Machine Learning 1

Technical University Berlin
Winter semester 2020/21

Exercise Sheet 2

(16. November 2020)

Students:	Mattes Ohse	337356
	Florian Ebert	391837
	Rodrigo Alexis Pardo Meza	
	Bertty Contreras Rojas	

1 Exercise 1: Maximum-Likelihood Estimation

$$p(x,y) = \lambda\eta e^{-\lambda x - \eta y}$$
$$\lambda,\ \eta > 0$$

1a) Show that x and y are independent

x and y are independent iff $p(x,y) = g(x)h(y)$ for all x,y for some functions g and h.
$\Rightarrow$ We choose $g(x) = \lambda e^{-\lambda x}$ and $h(y) = \eta e^{-\eta y}$

$$\begin{aligned}
&g(x) * h(y) \\
&\Leftrightarrow (\lambda e^{-\lambda x}) * (\eta e^{-\eta y}) \\
&\Leftrightarrow \lambda\eta e^{-\lambda x} * e^{-\eta y} \\
&\Leftrightarrow \lambda\eta e^{-\lambda x - \eta y} \\
&\Leftrightarrow p(x,y) \\
&\Rightarrow x\ and\ y\ are\ independent
\end{aligned}$$

1b) Derive a maximum likelihood estimator of the parameter λ based on D.

$\theta_1 = \lambda \ and \ \theta_2 = \eta$

$$
\begin{aligned}
& ln(p(D \mid \theta)) = ln(\prod_{k=1}^{N} p(x_k, y_k \mid \theta)) \\
\Leftrightarrow \ & ln(p(D \mid \theta)) = ln(\prod_{k=1}^{N} \theta_1 \theta_2 e^{-\theta_1 x_k - \theta_2 y_k}) \\
\Leftrightarrow \ & ln(p(D \mid \theta)) = \sum_{k=1}^{N} ln(\theta_1) + ln(\theta_2) - \theta_1 x_k - \theta_2 y_k \\
\Leftrightarrow \ & ln(p(D \mid \theta)) = N(ln(\theta_1) + ln(\theta_2)) - \sum_{k=1}^{N} \theta_1 x_k + \theta_2 y_k \\
\Rightarrow \ & \nabla_\theta (ln(p(D \mid \theta))) = \left[\frac{N}{\theta_1} - \sum_{k=1}^{N} x_k, \ \frac{N}{\theta_2} - \sum_{k=1}^{N} y_k \right] \wedge \ best \ \hat{\theta} \ at \ \nabla_\theta (ln(p(D \mid \theta))) = 0 \\
& \Rightarrow \hat{\lambda} = \frac{\sum_{k=1}^{N} x_k}{N} = \bar{x} \\
& \hat{\eta} = \frac{\sum_{k=1}^{N} y_k}{N} = \bar{y}
\end{aligned}
$$

1c) Derive a maximum likelihood estimator of the parameter λ based on D under the constraint $\eta = \frac{1}{\lambda}$.

$\theta_1 = \lambda \; and \; \theta_2 = \eta = \frac{1}{\lambda}$

$$\begin{aligned}
ln(p(D \mid \theta)) &= ln(\prod_{k=1}^{N} p(x_k, y_k \mid \theta)) \\
\Leftrightarrow ln(p(D \mid \theta)) &= ln(\prod_{k=1}^{N} \theta_1 \theta_2 e^{-\theta_1 x_k - \theta_2 y_k}) \\
\Leftrightarrow ln(p(D \mid \theta)) &= ln(\prod_{k=1}^{N} \theta_1 \frac{1}{\theta_1} e^{-\theta_1 x_k - \frac{1}{\theta_1} y_k}), because \; \theta_2 = \eta = \frac{1}{\lambda} = \frac{1}{\theta_1} \\
\Leftrightarrow ln(p(D \mid \theta)) &= -\sum_{k=1}^{N} \theta_1 x_k + \frac{1}{\theta_1} y_k \\
\Leftrightarrow ln(p(D \mid \theta)) &= -\theta_1 \sum_{k=1}^{N} x_k - \frac{1}{\theta_1} \sum_{k=1}^{N} y_k \\
\Rightarrow \; \nabla_\theta(ln(p(D \mid \theta))) &= \left[\frac{1}{\theta_1^2} \sum_{k=1}^{N} y_k - \sum_{k=1}^{N} x_k \right] \wedge \; best \; \hat{\theta} \; at \; \nabla_\theta(ln(p(D \mid \theta))) = 0 \\
\Rightarrow \hat{\theta_1} = \hat{\lambda} &= \sqrt{\frac{\sum_{k=1}^{N} y_k}{\sum_{k=1}^{N} x_k}} \\
\hat{\theta_2} = \hat{\eta} &= \sqrt{\frac{\sum_{k=1}^{N} x_k}{\sum_{k=1}^{N} y_k}}
\end{aligned}$$

1d) Derive a maximum likelihood estimator of the parameter λ based on D under the constraint $\eta = 1 - \lambda$.

$\theta_1 = \lambda \; and \; \theta_2 = \eta = 1 - \lambda$

$$
\begin{aligned}
ln(p(D \mid \theta)) &= ln(\prod_{k=1}^{N} p(x_k, y_k \mid \theta)) \\
\Leftrightarrow ln(p(D \mid \theta)) &= ln(\prod_{k=1}^{N} \theta_1 \theta_2 e^{-\theta_1 x_k - \theta_2 y_k}) \\
\Leftrightarrow ln(p(D \mid \theta)) &= ln(\prod_{k=1}^{N} \theta_1 (1-\theta_1) e^{-\theta_1 x_k - (1-\theta_1) y_k}), because \; \theta_2 = \eta = 1 - \lambda = 1 - \theta_1 \\
\Leftrightarrow ln(p(D \mid \theta)) &= \sum_{k=1}^{N} ln(\theta_1) + ln(1-\theta_1) - \theta_1 x_k - (1-\theta_1) y_k \\
\Leftrightarrow ln(p(D \mid \theta)) &= N(ln(\theta_1) + ln(1-\theta_1)) - \theta_1 \sum_{k=1}^{N} x_k - (1-\theta_1) \sum_{k=1}^{N} y_k \\
\Rightarrow \nabla_\theta(ln(p(D \mid \theta))) &= \left[N(\frac{1}{\theta_1} - \frac{1}{1-\theta_1}) + \sum_{k=1}^{N} y_k - \sum_{k=1}^{N} x_k \right] \wedge \; best \; \hat{\theta} \; at \; \nabla_\theta(ln(p(D \mid \theta))) = 0 \\
\Rightarrow 0 &= N(\frac{1}{\hat{\theta}_1} - \frac{1}{1-\hat{\theta}_1}) + \sum_{k=1}^{N} y_k - \sum_{k=1}^{N} x_k \\
\Leftrightarrow \frac{\sum_{k=1}^{N} x_k - \sum_{k=1}^{N} y_k}{N} &= \frac{1}{\hat{\theta}_1} - \frac{1}{1-\hat{\theta}_1} \\
\Leftrightarrow \frac{\sum_{k=1}^{N} x_k - \sum_{k=1}^{N} y_k}{N} &= \frac{1-2\hat{\theta}_1}{\hat{\theta}_1(1-\hat{\theta}_1)} = \frac{2\hat{\theta}_1 - 1}{\hat{\theta}_1 - \hat{\theta}_1^{\,2}} \\
\Leftrightarrow 0 &= \frac{\sum_{k=1}^{N} x_k - \sum_{k=1}^{N} y_k}{N} * \left(\frac{\hat{\theta}_1 - \hat{\theta}_1^{\,2}}{2\hat{\theta}_1 - 1} \right) \\
\Leftrightarrow 0 &= \sum_{k=1}^{N} x_k - \sum_{k=1}^{N} y_k \\
\vee \; 0 &= \hat{\theta}_1 - \hat{\theta}_1^{\,2}
\end{aligned}
$$

$\Rightarrow$ *the solution set is empty for* $\hat{\theta}_1$, *because the only possible solutions* $\hat{\theta}_1 = \lambda = 0 \; \vee \; \hat{\theta}_1 = \lambda = 1$ *contradict with* $\eta = 1 - \lambda$ *and* $\lambda, \; \eta > 0$

2 Exercise 2

$$D = (x_1, x_2, ..., x_7) = (\text{head, head, tail, tail, head, head, head}) \tag{1}$$

$$P(x|\theta) = \begin{cases} \theta & \text{if } x = \text{head} \\ 1-\theta & \text{if } x = \text{tail} \end{cases}$$

where $\theta \in [0, 1]$.

2.1 a

Following the lecture, the likelihood function $P(D|\theta)$ is given by:

$$\begin{aligned} P(D|\theta) &= \prod_{k=1}^{N} P(x_k|\theta) \\ &= \prod_{k=1}^{N_{\text{tail}}} P(\text{tail}|\theta) \cdot \prod_{k=1}^{N_{\text{Head}}} P(\text{head}|\theta) \\ &= \prod_{k=1}^{2} (1-\theta) \cdot \prod_{k=1}^{5} \theta \\ &= (1-\theta)^2 \cdot \theta^5 \end{aligned}$$

2.2 b

At first we determine the maximum likelihood solution $\hat{\theta}$. To find $\hat{\theta}$ we derive $P(D|\theta)$ (from exercise 2a) by θ and by setting $\frac{\partial}{\partial\theta} P(D|\theta) = 0$ and solving this equation for θ we will find θ.

$$\begin{aligned} \frac{\partial}{\partial\theta} P(D|\theta) &= \frac{\partial}{\partial\theta} (1-\theta)^2 \cdot \theta^5 \\ &= -2 \cdot (1-\theta) \cdot \theta^5 + 5 \cdot (1-\theta)^2 \cdot \theta^4 \\ &= -12 \cdot \theta^5 + 7 \cdot \theta^6 + 5 \cdot \theta^4 \\ &= \theta^4 \cdot (7 \cdot \theta^2 - 12 \cdot \theta + 5) \end{aligned}$$

$$\begin{aligned} \Rightarrow \quad & \frac{\partial}{\partial\theta}P(D|\theta) = 0 \\ \Leftrightarrow \quad & \theta^4 \cdot (7\cdot\theta^2 - 12\cdot\theta + 5) = 0 \end{aligned}$$

So we find the first theta: $\hat{\theta}_1 = 0$.

$$\begin{aligned} \Rightarrow \quad & 7\cdot\theta^2 - 12\cdot\theta + 5 = 0 \\ \Leftrightarrow \quad & \theta^2 - \frac{12}{7}\cdot\theta + \frac{5}{7} = 0 \mid \text{p-q-formula} \\ \Leftrightarrow \quad & \hat{\theta}_{2,3} = \frac{12}{14} \pm \sqrt{\left(\frac{12}{14}\right)^2 - \frac{5}{7}} \\ \Leftrightarrow \quad & \hat{\theta}_2 = 1,\ \hat{\theta}_3 = \frac{5}{7} \end{aligned}$$

So we have three posible solutions for $\hat{\theta}$:

$$\hat{\theta}_1 = 0,\ \hat{\theta}_2 = 1,\ \hat{\theta}_3 = \frac{5}{7}. \tag{2}$$

To decide which $\hat{\theta}$ is the maximum in the intervall $[0,1]$ we take a look at the functionvalues of $\frac{\partial^2}{\partial\theta^2}P(D|\hat{\theta}_i)$.

$$\begin{aligned} \frac{\partial^2}{\partial\theta^2}P(D|\hat{\theta}) &= \frac{\partial}{\partial\theta}\theta^4 \cdot (7\cdot\theta^2 - 12\cdot\theta + 5) \\ &= 4\cdot\theta^3\cdot(7\cdot\theta^2 - 12\cdot\theta + 5) + \theta^4\cdot(14\cdot\theta - 12) \\ &= 42\cdot\theta^5 - 60\cdot\theta^4 + 20\cdot\theta^5 \end{aligned}$$

$$\begin{aligned} &\Rightarrow \frac{\partial^2}{\partial\theta^2}P(D|\theta_1 = 0) = 42\cdot\theta_1^5 - 60\cdot\theta_1^4 + 20\cdot\theta_1^5 = 0 \rightarrow \text{saddle point} \\ &\Rightarrow \frac{\partial^2}{\partial\theta^2}P(D|\theta_2 = 1) = 42\cdot\theta_2^5 - 60\cdot\theta_2^4 + 20\cdot\theta_2^5 = 2 \rightarrow \text{minima} \\ &\Rightarrow \frac{\partial^2}{\partial\theta^2}P(D|\theta_3 = \frac{5}{7}) = 42\cdot\theta_3^5 - 60\cdot\theta_3^4 + 20\cdot\theta_3^5 \approx 0,179 \rightarrow \text{maxima} \end{aligned}$$

So the maximum likelihood solution for $\hat{\theta}$ is:

$$\hat{\theta} = \frac{5}{7} \tag{3}$$

Now we can calculate the probability that the next two tosses are "headünder assuming that all tosses are generated independently:

$$\begin{aligned} P(x_8 = \text{head}, x_8 = \text{head}|\theta) &= P(x_8 = \text{head}) \cdot P(x_9 = \text{head}) \\ &= \hat{\theta} \cdot \hat{\theta} = \frac{25}{49} \approx 0,51 = 51 \end{aligned}$$

2.3 c

Now the prior distribution for the parameter θ is definded as:

$$p(\theta) = \begin{cases} 1 & \text{if } 0 \leq \theta \leq 1 \\ 0 & \text{else} \end{cases}$$

$$\Rightarrow p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta} \;|p(\theta) = 1$$

$$= \frac{p(D|\theta)}{\int p(D|\theta)d\theta}$$

$$= \frac{(1-\theta)^2 \cdot \theta^5}{\int_0^1 (1-\theta)^2 \cdot \theta^5 d\theta}$$

$$= \frac{(1-\theta)^2 \cdot \theta^5}{[\frac{\theta^6}{6} - \frac{2 \cdot \theta^7}{7} + \frac{\theta^8}{8}]_0^1}$$

$$= 168 \cdot (1-\theta)^2 \cdot \theta^5$$

So we got the following form for $p(\theta|D)$:

$$p(\theta|D) = \begin{cases} 168 \cdot (1-\theta)^2 \cdot \theta^5 & \text{for } 0 \leq \theta \leq 1 \\ 0 & \text{else} \end{cases}$$

With this result for $p(\theta|D)$ we can now calculate the probability that the next two tosses are head:

$$\begin{aligned}
&\Rightarrow \int P(x_8 = \text{head}, x_9 = \text{head}|\theta) \cdot p(\theta|D) d\theta \\
&= \int \underbrace{P(x_8 = \text{head}|\theta)}_{=\theta} \cdot \underbrace{P(x_9 = \text{head}|\theta)}_{=\theta} \cdot \underbrace{p(\theta|D)}_{=168\cdot(1-\theta)^2\cdot\theta^5} d\theta \\
&= 168 \cdot \int_0^1 \theta^7 (1-\theta)^2 d\theta \\
&= 168 \cdot \int_0^1 (\theta^9 - 2 \cdot \theta^8 + \theta^7) d\theta \\
&= 168 \cdot \left[\frac{\theta^{10}}{10} - \frac{2 \cdot \theta^9}{9} + \frac{\theta^8}{8} \right]_0^1 \\
&= \frac{7}{15} \approx 0,467 = 46,7
\end{aligned}$$

3 Exercise 3

Given formulas are:

$$\frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \tag{4}$$

$$\frac{\mu_n}{\sigma_n^2} = \frac{n}{\sigma_n^2} \hat{\mu}_n + \frac{\mu_0}{\sigma_0^2} \tag{5}$$

$$\hat{\mu}_n = \frac{1}{n} \sum_{k=1}^{n} x_k \tag{6}$$

3.1 a

We want to show that following relation holds:

$$\sigma_n^2 \leq min\left(\frac{\sigma^2}{n}, \sigma_0^2\right) \tag{7}$$

At first we need a expression for σ_n^2. This will give us formula (4):

$$\begin{aligned}
\Rightarrow \quad && \frac{1}{\sigma_n^2} &= \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \\
\Leftrightarrow \quad && 1 &= \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right) \cdot \sigma_n^2 \\
\Leftrightarrow \quad && \sigma_n^2 &= \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}} \\
\Leftrightarrow \quad && \sigma_n^2 &= \frac{1}{\frac{n\sigma_0^2}{\sigma^2\sigma_0^2} + \frac{\sigma^2}{\sigma_0^2\sigma^2}} \\
\Leftrightarrow \quad && \sigma_n^2 &= \frac{1}{\frac{n\sigma_0^2+\sigma^2}{\sigma^2\sigma_0^2}} \\
\Leftrightarrow \quad && \sigma_n^2 &= \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}
\end{aligned}$$

Now we have two cases to consider to show that $\sigma_n^2 \leq min\left(\frac{\sigma^2}{n}, \sigma_0^2\right)$ holds:

$$\text{case 1: } \sigma_0^2 \leq \frac{\sigma^2}{n} \tag{8}$$

$$\text{case 2: } \sigma_0^2 \geq \frac{\sigma^2}{n} \tag{9}$$

At first we take a look at case one:

$$
\begin{aligned}
\Rightarrow \qquad & \sigma_n^2 \leq min\left(\frac{\sigma^2}{n}, \sigma_0^2\right) \\
\Leftrightarrow \qquad & \sigma_n^2 \leq \frac{\sigma^2}{n} \\
\Leftrightarrow \qquad & \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2} \leq \frac{\sigma^2}{n} \\
\Leftrightarrow \qquad & \frac{\sigma_0^2}{n\sigma_0^2 + \sigma^2} \leq \frac{1}{n} \\
\Leftrightarrow \qquad & \frac{1}{n\sigma_0^2 + \sigma^2} \leq \frac{1}{n\sigma_0^2} \\
\Leftrightarrow \qquad & 1 \leq \frac{n\sigma_0^2 + \sigma^2}{n\sigma_0^2} \\
\Leftrightarrow \qquad & 1 \leq 1 + \frac{\sigma}{n\sigma_0^2} \\
\Leftrightarrow \qquad & 0 \leq \frac{\sigma}{n\sigma_0^2}
\end{aligned}
$$

This expression is true because σ, n and σ_0 are always greater than zero (or maybe sometimes equal to zero) and so we have shown thas the first case is true.
Now we take a look at case two:

$$
\begin{aligned}
\Rightarrow \qquad & \sigma_n^2 \leq min\left(\frac{\sigma^2}{n}, \sigma_0^2\right) \\
\Leftrightarrow \qquad & \sigma_n^2 \leq \sigma_0^2 \\
\Leftrightarrow \qquad & \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2} \leq \sigma_0^2 \\
\Leftrightarrow \qquad & \frac{\sigma^2}{n\sigma_0^2 + \sigma^2} \leq 1 \\
\Leftrightarrow \qquad & \sigma^2 \leq n\sigma_0^2 + \sigma^2 \\
\Leftrightarrow \qquad & 1 \leq \frac{n\sigma_0^2}{\sigma^2} - 1 \\
\Leftrightarrow \qquad & 0 \leq \frac{n\sigma_0^2}{\sigma^2}
\end{aligned}
$$

Like in case one is this expression true because σ, n and σ_0 are always greater than zero (or maybe sometimes equal to zero) and so we have shown thas the second case is true.
So we have shown that relation (7) holds.

3b

Prove that $min(\hat{\mu}_n, \mu_0) \leq \mu_n \leq max(\hat{\mu}_n, \mu_0)$

$$\begin{aligned}
\frac{\mu_n}{\sigma_n^2} &= \frac{n}{\sigma^2}\hat{\mu}_n + \frac{\mu_0}{\sigma_0^2} \\
\Leftrightarrow \frac{\mu_n}{\frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}} &= \frac{n}{\sigma^2}\hat{\mu}_n + \frac{\mu_0}{\sigma_0^2} \\
\Leftrightarrow \mu_n &= \frac{n\sigma_0^2\hat{\mu}_n + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}
\end{aligned}$$

If $\hat{\mu}_n \leq \mu_n$:

$$\begin{aligned}
&\hat{\mu}_n \leq \frac{n\sigma_0^2\hat{\mu}_n + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \\
&\Leftrightarrow 0 \leq \frac{n\sigma_0^2\hat{\mu}_n + \mu_0\sigma^2 - \hat{\mu}_n(n\sigma_0^2 + \sigma^2)}{n\sigma_0^2 + \sigma^2} \\
&\Leftrightarrow 0 \leq \frac{\mu_0\sigma^2 - \hat{\mu}_n\sigma^2}{n\sigma_0^2 + \sigma^2} \\
&\Leftrightarrow 0 \leq \sigma^2(\mu_0 - \hat{\mu}_n),\ \textit{because neither } n,\ \sigma_0^2 \textit{ or } \sigma^2 \textit{ can be negative} \\
&\Leftrightarrow 0 \leq \mu_0 - \hat{\mu}_n,\ \textit{because } \sigma^2 > 0 \\
&\Leftrightarrow \hat{\mu}_n \leq \mu_0
\end{aligned}$$

If $\hat{\mu}_n \leq \mu_n,\ \mu_0$ has to be greater or equals to μ_n to fulfill $min(\hat{\mu}_n, \mu_0) \leq \mu_n \leq max(\hat{\mu}_n, \mu_0)$ because of transitivity that would mean $\hat{\mu}_n \leq \mu_0$ is true.

Test whether $\mu_0 \geq \mu_n$ if $\hat{\mu}_n \leq \mu_n$:

$$\begin{aligned}
&\mu_0 \geq \mu_n \\
&\Leftrightarrow \mu_0 \geq \frac{n\sigma_0^2\hat{\mu}_n + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \\
&\Leftrightarrow 0 \geq \frac{n\sigma_0^2\hat{\mu}_n + \mu_0\sigma^2 - \mu_0(n\sigma_0^2 + \sigma^2)}{n\sigma_0^2 + \sigma^2} \\
&\Leftrightarrow 0 \geq \frac{n\sigma_0^2\hat{\mu}_n - \mu_0 n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \\
&\Leftrightarrow 0 \geq n\sigma_0^2\hat{\mu}_n - \mu_0 n\sigma_0^2,\ \textit{because } n,\ \sigma_0^2 \textit{ and } \sigma^2 > 0 \\
&\Leftrightarrow \mu_0 \geq \hat{\mu}_n
\end{aligned}$$

$\Rightarrow$ If $\mu_0 \geq \hat{\mu}_n$ then $min(\hat{\mu}_n, \mu_0) \leq \mu_n \leq max(\hat{\mu}_n, \mu_0)$ is true.

Same can be proven for $\hat{\mu}_n \geq \mu_n$:

$$\begin{aligned}
&\hat{\mu}_n \geq \frac{n\sigma_0^2\hat{\mu}_n + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \\
&\Leftrightarrow 0 \geq \frac{n\sigma_0^2\hat{\mu}_n + \mu_0\sigma^2 - \hat{\mu}_n(n\sigma_0^2 + \sigma^2)}{n\sigma_0^2 + \sigma^2} \\
&\Leftrightarrow 0 \geq \frac{\mu_0\sigma^2 - \hat{\mu}_n\sigma^2}{n\sigma_0^2 + \sigma^2} \\
&\Leftrightarrow 0 \geq \sigma^2(\mu_0 - \hat{\mu}_n),\ \textit{because neither } n,\ \sigma_0^2 \textit{ or } \sigma^2 \textit{ can be negative} \\
&\Leftrightarrow 0 \geq \mu_0 - \hat{\mu}_n,\ \textit{because } \sigma^2 > 0 \\
&\Leftrightarrow \hat{\mu}_n \geq \mu_0
\end{aligned}$$

Test whether $\mu_n \geq \mu_0$ if $\hat{\mu}_n \geq \mu_n$:

$$\begin{aligned}
&\mu_0 \leq \mu_n \\
&\Leftrightarrow \mu_0 \leq \frac{n\sigma_0^2\hat{\mu}_n + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \\
&\Leftrightarrow 0 \leq \frac{n\sigma_0^2\hat{\mu}_n + \mu_0\sigma^2 - \mu_0(n\sigma_0^2 + \sigma^2)}{n\sigma_0^2 + \sigma^2} \\
&\Leftrightarrow 0 \leq \frac{n\sigma_0^2\hat{\mu}_n - \mu_0 n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \\
&\Leftrightarrow 0 \leq n\sigma_0^2\hat{\mu}_n - \mu_0 n\sigma_0^2,\ \textit{because } n,\ \sigma_0^2 \textit{ and } \sigma^2 > 0 \\
&\Leftrightarrow \mu_0 \leq \hat{\mu}_n
\end{aligned}$$

$\Rightarrow$ If $\hat{\mu}_n \geq \mu_n$ then $min(\hat{\mu}_n, \mu_0) \leq \mu_n \leq max(\hat{\mu}_n, \mu_0)$ is true.

It is proven that $min(\hat{\mu}_n, \mu_0) \leq \mu_n \leq max(\hat{\mu}_n, \mu_0)$