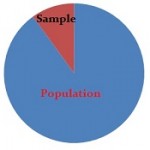
What is a Sufficient Statistic?

[](https://www.statisticshowto.datasciencecentral.com/wp-content/uploads/2015/07/sample-mean-small.jpg)A sufficient statistic is a statistic that summarizes all of the information in a sample about a chosen [parameter](https://www.statisticshowto.datasciencecentral.com/what-is-a-parameter-statisticshowto/). For example, the [sample mean](https://www.statisticshowto.datasciencecentral.com/sample-mean/), x̄, estimates the [population mean](https://www.statisticshowto.datasciencecentral.com/population-mean/), μ. x̄ is a sufficient statistic if it retains all of the information about the population mean that was contained in the original data points.

According to statistician Ronald Fisher,

*“…no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter.”*

In layman’s terms, a sufficient statistic is your best bet for summarizing your data; You can use it even if you don’t know any of the actual values in the sample.

Sufficient Statistic Example

You can think of a sufficient statistic as an [estimator](https://www.statisticshowto.datasciencecentral.com/estimator/)that allows you to estimate the population parameter as well as if you knew all of the data in all possible samples.

For example, let’s say you have the simple data set 1,2,3,4,5. You would calculate the sample mean as (1 + 2 + 3 + 4 + 5) / 5 = 3, which gives you the estimate of the population mean as 3. Let’s assume you don’t know those values (1, 2, 3, 4, 5), but you *only* know that the sample mean is 3. You would also estimate the population mean as 3, which would be just as good as knowing the whole data set. The sample mean of 3 is a sufficient statistic. To put this another way, **if you have the sample mean, then knowing all of the data items makes no difference in how good your estimate is**: it’s already “the best”.

[Order statistics](https://www.statisticshowto.datasciencecentral.com/order-statistics/) for *[iid](https://www.statisticshowto.datasciencecentral.com/iid-statistics/)* samples are also sufficient statistics. This does not hold for data that isn’t *iid* because only in these samples, can you re-order the data without losing meaning.

When collecting data, **the sufficiency principle justifies ignoring certain pieces of information** (Steel, 2007). For example, let’s say you were conducting an experiment recording the number of heads in a coin toss. You could record the number of heads and tails, along with their order: HTTHTTTHHH…. Or, you could just record the number of heads (e.g. 25 heads). For the purposes of a [binomial experiment](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/binomial-theorem/binomial-experiment/), the number of heads would be a sufficient statistic. Recording all of the tails, and their order, would give you no more information (assuming the variables are [independent and identically distributed](https://www.statisticshowto.datasciencecentral.com/iid-statistics/)).

Viewing the Sufficiency Principle as Data Reduction

We know from the sufficiency principle that if we have a sufficient statistic Y = T (X) and a statistical model, the inferences we can make about θ from our model and X (the data set) must be the same as from that model and Y.

This makes sufficiency a very strong property; a way of data reduction, or condensing all the important information in our sample into the statistic.

The Sufficiency Principle

Let’s consider for moment Y, a sufficient statistic, and X, a set of observations. We want to look at the pair (X, Y). Since Y is dependent on X, the pair (X, Y) will give us the same information about parameter θ that X does.

But since Y is sufficient, the conditional distribution of X given Y is independent of θ.

What exactly does that mean?

Let X be your last statistics lecture and the video recording of it, and Y be the notes you took about it. Parameter θ is information needed on question #7 in your class final. Y depends entirely on X, and the video *and* class notes includes exactly the same information as just the video did; nothing added. But if you’ve taken sufficient notes, the conditional distribution of the lecture given your notes is independent of that question #7 information. Conditional distribution here just means the probability distribution of the info in your notes, given the lecture. If the info is in the lecture, it’s in your notes. Once you’ve checked your memorized notes, going back and listening to the lecture won’t help you solve question #7.

Now let’s go back to the generic sufficiency principle and mathematical statistics. A focus on both bits of information (data set X and statistic Y) does not give us any more information about the distribution of θ than we’d have if we only focused on the statistic. And after looking at statistic Y, a look at X doesn’t give us any new information on θ it won’t enlighten us on whether a particular value of θ is more likely or less likely then another.

So if Y is a sufficient statistic, we don’t need to consider data set X any more, after using it to calculate Y; it becomes redundant.

**Sufficient Statistics**

Introduction

In the lesson on Point Estimation, we derived estimators of various parameters using two methods, namely, the method of maximum likelihood and the method of moments. The estimators resulting from these two methods are typically intuitive estimators. It makes sense, for example, that we would want to use the sample mean  and sample variance *S*2 to estimate the mean *μ* and variance *σ*2 of a normal population.

In the process of estimating such a parameter, we summarize, or reduce, the information in a sample of size *n*, *X*1, *X*2, ..., *Xn*, to a single number, such as the sample mean  The actual sample values are no longer important to us. That is, if we use a sample mean of 3 to estimate the population mean *μ*, it doesn't matter if the original data values were (1, 3, 5) or (2, 3, 4). Has this process of reducing the *n* data points to a single number retained all of the information about *μ* that was contained in the original *n* data points? Or has some information about the parameter been lost through the process of summarizing the data? In this lesson, we'll learn how to find statistics that summarize all of the information in a sample about the desired parameter. Such statistics are called **sufficient statistics**,

**Definition of Sufficiency**

|  |
| --- |
| **Definition.**L Let *X*1, *X*2, ..., *Xn* be a random sample from a probability distribution with unknown parameter *θ*. Then, the statistic:  Y=u(X1,X2,...,Xn)Y=u(X1,X2,...,Xn)  is said to be **sufficient** for *θ* if the conditional distribution of *X*1, *X*2, ..., *Xn*, given the statistic *Y*,  does not depend on the parameter *θ*. |

Example

Let *X*1, *X*2, ..., *Xn* be a random sample of *n* Bernoulli trials in which:

* *Xi*= 1 if the *ith* subject likes Pepsi
* *Xi* = 0 if the *ith* subject does not like Pepsi

If *p* is the probability that subject *i*likes Pepsi, for *i* = 1, 2, ..., *n*, then:

* *Xi* = 1 with probability *p*
* *Xi* = 0 with probability *q* = 1 – *p*

Suppose, in a random sample of *n* = 40 people, that  people like Pepsi. If we know the value of *Y*, the number of successes in *n* trials, can we gain any further information about the parameter *p* by considering other functions of the data *X*1, *X*2, ..., *Xn*? That is, is *Y* sufficient for *p*?

**Solution.** The definition of sufficiency tells us that if the conditional distribution of *X*1, *X*2, ..., *Xn*, given the statistic *Y*, does not depend on *p*, then *Y* is a sufficient statistic for *p*. The conditional distribution of *X*1, *X*2, ..., *Xn*, given *Y*, is by definition:

P(X1=x1,...,Xn=xn|Y=y)=P(X1=x1,...,Xn=xn,Y=y)P(Y=y)P(X1=x1,...,Xn=xn|Y=y)=P(X1=x1,...,Xn=xn,Y=y)P(Y=y)      (\*\*)

Now, for the sake of concreteness, suppose we were to observe a random sample of size *n* = 3 in which *x*1 = 1, *x*2 = 0, and *x*3 = 1. In this case:

P(X1=1,X2=0,X3=1,Y=1)=0P(X1=1,X2=0,X3=1,Y=1)=0

because the sum of the data values, ∑ni=1Xi∑i=1nXi, is 1 + 0 + 1 = 2, but *Y*, which is defined to be the sum of the *Xi*'s is  1. That is, because 2 ≠ 1, the event in the numerator of the starred (\*\*) equation is an impossible event and therefore its probability is 0.

Now, let's consider an event that is possible, namely (*X*1=1, *X*2 = 0, *X*3 = 1, *Y* = 2). In that case, we have, by independence:

P(X1=1,X2=0,X3=1,Y=2)=p(1−p)p=p2(1−p)P(X1=1,X2=0,X3=1,Y=2)=p(1−p)p=p2(1−p)

So, in general:

P(X1=x1,X2=x2,...,Xn=xn,Y=y)=0 if n∑i=1xi≠yP(X1=x1,X2=x2,...,Xn=xn,Y=y)=0 if ∑i=1nxi≠y

and:

P(X1=x1,X2=x2,...,Xn=xn,Y=y)=py(1−p)n−y if n∑i=1xi=yP(X1=x1,X2=x2,...,Xn=xn,Y=y)=py(1−p)n−y if ∑i=1nxi=y

Now, the denominator in the starred (\*\*) equation above is the binomial probability of getting exactly *y* successes in *n* trials with a probability of success *p*. That is, the denominator is:

P(Y=y)=(ny)py(1−p)n−yP(Y=y)=(ny)py(1−p)n−y

for *y* = 0, 1, 2, ..., *n*. Putting the numerator and denominator together, we get, if *y* = 0, 1, 2, ..., *n*, that the conditional probability is:

P(X1=x1,...,Xn=xn|Y=y)=py(1−p)n−y(ny)py(1−p)n−y=1(ny) if n∑i=1xi=yP(X1=x1,...,Xn=xn|Y=y)=py(1−p)n−y(ny)py(1−p)n−y=1(ny) if ∑i=1nxi=y

and:

P(X1=x1,...,Xn=xn|Y=y)=0 if n∑i=1xi≠yP(X1=x1,...,Xn=xn|Y=y)=0 if ∑i=1nxi≠y

Aha! We have just shown that the conditional distribution of *X*1, *X*2, ..., *Xn* given *Y* does not depend on *p*. Therefore, *Y* is indeed sufficient for *p*. That is, once the value of *Y* is known, no other function of *X*1, *X*2, ..., *Xn* will provide any additional information about the possible value of *p*.

**Factorization Theorem**

It is not always all that easy to find the conditional distribution of *X*1, *X*2, ..., *Xn* given *Y*.

Therefore, using the formal definition of sufficiency as a way of identifying a sufficient statistic for a parameter *θ* can be a difficult task. Factorization Theorem provides an easier alternative.

|  |
| --- |
| **Factorization Theorem.**  Let *X*1, *X*2, ..., *Xn* denote random variables with joint probability density function or joint probability mass function *f*(*x*1, *x*2, ..., *xn*;*θ*), which depends on the parameter *θ*. Then, the statistic Y=u(X1,X2,...,Xn)Y=u(X1,X2,...,Xn) is sufficient for*θ* if and only if the p.d.f (or p.m.f.) can be factored into two components, that is:  f(x1,x2,...,xn;θ)=ϕ[u(x1,x2,...,xn);θ]h(x1,x2,...,xn)f(x1,x2,...,xn;θ)=ϕ[u(x1,x2,...,xn);θ]h(x1,x2,...,xn)  where:   * *φ* is a function that depends on the data *x*1, *x*2, ..., *xn* only through the function *u*(*x*1, *x*2,..., *xn*), and * the function *h*(*x*1, *x*2, ..., *xn*) does not depend on the parameter *θ* |

**Example: Poisson**

Let *X*1, *X*2, ..., *Xn* denote a random sample from a Poisson distribution with parameter *λ* > 0. Find a sufficient statistic for the parameter *λ*.

**Solution.** Because *X*1, *X*2, ..., *Xn* is a random sample, the joint probability mass function of *X*1, *X*2, ..., *Xn* is, by independence:

f(x1,x2,...,xn;λ)=f(x1;λ)×f(x2;λ)×...×f(xn;λ)f(x1,x2,...,xn;λ)=f(x1;λ)×f(x2;λ)×...×f(xn;λ)

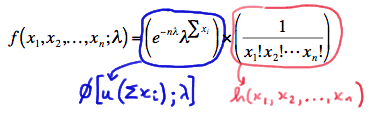
Inserting what we know to be the probability mass function of a Poisson random variable with parameter *λ*, the joint p.m.f. is therefore:

f(x1,x2,...,xn;λ)=e−λλx1x1!×e−λλx2x2!×...×e−λλxnxn!f(x1,x2,...,xn;λ)=e−λλx1x1!×e−λλx2x2!×...×e−λλxnxn!

Now, simpliyfing, by adding up all *n* of the *λ*s in the exponents, as well as all *n*of the *xi*'s in the exponents, we get:

f(x1,x2,...,xn;λ)=(e−nλλΣxi)×(1x1!x2!...xn!)f(x1,x2,...,xn;λ)=(e−nλλΣxi)×(1x1!x2!...xn!)

Hey, look at that! We just factored the joint p.m.f. into two functions, one (***φ***) being only a function of the statistic  Y=∑ni=1XiY=∑i=1nXi and the other (***h***) not depending on the parameter *λ*:



Therefore, the Factorization Theorem tells us that Y=∑ni=1XiY=∑i=1nXi is a sufficient statistic for *λ*. But, wait a second! We can also write the joint p.m.f. as:

f(x1,x2,...,xn;λ)=(e−nλλn¯x)×(1x1!x2!...xn!)f(x1,x2,...,xn;λ)=(e−nλλnx¯)×(1x1!x2!...xn!)

Therefore, the Factorization Theorem tells us that Y=¯XY=X¯ is also a sufficient statistic for *λ*!

If you think about it, it makes sense that Y=¯XY=X¯ and Y=∑ni=1XiY=∑i=1nXi are both sufficient statistics, because if we know Y=¯XY=X¯, we can easily find Y=∑ni=1XiY=∑i=1nXi. And, if we know Y=∑ni=1XiY=∑i=1nXi, we can easily find Y=¯XY=X¯.

The previous example suggests that there can be more than one sufficient statistic for a parameter *θ*. In general, if *Y* is a sufficient statistic for a parameter *θ*, then every one-to-one function of *Y* not involving *θ* is also a sufficient statistic for *θ*.

**Example: Normal**

Let *X*1, *X*2, ..., *Xn* be a random sample from a normal distribution with mean *μ* and variance 1. Find a sufficient statistic for the parameter *μ*.

**Solution.** Because *X*1, *X*2, ..., *Xn* is a random sample, the joint probability density function of *X*1, *X*2, ..., *Xn* is, by independence:

f(x1,x2,...,xn;μ)=f(x1;μ)×f(x2;μ)×...×f(xn;μ)f(x1,x2,...,xn;μ)=f(x1;μ)×f(x2;μ)×...×f(xn;μ)

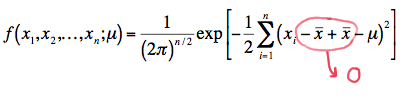
Inserting what we know to be the probability density function of a normal random variable with mean *μ* and variance 1, the joint p.d.f. is:

f(x1,x2,...,xn;μ)=1(2π)1/2exp[−12(x1−μ)2]×1(2π)1/2exp[−12(x2−μ)2]×...×1(2π)1/2exp[−12(xn−μ)2]f(x1,x2,...,xn;μ)=1(2π)1/2exp[−12(x1−μ)2]×1(2π)1/2exp[−12(x2−μ)2]×...×1(2π)1/2exp[−12(xn−μ)2]

Collecting like terms, we get:

f(x1,x2,...,xn;μ)=1(2π)n/2exp[−12n∑i=1(xi−μ)2]f(x1,x2,...,xn;μ)=1(2π)n/2exp[−12∑i=1n(xi−μ)2]

A trick to making the factoring of the joint p.d.f. an easier task is to add 0 to the quantity in parentheses in the summation. That is:



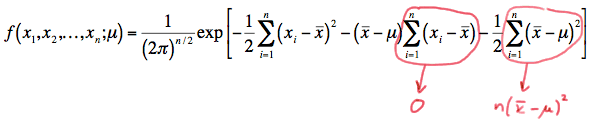
Now, squaring the quantity in parentheses, we get:

f(x1,x2,...,xn;μ)=1(2π)n/2exp[−12n∑i=1[(xi−¯x)2+2(xi−¯x)(¯x−μ)+(¯x−μ)2]]f(x1,x2,...,xn;μ)=1(2π)n/2exp[−12∑i=1n[(xi−x¯)2+2(xi−x¯)(x¯−μ)+(x¯−μ)2]]

And then distributing the summation, we get:

f(x1,x2,...,xn;μ)=1(2π)n/2exp[−12n∑i=1(xi−¯x)2−(¯x−μ)n∑i=1(xi−¯x)−12n∑i=1(¯x−μ)2]f(x1,x2,...,xn;μ)=1(2π)n/2exp[−12∑i=1n(xi−x¯)2−(x¯−μ)∑i=1n(xi−x¯)−12∑i=1n(x¯−μ)2]

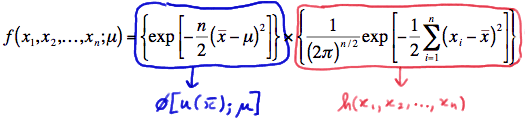
But, the middle term in the exponent is 0, and the last term, because it doesn't depend on the index *i*, can be added up *n* times:



So, simplifying, we get:

f(x1,x2,...,xn;μ)={exp[−n2(¯x−μ)2]}×{1(2π)n/2exp[−12n∑i=1(xi−¯x)2]}f(x1,x2,...,xn;μ)={exp[−n2(x¯−μ)2]}×{1(2π)n/2exp[−12∑i=1n(xi−x¯)2]}

In summary, we have factored the joint p.d.f. into two functions, one (***φ***) being only a function of the statistic Y=¯XY=X¯ and the other (***h***) not depending on the parameter*μ*:



Therefore, the Factorization Theorem tells us that Y=¯XY=X¯ is a sufficient statistic for*μ*. Now, Y=¯X3Y=X¯3 is also sufficient for *μ*, because if we are given the value of ¯X3X¯3, we can easily get the value of ¯XX¯ through the one-to-one function w=y1/3w=y1/3. That is:

W=(¯X3)1/3=¯XW=(X¯3)1/3=X¯

On the other hand, Y=¯X2Y=X¯2 is not a sufficient statistic for *μ*, because it is not a one-to-one function. That is, if we are given the value of ¯X2X¯2, using the inverse function:

w=y1/2w=y1/2

we get two possible values, namely:

−¯X−X¯  and +¯X+X¯

**Example : Exponential**

Let *X*1, *X*2, ..., *Xn* be a random sample from an exponential distribution with parameter *θ*. Find a sufficient statistic for the parameter *θ*.

**Solution.** Because *X*1, *X*2, ..., *Xn* is a random sample, the joint probability density function of *X*1, *X*2, ..., *Xn* is, by independence:

f(x1,x2,...,xn;θ)=f(x1;θ)×f(x2;θ)×...×f(xn;θ)f(x1,x2,...,xn;θ)=f(x1;θ)×f(x2;θ)×...×f(xn;θ)

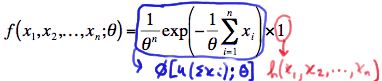
Inserting what we know to be the probability density function of an exponential random variable with parameter *θ*, the joint p.d.f. is:

f(x1,x2,...,xn;θ)=1θexp(−x1θ)×1θexp(−x2θ)×...×1θexp(−xnθ)f(x1,x2,...,xn;θ)=1θexp(−x1θ)×1θexp(−x2θ)×...×1θexp(−xnθ)

Now, simpliyfing, by adding up all *n* of the *θ*s and the *n* *xi*'s in the exponents, we get:

f(x1,x2,...,xn;θ)=1θnexp(−1θn∑i=1xi)f(x1,x2,...,xn;θ)=1θnexp(−1θ∑i=1nxi)

We have again factored the joint p.d.f. into two functions, one (***φ***) being only a function of the statistic Y=∑ni=1XiY=∑i=1nXi and the other (***h***) not depending on the parameter*θ*:



Therefore, the Factorization Theorem tells us that Y=∑ni=1XiY=∑i=1nXi is a sufficient statistic for *θ*. And, since Y=¯XY=X¯ is a one-to-one function of Y=∑ni=1XiY=∑i=1nXi, it implies that Y=¯XY=X¯ is also a sufficient statistic for *θ*.

Two or More Parameters

In each of the examples we considered so far in this lesson, there is one and only one parameter. What happens if a probability distribution has two parameters, *θ*1 and *θ*2, say, for which we want to find sufficient statistics, *Y*1 and *Y*2? Fortunately, the definitions of sufficiency can easily be extended to accommodate two (or more) parameters. Let's start by extending the Factorization Theorem.

|  |
| --- |
| **Definition (Factorization Theorem).** Let *X*1, *X*2, ..., *Xn* denote random variables with a joint p.d.f. (or joint p.m.f.):  f(x1,x2,...,xn;θ1,θ2)f(x1,x2,...,xn;θ1,θ2)  which depends on the parameters *θ*1 and *θ*2. Then, the statistics Y1=u1(X1,X2,...,Xn)Y1=u1(X1,X2,...,Xn) and Y2=u2(X1,X2,...,Xn)Y2=u2(X1,X2,...,Xn) are **joint sufficient statistics** for *θ*1 and *θ*2 if and only if:  f(x1,x2,...,xn;θ1,θ2)=ϕ[u1(x1,...,xn),u2(x1,...,xn);θ1,θ2]h(x1,...,xn)f(x1,x2,...,xn;θ1,θ2)=ϕ[u1(x1,...,xn),u2(x1,...,xn);θ1,θ2]h(x1,...,xn)  where:   * ϕϕ is a function that depends on the data  (x1,x2,...,xn)(x1,x2,...,xn) only through the functions u1(x1,x2,...,xn)u1(x1,x2,...,xn) and u2(x1,x2,...,xn)u2(x1,x2,...,xn), and * the function h(x1,...,xn)h(x1,...,xn) does not depend on either of the parameters *θ*1 or *θ*2. |

Example

Let *X*1, *X*2,..., *Xn* denote a random sample from a normal distribution *N*(*θ*1,*θ*2). That is, *θ*1 denotes the mean *μ* and *θ*2 denotes the variance *σ*2. Use the Factorization Theorem to find joint sufficient statistics for *θ*1 and *θ*2.

**Solution.** Because *X*1, *X*2, ..., *Xn* is a random sample, the joint probability density function of *X*1, *X*2, ..., *Xn* is, by independence:

f(x1,x2,...,xn;θ1,θ2)=f(x1;θ1,θ2)×f(x2;θ1,θ2)×...×f(xn;θ1,θ2)×f(x1,x2,...,xn;θ1,θ2)=f(x1;θ1,θ2)×f(x2;θ1,θ2)×...×f(xn;θ1,θ2)×

Inserting what we know to be the probability density function of a normal random variable with mean *θ*1and variance *θ*2, the joint p.d.f. is:

f(x1,x2,...,xn;θ1,θ2)=1√2πθ2exp[−12(x1−θ1)2θ2]×...×=1√2πθ2exp[−12(xn−θ1)2θ2]f(x1,x2,...,xn;θ1,θ2)=12πθ2exp[−12(x1−θ1)2θ2]×...×=12πθ2exp[−12(xn−θ1)2θ2]

Simplifying by collecting like terms, we get:

f(x1,x2,...,xn;θ1,θ2)=(1√2πθ2)nexp[−12∑ni=1(xi−θ1)2θ2]f(x1,x2,...,xn;θ1,θ2)=(12πθ2)nexp[−12∑i=1n(xi−θ1)2θ2]

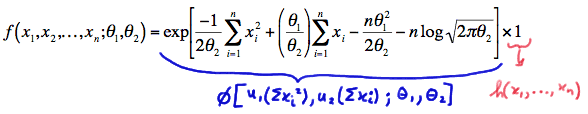
Rewriting the first factor, and squaring the quantity in parentheses, and distributing the summation, in the second factor, we get:

f(x1,x2,...,xn;θ1,θ2)=exp[log(1√2πθ2)n]exp[−12θ2{n∑i=1x2i−2θ1n∑i=1xi+n∑i=1θ21}]f(x1,x2,...,xn;θ1,θ2)=exp[log(12πθ2)n]exp[−12θ2{∑i=1nxi2−2θ1∑i=1nxi+∑i=1nθ12}]

Simplifying yet more, we get:

f(x1,x2,...,xn;θ1,θ2)=exp[−12θ2n∑i=1x2i+θ1θ2n∑i=1xi−nθ212θ2−nlog√2πθ2]f(x1,x2,...,xn;θ1,θ2)=exp[−12θ2∑i=1nxi2+θ1θ2∑i=1nxi−nθ122θ2−nlog2πθ2]

Look at that! We have factored the joint p.d.f. into two functions, one (***φ***) being only a function of the statistics Y1=∑ni=1X2iY1=∑i=1nXi2 and Y2=∑ni=1XiY2=∑i=1nXi, and the other (***h***) not depending on the parameters*θ*1 and*θ*2:



Therefore, the Factorization Theorem tells us that Y1=∑ni=1X2iY1=∑i=1nXi2 and Y2=∑ni=1XiY2=∑i=1nXi are joint sufficient statistics for*θ*1 and*θ*2. And, the one-to-one functions of *Y*1 and *Y*2, namely:

¯X=Y2n=1nn∑i=1XiX¯=Y2n=1n∑i=1nXi

and

S2=Y1−(Y22/n)n−1=1n−1[n∑i=1X2i−n¯X2]S2=Y1−(Y22/n)n−1=1n−1[∑i=1nXi2−nX¯2]

are also joint sufficient statistics for *θ*1 and*θ*2. Aha! We have just shown that the intuitive estimators of *μ* and *σ*2 are also sufficient estimators. That is, the data contain no more information than the estimators ¯XX¯ and *S*2 do about the parameters *μ* and *σ*2! That seems like a good thing!

We have just extended the Factorization Theorem. Now, the Exponential Criterion can also be extended to accommodate two (or more) parameters. It is stated here without proof.

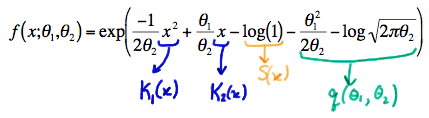
|  |
| --- |
| **Exponential Criterion.** Let *X*1, *X*2, ..., *Xn* be a random sample from a distribution with a p.d.f. or p.m.f. of the exponential form:  f(x;θ1,θ2)=exp[K1(x)p1(θ1,θ2)+K2(x)p2(θ1,θ2)+S(x)+q(θ1,θ2)]f(x;θ1,θ2)=exp[K1(x)p1(θ1,θ2)+K2(x)p2(θ1,θ2)+S(x)+q(θ1,θ2)]  with a support that does not depend on the parameters *θ*1 and *θ*2. Then, the statistics Y1=∑ni=1K1(Xi)Y1=∑i=1nK1(Xi) and Y2=∑ni=1K2(Xi)Y2=∑i=1nK2(Xi) are jointly sufficient for *θ*1 and *θ*2. |

Let's try applying the extended exponential criterion to our previous example.

Example (continued)

Let *X*1, *X*2,..., *Xn* denote a random sample from a normal distribution *N*(*θ*1,*θ*2). That is, *θ*1 denotes the mean *μ* and*θ*2 denotes the variance *σ*2. Use the Exponential Criterion to find joint sufficient statistics for *θ*1 and*θ*2.

**Solution.**The probability density function of a normal random variable with mean *θ*1 and variance*θ*2 can be written in exponential form as:



Therefore, the statistics  and  are joint sufficient statistics for*θ*1 and*θ*2