Introduction to Mathematical Statistics I
Final

Name: Freya Zou **OSU-ID:** 934-417-686

Instructions

- The exam is worth **30 points** and consists of **6 problems**r.
- You **must upload** your final solutions to Canvas. The link to upload your solutions will **expire** at 9:30 PM.
- You **must show your work** and give concise, yet complete answers to receive full credit.
- You **must clearly** identify the problem you are solving in your solutions.
- You don't need to type your answers, but your solutions **must be clear and well organized**. If we cannot read it, we cannot grade it.
- The exam is **open-notes and open-book**, so you **can** use your class notes, homeworks and recitations when working on your problems.
- You are **allowed** to use a calculator, but **not software** to do your calculations. Any calculator will work, a graphic calculator is not required.
- You **must work independently**. You are not allowed to communicate with any person in any way or form to discuss exam problems.
- You are **not allowed** to use the Canvas discussion board (or any other platform) to seek and/or offer any clarifications about any questions or class materials during the exam.
- Answer the questions to the **best of your knowledge** or interpretation. No clarifications will be given during the exam.

Problem 1

Determine whether the following statements are True or False. Justify your answer.

(i) If X and Y are independent, then $E(\frac{X}{Y}) = \frac{E(X)}{E(Y)}$.

(ii) If $X \sim$ Normal$(1, 1)$, then the second moment of X is equal to 2.

(iii) If X and Y satisfy $E(XY) = E(X)E(Y)$, then X and Y are independent.

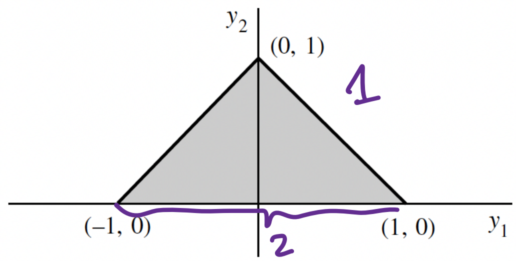
(i) False, If x and Y are Independent, then $E(XY) = E(X)E(Y)$
but expectation is Linear, but division is not. Thus $E(\frac{X}{Y}) \neq \frac{E(X)}{E(Y)}$
due to non-linearity of the expectation of a ratio.

(ii) True, $X \sim \text{Normal}(\mu=1, \sigma^2=1)$, the second moment is given
$E(X^2) = \mu^2 + \sigma^2$, $E(X^2) = 1^2 + 1^2 = 2$.

(iii) False. $E(XY) = E(X)E(Y)$ is necessary but not
sufficient. Independance also requires all events and conditional
probabilities also factorize accordingly, not just the expectation
of product.

Problem 2

Suppose that Y_1 and Y_2 have joint density function $f(y_1, y_2) = c$ (constant) over the region indicated in the figure:



(i) Find the value of the constant c.

(ii) Find $P(Y_1 - Y_2 \geq 0)$.

(iii) Find $E(Y_1 Y_2)$.

(i) Given: Length of triangle is 2, Hight is 1.
we know: $A = \frac{1}{2} * \text{base} * \text{Hight} = \frac{1}{2} \times 2 \times 1 = 1$. Thus, to Find c, we can use a probability density function: $\iint f(y_1, y_2)\, dy_1 dy_2 = 1$
$\iint c\, dy_1 dy_2 = 1c \times A = 1c \times 1 = 1 \Rightarrow c = 1$

(ii) we need to find $P(Y_1 - Y_2 \geq 0)$, the line $Y_1 = Y_2$ intersects the $(0,0)$, and we can the line $Y_1 = Y_2$ bisects the triangle into two smaller triangles of equal area. we already know the area of triangle equals 1, thus $P(Y_1 - Y_2 \geq 0) = \frac{1}{2}$

(iii) $E(Y_1, Y_2) = \iint y_1 y_2 f(y_1, y_2)\, dy_1 dy_2$, Since $f(y_1, y_2) = 1$
$\Rightarrow E(Y_1, Y_2) = \iint y_1 y_2\, dy_1 dy_2$. we can Integrate over this triangular region: (break into two parts)

The first triangle ($0 \leq y_1 \leq 1$, $0 \leq y_2 \leq y_1$): $\int_0^1 \int_0^{y_1} y_1 y_2\, dy_1 dy_2 = \frac{1}{8}$

The second triangle ($-1 \leq y_1 \leq 0$, $0 \leq y_2 \leq -y_1$) $\int_{-1}^{0} \int_0^{-y_1} y_1 y_2\, dy_1 dy_2 = -\frac{1}{8}$

$E(Y_1, Y_2) = \int_0^1 \int_0^{y_1} y_1 y_2\, dy_1 dy_2 + \int_{-1}^{0} \int_0^{-y_1} y_1 y_2\, dy_1 dy_2 = \frac{1}{8} + (-\frac{1}{8}) = 0$

thus $E(Y_1, Y_2) = 0$

Problem 3

The SAT and ACT college entrance exams are taken by thousands of students each year. The mathematics portions of each of these exams produce scores that are approximately normally distributed. In recent years, SAT mathematics exam scores have averaged 480 with a standard deviation of 100. The average and standard deviation for ACT mathematics scores are 18 and 6, respectively.

(i) An engineering school sets 550 as the minimum SAT score for new students. What percentage of students will score below 550 in a typical year?

(ii) What score should the engineering school set as a comparable standard of the ACT math test?

(iii) A fellowship program requires the applicants the be among the top 10% of scores in the SAT exam. What should be your minimum score to be eligible for the fellowship program?

(i) $Z = \frac{x-\mu}{\sigma} = \frac{550-480}{100} = 0.7$. use the standard Normal distribution table provided below:

$P(z < 0.7) = 1 - P(z > 0.7) = 1 - 0.2402 = 0.7580$

Thus, 75.8% of students will score below 550 in a typical year

(ii) From part (i) we know a SAT score of 550 corresponds to the 75.8%, we can use the value and z sore formula to solve for x: $0.7 = \frac{x-\mu}{\sigma} \Rightarrow x-\mu = 0.7\sigma \Rightarrow x = 0.7\sigma + \mu$

$= 0.7 * 6 + 18 = 22.2$

Thus, the engineering school set the ACT standard to 22 to match the percentile of the SAT score of 550.

(iii) To be top 10%, we look at $P(z > 2) = 0.1$ or $P(z < 2) = 0.9$

$z = 1.28$ (Form provided z score table), use the formula:

$z = \frac{x-\mu}{\sigma} \Rightarrow 1.28 = \frac{x-480}{100} \Rightarrow x = 1.28 \times 100 + 480 = 608$

Therefore, mininum SAT score of 608 would be eligible for the fellowship program.

Problem 4
Let X, Y be random variables with joint pdf

$$f(x,y) = \begin{cases} 6(1-y) & ,0 < x < y < 1 \\ 0 & ,\text{otherwise} \end{cases}$$

(i) Are X and Y independent? (Explain why or why not)

(ii) Find $P\left(\frac{Y}{X} > 3\right)$.

(iii) Find $E(X)$.

(i) X and Y are not independent, If X and Y are independent their joint probability (PDF) must factorize into the product of their marginal pdfs. The joint PDF $f(x,y) = 6(1-y)$ only defined of $0 < x < y < 1$. This dependence between X and Y makes it impossible to factorize the joint PDF into seperate functions of X and Y.

(ii) $\frac{Y}{X} > 3 \Rightarrow Y > 3X$: $P\left(\frac{Y}{X} > 3\right) = \int_{x=0}^{1/3}\int_{y=3x}^{1} 6(1-y)\,dy\,dx$

$$= \int_0^{1/3} 6\left(1-3x-\frac{1-9x^2}{2}\right)dx = \frac{1}{3}$$

(iii) $E(X) = \int_0^1 x f_X(x)\,dx$

Let find $f_X(x) = 6\int_{x}^{1}(1-y)\,dy = 6\left[y - \frac{y^2}{2}\right]_x^1 = 3 - 6x + 3x^2$

$$\Rightarrow E(X) = \int_0^1 x(3-6x+3x^2)\,dx = \int_0^1 (3x-6x^2+3x^3)\,dx$$

$$= 3\int_0^1 x\,dx - 6\int_0^1 x^2\,dx + 3\int_0^1 x^3\,dx = 3\left[\frac{x^2}{2}\right]_0^1 - 6\left[\frac{x^3}{3}\right]_0^1 + 3\left[\frac{x^4}{4}\right]_0^1$$

$$= 3\left(\frac{1^2}{2}\right) - 6\left(\frac{1^3}{3}\right) + 3\left(\frac{1^4}{4}\right) = \frac{6}{4} + \frac{3}{4} - 2 = \frac{1}{4} = 0.25$$

Problem 5

Let $N \sim \text{Binomial}(n, p)$, and let $X_1, X_2, X_3, \ldots$ be i.i.d Poisson(λ). Define

$$X = \sum_{i=1}^{N+1} X_i,$$

so that X is the sum of a random number of random variables.

i) Find E(X), the unconditional expectation of X.

ii) Find Var(X), the unconditional variance of X.

(i) Given: $E(X|N) = E\left(\sum_{i=1}^{N+1} X_i\right)$

Since the x_i is independent: $E(X|N) = (N+1)E(X_i) = (N+1)\lambda$

law of total expectation: $E(X) = E[E(X|N)]$

$\Rightarrow E(X) = E[(N+1)\lambda]$ subsitute $E(X|N)$.

$\Rightarrow E(X) = \lambda \cdot E(N+1) = \lambda(E(N)+1)$

Since Binomial (n, p). we know $E(N) = n \cdot p$. Thus:

$E(X) = \lambda(np+1)$

(ii) The law of total variance: $Var(X) = E[var(X|N)] + var[E(X|N)]$.

$E[var(X|N)] = E((N+1)\lambda) = \lambda E(N)+1 = \lambda(np+1)$

$var[E(X|N)] = var[(N+1)\lambda] = \lambda^2 \underline{var(N)}$

binomial $var(N) = np(1-p)$

$= \lambda^2 [np(1-p)]$

$= \lambda^2 np(1-p)$

$\therefore Var(X) = \lambda(np+1) + \lambda^2 np(1-p)$

Problem 6

Suppose that $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed Bernoulli(p) random variables. That is, $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$ for $i = 1, \ldots, n$.

(i) Find the moment generating function $m_{Y_1}(t)$ of Y_1 (the first observation).

(ii) Find the moment generating function of $W = Y_1 + \ldots + Y_n$

(iii) What is the distribution of W? Justify your answer.

(i) $M_{Y_1}(t) = E(e^{tY_1}) = \sum_{k=0}^{1} e^{tk} P(Y_1 = k)$, since Y_1 is Bernoulli:

$$M_{Y_1}(t) = e^{t\cdot 0} P(Y_1 = 0) + e^{t\cdot 1} P(Y_1 = 1)$$

$$= (1-p) + p\cdot e^t = pe^t + 1 - p$$

(ii) From Part(i) $M_{Y_1}(t) = pe^t + 1 - p$

$$M_W(t) = (M_{Y_1}(t))^n = (pe^t + 1 - p)^n$$

(iii) The distribution of $W \sim$ Binomial $(n \cdot p)$

W counts the ~~total~~ number of successes in n independent Bernoulli trials, which is exactly the definition of Binomial $(n \cdot p)$ random variable.

The MGF of Binomial $(n \cdot p)$: $m_W(t) = (pe^t + 1 - p)^n$ which matches the MGF derived in part(ii)

Problem 7

Suppose that Y_1 and Y_2 are independent random variables with distribution Uniform$(0, 1)$. Define $X = \min\{Y_1, Y_2\}$.

(i) Derive the pdf of X. (You need to show all the steps to obtain the pdf and not just writing the final formula)

(ii) Find $E(X)$.

(iii) Find $P(X > 0.5)$.

Given $Y \sim$ uniform$(0,1)$

(i) $F_X(x) = P(\min\{Y_1, Y_2\} \leq x) = 1 - P(\min\{Y_1, Y_2\} > x)$

$\Rightarrow P(\min\{Y_1, Y_2\} > x) = P(Y_1 > x)\, P(Y_2 > x)$

$= (1-x)(1-x) = (1-x)^2$

thus, $F_X(x) = 1 - (1-x^2) = 2x - x^2 \; (0 \leq x \leq 1)$

Derive the pdf of X:

$f_X(x) = \frac{d}{dx}(2x - x^2) = 2 - 2x \; (0 \leq x \leq 1)$

(ii) $E(X) = \int_0^1 x f_X(x)\, dx = \int_0^1 x(2-2x)\, dx = \int (2x - 2x^2) dx$

$= \int_0^1 2x\, dx - \int_0^1 2x^2 dx = \left[\frac{2x^2}{2} - \frac{2x^3}{3}\right]_0^1 = \frac{1}{3} \approx 0.3333$

(iii) $P(X > 0.5) = 1 - \underline{F_X}(0.5) \rightarrow$ part(i) $F_X(x) = 2x - x^2$

$= 1 - (2(0.5) - (0.5)^2)$

$= 1 - (1 - 0.25)$

$= 0.25$

Problem 8
The lifetime (in hours) Y of an electronic component is a random variable with pdf

$$f(y) = \begin{cases} \frac{1}{200}e^{-y/200} & , 0 < y < \infty \\ 0 & , \text{otherwise} \end{cases}$$

Suppose that four of these components operate independently in a certain equipment, and that the equipment fails if at least two of these components fail.

Find the probability that the equipment will operate for at least 400 hours without failure.

Given: $f(y) = \frac{1}{200} e^{-y/200}$, $y > 0$

$\lambda = \frac{1}{200}$, $E(Y) = \frac{1}{\lambda} = \frac{1}{1/200} = 200$

$P(Y > t) = e^{-\lambda t}$. subsitue λ and $t = 400$:

$P(Y > 400) = e^{-\frac{400}{200}} = e^{-2} \approx 0.1353$

$P(\text{failure}) = 1 - P(Y > 400) = 1 - 0.1353 = 0.8647$

thus $Z \sim \text{Binomial}(4, 0.8647)$

use Binomial Probability mass function:

$P(Z=k) = \binom{4}{k} p^k (1-p)^{4-k}$, $k = 0, 1$

$P(Z \leq 1) = P(Z=0) + P(Z=1)$

$= \binom{4}{0}(0.8647)^0(0.1353)^4 + \binom{4}{1}(0.8647)^1(0.1353)^3$

$= (0.1353)^4 + 4(0.8647)(0.1353)^3$

$= 0.000335 + 0.00856$

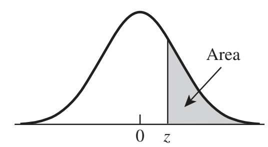
≈ 0.0089

thus: The equipment will operate for at least 400 hours with failures is 0.89%

Distribution	Probability Function	Mean	Variance	Moment-Generating Function
Binomial	$p(y) = \binom{n}{y} p^y (1-p)^{n-y};$ $y = 0, 1, \ldots, n$	np	$np(1-p)$	$[pe^t + (1-p)]^n$
Geometric	$p(y) = p(1-p)^{y-1};$ $y = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Hypergeometric	$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}};$ $y = 0, 1, \ldots, n$ if $n \le r$, $y = 0, 1, \ldots, r$ if $n > r$	$\frac{nr}{N}$	$n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$	does not exist in closed form
Poisson	$p(y) = \frac{\lambda^y e^{-\lambda}}{y!};$ $y = 0, 1, 2, \ldots$	λ	λ	$\exp[\lambda(e^t - 1)]$
Negative binomial	$p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r};$ $y = r, r+1, \ldots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$

Distribution	Probability Function	Mean	Variance	Moment-Generating Function
Uniform	$f(y) = \frac{1}{\theta_2 - \theta_1}; \theta_1 \le y \le \theta_2$	$\frac{\theta_1 + \theta_2}{2}$	$\frac{(\theta_2 - \theta_1)^2}{12}$	$\frac{e^{t\theta_2} - e^{t\theta_1}}{t(\theta_2 - \theta_1)}$
Normal	$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\left(\frac{1}{2\sigma^2}\right)(y-\mu)^2\right]$ $-\infty < y < +\infty$	μ	σ^2	$\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$
Exponential	$f(y) = \frac{1}{\beta} e^{-y/\beta}; \quad \beta > 0$ $0 < y < \infty$	β	β^2	$(1 - \beta t)^{-1}$
Gamma	$f(y) = \left[\frac{1}{\Gamma(\alpha)\beta^\alpha}\right] y^{\alpha-1} e^{-y/\beta};$ $0 < y < \infty$	$\alpha\beta$	$\alpha\beta^2$	$(1 - \beta t)^{-\alpha}$
Chi-square	$f(y) = \frac{(y)^{(v/2)-1} e^{-y/2}}{2^{v/2}\Gamma(v/2)};$ $y > 0$	v	$2v$	$(1 - 2t)^{-v/2}$
Beta	$f(y) = \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right] y^{\alpha-1}(1-y)^{\beta-1};$ $0 < y < 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	does not exist in closed form

Table 4 Normal Curve Areas
Standard normal probability in right-hand tail
(for negative values of z, areas are found by symmetry)



	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000 233									
4.0	.000 031 7									
4.5	.000 003 40									
5.0	.000 000 287									

From R. E. Walpole, *Introduction to Statistics* (New York: Macmillan, 1968).