CSC412 Tutorial 1

Distribution over Discrete Random Variables

&

MLE and Exponential Families

Chen-Hao (Lance) Chao

Distribution over Discrete Random Variables

- COVID-19 diagnostic test; two variables: H and Y
- **Infected**: H = 1; not infected: H = 0
- **Test positive**: Y = 1; negative: Y = 0

Target: **p(H = h | Y = y)**   e.g. if the test is positive (Y=1), how likely are you infected (H)?

COVID-19 Diagnostic Test

- **Known Quantities:**
  - **Sensitivity** (true positive rate): $p(Y = 1 \mid H = 1) = 0.875$
  - **Specificity** (true negative rate): $p(Y = 0 \mid H = 0) = 0.975$
  - **Prevalence**: $p(H = 1) = 0.1$

How to find $p(H = h \mid Y = y)$?

- First find the joint distribution (Y, H):
  - $p(Y = 0, H = 0) =$
  - $p(Y = 0, H = 1) =$
  - $p(Y = 1, H = 0) =$
  - $p(Y = 1, H = 1) =$

COVID-19 Diagnostic Test

- Now suppose you test positive
- What is $p(H = 1 \mid Y = 1)$?
- $p(H = 1 \mid Y = 1) =$

- What about $p(H = 1 \mid Y = 0)$?

Y \ H	0	1	
0	0.8775	0.0125	0.89
1	0.0225	0.0875	0.11
	0.9	0.1	

Rows: Y (test result) · Columns: H (infection status)

MLE and Exponential Families

N-sample example: the Multinomial distribution

A random variable $X \sim \mathrm{Multinomial}(\mathbf{q})$ where $\mathbf{q} \in \mathbb{R}^K$, which takes on $i \in [1, \ldots, K]$ discrete states each with probability q_i. That is $p(X = i) = q_i$, where $q_i \geq 0$ and $\sum_i q_i = 1$.

e.g. K-bit pixels, classes, unfair dice.

We observe *N* i.i.d. samples from Multinomial(**q**), i.e. $\mathcal{D}$ = {1, 3, K, 2, …} for *N* observations.

Multinomial Distribution

Recall that the exponential family natural form is

$$p(x|\eta) = h(x)\exp\{\eta^\top T(x) - A(\eta)\}$$

Our multinomial distribution is $p(x^{(n)} = i \mid q) = q_i$ with constraint $\Sigma_i\, q_i = 1$

$$p(\mathcal{D}; \mathbf{q}) = \prod_{n=1}^{N} q_1^{1[x^{(n)}=1]} \cdots q_K^{1[x^{(n)}=K]}$$

Multinomial Distribution

$$
\begin{aligned}
p(\mathcal{D};\mathbf{q}) &= \prod_{n=1}^{N} q_1^{1[x^{(n)}=1]} \cdots q_K^{1[x^{(n)}=K]} \\
&= \exp \log \prod_{n=1}^{N} \prod_{i=1}^{K} q_i^{1[x^{(n)}=i]} \\
&= \exp \sum_{n=1}^{N} \sum_{i=1}^{K} 1[x^{(n)} = i] \log q_i \\
&= \exp \sum_{i=1}^{K} \log q_i \sum_{n=1}^{N} 1[x^{(n)} = i] \\
&= \exp \sum_{i=1}^{K} \log q_i N_i
\end{aligned}
$$

Multinomial Distribution

$$p(\mathcal{D};\mathbf{q}) = \prod_{n=1}^{N} q_1^{1[x^{(n)}=1]} \cdots q_K^{1[x^{(n)}=K]}$$

$$= \exp \log \prod_{n=1}^{N} \prod_{i=1}^{K} q_i^{1[x^{(n)}=i]}$$

$$= \exp \sum_{n=1}^{N} \sum_{i=1}^{K} 1[x^{(n)} = i] \log q_i$$

$$= \exp \sum_{i=1}^{K} \log q_i \sum_{n=1}^{N} 1[x^{(n)} = i]$$

$$= \exp \sum_{i=1}^{K} \log q_i N_i$$

$$T(\mathcal{D}) = N_i = \sum_{n=1}^{N} 1[x^{(n)} = i]$$

$$\eta_i = \log q_i$$

$$A(\eta_i) = 0$$

Which means the distribution mean and variance are zero — which is not correct.

This is due to the constraint of $\sum_{i=1}^{K} q_i$

We can fix it using $q_K = 1 - \sum_{i=1}^{K-1} q_i$

Multinomial Distribution

$$\begin{aligned}
p(\mathcal{D};\mathbf{q}) &= \exp\sum_i^K N_i \log q_i \\
&= \exp(\sum_i^{K-1} N_i \log q_i + N_K \log q_K) \\
&= \exp(\sum_i^{K-1} N_i \log q_i + (N - \sum_i^{K-1} N_i) \log(1 - \sum_i^{K-1} q_i)) \\
&= \exp\left(\sum_i^{K-1} N_i \log \frac{q_i}{1 - \sum_i^{K-1} q_i} + N \log(1 - \sum_i^{K-1} q_i)\right)
\end{aligned}$$

Multinomial Distribution

$$
\begin{aligned}
p(\mathcal{D}; \mathbf{q}) &= \exp \sum_{i}^{K} N_i \log q_i \\
&= \exp(\sum_{i}^{K-1} N_i \log q_i + N_K \log q_K) \\
&= \exp(\sum_{i}^{K-1} N_i \log q_i + (N - \sum_{i}^{K-1} N_i) \log(1 - \sum_{i}^{K-1} q_i)) \\
&= \exp\left(\sum_{i}^{K-1} N_i \log \frac{q_i}{1 - \sum_{i}^{K-1} q_i} + N \log(1 - \sum_{i}^{K-1} q_i)\right)
\end{aligned}
$$

$$T(\mathcal{D}) = N_i = \sum_{n=1}^{N} 1[x^{(n)} = i]$$

$$\eta_i = \log \frac{q_i}{1 - \sum_{i=1}^{K-1} q_i} = \log \frac{q_i}{q_K}$$

$$h(\mathcal{D}) = 1$$

$$
\begin{aligned}
A(\eta) &= -N \log\left(1 - \sum_{i=1}^{K-1} q_i\right) \\
&= N \log\left(\frac{1}{q_K}\right) \\
&= N \log\left(1 + \sum_{i=1}^{K-1} e^{\eta_i}\right)
\end{aligned}
$$

MLE of the Distribution

Log-likelihood: $$\ell(\mathbf{q}; \mathcal{D}) = \sum_i \log q_i N_i$$

To include the constraint, we write the Lagrangian:

$$L(\mathbf{q}, \lambda) = \sum_i \log q_i N_i + \lambda(1 - \sum_i q_i)$$

$$\frac{dL}{dq_i} =$$

Sufficient Statistics and MLE for Univariate Normal

We assume the data is N i.i.d. samples $\{x^{(i)}\} \in \mathbb{R}$ from the Gaussian distribution, i.e.

$$\begin{aligned} x^{(i)} \sim p(x|\theta) &= \mathcal{N}(x|\mu, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\} \end{aligned}$$

The Gaussian distribution is a member of the exponential family, so we can put it into a natural form:

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

Sufficient Statistics and MLE for Univariate Normal

$$
\begin{aligned}
p(x|\theta) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\} \\
&= \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}\} \\
&= \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}\mu^2\}\exp\{\left[\tfrac{\mu}{\sigma^2}\ \tfrac{1}{\sigma^2}\right]\begin{bmatrix} x \\ -\frac{x^2}{2} \end{bmatrix}\}
\end{aligned}
$$

Sufficient Statistics for Univariate Normal

We can rewrite the likelihood in terms of η:

$$p(x|\eta) = \sqrt{\frac{\eta_2}{2\pi}} \cdot \exp\{-\frac{\eta_1^2}{2\eta_2}\} \cdot \exp\{\eta^T T(x)\}$$

- $h(x) = (2\pi)^{-\frac{1}{2}}$
- $A(\eta) = -\frac{1}{2}\log(\eta_2) + \frac{\eta_1^2}{2\eta_2}$

MLE for Univariate Normal

- We can use the general result from the exponential family, or take derivatives in this form to find the MLE for those natural statistics.
- However, we often prefer to work with $\theta = [\mu, \sigma^2]$
- Log-likelihood:

$$\ell(\theta; \mathcal{D}) = \log p(\mathcal{D}|\theta) = \log \prod_n p(x^{(n)}|\theta)$$

MLE for Univariate Normal

$$\begin{aligned}
\ell(\theta; \mathcal{D}) = \log p(\mathcal{D}|\theta) &= \log \prod_n p(x^{(n)}|\theta) \\
&= \log \prod_n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x^{(n)} - \mu)^2\} \\
&= \sum_n -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x^{(n)} - \mu)^2 \\
&= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_n \frac{(x^{(n)} - \mu)^2}{\sigma^2}
\end{aligned}$$

Derivatives of $\theta = [\mu, \sigma^2]$:

$$\frac{\partial \ell}{\partial \mu} = \qquad\qquad \frac{\partial \ell}{\partial \sigma^2} =$$

MLE for Univariate Normal

Derivatives of $\theta = [\mu, \sigma^2]$:

$$\frac{\partial \ell}{\partial \mu} = 0 + \frac{1}{\sigma^2} \sum_n x^{(n)} - \mu = 0 \Rightarrow \mu_{\text{MLE}} = \frac{1}{N} \sum_n x^{(n)}$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_n ((x^{(n)})^2 - \mu_{\text{MLE}}^2) = 0 \Rightarrow \sigma_{\text{MLE}}^2 = \frac{1}{N} \sum_n ((x^{(n)})^2 - \mu_{\text{MLE}}^2)$$