

A unified view on lifetime distributions arising from selection mechanisms

Josemar Rodrigues

Universidade Federal de São Carlos

joint work with N. Balakrishnan, G. M. Cordeiro and M. de Castro

Estatística Indutiva

01-03/2011-Setembro-São Carlos

Topics

- Selection mechanism from the carcinogenesis viewpoint

Topics

- Selection mechanism from the carcinogenesis viewpoint
- Personal probability

Topics

- Selection mechanism from the carcinogenesis viewpoint
- Personal probability
- Some special models

Topics

- Selection mechanism from the carcinogenesis viewpoint
- Personal probability
- Some special models
- Concluding Remarks

Topics

- Selection mechanism from the carcinogenesis viewpoint
- Personal probability
- Some special models
- Concluding Remarks
- References

Introduction

We have used the selection mechanism proposed by Arellano-Valle *et al.* (2006) to formulate a very flexible lifetime distribution. This distribution contains many of the recently proposed lifetime models as special cases and also facilitates in giving a biological interpretation for them.

Carcinogenesis process

- Selection mechanism from the carcinogenesis viewpoint:

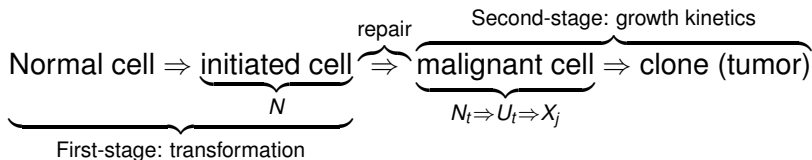


Figure: Two-stage carcinogenesis process

U_t : selection variable

X_j : the promotion time for the j th damaged cell (clonogens).

Latent cure rate model

- **First stage (transformation) $\Rightarrow N$ (damaged cells)**

$$p_n = P(N = n), \quad n = 0, 1, \dots \quad (1)$$

$$A_N(s) = \sum_{n=0}^{\infty} p_n s^n \Rightarrow (\text{pgf}) \Rightarrow p_0 = P[N = 0]. \quad (2)$$

Feller (1967): "The power and the possibilities of the pgf are rarely fully utilized."

p_0 : cure rate

Latent cure rate model

- **Second stage (growth kinetics):** Given

$N = n \Rightarrow X_j \perp X_k \mid N = n \quad (j \neq k)$ having the pdf $g(x)$ and $S(x) = 1 - G(x)$.

$$N_t = \begin{cases} Z_1 + Z_2 + \cdots + Z_N, & \text{if } N > 0, \\ 0, & \text{if } N = 0, \end{cases} \quad (3)$$

$$Z_j = \begin{cases} 1, & \text{if } X_j \leq t \Leftrightarrow \text{jth cell is activated by time } t, \\ 0, & \text{if } X_j > t \Leftrightarrow \text{jth cell is not activated by time } t, \end{cases}$$

$Z_j \approx \text{Bern}[G(t)], j = 1, \dots, n$: the presence of the j th clone by time t .

N_t : latent damage variable

$$(Z_j \perp N)$$

Flexible model for the lifetime T

- R-activation scheme by time t :

$$N_t = R \quad (R \geq 1) \Rightarrow T = X_{(R)}$$

Flexible model for the lifetime T

- R-activation scheme by time t :

$$N_t = R \quad (R \geq 1) \Rightarrow T = X_{(R)}$$

- First-activation by time t : $R = 1 \Rightarrow T = X_{(1)}$.

Flexible model for the lifetime T

- R-activation scheme by time t :

$$N_t = R \quad (R \geq 1) \Rightarrow T = X_{(R)}$$

- First-activation by time t : $R = 1 \Rightarrow T = X_{(1)}$.
- Last-activation by time t : $R = N \Rightarrow T = X_{(N)}$.

Flexible model for the lifetime T

- R-activation scheme by time t :

$$N_t = R \quad (R \geq 1) \Rightarrow T = X_{(R)}$$

- First-activation by time t : $R = 1 \Rightarrow T = X_{(1)}$.
- Last-activation by time t : $R = N \Rightarrow T = X_{(N)}$.
- **Problem**: To flexibilize the pdf $g(t)$ of T ($R = 1$ or $R = N$) of patients exposed to carcinogenesis process by time t .

Continuation...

- It follows from the fundamental formula for conditional probabilities that

$$P(N_t = j) = \sum_{n=j}^{\infty} p_n \overbrace{P(N_t = j | N = n)}^{\text{Binomial}(n, G(t)): \text{damaged mechanism}},$$

and its corresponding pgf (Feller, 1968) is

$$A_{N_t}(s) = A_N[1 - (1 - s)G(t)]. \quad (4)$$

The long-term survival function (Rodrigues *et al.*, 2008) can be obtained from (4) as

$$S_{\text{Pop}}(t) = P(T \geq t) = P[N_t = 0] = A_{N_t}(0) = A_N[S(t)]. \quad (5)$$

Continuation...

- Motivated by the work of Arellano-Valle *et al.* (2006), we start with a definition of a selection distribution and its association with the pgf $A_{N_t}(s)$ and density function $g(x)$ of the promotion time random variable X . First, we assume that the population is divided into two sub-populations of cured and non-cured patients defined by the following binary random variable (**selection mechanism**) for any time t :

$$U_t = \begin{cases} 1, & \text{if } N_t \geq 1, \\ 0, & \text{if } N_t = 0, \end{cases} \quad (6)$$

where $P(U_t = 1) = 1 - P(N_t = 0) = 1 - p_0$.

Definitions

Definition

We define **the selection distribution of T** as the conditional distribution of X given $U_t = 1$.

- This definition simply states that the selection probability distribution of T is the probability distribution of X , truncated by non-cured patients.

Definitions

Definition

We define **the selection distribution of T** as the conditional distribution of X given $U_t = 1$.

- This definition simply states that the selection probability distribution of T is the probability distribution of X , truncated by non-cured patients.



$$f_T(t) = \frac{g(t) P(U_t = 1 | X \leq t)}{P(U_t = 1)} = \frac{g(t) P(U_t = 1 | X \leq t)}{1 - p_0}. \quad (7)$$

Definitions

Definition

We define **the selection distribution of T** as the conditional distribution of X given $U_t = 1$.

- This definition simply states that the selection probability distribution of T is the probability distribution of X , truncated by non-cured patients.



$$f_T(t) = \frac{g(t) P(U_t = 1 | X \leq t)}{P(U_t = 1)} = \frac{g(t) P(U_t = 1 | X \leq t)}{1 - p_0}. \quad (7)$$

- In fact, (7) can be expressed as a **weighted distribution (Bayarri and DeGroot, 1992)**

$$f_T(t) = \frac{w(t) g(t)}{E[w(X)]}, \quad w(t) = P(U_t = 1 | X \leq t). \quad (8)$$

Personal probability

Definition

(Conditional personal non-cure rate under the first-activation) The conditional probability of the patient dying from the damaged or initiated cells (clonogens), given that $X \leq t$, called the “conditional personal non-cure rate”, is defined as

$$\gamma_{np}(t) = w(t) = P(U_t = 1 \mid X \leq t). \quad (9)$$

$$\gamma_p = 1 - \gamma_{np} \Rightarrow \text{personal probability}$$

Theorem

The crude cumulative distribution and the net survival at time t are given by

$$P(N_t = 1) = \frac{G(t) dA_N(s)}{ds} \Big|_{s=S(t)}, \quad (10)$$

$$P(N_t = N) = A_N[G(t)], \text{ respectively.}$$

Continuation

Theorem

Under the first activation and last activation we have that

$$\gamma_{np} = \left. \frac{dA_N(s)}{ds} \right|_{s=S(t)}, \quad (11)$$

$$\gamma_{np} = \left. \frac{dA_N(s)}{ds} \right|_{s=G(t)}, \text{ respectively.}$$

Continuation...

Table: Selection mechanisms and personal cure rates.

| Selection distribution | First-activation | Last-activation |
|------------------------|---|---|
| $f_T(t)$ | $\frac{g(t)}{1-p_0} \left\{ \frac{dA_N(s)}{ds} \Big _{s=S(t)} \right\}$ | $\frac{g(t)}{1-p_0} \left\{ \frac{dA_N(s)}{ds} \Big _{s=G(t)} \right\}$ |
| $S_T(t)$ | $\frac{A_N[S(t)]-p_0}{1-p_0}$ | $\frac{1-A_N[G(t)]}{1-p_0}$ |
| $h_T(t)$ | $\frac{g(t) \left\{ \frac{dA_N(s)}{ds} \Big _{s=S(t)} \right\}}{A_N[S(t)]-p_0}$ | $\frac{g(t) \left\{ \frac{dA_N(s)}{ds} \Big _{s=G(t)} \right\}}{1-A_N[G(t)]}$ |
| $\gamma_p(t)$ | $1 - \frac{dA_N(s)}{ds} \Big _{s=S(t)}$ | $1 - \frac{dA_N(s)}{ds} \Big _{s=G(t)}$ |

Some special models

- Generalized Exponential Poisson (GEP) distribution: Barreto-Souza and Cribari-Neto (2009) introduced the GEP distribution with two parameters α and λ and they showed that it has a desirable physical interpretation.

Some special models

- Generalized Exponential Poisson (GEP) distribution: Barreto-Souza and Cribari-Neto (2009) introduced the GEP distribution with two parameters α and λ and they showed that it has a desirable physical interpretation.
- Consider a sequence of independent Bernoulli trials where the k th trial has probability of success is α/k for $k = 1, 2, \dots$, $0 < \alpha < 1$. The trial number X for which the first success occurs follows the so-called Sibuya distribution with parameter α , say Sibuya(α) (Christoph and Schreiber, 2000; Devroye, 1993), given by $P(X = r) = (-1)^{r-1} \alpha (\alpha - 1) \dots (\alpha - r + 1) / r!$. The pgf of X (Pillai and Jayakumar, 1995) is

$$A_X(s) = 1 - (1 - s)^\alpha. \quad (12)$$

Some special models

- Generalized Exponential Poisson (GEP) distribution: Barreto-Souza and Cribari-Neto (2009) introduced the GEP distribution with two parameters α and λ and they showed that it has a desirable physical interpretation.
- Consider a sequence of independent Bernoulli trials where the k th trial has probability of success is α/k for $k = 1, 2, \dots$, $0 < \alpha < 1$. The trial number X for which the first success occurs follows the so-called Sibuya distribution with parameter α , say Sibuya(α) (Christoph and Schreiber, 2000; Devroye, 1993), given by $P(X = r) = (-1)^{r-1} \alpha (\alpha - 1) \dots (\alpha - r + 1) / r!$. The pgf of X (Pillai and Jayakumar, 1995) is

$$A_X(s) = 1 - (1 - s)^\alpha. \quad (12)$$

- Now, define $M \sim \text{Sibuya}(\alpha)$ and $X_j \sim P(\lambda)$, and

$$N = \begin{cases} X_1 + \dots + X_M & : \text{ if } M > 1 \\ 0 & : \text{ if } M = 0. \end{cases}$$

Some special models

- Then, we have

$$A_N(s) = 1 - [1 - \exp\{-\lambda(1 - s)\}]^\alpha. \quad (13)$$

From the first-activation mechanism in equation (13) by taking $S(x) = \exp(-\beta x)$, we obtain the GEP distribution

$$f_T(t; \theta) = \frac{\alpha\lambda\beta}{(1 - e^{-\lambda})^\alpha} \{1 - e^{-\lambda + \lambda \exp(-\beta t)}\}^{\alpha-1} e^{-\lambda - \beta t + \lambda \exp(\beta t)}, \quad (14)$$

where $\theta = (\alpha, \beta, \lambda)$. Further, if $\alpha = 1$, we have the EP distribution (?). Various properties and inferential methods for this two-parameter distribution with decreasing failure rate are discussed by ?.

The exponential-power series (EPS) distribution

- Chahkandi and Ganjali (2009) introduced a new lifetime family of distributions with DFR by combining a truncated at zero power series with some exponential distributions. Consequently, let us assume $S(t) = \exp(-\beta t)$ and the power series distribution with pdf

$$p_n(\alpha) = P(N = n; \alpha) = \frac{a_n \alpha^n}{A(\alpha)}, n = 0, 1, \dots, \quad (15)$$

where $a_n > 0$, $A(\alpha) = \sum_n a_n \alpha^n$ and $\alpha > 0$. This family of distributions includes the binomial, Poisson, negative binomial and logarithmic distributions, among others.

- The corresponding pgf is $A_N(s; \alpha) = \frac{A(\alpha s)}{A(\alpha)}$ and $p_0 = \frac{a_0}{A(\alpha)}$.
- Under the first-activation mechanism given in Table 1, we obtain the density function

$$f_T(t; \theta) = \frac{\alpha \beta \exp(-\beta t) \frac{dA_N(s; \alpha)}{ds} \Big|_{s=\exp(-\beta t)}}{A(\alpha) - a_0}, \quad (16)$$

Other recent lifetime distributions

- Classical Lehmann alternative distributions.
- Exponential Conway–Maxwell Poisson (ECOMP) distribution.
- The Exponentiated Weibull (EW) distribution.
- The Kumaraswamy G family of distributions.
- The Kumaraswamy Weibull (KwW) distribution.
- The exponential-power series (EPS) distribution.
- Beta generalized (BG) family.

Remarks

- We have used the selection mechanism proposed by Arellano-Valle *et al.* (2006) to formulate a very flexible distribution.

Remarks

- We have used the selection mechanism proposed by Arellano-Valle *et al.* (2006) to formulate a very flexible distribution.
- This unified distribution contains many of the recently proposed lifetime models as special cases and also facilitates in giving a biological interpretation for them.

Remarks

- We have used the selection mechanism proposed by Arellano-Valle *et al.* (2006) to formulate a very flexible distribution.
- This unified distribution contains many of the recently proposed lifetime models as special cases and also facilitates in giving a biological interpretation for them.
- Also, the idea of personal probability presented gives an important interpretation for the weight function, which we feel will be of interest in survival analysis.

Remarks

- We have used the selection mechanism proposed by Arellano-Valle *et al.* (2006) to formulate a very flexible distribution.
- This unified distribution contains many of the recently proposed lifetime models as special cases and also facilitates in giving a biological interpretation for them.
- Also, the idea of personal probability presented gives an important interpretation for the weight function, which we feel will be of interest in survival analysis.
- However, much more research needs to be carried out to investigate unexplored aspects of this mechanism and especially in inference problems. We hope to motivate many important applications of this selection lifetime distribution in the future.

Future research: Carcinogenesis process

- Carcinogenesis process:

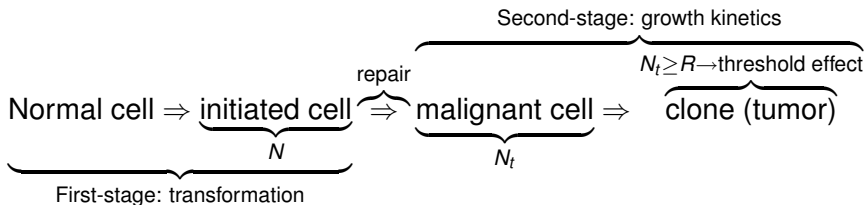


Figure: Two-stage carcinogenesis process

- **Problem:** To formulate a simple latent cure rate modeling with repair mechanism of a cell exposed to radiation in order to estimate de cure rate, volume of the tumor and the first passage time of the threshold effect.

- Arellano-Valle, R. B., Branco, M. D., and Genton, M. G. (2006). A unified view on skewed distributions arising from selections. *The Canadian Journal of Statistics*, **34**, 581–601.
- Barreto-Souza, W. and Cribari-Neto, F. (2009). A generalization of the exponential-Poisson distribution. *Statistics and Probability Letters*, **79**, 2493–2500.
- Bayarri, M. J. and DeGroot, M. (1992). A bad view of weighted distributions and selection models. *In: Bayesian Statistics 4, Eds: J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, London: Oxford University Press*, pages 17–29.
- Chahkandi, M. and Ganjali, M. (2009). On some lifetime distributions with decreasing failure rate. *Computational Statistics and Data Analysis*, **53**, 4433–4440.
- Christoph, G. and Schreiber, K. (2000). Scaled Sibuya distribution and discrete self-decomposability. *Statistics and Probability Letters*, **48**, 181–187.
- Devroye, L. (1993). A triptych of discrete distributions related to the stable law. *Statistics and Probability Letters*, **18**, 349–351.

- Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, volume I. Wiley, John Wiley & Sons, New York, 3rd edition.
- Pillai, R. N. and Jayakumar, K. (1995). Discrete Mittag-Leffler distributions. *Statistics and Probability Letters*, **23**, 271–274.
- Rodrigues, J., Cancho, V. G., de Castro, M., and Louzada-Neto, F. (2008). On the unification of the long-term survival models. *Statistics and Probability Letters*, **39**, 753–759.