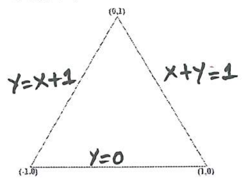
Statistics 511 Final Exam
Summer 2017

Name: PIYAS CHAKRABORTY

Purdue ID:

You may have two sheets of paper with notes handwritten in your own handwriting to take this test. Please show all your work to receive full credit. You may use the attached tables but please mention that whenever you do so. Report all numerical answers to at least three correct decimal places. You need to complete problems 1 and 4, any one out of problems 2 and 3 and any two out of problems 5, 6 and 7. When you finish your exam, turn it in and show your Purdue ID. You are expected to uphold the Honor Code of Purdue University.

1. (12 points) X and Y are jointly Uniformly distributed on the triangular region D enclosed by the three straight lines Y=0, X+Y=1 and Y=X+1 as shown below.



(a) (1) Write down the marginal supports of X and Y.
(b) (1) Write down the joint probability density function f(x,y).
(c) (1) Are X and Y independent? Justify your answer.
(d) (1) Calculate P(XY<0).
(e) (2) Calculate P(Y<X).
(f) (2) Compute the marginal density of Y.
(g) (1) Calculate P(Y>0.5).
(h) (3) Calculate P(X>0.2|Y=0.3), E(X|Y=0.3) and the conditional median of X given Y=0.3.

(a) $D_X = (-1, 1)$, $D_Y = (0, 1)$

(b) Area of $D = 2\cdot\left(\frac{1}{2}\times 1\times 1\right) = 1$

$\therefore f(x,y) = \begin{cases} 1, & \text{if } (x,y)\in D \\ 0, & \text{o.w.} \end{cases}$

(c) Joint support D is NOT rectangular ⇒ X, Y are NOT indep.

(d) $P(XY<0) = P(X<0)$ (as $Y>0$ w.p. 1)

$= 1/2$ (By symmetry)

$= 0.5$

Or, derive as: $\int_{-1}^{0}\int_{0}^{x+1} dy\,dx = 0.5$

(e) $P(Y<X) =$ Area of shaded region

$= 1/4 = 0.25$ (Symmetry within the right half of D)

[Can be also derived as: $\int_0^{1/2}\int_0^{x} dy\,dx + \int_{1/2}^{1}\int_0^{1-x} dy\,dx = 0.25$]

(f) $f_Y(y) = \int_{x\in D_X} f(x,y)\,dx = \int_{y-1}^{1-y} 1\cdot dx = 2(1-y)$, for $0<y<1$

or, $f_Y(y) = 2(1-y)\, I_{(0,1)}(y)$.

(g) $P(Y>0.5) = \int_{0.5}^{1} 2(1-y)\,dy = 2\cdot\left[(1-0.5) - \frac{1}{2}(1-.25)\right] = 1-.75 = 0.25$

(h) $f_{X|Y}(x|0.3) = \frac{f(x,0.3)}{f_Y(0.3)} = \frac{1}{2\times 0.7}$ for $x\in(0.3-1, 1-0.3) = (-0.7, 0.7)$

$\therefore X|Y=0.3 \sim U(-0.7, 0.7) \Rightarrow P(X>0.2|Y=0.3) = \frac{0.7-0.2}{2\times 0.7} = \frac{5}{14} = 0.357$ (Using uniformity)

and, $E(X|Y=0.3) = \eta_{X|Y}(0.5|0.3) = \frac{0.7-0.7}{2} = 0$ (Also, by symmetry)

2. (10 points) Let X and Y have a joint probability density function: $f(x,y)=k/(e^x y^5)$ for $x>0$, $y>2$.

(a) (1) Find out the constant k.
(b) (1) Are X and Y independent? Justify your answer.
(c) (3) Find out V(3X-2Y).
(d) (1) Find out P(Y>3|X>1).
(e) (1) Find out P(X>3|X>1).
(f) (2) Find out the conditional median of X given Y>3 and the unconditional median of X.
(g) (1) Compare the median of X with its mean. What does that say about the skewness of X?

(a) $K\int_0^\infty e^{-x}dx \int_2^\infty \frac{dy}{y^5} = 1 \Rightarrow K\left[-e^{-x}\right]_0^\infty \left[\frac{-1}{4y^4}\right]_2^\infty = 1 \Rightarrow K\cdot 1\cdot \frac{1}{64} = 1$
$\Rightarrow K=64.$

(b) $f(x,y) \propto (e^{-x})(y^{-5}) = g(x)h(y)$ and $D=(0,\infty)\times(2,\infty)$ is rectangular $\Rightarrow$ X and Y are indep.
In fact, $f(x,y)=f_X(x)\cdot f_Y(y)$, where
$f_X(x)=e^{-x}I_{(0,\infty)}(x)$ and $f_Y(y)=\frac{4\cdot(2)^4}{y^{4+1}} I_{(2,\infty)}(y)$
$\therefore X\sim Exp(1)$ and $Y\sim Pareto(2,4)$

(c) $V(X)=\frac{1}{(1)^2}=1$, $V(Y)=\frac{(2)^2\cdot 4}{(4-1)^2\cdot(4-2)} = \frac{8}{9}$. (Can derive these from direct integration also)
Now, $V(3X-2Y)$
$=9V(X)+4V(Y)$ (By indep.)
$=9\cdot 1+4\cdot 8/9 = 113/9 = 12.556$

(d) $P(Y>3|X>1) = P(Y>3)$ (By indep.)
$= \int_3^\infty \frac{64}{y^5}dy = 64\left[\frac{-1}{4y^4}\right]_3^\infty = 16\cdot\frac{1}{3^4} = \frac{16}{81}$
$=0.198$

(e) $P(X>3|X>1)$
$=P(X>2)$ (By memorylessness)
$=e^{-1\cdot 2}=e^{-2}=0.135$

(f) Unconditional median: $\tilde{\mu}_X$: Soln. of $F(x)=0.5 \Rightarrow 1-e^{-x}=0.5$
$\Rightarrow e^{-x}=0.5$
$\Rightarrow x=\ln(2)$
$\therefore \tilde{\mu}_X = \ln(2)=0.693$
Conditional median given $Y>3$: $\tilde{\mu}_X=0.693$ (By indep.)

(g) $E(X)=1>\tilde{\mu}_X$, So dist^n of X is right skewed.

3. (10 points) Consider a simple random sample $X_1, X_2, \ldots, X_n$ of size n = 45 from the Poisson distribution with parameter $\lambda>0$.
The sample mean $(X_1+X_2+\ldots+X_{45})/45$ turned out to be 5.75.

(a) (1) Find out the method of moments estimate of λ based on this sample.
(b) (2) What are the bias and standard error of the above estimator (in terms of λ)?
(c) (1) Find out an estimate of the above standard error.
(d) (1) What is the approximate distribution of the above estimator of λ?
(e) (4) Perform an approximate test at level $\alpha=0.01$ for the null hypothesis H_0: $\lambda = 5$ against the two-sided alternative based on the sample. State the test statistic, the rejection region and your conclusion. Would your conclusion change if α was 0.05?
(f) (1) Without doing any calculation, what can you say about the approximate 99% confidence interval of λ based on the sample?

(a) MoM estimator: $E(X)=\bar{X} \Rightarrow \lambda=\bar{X} \Rightarrow \hat{\lambda}=\bar{X}$ ($E(X)=\lambda$ as $X\sim Poi(\lambda)$)
$\therefore$ MoM estimate: $\hat{\lambda}=5.75$.

(b) Bias of $\hat{\lambda}$: $E(\hat{\lambda})-\lambda = E(\bar{X})-\lambda = E(X)-\lambda = \lambda-\lambda=0$
SE of $\hat{\lambda}$: $SE(\hat{\lambda})=\sqrt{Var(\hat{\lambda})}=\sqrt{V(\bar{X})}=\sqrt{\sigma^2/n}=\frac{\sigma}{\sqrt{n}}=\frac{\sqrt{\lambda}}{\sqrt{n}}=\frac{\sqrt{\lambda}}{\sqrt{45}}=\frac{\sqrt{\lambda}}{6.708}$
($\sigma^2=Var(X)=\lambda$ Since $X\sim Poi(\lambda)$)

(c) $\hat{SE}(\hat{\lambda})=\frac{\sqrt{\hat{\lambda}}}{\sqrt{n}}=\frac{\sqrt{5.75}}{\sqrt{45}}=0.357$

(d) $\hat{\lambda}=\bar{X} \underset{approx.}{\sim} N\left(\mu,\frac{\sigma^2}{n}\right) \equiv N\left(\lambda,\frac{\lambda}{n}\right) = N\left(\lambda,\frac{\lambda}{45}\right)$ (by CLT as $n>30$)

(e) $H_0: \lambda=5$ vs. $H_a: \lambda\neq 5$
By part d, for an approx. test,
TS: $Z = \frac{\bar{X}-\lambda_0}{\hat{SE}(\bar{X})} = \frac{5.75-5}{0.357} = 2.101$.
RR: $\{|Z|>z_{1-\frac{\alpha}{2}}\} = \{|Z|>z_{0.995}\} = \{|Z|>2.576\}$ ($\alpha=.01$) (From Table)
Since $|Z|<2.576$, Don't reject H_0.
If $\alpha=0.05$, RR: $\{|Z|>z_{0.975}\}=\{|Z|>1.96\}$, since $|Z|>1.96$, we reject H_0 if $\alpha=0.05$.

(f) Since $H_0: \lambda=5$ is not rejected for $\alpha=0.01=1\%$, the 99% CI (approx.) of λ will contain the value 5.

4. (8 points) A simple random sample of 30 automobiles was obtained and the CO_2 emissions from each was measured (in g/mi). The sample mean of these measurements was 325.65 g/mi. Assume that CO_2 emissions follow a normal distribution with population standard deviation 65.

(a) (2) Calculate a 90% confidence interval for the mean CO_2 emissions of all automobiles.
(b) (2) How will the interval change (in terms of its width) if you make the confidence level 95%? How would it change if you sampled 40 automobiles?
(c) (4) The mean CO_2 emissions of all automobiles was reported last year as 335 g/mi by the U.S. Environmental Protection Agency. Is there any evidence in your data that the mean CO_2 emissions has decreased this year? State the hypotheses, the test statistic, the rejection region and a conclusion. Use α=0.05.

$\bar{x} = 325.65$, $n=30$, $\sigma=65$.

(a) 90% CI for μ: $\left(\bar{x} - z_{.95}\cdot\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{.95}\cdot\frac{\sigma}{\sqrt{n}}\right)$ $(\alpha=0.1)$

$= \left(325.65 - 1.645\cdot\frac{65}{\sqrt{30}}, 325.65 + 1.645\cdot\frac{65}{\sqrt{30}}\right)$

$= (306.128, 345.172)$ (From Table)

(b) Width $= 2\cdot z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$

If level is 95%, α goes down (0.1 to .05) $\Rightarrow z_{1-\frac{\alpha}{2}}$ goes up $\Rightarrow$ Width goes up.

If $n=40$, n goes up (from 30) $\Rightarrow$ Width goes down.

(c) $H_0: \mu = 335$ vs. $H_a: \mu < 335$

TS: $Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{325.65-335}{65/\sqrt{30}} = -0.788$

RR: $\{Z < -z_{1-\alpha}\} = \{Z < -z_{0.95}\} = \{Z < -1.645\}$ (From Table)

Since, $Z > -1.645 \Rightarrow$ Don't reject H_0.

Conclusion: There isn't enough evidence in the data to say that the mean CO_2 emissions has decreased this year.

5. (5 points) Assume the same setup as problem 4.
(a) (1) How large a sample size is necessary if the width of the 90% confidence interval in 4(a) is to be at most 30?
(b) (1) What is the probability of type-II error $\beta(330)$ for the test in 4(c) at the alternative value of 330 for the true mean CO_2 emissions?
(c) (3) Perform a test similar to the one in 4(c) when the population standard deviation is unknown and the sample of 30 automobiles had a standard deviation 65. State the hypotheses, the test statistic, the rejection region and a conclusion. Use α=0.05.

(a) $n_{min} = \left(\frac{2\sigma z_{1-\frac{\alpha}{2}}}{w_0}\right)^2 = \left(\frac{2\times 65\times z_{.95}}{30}\right)^2$ $(\alpha=0.1)$

$= 50.81 \nearrow 51.$

(b) Since H_a is $\{\mu < 335\}$,

$\beta(330) = 1 - \Phi\left(-z_{1-\alpha} + \frac{\mu_0-\mu_a}{\sigma/\sqrt{n}}\right)$

$= 1 - \Phi\left(-z_{.95} + \frac{335-330}{65/\sqrt{30}}\right)$ $(\alpha=0.05)$

$= 1 - \Phi(-1.22)$

$= 1 - 0.1112 = 0.8888$

(c) $H_0: \mu = 335$ vs. $H_a: \mu < 335$

σ is unknown, $S=65$.

TS: $T = \frac{\bar{x}-\mu_0}{S/\sqrt{n}} = -0.788$

RR: $\{T < -t_{\alpha;n-1}\} = \{T < -t_{0.05;29}\} = \{T < -1.699\}$ (From Table)

Since $T > -1.699 \Rightarrow$ Don't reject H_0.

Conclusion: There isn't enough evidence in the data to say that the mean CO_2 emissions has decreased this year.

6. (5 points) A simple random sample from a Binomial distribution with parameters n and p is: {5,6,7}. You don't know n and p and want to estimate both of them. Find out the method of moments estimates of n and p. It's okay if you have don't have an integer estimate for n.

Two parameters: n, p

MOM: $E(X) = \bar{X}$

$\Leftrightarrow np = \frac{5+6+7}{3} = 6 \rightarrow ①$ $[E(X) = np]$

and $E(X^2) = \overline{X^2}$

$\Leftrightarrow np(1-p) + n^2p^2 = \frac{5^2+6^2+7^2}{3} = \frac{110}{3} \rightarrow ②$ $\left[\begin{matrix} Var(X) \\ = np(1-p) \end{matrix}\right]$

$② - ①^2 : np(1-p) = \frac{110}{3} - 36 = \frac{2}{3} \rightarrow ③$

$③ \div ① : 1-p = \frac{2}{3 \times 6} = \frac{1}{9}$

$\Rightarrow p = 8/9$

Now, from ①, $n = 6/p = \frac{6}{8} \times 9 = \frac{27}{4}$

$\therefore$ MOM estimates: $\hat{n} = 6.75 \uparrow 7$

$\hat{p} = 8/9 = 0.889$

7. (5 points) Suppose that the number of pages in an issue of a journal is Uniformly distributed between 256 and 384. During 40 years of publication, the journal publishes a total of 160 issues. Give an estimate for the probability that between 51,000 and 52,000 pages (inclusive) were used for these 160 issues. [Hint: Write the number of pages in the ith issue as X_i and use the CLT.]

Let X_i = # of Pages in the i^{th} issue

$\therefore X_i$'s are iid $U(256, 384)$

Popl^n Mean $= \mu = \frac{256+384}{2} = 320$

Popl^n Var. $= \sigma^2 = \frac{(384-256)^2}{12} = 1365.333$

Total # Pages: $T_0 = X_1 + X_2 + \ldots + X_{160}$

Sample Mean: $\bar{X} = T_0/160$.

Now, with $n = 160$, we apply CLT,

$\bar{X} \underset{approx.}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \equiv N\left(320, \underbrace{\frac{1365.333}{160}}_{8.533}\right)$

$\therefore P(51000 \le T_0 \le 52000)$

$= P\left(\frac{51000}{160} \le \bar{X} \le \frac{52000}{160}\right)$

$= P\left(\frac{\frac{51000}{160} - 320}{\sqrt{8.533}} \le \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le \frac{\frac{52000}{160} - 320}{\sqrt{8.533}}\right)$

$\approx P(-0.428 \le Z \le 1.712)$

$= \Phi(1.71) - \Phi(-0.428)$

$= 0.9564 - 0.3336$ (From Table)

$= 0.6228$