MEAN AND HISTOGRAM APPLICATIONS

Recording a CD: Laser carves pits, showing bits 0 or 1, from:

Music $x(t) \rightarrow$ |**ANTIALIAS FILTER**| $\rightarrow$ |**SAMPLE @44.1 kHz**| $\rightarrow$ |**QUAN– TIZER**| $\rightarrow$ 011001100011

Playing a CD: Laser reads pits, outputing bits 0 or 1. **However:**
We observe $y(t) = x(t) + v(t)$ where $x(t)$=0 or 1 and $v(t) = noise$.
$x(t)$ switches back and forth from 0 to 1, which are equally likely.

Problem: Given data $\{y(t)\}$, compute $x(t)$=0 or 1. **Solution:**

1. Partition $y(t)$ into blocks, each T seconds long ($T \approx 1\mu$sec).
   Within each block, $x(t)$ is *constant* (either $x(t)$=0 or $x(t)$=1).
2. Within each block, compute *mean* $M(y) = \frac{1}{T}\int_{BLOCK} y(t)dt$.
3. If $M(y) > \frac{1}{2}$, decide $x(t) = 1$ within that block.
   If $M(y) < \frac{1}{2}$, decide $x(t) = 0$ within that block.

How well does this work? Compute *probability of error*:
Pr[error]=0.5Pr[choose 1 when $x(t)$=0]+0.5Pr[choose 0 when $x(t)$=1]
First error is "false alarm"; second error is "fail to detect."
The (unknown) *mean* of $v(t)$ is $M(v) = \frac{1}{T}\int_{BLOCK} v(t)dt$.
Pr[choose 1 when $x(t)$=0]=Pr[M(v)$> +\frac{1}{2}$] since it is known that
Pr[choose 0 when $x(t)$=1]=Pr[M(v)$< -\frac{1}{2}$] $M(y) = M(x) + M(v)$

Problem: How do we compute probabilities when we don't know $v(t)$?
Solution: We don't *need* to know $v(t)$, only its *distribution*!
Observe $v(t)$ for awhile, sample it, and compute its *histogram*.
Use histogram to estimate distribution of $v(t)$ when $v(t)$ unknown.
If we think $v(t)$ has a Gaussian distribution, we estimate its *variance*
by computing $\sigma_v^2 = \frac{1}{T}\int_0^T v(t)^2 dt - (\frac{1}{T}\int_0^T v(t)dt)^2$ from data $\{v(t)\}$.

Problem: Suppose we don't know $x(t)$ either, only its distribution!
Want to decide if $x(t)$ is present or absent (i.e., $x(t)$=0) from $\{y(t)\}$.
Called "detection of unknown signal in unknown noise." Sounds hard!
Solution: Yet, if $x(t)$ and $v(t)$ have *correlation* $C(x, v) = 0$, we can:

1. Within each block, compute *rms* $RMS(y) = \sqrt{\frac{1}{T}\int_{BLOCK} y(t)^2 dt}$.
2. If $RMS(y) >$ threshold, decide $x(t)$ present within that block.
   If $RMS(y) <$ threshold, decide $x(t)$ absent within that block.

This works because $\sigma_y^2 = \sigma_x^2 + \sigma_v^2$ if their *correlation* $C(x, v) = 0$.
Threshold is set using distribution (estimated using histogram) of $y(t)$.
This is something like what you will be doing in labs one and two.