Probability Model Components

Ω = sample space = set of all possible outcomes of an experiment

An event is a subset of Ω.

$P(E)$ = probability of the event E. $P(\cdot)$ is a function that maps events to the unit interval $[0, 1]$.

Probability Axioms

1. For any event E,
$$0 \leq P(E) \leq 1.$$
2. $P(\Omega) = 1$
3. For disjoint events $E_1, E_2, \ldots$,
$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

Other Probability Rules and Concepts

1. *Rule of Complements.* $P(E) = 1 - P(E^C)$
2. *General Addition Rule.* For any n events $E_1, E_2, \ldots, E_n$,
$$\begin{aligned} P(\cup_i E_i) &= \sum_i P(E_i) - \sum_{i<j} P(E_i \cap E_j) + \sum_{i<j<k} P(E_i \cap E_j \cap E_k) - \sum_{i<j<k<\ell} P(E_i \cap E_j \cap E_k \cap E_\ell) \\ &\quad + \cdots + (-1)^{n+1} P(\cap_{i=1}^{n} E_i) \end{aligned}$$
3. *Conditional Probability.* Think of this whenever you need to compute a conditional probability. For events A and B, as long as $P(B) \neq 0$, define the conditional probability of A given that B occurs to be
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$
4. *Law of Total Probability* This is VERY USEFUL. Think of this when you try to "condition on" an event to calculate the probability of another. Let $E_1, \ldots, E_n$ partition Ω. For any event E we have
$$P(B) = \sum_{k=1}^{n} P(B \mid E_k) P(E_k)$$
5. *Bayes's Rule.* Also very useful. Think of this whenever you need to compute a conditional probability.
$$P(E^* \mid B) = \frac{P(B \mid E^*) P(E^*)}{\sum_{k=1}^{n} P(B \mid E_k) P(E_k)}$$

6. *Independence.* Two events E_1 and E_2 are independent if

$$P(E_1|E_2) = P(E_1)$$

$$\iff P(E_2|E_1) = P(E_2) \iff P(E_1 \cap E_2) = P(E_1)P(E_2)$$

Events $E_1, \ldots, E_n$ are mutually independent if $\forall\ S \subset \{1, \ldots n\}$

$$P\left(\bigcap_{k=1}^{n} E_k\right) = \prod_{k=1}^{n} P(E_k)$$

Pairwise independence does not imply mutual independence in general.

7. *A Multiplication Rule.*

$$P\left(\cap_{i=1}^{n} E_i\right) = P(E_1)P(E_2|E_1)P(E_3|E_2, E_1)\cdots P(E_n|E_1, E_2, \ldots, E_{n-1})$$

Example 1 What's the chance current flows from A to B if switches are connected like this? Given the system fails, what is the probability that switch #5 is down?

Random Variables

Ω : Sample space
$\mathbb{R}$: Set of all real numbers.

A real random variable is a function

$$X : \Omega \to \mathbb{R}$$

The cumulative distribution function (cdf) of X is

$$F(x) = P(X \leq x) \quad \forall x \in \mathbb{R}$$

and has the properties

(i) $\lim_{x \to -\infty} F(x) = 0$ (ii) $\lim_{x \to +\infty} F(x) = 1$

(iii) $F(x)$ is non decreasing

(iv) $F(x)$ is right continuous.

A random variable whose possible values form a countable set is called discrete. The probability mass function (pmf) of the discrete random X is

$$P(X = x) = p(x) \leftarrow \text{pmf of } X.$$

The cdf of a discrete random variable is a step function:

A random variable X is a continuous random variable if there exists a function f such that the cdf F can be expressed as

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

The function f is called the probability density function of the random variable, and if F is differentiable, then $f(x) = F'(x)$. The density f satisfies

1. $f(x) \geq 0 \qquad \forall x \in \mathbb{R}$
2. $\int_{\mathbb{R}} f(x)\, dx = 1$

The expected value (or mean) of X is

$$\mu_X = \begin{cases} \sum x\, p(x) & \text{if } X \text{ is discrete.} \\ \int_{-\infty}^{\infty} x f(x)\, dx & \text{if } X \text{ is Continuous.} \end{cases}$$

The n^{th} moment of the random variable X is

$$E(X^n) = \begin{cases} \sum x^n p(x) & ; \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x^n f(x)\, dx & ; \text{if } X \text{ is Continuous.} \end{cases}$$

and the n^{th} central moment is

$$E[(X-\mu_X)^n]$$

The second central moment is called the variance:

$$\begin{aligned} Var(X) &= E[(X-\mu)^2] \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

The random variables $X_1, \ldots, X_n$ are independent if

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i) \qquad \text{if discrete}$$

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i) \qquad \text{if continuous}$$

Common Discrete Distributions

Distribution	pmf $p(x)$	mgf $\phi(t)$	Mean	Variance
Uniform(a,b) *Models outcome of experiment with equally likely outcomes in*	$\frac{\mathbf{1}\{x\in\{a,a+1,\ldots,b\}\}}{b-a+1}$	$\frac{e^{at}-e^{(b+1)t}}{(b-a+1)(1-e^t)}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Bernoulli(p) *Models number of successes in 1 trial*	$p^x(1-p)^{1-x}\mathbb{1}\{x\in\{0,1\}\}$	$(pe^t+(1-p))$	p	$p(1-p)$
Binomial(n,p) *Models number of successes in n trials*	$\binom{n}{x}p^x(1-p)^{n-x}\mathbb{1}\{x\in\{0,1,\ldots,n\}\}$	$(pe^t+1-p)^n$	np	$(1-p)$
Poisson(λ) *Models number of events occurring in a continuous time*	$\frac{e^{-\lambda}\lambda^x}{x!}\mathbb{1}\{x\in\{0,1,2,\ldots\}\}$	$e^{-\lambda(1-e^t)}$	λ	λ
Geometric(p) *Models number of trials until the first success*	$p(1-p)^{x-1}\mathbb{1}\{x\in\{1,2,3,\ldots\}\}$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial(r,p) *Models number of trials until r^{th} success*	$\binom{x-1}{r-1}p^r(1-p)^{x-r}$, $r\in\mathbb{Z}^+$, $x\in\{r,r+1,\ldots\}$	$\left(\frac{pe^t}{1-(1-p)e^t}\right)^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

~~**Theorem 2** *The geometric distribution is the only discrete distribution with the memoryless property.*~~

~~**Proof**~~

∎

Common Continuous Distributions

Distribution	pdf $f(x)$	mgf $\phi(t)$	Mean	Variance
Uniform(a,b) *Models outcome of experiment that's equally likely to occur anywhere in* (a,b)	$\frac{1}{b-a}\mathbb{1}\{a \le x \le b\}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential(λ) *Continuous analog of the geometric*	$\lambda e^{-\lambda x}\mathbb{1}\{x \ge 0\}$	$\frac{\lambda}{\lambda-t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma(k,λ) *also called* Erlang(k,λ) *when* $k \in \mathbb{Z}^+$. *Continuous analog of the negative binomial*	$\frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!}\mathbb{1}\{x \ge 0\}$	$\left(\frac{\lambda}{\lambda-t}\right)^n$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$
Normal(μ,σ^2)	$\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}$, $x \in \mathbb{R}$	$e^{\mu t+\sigma^2 t^2/2}$	μ	σ^2

~~**Theorem 3** *The exponential distribution is the only continuous distribution*~~ *with the memoryless property.*

~~**Proof** See homework~~ ∎

Functions of Random Variables

If X is a random variable and g is a function such that $g : \mathbb{R} \longrightarrow S$ where $S \subset \mathbb{R}$, then $Y = g(X)$ is also a random variable. If X and Y are continuous random variables, we get the pdf of Y by

1. finding the cdf of Y,
2. differentiating the cdf of Y to find the pdf of Y

Example 4 Let U be a Uniform$(0,1)$ random variable. What is the distribution of $X = -\log U$? What is the distribution of $Y = \lambda^{-1}X$?

Multivariate Distributions

The distribution of the discrete random vector $(X_1, \ldots, X_n)$ is characterized by the joint probability mass function

$$p(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots X_n = x_n)$$

Example 5 Suppose an experiment can result in exactly one of r distinct outcomes, and the experiment is performed n times where the outcomes are independent. Let

$$p_i = P(\text{experiment results in outcome } i), \qquad \sum_{i=1}^{r} p_i = 1$$

and

$$X_i = \text{number of times outcome } i \text{ occurs}$$

Then

$$P(X_1 = x_1, \ldots, X_r = x_r) = n! \prod_{i=1}^{r} \frac{p_i^{x_i}}{x_i!}$$

The distribution of the continuous random vector is characterized by the joint cumulative distribution function F:

$$F(x_1, \ldots, x_n) = P(X_1 \le x_1, \ldots, X_n \le x_n) = \int_{(-\infty, x_1]} \cdots \int_{(-\infty, x_1]} f(s_1, \ldots, x_n)\, ds_1 \cdots ds_n$$

Some Nice Relationships Between Random Variables

Assuming independence,

1. $X_i \sim \text{Bernoulli}(p) \Rightarrow \sum_i X_i \sim \text{binomial}(n, p)$
2. $X_i \sim \text{binomial}(n_i, p) \Rightarrow \sum_{i=1}^{n} X_i \sim \text{binomial}(\sum_i n_i, p)$
3. $X_i \sim \text{Poisson}(\lambda_i) \Rightarrow \sum_i X_i \sim \text{Poisson}(\sum_i \lambda_i)$
4. $X_1, \ldots, X_r$ iid geometric$(p) \Rightarrow \sum_{i=1}^{r} X_i \sim$negative binomial$(r, p)$
5. $X_1, \ldots, X_k \sim$ iid $\exp(\lambda) \Rightarrow \sum_{i=1}^{k} X_i \sim \text{Erlang}(k, \lambda)$

Useful Laws/Rules/etc.

1. $E\left[\sum_{i=1}^{n} a_i g_i(X_i) + b\right] = \sum_{i=1}^{n} a_i E g_i(X_i) + nb$, regardless of whether or not the X_i are independent.
2. $\text{Var}(\sum_{i=1}^{n} a_i g_i(X_i) + b) = \sum_{i=1}^{n} a_i^2 \text{Var}(g_i(X_i)) + 2\sum_{1 \le i < j \le n} a_i a_j \text{Cov}(g_i(X_i), g_j(X_j))$
3. *Law of Total Expectation.*
$$E[X] = \sum_k E[X|A_k]P(A_k) \text{ for disjoint and exhaustive } A_k.$$
4. *Conditional Variance Formula.*
5. *Tower Property.* $E[X|Y] = E[E[X|Y, Z]|Y]$. Also
$$EX = E[E[X|Y]] = \sum_y E[X|Y = y]P(Y = y) \text{ if } Y \text{ is discrete.}$$
$$EX = E[E[X|Y]] = \int_{\mathbb{R}} E[X|Y = y] f_Y(y) dy \text{ if } Y \text{ is continuous.}$$

6. If $X_1, \ldots, X_n$ are independent,

   (a) $E\left[\prod_{i=1}^n g_i(X_i)\right] = \prod_{i=1}^n Eg_i(X_i)$

   (b) $\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i)$

7. *Law of Total Probability for Random Variables.*

$$P(E) = \sum_{i=1}^{\infty} P(E|X = x)P(X = x) \text{ if } X \text{ is discrete.}$$

$$P(E) = \int_{\mathbf{R}} P(E|X = x)f(x)dx \text{ if } X \text{ is continuous.}$$

8. For continuous random variables, the conditional density function of X_1 given $X_2 = x_2$ is

$$f_{X_1|X_2}(x_1|x_2) = \frac{f(x_1, x_2)}{f_{X_2}(x_2)}$$

   and

$$E[X_1|X_2 = x_2] = \int_{\mathbf{R}} x_1 f_{X_1|X_2}(x_1|x_2)dx_1$$

Transforms

The characteristic function of the random variable X is defined to be

$$\psi_X(t) = Ee^{itX}$$

where $i = \sqrt{-1}$. Characteristic functions always exist and are unique for each random variable: if you know the characteristic function, you know the distribution, and vice-versa.

The moment generating function of the random variable X is defined to be

$$\phi_X(t) = Ee^{tX}$$

They do not exist for some random variables, but when they do, they are unique. The moment generating function and characteristic function of X are related through

$$\phi_X(it) = \psi_X(t)$$

Moment generating functions have the useful property that the nth derivative of the moment generating function evaluated at 0 is the nth moment of the random variable:

$$\frac{d^n}{dt^n}\phi_X(t)|_{t=0} = E[X^n]$$

The Laplace transform L_X of the random variable X is defined to be

$$L_X(t) := Ee^{-tX}$$

Laplace transforms exist for all non-negative random variables and are related to characteristic functions and moment generating functions through

$$L_X(t) = \psi_X(it) \qquad L_X(t) = \phi_X(-t)$$

Limit Theorems

Theorem 7 *Markov's Inequality. If X is a nonnegative random variable, then for any $a > 0$*

$$P(X \geq a) \leq \frac{EX}{a}$$

Proof Pages 77 and 78. ∎

Theorem 8 *Chebyshev's Inequality. Let X be* **any** *random variable with mean μ and variance σ^2. Then the chance that X is at least k standard deviations from its mean is bounded from above like this:*

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Proof Page 78. ∎

Theorem 9 Classical Central Limit Theorem.*Let $X_1, X_2, \ldots, X_n$ be independent and identically distribution random variables with mean μ and variance $\sigma^2 < \infty$. Then*

$$P\left(\frac{n^{-1}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}} \leq z\right) \to P(Z \leq z)$$

as $n \to \infty$. Here, Z is normal random variable with mean $\mu = 0$ and variance $\sigma^2 = 1$ (called a standard normal).

Example 10 Pianos weigh 400 lbs on average, with a standard deviation of 25 lbs. Matt's living room floor will support 20,300 lbs. What's the chance his floor will support the 50 pianos he has stored in the living room?

Theorem 11 Strong Law of Large Numbers. *Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with mean μ. Then*

$$P\left(\left|\lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} X_k - \mu\right| > \epsilon\right) = 0$$

In other words,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \quad \text{as } n \to \infty$$

with probability 1.

Each distribution has a unique Laplace transform when it exits. Also, note $e^{-t} \in [0,1]$ for all $t > 0$, so Laplace transforms are always in $[0,1]$.

The probability generating function of the discrete random variable X is defined to be

$$P_X(t) = Es^X$$

When they exist, they uniquely determine distributions of random variables. Probability generating functions have the useful property that probabilities of random variables can be obtained by successive differentiation:

$$\frac{d^k}{ds^k} P_X(s) = k!P(X = k)$$

~~**Example 6** *The sum of independent Poissons is Poisson.*~~