1 On these review notes and the exam

1. You are responsible for the correctness of all of the formulae on this review sheet. (There are undoubtedly ytopgraphical errors :-).
2. You should know, *and understand*, everything in these review notes.
3. The exam format will be a series of multiple choice and short answer questions. Tedious calculations will be avoided.
4. You can bring a *non-fancy* (you know what I mean) scientific calculator. It must be able to take logs and raise numbers to exponents.
5. You can bring in one sheet of 8.5×11 paper filled, front and back, with formulae and notes.

2 Set theory

1. Notation - $\subset$ means "is a subset of", $\in$ 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 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 **compliment** 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 $(A \cap B)^c = A^c \cup B^c$.

3 Probability basics

1. A **probability measure**, say P, is a function on the collection of events to $[0, 1]$ so that:
   a. $P(\Omega) = 1$.
   b. If $A \subset \Omega$ then $P(A) \geq 0$.
   c. If $A_1, \ldots, A_n$ are disjoint then (**finite additivity**) $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.
2. $P(\bar{A}) = 1 - P(A)$.
3. The **odds** of an event, A, are $P(A)/(1 - P(A)) = P(A)/P(\bar{A})$.
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, are **mutually independent** if $P(\cap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i)$.
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. **Baye'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 $\bar{A}$ 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 $P(-|\bar{D})$.

13. Baye'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 **likelihood ratio** of a positive test result is $P(+|D)/P(+|\bar{D})$ = sensitivity/(1 − specificity). The likelihood ratio of a negative test result is $P(-|\bar{D})/P(-|D)$ = specificity/(1− sensitivity).

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,
$$\frac{P(D^c|-)}{P(D|-)} = \frac{P(-|D^c)}{P(-|D)}\frac{P(D^c)}{P(D)}.$$
This yields a method for evaluating the results of a diagnostic test without knowledge of the disease prevalence.

4 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.

5 Continuous random variables

1. **Continuous** random variables take values on a continuum.

2. The probability that a continuous random variable takes on any specific value is 0.

3. Probabilities associated with continuous random variables are governed by **probability density functions** (pdfs). Areas under probability density functions correspond to probabilities. For example, if f is a pdf corresponding to random variable X, then
$$P(a\leq X\leq b)=\int_a^b f(x)dx.$$
To be a pdf, a function must be positive and integrate to 1. That is, $\int_{-\infty}^{\infty} f(x)dx=1$

4. 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.

5. The expected value, or mean, of a continuous random variable, X, with pdf f, is
$$\mu = E[X] = \int_{-\infty}^{\infty} t f(t)dt.$$

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

7. 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.)

8. 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.

6 Properties of expected values and variances

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

1. Expected values commute across sums: $E[X+Y] = E[X] + E[Y]$.

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 commute across sums *for 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: $\mathrm{Var}(c+X) = \mathrm{Var}(X)$.

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.

8 Likelihood

1. Given a statistical probability mass function or density, say $f(x, \theta)$, where θ is an unknown parameter, the **likelihood** is f viewed as a function of θ for a fixed, observed value of x.

2. The likelihood has the following properties:

   a. Ratios of likelihood values measure the relative **evidence** of one value of the unknown parameter to another.

   b. Given a statistical model and observed data, all of the relevant information contained in the data regarding the unkown parameter is contained in the likelihood.

   c. If $\{X_i\}$ are independent events, then their likelihoods multiply. That is, the likelihood of the parameters given all of the X_i is simply the produce of the individual likelihoods.

3. A **likelihood plot** displays θ by $\mathcal{L}(\theta, x)$. Usually, it is divided by its maximum value so that its height is 1.

4. The value of θ where the curve reaches its maximum has a special meaning as the value of θ that is most well supported by the data. This point is called the **maximum likelihood estimate** (or MLE) of θ

$$MLE = \mathrm{argmax}_\theta \mathcal{L}(\theta, x).$$

5. Another interpretation of the MLE is that it is the value of θ that would make the data that we observed most probable.

6. For iid coin flips with a common success probability (of a head) p, the MLE of p is the sample proportion of heads.

7. A **likelihood ratio** is the value of the likelihood at one parameter over the value of the likelihood at another. Likelihood ratios larger than 1 support the parameter in the numerator more than the parameter in the denominator. (Likelihood ratio values less than one support the parameter in the denominator.)

8. We interpret a likelihood ratio

   of 8 as being weak evidence

   of 16 as being moderate evidence

   of 32 as being strong evidence

   of the hypothesized value of the parameter in the numerator over the one in the denominator.