FINAL SIMULATION THEORY – LT

Question 1: Differentiate Chi-square and K-S Test

K-S Test	Chi-square Test
Small samples	Large sample
Continuous distribution	Discrete distribution
Differences between observed and expected cumulative probabilities (CDFs)	Differences between observed and hypothesized probabilities (PDFs or PMFs)
Uses each observation in the sample without any grouping ➔ makes a better use of the data	Groups observations into a small number of cells
Cell size is not a problem	Cell sizes affect the conclusion but no firm guidelines
Exact	Approximate

Question 2: Present the parameter and application context of the following distribution:

Distribution	Parameter	Application
Discrete distribution	DISC($CumP_1$, Val_1, ...,$CumP_n$, Val_n)	- Incorporate discrete empirical data directly into the model. - Frequently used for discrete assignments such as the job type, the visitation sequence, or the batch size for an arriving entity
Contiuous distribution	CONT($CumP_1$, Val_1, ...,$CumP_n$, Val_n)	- Incorporate empirical data for continuous random variables directly into the model. - This distribution can be used to a theoretical distribution that has been fitted to the data, such as in data that have a multimodal profle or where there are signifcant outliers
Bernoulli distribution	BER(p) p is for success	Observe the successes in a trial
Binomial distribution	B(n, p) n – number of trials, p: probability of success, q (= 1 – p): probability of failure	Observe the successes in a trial
Poisson distribution	POIS(α) represents the average rate of occurrence of the event (α – number of arrivals in 1 unit of time) The mean (α) is a positive real number.	- Model the number of random events occurring in a fixed interval of time. - Model random batch sizes
Uniform distribution	UNIF(a, b) The minimum (a) and maximum (b) values for the distribution are real numbers with $a < b$.	- Used when all values over a finite range are considered to be equally likely. - It is sometimes used when no information other than the range is available
Triangular distribution	TRIA(a, c, b) The minimum (a), mode (c), and maximum (b) values for the distribution are real numbers with $a < c < b$	Used in situations in which the exact form of the distribution is not known, but estimates (or for the minimum, maximum, and most likely values are available
Exponential distribution	EXPO(λ) λ is the mean specified as a positive real number	Model interevent times in random arrival and breakdown processes, but it is generally inappropriate for modeling process delay times

Normal distribution	$\text{NORM}(\mu, \sigma^2)$ The mean (μ) is a real number and standard deviation (σ) is a positive real number	- Used in situations in which the central limit theorem applies — that is, quantities that are sums of other quantities. - It is also used empirically for many processes that appear to have a symmetric distribution.
Weibull distribution	$\text{WEIB}(\alpha, \beta)$ Scale parameter (α) and shape parameter (β) are positive real numbers	- Used in reliability models to represent the lifetime of a device. If a system consists of a large number of parts that fail independently, and if the system fails when any single part fails, then the time between successive failures can be approximated by the Weibull distribution. - Represent nonnegative task times
Erlang distribution	$\text{ERLA}(\beta, \text{k})$ The mean (β) of each of the component exponential distributions and the number of exponential random variables (k) are the parameters of the distribution. The exponential mean is a positive real number, and k is a positive integer.	- Used in situations in which an activity occurs in successive phases and each phase has an exponential distribution. - Represent the time required to complete a task.
Gamma distribution	$\text{GAMM}(\beta, \alpha)$ Scale parameter (β) and shape parameter (α) are positive real values	- For integer shape parameters, the gamma is the same as the Erlang distribution. - Represent the time required to complete some task (for example, a machining time or machine repair time).
Johnson distribution	$\text{JOHN}(\gamma, \delta, \lambda, \xi)$ Gamma shape parameter (γ), Delta shape parameter ($\delta > 0$), Lambda scale parameter ($\lambda > 0$), and Xi location parameter (ξ).	The fexibility of the Johnson distribution allows it to fit many data sets. Arena can sample from both the unbounded and bounded form of the distribution. If Delta (δ) is passed as a positive number, the bounded form is used
Lognormal distribution	LOGN(LogMean, LogStd) Mean LogMean ($\mu_l > 0$) and standard deviation LogStd ($\sigma_l > 0$)	- Used in situations in which the quantity is the product of a large number of random quantities. - Represent task times that have a distribution skewed to the right
Beta distribution	$\text{BETA}(\beta, \alpha)$ Shape parameters Beta (β) and Alpha (α) are positive real numbers	- Used as a rough model in the absence of data. - Represent random proportions, such as the proportion of defective items in a lot. - Rpresent many input quantities that can be assumed to have a range bounded on both ends.

<u>Quesiton 3</u>: What are the CDF, PDF of such distribution

Distribution	CDF (cummulative distribution function)	PDF or PMF (proba distribution function)
Discrete distribution	$F(x) = P(X \le x) = \sum_{x_i \le x} P(x_i)$	$P(x_i) = P(X=x_i)$
Contiuous distribution	$f(x) = \begin{cases} c_1 & \text{if } x = x_1 \text{ (a mass of probability } c_1 \text{ at } x_1) \\ c_j - c_{j-1} & \text{if } x_{j-1} \le x < x_j \text{, for } j = 2, 3, \ldots, n \\ 0 & \text{if } x < x_1 \text{ or } x \ge x_n \end{cases}$	
Bernoulli distribution		$P(X=0) = p \quad ; \quad P(X=1) = 1-p$

Binomial distribution		
Poisson distribution	$F(x) = \sum_{i=0}^{x} \frac{e^{-\alpha}\alpha^i}{i!}$	$p(x) = \begin{cases} \frac{e^{-\alpha}\alpha^x}{x!}, & x = 0, 1, \ldots \\ 0, & \text{otherwise} \end{cases}$
Uniform distribution	$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases}$	$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$
Triangular distribution	$\begin{cases} 0 & \text{for } x \le a, \\ \frac{(x-a)^2}{(b-a)(c-a)} & \text{for } a < x \le c, \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & \text{for } c < x < b, \\ 1 & \text{for } b \le x. \end{cases}$	$\begin{cases} 0 & \text{for } x < a, \\ \frac{2(x-a)}{(b-a)(c-a)} & \text{for } a \le x < c, \\ \frac{2}{b-a} & \text{for } x = c, \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{for } c < x \le b, \\ 0 & \text{for } b < x. \end{cases}$
Exponential distribution	$F(x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 1 - e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$	$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$
Normal distribution		$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$ for all real x
Weibull distribution	$F(x) = 1 - e^{-(\frac{x}{\alpha})^\beta}, x \ge 0$	$f(x) = \begin{cases} \frac{\beta}{\alpha^\beta} x^{\beta-1} e^{-(x/\alpha)^\beta}, & x \ge 0 \\ 0, & \textit{otherwise} \end{cases}$
Erlang distribution		$f(x) = \begin{cases} \frac{\beta^{-k} x^{k-1} e^{-x/\beta}}{(k-1)!} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$
Gamma distribution		$f(x) = \begin{cases} \frac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$ where Γ is the complete gamma function given by $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\,dt$
Lognormal distribution		$f(x) = \begin{cases} \frac{1}{\sigma x\sqrt{2\pi}} e^{-(\ln(x)-\mu)^2/(2\sigma^2)} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$
Beta distribution		$f(x) = \begin{cases} \frac{x^{\beta-1}(1-x)^{\alpha-1}}{B(\beta, \alpha)} & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$ where B is the complete beta function given by $B(\beta, \alpha) = \int_0^1 t^{\beta-1}(1-t)^{\alpha-1}\,dt$

Discrete Uniform Distribution	$F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{k}, & 1 \le x < 2 \\ \frac{2}{k}, & 2 \le x < 3 \\ \vdots & \vdots \\ \frac{k-1}{k}, & k-1 \le x < k \\ 1, & k \le x \end{cases}$	$p(x) = \frac{1}{k}, \quad x = 1, 2, \ldots, k$
Geometric distribution	$F(x) = \sum_{j=0}^{x} p(1-p)^j = \frac{p\{1-(1-p)^{x+1}\}}{1-(1-p)} = 1-(1-p)^{x+1}$, $0 < p < 1$	$p(x) = p(1-p)^x, \quad x = 0, 1, 2, \ldots$

<u>NOTE</u>: Nếu cần Reverse thì lấy CT CDF xong suy ngược lại tìm x

VD: Exponential CDF: $F(x) = 1 - e^{-\lambda x} \rightarrow X = -\frac{1}{\lambda}\ln(1 - F(x))$