~~Re~~ Estimating distributions

So far we have assumed that we have an analytical model for the probability densities of interest.

This was reasonable for quantization and is reasonable for other cases as well.

However, suppose we don't know the underlying density. Can we estimate it from observations of the signal?

Suppose we observe a sequence of i.i.d r.v.'s

$$X_1, \ldots, X_N$$

let $E(X) = \mu_X$

How do we estimate μ_X?

$$\hat{\mu}_X = \frac{1}{N} \sum_{n=1}^{N} X_n$$

How good is the estimate?

- $\hat{\mu}_X$ is actually a random variable — call it Y

$$E\{Y\} = E\left\{\frac{1}{N}\sum_{n=1}^{N} X_n\right\} = E\{X\} = \mu_X$$

$$E\{|Y-\mu_X|^2\} = E\left\{\left(\frac{1}{N}\sum_{n=1}^{N} (X_n - \mu_X)\right)^2\right\}$$

$$= \frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N} E\{[X_m-\mu_X][X_n-\mu_X]\}$$

$$= \frac{1}{N^2}\sum_{m=1}^{N}\sigma_X^2$$

$$= \frac{1}{N}\sigma_X^2$$

How do we estimate variance?

$$\hat{\sigma}_X^2 = \frac{1}{N}\sum_{n=1}^{N}\left|X_n - \hat{\mu}_X\right|^2$$

$$= \frac{1}{N-1}\sum_{n=1}^{N}\left|X_n\left(1-\frac{1}{N}\right) - \sum_{\substack{m=1\\m\neq n}}^{N} X_m\right|^2$$

$$= \frac{1}{N-1}\sum_{n=1}^{N}\left\{X_n^2\left(1-\frac{1}{N}\right)^2 - 2X_n\left(1-\frac{1}{N}\right)\sum_{\substack{m=1\\m\neq n}}^{N} X_m\right.$$

(4)

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{n=1}^{N} \left[[X_n - \mu_x] - [\hat{\mu}_x - \mu_x] \right]^2$$

$$= \frac{1}{N} \sum_{n=1}^{N} |X_n - \mu_x|^2$$

$$- 2[\hat{\mu}_x - \mu_x] \underbrace{\frac{1}{N} \sum_{n=1}^{N} [X_n - \mu_x]}_{\hat{\mu}_x - \mu_x}$$

$$+ \frac{1}{N} \sum_{n=1}^{N} [\hat{\mu}_x - \mu_x]^2$$

$$= \frac{1}{N} \sum_{n=1}^{N} |X_n - \mu_x|^2 - [\hat{\mu}_x - \mu_x]^2$$

call this r.v. Z

$$E\{Z\} = \frac{1}{N} \sum_{n=1}^{N} \sigma_X^2 - E\{|\hat{\mu}_x - \mu_x|^2\}$$

$$E\{|\hat{\mu}_x - \mu_x|^2\} = E\left\{\left(\frac{1}{N} \sum_{n=1}^{N} [X_n - \mu_x]\right)^2\right\}$$

$$= \frac{1}{N^2} \sum_{m=1}^{N} \sum_{n=1}^{N} \underbrace{E\{[X_m - \mu_x][X_n - \mu_x]\}}_{\begin{cases} \sigma_X^2, & m = n \\ 0, & \text{esle} \end{cases}}$$

$$E\{|\hat{\mu}_x - \mu_x|^2\} = \frac{1}{N^2} N\sigma_x^2 = \frac{1}{N}\sigma_x^2$$

$$\therefore E\{z\} = \sigma_x^2 - \frac{1}{N}\sigma_x^2$$

$$= \sigma_x^2\left(\frac{N-1}{N}\right)$$

So estimate is not unbiased

define unbiased estimate as

$$\hat{\sigma}^2_{x\,unb} = \frac{N}{N-1}\hat{\sigma}_x^2$$

$$= \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \hat{\mu}_x)^2$$

let this r.v. be U

$$\sigma_u^2 = E\left\{\left|\frac{1}{N-1}\sum_{n=1}^{N}(X_n - \hat{\mu}_x)^2 - \sigma_x^2\right|^2\right\}$$

$$= E\left\{\left|\frac{1}{N-1}\sum_{n=1}^{N}\left[(X_n - \hat{\mu}_x)^2 - \frac{N-1}{N}\sigma_x^2\right]\right|^2\right\}$$

$$= \left(\frac{1}{N-1}\right)^2 \sum\sum$$

5a

$$\sigma_{x}^{2} = \frac{1}{N} \sum_{n=1}^{N} |x_n - \hat{\mu}_x|^2$$

$$= \frac{1}{N} \left\{ \sum_{n=1}^{N} |x_n|^2 - 2\hat{\mu}_x \sum_{n=1}^{N} x_n + \hat{\mu}_x^2 \right\}$$

$$= \frac{1}{N} \sum_{n=1}^{N} |x_n|^2 - \hat{\mu}_x^2$$

$$\therefore \quad \sigma_{x_{unb}}^{2} = \frac{N}{N-1} \hat{\sigma}_x^2$$

So far we have seen how to estimate statistics such as mean & variance of a random variable.

These ideas can be extended to estimation of any moments of the random variable. But suppose we really want to actually know the density itself.

How do we do this?

Basically what we need to compute is a histogram.

Consider a seq. of i.i.d r.v.s with density $f_X(x)$.

We divide the domain of X into intervals

$$I_k = \{ x : (k-\tfrac{1}{2})\Delta < x \le (k+\tfrac{1}{2})\Delta \}$$

We then define r.v.'s

$$U^k = \begin{cases} 1, & X \in I_k \\ 0, & \text{else} \end{cases}$$

We observe an i.i.d sequence X_n $n=1,\dots,N$ and compute the corresponding r.v.s U_n^k.

Finally we let

$$z^k = \frac{1}{N}\sum_{n=1}^{N} U_n^k$$

Then since this r.v. has same form as estimator of mean which we analyzed earlier, we know that

$$E\{z^k\} = E\{U_n^k\}$$

&

$$\sigma_{z^k}^2 = \frac{1}{N}\sigma_{U^k}^2$$

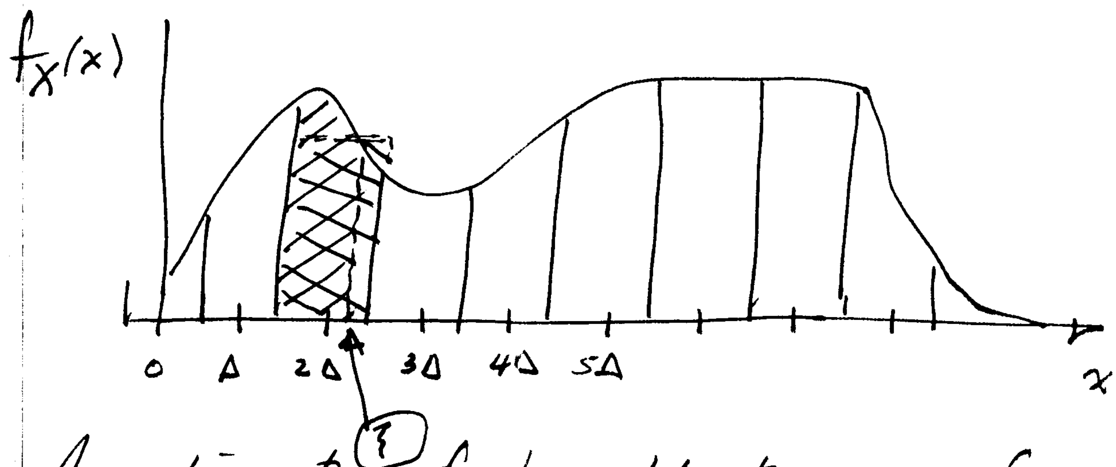
Now what is $E\{U_n^k\}$?

$$E\{U_n^k\} = 1 \cdot P\{X_n \in I_k\} + 0 \cdot P\{X_n \notin I_k\}$$

$$= P\{X_n \in I_k\}$$

$$= \int_{(k-\frac{1}{2})\Delta}^{(k+\frac{1}{2})\Delta} f_X(x)\,dx$$

So we see that we are estimating the integral of the density function over the range $(\cancel{k-\frac{1}{2}})\Delta < x \le (k+\frac{1}{2})\Delta$.



According to fundamental theorem of calculus, ~~∃~~ ~~st~~ there exists

$$(k-\tfrac{1}{2})\Delta < \xi \le (k+\tfrac{1}{2})\Delta$$

such that

$$f_X(\xi)\Delta = \int_{(k-\frac{1}{2})\Delta}^{(k+\frac{1}{2})\Delta} f_X(x)\,dx = E\{z^k\}$$

Assuming $f_X(x)$ is continuous,

As $\Delta \to 0$

$$E\{z^k\} = f_X(\xi)\Delta \longrightarrow f_X(k\Delta)\Delta$$

defined as

$$\cancel{SNR}\ \Phi = \frac{\left(\overline{U^k}\right)^2}{\sigma^2_{U^k}} = \overline{U^k}$$

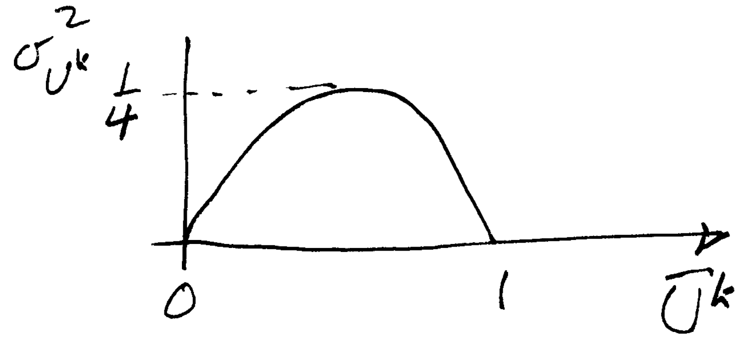
For fixed $f_X(x)$, $\overline{U^k} = P\{X \in I_k\}$ will increase with increasing Δ so there is a tradeoff between measurement noise & measurement resolution.

Of course, for fixed Δ & $f_X(x)$, we can always improve our results by increasing N.

What about variance of U_n^k?

$$E\{(U_n^k)^2\} = 1^2 P\{X \in I_k\} + 0^2 \cdot P\{X \notin I_k\}$$
$$= P\{X \in I_k\}$$
$$= E\{U_n^k\}$$
$$= \overline{U^k}$$

so
$$\sigma^2_{U^k} = \overline{U^k} - \left(\overline{U^k}\right)^2$$
$$= \overline{U^k}\left(1 - \overline{U^k}\right)$$



Would expect to operate in range where Δ is small so

$$\sigma^2_{U^k} \approx \overline{U^k}$$

let SNR of measurement be

defined as

$$\cancel{\mathrm{SNR}}\ \Phi = \frac{\left(\overline{U^k}\right)^2}{\sigma^2_{U^k}} = \overline{U^k}$$

For fixed $f_X(x)$, $\overline{U^k} = P\{X \in I_k\}$ will increase with increasing Δ so there is a tradeoff between measurement <u>noise</u> & measurement <u>resolution</u>.

Of course, for fixed Δ & $f_X(x)$, we can always improve our results by increasing N.