Handout - MSIS 5503 – Exam 1

For all distributions, X is the *random variable* and x are the values taken by it.

Discrete Distributions

1. **Binomial Distribution**
   - $X \sim$ Binomial(n,p)
   - n independent trials with 2 outcomes in each trial ("successes " or "failure")
   - X represents x "successes " in n trials, with probability of "success" p in each trial
   - pmf = Probability (x "successes" in n trials) = $\binom{n}{x} \cdot p^x(1-p)^{n-x}$, where $\binom{n}{x} = \frac{n!}{x!(n-x)!}$
   - Expected value of X = E(X) = np
   - Variance of X = V(X) = $np(1-p)$
   - Standard Deviation of X = SD(X) = $\sqrt{np(1-p)}$
   - When $n = 1$ we get the Bernoulli distribution

2. **Multinomial Distribution**
   - X $\sim$ Multinomial(n,p_i,k)
   - n independent trials with k outcomes in each trial
   - X represents outcome x_i in n trials, with probability of outcome i being p_i in each trial, where $i = 1, ..., k$ and $\sum_{i=1}^{k} x_i = n$
   - pmf = Probability (x_i outcomes in n trials) = $\frac{n!}{x_1!x_2!,...,x_k!} p_1^{x_1} p_2^{x_2}, ..., p_k^{x_k}$
   - Expected value of X_i = E($X = x_i$) = np_i
   - Variance of X_i = V($X = x_i$) = $np_i(1-p_i)$
   - Standard Deviation of X_i = SD($X = x_i$) = $\sqrt{np_i(1-p_i)}$
   - When $k = 2$ we get the Binomial distribution

3. **Hypergeometric Distribution**
   - $X \sim$ Hypergeometric(N, K, n)
   - N is the population size, K is the number of "successes" in the population, n is sample size
   - X represents number of "successes" in the sample
   - pmf = Probability (x "successes " in sample of size n) = $\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$

4. **Negative Binomial Distribution**

   - $X \sim$ Negative Binomial(p,r)
   - independent trials with 2 outcomes in each trial ("successes" or "failure")
   - X represents x **failures** before r^{th} "success", with probability of "success" p in each trial
   - pmf = Probability (x **failures** before r^{th} "success ") = $\binom{x+r-1}{r-1} \cdot p^r(1-p)^x$, where $\binom{x+r-1}{r-1} = \frac{(x+r-1)!}{x!(r-1)!}$
   - Expected value of X (i.e., expected number of failures before r^{th} success) = E(X) = $\frac{r(1-p)}{p}$
   - Variance of X = V(X) = $\frac{r(1-p)}{p^2}$
   - Standard Deviation of X = SD(X) = $\sqrt{\frac{r(1-p)}{p^2}}$
   - When $r = 1$ we get the **Geometric distribution**

5. **Geometric Distribution**

   - $X \sim$ Geometric(p)
   - independent trials with 2 outcomes in each trial ("successes" or "failure")
   - X represents x **failures** before **first** "success", with probability of "success" p in each trial
   - pmf = Probability (x **failures** before first "success ") = $p(1-p)^x$
   - Expected value of X (i.e., expected number of failures before first success) = E(X) = $\frac{(1-p)}{p}$
   - Variance of X = V(X) = $\frac{(1-p)}{p^2}$
   - Standard Deviation of X = SD(X) = $\sqrt{\frac{(1-p)}{p^2}}$

6. **Poisson Distribution**

   - $X \sim$ Poisson(λ)
   - independent occurrences of events over time or space with λ = average number of occurrences of event per time (or space)
   - X represents number of occurrences of event over specified time (or space)
   - pmf = Probability (x occurrences of event) = $\frac{e^{-\lambda}\lambda^x}{x!}$
   - Expected value of X = E(X) = λ
   - Variance of X = V(X) = λ
   - Standard Deviation of X = SD(X) = $\sqrt{\lambda}$

Continuous Distributions

1. **Uniform Distribution**
   - $X \sim \text{Uniform}(a, b)$
   - a is the lowest value and b is the highest value
   - pf $= f(x) = \frac{1}{(b-a)}$ for $a \leq x \leq b$, and 0 elsewhere
   - cdf $= F(x) = \frac{(x-a)}{(b-a)}$ for $a \leq x \leq b$
   - Expected value of $X = \text{E}(X) = \frac{(a+b)}{2}$
   - Variance of $X = \text{V}(X) = \frac{(b-a)^2}{12}$
   - Standard Deviation of $X = \text{SD}(X) = \sqrt{\frac{(b-a)^2}{12}}$
   - kth percentile: $x = a + (b–a) * k/100$
2. **Exponential Distribution**
   - $X \sim \text{Exponential}(\lambda)$
   - $\lambda \geq 0$ is called the *rate parameter* and is the same parameter as the **Poisson distribution** i.e., it is the average number of occurrences of event per unit time (or space)
   - X is the (time or space) interval between occurrences of the event
   - pdf $= f(x) = \lambda e^{-\lambda x}$
   - cdf $= F(x) = 1 - e^{-\lambda x}$
   - Expected value of $X = \text{E}(X) = \frac{1}{\lambda}$
   - Variance of $X = \text{V}(X) = \frac{1}{\lambda^2}$
   - Standard Deviation of $X = \text{SD}(X) = \frac{1}{\lambda}$
   - kth percentile: $x = -ln(1–(k/100))/\lambda$
3. **Normal Distribution - see Standard Normal Table**
   - You need to know how to convert a normal distribution with mean μ and Standard Deviation of σ, into a Standard Normal Distribution with mean 0 and Standard Deviation 1. Then you use the standard normal tables to find the cumulative probability. You should also be able to convert a cumulative probability from the tables back into values of the random variable for both the Standard Normal and the Normal cases.
4. **Student t Distribution - see Student t Table**
   - You only need to know how to use tables with k degrees of freedom to obtain the probability of the random variable and convert certain probabilities back to the value of the random variable (using the tables)

General Strategies for Identifying Distributions

Discrete-valued Random Variables

1. If the problem statement has a finite sample size (n) from a **large(infinite) population**:

   - The events in the population take on binary values ("yes", "no") or ("0","1") or ("success", "failure") with associated probabilities p and $(q = 1 - p)$, respectively
     - If the random variable X = number of "successes"
       * Then use **Binomial(n,p)**
     - If the random variable X = number of "failures" till first "success"
       * Then use **Geometric(p)**
     - If the random variable X = number of "failures" till r^{th} "success"
       * Then use **Negative Binomial(p,r)**

2. If the problem statement has a finite sample size (n) from a **large(infinite) population**:

   - The events in the population take on k values with associated probabilities p_i with $(i = 1, ..., k)$
     - If the random variable X_i = number of "successes" for each $(i = 1, ..., k)$
       * Then use **Multinomial**(n,p_i,k) where $n = \sum_{i=1}^{k} x_i$

3. If the problem statement has a finite sample size (n) from a **finite population of size N**:

   - There are K "successes" in the population and $N - K$ "failures"
     - If the random variable X = number of "successes" (with $n - X$ "failures") in a sample of size n taken from the population (*without replacement*)
       * Then use **Hypergeometric**(N,K,n,x)

4. If the problem statement deals with an interval (in time or space) of interest:

   - The probability of each event of interest is independent of the previous occurrence of that event
     - If the random variable X=number of events in the an interval (in time or space) of interest
       * Then use **Poisson** (λ), where λ is the average number of events in the interval of interest
         Note: Care must be taken to establish the interval of interest for X relative to λ.

Continuous Random Variables

The distribution will usually be specified. The exception is that, for the Exponential distribution, you should know its relationship with the Poisson. For normal distribution, you should be able to convert to standard normal and *vice versa*.

Calculate Sample Statistics for Numeric and Grouped Data

Sample Mean and Standard Deviation Formulas - Numerical Data

Sample Mean	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Sample Variance	$s^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \overline{X})^2$
Sample Standard Deviation	$s = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \overline{X})^2}$

Sample Mean and Standard Deviation Formulas - Grouped Data
There are n groups - The i^{th} group has frequency f_i

Sample Mean	$\overline{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$
Sample Variance	$s^2 = \frac{1}{(\sum_{i=1}^{n} f_i)-1}\sum_{i=1}^{n} f_i(x_i - \overline{X})^2$
Sample Standard Deviation	$s = \sqrt{\frac{1}{(\sum_{i=1}^{n} f_i)-1}\sum_{i=1}^{n} f_i(x_i - \overline{X})^2}$

Sample Mean to Estimate Population Mean

- **Population: Normal, Population Mean (unknown) μ, Population SD (known) σ, Sample size n: Any**
  - Sampling Distribution of $\overline{X}$: $\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$
  - $(1-\alpha)\%$ Confidence Interval: $\overline{X} \pm Z_{\alpha/2}(\frac{\sigma}{\sqrt{n}})$
  - Sample Test Statistic (Z-distribution): $\frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$, where μ_0 is the Null Hypothesis value for μ

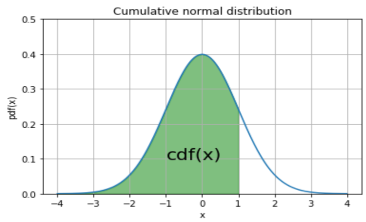
- **Population: Normal, Population Mean (unknown) μ, Population SD (Unknown), Sample size n:** ≥ 30
  - Sample Standard Deviation $= s$
  - Sampling Distribution of $\overline{X}$: $\overline{X} \sim N(\mu, \frac{s}{\sqrt{n}})$
  - $(1-\alpha)\%$ Confidence Interval: $\overline{X} \pm Z_{\alpha/2}(\frac{s}{\sqrt{n}})$
  - Sample Test Statistic (Z-distribution): $\frac{\overline{X}-\mu_0}{\frac{s}{\sqrt{n}}}$, where μ_0 is the Null Hypothesis value for μ

- **Population: Normal, Population Mean (unknown) μ, Population SD (Unknown), Sample size n:** < 30
  - Sample Standard Deviation $= s$
  - Sampling Distribution of $\overline{X}$: $\overline{X} \sim t(\mu, \frac{s}{\sqrt{n}})$ with $(n-1)$ degrees of freedom
  - $(1-\alpha)\%$ Confidence Interval: $\overline{X} \pm t_{\alpha/2}(\frac{s}{\sqrt{n}})$ with $(n-1)$ degrees of freedom
  - Sample Test Statistic (t-distribution) with $(n-1)$ degrees of freedom: $\frac{\overline{X}-\mu_0}{\frac{s}{\sqrt{n}}}$, where μ_0 is the Null Hypothesis value for μ

- **Population: NOT Normal or Unknown, Population Mean (unknown) μ, Sample size n:** ≥ 30
  - **Using Central Limit Theorem**
  - Sampling Distribution of $\overline{X}$: $\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$, if Population SD σ is known.
  - Sampling Distribution of $\overline{X}$: $\overline{X} \sim t(\mu, \frac{s}{\sqrt{n}})$ with $(n-1)$ degrees of freedom, if population SD σ NOT known. s is sample standard deviation.
  - $(1-\alpha)\%$ Confidence Interval: $\overline{X} \pm Z_{\alpha/2}(\frac{\sigma}{\sqrt{n}})$ **OR** $\overline{X} \pm t_{\alpha/2}(\frac{s}{\sqrt{n}})$ with $(n-1)$ degrees of freedom if σ NOT known
  - Sample Test Statistic (Z-distribution): $\frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ **OR** t-distribution with $(n-1)$ degrees of freedom: $\frac{\overline{X}-\mu_0}{\frac{s}{\sqrt{n}}}$ if σ NOT known, where μ_0 is the Null Hypothesis value for μ

Sample Proportion to Estimate Population Proportion

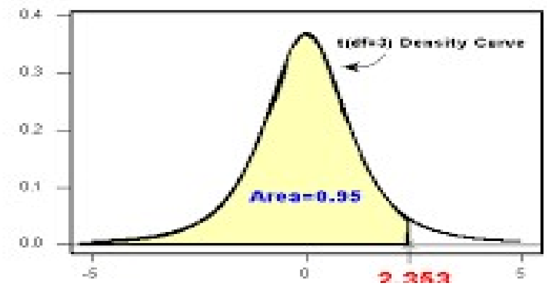
- **Population Proportion (unknown)** p**, Generally need** $np' > 5$ **and** $n(1-p') > 5$
  - Sample Proportion: p'
  - Sampling Distribution of p': $p' \sim N(p, \sqrt{\frac{p(1-p)}{n}})$
  - $(1-\alpha)\%$ Confidence Interval: $p' \pm Z_{\alpha/2}(\sqrt{\frac{p'(1-p')}{n}})$ for $n \geq 30$ or $p' \pm t_{\alpha/2}(\sqrt{\frac{p'(1-p')}{n}})$ for $n < 30$
  - Sample Test Statistic (Z-distribution for $n \geq 30$): $\frac{p'-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, where p_0 is the Null Hypothesis value for p
  - Sample Test Statistic (t-distribution with $(n-1)$ degrees of freedom for $n < 30$): $\frac{p'-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, where p_0 is the Null Hypothesis value for p



STANDARD NORMAL CDF - AREA TO THE LEFT

CDF FOR *NEGATIVE* Z VALUES: *TAKE (1-CELL VALUE)*

Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000



STUDENT'S t PERCENTILES: TABLE SHOWS t-values

DF	**60.0**	**66.7**	**75.0**	**80.0**	**87.5**	**90.0**	**95.0**	**97.5**	**99.0**	**99.5**	**99.9**
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.436	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.436	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.435	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.435	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.435	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.435	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385
35	0.255	0.434	0.682	0.852	1.170	1.306	1.690	2.030	2.438	2.724	3.340
40	0.255	0.434	0.681	0.851	1.167	1.303	1.684	2.021	2.423	2.704	3.307
45	0.255	0.434	0.680	0.850	1.165	1.301	1.679	2.014	2.412	2.690	3.281
50	0.255	0.433	0.679	0.849	1.164	1.299	1.676	2.009	2.403	2.678	3.261
55	0.255	0.433	0.679	0.848	1.163	1.297	1.673	2.004	2.396	2.668	3.245
60	0.254	0.433	0.679	0.848	1.162	1.296	1.671	2.000	2.390	2.660	3.232
∞	0.253	0.431	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090

Rough Paper

Rough Paper

Rough Paper