Math 4397/6397 Review

December 8, 2009

1 Set theory

1. The symbol $\subset$ means "is a subset of", and $\in$ means "is an element of".
2. The **sample space**, Ω, is the space of all possible outcomes of an experiment.
3. An **event**, say $A \subset \Omega$, is a subset of Ω.
4. The **union** of two events, $A \cup B$, is the collection of elements that are in A, B or both.
5. The **intersection** of two events, $A \cap B$, is the collection of elements that are in both A and B.
6. The **complement** of an event, say $\bar{A}$ or A^c, is all of the elements of Ω that are not in A.
7. The **null** or **empty** set is denoted $\emptyset$.
8. Two sets are **disjoint** or **mutually exclusive** if their intersection is empty, $A \cap B = \emptyset$.
9. **DeMorgan's laws** state that $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$.

2 Probability essentials

1. A **probability measure**, say P, is a function on the collection of events to $[0, 1]$ so that the following three properties hold:
   a. $P(\Omega) = 1$.
   b. If $A \subset \Omega$ then $P(A) \geq 0$.
   c. If a sequence of events $A_1, A_2 \ldots$, is disjoint then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.
2. $P(A^c) = 1 - P(A)$.
3. The **odds** of an event, A, are $P(A)/(1 - P(A)) = P(A)/P(A^c)$.
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
5. If $A \subset B$ then $P(A) \leq P(B)$.

6. Two events A and B are **independent** if $P(A \cap B) = P(A)P(B)$. A collection of events, $\{A_i\}_{i=1}^n$, are **mutually independent** if for any subset $J \subset \{1, 2, \ldots n\}$, we have $P(\cap_{i \in J} A_i) = \prod_{i \in J} P(A_i)$. If this holds for all sets J with size $|J| = 2$ then we say the collection is **pairwise independent**.

7. Pairwise independence of a collection of events does not imply mutually independence, though the reverse is true.

8. Given that $P(B) > 0$, the conditional probability of A given that B has occurred is $P(A|B) = P(A \cap B)/P(B)$.

9. Two events A and B are **independent** if $P(A|B) = P(A)$.

10. The **law of total probability** states that if A_i are a collection of *mutually exclusive events* so that $\Omega = \cup_{i=1}^n A_i$, then $P(C) = \sum_{i=1}^n P(C|A_i)P(A_i)$ for any event C.

11. **Bayes's rule** states that if A_i are a collection of *mutually exclusive events* so that $\Omega = \cup_{i=1}^n A_i$, then
$$P(A_j|C) = \frac{P(C|A_j)P(A_j)}{\sum_{i=1}^n P(C|A_i)P(A_i)}.$$
for any set C (with positive probability). Notice A and A^c are disjoint and $A \cup A^c = \Omega$ so that we have
$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}.$$

12. The **sensitivity** of a diagnostic test is defined to be $P(+|D)$ where $+$ $(-)$ is the event of a positive (negative) test result and D is the event that a subject has the disease in question. The **specificity** of a diagnostic test is $P(-|D^c)$.

13. Bayes's rule yields that
$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)},$$
and
$$P(D^c|-) = \frac{P(-|D^c)P(D^c)}{P(-|D^c)P(D^c) + P(-|D)P(D)}.$$

14. The **diagnostic likelihood ratio** of a positive test result is $P(+|D)/P(+|D^c)$ = sensitivity/(1− specificity). The likelihood ratio of a negative test result is $P(-|D)/P(-|D^c)$ = 1-sensitivity/specificity.

15. The odds of disease after a positive test are related to the odds of disease before the test by the relation
$$\frac{P(D|+)}{P(D^c|+)} = \frac{P(+|D)}{P(+|D^c)} \frac{P(D)}{P(D^c)}.$$
That is, the posterior odds equal the prior odds times the likelihood ratio. Correspondingly,
$$\frac{P(D^c|-)}{P(D|-)} = \frac{P(-|D^c)}{P(-|D)} \frac{P(D^c)}{P(D)}.$$

3 Random variables

1. A **random variable** is a function from Ω to the real numbers. A random variable is a random number that is the result of an experiment governed by a probability distribution.

2. A **Bernoulli** random variable is one that takes the value 1 with probability p and 0 with probability $(1-p)$. That is, $P(X=1)=p$ and $P(X=0)=1-p$.

3. A **probability mass function** (pmf) is a function that yields the various probabilities associated with a random variable. For example, the probability mass function for a Bernoulli random variable is $f(x)=p^x(1-p)^{1-x}$ for $x=0,1$ as this yields p when $x=1$ and $(1-p)$ when $x=0$.

4. The **expected value** or (population) **mean** of a discrete random variable, X, with pmf $f(x)$ is
$$\mu = E[X] = \sum_x xf(x).$$
The mean of a Bernoulli variable is then $1f(1)+0f(0)=p$.

5. The **variance** of any random variable, X, (discrete or continuous) is
$$\sigma^2 = E\left[(X-\mu)^2\right] = E[X^2] - E[X]^2.$$
The latter formula being the most convenient for computation. The variance of a Bernoulli random variable is $p(1-p)$.

6. The (population) **standard deviation**, σ, is the square root of the variance.

7. **Chebyshev's inequality** states that for any random variable $P(|X-\mu| \geq K\sigma) \leq 1/K^2$. This yields a way to interpret standard deviations.

8. A **binomial** random variable, X, is obtained as the sum of n Bernoulli random variables and has pmf
$$P(X=k) = \binom{n}{k} p^k(1-p)^{n-k}.$$
Binomial random variables have expected value np and variance $np(1-p)$.

4 Continuous random variables

1. **Continuous** random variables take values on the continuum of the real numbers or even higher-dimensional real vector spaces.

2. A continuous random variable X has a **probability density function** (pdf) f if for all $a<b$,
$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$
To be a pdf, a function must be positive and integrate to 1. That is, $\int_{-\infty}^{\infty} f(x)dx = 1$

3. If h is a positive function such that $\int_{-\infty}^{\infty} h(x)dx \leq \infty$ then $f(x) = h(x)/\int_{-\infty}^{\infty} h(x)dx$ is a valid density. Therefore, if we only know a density up to a constant of proportionality, then we can figure out the exact density.

4. The expected value, or mean, of a continuous random variable, X, with pdf f, is

$$\mu = E[X] = \int_{-\infty}^{\infty} tf(t)dt.$$

5. The variance is $\sigma^2 = E[(X-\mu)^2] = E[X^2] - E[X]^2$.

6. The **distribution function**, say F, corresponding to a random variable X with pdf, f, is

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)dt.$$

(Note the common convention that X is used when describing an unobserved random variable while x is for specific values.)

7. The p^{th} **quantile** (for $0 \leq p \leq 1$), say X_p, of a distribution function, say F, is the point so that $F(X_p) = p$. For example, the $.025^{th}$ quantile of the standard normal distribution is -1.96.

5 Properties of expected values and variances

The following properties hold for all expected values (discrete or continuous)

1. Expected values are additive: $E[X+Y] = E[X] + E[Y]$.

2. Multiplicative and additive constants can be pulled out of expected values $E[cX] = cE[X]$ and $E[c+X] = c + E[X]$.

3. For independent random variables, X and Y, $E[XY] = E[X]E[Y]$.

4. In general, $E[h(X)] \neq h(E[X])$.

5. Variances are additive for sums of *independent variables* $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

6. Multiplicative constants are squared when pulled out of variances $\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$.

7. Additive constants do not change variances: $\mathrm{Var}(c+X) = \mathrm{Var}(X)$.

6 The normal distribution

1. The **normal** or **Gaussian** density, often also called "bell curve", is a very common density. It is specified by its mean, μ, and variance, σ^2. The density is given by $f(x) = (2\pi\sigma^2)^{-1/2}\exp\{-(x-\mu)^2/2\sigma^2\}$. We write $X \sim \mathrm{N}(\mu, \sigma^2)$ to denote that X is normally distributed with mean μ and variance σ^2.

2. The **standard normal** density, labeled ϕ, corresponds to a normal density with mean $\mu = 0$ and variance $\sigma^2 = 1$.

$$\phi(z) = (2\pi)^{-1/2} \exp\{-z^2/2\}.$$

The standard normal distribution function is usually labeled Φ.

3. If f is the pdf for a $\mathrm{N}(\mu, \sigma^2)$ random variable, X, then note that $f(x) = \phi\{(x-\mu)/\sigma\}/\sigma$. Correspondingly, if F is the associated distribution function for X, then $F(x) = \Phi\{(x-\mu)/\sigma\}$.

4. If X is normally distributed with mean μ and variance σ^2 then the random variable $Z = (X-\mu)/\sigma$ is standard normally distributed. Taking a random variable subtracting its mean and dividing by its standard deviation is called "standardizing" a random variable.

5. If Z is standard normal then $X = \mu + Z\sigma$ is normal with mean μ and variance σ^2.

6. Approximately 68%, 95% and 99% of the mass of any normal distribution lies within 1, 2 and 3 (respectively) standard deviations from the mean.

7. Henceforth, the quantity z_α refers to the α^{th} quantile of the standard normal distribution. $z_{.90}$, $z_{.95}$, $z_{.975}$ and $z_{.99}$ are 1.28, 1.645, 1.96 and 2.32, respectively.

8. Sums and means of normal random variables are normal (regardless of whether or not they are independent). You can use the rules for expectations and variances to figure out μ and σ.

9. The sample standard deviation of iid normal random variables, appropriated normalized, is a Chi-squared random variable (see below).

7 Sample means and variances

Throughout this section let X_i be a collection of iid random variables with mean μ and variance σ^2.

1. We say random variables are **iid** if they are independent and identically distributed.

2. For random variables, X_i, the **sample mean** is $\bar{X} = \sum_{i=1}^n X_i/n$.

3. $E[\bar{X}] = \mu = E[X_i]$ (does not require the independence or constant variance).

4. If the X_i are iid with variance σ^2 then $\mathrm{Var}(\bar{X}) = \mathrm{Var}(X_i)/n = \sigma^2/n$.

5. The **sample variance** is defined to be

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}.$$

6. $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$ is a shortcut formula for the numerator.

7. $\sigma/\sqrt{n}$ is called the **standard error** of $\bar{X}$. The estimated standard error of $\bar{X}$ is $S/\sqrt{n}$. Do not confuse dividing by this $\sqrt{n}$ with dividing by $n-1$ in the calculation of S^2.

8. An estimator is **unbiased** if its expected value equals the parameter it is estimating.

9. $E[S^2] = \sigma^2$, which is why we divide by $n-1$ instead of n. That is, S^2 is unbiased. However, dividing by $n-1$ rather than n does increase the variance of this estimator slightly, $\text{Var}(S^2) \geq \text{Var}((n-1)S^2/n)$.

10. If the X_i are normally distributed with mean μ and variance σ^2, then $\bar{X}$ is normally distributed with mean μ and variance σ^2/n.

11. The **Central Limit Theorem**. If the X_i are iid with mean μ and (finite) variance σ^2 then

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

will limit to a standard normal distribution. The result is true for small sample sizes, if the X_i iid normally distributed.

12. If we replace σ with S; that is,

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

then Z still limits to a standard normal. If the X_i are iid normally distributed, then Z follows the Students t distribution for small n.

8 Confidence intervals for a mean using the CLT

1. Using the CLT, we know that

$$P\left(-z_{1-\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq z_{1-\alpha/2}\right) \approx 1 - \alpha$$

for large n. Solving the inequalities for μ, we calculated that in repeated sampling, the interval

$$\bar{X} \pm z_{1-\alpha/2}\frac{S}{\sqrt{n}}$$

will contain μ approximately $100(1-\alpha)\%$ of the time.

2. Prior to conducting a study, you can fix the **margin of error** (half width), say δ, of the interval by setting $n = (Z_{1-\alpha/2}\sigma/\delta)^2$. Round up. Requires an estimate of σ.

9 Confidence intervals for a variance and t confidence intervals

1. If X_i are iid normal random variables with mean μ and variance σ^2 then $\frac{(n-1)S^2}{\sigma^2}$ follows what is called a Chi-squared distribution with $n-1$ degrees of freedom.

2. Using the previous item, we know that

$$P\left(\chi^2_{n-1,\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1,1-\alpha/2}\right) = 1-\alpha,$$

where $\chi^2_{n-1,\alpha}$ denotes the α^{th} quantile of the Chi-squared distribution. Solving these inequalities for σ^2 yields

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right]$$

is a $100(1-\alpha)\%$ confidence interval for σ^2. Recall this assumes that the X_i are iid Gaussian random variables.

3. The fact that $(n-1)S^2 \sim \text{Gamma}((n-1)/2, 2\sigma^2)$ can be used to create a likelihood function for σ or σ^2.

4. Chi-squared confidence intervals and the likelihood function depend heavily on the normality assumption.

5. If Z is standard normal and X is and independent Chi-squared with df degrees of freedom then $\frac{Z}{\sqrt{X/df}}$ follows what is called a Student's t distribution with df degrees of freedom.

6. The Student's t density looks like a normal density with heavier tails (so it looks more squashed down).

7. By the previous item, if the X_i are iid $\text{N}(\mu, \sigma^2)$ then

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

follows a Student's t distribution with $(n-1)$ degrees of freedom. Therefore if $t_{n-1,\alpha}$ is the α^{th} quantile of the Student's t distribution then

$$\bar{X} \pm t_{n-1,1-\alpha/2}\frac{S}{\sqrt{n}}$$

is a $100(1-\alpha)\%$ confidence interval for μ.

8. The Student's t confidence interval assumes normality of the X_i. However, the t distribution has quite heavy tails and so the interval is conservative and works well in many situations.

9. For large sample sizes, the Student's t and CLT based intervals are nearly the same because the Student's t quantiles become more and more like standard normal quantiles as n increases.

10. For small sample sizes, it is difficult to diagnose normality/lack of normality. Regardless, the robust t interval should be your default option.

10 Summarizing and displaying data

1. The p^{th} **empirical quantile** of a data set is that point so that $100p\%$ of the data lies below it. The sample **median** is the $.50^{th}$ quantile. Empirical quantiles estimate population quantiles.

2. A **boxplot** plots a box with a centerline at the sample median and the box edges at the lower and upper quartiles. "Whiskers" extend to the largest data point that is within 1.5 of the IQR (inter quartile range). Side by side boxplots are useful to compare groups.

3. A **quantile-quantile** (qq) plot, plots empirical quantiles versus the theoretical quantiles. For normal random variables with mean μ and variance σ^2, let X_p be the p^{th} quantile. Then, $X_p = \mu + Z_p\sigma$. Therefore plotting the empirical quantiles versus the standard normal quantiles can be used to diagnose non-normality (a **normal qq** plot). Any deviation from a straight line indicates non-normality.

4. **Histograms** and **stem and leaf** plots give information about the density.

11 Hypothesis testing for a single mean

1. The null, or status quo, hypothesis is labeled H_0, the alternative H_a or H_1 or H_2 ...

2. A **type I error** occurs when we falsely reject the null hypothesis. The probability of a type I error is usually labeled α.

3. A **type II error** occurs when we falsely fail to reject the null hypothesis. A type II error is usually labeled β.

4. A **Power** is the probability that we correctly reject the null hypothesis, $1 - \beta$.

5. The Z test for $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$ or $H_2 : \mu \neq \mu_0$ or $H_3 : \mu > \mu_0$ constructs a test statistic $TS = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ and rejects the null hypothesis when

   H_1 $TS \leq -z_{1-\alpha}$

   H_2 $|TS| \geq z_{1-\alpha/2}$

   H_3 $TS \geq z_{1-\alpha}$

   respectively.

6. The Z test requires the assumptions of the CLT and for n to be large enough for it to apply.

7. If n is small, then a Student's t test is performed exactly in the same way, with the normal quantiles replaced by the appropriate Student's t quantiles and $n - 1$ df.

8. Tests define confidence intervals by considering the collection of values of μ_0 for which you fail to reject a two sided test. This yields exactly the t and z confidence intervals respectively.

9. Conversely, confidence intervals define tests by the rule where one rejects H_0 if μ_0 is *not in* the confidence interval.

10. A **p-value** is the probability of getting evidence as extreme or more extreme than we actually got under the null hypothesis. For H_3 above, the p-value is calculated as $P(Z \geq TS_{obs}|\mu = \mu_0)$ where TS_{obs} is the observed value of our test statistic. To get the P-value for H_2, calculate a one sided P-value and double it.

11. The p-value is also called the **attained significance level**. This is the smallest α value for which we would have rejected the null hypothesis. Therefore, rejecting the null hypothesis if a P-value is less than α is the same as performing the rejection region test.

12. The power of a Z test for H_3 is given by the formula (know how this is obtained)

$$P(TS > Z_{1-\alpha}|\mu = \mu_1) = P\left(Z \geq \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + Z_{1-\alpha}\right).$$

    Notice that power required a value for μ_1, the value under the null hypothesis. Correspondingly for H_1 we have

$$P\left(Z \leq \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - Z_{1-\alpha}\right).$$

    For H_2, the power is approximately the appropriate one sided power using $\alpha/2$.

13. Some facts about power.

    a. Power goes up as α goes down.

    b. Power of a one sided test is greater than the power of the associated two sided test.

    c. Power goes up as μ_1 gets further away from μ_0.

    d. Power goes up as n goes up.

14. The power formula can be used to calculate the sample size. For example, using the power formula for H_1, setting $Z_{1-\beta} = \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - Z_{1-\alpha}$ yields

$$n = \frac{(Z_{1-\beta} + Z_{1-\alpha})^2\sigma^2}{(\mu_0 - \mu_1)^2},$$

    which gives the sample size to have power $= 1 - \beta$. This formula applies for H_3 also. For the two sided test, H_2, replace α by $\alpha/2$.

15. Determinants of sample size.

    a. n gets larger as α gets smaller.

    b. n gets larger as the power you want gets larger.

    c. n gets lager the closer μ_1 is to μ_0.

12 Binomial confidence intervals

1. Binomial distributions are used to model proportions. If $X \sim \text{Binomial}(n, p)$ then $\hat{p} = X/n$ is a sample proportion.

2. $\hat{p}$ has the following properties.

   a. It is a sample mean of Bernoulli random variables.

   b. It has expected value p.

   c. It has variance $p(1-p)/n$. Note that the largest value that $p(1-p)$ can take is $1/4$ at $p = 1/2$.

   d. $Z = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}}$ follows a standard normal distribution for large n by the CLT.

3. The **Wald confidence interval** for a binomial proportion is

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}.$$

13 The likelihood for a binomial parameter p

1. The **likelihood** for a parameter is the probability density of a given outcome *viewed as a function of the parameter.*

2. The binomial likelihood for observed data x is proportional to $p^x(1-p)^{n-x}$.

3. The **principle of maximum likelihood** states that a good estimate of the parameter is the one that makes the data that was actually observed most probable. That is, the principle of maximum likelihood says that a good estimate of the parameter is the one that maximizes the likelihood.

   a. The maximum likelihood estimate for p is $\hat{p} = X/n$.

   b. The maximum likelihood estimate for μ for iid $\text{N}(\mu, \sigma^2)$ data is $\bar{X}$. The maximum likelihood estimate for σ^2 is $(n-1)S^2/n$ (the biased sample variance).

4. **Likelihood ratios** represent the relative evidence comparing one hypothesized value of the parameter to another.

5. Likelihoods are usually plotted so that the maximum value (the value at the ML estimate) is 1. Where reference lines at $1/8$ and $1/32$ intersect the likelihood depict **likelihood intervals**. Points lying within the $1/8$ reference line, for example, are such that no other parameter value is more than 8 times better supported given the data.

14 Group comparisons

1. For group comparisons, make sure to differentiate whether or not the observations are paired (or matched) versus independent.

2. For paired comparisons for continuous data, one strategy is to calculate the **differences** and use the methods for testing and performing hypotheses regarding a single mean. The resulting tests and confidence intervals are called **paired Student's** t tests and intervals respectively.

3. For independent groups of iid variables, say X_i and Y_i, *with a constant variance* σ^2 across groups

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

limits to a standard normal random variable as both n_x and n_y get large. Here

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$$

is the **pooled estimate** of the variance. The quantities $\bar{X}$, S_x, n_x are the sample mean, sample standard deviation and sample size for the X_i and $\bar{Y}$, S_y and n_y are defined analogously.

4. If the X_i and Y_i happen to be normal, then Z follows the Student's t distribution with $n_x + n_y - 2$ degrees of freedom. Therefore a $(1-\alpha)\times 100\%$ confidence interval for $\mu_y - \mu_x$ is

$$\bar{Y} - \bar{X} \pm t_{n_x+n_y-2,1-\alpha/2} S_p \left(\frac{1}{n_x} + \frac{1}{n_y}\right)^{1/2}$$

5. To **test** whether the **variances** of both groups are **equal**, we use that $\frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2}$ follows what is called the F distribution with $n_x - 1$ **numerator degrees of freedom** and $n_y - 1$ denominator degrees of freedom.

6. We test the hypothesis $H_0 : \sigma_x^2 = \sigma_y^2$ versus either of $H_1 : \sigma_x^2 < \sigma_y^2$, $H_2 : \sigma_x^2 \neq \sigma_y^2$ and $H_3 : \sigma_x^2 > \sigma_y^2$ compare the statistic $TS = S_1^2/S_2^2$ to the F distribution. We reject H_0 in favor of

   H_1 if $TS < f_{n_x-1,n_y-1,\alpha}$,

   H_2 if $TS < f_{n_x-1,n_y-1,\alpha/2}$ or $TS > f_{n_x-1,n_y-1,1-\alpha/2}$,

   H_3 if $TS > f_{n_x-1,n_y-1,1-\alpha}$.

7. The F distribution satisfies the property that $f_{n_x-1,n_y-1,\alpha} = (f_{n_y-1,n_x-1,1-\alpha})^{-1}$. So, if H_0 is true, then acceptance or rejection is unaffected, whether we put S_x^2 on the top or bottom.

8. Using the fact that

$$1 - \alpha = P\left(F_{n_x-1,n_y-1,\alpha/2} \leq \frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} \leq F_{n_x-1,n_y-1,1-\alpha/2}\right)$$

we can calculate a confidence interval for $\frac{\sigma_y^2}{\sigma_x^2}$ as $\left[F_{n_x-1,n_y-1,\alpha}\frac{S_x^2}{S_y^2}, F_{n_x-1,n_y-1,1-\alpha/2}\frac{S_x^2}{S_y^2}\right]$. Of course, the confidence interval for $\frac{\sigma_x^2}{\sigma_y^2}$ is $\left[F_{n_y-1,n_x-1,\alpha}\frac{S_y^2}{S_x^2}, F_{n_y-1,n_x-1,1-\alpha/2}\frac{S_y^2}{S_x^2}\right]$.

9. F tests heavily depend on the normality assumption.

10. If we conclude that the **variances of both groups are unequal**, then we use that the statistic

$$\frac{\bar{X}-\bar{Y}-(\mu_x-\mu_y)}{\sqrt{\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}}}$$

follows a standard normal distribution for large n_x and n_y. It follows an approximate Students T distribution if the X_i and Y_i are normally distributed. The degrees of freedom are given below.

11. For testing $H_0 : \mu_x = \mu_y$ in the event where there is evidence to suggest that $\sigma_x \neq \sigma_y$, the test statistic $TS = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}}}$ follows an approximate Student's T distribution under the null hypothesis when X_i and Y_i are normally distributed. The degrees of freedom are approximated with

$$\frac{(S_x^2/n_x + S_y^2/n_y)^2}{(S_x^2/n_x)^2/(n_x-1)+(S_y^2/n_y)^2/(n_y-1)}.$$

12. The power for a Z test of $H_0 : \mu_x = \mu_y$ versus $H_3 : \mu_x > \mu_y$ is given by

$$P\left(Z \geq Z_{1-\alpha} - \frac{\mu_x-\mu_y}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}}\right)$$

while for $H_1 : \mu_x < \mu_y$ it is

$$P\left(Z \leq -Z_{1-\alpha} - \frac{\mu_x-\mu_y}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}}\right).$$

13. Sample size calculation assuming $n_x = n_y = n$

$$n = \frac{(Z_{1-\alpha}+Z_{1-\beta})^2(\sigma_x^2+\sigma_y^2)}{(\mu_x-\mu_y)^2}.$$

14. Note that under unequal variances

$$\bar{Y}-\bar{X} \sim N\left(\mu_y-\mu_x, \frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}\right)$$

15. The statistic

$$\frac{\bar{Y} - \bar{X} - (\mu_y - \mu_x)}{\left(\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}\right)^{1/2}}$$

approximately follows Gosset's t distribution with degrees of freedom equal to

$$\frac{\left(S_x^2/n_x + S_y^2/n_y\right)^2}{\left(\frac{S_x^2}{n_x}\right)^2/(n_x-1) + \left(\frac{S_y^2}{n_y}\right)^2/(n_y-1)}$$

15 Comparing two binomials

1. Let $X \sim \text{Binomial}(n_1, p_1)$ and $\hat{p}_1 = X/n_1$
2. Let $Y \sim \text{Binomial}(n_2, p_2)$ and $\hat{p}_2 = Y/n_2$
3. We use the following notation

$n_{11} = X$	$n_{12} = n_1 - X$	n_1
$n_{21} = Y$	$n_{22} = n_2 - Y$	n_2
n_+	n_-	

4. We test $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$, $H_2 : p_1 > p_2$, $H_3 : p_1 < p_2$ with the statistic

$$TS = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

where $\hat{p} = \frac{X+Y}{n_1+n_2}$ is the estimate of the common proportion under the null hypothesis. This statistic is approximately normally distributed for large n_1 and n_2.

5. To estimate $p_1 - p_2$ we can use $\hat{p}_1 - \hat{p}_2$, which has an estimated standard error $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$, and construct a Wald confidence interval:

$$\hat{p}_1 - \hat{p}_2 \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

6. An easy fix to improve the performance of the Wald interval is to use $\tilde{p}_1 = (X+1)/(n_1+2)$ nad $\tilde{p}_2 = (Y+1)/(n_2+2)$ instead of $\hat{p}_1$ and $\hat{p}_2$.
7. The **relative risk** is defined as p_1/p_2 with estimate $\hat{p}_1/\hat{p}_2$.
8. The standard error for the *log relative risk* is

$$SE_{\log \hat{RR}} = \sqrt{\frac{1-p_1}{p_1 n_1} + \frac{1-p_2}{p_2 n_2}}$$

a. $\frac{\log \hat{RR} - \log RR}{\hat{SE}_{\log \hat{RR}}}$ is normally distributed for large n_1 and n_2

b. For hypothesis testing, use the null estimate of p

c. For intervals, use $\hat{p}_1$ and $\hat{p}_2$ in $\hat{SE}_{\log \hat{RR}}$. Exponentiate the interval to get one for the RR

9. The **odds ratio** is defined as $OR = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$

10. An estimate of the odds ratio is $\hat{OR} = \frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$

11. An estimated standard error for the odds ratio is $\hat{SE}_{\log \hat{OR}} = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$

12. For large sample sizes $\frac{\log \hat{OR} - \log OR}{\hat{SE}_{\log \hat{OR}}}$ follows a standard normal distribution. You can use this to get a Wald confidence interval and perform hypothesis test for the OR.

13. Exponentiate to get a CI for the odds ratio.

14. The odds ratio is invariant to transposing rows and columns

15. Taking logs for the RR and OR is done b/c it their finite sample distributions are often quite skewed and convergence to normality is faster on the log scale.

16 The delta method

1. The **delta method** is a useful tool for obtaining asymptotic standard errors.

2. The delta method states the following. If

$$\frac{\hat{\theta} - \theta}{\hat{SE}_{\hat{\theta}}} \to \mathrm{N}(0,1)$$

then

$$\frac{f(\hat{\theta}) - f(\theta)}{f'(\hat{\theta})\hat{SE}_{\hat{\theta}}} \to \mathrm{N}(0,1).$$

3. The delta method is motivated by noting that when $\hat{\theta}$ is close to θ then

$$\frac{f(\hat{\theta}) - f(\theta)}{\hat{\theta} - \theta} \approx f'(\hat{\theta})$$

so that

$$\frac{f(\hat{\theta}) - f(\theta)}{f'(\hat{\theta})\hat{SE}_{\hat{\theta}}} \approx \frac{\hat{\theta} - \theta}{\hat{SE}_{\hat{\theta}}}.$$

4. Therefore the asymptotic standard error for $f(\hat{\theta})$ is $f'(\hat{\theta})\hat{SE}_{\hat{\theta}}$.

17 Chi squared testing for contingency tables

1. Use the notation from Section 15.

2. The chi-squared statistic is written as

$$\sum \frac{(O-E)^2}{E}$$

the sum is taken over all four cells. The **expected** cell counts are calculated under the null hypothesis.

3. An easy computational form for this statistic is

$$\frac{n(n_{11}n_{22}-n_{12}n_{21})}{n_{1+}n_{2+}n_{+1}n_{+2}}.$$

4. We reject $H_0 : p_1 = p_2$ if the statistic is large. It is a two sided test. Compare to a $.95^th$ quantile of the Chi-squared distribution with 1 degree of freedom.

5. The chi-squared statistic is the square of the difference in proportions statistic with the common p in the denominator.

6. The chi-squared statistic is unchanged when transposing the rows and columns.

7. The chi-squared statistic also applies if the sampling is **multinomial** instead of binomial. That is if only the total sample size is fixed (and hence none of the margins).

8. In the multinomial case, the null hypothesis is that the row and column classifications are **independent**.

18 Fisher's exact test

1. Use the notation from Section 15.

2. Fisher's exact test is "exact" because it guarantees the α rate, regardless of the sample size

3. Under the null hypothesis, the distribution of $X \mid X+Y$ is the so called **hypergeometric** distribution. The PMF for the hypegeometric distribution is

$$P(X = x \mid X+Y = z) = \frac{\binom{n_1}{x}\binom{n_2}{z-x}}{\binom{n_1+n_2}{z}}.$$

The possibly values for x are $\max(0, z+n_1-n) \leq x \leq \min(z, n_1)$.

4. This distribution can be simulated by taking n_{1+} red balls and n_{2+} white balls and randomly allocating to two bins that can hold n_{+1} and n_{+2} balls respectively.

5. For a one sided hypothesis, you can perform Fisher's exact test by calculating the hypergeometric probabilities for all tables that are as or more supportive of the alternative hypothesis. Remember to constrain the margins. To obtain the two sided P-value, sum the probabilities for all tables with a probability less than or equal to that of the observed table.

6. Like the chi-squared test, Fisher's exact test applies to binomial, multinomial or Poisson sampling.

19 Chi-squared testing for binomial observations

1. The chi-squared test can be used to test $p_1 = p_2 = \ldots = p_k$ for k binomial observations, $X_i \sim \text{Binomial}(n_i, p_i)$.

2. The test statistic is $\sum \frac{(O-E)^2}{E}$ where O are the observed counts (successes and failures) and E are the estimated expected counts under the null hypothesis. This statistic is a chi-square with $k-1$ degrees of freedom.

3. A followup test would compare the proportions individually, two at a time.

4. The test can be generalized to multicategory settings where we would want to test whether or not the distribution of the counts in each row are the same. This test would have $(rows - 1)(cols - 1)$ degrees of freedom.

5. For multinomial sampling (only the overall sample size is constrained) a test of independence of the row and column classifications can be done. If n_{ij} are the observed counts in cell i, j, then the expected counts are $n_{i+}n_{+j}/n$. (Here n_{i+} refers to the i^{th} row total and n_{+j} refers to the j^{th} column total). The resulting statistic has degrees of freedom $(rows-1)(cols-1)$.

6. The test statistic for independence and the test for equal distributions in each row are mathematically the same and follow a chi-squared distribution with $(rows-1)(cols-1)$ degrees of freedom. The only difference is in the intepretation of the test.

7. Exact tests of independence (generalizations of Fisher's exact test) can be performed using Monte Carlo simulation.

8. Goodness of fit testing tests whether or not a series of counts follow a specified distribution. That is $H_0 : p_1 = p_{01}, p_2 = p_{02}, \ldots, p_k = p_{0k}$ where p_{0i} are specified. The expected count for cell i is $n * p_{0i}$. The resulting statistic has $k-1$ degrees of freedom.

20 Multiple comparisons

1. When conducting k hypothesis tests, the **familywise error rate** refers to the probability of falsely rejecting the null hypothesis in any of the k tests.

2. Bonferroni's inequality implies that the familywise error rate is no larger than $k\alpha$ where α is the Type I error rate (applied to each test individually). Therefore a **Bonferroni adjustment** uses the Type I error rate $\alpha^* = \alpha/k$ for each test. Under this adjustment the familywise error rate is no larger than α.

3. If there are a large number tests whos outcomes are independent (which is rarely the case), then the Bonferroni bound on the family wise error rate is nearly attained.

4. The **false discovery rate** is defined as the proportion of tests that are falsely declared significant.

5. The Benjamini and Hochberg procedure to control the FDR follows as

   i. Order your p-values so that $p_1 < \ldots < p_k$

   ii. Define $q_i = kp_i/i$

   iii. Define $F_i = \min(q_i, \ldots, q_k)$

   iv. Reject H_0 for all i so that F_i is less than the desired FDR. (Because the F_i are increasing, one need only find the largest i so that F_i <FDR).

21 Non-parametric testing

1. Non-parametric testing relaxes the assumptions of parametric tests. There are also referred to as "distribution free" tests. Note that these tests are not "assumption free".

2. For paired continuous data, consider taking the differences (as in the paired T-test); denote these differences by D_i. If the median difference is 0, then $p = P(D_i > 0) = .5$; if the median difference is greater than 0, then $P(D_i > 0) > .5$, and so on. The **sign test** tests $H_0 : p = .5$ versus the three alternative using the indicators of whether each D_i is larger than 0. Let D_+ be the total number of positive differences. Then D_+ is Binomial with success probability p. All of the usual binomial procedures can then be used to carry out the tests. Instances where $D_i = 0$ are thrown out and the overall sample size reduced.

3. The sign test disregards a lot of information contained in the observations. The **signed rank test** overcomes this to a large degree by also incorporating the **ranks** of the observations. The signed rank procedure is as follows

   a. Take the paired differences

   b. Take the absolute values of the differences

   c. Rank these absolute values, throwing out the 0s

   d. Multiply the ranks by the sign of the difference (+1 for a positive difference and -1 for a negative difference)

   e. Calcuate the rank sum W_+ of the positive ranks

4. For small sample sizes, the distribution of W_+ under the null hypothesis can be computed explicitly or by Monte-Carlo simulation. Critical values can also be obtained from tables. If the alternative is that the median difference is larger than 0, then W_+ should be large (hence reject if it is larger than the critical value). Vice-versa for the median difference being smaller than 0.

5. A large sample test statistic can be constructed as follows

$$E(W_+) = n(n+1)/4$$
$$Var(W_+) = n(n+1)(2n+1)/24$$
$$TS = \{W_+ - E(W_+)\}/Sd(W_+) \rightarrow \text{Normal}(0,1)$$

6. For unpaired data the relevant test is called the **rank sum** test.

7. Procedure

   (a) Discard the treatment labels

   (b) Rank the observations

   (c) Calculate the sum of the ranks in the first treatment

   (d) Either

      * calculate the asymptotic normal distrubtion of this statistic
      * compare with the exact distribution under the null hypothesis

8. Let W be the sum of the ranks for the first treatment (A)
   Let n_A and n_B be the sample sizes
   Then

   - $E(W) = n_A(n_a + n_b + 1)/2$
   - $\text{Var}(W) = n_A n_B(n_A + n_B + 1)/12$
   - $TS = \{W - E(W)\}/Sd(W) \rightarrow N(0,1)$

   This means, we can perform this test based on the Z-score TS.