

Homework 6 Solutions, 09 Dec

- Question 6.82, page 339. If Y is a continuous random variable and m is the median of the distribution, then m is such that $P(Y \leq m) = P(Y \geq m) = 1/2$. If Y_1, Y_2, \dots, Y_n are independent, exponentially distributed random variables with mean β and median m , Example 6.17 implies that $Y_{(n)} = \max(Y_1, \dots, Y_n)$ does not have an exponential distribution. Use the general form of $F_{Y_{(n)}}$ to show that $P(Y_{(n)} > m) = 1 - (.5)^n$.

$F_{Y_{(n)}}(m) = F_{Y_1}(m)^n = (1/2)^n$, and so $P(Y_{(n)} > m) = 1 - F_{Y_{(n)}}(m) = 1 - (.5)^n$.

- Question 6.85, page 339. Let Y_1 and Y_2 be independent and uniformly distributed over the interval $(0, 1)$. Find $P(2Y_{(1)} < Y_{(2)})$.

One can use Theorem 6.5 for the bivariate distribution of two order statistics.

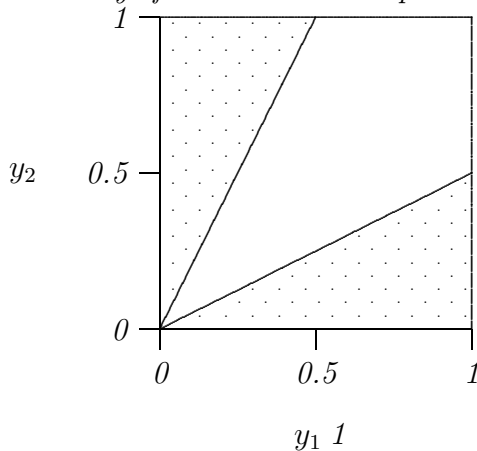
Let $U = Y_{(1)}$ and $V = Y_{(2)}$. By the theorem, $f_{U,V}(u, v) = \binom{2}{0,1,0,1,0} =$

$F(u)^0(F(v) - F(u))^0(1 - F(v))^0 f(u)f(v)$ for $f(u) = 1$ for $0 \in [0, 1]$ and 0 otherwise.

Then $f_{U,V}(u, v) = 1$ if $0 < u < v < 1$. You could have derived this density more directly by noting that probabilities of any set defined in terms of $Y_{(1)}$ and $Y_{(2)}$ come from identical sets in terms of Y_1 and Y_2 , and in terms of Y_2 and Y_1 .

Then $P(2Y_{(1)} < Y_{(2)}) = \int_0^1 \int_0^{v/2} 2 du dv = \int_0^1 v dv = 1/2$.

Alternatively, the probability of interest can be represented as the shaded area in the figure



below; the area is .5.

3. Question 7.11, page 364f. A forester studying the effects of fertilization on certain pin forests in the Southeast is interested in estimating the average basal area of pine trees. In studying basal areas of similar trees for many years, he has discovered that these measurements (in square inches) are normally distributed with standard deviation approximately 4 square inches. If the forester samples $n = 9$ trees, find the probability that the sample mean will be within 3 square inches of the population mean.

This problem's premise is untenable, in that areas cannot be negative, but the normal approximation attributes positive probability to negative values. Let \bar{X} represent the average area. Then $P(-3 \leq \bar{X} - \mu \leq 3) = P(-3\sqrt{9}/4 \leq \sqrt{n}(\bar{X} - \mu)/4 \leq 3\sqrt{9}/4) = P(-2.25 \leq \sqrt{n}(\bar{X} - \mu)/\sigma \leq 2.25) = 0.988 - 0.012 = 0.976$.

4. Question 7.20, page 366.

a. If U has a χ^2 distribution with ν df, find $E(U)$ and $V(U)$.

Since $U = \sum_{i=1}^{\nu} X_i^2$ for the X_i independent standard normals, with $E(X_i^2) = 1$ and $V(X_i^2) = E(X_i^4) - E(X_i^2)^2 = 2$, then $E(U) = \nu$ and $V(U) = 2\nu$.

b. Using the results of Theorem 7.3, find $E(S^2)$ and $V(S^2)$ when Y_1, Y_2, \dots, Y_n is a random sample from a normal distribution with mean μ and variance σ^2 .

By the cited theorem, $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$, and so $E((n-1)S^2/\sigma^2) = n-1$ and $V((n-1)S^2/\sigma^2) = 2(n-1)$. Hence $E(S^2) = \sigma^2$ and $V(S^2) = 2\sigma^4/(n-1)$.

5. Question 7.37, page 369. Let Y_1, Y_2, \dots, Y_5 be a random sample of size 5 from a normal population with mean zero and variance 1 and let $\bar{Y} = (1/5)\sum_{i=1}^5 Y_i$. Let Y_6 be another independent observation from the same population. What is the distribution of

a. $W = \sum_{i=1}^5 Y_i^2$? Why?

$W \sim \chi_5^2$, by definition.

b. $U = \sum_{i=1}^5 (Y_i - \bar{Y})^2$? Why?

$U \sim \chi_4^2$ by Theorem 7.3.

c. $\sum_{i=1}^5 (Y_i - \bar{Y})^2 + Y_6^2$? Why?

This is χ_5^2 , by adding an independent squared normal to the previous part.

960-381– Theory of Probability– Fall, 2021

6. Question 7.74, page 384. According to a survey conducted by the American Bar Association, 1 of every 410 American is a lawyer, but 1 of every 64 residents of Washington, D.C., is a lawyer.
- a. If you select a random sample of 1500 Americans, what is the approximate probability that the sample contains at least one lawyer?

Let N be the number of lawyers in the sample. $N \sim \text{Bin}(1500, 1/410)$, and the probability is $1 - (409/410)^{1500} = 0.9743$.

- b. If the sample is selected from among the residents of Washington, D.C., what is the approximate probability that the sample contains more than 50 lawyers?

If M is the number of lawyers in the sample, then $M \sim \text{Bin}(1500, 1/64)$.

Exactly, $1 - \text{pbinom}(50, 1500, 1/64) = 4.36339e-07$. Approximately,

$Z = (M - 1500 \times (1/64)) / \sqrt{1500 \times 63/64^2}$. The probability is

$1 - \Phi((50 + 1/2 - 23.4375) / 6.682794) = 1 - \Phi(4.04958)$. Evaluate using $1 - \text{pnorm}(4.04958) = 2.565482e-05$.

- c. If you stand on a Washington, D.C., street corner and interview the first 1000 persons who walked by and 30 say that they are lawyers, does this suggest that the density of lawyers passing the corner exceeds the density within the city?

The proportion in this sample is 0.030. The city-wide proportion is $1/64 = 0.0156$, significantly lower. Alternatively, the probability of seeing 30 or more lawyers if the overall proportion is $1/64$ is $1 - \text{pbinom}(29, 1000, 1/64) = 0.0007119629$, which is very small. It is implausible that the proportion of lawyers walking past this corner is $1/64$.

7. Question 7.77, page 384. The manager of a supermarket wants to obtain information about the proportion of customers who dislike a new policy on cashing checks. How many customers should he sample if he wants the sample fraction to be within .15 of the true fraction with probability .98?

Let X represent the number of respondents reporting dissatisfaction.

Let π represent the population proportion of dissatisfied customers. Then approximately, $P(-2.326 \leq (X - n\pi) / \sqrt{n\pi(1 - \pi)} \leq 2.326) = .98$, and

$P(-2.326\sqrt{\pi(1 - \pi)/n} \leq X/n - \pi \leq 2.326\sqrt{\pi(1 - \pi)/n}) = .98$. Find n such that $2.326\sqrt{\pi(1 - \pi)/n} \leq 0.15$; $n \geq 2.326^2\pi(1 - \pi)/.15^2 =$. The worst-case value of π is half, and so $n \geq 0.15 \times 0.5 = 36.0375$. Round up to $n = 37$.

8. Question 7.80, page 384. The median age of residents of the United States is 31 years. If a survey of 100 randomly selected U.S. residents is to be taken, what is the approximate probability that at least 60 will be under 31 years of age.

Let X be the number of respondents under 31. Then $X \sim \text{Bin}(.5, 100)$.

Then $P(X \geq 60)$. This quantity is exactly 0.0287, and approximately

$$1 - \Phi((59.5 - .5 \times 100)/\sqrt{100 \times .5 \times .5}) = 1 - \Phi(9.5/5) = 1 - \Phi(1.9) = 0.0284.$$