MH1820 Week 13 (Review)

QUESTION 1. **(30 Marks)**

(a) Let X be a continuous random variable with PDF given by

$$f(x) = \begin{cases} C(1-x^2), & \text{for } -1 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

(i) What is the value of C?

(ii) Compute $\mathbb{E}[X]$ and $\mathrm{Var}[X]$.

(iii) Find the PDF of $Y = e^X$.

(b) If X has a normal distribution with mean $\mu = 3$ and variance $\sigma^2 = 9$, find $\mathbb{P}(|X-3| > 6)$ in terms of $\Phi(z)$, the CDF of the standard normal random variable Z.

(a) Recall: $$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\int_{-1}^{1} c(1-x^2)\,dx = 1$$

$$c\left[x - \frac{x^3}{3}\right]_{-1}^{1} = 1$$

$$c\left[\left(1-\frac{1}{3}\right) - \left(-1-\frac{(-1)}{3}\right)\right] = 1$$

$$\frac{4}{3}c = 1 \Rightarrow c = \frac{3}{4} \quad \#$$

(iii) $\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$ (definition)

$$= \int_{-1}^{1} x \, \frac{3}{4}(1-x^2)\, dx$$

$$\vdots$$

$$= 0$$

$$Var[X] = \mathbb{E}[X^2] - \cancel{\mathbb{E}[X]}^2 \quad 0$$

$$= \mathbb{E}[X^2]$$

$$= \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

$$= \int_{-1}^{1} x^2 \frac{3}{4}(1-x^2)\, dx$$

$$= \frac{3}{4}\left[\frac{x^3}{3} - \frac{x^5}{5}\right]_{-1}^{1} = 0.2$$

$$= \frac{3}{4}\left[\left(\frac{1}{3}-\frac{1}{5}\right)-\left(\frac{-1}{3}+\frac{1}{5}\right)\right] = \frac{3}{4}\cdot 2\left[\frac{1}{3}-\frac{1}{5}\right]$$

$$\frac{3}{4}\cdot 2\left[\frac{2}{15}\right] = \frac{3}{15}$$

(ii) $Y = e^X$

CDF: $\mathbb{P}(Y \leq y)$

$= \mathbb{P}(e^X \leq y)$

$= \mathbb{P}(X \leq \ln y)$ ✓

$$P(X \le x) = \int_{-\infty}^{x} f(t)\, dt \qquad x$$

$$= \int_{-1}^{x} \frac{3}{4}(1 - t^2)\, dt$$

$$= \begin{cases} \frac{3}{4}x - \frac{x^3}{4} + \frac{1}{2} & \text{if } -1 \le x \le 1 \\ 0 & \text{if } x < -1 \\ 1 & \text{if } x > 1 \end{cases}$$

Hence

CDF of Y, $F_Y(y) = \mathbb{P}(Y \leq y)$

$$= \mathbb{P}(X \leq \ln y)$$

$$= \begin{cases} \frac{3}{4}(\ln y) - \frac{(\ln y)^3}{4} + \frac{1}{2}, & -1 \leq \ln y \leq 1 \\ 0, & \ln y < -1 \\ 1 & \ln y > 1 \end{cases}$$

PDF $\frac{dF_Y}{dy}(y) = \begin{cases} \frac{3}{4}\frac{1}{y} - \frac{1}{4}3(\ln y)^2\frac{1}{y}, & -1 \leq \ln y \leq 1 \\ 0 \end{cases}$

(b) $X \sim N(\mu = 3, \sigma^2 = 9)$

$$\mathbb{P}(|X-3| > 6)$$

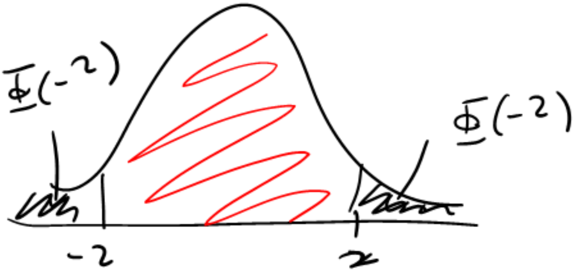
$$= \mathbb{P}(X-3 > 6 \text{ or } X-3 < -6)$$

$$= \mathbb{P}(X > 9 \text{ or } X < -3)$$

$$= \mathbb{P}\left(\frac{X-3}{3} > \frac{9-3}{3} \text{ or } \frac{X-3}{3} < \frac{-3-3}{3}\right)$$

$$= \mathbb{P}(\phi > 2 \text{ or } \phi < -2)$$

$$= 2\,\Phi(-2). \checkmark$$



$\Phi(-2)$
$\Phi(-2)$
-2
2

(c) Suppose X has the uniform distribution $U(1,3)$ on the interval $[1,3]$. Using the definition of moment generating function (MGF), find the MGF $M_X(t)$ of X.

(d) Each game you play is a win with probability 0.6. You plan to play 5 games, but if you win the fifth game, then you will keep on playing until you lose. Find the expected number of games that you will play.

Game $\boxed{1}$ $\boxed{2}$ $\boxed{3}$ $\boxed{4}$ W W .... W L

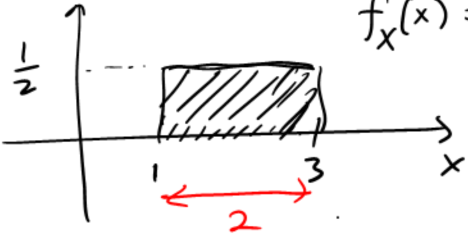
L

W L

W W L

(c) $$M_X(t) = \mathbb{E}\left[e^{tX}\right]$$

$$X \sim U(1,3)$$



$$f_X(x) = \begin{cases} \frac{1}{2}, & 1 \le x \le 3 \\ 0 & \text{otherwise.} \end{cases}$$

$$M_X(t) = \mathbb{E}[e^{tX}] \qquad t \neq 0$$

$$= \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$$

$$= \int_1^3 e^{tx} \frac{1}{2}\, dx$$

$$= \frac{1}{2}\left[\frac{e^{tx}}{t}\right]_1^3 = \frac{1}{2t}\left[e^{3t} - e^{t}\right]$$

#

RECALL...

A random variable X has a **Geometric distribution**, denoted by $X \sim Geom(p)$, if X counts the number of experiments until the first success in a sequence of independent experiments with success probability p.

Theorem 14 (Geometric distribution)

If $X \sim Geom(p)$, then

$$\text{PMF: } p(x) = (1-p)^{x-1}p, \quad x = 1, 2, \ldots,$$

$$\mathbb{E}[X] = \frac{1}{p}, \quad \text{Var}[X] = \frac{1-p}{p^2}.$$

Let X = # games played starting from the 5-th game.

$$X \sim Geom(p = 0.4)$$

$$\mathbb{E}[X] = \frac{1}{p} = \frac{1}{0.4}$$

Hence expected # games $= 4 + \mathbb{E}[X]$

$$= 4 + \frac{1}{0.4}$$

#

QUESTION 2. **(20 Marks)**

(a) The weight X (in grams) of a randomly selected chocolate bar produced by a company is normally distributed with mean μ and variance σ^2 which is unknown. Due to a potential fault in a machine, the company suspects that the mean weight is less than 300 grams. We shall test the null hypothesis H_0: $\mu = 300$ against the alternative hypothesis H_1: $\mu < 300$, with a significance level of $\alpha = 0.05$. A random sample of $n = 30$ yielded a mean of $\overline{x} = 280$ and standard deviation $s = 60$.

(i) What is the p-value of the test?

(ii) What is the conclusion of the test?

RECALL...

Procedue for Hypothesis Testing:

- Given are observations $x_1, \ldots, x_n$.
- Formulate **null hypothesis** H_0 describing the population distribution from which observations were drawn.
- Choose significance level α (often $\alpha = 0.05$)
- Choose test statistic $T(X_1, \ldots, X_n)$ that contains information on the parameters involved in H_0 and whose distribution is known under H_0.
- Assuming H_0, compute probability (p-value) to observe $t = T(x_1, \ldots, x_n)$ or something "**at least as extreme as t" (in the direction of rejection of H_0)**.
- If the p-value is smaller than α, reject null hypothesis

$$\begin{cases} T \geq t \\ T \leq t \\ |T - \mathbb{E}[T]| \geq |t - \mathbb{E}[T]| \end{cases}$$

$$H_0: \mu = 300$$
$$H_1: \mu < 300.$$

Use: $$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim N(0,1)$$

$$p\text{-value} = \mathbb{P}\left(T < t = \frac{280-300}{60/\sqrt{30}}\right)$$
$$= \mathbb{P}(T < -1.826)$$
$$= 1 - \mathbb{P}(T < 1.83)$$
$$= 1 - 0.9664 \simeq 0.0336 \ \#$$

(b) Reject H_0 since p-value < 0.05

AY2223 Exam Q2

(b) Let X_1, X_2, ..., X_{12} be a random sample of size $n = 12$ from the normal distribution $N(\mu, \sigma^2)$. We shall test the null hypothesis H_0: $\sigma^2 = 10$ against the alternative hypothesis H_1: $\sigma^2 = 35$.

(i) Find a rejection criteria for the test, where the size of the test is $\alpha = 0.05$.

(ii) Estimate the probability of a Type II Error with the rejection criteria in (i).

RECALL....

Suppose $X_1, \ldots, X_n$ i.i.d $\sim N(\mu, \sigma^2)$. Notice that

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma}\right)^2$$

- Recall that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ (chi-square distribution with degree of freedom $n-1$) (Week 9).
- We can use this to construct confidence intervals.

b(i) Use sample variance S^2

· Reject if $S^2 \geq C$ ✓ for some C

Test statistic $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

Want the size to be $\alpha = 0.05$.

RECALL...

Size of a Test

- If the null hypothesis H_0 is true, but **rejected**, then a **Type I Error** occurs.
- The probability of a Type I Error is

$$\alpha = \mathbb{P}(H_0 \text{ rejected} | H_0).$$

- $\mathbb{P}(H_0 \text{ rejected} | H_0)$ is also called the **size** of the test.
- The smaller the size, the more conclusive is the test – the size measures how conclusive a test is.

$$\mathbb{P}(S^2 \geq c \mid H_0) = 0.05 .$$

$$\mathbb{P}\left(\frac{(n-1)S^2}{\sigma^2} \geq \frac{(n-1)c}{\sigma^2} \,\middle|\, \sigma^2 = 10\right) = 0.05$$

$$\mathbb{P}\left(X \geq \boxed{\frac{11}{10}c}\right) = 0.05 .$$

$$\chi^2_{0.05}(11) = 19.68$$

$$\frac{11}{10}c = 19.68$$

$$c = 17.89 \quad \#$$

RECALL...

Power of a Test

- The probability of a Type II Error is denoted by β:

$$\beta = \mathbb{P}(H_0 \text{ not rejected} | H_1)$$

- The probability that H_0 is rejected if it is wrong is the **power** of the test, i.e.

$$\textbf{Power} = \mathbb{P}(H_0 \text{ rejected} | H_1) = 1 - \beta.$$

$$\text{power} = 1 - \beta$$

$$= 1 - \mathbb{P}(S^2 < c \mid H_1)$$

$$= 1 - \mathbb{P}\left(\frac{(n-1)S^2}{\sigma^2} < \frac{(n-1)c}{\sigma^2} \;\middle|\; \underline{\underline{\sigma^2 = 35}}\right)$$

$$= 1 - \mathbb{P}\left(\chi < \frac{11 \cdot 17.89}{35}\right)$$

$$= 1 - \mathbb{P}(\chi < 5.622)$$

$$= 1 - 0.1 = 0.9$$

$\beta = 0.1$

#

AY2223 Exam Q3

QUESTION 3. **(25 Marks)**

(a) The joint PDF of two random variables X and Y is given by

$$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1; \\ 0, & \text{elsewhere.} \end{cases}$$

(i) Find the marginal PDFs of X and Y.

(ii) Are X and Y independent? Justify your answer.

(iii) Compute $\mathbb{P}\left(\frac{1}{4} < X < \frac{1}{2} \mid \underline{Y = \frac{1}{3}}\right)$.

(iv) Compute $\mathbb{P}(X > Y)$.

(a)(i) $$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$$

$$= \int_0^1 \frac{x}{4}(1+3y^2)\, dy$$

$$= \frac{x}{4}\int_0^1 (1+3y^2)\, dy$$

$$= \frac{x}{4}\left[y+y^3\right]_0^1$$

$$= \frac{x}{4}\cdot 2 = \frac{x}{2} \qquad \boxed{0 < x < 2}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$$

$$= \int_0^2 \frac{x}{4}(1+3y^2)\, dx$$

$$= \left(\frac{1+3y^2}{4}\right) \int_0^2 x \, dx$$

$$= \left(\frac{1+3y^2}{4}\right) \left[\frac{x^2}{2}\right]_0^2 = \frac{1+3y^2}{2}, \quad 0 < y < 1$$

(ii) X & Y independent if $f_X(x) f_X(y)$

$$= f(x,y)$$

$$f_X(x) f_Y(y) = \frac{x}{2} \cdot \frac{1+3y^2}{2}$$

$$= \frac{x(1+3y^2)}{4}$$

$$= f(x,y)$$

<u>Yes</u>.

RECALL...

- The **conditional PMF/PDF** of X, given that $Y = y$, is defined by

$$g(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

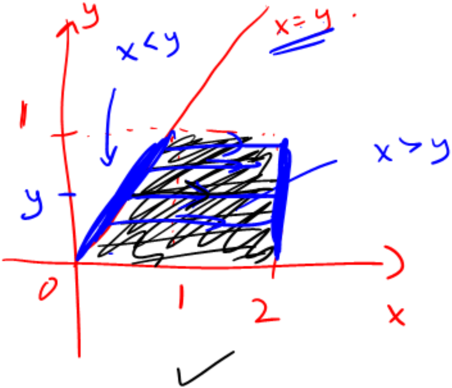
- The **conditional PMF/PDF** of Y, given that $X = x$, is defined by

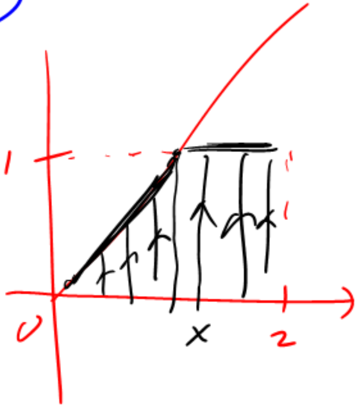
$$h(y|x) = \frac{f(x,y)}{f_X(x)}.$$

(a)(iii) $\mathbb{P}\left(\frac{1}{4} < X < \frac{1}{2} \,\middle|\, Y = \frac{1}{3}\right) = \int_{1/4}^{1/2} \frac{x}{2}\, dx = \ldots = \frac{3}{64}$

(a)(iv)

$$\mathbb{P}(X > Y) = \int_0^1 \int_y^2 f(x,y)\, dx\, dy$$





$$P(X > Y) = \int_0^1 \int_y^2 \frac{x}{4}(1+3y^2)\,dx\,dy$$

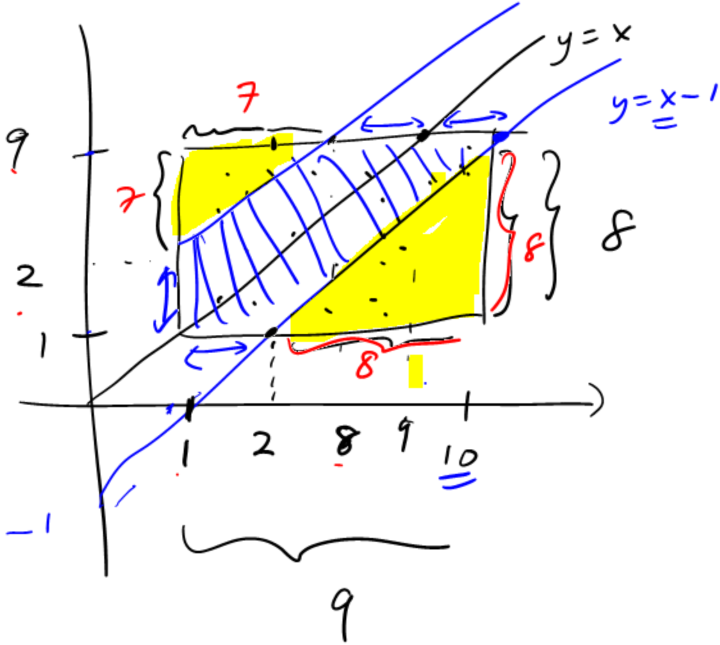
$$= \int_0^1 \left[\frac{x^2}{8}(1+3y^2)\right]_y^2 dy$$

$$= \int_0^1 \frac{(4-y^2)}{8}(1+3y^2)\,dy.$$

$$= \cdots\cdots = 0.8833 \quad \#$$

AY2223 Exam Q3

(b) In a number game, two participants make guesses of X and Y respectively. The joint PDF of X and Y is **uniform** (i.e. constant) on the region $1 \leq x \leq 10$, $1 \leq y \leq 9$. If $|X - Y| < 1$, then the two participants will be asked to guess again. What is the probability that they will be asked to guess again?



y = x
y = x − 1
7
7
8
8
8
9
2
1
−1
1
2
8
9
10
9

$\mathbb{P}(|X-Y| < 1)$ = prob. of guessing again.

$$|X-Y| < 1 \iff -1 < X-Y < 1$$

$$\underline{\underline{X-1 < Y < X+1}}$$

$$f(x,y) = \begin{cases} \frac{1}{72} & 1 < x < 10 \\ & 1 < y < 9 \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbb{P}(|X-Y|<1) = 1 - \mathbb{P}(|X-Y| \geq 1)$$
$$= 1 - (\text{area shaded in yellow}) \times \frac{1}{72}$$
$$= 1 - \left(\frac{49}{2} + \frac{64}{2}\right) \times \frac{1}{72} = \frac{31}{144} \ \#$$

QUESTION 4. **(25 Marks)**

(a) Let $X_1, \ldots, X_n$ be i.i.d from $Poisson(\lambda)$, where λ is unknown. Find the maximum likelihood estimator for λ based on the observations $x_1 = 13$, $x_2 = 5$, $x_3 = 6$, $x_4 = 7$ (here $n = 4$). (Recall that if $X \sim Poisson(\lambda)$, then $\mathbb{P}(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}$.)

RECALL...

Set-up for maximum likelihood:

- Let $X_1, \ldots, X_n$ be i.d. with PMF or PDF $f(x|\theta)$, depending on an unknown parameter θ.
- Observations $x_1, \ldots, x_n$ are given.
- The idea of the maximum likelihood method is to choose the value for θ as estimator which maximizes the following **maximum likelihood function**:

$$L(X_1 = x_1, \ldots, X_n = x_n|\theta) \quad = \quad \prod_{i=1}^{n} f(x_i|\theta)$$

RECALL...

Finding Maximum Likelihood Estimator (MLE):

The maximum likelihood estimator i.e. the value of θ that maximizes $L(x_1, \ldots, x_n|\theta)$ can be found by solving

- $\dfrac{d}{dp}(L) = 0$ or
- $\dfrac{d}{dp}(\ln L) = 0$

Both solution methods are valid, but sometimes the second method often is faster. There are likelihood functions for which the maximizer cannot be found in this way, but such cases will not occur in this course.

AY2223 Exam Q4

(b) Let D_θ, $0 < \theta < 1$, be the discrete distribution with the following PMF:

x	1	2	3
$p(x)$	$\theta/3$	$2\theta/3$	$1-\theta$

Let $X_1, \ldots, X_n$ be i.i.d drawn from D_θ and let $\overline{X}$ be the sample mean. Consider an estimator for θ given by $\widehat{\theta} = \frac{1}{3}\overline{X}$.

(i) Compute the bias and standard error for $\widehat{\theta}$.

(ii) Find $\widehat{\theta}$ using the observations $x_1 = 2$, $x_2 = 2$, $x_3 = 1$, $x_4 = 3$ (here $n = 4$).

(iii) Find an estimator of θ which is unbiased, i.e. it has zero bias.

RECALL...

Let $\widehat{\theta}$ be an estimator of θ. The **bias** of $\widehat{\theta}$ is defined by

$$\mathrm{Bias}(\widehat{\theta}) = \mathbb{E}[\widehat{\theta}] - \theta.$$

Here, the expectation is computed under the population distribution parametrized by θ.

Standard error of $\widehat{\theta}$:

$$SE(\widehat{\theta}) = \sqrt{\mathrm{Var}[\widehat{\theta}]}.$$

Here the variance is computed under the population distribution parametrized by θ.

$X_1, X_2, \ldots, X_n$ i.i.d.

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}.$$

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{X_1 + \cdots + X_n}{n}\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \mathbb{E}[X_i]$$

$$\mathrm{Var}[\bar{X}] = \mathrm{Var}\left[\frac{\sum_{i=1}^{n} X_i}{n}\right]$$

$$= \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \underline{\underline{\mathrm{Var}[X_i]}}$$

$$= \frac{1}{n^2}\, n \,\mathrm{Var}[X_i] = \frac{1}{n}\mathrm{Var}(X_i).$$