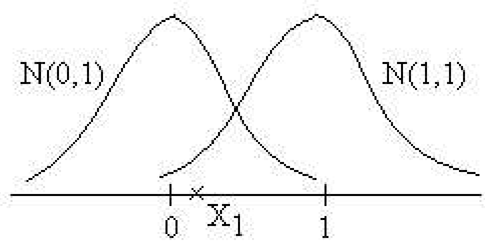
18.05 Lecture 29
April 25, 2005

Score distribution for Test 2:
70-100 A, 40-70 B, 20-40 C, 10-20 D
Average = 45

Hypotheses Testing.
$X_1, ..., X_n$ with unknown distribution $\mathbb{P}$
Hypothesis possibilities:
$H_1 : \mathbb{P} = \mathbb{P}_1$
$H_2 : \mathbb{P} = \mathbb{P}_2$
...
$H_k : \mathbb{P} = \mathbb{P}_k$
There are k simple hypotheses.
A simple hypothesis states that the distribution is equal to a particular probability distribution.

Consider two normal distributions: N(0, 1), and N(1, 1).



There is only 1 point of data: X_1
Depending on where the point is, it is more likely to come from either N(0, 1) or N(1, 1).
Hypothesis testing is similar to maximum likelihood testing $\rightarrow$
Within your k choices, pick the most likely distribution given the data.
However, hypothesis testing is NOT like estimation theory, as there is a different goal:

Definition: Error of type i
$\mathbb{P}$(make a mistake $|H_i$ is true) $= \alpha_i$
Decision Rule: $\delta : \mathcal{X}^n \rightarrow (H_1, H_2, ..., H_k)$
Given a sample $(X_1, ..., X_n), \delta(X_1, ..., X_n) \in \{H_1, ..., H_k\}$
$\alpha_i = \mathbb{P}(\delta \neq H_i | H_i)$ - error of type i
"The decision rule picks the wrong hypothesis" = error.

Example: Medical test, H_1 - positive, H_2 - negative.
Error of Type 1: $\alpha_1 = \mathbb{P}(\delta \neq H_1 | H_1) = \mathbb{P}(negative|positive)$
Error of Type 2: $\alpha_2 = \mathbb{P}(\delta \neq H_2 | H_2) = \mathbb{P}(positive|negative)$
These are very different errors, have different severity based on the particular situation.

Example: Missile Detection vs. Airplane
Type 1 $\rightarrow \mathbb{P}(airplane|missile)$, Type 2 $\rightarrow \mathbb{P}(missile|airplane)$
Very different consequences based on the error made.

Bayes Decision Rules
Choose a prior distribution on the hypothesis.

Assign a weight to each hypothesis, based upon the importance of the different errors.
$\xi(1), ..., \xi(k) \geq 0, \sum \xi(i) = 1$
Bayes error $\alpha(\xi) = \xi(1)\alpha_1 + \xi(2)\alpha_2 + ... + \xi(k)\alpha_k$
Minimize the Bayes error, choose the appropriate decision rule.
Simple solution to finding the decision rule:
$\mathbf{X} = (X_1, ..., X_n)$, let $f_i(x)$ be a p.f. or p.d.f. of $\mathbb{P}_i$
$f_i(\mathbf{x}) = f_i(x_1) \times ... \times f_i(x_n)$ - joint p.f./p.d.f.

Theorem: Bayes Decision Rule:

$$\delta = \{H_i : \xi(i)f_i(\mathbf{x}) = \max_{i \leq j \leq k} \xi(j)f_j(\mathbf{x})$$

Similar to max. likelihood.
Find the largest of joint densities, but **weighted** in this case.

$\alpha(\xi) = \sum \xi(i)\mathbb{P}_i(\delta \neq H_i) = \sum \xi(i)(1 - \mathbb{P}_i(\delta = H_i)) =$
$= 1 - \sum \xi(i)\mathbb{P}_i(\delta = H_i) = 1 - \sum \xi(i) \int I(\delta(\mathbf{x}) = H_i)f_i(\mathbf{x})d\mathbf{x} =$
$= 1 - \int (\sum \xi(i) I(\delta(\mathbf{x}) = H_i)f_i(\mathbf{x}))d\mathbf{x}$ - minimize, so maximize the integral:
Function within the integral:

$$I(\delta = H_1)\xi(1)f_1(\mathbf{x}) + ... + I(\delta = H_k)\xi(k)f_k(\mathbf{x})$$

The indicators pick the term $\rightarrow$
$\delta = H_1 : 1\xi(1)f_1(\mathbf{x}) + 0 + 0 + ... + 0$
So, just choose the largest term to maximize the integral.
Let δ pick the largest term in the sum.

Most of the time, we will consider 2 simple hypotheses:

$$\delta = \{H_1 : \xi(1)f_1(\mathbf{x}) > \xi(2)f_2(\mathbf{x}), \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} > \frac{\xi(2)}{\xi(1)}; H_2 \text{ if } <; H_1 \text{ or } H_2 \text{ if } =\}$$

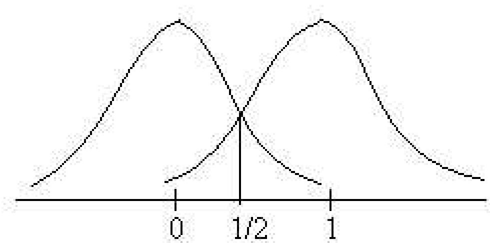
Example:
$H_1 : N(0, 1), H_2 : N(1, 1)$
$\xi(1)f_1(\mathbf{x}) + \xi(2)f_2(\mathbf{x}) \rightarrow$ minimize

$$f_1(\mathbf{x}) = (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}\sum x_i^2}; f_2(\mathbf{x}) = (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}\sum (x_i - 1)^2}$$

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} = e^{-\frac{1}{2}\sum x_i^2 + \frac{1}{2}\sum (x_i-1)^2} = e^{\frac{n}{2} - \sum x_i} > \frac{\xi(2)}{\xi(1)}$$

$$\delta = \{H_1 : \sum x_i < \frac{n}{2} - \log \frac{\xi(2)}{\xi(1)}; H_2 \text{ if } >; H_1 \text{ or } H_2 \text{ if } =\}$$

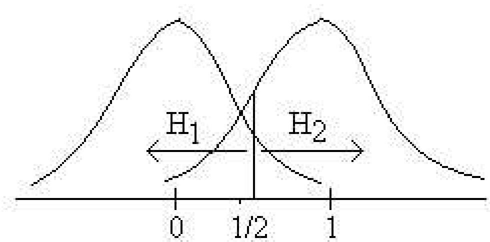
Considering the earlier example, N(0, 1) and N(1, 1)



$X_1, n = 1, \xi(1) = \xi(2) = \frac{1}{2}$

$$\delta = \{H_1 : x_1 < \frac{1}{2}; H_2 x_1 > \frac{1}{2}; H_1 \text{ or } H_2 \text{ if } =\}$$

However, if 1 distribution were more important, it would be weighted.



If N(0, 1) were more important, you would choose it more of the time, even on some occasions when $x_i > \frac{1}{2}$

Definition: H_1, H_2 - two simple hypotheses, then:
$\alpha_1(\delta) = \mathbb{P}(\delta \neq H_1 | H_2)$ - level of significance.
$\beta(\delta) = 1 - \alpha_2(\delta) = \mathbb{P}(\delta = H_2 | H_2)$ - power.
For more than 2 hypotheses,
$\alpha_1(\delta)$ is always the level of significance, because H_1 is always the
Most Important hypothesis.
$\beta(\delta)$ becomes a power function, with respect to each extra hypothesis.

Definition: H_0 - null hypothesis
Example, when a drug company evaluates a new drug,
the null hypothesis is that it doesn't work.
H_0 is what you want to disprove first and foremost,
you don't want to make that error!

Next time: consider class of decision rules.
$K_\alpha = \{\delta : \alpha_1(\delta) \leq \alpha\}, \alpha \in [0, 1]$
Minimize $\alpha_2(\delta)$ within the class K_α

** End of Lecture 29