Instructions

- The homework is due on **Friday 3/17 at 5pm ET**.
- No extension will be provided, unless for serious documented reasons.
- **Despite having two weeks for this HW, better start early than late!**
- Study the material taught in class, and feel free to do so in small groups, but the solutions should be a product of your own work.
- This is not a multiple choice homework; reasoning, and mathematical proofs are required before giving your final answer.

1 MLE and MoM [30 points]

1. (5pts) Let $X_1, \ldots, X_n$ be iid Bernoulli(p) samples. In class we sketched the proof that the maximum likelihood estimator of p is $p_{MLE} = \frac{\sum_{i=1}^n x_i}{n}$. Write a complete proof.
2. (5pts) Assume you have a prior p that is a $beta(\alpha, \beta)$ and let $Y = \sum_{i=1}^n X_i$. Write down the joint distribution of Y, p.
3. (5pts) Let $X_1, \ldots, X_n$ be iid $\mathcal{N}(\mu, \sigma^2)$ where both μ, σ are unknown. What are the MLEs for μ, σ^2?
4. (5pts) Let $X_1, \ldots, X_n$ be iid $Exponential(\lambda)$. Find the method of moments estimator for λ.
5. (5pts+5pts) Let $X_1, \ldots, X_n$ be iid $\beta(\theta, 1)$. Find (a) the MLE and the (b) MoM estimator for θ.

   Hint: You may use the fact that the expected value of a $beta(\alpha, \beta)$ is equal to $\frac{\alpha}{\alpha+\beta}$ without proof.

Solution:

1.1

∵ Bernoulli

$$\therefore f_X(x;\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$\log L(\theta|x) = \log\left(\prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}\right)$$

$$= \sum_{i=1}^{n} \log\left(\theta^{x_i}(1-\theta)^{1-x_i}\right)$$

$$= \sum_{i=1}^{n} x_i \cdot \log\theta + (1-x_i)\log(1-\theta)$$

$$= \log\theta \sum_{i=1}^{n} x_i + \log(1-\theta)(n-\sum_{i=1}^{n} x_i)$$

$$\hat{\theta} = \arg\max_{\theta}\left\{\log\theta \sum_{i=1}^{n} x_i + \log(1-\theta)(n-\sum_{i=1}^{n} x_i)\right\}$$

$$\frac{d}{d\theta}\left\{\log\theta \sum_{i=1}^{n} x_i + \log(1-\theta)(n-\sum_{i=1}^{n} x_i)\right\} = 0$$

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

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

$$(1-\theta)\sum_{i=1}^{n} x_i = n\cdot\theta - \theta\cdot\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i - \theta\sum_{i=1}^{n} x_i = n\theta - \theta\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i = n\theta$$

$$\hat{\theta}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}$$

To show that $\hat{\theta}$ is the maximum, we need to show the graph concaves down, which means $f'' < 0$.

$$f' = \frac{1}{\theta}\sum_{i=1}^{n} x_i + \frac{(-1)}{1-\theta}(n-\sum_{i=1}^{n} x_i)$$

$$f'' = \theta^{-1}\sum_{i=1}^{n} x_i + \frac{-1}{1-\theta}(n-\sum_{i=1}^{n} x_i)$$

$$-1\,\theta^{-2}\sum x_i + (-1)\frac{1}{(1-\theta)^2}(n-\sum x)$$

$$\Rightarrow \theta \in [0,1]$$

$$\therefore \theta^{-2} \geq 0, \quad \frac{1}{(1-\theta)^2} > 0 \quad \text{OR} \quad \theta^{-2} > 0, \quad \frac{1}{(1-\theta)^2} \geq 0 \quad \therefore (-1)\cdot \text{positive} = \text{negative}$$

$$(-1)\theta^{-2}\sum x_i + (-1)\frac{1}{(1-\theta)^2}(n-\sum x) < 0$$

$\therefore \hat{\theta}$ is the maximum.

1.2

$\because$ Beta distribution is Binomial's conjugate prior.

prior $beta(\alpha, \beta) = \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} p^{\alpha-1}(1-p)^{\beta-1}$

likelihood $bin(n, p) = \binom{n}{x} p^{x}(1-p)^{n-x}$

posterior $beta(\alpha+x, \beta+n-x)$

$$= \frac{(\alpha+\beta+n-1)!}{(\alpha+x-1)!(\beta+n-x-1)!} \cdot p^{\alpha+x-1}(1-p)^{\beta+n-x-1}$$

1.3

$\because$ iid $N(\mu, \sigma^2)$

$$\therefore f(X; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$\log L(\mu, \sigma^2 | X) = \log\left\{\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}\right\}$$

$$= \sum_{i=1}^{n} \log\left\{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}\right\}$$

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

$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\hat{\mu} = \frac{d}{d\mu}\left\{\underbrace{-\frac{n}{2}\log(2\pi\sigma^2)}_{\text{Constant}} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$= -\frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n} \frac{d}{d\mu}(x_i-\mu)^2$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{n} -1\cdot 2(x_i-\mu)$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\mu = 0$$

$$\sum_{i=1}^{n} x_i = n\mu$$

$$\boxed{\hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

$$\hat{\sigma}^2 = \frac{d}{d\sigma}\left\{-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

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

$$= -\frac{n}{2}\cdot\frac{1}{\sigma^2} - \frac{1}{2}\cdot(-1)\frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= \frac{-n}{2\sigma^2} + \frac{1}{2}\frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= \frac{1}{2\sigma^2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - n\right]$$

$$\frac{1}{2\sigma^2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu^2) - n\right] = 0$$

$$\frac{1}{2}\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu^2) - \frac{n}{2\sigma^2} = 0$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu^2) = n$$

$$\boxed{\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu^2)}$$

1.4

$$E[\text{exponential}(\lambda)] = \frac{1}{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$E(X) = \frac{1}{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\bar{X} = \frac{1}{\lambda}$$

$$\hat{\lambda}_{MOM} = \frac{1}{\bar{X}}$$

1.5

a) $$f_X(x;\theta) = \prod_{i=1}^{n} \frac{x_i^{\theta-1}(1-x_i)^{1-1}}{B(\theta,1)} = \prod_{i=1}^{n} \frac{x_i^{\theta-1}}{B(\theta,1)}$$

$$\log L(\theta|x) = \log\left(\prod_{i=1}^{n} \frac{x_i^{\theta-1}}{B(\theta,1)}\right) = \sum_{i=1}^{n} \log \frac{x_i^{\theta-1}}{B(\theta,1)} = \sum_{i=1}^{n} \log X_i^{\theta-1} \cdot \frac{\Gamma(\theta+1)}{\Gamma(\theta)\cdot\Gamma(1)}$$

$$\frac{d}{d\theta}\left\{\sum_{i=1}^{n} \log X_i^{\theta-1} \cdot \frac{\Gamma(\theta+1)}{\Gamma(\theta)\Gamma(1)}\right\}$$

$$= \frac{d}{d\theta}\left\{(\theta-1)\sum_{i=1}^{n} \log x_i \cdot \frac{\Gamma(\theta+1)}{\Gamma(\theta)\Gamma(1)}\right\} = 0$$

b) $$E(X) = \frac{\theta}{\theta+1}$$

$$\bar{x} = \frac{\theta}{\theta+1}$$

$$\theta\bar{x} + \bar{x} = \theta$$

$$\hat{\theta} = \frac{\bar{x}}{1-\bar{x}}$$

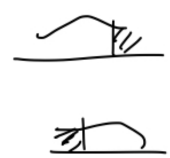
2 To Handshake or Not? [20 points]

Suppose n people walk into a party. Due to covid-19, each pair $\{i, j\}$ shakes hands with probability only $\frac{1}{10}$. Prove that with probability that tends to 1 as $n \to +\infty$ **every** person from that party shook hands in the range $[0.95\frac{n}{10}, 1.05\frac{n}{10}]$.

Solution:

2. $Bin(n-1, \frac{1}{10})$ for each person.

Using chernoff bounds, we get



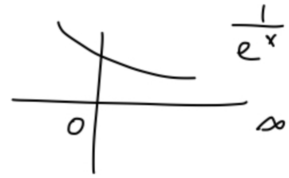
$$P(X \geq (1+0.05)\mu) \leq \exp\left(-\frac{0.05^2\mu}{3}\right)$$

$$P(X \leq (1-0.05)\mu) \leq \exp\left(-\frac{0.05^2\mu}{2}\right)$$

since $\mu = (n-1)\cdot\frac{1}{10}$, when $n \to \infty$, $\mu = \infty$.

$$\lim_{n\to\infty} \exp\left(-\frac{0.05^2\mu}{3}\right) + \exp\left(-\frac{0.05^2\mu}{2}\right)$$

$$= \lim_{n\to\infty} \frac{1}{e^{\infty}} \approx 0$$



Therefore, the probability of shaking hands in the range $(0, 0.95\mu)$ and $(1.05\mu, 1)$ is almost 0, and as $n \to \infty$, $n-1 \doteq n$.

So it is true that as $n \to +\infty$, every person from the party shook hands in the range $[0.95\frac{n}{10}, 1.05\frac{n}{10}]$

3 Mixture of Gaussians [25 points]

Let X, Y be two independent normal RVs, with means $\mu_x = 100, \mu_y = 300$ and standard deviations $\sigma_x = \sigma_y = 10$. Consider the RV U defined by

$$U = \frac{1}{2}(X + Y).$$

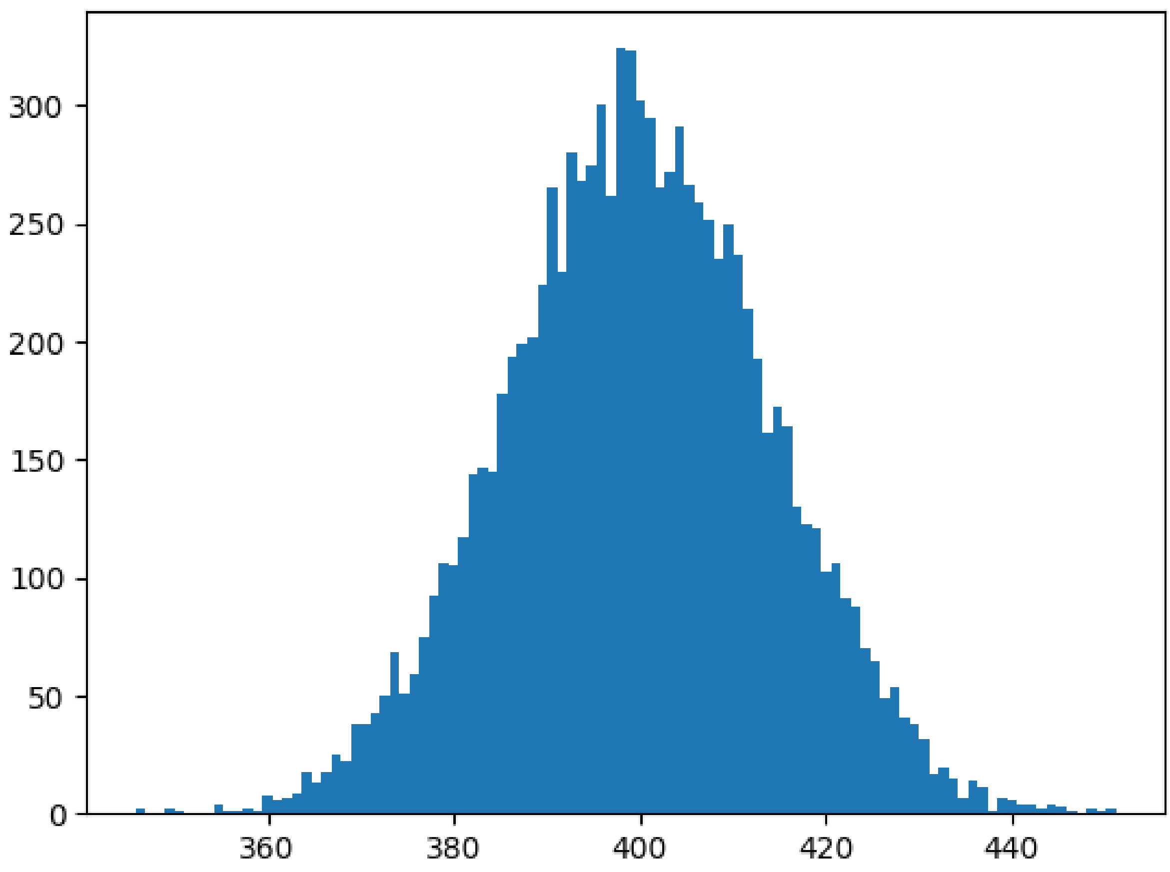
Alternatively, consider the RV Z that is generated as follows:

(a) With probability $\frac{1}{2}$ we sample Z from $N(\mu = 100, \sigma^2 = 100)$.

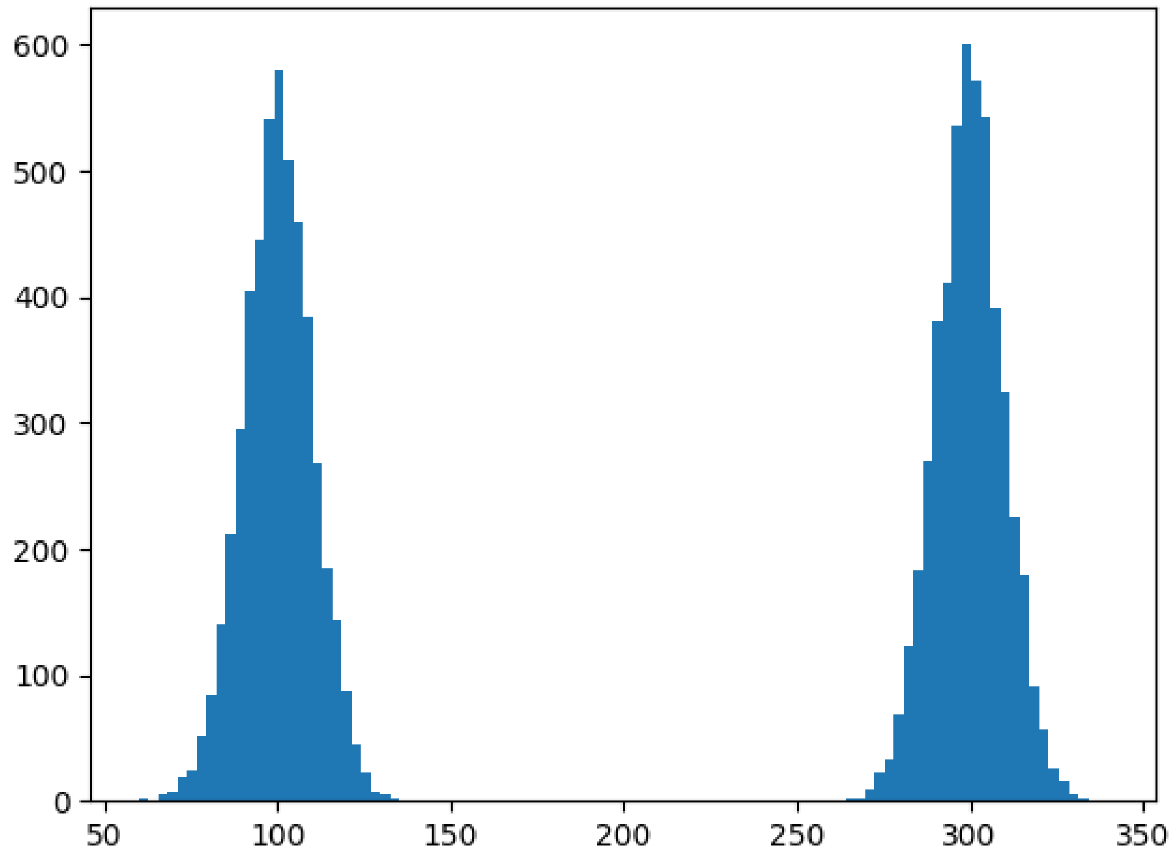
(b) With probability $\frac{1}{2}$ we sample Z from $N(\mu = 300, \sigma^2 = 100)$.

1. **[5 points]** Simulate the sampling, and produce two histograms (one for U and one for Z) over 10 000 samples for each U, Z.
2. **[10 points]** Compute the expected values of U, Z.
3. **[10 points]** Compute the variances of U, Z.

Solution:
This is the graph for U.



This is the graph for Z.



3.2

$\because$ linearity

$$\Rightarrow E[U] = E[\tfrac{1}{2}(X+Y)] = \tfrac{1}{2}[E(X)+E(Y)]$$
$$= \tfrac{1}{2} \cdot 400 = 200$$

$$E(Z) = \tfrac{1}{2} E(X) + \tfrac{1}{2} E(Y) = \tfrac{1}{2} \cdot (100+300)$$
$$= 200$$

3.3

$\because$ linearity

$$\Rightarrow Var[U] = \tfrac{1}{4}[Var(X)+Var(Y)]$$
$$= \tfrac{200}{4} = 50$$

$$Var[Z] = \tfrac{1}{2} Var(X) + \tfrac{1}{2} Var(Y)$$
$$= \tfrac{1}{2} \times 100 + \tfrac{1}{2} \times 100$$
$$= 100$$

4 Coding EM for Mixture of Gaussians [25 points]

Check the Jupyter notebook on our Git repo.