1.7.6. Random Sequences.

Def. A DT random (stochastic) process (or DT random signal, or random sequence)

$$X(n), \quad -\infty < n < \infty$$

is a sequence of random variables

$$\ldots X(-1), X(0), X(1), \ldots$$

(defined on the same probability space).

Alternative notation: X_n

$$\ldots X_{-1}, X_0, X_1, \ldots$$

Example 1 Flip a coin every second, and let

$$X(n) = \begin{cases} 1 & \text{if } n\text{-th flip is heads} \\ -1 & \text{" " " " tails} \end{cases}$$

This sequence of binary random variables is a random process

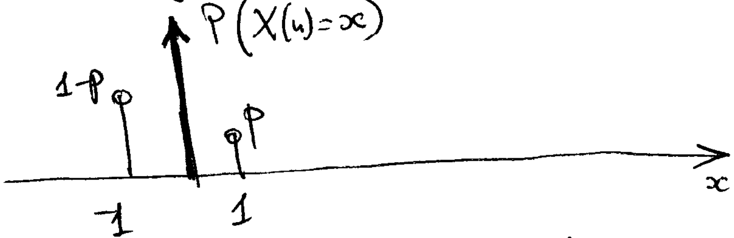
The observations associated with physical processes are often most appropriately modeled as random. It is typically the case, however, that their probability distribution is unknown and has to be learned from the data.

For example, when we flip a coin, we do not know a priori that $P(\text{heads}) = \frac{1}{2}$: if the coin is unfair, then we may have $P(\text{heads}) \neq \frac{1}{2}$.

1.7.7. Estimating Distributions.

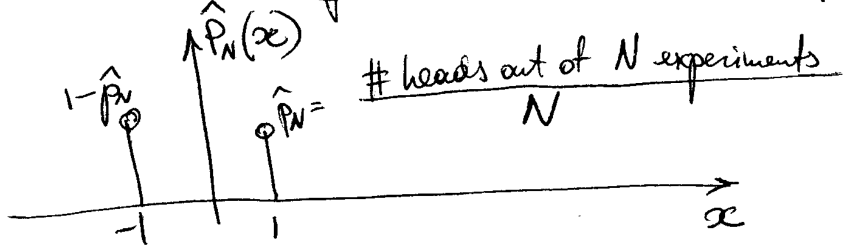
Example 1. Flip a coin [independently, multiple times]

$$X(n) = \begin{cases} 1 & \text{if } n\text{-th observation is heads} \\ -1 & \text{if } n\text{-th observation is tails} \end{cases}$$



Suppose we don't know p. How can we learn it from observing $X(n)$'s?

Solution. Construct empirical distribution $\hat{P}_N(x)$



heads	tails
✓	✓
✓	
✓	
⋮	⋮

How good is this estimate?

Let $H(n) = \begin{cases} 1 & \text{if } X(n) = 1 \\ 0 & \text{else} \end{cases}$.

$E(H(n)) = p \cdot 1 + (1-p) \cdot 0 = p$; $E((H(n))^2) = p \Rightarrow Var(H(n)) = p - p^2$.

Note that $\hat{p}_N = \dfrac{\sum_{n=1}^{N} H(n)}{N}$ - also a random variable!

$$\boxed{E(\hat{p}_N) = E\left\{\frac{\sum_{n=1}^{N} H(n)}{N}\right\} \stackrel{\text{linearity of expectation}}{=} \frac{1}{N}\sum_{n=1}^{N} \underbrace{E(H(n))}_{p} = p}$$

- this is good: the expected value of our estimator is exactly equal to the quantity we are trying to estimate. <u>Unbiased estimator</u>.

$\mathrm{Var}(\hat{p}_N) = ?$

Useful properties of Var:

1. $\cancel{\mathrm{Var}(\alpha Z)}$ $\mathrm{Var}(\alpha W) = E\left[\left(\alpha W - E(\alpha W)\right)^2\right] = E\left[\alpha^2 (W - E(W))^2\right]$

number α, r.v. W

$= \alpha^2 E\left[(W - E(W))^2\right]$

$= \alpha^2 \mathrm{Var}(W)$

2. $\mathrm{Var}\left(\sum_{n=1}^{N} W_n\right) = ?$ It depends. E.g.,

(a) Suppose $W_1 = W_2 = \dots = W_N = W$ are all the same.
then $\mathrm{Var}\left(\sum_{n=1}^{N} W_n\right) = \mathrm{Var}(N W_n) = N^2 \mathrm{Var}(W)$

(b) Suppose $W_1, \dots, W_N$ are independent. Then

$$\mathrm{Var}\left(\sum_{n=1}^{N} W_n\right) = E\left\{\left[\sum_{n=1}^{N}(W_n - E(W_n))\right]^2\right\}$$

$$= E\left\{\sum_{n=1}^{N} (W_n - E(W_n))^2 + 2\sum_{n \neq m} (W_n - E(W_n))(W_m - E(W_m))\right\}$$

$$= \sum_{n=1}^{N} E\left\{(W_n - E(W_n))^2\right\} + 2\sum_{n \neq m} E(W_n - E(W_n)) E(W_m - E(W_m))$$

$$= \sum_{n=1}^{N} \mathrm{Var}(W_n)$$

Thus, if we only make 1 experiment but count its result N times,

$$\mathrm{Var}(\hat{p}_N) = \mathrm{Var}\left\{\frac{\sum_{n=1}^{N} H(n)}{N}\right\} = \frac{1}{N^2}\mathrm{Var}\left\{\sum_{n=1}^{N} H(n)\right\} = \mathrm{Var}(H(n))$$

If, however, the N experiments are independent, then

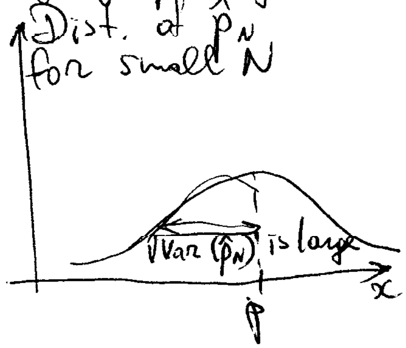
$$\boxed{\mathrm{Var}(\hat{p}_N) = \frac{1}{N^2}\mathrm{Var}\left\{\sum_{n=1}^{N} H(n)\right\} = \frac{1}{N^2}\sum_{n=1}^{N}\mathrm{Var}(H(n)) = \frac{1}{N}\mathrm{Var}(H(n)) = \boxed{\frac{p-p^2}{N}}}$$

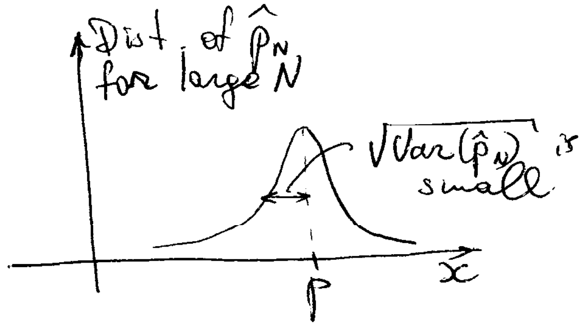
<u>Consistent</u> estimator: $\mathrm{Var} \to 0$ as $N \to \infty$.

I.e., as $N \to \infty$, our estimate becomes more reliable.

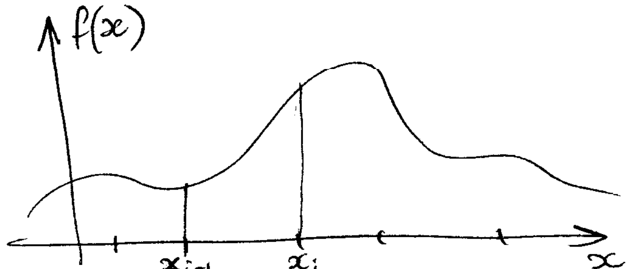
Variance ~~weakness~~ of the estimate is a measure of its reliability.

Roughly speaking,





Example 2. $X(1), \ldots, X(N)$ are independent, identically distributed (iid) r.v.'s, with unknown pdf $f(x)$.
How to estimate $f(x)$?



One possible method (Lab 7):

- partition the x-axis into L intervals
  $[x_0, x_1], (x_1, x_2], \ldots, (x_{L-1}, x_L]$
- ~~[illegible]~~ estimate $p^{(i)} = P(x_{i-1} \le X(n) \le x_i) = \int_{x_{i-1}}^{x_i} f(x)\,dx$ as:
  $$\hat{p}_N^{(i)} = \frac{\#\text{ of outcomes that fall into the } i\text{-th interval}}{N}$$
- note that, if $f(x) \approx$ const on $[x_{i-1}, x_i]$, then $\int_{x_{i-1}}^{x_i} f(x)\,dx \approx f(x)(x_i - x_{i-1})$

$\Rightarrow$ estimate $f(x) \approx \dfrac{\hat{p}_N^{(i)}}{x_i - x_{i-1}}$ on $[x_{i-1}, x_i]$.

[Unless something is known a priori about how fast $f(x)$ changes, this method will not necessarily be successful in estimating $f(x)$.]

Again, let $H_i(n) = \begin{cases} 1 & \text{if } x_{i-1} \le X(n) \le x_i \\ 0 & \text{else} \end{cases}$

$\Rightarrow E(H_i(n)) = p^{(i)}$

$\Rightarrow E(\hat{p}_N^{(i)}) = E\left(\frac{\sum_{n=1}^{N} H_i(n)}{N}\right) = p^{(i)}$

$Var(\hat{p}_N^{(i)}) = \dfrac{p^{(i)} - (p^{(i)})^2}{N}$

$Var\left(\dfrac{\hat{p}_N^{(i)}}{x_i - x_{i-1}}\right) = \dfrac{p^{(i)} - (p^{(i)})^2}{N (x_i - x_{i-1})^2}$

$\Rightarrow$ if the bin size $x_i - x_{i-1}$ is small, need large # of experiments N to get a reliable estimate