Probability Review for Final Exam
18.05 Spring 2014

Jeremy Orloff and Jonathan Bloom

Unit 1: Probability

1. Sets.
2. Counting.
3. Sample space, outcome, event, probability function.
4. Probability: conditional probability, independence, Bayes theorem.
5. Discrete random variables: events, pmf, cdf.
6. Bernoulli(p), binomial(n, p), geometric(p), uniform(n)
7. $E(X)$, Var(X), σ
8. Continuous random variables: pdf, cdf.
9. uniform(a,b), exponential(λ), normal(μ,σ)
10. Transforming random variables.
11. Quantiles.
12. Central limit theorem, law of large numbers, histograms.
13. Joint distributions: pmf, pdf, cdf, covariance and correlation.

Sets and counting

- Sets: $\emptyset$, union, intersection, complement Venn diagrams, products
- Counting: inclusion-exclusion, rule of product, permutations ${}_nP_k$, combinations ${}_nC_k = \binom{n}{k}$

Problem 1. Consider the nucleotides A, G, C, T.

(a) How many ways are there to make a sequence of 5 nucleotides.

(b) How many sequences of length 5 are there where no adjacent nucleotides are the same

(c) How many sequences of length 5 have exactly one A?

Problem 2. **(a)** How many 5 card poker hands are there?

(b) How many ways are there to get a full house (3 of one rank and 2 of another)?

(c) What's the probability of getting a full house?

Problem 3. How many arrangements of the letters in the word probability are there?

Probability

- Sample space, outcome, event, probability function. Rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
  Special case: $P(A^c) = 1 - P(A)$
  (A and B disjoint $\Rightarrow P(A \cup B) = P(A) + P(B)$.)
- Conditional probability, multiplication rule, trees, law of total probability, independence
- Bayes' theorem, base rate fallacy

Problem 4. Let E and F be two events for which one knows that the probability that at least one of them occurs is 3/4. What is the probability that neither E nor F occurs?

Problem 5. Let C and D be two events for which one knows that $P(C) = 0.3$, $P(D) = 0.4$, and $P(C \cap D) = 0.2$.

What is $P(C^c \cap D)$?

Problem 6. We toss a coin three times. For this experiment we choose the sample space

$$\Omega = \{HHH, THH, HTH, HHT, TTH, THT, HTT, TTT\}.$$

(a) Write down the set of outcomes corresponding to each of the following events:
A = 'we throw tails exactly two times'
B = 'we throw tails at least two times'
C = 'tails did not appear *before* a head appeared'
D = 'the first throw results in tails'

(b) Write down the set of outcomes corresponding to each of the following events: A^c, $A \cup (C \cap D)$, and $A \cap D^c$.

Problem 7. Suppose we have 8 teams labeled $T_1, \ldots, T_8$. Suppose they are ordered by placing their names in a hat and drawing the names out one at a time.

(a) How many ways can it happen that all the odd numbered teams are in the odd numbered slots and all the even numbered teams are in the even numbered slots?

(b) What is the probability of this happening?

Problem 8. Suppose you want to divide a 52 card deck into four hands with 13 cards each. What is the probability that each hand has a king?

Problem 9. A fair die is thrown twice. A is the event 'sum of throws equals 4," B is "at least one of the throws is 3."

(a) Calculate $P(A|B)$.

(b) Are A and B independent?

Problem 10. A Dutch cow is tested for BSE (mad cow disease), using a test with $P(T|B) = 0.70$ and $P(T|B^c) = 0.10$. Here T is the event of a positive test and B is the event of having BSE. The risk of BSE is $P(B) = 1.3 \times 10^{-5}$. Compute $P(B|T)$ and $P(B|T^c)$.

Problem 11. A student takes a multiple-choice exam. Suppose for each question he either know the answer or gambles and chooses an option at random. Further suppose that if he knows the answer, the probability of a correct answer is 1, and if he gambles this probability is 1/4. To pass, students need to answer at least 60% of the questions correctly. The student has "studied for a minimal pass," i.e., with probability 0.6 he knows the answer to a question. Given that he answers a question correctly, what is the probability that he actually *knows* the answer?

Problem 12. Suppose you have an urn containing 7 red and 3 blue balls. You draw three balls at random. On each draw, if the ball is red you set it aside and if the ball is blue you put it back in the urn. What is the probability that the third draw is blue?
(If you get a blue ball it counts as a draw even though you put it back in the urn.)

Problem 13. Independence
Suppose that $P(A) = 0.4, P(B) = 0.3$ and $P((A \cup B)^C) = 0.42$. Are A and B independent?

Problem 14. Suppose that events A, B and C are *mutually independent* with

$$P(A) = 0.3, \quad P(B) = 0.4, \quad P(C) = 0.5.$$

Compute the following: (Hint: Use a Venn diagram)
(i) $P(A \cap B \cap C^c)$ (ii) $P(A \cap B^c \cap C)$ (iii) $P(A^c \cap B \cap C)$

Problem 15. We choose a month of the year, in such a manner that each month has the same probability. Find out whether the following events are independent:

(a) The event 'outcome is an even numbered month' and the event 'outcome is in the first half of the year.'

(b) The event 'outcome is an even numbered month' and the event'outcome is a summer month' (i.e.,June, July, August).

Problem 16. Suppose A and B are events with $0 < P(A) < 1$ and $0 < P(B) < 1$.
(a) If A and B are disjoint, can they be independent?
(b) If A and B are independent, can they be disjoint?
(c) If $A \subset B$, can A and B be independent?
(d) If A and B are independent, can A and $A \cup B$ be independent?

Random variables, expectation and variance

- Discrete random variables: events, pmf, cdf
- Bernoulli(p), binomial(n, p), geometric(p), uniform(n)
- $E(X)$, meaning, algebraic properties, $E(h(X))$
- Var(X), meaning, algebraic properties
- Continuous random variables: pdf, cdf
- uniform(a,b), exponential(λ), normal(μ,σ)
- Transforming random variables
- Quantiles

Problem 17. Directly from the definitions of expected value and variance, compute $E(X)$ and Var(X) when X has probability mass function given by the following table:

X	-2	-1	0	1	2
p(X)	1/15	2/15	3/15	4/15	5/15

Problem 18. Suppose that X takes values between 0 and 1 and has probability density function $2x$. Compute Var(X) and Var(X^2).

Problem 19. The pmf of X is given by

$$P(X = -1) = \frac{1}{5}, \quad P(X = 0) = \frac{2}{5}, \quad P(X = 1) = \frac{2}{5}.$$

(a) Compute $E(X)$.
(b) Give the pdf of $Y = X^2$ and use it to compute $E(Y)$.
(c) Instead, compute $E(X^2)$ directly from an extended table.
(d) Determine $\text{Var}(X)$.

Problem 20. For a certain random variable X it is known that $E(X) = 2$ and $\text{Var}(X) = 3$. What is $E(X^2)$?

Problem 21. Determine the expectation and variance of a Bernoulli(p) random variable.

Problem 22. Suppose 100 people all toss a hat into a box and then proceed to randomly pick out a hat. What is the expected number of people to get their own hat back.

Hint: express the number of people who get their own hat as a sum of random variables whose expected value is easy to compute.

pmf, pdf, cdf

Probability Mass Functions, Probability Density Functions and Cumulative Distribution Functions

Problem 23. Suppose that $X \sim \text{Bin}(n, 0.5)$. Find the probability mass function of $Y = 2X$.

Problem 24. Suppose that X is uniform on $[0, 1]$. Compute the pdf and cdf of X. If $Y = 2X + 5$, compute the pdf and cdf of Y.

Problem 25. Now suppose that X has probability density function $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. Compute the cdf, $F_X(x)$. If $Y = X^2$, compute the pdf and cdf of Y.

Problem 26. Suppose that X is a random variable that takes on values 0, 2 and 3 with probabilities 0.3, 0.1, 0.6 respectively. Let $Y = 3(X - 1)^2$.

(a) What is the expectation of X?

(b) What is the variance of X?

(c) What is the expection of Y?

(d) Let $F_Y(t)$ be the cumulative density function of Y. What is $F_Y(7)$?

Problem 27. Suppose you roll a fair 6-sided die 25 times (independently), and you get \$3 every time you roll a 6.

Let X be the total number of dollars you win.

(a) What is the pmf of X.

(b) Find $E(X)$ and $\mathrm{Var}(X)$.

(c) Let Y be the total won on another 25 independent rolls. Compute and compare $E(X+Y)$, $E(2X)$, $\mathrm{Var}(X+Y)$, $\mathrm{Var}(2X)$.

Explain briefly why this makes sense.

Problem 28. A continuous random variable X has PDF $f(x) = x + ax^2$ on [0,1] Find a, the CDF and $P(.5 < X < 1)$.

Problem 29. (**PMF of a sum**)

Let X and Y be two independent random variables, where $X \sim \mathrm{Ber}(p)$ and $Y \sim \mathrm{Ber}(q)$. When $p = q = r$, we know that $X + Y$ has a $\mathrm{bin}(2, r)$ distribution. Suppose $p = 1/2$ and $q = 1/4$. Determine $P(X + Y = k)$, for $k = 0, 1, 2$, and conclude that $X + Y$ does not have a binomial distribution.

Problem 30. Let X be a discrete random variable with pmf p given by:

a	-1	0	1	2
$p(a)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$

(a) Let $Y = X^2$. Calculate the pmf of Y.
(b) Calculate the value of the cdf's of X and Y at $a = 1$, $a = 3/4$, and $a = \pi - 3$.

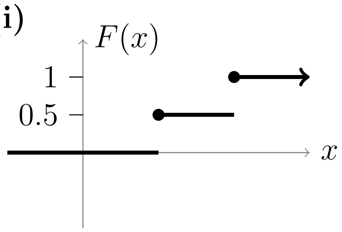
Problem 31. Suppose that the cdf of X is given by:

$$F(a) = \begin{cases} 0 & \text{for } a < 0 \\ \frac{1}{3} & \text{for } 0 \le a < \frac{1}{2} \\ \frac{1}{2} & \text{for } \frac{1}{2} \le a < \frac{3}{4} \\ 1 & \text{for } a \ge \frac{3}{4}. \end{cases}$$

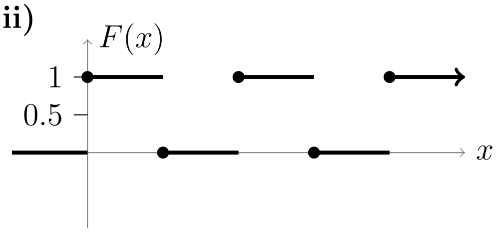
Determine the pmf of X.

Problem 32. For each of the following say whether it can be the graph of a cdf. If it can be, say whether the variable is discrete or continuous.

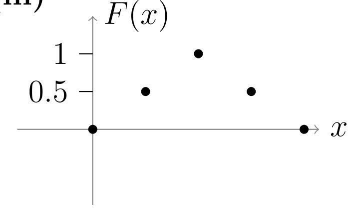
(i)



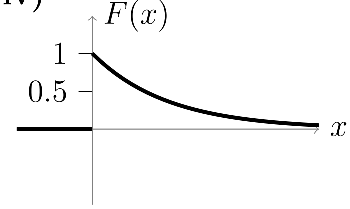
(ii)



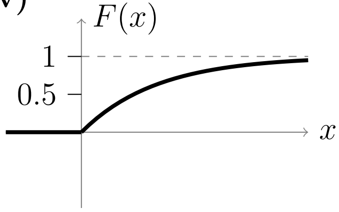
(iii)



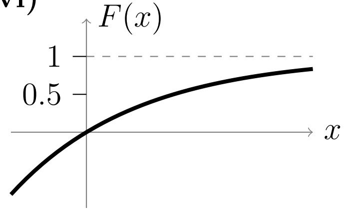
(iv)



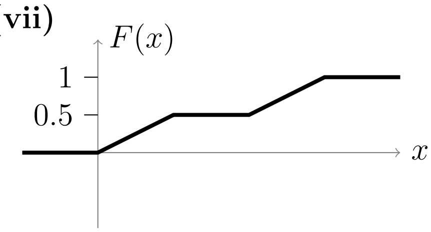
(v)



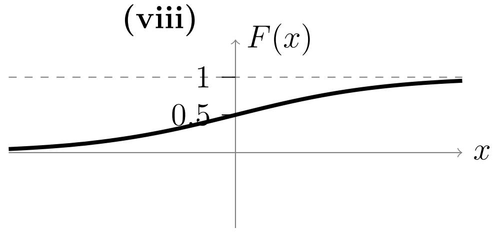
(vi)



(vii)



(viii)



Problem 33. Suppose X has range [0,1] and has cdf

$$F(x) = x^2 \quad \text{for } 0 \leq x \leq 1.$$

Compute $P(\frac{1}{2} < X < \frac{3}{4})$.

Problem 34. Let X be a random variable with range $[0, 1]$ and cdf

$$F(X) = 2x^2 - x^4 \quad \text{for } 0 \leq x \leq 1.$$

(a) Compute $P(\frac{1}{4} \leq X \leq \frac{3}{4})$.
(b) What is the pdf of X?

Distributions with names

Problem 35. Suppose that buses arrive are scheduled to arrive at a bus stop at noon but are always X minutes late, where X is an exponential random variable with probability density function $f_X(x) = \lambda e^{-\lambda x}$. Suppose that you arrive at the bus stop precisely at noon.

(a) Compute the probability that you have to wait for more than five minutes for the bus to arrive.

(b) Suppose that you have already waiting for 10 minutes. Compute the probability that you have to wait an additional five minutes or more.

Problem 36. Normal Distribution: Throughout these problems, let ϕ and Φ be the pdf and cdf, respectively, of the standard normal distribution Suppose Z is a standard normal random variable and let $X = 3Z + 1$.

(a) Express $P(X \leq x)$ in terms of Φ

(b) Differentiate the expression from (a) with respect to x to get the pdf of X, $f(x)$. Remember that $\Phi'(z) = \phi(z)$ and don't forget the chain rule

(c) Find $P(-1 \leq X \leq 1)$

(d) Recall that the probability that Z is within one standard deviation of its mean is approximately 68%. What is the probability that X is within one standard deviation of its mean?

Problem 37. Transforming Normal Distributions
Suppose $Z \sim \mathrm{N}(0{,}1)$ and $Y = \mathrm{e}^Z$.

(a) Find the cdf $F_Y(a)$ and pdf $f_Y(y)$ for Y. (For the CDF, the best you can do is write it in terms of Φ the standard normal cdf.)

(b) We don't have a formula for $\Phi(z)$ so we don't have a formula for quantiles. So we have to write quantiles *in terms* of Φ^{-1}.
(i) Write the .33 quantile of Z in terms of Φ^{-1}
(ii) Write the .9 quatntile of Y in terms of Φ^{-1}.
(iii) Find the median of Y.

Problem 38. (Random variables derived from normal r.v.)

Let X_1, X_2, $\ldots X_n$ be i.i.d. $\mathrm{N}(0, 1)$ random variables.
Let $Y_n = X_1^2 + \ldots + X_n^2$.

(a) Use the formula $\mathrm{Var}(X_j) = E(X_j^2) - E(X_j)^2$ to show $E(X_j^2) = 1$.

(b) Set up an integral in x for computing $E(X_j^4)$.

For 3 extra credit points, use integration by parts show $E(X_j^4) = 3$.
(If you don't do this, you can still use the result in part c.)

(c) Deduce from parts (a) and (b) that $\mathrm{Var}(X_j^2) = 2$.

(d) Use the Central Limit Theorem to approximate $P(Y_{100} > 110)$.

Problem 39. More Transforming Normal Distributions
(a) Suppose Z is a standard normal random variable and let $Y = aZ + b$, where $a > 0$ and b are constants.

Show $Y \sim \mathrm{N}(b, a^2)$.

(b) Suppose $Y \sim \mathrm{N}(\mu, \sigma^2)$. Show $\dfrac{Y - \mu}{\sigma}$ follows a standard normal distribution.

Problem 40. (**Sums of normal random variables**)

Let X be independent random variables where $X \sim N(2, 5)$ and $Y \sim N(5, 9)$ (we use the notation $N(\mu, \sigma^2)$). Let $W = 3X - 2Y + 1$.

a) Compute $E(W)$ and $\text{Var}(W)$.

b) It is known that the sum of independent normal distributions is normal. Compute $P(W \le 6)$.

Problem 41. Let $X \sim \text{U}(a, b)$. Compute $E(X)$ and $\text{Var}(X)$.

Problem 42. In $n + m$ independent Bernoulli(p) trials, let S_n be the number of successes in the first n trials and T_m the number of successes in the last m trials.

(a) What is the distribution of S_n? Why?

(b) What is the distribution of T_m? Why?

(c) What is the distribution of $S_n + T_m$? Why?

(d) Are S_n and T_m independent? Why?

Problem 43. Compute the median for the exponential distribution with parameter λ.

Joint distributions

- Joint pmf, pdf, cdf.
- Marginal pmf, pdf, cdf
- Covariance and correlation.

Problem 44. Correlation

Flip a coin 3 times. Use a joint pmf table to compute the covariance and correlation between the number of heads on the first 2 and the number of heads on the last 2 flips.

Problem 45. Correlation

Flip a coin 5 times. Use properties of covariance to compute the covariance and correlation between the number of heads on the first 3 and last 3 flips.

Problem 46. (**Another Arithmetic Puzzle**)

Let X and Y be two independent Bernoulli(.5) random variables. Define U and V by:

$$U = X + Y \quad \text{and} \quad V = |X - Y|.$$

(a) Determine the joint pmf and marginal pmf's of U and V.
(b) Find out whether U and V are dependent or independent.

Problem 47. To investigate the relationship between hair color and eye color, the hair color and eye color of 5383 persons was recorded. Eye color is coded by the values 1 (Light) and 2 (Dark), and hair color by 1 (Fair/red), 2 (Medium), and 3 (Dark/black). The data are given in the following table:

Eye \ Hair	1	2	3
1	1168	825	305
2	573	1312	1200

The table is turned into a joint pdf for X (hair color) and Y (eye color).
(a) Determine the joint and marginal pmf of X and Y.
(b) Are X and Y independent?

Problem 48. Let X and Y be two continuous random variables with joint pdf

$$f(x,y) = \frac{12}{5}xy(1+y) \quad \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1,$$

and $f(x) = 0$ otherwise.
(a) Find the probability $P(\frac{1}{4} \le X \le \frac{1}{2}, \frac{1}{3} \le Y \le \frac{2}{3})$.
(b) Determine the joint cdf of X and Y for a and b between 0 and 1.
(c) Use your answer from (b) to find marginal cdf $F_X(a)$ for a between 0 and 1.
(d) Find the marginal pdf $f_X(x)$ directly from $f(x,y)$ and check that it is the derivative of $F_X(x)$.
(e) Are X and Y independent?

Problem 49. Let X and Y be two random variables and let r, s, t, and u be arbitrary real numbers.
(a) Derive from the definition that $\mathrm{Cov}(X+s, Y+u) = \mathrm{Cov}(X,Y)$.
(b) Derive from the definition that $\mathrm{Cov}(rX, tY) = rt\mathrm{Cov}(X,Y)$.
(c) Combine parts (a) and (b) to show $\mathrm{Cov}(rX+s, tY+u) = rt\mathrm{Cov}(X,Y)$.

Problem 50. Derive the alternative expression for the covariance: $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$.

Problem 51. (**Arithmetic Puzzle**)
The joint pmf of X and Y is partly given in the following table.

$X \setminus Y$	0	1	2	
-1	...	...	...	1/2
1	...	1/2	...	1/2
	1/6	2/3	1/6	1

(a) Complete the table.
(b) Are X and Y independent?

Problem 52. (**Simple Joint Probability**)
Let X and Y have joint pmf given by the table:

$X \setminus Y$	1	2	3	4
1	16/136	3/136	2/136	13/136
2	5/136	10/136	11/136	8/136
3	9/136	6/136	7/136	12/136
4	4/136	15/136	14/136	1/136

Compute:
(a) $P(X = Y)$.
(b) $P(X + Y = 5)$.
(c) $P(1 < X \leq 3, 1 < Y \leq 3)$.
(d) $P((X, Y) \in \{1, 4\} \times \{1, 4\})$.

Problem 53. Toss a fair coin 3 times. Let X = the number of heads on the first toss, Y the total number of heads on the last two tosses, and Z the number of heads on the first two tosses.

(a) Give the joint probability table for X and Y. Compute $\mathrm{Cov}(X, Y)$.

(b) Give the joint probability table for X and Z. Compute $\mathrm{Cov}(X, Z)$.

Problem 54. Let X be a random variable that takes values -2, -1, 0, 1, 2; each with probability 1/5. Let $Y = X^2$.

(a) Fill out the following table giving the joint frequency function for X and Y. Be sure to include the marginal probabilities.

X / Y	-2	-1	0	1	2	total
0						
1						
4						
total						

(b) Find $E(X)$ and $E(Y)$.

(c) Show X and Y are not independent.

(d) Show $\mathrm{Cov}(X, Y) = 0$.
This is an example of uncorrelated but non-independent random variables. The reason this can happen is that correlation only measures the linear dependence between the two variables. In this case, X and Y are not at all linearly related.

Problem 55. Continuous Joint Distributions

Suppose X and Y are continuous random variables with joint density function $f(x,y) = x+y$ on the unit square $[0,1] \times [0,1]$.

(a) Let $F(x,y)$ be the joint CDF. Compute $F(1,1)$. Compute $F(x,y)$.

(b) Compute the marginal densities for X and Y.

(c) Are X and Y independent?

(d) Compute $E(X)$, (Y), $E(X^2+Y^2)$, $\mathrm{Cov}(X,Y)$.

Law of Large Numbers, Central Limit Theorem

Problem 56. Suppose $X_1, \ldots, X_{100}$ are i.i.d. with mean 1/5 and variance 1/9. Use the central limit theorem to estimate $P(\sum X_i < 30)$.

Problem 57. All or None
You have \$100 and, never mind why, you must convert it to \$1000. Anything less is no good. Your only way to make money is to gamble for it. Your chance of winning one bet is p.

Here are two extreme strategies:

Maximum strategy: bet as much as you can each time. To be smart, if you have less than \$500 you bet it all. If you have more, you bet enough to get to \$1000.

Minimum strategy: bet \$1 each time.

If $p < .5$ (the odds are against you) which is the better strategy?

What about $p > .5$ or $p = .5$?

Problem 58. (**Central Limit Theorem**)

Let $X_1, X_2, \ldots, X_{144}$ be i.i.d., each with expected value $\mu = E(X_i) = 2$, and variance $\sigma^2 = \mathrm{Var}(X_i) = 4$. Approximate $P(X_1 + X_2 + \cdots X_{144} > 264)$, using the central limit theorem.

Problem 59. (**Binomial $\approx$ normal**)

Let $X \sim \mathrm{bin}(n,p)$.
(a) An exact computation yields $P(X \leq 25) = 0.55347$, when $n = 100$ and $p = 1/4$. Use the central limit theorem to give an approximation of $P(X \leq 25)$ and $P(X < 26)$.

(b) When $n = 100$ and $p = 1/4$, then $P(X \leq 2) = 1.87 \times 10^{-10}$. Use the central limit theorem to give an approximation of this probability.

Problem 60. (**More Central Limit Theorem**)

The average IQ in a population is 100 with standard deviation 15 (by definition, IQ is normalized so this is the case). What is the probability that a randomly selected group of 100 people has an average IQ above 115?

Hospitals (binomial, CLT, etc)

- A certain town is served by two hospitals.
- Larger hospital: about 45 babies born each day.
- Smaller hospital about 15 babies born each day.
- For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys.

(a) Which hospital do you think recorded more such days?

(i) The larger hospital. (ii) The smaller hospital.
(iii) About the same (that is, within 5% of each other).

(b) Let L_i (resp., S_i) be the Bernoulli random variable which takes the value 1 if more than 60% of the babies born in the larger (resp., smaller) hospital on the i^{th} day were boys. Determine the distribution of L_i and of S_i.

(c) Let L (resp., S) be the number of days on which more than 60% of the babies born in the larger (resp., smaller) hospital were boys. What type of distribution do L and S have? Compute the expected value and variance in each case.

(d) Via the CLT, approximate the .84 quantile of L (resp., S). Would you like to revise your answer to part (a)?

(e) What is the correlation of L and S? What is the joint pmf of L and S? Visualize the region corresponding to the event $L > S$. Express $P(L > S)$ as a double sum.