CHAPTER 3 (DISCRETE RANDOM VARS) AND CHAPTER 4 (CONT RAND VAR)

So much similarity b/w continuous and discrete makes sense to do simultaneously.

SECTION 3.1: PROBABILITY MASS FN; SECTION 4.1: PROB DENSITY FNS

Defn: Prob mass function of a discrete random variable X is a fn $f: \mathbb{R} \to [0,1]$ given by $f(x) = \mathbb{P}(X = x)$

Prob density function of a cont random variable X is the f st $F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)\,du$

↳ note density fn not unique (only unique a.e.)

↳ if continuous, $\mathbb{P}(X = x) = 0$ for all x!

LEMMA: Standard properties

- Discrete: $F(x) = \sum_{x_i \le x} f(x_i)$, $f(x) = F(x) - \lim_{y \to x^-} F(y)$

  $\{x : f(x) \neq 0\}$ is at most countable

  $\sum_i f(x_i) = 1$ when $\{x_1, x_2, \dots\}$ where f non-zero

- Continuous: $\int_{-\infty}^{\infty} f(x)\,dx = 1$

  $\mathbb{P}(X = x) = 0 \quad \forall x \in \mathbb{R}$

  $\mathbb{P}(a \le X \le b) = \int_a^b f(x)\,dx$

Examples:

- Binomial $\quad Bin(n,p)$: $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$
- Poisson $\quad Pois(\lambda)$: $f(k) = \lambda^k e^{-\lambda}/k!$

  ↳ say a bit on e^x: $e^{x+y} = e^x e^y$ involves combinatorics!
- Uniform $\quad Unif(a,b)$ $\quad f(x) = \frac{1}{b-a}$ if $a \le x \le b$, 0 otherwise

HW: do (3.1) #1ac (hint: famous sum), #3; (4.1) #1b

suggested: (3.1) #5, (4.1) #1a, #4

Section 3.2 and 4.2: Independence

Events A and B independent if knowledge of one happening does not affect knowledge of other happening: $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Defn: <u>discrete</u>: X and Y <u>independent</u> if events $\{X = x\}$ and $\{Y = y\}$ are independent for all x, y

<u>Continuous</u>: X and Y independent if events $\{X \le x\}$ and $\{Y \le y\}$ are independent for all x, y.

<u>Example</u>: Toss coin with prob p of heads, $1-p$ of tails, N times. Let X = # heads Y = # tails. Shouldn't be indep as $X + Y = N$; in fact, $\mathbb{P}(X = Y = N) = 0 \neq p^N (1-p)^N = \mathbb{P}(X = N)\,\mathbb{P}(Y = N)$

<u>Example</u>: Same as above, but now N is $\mathrm{Pois}(\lambda)$. Now X and Y indep: still sum to N, but N is <u>not</u> fixed, so knowing X doesn't give Y.

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x, Y = y \mid N = x+y)\,\mathbb{P}(N = x+y)$$

$$= \binom{x+y}{x} p^x (1-p)^y \cdot \frac{\lambda^{x+y} e^{-\lambda}}{(x+y)!}$$

$$= \frac{(\lambda p)^x (\lambda p)^y}{x!\;y!} e^{-\lambda}$$

$$\mathbb{P}(X = x) = \sum_{n \ge x} \mathbb{P}(X = x \mid N = n)\,\mathbb{P}(N = n)$$

$$= \sum_{n \ge x} \binom{n}{x} p^x (1-p)^{n-x} \frac{\lambda^n e^{-\lambda}}{n!} = \frac{(\lambda p)^x e^{-\lambda p}}{x!}$$

$\hookrightarrow$ algebra: $\binom{n}{x} = n! / x!\,(n-x)!$

$$\text{Get } \underbrace{\left(\sum_{n \ge x} \frac{(\lambda(1-p))^{n-x}}{(n-x)!}\right)}_{\text{This is } e^{\lambda(1-p)}} \frac{(\lambda p)^x e^{-\lambda \cancel{p}}}{x!}$$

and $e^{\lambda(1-p)} e^{-\lambda} = e^{-\lambda p}$

SECTIONS 3.2 AND 4.2 CONT

LEMMA: Let $g, h : \mathbb{R} \to \mathbb{R}$ and X, Y independent. Then $g(X)$ and $g(Y)$ are independent.

↳ Proof is basically real analysis (properties of $\mathcal{F}$ and σ-algebras)

HW: do: (3.2) : #1, #4 (4.2): #1 (also do when F uniform on [0,1] and $K = .9$)

suggested (3.2) : #2, #5 (4.2): #2

SECTIONS 3.3 ~~and 3.4~~: EXPECTATION

Note: long section; also doing Differentiating identities

Mean value (also called expectation or expected value):

$$\mathbb{E}[X] = \sum_{x : f(x) > 0} x f(x) \quad \text{or} \quad \int_{-\infty}^{\infty} x f(x)\, dx \quad \text{if converges absolutely}$$

↳ to refer to both at same time, write $\int_{-\infty}^{\infty} x f(x) dx$

↳ $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$ (if converges absolutely)

Caveats: $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ (Cauchy Distribution)

↳ does it have a mean? Is symmetric...

note $\lim_{A \to \infty} \int_{-A}^{A} \frac{dx}{\pi(1+x^2)} \neq \lim_{A \to \infty} \int_{-A}^{2A} \frac{dx}{\pi(1+x^2)}$

↳ other example: $f(x) = \begin{cases} 1/x^2 & x \geq 1 \\ 0 & \text{otherwise} \end{cases}$ has no mean

Think of as average value:

↳ toss fair coin n times: $E[X] = \sum_{k=0}^{n} k \cdot \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k}$

↳ This should be $n/2$.... How to see this? If not fair replace with $p^k (1-p)^{n-k}$

What do we expect?

Sections 3.3 and 4.3: Expectation (continued)

Lemma: Expectation $\mathbb{E}$ satisfies
- if $X \geq 0$, $\mathbb{E}[X] \geq 0$
- $a, b \in \mathbb{R} \Rightarrow \mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
  - ↳ This means $\mathbb{E}$ is a linear operator
- $\mathbb{E}[\mathbb{1}] = 1$, where $\mathbb{1}$ is the random var always $= 1$

Proof: $\mathbb{E}[aX + bY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax+by)\, f_X(x)\, f_Y(y)\, dx\, dy$

$$= a\int_{-\infty}^{\infty} x\, f_X(x)\left[\int_{-\infty}^{\infty} f_Y(y)\, dy\right]dx + b\int_{-\infty}^{\infty} y\, f_Y(y)\left[\int_{-\infty}^{\infty} f_X(x)\, dx\right]dy$$

$$= a\int_{-\infty}^{\infty} x\, f_X(x)\cdot 1\, dx + b\int_{-\infty}^{\infty} y\, f_Y(y)\cdot 1\, dy$$

$$= a\mathbb{E}[X] + b\mathbb{E}[Y]$$

↳ must use Fubini-Tonelli (Multi Var Calc) to justify interchange

need $\iint |ax+by|\, f_X(x)\, f_Y(y)\, dx\, dy$ finite

↳ use $|ax+by| \leq |a|\cdot|x| + |b|\cdot|y|$

use linearity of integral

by assumption $\int |x|\, f_X(x)\, dx$ and $\int |y|\, f_Y(y)\, dy$ finite

↳ Now see why require $\int_{-\infty}^{\infty} |x|\, f_X(x)\, dx$ to be finite and not just $\int_{-\infty}^{\infty} x\, f(x)\, dx$ to conditionally exist — allows us to interchange orders of integration.

Caveat: Cannot always interchange; nice example with $\sum_m \sum_n \neq \sum_n \sum_m$

0	0	-1	…
0	-1	+1	…
-1	+1	0	…
+1	0	0	→

each column sums to 0, sum columns get 0

all rows sum to zero but first, which sums to 1; sum rows get 1

Problem: $\sum_m \sum_n |a_{mn}| = +\infty$

Sections 3.3 and 3.4: Continued (Expectation)

Example: Biased coin: $E[X] = \sum_{k=0}^{n} k\binom{n}{k} p^k(1-p)^{n-k}$

Soln: Let X_k be the random variable which is 1 if the k^{th} toss is heads, 0 if k^{th} toss is tails.

↳ Note $E[X_k] = 1\cdot p + 0\cdot(1-p) = p$

↳ Clearly the X_k's independent, $X = X_1 + \cdots + X_n$

Thus $E[X] = \sum_{k=1}^{n} E[X_k] = \sum_{k=1}^{n} p = np$

↳ note how much simpler this is!

New Technique: Differentiating Identities

Identities bread and butter of math; if can generate more identities from a given identity, that's good!

↳ Example: Binomial Thm

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Take $x\,d/dx$ of both sides (finite sum so okay)

↳ $n\,x(x+y)^{n-1} = \sum_{k=0}^{n} k\binom{n}{k} x^k y^{n-k}$

↳ NOW set $x = p$, $y = 1-p$ (essentially x and y were indep vars when differentiate)

↳ yields $\sum_{k=0}^{n} k\binom{n}{k} p^k(1-p)^{n-k} = np(p+(1-p))^{n-1}$

$= np$

↳ recover answer!

SECTIONS 3.3 and 4.3: EXPECTATION (CONTINUED)

More on differentiating identities

What is $\sum_{n=0}^{\infty} n \cdot \left(\frac{1}{2}\right)^n$?

↳ arises in geometric random variables: prob first head occurs after n tosses of a fair coin is $(1/2)^n$

↳ arises in heuristics for $3x+1$ problem (more on this later)

Consider $\sum_{n=0}^{\infty} x^n = (1-x)^{-1}$ if $|x| < 1$

↳ This is the geometric series formula: will prove later

↳ Apply $x \frac{d}{dx}$: $\sum_{n=0}^{\infty} n x^n = \frac{x}{(1-x)^2}$ $\left(x = \frac{1}{2} \Rightarrow 2\right)$

CHALLENGE PROBLEM: Justify $\frac{d}{dx} \sum = \sum \frac{d}{dx}$;
why can we interchange orders?

Proof of Geometric Series formula:

① $S_N = 1 + x + \cdots + x^N$

$x S_N = \phantom{1+{}} x + \cdots + x^N + x^{N+1}$ subtract

$(1-x) S_N = 1 - x^{N+1} \Longrightarrow S_N = \frac{1-x^{N+1}}{1-x}$

↳ if $|x| < 1$, $\lim_{N \to \infty} S_N = \frac{1}{1-x}$ and this is our sum

② Playing hoops: Alice + Bob shoot, first to make basket wins, Alice gets basket with prob p, Bob with prob q, let $x = (1-p)(1-q)$

Let $W = \text{Prob(Alice wins)}$

Note $W = p + (1-p)(1-q)p + [(1-p)(1-q)]^2 p + \cdots = p \sum_{n=0}^{\infty} x^n$

Also $W = p + (1-p)(1-q) W$ as if both miss, like start over

Thus $W = p + xW$ ~~and $\sum_{n=0}^{\infty} x^n$~~ so

SECTIONS 3.3 AND 3.4: EXPECTATION (CONT)

Proof of Geometric Series Formula

(2 cont) $\omega = p + x\omega \Rightarrow (1-x)\omega = p$ or $\omega = p/(1-x)$

as $\omega = p\sum_{n=0}^{\infty} x^n$ as well, $\Rightarrow \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$

CHALLENGE PROBLEM: The above only works if $0 \le x < 1$: Can you extend to $-1 < x < 1$? What about complex x?

3X+1 Problem (Aside)

$a_{n+1} = \frac{3a_n + 1}{2^k}$ where k is highest power of 2 dividing $3a_n+1$

Ex: $11 \underset{1}{\to} 17 \underset{2}{\to} 13 \underset{3}{\to} 5 \underset{4}{\to} 1 \circlearrowleft$

Conj: Any odd positive integer eventually iterates to $\cdots 1 \circlearrowleft$

↳ See papers by Lagarias for results/description/bibliography

Quotes: Kakutani: Soviet conspiracy to slow down American mathematics

Erdös: Mathematics not yet ready for such questions

Structure Thm (Sinai - Kontorovich)

~~m-path~~ study all seeds such that $(k_1, \ldots, k_m)$ describes how many powers of 2 remove in i^{th} step. Call this the m-path, each k_i is a geometric random variable with parameter 2, all independent.

Heuristic: $b_n = \log a_{n+1}$, $a_{n+1} \approx 3a_n/2^k$

$$\mathbb{E}[\log a_{n+1}] = \sum_{k=1}^{\infty} \frac{1}{2^k} \log\left(\frac{3a_n}{2^k}\right) = \log a_n + \log 3 - \log 2 \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{2}\right)^k$$

$$= \log a_n + \log(3/4) \text{ by differentiating identities}$$

SECTIONS 3.3 and 3.4: INDEPENDENCE (CONT.)

3x+1 Problem (Aside) (Continued)

Get $E[\log a_{n+1}] = \log a_n + \log(3/4)$, $\log(3/4) < 0$

↳ expect to decay, geometric Brownian motion (geometric as logarithm of a_n is Brownian motion).

↳ How long to decay from a_0 to 1?

Well, $\log a_0 + (n+1)\log 3/4 = 0 \Rightarrow$ $\log a_{n+1} \text{ "=" } \log a_n + \log 3/4$
$\text{"="} \log a_{n-1} + 2\log 3/4 + \cdots$

so $(n+1)\log 3/4 \approx -\log a_0 \Rightarrow (4/3)^{n+1} \approx a_0$

$\Rightarrow n \approx \log a_0 / \log 4/3$

Question: how accurate is this?
size of fluctuations?
prob of much larger/shorter path?

Challenge Problems

① Are there infinitely many a_0 such that $a_1 = 1$?
What about $a_2 = 1$?

② Consider instead a starting seed a_0 and set

$$a_{n+1} = \begin{cases} 3a_n/2^k & \text{with probability } 1/2^k \end{cases}$$

What is the expected value of a_n?
How are the a_n's distributed about this?

Sections 3.3 and 3.4: Expectation (continued)

<u>Defn</u>: <u>Moments</u>: m_k: k^{th} moment $E[X^k] = \int x^k f(x)dx$ or $\sum x_i^k f(x_i)$

σ_k: k^{th} central moment $E[(X-m_1)^k] = \int (x-m_1)^k f(x)\,dx$ or ...

↳ note: often write μ for <u>mean</u> (ie, first moment, m_1).
often write σ^2 for 2nd central moment, called <u>variance</u>
Some books use μ and μ' for moments / central moments

↳ note: $\sigma_k \geq 0$, $\sigma = \sqrt{\sigma^2}$ the standard deviation, measures how spread out

↳ Useful Expression: $\sigma^2 = E[(X-\mu)^2] = E[X^2] - E[X]^2$

↳ note: typically $E[X^2] \neq E[X]^2$

↳ can you give an example where equal?

$$\boxed{\begin{array}{l}\text{LEMMA (IMPORTANT)} \\ \text{If } X, Y \text{ independent then } E[XY] = E[X]E[Y]\end{array}}$$

<u>Notes</u>: $E[\sum X_i] = \sum E[X_i]$ for any $\{X_i\}$; for product we need the X_i's independent

<u>Challenge</u>: can $E[XY] = E[X]E[Y]$ without X, Y indep?

<u>Proof</u>: As <u>indep</u>, $f_{XY}(x,y) = f_X(x) f_Y(y)$ (using <u>indep</u> <u>here</u>)

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{XY}(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} x f_X(x)\,dx \int_{-\infty}^{\infty} y f_Y(y)\,dy = E[X]\,E[Y]$$

Sections 3.3 and ?4.3: Independence (Cont)

Lemma: X, Y random variables:

$Var(X) = E[(X-\mu)^2]$

- $Var(aX) = a^2 Var(X)$,
- $Var(X+Y) = Var(X) + Var(Y)$ if uncorrelated (means $E[XY] = E[X]E[Y]$). More Generally

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{1 \le i < j \le n} CoVar(X_i, X_j)$$

with $CoVar(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$

Proof: algebra: do one on board

Ex: $Var(X+Y) = E\left[\left((X+Y) - E(X+Y)\right)^2\right]$

$$= E\left[\left((X-\mu_X) + (Y-\mu_Y)\right)^2\right]$$

$$= E\left[(X-\mu_X)^2 + 2(X-\mu_X)(Y-\mu_Y) + (Y-\mu_Y)^2\right]$$

$$= Var(X) + 2\underbrace{E\left[(X-\mu_X)(Y-\mu_Y)\right]}_{} + Var(Y)$$

Zero: lots of ways: fastest:
X, Y indep $\Rightarrow g(X), h(Y)$ indep
let $g(X) = X - \mu_X$
$h(Y) = Y - \mu_Y$ ...

Application: Portfolio Theory (Economics)

Say $X_1, \dots, X_n$ all have same mean and variance, say σ^2, and $\boxed{\text{indep}}$

Then $Var(X_1 + \dots + X_n) = n\sigma^2$

and $Var\left(\frac{1}{n}X_1 + \dots + \frac{1}{n}X_n\right) = \frac{1}{n^2}\sum Var(X_i) = \frac{n\sigma^2}{n^2} = \sigma^2/n$

so $\overline{X} = \sum \frac{1}{n} X_i$ has same mean return as before, but much less risky as standard deviation is now $\sigma/\sqrt{n}$

Section 3.3 and 4.3: Independence (cont)

Challenge problem: $X_1, \dots, X_n$ indep, each mean μ and $Var(X_i) = \sigma_i^2$
Can you find weights $w_1, \dots, w_n$ st $0 \le w_i \le 1$, $\sum w_i = 1$ and $Var(\sum w_i X_i)$ is as small as possible? Can you do it for certain nice choices of $\{\sigma_i\}_{i=1 \text{ to } n}$?

HW: do: (3.3): #1, #2, #7 (4.3) #1a, #2
sugg: (3.3): #3, #4, #8 (4.3) #1b, #4, #5

Additional Topic: Chebyshev's Thm (see also Section 7.3)

Chebyshev's Inequality: Let X have mean μ and finite variance σ^2. Then $Prob(|X - \mu| \geq k\sigma) \le 1/k^2$.

Notes:
- Weak conditions: if assume more get much better upper bound for probability. Advantage is how applicable (talk about my MSTD paper)
- Natural scale to measure fluctuations of X about μ is in multiples of σ: VERY IMPORTANT in math'l science to study things at right scale
  - ↳ twin primes so hard as ave spacing b/w primes of size x is $\log x$, and $2/\log x \to 0$

↳ Compare this to Central Limit Theorem (CLT) later in course: weaker result analogous to Divide and conquer versus Newton's Method.

Challenge Problem: Is there a random variable st Chebyshev's inequality is an equality for all k? For all $k \geq 1$? For all integral $k \geq 1$?

Challenge Problem: Distinct Prime Divisors

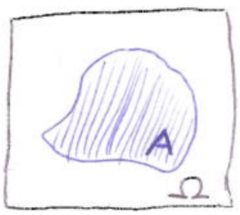
Additional Topic: Chebyshev's Thm

Statement: $\mathbb{P}(|X - \mu| \geq k\sigma) \leq 1/k^2$ (does make sense if $\sigma = \infty$)

Proof:
$$\mathbb{P}(|X-\mu| \geq k\sigma) = \int_{|x-\mu|\geq k\sigma} f(x)\,dx$$
$$\leq \int_{|x-\mu|\geq k\sigma} \left(\frac{x-\mu}{k\sigma}\right)^2 f(x)\,dx$$
$$\leq \frac{1}{k^2\sigma^2}\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$
$$= \sigma^2 / k^2\sigma^2 = 1/k^2$$

Application: Monte Carlo Integration

↳ Will get better results with CLT, but enough to get something nice



Assume want to find area (volume, ...) of some region A. Assume:

(1) it is bounded, for convenience in n-dim hypercube Ω

(2) it is easy to tell if a given point is in A

Choose N points uniformly in Ω; probability any in A is $\frac{Vol(A)}{Vol(\Omega)}$

Let $X_i = \begin{cases} 1 & \text{if } i^{th} \text{ point is in } A \\ 0 & \text{if } i^{th} \text{ point is not in } A \end{cases}$, $\mathbb{E}[X_i] = \frac{Vol(A)}{Vol(\Omega)} = Vol(A)$

$$Var(X_i) = (1 - Vol(A))^2 Vol(A) + (0 - Vol(A))^2 (1 - Vol(A))$$
$$= Vol(A)(1 - Vol(A))$$

$X = \frac{1}{N}\sum_{i=1}^{N} X_i$, Then $\mathbb{E}[X] = Vol(A)$, $Var(X_i) = \frac{Vol(A)(1-Vol(A))}{N}$

Apply Chebyshev: with high probability, error in estimating volume with N points is of the order $1/\sqrt{N}$

Section 3.4: Indicators and Matching

Inclusion/Exclusion

$$\boxed{\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \cdots + (-1)^{n+1}\,\mathbb{P}(A_1 \cap \cdots \cap A_n)}$$

↳ lots of applications:

↳ frequently $\mathbb{P}(A_i)$ indep of i, $\mathbb{P}(A_i \cap A_j)$ indep i,j, ...

↳ Probability a number is square-free, derangements, digits in continued fractions, ...

Matching Problem: n letters, randomly shuffled, what is prob exactly r are matched correctly? Book does general r, we do simpler $r=0$.

Let A_i be event i^{th} letter put back in i^{th} slot.

We want $1 - \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right)$, use inclusion/exclusion

↳ $\mathbb{P}(A_i) = \frac{1}{n} = \frac{(n-1)!}{n!}$

$\mathbb{P}(A_i \cap A_j) = \frac{1}{n(n-1)} = \frac{(n-2)!}{n!}$ and so on.

Note $\#\{(i_1, i_2, \ldots, i_\ell) : i_1 < i_2 < \cdots < i_\ell\}$ is $\binom{n}{\ell}$

Answer is $1 - \left[\binom{n}{1} \cdot \frac{1}{n} - \binom{n}{2}\frac{1}{n(n-1)} + \binom{n}{3}\frac{1}{n(n-1)(n-2)} - \cdots + (-1)^{n-1}\binom{n}{n}\frac{1}{n(n-1)\cdots(2)(1)}\right]$

$$= 1 - \left[n \cdot \frac{1}{n} - \frac{n(n-1)}{2!}\frac{1}{n(n-1)} + \frac{n(n-1)(n-2)}{3!}\frac{1}{n(n-1)(n-2)} - \cdots + (-1)^{n-1}\frac{1}{n!}\right]$$

$$= 1 - \left[1 - \frac{1}{2!} + \frac{1}{3!} + \cdots + (-1)^{n-1}\frac{1}{n!}\right]$$

$$= 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}$$

Note $e^{-1} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \cdots$

HW: Read Example 11, Page 59

Suggested: #1, #3, #9

Appendix V

Table of distributions

	mass/density function	domain	mean	variance	skewness	characteristic function
Bernoulli	$f(1) = p,\ f(0) = q = 1 - p$	$\{0, 1\}$	p	pq	$\dfrac{q-p}{\sqrt{pq}}$	$q + pe^{it}$
Uniform (discrete)	n^{-1}	$\{1, 2, \dots, n\}$	$\frac{1}{2}(n+1)$	$\frac{1}{12}(n^2 - 1)$	0	$\dfrac{e^{it}(1 - e^{int})}{n(1 - e^{it})}$
Binomial bin(n, p)	$\binom{n}{k} p^k (1-p)^{n-k}$	$\{0, 1, \dots, n\}$	np	$np(1-p)$	$\dfrac{1-2p}{\sqrt{np(1-p)}}$	$(1 - p + pe^{it})^n$
Geometric	$p(1-p)^{k-1}$	$k = 1, 2, \dots$	p^{-1}	$(1-p)p^{-2}$	$\dfrac{2-p}{\sqrt{1-p}}$	$\dfrac{p}{e^{-it} - 1 + p}$
Poisson	$e^{-\lambda}\lambda^k / k!$	$k = 0, 1, 2, \dots$	λ	λ	$\lambda^{-\frac{1}{2}}$	$\exp\{\lambda(e^{it} - 1)\}$
Negative binomial	$\binom{k-1}{n-1} p^n (1-p)^{k-n}$	$k = n, n+1, \dots$	np^{-1}	$n(1-p)p^{-2}$	$\dfrac{2-p}{\sqrt{n(1-p)}}$	$\left(\dfrac{p}{e^{-it} - 1 + p}\right)^n$
Hypergeometric	$\dfrac{\binom{b}{k}\binom{N-b}{n-k}}{\binom{N}{n}},\ p = \dfrac{b}{N},\ q = \dfrac{N-b}{N}$	$\{0, 1, 2, \dots, b \wedge n\}$	np	$\dfrac{npq(N-n)}{N-1}$	$\dfrac{q-p}{\sqrt{npq}}\sqrt{\dfrac{N-1}{N-n}}\left(\dfrac{N-2n}{N-2}\right)$	$\dfrac{\binom{N-b}{n}}{\binom{N}{n}} F(-n, -b; N-b-n+1; e^{it})$†
Uniform (continuous)	$(b-a)^{-1}$	$[a, b]$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$	0	$\dfrac{e^{ibt} - e^{iat}}{it(b-a)}$
Exponential	$\lambda e^{-\lambda x}$	$[0, \infty)$	λ^{-1}	λ^{-2}	2	$\dfrac{\lambda}{\lambda - it}$
Normal $N(\mu, \sigma^2)$	$\dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$	$\mathbb{R}$	μ	σ^2	0	$e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$
Gamma $\Gamma(\lambda, \tau)$	$\dfrac{1}{\Gamma(\tau)}\lambda^\tau x^{\tau-1} e^{-\lambda x}$	$[0, \infty)$	$\tau\lambda^{-1}$	$\tau\lambda^{-2}$	$2\tau^{-\frac{1}{2}}$	$\left(\dfrac{\lambda}{\lambda - it}\right)^\tau$
Cauchy	$\dfrac{1}{\pi(1+x^2)}$	$\mathbb{R}$	–	–	–	$e^{-\lvert t \rvert}$
Beta $\beta(a, b)$	$\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$	$[0, 1]$	$\dfrac{a}{a+b}$	$\dfrac{ab(a+b)^2}{a+b+1}$	$\dfrac{2(a-b)}{a+b+2}$	$M(a, a+b, it)$†
Doubly exponential	$\exp(-x - e^{-x})$	$\mathbb{R}$	γ†	$\frac{1}{6}\pi^2$	$1.29857\dots$	$\Gamma(1 - it)$
Rayleigh	$xe^{-\frac{1}{2}x^2}$	$[0, \infty)$	$\sqrt{\dfrac{\pi}{2}}$	$2 - \dfrac{\pi}{2}$	$\dfrac{2\sqrt{\pi}(\pi - 3)}{(4-\pi)^{3/2}}$	$1 + \sqrt{2\pi} it(1 - \Phi(-it))e^{-\frac{1}{2}t^2}$†
Laplace	$\frac{1}{2}\lambda e^{-\lambda\lvert x \rvert}$	$\mathbb{R}$	0	$2\lambda^{-2}$	0	$\dfrac{\lambda^2}{\lambda^2 + t^2}$

†The letter γ denotes Euler's constant.

† $F(a, b; c; z)$ is Gauss's hypergeometric function and $M(a, a+b, it)$ is a confluent hypergeometric fi
The $N(0, 1)$ distribution function is denoted by Φ.

Section 3.5: Examples of Discrete Variables

- Bernoulli: $X = \begin{cases} 1 & \text{prob } p \\ 0 & \text{prob } 1-p \end{cases}$ $\quad Bern(p)$

- Binomial: $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$ $\quad Bin(n,p)$

- Poisson: $f(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ $\quad Pois(\lambda)$

- Geometric $f(k) = p(1-p)^{k-1}$ $\quad Geo(p)$

Note: use distributions to model world; more we know, more can do (assuming can do the sums / integrals!

HW: do: #2 Suggested: #1, #4

Section 4.4: Examples of Continuous Variables

- Uniform: $X = \begin{cases} 1/(b-a) & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$ $\quad Unif(a,b)$

- Exponential: $f(x) = \frac{1}{\lambda} e^{-x/\lambda}$ $\quad Exp(\lambda)$

  ↳ opposite of book: book has $f(x) = \lambda e^{-\lambda x}$ with mean $1/\lambda$

- Normal: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$ $\quad N(\mu, \sigma^2)$

  ↳ polar coords to show integrates to 1, wlog take $\mu = 0, \sigma = 1$

$$I = \int_{-\infty}^{\infty} f(x)dx \rightarrow I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x) f(y) dx dy$$
$$= \int_0^{2\pi}\int_0^{\infty} \frac{1}{2\pi} e^{-r^2/2} r dr d\theta$$
$$= \frac{1}{2\pi} \cdot 2\pi \cdot 1 = 1$$

- Cauchy: $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$

- Weibull: $\frac{\gamma}{\alpha}\left(\frac{x-\beta}{\alpha}\right)^{\gamma-1} \exp\left(-\left(\frac{x-\beta}{\alpha}\right)^{\gamma}\right)$ note different than book

  ↳ application: baseball (Pythagorean paper)

HW: do: #5 Suggested: #2, #3, #4

Section 3.6 and 4.5: Dependence

> Needed input: Cauchy - Schwarz Inequality:
>
> $$E[XY]^2 \leq E[X]^2 E[Y]^2 \text{ or } \left(\int |fg|\right)^2 \leq \left(\int |f|^2\right)\left(\int |g|^2\right)$$

Proof: trivial if $\int |f|^2$ or $\int |g|^2$ is infinite

↳ integral proof: wlog assume $f, g \geq 0$

$\int (af + bg)^2 \, dx \geq 0$ (one of most important inequalities)

$\Rightarrow 0 \leq a^2 \int f^2 + 2ab \int f \int g + b^2 \int g^2$

↳ Thus, regarded as a fn of a, at most one real root as ≥ 0

Thus discriminant is non-positive

↳ $4b^2 \left(\int f \int g\right)^2 - 4 \int f^2 \, b^2 \int g^2 \leq 0$

Thus if take $b = 1$ get claim

↳ similar proof in book for expected value formulation

Defn: Joint Distribution Fn $F : \mathbb{R}^2 \to [0,1]$ of X, Y is

$$F(x,y) = P(X \leq x \text{ and } Y \leq y)$$

↳ discrete: joint mass fn $f(x,y) = P(X = x \text{ and } Y = y)$

Continuous: joint prob density fn $F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u,v) \, du \, dv$

> LEMMA: X, Y indep if and only if $f_{X,Y}(x,y) = f_X(x) f_Y(y)$

Proof: $\Leftarrow$ Clear as integral of products becomes prod of integrals

$\Rightarrow$ sketch:

Sections 3.6 and 4.5: Dependence (cont)

Marginals from $f_{X,Y}$

$$f_X(x) = \sum_y f_{X,Y}(x,y) \quad \text{or} \quad \int_{y=-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

Ex: $f(x,y) = \frac{\alpha^x \beta^y}{x!\,y!} e^{-(\alpha+\beta)} \qquad x,y \in \{0,1,2,\dots\}$

↳ $f_X(x) = \left(\sum_{y=0}^{\infty} \frac{\beta^y e^{-\beta}}{y!}\right) \frac{\alpha^x e^{-\alpha}}{x!} = \frac{\alpha^x e^{-\alpha}}{x!}$, ← Poisson

↳ see X, Y both Poisson and indep

Defn: Covariance of X and Y is $\mathrm{CoVar}(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$

Correlation Coefficient is $\rho(X,Y) = \mathrm{CoVar}(X,Y) / \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$

Lemma:
- $\mathrm{CoVar}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ (= 0 if independent)
- $|\rho(X,Y)| \le 1$ with equality iff $\exists\, a,b,c$ s.t. $P(aX+bY=c)=1$

Proof: first is algebra, second is Cauchy-Schwarz

↳ $\left|\mathbb{E}[(X-\mu_x)(Y-\mu_y)]\right| = \left|\iint (x-\mu_x)\cdot(y-\mu_y)\, f_{X,Y}(x,y)\,dx\,dy\right|$

$$\le \iint |x-\mu_x|\sqrt{f_{X,Y}(x,y)} \cdot |y-\mu_y|\sqrt{f_{X,Y}(x,y)}\;dx\,dy$$

$$\le \left[\iint (x-\mu_x)^2 f_{X,Y}(x,y)\,dx\,dy\right]^{\frac12} \left[\iint (y-\mu_y)^2 f_{X,Y}(x,y)\,dx\,dy\right]^{\frac12}$$

$$= \left[\int (x-\mu_X)^2 f_X(x)\,dx\right]^{1/2} \left[\int (y-\mu_y)^2 f_Y(y)\,dy\right]^{\frac12}$$

$$= \mathrm{Var}(X)^{1/2}\,\mathrm{Var}(Y)^{1/2}$$ ∎

Important Note

↳ to use Cauchy-Schwarz, split $f_{X,Y}$ into $\sqrt{f_{X,Y}} \cdot \sqrt{f_{X,Y}}$ need so that when square have a finite integral: $\int x^2\,dx$ is infinite... See this technique numerous times (Cramer-Rao inequality for instance).

HW: do (3.6): #2, #7 — do (4.5): #6
sugg (3.6): #1, #5, #6 — sugg (4.5): #2

SECTION 3.7 and 4.6: CONDITIONAL DISTR AND CONDITIONAL EXPECTATIONS

Defn: X, Y variables on $(\Omega, \mathcal{F}, \mathbb{P})$

Conditional distribution fn of Y given $X = x$ is $F_{Y|X}(y|x) = \mathbb{P}(Y \le y \mid X = x)$ for any x st $\mathbb{P}(X = x) > 0$. The conditional (prob) mass function of Y given $X = x$ is $f_{Y|X}(y|x) = \mathbb{P}(Y = y \mid X = x)$.

Conditional distribution fn of Y given $X = x$ is

$$F_{Y|X}(y|x) = \int_{-\infty}^{y} \frac{f(x,v)}{f_X(x)}\,dv \quad \text{for } x \text{ st } f_X(x) > 0;$$

sometimes denote $\mathbb{P}(Y \le y \mid X = x)$. The conditional density fn is $f_{Y|X}(y|x) = f(x,y) / f_X(x)$ for any x st $f_X(x) > 0$.

↳ X, Y indep if $f_{Y|X} = f_{X,Y} / f_X$.

↳ $\int y\, f_{Y|X}(y|x)$ is called the conditional expectation of Y given $X = x$

↳ Let $\psi(x) = \mathbb{E}[Y \mid X = x]$. $\psi(X)$ is the conditional expectation of Y given X, and written $\mathbb{E}[Y \mid X]$.

$$\boxed{\text{Thm: } \mathbb{E}[\psi(X)] = \mathbb{E}[Y]}$$

↳ Proof: $\mathbb{E}[\psi(X)] = \int \psi(x) f_X(x)\,dx$

$$= \iint y\, f_{Y|X}(y|x)\, f_X(x)\,dx\,dy$$

$$= \iint y\, f_{X,Y}(x,y)\,dx\,dy = \int y\, f_Y(y)\,dy = \mathbb{E}[Y]$$

↳ $\mathbb{E}[Y] = \sum_x \mathbb{E}[Y \mid X = x]\, \mathbb{P}(X = x) = \int \mathbb{E}[Y \mid X = x]\, f_X(x)\,dx$

See Example 5, page 68

Note: due to time constraints, high probability we'll skip this section

Section 4.7: Functions of Random Variables

As mentioned, using CDFs can easily get densities new rand vas.

Ex: X random variable with density f_X, $g: \mathbb{R} \to \mathbb{R}$ nice (say monotone increasing, and maybe differentiable). Let $Y = g(X)$.

Then $\mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y) = \mathbb{P}(X \le g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X(x)\,dx$

So $\mathbb{P}(Y \le y) = F_X(g^{-1}(y)) - 0$

$\Rightarrow f_Y(y) = f_X(g^{-1}(y)) \cdot \frac{d}{dy}(g^{-1}(y))$ by chain rule

↳ Note: do not need to know F_X

Generalization: X_1, X_2 joint density fn $f_{X,Y}$; let $(Y_1, Y_2) = T(X_1, X_2)$

So $T: (x_1, x_2) \to (y_1, y_2)$. Have maps $y_i = y_i(x_1, x_2)$ which for nice T can be inverted to $x_i = x_i(y_1, y_2)$ with Jacobean $J = \begin{vmatrix} \partial x_1/\partial y_1 & \partial x_2/\partial y_1 \\ \partial x_1/\partial y_2 & \partial x_2/\partial y_2 \end{vmatrix}$

Then $f_{Y_1, Y_2}(y_1, y_2) = \begin{cases} f_{X_1, X_2}(x_1(y_1, y_2), x_2(y_1, y_2))\,|J(y_1, y_2)| & \text{if } (y_1, y_2) \text{ in the range of } T \\ 0 & \text{otherwise} \end{cases}$

Ex: X_1, X_2 indep with density $\lambda e^{-x\lambda}$, $Y_1 = X_1 + X_2$, $Y_2 = X_1 / X_2$

$T(x_1, x_2) = (x_1 + x_2, x_1/x_2)$ and $T^{-1}(y_1, y_2) = \left(y_1 y_2/(1+y_2), \ y_1/(1+y_2)\right)$

Jacobean is $-y_1/(1+y_2)^2$

Density is $f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}\left(\frac{y_1 y_2}{1+y_2}, \frac{y_1}{1+y_2}\right) \frac{|y_1|}{(1+y_2)^2} = \frac{\lambda^2 e^{-\lambda y_1} y_1}{(1+y_2)^2}$

(for $y_1, y_2 \geq 0$, of course!)

↳ factors as $g(y_1) h(y_2)$ so indep

↳ Read Example (9) on pg 111

HW: do: #2

sugg: #8, #11

SECTIONS 3.8 and 4.8: SUMS OF RANDOM VARIABLES

Convolution: $(f * g)(x) = \int_{-\infty}^{\infty} f(t)\, g(x-t)\, dt$

↳ If X, Y indep with densities f, g then $f_{X+Y} = f * g$

Example: $X_i \sim Unif(0,1)$ then $X_1 + X_2$ has density

$$\int_{-\infty}^{\infty} f(t)\, f(x-t)\, dt \quad \text{with} \quad f(u) = \begin{cases} 1 & 0 \le u \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$= \int_0^1 f(x-t)\, dt$; clearly 0 unless $0 \le x \le 2$

↳ if $2 \ge x \ge 1$, have $\int_{x-1}^{1} 1\, dt = 2 - x$

if $0 \le x \le 1$ have $\int_0^x 1\, dt = x$

↳ note answer is symm about $x = 1$.

LEMMA: CONVOLUTION PROPERTIES

$f * g = g * f$

$\widehat{f * g} = \hat{f} \cdot \hat{g}$, where $\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i x \xi}\, dx$ is the Fourier Transform

$f * \delta = f$ where δ is the Dirac Delta Functional

$f * (g * h) = (f * g) * h$

Ex: X_1, X_2 $Pois(\lambda_1)$ and $Pois(\lambda_2)$. Then if $X = X_1 + X_2$, density is

$$\sum_n f_{X_1}(n)\, f_{X_2}(x-n) = \sum_n \frac{\lambda_1^n e^{-\lambda_1}}{n!} \cdot \frac{\lambda_2^{x-n} e^{-\lambda_2}}{(x-n)!}$$

$$= \frac{1}{x!} \left[\sum_{0 \le n \le x} \binom{x}{n} \lambda_1^n \lambda_2^{x-n} \right] e^{-(\lambda_1 + \lambda_2)}$$

$$= \frac{(\lambda_1 + \lambda_2)^x\, e^{-(\lambda_1 + \lambda_2)}}{x!} \sim Pois(\lambda_1 + \lambda_2)$$

↳ note how easy proofs are with generating fns

HW: do (3.8) #4, #5 do: (4.8) #1, #4

Sugg (3.8) #3, #6 Sugg (4.8) #2, #5

SECTION 4.10: DISTRIBUTIONS ARISING FROM NORMAL

- Sample mean: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

↳ Note divide by $n-1$ not n

↳ Can't compute sample variance if $n=1$, so reasonable

Thm: X_1, X_2 independent $N(\mu, \sigma^2)$. Then $\overline{X}$ and S^2 are indep, with $\overline{X} \sim N(\mu, \sigma^2/n)$ and $(n-1)S^2 \sim \chi^2(n-1)$

↳ here $\chi^2(d)$ has density $x^{\frac{d}{2}-1} e^{-x/2} \big/ 2^{\frac{d}{2}} \Gamma\left(\frac{d}{2}\right)$

with $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\,dx$

↳ call this a χ^2 distribution with d degrees of freedom

HW: do: #1, Show if $X \sim N(0,1)$ then X^2 is $\chi^2(1)$

PROBLEMS

Section 3.11: do #1a, #7, #9, #13, #14, #21, #30, #33

need prereq: §3.2 §3.5 §3.1 §3.3, §3.6 §3.5, 3.8 §3.7 §3.7 §3.8

Suggested: #2, #4, #10, #11, #17, #18, #22, #25

Section 4.14: do: #12, #16

needed prereq: §4.4, 4.10 §4.4, 4.10

Suggested: #18, #21, #27, #28, #35, #44, #45