Sequential Conformalized Density Regions (SCDR)

Max Sampson[1] and Kung-Sik Chan[1]

1. Department of Statistics and Actuarial Science, University of Iowa.

For the preprint, scan the QR code:

Contributions

- Extended conformal prediction for time series to include conditional density estimators, which achieve asymptotic conditional coverage.
- Showed that Normal mixture density estimators are consistent for stationary, invertible, and regular β-mixing processes.
- Achieved significantly smaller prediction sets than competing methods, while maintaining conditional coverage, not just marginal coverage.

Existing Methods

1. Methods that adaptively adjust the coverage rate for prediction intervals produced by the forecasting algorithm, including conformal PID control (PID) and Bellman Conformal Inference (BCI).
   - PID tracks the $1-\alpha$ quantile of the absolute or signed residual and updates at each step, growing or shrinking. It also can incorporate a scorecaster, a model to capture the remaining dependency in the residuals.
   - BCI optimizes a function which controls the prediction interval length and coverage, attempting to create sharp prediction intervals,

     $$\min_{\alpha_t} \mathbb{E}_{\beta_t \sim F_t}\left[L_t(\alpha_t) + \lambda_t \times \max(\mathrm{err}_t - \bar{\alpha}, 0)\right],$$

     where $\mathrm{err}_t = \mathbb{1}(\alpha_t > \beta_t)$ denotes the coverage error indicator at time t, λ_t denotes the relative weight on the miscoverage level that is used to guarantee coverage, and $L_t(\alpha_t)$ measures the length of a $1-\alpha_t$ prediction interval.
2. Methods that model the quantiles of the past T residuals, which are then combined with a point estimate to form prediction intervals, including sequential predictive conformal inference (SPCI).
   - SPCI starts with a point predictor, $\hat{f}$. It then uses a leave-one-out approach to compute the signed residuals, and models them with a quantile random forest using the past w residuals as features to capture dependence. The point predictor then adds and subtracts the estimated quantiles to form prediction intervals.

Leave One Out Algorithm for SCDR

Input: miscoverage level α, data $=(Y_i, \boldsymbol{X}_i)_{t=1}^T$, conditional density algorithm $\mathcal{B}$, quantile regression algorithm $\mathcal{Q}$, conditional density sliding window k_1, score sliding window k_2, and nonlinear autoregressive order of the scores w.

Procedure:

Initialize $\boldsymbol{V} = \{\}$
for $(j = 1, \ldots T)$ **do**
  Fit $\hat{f}_j = \mathcal{B}(\{(\boldsymbol{X}_i, Y_i) : i = j-1, , \ldots, j-k_2-1\})$
  Calculate $\hat{c}(\boldsymbol{X}_j)$, where $\hat{c}(\boldsymbol{x})$is the estimated $1-\alpha$ upper level density set cutoff for $Y \mid \boldsymbol{X} = \boldsymbol{x}$ found using $\hat{f}_j$
  $V_j = \hat{f}_j(Y_j \mid \boldsymbol{X}_j)/\hat{c}(\boldsymbol{X}_j)$
  $\boldsymbol{V} = \boldsymbol{V} \cup V_j$
end for
for $(t = T+1, \ldots T+m)$ **do**
  Fit $\hat{f}_t = \mathcal{B}(\{(\boldsymbol{X}_i, Y_i) : i = t-1, \ldots, t-k_2-1\})$
  Fit $\hat{q} = \mathcal{Q}(\{(V_i, [V_{i-1}, \ldots, V_{i-w}], \alpha) : i = t-1, \ldots, t-k_1-1\})$
  $\hat{q}_t = \hat{q}(V_{t-1}, \ldots, V_{t-w})$
  Calculate $\hat{c}(\boldsymbol{X}_t)$ using $\hat{f}_t$
  Output: $\hat{C}(\boldsymbol{X_t}) = \{y : \hat{f}_t(y|\boldsymbol{X}_t) > \hat{c}(\boldsymbol{X}_t) \times \hat{q}_t\}$
  $V_t = \hat{f}_t(Y_t \mid \boldsymbol{X}_t)/\hat{c}(\boldsymbol{X}_t)$
  $\boldsymbol{V} = \boldsymbol{V} \cup V_t$
end for

Theoretical Guarantees

Theorem 1. *Under mild conditions, when either the conditional density estimator or the QRF (or both) are correctly specified, SCDR achieves asymptotic $1-\alpha$ conditional coverage,*

$$|\mathbb{P}(Y_t \in \hat{C}(\boldsymbol{X}_t \mid \boldsymbol{X}_t)) - (1-\alpha)| \to 0 \quad \forall t.$$

- This result can be seen in the numerical experiments, where SCDR has near nominal marginal and conditional coverage.
- Our second theorem proves the density is consistent under the following conditions. When the true joint generating process, $\{\boldsymbol{W}_t = (Y_t, \boldsymbol{X}_t)\}$, is a convolution of Normals which are a stationary β-mixing sequence with the mixing rate $\beta(j) \leq \beta_0 j^{-\zeta}$ for some $\beta_0 > 0, \zeta > 2$, the eigenvalues of the covariance matrix lie between, $0 < \underline{\lambda} < \overline{\lambda} < \infty\}$, and the support of both the response and features is bounded, the MLE is consistent when the number of mixtures in the estimation grows at a rate of the order $C\log(n)^{p+1}$

Theorem 2. *$\int_B \hat{f}_n(y \mid \boldsymbol{x})dy \to \int_B f_0(y \mid \boldsymbol{x})dy$ as $n \to \infty$, in probability.*

Simulation Results

The simulation data generating process is

$$Y_t = f(X_t) + \epsilon_t,$$

where

$$f(X_t) = g(t)h(X_t),\ g(t) = \log(t')\sin(2\pi t'/12),\ t' = \mathrm{mod}(t, 12),$$
$$X_t = Y_{t-1}, h(t) = (X_t + X_t^2 + X_t^3)^{1/4},$$
$$\epsilon_t = 0.6\epsilon_{t-1} + e_t,\ \text{where } e_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1).$$

The simulation size was 2,000 with 500 observed data points and 1 unobserved data point that prediction sets were formed for. A Mixture density with 4 mixtures was used for SCDR, An AR(5) model assuming Normal innovations was used for PID and BCI. A random forest model was used for SPCI. The features used in all models were X_t and t'. The target coverage was 90%.

Method	Coverage	Size
SPCI	0.833 (0.012)	3.872 (0.024)
PID	0.873 (0.011)	6.721 (0.057)
BCI	0.951 (0.007)	3.923 (0.034)
SCDR	0.894 (0.010)	4.828 (0.042)

Old Faithful Geysey Study

We compared four methods on Old Faithful Geyser eruption durations and waiting times collected continuously from August 1st until August 15th, 1985, which contains 298 observations. The response is the eruption duration, the features used are the previous eruption duration and the previous waiting time. All models used were the same as in the simulation study. 200 points were treated as initially observed and 98 were sequentially observed.

	SPCI	PID	BCI	SCDR
Coverage	0.837 (0.038)	0.898 (0.031)	0.837 (0.038)	0.908 (0.029)
Size	2.909 (0.079)	4.069 (0.099)	2.199 (0.056)	1.837 (0.078)
$\mathrm{Cov}_{D>3.5}$	0.844 (0.066)	0.922 (0.034)	0.938 (0.030)	0.922 (0.034)
$\mathrm{Cov}_{D\leq 3.5}$	0.824 (0.046)	0.853 (0.062)	0.647 (0.083)	0.882 (0.056)