Math4310/Biol6317 Midterm Review

October 15, 2011

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
$$\mu = E[X] = \sum_x x f(x).$$
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
$$P(X=k) = \left(\begin{array}{c} n \\ k \end{array}\right) p^k(1-p)^{n-k}.$$
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.

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, $\mathrm{Var}(S^2) \geq \mathrm{Var}((n-1)S^2/n)$.

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. Chi-squared confidence intervals depend heavily on the normality assumption.

4. 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.

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

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

7. 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.

8. 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.

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

10 Binomial confidence intervals

1. Binomial distributions are used to model proportions. If $X \sim \mathrm{Binomial}(n,p)$ 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

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

11 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 $\mathrm{N}(\mu, \sigma^2)$ 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.

12 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.

5. Therefore a $(1-\alpha) \times 100\%$ confidence interval for $\mu_y - \mu_x$ is

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

6. Note that under unequal variances

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

7. 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

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

13 Comparing two binomials

(a) Let $X \sim \text{Binomial}(n_1, p_1)$ and $\hat{p}_1 = X/n_1$

(b) Let $Y \sim \text{Binomial}(n_2, p_2)$ and $\hat{p}_2 = Y/n_2$

(c) 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:

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