Math 376 LM (36768) Fall 2021

Test 2 Solutions

1 (10 points)- Let $X_1, \ldots, X_{10}$ be an iid sequence of continuous rv's with median ξ.

(a) Compute the probability that exactly 3 of the X_i's satisfy $X_i < \xi$.

(b) Compute the probability that all of the X_i's satisfy $X_i < \xi$.

1- The indicator rv of the event $\{X_i < \xi\}$ is a $Bern(1/2)$ rv. So the number N of the X_i's satisfy $X_i < \xi$ is a $Bin(10, 1/2)$ rv.

(a) $P(N = 3) = \binom{10}{3}(1/2)^{10} = \frac{120}{1024}$.

(b) $P(N = 10) = (1/2)^{10} = \frac{1}{1024}$.

2 (10 points)-(a) Compute the MGF for the Poisson distribution with mean m

(b) Let X_1 and X_2 have Poisson distributions each with mean m.

(i) Show that if X_1 and X_2 are independent, then $X = X_1 + X_2$ has a Poisson distribution. Explain where independence is used.

(ii) Show that if $X_1 = X_2$ then $X = X_1 + X_2$ does not have a Poisson distribution. Compute $P(X = 1)$ and $P(X = 2)$.

2- (a) $f_X(n) = e^{-m}\frac{m^n}{n!}$. So

$$M_X(t) = E(e^{tX} = \sum_{k=0}^{\infty} e^{tk} e^{-m}\frac{m^k}{k!} = e^{-m}\sum_{k=0}^{\infty}\frac{(me^t)^k}{k!}.$$

So $M_X(t) = exp(m(e^t - 1))$.

(b) (i) If $X_1 \sim Poiss(m_1), X_2 \sim Poiss(m_2)$ then $M_X(t) = E(e^{tX_1+tX_2}) = E(e^{tX_1}e^{tX_2})$. For independent rv's the expectation of the product is the product of the expectations and so

$$M_X(t) = M_{X_1}(t)M_{X_2}(t) = exp((m_1 + m_2)(e^t - 1)).$$

Because the MGF characterizes the distribution, $X \sim Poiss(m_1+m_2)$.

Alternatively, $P(X = n) = sum_{k=0}^{n} P(X_1 = k \& X_2 = n - k)$. By independence this equals

$$\sum_{k=0}^{n} e^{-m_1}\frac{m_1^k}{k!}e^{-m_2}\frac{m_1^{n-k}}{(n-k)!} = \frac{e^{-(m_1+m_2)}}{n!}\sum_{k=0}^{n}\binom{n}{k}m_1^k m_2^{n-k}.$$

By the Binomial Theorem this equals $e^{-(m_1+m_2)}\frac{(m_1+m_2)^n}{n!}$. So $X \sim Poiss(m_1+m_2)$.

If $m_1 = m_2 = m$ then $X \sim Poiss(2m)$.

(ii) If $X_1 = X_2$ then $X = 2X_1$ has range only the even nonnegative integers and so is not Poisson. In particular, $P(X = 1) = 0$ and $P(X = 2) = P(X_1 = 1) = me^{-m}$.

3 (15 points)- If $X \sim \Gamma(\alpha, \beta)$ then $f_X(x)dx = \frac{1}{\Gamma(\alpha)}(x/\beta)^\alpha e^{-(x/\beta)}\frac{dx}{x}$.. Also the MGF is given by: $M_X(t) = (1-\beta)^{-\alpha}$ for $t < \beta^{-1}$. Recall that a $\chi^2(r)$ distribution has a $\Gamma(r/2, 2)$ distribution.

(a) Assume $Z \sim \mathcal{N}(0,1)$, i.e. it is a standard normal.

(i) Compute $E(|Z|)$.

(ii) Compute the cdf and pdf of Z^2. Explain why this shows that $Z^2 \sim \chi^2(1)$ and that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

(b) If $X_1, \ldots, X_n$ are independent rvs each with distribution $\Gamma(1,\theta)$, explain why $\frac{2}{\theta}\sum_{i=1}^{n} X_i \sim \chi^2(2n)$.

3- (a) (i) $E(|Z|) = \int_{-\infty}^{\infty}|x|(2\pi)^{-1/2}e^{-\frac{x^2}{2}}dx$.

$$= (\frac{2}{\pi})^{1/2}\int_{x=0}^{x=\infty} xe^{-\frac{x^2}{2}}dx = (\frac{2}{\pi})^{1/2}\int_{u=0}^{u=\infty} e^{-u}du.$$

So $E(|X|) = (\frac{2}{\pi})^{1/2}$.

(ii) For $x \geq 0$, $P(Z^2 < x) = P(-\sqrt{x} < Z < \sqrt{x}) = \frac{2}{\pi})^{1/2}\int_{t=0}^{t=\sqrt{x}} e^{-\frac{t^2}{2}}dt$. Let $t^2 = u$ so that $dt = du/2t = du/2\sqrt{u}$. So

$$P(Z^2 < x) = (\frac{1}{\pi})^{1/2}\int_{u=0}^{u=x}(u/2)^{1/2}e^{-(u/2)}\frac{du}{u}.$$

So $Z^2 \sim \Gamma(1/2, 2)$ and so the constant $\Gamma(1/2) = \sqrt{\pi}$.

(b) If $X_1 \sim \Gamma(\alpha_1, \beta), X_2 \sim \Gamma(\alpha_2, \beta)$ are independent, then for $X = X_1 + X_2$

$$M_X(t) = M_{X_1}(t)M_{X_2}(t) = (1-\beta)^{-\alpha_1}(1-\beta)^{-\alpha_2} = (1-\beta)^{-(\alpha_1+\alpha_2)}.$$

So $X \sim \Gamma(\alpha_1 + \alpha_2, \beta)$.

In particular (by induction) the sum of n independent $\Gamma(1, \theta)$ distributions has a $\Gamma(n, \theta)$.

If $X \sim \Gamma(\alpha, \beta)$ and $k > 0$, then $kX \sim \Gamma(\alpha, k\beta)$. Hence,

$$\frac{2}{\theta} \sum_{i=1}^{n} X_i \sim \Gamma(n, 2) \sim \chi^2(2n).$$

4 (15 points)- Assume that $X_1, ..., X_n$ are i. i. d. with the normal density $f(x|\theta) = (2\pi\theta)^{-\frac{1}{2}} e^{-\frac{x^2}{2\theta}}$.

(a)Compute the MLE estimate for θ and determine whether or not it is unbiased.

(b)Assume that the sample size $n = 10$.

Construct a 90% confidence interval for θ, using the realized value of the MLE $\hat{\theta}$ from (a).

4- (a) With $L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\theta) - \frac{\Sigma_{i=1}^n x_i^2}{2\theta}$, $\partial L/\partial\theta = 0$ when $\hat{\theta} = \frac{\Sigma_{i=1}^n x_i^2}{n}$. Since the mean is zero, $E(\hat{\theta})$ is the variance which is θ. Hence, the estimate is unbiased.

(b) Each $X_i/\sqrt{\theta} \sim \mathcal{N}(0, 1)$. So $X_i^2/\theta \sim \chi^2(1)$ (see problem 3). Hence $W = \frac{n}{\theta}\hat{\theta} \sim \chi^2(n)$. We choose $q_{.05}$ and $Q_{.95}$ so that $P(W < q_{.05}) = 1 - P(W < Q_{.95}) = .05$. From the table with $n = 10$, $q_{.05} = 3.95, Q_{.95} = 20.48$. So the 90% confidence interval for θ is $(\frac{\hat{\theta}}{2}, \frac{\hat{\theta}}{.4})$.

5 (10 points)- Let $\bar{X}$ be the sample mean of a sample of size n from a distribution that is normal with $\mathcal{N}(\mu, 16)$. What is the smallest sample size n so that $(\bar{x} - 1, \bar{x} + 1)$ is a 90% confidence interval.

5- With a known standard deviation σ, the $(1 - \alpha)100\%$ confidence interval is $(\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} - z_{\alpha/2}\sigma/\sqrt{n})$. We use the $z-$score with $Prob(Z \leq z_{.05}) = .95$ which is $z_{.05} = 1.65$ and we use the given standard deviation $\sigma = 4$. We want the smallest integer n such that $z_{\alpha/2}\sigma/\sqrt{n} \leq 1$ and so that n is the smallest integer at least as large as $(4 \cdot 1.65)^2$.

6 (10 points)- With a sample of 10 trials of a $Bern(p)$ rv, let X be the number of successes. We use the critical region $C = \{X \leq 2\} \cup \{X \geq 8\}$

to test $H_0 : p = 1/2$ against the alternative $H_1 : p \neq 1/2$. Compute the size and power function of the test.

6- The power function $\gamma_p = P_p(C)$. $X \sim Bin(10, p)$.

$$P_p(C) = (1-p)^{10} + 10p(1-p)^9 + 45p^2(1-p)^8 + 45p^8(1-p)^2 + 10p^9(1-p) + p^{10}.$$

The size is the value with $p = 1/2$ which is $112(1/2)^{10} = \frac{112}{1024}$.

7 (15 points)- $X_1, \ldots, X_n$ is a sample with distribution $Unif(0, \theta)$.

(a) Compute the MLE $\hat{\theta}$ for θ and the MLE for the mean.

(b) For $M = \hat{\theta}$, compute the range and the cdf of $Z = \frac{M}{\theta}$ and show that this is a known distribution, i.e. it does not depend on θ. Use it as a pivot variable to obtain a 90% confidence interval for θ (Hint: Compute q so that $P(Z \leq q) = .1$)

7- (a) $L(\theta) = (1/\theta^n)$ for $0 \geq x_1, \ldots, x_n \leq \theta$ and $= 0$ otherwise. Since $(1/\theta^n)$ is decreasing in θ, the MLE $\hat{\theta} = M = \max(X_1, \ldots, X_n)$.

The mean value of is $\theta/2$ and so the MLE for the mean is $M/2$ and not $\bar{X}$.

(b) The range of M is $(0, \theta)$ and so the range of Z is $(0, 1)$. For $t \in (0, 1)$, $F_Z(t) = F_M(t\theta) = F_X(t\theta)^n$. $F_X(x) = x/\theta$ for $0 < x < \theta$. So $F_Z(t) = t^n$. With $q = .1^{1/n}$, $P(Z \leq q) = .1$ and so $P(Z > q) = .9$. That is, $P(\frac{M}{q} > \theta) = .9$. On the other hand, $P(\theta > M) = 1$. Thus, the 90% confidence interval is

$$(\max(x_1, \ldots, x_n)/.1^{1/n}, \max(x_1, \ldots, x_n)).$$

Notice that as n tends to infinity, $.1^{1/n} < 1$ increases with limit 1, and so $1/.1^{1/n} > 1$ decreases with limit 1.

8- Assume that $f(\theta; x) = c_\theta x^\theta (1 - x)$ for $0 < x < 1$, and equal to zero otherwise ($0 < \theta$). Compute c_θ.

(a) Using a sample $X_1, \ldots X_n$ describe a UMP test of the hypothesis $\theta = \theta_0$ against $\theta > \theta_0$. That is, describe a best critical region of its size. Justify your description.

(b) With $n = 2$ write the definite integral which computes its power function (you need not evaluate the integral).

$\int_0^1 x^\theta(1-x)dx = \frac{1}{\theta+1} - \frac{1}{\theta+2} = \frac{1}{(\theta+1)(\theta+2)}$. So $c_\theta = (\theta+1)(\theta+2)$.

(a) $\Lambda(\theta_0, \theta_1) = [\frac{(\theta_1+1)(\theta_1+2)}{(\theta_0+1)(\theta_0+2)}]^n (\prod_{i=1}^n x_i)^{\theta_0-\theta_1}$.

With $\theta_1 > \theta_0$ this is a decreasing function of the statistic $T = \prod_{i=1}^n X_i$. That is, Λ satisfies the monotone likelihood condition and so for any $1 > c > 0$ there exists a k such that $C = \{T > c\} = \{\Lambda < k\}$. Therefore, C is a UMP test for θ_0 against any $\theta > \theta_0$. It is a best critical region of size $P_{\theta_0}(C)$.

(b) With $n = 2$ the power function of C is

$$\gamma_C(\theta) = P_\theta(x_1 x_2 > c) = c_\theta^2 \int_c^1 \int_{c/x}^1 x^\theta(1-x)y^\theta(1-y)dydx.$$

For $Z \sim \mathcal{N}(0,1)$

z	$Prob(Z \leq z)$
.53	.70
.68	.75
.85	.80
1.04	.85
1.29	.90
1.65	.95

For $W \sim \chi^2(10)$

q	$Prob(W \leq q)$
3.94	.05
4.87	.10
15.99	.90
20.48	.95

For $W \sim \chi^2(20)$

q	$Prob(W \leq q)$
10.45	.05
12.44	.10
28.41	.90
31.41	.95