EE3731C Statistical Signal 2

BT Thomas Yeo

ECE, CIRC, Sinapse, Duke-NUS, HMS

Last Week Recap

- Discrete and continuous random variables
- Joint distributions p(X, Y)
  - Read as probability distribution of X and Y
  - Marginalization
  - Conditional probability
  - Bayes' Rule
  - Independence

$$E_{p(x|y)}[f(x,y)] \triangleq \int_x f(x,y)p(x|y)dx$$

- Expectation
  - Single variable: mean, variance
  - Two variables: covariance
  - Rules of expectation
  - Conditional expectation

Probabilistic Signal Detection

Probabilistic Signal Detection

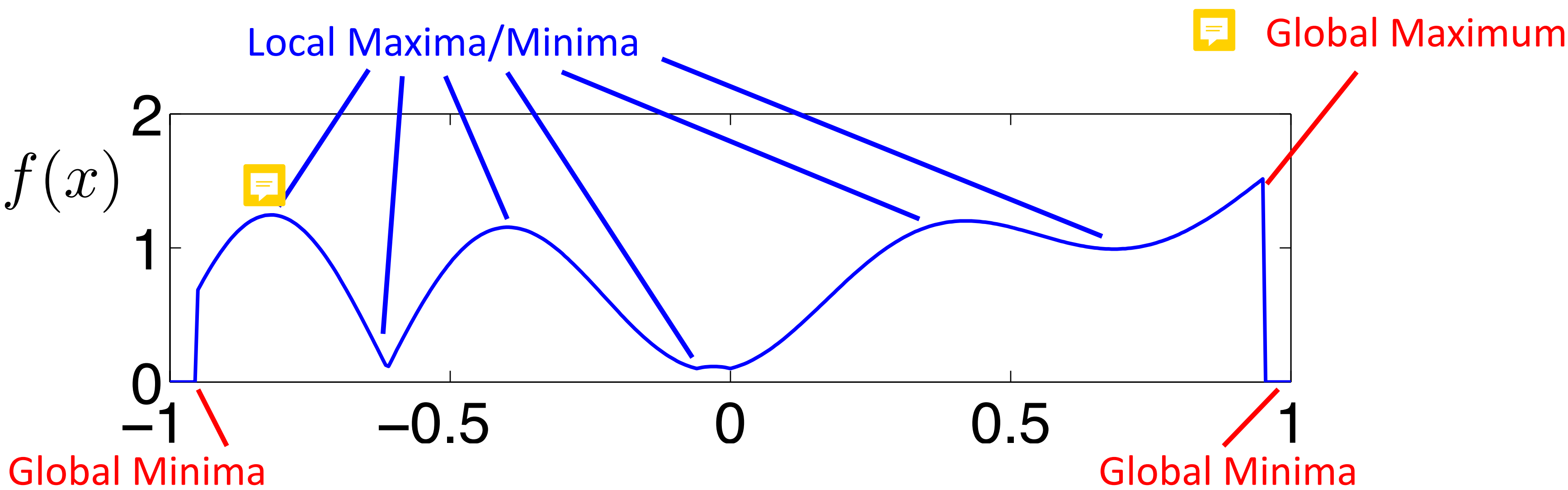
- Given observation y, estimate x by computing $x^* = d(y)$
  - "d" stands for "detection"
- In real world, have to take action based on what we think x is
  - y is cancer test results & x is whether patient has cancer
  - y is radar signal & x is whether there is an incoming missile
  - y is facebook photo & x is name of person in photo
- Maximum-A-Posteriori (MAP) Estimate:

$$x^* = \operatorname*{argmax}_x p(x|y),$$

where $p(x|y)$ is posterior probability

What is argmax and argmin?

- $\text{argmax}_x\, f(x)$ is value of x where $f(x)$ is biggest
- $\text{argmin}_x\, f(x)$ is value of x where $f(x)$ is smallest
- $f(x) = \begin{cases} |\sin(4x)^2 + x|\exp(-x) + x^2 + 0.1, & -0.95 \leq x \leq 0.95 \\ 0 & \text{otherwise} \end{cases}$
  - Generally easier to evaluate $f(x)$ than find maximum or minimum
  - Real problems: may have to live with local maximum or minimum



MAP Estimate for Cancer Example

- $y =$ cancer test result, x is whether student has cancer

0.001 0.99

$$P(+cancer| + test)$$
$$= \frac{P(+cancer)P(+test| + cancer)}{P(+cancer)P(+test| + cancer) + P(-cancer)P(+test| - cancer)}$$

0.001 0.99 0.999 0.05

MAP Estimate for Cancer Example

- $y =$ cancer test result, x is whether student has cancer

$$P(+cancer|+test)$$
$$= \frac{P(+cancer)P(+test|+cancer)}{P(+cancer)P(+test|+cancer) + P(-cancer)P(+test|-cancer)}$$
$$= 0.02$$

$$P(-cancer|+test) = 1 - P(+cancer|+test) = 0.98$$

- Therefore MAP estimate of x is

$$-cancer = \underset{x = \pm cancer}{\text{argmax}} \; p(x|+test)$$

MAP Properties

MAP minimizes Probability of Error

- Suppose we want detector $d(y)$ that minimizes probability of error

True value of x

$$p_e \triangleq p(d(y) \neq x)$$
$$= 1 - p(d(y) = x)$$
$$= 1 - \sum_y p(y, d(y) = x)$$
$$= 1 - \sum_y p(y) p(d(y) = x | y)$$

$$p(a) = 1 - p(\text{not } a)$$

$$p(a) = \sum_b p(a, b)$$

$$p(a, b) = p(b) p(a|b)$$

- $p(y) \geq 0$ for all y. Therefore, probability of error minimized by picking largest $p(d(y) = x | y)$ for each y, i.e., MAP estimation

Special Case: Maximum Likelihood Estimation

- MAP estimation:

$$
\begin{aligned}
x_{MAP} &= \underset{x}{\operatorname{argmax}}\, p(x|y) \\
&= \underset{x}{\operatorname{argmax}}\, \frac{p(x)p(y|x)}{p(y)} \\
&= \underset{x}{\operatorname{argmax}}\, p(x)p(y|x)
\end{aligned}
$$

Bayes' rule

p(y) not function of x

- $p(x)$ is constant $\implies$ maximum likelihood (ML) estimate

$$
x_{MAP} = \underset{x}{\operatorname{argmax}}\, p(y|x) \stackrel{\triangle}{=} x_{ML}
$$

- ML computationally easier; often use when prior unknown

MAP Example

MAP Game Example

- Let y have the distribution $p(y = 0) = 1/4$ and $p(y = 1) = 3/4$. Let

$$p(x|y = 0) = \begin{cases} 3/4 & x = 0 \\ 1/4 & x = 1 \end{cases} \quad \text{and} \quad p(x|y = 1) = \begin{cases} 1/8 & x = 0 \\ 7/8 & x = 1 \end{cases}$$

- Suppose I play the following game with a student everyday from day $n = 1$ to $n = \infty$. On day n, I independently sampled (x_n, y_n) from the above distribution $p(x, y)$. Given y_n, the student has to guess the value of x_n. Suppose the student guesses $\hat{x}_n$ on the n-th day. Then I pay the student $5\delta(x_n - \hat{x}_n)$.

  (i) How should the student play this game to maximize the payout on average?

  (ii) Under this optimal playing strategy, how much do I pay the student on average each day?

MAP Game Example

- How to maximize payout on average?
- Observe decisions for days when $y = 0$ and $y = 1$ can be separated
- Let $\hat{x} =$ decision when $y = 0$. To maximize average payout:

$$\operatorname*{argmax}_{\hat{x}} E_{p(x|y=0)}(5\delta(x - \hat{x})), \qquad \delta(x - \hat{x}) = 1 \text{ if } x = \hat{x} \text{ and } 0 \text{ otherwise}$$

$$= \operatorname*{argmax}_{\hat{x}} \sum_{x} 5\delta(x - \hat{x}) p(x|y = 0) \quad \text{by definition of conditional expectation}$$

$$= \operatorname*{argmax}_{\hat{x}} \Big(5\delta(0 - \hat{x}) p(x = 0|y = 0) + 5\delta(1 - \hat{x}) p(x = 1|y = 0) \Big)$$

$$= \operatorname*{argmax}_{x} p(x|y = 0) \quad \text{i.e., } \textbf{MAP estimate}$$

$$= 0 \qquad p(x = 0|y = 0) = 3/4; p(x = 1|y = 0) = 1/4$$

- Using MAP estimate, average payout (when $y = 0$) is $5 \times 3/4 = 15/4$

MAP Game Example

- How to maximize payout on average?
- Observe decisions for days when $y = 0$ and $y = 1$ can be separated
- Let $\hat{x}$ = decision when $y = 0$. To maximize average payout:

$$\begin{aligned}
&\underset{\hat{x}}{\operatorname{argmax}}\, E_{p(x|y=0)}(5\delta(x - \hat{x})), \qquad \delta(x - \hat{x}) = 1 \text{ if } x = \hat{x} \text{ and } 0 \text{ otherwise} \\
&= \underset{\hat{x}}{\operatorname{argmax}} \sum_x 5\delta(x - \hat{x})p(x|y = 0) \qquad \text{by definition of conditional expectation} \\
&= \underset{\hat{x}}{\operatorname{argmax}} \Big(5\delta(0 - \hat{x})p(x = 0|y = 0) + 5\delta(1 - \hat{x})p(x = 1|y = 0)\Big) \\
&= \underset{x}{\operatorname{argmax}}\, p(x|y = 0) \quad \text{i.e., } \textbf{MAP estimate} \\
&= 0 \qquad\qquad p(x = 0|y = 0) = 3/4; p(x = 1|y = 0) = 1/4
\end{aligned}$$

- Using MAP estimate, average payout (when $y = 0$) is $5 \times 3/4 = 15/4$
- If student guessed 1 instead, average payout only $5 \times 1/4 = 5/4$, justifying MAP estimate

MAP Game Example

- How to maximize payout on average?
- Observe decisions for days when $y = 0$ and $y = 1$ can be separated
- Let $\hat{x}$ = decision when $y = 1$. To maximize average payout:

$$\underset{\hat{x}}{\operatorname{argmax}} E_{p(x|y=1)}(5\delta(x - \hat{x})), \qquad \delta(x - \hat{x}) = 1 \text{ if } x = \hat{x} \text{ and } 0 \text{ otherwise}$$

$$= \underset{\hat{x}}{\operatorname{argmax}} \sum_x 5\delta(x - \hat{x})p(x|y = 1) \qquad \text{by definition of conditional expectation}$$

$$= \underset{\hat{x}}{\operatorname{argmax}} \left(5\delta(0 - \hat{x})p(x = 0|y = 1) + 5\delta(1 - \hat{x})p(x = 1|y = 1)\right)$$

$$= \underset{x}{\operatorname{argmax}}\, p(x|y = 1) \quad \text{i.e., } \textbf{MAP estimate}$$

$$= 1 \qquad\qquad p(x = 0|y = 1) = 1/8; p(x = 1|y = 1) = 7/8$$

- Using MAP estimate, average payout (when $y = 1$) is $5 \times 7/8 = 35/8$
- If student guessed 0 instead, average payout only $5 \times 1/8 = 5/8$, justifying MAP estimate

MAP Game Example

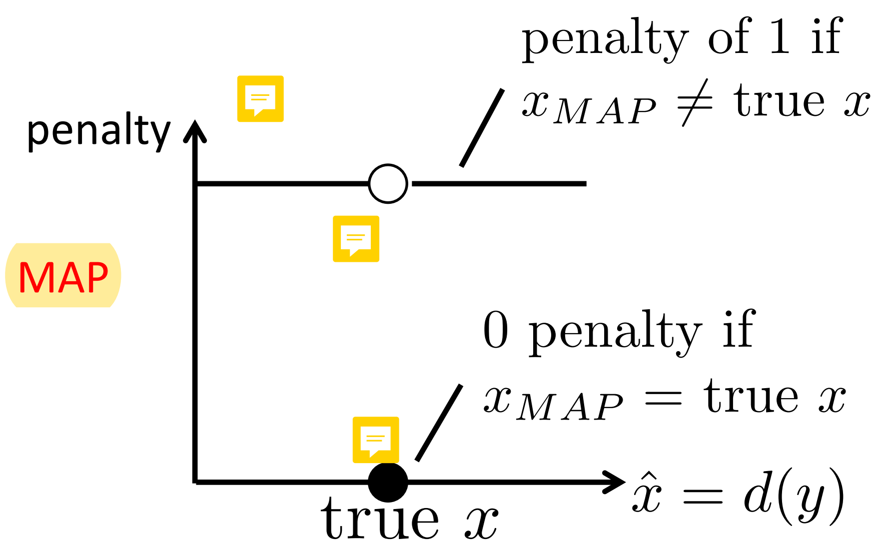
- Optimal strategy is MAP, resulting in $\hat{x} = \begin{cases} 0, & \text{if } y = 0 \\ 1, & \text{if } y = 1 \end{cases}$
- When $y = 0$, student receives $15/4$ dollars on average. Since $y = 0$ appears $1/4$ of days, I will pay student $15/4$ dollars for $1/4$ of days
- When $y = 1$, student receives $35/8$ dollars on average. Since $y = 1$ appears $3/4$ of days, I will pay student $35/8$ dollars for $3/4$ of days
- Therefore on average, I pay the student $1/4 \times 15/4 + 3/4 \times 35/8 =$ $135/32$ dollars each day
- "Formal" derivation at end of lecture notes
- If fixed reward R when estimate equals true x and zero reward when estimate not equal true x, then x_{MAP} gives highest reward on average
- If no penalty when estimate equals true x and fixed penalty P when estimate not equal true x, then x_{MAP} gives lowest penalty on average

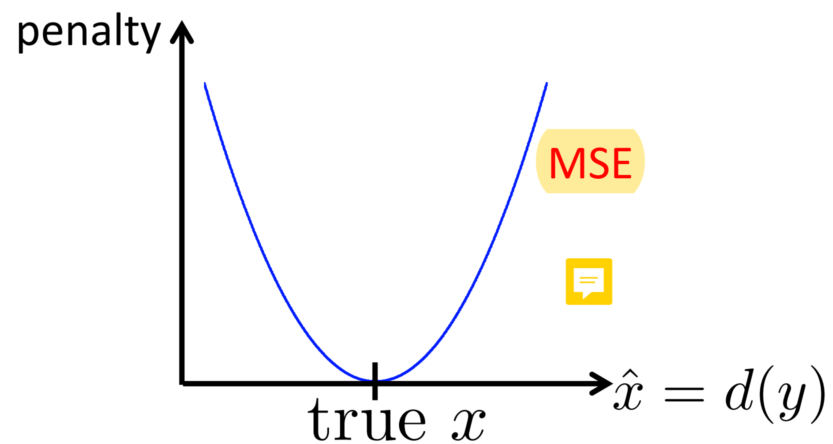
Minimum Mean Square Error (MMSE) Estimate

Minimum Mean Square Error (MMSE) Estimate

Let $\hat{x} = d(y)$ be estimate of x

- Suppose constant penalty if $x \neq \hat{x}$ and 0 penalty if $x = \hat{x}$, then MAP estimation is optimal
- Suppose penalty is $(x - \hat{x})^2$ (mean square error or MSE), then optimal estimator is called Minimum Mean Square Error (MMSE) estimate





Minimum Mean Square Error (MMSE) Estimate

- Recall $x_{MAP} = \text{argmax}_{\hat{x}} E_{p(x|y)}\delta(x - \hat{x}) = \text{argmax}_x p(x|y)$
- $x_{MMSE} = \text{argmin}_{\hat{x}} E_{p(x|y)}(x - \hat{x})^2 = ??$
- Conditional expectation of x given y is MMSE estimate of x given y:

$$E(x|y) = E_{p(x|y)}(x) = \underset{\hat{x}}{\text{argmin}}\, E_{p(x|y)}(x - \hat{x})^2$$

- Proof:
  - Last tutorial: $a = E_{p(x)}(x)$ minimizes $E_{p(x)}(x - a)^2$
  - Replace $p(x)$ with $p(x|y)$ and a with $\hat{x}$, we get $\hat{x} = E_{p(x|y)}(x)$ minimizes $E_{p(x|y)}(x - \hat{x})^2$
  - Note that $\hat{x} = E_{p(x|y)}(x)$ is a function of y

MMSE Example

MMSE Game Example

- Let y have the distribution $p(y = 0) = 1/4$ and $p(y = 1) = 3/4$. Let

$$p(x|y=0) = \begin{cases} 3/4 & x = 0 \\ 1/4 & x = 1 \end{cases} \quad \text{and} \quad p(x|y=1) = \begin{cases} 1/8 & x = 0 \\ 7/8 & x = 1 \end{cases}$$

- Suppose I play the following game with a student everyday from day $n = 1$ to $n = \infty$. On day n, I independently sampled (x_n, y_n) from the above distribution $p(x, y)$. Given y_n, the student has to guess the value of x_n. Suppose the student guesses $\hat{x}_n$ on the n-th day. Then the student has to pay me $(x_n - \hat{x}_n)^2$ dollar.

  (i) How should the student play this game to minimize the payout on average?

  (ii) Under this optimal playing strategy, how much does the student pay me on average each day?

MMSE Game Example

- How to minimize payout on average?
- Observe decisions for days when $y=0$ and $y=1$ can be separated
- Let $\hat{x}$ = decision when $y=0$. To minimize average payout:

$$\underset{\hat{x}}{\operatorname{argmin}} E_{p(x|y=0)}\left((\hat{x}-x)^2\right) = E_{p(x|y=0)}(x) \quad \textbf{i.e., MMSE estimator!}$$

$$= \sum_x x p(x|y=0) \quad \text{conditional expectation}$$

$$= 0 \times p(x=0|y=0) + 1 \times p(x=1|y=0)$$

$$= 1/4$$

- Using MMSE estimate, average payout (when $y=0$) is

$$E_{p(x|y=0)}\left((x_{MMSE}-x)^2\right) = \sum_x (x_{MMSE}-x)^2 p(x|y=0)$$

$$= (1/4-0)^2 p(x=0|y=0) + (1/4-1)^2 p(x=1|y=0)$$

$$= 1/16 \times 3/4 + 9/16 \times 1/4 = 3/16$$

MMSE Game Example

- How to minimize payout on average?
- Observe decisions for days when $y = 0$ and $y = 1$ can be separated
- Let $\hat{x} =$ decision when $y = 1$. To minimize average payout:

$$\operatorname*{argmin}_{\hat{x}} E_{p(x|y=1)}\left((\hat{x} - x)^2\right) = E_{p(x|y=1)}(x) \qquad \textbf{i.e., MMSE estimator!}$$

$$= \sum_x x p(x|y=1) \qquad \text{conditional expectation}$$

$$= 0 \times p(x=0|y=1) + 1 \times p(x=1|y=1)$$

$$= 7/8$$

- Using MMSE estimate, average payout (when $y = 1$) is

$$E_{p(x|y=1)}\left((x_{MMSE} - x)^2\right) = \sum_x (x_{MMSE} - x)^2 p(x|y=1)$$

$$= (7/8 - 0)^2 p(x=0|y=1) + (7/8 - 1)^2 p(x=1|y=1)$$

$$= 49/64 \times 1/8 + 1/64 \times 7/8 = 7/64$$

MMSE Game Example

- Optimal strategy is MMSE, resulting in $\hat{x} = \begin{cases} 1/4, & \text{if } y = 0 \\ 7/8, & \text{if } y = 1 \end{cases}$
- When $y = 0$, student pays $3/16$ dollars on average. Since $y = 0$ appears $1/4$ of days, student will pay $3/16$ dollars for $1/4$ of days
- When $y = 1$, student pays $7/64$ dollars on average. Since $y = 1$ appears $3/4$ of days, student will pay $7/64$ dollars for $3/4$ of days
- Therefore on average, the student pays $3/16 \times 1/4 + 7/64 \times 3/4 = 33/256$ dollars each day
- "Formal" derivation at end of lecture notes
- If penalty is $(x - \hat{x})^2$, then x_{MMSE} gives lowest penalty on average

N random variables (aka random vector)

- Have focused on 2 random variables x and y
- In real applications, usually more than 2 variables (e.g., photo has > 1M pixels)
- If we observe $x_1, x_2, \ldots, x_N$ multiple times, some combinations of outcomes more likely than others
- This information captured by joint probability distribution function
- Written as $p(x_1, x_2, \ldots, x_N)$, read as probability distribution of x_1 to x_N
- If $x_1, x_2, \ldots, x_N$ continuous, then p refers to joint probability distribution function (pdf). If discrete, then refers to joint probability mass function (pmf)
- Many properties for two random variables generalize naturally to more variables

Marginalization / Law of Total Probability

We can recover probability distribution of any variable in a joint distribution by integrating (or summing) over the other variables

$$Pr(x) = \int Pr(x,y)\,dy$$

$$Pr(y) = \int Pr(x,y)\,dx$$

Works in higher dimensions as well – leaves joint distribution between whatever variables are left

$$Pr(x,y) = \sum_{w} \int Pr(w,x,y,z)\,dz$$

Conditional Probability

- Two variables

$$p(x, y) = p(x)p(y|x)$$

- Three variables

$$p(a, b, c) = p(a)p(b, c|a) = p(a)p(b|a)p(c|a, b)$$

- N variables

$$\begin{aligned}
p(x_1, \cdots, x_N) &= p(x_1)p(x_2, \cdots, x_N|x_1) \\
&= p(x_1)p(x_2|x_1)p(x_3, \cdots, x_N|x_1, x_2) \\
&= p(x_1)p(x_2|x_1)p(x_3|x_1, x_2) \cdots p(x_N|x_1, \cdots, x_{N-1})
\end{aligned}$$

Independence

- If $x_1, \cdots, x_N$ are independent, then knowing any subset of x's tells us nothing about the remaining x's
- If $x_1, \cdots, x_N$ are independent if and only if the joint distribution factorizes into product of marginal distributions:

$$p(x_1, \cdots, x_N) = p(x_1)p(x_2)\cdots p(x_N) \triangleq \prod_{n=1}^{N} p(x_n)$$

- $x_1, \cdots, x_N$ are independently and identically distributed (i.i.d.) if they are independent and $p(x_1) = p(x_2) = \cdots = p(x_N)$

Conditional Independence

- x_1 and x_2 are conditionally independent given x_3 if and only if

$$p(x_1, x_2|x_3) = p(x_1|x_3)p(x_2|x_3)$$

Knowing x_2 tells us nothing about x_1 (and vice versa) if we already know x_3

Special Case I

- If x_1 and x_2 are independent, it does NOT imply x_1 and x_2 are conditionally independent given x_3
- Example: Let x = 1 if Thomas comes to class wet (0 otherwise). Let y = 1 if raining (0 otherwise). Let z = 1 if Thomas involved in water fight (0 otherwise)
  - Then y and z are independent (presumably)
  - But y and z are not conditionally independent given x
  - Because suppose I come to class wet. Then knowing it's not raining would suggest I was in a water fight

Special Case II

- If x_1 and x_2 are conditionally independent given x_3, it does NOT imply x_1 and x_2 are independent
- Example: Toss coin 99 times. Let q = probability of head. Let x_n = 1 if n-th coin toss = head and x_n= 0 if n-th coin toss = tail
  - x_1, x_2, … x_{99} are independent conditioned on knowing q (e.g., q = 0.7)
  - If q is NOT known, then x_1, x_2, … x_{99} are not independent
  - Imagine if all 99 coin tosses are equal to head, what would you guess about the 100th coin toss?
- Probably confusing because in your previous statistics class, you might have seen equation like $p(HHT) = q^2(1-q)$
  - This cannot be right since q does not appear on left hand side? How can it appear on right hand side?
  - Instead it should actually be $p(HHT \mid q) = q^2(1-q)$. The term "| q" is implicit and is often dropped to reduce clutter

Summary

Summary

- Estimating x given observation y
  - $x_{MAP} = \text{argmax}_{\hat{x}}\, E_{p(x|y)}\delta(x - \hat{x}) = \text{argmax}_x\, p(x|y)$
  - x_{MAP} minimizes probability of error
  - $x_{ML} = x_{MAP}$ when $p(x)$ is constant
  - $x_{MMSE} = \text{argmin}_{\hat{x}}\, E_{p(x|y)}(x - \hat{x})^2 = E_{p(x|y)}(x)$
- N random variables (aka random vector)
  - Conditional independence
  - Next week: probabilistic signal detection for N variables

Further Optional Readings

- Chapters 3 and 4 of Computer Vision: models, learning and inference. Free Download: http://www.computervisionmodels.com/

- Search for terms on Wikipedia like “maximum a posteriori”, “maximum likelihood”, “minimum mean square error”,

Additional Material

MAP Game

- Average payout for MAP game is

$$
\begin{aligned}
&\sum_x \sum_y p(x,y) 5\delta(x - x_{MAP}(y)) \\
&= \sum_x \sum_y p(y) p(x|y) 5\delta(x - x_{MAP}(y)) \quad \text{by definition of conditional probability} \\
&= \sum_y p(y) \sum_x p(x|y) 5\delta(x - x_{MAP}(y)) \quad \text{by swapping ordering of the sum} \\
&= p(y=0) \sum_x p(x|y=0) 5\delta(x - x_{MAP}(y=0)) \\
&\qquad + p(y=1) \sum_x p(x|y=1) 5\delta(x - x_{MAP}(y=1)) \\
&= p(y=0) E_{p(x|y=0)}(5\delta(x - x_{MAP}(y=0))) \\
&\qquad + p(y=1) E_{p(x|y=1)}(5\delta(x - x_{MAP}(y=1))) \\
&= 1/4 \times 15/4 + 3/4 \times 35/8 = 135/32
\end{aligned}
$$

MMSE Game

- Average payout for MMSE game is

$$
\begin{aligned}
&\sum_x \sum_y p(x,y)(x_{MMSE}(y) - x)^2 \\
&= \sum_x \sum_y p(y)p(x|y)(x_{MMSE}(y) - x)^2 \quad \text{by definition of conditional probability} \\
&= \sum_y p(y) \sum_x p(x|y)(x_{MMSE}(y) - x)^2 \quad \text{by swapping ordering of the sum} \\
&= p(y=0) \sum_x p(x|y=0)(x_{MMSE}(y=0) - x)^2 \\
&\qquad\qquad + p(y=1) \sum_x p(x|y=1)(x_{MMSE}(y=1) - x)^2 \\
&= p(y=0) E_{p(x|y=0)} \left((x_{MMSE}(y=0) - x)^2 \right) \\
&\qquad\qquad + p(y=1) E_{p(x|y=1)} \left((x_{MMSE}(y=0) - x)^2 \right) \\
&= 1/4 \times 3/16 + 3/4 \times 7/64 = 33/256
\end{aligned}
$$