EEL 4930 Stats – Lecture 29

EEL 4930 Stats – Lecture 29

ERRORS AND PERFORMANCE TRADEOFFS IN HYPOTHESIS TESTING

EEL 4930 Stats – Lecture 29

Errors and Performance Tradeoffs in Hypothesis Testing

- In binary hypothesis testing, there are two types of errors:

EEL 4930 Stats – Lecture 29

ERRORS AND PERFORMANCE TRADEOFFS IN HYPOTHESIS TESTING

- In binary hypothesis testing, there are two types of errors:

1. **False Alarm** (Type I Error, also called False Positive)

EEL 4930 Stats – Lecture 29

ERRORS AND PERFORMANCE TRADEOFFS IN HYPOTHESIS TESTING

- In binary hypothesis testing, there are two types of errors:

1. **False Alarm** (Type I Error, also called False Positive)
   – occurs if we accept a hypothesis when it is not true

EEL 4930 Stats – Lecture 29

Errors and Performance Tradeoffs in Hypothesis Testing

- In binary hypothesis testing, there are two types of errors:

1. **False Alarm** (Type I Error, also called False Positive)
   - occurs if we accept a hypothesis when it is not true
   - we will use the notation

$$P_{fa} = P(\text{false alarm})$$

2. **Miss** (Type II Errors, also called False Negative)

2. **Miss** (Type II Errors, also called False Negative)
   - occurs if we reject a hypothesis when it is actually true

2. **Miss** (Type II Errors, also called False Negative)
   - occurs if we reject a hypothesis when it is actually true
   - we will use the notation

$$P_m = P(\text{miss})$$

- when performing a hypothesis test, there is always a tradeoff between these two types of errors

- when performing a hypothesis test, there is always a tradeoff between these two types of errors
- the tradeoff is controlled by choosing the significance level, a, to which the p-value is compared

- when performing a hypothesis test, there is always a tradeoff between these two types of errors
- the tradeoff is controlled by choosing the significance level, a, to which the p-value is compared
  - the value a is the probability that we will reject the null hypothesis, H_0 when it is in fact true

- when performing a hypothesis test, there is always a tradeoff between these two types of errors
- the tradeoff is controlled by choosing the significance level, a, to which the p-value is compared
  - the value a is the probability that we will reject the null hypothesis, H_0 when it is in fact true
  - equivalently, it is the probability of accepting the alternative hypothesis, H_A, when H_A is false

- even though the binary hypothesis test is usually conducted under the assumptions of H_0, we are usually conducting it to determine whether H_A is the cause of the observed difference

- even though the binary hypothesis test is usually conducted under the assumptions of H_0, we are usually conducting it to determine whether H_A is the cause of the observed difference
  - thus, we will consider the implications with respect to H_A when labeling errors

- even though the binary hypothesis test is usually conducted under the assumptions of H_0, we are usually conducting it to determine whether H_A is the cause of the observed difference
  - thus, we will consider the implications with respect to H_A when labeling errors
  - so for the case that we accept H_A when it is false, we call that a **false alarm**/Type I error

- even though the binary hypothesis test is usually conducted under the assumptions of H_0, we are usually conducting it to determine whether H_A is the cause of the observed difference
  - thus, we will consider the implications with respect to H_A when labeling errors
  - so for the case that we accept H_A when it is false, we call that a **false alarm**/Type I error
  - then $P_{fa} = a$

- note that if we decrease a, then we decrease P_{fa}, but we also decide that the null hypothesis could be true when it is in fact false

- note that if we decrease a, then we decrease P_{fa}, but we also decide that the null hypothesis could be true when it is in fact false
  - i.e., we increase the **Probability of Miss/Type II error**, P_m

- note that if we decrease a, then we decrease P_{fa}, but we also decide that the null hypothesis could be true when it is in fact false
  - i.e., we increase the **Probability of Miss/Type II error**, P_m
- the converse is also true

BINARY DECISIONS FROM CONTINUOUS DATA

Binary Decisions from Continuous Data

- We have many situations where we have a continuous measurement that depends on some underlying binary phenomena

Binary Decisions from Continuous Data

- We have many situations where we have a continuous measurement that depends on some underlying binary phenomena

- For example, we may wish to determine the presence of a disease based on the measurement of some chemical

BINARY DECISIONS FROM CONTINUOUS DATA

- We have many situations where we have a continuous measurement that depends on some underlying binary phenomena
- For example, we may wish to determine the presence of a disease based on the measurement of some chemical
  - Then the distribution of the data depends on whether the disease is present or not

- More generally, we assume the data comes from one of two continuous densities, $f_0(x|H_0)$ or $f_1(x|H_1)$

- More generally, we assume the data comes from one of two continuous densities, $f_0(x|H_0)$ or $f_1(x|H_1)$, and we wish to make a decision between H_0 and H_1 based on an observed value x

- We will choose H_i if $x \in R_i$, where R_0, R_1 partition the real line

- More generally, we assume the data comes from one of two continuous densities, $f_0(x|H_0)$ or $f_1(x|H_1)$, and we wish to make a decision between H_0 and H_1 based on an observed value x

- We will choose H_i if $x \in R_i$, where R_0, R_1 partition the real line

- The probability of false alarm and probability of miss then depend on the decision regions R_0 and R_1

- More generally, we assume the data comes from one of two continuous densities, $f_0(x|H_0)$ or $f_1(x|H_1)$, and we wish to make a decision between H_0 and H_1 based on an observed value x

- We will choose H_i if $x \in R_i$, where R_0, R_1 partition the real line

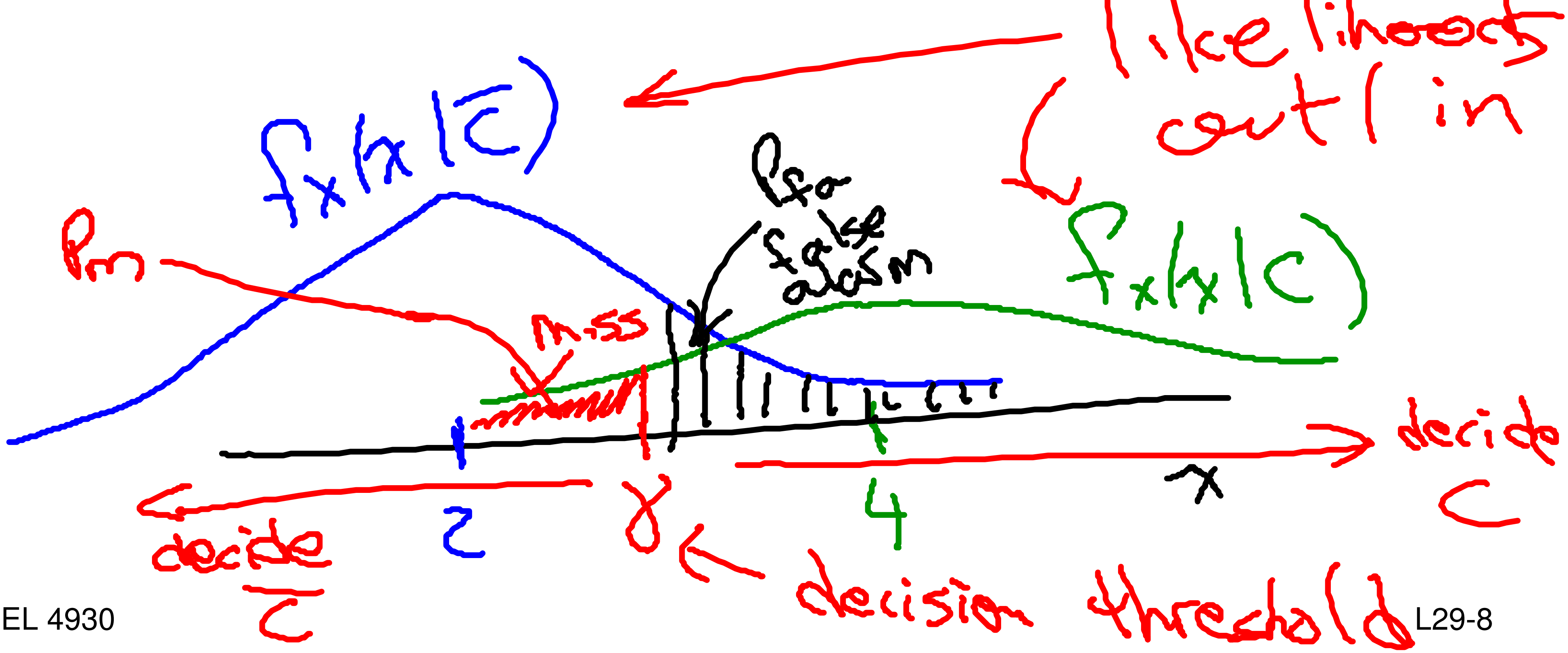
- The probability of false alarm and probability of miss then depend on the decision regions R_0 and R_1

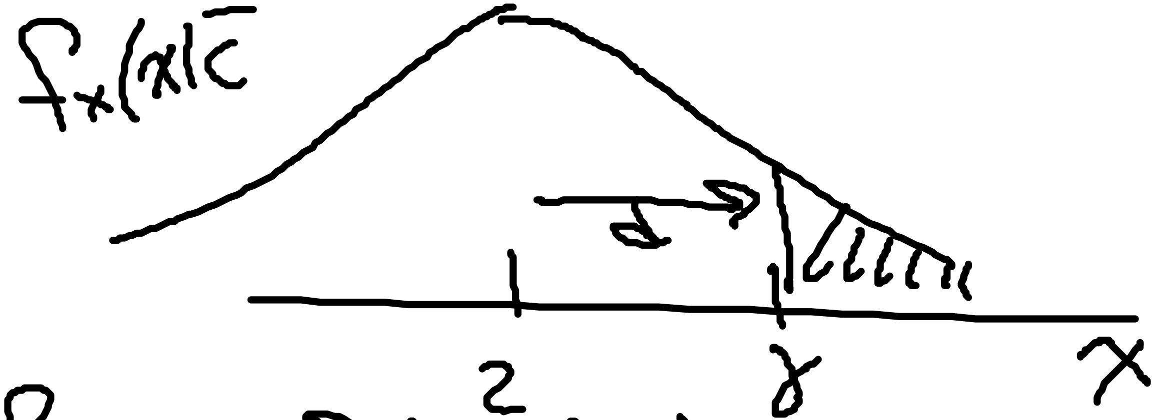
- In many cases, the decision regions are determined by a single threshold γ, like $R_0 = x < \gamma$ and $R_1 = x \geq \gamma$

EX The PSA values for men in their 60s without cancer are approximately Gaussian(2,$\sigma^2 = 1$). The PSA values are values for men in their 60s with cancer are approximately Gaussian(4, $\sigma^2 = 2$).

C = cancer

1. For $P_{fa} = 10\%$, find P_m.

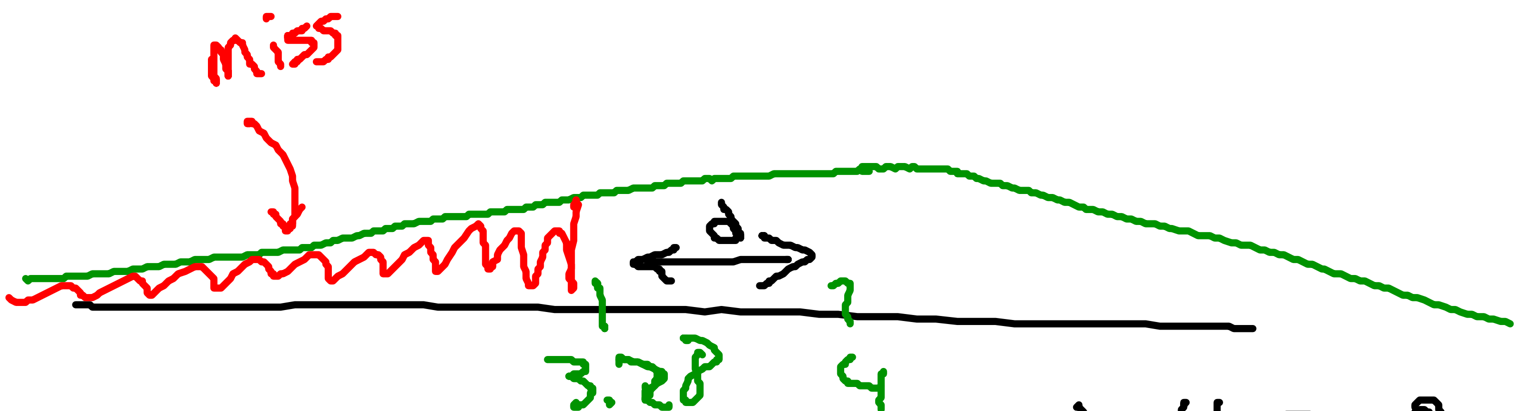




$$P_{fa} = Q\left(\frac{\gamma-2}{\cancel{\sigma}\,1}\right) = 0.1$$

$$\gamma - 2 = Q^{-1}(0.1) = 1.28$$

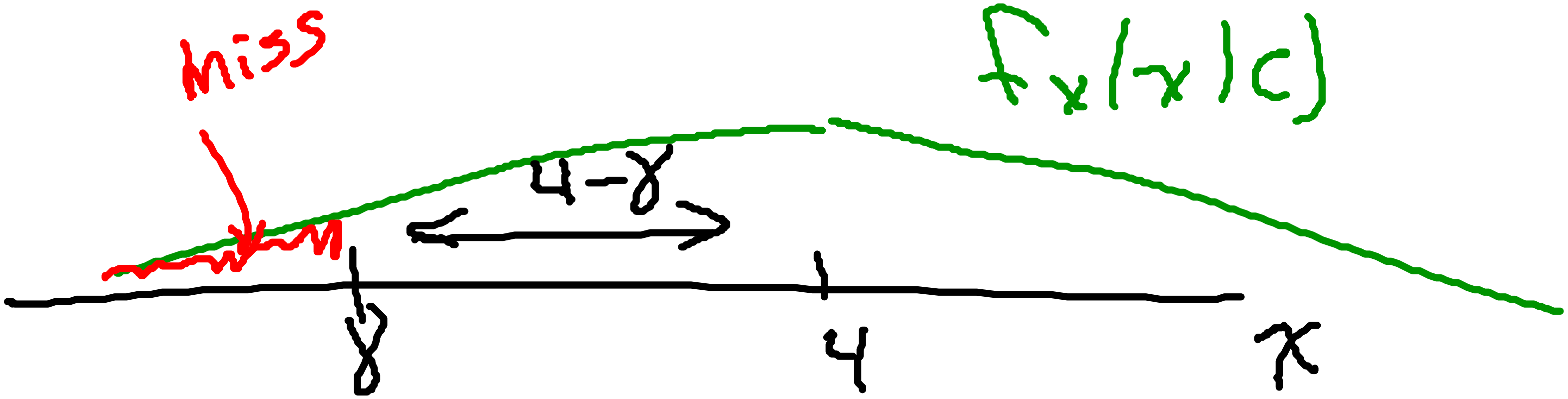
$$\gamma = 3.28$$



$$d = 4 - 3.28 = 0.72$$

$$P_m = Q\left(\frac{d}{\sigma}\right) = Q\left(\frac{0.72}{\sqrt{2}}\right) \approx 0.305$$

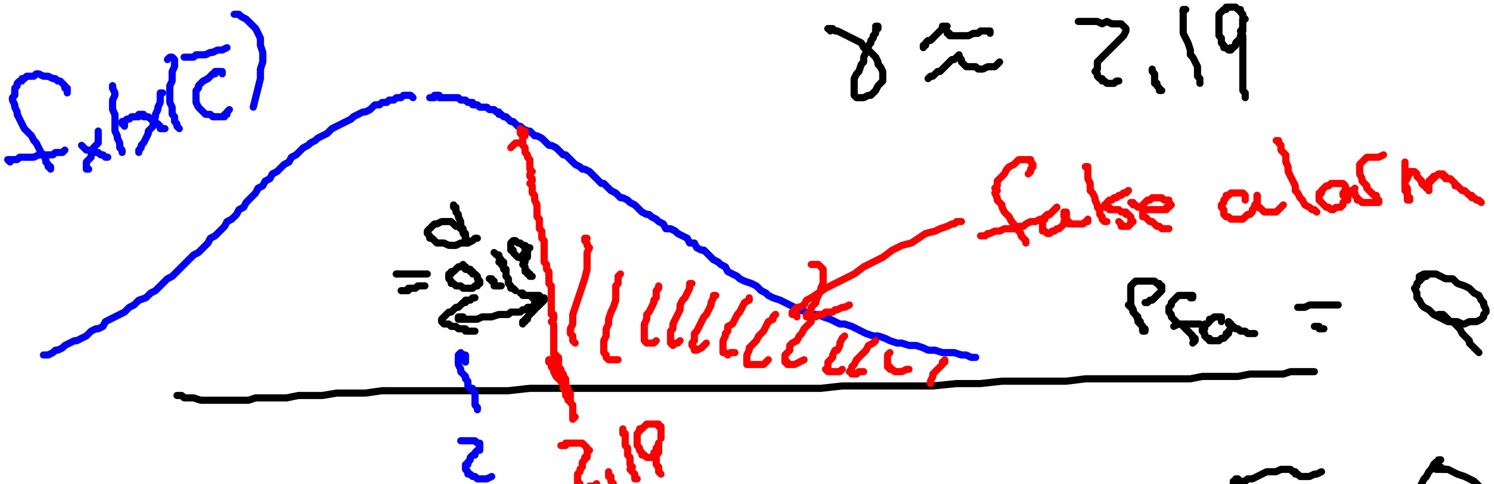
2. For $P_m = 10\%$, find P_{fa}.



$$P_m = Q\left(\frac{4-\gamma}{\sqrt{2}}\right) = 0.1$$

$$4-\gamma/\sqrt{2} = 1.28$$

$$\gamma \approx 2.19$$



$$P_{fa} = Q\left(\frac{0.19}{1}\right)$$

$$\approx 0.425$$

Visualizing Tradeoffs in Hypothesis Testing: ROC Curves

Visualizing Tradeoffs in Hypothesis Testing: ROC Curves

- We can visualize the relation between these types of errors using a ROC curve

Visualizing Tradeoffs in Hypothesis Testing: ROC Curves

- We can visualize the relation between these types of errors using a ROC curve
  - ROC= receiver operating characteristic

Visualizing Tradeoffs in Hypothesis Testing: ROC Curves

- We can visualize the relation between these types of errors using a ROC curve
  - ROC= receiver operating characteristic
  - ROC curves were developed for RADAR systems but are widely used in fields that do statistical tests, such as biomedicine

- In ROC curves, we do not plot P_{fa} vs P_m

- In ROC curves, we do not plot P_{fa} vs P_m
- Instead:

- In ROC curves, we do not plot P_{fa} vs P_m
- Instead:
  - the x-axis is FPR= false positive rate = P_{fa},

- In ROC curves, we do not plot P_{fa} vs P_m
- Instead:
  - the x-axis is FPR= false positive rate = P_{fa}, and
  - the y-axis is TPR= true positive rate = $1 - P_m$