MMR'2000

Bordeaux, July 2000

Exact Statistical Inference for Nonhomogenous Poisson Processes

Bo Lindqvist

Department of Mathematical Sciences,
Norwegian University of Science and Technology,

Trondheim

bo@math.ntnu.no

http://www.math.ntnu.no/~bo/

EXAMPLE

RELIABILITY GROWTH PROGRAM

$n = 10$ **failures**

QUESTION: RELIABILITY GROWTH?

- **By eye: YES (?)**

- **By statistical modeling and significance testing?**

TOPIC OF THIS TALK

- Exact goodness-of-fit testing for parametric non-homogeneous Poisson process models (NHPP)

- Exact statistical inference in parametric NHPP models

MOTIVATION

- Events occuring in time, e.g. failures of a repairable system, are commonly modeled by NHPP

- Goodness-of-fit tests based on asymptotic distributions require large number n of observations.

- Exact inference is well developed for power law ("Weibull") process. Need for considering exact procedures valid for other models (e.g. log linear).

OUTLINE

- Statistical modeling by NHPP, in particular power law and log linear law

- Exact goodness-of-fit testing for parametric NHPP models, using
  - conditional testing given sufficient statistic
  - Monte Carlo simulation of conditional distributions

- Exact confidence interval for growth parameter

GENERAL MODEL

- System observed from time $t = 0$, until n failures have occurred (right failure truncation).
- Failure times $T = (T_1, \ldots, T_n)$
- Model: NHPP with intensity $\lambda(t)$

(LOG) LIKELIHOOD FUNCTION FOR NHPP($\lambda(t)$)

$$L = \left\{ \prod_{i=1}^{n} \lambda(t_i) \right\} e^{-\Lambda(t_n)}$$

$$l = \sum_{i=1}^{n} \log \lambda(t_i) - \Lambda(t_n)$$

where

$$\Lambda(t) = \int_0^t \lambda(u) du$$

PARAMETRIC NHPP MODELS

POWER LAW

$$\lambda(t) = abt^{b-1}$$

$$\Lambda(t) = at^b$$

$$l = n \log a + n \log b + (b-1) \sum_{i=1}^{n} \log t_i - at_n^b$$

- **Sufficient statistic:**

$$(\log T_n, \sum_{i=1}^{n-1} \log T_i)$$

LOG LINEAR LAW

$$\lambda(t) = e^{a+bt}$$

$$\Lambda(t) = e^a(e^{bt} - 1)/b$$

$$l = na + b \sum_{i=1}^{n} t_i - e^a(e^{bt_n} - 1)/b$$

- **Sufficient statistic:**

$$(T_n, \sum_{i=1}^{n-1} T_i)$$

EXAMPLE OF GOODNESS-OF-FIT TESTING

Basic property: If $T_1, \ldots, T_n$ is NHPP($\lambda(\cdot)$), then

$\Lambda(T_1), \ldots, \Lambda(T_n)$ is HPP(1)

Thus

$$V_j = \frac{\Lambda(T_j)}{\Lambda(T_n)};\ j = 1, \ldots, n-1$$

are distributed as the ordering of n-1 i.i.d. uniforms on $(0,1)$.

In practice $\Lambda(\cdot)$ are functions of a, b. Substituting maximum likelihood estimates a^*, b^* we define

$$V_j^* = \frac{\Lambda^*(T_j)}{\Lambda^*(T_n)};\ j = 1, \ldots, n-1$$

- Power law: $\Lambda(t) = at^b$, so

$$V_j^* = (T_j/T_n)^{b^*}$$

- Log linear: $\Lambda(t) = e^a(e^{bt} - 1)/b$ so

$$V_j^* = \frac{e^{b^*T_j} - 1}{e^{b^*T_n} - 1}$$

- The set $(V_1^*, \ldots, V_{n-1}^*)$ is expected to behave "like" uniforms $(0,1)$ IF MODEL IS CORRECT.

GOODNESS-OF-FIT STATISTICS

Baker (TECH 96):

$$\boldsymbol{W = \sum_{i=1}^{n} (V_j^* - V_{j-1}^*)^2}$$

where $V_0^* = 0$, $V_n^* = 1$.

- **W too large: too much "clumping" of V_j^***
- **W too small: too "regularly" distributed V_j^***

NOTE: A goodness-of-fit statistic W need not be a function of the V_j^*, but may in principle be any function of observations $T_1, \ldots, T_n$

CONDITIONAL TESTING

- H_0 : $T_1, \ldots, T_n$ are from particular NHPP (e.g. power law/log linear law)

- Data $T = (T_1, \ldots, T_n)$

- Test statistic $W = W(T)$

- Sufficient statistic

$$S = ((\log)T_n, \sum_{i=1}^{n-1} (\log)T_j)$$

  (Power law: Use "log", Log linear: No "log")

- By definition of sufficiency:

  Conditional distribution of T given $S = s$ does not depend on unknown parameters

TEST:

If $W = w_o, S = s_o$ are observed:

- Reject H_0 if $P(W \leq w_o | S = s_o) \leq \alpha/2$

- or if $P(W \geq w_o | S = s_o) \leq \alpha/2$.

- Unconditionally this gives an exact α-level test.

COMPUTATION OF CONDITIONAL DISTRIBUTION

How to compute conditional distribution of $T = (T_1, \ldots, T_n)$ given $S = s$?

- **In simple cases: This distribution is the same for all values of parameters, so choose parameters which give some simple structure**

  Power law: $a = b = 1$ gives an HPP(1)

  Log linear law: $a = b = 0$ gives an HPP(1)

- **General method:**

  Monte Carlo conditioning on a sufficient statistic (Engen & Lillegård BIOMA 1997, Lindqvist & Taraldsen 2000)

MONTE CARLO CONDITIONING ON A SUFFICIENT STATISTIC

- **Observation:** $T = (T_1, \ldots, T_n)$
- **Sufficient statistic:** $S = S(T)$
- **Model:** $T \sim P_\theta;\ \theta \in \Omega$
- **Want to compute**

  $$P(W(T) \geq w_o | S = s_o)$$

  or more generally

  $$E[\phi(T) | S = s]$$

  for given functions ϕ.

Basic assumption of method:

- T **under** P_θ **can be simulated as**

  $$T = \chi(U, \theta)$$

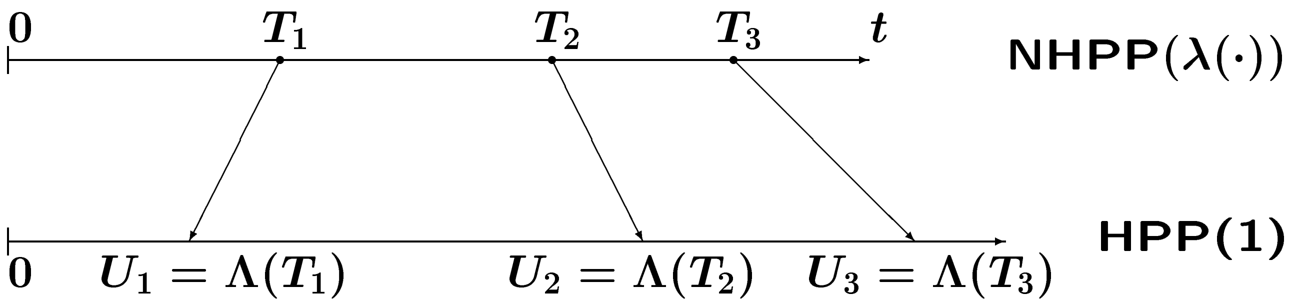
  for a random vector U with known distribution.

- S **can be simulated by**

  $$S = S(\chi(U, \theta)) \equiv \tau(U, \theta)$$

APPLICATION TO NHPP

SIMULATION OF NHPP



Simulate $T = (T_1, \ldots, T_n)$ by first simulating HPP(1)

$$U = (U_1, \ldots, U_n)$$

Define

$$T_i = \Lambda^{-1}(U_i)$$

POWER LAW

- $\Lambda(t) = at^b$
- $\Lambda^{-1}(u) = (u/a)^{1/b}$

$$T = \chi(U; a, b) = \left(\left(\frac{U_1}{a}\right)^{1/b}, \ldots, \left(\frac{U_n}{a}\right)^{1/b} \right)$$

$$S = \tau(U; a, b) = \left(\frac{1}{b} \log\left(\frac{U_n}{a}\right), \frac{1}{b} \sum_{i=1}^{n-1} \log\left(\frac{U_j}{a}\right) \right)$$

Can we simulate samples from conditional distribution of

$$T = (T_1, \ldots, T_n)$$

given $S = s$ where

$$S = (\log T_n, \sum_{i=1}^{n-1} \log T_j)$$

Idea:

(1) Draw $U = (U_1, \ldots, U_n)$ as HPP(1)

(2) Adjust $\hat{a}, \hat{b}$ so that

$\tau(U; \hat{a}, \hat{b}) = s_o$ (observed value of S)

(3) Use $\chi(U; \hat{a}, \hat{b})$ as sample of T given $S = s$.

Thus

$$\begin{aligned}
E[\phi(T)|S=s] &= E_U[\phi(\chi(U;\hat{a},\hat{b}))] \\
&\approx \frac{1}{m}\sum_{k=1}^{m}\phi(\chi(U^{(k)};\hat{a}^{(k)},\hat{b}^{(k)}))
\end{aligned}$$

Recall that $\hat{a},\hat{b}$ depend on U, s, and are obtained by solving

$$\tau(U;\hat{a},\hat{b}) = s$$

Warning:

Unfortunately such adjustment does not always give the correct conditional distribution

THEOREM 1 (Lindqvist and Taraldsen, 2000)

- Parameter: $\theta \in \Omega \subseteq R^k$
- Observation: $T = \chi(U, \theta)$
- Sufficient: $S = \tau(U, \theta) \in R^k$
- Wanted: $E[\phi(T)|S = s]$
- Assumption: Equation $\tau(u, \theta) = s$ has unique solution $\theta = \hat{\theta}(u, s)$
- Choose: Nonnegative function $f(\theta)$

Then $E(\phi(T)|S = s)$

$$= \frac{E_U\left[\phi(\chi(U, \hat{\theta}(U, s)))f(\hat{\theta}(U, s))|\det \partial_s \hat{\theta}(U, s)|\right]}{E_U\left[f(\hat{\theta}(U, s))|\det \partial_s \hat{\theta}(U, s)|\right]}$$

Note:

$$|\det \partial_s \hat{\theta}(U, s)| = 1/|\det \partial_\theta \tau(U, \theta)|_{\theta = \hat{\theta}(U,s)}$$

Formula SIMPLIFIES if there exists $f(\cdot)$ with

$$f(\hat{\theta}(U, s)) \propto |\det \partial_\theta \tau(U, \theta)|_{\theta = \hat{\theta}(U,s)}$$

This holds for power law, but NOT for log linear law.

THEOREM 2 (Lindqvist and Taraldsen, 2000)

- $\tau(u,\theta) \equiv \tilde{\tau}(r(u),\theta)$
- $\tilde{\tau}(r(u),\theta) = s$ has unique solution $r(u) = \tilde{v}(\theta, s)$

Then:

"Adjustment" algorithm is valid, i.e.

- $\chi(U, \hat{\theta}(U,s))$ is distributed as the conditional distribution of T given $S = s$.

Note: $\tilde{v}(\theta, S)$ $(= r(U))$ is a pivotal quantity in the classical sense.

POWER LAW:

$$\tau(U; a, b) = \left(\frac{1}{b}\log\left(\frac{U_n}{a}\right), \frac{1}{b}\sum_{i=1}^{n-1}\log\left(\frac{U_j}{a}\right)\right)$$

Assumptions hold with

$$r(U) = (\log U_n, \sum_{j=1}^{n-1}\log U_j)$$

LOG LINEAR LAW:

Assumptions of Theorem 2 do NOT hold

LOG LINEAR MODEL

- $\lambda(t) = e^{a+bt}$
- $\Lambda(t) = e^{a}(e^{bt} - 1)/b$
- $\Lambda^{-1}(u) = (1/b)\log(1 + be^{-a}u)$
- $S = (T_n, \Sigma_{i=1}^{n-1} T_j)$

so

$$\chi(U; a, b) = \Big(\log(1 + be^{-a}U_1)/b, \ldots, \log(1 + be^{-a}U_n)/b\Big)$$

$$\tau(U; a, b) = \left(\log(1 + be^{-a}U_n)/b,\ \sum_{j=1}^{n-1} \log(1 + be^{-a}U_j)/b\right)$$

- NOT possible to find "$r(U)$"
- But equation $\tau(u; \hat{a}, \hat{b}) = (s_1, s_2)$ can be solved uniquely for a, b. Thus Theorem 1 applies.
- Need determinant

$$\det \partial_{a,b}\tau(u; a, b)$$

EXAMPLE – RELIABILITY GROWTH PROGRAM

$n = 10$ failures

103, 315, 801, 1183, 1345, 2957, 3909, 5702, 7261, 8245.

GOODNESS-OF-FIT TESTING:

$$W = \sum_{i=1}^{n} (V_j^* - V_{j-1}^*)^2$$

- Power law: $W = 0.1263$
- Log linear; $W = 0.1466$

Log linear law (Theorem 1):

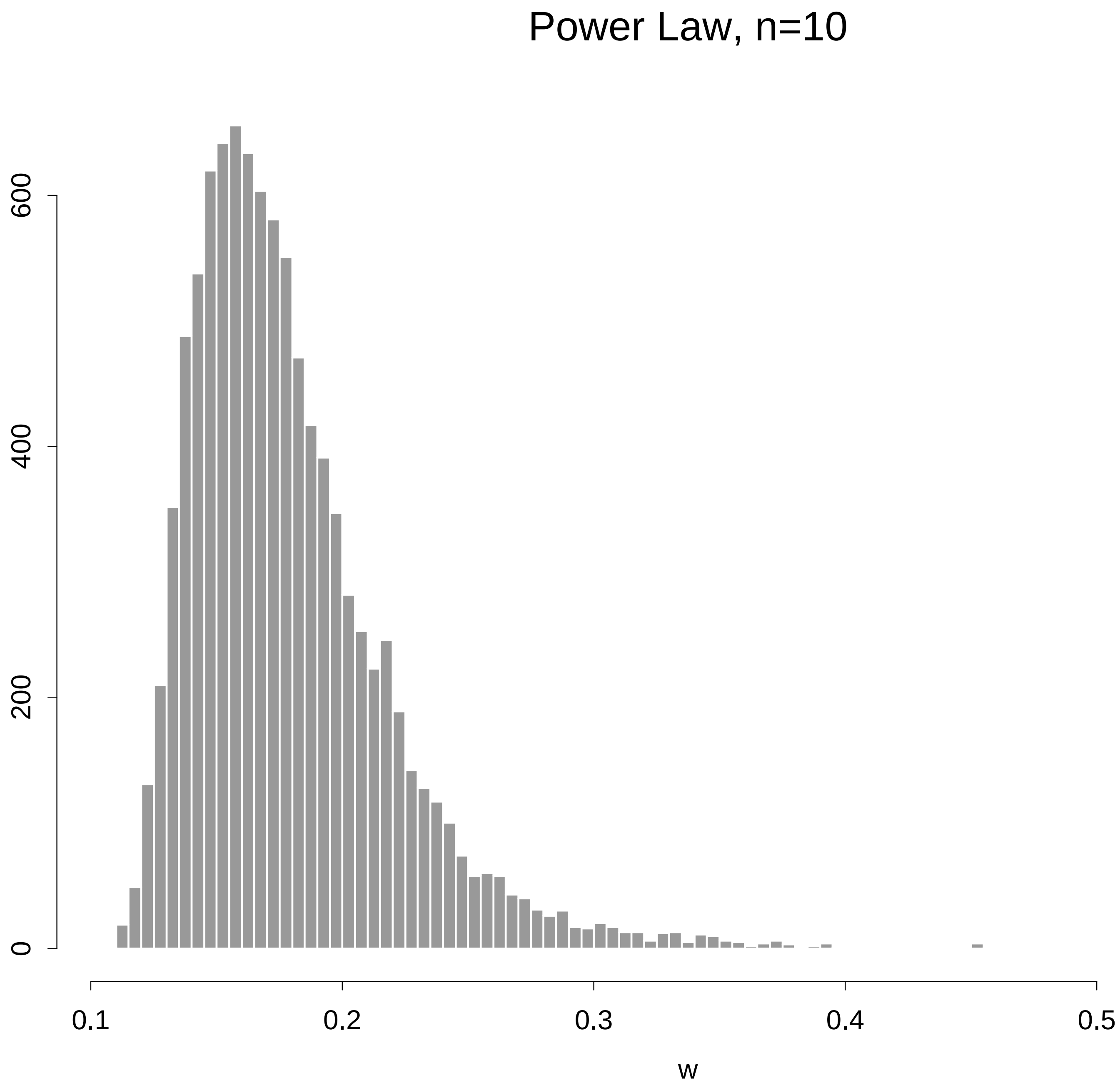
$P(W \leq 0.1466 | S = (8245, 23576)) = 0.217$

Power law:

W is itself a pivotal quantity and hence independent of S under H_0.

By simulation:

$P(W \leq 0.1263) = 0.024$


Power Law, n=10
0
200
400
600
0.1
0.2
0.3
0.4
0.5
w

CONFIDENCE INTERVALS FOR b

- Power law:

  Solution is well known

  $$\frac{2nb^*}{b} \sim \chi^2_{2(n-1)}$$

  $b^* = 0.6249$

  $$\Downarrow$$

  Exact 90 % confidence interval for b

  (0.2933,0.9020)

  (does not cover 1, so conclude $b < 1$)

- Log linear law:

  $b^* = -1.75 \cdot 10^{-4}$

  No pivotal quantity involving b^*

EXACT CONFIDENCE INTERVALS

Lillegård and Engen (J APP STAT 99):

(1) Draw a (large) number of independent realizations $u_1, \ldots, u_m$ of U.

(2) Solve $\tau(u_j; a, b) = s_o$, get $\hat{a}_j, \hat{b}_j$

(3) Order $\hat{b}_{(1)} < \cdots < \hat{b}_{(m)}$

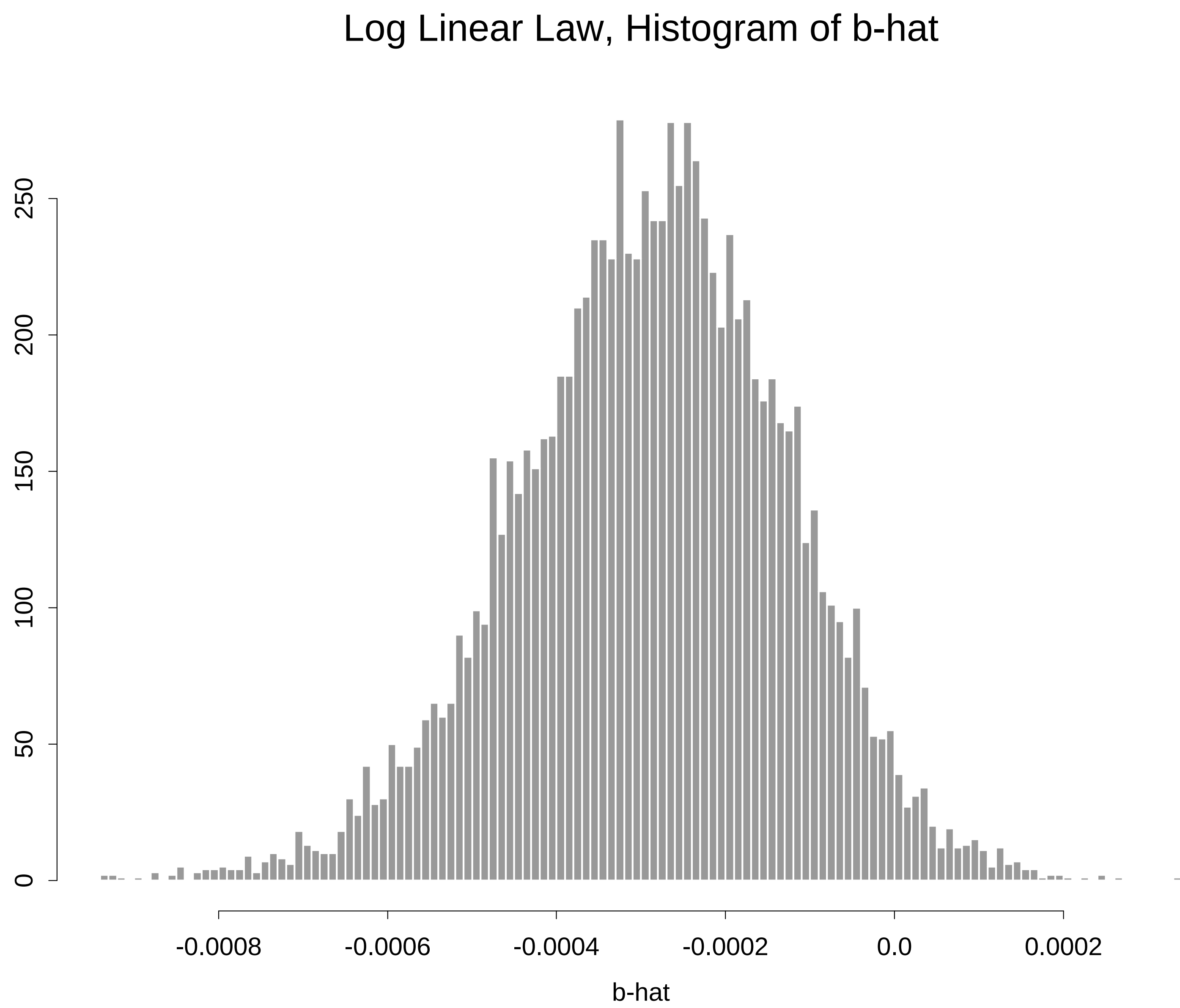
(4) Then $(\hat{b}_{(k)}, \hat{b}_{(m-k+1)})$ is an exact $1 - 2k/(m+1)$ confidence interval for b.

LOG LINEAR LAW:

90 % interval for b:

$$(-5.60 \cdot 10^{-4}, -3.83 \cdot 10^{-5})$$

Does not contain 0, so reliability GROWTH


Log Linear Law, Histogram of b-hat
0
50
100
150
200
250
-0.0008
-0.0006
-0.0004
-0.0002
0.0
0.0002
b-hat

CONCLUSIONS

- Parametric NHPP models
  - Exact goodness-of-fit testing
  - Exact inference for parameters

- Need:
  - Sufficiency reduction
  - Workable functions $\chi(U, \theta) = (\Lambda^{-1}(U_j))$
  - Efficient solving of the equation $\tau(U, \theta) = s$ (non-unique solutions can be handled)

- Power law case: Leads to known methods based on pivotal quantities

- Log liner law: Leads to exact procedures without using pivotal quantities

- Possible extensions:
  - Time truncation
  - Several identical systems
  - More general processes, e.g. modulated power law