10 Non-Random Parameter Estimations

Criteria for Non-Random Parameters

- For random parameter estimation problems, we assume that one has a prior distribution on θ and we can define a global estimation error criterion, the *Bayes risk*, which depends on the prior but not on any particular value of the parameter.

- In some cases, however, there is no such prior information. This is called *non-random parameter estimation*. For this problem, the formulation of optimal non-random parameter estimation requires a completely different approach. This is because if we do not have a prior distribution on the parameter virtually any reasonable estimation error criterion will be local, i.e., it will depend on the true parameter value.

- One possible approach is to consider the *minimax criteria* for the worst-case, i.e.,

$$\hat{\theta} = \arg\min_{\hat{\theta}} \left\{ \max_{\theta\in\Theta} \left\{ E_\theta(\mathsf{cost}(\hat{\theta},\theta)) \right\} \right\}, \text{ where } E_\theta(\mathsf{cost}(\hat{\theta},\theta)) = \int_{x\in\mathcal{X}} \mathsf{cost}(\hat{\theta}(x),\theta) f(x;\theta)dx$$

- However, the above minmax formulation is in general very difficult to solve. Here we consider several weaker conditions that a "good" estimator should satisfy. Formally, given an estimator $\hat{\theta} : X \to \Theta$, we define

  - the **estimator bias** at a point θ to be $\mathsf{b}_\theta(\hat{\theta}) = E_\theta[\hat{\theta}] - \theta$;
  - the **estimator variance** at a point θ to be $\mathsf{var}_\theta(\hat{\theta}) = E_\theta[(\hat{\theta} - E_\theta(\hat{\theta}))^2]$;

  We say an estimator $\hat{\theta}$ is **unbiased** if $\forall\theta \in \Theta : \mathsf{b}_\theta(\hat{\theta}) = 0$. Note that for MSE, we have

$$\mathsf{MSE}_\theta(\hat{\theta}) = E[(\hat{\theta}-\theta)^2] = E[(\hat{\theta}-E_\theta(\hat{\theta}))^2]+[E_\theta(\hat{\theta})-\theta]^2+2E_\theta[\hat{\theta}-E_\theta(\hat{\theta})]\mathsf{b}_\theta(\hat{\theta}) = \mathsf{var}_\theta(\hat{\theta})+\mathsf{b}^2_\theta(\hat{\theta})$$

- It is natural to require that a good estimator is unbiased. This suggests a reasonable design approach: constrain the class of admissible estimators to be unbiased and try to find one that minimizes variance over this class.

- In some cases such an approach leads to a really good, in fact optimal, unbiased estimator called the **Uniform Minimum Variance Unbiased** (UMVU) estimator $\hat{\theta}_{\text{UMVU}}$, i.e., $\hat{\theta}_{\text{UMVU}}$ is unbiased and $\forall\theta \in \Theta : \mathsf{var}_\theta(\hat{\theta}_{\text{UMVU}}) \leq \mathsf{var}_\theta(\hat{\theta})$ for any unbiased $\hat{\theta}$.

Asymptotic Properties of "Good" Estimators

- Consider estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$. In some cases, $\hat{\theta}_n$ may not be "good" enough when observing finite n observations, but it may converge to some "good" properties asymptotically. Formally, we say $\hat{\theta}_n$ is

  - **asymptotically unbiased**: if $\forall\theta \in \Theta : \lim_{n\to\infty} \mathsf{b}_\theta(\hat{\theta}_n) = 0$;
  - **consistent**: if $\forall\theta \in \Theta : \lim_{n\to\infty} \mathsf{MSE}_\theta(\hat{\theta}_n) = 0$;
  - **weakly consistent**: if $\forall\theta \in \Theta, \forall\epsilon > 0 : \lim_{n\to\infty} P_\theta(|\hat{\theta}_n - \theta| > \epsilon) = 0$;

- Next, we will consider how to design good estimators for non-random parameters. We consider two classes of estimations: method of moments and maximum likelihood.

Method of Moments Estimators (Scalar Parameter)

- We consider the scalar case, i.e., $\theta = (\theta_1)$. The basic idea of the method of moments is as follows. Suppose $X = (X_1, \dots, X_n)$ are i.i.d. and each X_i follows from the statistic model $f(x;\theta), \theta \in \Theta$. We can first compute kth order moment by

$$m_k(\theta) = E_\theta[X_i^k] = \int_{x\in\mathcal{X}} x^k f(x;\theta)dx$$

- Note that m_k is a function of θ, say $m_k = g_k(\theta)$. Suppose g_k is *invertible* for some order k. Then we can compute θ by $\theta = g_k^{-1}(m_k)$

- The problem is that we do not know m_k because we do not know the parameters $\theta \in \Theta$. However, if observations $X = (X_1, \dots, X_n)$ are i.i.d. from $f(x;\theta)$, then we can *estimate* the kth order moment m_k by its kth sample moment

$$\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

  Note that $\hat{m}_k$ is an unbiased, consistent estimator for $m_k(\theta)$ because

$$\mathsf{b}_\theta(\hat{m}_k) = E_\theta\left[\frac{1}{n}\sum_{i=1}^{n} X_i^k\right] - m_k = \frac{1}{n}\left(E_\theta\left[X_1^k\right] + \cdots + E_\theta\left[X_n^k\right]\right) - m_k = 0$$

$$\mathsf{var}_\theta(\hat{m}_k) = \mathsf{var}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i^k\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathsf{var}_\theta(X_i^k) = \frac{1}{n}\mathsf{var}_\theta(X_i^k) \to 0$$

- Therefore, by replacing m_k by its unbiased estimate $\hat{m}_k$, we get the moment estimator

$$\hat{\theta}_{\text{MOM}} = g_k^{-1}(\hat{m}_k)$$

Maximum Likelihood Estimators (Scalar Parameter)

- The basic idea of the maximum likelihood method is as follows. For any observation $x = (x_1, \dots, x_n)$, we define the likelihood function

$$L(\theta) = f(x;\theta)$$

  which is just the joint PDF of $X = (X_1, \dots, X_n)$. Sometimes it is more convenient to use the log-likelihood function $l(\theta) = \ln f(x;\theta)$.

- Then the maximum likelihood estimator simply maximizes the likelihood function by

$$\hat{\theta}_{\text{ML}} = \arg\max_\theta f(X;\theta) = \arg\max_\theta L(\theta) = \arg\max_\theta l(\theta)$$

  Intuitively, it chooses parameter θ under which the observation x is the most likely.

- Actually, the maximum likelihood estimator only depends on the sufficient statistic. For example, when $T = T(X)$ is an SS, we can write $f(X;\theta) = g(T;\theta)h(X)$. Therefore,

$$\hat{\theta}_{\text{ML}} = \arg\max_\theta f(X;\theta) = \arg\max_\theta g(T;\theta)$$

Examples for Methods of Moments

- **Example 1: Bernoulli Random Variables**
  Bernoulli measurements arise anytime one deals with binary quantized versions of continuous variables, e.g., exceed a threshold. Then the parameter of interest is typically the probability of success, i.e., the probability that the measured variable is a "logical one".

  Then the model is $X = (X_1, \ldots, X_n)$, where X_i are i.i.d. and $X_i \sim \theta^x(1-\theta)^{1-x}, x \in \{0, 1\}$. Since $m_1(\theta) = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = \theta$, we can compute an MOM estimator by:

$$\hat{\theta}_{\text{MOM}} = \hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

- **Example 2: Poisson Random Variables**
  Poisson measurements are ubiquitous in many scenarios where there are counting measurements. That is the total number of counts registered over a finite time interval is a Poisson random variable with a rate parameter.

  Then the model is $X = (X_1, \ldots, X_n)$, where X_i are i.i.d. and $X_i \sim f(x; \theta) = \frac{\theta^x}{x!}e^{-\theta}, x \in \{0, 1, 2, \ldots\}$. As we have already verified before, $m_1(\theta) = E(X_i) = \theta$. Therefore, like in the Bernoulli example a possible MOM estimator of θ is the sample mean

$$\hat{\theta}_1 = \hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

  Alternatively, we can compute the second moment, which is $m_2(\theta) = E(X_i^2) = \theta + \theta^2$. Therefore, another MOM estimator is the (positive) value of $\hat{\theta}_2$ which satisfies the equation

$$\hat{\theta}_2 + \hat{\theta}_2^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \overline{X^2} \quad \Rightarrow \quad \hat{\theta}_2 = \frac{-1 \pm \sqrt{1 + 4\overline{X^2}}}{2}$$

  As yet another example, we can express $m_2(\theta)$ as $m_2(\theta) = \theta + m_1^2(\theta)$. Hence, another MOM estimator is

$$\hat{\theta}_3 = \overline{X^2} - (\overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

  If we compare $\hat{\theta}_1$ and $\hat{\theta}_3$, we have

$$E_\theta(\hat{\theta}_1) = \theta, \mathsf{var}_\theta(\hat{\theta}_1) = \frac{\theta}{n} \quad \text{and} \quad E_\theta(\hat{\theta}_3) = \frac{n-1}{n}\theta, \mathsf{var}_\theta(\hat{\theta}_3) \approx \frac{2\theta^2 + \theta}{n}$$

  Then we note that $\hat{\theta}_1$ is unbiased while $\hat{\theta}_2$ and $\hat{\theta}_3$ are asymptotically unbiased. Furthermore, $\hat{\theta}_1$ compares favorably to $\hat{\theta}_3$ since it has both lower bias and lower variance.

General Properties for MOM Estimators

- The MOM estimator has the following properties:
  - MOM estimators are asymptotically unbiased as $n \to \infty$
  - MOM estimators are consistent.
  - MOM estimator is not unique, i.e., it depends on what order moment is used.
  - MOM is inapplicable in cases where moments do not exist (e.g. Cauchy r.v.).

Examples for Methods of Maximum Likelihood

- **Example 1: Poisson Random Variables**
  For $X = (X_1, \ldots, X_n)$, where each $X_i \sim f(x;\theta) = \frac{\theta^x}{x!}e^{-\theta}, x = 0, 1, \ldots$ is i.i.d., by considering sufficient statistic $T = \sum_{i=1}^n X_i$, we have

$$f(X;\theta) = \prod_{i=1} \frac{\theta^{X_i}}{X_i!}e^{-\theta} = \theta^{\sum_{i=1}^n X_i} e^{-n\theta} \prod_{i=1}^n \frac{1}{X_i!} = \underbrace{\theta^T e^{-n\theta}}_{g(T,\theta)} \underbrace{\prod_{i=1}^n \frac{1}{X_i!}}_{h(X)}$$

  Then, by maximizing $g(T,\theta)$, we compute the ML estimator as

$$\frac{d}{d\theta}\ln g(T,\theta) = \frac{d}{d\theta}(T\ln\theta - n\theta) = \frac{T}{\theta} - n = 0 \quad \Rightarrow \quad \hat{\theta}_{\text{ML}} = \frac{T}{n} = \overline{X}$$

  Also, the ML estimator is unbiased and we have

$$\mathsf{b}_\theta(\hat{\theta}_{\text{ML}}) = E_\theta(\hat{\theta}_{\text{ML}}) - \theta = 0 \text{ and } \mathsf{var}_\theta(\hat{\theta}_{\text{ML}}) = \frac{1}{n^2}\mathsf{var}_\theta(T) = \frac{1}{n}\mathsf{var}_\theta(X_i) = \frac{\theta}{n}$$

- **Example 2: Uniform Random Variables**
  For $X = (X_1, \ldots, X_n)$, where each $X_i \sim f(x;\theta) = \frac{1}{\theta}\mathbf{1}_{[0,\theta]}(X_i)$ is i.i.d., by considering sufficient statistic $T = \max_{i=1}^n X_i$, we have

$$f(X;\theta) = \prod_{i=1} \frac{1}{\theta}\mathbf{1}_{[X_i,\infty)}(\theta) = \underbrace{\frac{1}{\theta^n}\mathbf{1}_{[T,\infty)}(\theta)}_{g(T,\theta)}$$

  Since $g(T,\theta)$ is decreasing in θ from T, to maximize $g(T,\theta)$, the ML estimator is

$$\hat{\theta}_{\text{ML}} = T = \max_{i=1,\ldots,n} X_i$$

  In fact, we can compute the PDF of $f(t;\theta)$ by

$$f(t;\theta) = \frac{d}{dt}P_\theta(T \le t) = \frac{d}{dt}P_\theta(X_1 \le t, \ldots, X_n \le t) = \frac{d}{dt}\left(\frac{t}{\theta}\right)^n = \frac{nt^{n-1}}{\theta^n}\mathbf{1}_{[t,\infty)}(\theta)$$

  Actually, we can see that the ML estimator is biased, but asymptotically unbiased

$$\mathsf{b}_\theta(\hat{\theta}_{\text{ML}}) = E_\theta(\hat{\theta}_{\text{ML}}) - \theta = \int_{-\infty}^{\infty} t f(t;\theta)dt - \theta = \frac{n}{\theta^n}\int_0^\theta t^n dt - \theta = \frac{n}{n+1}\theta - \theta$$

  Also, we can show that the ML estimator is consistent because

$$\mathsf{var}_\theta(\hat{\theta}_{\text{ML}}) = E_\theta(\hat{\theta}_{\text{ML}}^2) - E_\theta^2(\hat{\theta}_{\text{ML}}) \propto \frac{\theta^2}{n} \to 0 \text{ a.s. } n \to \infty$$

Examples for Methods of Maximum Likelihood (Cont.)

- **Example 3: None-Existence of Unbiased Estimator**
  Let us consider i.i.d. $X = (X_1, \ldots, X_n)$, where each $X_i \sim f(x;\theta) = \frac{\theta^x}{1+\theta}, x = \{0, 1\}$. We can show that $T = \sum_{i=1}^n X_i$ is a sufficient statistic

$$f(X;\theta) = \frac{1}{(1+\theta)^n}\theta^{\sum_{i=1}^n X_i} = \underbrace{\frac{1}{(1+\theta)^n}\theta^T}_{g(T,\theta)}$$

  Then, by maximizing $g(T,\theta)$, we compute the ML estimator as

$$\frac{d}{d\theta}\ln g(T,\theta) = \frac{d}{d\theta}(-n\ln(1+\theta) + T\ln\theta) = -\frac{n}{1+\theta} + \frac{T}{\theta} = 0 \quad \Rightarrow \quad \hat{\theta}_{\text{ML}} = \frac{T}{n-T}$$

  One can verify that the above ML estimator is biased. In fact, we argue a stronger result that there is no unbiased estimator for this parameter! To this end, we consider the case of $n = 1$, i.e., we only make a single observation X_1. Let $\hat{\theta}$ be any function of X_1 and assume $\hat{\theta}(1) = a$ and $\hat{\theta}(0) = b$. If $\hat{\theta}$ is unbiased, then we should have the following relation

$$E_\theta(\hat{\theta}) = a \cdot \frac{\theta}{1+\theta} + b \cdot \frac{1}{1+\theta} = \theta$$

  However, we cannot find such a and b because

$$\underbrace{a\theta + b}_{\text{linear}} = \underbrace{\theta(1+\theta)}_{\text{quadratic}}$$

General Properties for ML Estimators

- MLEs are **asymptotically unbiased**, i.e.,

$$\lim_{n\to\infty}(\mathsf{b}_\theta(\hat{\theta}_{\text{ML}})) = \lim_{n\to\infty}(E_\theta(\hat{\theta}_{\text{ML}}) - \theta) = 0, \forall\theta \in \Theta$$

- MLEs are **consistent** , i.e.,

$$\forall\theta \in \Theta : \lim_{n\to\infty}\mathsf{MSE}_\theta(\hat{\theta}_{\text{ML}}) = \lim_{n\to\infty}\left(\mathsf{var}_\theta(\hat{\theta}_{\text{ML}}) + \mathsf{b}_\theta^2(\hat{\theta}_{\text{ML}})\right) = 0$$

- MLEs are invariant under any transformation of parameters, i.e.,

$$\varphi = g(\theta) \quad \Rightarrow \quad \hat{\varphi} = g(\hat{\theta})$$

- MLEs are **asymptotically UMVU** in the sense that

$$\lim_{n\to\infty} n\mathsf{var}_\theta(\hat{\theta}_{\text{ML}}) = \frac{1}{F_1(\theta)}$$

  where $1/F_1(\theta)$ is a quantity known as the Fisher information, which will be introduced soon, and $1/F_1(\theta)$ specifies the fastest possible asymptotic rate of decay of any unbiased estimator's variance.

- MLEs are **asymptotically Gaussian** in the sense that

$$\sqrt{n}(\hat{\theta}_{\text{ML}} - \theta) \to N\left(0, \frac{1}{F_1(\theta)}\right)$$

The Cramér-Rao Bound on Estimator Variance

- We naturally has the following question: "given a model $f(x;\theta)$, does there exist an unbiased estimator with minimum variance?" The answer is actually provided by the well-known **Cramér-Rao Lower Bound** (CRLB).
- To present the CRLB, we first introduce the notion of **Fisher Information** $F(\theta)$ associated with a scalar parameter θ

$$F(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right] = E_\theta\left[-\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)\right]$$

- Then the main theorem is as follows

Theorem: Cramér-Rao Lower Bound

Let $\theta \in \Theta$ be a non-random scalar parameter and assume Θ is an open subset of $\mathbb{R}$ and $f(x;\theta)$ is differentiable in θ. For any unbiased estimator $\hat{\theta}$ of θ, we have

$$\mathsf{var}_\theta(\hat{\theta}) \geq \frac{1}{F(\theta)}$$

where "=" holds if and only if for some non-random scalar k_θ such that

$$\frac{\partial}{\partial\theta}\ln f(x;\theta) = k_\theta(\hat{\theta}-\theta)$$

Then the quantify $1/F(\theta)$ is called the Cramér-Rao Bound (CRB). When the CRB is attainable it is said to be a tight bound.

Proof of the CRB

- The first step is to notice that the mean of the derivative of the log-likelihood is equal to zero:

$$E_\theta\left[\frac{\partial \ln f(X;\theta)}{\partial\theta}\right] = E_\theta\left[\frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta}f(X;\theta)\right] = \int \frac{\partial f(x;\theta)}{\partial\theta}dx = \frac{\partial}{\partial\theta}\int f(x;\theta)dx = 0$$

- The second step is to show that the correlation between the derivative of the log-likelihood and the estimator is zero:

$$E_\theta\left[(\hat{\theta}-\theta)(\frac{\partial \ln f(X;\theta)}{\partial\theta})\right] = E_\theta\left[\hat{\theta}(\frac{\partial \ln f(X;\theta)}{\partial\theta})\right] - \theta\underbrace{E_\theta\left[\frac{\partial \ln f(X;\theta)}{\partial\theta}\right]}_{=0} = \int \hat{\theta}\frac{\partial f(x;\theta)}{\partial\theta}dx = \frac{\partial}{\partial\theta}\underbrace{\int \hat{\theta}(x)f(x;\theta)dx}_{=E_\theta(\hat{\theta})=\theta} = 1$$

- Recall the Cauchy-Schwarz inequality $E^2(UV) \leq E(U^2)E(V^2)$, where "=" when $U = kV$.

$$1 = \left(E_\theta\left[(\hat{\theta}-\theta)(\frac{\partial \ln f(X;\theta)}{\partial\theta})\right]\right)^2 \leq E[(\hat{\theta}-\theta)^2]E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right] = \mathsf{var}_\theta(\hat{\theta})F(\theta)$$

This gives us the CRLB $\frac{1}{F(\theta)} \leq \mathsf{var}_\theta(\hat{\theta})$

Examples of the CRB

- **Example 1:** For $X = (X_1, \ldots, X_n)$, where each $X_i \sim \frac{\theta^x}{x!}e^{-\theta}$ is i.i.d Poisson

$$\frac{\partial}{\partial\theta}\ln f(X;\theta) = \frac{\partial}{\partial\theta}\left[(\sum_{i=1}^{n} X_i)\ln(\theta) - n\theta + c\right] = \frac{1}{\theta}\sum_{i=1}^{n} X_i - n = \underbrace{\frac{n}{\theta}}_{k_\theta}\left(\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\hat{\theta}_{\text{ML}}} - \theta\right)$$

Therefore, we can conclude that $\hat{\theta}_{\text{ML}} = \overline{X}$ is actually optimal among all unbiased estimators. Furthermore, the Fisher information is

$$F(\theta) = E_\theta\left[-\frac{\partial^2}{\partial\theta_2}\ln f(X;\theta)\right] = E_\theta\left[\frac{1}{\theta^2}\sum_{i=1}^{n} X_i\right] = \frac{n}{\theta}$$

Therefore, for any unbiased $\hat{\theta}$, we have $\mathsf{var}_\theta(\hat{\theta}) \geq \frac{\theta}{n}$ and "=" is achieved by $\hat{\theta} = \hat{\theta}_{\text{ML}}$.

- **Example 2:** For $X = (X_1, \ldots, X_n)$, where each $X_i \sim \binom{m}{x}\theta^x(1-\theta)^{m-x}$ is i.i.d. binomial

$$f(X;\theta) = \prod_{i=1}^{n} f(X_i;\theta) = \theta^{\sum_{i=1}^{n} X_i}(1-\theta)^{mn-\sum_{i=1}^{n} X_i}\prod_{i=1}^{n}\binom{m}{X_i} = \theta^{n\overline{X}}(1-\theta)^{n(m-\overline{X})}\prod_{i=1}^{n}\binom{m}{X_i}$$

and

$$\frac{\partial}{\partial\theta}\ln f(X;\theta) = \frac{n}{\theta}\overline{X} - \frac{n(m-\overline{X})}{1-\theta} = \underbrace{\frac{nm}{\theta(1-\theta)}}_{k_\theta}(\underbrace{\frac{1}{m}\overline{X}}_{\hat{\theta}} - \theta)$$

One can verify that $\frac{1}{m}\overline{X}$ is an unbiased estimator of θ

$$E_\theta(\hat{\theta}) = E_\theta(\frac{1}{m}\overline{X}) = \frac{1}{nm}\sum_{i=1}^{n} E_\theta(X_i) = \frac{1}{nm}\cdot n\cdot E_\theta(X_i) = \theta$$

Therefore, it achieves the CRB and the Fisher information is

$$F(\theta) = E_\theta\left[-\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)\right] = \frac{nm}{\theta(1-\theta)}$$

which means that $\mathsf{var}_\theta(\hat{\theta}) = \frac{\theta(1-\theta)}{nm}$.

Exponential Family

- We are interested in finding efficient estimators that are UMVU. However, as we have already shown, unbiased estimator may not even exist. The concept of the fexponential family provides us a sufficient condition for the existence of efficient estimators. Formally, we say a random variable X is in the **exponential family** if its PDF is of form

$$f(x;\theta) = a(\theta)b(x)e^{c(\theta)T(x)}$$

- Examples of distributions in the exponential family include: Gaussian with unknown mean or variance, Poisson with unknown mean, exponential with unknown mean, Bernoulli with unknown success probability, binomial with unknown success probability.
- Then the main theorem is as follows

Theorem: Exponential Family and Efficient Estimators

Efficient estimators only exist when $f(x;\theta)$ is a member of exponential family.

Proof: Based on the CRLB, for any efficient estimator $\hat{\theta}$, we have

$$\int_{\theta_0}^{\theta'} \frac{\partial}{\partial \theta} \ln f(x;\theta) d\theta = \int_{\theta_0}^{\theta'} k_\theta(\hat{\theta} - \theta) d\theta$$

This gives us

$$\ln f(x;\theta') - \underbrace{\ln f(x;\theta_0)}_{b(x)} = \underbrace{\hat{\theta}}_{T(x)} \underbrace{\int_{\theta_0}^{\theta'} k_\theta d\theta}_{c(\theta')} - \underbrace{\int_{\theta_0}^{\theta'} k_\theta \theta d\theta}_{d(\theta')} \quad \Rightarrow \quad f(x;\theta) = \underbrace{e^{-d(\theta)}}_{a(\theta)} b(x) e^{c(\theta)T(x)}$$

Therefore, $f(x;\theta)$ is a member of exponential family.