Lecture 10 - Inference

So far, given a distr. F $\longrightarrow$

What is its pdf/pmf, cdf?

" $E[X]$ or $Var(X)$ for $X \sim F$?

Ex: If $X \sim$ Exponential(λ), find $E[X]$ & $Var(X)$.

$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda}, \; x \geq 0, \lambda > 0$

Now Given data from an unknown distribution

What is the pdf/pmf, cdf, $E[X]$, $Var(X)$?

i.e., data D is available, we want to infer the underlying distr & relevant quantities or parameters.
e.g. mean.

Ex: $D = \{0.9, 0.7, 0.91, 0.01, 2.3, \ldots\}$

Qu: If data is from a distr F, what is it?

" : What is $E[X]$ for $X \sim F$? (F generates D)

Ex: Suppose we flip a coin and all we know is we flip it 10 times. $p = Prob(Heads)$. What is p?

(a) $D_1 = \{H, T, H, T, H, T, H, H, T, T\}$

(b) $D_2 = \{H, H, H, T, T, H, H, H, H, H\}$

Based on data we can infer the value of p, i.e. we can "estimate" p, denoted $\hat{p}$, from data.

In (a), our "best guess for p" is $\hat{p} = \frac{5}{10} = 0.5$

" (b), " " " " " $\hat{p} = \frac{8}{10} = 0.8$

"best guess based on data" ≡ "(point) estimate"

Let $D = \{X_1, X_2, \ldots, X_n\}$ and $\boxed{X_i \overset{iid}{\sim} F}$, which is not known, but known to have mean and var.
i.e. $E[X] = \mu$ & $Var(X) = \sigma^2$ (but μ, σ^2 still unknown)

(1) Q: What is a good estimate of μ?
(i.e. what is $\hat{\mu}$?)

(2) Q: " " " " " σ^2: $\hat{\sigma}^2$?

(3) Q: " " " " " $F_X(x)$: $\hat{F}_X(x)$?

(4) Is there evidence in the data D for $\mu > 0$?

(5) " " " " " " $\sigma^2 \neq 10$?

(6) " " " " " for $F_X \equiv$ normal?

All questions (1)–(6) are examples of inference
(1)–(3) : Estimation
(4)–(6) : (Hypothesis) Testing

Also,

Inference → Parametric Inference: distr. family is available or known, but parameters unknown

Inference → Nonparametric Inference: Nothing known about the underlying data.

- Both parametric & non-parametric inference assumes random samples or iid observations.

Parametric Inference

$D = \{X_1, X_2, \ldots, X_n\}$

(i) $X_i \overset{iid}{\sim} F(x|\theta)$, F known, θ unknown.

$\hat{\theta} = ?$ Is $\theta = 10$?

Ex: (1) $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $\hat{\mu}$? $\hat{\sigma}$? $\mu \neq 0$?

(2) $X_i \overset{iid}{\sim} Exp(\lambda)$, $\hat{\lambda}$? $\lambda > 1$?

(3) $X_i \overset{iid}{\sim} N(\mu, 10^2)$ $\hat{\mu}$?

(4) $X_i \overset{iid}{\sim} Uniform(a, b)$ $\hat{a}, \hat{b}$?

Non-parametric Inference

$D = \{X_1, X_2, \ldots, X_n\}$ $\quad F \in \mathcal{F} = \{\text{all cdfs}\}$

(ii) $X_i \overset{iid}{\sim} F$, nothing is known on F, even the # of parameters it has (if any) is unknown.

How to estimate F? i.e. $\hat{F}_X$?

" " " $E[X]$ for $X \sim F$? i.e. $\widehat{E[X]}$?

Non-parametric Inference for Discrete Data

Data is from F, unknown, want to estimate cdf or pmf.

we will estimate pmf (as cdf & pmf are both probabilities & carry the equivalent info about the distr.).

i.e. what is our best guess at pmf P_X? i.e. $\hat{P}_X(a)$?

$D = \{X_1, X_2, \ldots, X_n\}$ $X_i \overset{iid}{\sim} F_X$ with pmf P_X for a given $a \in \mathbb{R}$

$\mathcal{X} = \{a_1, a_2, \ldots\}$ be the support of X i.e. only $P_X(a_i) > 0$

Qu: Can X_i be α where $P_X(\alpha) = 0$? Ans: $D \subseteq \mathcal{X}$

Ex1: $D = \{4\} \Rightarrow \hat{P}_X(\alpha) = \begin{cases} 1 & \text{if } \alpha = 4 \\ 0 & \text{o.w.} \end{cases}$

Ex2: $D = \{13, 31\} \Rightarrow \hat{P}_X(\alpha) = \begin{cases} 1/2 & \alpha = 13 \\ 1/2 & \alpha = 31 \\ 0 & \text{o.w.} \end{cases}$

Ex 3: $D = \{1, 2, 1, 1\} \Rightarrow \hat{P}_X(\alpha) = \begin{cases} 3/4 & \text{if } \alpha = 1 \\ 1/4 & \text{if } \alpha = 2 \\ 0 & \text{o.w.} \end{cases}$

In general, for $D = \{X_1, X_2, \ldots X_n\}$ iid from a distr.

$$\hat{P}_X(\alpha) = \begin{cases} \frac{\#(X_i = \alpha)}{n} & \text{for all } X_i = \alpha,\ i = 1, 2, \ldots, n \\ 0 & \text{o.w.} \end{cases}$$

e.g.

for unique X_i, $\hat{P}_X(\alpha) = \begin{cases} 1/n & \text{if } \alpha = X_i,\ i = 1, 2, \ldots, n \\ 0 & \text{o.w.} \end{cases}$

more formally,

$$\boxed{\hat{P}_X(\alpha) = \frac{\sum_{i=1}^{n} I(X_i = \alpha)}{n}}$$

empirical pmf (or epmf)

Notes: X_i : RV

for a given α, $I(X_i = \alpha)$: RV (takes on values 0 or 1)

$\hat{P}_X(\alpha) = \frac{\sum_{i=1}^{n} I(X_i = \alpha)}{n}$ = epmf : RV (average of indicator RVs).

Ex: $D = \{X_1, \ldots X_{10}\} \rightarrow \hat{P}_X(0.1) = \frac{\sum_{i=1}^{10} I(X_i = 0.1)}{10}$

e.g. if only one $X_i = 0.1$, $\hat{P}_X(0.1) = 1/10$

" " 3 X_i's $= 0.1$, $\hat{P}_X(0.1) = 3/10$

" none $X_i = 0.1$ $\hat{P}_X(0.1) = 0$

Since $\hat{P}_X(\alpha)$ is a RV, it has its own distr.
so, it makes sense to talk about its expectations:

$$E[\hat{P}_X(\alpha)] = E\left[\frac{\sum_{i=1}^{n} I(X_i=\alpha)}{n}\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^{n} I(X_i=\alpha)\right]$$

$$\overset{LOE}{=} \frac{1}{n} \sum_{i=1}^{n} E[I(X_i=\alpha)]$$

$$= \frac{1}{n} \sum_{i=1}^{n} P(X_i=\alpha)$$

$$\overset{iid}{=} \frac{1}{n} \sum_{i=1}^{n} P(X_1=\alpha)$$

$$= \frac{1}{\cancel{n}} \cancel{n} P(X_1=\alpha) = P(X_1=\alpha)$$

$$= P_X(\alpha) \quad \text{since } X_1 \sim F$$

(def'n of pmf)

In general if $\hat{\theta}$ is an estimator of θ, then
$\hat{\theta}$ is a RV,
and if $E[\hat{\theta}]=\theta$, we say $\hat{\theta}$ is <u>unbiased estimator</u> of θ.

So, here we just showed that epmf, $\hat{P}_X(\alpha)$, is an unbiased estimator of $P_X(\alpha)$.

- Recall for $D=\{X_1, X_2, \rightarrow X_n\}$ epmf is

$$\hat{P}_X(\alpha) = \frac{\sum_{i=1}^{n} I(X_i=\alpha)}{n}$$, which is itself a pmf, with support $= D$.

So, mean of $\hat{P}_X(\alpha)$ can also be found as

$$\text{Mean of epmf} = \sum_{\alpha \in D} \alpha \hat{P}_X(\alpha)$$
$$= \sum_{i=1}^{n} x_i \hat{P}_X(x_i)$$
$$= \sum_{i=1}^{n} x_i \frac{1}{n} = \frac{\sum x_i}{n} = \overline{X}_n \text{ (sample mean)}$$

- What is the mean of the pmf $P_X(\alpha)$?
(i.e. for $X \sim P_X$, $\mu = E[X] = ?$)

$\mu = E[X] = \sum_{a \in \mathcal{X}} a\, P_X(a)$ where $\mathcal{X}$ is support of X.

- Note that in finding the mean of epmf, we just "plug-in" $\hat{P}_X(\alpha)$ for $P_X(\alpha)$, and D for $\mathcal{X}$

and estimate $\mu = E[X]$ as
$$\hat{\mu} = \widehat{E[X]} = \sum_{a \in D} a\, \hat{P}_X(a).$$

- Such an estimator is called a "<u>plug-in estimator</u>"
- We just showed that plug-in estimator of mean of $X \sim P_X(\alpha)$ is the "sample mean" $= \overline{X}_n$.
- In general, if $\hat{\theta}$ is a plug-in estimator of θ, a plug-in estimator of $g(\theta)$ is $g(\hat{\theta})$.

<u>EX:</u> $\hat{\theta}$ is a plugin estimator of θ, so plug-in estimator of $\sin^2(\theta) + \cos(5\theta)$ is
$$\sin^2(\hat{\theta}) + \cos(5\hat{\theta})$$

Suppose $\hat{\theta}$ is an estimator of θ.
Q: How to assess $\hat{\theta}$? or if $\tilde{\theta}$ is another estimator of θ, how does $\hat{\theta}$ compare to $\tilde{\theta}$?

Properties of Estimators

Ideally (i.e. desirable)

1. $\text{bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$ — $\text{bias}(\hat{\theta}) = 0$ or at least $\text{bias}(\hat{\theta}) \to 0$ as $n \to \infty$
   - smaller bias is better

   If $\text{bias}(\hat{\theta}) = 0$ then $E[\hat{\theta}] = \theta \Rightarrow \hat{\theta}$ is an "unbiased estimator" of θ

2. standard error
   $se(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$.
   - smaller std error is better (more precise estimator)
   - ideally $se(\hat{\theta}) \to 0$ as $n \to \infty$
   - easier to work with $Var(\hat{\theta})$ than $s.e(\hat{\theta})$.

3. $MSE(\hat{\theta})$: mean squared error
   $= E[(\hat{\theta} - \theta)^2]$
   - ideally, $MSE(\hat{\theta}) \to 0$ as $n \to \infty$ (qm convergence)
   - smaller MSE is better (estimator closer to θ "on average")

- $MSE(\hat{\theta}) = Var(\hat{\theta}) + (\text{bias}(\hat{\theta}))^2$

Pf. $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$
$= E[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2]$
$= E[(\hat{\theta} - E(\hat{\theta}))^2] + (E(\hat{\theta}) - \theta)^2$ (cross product term has 0 expectation)
$= Var(\hat{\theta}) + (\text{bias}(\hat{\theta}))^2$

<u>Ex:</u> $\underset{\text{epmf}}{\hat{P}_X(a)} = \dfrac{\sum_{i=1}^{n} I(X_i=a)}{n}$ is an estimator of pmf $P_X(a)$.

1. $\text{bias}(\hat{P}_X(a)) = E[\hat{P}_X(a)] - P_X(a)$

$= P_X(a) - P_X(a) = 0$ (unbiased)

2. $Var(\hat{P}_X(a)) = Var\left(\dfrac{\sum_{i=1}^{n} I(X_i=a)}{n}\right)$

$= \dfrac{1}{n^2} Var\left(\sum_{i=1}^{n} I(X_i=a)\right)$

$\underset{=}{LOVI} \dfrac{1}{n^2} \sum_{i=1}^{n} Var(I(X_i=a))$

(Recall $Var(I(X\in A)) = P(X\in A)(1-P(X\in A))$)

$= \dfrac{1}{n^2} \sum_{i=1}^{n} P(X_i=a)(1-P(X_i=a))$

$\overset{ind}{=} \dfrac{1}{n^2} \sum_{i=1}^{n} P_X(a)(1-P_X(a))$

$= \dfrac{1}{n^{\cancel{2}}} \cancel{n} P_X(a)(1-P_X(a))$

$= \dfrac{P_X(a)(1-P_X(a))}{n}$

$\Rightarrow se(\hat{p}_X(a)) = \sqrt{\dfrac{P(a)(1-P_X(a))}{n}}$

Note $Var(\hat{P}_X(a)) = \dfrac{P_X(a)(1-P_X(a))}{n} \longrightarrow 0$ as $n\to\infty$

(i.e. $se(\hat{P}_X(a)) \to 0$ as $n\to\infty$.)

3) $mse(\hat{P}_X(a)) = Var(\hat{P}_X(a)) + \cancel{\text{bias}(\hat{P}_X(a))^2} \to 0$

$= Var(\hat{P}_X(a)) \longrightarrow 0$ as $n\to\infty$

<u>Ex:</u> Do the same for $\hat{\mu} = \frac{\sum X_i}{n}$.

1. $E[\hat{\mu}] = E\left[\frac{\sum X_i}{n}\right] = \cdots = \mu \Rightarrow \text{bias}(\hat{\mu}) = E[\hat{\mu}] - \mu =$
$= \mu - \mu = 0.$
(unbiased)

2. $Var(\hat{\mu}) = Var\left(\frac{\sum X_i}{n}\right) = \cdots = \frac{Var(X_1)}{n} = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$

3. $MSE(\hat{\mu}) = Var(\hat{\mu}) + (\text{bias}(\hat{\mu}))^2 = Var(\hat{\mu}) \to 0$ as $n \to \infty$

<u>Consistency</u> (of estimators) An estimator $\hat{\theta}$ of the parameter or quantity θ is called <u>consistent</u> if $\hat{\theta} \xrightarrow{P} \theta$ as $n \to \infty$.

<u>Interpretation:</u> $\hat{\theta}$ is close to θ wp ≈ 1 as $n \to \infty$

<u>implication:</u> we can replace θ with $\hat{\theta}$ for large n
(i.e. $\theta \approx \hat{\theta}$)
i.e. $\hat{\theta}$ is a "good estimator" of θ

<u>Thm:</u> If $MSE(\hat{\theta}) \to 0$ as $n \to \infty$, then $\hat{\theta}$ is a consistent estimator of θ (i.e. $\hat{\theta} \xrightarrow{P} \theta$)
($qm \Rightarrow P$)

<u>Note:</u> "$MSE(\hat{\theta}) \to 0$ is equivalent to
"$\text{bias}(\hat{\theta}) \to 0$ & $Var(\hat{\theta}) \to 0$" as $n \to \infty$.

<u>Ex:</u> Since $MSE(\hat{P}_X(a)) \to 0$ as $n \to \infty$,
$\hat{P}_X(a)$ is a <u>consistent estimator</u> of $P_X(a)$

<u>Exer:</u> Show that $\hat{\mu} = \frac{\sum X_i}{n}$ is a consistent estimator of μ.
($\mu = E[X]$)

<u>Notes</u> (1) We can also estimate ecdf for a discrete distr as

$$\hat{F}_X(a) = \hat{P}(X \le a) = \frac{\sum_{i=1}^{n} I(X_i \le a)}{n}$$

- But as mentioned before epmf is easier and simpler to work with.

(2) $\hat{P}_X(a) = \frac{\sum I(X_i = a)}{n} \Rightarrow \underbrace{n\hat{P}_X(a)}_{RV} = \sum_{i=1}^{n} I(X_i = a)$

Since $I(X_i = a) \overset{iid}{\sim}$ Bernoulli(p) where

$p = P(X_i = a) = P_X(a)$

we have $n\hat{P}_X(a) \sim \text{Binomial}(n, p = P_X(a))$

So, we can derive mean, variance etc for $\hat{P}_X(a)$ using Binomial distr. properties.

<u>Ex:</u> (1) $E[n\hat{P}_X(a)] = nP_X(a)$

$\Rightarrow \cancel{n} E[\hat{P}_X(a)] = \cancel{n} P_X(a)$

$\Rightarrow E[\hat{P}_X(a)] = P_X(a)$

(2) $Var(n\hat{P}_X(a)) = nP_X(a)(1 - P_X(a))$

$\Rightarrow n^2 Var(\hat{P}_X(a)) = nP_X(a)(1 - P_X(a))$

$\Rightarrow Var(\hat{P}_X(a)) = \frac{P_X(a)(1 - P_X(a))}{n}$

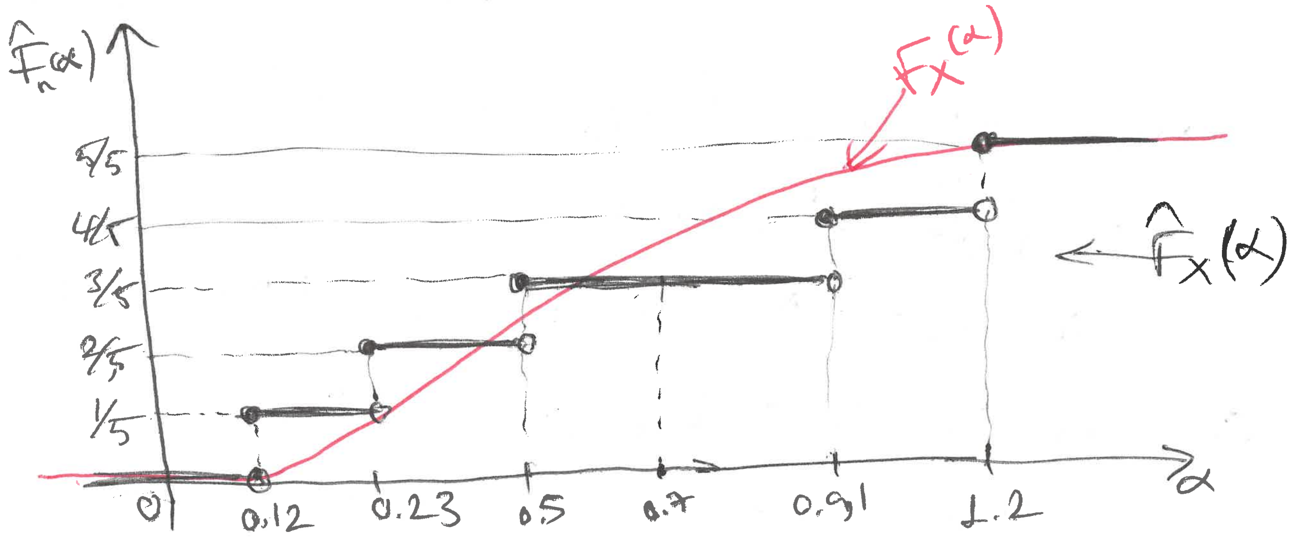
Nonparametric Inference for Continuous Distributions

Suppose the data $D = \{X_1, X_2, \dots, X_n\}$ where $X_i \overset{iid}{\sim} F$,
F is continuous

Qu: Can we estimate the pdf f_X as
$$\hat{f}_X(a) = \frac{\sum_{i=1}^{n} I(X_i = a)}{n} \text{ ? (No, why not?)}$$

But we can estimate the cdf $F_X(a)$ with
$$\hat{F}_X(a) = \frac{\sum_{i=1}^{n} I(X_i \le a)}{n} \leftarrow \text{ecdf}$$

Ex: $D = \{0.91, 0.23, 0.5, 0.12, 1.2\}$ $\hat{F}_X(a) = ?$ $\hat{F}_X(0.7) = ?$



Here, ecdf is "estimating" the cdf ← cts
has jumps

Note: plug-in estimators for cts distributions are obtained by replacing $F_X(a)$ with $\hat{F}_X(a)$.
Ex: $\mu = \int x f_X(x)\, dx = \int x F_X'(x)\, dx \Rightarrow \hat{\mu} = \sum x \hat{p}_X(x)$ "subtle"
$F_X(x) \approx \hat{F}_X(x) \equiv$ data ecdf (discrete) } so $E(X) \approx \sum_{x \in D} x \hat{p}_X(x) \leftarrow$ pmf for ecdf

<u>Properties</u> for a given $a \in \mathbb{R}$

1) $E[\hat{F}_X(a)] = ?$

$$\hat{F}_X(a) = \frac{\sum I(X_i \le a)}{n} \Rightarrow \underbrace{n\hat{F}_X(a)}_{RV} = \sum_{i=1}^{n} I(X_i \le a)$$

Since $I(X_i \le a) \sim \text{Bernoulli}(p)$ where
$p = P(X_i \le a) = F_X(a)$

$$n\hat{F}_X(a) = \sum_{i=1}^{n} I(X_i \le a) \sim \text{Binomial}(n, \underset{\text{prob.}}{F_X(a)})$$

So, $E[n\hat{F}_X(a)] = nF_X(a)$

$\Rightarrow E[\hat{F}_X(a)] = F_X(a)$ (unbiased) ✓

2) $Var(n\hat{F}_X(a)) = nF_X(a)(1 - F_X(a))$

$$\Rightarrow Var(\hat{F}_X(a)) = \frac{F_X(a)(1 - F_X(a))}{n} \xrightarrow[n\to\infty]{} 0 \text{ as}$$

3) 1) & 2) $\Rightarrow MSE(\hat{F}_X(a)) \to 0$ as $n \to \infty$

So $\hat{F}_X(a) \xrightarrow{P} F_X(a)$

i.e. $\hat{F}_X(a)$ is a consistent estimator of $F_X(a)$

<u>Glivenko-Cantelli Thm</u>

$$\sup_a |\hat{F}_X(a) - F_X(a)| \to 0 \text{ as } n \to \infty$$

"maximum" difference over a

$\Rightarrow \hat{F}_X(a) \xrightarrow{P} F_X(a)$ (consistency)

Drovetzky-Kifer-Wolfowitz (DKW) Th'm:

$$P\left(\sup_a |\hat{F}_X(a) - F_X(a)| > \varepsilon\right) \le 2e^{-2n\varepsilon^2} = \frac{2}{e^{2n\varepsilon^2}}$$

for all $\varepsilon > 0$.

Ex: Let $\theta = P(X \in (a,b))$ for cts $X \sim F_X$.

$D = \{X_1, X_2, \ldots, X_n\}$ with $X_i \overset{iid}{\sim} F_X$.

One can estimate $\theta = P(X \in (a,b)) = \int_a^b f_X(x)\,dx$ with

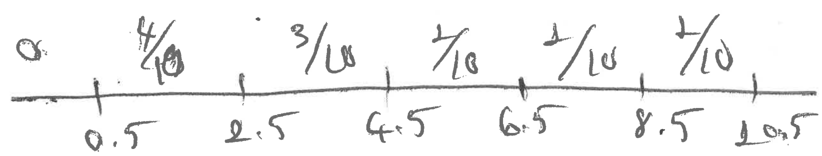
$$\hat{\theta} = \frac{\sum_{i=1}^{n} I(X_i \in (a,b))}{n}.$$

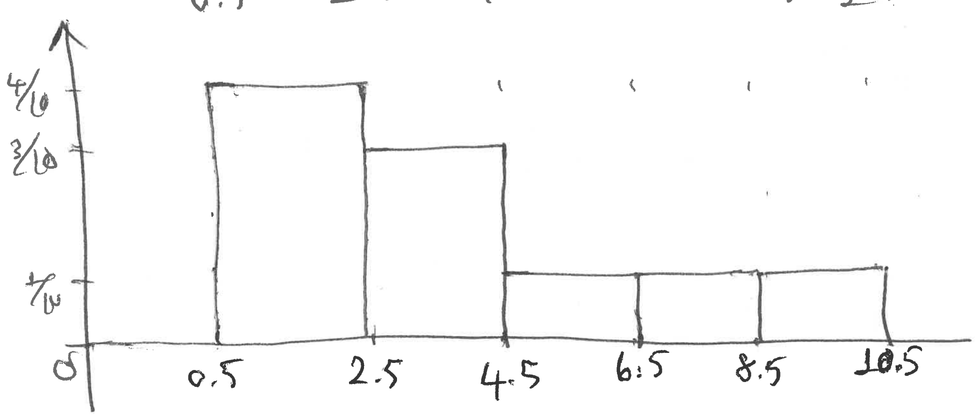
Notice also that $\theta = F_X(b) - F_X(a)$.

Qu: How to estimate the curve of pdf?

Histogram Estimator:

$D = \{3, 3.2, 1.8, 3.5, 2, 1.5, 7, 9, 2, 6\}$

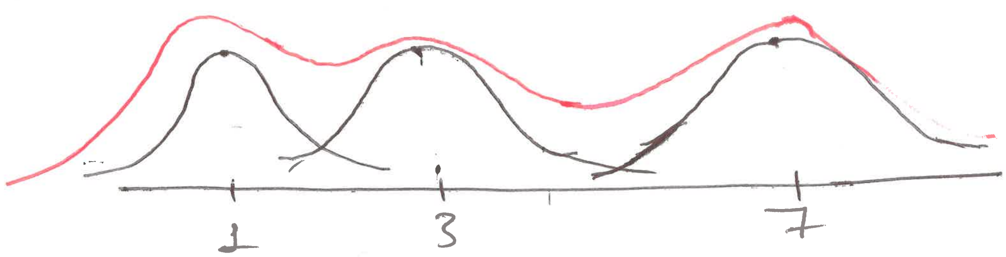




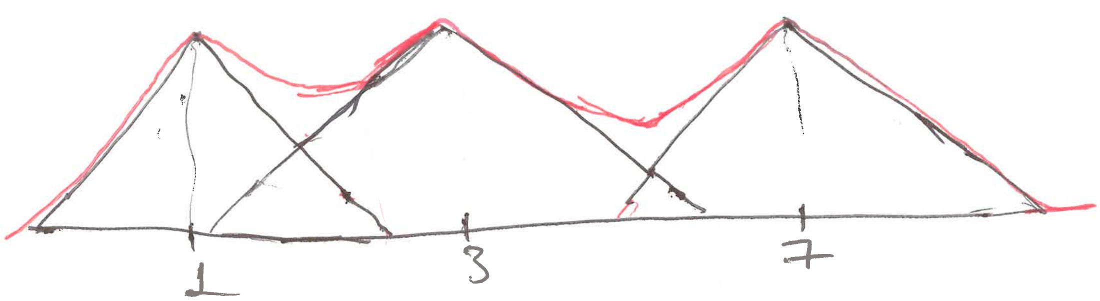
- **Kernel density estimator:**

$D = \{3, 1, 7\}$ For any $a \in \mathbb{R}$,

kernel $\equiv$ Normal $\hat{f}_X(a) = \sum_{i=1}^{n} \phi_i(a)$ ϕ_i is normal pdf centered at x_i with std dev $= \sigma$.



kernel $\equiv$ Triangular



Statistical Functional

Any function of $P_X(a)$, $f_X(a)$ or $F_X(a)$

Ex: For discrete X $E[X] = \sum_{x \in \Omega} x\, P_X(x)$ is a function of $P_X(x)$

For cts X
$E[X] = \int x f_X(x)\, dx$ "   "   "   " $f_X(x)$

$$E[X^k] = \sum_{x \in \Omega} x^k p_X(x), \text{ for discrete } X,$$

is a fn of $p_X(x)$

$$E[X^k] = \int x^k f_X(x)\,dx \text{ for cts } X.$$

So, $Var(X) = E[X^2] - (E[X])^2$

$$Var(X) = \sum x^2 p_X(x) - \left(\sum x p_X(x)\right)^2 \text{ for discrete } X$$

$$Var(X) = \int x^2 f_X(x)\,dx - \left(\int x f_X(x)\right)^2 \quad \text{" cts } X$$

Ex: For discrete X, with $E[X] = \mu$ & $Var(X) = \sigma^2$, consider $\theta = \sin^3(\mu) + \cos(\sigma^2)$

Then $\theta = \sin^3\left(\sum_{x \in \Omega} x p_X(x)\right) + \cos\left(\sum_{x \in \Omega} x^2 p_X(x) - \left(\sum x p_X(x)\right)^2\right)$

Plug-in Estimation:

- For discrete X, if $\theta = \theta(p_X)$, then an (plug-in) estimate of θ is $\hat{\theta} = \theta(\hat{p}_X)$

- For cts, X, if $\theta = \theta(F_X)$, then the plug-in estimate of θ is
$$\hat{\theta} = \theta(\hat{F}_X)$$

Ex: For discrete $D = \{x_1, x_2, \ldots, x_n\}$, $X_i \overset{iid}{\sim} p_X$, with $E[X_1] = \mu$, $Var(X_1) = \sigma^2$

Plug-in estimate of $\theta = \sin^3(\mu) + \cos(\sigma^2)$

$$\hat{\theta} = \sin^3\Big(\sum_{x\in D} x\,\hat{p}_X(x)\Big) + \cos\Big(\sum_{x\in D} x^2\hat{p}_X(x) - \big(\sum_{x\in D} x\,\hat{p}_X(x)\big)^2\Big)$$

i.e. $\hat{\theta} = \sin^3(\hat{\mu}) + \cos(\hat{\sigma}^2)$

where $\hat{\mu} = \sum_{x\in D} x\,\hat{p}_X(x)$ plug-in estimator of $\mu = E[X]$

& $\hat{\sigma}^2 = \sum_{x\in D} x^2\hat{p}_X(x) - \big(\sum_{x\in D} x\,\hat{p}_X(x)\big)^2$ plug-in estimator of $\sigma^2 = E[X^2] - (E[X])^2$

Ex: $D = \{3, 1, 2\}$, so plug-in estimator of μ & σ^2

are $\hat{\mu} = \sum x_i \frac{1}{n} = \frac{3+1+2}{3} = 2$

and $\hat{\sigma}^2 = \sum x_i^2 \cdot \frac{1}{n} - 2^2 = \frac{3^2+1^2+2^2}{3} - 4 = \frac{14}{3} - 4 = \frac{2}{3}$

then plug-in estimator of θ is

$$\hat{\theta} = \sin^3(2) + \cos\left(\tfrac{2}{3}\right) = 1.6127$$

Plug-in Estimation

Goal: provide a numerical estimate of a quantity or parameter, θ, in terms of data D

1) $\theta = \theta(p_X)$

2) $\hat{\theta} = \theta(\hat{p}_X)$

   (a) Replace Ω with D

   (b) " $p_X(a)$ with $\hat{p}_X(a) = \begin{cases} \frac{1}{n} & \text{if } a \in D \\ 0 & \text{o.w.} \end{cases}$

   epmf

Ex: Plug-in estimator of σ^2 is related to sample variance as $\hat{\sigma}^2 = \frac{n-1}{n} s^2$ where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Note: For cts case, replace $F_X(a)$ with $\hat{F}_X(a)$. Gives same estimates as in discrete case.