Supplementary Material: A Robust Approach to Sequential Information Theoretic Planning

Sue Zheng [1] Jason Pacheco [1] John W. Fisher, III [1]

1. Proofs of Estimator Properties

1.1. Proof of Prop. 1

Here we show how the bias of the empirical estimator depends on N, M. Since we use a plug in estimate of the predictive posterior $\hat{p}^i \approx p(y^i|\mathcal{Y})$ to estimate MI, we first determine the mean and variance of the samples θ^i for MI estimation which incorporate the plug in estimate $\theta^i = \log \frac{p(y^i|x^i)}{\hat{p}^i}$.

Let $\{x^i, y^i\}_{i=1}^N \sim p(x, y \mid \mathcal{Y})$ and for each y^i let $\{x^{ij}\}_{j=1}^M \sim p(x \mid \mathcal{Y})$ be independent samples. Going forward, we will drop the explicit dependence on the observed data, $\mathcal{Y}$, for clarity but all expressions are conditioned on them. The empirical estimate of the posterior predictive takes form $\hat{p}^i = \frac{1}{M}\sum_j p(y^i|x^{ij})$. From Thm. 3.9 of (DasGupta, 2008), if we have iid observations $X^1, X^2, \dots, X^M$ with finite fourth moment, mean μ, and variance σ^2, and a scalar transformation g with four uniformly bounded derivatives, then

$$E[g(\overline{X})] = g(\mu) + \frac{g^{(2)}(\mu)\sigma^2}{2M} + O(n^{-2})$$
$$\text{var}\big(g(\overline{X})\big) = \frac{(g'(\mu))^2\sigma^2}{M} + O(n^{-2})$$

where $\overline{X}$ is the empirical mean $\overline{X} = \frac{1}{M}\sum_i X^i$. In our case, we have M samples $p(y^i|x^{ij})$ which are iid when conditioned on y^i with conditional mean $p(y^i)$ and conditional variance $\sigma^2_{p(y^i|X)} \triangleq \int_x p(y^i|x)^2 p(x)dx - p(y^i)^2$. We are interested in the transformation $g = -\log$. Consequently, we have

$$\mathbb{E}[-\log(\hat{p}^i)|y^i] = -\log(p(y^i)) + \frac{\sigma^2_{p(y^i|X)}}{2Mp(y^i)^2} + O(M^{-2})$$
$$\text{var}\big(-\log(\hat{p}^i)|y^i\big) = \frac{\sigma^2_{p(y^i|X)}}{Mp(y^i)^2} + O(M^{-2}).$$

Using the law of total expectation to remove the conditioning on y^i in the mean and dropping the superscript i since measurements are sampled iid, we have:

$$\mathbb{E}[-\log(\hat{p})] = H(Y) + \mathbb{E}\left[\frac{\sigma^2_{p(y|X)}}{2Mp(y)^2}\right] + O(M^{-2}). \quad (1)$$

Expanding the second term on the RHS, we obtain:

$$\begin{aligned}\mathbb{E}\left[\frac{\sigma^2_{p(y|X)}}{2Mp(y)^2}\right] &= \int_y p(y)\frac{\int_x p(y|x)^2 p(x)dx - p(y)^2}{2Mp(y)^2}dy \\ &= \frac{1}{2M}\left[\int_y \frac{\int_x p(y|x)^2 p(x)dx}{p(y)}dy - 1\right] \\ &= \frac{1}{2M}\left[\int_y \int_x \frac{p(y|x)^2 p(x)^2 dx}{p(y)p(x)}dy - 1\right] \\ &= \frac{1}{2M}\chi^2(p(y,x)||p(y)p(x)) \quad (2)\end{aligned}$$

where $\chi^2(p(y,x)||p(y)p(x))$ denotes the chi-square divergence of $p(y,x)$ from $p(y)p(x)$. Plugging this back into 1, we obtain

$$\mathbb{E}[-\log(\hat{p})] = H(Y) + \frac{\chi^2(p(y,x)||p(y)p(x))}{2M} + O(M^{-2}). \quad (3)$$

From Eqn. 3 we can now obtain the mean of each sample $\theta_i = \log \frac{p(y^i|x^i)}{\hat{p}^i}$ used in the empirical MI estimate $\hat{I}_{NM} = \frac{1}{N}\sum_i^N \theta_i$ and therefore, the mean of the MI estimate:

$$\begin{aligned}\mathbb{E}[\hat{I}_{NM}] = \mathbb{E}[\theta] &= \mathbb{E}[\log p(y|x)] - \mathbb{E}[\log \hat{p}] \\ &= -H(Y|X) + H(Y) + \frac{\chi^2(p(y,x)||p(y)p(x))}{2M} \\ &\quad + O(M^{-2}) \\ &= I(X;Y) + \frac{\chi^2(p(y,x)||p(y)p(x))}{2M} + O(M^{-2}). \quad (4)\end{aligned}$$

We have thus shown how the bias of the empirical estimator depends on N, M.

*Equal contribution [1]Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Boston, MA, USA. Correspondence to: Sue Zheng <szheng@csail.mit.edu>.

Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018. Copyright 2018 by the author(s).

1.2. Proof of Prop. 2

We first show that the empirical plug-in estimator is consistent and asymptotically normal. We will next show that the influence function of the robust estimator converges to the influence function of the empirical estimator, such that the asymptotic analysis of the empirical estimator holds for the robust estimate as well. We begin by finding expressions for the mean and variance of the iid θ^i samples used in MI estimation. Then, using these expression show that the empirical mean converges to a Normal distribution.

In Sec. 1.1, we have an expression for the variance of the log predictive posterior conditioned on y^i. We can remove the conditioning on y^i though the law of total variance. Again removing the superscript i for clarity, we have the following expression for variance:

$$\begin{aligned}
&\text{var}\big(-\log(\hat{p})\big) \qquad (5)\\
&\quad= \mathbb{E}[\text{var}\big(\log(\hat{p})|y\big)] + \text{var}\big(\mathbb{E}[\log(\hat{p})|y]\big)\\
&\quad= \mathbb{E}\left[\frac{\sigma^2_{p(y|X)}}{Mp(y)^2} + O(M^{-2})\right] + \text{var}\big(\mathbb{E}[\log(\hat{p})|y]\big).
\end{aligned}$$

We can simplify the first term on the RHS through Eqn. 2. Keeping track of only the order M^{-1} terms and higher, we expand the last term on the RHS as

$$\begin{aligned}
&\text{var}\big(-\mathbb{E}[\log(\hat{p})|y]\big)\\
&\quad= \mathbb{E}\left[\left(-\log(p(y)) + \frac{\sigma^2_{p(y|X)}}{2Mp(y)^2} + O(M^{-2})\right)^2\right]\\
&\qquad - \mathbb{E}[\log(\hat{p})]^2\\
&\quad= \int_y p(y)\left[(\log p(y))^2 - \frac{\log p(y)\sigma^2_{p(y|X)}}{Mp(y)^2} + O(M^{-2})\right]dy\\
&\qquad - (H(Y) + \frac{1}{2M}\chi^2(p(y,x)||p(y)p(x)) + O(M^{-2}))^2\\
&\quad= \int_y p(y)(\log p(y))^2 - \frac{\log p(y)\sigma^2_{p(y|X)}}{Mp(y)}dy + O(M^{-2})\\
&\qquad - \left(H(Y)^2 + \frac{1}{M}H(Y)\chi^2(p(y,x)||p(y)p(x))\right)\\
&\quad= \sigma^2(\log p(y)) - \int_y \frac{\log p(y)\sigma^2_{p(y|X)}}{Mp(y)}dy\\
&\qquad - \left(\frac{1}{M}H(Y)\chi^2(p(y,x)||p(y)p(x))\right) + O(M^{-2}).
\end{aligned}$$

We plug the above back into the variance expression (Eqn. 5) to get the following dependence on M:

$$\begin{aligned}
\text{var}\big(-\log\hat{p}\big) &= \sigma^2(\log p(y)) \qquad (6)\\
&\quad+ \frac{1}{M}\bigg[(1-H(Y))\chi^2(p(y,x)||p(y)p(x))\\
&\qquad \int_y \log p(y)\sigma^2_{p(y|X)}/p(y)dy\bigg] + O(M^{-2}).
\end{aligned}$$

We can now find the variance of the samples used in MI estimation:

$$\begin{aligned}
\text{var}\big(\theta\big) &= \sigma^2(\log p(y|x)) - \text{cov}\big(\log p(y|x), \log\hat{p}\big)\\
&\quad+ \text{var}\big(-\log\hat{p}\big). \qquad (7)
\end{aligned}$$

To simplify our analysis, let us assume that we use independent sets of samples $\{x^i, y^i\}_{i=1}^N$ and $\{x^i, y^i\}_{i=N+1}^{2N}$ for estimating $\mathbb{E}[\log p(y|x)]$ and $\mathbb{E}[\log\hat{p}]$ respectively. Thus, the covariance term in Eqn. 7 is zero. Substituting the log posterior predictive variance Eqn. 6 into the MI sample variance Eqn. 7 yields the full expression for the MI sample variance

$$\begin{aligned}
\text{var}\big(\theta\big) &= \sigma^2(\log p(y|x)) + \sigma^2(\log p(y))\\
&\quad+ \frac{1}{M}\bigg[(1-H(Y))\chi^2(p(y,x)||p(y)p(x))\\
&\qquad \int_y \log p(y)\sigma^2_{p(y|X)}/p(y)dy\bigg] + O(M^{-2})\\
&= \sigma^2\left(\log\frac{p(y|x)}{p(y)}\right) \qquad (8)\\
&\quad+ \frac{1}{M}\bigg[(1-H(Y))\chi^2(p(y,x)||p(y)p(x))\\
&\qquad \int_y \log p(y)\sigma^2_{p(y|X)}/p(y)dy\bigg] + O(M^{-2}).
\end{aligned}$$

We now have the variance and mean of θ_i, which compose the samples for the plug-in empirical estimate of MI. We will now show that this MI estimate admits a CLT. Let us define $\widetilde{\theta}_i = \theta_i - I(X;Y)$ and let each $\widetilde{\theta}_i$ have cdf $F(\widetilde{\theta}^i) = P(\widetilde{\Theta}^i \leq \widetilde{\theta}^i)$. Let us assume that it has finite moment and its moment generating function (mgf) $M_{\widetilde{\theta}}(t) = \mathbb{E}[\exp(t\widetilde{\theta}^i)]$ exists. We will show that moment generating function for $Z_N \triangleq \sqrt{N}(\hat{I}_N - I(X;Y)) = \sqrt{N}\left(\frac{\sum_i \widetilde{\theta}^i}{N}\right)$ approaches that of a zero-mean normal with variance $\sigma^2_{\hat{I}} \triangleq \sigma^2\left(\log\frac{p(y|x)}{p(y)}\right)$ so that the estimator is consistent and a CLT holds. Because $\widetilde{\theta}^i$ are iid, the mgf for Z_N

can be written as

$$\begin{aligned} M_{Z_N}(t) &= \left(M_{\sum_{i=1}^{N} \widetilde{\theta}^i/\sqrt{N}}(t)\right) \\ &= \left(M_{\widetilde{\theta}/\sqrt{N}}(t)\right)^N \\ &= \left(M_{\widetilde{\theta}}(t/\sqrt{N})\right)^N \\ &= \left(\mathbb{E}[\exp(t\widetilde{\theta}/\sqrt{N})]\right)^N . \end{aligned} \tag{9}$$

Using Taylor's theorem on the exponential, we obtain

$$\begin{aligned} &M_{Z_N}(t) \\ &= \left(\mathbb{E}[1 + \frac{t\widetilde{\theta}}{\sqrt{N}} + \frac{(t\widetilde{\theta})^2}{2N} + \frac{(t\widetilde{\theta})^3}{3!N^{3/2}} + \ldots]\right)^N \\ &= \left(1 + \frac{t}{\sqrt{N}}\left(\frac{\chi^2(p(y,x)||p(y)p(x))}{2M} + O(M^{-2})\right)\right. \\ &\qquad \left. + \frac{t^2}{2N}\mathbb{E}[\widetilde{\theta}^2] + \frac{t^3}{6N^{3/2}}\mathbb{E}[\widetilde{\theta}^3] + \ldots\right)^N \end{aligned} \tag{10}$$

Note that, since $\widetilde{\theta} = \theta - I(X;Y)$, we have that $\mathrm{var}\big(\widetilde{\theta}\big) = \mathrm{var}\big(\theta\big)$ and $\mathbb{E}[\widetilde{\theta}] = O(M^{-1})$ such that $\mathbb{E}[\widetilde{\theta}^2] = O(M^{-2})$. We now expand the third term in Eqn. 10 using our expression for variance (Eqn. 8) of θ:

$$\begin{aligned} \mathbb{E}[\widetilde{\theta}^2] &= \mathrm{var}\big(\widetilde{\theta}\big) + \mathbb{E}[\widetilde{\theta}]^2 \\ &= \mathrm{var}\big(\theta\big) + O(M^{-2}) \\ &= \sigma^2\left(\log\frac{p(y|x)}{p(y)}\right) \\ &\quad + \frac{1}{M}\Bigg[(1 - H(Y))\chi^2(p(y,x)||p(y)p(x)) \\ &\qquad \int_y \log p(y)\sigma^2_{p(y|X)}/p(y)dy\Bigg] + O(M^{-2}). \end{aligned} \tag{11}$$

Furthermore, note that since all the moments of $\widetilde{\theta}$ are finite, we can ignore the $N^{-3/2}$ or lower terms since they disappear as $N \to \infty$. For the limit to exist, the coefficient for $N^{-1/2}$ must decay at a rate at least $1/\sqrt{N}$; thus, we must have $M = \Omega(\sqrt{N})$. Finally, recall that the mgf for a Gaussian with mean μ and variance σ^2 is $\exp(t\mu + \frac{1}{2}\sigma^2 t^2)$. Therefore, to have a consistent estimator (Z_N has zero mean in the limit), we require the coefficient on the t term to disappear in the limit. This requires that M grows *strictly faster* than $\sqrt{N}$: $M = \omega(\sqrt{N})$. Letting $M = \omega(\sqrt{N})$, the limit of Eqn. 10 is:

$$\lim_{N\to\infty} M_{Z_N}(t) = \exp\left(\sigma^2\left(\log\frac{p(y|x)}{p(y)}\right)\frac{t^2}{2}\right). \tag{12}$$

Since the mgf approaches that of a zero mean Gaussian with variance $\sigma^2\left(\log\frac{p(y|x)}{p(y)}\right)$ we can conclude that the empirical plug-in MI estimator is consistent and satisfies the following CLT when $M = \omega(\sqrt{N})$:

$$\lim_{N\to\infty} \sqrt{N}(\hat{I}_N - I(X;Y)) \to \mathcal{N}(0, \sigma^2_{\hat{I}}). \tag{13}$$

Lastly, we demonstrate that as $N \to \infty$, the robust influence function approaches that of the empirical such that the consistency and CLT guarantees also hold for our robust estimator. Let $\alpha = \sqrt{\frac{2}{N\sigma^2}}$. Note that the influence function corresponding to an empirical estimate is $\psi_{empir}(x) = cx$ where c is any constant. The robust cost function, parameterized by α, is given as

$$\psi(x;\alpha) = \begin{cases} \log(1 + \alpha x + (\alpha x)^2/2), & x \geq \theta \\ -\log(1 - \alpha x + (\alpha x)^2/2), & x < \theta. \end{cases}$$

For simplicity of analysis, we will only consider $x \geq 0$ however, the proof for $x < 0$ follows similarly. Note that $\lim_{N\to\infty} \alpha x = \lim_{N\to\infty} \sqrt{\frac{2}{N}}\frac{x}{\sigma} = 0$. Therefore, we take the Taylor series of $\log(1 + y + y^2/2)$ at 0, where $y = \alpha x$, and obtain

$$\psi(x;\alpha) = \alpha x + 0 \cdot \frac{(\alpha x)^2}{2} - \frac{(\alpha x)^3}{3!} + \cdots$$

As $N \to \infty$, the first term dominates and we get that the robust influence function approaches the empirical function with $c = \alpha$.

1.3. Proof of Prop. 3

Here, we first show the finite sample deviation bound on the robust estimator, then secondly the bound on the empirical. Let $\{x^i, y^i\}_{i=1}^N \sim p(x, y \mid \mathcal{Y})$ and for each y^i let $\{x^{ij}\}_{j=1}^M \sim p(x \mid \mathcal{Y})$ be independent samples. We define $\boldsymbol{x}^i \triangleq \{x^{ij}\}_{j=1}^M$. Recall that $\hat{I}_{NM} = \textsc{Root}(\sum_i \psi(\alpha(\theta^i - \hat{I}_{NM}))$ where $\theta^i = \log\frac{p(y^i|x^i)}{\hat{p}(y^i;\boldsymbol{x}^i)}$ and $\hat{p}(y^i;\boldsymbol{x}^i)$ is the posterior predictive estimator. Assuming $\alpha = \sqrt{2/N\sigma^2_{\hat{I}_{NM}}}$ and $N > 2 + 2\log\epsilon^{-1}$, we have from Prop. 2.4 of (Catoni, 2012), the robust estimator satisfies with probability at least $1 - 2\epsilon$

$$c \leq \hat{I}_{NM} - m \leq c \tag{14}$$

where

$$m = \mathbb{E}[\theta^i] = \mathbb{E}\left[\log\frac{p(y \mid x)}{\hat{p}(y;\boldsymbol{x})}\right]$$

and $c = \frac{2(1+\log\epsilon^{-1})\sqrt{\sigma^2_{\hat{I}_{NM}}/2N}}{1+\sqrt{1-2(1+\log\epsilon^{-1})/N}}$. Since the samples are identically distributed, m does not depend on sample index i. Note that, because the denominator contains the *estimated* posterior predictive, this currently does not bound deviation from the true MI value, I. We can rewrite m as follows:

$$\begin{aligned} m &= \mathbb{E}\left[\log \frac{p(y \mid x)}{\hat{p}(y; \boldsymbol{x})} \frac{p(y \mid \mathcal{Y})}{p(y \mid \mathcal{Y})}\right] \\ &= \mathbb{E}\left[\log \frac{p(y \mid x)}{p(y \mid \mathcal{Y})}\right] + \mathbb{E}\left[\log \frac{p(y \mid \mathcal{Y})}{\hat{p}(y; \boldsymbol{x})}\right] \\ &= I + \mathbb{E}\left[\log \frac{p(y \mid \mathcal{Y})}{\hat{p}(y; \boldsymbol{x})}\right] \\ &= I + \mathbb{E}_{\boldsymbol{x}}\left[KL(p(y \mid \mathcal{Y}) || \hat{p}(y; \boldsymbol{x}))\right]. \end{aligned} \tag{15}$$

Note that the above derivation holds for both the robust and empirical estimators. Letting $b = \mathbb{E}_{\boldsymbol{x}}\left[KL(p(y \mid \mathcal{Y}) || \hat{p}(y; \boldsymbol{x}))\right]$, we obtain the presented deviation bounds on the robust estimator:

$$b - c \leq \hat{I}_{NM} - m \leq b + c \tag{16}$$

Similarly, the empirical estimate can be bounded as in Eqn. 14 but with $c = \sqrt{\frac{\sigma^2_{\hat{i}_{NM}}}{2N\epsilon}}$ by Chebyshev's inequality. Since the same derivation relating the true mean to the plug-in mean in Eqn. 15 holds for the empirical, we have that Eqn. 16 also carries through. Although the high-level expression stays the same, we must be careful to note that both b and c differ for the two estimators. The difference in b is a bit more subtle. b is the expected KL divergence between the posterior predictive and the estimated posterior predictive; the estimated posterior predictive differs between robust and empirical.

1.4. Proof of Prop. 4

We now show how we approximate the probability that the correct action is selected. Without loss of generality, let $I_1 \geq I_2 \geq \ldots \geq I_A$ and let $f_a(\hat{I}_a)$ and $F_a(\hat{I}_a)$ denote the pdf and cdf of the estimate. Since independent samples are drawn to estimate MI under each candidate action, their estimates are independent. This allows us to decompose the probability of selecting the correct action as:

$$\begin{aligned} \mathbb{P}(a^* = 1) &= \int_{-\infty}^{+\infty} f(\hat{I}_1) \left(\prod_{a=2}^{A} \int_{-\infty}^{\hat{I}_1} f(\hat{I}_a) d\hat{I}_a \right) d\hat{I}_1 \\ &= \int_{-\infty}^{+\infty} f(\hat{I}_1) \left(\prod_{a=2}^{A} F_a(\hat{I}_1) \right) d\hat{I}_1 \end{aligned} \tag{17}$$

For $N \gg 1$, we have that

$$\mathbb{P}(a^* = 1) \approx \int_{-\infty}^{+\infty} \mathcal{N}(\hat{I}_1; I_1, \sigma_1^2) \prod_{a=2}^{A} \Phi\left(\frac{\hat{I}_1 - I_a}{\sigma_a}\right) d\hat{I}_1$$

where $\sigma_a^2 = \frac{1}{N}\sigma^2(\log \frac{p_a(y|x)}{p_a(y|\mathcal{Y})})$ and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

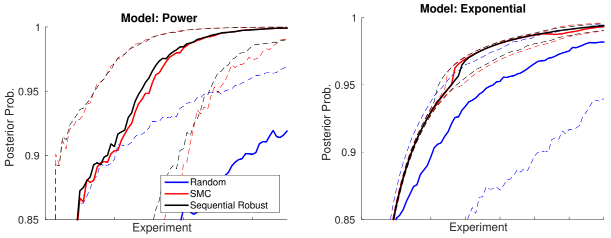

Figure 1. **Memory retention model.** Plots of posterior model probability of the *correct* retention model for 500 random trials of 100 experiments with 500 particles. Data are sampled from POW (left) and EXP (right). Plots show median (solid) and quartiles (dashed) over the same set of runs.

2. Experiment: Memory Retention Model Selection

We apply our method to sequential experiment design for Bayesian model discrimination. We adopt the analysis posed in (Cavagnaro et al., 2010) where the aim is to discriminate among candidate models of memory retention. At each stage an experiment is conducted where a simulated participant is given a list of words to memorize. After a chosen lag time d_t participants are asked to recall the words. At stage t the aim is to select lag time d_t which maximally discriminates between two retention models, *power* (POW) and *exponential* (EXP):

$$p_0(d_t \mid \theta) = \theta_0(d_t + 1)^{-\theta_1}, \;\; p_1(d_t \mid \theta) = \theta_0 e^{-\theta_1 d_t}.$$

The joint distribution combines the retention model with a response variable y_t where $y_t = 1$ indicates the participant remembers the words and $y_t = 0$ otherwise,

$$\begin{gathered} m \sim \text{Unif}(\cdot), \;\; \theta_0 \sim \text{Beta}(\alpha_0, \beta_0), \;\; \theta_1 \sim \text{Beta}(\alpha_1, \beta_1), \\ y_t \mid m, \theta; d_t \sim \text{Bernoulli}\big(p_m(d_t \mid \theta)\big). \end{gathered}$$

At each stage we select the lag time which maximizes mutual information between the model and observation, $\arg\max_{d_t} I_{d_t}(M; Y_t \mid \mathcal{Y}_{t-1})$. Expectations over model M and response Y_t are discrete, and can be computed efficiently. Furthermore, under a uniform prior we have that the model posterior is proportional to the evidences,

$$Z_{mt}(y_t, d_t) \propto \int p(\theta \mid m) \prod_{i=1}^{t} p(y_i \mid \theta, m; d_1^t) \; \mathrm{d}\theta \tag{18}$$

We compare our integrated inference and planning approach to the sequential Monte Carlo (SMC) approach of (Drovandi et al., 2014; Chopin, 2002), which utilizes the SMC estimate of the evidence for planning: $\log \hat{Z}_{mt} = \log \hat{Z}_{m,t-1} + \log \sum_{i=1}^{N} w_{mt}^i$.

We observe moderate slightly improved median performance and tighter quartiles as shown in Fig. 1. Improvements over SMC are modest, likely due to the low-dimensional 2D integration over θ. We also find that when data are sampled from the exponential model posterior concentrates quickly under both methods, which confirms findings of the original authors. Both our method and SMC significantly outperform random design.

References

Catoni, O. Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. H. Poincar Probab. Statist.*, 48(4):1148–1185, 11 2012. doi: 10.1214/11-AIHP454.

Cavagnaro, D. R., Myung, J. I., Pitt, M. A., and Kujala, J. V. Adaptive design optimization: A MI-based approach to model discrimination in cognitive science. *Neural-Comp.*, 22(4):887–905, 2010.

Chopin, N. A sequential particle filter method for static models. *Biometrika*, 89(3):539–552, 2002.

DasGupta, A. *Asymptotic Theory of Statistics and Probability*. Springer Texts in Statistics. Springer New York, 2008. ISBN 9780387759715.

Drovandi, C. C., McGree, J. M., and Pettitt, A. N. A sequential monte carlo algorithm to incorporate model uncertainty in bayesian sequential design. *JCGS*, 23(1): 3–24, 2014.