Chapter 4 Sols + 5.1 Sols

(1) a) ⇝

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

$$f_{\underline{x}}(\theta,\sigma) = (2\pi\sigma^2)^{-n/2} \cdot \exp\left\{\sum_{i=1}^{n} -\frac{1}{2}\frac{(x_i-\theta)^2}{\sigma^2}\right\}$$ (likelihood of sample)

$$l_{\underline{x}}(\theta,\sigma) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\cdot\sum_{i=1}^{n}\frac{(x_i-\theta)^2}{\sigma^2}$$ (log-likelihood of sample)

Optimality condition

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

Derivative w.r.t. θ

$$\Rightarrow \boxed{\hat{\theta}^{mle} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{X}}$$ (mean estimator)

Optimality condition

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

Derivative w.r.t. σ

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

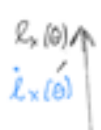
$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{n}$$ (Estimator for variance.)

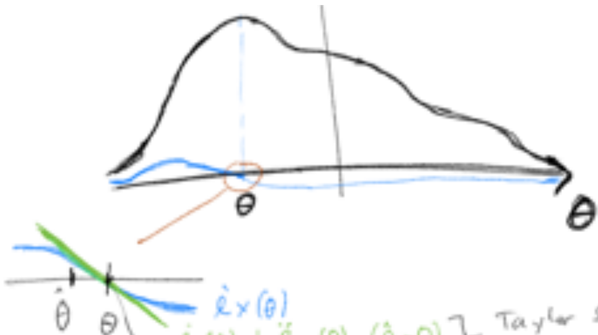
$$\boxed{\sigma = \left[\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n}\right]^{1/2}}$$ (Using plug-in principle $\theta = \overline{x}$ and taking square root.)

b) ⇝

Estimator for variance only takes into account current sample when maximizing likelihood. It ignores the fact that $\overline{x}$, our plug-in estimate for θ, results in loss of 1 degree of freedom for the residuals.

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





$\dot{\ell}_x(\theta)$

$\dot{\ell}(\theta) + \ddot{\ell}_x(\theta) \cdot (\hat{\theta} - \theta)$ } Taylor series around the parameter θ

$$\Rightarrow 0 = \dot{\ell}(\theta) + \ddot{\ell}_x(\theta) \cdot (\hat{\theta} - \theta)$$

$$(\theta - \hat{\theta})\, \ddot{\ell}_x(\theta) = \dot{\ell}_x(\theta)$$

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

3 Cramer-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.

$X_1 \sim Bin(20, \theta)$
$X_2 \sim Poi(10 \cdot \theta)$ sampled Indep.

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

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

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

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

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

$$= \frac{n\cancel{(1-\theta)}}{(1-\theta)^{\cancel{2}}} + \frac{n}{\theta} = \frac{\cancel{n\theta} + n - \cancel{n\theta}}{\theta(1-\theta)} = \frac{n}{\theta(1-\theta)}$$

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

Let $\lambda = 10 \cdot \theta$

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

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

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

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

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

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

Finally, $\boxed{I_{\theta|x_1,x_2} = \frac{20}{\theta(1-\theta)} + \frac{1}{10\theta}}$ 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 $0 \le \theta \le 1$. Making a grid in increments of 0.1, we get

θ	$I_{\theta\|x_1,x_2}$	$1/I_\theta$
0.1	223	0.0044
0.2	125	0.0080
0.3	95	0.0104
0.4	84	0.0119
0.5	96	0.0125
0.6	84	0.0119
0.7	95	0.105
0.8	125	0.0080
0.9	222	0.0045

Information gain is greater at the edges :o!

Variance of any unbiased estimator $\ge 1/I_\theta$ for this 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_i).

(5) Just Change B from 10000 to 1000 in our Jupyter from Class.

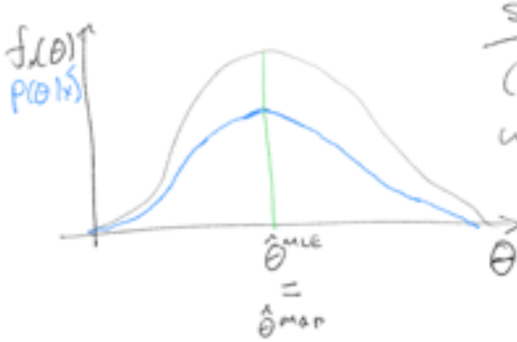
(6) Since the posterior computed by statistician B is proportional to the likelihood

$$P(\theta \mid x) \propto f_x(\theta) \cdot g(\theta)$$
$$\propto f_x(\theta)$$

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).

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

a) Before getting the marginal density of X, let's get the joint density for x, μ.

$$f(x, \mu \mid V) = f_x(\mu) \cdot \pi(\mu) \quad \left(\text{Using conditional probability}\right)$$

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

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

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

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

$$f(x \mid V) = \frac{1}{\quad} \int^{\infty} \mu^{(v+x)-1} e^{-\mu/(1/2)} d\mu$$

$$\cdots \frac{\cdots}{\Gamma(x-1)} \cdot \int_0 \frac{\cdots}{\Gamma(v)} \, d\mu$$

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

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

b) Using Bayes Rule

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

So just substitute them & simplify ☺.