Statistics 200 - Winter 2001
Stat 116 Review
Victoria Stodden

The handout largely follows Ross's *A First Course In Probability*, 1998.

1. **Counting**

   (a) Permutations

   How many different ordered arrangements of n items are there? $n!$

   Example: *abc*
   Possible permutations: *abc*, *acb*, *bac*, *bca*, *cab*, *cba*.
   There are 6 permutations: 3!

   (b) Combinations

   How many different groups of r objects can be formed from n objects?

   $\binom{n}{r} = \frac{n!}{(n-r)!r!}$

   Example: There are 5 women and 7 men. How many different groups of 2 women and 3 men can be formed?

   Solution: $\binom{5}{2}\binom{7}{3} = \frac{5 \cdot 4}{2} \cdot \frac{7 \cdot 6 \cdot 5}{3 \cdot 2} = 350$ possibilities.

   (c) Binominal Theorem
   $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$

   Please see Ross p. 8 for the proof.

2. **Probability**

   (a) The set of all possible outcomes of an experiment is called the *sample space*, usually denoted Ω. An *event* is a subset of Ω. For any event E, with equally likely outcomes, $P(E) = \frac{\#E}{\#\Omega}$.
   Since we are dealing with sets, we will need to know set operations.

For any sets E, F, and G:

$$E \cup F = F \cup E$$

$$(E \cup F) \cup G = E \cup (F \cup G)$$

$$(E \cup F)G = EG \cup FG$$

and DeMorgan's Laws:

$$(\cup_{i=1}^n E_i)^c = \cap_{i=1}^n E_i^c$$

$$(\cap_{i=1}^n E_i)^c = \cup_{i=1}^n E_i^c$$

Inclusion-Exclusion Formula:

$$\begin{aligned} P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i=1}^n P(E_i) - \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) + \cdots \\ +(-1)^{r+1} \sum_{i_1<i_2<\cdots<i_r} P(E_{i_1}E_{1_2}\cdots E_{i_r}) + \cdots \\ +(-1)^{n+1} P(E_1E_2\cdots E_n) \end{aligned}$$

Here, the summation $\sum_{i_1<i_2<\cdots<i_r} P(E_{i_1}E_{1_2}\cdots E_{i_r})$ is taken over all of the $\binom{n}{r}$ possible subsets of size r of the set $\{1, 2, \cdots, n\}$.

Example: Suppose we toss two fair coins. Here $\Omega = (H, H), (H, T), (T, H), (T, T)$. What is the probability that either the first coin or the second coin falls heads?

Solution: Let $E = \{(H, H), (H, T)\}$ and $F = \{(H, H), (T, H)\}$. Now we want $P(E \cup F)$, so:

$P(E \cup F) = P(E) + P(F) - P(EF)$ by the inclusion-exclusion formula
$= \frac{1}{2} + \frac{1}{2} - P(\{(H, H)\}) = 1 - \frac{1}{4} = \frac{3}{4}$.

This is also obvious by inspection: $P(E \cup F) = P(\{(H, H), (H, T), (T, H)\}) = \frac{3}{4}$.

3. **Conditional Probability**

(a) For $P(F) > 0, P(E|F) = \frac{P(EF)}{P(F)}$

Example Back to our flipping two coins setup. What is the probability that both flips are heads, given that the first flip is a head?

Solution: Now define $E = \{(H,H)\}$ and $F = \{(H,H),(H,T)\}$. Thus $P(E|F) = \frac{P(\{(H,H)\})}{P(\{(H,H),(H,T)\})} = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2}$.

(b) The Multiplication Rule:

$$P(E_1E_2\cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1\cdots E_{n-1})$$

(c) Bayes' Formula: Suppose E has occurred and now we are interested in which of F_j has also occured, where $F_1, F_2, \ldots F_n$ are mutually exclusive events with $\cup_{i=1}^n F_i = \Omega$ (thus one of the F_i must occur). Bayes' Formula tells us how to modify our opinions about the F_i held before the experiment E occurred, based on the evidence provided by the experiment.

$P(F_j|E) = \frac{P(EF_j)}{P(E)} = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^n P(E|F_i)P(F_i)}$.

Example:
A laboratory test is 95% effective in detecting a certain disease when the disease is present. The test also yields a false positive for 1% of the healthy people tested. If .5% of the population has the disease, what is the probability that a person actually has the disease, D, given a positive test result, E?

Solution:
$P(D|E) = \frac{P(DE)}{P(E)} = \frac{P(E|D)P(D)}{P(E|D)P(D)+P(E|D^c)P(D^c)} = \frac{(.95)(.005)}{(.95)(.005)+(.01)(.995)} = \frac{95}{294} \approx .323$.

(d) Two events E and F are independent if $P(EF) = P(E)P(F)$. Three events are independent if $P(EFG) = P(E)P(F)P(G)$, $P(EF) = P(E)P(F)$, $P(EG) = P(E)P(G)$ and $P(FG) = P(F)P(G)$.

(e) An *odds ratio* of an event E is defined by $\frac{P(E)}{P(E^c)} = \frac{P(E)}{1-P(E)}$.

4. **Random Variables**

(a) A *random variable* is a real-valued function defined on the sample space of an experiment. For example, the total number of heads that occurs in a series of coin flips may be of interest, rather than simply a particular sequence. Or, we may want to know that the sum of two rolls of a die is 7 rather than whether it was $(1,6)$ or $(2,5)$, etc. Random variables can be discrete, as in the two above example, or continuous. Examples of continuous random variables will follow.

(b) The *cumulative distribution function* or CDF of the random variable X is defined for all real numbers b, $-\infty < b < \infty$, by $F(b) = P\{X \leq b\}$.
The function has certain properties: F is non-decreasing; $\lim_{b\to\infty} F(b) = 1; \lim_{b\to-\infty} F(b) = 0$; and F is right continuous.

Note that $P\{a < X \leq b\} = F(b) - F(a)$ for discrete random variables.

For a discrete random variable we can define the *probability mass function* as $p(a) = P\{X = a\}$, where $F(a) = \sum_{all x \leq a} p(x)$.
For a continuous random variable we define the *probability density function* to be a non-negative function $f(x)$ such that $P\{X \in B\} = \int_B f(x)dx$. Note that $P\{A < X \leq b\} = \int_a^b f(x)dx$.

(c) Expected Value:

Discrete case: $E[X] = \sum_{x:p(x)>0} xp(x)$. Note also $E[g(X)] = \sum_i g(x_i)p(x_i)$ for any real valued function g.

Continuous case: $E[X] = \int_{-\infty}^{\infty} xf(x)dx$ and similarly $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$.

Expectation is *linear*, meaning that for a and b constants, $E[aX+b] = aE[X]+b$. Also, $E[X+Y] = E[X] + E[Y]$.

(d) Variance:

Discrete and continuous case: $Var[X] = E[(X-\mu)^2] = E[X^2] - (E[X])^2$.
The *standard deviation* of X is $\sqrt{Var(X)}$. Notice that $Var[aX+b] = a^2Var[X]$ and $Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y]$.

The *covariance* between two random variables is defined as $Cov[X,Y] = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$.
Examples follow.

5. **Moment Generating Functions**

These mathematical devices are useful for finding moments of the random variable X. Differentiating the moment generating function and evaluating the result at $t = 0$ produces the moments of X.

$M(t) = E[e^{tX}]$

Discrete case: $M(t) = \sum_x e^{tx}p(x)$
Continuous case: $M(t) = \int_{-\infty}^{\infty} e^{tx}f(x)dx$

6. **Discrete Distributions**

   (a) Bernoulli(p)

   pmf: $P(X = x|p) = p^x(1-p)^{1-x}; x = 0, 1; 0 \leq p \leq 1$
   mean: $E[X] = p$
   var: $Var[X] = p(1-p)$
   mgf: $M(t) = (1-p) + pe^t$

   (b) Binomial(n,p)

   pmf: $P(X = x|n, p) = \begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{n-x}; x = 0, 1, 2, \ldots, n; o \leq p \leq 1$
   mean: $E[X] = np$
   var: $Var[X] = np(1-p)$
   mgf: $M(t) = [pe^t + (1-p)]^n$

   (c) Poisson(λ)

   pmf: $P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}; x = 0, 1, \ldots; 0 \leq \lambda < \infty$
   mean: $E[X] = \lambda$
   var: $Var[X] = \lambda$
   mgf: $M(t) = e^{\lambda(e^t-1)}$

   (d) Geometric(p)

   pmf: $P(X = x|p) = p(1-p)^{x-1}; x = 1, 2, \ldots; 0 \leq p \leq 1;$
   mean: $E[X] = \frac{1}{p}$
   var: $Var[X] = \frac{1-p}{p}$
   mgf: $M(t) = \frac{pe^t}{1-(1-p)e^t}, t < log(1-p)$

(e) Negative Binomial(r,p)

pmf: $P(X = x|r,p) = \begin{pmatrix} r+x-1 \\ x \end{pmatrix} p^r(1-p)^x; x = 0, 1, \cdots; 0 \le p \le 1$
mean: $E[X] = \frac{r(1-p)}{p}$
Var: $Var[X] = \frac{r(1-p)}{p^2}$
mgf: $M(t) = (\frac{p}{(1-(1-p)e^t)})^r, t < -log(1-p)$

(f) Hypergeometric

pmf: $P(X = x|N,M,K) = \frac{\begin{pmatrix} M \\ x \end{pmatrix}\begin{pmatrix} N-M \\ K-x \end{pmatrix}}{\begin{pmatrix} N \\ K \end{pmatrix}}; x = 0, 1, 2, \ldots, K;$
$M - (N-K) \le x \le M; N, M, K \ge 0$
mean: $E[X] = \frac{KM}{N}$
var: $Var[X] = \frac{KM}{N}\frac{(N-M)(N-k)}{N(N-1)}$

7. **Continuous Distributions**

(a) Uniform(a,b)

pdf: $F(a|b) = \frac{1}{b-a}, a \le x \le b$
mean: $E[X] = \frac{b+a}{2}$
var: $Var[X] = \frac{(b-a)^2}{12}$
mgf: $M(t) = \frac{e^{bt}-e^{at}}{(b-a)t}$

(b) Normal(μ, σ^2)

pdf: $f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}, -\infty < x < \infty$
mean: $E[X] = \mu$
var: $Var[X] = \sigma^2$
mgf: $M(t) = e^{\mu t + \sigma^2 t^2/2}$

(c) Beta(a,b)

pdf: $f(x|a,b) = \frac{1}{B(a,b)}x^{a-1}(1-x)^{\beta-1}, 0 \le x \le 1, a > 0, b > 0$
mean: $E[X] = \frac{a}{a+b}$
Var: $Var[X] = \frac{ab}{(a+b)^2(a+b+1)}$
mgf: $M(t) = 1 + \sum_{k=1}^{\infty}(\prod_{r=0}^{k-1}\frac{a+r}{a+b+r})\frac{t^k}{k!}$

Here $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ and $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx$. Notice that if a is a natural number then $\Gamma(a) = (a-1)!$

(d) Exponential(β)

pdf: $f(x|\beta) = \frac{1}{\beta}e^{-x/\beta}, 0 \le x < \infty, \beta > 0$
mean: $E[X] = \beta$
var: $Var[X] = \beta^2$
mgf: $M(t) = \frac{1}{1-\beta t}, t < \frac{1}{\beta}$

(e) Gamma(α, β)

pdf: $f(X|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}, 0 \le x < \infty, \alpha, \beta > 0$
mean: $E[X] = \alpha\beta$
var: $Var[X] = \alpha\beta^2$
mgf: $M(t) = (\frac{1}{(1-\beta t)})^\alpha, t < \frac{1}{\beta}$

(f) Chi Squared

pdf: $f(x|p) = \frac{1}{\Gamma(p/2)2^{p/2}}x^{(p/2)-1}e^{-x/2}; 0 \le x < \infty; p-1,2,\ldots$
mean: $E[X] = p$
var: $Var[X] = 2p$
mgf: $M(t) = (\frac{1}{1-2t})^{p/2}, t\frac{1}{2}$

(g) F

pdf: $f(x|\nu_1,\nu_2) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})}(\frac{\nu_1}{\nu_2})^{\nu_1/2}\frac{x^{(\nu_1-2)/2}}{(1+(\frac{\nu_1}{\nu_2})x)^{(\nu_1+\nu_2)/2}}; 0 \le x < \infty;$
$\nu_1, \nu_2 = 1, \ldots$
mean: $E[X] = \frac{\nu_1}{\nu_2-2}, \nu_2 > 2$
var: $Var[X] = 2(\frac{\nu_2}{\nu_2-2})^2\frac{(\nu_1+\nu_2-2)}{\nu_1(\nu_2-4)}, \nu_2 > 4$

8. **Joint Distributions**

We define the *joint cumulative distribution function* of X and Y by $F(a,b) = P\{X \le a, Y \le b\}$ for $-\infty < a, b < \infty$.

Random variables X and Y are *independent* if and only if their joint probability density function can be expressed as $F_{X,Y}(x,y) = h(x)g(y)$ with $-\infty < x, y < \infty$.

Let $X_1, X_2, \ldots, X_n$ be n independent and identically distributed continuous random variables with density function f. Define $X_{(1)}$ to be the smallest value of $X_1, X_2, \ldots, X_n$; $X_{(2)}$ to be the second smallest value of $X_1, X_2, \ldots, X_n$; ... and $X_{(n)}$ to be the largest of $X_1, X_2, \ldots, X_n$. The ordered values $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ are called the *order statistics* corresponding to $X_1, X_2, \ldots, X_n$. The joint density of the order statistics is $f_{X_{(1)} \ldots X_{(n)}}(x_1, x_2, \ldots, x_n) = n! f(x_1) \cdots f(x_n)$ for $x_1 < x_2 < \cdots < x_n$. Please Ross p. 277 for the proof

9. **Expectation Formulas**

   (a) If X and Y have a joint probability density function $f(x, y)$ then $E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy$.

   (b) $E[X] = E[E[X|Y]]$. Please see Ross p 338 for proof and examples.

   (c) $Var(X) = E[Var(X|Y)] + Var(E[X|Y])$. Ross p. 348.

10. **Limit Theorems** (for proofs please see Ross)

    (a) Markov's Inequality: If X is a random variable that takes only nonnegative values, then for any value $a > 0, P\{X \geq a\} \leq \frac{E[X]}{a}$.

    (b) Chebyshev's Inequality: If X is a random variable with finite mean μ and variance σ^2, then for any value $K > 0, P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$.

    (c) Weak Law of Large Numbers: Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$. Then, for any $\epsilon > 0$,

    $$P\{|\frac{X_1 + \cdots + X_n}{n} - \mu| \geq\} \rightarrow 0$$

    as $n \rightarrow \infty$.

    (d) The Central Limit Theorem: Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables each having mean μ and variance σ^2. Then the distribution of

    $$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma \sqrt{n}}$$

    tends to the standard normal as $n \rightarrow \infty$.

(e) The Strong Law of Large Numbers: Let $X_1, X_2, \ldots$ be a sequnce of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$. Then, with probability 1,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu$$

as $n \to \infty$.