DISCRETE STOCHASTIC PROCESSES

Lecture 1

Review probability models and Random Variables.

Define discrete stochastic processes.

Discuss the philosophy of modeling and the objectives of this course.

Present a clever result on expectations.

Sums of IID random variables.

Convergence in mean square

Central limit theorem.

Markov and Chebyshev inequalities

Moment Generating Functions and the Chernoff Bound

Probability

A Stochastic Process is a special case of a probability model/experiment

Probability Model (Quick, very basic review):

Sample space, Ω (set of possible "elementary outcomes" $\omega \in \Omega$)

Events (subsets of sample space) Σ

Probability associated with each event, $P : \Sigma \longrightarrow R$

"Experiment" results in one & only one sample point

Probability of the union of mutually exclusive events (disjoint sets) is the sum of the probabilities of the events.

EXAMPLE:

Digital Cellular Radio receives string of 4 binary digits; errors are made independently with probability p.

Experiment records particular string of correct (0) bits and bits in error (1).

16 sample points: 0000, 0001, 0010, 0011, ...

P(0000) = $(1-p)^4$, P(1110) = $p^3(1-p)$, ...

Example of event: first bit correct, second in error. This event is the union of all sample points starting with 01; probability of this event is $(1-p)p$.

Discrete Probability Model—Finite or Countable Sample Space

In general, for finite or countable sample space, each sample point a_i has probability $P(a_i)$ and for each event E, $P(E) = \sum_{a_i \in E} P(a_i)$.

Continuous Probability Models—Uncountable Sample Space

Assigning probabilities to sample points doesn't always work:

Ex. 1 Uniform distribution over real numbers 0 to 1 – each number has 0 probability;

Ex. 2 Unending binary sequence $X_1, X_2, \ldots$ where each bit is 1 with probability p; bits are statistically independent -- each sequence has 0 probability .

We must work with events (intervals, unions of intervals, ...) directly. Requirements on limits of series of events is what makes formal probability theory (measure theory) tricky.

Possible events for Ex. 2.

Event that first hundred bits are all 0.

Event that $\lim_{n\to\infty} \frac{X_1 + X_2 + \ldots + X_n}{n} = p$ (all sequences of {0,1} that satisfy equation)

Stochastic Processes and Discrete Stochastic Processes

A **stochastic process** is a probability model in which each sample point (i.e. elementary outcome) is actually a function of "time".

The above binary sequence is a stochastic process over discrete (integer) time.

For any probabilistic model representing the times at which customers or thigamajigs arrive, one can take the number of arrivals up to t (as a function of t) as a sample function.

The probabilistic model must in principle specify the probability of each event (e.g. having 57 arrivals between time 10 and 100 is an event).

A **discrete stochastic process** is one in which the sample functions change by discrete amounts (discontinuously) as functions of time.

(You may be used to Gaussian process where functions are continuous in time.)

In discrete stochastic processes, we concentrate on when the function changes.

In this course we will study the following types of discrete random processes:

- **Poisson Processes** (continuous time, discrete values)
- **Finite State Markov Chains** (discrete time, discrete values)
- **Renewal Processes** (continuous and discrete time, discrete values)
- **Markov Chains with Countably Infinite State Spaces** (discrete time, discrete values)
- **Semi-Markov Processes** (discrete and continuous time, discrete values)
- **Continuous-Time Markov Processes** (continuous time, discrete values)

We also study, at the end

- **Random Walks** (take on discrete or continuous set of values, discrete time)
- **Martingales** (take on discrete or continuous set of values, discrete time)

Models

It is important in all fields to separate models from reality. **In probability, it is particularly important, since only one sample point is "real"**. Models compromise between simplicity and fidelity, and often many models are used to study one issue.

Given a model, all questions about model are "math" and thus have precise answers.

Engineers & Operations Researchers are concerned with finding the right model.

Applied mathematicians concerned with analyzing the model.

Structural properties (what the model really says) should be of concern to all.

Building a model for given applications requires detailed knowledge about the real problem (both the physical basis and the intended use) and an understanding of classes of models that are tractable.

There are usually iterations between modeling and analysis (this part often involves statistics).

What this course is about

These comments on modeling are true in particular for stochastic processes.

The purpose of this course is to build a conceptual understanding of simple tractable classes of models.

This is best done by focusing on the structure of the models, rather than on applications.

The exercises are to help understand the structure of the model, not to turn you into computing machines.

You will also learn a *great deal* of applied probability theory (but not measure theory) along the way.

Our focus on insight and conceptual understanding might be frustrating at first –

For the first year graduate student since it breaks undergraduate mold of "plug and chug" or "simulating on a computer"

For the engineer since it breaks focus on bottom line-itis and requires conceptual rather than computational approaches

For the mathematician since insight is more important than generality or formalism.

All models are wrong, but some models are useful !

Some Application areas for Discrete Stochastic Processes

Operations Research

Queueing in any area

Failures in manufacturing systems

Finance

Risk modelling

Network models

Biology and Medicine

Epidemiology

Genetics and DNA studies

Cell modelling

Bioinformatics

Medical screening

Neurophysiology

Computer Systems

Communication networks

Data Compression

Job flow in computer systems

Intelligent control systems

Detection of signals

Physics – statistical mechanics

RANDOM VARIABLES

A random variable is neither random nor a variable. (Discuss amongst yourselves.)

A **random variable** X in a probability model is a function from the sample space to the (finite) real numbers.

$P_X(x)$ = **P(X=**x) = $\sum_{a \in \Omega \text{ such that } X(a) = x} P(a)$ for countable sample space Ω;

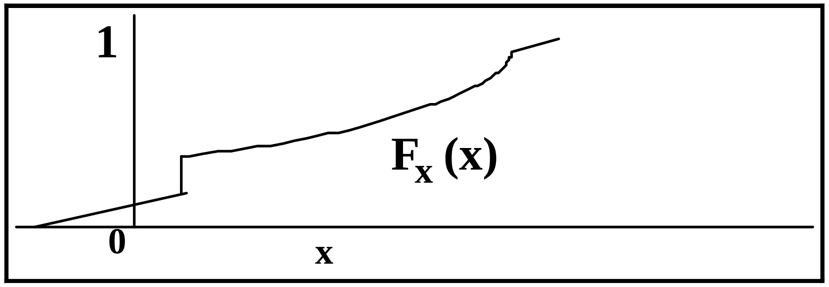
$P_X(x)$ called the **Probability Mass Function** of *X*. In general, $P(X \le x) = P(\{a : X(a) \le x\})$.
$\sum P(a)$

$F_X(x) = P(X \le x)$ is called the **Distribution Function** of *X*; for given r.v. *X*, it is a function of a real variable x.

If it exists, $f_X(x) = dF_X(x)/dx$ is called the **density** of *X* and *X* is called a **continuous random variable**.

If *X* takes on only a countable set of values, it is called **discrete**.

The Distribution Function



$F_X(x)$ is **monotonic non-decreasing** and always exists for all random variables.

What does the height and placement of a jump tell you?

The density function and the PMF are usually more convenient for hand calculations and thus more familiar in elementary subjects; the distribution function is more useful conceptually since discrete, continuous, and arbitrary random variables are all described by probability distribution functions.

If $X, Y, Z, \ldots$ are random variables in a probability model with sample space *S*, then the **Joint Distribution Function** is:

$$F_{XYZ}(x, y, z) = P(X \le x, Y \le y, Z \le z) = P(\{a \in \Omega : X(a) \le x, Y(a) \le y, Z(a) \le z\})$$

The joint density (if it exists) is $f(x,y,z) = d^3F(x,y,z)/dxdydz$.

X,Y,Z are (statistically) independent if $F_{XYZ}(x,y,z) = F_X(x)F_Y(y)F_Z(z)$ for all x,y,z.

Pairwise independence does **not** imply independence:

EXPECTATIONS

The expectation of X is $E[X] = \sum_x xP_X(x)$ for a discrete random variable X

$$E[X] = \int_{-\infty}^{\infty} xf_X(x)dx \quad \text{for continuous variables}$$

$$E[X] = \int_{-\infty}^{\infty} xdF_X(x) \quad \text{in general}$$

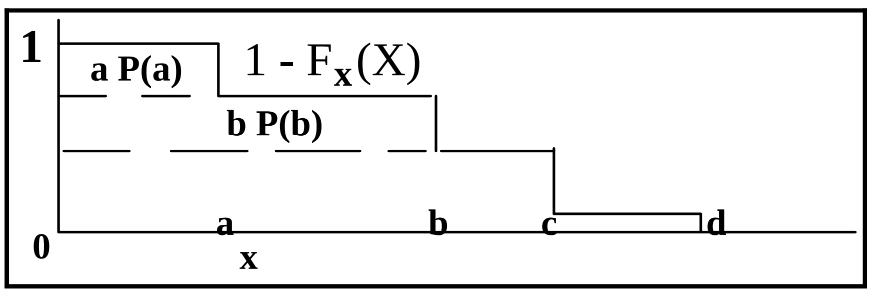
This is a Stieltjes integral and can be regarded as shorthand for any sensible way of integrating (view $dF_X(x)$ as $f_X(x)dx$ with the density function including impulses, etc). $E[X]$ exists iff $E[|X|]$ exists.

Stieltjes Integral

$$\int_{-\infty}^{\infty} g(x)dF(x) = \lim_{\delta\to 0}\sum_{n=-\infty}^{\infty} \sup_{n\delta<x\le n\delta+\delta} g(x)\left[F(n\delta+\delta) - F(n\delta)\right] = \lim_{\delta\to 0}\sum_{n=-\infty}^{\infty} \inf_{n\delta<x\le n\delta+\delta} g(x)\left[F(n\delta+\delta) - F\right.$$

For **non-negative random variables**, a convenient way to calculate expectation (both for theoretical and practical applications) is

$$E[X] = \int_0^{\infty}\left[1 - F_X(x)\right]dx$$



This thinks about an integral as a limit of horizontal strips rather than a limit of vertical strips. (The same way of thinking is used when defining the Lebesgue integral instead of the Reimann integral in a real analysis course.)

FUNCTIONS OF RANDOM VARIABLES

If X is a random variable (rv), $(X(\omega)$ is a function from sample space Ω to reals), and $g(x)$ is a function (reals to reals), then $g(X)$ is a rv mapping each sample point $\omega \in \Omega$ to $g(X(w))$.

$$E[g(X)] = \int y \, dF_{g(X)}(y) = \int g(x) \, dF_X(x)$$

Examples:

$$E[X^2] = \int x^2 dF_X(x) \quad \text{(second moment)}$$

$$\text{VAR}(X) = E\left[(X - E[X])^2\right] = E[X^2] - (E[X])^2$$

$$E[X^n] = \int x^n dF_X(x) \quad \text{(n}^{\text{th}} \text{ moment)}$$

$$E[e^{rX}] = \int e^{rx} dF_X(x) \quad \text{(moment generating function)}$$

Recall that if X, Y are **independent (**or just **uncorrelated)** rv's, then

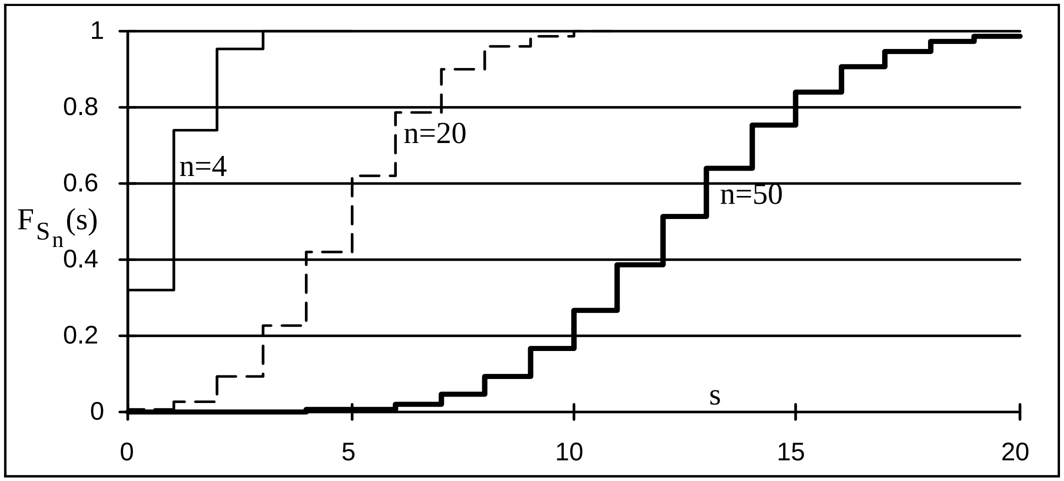
$$\text{VAR}(X + Y) = \text{VAR}(X) + \text{VAR}(Y).$$

Important Properties of the Sum of IID Random Variables

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (IID)

Define the sum $S_n = X_1 + X_2 + \ldots + X_n$, then

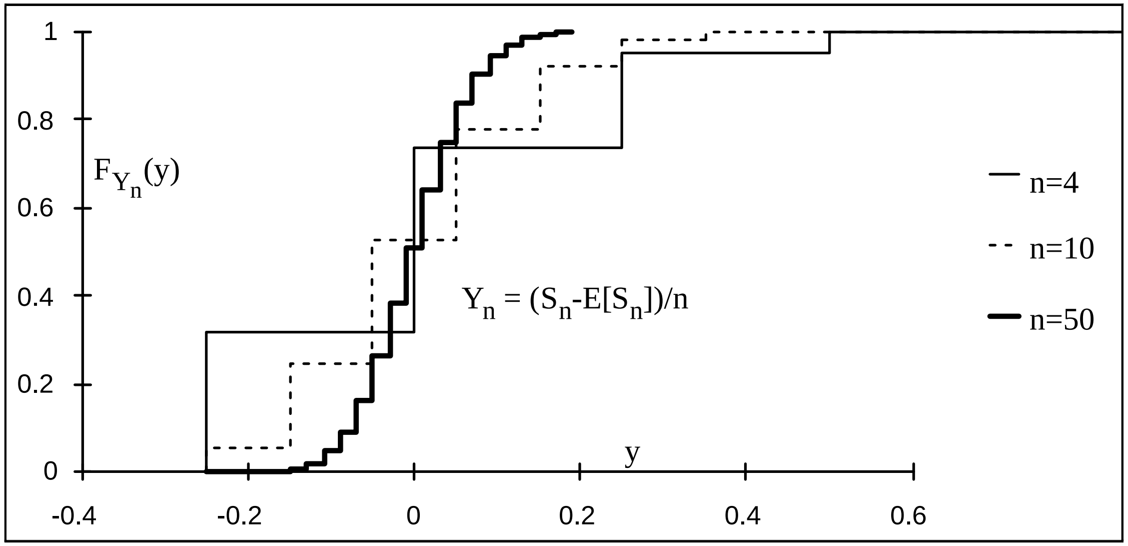
$$\mathrm{VAR}(S_n) = n\,\mathrm{VAR}(X_1); \quad \sigma_{S_n} = \sqrt{n}\,\sigma_{x_1}$$



X_i is binary, $P_X(0) = 3/4$, $P_X(1) = 1/4$. (There are unseen steps near the top.)

Sample Averages (Discrete Time Averages)

$$VAR\left(\frac{S_n}{n}\right) = E\left[\left(\frac{S_n - \overline{S}_n}{n}\right)^2\right] = \frac{VAR(S_n)}{n^2} = \frac{VAR(X_n)}{n}$$

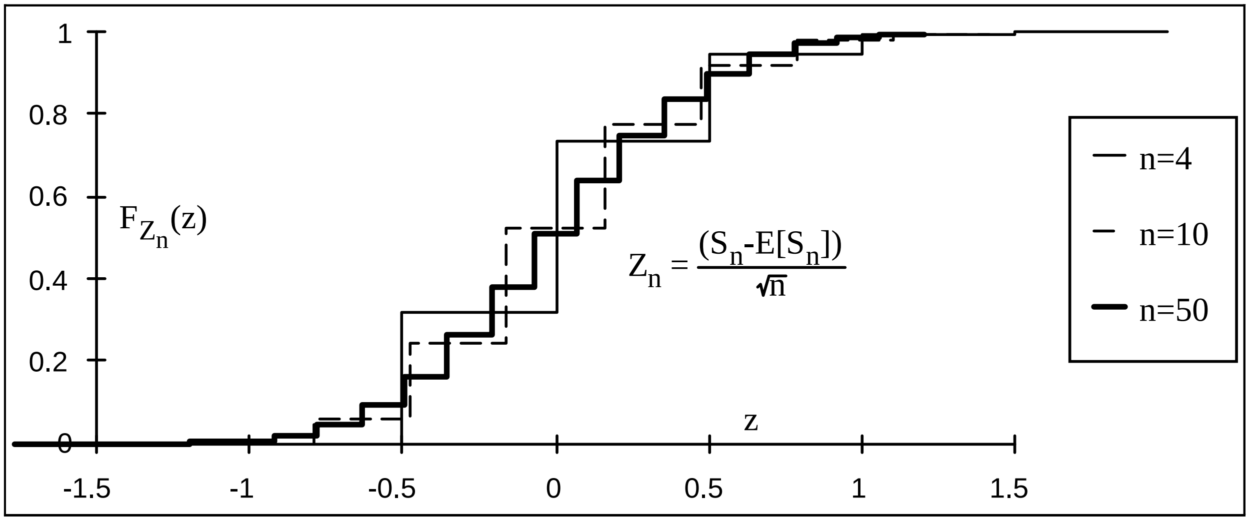


$$\lim_{n\to\infty} VAR(S_n / n) = 0, \qquad \lim_{n\to\infty} E\left\{\left(\frac{S_n}{n} - E\left(\frac{S_n}{n}\right)\right)^2\right\} = \lim_{n\to\infty} E\left\{\left(\frac{S_n}{n} - E(X_i)\right)^2\right\} = 0$$

We say that the sample average, S_n / n, approaches $E[X_i]$ **in the mean square**.

Central Limit Theorem

$$VAR\left(\frac{S_n}{\sqrt{n}}\right) = E\left[\left(\frac{S_n - \overline{S}_n}{\sqrt{n}}\right)^2\right] = \frac{VAR(S_n)}{n} = VAR(X_i)$$



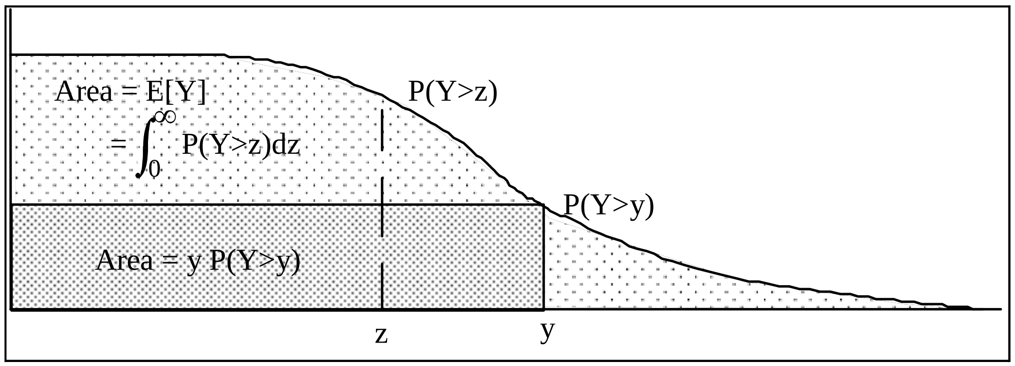
The distribution function of Z_n approaches a Gaussian **distribution** with mean 0 and variance $\mathrm{VAR}(X_i)$ (Central Limit Theorem).

For a given n, percentage error of Gaussian approximation is likely to be large when working with the distribution's tails.

Densities don't converge. Density of Z_n doesn't exist. (It's all impulses.)

Two Useful Bounds: The Markov and Chebyshev Inequalities

For any non-negative *rv* Y, the **Markov inequality** says that $P(Y \geq y) \leq E[Y]/y$ for any $y > 0$.



Let Z be any *rv* with a finite variance, and let $Y = (Z - E[Z])^2$

$$P\left((Z - E[Z])^2 \geq y\right) \leq \frac{\sigma_z^2}{y};\ P\left(|Z - E[Z]| \geq \varepsilon\right) \leq \frac{\sigma_z^2}{\varepsilon^2}$$

This is the **Chebyshev inequality**.

<u>Transforms</u>

The ***moment generating function*** for a random variable X is

$$g_X(r) = E[e^{rX}] = \int_{-\infty}^{\infty} e^{rx} dF_X(x)$$

If r is real, this may exist only for a certain range of values of r.
If r is pure imaginary, it always exists.

<u>Useful property #1:</u> Finding Central Moments

If $g_X(r)$ converges in a range around zero, then

$$\left.\frac{\partial^n g_X(r)}{\partial r^n}\right|_{r=0} = \left.E[\frac{\partial^n}{\partial r^n} e^{rX}]\right|_{r=0} = \left.E[X^n e^{rX}]\right|_{r=0} = E[X^n]$$

Transforms

Useful property #2: Sums of Independent random variables.

If $X_1, \cdots, X_n$ are independent and S = $X_1 + \cdots + X_n$, then

$$g_S(r) = E[e^{rS}] = E[e^{\sum_{k=1}^{n} rX_k}] = E[\prod_{k=1}^{n} e^{rX_k}] =$$

(by independence)

$$\prod_{k=1}^{n} E[e^{rX_k}] = \prod_{k=1}^{n} g_{X_k}(r)$$

If furthermore $X_1, \cdots X_n$ are iid, then

$$g_S(r) = [g_X(r)]^n$$

Transforms

Simple Example:

$X_1, \cdots, X_n$ iid, S = $X_1 + \cdots + X_n$, $n \geq 2$,

$$E[S^2] = \frac{\partial^2}{\partial r^2} g_S(r)\Big|_{r=0} = \frac{\partial^2}{\partial r^2}[g_X(r)]^n\Big|_{r=0} =$$

$$n\frac{\partial}{\partial r}\{g_X^{n-1}(r)g_X{}'(r)\Big|_{r=0} = n(n-1)g_X^{n-2}(0)(g_X{}'(0))^2 + ng_X^{n-1}(0)g''_X(0) =$$

$$\left(\text{ since } g_X(0) = 1,\ g'_X(0) = E[X],\ g''_X(0) = E[X^2]\ \right)$$

$$n(n-1)(E[X])^2 + nE[X^2].$$

A Third Useful Bound: The Chernoff Bound

Let X be a random variable with a moment generating function.
Apply the Markov bound to the non-negative r.v. e^{rX} (with real r):

For any real r where MGF converges,

$$P(e^{rX} \geq y) \leq \frac{E[e^{rX}]}{y} = \frac{g_X(r)}{y}.$$

Let $y = e^{ra}$ for any real a. Then we have two exponential bounds:

For all r > 0 where MGF converges, $P(X \geq a) \leq g_X(r)e^{-ra}$

For all r < 0 where MGF converges, $P(X \leq a) \leq g_X(r)e^{-ra}$

Must optimize over r to find best bounds.