Probability Recap: CRV

Klejdi Sevdari

May 2021

LN12: Continuous Random Variables (CRV)

- Definition: A random variable X is continuous if its distribution function $F(x) = \int_{-\infty}^{x} f(u)du$.
- PDF: Every function f that satisfies $\int_{-\infty}^{\infty} f(u)du = 1$ and $f(x) \geq 0 \ \forall x$ is called a probability density function.
- Basic properties: Let X be a CRV with PDF $f(x)$. Then:
  - $\int_{-\infty}^{\infty} f(u)du = \lim_{x \to \infty} F(x) = 1$
  - $P(X = x) = 0 = F(x) - F(x^-)$ *(left limit)*
  - $P(a \leq X \leq b) = \int_{a}^{b} f(u)du = F(b) - F(a)$
- Analysis 1: $\frac{d}{dx}F(x) = f(x)$ if F differentiable at x. if f continuous in $[a, b]$ then F is differentiable in (a, b).

Independence of CRV

- Definition: $X \perp\!\!\!\perp Y$ if $\forall x, y \ \ P((X \leq x) \cap (Y \leq y)) = P(X \leq x)P(Y \leq y)$.
- Borel sets: If $X \perp\!\!\!\perp Y$ and A, B are two borel sets: $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$.
- Transformations: if $X \perp\!\!\!\perp Y$ then $h(X) \perp\!\!\!\perp g(Y)$.
- Several variables: $X_1 \perp\!\!\!\perp X_2 \ldots \perp\!\!\!\perp X_n$ if $\forall x_1, \ldots x_n \ \ P(X_1 \leq x_1, X_2 \leq x_2, \ldots X_n \leq x_n) = \Pi_i P(X_i \leq x_i)$. The same would hold for Borel sets.

Transformations of CRV

Distribution Functions

- Base case: X is a CRV, $g : \mathbb{R} \to \mathbb{R}$, and $Y = g(X)$ (Y is not necessarily continuous). Then:
$$F_y(y) = P(g(X) \leq y) = \int_{\{x:\, g(X) \leq y\}} f_x(u)du$$
- g strictly increasing and continuous:
$$F_y(y) = P(g(X) \leq y) = P(x \leq g^{-1}(y)) = F_x(g^{-1}(y))$$
- g strictly increasing and continuous:
$$F_y(y) = P(g(X) \leq y) = P(x \geq g^{-1}(y)) = 1 - F_x(g^{-1}(y))$$
- g strictly monotonic but not continuous: The same equations above would still hold if we substitute g^{-1} with the generalized inverse functions:
  - g increasing: $g^{-1}(y) = inf\{x : g(x) \geq y\}$
  - g increasing: $g^{-1}(y) = inf\{x : g(x) \leq y\}$

PDF

- Base case: X is a CRV with PDF f_x, $g : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function with $g'(x) > 0\ \forall x$ or $g'(x) < 0\ \forall x$. Then $Y = g(X)$ is a CRV with pdf:

$$f_y(y) = f_x\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|$$

- Restriction: If X takes values in an open interval Um g is continuously differentiable with $g(x) > 0$ or $g(x) < 0 \forall\ x \in U$, $V = g(U)$, then:

$$f_y(y) = f_x\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right| I_v(y)$$

LN13: Expectation and Variance

Expectation

- Definition: $E(X) = \int_{-\infty}^{\infty} xf(x)dx$, provided $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$ (converges).
- Behaviour of PDF and CDF at the tails: The existence of the Expectation puts the following constraints on the behaviour of f(x) and F(x):

$$\int_{-\infty}^{\infty} |x|f(x)dx < \infty \implies \lim_{x\to\infty}\int_{x}^{\infty} |u|f(u)du = 0,\ \lim_{x\to-\infty}\int_{-\infty}^{x} |u|f(u)du = 0$$

$$\int_{-\infty}^{\infty} |x|f(x)dx < \infty \implies F(x) = o\left(\frac{1}{|x|}\right) x \to -\infty,\ 1 - F(x) = o\left(\frac{1}{|x|}\right) x \to \infty$$

- The converse implication: If $\int_0^\infty 1 - F(x) < \infty$ and $\int_{-\infty}^0 F(x) < \infty$ then $\int_{-\infty}^{\infty} |x|f(x) < \infty$.
- Expectation in terms of distribution: $E(X) = -\int_{-\infty}^0 F(x) + \int_0^\infty 1 - F(x)$.
- Transformations: if $Y = g(X)$, then $E(Y) = \int_{-\infty}^{\infty} g(x)f_x(x)dx$.
  $E(aX + b) = aE(X) + b$ (follows from above).

Moments and Variance

- The k^{th} moment of X: $m_k = E(X^k) = \int_{-\infty}^{\infty} x^k f_x(x)dx$, provided $\int_{-\infty}^{\infty} |x|^k f_x(x)dx < \infty$
- Lower moments: if m_k exists finite, then $\forall j :\ 1 \le j \le k$, m_j exists finite.
- k^{th} **centred** moment: $\sigma_k = E((X - u)^k)$.
- **Variance**: $Var(X) = \sigma_2 = E((X - u)^2) = \int_{-\infty}^{\infty}(x - u)^2 f_x(x)dx$, provided $\int_{-\infty}^{\infty}(x - u)^2 f_x(x)dx < \infty$
- Properties of Variance:
  - $V(X) = E(X^2) - E(X)^2$
  - $V(aX + b) = a^2V(X)$
- The standard deviation: $\sigma = \sqrt{\sigma^2} = \sqrt{Var(X)}$
- Non-Zero Variance: $Var(X) \neq 0$ if X is a CRV.

LN14: Noteworthy Distributions

Uniform Distribution

$$f(x) = \frac{1}{b-a} I_{[a,b]}(x), \quad F(x) = \frac{x-a}{b-a} \quad \forall x : a \leq x \leq b$$

$$E(X) = \frac{b+a}{2}, \quad V(X) = \frac{(b-a)^2}{12}$$

..

Exponential Distribution

$$f(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x), \quad F(x) = \left(1 - e^{\lambda x}\right) I_{(0,\infty)}(x)$$

$$E(X) = \lambda, \; Var(X) = \frac{1}{\lambda^2}$$

- Lack of memory property: $P(X > x + z \mid X > z) = P(X > x)$
- Usage: The Exponential Distribution is generally used for waiting times, when time is measured continuously. The Exponential is the limit of a geometric distribution.
- Connection to Poisson: If the number of events in an interval of length t has $Poisson(\lambda t)$. The probability that no events occur in this interval is $P(N = 0) = e^{-\lambda t}$, which is exponential.

Normal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), \; E(X) = \mu, \; V(X) = \sigma^2$$

- Symmetry of PDF (φ): $\varphi(\mu + x) = \varphi(\mu - x)$.
- Symmetry of CDF (ϕ): $\phi(x, 0, 1) = 1 - \phi(-x, 0, 1), \; \phi(\mu + x, \mu, \sigma^2) = 1 - \phi(\mu - x, \mu, \sigma^2)$.
- Standard Normal Distribution: It is the normal distribution with parameters $\mu = 0, \; \sigma^2 = 1$
- Stable under Linear Transformations: $X \sim N(\mu, \sigma^2)$, then $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$
- Standardization: if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.
- Using the Normal table: Say $X \sim N(\mu, \sigma^2)$, then $P(X \leq x) = P\left(\frac{X-u}{\sigma} \leq \frac{x-u}{\sigma}\right) = \phi(\frac{x-u}{\sigma}, 0, 1)$.

Gamma Distribution

$$X \sim \Gamma(k, \lambda) \implies f(x) = \frac{\lambda^k}{\Gamma(k)} e^{-\lambda x} x^{k-1} I_{(0,\infty}(x), \; E(X) = \frac{k}{\lambda}, \; Var(X) = \frac{k}{\lambda^2}$$

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$$

- Properties of Gamma function:
  - $\Gamma(1) = 1$
  - $\Gamma(x) = (x-1)\Gamma(x-1)$
  - $\Gamma(n) = (n-1)!$
  - $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
- λ is a scale parameter. If $X \sim \Gamma(k, \lambda)$, then $Y = \lambda X \sim \Gamma(k, 1)$.
- Sum of Exponentials: if $X_1, X_2, \ldots X_k$ are independent identically distributed random variables with $\xi xp(\lambda)$ distribution, then $\sum_i X_i \sim \Gamma(k, n)$.
- Relation with Normal Distribution: if $Z \sim N(0, 1)$, then $Z^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$

Cauchy Distribution

$$f(x) = \frac{1}{\pi(1+x^2)} \quad F(x) = \frac{1}{2} + \frac{1}{\pi} arctan(x)$$

The Cauchy Distribution has no finite moments.

Beta Distribution

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} I_{[0,1]}(x),\ E(X) = \frac{a}{a+b},\ Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Betta(1, 1) = Uniform([0, 1])

LN15: Dependence of CRV

- Joint Distribution of X, Y: $F(x, y) = P(X \leq x, Y \leq y)$.
- Jointly Continuous: X, Y are jointly continuous if $\exists\ f(x, y) : P((X, Y) \in A) = \iint_A f(x, y)$. f is called the joint density function.
- Marginal Density functions: If X, Y are jointly continuous, then X, Y are continuous and:

$$f_x(x) = \int_{-\infty}^{\infty} f_{x,y}(x, y)dy, \quad f_y(y) = \int_{-\infty}^{\infty} f_{x,y}(x, y)dx$$

  Note: X, Y continuous $\not\Longrightarrow$ X, Y jointly continuous.

- Density from Distribution: In the points where $f(x, y)$ is continuous we can find $f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$.
- Expectation: if X, Y jointly continuous, then for any measurable function $g : \mathbb{R}^2 \to \mathbb{R}$

$$E(g(X, Y) = \iint_{-\infty}^{\infty} g(x, y)f(x, y)dxdy$$

- Properties of the Expectation:
  - Linearity: $E(aX + bY) = aE(X) + bE(Y)$.
  - Product: $E(XY) = E(X)E(Y)$ if $X \perp\!\!\!\perp Y$.
- Independence and Density: $X \perp\!\!\!\perp Y$ if and only if $f_{x,y}(x, y) = f_x(x)f_y(y)$.
  Note: if $f(x, y) = g(x)h(y)$ then $\exists c : f_x(x) = cg(x),\ f_y(y) = \frac{1}{c}h(y)$.
- Covariance: $Cov(X, Y) = E((X - \mu_x)(Y - \mu_y)) = E(XY) - E(X)E(Y)$.
- Covariance is a Multilinear form: $Cov(aX + bY, Z) = a\ Cov(X, Z) + b\ Cov(Y, Z)$.
- Correlation: $\rho(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$
- Cauchy Schwarz: if X, Y jointly continuous $E(XY)^2 < E(X^2)E(Y^2)$ (strict inequality).
  This means that $|Cov(X, Y)| < \sqrt{Var(X)Var(Y)}$ and $-1 < \rho(X, Y) < 1$.
- Bivariate Normal Distribution: If X, Y are jointly continuous with Bivariate Normal Distribution then:
  - $X \sim N(\mu_1, \sigma_1^2)$
  - $Y \sim N(\mu_2, \sigma_2^2)$
  - $\rho = \rho(X, Y)$

LN16: Conditioning

- Conditional density function: Let X, Y be jointly continuous with joint density function $f_{x,y}(x,y)$ and marginals f_x, f_y. Then the conditional density of Y given $X = x$ is **the function of y**:

$$f_{(y|x)}(y \mid x) = \frac{f_{x,y}(x,y)}{f_x(x)}$$

- $f_{(t|x)}$ is a density function:
  - $f_{(y|x)}(y \mid x) \geq 0 \ \forall y$.
  - $\int_{-\infty}^{\infty} f_{(y|x)} dy = 1$
- Conditional Distribution of Y given $X = x$: $F_{Y|X}(y \mid x) = \int_{-\infty}^{y} f_{(y|x)} dy$
  Note: $P(Y \in A \mid X = x) = \int_A f_{(y|x)} dy$
- Product Rule: $f_{x,y}(x,y) = f_x(x) f_{(y|x)}(y \mid x)$
- Conditional Expectation: $\psi(X) = E(Y \mid X = x) = \int_{-\infty}^{\infty} y f_{(y|x)} dy$
- Expectation of Expectation: $E(E(Y \mid X)) = E(Y)$
  Note: This follows from $E(g(X)E(Y \mid X)) = E(g(X)Y)$
- Conditional Variance: $Var(Y \mid X) = \int_{-\infty}^{\infty} (Y - \psi(X))^2 f_{(y|x)} dy$
- Variance Decomposition:
  - $Var(Y \mid X) = E(Y^2 \mid X) - E(Y \mid X)^2$
  - $Var(Y \mid X) = V(E(Y \mid X)) + E(V(Y \mid X))$
- Conditioning the Bivariate: Let X, Y have Bivariate joint density function. Then:

$$Y \mid X \sim N\left(\mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x),\ \sigma_y^2(1 - \rho^2)\right)$$

LN17: Transformations

- Main Theorem: Let X, Y be CRVs with joint density function $f_{(x,y)}(x, y)$. Let $U \subset \mathbb{R}^2$ be an open set s.t $P((X, Y) \in U) = 1$. Let $T : U \to V$ be one-to-one and let T^{-1} be its inverse. If all the partial derivatives of T^{-1} exist and are continuous with $detJ_{T^{-1}}(z, w) \neq 0 \; \forall (z, w) \in V$, then Z and W are jointly continuous with density:

$$f_{z,w}(z, w) = f_{x,y}(T^{-1}(z, w)) \mid detJ_{T^{-1}}(z, w) \mid I_v(z, w)$$

- Sums of Random Variables: Let X and Y be continuous random variables with joint density function $f_(x, y)$. Let $Z = X + Y$. Z is continuous with density function:

$$f_z(z) = \int_{-\infty}^{\infty} f_{x,y}(x, z - x)dx$$

If $X \perp\!\!\!\perp Y$, then $f_z = \int_{-\infty}^{\infty} f_x(x) f_y(z - x)dx$

LN18: Random Vectors and the Multivariate Distribution

- Random Vectors: Let $X_1, \ldots, X_k$ be CRVs on the same Probability Space. Then $\overline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix}$ is a random vector.

- Continuity: $\overline{X}$ is said to be continuous with density function $f(x_1, \ldots, x_k)$ if $P(X \in A) = \int \cdots \int_A f(x_1, \ldots, x_k)$.
If $\overline{X}$ is continuous, then every subvector is continuous.

- Subvectors: The density function of a subvector can be obtained by integrating out the other variables.

- Independence: Let $X_1, \ldots, X_k$ be jointly continuous random variables. Then they are independent iff $f(x_1, \ldots, x_k) = \Pi_i f_{x_i}(x_i)$

- Expectation: $E(\overline{X}) = \overline{u} = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_k) \end{bmatrix}$, $Var(X) = \Sigma = [Cov(X_i, X_j)]_{i,j=i,2\ldots k}$

- Non-Singularity of Σ: If $X_1, \ldots X_k$ are jointly continuous, then Σ is non-singular and there exists the inverse Σ^{-1}. It is possible to find $\Sigma^{\frac{1}{2}}$ and $\Sigma^{-\frac{1}{2}}$ such that $\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} = \Sigma$, $\Sigma^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}} = \Sigma^{-1}$ and $\left(\Sigma^{1/2}\right)^{-1} = \Sigma^{-\frac{1}{2}}$.

- Variance in term of Expectation: $\Sigma = E\left(\left(\overline{X} - \overline{u}\right)\left(\overline{X}^T - \overline{u}^T\right)\right)$.

- Linearity of the Expectation: $E(A\overline{X} + b) = AE(\overline{X}) + b$
In particular if A is a row vector, $E(a_i X_i + \ldots a_n X_n) = \sum_i a_i E(X_i)$

- Properties of the Variance:
  - $V(\overline{X}) = E(X\ X^T) - E(X)E(X)^T$
  - $V(AX + b) = AV(X)A^T$
  - $V(a_i X_i + \ldots a_n X_n) = \sum_i a_i^2 V(X_i) + \sum_{i \neq j} a_i a_j Cov(X_i, X_j)$.

- Conditioning: The conditional density of X_1 given that $X_2 = x_2, \ldots, X_k = x_k$ is defined as:

$$f_{x_1|x_2,\ldots x_k}(x_1 \mid x_2, \ldots x_k) = \frac{f_{x_1,\ldots x_k}(x_1, \ldots, x_k)}{f_{x_2,\ldots x_k}(x_2, \ldots, x_k)}$$

Transformations of Random Vectors

- Main Theorem: Let X be a continuous random vector with density function $f(x_1, \ldots, x_k)$. Let $U \subset \mathbb{R}^k$ be a one-to-one function and let T^{-1} be its inverse. If all the partial derivatives of T^{-1} exist and are continuous with $\det J_{T^{-1}}(\overline{y}) \neq 0$, then $\overline{Y}$ is a continuous random variable with density function:

$$f_Y(y) = f_x(T^{-1}(y)) \mid detJ_{T^{-1}}(y) \mid I_v(y)$$

- Linear Transformations: $Y = AX + b$ where A is a $k \times k$ non-singular matrix, then Y is also continuous with
$$f_y(y) = f_x(A^{-1}(y-b)) \mid detA^{-1} \mid$$
Since every subvector of a continuous vector is continuous vector is continuous, for every $h \times k$ matrix A of rank h and for every vector b of dimension h, $Y = AX + b$ is continuous.

Multivariate Distribution

- Let μ be a vector of dimension k and let Σ be a $k \times k$ symmetric matrix with positive determinant of all positive leading principal minors. $\overline{X}$ as multivariate normal distribution with parameters μ and Σ ($X \sim N(\mu, \Sigma)$) if $\overline{X}$ has density:

$$f_x(x) = \frac{1}{(\sqrt{2\pi})^k} \frac{1}{\sqrt{det(\Sigma)}} exp\left(-\frac{1}{2}(x-u)^T \Sigma^{-1}(x-u)\right)$$

- Independence and Uncorrelation: $\overline{X} \sim N(\overline{0}, I)$ (where $\overline{0}$ is the zero vector and I is the identity matrix) if and only if $X_1, \ldots, X_k$ are independent and identically distributed CRVs with univariate standard normal distribution.
Analogously it can be proved that if $X \sim N(\mu, \Sigma)$ then $X_1, \ldots X_k$ are mutually independent iff Σ is a diagonal matrix. That is all covariances are 0. In this case Independence $\iff$ Uncorrelation.

- Linear Transformations: Let $X \sim N(\mu, \Sigma)$, A a $k \times k$, non-singular matrix and b a vector of dimension k. Then $Y = AX + b \sim N(A\mu + b, A\Sigma A^T)$.
Note: This also works if A is a $h \times k$ matrix of rank h.

- Standardization: if $X \sim N(\mu, \Sigma)$, then $Z = \Sigma^{-1/2}(X - u) \sim N(\overline{0}, I)$.

LN19: MGF

- Definition: $M_x(t) = E\left(e^{tx}\right) = E(\sum_{k=0}^{\infty}(tx)^k/k!) = \sum_k t^k m_k/k!$

- MGFs and Distributions: If two random variables have the same MGF, then they have the same distribution.

- Convergence: If $X, X_1, X_2, \ldots X_n$ is a sequence of random variables such that the MGFs are defined, then $M_{X_n} \to M_X$ implies $F_{X_n} \to F_X$ at least in the points where F_x is continuous.

- Maclaurin Expansion: If $t \to 0$, then $M_x(t) = 1 + M_x'(0)t + M_x''(0)t^2/2 + o(t^2) = 1 + m_1 t + m_2 t^2/2 + o(t^2)$.
Remark: $M_x^{(k)}(0) = E(X^k)$. (the k^{th} derivative).

- Linear Transformations: $M_{aX+b} = e^{tb} M_x(at)$.

- Independence: $M_{X+Y}(t) = M_X(t) M_Y(t)$.

- Central limit theorem: If $X_1, X_2 \ldots X_n$ are independent and identically distributed such that $E(X_i^2)$ exists finite, then:

$$\frac{\sum_i X_i - nu}{\sqrt{n}\sigma} \xrightarrow{\text{d}} Z \sim N(0,1)$$

LN20: Convergence and the Laws of Large Numbers

Inequalities

- Markov's inequality: $P(|X| \geq k)k^r \leq E(|X|^r)$
- Chebyshew's inequality: $P(|X-\mu| \geq k)k^2 \leq \sigma^2$, where $\sigma^2 = Var(X)$.
- Jensen's inequality: if $g : \mathbb{R} \to \mathbb{R}$ is a convex function: $E(g(X)) \geq g(E(X))$.

Types of Convergence

- Probability: $X_n \to X$ in probability if $P(|X_n - X| \geq \epsilon) \to 0$ as $n \to \infty$.
- Almost surely (a.s): $X_n \to X$ a.s if $P(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}) = 1$
- Rth Mean: $X_n \to X$ in r^{th} mean if $E(|X_n - X|^r) \to 0$ as $n \to \infty$
- Implications:
  - r^{th} mean convergence implies probability convergence.
  - a.s convergence implies probability convergence.
  - probability convergence implies distribution convergence.
  - if X_n converges to a constant in distribution, then X_n converges to the same constant in probability.
- Sufficient Condition for a.s convergence: $\sum_{n=1}^{\infty} P(|X_n - X| \geq \epsilon) < \infty$.

Laws of Large Numbers

- Sample Mean: $\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k$.
- Weak Law of Large Numbers: if $X_1, X_2, \ldots X_n$ are uncorrelated with finite second moment then $\overline{X} \to \mu = E(X_i)$ in probability.
- Strong Law of Large Numbers: If $X_1, X_2, \ldots X_n$ are i.i.d with finite expectation, then $\overline{X_n} \to \mu$ almost surely.

Transformations and Convergence

- Continuous mappings: Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then:
  - $X_n \to X$ a.s $\implies g(X_n) \to g(X)$ a.s.
  - $X_n \to X$ in probability $\implies g(X_n) \to g(X)$ in probability.
  - $X_n \to X$ in distribution $\implies g(X_n) \to g(X)$ in distribution.
- Continuous mappings $\mathbb{R}^2$ case: Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then:
  - $X_n \to X$ and $Y_n \to Y$ a.s $\implies g(X_n, Y_n) \to g(X, Y)$ a.s.
  - $X_n \to X$ and $Y_n \to Y$ in probability $\implies g(X_n, Y_n) \to g(X, Y)$ in probability.
  - The property does not hold for convergence in distribution !
- Slutski's Thoerem: $X_n \to X$ in distribution and $Y_n \to c$ in probability, then:
  - $X_n + Y_n \to X + c$ in distribution.
  - $X_n Y_n \to Xc$ in distribution.

Useful facts

- $\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$
- $\sum_{x=0}^\infty \frac{a^x}{x!} = e^a$
- $\lim_{n\to\infty}(1 + \frac{a}{n})^n = e^a$
- Geometric Sum= $\frac{u_1(r^n-1)}{r-1}$