Counting

	order	order doesn't matter
with replacement	n^k	$\binom{n+k-1}{k}$ → Bose-Einstein or Stars n Bars
w/o replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$ → $\frac{n!}{(n-k)!k!}$ ← Binom. coefficient

Ex Birthday Problem
k ppl in room, 1 pair has same bday:
$1 - \frac{365 \cdot 364 \cdots (365-k+1)}{365^k}$
Logic: find complement s.t. no 2 ppl share the same birthday.

Ex Bose-Einstein
$\binom{n+k-1}{n}$ ways to arrange n balls to k boxes (box divisions w/ k-1 lines), permute n+k-1 balls + box divisions

Math

Taylor Series
$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$

Vandermonde's Identity
$$\sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j} = \binom{m+n}{k}$$
Story: n girls, m guys, form k teams of students.
Left: choose m guys $\binom{m}{j}$, then k-j girls $\binom{n}{k-j}$, so we sum these cases. Right: choose k out of total (m+n)

Disjointness + Independence

Disjoint - cannot happen simultaneously
$P(A \cap B) = 0$; $A \cap B = \emptyset$
Independent - knowing 1 gives you no info about the other
$P(A \cap B) = P(A)P(B)$; $P(A|B) = P(A)$
Conditional Independence - A|C and B|C
$P(A \cap B|C) = P(A|C)P(B|C)$
* Condit. indpt. ≠ indpt. *

Monty Hall (Bayes + LOTP)

If we chose door 1 and monty opens door 2
Planning if we switch?
· use Bayes $\frac{1}{3}$ + LOTP
$P(C_3|M_2) = \frac{P(M_2|C_3)P(C_3)}{P(M_2)}$ ← LOTP to know car door · would be nice
$P(M_2) = P(M_2|C_1)P(C_1) + P(M_2|C_2)P(C_2) + P(M_2|C_3)P(C_3)$
$= \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$
$= \frac{1}{6} + \frac{2}{6} = \frac{1}{2}$
$\frac{1}{3} \div \frac{1}{2} = \boxed{\frac{2}{3}}$

Probability

Bayesian: based on info.
Frequentist: large # of times

Principle of Inclusion-Exclusion (PIE)

* use venn diagram *
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$
This stays with conditioning:
ex $P(A \cup B | C) = P(A|C) + P(B|C) - P(A \cap B|C)$

De Morgan's
- $(A \cup B)^c = A^c \cap B^c$
- $(A \cap B)^c = A^c \cup B^c$

Properties
- $P(A^c) = 1 - P(A)$
- $A \subseteq C$ implies $P(A) \leq P(B)$
- * PIE *

Defn. Conditional

$P(A|B) = \frac{P(A \cap B)}{P(B)}$
or
$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$
ex: like $P(P(P(A|B),C),D)$
$P(A,B,C,D) = P(A)P(B|A)P(C|B,A)P(D|C,B,A)$
note: $P(A|A) = 1$ is always true
here, A is "prior" prob. w/o updating based on evid.

Geometric: $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$ $|x| < 1$

Bayes' Rule

Discrete: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ ← usually use LOTP
$P(A|B,C) = \frac{P(B|A,C)P(A|C)}{P(B|C)}$
Continuous: $f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)}$

Variance

$Var(X) = E(X - EX)^2 = E(X^2) - (E(X))^2$ ← usually use LOTUS
Properties
- $Var(X) \geq 0$
- $Var(cX) = c^2 Var(X)$
- $Var(X + c) = Var(X)$
- if X,Y indpt., $Var(X+Y) = Var(X) + Var(Y)$

Standard dev. is square of var.

LOTP (Law of total prob.)

remember to use for wishful thinking
Discrete: $P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$
ex: $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$
basically condition on every possibility.
Xtra conditioning:
$P(A|C) = \sum_{i=1}^{n} P(A|B_i, C)P(B_i|C)$
ex: $P(A|B)$ conditioned on C: $P(A|B,C)P(B|C)$

Gambler's Ruin (1st-step analysis)

Typically used for recursive.
Steps:
apply LOTP → Rewrite in terms of itself
→ solve resulting equation
ans:
$$p_i = \begin{cases} \frac{1 - (\frac{q}{p})^i}{1 - (\frac{q}{p})^N} & p \neq 1/2 \\ \frac{i}{N} & p = 1/2 \end{cases}$$

Gambler A has i dollars, wins prob. p
B has N-i, wins 1-p.
$\forall k \in \{1, \ldots, N-1\}$ (where we start)
$p_k = p \cdot p_{k+1} + (1-p) p_{k-1}$
↑ A wins ↑ A loses

Random Variables

CDF $F_X(x) = P(X \leq x)$
3 properties:
1) F is increasing $x_1 \leq x_2$, $F(x_1) \leq F(x_2)$
2) Right-continuous $F(a) = \lim_{x \to a^+} F(x)$
3) Convergence to 0 and 1 in limits $F(x) \to 1$ as $x \to \infty$, $F(x) \to 0$ as $x \to -\infty$

Continuous:
PDF $f_X(x) = F'_X(x)$
CDF $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$
* Integrate PDF to get CDF
Differentiate CDF to get PDF *

Properties of PDF:
Defn: relative freq of getting value within ε of x value.
- nonnegative
- f integrates to 1 since $\int_{-\infty}^{\infty} f(x) = 1$

Discrete: (positive if x is in support of X, 0 o/w)
PMF $P_X(x) = P(X = x)$
CDF $F_X(x) = \sum_{y \leq x} P_X(y)$
Independence for all x,y if
$P(X = x, Y = y) = P(X = x)P(Y = y)$

Independence if
$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$

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

Indicator r.v.s
I_A / $I(A) = 1$ if A occurs
$= 0$ if A doesn't
Properties:
- $I_{A^c} = 1 - I_A$
- $I_{A \cap B} = I_A I_B$ (and)
- $(I_A)^k = I_A$, $k \in \mathbb{Z}^+$
- $I_{A \cup B} = I_A + I_B - I_A I_B$ (or)

Expectation

Discrete: $E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i)$
Continuous: $E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$
Properties:
- Linearity of expectation
$E(aX + bY + c) = aE(X) + bE(Y) + c$
- Symmetry - if 2 r.v. have same dist, their values are equal
- Fundamental Bridge - property between prob. and expectation.
$P(A) = E(I_A)$ ← with indicator variables

ex: N=r.v. counting $A_1, \ldots, A_n$. Let I_j be indicator for event A_j occurring.
$N = I_1 + \ldots + I_n \Rightarrow E(N) = E(I_1 + \ldots + I_n)$
with linearity and fundamental bridge:
$EN = EI_1 + \ldots + EI_n = P(A_1) + \ldots + P(A_n)$

LOTUS - if X is a r.v. and g is some func:
$E(g(X)) = \sum_x g(x)P(X = x)$
Continuous:
$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)\,dx$
ex: $E(X) = \sum_{n=0}^{\infty} n p_n$ then $E(X^3) = \sum_{n=0}^{\infty} n^3 p_n$
$E(X^2) = \sum_x x^2 P(X = x)$

- $E(X) = \mu$ · $Var(X) = M_2 - M_1^2$
- Skew(X) $= m_3 = E\left(\frac{X - \mu}{\sigma}\right)^3$
- Kurtosis: $Kurt(X) = E\left(\frac{X - \mu}{\sigma}\right)^4 - 3$
- Symmetry: $\mu - X$ same as $X - \mu$

Sampling moment: $M_k = \frac{1}{n}\sum_{j=1}^{n} X_j^k$
Sample variance: $S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2$
Sample mean is $\frac{1}{n}\sum_{j=1}^{n} X_j$

Moments + MGFs

Provide info. abt. shape of dist.
- n^{th} moment of X is $E(X^n)$
$M(t) = E(e^{tX})$
if MGF equal, same dist.
- n^{th} derivative of MGF:
$m^{(n)}(t) = E(X^n e^{tX})$
To find the n^{th} moment, find n^{th} deriv. at 0:
ex: $M(0) = 1$
$M'(0) = E(Xe^0) = E(X)$
$M''(0) = E(X^2 e^0) = E(X^2)$

Mean + Var of Sample mean:
$E(\bar{X}_n) = \mu$
$Var(\bar{X}_n) = \frac{\sigma^2}{n}$

Properties:
- Determines distribution.
If $M_X(t) = M_Y(t)$, then $X \sim Y$
- With $M_X(t)$ as MGF of X.
If $Y = aX + b$, then $M_Y(t) = e^{bt}M_X(at)$
- Independence of X and Y
$M_{X+Y}(t) = M_X(t)M_Y(t)$

Order Statistics

j^{th} largest value of a set of iid r.vs
Defn: n iid r.v.s $X_1, \ldots, X_n$, allow to crystallize and arrange from smallest → largest, then dist of j^{th} val is $\boxed{j^{th} \text{ order stat}}$ X_j.
usually involves finding CDF of X_j.
$X_{(j)} \leq x$ is "at least j of $X_1 \& X_2, \ldots, X_n$ fall to the left of x.
CDF $P(X_{(j)} \leq x) = \sum_{k=j}^{n} \binom{n}{k} F(x)^k (1 - F(x))^{n-k}$
PDF
$f_{X_{(j)}}(x) = n\binom{n-1}{j-1} f(x)F(x)^{j-1}(1 - F(x))^{n-j}$
Order statistic of Uniform $U_1, \ldots, U_n \overset{iid}{\sim} Unif(0,1)$
$U_{(j)} \sim Beta(j, n-j+1)$

Change of Variables

find PDFs of complex r.v.s in terms of PDFs of r.v.s we know.
X is continuous r.v. w/ PDF f_X
$Y = g(X)$ → g is differentiable AND strictly increasing/decreasing
Then PDF of Y
$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$ where $x = g^{-1}(y)$
To get $\frac{dx}{dy}$, we can take its reciprocal
ex: Log-Normal PDF:
$X \sim N(0,1)$, $Y = e^X$ → strictly increasing
$y = e^x$ $x = \log y$ $\frac{dy}{dx} = e^x$ derivative
$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \varphi(x)\frac{1}{e^x} = \varphi(\log y)\frac{1}{y}$
solve for x in e^x

Conditional Expectation

Conditional given event ⇒ some quantity
given r.v. ⇒ r.v.
$E(Y|X)$ is function of r.v. X
Properties
- if X,Y indpt, $E(Y|X) = E(Y)$
- can pull out functions of X
$E(h(X)Y|X) = h(X)E(Y|X)$
- linearity
$E(Y_1 + Y_2|X) = E(Y_1|X) + E(Y_2|X)$
$E(cY|X) = cE(Y|X)$

LOTE: $E(X) = \sum_{i=1}^{n} E(X|A_i)P(A_i)$
Adam's Law: $E(E(Y|X)) = E(Y)$ ← event
Xtra conditioning: $E(E(Y|X,Z)|Z) = E(Y|Z)$
Eve's Law: $Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$

Convergence

Law of Large Numbers:

- iid X_i with finite mean μ and finite variance σ^2
- Sample: $\bar{X}_n = \frac{X_1 + \ldots + X_n}{n}$
- Strong LLN: $\bar{X}_n \to \mu$ (true mean) as $n \to \infty$
- Weak LLN: $\forall \epsilon > 0$, $P(|\bar{X}_n - \mu| > \epsilon) \to 0$
"convergence in probability" → Prob sample mean is close to μ goes to 1 as $n \to \infty$

Central Limit Theorem:

- as $n \to \infty$:
$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \overset{D}{\to} N(0,1)$
- another way for large n:
$\bar{X}_n \mathrel{\dot\sim} N(\mu, \sigma^2/n)$

Inequalities

Cauchy-Schwarz
- for any 2 r.v. X and Y w/ finite var.
$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$
useful for correlated X,Y, don't have to know joint

Jensen's Inequality
- for g as convex,
$E(g(X)) \geq g(E(X))$
- concave:
$E(g(X)) \leq g(E(X))$

Convex: x^2 → even; $g''(x) \geq 0$ 2nd deriv.
Concave: $-x^2$; log, sqrt; $g''(x) \leq 0$

Bounds on Tail Probabilities
- markov $P(|X| \geq a) \leq \frac{E|X|}{a}$
- Chebyshev $P(|X - \mu| \geq a) \leq \frac{Var(X)}{a^2}$
- Chernoff $P(X \geq a) \leq \frac{E(e^{tX})}{e^{ta}}$

Moments Example:
$X \sim Expo(1)$
$M(t) = E(e^{tX})$
$= \int_0^{\infty} e^{tx}e^{-x}dx$
$= \frac{1}{1-t}$ for $t < 1$

Joint Distributions

- Joint CDF: $F_{X,Y}(x,y) = P(X\le x, Y\le y)$
- Joint PMF of discrete: $p_{X,Y}(x,y) = P(X=x, Y=y)$
- Joint PDF of continuous: $f_{X,Y}(x,y) = \frac{d^2}{dx\,dy} F_{X,Y}(x,y)$

} nonneg + sum to 1

Marginal Distributions

getting marginal from joint:

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

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

Independence:

joint decomposes to marg.

- $F_{X,Y}(x,y) = F_X(x)F_Y(y)$
- $P(X=x, Y=y) = P(X=x)P(Y=y)$
- $P(Y=y \mid X=x) = P(Y=y)$

Conditional Distributions

Discrete r.v.s: $P(Y=y \mid X=x) = \frac{P(X=x, Y=y)}{P(X=x)}$

Continuous: $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$

we can find marginal given conditional:

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

2D-LOTUS

Discrete:

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

Continuous:

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

Bayes

	Y discrete	Y contin.
X discrete	$P(Y=y \mid X=x) = \frac{P(X=x \mid Y=y)P(Y=y)}{P(X=x)}$	$f_Y(y \mid X=x) = \frac{P(X=x \mid Y=y) f_Y(y)}{P(X=x)}$
X cont.	$P(Y=y \mid X=x) = \frac{f_X(x \mid Y=y)P(Y=y)}{f_X(x)}$	$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) f_Y(y)}{f_X(x)}$

LOTP

	Y discrete	Y contin.
X discrete	$\sum_y P(X=x \mid Y=y)P(Y=y)$	$\int_{-\infty}^{\infty} P(X=x \mid Y=y) f_Y(y)\,dy$
X cont.	$\sum_y f_X(x \mid Y=y)P(Y=y)$	$\int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\,dy$

Covariance

If $X \perp\!\!\!\perp Y$, then $Cov(X,Y)=0$

Captures how X and Y ↑↓ together

$$Cov(X,Y) = E((X-E(X))(Y-E(Y))) = E(XY) - E(X)E(Y)$$

Rules

- $Cov(X,X) = Var(X)$
- $Cov(X,Y) = Cov(Y,X)$
- $Cov(X,c) = 0$
- $Cov(aX,Y) = a\,Cov(X,Y)$
- $Cov(X+Y, Z+W) = Cov(X,Z) + Cov(X,W) + Cov(Y,Z) + Cov(Y,W)$
- $Var(X+Y) = Cov(X+Y, X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
- $Corr(aX+b, cY+d) = Corr(X,Y)$
- $Var(X_1 + \dots + X_n) = Cov(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n) + 2\sum_{i<j} Cov(X_i, X_j)$

Correlation

Standardized covariance.

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

$-1 \le Corr(X,Y) \le 1$

Multinomial $\vec{X} \sim Mult_k(n, \vec{p})$

generalization of binomial (either in category or not)

Story: n items can fall into one of k buckets indpt. w/ prob $\vec{p} = (p_1, p_2, \dots, p_k)$.

ex: 100 students sorted at Hogwarts. # ppl per house $\sim Mult_4(100, \vec{p})$

$p_i = \frac{1}{4}$. $X_1+X_2+X_3+X_4 = 100$, dependent.

Joint PMF:

$$P(X_1=n_1, X_2=n_2, \dots, X_k=n_k) = \underbrace{\frac{n!}{n_1! \dots n_k!}}_{\text{multinom coefficient } \binom{n}{n_1 \dots n_k}} p_1^{n_1} \dots p_k^{n_k}$$

Properties:

- subvector of $\vec{X}$ is also multinomial
- Conditioning: if $\vec{X} \sim Mult_k(n, \vec{p})$ then $(X_2, \dots, X_k) \mid X_1 = n_1 \sim Mult_{k-1}\left(n-n_1, \left(\frac{p_1}{1-p_k}, \dots, \frac{p_{k-1}}{1-p_k}\right)\right)$
- Let $\vec{X} \sim Mult_k(n, \vec{p})$. Then, $Cov(X_i, X_j) = -np_ip_j$

Marginal PMF:

$X_i \sim Bin(n, p_i)$

Lumping:

"lumping" together 2 multivariates gives you a multivariate.

2 dimensions = binomial.

Cases:

- $X_i \sim Bin(n, p_i)$
- $X_i + X_j \sim Bin(n, p_i + p_2)$
- $X_1, X_2, X_3 \sim_3 (n, (p_1, p_2, p_3)) \Rightarrow X_1, X_2 + X_3 \sim_2 (n, (p_1, p_2 + p_3))$
- $X_1, X_2, \dots, X_{k-1} \mid X_k = n_k \sim_{k-1} \left(n - n_k, \left(\frac{p_1}{1-p_k}, \dots, \frac{p_{k-1}}{1-p_k}\right)\right)$

Multivariate Normal

if every linear combination of X_j has a Normal distribution

$t_1X_1 + \dots + t_kX_k \sim Norm(\mu, \sigma^2)$

when k=2, it's Bivariate Normal

MVN is denoted: $\vec{X} \sim N_k(\vec{\mu}, V)$

$\vec{\mu} = (\mu_1, \dots, \mu_k) = (E(X_1) \dots E(X_k))$, $\mu_i = |X_i|$

V is covariance matrix $V = (v_{ij})$,

$v_{ij} = Cov(X_i, X_j)$ This is symmetric ($V = V^T$)

Covariate matrix:

$$Cov(\vec{X}) = \begin{bmatrix} Cov(X_1,X_1) \dots Cov(X_1,X_n) \\ \vdots \\ Cov(X_n,X_1) \dots Cov(X_n,X_n) \end{bmatrix}$$

Properties:

- Normal r.v.s isn't necessarily MVN

ex: $X \sim N(0,1)$, $S \begin{cases} 1 & p=1/2 \\ -1 & p=1/2 \end{cases}$ then $Y = SX$ due to symmetry on normal.

not BVN since $P(X+Y=0) = P(S=-1) = 1/2$ can't be BVN since constant

- Suppose $\vec{X} = (X_1, \dots, X_n)$ is MVN, then any subvector is also MVN.

Markov Chain

A type of stochastic process, is a random walk in a finite space $\{1, 2, \dots\}$ $\forall\, n \ge 0$

" The transition matrix is symmetric, so the stationary distribution is uniform over the state space $\{1, \dots, M\}$

if colums sum to 1, then symmetric

Markov Property: given past history, use only most recent term

$$P(X_{n+1}=j \mid X_n=i) = P(X_{n+1}=j \mid X_n=i, X_{n-1}=i_{n-1}, \dots, X_1=i_1, X_0=i_0)$$

Transition Matrix:

$q_{ij} = P(X_{n+1}=j \mid X_n=i)$ is a transition probability

n^{th}-step transition property is prob. of being at state j exactly n steps after being at step i. (n^th power of trans. matrix)

marginal distribution of X_n is tQ^n where t is vector of initial prob.

Properties:

- if markov chain is irreducible, all states must be recurrent.
- Let i be a transient state. Suppose prob. of never returning to i starting from i, # times chain returns to i before leaving forever is dist. (p).

States

- recurrent: start at i, can always return to i.
  - else transient.
- aperiodic: if gcd of possible # of steps to return to i from i = 1
  - else periodic
- irreducible: can get from anywhere to anywhere (all states recurrent)
  - else reducible

ex)

trans. matrix

$$\begin{array}{c|cc} & A & B \\ \hline A & 2/3 & 1/3 \\ B & 1/4 & 3/4 \end{array}$$

square matrix 2nd step

$$\begin{array}{c|cc} & A & B \\ \hline A & 19/36 & 17/36 \\ B & 17/48 & 31/48 \end{array}$$

rows sum to 1

Stationary dist:

long-run prob. of being in any state regardless of starting position

row vector $s = (s_1, \dots, s_M)$ $s_i \ge 0$, $\sum_i s_i = 1$ if

$\sum_i s_i q_{ij} = s_j \;\forall j$

or

$sQ = S$

if we move from s, we will stay in s forever

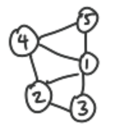
- any irreducible markov chain has unique stationary distribution, each state has positive prob.

Random Walk on Undir. Network

let $\vec{d} = (d_1, \dots, d_n)$ be degree (# edges attach)

if markov chain is walk on undir. network, then $d_i q_{ij} = d_j q_{ji}$.

ex



stationary proportion to deg sequence

$$s = \left(\frac{4}{14}, \frac{3}{14}, \frac{2}{14}, \frac{3}{14}, \frac{2}{14}\right)$$

Symmetric Matrix

- if Q symmetric ($q_{ij} = q_{ji}$) then reversible and stationary dist is uniform

$s = (1/M, \dots, 1/M)$

- if each column sums to 1, then is stationary

Reversibility

$\exists\, \vec{s}$ s.t. $s_i q_{ij} = s_j q_{ji}$ for all i,j

↳ if reversible, then stationary

* intuitively- chain behaves the same way forward or backward. *

Expected time to return:

Stationary dist s → $s_i = \frac{1}{r_i}$ ← Expected time

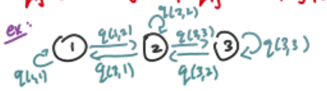
r_i: return to state i, given it starts at i

Birth-Death Chain

can only go 1 step to left/right

$q_{ij} > 0$ if $|i-j| = 1$, $q_{ij} = 0$ if $|i-j| \ge 2$

ex:



- BirthDeath is reversible:

$$s_j = \frac{s_1 q_{12} q_{23} \cdots q_{j-1,j}}{q_{j,j-1} q_{j-1,j-2} \dots q_{21}}$$

$\forall$ states $2 \le j \le M$. choose s, s.t. s_j sums to 1

Discrete Distributions

Distribution	Story	PDF $f(x)$ or $(P(X=x)=?)$	CDF $F(x)$ or $(P(X\le x))$	Support $(K \in ?)$	$E(X)$	$Var(X)$
Bernouli $(Bern(p))$ same as $Bin(1,p)$	X "succeeds" (is 1) with prob. p and "fails" (is 0) w/ prob. $1-p$. ex: fair coin flip is $Bern(\frac{1}{2})$	$p^x(1-p)^{1-x}$ or $\begin{cases} p & x=1 \\ 1-p & x=0 \end{cases}$ aka $P(X=1)=p$, $P(X=0)=1-p$	$\begin{cases} 0 & \text{if } k<0 \\ 1-p & \text{if } 0\le k<1 \\ 1 & \text{if } k\ge 1 \end{cases}$	$\{0,1\}$	p	pq
Binomial $(Bin(n,p))$ Σn iid $Bern(p)$	X is # of "success" in n indpt. trials also sum of Bern ($X=X_1+\dots+X_n \sim Bin(n,p)$ when $X_i \sim Bern(p)$) ex make 10 free throws w/ 3/4 prob. of getting in. X = # free throws he makes $\sim Bin(10,\frac{3}{4})$	$\binom{n}{x}p^x(1-p)^{n-x}$	$\sum_{i=0}^{x}\binom{n}{i}p^i(1-p)^{n-i}$	$\{0,1,\dots,n\}$	np	npq
Geometric $(Geom(p))$ $NBin(1,p)$	X = # of failures before the 1st success ex each pokeball has prob. $\frac{1}{10}$ to catch, # failed poke balls is $Geom(\frac{1}{10})$	$q^k p$	$\begin{cases} 1-(1-p)^{x+1} & \text{if } x\ge 0 \\ 0 & \text{if } x<0 \end{cases}$	$\{0,\dots\}$	$\frac{q}{p}$	$\frac{q}{p^2}$
First Success $(FS(p))$	Same as Geom, but includes 1st success, $X \sim Geom(p)+1$	$q^{k-1}p$		$\{0,\dots\}$	$\frac{1}{p}$	$\frac{q}{p^2}$
Neg. Binomial $(NBin(r,p))$ Σr $Geom(p)$	X is # of failures before r^{th} success ex Thundershock has 60% accuracy + faints Raticate in 3 hits. # of misses before fainting $\sim NBin(3,0.6)$	$\binom{n+r-1}{r-1}p^r q^n$	do math	$\{0,\dots\}$	$\frac{rq}{p}$	$\frac{rq}{p^2}$
Hypergeometric $(HGeom(w,b,n))$	w desired, b undesired, X = # "success" in simple random sample of n obj. w/o replacement. ex 1) w white balls, b black, draw b balls. X = # of white balls you draw 2) N elk, capture n, tag, then release. collect new sample m, X = # tagged elk.	$\frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$	math	$\{0,\dots\}$	$\frac{nw}{w+b}$	$\left(\frac{w+b-n}{w+b-1}\right)M\left(1-\frac{M}{n}\right)$
Poisson $(Pois(\lambda))$	λ = rate parameter (λ successes in unit) low probability, high occurance X = # events occuring in this time. Don't forget chicken-egg	$\frac{e^{-\lambda}\lambda^k}{k!}$	$P(X\le t) = 1-e^{-\lambda x}$	$\{0,\dots\}$	λ	λ

differentiate

integrate

1) Poisson approximation: for a large number n of indpt / weakly dependent trials $A_1,\dots,A_n$ where $p_k = P(A_k)$ for each small p_k, then $\sim Pois(\lambda)$ where $\lambda = \sum_k p_k$

2) Chicken-Egg: Chicken lays $N \sim Pois(\lambda)$ eggs, each w/ prob. p of hatching independently. X = # eggs that hatch, Y = # eggs that don't hatch, then $X \sim Pois(\lambda p)$, $Y \sim Pois(\lambda q)$ and X, Y are indpt.

joint PMF:

eggs are indpt. Bernouli trials w/ success p, so $X|N=n \sim Bin(n,p)$, $Y|N=n \sim Bin(n,q)$.

wishful thinking to know total # of eggs:

$$P(X=i, Y=j) = \sum_{n=0}^{\infty} P(X=i, Y=j \mid N=n)\,P(N=n)$$

$$P(X=i \mid N=i+j)\,P(N=i+j)$$

$$= \binom{i+j}{i}p^i q^j \cdot \frac{e^{-\lambda}\lambda^{i+j}}{(i+j)!} = \frac{e^{-\lambda p}(\lambda p)^i}{i!} \cdot \frac{e^{-\lambda q}(\lambda q)^j}{j!}$$

Metropolis (for Markov)

Metropolis-Hastings The Metropolis-Hastings algorithm is a general recipe that lets us start with any irreducible Markov chain on the state space of interest and then modify it into a new Markov chain that has the desired stationary distribution.

Given a probability distribution vector $s = (s_1, \dots, s_M)$, we want to construct a Markov chain that has s as a stationary distribution (assume $s_j > 0$ for all j). Then, our algorithm is as follows:

1. Take any Markov chain transition matrix P. If the chain is currently at $X_n = i$, then generate a new *proposal* destination j, using the ith row of the transition matrix P.
2. Compute the transition probability
$$a_{ij} = \min\left(\frac{s_j p_{ji}}{s_i p_{ij}}, 1\right)$$
3. With probability a_{ij}, *accept* the proposal and transition to $X_{n+1} = j$. Otherwise, *reject* the proposal and transition to (i.e., stay at) $X_{n+1} = i$.

The main takeaway is that the Metropolis-Hastings algorithm is an extremely general way to construct a Markov chain with a desired stationary distribution.

Continuous Distributions

Uniform $U \sim Unif(a,b)$

prob. of a draw from any interval of Unif is proportional to length of unif.

note: PDF is constant, so integrating gives u area proportional to length of interval

PDF:
$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & \text{otherwise} \end{cases}$$

CDF
$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{x-a}{b-a} & x \in (a,b) \\ 1 & x > 0 \end{cases}$$

$E(x) = \frac{a+b}{2}$ $\quad Var(x) = \frac{(b-a)^2}{12}$

Order statistic of Uniform $U_1, \dots, U_n \overset{iid}{\sim} Unif(0,1)$
$U_{(j)} \sim Beta(j, n-j+1)$

Universality of the Uniform

for continuous r.v. X, can transform to uniform and back with CDF.

F is strictly increasing + continuous on support

1. Let $U \sim Unif(0,1)$ and $X = F^{-1}(U)$. X is r.v. w/ CDF of F.
2. X be r.v. w/ CDF F. $F(X) \sim Unif(0,1)$

Exponential $X \sim Expo(\lambda)$

PDF: $f(x) = \lambda e^{-\lambda x}$ $\quad$ CDF: $1 - e^{-\lambda x}$

$E(x) = \frac{1}{\lambda}$ $\quad Var(x) = \frac{1}{\lambda^2}$

"Standard" distribution $Expo(1)$

$Expo(\lambda) = \frac{1}{\lambda} Expo(1)$

Story:
1) Shooting stars come every 15 min on average, but never "due" even if you wait a while. memoryless

$X \sim Expo(4)$, where λ = rate

$E(x) = \frac{1}{\lambda} = \frac{1}{4}$ since you see avg 1 every 15 min.

Memoryless "doesn't remember prev. values"

$$P(X \geq a+b \mid X \geq a) = P(X \geq b)$$

given you've waited a min, P(wait xtra b min) = P(wait b min). $X-a \mid X \geq a \sim Expo(\lambda)$

$E(X \mid X > 5) = 5 + E(X) = 5 + \frac{1}{\lambda}$

Properties:
- Given indpt. $X_i \sim Expo(\lambda_i)$, $\min(X_1, \dots, X_k) \sim Expo(\lambda_1 + \dots + \lambda_k)$
- Given iid $X_i \sim Expo(\lambda)$, $\max(X_1, \dots, X_k)$, $E(X) = \frac{1}{\lambda_1 + \dots + \lambda_n} + \dots + \frac{1}{\lambda_1}$

Given time period, # occurrences of Expo is $X \sim Pois(\lambda t)$

Beta $X \sim Beta(a,b)$

PDF for $x \in (0,1)$

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

or

$$f(x) = \frac{1}{\underbrace{\beta(a,b)}_{\text{constant}}} x^{a-1}(1-x)^{b-1}$$

aka $\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$

$E(X) = \frac{a}{a+b}$ $\quad Var(X) = \frac{\mu(1-\mu)}{a+b+1}$

Story: Bank / Post Office

Waiting time @ Bank $X \sim Gamma(a, \lambda)$
@ PO $Y \sim Gamma(b, \lambda)$

Total wait time is $X+Y \sim Gamma(a+b, \lambda)$

Fraction is $\frac{X}{X+Y} \sim Beta(a,b)$ ← not dependent on λ,

Properties: $\frac{X}{X+Y} \perp X+Y$
- $Beta(1,1) \sim Unif(0,1)$
- Beta is conjugate prior of Binomial.

$X \mid p \sim Bin(n,p)$, $p \sim Beta(a,b)$
$p \mid (X=x) \sim Beta(a+x, b+n-x)$

- Bayes' Billiards

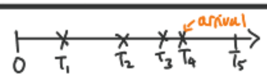
$$\int_0^1 \binom{n}{k} x^k (1-x)^{n-k}\,dx = \frac{1}{n+1}$$

1) n+1 balls, n white, 1 gray throw onto $[0,1] \sim Unif(0,1)$
X = # whites left of gray → pos. of gray ball

$P(X=k) = P(X=k \mid B=p) f(p)\,dp$
$= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k}\,dp$

2) n+1 white, paint 1 gray

Poisson Process



- random process; counts # of events that occurs within a time interval.

related to Exponential since arrival times for events are iid Expo(λ).

# of arrivals N in time interval $[0,t]$ is $Pois(\lambda t)$

Properties:
- # of arrivals in disjoint intervals are indpt. (i.e. (0,10), [10,12), [15,∞))
- Count-time duality
  N_t = # arrivals between time t
  T_n: time before n^{th} arrival
  $P(N_t < n) = P(T_n > t)$

Over time,
- time between transitions is Expo(λ)
- time that they occur are Pois(λt)

- $T_1 > t = N = 0$
Think of email sending story!

Ex) machines break down, work for Expo(λ) time. now 5 components work, E(machine breaks down)=
Breaks down if >1 component fails: memoryless
$\frac{1}{5\lambda} + \frac{1}{4\lambda} = \frac{9}{20\lambda}$

Normal $N \sim Norm(\mu, \sigma^2)$ $\quad E(X) = \mu$ $\quad Var(X) = \sigma^2$

PDF:
$$f(x) = \varphi(x) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$
→ even function

CDF:
- Standard Normal (CDF: Φ):
$Z \sim N(0,1)$ w/ $\mu = 0$, $\sigma^2 = 1$

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)\,dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt$$

Properties:
- Symmetric tail areas: $\Phi(z) = 1 - \Phi(-z)$
- Symmetry of Z and -Z: For $Z \sim Norm(0,1)$:
  $P(-Z \leq z) = P(Z \geq -z) = 1 - \Phi(-z)$
- Symmetry of PDF:
  $\varphi(z) = \varphi(-z)$ since even

Standardization

$N(\mu, \sigma^2) \to N(0,1)$

$\frac{X-\mu}{\sigma} \sim Norm(0,1)$

aka. for $X \sim N(\mu, \sigma^2)$, CDF of X is:

$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

PDF:
$f(x) = \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$

68-95-99.7 rule

for $Z \sim Norm(0,1)$:
$P(|Z| < 1) \approx 0.68$
$P(|Z| < 2) \approx 0.95$
$P(|Z| < 3) \approx 0.997$

ex: Folded Normal for E(X), Var(X), PDF(CDF)

let $Y = |Z|$ w/ $Z \sim N(0,1)$

a) E(Y): use LOTUS:

$$E(Y) = E(|Z|) = \int_{-\infty}^{\infty} |z| \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz$$
goes away
$$= 2\int_0^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz$$
$$= \sqrt{\frac{2}{\pi}}$$

$2 \cdot \frac{\sqrt{2\pi}}{2\pi} = \frac{\sqrt{2\pi}}{\pi} = \sqrt{\frac{2}{\pi}}$

b) Var(Y)

$Y^2 = Z^2$, so $E(Y^2) = E(Z^2) = Var(Z) = 1$

c) CDF and PDF

CDF: standardize

$y \leq 0$, CDF of Y is $F_Y(y) = P(Y \leq y) = 0$

$y > 0$:
$F_Y(y) = P(Y \leq y) = P(|Z| \leq y) = P(-y \leq Z \leq y)$
$= \Phi(y) - \Phi(-y) = 2\Phi(y) - 1$

ex: Standardization

Let $X \sim N(-1, 4)$. $P(|X| < 3) = ?$

$|X| < 3 = -3 < X < 3$. Using standardiz. and 68-95-99.7% rule:

$$P(-3 < X < 3) = P\left(\frac{-3-(-1)}{2} < \frac{X-(-1)}{2} < \frac{3-(-1)}{2}\right)$$
$$= P(-1 < Z < 2)$$
$$= \Phi(2) - \Phi(-1)$$

for approx. recall
$P(-1 < Z < 1) \approx 0.68$; $P(-2 < Z < 2) \approx 95$

from ±1 SD to ±2 SD =
$0.95 - 0.68 = 0.27$ (evenly divided)

$P(-1 < Z < 2) = P(-1 < Z < 1) + P(1 < Z < 2)$
$\approx 0.68 + \frac{0.27}{2} = 0.815$

$\boxed{\text{Odd moment of a normal is 0}}$

Gamma $X \sim Gamma(a, \lambda)$

Gamma function:
$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t}\,dt$$

Properties:
- n is positive, then $\Gamma(n) = (n-1)!$
- $n \in$ and $n \neq -1, 0$: $\Gamma(n+1) = n\Gamma(n)$

Gamma Distribution

PDF for $x \in [0, \infty)$
$$f(x) = \frac{1}{\Gamma(a)} (\lambda x)^a e^{-\lambda x} x^{-1}$$

$E(X) = \frac{a}{\lambda}$ $\quad Var(X) = \frac{a}{\lambda^2}$

Also,
$Gamma(1, \lambda) \sim Expo(\lambda)$

Convolutions of indpl. r.v.s

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

Chi-Squared and Student-t

(Chi-Squared (n)) $\quad \chi_n^2 \sim Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$

PDF
$$\frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}$$
for $x > 0$

$E(X) = n$
$Var(X) = 2n$

What is Chi-Sq:
$V = Z_1^2 + \dots + Z_n^2$ where Z_i are iid $N(0,1)$.
$V \sim \chi_n^2$, or V is Chi-Sq random w/ n deg of freedom.

Student-t:
$$T = \frac{Z}{\sqrt{V/n}}$$
where $Z \sim N(0,1)$ indpt $V \sim \chi_n^2$

Properties:
- Symmetry: if $T \sim t_n$, $-T \sim t_n$
- Cauchy $t_1 \sim$ Cauchy dist
- Convergence to Normal
  $n \to \infty$, $t_n \to N(0,1)$