Math 10A with Professor Stankova
Worksheet, Discussion #42; Friday, 12/8/2017
GSI name: Roy Zhao

Problems

1. For each of the following distributions, derive/find all of the following: PMF/PDF, CDF, median, μ, σ, σ^2. Distributions are: Uniform Distribution, Birthday Distribution, Exponential Distribution, Laplacian Distribution, Pareto Distribution, Normal Distribution, Logistic Distribution, Dart Throw Distribution, Die Throw, Coin Flips, Number of Boys-Girls in a Family, Binomial Distribution.

True/False

2. True False Histograms are always defined over intervals of the form $[a, b]$ with $a \geq 0$ because probabilities are always non-negative.

3. True False The height of a rectangle in a histogram equals $\frac{\text{the amount of the data corresponding to the subinterval}}{\text{the width of the subinterval}}$, because all rectangular areas in a histogram must sum up to 1.

4. True False $P(x) = e^{-x^2}$ is a PDF on $\mathbb{R}$.

5. True False Uniform PDFs are defined by $f(x) = c$ for all $x \in \mathbb{R}$.

6. True False Shifting the bell-shaped PDF $f(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$ to the left by 5 units results in another PDF $g(x) = \frac{1}{\sqrt{\pi}} e^{-(x+5)^2}$ centered at $x = 5$.

7. True False CDFs are continuous functions on $\mathbb{R}$ but PDFs could be piecewise continuous on $\mathbb{R}$.

8. True False For a PDF to be centered at $x = a$, it means that a is the median of the CDF.

9. True False We adjusted the solution $y(t) = y_0 e^{-ct}$ of the exponential decay DE $y'(t) = -cy(t), y(0) = y_0$ with rate constant c to the PDF $Y(t) = ce^{-ct}$ for $x \geq 0$ and $y(t) = 0$ for $x < 0$ because we wanted to simplify the initial condition y_0 to match c.

10. True False CDFs behave in general like antiderivaties of their PDFs, but there are some situations where the CDF is not continuous and hence there is no way it can be an antiderivative of a PDF.

11. True False The formula $P(a \leq X \leq b) = F(b) - F(a)$ for a CDF $F(x)$ works because $F(x)$ can be essentially thoughout of as the "area-so-far" function for a PDF.

12. True False To prove that a function $F(x)$ on $\mathbb{R}$ is a CDF, we need only to confirm that it is non-decreasing on $\mathbb{R}$, attains only non-negative values, and that its value tends to 1 and 0 as $x \to \infty$ and $x \to -\infty$ correspondingly.

13. True False The CDF of the bell-shaped PDF $f(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$ has a graph that is increasing and looks like a solution to the classic logistic model with carrying capacity $K = 1$, but it cannot be exactly equal to it because there is no elementary function of an antiderivative of e^{-x^2} yet we have derived the formula $g(x) = \frac{1}{1+Ae^{-kt}}$ for these logistic model solutions.

14. True False There are at least three ways to compute the probability of a person to be born in May: using a discrete random variable, and using the PDF or the CDF of a continuous random variable.

15. True False The formula for the mean of a continuous random variable is a limit version of the mean for a discrete random variable; but while the latter always exists for a finite amount of data, the former may not exist for certain continuous random variables.

16. True False If the mean is larger than the median, the distribution tends to be more spread away on the right and more clustered together on the left.

17. True False Both the median and the mean of an exponential distribution directly depend on the initial condition $f(0)$ of the DE $f'(x) = Cf(x)$ and on nothing else.

18. True False During the process of drug decay (or extinction of species), it takes longer for half of the drug (or species) to be gone than for a randomly chosen average molecule (species) to be gone.

19. True False If we make the area between a PDF and the x-axis out of uniform cardboard material and make an infinite seesaw out of the x-axis, the point on the x-axis where the seesaw will balance is the median of the distribution because there is an equal material to the left and to the right of the median.

20. True False Exam distributions of large classes tend to have smaller means than medians when the medians are higher than 50% of the maximum possible score.

21. True False The Pareto distribution fails to have a well-defined mean when the constant $a \geq 2$.

22. True False Improper integrals resurface when we want to compute probabilities of discrete random variables with finitely many values.

23. True False For a symmetric distribution, we do not have to calculate the mean because it will always equal the median.

24. True False Integration by Parts and Substitution Rule can be safely forgotten as Statistics uses ready formulas from Mathematics in all basic examples.

25. True False The formula for the standard deviation of a continuous random variable is a limit version of the standard deviation for a discrete random variable; but while the latter always exists for a finite amount of data, the former may not exist for certain continuous random variables.

26. True False If the standard deviation is larger than the mean for an exam distribution with only non-negative scores, then Chebychev's inequality will for sure predict negative scores on this exam.

27. True False All of the median, mean, and standard deviation for an exponential distribution and for the Laplacian distribution directly depend on the initial condition $f(0)$ of the PDF and on nothing else.

28. True False The $\sqrt{12}$ that appears in the denominator of the standard deviation for the uniform birthday distribution will also appear in the denominator of σ for any uniform distribution.

29. True False If we make the area between a PDF and the x-axis out of uniform cardboard material, more than 55% of the material will be dedicated to the interval within 1.5μ of σ

30. True False For exam distributions of large classes statisticians use a modified formula for σ, essentially "pretending" that there is one more sample item, in order to get a more accurate "unbiased" statistics.

31. True False Improper integrals resurface when we want to justify the shortcut formula for σ^2.

32. True False For a symmetric distribution centered at 0, we do not have to calculate σ because it will always be 0 or not well-defined.

33. True False For any continuous or discrete random variable X with a well-defined mean and variance, it is true that $\sigma^2(X) + \mu^2(X) = \mu(X^2)$, where X^2 is a random variable whose values are the squares of the corresponding values of X.

34. True False The PDF of the logistic distribution is given by the solution to the logistic DE $y'(t) = ry(t)(1 - y(t))$.

35. True False When we increase the initial condition $y(0)$ in the logistic model but keep everything else the same, the logistic distribution tends to move to the left; and similarly, when we increase r, the center of the logistic distribution tends to move to the left.

36. True False After we establish that the logistic distribution is symmetric about $x = a/r$ and that the normal distribution is symmetric about $x = \mu$, we can automatically conclude that their means are correspondingly $x = a/r$ and μ without any further integration considerations.

37. True False The log-plot transforms the points on a logistic distribution to lie on a line.

38. True False σ appears twice in denominators in the formula for normal distribution.

39. True False $\frac{1}{\sqrt{2\pi}}$ ensures that the formula for the normal distribution indeed represents a valid PDF.

40. True False The z-scores are made possible by a linear change of variables that converts any normal distribution into the standard normal distribution; and hence, to know everything about normal distributions it suffices to study only the standard normal distribution.

41. True False z scores are not suitable for computing probabilities of the type $P(-\infty \leq X \leq a)$ or $P(b \leq X \leq \mu)$ for arbitrary normal distributions.

42. True False Normal distributions are defined only for positive X; yet, when converted to the standard normal distribution, they may be defined for negative X too.

43. True False $\pi^2 = \int_{-\infty}^{\infty} \frac{3t^2 e^{-t}}{(1+e^{-t})^2} dt$.

44. True False PMFs replace PDFs when moving from continuous to discrete random variables.

45. True False CDFs can be defined by the same probability formula for both discrete and continuous variables; however, at the next step when actually computing the CDFs, one must be careful to use correspondingly PMFs with integrals and PDFs with appropriate summations.

46. True False The target spaces for PDFs, PMFs, and probability functions are all the same.

47. True False The domains of PDFs and PMFs are the corresponding outcome spaces Ω.

48. True False A set with 10 elements has 2^{10} number of subsets, and hence there will be 2^{10} inputs for any probability function P on any outcome space Ω with 10 points.

49. True False While we can define a discrete random variable without using a PMF, a continuous random variable has a PDF in its definition.

50. True False $P(A \cup B) = P(A) + P(B)$ as long as A and B are independent events in different outcome spaces.

51. True False Every row in Pascal's triangle represents the PMF of some binomial distribution.

52. True False The probability of having 3 boys in a family of 7 children (assuming equal probability of having boys and girls) is equal to the probability of having 4 boys in a family of 7 children.

53. True False To each probability space we can associate different random variables, but each probability space has a unique probability function.

54. True False The formula for the mean of a uniform distribution on $[a, b]$ has $\sqrt{12}$ in the denominator, while the formula for its standard deviation has 2 in the denominator.

55. True False For any R.V. X it is true that $E(5X) = 5E(X)$, $E(5 + X) = E(5) + E(X)$, and $E(5X) = E(5)E(X)$.

56. True False The shortcut formula for variance $Var(X) = E[X^2] - E^2[X]$ works for both continuous and discrete R.V.s X.

57. True False The standard deviation tells us how risky (or not risky) it is to play certain games.

58. True False The size of the outcome space Ω for rolling 5 die is 5^6.

59. True False The PDFs and PMFs play analogous roles in the formulas for mean (expected value) and standard deviation (standard error).

60. True False $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} te^{-\frac{t^2}{2}} dt = 1$ and $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = 0$.

61. True False Var(X) can be defined as the expected value of the square of how far X is from its own mean $E(X) = \mu$; i.e., $Var(X) = E[(X - E(X))^2]$, or using an integral formula for continuous X, and as a consequence $Var(cX) = c^2 Var(X)$ and $SE(cX) = |c|SE(X)$ for any $c \in \mathbb{R}$.

62. True False The binomial coefficient $\binom{n}{k}$ will not change if we replace k by $n - k$.

63. True False The binomial coefficients $\binom{n}{k}$ increase for the first half of the values of k and then they decrease.

64. True False The formula for the mean of the average $\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$ of independent identically distributed RV's has $\sqrt{n}$ in the denominator.

65. True False For any RV's X and Y, it is true that $E(5X - 7Y) = 5E(X) - 7E(Y)$ and $E(XY) = E(X)E(Y)$.

66. True False The formula for the variance $Var(X + Y) = Var(X) + Var(Y)$ works regardless of whether the RV's X and Y are independent or not.

67. True False To approximate the height of a tall tree (using similar triangles and measurements along the ground – without climbing the tree!) it is better to ask several people to do it independently of each other and then to average their results, than to do it once just by yourself.

68. True False The z-scores can be used to reduce probability calculations on any normal distribution to using a table in the textbook, as long as we are not too many standard deviations σ from the mean μ in our original normal distribution.

69. True False According to the Central Limit Theorem, the normalized distribution $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ is the standard normal distribution for n large, where $\overline{X}$ is the average of n independent, identically distributed variables, each with mean μ and standard error $\sigma/\sqrt{n}$.

70. True False The Law of Large Numbers can be viewed as part of the Central Limit Theorem.

71. True False To prove that $Var(X) = E(X^2) - E^2(X)$ for any RV X, one needs to use that $E(X_1 + X_2) = E(X_1) + E(X_2)$ for some particular dependent RV's X_1 and X_2.

72. True False The Maximum Likelihood (M-L) method uses a probabilistic experiment to estimate a real world parameter θ, by considering all possible outcomes of the experiment.

73. True False The formula for the PMF $f(k)$ of the binomial distribution for k successes out of n independent trials is a polynomial of degree n in the probability p of a success for each trial.

74. True False To determine the Maximum Likelihood of a continuous variable, we can write a table for all values of the individual likelihoods $L(x|\theta)$ and choose the θ that yields the largest value of $L(x|\theta)$.

75. True False When using a random variable X and one experiment with it to estimate a parameter θ, we compare all values of $L(x|\theta)$ for $X = x$ fixed and θ varying.

76. True False The log-derivative of $f(\theta) = L(x|\theta)$ can used to shortcut calculations with the M-L method.

77. True False The M-L method applied to estimate the standard error in a normal distribution by taking n independent measurements $x_1, x_2, \ldots, x_n$ leads to the standard deviation formula for these same n measurements.

78. True False The mean μ and standard error σ of a random variable X can be viewed as parameters and hence the M-L method can be applied to estimate them.

79. True False If you interview n random people in the U.S. about their preference to saying "tom-ei-to" v.s. "tom-a-to" and you get k of them saying they prefer "tom-ei-to", you can conclude that based on this data alone your best prediction for the fraction of all people in the U.S. who indeed prefer "tom-ei-to" over "tom-a-to" is $\frac{n}{k}$.

80. True False In the class examples of binomial distribution for discrete variables and normal distributions for continuous variables, in order to apply the M-L method, it was crucial that we could find the critical points for the corresponding PMF/PDFs.

81. True False The formula $(\ln(f(\theta)))' = \frac{f'(\theta)}{f(\theta)}$ makes it possible to shift the process from finding critical points for the original $f(\theta)$ to finding critical points of $\ln(f(\theta))$.

82. True False The "score equation" for a parameter θ estimated by the M-L method is essentially setting up the log-derivative of the likelihood function $f(\theta) = L(x|\theta)$ equal to 0.

83. True False Working with a larger sample of the population will make the "biased" estimate for σ^2 less biased.

84. True False The Null hypothesis is a theory that we believe is true.

85. True False The higher the significance of a test, the higher the probability of rejecting a true Null hypothesis.

86. True False Adding up the power and the significance of a test yields 1.

87. True False A type-2 error made by a road patrol may result in letting drunken drivers continue driving.

88. True False A criterion is used on data from experiments to accept the Alternative theory or to keep the Null hypothesis.

89. True False A number of important questions about hypothesis testing can be reformulated eventually about using z-scores.

90. True False The significance of a test shows how often, on the average, we can make a Type 1 error.

91. True False Using two-sided Alternative Hypotheses H_1 may lead to twice as large significance as their one-sided analogs.

92. True False Deciding on a rejection region lead to making global policies that affect many local decisions.

93. True False In class we solved at least one problem in finding the power of a test $1 - \beta$.

94. True False The p-value of a possible test result r is the probability that the experiment produces a result that is equally or more extreme (towards H_1) than r, assuming H_0 is true.