Prediction Interval Methods for Reliability Data

Qinglong Tian

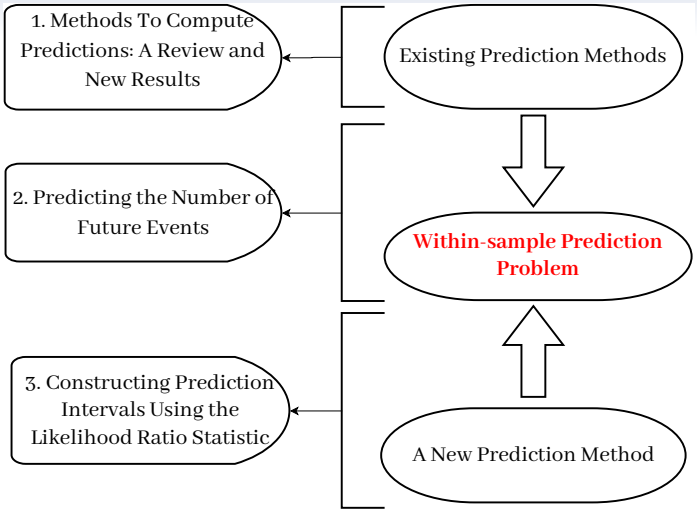
Department of Statistics
Iowa State University

April 15, 2021

Overview

1. Methods to Compute Prediction Intervals: A Review and New Results
2. Predicting the Number of Future Events
3. Constructing Prediction Intervals Using the Likelihood Ratio Statistic

The Structure of the Dissertation



Methods to Compute Prediction Intervals: A Review and New Results

Motivations and Setting

Motivations:

- [Patel, 1989] gave a review of prediction interval methods. Some important methods were not included and more research has been done on prediction methods since then.
- To find solutions to prediction problems.

Setting:

- We consider i.i.d. sample $\boldsymbol{X}_n \equiv (X_1, \ldots, X_n)$ from common density $f(x; \boldsymbol{\theta})$.
- The future random variable Y is independent of $\boldsymbol{X}_n$ and has density $g(y; \boldsymbol{\theta})$.
- We also consider the situation where $\boldsymbol{X}_n$ and Y are dependent.

Prediction Intervals and Coverage Probability

- (One-sided) prediction bound versus (two-sided) prediction interval.
- Conditional coverage probability versus unconditional coverage probability.

Pivotal Method

1. Find a function $q(\cdot, \cdot)$ of the data $\boldsymbol{X}_n$ and the future random variable Y.

2. The distribution $q(\boldsymbol{X}_n, Y)$ does not depend on any parameters (i.e., a pivotal quantity).

3. The prediction region $\{y : q(\boldsymbol{x}_n, y) \leq q_{1-\alpha}\}$ satisfies
$$\Pr\left[y : q(\boldsymbol{x}_n, y) \leq q_{1-\alpha}\right] = 1 - \alpha,$$
where $q_{1-\alpha}$ is the $1-\alpha$ quantile of $q(\boldsymbol{X}_n, Y)$.

Example: $X \sim \text{Norm}(\mu, 1/2)$ and $Y \sim \text{Norm}(\mu, 1/2)$.

Two special types of pivotal methods:
1. Inverting a hypothesis test.
2. Pivotal (conditional) cdf method

Pivotal Method: Inverting a Hypothesis Test

Suppose $\boldsymbol{X}_n \sim f(x;\boldsymbol{\theta})$ and $Y \sim g(y;\boldsymbol{\theta}^\dagger)$, let w_α be a size α critical region for testing $\boldsymbol{\theta} = \boldsymbol{\theta}^\dagger$; that is, $\Pr\left[(\boldsymbol{X}_n, Y) \in w_\alpha; \boldsymbol{\theta} = \boldsymbol{\theta}^\dagger\right] = \alpha$. Then a $1-\alpha$ prediction set can be defined as

$$\{y : (\boldsymbol{x}_n, y) \notin w_\alpha\}.$$

Pivotal Method: Pivotal (Conditional) Cdf Method

Suppose

- $T(\boldsymbol{X}_n)$ is a statistic from data $\boldsymbol{X}_n$
- $R(\boldsymbol{X}_n, Y)$ is a statistic from both the data $\boldsymbol{X}_n$ and the future random variable Y
- $G[t|R(\boldsymbol{x}_n, y)]$ is the conditional cdf of $T(\boldsymbol{X}_n)$ given $R(\boldsymbol{X}_n, Y)$, and $G[t|R(\boldsymbol{x}_n, y)]$ does not depend on any parameters
- $G[t|R(\boldsymbol{x}_n, y)]$ is a monotone function of y (we assume it is non-increasing here)

because $G\left[T(\boldsymbol{X}_n)|R(\boldsymbol{X}_n, Y)\right]$ has a unif$(0, 1)$ distribution, the $1-\alpha$ lower and upper prediction bounds are defined as

$$\underset{\sim}{Y}_{1-\alpha} = \inf\{y : G\left[T(\boldsymbol{x}_n)|R(\boldsymbol{x}_n, y)\right] < 1-\alpha\},$$
$$\widetilde{Y}_{1-\alpha} = \sup\{y : G\left[T(\boldsymbol{x}_n)|R(\boldsymbol{x}_n, y)\right] > \alpha\}.$$

Approximate Pivotal Method

- In many situations, a pivotal quantity is not available.
- An approximate pivotal quantity has a sampling distribution $Q_n(\cdot;\boldsymbol{\theta})$ that asymptotically converges to a limit distribution $Q(\cdot)$.
- A useful approximate pivotal quantity is

$$U \equiv G(Y;\widehat{\boldsymbol{\theta}}_n) \overset{d}{\sim} \text{unif}(0,1),$$

  where $Y \sim G(\cdot;\boldsymbol{\theta})$ and $\widehat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}$.
- Let $u_{1-\alpha}$ be the $1-\alpha$ quantile of U, then we have

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

- We need to approximate $u_{1-\alpha}$.

Approximate Pivotal Method: Plug-in

- The plug-in method uses $1-\alpha$ (i.e., $1-\alpha$ quantile of unif$(0,1)$) to approximate $u_{1-\alpha}$.
- The plug-in method uses $G(\cdot;\widehat{\boldsymbol{\theta}}_n)$ as the distribution for Y.
- The plug-in method is easy to implement but the error of the coverage probability is of order $O(1/n)$.

Approximate Pivotal Method: Calibration-bootstrap Method

- Let the cdf of $U \equiv G(Y; \widehat{\boldsymbol{\theta}}_n)$ be $H_n(\cdot; \boldsymbol{\theta})$.
- The calibration method uses $H_n^{-1}(1-\alpha; \widehat{\boldsymbol{\theta}}_n)$ to approximate $u_{1-\alpha}$.
- When the closed form of $H_n(\cdot; \boldsymbol{\theta})$ (as well as $H_n^{-1}(\cdot; \boldsymbol{\theta})$) is not available, we can use bootstrap method to approximate $H_n^{-1}(1-\alpha; \widehat{\boldsymbol{\theta}}_n)$.
- The error of the coverage probability is at rate $O(n^{-2})$.

Approximate Pivotal Method: Calibration Using an Asymptotic Expansion

Let $y_{1-\alpha_c}(\widehat{\boldsymbol{\theta}}_n)$ be the $1-\alpha_c$ upper prediction bound using plug-in method.

$$\begin{aligned}
&\Pr\left[Y \le y_{1-\alpha_c}(\widehat{\boldsymbol{\theta}}_n)\right] = \mathrm{E}\left\{1-\alpha_c + (\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})\kappa'_{\alpha_c}(\boldsymbol{\theta}) + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})^2\kappa''_{\alpha_c}(\boldsymbol{\theta})\right\} + o\left(\frac{1}{n}\right)\\
=&\mathrm{E}\left\{1-\alpha_c + (\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})\kappa'_{\alpha}(\boldsymbol{\theta}) + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})^2\kappa''_{\alpha}(\boldsymbol{\theta})\right\} + O\left(\frac{1}{n^2}\right) + o\left(\frac{1}{n}\right)\\
=&1-\alpha_c + c(\boldsymbol{\theta})/n + o\left(\frac{1}{n}\right),
\end{aligned}$$

By letting $\alpha_c = \alpha + c(\widehat{\boldsymbol{\theta}}_n)/n$, we can reduce the error from $O(n^{-1})$ to $o(n^{-1})$.

Approximate Pivotal Method: Predictive Distribution

- For pivotal method where $q(\boldsymbol{X}_n, Y) \sim Q_n(\cdot)$, the corresponding predictive distribution is
$$\widetilde{G}_p(y|\boldsymbol{x}_n) = Q_n\left[q\left(\boldsymbol{x}_n, y\right)\right].$$
- For approximate pivotal method where $q(\boldsymbol{X}_n, Y) \sim Q_n(\cdot; \boldsymbol{\theta})$, the corresponding predictive distribution is
$$\widetilde{G}_p(y|\boldsymbol{x}_n) = \widetilde{Q}_n[q(\boldsymbol{x}_n, y)],$$
where $\widetilde{Q}_n(\cdot)$ is an estimate of $Q_n(\cdot; \boldsymbol{\theta})$.
  1. For the plug-in method, $\widetilde{G}_p(y|\boldsymbol{x}_n) = G(y; \widehat{\boldsymbol{\theta}}_n)$ because $\widetilde{Q}_n(u) = u$ for $u \in (0, 1)$ and $q(\boldsymbol{x}_n, y) = G(y; \widehat{\boldsymbol{\theta}}_n)$
  2. For calibration-bootstrap method, $\widetilde{G}_p(y|\boldsymbol{x}_n) = H[G(y; \widehat{\boldsymbol{\theta}}_n); \widehat{\boldsymbol{\theta}}_n]$

Predictive Distribution I

To construct a predictive distribution, we need to get rid of the unknown parameters from the joint likelihood of $(\boldsymbol{X}_n, Y)$.

- Maximizing:
$$\widetilde{L}_p(\boldsymbol{x}_n, y) = \sup_{\boldsymbol{\theta}} p(\boldsymbol{x}_n, y; \boldsymbol{\theta}).$$
- Conditioning:
Let $R = R(S, T)$, S, and T respectively denote minimal sufficient statistics for $(\boldsymbol{X}_n, Y)$, $\boldsymbol{X}_n$, and Y. The predictive likelihood for $T = t$ given $S = s$ is defined as
$$\widetilde{L}_p(t|s) = p(s|r) = p\left[s|r(s,t)\right].$$

Predictive Distribution II

- Integrating:
  1. The Bayesian method.
  2. Direct/GPQ-bootstrap.
  3. Using a fiducial distribution.
  4. Using a confidence distribution.

Special Results for (Log-)Location-Scale Distribution

For (Log-)location-scale distribution, the calibration-bootstrap method

- is equivalent to a predictive distribution based on integrating out the parameters with the GPQ distribution.
- is equivalent to a pivotal method for complete or Type-II censored data; thus having exact coverage probability.

For (Log-)location-scale distribution with complete or Type-II censored data, the Bayesian prediction interval using the prior $\pi(\mu, \sigma) \propto \sigma^{-1}$

- has exact coverage probability.
- is equivalent to a predictive distribution based on the generalized fiducial distribution (GFD).

Prediction Intervals for Discrete Distributions

The prediction interval methods for discrete distribution follow the same two prediction principles discussed in the previous slides (i.e., (approximate) pivotal methods and predictive distribution (likelihood) methods).

The Pivotal (Conditional) Cdf Method for Discrete Distributions

- $T(\boldsymbol{X}_n)$ is a statistic whose conditional distribution is a discrete function of $\boldsymbol{X}_n$ and Y, say $R(\boldsymbol{X}_n, Y)$, that does not depend on any unknown parameters.
- The pivotal (conditional) cdf method described before cannot be used directly because $G\left[T(\boldsymbol{X}_n|R(\boldsymbol{X}_n, Y))\right]$ is no longer a pivotal quantity.
- Nevertheless, $G\left[T(\boldsymbol{X}_n|R(\boldsymbol{X}_n, Y))\right]$ is stochastically ordered w.r.t. the unif$(0, 1)$ distribution.
- Suppose $G[T(\boldsymbol{x}_n)|R(\boldsymbol{x}_n, y)]$ is a non-increasing function of y, the lower and upper bounds are defined as

$$\underset{\sim}{Y}_{1-\alpha} = \inf\left\{y : 1 - G\left[T(\boldsymbol{x}_n) - 1|R(T(\boldsymbol{x}_n), y)\right] > \alpha\right\},$$
$$\widetilde{Y}_{1-\alpha} = \sup\left\{y : G\left[T(\boldsymbol{x}_n)|R(T(\boldsymbol{x}_n), y)\right] > \alpha\right\}.$$

- This method is guaranteed to have coverage probability that is greater or equal to the nominal level.

Prediction Methods for Dependent Data

- Similar prediction methods exist for dependent data along the lines of the plug-in, calibration-bootstrap, and (integration-based) predictive distribution methods.
- Prediction strategies developed for independent observations may fail if directly applied to dependent data.
- There exists a variety of ways to generate bootstrap samples under dependence, particularly for the time series.
- The prediction procedures have been largely studied and justified in the context of Gaussian process models.
- When parametric bootstrap methods are hard to justify, we can use residual-based bootstraps to formulate prediction interval for future random variable.

Predicting the Number of Future Events

Predicting the Number of Future Events

Overview

- Explain a type of prediction problem, called *within-sample prediction*, where the future target variable depends on (or is not independent from) the same sample that provides the original data.
- Show that the standard plug-in method is not valid for within-sample prediction.
- Provide valid alternatives for constructing prediction intervals for within-sample prediction, based on bootstrap resampling

Within-sample Prediction I

- It is useful to distinguish between "new-sample" prediction and "within-sample" prediction.
- For "new-sample" prediction, data from a past sample is used to make predictions on a future unit or sample of units from the same process or population.
- For within-sample prediction, the problem is to predict future events in a sample or process based on early data from that sample or process.

Within-sample Prediction II

Examples

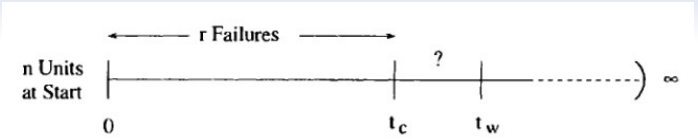
$n = 10{,}000$ units of product were put into service.

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.

Basic problem structure is to predict a future count (i.e., how many failures will occur in a future time period) among units that have not yet failed.

Within-sample Prediction III

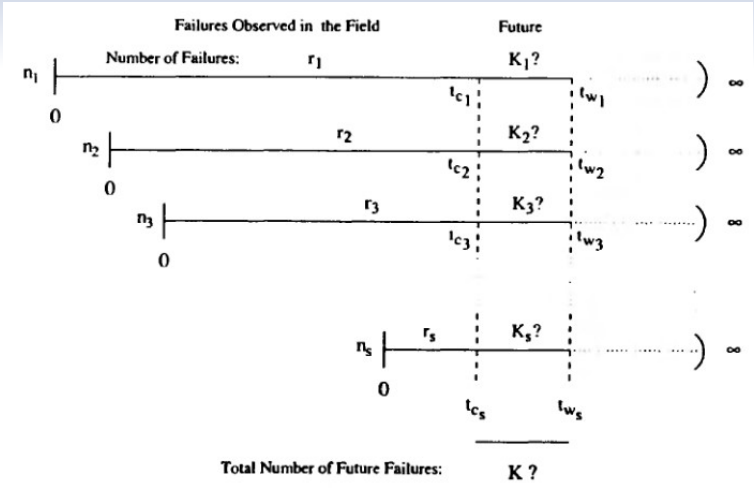


Source: [Meeker and Escobar, 1998]

Figure: Within-sample prediction.

- Time of the next failure.
- Time until k additional failures.
- Number of additional failures in a future interval (t_c, t_w).

Within-sample Prediction IV



Source: [Meeker and Escobar, 1998]

Figure: Illustration of staggered entry prediction.

Evaluating Different Prediction Methods

- Let $\mathrm{PI}_{1-\alpha}(\boldsymbol{X}_n)$ be a prediction interval with nominal confidence level $1-\alpha$.
- The (unconditional) coverage probability is

$$\mathrm{CP}\left[\mathrm{PI}_{1-\alpha}(\boldsymbol{X}_n)\right] = \Pr\left[Y \in \mathrm{PI}_{1-\alpha}(\boldsymbol{X}_n)\right].$$

- If $\mathrm{CP}[\mathrm{PI}_{1-\alpha}(\boldsymbol{X}_n)] = 1-\alpha$, we say such prediction method is exact.
- If $\mathrm{CP}[\mathrm{PI}_{1-\alpha}(\boldsymbol{X}_n)] \to 1-\alpha$ as $n \to \infty$, we say such prediction method is asymptotically correct.

Statistical Formulation

- Suppose $T_1, ..., T_n$ denote a random sample of "event times" following a parametric distribution $F(t; \boldsymbol{\theta})$ on the positive real line.

- At a freeze-time $t_c > 0$, the observed data $D_1, \ldots, D_n$ where

$$D_i \equiv \begin{cases} T_i & \text{if } T_i \leq t_c \\ t_c & \text{if } T_i > t_c, \end{cases} \qquad i = 1, \ldots, n.$$

  & $r_n \equiv \sum_{i=1}^{n} \mathrm{I}(T_i \leq t_c)$ is number of observed events

- For a future time $t_w > t_c$, we wish to predict the future count

$$Y_n \equiv \sum_{i=1}^{n} \mathrm{I}(T_i \in (t_c, t_w]) \qquad \text{i.e., } Y_n \in [0, n - r_n]$$

  as the number of values from $T_1, ..., T_n$ occurring in $(t_c, t_w]$, based on the observed data $\boldsymbol{D}_n = (D_1, ..., D_n)$ when $\boldsymbol{\theta}$ is unknown.

True Distribution of Y_n & Prediction Bounds

Conditional Distribution of Count Y_n given Data $\boldsymbol{D}_n = (D_1, ..., D_n)$

Y_n has a $\text{binomial}(n - r_n, p)$ distribution where p is the conditional probability that $T_i \in (t_c, t_w]$ given that $T_i > t_c$, or

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

in terms of the event-time distribution $F(t; \boldsymbol{\theta}) \equiv P(T_i \leq t; \boldsymbol{\theta})$, $t > 0$.

- We'd like (say) a $1 - \alpha$ upper prediction bound $\tilde{Y}_{n,1-\alpha}$ for Y_n, where $\tilde{Y}_{n,1-\alpha}$ is some function of data $\boldsymbol{D}_n$ that satisfies

$$\Pr\left[Y_n \leq \tilde{Y}_{n,1-\alpha}(\boldsymbol{D}_n); \boldsymbol{\theta}\right] \approx 1 - \alpha$$

- If $\boldsymbol{\theta}$ were known, take $\tilde{Y}_{n,1-\alpha}$ as quantile of $\text{binomial}(n - r_n, p)$

Standard Plug-In Method for Prediction

- Let $\widehat{\boldsymbol{\theta}}_n$ denote an estimator of $\boldsymbol{\theta}$ based on data $\boldsymbol{D}_n$
- Estimate the conditional probability $p \equiv p(\boldsymbol{\theta})$ as $\widehat{p}_n = p(\widehat{\boldsymbol{\theta}}_n)$
- Approximate the distribution of Y as $\mathrm{binomial}(n - r_n, \widehat{p}_n)$
- Get an upper $1 - \alpha$ Plug-in prediction bound for Y as

$$\tilde{Y}^{PL}_{n,1-\alpha} \equiv 1 - \alpha \text{ quantile of } \mathrm{binomial}(n - r_n, \widehat{p}_n)$$

The Plug-in Method

Let $G(\cdot|\boldsymbol{X}_n;\widehat{\boldsymbol{\theta}}_n)$ be the conditional cdf of the future random variable Y given data $\boldsymbol{X}_n$. For the plug-in method to work,

- Using the probability integral transform, we need the following convergence (for continuous random variable)

$$G(Y|\boldsymbol{X}_n,\widehat{\boldsymbol{\theta}}_n) \xrightarrow{d} G(Y|\boldsymbol{X}_n,\boldsymbol{\theta}_0) \stackrel{d}{=} \text{Unif}(0,1),$$

  as $n \to \infty$.

- More generally, the following convergence needs to hold uniformly,

$$\sup_{y\in\mathbb{R}} |G(y|\boldsymbol{X}_n;\boldsymbol{\theta}_0) - G(y|\boldsymbol{X}_n;\widehat{\boldsymbol{\theta}}_n)| \xrightarrow{p} 0$$

  as $n \to \infty$.

Failure of the Plug-in Method in Within-sample Prediction I

- The convergence above does not hold for within-sample prediction.
- Here, we assume that $\hat{p}_n \equiv \pi(\widehat{\boldsymbol{\theta}}_n) = p_0 + O_p(n^{-1/2})$, which is the standard $\sqrt{n}$-consistency.
- Let $G_n(\cdot|D_n, \boldsymbol{\theta}_0) \sim \text{binomial}(n - r_n, p)$ be the conditional cdf for Y_n, we can show that

$$\sup_{y \in \mathbb{R}} \left| G_n(y|\boldsymbol{D}_n, \boldsymbol{\theta}_0) - G_n(y|\boldsymbol{D}_n, \widehat{\boldsymbol{\theta}}_n) \right| \xrightarrow{d} 1 - 2\Phi_{\text{nor}}(\sqrt{v_1}|Z_1|/2),$$

  for $Z_1 \sim N(0, 1)$ and v_1 is a function of t_c and $\boldsymbol{\theta}_0$.

Failure of the Plug-in Method in Within-sample Prediction II

- The $1-\alpha$ upper prediction bound is then defined as

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

- We can show that $\tilde{Y}_{n,1-\alpha}^{PL}$ generally fails as

$$\lim_{n\to\infty} \Pr(Y_n \leq \tilde{Y}_{n,1-\alpha}^{PL}) = \Lambda_{1-\alpha}(\nu_1) \in (0,1) \quad \text{such that}$$

$$\mathrm{sgn}\left[\Lambda_{1-\alpha}(\nu_1) - (1-\alpha)\right] = \begin{cases} 1 & \text{if } \alpha \in (1/2, 1) \\ 0 & \text{if } \alpha = 1/2 \\ -1 & \text{if } \alpha \in (0, 1/2) \end{cases}.$$

Failure of the Plug-in Method in Within-sample Prediction III

- For large n, $Y_n \sim \text{binomial}(n - r_n, p)$ acts $\dfrac{Y_n - p(n - r_n)}{\sqrt{n - r_n}} \overset{d}{\approx} \text{Normal}$,

  but plug-in estimation produces an approximation

$$\frac{Y_n - \widehat{p}_n(n - r_n)}{\sqrt{n - r_n}} = \frac{Y_n - p(n - r_n)}{\sqrt{n - r_n}} + \sqrt{n - r_n}(\widehat{p}_n - p)$$

- The second part doesn't converge to zero (in fact, has a normal limit)

- As $n \to \infty$ (more data), the mean $p(n - r_n)$ of Y also diverges & an estimator $\widehat{p}_n$ of p is never "precise enough" to capture the mean of Y

Calibrating the Plug-in Prediction Intervals

- If $\boldsymbol{\theta}$ is known, we can calibrate the nominal confidence level $1-\alpha$ by finding a $1-\alpha_c$ such that for a $1-\alpha_c$ plug-in upper prediction bound $\tilde{Y}^{PL}_{1-\alpha_c}$,

$$\text{CP}\left[\text{PI}(1-\alpha_c);\boldsymbol{\theta}\right] = \text{Pr}_{\boldsymbol{\theta}}\left(Y \leq \tilde{Y}^{PL}_{1-\alpha_c}\right) = 1-\alpha.$$

- The basic idea of the calibration method is to evaluate $\text{CP}\left[\text{PI}(1-\alpha);\boldsymbol{\theta}\right]$ at $\widehat{\boldsymbol{\theta}}_n$ and find a calibration value $1-\alpha_c$ such that for a $1-\alpha_c$ plug-in upper prediction bound $\tilde{Y}^{PL}_{1-\alpha_c}$,

$$\text{CP}\left[\text{PI}(1-\alpha_c);\widehat{\boldsymbol{\theta}}_n\right] = \text{Pr}_{\widehat{\boldsymbol{\theta}}_n}\left(Y \leq \tilde{Y}^{PL}_{1-\alpha_c}\right) = 1-\alpha.$$

Calibration Method for the Within-sample Prediction Problem

- To implement the calibration idea, first we need to construct a predictive root $U_n = \text{pbinom}(Y_n, n - r_n, \hat{p}_n)$.
- The calibrated nominal confidence level $1 - \alpha_c$ is given by

$$1 - \alpha_c = \inf\{u \in [0, 1] : \Pr_{\widehat{\theta}_n}(U_n \leq u) \geq 1 - \alpha\}.$$

- The calibrated $1 - \alpha$ upper prediction bound is $\tilde{Y}^C_{1-\alpha} = \tilde{Y}^{PL}_{1-\alpha_c}$, where

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

- The calibration method is asymptotically correct (cf. [Tian et al., 2021]),

$$\lim_{n\to\infty} \Pr_{\boldsymbol{\theta}}(Y \leq \tilde{Y}^C_{1-\alpha}) = \lim_{n\to\infty} \Pr_{\boldsymbol{\theta}}(Y \leq \tilde{Y}^{PL}_{1-\alpha_c}) = 1 - \alpha.$$

Calibration by Simulation of the Sampling/Prediction Process

To solve u from $\inf\{u \in [0,1] : \Pr_{\widehat{\theta}_n}(U_n \leq u) \geq 1-\alpha\}$ using bootstrap:

1. Pretend $F(\cdot;\boldsymbol{\theta})$ is $F(\cdot;\widehat{\boldsymbol{\theta}}_n)$ and generate sample $\boldsymbol{D}_n^* = (D_1^*, \ldots, D_n^*)$.
2. Compute analogs of r_n^*, $\widehat{\boldsymbol{\theta}}_n^*$, and $\hat{p}_n^* = p(\widehat{\boldsymbol{\theta}}_n^*)$.
3. Compute $U_n^* = \text{pbinom}(Y_n^*, r_n^*, \hat{p}_n^*)$.
4. Repeat steps 1-3 for many times ($b = 1, \ldots, B$) and get a sample $\left\{U_{n,b}^*\right\}_{b=1}^B$.
5. The calibrated confidence level $1-\alpha_c$ is the $1-\alpha$ sample quantile of $\left\{U_{n,b}^*\right\}_{b=1}^B$.

Predictive Distributions and Prediction Intervals

- Let $\tilde{y}_{1-\alpha}(\boldsymbol{x}_n)$ be a $1-\alpha$ upper prediction bound computed using some prediction interval procedure.
- We can define the corresponding predictive distribution in form of a cdf as $\tilde{G}_p(\cdot|\boldsymbol{x}_n)$ by
$$\widetilde{G}_p[\widetilde{y}_{1-\alpha}(\boldsymbol{x}_n)|\boldsymbol{x}_n] = 1-\alpha \text{ for } \alpha \in (0,1).$$
- 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.
- There is a matching predictive distribution for any prediction intervals, and vice versa.

Alternative Methods based on Predictive Distributions

The strategy is to construct a predictive distribution so that its corresponding prediction interval has asymptotically correct coverage.

- A predictive distribution does not depend on any unknown parameters and is based on observed data $\boldsymbol{X}_n = \boldsymbol{x}_n$.
- We can obtain a predictive distribution by integrating out the unknown parameters with some distribution based on the observed data.
- In within-sample prediction problem, Y_n has a binomial distribution,

$$G(y|\boldsymbol{d}_n; \boldsymbol{\theta}) \sim \text{binomial}(n - r_n, p).$$

Direct-bootstrap Method I

- In the calibration method, we use the bootstrap version of $\hat{p}$ (i.e., $\hat{p}_n^*$) to compute U_n^*.
- The distribution of $\hat{p}_n^* \equiv \hat{p}_n^*(\boldsymbol{d}_n)$ is based on the observed data $\boldsymbol{D}_n = \boldsymbol{d}_n$ and does not depend on any unknown parameters.
- We can use the distribution of $\hat{p}_n^*$ to integrate out the unknown p in the distribution of Y_n.
- The direct-bootstrap predictive distribution is

$$\tilde{G}_{Y_n}^{Direct}(y|\boldsymbol{d}_n) = \int \text{pbinom}(y, n - r_n, \widehat{p}_n^*) \Pr_*(d\widehat{p}_n^*) \approx \frac{1}{B} \sum_{b=1}^{B} \text{pbinom}(y, n - r_n, \widehat{p}_{n,b}^*).$$

Here $\Pr_*(\cdot)$ is the bootstrap probability induced by the bootstrap sample $\boldsymbol{D}_n^*$.

Direct-bootstrap Method II

- With the direct-bootstrap predictive distribution, we can compute the corresponding prediction interval. The $1-\alpha$ upper prediction bound is given by

$$\tilde{Y}_{n,1-\alpha}^{Direct} = \inf\left\{y \in \{0\} \cup \mathbb{Z}^+ : \tilde{G}_{Y_n}^{Direct}(y|\boldsymbol{d}_n) \geq 1-\alpha\right\}.$$

- The resulting prediction interval is asymptotically correct (cf. [Tian et al., 2021]),

$$\lim_{n\to\infty} \Pr\left(Y_n \leq \tilde{Y}_{n,1-\alpha}^{Direct}\right) = 1-\alpha.$$

GPQ-bootstrap Method I

- We usually use a log-location-scale distribution to model the lifetime of a unit: lognormal, Weibull, and Fréchet distributions. For a log-location-scale distribution, we have $\boldsymbol{\theta} = (\mu, \sigma)$ and $p \equiv \pi(\mu, \sigma)$.
- Letting $\widehat{\boldsymbol{\theta}}_n^* = (\widehat{\mu}_n^*, \widehat{\sigma}_n^*)$ denote a bootstrap version of $\widehat{\boldsymbol{\theta}}_n = (\widehat{\mu}_n, \widehat{\sigma}_n)$,

$$\text{approximate dist. of "pivot"} \quad \left(\frac{\widehat{\mu}_n - \mu}{\widehat{\sigma}_n}, \frac{\widehat{\sigma}_n}{\sigma} \right) \quad \approx \quad \left(\frac{\widehat{\mu}_n^* - \widehat{\mu}_n}{\widehat{\sigma}_n^*}, \frac{\widehat{\sigma}_n^*}{\widehat{\sigma}_n} \right)$$

- "Solving" the above for parameters (μ, σ) gives the GPQ bootstrap distribution of $\widehat{\boldsymbol{\theta}}_n^{**} = (\widehat{\mu}_n^{**}, \widehat{\sigma}_n^{**})$, where

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

- The GPQ-bootstrap version of p is given by $\hat{p}_n^{**} \equiv \pi(\widehat{\mu}_n^{**}, \widehat{\sigma}_n^{**})$.

GPQ-bootstrap Method II

- The GPQ-bootstrap predictive distribution is

$$\tilde{G}_{Y_n}^{GPQ}(y|\boldsymbol{d}_n) = \int \text{pbinom}(y, n - r_n, \widehat{p}_n^{**}) \, \text{Pr}_*\left(d\widehat{p}_n^{**}\right) \approx \frac{1}{B} \sum_{b=1}^{B} \text{pbinom}(y, n - r_n, \widehat{p}_{n,b}^{**}).$$

- The $1 - \alpha$ upper prediction bound is given by

$$\tilde{Y}_{n,1-\alpha}^{GPQ} = \inf\left\{ y \in \{0\} \cup \mathbb{Z}^+ : \tilde{G}_{Y_n}^{GPQ}(y|\boldsymbol{d}_n) \geq 1 - \alpha \right\}.$$

- The resulting prediction interval is asymptotically correct (cf. [Tian et al., 2021]),

$$\lim_{n\to\infty} \text{Pr}\left(Y_n \leq \tilde{Y}_{n,1-\alpha}^{GPQ} \right) = 1 - \alpha.$$

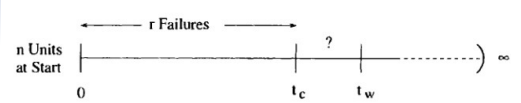
Simulation Study I

We have three alternative methods that are asymptotically correct:

- Calibration method (via bootstrap).
- Direct-bootstrap method.
- GPQ-bootstrap method.

The goal of the simulation study is to investigate their performances when the sample size is limited.

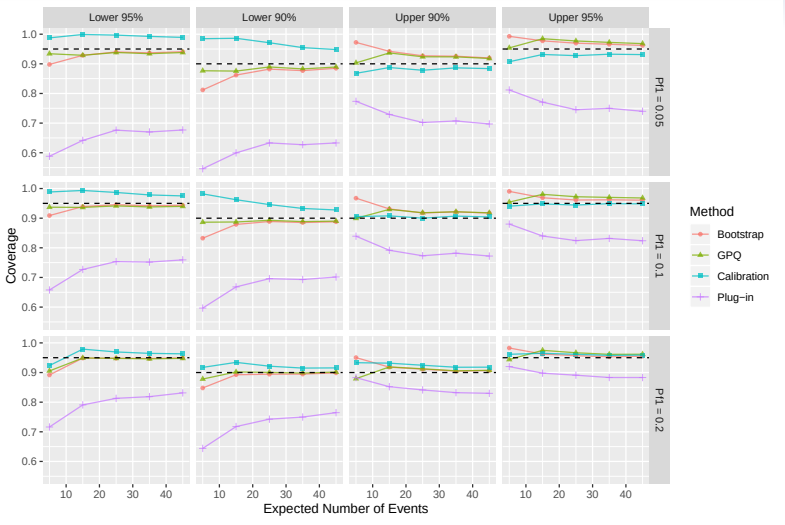
Simulation Study II



We used the Weilbull distribution to model the lifetime of the units. Here are the factors we used in the simulation

- Expected proportion of failures at the censoring time (p_{f1}): $p_{f1} = F(t_c; \boldsymbol{\theta})$.
- Expected number of failures at the censoring time (E(r)): E(r) $= np_{f1}$.
- The length of the prediction window $(t_c, t_w]$ (d): $d = p_{f2} - p_{f1} = F(t_w; \boldsymbol{\theta}) - F(t_c; \boldsymbol{\theta})$.
- The Weibull shape parameter β.
- WLOG, we set the Weibull scale parameter as $\eta = 1$.

Simulation Study III



- $d = 0.1$
- $\beta = 2$

Within-Sample Prediction

Example 1

$n = 10{,}000$ 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.

Fit Weibull model $\hat{\beta} = 1.52$, $\eta = 1152$, $\hat{p} = 0.003$ (next 12 months)

Confidence Level	Bound Type	Plug-in	Direct	GPQ	Calibration
95%	Lower	23	20	20	20
90%	Lower	25	23	23	23
90%	Upper	39	43	43	43
95%	Upper	42	47	47	46

Within-Sample Prediction

Example 2 (few failures without exact times)

A nuclear power plant has steam generators with $n = 20{,}000$ heat-exchanger tubes. Annual inspections found

1 cracked tube in 1st year+1 tube in 2nd year+6 more tubes in 3rd year.

A prediction was needed for the number of remaining 19,992 tubes that would crack in next 7 years.

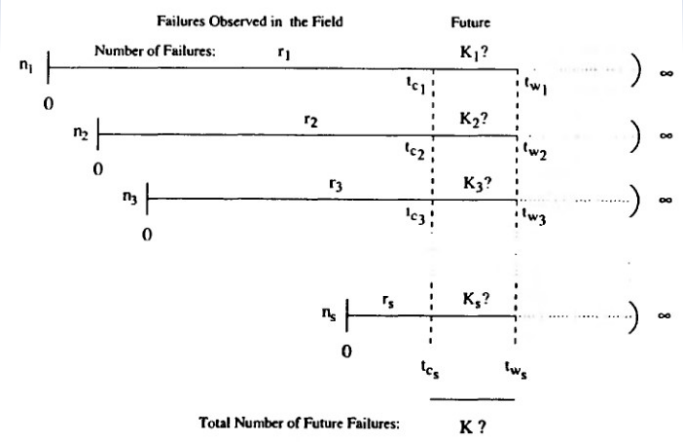
Fit Weibull model $\hat{\beta} = 2.53$, $\eta = 66.1$, $\hat{p} = 0.007$ (next 7 years)

Confidence Level	Bound Type	Plug-in	Direct	GPQ	Calibration
95%	Lower	138	28	23	NA
90%	Lower	142	43	34	NA
90%	Upper	176	1627	888	NA
95%	Upper	180	4343	1890	NA

Recommendations

- The direct/GPQ bootstrap methods are preferred over the calibration method because of
  1. Better coverage probability.
  2. Simpler procedure.
  3. Computational stability (cf. [Tian et al., 2021, Sec. 9.1]).
- Direct versus GPQ bootstrap methods:
  1. When the expected number of events is moderately large, these two methods have similar performances.
  2. Numerical study shows that when both the number of events and proportion of failing are small, GPQ-bootstrap method has better coverage probability than the direct-bootstrap method.
  3. Under the situation described in 2, the direct bootstrap method is too conservative in the upper bound but under-coverage for in the lower bound.

Staggered Entry Prediction (Multiple Cohort)



Source: [Meeker and Escobar, 1998]

Figure: Illustration of staggered entry prediction.

Multiple Cohort versus Single Cohort

- The multiple cohort prediction problem can be seen as a collection of independent single cohort prediction problems.
- The future random variable Y_n in the multiple cohort prediction problem is the sum of K independent but heterogeneously distributed binomial random variables, where K is the number of cohorts.

'The Poisson binomial distribution is the discrete probability distribution of a sum of independent Bernoulli trials that are not necessarily identically distributed.'

Wikipedia

Extending the Calibration Method

Single-cohort:

- The predictive root is $U_n = \text{pbinom}(Y_n, n - r_n, \hat{p}_n)$.
- The calibrated nominal confidence level is $1 - \alpha_c = \inf\{u \in [0,1] : \Pr_{\widehat{\boldsymbol{\theta}}_n}(U_n \leq u) \geq 1 - \alpha\}$.
- The calibrated $1 - \alpha$ upper prediction bound is $\tilde{Y}^{PL}_{n,1-\alpha_c} = \inf\{y \in \{0\} \cup \mathbb{Z}^+; \text{pbinom}(y, n - r_n, \widehat{p}_n) \geq 1 - \alpha_c\}$.

Multiple-cohort:

- The predictive root is $U_n = \text{ppoibin}(Y_n, \boldsymbol{n} - \boldsymbol{r}_n, \hat{\boldsymbol{p}}_n)$.
- The calibrated nominal confidence level is $1 - \alpha_c = \inf\{u \in [0,1] : \Pr_{\widehat{\boldsymbol{\theta}}_n}(U_n \leq u) \geq 1 - \alpha\}$.
- The calibrated $1 - \alpha$ upper prediction bound is $\tilde{Y}^{PL}_{n,1-\alpha_c} = \inf\{y \in \{0\} \cup \mathbb{Z}^+; \text{ppoibin}(y, \boldsymbol{n} - \boldsymbol{r}_n, \hat{\boldsymbol{p}}_n) \geq 1 - \alpha_c\}$.

Extending the Direct/GPQ Bootstrap Method

Here we use the direct-bootstrap method as an example:

Single-cohort:

- The predictive distribution is $\tilde{G}_{Y_n}^{Direct}(y|\boldsymbol{d}_n) = \int \text{pbinom}(y, n - r_n, \widehat{p}_n^*) \Pr_*(d\widehat{p}_n^*) \approx \frac{1}{B}\sum_{b=1}^{B} \text{pbinom}(y, n - r_n, \widehat{p}_{n,b}^*)$.

Multiple-cohort:

- The predictive distribution is $\tilde{G}_{Y_n}^{Direct}(y|\boldsymbol{d}_n) = \int \text{ppoibin}(y, \boldsymbol{n}-\boldsymbol{r}_n, \widehat{\boldsymbol{p}}_n^*) \Pr_*(d\widehat{\boldsymbol{p}}_n^*) \approx \frac{1}{B}\sum_{b=1}^{B} \text{ppoibin}(y, \boldsymbol{n}-\boldsymbol{r}_n, \widehat{\boldsymbol{p}}_{n,b}^*)$.

Choice of a Distribution I

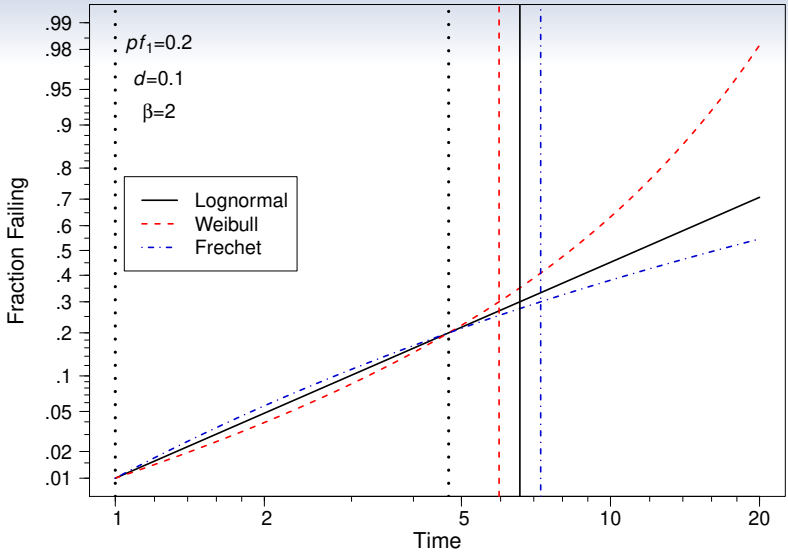
- When extrapolation is required, predictions can be strongly dependent on the distribution choice.
- It is best to choose a failure-time distribution based on knowledge of the failure mechanism and the related physics/chemistry of failure.

Examples

- The lognormal distribution is often used to model failure from crack initiation and growth due to cyclic stressing of metal components and chemical degradation like corrosion.
- The Weibull distribution may provide a suitable model for the time to first failure of a large number of similar components in a system.

Choice of a Distribution II

- Choosing a distribution based on failure-mechanism knowledge is not always possible. The alternative is to do sensitivity analysis using different distributions.
- When in doubt, the Weibull distribution is often used because it is a conservative choice.


pf_1=0.2
d=0.1
β=2
Lognormal
Weibull
Frechet
Fraction Failing
.99
.98
.95
.9
.8
.7
.6
.5
.4
.3
.2
.1
.05
.02
.01
1
2
5
10
20
Time

Future Research

We can extend current work by relaxing the i.i.d. assumption

- Units are exposed to different operating or environmental conditions, resulting in different time-to-event distributions.
- Covariates can be either constant or time-varying.
- Samples from different cohorts may be dependent.

Constructing Prediction Intervals Using the Likelihood Ratio Statistic

Motivations I

- [Cox, 1975] suggested to construct prediction intervals by inverting a hypothesis test.
- Let $\boldsymbol{X}_n$ be the data and Y be the predictand with densities $f(\boldsymbol{x}_n; \boldsymbol{\theta})$ and $g(y; \boldsymbol{\theta}^\dagger)$ respectively.
- Suppose we can form a hypothesis test for the null hypothesis $\boldsymbol{\theta} = \boldsymbol{\theta}^\dagger$ and ω_α is a critical region having size α.
- A $1 - \alpha$ prediction region for Y can be defined as

$$\mathcal{P}_{1-\alpha}(\boldsymbol{x}_n) = \{y : (\boldsymbol{x}_n, y) \notin w_\alpha\}.$$

Motivations II

Examples

$\boldsymbol{X}_n \sim \mathrm{N}(\mu_1, \sigma)$ and $Y \sim \mathrm{N}(\mu_2, \sigma)$, a t-test statistic for the null hypothesis $H_0 : \mu_1 = \mu_2$ is

$$t = \frac{\bar{X}_n - Y}{s\sqrt{(n+1)/n}} \sim t_{n-1},$$

where $s^2 = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2/(n-1)$ and t_{n-1} denotes a t-random variable with $n-1$ degrees of freedom. Then a $1-\alpha$ equal-tailed (i.e., equal probability of being outside either endpoint) prediction interval based on inverting the t-test is

$$\mathrm{PI}_{1-\alpha}(\boldsymbol{X}_n) = \left[\bar{X}_n - t_{n-1,\alpha/2}s\sqrt{(n+1)/n},\ \bar{X}_n + t_{n-1,\alpha/2}s\sqrt{(n+1)/n}\right],$$

where $t_{n-1,\alpha}$ denotes the α quantile of a t_{n-1} distribution.

Motivations III

- The construction of such tests often needs to be tailored to each problem.
- In many cases, there is no well-known or clear hypothesis test.
- The purpose of this paper is to propose a general prediction method based on inverting a type of LR test.
- The proposed method can be applied broadly to different settings.

The Likelihood Ratio Test I

- The traditional use of the LR test is to provide a general approach for comparing two nested models based on the observed data $\boldsymbol{X}_n = \boldsymbol{x}_n$.
- Let $\mathcal{L}_n(\boldsymbol{\theta}; \boldsymbol{x}_n)$ be the likelihood function, the LR for testing the null hypothesis $H_0 : \boldsymbol{\theta} \in \Theta_0$ is then

$$\Lambda_n = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} \mathcal{L}_n(\boldsymbol{\theta}; \boldsymbol{x}_n)}{\sup_{\boldsymbol{\theta} \in \Theta} \mathcal{L}_n(\boldsymbol{\theta}; \boldsymbol{x}_n)},$$

  where Θ is the parameter space for $\boldsymbol{\theta}$.
- If H_0 is true and Wilks' theorem applies,

$$-\log \Lambda_n \xrightarrow{d} \chi^2_d,$$

  as $n \to \infty$, and d is the difference in the lengths of Θ and Θ_0.

The Likelihood Ratio Test II

- We need to frame a LR statistic by construct a comparison involving full vs. reduced models for the joint distribution $(\boldsymbol{X}_n, Y)$.
- Let $\boldsymbol{X}_n \sim f(\cdot;\boldsymbol{\theta})$, and $Y \sim f(\cdot;\boldsymbol{\theta})$ is independent of $\boldsymbol{X}_n$, the original parameter space is $\boldsymbol{\theta} \in \Theta$.
- The main idea is to define a hypothesis test regarding an enlarged (and fictional) parameter space $\Theta_E \equiv \{(\boldsymbol{\theta}, \boldsymbol{\theta}_y)\}$, where $\boldsymbol{X}_n \sim f(\cdot;\boldsymbol{\theta})$ and $Y \sim f(\cdot;\boldsymbol{\theta}_y)$.
- $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_y$ are allowed to differ in exactly one pre-selected component when $(\boldsymbol{\theta}, \boldsymbol{\theta}_y) \in \Theta_E$.
- Full model: $\boldsymbol{X}_n \sim f(\cdot;\boldsymbol{\theta})$ and $Y \sim f(\cdot;\boldsymbol{\theta}_y)$ for $(\boldsymbol{\theta}, \boldsymbol{\theta}_y) \in \Theta_E$.
- Reduced model: $\boldsymbol{\theta}_y = \boldsymbol{\theta} \in \Theta$, where the parameter space of reduced model is nested within Θ_E with the constraint $\boldsymbol{\theta} = \boldsymbol{\theta}_y$.

The Likelihood Ratio Test III

The joint LR statistic for the test of $\boldsymbol{\theta} = \boldsymbol{\theta}_y$ based on $(\boldsymbol{X}_n, Y)$ is

$$\Lambda_n(\boldsymbol{X}_n, Y) = \frac{\sup_{\boldsymbol{\theta}=\boldsymbol{\theta}_y \in \Theta} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\theta}_y; \boldsymbol{X}_n, Y)}{\sup_{(\boldsymbol{\theta},\boldsymbol{\theta}_y)\in\Theta_E} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\theta}_y; \boldsymbol{X}_n, Y)}.$$

- The purpose of this LR test is NOT to conduct hypothesis tests of parameters.
- The goal is to construct a predictive root (i.e., a test statistic containing $\boldsymbol{X}_n$ and Y), which will be used to predict Y.

Determining the Distribution of the LR I

- Determine the distribution of $-2\log\Lambda_n(\boldsymbol{X}_n, Y)$ analytically.
- Let $\boldsymbol{X}_n \sim \text{Norm}(\theta, \sigma)$ and $Y \sim \text{Norm}(\theta, \sigma)$, where the location parameter is unknown but σ is known.
  Full model: $\boldsymbol{X}_n \sim \text{Norm}(\mu, \sigma)$ and $Y \sim \text{Norm}(\mu_y, \sigma)$, where $(\mu, \mu_y) \in \mathbb{R}^2$.
  Reduced model: $\mu = \mu_y$.
- The corresponding log-LR statistic for Y based on $\boldsymbol{X}_n$ is then

$$-2\log\Lambda_n(\boldsymbol{X}_n, Y) = \frac{n}{n+1}\left(\frac{Y - \bar{X}_n}{\sigma}\right)^2 \sim \chi^2_1.$$

Determining the Distribution of the LR II

- A $1-\alpha$ prediction region for Y given $\boldsymbol{X}_n = \boldsymbol{x}_n$ is

$$\begin{aligned}\mathcal{P}_{1-\alpha}(\boldsymbol{x}_n) &= \left\{y : -2\log\Lambda_n(\boldsymbol{x}_n, y) \leq \chi^2_{1,1-\alpha}\right\} \\ &= \left\{y : \bar{x}_n - z_{1-\alpha/2}\sigma\sqrt{\frac{n+1}{n}} \leq y \leq \bar{x}_n + z_{1-\alpha/2}\sigma\sqrt{\frac{n+1}{n}}\right\}\end{aligned}$$

Determining the Distribution of the LR III

- Use Wilks' theorem.
- Because the difference in the dimensions of the full and reduced models is $d = 1$, when Wilks' theorem applies, we have $-2 \log \Lambda_n(\boldsymbol{X}_n, Y) \xrightarrow{d} \chi^2_1$.
- A $1 - \alpha$ prediction region base on Wilk's theorem is

$$\mathcal{P}_{1-\alpha}(\boldsymbol{x}_n) = \left\{ y : -2 \log \Lambda_n(\boldsymbol{x}_n, y) \leq \chi^2_{1,1-\alpha} \right\}.$$

Determining the Distribution of the LR IV

- Use bootstrap.
- The idea is to use a parametric bootstrap re-creation of the data $(\boldsymbol{X}_n^*, Y^*)$ to approximate $\lambda_{1-\alpha}$ (the $1-\alpha$ quantile of$-2\log\lambda_n(\boldsymbol{X}_n, Y)$).
- The bootstrap approximate $\lambda_{1-\alpha}^*$ is used to construct the prediction region

$$\mathcal{P}_{1-\alpha}(\boldsymbol{x}_n) = \left\{y : -2\log\Lambda_n(\boldsymbol{x}_n, y) \leq \lambda_{1-\alpha}^*\right\}.$$

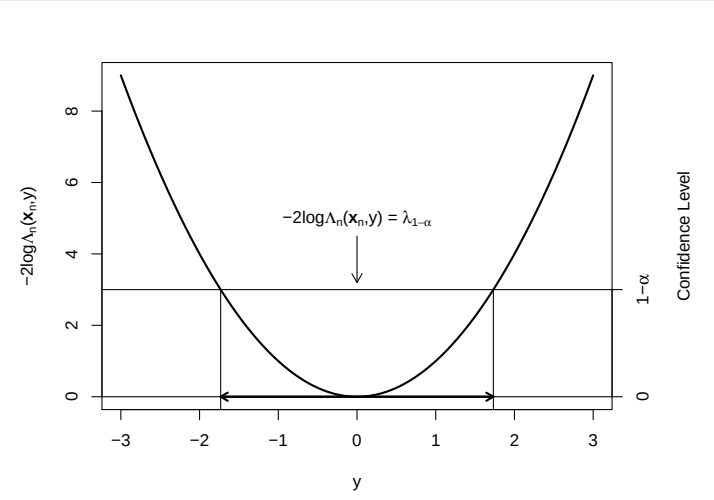
One-Sided Prediction Bounds I

- For most prediction problems, the LR statistic $\Lambda_n(\boldsymbol{X}_n, Y)$ is a unimodal function of y given $\boldsymbol{X}_n = \boldsymbol{x}_n$.
- Prediction regions defined in the form of

$$\mathcal{P}_{1-\alpha}(\boldsymbol{x}_n) = \left\{ y : -2 \log \Lambda_n(\boldsymbol{x}_n, y) \leq \lambda^*_{1-\alpha} \right\}$$

  becomes a two-sided prediction interval.

One-Sided Prediction Bounds II



One-Sided Prediction Bounds III

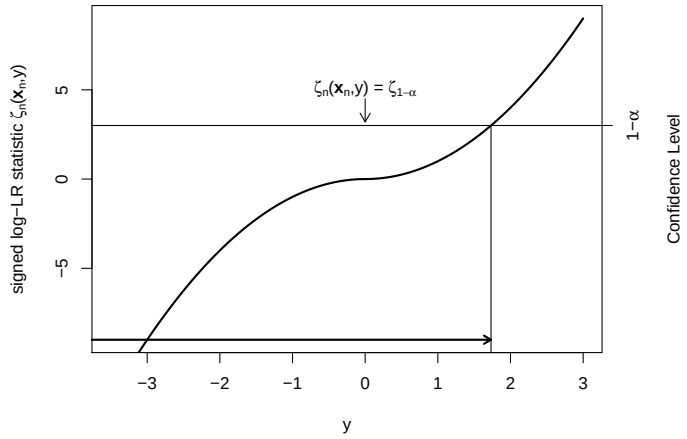
- In many applications, the cost of the predictand being greater than the upper bound is different than having it being less than the lower bound.
- We define a new statistic by modifying the LR statistic.
- For observed data $\boldsymbol{X}_n = \boldsymbol{x}_n$, let $y_0 \equiv y_0(\boldsymbol{x_n})$ be the value of y that maximizes $\Lambda_n(\boldsymbol{x}_n, y)$ so that $\Lambda_n(\boldsymbol{x}_n, y_0) = 1$.
- Define a signed LR statistic

$$\zeta_n(\boldsymbol{x}_n, y) \equiv (-1)^{\mathsf{I}(y \leq y_0)} \left[-2 \log \Lambda_n(\boldsymbol{x}_n, y)\right].$$

- Use bootstrap to approximate the $1 - \alpha$ quantile of $\zeta_n(\boldsymbol{X}_n, Y)$, denoted by $\zeta^*_{1-\alpha}$, the $1 - \alpha$ upper prediction bound

$$\tilde{y}_{1-\alpha}(\boldsymbol{x}_n) = \sup_{y \in \mathbb{R}} \left\{ y : \zeta_n(\boldsymbol{x}_n, y) \leq \zeta^*_{1-\alpha} \right\}.$$

One-Sided Prediction Bounds IV



Exact Results

The LR prediction method has exact coverage probability if $\boldsymbol{X}_n, Y \sim f(\cdot; \mu, \sigma)$ are from a location-scale distribution, including the uniform distribution which does not satisfy the regularity conditions.

Examples

Suppose $X_1, \ldots, X_n, Y$ are iid Unif$(0, \theta)$, the LR statistic is

$$\Lambda_n(\boldsymbol{x}_n, y) = \frac{(x_{(n)}/y)^n}{[\max(x_{(n)}/y, 1)]^{n+1}},$$

where $x_{(n)} \equiv \{x_1, \ldots, x_n\}$. Because $X_{(n)}/Y = \left(X_{(n)}/\theta\right) / \left(Y/\theta\right)$ is a pivotal quantity, $\Lambda_n(\boldsymbol{X}_n, Y)$ is also a pivotal quantity.

General Results

Under mild conditions,

- As $n \to \infty$,

$$-2 \log \Lambda_n(\boldsymbol{X}_n, Y) \xrightarrow{d} -2 \log \left[\frac{f(Y; \boldsymbol{\theta}_0)}{\sup_{\theta_y} f(Y; \theta_y, \boldsymbol{\theta}_0')} \right],$$

  where $\boldsymbol{\theta}_0 = (\theta_0, \boldsymbol{\theta}_0')$ denotes the true value of the parameter vector $\boldsymbol{\theta}$ and θ_y is where the reduced (or true) and full (or contrived) models differ.

- The bootstrap version of the LR (denoted by Λ_n^*) captures the distribution of the LR,

$$\sup_{\lambda \in \mathbb{R}} |\Pr_*(-2 \log \Lambda_n^* \leq \lambda) - \Pr(-2 \log \Lambda_n \leq \lambda)| \xrightarrow{p} 0,$$

  where $\Pr_*(\cdot)$ is the bootstrap induced probability and $\Lambda_n^* \equiv \Lambda_n(\boldsymbol{X}_n, Y^*)$.

Discrete Data

- Many methods that generally work in continuous settings are not applicable for certain discrete data models.
- The LR prediction method not only works but also has performance that is comparable to those especially tailored methods in binomial and Poisson distributions.
- The LR method also works for the within-sample prediction problem with good coverage probability.

Within-Sample Prediction I

- Let the probability of failing between $(t_c, t_w]$ conditional on surviving at t_c be p.
- Under the true model, the constraint on p is that

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

- We construct the LR by testing the null hypothesis

$$H_0 : p = \frac{F(t_w; \boldsymbol{\theta}) - F(t_c; \boldsymbol{\theta})}{1 - F(t_c; \boldsymbol{\theta})}$$

Within-Sample Prediction II

- The joint likelihood function is

$$\mathcal{L}(\boldsymbol{\theta}, p; \boldsymbol{t}_n, y) = \binom{n - r_n}{y} \prod_{i=1}^{r} f(t_{(i)}; \boldsymbol{\theta}) p^y (1-p)^{n-y-r_n}.$$

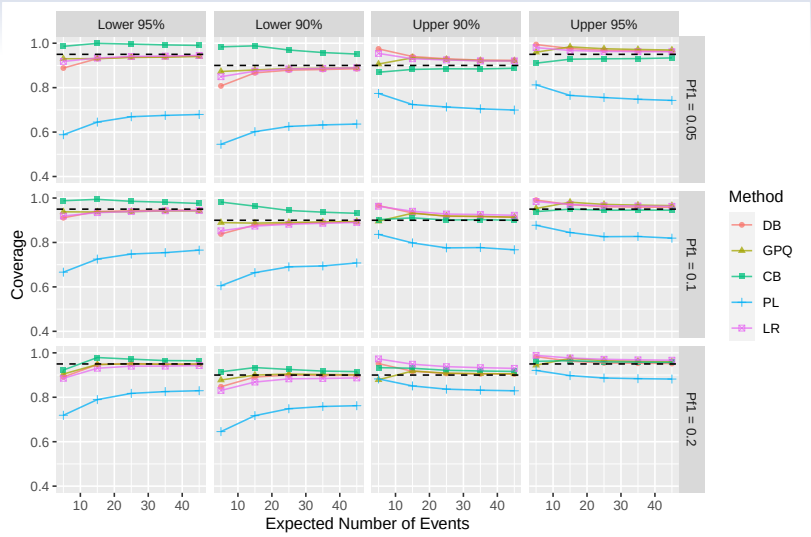
- The LR statistic is

$$\Lambda_n(\boldsymbol{t}_n, y) = \frac{\sup_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, p(\boldsymbol{\theta}); \boldsymbol{t}_n, y)}{\sup_{\boldsymbol{\theta}, p} \mathcal{L}(\boldsymbol{\theta}, p; \boldsymbol{t}_n, y)}.$$

- A $1 - \alpha$ prediction region is defined as

$$\left\{ y : -2 \log \Lambda_n(\boldsymbol{t}_n, y) \leq \chi^2_{1,1-\alpha} \right\}$$

Simulation Results



Future Research

- The construction of LR does not require the predictand to be univariate.
- LR method for dependent data.

The End

Cox, D. R. (1975).
Prediction intervals and empirical Bayes confidence intervals.
Journal of Applied Probability, 12:47–55.

Meeker, W. Q. and Escobar, L. A. (1998).
Statistical methods for reliability data.
Wiley.

Patel, J. (1989).
Prediction intervals-a review.
Communications in Statistics-Theory and Methods, 18:2393–2465.

Tian, Q., Meng, F., Nordman, D. J., and Meeker, W. Q. (2021).
Predicting the number of future events.
Journal of the American Statistical Association.
DOI:10.1080/01621459.2020.1850461.