Class 27: review for final exam
18.05, Spring 2014

This does not cover everything on the final.
Look at the posted practice problems for other topics.

To save time in class: set up, but do not carry out computations.

Problem 1. (Counting)
(a) How many arrangements of the letters in the word probability are there?

(b) Suppose all of these arrangements are written in a list and one is chosen at random. What is the probability it begins with 'b'.

Problem 2. (Probability)
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 3. (Counting)
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 4. (Conditional probability)
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 5. (Bayes formula)
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 6. (Bayes formula)
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 7. (Expected value and variance)
Directly from the definitions of expected value and variance, compute $E(X)$ and $\mathrm{Var}(X)$ 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 8. (Expected value and variance)
Suppose that X takes values between 0 and 1 and has probability density function $2x$. Compute $\mathrm{Var}(X)$ and $\mathrm{Var}(X^2)$.

Problem 9. (Expected value and variance)
For a certain random variable X it is known that $E(X) = 2$ and $\mathrm{Var}(X) = 3$. What is $E(X^2)$?

Problem 10. (Expected value and variance)
Determine the expectation and variance of a Bernoulli(p) random variable.

Problem 11. (Expected value)
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.

Problem 12. (Expected value and variance)
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 13. (Continuous random variables)
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 14. (Exponential distribution)
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 15. Transforming Normal Distributions
Suppose $Z \sim$ N(0,1) and $Y = e^Z$.

(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 16. 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 17. (Quantiles)
Compute the median for the exponential distribution with parameter λ.

Problem 18. (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 19. (Joint distributions)
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 $f(x,y) = 0$ 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 20. (Joint distributions)
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 21. 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 22. (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?

Problem 23. (NHST chi-square)
A study of recidivism (repeat offenses) of juvenile offenders used an experimental design with random assignment of juveniles to experimental intervention (Family Group Counseling) or control group (diversion programs). 70 out of 200 people in the control group re-offended and 30 out of 200 people in the experimental group re-offended.

Use a chi-square significance test to test whether the recidivism rates within 6 months for the two experimental groups are significantly different at a significance level of 0.05.

Problem 24. (Confidence intervals)
Suppose that against a certain opponent the number of points the MIT basketaball team scores is normally distributed with unknown mean θ and unknown variance, σ^2.

Suppose that over the course of the last 10 games between the two teams MIT scored the following points:

$$59, 62, 59, 74, 70, 61, 62, 66, 62, 75$$

Compute a 95% t–confidence interval for θ. Does 95% confidence mean that the probability θ is in the interval you just found is 95%?

Problem 25. (Confidence intervals)
The volume in a set of wine bottles is known to follow a $\mathrm{N}(\mu, 25)$ distribution. You take a sample of the bottles and measure their volumes. How many bottles do you have to sample to have a 95% confidence interval for μ with width 1?

Problem 26. (Confidence intervals)
Let p be the fraction of the population who prefer candidate A. If you poll 400 people, how many have to prefer candidate A to make the 90% confidence interval entirely in the range where A is preferred.

Problem 27. (Confidence intervals)
Suppose you made 40 confidence intervals with confidence level 95%. About how many of them would you expect to be "wrong'? That is, how many would not actually contain the parameter being estimated? Should you be surprised if 10 of them are wrong?

Problem 28. (Confidence intervals)
A statistician chooses 20 randomly selected class days and counts the number of students present in 18.05. The find a standard deviation of 4.56 students If the number of students present is normally distributed, find the 95% confidence interval for the population standard deviation of the number of students in attendance.

Problem 29. Parametric bootstrap
Suppose we have a sample of size 100 drawn from a geom(p) distribution with unknown p. The MLE estimate for p is given by by $\hat{p} = 1/\bar{x}$. Assume for our data $\bar{x} = 3.30$, so $\hat{p} = 1/\bar{x} = 0.30303$.

(a) Outline the steps needed to generate a parametric bootstrap 90% confidence interval.

(b) Suppose the following sorted list consists of 200 bootstrap means computed from a sample of size 100 drawn from a geometric(0.30303) distribution. Use the list to construct a 90% CI for p.

```
2.68 2.77 2.79 2.81 2.82 2.84 2.84 2.85 2.88 2.89
2.91 2.91 2.91 2.92 2.94 2.94 2.95 2.97 2.97 2.99
3.00 3.00 3.01 3.01 3.01 3.03 3.04 3.04 3.04 3.04
3.04 3.05 3.06 3.06 3.07 3.07 3.07 3.08 3.08 3.08
3.08 3.09 3.09 3.10 3.11 3.11 3.12 3.13 3.13 3.13
3.13 3.15 3.15 3.15 3.16 3.16 3.16 3.16 3.17 3.17
3.17 3.18 3.20 3.20 3.20 3.21 3.21 3.22 3.23 3.23
3.23 3.23 3.23 3.24 3.24 3.24 3.24 3.25 3.25 3.25
3.25 3.25 3.25 3.26 3.26 3.26 3.26 3.27 3.27 3.27
3.28 3.29 3.29 3.30 3.30 3.30 3.30 3.30 3.30 3.31
3.31 3.32 3.32 3.34 3.34 3.34 3.34 3.35 3.35 3.35
3.35 3.35 3.36 3.36 3.37 3.37 3.37 3.37 3.37 3.37
3.38 3.38 3.39 3.39 3.40 3.40 3.40 3.40 3.41 3.42
3.42 3.42 3.43 3.43 3.43 3.43 3.44 3.44 3.44 3.44
3.44 3.45 3.45 3.45 3.45 3.45 3.45 3.45 3.46 3.46
3.46 3.46 3.47 3.47 3.49 3.49 3.49 3.49 3.49 3.50
3.50 3.50 3.52 3.52 3.52 3.52 3.53 3.54 3.54 3.54
3.55 3.56 3.57 3.58 3.59 3.59 3.60 3.61 3.61 3.61
3.62 3.63 3.65 3.65 3.67 3.67 3.68 3.70 3.72 3.72
3.73 3.73 3.74 3.76 3.78 3.79 3.80 3.86 3.89 3.91
```

Problem 30. Empirical bootstrap
Suppose we had 100 data points $x_1, \ldots x_{100}$ with sample median $\hat{q_{0.5}} = 3.3$.

(a) Outline the steps needed to generate an empirical bootstrap 90% confidence interval for the median $q_{0.5}$.

(b) Suppose now that the sorted list in the previous problems consists of 200 empirical bootstrap medians computed from resamples of size 100 drawn from the original data. Use the list to construct a 90% CI for $q_{0.5}$.

Problem 31. Linear regression (least squares)

Set up fitting the least squares line through the points $(1, 1)$, $(2, 1)$, and $(3, 3)$.

MIT OpenCourseWare
https://ocw.mit.edu

18.05 Introduction to Probability and Statistics
Spring 2014

For information about citing these materials or our Terms of Use, visit: https://ocw.mit.edu/terms.