Review of Probability Theory I

January 22-24, 2008

1 Probability

A probability model has two essential pieces of its description.

- Ω, the **probability** or **sample space**, is the set of possible outcomes.
  - An **event** is a subset of the sample space
  $$A \subset \Omega.$$
  - a collection of **outcomes**, $\{\omega; \omega \in A\}$
- P, the **probability** assigns a number to each event.

Let Ω be a sample space $\{\omega_1, \ldots, \omega_n\}$ and for $A \subset \Omega$, let $|A|$ denote the number of elements in A. Then the probability associated with *equally likely events*

$$P(A) = \frac{|A|}{|\Omega|}.$$

reports the fraction of outcomes in Ω that are also in A.

Some facts are immediate:

1. $$0 \leq P(A) \leq 1 \tag{1}$$
2. If $A \cap B = \emptyset$, then
$$P(A \cup B) = P(A) + P(B). \tag{2}$$
3. $$P(\Omega) = 1. \tag{3}$$

From these facts, we can derive several others:

Exercise 1. *1. If $A_1, \ldots, A_k$ are* **pairwise disjoint** *or* **mutually exclusive***, ($A_i \cap A_j = \emptyset$ if $i \neq j$.) then*

$$P(A_1 \cup A_2 \cup \cdots \cup A_k) = P(A_1) + P(A_2) + \cdots + P(A_k).$$

2. *(inclusion-exlusion) For any two events A and B,*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

3. *If* $A \subset B$ *then* $P(A) \leq P(B)$.

4. *Letting* A^c *denote the* complement *of* A, *then* $P(A^c) = 1 - P(A)$.

So, now we know that the range of the function we call the probability is a subset of the interval [0,1].

In order to use calculus ideas, we extend equation (3) to a countable collection if $\{A_j; j \geq 1\}$ ofpairwise disjoint events and require that

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j). \tag{2'}$$

Any function P that accepts events as its domain and returns numbers for its range and satisfies (1), (2), and (3') is a called a probability.

Exercise 2 (Continuity property of probability)**.** *If the events* $B_1 \subset B_2 \subset \cdots$ *and* $B = \cup_j B_j$, *then*

$$P(B) = \lim_{j \to \infty} P(B_j). \tag{4a}$$

and if the events $C_1 \supset C_2 \subset \cdots$ *and* $C = \cap_j C_j$, *then*

$$P(C) = \lim_{j \to \infty} P(C_j). \tag{4b}$$

2 Independence and Conditional Probability

Definition 3. 1. *Two events A and B are called* **independent** *if*

$$P(A \cap B) = P(A) \cdot P(B) \tag{5}$$

The **conditional probability** *of B given A,*

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \tag{6}$$

Often this formula is used in the form $P(B \cap A) = P(B|A)P(A)$.
Thus, if B and A are independent $P(B|A) = P(B)$.

Exercise 4 (Bayes formula)**.**

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{7}$$

3 Random Variables and Distribution Functions

Definition 5. *A* **random variable** X *is a function whose domain is the probability space. The range space* S *of* X *is called the* **state space***.*

Definition 6. *If* $X : \Omega \to \mathbb{R}$*, then the* **distribution function** *is given by*

$$F_X(x) = P\{X \leq x\}. \qquad (8).$$

Theorem 7. *Any distribution function has the following properties.*

1. F_X *is nondecreasing.*
2. $\lim_{x\to\infty} F_X(x) = 1$, $\lim_{x\to-\infty} F_X(x) = 0$.
3. F_X *is right continuous.*
4. *Set* $F_X(x-) = \lim_{p\to x-} F_X(p)$*. Then* $F_X(x-) = P\{X < x\}$*.*
5. $P\{X = x\} = F(x) - F(x-)$*.*

The **multidimensional** or **joint distribution function** F for the random vector $(X_1, \ldots, X_n)$, defined by

$$F(x_1, \ldots, x_n) = P\{X_1 \leq x_1, \ldots, X_n \leq x_n\},$$

has properties analogous to the one dimensional distribution function.

Definition 8. *Let* $X : \Omega \to \mathbb{R}$ *be a random variable. Call* X

1. **discrete** *if there exists a countable set* D *so that* $P\{X \in D\} = 1$*,*
2. **continuous** *if the distribution function* F *has a derivative.*

For discrete random variable define the **mass function**,

$$p(x) = P\{X = x\} \qquad (8)$$

In this case,

$$F(x) = \sum_{s\in D, s\leq x} p(s).$$

Thus, the distribution is constant except for jumps.
The requirements for a mass function are that $p(x) \geq 0$ for all $x \in D$ and

$$1 = \sum_{s\in D} p(s).$$

Discrete random variables will be defined by giving their mass function.

Continuous random variable have distributions with a derivative, which we also denote by f and call the **density function**. in this case,

$$F(x) = \int_{-\infty}^{x} f(s)\, ds.$$

Thus, the requirements for a density are that $f(x) \geq 0$ for all $x \in \mathbb{R}$ and

$$1 = \int_{-\infty}^{\infty} f(s)\, ds.$$

Generally speaking, we shall use the density function to describe the distribution of a random variable.

Exercise 9. *If X and Y are independent integer valued random variables with mass functions p_X and p_Y, then $X+Y$ has mass function*

$$p_{X+Y}(s) = \sum_x p_X(x) p_Y(s-x).$$

Exercise 10. *If X and Y are real valued random variables with density functions f_X and f_Y, then $X+Y$ has density function*

$$f_{X+Y}(s) = \int_{-\infty}^{+\infty} f_X(x) f_Y(s-x)\ dx.$$

4 Examples of Discrete Random Variables

1. (Bernoulli) $Ber(p)$, $D = \{0,1\}$

   $$p(x) = p^x (1-p)^{1-x}.$$

2. (binomial) $Bin(n,p)$, $D = \{0,1,\ldots,n\}$

   $$p(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

   So $Ber(p)$ is $Bin(1,p)$.

3. (geometric) $Geo(p)$, $D = \mathbb{N}$

   $$p(x) = p(1-p)^x.$$

4. (hypergeometric) $Hyp(N,n,k)$, $D = \{\max\{0, n-N+k\}, \ldots, \min\{n,k\}\}$

   $$p(x) = \frac{\binom{n}{x}\binom{N-n}{k-x}}{\binom{N}{n}}.$$

   For a hypergeometric random variable, consider an urn with N balls, k green. Choose n and let X be the number of green under equally likely outcomes for choosing each subset of size n.

5. (negative binomial) $Negbin(n,p)$, $D = \mathbb{N}$

   $$p(x) = \binom{n+x-1}{x} p^n (1-p)^x.$$

   Note that $Geo(p)$ is $Negbin(1,p)$.

6. (Poisson) $Pois(\lambda)$, $D = \mathbb{N}$,

   $$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}.$$

7. (uniform) $U(a,b)$, $D = \{a, a+1, \dots, b\}$,

$$p(x) = \frac{1}{b-a+1}.$$

Exercise 11. *Check that* $\sum_{x \in D} f(x) = 1$ *in the examples above.*

5 Examples of Continuous Random Variables

1. (beta) $Beta(\alpha, \beta)$ on $[0,1]$,
$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}.$$

2. (Cauchy) $Cau(\mu, \sigma^2)$ on $(-\infty, \infty)$,
$$f(x) = \frac{1}{\sigma\pi} \frac{1}{1+(x-\mu)^2/\sigma^2}.$$

3. (chi-squared) χ^2_a on $[0, \infty)$
$$f(x) = \frac{x^{a/2-1}}{2^{a/2}\Gamma(a/2)} e^{-x/2}.$$

4. (exponential) $Exp(\theta)$ on $[0, \infty)$,
$$f(x) = \theta e^{-\theta x}.$$

5. (Fisher's F) $F_{q,a}$ on $[0, \infty)$,
$$f(x) = \frac{\Gamma((q+a)/2) q^{q/2} a^{a/2}}{\Gamma(q/2)\Gamma(a/2)} x^{q/2-1}(a+qx)^{-(q+a)/2}.$$

6. (gamma) $\Gamma(\alpha, \beta)$ on $[0, \infty)$,
$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}.$$
Observe that $Exp(\theta)$ is $\Gamma(1, \theta)$.

7. (inverse gamma) $\Gamma^{-1}(\alpha, \beta)$ on $[0, \infty)$,
$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}.$$

8. (Laplace) $Lap(\mu, \sigma)$ on $\mathbb{R}$,
$$f(x) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}.$$

9. (normal) $N(\mu, \sigma^2)$ on $\mathbb{R}$,
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

10. (Pareto) $Par(\alpha, c)$ on $[c, \infty)$,
$$f(x) = \frac{c^\alpha \alpha}{x^{\alpha+1}}.$$

11. (Student's t) $t_a(\mu, \sigma^2)$ on $\mathbb{R}$,
$$f(x) = \frac{\Gamma((a+1)/2)}{\sqrt{\alpha\pi}\Gamma(\alpha/2)\sigma}\left(1 + \frac{(x-\mu)^2}{a\sigma^2}\right)^{-(a+1)/2}.$$

12. (uniform) $U(a, b)$ on $[a, b]$,
$$f(x) = \frac{1}{b-a}.$$

Example 12 (probability transform)**.** *Let the distribution function F for X be continuous and strictly increasing, then $F(X)$ is a $U(0,1)$ random variable.*

If $0 \le u \le 1$, then
$$P\{F(X) \le u\} = P\{X \le F^{-1}(u)\} = F(F^{-1}(u)) = u.$$

Exercise 13. *Let X be a continuous random variable with density function f_X. Let $g : \mathbb{R} \to \mathbb{R}$ be a monotonic differentiable function, then $Y = g(X)$ has density*
$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|.$$