

Chapter 4 Sols + 5.1 Sols

1) a) ~s

$$f_x(\theta, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(x-\theta)^2}{\sigma^2}}$$

(Density for 1 observation)

$$f_x(\theta, \sigma) = (2\pi\sigma^2)^{-n/2} \cdot \exp\left\{\sum_{i=1}^n -\frac{1}{2} \frac{(x_i-\theta)^2}{\sigma^2}\right\}$$

(likelihood of sample)

$$l_x(\theta, \sigma) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \cdot \sum_{i=1}^n \frac{(x_i-\theta)^2}{\sigma^2}$$

(log-likelihood of sample)

optimality condition

$$\frac{\partial l_x}{\partial \theta} = \sum_{i=1}^n \frac{(x_i-\theta)}{\sigma^2} = 0 \Rightarrow \sum_{i=1}^n x_i - n\theta = 0$$

↑

Derivative wr.t. θ

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

(mean estimator)

$$\frac{\partial L(\sigma)}{\partial \sigma} = -n \cdot \frac{2\pi \sigma}{2\pi \sigma^2} + \frac{\sum_{i=1}^n (x_i - \theta)^2}{\sigma^3} = 0$$

Optimality condition

Derivative w.r.t. σ

$$\Rightarrow \frac{\sum_{i=1}^n (x_i - \theta)^2}{\sigma^3} = \frac{n}{\sigma}$$

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

(Estimator for variance.)

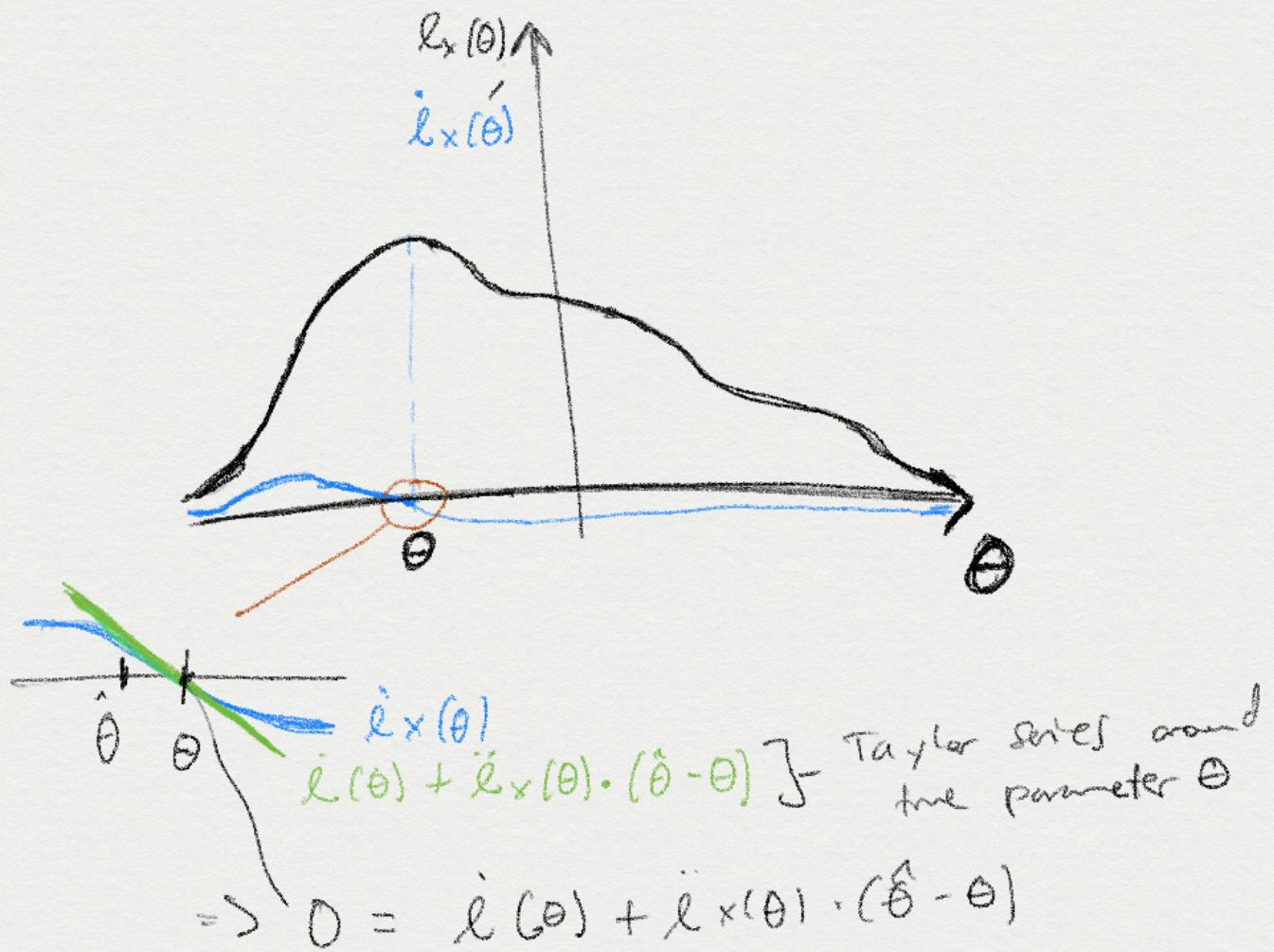
$$\sigma = \left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \right]^{1/2}$$

Using plugin principle
 $\hat{\theta} = \bar{x}$ and taking square root.

b) $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
Estimator for variance only takes into account current sample when maximizing likelihood. It ignores the fact that \bar{x} , our plug-in estimate for θ , results in loss of 1 degree of freedom for the residuals.

- ② Given a log-likelihood function $\ell_X(\theta)$, we observe that score function $\dot{\ell}_X(\theta)$ is zero at the value of θ where $\ell_X(\theta)$ is maximized.

zero at the value of θ where $\hat{\ell}_x(\theta)$ is maximized.



$$(\theta - \hat{\theta}) \tilde{I}_x(\theta) = \dot{L}(\theta)$$

$$\hat{\theta} = \theta - \frac{\dot{L}_x(\theta)}{\ddot{L}_x(\theta)} = \theta - \frac{\dot{L}_x(\theta)/n}{\ddot{L}_x(\theta)/n}$$

③ Cramér-Rao Bound

$$V_{\theta}\{\tilde{\theta}\} \geq -\frac{1}{I_{\theta}}$$

Total Fisher Information,
if observations are iid
then $= n \cdot I_{\theta}^{(1)}$
where $I_{\theta}^{(1)}$ is the Fisher Inf. of
1 observation.

$$X_1 \sim \text{Bin}(20, \theta)$$

samp led Indep.

$$X_2 \sim \text{Poi}(10 \cdot \theta)$$

$$f(x_1) = \binom{n}{x_1} \cdot \theta^{x_1} \cdot (1-\theta)^{n-x_1}$$

$$\ell_{x_1}(\theta) = \log \left[\binom{n}{x_1} \right] + x_1 \log(\theta) + (n-x_1) \log(1-\theta)$$

$$\dot{\ell}_{x_1}(\theta) = \frac{x_1}{\theta} + \frac{(n-x_1)}{1-\theta}$$

$$\ddot{\ell}_{x_1}(\theta) = -\frac{(n-x_1)}{(1-\theta)^2} - \frac{x_1}{\theta^2}$$

$$\mathcal{L}_\theta = -E[\ddot{\ell}_{x_1}(\theta)] = \frac{(n-n\cdot\theta)}{(1-\theta)^2} + \frac{n\cancel{\theta}}{\theta^2}$$

$$= \frac{n(1-\theta)}{(1-\theta)^x + \frac{n}{\theta}} = \frac{\cancel{n\theta} + n - \cancel{n\theta}}{\theta(1-\theta)} = \frac{n}{\theta(1-\theta)}$$

$\boxed{\frac{20}{\theta(1-\theta)}}$

Let $\lambda = 10 \cdot \theta$

$$f_{X_2}(\lambda) = \frac{e^{-\lambda} \cdot \lambda^{x_2}}{x_2!}$$

$$\ell_{X_2}(\lambda) = -\lambda + x_2 \log(\lambda) - \log(x_2!)$$

$$\dot{l}_{x_2}(\lambda) = -1 + \frac{x_2}{\lambda}$$

$$\ddot{l}_{x_2}(\lambda) = \frac{x_2}{\lambda^2}$$

Then, taking the expectation & substituting Θ back...

$$I_\theta = -E[\ddot{l}_{x_2}(\theta)] = -E\left[\frac{-x_2}{(10\theta)^2}\right] = \cancel{\frac{-(10\cdot\theta)}{(10\cdot\theta)^2}} = \frac{1}{10\cdot\theta}$$

Finally, $\mathcal{L}_\theta | x_1, x_2$ = $\frac{2\theta}{\theta(1-\theta)} + \frac{1}{10\theta}$] Total Fisher inf. of both observations.

Finally, $I_{\theta|x_1, x_2} = \frac{20}{\theta(1-\theta)} + \frac{1}{10\theta}$] Total Fisher Inf. of both observations.

let's now compute numerically, as requested by problem. Since θ is the success probability of a Binomial distribution, we know $0 \leq \theta \leq 1$.
 Making a grid in increments of 0.1, we get

θ	$I_{\theta x_1, x_2}$	$1/I_{\theta}$	In Information gain
			$I_{\theta=0} - I_{\theta=1}$

θ	$\bar{I}_\theta x_1, x_2$
0.1	223
0.2	125
0.3	95
0.4	84
0.5	96
0.6	84
0.7	95
0.8	128
0.9	222

$$\bar{I}_\theta | x_1, x_2$$

$$1 / \bar{I}_\theta$$

0.0044

0.0080

0.0104

0.0119

0.0125

0.0119

0.105

0.0080

0.0045

Information gain
is greater at
the edges "!"

Variance of any
unbiased estimator
 $\geq 1 / \bar{I}_\theta$ for fair
sample.

④

a) $\frac{x_1 + x_2}{n_1 + n_2}$, the total amount of successes over the total amount of trials.

b) Conditional Inference, because we considered the outcomes of the individual samples, regardless of how the second sample was generated (i.e., ignoring the distribution of x_1).

⑤ Just change B from 10000 to 1000 in our Jupyter from Class.

⑥ Since the posterior computed by statistician
B is proportional to the likelihood,

$$\begin{aligned} P(\theta | x) &\propto f_x(\theta) \cdot g(\theta) \\ &\propto \underline{f_x(\theta)} \end{aligned}$$

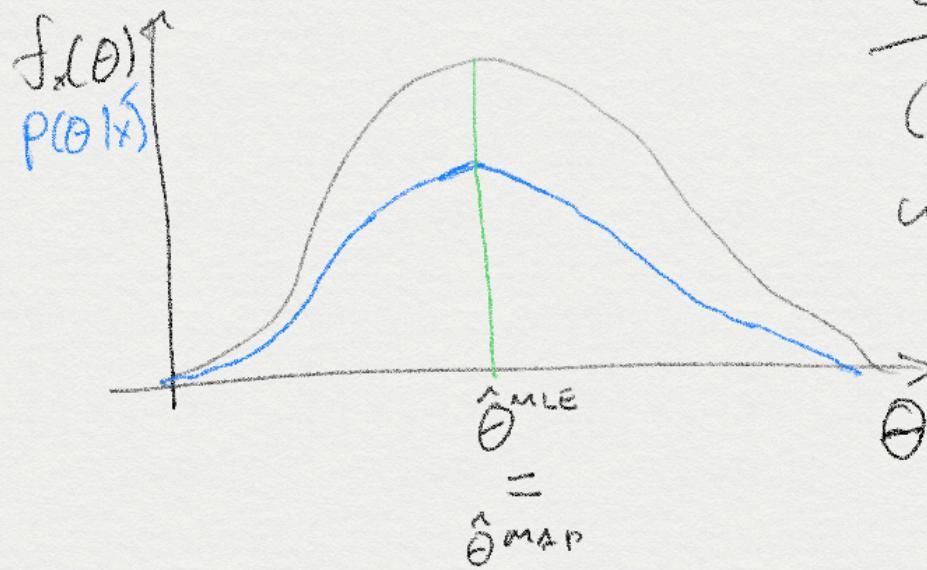
which is the function statistician A is trying to
maximize. Thus, under a flat prior, they have the

* which is the function Statistician A is trying to maximize. Thus, under a flat prior, they have the

same maximizer.

$$(\hat{\theta}_{MLE} = \hat{\theta}_{MAP})$$

where $\hat{\theta}$ is the maximum a posteriori estimate.



Chapter 5, ①.

$$X \sim \text{Poi}(\mu), \pi(\mu) \sim \text{Ga}(V, 1)$$

a) Before getting the marginal density of X , let's get the joint density for $\underline{x, \mu}$.

$$f(x, \mu | V) = f_x(x) \cdot \pi(\mu) \quad (\text{using conditional probability})$$

$$= \frac{\mu^x \cdot e^{-\mu}}{x!} \cdot \frac{\mu^{V-1} \cdot e^{-\mu/1}}{1^V \cdot \Gamma(V)}$$

$$\frac{\mu^{(V+x)-1} \cdot e^{-2\mu}}{\Gamma(x+1) \cdot \Gamma(V)}$$

Recall
 $\Gamma(y+1) = y!$

Now, get marginal density $f(x)$ by integrating over μ .

$$f(x|N) = \frac{1}{T(x-1)} \cdot \int_0^{\infty} \frac{\mu^{(v+x)-1} \cdot e^{-\mu/(1/2)}}{\pi(v)} d\mu$$

$$f(x|N) = \frac{(1/2)^{v+x}}{T(x-1)} \cdot \underbrace{\int_0^{\infty} \frac{\mu^{(v+x)-1} \cdot e^{-\mu/(1/2)}}{(1/2)^{v+x} \cdot T(v)} d\mu}_{\text{Gamma density } \text{Ga}(v+x, 1/2)}$$

$$\Rightarrow f(x|N) = \frac{(1/2)^{v+x}}{T(x-1)}$$

b) Using Bayes Rule

$$f(w|x,v) = \frac{f(w,x|v)}{f(x|v)}$$

So just substitute them & simplify !!.