Prob 3, Chap 3

Going step-by-step, let's first compute the Score function.

$$f(k) = \binom{n}{k} p^{k} (1-p)^{n-k}$$

Log-likelihood

$$\ell(k) = \log\left[\binom{n}{k}\right] + k \cdot \log(p) + (n-k) \cdot \log(1-p)$$

Score function

$$\dot{\ell}_k(p) = \frac{\partial}{\partial p}\left[\ell(k)\right]$$

$$= 0 + \frac{k}{p} - \frac{(n-k)}{1-p}$$

$$= \frac{k(1-p) - (n-x)\cdot p}{p(1-p)} = \frac{k - np}{p(1-p)}$$

Now, using the definition of Fisher Information of $\dot{\ell}_k(p)$, we'll compute the variance of $\dot{\ell}_k(p)$

$$I_p = \underbrace{\sum_{k=0}^{n} \left(\dot{\ell}_k(p) - E\left[\dot{\ell}_k(p)\right]\right)^2 \cdot \binom{n}{k} p^k (1-p)^{n-k}}_{\text{Variance of } \dot{\ell}_k(p)}$$

$$= \sum_{k=0}^{n} \left(\frac{k - np}{p(1-p)}\right)^2 \cdot \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \frac{1}{[p(1-p)]^2} \cdot \underbrace{\sum_{K=0}^{n} \binom{n}{k} p^K (1-p)^{n-K} [K - E(K)]^2}$$

This is the definition of the variance for a Binomial(n, p) distr., which is known and equal to $np(1-p)$

$$= \frac{n\,\cancel{p(1-p)}}{[p(1-p)]^{\cancel{2}}} = \frac{n}{p(1-p)}.$$

Part 1

Part 2. You can estimate $\hat{p} = \frac{K}{n}$ to get observed Fisher Info.

$$I(K) = \frac{n}{\frac{K}{n} \cdot \left[\frac{(n-K)}{n}\right]}$$

$$= \frac{n}{\frac{K(n-K)}{n^2}} = \frac{n^3}{K(n-K)}$$

Chapter 4

- want to verify $\hat{\theta}_{MLE} = \bar{x}$

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

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

$$f_{\underline{X}}(\theta,\sigma) = (2\pi\sigma^2)^{-n/2} \cdot e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\theta)^2}{\sigma^2}}$$ (Likelihood)

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

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

(with $+$ and 1 written above the crossed-out term)

Let $$\frac{\partial l_{\underline{x}}}{\partial\theta} = \frac{\sum_{i=1}^{n}(x_i-\theta)}{\cancel{\sigma^2}} = 0$$ (Because we want to get $\hat{\theta}_{MLE}$)

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

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

Now, let's get $\hat{\sigma}_{MLE}$

$$\frac{\partial}{\partial\sigma}(2\pi\sigma^2) = 4\pi\sigma \qquad \frac{\partial}{\partial\sigma}\left(\frac{1}{\sigma^2}\right) = -2\cdot\frac{1}{\sigma^3}$$

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

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

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

To get $\hat{\sigma}^{MLE}$, we let $\frac{\partial l_x}{\partial \sigma} = 0$

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

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

Finally, $\hat{\sigma}^{MLE} = \left[\frac{\sum_{i=1}^{n} (x_i - \theta)^2}{n} \right]^{1/2}$

All these steps assumed we knew θ.
If in practice we don't know it, then use
plug-in principle to substitute $\bar{x}$ in place
of θ.

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

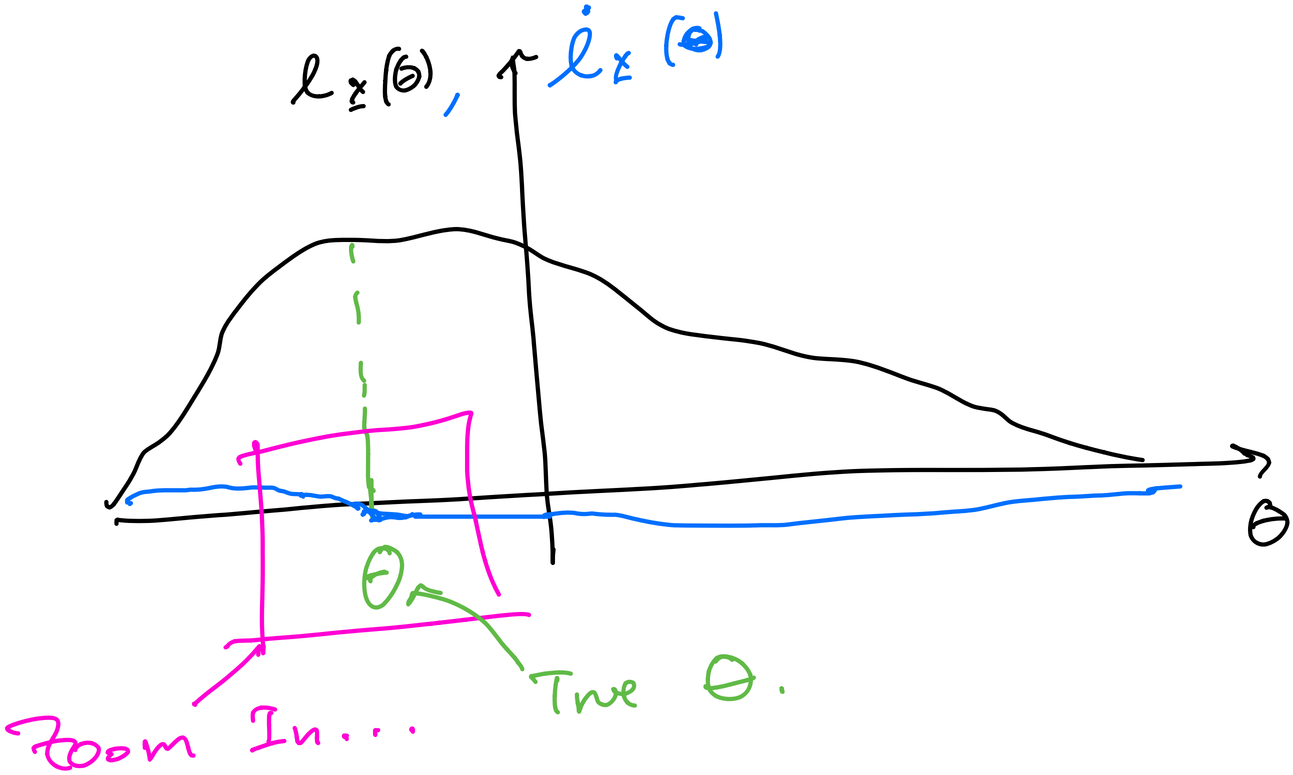
b) Because of $\hat{\sigma}^{MLE}$ assumed we knew θ,
If we need to estimate it, then, we'll lose
a degree of freedom and will to apply
correction factor $\sqrt{\frac{n}{n-1}}$ to get an unbiased

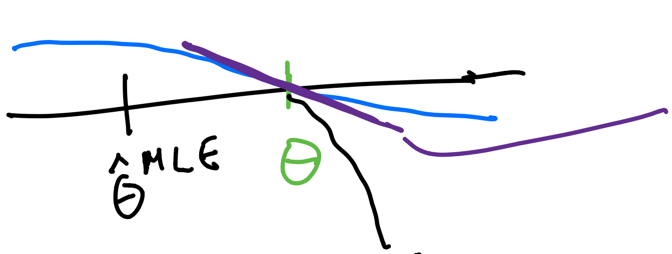
estimator.

$$\hat{\sigma}_{\text{unbiased}} = \hat{\sigma}^{MLE}\left(\sqrt{\frac{n}{n-1}}\right)$$

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

② Given a log-likelihood function $\ell_X(\theta)$, we observe that the score function $\dot{\ell}_X(\theta)$ is <u>zero</u> at $\hat{\theta}^{MLE}$.





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

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

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

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

(3) You observe $x_1 \sim \text{Binomial}(20, \theta)$ and independently $x_2 \sim \text{Poi}(10 \cdot \theta)$

Numerically compute the Cramér–Rao Bound.

Hint: Fisher Information adds for independent observations.

$$\boxed{V_\theta\{\hat{\theta}\} \geq \frac{1}{I_\theta}}$$

total Fisher Information

if observations are iid, then $= n I_\theta^{(1)}$

where $I_\theta^{(1)}$ is the Fisher Info of 1 observation

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

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

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

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

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

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

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

Let $\lambda = 10 \cdot \theta$

$$f_{x_2}(\lambda) = \frac{e^{-\lambda} \cdot \lambda^{x_2}}{x_2!} \quad \left(\text{Poisson Density}\right)$$

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

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

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

Now, taking the expectation...

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

$$= \frac{1}{10\theta}$$

Finally, $\mathcal{I}_\theta^{x_1+x_2} = \frac{20}{\theta(1-\theta)} + \frac{1}{100}$ ✓

Total Fisher Information for both observations.

θ	$\mathcal{I}_\theta^{x_1+x_2}$	$1/\mathcal{I}_\theta$
0.1	223	0.0044
0.2	125	0.0080
0.3	95	0.0104
0.4	84	0.0119
0.5	76	0.0125
0.6	84	0.0119
0.7	95	0.0105
0.8	125	0.0080
0.9	222	0.0045

Cramér Rao Bound.
Lower bound for Variance of any unbiased estimator $\tilde{\theta}$ of θ increases for values of θ that are farther away from the edges.

④ A coin w/ probability of heads θ is flipped n_1 times, yielding x_1 heads. Then, it is flipped another x_1 times yielding x_2 heads.

a) What is an intuitively plausible estimate of θ?

(1) Flip a coin n_1 times $\rightarrow$ get $\underbrace{x_1}_{x_1 \leq n_1}$ successes

(2) Flip a coin X_1 times $\longrightarrow$ get x_2 successes
$x_2 \leq x_1$

$$\underbrace{\hat{\theta} = \frac{x_1 + x_2}{n_1 + x_1}}$$

$$\cancel{\frac{x_1}{n_1} + \frac{x_2}{x_1}}$$

The total amount of successes over the total amount of trials.

b) What Fisherian principle have you invoked?

Conditional Inference, because we considered the outcomes of the individual samples, regardless of how the second sample was generated.

(5) Recreate a version of figure 4.3 based on 1000 permutations.

→ Jupyter from yesterday.

(6) A one-parameter family of densities $f_\theta(x)$ gives an observed value x.

Statistician A computes the MLE $\hat{\theta}$

Statistician B uses a prior $\boxed{g(\theta) = 1}$ flat prior to compute $\hat{\theta}^{MAP}$.

How is $\hat{\theta}^{MAP}$ related to $\hat{\theta}^{MLE}$ in this setting?

$\hat{\theta}^{MAP} = \hat{\theta}^{MLE}$.