Solution by A. Dieuleveut

Problem 1:

1) The Poisson distribution is a discrete distribution: its support is $\mathbb{N}$.

For example, we could model by a Poisson distribut°:

* the number of people coming to a post office on one day
* the number of shooting stars during a night.
* the number of bikes stolen during a day.

2) We have, if we denote $\mathcal{P}(\theta)$ a Poisson distribution with parameter θ; and $X \sim \mathcal{P}(\theta)$:

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \, P(X=k) = \sum_{k=0}^{\infty} k \, e^{-\theta} \frac{\theta^k}{k!}$$

$$= \theta \underbrace{\boxed{\sum_{k=0}^{\infty} e^{-\theta} \frac{\theta^{k-1}}{(k-1)!}}}_{=1} = \theta$$

$$\mathbb{E}[X^2] = \sum_{k=0}^{\infty} k^2 \, P(X=k) = \sum_{k=1}^{\infty} k \, e^{-\theta} \frac{\theta^k}{(k-1)!}.$$

$$= \theta^2 \sum_{k=2}^{\infty} (k-1) \, e^{-\theta} \frac{\theta^{k-2}}{(k-1)!} + \theta \sum_{k=1}^{\infty} \frac{e^{-\theta} \theta^{k-1}}{(k-1)!}$$

$$= \theta^2 \underbrace{\boxed{\sum_{k=2}^{\infty} \frac{e^{-\theta} \theta^{k-2}}{(k-2)!}}}_{=1} + \theta \underbrace{\boxed{\sum_{k=1}^{\infty} \frac{e^{-\theta} \theta^{k-1}}{(k-1)!}}}_{=1} = \theta^2 + \theta.$$

Donc $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \theta.$

3) We observe $(X_1, \dots, X_n)$ n independant Poisson r.v.

$$(X_1, \dots, X_n) \sim \mathcal{P}(\theta)^{\otimes n}, \text{ i.e., } X_i \sim \mathcal{P}(\theta) \quad \forall i \in [\![1; n]\!]$$

and $(X_i)_{i=1\dots n}$ are independant

The statistical model is the set of possible distributions in our model:

$$\mathcal{M} = \left\{ \mathcal{P}(\theta)^{\otimes n}, \ \theta \in \mathbb{R}_+ \right\}$$

[rk: we accept θ in the set of parameters, in order to have that $\frac{\sum X_i}{n}$ is an estimator.

We sometimes write $\mathcal{M} = \left(\underbrace{\mathbb{R}^n}_{\text{set in which observations live}}, \underbrace{\mathcal{B}(\mathbb{R}^n)}_{\sigma\text{-field}}, \underbrace{\left\{ \mathcal{P}(\theta)^{\otimes n}, \theta \in \mathbb{R}_+ \right\}}_{\text{set of distributions}} \right)$

to also describe :

We are estimating $\theta \in \mathbb{R}_+$.

4) The likelihood function is :

$$L_n(\theta, X_1, \dots, X_n) = \prod_{i=1}^{n} p(\theta, X_i)$$

$$= \prod_{i=1}^{n} e^{-\theta} \frac{\theta^{X_i}}{X_i!} = e^{-n\theta} \frac{\theta^{\sum_{i=1}^{n} X_i}}{\prod_{i=1}^{n} X_i!}$$

Thus the log-likelihood is :

$$\ell_n(\theta, X_1 \dots X_n) = \ln\left(L_n(\theta, X_1, \dots, X_n)\right)$$

$$= -n\theta + \sum_{i=1}^{n} X_i \ln(\theta) - \ln\left(\prod_{i=1}^{n} X_i!\right)$$

If $\sum_{i=1}^{n} X_i > 0$:

$$\lim_{\theta \to 0} \ell_n(\theta, X_1 \dots, X_n) = \lim_{\theta \to +\infty} \ell_n(\theta, X_1, \dots, X_n) = -\infty$$

and $l_n(\cdot, X_1, \dots, X_n)$ is a differentiable function: to find $\hat{\theta}_{ML} = \underset{\theta \in \mathbb{R}_+}{\text{argmax}}\ l_n(\theta, X_1, \dots, X_n)$, we solve the following equation

$$l_n'(\theta, X_1, \dots, X_n) = 0 \iff -n + \frac{\sum_{i=1}^n X_i}{\theta} = 0$$

$$\iff \theta = \frac{\sum_{i=1}^n X_i}{n}$$

Thus $\boxed{\hat{\theta}^{MV} = \frac{\sum_{i=1}^n X_i}{n}}$

$\boxed{\text{If } \sum_{i=1}^n X_i = 0}$ then $L_n(\theta, X_1 \dots X_n)$ is decreasing on $\mathbb{R}_+$, thus $\boxed{\hat{\theta}_{ML} = 0 = \frac{\sum X_i}{n}}$

In both cases, $\boxed{\hat{\theta}_{ML} = \frac{1}{n} \sum_{i=1}^n X_i.}$

If we consider $\Theta = \mathbb{R}_+^*$ instead of $\mathbb{R}_+$, then the MLE is not defined when $\sum_{i=1}^n X_i = 0$, i.e., $\frac{1}{n}\sum_{i=1}^n X_i \notin \Theta$.

5) By the central limit theorem: (as $\mathbb{E}(X_1^2) < +\infty$)

$$\boxed{\sqrt{n}\left(\hat{\theta}_{ML} - \theta\right) = \sqrt{n}\left(\frac{\sum_{i=1}^n X_i}{n} - \underbrace{\theta}_{\mathbb{E}[X_1]}\right) \underset{n \to +\infty}{\overset{\text{distribution}}{\rightsquigarrow}} \mathcal{N}(0, \underbrace{\theta}_{\text{Var}(X_1)})}$$

6) By the law of large numbers; (as $\mathbb{E}|X| < +\infty$)

$$\hat{\theta}_{ML} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[\text{in probability}]{} \mathbb{E}[X_1] = \theta$$

Thus $\boxed{\hat{\theta}_{ML} \xrightarrow{\text{probab.}} \theta}$.

In the following, we assume $\theta \neq 0$.

Then $\mathbb{P}(\hat{\theta}_{ML} = 0) = \mathbb{P}(\forall i, X_i = 0) = e^{-n\theta}$

Thus $\mathbb{1}_{\hat{\theta}_{ML} > 0} \xrightarrow{\text{probab}} 1$

see remark at the last page!

As a consequence; using **Slutsky's Lemma:**

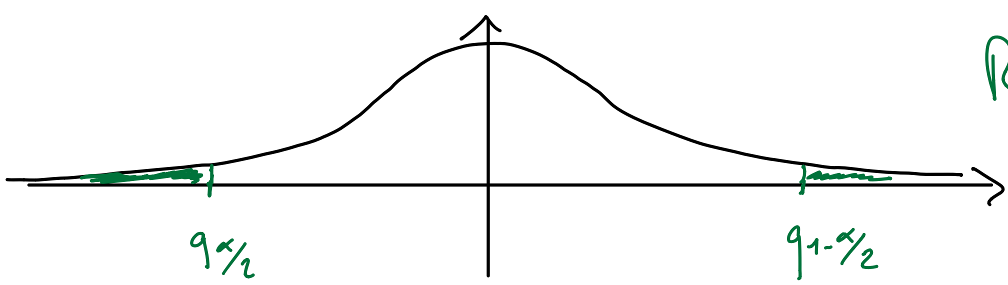
$$\left.\begin{array}{l} * \ \sqrt{n}\,(\hat{\theta}_{ML} - \theta) \overset{\text{dist}}{\rightsquigarrow} \mathcal{N}(0, \theta) \\ \\ * \ \sqrt{\hat{\theta}_{ML}} \xrightarrow[\text{proba}]{} \sqrt{\theta} \end{array}\right\} \underset{\text{Slutsky}}{\Longrightarrow} \left(\sqrt{n}(\hat{\theta}_{ML} - \theta), \sqrt{\hat{\theta}_{ML}}\right) \overset{\text{dist}}{\rightsquigarrow} \left(\mathcal{N}(0,\theta), \sqrt{\theta}\right)$$

As a consequence

$$\mathbb{1}_{\hat{\theta}_{ML} > 0} \frac{\sqrt{n}\,(\hat{\theta}_{ML} - \theta)}{\sqrt{\hat{\theta}_{ML}}} \overset{\text{dist}}{\rightsquigarrow} \mathcal{N}(0, 1)$$

because $f : x, y \longmapsto \frac{x}{y}$ is a continuous function on $\mathbb{R} \times \mathbb{R}_+^*$.

7) Define $q_{\alpha/2}$ and $q_{1-\alpha/2}$ the quantiles $\alpha/2$ and $1-\alpha/2$ of a standard gaussian distribution $\mathcal{N}(0, 1)$:



Remember $q_{\alpha/2} = -q_{1-\alpha/2}$

By question 6, $\mathbb{P}\left(\frac{\sqrt{n}(\hat{\theta}_{ML} - \theta)}{\sqrt{\hat{\theta}_{ML}}} \in [q_{\alpha/2} ; q_{1-\alpha/2}] \right) \xrightarrow[n \to +\infty]{} 1 - \alpha$

Thus $\mathbb{P}\left(\sqrt{n} \left(\hat{\theta}_{ML} - \theta \right) \in \left[q_{\alpha/2} \sqrt{\hat{\theta}_{ML}} ; q_{1-\alpha/2} \sqrt{\hat{\theta}_{ML}} \right] \right) \xrightarrow[n \to +\infty]{} 1 - \alpha$

i.e., $\mathbb{P}\left(\theta \in \left[\hat{\theta}_{ML} + \frac{\sqrt{\hat{\theta}_{ML}}}{\sqrt{n}} q_{\alpha/2} ; \hat{\theta}_{ML} + \frac{\sqrt{\hat{\theta}_{ML}}}{\sqrt{n}} q_{1-\alpha/2} \right] \right) \xrightarrow[n \to +\infty]{} 1 - \alpha$

$$\boxed{\left[a_n(\alpha, (X_i)_i) ; b_n(\alpha, (X_i)_i) \right] := \left[\hat{\theta}_{ML} + \frac{\sqrt{\hat{\theta}_{ML}}}{\sqrt{n}} q_{\alpha/2} ; \hat{\theta}_{ML} + \frac{\sqrt{\hat{\theta}_{ML}}}{\sqrt{n}} q_{1-\alpha/2} \right]}$$

is thus an asymptotic confidence interval of level α.

(remark that if $\theta = 0$, then $\hat{\theta}_{ML} = 0$ a.s., and this is also a confidence interval)

8) $\sqrt{n} \left(\hat{\theta}_{ML} - \theta \right) \xrightarrow{d} \mathcal{N}(0, \theta)$.

Thus, by δ method, with $g(x) = 2\sqrt{x}$, and $(g'(x))^2 = \frac{1}{x}$, $\forall x \in \mathbb{R}_+^*$

$$\boxed{\sqrt{n} \left(2\sqrt{\hat{\theta}_{ML}} - 2\sqrt{\theta} \right) \xrightarrow{d} \mathcal{N}(0, 1).}$$

9) Using question 8, we have:

$$\mathbb{P}\left(2\sqrt{n} \left(\sqrt{\hat{\theta}_{ML}} - \sqrt{\theta} \right) \in \left[q_{\alpha/2} ; q_{1-\alpha/2} \right] \right) \xrightarrow[n \to +\infty]{} 1 - \alpha$$

i.e., $\lim\limits_{n \to +\infty} \mathbb{P}\left(\sqrt{\theta} \in \left[\sqrt{\hat{\theta}_{ML}} + \frac{q_{\alpha/2}}{2\sqrt{n}} ; \sqrt{\hat{\theta}_{ML}} + \frac{q_{1-\alpha/2}}{2\sqrt{n}} \right] \right) = 1 - \alpha$

thus, $\liminf\limits_{n \to +\infty} \mathbb{P}\left(\theta \in \left[\left(\sqrt{\hat{\theta}_{ML}} + \frac{q_{\alpha/2}}{2\sqrt{n}} \right)^2 ; \left(\sqrt{\hat{\theta}_{ML}} + \frac{q_{1-\alpha/2}}{2\sqrt{n}} \right)^2 \right] \right) \geq 1 - \alpha$

Technical details

here we have used that $\sqrt{x} \in [\alpha ; \beta]$ and $\alpha \geq 0 \implies x \in [\alpha^2 ; \beta^2]$, and used

$\mathbb{P}\left(\sqrt{\theta} \in [\alpha ; \beta] \right) \leq \mathbb{P}\left(\sqrt{\theta} \in [\alpha ; \beta] \text{ and } \underbrace{\sqrt{\hat{\theta}_{ML}} + \frac{q_{\alpha/2}}{2\sqrt{n}}}_{\text{"}\alpha\text{"}} > 0 \right) + \mathbb{P}\left(\sqrt{\hat{\theta}_{ML}} + \frac{q_{\alpha/2}}{2\sqrt{n}} < 0 \right)$

$\leq \mathbb{P}\left(\theta \in [\alpha^2 ; \beta^2] \right) + \mathbb{P}\left(\sqrt{\hat{\theta}_{ML}} + \frac{q_{\alpha/2}}{2\sqrt{n}} < 0 \right)$

Thus $\mathbb{P}\left(\theta \in [\alpha^2 ; \beta^2] \right) \geq \underbrace{\mathbb{P}\left(\sqrt{\theta} \in [\alpha ; \beta] \right)}_{\to 1 - \alpha} - \underbrace{\mathbb{P}\left(\sqrt{\hat{\theta}_{ML}} + \frac{q_{\alpha/2}}{2\sqrt{n}} < 0 \right)}_{\to 0}$

As a consequence, the interval

$$\left[c_n\left(\alpha,(X_i)_i\right) ; d_n\left(\alpha,(X_i)\right)\right] := \left[\left(\sqrt{\hat{\theta}_{ML}} + \frac{q_{\alpha/2}}{2\sqrt{n}}\right)^2 ; \left(\sqrt{\theta_{ML}} + \frac{q_{1-\alpha/2}}{2\sqrt{n}}\right)^2\right]$$

is an asymptotic confidence interval of level α.

10) We have that, if $X \sim \mathcal{P}(\theta)$, then:

$\mathbb{E}[X] = \theta$

as a consequence, we can use $\hat{\theta}_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ as an estimator of θ derived from methods of moments using the first order moment.

Using the second order moment, we note that: $\mathbb{E}[X^2] = \theta + \theta^2$ and $Var(X) = \theta$.

We can thus use the following 3 estimators

* $\hat{\theta}_{2,1} = \hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$

Empirical Variance estimator

remark: intuitively, we have use $\theta = \mathbb{E}(X^2) - \theta^2$

replaced by 1st moment estimator. $\overline{X}$

or * $\hat{\theta}_2$ / $\hat{\theta}_2^{\,2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\sum_{i=1}^{n} X_i$

$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - X_i$

Thus $\boxed{\hat{\theta}_{2,2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\underbrace{(X_i^2 - X_i)}_{\geq 0}}}$ because $X_i \in \mathbb{N}$.

or $x \quad \hat{\theta}_2^2 + \hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \hat{m}_2$

We then need to solve $x^2 + x - \hat{m}_2 = 0 \qquad \Delta = 1 + 4\hat{m}_2$

$$x_+ = \frac{-1 + \sqrt{1 + 4\hat{m}_2}}{2} > 0 \qquad x_- = \frac{-1 - \sqrt{1+4\hat{m}_2}}{2} < 0$$

$$\Longrightarrow \boxed{\hat{\theta}_{2,3} = \frac{-1 + \sqrt{1 + 4\hat{m}_2}}{2}}.$$

We note that $\hat{\theta}_1 = \hat{\theta}_{ML}$

11) With $\hat{\theta}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i$, we have:

$$\mathbb{E}\left[\hat{\theta}_{ML}\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mathbb{E}[X_1] = \theta.$$

Thus $\hat{\theta}_{ML}$ is an unbiased estimator.

Moreover $\mathrm{Var}\left(\hat{\theta}_{ML}\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \underset{(X_i)_{i=1..n} \text{ independant}}{=} \frac{1}{n}\mathrm{Var}(X_1) = \frac{\theta}{n}$

As a consequence, we have that the quadratic risk of $\hat{\theta}_{ML}$ is:

$$\boxed{QR\left(\hat{\theta}_{ML}\right) = \left(\mathrm{Bias}\left(\hat{\theta}_{ML}\right)\right)^2 + \mathrm{Var}\left(\hat{\theta}_{ML}\right) = \frac{\mathrm{Var}\, X_1}{n} = \frac{\theta}{n}}$$

[2] The Fisher information for one observation, $I_1(\theta)$, can be computed as:

$$I_1(\theta) = \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2} \ln p(x,\theta)\right] = \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\left(-\theta + X\ln\theta\right)\right]$$

$$= \mathbb{E}\left[+\frac{X}{\theta^2}\right] = \frac{1}{\theta} \quad , \quad \text{as } \mathbb{E}[X] = \theta$$

The Fisher information of an n-sample $(X_1, \dots X_n)$ is thus $I_n(\theta) = \frac{n}{\theta}$.

As a consequence, $\hat{\theta}_{ML}$ achieves the Cramer Rao bound:

1. it is an unbiased estimator.
2. it has minimal variance : $\operatorname{Var}(\hat{\theta}_{ML}) = \frac{1}{I_n(\theta)} = \frac{\theta}{n}$.

[3] We consider $\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

This is one of the 3 possible estimators using the second order moment of X.

$$\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \theta + \theta - \overline{X}\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i-\theta)^2 + \frac{2}{n}\sum_{i=1}^{n}(X_i-\theta)(\theta-\overline{X}) + \frac{1}{n}\sum_{i=1}^{n}(\theta-\overline{X})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i-\theta)^2 + 2(\overline{X}-\theta)(\theta-\overline{X}) + (\theta-\overline{X})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i-\theta)^2 - (\overline{X}-\theta)^2$$

Which is the desired inequality

14) We have already computed
$$\mathbb{E}\left(\theta - \overline{X}_n\right)^2 = \operatorname{Var}\left(\overline{X}_n\right) = \frac{\theta}{n}.$$

$$\mathbb{E}\left[\widehat{\theta}_2\right] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\mathbb{E}\left(X_i - \theta\right)^2}_{\operatorname{Var}(X_i) = \theta.} - \frac{\theta}{n}$$
$$= \theta - \frac{\theta}{n} = \frac{n-1}{n}\,\theta.$$

This proves that $\widehat{\theta}_2$ is a biased estimator if $\theta \neq 0$.
To get an unbiased estimator, we consider:
$$\widetilde{\theta}_2 = \frac{n}{n-1}\widehat{\theta}_2 = \underbrace{\frac{1}{n-1}}_{\text{"classical" correction for unbiased variance estimation.}} \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2.$$

15) We have that:
$$\widehat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \theta\right)^2 - \left(\overline{X} - \theta\right)^2$$

By the central limit theorem, we have that, for the first term in the equation above:
$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \theta\right)^2 - \underbrace{\mathbb{E}\left[\left(X_1 - \theta\right)^2\right]}_{\theta}\right) \xrightarrow{\text{dist}} \mathcal{N}\left(0, \operatorname{Var}\left(X_1 - \theta\right)^2\right)$$

Moreover, for the second part of the decomposition,

$$\sqrt{n}\,(\bar{X}-\theta)^2 = \underbrace{\underbrace{\sqrt{n}\,(\bar{X}_n-\theta)}_{\overset{\text{distrib}}{\rightsquigarrow}\ \mathcal{N}(0,\theta)}}_{\text{by CLT}} \times \underbrace{\underbrace{(\bar{X}_n-\theta)}_{\xrightarrow{\text{proba}}\ 0}}_{\text{by LLN.}}$$

By Slutsky's lemma (one convergence in probability and one convergence in distribution), and that $x, y \mapsto xy$ is continuous on $\mathbb{R}\times\mathbb{R}$, we have that:

$$\sqrt{n}\,(\bar{X}-\theta)^2 \overset{\text{dist}}{\rightsquigarrow} 0. \quad \text{and thus} \quad \sqrt{n}\,(\bar{X}-\theta)^2 \xrightarrow[\text{probability}]{} 0$$

using the fact that a convergence in distribution to a constant is equivalent to a convergence in probability to this constant.

We can thus use Slutsky's Lemma again:

* $\sqrt{n}\,(\bar{X}-\theta)^2 \xrightarrow[\text{probability}]{} 0$
* $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\theta)^2 - \theta\right) \overset{\text{dist}}{\rightsquigarrow} \mathcal{N}\left(0, \mathrm{Var}(X_1-\theta)^2\right)$

and the fact that $x, y \mapsto x+y$ is a continuous function on $\mathbb{R}\times\mathbb{R}$

thus $\sqrt{n}\,(\hat{\theta}_2-\theta) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\theta)^2 - \theta\right) + \sqrt{n}\,(\bar{X}-\theta)$

$$\sqrt{n}\,(\hat{\theta}_2-\theta) \overset{\text{distrib}}{\rightsquigarrow} \mathcal{N}(0,\ 2\theta^2+\theta)$$

using $\mathrm{Var}\left((X_1-\theta)^2\right) = 2\theta^2+\theta$.

Finally, we remark that the limit variance, $2\theta^2 + \theta$ is larger that the minimal variance θ.

16) We compute:

$$G_X(s) = \mathbb{E}\left[e^{sX}\right] = \sum_{i=0}^{\infty} \mathbb{P}(X=i)\, e^{si}$$

$$= \sum_{i=0}^{\infty} \frac{e^{-\theta}\, \theta^i e^{si}}{i!} = e^{-\theta} \sum_{i=0}^{\infty} \frac{(\theta e^s)^i}{i!}$$

$$= e^{-\theta} e^{\theta e^s} = e^{\theta(e^s - 1)}$$

We have that ⊛ $G_X'(s) = \left(e^{\theta(e^s-1)}\right) \theta e^s$

and $G_X'(0) = \mathbb{E}(X) = \theta.$

⊛ $G_X''(s) = \theta e^s \left(e^{\theta(e^s-1)}\right) + (\theta e^s)^2 \left(e^{\theta(e^s-1)}\right)$

and $G_X''(0) = \mathbb{E}(X^2) = \theta + \theta^2$

⊛ $G_X'''(s) = \theta e^s \left(e^{\theta(e^s-1)}\right) + (\theta e^s)^2 \left(e^{\theta(e^s-1)}\right) + 2\theta^2 e^{2s} e^{\theta(e^s-1)} + (\theta e^s)^3 e^{\theta(e^s-1)}$

$G_X'''(0) = \mathbb{E}[X^3] = \theta + \theta^2 + 2\theta^2 + \theta^3 = \theta^3 + 3\theta^2 + \theta.$

⊛ $G_X^{(4)}(s) = \theta e^s e^{\theta(e^s-1)} + (\theta e^s)^2 e^{\theta(e^s-1)} + 2\theta^2 e^{2s} e^{\theta(e^s-1)}$
$+ (\theta e^s)^3 e^{\theta(e^s-1)} + 4\theta^2 e^{2s} e^{\theta(e^s-1)} + 2\theta^3 e^{3s} e^{\theta(e^s-1)}$
$+ 3(\theta e^s)^3 e^s e^{\theta(e^s-1)} + (\theta e^s)^3 \theta e^{\theta(e^s-1)}$

$$G_X^{4}(0) = \theta + \theta^2 + 2\theta^2 + \theta^3 + 4\theta^2 + 2\theta^3 + 3\theta^3 + \theta^4$$

$$\boxed{\begin{aligned} &= \theta^4 + 6\theta^3 + 7\theta^2 + \theta. \\ &= \mathbb{E}[X^4] \end{aligned}}$$

<u>And</u> $\mathrm{Var}\left((X-\theta)^2\right) = \mathbb{E}(X-\theta)^4 - \Big(\overbrace{\mathbb{E}(X-\theta)^2}^{\theta}\Big)^2$

$$= \mathbb{E}X^4 - 4\theta\,\mathbb{E}X^3 + 6\theta^2\,\mathbb{E}X^2 - 4\theta^3\,\mathbb{E}(X) + \theta^4 - \theta^2$$

$$\begin{aligned} = {} & \theta^4 + 6\theta^3 + 7\theta^2 + \theta. \\ & - 4\theta(\theta^3 + 3\theta^2 + \theta) \\ & + 6\theta^2(\theta^2 + \theta) \\ & - 4\theta^3\theta \\ & + \theta^4 - \theta^2 \end{aligned}$$

$$\begin{aligned} = {} & \theta^4 + 6\theta^3 + 7\theta^2 + \theta \\ & - 4\theta^4 - 12\theta^3 - 4\theta^2 \\ & + 6\theta^4 + 6\theta^3 \\ & - 4\theta^4 \\ & + \theta^4 - \theta^2 \end{aligned}$$

$$= 0\theta^4 - 0\theta^3 + 2\theta^2 + \theta = 2\theta^2 + \theta.$$

We have proved $\boxed{\mathrm{Var}\left((X-\theta)^2\right) = 2\theta^2 + \theta}$

Remark: Slutsky's Lemma.

There are several ways of describing Slutsky's Lemma: we use it as follows:

If $\begin{cases} X_n \overset{\text{distr}}{\rightsquigarrow} X \\ Y_n \xrightarrow{\text{probab}} c, \text{ with } c \text{ a constant} \end{cases}$. $\left[Y_n \xrightarrow{\text{probab}} c \Longrightarrow Y_n \overset{\text{dist}}{\rightsquigarrow} c \right]$

Then $(X_n, Y_n) \overset{\text{dist}}{\rightsquigarrow} (X, c)$

Informally, if we have * one cvgce in distrib, * one cvgce in probab. to a constant, then the pair converges in distribution

As a consequence, for any $f: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ continuous

$$f(X_n, Y_n) \overset{\text{dist}}{\rightsquigarrow} f(X, c)$$ (continuous mapping theorem)

For example,

$f_1(x, y) = x + y \Longrightarrow X_n + Y_n \overset{\text{dist}}{\rightsquigarrow} X + c$

$f_2(x, y) = x\, y \Longrightarrow X_n Y_n \overset{\text{dist}}{\rightsquigarrow} X c$

$f_3(x, y) = \frac{x}{y} \Longrightarrow \frac{X_n}{Y_n} \overset{\text{dist}}{\rightsquigarrow} \frac{X}{c}$

We have used the 3 examples in the homework.

we need $c \neq 0$, and f_3 is C^0 over $\mathbb{R} \times \mathbb{R}^*$