Solutions to Problem Sheet 1: Maximum Likelihood and Bayesian Inference

1. **Gaussian / Exponential model**

   (a) *Gaussian Model:*
   We have a single observation $x \sim \mathcal{N}(\theta, \sigma^2)$ (likelihood) and a prior $\theta \sim \mathcal{N}(m, \rho\sigma^2)$.
   The posterior density $p(\theta|x)$ is proportional to the product of the likelihood and the prior:

   $$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)d\theta} \propto p(x|\theta)p(\theta).$$

   Substituting the Gaussian densities (ignoring constants independent of θ):

   $$p(\theta|x) \propto \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right) \exp\left(-\frac{(\theta-m)^2}{2\rho\sigma^2}\right)$$

   Combining the terms in the exponent:

   $$\begin{aligned}
   \text{Exponent} &= -\frac{1}{2\sigma^2}\left[(x-\theta)^2 + \frac{(\theta-m)^2}{\rho}\right] \\
   &= -\frac{1}{2\sigma^2}\left[\theta^2 - 2x\theta + x^2 + \frac{\theta^2 - 2m\theta + m^2}{\rho}\right] \\
   &= -\frac{1}{2\sigma^2}\left[\theta^2\left(1+\frac{1}{\rho}\right) - 2\theta\left(x + \frac{m}{\rho}\right)\right] + \text{const} \\
   &= -\frac{1+\frac{1}{\rho}}{2\sigma^2}\left[\theta^2 - 2\theta\frac{x+\frac{m}{\rho}}{1+\frac{1}{\rho}}\right] + \text{const} \\
   &= -\frac{1+\frac{1}{\rho}}{2\sigma^2}\left[\theta^2 - 2\theta\frac{x+\frac{m}{\rho}}{1+\frac{1}{\rho}} + \left(\frac{x+\frac{m}{\rho}}{1+\frac{1}{\rho}}\right)^2 - \left(\frac{x+\frac{m}{\rho}}{1+\frac{1}{\rho}}\right)^2\right] + \text{const} \\
   &= -\frac{1+\frac{1}{\rho}}{2\sigma^2}\left[\theta^2 - 2\theta\frac{x+\frac{m}{\rho}}{1+\frac{1}{\rho}} + \left(\frac{x+\frac{m}{\rho}}{1+\frac{1}{\rho}}\right)^2\right] + \text{const} \\
   &= -\frac{1+\frac{1}{\rho}}{2\sigma^2}\left[\theta - \frac{x+\frac{m}{\rho}}{1+\frac{1}{\rho}}\right]^2 + \text{const}
   \end{aligned}$$

   Therefore, $p(\theta|x) \propto e^{-\frac{1+\frac{1}{\rho}}{2\sigma^2}\left[\theta - \frac{x+\frac{m}{\rho}}{1+\frac{1}{\rho}}\right]^2}$ and we recognize the Gaussian distribution $\mathcal{N}(\mu_{post}, \sigma^2_{post})$ with:

   $$\sigma^2_{post} = \frac{\rho}{1+\rho}\sigma^2,$$

$$\mu_{post} = \frac{\rho x + m}{1+\rho}.$$

Solution: The posterior is $\theta|x \sim \mathcal{N}\left(\frac{m+\rho x}{1+\rho}, \frac{\rho}{1+\rho}\sigma^2\right)$.

(b) *Exponential Model:*
Suppose $X \sim \mathcal{E}(\lambda)$ (Exponential) and prior $\lambda \sim \mathcal{G}(a, b)$ (Gamma).
Likelihood: $p(x|\lambda) = \lambda e^{-\lambda x}$.
Prior: $p(\lambda) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda} \propto \lambda^{a-1}e^{-b\lambda}$.
Posterior:

$$p(\lambda|x) \propto (\lambda e^{-\lambda x})(\lambda^{a-1}e^{-b\lambda}) = \lambda^{(a+1)-1}e^{-(b+x)\lambda}$$

This is the density of a Gamma distribution with parameters $a' = a+1$ and $b' = b+x$.
Solution: The posterior is $\lambda|x \sim \mathcal{G}(a+1, b+x)$.

2. **Maximum Likelihood Estimator (MLE)**

We observe two values (z_1, z_2).

(a) *Case (i):* $X_1, X_2 \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$.
The likelihood function is:

$$L(\theta) = \prod_{i=1}^{2} \frac{1}{\sqrt{2\pi}} e^{-(z_i-\theta)^2/2} \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{2}(z_i-\theta)^2\right)$$

Maximizing the likelihood is equivalent to minimizing the sum of squared differences $(z_1 - \theta)^2 + (z_2-\theta)^2$, which is convex as a function of θ. The derivative is $4\theta - 2z_1 - 2z_2$ and is equal to 0 for

$$\hat{\theta}_{MLE} = \frac{z_1+z_2}{2}$$

(b) *Case (ii)*
The joint density is given by:

$$g(x_1, x_2|\theta) = \pi^{-3/2}\frac{\exp\{-(x_1+x_2-2\theta)^2/4\}}{1+(x_1-x_2)^2}$$

To find the MLE, we maximize $g(x_1, x_2|\theta)$ with respect to θ. Notice that the denominator $1+(z_1-z_2)^2$ does not depend on θ. Thus, we only need to maximize the numerator:

$$\text{Numerator} \propto \exp\left\{-\frac{(z_1+z_2-2\theta)^2}{4}\right\}$$

This exponential is maximized when the exponent is zero (since the exponent is non-positive):

$$(z_1+z_2-2\theta)^2 = 0 \implies 2\theta = z_1+z_2 \implies \hat{\theta}_{MLE} = \frac{z_1+z_2}{2}$$

Remark: Even though the joint distributions in (i) (independent Gaussian) and (ii) (dependent, heavy-tailed Cauchy-like term) are very different, they yield the exact same Maximum Likelihood Estimator for θ.

3. **Bernoulli model**

(a) *Show uniqueness.*
The likelihood for n i.i.d. observations is $L(\theta) = \prod_i \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$. The log-likelihood is:

$$\ell(\theta) = \left(\sum_{i=1}^{n} x_i\right)\log\theta + \left(n - \sum_{i=1}^{n} x_i\right)\log(1-\theta)$$

The second derivative with respect to θ is:

$$\ell''(\theta) = -\frac{\sum x_i}{\theta^2} - \frac{n-\sum x_i}{(1-\theta)^2} < 0$$

Since $x_i \in \{0,1\}$ and $\theta \in (0,1)$, both terms are negative (assuming at least one 0 and one 1 are observed, or treating boundary cases as limits). Thus, the function is strictly concave, implying that any critical point found is a unique global maximum.

(b) *Show* $\hat{\theta}(X) = \overline{X}_n$*.*
Setting the first derivative to zero:

$$\ell'(\theta) = \frac{\sum x_i}{\theta} - \frac{n-\sum x_i}{1-\theta} = 0$$

$$(1-\theta)\sum x_i = \theta(n - \sum x_i)$$

$$\sum x_i - \theta\sum x_i = n\theta - \theta\sum x_i$$

$$\sum x_i = n\theta \implies \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{X}_n$$

4. **Genetics**

Let B denote the allele for Brown (dominant) and b denote Blue (recessive, denoted X and x in the prompt). Genotypes: xx (blue), xX (brown), XX (brown). Frequencies: $P(xx) = p^2$, $P(xX) = 2p(1-p)$, $P(XX) = (1-p)^2$. Note: Total Brown freq $= 1 - p^2$.

(a) *Expected proportion of heterozygotes among brown-eyed children of brown-eyed parents.*
We are looking for $P(\text{Heterozygote child}|\text{ Brown parents and Brown child})$ which is:

$$\begin{aligned}P(xX|\text{ Brown parents and Brown child}) &= \frac{P(\text{ Brown parents, Brown child with xX})}{P(\text{ Brown parents and Brown child})}\\ &= \frac{P(\text{ Brown parents, child with xX})}{P(\text{ Brown parents and Brown child})},\end{aligned}$$

using the fact that heterozygotes always have brown eyes.
The numerator is equal to:

$$\begin{aligned}&\sum_{\substack{\text{Parent }1\in\{xX,XX\},\\ \text{Parent }2\in\{xX,XX\}}} P(\text{ Parent 1, Parent 2, child with xX})\\ =&\sum_{\substack{\text{Parent }1\in\{xX,XX\},\\ \text{Parent }2\in\{xX,XX\}}} P(\text{ child with xX}|\text{Parent 1, Parent 2})P(\text{Parent 1, Parent 2}).\end{aligned}$$

If both parents are homozygote (resp. heterozygote), the above probability is 0 (resp. 1/2), and it is 1/2 in the last case. So the above gives

$$\frac{1}{2}P(\text{xX, XX})+\frac{1}{2}P(\text{XX,xX})+\frac{1}{2}P(\text{xX,xX}) = 2p(1-p)\cdot(1-p)^2+\frac{(2p(1-p))^2}{2} = 2p(1-p)^2.$$

The denominator is equal to

$$P(\text{ Brown parents, child with xX}) + P(\text{ Brown parents, child with XX}).$$

Similarly, the second term is

$$\sum_{\substack{\text{Parent } 1\in\{xX,XX\},\\ \text{Parent } 2\in\{xX,XX\}}} P(\text{ child with XX}|\text{Parent 1, Parent 2})P(\text{Parent 1, Parent 2}),$$

and If both parents are homozygote (resp. heterozygote), the above probability is 1 (resp. 1/2), and it is 1/2 in the last case, so the above is

$$\frac{1}{2}P(\text{xX, XX})+\frac{1}{2}P(\text{XX,xX})+\frac{1}{2}P(\text{xX,xX})+P(\text{XX,XX}) = 2p(1-p)^2+(1-p)^4 = (1+p^2)(1-p)^2.$$

In the end, we obtain

$$2p/(1+p^2).$$

(b) *Posterior probability Judy is a heterozygote.*
Let H be the event Judy is Heterozygous (xX) and D be the event Judy is Dominant Homozygous (XX). Since Judy is the brown-eyed child of brown-eyed parents, we use the result from (a) as her prior:

$$P(H) = \frac{2p}{1+p^2}, \quad P(D) = 1 - P(H) = \frac{(1-p)^2}{1+p^2}$$

Evidence E: She marries a heterozygote (xX) and has n brown-eyed children.
Likelihoods:

- If Judy is H (xX): Mating $xX \times xX$. Prob of brown child is 3/4, since the only way for the child to have blue eyes is for both parents to provide allele x, with probability 1/2.

$$P(E|H) = (3/4)^n$$

- If Judy is D (XX): Mating $XX \times xX$. All children are brown (XX or xX).

$$P(E|D) = 1^n = 1$$

Posterior $P(H|E)$:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|D)P(D)}$$

$$= \frac{(3/4)^n \frac{2p}{1+p^2}}{(3/4)^n \frac{2p}{1+p^2} + 1 \cdot \frac{(1-p)^2}{1+p^2}} = \frac{2p(3/4)^n}{2p(3/4)^n + (1-p)^2}$$

(c) *Probability Judy's first grandchild has blue eyes.*
Let K be Judy's child (one of the n brown children). Let G be the grandchild. For G to be blue (xx), both parents must contribute x. Assume K marries a random person from the population. The probability the partner contributes x is 1 if it is xx (with probability p^2) and 1/2 if it is xX (with probability $2p - 2p^2$), so it is $p^2 + p - p^2 = p$ in total.

Now we need the probability K contributes x. Let $\pi_n = P(H|E)$ be Judy's posterior from (b).

- Case 1: Judy is D (XX). Then K is from $XX \times xX$. K is XX or xX with prob 0.5 each. $P(K \text{ passes } x|\text{Judy } D) = 0.5 \times 0 + 0.5 \times 0.5 = 0.25$.
- Case 2: Judy is H (xX). Then K is from $xX \times xX$, but we know K is brown. Among brown children of $xX \times xX$, the genotypes are XX (1/3) and xX (2/3). $P(K \text{ passes } x|\text{Judy } H) = (1/3) \times 0 + (2/3) \times 0.5 = 1/3$.

Total prob K passes x: $P(K_x) = \pi_n(1/3) + (1 - \pi_n)(1/4)$.
Finally, $P(G \text{ is blue}) = P(K_x) \times P(\text{Partner}_x) = \left[\frac{\pi_n}{3} + \frac{1-\pi_n}{4}\right] p$.

5. **Twins and Elvis Presley** Let I be the event of Identical twins and F be Fraternal.

$$P(I) = 1/300, \qquad P(F) = 1/125.$$

Event E: Elvis had a twin brother.
Note: We must condition on the fact that the birth was a twin birth.

$$P(I|\text{Twins}) = \frac{P(I,\ \text{Twins})}{P(\text{Twins})} = \frac{P(I,\ \text{Twins})}{P(\text{Twins})} = \frac{P(I)}{P(I) + P(F)} = \frac{1/300}{1/300 + 1/125} = \frac{125}{125 + 300} = \frac{5}{17}.$$

We also deduce that $P(F|\text{Twins}) = \frac{12}{17}$.

Now, let's use the evidence that the second twin was a brother. Likelihood of observing Male-Male twins (MM):

- If Identical (I): Sex is always same. Assuming 50/50 boys/girls, $P(MM|I) = 1/2$.
- If Fraternal (F): Sexes are independent. $P(MM|F) = 1/2 \times 1/2 = 1/4$.

Posterior probability that Elvis was an identical twin:

$$P(I|Twins,\ MM) = \frac{P(MM|I) * 5/17}{P(MM|I) * 5/17 + P(MM|F) * 12/17}$$

(Using raw probabilities or conditional on twins cancels out, let's use raw).

$$= \frac{(1/2) * 5}{(1/2) * 5 + (1/4) * 12} = \frac{5}{5 + 6} = \frac{5}{11}.$$

So there is a 5/11 (approx 45.4%) chance Elvis was an identical twin.

6. **Monty Hall**

(a) *Should the contestant switch?*
Yes. Note C the event that you chose the correct door: $P(C) = 1/3$ (there are three doors). Note O the event that the host revealed the door he chose. Given C, the host had the choice between the last two doors to show you what is behind, so $P(O|C) = 1/2$.

Given C^c, the host could not reveal your own door or the one hiding the big prize, so it did not have a choice and had to reveal the door he chose: $P(O|C^c) = 1$. Therefore,

$$P(C|O) = \frac{P(C)P(O|C)}{P(C)P(O|C) + P(C^c)P(O|C^c)} = \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot 1} = 1/3.$$

Therefore, there is probability 2/3 the last door is the correct one.

(b) *Calculate probabilities.*

- **Strategy Stay:** You win only if your initial choice was correct. Since there are 3 boxes and 1 prize, $P(\text{Win}) = 1/3$.
- **Strategy Switch:** You win if your initial choice was wrong. (If you pick a losing box, Monty reveals the other loser, leaving the winner as the switch option). Since there are 2 losing boxes, $P(\text{Win}) = 2/3$.

(c) **Evil Monty Variant**

Let W be the event "Original choice is Winner" ($P(W) = 1/3$) and L be "Original choice is Loser" ($P(L) = 2/3$). Let O be the event "Monty Offers a switch".

Rules:

- If W, Monty offers switch with prob p: $P(O|W) = p$.
- If L, Monty always offers switch: $P(O|L) = 1$.

We want to know if we should switch given we are offered the chance. This is equivalent to comparing $P(L|O)$ (Switch wins) vs $P(W|O)$ (Stay wins).

$$P(L|O) = \frac{P(O|L)P(L)}{P(O|L)P(L) + P(O|W)P(W)} = \frac{1 \cdot (2/3)}{1 \cdot (2/3) + p \cdot (1/3)}$$

$$P(L|O) = \frac{2/3}{2/3 + p/3} = \frac{2}{2 + p}$$

If $p < 1$, the probability of winning by switching $(2/(2 + p))$ is **greater** than the standard 2/3. If $p = 0$, switching guarantees a win ($P = 1$). Even if $p = 1$ (standard game), prob is 2/3. Since $p \in [0, 1]$, $P(L|O) \geq 2/3$.

Conclusion: You should still switch. In fact, the "Evil" behavior makes switching even more advantageous because being offered a switch is now evidence suggesting you likely picked a losing box initially (since he might not have offered it if you had picked the winner).