Ch 5 1] * p129 Let $f(\underset{\sim}{y}|\underset{\sim}{\theta})$ be the pmf or pdf of $\underset{\sim}{Y}$ for $\underset{\sim}{\theta} \in \Theta$. If $\underset{\sim}{Y} = \underset{\sim}{y}$ is observed, the likelihood function

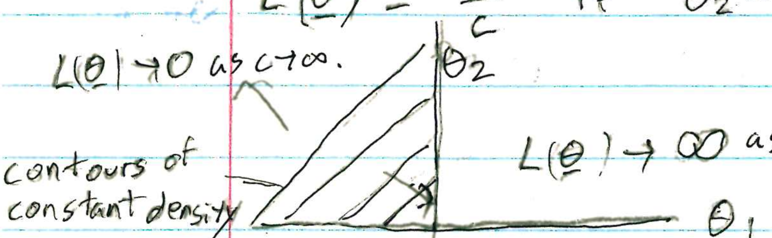
$$L(\underset{\sim}{\theta}) \equiv L(\underset{\sim}{\theta}|\underset{\sim}{y}) = f(\underset{\sim}{y}|\underset{\sim}{\theta}).$$

2] If $Y_1, \ldots, Y_n$ are iid, $L(\underset{\sim}{\theta}) = \prod_{i=1}^{n} f(y_i|\underset{\sim}{\theta})$.

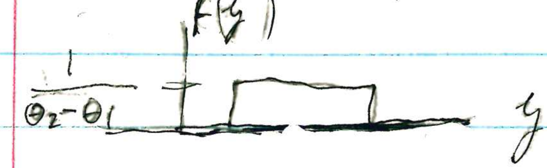
3] As a function of $\underset{\sim}{\theta}$, the function $L(\underset{\sim}{\theta})$ is not necessarily a pdf or pmf.

ex] $f(y|\theta_1, \theta_2) = \dfrac{1}{\theta_2 - \theta_1}$, $\theta_1 \le x \le \theta_2$ so $Y \sim U(\theta_1, \theta_2)$.

For $n=1$, $f(y|\theta_1, \theta_2)$ is a fn of 1 variable $y$.
$L(\theta_1, \theta_2)$ is a fn of 2 variables $\theta_1$ and $\theta_2$.
Suppose $y = 0$. Then $\theta_1 < 0$ and $\theta_2 > 0$.
$L(\underset{\sim}{\theta}) = \frac{1}{c}$ if $\theta_2 - \theta_1 = c$ or $\theta_2 = \theta_1 + c$.




$L(\underset{\sim}{\theta}) \to \infty$ as $c \to 0$. Take the SE corner of your paper and lift it until it is vertical. The paper is very similar to $L(\underset{\sim}{\theta})$. Note that $L(\underset{\sim}{\theta})$ and $f(y)$ are very different functions.




4] A point estimator is a statistic $\Leftrightarrow$ any function of $Y_1, \ldots, Y_n$ free of $\underset{\sim}{\theta}$.

5] An estimator is a statistic $\underset{\sim}{T}(\underset{\sim}{Y})$ while an estimate $\underset{\sim}{T}(\underset{\sim}{y})$ is the value of the estimator evaluated at the observed data.

ex] $T(\underset{\sim}{Y}) = Y_{(1)}$ is an estimator. If data is $y_1 = 1$, $y_2 = -10$, $T(\underset{\sim}{y}) = y_{(1)} = -10$ is the estimate.

6] Know p129 For each sample point $\underset{\sim}{y}$, let $\hat{\underset{\sim}{\theta}}(\underset{\sim}{y}) \in \Theta$ be a parameter value at which $L(\underset{\sim}{\theta})$ attains its maximum as a function of $\underset{\sim}{\theta}$ with $\underset{\sim}{y}$ treated as a constant. A maximum likelihood estimator MLE of $\underset{\sim}{\theta}$ based on a sample $\underset{\sim}{Y}$ is $\hat{\underset{\sim}{\theta}}(\underset{\sim}{Y})$.

7] Range $\hat{\underset{\sim}{\theta}}(\cdot) = \Theta$. So if you find an MLE $\hat{\underset{\sim}{\theta}} \notin \Theta$, you made a mistake.

8] Know P130 Invariance Principle If

$\hat{\underset{\sim}{\theta}}$ is the MLE of $\underset{\sim}{\theta}$, then $h(\hat{\underset{\sim}{\theta}})$ is the MLE of $h(\underset{\sim}{\theta})$ where $h$ is a function with domain $\Theta$.

9] Proof II on p146:

If $h$ is one to one, then $h(\underset{\sim}{\theta}) = \underset{\sim}{\eta}$ and $f(\underset{\sim}{y} \mid \underset{\sim}{\theta}) = f(\underset{\sim}{y} \mid h^{-1}(\underset{\sim}{\eta}))$ and $L^*(\underset{\sim}{\eta}) = K(\underset{\sim}{\eta}) = L(h^{-1}(\underset{\sim}{\eta}))$. If $\hat{\underset{\sim}{\theta}}$ is the

MLE, then $\sup_{\underline{\theta}} L(\underline{\theta}) = L(\hat{\underline{\theta}})$. So

$\sup_{\underline{\eta}} K(\underline{\eta}) = L(\hat{\underline{\theta}})$ and the MLE $\hat{\underline{\eta}} = h(\hat{\underline{\theta}})$.

If $h$ is not one to one, there is a function $w(\underline{\theta}) = (h(\underline{\theta}), u(\underline{\theta})) = (\underline{\eta}, \underline{\gamma}) = \underline{\xi}$ that is one to one, eg $w(\underline{\theta}) = (h(\underline{\theta}), \underline{\theta})$.

So $w(\hat{\underline{\theta}}) = (h(\hat{\underline{\theta}}), u(\hat{\underline{\theta}}))$ is the MLE of $(h(\underline{\theta}), u(\underline{\theta}))$.

10) p132 Let $\underline{\theta} = (\underline{\theta}_1, \underline{\theta}_2)$ and MLE

$\hat{\underline{\theta}}(\underline{Y}) = (\hat{\underline{\theta}}_1(\underline{Y}), \hat{\underline{\theta}}_2(\underline{Y}))$. We will say

$\hat{\underline{\theta}}_i(\underline{Y})$ is the MLE of $\underline{\theta}_i$, $i = 1, 2$.

Note) Although $w(\underline{\theta}) = (h(\underline{\theta}), \underline{\theta})$ is one to one, often there is an obvious $u(\underline{\theta}) \neq \underline{\theta}$ that makes the dimension of $w(\underline{\theta})$ smaller. Need a one to one function $\underline{\eta} = w(\underline{\theta})$ so $\underline{\theta}$ is being renamed.

11] Know 2nd big type of question on Exam 3, final and qual is finding the MLE of $\theta$ and of $h(\theta)$ or $\tau(\theta)$.

12] p130 * Techniques for finding potential candidates:

i) Differentiate the log likelihood $\log L(\theta)$.

ii) Directly maximize the likelihood $L(\theta)$, often using a plot.

iii) If $\hat{\theta}$ is the MLE of $\theta$, $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$.

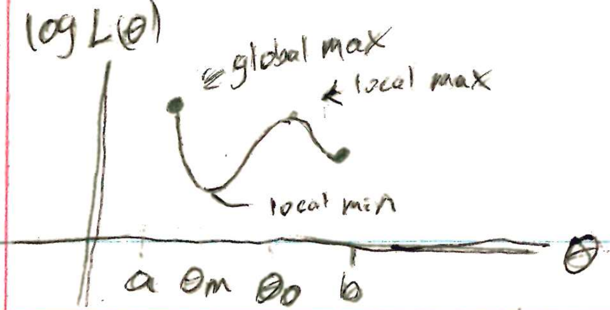
13] If $L(\theta) > 0 \ \forall \ \theta \in \Theta$ and $\hat{\theta}$ maximizes $\log L(\theta)$, then $\hat{\theta}$ maximizes $L(\theta)$.

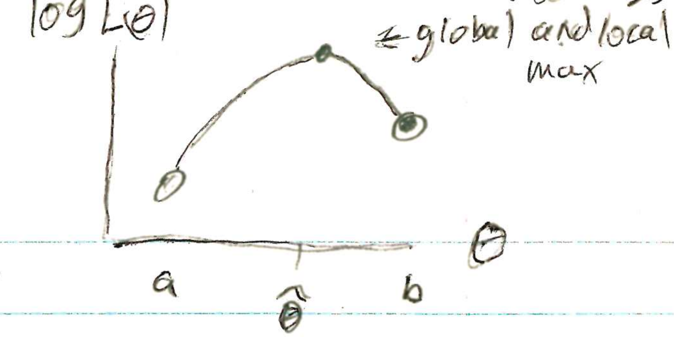
14] p130 Suppose $\theta$ is a scalar, $L(\theta) > 0$ is continuous on $\Theta$ and $\Theta = [a,b]$ is closed and bounded. Then $\exists \ \hat{\theta} \in \Theta$ that maximizes $L(\theta)$. If $\log L$ is differentiable on $(a,b)$ and if $\frac{d}{d\theta} \log L(\theta) \overset{set}{=} 0$ has a unique solution $\theta_0(\underline{y}) \in \Theta$, then

$$\hat{\theta}(\underline{Y}) = \begin{cases} \theta_0(\underline{Y}) & \\ a & \text{depending on which maximizes} \\ b & \log L(\theta). \end{cases}$$

For general $\Theta$, eg $\Theta = (-\infty, \infty)$, we still need to check the boundaries.

15) Know p131 Remark 5.1 ✓ Suppose $\Theta$ is an interval with endpoints $a < b$, not necessarily finite. Suppose $\log L(\theta)$ is continuous on the interval $\Theta$ and differentiable on $(a,b)$. If the critical point $\hat{\theta} \equiv \hat{\theta}(\underline{y})$ (soln to $\frac{d}{d\theta} \log L(\theta) \overset{set}{=} 0$) is unique and if $\hat{\theta}$ is a local max (eg $\left.\frac{d^2}{d\theta^2} \log L(\theta)\right|_{\hat{\theta}} < 0$), then $\hat{\theta}(\underline{Y})$ is the MLE ($\hat{\theta}$ is the global max). This result holds since if $\hat{\theta}$ is not the global max then there would be a local min and




Remark always check boundaries because it is easy to make a mistake.

16) know p 131
If $\log L(\theta)$ is strictly concave:
$\frac{d^2}{d\theta^2} \log L(\theta) < 0 \ \forall \ \theta \in \Theta$, then
any local max $\theta_0 = \theta_0(\underline{y})$ is the global max,
so $\hat{\theta}(\underline{Y}) = \theta_0(\underline{Y})$.

17) Let $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$, solve
$(*)$ $\frac{\partial}{\partial \theta_i} \log L(\underline{\theta}) \overset{set}{=} 0$ or $(**)$ $\frac{\partial}{\partial \theta_i} L(\underline{\theta}) \overset{set}{=} 0$

for $i = 1, \ldots, k$. The solutions to $(*)$ or $(**)$ are potential candidates for the MLE. Other candidates are on the boundary of $\Theta$ and where the derivatives don't exist. The solutions of $(*)$ could be local min's, local max's or the global max (want global max).

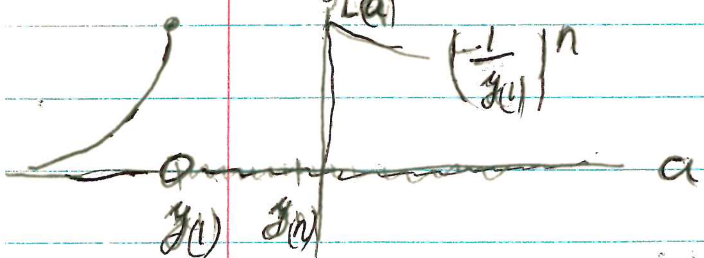
ex) Direct maximization see ex 5.8, p141 uniform

ex) Direct maximization with a plot. see HW 8 3), 4), 5).
Let $Y_1, \ldots, Y_n$ be iid $U(a, 0)$, $a < 0$.
So $\Theta = (-\infty, 0)$. Then $f(y|a) = \frac{-1}{a} I(a \le y \le 0)$.

So $L(a) = f(\underline{y}|a) = \prod_{i=1}^{n} f(y_i|a) = \prod_{i=1}^{n} \frac{-1}{a} I(a \le y_i \le 0)$

$= \left(\frac{-1}{a}\right)^n I(a \le y_{(1)})\ I(y_{(n)} \le 0)$.

using a minimal suff stat $(Y_{(1)})$ in $f(\underline{y})$ is often useful.




$L(a)$ is maximized by $\hat{a}(\underline{y}) = y_{(1)}$ since $\frac{-1}{y_{(1)}} > \frac{-1}{a}$ for any $a < y_{(1)}$

(eg $y_{(1)} = -10$, $a = -100$, $\frac{-1}{-10} > \frac{-1}{-100}$).

So $\hat{a} = Y_{(1)}$ is the MLE of $a$.

18] Memorize how to get the MLE if $Y_1, \ldots, Y_n$ are iid i) $U(a,0)$, ii) $U(0,b)$, iii) $N(\mu, \sigma^2)$, $\sigma^2$ known, iv) $N(\mu,\sigma^2)$ $\mu$ known and v) $N(\mu, \sigma^2)$ both $\mu$ and $\sigma^2$ unknown. See p 338-339 and example 5.2.

19) p132 ✗ Suppose $Y_1, \ldots, Y_n$ are iid with $f(\underline{y}|\underline{\theta})$ where $\underline{\theta} = (\theta_1, \theta_2)$. Suppose the MLE of $\underline{\theta}$ is $\hat{\underline{\theta}}(\underline{y}) = (\hat{\theta}_1(\underline{y}), \hat{\theta}_2(\underline{y}))$. The profile likelihood function $L_p(\theta_1) = L(\theta_1, \hat{\theta}_2(\underline{y}))$ with domain $\{\theta_1 | (\theta_1, \hat{\theta}_2) \in \Theta\}$.

20] Since $\hat{\theta}_1(\underline{y})$ maximizes $L_p(\theta_1)$, $\log L_p(\theta_1)$ is useful for finding $\hat{\theta}_1$ if $\hat{\theta}_2$ can be found by direct maximization.

55C 36

see ex 5.3 half normal

p 311 2 parameter $EXP(\theta, \lambda)$

ex] ex 5.2 p135

ex] $Y_1, \ldots, Y_n$ are iid $N(\mu, \sigma^2)$.

(*) Fact $\bar{y}$ is the global minimizer of $h(\mu) = \sum_{i=1}^{n} (y_i - \mu)^2$.

proof $\frac{d}{d\mu} \sum (y_i - \mu)^2 = \sum -2 (y_i - \mu) \overset{set}{=} 0$

or $\sum y_i - n\mu = 0$ or $\hat{\mu} = \bar{y}$ unique

$\frac{d^2}{d\mu^2} \sum (y_i - \mu)^2 = \frac{d}{d\mu} 2n\mu = 2n > 0$

so $\hat{\mu} = \bar{y}$ is global min (much like remark 5.1 v where unique local max is global max).

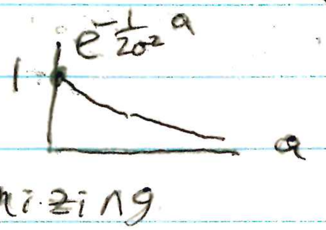
Now $L(\mu, \sigma^2) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[\frac{-1}{2\sigma^2}(y_i - \mu)^2\right]$

$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{(\sigma^2)^{n/2}} \exp\left[\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right]$.

By (*) $\bar{y}$ is the global maximizer of $\exp\left[\frac{-1}{2\sigma^2} \sum (y_i - \mu)^2\right]$

for any $\sigma^2 > 0$. So $\bar{Y}$ is the MLE of $\mu$

( $e^{-\frac{1}{2\sigma^2} a}$ is decreasing for $a \geq 0$



)

So find the MLE of $\sigma^2$ by maximizing the profile likelihood $L_p(\sigma^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left[\frac{-1}{2\sigma^2} \sum (y_i - \bar{y})^2\right]$

Let $\tau = \sigma^2$ to help prevent errors.

$\log L_p(\tau) = d - \frac{n}{2} \log(\tau) + \frac{-1}{2\tau} \sum (y_i - \bar{y})^2$.

$$\frac{d}{d\tau} \log L_p(\tau) = -\frac{n}{2}\frac{1}{\tau} + \frac{1}{2\tau^2}\sum(y_i - \bar{y})^2 \stackrel{set}{=} 0$$

or $n\tau = \sum(y_i - \bar{y})^2$ or $\hat{\tau} = \frac{1}{n}\sum(y_i - \bar{y})^2$ unique

$$\frac{d^2}{d\tau^2} \log L_p(\tau) = \frac{n}{2\tau^2} - \frac{\sum(y_i-\bar{y})^2}{\tau^3}\Big|_{\tau=\hat{\tau}} =$$

$$\frac{n}{2\hat{\tau}^2} - \frac{n\hat{\tau}}{\hat{\tau}^3}\frac{2}{2} = \frac{-n}{2\hat{\tau}^2} < 0$$

so $\hat{\tau}$ is the MLE by Remark 5.1 V.

so $\left(\bar{Y}, \frac{1}{n}\sum(Y_i - \bar{Y})^2\right)$ is the MLE of $(\mu, \sigma^2)$.

21) Know If $Y_1, ..., Y_n$ are iid from a kPREF with $f(y|\underline{\eta}) = h(y)\, b(\underline{\eta}) \exp\left[\sum_{i=1}^{k} \eta_i\, t_i(y)\right]$,

then the log likelihood function $\log L(\underline{\eta})$ is a strictly concave function of $\underline{\eta}$. Hence if $\hat{\underline{\eta}} \in \Omega$ and if $\hat{\underline{\eta}}$ is the unique critical point of $\log L(\underline{\eta})$, then $\hat{\underline{\eta}}$ is the unique MLE of $\underline{\eta}$.

Note] If $k=1$ find $\frac{d}{d\theta}\log L(\theta)$ and $\frac{d^2}{d\theta^2}\log L(\theta)$

since finding derivatives is easier than finding $f(y|\underline{\eta})$. If $k=2$, 21) can be useful.

Note] $\underline{\eta}$ is usually a one to one function of $\underline{\theta}$, so $\hat{\underline{\theta}}$ is the MLE of $\underline{\theta}$ by invariance.

Note) If $Y_1, \ldots, Y_n$ are from a pmf, often $\hat\theta \notin \Theta$. eg $Y_i$ bin$(k,\rho)$ or $Y_i$ Pois$(\theta)$

$\bar Y$ is the MLE if $\bar Y \neq 0$ since $\Theta = (0, \infty)$.
so for pmf, should add line "so $\hat\theta$ is the MLE of $\theta$ if $\hat\theta \in \Theta$."

ex] $Y_1, \ldots, Y_n$ iid Pois$(\theta)$, $\theta > 0$.

$$L(\theta) = \underbrace{\prod_{i=1}^{n} \frac{1}{y_i!} I[y_i \in \{0,1,\ldots\}]} e^{-\theta} \theta^{y_i}$$

$$= c \; e^{-n\theta} \exp\left(\log(\theta) \sum y_i\right)$$

$$\theta^{\sum y_i} = e^{\log \theta^{\sum y_i}} = e^{[\log(\theta) \sum y_i]}$$

$$\log L(\theta) = d \quad -n\theta + \log(\theta) \sum y_i$$

$$\frac{d}{d\theta} \log L(\theta) = -n + \frac{1}{\theta} \sum y_i \overset{set}{=} 0$$

$$\text{or } n\theta = \sum y_i \quad \text{or } \hat\theta = \bar y \quad \underline{\text{unique}}$$

$$\frac{d^2}{d\theta^2} \log L(\theta) = \frac{-\sum y_i}{\theta^2} < 0 \text{ if } \sum y_i \neq 0.$$

So $\bar Y$ is the MLE of $\theta$ $\underbrace{\text{if } \bar Y \in \Theta = (0,\infty)}_{\text{crucial for MLE of discrete } Y_i}$.

By invariance, $\hat\eta = \log(\bar Y)$ is the MLE of $\eta = \log(\theta)$ if $\bar Y > 0$.

ex] $Y_1, \ldots, Y_n$ iid $N(0, v = \sigma^2)$. Find the MLE of $v$ and $\sqrt{v}$.

$$L(v|\underline{y}) = \prod f(y_i|v) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{v}} \exp\left[\frac{-1}{2v} y_i^2\right]$$

37.5

$$= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{v}\right)^{n/2} \exp\left[-\frac{1}{2v} \sum y_i^2\right]$$

$$\log L(v) = \log\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} - \frac{n}{2}\log(v) - \frac{1}{2v} \sum y_i^2$$

$$\frac{d}{dv} \log L(v) = -\frac{n}{2}\frac{1}{v} + \frac{1}{2v^2} \sum y_i^2 \overset{set}{=} 0$$

(mult both sides by $2v^2$)

or $-nv + \sum y_i^2 = 0$ or $\frac{n}{2} v = \sum y_i^2$

so $\hat{v} = \frac{\sum y_i^2}{n}$ unique $\Theta = (0, \infty)$

$$\frac{d^2}{dv^2} \log L(v) = \frac{n}{2v^2} - \frac{\sum y_i^2}{v^3}\Big|_{\hat{v}} = \frac{n}{2\hat{v}^2} - \frac{n\hat{v}}{\hat{v}^3}\frac{2}{2}$$

$= \frac{-n}{2\hat{v}^2} < 0$. so $\hat{v} = \frac{\sum_{i=1}^n Y_i^2}{n}$ is

the MLE by Remark 5.1 V.

The MLE of $\sqrt{v} = \sqrt{\hat{v}} = \sqrt{\frac{\sum_{i=1}^n Y_i^2}{n}}$ by invariance.

Read ex's 5.6 and 5.7.

do the maximization on the support

**Note** $L(\underline{\theta}) = f(\underline{y}|\underline{\theta}) = g(\underline{y}|\underline{\theta})\, I(\underline{y} \in \mathcal{Y}_{\underline{\theta}})$

$= g(\underline{y}|\underline{\theta})\, I(\underline{y} \in \mathcal{Y}_{\underline{\theta}}) + 0\, I(\underline{y} \notin \mathcal{Y}_{\underline{\theta}})$

$$\log L(\underline{\theta}) = \log[g(\underline{y}|\underline{\theta})]\, I(\underline{y} \in \mathcal{Y}_{\underline{\theta}}) + \underbrace{(-\infty)\, I(\underline{y} \notin \mathcal{Y}_{\underline{\theta}})}_{\underline{\theta} \text{ where } L(\underline{\theta}) = -\infty \text{ wont be an MLE}}$$

Convention: write $\log L(\underline{\theta}) = \log[g(\underline{y}|\underline{\theta})]\, I(\underline{y} \in \mathcal{Y}_{\underline{\theta}})$

ex] If $f(\underline{y}|\theta) = g(\underline{y}|\theta)\, I(y_{(1)} \ge \theta)$ then $\log L(\theta) = \log[g(\underline{y}|\theta)]\, I(y_{(1)} \ge \theta) \overbrace{- \infty\, I(y_{(1)} < \theta)}^{\text{always omitted}}$