Conditional CDF and pdf of X

- Given that event A (with P(A) > 0) has occurred, the conditional CDF of **X** is

$$F_{\mathbf{X}|A}(u|A) = P\{\mathbf{X} \le u|A\}$$

- $P\{\mathbf{X} \le u|A\} = P(\{\mathbf{X} \le u\} \cap A)/P(A)$ is the conditional probability that $\{\mathbf{X} \le u\}$ given that the event A occurred
- If **X** is a continuous random variable, then $f_{\mathbf{X}|A}(u|A)$, the conditional pdf of **X** is the derivative of the conditional CDF $F_{\mathbf{X}|A}(u|A)$

Just as good as the regular kind …

- Conditional CDFs have all the properties of ordinary CDFs — they are monotone nondecreasing, right-continuous functions with limits 0 at –∞ and 1 at ∞
- Conditional pdfs have all the properties of ordinary pdfs — they are non-negative functions of total area 1, and

$P\{u \le \mathbf{X} \le u+\Delta u\ |A\} = P\{u \le \mathbf{X} \le u+\Delta u\}/P(A)$
$\approx f_{\mathbf{X}|A}(u|A)\bullet\Delta u$

When A is specified in terms of X…

- Sometimes event A is specified in terms of the random variable **X**, e.g. $A = \{c < \mathbf{X} < d\}$
- P(A) = area under the unconditional pdf $f_{\mathbf{X}}(u)$ between c and d
  - $f_{\mathbf{X}|A}(u|A) = f_{\mathbf{X}}(u)/P(A)$ if u   A
  - $f_{\mathbf{X}|A}(u|A) = 0$ if u A^c
- P(A) = area under $f_{\mathbf{X}}(u)$ in region A is a "normalizing factor" that makes $f_{\mathbf{X}}(u)/P(A) = f_{\mathbf{X}|A}(u|A)$ a valid density

Conditional mean and variance of X

- We can calculate means and variances using conditional pdfs
  - The mean of the conditional pdf is called the conditional mean $E[\mathbf{X}|A]$
  - The variance of the conditional pdf is called the conditional variance $\mathrm{var}(\mathbf{X}|A)$
- Instead of using the unconditional pdf $f_{\mathbf{X}}(u)$, use the conditional pdf $f_{\mathbf{X}|A}(u|A)$ in the various integrals

Unconditional pdf from conditional

- Let $\{A_1, A_2, \ldots, A_n, \ldots\}$ be a (countable) partition of Ω
- The theorem of total probability gives the unconditional distributions in terms of the conditional distributions (cf. Lecture 14)
- $F_{\mathbf{X}}(u) = \sum F_{\mathbf{X}|A_i}(u|A_i)\bullet P(A_i)$
- $f_{\mathbf{X}}(u) = \sum f_{\mathbf{X}|A_i}(u|A_i)\bullet P(A_i)$
- $E[\mathbf{X}] = \sum E[\mathbf{X}|A_i]\bullet P(A_i)$

Unconditional variance of X

- The unconditional variance of **X** is found via the result $E[(\mathbf{X} - a)^2] = \mathrm{var}(\mathbf{X}) + (\mu - a)^2$
- $\mathrm{var}(\mathbf{X}) = \int (u-\mu)^2\bullet f_{\mathbf{X}}(u)\bullet du$

$$= \int (u-\mu)^2\bullet \sum_k f_{\mathbf{X}|A_k}(u|A_k)\bullet P(A_k)\bullet du$$

$$= \sum_k P(A_k)\bullet \int (u-\mu)^2\bullet f_{\mathbf{X}|A_k}(u|A_k)\bullet du$$

$$= \sum_k P(A_k)\bullet[\mathrm{var}(\mathbf{X}|A_k) + (\mu_k - \mu)^2]$$

where μ_k denotes $E[\mathbf{X}|A_k]$

Another way of looking at it

- var(**X**) $= \sum P(A_k)\bullet[\mathrm{var}(\mathbf{X}|A_k) + (\mu_k - \mu)^2]$
  $= \sum \mathrm{var}(\mathbf{X}|A_k)\bullet P(A_k) + \sum(\mu_k - \mu)^2\bullet P(A_k)$
- The first sum is the weighted sum of the conditional variances of **X**
- The second sum is the variance of a discrete random variable taking on values $\mu_1, \mu_2, \ldots$ with probabilities $P(A_1), P(A_2), \ldots$
- This discrete random variable can be thought of as E[**X**|A], the conditional mean

Slick formulas … as in Lecture 14

- The unconditional mean of **X** is the weighted sum of the conditional means, i.e., mean of the conditional means of **X**
- $\mu = E[\mathbf{X}] = \sum P(A_k)\bullet E[\mathbf{X}|A_k] = E[E[\mathbf{X}|A]]$
- The unconditional variance of **X** is the mean of the conditional variances plus the variance of the conditional means
- $\mathrm{var}(\mathbf{X}) = \sum \mathrm{var}(\mathbf{X}|A_k)\bullet P(A_k) + \sum(\mu_k - \mu)^2\bullet P(A_k)$
  $\mathrm{var}(\mathbf{X}) = E[\mathrm{var}(\mathbf{X}|A)] + \mathrm{var}(E[\mathbf{X}|A])$

Example

- $f_{\mathbf{X}}(u) = (\lambda/2)\bullet\exp(-\lambda|u|)$, $-\infty < u < \infty$
- A = {**X** ≥ 0}; A^c = {**X** < 0}. $P(A) = P(A^c)$
- $f_{X|A}(u|A) = \lambda\bullet\exp(-\lambda u)$ for $u \ge 0$ is an exponential pdf with parameter λ
- $f_{X|A^c}(u|A^c) = \lambda\bullet\exp(\lambda u)$ for $u < 0$ is just $f_{X|A}(u|A)$ “flipped over”
- $E[\mathbf{X}|A] = 1/\lambda = -E[\mathbf{X}|A^c]$, i.e. conditional mean equally likely to have value $\pm 1/\lambda$

Example (continued)

- Conditional mean equally likely to be $\pm 1/\lambda$
- $E[\mathbf{X}] = (1/\lambda)\bullet(1/2) + (-1/\lambda)\bullet(1/2) = 0$
- Conditional variance has same value $\mathrm{var}(\mathbf{X}|A) = 1/\lambda^2 = \mathrm{var}(\mathbf{X}|A^c)$
- var(**X**) = mean of conditional variance + variance of conditional mean
  $= (1/\lambda^2)\bullet(1/2) + (1/\lambda^2)\bullet(1/2)$
  $+ (1/\lambda)^2\bullet(1/2) + (-1/\lambda)^2\bullet(1/2)$
  $= 2/\lambda^2$

Another Example

- Conditioned on A, **X** is $\mathbf{N}(\mu_0, (\sigma_0)^2)$
- Conditioned on A^c, **X** is $\mathbf{N}(\mu_1, (\sigma_1)^2)$
- $E[\mathbf{X}] = \mu_0\bullet P(A) + \mu_1\bullet P(A^c) = \mu$
- var(**X**) = mean of conditional variance + variance of conditional mean
  $= (\sigma_0)^2\bullet P(A) + (\sigma_1)^2\bullet P(A^c)$
  $+ (\mu_0 - \mu)^2\bullet P(A) + (\mu_1 - \mu)^2\bullet P(A^c)$
  since conditional mean takes on values μ_0 and μ_1 with probabilities P(A) and $P(A^c)$

Decision making revisited

- We observe the value of a continuous random variable **X** and need to decide which of two hypotheses H_0 and H_1 is true
- A decision rule partitions the set of all real numbers into two sets Γ_0 and Γ_1
- If the observed value of **X** is in Γ_0, decide that H_0 is true
- If the observed value of **X** is in Γ_1, decide that H_1 is true

Decision regions

- Γ_0 and Γ_1 are called the decision regions associated with H_0 and H_1 respectively
  — if $\mathbf{X}$ Γ_i, we decide in favor of H_i
- Decision regions Γ_0 and Γ_1 can be chosen arbitrarily, or we can use various criteria
- If H_0 is the true hypothesis but $\mathbf{X}$ Γ_1, we have a false alarm or Type I error
- If H_1 is the true hypothesis but $\mathbf{X}$ Γ_0, we have a missed detection or false dismissal or Type II error

Probability density functions

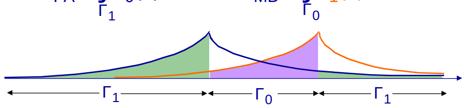
- Let $f_0(u)$ and $f_1(u)$ denote the pdfs of $\mathbf{X}$ when respectively H_0 and H_1 are true
- False alarm occurs if H_0 is true but $\mathbf{X}$ Γ_1
- $P_{FA} = P\{\mathbf{X}$ Γ_1 when H_0 is true$\}$
- When H_0 is true, the pdf of $\mathbf{X}$ is $f_0(u)$
- If H_1 is the true hypothesis but $\mathbf{X}$ Γ_0, we have a missed detection or false dismissal
- When H_1 is true, the pdf of $\mathbf{X}$ is $f_1(u)$
- $P_{MD} = P\{\mathbf{X}$ Γ_0 when H_1 is true$\}$

Error probabilities

- $P_{FA} = P\{\mathbf{X}$ Γ_1 when H_0 is true$\}$
- When H_0 is true, the pdf of $\mathbf{X}$ is $f_0(u)$
- Hence, $P_{FA} = \int_{\Gamma_1} f_0(u)\,du$ where the integral is over the decision region Γ_1
- $P_{MD} = P\{\mathbf{X}$ Γ_0 when H_1 is true$\}$
- When H_1 is true, the pdf of $\mathbf{X}$ is $f_1(u)$
- Hence, $P_{MD} = \int_{\Gamma_0} f_1(u)\,du$

Making mistakes is a big mistake!

- $P_{FA} = \int_{\Gamma_1} f_0(u)\,du$ $\quad P_{MD} = \int_{\Gamma_0} f_1(u)\,du$



- P_{FA} = green area $\quad P_{MD}$ = mauve area
- Reduce P_{FA} by decreasing the size of Γ_1? But this increases the size of Γ_0 and hence P_{MD} will increase

Average error probability

- Let hypotheses H_0 and H_1 be true with probabilities π_0 and $\pi_1 = 1 - \pi_0$ respectively
- Many statisticians deny the existence of such prior probabilities of hypotheses
- The average probability of error is
  $P(E) = \pi_0 \cdot P_{FA} + \pi_1 \cdot P_{MD}$
  $= \pi_0 \cdot \int_1 f_0(u)\,du + \pi_1 \cdot \int_0 f_1(u)\,du$
  where the integrals are over Γ_1 and Γ_0

How to choose decision regions?

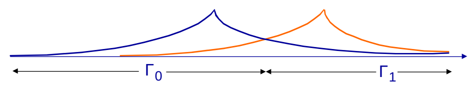
- The maximum-likelihood (ML) rule says that we should decide in favor of the hypothesis that maximizes the likelihood (probability) of the observation $\{\mathbf{X} = u\}$
- Since $\mathbf{X}$ is a continuous random variable under both hypotheses, $P\{\mathbf{X} = u\} = 0$
- Physically, we observe the value of $\mathbf{X}$ subject to some measurement error, i.e. we observe $\{u - \Delta u/2 \le \mathbf{X} \le u + \Delta u/2\}$

Maximum-likelihood decision rule

- We actually observe the event
  $\{u–\Delta u/2 \le \mathbf{X} \le u+\Delta u/2\}$
- $P\{u–\Delta u/2 \le \mathbf{X} \le u+\Delta u/2\} \approx f_0(u)\bullet\Delta u$ if H_0 is the true hypothesis
- $P\{u–\Delta u/2 \le \mathbf{X} \le u+\Delta u/2\} \approx f_1(u)\bullet\Delta u$ if H_1 is the true hypothesis
- Maximum-likelihood (ML) decision rule:
  If $f_1(u) > f_0(u)$, decide in favor of H_1
  If $f_1(u) < f_0(u)$, decide in favor of H_0

Decision regions for the ML rule

- Maximum-likelihood (ML) decision rule:
  If $f_1(u) > f_0(u)$, decide in favor of H_1
  If $f_1(u) < f_0(u)$, decide in favor of H_0
- Hence, $\Gamma_1(ML) = \{u: f_1(u) > f_0(u)\}$
  $\Gamma_0(ML) = \{u: f_1(u) < f_0(u)\}$



The ML rule as an LRT

- The ML decision rule can be stated in terms of a likelihood ratio test (LRT)
- Compute the value of the likelihood ratio $\Lambda(u) = f_1(u)/f_0(u)$ at the value u that was observed

$$\Lambda(u) \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

- Choose H_1 if $\Lambda > 1$ and choose H_0 if $\Lambda < 1$

Minimum average error probability

- $P(E) = \pi_0\bullet P_{FA} + \pi_1\bullet P_{MD}$
  $= \pi_0\bullet\int_{\Gamma_1} f_0(u)\,du + \pi_1\bullet\int_{\Gamma_0} f_1(u)\,du$
- Add and subtract $\pi_0\bullet\int f_0(u)\,du$ over region Γ_0 and remember $\Gamma_0 \quad \Gamma_1$ = entire real line
- $P(E) = \pi_0 + \int_{\Gamma_0} [\pi_1\bullet f_1(u) – \pi_0\bullet f_0(u)]\,du$
- To minimize P(E), choose the set Γ_0 to be those and only those real numbers u for which $\pi_1\bullet f_1(u)–\pi_0\bullet f_0(u)$ is negative!!

Decision regions for the MEP rule

- Bayes' or minimum error probability (MEP) or maximum ***a posteriori*** probability (MAP) decision rule:
  If $\pi_1\bullet f_1(u) > \pi_0\bullet f_0(u)$, decide in favor of H_1
  If $\pi_1\bullet f_1(u) < \pi_0\bullet f_0(u)$, decide in favor of H_0
- Hence, $\Gamma_1(MEP) = \{u: \pi_1\bullet f_1(u) > \pi_0\bullet f_0(u)\}$
  $\Gamma_0(MEP) = \{u: \pi_1\bullet f_1(u) < \pi_0\bullet f_0(u)\}$
- If $\pi_1 = \pi_0 = 1/2$, the MEP rule is the same as the ML rule. We have seen this before!

ML versus MEP decision regions

- Suppose that $\pi_1 < \pi_0$ $(\pi_0/\pi_1) = \alpha > 1$
- $\Gamma_1(MEP) = \{u: \pi_1\bullet f_1(u) > \pi_0\bullet f_0(u)\}$
  $= \{u: f_1(u) > \alpha\bullet f_0(u)\}$
- $\Gamma_1(ML) = \{u: f_1(u) > f_0(u)\}$
- If u $\quad \Gamma_1(MEP)$, then u $\quad \Gamma_1(ML)$ also
- $\Gamma_1(MEP) \quad \Gamma_1(ML)$; $\Gamma_0(MEP) \quad \Gamma_0(ML)$
- $P_{FA} = \int_{\Gamma_1} f_0(u)\,du \qquad P_{MD} = \int_{\Gamma_0} f_1(u)\,du$

$P_{FA}(MEP) < P_{FA}(ML)$; $P_{MD}(MEP) > P_{MD}(ML)$

Playing the odds…

- Remember that $\pi_1 < \pi_0$
- Γ_1(MEP) Γ_1(ML); Γ_0(MEP) Γ_0(ML)
- P_{FA}(MEP) < P_{FA}(ML)
- P_{MD}(MEP) > P_{MD}(ML)
- $P(E) = \pi_0 \cdot P_{FA} + \pi_1 \cdot P_{MD}$
- The MEP rule reduces P_{FA} while allowing P_{MD} to increase because P_{FA} is weighted more than P_{MD} in the P(E) expression

ECE 313 - Lecture 29 Slide 25 of 36

The MEP/MAP rule as an LRT

- The MEP/MAP decision rule can also be stated as an LRT
- Compute the value of the likelihood ratio $\Lambda(u) = f_1(u)/f_0(u)$ at the value u that was observed

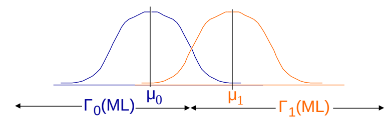
$$\Lambda \underset{H_0}{\overset{H_1}{\gtrless}} \pi_0/\pi_1$$

- This says choose H_1 if $\Lambda > \pi_0/\pi_1$ and choose H_0 if $\Lambda < \pi_0/\pi_1$

ECE 313 - Lecture 29 Slide 26 of 36

Example

- H_0 : $\mathbf{X} \sim \mathbf{N}(\mu_0, \sigma^2)$
- H_1 : $\mathbf{X} \sim \mathbf{N}(\mu_1, \sigma^2)$

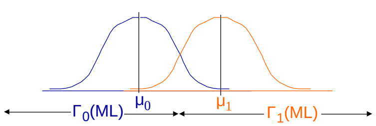


- It is easy to see by inspection that
  - $\Gamma_0(ML) = \{u: u < (\mu_0 + \mu_1)/2\}$
  - $\Gamma_1(ML) = \{u: u > (\mu_0 + \mu_1)/2\}$

ECE 313 - Lecture 29 Slide 27 of 36

Example (error probs. for ML rule)

- H_0 : $\mathbf{X} \sim \mathbf{N}(\mu_0, \sigma^2)$
- H_1 : $\mathbf{X} \sim \mathbf{N}(\mu_1, \sigma^2)$

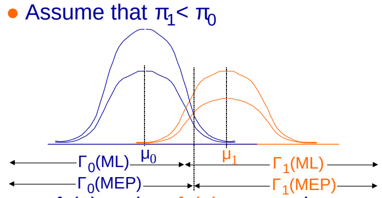


- $\Gamma_0(ML) = \{u: u < (\mu_0 + \mu_1)/2\}$
- $\Gamma_1(ML) = \{u: u > (\mu_0 + \mu_1)/2\}$
- $P_{FA}(ML) = P_{MD}(ML) = Q((\mu_1 - \mu_0)/\sigma) = Q(S)$

ECE 313 - Lecture 29 Slide 28 of 36

Example (MEP rule)

- H_0 : $\mathbf{X} \sim \mathbf{N}(\mu_0, \sigma^2)$ • H_1 : $\mathbf{X} \sim \mathbf{N}(\mu_1, \sigma^2)$
- Assume that $\pi_1 < \pi_0$



- $\pi_0 \cdot f_0(u)$ and $\pi_1 \cdot f_1(u)$ are as shown
- Γ_1(MEP) Γ_1(ML); Γ_0(MEP) Γ_0(ML)

ECE 313 - Lecture 29 Slide 29 of 36

Example (error probs. for MEP rule)

- H_0 : $\mathbf{X} \sim \mathbf{N}(\mu_0, \sigma^2)$ • H_1 : $\mathbf{X} \sim \mathbf{N}(\mu_1, \sigma^2)$
- $\Lambda(u) = \exp[\{-(u-\mu_1)^2 + (u-\mu_0)^2\}/2\sigma^2]$
  $= \exp[(\mu_1 - \mu_0) \cdot \{u - (\mu_1 + \mu_0)/2\}/\sigma^2]$
  is compared to $\pi_0/\pi_1 > 1$
- Take natural logarithms
- Compare $\ln(\Lambda(u))$ to $\ln(\pi_0/\pi_1)$
- $\Gamma_0(MEP) = \{u: u < (\mu_1 + \mu_0)/2 + \Theta\}$
- $\Gamma_1(MEP) = \{u: u > (\mu_1 + \mu_0)/2 + \Theta\}$
  where $\Theta = [\sigma^2/(\mu_1 - \mu_0)] \cdot \ln(\pi_0/\pi_1)$

ECE 313 - Lecture 29 Slide 30 of 36

Crossing the i's and dotting the t's

- The MEP rule compares the observed value of **X** to

  $(\mu_1+\mu_0)/2 + [\sigma^2/(\mu_1-\mu_0)]\bullet\ln(\pi_0/\pi_1)$

  and decides in favor of H_1 if **X** is larger
- We had defined $S = (\mu_1-\mu_0)/\sigma$
- Let $Y = [\sigma/(\mu_1-\mu_0)]\bullet\ln(\pi_0/\pi_1) = S^{-1}\bullet\ln(\pi_0/\pi_1)$
- $P_{FA}(MEP) = Q(S + Y) < Q(S) = P_{FA}(ML)$
- $P_{MD}(MEP) = Q(S - Y) > Q(S) = P_{MD}(ML)$

Some remarks

- Example: digital communication systems
- $S = (\mu_1-\mu_0)/\sigma$ is called the signal-to-noise ratio (SNR)
- Exercise: Show that

  $P(E) = \pi_0\bullet Q(S + S^{-1}\bullet\ln(\pi_0/\pi_1))$
  $+ \pi_1\bullet Q(S - S^{-1}\bullet\ln(\pi_0/\pi_1))$

  is a decreasing function of S
- Thus, designing a system to maximize the signal-to-noise S minimizes P(E)

General Bayes' rules

- Bayesian decision theory can also be used to minimize the average costs (also called risks) in making decisions
- $C_{i,j}$ is the cost if we decide that hypothesis H_j is true when in fact hypothesis H_i is true
- $C_{i,i}$ could be negative (indicating a reward) or it could be positive (representing the fixed costs associated with experiment)
- Assumption: $C_{i,j} > C_{i,i}$

The average risk or cost

- The cost is a discrete random variable **C** taking on values $C_{0,1}$, $C_{1,0}$, $C_{0,0}$, and $C_{1,1}$ with probabilities $\pi_0\bullet P_{FA}$, $\pi_1\bullet P_{MD}$, $\pi_0\bullet(1-P_{FA})$, and $\pi_1\bullet(1-P_{MD})$ respectively
- The average cost E[**C**] depends on P_{FA} and P_{MD}, i.e., the choice of the decision regions Γ_0 and Γ_1
- We can minimize E[**C**] by careful choice of Γ_0 and Γ_1

Minimum average risk or cost

- Same trick of adding and subtracting works to show how to choose Γ_0 and Γ_1
- The minimum average cost rule compares the likelihood ratio Λ to the threshold

  $(C_{0,1} - C_{0,0})\bullet\pi_0/(C_{1,0} - C_{1,1})\bullet\pi_1$
- Its average cost is called the Bayes' risk
- Uniform costs are said to be incurred when $C_{1,0} = C_{0,1}$ and $C_{0,0} = C_{1,1}$. In this case, we get the MEP rule!

Summary

- We have studied decision making with continuous random variables
- The maximum-likelihood (ML) decision rule, the minimum-error-probability (MEP) decision rule, and the more general minimum average cost rule can be formulated as likelihood ratio tests
- We looked at an example drawn from digital communication systems analysis