STT351 - 005
Exam 2 **PRINT YOUR NAME HERE →**

Evaluate to the extent requested.
SHOW HOW YOU GET YOUR ANSWERS.
No notes, no extra paper, electronics, books, writings on one's person, are allowed.

Section on discrete models.

1. A joint pmf is given below. For example, $P(X = 2, Y = 4) = 0.1$.

x \ y	0	4
0	.2	.1
2	.3	.1
6	.1	.2

a. $p_X(0) = P(X = 0) =$ $.3 = .2 + .1$

b. $p_{Y|X}(4 \mid 2) = P(Y = \cancel{0}\ 4 \mid X = 2) =$ $\frac{.1}{.4} = 1/4$

c. $E(Y \mid X = 2) =$ $0 \cdot \frac{.3}{.4} + 4 \cdot \frac{.1}{.4} = 1$

d. $E(Y^2 \mid X = 2) =$ $0^2 \cdot \frac{.3}{.4} + 4^2 \cdot \frac{.1}{.4} = \frac{4^2}{4} = 4$

e. If (d) is 6 and (c) is 2 (they are not) then $\mathrm{Var}(Y \mid X = 2) =$ $E(Y^2 \mid X=2) - (E(Y \mid X=2))^2$
$= 6 - 2^2 = 2$

f. $P(X + Y < 3) =$ This is true for $y=4$, all x and $x=6$, $y=0$
So $.1 + .1 + .2 + .1 = .5$

2. A population has two sub-populations

	women	men	
number	300	200	(population sizes each sub-population)
mean age	35	30	(mean scores each sub-population)
sd age	10	8	(these are sd not variances)

Display the full numerical calculations but do not reduce answers.

a. Within component of variance of age = E Var(age | sex) =

$P(w) = \frac{300}{500} = \frac{3}{5}$

$P(m) = \frac{200}{500} = \frac{2}{5}$

$Var(age \mid w) = sd(age \mid w)^2 = 100$

$Var(age \mid m) = sd(age \mid m)^2 = 64$

128

$E(Var(age \mid sex)) = 100 \cdot \frac{3}{5} + 64 \cdot \frac{2}{5} = \frac{428}{5}$

b. Between component of variance of age = Var(E(age | sex)) =

$E(E(age \mid sex)^2) - E(E(age \mid sex))^2 = 35^2 \cdot \frac{3}{5} + 30^2 \cdot \frac{2}{5} - \left(35 \cdot \frac{3}{5} + 30 \cdot \frac{2}{5}\right)^2$

$7^2 \cdot 5 \cdot 3 + 6^2 \cdot 5 \cdot 2 - 7^2 \cdot 3^2 - 7 \cdot 3 \cdot 6 \cdot 2 - 6^2 \cdot 2^2$

$E(age \mid w) = 35 \quad E(age \mid m) = 30$

$7^2(5 \cdot 3 - 3^2) + 6^2(10 - 4) - 7 \cdot 3 \cdot 6 \cdot 2 = 49 \cdot 6 + 36 \cdot 6 - 12 \cdot 21$

c. Determine the variance of age over the entire population.

$= 294 + 216 - 252$

$= 510 - 252 = 258$

$Var(age) = E\,Var(age \mid sex) + Var\,E(age \mid sex) = \frac{428}{5} + 258 = \frac{1718}{5}$

$1290 + 428 = 1718$

d. If you allocate a sample of size n = 36 according to the method of proportional sampling, how many of the 36 will be selected randomly from the sub-population of women?

$\frac{3}{5} \cdot 36 = \frac{108}{5}$ is 21 or 22, neither perfect, 22 closer

e. If you employ the method mentioned in (d) (proportionally stratified sampling) the variance of the sample mean is not $\sigma^2/\sqrt{n}$. What is it? You claim to remove the between component of the variance; so $\sigma^2/\sqrt{n} - Var\,E(age \mid sex)$ if you sampled w/ equal probability (not proportional)

BONUS

Section on "continuous" models.

3. Define f(x, y) = (x - y), for 2 < x < 3, 1 < y < 2 (zero elsewhere).

a. Verify that f(x, y) is a density (two conditions must be satisfied).

$\iint_{-\infty} f(x,y)\,dx\,dy = 1$ on the interval $\Rightarrow \int_2^3 \int_1^2 (x-y)\,dy\,dx = \int_2^3 xy - \frac{y^2}{2}\Big|_1^2$

$= \int_2^3 x - \frac{3}{2}\,dx$

$0 \le x - y \le 2$ on the interval, so $f(x,y)$ is non negative

$= \frac{x^2}{2} - \frac{3}{2}x\Big|_2^3 = \frac{9}{2} - \frac{9}{2} - (2-3) = 1$ ✓

BONUS

b. Determine the marginal density $f_X(x)$ =

$\int f(x,y)\,dy = \int_1^2 x - y\,dy = xy - \frac{y^2}{2}\Big|_1^2 = x - \frac{3}{2}$

+1

$\int_1^2 2.5 - y = 2.5y - \frac{y^2}{2}\Big|_1^2 \quad 5-2-(2.5-1/2) = 1$

c. Determine the conditional density $f_{Y|X}(y \mid x = 2.5)$.

$f_{Y|X}(y|x=2.5) = 2.5 - y$

Bonus

d. Determine E(Y | X = 2.5).

$\int y f_{Y|X}(y|x=2.5)\,dy = \int_1^2 y(2.5-y)\,dy$

$= \frac{2.5y^2}{2} - \frac{y^3}{3}\Big|_1^2 = 5 - \frac{8}{3} - \left(1.25 - \frac{1}{3}\right)$

$\frac{60-32-15+4}{12} = \frac{17}{12}$

Section on Confidence Intervals.

z OK AS $n \to \infty$
$t \sim z$ AS $n \to \infty$

4. A sample of n = 400 items is selected at random, with equal probability, from a large batch of items. So we may as well assume sampling is with-replacement. The sample mean for score x = "smoothness of finish" is 78 and the sample sd of this score is s = 8.

NOTE: t ~~or z~~ findings are meaningless if pop. is not normal...

a. What is your estimate of the population mean finish score μ?

$78 = \bar{X}$

b. Determine a z-based (i.e. large n type) 90% CI for the population mean μ (use the appropriate entry from the t-table appended to this exam). df = n−1 = 399... much larger than the available 120, so use infinity df. t=1.645

so $\bar{X} \pm t \cdot \frac{s}{\sqrt{n}} = 78 \pm 1.645 \frac{8}{20}$ is 90% conf. interval

c. Give the estimated margin of error of the sample mean.

$\bar{X} \pm 1.96 \frac{s}{\sqrt{n}} = 78 \pm 1.96 \frac{8}{20}$

or just $1.96 \cdot \frac{8}{20}$

5. Refer to (4). Suppose the distribution of score x over the population is approximately normal (i.e. the score x is "in control"). This entitles us to employ a t-based 95% CI for μ even for a sample of n = 3.

a. Calculate (and reduce to a number) the sample sd of a sample {36, 40, 44}.

$\bar{X} = \frac{36+40+44}{3} = 40 \quad \frac{(36-40)^2 + (40-40)^2 + (44-40)^2}{n-1=2} = \frac{16+16}{2} = 16$ Var

$\sqrt{16} = 4 = s$, sample sd.

b. Using s from (a) determine a 95% t-based CI for population mean μ.

df=2 4.303 is the t score for df=2, $\frac{\alpha}{2} = .025$

so $40 \pm 4.303 \cdot \frac{4}{\sqrt{3}}$

note that (4) and (5) seem to not give a very good agreement on possible N

6. Bootstrap CI. An equal probability random sample X_1, ..., X_{30} is selected with-replacement from a population of thousands of business accounts. These 30 sample accounts are audited in order to estimate the mean (per account) dollar amount μ owed our company from these thousands of accounts.

a. Illustrate bootstrap by supposing four bootstrap sample means $\overline{X^*}$ (indiated by asterisks) and the sample mean $\overline{X}$ (indicated by +). In the plot below, indicate the bootstrap 75% CI (symmetric about $\overline{X}$)with left and right brackets [ ]. Four bootstrap replications is a toy number chosen for illustration only.

(* * + *) *

X

CUT ON 15TH PERCENTILE OF $|\bar{X}^* - \bar{X}|$

b. Around what percentage of 75% bootstrap CI cover the population mean μ? Assume each bootstrap CI is prepared using many thousands of replications, not the four of the toy illustration above. 75%; this the definition of sorts for CI.

or write

$$P\left(\mu \in \left\{\bar{x} \pm \underset{\text{bootstrap}}{75\% \widehat{CI}}\right\}\right) = .75 + O\left(\tfrac{1}{n}\right)$$

c. Repeat the illustration if instead we wish a 75% bootstrap CI for the population median and the following are the bootstrap sample medians (asterisks) and sample median (+) respectively.

exact same as (a)

* (* + *)

X

d. Express mathematically the claim made for (any worthy) 95% CI as n→∞.

as in (b)...

$$P\left(\theta \in \left\{\hat{\theta} \pm 95\% \text{ bootstrap CI}\right\}\right) = .95 + O\left(\tfrac{1}{n}\right)$$

for CI for μ — More generally see $O\left(\frac{1}{\sqrt{n}}\right)$

Section on unbiased estimation.

7. A sample of n is selected from a population having mean μ and sd σ.

a. If the sample is with equal probability and with-replacement then

$E\,\overline{X} = \mu$

b. If the sample is with equal probability and **without** rplacement then

$E\,\overline{X} = \mu$

c. If the sample is with equal probability and with-replacement then

$E\,s^2 = \sigma^2$

-2

8. Independent samples $X_1, \ldots, X_{30}$ are selected from the uniform density on the interval $[0, \theta]$ where $\theta > 0$ is not known. The density is

$f(x \mid \theta) = 1/\theta$, for $0 < x < \theta$. Uniform based on θ

a. Give an unbiased estimator of θ based on only the first sample X_1. Show that it is unbiased.

$2 \cdot X_1$ Not sure on showing ... but, if you consider

$E(2 \cdot X_1) = 2\int_0^\theta x \cdot \frac{1}{\theta}\,dx = \frac{2}{\theta}\left(\frac{x^2}{2}\right)\Big|_0^\theta = \frac{\theta^2}{\theta} - 0 = \theta$ ← what you want estimator to do

b. One of the estimators below is the conditional expectation of (a) relative to a sufficient statistic. It is also unbiased for θ and also has minimum possible variance, at every θ, of any unbiased estimator of θ. Which one is it?

$\text{Max}\{X_1, \ldots, X_{30}\}$

$(30/29)\,\text{Max}\{X_1, \ldots, X_{30}\}$ Why? Max is biased, $2\bar{X}$ has no sufficiency,

$2\bar{X}$ and $14X_1 - 12X_1$ has no sufficiency...

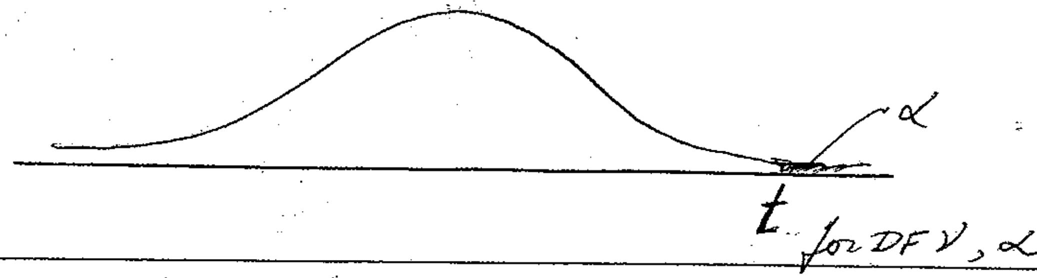
$14X_1 - 12X_1$ just $2X_1$, above

c. Determine $E(14X_1 - 12X_1)$. $= 14E(X_1) - 12E(X_1) = 2E(X_1)$

which was calculated in (a) to be θ

scaling Max in this way will be a closer guess than Max (unbiased) while having very low variance (decreases quickly as $n \to \infty$)

Bonus +1



ν	α = .10	.05	.025	.01	.005	.001	.0005
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
32	1.309	1.694	2.037	2.449	2.738	3.365	3.622
34	1.307	1.691	2.032	2.441	2.728	3.348	3.601
36	1.306	1.688	2.028	2.434	2.719	3.333	3.582
38	1.304	1.686	2.024	2.429	2.712	3.319	3.566
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	3.262	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291