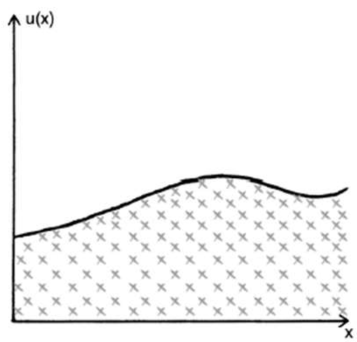
Module 2

Probability Distributions

Mini-Project of Module 2

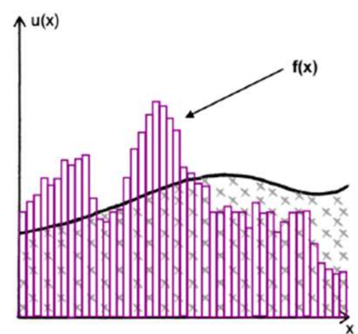
Predicting Insurance Claim Aggregates
during a
Policy Period

Geometrical Interpretation of Integration with respect to a Distribution Function



$$\int_{\mathfrak{D}} u(x)dx \cong \sum_{x_i \in \mathfrak{D}} u(x_i^*)\Delta x_i$$

Figure 3.5: Profile of the function $u(x)$ along x. The shaded area under the curve $u(x)$ is given by $\int u(x)dx$.



$$\int_{\mathfrak{D}} u(x)f(x)dx \cong \sum_{x_i \in \mathfrak{D}} u(x_i^*)f(x_i^*)\Delta x_i$$

$$\int_{\mathfrak{D}} u(x)dF(x)$$

where

$$f(x) = \frac{dF(x)}{dx}$$

Figure 3.6: Profile of the weight function $f(x)$ demonstrates the relative importance of the observables x in $\mathfrak{D}$.

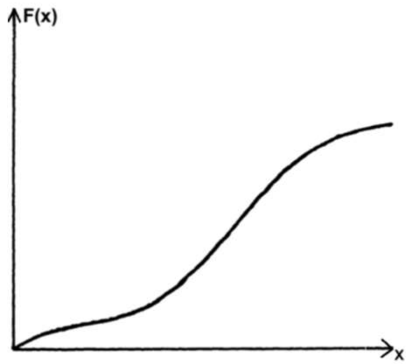


Figure 3.7: Distribution profile of the observables x is prescribed by some function $F(x)$

$F(x)$: Cumulative Distribution Function

with

$f(x)$: Probability Density Function

$$\int_{\mathfrak{D}} u(x)f(x)dx = \int_{\mathfrak{D}} u(x)dF(x)$$

Over the range of x between (-∞, ∞), $F(x)$ varies between 0 and 1.

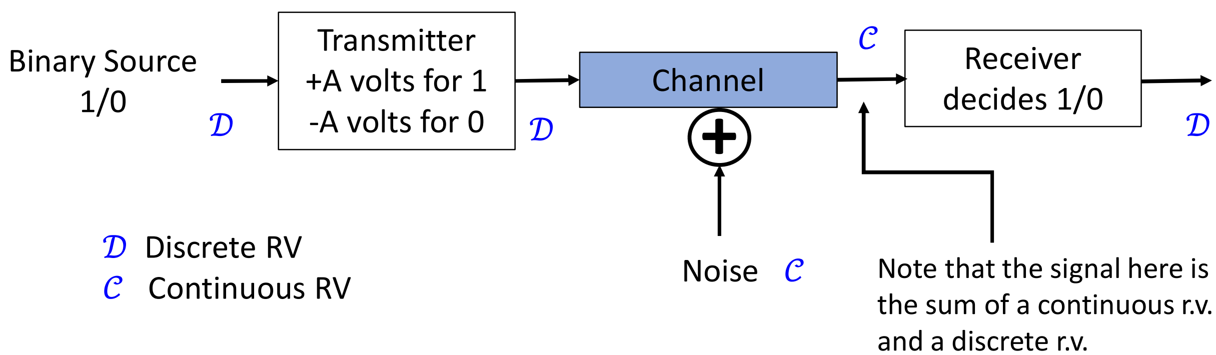
Discontinuities (positive jumps in $F(x)$) will show up as delta functions in $f(x)$, e.g., of the type $a\delta(x - x_0)$ for a jump of a in $F(x)$ at $x = x_0$

Discrete vs Continuous Probability Distributions

Random Variable

- **Discrete Random Variable**: Random variable which takes on distinct values, e.g., tossing a coin (H or T), throwing a dice (1, 2, 3, 4, 5, 6)
- **Continuous Random Variable**: Random variables may take on a continuum of values over some specified range (e.g., in a sample space Ω), e.g., Temperature at a particular day, time and place

Example of a Communication System

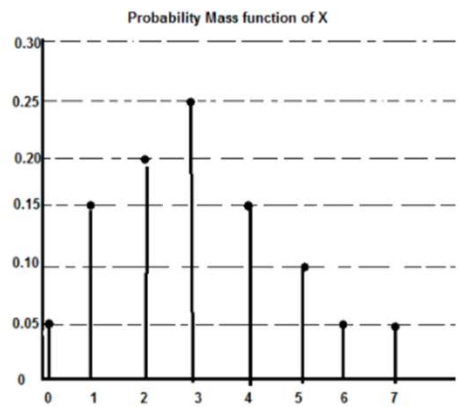


Probability Distribution Profile

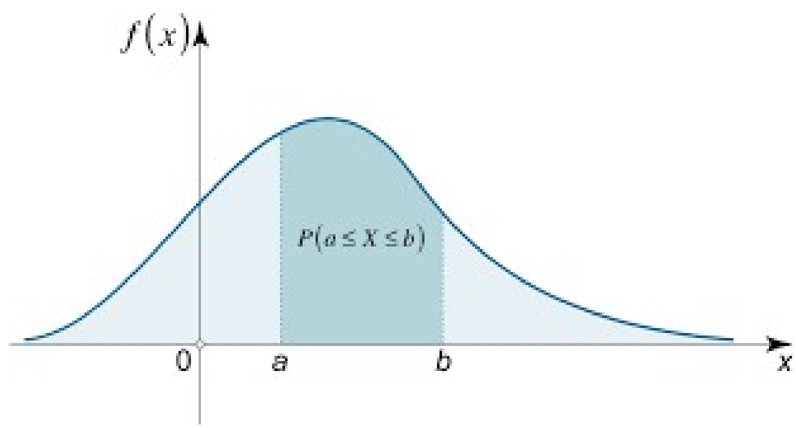
Discrete Random Variable

Continuous Random Variable

Probability Mass Function



Probability Density Function



Probability Mass Function

For a *Discrete Random Variable*, each possible observable $x_i \in \Omega$ has a certain probability of occurrence $p_i \coloneqq P(X = x_i)$ which we can think of as its *probability mass*

Axiom of Unitarity $\Rightarrow$ $\sum_{x_i \in \Omega} P(X = x_i) = 1$

Convenient Notation: $P(x_i) \equiv p_i$

This is to be read as the "probability mass function of the random variable X for the value x_i"

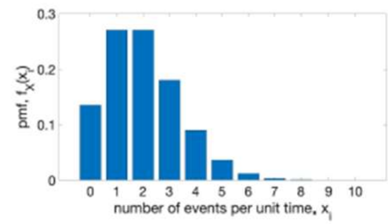


Figure 3.9: Probability mass function $f_X(x_i)$ of a certain discrete random variable. It may be verified that $\sum_{x_i \in \Omega} f_X(x_i) = 1$.

Probability Density Function (PDF)

In the case of a continuous random variable X, the probability mass is spread continuously over the range of the observables.

Therefore, it is appropriate to use the notion of a density function $f_X(x)$, instead of a probability mass.
This is interpreted as "$f_X(x)dx$ is the probability of the random variable X lying between x and $x + dx$".

The unitarity axiom of probability then enforces the normalization of the *probability density function* (pdf) as –

$$\int_{x\in\Omega} f_X(x)dx = 1$$

It also follows that $P(a \leq X \leq b) = \int_a^b f_X(x)dx$

Area under the curve f between a and b as in Fig. 3.10

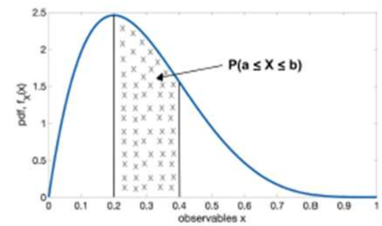


Figure 3.10: Probability density function $f_X(x)$ of a certain continuous random variable.

$f_X(x)dx$ is a **probability**, but $f_X(x)$ is not! $f_X(x)$ is the **probability density**

8

Cumulative Distribution Function (CDF)

The cumulative distribution function (cdf) $F_X: \mathbb{R} \to [0, 1]$ is defined as

$$F_X(x) \equiv F(x) := P(X \le x), x \in \mathbb{R}$$

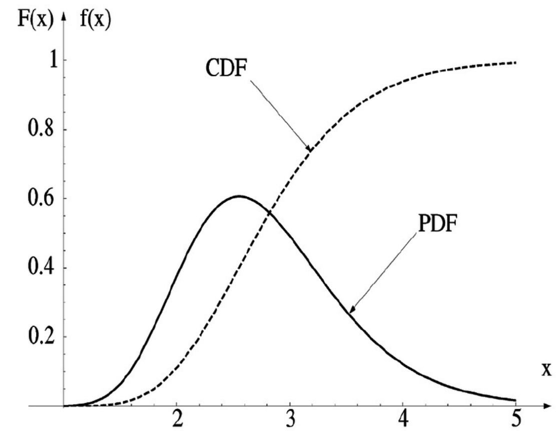
It follows that $P(a \le X \le b) = \int_a^b f(x)dx = F(b) - F(a)$

The cdf F must also satisfy the following properties –

(i) $lim_{y\downarrow -\infty} F(y) = 0$ y tends to -∞ from the right
(ii) $lim_{y\uparrow \infty} F(y) = 1$ y tends to +∞ from the left
(iii) $lim_{y\downarrow x} F(y) = F(x), \forall x \in R$ ($i.e. F_X$ is right continuous)
y tends to x from the right

The properties (i) and (ii) imply that F is a non-decreasing function going from 0 to 1.

For a Continuous Random Variable, $F(x) = \int_{-\infty}^{x} f(\alpha)d\alpha$
and $\frac{dF(x)}{dx} = f(x) \Rightarrow dF(x) = f(x)dx$



Cumulative Distribution Function (CDF) …… *continued*……

There are two main interpretation of the distribution function $F_X(x)$ that is noteworthy to mention here.

(I) $F_X(x)$ is the distribution of unit mass on the real line. Therefor, $F(b) - F(a)$ is the mass concentrated in the interval $(b - a)$.

For the discrete case, locations of concentrated point mass on the real line (x_i) are points of discontinuity of F_X with jumps proportional to $p_i \equiv F_X(x_i + 0) - F_X(x_i - 0)$. There are a finite of countable number of such jumps and F_X is continuous everywhere else. The corresponding PDF has delta functions $\delta(x - x_i)$ with weight p_i at each such x_i, i.e., $x_i \delta(x - x_i)$.

(ii) $F_X(x)$ encompasses the accumulation of probability masses (or density) up to x. Therefore, it is *additive*, non-negative, and has a unit maximum value.

Statistical Moments and their Significance

$\boldsymbol{X}$: Random Variable (Discrete or Continuous) and the observables $x \in \Omega$

Mean ($\mu, \mu_X, E(X), \bar{X}$) **First Moment of the random variable $\boldsymbol{X}$**

$$E(X) = \sum_{x\in\Omega} xP(X = x) \qquad \textit{Discrete Case}$$

$$E(X) = \int_{x\in\Omega} xf(x)dx = \int_{x\in\Omega} x\,dF(x)dx \qquad \textit{Continuous Case}$$

$P(X)$: 0 X=1, 0.5 X=2, 0.25 X=3, 0.25 X=4

$\mu_X = 0 + 1 + 0.75 + 1 = 2.75$

$$f_X(x) = \lambda e^{-\lambda x} \qquad x \geq 0$$

$$\mu_X = \int_0^\infty x\left(\lambda e^{-\lambda x}\right) dx = \frac{1}{\lambda}$$

Poisson Distribution

Statistical Moments and their Significance

$\boldsymbol{X}$: Random Variable (Discrete or Continuous) and the observables $x \in \Omega$

Variance: $(\sigma^2, \sigma_X^2, Var(X))$ **Second Statistical Moment of the random variable $\boldsymbol{X}$**

$$Var(X) = E\left(\left(X-\mu\right)^2\right) = \sum_{x\in\Omega}\left(x-\mu\right)^2 P\left(X=x\right) \qquad \textit{Discrete Case}$$

$$Var(X) = \int_{x\in\Omega}\left(x-\mu\right)^2 f(x)dx = \int_{x\in\Omega}\left(x-\mu\right)^2 dF(x)dx \qquad \textit{Continuous Case}$$

Note that –
$\sigma_X^2 = \overline{X^2} - \bar{X}^2$

$P(X)$: 0 X=1, 0.5 X=2, 0.25 X=3, 0.25 X=4
$\overline{X^2} = 0 + 2 + 2.25 + 4 = 8.25$
$\sigma_X^2 = 8.25 - 2.75^2 = 0.6875$

$$f_X(x) = \lambda e^{-\lambda x} \qquad x \geq 0$$

$$\sigma_X^2 = \int_0^\infty \left(x - \frac{1}{\lambda}\right)^2 \left(\lambda e^{-\lambda x}\right) dx = \frac{1}{\lambda^2}$$

Statistical Moments and their Significance .... *continued* ......

$\boldsymbol{X}$: Random Variable (Discrete or Continuous) and the observables $x \in \Omega$

Skewness: (μ_3) **Third Standardized Moment of the random variable X**

$$\mu_3 = E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right) = \frac{E\left((X-\mu)^3\right)}{(Var(X))^{\frac{3}{2}}}$$

μ_3 measures the *Degree of Asymmetry* of the pdf.

For example, a pdf that is symmetric about the mean has zero skewness and all its higher order moments about the mean will also be obviously zero.

Data with positive skewness has a pdf with a longer tail for $X - \mu_X > 0$ than for $X - \mu_X < 0$

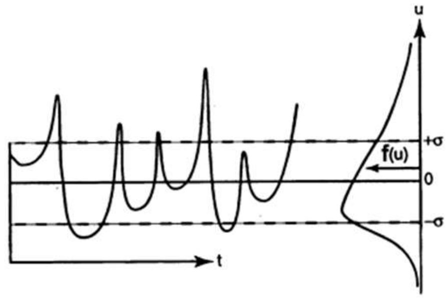


Figure 3.12: Time series data $u(t)$ with positive skewness ($\mu_3 > 0$).

13

Statistical Moments and their Significance ……. *continued* ……

$\boldsymbol{X}$: Random Variable (Discrete or Continuous) and the observables $x \in \Omega$

Skewness: (μ_3) **Third Standardized Moment of the random variable** $\boldsymbol{X}$

$$\mu_3 = E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right) = \frac{E\left((X-\mu)^3\right)}{\left(Var(X)\right)^{\frac{3}{2}}}$$

μ_3 measures the *Degree of Asymmetry* of the pdf.

For example, a pdf that is symmetric about the mean has zero skewness and all its higher order moments about the mean will also be obviously zero.

Data with positive skewness has a pdf with a longer tail for $X - \mu_X > 0$ than for $X - \mu_X < 0$

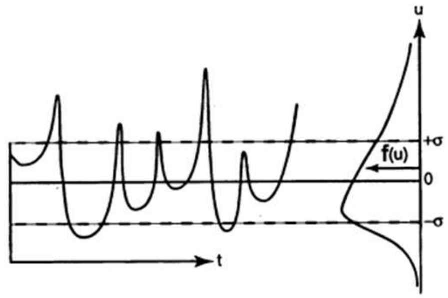
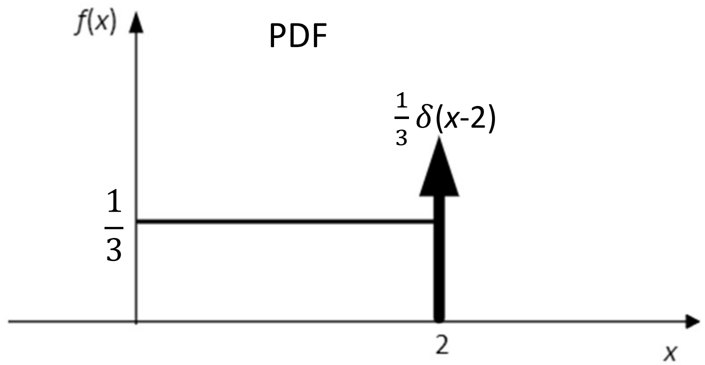
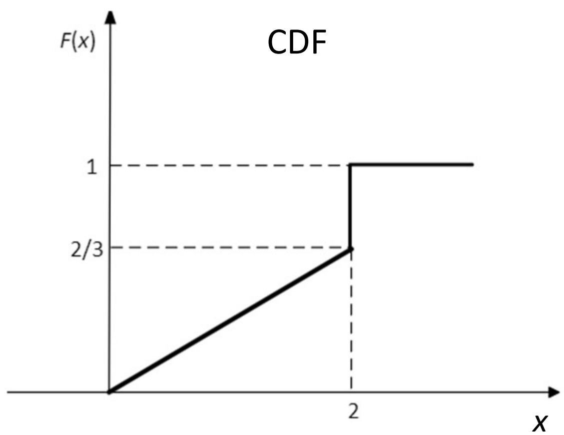


Figure 3.12: Time series data $u(t)$ with positive skewness ($\mu_3 > 0$).

Example 1 Consider a random variable X with the cdf as shown

$$F_X(x) = 0 \quad x < 0$$
$$= \frac{x}{3} \quad 0 \le x < 2$$
$$= 1 \quad x \ge 2$$





$$\frac{1}{3}\int_2^{\infty}(2)\delta(x-2)dx$$

Mean: $\bar{X} = \int_0^2 \frac{1}{3} x dx + \left(\frac{1}{3}\right)(2) = \frac{4}{3}$

Second Moment: $\overline{X^2} = \int_0^2 \frac{1}{3} x^2 dx + \left(\frac{1}{3}\right)(4) = \frac{20}{9}$

Variance: $\sigma^2 = E\left[(X-\bar{X})^2\right] = \overline{X^2} - \left(\bar{X}\right)^2 = \frac{4}{9}$

15

Example 2 Operation of an Insurance Policy (*insuring against a business loss*)

An insurance policy reimburses a loss up to a benefit limit of C but has a deductible of d.

Suppose that the policyholder's loss, X has the pdf $f_X(x) = \frac{1}{5}e^{-\frac{x}{5}}, x \geq 0$.

Let Y denote the benefit paid under the insurance policy.

Find the distribution of Y.

For $0 \leq X \leq d$, no benefit will be paid, i.e.,

$Y = 0$ with probability $\int_0^d f_X(x)dx = 1 - e^{-\frac{d}{5}}$

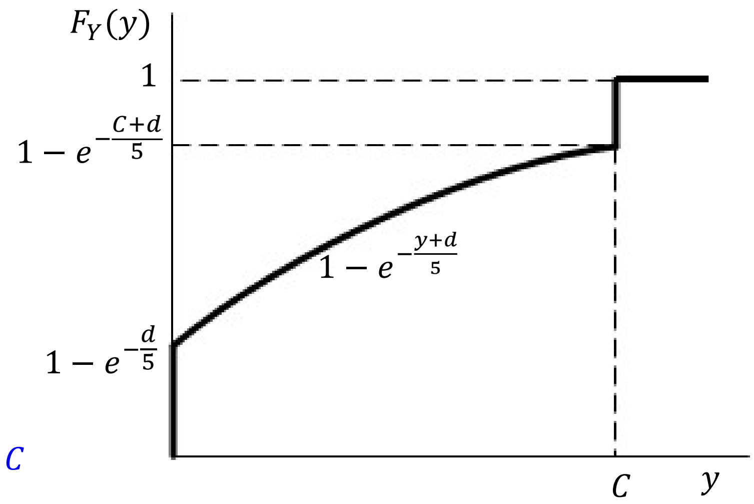
For $X \geq C + d$, the benefit is fixed at

$Y = C$ with probability $\int_{C+d}^{\infty} f_X(x)dx = e^{-\frac{C+d}{5}}$

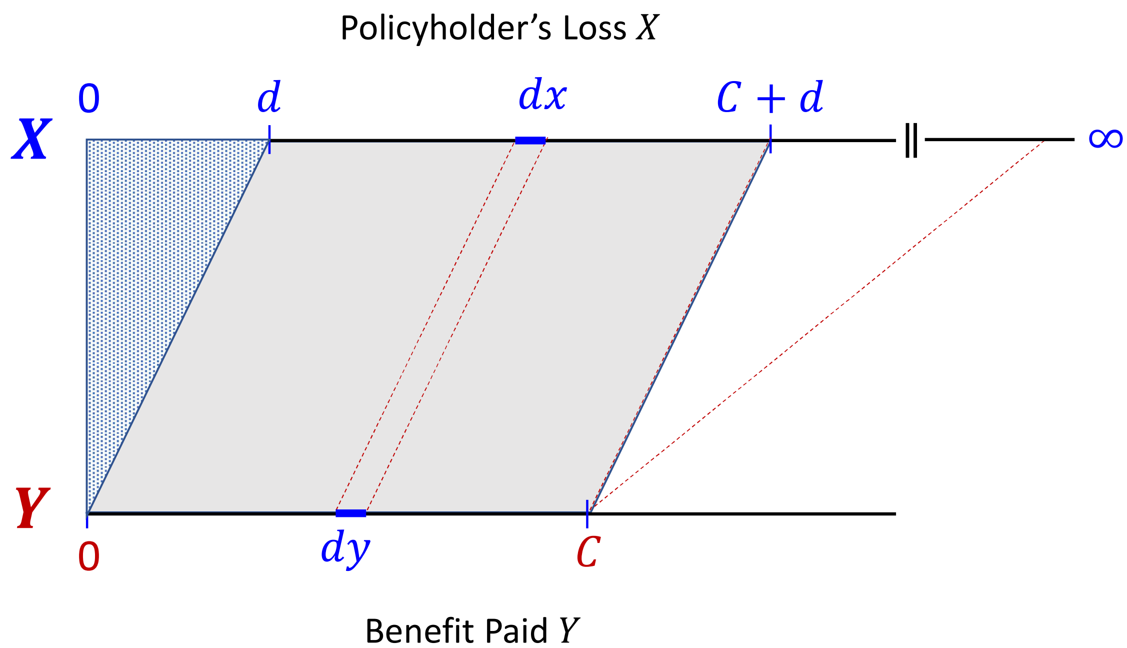
For $d \leq X < C + d$, the benefit varies as

$Y = X - d$ with probability $f_Y(y) = \frac{1}{5}e^{-\frac{y+d}{5}} \quad 0 < y < C$

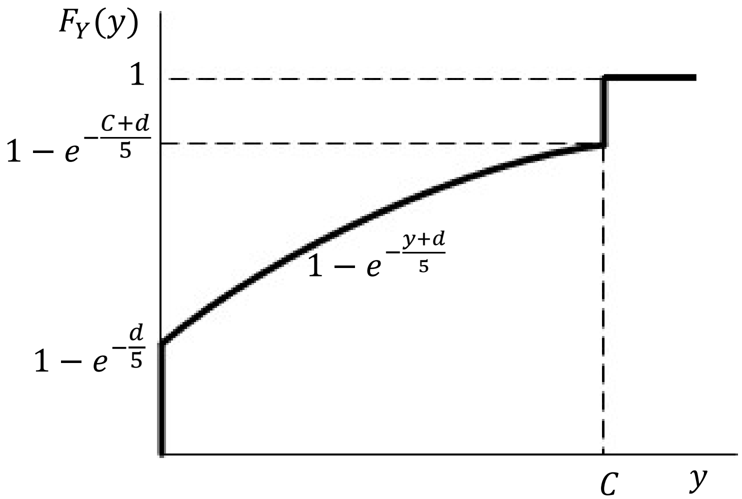
and cdf $F_Y(y) = 1 - e^{-\frac{d}{5}} + \int_0^y \frac{1}{5}e^{-\frac{y+d}{5}}dy = 1 - e^{-\frac{y+d}{5}} \quad 0 < y < C$



See Next Slide for a more graphical interpretation



Policyholder's Loss X
0
d
dx
$C + d$
∞
X
Y
0
dy
C
Benefit Paid Y

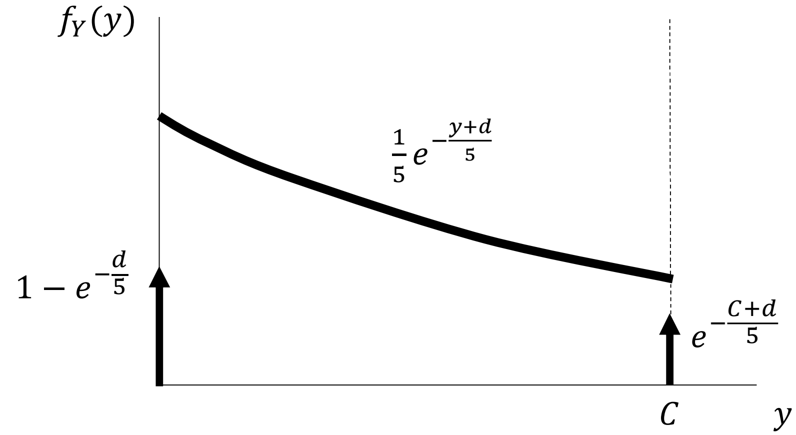


$\Rightarrow$

$$f_Y(y) = \left(1-e^{-\frac{d}{5}}\right)\delta(y) \qquad y=0$$

$$= \frac{1}{5}e^{-\frac{y+d}{5}} \qquad 0<y<C$$

$$= e^{-\left(\frac{C+d}{5}\right)}\delta(y-C) \qquad y=C$$



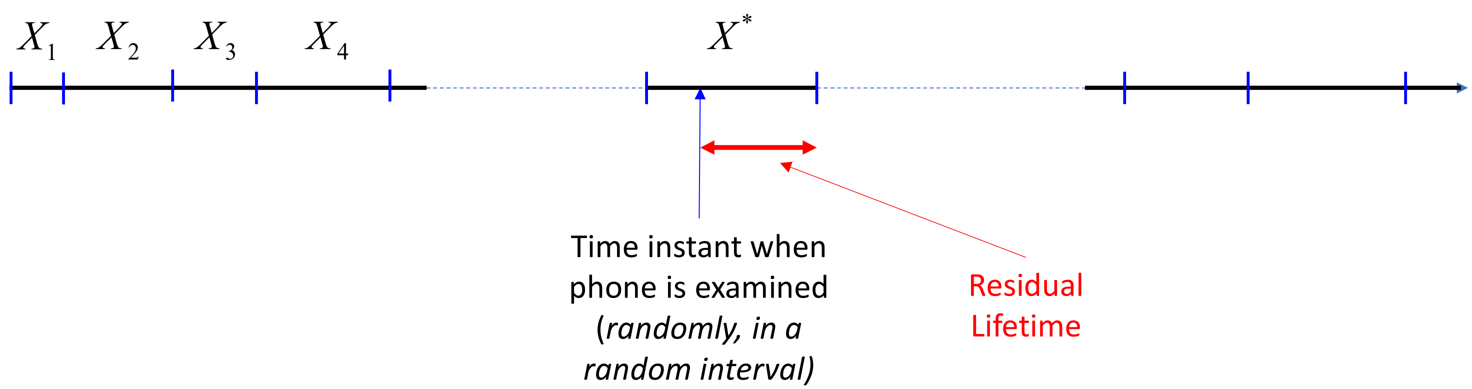
Paradox of Residual Life

Your experiences with a cheap mobile phone and a super efficient repair person!

The phone has a lifetime given by the random variable X with pdf $f_X(x), 0 \leq x < \infty$ and mean $\bar{X}$. Your repair person is super-good and can immediately fix the phone and put it back in service once again!

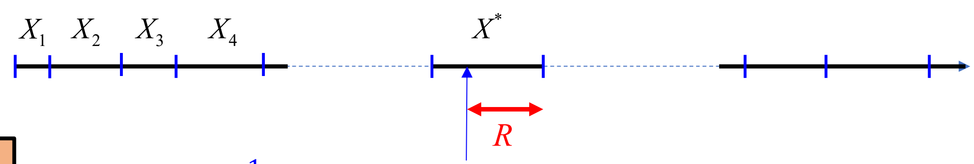
Your father/mother wants to decide whether you have wasted your money or not and wants to check (at a random time instant) to see what is the time from that instant to when the phone fails next (Residual Life)

Time Line of Your Phone



What is the Mean Residual Life Time, i.e. Mean Time to Next Breakdown ?

Mean Life Time $\overline{X} = E(X) = \int_0^\infty x f_X(x)dx$



"Paradox of Residual Life" – The Mean Residual Time $\overline{R}$ you would see is **not** $\frac{1}{2}\overline{X}$!

Actually, $\overline{R} \geq \frac{1}{2}\overline{X}$

We can see that $\overline{R} = \frac{1}{2}E(\overline{X^*})$ but to find that we need to find $f_{X^*}(x)$, the pdf of the selected lifetime

We can argue from simple logic that $f_{X^*}(x) = Kxf_X(x)$ and the normalization condition requires

$$\Rightarrow \int_0^\infty f_{X^*}(x)dx = 1 \quad \Rightarrow \quad K = \frac{1}{\overline{X}}, \quad f_{X^*}(x) = \frac{x}{\overline{X}} f_X(x)$$

Therefore,

$$\overline{R} = \int_0^\infty \left(\frac{1}{2}x\right)\left(\frac{x}{\overline{X}} f_X(x)\right)dx = \frac{\overline{X^2}}{2\overline{X}} = \frac{1}{2}\overline{X} + \frac{\sigma_X^2}{2\overline{X}} > \frac{1}{2}\overline{X}$$

This should make your parents very happy as they will see that you are doing better than the average lifetime written on the phone!

Bernoulli Distribution $X \sim Bernoulli(p)$

The Bernoulli distribution is a discrete probability distribution for a Bernoulli trial — a random experiment that has only two outcomes (usually called a “Success” or a “Failure”)

If we associate the random variable X with it as $X = 1$ for, say, Success or Heads and $X = 0$ for Failure or Tails, then the corresponding Probability Mass Function will be given as –

$P(X = 1) = p$ Probability of Success

and $P(X = 0) = 1 - p$ Probability of Failure

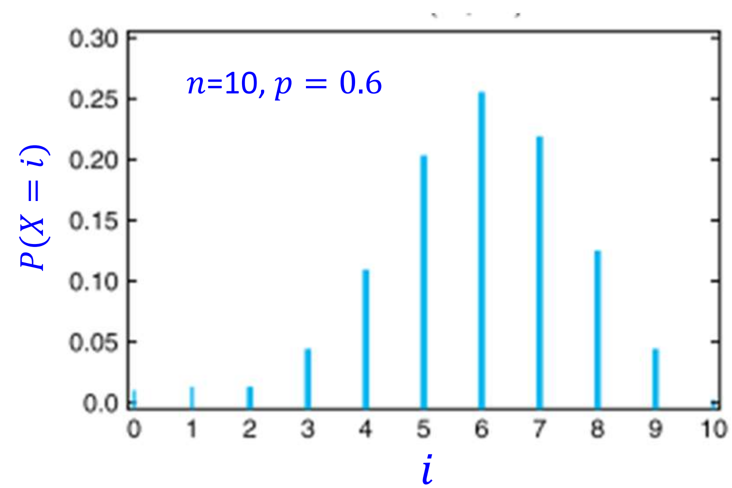
$$\bar{X} = E(X) = p$$

$$Var(X) = \sum_{x=(0,1)} (x - E(X))^2 P_X(x) = p(1-p)$$

For multiple independent Bernoulli trials (say n trials), the probability mass function will be given by the Binomial Distribution in the next slide

Binomial Distribution $X \sim Bin(n, p)$

The Binomial Distribution with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent experiments, each with its own Boolean-valued outcome: success (with probability p) or failure (with probability q=1- p).



P(i successes in n trials) =

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i=0, 1, \ldots, n$$

and $\binom{n}{i} = \frac{n!}{i!(n-i)!}$

Equivalently, we can see that $X = \sum_{i=1}^{n} X_i$

where $X_i = \begin{cases} 1 & probability\ p \\ 0 & probability\ (1-p) \end{cases}$

$E(X_i) = p, \qquad Var(X_i) = E(X_i^2) - p^2 = p(1-p)$

$E(X) \quad = \sum_{i=1}^{n} E(X_i) = np$

$Var(X) = \sum_{i=1}^{n} Var(X_i)$ since the X_i are independent

$= np(1-p)$

Geometric Distribution of Type-0 $X \sim geom_0(p)$

The Geometric Distribution of Type 0 is a type of discrete probability distribution that represents the probability of the number of successive failures before a success is obtained in a Bernoulli trial. Note that the probability of failure in a given trial is $(1-p)$.

$$P(X=x) = \begin{cases} (1-p)^x p & x = 0,1,\ldots\ldots \\ 0 & otherwise \end{cases}$$

(x +1)trials overall with x failures and one success at the end

$$E(X) = \frac{1-p}{p} \qquad Var(X) = \frac{1-p}{p^2}$$

Geometric Distribution of Type-1 $\quad Y \sim geom_1(p)$

The Geometric Distribution of Type-1 is a type of discrete probability distribution that represents the probability of the number of Bernoulli trials until first success

Therefore, in a geometric distribution, a Bernoulli trial is repeated until a success (with probability p) is obtained and then stopped. (Note that the probability of failure in a given trial is $(1-p)$.

x trials overall with $x-1$ failures and one success at the end

$$P(X = x) = \begin{cases} (1-p)^{x-1}p & x = 1,2 \ldots\ldots \\ 0 & otherwise \end{cases}$$

$$E(X) = \frac{1}{p} \qquad Var(X) = \frac{1-p}{p^2}$$

Revise your High School "tricks" on how to sum expressions like $\sum_{n=1}^{\infty} np^{n-1}$ (There are two simple ways of doing this)

$$\begin{aligned} P(X > k) &= P(X = k+1) + P\,(X = k+2) + \cdots \\ &= (1-p)^k p + (1-p)^{k+1} p + \ \ldots. \\ &= (1-p)^k \end{aligned}$$

$$\begin{aligned} F_X(k) &= P(X \le k) = 1 - P(X > k) \\ &= 1 - (1-p)^k \end{aligned}$$

Memoryless Property of a Random Variable

A random variable X is said to be Memoryless if $P(X > n + m|X > m) = P(X > n)$

i.e. "The conditional probability of X being greater than $(n + m)$, given that it is greater than m is the same as the probability of X being greater than n"

Note that, $$P(X > n + m|X > m) = \frac{P(\{X>n+m\}\cap\{X>m\})}{P(X>m)} = \frac{P(X>n+m)}{P(X>m)}$$

Show that $X \sim geom_0(p)$ is NOT a memory less distribution.

For the geometric random variable $X \sim geom_1(p)$, this implies that -

$$P(X > n + m|X > m) = \frac{(1-p)^{n+m}}{(1-p)^{m}} = (1-p)^n = P(X > n)$$

If a random variable of this type has crossed m levels, then the probability of it crossing an additional n levels is the same as its probability of crossing n levels starting from the initial state. $X \sim geom_1(p)$ is the only example of a Discrete Memoryless Distribution

Memoryless Distribution for Continuous Random Variables

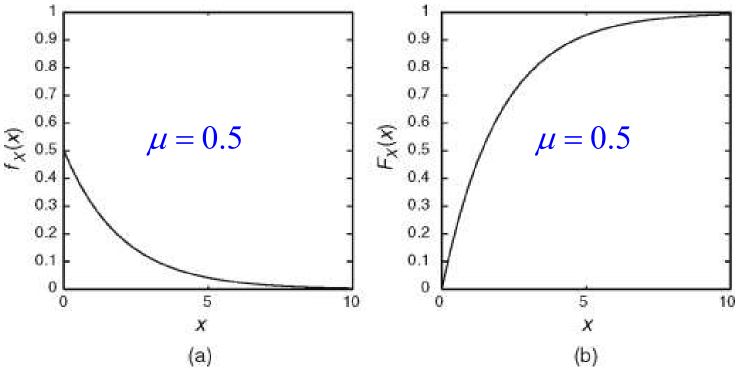
The *continuous* analogue of the discrete geometric $geom_1(p)$ distribution is the exponential distribution.

PDF $f_X(x)$ and CDF $F_X(x)$ of an exponentially distributed random variable are -

$$f_X(x) = \mu e^{-\mu x} \qquad 0 \le x < \infty$$

$$F_X(x) = 1 - e^{-\mu x} \qquad 0 \le x < \infty$$

with Mean = $1/\mu$ and Variance = $1/\mu^2$



Prove that $P(X > T + S | X > S) = P(X > T)$ in this case which shows that the exponential distribution is a *Memoryless Distribution* .

Shown in the next slide

Exponential Distribution $X \sim \exp(\mu)$ has pdf and cdf as given earlier

$$f_X(x) = \mu e^{-\mu x} \qquad 0 \le x < \infty$$
$$F_X(x) = 1 - e^{-\mu x} \qquad 0 \le x < \infty$$

with Mean = $1/\mu$ and Variance = $1/\mu^2$
and $P(X > t) = 1 - F_X(t) = e^{-\mu}$, $t > 0$

It follows that -

$$P\left(X > t+s \mid X > s\right) = \frac{P(X > t+s)}{P(X > s)}$$
$$= \frac{e^{-\mu(t+s)}}{e^{-\mu s}} = e^{-\mu t} = P(X > s)$$

The length of phone calls is commonly modelled as having an exponential distribution!

So, now you know why your brother/sister/son/daughter never seem to end their phone calls, when you also want to use the landline at home

The gap between successive cars on a highway is also modelled as having an exponential distribution.

What implication does it have for what happens when “a chicken wants to cross the road”.

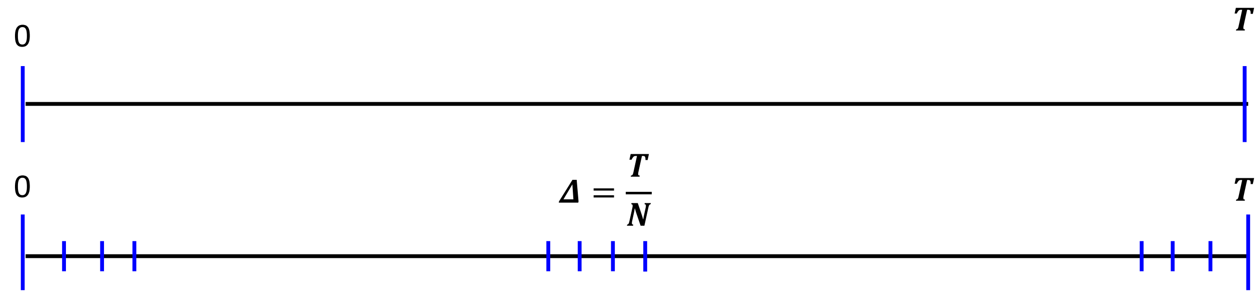
“Why should the chicken never be in a hurry to cross the road?”

Because he/she will always find a gap in the traffic which is as large as anything he/she wants and then should use that to cross the road safely!

Correspondence between the Geometric Distribution for Discrete Random Variables and the Continuous Distribution for Continuous Random Variables

Continuous
$T \sim \exp(\mu)$

Discrete
$N \sim geom_0(p)$



Consider a system which you start observing at time $t = 0$ where the event that you are observing for happens at time $t = T$.
For the Continuous Time model, let us say we observe "No event in time $(0, T)$ and then the event happening in $(T, T + dT)$ with probability $e^{-\mu T}\mu(dT)$
In the Discrete Time Model, this would be equivalent to saying that the event does not happen for N slots and then happens in the $(N + 1)^{th}$ slot. Note that the probability p of the event happening in a slot will be $p = \mu\Delta = \frac{\mu T}{N}$ while the probability of the event not happening will be (1-p). With $N \to \infty$, $\Delta \to 0$, these will be the only two things that can happen in a slot, i.e., multiple events cannot occur (their probability will tend to zero))

29

Probability of No Event in $(0, T)$ and the event happening in $(T, T + dT)$ or equivalently in the interval $(T, T + \Delta T)$

Continuous Case

$$e^{-\mu T} \mu(dT)$$

Discrete Case $p = \mu\Delta = \frac{\mu T}{N}$

$$\left(1 - \frac{\mu T}{N}\right)^N \left(\frac{\mu T}{N}\right)$$

$$N \to \infty \quad \frac{T}{N} \to \Delta T \text{ or } dT$$

$$\left(1 - \frac{\mu T}{N}\right)^N \to e^{-\mu T}$$

Poisson Random Variable $X \sim Poisson(\lambda)$

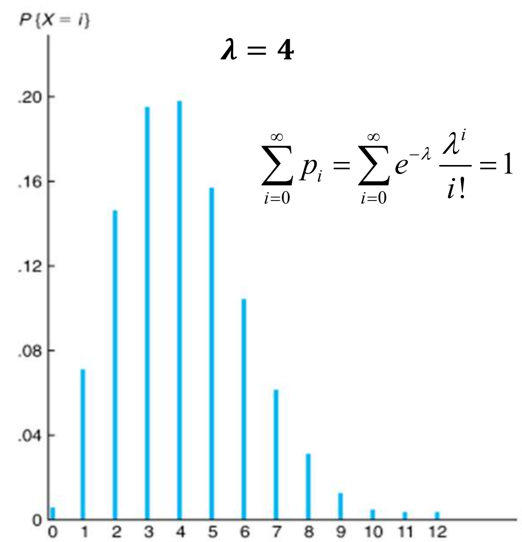
A random variable X, taking on one of the values 0, 1, 2,…, is said to be a Poisson random variable with parameter λ, $\lambda > 0$, if its probability mass function is given by –

$$\boldsymbol{X \sim Poisson(\lambda)} \qquad P(X = i) = p_i = e^{-\lambda}\frac{\lambda^i}{i!}, \qquad i = 0, 1, \dots \dots . \infty$$

Mean: $\bar{X} = \sum_{i=0}^{\infty} i p_i = \lambda$

Second Moment: $\overline{X^2} = \sum_{i=0}^{\infty} i^2 p_i = \lambda^2 + \lambda$

Variance: $\sigma_X^2 = \overline{X^2} - (\bar{X})^2 = \lambda$



The Poisson distribution is very popular in analytical modelling and simulations as it is *described by just one variable* λ.

Properties of the Poisson Distribution

Homogeneity

The arrival rate λ is constant with respect to time. The expected number of arrivals in any given interval of time Δt is $\lambda \Delta t$.
(*Weak Stationarity* also holds ⇒ Mean and Variance does not change with time)

Independence

The number of arrivals in disjoint intervals are independent of each other.
The number of arrivals in one interval will not have any effect on the number of arrivals in any other disjoint interval

We show in the next slide that these properties are enough to prove that the arrival process for this will have the Poisson distribution

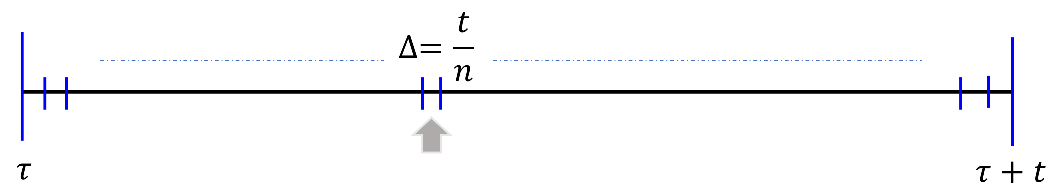
Let N_t be the number of arrivals in $[\tau, \tau + t]$, τ>0.   Homogeneity $\Rightarrow E(N_t)= \lambda t$

Divide t into n non-overlapping intervals with $n \to \infty$ and let M_j be the number of arrivals (0 or 1; 1 with probability p_j) arriving in the j^{th} interval.

We can also conclude that for any j,

$$P(M_j = 1) = \lambda\left(\frac{t}{n}\right), \quad P(M_j = 0) = 1 - \lambda\left(\frac{t}{n}\right)$$

and $P(M_j > 1) \to 0$ as $\frac{1}{n^2}$ and higher powers of n and can be ignored as $n \to \infty$



P(one arrival)=$\lambda\left(\frac{t}{n}\right)$

P(no arrival)=1-$\lambda\left(\frac{t}{n}\right)$

$P(k$ arrivals in $[\tau, \tau + t])$ =P_k

= Probability that any k of the n slots have an arrival and there are no arrivals in the other $(n-k)$ slots

$$= \frac{n!}{k!(n-k)!}\left[\lambda\left(\frac{t}{n}\right)\right]^k \left[1-\lambda\left(\frac{t}{n}\right)\right]^{n-k}$$

Poisson Distribution

$$\to \quad e^{-\lambda t}\frac{(\lambda t)^n}{n!} \quad \text{as } n \to \infty$$

See next slide for the steps, if needed

Binomial → Poisson

The Poisson Distribution may be used as an approximation to the Binomial Distribution with parameters (n, p) when n is large and p is small and $\lambda = np$ remains the mean of both the distributions

$$P(X=i) = \frac{n!}{(n-i)!\,i!} p^i (1-p)^{n-i}$$

$$= \frac{n!}{(n-i)!\,i!} \left(\frac{\lambda}{n}\right)^i \left(1-\frac{\lambda}{n}\right)^{n-i}$$

$$= \frac{n(n-1).....(n-i+1)}{n^i} \times \frac{\lambda^i}{i!} \times \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^i}$$

$$\rightarrow \quad e^{-\lambda} \frac{\lambda^i}{i!} \quad as\ n \rightarrow \infty$$

For large n and small p, we have -

$$\left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda}$$

$$\frac{n(n-1)......(n-i+1)}{n^i} \approx 1$$

$$\left(1-\frac{\lambda}{n}\right)^i \approx 1$$

These were also the approximations needed in the previous slide

Another useful property of the Poisson Distribution

The sum of **Independent** Poisson random variables is also a Poisson random variable

Let $X \sim Poisson(\lambda_1)$ and $Y \sim Poisson(\lambda_2)$ be two Independent Poisson Random Variables, i.e., $X \perp Y$

Then $X + Y \sim Poisson(\lambda_1 + \lambda_2)$

The Sum of Independent Poisson Random Variables is also a Poisson Random Variable

Proof: Taken from https://llc.stat.purdue.edu/2014/41600/notes/prob1805.pdf

Phew!!!

Remind me to show you later how a little bit of clever thinking will let you show this in about two and a half lines!!

Sums of independent Poisson random variables are Poisson random variables. Let X and Y be independent Poisson random variables with parameters λ_1 and λ_2, respectively.

Define $\lambda = \lambda_1 + \lambda_2$ and $Z = X + Y$. Claim that Z is a Poisson random variable with parameter λ. Why?

$$
\begin{aligned}
p_Z(z) &= P(Z = z) \\
&= \sum_{j=0}^{z} P(X = j \;\&\; Y = z - j) && \text{so } X + Y = z \\
&= \sum_{j=0}^{z} P(X = j)P(Y = z - j) && \text{since } X \text{ and } Y \text{ are independent} \\
&= \sum_{j=0}^{z} \frac{e^{-\lambda_1}\lambda_1^j}{j!} \frac{e^{-\lambda_2}\lambda_2^{z-j}}{(z-j)!} \\
&= \sum_{j=0}^{z} \frac{1}{j!(z-j)!} e^{-\lambda_1}\lambda_1^j e^{-\lambda_2}\lambda_2^{z-j} \\
&= \sum_{j=0}^{z} \frac{z!}{j!(z-j)!} \frac{e^{-\lambda_1}\lambda_1^j e^{-\lambda_2}\lambda_2^{z-j}}{z!} && \text{multiply and divide by } z! \\
&= \sum_{j=0}^{z} \binom{z}{j} \frac{e^{-\lambda_1}\lambda_1^j e^{-\lambda_2}\lambda_2^{z-j}}{z!} && \text{using the form of binominal coefficients} \\
&= \frac{e^{-\lambda}}{z!} \sum_{j=0}^{z} \binom{z}{j} \lambda_1^j \lambda_2^{z-j} && \text{factoring out } z! \text{ and } e^{-\lambda_1}e^{-\lambda_2} = e^{-\lambda_1-\lambda_2} = e^{-\lambda} \\
&= \frac{e^{-\lambda}}{z!}(\lambda_1 + \lambda_2)^z && \text{using binomial expansion (in reverse)} \\
&= \frac{e^{-\lambda}\lambda^z}{z!}
\end{aligned}
$$

So altogether we showed that $p_Z(z) = \frac{e^{-\lambda}\lambda^z}{z!}$. So $Z = X + Y$ is Poisson, and we just sum the parameters.

36

We saw that the sum of two **Independent** Poisson random variables is also a Poisson random variable

Note that this result is not limited to the sum of just two independent random variables!

We can obviously extend this to N independent random variables for $N \geq 2$

Uniform Distribution (Discrete) $X{\sim}Unif(f[1,m])$

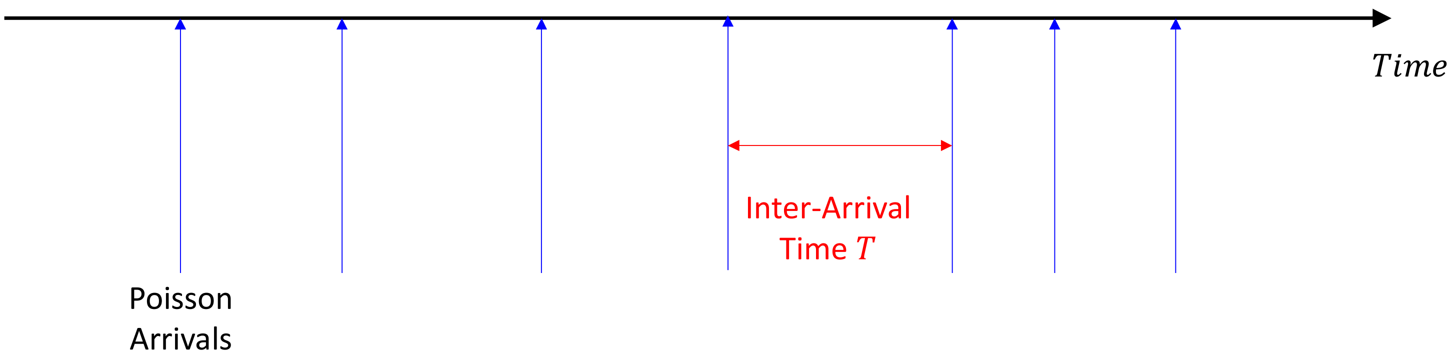
$$p_i \text{ or } p_X(i) = P(X = i) = \frac{1}{m}, \forall i \in [1, m]$$

$$E(X) = \overline{X} = \sum_{x=1}^{m} x\left(\frac{1}{m}\right) = \frac{(m+1)}{2}$$

$$Var(X) = \overline{X^2} - \left(\overline{X}\right)^2 = \frac{m^2 - 1}{12}$$

Roll a six-sided dice to get one of {1, 2, 3, 4, 5, 6} each with probability $\frac{1}{m} = \frac{1}{6}$

Poisson Arrivals have Exponentially Distributed Inter-Arrival Times



Poisson Arrivals come from a Poisson Process with rate λ

$\Rightarrow \quad P(k\ arrivals\ in\ a\ time\ interval\ of\ length\ T)$
$$= e^{-\lambda T}\frac{(\lambda T)^k}{k!} \quad \text{for } k = 0, 1,2, \ldots. \infty$$

$$k \sim Poisson(\lambda T)$$

The Inter-Arrival Times are independent, exponentially distributed random variables which have identical distributions with mean $1/\lambda$

$$\Rightarrow \quad f_T(t) = \lambda e^{-\lambda t} \quad 0 \le t < \infty$$

$$T \sim \exp(\lambda)$$

If phone calls are exponentially distributed in length, then they are also assumed to be coming from a Poisson process

39

Compound Probability Distribution

Consider the random variable Y defined as

$$Y = X_1 + X_2 + \cdots + X_N$$

where –

(i) N is a random number

(ii) $X_i,\ i = 1,2,\ldots,N$ are independent, identically distributed (i.i.d.) random variables with c.d.f. F_x, mean μ_X and variance σ_X^2

(iii) Each X_i is independent of N $\quad$ N is a discrete r.v. with mean μ_N and variance σ_N^2

Using the Law of Total Probability, the *Compounded Distribution* of Y is given as -

$$F_Y(y) = P(Y = y) = \sum_{n=0}^{\infty} P\left(X_1 + X_2 + \ldots.. + X_N = y \mid N = n\right) P(N = n) = \sum_{n=0}^{\infty} F_Y^{(n)} P(N = n)$$

where $F_Y^{(n)}$ is the *n-fold convolution* of F_{X_i}

$$F_Y^{(n)} = F_{X_1} * F_{X_2} * \cdots * F_{X_N}$$

$$Z = X + Y \quad X \perp Y$$

$$P(Z = z) = \sum_{k=-\infty}^{\infty} P(X = k) P(Y = z - k)$$

Consider the first two moments of a random variable with this compound distribution

$Y = X_1 + X_2 + \cdots + X_N$ $\quad X_i s$ are i.i.d.

$$\begin{aligned}E(Y) &= E_N\left(E_Y(Y \mid N)\right) \\ &= \sum_{n=0}^{\infty} E(Y \mid N = n) P(N = n) \\ &= \sum_{n=0}^{\infty} n E(X) P(N = n) = \mu_X \sum_{n=0}^{\infty} n P(N = n) \\ &= \mu_X \mu_N\end{aligned}$$

From Eq. 2.19 "Law of Total Variance"

$$Var(Y) = E_X\left[Var_Y\left(Y \mid X\right)\right] + Var_X\left(E_Y\left(Y \mid X\right)\right)$$

$$\begin{aligned}Var(Y) &= E_N\left(Var_Y(Y \mid N)\right) + Var_N\left(E_Y(Y \mid N)\right) \\ &= E_N\left(N Var_X(X)\right) + Var_N\left(N E_X(X)\right) \\ &= \mu_N Var_X(x) + Var_N\left(N \mu_X\right) \\ &= \mu_N \sigma_X^2 + \left(\mu_X\right)^2 Var_N(N) \\ &= \mu_N \sigma_X^2 + \mu_X^2 \sigma_N^2\end{aligned}$$

Negative Binomial (Discrete) Distribution $X \sim NB(k; r, p)$
also known as the Pascal Distribution $X \sim Pa(k; r, p)$

The negative binomial experiment is almost the same as a binomial experiment with one difference: a binomial experiment has a fixed number of trials but the number of trials is not fixed in the negative binomial case.

Recall that if the following five conditions are true, then the experiment is **binomial**:

1. Fixed number of n trials **2.** Each trial is independent **3.** Only outcomes are Success/Failure

4. Probability of Success (p) for each trial is constant **5.** Random variable X = the number of successes.

The **negative binomial** is similar to the binomial with two differences (specifically to numbers 1 and 5 in the list above):

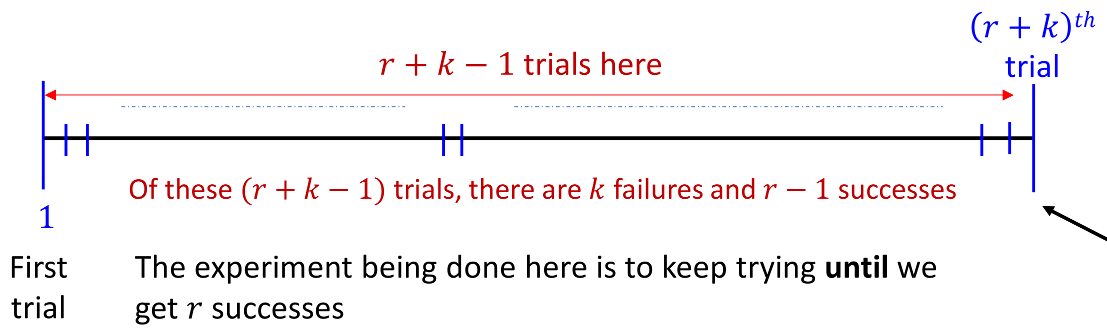
•The number of trials, n **is not fixed**.

•Random variable X differenty defined (**see subsequent slides**)

See graphical description in the next slide

Negative Binomial (Discrete) Distribution $X \sim NB(k; r, p)$

Following the notation in the online notes provided to you



$P(success) = p$
$P(failure) = 1 - p$

The last one at $r + k$ is a success; it is the r^{th} success

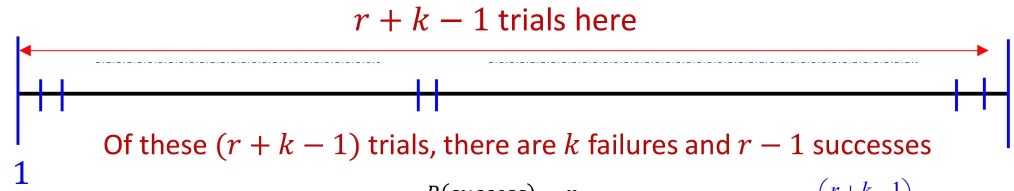
The experiment being done here is to keep trying **until** we get r successes

If we define our random variable X to be the number of failures that we will encounter in that case, then -

$$P(X = k) = \binom{r+k-1}{k}(1-p)^k p^r$$

Read the example on page 58 of the lecture material given for this module with this figure in front of you

Let’s try to think a little differently!



$P(success) = p$
$P(failure) = 1 - p$

$$P(X=k) = \binom{r+k-1}{k}(1-p)^k p^r$$

Y: The number of trials for r successes with the **last trial being a success**

Using the results of the previous slide, we can easily conclude that -

$$\boxed{P(Y=n) = \binom{n-1}{n-r}(1-p)^{n-r} p^r \quad n = r, r+1, r+2....\infty}$$

This is the probability distribution (actually, the Probability Mass Function) for the number of trials n needed for r successes when the **last trial is a success** and the probability of success in any one trial is p.

Example of Page 50

P(Pen wins a rally)=p=0.6 P(Hart wins a rally)=h=0.4

Winning Patterns for Li Pen

20+0+1 $\binom{20}{20}(0.6)^{20}(0.4)^{0}(0.6)$ Li Pen winning in 21 rallies, 20 of the first 20 and then one more, 21/0

20+1+1 $\binom{21}{20}(0.6)^{20}(0.4)^{1}(0.6)$

20+2+1 $\binom{22}{20}(0.6)^{20}(0.4)^{2}(0.6)$

....

....

20+5+1 $\binom{25}{20}(0.6)^{20}(0.4)^{5}(0.6)$ Li Pen winning in 26 rallies, 20 of the first 25 and then one more, 21/5

....

....

20+19+1 $\binom{39}{20}(0.6)^{20}(0.4)^{19}(0.6)$

20+20+1 $\binom{40}{20}(0.6)^{20}(0.4)^{20}(0.6)$ Li Pen winning in 41 rallies, 20 of the first 40 and then one more, 41/20

Deuce not considered in this model

Example of Page 50

No Deuce in this model!

How to calculate the probabilities of Li Pen winning with scores like 22-20, 23-21, 24-22….?

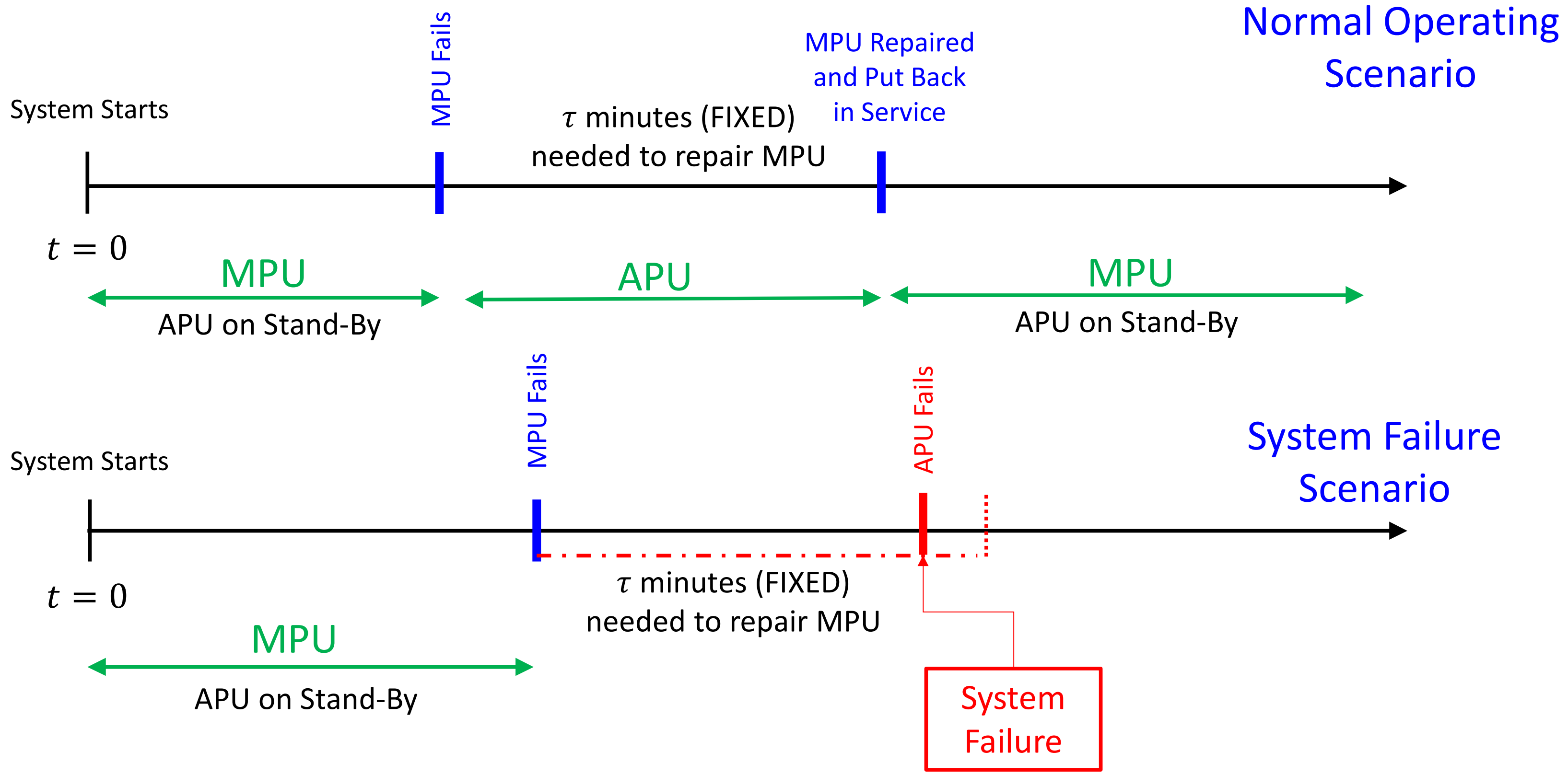
If you deuce at 29-29 then the game would end at 30-29.

Example (page60): System Failure because of the failure of both Main Power Unit (MPU) and Auxiliary Power Unit (APU)

Assume that the time to failure of both are independent exponentially distributed (i.i.d.) random variables (r.v.s) with mean $1/\mu$ minutes.
The mode of operation followed is described below.

"The system is started with the MPU. When the MPU fails, we immediately move to the APU while the MPU is being repaired. We need τ minutes (FIXED) to repair the MPU and put it back in service. If the APU fails before the MPU is fixed, then the system fails. If we can repair the MPU before the APU fails, then the system resumes normal operation as before. In that case, there is no system failure until the next time the failure sequence repeats itself."

Operation Scenarios



48

X: Time (random) to First System Failure, τ: Time (Fixed) to repair failed MPU and put it back in service
L_M: Operating Time of MPU, L_A: Operating time of APU Both are *i.i.d.* exponentially distributed with mean $1/\mu$

$$E(X) = E\left(X \mid L_A \le \tau\right) P\left(L_A \le \tau\right) + E\left(X \mid L_A > \tau\right) P\left(L_A > \tau\right)$$

System does not fail when $L_A > \tau$

$$E\left(X \mid L_A > \tau\right) = E\left(L_M + \tau + X\right) = \frac{1}{\mu} + \tau + E(X)$$

Why?

System Fails when $L_A \le \tau$

$$E\left(X \mid L_A \le \tau\right) = \frac{1}{\mu} + E\left(L_A \mid L_A \le \tau\right) = \frac{1}{\mu} + \frac{\int_0^\tau (x)\left(\mu e^{-\mu x}\right) dx}{P\left(L_A \le \tau\right)} = \frac{1}{\mu} + \frac{1 - e^{-\mu\tau} - \mu\tau e^{-\mu\tau}}{\mu(1 - e^{-\mu\tau})}$$

Note that

$$P\left(L_A \le \tau\right) = \int_0^\tau \mu e^{-\mu x} dx = \left(1 - e^{-\mu\tau}\right)$$

Putting everything together -

$$E(X) = e^{-\mu\tau} E\left(X \mid L_A > \tau\right) + \left(1 - e^{-\mu\tau}\right) E\left(X \mid L_A \le \tau\right) = \frac{1}{\mu} + e^{-\mu\tau}\left[\tau + E(X)\right] + \frac{1 - e^{-\mu\tau} - \mu\tau e^{-\mu\tau}}{\mu}$$

$$\Rightarrow \boxed{E(X) = \frac{2 - e^{-\mu\tau}}{\mu(1 - e^{-\mu\tau})}}$$

Check this to see what happens when you do $\tau \to 0$ and $\tau \to \infty$

49

The actual problem in the book is a lot simpler!

- It is stated that "The exponential model is often used as the probability model for the *time until a rare event*"
- Also, if the random variable X is the time until the first system failure, then under fairly general conditions, $P(X > t) \approx e^{-\frac{t}{E(X)}}$ holds
- In this problem, you are given that $E(X) = 500$ hours.
- Therefore, the probability of failure after 100 hours is $P(X > 100) \approx e^{-\frac{100}{500}} = 0.8187$