Derivation of $\sigma^2_{\mathrm{loo,diff}}$

Proofs

The main quantity of interest is the mean expected log pointwise predictive density, which we want to use for model evaluation and comparison.

Definition 1 ($\overline{elpd}$)**.** *The* mean expected log pointwise predictive density *for a model p is defined as*

$$\overline{\mathrm{elpd}} = \int p_t(x) \log p(x)\, dx$$

where $p_t(x) = p(x|\theta_0)$ is the true *density at a new unseen observation x and $\log p(x)$ is the log predictive density for observation x.*

We estimate $\overline{\mathrm{elpd}}$ using *leave-one-out cross-validation (loo).*

Definition 2 (Leave-one-out cross-validation)**.** *The loo estimator $\overline{\mathrm{elpd}}_{\mathrm{loo}}$ is given by*

$$\overline{\mathrm{elpd}}_{\mathrm{loo}} = \frac{1}{n} \sum_{i=1}^{n} \pi_i, \tag{1}$$

where $\pi_i = \log p(y_i|y_{-i}) = \int \log p(y_i|\theta) p(\theta|y_{-i}) d\theta$.

To estimate $\overline{elpd}_{\mathrm{loo}}$ in turn, we use difference estimator. Definitions follow.

Definition 3. *Let $\tilde{\pi}_i$ be any approximation of π_i. The difference estimator of $\overline{elpd}_{\mathrm{loo}}$ based on $\tilde{\pi}_i$ is given by*

$$\widehat{\overline{\mathrm{elpd}}}_{\mathrm{loo,diff}} = \frac{1}{n} \left(\sum_{i=1}^{n} \tilde{\pi}_i + \frac{n}{m} \sum_{j \in \mathcal{S}} (\pi_j - \tilde{\pi}_j) \right) = \frac{1}{n} \left(t_{\tilde{\pi}} + \hat{t}_e \right) ,$$

where $\mathcal{S}$ is the subsample set, m is the subsampling size, and the probability of subsampling observation i is $1/n$, i.e. the subsample is uniform with replacement.

One important estimator of π_i among others is the importance sampling estimator

$$\log \hat{p}(y_i|y_{-i}) = \log \left(\frac{\frac{1}{S} \sum_{s=1}^{S} p(y_i|\theta_s) r(\theta_s)}{\frac{1}{S} \sum_{s=1}^{S} r(\theta_s)} \right) , \tag{2}$$

where $r(\theta)$ is any suitable weight function such that $0 < r(\theta) < \infty$ for all $\theta \in \Theta$ and $(\theta_1, \dots, \theta_S)$ is a sample from a suitable approximation of the posterior $p(\theta|y)$. We are in particular interested in the weight function

$$
\begin{aligned}
r(\theta_s) =& \frac{p(\theta_s|y_{-i})}{p(\theta_s|y)} \frac{p(\theta_s|y)}{q(\theta_s|y)} \\
\propto& \frac{1}{p(y_i|\theta_s)} \frac{p(\theta_s|y)}{q(\theta_s|y)}
\end{aligned} \tag{3}
$$

and where $q(\cdot|y)$ is an approximation of the posterior distribution that satisfies for each y that $q(\theta|y)$ iff $\theta \in \Theta$, θ_s is a sample point from q and S is the total posterior sample size. (The condition on q makes sure that $0 < r(\theta) < \infty$ for all θ.)

In the case of truncated importance sampling, we instead truncate these weights and replace r with r_τ given by

$$
r_\tau(\theta_s) = \min(r(\theta_s), \tau)\,, \tag{4}
$$

where $\tau > 0$ is the weight truncation [see Ionides, 2008, for a more elaborate discussion on the choice of τ].

Proof of Proposition 1

Proposition 1. *The estimators* $\widehat{\mathrm{elpd}}_{\mathrm{diff}}$ *and* $\hat{\sigma}^2_{\mathrm{loo}}$ *are unbiased with regard to* $\mathrm{elpd}_{\mathrm{diff}}$ *and* σ^2_{loo}.

Proof. We start out by proving unbiasedness for the general estimator. Write the difference estimator as

$$
\widehat{\mathrm{elpd}}_{\mathrm{loo,diff}} = \hat{t}_\pi = \sum_{i=1}^{n} \tilde{\pi}_i + \frac{n}{m} \sum_{i=1}^{n} \sum_{j \in \mathcal{S}} I_{ij}(\pi_j - \tilde{\pi}_j),
$$

where I_{ij} is the indicator that data point i is chosen as the j'th point of the subsample. Since $\mathbb{E}[I_{ij}] = 1/n$, the expectation of the double sum is $\sum_i (\pi_i - \tilde{\pi}_i)$ and $\mathbb{E}[\widehat{\mathrm{elpd}}_{\mathrm{loo,diff}}] = \sum_i \pi_i$ as desired.

Next we prove unbiasedess of $\hat{\sigma}^2_{\mathrm{loo,diff}}$. We are interested in estimating the finite sampling variance using the difference estimator. This can be done as

$$
\begin{aligned}
\sigma^2_{\mathrm{loo}} =& \frac{1}{n} \sum_{i=1}^{n} (\pi_i - \bar{\pi})^2 & (5) \\
=& \underbrace{\frac{1}{n} \sum_{i=1}^{n} \pi_i^2}_{a} - \underbrace{\left(\frac{1}{n} \sum_{i=1}^{n} \pi_i \right)^2}_{b} & (6)
\end{aligned}
$$

We can estimate a and b separately as follows. The first part can be estimated using the difference estimator with $\tilde{\pi}_i^2$ as auxiliary variable. Let $t_\epsilon = \sum_i^n \epsilon_i = \sum_i^n \pi_i^2 - \tilde{\pi}_i^2 = t_{\pi^2} - t_{\tilde{\pi}^2}$, the we can estimate a as

$$\hat{a} = \frac{1}{n}(t_{\tilde{\pi}^2} + \hat{t}_\epsilon)\,, \tag{7}$$

where

$$\hat{t}_\epsilon = \frac{n}{m}\sum_{j\in\mathcal{S}}\left(\pi_j^2 - \tilde{\pi}_j^2\right)\,.$$

From the previous section, it follows directly that

$$E(\hat{a}) = \frac{1}{n}t_{\pi^2} = \frac{1}{n}\sum_{i=1}^{n}\pi_i^2,$$

The second part, b, can then be estimated as

$$\hat{b} =\frac{1}{n^2}\left[\hat{t}_e^2 - v(\hat{t}_e) + 2t_{\tilde{\pi}}\hat{t}_\pi - t_{\tilde{\pi}}^2\right]\,, \tag{8}$$

with the expectation

$$\begin{aligned}
E(\hat{b}) =& \frac{1}{n^2}\left[E(\hat{t}_e^2) - E(v(\hat{t}_e)) + 2t_{\tilde{\pi}}E(\hat{t}_\pi) - t_{\tilde{\pi}}^2\right] && (9)\\
=& \frac{1}{n^2}\left[V(\hat{t}_e) + E(\hat{t}_e)^2 - V(\hat{t}_e) + 2t_{\tilde{\pi}}t_\pi - t_{\tilde{\pi}}^2\right] && (10)\\
=& \frac{1}{n^2}\left[t_e^2 + 2t_{\tilde{\pi}}t_\pi - t_{\tilde{\pi}}^2\right] && (11)\\
=& \frac{1}{n^2}\left[(t_\pi - t_{\tilde{\pi}})^2 + 2t_{\tilde{\pi}}t_\pi - t_{\tilde{\pi}}^2\right] && (12)\\
=& \frac{1}{n^2}t_\pi^2 = \left(\frac{1}{n}\sum_i^n \pi_i\right)^2 && (13)
\end{aligned}$$

Using that

$$E(v(\hat{t}_e)) = n^2\left(1 - \frac{m}{n}\right)\frac{E(s_e^2)}{m} = n^2\left(1 - \frac{m}{n}\right)\frac{S_e^2}{m} = V(\hat{t}_e)\,. \tag{14}$$

Combining the results we have that

$$E(\hat{a} - \hat{b}) = \frac{1}{n}\sum_{i=1}^{n}\pi_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\pi_i\right)^2 = \sigma_{\text{loo}}^2\,. \tag{15}$$

□

Remark. We believe this has probably been proven before, and hence this is probably not a new theoretical result.

Derivation of $\sigma^2_{\mathbf{loo,diff}}$

Based on Eq. 6, 7 and 8, we can construct the estimator for σ^2_{loo} as

$$\sigma^2_{\text{loo}} = \hat{a} - \hat{b} \tag{16}$$

$$= \underbrace{\frac{1}{n}(t_{\tilde{\pi}^2} + \hat{t}_\epsilon)}_{\hat{a}} - \tag{17}$$

$$\underbrace{\frac{1}{n^2}\left[\hat{t}_e^2 - v(\hat{t}_e) + 2t_{\tilde{\pi}}\hat{t}_\pi - t_{\tilde{\pi}}^2\right]}_{\hat{b}} \tag{18}$$

$$= \underbrace{\frac{1}{n}\left(\sum_{i=1}^{n}\tilde{\pi}_i^2 + \frac{n}{m}\sum_{j\in\mathcal{S}}\left(\pi_j^2 - \tilde{\pi}_j^2\right)\right)}_{\hat{a}} - \tag{19}$$

$$\underbrace{\frac{1}{n^2}\left[\hat{t}_e^2 - v(\hat{t}_e)\right]}_{\hat{b}} - \tag{20}$$

$$\underbrace{\frac{1}{n^2}\left[2t_{\tilde{\pi}}\hat{t}_\pi - t_{\tilde{\pi}}^2\right]}_{\hat{b}} \tag{21}$$

$$= \underbrace{\frac{1}{n}\left(\sum_{i=1}^{n}\tilde{\pi}_i^2 + \frac{n}{m}\sum_{j\in\mathcal{S}}\left(\pi_j^2 - \tilde{\pi}_j^2\right)\right)}_{\hat{a}} - \tag{22}$$

$$\underbrace{\frac{1}{n^2}\left[\left(\frac{n}{m}\sum_{j\in\mathcal{S}}(\pi_j - \tilde{\pi}_j)\right)^2 - v(\hat{t}_\pi)\right]}_{\hat{b}} - \tag{23}$$

$$\underbrace{\frac{1}{n^2}\left[2\left(\sum_{i=1}^{n}\tilde{\pi}_i\right)\hat{t}_\pi - \left(\sum_{i=1}^{n}\tilde{\pi}_i\right)^2\right]}_{\hat{b}} \tag{24}$$

Here, we use that

$$v(\hat{t}_e) = v(\hat{t}_\pi) = V(\widehat{\text{elpd}}_{\text{diff,loo}})\,,$$

that

$$\hat{t}_\pi = \widehat{\text{elpd}}_{\text{diff,loo}}\,,$$

$$
\begin{aligned}
\hat{\sigma}^2_{\text{diff,loo}} &= n\sigma^2_{\text{loo}} && (25)\\
&= \sum_{i=1}^{n} \tilde{\pi}_i^2 + \frac{n}{m}\sum_{j\in\mathcal{S}} \left(\pi_j^2 - \tilde{\pi}_j^2\right) - && (26)\\
&\quad \frac{1}{n}\left[\left(\frac{n}{m}\sum_{j\in\mathcal{S}}(\pi_j - \tilde{\pi}_j)\right)^2 - V(\widehat{\text{elpd}}_{\text{diff,loo}})\right] - \\
&\quad \frac{1}{n}\left[2\left(\sum_{i=1}^{n}\tilde{\pi}_i\right)\widehat{\text{elpd}}_{\text{diff,loo}} - \left(\sum_{i=1}^{n}\tilde{\pi}_i\right)^2\right].
\end{aligned}
$$

References

Edward L Ionides. Truncated importance sampling. *Journal of Computational and Graphical Statistics*, 17(2):295–311, 2008.