PhD Preliminary Oral Exam
Prediction Methods for Reliability Applications

Qinglong Tian

Department of Statistics, Iowa State University

July 7, 2020

Overview

1. Paper 1: Methods to Compute Prediction Intervals: A Review and New Results
2. Paper 2: Predicting the Number of Future Events
3. Paper 3: Constructing Prediction Interval Based on Likelihood Ratio (a future project)

Paper 1: Methods to Compute Prediction Intervals: A Review and New Results

Overview:

- The purposes of this work are to
  1. provide a review of important prediction methods.
  2. investigate the connections among different methods.
  3. provide some new results.
- Manuscript in preparation.

Notation:

- Independent and identically distributed observable sample X_n from the distribution $f(x; \boldsymbol{\theta})$.
- The future random variable Y is independent of X_n and has distribution $g(y; \boldsymbol{\theta})$.
- The parameter (vector) $\boldsymbol{\theta}$ is common to both distributions.

Prediction Interval

Let $y_p \equiv y_p(\boldsymbol{\theta})$ be the p quantile of $G(y;\boldsymbol{\theta})$.

- When $\boldsymbol{\theta}$ is known, the quantile can be used.
- When $\boldsymbol{\theta}$ is unknown, a prediction method will use information about $\boldsymbol{\theta}$ from the observed data $X_n = x_n$ to construct prediction intervals.
- We focus on the upper (one-sided) bound because usually the cost of being outside the prediction bound on one side is higher than the other side so the primary interest is on one side or the other, and an equal tail prediction interval can be properly interpreted as one-sided prediction bounds.

Coverage Probability

- Different prediction methods may lead to different prediction bounds. To evaluate prediction method, coverage probability is used to assess different prediction methods.
- There are two types of coverage probabilities: conditional coverage probability and unconditional (overall) coverage probability.
  1. Given the sample X_n, the conditional coverage probability is

     $$\mathrm{CP}[\mathrm{PI}(1-\alpha)|X_n;\boldsymbol{\theta}] = \Pr[Y \in \mathrm{PI}(1-\alpha)|X_n;\boldsymbol{\theta}].$$

  2. The unconditional coverage probability is the expectation of the conditional coverage,

     $$\mathrm{CP}[\mathrm{PI}(1-\alpha);\boldsymbol{\theta}] = \Pr[Y \in \mathrm{PI}(1-\alpha)] = \mathrm{E}_{X_n}\{\Pr[Y \in \mathrm{PI}(1-\alpha)|X_n;\boldsymbol{\theta}]\}.$$

Pivotal Based Method

Pivotal Quantity: The distribution of $q(X_n, Y)$ does not depend on any parameter.

Then the $1-\alpha$ prediction set of Y is given by

$$\{y : q(x_n, y) \leq q_{1-\alpha}\},$$

where $q_{1-\alpha}$ is the $1-\alpha$ quantile of $q(X_n, Y)$.

- What if a closed form distribution of $q(X_n, Y)$ is not available?
- What if $q(X_n, Y)$ is only an approximate pivotal quantity?

Construct Prediction Interval: Plug-in Method

The plug-in method, also known as the naive or estimative method, replaces the unknown parameters $\boldsymbol{\theta}$ with a consistent estimator $\widehat{\boldsymbol{\theta}}_n$ (usually the maximum likelihood (ML) estimator) based on the sample data X_n. The plug-in upper prediction bound is then defined as $(-\infty, y_{1-\alpha}(\widehat{\boldsymbol{\theta}}_n)]$, where $y_{1-\alpha}(\widehat{\boldsymbol{\theta}}_n)$ is defined as

$$y_{1-\alpha}(\widehat{\boldsymbol{\theta}}_n) \equiv G^{-1}(1-\alpha; \widehat{\boldsymbol{\theta}}_n) \equiv \inf\{y : G(y; \widehat{\boldsymbol{\theta}}_n) \geq 1-\alpha\}.$$

Constructing Prediction Interval: Plug-in Method

The plug-in method, though easy to implement, ignores the uncertainty in $\widehat{\boldsymbol{\theta}}_n$ and typically has a coverage probability that is different from the nominal confidence level.

Under regularity assumptions, the coverage probability of the plug-in method has an error of the order of $O(1/n)$.

$$\Pr\left[Y \leq y_{1-\alpha}(\widehat{\boldsymbol{\theta}}_n)\right] = 1 - \alpha + c(\boldsymbol{\theta})/n + o\left(\frac{1}{n}\right).$$

- The approximate pivotal quantity is $G(Y; \widehat{\boldsymbol{\theta}}_n) \dot{\sim}$ **unif**$(0,1)$. The $1-\alpha$ quantile of **unif**$(0,1)$, which is $1-\alpha$ is used to replace the $1-\alpha$ quantile of $G(Y; \widehat{\boldsymbol{\theta}}_n)$.

Bootstrap Calibration

The bootstrap calibration prediction method starts by defining a quantity

$$U = G\left(Y; \widehat{\boldsymbol{\theta}}_n\right).$$

Let $u_{1-\alpha}$ be $1-\alpha$ quantile of U. Then we have

$$1-\alpha = \Pr(U \leq u_{1-\alpha}) = \Pr[Y \leq G^{-1}(u_{1-\alpha}; \widehat{\boldsymbol{\theta}}_n)].$$

Taking $1-\alpha + c(\boldsymbol{\theta})/n \equiv \alpha_c = 1 - u_{1-\alpha}$, the coverage probability of the $1-\alpha_c$ plug-in upper prediction bound will be $1-\alpha$.

- The distribution of U is denoted as $H(\cdot; \boldsymbol{\theta})$. Because $\boldsymbol{\theta}$ is unknown, $H(\cdot; \widehat{\boldsymbol{\theta}}_n)$ is used to get the $1-\alpha$ quantile as $\tilde{u}_{1-\alpha} = H^{-1}(1-\alpha; \widehat{\boldsymbol{\theta}}_{\boldsymbol{n}})$ (**unif**$(0,1)$ used in plug-in method).
- Because the function $H^{-1}(\cdot;\cdot)$ is usually not available, bootstrap method can be used to approximate $\tilde{u}_{1-\alpha}$.

A Different Viewpoint: Predictive Distribution

Definition: We use $\widetilde{y}_{1-\alpha}(x_n)$ as a generic notation to denote a $1-\alpha$ upper prediction bound using some prediction method based on data $X_n = x_n$. Then the corresponding predictive distribution $\widetilde{G}_p(\cdot|x_n)$ satisfies

$$\widetilde{G}_p[\widetilde{y}_{1-\alpha}(x_n)|x_n] = 1-\alpha \text{ for } \alpha \in (0,1).$$

- The predictive distribution defines the $1-\alpha$ upper prediction bound as the $1-\alpha$ quantile of the predictive distribution $\widetilde{G}_p(y|x_n)$.
- Correspondingly, given a predictive distribution $\widetilde{G}_p(y|x_n)$, we can treat the $1-\alpha$ quantile of $\widetilde{G}_p(y|x_n)$ as $1-\alpha$ upper prediction bound.

The Predictive Distribution for Bootstrap Calibration

- For the pivotal prediction method, [Lawless and Fredette, 2005] showed that the associated predictive distribution has the form $\widetilde{G}_p(y|x_n) = Q[q(x_n, y)]$, where $Q(\cdot)$ is the cdf of the pivotal quantity $q(X_n, Y)$.
- Similarly, when $q(X_n, Y)$ is an approximate pivotal quantity, the corresponding predictive distribution is

$$\widetilde{G}_p(y|x_n) = \widetilde{Q}[q(x_n, y)], \tag{1}$$

- By applying (1) to bootstrap calibration, the associated predictive distribution can be computed from

$$\widetilde{G}_p(y|x_n) = H[G(y; \widehat{\boldsymbol{\theta}}_n); \widehat{\boldsymbol{\theta}}_n] \approx \frac{1}{B}\sum_{b=1}^{B} G\left\{G^{-1}\left[G(y; \widehat{\boldsymbol{\theta}}_n); \widehat{\boldsymbol{\theta}}^{*}_{n,b}\right]; \widehat{\boldsymbol{\theta}}_n\right\}.$$

Construct a Predictive Distribution (Likelihood)

[Bjørnstad, 1990] reviewed that there are three types of operations to construct predictive likelihood (distribution) $\tilde{L}_p(x_n, y)$:

- maximization: $\widetilde{L}_p(x_n, y) = \sup_{\boldsymbol{\theta}} f_{X_n,Y}(x_n, y; \boldsymbol{\theta})$.
- conditioning: use sufficient statistic.
- integration: assign a data-based distribution to the parameter.

Predictive Likelihood via Integration I

The integration operation is to remove the unknown parameters by integrating out them with a data-based distribution.

1. A Bayesian predictive distribution is defined by

$$G_B(y|x_n) = \int_{\boldsymbol{\theta} \in \Theta} G(y; \boldsymbol{\theta}) p(\boldsymbol{\theta}|x_n) d\boldsymbol{\theta},$$

   where $p(\boldsymbol{\theta}|x_n)$ is the joint posterior distribution of the parameter $\boldsymbol{\theta}$ given the data $\boldsymbol{X}_n = x_n$ and $G(\cdot; \theta)$ is the cdf of the predictand Y. The Bayesian $1-\alpha$ (upper) prediction bound is defined as the $1-\alpha$ quantile of $G_B(y|x_n)$ as

$$\widetilde{Y}_{1-\alpha}^{Bayes} = \inf\{y : G_B(y|x_n) \geq 1-\alpha\}.$$

2. Confidence distribution [Xie and Singh, 2013].

Predictive Likelihood via Integration II

3. Bootstrap distribution [Harris, 1989].
4. Fiducial distribution [Wang et al., 2012].

Special Results for (Log-)Location-Scale Distribution (1)

The predictive distribution of the bootstrap calibration is

$$\widetilde{G}_p(y|x_n) = \mathrm{E}_*\Phi\left\{\frac{y-[\widehat{\mu}+\frac{\widehat{\sigma}}{\widehat{\sigma}^*}(\widehat{\mu}-\widehat{\mu}^*)]}{\widehat{\sigma}\frac{\widehat{\sigma}}{\widehat{\sigma}^*}}\right\} \approx \frac{1}{B}\sum_{b=1}^{B}\Phi\left(\frac{y-\widehat{\mu}_b^{**}}{\widehat{\sigma}_b^{**}}\right).$$

We define two new quantities $(\widehat{\mu}^{**}, \widehat{\sigma}^{**})$ using $(\widehat{\mu}, \widehat{\sigma})$ and $(\widehat{\mu}^*, \widehat{\sigma}^*)$,

$$\widehat{\mu}^{**} = \widehat{\mu} + \widehat{\sigma}\frac{\widehat{\mu}-\widehat{\mu}^*}{\widehat{\sigma}^*} \text{ and } \widehat{\sigma}^{**} = \widehat{\sigma}\frac{\widehat{\sigma}}{\widehat{\sigma}^*}.$$

Under complete or Type-II censored data, $(\widehat{\mu}^{**}, \widehat{\sigma}^{**})$ are the generalized pivotal quantities (GPQs) for (μ, σ).

Special Results for (Log-)Location-Scale Distribution (2) I

For location-scale distribution, it is well known that the probability matching prior is $\pi(\mu, \sigma) = \sigma^{-1}$. The credible interval has exact coverage, either for μ or σ.

Theorem 1: Under a complete sample X_n or a Type-II (time) censored sample X_n^{II} from a location-scale distribution with location parameter μ and scale parameter σ, the ML estimators are $\widehat{\mu}$ and $\widehat{\sigma}$. Y is an independent random variable from the same distribution as X_n. The quantity (U_1, U_2) is defined as $((\widehat{\mu} - \mu)/\widehat{\sigma}, \widehat{\sigma}/\sigma)$. Then,

1. the joint posterior distribution of (U_1, U_2) using prior $\pi(\mu, \sigma) \propto \sigma^{-1}$ is the same as the frequentist conditional distribution of (U_1, U_2) conditioned on ancillary statistic $A = ((X_1 - \widehat{\mu})/\widehat{\sigma}, \ldots, (X_{n-2} - \widehat{\mu})/\widehat{\sigma})$.

Special Results for (Log-)Location-Scale Distribution (2) II

2. The Bayesian upper prediction bound, which is defined as

$$\widetilde{Y}_{1-\alpha}^{Bayes} \equiv \inf\left\{ y : \int_{(\mu,\sigma)\in\boldsymbol{\Theta}} F(y;\mu,\sigma)p(\mu,\sigma|X_n)d\mu d\sigma \geq 1-\alpha \right\},$$

has exact coverage probability, i.e., $\Pr\left(Y \leq \widetilde{Y}_{1-\alpha}^{Bayes}\right) = 1-\alpha$, where $p(\mu,\sigma|\boldsymbol{X}_n = x_n)$ is the joint posterior distribution using prior $\pi(\mu,\sigma) = \sigma^{-1}$.

Special Results for (Log-)Location-Scale Distribution (2) III

3. By theorem 1 of [Hannig et al., 2016], the generalized fiducial distribution of (μ, σ) for a location-scale distribution is proportional to

$$r(\mu, \sigma | \boldsymbol{x}_n) \propto \frac{1}{\sigma^n} \prod_{i=1}^{n} \phi\left(\frac{x_n - \mu}{\sigma}\right) J(\mu, \sigma, \boldsymbol{x}_n),$$

where $J(\mu, \sigma, \boldsymbol{x}_n) = \sum_{1 \le i < j \le n} |x_i - x_j| \sigma^{-1}$. So, the GFD of (μ, σ) is the same as the posterior distribution using prior $\pi(\mu, \sigma) = \sigma^{-1}$.

Generalized Fiducial Distribution and GPQ

- GFD depends on the conditional distribution of (U_1, U_2) given the ancillary statistic.
- GPQ depends on the unconditional distribution of (U_1, U_2).

Conclusions

- Pivotal, plug-in, and bootstrap calibration methods are essentially the same class of prediction method.
- For location-scale distribution, bootstrap calibration is equivalent to integrating out parameters with the GPQ distribution.
- For location-scale distribution, the Bayesian prediction interval using modified Jeffreys prior has exact coverage probablity and it is equivalent to the fiducial prediction interval using the fiducial formula described in [Hannig et al., 2016].
- Simulation study shows that the bootstrap calibration has the best coverage even in small sample.
- Different prediction methods for binomial and Poisson distributions are compared and a pivotal-based method proposed by [Krishnamoorthy and Peng, 2011] has good coverage property.

Paper 2: Predicting the Number of Future Events

Overview

- This is a different type of prediction problem (within sample prediction), where the predictand (number of future events) depends on the observed data.
- We show that the naive plug-in method is not asymptotically correct for within-sample prediction.
- We provide solutions for constructing prediction intervals in within sample prediction and prove that they are asymptotically correct.
- This work has been submitted and is being revised.

Within Sample Prediction

Examples

$n = 10{,}000$ units of product were put into service and over the next 48 months, 80 failures occurred and the failure times were recorded. Management requested an upper prediction bound on the number of failures among the remaining 9920 units during the next 12 months.

- In new-sample prediction, past data are used, for example, to compute a prediction interval for the lifetime of a single unit from a new and completely independent sample.
- For within-sample prediction, however, the sample has not changed; the future random variable that we wish to predict (i.e., a count) relates to the same sample that provided the original (censored) data.

Notations

- Let $(T_1, ..., T_n)$ be a sample from a parametric distribution $F(t; \boldsymbol{\theta})$ having support on the positive real line.
- The available data may then be expressed by $D_i = (\delta_i, T_i^{obs}), i = 1, ..., n$, where $\delta_i = \mathrm{I}(T_i \leq t_c)$ and $T_i^{obs} = T_i\delta_i + t_c(1 - \delta_i)$.
- The observed number of events (uncensored units) in the sample will be denoted by $r_n = \sum_{i=1}^{n} \mathrm{I}(T_i \leq t_c)$. For a future time $t_w > t_c$, let $Y_n = \sum_{i=1}^{n} \mathrm{I}(T_i \in (t_c, t_w])$ denote the (future) number of values from $T_1, ..., T_n$, that occur in the interval $(t_c, t_w]$, $\boldsymbol{\theta} \in \mathbb{R}^q$.
- The goal is to construct a prediction interval for Y_n based on the observed data $\boldsymbol{D}_n = (D_1, ..., D_n)$ when $\boldsymbol{\theta}$ is unknown.

The Plug-in Method

The conditional distribution of Y_n is then $\mathrm{Binomial}(n - r_n, p)$ given the observed data $\boldsymbol{D}_n = (D_1, ..., D_n)$, where p is the conditional probability that $T_i \in (t_c, t_w]$ given that $T_i > t_c$. As a function of $\boldsymbol{\theta}$, we may define p by

$$p \equiv \pi(\boldsymbol{\theta}) = \frac{F(t_w; \boldsymbol{\theta}) - F(t_c; \boldsymbol{\theta})}{1 - F(t_c; \boldsymbol{\theta})}. \tag{2}$$

Let $\widehat{\boldsymbol{\theta}}_n$ denote an estimator of $\boldsymbol{\theta}$ based on D_n, then a plug-in estimator $\widehat{p}_n = \pi(\widehat{\boldsymbol{\theta}}_n)$ of the conditional probability p follows from (2). The plug-in upper prediction bound is

$$\tilde{Y}^{PL}_{n,1-\alpha} = \inf\{y \in \{0\} \cup \mathbb{Z}^{+}; \mathrm{pbinom}(y, n - r_n, \widehat{p}_n) \geq 1 - \alpha\}.$$

Regular Prediction Problem

However, plug-in estimation of prediction distributions has only been considered (& shown to be valid) for regular prediction problems by our following definition:

Definition

A prediction problem is called regular if

$$\sup_{y \in \mathbb{R}} \left| G_n(y|\boldsymbol{D}_n; \boldsymbol{\theta}) - G_n(y|\boldsymbol{D}_n; \widehat{\boldsymbol{\theta}}_n) \right| \xrightarrow{p} 0$$

holds as $n \to \infty$ for any consistent estimator $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ (i.e., $\widehat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}$) & $G_n(y|D_n; \theta)$ is the conditional cdf of $Y|D_n$.

Failure of the Plug-in Method

Theorem 1: The within sample prediction is **not regular** and the plug-in method is **not** asymptotically correct:

1. $\sup_{y\in\mathbb{R}} \left| G_n(y|\boldsymbol{D}_n, \boldsymbol{\theta}_0) - G_n(y|\boldsymbol{D}_n, \widehat{\boldsymbol{\theta}}_n)\right| \xrightarrow{d} 1 - 2\Phi_{\mathrm{nor}}(\sqrt{v_1}|Z_1|/2)$, for $Z_1 \sim N(0,1)$ and v_1 is a function of t_c and $\boldsymbol{\theta}_0$.
2. The plug-in upper prediction bound $\tilde{Y}^{PL}_{n,1-\alpha}$ generally fails to have an asymptotically correct coverage:

$$\lim_{n\to\infty} \Pr(Y_n \le \tilde{Y}^{PL}_{n,1-\alpha}) = \Lambda_{1-\alpha}(v_1) \in (0,1) \quad \text{such that}$$

$$\operatorname{sgn}\left[\Lambda_{1-\alpha}(v_1) - (1-\alpha)\right] = \begin{cases} 1 & \text{if } \alpha \in (1/2, 1) \\ 0 & \text{if } \alpha = 1/2 \\ -1 & \text{if } \alpha \in (0, 1/2), \end{cases}$$

Bootstrap Calibration for Within Sample Prediction

Proof in [Beran, 1990] does not apply to within-sample prediction.
We have established that the bootstrap calibration method is asymptotically correct.
To implement bootstrap calibration, bootstrap method is used to approximate the distribution of $U = \textbf{pbinom}(Y_n, n - r_n, \widehat{p}_n)$.
For the $100(1-\alpha)\%$ upper prediction bound, the calibrated confidence level is

$$1-\alpha_c = \inf\{u \in [0,1] : \Pr_*\left[\mathrm{pbinom}(Y_n^{\dagger}, n - r_n^*, \widehat{p}_n^*) \le u\right] \ge 1-\alpha\},$$

so that the calibrated $100(1-\alpha)\%$ upper prediction bound is $\tilde{Y}^{C}_{n,1-\alpha} = \tilde{Y}^{PL}_{n,1-\alpha_c}$.

Alternative: Direct/GPQ Bootstrap Distribution I

- A different type of approach is to construct prediction intervals by constructing the predictive distribution using integration operation. The following two methods are proven to be asymptotically correct.
- Direct Bootstrap: Letting $\Pr_*$ denote bootstrap probability (probability induced by a bootstrap sample $\boldsymbol{D}_n^*$), the direct bootstrap predictive distribution is

$$
\begin{aligned}
F_{Y_n}^{Boot}(y|\boldsymbol{D}_n) &= \int \mathrm{pbinom}(y, n - r_n, \widehat{p}_n^*) \Pr_*(d\widehat{p}_n^*) \\
&\approx \frac{1}{B}\sum_{b=1}^{B} \mathrm{pbinom}(y, n - r_n, \widehat{p}_b^*).
\end{aligned}
$$

Alternative: Direct/GPQ Bootstrap Distribution II

- For log-location-scale distribution, reparameterize the distribution to obtain the location and scale parameters. For example, the Weibull distribution with shape parameter β and scale parameter η, after taking logarithm,the location parameter is $1/\beta$ and the scale parameter is $\exp(\eta)$.
- Letting $\widehat{\boldsymbol{\theta}}_n^* = (\widehat{\mu}_n^*, \widehat{\sigma}_n^*)$ denote a bootstrap version of $\widehat{\boldsymbol{\theta}}_n = (\widehat{\mu}_n, \widehat{\sigma}_n)$, the GPQ bootstrap distribution is the resampling distribution of $\widehat{\boldsymbol{\theta}}_n^{**} = (\widehat{\mu}_n^{**}, \widehat{\sigma}_n^{**})$, where

$$\widehat{\mu}_n^{**} = \widehat{\mu}_n + \left(\frac{\widehat{\mu}_n - \widehat{\mu}_n^*}{\widehat{\sigma}_n^*}\right)\widehat{\sigma}_n \quad \text{and} \quad \widehat{\sigma}_n^{**} \quad = \left(\frac{\widehat{\sigma}_n}{\widehat{\sigma}_n^*}\right)\widehat{\sigma}_n.$$

  Similarly, $\widehat{p}_n^{**} = \pi(\widehat{\mu}_n^{**}, \widehat{\sigma}_n^{**})$ can be used to compute the GPQ bootstrap predictive distribution.

Simulation Study

- Proportion of Failure: $p_{f1} = F(t_c; \boldsymbol{\theta})$.
- Expected Number of Failures: $E(r) = np_{f1}$.
- The Prediction Time Window:
  $d = p_{f2} - p_{f1} = F(t_w; \boldsymbol{\theta}) - F(t_c; \boldsymbol{\theta})$.
- The Weibull Shape parameter β (the scale parameter is set as 1).

Simulation Study: Weibull Distribution, $d = 0.1$

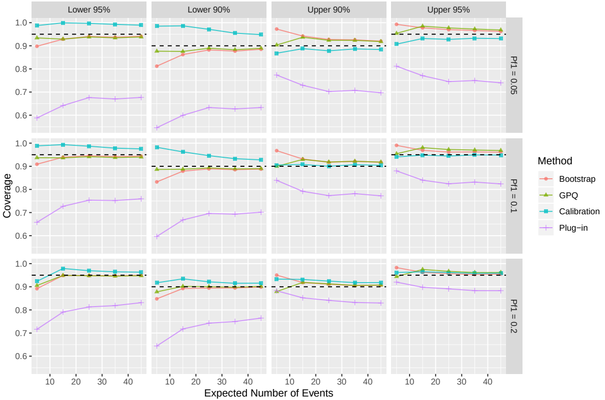

Figure: Coverage probabilities vs. expected number of events for the calibration, bootstrap and approximate GPQ prediction methods when $d = p_{f2} - p_{f1} = 0.1$ and $\beta = 2$.(Plug-in seen as invalid, direct/GPQ work well)

Multiple Cohort Within-Sample Prediction

Group i	Hours in Service	Group Size n_i	Failed r_i	At Risk $n_i - r_i$	$\hat{p}_i$	$(n_i - r_i) \times \hat{p}_i$
1	50	288	0	288	0.000763	0.2196
2	150	148	0	148	0.001158	0.1714
3	250	125	1	124	0.001558	0.1932
4	350	112	1	111	0.001962	0.2178
5	450	107	1	106	0.002369	0.2511
6	550	99	0	99	0.002778	0.2750
7	650	110	0	110	0.003189	0.3508
8	750	114	0	114	0.003602	0.4106
9	850	119	0	119	0.004016	0.4779
10	950	128	0	128	0.004432	0.5673
11	1050	124	2	122	0.004848	0.5915
12	1150	93	0	93	0.005266	0.4898
13	1250	47	0	47	0.005685	0.2672
14	1350	41	0	41	0.006105	0.2503
15	1450	27	0	27	0.006525	0.1762
16	1550	12	1	11	0.006946	0.0764
17	1650	6	0	6	0.007368	0.0442
18	1750	0	0	0	0.007791	0
19	1850	1	0	1	0.008214	0.0082
20	1950	0	0	0	0.008638	0
21	2050	2	0	2	0.009062	0.0181
Total	1703		6			5.062

Operationally, the **binom** functions (**pbinom**, **qbinom**, **rbinom**) are replaced by the **poibin** functions (**ppoibin**,**qpoibin**, **rpoibin**).

Recommendations

- The direct/GPQ bootstrap methods are preferred over bootstrap calibration because of better coverage probability and the bootstrap calibration is computationally unstable.
- Numerical study shows that when number of failures and proportion of failing are small, GPQ bootstrap has better coverage probability than direct bootstrap. The direct bootstrap method tends to be more conservative than the GPQ method on the upper prediction bound but under-coverage compared to the GPQ method on the lower prediction bound.

Constructing Prediction Intervals Using the Likelihood Ratio Statistic

Overview

- The goal of this project is to develop a general prediction method by using likelihood ratio statistic.
- This project is at an early stage.

Motivations

- In "*Predict the Number of Future Events*", all three proposed methods are based on bootstrap samples, which will cost extra computational resources and time.
- [Nelson, 2000] and [Nordman and Meeker, 2002] described a prediction method based on likelihood ratio for a similar prediction problem where the Weibull shape parameter is known.

Within Sample Prediction: Recap

t_c is the censoring time, r_n failures occurred in interval $(0, t_c]$. At time t_c, the numer of units at risk is $n - r_n$. The predictand is the number of failures in interval $(t_c, t_w]$.

Full Model

The distribution during interval $(0, t_c]$ may be different from that of interval $(t_c, +\infty)$.

Reduced Model

The time-to-failure process is governed by the same distribution in intervals $(0, t_c]$ and $(t_c, +\infty)$.

Suppose the distribution for the time-to-failure process in $(0, t_c]$ is $F(t; \boldsymbol{\theta})$. Under the **reduced** model, the conditional probability for a surviving unit to fail within $(t_c, t_w]$ is $p = (F(t_w; \boldsymbol{\theta}) - F(t_c; \boldsymbol{\theta}))/(1 - F(t_c; \boldsymbol{\theta}))$ while under **full** model, p is a free parameter and does not depend on $\boldsymbol{\theta}$.

Within Sample Prediction: Likelihood Ratio Statistic

Given the observated data $\boldsymbol{X}_n = \boldsymbol{x}_n$, suppose the realization of the predictand is $Y = y$.

Full Model Likelihood Function

$L_2(\boldsymbol{\theta}, p, \boldsymbol{x}_n, y) = C_{r,n} \prod_{i=1}^{r} f(x_{(i)}; \boldsymbol{\theta}) p^y (1-p)^{n-y-r}$

Reduced Model Likelihood Function

$L_1(\boldsymbol{\theta}, \boldsymbol{x}_n, y) =$
$C_{r,n} \prod_{i=1}^{r} f(x_{(i)}; \boldsymbol{\theta}) [F(t_w; \boldsymbol{\theta}) - F(t_c; \boldsymbol{\theta})]^y [1 - F(t_w; \boldsymbol{\theta})]^{n-y-r}$

Likelihood Ratio Statistic

$\Lambda_n(\boldsymbol{x}_n, y) = \frac{\max_{\boldsymbol{\theta}} L_1(\boldsymbol{\theta}, \boldsymbol{x}_n, y)}{\max_{\boldsymbol{\theta}, p} L_2(\boldsymbol{\theta}, p, \boldsymbol{x}_n, y)}$ and $-2 \log \Lambda_n \xrightarrow{d} \chi^2_1$ as $n \to \infty$.

Within Sample Prediction: Prediction Interval

- Using the asymptotic distribution of $\Lambda_n(\boldsymbol{X}_n, Y)$, given $\boldsymbol{X}_n = \boldsymbol{x}_n$ a $1-\alpha$ prediction set for Y can be constructed as

$$\{y : -2\log\Lambda_n(\boldsymbol{x}_n, y) \le \chi^2_{1,1-\alpha}\}.$$

- Because $-2\log\Lambda_n(\boldsymbol{x}_n, y)$ is a unimodal function of y, the prediction set could be reduced to a prediction interval by solving the two roots of the equation $-2\log\Lambda_n(\boldsymbol{x}_n, y) = \chi^2_{1,1-\alpha}$. Asymptotically, the two roots are the lower and upper $1-\alpha/2$ prediction bounds.
- This method does not need bootstrap samples and simulation study showed that it has comparable coverage probability as the direct bootstrap method/GPQ bootstrap method.

Likelihood Ratio Prediction as a General Method

The likelihood ratio prediction method has been applied to several examples using the "free one parameter" trick.

Examples

$\boldsymbol{X}_n$ is an independent sample from $\text{Norm}(\mu, \sigma)$ and Y is an independent random variable from the same distribution. Prediction interval for Y is needed.

Full Model:
We assume that $\boldsymbol{X}_n$ and Y share the same scale parameter but different location parameter. $\boldsymbol{X}_n \sim \text{Norm}(\mu_1, \sigma)$, $Y \sim \text{Norm}(\mu_2, \sigma)$

Reduced Model:
We assume that $\boldsymbol{X}_n$ and Y have the same distribution. $\boldsymbol{X}_n \sim \text{Norm}(\mu_0, \sigma)$, $Y \sim \text{Norm}(\mu_0, \sigma)$

For the **full model**, the ML estimators are

$$\widehat{\mu}_1 = \bar{X}_n, \widehat{\mu}_2 = Y, \widehat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n+1}}.$$

For the **reduced mode**, the ML estimators are

$$\widehat{\mu}_0 = \frac{\sum_{i=1}^n X_i + Y}{n+1}, \widehat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (X_i - \widehat{\mu}_0)^2 + (Y - \widehat{\mu}_0)^2}{n+1}}.$$

The likelihood ratio is

$$\Lambda_n(\boldsymbol{X}_n, Y) = \left[\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\sum_{i=1}^n (X_i - \widehat{\mu}_0)^2 + (Y - \widehat{\mu}_0)^2} \right]^{\frac{n+1}{2}}.$$

What is the distribution of $\Lambda_n(\boldsymbol{X}_n, Y)$?

Distribution of $\Lambda_n(\boldsymbol{X}_n, Y)$

- We cannot use χ_1^2 to approximate the distribution of $\Lambda_n(\boldsymbol{X}_n, Y)$ because $\widehat{\mu}_2 = Y$ is not consistent.
- Instead, the bootstrap distribution of $\Lambda_n(\boldsymbol{X}_n^*, Y^*)$ is used to determine the quantile of $\Lambda_n(\boldsymbol{X}_n, Y)$. Here $\boldsymbol{X}_n^*, Y^* \sim \text{Norm}(\widehat{\mu}, \widehat{\sigma})$, where $\widehat{\mu} = \bar{X}_n$ and $\widehat{\sigma} = \sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2/n}$.
- $\Lambda_n(\boldsymbol{X}_n, Y)$ is a pivotal quantity as
$$\Lambda_n(\boldsymbol{X}_n, Y) = \left[\frac{\sum_{i=1}^n \left(\frac{X_i-\mu}{\sigma} - \frac{\bar{X}_n-\mu}{\sigma}\right)^2}{\sum_{i=1}^n \left(\frac{X_i-\mu}{\sigma} - \frac{\widehat{\mu}_0-\mu}{\sigma}\right)^2 + \left(\frac{Y-\mu}{\sigma} - \frac{\widehat{\mu}_0-\mu}{\sigma}\right)^2}\right]^{\frac{n+1}{2}}, \text{ implying}$$
that the prediction interval has exact coverage probability.

Remaining Work I

We have considered the likelihood ratio prediction method in several examples and leads to the pivotal quantities when exist.

- In there a general method/principle here rather than a collection of examples?
- Is it possible to apply this method to more complicated prediction problems?

Remaining Work II

- How does likelihood ratio relate to other profile predictive distributions (i.e., likelihood based) mentioned earlier? In some 1-parameter distributional examples, these seem to match... but not in other examples e.g., iid $X_1, \ldots, X_n, Y \sim N(\mu, \sigma^2)$, LRT is equivalent to prediction intervals for Y from

$$T = \frac{Y - \bar{X}}{S_x \sqrt{\frac{1}{n} + 1}} \sim t_{n-1},$$

  but predictive distributions entail approximating $T \sim t_{n-3}$ (wrong df)

Remaining Work III

- New prediction intervals have been recently tailored for certain discrete distributions (Binomial, Poisson) using "joint sampling" approach ([Krishnamoorthy and Peng, 2011]) similar intervals would seem to follow from LR for these discrete distributions (and others) without tailoring.

References I

Beran, R. (1990).
Calibrating prediction regions.
Journal of the American Statistical Association, 85:715–723.

Bjørnstad, J. F. (1990).
Predictive likelihood: A review.
Statistical Science, 5:262–265.

Hannig, J., Iyer, H., Lai, R. C., and Lee, T. C. (2016).
Generalized fiducial inference: A review and new results.
Journal of the American Statistical Association, 111:1346–1361.

References II

- Harris, I. R. (1989).
  Predictive fit for natural exponential families.
  Biometrika, 76:675–684.

- Krishnamoorthy, K. and Peng, J. (2011).
  Improved closed-form prediction intervals for binomial and poisson distributions.
  Journal of Statistical Planning and Inference, 141:1709–1718.

- Lawless, J. F. and Fredette, M. (2005).
  Frequentist prediction intervals and predictive distributions.
  Biometrika, 92:529–542.

References III

Nelson, W. (2000).
Weibull prediction of a future number of failures.
Quality and Reliability Engineering International, 16:23–26.

Nordman, D. J. and Meeker, W. Q. (2002).
Weibull prediction intervals for a future number of failures.
Technometrics, 44:15–23.

Wang, C., Hannig, J., and Iyer, H. K. (2012).
Fiducial prediction intervals.
Journal of Statistical Planning and Inference, 142:1980–1990.

References IV

Xie, M.-G. and Singh, K. (2013).
Confidence distribution, the frequentist distribution estimator of a parameter: A review.
International Statistical Review, 81:3–39.