Keener, Theoretical Statistics
Chapter 5: Curved Exponential Families

Junhyeok Kil

25 Sep 2025

Overview

- Curved exponential families can emerge when the parameters of a standard exponential family are subject to certain constraints.
- A key consequence is that for these families, the minimal sufficient statistic might not be complete.
- This has significant implications, as Uniformly Minimum-Variance Unbiased (UMVU) estimation may not be possible.

Constrained Families: The Setup

- Let $\mathcal{P} = \{P_\eta : \eta \in \Xi\}$ be a full rank s-parameter canonical exponential family.
- This family has a complete sufficient statistic T.

- Consider a submodel $\mathcal{P}_0$ parameterized by $\theta \in \Omega$.
- The canonical parameter η is now a function of θ, denoted as $\tilde{\eta}(\theta)$.
- The submodel is thus defined as:

$$\mathcal{P}_0 = \{P_{\tilde{\eta}(\theta)} : \theta \in \Omega\}$$

- Often, $\tilde{\eta} : \Omega \to \Xi$ is a one-to-one and onto mapping. In such cases, $\mathcal{P}_0 = \mathcal{P}$, and we are just reparameterizing.

- Curved families arise when $\mathcal{P}_0$ is a strict subset of $\mathcal{P}$, which typically happens when the dimension of Ω is less than the dimension of Ξ (i.e., $r < s$, for $\Omega \subset \mathbb{R}^r$).

Two Possibilities for Constrained Models

When the submodel $\mathcal{P}_0$ is a strict subset of the full exponential family $\mathcal{P}$, two main scenarios can occur based on the nature of the constraint on the canonical parameter η.

1. The points η in the range of $\tilde{\eta}(\theta)$ satisfy a **nontrivial linear constraint**.

2. The points η in the range of $\tilde{\eta}(\theta)$ **do not satisfy a linear constraint**.

Case 1: Linear Constraint

Scenario

The set of canonical parameters $\{\tilde{\eta}(\theta) : \theta \in \Omega\}$ satisfies a nontrivial linear constraint.

- In this case, the submodel $\mathcal{P}_0$ is itself a canonical exponential family, but with fewer parameters ($q < s$).
- The original statistic T is still sufficient for the submodel $\mathcal{P}_0$.
- However, T is **no longer minimal sufficient** for this constrained family.

Case 2: No Linear Constraint (Curved Family)

Scenario

The set of canonical parameters $\{\tilde{\eta}(\theta) : \theta \in \Omega\}$ does not satisfy any linear constraint.

- This is the defining characteristic of a **curved exponential family**. The name comes from the fact that the parameter space of η forms a "curve" within the larger space Ξ.
- In this situation, the original statistic T **remains minimal sufficient**.
- However, a critical property is lost: T **may not be complete**.

Example 5.1: The Normal Distribution Family

- Consider a sample $X_1, \ldots, X_n$ from a Normal distribution $N(\mu, \sigma^2)$.

- The joint distributions form a two-parameter exponential family.

- The canonical parameter η is:

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

- The complete sufficient statistic T is:

$$T = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2 \right)$$

Case 1: Linear Constraint ($\mu = \sigma^2$)

- Let's impose the constraint that the mean and variance are equal.
- Let $\theta = \mu = \sigma^2$, for $\theta > 0$. The sample is from $N(\theta, \theta)$.
- The canonical parameter becomes a function of θ:

$$\tilde{\eta}(\theta) = \left(\frac{\theta}{\theta}, -\frac{1}{2\theta}\right) = \left(1, -\frac{1}{2\theta}\right)$$

- The range of $\tilde{\eta}$ satisfies the linear constraint $\eta_1 = 1$. The parameter space is restricted to a vertical line.
- **Conclusion**: The subfamily is a one-parameter exponential family, and $\sum X_i^2$ becomes the canonical complete sufficient statistic.

Case 2: Non-Linear Constraint ($\sigma = |\mu|$)

- Now, impose the constraint $\sigma = |\mu|$.
- Let $\theta = \mu$, so the sample is from $N(\theta, \theta^2)$, for $\theta \in \mathbb{R}$.
- The canonical parameter is:

$$\tilde{\eta}(\theta) = \left(\frac{\theta}{\theta^2}, -\frac{1}{2\theta^2}\right) = \left(\frac{1}{\theta}, -\frac{1}{2\theta^2}\right)$$

- The range space $\tilde{\eta}(\Omega)$ is the parabola $\eta_2 = -\frac{1}{2}\eta_1^2$. This is a non-linear constraint.
- **Conclusion**: This is a **curved exponential family**, and the original statistic $T = (\sum X_i, \sum X_i^2)$ is minimal sufficient.

Proving Incompleteness for the Curved Case

For the curved family $N(\theta, \theta^2)$, the statistic $T = (T_1, T_2)$ is not complete.

- To prove that we need to find a non-zero function $g(T)$ such that $E_\theta[g(T)] = 0$ for all θ.
- Calculate the expected values of components of T:

$$\begin{aligned} E_\theta[T_1^2] &= (E_\theta[T_1])^2 + \mathrm{Var}_\theta(T_1) \\ &= (n\theta)^2 + n\theta^2 = n^2\theta^2 + n\theta^2 \\ E_\theta[T_2] &= nE_\theta[X_i^2] = n\left((E_\theta[X_i])^2 + \mathrm{Var}_\theta(X_i)\right) \\ &= n(\theta^2 + \theta^2) = 2n\theta^2 \end{aligned}$$

Proving Incompleteness for the Curved Case

- Consider the function $g(T) = 2T_1^2 - (n+1)T_2$.

$$\begin{aligned} E_\theta[g(T)] &= 2E_\theta[T_1^2] - (n+1)E_\theta[T_2] \\ &= 2(n^2\theta^2 + n\theta^2) - (n+1)(2n\theta^2) \\ &= (2n^2+2n)\theta^2 - (2n^2+2n)\theta^2 = 0 \end{aligned}$$

- Since $g(T)$ is not the zero function but its expectation is zero for all θ, T is **not complete**.

Example 5.2: Two Independent Samples

- Consider two independent random samples:
  - $X_1, \ldots, X_m \sim N(\mu_x, \sigma_x^2)$
  - $Y_1, \ldots, Y_n \sim N(\mu_y, \sigma_y^2)$
- The joint distributions form a **four-parameter exponential family**.
- The canonical parameter η is:

$$\eta = \left(\frac{\mu_x}{\sigma_x^2}, \frac{\mu_y}{\sigma_y^2}, -\frac{1}{2\sigma_x^2}, -\frac{1}{2\sigma_y^2} \right)$$

- A canonical sufficient statistic T is:

$$T = \left(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j, \sum_{i=1}^{m} X_i^2, \sum_{j=1}^{n} Y_j^2 \right)$$

- An equivalent statistic is $(\bar{X}, \bar{Y}, S_x^2, S_y^2)$.

Case 1: Equal Variances ($\sigma_x^2 = \sigma_y^2$)

Constraint

The variances of the two samples are equal: $\sigma_x^2 = \sigma_y^2 = \sigma^2$.

- This imposes the linear constraint $\eta_3 = \eta_4$ on the canonical parameters.
- The joint distributions now form a **three-parameter exponential family** and the complete sufficient statistic is $(T_1, T_2, T_3 + T_4)$.
- An equivalent sufficient statistic is $(\bar{X}, \bar{Y}, S_p^2)$, where S_p^2 is the **pooled sample variance**:

$$S_p^2 = \frac{\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{n + m - 2} = \frac{(m-1)S_x^2 + (n-1)S_y^2}{n + m - 2}$$

- The distribution of the scaled pooled variance is Chi-square (and we can find UMVUEs):

$$\frac{(n + m - 2)S_p^2}{\sigma^2} \sim \chi^2_{n+m-2}$$

Case 2: Equal Means ($\mu_x = \mu_y$)

Constraint

The means of the two samples are equal: $\mu_x = \mu_y$.

- In this case, the joint distributions form a **curved exponential family**.
- The original four-component statistic T (or its equivalent, $(\bar{X}, \bar{Y}, S_x^2, S_y^2)$) is **minimal sufficient**.
- However, this minimal sufficient statistic is **not complete**.

Case 2: Equal Means ($\mu_x = \mu_y$)

Proof of Incompleteness

Consider the function $g(T) = \bar{X} - \bar{Y}$. This function is not identically zero.
The expected value is:

$$E[\bar{X} - \bar{Y}] = E[\bar{X}] - E[\bar{Y}] = \mu_x - \mu_y$$

Under the constraint $\mu_x = \mu_y$, this expectation is always zero:

$$E[\bar{X} - \bar{Y}] = 0$$

Since we found a non-zero function of the statistic whose expectation is zero for all parameters in the subfamily, the statistic is not complete.

What are Sequential Experiments?

Definition

An experiment's protocol is **sequential** if the data are observed as they are collected, and the information from these observations influences how the experiment is performed.

- Examples of influence:
  - Deciding whether to terminate a study or continue collecting more data.

Why use sequential over classical experiment?

1. **Efficiency**: They may reduce decision-theoretic risk (e.g., lower running cost).
2. **Necessity**: Certain experimental objectives can only be met with a sequential design.

Example 5.3: Estimating Population Size

The Capture-Recapture Experiment

- **Goal**: Estimate M, the number of fish in a lake.
- **Phase 1**: Capture k fish, tag them, and release them.
- **Phase 2**: Sample fish randomly from the lake. The probability a sampled fish is tagged is $\theta = k/M$.
- **Inference Goal**: Estimating M is equivalent to estimating $1/\theta$.
- The data from Phase 2 can be viewed as i.i.d. Bernoulli variables $X_1, X_2, \ldots$ where $X_i = 1$ if the fish is tagged.

The Problem with a Fixed Sample Size

- If we use a fixed sample size $N = n$, the sufficient statistic is $T = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, \theta)$.
- The statistic T/n is an unbiased (and UMVU) estimator for θ.
- **But there is no unbiased estimator of $1/\theta$ in this model**
- **Why?** The expectation of any estimator, $E_\theta[\delta(T)]$, is bounded. However, $1/\theta$ is unbounded as $\theta \to 0$.
- If θ is very small, we are very likely to observe $T = 0$, which gives little information about the true population size.

Sequential Solution: Inverse Binomial Sampling

- To avoid the problem of observing zero successes, we change the sampling rule.
- **Method**: Continue sampling until we have observed a fixed number, m, of successes (tagged fish).
- Now, the total sample size N is a random variable.
- This is a sequential experiment because the decision to stop sampling depends directly on the observed data.

Inverse Binomial Sampling

- Let Y_i be the number of failures before the i-th success.
- The variables $Y_1, \ldots, Y_m$ are i.i.d. with a Geometric distribution:
$$P_\theta(Y_i = y) = (1-\theta)^y \theta$$
- The Geometric distribution is a one-parameter exponential family.
- The joint distribution is
$$P_\theta(y_1, \ldots, y_m) = (1-\theta)^{\sum y_i} \theta^m,$$
which depends on the data only through $T = \sum Y_i$.
- $\Rightarrow$ By the factorization theorem, T is a sufficient statistic.
- Since this is a full-rank exponential family, T is **complete sufficient**.
- Here $T = \sum_{i=1}^m Y_i = N - m$, which has a Negative Binomial distribution.

General Density for Sequential Experiments

Theorem (5.4)

Suppose $X_1, X_2, \ldots$ are i.i.d. with density f_θ, and sampling continues until a stopping time N (which may depend on the data). If $P_\theta(N < \infty) = 1$, then the joint density of the data $(N, X_1, \ldots, X_N)$ is:

$$\prod_{i=1}^{n} f_\theta(x_i)$$

- So the **likelihood function** (joint density seen as a function of θ) is the same as if we had ignored the optional stopping and fixed N as a fixed constant in advance. (i.e., optional stopping does not change how the data inform you about θ)
- If $f_\theta(x) = e^{\eta(\theta)\cdot T(x) - B(\theta)} h(x)$, the joint density is:

$$\exp\left(\eta(\theta) \cdot \sum_{i=1}^{n} T(x_i) - nB(\theta)\right) \prod_{i=1}^{n} h(x_i)$$

Consequences and Summary

- When the original observations come from an exponential family, the sequential experiment also results in an exponential family.
- The sufficient statistic becomes $(\sum_{i=1}^{N} T(X_i), N)$.
- The canonical parameters include an extra term, $-B(\theta)$.
- The presence of this extra parameter usually means that sequential experiments lead to **curved exponential families**.
  - The inverse binomial example was a special case where this didn't happen because $\sum X_i$ was fixed at m.
- While the likelihood is unchanged, distributional properties of estimators are affected by optional stopping. For instance, the sample average $(\sum X_i)/N$ is generally **biased**.

The Multinomial Distribution

- A generalization of the binomial distribution for n independent trials with outcomes in a finite set $\{a_0, \ldots, a_s\}$.
- Let p_j be the probability of outcome a_j.
- Let $Y_i = e_j$ if trial i has outcome a_j, where e_j is a standard basis vector in $\mathbb{R}^{s+1}$. The Y_i are i.i.d.
- The vector of counts is $X = \sum_{i=1}^{n} Y_i$, where X_j is the number of trials with outcome a_j.

Joint Mass Function of Trials

The joint probability for a specific sequence of outcomes $(y_1, \ldots, y_n)$ is:

$$P(Y_1 = y_1, \ldots, Y_n = y_n) = \prod_{j=0}^{s} p_j^{x_j} = \exp\left(\sum_{j=0}^{s} x_j \log p_j\right)$$

where $x = \sum y_i$ is the vector of total counts.

The Multinomial Mass Function

- The distribution of the count vector X is found by summing the probabilities of all sequences $(y_1, \ldots, y_n)$ that result in the same counts x.
- Each such sequence has the same probability, $\prod p_j^{x_j}$.
- The number of such sequences is given by the multinomial coefficient:
$$\binom{n}{x_0, \ldots, x_s} = \frac{n!}{x_0! \times \cdots \times x_s!}$$

Multinomial Distribution

The probability mass function for the count vector X is:

$$P(X_0 = x_0, \ldots, X_s = x_s) = \binom{n}{x_0, \ldots, x_s} p_0^{x_0} \times \cdots \times p_s^{x_s}$$

We write $X \sim \text{Multinomial}(p_0, \ldots, p_s; n)$.

UMVU Estimation for Multinomial

- The marginal distribution of a single count X_j is Binomial(n, p_j).
- We can also show that X is complete sufficient, and thus $\frac{X_j}{n}$ is UMVUE for p_j (Since $E[X_j] = np_j$ and $E[\frac{X_j}{n}] = p_j$).
- We can find the UMVUE of the product of two multinomial probabilities $p_j p_k$ (for $j \neq k$) using Rao-Balckwellization.
- We first define an unbiased estimator of $p_j p_k$: $\delta = I(Y_1 = a_j, Y_2 = a_k)$.
- $E[\delta] = P(Y_1 = a_j, Y_2 = a_k) = p_j p_k$, so δ is an unbiased estimator.
- But this only uses two trials, not all n. To improve it, we condition on the complete sufficient statistic X.

Derivation of UMVU for $p_j p_k$

By definition of conditional probability:

$$E[\delta|X = x] = P(Y_1 = a_j, Y_2 = a_k|X = x) = \frac{P(Y_1 = a_j, Y_2 = a_k, X = x)}{P(X = x)}$$

The denominator is the standard Multinomial PMF:

$$P(X = x) = \binom{n}{x_0, \ldots, x_s} \prod_{i=0}^{s} p_i^{x_i}$$

The numerator is the probability that the first trial is a_j, the second is a_k, and the remaining $n - 2$ trials result in the remaining counts $(x_0, \ldots, x_j - 1, \ldots, x_k - 1, \ldots, x_s)$:

$$P(Y_1 = a_j, Y_2 = a_k, X = x) = p_j p_k \cdot \binom{n-2}{\ldots} \prod_{i \neq j,k} p_i^{x_i} \cdot p_j^{x_j - 1} \cdot p_k^{x_k - 1}$$

Derivation of UMVU for $p_j p_k$

Simplifying the Numerator's Probability Term

$$
\begin{aligned}
& p_j p_k \cdot \left(\prod_{i \neq j,k} p_i^{x_i} \cdot p_j^{x_j - 1} \cdot p_k^{x_k - 1} \right) \\
&= \left(p_j \cdot p_j^{x_j - 1} \right) \cdot \left(p_k \cdot p_k^{x_k - 1} \right) \cdot \left(\prod_{i \neq j,k} p_i^{x_i} \right) \\
&= p_j^{x_j} \cdot p_k^{x_k} \cdot \left(\prod_{i \neq j,k} p_i^{x_i} \right) \\
&= \prod_{i=0}^{s} p_i^{x_i}
\end{aligned}
$$

Derivation of UMVU for $p_j p_k$

Now, we take the ratio and simplify.

$$
\begin{aligned}
E[\delta|X = x] &= \frac{p_j p_k \frac{(n-2)!}{x_0!\ldots(x_j-1)!\ldots(x_k-1)!\ldots x_s!} \prod_{i=0}^{s} p_i^{x_i} / (p_j p_k)}{\frac{n!}{x_0!\ldots x_j!\ldots x_k!\ldots x_s!} \prod_{i=0}^{s} p_i^{x_i}} \\
&= \frac{\frac{(n-2)!}{x_0!\ldots(x_j-1)!\ldots(x_k-1)!\ldots x_s!}}{\frac{n!}{x_0!\ldots x_j!\ldots x_k!\ldots x_s!}} \\
&= \frac{(n-2)!}{n!} \cdot \frac{x_0! \ldots x_j! \ldots x_k! \ldots x_s!}{x_0! \ldots (x_j - 1)! \ldots (x_k - 1)! \ldots x_s!} \\
&= \frac{(n-2)!}{n(n-1)(n-2)!} \cdot \frac{x_j!}{(x_j - 1)!} \cdot \frac{x_k!}{(x_k - 1)!} \\
&= \frac{1}{n(n-1)} \cdot x_j \cdot x_k \\
&= \frac{x_j x_k}{n(n-1)}
\end{aligned}
$$

Example 5.5: Two-Way Contingency Tables

- Consider n trials where two characteristics are observed for each:
  - Characteristic A with possibilities $\{A_1, \ldots, A_I\}$.
  - Characteristic B with possibilities $\{B_1, \ldots, B_J\}$.

- Let N_{ij} be the count of outcome (A_i, B_j), and p_{ij} its probability. The vector of counts $N = (N_{11}, \ldots, N_{IJ})$ is Multinomial.

- Data is often presented in a table with marginal sums:

	B_1	$\cdots$	B_J	Total
A_1	N_{11}	$\cdots$	N_{1J}	N_{1+}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_I	N_{I1}	$\cdots$	N_{IJ}	N_{I+}
Total	N_{+1}	$\cdots$	N_{+J}	n

Contingency Tables with Independence

The Independence Assumption

If characteristics A and B are independent, then $p_{ij} = p_{i+}p_{+j}$, where p_{i+} and p_{+j} are the marginal probabilities.

- Under independence, the mass function for N simplifies. The probability term of the multinomial distribution becomes:

$$\prod_{i=1}^{I}\prod_{j=1}^{J}(p_{i+}p_{+j})^{n_{ij}} = \left(\prod_{i=1}^{I} p_{i+}^{n_{i+}}\right)\left(\prod_{j=1}^{J} p_{+j}^{n_{+j}}\right)$$

- This model can be written as a **full rank** $(I+J-2)$-**parameter exponential family**.

- The **complete sufficient statistic** is the set of marginal totals:

$$(N_{1+},\dots,N_{I+},N_{+1},\dots,N_{+J})$$

UMVU Estimation under Independence

- In the independence model, the vectors of marginal counts are independent:
  - $(N_{1+}, \ldots, N_{I+}) \sim \text{Multinomial}(p_{1+}, \ldots, p_{I+}; n)$
  - $(N_{+1}, \ldots, N_{+J}) \sim \text{Multinomial}(p_{+1}, \ldots, p_{+J}; n)$

- The UMVU estimator for the marginal probability p_{i+} is $\hat{p}_{i+} = N_{i+}/n$.

Proof of Unbiasedness for $\hat{p}_{i+}$

The expectation of a Multinomial count N_{i+} is $E[N_{i+}] = np_{i+}$.

$$E[\hat{p}_{i+}] = E\left[\frac{N_{i+}}{n}\right] = \frac{1}{n}E[N_{i+}] = \frac{1}{n}(np_{i+}) = p_{i+}$$

Since the estimator is an unbiased function of the complete sufficient statistic, it is the UMVUE by the Lehmann-Scheffé theorem.

UMVU Estimation under Independence

- Similarly, the UMVU estimator for p_{+j} is $\hat{p}_{+j} = N_{+j}/n$.

- Because the estimators for the marginals are based on independent statistics, the UMVU estimator for the joint probability $p_{ij} = p_{i+}p_{+j}$ is the product of the estimators:

$$\boldsymbol{\hat{p}_{ij} = \hat{p}_{i+}\hat{p}_{+j} = \frac{N_{i+}N_{+j}}{n^2}}$$

Proof of Unbiasedness for $\hat{p}_{ij}$

Using the independence of the marginal statistics N_{i+} and N_{+j}:

$$\begin{aligned}
E[\hat{p}_{ij}] &= E\left[\frac{N_{i+}N_{+j}}{n^2}\right] = \frac{1}{n^2}E[N_{i+}N_{+j}] \\
&= \frac{1}{n^2}E[N_{i+}]E[N_{+j}] \quad \text{(by independence)} \\
&= \frac{1}{n^2}(np_{i+})(np_{+j}) \\
&= p_{i+}p_{+j} = p_{ij}
\end{aligned}$$