Answers to the 8th problem set

The likelihood ratio with which we worked in this problem set is:

$$\Lambda(x) = \frac{f(x \mid \theta = \theta_1)}{f(x \mid \theta = \theta_0)} = \frac{L(\theta_1)}{L(\theta_0)}.$$

With a lower-case "x", this defines a function. With a capital "X", this is a random variable. We assumed X is a continuous random variable whose density, given that $\theta = \theta_0$, is $f(x \mid \theta = \theta_0)$. We used "$\mathbf{E}_{\theta_0}(\quad)$", "$\mathbf{var}_{\theta_0}(\quad)$", "$\mathbf{cov}_{\theta_0}(\quad)$", etc., to mean these operators are to be evaluated assuming $\theta = \theta_0$.

1. Show that $\mathbf{E}_{\theta_0}(\Lambda(X)) = 1$.

   Answer:

$$\begin{aligned}\mathbf{E}_{\theta_0}(\Lambda(X)) = \int_{-\infty}^{\infty} \Lambda(x)\, f(x \mid \theta = \theta_0)\, dx &= \int_{-\infty}^{\infty} \frac{f(x \mid \theta = \theta_1)}{f(x \mid \theta = \theta_0)}\, f(x \mid \theta = \theta_0)\, dx \\ &= \int_{-\infty}^{\infty} f(x \mid \theta = \theta_1)\, dx = 1.\end{aligned}$$

2. (a) The following may or may not be useful in doing part (b), perhaps depending on your tastes: Suppose $\mathbf{E}(U) = \mu$ and $\mathbf{E}(V) = \nu$. Show that $\mathbf{cov}(U, V)$ can be written in any of these forms (the second one is the definition, so there is nothing to show):

$$\mathbf{E}(U(V-\nu)) = \mathbf{E}((U-\mu)(V-\nu)) = \mathbf{E}((U-\mu)V).$$

   Answer:

$$\begin{aligned}\mathbf{E}((U-\mu)(V-\nu)) &= \mathbf{E}(U(V-\nu)) - \mathbf{E}(\mu(V-\nu)) \\ &= \mathbf{E}(U(V-\nu)) - \mu\,\mathbf{E}(V-\nu) \\ &= \mathbf{E}(U(V-\nu)) - \mu(\mathbf{E}(V)-\nu) \\ &= \mathbf{E}(U(V-\nu)) - \mu(\nu-\nu) \\ &= \mathbf{E}(U(V-\nu)).\end{aligned}$$

   We could of course say that the other identity is proved by the same method, but even that is needlessly complicated, since we can just observe that the other identity says the same thing as this one.

(b) Suppose $T(X)$ is an unbiased estimator of θ.
Show that $\mathbf{cov}_{\theta_0}(T(X), \Lambda(X)) = \theta_1 - \theta_0$.

Answer:

$$\begin{aligned}
\mathbf{cov}_{\theta_0}(T(X), \Lambda(X)) &= \mathbf{E}_{\theta_0}(T(X)(\Lambda(X) - 1)) \\
&\quad \text{(We have used the results of part \#2(a) and of \#1.)} \\
&= \int_{-\infty}^{\infty} T(x)(\Lambda(x) - 1)\, f(x \mid \theta = \theta_0)\, dx \\
&= \int_{-\infty}^{\infty} T(x) \left(\frac{f(x \mid \theta = \theta_1)}{f(x \mid \theta = \theta_0)} - 1 \right) f(x \mid \theta = \theta_0)\, dx \\
&= \int_{-\infty}^{\infty} T(x) \left(f(x \mid \theta = \theta_1) - f(x \mid \theta = \theta_0) \right) dx \\
&= \int_{-\infty}^{\infty} T(x)\, f(x \mid \theta = \theta_1)\, dx - \int_{-\infty}^{\infty} T(x)\, f(x \mid \theta = \theta_0)\, dx \\
&= \mathbf{E}_{\theta_1}(T(X)) - \mathbf{E}_{\theta_0}(T(X)) \\
&= \theta_1 - \theta_0 \text{ because } T(X) \text{ is an unbiased estimator of } \theta.
\end{aligned}$$

3. Use the Cauchy-Schwartz inequality in the form $|\mathbf{cov}(U, V)| \leq \mathbf{SD}(U)\,\mathbf{SD}(V)$ and the result of #2(b) to show that when $\theta = \theta_0$ then the mean squared error of $T(X)$ as an estimator of θ cannot be less than

$$\frac{(\theta_1 - \theta_0)^2}{\mathbf{var}_{\theta_0}(\Lambda(X))}.$$

Answer:

$$|\theta_1 - \theta_0| \underset{\text{By \#2(b)}}{=} |\mathbf{cov}_{\theta_0}(T(X), \Lambda(X))| \underset{\text{Cauchy-Schwartz}}{\leq} \mathbf{SD}_{\theta_0}(T(X))\,\mathbf{SD}_{\theta_0}(\Lambda(X)).$$

$$\text{Therefore } (\theta_1 - \theta_0)^2 \leq \mathbf{var}_{\theta_0}(T(X))\,\mathbf{var}_{\theta_0}(\Lambda(X)),$$

$$\text{and so } \frac{(\theta_1 - \theta_0)^2}{\mathbf{var}_{\theta_0}(\Lambda(X))} \leq \mathbf{var}_{\theta_0}(T(X)) \underset{\text{Because } T(X) \text{ is unbiased.}}{=} \text{mean squared error.}$$

4. Briefly comment on the justification for, and validity of, each step below.

Answer:

$$\frac{\mathsf{var}_{\theta_0}(\Lambda(X))}{(\theta_1-\theta_0)^2} = \mathsf{var}_{\theta_0}\left(\frac{\Lambda(X)-1}{\theta_1-\theta_0}\right)$$

$\mathsf{var}(\Lambda(X)) = \mathsf{var}(\Lambda(X)-1)$ since adding a constant to a random variable does not alter its variance.

$$\frac{\mathsf{var}(\Lambda(X)-1)}{(\theta_1-\theta_0)^2} = \mathsf{var}\left(\frac{\Lambda(X)-1}{\theta_1-\theta_0}\right)$$

because $c\,\mathsf{var}(Y) = \mathsf{var}(c^2 Y)$ if c is constant.

$$= \mathsf{var}_{\theta_0}\left(\frac{1}{f(X\mid\theta=\theta_0)}\cdot\frac{f(X\mid\theta=\theta_1)-f(X\mid\theta=\theta_0)}{\theta_1-\theta_0}\right)$$

This is just the definition of Λ, which says $\Lambda(X) = \dfrac{f(X\mid\theta=\theta_1)}{f(X\mid\theta=\theta_0)}$, followed by a bit of algebra.

$$\longrightarrow \mathsf{var}_{\theta_0}\left(\frac{\partial}{\partial\theta}\log f(X\mid\theta)\bigg|_{\theta=\theta_0}\right) \quad \text{as } \theta_1\longrightarrow\theta_0.$$

This last step involves some subtlety, but the conjunction of the words "briefly comment" with the prerequisites for this course entails that **you don't need to go into that** in order to get credit for this problem. Most of what you need to do is to observe that

$$\frac{f(X\mid\theta=\theta_1)-f(X\mid\theta=\theta_0)}{\theta_1-\theta_0} \longrightarrow \frac{\partial}{\partial\theta}f(X\mid\theta)\bigg|_{\theta=\theta_0} \quad \text{as } \theta_1\longrightarrow\theta_0$$

because that is how derivatives are defined. The subtlety resides in the following step, which the notation we used above camouflages, and the notation on the displayed line below this line makes explicit:

$$\lim_{\theta_1\longrightarrow\theta_0}\mathsf{var}\left(\cdots\cdots\cdots\cdots\right) = \mathsf{var}\left(\lim_{\theta_1\longrightarrow\theta_0}\cdots\cdots\cdots\cdots\right).$$

The variance is an integral, so the question is: When is the following true:

$$\lim_{\theta_1\longrightarrow\theta_0}\int_{-\infty}^{\infty}\cdots\cdots\cdots\cdots = \int_{-\infty}^{\infty}\lim_{\theta_1\longrightarrow\theta_0}\cdots\cdots\cdots\cdots\ ?$$

Answers to questions of this sort may be found in Walter Rudin's book *Principles of Mathematical Analysis.*

5. Let $I(\theta) = \textsf{\textbf{var}}_\theta\left(\dfrac{\partial}{\partial\theta}\log f(X \mid \theta)\right)$. This quantity $I(\theta)$ is called the "Fisher information" in the sample X. It is thought of as a measure of how much information about θ is conveyed by the sample X. If X happens to be a sample of size n, so that $X = (X_1, \ldots, X_n)$, then let us denote this by $I_n(\theta)$, so that, in particular, $I_1(\theta)$ would be the amount of information in a sample of size 1. Show that n times as much information is in a sample of size n than in a sample of size 1, i.e., show that $I_n(\theta) = nI_1(\theta)$.

Answer:

$$
\begin{aligned}
I_n(\theta) &= \textsf{\textbf{var}}_\theta\left(\frac{\partial}{\partial\theta}\log f_{X_1,\ldots,X_n}(X_1,\ldots,X_n \mid \theta)\right) \\
&= \textsf{\textbf{var}}_\theta\left(\frac{\partial}{\partial\theta}\log \prod_{i=1}^n f_{X_i}(X_i \mid \theta)\right) \text{ because of independence,} \\
&= \textsf{\textbf{var}}_\theta\left(\frac{\partial}{\partial\theta}\sum_{i=1}^n \log f_{X_i}(X_i \mid \theta)\right) \\
&= \textsf{\textbf{var}}_\theta\left(\sum_{i=1}^n \frac{\partial}{\partial\theta}\log f_{X_i}(X_i \mid \theta)\right) \\
&= \sum_{i=1}^n \textsf{\textbf{var}}_\theta\left(\frac{\partial}{\partial\theta}\log f_{X_i}(X_i \mid \theta)\right) \text{ because of independence,} \\
&= \sum_{i=1}^n \textsf{\textbf{var}}_\theta\left(\frac{\partial}{\partial\theta}\log f_{X_1}(X_1 \mid \theta)\right) \text{ because of identity of distributions,} \\
&= n\,\textsf{\textbf{var}}_\theta\left(\frac{\partial}{\partial\theta}\log f_{X_1}(X_1 \mid \theta)\right) \text{ because all } n \text{ terms are the same,} \\
&= nI_1(\theta).
\end{aligned}
$$

6. Use conclusions from #3 and #4 to show that the mean squared error of an unbiased estimator $T(X)$ of θ cannot be less than $1/I(\theta)$. (This is the "Cramer-Rao inequality" or the "information inequality." The quantity $1/I(\theta)$ is the "Cramer-Rao lower bound.")

Answer: According to #4 we have

$$\frac{\textsf{\textbf{var}}_{\theta_0}(\Lambda(X))}{(\theta_1 - \theta_0)^2} \longrightarrow I(\theta_0) \text{ as } \theta_1 \to \theta_0.$$

Since the reciprocal function is continuous (except at 0) we can infer that

$$\frac{(\theta_1 - \theta_0)^2}{\textsf{\textbf{var}}_{\theta_0}(\Lambda(X))} \longrightarrow \frac{1}{I(\theta_0)} \text{ as } \theta_1 \to \theta_0.$$

Then

Continued⟶

bring in the result of #3:

$$\text{mean squared error} \geq \frac{(\theta_1 - \theta_0)^2}{\mathsf{var}_{\theta_0}(\Lambda(X))} \longrightarrow \frac{1}{I(\theta_0)} \text{ as } \theta_1 \to \theta_0.$$

Since the mean squared error does not depend on θ_1, it does not change as θ_1 is approaching θ_0, and so if it is $\geq$ something that depends on θ_1, then it is $\geq$ the limit of that "something," as θ_1 approaches θ_0, and we conclude

$$\text{mean squared error} \geq \frac{1}{I(\theta_0)}.$$

7. The Fisher information $I(\theta)$ can sometimes be more readily computed by the result you will derive in this problem than by using the definition of it given in #5. In this problem you may assume that

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} \cdots\cdots \, dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} \cdots\cdots \, dx.$$

(This is valid if the function being integrated is sufficiently well-behaved.)

(a) Show that $\int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(x \mid \theta)\, dx = 0$.

Answer:

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(x \mid \theta)\, dx = \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(x \mid \theta)\, dx = \frac{\partial}{\partial\theta} 1 = 0.$$

(b) Show that $\int_{-\infty}^{\infty} \frac{\partial^2}{\partial\theta^2} f(x \mid \theta)\, dx = 0$.

Answer:

$$\int_{-\infty}^{\infty} \frac{\partial^2}{\partial\theta^2} f(x \mid \theta)\, dx = \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(x \mid \theta)\, dx = \frac{\partial}{\partial\theta} 0 = 0.$$

(c) Show that

$$\frac{\partial^2}{\partial\theta^2} \log f(x \mid \theta) = \frac{\frac{\partial^2}{\partial\theta^2} f(x \mid \theta)}{f(x \mid \theta)} - \left(\frac{\partial}{\partial\theta} \log f(x \mid \theta)\right)^2.$$

Then take expected values of all three terms, use the result of part (b) to show that one of the expected values is 0, and finally draw the conclusion that

$$I(\theta) = -\,\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2} \log f(X \mid \theta)\right).$$

Answer:

$$\frac{\partial^2}{\partial\theta^2}\log f(x\mid\theta) = \frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta}\log f(x\mid\theta)) \overset{\substack{\boxed{\text{Chain rule}}\\ \downarrow}}{=} \frac{\partial}{\partial\theta}\,\frac{(\partial/\partial\theta)f(x\mid\theta)}{f(x\mid\theta)}$$

$$\underset{\substack{\uparrow\\ \boxed{\text{Quotient rule}}}}{=} \frac{f(x\mid\theta)(\partial^2/\partial\theta^2)f(x\mid\theta) - ((\partial/\partial\theta)f(x\mid\theta))^2}{f(x\mid\theta)^2}$$

$$= \frac{\frac{\partial^2}{\partial\theta^2}f(x\mid\theta)}{f(x\mid\theta)} - \left(\frac{(\partial/\partial\theta)f(x\mid\theta)}{f(x\mid\theta)}\right)^2$$

$$= \frac{\frac{\partial^2}{\partial\theta^2}f(x\mid\theta)}{f(x\mid\theta)} - \left(\frac{\partial}{\partial\theta}\log f(x\mid\theta)\right)^2.$$

In order that it make sense to take expected values, we need "X" rather than "x":

$$\frac{\partial^2}{\partial\theta^2}\log f(X\mid\theta) = \frac{\frac{\partial^2}{\partial\theta^2}f(X\mid\theta)}{f(X\mid\theta)} - \left(\frac{\partial}{\partial\theta}\log f(X\mid\theta)\right)^2.$$

$$\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X\mid\theta)\right) = \underbrace{\mathbf{E}_\theta\left(\frac{\frac{\partial^2}{\partial\theta^2}f(X\mid\theta)}{f(X\mid\theta)}\right)}_{\substack{\uparrow\\ \boxed{\text{We will see that this term} = 0.}}} - \mathbf{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X\mid\theta)\right)^2\right).$$

The term immediately to the right of "=" is

$$\mathbf{E}_\theta\left(\frac{\frac{\partial^2}{\partial\theta^2}f(X\mid\theta)}{f(X\mid\theta)}\right) = \int_{-\infty}^{\infty}\frac{\frac{\partial^2}{\partial\theta^2}f(x\mid\theta)}{f(x\mid\theta)}\,f(x\mid\theta)\,dx = \int_{-\infty}^{\infty}\frac{\partial^2}{\partial\theta^2}f(x\mid\theta)\,dx$$

and we showed in (b) that that is 0. Using the identity $\mathbf{E}(V^2) = \mathbf{var}(V) + (\mathbf{E}(V))^2$ we see that

$$-\mathbf{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X\mid\theta)\right)^2\right) = -\,\mathbf{var}_\theta\left(\frac{\partial}{\partial\theta}\log f(X\mid\theta)\right) - \left(\mathbf{E}_\theta\left(\frac{\partial}{\partial\theta}\log f(X\mid\theta)\right)\right)^2.$$

The second term above is $-I(\theta)$, so it suffices to show that the last term is 0. We have

$$\mathbf{E}_\theta\left(\frac{\partial}{\partial\theta}\log f(X\mid\theta)\right) = \int_{-\infty}^{\infty}\frac{(\partial/\partial\theta)f(x\mid\theta)}{f(x\mid\theta)}\,f(x\mid\theta)\,dx = 0 \text{ by part (a).}$$

8. Suppose $X_1,\ldots,X_n \sim \text{i.i.d.}\,N(\theta,1^2)$. The result of #6 then says no function of $X_1,\ldots,X_n$ that is an unbiased estimator of θ can have smaller variance that <u>something</u>. Find and simplify what goes in the role of "something" in the previous sentence. You can use the result of #7(c), but you are not required to do so.

   <u>Answer</u>:

   By the result of # 5.

$$\begin{aligned}
I_n(\theta) = nI_1(\theta) &= n\,\mathbf{var}_\theta\left(\frac{\partial}{\partial\theta}\log f_{X_1}(X_1\mid\theta)\right) = n\,\mathbf{var}_\theta\left(\frac{\partial}{\partial\theta}\log f_{X_1}(X_1\mid\theta)\right) \\
&= n\,\mathbf{var}_\theta\left(\frac{\partial}{\partial\theta}\log\left\{[\text{constant}]\cdot\exp\left(-1(X_1-\theta)^2/2\right)\right\}\right) \\
&= n\,\mathbf{var}_\theta\left(\frac{\partial}{\partial\theta}\left\{[\text{constant}]-\frac{1}{2}(X_1-\theta)^2\right\}\right) \\
&= n\,\mathbf{var}_\theta\left(X_1-\theta\right) = n\cdot 1^2 = n.
\end{aligned}$$

   Consequently the lower bound on the variance of unbiased estimators of θ that are functions of $X_1,\ldots,X_n$ given by the result of #6 is

$$\frac{1}{I_n(\theta)} = \frac{1}{nI_1(\theta)} = \frac{1}{n\cdot 1^2} = \frac{1}{n}.$$

   If we had used the result of #7, we would have done this:

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\log f_{X_1}(X_1\mid\theta) &= \frac{\partial^2}{\partial\theta^2}\log\left\{[\text{constant}]\cdot\exp\left(-1(X_1-\theta)^2/2\right)\right\} \\
&= \frac{\partial^2}{\partial\theta^2}\left\{[\text{constant}]-\frac{1}{2}(X_1-\theta)^2\right\} \\
&= -1.
\end{aligned}$$

   Then $I_1(\theta) = -E_\theta(-1) = 1$, and so $I_n(\theta) = nI_1(\theta) = n\cdot 1 = n$.

9. Suppose $X_1,\ldots,X_n \sim \text{i.i.d.}\,\text{Bin}(10,\theta)$. Answer the same question as in #8. (The results you use work just as well if the probability distributions involved are discrete.)

Answer:

$$I_n(\theta) = nI_1(\theta) = -n\,\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f_{X_1}(X_1 \mid \theta)\right) = -n\,\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log\binom{10}{X_1}\theta^{X_1}(1-\theta)^{10-X_1}\right)$$

$$= -n\,\mathbf{E}_\theta\left(-\frac{X_1}{\theta^2} - \frac{10-X_1}{(1-\theta)^2}\right) = n\left(\frac{\mathbf{E}_\theta(X_1)}{\theta^2} + \frac{10-\mathbf{E}_\theta(X_1)}{(1-\theta)^2}\right) = n\left(\frac{10\theta}{\theta^2} + \frac{10-10\theta}{(1-\theta)^2}\right)$$

$$= 10n\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) = \frac{10n}{\theta(1-\theta)}. \qquad \text{So the lower bound is } \frac{\theta(1-\theta)}{10n}.$$

No unbiased estimator of θ can have a mean squared error smaller than that.

10. Suppose $X_1, \ldots, X_n \sim \text{i.i.d. Poisson}(\theta)$. Answer the same question as in #8.

Answer: Again we are working with a discrete distribution.

$$I_n(\theta) = nI_1(\theta) = -n\,\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f_{X_1}(X_1 \mid \theta)\right) = -n\,\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log\frac{e^{-\theta}\theta^{X_1}}{X_1!}\right)$$

$$= -n\,\mathbf{E}_\theta\left(\frac{-X_1}{\theta^2}\right) = \frac{n\,\mathbf{E}_\theta(X_1)}{\theta^2} = \frac{n\theta}{\theta^2} = \frac{n}{\theta}.$$

Therefore no unbiased estimator of θ can have a smaller mean squared error than $\frac{\theta}{n}$.

11. Suppose $X_1, \ldots, X_n \sim$ i.i.d. have a "memoryless" (continuous) exponential distribution with expected value θ, i.e., each X_i is distributed as the (continuous) waiting time in a Poisson process with intensity $1/\theta$ occurrences per unit time. Answer the same question as in #8.

Answer: We're back to continuous distributions.

$$I_n(\theta) = nI_1(\theta) = -n\,\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f_{X_1}(X_1 \mid \theta)\right) = -n\,\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log\left(\frac{1}{\theta}e^{-X_1/\theta}\right)\right)$$

$$= -n\,\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\left(-\log\theta - \frac{X_1}{\theta}\right)\right) = n\,\mathbf{E}_\theta\left(\frac{-1}{\theta^2} + \frac{2X_1}{\theta^3}\right) = n\left(\frac{-1}{\theta^2} + \frac{2\,\mathbf{E}_\theta(X_1)}{\theta^3}\right)$$

$$= n\left(\frac{-1}{\theta^2} + \frac{2\theta}{\theta^3}\right) = \frac{n}{\theta^2}.$$

Therefore no unbiased estimator of θ can have a smaller mean squared error than $\frac{\theta^2}{n}$.

12. Consider testing the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta = \theta_1$.

The likelihood-ratio test rejects the null hypothesis if and only if $\Lambda(X) > c$ (where $c =$ the "critical value"). Suppose this test's probability of Type I error is 3%, i.e.,

$\mathbf{Pr}(\text{this test rejects } H_0 \mid H_0) = 0.03$. The power of the test is $\mathbf{Pr}(\text{this test rejects } H_0 \mid H_1)$.

Some other test, based on some statistic $K(X)$ different from $\Lambda(X)$, rejects the null hypothesis if and only if $K(X) > d$. Suppose this test's probability of Type I error is also 3%, i.e., $\mathbf{Pr}(\text{this test rejects } H_0 \mid H_0) = 0.03$. The power of the test is $\mathbf{Pr}(\text{this test rejects } H_0 \mid H_1)$.

These are different tests. That means one test could reject H_0 while the other does not, with the same data, and the probability that the two tests disagree on whether to reject H_0 is more than 0.

$$\begin{aligned}\text{Let } A &= [\text{the likelihood-ratio test rejects } H_0 \text{ and the other test does not}],\\ B &= [\text{the other test rejects } H_0 \text{ and the likelihood-ratio test does not}],\\ \text{and } C &= [\text{both tests reject } H_0].\end{aligned}$$

(a) Explain and justify each of the seven relations marked by "?" below.

Answer: I have changed "?" to A, B, C, D, E, F, and G.

$$\begin{aligned}\mathbf{Pr}(A \mid H_1) &\overset{\boxed{A}}{=} \int_A f(x \mid \theta=\theta_1)\,dx \overset{\boxed{B}}{>} \int_A c\,f(x \mid \theta=\theta_0)\,dx \overset{\boxed{C}}{=} c\,\mathbf{Pr}(A \mid H_0)\\ &\underset{\boxed{D}}{=} c\,\mathbf{Pr}(B \mid H_0) \underset{\boxed{E}}{=} \int_B c\,f(x \mid \theta=\theta_0)\,dx \underset{\boxed{F}}{\geq} \int_B f(x \mid \theta=\theta_1)\,dx \underset{\boxed{G}}{=} \mathbf{Pr}(B \mid H_1).\end{aligned}$$

A: The symbol "$\int_A$" must be taken to mean the integral over the set of all values of x for which the event A occurs. The integral of the probability density function of a random variable X over any set, is the probability that that random variable is within that set. The density is conditional upon the hypothesis that $\theta = \theta_1$ because the probability to the left of "=" is conditional upon H_1.

B: The event A is the event that the likelihood-ratio test reject H_0 and the other test does not. Therefore, if $x \in A$, then the value of the likelihood-ratio test statistic is such that that test rejects H_0. That means that when $x \in A$ then $\dfrac{f(x \mid \theta=\theta_1)}{f(x \mid \theta=\theta_0)} > c$, and hence $f(x \mid \theta=\theta_1) > c\,f(x \mid \theta=\theta_0)$.

C: The justification of this equality is the same as in "**A**" above.

D: We know that

$$\mathbf{Pr}(\text{the likelihood-ratio test rejects } H_0 \mid H_0) = P(\text{the other test rejects } H_0 \mid H_0).$$

The verbal statement "The likelihood-ratio test rejects H_0" is equivalent to

"$\left[\begin{array}{c}\text{The likelihood-ratio test rejects } H_0\\ \text{and the other test does not}\end{array}\right]$ **or** $\big[$ both tests reject $H_0\big]$"

which is to say, it is equivalent to "A or C". Similarly, the statement "The other test rejects H_0" is equivalent to "B or C". Consequently we conclude that

$$\mathbf{Pr}(A \text{ or } C \mid H_0) = \mathbf{Pr}(B \text{ or } C \mid H_0).$$

Then, by mutual exclusivity,

$$\mathbf{Pr}(A \mid H_0) + \mathbf{Pr}(C \mid H_0) = \mathbf{Pr}(B \mid H_0) + \mathbf{Pr}(C \mid H_0).$$

Consequently

$$\mathbf{Pr}(A \mid H_0) = \mathbf{Pr}(B \mid H_0).$$

E: Again, the justification is the same as in "**A**".

F: Just as in "**B**", except that the test does not reject H_0.

G: Just as in "**A**".

(b) By thinking about $\mathbf{Pr}(A \text{ or } C \mid H_1)$ and $\mathbf{Pr}(B \text{ or C} \mid H_1)$, explain why the conclusion of part (a) can be summarized by saying that the likelihood-ratio test is better than the other test. (This is the **"Neyman-Pearson lemma."**)

Answer: We just showed

$$\mathbf{Pr}(A \mid H_1) > \mathbf{Pr}(B \mid H_1).$$

It follows that

$$\mathbf{Pr}(A \text{ or } C \mid H_1) > \mathbf{Pr}(B \text{ or } C \mid H_1).$$

We saw above that "A or C" is equivalent to "The likelihood-ratio test rejects H_0" and "B or C" is equivalent to "The other test rejects H_0." Consequently we conlude:

$$\mathbf{Pr}(\text{The likelihood-ratio test rejects } H_0 \mid H_1) > \mathbf{Pr}(\text{The other test rejects } H_0 \mid H_1).$$

The two tests have equal probabilities of Type I error, but the likelihood-ratio test has a smaller probability of Type II error. In other words, the likelihood-ratio test is the most powerful test among all tests with a particular probability of Type I error.