Chapter 3: Facts about Common Statistical Models *

3 Facts about common statistical models

3.1 Bayes Models

- Probability Model on Data We have distributions $\mathcal{P} \equiv \{P_\theta : \theta \in \Theta\}$ for X on $(\mathcal{X}, \mathcal{B})$, where $\mathcal{P} \ll \mu$ ($\sigma-$ finite measure) and R -N derivatives

$$\frac{dP_\theta}{d\mu}(x) = f_\theta(x)$$

- Prior on Parameter We now add an assumption of a distribution G on $(\Theta, \mathcal{C})$ with $G \ll \nu$(σ -finite measure) and R-N derivative

$$\frac{dG}{d\nu}(\theta) = g(\theta)$$

- Joint Distribution for (X, θ) : Here we consider $f_\theta(x)$ as a function of both x and θ (i.e., measurable in (x, θ)). If $f_\theta(x)$ is $\mathcal{B} \times \mathcal{C}$ -measurable, then there exists a joint probability distribution for (X, θ) on $(\mathcal{X} \times \Theta, \mathcal{B} \times \mathcal{C})$ defined, for $A \in \mathcal{B} \times \mathcal{C}$, by

$$\pi^{X,\theta}(A) \equiv P((X, \theta) \in A) = \int_A f_\theta(x) d(\mu \times G)(x, \theta) = \int f_\theta(x) g(\theta) d(\mu \times \nu)(x, \theta)$$

where
$$\frac{d\pi^{X,\theta}}{d(\mu \times G)} \equiv f_\theta(x), \quad \frac{d\pi^{X,\theta}}{d(\mu \times \nu)} \equiv f_\theta(x) g(\theta)$$

- Marginal Distributions
  - for X $(B \in \mathcal{B})$

$$\pi^X(B) \equiv P(X \in B) = \pi^{X,\theta}(B \times \Theta) = \int_{B \times \Theta} f_\theta(x) d(\mu \times G)(x, \theta) \overset{Fubini}{=} \int_B \left[\int_\Theta f_\theta(x) dG(\theta) \right] d\mu(x)$$

$$= \int_B \left[\int_\Theta f_\theta(x) g(\theta) d\nu \right] d\mu(x)$$

$$0 \leq \frac{d\pi^X(x)}{d\mu} = \int_\Theta f_\theta(x) dG(\theta) = \int_\Theta f_\theta(x) g(\theta)$$

  - for θ $(C \in \mathcal{C})$

$$\pi^\theta(C) \equiv P(\theta \in C) = \pi^{X,\theta}(\mathcal{X} \times C) = \int_{\mathcal{X} \times C} f_\theta(x) d(\mu \times G)(x, \theta) = \int_C \left[\int_{\mathcal{X}} f_\theta(x) d\mu(x) \right] dG(\theta) = G(C)$$

  Marginal distribution of θ is prior distribution G.

- Conditional distributions
  - for $X \mid \theta$

$$\pi^{X|\theta}(B \mid \theta) \equiv P_{X|\theta}(X \in B \mid \theta) = \int_B f_\theta(x) d\mu(x) = P_\theta(B), \quad B \in \mathcal{B}$$

$$\frac{d\pi^{X|\theta}(x)}{d\mu} = \frac{dP_\theta(x)}{d\mu} = f_\theta$$

*STA643: Advanced Theory of Statistical Inference. Instructed by Dr. Daniel Nordman. Arranged by Zhiling Gu

- for $\theta \mid X$

$$\pi^{\theta|X}(C \mid x) \equiv P_{\theta|X}(\theta \in C \mid X = x) = \int_C \left[\frac{f_\theta(x)}{\int_\Theta f_\theta(x) dG(\theta)}\right] dG(\theta) = \int_C \frac{f_\theta(x) g(\theta)}{\int_\Theta f_\theta(x) g(\theta) d\nu(\theta)} d\nu(\theta)$$

$$\frac{d\pi^{\theta|X}(\theta)}{dG} = \frac{f_\theta(x) g(\theta)}{\int_\Theta f_\theta(x) dG(\theta)}, \quad \frac{d\pi^{\theta|X}(\theta)}{d\nu} = \frac{f_\theta(x) g(\theta)}{\int_\Theta f_\theta(x) g(\theta) d\nu(\theta)}, G \ll \nu$$

Note priors does not necessarily have a density, i.e. G is not necessarily dominated by some ν. But you can always write the density of posterior with respect to G.

3.2 Exponential Family of Distributions

- Exponential family: Definition 16 : $\mathcal{P} \equiv \{P_\theta : \theta \in \Theta\} \ll \mu(\sigma$ -finite measure) is an exponential family if, for some $h(x) \geq 0$, it holds that

$$f_\theta(x) \equiv \frac{dP_\theta}{d\mu}(x) = \exp\left(\alpha(\theta) + \sum_{i=1}^k \eta_i(\theta) T_i(x)\right) h(x), \quad x \in \mathcal{X}$$

for any $\theta \in \Theta$

- Identifiable: Definition 17 : A family of distributions, $\mathcal{P} \equiv \{P_\theta : \theta \in \Theta\}$ is identifiable if $\theta_1 \neq \theta_2$ implies $P_{\theta_1} \neq P_{\theta_2}$
- Natural parameter space: Let $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k) \in \mathbb{R}^k$ and let

$$\Gamma \equiv \left\{\boldsymbol{\eta} \in \mathbb{R}^k : \int_{\mathcal{X}} h(x) \exp\left(\sum_{i=1}^k \eta_i T_i(x)\right) d\mu(x) < \infty\right\}$$

note: this kernel with linear combinations of $T_i(X)$ using real numbers $\eta_i, i = 1, \ldots, k$.

Define a new distributional family

$$\mathcal{P}^* \equiv \left\{P_\eta \text{ has R-N derivative as } f_\eta(x) \equiv \frac{dP_\eta}{d\mu}(x) = K(\boldsymbol{\eta}) h(x) \exp\left(\sum_{i=1}^k \eta_i T_i(x)\right) : \boldsymbol{\eta} \in \Gamma\right\}$$

- $\mathcal{P} \subset \mathcal{P}^*$
- Γ is called the natural parameter space for $\mathcal{P}^*$ and Γ is a convex subset of $\mathbb{R}^k$
- If Γ lies in a subspace of dimension less than k, then $f_\eta(x)$ (and $f_\theta(x)$) can be rewritten in a form involving fewer than k statistics $T_i(x)$. (We'll assume Γ to be fully k -dimensional.)
- $\mathcal{P}$ may be a proper subset of $\mathcal{P}^*$ or

$$\Gamma_\theta \equiv \left\{(\eta_1(\theta), \eta_2(\theta), \ldots, \eta_k(\theta)) \in \mathbb{R}^k : \theta \in \Theta\right\}$$

  can be a proper subset of Γ.

  * For example, for $f_\theta \propto \exp(\theta, -\theta^2)$,

  $$\Gamma_\theta = \{(\theta, -\theta^2) : \theta \in \mathbb{R}\} \subset \Gamma \equiv \{(\eta_1, \eta_2) : \eta \in \mathcal{T}, \eta_2 < 0\}$$

  * The most useful results/theorems about the $\boldsymbol{\eta}$ -parameterization are the ones where Γ contains an open set, i.e. Γ is rich/big enough.
  * If we want to translate results about the $\boldsymbol{\eta}$ -parameterization to θ, then we want Γ_θ to contain an open set in $\mathbb{R}^k$.
  * To use the θ -parameterization, we must want $\boldsymbol{\eta}(\cdot)$ to be 1 -to-1 on Θ.
- Claim 19: The support of P_θ is

$$\{x \in \mathcal{X} : f_\theta(x) > 0\} = \{x \in \mathcal{X} : h(x) > 0\}$$

  The distributions in $\mathcal{P} \equiv \{P_\theta : \theta \in \Theta\}$ are mutually absolutely continuous.
- Claim 20: The statistic $T = (T_1, \ldots, T_k)$ is sufficient for the exponential family $\mathcal{P}$.
- Claim 21 : $T = (T_1, \ldots, T_k)$ has induced distributions $\left\{P_\theta^T : \theta \in \Theta\right\}$, where

$$P_\theta^T(B) = P_\theta(T \in B), \quad B \in \mathcal{B}\left(\mathbb{R}^k\right)$$

  which is also an exponential family.

- Claim 22: If Γ_θ contains an *open rectangle* in $\mathbb{R}^k$, then $T = (T_1, \ldots, T_k)$ is complete for the exponential family $\mathcal{P}$.
- Claim 23: If Γ_θ contains an *open rectangle* in $\mathbb{R}^k$ (or under a much weaker assumption by Lehmann (1983)), then $T = (T_1, \ldots, T_k)$ is minimal sufficient for $\mathcal{P}$.

  Lehmann's Geometric Condition: If there exists $k+1$ points $v_0, \ldots, v_k \in \Gamma_\theta \subset \mathbb{R}^k$ *convex hull*
  $$\left\{ \sum_{i=0}^{k} p_i v_i, v_i \in \mathbb{R}^k, p_i \geq 0, \sum_{i=0}^{k} p_i = 1 \right\}$$
  contains an open set in $\mathbb{R}^k$ then T is minimally sufficient.
- Claim 24: If $g : \mathcal{X} \to \mathbb{R}$ is a measurable real-valued function with $E_\eta |g(X)| < \infty$ then
  $$\mathrm{E}_{\boldsymbol{\eta}} g(X) = \int_{\mathcal{X}} g(x) f_{\boldsymbol{\eta}}(x) d\mu(x)$$
  is continuous on Γ and has continuous partial derivatives of all orders on the interior of Γ. Also,
  $$\frac{\partial^{\alpha_1+\alpha_2+\cdots+\alpha_k}}{\partial \eta_1^{\alpha_1} \partial \eta_2^{\alpha_2} \cdots \partial \eta_k^{\alpha_k}} \mathrm{E}_{\boldsymbol{\eta}} g(X) = \int_{\mathcal{X}} g(x) \frac{\partial^{\alpha_1+\alpha_2+\cdots+\alpha_k}}{\partial \eta_1^{\alpha_1} \partial \eta_2^{\alpha_2} \cdots \partial \eta_k^{\alpha_k}} f_{\boldsymbol{\eta}}(x) d\mu(x)$$
  holds for $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{Z}_+ \equiv \{0, 1, 2, \ldots\}$

  It is okay to swap partials and expectations.
- Claim 25: Recall the exponential form $f_\eta(x) = K(\boldsymbol{\eta}) \exp\left(\sum_{i=1}^k \eta_i T_i(x)\right) h(x)$ of densities in $\mathcal{P}^*$ where $K(\eta)$ is normalizing constant. If $\boldsymbol{\eta}_0, \boldsymbol{\eta}_0 + \boldsymbol{u} \in \Gamma$ for $\boldsymbol{u} = (u_1, \ldots, u_k)$, then the moment generating function of statistic $T(X)$ is
  $$\mathrm{E}_{\boldsymbol{\eta}_0} \exp\left[u_1 T_1(X) + \cdots + u_k T_k(X)\right] = \frac{K(\boldsymbol{\eta}_0)}{K(\boldsymbol{\eta}_0 + \boldsymbol{u})}$$
  and the moments can be calculated by taking derivatives wrt u evaluated at $u = 0$.
  $$\mathrm{E}_{\boldsymbol{\eta}_0} \left[T_1^{\alpha_1}(X) T_2^{\alpha_2}(X) \cdots T_k^{\alpha_k}(X)\right] = K(\boldsymbol{\eta}_0) \frac{\partial^{\alpha_1+\alpha_2+\cdots+\alpha_k}}{\partial \eta_1^{\alpha_1} \partial \eta_2^{\alpha_2} \cdots \partial \eta_k^{\alpha_k}} \frac{1}{K(\boldsymbol{\eta})} \Bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0}$$