Probability review

This review sheet provides a summary of some of the important definitions and properties from probability which will be useful in computational statistics. It is by no means complete. For full details, see *Statistical Inference* (2nd ed.) by Casella & Berger, chapters 1, 2, and 4.

CDFs, density functions, and probability mass functions

- *Cumulative distribution function (cdf)*: Let X be a random variable. The cdf of X is defined by
$$F_X(x) = \mathbb{P}(X \leq x).$$
- X is a *continuous* random variable if $F_X(x)$ is a continuous function of x, and X is a *discrete* random variable if $F_X(x)$ is a step function of x.
- *Probability mass function (pmf)*: The pmf of a discrete random variable X is $f(x) = \mathbb{P}(X = x)$.
- *Probability density function (pdf)*: The pdf of a continuous random variable X is the function which satisfies
$$F_X(x) = \int\limits_{-\infty}^{x} f(x)dx.$$

Joint, marginal, and conditional distributions

Let X and Y be two random variables.

- *Joint cdf:* The joint cdf of X and Y is
$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$
- *Joint mass function:* If X and Y are discrete, their joint mass function is $f(x, y) = \mathbb{P}(X = x, Y = y)$.
- *Joint pdf*: If X and Y are continuous, their joint pdf is the function $f(x, y)$ such that for every set $A \subset \mathbb{R} \times \mathbb{R}$,
$$\mathbb{P}((X, Y) \in A) = \int\limits_{A}\int f(x, y)dxdy$$
- *Marginal distributions:* Given a joint pdf $f(x, y)$, the marginal pdf of X is given by
$$f_X(x) = \int\limits_{-\infty}^{\infty} f(x, y)dy$$
and the marginal pdf of Y is given by
$$f_Y(y) = \int\limits_{-\infty}^{\infty} f(x, y)dx$$
(For discrete random variables, the definitions are similar, just replace integrals with sums)

- *Conditional distributions:* Given a joint pdf or pmf $f(x, y)$, the conditional pdf/pmf of $X|Y = y$ is defined by

$$f(x|y) = \frac{f(x, y)}{f_Y(y)},$$

for any y such that $f_Y(y) > 0$.

Probability, expectation, and variance

- *Expectation*: The *expectation*, or *mean*, of a random variable X is

$$\mathbb{E}[X] = \begin{cases} \sum\limits_x xf(x) & X \text{ is discrete} \\ \int\limits_{-\infty}^{\infty} xf(x)dx & X \text{ is continuous} \end{cases}$$

- *Variance*: The *variance*, or second central moment, of a random variable X is

$$Var(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$$

- *Covariance*: If X and Y are two random variables, the *covariance* of X and Y is

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))].$$

- *Conditional expectation*: The conditional expectation of X given $Y = y$, denoted $\mathbb{E}[X|Y = y]$, is given by

$$\mathbb{E}[X|Y = y] = \begin{cases} \sum\limits_x xf(x|y) & X \text{ is discrete} \\ \int\limits_{-\infty}^{\infty} xf(x|y)dx & X \text{ is continuous} \end{cases}$$

- *Law of total probability*: Let A be an event and $B_1, ..., B_k$ be disjoint event which partition the space (i.e, $P(B_i \cap B_j) = 0$ if $i \neq j$, and $\sum_i P(B_i) = 1$). Then,

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

- *Law of total expectation* (aka law of iterated expectation):

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$$

(Note here that $\mathbb{E}[X|Y]$ is a random variable which is a function of Y). We can apply this rule to conditional expectations, too:

$$\mathbb{E}[X|Y_1] = \mathbb{E}[\mathbb{E}[X|Y_1, Y_2]|Y_1]$$

- *Law of total variance* (aka law of iterated variance):

$$Var(X) = \mathbb{E}[Var(X|Y)] + Var(\mathbb{E}[X|Y])$$

Functions of random variables

- *Law of the unconscious statistician*: Let X be a random variable with pdf or pmf $f(x)$ (depending on whether X is continuous or discrete). Let $g(X)$ be a function of X. Then

$$\mathbb{E}[g(X)] = \sum_x g(x)f(x) \qquad X \text{ is discrete}$$

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \qquad X \text{ is continuous}$$

- *Finding the distribution of a transformation*: Let X be a continuous random variable with pdf $f_X(x)$, and let $Y = g(X)$ be a function of X. To find the distribution of Y:

  1. For each y, find the set $A_y = \{x : g(x) \leq y\}$
  2. Find the cdf:
  $$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \int_{A_y} f_X(x)dx$$
  3. The pdf is $f_Y(y) = \frac{d}{dy}F_Y(y)$

  There is a special case when g is a monotone function. If X is continuous and g is monotone, then
  $$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.$$
  This special case can be extended if there exists a partition such that g is monotone on each piece of the partition (see Theorem 2.1.8 in Casella & Berger).

Moment generating functions

- *Moments:* Let X be a random variable. The nth *moment* of X is $\mathbb{E}[X^n]$.

- *Moment generating function (mgf)*: The mgf of X is $M_X(t) = \mathbb{E}[e^{tX}]$, provided that the expectation exists for t in some neighborhood of 0.

- *Properties of mgfs*:

  (a) If X has mgf $M_X(t)$, then $\mathbb{E}[X^n] = \frac{d^n}{dt^n} M_X(t) \Big|_{t=0}$

  (b) If X and Y are independent, with mgfs $M_X(t)$ and $M_Y(t)$, then the mgf of $X + Y$ is
  $$M_{X+Y}(t) = M_X(t)M_Y(t)$$

  (c) Let X and Y be random variables with cdfs F_X and F_Y. If $M_X(t) = M_Y(t)$ for all t in an open interval around 0, then $F_X(u) = F_Y(u)$ for all u.

  (d) Let $a, b \in \mathbb{R}$, and let $Y = a + bX$. The mgf of Y is
  $$M_Y(t) = e^{at}M_X(bt).$$

Statistics with matrix algebra

- *Definition of expectation and variance*: Let $X = (X_1, ..., X_k)^T \in \mathbb{R}^k$ be a random vector. Then
$$\mathbb{E}[X] = (\mathbb{E}[X_1], ..., \mathbb{E}[X_k])^T,$$
and
$$Var(X) = \Sigma$$
where $\Sigma \in \mathbb{R}^{k \times k}$ is the covariance matrix for X, with entries $\Sigma_{ij} = Cov(X_i, X_j)$. (This implies that the diagonal entries are $\Sigma_{ii} = Var(X_i)$).

- *Expectation and variance of linear combinations*: Let $X \in \mathbb{R}^k$ be a random vector, and let $\mathbf{a} \in \mathbb{R}^k$ and $\mathbf{B} \in \mathbb{R}^{m \times k}$. Then
$$\mathbb{E}[\mathbf{a} + \mathbf{B}X] = \mathbf{a} + \mathbf{B}\mathbb{E}[X]$$
$$Var(\mathbf{a} + \mathbf{B}X) = \mathbf{B}Var(X)\mathbf{B}^T$$

- *Matrix square roots*: If M is a positive semi-definite matrix, then $M^{\frac{1}{2}}$ is the unique positive semi-definite matrix such that $M = M^{\frac{1}{2}} M^{\frac{1}{2}}$. If $M = \text{diag}(m_1, ..., m_k)$, then $M^{\frac{1}{2}} = \text{diag}(\sqrt{m_1}, ..., \sqrt{m_k})$.

- *Block matrix inverses*: Let
$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$
be a block matrix with $A \in \mathbb{R}^{p \times p}$, $B \in \mathbb{R}^{p \times q}$, $C \in \mathbb{R}^{q \times p}$, and $D \in \mathbb{R}^{q \times q}$. Assuming that A and D are invertible, then
$$M^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}$$