AMAT 362 - PROBABILITY for STATISTICS

LECTURE 19: Parallels between GEOMETRIC & EXPONENTIAL and "The CDF Trick"

RECALL... → GEOMETRIC RVs

Two Formulations

1) $P(T=k)$ = Probability that you need k tries to obtain 1st success

$E[T] = 1/p$
$V[T] = q/p^2$

$$\boxed{P(T=k) = q^{k-1}p = (1-p)^{k-1}p}$$

[supp = {1, 2, 3, ... ∞}]

"ordinary Geom."

2) $P(F=k)$ = Probability that you fail k times before succeeding

$E[F] = \frac{q}{p}$
$V[F] = \frac{q}{p^2}$

$$\boxed{\begin{array}{c} P(F=k) = q^{k}p = (1-p)^{k}p \\ \text{PMF for Geom}(p) \text{ distribution} \end{array}}$$

[supp = {0, 1, 2, 3, ...}]

N.B. Slightly different formula for Expectation but Variance is the same.

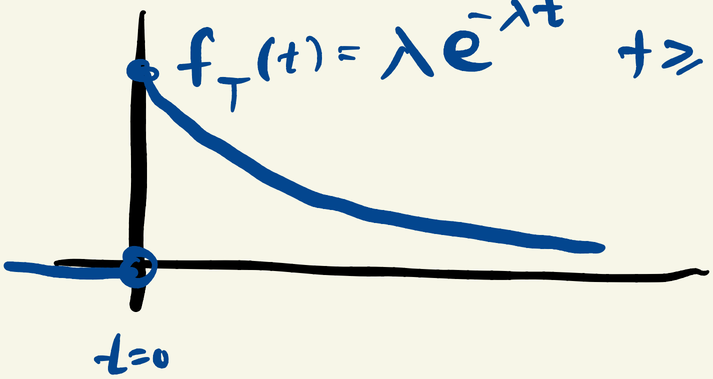
NOTICE

(values k when $P(k) \neq 0$)

$F \sim Geom(p)$ has "support" $= \{0, 1, 2, 3, \ldots\}$

COMPARE w/ EXPONENTIAL

$f_T(t) = \lambda e^{-\lambda t}$ $t \geq 0$



Facts about Exponential | $T \sim Exp(\lambda)$

1) $E[T] = \frac{1}{\lambda}$

2) $V[T] = \frac{1}{\lambda^2}$

$\rightsquigarrow SD[T] = 1/\lambda$

Geometric F

$q^k p \rightsquigarrow (1-p)^k p$

$1-x \sim e^{-x}$

$e^{-pk} \cdot p$

LOOKS A LOT LIKE

$e^{-\lambda t} \cdot \lambda$

w/ $p = \lambda$
$k = t$

$E[F] = \frac{q}{p} = \frac{1-p}{p} = \frac{1}{p} - \cancel{1}$

If p is very small then $\frac{1}{p}$ is very large

$E[F] \approx \frac{1}{p} \xleftrightarrow[p=\lambda]{} \frac{1}{\lambda} = E[T]$

$V[F] = \frac{q}{p^2} = \frac{1-p}{p^2} = \frac{1}{p^2} - \frac{1}{p}$ .. then $\frac{1}{p^2}$ dominates $\sim \frac{1}{p^2} \sim \frac{1}{\lambda^2}$

Half-Life Analogy

Suppose I draw out balls one at a time w/ replacement. After how many draws do I have less than a 50% chance of seeing no red balls?

Prob of no red balls in n draws $= \left(1 - \frac{3}{100}\right)^n < \frac{1}{2}$

$\nearrow \left(\frac{97}{100}\right)^n$

$\sim e^{-\frac{3n}{100}} < 1/2$

$\Rightarrow \quad -\frac{3n}{100} < -\ln 2$

$n > \boxed{\frac{100}{3}} \cdot \ln 2 \doteq \boxed{23.1}$

$E(T) \text{ "=" } \frac{1}{\lambda}$

Half-life for Exp(λ) $= \frac{\ln 2}{\lambda}$

FINAL PARALLEL = MEMORYLESS

$P(F > a+b \mid F > a) = P(F > b)$

$\left\{ \begin{array}{l} P(T > t+s \mid T > t) \\ \quad = P(T > s) \end{array} \right.$

Now onto ... **THE CDF TRICK**

Q: Suppose I buy 10 lightbulbs each of which have an avg lifetime of 1000 hours.

→ What is the expected time of the first failed lightbulb? (Assuming $X_i \sim Exp(\lambda = \frac{1}{1000})$)

A: $X_1, X_2, \ldots, X_{10} \sim Exp(\frac{1}{1000} = \lambda)$ IID Exponentials where

X_1 = First lightbulb's time of failure

$\vdots$

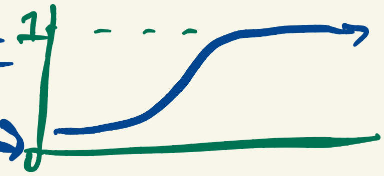
X_{10} = Tenth lightbulb's time of failure

Define $M = \min\{X_1, X_2, \ldots, X_{10}\}$

What we want is first, i.e. earliest failure

$$\boxed{\text{Q: HOW DO WE FIND THE PDF OF } M?}$$

CONSIDER THE CDF



Recall Given __any__ RV X

☆ Define $F(t) = P(X \leq t) = \int_{-\infty}^{t} f_X(s)\,ds$

to be the __CUMULATIVE__ __DENSITY__ __FUNCTION__
(or DISTRIBUTION)

APPLIED TO OUR PROBLEM

Computing $P(M \leq t)$ is hard, BUT

$P(M > t)$ is easy...

$$\parallel$$

$P(X_1 > t \text{ AND } X_2 > t \text{ AND } X_3 > t \cdots X_{10} > t)$

BY INDEP. $= P(X_1 > t)P(X_2 > t) \cdots P(X_{10} > t)$

$$= e^{-\lambda_1 t} e^{-\lambda_2 t} \cdots e^{-\lambda_{10} t}$$

$$= e^{-(\lambda_1 + \lambda_2 + \cdots + \lambda_{10})t}$$

CDF of $M = \text{Min}\{X_1, \ldots, X_{10}\}$ $X_i \sim \text{Exp}(\lambda_i)$

$$F_M(t) = 1 - e^{-(\lambda_1 + \cdots + \lambda_{10})t}$$

$$\rightsquigarrow \frac{d}{dt} F_M(t) = f_M(t) = (\lambda_1 + \cdots + \lambda_{10}) e^{-(\lambda_1 + \cdots + \lambda_{10})t}$$

PDF for $M \sim \text{Exp}(\lambda_1 + \cdots + \lambda_{10})$

ANSWER TO OUR QUESTION

What's the average time of the first failure?

$$\frac{1}{\lambda_1 + \cdots + \lambda_{10}} = E[M] \boxed{= 100 \text{ hours}}$$

When $\lambda_1 = \lambda_2 = \cdots = \lambda_{10} = \frac{1}{1000}$

$$\Rightarrow \lambda_1 + \cdots + \lambda_{10} = \frac{10}{1000} = \frac{1}{100}$$

UPSHOT → If you had 1000 1000-hr lightbulbs then you'd expect to replace one every hour!

ANOTHER EXAMPLE

Time of 2nd arrival / failure in Poisson Point process

$P(T_2 > t)$ = Prob of 0 or 1 arrivals in $[0,t]$

$$\sim P(N([0,t]) = 0 \text{ or } 1)$$

Poisson RV w/ λt parameter

$$= \boxed{\frac{(\lambda t)^0}{0!} e^{-\lambda t} + \frac{(\lambda t)^1}{1!} e^{-\lambda t}} \quad \star$$

$\leadsto$ CDF for $T_2 = P(T_2 \leq t) = 1 - \star$

So PDF = Derivative of CDF

$$\boxed{f_{T_2}(t) = \lambda^2 t e^{-\lambda t}}$$

PDF for time of 2nd arrival

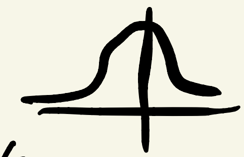
RECALL → $\text{Gamma}(r, \lambda) = \dfrac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}$

when $\Gamma(r) = \int_0^\infty x^r e^{-x}\,dx \underset{\uparrow}{=} (r-1)!$

when r is an integer

FUNCTIONS of a RV & THEIR PDFs

Consider $X \sim \text{Norm}(0,1)$



mean $= 0$
var $= 1$

PDF for $X \rightsquigarrow \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} = f_X(x) = \varphi$

Considr $Y = X^2$. HOW TO FIND PDF?

$\Rightarrow$ Compute CDF $\rightsquigarrow P(Y \le y) = P(X^2 \le y)$

$= P(-\sqrt{y} \le X \le \sqrt{y})$

Δ

$\Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = \boxed{2\Phi(\sqrt{y}) - 1 = \text{CDF}}$

$\Rightarrow$ Now differentiate! $\dfrac{d}{dy}\left(2\Phi(\sqrt{y}) - 1\right) = 2\dfrac{d}{dy}\Phi(\sqrt{y})$

$\Rightarrow 2\varphi(\sqrt{y}) \cdot \dfrac{d}{dy}(\sqrt{y}) = \cancel{2}\varphi(\sqrt{y}) \cdot \dfrac{1}{\cancel{2}\sqrt{y}} = \dfrac{\varphi(\sqrt{y})}{\sqrt{y}}$

For $X \sim \text{Norm}(0,1)$ PDF $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

For $Y = X^2$ PDF $\frac{\varphi(\sqrt{y})}{\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-y/2}$

Given $\Gamma(1/2) = \sqrt{\pi}$ $\rightsquigarrow$ $\boxed{= \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)}$

Observe that X^2 for $X \sim$ Norm & Exp(λ) all fit into the single family of Gamma!

Linear (Affine) FUNCTIONS of a RV

$$Y = aX + b$$

Compute CDF for Y

$$P(Y \le y) = P(aX + b \le y)$$
$$= P\left(X \le \frac{y-b}{a}\right)$$
$$= F_X\left(\frac{y-b}{a}\right) = \text{CDF for } X$$

$\rightarrow$ Differentiate

$$f_Y = \frac{d}{dy} F_X\left(\frac{y-b}{a}\right)$$
$$= F_X'\left(\frac{y-b}{a}\right) \frac{d}{dy}\left(\frac{y-b}{a}\right)$$

$$\boxed{\text{For } a > 0 \quad Y = aX + b \qquad f_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}}$$

FINAL FORMULA

If X is a RV w/ PDF f_X and g is differentiable w/ $g'(x)=0$ only at finitely many x then

$$f_Y(y) = \sum_{\substack{x \mid g(x)=y \\ g'(x)\neq 0}} f_X(x) \cdot \frac{1}{|g'(x)|}$$

is the PDF for $Y=g(X)$.

Example $g(x)=ax+b$ $\quad a\neq 0$

slope $\neq 0$

$$f_Y(y) = \sum_{x \mid ax+b=y} f(x)\cdot\frac{1}{|a|}$$

only 1 value

$x=\frac{y-b}{a}$ so sum disappears

$$= \boxed{f\left(\frac{y-b}{a}\right)\cdot\frac{1}{|a|}}$$