Lecture 12: Hazard & Change of Variables

Sta230 / Mth 230

Colin Rundel

March 5, 2014

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 occurred 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

Erlang Distribution

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

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

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

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

Gamma Distribution

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

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

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

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

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

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

Background

A question that comes up often in many fields is if I have some item with a lifetime modeled by the positive 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.

$$P(t < X < t+\epsilon | X > t) = \frac{P\big(\{t < X < t+\epsilon\} \cap \{X > t\}\big)}{P(X > t)} = \frac{P(t < X < t+\epsilon)}{P(X > t)} \approx \frac{f(t)\ \epsilon}{1 - F(t)}$$

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, time to event models, etc.

Hazard Rate

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

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

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

$$\overline{F}(x) = 1 - F(x)$$

Hazard Rate - Uniform

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

Hazard Rate, cont.

Just like the MGF, the hazard rate is enough to unique identify a distribution

Hazard Rate - Constant Hazard

Based on the preceding result what distribution has $\lambda(t) = \lambda$?

Hazard Rate - Linear Hazard

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

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

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

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

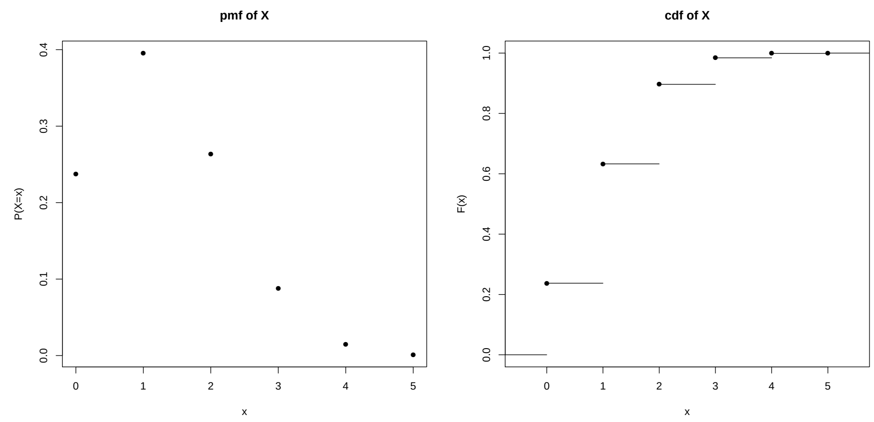
Discrete RV, cont.

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.



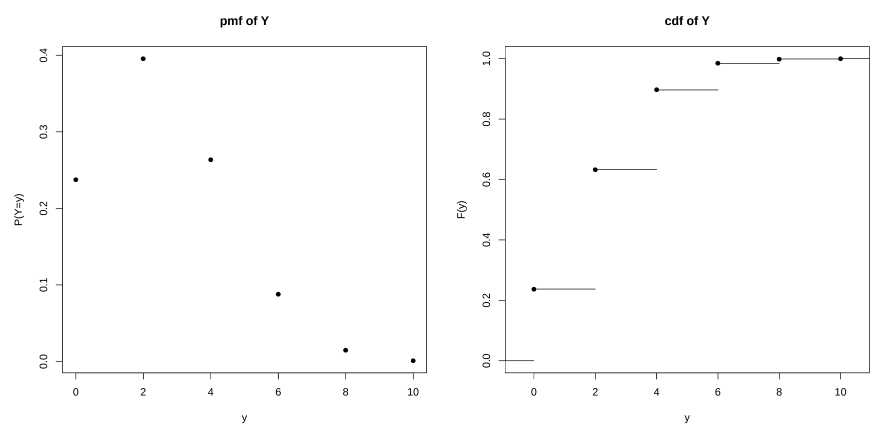
Discrete RV, cont.

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:

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

Monotonically decreasing (decreasing, non-increasing) function:

$$x \le y \implies f(x) \ge f(y)$$

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

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

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 2 (4.4.4)

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

Example 3

Let $X \sim \text{Exp}(1)$ find the pdf of the random variable Y where $Y = \log X$