01:640:481 - Exam 1 Review

Pranav Tikkawar

October 14, 2024

1. Question 1. Suppose $X_1, X_2, X_3, X_4 \sim N(7,5)$ are independent.

   (a) Express r.v $A = (X_1 - 7)^2 + (X_2 - 7)^2 + (X_3 - 7)^2 + (X_4 - 7)^2$ as a multiple of a χ^2 random variable. That is find $A = cY$ where $Y \sim \chi^2_{\nu=\nu_1}$.

   (b) Express r.v $B = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2 + (X_4 - \bar{X})^2$ as a multiple of a χ^2 random variable. That is find $B = kM$ where $M \sim \chi^2_{\nu=\nu_2}$.

   > **Solution:** Part (a):
   > We can first consider the standard normal random variable $Z \sim N(0,1)$. where $Z = \frac{X-7}{\sqrt{5}}$. Then $Z^2 = \frac{(X-7)^2}{5}$. We can then consider the sum of squares of Z random variables. That is $Y = Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2$. We know that $Y \sim \chi^2_{\nu=4}$. We can then substitute $Z = \frac{X-7}{\sqrt{5}}$ to get $A = 5Y$.

   > **Solution:** Part (b):
   > We can first recall that the definition of sample variance is $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$. In this case it is $S^2 = \frac{1}{3} \sum_{i=1}^{4} (X_i - 7)^2$. Further we know that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{\nu=n-1}$. We can then substitute $n = 4$ and $\sigma^2 = 5$ to see that $Y = \frac{3B}{5 \cdot 3}$ therefore $B = 5Y$.

2. Question 2. Recall the MFG of Gamma distribution is $\frac{1}{(1-\beta t)^\alpha}$.

   (a) Let X_2, X_3 be independent χ^2 random variables with degrees of freedom $\nu_1 = 2$ and $\nu_2 = 3$. Write down the MFG of X_2, X_3.

   (b) Using the properties of MGF, find the MFG of $X_2 + X_3$.

   (c) is $X_2 + X_3$ a χ^2 random variable? Why?

   (d) Using the same ideas determine the MGF of $B = 7X_2$ and decide if it is a χ^2. If yes, what ν if no, is it gamma distributed? With what α and β?

Solution: Part (a):
We know that the MFG of a χ^2 random variable is generally $\frac{1}{(1-2t)^{\nu/2}}$. Therefore the MFG of X_2 and X_3 are $\frac{1}{(1-2t)^{2/2}}$ and $\frac{1}{(1-2t)^{3/2}}$.

Solution: Part (b):
We know that the MFG of the sum of independent random variables is the product of their MFG. Therefore the MFG of $X_2 + X_3$ is $\frac{1}{(1-2t)^{2/2}(1-2t)^{3/2}} = \frac{1}{(1-2t)^{5/2}}$.

Solution: Part (c):
$X_2 + X_3$ is a χ^2 random variable. This is because the MFG of $X_2 + X_3$ is $\frac{1}{(1-2t)^{5/2}}$ which is the MFG of a χ^2 random variable with $\nu = 5$.

Solution: Part (d):
We know that the MFG of $B = 7X_2$ is $MFG_B(t) = MFG_{X_2}(7t) = \frac{1}{(1-2(7)t)^{2/2}} = \frac{1}{(1-14t)^1}$. This is not a χ^2 random variable. This is a gamma distributed random variable with $\alpha = 1$ and $\beta = 14$.

3. Question 3. Suppose $X_1, X_2, \ldots, X_n$ are IID random variables with an unknown distribution with mean $\mu = 5$, variance $\sigma^2 = 2$. Applying Chebyshev's theorem, determine the smallest sample size n that will guarantee that the sample mean is within the interval $(5 - 0.1, 5 + 0.1)$ with probability at least 0.95.

Solution: We can see Chebyshev's Theorem as $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$.
We can manipulate this to get $P(|X - \mu| < k) \geq 1 - \frac{\sigma^2}{k^2}$.
Applying this to the sample with $\mu = \mu_{\bar{x}} = 5, \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{2}{n}$. We can see that $P(|\bar{X} - 5| < k) \leq 1 - \frac{2}{nk^2}$.
Since we want probability at least 0.95, we have $k = .1$ we can set $1 - \frac{2}{n\cdot.01} \geq 0.95$.
Thus $n \geq \frac{2}{.05\cdot.01}$
Therefore the smallest sample size n that will guarantee that the sample mean is within the interval $(5 - 0.1, 5 + 0.1)$ with probability at least 0.95 is $n = 4000$.

4. Question 4.

   (a) Consider a Random sample of size $n = 10$ from a normal population $N(2,3)$ Find the probability that the sample variance is greater than 1. Leave the answer in the form of $P(Y >?)$ where Y is a χ^2 random variable.

   (b) Continuing previous part, use the relation beween Y and S to determine $Var[S^2]$.

Solution: Part (a):
We want the probability of $P(S^2 > 1)$.
We know that the sample variance is a χ^2 random variable with $n-1$ degrees of freedom.
In this case we can see that $Y = \frac{(n-1)S^2}{\sigma^2} = \frac{9S^2}{3} = 3S^2$.
Therefore $P(S^2 > 1) = P(3S^2 > 3) = P(Y > 3)$.

Solution: Part (b):
We know that $Var[Y] = 2\nu = 2(n-1) = 18$.
We also know that $Var[S^2] = Var[\frac{Y}{3}] = \frac{1}{9}Var[Y] = 2$

---

5. Consider a population with PDF

$$f(x) = \frac{1}{x^2}, x > 1$$

Consider a random sample of size $n = 5$ from this population find the expection of the second order statistic after computing the pdf.

Solution: We know the formula for the rth order statistic is
$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!}f(x)[F_l(x)]^{r-1}[F_u(x)]^{n-r}$.
Where $F_l(x) = \int_{-\infty}^{x} f(x)dx$.
And $F_u = \int_{x}^{\infty} f(x)dx$.
In this case we have

$$f(x) = \frac{1}{x^2}, x > 1$$
$$F_l(x) = \int_{1}^{x} \frac{1}{x^2}dx = 1 - \frac{1}{x}$$
$$F_u(x) = \int_{x}^{\infty} \frac{1}{x^2}dx = \frac{1}{x}$$

Thus the second order statistic of a sample of 5 is $f_2(x) = \frac{5!}{1!(5-2)!}\frac{1}{x^2}[-\frac{1}{x}+1]^1[\frac{1}{x}]^3$.
Simplified it is

$$f_2(x) = 20 \cdot \left[\frac{1}{x^5} - \frac{1}{x^6}\right]$$

Thus to find the expection of this we can integrate over the range $[1, \infty)$.

$$E[X_{(2)}] = \int_1^\infty y f_2(y) dy = \int_1^\infty y \cdot 20 \cdot \left[\frac{1}{y^5} - \frac{1}{y^6}\right] dy$$

$$\begin{aligned}
E[X_{(2)}] &= 20 \int_1^\infty \frac{1}{y^4} - \frac{1}{y^5} dy \\
&= 20 \left[-\frac{1}{3y^3} + \frac{1}{4y^4}\right]_1^\infty \\
&= 20 \left[0 - (-\frac{1}{3} + \frac{1}{4})\right] \\
&= 20 \cdot \frac{1}{12} \\
&= \frac{5}{3}
\end{aligned}$$

6. SKIP FOR NOW

7. Consider a population with PDF $f(x) = -ae^{ax}, x > 0$. where $a < 0$ is a parameter. Determine the estimator for a using the method of maximum likelihood. Your answer will be an expression of the form $\hat{a}$ in terms of $x_1, x_2, \ldots, x_n$.

Solution: We know that the likelihood function is $L(a) = \prod_{i=1}^n f(x_i)$.
In this case

$$L(a) = \prod_{i=1}^n -ae^{ax_i}$$

We can rewrite this as

$$L(a) = -a^n e^{a \sum_{i=1}^n x_i}$$

We can then take the log likelihood function

$$l(a) = \log(L(a))$$
$$= \log(-a^n e^{a \sum_{i=1}^n x_i})$$
$$= \log(-a^n) + \log(e^{a \sum_{i=1}^n x_i})$$
$$= n \log(-a) + a \sum_{i=1}^n x_i$$

Then taking the derivative with respect to a we get

$$\frac{dl(a)}{da} = \frac{d}{da}(n \log(-a) + a \sum_{i=1}^n x_i)$$
$$= \frac{n}{a} + \sum_{i=1}^n x_i$$

Then we can set this to zero and solve for a.

$$0 = \frac{n}{a} + \sum_{i=1}^n x_i$$
$$-\sum_{i=1}^n x_i = \frac{n}{a}$$
$$a = -\frac{n}{\sum_{i=1}^n x_i}$$

We can notice that this is in the form of

$$\hat{a} = -\frac{1}{\bar{X}}$$

8. Through parts below we determine $\Gamma(1/2)$. You may already know what the value is but the point of this question is how to obtain the value. First two parts are reasonably question. Most work will be in the last part.

    (a) Write down the PDF of the standard normal distribution. Use that to conclude the numver of $I = \int_0^\infty e^{-x^2/2} dx$.

    (b) Recall that $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$. substitute and write the integral for $\Gamma(1/2)$.

    (c) Use the substitution $y = x^2/2$ in the previous part and write down the resulting integral in x.

Solution: Part (a):

The PDF of the standard normal distribution is $f(x) = \frac{1}{\sqrt{2\pi}}e^{-(x)^2/2}$.

We can see that $I = \int_0^\infty e^{-x^2/2}dx = \frac{\sqrt{2\pi}}{2}$.

Solution: Part (b):

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy$$

$$\Gamma(1/2) = \int_0^\infty y^{-1/2}e^{-y}dy$$

Solution: Part (c):

We can substitute $y = x^2/2$ to get $dy = xdx$.

We can then rewrite the integral in terms of x.

$$\begin{aligned}
\Gamma(1/2) &= \int_0^\infty (x^2/2)^{-1/2}e^{-x^2/2}xdx \\
&= \int_0^\infty x^{-1}\sqrt{2}e^{-x^2/2}xdx \\
&= \sqrt{2}\int_0^\infty e^{-x^2/2}dx \\
&= \sqrt{2}I \\
&= \sqrt{\pi}
\end{aligned}$$

9. For the uniform distribution in $[0, a]$ (with parameter $a > 0$), is the sample max Y_n an unbiased estimator of a? Show all work and justification.

Solution: We know that the PDF of the uniform distribution is $f(x) = \frac{1}{a}$.

We know that the sample max is $Y_n = \max(X_1, X_2, \ldots, X_n)$.

The pdf of the sample max is $f_{Y_n}(y) = nf(y)[F_u(y)]^{n-1}$. where $F_u(y) = \int_y^\infty f(z)dz$

In this case $F_u(y) = \int_y^a \frac{1}{a}dz = 1 - \frac{y}{a}$.

Therefore $f_{Y_n}(y) = n\frac{1}{a}[1 - \frac{y}{a}]^{n-1}$.

We can then find the expectation of the sample max. $E[Y_n] = \int_0^a yf_{Y_n}(y)dy =$

$\int_0^a y \cdot n \frac{1}{a} [1 - \frac{y}{a}]^{n-1} dy$

We can then substitute $u = 1 - \frac{y}{a}, y = a - au$ to get $du = -\frac{1}{a} dy$

We can then rewrite the integral as $\int_1^0 a(1-u) n u^{n-1} du$

Simplifying this we get $E[Y_n] = a \int_1^0 n u^{n-1} - n u^n du$

Evaluating gets $E[Y_n] = a(u^n - \frac{n u^{n+1}}{n+1}) \Big|_1^0$

This simplifies to $E[Y_n] = a[(0-0) - (1 - \frac{n}{n+1})] = \frac{a}{n+1}$

Clealry $E[Y_n] \neq a$ therefore the sample max is not an unbiased estimator of a.