Hands on session 1: Statistics 101

Exercise 1: Maximum likelihood estimators

$$\text{MLE}: \hat{\theta} = \underset{\theta}{\operatorname{argmax}} \left\{ \mathbb{P}_\theta (X) \right\}$$

data

$$L^2 \text{ risk}: R_2(\hat{\theta}) = \mathbb{E}_{P_\theta}\left((\theta - \hat{\theta})^2\right)$$

1. $\mathcal{X} = \mathbb{R}$, $Q_\theta = \mathcal{N}(\theta, \sigma^2)$ σ known $X = (X_1, \dots, X_n)$ data

$$\mathcal{L}_\theta(x) = \mathbb{P}_\theta(X) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i) \propto \prod_{i=1}^{n} e^{-\frac{(x_i-\theta)^2}{4\sigma^2}}$$

$$\ell_\theta(x) = \text{cte} + \sum_{i=1}^{n} - \frac{(x_i - \theta)^2}{4\sigma^2}$$

$\ell_\theta(x)$ is concave in θ. In order to maximize it, we need to impose the first order condition $\nabla_\theta \ell_\theta(x) = 0$

$$\sum_{i=1}^{n} (x_i - \theta) = 0 \text{ iff } \theta = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\boxed{\text{MLE}: \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i}$$

$E(\hat{\theta}) = \theta$: without bias

SLLN: $\hat{\theta} \xrightarrow{as} \theta$

Same as Moment estimator

2. Now, σ is unknown, $\mathcal{L}_{\mu,\sigma}(x) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2}$

$$\ell_{\mu,\sigma}(x) = \sum_{i=1}^{n} \left(\log\left(\frac{1}{\sigma}\right) - \frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2 \right)$$

First order condition: $\nabla_\mu \ell_{\mu,\sigma}(x) = 0$ iff $\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$ $\quad \boxed{\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i}$

$$\nabla_\sigma \ell_{\mu,\sigma}(x) = 0 \text{ iff } \sum_{i=1}^{n} \frac{-1}{\sigma} - \frac{1}{2} \frac{-(x_i - \mu)}{\sigma^2} 2\left(\frac{x_i - \mu}{\sigma}\right) = 0$$

$\partial_\sigma \ell_{\mu,\sigma}(x) = 0$ iff $\dfrac{\sum (x_i - \mu)^2}{\sigma^2} = n$

So, $\boxed{\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \hat{\mu})^2}$ Same as moment estimator

$\hat{\sigma}^2 = \frac{1}{n} \sum_i x_i^2 - \frac{1}{n^2} \sum_i \sum_j x_i x_j$

so $E(\hat{\sigma}^2) = \frac{1}{n} \sum_i E(x_i^2) - \frac{1}{n^2} \sum_i E(x_i^2) - \frac{1}{n^2} n(n-1)\mu^2$

$E(x_i^2) = E(((x_i - \mu) + \mu)^2) = \sigma^2 + \mu^2$

so $\boxed{E(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2}$ the estimator is biased

3 $X = (X_1, \dots, X_n)$ iid $X_i \sim \mathcal{B}(\theta)\ \forall i$ $\theta \in (0,1)$

$\mathcal{L}_\theta(x) = P_\theta(x) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}$

$\ell_\theta(x) = \log \mathcal{L}_\theta(x) = \sum_{i=1}^{n} x_i \log \theta + (1 - x_i) \log(1-\theta)$

$\frac{\partial}{\partial \theta} \ell_\theta(x) = \sum_{i=1}^{n} \frac{x_i}{\theta} + \frac{(x_i - 1)}{1-\theta}$

$\frac{\partial}{\partial \theta} \ell_\theta(x) = 0$ iff $\sum_{i=1}^{n} x_i - \theta \sum_{i=1}^{n} x_i - \theta \sum_{i=1}^{n} (1 - x_i)$ iff $\boxed{\underset{\hat{\theta}}{\theta} = \frac{1}{n} \sum_{i=1}^{n} x_i}$

Furthermore, $\frac{\partial^2}{\partial \theta^2} \ell_\theta(x) = \sum_{i=1}^{n} - \frac{x_i}{\theta^2} - \frac{(1-x_i)}{(1-\theta)^2} \leq 0$ always

Same as moment estimator

$\boxed{E(\hat{\theta}) = \theta}$ Unbiased

4 $X = (X_1, \dots, X_n)$ iid $X_i \sim U(0, \theta)\ \forall i$, $\theta > 0$

$\mathcal{L}_\theta(x) = P_\theta(x) = \prod_{i=1}^{n} \frac{1}{\theta^n} \mathbb{1}_{x_i \leq \theta}$ so, $\underset{\theta}{\operatorname{argmax}}\, \mathcal{L}_\theta(x) = \max_i (x_i)$

$\boxed{\hat{\theta} = \max_i (X_i)}$ Let's look at the law of $\hat{\theta}$. Dif from moment estimator

$$P(\hat{\theta} \leq t) = P(\max(X_1, \dots, X_n) \leq t) = P(X_1 \leq t, \dots, X_n \leq t)$$
$$= P(X_1 \leq t) \dots P(X_n \leq t) = P(X_1 \leq t)^n$$

$$\text{so } P(\hat{\theta} \leq t) = \begin{cases} 0 & \text{if } t \leq 0 \\ \left(\frac{t}{\theta}\right)^n & \text{if } t \in (0,\theta) \\ 1 & \text{if } t \geq \theta \end{cases}$$

So $\hat{\theta}$ has a density p_θ w.r.t Lebesgue measure and $p_\theta(\hat{\theta}) = \begin{cases} 0 & \text{if } \hat{\theta} \leq 0 \\ \frac{n\hat{\theta}^{n-1}}{\theta^n} & \text{if } \hat{\theta} \in (0,\theta) \\ 0 & \text{if } \hat{\theta} \geq \theta \end{cases}$

$$\underline{E(\hat{\theta})} = \int_0^\theta n \frac{\hat{\theta}^n}{\theta^n} \, d\hat{\theta} = \frac{n}{(n+1)\theta^n} \left[\hat{\theta}^{n+1}\right]_0^\theta = \underline{\frac{n}{n+1}\theta}$$

$$E(\hat{\theta}^2) = \int_0^\theta n \frac{\hat{\theta}^{n+1}}{\theta^n} \, d\hat{\theta} = \frac{n}{n+2}\theta^2$$

$$\underline{Var(\hat{\theta})} = E(\hat{\theta}^2) - E(\theta)^2 = \underline{\left[\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right]\theta^2}$$

5. $X = (X_1, \dots, X_n)$ iid $X_i \sim \mathcal{E}(\theta)$ U. $\theta > 0$

$$\mathcal{L}_\theta(X) = p_\theta(X) = \prod_{i=1}^n \theta e^{-\theta X_i}$$

$$\ell_\theta(X) = \sum_{i=1}^n \log\theta - \theta X_i \qquad \frac{\partial}{\partial\theta} \ell_\theta(X) = \sum_{i=1}^n \frac{1}{\theta} - X_i$$

$$\frac{\partial}{\partial\theta} \ell_\theta(X) = 0 \text{ iff } \theta = \frac{1}{\frac{1}{n}\sum_{i=1}^n X_i} \qquad \text{Furthermore, } \frac{\partial^2}{\partial\theta^2} \ell_\theta(X) = \frac{-n}{\theta^2} \leq 0$$

So $\boxed{\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{i=1}^n X_i}}$ Same as moment estimator

6 $X = (X_1, \dots, X_n)$ iid $X_i \sim \mathcal{L}(\theta)\ \forall i$ $\quad \theta \in \mathbb{R}$

$$f_\theta(x) = \exp(-|x-\theta|)/2$$

$$\mathcal{L}_\theta(x) = \prod_{i=1}^{n} \exp(-|X_i - \theta|)/2$$

$$\ell_\theta(x) = \sum_{i=1}^{n} -|X_i - \theta| \quad + \text{cte}$$

We want to minimize $\sum_{i=1}^{n} |X_i - \theta|$ convex function of θ $(g(\theta))$ ✓

FOC: $0 \in \underset{\substack{\uparrow \\ \text{subgradient}}}{\partial g(\theta)}$ $\quad \partial \sum_{i=1}^{n} |X_i - \theta| = \partial \sum_{X_i < \theta} |X_i - \theta| + \partial \sum_{X_i > \theta} |X_i - \theta| + \partial \sum_{X_i = \theta} |X_i - \theta|$

$$= |\{X_i : X_i < \theta\}| \{-1\} + |\{X_i : X_i > \theta\}| \{1\} + |\{X_i : X_i = \theta\}| [-1, 1]$$

So $0 \in \partial g(\text{median}(X_1, \dots, X_n))$

$$\boxed{\hat{\theta} = \text{median}(X_1, \dots, X_n)}$$ Diff from moment estimator.

Risk of 1

$$\boxed{\text{Bias Variance Tradeoff}: E_\theta\left((\theta-\hat\theta)^2\right) = E_\theta(\theta-\hat\theta)^2 + Var_\theta(\hat\theta)}$$

proof: $E_\theta\left((\theta-\hat\theta)^2\right) = E_\theta\left((\theta - E(\hat\theta) + E(\hat\theta) - \hat\theta)^2\right)$

$$= (\theta - E(\hat\theta))^2 + E\left((\hat\theta - E(\hat\theta))^2\right) + 2\underbrace{E\left((\theta - E(\hat\theta))(E(\hat\theta) - \hat\theta)\right)}_{0}$$

$\hat\theta = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\quad$ $\left.\begin{aligned} bias(\hat\theta) &= 0 \\ Var(\hat\theta) &= \frac{\sigma^2}{n}\end{aligned}\right\} \Rightarrow R_\theta(\hat\theta) = \frac{\sigma^2}{n}$

Risk of 3

$$R_\theta(\hat\theta) = bias(\hat\theta)^2 + Var_\theta(\hat\theta) = 0 + \frac{1}{n^2} n\theta(1-\theta) = \frac{\theta(1-\theta)}{n}$$

$$\boxed{R_\theta(\hat\theta) = \frac{\theta(1-\theta)}{n}}$$

Risk of 4

$$R_\theta(\hat\theta) = bias(\hat\theta)^2 + Var(\hat\theta) = \frac{n^2}{(n+1)^2}\theta^2 + \left[\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right]\theta^2$$

$$\boxed{R_\theta(\hat\theta) = \frac{n}{n+2}\theta^2}$$

Exercise 2: Confidence intervals

Definition: θ is a param $\qquad \theta \in [f(\hat{\theta}), g(\hat{\theta})]$ with good P

Let I be an interval and $\alpha \in [0,1]$

We say that I is an CI for θ with level α if

$$P(\theta \in I) \geq \alpha$$

We say that I is an asymptotic CI for θ with level α if

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n \in \mathbb{N}, n \geq n_0 \Rightarrow P(\theta \in I) \geq \alpha - \varepsilon$$

1. $X = (X_1, \dots, X_n)$ $X_i \sim N(\theta, \sigma^2)\ \forall i$, σ known.

$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (MLE estimator, estimator of moments)

$E(\hat{\theta}) = \theta$, $Var(\hat{\theta}) = \frac{\sigma^2}{n}$ $\qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Bienaymé-Tchebychev: $P(|\bar{X} - E(\bar{X})| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2} \quad \forall \varepsilon > 0$

So, let $\alpha \in (0,1)$, $P\left(|\hat{\theta} - \theta| \geq \sqrt{\frac{\sigma^2}{n(1-\alpha)}}\right) \leq 1-\alpha$

So, $\left[\hat{\theta} - \sqrt{\frac{\sigma^2}{n(1-\alpha)}} ; \hat{\theta} + \sqrt{\frac{\sigma^2}{n(1-\alpha)}}\right]$ is a CI for θ of level α

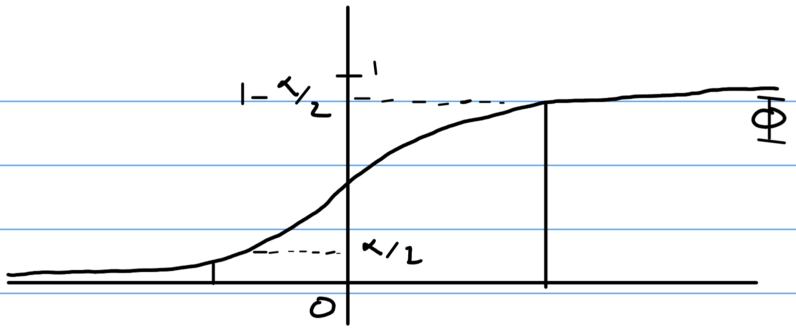
Central Limit Theorem: $X_1, \dots, X_n$ iid RVs st $E(X_1) = \mu$ $Var(X_1) = \sigma^2$

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

Then $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$

Here: $\frac{\sqrt{n}}{\sigma}(\hat{\theta} - \theta) \xrightarrow{d} N(0,1)$ $\qquad X_n \xrightarrow{d} X$ then $E(f(X_n)) \to E(f(X))$ f nice

In particular, $P\left(\frac{\sqrt{n}}{\sigma}|\hat{\theta} - \theta| \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) \xrightarrow[n \to +\infty]{} \alpha$ $\forall \alpha \in (0,1)$ where Φ CDF of $N(0,1)$



So, $\left[\hat{\theta} \pm \phi^{-1}(\alpha/2)\frac{\sigma}{\sqrt{n}}\right]$ is an asymptotic CI for θ of level $1-\alpha$

Sum of gaussians: If X and Y are two gaussian RV and are independent,
Then $X+Y \sim \mathcal{N}(E(X)+E(Y), Var(X)+Var(Y))$

Here, $\hat{\theta} \sim \mathcal{N}(\theta, \frac{\sigma^2}{n})$ and $\frac{\sqrt{n}}{\sigma}(\hat{\theta}-\theta) \sim \mathcal{N}(0,1)$
We obtain the same CI as with CLT but it's not asymptotic

2 $X=(X_1, \dots, X_n)$ iid $X_i \sim B(\theta)\ \forall i$. $\quad \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$E(\hat{\theta}) = \theta$, $Var(\hat{\theta}) = \frac{\theta(1-\theta)}{n}$

BT $\Rightarrow \left[\hat{\theta} \pm \sqrt{\frac{\theta(1-\theta)}{n(1-\alpha)}}\right]$ is a CI for θ. Problem: It depends on θ!

CLT $\Rightarrow \sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} \mathcal{N}(0, \theta(1-\theta))$ Problem: Still depends on θ

Strong law of large numbers: $\hat{\theta} \xrightarrow{a.s.} \theta$ so $\hat{\theta} \xrightarrow{P} \theta$

Lemma (Slutsky) If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c \in \mathbb{R}$, $(X_n, Y_n) \xrightarrow{d} (X, c)$

So $\left(\frac{\sqrt{n}(\hat{\theta}-\theta)}{\sqrt{\theta(1-\theta)}}, \hat{\theta}\right) \xrightarrow{d} (\mathcal{N}(0,1), \theta)$

So, $P\left(\frac{\sqrt{n}\,|\hat{\theta}-\theta|}{\sqrt{\hat{\theta}(1-\hat{\theta})}} \geq \Phi^{-1}(1-\alpha/2)\right) \xrightarrow[n\to+\infty]{} \alpha$

So $\left[\hat{\theta} \pm \frac{\phi^{-1}(1-\alpha/2)}{\sqrt{n}}\sqrt{\hat{\theta}(1-\hat{\theta})}\right]$ is a CI of level $1-\alpha$ for θ

Hoeffding Inequality $Z_1, \dots, Z_n$ independent real valued RVs
$\exists a, b$ st $\forall i$, $a \leq Z_i \leq b$ a.s.

Then, $P\left(\left|\sum_{i=1}^{n} (Z_i - E(Z_i))\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{n(b-a)^2}\right) \quad \forall t > 0$

So here it gives: $P\left(n|\hat{\theta} - \theta| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{n}\right) \quad \forall t > 0$

Let $\alpha \in (0,1)$, $2\exp\left(-\frac{2t^2}{n}\right) \leq 1-\alpha$ iff $t \geq \sqrt{\log\left(\frac{2}{1-\alpha}\right)\frac{n}{2}}$

So, $\left[\hat{\theta} \pm \sqrt{\log\left(\frac{2}{1-\alpha}\right)\frac{1}{2n}}\right]$ is a CI of level α for θ

3. $X = (X_1, \dots, X_n)$, $X_i \sim \mathcal{U}([0,\theta])$ $\forall i$ $\quad \hat{\theta} = \max_i (X_i)$

$$\forall t \in \mathbb{R}, \; P(\hat{\theta} \leq t) = \begin{cases} 0 & \text{if } t \leq 0 \\ \left(\frac{t}{\theta}\right)^n & \text{if } 0 < t < \theta \\ 1 & \text{if } t \geq \theta \end{cases}$$

$\left(\frac{t}{\theta}\right)^n = \alpha$ iff $t = \theta \sqrt[n]{\alpha}$

So $P\left(\hat{\theta} \leq \theta\sqrt[n]{\alpha}\right) = \alpha$

So $\left[\frac{\hat{\theta}}{\sqrt[n]{\alpha}}, +\infty\right)$ is a CI of level α for θ

4. $X = (X_1, \dots, X_n)$, $\forall i$, $X_i \sim \mathcal{L}(\theta)$

* Moment estimator: $\hat{\theta} = \frac{1}{n}\sum_i X_i$ BC, CLT: OK

* MLE estimator: $\hat{\theta} = \text{median}\{X_i : i\}$

Let's compute $P(\hat{\theta} < t)$.

Without loss of generality $\mu = 0$. Furthermore $\hat{\theta}$ is symmetric wrt 0 so we only have to look at $t < 0$,

- case $n = 2p+1$:

$\hat{\theta} < t$ iff at least $p+1$ of the X_i's are $< t$

$$P(\hat{\theta} < t) = \sum_{i=p+1}^{2p+1} \binom{2p+1}{i} \left(\frac{1}{2} e^{t}\right)^{i} \left(1 - \frac{1}{2} e^{t}\right)^{2p+1-i}$$

- case $n = 2p$

$$P(\hat{\theta} < t) = \sum_{i=p+1}^{2p} \binom{2p}{i} \left(\frac{1}{2} e^{t}\right)^{i} \left(1 - \frac{1}{2} e^{t}\right)^{2p-i}$$

Can be solved with a computer

Hands on 1: Mean or Median?

$M_1 = (\mathbb{R}^n, \{\mathcal{N}(\mu, 1)^{\otimes n} : \mu \in \mathbb{R}\})$

- $\hat{\mu}_n = \frac{1}{n} \sum_i X_i$

$$E(\hat{\mu}_n) = \frac{1}{n} \sum_i E(X_i) = \frac{1}{n} \sum_i \mu = \mu$$

μ_n is unbiased

SLLN: $\hat{\mu}_n \xrightarrow{a.s.} \mu$ $\quad$ $\hat{\mu}_n$ is strongly consistent

- $\tilde{\mu}_n = \text{median}\{X_i\}_i$

Median Theorem: $X_1, \dots, X_{2m+1}$ iid samples with density f (cdf F)

$Y_1, \dots, Y_{2m+1} = \mathrm{sort}(X_1, \dots, X_{2m+1})$

$\tilde{X} = \mathrm{median}(X_1, \dots, X_{2m+1}) = Y_{m+1}$

$\tilde{\mu} = F^{-1}(1/2)$

if $\begin{cases} f(\tilde{\mu}) \neq 0 \\ f_{|N(\tilde{\mu})} \in C^1(N(\tilde{\mu})) \end{cases}$ then $\tilde{X} \approx \mathcal{N}\left(\tilde{\mu}, \frac{1}{8 f(\tilde{\mu})^2 m}\right)$

Here, up to a translation, let's suppose that $\mu = 0$

when $n = 2m+1$, $\forall x$, $P(\tilde{\mu}_n = x) = P(\tilde{\mu}_n = -x)$

indeed, let $f : X \mapsto -X$

then, X is bijective and

$\forall X$, $P(X) = P(f(X))$ and $\tilde{\mu}_n(f(X)) = -\tilde{\mu}_n(X)$

so, $P(\tilde{\mu}_n = x) = \int \mathbb{1}_{\tilde{\mu}_n(X) = x} P(dX) = \int \mathbb{1}_{\mu_n(f(X)) = -x} P(d(f(X)))$

$= \int \mathbb{1}_{\tilde{\mu}_n(X)} P(dX) = P(\tilde{\mu}_n = x)$

So, $E(\tilde{\mu}_n) = \int \tilde{\mu}_n P(d\tilde{\mu}_n) = \mu - \int (\tilde{\mu}_n - \mu) P(d\tilde{\mu}_n) = \mu$

since $\tilde{\mu}_n$ is symmetrical wrt μ.

$\tilde{\mu}_n$ is consistent

* $M_2 = (\mathbb{R}^n, \{\mathcal{L}(\mu)^{\otimes n}; \mu \in \mathbb{R}\})$

Some results