Before the experiment is performed we have no data, so we cannot obtain the observed information. However we can calculate the expected or Fisher information. –A. C. Davison

Solutions to Problems 41-50

41. If the likelihood is $L(\theta; x)$, what is its score function $S(\theta; x)$? Regarding S as a transformation of the random variable X (by capitalizing the data argument), and fix $\theta = \theta_0$, then $S(\theta_0; X)$ should have an expectation. Find this expectation.

Solution. $S(\theta; x) := \frac{\partial}{\partial\theta}\ln L(\theta; x) \equiv \frac{\partial\ell(\theta;x)}{\partial\,\theta}$. $\mathbb{E}\left(\frac{\partial \ln L}{\partial\theta}\right) = \mathbb{E}\left(\frac{1}{L}\frac{\partial L}{\partial\theta}\right) = \int \frac{1}{f}\frac{\partial f}{\partial\theta} f dx = \int \frac{\partial f}{\partial\theta} dx \overset{!!}{=} \frac{\partial}{\partial\theta}\int f dx = 0$ where all $\frac{\partial}{\partial\theta}$ is carried out at $\theta = \theta_0$, which is an arbitrary location; despite of this, the result is void of θ_0.

42. If the likelihood is $L(\theta; x)$, what is its observed information $J(\theta; x)$? What is its Fisher information $I(\theta; x)$? Why do people say the Fisher information is the same as the variance of the score function?

Solution. $J(\theta; x) := -\frac{\partial^2}{\partial\theta^2}\ln L(\theta; x)$. $I(\theta\,; x) := \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\ln L(\theta\,; x)\right]$. The variance of $S(\theta; x)$ is $\mathbb{E}\left[\left(\frac{\partial \ln L}{\partial\theta}\right)^2\right]$. The trick for showing that this is equal to $I(\theta; x)$ is to start with $I(\theta; x)$, the one that is more readily expandable: $-(\ln L)'' = -\frac{\partial}{\partial\theta}\left(\frac{L'}{L}\right) = -\frac{L''L - L'L'}{L^2} = \left(\frac{L'}{L}\right)^2 - \frac{L''}{L}$ (prime notation denotes derivative w.r.t. θ) $\rightarrow$ $\mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\ln L(\theta\,; x)\right] = \mathbb{E}\left[\left(\frac{\partial \ln L}{\partial\theta}\right)^2\right] - \mathbb{E}\left[\frac{1}{L}\frac{\partial^2 L}{\partial\theta^2}\right]$, but $\mathbb{E}\left[\frac{1}{L}\frac{\partial^2 L}{\partial\theta^2}\right] = \int \frac{1}{f}\frac{\partial^2 f}{\partial\theta^2} f dx = \int \frac{\partial^2 f}{\partial\theta^2} dx \overset{!!}{=} \frac{\partial^2}{\partial\theta^2}\int f dx = 0 \rightarrow \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\ln L(\theta\,; x)\right] = \mathbb{E}\left[\left(\frac{\partial \ln L}{\partial\theta}\right)^2\right]$.

43. If the likelihood is $L(\theta; x)$, let S and I denote the score function and the Fisher information and let $\hat\theta$ be an estimator for θ, show that $\mathbb{V}(\hat\theta) \geq \frac{\mathbb{E}(\hat\theta S)^2}{I}$. Then show that $\mathbb{E}(\hat\theta S) = 1 + \frac{\partial b(\hat\theta)}{\partial\theta}$ where $b(\hat\theta)$ is the bias of the estimator $\hat\theta$. Now if $\hat\theta$ is required to be unbiased for θ, then what is a lower bound for the variance of the unbiased $\hat\theta$?

Solution. $1 \geq \rho(\hat\theta, S)^2 = \frac{\mathrm{Cov}(\hat\theta, S)^2}{\mathbb{V}(\hat\theta)\mathbb{V}(S)} = \frac{\mathbb{E}(\hat\theta S)^2}{\mathbb{V}(\hat\theta)\mathbb{V}(S)} \rightarrow \mathbb{V}(\hat\theta) \geq \frac{\mathbb{E}(\hat\theta S)^2}{I}$. The trick to show that $\mathbb{E}(\hat\theta S) = 1 + \frac{\partial b(\hat\theta)}{\partial\theta}$ is to start with the RHS:

$$1 + \frac{\partial b(\hat\theta)}{\partial\theta} = \frac{\partial[\theta + b(\hat\theta)]}{\partial\theta} \overset{!!}{=} \frac{\partial}{\partial\theta}\mathbb{E}[\hat\theta] = \frac{\partial}{\partial\theta}\int \hat\theta f dx \overset{!!}{=} \int \frac{\partial}{\partial\theta}[\hat\theta f]dx = \int \left[\frac{\partial\hat\theta}{\partial\theta} f + \frac{\partial f}{\partial\theta}\hat\theta\right] dx \overset{!!}{=} \int \frac{\partial f}{\partial\theta}\hat\theta dx \overset{!!}{=} \int \frac{\partial \ln f}{\partial\theta}\hat\theta f dx.$$

Now if $\hat\theta$ is constructed as unbiased, then for all θ, $b(\hat\theta) = 0$, therefore $\frac{\partial b(\hat\theta)}{\partial\theta} = 0$. Therefore $\mathbb{E}(\hat\theta S) = 1$ and $\mathbb{V}(\hat\theta) \geq I^{-1}$.

44. What is a *uniformly minimum variance unbiased estimator (UMVUE)*? Why is it *unique*, if exists? What is a lower *bound* for that minimum variance? What is the role of the parameter's Ln-Likelihood to play with this UMVUE?

Answer. This question clarifies the use of this lower bound. A uniformly minimum variance unbiased estimator (UMVUE) is an unbiased estimator possessing the minimum variance among all unbiased estimators at all possible values of the target parameter.

Lemma. UMVUE is unique, if exists. **Proof of Lemma.** Let T and T' be two UMVUEs. The trick is to consider $\frac{1}{2}(T+T')$ and its variance: $\mathbb{V}(T) \leq \mathbb{V}\left[\frac{1}{2}(T+T')\right] = \frac{1}{4}\mathbb{V}(T)+\frac{1}{4}\mathbb{V}(T')+\frac{1}{2}\mathrm{Cov}(T,T') \leq \frac{1}{4}\mathbb{V}(T)+\frac{1}{4}\mathbb{V}(T')+\frac{1}{2}\sqrt{\mathbb{V}(T)\mathbb{V}(T')} = \mathbb{V}(T)$. Hence $\rho(T,T')=1$ which happens only if $T'=kT+b$. Note also that $\mathrm{Cov}(T,T')=\mathbb{V}(T)$ is implied in the equality attainment. Therefore $\mathrm{Cov}(T,kT+b)=\mathbb{V}(T) \rightarrow k=1, b=0$, or $T=T'$.

The previous 3 problems which arrived at the famous result known as the Fréchet-Darmois-Cramér-Rao Lower Bound, I^{-1}, for all unbiased estimators' variances. The role of Ln-Likelihood is the backdrop vessel of information from raw data from which grants the score function, the Fisher information, etc..

45. Describe the sufficiency of a statistic verbally without citing Neyman's factorization. Then, define the sufficiency of a statistic via Neyman's factorization.

Answer. Sufficiency of a statistic means, for the purpose of inferring about the parameter, full knowledge about the statistic shall waive the need to look at the sample, whereever is the parameter's location (of course, asserting the operating statistical model is correct). We say that a statistic is sufficient for a parameter (not for the data nor the model). A more precise definition without citing Neyman's factorization is using the statistical independence: T is sufficient for θ iff $X|T \perp\!\!\!\perp \theta$. In words, which ever θ_0 is θ located at, the distribution of $X|T$ is unaffected and remains the same.

Using Neyman's factorization criterion (a theorem proved by Jerzy Neyman), T is sufficient for θ iff $L(\theta;x)=h(x)g(T(x),\theta)$. In words, there is a (multiplicative) factor of the likelihood that outside that factor lies no θ and inside that factor all data is interacting with θ via the intermediation of T.

46. Describe completeness of a statistic verbally based on the basic requirement of identifiability of any parameter's true value by the statistic. Then, define the completeness of a statistic, formally, using expectation.

Answer. Completeness of a statistic means, for the purpose of inferring about the parameter, full knowledge about the statistic is sophisticated enough to tell apart any two possible values of the parameter so that it can completely pin down to the one value of the parameter, wherever it is located. We say that a statistic is complete for a parameter (not for the data nor the model).

Using expectation, a precise definition is the following: T is complete for θ iff any function g such that $\mathbb{E}\big(g(T)\big)=0$ requires $g(T)\equiv 0$ (with probability 1).

47. Let X and Y be two random variables. Show that $\mathbb{E}(Y|X)$, as a random variable, reduces variance of Y while maintaining its mean, i.e., show that $\mathbb{V}(\mathbb{E}(Y|X)) \leq \mathbb{V}(Y)$ and that $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$. When is $\mathbb{V}(\mathbb{E}(Y|X)) = \mathbb{V}(Y)$?

Solution. Law of Iterated Expectation $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$, is proved in Problem 1. **Lemma.** $\mathbb{V}(Y) = \mathbb{E}(\mathbb{V}(Y|X)) + \mathbb{V}(\mathbb{E}(Y|X))$. **Proof of Lemma.** $\mathbb{E}[\mathbb{V}(Y|X)] = \mathbb{E}[\mathbb{E}(Y^2|X)] - \mathbb{E}[[\mathbb{E}(Y|X)]^2]$, $\mathbb{V}[\mathbb{E}(Y|X)] = \mathbb{E}[[\mathbb{E}(Y|X)]^2] - [\mathbb{E}(Y)]^2$, therefore $\mathbb{E}[\mathbb{V}(Y|X)] + \mathbb{V}[\mathbb{E}(Y|X)] = \mathbb{E}[\mathbb{E}(Y^2|X)] - [\mathbb{E}(Y)]^2$. But $\mathbb{E}[\mathbb{E}(Y^2|X)] = \mathbb{E}(Y^2)$, hence the result. (**End of Proof of Lemma**). Since $\mathbb{E}(\mathbb{V}(Y|X)) \geq 0$, therefore $\mathbb{V}(\mathbb{E}(Y|X)) \leq \mathbb{V}(Y)$, and $\mathbb{V}(\mathbb{E}(Y|X)) = \mathbb{V}(Y)$ iff $\mathbb{E}(\mathbb{V}(Y|X)) = 0$ iff $\mathbb{V}(Y|X) = 0$ with probability 1.

48~~. Let T be a complete sufficient statistic on the data. Show that for any function h of T, its variance, $\mathbb{V}(h(T))$, cannot be reduced by conditioning on T.~~

49. Establish a way for constructing the UMVUE by considering the connection between a complete sufficient statistic and UMVUE. That is, please prove the Lehmann-Scheffé Theorem.

Solution. Lemma. For any two unbiased estimators of the parameter θ, U and U', it holds that $\mathbb{E}(U|T) = \mathbb{E}(U'|T)$ with probability 1, where T is complete for θ. **Proof of Lemma.** First note that $\mathbb{E}[\mathbb{E}(U|T)] = \mathbb{E}[\mathbb{E}(U'|T)] = \theta$. Since both conditional expectations are functions of the complete statistic T, and their difference has an expected value equal to zero: $\mathbb{E}[\mathbb{E}(U|T) - \mathbb{E}(U'|T)] = 0$, therefore they equal to each other with probability 1 (directly from definition of completeness). (End of Proof of Lemma)

Then we only have to note that for any unbiased U, it holds that $\mathbb{V}(U) \geq \mathbb{V}[\mathbb{E}(U|T)]$ and that $\mathbb{E}(U) = \mathbb{E}[\mathbb{E}(U|T)]$, implying that $\mathbb{E}(U|T)$ is UMVUE.

50. Let $X \sim Poisson(\lambda)$, $\lambda > 0$. The statistic $T(X) := \mathbb{I}(X = 0)$ (i.e., T indicates whether $X = 0$) is used to estimate $q(\lambda) = e^{-\lambda}$. Show that T is the UMVUE (Hint: by Lehmann-Scheffé) and yet $\mathbb{V}(T)$ does not attain the Cramér-Rao Lower Bound. Note that the sample size here is only 1.

Solution. Lemma. X is complete for $e^{-\lambda}$. **Proof of Lemma.** Suppose g is a function of X such that

$$0 = \mathbb{E}(g(X)) = e^{-\lambda} \sum_{x=0}^{\infty} g(x) \frac{\lambda^x}{x!},$$

then the only choice of g is the constant zero function. This means, by definition, that X is complete for $e^{-\lambda}$. (End of Proof of Lemma)

Note that $\mathbb{E}(T) = \mathbb{P}(X = 0) = e^{-\lambda}$. And since T is a function only of X, $\mathbb{E}(T|X) = T$, by Lehmann- Scheffé, T is the UMVUE. $\mathbb{V}(T) = e^{-\lambda}(1 - e^{-\lambda})$.

The Fisher Information for $e^{-\lambda}$ is equal to the variance of the score function: $I(e^{-\lambda}) = \mathbb{E}\left[\left(\frac{\partial \ln f}{\partial(e^{-\lambda})}\right)^2\right] = \mathbb{E}\left[\left(\frac{\partial\lambda}{\partial e^{-\lambda}}\frac{\partial \ln f}{\partial\lambda}\right)^2\right] = \mathbb{E}\left[e^{2\lambda}\left(\frac{\partial}{\partial\lambda}(-\lambda + X\ln\lambda - \ln X!)\right)^2\right] = e^{2\lambda}\mathbb{E}\left[\left(-1+\frac{X}{\lambda}\right)^2\right] = e^{2\lambda}\left[1 - \frac{2}{\lambda}\mathbb{E}(X) + \frac{1}{\lambda^2}\mathbb{E}(X^2)\right] = \frac{1}{\lambda}e^{2\lambda}$. Thus the CRLB for the variance of any unbiased estimator is $\frac{1}{I} = \lambda e^{-2\lambda}$ and $e^{\lambda} \geq 1 + \lambda \Rightarrow 1 \geq e^{-\lambda} + \lambda e^{-\lambda} \Rightarrow e^{-\lambda} \geq e^{-2\lambda} + \lambda e^{-2\lambda} \Rightarrow e^{-\lambda}(1 - e^{-\lambda}) \geq \lambda e^{-2\lambda}$ or $\mathbb{V}(T) \geq I^{-1}$, attaining equality only if $\lambda = 0$.