The Breslow Estimator, the Breslow-Crowley Theory, and Their Modern Impacts

A Personal History

Jon A. Wellner [1]

[1] Departments of Statistics and Biostatistics
University of Washington

August 4 / In honor of Norman E. Breslow

Outline

1. Introduction: the facts
   - The Breslow estimator
   - The Breslow - Crowley theory
2. The Impacts, Part I
   - Citations
   - Developments
3. The Impacts, Part II

Outline

Outline

The Breslow estimator.
of the baseline hazard, Cox model

- $X = (\Delta, T, Z)$,
  $T = \min\{\tilde{T}, C\}$, $\Delta = 1\{\tilde{T} \leq C\}$, $Z =$ vector of covariates.
- $\theta =$ finite-dimensional parameter;
- three infinite - dimensional parameters;
  $\Lambda = \int_0^{\cdot} \lambda(s)ds =$ baseline hazard function,
  $G(t|Z) = P(C \leq t|Z) =$ conditional distribution of censoring time given covariates Z
  G_Z, the marginal distribution of the covariate vector Z.
- $\Lambda(t|Z) = e^{\theta' Z}\Lambda(t) \equiv$ "the Cox model"

The Breslow estimator.
of the baseline hazard, Cox model

- $X = (\Delta, T, Z)$,
  $T = \min\{\tilde{T}, C\}$, $\Delta = 1\{\tilde{T} \le C\}$, $Z =$ vector of covariates.
- $\theta =$ finite-dimensional parameter;
- three infinite - dimensional parameters;
  $\Lambda = \int_0^{\cdot} \lambda(s) ds =$ baseline hazard function,
  $G(t|Z) = P(C \le t|Z) =$ conditional distribution of censoring time given covariates Z
  G_Z, the marginal distribution of the covariate vector Z.
- $\Lambda(t|Z) = e^{\theta' Z} \Lambda(t) \equiv$ "the Cox model"

The Breslow estimator.
of the baseline hazard, Cox model

- $X = (\Delta, T, Z)$,
  $T = \min\{\tilde{T}, C\}$, $\Delta = 1\{\tilde{T} \le C\}$, $Z =$ vector of covariates.
- $\theta =$ finite-dimensional parameter;
- three infinite - dimensional parameters;
  $\Lambda = \int_0^{\cdot} \lambda(s)ds =$ baseline hazard function,
  $G(t|Z) = P(C \le t|Z) =$ conditional distribution of censoring time given covariates Z
  G_Z, the marginal distribution of the covariate vector Z.
- $\Lambda(t|Z) = e^{\theta' Z}\Lambda(t) \equiv$ "the Cox model"

The Breslow estimator.
of the baseline hazard, Cox model

- $X = (\Delta, T, Z)$,
  $T = \min\{\tilde{T}, C\}$, $\Delta = 1\{\tilde{T} \le C\}$, $Z =$ vector of covariates.
- $\theta =$ finite-dimensional parameter;
- three infinite - dimensional parameters;
  $\Lambda = \int_0^{\cdot} \lambda(s)ds =$ baseline hazard function,
  $G(t|Z) = P(C \le t|Z) =$ conditional distribution of censoring time given covariates Z
  G_Z, the marginal distribution of the covariate vector Z.
- $\Lambda(t|Z) = e^{\theta' Z}\Lambda(t) \equiv$ "the Cox model"

David R. Cox (1975)

Sir David R. Cox (March, 2006)

- $N(t) = \Delta 1_{[T \le t]}$ = event counting process;
  $Y(t) = 1_{[T \ge t]}$ = at risk process.
- Density of $X = (\Delta, T, Z)$ at $x = (\delta, t, z)$ is given by

$$\exp(-e^{\theta' z}\Lambda(t))[e^{\theta' z}\lambda(t)(1 - G(t|z))]^{\delta}[g(t|z)]^{1-\delta} g_Z(z).$$

- Suppose $X_1, \ldots, X_n$ i.i.d. as $X = (\Delta, T, Z)$.
-

$$\overline{N}_n(t) = \mathbb{P}_n \Delta 1\{T \le t\} = n^{-1}\sum_{i=1}^{n} N_i(t),$$

$$\widehat{S}_n(t; \theta) = \mathbb{P}_n e^{\theta' Z} Y(t) = n^{-1}\sum_{i=1}^{n} e^{\theta' Z_i} Y_i(t).$$

- $N(t) = \Delta 1_{[T \le t]}$ = event counting process;
  $Y(t) = 1_{[T \ge t]}$ = at risk process.
- Density of $X = (\Delta, T, Z)$ at $x = (\delta, t, z)$ is given by

$$\exp(-e^{\theta' z}\Lambda(t))[e^{\theta' z}\lambda(t)(1 - G(t|z))]^{\delta}[g(t|z)]^{1-\delta} g_Z(z).$$

- Suppose $X_1, \ldots, X_n$ i.i.d. as $X = (\Delta, T, Z)$.
-

$$\overline{N}_n(t) = \mathbb{P}_n \Delta 1\{T \le t\} = n^{-1}\sum_{i=1}^{n} N_i(t),$$

$$\widehat{S}_n(t;\theta) = \mathbb{P}_n e^{\theta' Z} Y(t) = n^{-1}\sum_{i=1}^{n} e^{\theta' Z_i} Y_i(t).$$

- $N(t) = \Delta 1_{[T \le t]}$ = event counting process;
  $Y(t) = 1_{[T \ge t]}$ = at risk process.
- Density of $X = (\Delta, T, Z)$ at $x = (\delta, t, z)$ is given by

$$\exp(-e^{\theta' z}\Lambda(t))[e^{\theta' z}\lambda(t)(1 - G(t|z))]^{\delta}[g(t|z)]^{1-\delta} g_Z(z).$$

- Suppose $X_1, \ldots, X_n$ i.i.d. as $X = (\Delta, T, Z)$.
-

$$\overline{N}_n(t) = \mathbb{P}_n \Delta 1\{T \le t\} = n^{-1}\sum_{i=1}^{n} N_i(t),$$

$$\widehat{S}_n(t;\theta) = \mathbb{P}_n e^{\theta' Z} Y(t) = n^{-1}\sum_{i=1}^{n} e^{\theta' Z_i} Y_i(t).$$

- $N(t) = \Delta 1_{[T \le t]} =$ event counting process;
  $Y(t) = 1_{[T \ge t]} =$ at risk process.
- Density of $X = (\Delta, T, Z)$ at $x = (\delta, t, z)$ is given by

$$\exp(-e^{\theta' z}\Lambda(t))[e^{\theta' z}\lambda(t)(1 - G(t|z))]^{\delta}[g(t|z)]^{1-\delta} g_Z(z).$$

- Suppose $X_1, \ldots, X_n$ i.i.d. as $X = (\Delta, T, Z)$.
-

$$\overline{N}_n(t) = \mathbb{P}_n \Delta 1\{T \le t\} = n^{-1}\sum_{i=1}^{n} N_i(t),$$

$$\widehat{S}_n(t; \theta) = \mathbb{P}_n e^{\theta' Z} Y(t) = n^{-1}\sum_{i=1}^{n} e^{\theta' Z_i} Y_i(t).$$

The Breslow estimator

- The Breslow estimator of $\Lambda =$ baseline hazard function is:

$$\widehat{\Lambda}_n(t) = \int_0^t \frac{1}{S_n^0(s;\widehat{\theta})} d\overline{N}_n(s).$$

- where $\widehat{\theta} =$ Cox partial likelikhood estimator of θ.
- 38 lines on pages 216-217; less than a full page.

The Breslow estimator

- The Breslow estimator of $\Lambda =$ baseline hazard function is:

$$\widehat{\Lambda}_n(t) = \int_0^t \frac{1}{S_n^0(s;\widehat{\theta})} d\overline{N}_n(s).$$

- where $\widehat{\theta} =$ Cox partial likelikhood estimator of θ.
- 38 lines on pages 216-217; less than a full page.

The Breslow estimator

- The Breslow estimator of Λ = baseline hazard function is:

$$\widehat{\Lambda}_n(t) = \int_0^t \frac{1}{S_n^0(s; \widehat{\theta})} d\overline{N}_n(s).$$

- where $\widehat{\theta}$ = Cox partial likelikhood estimator of θ.
- 38 lines on pages 216-217; less than a full page.

Breslow - Crowley theory

limit theory for the Kaplan - Meier estimator

Kaplan-Meier estimator

- With no covariates, let

$$\begin{aligned}\overline{Y}_n(t) &= \mathbb{P}_n Y(t) = \mathbb{P}_n 1\{T \ge t\} \to_{a.s.} P1\{T \ge t\} \\ &= (1 - F(t-))(1 - G(t-)) \equiv 1 - H(t-).\end{aligned}$$

- $\widehat{\Lambda}_n(t) = \int_0^t \frac{1}{\overline{Y}_n(s)} d\overline{N}_n(s)$ estimates $\Lambda(t) = \int_0^t \frac{1}{1-F(s-)} dF(s)$.
- The survival function $1 - F$ is given in terms of Λ by

$$\begin{aligned}1 - F(t) &= \exp(-\Lambda_c(t)) \prod_{s \le t} (1 - \Delta\Lambda(s)) \\ &= \begin{cases} \exp(-\Lambda(t)), & \Lambda \ \text{continuous} \\ \prod_{s \le t}(1 - \Delta\Lambda(s)), & \Lambda \ \text{discrete.}\end{cases}\end{aligned}$$

Breslow - Crowley theory

limit theory for the Kaplan - Meier estimator

Kaplan-Meier estimator

- With no covariates, let

$$\begin{aligned}\overline{Y}_n(t) &= \mathbb{P}_n Y(t) = \mathbb{P}_n 1\{T \ge t\} \rightarrow_{a.s.} P1\{T \ge t\} \\ &= (1 - F(t-))(1 - G(t-)) \equiv 1 - H(t-).\end{aligned}$$

- $\widehat{\Lambda}_n(t) = \int_0^t \frac{1}{\overline{Y}_n(s)} d\overline{N}_n(s)$ estimates $\Lambda(t) = \int_0^t \frac{1}{1-F(s-)} dF(s)$.
- The survival function $1 - F$ is given in terms of Λ by

$$\begin{aligned}1 - F(t) &= \exp(-\Lambda_c(t)) \prod_{s \le t}(1 - \Delta\Lambda(s)) \\ &= \begin{cases} \exp(-\Lambda(t)), & \Lambda \text{ continuous} \\ \prod_{s \le t}(1 - \Delta\Lambda(s)), & \Lambda \text{ discrete.} \end{cases}\end{aligned}$$

Breslow - Crowley theory

limit theory for the Kaplan - Meier estimator

Kaplan-Meier estimator

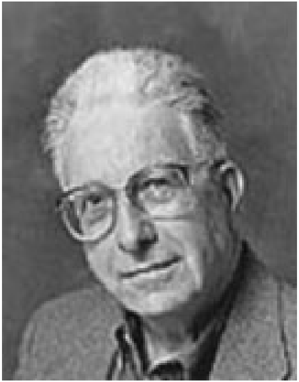
- With no covariates, let

$$\begin{aligned} \overline{Y}_n(t) &= \mathbb{P}_n Y(t) = \mathbb{P}_n 1\{T \ge t\} \rightarrow_{a.s.} P1\{T \ge t\} \\ &= (1 - F(t-))(1 - G(t-)) \equiv 1 - H(t-). \end{aligned}$$

- $\widehat{\Lambda}_n(t) = \int_0^t \frac{1}{\overline{Y}_n(s)} d\overline{N}_n(s)$ estimates $\Lambda(t) = \int_0^t \frac{1}{1-F(s-)} dF(s)$.
- The survival function $1 - F$ is given in terms of Λ by

$$\begin{aligned} 1 - F(t) &= \exp(-\Lambda_c(t)) \prod_{s \le t} (1 - \Delta\Lambda(s)) \\ &= \begin{cases} \exp(-\Lambda(t)), & \Lambda \text{ continuous} \\ \prod_{s \le t}(1 - \Delta\Lambda(s)), & \Lambda \text{ discrete.} \end{cases} \end{aligned}$$

Paul Meier photo



Paul Meier

- Kaplan - Meier (1958) estimator $\widehat{F}_n$ of F is given by

$$1 - \widehat{F}_n(t) = \prod_{s \le t} (1 - \Delta \widehat{\Lambda}_n(s)) = \prod_{s \le t} \left(1 - \frac{\Delta \overline{N}_n(s)}{\overline{Y}_n(s)} \right) .$$

- Breslow & Crowley (1974) proved:

Theorem

For each $\tau < \inf\{t : H(t) = 1\}$

$$\sqrt{n}(\widehat{F}_n - F) \Rightarrow (1 - F)\mathbb{B}(C) \qquad in \;\; D[0, \tau]$$

where $\mathbb{B}$ is standard Brownian motion on $[0, \infty)$ and

$$C(t) = \int_0^t \frac{1}{(1 - H)} d\Lambda .$$

- Kaplan - Meier (1958) estimator $\widehat{F}_n$ of F is given by

$$1 - \widehat{F}_n(t) = \prod_{s \le t}(1 - \Delta\widehat{\Lambda}_n(s)) = \prod_{s \le t}\left(1 - \frac{\Delta \overline{N}_n(s)}{\overline{Y}_n(s)}\right).$$

- Breslow & Crowley (1974) proved:

Theorem

For each $\tau < \inf\{t : \ H(t) = 1\}$

$$\sqrt{n}(\widehat{F}_n - F) \Rightarrow (1 - F)\mathbb{B}(C) \qquad in \ \ D[0, \tau]$$

where $\mathbb{B}$ *is standard Brownian motion on* $[0, \infty)$ *and*

$$C(t) = \int_0^t \frac{1}{(1 - H)} d\Lambda.$$

- Kaplan - Meier (1958) estimator $\widehat{F}_n$ of F is given by

$$1 - \widehat{F}_n(t) = \prod_{s \le t} (1 - \Delta \widehat{\Lambda}_n(s)) = \prod_{s \le t} \left(1 - \frac{\Delta \overline{N}_n(s)}{\overline{Y}_n(s)} \right) .$$

- Breslow & Crowley (1974) proved:

Theorem

For each $\tau < \inf\{t : \; H(t) = 1\}$

$$\sqrt{n}(\widehat{F}_n - F) \Rightarrow (1 - F)\mathbb{B}(C) \qquad \text{in} \;\; D[0, \tau]$$

where $\mathbb{B}$ *is standard Brownian motion on* $[0, \infty)$ *and*

$$C(t) = \int_0^t \frac{1}{(1 - H)} d\Lambda.$$

- Cites of 1972 Breslow's discussion of Cox: 143
- Cites of 1974 Breslow - Crowley: 354
- Context:
  - Cites of Cox, 1972: > 21183
    (ISI Web of Science, 24 July 2006)
  - Cites of Kaplan - Meier: > 31256
    (ISI Web of Science, 24 July 2006).

- Cites of 1972 Breslow's discussion of Cox: 143
- Cites of 1974 Breslow - Crowley: 354
- Context:
  - Cites of Cox, 1972: > 21183
    (ISI Web of Science, 24 July 2006)
  - Cites of Kaplan - Meier: > 31256
    (ISI Web of Science, 24 July 2006).

- Cites of 1972 Breslow's discussion of Cox: 143
- Cites of 1974 Breslow - Crowley: 354
- Context:
  - Cites of Cox, 1972: > 21183
    (ISI Web of Science, 24 July 2006)
  - Cites of Kaplan - Meier: > 31256
    (ISI Web of Science, 24 July 2006).

- Cites of 1972 Breslow's discussion of Cox: 143
- Cites of 1974 Breslow - Crowley: 354
- Context:
  - Cites of Cox, 1972: > 21183
    (ISI Web of Science, 24 July 2006)
  - Cites of Kaplan - Meier: > 31256
    (ISI Web of Science, 24 July 2006).

- Cites of 1972 Breslow's discussion of Cox: 143
- Cites of 1974 Breslow - Crowley: 354
- Context:
  - Cites of Cox, 1972: > 21183
    (ISI Web of Science, 24 July 2006)
  - Cites of Kaplan - Meier: > 31256
    (ISI Web of Science, 24 July 2006).

Impacts & consequences

- Martingale theory:
  Le Cam → Aalen → Gill → Anderson & Gill ...
- Martingale approach to survival analysis: two books, Andersen, Borgan, Gill, and Keiding; Fleming and Harrington
- Dutch school of semiparametrics
  Groeneboom → Gill → van Zwet, Klaassen, van der Vaart
- Compact differentiability:
  Reeds → Groeneboom → Gill →
- Empirical process approaches to Breslow-Crowley:
  Meier (1975), Pollard (1990), Csörgo and Horvath, ...

Impacts & consequences

- Martingale theory:
  Le Cam → Aalen → Gill → Anderson & Gill ...
- Martingale approach to survival analysis: two books, Andersen, Borgan, Gill, and Keiding; Fleming and Harrington
- Dutch school of semiparametrics
  Groeneboom → Gill → van Zwet, Klaassen, van der Vaart
- Compact differentiability:
  Reeds → Groeneboom → Gill →
- Empirical process approaches to Breslow-Crowley:
  Meier (1975), Pollard (1990), Csörgo and Horvath, ...

Impacts & consequences

- Martingale theory:
  Le Cam → Aalen → Gill → Anderson & Gill ...
- Martingale approach to survival analysis: two books, Andersen, Borgan, Gill, and Keiding; Fleming and Harrington
- Dutch school of semiparametrics
  Groeneboom → Gill → van Zwet, Klaassen, van der Vaart
- Compact differentiability:
  Reeds → Groeneboom → Gill →
- Empirical process approaches to Breslow-Crowley:
  Meier (1975), Pollard (1990), Csörgo and Horvath, ...

Impacts & consequences

- Martingale theory:
  Le Cam → Aalen → Gill → Anderson & Gill ...
- Martingale approach to survival analysis: two books, Andersen, Borgan, Gill, and Keiding; Fleming and Harrington
- Dutch school of semiparametrics
  Groeneboom → Gill → van Zwet, Klaassen, van der Vaart
- Compact differentiability:
  Reeds → Groeneboom → Gill →
- Empirical process approaches to Breslow-Crowley:
  Meier (1975), Pollard (1990), Csörgo and Horvath, ...

Impacts & consequences

- Martingale theory:
  Le Cam → Aalen → Gill → Anderson & Gill ...
- Martingale approach to survival analysis: two books, Andersen, Borgan, Gill, and Keiding; Fleming and Harrington
- Dutch school of semiparametrics
  Groeneboom → Gill → van Zwet, Klaassen, van der Vaart
- Compact differentiability:
  Reeds → Groeneboom → Gill →
- Empirical process approaches to Breslow-Crowley:
  Meier (1975), Pollard (1990), Csörgo and Horvath, ...

Odd Aalen

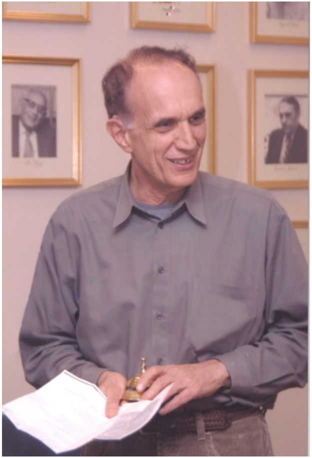
Richard Gill

Per Kragh Andersen

- Information bounds:
  Efron (1977) → Begun, Hall, Huang, Wellner (1983) → BKRW (1993)
- Confidence bands:
  Gillespie & Fisher, Hall & Wellner, Nair, Csörgo & Horvath ...
- Multivariate generalizations: Gill, Dabrowska, van der Laan, Prentice ...

- Information bounds:
  Efron (1977) → Begun, Hall, Huang, Wellner (1983) → BKRW (1993)
- Confidence bands:
  Gillespie & Fisher, Hall & Wellner, Nair, Csörgo & Horvath ...
- Multivariate generalizations: Gill, Dabrowska, van der Laan, Prentice ...

- Information bounds:
  Efron (1977) $\rightarrow$ Begun, Hall, Huang, Wellner (1983) $\rightarrow$ BKRW (1993)
- Confidence bands:
  Gillespie & Fisher, Hall & Wellner, Nair, Csörgo & Horvath ...
- Multivariate generalizations: Gill, Dabrowska, van der Laan, Prentice ...



Brad Efron

Impacts - personal I

- John Crowley's (1973) request from Lyon for

  Dudley, R. M. (1966). Weak convergence of probabilities on non-separable metric spaces and empirical measures on Euclidean spaces. Illinois Journal of Math. **10**, *109-126.*

Impacts - personal I

- John Crowley's (1973) request from Lyon for

  Dudley, R. M. (1966). Weak convergence of probabilities on non-separable metric spaces and empirical measures on Euclidean spaces. Illinois Journal of Math. **10**, *109-126.*

Dick Dudley (1971)

Impacts - personal II

- Discussions with Crowley about work with Norm
- Rochester years: 1975 - 1983.
- 1979 - 1983: efficiency theory for semiparametric models
- one point of origin: trying to understand
  Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72**, 557 - 565.
- Different set of models and problems to test statistical theory against.
- Need and motivation for the more general empirical process theory being developed by Dudley, Giné, Zinn, Pollard, and others.

Impacts - personal II

- Discussions with Crowley about work with Norm
- Rochester years: 1975 - 1983.
- 1979 - 1983: efficiency theory for semiparametric models
- one point of origin: trying to understand Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72**, 557 - 565.
- Different set of models and problems to test statistical theory against.
- Need and motivation for the more general empirical process theory being developed by Dudley, Giné, Zinn, Pollard, and others.

Impacts - personal II

- Discussions with Crowley about work with Norm
- Rochester years: 1975 - 1983.
- 1979 - 1983: efficiency theory for semiparametric models
- one point of origin: trying to understand
  Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72**, 557 - 565.
- Different set of models and problems to test statistical theory against.
- Need and motivation for the more general empirical process theory being developed by Dudley, Giné, Zinn, Pollard, and others.

Impacts - personal II

- Discussions with Crowley about work with Norm
- Rochester years: 1975 - 1983.
- 1979 - 1983: efficiency theory for semiparametric models
- one point of origin: trying to understand
  Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72**, 557 - 565.
- Different set of models and problems to test statistical theory against.
- Need and motivation for the more general empirical process theory being developed by Dudley, Giné, Zinn, Pollard, and others.

Impacts - personal II

- Discussions with Crowley about work with Norm
- Rochester years: 1975 - 1983.
- 1979 - 1983: efficiency theory for semiparametric models
- one point of origin: trying to understand
  Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72**, 557 - 565.
- Different set of models and problems to test statistical theory against.
- Need and motivation for the more general empirical process theory being developed by Dudley, Giné, Zinn, Pollard, and others.

Impacts - personal II

- Discussions with Crowley about work with Norm
- Rochester years: 1975 - 1983.
- 1979 - 1983: efficiency theory for semiparametric models
- one point of origin: trying to understand
  Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72**, 557 - 565.
- Different set of models and problems to test statistical theory against.
- Need and motivation for the more general empirical process theory being developed by Dudley, Giné, Zinn, Pollard, and others.

Impacts - personal III

- 1986 - 1996
- getting: out of the office and into the mountains
- three treks in Nepal (1989), (1996), (1999)
- skiing the the Coast Range and Cascades

Impacts - personal III

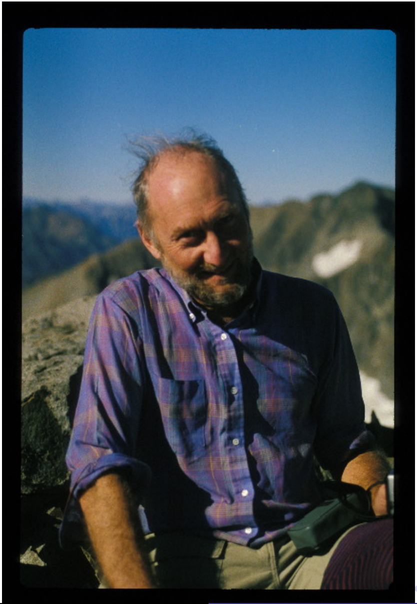
- 1986 - 1996
- getting: out of the office and into the mountains
- three treks in Nepal (1989), (1996), (1999)
- skiing the the Coast Range and Cascades

Impacts - personal III

- 1986 - 1996
- getting: out of the office and into the mountains
- three treks in Nepal (1989), (1996), (1999)
- skiing the the Coast Range and Cascades

Impacts - personal III

- 1986 - 1996
- getting: out of the office and into the mountains
- three treks in Nepal (1989), (1996), (1999)
- skiing the the Coast Range and Cascades



Impacts - personal IV

In his (1972) discussion of D. R. Cox's famous paper on the Cox proportional hazards model, Norman Breslow showed how the baseline cumulative hazard function Λ could be estimated in a natural way in parallel with the estimators of the regression parameters proposed by Cox. The resulting "Breslow estimator", together with Breslow's (1974) joint paper with John Crowley on the asymptotic behavior of the Kaplan-Meier estimator, set the stage for an extraordinary collection of developments in survival analysis, semiparametric models, and counting processes. These developments continue to have significant and profound consequences for present day statistical research and practice.

Impacts - personal IV

Norm has also had a profound effect on my own research directly (and indirectly) over a period of more than 30 years.

THANK YOU NORM