Study guide for midterm:

Schedule:

Spring break: March 2 & March 4

Monday March 9th: Mid-term

Questions on the mid-term will be a mix of questions on this study guide and questions you haven't seen before, but no new topics will be introduced.

In general, you are responsible for the material in the lectures (on the PowerPoint slides).

---

(1) What is the mean of these numbers { -10,-5,0,5,10}.

(2) What is the median of these numbers : { -30,-20,4,10,15}

(3) True or false: The median calculation is more robust to outliers than the mean. Why?

(4) The variance of a set of numbers is 16. What is the standard deviation?

(5) Define the following data-types in R: lists, vectors, arrays, data frames and matrices.

(6) Draw the following probability density functions:

The uniform distribution;

The binomial distribution with p = .25 and N = 100;

The beta distribution with alpha = 1, beta =1

The beta distribution with alpha = 25, beta = 25

The beta distribution with alpha = 100, beta = 25

(7) Define mean and variance for a discrete distribution.

(8) In your own words, describe the law of large numbers.

(9) Define the mean and variance for continuous distributions.

(10) Derive Bayes' law

(11) Understand the Bayesian universe for a discrete two state test (i.e. the Cylon detector)

(12) Define the following terms: posterior probability, prior probability, likelihood probability, marginal probabilites,Bayesian update, joint probability, conditional probability.

(13) Our prior belief is that 20% of all the people on a given ship are cylons. We have a cylon detector where p(positiveTest|cylon) = 0.90 and p(positiveTest|not a cylon) = 0.1

Fill out the Bayesian universe for this situation (you don't have to work out the math). So you can write "(.2) * (.9)" and not .18

	prior	Test positive	Test negative	Marginal probs
cylon				p(cylon) =
not a cylon				p(not a cylon) =
		P(testPositive)=	P(testNegative)=	

We see a positive test in the above situation. What is our posterior probability that the person is a cylon? (Again, you don't have to work out the math: so for example you could say

(0.2) * (0.3) / (0.2 + 0.1)

rather than 0.2)

(14) What is "5 choose 2"

(15) What is the expected mean and variance of the binomial distribution?

(16) What is the difference between pbinom, dbinom and rbinom in R?

(17) Consider three coins with a probability of getting a head of 0.3, 0.5 and 0.7. We pick up one of the coins (we don't know which one). Starting with a uniform prior, fill out the Bayesian universe below… (again, you don't need to work through all the math, so you can say "0.2*0.3" rather than "0.06".

Y

		prior	“H”	“T”	Marginal probs
	$\Pi_1 = 0.3$				
X	$\Pi_2 = 0.5$				
	$\Pi_3 = 0.7$				1/3
Marginal probs			p(H) =	p(T) =	

(18) In the above example, you see a head. What is the posterior probability that you’ve picked up the fair coin (the coin with p(head) =0.5). Show your work.

(19) How is the beta distribution defined? What is the relationship between the beta distribution and the binomial distribution. Define the term conjugate prior.

(20) The beta distribution is used as a prior with alpha=5 and beta=5. Then 26 heads and 24 tails are observed. What are the values for alpha and beta for the beta posterior?

(21) What values for alpha and beta should be used to use the beta distribution with a uniform distribution. Why is the beta distribution not uniform whenever alpha and beta are equal?

(22) You pick up a coin and flip it 50 times, observing 34 heads. To determine if the coin is fair, you type:

```
> binom.test(34,50)

        Exact binomial test

data:  34 and 50
number of successes = 34, number of trials = 50, p-value = 0.01535
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.5330062 0.8047958
sample estimates:
probability of success 
                  0.68 

> 
```

Is this is a two sided or one-sided test?

In your own words, explain how the p-value of 0.01535 is calculated.

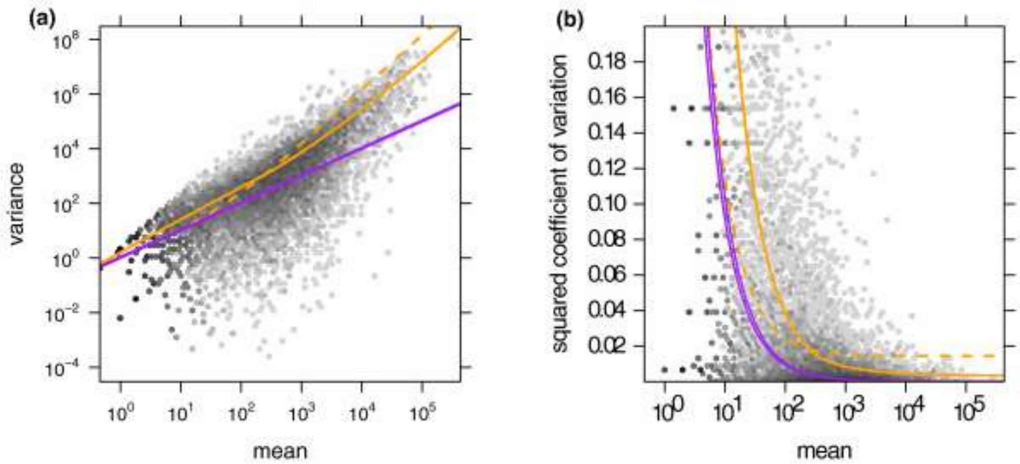
What is the probability of success (0.68). In your own words, how is that defined?

What is the relationship between the hypergeometric distribution and the binomial distribution? Under what circumstances do the hypergeometric and binomial distributions converge?

What is the relationship between the Poisson distribution and the binomial distribution? Under what circumstances do the Poisson distribution and the binomial distribution converge?

(23) In your own words, explain (a) in this figure:

Figure 1.



Dependence of the variance on the mean for condition *A* in the fly RNA-Seq data. (a) The scatter plot shows the common-scale sample variances (Equation (7)) plotted against the common-scale means (Equation (6)). The orange line is the fit $w(q)$. The purple lines show the variance implied by the Poisson distribution for each of the two samples, that is, $\hat{s}_j\hat{q}_{i,A}$. The dashed orange line is the variance estimate used by *edgeR*. (b) Same data as in (a), with the y-axis rescaled to show the squared coefficient of variation (SCV), that is all quantities are divided by the square of the mean. In (b), the solid orange line incorporated the bias correction described in Supplementary Note C in Additional file 1. (The plot only shows SCV values in the range [0, 0.2]. For a zoom-out to the full range, see Supplementary Figure S9 in Additional file 1.)

Anders and Huber *Genome Biology* 2010 **11**:R106 doi:10.1186/gb-2010-11-10-r106
Download authors' original image

(24) What is the relationship between the negative binomial distribution and the binomial distribution?

(25) Choosing among the following distributions: binomial, beta, uniform, poisson, hypergeometric and negative binomial describe which distribution would be most appropriate for each of the following datasets. Where appropriate, indicate what the parameters would be when using the "d" series of functions in R to generate the probability density function.

(A) The results of rolling a fair 100 sided die a large number of times

(B) The results of flipping a fair coin 50 times.

(C) The results of flipping a coin 50 times in which the p(head) = 0.25.

(D) The expected p-values for each gene from an RNA-seq experiment comparing two technical replicates (i.e. where the null hypothesis is always true).

(E) Describe a prior probability capturing our belief about a coin after observing 10 heads and 5 tails.

(F) Describe how many marked cards that would be drawn when drawing 4 cards from a deck with 20 marked and 40 unmarked cards.

(G) Describe how many heads would be expected to be observed from a coin with p(head)=0.9 before observing 3 tails.

(H) For two genomes of 5 billion nucleotides each, the expected number of SNP differences if a SNP difference was expected to occur once every 200,000 nucleotides.

(I) A disease kills 40% of the people who have it. In a cohort of 75 people, the number of people we would expect to survive.

(J) The number of wins that would be expected in a tournament in which a player plays until they lose 5 games with a p(win) = 0.6

(26) The prevalence of a disease within a population is 1 in 10,000. A test is developed with:

p(positive result|a person has the disease) = .99 and
p(positive result|a person does not have the disease) = 0.01

A person takes the test and gets a positive result. What is the posterior probability of our belief that the person has the disease.

(27) What is the average (expected) value for a fair 5 sided die?

(28) What is the difference between the [] and [[]] operators in R?

(29) What is the expected (average) number of heads in 50 flips when the probability of a head = 0.0? When it is 0.5? When it is 1.0?

(30) In your own words, describe how the Metropolitan algorithm works? How does grid approximation work? Given that grid approximation can produce a result that is not dependent on chance, why not always use grid approximation to sample a posterior distribution?