Sampling Plans and Initial Condition Problems For Continuous Time Duration Models

James J. Heckman
University of Chicago

Econ 312, Spring 2023

Sampling Plans and Initial Condition Problems For Continuous Time Duration Models

James J. Heckman
University of Chicago

Econ 312, Spring 2019

Sampling plans and initial condition problems: Duration Models

Consider a random sample of unemployment spells in progress. For sampled spells, one of the following duration times may be observed:

- time in state up to sampling date (T_b) (recall of time spent)
- time in state after sampling date (T_a) (sampling forward)
- total time in completed spell observed at origin of sample ($T_c = T_a + T_b$)

Duration of spells beginning after the origin date of the sample, denoted T_d, are not subject to initial condition problems. The intake rate at time $-t_b$ (assuming sample occurs at time 0: the proportion of the population entering a spell at $-t_b$.

Assume:

- A time homogenous environment, i.e. constant intake rate, $k(-t_b) = k, \forall b$
- A model without observed or unobserved explanatory variables.
- No right censoring, so $T_c = T_a + T_b$
- Underlying distribution $f(x)$ is nondefective
- $m = \int_0^{\infty}(x)dx < \infty$

The proportion of the population experiencing a spell at $t = 0$, the origin date of the sample, is

$$\begin{aligned} P_0 &= \int_0^{\infty} k(-t_b)(1 - F(t_b))dt_b = k\int_0^{\infty}(1 - F(t_b))dt_b \\ &= k\left[t_b(1 - F(t_b))|_0^{\infty} - \int_0^{\infty} t_b d(1 - F(t_b))\right] \\ &= k\int_0^{\infty} t_b f(t_b)dt_b = km \end{aligned}$$

where $1 - F(t_b)$ is the probability the spell lasts from $-t_b$ to 0 (or equivalently, from 0 to $-t_b$).

So the density of a spell of length t_b interrupted at the beginning of the sample ($t = 0$) is

$$\begin{aligned} g(t_b) &= \frac{\text{proportion surviving til } t = 0 \text{ from batch } t_b}{\text{total surviving til } t = 0} \\ &= \frac{k(-t_b)(1 - F(t_b))}{P_0} = \frac{1 - F(t_b)}{m} \neq f(t_b) \end{aligned}$$

Notice: g is the distribution of T_b in the population constructed by sampling rule of source population.
Distinguish from F : *cdf* of the true population. G : *cdf* of the sampled spells.

The probability that a spell lasts until t_c given that it has lasted from $-t_b$ to 0, is the conditional density of t_c given $0 < t_b < t_c$.

$$f(t_c|t > t_b > 0) = \frac{f(t_c)}{1 - F(t_b)}; t_c \geq t_b \geq 0$$

So the density of a spell **in the sampled population** that lasts, t_c is

$$\begin{aligned} g(t_c) &= \int_0^{t_c} f(t_c|t \geq t_b) f(t \geq t_b) dt_b \\ &= \int_0^{t_c} \frac{f(t_c)}{m} dt_b = \frac{f(t_c) t_c}{m} \end{aligned}$$

Likewise, the density of a sampled spell that lasts until t_a is

$$\begin{aligned} g(t_a) &= \int_0^\infty f(t_a + t_b | t_b) Pr(t \geq t_a \geq 0)) dt_b \\ &= \int_0^\infty \frac{f(t_a + t_b)}{m} dt_b \\ &= \frac{1}{m} \int_{t_a}^\infty f(t_b) dt_b \\ &= \frac{1 - F(t_a)}{m} \end{aligned}$$

(Stationarity, mirror images have same densities). So the functional form of $f(t_b) = f(t_a)$: Consequences of stationarity.

Some useful results that follow from this model:

- If $f(t) = \theta e^{-t\theta}$, then $g(t_b) = \theta e^{-t_b\theta}$ and $g(t_a) = \theta e^{-t_a\theta}$.
  Proof:

$$
\begin{aligned}
f(t) &= \theta e^{-t\theta} \rightarrow m = \frac{1}{\theta}, \\
F(t) &= 1 - e^{-t\theta} \rightarrow g(t_a) = \frac{1 - F(t)}{m} = \theta e^{-t\theta}
\end{aligned}
$$

- $E(T_a) = \frac{m}{2}(1 + \frac{\sigma^2}{m^2})$.

Proof:

$$
\begin{aligned}
E(T_a) &= \int t_a f(t_a) dt_a = \int t_a \frac{1 - G(t_a)}{m} dt_a \\
&= \frac{1}{m}\left[\frac{1}{2} t_a^2 (1 - F(t_a))|_0^\infty - \int \frac{1}{2} t_a^2 d(1 - F(t_a))\right] \\
&= \frac{1}{m} \int \frac{1}{2} t_a^2 F(t_a) dt_a = \frac{1}{2m}[var(t_a) + E^2(t_a)] \\
&= \frac{1}{2m}[\sigma^2 + m^2]
\end{aligned}
$$

- $E(T_b) = \frac{m}{2}(1 + \frac{\sigma^2}{m^2})$. **Proof**: See proof of Proposition 2.
- $E(T_c) = m(1 + \frac{\sigma^2}{m^2})$. **Proof**:

$$E(T_c) = \int \frac{t_c^2 F(t_c)}{m} dt_c = \frac{1}{m}(var(t_c) + E^2(t_c))$$

$$\rightarrow E(T_c) = 2E(T_a) = 2E(T_b), E(T_c) > m \text{ unless } \sigma^2 = 0$$

Some Additional Results:

$$h(t) = \text{ hazard} : h(t) = \frac{f(t)}{1 - F(t)}.$$

- $h'(t) > 0 \rightarrow E(T_a) = E(T_b) < m$. **Proof:** See Barlow and Proschan.
- $h'(t) < 0 \rightarrow E(T_a) = E(T_b) > m$. **Proof**: See Barlow and Proschan.

Examples

Specification of the Distribution

Weibull Distribution

- Parameters: $\lambda > 0, k > 0$
- Probability Density Function (PDF):

$$\frac{\lambda}{k}\left(\frac{t}{\lambda}\right)^{k-1}\exp\left(-\left(\frac{t}{k}\right)^{k}\right)$$

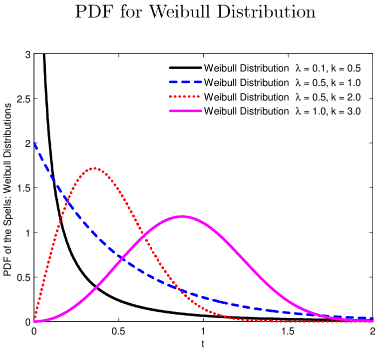
- Cumulative Density Function:

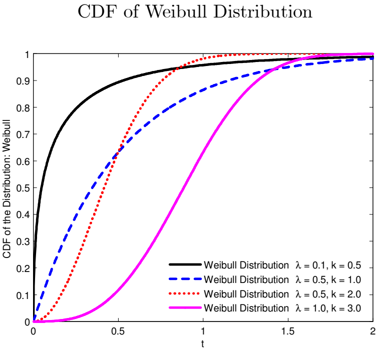
$$1-\exp\left(-\left(\frac{t}{k}\right)^{k}\right)$$

- Set of Parameters:

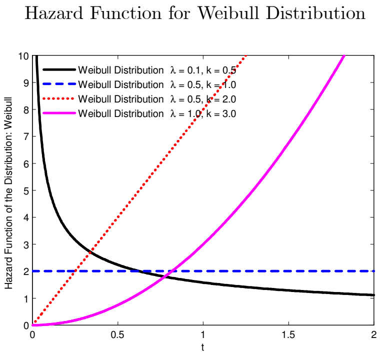
$$\left(\begin{array}{l}\lambda_1, k_1 = 0.5\\ \lambda_2, k_1 = 1.0\\ \lambda_3, k_1 = 2.0\\ \lambda_3, k_1 = 3.0\end{array}\right), \quad \text{respectively}$$

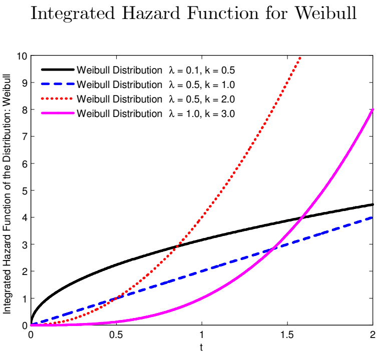
Basic Distribution Graphs



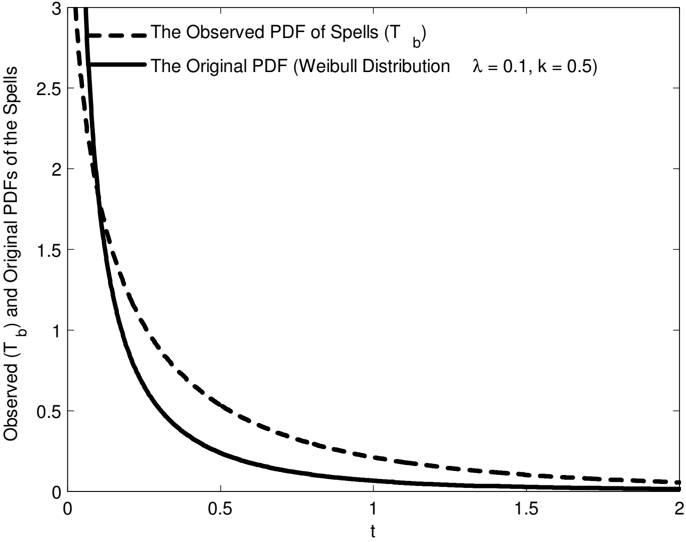


Basic Duration Graphs

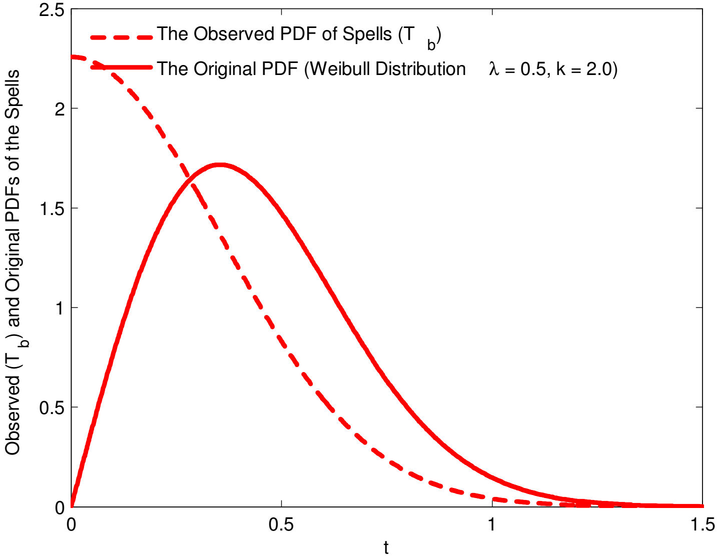




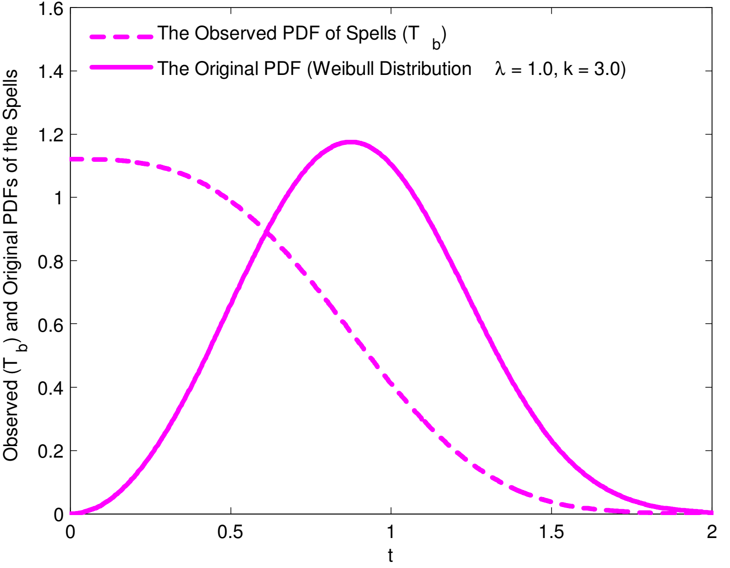
Observed and Original Distribution for T_b (Example 1)



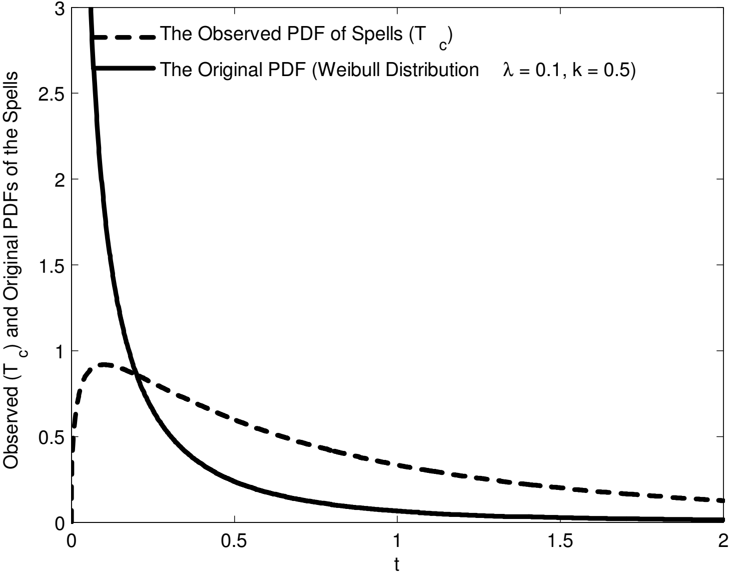
Observed and Original Distribution for T_b (Example 3)



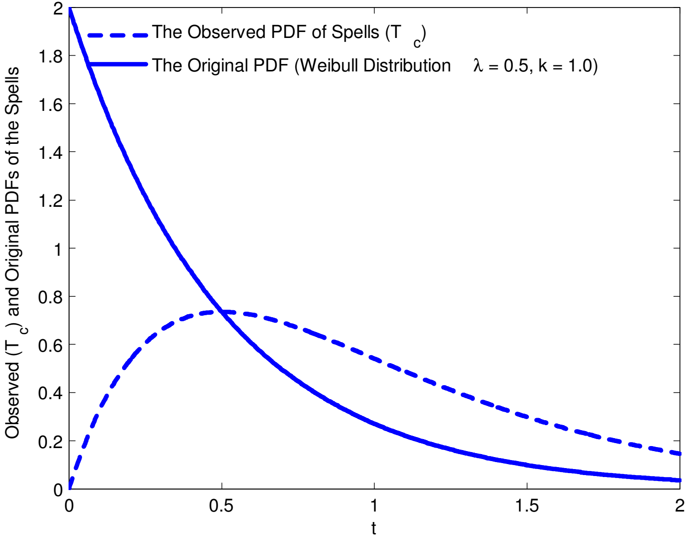
Observed and Original Distribution for T_b (Example 4)



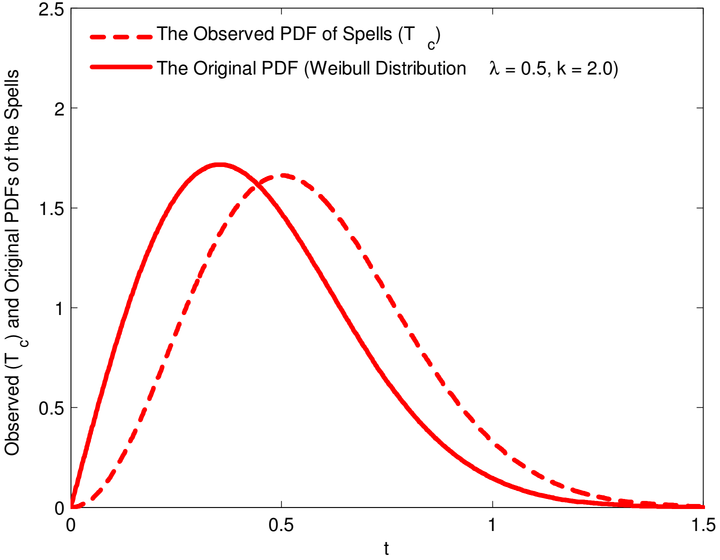
Observed and Original Distribution for T_c (Example 1)



Observed and Original Distribution for T_c (Example 2)



Observed and Original Distribution for T_c (Example 3)



Observed and Original Distribution for T_c (Example 4)