A speedy review of probability

Moty Katzman

November 1, 2012

Probability spaces

A probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is a triple,
where
(1) Ω is a set,
(2) $\mathcal{F}$ is a collection of subsets of Ω which contains Ω, and is closed under complements, intersections and countable unions, and
(3) $\mathrm{P} : \mathcal{F} \to [0, 1]$ satisfies $\mathrm{P}(\emptyset) = 0$, $\mathrm{P}(\Omega) = 1$, and, if $\{A_i\}_{i=1}^{\infty}$ is a sequence of disjoint sets in $\mathcal{F}$,

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

Example: tossing independent fair coins

We toss 2 fair coins:
we can describe all 4 possible outcomes with length-two sequences of H's and T's. So here we might want to define

$$\Omega = \{HH, HT, TH, TT\}$$

now we can take $\mathcal{F}$ to be the set of all possible subsets of Ω (there are $2^4 = 16$ of those)
and we since the coins are fair and the tosses independent, we probably want to define $\mathsf{P} : \mathcal{F} \to [0, 1]$ by declaring

$$\mathsf{P}(A) = \frac{\text{number of elements in } A}{4}.$$

Example: a uniform distribution

Suppose we want to describe a uniform distribution, e.g., we want to pick a random number in the interval $[0, 1]$ so that for $0 \leq a \leq b \leq 1$ the probability of picking a number in the interval $[a, b]$ is given by $b - a$.
Here $\Omega = [0, 1]$;
one can prove that there exists a smallest collection $\mathcal{F}$ of subsets of Ω containing all intervals of the form $[a, b]$ and satisfying condition (2): we take this to be our $\mathcal{F}$;
it can also be shown that the exists a function $\mathrm{P} : \mathcal{F} \to [0, 1]$ with the property that $\mathrm{P}([a, b]) = b - a$.

Random variables

Let $\mathcal{P} = (\Omega, \mathcal{F}, \mathrm{P})$ be a probability space.
Definition: A random variable on $\mathcal{P}$
is a *function*
$X : \Omega \to \mathbb{R}$ with the property that, for all $a \in \mathbb{R}$,

$$X^{-1}\left((-\infty, a)\right) \in \mathcal{F}.$$

Notice that random variables are not random and are not variables, they are functions!

Back to first example: (tossing 2 coins)

Define $X : \Omega \to \mathbb{R}$ by declaring

$$X(\omega) = \text{number of H's in } \omega.$$

Now the probability of obtaining exactly one H is

$$\mathsf{P}\, X^{-1}(1) = \mathsf{P}(HT, TH) = 2/4 = 1/2.$$

Back to second example: (uniform distribution)

Define $X : \Omega \to \mathbb{R}$ by declaring

$$X(\omega) = \omega^2 + 1.$$

Discrete probability spaces

Definition: A probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is *discrete* if Ω is either finite or countable, i.e., if $\Omega = \{\omega_1, \dots, \omega_n\}$ or $\Omega = \{\omega_1, \dots, \omega_n, \dots\}$.
When $(\Omega, \mathcal{F}, \mathrm{P})$ is discrete $\mathcal{F}$ normally consists of all subsets of Ω, i.e., we can assign a probability to any event $F \subset \Omega$.
This assignment is usually not possible when $(\Omega, \mathcal{F}, \mathrm{P})$ is not discrete.
Example 1 is an example of a discrete probability space, Example 2 is not.

Independent random variables

Definition: Random variables $X_1, \ldots, X_n$ are independent if, for all $a_1, \ldots, a_n$,

$$\mathrm{P}\left(X_1 < a_1, \ldots, X_n < a_n\right) = \prod_{i=1}^{n} \mathrm{P}\left(X_i < a_i\right).$$

Expected values of discrete random variables

Definition: The *expected value* of a random variable X defined on a discrete probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is

$$E(X) = \sum_{\omega \in \Omega} X(\omega)\, \mathrm{P}(\omega)$$

provided that this series converges absolutely (i.e., in any order).

Example: a fair dice

We can model a fair dice as follows:
let $\Omega = \{1, 2, 3, 4, 5, 6\}$,
take $\mathcal{F}$ to be the set of all subsets of Ω (there are $2^6 = 64$ of these)
and define P to be the function which assigns to every subset E of Ω the number

$$\frac{\text{number of elements in } E}{6}.$$

Define the random variables $X(\omega) = \omega$, $Y(\omega) = \omega^2$.
We have $E(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6}$,
$E(Y) = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{91}{6}$.

Exercise:

Let X and Y be random variables defined on the same discrete probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and let $\lambda, \mu \in \mathbb{R}$.
Show that, if $E(X)$ and $E(Y)$ exist then so does $E(\lambda X + \mu Y)$ and

$$E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y).$$

Density and distribution functions

Definition: A function $f : \mathbb{R} \to [0, \infty)$ is a density function for the random variable X if, for all $a \in \mathbb{R}$,

$$\mathrm{P}(X < a) = \int_{-\infty}^{a} f(t)dt.$$

Any continuous $f : \mathbb{R} \to [0, \infty)$ with the property that $\displaystyle\int_{-\infty}^{\infty} f(t)dt = 1$ is the density function of a random variable. (Why?)

Definition: The distribution function of a random variable X with density function $f(x)$ is the function

$$F(x) = \int_{-\infty}^{x} f(t)dt = \mathrm{P}(X < x).$$

The *expected value* of a random variable X with density function f is

$$E(X) = \int_{-\infty}^{\infty} tf(t)dt$$

provided that this integral exists.

Fact: Let X be a random variable with density function f and let $G : \mathbb{R} \to \mathbb{R}$ be any function.

Then $G(X)$ is a new random variable and

$$E(G(X)) = \int_{-\infty}^{\infty} G(t)f(t)dt.$$

provided that this integral exists.

Variance, covariance and correlation

Definition: The <u>variance</u> of the random variable X is

$$\mathrm{Var}(X) = \mathrm{E}((X - \mathrm{E}(X))^2) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$$

whenever this expression exists.
The *standard deviation* of X is defined as $\sqrt{\mathrm{Var}(X)}$.
Exercise: If c is any number, $\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X)$.
Exercise: Let X be a random variable. Show that the random variable

$$\frac{X - E(X)}{\sqrt{\mathrm{Var}(X)}}$$

has expected value 0 and variance 1.

For random variables X and Y we define the *covariance of X and Y*

$$\mathrm{Covar}(X,Y) = \mathrm{E}\Big(\big(X - \mathrm{E}(X)\big)\big(Y - \mathrm{E}(Y)\big)\Big).$$

Exercise: Show that

$$\mathrm{Covar}(X,Y) = E(XY) - E(X)E(Y)$$

and

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + 2\,\mathrm{Covar}(X,Y) + \mathrm{Var}(Y).$$

Exercise: Show that $\mathrm{Covar}(-,-)$ is a bilinear function, i.e., for random variables X, Y, Z and numbers $\lambda, \mu \in \mathbb{R}$ we have

$$\mathrm{Covar}(\lambda X + \mu Y, Z) = \lambda \, \mathrm{Covar}(X, Z) + \mu \, \mathrm{Covar}(Y, Z)$$

and

$$\mathrm{Covar}(X, \lambda Y + \mu Z) = \lambda \, \mathrm{Covar}(X, Y) + \mu \, \mathrm{Covar}(X, Z).$$

Exercise: Find the covariance between the number of ones and the number of sixes in n throws of a fair die.
(Hint: represent each of the two random variables as a sum of n random variables.)

Exercise: Show that for random variables X, Y and numbers $a, b \in \mathbb{R}$ we have

$$\mathrm{Covar}(X + a, Y + b) = \mathrm{Covar}(X, Y).$$

Correlation

Definition: The correlation between the random variables X and Y is

$$\rho(X, Y) = \frac{\mathrm{Covar}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$

Exercise: Show that for random variables X and Y and any $a, b, c, d \in \mathbb{R}$ with $a, c > 0$ we have

$$\rho(a(X - b), c(Y - d)) = \rho(X, Y).$$

Exercise: Show that for random variables X and Y, $-1 \leq \rho(X, Y) \leq 1$. Also show that, if $\rho(X, Y) = \pm 1$, $X = aY + b$ for some $a, b \in \mathbb{R}$.
(Hint: use the previous exercise to show that it is enough to prove that $-1 \leq \rho(X', Y') \leq 1$ where

$$X' = \frac{X - E(X)}{\sqrt{\mathrm{Var}(X)}}, \quad Y' = \frac{Y - E(Y)}{\sqrt{\mathrm{Var}(Y)}}.$$

Now

$$0 \leq \mathrm{Var}(X' + Y') = \mathrm{Var}(X') + \mathrm{Var}(Y') + 2\,\mathrm{Covar}(X', Y')$$

and

$$0 \leq \mathrm{Var}(X' - Y') = \mathrm{Var}(X') + \mathrm{Var}(Y') - 2\,\mathrm{Covar}(X', Y').)$$

Important formula

For any two random variables X and Y
$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Covar}(X, Y)$
$= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\rho(X, Y)\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$

The normal distribution

A random variable X is *normally distributed with mean* μ *and variance* σ^2 if

$$P(a \leq X \leq b) = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx,$$

i.e., if its density function is

$$f(\mu, \sigma^2; x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

and its distribution function is

$$N(\mu, \sigma^2; x) := \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt.$$

Standard normal random variables

Definition: X has a *standard normal distribution* if it is normally distributed with mean $\mu = 0$ and variance $\sigma^2 = 1$.
We denote $N(x)$ (or $\Phi(x)$) the distribution function of such a random variable, i.e.,

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

Exercise: Show that $N(x) = 1 - N(-x)$.
If X is a normally distributed random variable with mean μ and standard deviation σ then

$$\mathrm{P}(X \leq a) = N\left(\frac{a - \mu}{\sigma}\right)$$

Normal random variables: confidence intervals

We have the following confidence intervals

$$\mathsf{P}(\mu - 1.645\sigma < X < \mu + 1.645\sigma) \approx 90\%,$$

$$\mathsf{P}(\mu - 1.96\sigma < X < \mu + 1.96\sigma) \approx 95\%$$

$$\mathsf{P}(\mu - 2.58\sigma < X < \mu + 2.58\sigma) \approx 99\%.$$

Exercise: Let the current price of a stock be £10 and let S be the random variable representing its price one year in the future. Suppose that we are told that $\log S_1$ is a normal random variable with mean $\log(12)$ and standard deviation 0.2. What is the probability that he value of the stock in one year will be below £5?