STAT 511 Summary Sheet

Karl B. Gregory

Throughout, let A and B be events, let X and Y be random variables, and let a and b be constants.

1. Set theory and beginning of probability theory
   - $P(\emptyset) = 0$.
   - $P(A) \leq 1$ for any A.
   - $P(A^c) = 1 - P(A)$.
   - $P(A \cap B^c) = P(A) - P(A \cap B)$.
   - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
   - If $A \subset B$ then $P(A) \leq P(B)$.
   - Bonferroni's inequality: $P(A \cap B) \geq P(A) + P(B) - 1$.
   - $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ for any partition $C_1, C_2 \dots$
   - Boole's inequality: $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{n} P(A_i)$ for any sets $A_1, A_2, \dots$
2. Counting rules
   - Number of ways to draw r things from N things
     (a) ordered, without replacement: $N!/(N-r)!$
     (b) unordered, without replacement: $\binom{N}{r} = N!/[(N-r)!r!]$
     (c) ordered, with replacement: N^r
     (d) unordered, with replacement: $\binom{N+r-1}{r}$
   - The number of ways to partition N things into K groups of sizes $n_1, \dots, n_K$, where $n_1 + \dots + n_K = N$ is $N!/(n_1! \cdot \dots \cdot n_K!)$.
3. Conditional probability and independence
   - $P(A|B) = P(A \cap B)/P(B)$.
   - Bayes' Rule: For an event B with $P(B) > 0$ and a partition $A_1, A_2, \dots$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)}.$$

- Simple version of Bayes' rule using A and A^c as the partition:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}.$$

- The events A and B are called independent if $P(A \cap B) = P(A)P(B)$.
- The following are equivalent:
  (a) $P(A \cap B) = P(A)P(B)$
  (b) $P(A|B) = P(A)$
  (c) $P(B|A) = P(B)$

4. Random variables and the cdf

- A random variable (rv) X is a function from the sample space to the real numbers.
- The cumulative distribution function (cdf) F_X of a rv X is the function

$$F_X(x) = P(X \leq x) \quad \text{for all} \quad x \in \mathbb{R}.$$

- A function $F_X(x)$ is a cdf if and only if the following hold:
  (a) $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.
  (b) $F_X(x)$ is non-decreasing in x.
  (c) $F_X(x)$ is right-continuous, i.e. for every x_0, $\lim_{x \downarrow x_0} = F_X(x_0)$.
- For discrete rvs the cdf is a step function; for continuous rvs the cdf is continuous.
- For a continuous rv X, $P(X = x) = 0$ for all $x \in \mathbb{R}$, i.e. no probability is assigned to points.
- For a continuous rv X

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq B) = F_X(b) - F_X(a).$$

- Two rvs X and Y with cdfs F_X and F_Y, respectively, are called identically distributed if $F_X(u) = F_Y(u)$ for all $u \in \mathbb{R}$.

5. Probability mass and density functions

- Discrete rvs have probability mass functions (pmfs) and continuous rvs have probability density functions (pdfs).
- The pmf p_X of a discrete rv X is the function $p_X(x) = P(X = x)$ for all $x \in \mathbb{R}$.
- The pdf f_X of a continuous rv X is the function f_X which satisfies

$$P_X(a < X < b) = \int_a^b f_X(x)dx, \quad \text{for all} \quad a < b \in \mathbb{R}.$$

- If the cdf F_X of a continuous rv X has a continuous first derivative, then $f_X(x) = \frac{d}{dx}F(x)$.
- If X is discrete with pmf p_X, the support of X, denoted $\mathcal{X}$, is the set of values for which the pmf is positive, that is $\mathcal{X} = \{x : p_X(x) > 0\}$.
- If X is continuous with pdf f_X, the support of X, denoted $\mathcal{X}$, is the set of values for which the pdf is positive, that is $\mathcal{X} = \{x : f_X(x) > 0\}$.
- The function p_X is a pmf if and only if the following hold:
  (a) $p_X(x) \geq 0$ for all x.
  (b) $\sum_{x\in\mathcal{X}} p_X(x) = 1$.
- The function f_X is a pdf if and only if the following hold:
  (a) $f_X(x) \geq 0$ for all x.
  (b) $\int_{\mathbb{R}} f_X(x) = 1$.
- We write $X \sim p_X$ when X is a rv with pmf p_X.
- We write $X \sim f_X$ when X is a rv with pdf f_X.
- We write $X \sim F_X$ when X is a rv with cdf F_X.

6. Expected value and variance of rvs

- The expected value $\mathbb{E}X$ of a rv X is

$$\mathbb{E}X = \begin{cases} \displaystyle\sum_{x\in\mathcal{X}} x \cdot p_X(x) & \text{if } X \sim p_X \\ \displaystyle\int_{\mathbb{R}} x \cdot f_X(x)dx & \text{if } X \sim f_X. \end{cases}$$

- Moreover, for any function $g : \mathbb{R} \to \mathbb{R}$, the expected value $\mathbb{E}g(X)$ of $g(X)$ is

$$\mathbb{E}g(X) = \begin{cases} \displaystyle\sum_{x\in\mathcal{X}} g(x) \cdot p_X(x) & \text{if } X \sim p_X \\ \displaystyle\int_{\mathbb{R}} g(x) \cdot f_X(x)dx & \text{if } X \sim f_X. \end{cases}$$

- The variance $\operatorname{Var} X$ of a rv X is defined as $\operatorname{Var} X = \mathbb{E}(X - \mathbb{E}X)^2$.
- Useful expression: $\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2$.
- $\mathbb{E}(aX + b) = a\mathbb{E}X + b$.
- $\operatorname{Var}(aX + b) = a^2 \operatorname{Var} X$.
- Chebychev's Inequality: For any rv X with mean μ_X and variance σ_X^2 and any constant K,

$$P_X(|X - \mu_X| < K\sigma_X) \geq 1 - \frac{1}{K^2}.$$

7. Suite of discrete rv probability distributions (see table at end).
8. Suite of continuous rv probability distributions (see table at end).
9. Quantiles
   - For any rv X with cdf F_X, the θ quantile of X is defined as $\inf\{x : F_X(x) \geq \theta\}$, for $\theta \in (0, 1)$. This is the definition we need to use when the cdf of X has jumps or flat parts, which is the case when X is discrete.
   - If X is a continuous rv with pdf f_X and a strictly increasing cdf F_X, then the θ quantile of X is the (unique) value q which satisfies $F_X(q) = \theta$.
10. Moments and moment generating functions
    - The kth moment about the origin is $\mathbb{E}X^k$.
    - The kth moment about the mean, also called the kth central moment, is $\mathbb{E}(X - \mathbb{E}X)^k$.
    - The moment generating function (mgf) M_X of a rv X is the function given by $M_X(t) = \mathbb{E}e^{tX}$, as long as this expected value is finite for all values of t in a neighborhood of zero.
    - The kth moment of X may be found by taking the kth-order derivative of $M_X(t)$ and evaluating the resulting function of t at $t = 0$. That is

$$\mathbb{E}X^k = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0}.$$

    - If M_X is the mgf of X and M_Y is the mgf of Y and $M_X(t) = M_Y(t)$ for all t in a neighborhood of zero, then X and Y are identically distributed.
    - The mgf of $aX + b$ is $M_{aX+b}(t) = e^{tb} M_X(at)$.
11. Joint and marginal distributions
    - For a pair of discrete rvs (X, Y), the joint pmf p is the function given by $p(x, y) = P(X = x, Y = y)$ for all $x, y \in \mathbb{R}$.
    - For a pair of continuous rvs (X, Y), the joint pdf f is the function satisfying

$$P((X, Y) \in A) = \iint_A f(x, y)dxdy \quad \text{for all } A \subset \mathbb{R}^2,$$

      where $\iint_A$ denotes integration over all $(x, y) \in A$.
    - We get the marginal pmf/pdf of X by summing/integrating the joint pmf/pdf of (X, Y) over Y:

$$p_X(x) = \sum_{y \in \mathcal{Y}} p(x, y)$$

$$f_X(x) = \int_{\mathbb{R}} f(x, y)dy$$

- The joint cdf F of a pair of rvs (X, Y) is the function given by

$$F(x, y) = P(X \le x, Y \le y) \quad \text{for all } (x, y) \in \mathbb{R}^2.$$

12. Conditional distributions and conditional expectation

- For any value $y \in \mathbb{R}$ for which $p_Y(y) > 0$, the conditional pmf of $X|Y = y$ is

$$p(x|y) = \frac{p(x, y)}{p_Y(y)} \quad \text{for all } x \in \mathbb{R}.$$

- Likewise, for any $y \in \mathbb{R}$ for which $f_Y(y) > 0$, the conditional pdf of $X|Y = y$ is

$$f(x|y) = \frac{f(x, y)}{f_Y(y)} \quad \text{for all } x \in \mathbb{R}.$$

- For any function $g : \mathbb{R} \to \mathbb{R}$ and any value y such that $p_Y(y) > 0$ or $f_Y(y) > 0$, the conditional expectation of $g(X)$ given that $Y = y$ is

$$\mathbb{E}[g(X)|Y = y] = \begin{cases} \displaystyle\sum_{x \in \mathcal{X}} g(x) \cdot p(x|y) & \text{if } X \text{ is discrete} \\ \displaystyle\int_{\mathbb{R}} g(x) \cdot f(x|y) dx & \text{if } X \text{ is continuous.} \end{cases}$$

- The conditional variance of X given that $Y = y$ is

$$\text{Var}[X|Y = y] = \mathbb{E}[(X - \mathbb{E}[X|Y = y])^2 | Y = y].$$

- We often choose not to specify a value for the variable on which we condition, leaving it to be random; that is we write $\mathbb{E}[X|Y]$ and $\text{Var}[X|Y]$ instead of $\mathbb{E}[X|Y = y]$ and $\text{Var}[X|Y = y]$ for the conditional mean and variance of X given Y. The quantities $\mathbb{E}[X|Y]$ and $\text{Var}[X|Y]$, which depend on the value Y takes, are themselves random variables.
- Useful expression: $\text{Var}[X|Y] = \mathbb{E}[X^2|Y] - (\mathbb{E}[X|Y])^2$.

13. Independence of random variables

- If (X, Y) is a discrete pair of rvs with joint pmf p and marginal pmfs p_X and p_Y, then X and Y are independent if and only if

$$p(x, y) = p_X(x) p_Y(y) \text{ for all } (x, y) \in \mathbb{R}^2.$$

- If (X,Y) is a continuous pair of rvs with joint pdf f and marginal pdfs f_X and f_Y, then X and Y are independent if and only if

$$f(x,y) = f_X(x)f_Y(y) \text{ for all } (x,y) \in \mathbb{R}^2.$$

- Shortcut independence check: Let (X,Y) be a pair of discrete or continuous rvs with joint pmf p or joint pdf f. Then X and Y are independent if and only if there exist functions g and h such that

$$p(x,y) = g(x)h(y) \text{ for all } (x,y) \in \mathbb{R}^2$$

or

$$f(x,y) = g(x)h(y) \text{ for all } (x,y) \in \mathbb{R}^2.$$

- If X and Y are independent rvs then for any functions $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbb{E}g(X)h(Y) = \mathbb{E}g(X)\mathbb{E}h(Y),$$

so that the expectation of a product is the product of expectations.

- If X and Y are independent rvs with mgfs M_X and M_Y, then the mgf of $X+Y$ is given by

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

14. Covariance and correlation and bivariate Normal distribution

- The covariance between two rvs X and Y with means μ_X and μ_Y is

$$\text{Cov}(X,Y) = \mathbb{E}(X-\mu_X)(Y-\mu_Y) =: \sigma_{XY}.$$

- The correlation between two rvs X and Y with variances σ_X^2 and σ_Y^2 and covariance σ_{XY} is

$$\text{corr}(X,Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} =: \rho_{XY} \in [-1,1].$$

- Useful expression: $\text{Cov}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$.
- If X and Y are independent then $\text{Cov}(X,Y) = 0$.
- If $\text{Cov}(X,Y) = 0$, X and Y are not necessarily independent.
- Exceptionally, if the pair (X,Y) has the bivariate Normal distribution and $\text{Cov}(X,Y) = 0$, then X and Y *are* independent.
- $\mathbb{E}(aX+bY) = a\mathbb{E}X + b\mathbb{E}Y$.
- $\text{Var}(aX+bY) = a^2 \text{Var}\, X + b^2 \text{Var}\, Y + 2ab\, \text{Cov}(X,Y)$.

- For random variables $X_1, \ldots, X_n$ and constants $a_1, \ldots, a_n$,

$$\mathrm{Var}(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 \,\mathrm{Var}\, X_i + \sum_{i \neq j} a_i a_j \,\mathrm{Cov}(X_i, X_j).$$

15. Hierarchical models

- $\mathbb{E}Y = \mathbb{E}(\mathbb{E}[Y|X])$
- $\mathrm{Var}\, Y = \mathbb{E}(\mathrm{Var}[Y|X]) + \mathrm{Var}(\mathbb{E}[Y|X])$
- Example: For $Y|X \sim \mathrm{Binom}(X, p)$ and $X \sim \mathrm{Poisson}(\lambda)$,

$$\begin{aligned} \mathbb{E}Y &= \mathbb{E}(Xp) = \lambda p \\ \mathrm{Var}\, Y &= \mathbb{E}(Xp(1-p)) + \mathrm{Var}(Xp) = \lambda p(1-p) + \lambda p^2 = \lambda p. \end{aligned}$$

Commonly encountered pmfs and pdfs along with their mgfs, expected values, and variances:

pmf/pdf	$\mathcal{X}$	$M_X(t)$	$\mathbb{E}X$	Var X
$p_X(x;p) = p^x(1-p)^{1-x}$,	$x = 0, 1$	$pe^t + (1-p)$	p	$p(1-p)$
$p_X(x;n,p) = \binom{n}{x}p^x(1-p)^{n-x}$,	$x = 0, 1, \ldots, n$	$[pe^t + (1-p)]^n$	np	$np(1-p)$
$p_X(x;p) = (1-p)^{x-1}p$,	$x = 1, 2, \ldots$	$\frac{pe^t}{1-(1-p)e^t}$	p^{-1}	$(1-p)p^{-2}$
$p_X(x;p,r) = \binom{x-1}{r-1}(1-p)^{x-r}p^r$,	$x = r, r+1, \ldots$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	rp^{-1}	$r(1-p)p^{-2}$
$p_X(x;\lambda) = e^{-\lambda}\lambda^x/x!$	$x = 0, 1, \ldots$	$e^{\lambda(e^t-1)}$	λ	λ
$p_X(x;N,M,K) = \binom{M}{x}\binom{N-M}{K-x}/\binom{N}{K}$	$x = 0, 1, \ldots, K$	¡complicadísimo!	$\frac{KM}{N}$	$\frac{KM}{N}\frac{(N-K)(N-M)}{N(N-1)}$
$p_X(x;K) = \frac{1}{K}$	$x = 1, \ldots, K$	$\frac{1}{K}\sum_{x=1}^{K} e^{tx}$	$\frac{K+1}{2}$	$\frac{(K+1)(K-1)}{12}$
$p_X(x;x_1,\ldots,x_n) = \frac{1}{n}$	$x = x_1, \ldots, x_n$	$\frac{1}{n}\sum_{i=1}^{n} e^{tx_i}$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$-\infty < x < \infty$	$e^{\mu t + \sigma^2 t^2/2}$	μ	σ^2
$f_X(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)$	$0 < x < \infty$	$(1-\beta t)^{-\alpha}$	$\alpha\beta$	$\alpha\beta^2$
$f_X(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$0 < x < 1$	$1 + \sum_{k=1}^{\infty}\frac{t^k}{k!}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

The "multinoulli" and multinomial pmfs and the bivariate Normal pdf:

$$p_{(X_1,\ldots,X_K)}(x_1,\ldots,x_K;p_1,\ldots,p_K) = p_1^{x_1}\cdots p_K^{x_K}\cdot\mathbf{1}\left\{(x_1,\ldots,x_K)\in\{0,1\}^K : \sum_{k=1}^K x_k = 1\right\}$$

$$p_{(Y_1,\ldots,Y_K)}(y_1,\ldots,y_K;n,p_1,\ldots,p_K) = \left(\frac{n!}{y_1!\cdots y_K!}\right)p_1^{y_1}\cdots p_K^{y_K}\cdot\mathbf{1}\left\{(y_1,\ldots,y_K)\in\{0,1,\ldots,n\}^K : \sum_{k=1}^K y_k = n\right\}$$

$$f_{(X,Y)}(x,y;\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2,\rho) = \frac{1}{2\pi}\frac{1}{\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right)$$