5.2. İstatistik Parametre Kestirim Yöntemleri

Bilindiği gibi, parametre tahmini için en iyi kestirim yöntemi yukarıda sözü edilen dört özelliğe sahip bir kestirim yöntemidir. Ne var ki, bir tahminde, ancak bunlardan bazıları geçerliliğini korurken bazıları da asimtotik olarak geçerli olabilir. Böyle durumlarda bu kestirim, n örnekleme veri değerini göstermek üzere, $n \to \infty$ için *asimtotik* bir tahmin yöntem olur.

Günümüzde çeşitli amaçlar için kullanılmak üzere geliştirilmiş birçok parametre kestirim yöntemi mevcuttur. Bunlardan bazıları istatistik anlamda direkt çözümler olmasına rağmen, diğer bir kısmı da sadece parametre tahmini için kullanılan iteratif ya da koşullu olasılık çözümleri olmaktadır. Burada, bu kestirim yöntemlerden istatistik anlamda direkt çözüm veren ve aynı zamanda konuyla yakından ilişkili olan birkaç yöntemden kısaca söz edilecektir.

5.2.1. Moment Yöntemi

Moment yöntemine göre parametre kestiriminde, bir değişkenin *k.* '*inci* dereceden momentlerinden faydalanılır. Böyle bir işlem paragraf 4.1.1. 'deki bigilerlerden faydalanılarak,

$$\mu_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k \qquad 5\text{-}14$$

şeklindeki bir formülle hesaplanabilir. Bu tahmin, aynı zamanda ayrık türden bir rastgele değişken için olasılık dağılımının *k. 'inci* dereceden momentinin bir kestirim değeri olmaktadır.

Bu şekildeki bir tahmin için, $k=1$ alınırsa,

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad 5\text{-}15$$

bağıntısından hesaplanan birinci dereceden moment değeri ya da ölçülerin aritmetik ortalaması buna bir örnek teşkil etmektedir.

Burada, x_i rastgele değişkenleri yansız olduğu sürece, buradan hesaplanacak kestirim parametresi olan ortalama değeri de *asimtotik* ve *yansız* birer tahmin olur. Aksi halde yanlı ve *asimtotik* bir tahmin değeri olmaz.

Benzer şekilde, (5-14) denkleminde x_i^k yerine; bunun μ parametresi kadar ötelenmiş $(x-\mu)$ değerinin alınması ile Paragraf 4.1.1. de anlatılanlara benzer şekilde, ayrık türden değişkenler için *k.'inci dereceden* merkezsel moment değerleri,

$$E\{(x-\mu)k\} = \frac{1}{n}\sum_{i=1}^{n}(x-\mu)^k \qquad 5\text{-}16a$$

bağıntısı ile ifade edilebilir. Bu şekilde ifade edilen bir merkezsel moment bağıntısında; $k=2$ alınarak,

$$\sigma^2 = E\{(x-\mu)^2\} = \frac{1}{n}\sum_{i=1}^{n}(x-\mu)^2 \qquad 5\text{-}16b$$

şeklindeki hesaplanan ikinci derecen merkezsel moment, diğer bir ifade ile *varyans* değeri de benzer şekilde, *asimtotik* ve *yansız* ya da *kayık* olmayan bir tahmin değeri olmaktadır.

5.2.2. En Küçük Kareler Kestirim Yöntemi

C. F. GAUSS yöntemi ile *En küçük kareler* yöntemi arasındaki yakın ilişkiyi anlamak için, *GAUSS'un 1809* yılında yayınlamış olduğu "*Theoria Motus Corporum Coelestium*" adlı eserinde verilen aşağıdaki paragrafı okumak gerekir.

> *If the astronomical observations and other quantities on which the computation of orbits is based were absolutely correct, the elements also, whether deduced from three or four observations, would be strictly accurate (so far indeed as the motion is supposed to take place exactly according to the law of Kepler) and, therefore, if other observations were used, they might be confirmed but not corrected. But since all our measurements and observations are nothing more than approximations to the truth, the same must be true of all calculations resting upon them, and the highest aim of all computations made concerning concrete phenomena must be to approximate, as nearly as practicable, to the truth. But this can be accomplished in no other way than by a suitable combination of more observations than the number absolutely requisite for the determination of the unknown quantities. This problem can only be properly undertaken when an approximate knowledge of the orbit has been already attained, which is afterwards to be corrected so as to satisfy all the observations in the most accurate manner possible"*

Bugün itibariyle, bundan yaklaşık *200* yıl önce yazılan bu paragrafın incelenmesinden aşağıdaki kavramları somutlaştırmak mümkündür.

- *Matematik modeller eksik olabilir,*
- *Fiziksel gözlemler tutarsız ya da çelişkili, diğer bir deyişle hatalı (inconsistent) olabilir,*
- *Tutarsız ölçülere dayanan hesaplamalardan tahmin edilen her şey gerçek değerin bir tahminidir,*
- *Fazla sayıda yapılan ölçmeler, ölçüler arasındaki tutarsızlıkların etkilerini de azaltacaktır,*
- *Kesin tahmin değerleri için herhangi bir başlangıç yaklaşımları kullanılmalı, bu başlangıç değerleri gözlemler arasındaki tutarsızlıkları minimum kılacak şekilde düzeltilmelidir.*

Burada yapılan açıklamalardan da görüldüğü gibi, *C. F. GAUSS* daha *1795* yılında kullandığı ve halen bugünkü anlamda da kullanılmakta olan *En küçük kareler* yöntemini ilk defa o yıllarda keşfetmiş ve ancak bunu daha sonra, *1809* yılında yazılı eser haline getirerek yayınlamıştır *(Wells – Krakıwsky, 1971)*. Bu tarihten sonra, yaklaşık bir ya da iki yüzyıldır, *En küçük kareler yöntemi* mühendislik ve bilimin çeşitli branşlarında başarı ile uygulanmaktadır.

Özet olarak, orijinal şekliyle verilmiş olan yukarıdaki paragrafta görüldüğü gibi, *En küçük kareler* yöntemi sadece, gözlemlerdeki tutarsızlıklara neden olan, *"ölçü hatalarının karelerinin toplamının minimum yapılması"* prensibine dayanmaktadır. Böyle bir kestirim ilkesinin ilk orijinal şekliyle analitik ifadesi de: l_i ölçü değerini, μ ölçülen büyüklüğün gerçek değerini ve $-\varepsilon_i = l_i - \mu$ ölçülerin gerçek hatasını göstermek üzere hataların kareleri toplamının minimum kılınması ilkesi;

$$[\varepsilon\varepsilon] \rightarrow \min.$$

şeklinde yapılabilir. En küçük kareler ilkesiyle ilgili bu şekilde yapılmış bir tanım; ölçülerin korelasyonsuz ve farklı duyarlıkta olmaları halinde bu ifadeye; daha sonraki yıllar; p_i her bir ölçünün ne derece güvenilir olduğunu gösteren ölçü ağırlıkları da dahil edilerek bu tür ölçüler için,

$$[p\varepsilon\varepsilon] \rightarrow \min.$$

"*Ölçü hatalarının ağırlıklı karelerinin toplamının minimum olması"* biçiminde bir ilave yapılarak daha da genişletilmiş biçimidir.

Ancak, ne var ki, bu tanım ölçüler arasındaki korelayonları içermediğinden bu haliyle korelasyonlu gözlemler için doğrudan kullanılamamaktadır. Daha sonraki yıllar Hata teorisi konusuyla ilgili tanımlarda artan gereksinimler karşısında *ters ağırlık katsayıları* kavramının tanımlanıp hata teorisinde kullanılması ile, *En küçük kareler ilkesi* tanımı daha da genişletilerek, ölçülerin korelasyonlu ve farklı duyarlıklarda olmaları halinde; aralarındaki olası düşünceye dayalı *stokastik* ilişkileri gösteren *Varyans-Kovaryanslar*

$$\Sigma_{ll} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & ..... & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & ...... & \sigma_{2n} \\ : & : & & : \\ \sigma_{n1} & \sigma_{n2} & ..... & \sigma_n^2 \end{bmatrix} \qquad \text{5-17a}$$

matrisi ile gerçek ölçü hatalarından oluşan gerçek hata vektörü,

$$\varepsilon^T = [\varepsilon_1 \quad \varepsilon_2 \quad ..... \quad \varepsilon_n] \qquad \text{5-17b}$$

olmak üzere en genel anlamdaki tanım,

$$\varepsilon^T \Sigma_{ll}^{-1} \varepsilon \rightarrow \min .$$

şeklinde ifade edilmektedir. Burada, ölçüler arasındaki *korelasyonları* ve ölçülerin varyanslarını içeren, Σ_{ll} *varyans-kovaryans* matrisi; n ölçü sayısını göstermek üzere nxn boyutunda simetrik pozitif tanımlı bir kare matris olmaktadır.

Ne var ki, ilk yıllarda bu şekilde tanımlanan *En küçük kareler yöntemi;* ölçüler arasında tutarsızlığa neden olan gerçek ölçü hatalarının çoğu zaman bilinemediğinden dolayı bu tanım, daha sonraki *1960*'lı yıllarda, her durumda belli bir matematik modele göre hesaplanabilen "*düzeltmeler*" sözcüğü kullanılarak değiştirilmiştir.

Neticede bu tarihten sonra, *En küçük kareler kestirim yöntemi*, hatalara göre değil de, düzeltme değerlerine göre, *"düzeltmelerin kareleri toplamının minimum olması"* şeklinde tanımlanmıştır. O zaman, gerçek ölçü hatalarına göre matematik olarak ifade edilen En küçük kareler yöntemi; μ yerine $\hat{x}$ tahmin değeri kullanılarak bir ölçü kümesinden $v_i = \hat{x} - l_i$ olarak hesaplanabilen görünen düzeltmeler cinsinden deneysel değerlere göre; korelasyonsuz ve eşit duyarlıklı gözlemler için

$$[vv] \rightarrow \min .$$

farklı duyarlıklı ve korelasyonsuz gözlemler için,

$$[pvv] \rightarrow \min .$$

ve farklı duyarlıklı korelasyonlu gözlemler için de;

$$K_{ll} = \begin{bmatrix} m_1^2 & m_{12} & ..... & m_{1n} \\ m_{21} & m_2^2 & ...... & m_{2n} \\ : & : & & : \\ m_{n1} & m_{n2} & ..... & m_n^2 \end{bmatrix} \quad ; \quad v = \begin{bmatrix} v_1 \\ v_2 \\ : \\ v_n \end{bmatrix}$$

K_{ll} : *ölçülerin varyans-kovaryansları matrisi,* v : *düzeltme vektörü* ve

$$Q_{ll} = \frac{1}{\sigma_0^2} K_{ll}$$

ölçülerin ters ağırlık katsayıları matrisi olmak üzere

$$v^T Q_{ll}^{-1} v \rightarrow \min . \qquad 5\text{-}18$$

olarak değiştirilerek ifade edilmiştir. Bu tanım halen bugün de aynı şekliyle kullanılmaktadır.

GAUSS 'un *1809* yılında kaleme alarak yazmış olduğu yukarıda metni verilmiş olan eserinden elde edilmiş böyle bir sonuca rağmen, yine aynı paragrafın uğraşılan diğer konular yönünden tekrar incelenmesinden çıkartılabilecek bir başka sonuç; *En Küçük Kareler Yönteminin* ilk uygulandığı yıllarda gözlemlerin istatistik dağılımları ile hiç ilgilenilmemiş olmasıdır. Bu yıllarda sadece, basit bir analitik denklem eşitsizliği çözümü olan *En Küçük Kareler Parametre Kestirim Yöntemi* ile uğraşılmıştır. Buna göre, günümüzde halen parametre kestiriminde kullanılmakta olan "*Toplam En Küçük Kareler Parametre Kestirimi Yöntemi"* daha sonraları gündeme gelmiş bir diğer yeni parametre tahmin yöntemi olmaktadır.

Sonuçta buradan rahatlıkla söylenebilir ki, *En Küçük Kareler Tahmin Yönteminde*, bu gibi dağılım bilgilerine, sadece parametre kestirimi sonucunda elde edilmiş olan tahmin değerlerinin istatistik yönden irdelenmesinde ihtiyaç duyulmaktadır. Bu haliyle de *En Küçük kareler Yöntemi* diğer kestirim yöntemlerine oranla en basit bir tahmin yöntemi olmaktadır. Ancak, ölçülerdeki bir kaba hatanın kestirim sonucunda bütün tahmin parametrelerini olumsuz yönde fazla etkileyeceğinden bu yöntemin kırılma noktası da oldukça düşük olmaktadır. Bu özellik, bir parametre tahmin yöntemi için istenmeyen olumsuz ve tenkit edilir bir durum olmaktadır. Buna karşılık, parametre kestiriminde gözlemlerin dağılımlarının bilinmesine ihtiyaç duyulmamış olması, istatistik dağılım bilgilerinin sadece sonuç ürünlerinin geçerliliğinin ve güvenirliklerinin irdelenmesinde kullanılmış olması *En Küçük Kareler Yönteminin* basitliği yanında bir diğer olumlu önemli yanını oluşturmaktadır.

5.2.3. Maximum Likelihood Kestirim Yöntemi

Maximum Likelihood Metodu ya da diğer adıyla *"Maksimum Benzerlik Kestirimi"* Matematik-istatistikte parametre kestirimi için, geniş ölçüde kullanılan bir diğer parametre tahmin yöntemidir. Bu yöntemde maksimum benzerliğe sahip en olasılıklı parametre kestirim değeri olarak; bir veri kümesinde en fazla olasılıklı değere sahip parametre o veri kümesinin en uygun ya da olasılıklı değeri olarak alınmaktadır. Bu nedenle, pratikte çoğu zaman bu yönteme *"En Olasılıklı Parametre Kestirim Yöntemi"* de denmektedir. Parametre kestirim kuramında böyle bir işlem; eğer $x_i = \{x_1, x_2, x_3, \ldots\ldots, x_n\}$ rastgele değişkenler kümesindeki her bir x_i rastgele değişkeni birbirinden bağımsız ve aynı $f(x_i)$ olasılık fonksiyonuna sahip iseler, o zaman bunların bileşik olasılık fonksiyonu; bağımsız rastgele değişkenlerin birleşik olasılık fonksiyonu ile her bir rastgele değişkenin olasılık fonksiyonları arasında var olan,

$$f(x_1, x_2, \ldots. x_n = f(x_1) f(x_2) \ldots\ldots. f(x_n) \qquad 5\text{-}19$$

olasılık bağıntısından faydalanılarak *(Bayes kuralı),*

$$L(x_1, x_2, \ldots\ldots, x_n \,;\, p_1, p_2, \ldots\ldots, p_m) = \prod_{i=1}^{n} f(x_i \,;\, p_1, p_2, \ldots\ldots, p_m) \qquad 5\text{-}20$$

şeklinde ifade edilebilir. Bu şekilde ifade edilmiş olan (5-20) bileşik olasılık fonksiyonu aynı zamanda *Likelihood Fonksiyonu* olarak da isimlendirilmektedir.

Sonuçta, (5-20) bileşik olasılık fonksiyonu; her bir rastgele değişkene ilişkin $f(x_i)$ olasılık fonksiyonlarının ve rastgele değişkenlerden,

$$p_i = g_i(x_1, x_2,......,x_n)$$

şeklinde elde edilen $p_1, p_2,.....,p_m$ kestirim parametreleri olan bilinmeyen parametrelerinin de bir fonksiyonu olmaktadır. Bu özelliklere sahip $p_1, p_2,.....p_m$ parametre değerlerinden (5-20) *Likelihood* fonksiyonunu maksimum yapan herhangi bir $\hat{p}_i$ tahmin değeri de *Maximum Likelihood Kestirim* yöntemi için en uygun kestirim değeri olur. Bu durum, *Likelihood fonksiyonunun,*

$$\frac{\partial L}{\partial p} = 0 \qquad \text{ya da} \qquad \frac{\partial \ln L}{\partial p} = 0 \qquad\qquad 5\text{-}21$$

p parametrelerine göre diferansiyelinin sıfıra eşitlenmesine denk bir ifade olmaktadır.
Buradan da görüldüğü gibi, *En Küçük Kareler Yönteminin* aksine, böyle bir parametre tahmini işlemi için her zaman rastgele değişkenlerin olasılık ya da dağılım fonksiyonlarının bilinmesi gerekmektedir. Bu amaçla, genel teorisi bu şekilde özetlenebilen *Maksimum Likelihood Kestirimi,* biraz daha özelleştirilerek, $x_1, x_2, x_3,.....,x_n$ rastgele değişkenlerden her biri;

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \qquad\qquad 5\text{-}22$$

şeklinde verilen bir olasılık fonksiyonuna sahip normal dağılımlı ve birbirinden bağımsız rastgele değişkenler iseler; bu durumda, bunlara ilişkin birleşik olasılık fonksiyonunu temsil eden *Likelihood fonksiyonu*; (5-22) olasılık fonksiyonlarını (5-20) özelliğine göre işleme sokarak,

$$L = \left[\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(x_1-\mu)^2}{\sigma^2}}\right]\left[\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(x_2-\mu)^2}{\sigma^2}}\right]\,.........\left[\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(x_n-\mu)^2}{\sigma^2}}\right]$$

ya da kısaca,

$$L = \frac{1}{(2\pi\sigma^2)^{n/2}}\, e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}} \qquad\qquad 5\text{-}23$$

şeklinde ifade edilebilir.

Böyle bir fonksiyonun basit bir şekilde diferansiyelini almak amacıyla; sadece normal dağılıma sahip birbirinden bağımsız rastgele değişkenlerin birleşik olasılık fonksiyonu olan (5-23) 'deki *Likelihood fonksiyonu'*nun öncelikle tabii *(Neper)* logaritması alınarak logaritmik olarak ifade edilmiş,

$$\ln L = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \qquad\qquad 5\text{-}24$$

fonksiyonu elde edilir. Daha sonra, bunun μ parametresine göre alınmış birinci dereceden diferansiyelinin,

$$\frac{\partial(\ln L)}{\partial\mu} = \frac{-2}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)(-1) = 0$$

şeklinde sıfıra eşitlenmesinden (5-23) deki bileşik olasılık fonksiyonu
maksimum yapan değer,

$$\sum_{i=1}^{n}(x_i - \mu) = \sum_{i=1}^{n}x_i - n\mu = 0 \qquad\qquad 5\text{-}25$$

ve

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

olarak elde edilir. Buradan, aynı normal dağılıma sahip ve bağımsız rastgele değişkenlerin aritmetik ortalaması *(birinci derece momenti)*, aynı zamanda birleşik olasılık fonksiyonunu *maksimum* yapan μ *Maksimum Likelihood kestirim* değeri olmaktadır. *Maximum Likelihood* kestirim yönteminde bu şekildeki bir parametre kestirimi için yapılan işlemlerde dikkat çeken önemli bir durum; μ kestirim parametresinden başka normal dağılım fonksiyonunu tanımlamada kullanılmış ikinci bir parametre değerini olan σ^2 varyans değerinin yapılan işlemlere dolaylı da olsa katılmamış olmasıdır. Bu amaçla, *Maksimum Likelihood* yöntemine göre yapılmış bu şekildeki bir parametre kestiriminde; σ^2 varyansının durumunun ne olduğunun incelenebilmesi için (5-23) de verilen *Likelihood* fonksiyonunun önce σ^2 varyans değerine göre birinci dereceden diferansiyelinin alınması gerekir. Bu amaçla, (5-23) bağıntısının σ^2 varyans değerine göre alınan diferansiyelinden elde edilmiş birinci dereceden diferansiyel bağıntısının,

$$\frac{\partial(\ln L)}{\partial(\sigma^2)} = -\frac{n\pi}{2\pi\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0 \qquad \text{5-26}$$

şeklinde sıfıra eşitlenerek, daha sonra gerekli ara işlemlerin yapılması neticesinde, bu (5-26) eşitliğinden, *Likelihood* fonksiyonunu maksimum yapan değer için σ^2 varyans değeri de,

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 = n$$

formülünden faydalanılarak,

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \qquad \text{5-27}$$

olarak bulunur.

Neticede; *En Küçük Kareler Yönteminde* temel prensip sadece ölçü hatalarının *(veya düzeltmelerinin)* kareleri toplamını minimum yapılması esas alınmaktadır. Buna karşılık, *Maksimum Likelihood yönteminde, Likelihood fonksiyonun* σ^2 varyans değerine göre değişiminin incelenmesinden elde edilen (5-27) varyans bağıntısından da görüldüğü gibi, *En küçük kareler prensibinin* esası olan; hataların diğer bir deyişle düzeltmelerin kareleri toplamı, bu bağıntıda da sadece

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(\varepsilon_i)^2 = [\varepsilon\varepsilon]$$

şekliyle yer almaktadır. Buna göre de (5-16) bağıntısında bu terime bağımlı olarak ifade edilmiş olan,

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

varyans değerinin *minimum* olması da, neticede *Maksimum Likelihood* kestirimindeki, (5-23) *Likelihood fonksiyonunu maksimum* yapan,

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

ortalama *(veya birinci derece moment)* değeri ile ilgili bir değer olmaktadır. Bu nedenle; *En küçük kareler* yönteminde hataların veya düzeltmelerin karelerinin toplamının minimum yapılması ile sonuçta hatalara veya diğer bir deyişle düzeltmelere ilişkin olasılık fonksiyonunu ortalama değerlerine göre maksimumlaştırmaktadır. İlgili bağıntıların incelenmesinden de görüleceği gibi uygulamada bu durum, *Maksimum Likelihood Kestirimine* eşdeğer bir sonuç olmaktadır.

Sonuçta buradan özet olarak söylenebilir ki, verilerin normal dağılımda olmaları halinde *En küçük kareler* kestirim yöntemi ile *Maksimum Likelihood Kestirim* yöntemi parametre kestirimindeki yaklaşımları yönünden eşdeğer neticeler vermektedir.

5.2.4. En Küçük Kareler Kestirim Yöntemi ve İstatistik

Bir büyüklüğe ait gerçek *(fiziksel ya da doğal*) değer, doğadaki şekliyle, hiçbir zaman tam ya da mükemmel şekliyle ölçülemez ya da belirlenemez. Çünkü, matematik modellendirilmelerden, ortamdan ve kişiden kaynaklanan birçok hatadan veya kullanılan aletlerin sınırlı duyarlıkta ölçü yapmalarından dolayı, ölçüler *(gözlemler)* daima sınırlı presizyonda yapılırlar.

Uygulamada bu gibi sınırlamalardan dolayı ölçüler, daima çok küçük miktarlarda olsalar bile, sürekli hatalı olacaklarından, tek anlamlı çözüm için gerekli olan sayıdan daha fazla sayıda yapıldıkları sürece ölçme duyarlığına bağlı olarak birbirleriyle aynı değerlerde olmazlar. Bunların aralarında az da olsa farklılıklar olur.

Örneğin, herhangi bir uzunluk milimetre incelikte ölçü yapabilen bir metre ile birkaç kez ölçülmüş olsun. Eğer her defasında yapılan ölçüler santimetre incelikte bir kenara yazılırsa bunların seçilen ölçü birimi cinsinden tutarlı *(consistent)* oldukları görülür. Bunun nedeni, bu şekilde elde edilen mevcut değerlerin ölçü aletinin ölçme duyarlığından daha kaba olması nedeniyle aralarındaki uyuşumsuzluklar da seçilen birim içinde saklandığından fark olunamazlar. Bu ölçüler seçilen birim cinsinden daima tutarlı olurlar. Buna karşılık, bu amaçla yapılmış olan uzunluk ölçüleri milimetre veya daha presizyonlu ölçü birimlerine göre karşılaştırıldıklarında, aralarındaki çelişkilerin veya uyuşumsuzlukların *(inconsistent)* varlığı açıkça görülebilir.

Böyle bir durumda, sözü edilen uzunluğun bu presizyondaki gerçek değeri de bilinemez. O zaman, bu uzunluğun doğal veya gerçek değeri yerine kullanılabilecek ve onu en iyi temsil eden tek anlamlı en muhtemel *(kesin, tahmin* veya *kestirim)* değeri belirlenmek istenir. Bu gibi isteklere cevap vermek için böyle bir değerin; "ne gibi özelliklere sahip olması ve nasıl hesaplanması veya elde edilmesi gerektiği?" sorusu ile karşılaşılır. Günümüzde, bu gibi uyuşumsuz *(inconsistent)* ölçü değerlerini irdelemek için kullanılan ilmi yöntem *istatistik* olarak adlandırılmaktadır. Bu özellikteki ölçü değerlerinden doğal büyüklüğün gerçek değeri yerine kullanılabilecek tek anlamlı en muhtemel *(kesin)* değerin belirlenmesi işlemine *istatistik tahmin* veya *kestirim (statistical estimation)* denir. En küçük kareler yöntemi de bu amaçla kullanılan yöntemlerden biri olup; "*tüm ölçülerdeki tutarsızlıkların kareleri toplamının minimum olması"* prensibine dayanır. Gözlemlerin kaba ve sistematik hatalardan ayıklandığı, sadece rastgele değişken özelliği gösterdikleri, ayrıca bunların doğrusal matematik modellerde *En küçük kareler* kestirim yöntemi kullanılarak;

- *Gözlem sayısı* $n \to \infty$ *olması durumunda,* $\hat{x}$ *kestirim değeri ve* μ_x *onun dağılımının parametresi ve* $\varepsilon \rangle 0$ *olmak üzere;* $\lim\limits_{n \to \infty} P(|\mu_x - \hat{x}| \langle \varepsilon) = 1$ *tutarlı (consistence) olması,*
- $E\{\hat{x}\} = \mu_x$*, Yansız (unbiased) olması,*
- *Minimum varyanslı olması,*
- *Etkin olması*

özelliklerini içeren bir kestirim değeri elde edilmiş olur. Bilindiği gibi istatistikte aynı amaca yönelik *(yani tek anlamlı parametre kestirim değeri belirlemede)* kullanılan başka birçok yöntemler daha mevcuttur. Bu yöntemlere bir örnek olarak, Tutarsızlıkların *(hataların)* mutlak değerlerinin toplamını minimum veya maksimum uyuşumsuzluğun minimum olması ilkesini esas alan yöntemler verilebilir. Bunlardan, ölçüler arasındaki tutarsızlıkların mutlak değerlerinin minimum yapılması prensibini içeren çözüm, gözlemlerin gerçek değeri ile parametreler arasında kurulan denklemlerde, rastgele ölçü hatalarını da içeren ölçü ya da gözlem değerleri kullanıldığında eşitlik sağlanamaz. Gerçek değerlere göre kurulmuş olan eşitlikler gözlem değerlerinin kullanılması ile eşitsizlikler haline dönüşmüş olur. Uygulamada, bu tür eşitsizliklerden parametrelerin kesin değerlerini hesaplayabilmek için, her bir eşitsizlik ifadesine aylak ve yapay değişkenler ilave edilip *simplex* çözüm yöntemleri kullanılmak suretiyle istenen sonuçlara ulaşılabilinir. Böyle bir işlem sonucunda, bilinmeyen parametreler için arzulanan bir çözüme ancak kavuşulmuş olunur.

Ancak, ne var ki, bu şekilde yapılacak parametre kestiriminin, çeşitli işlem güçlükleri yanında, tenkit edilebilir bir diğer yönü; *simplex* çözümler neticesinde elde edilecek sonuçların istatistik olarak özel bir irdelemelerinin olmayışıdır. Bu yöntemlerden bir diğeri olan *"Maksimum uyuşumsuzluğun minimum çözümün*"de ise; parametre kestirimine esas olacak gözlemlerin dağılımlarının bilinmesi gerekir. Bu dağılıma uymayan, uyuşumsuz gözlemlerin parametre kestirimi üzerindeki olumsuz etkileri, belli bir amaç fonksiyonuna göre tanımlanan etki ve ağırlık fonksiyonları kullanılarak kademeli olarak giderilir. *Robust kestirimi* olarak da isimlendirilen bu yöntem; kırılma noktasının yüksek olması nedeniyle son zamanlarda oldukça kullanılmaktadır. Ancak, bunun tenkit edilir yanı; *iteratif* çözüm olması yanında, her zaman ölçü dağılımlarının bilinmesi ve sonuçlarının istatistik yönden özel bir irdeleme tekniğine sahip olmamasıdır.

Buradan görüldüğü gibi; *En Küçük Kareler Kestirim yöntemine* alternatif olarak verilen bu yöntemlerin yukarıdaki dört kestirim özelliğini birden içermemeleri dahil, çözüm ve sonuçların irdelenmesindeki zorluklar gibi, başlıca bir çok tenkit edilir yönleri bulunmaktadır. Bu nedenle, yukarıda sözü edilen alternatif kestirim yöntemlerine oranla En küçük kareler yönteminin en iyi yanı;

- *Fonksiyonel modeller ister lineer olsun, isterse non-lineer olsun her durumda uygulanabilir modellerdir,*
- *Dağılımın istatistik parametreleri, sadece median (ortalama) ve standart sapmalardır,*
- *En küçük kareler yönteminde parametre kestirimi için ölçülerin dağılımının bilinmesine hiçbir zaman gerek yoktur,*
- *Bu yöntemin kestirim sonuçlarının istatistik yönden analiz edilmeleri daima olanaklıdır.*

Bu olumlu yönlere karşılık, *En Küçük Kareler Yönteminin* en olumsuz tarafı, kırılma noktasının *(break down)*; *1/n (burada n : ölçü sayısı)* küçük olması nedeniyle herhangi bir ölçüdeki uyuşumsuzluğun parametre kestirim sonuçlarını çok fazla etkileyecek olmasıdır. Bu özellik, *En Küçük Kareler Yönteminin* diğer kestirim yöntemlerine oranla birçok avantajının yanında bir dezavantajı olmaktadır. Kimilerine göre; istatistik rastgele değişkenlerin fonksiyonları teorisidir. Böyle bir rastgele değişken de sadece uygun değerlerden oluşan bir örneklemeyi özet olarak tanımlayan değişkendir. Bu durumda, herhangi bir olayla ilgili bir rastgele değişken, hem uygun örnekleme sonuçları, hem de bu değerlerin hangi sıklıkta meydana geldiklerini ifade eden yoğunluk *(olasılık)* fonksiyonu ile tanımlanır. Bilindiği gibi bu fonksiyonlar içerisinde, jeodezik gözlemler için en önemli olanı *normal dağılım fonksiyonu* diğer adıyla; *Gauss dağılım fonksiyonudur*.

Matematik-istatistikte; fiziksel gözlemler daima normal dağılıma sahip rastgele değişkenler oldukları kabul edilir. Yine matematik istatistiğin diğer bir yönüne göre, doğal bir olayla ilgili fiziksel büyüklüğün örnekleme değerlerinden tek anlamlı olarak tahmin edilen kestirim değeri, onun gerçek değerine ne derece yakın olduğu ancak bu olayla ilgili tanımlanan rastgele değişkenin yoğunluk fonksiyonunun bilinip, bilinmemesi durumuna göre istatistik olarak açıklanabilir. Bu durum, istatistik olarak sonuçları irdelenecek olan rastgele değişkenlerin dağılım fonksiyonlarının her zaman bilinmesini zorunlu kılmaktadır. Diğer bir ifade ile, bir kestirim değeri için böyle bir fonksiyonun bilinmemesi halinde, sonuçların *matematik-istatistik* yasalarına göre irdelenmesine ışık tutan, güven aralığı tahmini ve hipotez testleri de hiçbir zaman yapılamaz.

Neticede, *Rabust* ve *Bayes* kestirimleri böyle bir özelliğe sahip tahmin türleri olmadığı için, bunların sonuçları matematik istatistik yasalarına göre irdelenemez.

Bu nedenle, bu kestirim türleri konunun akışı içerisinde göz ardı edilerek, burada sadece sonuçları istatistik olarak irdelenebilen *En küçük kareler* ve normal dağılımlı rastgele değişkenler için benzer karaktere sahip *Maximum Likelihood* kestirim yöntemleri üzerinde konu yoğunlaştırılmıştır. Bu amaçla, burada sadece her iki yöntemlerle ilgili bazı temel kurallar ele alınarak açıklanmıştır.