Experiment $\rightarrow$ Outcome set of outcomes : S P is a function that maps any subset of S, called an "events", to $[0,1]$

Axiom 0: $P(A) \geq 0$ Axiom 1: $P(\phi) = 0$, $P(S) = 1$ Also, "$\{A_i\}$ are a partition of $A = \cup A_i$"

Axiom 2: $P(\cup A_i) = \sum P(A_i)$ iff $A_i \cap A_j = \phi \ \forall i \neq j$ i.e. A_i are disjoint.

$\boxed{\text{Foundations of Probability}}$

---

1. $P(A^c) = 1 - P(A)$ Since $A \cap A^c = \phi$, $A \cup A^c = S$ $P(S) = P(A) + P(A^c)$, $1 = P(A) + P(A^c)$

2. If $A \subseteq B$, $P(A) \leq P(B)$ $\because P(B) = P(A) + \underbrace{P(B \cap A^c)}_{\geq 0}$ [$B \cap A$ & $B \cap A^c$ are disjoint] $B \cap A = A$

Proof strategy 1: Disjointify your sets. Strategy 2: To show $P = Q$, show $P \subseteq Q$ & $Q \subseteq P$ by considering $\forall x \in P, x \in Q$ & then $\forall y \in Q, y \in P$.

$\boxed{\text{Set Theory}}$

3. $P(A \cup B) = P(A) + P(A^c \cap B)$, $P(B \cap A) + P(B \cap A^c) = P(B)$ $\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$

4. $P(\bigcup_{i \in [n]} A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \dots (-1)^{n+1} \bigcap_{i \in [n]} A_i$ [inclusion & exclusion] $\rightarrow$ 8. No of Derangements

5. $P(\bigcup_{i \in [n]} A_i) \leq \sum_{i \in [n]} P(A_i)$ [Union Bound] 6. $(\bigcup A_i)^c = \bigcap A_i^c$ [De Morgan's Laws] $\left[\frac{1}{2!} - \frac{1}{3!} + \dots \frac{(-1)^n}{n!}\right] n! \approx \frac{n!}{e}$

7. $\lim_{n \to \infty} P(\bigcup_{i \in [n]} A_i) = P(\bigcup_{i \in \mathbb{N}} A_i)$ "continuity of P" (not particularly relevant)

$\boxed{\text{Basic Counting}}$

	Order matters	Order doesn't matter
Replace	n^k	$\binom{n+k-1}{k}$
Don't Replace	${}^nP_k$	$\binom{n}{k}$

k distinguishable urns, n balls (indistinguishable)

$\underbrace{\circ\circ\circ}_{\text{urn 1}} | \underbrace{\circ\circ\circ}_{\text{urn 2}} | \circ || \underbrace{\circ}_{\text{urn n}}$ $n+k-1$ objects, $k-1$ lines

$\boxed{\Rightarrow \binom{n+k-1}{k-1}}$

$a_1 + a_2 + \dots a_r = n$, $a_i \in \mathbb{N} \cup \{0\}$ $\Rightarrow$ r urns, n indistinguishable balls $\Rightarrow \binom{n+r-1}{r-1}$ solutions exist

if $a_i \in \mathbb{N}$, then $\boxed{{}^{n-1}C_{r-1}}$ solutions exist. Give 1 to all. Now, $n-r$ balls to r urns. ${}^{n-r+r-1}C_{r-1} = {}^{n-1}C_{r-1}$

if $a_1 \leq k$ $(x_1^0 + x^1 \dots x^k)(x^0 \dots x^n) \dots \underbrace{(\ )}_{r^{th}\text{ set}}$ $\rightarrow$ The coefficient of x^n in the expansion. Use GP sums & discard high powers.

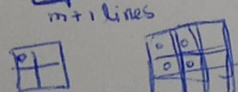
$a_1 + 4a_2 + 3a_3 = n \Rightarrow (x^0 + \dots x^n)((x^4)^0 + (x^4)^1 \dots (x^4)^{\lfloor \frac{n}{4} \rfloor})((x^3)^0 + \dots) \rightarrow$ coefficient of x^n

n pts, m collinear no of lines $= \boxed{{}^nC_2 - {}^mC_2 + 1}$ no of Δs $= \boxed{{}^nC_3 - {}^mC_3}$

The overcounted line has to be counted once.

In an $m \times n$ grid, there are ${}^{m+1}C_2 \, {}^{n+1}C_2$ rectangles $\underbrace{|||||}_{m+1 \text{ lines}}$ $\equiv \}$ $n+1$ lines and $\sum_{r=0}^{\min(m,n)} (m-r)(n-r)$ squares.

mn of size 1 $(m-1)(n-1)$ of size 2



Slide the dotted 4 tile around. Dotted tile cannot occupy the last column or row, thus, $(m-1)(n-1)$. Similarly, $(m-r)(n-r)$ of size r

---

$\boxed{\text{Conditioning}}$ Independence: A & B iff $P(A \cap B) = P(A)P(B)$ A, B & C iff pairwise independent, & $P(A,B,C) = P(A)P(B)P(C)$

$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = \boxed{P(A|B)P(B) = P(B|A)P(A)} \Rightarrow \boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$

$\boxed{P(A_1, A_2, \dots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1,A_2) \dots P(A_n|A_1,A_2,\dots,A_{n-1})}$

Conditional Independence $\Rightarrow P(A \cap B|C) = P(A|C)P(B|C)$ Ind $\not\Rightarrow$ Con ind; Con ind $\not\Rightarrow$ Ind

Conditioning can be very unintuitive: (i) P(2 aces | atleast one ace) = 1/33 P(2 aces | ace of ♠) = 1/17, almost twice.

(ii) Rare diseases & Baye's Rule; Sally Clark case; "The prosecutor's fallacy": confusing $P(A|B)$ for $P(B|A)$

(iii) Monty Hall problem; draw trees & condition, or split probabilities with law of total probability (LOTP) on conditional spaces.

(iv) Simpson's paradox → Dr. Nick vs Dr. Hilbert, Jelly bean jars, sporting averages.

$P(A|B,C) < P(A|B^c,C)$ but $P(A|B) > P(A|B^c)$

$P(A|B,C^c) < P(A|B^c,C^c)$ this is coz $P(A|B) = P(A|B,C)P(C|B) + P(A|B,C^c)P(C^c|B)$

Distribution	Range	PMF	Expectation, Variance	
1. Uniform (a,b)	$[a,b]\cap\mathbb{N}$	$\frac{1}{b-a+1}\ \forall x\in$ Range	$\frac{a+b}{2}$, $\frac{(b-a+1)^2-1}{12}$	
2. Bernoulli (p)	$\{0,1\}$	$\begin{cases}1: p\\ 0: 1-p\end{cases}$	p, $(1-p)p$	
3. Binomial (n,p)	$[0,n]\cap\mathbb{N}$	$P(r)=\binom{n}{r}p^r(1-p)^{n-r}$	np, $n(1-p)p$	← Number of successes in n Bernoulli Trials. Bino(n,p) = Sum of n iid Bern(p)
4. Geometric (p)	$\mathbb{N}$	$P(x)=(1-p)^{x-1}p$	$\frac{1}{p}$, $\frac{1-p}{p^2}$	← Number of trials for 1st success (iid Bern)
5. Negative Binomial (r,p)	$[r,\infty)\cap\mathbb{N}$	$P(k)=\binom{k-1}{r-1}p^r(1-p)^{k-r}$	$\frac{r}{p}$, $\frac{r(1-p)}{p^2}$	← Number of iid Bern(p) trials for r successes. = Sum of r iid Geo
6. Hypergeometric (n,r,k)	$[0,\min(k,r)]\cap\mathbb{N}$	$P(x)=\frac{{}^rC_x\ {}^{n-r}C_{k-x}}{{}^nC_k}$	$\frac{rk}{n}$, not worth it	← Urn with n balls, r red, rest blue. Pick k balls without replacement. No of red balls is the value.
7. Poisson (λ)	$\{0\}\cup\mathbb{N}$	$P(n)=e^{-\lambda}\frac{\lambda^n}{n!}$	λ, λ	← v. cool coz Poisson(μ)+Poisson(λ) = Poisson$(\mu+\lambda)$ (if independent)

Random Variables

$\langle x\rangle = \sum x P(X=x)$

$Var(x) = \langle x^2\rangle - \langle x\rangle^2$

Working out Sums

$n\binom{n-1}{k-1} = k\binom{n}{k}$ pick k members for a club, then a president, or pick a president & then the k−1 members.

$\sum_{i=0}^{k}\binom{m}{i}\binom{n}{k-i} = \binom{m+n}{k}$ pick k fruits from m apples & n oranges. Same as summing over possibilities.

3. $E(X) = \sum_{r=0}^{n} r\binom{n}{r}p^r q^{n-r} = \sum_{r=1}^{n} n\binom{n-1}{r-1}p^r q^{n-1-(r-1)} = np\,(p+q)^{n-1} = np$

$\frac{d}{dp}\left(\sum\binom{n}{r}p^r q^{n-r}\right) = \sum r\binom{n}{r}p^{r-1}q^{n-r} - \sum(n-r)\binom{n}{r}p^r q^{n-r-1}$ $\Rightarrow \frac{E(x)}{p} = \frac{n}{q} - \frac{E(x)}{q} \Rightarrow E(x) = np$

$\frac{d}{dp}\left(\sum_0^n\binom{n}{r}p^r q^{n-r}\right) = \sum_0^n\binom{n}{r}p^r(1-p)^{n-r}\left[\frac{r}{p}-\frac{n-r}{q}\right] = n\sum_0\binom{n}{r}p^{r-1}q^{n-r-1}r + \sum\binom{n}{r}p^{r-1}q^{n-r-1}r^2$

$0 = -\frac{n}{pq}E(X) + \frac{1}{pq}E(X^2) \Rightarrow E(X^2) = n^2p^2 + npq$ $\quad Var(X) = n^2p^2 + npq - n^2p^2 = npq$

4. $\sum_1^\infty(1-p)^{r-1}pr \Rightarrow \frac{d}{dp}\left[\sum_1^\infty(1-p)^r p\right] = \sum_1^\infty(1-p)^r - \sum_1^\infty(1-p)^{r-1}pr \Rightarrow \frac{d}{dp}(1-p) = \frac{1-p}{p} - E(x) \Rightarrow E(x) = \frac{1}{p} - 1 + 1 = \frac{1}{p}$

5. $\frac{d}{dp}\sum_{k=r}^\infty\binom{k-1}{r-1}p^r(1-p)^{k-r} = \sum_r\binom{k-1}{r-1}p^{r-1}(1-p)^{k-r}r - \sum_r\binom{k-1}{r-1}p^r(1-p)^{k-r-1}(k-r) = \frac{r}{p} - \frac{E(x)}{q} + \frac{r}{q} \Rightarrow \frac{r}{pq} = \frac{E(x)}{q}$ $\boxed{E(x) = \frac{r}{p}}$

6. $\sum_{x=0}^{k}\frac{\binom{r}{x}\binom{n-r}{k-x}x}{\binom{n}{k}} = \sum_1^k\frac{\binom{r-1}{x-1}\binom{n-r}{k-x}r}{\binom{n}{k}} = \frac{\binom{n-1}{k-1}r}{\binom{n}{k}} = \frac{rk}{n}$

7. $e^x = 1 + x + \frac{x^2}{2!} + \dots \frac{x^n}{n!} \dots \infty$ $\quad\Big|\quad \sum_{n=0}^\infty\frac{\lambda^n}{n!}n e^{-\lambda}$

$\sum_{n=1}^\infty\frac{\lambda^{n-1}}{(n-1)!} * \lambda e^{-\lambda} = \lambda$

3. Alt: Bin(n,p) = Sum of n Bern(p) $\boxed{E(X+Y) = E(X)+E(Y)}$ $\Rightarrow E(x) = np$

5. Neg Bin(r,p) = Sum of r Geo(p) $\Rightarrow E(x) = r/p$

6. X: no of red balls I_i: indicator of ith ball is red.

$X = \sum_1^k I_i \Rightarrow E(x) = \sum_1^k E(I_i) = k\left(\frac{r}{n}\right) = \frac{rk}{n}$

The argument in 6 is subtle: $E(I_2) = E(I_1)$ because in any picking, there's no reason for 2 to have better or worse odds in being red. This is by symmetry. However, I_1, I_2 etc. are dependent variables; Their distributions are not the same, as probabilities change with draws.

$\lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} \text{Bin}(n,p) = \text{Poisson}(\lambda)$ Eg: 3 Birthday problem

2 RVs are independent if $P(x_1=n_1, x_2=n_2, \dots) = P(x_1=n)P(x_2=n)\dots$ in the continuous case.
(Joint CDF) $f(x_1=n_1, x_2=n_2 \dots) = f(x_1=n)f(x_2=n)\dots$

This is equivalent to $P(x_1=n, x_2=n_2 \dots) = P(x_1=n)P(x_2=n_2)\dots$ in the discrete case only.

Joint & Marginal Independence & Conditioning

Joint CDF: $F_{XY}(x,y) = P(X\le x, Y\le y)$ ⇒ Joint PDF, for continuous, $f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)$

⇒ Marginal PDF, $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$, For discrete, $P_{XY}(x,y) = F_{XY}(x,y) - F_{XY}(x-1,y) - F_{XY}(x,y-1) + F_{XY}(x-1,y-1)$

⇒ Marginal PMF, $P_X(x) = \sum_{\forall y} P_{XY}(x,y)$ "Margins" but we typically work with joint PMF rather than CDF

Marginals ⇏ Joint
Joint ⇒ Marginals

For joint distributions, Area or nD area is probability. Lower dimensional entities are of measure zero & have 0 probability.

Conditional PDF: $f_{Y|X=x}(y) = \frac{f_{X|Y=y}(x)\, f_Y(y)}{f_X(x)}$ Similar process thing for functions $= f_{XY}(x,y)/f_X(x)$

Correlation $\rho(x,y) = \frac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\sqrt{\langle x^2\rangle - \langle x\rangle^2}\sqrt{\langle y^2\rangle - \langle y\rangle^2}}$

Independent ⇒ Uncorrelated
Uncorrelated ⇏ Independent

$\text{CoVariance}(X,Y) = \langle (X-\langle X\rangle)(Y-\langle Y\rangle)\rangle = \langle XY\rangle - \langle X\rangle\langle Y\rangle$

$\text{Uncorrelated} \equiv \langle XY\rangle = \langle X\rangle\langle Y\rangle$

Use conditioning & sum to find things with law of total probability. Very underappreciated tactic.

$\text{CoV}(X,Y) = \text{CoV}(Y,X)$ $\text{CoV}(cX,Y) = c\,\text{CoV}(X,Y)$ $\text{CoV}(X,Y+Z) = \text{CoV}(X,Y) + \text{CoV}(X,Z)$ $\text{CoV}(X,X) = \text{Var}(X)$

$\text{Var}(X_1+X_2) = \text{CoV}(X_1+X_2, X_1+X_2) = \text{CoV}(X_1,X_1) + \text{CoV}(X_1,X_2)\times 2 + \text{CoV}(X_2,X_2)$ ⇒ $\text{Var}(X_1+X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2\,\text{CoV}(X_1,X_2)$

Uncorrelated ⇒ $\text{Var}(X_1+X_2) = \text{Var}(X_1) + \text{Var}(X_2)$ $\text{Var}(X_1+X_2+\dots) = \sum_i \text{Var}(X_i) + 2\sum_{i<j} \text{CoV}(X_i,X_j)$ **CoV**

It's common to package Covariance information in a matrix, $M_{ij} = \text{CoV}(X_i,X_j)$. Diagonals are Variances, it's symmetric, & positive semidefinite.

$I_A \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$, ⇒ $\boxed{E(I_A) = P(A)}$ A bridge between probabilities & Expectation

$I_A = I_A^2 = I_A^n$ $I_A I_B = I_{A\cap B}$ Indicator RVs are super powerful, but great care needs to be taken in invoking symmetries.

3. $\text{Bino}(n,p) = X_1 + \dots X_n$ $X_i \sim \text{Bern}(p)$ $\text{Var}(X) = \sum \text{Var}(X_i) + 2\times 0 = npq$ Note that Bern = Indicator

6. $\text{HG}(n,x,k) = X_1 + \dots X_k$ X_i = Indicator (i^{th} ball is red) $\text{Var}(X) = k\times \frac{x}{n}\left(1-\frac{x}{n}\right) + 2\times\binom{k}{2}\text{CoV}(X_1,X_2)$

$\text{CoV}(X_1,X_2) = \langle X_1X_2\rangle - \langle X_1\rangle\langle X_2\rangle = \left(\frac{x}{n}\right)\left(\frac{x-1}{n-1}\right) - \left(\frac{x}{n}\right)^2$ ⇒ $\text{Var}(X) = kp(1-p)\cdot\left(\frac{n-k}{n-1}\right)$ ← Finite population correction.

4, 7 are easy to do computationally. 5 follows from 4 like 3 from 2.

Indicators & Variances

HG: pick balls without replacement
Binomial: with.
⇒ $\lim_{\frac{k}{n}\to 0} \text{HG} = \text{Bino}$
$k=1$ ⇒ HG = Bino
doesn't matter if you replace.

Functions of RVs Discrete: $P_Y(y) = \sum_{\forall x \mid f(x)=y} P_X(x)$ $E(Y) = \sum f(x)P(X=x)$ ← Same idea for continuous RVs.

Continuous: If invertible & differentiable (monotonic) map, $f_Y(y)\,dy = f_X(x)\,dx$ ⇒ $f_Y(y) = \frac{f_X(g^{-1}(y))}{\frac{dy}{dx}\big/_{x=g^{-1}(y)}}$ $y = g(x)$

$f_{\vec{Y}}(\vec{y}) = \frac{f_X(\vec{g}^{-1}(\vec{y}))}{\left|\frac{d\vec{y}}{d\vec{x}}\right|}$ ← abs of det of Jacobian $J_{ij} = \frac{\partial y_i}{\partial x_j}$

else, $F_Y(y)$ can be determined & $f_Y(y) = \frac{d}{dy}F_Y(y)$

Sum: $T = X+Y$ $\sum_x P(X=x)P(Y=T-x) = P(X)\otimes P(Y)$ continuous: $\int_{-\infty}^{\infty} f_X(x) f_Y(t-x)\,dx = f_X(x)\otimes f_Y(y)$

Convolution works only if X & Y are independent. Else sum/integrate on $x+y$ contour.

$T = \min(x,y)$ $\bar{F}_T(t) = \bar{F}_X(x)\bar{F}_Y(y)$ $T = \max(x,y)$ $(1-\bar{F}_T(t)) = (1-\bar{F}_X(x))(1-\bar{F}_Y(y))$

Random Remarks $E(X) = \sum P(X>k)$ if x is always +ve

$F_X(x) \to 0$ as $x\to-\infty$ Right continuous
$\to 1$ as $x\to\infty$ Strictly non decreasing

$E(X)$, $E(X^2)$ need not exist for well defined PMFs. Eg: $\frac{1}{x^2}$, $\frac{1}{x^3}$
Conditional convergence ≠ Convergence

MGFs • $M_X(s) = E(e^{sX}) = E\left[\sum \frac{s^n X^n}{n!}\right]$ ⇒ $\langle X^n\rangle = \frac{d^n}{ds^n}M_X(0)$

• MGF uniqueness → MGF determines the distribution completely & vice versa.

• For independent X, Y $M_{XY}(s) = E(e^{sX+sY}) = E(e^{sX})E(e^{sY}) = M_X(s)M_Y(s)$
convolution → product & vice versa. Laplace domain like. Note: $E(XY) = E(X)E(Y)$ if independent

2. $\text{Bern}(p)$ $M_X(s) = pe^s + q$ 3. $\text{Bin}(n,p)$ $M_X(s) = (pe^s+q)^n$ → Very easily get $\langle x^2\rangle$, $\langle x\rangle$

$M_{X+c}(s) = \int e^{(x+c)s} f_X(x)\,dx = e^{sc}M_X(s)$

$M_{kX}(s) = \int e^{skx} f_X(x)\,dx = M_X(ks)$

III Normal $\frac{e^{-z^2/2}}{\sqrt{2\pi}}$ → $\int_{-\infty}^{\infty} e^{+sz}\frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz = \int_{-\infty}^{\infty}\frac{e^{-\frac{1}{2}(z-s)^2}}{\sqrt{2\pi}}\,dz \times e^{+s^2/2} = \underline{\underline{e^{s^2/2}}}$ (Same functional form)

$X = \sigma Z + \mu$ (General normal) $M_X(s) = e^{\mu s}M_{\sigma Z}(s) = e^{\mu s}\boxed{M_Z(\sigma s)} = \underline{\underline{e^{\mu s + \frac{\sigma^2 s^2}{2}}}}$

II Exponential e^{-x} → $\int_0^{\infty} e^{sx}e^{-x}\,dx = \frac{1}{1-s} = \sum_{n=0}^{\infty} s^n = \sum_{n=0}^{\infty} n!\frac{s^n}{n!}$ ⇒ n^{th} moment of $\text{Exp}(1) = n!$
$\lambda = 1$

$e^{\frac{s^2}{2}} = \sum_{n=0}^{\infty}\frac{(s^2/2)^n}{n!} = \sum\frac{s^{2n}}{2^n n!}$

$Y = \text{Exp}(\lambda) = \lambda e^{-\lambda x} = \frac{1}{\lambda}\text{Exp}(1)$ ⇒ $M_Y(s) = \frac{1}{1-\frac{s}{\lambda}}$ ⇒ n^{th} moment is $\frac{n!}{\lambda^n}$
mean $= 1/\lambda$
$\lambda Y = X$ ← mean $= 1$

odd moments of normal are zero.
even moments are $\frac{(2n)!}{n!}\times\frac{1}{2^n}$ for $2n^{th}$ moment

7. Poisson(λ) $\sum_{n=0}^{\infty} e^{sn} e^{-\lambda} \frac{\lambda^n}{n!} = \sum \frac{(\lambda e^s)^n}{n!} \times e^{-\lambda} = e^{\lambda(e^s-1)}$

Poisson(μ) + Poisson(λ) $\Rightarrow M_{X+Y}(s) = e^{\lambda(e^s-1)} e^{\mu(e^s-1)} = e^{(\lambda+\mu)(e^s-1)}$ = Poisson($\mu+\lambda$)

(Assumed independence when $E(e^{X+Y}) = E(e^X)E(e^Y)$)

Some important Continuous RVs

	Mean	Variance	PDF, Range	MGF
I Uniform(a,b)	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{1}{b-a}$ $\forall x \in [a,b]$	
II Exponential (λ)	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\lambda e^{-\lambda x}$ $\forall x \in [0,\infty)$	$\frac{1}{1-s/\lambda}$
III Normal (μ,σ^2)	μ	σ^2	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $\forall x \in (-\infty,\infty)$	$e^{\mu s + \sigma^2 s^2/2}$
IV Beta (α,β)	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}\Gamma(\alpha+\beta)$ $\forall x \in [0,1]$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ is analogous to $\binom{\alpha+\beta}{\alpha}$
V Gamma(n,λ)	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$	$\frac{1}{\Gamma(n)}(\lambda x)^n e^{-\lambda x} \cdot \frac{1}{x}$ $\forall x \in [0,\infty)$	

I Uniform is universal. 1. $X = F^{-1}(U)$ then X has CDF F. $P(X \le x) = P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x)$

2. $F(X) \sim U$ $\quad c = P(X \le x) = P(F(X) \le c) \Rightarrow P(F(X) \le c) = c \Rightarrow F(X)$ is Uniform.

II 1. Poisson (λt) = no of customers entering shop in next hour. Find time till first customer. $P(t \geq T) = P(X_T = 0) = e^{-\lambda T} \frac{\lambda^0}{0!} = e^{-\lambda T}$

$\Rightarrow P(t \le T) = 1 - e^{-\lambda T} = F_T \quad \frac{dF_T}{dT} = \lambda e^{-\lambda T}$ = exponential(λ)

2. $Y = \lambda X \quad X \sim Exp(\lambda) \quad P(\lambda X \le t) = P(X \le \frac{t}{\lambda})$
$\hookrightarrow Y \sim Exp(1)$ $\quad$ CDF of Exp(1) $= 1 - e^{-\lambda(\frac{t}{\lambda})} \hookrightarrow = 1 - e^{-\lambda}$

3. The only memoryless distro is exp. (in discrete, it's geometric)

$P(X \geq t + t_0 | X \geq t_0) \overset{?}{=} P(X \geq t) \quad$ LHS $= \frac{P(X \geq t_0 + t,\ X \geq t_0)}{P(X \geq t_0)} = \frac{e^{-\lambda(t_0+t)}}{e^{-\lambda t_0}} = e^{-\lambda t} = P(X \geq t)$ QED

If $P(X \geq t + t_0 | X \geq t_0) = P(X \geq t)$, then, $\frac{P(X \geq t_0 + t)}{P(X \geq t_0)} = P(X \geq t) \Rightarrow G(t_0 + t) = G(t)G(t_0) \quad G = 1 - F$

Now $G(at) = G(t)^a$ $\forall a \in \mathbb{R}$ by subbing int, then $\frac{1}{\text{int}}$ then $\frac{p}{q}$ then $\lim \frac{p}{q}$ $\Rightarrow G(a) = e^{\ln G(1)\, a} = e^{-\lambda a}$ (say) = Exponential.

V 1. If Exp ≈ Geometric, Gamma ≈ Neg Bino. $\Gamma(a) = \int_0^\infty x^a e^{-x} \frac{dx}{x} \quad \Gamma(n+1) = n\Gamma(n) \quad \Gamma(1/2) = \sqrt{\pi}$ (cool fact) $= n!$ (for integral n) (Gaussian, actually)

$\frac{1}{\Gamma(a)} x^a e^{-x} \frac{1}{x}$ is a valid PDF, as it integrates to 1. It is Gamma(a, 1). $\quad X \sim$ Gamma(a,λ) $\lambda X = Y$

$\frac{1}{\Gamma(a)}(\lambda x)^a e^{-\lambda x} \frac{1}{x}$ $\sim$ Gamma(a,λ)

In a poisson process, t to 1st "success" is Exp(λ), t to n^{th} success is Gamma(n,λ)

$T_n = \sum X_i \quad X_i = t$ between i^{th} & $i-1^{th}$ successes; $M_{X_i} = \frac{1}{1-\frac{s}{\lambda}} \quad M_{T_n} = \left(\frac{1}{1-s/\lambda}\right)^n$

$M_{Gn\lambda}(s) = \frac{1}{\Gamma(n)} \int_0^\infty (\lambda x)^n e^{-\lambda x} \frac{1}{x} dx\, e^{sx}$ $\quad$ Let $u = (1 - \frac{s}{\lambda})\lambda x \Rightarrow \frac{1}{\Gamma(n)} \int_0^\infty \frac{(u)^n}{(1-s/\lambda)^n} e^{-u} \frac{du}{u} = \frac{\Gamma(n)}{\Gamma(n)(1-\frac{s}{\lambda})^n}$

$du = (1 - \frac{s}{\lambda})\lambda\, dx$

$= \frac{1}{(1-s/\lambda)^n}$ $\quad$ MGFs match, so T_n is $G_{n\lambda}$.

You go to a bank & need to go through a queue twice. The first time, there are n_1 people ahead of you. The next, n_2 ppl.

IV Consider $A \sim$ Gamma(n_1,λ) $B \sim$ Gamma(n_2,λ)

$X = A + B \sim$ Gamma($n_1 + n_2, \lambda$) by logic of story. $Y = \frac{A}{A+B}$, the fraction of time in 1st queue. Q: Is X independent of Y?

$f_{XY}(x,y) = f_{AB}(a,b) \left\| \frac{\partial(a,b)}{\partial(x,y)} \right\| = \frac{1}{\Gamma(n_1)\Gamma(n_2)} \times \frac{(\lambda a)^{n_1}}{a} e^{-\lambda a} \frac{(\lambda b)^{n_2}}{b} e^{-\lambda b} \times x$

$x = a + b \quad y = \frac{a}{a+b}$

$\left\| \begin{matrix} \frac{\partial x}{\partial a} & \frac{\partial y}{\partial a} \\ \frac{\partial x}{\partial b} & \frac{\partial y}{\partial b} \end{matrix} \right\| = \left\| \begin{matrix} 1 & \frac{1}{a+b} - \frac{a}{(a+b)^2} \\ 1 & \frac{-a}{(a+b)^2} \end{matrix} \right\| \quad \begin{matrix} a = xy \\ b = x(1-y) \end{matrix}$

$= \left| \frac{-a}{(a+b)^2} - \frac{b}{(a+b)^2} \right| = \left| -\frac{1}{a+b} \right| = \left| -\frac{1}{x} \right|$

$\Rightarrow f_{XY}(x,y) = \frac{1}{\Gamma(n_1)\Gamma(n_2)} \cdot \frac{(\lambda x)^{n_1} y^{n_1} (\lambda x)^{n_2} (1-y)^{n_2}}{xy\, x(1-y)} \times x e^{-\lambda x}$

$= \underbrace{\frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\Gamma(n_2)} y^{n_1-1}(1-y)^{n_2-1}}_{Y \sim \text{Beta}(n_1, n_2)} \times \underbrace{\frac{1}{\Gamma(n_1+n_2)} \frac{(\lambda x)^{n_1+n_2}}{x} e^{-\lambda x}}_{X \sim \text{Gamma}(n_1+n_2,\lambda)}$

Because f_{XY} factors nicely, it is independent.

$f_X(x) = \int_0^1 dy\, \underset{[\text{in } y]}{\text{Beta}(n_1,n_2)}\ \underset{[\text{in } x]}{\text{Gamma}(n_1+n_2,\lambda)}$

$=$ Gamma(n_1+n_2,λ) since normalisation works.

Thus, a single poisson process gives rise to the Exp, Beta, Gamma & of course, poisson distributions.

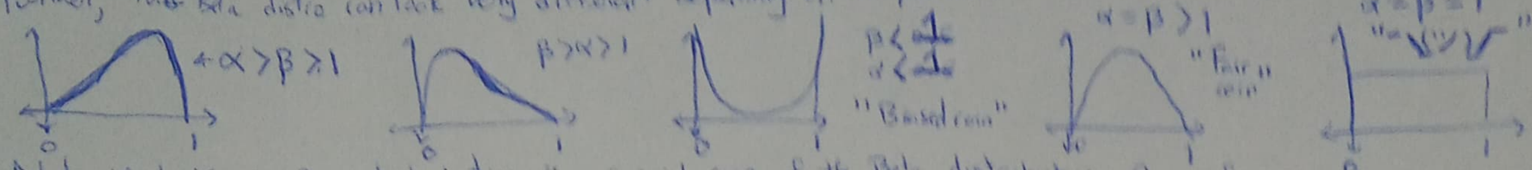
Yet another crucial property of the Beta distribution is that it is the conjugate prior for Binomial(n,p). If we do not know the p of an experiment, but we know that it is $|P_{,} \sim$ Bino(n,p), then we can use Bayesian updates to update our belief about the distribution of p. Let the prior $\sim$ Beta(α,β). Now, you observe k successes in n trials.

$$f_p(p|X=k) = \frac{f_x(X=k|p)\,f_p(p)}{f_x(X=k)} = \binom{n}{k} p^k (1-p)^{n-k}\, p^{\alpha-1}(1-p)^{\beta-1} \cdot \text{const} = \text{const}\cdot p^{\alpha+k-1}(1-p)^{\beta+n-k-1}$$

$f_x(X=k) \leftarrow$ const w.r.t. p

$$= \text{Beta}(\alpha+k,\ \beta+n-k)$$

The prior belief encoded that we had previously done the expt, got α successes & β failures. New data simply updates the parameters, without changing the form of the posterior, making Beta a very good choice of prior. Further, the Beta distr can look very different depending on the params.



Note that the Uniform distribution is a special case of the Beta distribution. A uniform prior is equal to having got one ghost failure & success. In a sense, the Beta can encode a hunch into the prior.

$\boxed{\text{Conditional Expectations}}$ ⚠ Handle with care. If 2 variables depend on each other, knowing the value of 1, tells you something about the other. $P(X|P) = \text{Bino}(n,P)$ is shorthand for $P(X|P=p) = \text{Bino}(n,p)$ (random variable, actual value)

In gen, keep in mind that $P(Y|X)$ is $P(Y|X=x)$ some known constant value.

1. $E(h(X)Y|X) = h(X)E(Y|X)$ 2. $E(Y|X) = E(Y)$ if independent 3. $E(E(Y|X)) = E(Y)$ Adam's law

4. Residual is uncorrelated with any function of X $\text{Cov}(Y - E(Y|X), h(X)) = 0 \Rightarrow E[(Y-E(Y|X))h(X)] = 0$ (residual, prediction)

$(\text{as } E[Y - E(Y|X)] = 0)$

$$\text{Var}(Y|X) = E(Y^2|X) - (E(Y|X))^2 = E((Y-E(Y|X))^2|X)$$

5. $\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X)) \leftarrow$ Eve's law

Adam's + Eve's Laws can be used to get μ, σ^2 for hard problems with dependence.

---

1. Cauchy Schwartz $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$ $E(XY)$ is the dot product of this vector space.
2. Jensen's inequality: if g is convex, $E(g(X)) \geq g(E(X))$ convex $= \cup$

   Eg: $E(X^2) \geq E(X)^2$ as $\text{Var} \geq 0$, $E(\frac{1}{X}) \geq \frac{1}{E(X)}$, $E(\ln X) \leq \ln(E(X))$

3. Markov's inequality $P(|X| \geq a) \leq \frac{E(|X|)}{a}$
4. Chebyshev: $P(|X-\mu| > k) \leq \frac{\sigma^2}{k^2}$

$g(x) \geq a + bE(x)$

$\Rightarrow g(x) \geq g(E(x))$

the function is above all tangents, particularly the tangent at $E(X)$ gives this.

$\boxed{\text{Inequalities}}$

Law of large nos: (strong) $\bar{X}_n \to \mu$ as $n \to \infty$ with probability 1.

Weak LLN: $P(|\bar{X}_n - \mu| > c) \to 0$ as $n \to \infty$

$$P(|\bar{X}_n - \mu| > c) \leq \frac{\text{Var}(\frac{\sum X}{n})}{c^2} = \frac{\frac{1}{n^2} n\sigma^2}{c^2}$$

$$\Rightarrow P(|\bar{X}_n - \mu| > c) \leq \frac{\sigma^2}{nc^2} \to 0 \text{ as } n \to \infty$$

$$\lim_{n\to\infty} \frac{n^{1/2}(\bar{X}_n - \mu)}{\sigma} \sim \text{Normal}(0,1)$$ "Central Limit Theorem" Equivalently, $\sum X_i$ (n of them) $= N(n\mu, n\sigma^2)$

$$\Rightarrow \frac{\sum X_i - n\mu}{\sqrt{n}\,\sigma} = N(0,1)$$

"Continuity correction" $P(X=a) \approx \int P(a-\frac{1}{2} < X < a+\frac{1}{2})\,dx$ (Discrete, Normal)

$\chi^2(n) = Z_1^2 + Z_2^2 \cdots Z_n^2$ $Z_i \sim$ iid Normal$(0,1)$ $\chi^2(1) = \text{Gamma}(\frac{1}{2},\frac{1}{2}) \Rightarrow \chi^2(n) = \text{Gamma}(\frac{n}{2},\frac{1}{2})$