

Recitation Session 7 Solutions1. *CLT and MLE.*

- a. For data
- x_1, \dots, x_n
- first find an estimator to the parameter
- p
- ,

$$\mathcal{L}_{\mathbf{x}}(\hat{p}) = \prod_{i=1}^n p_{X_i}(x_i, \hat{p}) \quad (1)$$

$$= \prod_{i=1}^n \binom{20}{x_i} \hat{p}^{x_i} (1 - \hat{p})^{20-x_i} \quad (2)$$

- b. Take logarithm of the above expression,

$$\log \prod_{i=1}^n \binom{20}{x_i} \hat{p}^{x_i} (1 - \hat{p})^{20-x_i} = \log(c) + \sum x_i \log \hat{p} + (20n - \sum x_i) \log(1 - \hat{p}) \quad (3)$$

- c. Take two derivatives with respect to
- \hat{p}
- ,

$$\frac{d \log \mathcal{L}_{\mathbf{x}}(\hat{p})}{d\hat{p}} = \frac{1}{\hat{p}} \sum x_i - \frac{1}{1 - \hat{p}} (20n - \sum x_i) \quad (4)$$

$$\frac{d^2 \log \mathcal{L}_{\mathbf{x}}(\hat{p})}{d\hat{p}^2} = \frac{-1}{\hat{p}^2} \sum x_i - \frac{1}{(1 - \hat{p})^2} (20n - \sum x_i) \quad (5)$$

Second derivative is negative, hence, we solve the first derivative for zero,

$$\hat{p} = \frac{\sum x_i}{2000} \quad (6)$$

- d. Since we want to estimate the mean of the distribution
- $\mu_{\text{ML}} = 20\hat{p} = \frac{\sum x_i}{100}$

- e. Unbiased, as expected, it is the average of samples.

- f. Note that we can apply central limit theorem to
- μ_{ML}
- , as
- $\sum x_i/n$
- is approximately normally distributed with mean
- $20p = 20 \cdot 0.65 = 13$
- and variance
- $20p(1-p)/n = 20 \cdot 0.65 \cdot 0.35/100 = 0.0455$
- (hence st dev is 0.21).

$$P(\mu_{\text{ML}} > 14) = P\left(\frac{\mu_{\text{ML}} - 13}{0.048} > \frac{14 - 13}{0.21}\right) = Q(4.762) \approx 0 \quad (7)$$

In many problems there is an intimate connection between MLE and the normal distribution.

2. *Normal priors.*

- a. The prior density is,

$$\pi(\mu) = \sqrt{\frac{s}{2\pi}} \exp\left(-\frac{s}{2}(\mu - m)^2\right). \quad (8)$$

b. Product of independent densities,

$$f_{\mathbf{X}|M}(\mathbf{x}|\mu) = \prod_{i=1}^n f_{X_i}(x_i, \mu) \quad (9)$$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \sum (x_i - \mu)^2\right) \quad (10)$$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{n}{2}(\mu - \hat{\mu})^2 - \frac{1}{2} \sum (x_i - \hat{\mu})^2\right) \quad (11)$$

where $\hat{\mu} = \frac{1}{n} \sum x_i$.

c. Posterior is proportional to the density times the prior, $f_{M|\mathbf{X}}(\mu|\mathbf{x}) = f_{\mathbf{X}|M}(\mathbf{x}|\mu) \pi(\mu)$:

$$f_{M|\mathbf{X}}(\mu|\mathbf{x}) = c(\mathbf{x}) \exp\left(-\frac{n}{2}(\mu - \hat{\mu})^2 - \frac{1}{2} \sum (x_i - \hat{\mu})^2 - \frac{s}{2}(\mu - m)^2\right) \quad (12)$$

$$= \tilde{c}(\mathbf{x}) \exp\left(-\frac{1}{2}(n(\mu - \hat{\mu})^2 + s(\mu - m)^2)\right) \quad (13)$$

d. The goal is to shape the above density to a normal density form. We can focus on the exponential part only:

$$-\frac{1}{2}(n(\mu - \hat{\mu})^2 + s(\mu - m)^2) = -\frac{1}{2}((n+s)\mu^2 - 2(n+s)\mu\hat{\mu} + n\hat{\mu}^2 + sm^2) \quad (14)$$

$$= -\frac{n+s}{2}(\mu - \mu'(\mathbf{x}))^2 \quad (15)$$

where $\mu'(\mathbf{x}) = \frac{s}{s+n}m + \frac{n}{s+n}\hat{\mu}$. Therefore the posterior is normal distribution with mean weighted average and variance inversely proportional to the total information $s+n$. Thus if n is small compared to s then the mean is near m and if n is large then the mean is near the sample mean.

3. *Which method?* Comparing the two groups, they have similar median but means are very different. Assuming data coming from approximately normal with some parameters will fail. It is more appropriate to use non-parametric density estimation methods.