Bayes Inference.

Bayesian Inference Framework

Ground truth (unknown to us) | data (known to us) | Estimator. (our guess about unknown)

dog

dog or cat ?

$$\Theta \rightarrow \boxed{p(x|\theta)} \rightarrow X \rightarrow \boxed{\text{A function related to posterior distribution } p(\theta|x)} \rightarrow \hat{\Theta}$$

$p(\theta)$ prior distribution

observation model

$\hat{\theta}(x)$

Goal: Given data X, find the best estimate $\hat{\Theta}$ of Θ.

https://www.medicalnewstoday.com/articles/322868

Estimation Criterion I

- Goal: Given data X, find the best estimate $\hat{\Theta}$ of Θ.
- What is your criterion for "best"?
  - Want $\hat{\Theta} = \Theta$, but this is not guaranteed, as both are r.v.s.

$$\boxed{\text{Criterion: Minimize the probability of error } \mathbb{P}(\Theta \neq \hat{\Theta}).}$$

- What can be a good estimator?

Example: COVID-19 test. https://nyti.ms/31MTZgV

Suppose you think you may have contracted COVID-19.

You decide to take a diagnostic test to determine if you are infected.

The probability of contracting Covid-19 is 10%.

The false negative rate is 12.5%; The false positive rate is 2.5%.

What is the probability of being infected given a positive test result?

- $\Theta = 1$: infected, $\Theta = 0$: not infected. $P_\Theta(1) = 10\%$
- $X = 1$: test positive. $X = 0$: test negative. $P_{X|\Theta}(0|1) = 12.5\%$, $P_{X|\Theta}(1|0) = 2.5\%$

$$\mathbb{P}(\Theta = 1 \mid X = 1) = \frac{P_\Theta(1) P_{X|\Theta}(1|1)}{P_X(1)} = \frac{P_\Theta(1) P_{X|\Theta}(1|1)}{P_\Theta(1) P_{X|\Theta}(1|1) + P_\Theta(0) P_{X|\Theta}(1|0)} = 0.795.$$

- Given a positive test result $X=1$, what would be your guess of Θ?

(HW: Given a negative test result, what would be your guess of Θ?)

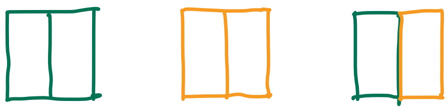
Example: Three cards.

- There are 3 cards

① green on both sides

② yellow on both sides

③ green on one side and yellow on the other



1 2 3

- Pick a card and a side uniform at random. Let X be the color you get
- Let Y be the color on the back.
- Q: What is $\mathbb{P}(Y = \text{green} \mid X = \text{green})$? A. $> \frac{1}{2}$ ✓ B. $< \frac{1}{2}$ C. $= \frac{1}{2}$.
- card number $\Theta \sim \text{Unif}\{1,2,3\}$.
- $P_{X|\Theta}(\text{green} \mid 1) = 1$, $P_{X|\Theta}(\text{green} \mid 2) = 0$, $P_{X|\Theta}(\text{green} \mid 3) = \frac{1}{2}$

$$\mathbb{P}(Y = \text{green} \mid X = \text{green}) = P_{\Theta|X}(1 \mid \text{green}) = \frac{P_\Theta(1)\, P_{X|\Theta}(\text{green} \mid 1)}{\sum_\theta P_{X|\Theta}(\text{green} \mid \theta)\, P(\theta)} = \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3}} = \frac{2}{3}$$

- Given $X = \text{green}$, what's your guess of the color on the back?

Estimation Criterion I

- Goal: Given data X, find the best estimate $\hat{\Theta}$ of Θ.
- What is your criterion for "best"?

$\boxed{\text{Criterion: Minimize the probability of error } \mathbb{P}(\Theta \neq \hat{\Theta}).}$

- What can be a good estimator?

Optimal estimator: MAP (Maximum a posteriori) estimator

$$\hat{\theta}_{MAP}(x) = \arg\max_{\theta} p(\theta | x) \quad \text{(break the tie arbitrarily)}$$

MAP (Maximum a Posteriori)

Estimation

Example: COVID-19 test. https://nyti.ms/31MTZgV

Suppose you think you may have contracted COVID-19.

You decide to take a diagnostic test to determine if you are infected.

The probability of contracting Covid-19 is 10%.

The false negative rate is 12.5%; The false positive rate is 2.5%.

What's the MAP estimator given X? What's the probability of error?

- $\Theta = 1$: infected, $\Theta = 0$: not infected. $P_\Theta(1) = 10\%$
- $X = 1$: test positive. $X = 0$: test negative. $P_{X|\Theta}(0|1) = 12.5\%$, $P_{X|\Theta}(1|0) = 2.5\%$

$$\mathbb{P}(\Theta = 1 \mid X = 1) = \frac{P_\Theta(1) P_{X|\Theta}(1|1)}{P_X(1)} = \frac{P_\Theta(1) P_{X|\Theta}(1|1)}{P_\Theta(1) P_{X|\Theta}(1|1) + P_\Theta(0) P_{X|\Theta}(1|0)} = 0.795.$$

$\Rightarrow$ Given a positive test result, it's more likely to be infected (79.5% chance) than not infected (20.5% chance) $\Rightarrow \hat{\theta}_{MAP}(1) = 1$

- $\mathbb{P}(\Theta=1 \mid X=0) = \dfrac{p_\Theta(1)\, p_{X|\Theta}(1|0)}{p_X(0)} = \dfrac{0.1 \times 0.025}{0.1 \times 0.025 + 0.9 \times 0.975} = 0.00284$

$\Rightarrow$ Given a negative test result, it's more likely to be not infected ($>99\%$ chance) than infected ($<1\%$ chance).

$\Rightarrow \hat{\theta}_{MAP}(0) = 0$.

- $\mathbb{P}(\hat{\Theta}_{MAP} \neq \Theta) = \mathbb{P}\left((\hat{\Theta}_{MAP} = 0, \Theta = 1) \text{ or } (\hat{\Theta}_{MAP} = 1, \Theta = 0)\right)$

$= \mathbb{P}(\hat{\Theta}_{MAP} = 0, \Theta = 1) + \mathbb{P}(\hat{\Theta}_{MAP} = 1, \Theta = 0)$

$= \mathbb{P}(X = 0, \Theta = 1) + \mathbb{P}(X = 1, \Theta = 0)$

$= \mathbb{P}(\Theta = 1)\,\mathbb{P}(X = 0 \mid \Theta = 1) + \mathbb{P}(\Theta = 0)\,\mathbb{P}(X = 1 \mid \Theta = 0)$

$= 0.1 \times 0.125 + 0.9 \times 0.025$

$= 0.035$

Example: Infer the unknown bias in choosing lab problems.

- Θ is drawn uniform at random in the interval $[0,1]$.
- Given $\Theta=\theta$, each problem is chosen to be graded with probability θ. independent of each other
- There are n problems in total (fixed). K were chosen to be graded.
- Find the MAP estimator $\hat{\theta}_{MAP}(k)$

- Unknown $\Theta \sim Unif[0,1]$.
- Observation model: for $i=1,2,\cdots,n$, denote $X_i = \begin{cases} 1, & \text{if problem } i \text{ is graded.} \\ 0, & \text{o.w.} \end{cases}$

  $X_i \mid \Theta=\theta \sim Bern(\theta)$. $K = X_1+X_2+\cdots+X_n$. $K \mid \Theta=\theta \sim Binom(n,\theta)$

  $p(k|\theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$. for $\theta\in[0,1]$ $k=0,1,\cdots,n$.

- posterior $f(\theta|k) = \dfrac{f(\theta)\,p(k|\theta)}{p(k)} = \dfrac{1\cdot\binom{n}{k}\theta^k(1-\theta)^{n-k}}{p(k)}$ for $\theta\in[0,1]$, $k=0,1,\cdots,n$.

- MAP estimator $\hat{\Theta}_{MAP}(k) = \underset{\theta}{\arg\max} f(\theta|k) = \underset{\theta}{\arg\max} \frac{\binom{n}{k}\theta^k(1-\theta)^{n-k}}{p(k)}$

Θ is uniform

$$= \underset{\theta}{\arg\max}\ p(k|\theta) = \underset{\theta}{\arg\max} \binom{n}{k} \underbrace{\theta^k(1-\theta)^{n-k}}_{\triangleq g(\theta)}.$$

- Find maximum: $g'(\theta) = k\theta^{k-1}(1-\theta)^{n-k} - (n-k)\theta^k(1-\theta)^{n-k-1} \overset{set}{=} 0 \Rightarrow \hat{\Theta}_{MAP}(k) = \frac{k}{n}$.

Takeaway message: when Θ is uniform, $\hat{\Theta}_{MAP}(x) = \hat{\Theta}_{ML}(x)$

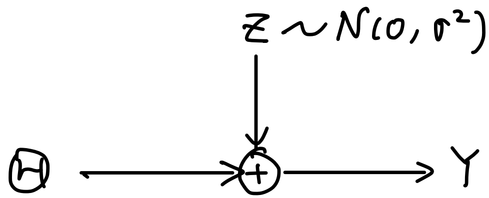
- When the prior distribution is uniform ($p(\theta)$ = constant for all θ)

$$\hat{\theta}_{MAP}(x) = \underset{\theta}{\arg\max}\ p(\theta|x) = \underset{\theta}{\arg\max} \frac{p(\theta)\,p(x|\theta)}{p(x)} = \underset{\theta}{\arg\max} \overbrace{p(x|\theta)}^{\text{likelihood}}.$$

- ML estimate : $\boxed{\hat{\theta}_{ML}(x) = \underset{\theta}{\arg\max}\ p(x|\theta)}$.

Example: Additive Gaussian Noise Channel.

Consider the following communication channel:



The signal transmitted is a binary random variable Θ:

$$\Theta = \begin{cases} +1, & \text{with probability } \frac{1}{2} \\ -1, & \text{with probability } \frac{1}{2}. \end{cases}$$

The received signal, also called the observation, is $Y = \Theta + Z$

We assume Θ and Z are independent

Find the best estimator $\hat{\Theta}$ that minimizes $\mathbb{P}(\hat{\Theta} \neq \Theta)$.

Find the probability of error $\mathbb{P}(\hat{\Theta} \neq \Theta)$

Solution

- MAP estimator minimizes $\mathbb{P}(\hat{\Theta} \neq \Theta)$.
- The prior is uniform, $p_\Theta(-1) = p_\Theta(1) = 0.5 \Rightarrow \hat{\theta}_{MAP}(y) = \hat{\theta}_{ML}(y)$
- $\hat{\theta}_{ML}(y) = \underset{\theta}{\arg\max}\, f(y|\theta) = \underset{\theta}{\arg\max}\, f_Z(y-\theta) = \begin{cases} 1, & \text{if } y > 0 \\ -1, & \text{if } y \leq 0 \end{cases}$
- $\mathbb{P}(\hat{\Theta} \neq \Theta) = \mathbb{P}(\hat{\Theta} = -1, \Theta = 1) + \mathbb{P}(\hat{\Theta} = 1, \Theta = -1)$

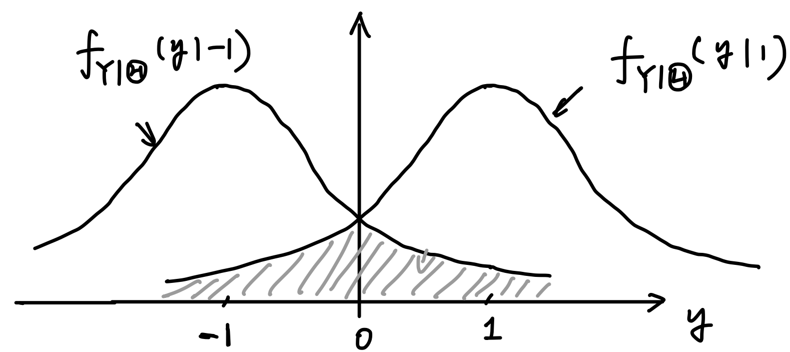
$= \mathbb{P}(\Theta = 1, Y \leq 0) + \mathbb{P}(\Theta = -1, Y > 0) = \mathbb{P}(\Theta = 1)\mathbb{P}(Y \leq 0 | \Theta = 1) + \mathbb{P}(\Theta = -1)\mathbb{P}(Y > 0 | \Theta = -1)$

$= \frac{1}{2}\int_{-\infty}^{0} f_{Y|\Theta}(y|1)\,dy + \frac{1}{2}\int_{0}^{\infty} f_{Y|\Theta}(y|-1)\,dy$

$= \frac{1}{2}\phi(-\frac{1}{\sigma}) + \frac{1}{2}(1 - \phi(\frac{1}{\sigma}))$

$= 1 - \phi(\frac{1}{\sigma})$

$\phi(z) \triangleq \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,dt$



Example: Estimating Gaussian signal in Gaussian noise.

- Signal $\Theta \sim N(0,1)$
- Observation $X = \Theta + W$, $W \sim N(0,1)$ indep. of Θ. Find $\hat{\theta}_{MAP}(x)$.

- Unknown: $\Theta \sim N(0,1)$

recall our trick in BSC example.

- Observation model: $f(x|\theta) = f_W(x-\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$

- Posterior: $f(\theta|x) = \frac{f(\theta) f(x|\theta)}{f(x)} = \frac{1}{f(x)} \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}} \triangleq c(x)\, e^{-\frac{1}{2}\left(\theta^2 + (x-\theta)^2\right)}$

$\Rightarrow$ Want $\min_{\theta} [\theta^2 + (x-\theta)^2]$ for fixed x.

$$\frac{\partial [\theta^2 + (x-\theta)^2]}{\partial \theta} \triangleq 0 \Rightarrow \theta^* = \frac{x}{2}. \qquad \Rightarrow \hat{\theta}_{MAP}(x) = \frac{x}{2}.$$

How to compute the MAP estimator?

- MAP estimator : $\hat{\theta}_{MAP}(x) = \arg\max_{\theta} p(\theta|x)$.

  - First fix x, treat $p(\theta|x)$ as a function $g(\theta)$ in θ.

  - Then find the θ^* that maximizes $g(\theta)$, this θ^* is called $\hat{\theta}_{MAP}(x)$.

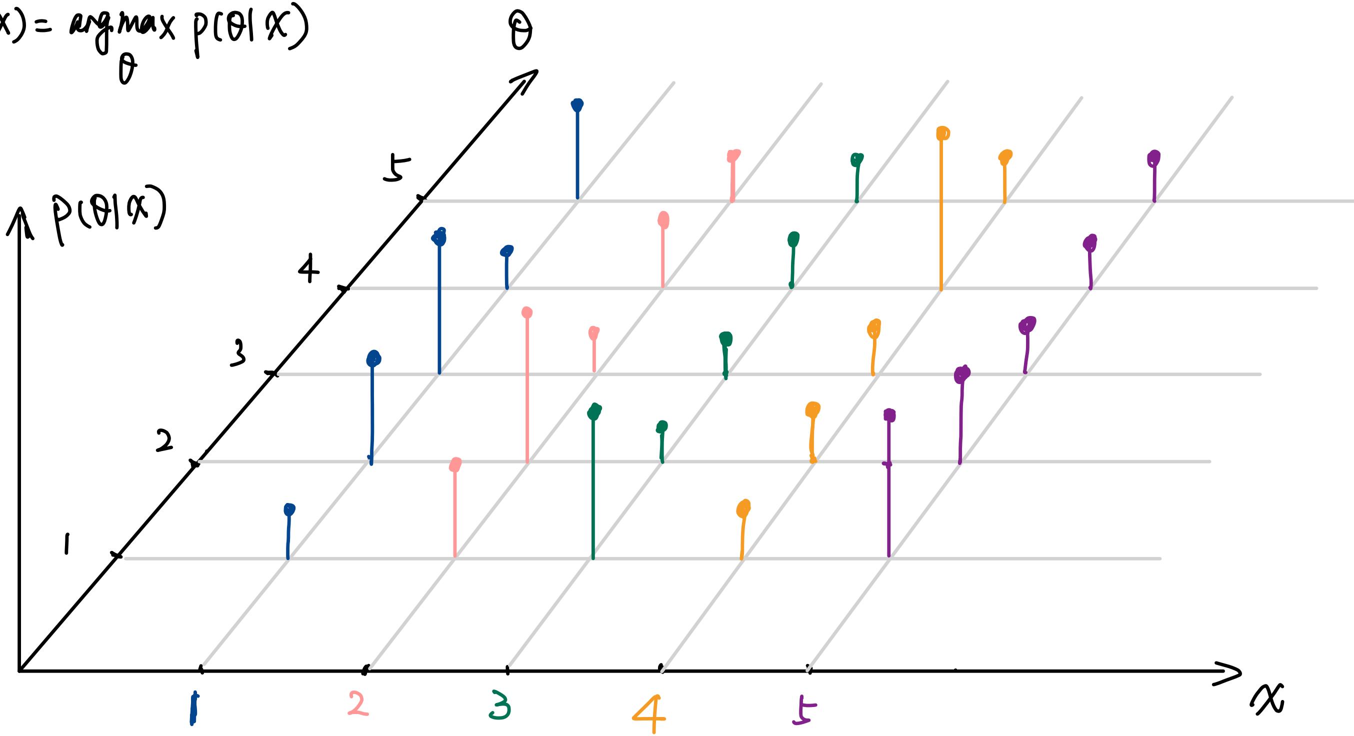
    (use the function property or calculus)

  - Do this for every x.

    $\hat{\theta}_{MAP}(x)$ is now viewed as a function in x

Illustration of the MAP estimator.

$$\hat{\theta}_{MAP}(x) = \arg\max_{\theta} p(\theta|x)$$



$\hat{\theta}_{MAP}(1) = 3$ $\hat{\theta}_{MAP}(2) = 2$ $\hat{\theta}_{MAP}(3) = 1$ $\hat{\theta}_{MAP}(4) = 4$ $\hat{\theta}_{MAP}(5) = 1$.

$\hat{\theta}_{MAP}(x)$ is a function in x!

How to compute $\mathbb{P}(\hat{\Theta} \neq \Theta)$?

- Relation among Θ, X, $\hat{\Theta}$:

a function of x

$$\xrightarrow{\Theta} \boxed{p(x|\theta)} \xrightarrow{X} \boxed{\hat{\theta}(x)} \xrightarrow{\hat{\Theta}}$$

- $\hat{\Theta}$ is a function of X. Given $X=x$, $\hat{\Theta} = \hat{\theta}(x)$ is not random.

- $$\mathbb{P}(\hat{\Theta} \neq \Theta) = \sum_x p_X(x)\, \mathbb{P}(\hat{\Theta} \neq \Theta \mid X=x) = \sum_x p_X(x) \underbrace{\mathbb{P}(\hat{\theta}(x) \neq \Theta \mid X=x)}_{= 1 - \mathbb{P}(\Theta = \hat{\theta}(x) \mid X = x).}$$

two r.v.s

a number.

alternatively $$= \sum_\theta \mathbb{P}(\Theta = \theta, \hat{\Theta} \neq \theta) = \sum_\theta p_\Theta(\theta) \sum_{x:\, \hat{\theta}(x) \neq \theta} \mathbb{P}(X=x \mid \Theta = \theta).$$

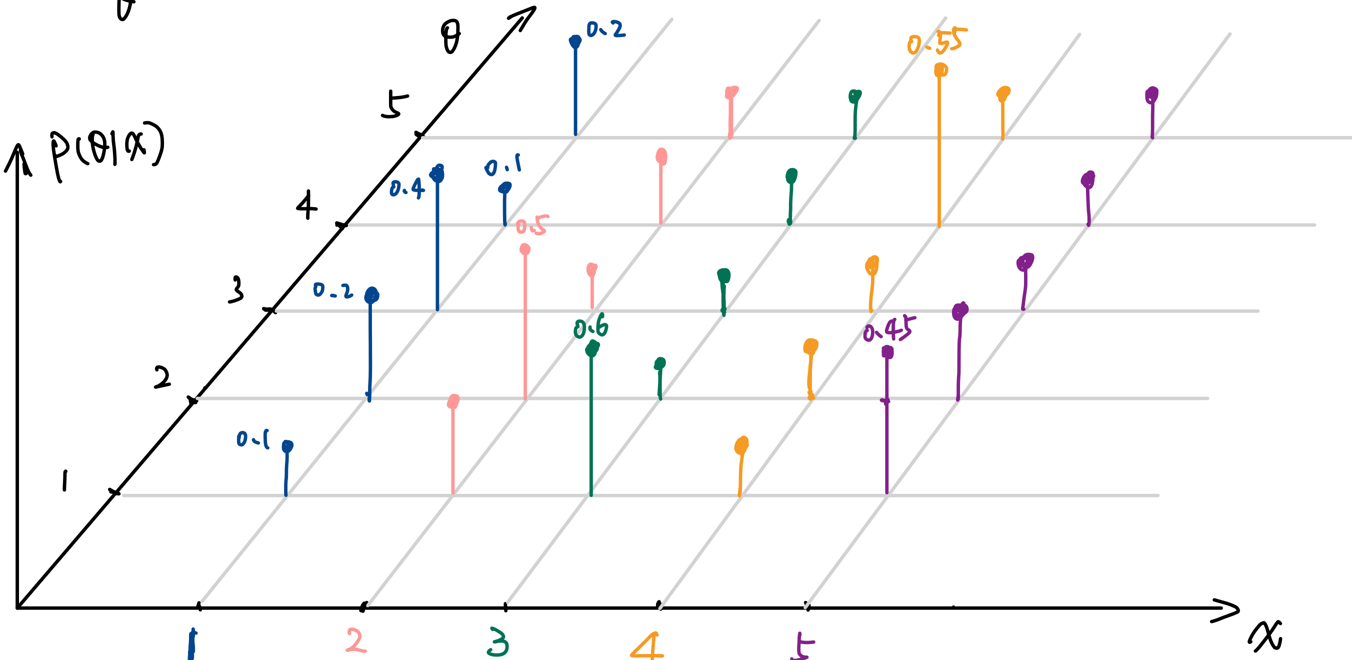
- Conditional probability of error given data x : $\mathbb{P}(\hat{\Theta} \neq \Theta \mid X = x)$

(Overall) probability of error : $\mathbb{P}(\hat{\Theta} \neq \Theta)$ (the default)

Illustration of the probability of error.

$\hat{\theta}_{MAP}(x) = \arg\max_{\theta} p(\theta|x)$

$\hat{\theta}_{MAP}(1)=3$ $\hat{\theta}_{MAP}(2)=2$ $\hat{\theta}_{MAP}(3)=1$ $\hat{\theta}_{MAP}(4)=4$ $\hat{\theta}_{MAP}(5)=1$.



$\mathbb{P}(\hat{\Theta}\neq\Theta|X=1) = \mathbb{P}(\hat{\theta}_{MAP}(1)\neq\Theta|X=1) = \mathbb{P}(3\neq\Theta|X=1) = 1-\mathbb{P}(\Theta=3|X=1) = 1-0.4 = 0.6$

$\mathbb{P}(\hat{\Theta}\neq\Theta|X=2) = 1-0.5=0.5$ $\mathbb{P}(\hat{\Theta}\neq\Theta|X=3)=1-0.6=0.4$ $\mathbb{P}(\hat{\Theta}\neq\Theta|X=4)=1-0.55=0.45$ $\mathbb{P}(\hat{\Theta}\neq\Theta|X=5)=1-0.45=0.55$

$\mathbb{P}(\Theta\neq\Theta) = 0.6\,P_X(1) + 0.5\,P_X(2) + 0.4\,P_X(3) + 0.45\,P_X(4) + 0.55\,P_X(5)$

MAP estimator minimizes $\mathbb{P}(\hat{\Theta} \neq \Theta)$

- Relation among Θ, X, $\hat{\Theta}$

$$\xrightarrow{\ \Theta\ } \boxed{p(x|\theta)} \xrightarrow{\ X\ } \boxed{\hat{\theta}_{MAP}(x)} \xrightarrow{\ \hat{\Theta}_{MAP}\ }$$

- $$\begin{aligned}
\mathbb{P}(\hat{\Theta}_{MAP} \neq \Theta) &= \sum_x \mathbb{P}(\hat{\Theta}_{MAP} \neq \Theta \mid X = x)\, p(x) \\
&= \sum_x \mathbb{P}(\underbrace{\hat{\theta}_{MAP}(x)}_{\text{a number}} \neq \Theta \mid X = x)\, p(x). \\
&= \sum_x \left[1 - \mathbb{P}(\Theta = \hat{\theta}_{MAP}(x) \mid X = x)\right] p(x) \\
&\leq \sum_x \left[1 - \mathbb{P}(\Theta = \hat{\theta}(x) \mid X = x)\right] p(x). \\
&= \mathbb{P}(\hat{\Theta} \neq \Theta) \qquad \text{for any } \hat{\theta}(x).
\end{aligned}$$

$\hat{\Theta}$ is a function of X

$\hat{\theta}_{MAP}(x) = \arg\max_{\theta} p(\theta|x)$

$\Leftrightarrow \mathbb{P}(\Theta = \hat{\theta}_{MAP}(x) | X = x)$

$\geq \mathbb{P}(\Theta = \hat{\theta}(x) | X = x)$

for any $\hat{\theta}(x)$