Safe models for risky decisions

Steingröver, H.M.

Appendix F

Appendix to Chapter 8: "A Tutorial on Bridge Sampling"

F.1 The Bridge Sampling Estimator as a General Case of Methods 1 – 3

In this section we show that the naive Monte Carlo, the importance sampling, and the generalized harmonic mean estimators are special cases of the bridge sampling estimator under specific choices of the bridge function $h(\theta)$ and the proposal distribution $g(\theta)$.[1] An overview is provided in Table F.1.

To prove that the bridge sampling estimator reduces to the naive Monte Carlo estimator, consider bridge sampling, choose the prior distribution as the proposal distribution (i.e., $g(\theta) = p(\theta)$), and specify the bridge function as $h(\theta) = 1/g(\theta)$. Inserting these specifications into Equation 8.12 yields:

$$\hat{p}_4\Big(y \mid h(\theta) = \frac{1}{g(\theta)},\ g(\theta) = p(\theta)\Big) = \frac{\frac{1}{N_2}\sum_{i=1}^{N_2} \dfrac{1}{p(\tilde{\theta}_i)}\, p(y \mid \tilde{\theta}_i)\, p(\tilde{\theta}_i)}{\frac{1}{N_1}\sum_{j=1}^{N_1} \dfrac{1}{p(\theta_j^*)}\, p(\theta_j^*)}, \quad \tilde{\theta}_i \sim p(\theta), \quad \theta_j^* \sim p(\theta \mid y)$$

$$= \frac{\frac{1}{N_2}\sum_{i=1}^{N_2} p(y \mid \tilde{\theta}_i)}{\frac{1}{N_1} N_1} = \frac{1}{N_2}\sum_{i=1}^{N_2} p(y \mid \tilde{\theta}_i)\ , \quad \tilde{\theta}_i \sim p(\theta),$$

which is equivalent to the naive Monte Carlo estimator shown in Equation 8.6.

To prove that the bridge sampling estimator reduces to the importance sampling estimator, consider bridge sampling, choose the importance density as the proposal distribution (i.e., $g(\theta) =$

[1]Note that bridge sampling is also a general case of the Chib and Jeliazkov (2001) method of estimating the marginal likelihood using the Metropolis-Hastings acceptance probability (Meng & Schilling, 2002; Mira & Nicholls, 2004).

Table F.1: *Summary of the Bridge Sampling Estimators for the Marginal Likelihood, and Its Special Cases: the Naive Monte Carlo, Importance Sampling, and Generalized Harmonic Mean Estimator*

Method	Estimator	Samples	Bridge Function $h(\theta)$
Bridge sampling	$\dfrac{\frac{1}{N_2}\sum_{i=1}^{N_2} p(y \mid \tilde{\theta}_i)\, p(\tilde{\theta}_i)\, h(\tilde{\theta}_i)}{\frac{1}{N_1}\sum_{j=1}^{N_1} h(\theta_j^*)\, g(\theta_j^*)}$	$\tilde{\theta}_i \sim g(\theta)$ $\theta_j^* \sim p(\theta \mid y)$	$\dfrac{C}{\frac{N_1}{N_2+N_1} p(y \mid \theta) p(\theta) + \frac{N_2}{N_2+N_1} p(y) g(\theta)}$
Naive Monte Carlo	$\dfrac{1}{N}\sum_{i=1}^{N} p(y \mid \tilde{\theta}_i)$	$\tilde{\theta}_i \sim p(\theta)$	$\dfrac{1}{g(\theta)}$ and $g(\theta) = p(\theta)$
Importance sampling	$\dfrac{1}{N}\sum_{i=1}^{N} \dfrac{p(y \mid \tilde{\theta}_i)\, p(\tilde{\theta}_i)}{g_{IS}(\tilde{\theta}_i)}$	$\tilde{\theta}_i \sim g_{IS}(\theta)$	$\dfrac{1}{g_{IS}(\theta)}$ and $g(\theta) = g_{IS}(\theta)$
Generalized harmonic mean	$\left(\dfrac{1}{N}\sum_{i=1}^{N} \dfrac{g_{IS}(\theta_i^*)}{p(y \mid \theta_i^*)\, p(\theta_i^*)}\right)^{-1}$	$\theta_i^* \sim p(\theta \mid y)$	$\dfrac{1}{p(y \mid \theta) p(\theta)}$ and $g(\theta) = g_{IS}(\theta)$

Note. $p(\theta)$ is the prior distribution, $g_{IS}(\theta)$ is the importance density, $p(\theta|y)$ is the posterior distribution, $g(\theta)$ is the proposal distribution, $h(\theta)$ is the bridge function, and C is a constant. The last column shows the bridge function needed to obtain the special cases.

$g_{IS}(\theta)$), and specify the bridge function as $h(\theta) = 1/g_{IS}(\theta)$. Inserting these specifications into Equation 8.12 yields:

$$\hat{p}_4\Big(y \mid h(\theta) = \frac{1}{g_{IS}(\theta)},\ g(\theta) = g_{IS}(\theta)\Big) = \frac{\frac{1}{N_2}\sum_{i=1}^{N_2} \dfrac{1}{g_{IS}(\tilde{\theta}_i)}\, p(y \mid \tilde{\theta}_i)\, p(\tilde{\theta}_i)}{\frac{1}{N_1}\sum_{j=1}^{N_1} \dfrac{1}{g_{IS}(\theta_j^*)}\, g_{IS}(\theta_j^*)}, \quad \tilde{\theta}_i \sim g_{IS}(\theta), \quad \theta_j^* \sim p(\theta \mid y)$$

$$= \frac{\frac{1}{N_2}\sum_{i=1}^{N_2} \dfrac{p(y \mid \tilde{\theta}_i)\, p(\tilde{\theta}_i)}{g_{IS}(\tilde{\theta}_i)}}{\frac{1}{N_1} N_1} = \frac{1}{N_2}\sum_{i=1}^{N_2} \frac{p(y \mid \tilde{\theta}_i)\, p(\tilde{\theta}_i)}{g_{IS}(\tilde{\theta}_i)}\,, \quad \tilde{\theta}_i \sim g_{IS}(\theta),$$

which is equivalent to the importance sampling estimator shown in Equation 8.7.

To prove that the bridge sampling estimator reduces to the generalized harmonic mean estimator, consider bridge sampling, choose the importance density as the proposal distribution (i.e., $g(\theta) = g_{IS}(\theta)$), and specify the bridge function as $h(\theta) = 1/(p(y \mid \theta)\, p(\theta))$. Inserting these specifications into Equation 8.12 yields:

$$\hat{p}_4\Big(y \mid h(\theta) = \frac{1}{p(y \mid \theta)\, p(\theta)},\ g(\theta) = g_{IS}(\theta)\Big)$$

$$= \frac{\frac{1}{N_2}\sum_{i=1}^{N_2} \dfrac{1}{p(y \mid \tilde{\theta}_i)\, p(\tilde{\theta}_i)}\, p(\tilde{\theta}_i)\, p(y \mid \tilde{\theta}_i)}{\frac{1}{N_1}\sum_{j=1}^{N_1} \dfrac{1}{p(y \mid \theta_j^*)\, p(\theta_j^*)}\, g_{IS}(\theta_j^*)}, \quad \tilde{\theta}_i \sim g(\theta), \quad \theta_j^* \sim p(\theta \mid y)$$

$$= \frac{\frac{1}{N_2}\sum_{i=1}^{N_2} 1}{\frac{1}{N_1}\sum_{j=1}^{N_1} \dfrac{g_{IS}(\theta_j^*)}{p(y \mid \theta_j^*)\, p(\theta_j^*)}} = \left(\frac{1}{N_1}\sum_{j=1}^{N_1} \frac{g_{IS}(\theta_j^*)}{p(y \mid \theta_j^*)\, p(\theta_j^*)}\right)^{-1}, \ \theta_j^* \sim p(\theta \mid y),$$

which is equivalent to the generalized harmonic mean estimator shown in Equation 8.8.

F.2 Bridge Sampling Implementation: Avoiding Numerical Issues

In order to avoid numerical issues, we can rewrite Equation 8.15 in the following way:

$$\hat{p}_4(y)^{(t+1)} = \frac{\frac{1}{N_2}\sum\limits_{i=1}^{N_2} \frac{l_{2,i}}{s_1\, l_{2,i} + s_2\, \hat{p}_4(y)^{(t)}}}{\frac{1}{N_1}\sum\limits_{j=1}^{N_1} \frac{1}{s_1\, l_{1,j} + s_2\, \hat{p}_4(y)^{(t)}}}$$

$$= \frac{\frac{1}{N_2}\sum\limits_{i=1}^{N_2} \frac{\exp\big(\log(l_{2,i})\big)}{s_1 \exp\big(\log(l_{2,i})\big) + s_2 \hat{p}_4(y)^{(t)}}}{\frac{1}{N_1}\sum\limits_{j=1}^{N_1} \frac{1}{s_1 \exp\big(\log(l_{1,j})\big) + s_2 \hat{p}_4(y)^{(t)}}}$$

$$= \frac{\frac{1}{N_2}\sum\limits_{i=1}^{N_2} \frac{\exp\big(\log(l_{2,i})\big)\exp\big(-l^*\big)}{s_1 \exp\big(\log(l_{2,i})\big)\exp\big(-l^*\big) + s_2 \hat{p}_4(y)^{(t)} \exp\big(-l^*\big)}}{\frac{1}{N_1}\sum\limits_{j=1}^{N_1} \frac{\exp\big(-l^*\big)}{s_1 \exp\big(\log(l_{1,j})\big)\exp\big(-l^*\big) + s_2 \hat{p}_4(y)^{(t)} \exp\big(-l^*\big)}}$$

$$= \frac{1}{\exp\big(-l^*\big)} \frac{\frac{1}{N_2}\sum\limits_{i=1}^{N_2} \frac{\exp\big(\log(l_{2,i}) - l^*\big)}{s_1 \exp\big(\log(l_{2,i}) - l^*\big) + s_2 \hat{p}_4(y)^{(t)} \exp\big(-l^*\big)}}{\frac{1}{N_1}\sum\limits_{j=1}^{N_1} \frac{1}{s_1 \exp\big(\log(l_{1,j}) - l^*\big) + s_2 \hat{p}_4(y)^{(t)} \exp\big(-l^*\big)}}$$

$$= \exp\left(l^*\right) \frac{\frac{1}{N_2} \sum_{i=1}^{N_2} \frac{\exp\left(\log(l_{2,i}) - l^*\right)}{s_1 \exp\left(\log(l_{2,i}) - l^*\right) + s_2 \hat{p}_4(y)^{(t)} \exp\left(-l^*\right)}}{\frac{1}{N_1} \sum_{j=1}^{N_1} \frac{1}{s_1 \exp\left(\log(l_{1,j}) - l^*\right) + s_2 \hat{p}_4(y)^{(t)} \exp\left(-l^*\right)}}.$$

l^* is a constant which we can choose in a way that keeps the terms in the sums manageable. We used $l^* = \text{median}(\log(l_{1,j}))$. Let

$$\hat{r}^{(t)} = \hat{p}_4(y)^{(t)} \exp\left(-l^*\right),$$

so that

$$\hat{p}_4(y)^{(t)} = \hat{r}^{(t)} \exp\left(l^*\right).$$

Then we obtain

$$\hat{p}_4(y)^{(t+1)} = \exp\left(l^*\right) \frac{\frac{1}{N_2} \sum_{i=1}^{N_2} \frac{\exp\left(\log(l_{2,i}) - l^*\right)}{s_1 \exp\left(\log(l_{2,i}) - l^*\right) + s_2 \hat{r}^{(t)}}}{\frac{1}{N_1} \sum_{j=1}^{N_1} \frac{1}{s_1 \exp\left(\log(l_{1,j}) - l^*\right) + s_2 \hat{r}^{(t)}}}$$

$$\hat{p}_4(y)^{(t+1)} \exp\left(-l^*\right) = \frac{\frac{1}{N_2} \sum_{i=1}^{N_2} \frac{\exp\left(\log(l_{2,i}) - l^*\right)}{s_1 \exp\left(\log(l_{2,i}) - l^*\right) + s_2 \hat{r}^{(t)}}}{\frac{1}{N_1} \sum_{j=1}^{N_1} \frac{1}{s_1 \exp\left(\log(l_{1,j}) - l^*\right) + s_2 \hat{r}^{(t)}}}$$

$$\hat{r}^{(t+1)} = \frac{\frac{1}{N_2} \sum_{i=1}^{N_2} \frac{\exp\left(\log(l_{2,i}) - l^*\right)}{s_1 \exp\left(\log(l_{2,i}) - l^*\right) + s_2 \hat{r}^{(t)}}}{\frac{1}{N_1} \sum_{j=1}^{N_1} \frac{1}{s_1 \exp\left(\log(l_{1,j}) - l^*\right) + s_2 \hat{r}^{(t)}}}.$$

Hence, we can run the iterative scheme with respect to $\hat{r}$ which is more convenient because it keeps the terms in the sums manageable and multiply the result by $\exp\left(l^*\right)$ to obtain the estimate of the marginal likelihood or, equivalently, we can take the logarithm of the result and add l^* to obtain an estimate of the logarithm of the marginal likelihood.

F.3 Correction for the Probit Transformation

In this section we describe how the probit transformation affects our expression of the generalized harmonic mean estimator (Equation 8.8) to yield Equation 8.9. Recall that we derived the generalized harmonic mean estimator using the following equality:

$$\frac{1}{p(y)} = \int \frac{g_{IS}(\theta)}{p(y \mid \theta) p(\theta)}\, p(\theta \mid y)\, \mathrm{d}\theta. \tag{F.1}$$

For practical reasons, in the running example, we used a normal distribution on ξ as importance density. ξ was obtained by probit-transforming θ (i.e., $\xi = \Phi^{-1}(\theta)$). In particular, the normal importance density was given by $\frac{1}{\hat{\sigma}} \phi\left(\frac{\xi - \hat{\mu}}{\hat{\sigma}}\right)$. Note that this importance density is a function of ξ, whereas the general importance density g_{IS} in Equation F.1 is specified in terms of θ. Therefore, to include our specific importance density to Equation F.1, we need to write it in terms of θ. This yields $\frac{1}{\hat{\sigma}}\phi\left(\frac{\Phi^{-1}(\theta)-\hat{\mu}}{\hat{\sigma}}\right)\frac{1}{\phi(\Phi^{-1}(\theta))}$, where the latter factor comes from applying the change-of-variable method. Replacing $g_{IS}(\theta)$ in Equation F.1 by this expression, results in:

$$\begin{aligned}\frac{1}{p(y)} &= \int \frac{\frac{1}{\hat{\sigma}}\phi\left(\frac{\Phi^{-1}(\theta)-\hat{\mu}}{\hat{\sigma}}\right)\frac{1}{\phi(\Phi^{-1}(\theta))}}{p(y \mid \theta)p(\theta)}\, p(\theta \mid y)\, \mathrm{d}\theta \\ &= \mathbb{E}_{\text{post}}\left(\frac{\frac{1}{\hat{\sigma}}\phi\left(\frac{\Phi^{-1}(\theta)-\hat{\mu}}{\hat{\sigma}}\right)\frac{1}{\phi(\Phi^{-1}(\theta))}}{p(y \mid \theta)\, p(\theta)}\right).\end{aligned} \tag{F.2}$$

Rewriting results in:

$$p(y) = \left(\mathbb{E}_{\text{post}}\left(\frac{\frac{1}{\hat{\sigma}}\phi\left(\frac{\Phi^{-1}(\theta)-\hat{\mu}}{\hat{\sigma}}\right)\frac{1}{\phi(\Phi^{-1}(\theta))}}{p(y \mid \theta)\, p(\theta)}\right)\right)^{-1},$$

which can be approximated as:

$$\begin{aligned}\hat{p}_3(y) &= \left(\frac{1}{N}\sum_{j=1}^{N}\frac{\overbrace{\frac{1}{\hat{\sigma}}\phi\left(\frac{\Phi^{-1}(\theta_j^*)-\hat{\mu}}{\hat{\sigma}}\right)\frac{1}{\phi\left(\Phi^{-1}(\theta_j^*)\right)}}^{\text{importance density}}}{\underbrace{p(y \mid \theta_j^*)}_{\text{likelihood}}\ \underbrace{p(\theta_j^*)}_{\text{prior}}}\right)^{-1}, \quad \underbrace{\theta_j^* \sim p(\theta \mid y)}_{\substack{\text{samples from the}\\ \text{posterior distribution}}}. \\ &= \left(\frac{1}{N}\sum_{j=1}^{N}\frac{\overbrace{\frac{1}{\hat{\sigma}}\phi\left(\frac{\xi_j^*-\hat{\mu}}{\hat{\sigma}}\right)}^{\text{importance density}}}{\underbrace{p\left(y \mid \Phi\left(\xi_j^*\right)\right)}_{\text{likelihood}}\ \underbrace{p\left(\Phi\left(\xi_j^*\right)\right)\phi\left(\xi_j^*\right)}_{\text{prior}}}\right)^{-1}, \quad \underbrace{\xi_j^* = \Phi^{-1}(\theta_j^*) \text{ and } \theta_j^* \sim p(\theta \mid y)}_{\substack{\text{probit-transformed samples}\\ \text{from the posterior distribution}}},\end{aligned} \tag{F.3}$$

which shows that the generalized harmonic estimate can be obtained using the samples from the posterior distribution, or the probit-transformed ones. In the online-provided code, we use the latter approach (see also Overstall & Forster, 2010).

F.4 Details on the Application of Bridge Sampling to the Individual-Level EV Model

In this section, we provide more details on how we obtained the unnormalized marginal likelihood for a specific participant s, $s \in \{1, 2, \ldots, 30\}$, with choices $Ch_s(T)$ and corresponding payoffs

$X_s(T)$. Since we focus on one specific participant, we drop the subscript s in the remainder of this section. As explained in Appendix F.2, we run the iterative scheme with respect to $\hat{r}$ to avoid numerical issues. Consequently, we have to compute $\log(l_{1,j})$ and $\log(l_{2,i})$. Using $\tilde{\boldsymbol{\kappa}}_{\boldsymbol{i}} = (\tilde{\omega}_i, \tilde{\alpha}_i, \tilde{\gamma}_i)$ for the i^{th} sample from the proposal distribution, we get for $\log(l_{2,i})$ ($\log(l_{1,j})$ works analogously):

$$\log(l_{2,i}) = \log\left(\frac{p(Ch(T) \mid \Phi(\tilde{\boldsymbol{\kappa}}_i), X(T))\; p(\Phi(\tilde{\boldsymbol{\kappa}}_i))\; \phi(\tilde{\boldsymbol{\kappa}}_{\boldsymbol{i}})}{g(\tilde{\boldsymbol{\kappa}}_i)}\right).$$

Therefore, instead of computing the unnormalized posterior distribution directly, we compute the logarithm of the unnormalized posterior distribution:

$$\begin{aligned}\log(p(Ch(T) \mid \Phi(\tilde{\boldsymbol{\kappa}}_i), X(T))\; p(\Phi(\tilde{\boldsymbol{\kappa}}_i))\; \phi(\tilde{\boldsymbol{\kappa}}_i)) \; &= \log(p(Ch(T) \mid \Phi(\tilde{\boldsymbol{\kappa}}_i), X(T)))+ \\ &\log(\phi(\tilde{\omega}_i)) + \log(\phi(\tilde{\alpha}_i)) + \log(\phi(\tilde{\gamma}_i)),\end{aligned}$$

because we assumed independent priors on each model parameter w, a, c. $\log(p(\Phi(\tilde{\boldsymbol{\kappa}}_i))) = 0$ because p refers to the uniform prior on $[0, 1]$.

F.5 Details on the Application of Bridge Sampling to the Hierarchical EV Model

Analogous to the last section, we explain here how we obtained the logarithm of the unnormalized posterior for the hierarchical implementation of the EV model. Using $Ch_s(T)$ to refer to all choices of subject s, $X_s(T)$ for the corresponding net outcomes, $\tilde{\boldsymbol{\kappa}}_{\boldsymbol{s,i}} = (\tilde{\omega}_{s,i}, \tilde{\alpha}_{s,i}, \tilde{\gamma}_{s,i})$ for the i^{th} sample from the proposal distribution for the individual-level parameters of subject s, and $\tilde{\boldsymbol{\zeta}}_i$ for the i^{th} sample from the proposal distribution for all group-level parameters (e.g., $\tilde{\boldsymbol{\zeta}}_i = (\tilde{\mu}_{\omega,i}, \tilde{\tau}_{\omega,i}, \tilde{\mu}_{\alpha,i}, \tilde{\tau}_{\alpha,i}, \tilde{\mu}_{\gamma,i}, \tilde{\tau}_{\gamma,i})$), we get:

$$\log\left(\left(\prod_{s=1}^{30} p(Ch_s(T) \mid \Phi(\tilde{\boldsymbol{\kappa}}_{\boldsymbol{s,i}}), X_s(T))\; p(\tilde{\boldsymbol{\kappa}}_{\boldsymbol{s,i}} \mid \tilde{\boldsymbol{\zeta}}_i)\right)\; p(\tilde{\boldsymbol{\zeta}}_i)\right)$$

$$= \sum_{s=1}^{N} [\log(p(Ch_s(T) \mid \Phi(\tilde{\boldsymbol{\kappa}}_{s,i}), X_s(T)))+$$

$$\log\left(\frac{1}{1.5\Phi(\tilde{\tau}_{\omega,i})}\,\phi\left(\frac{\tilde{\omega}_{s,i} - \tilde{\mu}_{\omega,i}}{1.5\Phi(\tilde{\tau}_{\omega,i})}\right)\right) + \log\left(\frac{1}{1.5\Phi(\tilde{\tau}_{\alpha,i})}\,\phi\left(\frac{\tilde{\alpha}_{s,i} - \tilde{\mu}_{\alpha,i}}{1.5\Phi(\tilde{\tau}_{\alpha,i})}\right)\right) +$$

$$\left.\log\left(\frac{1}{1.5\Phi(\tilde{\tau}_{\gamma,i})}\,\phi\left(\frac{\tilde{\gamma}_{s,i} - \tilde{\mu}_{\gamma,i}}{1.5\Phi(\tilde{\tau}_{\gamma,i})}\right)\right)\right] +$$

$$\log\left(\phi(\tilde{\mu}_{\omega,i})\right) + \log\left(\phi(\tilde{\mu}_{\alpha,i})\right) + \log\left(\phi(\tilde{\mu}_{\gamma,i})\right) +$$

$$\log\left(\phi(\tilde{\tau}_{\omega,i})\right) + \log\left(\phi(\tilde{\tau}_{\alpha,i})\right) + \log\left(\phi(\tilde{\tau}_{\gamma,i})\right).$$