

# Probability

Keith A. Lewis

September 2, 2019

## Abstract

This note collects salient facts about probability theory.

## Contents

<b>Probability Model</b>	<b>2</b>
Sample Space . . . . .	2
Probability Measure . . . . .	2
Algebra . . . . .	2
Partition . . . . .	3
Measurable . . . . .	3
<b>Random Variable</b>	<b>3</b>
Expected Value . . . . .	4
Moments . . . . .	4
Cumulants . . . . .	5
Conditional Expectation . . . . .	5
Joint Distribution . . . . .	6
Independent . . . . .	6
Copulas . . . . .	6
Characteristic Function . . . . .	7
Fourier Transform . . . . .	7
Examples . . . . .	7
Poisson . . . . .	8
Infinitely Divisible . . . . .	8
Normal . . . . .	8
Poisson . . . . .	8
Infinitely Divisible . . . . .	8
Stochastic Processes . . . . .	8
Brownian Motion . . . . .	8
L'evy Processes . . . . .	8
Remarks . . . . .	8
Examples . . . . .	9

Discrete . . . . .	9
Uniform . . . . .	9
Normal . . . . .	9

Probability is an extension of logic. Instead of propositions being either true or false a degree of belief can be specified for events occurring. All probabilities are conditional on models of available information.

## Probability Model

A *probability model* specifies a *sample space* and a *probability measure*.

### Sample Space

A sample space is what can happen: heads or tails as the outcome of a coin toss, the integers from 1 to 6 as the outcomes of rolling a single die, the set of all sequences of not more than 280 characters as a model of possible Twitter tweets.

An *event* is a subset of a sample space.

People seem to be surprised probabilities are modeled using sets. Sets have no structure, they are just a bag of things (*elements*).

### Probability Measure

A *probability measure* assigns a number between 0 and 1 to events. If  $\Omega$  is a sample space and  $P$  is a probability measure then the measure of the union of sets is the sum of the measure of each set minus the measure of the intersection:  $P(E \cup F) = P(E) + P(F) - P(E \cap F)$  for events  $E$  and  $F$ . This is the mathematical way to say measures do not double count.

A probability measure must also satisfy  $P(\emptyset) = 0$  and  $P(\Omega) = 1$ .

Exercise. If  $Q$  is a measure with  $Q(\emptyset) = a$  and  $Q(\Omega) = b$ , show  $(Q - a)/(b - a)$  is a probability measure.

### Algebra

An *algebra of sets*, or *algebra*, on  $\Omega$  is a collection of subsets (events),  $\mathcal{A}$ , that is closed under complement and union. This lets us talk about an event not happening and whether event  $A$  or  $B$  occurred.

We also assume the empty set belongs to  $\mathcal{A}$ , hence also  $\Omega$ . By De Morgan's Laws an algebra is also closed under intersection. The *power set* of  $\Omega$ ,  $2^\Omega = \{E : E \subseteq \Omega\}$ , clearly satisfies these conditions.

An *atom* of an algebra is a member,  $A$ , of the algebra such that if  $B \subseteq A$  and  $B$  is in the algebra, then either  $B = A$  or  $B$  is the empty set.

## Partition

A *partition* of a set is a collection of pairwise disjoint subsets whose union is equal to the set.

Exercise. If an algebra on  $\Omega$  is finite its atoms form a partition of  $\Omega$ .

Hint: Show  $A_\omega = \cap\{B \in \mathcal{A} : \omega \in B\}$ ,  $\omega \in \Omega$ , is an atom

This shows there is a one-to-one correspondence between finite partitions and finite algebras of sets. A partition is the mathematical way of specifying partial information. Knowing the outcome,  $\omega \in \Omega$ , corresponds to complete knowledge. Knowing which atom the outcome belongs to corresponds to partial knowledge. For example, the partition  $\{\{1, 3, 5\}, \{2, 4, 6\}\}$  corresponds to knowing whether the roll of a die is odd or even.

The coarsest partition  $\{\Omega\}$  corresponds to no knowledge while the finest partition  $\{\{\omega\} : \omega \in \Omega\}$  corresponds to complete knowledge.

## Measurable

A function  $X : \Omega \rightarrow \mathbf{R}$  is  *$\mathcal{A}$ -measurable* if the sets  $X^{-1}((-\infty, x]) = \{\omega \in \Omega : X(\omega) \leq x\}$  belong to  $\mathcal{A}$  for  $x \in \mathbf{R}$ .

Exercise: If  $\mathcal{A}$  is finite, show that a function is measurable if and only if it is constant on atoms of  $\mathcal{A}$ .

In this case  $X : \mathcal{A} \rightarrow \mathbf{R}$  is indeed a function on the atoms.

## Random Variable

A *random variable* is a variable, a symbol that can be used in place of a number, with additional information: the probability of the values it can take on. The *cumulative distribution function* is  $F(x) = F^X(x) = P(X \leq x)$ . It tells you everything there is to know about  $X$ . For example,  $P(a < X \leq b) = F(b) - F(a)$ .

Exercise. Show  $P(a \leq X \leq b) = \lim_{x \uparrow a} F(b) - F(x)$ .

Hint:  $[a, b] = \cap_n (a - 1/n, b]$ .

In general,  $P(X \in A) = E1_A = \int 1_A(x) dF(x)$  for sufficiently nice  $A \subset \mathbf{R}$  where we are using Riemann–Stieltjes integration.

Exercise: Show for any cumulative distribution function,  $F$ , that  $F$  is non-decreasing,  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,  $\lim_{x \rightarrow \infty} F(x) = 1$  and  $F$  is right continuous with left limits.

Every such function defines a random variable.

The cdf  $F(x) = \max\{0, \min\{1, x\}\}$  defines the uniformly distributed random variable  $U$ . For  $0 \leq a < b \leq 1$ ,  $P(a < U < b) = b - a$ .

Two random variables,  $X$  and  $Y$ , have the same *law* if they have the same cdf.

Exercise. If  $X$  has cdf  $F$ , then  $X$  and  $F^{-1}(U)$  have the same law.

Exercise. If  $X$  has cdf  $F$ , then  $F(X)$  and  $U$  have the same law.

This shows a uniformly distributed random variable has sufficient randomness to generate any random variable. There are no random, random variables.

The mathematician's definition of a random variable is that it is a measurable function  $X: \Omega \rightarrow \mathbf{R}$ . Its cumulative distribution function is  $F(x) = P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\})$ . Given a cdf  $F$  we can define  $X: \mathbf{R} \rightarrow \mathbf{R}$  to be the identity function and let  $P$  be the probability measure defined by  $F$ :  $P(A) = \int 1_A(x) dF(x)$ .

## Expected Value

The *expected value* of a random variable is defined by the  $EX = \int_{-\infty}^{\infty} x dF(x)$ . The expected value of any function of a random variable is  $Ef(X) = \int_{-\infty}^{\infty} f(x) dF(x)$ .

The *indicator* (or *characteristic*) function  $1_A(\omega)$  is 1 if  $\omega \in A$  and 0 if  $\omega \notin A$ . If  $X = \sum a_i 1_{A_i}$  where  $a_i \in \mathbf{R}$  and  $A_i$  are events, Define the *expected value* of  $X$  by  $EX = \sum_i a_i P(A_i)$ .

Exercise. Show that if  $\sum_i a_i 1_{A_i} = 0$  then  $\sum_i a_i P(A_i) = 0$ .

Hint: Replace the  $A_i$  by disjoint  $B_j$  so  $b_j = 0$  for all  $j$ .

This shows expected value is well-defined.

Exercise. Show  $P(\cup_i A_i) = \sum_i P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) \cdots$ .

Hint: Use  $(1_A - 1_{A_1}) \cdots (1_A - 1_{A_n}) = 0$ , where  $A = \cup_{k=1}^n A_k$ .

## Moments

The *moments* of a random variable,  $X$ , are  $m_n = E[X^n]$ ,  $n = 0, 1, 2, \dots$ . They don't necessarily exist for all  $n$ , except for  $n = 0$ . They also cannot be an arbitrary sequence of values.

Suppose all moments of  $X$  exist, then for any complex numbers,  $(c_i)$ ,  $0 \leq E|\sum_i c_i X^i|^2 = E\sum_{j,k} c_j \bar{c}_k X^{j+k} = \sum_{j,k} c_j \bar{c}_k m_{j+k}$ . This says the Hankel matrix,  $M = [m_{j+k}]_{j,k}$ , is positive definite. The converse is also true: if the Hankel matrix is positive definite there exists a random variable with the corresponding moments. This is not a trivial result and the random variable might not be unique.

## Cumulants

The *cumulant* of a random variable,  $X$ , is  $\kappa(s) = \kappa^X(s) = \log E \exp(sX)$ . The *cumulants*,  $\kappa_n$ , are defined by  $\kappa(s) = \sum_{n \geq 0} \kappa_n s^n / n!$ .

It is easy to see  $\kappa_1 = EX$  and  $\kappa_2 = \text{Var } X$ . The third and fourth cumulants are related to skew and kurtosis. We will see the exact relationship below.

If  $c$  is a constant then  $\kappa^{cX}(s) = \kappa^X(cs)$  so  $\kappa_n^{cX} = c^n \kappa_n^X$ . If  $X$  and  $Y$  are independent then  $\kappa^{X+Y}(s) = \kappa^X(s) + \kappa^Y(s)$  so  $\kappa_n^{X+Y} = \kappa_n^X + \kappa_n^Y$ .

## Bell Polynomial

The relationship between moments and cumulants is given by *Bell polynomials*. They are defined by  $\exp(\sum_1^n n! a_n s^n / n!) = \sum_0^\infty B_n(a_1, \dots, a_n) s^n / n!$ . Taking the derivative with respect to  $s$  and equating powers of  $s$  shows  $B_0 = 1$  and  $B_{n+1}(a_1, \dots, a_{n+1}) = \sum_{k=0}^n \binom{n}{k} B_{n-k}(a_1, \dots, a_{n-k}) a_{k+1}$ .

Bell polynomials show the connection between the moments and the cumulants of a random variable since  $E \exp(sX) = \sum_0^\infty EX^n s^n / n! = \sum_0^\infty m_n s^n / n!$  where  $m_n$  is the  $n$ -th moment and  $E \exp(sX) = \exp(\kappa(s)) = \exp(\sum_{n=1}^\infty \kappa_n s^n / n!)$ .

Exercise: Show  $m_n = \sum_{k=1}^n B_k(\kappa_1, \dots, \kappa_n)$ .

Exercise: Find the first five Bell polynomials.

In particular  $m_1 = \kappa_1$  and  $m_2 = \kappa_1^2 + \kappa_2$  so  $\kappa_1$  is the mean and  $\kappa_2$  is the variance. If the mean is 0 and the variance is 1, then  $\kappa_3$  is the skew and  $\kappa_4$  is the excess kurtosis.

## Conditional Expectation

The *conditional expectation* of an event  $B$  given an event  $A$  is  $P(B|A) = P(B \cap A) / P(A)$ . In some sense, this reduces the sample space to  $A$ . In particular,  $P(A|A) = 1$ . Since  $P(A|B) = P(A \cap B) / P(B)$  we have  $P(A|B) = P(B|A)P(A) / P(B)$ . This is the simplest form of Bayes Theorem. It shows how to update your degree of belief based on new information. Every probability is conditional on given information.

Define  $E[X|A] = E[X1_A]/P(A)$  for any random variable  $X$ . If  $X = 1_B$  then this coincides with the definition of conditional expectation above.

This is how we define  $E[X|\mathcal{A}] : \mathcal{A} \rightarrow \mathbf{R}$  on atoms of  $\mathcal{A}$ .

## Joint Distribution

Two random variables,  $X$  and  $Y$ , are defined by their *joint distribution*,  $H(x, y) = P(X \leq x, Y \leq y)$ . For example, the point  $(X, Y)$  is in the square  $(a, b] \times (c, d]$  with probability  $P(a < X \leq b, c < Y \leq d) = P(X \leq b, Y \leq d) - P(X \leq a) - P(Y \leq c) + P(X \leq a, Y \leq c)$ .

The *marginal distributions* are  $F(x) = H(x, \infty)$  and  $G(y) = H(\infty, y)$ , where  $F$  and  $G$  are the cumulative distributions of  $X$  and  $Y$  respectively.

In general, the joint distribution of  $X_1, \dots, X_n$  is  $F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$ .

## Independent

The random variables  $X$  and  $Y$  are *independent* if  $F^{X,Y}(x, y) = F^X(x)F^Y(y)$  for all  $x$  and  $y$ . This is equivalent to  $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$  for any sets  $A$  and  $B$ .

We also have that  $Ef(X)g(Y) = Ef(X)Eg(Y)$  for any functions  $f$  and  $g$  whenever all expected values exist.

Exercise: Prove this for the case  $f = \sum_i a_i 1_{A_i}$  and  $g = \sum_j b_j 1_{B_j}$ .

In general,  $X_1, \dots, X_n$  are independent if  $F(x_1, \dots, x_n) = F_1(x_1) \cdots F_n(x_n)$ , where  $F_j$  is the law of  $X_j$ .

## Copulas

A *copula* is the joint distribution of uniformly distributed random variables on the unit interval. Let  $U$  and  $V$  be two uniformly distributed random variables. The copula of  $X$  and  $Y$  is the joint distribution of  $F^{-1}(X)$  and  $G^{-1}(Y)$  where  $F$  and  $G$  are the cumulative distributions of  $X$  and  $Y$  respectively:  $C^{X,Y}(u, v) = P(F^{-1}(X) \leq u, G^{-1}(Y) \leq v)$ .

Exercise: Show  $C^{X,Y}(u, v) = H(F(u), G(v))$  where  $C^{X,Y}$  is the copula of  $X$  and  $Y$ , and  $H$  is the joint distribution of  $X$  and  $Y$ .

Exercise: Show  $H(x, y) = C(F^{-1}(x), G^{-1}(y))$

This shows how to use the copula and marginal distributions to get the joint distribution.

If  $U$  and  $V$  are independent, uniformly distributed random variables on the unit interval then  $C(u, v) = uv$ .

If  $V = U$  then their joint distribution is  $C(u, v) = P(U \leq u, V \leq v) = P(U \leq u, U \leq v) = P(U \leq \min\{u, v\}) = \min\{u, v\} = M(u, v)$ .

If  $V = 1 - U$  then their joint distribution is  $C(u, v) = P(U \leq u, V \leq v) = P(U \leq u, 1 - U \leq v) = P(1 - v \leq U \leq u) = \max\{u - (1 - v), 0\} = \max\{u + v - 1, 0\} = W(u, v)$

Exercise: (Frechet-Hoeffding) For every copula,  $W \leq C \leq M$ .

### Characteristic Function

The *characteristic function* of a random variable,  $X$ , is  $\xi(t) = \kappa(it)$ .

### Fourier Transform

The *Fourier transform* is  $\psi(t) = \xi(-t) = \kappa(-it)$ . Clearly  $\psi(t) = \xi(-t)$ .

### Examples

Move!!! These can be used to prove the *central limit theorem*: if  $X_j$  are independent, identically distributed random variables with mean zero and variance one, then  $(X_1 + \dots + X_n)/\text{sqrtn}$  converges to a standard normal random variable.

If  $X$  is normal then  $E \exp(X) = \exp(EX + \text{Var}(X)/2)$  so the cumulants satisfy  $\kappa_n = 0$  for  $n > 2$ .

If  $X$  is Poisson with parameter  $\lambda$  then

$$\begin{aligned} E e^{sX} &= \sum_{k=0}^{\infty} e^{sk} e^{-\lambda} \lambda^k / k! \\ &= \sum_{k=0}^{\infty} (e^s \lambda)^k e^{-\lambda} / k! \\ &= \exp(\lambda(e^s - 1)) \end{aligned}$$

so  $\kappa(s) = \lambda(e^s - 1)$  and  $\kappa_n = \lambda$  for all  $n$ . ### Normal

## Poisson

### Infinitely Divisible

A random variable,  $X$ , is *infinitely divisible* if for any positive integer,  $n$ , there exist independent, identically distributed random variables  $X_1, \dots, X_n$  such that  $X_1 + \dots + X_n$  has the same law as  $X$ .

Characteristic function ...

moments, Hamburger moment problem.

cumulants, Bell polynomials

## Normal

## Poisson

### Infinitely Divisible

## Stochastic Processes

A *stochastic process* is ...

## Brownian Motion

reflection

## L'evy Processes

## Remarks

Cheval de Mere

Pascal

Bernoulli(s)

Kolmogorov

Willy Feller



## Examples

### Discrete

A *discrete* random variable is defined by  $x_i \in \mathbf{R}$  and  $p_i > 0$  with  $\sum p_i = 1$ . The probability the random variable takes on value  $x_i$  is  $p_i$ .

### Bernoulli

A *Bernoulli* random variable has  $P(X = 1) = p$ ,  $P(X = 0) = 1 - p$ .

### Binomial

A *Binomial* random variable ...  $P(X = k) = \binom{n}{k}/2^n$ .

### Uniform

The random variable,  $U$ , that is *uniformly distributed* on the *unit interval*,  $[0, 1]$ , has cdf  $F(x) = x$  if  $0 \leq x \leq 1$ ,  $= 0$  if  $x < 0$ , and  $= 1$  if  $x > 1$ .

### Normal