ENGG 2780A / ESTR 2020: Statistics for Engineers
The Chinese University of Hong Kong, Spring 2021 — Midterm Exam

---

You must work alone and cite any external references that you use as stipulated in the CUHK academic honesty guidelines. Please submit your solution by 2.15pm on March 8 electronically at `https://course.cse.cuhk.edu.hk/~engg2780a/`.

Each question is worth 10 points. Explain your answers clearly.

1. You are trying to estimate the fraction V of vegetarians in Hong Kong using Bayesian statistics. Your prior is that V is a Uniform$(0, 1/2)$ random variable.

   (a) You poll a random person and they are not a vegetarian. What is the posterior PDF of V?

   Solution: By Bayes's rule, $f_{V|X}(v|0) \propto \mathrm{P}(X = 0|V = v) f_V(v) \propto 1 - v$ for $v \in [0, 1/2]$. As $\int_0^{1/2}(1-v)dv = \frac{3}{8}$, the posterior PDF is $\frac{8}{3}(1-v)$ for $v \in [0, 1/2]$.

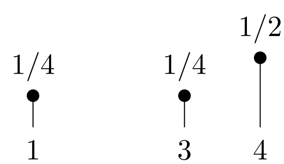
   (b) What is the expected posterior probability that the next polled person will be a vegetarian?

   Solution: It is $\mathrm{E}[V|X=0] = \int_0^{1/2} v \cdot \frac{8}{3}(1-v)dv = \frac{8}{3}(\frac{1}{8} - \frac{1}{24}) = \frac{2}{9} \approx 0.222$.

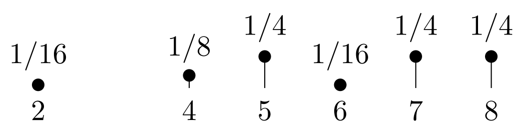
2. A food company produced 1000 boxes of biscuits, 500 of which contain 4 biscuits, 250 contain 3 biscuits and 250 contain one biscuit. You sample two boxes (with repetition) and record the sample mean $\overline{X}$ of the number of biscuits.

   (a) What is the PMF of $\overline{X}$?

   Solution: The marginal PMF of each sample is $f(4) = \frac{1}{2}$, $f(3) = \frac{1}{4}$, $f(1) = \frac{1}{4}$:



   We can calculate the PMF of $X_1 + X_2$ using the convolution formula to get:



   The PMF of $\overline{X} = (X_1 + X_2)/2$ is then obtained by scaling the value by $\frac{1}{2}$:

x	1	2	$2\frac{1}{2}$	3	$3\frac{1}{2}$	4
$\mathrm{P}(\overline{X} = x)$	1/16	1/8	1/4	1/16	1/4	1/4

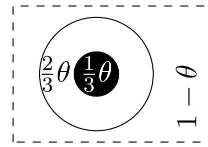
   (b) What is the probability that the sample mean equals the actual mean?

   Solution: The actual mean is $\mu = 1 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{2} = 3$, so $\mathrm{P}(\overline{X} = \mu) = 1/16$.

3. Four independent samples of a normal random variable of unknown mean and variance take values 83, 103, 93, 93. Calculate a 95% symmetric confidence interval for the sample mean. Explain which formula(s) you used in your calculation.

Solution: The sample mean $\overline{X}$ is 93 and the adjusted sample variance S^2 is $((83-93)^2 + (103-93)^2 + 2 \cdot (93-93)^2)/3 = 200/3$. As $(\overline{X}-\mu)/(S/2)$ is a $t(3)$ random variable, the desired confidence interval is $(\overline{X} - zS/2, \overline{X} - zS/2)$ where $z \approx 3.182$ is chosen so that $\mathrm{P}(-z \leq t(3) \leq z) = 95\%$. Plugging in the values for $\overline{X}$, S, z, and n we get the interval $(80, 106)$.

4. An archer hits the bull's eye with probability $\frac{1}{3}\theta$, the rest of the target with probability $\frac{2}{3}\theta$, and misses the target with probability $1-\theta$, where $\theta \in [0,1]$ is a parameter that models the archer's skill.



(a) The archer hits the bull's eye twice and misses the board once. What is the maximum likelihood estimate of their skill θ (assuming their shots are independent)?

Solution: The probability of this outcome is proportional to $(\frac{1}{3}\theta)^2 \cdot (1-\theta) \propto \theta^2 - \theta^3$. The critical points are those θ for which $(d/d\theta)(\theta^2-\theta^3) = 2\theta - 3\theta^2$ equals zero, namely 0 and $2/3$, together with the other endpoint $\theta = 1$. The maximum occurs at $\theta = 2/3$. This is the maximum likelihood estimate.

(b) Describe an unbiased estimator (3 points) of minimum variance (+2 points) for the player's skill θ from a single attempt. (**Hint:** The estimator assigns a "score" to each outcome.)

Solution: One unbiased estimator assigns a score S of 1 for hitting the board (bull's eye or not) and 0 for missing it. Then $\mathrm{E}[S] = 1 \cdot \frac{2}{3}\theta + 1 \cdot \frac{1}{3}\theta + 0 \cdot (1-\theta) = \theta$.

A general estimator S will assign scores a, b, c for the bull's eye, the rest of the board, and a miss. For the estimator to be unbiased we need that $\frac{1}{3}\theta a + \frac{2}{3}\theta b + (1-\theta)c = \theta$ for all possible skill levels θ, from where we must choose $c = 0$ and $a + 2b = 3$. The variance of the estimator is $\mathrm{Var}[S] = \mathrm{E}[S^2] - \mathrm{E}[S]^2 = \frac{1}{3}\theta a^2 + \frac{2}{3}\theta b^2 - \theta^2$, so we need to minimize $\frac{1}{3}a^2 + \frac{2}{3}b^2 = \frac{1}{3}(3-2b)^2 + \frac{2}{3}b^2$. This is an increasing quadratic function of b so it is minimized at the critical point where its first derivative $\frac{4}{3}(2b-3) + \frac{4}{3}b$ vanishes, namely $b = 1$ and $a = 1$. So the proposed unbiased estimator happens to be the one of minimum variance.