Lecture 14: Hazard

Statistics 104

Colin Rundel

March 12, 2012

Gamma Function

We have just shown the following that when $X \sim \text{Exp}(\lambda)$:

$$E(X^n) = \frac{n!}{\lambda^n}$$

Lets set $\lambda = 1$ and define an new value $\alpha = n + 1$

$$E(X^{\alpha-1}) = (\alpha - 1)!$$

$$\int_0^\infty x^{\alpha-1} e^{-x} dx = (\alpha - 1)!$$

$$\Gamma(\alpha) \equiv \int_0^\infty x^{\alpha-1} e^{-x} dx = (\alpha - 1)!$$

Using a tradition definition of the factorial it only makes sense when $n \in \mathbb{N}$ but we can use this new definition of the gamma function $\Gamma(\alpha)$ for any $\alpha \in \mathbb{R}^+$

Gamma/Erlang Distribution - CDF

Imagine instead of finding the time until an event occurs we instead want to find the distribution for the time until the nth event.

Let T_n denote the time at which the nth event occurs, then $T_n = X_1 + \cdots + X_n$ where $X_1, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\lambda)$. Let $N(t)$ be the number of events that have occured at time t.

$$\begin{aligned}
F(t) = P(T_n \le t) &= P(N(t) \ge n) \\
&= \sum_{j=n}^{\infty} P(N(t) = j) \\
&= \sum_{j=n}^{\infty} \frac{e^{-\lambda t}(\lambda t)^j}{j!}
\end{aligned}$$

Gamma/Erlang Distribution - pdf

$$
\begin{aligned}
f(t) &= \frac{d}{dt}F(t) = \frac{d}{dt}\sum_{j=n}^{\infty}\frac{e^{-\lambda t}(\lambda t)^j}{j!} \\
&= \sum_{j=n}^{\infty}\left(\lambda j\frac{e^{-\lambda t}(\lambda t)^{j-1}}{j!} - \lambda\frac{e^{-\lambda t}(\lambda t)^j}{j!}\right) \\
&= \sum_{j=n}^{\infty}\lambda\frac{e^{-\lambda t}(\lambda t)^{j-1}}{(j-1)!} - \sum_{j=n}^{\infty}\lambda\frac{e^{-\lambda t}(\lambda t)^j}{j!} \\
&= \lambda\frac{e^{-\lambda t}(\lambda t)^{j-1}}{(j-1)!} + \sum_{j=n+1}^{\infty}\lambda\frac{e^{-\lambda t}(\lambda t)^{j-1}}{(j-1)!} - \sum_{j=n}^{\infty}\lambda\frac{e^{-\lambda t}(\lambda t)^j}{j!} \\
&= \lambda\frac{e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!} + \sum_{j=n}^{\infty}\lambda\frac{e^{-\lambda t}(\lambda t)^j}{j!} - \sum_{j=n}^{\infty}\lambda\frac{e^{-\lambda t}(\lambda t)^j}{j!} \\
&= \frac{e^{-\lambda t}\lambda^n t^{n-1}}{(n-1)!} = \frac{e^{-\lambda t}\lambda^n t^{n-1}}{\Gamma(n)}
\end{aligned}
$$

Erlang Distribution

Let X reflect the time until the nth event occurs when the events occur according to a Poisson process with rate λ, $X \sim \text{Er}(n, \lambda)$

$$f(x|n,\lambda) = \frac{e^{-\lambda x}\lambda^n x^{n-1}}{(n-1)!}$$

$$F(x|n,\lambda) = \sum_{j=n}^{\infty} \frac{e^{-\lambda x}(\lambda x)^j}{j!}$$

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^n$$

$$E(X) = n/\lambda$$

$$Var(X) = n/\lambda^2$$

Gamma Distribution

We can generalize the Erlang distribution by using the gamma function instead of the factorial function, we also reparameterize using $\theta = 1/\lambda$, $X \sim \text{Gamma}(n, \theta)$.

$$f(x|n,\lambda) = \frac{e^{-x/\theta}x^{n-1}}{\theta^n \Gamma(n)}$$

$$F(x|n,\lambda) = \frac{\int_0^x e^{-t/\theta}t^{n-1}\,dt}{\theta^n \Gamma(n)} = \frac{\gamma(n, x/\theta)}{\Gamma(n)}$$

$$M_X(t) = \left(\frac{1}{1-\theta t}\right)^n$$

$$E(X) = n\theta$$

$$Var(X) = n\theta^2$$

Background

A question that comes up often in many fields is if I have some item with a lifetime modeled by the random variable X with distribution function F and density f what is the probability that the item fails in the next ϵ given it has lasted t already.

$$\begin{aligned} P(t < X < t + \epsilon | X > t) &= \frac{P(t < X < t + \epsilon \cup X > t)}{P(X > t)} \\ &= \frac{P(t < X < t + \epsilon)}{P(X > t)} \\ &\approx \frac{f(t)\ \epsilon}{1 - F(t)} \end{aligned}$$

Applicable to everything from car tires and jet engines to cancer and earthquakes. Many areas of research are based on this: survival analysis, reliability analysis, duration analysis, etc.

Hazard Rate

We define the hazard rate for a distribution function F with density f to be

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{\bar{F}(t)}$$

Note that this does not make any assumptions about F or f, therefore we can find the Hazard rate for any of the distributions we have discussed so far.

A related quantity is the Survival function which is defined to be $\bar{F}(x) = 1 - F(x)$

Hazard Rate - Uniform

Let $X \sim \text{Unif}(a, b)$ where $0 \le a \le b$ then the Hazard function is

$$\begin{aligned}
\lambda(t) &= \frac{f(t)}{1 - F(t)} \\
&= \frac{\frac{1}{b-a}}{1 - \frac{t-a}{b-a}} \\
&= \frac{1}{b-t} \qquad \text{when } a \le t \le b
\end{aligned}$$

Hazard Rate, cont.

The hazard rate is enough to unique identify a distribution

$$
\begin{aligned}
\lambda(x) &= \frac{f(x)}{1-F(x)} \\
\int_0^x \lambda(t)\; dt &= \int_0^t \frac{f(x)}{1-F(x)} dt \\
&= \int_0^t \frac{\frac{d}{dt}F(x)}{1-F(x)} dt \\
&= -\log(1-F(t)) + \log(1-F(0)) \\
\int_0^x \lambda(t)\; dt &= -\log(1-F(t)) \\
1-F(t) &= \exp(-\int_0^x \lambda(t)\; dt) \\
F(t) &= 1-\exp\left(-\int_0^x \lambda(t)\; dt\right)
\end{aligned}
$$

Hazard Rate - Constant Hazard

Based on the preceding result what distribution do we get when $\lambda(t) = \lambda$?

Hazard Rate - Constant Hazard

Based on the preceding result what distribution do we get when $\lambda(t) = \lambda$?

$$\begin{aligned} F(x) &= 1 - \exp\left(-\int_0^x \lambda(t)\ dt\right) \\ &= 1 - \exp\left(-\int_0^x \lambda\ dt\right) \\ &= 1 - \exp\left(-\lambda x\right) \end{aligned}$$

$$\begin{aligned} f(x) &= \frac{d}{dx}F(x) = \frac{d}{dx}\left(1 - e^{-\lambda x}\right) \\ &= \lambda e^{-\lambda x} \end{aligned}$$

Which is the exponential distribution.

Hazard Rate - Linear Hazard

Based on the preceding result what distribution do we get when $\lambda(t) = at$?

Hazard Rate - Linear Hazard

Based on the preceding result what distribution do we get when $\lambda(t) = at$?

$$\begin{aligned}
F(x) &= 1 - \exp\left(-\int_0^x \lambda(t)\; dt\right) \\
&= 1 - \exp\left(-\int_0^x at\; dt\right) = 1 - \exp\left(-\frac{ax^2}{2}\right) \\
f(x) &= \frac{d}{dx}F(x) = \frac{d}{dx}\left(1 - e^{-ax^2/2}\right) \\
&= axe^{-ax^2/2}
\end{aligned}$$

If we reparameterize such that $a = 1/\sigma^2$ this is known as the Rayleigh distribution.

$$f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$$

This is a special case of the Weibull distribution which you will see in the homework.

Discrete RV

Previously we have discussed changes of variables / functions of random variables in terms of the effect on things like expectation and variance. For example let X be a random variable with a pmf given by $f(x)$ then let Y be a random variable that is a linear transform of X such that $Y = aX + b$ then

$$
\begin{aligned}
E(Y) &= E(aX + b) \\
&= \sum_x (ax + b) f(x) = \sum_x ax f(x) + b f(x) \\
&= \sum_x ax f(x) + \sum_x b f(x) = a \sum_x x f(x) + b \sum_x f(x) \\
&= aE(X) + b
\end{aligned}
$$

But what if we want to know the pdf or cdf of Y?

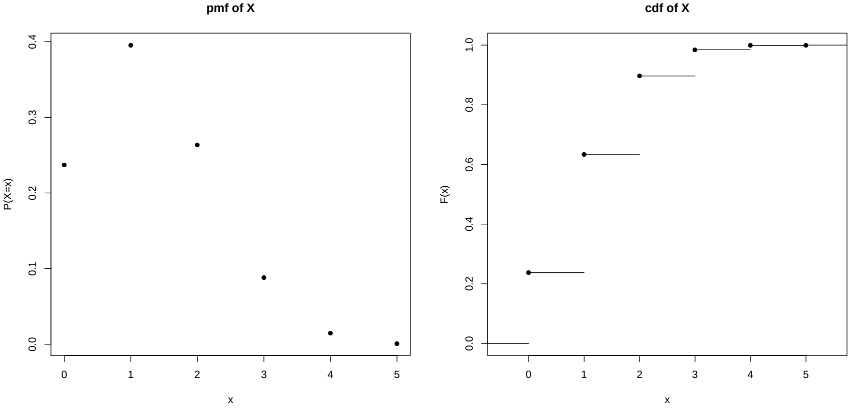
Discrete RV, cont.

Let $X \sim \text{Binom}(5, 0.25)$ what is the pmf and cdf of X?

$$f_X(x) = P(X = x) = \begin{cases} \binom{5}{x}(0.25)^x(0.75)^{5-x} & \text{If } x \in \{0,1,2,3,4,5\} \\ \\ 0 & \text{Otherwise} \end{cases}$$

$$F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} P(X = k) = \begin{cases} 0 & \text{If } x < 0 \\ \sum_{k=0}^{\lfloor x \rfloor} \binom{5}{k}(0.25)^k(0.75)^{5-k} & \text{If } 0 \le x \le 5 \\ 1 & \text{If } x > 5 \end{cases}$$

Discrete RV, cont.



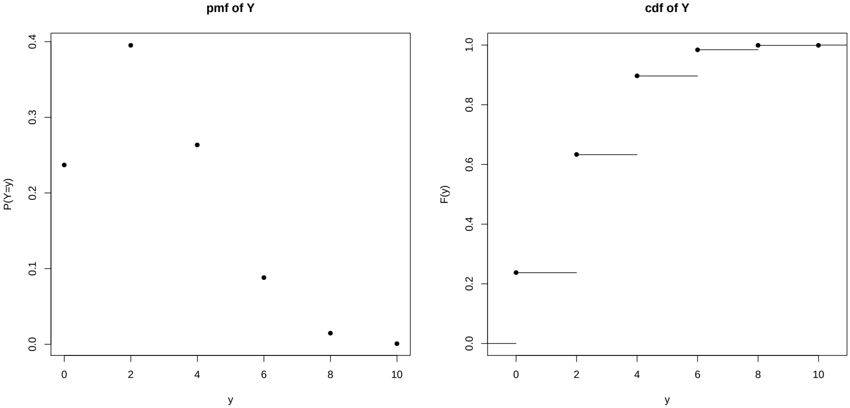
Discrete RV, cont.

Let $X \sim \text{Binom}(5, 0.25)$ and $Y = 2X$ what is the pmf and cdf of Y?

$$f_Y(y) = P(Y = y) = \begin{cases} \binom{5}{y/2}(0.25)^{y/2}(0.75)^{5-y/2} & \text{If } y \in \{0, 2, 4, 6, 8, 10\} \\ \\ 0 & \text{Otherwise} \end{cases}$$

$$F_Y(y) = \sum_{k=0}^{\lfloor y/2 \rfloor} P(Y = k) = \begin{cases} 0 & \text{If } y < 0 \\ \sum_{k=0}^{\lfloor y/2 \rfloor} \binom{5}{k}(0.25)^{k}(0.75)^{5-k} & \text{If } 0 \le y \le 10 \\ 1 & \text{If } y > 10 \end{cases}$$

Discrete RV, cont.



Continuous RV

Let $X \sim \text{Unif}(0,1)$ what is the pdf and cdf of X?

$$f_X(x) = \begin{cases} 1 & \text{If } x \in [0,1] \\ 0 & \text{Otherwise} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{If } x < 0 \\ x & \text{If } x \in [0,1] \\ 1 & \text{If } x > 1 \end{cases}$$

Continuous RV, cont.

Let $X \sim \text{Unif}(0,1)$ and $Y = 2X$ what is the pdf and cdf of Y?

The naive approach with the pdf would lead to the following:

$$f_Y(y) = \begin{cases} 1 & \text{If } y \in [0,2] \\ 0 & \text{Otherwise} \end{cases}$$

There is a problem with this:

$$\int_{-\infty}^{\infty} f_Y(y)\; dy = \int_0^2 1\; dy = y|_0^2 = 2 - 0 = 2$$

Continuous RV, cont.

Let $X \sim \text{Unif}(0,1)$ and $Y = 2X$ what is the pdf and cdf of Y?

Lets try using the cdf:

$$F_Y(y) = \begin{cases} 0 & \text{If } y < 0 \\ y/2 & \text{If } y \in [0,2] \\ 1 & \text{If } y > 2 \end{cases}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} 0 & \text{If } y < 0 \\ 1/2 & \text{If } y \in [0,2] \\ 0 & \text{If } y > 2 \end{cases}$$

Continuous RV, cont.

Let $X \sim \text{Unif}(0,1)$ and $Z = -2X$ what is the pdf and cdf of Z?

Lets try using the cdf:

$$F_Z(z) = \begin{cases} 0 & \text{If } z < -2 \\ z/2 & \text{If } z \in [-2,0] \\ 1 & \text{If } z > 0 \end{cases}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} 0 & \text{If } y < -2 \\ 1/2 & \text{If } y \in [-2,0] \\ 0 & \text{If } y > 0 \end{cases}$$

Some Quick Definitions

Monotonically increasing (increasing, non-decreasing) function:

$$x \leq y \implies f(x) \leq f(y)$$

Monotonically decreasing (decreasing, non-increasing) function:

$$x \leq y \implies f(x) \geq f(y)$$

Strictly increasing function:

$$x < y \implies f(x) < f(y)$$

Strictly decreasing function:

$$x < y \implies f(x) > f(y)$$

Continuous RV in general

Let X be a random variable with density $f_X(x)$ on the range (a,b) and let $Y = g(X)$ which will have the range $(g(a), g(b))$

$$
\begin{aligned}
F_Y(y) &= F_X(x) \\
\frac{d}{dx}F_Y(y) &= \frac{d}{dx}F_X(x) \\
f_Y(y)\frac{dy}{dx} &= f_X(x) \\
f_Y(y) &= f_X(x) \bigg/ \frac{dy}{dx}
\end{aligned}
$$

This results in a valid pdf only if we assume that $\frac{d}{dx}g(x) > 0$ on (a,b) which is only true if $g(x)$ is strictly increasing on (a,b).

Continuous RV in (more) general

More generally, let X be a random variable with density $f_X(x)$ on the range (a,b) and let $Y = g(X)$ which will have the range $(g(a), g(b))$, if $g(x)$ is either strictly increasing or decreasing on (a,b) then

$$f_Y(y) = f_X(x) \left/ \left|\frac{dy}{dx}\right| \right.$$

which takes care of the edge cases where $\frac{dy}{dx} = 0$ and $\frac{dy}{dx} < 0$

Example 1

If X is uniformly distributed over $(0, 1)$ find the density function of $Y = e^X$

Example 1

If X is uniformly distributed over $(0,1)$ find the density function of $Y = e^X$

$$\frac{dy}{dx} = \frac{d}{dx}e^x = e^x$$

$$\begin{aligned} f_Y(y) &= f_X(x) \left/ \left|\frac{dy}{dx}\right| \right. \\ &= \begin{cases} 1/e^x & \text{if } y \in (1,e) \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} 1/y & \text{if } y \in (1,e) \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Example 2 (4.4.4)

If X is uniformly distributed over $(0, 1)$ find the density function of $Y = -\lambda^{-1} \log X$

Example 2 (4.4.4)

If X is uniformly distributed over $(0,1)$ find the density function of $Y = -\lambda^{-1} \log X$

$$\frac{dy}{dx} = \frac{d}{dx} - \frac{\log x}{\lambda} = -\frac{1}{\lambda x}$$

$$\begin{aligned} f_Y(y) &= f_X(x) \bigg/ \left| \frac{dy}{dx} \right| \\ &= 1 \bigg/ \frac{1}{\lambda x} = \lambda x \\ &= \lambda e^{-\lambda y} \end{aligned}$$

Which just happens to be the exponential distribution, consequently if we can generate a uniform variable on $(0,1)$ we can easily transform it to an exponential with an arbitrary λ.