Feynman Liang
STAT 210A
Lecture #3
9/1/2016

1 Review

$$p_\theta(x) = e^{\eta(\theta)'T(x) - A(\eta(\theta))} h(x) \tag{1.1}$$
$$\theta = \eta \quad \text{“canonical form”} \tag{1.2}$$
$$A(\eta) = \log \int_X e^{\eta' T(x) - A(\eta)} h(x) d\mu(x) \quad \text{“CGF”} \tag{1.3}$$
$$\Xi = \{\eta : A(\eta) < \infty\} \quad \text{“natural param. space”} \tag{1.4}$$

If $\eta \in \Xi^\circ$, can exchange $\frac{\partial}{\partial \eta} \int_X p_\eta(x) d\mu(x) = \int_X \frac{\partial}{\partial \eta} p_\eta(x) d\mu(x)$.

2 Sufficiency

Definition 2.1. An *estimator* $\delta(x)$ is a statistic meant ot estimate $g(\theta)$.

Definition 2.2. The *Risk* $R(\theta, \delta) = \mathbb{E}_\theta[L(\theta, \delta(X))]$.

For squared error loss ($L(\theta, \delta(X)) = (\delta(X) - g(\theta))^2$), the risk is the mean squared error (MSE).

Say $X_1, \cdots, X_n \overset{\text{iid}}{\sim} \text{Bern}(\theta)$. $T(x) = \sum_{i=1}^n x_i \sim \text{Binom}(n, \theta)$. $\{X_i\}_{i=1}^n$ has more information than $T(x)$; is there any way to *justify throwing out this additional information* and summarizing the data with $T(X)$?

Definition 2.3. Let $\mathcal{P} = \{p_\theta : \theta \in \Theta\}$ be a model for data X. We say $T(X)$ is *sufficient* for $\mathcal{P}$ if $p_\theta(X \mid T(X))$ does not depend on θ.

Equivalently, T is sufficient iff $\theta \to T \to X$ is a Markov chain.

Example 2.4. $t \in \{0, \ldots, n\}$. Then $X \mid T = t$ is uniform on sequences $\{x \in \{0,1\}^n : \sum_i x_i = t\}$.

$$P_\theta(X = x) = \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum_i x_i} (1-\theta)^{n - \sum_i x_i} \tag{2.1}$$
$$P_\theta(X = x \mid T = t) = 1_{\sum_i x_i = t} \frac{\cancel{\theta^t (1-\theta)^{n-t}}}{\sum_{x : \sum_x i = t} \cancel{\theta^t (1-\theta)^{n-t}}} = \underbrace{\frac{1_{\sum_i x_i = t}}{\binom{n}{t}}}_{\text{indep of } \theta\text{, so } T \text{ sufficient}} \tag{2.2}$$

Example 2.5. Suppose $\delta(X)$ estimator of θ which is not just a function of T. We could generate $\tilde{X} \mid T(X)$.

Then, $\delta(X)$ and $\delta(\tilde{X})$ have the same distribution and hence have the same risk.

Therefore, $\delta(X)$ is no better than $\delta(\tilde{X})$, and $\tilde{X} \mid T(X) \to \delta(\tilde{X})$ is an estimator which is only a function of $T(X)$!

Sufficiency principle: If $T(X)$ is sufficient, then any statistical procedure should depend only on $T(X)$.

Theorem 2.6 (Factorization Theorem). *Let $\mathcal{P} = \{p_\theta : \theta \in\}$ be a family of densities wrt μ. T is sufficient for $\mathcal{P} \iff \exists$ functions $g_\theta, h \geq 0$ such that*

$$p_\theta(x) = g_\theta(T(x))h(x) \quad a.e. x \tag{2.3}$$

Proof. Rigorous proof in Keener 6.4.

"$\Leftarrow$" $p_\theta(x \mid T = t) = 1_{T(x)=t} \frac{\cancel{g_\theta(t)}h(x)}{\int_{T(x)=t} \cancel{g_\theta(t)}h(x)d\mu(x)}$ is independent of θ.

"$\Rightarrow$" Take $g_\theta(t) = \int_{T(z)=t} p_\theta(z)d\mu(z) = p_\theta(T = t)$ Then

$$h(x) = \frac{p_{\theta_0}(x)}{\int_{T(z)=T(x)} p_{\theta_0} p_{\theta_0}(z)d\mu(z)} = \underbrace{p_{\theta_0}(X = x \mid T = T(x))}_{\text{function of } X\text{, independent of } \theta} \tag{2.4}$$

$\square$

Example 2.7 (Exponential Family). $p_\theta(x) = \underbrace{e^{\eta(\theta)'T(x) - B(\theta)}}_{g_\theta(T(x))} h(x)$

Definition 2.8. $X_{(i)}$ denotes *order statistics*, indexes X_i by ordering i.e. $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ even if $X_1 > X_2$ etc

Example 2.9 (Non-parametric). $X_1, \cdots, X_n \overset{\text{iid}}{\sim} p_\theta^{(1)}$ on $\mathbb{R}$

For *any* model $\mathcal{P} = \{p_\theta^{(1)} : \theta \in \Theta\}$

$$p_\theta(x) = \prod_{i=1}^{n} p_\theta^{(1)}(x_i) = \prod_{i=1}^{n} p_\theta^{(1)}(X_{(i)}) \tag{2.5}$$

The ordering information has been thrown away; irrelevant for iid samples.

Example 2.10. $X_1, \ldots, X_n \overset{\text{iid}}{\sim} U[\theta, \theta + 1]$.

Then $p_\theta(x) = \prod_{i=1}^n 1_{\theta \leq x_i \leq \theta+1} = 1_{\theta \leq x_{(i)} \leq x_{(n)} \leq \theta+1}$

3 Minimal sufficiency

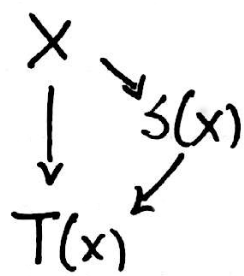
Consider coin flips $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(\theta)$. The following are all sufficient statistics:

(a) $T(X) = \sum_i X_i$

(b) $(X_{(1)}, \dots, X_{(n)})$ (in fact equivalent since $\operatorname{supp} X_i = \{0, 1\}$)

(c) $(X_1, \dots, X_n) = X$

(d) $\left(\sum_{i=1}^{n/2} X_i, \sum_{i=n/2+1}^{n} X_i\right) = S(X)$



Can $T(X)$ be compressed further?

Proposition 3.1. *If T is sufficient, $T = f(S) \implies S$ is sufficient.*

Proof. $p_\theta(X) = g_\theta(T(X))h(X) = (g_\theta \circ f)(S(X))h(X)$ so taking $g_\theta \circ f = \tilde{g}_\theta$ in the factorization theorem shows S is sufficient. $\square$

Definition 3.2. T is *minimal sufficient* if:

(i) T is sufficient

(ii) $\forall S$ sufficient, $\exists f$ such that $T \overset{\text{a.s.}}{=} f(S)$