the prior with the minimum risk, we take the derivative with respect to $P(\omega_1)$ and set it to 0, that is

$$\frac{d}{dP(\omega_1)}R(P(\omega_1)) = \int_{\mathcal{R}_2} p(x|\omega_1)dx - \int_{\mathcal{R}_1} p(x|\omega_2)dx = 0,$$

which gives the desired result:

$$\int_{\mathcal{R}_2} p(x|\omega_1)dx = \int_{\mathcal{R}_1} p(x|\omega_2)dx.$$

(b) This solution is not always unique, as shown in this simple counterexample. Let $P(\omega_1) = P(\omega_2) = 0.5$ and

$$p(x|\omega_1) = \begin{cases} 1 & -0.5 \le x \le 0.5 \\ 0 & \text{otherwise} \end{cases}$$

$$p(x|\omega_2) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to verify that the decision regions $\mathcal{R}_1 = [-0.5, 0.25]$ and $\mathcal{R}_1 = [0, 0.5]$ satisfy the equations in part (a); thus the solution is not unique.

4. Consider the minimax criterion for a two-category classification problem.

(a) The total risk is the integral over the two regions $\mathcal{R}_i$ of the posteriors times their costs:

$$\begin{aligned} R &= \int_{\mathcal{R}_1} [\lambda_{11}P(\omega_1)p(\mathbf{x}|\omega_1) + \lambda_{12}P(\omega_2)p(\mathbf{x}|\omega_2)]\ d\mathbf{x} \\ &\quad + \int_{\mathcal{R}_2} [\lambda_{21}P(\omega_1)p(\mathbf{x}|\omega_1) + \lambda_{22}P(\omega_2)p(\mathbf{x}|\omega_2)]\ d\mathbf{x}. \end{aligned}$$

We use $\int_{\mathcal{R}_2} p(\mathbf{x}|\omega_2)\ d\mathbf{x} = 1 - \int_{\mathcal{R}_1} p(\mathbf{x}|\omega_2)\ d\mathbf{x}$ and $P(\omega_2) = 1 - P(\omega_1)$, regroup to find:

$$\begin{aligned} R &= \lambda_{22} + \lambda_{12}\int_{\mathcal{R}_1} p(\mathbf{x}|\omega_2)\ d\mathbf{x} - \lambda_{22}\int_{\mathcal{R}_1} p(\mathbf{x}|\omega_2)\ d\mathbf{x} \\ &\quad + P(\omega_1)\Bigg[(\lambda_{11} - \lambda_{22}) + \lambda_{11}\int_{\mathcal{R}_2} p(\mathbf{x}|\omega_1)\ d\mathbf{x} - \lambda_{12}\int_{\mathcal{R}_1} p(\mathbf{x}|\omega_2)\ d\mathbf{x} \\ &\quad + \lambda_{21}\int_{\mathcal{R}_2} p(\mathbf{x}|\omega_1)\ d\mathbf{x} + \lambda_{22}\int_{\mathcal{R}_1} p(\mathbf{x}|\omega_2)\ d\mathbf{x}\Bigg] \end{aligned}$$

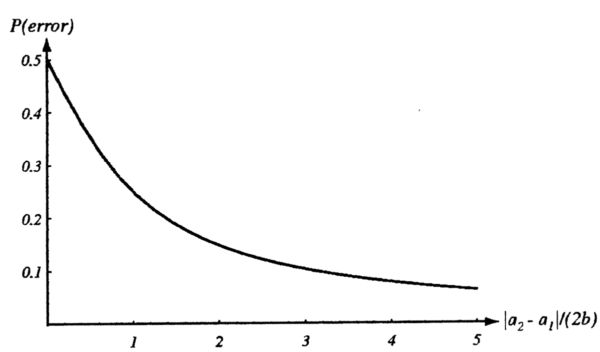
(a) Without loss of generality, we assume that $a_2 > a_1$, note that the decision boundary is at $(a_1 + a_2)/2$. The probability of error is given by

$$\begin{aligned}
P(error) &= \int_{-\infty}^{(a_1+a_2)/2} p(\omega_2|x)dx + \int_{(a_1+a_2)/2}^{\infty} p(\omega_1|x)dx \\
&= \frac{1}{\pi b}\int_{-\infty}^{(a_1+a_2)/2} \frac{1/2}{1+\left(\frac{x-a_2}{b}\right)^2}dx + \frac{1}{\pi b}\int_{(a_1+a_2)/2}^{\infty} \frac{1/2}{1+\left(\frac{x-a_1}{b}\right)^2}dx \\
&= \frac{1}{\pi b}\int_{-\infty}^{(a_1-a_2)/2} \frac{1}{1+\left(\frac{x-a_2}{b}\right)^2}dx = \frac{1}{\pi}\int_{-\infty}^{(a_1-a_2)/2} \frac{1}{1+y^2}dy,
\end{aligned}$$

where for the last step we have used the trigonometric substitution $y = (x-a_2)/b$ as in Problem 8. The integral is a standard form for $\tan^{-1}y$ and thus our solution is:

$$\begin{aligned}
P(error) &= \frac{1}{\pi}\left[\tan^{-1}\left|\frac{a_1-a_2}{2b}\right| - \tan^{-1}[-\infty]\right] \\
&= \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2-a_1}{2b}\right|.
\end{aligned}$$

(b) SEE FIGURE.



(c) The maximum value of the probability of error is $P_{max}(\frac{a_2-a_1}{2b}) = 1/2$, which occurs for $|\frac{a_2-a_1}{2b}| = 0$. This occurs when either the two distributions are the same, which can happen because $a_1 = a_2$, or even if $a_1 \neq a_2$ because $b = \infty$ and both distributions are flat.

10. We use the fact that the conditional error is

$$P(error|x) = \begin{cases} P(\omega_1|x) \text{ if we decide } \omega_2 \\ P(\omega_2|x) \text{ if we decide } \omega_1. \end{cases}$$

(a) Thus the decision as stated leads to:

$$P(error) = \int_{-\infty}^{\infty} P(error|x)p(x)dx.$$

(b) The error for the converse case is found similarly:

$$\begin{aligned} E_2 &= \frac{1}{\pi b}\int_{-\infty}^{x^*} \frac{1}{1+\left(\frac{x-a_2}{b}\right)^2} P(\omega_2)\, dx \\ &= \frac{1}{2\pi}\int_{\theta=-\pi}^{\theta=\tilde{\theta}} d\theta \\ &= \frac{1}{2\pi}\left\{\sin^{-1}\left[\frac{b}{\sqrt{b^2+(x^*-a_2)^2}}\right]+\pi\right\} \\ &= \frac{1}{2}+\frac{1}{\pi}\sin^{-1}\left[\frac{b}{\sqrt{b^2+(x^*-a_2)^2}}\right], \end{aligned}$$

where $\tilde{\theta}$ is defined in part (a).

(c) The total error is merely the sum of the component errors:

$$E = E_1 + E_2 = E_1 + \frac{1}{2} + \frac{1}{\pi}\sin^{-1}\left[\frac{b}{\sqrt{b^2+(x^*-a_2)^2}}\right],$$

where the numerical value of the decision point is

$$x^* = a_1 + b/\tan[2\pi E_1] = 0.376.$$

(d) We add the errors (for $b = 1$) and find

$$E = 0.1 + \frac{1}{2} + \frac{1}{\pi}\sin^{-1}\left[\frac{b}{\sqrt{b^2+(x^*-a_2)^2}}\right] = 0.2607.$$

(e) For the Bayes case, the decision point is midway between the peaks of the two distributions, i.e., at $x^* = 0$ (cf. Problem 6). The Bayes error is then

$$E_B = 2\int_{0}^{\infty} \frac{1}{1+\left(\frac{x-a}{b}\right)^2} P(\omega_2)\, dx = 0.2489.$$

This is indeed lower than for the Neyman-Pearson case, as it must be. Note that if the Bayes error were lower than $2 \times 0.1 = 0.2$ in this problem, we would use the Bayes decision point for the Neyman-Pearson case, since it too would ensure that the Neyman-Pearson criteria were obeyed *and* would give the lowest total error.

8. Consider the Cauchy distribution.

(a) We let k denote the integral of $p(x|\omega_i)$, and check the normalization condition, that is, whether $k = 1$:

$$k = \int_{-\infty}^{\infty} p(x|\omega_i)\, dx = \frac{1}{\pi b}\int_{-\infty}^{\infty} \frac{1}{1+\left(\frac{x-a_i}{b}\right)^2}\, dx.$$

We substitute $(x - a_i)/b = y$ into the above and get

$$k = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{1+y^2}\, dy,$$

and use the trigonometric substition $1/\sqrt{1+y^2} = \sin\theta$, and hence $dy = d\theta/\sin^2\theta$ to find

$$k = \frac{1}{\pi}\int_{\theta=-\pi}^{\theta=0} \frac{\sin^2\theta}{\sin^2\theta}\, d\theta = 1.$$

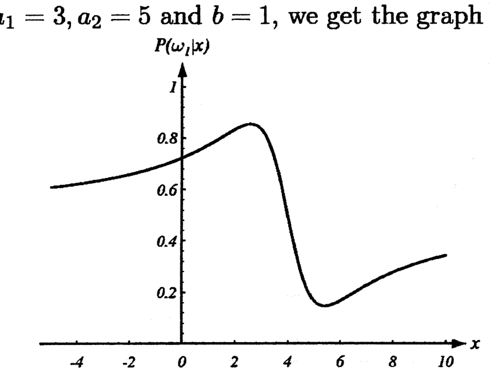
Indeed, $k = 1$, and the distribution is normalized.

(b) We let x^* denote the decision boundary (a single point) and find its value by setting $p(x^*|\omega_1)P(\omega_1) = p(x^*|\omega_2)P(\omega_2)$. We have then

$$\frac{1}{\pi b}\frac{1}{1+\left(\frac{x^*-a_1}{b}\right)^2}\frac{1}{2} = \frac{1}{\pi b}\frac{1}{1+\left(\frac{x^*-a_2}{b}\right)^2}\frac{1}{2},$$

or $(x^* - a_1) = \pm(x^* - a_2)$. For $a_1 \neq a_2$, this implies that $x^* = (a_1 + a_2)/2$, that is, the decision boundary is midway between the means of the two distributions.

(c) For the values $a_1 = 3, a_2 = 5$ and $b = 1$, we get the graph shown in the figure.



(d) We substitute the form of $P(\omega_i|x)$ and $p(x|\omega_i)$ and find

$$\begin{aligned} \lim_{x\to\infty} P(\omega_i|x) &= \lim_{x\to\infty} \frac{\frac{1}{2}\left[\frac{1}{\pi b}\frac{1}{1+\left(\frac{x-a_i}{b}\right)^2}\right]}{\frac{1}{2}\left[\frac{1}{\pi b}\frac{1}{1+\left(\frac{x-a_1}{b}\right)^2}\right]+\frac{1}{2}\left[\frac{1}{\pi b}\frac{1}{1+\left(\frac{x-a_2}{b}\right)^2}\right]} \\ &= \lim_{x\to\infty} \frac{b^2+(x-a_i)^2}{b^2+(x-a_1)^2+b^2+(x-a_2)^2} = \frac{1}{2}, \end{aligned}$$

and likewise, $\lim_{x\to-\infty} P(\omega_i|x) = 1/2$, as can be confirmed in the figure.

9. We follow the terminology in Section 2.3 in the text.

(b) The error for the converse case is found similarly:

$$\begin{aligned}
E_2 &= \frac{1}{\pi b}\int_{-\infty}^{x^*} \frac{1}{1+\left(\frac{x-a_2}{b}\right)^2} P(\omega_2)\, dx \\
&= \frac{1}{2\pi}\int_{\theta=-\pi}^{\theta=\tilde{\theta}} d\theta \\
&= \frac{1}{2\pi}\left\{\sin^{-1}\left[\frac{b}{\sqrt{b^2+(x^*-a_2)^2}}\right]+\pi\right\} \\
&= \frac{1}{2}+\frac{1}{\pi}\sin^{-1}\left[\frac{b}{\sqrt{b^2+(x^*-a_2)^2}}\right],
\end{aligned}$$

where $\tilde{\theta}$ is defined in part (a).

(c) The total error is merely the sum of the component errors:

$$E = E_1 + E_2 = E_1 + \frac{1}{2} + \frac{1}{\pi}\sin^{-1}\left[\frac{b}{\sqrt{b^2+(x^*-a_2)^2}}\right],$$

where the numerical value of the decision point is

$$x^* = a_1 + b/\tan[2\pi E_1] = 0.376.$$

(d) We add the errors (for $b = 1$) and find

$$E = 0.1 + \frac{1}{2} + \frac{1}{\pi}\sin^{-1}\left[\frac{b}{\sqrt{b^2+(x^*-a_2)^2}}\right] = 0.2607.$$

(e) For the Bayes case, the decision point is midway between the peaks of the two distributions, i.e., at $x^* = 0$ (cf. Problem 6). The Bayes error is then

$$E_B = 2\int_0^{\infty} \frac{1}{1+\left(\frac{x-a}{b}\right)^2} P(\omega_2)\, dx = 0.2489.$$

This is indeed lower than for the Neyman-Pearson case, as it must be. Note that if the Bayes error were lower than $2 \times 0.1 = 0.2$ in this problem, we would use the Bayes decision point for the Neyman-Pearson case, since it too would ensure that the Neyman-Pearson criteria were obeyed *and* would give the lowest total error.

8. Consider the Cauchy distribution.

(a) We let k denote the integral of $p(x|\omega_i)$, and check the normalization condition, that is, whether $k = 1$:

$$k = \int_{-\infty}^{\infty} p(x|\omega_i)\, dx = \frac{1}{\pi b}\int_{-\infty}^{\infty} \frac{1}{1+\left(\frac{x-a_i}{b}\right)^2}\, dx.$$

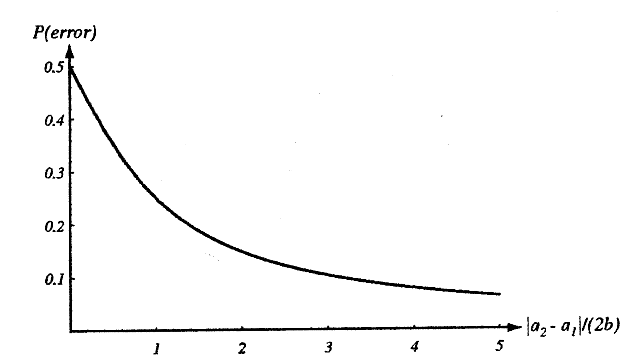
(a) Without loss of generality, we assume that $a_2 > a_1$, note that the decision boundary is at $(a_1 + a_2)/2$. The probability of error is given by

$$\begin{aligned}
P(error) &= \int_{-\infty}^{(a_1+a_2)/2} p(\omega_2|x)dx + \int_{(a_1+a_2)/2}^{\infty} p(\omega_1|x)dx \\
&= \frac{1}{\pi b}\int_{-\infty}^{(a_1+a_2)/2} \frac{1/2}{1+\left(\frac{x-a_2}{b}\right)^2} dx + \frac{1}{\pi b}\int_{(a_1+a_2)/2}^{\infty} \frac{1/2}{1+\left(\frac{x-a_1}{b}\right)^2} dx \\
&= \frac{1}{\pi b}\int_{-\infty}^{(a_1-a_2)/2} \frac{1}{1+\left(\frac{x-a_2}{b}\right)^2} dx = \frac{1}{\pi}\int_{-\infty}^{(a_1-a_2)/2} \frac{1}{1+y^2} dy,
\end{aligned}$$

where for the last step we have used the trigonometric substitution $y = (x-a_2)/b$ as in Problem 8. The integral is a standard form for $\tan^{-1} y$ and thus our solution is:

$$\begin{aligned}
P(error) &= \frac{1}{\pi}\left[\tan^{-1}\left|\frac{a_1 - a_2}{2b}\right| - \tan^{-1}[-\infty]\right] \\
&= \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2 - a_1}{2b}\right|.
\end{aligned}$$

(b) SEE FIGURE.



(c) The maximum value of the probability of error is $P_{max}(\frac{a_2-a_1}{2b}) = 1/2$, which occurs for $|\frac{a_2-a_1}{2b}| = 0$. This occurs when either the two distributions are the same, which can happen because $a_1 = a_2$, or even if $a_1 \neq a_2$ because $b = \infty$ and both distributions are flat.

10. We use the fact that the conditional error is

$$P(error|x) = \begin{cases} P(\omega_1|x) & \text{if we decide } \omega_2 \\ P(\omega_2|x) & \text{if we decide } \omega_1. \end{cases}$$

(a) Thus the decision as stated leads to:

$$P(error) = \int_{-\infty}^{\infty} P(error|x)p(x)dx.$$

Thus we can write the probability of error as

$$\begin{aligned}
P(error) &= P(x < \theta \text{ and } \omega_1 \text{ is the true state}) \\
&\quad + P(x > \theta \text{ and } \omega_2 \text{ is the true state}) \\
&= P(x < \theta|\omega_1)P(\omega_1) + P(x > \theta|\omega_2)P(\omega_2) \\
&= P(\omega_1)\int_{-\infty}^{\theta} p(x|\omega_1)\ dx + P(\omega_2)\int_{\theta}^{\infty} p(x|\omega_2)\ dx.
\end{aligned}$$

(b) We take a derivative with respect to θ and set it to zero to find an extremum, that is,

$$\frac{dP(error)}{d\theta} = P(\omega_1)p(\theta|\omega_1) - P(\omega_2)p(\theta|\omega_2) = 0,$$

which yields the condition

$$P(\omega_1)p(\theta|\omega_1) = P(\omega_2)p(\theta|\omega_2),$$

where we have used the fact that $p(x|\omega_i) = 0$ at $x \to \pm\infty$.

(c) No, this condition does not uniquely define θ.

1. If $P(\omega_1)p(\theta|\omega_1) = P(\omega_2)p(\theta|\omega_2)$ over a *range* of θ, then θ would be unspecified throughout such a range.
2. There can easily be multiple values of x for which the condition hold, for instance if the distributions have the appropriate multiple peaks.

(d) If $p(x|\omega_1) \sim N(1,1)$ and $p(x|\omega_2) \sim N(-1,1)$ with $P(\omega_1) = P(\omega_2) = 1/2$, then we have a *maximum* for the error at $\theta = 0$.

11. The deterministic risk is given by Bayes' Rule and Eq. 20 in the text

$$R = \int R(\alpha_i(\mathbf{x})|\mathbf{x})\ d\mathbf{x}.$$

(a) In a random decision rule, we have the *probability* $P(\alpha_i|\mathbf{x})$ of deciding to take action α_i. Thus in order to compute the full probabilistic or randomized risk, R_{ran}, we must integrate over all the conditional risks weighted by their probabilities, i.e.,

$$R_{ran} = \int \left[\sum_{i=1}^{a} R(\alpha_i(\mathbf{x})|\mathbf{x})P(\alpha_i|\mathbf{x})\right] p(\mathbf{x})\ d\mathbf{x}.$$

(b) Consider a fixed point $\mathbf{x}$ and note that the (deterministic) Bayes minimum risk decision at that point obeys

$$R(\alpha_i(\mathbf{x})|\mathbf{x}) \geq R(\alpha_{max}(\mathbf{x})|\mathbf{x}).$$