Experiment: Cross Validation and WAIC

Let's compare ISCV, PSIS, and WAIC with the generalization error in a simple regression problem. You might better see the result by your own eyes. Sumio Watanabe.

Training sample size: $n=10$,

Test sample size: $t=100*n$,

True parameters: $a_0=0.2$, $s_0=100$.

Training inputs: $x(i) = 0.1*i$ $(i=1,2,\ldots,n)$.

Random variables : subject to $N(0,1)$: $N(i)$ $(i=1,2,\ldots,n)$.

Training outputs: $y(i) = a_0*x(i) + (1/s_0)^{1/2}*N(i)$ $(i=1,2,\ldots,n)$.

Test inputs: $x(j) = 0.1*\{1+ (j-1)\%n\}$ $(i=1,2,\ldots,t)$. % means reminder

Random variables : subject to $N(0,1)$: $N(j)$ $(j=1,2,\ldots,t)$.

Test outputs: $y(j) = a_0*x(j) + (1/s_0)^{1/2}*N(j)$ $(j=1,2,\ldots,t)$.

Parameters: $s>0$, a. Hyperparameter: $\mu=0.01>0$.

Regression Model: $p(y|x,a,s)=(s / 2\pi)^{1/2} \exp(-(s/2)(y-ax)^2)$

Prior: $\varphi(a,s|\mu) = (1/C)\, s \exp(- (\mu/2)\, s(1+a^2))$ $(C = 4\pi^{1/2}\, \Gamma(3/2) / \mu^2)$

Training entropy: $S_n = - (1/n) \Sigma_i \log p(y(i)|x(i),a_0,s_0)$

Test entropy: $S_t = - (1/t) \Sigma_j \log p(y(j)|x(j),a_0,s_0)$

$E_w[\ ]$: posterior mean. $V_w[\ ]$: posterior variance

The generalization error: $G = - (1/t) \Sigma_{j=1}^{t} \log E_w [p(y(j)|x(j),a,s)] - S_t$

WAIC: $W = (1/n) \Sigma_{i=1}^{n} \{ - \log E_w [p(y(i)|x(i),a,s)] + V_w[\log p(y(i)|x(i),a,s)] \} - S_n$

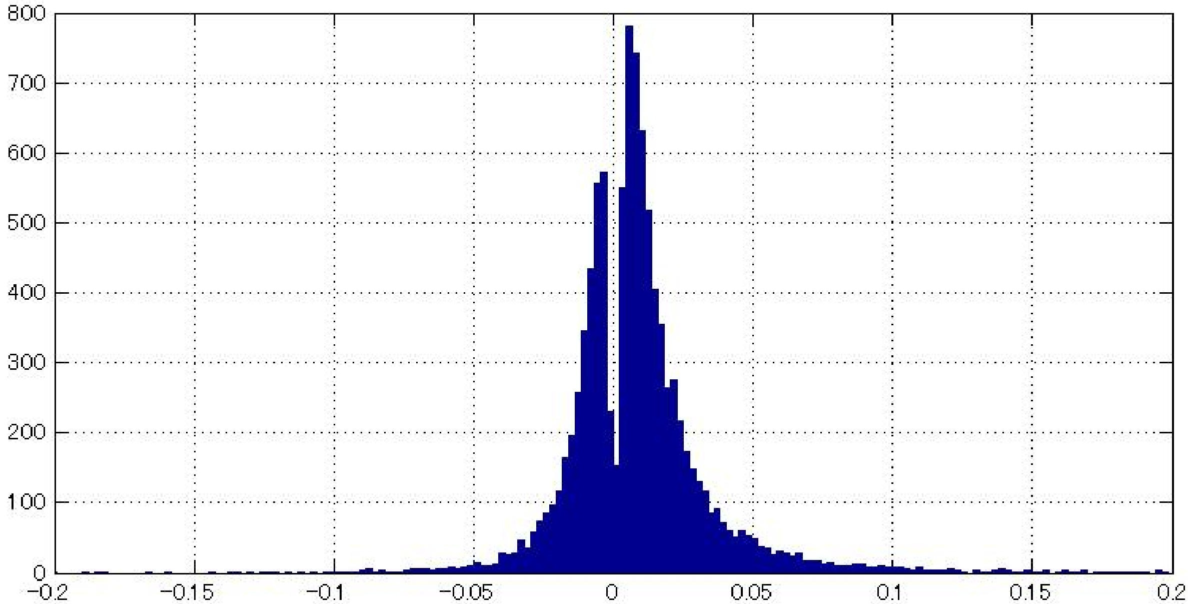
Importance sampling cross validation: $ISCV = (1/n) \Sigma_{i=1}^{n} \log E_w [1 / p(y(i)|x(i),a,s)] - S_n$

PSIS : Pareto smoothing cross validation : posterior weight is improved by Pareto dist.

Experiment : $E[\ ]$ is the expected value over all training data.

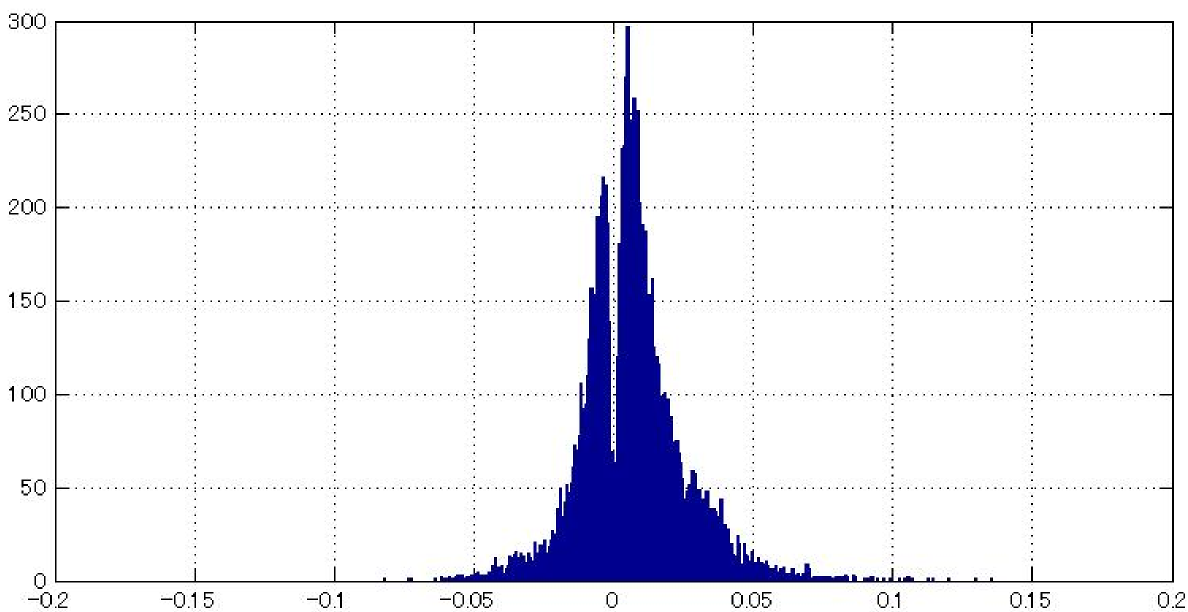
Compare $E\,|\,WAIC - GE|$, $E\,|\,ISCV - GE|$, and $E\,|\,PSIS - GE|$.

Experimental Results



This figure shows the histogram of |ISCV-GE| - |WAIC-GE| of 10000 trials.

WAIC is a better estimator of GE than ISCV.



This figure shows the histogram of |PSIS-GE| - |WAIC-GE| of 10000 trials.

WAIC is a better estimator of GE than PSIS.