Midterm exam for Fundamentals of Statistical Machine Learning (201910). Open book test.

Your Name: ____

1. Consider a computer program C which outputs 'a' or 'b' each time it runs. It is known that C uses a random generator to output a with probability p_a.

   No one knows the value of p_a, but you have a friend, Thomas, who plans to run the program some number of times and then predict the output of the next run of the program.

   Thomas loves Bayesian statistics, so his posterior probability estimate certainly will include a prior.

   Unfortunately Thomas is coy and will not tell you what his prior is.

   Instead he tells you that if he were to run the program 4 times and get an a each time, he would be 90% sure that next run would also output an a. Likewise, if he were to get 4 out of 4 b's, he would be 90% sure that the next run would output another b.

   Question: Is Thomas using a uniform prior? Explain your answer.

Your Name: ______________________________

2. This question asks about the entropy of a probability distribution in which two categories have been joined together.

   Let F denote a probability distribution over the first k natural numbers: $1, 2, \dots, k$. Denote the probability $\mathrm{F}(i)$ as p_i, so $\sum_1^k p_i \equiv 1$.

   (i.e. F is like a die (骰子) with k-sides, each side having its own probability p_i)

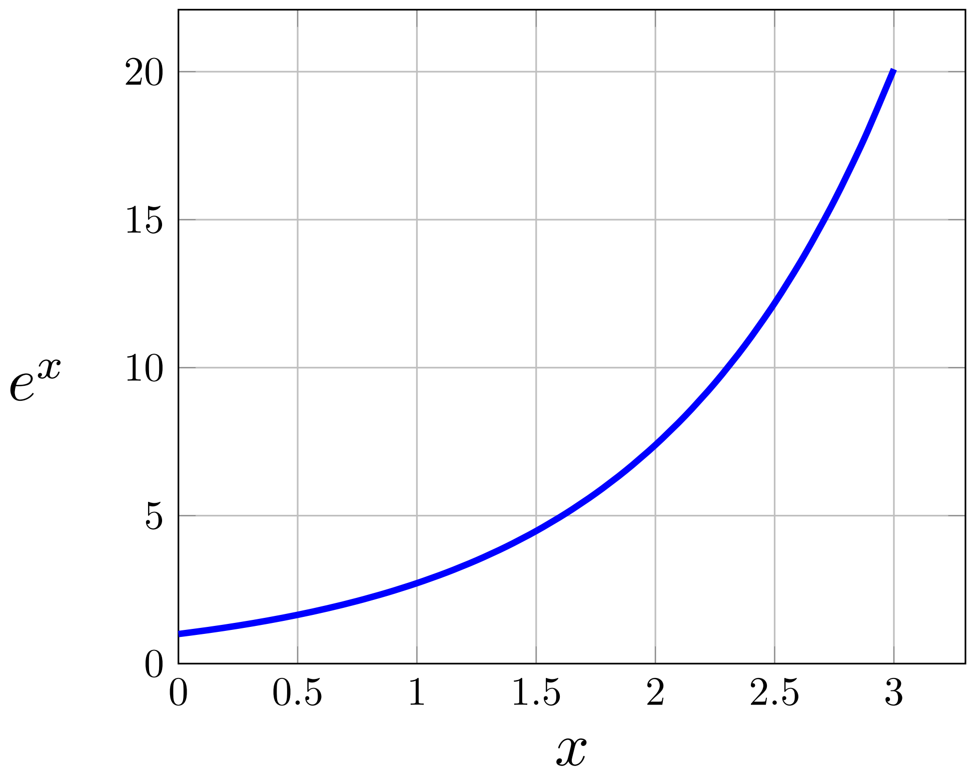
   Let F′ be a probability distribution over $1, 2, \dots, k-1$;
   almost the same as F, but with the last two elements ($k-1$ and k) merged.
   So, $p'_i = p_i$ $1 \leq i < k-1$, and $p'_{k-1} = p_{k-1} + p_k$. There is no p'_k (or equivalently $p'_k \stackrel{\text{def}}{=} 0$).

   Let H() denote information theoretic entropy.

   GIVEN: We know the theoretic entropy of F, H(F) = 5.0; and that $p_{k-1} = p_k = 0.01$.

   1. Give a lower bound on k and explain why k must be a least that large.
   2. What is H(F′), the entropy of F′?

Your Name: ______________________________



3.

Consider a mixture model of two normal (one-dimensional Gaussian) distributions:

$$M(x) = {}^2\!/_3 N_a + {}^1\!/_3 N_b$$

Denoted $N_a : N(\mu_a, \sigma_a^2)$ and $N_b : N(\mu_b, \sigma_b^2)$.

In other words, when generating points from M:
first one of the two components N_a or N_b are selected are random with N_a selected $2/3$ of the time, and then a point is generated according to the mean and variance of the selected component.

A data point x_1 is sampled from $M(x)$.

Consider the probability that x_1 "came from" N_a.

1. Give the general formula for that probability (a formula including x_1,...)
2. Give the approximate numerical value for the probability for the case:

x_1	μ_a	σ_a	μ_b	σ_b
4	-1	2	18	4

Approximation should be made *without* using a caculator (不可以用計算機) and should be within $\pm 10\%$.
The answer can be given as a fraction, e.g. $2/7$ instead of 0.285714 ....

Your Name: ______________________________

Consider a typically classification problem. For example one with m features $\mathrm{F}_1, \dots, \mathrm{F}_m$, which may be symptoms, e.g. fever, coughing, etc.; while the class to be predicted is the person's condition (healthy or disease, etc.).

Consider the following statements:

$$\mathrm{P}[\mathrm{F}_1, \dots, \mathrm{F}_m] \Leftrightarrow \prod_{i=1}^{m} \mathrm{P}[\mathrm{F}_i] \tag{1}$$

$$\mathrm{P}[\mathrm{F}_1, \dots, \mathrm{F}_m|\mathrm{C}] \Leftrightarrow \mathrm{P}[\mathrm{F}_1|\mathrm{C}]\,\mathrm{P}[\mathrm{F}_2|\mathrm{C}, \mathrm{F}_1]\,\mathrm{P}[\mathrm{F}_3|\mathrm{C}, \mathrm{F}_1, \mathrm{F}_2] \cdots \mathrm{P}[\mathrm{F}_m|\mathrm{C}, \mathrm{F}_1, \dots, \mathrm{F}_{m-1}] \tag{2}$$

$$\mathrm{P}[\mathrm{C}|\mathrm{F}_1, \dots, \mathrm{F}_m] \Leftrightarrow \mathrm{P}[\mathrm{C}, \mathrm{F}_1]\,\mathrm{P}[\mathrm{C}, \mathrm{F}_2, \mathrm{F}_1]\,\mathrm{P}[\mathrm{C}, \mathrm{F}_3, \mathrm{F}_1, \mathrm{F}_2] \cdots \mathrm{P}[\mathrm{C}, \mathrm{F}_m, \mathrm{F}_1, \dots, \mathrm{F}_{m-1}] \tag{3}$$

$$\mathrm{P}[\mathrm{C}|\mathrm{F}_1, \dots, \mathrm{F}_m] \Leftrightarrow \prod_{i=1}^{m} \mathrm{P}[\mathrm{C}, \mathrm{F}_i] \tag{4}$$

$$\mathrm{P}[\mathrm{F}_1, \dots, \mathrm{F}_m|\mathrm{C}] \Leftrightarrow \prod_{i=1}^{m} \mathrm{P}[\mathrm{F}_i|\mathrm{C}] \tag{5}$$

For each blank in the table below fill in one of {always, often, seldom, hardly}, where always means always equal, often means often approximately equal, seldom means usually unequal, and hardly means hardly ever equal.

Eq num:	(1)	(2)	(3)	(4)	(5)
Equal?					

Why? Give a the reason for your answers (可以用中文).

Your Name: ____

	start	unknowns	finish
example 1.	$(a-b)(a+b)$	$\equiv \underline{\ ?\ }^2 - \underline{\ ?\ }^2$	$\Rightarrow a^2 - b^2$
example 2.	$x!$	$\equiv \Gamma(\underline{\ ?\ })$	$\Rightarrow \Gamma(x+1)$
problem 1.	$\binom{n}{k}$	$\equiv \frac{\Gamma(\underline{\ ?\ })}{\Gamma(\underline{\ ?\ })\Gamma(\underline{\ ?\ })}$	$\Rightarrow$ ________________
problem 2.	$\mathrm{Binomal}(k\|p,n)$	$\underline{\ ?\ }\mathrm{Beta}(\underline{\ ?\ },\underline{\ ?\ },\underline{\ ?\ })$	$\Rightarrow$ ________Beta(________)

Problem 1. Fill in problem 1. in the table above relating $\binom{n}{k}$ to $\frac{\Gamma(\underline{\ ?\ })}{\Gamma(\underline{\ ?\ })\Gamma(\underline{\ ?\ })}$.

The Beta distribution is defined as:

$$\mathrm{Beta}(x|a,b) \stackrel{\mathrm{def}}{=} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

Problem 2. Express the Beta distribution parameters (a,b) and data point x in terms of (k, p and n), such that

$$\mathrm{Binomal}(k|p,n) = \mathrm{F}(k,p,n)\mathrm{Beta}(x|a,b)$$

Where $\mathrm{F}(,p,n)$ is a simple function of k, p, and n.