Catherine Yeo

COUNTING

- **Sampling table**:

	order	order doesn't matter
w/ replace	n^k	$\binom{n+k-1}{k}$ → Bose-Einstein
w/out replace	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

MATH

- Geometric series ($\lvert x\rvert<1$) $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$, $\sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x}$
- Taylor series of e^x $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$
- Binomial Thm $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$

PROBABILITY

- Independent events
  $P(A\cap B) = P(A)\cdot P(B)$
  $P(A\mid B) = P(A)$ $(P(B)\neq 0)$
- Conditional independence
  $P(A\cap B\mid C) = P(A\mid C)P(B\mid C)$
- Conditional indp. $\cancel{\Leftrightarrow}$ Indp.

Ex: Indp. $\not\Rightarrow$ Cond. indp.
A = 1st coin is H, B = 2nd coin is H
C = 2 coins are the same
A, B indp. But knowing C happened, knowing A would tell you B.

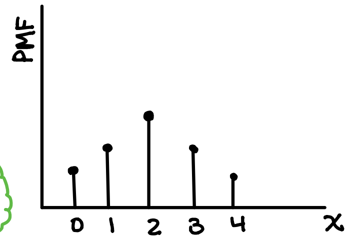
Ex: Indp. $\not\Leftarrow$ Cond. indp.
Coin 1 99% H, Coin 2 1% H
A = 1st flip H, B = 2nd flip H,
C = coin 1 drawn

- Inclusion-Exclusion (PIE)
  $P(A\cup B) = P(A) + P(B) - P(A\cap B)$
  $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{i<j} P(A_i\cap A_j) + \dots + (-1)^{n+1} P(A_1\cap\dots\cap A_n)$
- DeMorgan's Laws
  $(A\cup B)^c = A^c \cap B^c$
  $(A\cap B)^c = A^c \cup B^c$
- **LOTP** → used for wishful thinking
  $P(A) = \sum_{i=1}^{n} P(A\mid B_i)P(B_i)$
  $P(A\mid C) = \sum_{i=1}^{n} P(A\mid B_i, C)\, P(B_i\mid C)$ ← orig. event conditioned
- **Bayes' Rule**
  $P(A\mid B) = \frac{P(B\mid A)P(A)}{P(B)}$ $P(A\mid B,C) = \frac{P(B\mid A,C)P(A\mid C)}{P(B\mid C)}$

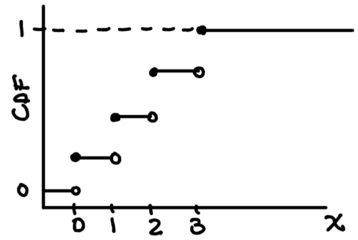
DISTRIBUTIONS

- **Probability Mass Function** (PMF)
  ↳ gives prob. that a discrete r.v. takes on value of x
  - $P_X(x) = P(X=x)$
  - PMF ≥ 0, $\sum$ prob. $= 1$

  remember to give support!



- **Cumulative Distribution Function** (CDF)
  ↳ gives prob. that r.v. $\leq x$
  - Always increasing $\begin{cases} F_X \to 0 & \text{as } x\to -\infty \\ F_X \to 1 & \text{as } x\to \infty \end{cases}$
  - Same CDF ⇒ same distr.



EXPECTATION, VARIANCE, INDICATORS

- Expected value (mean/avg/expectation)
  $E(X) = \sum_x x\cdot P(X=x)$
  $E(X) = \int_{-\infty}^{\infty} x\cdot f(x)\,dx$
- **LOTUS** (can find exp. val. of function of rv)
  $E(g(X)) = \sum_x g(x)\cdot P(X=x)$
  $= \int_{-\infty}^{\infty} g(x)\cdot f(x)\,dx$
- Linearity of Expectation ← holds even if X, Y dependent
  $E(aX+bY+c) = aE(X) + bE(Y) + c$
- Same distribution ⇒ same mean
  X, Y same distr. → $E(X)=E(Y)$ → $E(g(X)) = E(g(Y))$
- **Fundamental bridge**
  $E(I_A) = P(A) = P(I_A = 1)$
- Power of **indicators**
  $I_A^k = I_A$
- Product $I_A I_B = I_{A\cap B}$
- Union $I_{A\cup B} = I_A + I_B - I_{A\cap B}$
- Variance
  $\mathrm{Var}(X) = E(X^2) - (E(X))^2$ (LOTUS, indicators)
  $= E(X - E(X))^2$
  $SD(X) = \sqrt{\mathrm{Var}(X)}$
- $\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$
- If X, Y indp.:
  $\mathrm{Var}(X+Y) = \mathrm{Var}(X-Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

CONTINUOUS R.V.s

- **Probability density function** (PDF)
  f is derivative of CDF F
  - f is non-negative
  - $\int f = 1$
  - $F(x) = \int_{-\infty}^{x} f(t)\,dt$
- Universality of the uniform
  ↳ When you plug CRV into its CDF, $\sim \mathrm{Unif}(0,1)$. If $U\sim\mathrm{Unif}(0,1)$, $F^{-1}(U)$ has CDF F

Discrete vs Continuous

	Discrete	Continuous
$P(X\leq x) =$	$F(x)$ (CDF)	$F(x)$ (CDF)
To find prob.s,	Add over PMF $P(X=x)$	Integrate over $f(x) = F'(x)$
$E(X) =$	$\sum_x x\cdot P(X=x)$	$\int_{-\infty}^{\infty} x\cdot f(x)\,dx$
$E(g(X)) =$	$\sum_x g(x)\cdot P(X=x)$ ↳ LOTUS discrete	$\int_{-\infty}^{\infty} g(x)\cdot f(x)\,dx$ ↳ LOTUS continuous

DISCRETE DISTRIBUTIONS

Distributions for sampling schemes:

	Draw w/ Replace	No Replace
Fixed # of trials	Binomial (Bern if $n=1$)	HGeom
Draw until k successes	NBin (Geom if $k=1$)	*NHGeom

Bernoulli Distribution $X \sim \text{Bern}(p)$

- Only 1 trial ($n=1$)
- Story: X succeeds w/ prob p, fails w/ prob q ($q = 1-p$)
- Ex: fair coin toss $\sim \text{Bern}(\frac{1}{2})$. $1-X \sim \text{Bern}(\frac{1}{2})$ tails.
- PMF: $P(X=x) = p^x (1-p)^x$

$$= \begin{cases} p & x=1 \\ 1-p & x=0 \end{cases}$$

Binomial Distribution $X \sim \text{Bin}(n, p)$

- Story: X = # of successes achieved in n independent trials, each with success prob. p
- Properties: Let $X \sim \text{Bin}(n,p)$, $Y \sim \text{Bin}(m,p)$, indp.
  - $n - X \sim \text{Bin}(n, 1-p)$
  - $X + Y \sim \text{Bin}(n+m, p)$
  - $X \mid (X+Y=r) \sim \text{HGeom}(n, m, r)$
  - $\text{Bin}(n,p)$ approx. $\text{Pois}(\lambda)$ if p small
  - $\text{Bin}(n,p) \approx N(np, np(1-p))$ if n big, p far from 0 or 1

Geometric Distribution $X \sim \text{Geom}(p)$

- Story: X = # failures achieved before 1st success, success prob. p

$\Rightarrow$ First Success Distribution includes success, $E(X) = 1/p$

Negative Binomial Distribution $X \sim \text{NBin}(r, p)$

- Story: # of failures before r-th success, success prob. p

Poisson $X \sim \text{Pois}(\lambda)$

- Story: # of (rare) events that occur at an avg rate of λ occurrences per unit time
- Paradigm: $X = I_1 + \cdots + I_n$
  1) n is large
  2) $P(I_i = 1) = p_i$ small $\forall i$
  3) Events are indp. or weakly dep.

  $\Rightarrow$ X approx. $\sim$ Poisson w/ $\lambda = E(X)$

Hypergeometric Distribution $X \sim \text{HGeom}(w, b, n)$

- Story: w desired, b undesired, X = # "successes" in simple random sample of n objects w/out replacem.

If $X, Y \sim \text{Pois}(\lambda_1), \text{Pois}(\lambda_2)$, $X \perp\!\!\!\perp Y$, $X \mid (X+Y=n) \sim \text{Bin}(n, \frac{\lambda_1}{\lambda_1+\lambda_2})$.

CONTINUOUS DISTRIBUTIONS

Uniform Distribution $U \sim \text{Unif}(a,b)$

- $f(x) = \begin{cases} c & a<x<b \\ 0 & \text{else} \end{cases}$ $\quad c(b-a) = 1$, $\quad c = \frac{1}{b-a}$
- Prob. of a draw from any interval w/in support is proportional to length of interval
- PDF and CDF:

Unif$(0,1)$ $\quad f(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & \text{else} \end{cases}$ $\quad F(x) = \begin{cases} 0, & x<0 \\ x, & x \in [0,1] \\ 1, & x>1 \end{cases}$

Unif(a,b) $\quad f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{else} \end{cases}$ $\quad F(x) = \begin{cases} 0 \\ \frac{x-a}{b-a} \\ 1 \end{cases}$

Normal Distribution $X \sim N(\mu, \sigma^2)$

- Standard normal (CDF = Φ): $Z \sim N(0,1)$ has $\mu = 0$, Var $= 1$.
- Central Limit Thm: sample mean of iid r.v.s will approach N as sample size grows, regardless of initial distr.
- Location-scale transformation: every time we shift or rescale a r.v., we change it to another r.v.
  - ↳ $X \sim N(\mu, \sigma^2) \Rightarrow N(0,1)$ by $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

CONDITIONAL EXPECTATION

Conditioning on Event

- $E(X|A) = \sum_x x \cdot P(X=x|A)$ or $= \int_{-\infty}^{\infty} x \cdot f(x|A)\,dx$

a number

- Ex: Let $T \sim \text{Expo}(1/10)$ be how long to wait until shuttle comes. Given waited t mins, memoryless ⇒ expected wait time is 10 more mins. $E(T|T>t) = t+10$.

Conditioning on r.v.

- $E(Y|X)$ → function
- Find $E(Y|X=x)$, then plug in X for x

Properties:

- $E(Y|X) = E(Y)$ if $X \perp\!\!\!\perp Y$
- Taking out what's known
  - ↳ $E(h(X)\cdot Y|X) = h(X)\cdot E(Y|X)$
  - $E(h(X)|X) = h(X)$
- Law of Total Exp. (LOTE)
  - ↳ $E(X) = \sum_{i=1}^{n} E(X|A_i)\cdot P(A_i)$
- Adam's Law
  - ↳ $E(E(Y|X)) = E(Y)$
  - $E(E(Y|X,Z)|Z) = E(Y|Z)$
- Eve's Law
  - ↳ $\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$

MGFs

- $M_X(t) = E(e^{tX})$ is MGF of X if it $\exists\, \forall t$ in open interval containing 0.
- $\mu_k = E(X^k) = M_X^{(k)}(0)$
- $M_{X+Y}(t) = M_X(t)\cdot M_Y(t)$
- $M_Y(t) = e^{bt} M_X(at)$, if $Y = aX+b$
- Uniqueness: X and Y have same distr. iff their MGFs are equal
- Ex: Expo MGF

  $X \sim \text{Expo}(1)$

  $M(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} e^{-x}\,dx = \frac{1}{1-t}$ for $t<1$.

MOMENTS

Moments describe shape of distribution.

X has μ, σ, $Z = \frac{X-\mu}{\sigma}$ is standardized X.

- k-th moment of X is $\mu_k = E(X^k)$
- k-th standard moment of X is $m_k = E(Z^k)$
- Mean: $E(X) = \mu_1$
- Variance: $\text{Var}(X) = \mu_2 - \mu_1^2$
- Skewness: $\text{Skew}(X) = m_3 = E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right)$
- Symmetry: $X-\mu$ has same distr. as $\mu - X$ (PDF sym.)
- Kurtosis: $\text{Kurt}(X) = m_4 - 3$

- Sample Moment
  - ↳ $M_k = \frac{1}{n}\sum_{j=1}^{n} X_j^k$
- Sample variance
  - ↳ $S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2$

JOINT PDFs & CDFs

- Joint CDF: $F(x,y) = P(X\le x, Y\le y)$
- Joint PMF: $P_X(x,y) = P(X=x, Y=y)$
- Joint PDF: $f_{X,Y}(x,y) = \frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x,y)$

  ↳ non-neg., sum to 1

- Marginal from joint:

  PMF $P(X=x) = \sum_y P(X=x, Y=y)$

  PDF $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$

- LOTUS

  $E(g(X,Y)) = \sum_x \sum_y g(x,y)\, P(X=x, Y=y)$

  $E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$

COVARIANCE

→ measures r.v.s go ↑/↓ together

$\text{Cov}(X,Y) = E((X-EX)(Y-EY))$
$= E(XY) - (EX)(EY)$

CORRELATION

→ normalized Cov b/t −1 and 1

$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$

Properties of Cov and Corr:

- $\text{Cov}(X,X) = \text{Var}(X)$
- If $X \perp\!\!\!\perp Y$, $\text{Cov}(X,Y) = 0 \Rightarrow E(XY) = E(X)E(Y)$
- $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$
- If $X \perp\!\!\!\perp Y$, $\text{Var}(X) + \text{Var}(Y) = \text{Var}(X+Y)$
- $\text{Cov}(X,Y) = \text{Cov}(Y,X)$
- $\text{Cov}(X+a, Y+b) = \text{Cov}(X,Y)$
- $\text{Cov}(aX, bY) = ab\,\text{Cov}(X,Y)$
- $\text{Cov}(W+X, Y+Z) = \text{Cov}(X,Y) + \text{Cov}(W,Y) + \text{Cov}(W,Z) + \text{Cov}(X,Z)$
- $\text{Corr}(aX+b, cY+d) = \text{Corr}(X,Y)$

MULTINOMIAL DISTRIBUTION

- Story: n items can fall into any 1 of k buckets indp.ly with prob. $\vec{p} = (p_1, \dots, p_k)$
- $\vec{X} = (X_1, \dots, X_k) \sim \text{Mult}_k(n, \vec{p})$
- Ex: 100 students at Hogwarts sorted.

  # of ppl sorted per house $\sim \text{Mult}_4(100, \vec{p})$, $p_i = \frac{1}{4}$.

  $X_1 + X_2 + X_3 + X_4 = 100$, dependent.

Joint PMF:

$P(X_1 = n_1, \dots, X_k = n_k) = \frac{n!}{n_1!\,n_2!\cdots n_k!}\, p_1^{n_1}\cdots p_k^{n_k}$

- Marginal PMF: $X_i \sim \text{Bin}(n, p_i)$
- $X_i + X_j \sim \text{Bin}(n, p_i + p_j)$ for $i \ne j$
- Conditioning on X_j still gives a Multinomial:

  $X_1, \dots, X_{k-1} | X_k = n_k \sim \text{Mult}_{k-1}\left(n - n_k, \left(\frac{p_1}{1-p_k}, \dots, \frac{p_{k-1}}{1-p_k}\right)\right)$

- $\text{Cov}(X_i, X_j) = -np_ip_j$

MULTIVARIATE NORMAL DISTR.

- If every linear combo of X_j has a Normal distribution:

  $t_1X_1 + \dots + t_kX_k \sim \text{Norm}$

- Parameters:
  1) Mean vector $\vec{\mu} = (\mu_1, \mu_2, \dots, \mu_k)$
  2) Covariance matrix ⇒ $[i][j] = \text{Cov}(X_i, X_j)$
- Any subvector is also MVN
  - ↳ if $(X_1, X_2, X_3) \sim \text{MVN}$, $(X_1, X_2) \sim \text{MVN}$
- $X, Y \sim \text{MVN}$, $W = (X_1, \dots, X_n, Y_1, \dots, Y_m) \sim \text{MVN}$
- In MVN vector, if any 2 elements are uncorrelated, they are independent
- Joint MGF:

  $M(t) = E(e^{t'X}) = E(e^{t_1X_1 + \dots + t_kX_k})$

- Joint PDF of Bivar. Normal (X,Y) with $N(0,1)$ marg. distr. & correl. $\rho \in (-1,1)$:

  $f_{X,Y}(x,y) = \frac{1}{2\pi\tau}\exp\left(-\frac{1}{2\tau^2}(x^2+y^2-2\rho xy)\right)$

  with $\tau = \sqrt{1-\rho^2}$.

TRANSFORMATIONS

- 1D Change of variables

  ↳ $Y = g(X)$, g diff. & strictly incr./decr.

  $f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$ $x = g^{-1}(y)$

  $= f_X(g^{-1}(y))\left|\frac{d}{dy} g^{-1}(y)\right|$

- Convolutions

  $f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(x) f_Y(t-x)\,dx$

- Ex: $X, Y \sim N(0,1)$ i.i.d. Fixed t,

  $f_{X+Y}(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-x)^2}{2}}\,dx$

  ⇒ $N(0,2)$ PDF.

ORDER STATISTICS

- $X_{(i)}$ = i-th smallest element
- Dependent
- CDF:

  $P(X_{(j)} \le x) = \sum_{k=j}^{n}\binom{n}{k} F(x)^k (1-F(x))^{n-k}$

- PDF:

  $f_{X_{(j)}}(x) = n\binom{n-1}{j-1} f(x)\, F(x)^{j-1}(1-F(x))^{n-j}$

- $U_{(j)} \sim \text{Beta}(j, n-j+1)$, for $U_i \sim \text{Unif}(0,1)$

CONTINUOUS DISTRIBUTIONS

Standard Normal $N(0,1)$

- CDF $= \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$
- $\varphi(z) = \varphi(-z)$
- $\Phi(z) = 1 - \Phi(-z)$
- $P(-z \le Z \le z) = P(Z \ge -z) = 1 - \Phi(-z)$
- **Location-Scale**
  $$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

Exponential Distribution

- $F(x) = 1 - e^{-\lambda x}$, $x > 0$
- $f(x) = \lambda e^{-\lambda x}$, $x > 0$
- Story: Shooting stars avg appear every 15 mins. Wait time until next ★ $\sim$ Expo(4) hours. λ = rate
- Rescaling Expo:
  $Y \sim \text{Expo}(\lambda) \rightarrow X = \lambda Y \sim \text{Expo}(1)$

Uniform CDF

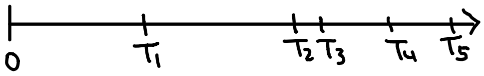
$\text{Unif}(0,1) \rightarrow F(x) = x$, $x \in (0,1)$

- **Memoryless property**
  - Expo is the only contin. mem.less distr.!
  - $P(X > s+t \mid X > s) = P(X > t)$
    $E(X \mid X > s) = s + E(X) = s + \frac{1}{\lambda}$
  - $X - a \mid (X > a) \sim \text{Expo}(\lambda)$
- Given independent $X_i \sim \text{Expo}(\lambda_i)$, $\min(X_1, \dots, X_k) \sim \text{Expo}(\lambda_1 + \dots + \lambda_k)$
- Given i.i.d. $X_i \sim \text{Expo}(\lambda)$, $\max(X_1, \dots, X_k)$ has same distr. as $Y_1 + \dots + Y_k$, where $Y_j \sim \text{Expo}(j\lambda)$, Y_j are indp.

Poisson Process

process with rate λ arrivals per unit time and:

1) # of arrivals in time interval of length t is $\text{Pois}(\lambda t)$ r.v.
2) # of arrivals in disjoint time intervals are indp. of each other



- **Count-Time Duality**
  - T_i = time i-th ✉ arrives
  - N_t = # of ✉s b/t $[0,t]$
  - Event $T_1 > t \iff N_t = 0$
    $P(T_1 > t) = P(N_t = 0) = e^{-\lambda t}$
    $\rightarrow P(T_1 \le t) = 1 - e^{-\lambda t}$
    Thus $T_1 \sim \text{Expo}(\lambda)$.
    $T_n - T_{n-1} \sim \text{Expo}(\lambda)$

Beta Distribution

- **Bank/Post Office**: bank time $X \sim \text{Gamma}(a, \lambda)$, PO time $Y \sim \text{Gamma}(b, \lambda)$. Total time at both is $X + Y \sim \text{Gamma}(a+b, \lambda)$. Fraction of time at bank is $\frac{X}{X+Y} \sim \text{Beta}(a,b)$. $\frac{X}{X+Y} \perp\!\!\!\perp X+Y$
- $\text{Beta}(1,1) \sim \text{Unif}(0,1)$
- **Bayes' billiards**: $\int_0^1 \binom{n}{k} x^k (1-x)^{n-k}\, dx = \frac{1}{n+1}$
  - ① $n+1$ balls, n white, 1 gray; throw onto $[0,1] \sim \text{Unif}(0,1)$; X = # whites left of gray; $P(X=k) = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k}\, dp$
  - ② $n+1$ white, paint 1 gray $\Rightarrow \frac{1}{n+1}$
- Beta is **conjugate prior** of Binomial
  $X \mid p \sim \text{Bin}(n,p)$, $p \sim \text{Beta}(a,b)$ $\Rightarrow p \mid (X = x) \sim \text{Beta}(a+x, b+n-x)$

Gamma Distribution $X \sim \text{Gamma}(a, \lambda)$

- Story: waiting time for shooting star $\sim \text{Expo}(\lambda)$. Want to see n ★s. Wait time for n-th ★ $\sim \text{Gamma}(n, \lambda)$.
- $X_1, \dots, X_n \overset{iid}{\sim} \text{Expo}(\lambda)$. Then $X_1 + \dots + X_n \sim \text{Gamma}(n, \lambda)$
- $\text{Gamma}(1, \lambda) \sim \text{Expo}(\lambda)$
- $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\, dx$, $a > 0$
- $\Gamma(a+1) = a \cdot \Gamma(a)$
- $\Gamma(n) = (n-1)!$, $n \in \mathbb{Z}^+$

χ^2 Distribution

- X is distributed χ_n^2
- Story: Sum of squares of n indp. std. Norm r.v.s

$\Rightarrow X$ distributed as $Z_1^2 + Z_2^2 + \dots + Z_n^2$, $Z_i \overset{iid}{\sim} N(0,1)$
$X \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$

CONVOLUTIONS OF R.V.s ($X \perp\!\!\!\perp Y$)

- $X \sim \text{Pois}(\lambda_1)$, $Y \sim \text{Pois}(\lambda_2) \rightarrow X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$
- $X \sim \underset{(NBin)}{\text{Bin}}(n_1, p)$, $Y \sim \underset{(NBin)}{\text{Bin}}(n_2, p) \rightarrow X + Y \sim \underset{(NBin)}{\text{Bin}}(n_1 + n_2, p)$
- $X \sim \text{Gamma}(a_1, \lambda)$, $Y \sim \text{Gamma}(a_2, \lambda) \rightarrow X + Y \sim \text{Gamma}(a_1 + a_2, \lambda)$
- $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2) \rightarrow X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

SPECIAL CASES OF DISTRIBUTIONS

- $\text{Bin}(1,p) \sim \text{Bern}(p)$
- $\text{Beta}(1,1) \sim \text{Unif}(0,1)$
- $\text{Gamma}(1, \lambda) \sim \text{Expo}(\lambda)$
- $\text{NBin}(1,p) \sim \text{Geom}(p)$

INEQUALITIES

Chebyshev	$P(\lvert X - \mu \rvert \ge a) \le \frac{\sigma^2}{a^2}$
Cauchy-Schwarz	$\lvert E(XY) \rvert \le \sqrt{E(X^2) E(Y^2)}$
Markov	$P(X \ge a) \le \frac{E\lvert X \rvert}{a}$ for $a > 0$
Jensen	$E(g(X)) \ge g(E(X))$ if g convex (ex: x^2) ; $\le$ if concave

LAW OF LARGE NUMBERS

- Sample mean: $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$
- LLN says as $n \to \infty$, $\bar{X}_n \to \mu$ with probability 1

CENTRAL LIMIT THEOREM

- Approximation $\Rightarrow Y \,\dot\sim\, N(\mu_Y, \sigma_Y^2)$
  $Y = X_1 + X_2 + \dots + X_n$; $\bar{X}_n \,\dot\sim\, N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$, $\mu_Y = n\mu_X$, $\sigma_Y^2 = n\sigma_X^2$
- Asymptotic distributions
  - As $n \to \infty$, $\sqrt{n}\left(\frac{\bar{X}_n - \mu_X}{\sigma_X}\right) \xrightarrow{D} N(0,1)$.

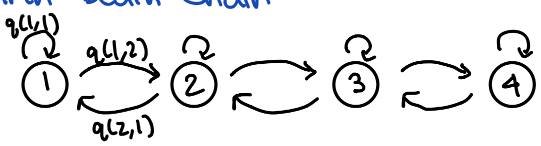
MORE MATH

- Gamma & Beta Integrals (good for pattern matching)
  $$\int_0^\infty x^{t-1} e^{-x}\, dx = \Gamma(t), \quad \int_0^1 x^{a-1} (1-x)^{b-1}\, dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$
- Harmonic: $1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \approx \log n + 0.577\ldots$
- Stirling's Approx. $n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$

MARKOV CHAINS

- A **Markov chain** is a random walk in a (finite) state space $\{1,2,\dots,M\}$, the sequence of r.v.s of where the walk is at all times in time.
- **Markov property**: given all past history, only the most recent term X_n matters for predicting X_{n+1}
  → when checking: would knowing more help? If yes, X.
- $P(X_{n+1}=j \mid X_0=i_0,\dots,X_n=i) = P(X_{n+1}=j \mid X_n=i)$

Birth-Death Chain



- $q_{ij}>0$ if $|i-j|=1$, $q_{ij}=0$ if $|i-j|\geq 2$
- Step to right = birth, left = death
- Reversible ⇒ $s_j = \dfrac{s_1 q_{12} q_{23} \cdots q_{j-1,j}}{q_{j,j-1}\, q_{j-1,j-2} \cdots q_{21}}$ → choose s_1 so $s_1+\dots+s_M=1$

State

- Recurrent: start at i, can always return to i
- Transient: else
- Aperiodic: if gcd of possible # of steps to return to i from $i=1$
- Irreducible: can get from anywhere to anywhere
  ↳ all states recurrent

Random Walk on Undirected Graph

$s_i = \dfrac{d_i}{\sum_i d_i}$

$d_1=3$

→ also reversible

Symmetric Transition Matrix

↳ stationary distr. is uniform over all states

Reversibility

- If $\exists \vec{s}$ s.t. $s_i q_{ij} = s_j q_{ji}$ $\forall i,j$.
  ↳ if reversible, stationary

Stationary Distribution

- $\vec{s}=(s_1,\dots,s_M)$ is s.d. for chain if $\vec{s}Q=\vec{s}$ ↔ $(Q^T-I)\vec{s}^T=0$
- Uniform if cols sum to 1
- Irred. + Aperiodic ⇒ s.d. ∃ & unique. Expected steps to return to i is $\frac{1}{s_i}$.
- If X_t has s.d., $X_{t+1}, X_{t+2},\dots$ also have s.d.

Tip: if not obvious how to find stationary distribution, likely need to prove reversible.

METROPOLIS ALGORITHM ⇒ reversible

Start at X_0. Suppose new chain is at X_n.

1) If $X_n=i$, propose new state j using transition prob. in i-th row of orig. trans. matrix P
2) Compute acceptance prob.
   $$a_{ij} = \min\left(\frac{s_j p_{ji}}{s_i p_{ij}},\ 1\right)$$
3) Flip a coin that lands H with prob. a_{ij}
4) Accept proposal (go to j) if H. Else, reject ($X_{n+1}=i$).

Ex. Problem (2011 #8)

a) Find q_{ij} $\forall i,j$.

$i\neq j$: If no edge $\{i,j\}$, $q_{ij}=0$.
Yes edge,
$$q_{ij} = \frac{1}{d_i}\cdot \min\left(\frac{d_i}{d_j},1\right) = \begin{cases} \frac{1}{d_i} & \text{if } d_i \geq d_j \\ \frac{1}{d_j} & \text{if } d_i < d_j \end{cases}$$

$i=j$: $q_{ii} = 1-\sum_{j\neq i} q_{ij}$ (every row sums to 1)

b) Find stat. distr.

WTS $q_{ij}=q_{ji}$ ⇒ reversible.
If $d_i \geq d_j$, $q_{ij}=\frac{1}{d_i}$, $q_{ji}=\frac{1}{d_j}\cdot\frac{d_j}{d_i}=\frac{1}{d_i}$.
If $d_i<d_j$, $q_{ij}=\frac{1}{d_j}$, $q_{ji} = \frac{1}{d_j}$.
∴ Reversible ⇒ $s_j=\frac{1}{M}$ $\forall j$, uniform.

LOG-NORMAL DISTRIBUTION

- If $Y\sim \text{Log-Normal}(\mu,\sigma^2)$, then $\log(Y)\sim N(\mu,\sigma^2)$.
- If $Y\sim N(\mu,\sigma^2)$, then $X=e^Y \sim \text{log-Normal}(\mu,\sigma^2)$.

Ex (2015 #7):
Let $X,Y,Z,W \sim N(0,1)$.
a) LOTUS ⇒ $E(\Phi(Z)e^Z) = \int_{-\infty}^{\infty} \Phi(z)e^z \varphi(z)\,dz$
b) By UoU, $\Phi(Z)\sim \text{Unif}(0,1)$, so $E(\Phi(Z))=\frac{1}{2}$.
Since e^Z is log-normal, $E(e^Z)=e^{1/2}$.
c) $P(X+Y<Z+W+1) = P(X+Y-Z-W<1)$
$= P\left(\frac{X+Y-Z-W}{2}<\frac{1}{2}\right)$
$= \Phi(\frac{1}{2})$, since $X+Y-Z-W\sim N(0,4)$

BUS PARADOX

Times b/t buses are 10 mins on avg. Fred waits >10 mins on avg. How is it possible?

⇒ length-biased sampling

- Can choose distribution s.t. most intervals are short, some super long
- More likely to arrive during long interval than short interval b/t buses

CLASS-SIZE PARADOX

16 courses with 10 students, 2 lectures with 100 students.

- Dean's eye view:
  $$\frac{16\cdot 10 + 2\cdot 100}{16+2} = 20$$
- Student's eye view:
  $$\frac{160\cdot 10 + 200\cdot 100}{16\cdot 10 + 2\cdot 100} = 60$$

Large class weighted more in students' eyes, same weights in dean's eye

BETA-BIN + BILLIARDS

(2012 #5)
V wins chess w/ $p\sim \text{Beta}(a,b)$, M wins w/ q.

a) $E(N)$? N = # games needed for V to win one

$N|p \sim FS(p)$, so
$E(N) = E(E(N|p)) = E\left(\frac{1}{p}\right)$
$= \int_0^1 \frac{1}{p}\cdot\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\cdot p^{a-1}(1-p)^{b-1}\,dp$

Beta(a-1,b)'s PDF integrates to 1, so
$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\cdot\frac{\Gamma(a-1)\Gamma(b)}{\Gamma(a+b-1)} = \frac{a+b-1}{a-1}$.

c) Cond. distr. of p given V wins 7 out of 1st 10 games.
Beta$(a+7, b+10-7)$

EXPO + MEMORYLESS

(2016 #5)
5 components, each works for Expo(λ) time, then fails. Expected time until 2 comp. die?

- Time until 1 comp. dies ~ Expo(5λ) due to min property
- Memoryless ⇒ $T_{(2)}\sim \text{Expo}(4\lambda)$
- $E(T) = \frac{1}{5\lambda}+\frac{1}{4\lambda} = \frac{9}{20\lambda}$.

(2013 #4)
$N\sim \text{Pois}(\lambda_1)$ qs posted, tmrw, unif., Expo(λ_2) time to answer. Find prob. a q posted tmrw won't be ans. tmrw.

$X\sim \text{Unif}(0,1)$ = time asked, T = time to answer
$P(X+T>1) = \int_0^1\int_{1-x}^{\infty} \lambda_2 e^{-\lambda_2 t}\,dt\,dx = \frac{1-e^{-\lambda_2}}{\lambda_2}$.

(2013 #6)
a) Given Fred has waited 20 mins, memoryless ⇒ still has to wait $1/\lambda_1$ mins.
b) Prob. of n buses pass before route 2 = $\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^n$.

GENERAL STRATEGY

1) What should the answer look like?
   ↳ Ex: if CDF, must be incr., right-cont, and $F(x)\to 0$ as $x\to -\infty$, $F(x)\to 1$ as $x\to\infty$.
2) Category errors?
3) Extreme cases
4) Simple cases
5) Solve

WHEN STUCK, REMEMBER:

- You can name (a)'s answer as a variable, then reference in later parts
- Adam's Law
- Eve's Law (Var!)
- Bayes' Billiards (esp. with k out of n options, integral, simplify fully)
- Complements