SCAN ME

CS5340:
Tutorial 2

Asst. Prof. Harold Soh

TA: Eugene Lim

Poll Everywhere!

https://pollev.com/haroldsohsoo986

SCAN ME

Upcoming Dates

- **Project**
  - Form teams: Feb 6th
  - Abstract due: Feb 27th
- Piazza teams

Course Schedule (Tentative)

Week	Date	Lecture Topic	Tutorial
1	16 Jan	Introduction to Uncertainty Modeling + Probability Basics	~~Introduction~~
2	23 Jan	Simple Probabilistic Models	Introduction and Probability Basics
3	30 Jan	Bayesian networks (Directed graphical models)	More Basic Probability
4	6 Feb	Markov random Fields (Undirected graphical models)	DGM modelling and d-separation
5	13 Feb	Variable elimination and belief propagation	MRF + Sum/Max Product
6	20 Feb	Factor graphs	**Quiz 1**
-	-	**RECESS WEEK**	
7	5 Mar	Mixture Models and Expectation Maximization (EM)	Linear Gaussian Models
8	12 Mar	Hidden Markov Models (HMM)	Probabilistic PCA
9	19 Mar	Monte-Carlo Inference (Sampling)	Linear Gaussian Dynamical Systems
10	26 Mar	Variational Inference	MCMC + Langevin Dynamics
11	2 Apr	Inference and Decision-Making	Diffusion Models + Sequential VAEs
12	9 Apr	Gaussian Processes (optional)	**Quiz 2**
13	16 Apr	Project Presentations	Closing Lecture

CS5340:

Tutorial 2

Asst. Prof. Harold Soh

TA: Eugene Lim

1. Uncorrelated Random Variables

SCAN ME

https://pollev.com/haroldsohsoo986

a. Two random vars X and Y where

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$$

$p(X,Y) = p(X)p(Y)$

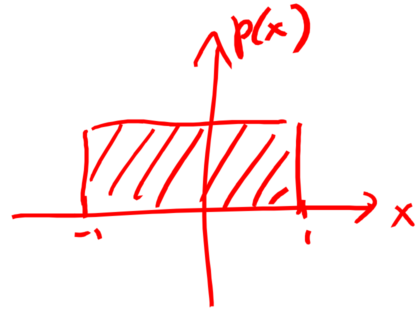
Are X and Y **independent**? Justify your answer.

if X and Y are independent $\Rightarrow \mathrm{Cov}[X,Y] = 0$

$\not\Leftarrow$

$X \sim \text{Uniform}[-1, 1]$

$Y = X^2$

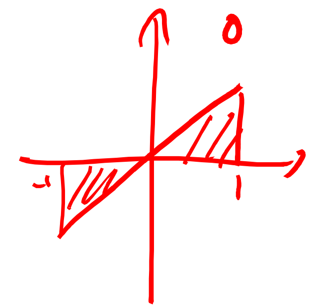


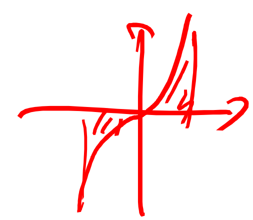
$$\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].$$

$$= \mathbb{E}[\underline{\underline{X}}^3] - \mathbb{E}[\underline{X}]\,\mathbb{E}[X^2] = 0.$$

$$\mathbb{E}[X] = \int_{-1}^{1} x\, p(x)\, dx = \int_{-1}^{1} x \frac{1}{2}\, dx = 0$$

$$\mathbb{E}[X^3] = \int_{-1}^{1} x^3 p(x)\, dx = \int_{-1}^{1} \frac{x^3}{2}\, dx = 0$$





1. Uncorrelated Random Variables

SCAN ME

https://pollev.com/haroldsohsoo986

b. Consider:

- samples $x_1, \ldots, x_N$ drawn from $p(X)$
- samples $y_1, \ldots, y_N$ drawn from $p(Y)$

We want to model the joint $p_\theta(X, Y)$

Can we just perform MLE by finding

$$\mathrm{argmax}_\theta \sum_i^N \log p_\theta(x_i, y_i) \text{ ?}$$

assume $Y = X$

X	Y
0	0
1	1
0	0
1	1
1	1
0	0

$X \sim \text{Bern}(0.5)$

$p(X, Y)$.

X	Y	
0	0	$\theta_{00} = 0.5$
0	1	$\theta_{01} = 0$
1	0	$\theta_{10} = 0$
1	1	$\theta_{11} = 0.5$

$(x, y) \sim p(X, Y)$.

$X \sim p(X)$ $\quad$ $Y \sim p(Y)$

X	Y
0	1
1	0
0	1
0	0
1	1
1	0
0	0
1	1
⋮	

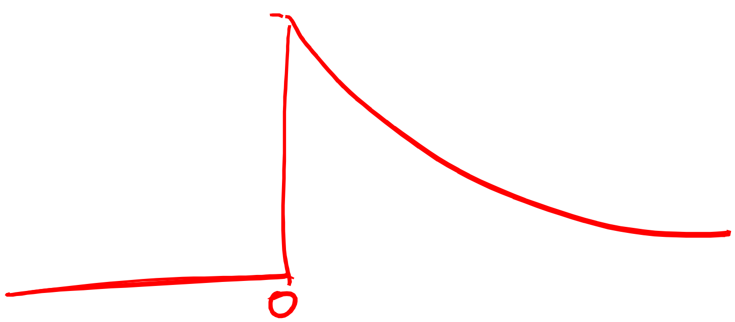
X	Y	
0	0	0.25
0	1	0.25
1	0	0.25
1	1	0.25.

Exponential Family (ExpFam)

2.a. Show the exponential distribution is exponential family (ExpFam).

$$p(x) = \frac{h(x)\exp(\eta^T s(x))}{Z(\eta)} = h(x)\exp\left(\eta^T s(x) - A(\eta)\right) \qquad A(\eta) = \log Z(\eta)$$

$$p(x) = 1 \times \lambda \exp(-\lambda x) \quad x \geq 0.$$



$s(x) = x$

$\eta = -\lambda$

$\rightarrow h(x) = 1$ if $x \geq 0$, 0 otherwise

$Z(\eta) = \frac{1}{\lambda} = \frac{-1}{\eta}$

$A(\eta) = \log\left(\frac{-1}{\eta}\right)$

ExpFam MLE

$$\mathbb{E}_{x \sim D(\eta)}[s(x)] = \nabla_\eta A(\eta) \quad \text{— } 1$$

$$\boxed{\nabla_\eta A(\eta_{MLE}) = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad x_1, \dots, x_N} \quad \text{— } 2$$

2.b. Derive the MLE for the exponential distribution using facts about ExpFam.

1. Find $\nabla_\eta A(\eta)$
2. Evaluate $\nabla_\eta A(\eta)$ at η_{MLE}.
3. Use 2 and solve.

1. Find $\nabla_\eta A(\eta)$:

$$\frac{d}{d\eta} \log\left(\frac{-1}{\eta}\right) = \frac{1/\eta^2}{(-1/\eta)} = \frac{1}{-\eta}$$

2. Evaluate … η_{MLE}, 3. Use 2 to solve

$$\frac{-1}{\eta_{MLE}} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\Rightarrow \eta_{MLE} = \frac{-1}{\frac{1}{N}\sum_{i=1}^{N} x_i}$$

$$\lambda_{MLE} = -\eta_{MLE} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} x_i}$$

Gaussian is ExpFam

$$p(x) = \underline{h(x)} \exp\left(\underline{\eta^T s(x)} - \underline{A(\eta)}\right)$$

$h(x) = 1 \qquad Z(\eta) = \sqrt{2\pi\sigma^2}$

2.c.1. Show that the Gaussian is ExpFam.

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(\log\frac{1}{\sigma}\right) \exp\left(\frac{-(x^2 - 2x\mu + \mu^2)}{2\sigma^2}\right)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(\log\frac{1}{\sigma} - \frac{x^2}{2\sigma^2} + \frac{x\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(\underbrace{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2}_{\eta^T s(x)} - \left(\underline{\log\sigma + \frac{\mu^2}{2\sigma^2}}\right)\right)$$

$$= \underbrace{\begin{bmatrix} \mu/\sigma^2 \\ -1/2\sigma^2 \end{bmatrix}^T}_{\eta} \underbrace{\begin{bmatrix} x \\ x^2 \end{bmatrix}}_{s(x)}$$

$$\begin{cases} h(x) = \dfrac{1}{\sqrt{2\pi}} \\ A(\eta) = \log\sigma + \dfrac{\mu^2}{2\sigma^2} \\ \quad = \dfrac{-\eta_1^2}{4\eta_2} - \dfrac{1}{2}\log(-2\eta_2) \\ \eta = \begin{bmatrix} \mu/\sigma^2 \\ -1/2\sigma^2 \end{bmatrix} \begin{matrix} \leftarrow \eta_1 \\ \leftarrow \eta_2 \end{matrix} \\ s(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \end{cases}$$

TRY VERIFY YOURSELF

$\hat{\eta}_1$: MLE

Gaussian MLE

$$\nabla_\eta A(\eta_{MLE}) = \frac{1}{N}\sum_{i=1}^{N} S(x_i)$$

$$S(x_i) = \begin{bmatrix} x_i \\ x_i \end{bmatrix}$$

$$A(\eta) = \frac{-\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)$$

2.c.2. Derive the MLE of the Gaussian's natural parameters.

$$\nabla_\eta A(\eta) = \begin{bmatrix} \frac{\partial}{\partial \eta_1} A(\eta) \\ \frac{\partial}{\partial \eta_2} A(\eta) \end{bmatrix}$$

$$\frac{\partial}{\partial \eta_1} A(\eta) = \frac{-2\eta_1}{4\eta_2} = \frac{-\eta_1}{2\eta_2}$$

$$\frac{\partial}{\partial \eta_2} A(\eta) = \frac{\eta_1^2}{4\eta_2^2} - \frac{1}{2}\left(\frac{-2}{-2\eta_2}\right)$$

$$= \frac{\eta_1^2}{4\eta_2^2} - \frac{1}{2\eta_2}$$

$$\frac{-2\hat{\eta}_1}{4\hat{\eta}_2} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\|$$

$$\frac{-2\left(\frac{\hat{\mu}}{\cancel{\sigma^2}}\right)}{2 \cdot 4\left(\frac{-1}{2\cancel{\sigma^2}}\right)} = \hat{\mu}$$

$$\frac{\tilde{\eta}_1^2}{4\hat{\eta}_2^2} - \frac{1}{2\hat{\eta}_2} = \frac{1}{N}\sum_{i=1}^{N} x_i^2$$

$$= \hat{\mu}^2 + \hat{\sigma}^2$$

$$\frac{\hat{\eta}_1^2}{4\hat{\eta}_2^2} = \frac{\left(\frac{\tilde{\mu}}{\hat{\sigma}^2}\right)^2}{4\left(\frac{-1}{2\hat{\sigma}^2}\right)^2} = \hat{\mu}^2$$

$+$

$$\frac{-1}{2\hat{\eta}_2} = \frac{-1}{2\left(\frac{-1}{2\hat{\sigma}^2}\right)} = \hat{\sigma}^2$$

$$\hat{\mu}^2 + \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2$$

$$\hat{\sigma}^2 = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right) - \hat{\mu}^2 \cdot \frac{N}{N}$$

$$\sum \frac{1}{N}\hat{\mu}^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(x_i^2 - \hat{\mu}^2\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(x_i^2 - 2\hat{\mu}^2 + \hat{\mu}^2\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(x_i^2 - 2x_i\hat{\mu} + \hat{\mu}^2\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \hat{\mu}\right)^2$$

$$\frac{1}{N}\sum x_i\hat{\mu}$$

$$= \frac{1}{N}\hat{\mu}\sum x_i$$

$$= \hat{\mu}\hat{\mu}$$

$$= \hat{\mu}^2$$

Questions?

https://pollev.com/haroldsohsoo986

SCAN ME

3. Meme of the Year

https://pollev.com/haroldsohsoo986

CS5340 student: Let me just skip solving tutorials.

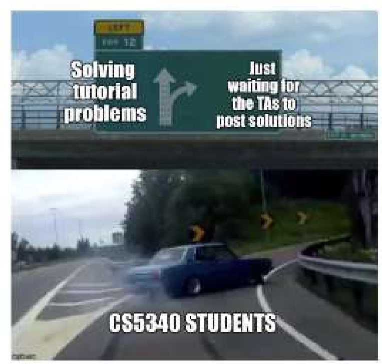
screws up in the final exam

CS5340 student:

(a) Surprised Pikachu

(b) Two Buttons Dilemma

(c) Distracted Boyfriend



(d) Left Exit 12

3. Meme of the Year

a. What model (distribution) can we use to model this data? Compute the parameters via MLE.

$x_1 = [1 \ 0 \ 0 \ 0]$

$x_2 = [0 \ 0 \ 1 \ 0]$

Categorical Distribution

- Discrete variables $\boldsymbol{X}$ that take on 1-of-*K* possible mutually exclusive states, e.g. a *K*-faced die.
- $\boldsymbol{x}$ is represented by a *K*-dimensional vector $\mathbf{e}_k$ in which one of the elements $x_k = 1$, and $\sum_{k=1}^{K} x_k = 1$.
- e.g. $K = 5$, and $\boldsymbol{x} = \mathbf{e}_3 = [0,0,1,0,0]^T$.
- *K* parameters $\lambda = [\lambda_1, \dots, \lambda_K]^T$, where $\lambda \geq 0$, $\sum_k \lambda_k = 1$.

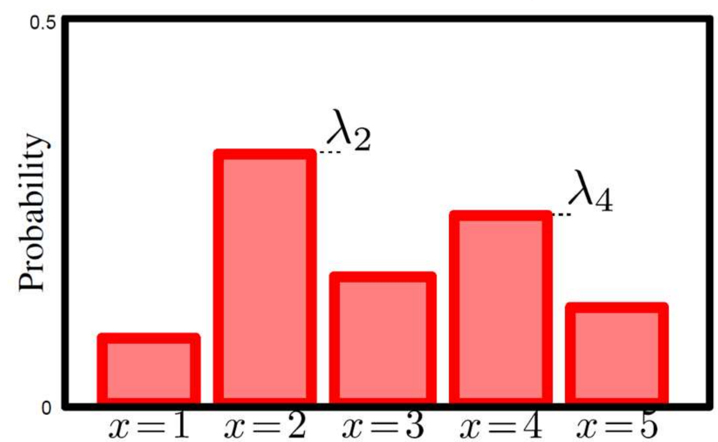
K-1

$$p(\boldsymbol{X} = \mathbf{e}_k \mid \lambda) = \lambda_k$$

Or

$$p(\boldsymbol{x}) = \prod_{k=1}^{K} \lambda_k^{x_k} = \lambda_k,$$

$$p(\boldsymbol{x}) = \mathrm{Cat}_{\boldsymbol{x}}[\lambda]$$



Images Source: “Computer Vision: Models, Learning, and Inference”, Simon Prince

3. Meme of the Year

a. What model (distribution) can we use to model this data? Compute~~r~~ the parameters via MLE.

$Cat[\vec{\lambda}]$ $\quad D = \{x_1, \ldots x_N\}.$

$$\boxed{\vec{\lambda}_{MLE} = \arg\max p(D|\vec{\lambda})}$$

$$p(D|\vec{\lambda}) = \underbrace{\prod_{i=1}^{N} \prod_{j=1}^{k} \lambda^{[x_{ij}=1]}}_{} \qquad \left(\sum_{j=1}^{k} \lambda_j = 1 \right)$$

$$= \prod_{j=1}^{k} \lambda_j^{N_j}$$

maximize $f(\theta)$

s.t. $g(\theta) = 0$

$$\arg\max \log p(D|\vec{\lambda}) = \sum_{j=1}^{k} N_j \log \lambda_j + v\left(\sum_{j=1}^{k} \lambda_j - 1\right).$$

$$\mathcal{L} = \sum_{j=1}^{k} N_j \log \lambda_j + v\left(\sum_{j=1}^{k} \lambda_j - 1\right)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_j} = \frac{N_j}{\lambda_j} + v = 0 \Rightarrow \lambda_j = -\frac{N_j}{\underline{\underline{v}}}$$

Const.: $\sum_{j=1}^{k} \lambda_j = 1 \Rightarrow \sum_{j=1}^{k} -\frac{N_j}{v} = 1 \Rightarrow v = -\sum_{j=1}^{k} N_j$

$$\boxed{\lambda_j = \frac{N_j}{\sum_{\ell=1}^{k} N_\ell}.}$$

3. Meme of the Year

b. What prior distribution can we use for our model?

$p(\vec{\lambda})$

Dirichlet Distribution

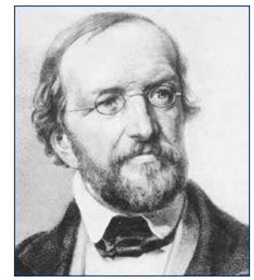
$\Sigma \lambda = 1$

$[\lambda_1, \lambda_2, \lambda_3]$

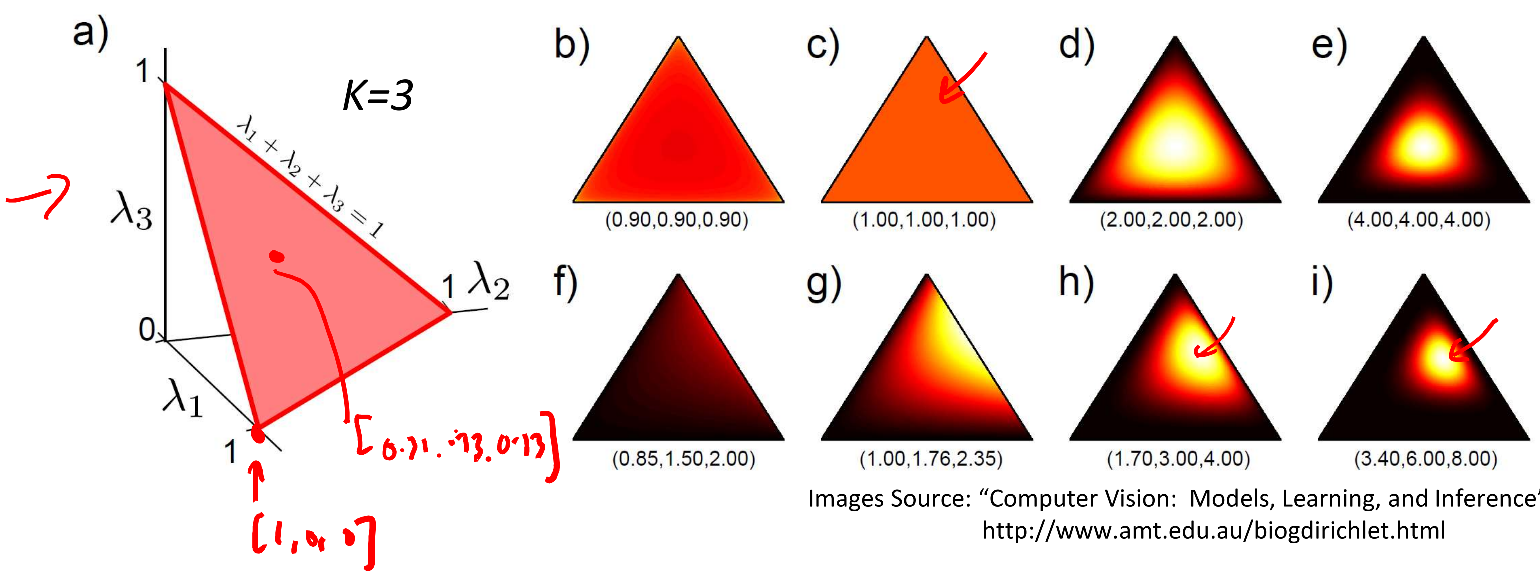
- Conjugate distribution of categorical distribution.
- Defined over K parameters of Categorical distribution, $\lambda_k \in [0,1]$, where $\sum_k \lambda_k = 1$.

$$p(\lambda_1, \ldots, \lambda_K) = \frac{\Gamma[\sum_{k=1}^{K} \alpha_k]}{\prod_{k=1}^{K} \Gamma[\alpha_k]} \prod_{k=1}^{K} \lambda_k^{\alpha_k - 1},$$

$$p(\lambda_1, \ldots, \lambda_K) = \mathrm{Dir}_{\lambda_{1 \ldots K}}[\alpha_1, \ldots \alpha_K]$$



Peter Gustav Lejeune Dirichlet (1805-1859)



Images Source: “Computer Vision: Models, Learning, and Inference”, Simon Prince
http://www.amt.edu.au/biogdirichlet.html

Dirichlet Distribution

Useful property:

$$\Gamma(1) = 1$$

$$\Gamma(z+1) = z\Gamma(z)$$

$$p(\lambda_1, \dots, \lambda_K) = \frac{\Gamma[\sum_{k=1}^{K} \alpha_k]}{\prod_{k=1}^{K} \Gamma[\alpha_k]} \prod_{k=1}^{K} \lambda_k^{\alpha_k - 1},$$

$$p(\lambda_1, \dots, \lambda_K) = \mathrm{Dir}_{\lambda_{1\dots K}}[\alpha_1, \dots \alpha_K]$$

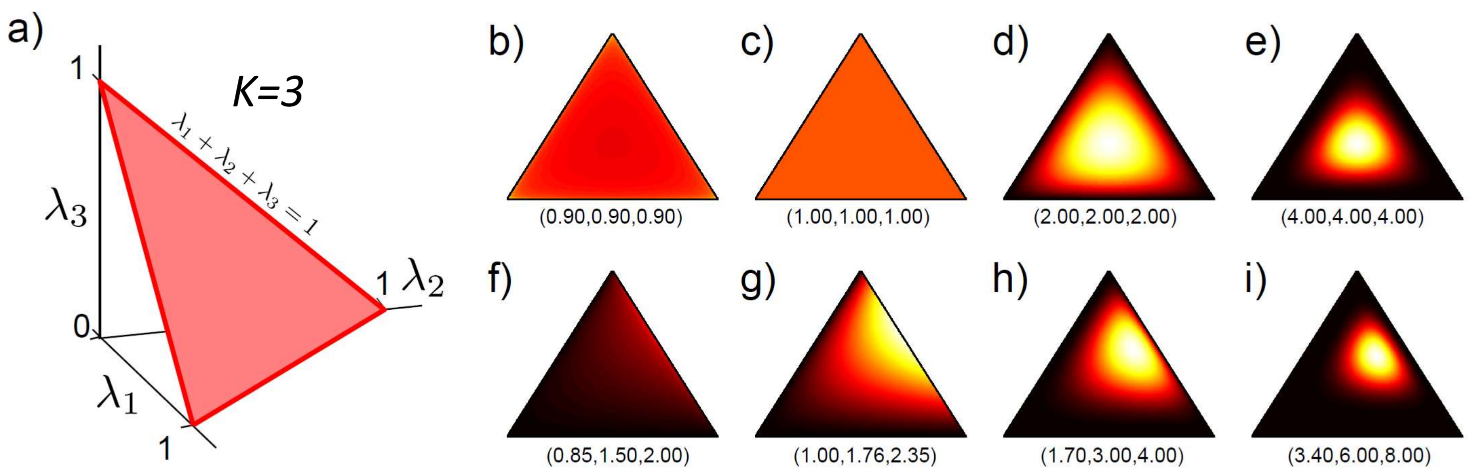
Gamma & Beta Functions:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$$

$\Gamma(n) = (n-1)!$

$$B(\underline{\boldsymbol{\alpha}}) = \frac{\prod_{k=1}^{K} \Gamma[\alpha_k]}{\Gamma[\sum_{k=1}^{K} \alpha_k]}$$

- *K* hyperparameters $\alpha_k > 0$.



Images Source: “Computer Vision: Models, Learning, and Inference”, Simon Prince

3. Meme of the Year

b. What is the posterior distribution?

$$p(\vec{\lambda}) = \frac{1}{B(\alpha)} \prod_{j=1}^{k} \lambda_j^{\alpha_j - 1}$$

$$p(X|\vec{\lambda}) = \prod_{j=1}^{k} \lambda_j^{N_j}$$

Bayes

$$p(\vec{\lambda}|X) \propto p(X|\vec{\lambda})\, p(\vec{\lambda})$$

$$\left[\prod_{j=1}^{k} \lambda_j^{N_j}\right]\left[\frac{1}{B(\vec{\alpha})} \prod_{l=1}^{k} \lambda_l^{\alpha_l - 1}\right]$$

$$= \underbrace{\frac{1}{B(\vec{\alpha})}}_{\text{const wrt } \lambda} \prod_{j=1}^{k} \lambda_j^{\alpha_j + N_j - 1} \Rightarrow \mathrm{Dir}[\tilde{\alpha}]$$

$$[\alpha_1 + N_1, \alpha_2 + N_2, \ldots, \alpha_k + N_k]^T$$

Useful property:

$$\Gamma(1) = 1$$
$$\Gamma(z+1) = z\Gamma(z)$$

3. Meme of the Year

b. What is the posterior predictive distribution?

$p(\hat{\theta} \mid X)$

$$p(x^* \mid X) = \int \overbrace{p(x^* \mid \vec{\lambda})}^{\text{Cat likelihood}} \overbrace{p(\vec{\lambda} \mid X)}^{\text{Dirch}[\tilde{\alpha}]} d\lambda$$

$$= \int \prod_{j=1}^{k} \lambda_j^{[x_j^* = 1]} \cdot \frac{1}{B(\tilde{\alpha})} \prod_{j=1}^{k} \lambda_j^{\tilde{\alpha}_j - 1} d\lambda.$$

$=$

multipy and divide by

$$\frac{\prod_{j=1}^{k} \Gamma(\tilde{\alpha}_j + [x_j^* = 1])}{\Gamma(\sum_{j=1}^{k} \tilde{\alpha}_j + [x_j^* = 1])}$$

Questions?

https://pollev.com/haroldsohsoo986

SCAN ME

Please Prepare for Next Week

- Watch the videos
  - Bayesian Networks!
- Do the tutorial