Betting Score and Testing Information

Xiao-Li Meng

Harvard University

RSS - September 9, 2020

A Bayesian Interpretation of the Betting Score

- $E_P[S(Y)] = 1$ implies that $Q(y) = S(y)P(y)$ is a PF/PDF.

A Bayesian Interpretation of the Betting Score

- $E_P[S(Y)] = 1$ implies that $Q(y) = S(y)P(y)$ is a PF/PDF.
- Testing using $S(y) =$ LR test against Q.

A Bayesian Interpretation of the Betting Score

- $E_P[S(Y)] = 1$ implies that $Q(y) = S(y)P(y)$ is a PF/PDF.
- Testing using $S(y) =$ LR test against Q.
- Suppose $P(y)$ represents a composite null $\{P_\theta(y), \theta \in \Theta\}$, then $Q(Y)$ is a composite alternative $\{Q_\theta(y), \theta \in \Theta\}$.

A Bayesian Interpretation of the Betting Score

- $E_P[S(Y)] = 1$ implies that $Q(y) = S(y)P(y)$ is a PF/PDF.
- Testing using $S(y) =$ LR test against Q.
- Suppose $P(y)$ represents a composite null $\{P_\theta(y), \theta \in \Theta\}$, then $Q(Y)$ is a composite alternative $\{Q_\theta(y), \theta \in \Theta\}$.
- Suppose we have a (proper) prior $\pi(\theta)$, then

$$\mathrm{E}_{P,\pi}\left[\frac{Q_\theta(y)}{P_\theta(y)}\bigg| y\right] = \int \frac{Q_\theta(y)}{P_\theta(y)} \frac{P_\theta(y)\pi(\theta)}{P(y)} d\theta$$

$$= \frac{Q(y)}{P(y)} = \text{Bayes Factor}$$

A Bayesian Interpretation of the Betting Score

- $E_P[S(Y)] = 1$ implies that $Q(y) = S(y)P(y)$ is a PF/PDF.
- Testing using $S(y) =$ LR test against Q.
- Suppose $P(y)$ represents a composite null $\{P_\theta(y), \theta \in \Theta\}$, then $Q(Y)$ is a composite alternative $\{Q_\theta(y), \theta \in \Theta\}$.
- Suppose we have a (proper) prior $\pi(\theta)$, then

$$\mathrm{E}_{P,\pi}\left[\frac{Q_\theta(y)}{P_\theta(y)}\middle| y\right] = \int \frac{Q_\theta(y)}{P_\theta(y)} \frac{P_\theta(y)\pi(\theta)}{P(y)} d\theta$$
$$= \frac{Q(y)}{P(y)} = \text{Bayes Factor}$$

- A Caveat: It requires $S(Y)$ to be first-order ancillary: $\mathrm{E}_{P_\theta}[S(Y)] = 1$.

Create Alternatives for Measuring "Testing Information" ...

- For an estimator $\hat{\theta}(y)$, we can bound/estimate its information under $P_\theta(y)$ by the Cramér-Rao bound/Fisher Information.

Create Alternatives for Measuring "Testing Information" ...

- For an estimator $\hat{\theta}(y)$, we can bound/estimate its information under $P_\theta(y)$ by the Cramér-Rao bound/Fisher Information.
- But how do we measure "information" in a test statistic $T(y)$, when all we have is $P_0(y)$?

Create Alternatives for Measuring "Testing Information" ...

- For an estimator $\hat{\theta}(y)$, we can bound/estimate its information under $P_\theta(y)$ by the Cramér-Rao bound/Fisher Information.
- But how do we measure "information" in a test statistic $T(y)$, when all we have is $P_0(y)$?
- Create "local alternative" via exponential titling

$$P_\theta(y) = P_0(y) exp\{\theta T(y)\}/c_\theta$$

Create Alternatives for Measuring "Testing Information" ...

- For an estimator $\hat{\theta}(y)$, we can bound/estimate its information under $P_\theta(y)$ by the Cramér-Rao bound/Fisher Information.
- But how do we measure "information" in a test statistic $T(y)$, when all we have is $P_0(y)$?
- Create "local alternative" via exponential titling

$$P_\theta(y) = P_0(y) exp\{\theta T(y)\}/c_\theta$$

- T then is the same as Log LR test.

Create Alternatives for Measuring "Testing Information" ...

- For an estimator $\hat{\theta}(y)$, we can bound/estimate its information under $P_\theta(y)$ by the Cramér-Rao bound/Fisher Information.
- But how do we measure "information" in a test statistic $T(y)$, when all we have is $P_0(y)$?
- Create "local alternative" via exponential titling

$$P_\theta(y) = P_0(y) exp\{\theta T(y)\}/c_\theta$$

- T then is the same as Log LR test.
- Nicolae, Meng and Kong (2008) **Quantifying the Fraction of Missing Information for Hypothesis Testing in Statistical and Genetic Studies.** *Statistical Science* 2008, 287–312.

Create Alternatives for Measuring "Testing Information" ...

- For an estimator $\hat{\theta}(y)$, we can bound/estimate its information under $P_\theta(y)$ by the Cramér-Rao bound/Fisher Information.
- But how do we measure "information" in a test statistic $T(y)$, when all we have is $P_0(y)$?
- Create "local alternative" via exponential titling

$$P_\theta(y) = P_0(y) exp\{\theta T(y)\}/c_\theta$$

- T then is the same as Log LR test.
- Nicolae, Meng and Kong (2008) **Quantifying the Fraction of Missing Information for Hypothesis Testing in Statistical and Genetic Studies.** *Statistical Science* 2008, 287–312.
- Jones and Meng (2020) **Duality-induced Information Design for Distinguishing Between Models or Hypotheses** Revision for *JRSSB*.

A Powerful Identity

- An Extended "EM Identity":

$$\mathrm{E}\left[\log \frac{P_{\theta_1}(Y_{com})}{P_{\theta_2}(Y_{com})}\middle| Y_{obs}, \theta\right] = \log \frac{P_{\theta_1}(Y_{obs})}{P_{\theta_2}(Y_{obs})} + \mathrm{E}\left[\log \frac{P_{\theta_1}(Y_{com}|Y_{obs})}{P_{\theta_2}(Y_{com}|Y_{obs})}\middle| Y_{obs}, \theta\right]$$

A Powerful Identity

- An Extended "EM Identity":

$$\mathrm{E}\left[\log \frac{P_{\theta_1}(Y_{com})}{P_{\theta_2}(Y_{com})}\middle| Y_{obs}, \theta\right] = \log \frac{P_{\theta_1}(Y_{obs})}{P_{\theta_2}(Y_{obs})} + \mathrm{E}\left[\log \frac{P_{\theta_1}(Y_{com}|Y_{obs})}{P_{\theta_2}(Y_{com}|Y_{obs})}\middle| Y_{obs}, \theta\right]$$

- By choosing $\theta = \theta_1$, we obtain the Gibbs's inequality.

A Powerful Identity

- An Extended "EM Identity":

$$\mathrm{E}\left[\log \frac{P_{\theta_1}(Y_{com})}{P_{\theta_2}(Y_{com})}\middle| Y_{obs}, \theta\right] = \log \frac{P_{\theta_1}(Y_{obs})}{P_{\theta_2}(Y_{obs})} + \mathrm{E}\left[\log \frac{P_{\theta_1}(Y_{com}|Y_{obs})}{P_{\theta_2}(Y_{com}|Y_{obs})}\middle| Y_{obs}, \theta\right]$$

- By choosing $\theta = \theta_1$, we obtain the Gibbs's inequality.
- It establishes multiple (relative) information measures as given in Nicole, Meng, and Kong (2018)

The Second C-R lower Bound

- Suppose $\mathrm{E}_{P_\theta}[T(Y)] = \tau(\theta)$. Then

$$\mathrm{Var}_\theta(T(Y)) \geq \sup_\psi \frac{[\tau(\psi) - \tau(\theta)]^2}{\mathrm{Var}(LR(\psi, \theta | Y) | \theta)}$$

The Second C-R lower Bound

- Suppose $\mathrm{E}_{P_\theta}[T(Y)] = \tau(\theta)$. Then

$$\mathrm{Var}_\theta(T(Y)) \geq \sup_\psi \frac{[\tau(\psi) - \tau(\theta)]^2}{\mathrm{Var}(LR(\psi, \theta | Y) | \theta)}$$

- The Statistical "Heisenberg's uncertainty principle"

The Second C-R lower Bound

- Suppose $\mathrm{E}_{P_\theta}[T(Y)] = \tau(\theta)$. Then

$$\mathrm{Var}_\theta(T(Y)) \geq \sup_\psi \frac{[\tau(\psi) - \tau(\theta)]^2}{\mathrm{Var}(LR(\psi, \theta | Y) | \theta)}$$

- The Statistical "Heisenberg's uncertainty principle"
- Chapman, D. C. and Robbins, H. (1951). Minimum variance estimation without regularity assumptions. *Ann. Math. Statist.* 22 581–586.

The Second C-R lower Bound

- Suppose $\mathrm{E}_{P_\theta}[T(Y)] = \tau(\theta)$. Then

$$\mathrm{Var}_\theta(T(Y)) \geq \sup_\psi \frac{[\tau(\psi) - \tau(\theta)]^2}{\mathrm{Var}(LR(\psi, \theta|Y)|\theta)}$$

- The Statistical "Heisenberg's uncertainty principle"
- Chapman, D. C. and Robbins, H. (1951). Minimum variance estimation without regularity assumptions. *Ann. Math. Statist.* 22 581–586.
- For betting score:

$$E_Q(S) = 1 + \mathrm{Var}_P(S) = 1 + \mathrm{Var}\left(LR(Q, P|Y)|P\right).$$

The Second C-R lower Bound

- Suppose $\mathrm{E}_{P_\theta}[T(Y)] = \tau(\theta)$. Then

$$\mathrm{Var}_\theta(T(Y)) \geq \sup_\psi \frac{[\tau(\psi) - \tau(\theta)]^2}{\mathrm{Var}(LR(\psi, \theta|Y)|\theta)}$$

- The Statistical "Heisenberg's uncertainty principle"
- Chapman, D. C. and Robbins, H. (1951). Minimum variance estimation without regularity assumptions. *Ann. Math. Statist.* 22 581–586.
- For betting score:

$$E_Q(S) = 1 + \mathrm{Var}_P(S) = 1 + \mathrm{Var}\left(LR(Q, P|Y)|P\right).$$

- The larger the $\mathrm{Var}\left(LR(Q, P|Y)|P\right)$, the more information for separating Q from P, and hence more evidence against P.