Probability

Keith A. Lewis

August 12, 2019

Abstract

This short note collects salient facts about probability theory.

Contents

Probability

Probability is an extension of logic. Instead of propositions being either true or false a degree of belief can be specified for events occuring.

A *probability model* specifies a *sample space* and a *probability measure*. The sample space is just a set of what can possibly happen: heads or tails as the outcome of a coin toss, the numbers 1 through 6 as the outcomes of rolling a singled die, the set of all sequences of not more than 280 characters as a model of possible Twitter tweets.

People seem to be surprised probabilities are modeled using a set. Sets have no structure, they are just a bag of things.

An *event* is a subset of a sample space. A *probability measure* assigns a number between 0 and 1 to events. If Ω is a sample space and P is a probability measure then $P(E\cup F) = P(E) + P(F) - P(E\cap F)$ for events E and F: the measure of the union of sets is the sum of the measure of each set minus the measure of the intersection. 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.

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

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

Exercise. (Inclusion-Exclusion principal) Let S be a finite set and let f be any function defined on subsets of S. Define $\phi f(T) = \sum_{Y\supseteq T} f(Y)$ and $\psi g(T) = \sum_{Y\supseteq T}(-1)^{|Y|-|T|} g(T)$. Show $\phi\psi g = g$ and $\psi\phi f = f$.

Hint: Group the sum by $|Y| - |T|$.

Algebras of Sets

An *algebra of sets* on Ω is a collection of subsets, $\mathcal{A}$, that is closed under complement and union. We also assume the empty set belongs to $\mathcal{A}$. By De Morgan's Law an algebra is also closed under intersection and Ω belongs to $\mathcal{A}$.

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.

Exercise. If an algebra on Ω is finite its atoms form a partition of Ω.

Recall a *partition* of a set is a collection of pairwise disjoint subsets whos union is equal to the set.

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 of a sample 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.

Probability Spaces

A triple $\langle \Omega, P, \mathcal{A} \rangle$ where Ω is a set, P is a probability measure with domain $\mathcal{A}$ is called a *probability space*. A *random variable* is a function $X \colon \Omega \to \mathbf{R}$ that is $\mathcal{A}$ measureable: 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 $X \colon \mathcal{A} \to \mathbf{R}$ is indeed a function on the atoms.

Random Variables

A *random variable* is a function $X \colon \Omega \to \mathbf{R}$. Its *cumulative distribution function* is $F(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x)$. It tells you everything there is to know about X. A random variable is a variable: a symbol that can be used in place of a number. It has additional information: the probability of the values it can take on.

Moments

The *moments* of a random variable, X, are $m_n = E[X^n]$, $n = 0, 1, 2, \ldots$. 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 $0 \leq E|\sum_i c_i X^i|^2 = E \sum_{j,k} c_j \bar{c_k} X^{j+k} = \sum_{j,k} c_j \bar{c_k} m_{j+k}$. This says the Hankel matrix, $[m_{j+k}]_{j,k}$, is positive definite. The converse is true: if the Hankel matrix is positive definite there exists a random variable with the corresponding moments.

Define the linear operator $M : \ell^2 \to \ell^2$ by $\ldots$

Spectral measure $\ldots$

Cumulants

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

Bell Polynomials

The relationship between moments and cumulants is given by Bell polynomials.

Examples

Uniform

every rv comes from uniform

Normal

Poisson

Infinitely Divisible

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.

If we write this as $E[X|A]P(A) = E[X1_A]$ then defining $E[X|\mathcal{A}]$ by $E[X|\mathcal{A}]P|_{\mathcal{A}} = (XP)_{\mathcal{A}}$ agrees on atoms of $\mathcal{A}$.

Random Variables

moments, Hamberger moment problem.

cumulants, Bell polynomials

Normal

Poisson

Infinitely Divisible

Stochastic Processes

Brownian Motion

reflection

Le'vy Processes

Remarks

Cheval de Mere

Pascal

Bernoulli(s)

Kolmogorov

Willy Feller