Chapter 3. Kaplan-Meier Curve and NPMLE

1. Identifiablity: noninformative versus independent censoring
2. Nonparametric estimation: Nelson-Aalen and Kaplan-Meier estimators
3. Asymptotic theory

- Likelihood contribution
  1. failure t_j: $\Pr\{T = t_j\} = \Pr\{T \geq t_j\} - \Pr\{T > t_j\} = S(t_j-) - S(t_j)$
  2. independent censoring t_{jl}: $\Pr\{T > t_{jl}\} = S(t_{jl})$, $l = 1, \ldots, m_j$
  3. total likelihood

$$\mathscr{L}(S) = \prod_{j=0}^{k} \left\{ [S(t_j-) - S(t_j)]^{d_j} \prod_{l=1}^{m_j} S(t_{jl}) \right\}$$

- Nonparametric maximum likelihood estimation (NPMLE)
  1. maximize $\mathscr{L}(S)$ with respect to S: $\widehat{S} = \arg\max_S \mathscr{L}(S)$
  2. $\widehat{S}(t)$ has to be discontinuous at t_j; otherwise $S(t_j) - S(t_j-) = 0$
  3. $t_{jl} \geq t_j$, $S(t_{jl}) \leq S(t_j) \Rightarrow \max \widehat{S}(t_{jl}) = \widehat{S}(t_j)$
  4. $\widehat{S}$ should be a step function with jumps at t_j

- How to choose jump sizes to maximize $\mathscr{L}(S)$?

1.

$$\begin{aligned} S(t_j) &= \Pr\{T > t_j\} = \Pr\{T \geq t_{j+1}\} \\ &= \frac{\Pr\{T \geq t_{j+1}\}}{\Pr\{T \geq t_j\}} \cdot \frac{\Pr\{T \geq t_j\}}{\Pr\{T \geq t_{j-1}\}} \cdot \cdots \cdot \frac{\Pr\{T \geq t_2\}}{\Pr\{T \geq t_1\}} \cdot \Pr\{T \geq t_1\} \\ &= \left[1 - \frac{\Pr\{T = t_j\}}{\Pr\{T \geq t_j\}}\right] \cdots \cdots \left[1 - \frac{\Pr\{T = t_1\}}{\Pr\{T \geq t_1\}}\right] \\ &= \prod_{l=1}^{j} (1 - \lambda_l) \end{aligned}$$

2. $S(t_j-) = \prod_{l=1}^{j-1}(1-\lambda_l) \Rightarrow [S(t_j-) - S(t_j)]^{d_j} = \lambda_j^{d_j} \prod_{l=1}^{j-1}(1-\lambda_l)^{d_j}$

3.

$$\mathscr{L}(S) = \prod_{j}^{k} \left[\lambda_j^{d_j} \prod_{l=1}^{j-1} (1-\lambda_l)^{d_j} \prod_{l=1}^{j} (1-\lambda_l)^{m_j}\right] = \prod_{j}^{k} \lambda_j^{d_j} (1-\lambda_j)^{n_j - d_j}$$

4. $\widehat{\lambda}_j = d_j/n_j \Rightarrow \widehat{S}(t) = \prod_{t_j \leq t}(1 - d_j/n_j)$ is the NPMLE.

- Kaplan-Meier estimator

$$\widehat{S}(t) = \prod_{t_j \le t} \left(1 - \frac{d_j}{n_j}\right) \approx \prod_{t_j \le t} e^{-\frac{d_j}{n_j}} = e^{-\sum_{t_j \le t} \frac{d_j}{n_j}}$$

- Assume T is continuous $\Rightarrow d_j = 1$ mostly
  1. $d_j = N(t_j) - N(t_j-) = dN(t_j)$
  2. $n_j = Y(t_j) > 0$
  3. $\sum_{t_j \le t} \frac{d_j}{n_j} = \int_{u \le t} I(Y(u) > 0) dN(u)/Y(u)$

- Nelson-Aalen estimator

$$\widehat{\Lambda}(t) = -\log \widehat{S}(t) = \int_0^t \frac{I(Y(u) > 0) dN(u)}{Y(u)}$$

3. Asymptotic theory

- Goals
  1. consistency: did we hit the target right?
  2. asymptotic distribution (normality): what is the error bound?
  3. efficiency: can we get the most of it?

- Nelson-Aalen estimator: $\widehat{\Lambda}(t) = \int_0^t I(Y(u) > 0)dN(u)/Y(u)$
  1. consistency
  2. asymptotically normal
  3. efficiency consideration

- What the Nelson-Aalen estimator estimates?
  - $E[dN_i(t) \mid \mathscr{F}_{t-}] = Y_i(t)\lambda(t)dt = Y_i(t)d\Lambda(t)$
  - $E[dN(t) \mid \mathscr{F}_{t-}] = Y(t)d\Lambda(t)$
  - $\widehat{\Lambda}(t) = \int_0^t I(Y(u) > 0)dN(u)/Y(u) \bumpeq \int_0^t I(Y(u) > 0)d\Lambda(t) = \Lambda^*(t)$
- Asymptotic normality
  - $n^{1/2}[\widehat{\Lambda}(t) - \Lambda^*(t)] \to \mathscr{N}(0, \sigma^2(t))$
- We will create a series of tools to establish the asymptotic properties

- Probability space $(\Omega, \mathscr{F}, P)$
  - $\Omega = \{\omega\}$: outcome space
  - $\mathscr{F}$: σ-algebra; a collection of subsets of Ω, satisying the closure of complementation and countable unions; large-enough to contain most of the events to have reasonable probility measure
  - P: probability measure, satisfying countably additivity $P(\bigcup_i A_i) = \sum_i P(A_i)$ for disjoint A_i's, $P(\Omega) = 1$ and $P(A) \geq 0$ for $A \in \mathscr{F}$.
- Random variable $X : \Omega \mapsto R$
  - X is $\mathscr{F}-$measurable if $\{\omega : X(\omega) \leq x\} \in \mathscr{F}$ for any x
  - distribution function: $F(x) = P[\omega; X(\omega) \leq x]$
- Stochastic process $\{X(t); t \in \Gamma\}$
  1. a collection of random variables indexed by $t \in \Gamma$
  2. for a fixed ω, $X(t; \omega)$ is a sample path as a function of t

- Example: Brownian Motion $X(t)$
  - $X(t)$ are continuous in time t
  - for any set of $t_1 < t_2 < \ldots < t_k$, $X(t_1)$, $X(t_2) - X(t_1)$, …, $X(t_k) - X(t_{k-1})$ are independent normal random variables
  - μ: drift; $EX(t_1) = \mu t_1$, $E[X(t_2) - X(t_1)] = \mu(t_2 - t_1)$, …, $E[X(t_k) - X(t_{k-1})] = \mu(t_k - t_{k-1})$
  - σ^2: diffusion; var$[(X(t_1)] = \sigma^2 t_1$, var$[X(t_2) - X(t_1)] = \sigma^2(t_2 - t_1)$, …, var$[X(t_k) - X(t_{k-1})] = \sigma^2(t_k - t_{k-1})$
  - standard Brownian motion: $\mu = 0$, $\sigma = 1$
  - time-transformed Brownian motion: $\mu = 0$, var$\{X(t)\} = \alpha(t)$

- Filtration $\mathscr{F}_t$
  - a family of sub-σ-algebra $\{\mathscr{F}_t; t \geq 0\}$
  - increasing $\mathscr{F}_s \subset \mathscr{F}_t$ for $s \leq t$
  - $\mathscr{F}_{t-} = \sigma(\bigcup_n \mathscr{F}_{t-1/n})$; $\mathscr{F}_{t+} = \bigcap_n \mathscr{F}_{t+1/n}$
  - right-continuous filtration: $\mathscr{F}_t = \mathscr{F}_{t+}$
- Histories
  1. $\mathscr{F}_t = \sigma(\{X(s); s \leq t\})$: information about $X(\cdot)$ up to t
  2. $\{W(t); t \geq 0\}$ is adapted to $\mathscr{F}_t$ if $W(t)$ is $\mathscr{F}_t-$measureable
- Counting processes $N_i(t)$
  1. right-continuous, piecewise constant, jump size of +1
  2. coupled with $Y_i(t)$

- Conditional expectation $E(Y \mid X)$
  1. $E(Y \mid X)$ is $\sigma(X)-$measurable
  2. $E(Y \mid X) = E[Y \mid \sigma(X)]$
  3. average prediction of Y given all the information we have on X
  4. $EY = E_X[E_{Y|X}(Y \mid X)]$
- Martingales $M(t)$
  1. $M(t)$ adapted to filtration $\mathscr{F}_t$
  2. $E|M(t)| < \infty$
  3. $E[M(t+s) \mid \mathscr{F}_t] = M(t) \Rightarrow E[M(t)] = 0$

- Example: Brownian motion

1. $\mathscr{F}_t = \sigma\{X(u); 0 \le u \le t\}$

$$\begin{aligned} E[X(t+s) \mid \mathscr{F}_t] &= E[X(t+s) \mid X(u); 0 \le u \le t] \\ &= E[X(t) + (X(t+s) - X(t)) \mid X(u); 0 \le u \le t] \\ &= X(t) + E(X(t+s) - X(t)) = X(t) + \mu s \end{aligned}$$

2. $\mu = 0$: martingale (fair)
3. $\mu > 0$: submartingale (winning)
4. $\mu < 0$: supermartingale (losing)

- Predictable processes $W(t)$
  1. $W(t)$ is $\mathscr{F}_{t-}-$measurable
  2. $E[W(t) \mid \mathscr{F}_{t-}] = W(t)$
  3. left-continuous process adapted to $\mathscr{F}_t$ is predictable
  4. example: at-risk process $E[Y(t) \mid \mathscr{F}_{t-}] = Y(t)$

3. Asymptotic theory

- Doob-Meyer Decomposition Theorem
  - If $X(t)$ is an adapted, right-continuous, non-negative submartingale, then there exists a unique right-continuous, increasing, predictable process $A(t)$ such that $A(0) = 0$, $E[A(t)] < \infty$ and $M(t) = X(t) - A(t)$ is a martingale
  - $dA(t) = E[dX(t) \mid \mathscr{F}_{t-}]$
- Example: $N(t)$ is a submartingale
  - right-continuous, non-negative, increasing, finite expectation
  - $A(t)$: compensator
  - $dA(t) = E[dN(t) \mid \mathscr{F}_{t-}]$
  - $A(t)$ depends on what $\mathscr{F}_t$ is.

- Martingale: $M(t)$
  - “time-dependent” residuals
  - variance: $\text{var}\{M(t)\} = EM(t)^2$
- Variance: $EM(t)^2$
  - $E[M(t+s)^2 \mid \mathscr{F}_t] \geq E[M(t+s) \mid \mathscr{F}_t]^2 = M(t)^2$: submartingale
  - Doob-Meyer: $M(t)^2 - < M, M > (t)$ is martingale

$$d < M, M > (t) = E[dM(t)^2 \mid \mathscr{F}_{t-}] = \text{var}\{dM(t) \mid F_{t-}\}$$

  - predictable variation: $< M, M > (t)$

$$\text{var}\{M(t)\} = EM(t)^2 =< M, M > (t)$$

- Example: $M(t) = N(t) - \int_0^t Y(u)\lambda(u)du$
  - $dM(t) = dN(t) - Y(t)\lambda(t)dt$
  - $$\begin{aligned} \text{var}\{dM(t) \mid F_{t-}\} &= \text{var}[dN(t) \mid \mathscr{F}_{t-}] \\ &= E[dN(t)^2 \mid \mathscr{F}_{t-}] - E[dN(t) \mid \mathscr{F}_{t-}]^2 \\ &= E[dN(t) \mid \mathscr{F}_{t-}] - E[dN(t) \mid \mathscr{F}_{t-}]^2 \\ &= Y(t)\lambda(t)dt - Y(t)[\lambda(t)dt]^2 \\ &= Y(t)\lambda(t)dt + o_p(dt) \end{aligned}$$
  - $o_p(dt)/dt \to 0 \Rightarrow \int o_p(dt) \to 0$
  - $d < M, M > (t) = Y(t)\lambda(t)dt$

- Covariance: $E[M_1(t)M_2(t)]$
  - Predictable variation $< M_1, M_2 > (t)$
    - $M_1(t)M_2(t)- < M_1, M_2 > (t)$ is martingale
    - $E[dM_1(t)M_2(t) \mid \mathscr{F}_{t-}] = d < M_1, M_2 > (t)$
  - $$\begin{aligned} d[M_1(t)M_2(t)] &= M_1(t)M_2(t) - M_1(t-)M_2(t-) \\ &= [M_1(t-) + dM_1(t)] \cdot [M_2(t-) + dM_2(t)] - M_1(t-)M_2(t-) \\ &= M_2(t-)dM_1(t) + M_1(t-)dM_2(t) + dM_1(t)dM_2(t) \end{aligned}$$
  - $$\begin{aligned} E[d(M_1(t)M_2(t)) \mid \mathscr{F}_{t-}] &= E[dM_1(t)dM_2(t) \mid \mathscr{F}_{t-}] \\ &= \text{cov}[dM_1(t), dM_2(t) \mid \mathscr{F}_{t-}] \end{aligned}$$
  - $< M_1, M_2 > (t) = 0$: orthogonal

- Example
  - $M_1(t) = N_1(t) - \int_0^t Y_1(u)\lambda(u)du$
  - $M_2(t) = N_2(t) - \int_0^t Y_2(u)\lambda(u)du$
  - $$\begin{aligned} d < M_1, M_2 > (t) &= \mathrm{cov}[dM_1(t), dM_2(t) \mid \mathscr{F}_{t-}] \\ &= E[(dN_1(t) - Y_1(t)\lambda(t)dt)(dN_2(t) - Y_2(t)\lambda(t)dt) \mid \mathscr{F}_{t-}] \\ &= E[dN_1(t)dN_2(t) \mid \mathscr{F}_{t-}] - Y_1(t)\lambda(t)dt \cdot E[dN_2(t) \mid \mathscr{F}_{t-}] \\ &\quad - E[dN_1(t) \mid \mathscr{F}_{t-}] \cdot Y_2(t)\lambda(t)dt + Y_1(t)Y_2(t)(\lambda(t)dt)^2 \\ &= E[dN_1(t)dN_2(t) \mid \mathscr{F}_{t-}] - Y_1(t)Y_2(t)(\lambda(t)dt)^2 \\ &= E[dN_1(t)dN_2(t) \mid \mathscr{F}_{t-}] + o_p(dt) \end{aligned}$$
  - $E[dN_1(t)dN_2(t) \mid \mathscr{F}_{t-}] = 0$ as long as $N_1(t)$ and $N_2(t)$ don't always jump at the same t

- Are we there yet?
  - Nelson-Aalen estimator

$$\widehat{\Lambda}(t) - \Lambda^*(t) = \int_0^t \frac{I(Y(u) > 0)}{Y(u)}[dN(u) - Y(u)\lambda(u)du]$$
$$= \sum_{i=1}^n \int_0^t \frac{I(Y(u) > 0)}{Y(u)} dM_i(u)$$

  - Parametric score function

$$l'(\theta) = \sum_{i=1}^n \int_0^\infty [\log \lambda(u \mid Z_i; \theta)]'_\theta \, [dN_i(u) - Y_i(u)\lambda(u \mid Z_i; \theta)du]$$
$$= \sum_{i=1}^n \int_0^\infty [\log \lambda(u \mid Z_i; \theta)]'_\theta dM_i(u)$$

  - what's (seemingly not) in common?

1. yes: $\sum$, $\int$, $dM_i(t)$
2. seemingly not but yes: predictable intergrand

- $L(t) = \int_0^t H(u)dM(u)$
  - $M(t)$ is $\mathscr{F}_t-$martingale
  - $H(t)$ is $\mathscr{F}_t-$predictable
  - $L(t)$ is $\mathscr{F}_t-$martingale
- $< L, L > (t)$
  - $L(t)^2$ is submartingale
  - $$\begin{aligned} d < L, L > (t) &= \mathsf{var}[dL(t) \mid \mathscr{F}_{t-}] \\ &= \mathsf{var}[H(t)dM(t) \mid \mathscr{F}_{t-}] \\ &= H(t)^2 d < M, M > (t) \end{aligned}$$
- $< L_1, L_2 > (t)$
  - $L_1(t)L_2(t)$ is submartingale
  - $$\begin{aligned} d < L_1, L_2 > (t) &= \mathsf{cov}[dL_1(t), dL_2(t) \mid \mathscr{F}_{t-}] \\ &= \mathsf{cov}[H_1(t)dM_1(t), H_2(t)dM_2(t) \mid \mathscr{F}_{t-}] \\ &= H_1(t)H_2(t)d < M_1, M_2 > (t) \end{aligned}$$

- Now it’s time to study the asymptotics of

$$U_n(t) = \sum_{i=1}^{n} \int_0^t H_i(u) dM_i(u)$$

- Are we there yet?
  - Wait a second, what do we mean by asymptotics?
  - that means “what if sample size gets big...,” i.e., $n \to \infty$
  - what n, i.e., the order of n?
  - $n^{-1}U_n$? $n^{-1/2}U_n$?

- Mode of convergence: $X_n \to X$
  - convergence in probablity $X_n \to_P X$: for any $\epsilon > 0$

$$\Pr\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\} \to 0$$

  - almost surely $X_n \to X, a.s.$:

$$\Pr\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 1$$

  - convergence in distribution $X_n \to_{\mathscr{D}} X$

$$\Pr\{\omega : X_n(\omega) \le x\} \to \Pr\{\omega : X(\omega) \le x\}$$

    at continuity points of $F_X(x)$

- Alternative view on convergence in distribution
  - any $F_n(\cdot)$ defines a measure: $\mathscr{P}_n = \mathscr{P}_{F_n}$
  - $\sigma\{(-\infty, x]; x \in R\}$ defines $\mathscr{B}(R)$
  - $X_n \rightarrow_{\mathscr{D}} X \Leftrightarrow \mathscr{P}_n(B) \rightarrow \mathscr{P}(B)$ for any $B \in \mathscr{B}$
- Weak convergence of stochastic processes
  - Γ: metric space of $X(\cdot)$
  - $\mathscr{S}_0$: σ-algebra generated by all open sets
  - $\mathscr{P}_n(A) \rightarrow \mathscr{P}(A)$ for any $A \in \mathscr{S}_0$

- Example
  - $D[0, \tau]$: cadlag functions
  - $d(x, y)$: Skorohod metric for $x(t), y(t) \in D[0, \tau]$
    1. $r(\cdot)$: strictly increasing time-transformation with $r(0) = 0$, $r(\tau) = \tau$
    2. $d(x, y) = \inf\{\epsilon; \sup_t |x(t) - y[r(t)]| \leq \epsilon, \sup_t |r(t) - t| \leq \epsilon\}$
    3. neighborhood: $U(x; \delta) = \{y(t); d(x, y) < \delta\}$
    4. open set A: for any $x \in A$, there exists $U(x; \delta) \subset A$
  - $\mathscr{S}$: $\sigma-$algebra generated by all open sets
  - $X_n(\cdot) \Rightarrow X(\cdot)$ iff $\mathscr{P}_n(A) \rightarrow \mathscr{P}(A)$ for any $A \in \mathscr{S}$

- Properties on weak convergence
  1. Continuous mapping theorem: if $X_n \Rightarrow X$, then $g(X_n) \rightarrow_{\mathscr{D}} g(X)$,
     - $\sup_t\{X_n(t)\} \rightarrow_{\mathscr{D}} \sup_t\{X(t)\}$
     - $(X_n(t_1), \ldots, X_n(t_k)) \rightarrow_{\mathscr{D}} (X(t_1), \ldots, X(t_k))$ for any $(t_1, \ldots, t_k)$
     - $\int_0^\tau H(t) dX_n(t) \rightarrow_{\mathscr{D}} \int_0^\tau H(t) dX(t)$
  2. Slutsky's theorem: if $H_n \rightarrow_{\mathscr{D}} H$, i.e., $\Pr\{\sup |H_n(t) - H(t)| > \epsilon\} \rightarrow 0$:
     - $H_n X_n \Rightarrow HX$
     - $H_n + X_n \Rightarrow H + X$

- A theorem on weak convergence of cadlag processes
  - X_n and X are stochastic processes with cadlag sample paths on $(D[0,\tau],\mathscr{S})$, such that
    1. $(X_n(t_1),\ldots,X_n(t_k)) \to_{\mathscr{D}} (X(t_1),\ldots,X(t_k))$ for any $(t_1,\ldots,t_k)$
    2. X_n is tight

    then $X_n \Rightarrow X$ on $(D[0,\tau],\mathscr{S})$

- Tightness
  - X_n is said to be tight, if there exists a compact set K such that $\mathscr{P}_n(K) > 1-\epsilon$ for any $\epsilon > 0$.
  - Stone's sufficient condition: there exists $d > 0$ such that for any $\epsilon,\eta > 0$ and $0 \le s,t \le \tau$

$$\limsup_{n\to\infty} \Pr\{\sup_{|s-t|<d} |X_n(s) - X_n(t)| > \epsilon\} < \eta$$

- What is the asymptotic (limiting) distribution of

$$U_n(t) = \sum_{i=1}^{n} \int_0^t H_i(u) dM_i(u)$$

- Answer: Martingale Central Limit Theorem (MCLT)
  - $U_n(t) \Rightarrow U$ on $(D[0, \tau], \mathscr{S})$
  - U is a time-transformed Brownian motion
    1. $EU(t) = 0$
    2. var$\{U(t)\} = \lim_{n\to\infty} < U_n, U_n > (t)$
    3. independent increment: $U(s)$ and $U(t) - U(s)$ are independent for $s \leq t$

- Assumptions
  - convergent variance: $< U_n, U_n > (t) \rightarrow v(t)$
  - smoothness (no spikes): $< U_{n,\epsilon}, U_{n,\epsilon} > (\tau) \rightarrow 0$ for any $\epsilon > 0$, where

$$U_{n,\epsilon}(t) = \sum_{i=1}^{n} \int_0^t H_i(u) I(|H_i(u)| \geq \epsilon) dM_i(u)$$

- Two more lemmae (Tsiatis, 1981):
  1. Z_i are random variables such that $Eg(Z_i)^2 < \infty$ where $g(\cdot)$ is continuous, then
$$\sup_t \left| \sum_{i=1}^{n} \frac{g(Z_i)I(X_i \geq t)}{n} - E[g(Z)I(X \geq t)] \right| \rightarrow_P 0$$
  2. $Y_n(t) \rightarrow_P Y(t)$, $Z_n(t) \rightarrow_P Z(t)$ and $Q_n(t) \rightarrow_P Q(t)$. $Q_n(t)$ is positive and increasing. $Q(\tau)$ is bounded. $f(y,z)$ is continuous. Then
$$\sup_t \left| \sum_{i=1}^{n} \int_0^t f(Y_n, Z_n)dQ_n - \int_0^t f(Y,Z)dQ \right| \rightarrow_P 0$$

- Asymptotics of Nelson-Aalen estimator
  - $$U_n(t) = \sqrt{n}[\widehat{\Lambda}(t) - \Lambda(t)] = \sum_{i=1}^{n} \int_0^t \frac{\sqrt{n}}{Y(u)} dM_i(u)$$
    for $t < \tau = \sup_t\{t; \Pr(X \geq t) > 0\}$
  - $$\begin{aligned} < U_n, U_n > (t) &= \sum_i \int_0^t \frac{n}{Y(u)^2} Y_i(t)\lambda(u)du \\ &= \int_0^t \frac{n}{Y(u)^2} Y(u)\lambda(u)du \\ &= \int_0^t \frac{\lambda(u)du}{Y(u)/n} \longrightarrow v(t) = \int_0^t \frac{\lambda(u)dt}{EY_1(u)} \end{aligned}$$
  - $$\begin{aligned} < U_{n,\epsilon}, U_{n,\epsilon} > (t) &= \int_0^t \frac{n}{Y(u)} I\left(\frac{\sqrt{n}}{Y(u)} \geq \epsilon\right) \lambda(u)du \\ &= \int_0^t \frac{1}{Y(u)/n} I\left(\frac{n^{-1/2}}{Y(u)/n} \geq \epsilon\right) \lambda(u)du \longrightarrow 0 \end{aligned}$$

- Back-of-the-envelop calculations on

$$U_n(t) = \sum_{i=1}^{n} \int_0^t H_i(u) dM_i(u)$$

1. $\mathrm{cov}\{dM_i(t), dM_j(t) \mid \mathscr{F}_{t-}\} = 0$
2. $\mathrm{var}\{dM_i(t) \mid \mathscr{F}_{t-}\} = Y_i(t)\lambda(t)dt$

$$\int_0^t \sum_{i=1}^{n} H_i(u)^2 Y_i(u)\lambda(u)du \to v(t)$$