STAT 801=830
Probability Basics

Richard Lockhart

Simon Fraser University

STAT 801=830 — Fall 2012

Purposes of These Notes

- Cover the material in Wasserman Chapter 1, 2 pp 1-30; you are responsible for reading.
- This chapter is mostly review.
- Define probability spaces, σ-field.
- Define cdf, pmf, density, etc.
- Discuss coverage in text.

Probability Definitions pp3-4

- **Probability Space** (or **Sample Space**): ordered triple $(\Omega, \mathcal{F}, P)$.
- Ingredients are a set Ω, $\mathcal{F}$ which is a family of **events** (subsets of Ω), and P the probability.
- Required properties on next slide.

Outcomes, Events, Probabilities pp3-6,13

- Ω is a set (possible outcomes); elements are ω called elementary outcomes.
- $\mathcal{F}$ is a family of subsets (**events**) of Ω with the property that $\mathcal{F}$ is a σ-field (or Borel field or σ-algebra):
  1. The empty set $\emptyset$ and Ω are members of $\mathcal{F}$.
  2. $A \in \mathcal{F}$ implies $A^c = \{\omega \in \Omega : \omega \notin A\} \in \mathcal{F}$
  3. $A_1, A_2, \cdots$ all in $\mathcal{F}$ implies $A = \cup_{i=1}^{\infty} A_i$.
- P a function, domain $\mathcal{F}$, range a subset of $[0, 1]$ satisfying:
  1. $P(\emptyset) = 0$ and $P(\Omega) = 1$.
  2. **Countable additivity**: $A_1, A_2, \cdots$ **pairwise disjoint** ($j \neq k$ $A_j \cap A_k = \emptyset$)

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Consequences of axioms pp 6-7

- Can compute probabilities by usual rules, including approximation.
- Closure under countable intersections:

$$A_i \in \mathcal{F} \text{ implies } \cap_i A_i \in \mathcal{F}$$

- "Continuity" of P:

$$A_1 \subset A_2 \subset \cdots \text{ all in } \mathcal{F} \text{ implies } P(\cup A_i) = \lim_{n\to\infty} P(A_n)$$

and

$$A_1 \supset A_2 \supset \cdots \text{ all in } \mathcal{F} \text{ implies } P(\cap A_i) = \lim_{n\to\infty} P(A_n)$$

Vector valued random variable pp 19-22($p = 1$)

- A random vector is a function $X : \Omega \mapsto R^p$ such that, writing $X = (X_1, \ldots, X_p)$,

$$P(X_1 \le x_1, \ldots, X_p \le x_p)$$

  is defined for any constants $(x_1, \ldots, x_p)$.

- Formally the notation

$$X_1 \le x_1, \ldots, X_p \le x_p$$

  is a subset of Ω or **event**:

$$\{\omega \in \Omega : X_1(\omega) \le x_1, \ldots, X_p(\omega) \le x_p\}$$

- Remember X is a function on Ω so X_1 is also a function on Ω.
- Dependence of rv on ω is hidden! Almost always see X not $X(\omega)$.

Borel sets Not in Text; cf pp13,43

- **Borel** σ-field in R^p: smallest σ-field in R^p containing every open ball.
- Intersection of all σ fields containing all open balls.
- Every common set is a Borel set, that is, in the Borel σ-field.
- An R^p valued **random variable** is a map $X : \Omega \mapsto R^p$ such that when A is Borel then $\{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}$.
- Fact: this is equivalent to

$$\{\omega \in \Omega : X_1(\omega) \le x_1, \ldots, X_p(\omega) \le x_p\} \in \mathcal{F}$$

for all $(x_1, \ldots, x_p) \in R^p$.

The Distribution of a random variable cf p 20

- Jargon and notation: we write $P(X \in A)$ for $P(\{\omega \in \Omega : X(\omega) \in A\})$
- We define the **distribution** of X to be the map

$$A \mapsto P(X \in A)$$

- This is a probability on the set R^p with the Borel σ-field rather than the original Ω and $\mathcal{F}$.
- Talk about normal, Gamma, Weibull, Binomial, etc distributions.
- This is why we rarely see ω.

Cumulative Distribution Functions pp20-22 $p = 1$

- The **Cumulative Distribution Function** (CDF) of X: function F_X on R^p defined by

$$F_X(x_1, \ldots, x_p) = P(X_1 \le x_1, \ldots, X_p \le x_p)$$

- Properties of F_X (usually just F) for $p = 1$:
  1. $0 \le F(x) \le 1$.
  2. $x > y \Rightarrow F(x) \ge F(y)$ (monotone non-decreasing).
  3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
  4. $\lim_{x \searrow y} F(x) = F(y)$ (right continuous).
  5. $\lim_{x \nearrow y} F(x) \equiv F(y-)$ exists.
  6. $F(x) - F(x-) = P(X = x)$.
  7. $F_X(t) = F_Y(t)$ for all t implies that X and Y have the same distribution, that is, $P(X \in A) = P(Y \in A)$ for any (Borel) set A.

Discrete Distributions pp 20-25

- Distribution of a random variable X is **discrete** (also call rv discrete) if there is a countable set $x_1, x_2, \cdots$ such that

$$P(X \in \{x_1, x_2 \cdots\}) = 1 = \sum_i P(X = x_i)$$

- Then **discrete density** or **probability mass function** of X is

$$f_X(x) = P(X = x)$$

- $\sum_x f(x) = 1.$

Absolutely Continuous Distributions pp 20-25

- Rv X is **absolutely continuous** if there is a function f such that for any (Borel) set A:
$$P(X \in A) = \int_A f(x)dx. \tag{1}$$
- This is a p dimensional integral in general. Equivalently (for $p = 1$)
$$F(x) = \int_{-\infty}^{x} f(y)\, dy$$
- Any function f satisfying (1) is a **density** of X.
- Unique (up to *null* sets).
- For almost all values of x F is differentiable at x and
$$F'(x) = f(x)\,.$$
- Text calls these *continuous* distributions.

Example distributions see pp26-30,39-40

- X is Uniform[0,1] means

$$F(x) = \begin{cases} 0 & x \le 0 \\ x & 0 < x < 1 \\ 1 & x \ge 1 \end{cases}$$

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ \text{undefined} & x \in \{0,1\} \\ 0 & \text{otherwise} \end{cases}$$

- X is exponential:

$$F(x) = \begin{cases} 1 - e^{-x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

$$f(x) = \begin{cases} e^{-x} & x > 0 \\ \text{undefined} & x = 0 \\ 0 & x < 0 \end{cases}$$