Detection and Estimation

Xiang-Gen Xia

Outline

Chapter 1: Introduction and Probability Theory Preparation

Chapter 2: Elements of Hypothesis Testing

1. Bayesian Hypothesis Testing
2. Minimax Hypothesis Testing
3. Neyman-Pearson Hypothesis Testing
4. Composite Hypothesis Testing

Chapter 3: Signal Detection In Discrete Time

1. Deterministic Signal Detection
2. Random Signal Detection

Chapter 4: Elements of Parameter Estimation

1. Bayesian Parameter Estimation
2. Nonrandom Parameter Estimation, Cramer-Rao Bound
3. Maximum-Likelihood Estimation

Chapter 5: Elements of Signal Estimation

1. Kalman-Bucy Filtering
2. Linear Estimation
3. Wiener-Kolmogorov Filtering
   1) Noncausal
   2) Causal

Chapter 6: Signal Detection in Continuous Time

Chapter 7: Signal Estimation in Continuous time

Chapter 1: Introduction and Probability Theory

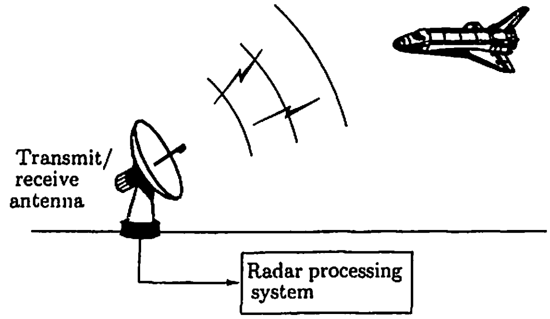
§1 Detection theory and Estimation theory are fundamental to the designs of electronic signal processing systems for decision making and information extraction:

Radar
Communications
Speech
Sonar
Image processing
Biomedicine
Control
Seismology

To decide when/whether an interesting event occurs

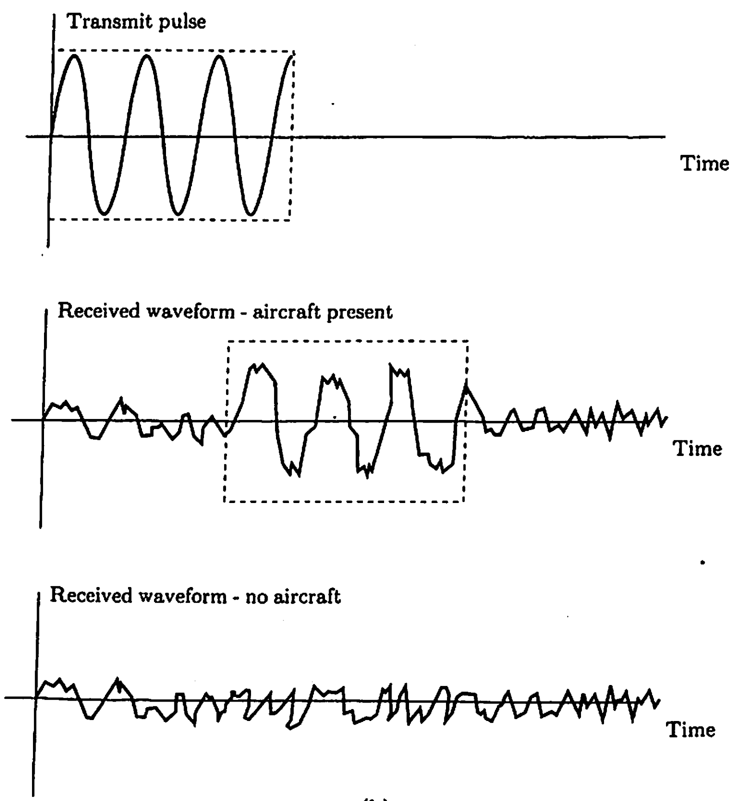
To determine what the event is

Radar system

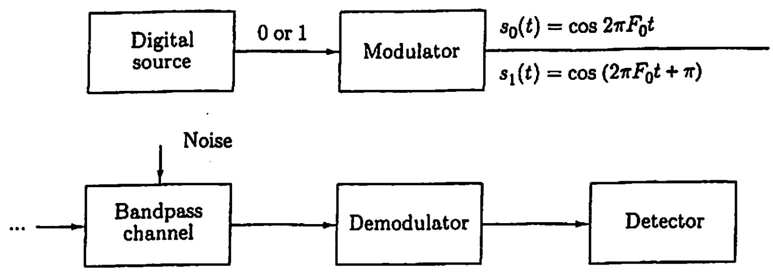


(a)

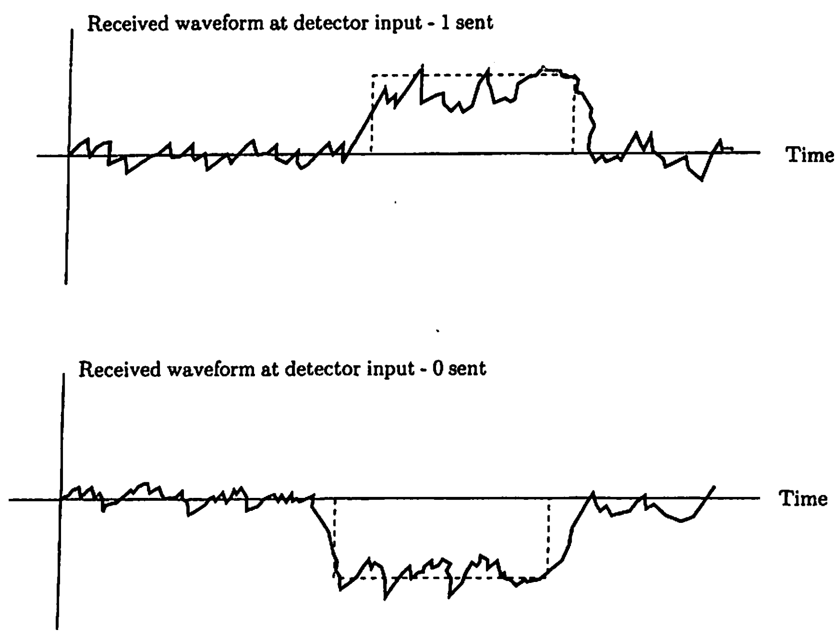
Radar Transmit waveform



(b)



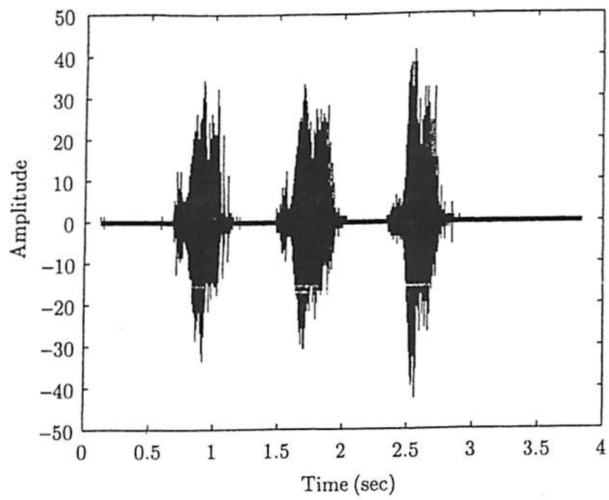
(a)



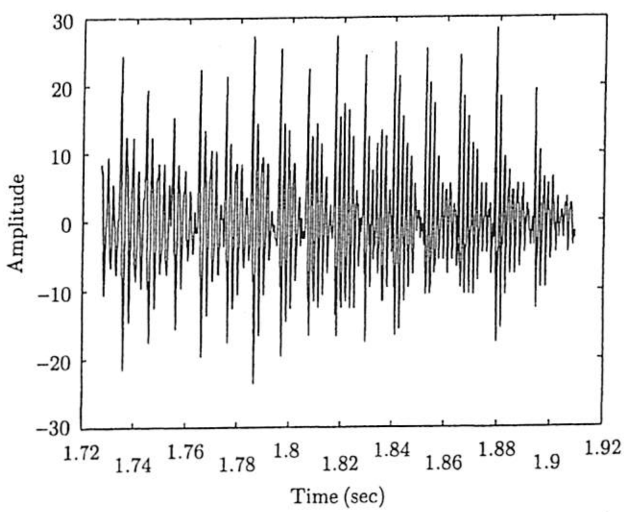
(b)

Figure 1.2. Binary phase shift keyed digital communication system (a) Basic system (b) BPSK baseband waveforms.

Communication signals



(a)



(b)

Figure 1.3. Speech waveforms for digits "zero" and "one" (a) "Zero" spoken three times (b) "Zero"–portion of utterance.

Speech signals

The detection problem is, if the signal of interest, 1:

$$x = \begin{cases} \text{either } 1+w, & \text{if signal occurs} \\ w, & \text{if signal does not occur} \end{cases}$$

where w is the noise of mean 0.

If one wants to make decision based on observation x, one may

$$x \begin{cases} > \frac{1}{2}, & \text{signal occurs} \\ < \frac{1}{2}, & \text{signal doesn't occur} \end{cases}$$

Hypothesis

$$\begin{cases} H_0 : & x = w \\ H_1 : & x = 1+w \end{cases}$$

We want to study the optimality!

If noise w is Gaussian, then the pdf are

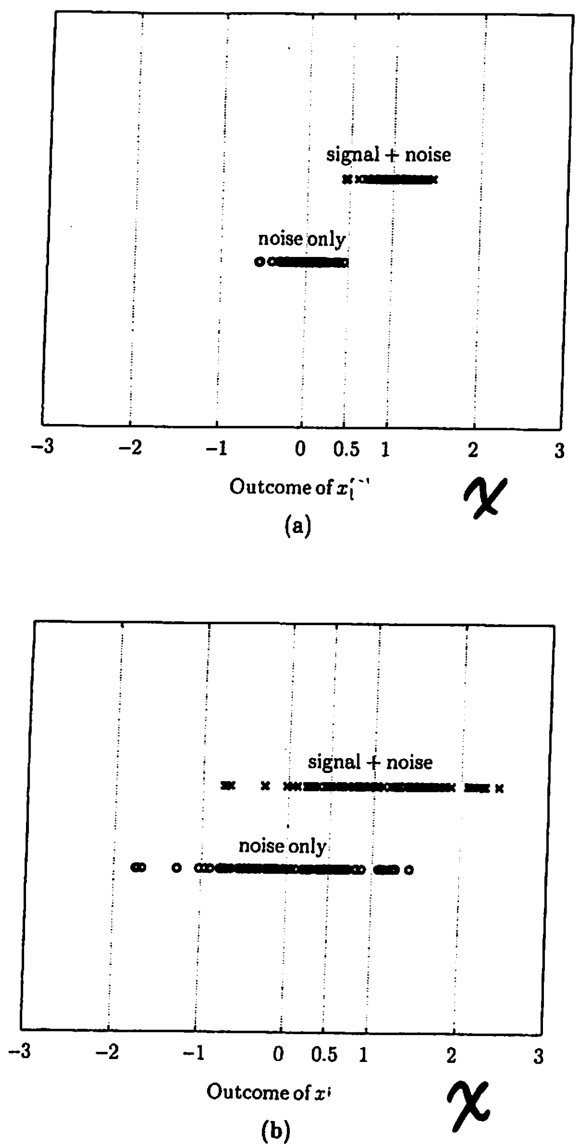
$$p(x; H_0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} x^2}$$

$$p(x; H_1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} (x-1)^2}$$

In general,

$$p(x; A) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} (x-A)^2}$$

where A is either 0 or 1.



noise variance

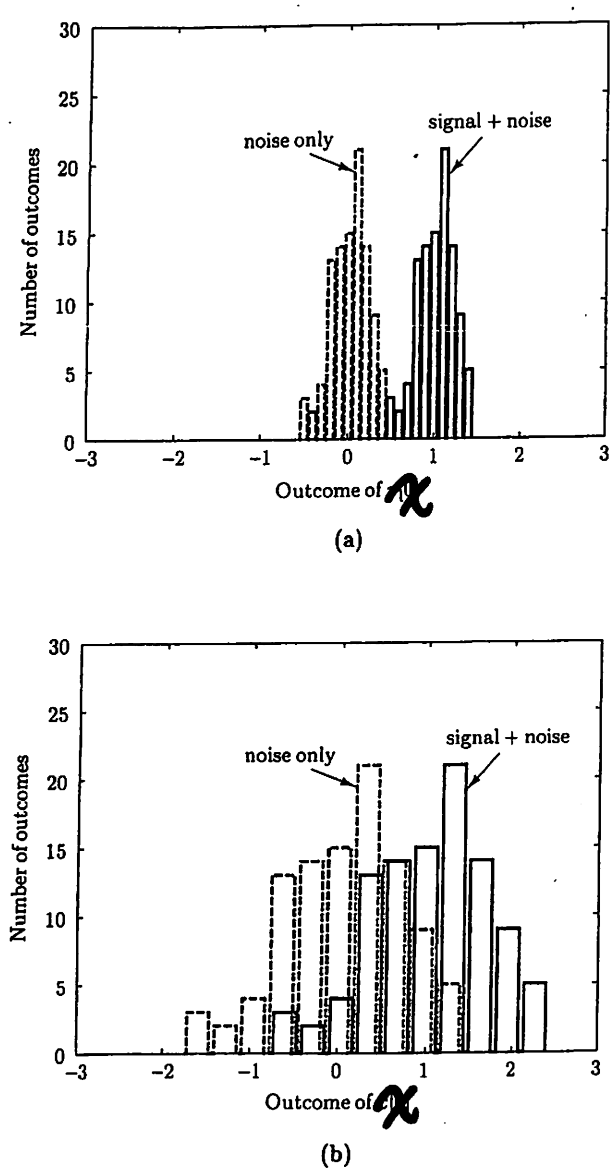
$\sigma^2 = 0.05$

noise variance

$\sigma^2 = 0.5$

Figure 1.4. Realizations of $x[0]$ for signal present and signal absent (a) $\sigma^2 = 0.05$ (b) $\sigma^2 = 0.5$.

x value

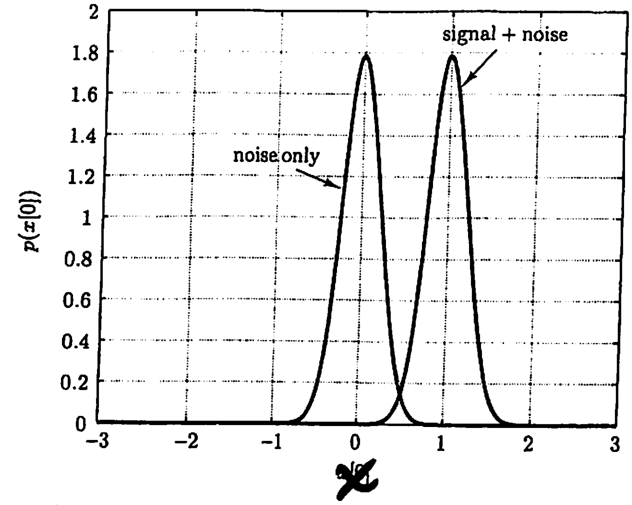


$\sigma^2 = 0.05$

$\sigma^2 = 0.5$

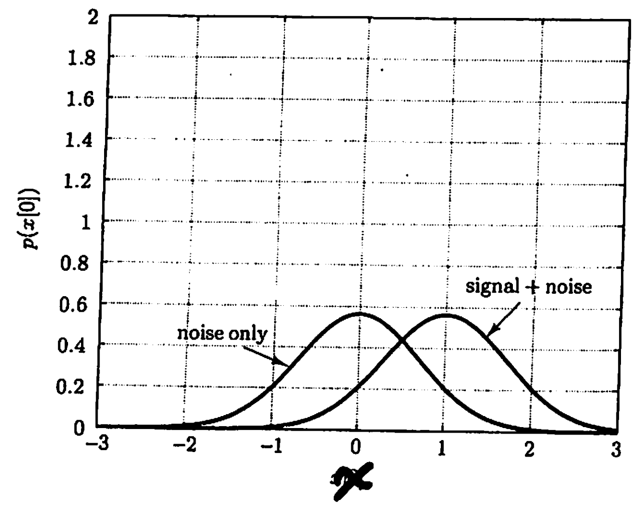
Figure 1.5. Histograms of x for signal present and signal absent (a) $\sigma^2 = 0.05$ (b) $\sigma^2 = 0.5$.

Histogram of x



(a)

$\sigma^2 = 0.05$



(b)

$\sigma^2 = 0.5$

Figure 1.6. PDFs of $\cancel{x[0]}$ x for signal present and signal absent (a) $\sigma^2 = 0.05$ (b) $\sigma^2 = 0.5$.

probability density function of x (pdf)

Signal Estimation Problem

mean estimation : $x = A + W$

to use multiple observations

$$x(n) = A + W(n)$$

$$\tilde{A} = \frac{1}{N}\sum_{n=1}^{N} x(n)$$

Signal estimation from noise

noisy signal

Estimated signal

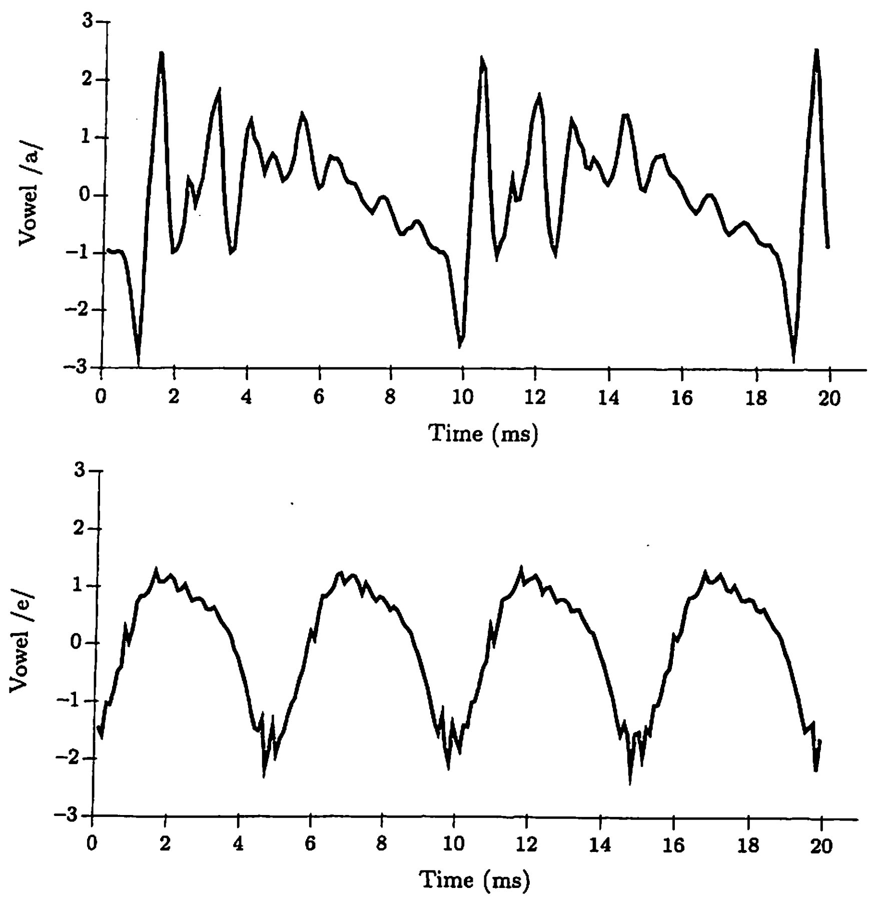


Figure 1.3 Examples of speech sounds

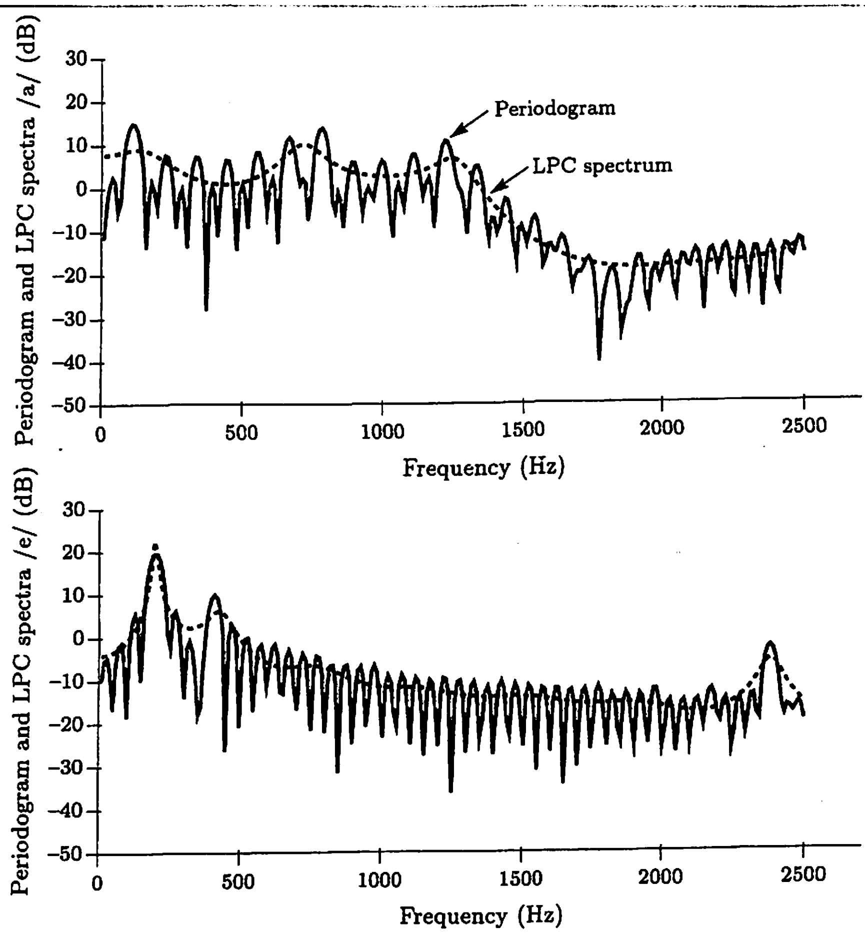


Figure 1.4 LPC spectral modeling

spectrum estimation

LPC spectrum

Linear predictive coding (LPC) is a method used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model. LPC is the most widely used method in speech coding and speech synthesis.

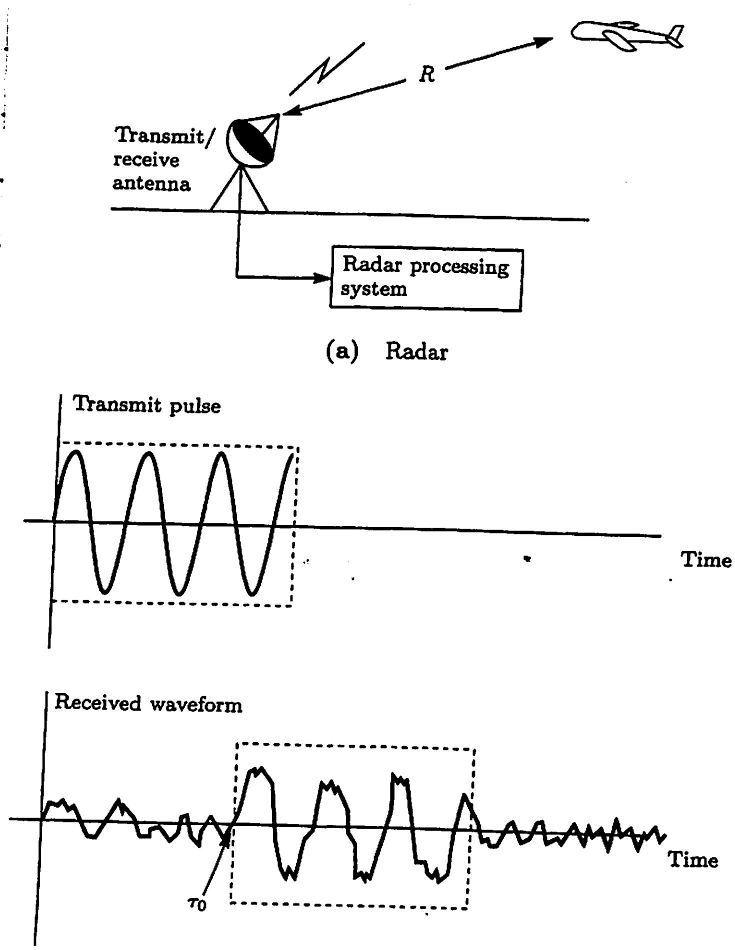
Parameter Estimation Problem

Radar range estimation (R)

τ_0 : time delay (time used for the wave propagation)

$\tau_0 = 2R/c$

c ← light speed



(a) Radar

(b) Transmit and received waveforms

Figure 1.1 Radar system

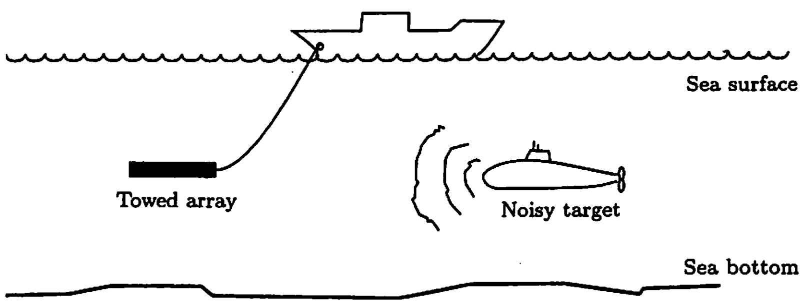
Target Positioning Using Sonar (submarine)

Measuring the time delays from an array of sensors, we determine the bearing β:

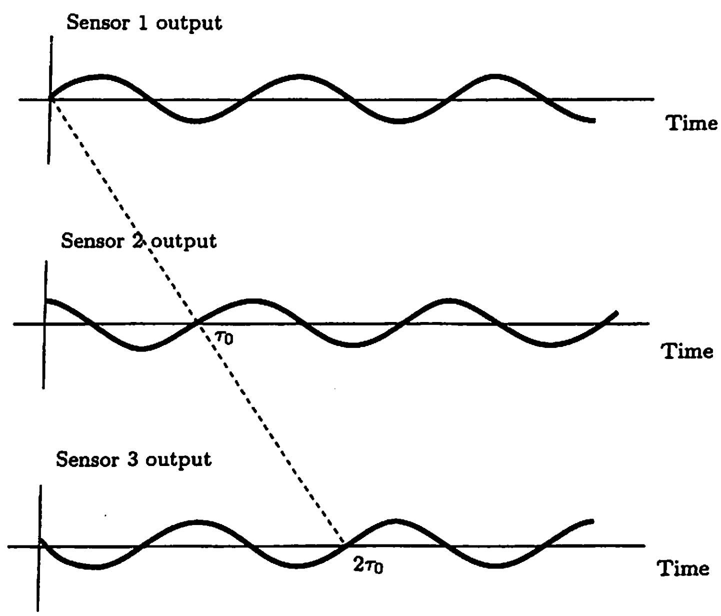
$$\beta = \arccos\left(\frac{c\,\tau_0}{d}\right)$$

β ← target direction

d ← distance between sensors



(a) Passive sonar



(b) Received signals at array sensors

Figure 1.2 Passive sonar system

§2. Probability Theory

1. Probability space

Γ: observation set

$\mathcal{G}$: a set of all the subsets of Γ
(Subset)

Called observation events
Mathematically, a σ-algebra
(closed to compliments, union, intersection)

$(\Gamma, \mathcal{G})$: observation space
probability is assigned on $\mathcal{G}$

Two cases for Γ: ① $\mathbb{R}^n$, n dimensional real space
(continuous case)
② $\{\gamma_1, \gamma_2, \cdots\}$,
(discrete case)

Events when $\Gamma = \mathbb{R}^n$:
$\{y = (y_1, \cdots, y_n)^T : a_1 \le y_1 \le b_1, \cdots, a_n \le y_n \le b_n)\}$
in this case $\mathcal{G}$ is the Borel set
(union, intersection, compliments
of intervals, regions)

Events when $\Gamma = \{\gamma_1, \gamma_2, \cdots\}$:
$\mathcal{G}$ includes any subset of Γ
denoted by 2^Γ called power set of Γ.

2. Probability

1) Discrete Case: probability mass function

$$p: \Gamma \rightarrow [0,1] \text{ with } \sum_i p(\gamma_i) = 1$$

Probability of event $A \in 2^{\Gamma}$ is

$$P(A) = \sum_{\gamma_i \in A} p(\gamma_i) \overset{\text{for simplicity}}{\underset{\Delta}{=\!=}} \int_A p(x)\,\mu(dx)$$

discrete measure

2) Continuous case: probability density function (PDF)

$$p: \mathbb{R}^n \rightarrow [0, \infty) \ \text{ s.t. } \int_{\mathbb{R}^n} p(\underline{x})\, d\underline{x} = 1$$

probability of event $A \subset \mathbb{R}^n$ is

$$P(A) = \int_A p(\underline{x})\, d\underline{x}$$

3) A random variable is a function $g: \Gamma \rightarrow \mathbb{R}$ acting on a random observation Y
$g(Y)$

Expected value of $g(Y)$

$$E(g(Y)) = \begin{cases} \sum_{i=1}^{\infty} g(\gamma_i)\, p(\gamma_i), & \text{discrete} \\ \int_{\mathbb{R}^n} g(y)\, p(y)\, dy, & \text{continuous} \end{cases}$$

$$\overset{\text{or}}{=} \int_{\Gamma} g(y)\,p(y)\,\mu(dy) = \int_{\Gamma} g\,p\,d\mu$$

Upper case letters Y denote random quantities

Lower case letters y denote specific values (realizations) of the random quantity Y, i.e., $Y = y$, for example.

3. Some Important PDFs

1) Gaussian (Normal) : $\mathcal{N}(\mu, \sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

mean: μ, variance: σ^2

when $\mu = 0$, its moments

$$E(x^n) = \begin{cases} 1\cdot 3\cdot 5 \cdots (n-1)\,\sigma^n & \text{for even } n \\ 0, & \text{for odd } n \end{cases}$$

When $\mu = 0$, $\sigma^2 = 1$, it is called standard normal

Right-tail probability (Q-function)

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}t^2\right\} dt$$

$$\approx \frac{1}{\sqrt{2\pi}\,x} \exp\left\{-\frac{1}{2}x^2\right\} \quad \text{for not small } x$$

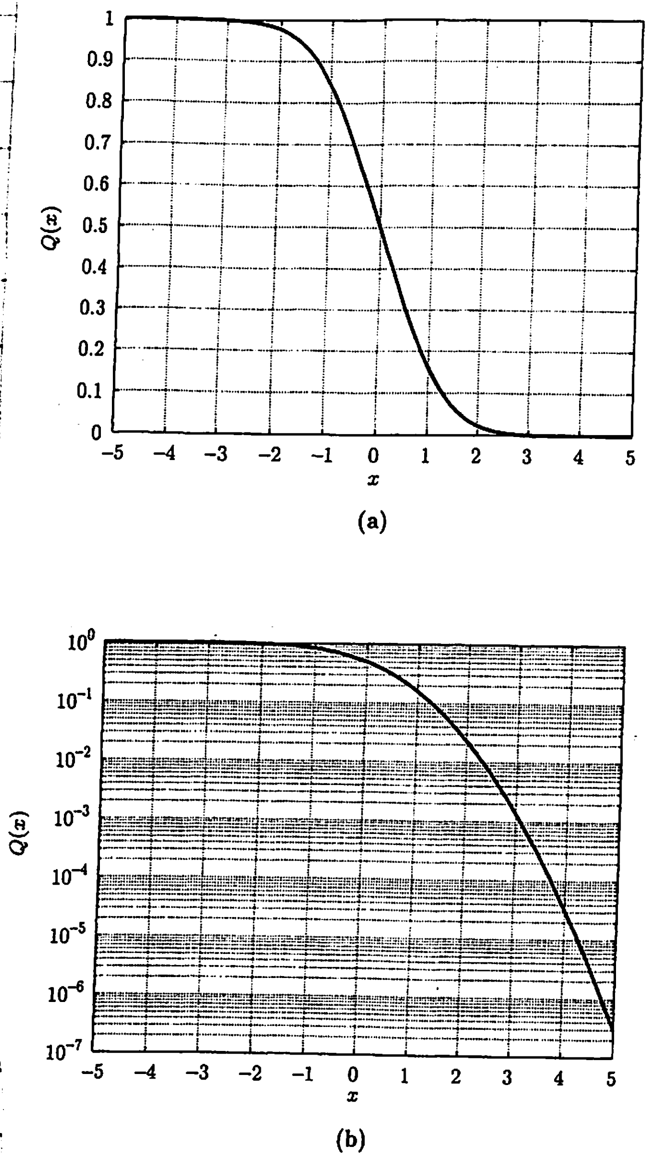


Figure 2.1. Right-tail probability for standard normal PDF (a) Linear vertical axis (b) Logarithmic vertical axis.

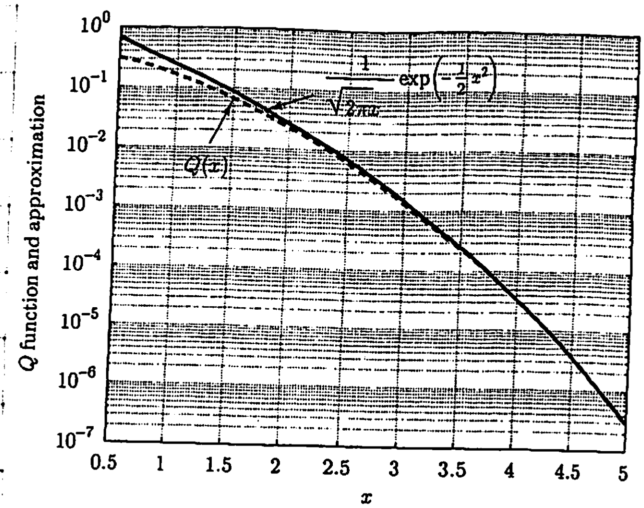
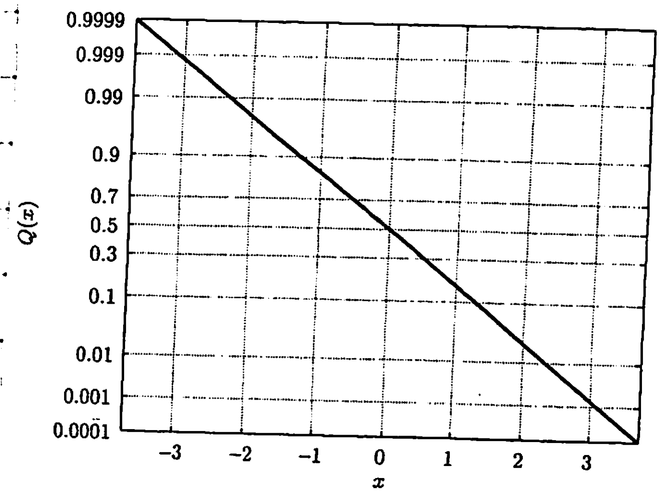


Figure 2.2. Approximation to Q function.



ure 2.3. Q function plotted on normal probability paper.

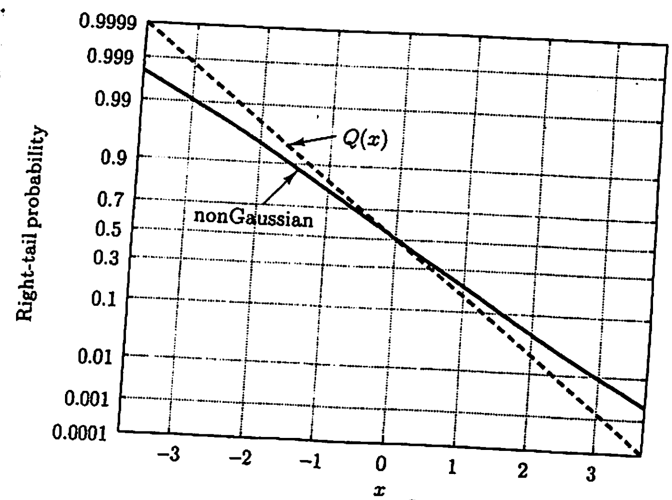


Figure 2.4. Right-tail probability for nongaussian PDF on normal probability paper.

2) Multivariate Gaussian : $x \in \mathbb{R}^n$

$$\mathcal{N}(\underline{\mu}, \underline{C})$$

$$p(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det^{\frac{1}{2}}(\underline{C})} \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \underline{C}^{-1}(\underline{x}-\underline{\mu})\right\}$$

where $\underline{\mu}$ is the mean vector, $\underline{C}$ is the covariance matrix such that $\underline{C}$ is positive definite and $\underline{C}^{-1}$ exists

$$\mu_i = [\underline{\mu}]_i = E(x_i)$$

$$[C]_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$$

$$\text{or } \underline{C} = E[(\underline{x} - E(\underline{x}))(\underline{x} - E(\underline{x}))^T]$$

3) Chi-Squared (Central) : χ^2_ν

with ν degrees of freedom

$x = \sum_{i=1}^{\nu} x_i^2$ where $x_1, \ldots, x_\nu$ are i.i.d. normal $\mathcal{N}(0,1)$

$$p(x) = \begin{cases} \dfrac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} \exp\{-\frac{1}{2}x\}, & x > 0 \\ 0, & x < 0 \end{cases}$$

where $\Gamma(u) = \int_0^\infty t^{u-1} \exp\{-t\}\, dt$

(Γ - function) $= (u-1)\Gamma(u-1)$

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}, \quad \Gamma(n) = (n-1)!$$

$$E(x) = \nu, \quad Var(x) = 2\nu$$

The right-tail probability

Even ν

$$Q_{\chi_\nu^2}(x) = \int_x^\infty p(t)\,dt = \exp\{-\tfrac{1}{2}x\} \sum_{k=0}^{\frac{\nu}{2}-1} \frac{\left(\frac{x}{2}\right)^k}{k!}, \quad \nu \geq 2$$

odd ν

$$Q_{\chi_\nu^2}(x) = \begin{cases} 2Q(\sqrt{x}), & \nu = 1 \\ 2Q(\sqrt{x}) + \dfrac{\exp\{-\frac{1}{2}x\}}{\sqrt{\pi}} \displaystyle\sum_{k=1}^{\frac{\nu-1}{2}} \frac{(k-1)!\,(2x)^{k-\frac{1}{2}}}{(2k-1)!}, & \nu \geq 3 \end{cases}$$

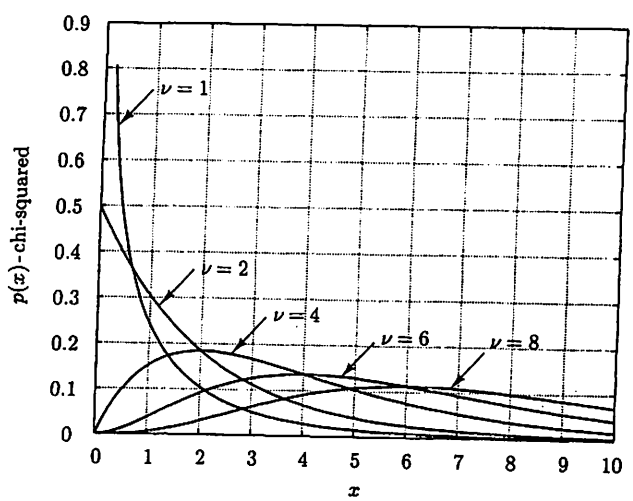


Figure 2.5. PDF for chi-squared random variable.

4) Chi-Squared (Noncentral): $\chi_\nu'^2(\lambda)$

with ν degrees of freedom

$$x = \sum_{i=1}^{\nu} x_i^2 \quad \text{where } x_i \sim N(\mu_i, 1)$$

$\{x_i\}$ are independent

non centrality parameter $\lambda = \sum_{i=1}^{\nu} \mu_i^2$

$$p(x) = \begin{cases} \frac{1}{2}\left(\frac{x}{\lambda}\right)^{\frac{\nu-2}{4}} \exp\{-\frac{1}{2}(x+\lambda)\} I_{\frac{\nu}{2}-1}(\sqrt{\lambda x}), & x>0 \\ 0 & , x<0 \end{cases}$$

$I_r(u)$ is the modified Bessel function of the first kind and order r.

$$I_r(u) = \frac{(\frac{1}{2}u)^r}{\sqrt{\pi}\,\Gamma(r+\frac{1}{2})} \int_0^{\pi} \exp(u\cos\theta) \sin^{2r}\theta \, d\theta$$

$$= \sum_{k=0}^{\infty} \frac{(\frac{1}{2}u)^{2k+r}}{k!\,\Gamma(r+k+1)}$$

$$\Rightarrow \quad p(x) = \frac{x^{\frac{\nu}{2}-1}\exp\{-\frac{1}{2}(x+\lambda)\}}{2^{\frac{\nu}{2}}} \sum_{k=0}^{\infty} \frac{\left(\frac{\lambda x}{4}\right)^k}{k!\,\Gamma(\frac{\nu}{2}+k)}$$

$$E(x) = \nu + \lambda$$

$$Var(x) = 2\nu + 4\lambda$$

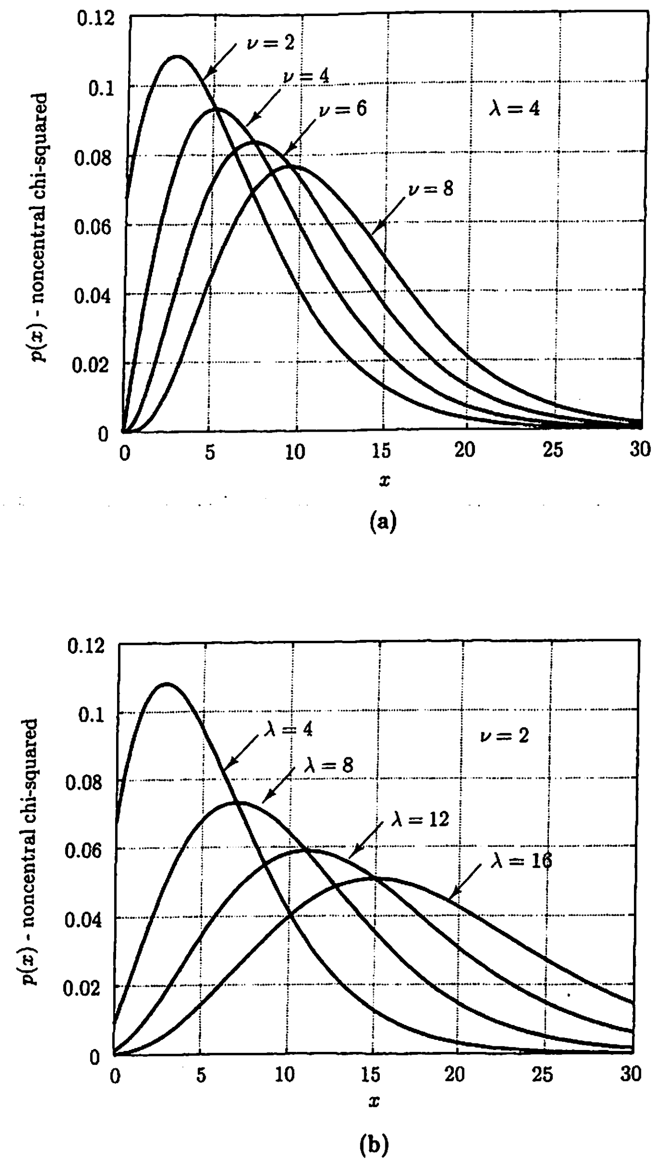


Figure 2.6. PDF for noncentral chi-squared random variable (a) Varying degrees of freedom (b) Varying noncentrality parameter.

5) F(central) : F_{ν_1, ν_2}

The ratio of two independent χ^2 random variables :

$$x = \frac{x_1/\nu_1}{x_2/\nu_2}$$

$x_1 \sim \chi^2_{\nu_1}$, $x_2 \sim \chi^2_{\nu_2}$, x_1 and x_2 are independent

$$p(x) = \begin{cases} \dfrac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \dfrac{x^{\frac{\nu_1}{2}-1}}{\left(1+\frac{\nu_1}{\nu_2}x\right)^{\frac{\nu_1+\nu_2}{2}}}, & x > 0 \\ 0, & x < 0 \end{cases}$$

where $B(u, v)$ is the Beta function

$$B(u, v) = \frac{\Gamma(u)\,\Gamma(v)}{\Gamma(u+v)}$$

$$E(x) = \frac{\nu_2}{\nu_2 - 2} \quad , \nu_2 > 2$$

$$\mathrm{Var}(x) = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)} \quad , \nu_2 > 4$$

as $\nu_2 \to \infty$, $x \to \frac{x_1}{\nu_1} \sim \frac{\chi^2_{\nu_1}}{\nu_1}$

since $\frac{x_2}{\nu_2} \to 1$

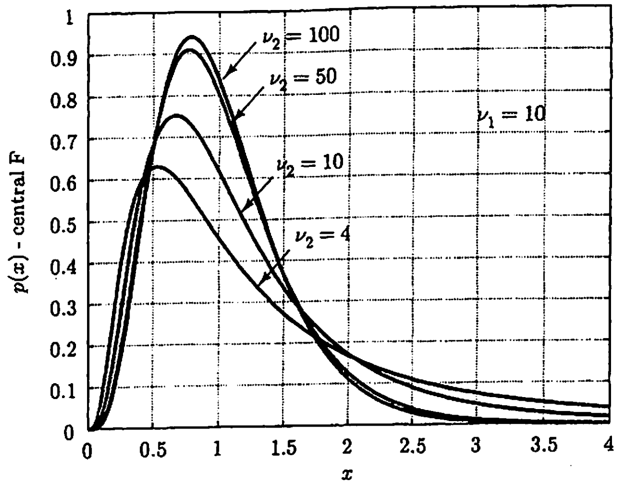


Figure 2.7. PDF for F random variable.

6) F (non central) : $F'_{\nu_1,\nu_2}(\lambda)$

$$x = \frac{\chi_1/\nu_1}{\chi_2/\nu_2}$$

$\chi_1 \sim \chi'^2_{\nu_1}(\lambda)$, $\chi_2 \sim \chi^2_{\nu_2}$.

χ_1 and χ_2 are independent

$$p(x) = \exp\left(-\frac{\lambda}{2}\right) \sum_{k=0}^{\infty} \frac{\left(\frac{\lambda}{2}\right)^k \left(\frac{\nu_1}{\nu_2}\right)^{\frac{1}{2}\nu_1 + k}}{k! \; B\left(\frac{\nu_1+2k}{2}, \frac{\nu_2}{2}\right)}$$

$$\cdot\, x^{\frac{\nu_1}{2}+k-1} \left(1 + \frac{\nu_1}{\nu_2} x\right)^{-\frac{1}{2}(\nu_1+\nu_2)-k}$$

$$E(x) = \frac{\nu_2(\nu_1+\lambda)}{\nu_1(\nu_2-2)}, \quad \nu_2 > 2$$

$$\mathrm{Var}(x) = 2\left(\frac{\nu_2}{\nu_1}\right)^2 \frac{(\nu_1+\lambda)^2 + (\nu_1+2\lambda)(\nu_2-2)}{(\nu_2-2)^2\,(\nu_2-4)}, \quad \nu_2 > 4$$

7) Rayleigh

$x = \sqrt{x_1^2 + x_2^2}$

$x_1 \sim \mathcal{N}(0, \sigma^2), \quad x_2 \sim \mathcal{N}(0, \sigma^2)$

x_1 and x_2 are independent

$$p(x) = \begin{cases} \frac{x}{\sigma^2} \exp\left\{-\frac{1}{2\sigma^2} x^2\right\}, & x > 0 \\ 0, & x < 0 \end{cases}$$

$E(x) = \sqrt{\frac{\pi \sigma^2}{2}}, \quad \text{var}(x) = (2 - \frac{\pi}{2})\sigma^2$

Right-tail prob. $\int_0^\infty p(x)dx = \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$

$P_r(x > \sqrt{\gamma'}) = Q_{\chi_2^2}\left(\frac{\gamma'}{\sigma^2}\right)$

$\Rightarrow \quad Q_{\chi_2^2}(x) = \exp\left\{\frac{x}{2}\right\}$

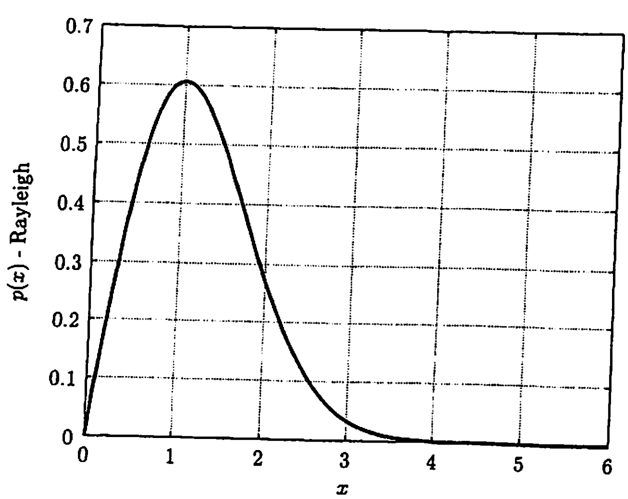


Figure 2.8. PDF for Rayleigh random variable ($\sigma^2 = 1$).

8) Rician

$$x = \sqrt{x_1^2 + x_2^2}$$

$$x_1 \sim N(\mu_1, \sigma^2), \quad x_2 \sim N(\mu_2, \sigma^2)$$

x_1 and x_2 are independent

$$p(x) = \begin{cases} \frac{x}{\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}(x^2+\alpha^2)\right\} I_0\left(\frac{\alpha x}{\sigma^2}\right), & x > 0 \\ 0, & x < 0 \end{cases}$$

$$I_0(u) = \frac{1}{\pi}\int_0^{\pi} \exp(u\cos\theta)\,d\theta$$
$$= \int_0^{2\pi} \exp(u\cos\theta)\frac{d\theta}{2\pi}$$

$$\Pr(x > \gamma) = Q_{\chi_2'^2(\lambda)}\left(\frac{\gamma^2}{\sigma^2}\right), \quad \lambda = \frac{\mu_1^2 + \mu_2^2}{\sigma^2}$$

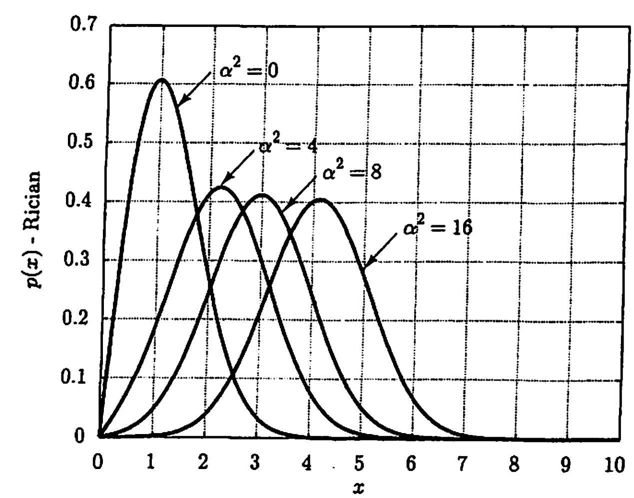


Figure 2.9. PDF for Rician random variable ($\sigma^2 = 1$).

$$\alpha^2 = \mu_1^2 + \mu_2^2$$

9) Quadratic Forms of Gaussian Random Variables

$$y = \underline{x}^T A \underline{x}$$

where A is a symmetric $n \times n$ matrix

$\underline{x}$ is $n \times 1$ Gaussian vector

$\sim \mathcal{N}(\underline{\mu}, \underline{C})$

1. If $A = \underline{C}^{-1}$ and $\underline{\mu} = \underline{0}$, $\underbrace{\underline{x}^T \underline{C}^{-1} \underline{x}}_{y} \sim \chi_n^2$

2. If $A = \underline{C}^{-1}$, $\underline{\mu} \neq 0$, $\underline{x}^T \underline{C}^{-1} \underline{x} \sim \chi_n'^2(\lambda)$

$$\lambda = \underline{\mu}^T \underline{C}^{-1} \underline{\mu}$$

3. If $A^2 = A$ (idempotent), $\underline{C} = I$, $\underline{\mu} = 0$,

$$\underline{x}^T A \underline{x} \sim \chi_r^2$$

r is the rank of A.

Chapter 2. Elements of Hypothesis Testing

In many applications, one needs to decide one from a few choices, such as 0 or 1, in communication systems. Since the data set may be huge, it is critically important to make the decision statistically optimal. Such a task is the goal of this Chapter.

§1. Bayesian Hypothesis Testing

* H_0 and H_1 are two possible hypotheses corresponding to two possible probability distributions P_0 and P_1, respectively, on the observation space $(\Gamma, \mathcal{G})$:

$$\left.\begin{aligned} H_0 &: Y \sim P_0 \\ \text{vs. } H_1 &: Y \sim P_1 \end{aligned}\right\} \text{(II.B.1)},$$

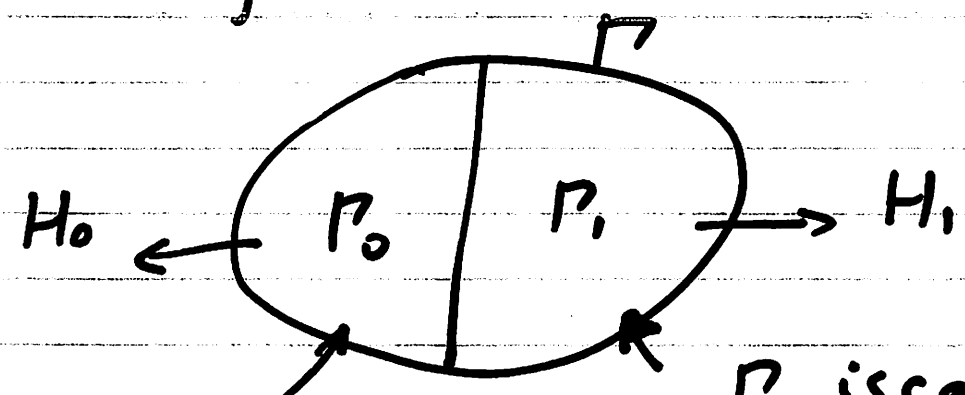
H_0: null hypothesis

H_1: alternative hypothesis

* A decision rule (or hypothesis test) δ for H_0 vs. H_1 is a partition of the observation set Γ into sets $\Gamma_1 \in \mathcal{G}$ and $\Gamma_0 = \Gamma_1^c = \Gamma \setminus \Gamma_1$ such that

  Choose H_j when $y \in \Gamma_j$ for $j = 0, 1$,

  where y is an observed value of Y.



Γ_0 is called acceptance region

Γ_1 is called the rejection region (or critical region)

* The decision rule δ can be thought of as a function on Γ:

$$\delta(y) = \begin{cases} 1 & \text{if } y \in \Gamma_1 \\ 0 & \text{if } y \in \Gamma_1^c . \end{cases}$$

The value of δ for a given $y \in \Gamma$ is the index of the hypothesis accepted by δ.

* It is clear that good or bad the decision is completely depending on the partition Γ_1 and $\Gamma_1^c = \Gamma_0$. The goal of this chapter is to do this partition in "some optimal" way.

To do so, we first need to have the concept what is called "good" or "bad" of a decision: costs:

How much does it cost if this decision is made?

Note: a partition Γ_1, Γ_1^c
$\Longleftrightarrow$ a decision rule δ

* Costs are non-negative numbers C_{ij} for $i,j = 0, 1$, and C_{ij} is the cost incurred by choosing hypothesis H_i when hypothesis H_j is true.

* Conditional risk for each hypothesis is defined as the average (or expected) cost incurred by decision rule δ when that hypothesis is true: The conditional risk for hypothesis j is

$$R_j(\delta) = C_{1j} P_j(\Gamma_1) + C_{0j} P_j(\Gamma_0), \quad j = 0, 1,$$

where P_0 and P_1 are the two probability distributions associated with the hypotheses H_0 & H_1, respectively.

The meaning of $R_j(\delta)$: the cost of choosing H_1 when H_j is true times the probability of doing so, plus the cost of choosing H_0 when H_j is true times the probability of doing it.

* Notice that we want to have a kind of optimal decision rule δ (or equivalently partition Γ_i). To do so, we need to have a single objective function to optimize.

In applications, the two hypotheses H_0 and H_1 may occur in some probabilities.

* Assume π_0 and $\pi_1 = 1 - \pi_0$ are the probabilities of the occurences of H_0 and H_1, respectively. (unconditioned on the value of an observation Y)

These probabilities π_0 & π_1 are known in prior as a priori probabilities of the two hypotheses.

Example: in a communication system to transmit binary digits 0 and 1 sequence,

$\pi_0 = P(0)$, $\pi_1 = P(1)$

are the probabilities of the occurences of 0 and 1 in the sequence to transmit, which are determined by a source.

* **Bayes risk** (an average risk) is defined as the overall average cost incurred by decision rule δ:

$$r(\delta) = \pi_0 R_0(\delta) + \pi_1 R_1(\delta),$$

which gives a unique (single) objective function to optimize for choosing the decision rule δ.

* **Bayes rule** for H_0 vs. H_1 is the optimum decision rule for H_0 vs. H_1 that minimizes $r(\delta)$ over all decision rules : $\arg\min_{\delta} r(\delta)$

This clearly depends on the costs C_{ij} defined earlier.

We next want to see what a Bayes rule is (likelihood ratio test).

* Let p_j be the densities of the distributions P_j, $j=0,1$, respectively.

$$r(\delta) = \sum_{j=0}^{1} \pi_j [C_{0j}(1-P_j(\Gamma_1)) + C_{1j} P_j(\Gamma_1)]$$

$$= \sum_{j=0}^{1} \pi_j C_{0j} + \sum_{j=0}^{1} \pi_j (C_{1j} - C_{0j}) P_j(\Gamma_1)$$

$$= \sum_{j=0}^{1} \pi_j C_{0j} + \sum_{j=0}^{1} \pi_j (C_{1j} - C_{0j}) \int_{\Gamma_1} p_j(y) \mu(dy)$$

$$= \underbrace{\sum_{j=0}^{1} \pi_j C_{0j}}_{\text{non negative}} + \underbrace{\int_{\Gamma_1} \left[\sum_{j=0}^{1} \pi_j (C_{1j} - C_{0j}) p_j(y) \right] \mu(dy)}, \quad (*)$$

non negative

The more negative it is, the smaller $r(\delta)$ is

$\Downarrow$

The most negative this will be if its integrant function is always negative on Γ_1, i.e.

$$\sum_{j=0}^{1} \pi_j (C_{1j} - C_{0j}) p_j(y) \le 0 \quad \forall y \in \Gamma_1$$

$\Downarrow$

$r(\delta)$ becomes smallest

Thus, $r(\delta)$ is a minimum over all Γ_1 (or over all δ)

if we choose

$$\Gamma_1 = \left\{ y \in \Gamma \,\middle|\, \sum_{j=0}^{1} \pi_j (C_{1j} - C_{0j}) p_j(y) \le 0 \right\}$$

$$= \left\{ y \in \Gamma \,\middle|\, \pi_1 (C_{11} - C_{01}) p_1(y) \le \pi_0 (C_{00} - C_{10}) p_0(y) \right\}$$

It is nature to assume that $C_{11} < C_{01}$, i.e., the cost of correctly choosing H_1 is less than the cost of incorrectly rejecting H_1.

Then,

$$\Gamma_1 = \{ y \in \Gamma \mid p_1(y) \geq \tau\, p_0(y) \}, \quad (\text{II.B.7})$$

where

$$\tau \triangleq \frac{\pi_0 (C_{10} - C_{00})}{\pi_1 (C_{01} - C_{11})}, \quad (\text{II.B.9})$$

This can be determined in a priori

The decision rule described by the above rejection region is called a likelihood-ratio test (or probability ratio test). Very important

$$\Gamma_1 = \{ y \in \Gamma \mid p_1(y) \geq \tau\, p_0(y) \}$$
$$= \{ y \in \Gamma \mid \frac{p_1(y)}{p_0(y)} \geq \tau \}. \quad \left(\begin{array}{l} \text{when } p_0(y) = 0, \\ k/0 = \infty \text{ for} \\ \text{any } k \geq 0 \end{array} \right)$$

$\Rightarrow$ The quantity $L(y) \triangleq \dfrac{p_1(y)}{p_0(y)}, \; y \in \Gamma$

is called the likelihood ratio (or the likelihood-ratio statistic) between H_0 and H_1.

Note: The boundary region $\{ y \in \Gamma \mid p_1(y) = \tau\, p_0(y) \}$ does not contribute to $r(\delta)$, since the integrant in (*) is 0 on the boundary.

$\Rightarrow$ The Bayes decision rule computes the likelihood ratio for the observed value y of Y and then makes its decision by comparing this ratio with the threshold τ:

$$\delta_B(y) = \begin{cases} 1, & \text{if } L(y) \geq \tau \\ 0, & \text{if } L(y) < \tau \end{cases}$$

B stands for Bayes

* Minimum-Probability-of-Error Decision

A commonly used cost is the uniform cost assignment:

$$C_{ij} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \end{cases} \qquad (\text{II.B.13})$$

With this cost,

$$r(\delta) = \pi_0 P_0(\Gamma_1) + \pi_1 P_1(\Gamma_0)$$

Note $P_i(\Gamma_j)$ is the probability of choosing H_j when H_i is true.

$\Rightarrow$ $P_i(\Gamma_j)$ when $i \neq j$ is the conditional probability of making an error given that H_i is true.

$\Rightarrow$ $r(\delta)$ is the average probability of error incurred by the decision rule δ.

Since the likelihood-ratio test with $\tau = \frac{\pi_0}{\pi_1}$ minimizes $r(\delta)$, the Bayes decision rule is also a minimum-probability-of-error decision scheme

* Using Bayes' formula

$$\pi_j(y) \triangleq P(H_j \text{ true} \mid Y = y) = \frac{p_j(y)\,\pi_j}{p(y)},$$

where $p(y)$ is the average or overall density of Y:

$$p(y) = \pi_0\, p_0(y) + \pi_1\, p_1(y)$$

The probabilities $\pi_0(y)$ and $\pi_1(y)$ are called the posterior or a posteriori probabilities of the two hypotheses.

Going back to (II.B.7), the critical (rejection) region of the Bayes rule can be rewritten

$$\Gamma_1 = \{y \in \Gamma \mid C_{10}\pi_0(y) + C_{11}\pi_1(y) \le C_{00}\pi_0(y) + C_{01}\pi_1(y)\}$$

based on the posterior probabilities.

* $C_{i0}\pi_0(y) + C_{i1}\pi_1(y)$ is the average cost incurred by choosing hypothesis H_i given $Y = y$.

⇒ The Bayes rule makes its decision by choosing the hypothesis that yields the minimum posterior cost

When the uniform cost (II.B.13) is used, the Bayes rule is

$$\delta_B(y) = \begin{cases} 1, & \text{if } \pi_1(y) \geq \pi_0(y) \\ 0, & \text{if } \pi_1(y) < \pi_0(y) \end{cases}$$

⇒ The minimum-probability-of-error decision rule chooses the hypothesis that has the maximum a posteriori probability of having occured given $Y = y$.

⇒ It is called the MAP decision rule for the binary hypothesis test (II.B.1)

⇒ MAP is the Bayes rule when the cost is uniform.

★ Example I : The Binary Channel

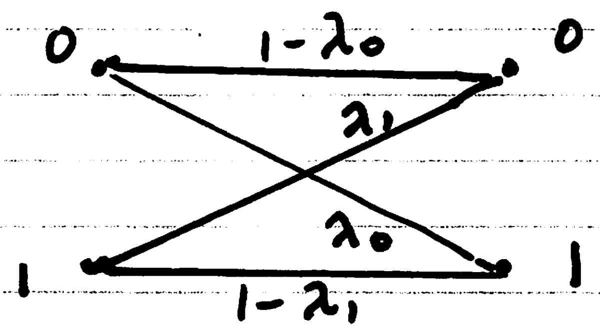
Suppose a binary digit 0 or 1, is transmitted over a communication channel.

Observation Y is the output of the channel, either 0 or 1.

The channel :

✓ transmitted 0 is received as 1 with probability λ_0 and as 0 . . . . . $1-\lambda_0$, $0<\lambda_0<1$.

✓ transmitted 1 is received as 0 with probability λ_1 . . . 1 — . . — — $1-\lambda_1$, $0<\lambda_1<1$.



The observation Y does not tell exactly whether the transmitted digit 0 or 1. We need to find an optimal way to decide what was transmitted.

This problem can be modeled as a binary hypothesis testing problem, where the hypothesis

H_j means that a "j" was transmitted, $j=0,1$,

H_0 : 0 was transmitted,
H_1 : 1 was transmitted.

The observation set $\Gamma = \{0, 1\}$

The observation Y has densities (probability mass function)

$$p_j(y) = \begin{cases} \lambda_j & \text{if } y \neq j \\ 1-\lambda_j & \text{if } y = j \end{cases}, \quad j=0,1.$$

The likelihood ratio:

$$L(y) = \frac{p_1(y)}{p_0(y)} = \begin{cases} \dfrac{\lambda_1}{1-\lambda_0} & \text{if } y=0 \\ \dfrac{1-\lambda_1}{\lambda_0} & \text{if } y=1 \end{cases}$$

For a Bayes test, the test threshold τ is determined by the cost and prior probabilities (II.B.9).

Received 0 $\longrightarrow$ transmitted 1, if $\lambda_1 \geq \tau(1-\lambda_0)$
" 0, otherwise

Received 1 $\longrightarrow$ transmitted 1, if $1-\lambda_1 \geq \tau\lambda_0$
" 0, otherwise

A special case when the cost is uniform and equal priors ($\pi_0 = \pi_1 = \frac{1}{2}$, equal probable signals)

$\Rightarrow$ the threshold $\tau = 1$

$\Rightarrow$ The Bayes rule is

$$\delta_B(0) = \begin{cases} 1, & \text{if } \lambda_1 \geq 1-\lambda_0 \\ 0, & \text{if } \lambda_1 < 1-\lambda_0 \end{cases}$$

$$\delta_B(1) = \begin{cases} 1, & \text{if } 1-\lambda_1 \geq \lambda_0 \\ 0, & \text{if } 1-\lambda_1 < \lambda_0 \end{cases}$$

Note: the boundary points $L(y) = \tau$ can be assigned to either Γ_1 or Γ_0

overall $\Rightarrow$
$$\delta_B(y) = \begin{cases} y & \text{if } 1-\lambda_1 \geq \lambda_0 \\ 1-y & \text{if } 1-\lambda_1 < \lambda_0. \end{cases}$$

When the channel is symmetric, i.e., $\lambda_1 = \lambda_0 = \lambda$, then

$$\delta_B(y) = \begin{cases} y, & \text{if } \lambda \leq \frac{1}{2} \\ 1-y, & \text{if } \lambda > \frac{1}{2} \end{cases}$$

<u>Physical meaning</u>: If the channel is more likely than not to converts bits ($\lambda > \frac{1}{2}$), we make our decision by flipping the received bit; Otherwise, we accept the received bit as it is.

$\Rightarrow r(\delta_B) = \min\{\lambda, 1-\lambda\}$

$\Rightarrow$ The performance improves as the channel becomes more reliable in either transmitting the bit directly or in flipping the transmitted bit

In this case, when $\lambda = \frac{1}{2}$ (50% wrong), the observation is worthless.

* Example II: Location Testing with Gaussian Error

Consider the following two hypotheses concerning a real-valued observation Y:

$$H_0 : Y = \varepsilon + \mu_0$$
vs.
$$H_1 : Y = \varepsilon + \mu_1$$
where ε is a Gaussian random variable with mean 0 and variance σ^2, and μ_0 and μ_1 are two fixed real numbers with $\mu_1 > \mu_0$.

μ_0, μ_1 only change the mean value of the observation. We are testing about which of two possible values or "locations" the observation is distributed.

Going back to the original hypothesis testing model (II. B.1):

$$H_0 : Y \sim \mathcal{N}(\mu_0, \sigma^2)$$

vs.

$$H_1 : Y \sim \mathcal{N}(\mu_1, \sigma^2)$$

The likelihood ratio is

$$L(y) = \frac{p_1(y)}{p_0(y)} = \frac{\frac{1}{\sqrt{2\pi}\,\sigma} e^{-(y-\mu_1)^2/2\sigma^2}}{\frac{1}{\sqrt{2\pi}\,\sigma} e^{-(y-\mu_0)^2/2\sigma^2}}$$

$$= \exp\left\{\frac{\mu_1-\mu_0}{\sigma^2}\left(y-\frac{\mu_0+\mu_1}{2}\right)\right\}$$

$\Rightarrow$ A Bayes test (rule) is

$$\delta_B(y) = \begin{cases} 1, & \text{if } \exp\left\{\frac{\mu_1-\mu_0}{\sigma^2}\left(y-\frac{\mu_0+\mu_1}{2}\right)\right\} \geq \tau, \\ 0, & \text{otherwise} \end{cases}$$

where τ is the threshold.

Due to $\mu_1 > \mu_0$, $L(y)$ is a strictly increasing function of y.

$\Rightarrow$ Comparing $L(y)$ to the threshold τ
$\Leftrightarrow$ comparing y itself to another threshold
$\tau' = L^{-1}(\tau)$

$$\Rightarrow \quad \delta_B(y) = \begin{cases} 1, & \text{if } y \geq \tau' \\ 0, & \text{if } y < \tau', \end{cases}$$

where $\tau' = \dfrac{\sigma^2}{\mu_1 - \mu_0} \log(\tau) + \dfrac{\mu_0 + \mu_1}{2}$

In particular when the uniform cost and equal priors ($\pi_0 = \pi_1 = \frac{1}{2}$) are used, we have
$$\tau = 1 \quad \text{and} \quad \tau' = \frac{\mu_0 + \mu_1}{2}$$
In this case, the Bayes rule compares the observation to the average of μ_0 and μ_1

$$H_1 \quad \text{if } y \geq \frac{\mu_0 + \mu_1}{2}$$
$$H_0 \quad \text{if } y < \frac{\mu_0 + \mu_1}{2}$$

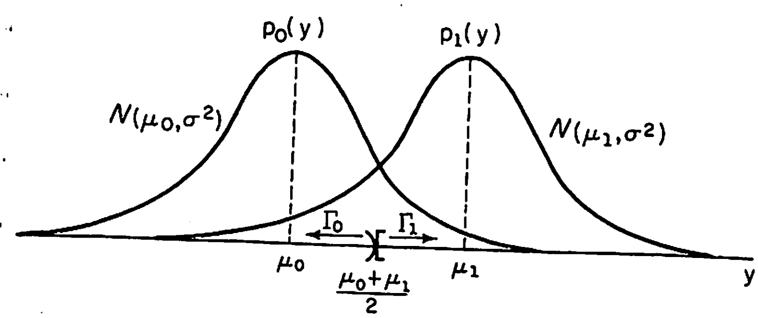


FIGURE II.B.2. Illustration of location testing with Gaussian errors, uniform costs, and equal priors.