MATH20802: STATISTICAL METHODS EXAMPLES

1. If $X \sim N(\mu, \sigma^2)$ show that its mgf is

$$M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right).$$

2. If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent then show that $aX_1 + bX_2 + c \sim N(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2)$.

3. If $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ are iid then show that $\overline{X} \sim N(\mu, \sigma^2/n)$, where

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

   is the sample mean.

4. If $X \sim Uni(a, b)$ then show that its mgf is $M_X(t) = \{\exp(bt) - \exp(at)\}/((b-a)t)$.

5. If $X \sim Exp(\lambda)$ then show that its mgf is:

$$M_X(t) = \frac{\lambda}{\lambda - t}.$$

6. If $X \sim Ga(a, \lambda)$ then show that its mgf is:

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^a.$$

7. Show that as $\nu \to \infty$, the pdf of the Student's t distribution given by

$$f(x) \quad = \quad \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \qquad -\infty < x < \infty$$

   approaches the standard normal pdf (Hint: use Stirling's approximation that as $n \to \infty$, $\Gamma(n+1) \sim \sqrt{2\pi n} n^n \exp(-n)$).

8. How is a random variable T with a t-distribution related to a random variable with a normal distribution and an independent random variable with a chi-square distribution. If U and V are independent, U being distributed $N(3, 16)$ and V being distributed as chi-square on 9 degrees of freedom, find $\Pr(U - 3 < 4.33\sqrt{V})$.

9. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ where σ^2 is known. Consider the estimator $\overline{X} = (1/n)\sum_{i=1}^{n} X_i$ for μ. Find its bias, MSE, variance and check whether it is consistent.

10. Let $X_1, X_2, \ldots, X_n$ be a random sample from $Uni(\theta, \theta+1)$. Consider the estimator $\overline{X} = (1/n)\sum_{i=1}^{n} X_i$ for θ. Find its bias, MSE, variance and check whether it is consistent.

11. Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution

$$f(x) = \begin{cases} 2x/\theta^2, & \text{if } 0 < x < \theta, \\ 0, & \text{otherwise,} \end{cases}$$

    where θ is an unknown parameter. Consider the estimator $\widehat{\theta} = \overline{X}$. Find the bias and MSE of $\widehat{\theta}$. Is $\widehat{\theta}$ MSE consistent for θ? Using your results above, modify the estimator product a new estimator of θ which is unbiased. What is the MSE of the new estimator?

12. Let $\widehat{\theta}_1$ and $\widehat{\theta}_2$ be two independent and unbiased estimators of a given parameter θ. If $Var(\widehat{\theta}_1) = 3Var(\widehat{\theta}_2)$ find the values of the two constants a_1 and a_2 such that the linear combinations $a_1\widehat{\theta}_1 + a_2\widehat{\theta}_2$ is both an unbiased estimator of θ and also has the smallest variance of such linear combinations.

13. For the independent samples design, assume $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are mutually independent random samples from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively. For the matched pairs design, assume also that $\text{Corr}(X_i, Y_i) = \rho$ and let $D_i = X_i - Y_i$. To estimate $\mu_1 - \mu_2$, we use $\overline{X} - \overline{Y}$ for the independent samples design and $\overline{D} = \overline{X} - \overline{Y}$ for the matched pairs design. Show that the ratio of the variances of $\overline{X} - \overline{Y}$ for the two designs equals

$$\frac{Var_{\text{matched}}\left(\overline{X} - \overline{Y}\right)}{Var_{\text{independent}}\left(\overline{X} - \overline{Y}\right)} = 1 - \rho.$$

14. In a series of m independent Bernoulli trials there are X successes. In a further series of n trials there are Y successes. Assuming that the probability of success, p, is the same for both sets of trials, show that

$$\widehat{p}_1 = \frac{1}{2}\left(\frac{X}{m} + \frac{Y}{n}\right) \tag{1}$$

and

$$\widehat{p}_2 = \frac{X + Y}{m + n} \tag{2}$$

are both unbiased estimators of p. Which of these two estimators do you prefer and why?

15. If X_1, X_2, X_3 constitute a random sample from a $N(\mu, \sigma^2)$ distribution with μ unknown then consider the following estimators of μ:

$$\widehat{\mu}_1 = (X_1 + 2X_2 + X_3)/4$$

and

$$\widehat{\mu}_2 = (X_1 + X_2 + X_3)/3.$$

Which of these two estimators would it be best to use in practice and why?

16. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2.

    (i) Show that $\sum_{i=1}^n a_i X_i$ is an unbiased estimator of μ for any set of known constants $a_1, a_2, \ldots, a_n$ with $\sum_{i=1}^n a_i = 1$.

    (ii) If $\sum_{i=1}^n a_i = 1$ show that $Var(\sum_{i=1}^n a_i X_i)$ is minimized for $a_i = 1/n$, $i = 1, 2, \ldots, n$ (Hint: prove that $\sum_{i=1}^n a_i^2 = \sum_{i=1}^n (a_i - 1/n)^2 + 1/n$ when $\sum_{i=1}^n a_i = 1$.)

17. Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. Find $E(S^2)$ and $Var(S^2)$ where

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2.$$

If S^2 a MSE consistent estimator of σ^2? (Hint: Recall that $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$.)

18. Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution

$$f(x) = \begin{cases} (x/\gamma)\exp(-x^2/(2\gamma)), & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

   (i) Show that $\widehat{\gamma} = (1/(2n)) \sum_{i=1}^{n} X_i^2$ is an unbiased estimator of γ.
   (ii) Fine the MSE of $\widehat{\gamma}$.
   (iii) What is the approximate distribution of $\widehat{\gamma}$ for large n.

19. If the random variable $X \sim Bin(n, p)$ show that the sample proportion $\widehat{p} = X/n$ is an unbiased estimator of p. Calculate $Var(\widehat{p})$ and hence the MSE $(\widehat{p})$. Is $\widehat{p}$ MSE consistent for p?

20. Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(0, \sigma^2)$ distribution. Show that $\sum_{i=1}^{n} X_i^2/n$ is an unbiased estimator of σ^2. Show that it is also consistent.

21. Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution

$$f(x) = \begin{cases} \exp(\delta - x), & \text{if } x \geq \delta, \\ 0, & \text{otherwise,} \end{cases}$$

   where δ is an unknown parameter. Show that $\overline{X}$ is a biased estimator of δ. Hence, find a linear transformation of $\overline{X}$ which will provide an unbiased estimator of δ. Is your new estimator MSE consistent for δ?

22. When have seen in the notes that the maximum likelihood estimator of the parameter λ in a Poisson distribution is $\widehat{\lambda} = \overline{X}$. Find the expected value and the variance of $\widehat{\lambda}$ and show that it is a consistent estimator for λ.

23. Write down the likelihood function if $X_1, X_2, \ldots, X_n$ is a random sample from $Po(\lambda)$.

24. Write down the likelihood function if $X_1, X_2, \ldots, X_n$ is a random sample from $Exp(\lambda)$.

25. If $X_1, X_2, \ldots, X_n$ is a random sample from $Po(\lambda)$ find the mle of λ.

26. If $X_1, X_2, \ldots, X_n$ is a random sample from $Exp(\lambda)$ find the mle of λ.

27. If $X_1, X_2, \ldots, X_n$ is a random sample from $U(0, \theta)$ find the mle of θ.

28. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from $U[\theta - 1/2, \theta + 1/2]$. Find the mle of θ.

29. If $X_1, X_2, \ldots, X_n$ is a random sample from $Exp(\lambda)$ find the mles of $\Pr(X < 1)$ and the mean of the distribution.

30. If $X_1, X_2, \ldots, X_n$ is a random sample from $Exp(1/\lambda)$ show that the mle of λ is unbiased and consistent.

31. If $X_1, X_2, \ldots, X_n$ is a random sample from the distribution

$$f(x) = \begin{cases} \exp(\delta - x), & \text{if } x \geq \delta, \\ 0, & \text{otherwise} \end{cases}$$

   find the mle of δ.

32. If $X_1, X_2, \ldots, X_n$ is a random sample from the distribution

$$f(x) = \begin{cases} (\theta+1)x^\theta, & \text{if } 0 < x < 1, \\ 0, & \text{otherwise} \end{cases}$$

find the mle of θ.

33. If $X_1, X_2, \ldots, X_n$ is a random sample from the geometric distribution (with parameter p) find the mle of p.

34. Among N independent random variable having the $Bin(2, p)$ distribution, n_0 take on the value 0, n_1 take on the value 1 while n_2 take on the value 2. We have $n_0 + n_1 + n_2 = N$. Find the mle of p.

35. The proportion p of the breakfast cereal Cocobix bought by men rather than women is unknown. In a random sample of 70 purchases of the cereal it was found that 58 were made by men.

    (i) Write down the likelihood function $L(p)$.

    (ii) Find the mle of p and an approximate 95% confidence interval for its true value.

    (iii) Sketch $L(p)$ for $0 \leq p \leq 1$ and find the mle of p if we know its true values lies in the interval $1/2 \leq p \leq 2/3$.

36. If $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with the pdf $f(x) = \theta x^{\theta-1}$ find the mle of θ.

37. If $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with the pdf $f(x) = \theta^2 x \exp(-\theta x)$ find the mle of θ.

38. The following is a random sample from $N(\mu, 1)$ where μ is unknown: 1.466, 1.791, 1.353, 0.059, 1.499, 1.209, 0.087, -0.237. Find the mle of $\Pr(X < 0)$, where $X \sim N(\mu, 1)$.

39. Given the independent random samples: $X_1, X_2, \ldots, X_n \sim N(\alpha+\beta, 1)$ and $Y_1, Y_2, \ldots, Y_n \sim N(\alpha-\beta, 1)$ find the mles of α and β.

40. If $X_1, X_2, \ldots, X_n$ be a random sample from the $Ga(r, \lambda)$ (where r is known) find the mles of λ and $\tau = (2\lambda - 1)^2$.

41. Given the independent random samples: $X_1, X_2, \ldots, X_n \sim Exp(\lambda_1)$ and $Y_1, Y_2, \ldots, Y_m \sim Exp(\lambda_1\lambda_2)$ find the mles of λ_1 and λ_2.

42. Consider the beta distribution with the probability density function

$$f(x) = \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)} x^{a_1-1}(1-x)^{a_2-1},$$

where $0 < x < 1$, $a_1 > 0$ and $a_2 > 0$. If $x_1, \ldots, x_n$ is a random sample from this distribution then show the estimates of a_1 and a_2 under the method of maximum likelihood are the solutions of the equations:

$$\Psi(a_1) - \Psi(a_1+a_2) = \frac{1}{n}\sum_{i=1}^{n} \log(x_i)$$

and

$$\Psi(a_2) - \Psi(a_1 + a_2) = \frac{1}{n} \sum_{i=1}^{n} \log(1 - x_i),$$

where

$$\Psi(x) = \frac{d \log \Gamma(x)}{dx}$$

denotes the Euler's psi function.

43. Consider the simple linear regression model given by

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n,$$

where the errors ϵ_i are independent of each other, and are normally distributed with mean 0 and standard deviation σ. Establish the following:

(a) the mles of β_1 and β_0 are

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \left(\sum_{i=1}^{n} x_i\right) \left(\sum_{i=1}^{n} y_i\right)}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i\right)^2};$$

and

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}.$$

(b) the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are unbiased, i.e.

$$E\left(\widehat{\beta}_0\right) = \beta_0$$

and

$$E\left(\widehat{\beta}_1\right) = \beta_1.$$

(c) the variance of $\widehat{\beta}_1$ is given by

$$Var\left(\widehat{\beta}_1\right) = \frac{n\sigma^2}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}.$$

44. Often the conditions of the problem dictate that the intercept β_0 must be zero, e.g. the sales revenue as a function of the number of units sold or the gas mileage of a car as a function of the weight of the car. This is called regression through the origin. Show that the mle of the slope β_1 when fitting the line $y = \beta_1 x$ based on the data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ is

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

45. Consider the two independent random samples: $X_1, X_2, \ldots, X_n$ iid from $N(\mu_X, \sigma^2)$ and $Y_1, Y_2, \ldots, Y_m$ iid from $N(\mu_Y, \sigma^2)$, where μ_X and μ_Y are assumed known. Find the mle of σ.

46. If $X_1, X_2, \ldots, X_n$ is a random sample from $Exp(\lambda)$ where $0 < \lambda \leq 2$. Find the mle of λ.

47. If $X_1, X_2, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$ find the mles of μ and σ.

48. Given the pdf

$$f(x) = \begin{cases} 1/4 & \text{if } x = 0, 1, 2, 3, \\ 0 & \text{otherwise,} \end{cases}$$

find the probability that a random sample of size 36, selected with replacement, will yield an average greater than 1.4 but less than 1.8.

49. Verify the identity

$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 + n(\overline{X} - \mu)^2.$$

50. A random sample of 20 observations are taken from $N(\mu, 1.4)$ distribution when μ is unknown. Find two numbers a and b such that $\Pr(a \leq S^2 \leq b) = 0.95$, where S^2 denotes the sample variance.

51. In 16 test runs the petrol consumption of an experimental engine has a standard deviation of 2.2 gallons. Stating any distributional assumptions you make, test whether σ, the true standard deviation of the petrol consumption of the engine, is equal to 4.5 gallons.

52. If $X \sim F_{\nu_1,\nu_2}$ explain why $1/X \sim F_{\nu_2,\nu_1}$. Show that $F_{\nu_1,\nu_2,1-\alpha} = 1/F_{\nu_2,\nu_1,\alpha}$.

53. $n_1 = 12$ mature citrus trees, all of a particular variety, had their heights measured and the standard deviation of these values was found to be $s_1 = 0.35$. An independent random sample of $n_2 = 15$ mature citrus tree all of another variety also had their heights measured and the sample standard deviation was found to be $s_2 = 0.41$. Assuming that the random samples were selected from normal populations show that a test at the 5% level of whether the variance in the two populations may be regarded as being equal concludes that they are. Construct a 95% confidence interval for the value of the common variance.

54. An opinion poll is carried out in which $n = 15$ people are asked whether they will support candidate Jones or not in a forthcoming election. Test the hypothesis whether Jones is favored by at least 50% of the electorate.

55. Suppose we wish to test the H_0 that the mean μ of a normal population with $\sigma^2 = 1$ is μ_0 against the alternative that it is μ_1 where $\mu_1 > \mu_0$. Find the value of k such that $\overline{X} > k$ provides a rejection region with significance level $\alpha = 0.05$ for a random sample size of n. Also, determine the sample size n needed for testing $H_0 : \mu = 10$ versus $H_1 : \mu = 11$ by this procedure so that $\beta = 0.05$.

56. Suppose we have a random sample of size 25 from $N(\mu, \sigma^2)$ where both parameters are unknown. We find that the sample variance is 14.7. Test at the 5% significance level $H_0 : \sigma^2 = 9$ versus $H_1 : \sigma^2 > 9$.

57. The weights were obtained for 16 male babies born to mothers on a special vitamin supplement. The sample standard deviation of these weights was found to be 0.657kg. Test the claim that the sample comes from a population with a standard deviation equal to 0.470kg at the 5% significance level.

58. A water official insists that the average daily household water use in a certain county is at least 400 gallons. To check this claim, a random sample of 25 households was checked. The average of those sampled was 367 with a sample standard deviation of 62. Is this consistent with the official's claim?

59. The mean response time of a species of pigs to a stimulus is 0.8 second. Twenty-eight pigs were given 2 ounces of alcohol and then tested. If their average response time was 1.0 second with a standard deviation of 0.3 second, can we conclude that alcohol affects the mean response time? Use the 5 percent level of significance.

60. A standard drug is known to be effective in 72 percent of cases in which it is used to treat a certain infection. A new drug has been developed, and testing has found it to be effective in 42 cases out of 50. Is this strong enough evidence to prove that the new drug is more effective than the old one?

61. It has been "common wisdom" for some time that 22 percent of the population have a firearm at home. In a recently concluded poll, 54 out of 200 randomly chosen people were found to have a firearm in their homes. Is this strong enough evidence, at the 5 percent level of significance, to disprove common wisdom?

62. A packaging line fills nominal 32-ounce tomato juice jars with an actual mean of 32.3 ounces. The process should have a standard deviation smaller than 0.15 ounces per jar. Samples of 61 jars are regularly taken to test the process. One such sample yields a sample mean of 32.28 ounces and a standard deviation of 0.132 ounce. Does this indicate that $\sigma < 0.15$ (use 5 percent level of significance)?

63. A certain part of a small assembly should have a diameter of 4mm, and a maximum standard deviation of 0.011mm is allowed by specifications. A random sample of 5 parts show the following: 3.952, 3.978, 3.979, 3.984, 3.987. Can the hypothesis that $\sigma > 0.011$ be supported at 5 percent level of significance?

64. The following data are the scores obtained in a certain test for 15 randomly selected males: 51, 62, 53, 64, 58, 61, 63, 73, 66, 59, 67, 50, 62, 62, 59. Assuming the data to have the $N(\mu, \sigma^2)$ distribution, test at the 5% significance level $H_0 : \sigma^2 = 100$ versus $H_0 : \sigma^2 < 100$.

65. The following are the values of independent samples from two different populations.

    Sample X: 122, 114, 130, 165, 144, 133, 139, 142, 150.
    Sample Y: 108, 125, 122, 140, 132, 120, 137, 128, 138.

    Let μ_X and μ_Y be the respective means of the two populations. Find the p-value of the test of the null hypothesis $H_0 : \mu_X \leq \mu_Y$ against the alternative $H_1 : \mu_X > \mu_Y$ when the population standard deviations are $\sigma_X = 10$ and: a) $\sigma_Y = 5$, b) $\sigma_Y = 10$, c) $\sigma_Y = 20$.

66. The value received at a certain message receiving station is equal to the value sent plus a random error that is normal with mean 0 and standard deviation 2. Two messages, each consisting of a single value, are to be sent. Because of the random error, each message will be sent 9 times. Before reception, the receiver is fairly certain that the first message value will be less than or equal to the second. Should this hypothesis be rejected if the average of the values relating to message 1 is 5.6 whereas the average of those relating to message 2 is 4.1? Use the 1 percent level of significance.

67. Data were collected to determine if there is a difference between the mean IQ scores of urban and rural students in upper Michigan. A random sample of 100 urban students yielded a sample mean score of 102.2 and a sample standard deviation of 11.8. A random sample of 60 rural students yielded a sample mean score of 105.3 with a sample standard deviation of 10.6. Are the data significant enough, at the 5 percent level, for us to reject the hypothesis that the mean scores of urban and rural students are the same? Perform the test for the case where the variances are equal and for the case where they are not equal.

68. In the above problem, are the data significant enough, at the 1 percent level, to conclude that the mean score of rural students in upper Michigan is greater than that of urban students?

69. To see whether there are any differences in starting salaries for women and men law school graduates, a set of eight law firms was selected. For each of these firms a recently hired woman and a recently hired man were randomly chosen. The following starting salary information resulted from interviewing those chosen.

Company	1	2	3	4	5	6	7	8
Woman's salary	52	53.2	78	75	62.5	72	39	49
Man's salary	54	55.5	78	81	64.5	70	42	51

Use the above data to test the hypothesis, at the 10 percent level of significance, that the starting salary is the same for both sexes.

70. The following data give the marriage rates per 1000 population in a random sample of countries for 1987 and 1989.

Country	1987 Rate	1989 Rate
Belgium	5.8	6.4
Finland	5.4	5.1
Greece	6.3	6.0
Israel	6.9	7.0
New Zealand	6.0	6.1
Norway	5.0	4.9
Switzerland	6.6	6.8
United States	9.9	9.7
Yugoslavia	7.0	6.7

Test the hypothesis that the worldwide marriage rates in 1987 are no greater than those in 1989. Use the 5 percent level of significance.

71. The American Cancer Society recently sampled 2500 adults and determined that 738 of them were smokers. A similar poll of 2000 adults carried out in 1986 yielded a total of 640 smokers. Do these figures prove that the proportion of adults who smoke has decreased since 1986? Use the 5 percent level of significance.

72. Suppose a random sample of 480 heart-bypass operations at hospital A showed that 72 patients did not survive, whereas a random sample of 360 operations at hospital B showed that 30 patients did not survive. Find the p-value of the test of the hypothesis that the survival probabilities are the same at the two hospitals.

73. Life threatening heart arrhythmias can be predicted from an electrocardiogram by measuring the lengths of QT intervals (the distances from the starts of the Q-waves to the starts of the T-waves). Two different calipers, A and B, were used to measure a set of 10 QT intervals. The sample variances for the two calipers were 831 and 592, respectively. Do these data suggest that the calipers A and B have different variability? Use the 5 percent level of significance.

74. A restaurant adds a new commercial oven to its kitchen. It is hoped that the new oven has more evenly distributed heat than the current oven. The ovens are heated to 350F, using a thermostat control, and temperature readings are obtained from thermometers placed at 9 locations in each oven, yielding the following data:

    Current oven: $m = 9$, $\bar{x} = 352.4$ and $s_X = 2.3$
    New oven: $n = 9$, $\bar{y} = 350.2$ and $s_Y = 1.1$

    Test $H_0 : \sigma_X^2 = \sigma_Y^2$ versus $H_1 : \sigma_X^2 > \sigma_Y^2$ using the 5 percent level of significance.

75. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ where σ^2 is known. Consider testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$ using the procedures:

    - reject H_0 if $\sqrt{n}(\overline{X} - \mu_0)/\sigma > z_\alpha$;
    - reject H_0 if $(X_1 - \mu_0)/\sigma > z_\alpha$;
    - select an integer i at random between 0 and 99 and reject H_0 if $0 \leq i \leq 4$.

    Find α, β and the power function for each of the procedures.

76. The proportion p of defective items in a large population is unknown. We wish to test $H_0 : p = 0.2$ versus $H_1 : p \neq 0.2$. Suppose a random sample of 20 items is drawn from the population. Let Y be the number of defective items in the sample. Consider the test which rejects H_0 if and only if $Y \geq 7$ or $Y \leq 1$. Find the power function $\Pi(p)$ of this test when $p = 0, 0.2, 0.4, 0.6, 0.8, 1$. What is the level of significance of the test?

77. In a sequence of Bernoulli trials the probability p of success at each trial is unknown. Let X be the number of trials up to and including the first success (X has the geometric distribution). We wish to test $H_0 : p = 0.1$ versus $H_1 : p = 0.2$. Consider the test procedure that rejects H_0 if and only if $X \leq 5$. Find the probabilities of a type I error and a type II error of this test.

78. Suppose we wish to test $H_0 : \mu = 0$ versus $H_1 : \mu = 4$ on the basis of a random sample $X_1, X_2, \ldots, X_n \sim N(\mu, 4)$ at the level $\alpha = 0.05$. Find an expression for the probability of a type II error using the usual test procedure.

79. Let $X_1, X_2, \ldots, X_n$ be a random sample from the $N(\mu, 4)$ distribution. Show that the power function for a test of $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$ at level α is $\Pi(\mu^{'}) = \Phi(-z_{\alpha/2} - \mu^{'}\sqrt{n}/2) + 1 - \Phi(z_{\alpha/2} - \mu^{'}\sqrt{n}/2)$, where $\Phi(\cdot)$ is the cdf of the standard normal distribution.

80. Suppose we wish to test $H_0 : \mu = 100$ versus $H_1 : \mu < 100$ on the basis of a random sample $X_1, X_2, \ldots, X_n \sim N(\mu, 20)$ at the level $\alpha = 0.05$. If we require that with probability at least 0.9 the standard test procedure will reject H_0 when in fact the true value of μ is 95, how large should n be?

81. A machine manufactures bolts whose diameters have the $N(\mu, \sigma^2)$ distribution. Write down the standard procedure for testing $H_0 : \sigma = 10^{-2}$cm versus $H_1 : \sigma > 10^{-2}$cm at the level $\alpha = 0.05$ based on a random sample of size $n = 35$. If in fact $\sigma = 2 \times 10^{-2}$cm express the probability of a type II error in terms of the cdf of a χ^2 random variable.

82. Suppose that a random sample from a normal distribution with known variance σ^2 is to be used to test the hypothesis $H_0 : \mu = \mu_0$ against the alternative $H_1 : \mu = \mu_1$ (where $\mu_1 < \mu_0$) and that the probabilities of Type I and Type II errors have pre-assigned values a and B, respectively. Show that the required sample size is

$$n \quad = \quad \frac{\sigma^2 \left(z_\alpha + z_B\right)^2}{\left(\mu_0 - \mu_1\right)^2}.$$

83. Let $\overline{X}$ be the mean of a random sample of size n from an $N(\mu, \sigma^2)$ distribution and suppose that σ^2 is known. Show that

$$\left(\overline{X} - z_{\alpha_1}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha_1}\frac{\sigma}{\sqrt{n}}\right)$$

is a $(1-\alpha)$-level confidence interval for μ if α_1 and α_2 satisfy $\alpha_1 + \alpha_2 = \alpha$. Show further that $\alpha_1 = \alpha$, $\alpha_2 = 0$ yields the lower one-sided confidence interval for μ, while $\alpha_1 = 0$, $\alpha_2 = \alpha$ yields the upper one-sided confidence interval for μ.

84. Consider the problem of testing $H_0 : \sigma^2 = \sigma_0^2$ versus $H_0 : \sigma^2 > \sigma_0^2$ at level α using the test statistic that rejects H_0 if

$$\frac{(n-1)S^2}{\sigma_0^2} \quad > \quad \chi^2_{n-1,\alpha},$$

where S^2 is the sample variance. Find an expression for the power of this test in terms of the χ^2_{n-1} distribution if the true $\sigma^2 = c\sigma_0^2$ where $c > 1$.

85. Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with the pdf $f(x) = \lambda \exp(-\lambda x)$. The mean of this distribution is $\mu = 1/\lambda$. Show that $Z = 2(X_1 + \cdots + X_n)/\mu$ has the pdf

$$f(z) \quad = \quad \frac{1}{2^n \Gamma(n)} z^{n-1} \exp\left(-z/2\right)$$

and thus conclude that Z has the chi-squared distribution with degrees of freedom $2n$. Hence, show that

$$\frac{2(X_1 + \cdots + X_n)}{\chi^2_{2n,\alpha/2}} \leq \mu \leq \frac{2(X_1 + \cdots + X_n)}{\chi^2_{2n,1-\alpha/2}}$$

is a $1 - \alpha$ level confidence interval for μ.

86. Suppose $X_1, \ldots, X_n$ are counts of defective items produced on n different days. The number of defectives on any given day is modeled as a Poisson random variable with parameter λ, which is the unknown population mean defectives per day. Since $E(X_i) = \mu = \lambda$ and $Var(X_i) = \sigma^2 = \lambda$ for the Poisson distribution, it follows that $E(\overline{X}) = \lambda$ and $Var(\overline{X}) = \lambda/n$. Furthermore, by the central limit theorem, $(\overline{X} - \lambda)/(\sqrt{\lambda/n})$ has the standard normal distribution. Beginning with the probability statement

$$\Pr\left[-z_{\alpha/2} \leq \frac{\overline{X} - \lambda}{\sqrt{\lambda/n}} \leq z_{\alpha/2}\right] \quad \approx \quad 1 - \alpha$$

show that $[\lambda_L, \lambda_U]$ is a large sample $(1-\alpha)$-level confidence interval for λ, where

$$\lambda_L = \overline{X} + \frac{z_{\alpha/2}^2}{2n} - \sqrt{\frac{\overline{X} z_{\alpha/2}^2}{n} + \frac{z_{\alpha/2}^4}{4n^2}}$$

and

$$\lambda_U = \overline{X} + \frac{z_{\alpha/2}^2}{2n} + \sqrt{\frac{\overline{X} z_{\alpha/2}^2}{n} + \frac{z_{\alpha/2}^4}{4n^2}}.$$

Show that if the the terms of the order of $1/n$ are ignored in comparison to the terms of the order of $1/sqrtn$ when n is large, the confidence interval simplifies to

$$\left[\overline{X} - z_{\alpha/2} \sqrt{\frac{\overline{X}}{n}}, \overline{X} + z_{\alpha/2} \sqrt{\frac{\overline{X}}{n}} \right].$$

87. Let X represent a single observation from the $B(a, 1)$ distribution. Find the most powerful test with significance level $\alpha = 0.05$ to test $H_0 : a = 1$ versus $H_1 : a = 2$.

88. Let $X_1, X_2, \ldots, X_{20}$ be a random sample from the $Bernoulli(p)$ distribution. Find the most powerful test with significance level $\alpha = 0.05$ to test $H_0 : p = 0.2$ versus $H_1 : p = 0.3$.

89. Let $X_1, X_2, \ldots, X_n$ be a random sample from the $N(\mu, 1)$ distribution. Find the most powerful test with significance level $\alpha = 0.05$ to test $H_0 : \mu = 0$ versus $H_1 : \mu = 1$.

90. Let X_1, X_2, X_3 be a random sample from the $Po(\lambda)$ distribution.

    (i) Find the most powerful test at level α for $H_0 : \lambda = 1$ versus $\lambda = 2$.

    (ii) Show the this test is equivalent to rejecting H_0 if and only if $T = X_1 + X_2 + X_3 \geq k$ for some k.

    (iii) Find the value of k when $\alpha = 0.1$. You may assume that T has a $Po(3\lambda)$ distribution.

91. Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(0, \sigma^2)$ distribution.

    (i) Find the most powerful test at level α for $H_0 : \sigma = \sigma_0$ versus $\sigma = \sigma_1$, where $\sigma_0 < \sigma_1$ are constants. Show that the test rejects H_0 if and only if $\sum_{i=1}^{n} (X_i/\sigma_0)^2 > k$ for some k.

    (ii) Find the value of k when $\alpha = 0.05$.

    (iii) Find $\beta = \Pr$ (Type II error) when $n = 10$, $\sigma_0 = 2$ and $\sigma_1 = 3$ in terms of the cdf of a χ^2 random variable.