TASK 1

HPCSE HW2 – Beat Hurmann

/a

<u>Given:</u> $X \sim \mathcal{N}(\mu, \sigma^2)$

with PDF $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

<u>To do:</u>

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx =$$

$$= \left[u := \frac{x-\mu}{\sigma} \; ; \; du = \frac{dx}{\sigma} \right] =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (u\sigma + \mu)\sigma \, e^{-\frac{u^2}{2}} du$$

$$= \underbrace{\cancel{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u\sigma e^{-\frac{u^2}{2}} du}}_{\to 0} + \frac{\mu}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du$$

$$= \frac{\mu}{\cancel{\sqrt{2\pi}}} \cdot \cancel{\sqrt{2\pi}} = \underline{\underline{\mu}}$$

$$\mathbb{E}((X-\mu)^2) = \mathrm{Var}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu)^2 \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx =$$

$$= \left[u = \frac{x-\mu}{\sigma} \right] =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (u\sigma + \cancel{\mu} - \cancel{\mu})^2 \sigma \cdot e^{-\frac{u^2}{2}} du =$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2}} du = \frac{\sigma^2}{\cancel{\sqrt{2\pi}}} \cdot \frac{\sqrt{\pi} \cdot 2^{3/2}}{2} = \underline{\underline{\sigma^2}}$$

/b

GIVEN: LAPLACE DISTRIBUTION

W/ PDF: $f_X(x) = \frac{1}{2\beta} e^{-\frac{|x-\mu|}{\beta}}$

TODO:

i) $\triangleright$ $F_X(x) = \int_{-\infty}^{x} f_X(x')dx' = \int_{-\infty}^{x} \frac{1}{2\beta} e^{-\frac{|x'-\mu|}{\beta}} dx' =$

CASE 1: $x \leq \mu$: THEN $|x'-\mu| = \mu - x'$ FOR $-\infty < x' \leq x$

$\Rightarrow F_X(x) = \frac{1}{2\beta} \int_{-\infty}^{x} e^{\frac{x'-\mu}{\beta}} dx' = \frac{1}{2\cancel{\beta}} \cdot \cancel{\beta} \cdot e^{\frac{x-\mu}{\beta}} = \underline{\underline{\frac{1}{2} \cdot e^{\frac{x-\mu}{\beta}}}}$

CASE 2: $x \geq \mu$:

THEN: $F_X(x) = \int_{-\infty}^{x} \frac{1}{2\beta} e^{-\frac{|x'-\mu|}{\beta}} dx' = \underbrace{\int_{-\infty}^{\mu} \frac{1}{2\beta} e^{-\frac{|x'-\mu|}{\beta}} dx'}_{= F_X(x)|_{x=\mu} = \frac{1}{2}} + \int_{\mu}^{x} \frac{1}{2\beta} e^{-\frac{|x'-\mu|}{\beta}} dx'$

$= \frac{1}{2} + \int_{\mu}^{x} \frac{1}{2\beta} e^{-\frac{x'-\mu}{\beta}} dx' =$

$= \frac{1}{2} + \frac{1}{2\beta} (-\beta) \cdot \left[e^{-\frac{x'-\mu}{\beta}} \right]_{\mu}^{x} = \frac{1}{2} - \frac{1}{2}\left(e^{-\frac{x-\mu}{\beta}} - 1 \right) =$

$= \underline{\underline{1 - \frac{1}{2} e^{-\frac{x-\mu}{\beta}}}}$

$$\Rightarrow F_X(x) = \begin{cases} \frac{1}{2} e^{\frac{x-\mu}{\beta}} & \text{FOR } x \leq \mu \\ 1 - \frac{1}{2} e^{-\frac{x-\mu}{\beta}} & \text{FOR } x \geq \mu \end{cases}$$

ii) ▸ FOR MEDIAN m WE HAVE THAT $\int_{-\infty}^{m} f_X(x)\,dx = F_X(m) = 1/2$.

SO WITH THE ABOVE: $F_X(m) = \frac{1}{2} e^{\frac{m-\mu}{\beta}} = \frac{1}{2}$

$$\Rightarrow \quad \frac{m-\mu}{\beta} = 0 \quad \Leftrightarrow \quad \underline{\underline{m = \mu}}$$ (EITHER CASE IS FINE AS $m=\mu$)

1c GIVEN: $Q = \frac{X}{Y}$, $f_Q(q) = \int_{-\infty}^{\infty} |x| \cdot f_{X,Y}(qx, x)\,dx$

W/ $f_{X,Y}$ JOINT PDF OF X, Y

X, Y INDEPENDENT; $f_X(x) = \mathcal{N}(x \mid 0, \sigma_x^2)$, $f_Y(y) = \mathcal{N}(y \mid 0, \sigma_y^2)$

i) ▸ $f_{X,Y}(x,y) = f_X(x)\, f_Y(y) =$

$$= \frac{1}{\sqrt{2\pi\sigma_x^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_y^2}} \cdot e^{-\frac{x^2}{2\sigma_x^2}} \cdot e^{-\frac{y^2}{2\sigma_y^2}} =$$

$$= \underline{\underline{\frac{1}{2\pi\sigma_x\sigma_y} \cdot e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)}}}$$

ii) ▸ CAUCHY-DISTRIBUTION W/ $x_0 = 0$ (LOCATION PARAMETER), AND $\gamma = \sigma_x / \sigma_y$

$$f(x) = \frac{1}{\pi} \cdot \frac{\sigma_x}{\sigma_y} \cdot \frac{1}{x^2 + \frac{\sigma_x^2}{\sigma_y^2}}$$

FROM THE ABOVE:

$$f_Q(q) = \int_{-\infty}^{\infty} |y|\, f_{X,Y}(qy, y)\,dy =$$

$$= \int_{-\infty}^{\infty} |y| \cdot \frac{1}{2\pi\sigma_x\sigma_y} \cdot e^{-\frac{1}{2}\left(\frac{(qy)^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)} dy =$$

$$= \frac{1}{2\pi\sigma_x\sigma_y} \int_{-\infty}^{\infty} |y| \cdot e^{-\frac{1}{2}\left(\frac{q^2y^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)} dy =$$

$$= \frac{1}{2\pi\sigma_x\sigma_y} \cdot \left[\int_{-\infty}^{0} -y \cdot e^{-\frac{1}{2}\left(\frac{q^2(-y)^2}{\sigma_x^2} + \frac{(-y)^2}{\sigma_y^2}\right)} dy + \int_{0}^{\infty} y \cdot e^{-\frac{1}{2}\left(\frac{q^2y^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)} dy \right] =$$

$$= \frac{1}{\cancel{2}\pi\sigma_x\sigma_y} \cdot \cancel{2} \cdot \int_{0}^{\infty} y \cdot e^{-\frac{y^2}{2}\underbrace{\left(\frac{q^2}{\sigma_x^2} + \frac{1}{\sigma_y^2}\right)}_{=:\frac{1}{\sigma^2(q)}}} dy =$$

$$= \frac{1}{\pi\sigma_x\sigma_y} \cdot \int_{0}^{\infty} y\, e^{-\frac{y^2}{2\sigma^2(q)}} dy = \frac{1}{\pi\sigma_x\sigma_y} \left[-\sigma^2(q)\, e^{-\frac{y^2}{2\sigma^2(q)}} \right]_{0}^{\infty} =$$

$$= \frac{1}{\pi\sigma_x\sigma_y} \cdot \sigma^2(q)$$

FINALLY REINSERTING $\sigma^2(q) = \dfrac{1}{\frac{q^2}{\sigma_x^2} + \frac{1}{\sigma_y^2}}$ YIELDS:

$$f_Q(q) = \frac{1}{\pi\sigma_x\sigma_y} \cdot \frac{1}{\frac{q^2}{\sigma_x^2} + \frac{1}{\sigma_y^2}} = \frac{1}{\pi} \cdot \underset{\gamma}{\underbrace{\frac{\sigma_x}{\sigma_y}}} \cdot \frac{1}{q^2 + \underset{\gamma^2}{\underbrace{\frac{\sigma_x^2}{\sigma_y^2}}}}$$

WHICH IS THE DESIRED CAUCHY DISTRIBUTION WITH LOCATION PARAMETER $x_0 = 0$ AND SCALE $\gamma = \dfrac{\sigma_x}{\sigma_y}$

TASK 2

GIVEN: $\vec{d} = \{d_i\}_{i=1}^{N}$, $d_i \in \mathbb{R}$

$d_i \sim \mathcal{N}(\mu, 1)$; d_i INDEPENDENT

/a

TODO: $L(\mu) = p(d \mid \mu)$

AS d_i ARE INDEPENDENT AND WITH

$$f_X(d_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(d_i-\mu)^2}{2\sigma^2}}$$ AND $\sigma = 1$ WE GET:

$$L(\mu) = \prod_{i=1}^{N} f_X(d_i \mid \mu) = \prod_{i=1}^{N} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(d_i-\mu)^2}{2}} \right) =$$

$$= (2\pi)^{-\frac{n}{2}} \cdot e^{-\frac{1}{2} \cdot \sum_{i=1}^{N} (d_i-\mu)^2}$$

/b

TO DO: FIND $MLE(\mu)$

SET $\log L(\mu) =: \mathcal{L}(\mu) \equiv$ THE LOG-LIKELIHOOD FUNCTION OF μ;

WITH THE MONOTONY OF THE LOGARITHM: $\arg\max_{\mu} L(\mu) = \arg\max_{\mu} \mathcal{L}(\mu)$;

$$\hat{\mu} = MLE(\mu) = \arg\max_{\mu} \mathcal{L}(\mu) = \arg\max_{\mu} \left(-\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{N} (d_i - \mu)^2 \right)$$

WITH THE FIRST ORDER CONDITION FOR A MAXIMUM: $\frac{\partial}{\partial \mu} \mathcal{L} \stackrel{!}{=} 0$:

$$\frac{\partial}{\partial \mu} \mathcal{L}(\mu) = \frac{\partial}{\partial \mu} \left(-\frac{1}{2} \sum_{i=1}^{N} (d_i-\mu)^2 \right) = \sum_{i=1}^{N} (d_i - \mu) \stackrel{!}{=} 0 \iff N\mu \stackrel{!}{=} \sum_{i=1}^{N} d_i$$

$$\Rightarrow \underline{\underline{\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} d_i =: \bar{d}}}$$

ALSO, $\frac{\partial}{\partial \sigma^2}\left[\log\left(\prod_{i=1}^{N} f_x(d_i)\right)\right] = \frac{\partial}{\partial \sigma^2}\left[\log\left((2\pi\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(d_i-\mu)^2}\right)\right] =$

$$= \frac{\partial}{\partial \sigma^2}\left[-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(d_i-\mu)^2\right] =$$

$$= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{N}(d_i-\mu)^2 = \frac{1}{2\sigma^2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{N}(d_i-\mu)^2 - n\right] \overset{!}{=} 0$$

FOR $\sigma \neq 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(d_i-\hat{\mu})^2 = \frac{\sigma^2}{N} = \frac{1}{N}$ FOR $\sigma = 1$

IN SUMMARY, $d \sim \mathcal{N}(\bar{d}, \frac{1}{N})$

/c <u>TO DO</u>: UPDATE POSTERIOR $p(\mu|d)$,

FIND MEAN, VARIANCE OF $p(\mu|d)$.

USING BAYES' THEOREM: POSTERIOR $\propto$ LIKELIHOOD $\times$ PRIOR

SO: $p(\mu|d) \propto p(d|\mu) \cdot p(\mu)$

WITH $p(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \cdot e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}$ AS $\mu \sim \mathcal{N}(\mu_0, \sigma_0)$

AND THE LIKELIHOOD FROM ABOVE:

$$p(\mu|d) \propto \left(2\pi \cdot \frac{1}{n}\right)^{-\frac{n}{2}} \cdot e^{-\frac{n}{2}\cdot(\mu-\bar{d})^2} \cdot \frac{1}{\sqrt{2\pi\sigma_0^2}} \cdot e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} =$$

$$= (2\pi)^{-\frac{n+1}{2}} \cdot \frac{1}{\sigma_0} \cdot e^{-\frac{1}{2}\left(n(\mu-\bar{d})^2 + \frac{(\mu-\mu_0)^2}{\sigma_0^2}\right)} =$$

$$= \frac{(2\pi)^{-\frac{n+1}{2}}}{\sigma_0} \cdot \exp\left(-\frac{1}{2}\left[n\left(\mu^2 - 2\mu\bar{d} + \bar{d}^2\right) - \frac{1}{\sigma_0^2}\left(\mu^2 - 2\mu\mu_0 + \mu_0^2\right)\right]\right)$$

$$= \frac{(2\pi)^{-\frac{n+1}{2}}}{\sigma_0} \cdot \exp\left(-\frac{1}{2}\underbrace{\left[\mu^2\left(n+\frac{1}{\sigma_0^2}\right) - 2\mu\left(\bar{d}n + \frac{\mu_0}{\sigma_0^2}\right) + n\bar{d}^2 + \frac{\mu_0^2}{\sigma_0^2}\right]}\right)$$

FOR μ_n, σ_n^2 = MEAN, VARIANCE OF POSTERIOR $p(\mu|d)$, THIS SHOULD LOOK LIKE:

$$-\frac{1}{2\sigma_n^2}\left[\mu - \mu_n\right]^2 = -\frac{1}{2\sigma_n^2}\left[\mu^2 - 2\mu\mu_n + \mu_n^2\right]$$

$\Rightarrow$ COMPARE COEFFICIENTS:

$\boxed{\mu_n}$

$$\frac{\mu\mu_n}{\sigma_n^2} \overset{!}{=} \mu\left(\bar{d}n + \frac{\mu_0}{\sigma_0^2}\right) \Leftrightarrow \frac{\mu_n}{\sigma_n^2} = \frac{\bar{d}n\sigma_0^2 + \mu_0}{\sigma_0^2}$$

$$\Rightarrow \mu_n = \sigma_n^2 \cdot \left(\frac{\mu_0 + \bar{d}n\sigma_0^2}{\sigma_0^2}\right) \qquad (\mathrm{I})$$

$\boxed{\sigma_n^2}$

$$-\frac{1}{2\sigma_n^2} \cdot \mu^2 \overset{!}{=} -\frac{1}{2}\mu^2\left(n + \frac{1}{\sigma_0^2}\right) \Leftrightarrow \frac{1}{\sigma_n^2} = n + \frac{1}{\sigma_0^2}$$

$$\Rightarrow \sigma_n^2 = \frac{\sigma_0^2}{n\sigma_0^2 + 1} \qquad (\mathrm{II})$$

$$\Leftrightarrow \underline{\underline{\frac{1}{\sigma_n^2} = n + \frac{1}{\sigma_0^2}}}$$ FOR THE VARIANCE SOUGHT,

AND FINALLY FOR THE DESIRED EXPECTATION VALUE:

$$(\mathrm{II}) \text{ in } (\mathrm{I}): \quad \mu_n = \left(\frac{\sigma_0^2}{n\sigma_0^2+1}\right)\left(\frac{\mu_0 + \bar{d}n\sigma_0^2}{\sigma_0^2}\right) =$$

$$= \frac{\mu_0 + \bar{d}n\sigma_0^2}{n\sigma_0^2 + 1}$$

$$\Leftrightarrow \underline{\underline{\mu_n}} = \frac{\frac{1}{\sigma_0^2}\cdot\mu_0 + n\cdot\bar{d}}{n + \frac{1}{\sigma_0^2}} = \underline{\underline{\frac{\frac{1}{\sigma_0^2}\cdot\mu_0 + \sum_{i=1}^{N} d_i}{n + \frac{1}{\sigma_0^2}}}}$$

1d TO DO: FIND $MAP(\mu) = \underset{\mu}{\arg\max}\ p(\mu|d) = \hat{\mu}$:

MAXIMIZING:

$$p(\mu|d) \propto \left(2\pi \cdot \frac{1}{n}\right)^{-\frac{n}{2}} \cdot e^{-\frac{n}{2}\cdot(\mu-\bar{d})^2} \cdot \frac{1}{\sqrt{2\pi\sigma_0^2}} \cdot e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}$$

EQUALS MINIMIZING

$$f(\mu) := \frac{n}{2}(\mu-\bar{d})^2 + \frac{(\mu-\mu_0)^2}{2\sigma_0^2}$$

$$\Rightarrow \frac{\partial}{\partial\mu} f(\mu) = n(\mu-\bar{d}) + \frac{\mu-\mu_0}{\sigma_0^2} \overset{!}{=} 0$$

$$\Rightarrow \mu(n\sigma_0^2+1) = \bar{d}n\sigma_0^2 + \mu_0$$

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

$$\Rightarrow \underline{\underline{\hat{\mu}}} = \frac{\sigma_0^2\bar{d}\cdot n + \mu_0}{n\sigma_0^2+1} = \underline{\underline{\frac{\sigma_0^2 \cdot \sum_{i=1}^{n} d_i + \mu_0}{n\sigma_0^2+1}}}$$

WHICH MAKES SENSE AS IT'S A WEIGHTED SUM OF PRIOR & LIKELIHOOD

1e TO DO: FIND σ_n, μ_n, $\hat{\mu}$ USING UNIFORM PRIOR OVER $\mathbb{R}$:

$$p(\mu) \propto 1$$

WHICH IS IMPROPER AS $\int_{-\infty}^{\infty} p(\mu)\, d\mu = \infty$. THUS:

$$p(\mu|d) = \frac{p(d|\mu)\cdot p(\mu)}{\int_{-\infty}^{\infty} p(d|\mu)\cdot p(\mu)\, d\mu} =$$

$$= \frac{\left(2\pi\cdot\frac{1}{n}\right)^{-\frac{n}{2}} \cdot e^{-\frac{n}{2}\cdot(\mu-\bar{d})^2} \cdot 1}{\underbrace{1\cdot\int_{-\infty}^{\infty}\left(2\pi\cdot\frac{1}{n}\right)^{-\frac{n}{2}} \cdot e^{-\frac{n}{2}\cdot(\mu-\bar{d})^2}\, d\mu}_{=1 \text{ BY DEF}}} =$$

$$= \left(2\pi\cdot\frac{1}{n}\right)^{-\frac{n}{2}} \cdot e^{-\frac{n}{2}\cdot(\mu-\bar{d})^2}$$

WITH $\frac{1}{2\sigma_n^2} = \frac{n}{2} \Leftrightarrow \underline{\underline{\sigma_n^2}} = \underline{\underline{\frac{1}{n}}}$

AND $\underline{\underline{\mu_n}} = \bar{d} = \underline{\underline{\frac{1}{n}\sum_{i=1}^{n} d_i}}$

AND $\underline{\underline{\hat{\mu}}} = \arg\max_{\mu}\, p(\mu|d) = \arg\min_{\mu} \frac{n}{2}(\mu-\bar{d})^2 \Rightarrow \hat{\mu} = \bar{d} = \underline{\underline{\frac{1}{n}\sum_{i=1}^{n} d_i}}$

WHICH MEANS THAT IN THIS CASE $\underline{\underline{\text{MAP}(\mu) = \text{MLE}(\mu)}}$

WHICH WAS TO BE EXPECTED FROM A UNIFORM PRIOR.

TASK 3

GIVEN: LINEAR MODEL

$$y = \beta x + \varepsilon$$

$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$D = \{x_0, y_0\}$$

UNINFORMATIVE PRIOR FOR β

BY THE LINEARITY OF THE EXPECTATION VALUE, THE MODEL CAN BE LINEAR-TRANSFORMED INTO:

$$Y \sim \mathcal{N}(\beta x, \sigma^2) \quad \text{AS} \quad E[y \mid \beta, x] = \beta x$$

THUS:

$$p(y \mid \beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n} (y_i - \beta x_i)^2}$$

FOR THE UNINFORMATIVE PRIOR, WE CHOOSE AGAIN THE SIMPLEST PRIOR: $p(\beta, \sigma^2) \propto 1$. WE OBTAIN:

$$p(\beta, \sigma^2 \mid y) \propto p(y \mid \beta, \sigma^2) \cdot p(\beta, \sigma^2)$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta x_i)^2}$$

WITH $D = \{x_0, y_0\}$ AND $|D| = 1$:

$$p(\beta, \sigma^2 \mid D) \propto (2\pi\sigma^2)^{-\frac{1}{2}} \cdot \frac{1}{\sigma^2} \cdot e^{-\frac{1}{2\sigma^2}(y_0 - \beta x_0)^2}$$

$$\propto e^{-\frac{1}{2\sigma^2}(y_0^2 - 2y_0\beta x_0 + \beta^2 x_0^2)} \qquad (*)$$

AND THUS WITH THE INTERLUDE:

$$\hat{\beta} = \mathrm{MAP}(\beta) = \underset{\beta}{\operatorname{argmin}} \left(-2y_0 \beta x_0 + \beta^2 x_0^2\right)$$

$$\Rightarrow \frac{\partial}{\partial \beta}\left(-2y_0\beta x_0 + \beta^2 x_0\right) = -2y_0x_0 + 2\beta x_0^2 \stackrel{!}{=} 0$$

$$\Rightarrow \underline{\underline{\hat{\beta}}} = \frac{x_0 y_0}{x_0^2} = \underline{\underline{\frac{y_0}{x_0}}} = \underline{\underline{\mathrm{MAP}(\beta|D)}}$$

WHICH JUST EQUALS THE β_{OLS} FROM SIMPLE LINEAR REGRESSION AS EXPECTED FOR AN UNINFORMATIVE PRIOR.

WE CONTINUE:

$$(*) \quad \propto \quad e^{-\frac{x_0^2}{2\sigma^2}\left(\frac{y_0^2}{x_0^2} - 2\beta\frac{y_0}{x_0} + \beta^2\right)}$$

$$\propto \quad e^{-\frac{1}{2(\sigma/x_0)^2}\left(\beta - \frac{y_0}{x_0}\right)^2} \propto e^{-\frac{1}{2(\sigma/x_0)^2}\left(\beta - \hat{\beta}\right)^2}$$

$$\Rightarrow \beta|D \sim \mathcal{N}\left(\hat{\beta}, \frac{\sigma^2}{x_0^2}\right)$$

AND THUS $\underline{\underline{\sigma_\beta}} = \sqrt{\frac{\sigma^2}{x_0^2}} = \underline{\underline{\frac{\sigma}{x_0}}}$ FOR THE STD DEV OF $\beta|D$.

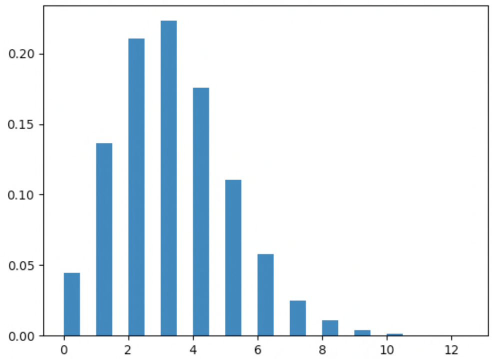
TASK 4

1/ N/A

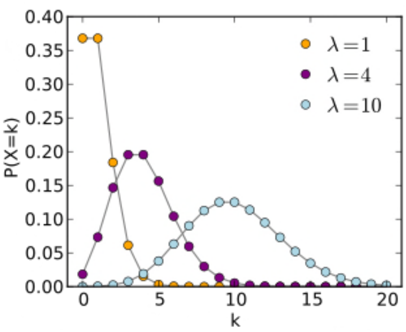
2/ N/A

3/ To get a proper PDF, the histogram needs to be normalized. This is achieved by dividing each bin count by the total number of observations (see code).

4/ Based on the normalized histogram



looking like the above and after perusing the numpy documentation and the page https://en.wikipedia.org/wiki/List_of_probability_distributions



I reckon we're dealing with a Poisson distribution with $\lambda \approx 3$. Not only does the shape match, but the Poisson distribution is common, fairly nice to handle and it fits the given data $k = 0, 1, 2, \ldots$ in

$$P(k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Finally, it is a single-parameter, single-mode distribution which is preferable when given a choice.

5/ $\underline{\underline{L(\theta)}} = \prod_{i=1}^{n} f(k_i \mid \theta) = \prod_{i=1}^{n} e^{-\theta} \cdot \frac{\theta^{k_i}}{k_i!} = \underline{\underline{e^{-n\theta} \prod_{i=1}^{n} \frac{\theta^{k_i}}{k_i!}}}$

$$\underline{\underline{l(\theta)}} = \log L(\theta) = \log\left(\prod_{i=1}^{n} e^{-\theta} \cdot \frac{\theta^{k_i}}{k_i!}\right) = \sum_{i=1}^{n} \log\left(e^{-\theta} \cdot \frac{\theta^{k_i}}{k_i!}\right) =$$

$$= \underline{\underline{-n\theta + \log(\theta) \cdot \sum_{i=1}^{n} k_i - \sum_{i=1}^{n} \log(k_i!)}}$$

SEE CODE FOR IMPLEMENTATION

6/ $\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\, L(\theta)$ CAN BE OBTAINED FROM $l(\theta)$ AS USUAL BY OBSERVING THAT log IS MONOTONIC:

$$\frac{\partial}{\partial\theta} L(\theta) \propto \frac{\partial}{\partial\theta} l(\theta) = -n + \frac{1}{\theta} \cdot \sum_{i=1}^{n} k_i \overset{!}{=} 0$$

$$\Leftrightarrow \quad \underline{\underline{\hat{\theta}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} k_i}}$$

SEE CODE FOR IMPLEMENTATION

7/ FOR THE GAUSSIAN DISTRIBUTION $\mathcal{N}(\mu, \sigma^2)$:

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (k_i - \mu)^2}$$

$$l(\mu, \sigma^2) = \log L(\mu, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (k_i - \mu)^2$$

WITH MLEs:

$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} k_i$ (WITH $E[\hat{\mu}] = \mu$) AND

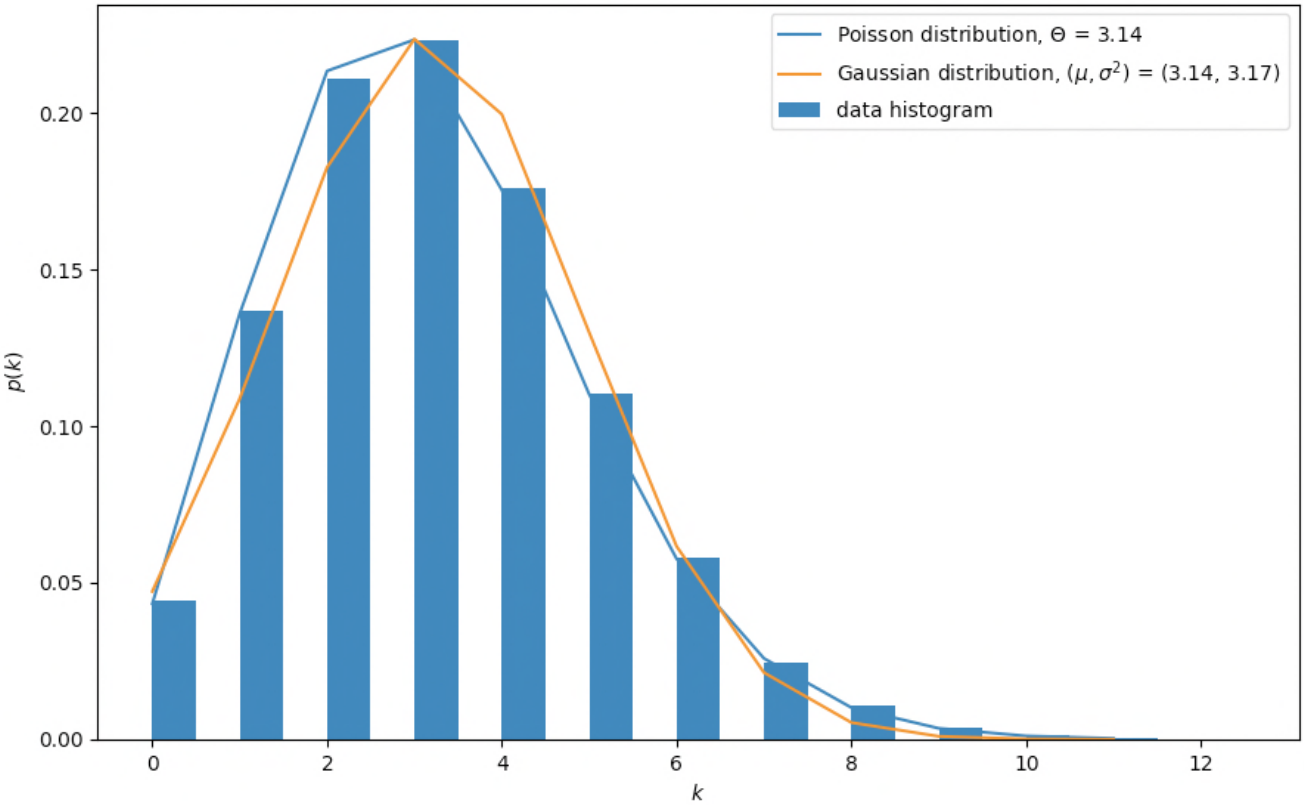
$\hat{\sigma}_2 = \frac{1}{n} \sum_{i=1}^{n} (k_i - \mu)^2$

SEE CODE FOR IMPLEMENTATION.

THE POISSON ($\lambda = \lambda_{MLE} = 3.14$) DISTRIBUTION OFFERS THE BETTER LOG-LIKELIHOOD OF -39580.99.

THE LIKELIHOODS OF BOTH DISTRIBUTIONS END UP BEING 0 — A FLOATING POINT PRECISION ERROR PROBLEM RESULTING FROM MULTIPLYING A LOT OF VERY SMALL NUMBERS. APART FROM SUMMATION BEING A CHEAPER OPERATION THAN MULTIPLICATION, NUMERICAL ERROR AVOIDANCE IS THE MAIN REASON FOR WORKING IN LOG SPACE.

8/ THE FINAL PLOT IS:



WHILE PEAKS & TAILS ARE SIMILAR, IT IS QUALITATIVELY EVIDENT THAT THE POISSON DISTRIBUTION IS A BETTER FIT (ESPECIALLY AROUND $k = 1, 2, 4$). THIS AGREES WITH THE LOG-LIKELIHOOD FINDINGS FROM 7 ABOVE.

TASK 5

1/ CAVEAT (FROM PIAZZA):

We have to restrict ourselves to pdf's with a unique global maximum (e.g. no uniform distribution) and that are C1 (continuously differentiable, e.g. no Laplace distribution)

	HOLDS
i)	N
ii)	Y
iii)	N
iv)	Y
v)	N
vi)	N

2/ $$\log p(x) = L(x) = L(\hat{x}) + \frac{1}{2}\frac{\partial^2 L}{\partial x^2}\bigg|_{\hat{x}}(x-\hat{x})^2 + O((x-\hat{x})^3)$$

WE CALCULATED AS FOLLOWS WHILE DISREGARDING HIGHER ORDER TERMS:

$$p(x) = e^{L(x)} = \exp\left(L(\hat{x}) + \frac{1}{2}\frac{\partial^2 L}{\partial x^2}\bigg|_{\hat{x}}(x-\hat{x})^2 + O((x-\hat{x})^3)\right) =$$

$$\approx \exp(L(\hat{x})) \cdot \exp\left(\frac{1}{2}\frac{\partial^2 L}{\partial x^2}\bigg|_{\hat{x}}(x-\hat{x})^2\right)$$

AND THUS $A = e^{\log p(\hat{x})} = e^{L(\hat{x})}$ FOR THE CONSTANT

3/ THIS LOOKS LIKE A GAUSSIAN DISTRIBUTION AROUND $\hat{x}$ WITH VARIANCE $\sigma^2 := -\left(\frac{\partial^2 L}{\partial x^2}\bigg|_{\hat{x}}\right)^{-1}$. THUS:

$$\underline{p(x)} \approx \underbrace{p(\hat{x})}_{=A} \cdot \overbrace{\sqrt{2\pi\sigma^2} \cdot \frac{1}{\sqrt{2\pi\sigma^2}}}^{=1} \underbrace{\cdot \exp\left(-\frac{1}{2\sigma^2}(x-\hat{x})^2\right)}_{=\mathcal{N}(\hat{x},\sigma^2)} = \underline{p(\hat{x})\sqrt{2\pi\sigma^2}\,\mathcal{N}(\hat{x},\sigma^2)}$$

4/ $$p(x|x_0,\gamma) = \frac{1}{\pi}\frac{\gamma}{(x-x_0)^2+\gamma^2}$$

AS DONE SEVERAL TIMES BEFORE FOR THE MAXIMUM, USING THE MONOTONIC PROPERTY OF LOG:

$$\frac{\partial}{\partial x} p(x|x_0\gamma) = \frac{\partial}{\partial x}\frac{1}{\pi}\frac{\gamma}{(x-x_0)^2+\gamma^2} =$$

$$\propto \frac{\partial}{\partial x}\log\left(\frac{1}{\pi}\cdot\frac{\gamma}{(x-x_0)^2+\gamma^2}\right) = \frac{\partial}{\partial x}\left(-\log(\pi)+\log(\gamma)-\log\left((x-x_0)^2+\gamma^2\right)\right) =$$

$$= -\frac{\partial}{\partial x}\log\left((x-x_0)^2+\gamma^2\right) = -\frac{1}{(x-x_0)^2+\gamma^2}\cdot 2\cdot(x-x_0) \stackrel{!}{=} 0$$

$$\Rightarrow (x-x_0) \stackrel{!}{=} 0 \Rightarrow x \stackrel{!}{=} x_0 \Rightarrow \underline{\hat{x} = x_0}$$

TO CHECK THAT THIS IS A MAXIMUM:

$$\frac{\partial^2}{\partial x^2}\bigg|_{\hat{x}=x_0}\log\left(\frac{1}{\pi}\cdot\frac{\gamma}{(x-x_0)^2+\gamma^2}\right) = \frac{\partial}{\partial x}\bigg|_{\hat{x}=x_0}\left(-\frac{1}{(x-x_0)^2+\gamma^2}\cdot 2\cdot(x-x_0)\right) =$$

$$= -2\frac{\partial}{\partial x}\bigg|_{\hat{x}=x_0}\left(\frac{x-x_0}{(x-x_0)^2+\gamma^2}\right) = -2\,\frac{(x-x_0)^2+\gamma^2-(x-x_0)\cdot 2\cdot(x-x_0)}{(x-x_0)^4+\gamma^2(x-x_0)^2+\gamma^4}\bigg|_{\hat{x}=x_0} =$$

$$= \frac{-2}{\gamma^2} < 0 \quad \text{AS SEEN @ 1/}$$

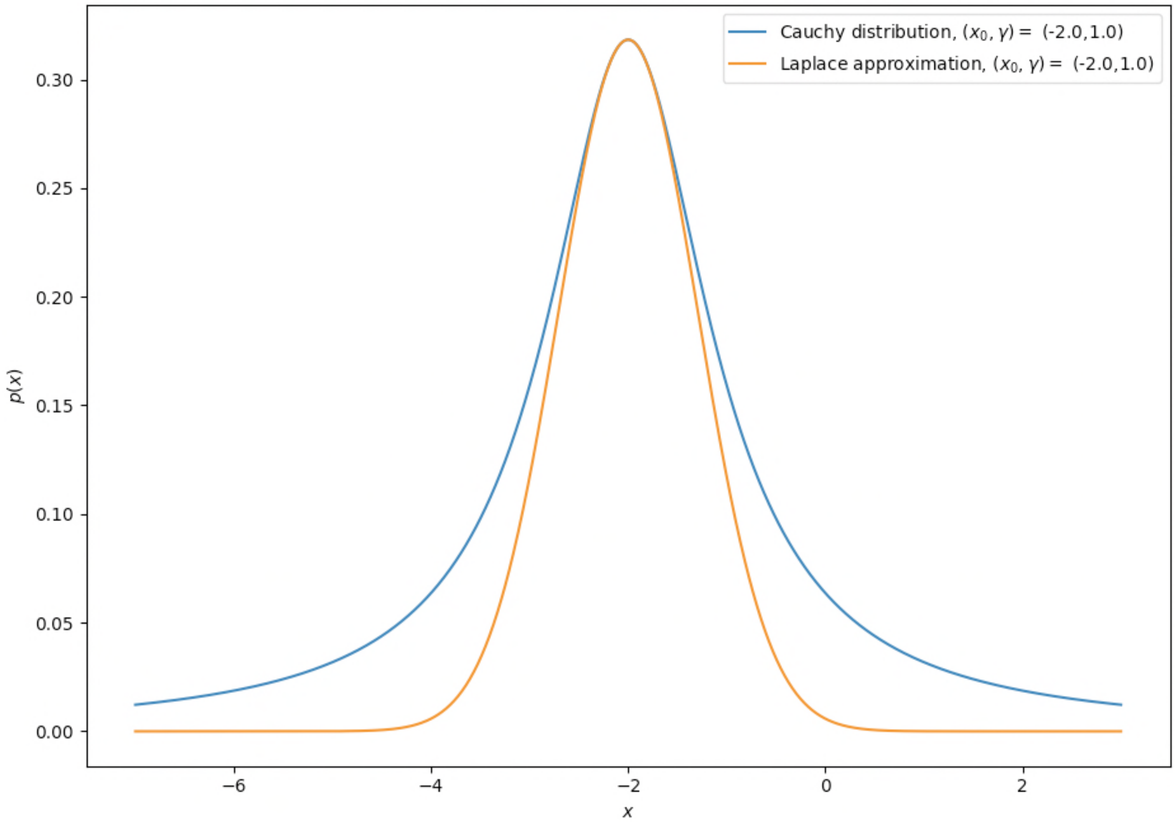
USING THE ABOVE, WE GET THE LAPLACIAN APPROXIMATION:

$$\underline{p(x)} \approx e^{L(\hat{x})}\cdot\exp\left(\frac{1}{2}\frac{\partial^2 L}{\partial x^2}\bigg|_{\hat{x}=x_0}(x-\hat{x})^2\right)$$

$$= \frac{1}{\pi\gamma}\cdot\exp\left(\frac{1}{2}\frac{-2}{\gamma^2}\cdot(x-\hat{x})^2\right) =$$

$$= \underline{\underline{\frac{1}{\pi\gamma}\cdot\exp\left(-\frac{1}{\gamma^2}(x-\hat{x})^2\right)}}$$

5/ THE PLOT GENERATED BY task5.py IS AS FOLLOWS:



CODE: task5.py

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import cauchy

x_0, gamma = -2.0, 1.0 # given by task

width = 5.0
n_points = 1000
x_range = np.linspace(x_0 - width, x_0 + width, n_points)

cauchy_distr = cauchy(x_0, gamma)
laplace = lambda x, x_0=x_0, gamma=gamma: np.exp(-(x - x_0)**2 /
gamma**2) / np.pi / gamma

plt.plot(x_range, cauchy_distr.pdf(x_range),
 label='Cauchy distribution, $(x_0, \gamma)=$ ({:.1f},{:.1f})'.format(x_0,
gamma))
plt.plot(x_range, laplace(x_range),
 label='Laplace approximation, $(x_0, \gamma)=$ ({:.1f},{:.1f})'.format(x_0,
gamma))
plt.xlabel('$x$')
plt.ylabel('$p(x)$')
plt.legend()
plt.show()
```