200B HW#8 solution

8.1 The Sampling Distribution of a Statistic

1. Suppose that a random sample $X_1, \dots, X_n$ is to be taken from the uniform distribution on the interval $[0, \theta]$ and that θ is unknown. How large a random sample must be taken in order that

$$\Pr(|\max\{X_1, \dots, X_n\} - \theta| \le 0.1\theta) \ge 0.95,$$

for all possible θ?

solution The c.d.f. of $U = \max\{X_1, \dots, X_n\}$ is

$$F(u) = \begin{cases} 0 & \text{for } u \le 0, \\ (u/\theta)^n & \text{for } 0 < u < \theta, \\ 1 & \text{for } u \ge \theta, \end{cases}$$

Since $\Pr(U \le \theta) = 1$, the event that $|U - \theta| \le 0.1\theta$ is the same as the event that $U \ge 0.9\theta$. The probability of this is $1 - F(0.9\theta) = 1 - 0.9^n$. In order for this to be at least 0.95, we need $n \ge \log(0.05)/\log(0.9) = 28.43$. So $n \ge 29$ is needed.

4. Suppose that a random sample is to be taken from normal distribution with unknown mean θ and standard deviation 2. How large a random sample must be taken in order that $\Pr(|\bar{X} - \theta| \le 0.1) \ge 0.95$ for every possible value of θ?

solution It is known that $\bar{X}_n$ has the normal distribution with mean θ and variance $4/n$. Hence the random variable $Z = (\bar{X} - \theta)/(2/\sqrt{n})$ will has the standard normal distribution. Therefore,

$$\Pr(|\bar{X} - \theta| \leq 0.1) = Pr(|Z| \leq 0.05\sqrt{n}) = 2\Phi(0.05\sqrt{n}) - 1.$$

Therefore, this value will be at least 0.95 if and only if $\Phi(0.05\sqrt{n}) \geq 0.975$. It is found from a table of values of Φ that we must have $0.05\sqrt{n} \geq 1.96$. Therefore, we must have $n \geq 1536.64$ or, since n must be an integer, 1537.

9. Let $X_1, \ldots, X_n$ be a random sample from the exponential distribution with parameter θ. Find the c.d.f. for the sampling distribution of the MLE of θ.

solution The joint pdf is $f_n(\mathbf{x}|\theta) = \theta^n e^{-\theta \sum_{i=1}^n X_i}$ and

$$\begin{aligned}
L(\theta) &= n \log(\theta) - \theta \sum_{i=1}^n X_i, \\
\frac{\partial L(\theta)}{\partial \theta} &= \frac{n}{\theta} - \sum_{i=1}^n X_i, \\
\frac{\partial^2 L(\theta)}{\partial \theta^2} &= -\frac{n}{\theta^2} < 0.
\end{aligned}$$

The MLE is $\hat{\theta} = n/T$, where $T = \sum_{i=1}^n X_i$ follows the gamma distribution with parameters n and θ by property 3 in lecture notes page 44. Let $G(\cdot)$ denote the c.d.f. of the sample distribution of T, that is $G(\cdot)$ is the c.d.f of a Gamma(n, θ) distribution. Let $H(\cdot)$ be the sampling distribution of $\hat{\theta}$. Then $H(t) = 0$ for $t \leq 0$, and for $t > 0$,

$$H(t) = \Pr(\hat{\theta} \leq t) = \Pr\left(\frac{n}{T} \leq t\right) = \Pr\left(T \geq \frac{n}{t}\right) = 1 - G\left(\frac{n}{t}\right).$$

8.2 The Chi-Square Distributions

4. Suppose that a point (X, Y) is to be chosen at random in the $xy-$plane, where X and Y are indeoendent random variables and each has the standard normal distribution. If a circle is drawn in the xy-plan with its center at the origin, what is the radius of the smallest circle that can be chosen in order for there to be probability 0.99 that the point (X, Y) will lie inside the circle?

solution Let r denote the radius of the circle. The point (X, Y) will lie inside the circle if and only if $X^2 + Y^2 < r^2$. Also, $X^2 + Y^2$ has a χ^2 distribution with degrees of freedom 2. It is found from the chi-square table that $\Pr(\chi_2^2 \leq 9.210) = 0.99$. Therefore, we must have $r^2 \geq 9.210$ and the smallest circle is of radius $r = \sqrt{9.21}$.

6. When the motion of a microscopic particle in a liquid or a gas is observed, it is seen that the motion is irregular because the particle collides frequently with other particles. The probability model for this motion, which is called *Brownian motion*, is as follows: A coordinate system is chosen in the liquid or gas. Suppose that the particle is at the origin of this coordinate system at time $t = 0$, and let (X, Y, Z) denote the coordinates of the particle at any time $t > 0$. The random variables X, Y, and Z are i.i.d., and each of them has the normal distribution with mean 0 and variance $\sigma^2 t$. Find the probability that at time $t = 2$ the particle will lie within a sphere whose center is at the origin and whose radius is 4σ.

solution We need to calculate $P(\sqrt{X^2 + Y^2 + Z^2} \leq 4\sigma)$. At time 2, each of the independent variables X, Y and Z has a normal distribution with mean 0 and variance $2\sigma^2$. Therefore, each of the variables $X/(\sqrt{2}\sigma)$, $Y/(\sqrt{2}\sigma)$ and $Z/(\sqrt{2}\sigma)$ has a standard normal distribution. Hence, $V = (X^2 + Y^2 + Z^2)/(2\sigma^2)$ has a χ^2 distribution with three degrees of freedom. It follows that

$$P(\sqrt{X^2 + Y^2 + Z^2} \leq 4\sigma) = P(X^2 + Y^2 + Z^2 \leq 16\sigma^2) = P(V \leq 8) = 0.954.$$

9. Suppose that $X_1, \ldots, X_n$ form a random sample from the normal distribution with mean μ and variance σ^2. Find the distribution of $\frac{n(\bar{X}_n - \mu)^2}{\sigma^2}$.

solution It is known that $\bar{X}_n$ has the normal distribution with mean μ and variance σ^2/n. Therefore, $(\bar{X}_n - \mu)/(\sigma/\sqrt{n})$ has a standard normal distribution and the square of this variable has the χ^2 distribution with one degree of freedom.

8.3 Joint Distribution of the Sample Mean and Sample Variance

4. Suppose that the random variables X_1, X_2 and X_3 are i.i.d., and that each has the standard normal distributon. Also suppose that

$$
\begin{aligned}
Y_1 &= 0.8X_1 + 0.6X_2, \\
Y_2 &= \sqrt{2}(0.3X_1 - 0.4X_2 - 0.5X_3), \\
Y_3 &= \sqrt{2}(0.3X_1 - 0.4X_2 + 0.5X_3).
\end{aligned}
$$

Find the joint distribution of Y_1, Y_2 and Y_3.

solution The 3×3 matrix of the transformation from (X_1, X_2, X_3) to (Y_1, Y_2, Y_3) is

$$
A = \begin{bmatrix} 0.8 & 0.6 & 0 \\ (0.3)\sqrt{2} & -(0.4)\sqrt{2} & -(0.5)\sqrt{2} \\ (0.3)\sqrt{2} & -(0.4)\sqrt{2} & (0.5)\sqrt{2} \end{bmatrix}
$$

Since the matrix A is orthogonal, it follows from Theorem 8.3.4 that Y_1, Y_2 and Y_3 are independent and each has a standard normal distribution.

8.4 The t Distributions

3. Suppose that the five random variables $X_1, \ldots, X_5$ are i.i.d. and that each has the standard normal distribution. Determine a constant c such that the random variable $\frac{c(X_1+X_2)}{(X_3^2+X_4^2+X_5^2)^{1/2}}$ will have a t distribution.

solution $X_1 + X_2$ has the normal distribution with mean 0 and variance 2. Therefore, $Y = (X_1 + X_2)/\sqrt{2}$ has a standard normal distribution. Also, $Z = X_3^2 + X_4^2 + X_5^2$ has the χ^2 distribution with 3 degrees of freedom, and Y and Z are independent. Therefore, $\frac{Y}{(Z/3)^{1/2}}$ has the t distribution with 3 degrees of freedom. Thus, $c = \sqrt{3/2}$.

5. Suppose that the random variables X_1 and X_2 are independent and that each has the normal distribution with mean 0 and variance σ^2. Determine the value of $\Pr\left[\frac{(X_1+X_2)^2}{(X_1-X_2)^2} < 4\right]$.
Hint:

$$(X_1 - X_2)^2 = 2\left[\left(X_1 - \frac{X_1 + X_2}{2}\right)^2 + \left(X_2 - \frac{X_1 + X_2}{2}\right)^2\right].$$

solution Let $\bar{X}_2 = (X_1 + X_2)/2$ and $S_2^2 = \sum_{i=1}^{2}(X_i - \bar{X}_2)^2 = (X_1 - X_2)^2/2$. Then

$$W = \frac{(X_1 + X_2)^2}{(X_1 - X_2)^2} = \frac{2\bar{X}_2^2}{S_2^2}.$$

It follows from Eq. (8.4.4) that $U = \sqrt{2}\bar{X}_2/\sqrt{S_2^2}$ has the t distribution with one degree of freedom. Since $W = U^2$, we have

$$P(W < 4) = P(-2 < U < 2) = 2P(U < 2) - 1 = 2(0.852) - 1 = 0.704.$$

The second equality holds since t-distribution is symmetric around 0. Hence $P(U < 2) = P(U > -2)$ and the c.d.f. can be calculated using the tables/ software.

8.9 Supplementary Exercises

4. Suppose that X_1 and X_2 are independent random variables, and that each has the normal distribution with mean 0 and variance σ^2. Show that $(X_1 + X_2)/(X_1 - X_2)$ has the t distribution with one degree of freedom.

solution Let $Z_i = \frac{X_i}{\sigma}$. Then Z_1 and Z_2 are independent and each has a standard normal distribution. Next, let $Y_1 = (Z_1 + Z_2)/\sqrt{2}$ and $Y_2 = (Z_1 - Z_2)/\sqrt{2}$. Then the transformation matrix is.

$$A = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}.$$

Since it is orthogonal, Y_1, Y_2 are independent and so are $X_1 + X_2$, $X_1 - X_2$. Further, if we let

$$Y_1 = \frac{1}{\sqrt{2}\sigma}(X_1 + X_2) \text{ and } Y_2 = \frac{1}{\sqrt{2}\sigma}(X_1 - X_2),$$

then Y_1 and Y_2 have standard normal distributions. It follows, therefore, from Exercise 18 of Section 5.6 that Y_1/Y_2 has a Cauchy distribution, which is the same as the t distribution with one degree of freedom. But

$$\frac{Y_1}{Y_2} = \frac{X_1 + X_2}{X_1 - X_2},$$

so the desired result has been established. This result could also have been established by a direct calculation of the desired p.d.f.

5. Suppose that $X_1, \dots, X_n$ form a random sample from the exponential distribution with parameter β. Show that $2\beta \sum_{i=1}^n X_i$ has the χ^2 distribution with $2n$ degrees of freedom.

solution Since X_i has the exponential distribution with parameter β, it follows that $2\beta X_i$ has the exponential distribution with parameter $1/2$. But this exponential distribution is the χ^2 distribution with 2 degrees of freedom. Therefore, the sum of the i.i.d. χ^2 random variables $\sum_{i=1}^n 2\beta X_i = 2\beta \sum_{i=1}^n X_i$ has a χ^2 distribution with $2n$ degrees of freedom.