Introduction to Stochastic Process
STAT 615 Course Notes

Daren Cline,[1*]

[1]Department of Statistics, Texas A&M University,
College Station 77843-3143, USA
E-mail: dcline@stat.tamu.edu.
Input by Tian Xia; Email: tianxia@cse.tamu.edu

0 Preliminaries

Definition. 0.1. A **sample space** $\mathbf{\Omega}$ is the collection of individual outcomes or realizations for a random experiment.

Subsets of $\mathbf{\Omega}$ are called events. For $A \subset \mathbf{\Omega}$, we say "A occurs" if the outcome $\omega \in A$

Ω can be countable (finite or countably infinite) or uncountable. It need not consisit of real-valued entities. Most often, its role is in the background and simply assured.

Example. 0.1. Toss a coin infinitely many times, observing 1 (heads) or 0 (tails) for each toss. The outcomes are inifinite sequences of 0's & 1's, such as $(1, 1, 0, 0, 0, 1, \cdots)$. The collection $\mathbf{\Omega}$ of all possible outcomes is uncountable in this case.

Events are subsets of $\mathbf{\Omega}$, often identified descriptively:

- "1st toss is H"$= \{(x_1, x_2, \cdots) : x_1 = 1\}$.

- "12 H's in the 1st 100 tosses" $= \{(x_1, x_2, \cdots) : \sum_{i=1}^{1} 00x_i = 12\}$.
- "the long tem proportion of H's is $\frac{1}{3}$" = $\{(x_1, x_2, \cdots) : \frac{1}{n}\sum_{i=1}^{n} x_i \to \frac{1}{3}\}$.

A singleton event is a subset of just one outcome, which is not quite the same as the outcome itself.

The empty set ϕ is considered to be an event also.

Definition. 0.2. A **probability measure** P is a set function (i.e. P is applied to events, not to outcomes) such that

(i) $0 \leq P(A) \leq 1$, with $P(\mathbf{\Omega}) = 1, P(\phi) = 0$

(ii) if $A_1, A_2, \cdots$ are disjoint then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ (similarly for finite unions)

when the A_i's are not disjoint, we have

Theorem 0.1. *(Boole's inequality)*

$$P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$$

.

Definition. 0.3. A collection of $\mathbf{A}$ of subsets of $\mathbf{\Omega}$ is called a **σ-algebra** (or σ-field) if

(i) $\phi, \mathbf{\Omega} \in \mathbf{A}$

(ii) $A \in \mathcal{A} \Longrightarrow A^{\mathrm{c}} \in \mathcal{A}$

(iii) $A_1, A_2, \cdots \in \mathcal{A} \Longrightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ (Some for finite unions) and hence also $\bigcap_{i=1}^{\infty} \in \mathcal{A}$

Note: the requirement is closure under countable unions and intersections. So a union of uncountably may subsets in $\mathcal{A}$ need not itself be in $\mathcal{A}$. (there are important mathematical reasons for allowing this.) **Additional Note:** complement $A^{\rfloor}$ ="A does not occur" union of events $\bigcup_i A_i$ = "some A_i (at least one) occurs" intersection $\bigcap_i A_i$ = "every A_i occurs"

Example. 0.2. If $\boldsymbol{\Omega}$ is countable (such as the integers $\mathbb{Z}$) then we can let $\mathcal{A}$ be all subsets.

Example. 0.3. If $\Omega = \mathbb{R}$ (all real values), we usually use the **Borel Sets** $B =$ smallest σ-algebra that contains all singletons and all intervals. So $\mathcal{B}$ would include countable unions of intervals, but this by no means comes close to all it has. There is no direct formula or description to characterize all arbitrary Borel sets.
On the other hand, given $B \in \mathcal{B}$ and a probability measure P, $P(B)$ can be approximated as close as we like with the probability of some finite union of intervals.
Borel sets on $\mathbb{R}^{\mathrm{d}}$ (d-dimensional vectors) are defined similaryly, using rectangles.

Example. 0.4. (cond.) Let $\Omega = \{(x_1, x_2, x_3, \cdots) : x_i = 0 \text{ or } x_i = 1 \text{ for each } i \geq 1\}$ be the sample space of infinite sequences of coin tosses by cylinder events (events where a finite number of the x_i's are fixed). For example,

$$\{(x_1, x_2, \cdots) : x_1 = 1, x_2 = 0\}$$

.

The triple $(\Omega, \mathcal{A}, P)$ (i.e. sample space, σ-algebra of events probability measure) is called a probability space.

Definition. 0.4. A **random variable** $X = X(\omega)$ is a real-valued function applied to outcomes (i.e., $X : \Omega \to \mathbb{R}$.
An extended random variable can possibly take infinite values (i.e., $X : \Omega \to \bar{\mathbb{R}} \overset{def}{=} \mathbb{R} \vee \{-\infty, \infty\}$ ("$\overset{def}{=}$" means "defined as", the same as "=:"))

Since random variables (rvs) are functions, it is highly recommended that you take care to distinguish them from actual (possible) values. We will usually, but not always, use upper case for rvs and lower case for actual values. (Also, try to avoid using the same name, such as X or Y, for everything, and centatinly make distinctions between rvs in the same problem/context.)

A random vector is a vector of rvs and a random (stochastic) process is a sequence of rvs (e.g. $(x_1, x_2, \cdots)$) or a function that takes random values (e.g. $x(t), t \in \mathbb{R}$, where each $X(t)$ is a random variable).

Definition. 0.5. Every rv X has a culultative distribution function (cdf)

$$F_X(x) = P(X \leq x)$$

(sensible even for extended rvs)

A rv X (or its cdf) is discrete if there is a countable set of values $\{a_i\}$ such that $\sum_i P(x = a_i) = 1, \{P_i\} = \{P(x = a_i)\}$ is called the probability mass function (pmf).

A rv X (or its cdf) is absolutely continuous if

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$$

for some function $f(x)$, called the probability density function (pdf)

However, we do not with to limit ourselves to just two types. There are, for example, rvs with neither a pmf or a pdf, and these are mixtures of types. Things become more general still for random vectors.

Furthermore, a stochastic process cannot be characterized so simply at all (More on this in the 1st chapter).

We will write $X \sim F_X$ (or $X \sim \{P_i\}$ or $X \sim f(x)$ when the context is clear) to signify that X has distribution given by F_x.

You should be familiar with all the commonly used distributions.

- ex. $Y \sim Poisson(\lambda)$ with pmf

$$P(y = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \cdots$$

.

- ex. $T \sim Gamma(\alpha, \beta)$ with pdf

$$f(t) = \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}, t \geq 0$$

.

Definition. 0.6. Random variables $X_1, \cdots, X_m$ are independent if

$$P(X_1 \in A_1, X_2 \in A_2, \cdots, X_m \in A_m) = \prod_{i=1}^{m} P(X_i \in A)$$

for all **Borel sets** $A_1, \cdots, A_m$.

Equivalently,

$$P(X_1 \leq x_1, \cdots, X_m \leq x_m) = \prod_{i=1}^{m} P(X_i \leq x_i)$$

for all real values $x_1, \cdots, x_m$.

(The 1st definition can be applied to random elements of any kind, with appropriate A_i's.)

A collection $\{X_t\}$ of rvs (possibly uncountably many) is independent if every finite subcollection is. That is, if $X_{t_1}, \cdots, X_{t_m}$ are independent, for any $t_1, \cdots, t_m$.

A sequence $\{X_n\}$ is independent and identically distributed (iid) if it is independent and all the rvs have the same distribution. We sometimes write $\{X_n\} \overset{iid}{\sim} F$ (or just $X_n \overset{iid}{\sim} F$) .

Stochastic processes are generally not independent (they wouldn't be interesting if they were). But they often can be constructed or described in terms of independent rvs.

Random variables and their distributions can also be characterized in terms of expectations. As before, we do not wish to be limited to special types (and in fact it is notationally more convenient not to be).

Definition. 0.7. Here is how the expectation $E(g(x))$ is defined (if possible).

For simplicity, just let $Y = g(X)$ and we'll define $E(Y)$.

1. indicator rv

   if
   $$Y = \mathbb{1}_A = \begin{cases} 0, & \text{if A does not occur} \\ 1, & \text{if A does occur} \end{cases}$$
   then
   $$E(\mathbb{1}_A) = P(A)$$
   . indicator rvs have Bernoulli distribution (which simply means discrete on $\{0, 1\}$)

2. simple rv

   Y is a finite linear combination of indicator rvs

   i.e. $Y = \sum_{i=1}^{m} c_i \mathbb{1}_{A_i}$ (hence discrete w/ finitely many values)

   then
   $$E(Y) = \sum_{i=1}^{m} c_i P(A_i)$$
   . (this works even if the A_i's are not disjoint or if the c_i's are not distinct.)

3. nonnegative rv

   $Y \geq 0$. There always exist nonnegative simple rvs $Y_n(\omega) \uparrow Y(\omega)$ for every $\omega \in \Omega$.

   Then
   $$E(Y) = \lim_{n\to\infty} E(Y_n) \quad \text{(which myght equal } \infty)$$
   (This does not depend on the choice of sequenc $\{Y_n\}$, nor is it necessarily the best way to compute $E(y)$.)

4. general rv Y

   Define the positive part of Y, $Y_+ = \max(Y, 0)$

and the negative part, $Y_- = \max(-Y, 0)$.

(both are nonnegative)

so $Y = Y_+ - Y_-$ and $|Y| = Y_+ + Y_-$

Then if either (or both) of $E(Y_+)\&E(Y_-)$ are finite, define

$$E(Y) = E(Y_+) - E(Y_-)$$

. (Again, the value can be infite - we just don't allow the case $\infty - \infty$)

The purpose of the above is just so we can define (and talk about) expectations without having to worry about the type of rv or how the expectation is to be computed. In fact, computing expectations will depend on context and often is not necessary in specific cases.

Definition. 0.8. (Lebesque-Stielties notation).

If $X \sim F$, it is conventional to express

$$E(x) = \int_{-\infty}^{\infty} xF(dx)$$

, and $E(g(x)) = \int_{-\infty}^{\infty} g(x)F(dx)$, when they can be defined.

A similar usage applies to random vectors, using a joint cdf (but not to stochastic processes) .

Other terminology.

- A random variable X is finite if $|X(\omega)| < \infty$ for all $\omega \in \Omega$.
- A random variable X is bounded if there exists a constant C such that $|X(\omega)| \leq C$ for all $\omega \in \Omega$.
- Do not confuse "finite" & "bounded".

The terms apply in similar fashion to any real valued function $g(x)$.

An event A occurs almost surely (a.s.) or with probability 1 (w.p.1) if P(A) = 1.
Note that "almost sure" is in the context of a specific probability measure or model, whereas the definition of A (as some event) is not.

ex. Y is almost surely finite if $P(|Y| < \infty) = 1$. However, the event $\{Y = \infty\}$ need not be empty. Indeed, we will want to discuss rvs taht are almost surely finite for some models but not for others (and to identify the distinguishing condition) .

ex. X_n converges to X almost surely ($X_n \to X$ a.s.), if $P(\lim_{n\to\infty} X_n = X) = 1$,
This is stronger than convergence in probability: $P(|X_n - X| > \epsilon) \to 0$ for all positive ϵ.

The basic limit theorems

Theorem 0.2. ***Monotone Convergence*** *(MCT)*
For $0 \le X_n \uparrow X : E(X_n) \to E(X)$ *(even if infinite)*
Likewise, if $0 \le g_n(x) \uparrow g(x)$ *for all* x*, then*

$$\int g_n(x)dx \to \int g(x)dx$$

, and

$$\sum_{i=1}^{\infty} g_n(i) \to \sum_{i=1}^{\infty} g(i)$$

Theorem 0.3. ***Dominated Convergence*** *(DCT)*
For $X_n \to X$ *such that* $|X_n| \le Y$ *and* $E(y) < \infty$*,* $E(X_n) \to E(X)$*.*
(again, similar statements for integrals & sums)

Theorem 0.4. ***Fatou's Lemma*** $X_n \ge 0, X_n \to X$

$$\liminf_{n\to\infty} E(X_n) \ge E(X)$$

(& similarly for sums and integrals)

(Note:

$$\liminf_{n\to a_n} = \lim_{n\to\infty} (\inf_{k\geq n} a_k) \quad \& \quad \limsup_{n\to\infty} a_n = \lim_{n\to\infty} (\sup_{k\geq n} a_k)$$

.)

Theorem 0.5. ***Strong Law of Large Numbers*** *(SLLN) For iid X_n,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to E(X) \qquad \textit{(if defined)}$$

almost surely, and

$$E(|\frac{1}{n}\sum_{i=1} nX_i - E(X)|) \to 0 \quad \textit{(if } E(|x|) < \infty)$$

.

Equalities & inequalities

Theorem 0.6.

- *If $X \geq 0$ then $E(X) = \int_0^\infty P(X > x)dx$, even if infinite (& valid even if $P(X = \infty) > 0$).*
- *If $X \geq 0$ and is integer-valued then*

$$E(X) = \sum_{n=0}^{\infty} P(X > n)$$

Theorem 0.7.

(i) (Markov) If $g(x) \geq 0$ & nondecreasing then

$$P(X \geq x) \leq \frac{E(g(X))}{g(x)}$$

e.g.

$$P(|X| > x) \leq \frac{E(|x|)}{x} \quad \& \quad P(|X| > x) \leq \frac{E(x^2)}{x^2}$$

.

(ii) $|E(x)| \leq E(|x|)$

(iii) *(Lyapunov)* $E(|X|^r) \leq (E(x^2)E(y^2))^{\frac{1}{2}}, \quad$ *if* $0 < r < s$.

(iv) *(Cauchy-Schwaiz)* $|E(xy)| \leq (E(x^2)(y^2))^{\frac{1}{2}}$

(v) *(Jenson) If $g(x)$ is convex then $E(g(x)) \geq g(E(x))$ with equality only if $g(x) = a + bx$ for some constants a, b.*

Note: "convex" means $g(ax + (1-a)y) \leq ag(x) + (1-a)g(y), 0 < a < 1$

Equivalently, $g(b) - g(a) = \int_a^b g'(u)du$ all a, b and $g'(u)$ is nondecreasing (the derivative need not exist at all values of u for this to hold)

Theorem 0.8. *(Kolmogorov maximal inequality)*

If X_n are independent, each $E(X_n^2) < \infty \quad \& \quad S_n = X_1 + \cdots + X_n$

then

$$P(\max_{k \leq n} |S_k - E(S_k)| > x) \leq \frac{\mathrm{Var}(S_n)}{x^2}$$

The next result takes many guises, but you should set the idea from examples shown.

Theorem 0.9. *(Fubini-tonelli)*

Suppose either that $g(x, y) \geq 0$ for all x, y or the quantities below are finite when $g(x, y)$ is replaced with $|g(x, y)|$.

(i)

$$\iint g(x, y)dxdy = \iint g(x, y)dydx$$

.

(ii)

$$\int \sum_{n=0}^{\infty} g(n,y)dy = \sum_{n=0}^{\infty} \int g(n,y)dy$$

.

(iii)

$$\int E(g(X,y))dy = E(\int g(x,y)dy)$$

.

(iv)

$$\sum_{n=0}^{\infty} E(g(X,n)) = E(\sum_{n=0}^{\infty} g(X,n))$$

.

(v)

$$\sum_{n=0}^{\infty} \sum_{k=0}^{\infty} g(n,k) = \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} g(n,k)$$

.

(vi)

$$\int\int g(x,y)F(dx)G(dy) = \int\int g(x,y)G(dy)F(dx)$$

, where the latter are lebesgue-stieltjes integrals. (When $g(x,y) \geq 0$ all x, y the double integrals/sums/etc above can equal ∞, on both sides.)

We rely heavily on conditional probabilities and conditional expectation. Like expectation, however, we want to discuss & use them without being restricted to joint discrete distributions or joint continuous distributions. We provide here, without elaboration, a suitably general definition.

Definition. 0.9. Let Y be a r.v. such that $E(Y)$ exists, and X be a random element (r.v., r. vector, stochastic process). The <u>conditional expectation</u> of Y given X, denoted $E(y|X)$ is a function of the random variable X, say $h(x)$, (Note: this also means $E(Y|X)$ is itself a r.v.) such that

$$E(g(x)Y) = E(g(x)E(Y|X)) \qquad \text{(*)}$$

for all (measurable) bounded real-valued functions $g(x)$. ("measurable" just means $\{x : g(x) \leq y\}$ is a Borel set, which is a technical restriction not usually of much concern to us.)

Conditional Probability $P(y \in A|X) = E(\mathbb{1}_A(y)|X)$.

The definition <u>does not</u> tell us how to compute $E(y|X)$; it merely says that we can define it. Note that $E(y|X)$ is <u>a random variable</u> (but it must be a function of X) . Often we may have a way to define or compute $h(x) = E(y|X = x)$ suitably and then set $E(y|X) = h(X)$.

Properties of conditional expectation

Theorem 0.10. *Given X, and r.v.'s $Y, Y_1, Y_2, \cdots$ (w/ means)*

(i) $Y_1 \leq Y_2 \implies E(Y_1|X) \leq E(Y_2|X)$.

(ii) $E(aY_1 + bY_2|X) = aE(Y_1|X) + bE(Y_2|X)$.

(iii) MCT, DCT & Fatou's lemma all hold, cond. on X.

*(iv) The formula (*) in **Theorem.0.6.** holds whenever whenever the expectations exist (not just for bounded $g(x)$).*

(v) (Pull-Out) $E(g(x)y/X) = g(X)E(y|x)$, if it exists.

(vi) (In)equalities for expectation (Jensen, etc.) all hold, cond. on X.

(vii) If Y is independent of X then $E(Y|X) = E(Y)$.

(viii) $y = h(x)$ then $E(y|x) = y$

(ix) (Tower) $E(y|x) = E(E(y|x,z)|x) = E(E(y|x)|x,z)$.

Basically, we can work with conditional expectations in an intuitive fashion based on familarity with definitions for the case (x, y) is jointly discrete or jointly absolutely cont.
But now we can do so without those restrictions, including the case X is a stochastic process (or a portion of one) consisting of infinitely many rv's.

Take care to use the tower property, (ix) above, correctly. It is not usually true that

$$E(E(y|x)|z) = E(E(y|z)|x)$$

One other caution: $E(y|x)$ is not originly defined since $P(bx = 0) = 1$ implies $E(y|x) + b(x)$ also satisfies the definition. We don't usually have a problem w/ this if there are only countably many r.v.s.

1 Introduction

- definitions
- examples (simple random walk, simple branching proc.)
- probability generating functions
- some analytic results

Definition. 1.1. A **stochastic process** is a collection of random variables, $\{X_t\}_{t \in \mathbb{T}}$, defined on the same probability space. The state space $\mathbb{X}$ is a set of possible values for X_t and the index set $\mathbb{T}$ is the set on which X_t evolves.
(a single instance of values $\{X_t\}$ is called a sample path.)
For us, $\mathbb{X}$ will usually be the integers $\mathbb{Z}$ or the real line $\mathbb{R}$ (or some other subset of $\mathbb{R}$ but it can be $\mathbb{R}^m$ or something more exotic.
$\mathbb{T}$ will usually be the nonnegative integers $\mathbb{Z}_+$ or the nonnegative real numbers $\mathbb{R}_+ = [0, \infty)$, but it also can be more general.

Example. 1.1. $X_t = \#$ of people infected on day t of an epidemic
$X_{s,t} =$ temperature at point s on the globe and time t
$X(A) = \#$ of fireant colonies in a region A (here $\mathbb{T}$ is a collection of sets, including all open & closed sets)

Stochastic processes have many applications:
e.g. signal process, random networks, queues, climate, earthquakes, insect populations, stock market, economics, epidemics, and so on.

Note: I generally use upper case for random variables and lower case for actual values.

Just as we do for a random variable & a random vector, we want to discuss the distribution of a stochastic process. But, with infinitely many rv's, we cannot simply express it as an ordinary function (such as a cdf). Instead we have

Theorem 1.1. *The distribution of a real-valued stochastic process $\{X_t\}$ (that is, how we can calculate probabilities & expectations for it) is uniquely determined by its finite-dimensional distributions*

$$F_{t_1,\cdots,t_n}(x_1,\cdots,x_n) = P(X_{t_1} \le x_1, \cdots, X_{t_n} \le x_n)$$

The idea is that any probability about the process as a whole can be computed in terms of limits involving finite-dim dist's (or prob's of finitely many rv's).
A corollary of this theorem is that, in principle, it is always possible to construct a process from its distribution in terms of functions of independent r.v's.

- ex.
$$P(\sup_{n\ge 0} X_n \le x) = \lim_{n\to\infty} P(\bigcap_{k=1}^{n}\{X_k \le x\}) = \lim_{n\to\infty} F_{1,\cdots,n}(x,\cdots,x)$$
.

-
$$P(\text{the limit until } X_t \text{ first exceeds } x \text{ is finite}) = \sum_{n=0}^{\infty} P(\max_{k<n} X_k \le x < X_n)$$
(Respectively: limit of sums w/ finitely many terms, probability for finitely many r.v's)

(Note: the concept applies to all stochastic processes, whether real-valued or not) key point The individual random variables in a stochastic process are not independent. Thus the primary problem is to model their dependence. This can be done in various ways.

- by construction fron independent random variables

  e.g. $X_t = \sum_{j=0}^{\infty} \lambda_j \epsilon_{t-j}$, where the ϵ_t's are iid.

- by conditional probabilities

  e.g. $P(X_t = j | X_{t-1} = i) = \binom{i}{j} p^j (1-p)^{i-j}, i \geq j \geq 0$

- by correlation structure (which may not be sufficient to characterize the process)

  e.g. $cov(X_t, X_{t-1}) = p^j$

Each approach offers insight into the process, so it is useful to study all three. Constructive versions of the process are also useful for simulations. Covariances are useful for statistics.

Example. 1.2. Simple Random Walk (SRW) Let S_n be a location at time n, on the integers $\mathbb{Z}$ and suppose each new location is either up one or down one, at random, form the last location. Specifically,

$$S_0 = 0$$

$$S_{n+1} = \begin{cases} S_n - 1 & \text{w.p. } 1-p \\ S_n + 1 & \text{w.p. } p \end{cases} \quad , 0 < p < 1$$

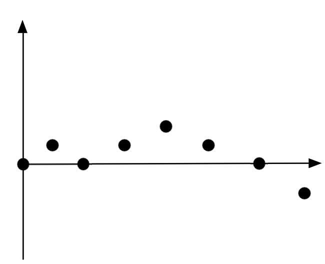

Figure 1: Simple Random Walk (SRW)

Note that, no matter what the past history of the process is, the next step depends only on the current location. We can express this by

$$P(S_{n+1} = j | S_1 = i_1, \cdots, S_n = i_n) = \begin{cases} 1-p & \text{if } j = i_n - 1 \\ p & \text{if } j = i_n + 1 \\ 0 & \text{o.w.} \end{cases}$$
$$= P(S_{n+1} = j | S_n = i_n)$$

A constructive definition is as follows: let $x_1, x_2, \cdots$ be iid r.v's with

$$P(X_n = -1) = 1 - p, \quad P(X_n = +1) = p$$

. Then define $S_n = X_1 + \cdots + X_n$. $\{S_n\}$ is a simple random walk as defined above. (Here, $S_0 = 0$.)

Example. 1.3. Simple Branching Process (SBP)

Suppose at time 0 one individual (the prognitor) is reproductive. At time 1, this individual bears a random number $Z_{1,1}$ of descendants then becomes non-reproductive. Let $\{P_k\}_{k=0}^{\infty}$ be the probability mass function (pmf) for $Z_{1,1}$, i.e.

$$P(Z_{1,1} = k) = p_k$$

At time 2, each newly born individual randomly bears offspring, independently of each other and with the same distribution as $Z_{1,1}$. Then each becomes non-reproductive. The process continues this way, with each new individual starting its own line of descent. To represent the process constructively let $Z_{i,j}, \quad i \geq 1, j \geq 1$, ne iid r.v's with pmf $\{p_k\}$, and let

$$Z_n = \text{total \# of individuals born at time } n.$$

So

$$
\begin{aligned}
Z_0 &= 1 \\
Z_1 &= Z_{1,1} \\
Z_2 &= Z_{2,1} + Z_{2,2} + \cdots + Z_{2,Z_1} \\
&\vdots \\
Z_n &= Z_{n,1} + \cdots + Z_{n,Z_{n-1}}
\end{aligned}
$$

Each Z_n is a random sum of Z_{n-1} id r.v's. (Define the sum to be 0 if $Z_{n-1} = 0$.) Note that each example relies on sums of independent r.v.' s. This is often the case for stochastic process (or approximately so) and at least makes a good starting point for our course. We review some basics about sums of independent r.v.'s.

Definition. 1.2. Let $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ be real valued sequences. Their convolution is the sequence (double $\{a_n\} * \{b_n\}$) with elements

$$\sum_{i=0}^{n} a_i b_{n-i} = \sum_{i=0}^{n} a_{n-i} b_i$$

.

Note that $\{a_n\} * \{b_n\} = \{b_n\} * \{a_n\}$.

Also, $(\{a_n\} * \{b_n\}) * \{c_n\} = \{a_n\} * (\{b_n\} * \{c_n\})$ (check)

Convolutions have an important probabilistic interpretation, given next. Note: we write $X \sim \{a_n\}$ to indicate X has pmf $\{a_n\}$.

Lemma 1.1. *Suppose X and Y are independent r.v.'s such that X has pmf $\{a_n\}$ and Y has pmf $\{b_n\}$. Then $X + Y$ has pmf $\{a_n\} * \{b_n\}$. ($n \geq 0$)*

Proof. *(By asumption both X & Y are nonnegative and integer valued.)*

The event $\{X + Y = n\}$ can be decomposed into disjoint events:

$$\{X + Y = n\} = \bigcap_{i=0}^{n} \{X = i, Y = n - i\}$$

So

$$P(X+Y=n) = \sum_{i=0}^{n} P(X=i)P(Y=n-i) \quad \textit{using independence}$$
$$= \sum_{i=0}^{n} a_i b_{n-i}$$

. *That is, the pmf for $X+Y$ is $\{a_n\} * \{b_n\}$.*

<u>Remark</u> If X and Y are independent and integer valued with pmf's $\{a_n\}_{n=-\infty}^{\infty}$, $\{b_n\}_{n=-\infty}^{\infty}$, rep. then

$$P(X+Y=n) = \sum_{i=-\infty}^{\infty} a_i b_{n-i} \quad \text{(which exists since each sequence is summable)}$$

(A probabilistic distribution is <u>absolutely continuous</u> if it has a nonnegative density $f(x)$ such that $\int_{-\infty}^{\infty} f(x)dx = 1$. [There are continuous distributions w/o densities.]) Or if X and Y are independent with absolutely continuous distributions having densities $f_X(x)$ and $f_Y(y)$, rep., then $X+Y$ has density

$$f_X * f_Y(Z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx$$

. (Note: convolution formulas are context dependent, so beware!)

<u>example</u> Suppose $X \sim Poisson(\lambda), Y \sim Poisson(\mu)$. So

$$a_n = P(X=n) = \frac{\lambda^n e^{-\lambda}}{n!}, n \geq 0, b_n = P(Y=n) = \frac{\mu_n e^{-\lambda}}{n!}, n \geq 0$$

Then,

$$P(X+Y=n) = \sum_{i=0}^{n} a_i b_{n-i} = \cdots = \frac{(\lambda+\mu)^n e^{-(\mu+\lambda)}}{n!}, n \geq 0, \text{(check)}, \quad \text{and } X+Y \sim Poisson(\lambda+\mu).$$

Note: If $X_1, \cdots, X_n$ are iid with pmf $\{p_k\}$ we will denote the pmf of $S_n = X_1 + \cdots + X_n$ by

$$\{P_k^{*n}\} = \{p_k\} * \cdots * \{p_k\} \quad \text{(n-fold convolution)}$$

. Computing convolutions can be tedious, difficult or even impossible. Alternatively, one may use a transform. (there several different kinds of transforms.)

Definition. 1.3. (a) The generating function (gf) for a sequence $\{a_n\}_{n\geq 0} a_n S^n$ is

$$A(S) = \sum_{n=0}^{\infty} a_n S^n,$$

(defined for all S such that $\sum_{n=0}^{\infty} |a_n| S^n < \infty$.)

(b) The probability generating function (pgf) for a pmf $\{p_n\}_{n\geq 0}$ is its gf:

$$P(S) = \sum_{n=0}^{\infty} p_n S^n$$

Remark If X is a nonnegative integer valued r.v. with pmf $\{P_n\}$. Then $P(s) = E(s^x), |s| \leq 1$. ($s \geq 0$ suffices here)

It often is more convenient & illuminating to use this expectation notation.

examples

- $X \sim Poisson(\lambda)$

$$E(s^x) = P(s) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda} s^k}{k!} = e^{\lambda s} e^{-\lambda} = e^{-\lambda(1-s)}$$

- $X \sim binomial(n, p)$

$$E(s^x) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} s^k = (ps + 1 - p)^n$$

- $X \sim geometric(p)$

$$E(s^x) = \sum_{n=0}^{\infty} p(1-p)^k s^k = \frac{p}{1-(1-p)s}$$

Some useful power series to know are

$$\sum_{n=k}^{\infty} z^n = \frac{z^k}{1-z} \quad \text{if } |z| < 1$$

$$\sum_{n=0}^{m} \binom{m}{n} b^{m-n} z^n = (z+b)^m$$

$$\sum_{n=0}^{\infty} \frac{1}{n!} z^n = e^z$$

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n = log(1+z) \quad \text{if } z > -1$$

An example solving a gf (rational polynomial, partial fractions)

$$A(s) = \frac{1+s+2s^2}{2-3s+s^2} = \frac{1+s+2s^2}{(2-s)(1-s)} = \frac{a+bs}{2-s} + \frac{c+ds}{1-s}$$

Solve the equation

$$(a+bs)(1-s) + (c+ds)(2-s) = 1+s+2s^2$$

We have

$$\begin{aligned} A(s) &= \frac{-11}{2-s} + \frac{6-2s}{1-s} \\ &= \frac{-11/2}{1-s/2} + \frac{6-2s}{1-s} \\ &= \frac{-11}{2} \sum_{n=0}^{\infty} (s/2)^n + (6-2s) \sum_{n=0}^{\infty} s^n \\ &= \sum_{n=0}^{\infty} [(-\frac{11}{2} 2^{-n} + 6)s^n - 2s^{n+1}] \qquad \text{(combining terms)} \\ &= \frac{1}{2} + \sum_{n=1}^{\infty} (4 - 11(2^{-n-1})) \qquad (n = 0 \text{ case handled separately}) \end{aligned}$$

Hence, $a_0 = \frac{1}{2}$ & $a_n = 4 - 11(2^{-n-1})$ for $n \geq 1$.

Note that all coefficients are nonnegative. Hence, all derivatives of $A(s)$ are nonnegative for $s \in [0, 1)$. However, $A(1) = \infty$, meaning $\sum_{n=0}^{\infty} a_n = \infty$.

recursion method

Fibonacci Sequence $f_0 = f_1 = 1$ and $f_n = f_{n-1} + f_{n-2}, n \geq 2$ let

$$\begin{aligned} F(s) &= \sum_{n=0}^{\infty} f_n s^n = 1 + s + \sum_{n=2}^{\infty} (f_{n-1} + f_{n-2}) s^n \\ &= 1 + s + s(F(s) - 1) + s^2 F(s) \\ &\implies F(s) = \frac{1}{1 - s - s^2} \end{aligned}$$

using the method on the previous page,

$$\begin{aligned} F(s) &= \frac{1}{s_2 - s_1}(\frac{1}{s - s_2} - \frac{1}{s - s_1}) = (\frac{s_1}{1 + s_1 s} - \frac{s_2}{1 + s_2 s})\frac{-1}{s_2 - s_1} \\ &= \frac{-1}{\sqrt{5}}(\sum_{n=0}^{\infty} s_1(-s_1 s)^n - \sum_{n=0}^{\infty} s_2(-s_2 s)^n) \quad \text{(geom. series)} \\ &= \sum_{n=0}^{\infty} \frac{1}{\sqrt{5}}((\frac{1+\sqrt{5}}{2})^{n+1} - (\frac{1-\sqrt{5}}{2})^{n+1}) \end{aligned}$$

Therefore, $f_n = \frac{1}{\sqrt{5}}((\frac{1+\sqrt{5}}{2})^{n+1} - (\frac{1-\sqrt{5}}{2})^{n+1})s^n$ for all $n \geq 0$. (doublecheck: $f_0 = 1, f_1 = 1$, etc.

Remark Note the relationship between a pgf and the moment generating function of X:

$$m(t) = E(e^{tx}) = P(e^t)$$

. (pgf's are relevant only for integer valued r.v's, but mgf's are more general)

Another Remark: for a doubly infinite sequence $\{a_n\}_{n=-\infty}^{\infty}$. We can have the gf $A(s) = \sum_{n=-\infty}^{\infty} a_n s^n$, as long as existence ($\sum_{-\infty}^{\infty} |a_n| s^n < \infty$) is not a problem.

We now recall some of the basic properties of gf's

Theorem 1.2. *Let* $A(s) = \sum_{n=0}^{\infty} s^n, B(s) = \sum_{n=0}^{\infty} b_n s^n$, *and assume existence* *(i) Suppose* $\sum_{n=0}^{\infty} |a_n| s_0^n < \infty$*, then for all* $s \in (-s_0, s_0)$ *[The condition allows us to bring derivative inside the sum.]*

$$\frac{d}{ds} A(s) = \sum_{n=1}^{\infty} n a_n s^{n-1}$$

.

(ii) $A(s)$ *is infinitely differentiable on* $(-s_0, s_0)$.

(iii) $\{a_n\}$ *is uniquely determined by its gf (This is a crucial point as it says there is a 1-1 correspondence between sequences & their gf's (assuming existence of the gf). This is why* $A(s)$ *is called "generating"):*

$$a_n = \frac{1}{n!} (\frac{d^n}{ds^n} A(s))_{s=0}$$

(iv) If every $a_n \geq 0$*, then* $A(s)$ *is non decreasing for* $s \geq 0$ *and all derivatives of* $A(s)$ *are non decreasing for* $s \geq 0$*.*

In particular, A(s) is <u>*convex*</u>*:*

$$\alpha A(x) + (1 - \alpha) A(y) \geq A(\alpha x + (1 - \alpha) y) \quad (0 \leq \alpha \leq 1)$$

(v) The gf for $\{a_n\} * \{b_n\}$ *is* $A(s)B(s)$*. In review of (iii),* $A(s)B(s)$ *thus can be used to identify the convolution.* $\square$

Now we want to look at how these properties are relevant to integer valued r.v.'s.

But first an extension.

Definition. 1.4. An <u>**extended**</u> r.v. is a random variable that can possibly take values $+\infty$ or $-\infty$.

In particular, suppose X is nonnegative but extended.

Then its cdf $F(x) = P(X \le x)$ satisfies $\lim_{x\to\infty} F(x) = P(X \le x)$, which could be less than 1. The "missing" probability is $P(X = \infty)$.

Why would we consider such rv's? Interestingly, they appear all too easily in stochastic process.

EX 1.2 (cont) SRW $S_n = X_1 + \cdots + X_n$ where $X_n \sim$ iid with $P(X_n = 1) = p = 1 - P(X_n = -1)$.

Let $\mathcal{T}_k = \min\{n : S_n = k\}$, the time that value k is first "hit" by the process.

The problem is that there are sequences for which k is never hit: For those cases, we naturally

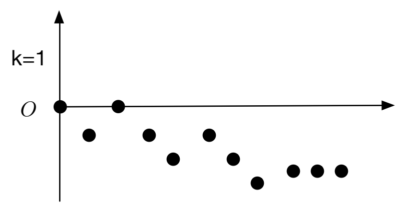

Figure 2: Simple Random Walk (SRW) never hit

define $\mathcal{T}_k = \infty$.

EX 1.3 (cont.) SBP Z_n. Recall $Z_0 = 1$. Since $Z_{n-1} = 0 \implies Z_n = 0$, we call $\mathcal{T}_0 = \min\{n : Z_n = 0\}$. The extinction time of the process. Again, this may never happen in which case $\mathcal{T}_0 = \infty$.

Batch to pdf's.

convention for pgf's: defined for $0 \le s < 1$. values at 1 are defined as limits as $s \uparrow 1$.

Theorem 1.3. *Suppose X is nonnegative integer valued and possibly extended, with pmf $P_n = P(X = n), n = 0, 1, 2, \cdots$*

(i)

$$P(1) = \sum_{n=0}^{\infty} P_n = P(X < \infty)$$

. Note: This is also $\lim_{s\uparrow 1} P(s)$. $P(X = \infty)$ *is not in the sum.* $s^\infty = 0$ *when* $0 \le s < 1$.

***Remark**:* $P(X < \infty) = 1$ *does not imply* X *is always finite.*

(ii) $P(s)$ *is infinitely differentiable, nondecreasing and all its derivatives are nondecreasing. In particular, it is convex.*

(iii) For all n*,*

$$P_n = \frac{1}{n!}(\frac{d^n}{ds^n}P(s))_{s=0}$$

. Thus P_n *is uniquely determined by its pgf.*

(iv) (Factorial Moments) Assume $P(1) = 1$.

Then

$$(\frac{d^n}{ds^n}P(s))_{s\uparrow 1} = E(X(X-1)\cdots(X-n+1)$$

. $s \uparrow 1$ *defined as limit since derivatives may not exist for* $s \ge 1$

prob. proof: (To be rigorous, we should also justify why it is okay to bring the derivative inside the expectation for this case. However, note how the prob. proof (using expectation) gives more insight than the analytic proof would.)

$$\begin{aligned}
\frac{d^n}{ds^n}P(s) &= \frac{d^n}{ds^n}E(s^X) \\
&= E(\frac{d^n}{ds^n}s^X) = E(X(X-1)\cdots(X-n+1)s^{X-n}) \\
&\overset{s\uparrow 1}{\to} E(X(X-1)\cdots(X-n+1)\mathbb{1}_{x<\infty}), \quad (\mathbb{1}_{X<\infty} = 1 \textit{ if } X < \infty, -0 \textit{ o.w.}) \\
&= E(X(X-1)\cdots(X-n+1)) \quad (\textit{since } P(1) = 1 \implies P(X<\infty) = 1)
\end{aligned}$$

analytic proof: use Thm.1.2(i)

(v) Suppose Y is also nonnegative integer valued with pdf $Q(s)$. If X and Y are independent then $X+Y$ has pgf $P(s)Q(s)$.

prob. proof: the pgf for $X+Y$ is

$$\begin{aligned} E(s^{X+Y}) &= E(s^X s^Y) \\ &= E(s^X)E(s^Y) \quad \textit{by independence} \\ &= P(S)Q(S) \end{aligned}$$

. (note how we avoided referring to the convolution.) $\square$

Corollary 1.1. If X has pgf $P(s)$ and $P(1) = 1$ then $E(x) = P'(1)$ and $Var(X) = P''(1) + P'(1) - (P'(1))^2$, (where the derivatives $P'(1)$ & $P''(1)$ are computed as limits with $s \uparrow 1$ if necessary).

Proof. *(check - revisit Thm 1.3 (iv).)*

(exercise -what is $P'(1)$ when $P(1) < 1$? It can still be finite.) $\square$

example. $X \sim Pois(\lambda).Y \sim Pois(\mu)$*, independent (both are finite w.p.1)*

We have seen

$$E(s^X) = e^{-\lambda(1-s)}, E(s^Y) = e^{-\mu(1-s)}$$

$$\frac{d}{ds}e^{-\lambda(1-s)} = \lambda e^{-lambda(1-s)} \to \lambda \quad \textit{as} \quad s \to 1. \textit{ So} \quad E(x) = \lambda$$

$$\frac{d^2}{ds^2}e{-}\lambda(1-s) = \lambda^2 e^{-\lambda(1-s)} \to \lambda^2 \quad \textit{as} \quad s \to 1. \textit{ So} \quad E(X(X-1)) = \lambda^2$$

and $Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

$X+Y$ has pgf

$$e^{-\lambda(1-s)}e^{-\mu(1-s)} \quad \textit{by Thm.1.3(v)}$$

$$= e^{-(\lambda+\mu)(1-s)}$$

$$\implies X+Y \sim Pois(\lambda+\mu) \quad \textit{by Thm.1.3 (iii)}$$

example. $X \sim bin(n,p), Y \sim bin(m,p)$, *indep.*

(use pgf's to show $E(X) = np$*,* $Var(X) = np(1-p)$ *and* $X + Y \sim bin(n+m, p)$*.)*

<u>**EX.1.2** cont.</u> SRW $S_n = X_1 + \cdots + X_n$.

The pgf for each X_μ is

$$\begin{aligned} P(s) = E(s^{X_k}) &= sP(X_k = 1) + s^{-1}P(X_k = -1) \\ &= ps + (1-p)s^{-1} \end{aligned}$$

There is no problem with existence, so we still have

$$\begin{aligned} E(s^{S_n}) &= E(S^{X_1 + \cdots + X_n}) \\ &= E(s^{X_1}) \cdots E(s^{X_n}) \quad \text{by independence} \\ &= (P(s))^n = (ps + (1-p)s^{-1})^n \end{aligned}$$

. Observe that this is

$$\sum_{k=-n}^{n} P(S_n = k)s^k \quad \text{, which includes both pos. \& neg. k.}$$

(Note: $\frac{S_n + n}{2} \sim bin(n,p)$ - check; there are several ways you can do this.)

Theorem 1.4. *Suppose* $X_1, X_2, \cdots$ *is a sequence of r.v.'s and* N *is a nonnegative integer valued r.v., independent of the sequence. The r.v.* $S_N = X_1 + \cdots + X_n$ *(*$= 0$ *if* $N = 0$ *by convention) is called a <u>random sum</u> of the sequence.* □

We have already seen an example of random sums in our construction of the simple branching process. (Ex.1.3).

Theorem 1.5. *Suppose* $X_1, X_2, \cdots$ *are iid with pgf* $P(s) = \sum_{k=0}^{\infty} P_k s^k$ *and* N *has pgf* $A(s) = \sum_{n=0}^{\infty} a_n s^n$*,* N *independent of* $\{X_n\}$*. Then* $S_N = X_1 + \cdots + X_N$ *has pgf* $A(P(s))$*,* $0 \le s \le 1$*.*

prob. proof. (by and expectation) Assume $0 \le s \le 1$.

Recall that the pgf of S_n of S_n is $E(s^{S_n}) = (P(s))^n$.

This even makes sense if $n = 0$ (since $S_0 = 0$ by convention). Thus $E(s^{S_N}|N) = (P(s))^N$ (note: this is a function of N)

(**Important**: Regardless of the type of rv's, we can think of $E(X|Y)$ as the function of Y such that

$$E(h(Y)X) = E(h(Y)E(X|Y)$$

, whenever the expectations make sense.)

By the law of iterated expectation

$$\begin{aligned} E(s^{S_N}) &= E(E(s^{S_N}|N)) \\ &= E((P(s))^N) = A(P(s)) \end{aligned}$$

(take limit as $s \uparrow 1$ to get case $s = 1$.)

Be sure to learn both these methods (↑↓)

prob. proof (by partitioning the sample space)

Using the indicator functions,

$$\begin{aligned} E(s^{S_N}) &= E(\sum_{n=0}^{\infty} s^{S_n} \mathbb{1}_{N=n}) \\ &= \sum_{n=0}^{\infty} E(s^{S_n}) E(\mathbb{1}_{N=n}) \quad \text{(by independence of } S_n \in N) \\ &= \sum_{n=\infty}^{\infty} (P(s))^n a_n \qquad \text{(since } E(\mathbb{1}_{N=n} = P(N=n) = a_n) \\ &= A(P(s)) \qquad \square \end{aligned}$$

Refer to Resnick for an analytical proof involving the pmf of S_N, given by

$$P(S_n = j) = \sum_{n=0}^{\infty} P_j^{*n} a_n$$

, where $P_j^{*n} = P(S_n = j)$ (n-fold convolution) and

$$P_j^{*0} = P(S_0 = j) = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{o.w.} \end{cases}$$

Corollary 1.2. Assume as in Thm.1.5. If $E(|X_1|) < \infty$ and $P(N < \infty) = 1$ then

$$E(S_N) = E(N)E(X_1) \quad \text{(this is valid even if } E(N) = \infty)$$

. If also $E(X_1^2) < \infty$ and $E(N^2) < \infty$ then

$$Var(S_N) = E(N)Var(X_1) + Var(N)(E(X_1))^2$$

. (Helpful result: variance partition formula $var(X) = E(var(X|Y)) + var(E(X|Y))$.)

Proof. *(exercise - use the pgf from Thm1.11, or compute directly using either of the methods in the proof of Thm.1.5.)* □

EX.1..3 (cont.) Simple Branching Process

Let Z_{ij} be iid with pmf $\{p_k\}$ and pgf $P(s)$.

Recall that $Z_0 = 1, Z_1 = Z_{1,1}, \cdots, Z_n = Z_{n,1} + \cdots + Z_{n,Z_{n-1}}, \cdots$

Note that Z_{n-1} depends only on Z_{ij}'s s.t. $i \le n-1$ and thus Z_{n-1} is independent of $Z_{n,j}$'s.

Therefore Z_n is a random sum of $Z_{n,j}$'s (w/ $N = Z_{n-1}$) and so its pgf is $P_n(s) = E(s^{Z_n})$ where $P_1(s) = P(s)$ and

$$\begin{aligned} P_n(s) &= P_{n-1}(P(s)) \\ &= \cdots = P(P(\cdots P(s)\cdots)). \quad \text{(n-fold composition)} \end{aligned}$$

(interpret $P_n(s) = P(P_{n-1}(s))$: think of z_1 independent process starting at time 1.) [Many problems can be solved by "conditioning". There may be several ways to do this. (In this situation, one cond. on Z_{n-1}, the other cond. on Z_1)]

example. *Let* $P(s) = 1 - p + ps$ *(pgf of* $Bernoulli(p)$*)*

Thus, at each generation the (sole) individual either generates one offspring, w.p. p, or none at all (which stops the process).

Thinking prpbabilistically: $Z_n = 1$ *iff* n *independent successes (births) have occurred (w.p* p^n*)*

$Z_n = 0$ *o.w. (w.p.* $1 - p^n$*)*

Thinking analytically: We claim $P_n(s) = 1 - p^n + p^n s$

Proof by induction: $P_1(s) = P(s) = 1 - p + ps$

If we claim holds for P_n *then*