5. Completeness and sufficiency

5.1. Complete statistics.

Definition 5.1. A statistic T is called *complete* if $Eg(T) = 0$ for all θ and some function g implies that $P(g(T) = 0; \theta) = 1$ for all θ.

This use of the word *complete* is analogous to calling a set of vectors $v_1, \ldots, v_n$ complete if they span the whole space, that is, any v can be written as a linear combination $v = \sum a_j v_j$ of these vectors. This is equivalent to the condition that if w is orthogonal to all v_j's, then $w = 0$. To make the connection with Definition 5.1, let's consider the discrete case. Then completeness means that $\sum g(t)P(T = t; \theta) = 0$ implies that $g(t) = 0$ for all possible values t of T. Since the sum may be viewed as the scalar product of the vectors $(g(t_1), g(t_2), \ldots)$ and $(p(t_1), p(t_2), \ldots)$, with $p(t) = P(T = t)$, this is the analog of the orthogonality condition just discussed.

We also see that the terminology is somewhat misleading. It would be more accurate to call the family of distributions $p(\cdot; \theta)$ complete (rather than the statistic T). In any event, completeness means that the collection of distributions for all possible values of θ provides a sufficiently rich set of vectors. In the continuous case, a similar interpretation works. Completeness now refers to the collection of densities $f(\cdot; \theta)$, and $\langle f, g \rangle = \int fg$ serves as the (abstract) scalar product in this case.

Example 5.1. Let's take one more look at the coin flip example. I claim that $T = X_1 + \ldots + X_n$ is a complete statistic. To check this, suppose that

$$Eg(T) = \sum g(t)P(T = t) = \sum_{t=0}^{n} \binom{n}{t} \theta^t (1-\theta)^{n-t} g(t) = 0$$

for all $0 \le \theta \le 1$. Observe that this is a polynomial in θ, which can only be identically equal to zero on an interval if all coefficients are zero. Now $Eg(T) = g(0)(1-\theta)^n + a_1\theta + \ldots + a_n\theta^n$, so it already follows that $g(0) = 0$ because otherwise we will get a non-zero constant term. But then

$$\sum_{t=1}^{n} \binom{n}{t} \theta^t (1-\theta)^{n-t} g(t) = 0,$$

and now we can divide through by θ and then repeat this argument to conclude that $g(1) = 0$. Continuing in this way, we see that $g(0) = g(1) = \ldots = g(n) = 0$, so $g(T) = 0$ identically, as desired.

Suppose now that we restrict θ to just the three values $\theta = 0, 1/2, 1$. Is T still complete then? The geometric interpretation discussed above shows that the answer to this is clearly no, for random samples of size $n \geq 3$. Indeed, we only have three vectors $p(\cdot;\theta)$, corresponding to the three possible values of θ, but these vectors lie in an $(n+1)$ dimensional space because t varies over $0, 1, \ldots, n$, so they cannot possibly span the whole space.

The following theorem is probably the main reason why we care about completeness.

Theorem 5.2 (Lehmann-Scheffé). *Let Y be a complete sufficient statistic. If there are unbiased estimators, then there exists a unique MVUE. We can obtain the MVUE as $T = E(U|Y)$, for any unbiased U.*

The MVUE can also be characterized as the unique unbiased *function $T = \varphi(Y)$ of the complete sufficient statistic Y.*

This is a wonderful result that will explain quite a few things that we noticed earlier, in our discussions of concrete examples. Also, and perhaps quite surprisingly, it now turns out that at least in principle it is surprisingly easy to find the MVUE once a complete sufficient statistic Y has been detected: just make Y unbiased by taking a suitable function.

We also obtain for the first time uniqueness statements about best estimators. (However, it is in fact true in general that MVUEs are unique when they exist.)

Finally, recall that MVUEs are optimal in a very strong sense: It is not possible to achieve a smaller variance *at a single* $\theta = \theta_0$ with a different unbiased estimator, even if we were willing to allow non-optimal values at other values $\theta \neq \theta_0$. (This kind of situation is quite unusual; most people and devices, for example, are good at some things and bad at others.)

Exercise 5.1. We proved in Theorem 4.5 that $\mathrm{Var}(E(X|Y)) \leq \mathrm{Var}(X)$. Show that we have equality here precisely if $X = E(X|Y)$, which holds precisely if $X = f(Y)$ is a function of Y. *Suggestion:* Extract this extra information from the proof that we gave earlier.

Proof. If U_1, U_2 are any two unbiased estimators and we define $T_j = E(U_j|Y)$, then $E(T_2 - T_1) = 0$. Since $T_2 - T_1$ is a function of Y, completeness shows that $T_1 = T_2$ with probability one. In other words, we always obtain the same $T = E(U|Y)$, no matter which unbiased estimator U we start out with. By Rao-Blackwell, T is an MVUE. If S is another MVUE, which is not assumed to be a function of Y at this point, then S and $E(S|Y)$ have the same variance, by the Rao-Blackwell Theorem again. Now the exercise shows that $S = E(S|Y) = T$, so the MVUE is unique, as claimed.

Finally, if $\varphi(Y)$ is unbiased, then $\varphi(Y) = E(\varphi(Y)|Y) = T$ is the unique MVUE, so this is indeed the only unbiased function of Y. □

This proof has also clarified the precise technical meaning of the uniqueness claims: we identify two statistics that are equal to one another with probability one, for all θ.

Example 5.2. We saw in Example 5.1 that $Y = X_1 + \ldots + X_n$ is complete and sufficient for the coin flip distribution. Lehmann-Scheffé now clarifies everything. Since $\overline{X} = Y/n$ is an unbiased function of Y, this is the unique MVUE; there is no other unbiased estimator that achieves the same variance. In particular, $\overline{X}$ is the only efficient estimator. Moreover, $\varphi(Y)$ is unbiased only for this specific function $\varphi(y) = y/n$. We observed earlier in a few concrete examples that no matter what unbiased U we start out with, we always seem to find that $E(U|Y) = Y/n$. This is now given a precise theoretical explanation.

Example 5.3. Consider the uniform distribution $f(x) = 1/\theta$, $0 < x < \theta$. We know that $Y = \max X_j$ is sufficient. Is this statistic also complete? To answer this, recall that Y has density $f_Y(y) = ny^{n-1}/\theta^n$. So $Eg(Y) = 0$ means that

$$\int_0^\theta g(y)y^{n-1}\,dy = 0$$

for all $\theta > 0$. By differentiating with respect to θ, we find from this that $g(y) = 0$ for all $y > 0$. So Y is indeed complete.

With this in place, the Lehmann-Scheffé Theorem again answers all our questions. We saw earlier that $T = \frac{n+1}{n}Y$ is unbiased, so this is the unique MVUE and in fact the only unbiased function of Y. We computed in Chapter 3 that $\text{Var}(T) = \theta^2/(n(n+2))$, and this now turns out to be best possible. As we observed earlier, this aymptotic behavior $\text{Var}(T) \sim 1/n^2$ is quite different from what we get from the CR bound when it applies.

Finally, we can now return to a perhaps cryptic remark I made in Example 4.7. Recall that I computed in somewhat informal style that $E(2\overline{X}|Y) = T$ (using the current notation), and I promised that an alternative, more rigorous argument would be forthcoming later. This we can now do: Lehmann-Scheffé says that there is only one unbiased function of Y, and we know that $E(2\overline{X}|Y)$ is such an unbiased function of Y, so it has to equal $T = \frac{n+1}{n}Y$.

This technique works quite generally and can sometimes be used to dramatically simplify the computation of conditional expectations. For example:

Exercise 5.2. Let Y be a complete sufficient statistic, and let X be an arbitrary statistic. Suppose that $E\varphi(Y) = EX$ for all θ for some (θ independent) function φ. Show that then $E(X|Y) = \varphi(Y)$.

Exercise 5.3. Use the method outlined in the previous exercise to find $E(X_1X_2|Y)$ in the coin flip example, with $Y = X_1 + \ldots + X_n$. *Suggestion:* Try a suitable combination of Y and Y^2.

Example 5.4. Another typical situation where these ideas come in handy deals with estimation of functions of previously estimated parameters. Let's make this more concrete. Consider again the coin flip example, but this time I want to estimate the variance $\delta = \theta(1-\theta)$. The statistic $Y = X_1 + \ldots + X_n$ is still complete and sufficient for the coin flip distribution, now viewed as parametrized by δ. More precisely, we solve for θ as a function of δ, and then write our distribution as

$$P(X_1 = 1) = \frac{1}{2} + \sqrt{\frac{1}{4} - \delta}, \qquad P(X_1 = 0) = \frac{1}{2} - \sqrt{\frac{1}{4} - \delta}.$$

Exercise 5.4. Can you explain in this more detail? What happened to the second solution of the quadratic equation for θ? Is Y really still complete?

So to find the MVUE, we only need to come up with an unbiased function of Y. A natural first try and starting point seems to be the statistic

$$T = \frac{Y}{n}\left(1 - \frac{Y}{n}\right).$$

Exercise 5.5. Show that T is also the MLE for δ.

Recall that $EY = n\theta$, $EY^2 = n\delta + n^2\theta^2$. This shows that

$$ET = \theta - \frac{\delta}{n} - \theta^2 = \delta - \frac{\delta}{n} = \frac{n-1}{n}\,\delta,$$

so T itself is not unbiased, but this is easy to fix:

$$U = \frac{n}{n-1}\,T = \frac{Y}{n-1}\left(1 - \frac{Y}{n}\right) \tag{5.1}$$

is the unbiased function of Y we are looking for, and thus U is the MVUE for δ.

This could have been derived more systematically. We then need an unbiased estimator for δ, which we will then condition on Y. The sample variance comes to mind, and in fact we can simplify matters by using only the first two data points:

$$S_2^2 = \left(X_1 - \frac{X_1+X_2}{2}\right)^2 + \left(X_2 - \frac{X_1+X_2}{2}\right)^2 = \frac{1}{2}(X_1 - X_2)^2 \tag{5.2}$$

Exercise 5.6. Show that

$$P(X_1 \neq X_2 | Y = y) = 2\frac{y(n-y)}{n(n-1)}.$$

If we now make use of this exercise, then (5.2) shows that

$$E(S_2^2|Y) = \frac{Y(n-Y)}{n(n-1)} = \frac{Y}{n-1}\left(1 - \frac{Y}{n}\right);$$

we have recovered the MVUE U from (5.1).

In fact, somewhat surprisingly perhaps, U *is* the (full) sample variance. To see this, recall that $X_j = 0$ or 1, so $X_j^2 = X_j$. Thus

$$\begin{aligned} S^2 &= \frac{1}{n-1}\sum_{j=1}^n \left(X_j - \frac{Y}{n}\right)^2 = \frac{1}{n-1}\sum_{j=1}^n \left(X_j - \frac{2X_jY}{n} + \frac{Y^2}{n^2}\right) \\ &= \frac{1}{n-1}\left(Y - \frac{Y^2}{n}\right) = \frac{Y}{n-1}\left(1 - \frac{Y}{n}\right) = U, \end{aligned}$$

as claimed.

Example 5.5. We are now also in a position to finally clarify everything in the discrete version of Example 5.3, the urn with an unknown number N of balls in it. So we consider the distribution $P(X_1 = x) = 1/N$ for $x = 1, 2, \ldots, N$. Recall that $Y = \max X_j$ is sufficient.

Exercise 5.7. Proceed as in the Example 5.3 to show that Y is complete. You will need the distribution $P(Y = y)$ of Y, so analyze this first (or go back to Exercise 2.1).

By the exercise, there is a unique MVUE, which can be found as the unique unbiased function of Y. As we discussed in Example 4.6, this is given by

$$T = Y + \frac{(Y-1)^n}{Y^n - (Y-1)^n}.$$

Example 5.6. Let's revisit Exercise 3.18. We consider the exponential distribution $f(x) = \theta e^{-\theta x}$ $(x > 0)$. We know that $Y = X_1 + \ldots + X_n$ is sufficient (show it again with the help of Neyman's theorem if you forgot). We're hoping that this statistic will be complete, too. Recall from our discussion in Example 3.4 that Y has density

$$f(y) = \frac{\theta^n}{(n-1)!}\, y^{n-1} e^{-\theta y}, \qquad y > 0. \tag{5.3}$$

(When you compare formulae, don't forget that what we're now calling θ corresponds to $1/\theta$ in Example 3.4.) So $Eg(Y) = 0$ gives

$$\int_0^\infty g(y) y^{n-1} e^{-\theta y}\, dy = 0,$$

and this condition for all $\theta > 0$ indeed implies that $g \equiv 0$. The transform $(Lh)(x) = \int_0^\infty h(y) e^{-xy}\, dy$ is called the *Laplace transform* of h, and completeness for the exponential distribution essentially follows from the uniqueness of Laplace transforms, so if you want to know more, this is what you should look up, but I'll leave the matter at that.

We saw earlier, in Example 3.4, that $E(1/Y) = \theta/(n-1)$. This identifies $T = (n-1)/Y$ as the unique unbiased function of Y, and thus this statistic is the unique MVUE for θ.

Something very interesting is going on here; this was hinted at in Exercise 3.18, but let me perhaps spell this out in full detail now (in particular, I'm going to solve Exercise 3.18 here). First of all, what is $\text{Var}(T)$ equal to? To answer this, I will need to find the density of $Z = 1/Y$ from that of Y, as given in (5.3). We apply the usual technique of going through the cumulative distribution: $P(1/Y \le z) = P(Y \ge 1/z) = \int_{1/z}^\infty f(t)\, dt$, and differentiation gives

$$f_Z(z) = \frac{1}{z^2} f(1/z) = \frac{\theta^n}{(n-1)!} z^{-n-1} e^{-\theta/z}, \qquad z > 0.$$

It follows that

$$\begin{aligned} EZ^2 &= \frac{\theta^n}{(n-1)!} \int_0^\infty z^{-n+1} e^{-\theta/z}\, dz = \frac{\theta^2}{(n-1)!} \int_0^\infty t^{n-3} e^{-t}\, dt \\ &= \frac{\theta^2}{(n-1)(n-2)}, \end{aligned}$$

and since $EZ = \theta/(n-1)$, we obtain $\text{Var}(Z) = \theta^2/((n-1)^2(n-2))$. Hence $\text{Var}(T) = \theta^2/(n-2)$.

On the other hand, the exponential density satisfies $\ln f = -\theta x + \ln \theta$, so

$$-\frac{\partial^2}{\partial \theta^2} \ln f(X, \theta) = \frac{1}{\theta^2},$$

and thus the Fisher information is given by $I(\theta) = 1/\theta^2$. So the estimator T does not achieve the CR bound:

$$\text{Var}(T) = \frac{\theta^2}{n-2} = \frac{n}{n-2} \frac{1}{nI(\theta)} > \frac{1}{nI(\theta)}$$

Since T is the (unique) MVUE, nothing does, and this variance that is (slightly) larger than the CR bound is the best we can do here.

Example 5.7. Consider the Poisson distribution $P(X_1 = x) = e^{-\theta}\theta^x/x!$. We already know that $\overline{X}$ is the MVUE, but let's take a look at this from the point of view suggested by the material of this section. We saw earlier, in Example 4.2, that $Y = X_1 + \ldots + X_n$ is sufficient. I claim that this statistic is also complete. We will need the distribution of Y to confirm this claim. The main ingredient to this is provided by the following fact.

Proposition 5.3. *Suppose Y_1, Y_2 are independent Poisson distributed random variables, with parameters λ_1 and λ_2, respectively. Then $S = Y_1 + Y_2$ is Poisson distributed with parameter $\lambda_1 + \lambda_2$.*

Proof.

$$\begin{aligned} P(S = x) &= \sum_{n=0}^{x} P(Y_1 = n, Y_2 = x - n) = e^{-\lambda_1-\lambda_2} \sum_{n=0}^{x} \frac{\lambda_1^n}{n!} \frac{\lambda_2^{x-n}}{(x-n)!} \\ &= \frac{e^{-\lambda_1-\lambda_2}}{x!} \sum_{n=0}^{x} \binom{x}{n} \lambda_1^n \lambda_2^{x-n} = \frac{(\lambda_1+\lambda_2)^x}{x!} e^{-\lambda_1-\lambda_2} \end{aligned}$$

□

By repeatedly applying this, we see that $Y = X_1 + \ldots + X_n$ is Poisson distributed with parameter $n\theta$. Therefore

$$Eg(Y) = e^{-n\theta} \sum_{y=0}^{\infty} g(y) \frac{(n\theta)^y}{y!}.$$

The sum is a power series in θ, so if this is equal to zero identically, then all coefficients are zero, which shows that $g(y) = 0$ for $y = 0, 1, \ldots$. In other words, Y is indeed complete.

Since $\overline{X} = Y/n$ is unbiased, as we saw earlier, this confirms one more time that $\overline{X}$ is the (unique) MVUE.

Example 5.8. This is a somewhat artificial example, intended as a cautionary tale. Consider again the Poisson distribution with parameter $\lambda > 0$; we would like to estimate $\theta = e^{-2\lambda}$. We will (quite unexpectedly) work with a random sample of size $n = 1$. I claim that $Y = X_1$ is a complete sufficient statistic. This is really the special case $n = 1$ of the previous example (neither sufficiency nor completeness is affected by taking a function of the old parameter as the new parameter), but of course we can also check it directly. The sufficiency can be confirmed by just unwrapping the definition: we need to check that $P(X_1 = x|Y = y)$ is independent of θ, but since $Y = X_1$, this probability is either always zero (if $x \neq y$) or always one (if $x = y$).

Completeness can be verified exactly as in the previous example; notice that taking a function of the parameter is indeed irrelevant because $Eg(Y) = 0$ for all λ is of course the same as requiring this for all $\theta = e^{-2\lambda}$.

Since Y is a complete sufficient statistic, an unbiased function of Y (and there will be at most one such function) is the MVUE. I now claim that

$$T = (-1)^Y = \begin{cases} 1 & Y = 0, 2, 4 \dots \\ -1 & Y = 1, 3, 5, \dots \end{cases}$$

is unbiased. Indeed,

$$ET = e^{-\lambda} \sum_{n=0}^{\infty} (-1)^n \frac{\lambda^n}{n!} = e^{-2\lambda} = \theta.$$

So, according to our self-imposed criteria of absence of bias and minimum variance, T is the best estimator of θ here, even though it definitely doesn't feel "right" intuitively. We are estimating the parameter θ which varies over $0 < \theta < 1$, and our estimate will be either 1 or -1, so we are not even making an attempt to hit the correct value, and, to add insult to injury, the guess -1 is guaranteed to be off by at least 1 whenever it occurs.

With the benefit of hindsight, it is now clear how we got into this: our insistence on unbiased estimators simply forces us to take this function of Y because there is no other unbiased function. In fact, since $Y = X_1$ is the complete random sample here, there is no other unbiased statistic.

If we drop this requirement, what would we typically have done then? We know that Y is the MVUE for λ and, moreover, Y feels natural as an estimator of λ, so perhaps $U = e^{-2Y}$ is the way to go if we want to estimate θ. Let's try to compare the *mean square errors* that we are making with these estimators:

$$\begin{aligned} E(U-\theta)^2 &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \left(e^{-2n} - e^{-2\lambda}\right)^2 \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \left(\frac{(e^{-4}\lambda)^n}{n!} - 2e^{-2\lambda} \frac{(e^{-2}\lambda)^n}{n!} + e^{-4\lambda} \frac{\lambda^n}{n!} \right) \\ &= e^{-(1-e^{-4})\lambda} - 2e^{-(3-e^{-2})\lambda} + e^{-4\lambda} \qquad (5.4) \end{aligned}$$

On the other hand,

$$E(T-\theta)^2 = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \left((-1)^n - \theta\right)^2 = 1 - 2\theta^2 + \theta^2 = 1 - \theta^2.$$

It is not very clear how to compare these as (5.4) is not a particularly transparent expression, but for example if $\theta > 0$ is small or, equivalently, λ gets large, then (5.4) is dominated by the first term $e^{-c\lambda} = \theta^{c/2}$, which is small while $\mathrm{Var}(T) \simeq 1$. So, as expected, at least for these θ, U performs considerably better than the so-called optimal estimator T. For θ close to 1 or, equivalently, small $\lambda > 0$, both estimators have a small mean square error. More precisely, a Taylor expansion shows that

$$E(U-\theta)^2 = \left(1 + e^{-4} - 2e^{-2}\right)\lambda + O(\lambda^2) \simeq 0.75\lambda,$$
$$E(T-\theta)^2 = 4\lambda + O(\lambda^2),$$

so again U is better.

Finally, let's return to larger random samples. We still want to estimate $\theta = e^{-2\lambda}$ for the Poisson distribution with parameter $\lambda > 0$. As we saw above, $Y = X_1 + \ldots + X_n$ is complete and sufficient, so we only need to find an unbiased function of Y. Since I'm out of good ideas for the moment, let's try to approach this systematically: we know that $(-1)^{X_1}$ is unbiased, so we can obtain the unique unbiased function of Y as $T = E((-1)^{X_1}|Y)$. To work out this conditional expectation, we will need the conditional probabilites $P(X_1 = x|Y = y)$. Let $S = X_2 + \ldots + X_n$. Then S is Poisson distributed with parameter $(n-1)\lambda$, by Proposition 5.3. Moreover, S and X_1 are independent, so

$$P(X_1 = x, Y = y) = P(X_1 = x)P(S = y - x) = e^{-n\lambda}\frac{\lambda^x}{x!}\frac{((n-1)\lambda)^{y-x}}{(y-x)!}.$$

Since Y is Poisson distributed with parameter $n\lambda$, by Proposition 5.3 again, this shows that

$$P(X_1 = x|Y = y) = \frac{\lambda^x}{x!}\frac{((n-1)\lambda)^{y-x}}{(y-x)!}\frac{y!}{(n\lambda)^y} = \binom{y}{x}\frac{1}{n^x}\left(1 - \frac{1}{n}\right)^{y-x}.$$

Thus if $Y = y$, then

$$E((-1)^{X_1}|Y) = \sum_{x=0}^{y}(-1)^x\binom{y}{x}\frac{1}{n^x}\left(1-\frac{1}{n}\right)^{y-x} = \left(-\frac{1}{n} + 1 - \frac{1}{n}\right)^y.$$

In other words, for a random sample of size n, the MVUE T for $\theta = e^{-2\lambda}$ is given by

$$T = E((-1)^{X_1}|Y) = \left(1 - \frac{2}{n}\right)^Y = \left(1 - \frac{2}{n}\right)^{n\overline{X}}. \tag{5.5}$$

Observe that this still contains the freakish estimator $(-1)^Y$ as the special case $n = 1$.

Exercise 5.8. Show that if $n = 2$, then (5.5) needs to be interpreted as

$$T = \begin{cases} 1 & X_1 = X_2 = 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

(the case $Y = 0$ needs special attention, otherwise this is just (5.5)). Then confirm (5.6) one more time by just computing ET.

More importantly, for large n, we can approximate $(1-2/n)^n \simeq e^{-2}$, so $T \simeq e^{-2\overline{X}}$, and this looks very sensible indeed as an estimator for $\theta = e^{-2\lambda}$, and all is well again.

Example 5.9. Consider the density $f(x) = \theta x^{\theta-1}$, $0 < x < 1$, for $\theta > 0$.

Exercise 5.9. (a) Show that the MLE is given by $\widehat{\theta} = -n/\sum_{j=1}^{n} \ln X_j$ (you did this earlier, in Exercise 3.8(a), but perhaps you forgot).
(b) Show that $I(\theta) = 1/\theta^2$.
(c) Show with the help of Neyman's theorem that $Z = X_1 X_2 \cdots X_n$ is a sufficient statistic.

Since we can take functions of sufficient statistics, $Y = -\ln Z = -\ln X_1 - \ldots - \ln X_n$ is sufficient, too. Moreover, I claim that Y is also complete. To confirm this, we will need the distribution of Y. Let's first take a look at

$$P(-\ln X_1 \le x) = P(X_1 \ge e^{-x}) = \theta \int_{e^{-x}}^{1} t^{\theta-1}\, dt.$$

By differentiating, we find that $-\ln X_1$ has density

$$f_{X_1}(x) = \theta e^{(1-\theta)x} e^{-x} = \theta e^{-\theta x};$$

in other words, $-\ln X_1$ is exponentially distributed. We can now refer to calculations we did earlier, for the first time in Example 3.4 (there's the usual small trap to avoid, what we are calling θ here corresponds to $1/\theta$ in Example 3.4). In particular, we found that Y has density

$$f(y) = \frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y};$$

compare also (5.3) above. Now completeness of Y follows as in Example 5.6. Moreover, we also saw in this example that

$$T = \frac{n-1}{Y} = -\frac{n-1}{\sum_{j=1}^{n} \ln X_j}$$

is the unique unbiased function of Y and thus the MVUE for θ. Still quoting from above, we also know that $\mathrm{Var}(T) = \theta^2/(n-2)$, and by comparing with Exercise 5.9(b), we see that T is not efficient and thus the CR bound is not attained.

Exercise 5.10. Consider the uniform distribution $f(x,\theta) = 1/(2\theta)$ on $-\theta < x < \theta$. Show that $Y = X_1 + \ldots + X_n$ is not complete. *Suggestion:* Use symmetry, or ignore this suggestion if it confuses you.

Exercise 5.11. Consider the density $f(x,\theta) = e^{\theta - x}$, $x > \theta$, for $\theta \in \mathbb{R}$.
(a) Show that $Y = \min(X_1, X_2, \ldots, X_n)$ is a sufficient statistic.
(b) Find the distribution of Y.
(c) Show that Y is complete.
(d) Find the MVUE.

5.2. **Exponential classes.** In this section, we identify a general class of distributions that come supplied with a complete sufficient statistic (which is a highly desirable feature to have, as we saw in the previous section).

Definition 5.4. A distribution of the form

$$L(x,\theta) = e^{p(\theta)K(x)+M(x)+q(\theta)}, \qquad (x \in S)$$

is called a *regular exponential class* if S is independent of θ and the functions p, K are not constant.

This covers both the discrete and the continuous case, with L being the density in the continuous case and $L(x) = P(X = x)$ in the discrete case. Later on, we will also impose differentiability assumptions on the functions p, q, K, M (which we won't make explicit, however).

Let's collect a few quick examples.

Example 5.10. The $N(0, \theta^{1/2})$ distribution $f(x) = (2\pi\theta)^{-1/2}e^{-x^2/(2\theta)}$ is a regular exponential class, with $S = \mathbb{R}$, and we can put $K(x) = x^2$, $p(\theta) = -1/(2\theta)$, $M(x) = 0$, $q(\theta) = -(1/2)\ln 2\pi\theta$.

Example 5.11. For a discrete example, return to the coin flip, that is, $L(x,\theta) = \theta^x(1-\theta)^{1-x}$. This is of the required form with $S = \{0, 1\}$ since we can write

$$L = e^{x\ln\theta + (1-x)\ln(1-\theta)},$$

so taking $K(x) = x$, $p(\theta) = \ln\theta - \ln(1-\theta)$, $q(\theta) = \ln(1-\theta)$, $M(x) = 0$ works.

Example 5.12. Now consider the uniform distribution $f(x,\theta) = 1/\theta$, $0 < x < \theta$. At first sight, this might appear to be of the required form with $q(\theta) = -\ln\theta$, $K = M = p = 0$, but this doesn't help because the support $S = (0,\theta)$ is not independent of θ. In fact, there is a second problem with the requirement that p, K must not be constant. So this is not a regular exponential class.

Now suppose we draw a random sample from a regular exponential class. Then the likelihood function is given by

$$L(x_1, \ldots, x_n; \theta) = e^{p(\theta)\sum_{j=1}^n K(x_j) + nq(\theta)} e^{\sum_{j=1}^n M(x_j)}. \tag{5.7}$$

Thus, by Neyman's criterion, $Y = K(X_1) + \ldots + K(X_n)$ is a sufficient statistic. It turns out we can say quite a bit more:

Theorem 5.5. *Let $Y = K(X_1) + \ldots + K(X_n)$. Then:*
(a) The distribution of Y has the form

$$L_Y(y, \theta) = R(y) e^{p(\theta)y + nq(\theta)}.$$

(b) Y is a complete sufficient statistic.
(c) $EY = -nq'(\theta)/p'(\theta)$ and

$$\text{Var}(Y) = \frac{n}{p'(\theta)^3} \left(p''(\theta)q'(\theta) - q''(\theta)p'(\theta) \right).$$

Sketch of proof. I'll do this in the discrete setting. Part (a) follows from (5.7) because

$$P(Y = y) = \sum_{x: Y(x) = y} e^{\sum M(x_j)} e^{p(\theta)y + nq(\theta)},$$

which is of the desired form, with $R(y) = \sum e^{\sum M(x_j)}$.

We already saw that Y is sufficient, and as for completeness, part (a) shows that $Eg(Y) = 0$ implies that

$$\sum_y g(y) R(y) e^{p(\theta)y} = 0. \tag{5.8}$$

Now if $p(\theta)$ takes values in at least an interval, we would indeed expect that (5.8) for all θ implies that $g(y) = 0$ for all y; we can reason by analogy to power series or Laplace transforms. Of course, the argument is not very convincing in this form, but I'll just leave the matter at that anyway.

As for (c), take derivatives in

$$\sum_y R(y) e^{p(\theta)y + nq(\theta)} = 1.$$

This gives

$$\sum_y R(y)(p'(\theta)y + nq'(\theta)) e^{p(\theta)y + nq(\theta)} = 0,$$

or, put differently, $p'(\theta)EY + nq'(\theta) = 0$, as claimed. To compute the variance, take one more derivative to obtain

$$\sum_y R(y) \left[(p'(\theta)y + nq'(\theta))^2 + p''(\theta)y + nq''(\theta) \right] e^{p(\theta)y + nq(\theta)} = 0.$$

In other words,

$$p'^2 EY^2 + n^2 q'^2 + 2np'q'EY + p''EY + nq'' = 0,$$

and the asserted formula for $\mathrm{Var}(Y)$ follows from this. □

Exercise 5.12. Do this calculation in more detail.

This theorem doesn't really tell us anything radically new, but it does unify quite a few previous results. If we return to Examples 5.10, 5.11, then we learn from Theorem 5.5(b) that $Y = \sum X_j^2$ is complete and sufficient for the $N(0, \theta^{1/2})$ distribution. You showed earlier, in Exercise 4.3, that Y is sufficient; the completeness is a new result. We also learn from Theorem 5.5(b) that $Y = \sum X_j$ is a complete and sufficient statistic for the coin flip. These statements are not new to us, of course.

Exercise 5.13. Show that the Poisson distribution $P(X = x; \theta) = e^{-\theta}\theta^x/x!$ forms a regular exponential class. Then use Theorem 5.5(b) to conclude one more time that $Y = X_1 + \ldots + X_n$ is complete and sufficient.

Example 5.13. For a new example, consider the *gamma distribution* with parameters $\alpha > 0$ (assumed to be known here) and $\theta > 0$ (unknown, as usual):

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}, \qquad (x > 0)$$

Exercise 5.14. Show that $\int f = 1$.

To identify this as a regular exponential class, we can put $K(x) = x$, $p(\theta) = -1/\theta$, $q(\theta) = -\ln \Gamma(\alpha)\theta^\alpha$, and $M(x) = (\alpha - 1)\ln x$. Now Theorem 5.5(b) shows that $Y = X_1 + \ldots + X_n$ is complete and sufficient. Moreover, EY can be conveniently obtained from part (c): notice that $q = -\ln \Gamma(\alpha) - \alpha \ln \theta$, so $q' = -\alpha/\theta$ and thus $EY = n\alpha\theta$. In particular, the unique unbiased function of Y is given by

$$T = \frac{\overline{X}}{\alpha} = \frac{Y}{n\alpha};$$

this is also the unique MVUE for θ. To find its variance, we can again refer to part (c) of the Theorem:

$$\mathrm{Var}(T) = \frac{1}{n\alpha^2}\theta^6 \left(\frac{2\alpha}{\theta^4} - \frac{\alpha}{\theta^4} \right) = \frac{\theta^2}{n\alpha}$$

Exercise 5.15. Compute the Fisher information for the gamma distribution. Is T efficient?

5.3. **Ancillary statistics.**

Definition 5.6. A statistic $T = T(X_1, \dots, X_n)$ is called *ancillary* if $P(T \in A; \theta)$ is independent of θ for all $A \subseteq \mathbb{R}$.

This condition is reminiscent of the one defining sufficient statistics, but in fact ancillary and sufficient statistics are at opposite ends of the spectrum: a sufficient statistic contains all the information worth knowing about if we want to estimate θ while the information contained in an ancillary statistic is completely irrelevant for this assignment, or at least it's safe to say that this interpretation is in line with the one we gave earlier for sufficient statistics. Indeed, recall that if T is sufficient, then $P(X_1 = x_1, \dots, X_n = x_n | T = t)$ is independent of θ, so once the value of T is known, additional information on the random sample does not support any inference whatsoever on θ. On the other hand, if T is ancillary, then the value of T itself does not support any inference on θ.

A constant random variable $T = c$ is an ancillary statistic. A more interesting example is given by the sample variance S^2 for an $N(\theta, 1)$ distribution. We know from Theorem 2.8(c) that $(n-1)S^2 \sim \chi^2(n-1)$, and this distribution is independent of θ, as required. In fact, we can also show directly that S^2 is ancillary, without knowing its distribution: it suffices to observe that

$$S^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \overline{X})^2$$

is invariant under simultaneous shifts $X_j \to X_j - a$. Moreover, the X_j are a random sample drawn from an $N(\theta, 1)$ distribution precisely if the $X_j - \theta$ are iid and standard normal. So it will not affect the distribution of S^2 if we pretend that the X_j are $N(0, 1)$-distributed, and thus the distribution of S^2 is indeed independent of θ.

For a general class of examples, suppose that Y is a sufficient statistic, and T is another statistic that is independent of Y for all θ. Then (in the discrete case, say)

$$P(T = t) = P(T = t | Y = y)$$

for any y with $P(Y = y) > 0$, and this conditional probability is independent of θ, by the sufficiency of Y. Note in this context that the event $T = t$ can be decomposed into events of the type $X_1 = x_1, \dots, X_n = x_n$ because T is a statistic. So, to sum this up, if T is a statistic that is independent of a sufficient statistic, then T is ancillary.

Interestingly, there is a partial converse to this observation:

Theorem 5.7 (Basu). *If T is an ancillary statistic and Y is complete and sufficient, then T and Y are independent for all θ.*

Proof. I'll assume that the distributions are discrete (but in fact the following argument works in general). For t in the range of T, let $f = \chi_{\{t\}}$, so $f(s) = 1$ if $s = t$ and $f(s) = 0$ otherwise, and consider $E(f(T)|Y)$. Observe that $E(E(f(T)|Y)) = Ef(T) = P(T = t)$, or, put differently,

$$E[E(f(T)|Y) - P(T = t)] = 0, \tag{5.9}$$

for all θ. Now since T is ancillary, $P(T = t)$ is just a constant. Moreover, $E(f(T)|Y) = g(Y)$ is also independent of θ, by the sufficiency of Y. So the expression in $[\ldots]$ is a θ independent function of Y, and thus completeness may be applied to (5.9). It follows that $E(f(T)|Y) = P(T = t)$.

Exercise 5.16. Show that (for the f from above) if $Y(\omega) = y$, then

$$E(f(T)|Y)(\omega) = P(T = t|Y = y). \tag{5.10}$$

Suggestion: Use the definition of conditional expectation, see (4.6).

From (5.10), we now see that $P(T = t|Y = y) = P(T = t)$ for all y with $P(Y = y) > 0$, and thus T, Y are indeed independent, as claimed. □

Example 5.14. Consider the $N(\theta, \sigma)$ distribution with known σ. We observed above that the sample variance S^2 is ancillary. Moreover, the sample mean $\overline{X}$ is sufficient. This we can confirm by identifying the normal distribution as a regular exponential class: the density

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{\theta^2 + 2\theta x - x^2}{2\sigma^2}\right)$$

is of the required form with $K(x) = x$, $p(\theta) = \theta/\sigma^2$, $M(x) = -x^2/(2\sigma^2)$, $q(\theta) = \theta^2/(2\sigma^2) - \ln\sqrt{2\pi}\sigma$. Therefore $Y = K(X_1) + \ldots + K(X_n) = n\overline{X}$ is complete and sufficient by Theorem 5.5(b), as desired. Now Basu's Theorem shows that $\overline{X}$ and S^2 are independent, a result we stated earlier (but did not prove then) in Theorem 2.8(b).

Example 5.15. Consider the density $f(x) = e^{-(x-\theta)}$, $x > \theta$. As above in the case of the $N(\theta, \sigma)$ distribution, the parameter θ only affects the location of the density, not its shape. Thus statistics that are insensitive to simultaneous shifts of the whole random sample are automatically ancillary. To state this more formally: if $T(X_1 + c, \ldots, X_n + c) =$

$T(X_1, \ldots, X_n)$, then T is ancillary. This follows because $X_j - \theta$ is a random sample drawn from the distribution $f(x) = e^{-x}$, $x > 0$, so

$$P(T = t; \theta) = P(T(X_1 - \theta, \ldots, X_n - \theta) = t; \theta) \\ = P(T(X_1, \ldots, X_n) = t; 0)$$

is independent of θ, as required.

Examples of such shift invariant statistics are given by S^2, by the sample range $R = \max X_j - \min X_j$, or by $T = \overline{X} - \min X_j$. On the other hand, I claim that $Y = \min X_j$ is complete and sufficient. Given this, Basu's Theorem now shows that each of S^2, R, T is independent of Y. These statements are interesting even if we specialize to particular values of θ, say $\theta = 0$.

Exercise 5.17. Use Neyman's criterion to show that Y is indeed sufficient.

To confirm that Y is complete, observe that (for $y > \theta$)

$$P(Y \geq y) = P(X_1 \geq y, \ldots, X_n \geq y) = P(X_1 \geq y)^n \\ = e^{n\theta} \left(\int_y^\infty e^{-x} \, dx \right)^n,$$

so $f_Y(y) = ne^{n(\theta - y)}$, $y > \theta$. Thus, if $Eg(Y) = 0$, then

$$\int_\theta^\infty g(y) e^{-ny} \, dy = 0,$$

and if this holds for all θ, then we can differentiate with respect to θ, and we conclude that $g = 0$, as required.

Example 5.16. Finally, consider again the density $f(x) = e^{-x/\theta}/\theta$, $x > 0$.

Exercise 5.18. Suppose that the random variable $X > 0$ has density $g(x)$, and let $c > 0$. Show that $Y = cX$ has density $g(x/c)/c$.

So by this exercise, varying the parameter θ has the same effect as rescaling the random variable according to $X \to \theta X$. In particular, in complete analogy to the previous example, any statistic that is insensitive to a rescaling of the random sample will be ancillary. Examples include $T_1 = \overline{X}/\max X_j$, $T_2 = X_2/X_1$ etc. By Basu's Theorem, any such statistic is independent of the complete and sufficient statistic $Y = X_1 + \ldots + X_n$.