Final examination for course of Fundamentals of Statistical Machine Learning (20190110).
This exam is closed book, closed notes.

Your Name: ______________________________________________

1. Let $f$ denote a probability density function over the real numbers, with mean $f_\mu$, standard deviation $f_\sigma$ and variance $f_{\sigma^2}$.

   Let $g$ be the sum of $n$ numbers indepently sampled from $f$. Denote its mean, standard deviation and variances as: $g_\mu$, $g_\sigma$ and $g_{\sigma^2}$ respectively.

   In general, which (if any) of the following hold? Multiple answers allowed.

   A. $g_\mu = nf_\mu$ B. $g_\sigma = nf_\sigma$ C. $g_{\sigma^2} = nf_{\sigma^2}$

   Answer:_______________

Your Name: ______________________________

2. In the year 20XX the population distribution of the continents may be approximately:

| Continent | Fraction |
|---|---|
| Asia | $1/2$ |
| Africa | $1/4$ |
| Europe | $1/8$ |
| Americas | $3/32$ |
| Oceania | $1/32$ |

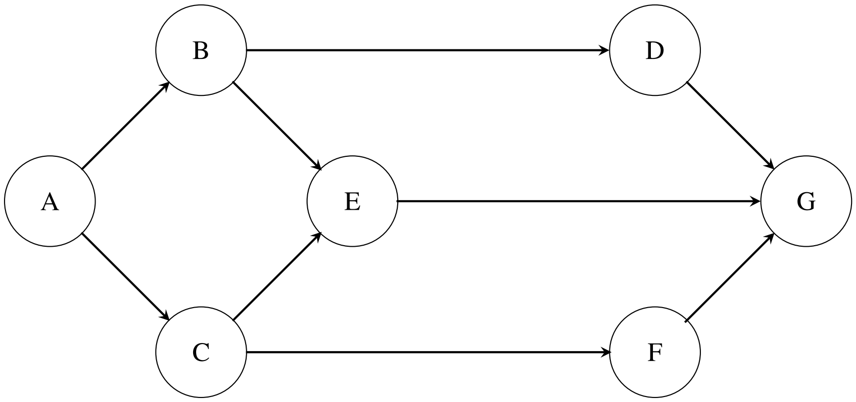
The information theoretic entropy of this distribution is approximately 1.85.

Here "Americas" means North and South America combined. Assuming that North and South America have equal populations, what would the information theoretic entropy be if they are treated separately (i.e. the category "Americas" is split into North and South America).

A. 1.85 B. 1.80 C. 1.65 D. 1.75 E. 1.90 F. 1.95 G. 1.60 H. 2.00

Answer:______________

Your Name: ______________________________




3.

Let $W \perp\!\!\!\perp X \mid Y,Z$ denote the statement: W and X are independent given Y and Z.

And $W \perp\!\!\!\perp X,Y \mid Z$ denote the statement: W and the set {X,Y} are independent given Z.

Which (if any) of the following choices is true? Multiple answers allowed.

1. $B \perp\!\!\!\perp F \mid A,E$ 2. $B \perp\!\!\!\perp F \mid A,E,G$ 3. $C \perp\!\!\!\perp D \mid B,F$ 4. $C \perp\!\!\!\perp D \mid B,G$ 5. $C \perp\!\!\!\perp B,D \mid A$

Answer:_______________

Your Name: ____________________________________________

4. The melting temperatures of magnesium (Mg) and aluminium (Al) are Mg:650 and Al:660 °C respectively.

   Suppose we will receive a sample of either Mg or Al, but do not know in advance which one we will receive. Based on past experience, we expect a 75% chance of Al (and therefore 25% of Mg).

   Further suppose we can measure melting temperature, but with measurement error. The error is unbiased (has a mean of zero) and is well approximated by a normal distribution with a standard deviation of 2°C.

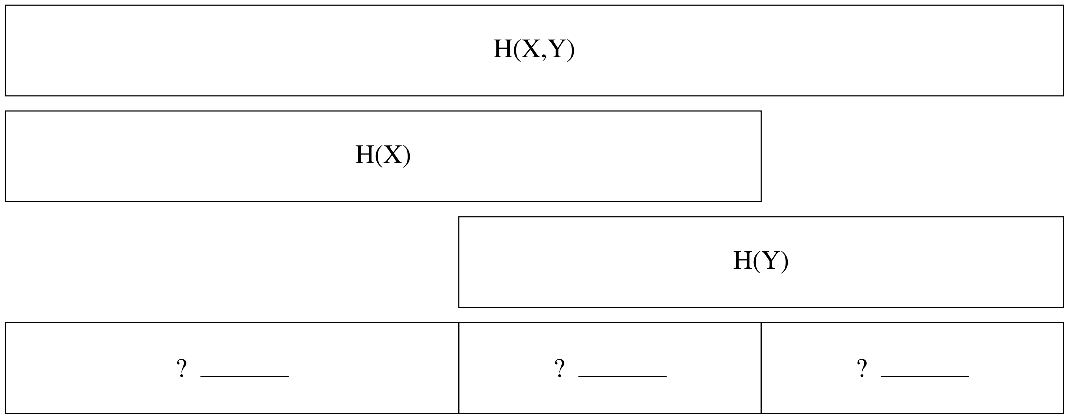
   One sample is received and its melting temperature is measured to be 654 ±2°C.

   Which one of the following is a reasonable estimate of the probability that the sample is Mg?

   A. $1/3e^2$ B. $1/3\sqrt{e}$ C. $1/3e^4$ D. $1/3e^{0.4}$ E. $1/4e^2$ F. $1/4\sqrt{e}$ G. $1/4e^4$ H. $1/4e^{0.4}$ I. None of the above.

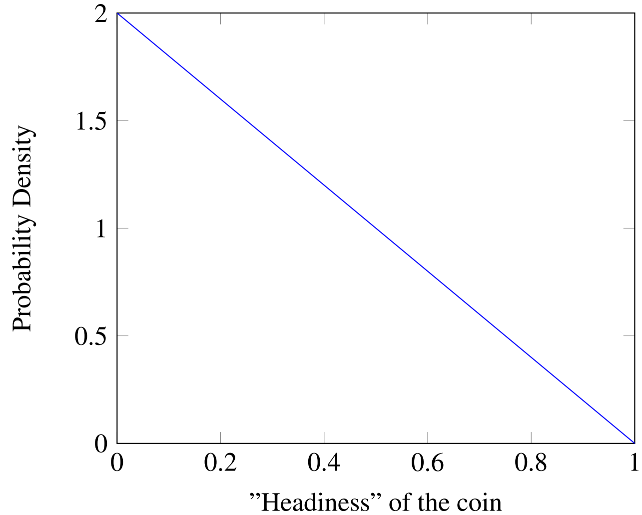
   Answer:________________________

Your Name: ____________________

5. 


In the diagram above, fill in the quantities marked by ? ______.

Your Name: ________________________________________




6. The graph above represents the posterior belief in the "headiness" of a coin after observing one sample of data which happened to be a **tail**. The prior belief regarding the headiness of the coin had the form of a beta distribution with parameters $(a, b)$.

   For your reference: the beta distribution Beta(r:a,b) is defined as: $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} r^{a-1}(1-r)^{b-1}$.

   Question: What values of $(a, b)$ did the prior have?

   $a$:_______________ $b$:_______________