Lecture : Other Useful Distributions

Last time:

- Normal distribution, continuous random variables
- Standardization and calculations using the Normal distribution

Today:

- Bernoulli Distribution, Binomial Distribution, Geometric Distribution
- Exponential Distribution
- Poisson Process
- Lognormal Distribution

Columbia University
IN THE CITY OF NEW YORK

Bernoulli Distribution

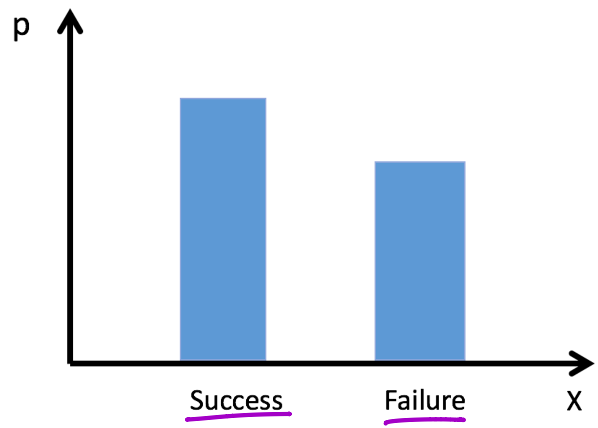
- Discrete distribution with two possible outcomes

- $$X = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } (1-p) \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1-p & \text{if } 0 \leqslant x < 1 \\ 1 & \text{if } x \geqslant 1 \end{cases}$$

$$E(X) = p$$

$$Var(X) = p(1-p)$$



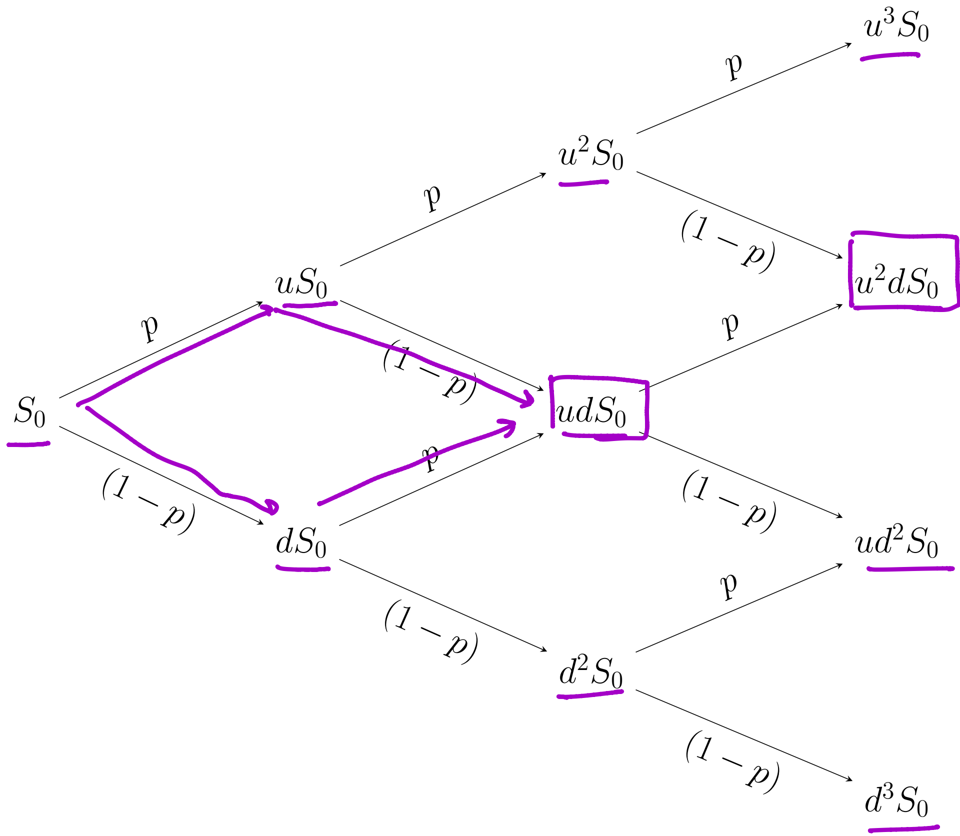
- **Examples**
  - probability of click in Display advertising
  - probability of stock price going up or down in a period

Bernoulli Distribution

- Building block for other richer discrete distributions
  - **Binomial Distribution** - number of successes in n trials
    (e.g. probability of k clicks out of n ads displayed)
  - **Geometric Distribution** - number of failures before the first success
  - **Negative Binomial Distribution** - number of failures before the x_{th} success

Binomial Distribution

- **Example: Binomial Option Pricing Model**



$\binom{3}{2} = 3$

$3p^2(1-p)$

Binomial Distribution

- k success in n independent trials

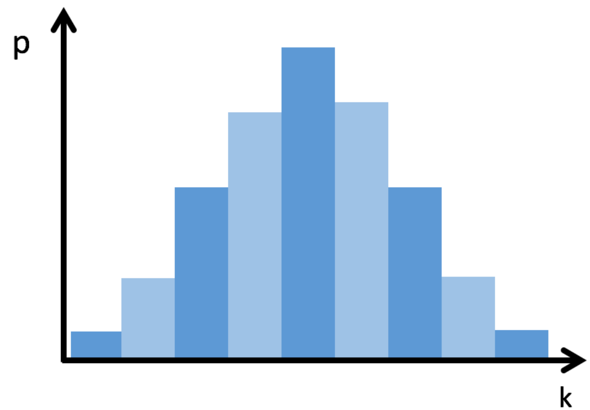
$$\text{Per trial} \begin{cases} \text{success (e.g. purchase) with probability } p \\ \text{failure (e.g. no purchase) with probability } 1-p \end{cases}$$

-

$$p_X(k) = Pr(\text{k success in n trials})$$
$$= \binom{n}{k} p^k (1-p)^{n-k}$$

$$E(X) = np$$

$$Var(X) = np(1-p)$$



$$E[X] = \sum_{k=0}^{n} P(X=k)\, k$$

$$Var[X] = E[X^2] - (E[X])^2$$

Binomial Distribution

Consider a random variable X distributed according to Binomial(n, p)

$$X = Y_1 + Y_2 + \cdots Y_n$$

$$Y_j = \begin{cases} 1 & \text{w/prob } p \\ 0 & 1-p \end{cases}$$

$$\mathbb{E}[X] = \sum_{j=1}^{n} \mathbb{E}[Y_j]$$

$$= np$$

$$\mathbb{E}[Y_j] = p \quad \forall j$$

$$\mathrm{Var}[Y_j] = p(1-p)$$

$$\mathrm{Var}[X] = \sum_{j=1}^{n} \mathrm{Var}[Y_j]$$

$$= np(1-p)$$

Binomial Distribution

Approximation: If n is large enough, $N(\mu, \sigma^2)$ is a good approximation for $B(n, p)$

- $\mu = np$
- $\sigma^2 = np(1-p)$

$$X = Y_1 + Y_2 + \cdots Y_n$$

Central limit theorem

(CLT)

Geometric Distribution

$X = 0, 1, 2 \dots$

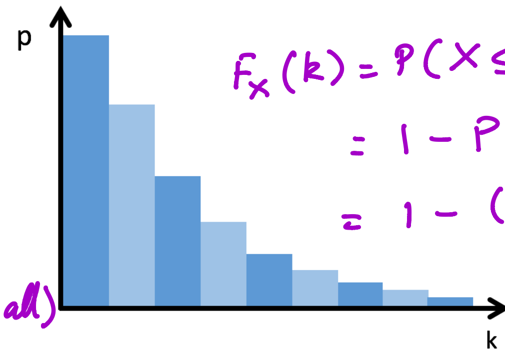
- Number of trials until first success

- $$p_X(k) = p(1-p)^{k-1}$$

  $$F_X(k) = 1 - (1-p)^k$$

  $$E(X) = \frac{1}{p}$$

  $$Var(X) = \frac{1-p}{p^2} \sim \frac{1}{p^2} \quad (p \text{ small})$$



$F_X(k) = P(X \leq k)$
$= 1 - P(X > k)$
$= 1 - (1-p)^k$

- **Example**: A certain basketball player has a 60% chance of making a free throw. Assume all free throws are independent. What is the probability that he makes his first free throw on the 3^{rd} try?

$p = 0.6$ $\quad P(X=3) = (1-p)^2 p = (0.16)(0.6) = 0.096$

Exponential Distribution

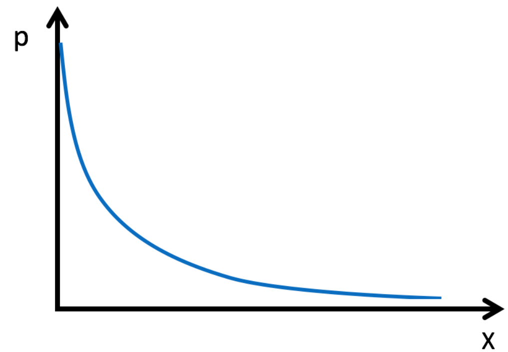
$$f_X(x) = \lambda e^{-\lambda x} \quad x \geqslant 0$$

(λ)

$$P(X \leq x) = F_X(x) = 1 - e^{-\lambda x} \quad x \geqslant 0$$

$$E(X) = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2} \quad , \quad \sigma(X) = \frac{1}{\lambda}$$



$$\mathbb{E}[X] = \int_{x=0}^{\infty} x \, f(x) \, dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

$$Var[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Exponential distribution: Properties

- Exponential distribution is the continuous analogue of the geometric distribution

$X \sim Exp(\lambda)$

$P(X > k) = 1 - F(k) = \underbrace{e^{-\lambda k}}$

$Y \sim Geom(p)$

$P(Y > k) = \underbrace{(1-p)^k}$

$$P(Z > t) = (1 - \lambda\delta)^{t/\delta} \xrightarrow{\lim \delta \to 0} \underline{\underline{e^{-\lambda t}}}$$

$\to \delta \leftarrow$

Z : denotes the stopping time

$(\lambda\delta)$ success $\longrightarrow$ stop

$(1 - \lambda\delta)$ failure $\longrightarrow$ continue

$T \sim Exp(\lambda)$

- **Memoryless** - $P(T > s + t | T > s) = P(T > t))$

$$P(T > s+t \mid T > s) = P(T > t)$$

$$= \frac{P(T > s+t \text{ and } T > s)}{P(T > s)} = e^{-\lambda t} = P(T > t)$$

Example of Exponential Distribution

- **Example**: On average number of people arriving at the bus station in an hour is 3. Probability the time till the next person arrive is less than one hour is:

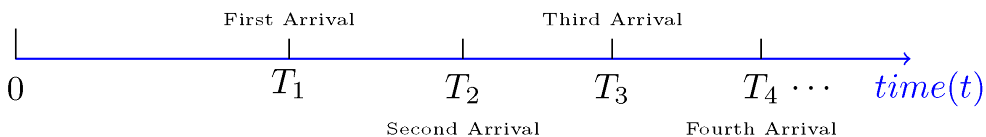
  $X \sim Exp(\lambda)$ $(\lambda = 3)$

  $E[X] = \frac{1}{3} = \frac{1}{\lambda}$

  $$F_X(1) = P(X \leqslant 1) = 1 - e^{-\lambda} = 1 - \frac{1}{e^3}$$

- **Call Center**
  Calls arrive at call center an average rate λ per hour. Customers wait in the queue until one of two things happen: an agent is allocated to serve them (through supporting software), or they become impatient and abandon the tele-queue. Service time and customer patience (time to abandonment) are both exponentially distributed.

- **Poisson Process**



Poisson Distribution

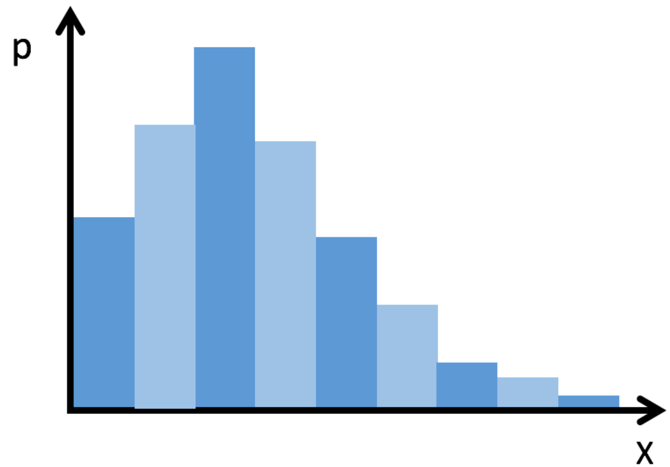
- Probability of a given number of events occurring in a fixed interval of time and/or space
-

$$p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

(λ)

$$E(X) = \lambda$$

$$Var(X) = \lambda^{2}$$



Example of Poisson Distribution

- **Example**: Which of the following is most likely to be well modeled by a Poisson distribution?
  1. Number of trains arriving at station every hour ✓
  2. Number of lottery winners each year that live in Manhattan
  3. Number of days between solar eclipses
  4. Number of days until a component fails

- **Example**: the mean number of people arriving per hour at a shopping center is 18.What is the probability that the number of customers arriving in an hour is 20.

$\lambda = 18$

$$P(20) = e^{-18}\frac{18^{20}}{20!}$$

$$\left(e^{-\lambda}\,\frac{\lambda^{20}}{20!}\right)$$

Poisson Process

- The counting process $\{N(t), t > 0\}$ with rates λ, $\lambda > 0$,

$$P\{N(t) = n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

$N(t) \sim Poisson(\lambda t)$

$\mathbb{E}[N(t)] = \lambda t$

- $$E[N(t)] = \lambda t$$

- The inter-arrival times X_1, X_2, ... are independent and $X_i \sim Exponential(\lambda)$

$N(0) = 0$

$P(N(t) = 0) = e^{-\lambda t}$

$P(X_1 > t) = e^{-\lambda t} \Rightarrow X_1 \sim Exp(\lambda)$

Example of Poisson Process

- **Traffic Model**

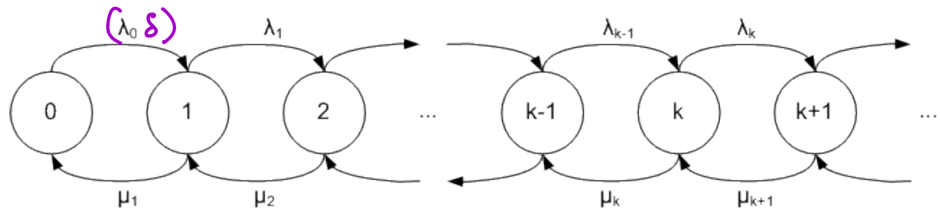
  Suppose the time between arrival of buses at the student center is exponentially distributed with a mean of 60 minutes. If we arrive at the student center at a randomly chosen instant, what is the average amount of time that we will have to wait for a bus?

  $\lambda = \frac{1}{60}$

  avg wait time = 60 minutes

  The Waiting Time Paradox: The memoryless property of the exponential distribution implies that whatever the time at which we arrive , the mean waiting time is the 60 min.

- **Birth and Death Process**



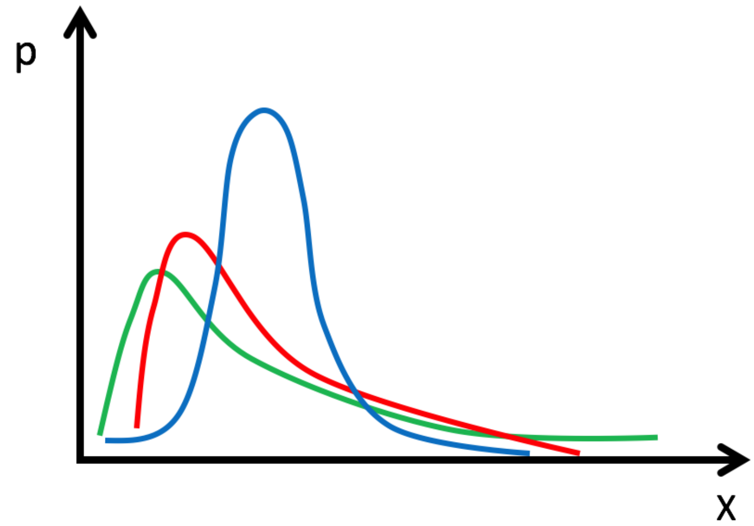
Lognormal Distribution

- If $ln(x)$ is normally distributed, x is lognormally distributed.

- $ln(X) \sim N(\mu, \sigma^2)$ $(x > 0)$

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(ln(x)-\mu)^2}{2\sigma^2}}$$

$$F(x) = \Phi\left(\frac{ln(x)-\mu}{\sigma}\right)$$



- Consequence of CLT on the logarithm of product of independent random variables
- Arises in many natural phenomenon. For instance:
  - Biological processes: size of a living tissue, blood pressure in adult human
  - Epidemic or rumor spreading: number of affected nodes

$Z = X_1 \cdot X_2 \cdots X_n$ $\qquad \log Z = \underbrace{\sum_{j=1}^{n} \log X_j}$