ML & Advanced Analytics For Biomedicine

Ishanu Chattopadhyay

Department of Medicine

University of Chicago

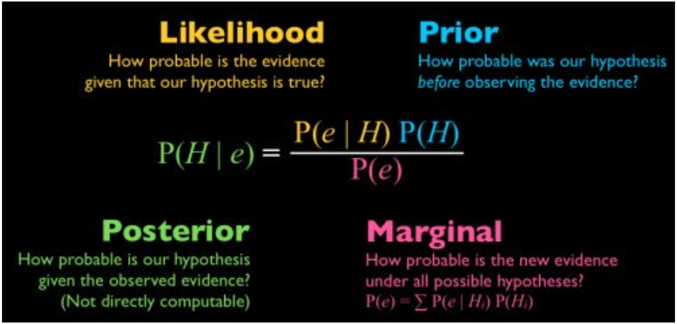
CCTS 40500 / CCTS 20500 / BIOS 29208

Winter 2019

Today

- **Bayesian Statistics (Contd.)**
- **MLE vs MAP Estimates**
- **Naive Bayes Classifier**
- **Nearest Neighbor Classifier**

Bayes Theorem



A Bayesian is one who, vaguely expecting a horse and catching a glimpse of a donkey, strongly concludes he has seen a mule.

– Senn, 1997–

Maximum Likelihood vs Maximum a Posteriori Probability Estimation

$$
\begin{aligned}
\theta_{MLE} &= \arg\max_{\theta} Pr(X|\theta) &(1)\\
\theta_{MAP} &= \arg\max_{\theta} Pr(\theta|X) \\
&= \arg\max_{\theta} \Big(\log P(X|\theta) + \log Pr(\theta) \Big) &(2)
\end{aligned}
$$

◦ **Intuitively, a sufficient statistic for a parameter is a statistic that captures all the information about a given parameter contained in the sample.**

◦ **Sufficiency Principle: If $T(X)$ is a sufficient statistic for θ, then any inference about θ should depend on the sample X only through the value of $T(X)$.**

◦ **That is, if x and y are two sample points such that $T(x) = T(y)$, then the inference about θ should be the same whether $X = x$ or $X = y$.**

◦ **Definition: A statistic $T(x)$ is a sufficient statistic for θ if the conditional distribution of the sample X given $T(x)$ does not depend on θ.**

Sufficient Statistic

Let $X_1, X_2, \cdots X_n$ denote a random sample of size n from a distribution that has a pdf $f(x, \theta), \theta \in \Theta$. Let $Y = g(X_1, \cdots, X_n)$ be a statistic whose pdf or pmf is $f_Y(g(x_1, \cdots, x_n); \theta)$.
Then Y is a sufficient statistic for θ if and only if

$$\frac{\prod_i f(x_i; \theta)}{f_y(x_1, \cdots, x_n; \theta)} = H(x_1, \cdots, x_n) \tag{3}$$

Let $X_1, X_2, \cdots X_n$ be independently and identically distributed $\mathcal{N}(\mu\sigma^2)$ where the variance is known. The sample mean

$$T(X) = \overline{X} = \frac{1}{n}\sum_i X_i \tag{4}$$

is a sufficient statistic for μ.

Let $f(x|\theta)$ denote the joint pdf or pmf of a sample X. A statistic $T(X)$ is a sufficient statistic for θ if and only if there exists functions $g(t|\theta)$ and $h(x)$ such that, for all sample points x and all parameter points θ, we have:

$$\boxed{f(x|\theta) = g(T(x)|\theta)h(x)} \tag{5}$$

Theorem: The posterior distribution depends only on sufficient statistics.

$$f(\theta|X) = f(\theta|T(X)) \tag{6}$$

Proof:

Let $X_1, X_2, \cdots X_n$ denote the vector of observations having joint density $f(X;\theta)$ with unknown parameter θ.
Let $\hat{\theta}$ be an estimator for θ. Then, the MSE of th estimator:

$$MSE(\hat{\theta}) = \mathrm{E}(\theta - \hat{\theta})^2 \tag{7}$$

$$= \int (\theta - \hat{\theta})^2 Pr(X|\theta) Pr(\theta) dx d\theta \tag{8}$$

$$\hat{\theta} = \mathrm{E}(\theta|X) \tag{9}$$

Theorem:
The Bayes estimator minimizes the mean square error

Thus:
Posterior mean is the best estimator under quadratic loss
Recall: Posterior depends only on sufficient statistics.
Hence:

$$\hat{\theta} = \mathrm{E}(\theta|T(X)) \tag{10}$$

Credible Intervals