WOMP 2008: Probability

Robert Masson. Edited to his liking by Shawn

November 16, 2008

1 Probability Spaces

Definition 1.1. A probability space is a measure space $(\Omega, \mathcal{F}, \mathbf{P})$ with total measure one. More precisely,

- Ω is a set (called the sample space) and elements in Ω are called outcomes.
- $\mathcal{F}$ is a σ-algebra on Ω; elements in $\mathcal{F}$ are called events.
- $\mathbf{P}$ is a measure on $\mathcal{F}$ such that $\mathbf{P}\{\Omega\} = 1$.

Example 1.2. When Ω is countable, we say that $(\Omega, \mathcal{F}, \mathbf{P})$ is a discrete probability space. One usually lets $\mathcal{F}$ consist of all subsets of Ω. In this case, we can assign a measure $\mathbf{P}\{\omega\}$ to each individual outcome $\omega \in \Omega$, such that

$$\sum_{\omega \in \Omega} \mathbf{P}\{\omega\} = 1.$$

Then, for each $E \in \mathcal{F}$,

$$\mathbf{P}\{E\} = \sum_{\omega \in E} \mathbf{P}\{\omega\}.$$

Example 1.3. An example of a discrete probability space is the space $\Omega_n = \{0,1\}^n$ consisting of n-tuples of 0's and 1's. One can think of these as the outcomes of flipping a coin n times with 0 denoting tails and 1 denoting heads. Let $p \in (0,1)$ denote the probability of getting heads, and for $\omega = (\omega_1, \ldots, \omega_n) \in \Omega_n$, let

$$\alpha(\omega) = \#\{\omega_k : \omega_k = 1\}.$$

Then for each $\omega \in \Omega_n$,

$$\mathbf{P}\{\omega\} = p^{\alpha(\omega)}(1-p)^{n-\alpha(\omega)}.$$

Thus, if $E_k = \{\omega : \alpha(\omega) = k\}$ is the event that one gets exactly k heads, then

$$\mathbf{P}\{E_k\} = \sum_{\omega : \alpha(\omega) = k} \mathbf{P}\{\omega\} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Remark One can define a probability space on the set of infinite coin tosses, but one has to be a bit more careful. Note that the set of infinite coin tosses is not a discrete probability space.

Example 1.4. Another canonical example of a probability space is $([0,1], \mathcal{L}[0,1], m)$: Lebesgue measure on $[0,1]$. It is often useful to think of this space when thinking of examples, or trying to come up with counterexamples.

2 Random Variables and Distributions

A random variable $X : \Omega \to \mathbb{R}$ is just a real measurable function on $(\Omega, \mathcal{F}, \mathbf{P})$. Note that X induces a measure μ_X on $\mathcal{B}$, the Borel subsets of $\mathbb{R}$, as follows. Given $B \in \mathcal{B}$, we define

$$\mu_X(B) = \mathbf{P}\left\{X^{-1}(B)\right\}.$$

μ_X is a probability measure and therefore $(\mathbb{R}, \mathcal{B}, \mu_X)$ is a probability space. μ_X is called the distribution of X.

If μ_X is supported on a countable set, then X is called a discrete random variable. If μ_X gives zero measure to every singleton set, then X is called a continuous random variable. Associated to *some* continuous random variables X is a density function f with the property that for any Borel set E,

$$\mu_X(E) = \int_E f(t)\, dt.$$

Example 2.1. If E is any subset of Ω, the indicator function of E is

$$\mathbf{1}_E(\omega) = \begin{cases} 1 & \text{if } \omega \in E; \\ 0 & \text{if } \omega \notin E. \end{cases}$$

Note that $\mathbf{1}_E$ is a Bernoulli random variable with parameter $\mathbf{P}\{E\}$ (see section 4.1 for the definition).

Example 2.2. Let Ω_n be as in example 1.3 and let $\omega = (\omega_1, \ldots, \omega_n) \in \Omega_n$. Let

$$X(\omega) = \#\{\omega_k : \omega_k = 1\}.$$

Then as we saw in that example, for $k = 0, \ldots, n$,

$$\mathbf{P}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}.$$

X is a Binomial random variable with parameters n and p (see section 4.1).

Definition 2.3. If X is an integrable random variable, the expectation of X is

$$\mathbf{E}\{X\} = \int X\, d\mathbf{P}.$$

The variance of X is defined to be

$$\mathrm{Var}(X) = \mathbf{E}\left\{(X - \mathbf{E}\{X\})^2\right\} = \mathbf{E}\left\{X^2\right\} - \mathbf{E}\{X\}^2,$$

as long as X^2 is integrable.

By the additivity of the integral, it follows that if $X_1, \ldots, X_n$ are any random variables,

$$\mathbf{E}\left\{\sum_{k=1}^n X_k\right\} = \sum_{k=1}^n \mathbf{E}\{X_k\}.$$

Theorem 2.4. *Suppose X is a random variable with distribution μ_X and that $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable. Then*

$$\int g(X)\, d\mathbf{P} = \int g(x)\, d\mu_X(x).$$

In particular,

$$\mathbf{E}\left\{X\right\} = \int x\, d\mu_X(x).$$

Remark The previous theorem shows that the distribution of a random variable X is usually all one needs to know about X. In particular, one usually does not care about the underlying probability space $(\Omega, \mathcal{F}, \mathbf{P})$. In fact, one usually makes statements like "let $X_1, X_2, \ldots$ be independent standard normal random variables" with the understanding that there exists a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ that admits such random variables, but without caring what $(\Omega, \mathcal{F}, \mathbf{P})$ is.

3 Independence

Definition 3.1. Fix a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Two events E_1 and E_2 are independent if

$$\mathbf{P}\left\{E_1 \cap E_2\right\} = \mathbf{P}\left\{E_1\right\} \mathbf{P}\left\{E_2\right\}.$$

More generally, a collection $\{E_\alpha\}$ of events is independent if for any finite sub-collection $E_{\alpha_1}, \ldots, E_{\alpha_n}$,

$$\mathbf{P}\left\{\bigcap_{k=1}^{n} E_{\alpha_k}\right\} = \prod_{k=1}^{n} \mathbf{P}\left\{E_{\alpha_k}\right\}.$$

Exercise 3.2. A collection $\{E_\alpha\}$ of events is called pairwise independent if for all pairs E_{α_1} and E_{α_2},

$$\mathbf{P}\left\{E_{\alpha_1} \cap E_{\alpha_2}\right\} = \mathbf{P}\left\{E_{\alpha_1}\right\} \mathbf{P}\left\{E_{\alpha_2}\right\}.$$

Show that pairwise independence does not necessarily imply independence.

Definition 3.3. Random variables $X_1, \ldots, X_n$ are said to be independent if for any collection of Borel sets $B_1, \ldots, B_n$,

$$\mathbf{P}\left\{\bigcap_{k=1}^{n} X_k^{-1}(B_k)\right\} = \prod_{k=1}^{n} \mathbf{P}\left\{X_k^{-1}(B_k)\right\}.$$

An infinite collection of random variables is said to be independent if every finite subcollection is independent. Note that if $X_1, \ldots, X_n$ are independent and $g_1, \ldots, g_n$ are any Borel measurable functions, then $g_1(X_1), \ldots, g_n(X_n)$ are also independent. This follows immediately from the definition.

Theorem 3.4. *1. If $X_1, \ldots, X_n$ are independent, integrable random variables then*

$$\mathbf{E}\left\{\prod_{k=1}^{n} X_k\right\} = \prod_{k=1}^{n} \mathbf{E}\left\{X_k\right\}.$$

2. If $X_1, \ldots, X_n$ are independent random variables whose variance is finite then

$$\mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n).$$

Exercise 3.5. Show that the converse to (1) in the above theorem is false. In other words, give an example of two random variables X and Y such that $\mathbf{E}\{XY\} = \mathbf{E}\{X\}\mathbf{E}\{Y\}$ but such that X and Y are not independent.

4 Some common distributions

4.1 Discrete distributions

Example 4.1 (Bernoulli distribution). If $p \in (0,1)$, a random variable X is said to be Bernoulli with parameter p if $\mathbf{P}\{X = 1\} = p$ and $\mathbf{P}\{X = 0\} = 1 - p$. This represents the outcome of a trial with probability p of success (the event that $X = 1$) and probability $1 - p$ of failure (when $X = 0$). It is fairly immediate that $\mathbf{E}\{X\} = p$ and that

$$\mathrm{Var}(X) = \mathbf{E}\left\{X^2\right\} - \mathbf{E}\{X\}^2 = p - p^2 = p(1-p).$$

Example 4.2 (Binomial distribution). A random variable X is said to be binomial with parameters n and p, if

$$\mathbf{P}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, \ldots, n.$$

Note that by exercise 1.3 a Binomial random variable can be thought of as the number of successes in n independent Bernoulli trials with probability p. Thus, the sum of n independent Bernoulli random variables $X_1, \ldots, X_n$ with parameter p is a binomial random variable with parameters n and p. Therefore,

$$\mathbf{E}\{X\} = \mathbf{E}\{X_1 + \ldots + X_n\} = \mathbf{E}\{X_1\} + \ldots + \mathbf{E}\{X_n\} = np,$$

and

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n) = np(1-p).$$

Example 4.3 (Poisson distribution). A random variable X has a Poisson distribution with parameter $\lambda > 0$ if

$$\mathbf{P}\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \ldots.$$

The Poisson distribution is the limit of binomial random variables with parameters n and λ/n as $n \to \infty$. One can use this to show that $\mathbf{E}\{X\} = \mathrm{Var}(X) = \lambda$.

Example 4.4 (Geometric distribution). A random variable X has a geometric distribution with parameter p if

$$\mathbf{P}\{X = k\} = (1-p)^{k-1} p, \qquad k = 1, 2, \ldots.$$

X can be interpreted as the number of Bernoulli trials needed until the first success occurs. One can check that $\mathbf{E}\{X\} = 1/p$ and $\mathrm{Var}(X) = (1-p)/p^2$.

4.2 Continuous distributions

Example 4.5 (Normal distribution)**.** A random variable X has a normal (or Gaussian) distribution with parameters μ and σ^2 if it has density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad -\infty < x < \infty.$$

One often writes $X \sim N(\mu, \sigma^2)$. When $\mu = 0$ and $\sigma^2 = 1$ we say that X is a standard normal. One can check that $\mathbf{E}\{X\} = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

Example 4.6 (Exponential distribution)**.** A random variable X has an exponential distribution with parameter $\lambda > 0$ if it has density

$$f(x) = \lambda e^{-\lambda x}, \qquad x > 0.$$

A simple calculation shows that $\mathbf{E}\{X\} = \lambda^{-1}$ and that $\mathrm{Var}(X) = \lambda^{-2}$.

Definition 4.7. The *characteristic function* of X is the function $\phi : \mathbb{R} \to \mathbb{C}$ defined by

$$\phi(t) = \mathbb{E}(e^{itX}).$$

Theorem 4.8. *Random variables X and Y have the same characteristic function if and only if they have the same distribution function.*

References

[1] G. Grimmett and D. Stirzaker (1982). *Probability and Random Processes*, Oxford University Press.

[2] Greg Lawler's Probability Notes. Available on the course webpage: http://www.math.uchicago.edu/%7Elawler/m312f07.html.

[3] L. C. G. Rogers and D. Williams (1979). *Diffusions, Markov Processes and Martingales*, Cambridge University Press.

[4] S Varadhan (2001). *Probability Theory*, Courant Lecture Notes in Mathematics #7.

[5] D Williams (1991). *Probability with Martingales*, Cambridge University Press.