cheatsheet

Sample Space: all the range of possible outcomes
event: any subset of the sample space
σ **algebra**: for set of events $\mathcal{F}$,sample space$S$: 1. nonempty $S \in \mathcal{F}$ 2. closed under complements $E \in \mathcal{F} \Rightarrow \bar{E} \in \mathcal{F}$ 3.closed under countable union $E_1, E-2 \cdots \in \mathcal{F} \Rightarrow \cup_i E_i \in \mathcal{F}$
probability measure: $(S, \mathcal{F})$
axioms 1. $\forall E \in \mathcal{F}, 0 \leq P(E) \leq 1$ 2. $P(S) = 1$ 3. if mutually exclusive$P\Big(\cup_i E_i\Big) = \sum_i P(E_i)$ 4. $P(\bar{E}) = 1 - P(\bar{E})$ 5. $P(\varnothing) = 0$
independent: $P(E \cap F) = P(E)P(F)$, if E and F independent, then $\bar{E}$ and F are independent
$P(E \cup F) = P(E) + P(F) - P(E \cap F)$
$P(E|F) = \frac{P(E \cap F)}{P(F)}$, if independent then additionally $= \frac{P(E)P(F)}{P(F)} = P(E)$
Law of total Probability$P(E) = \sum_i P(E|F_i)P(F_i)$
Bayes Theorem $P(E|F) = \frac{P(E)P(F|E)}{P(F)}$
Always consider whether two events are independent
Random Variable (r.v) | support of r.v. $supp(X) \equiv X(S) = \{x \in \mathbb{R} | \exists s \in S \text{ s.t. } X(s) = x\}$ (all the events with positive probability $supp(X) = S - x \in S | P(x) = 0$)
cdf,pdf: upper case for pdf, lower case for cdf
pmf the discrete version of pdf
expectation$E(X) = \sum_x xp(x)$, is also referred to as **mean** $E(aX + b) = aE(X) + b \forall a, b \in \mathbb{R}$ $E(g(X) + h(X)) = E(g(X)) + E(h(x))$
moment: expectation of$g(X) = X^n$, **central moment/variance**: expectation of $Var(X) = E[(X - E(X))^2] = E(X^2) - (E(X))^2$
linearity of variance $Var(aX + b) = a^2 Var(X)$
standard deviation: $sd(X) = \sqrt{Var_X(X)}$
skewness: $\gamma = \frac{E[(X-\mu)^3]}{\sigma^3}$

	Form	Mean	Variance	skewness
Bernoullli(p)	$p(x) = p^x(1-p)^{1-x}$	$\mu = p$	$\sigma^2 = p(1-p)$	
Binomial(n,p)	$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$	$\mu = np$	$\sigma^2 = np(1-p)$	$\gamma = \frac{1-2p}{\sqrt{np(1-p)}}$
Geometric(p)	$p(x) = p(1-p)^{x-1}$	$\mu = \frac{1}{p}$	$\sigma^2 = \frac{1-p}{p^2}$	
Poisson(λ)	$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$	$\mu = \lambda$	$\sigma^2 = \lambda$	$\gamma_1 = \frac{1}{\lambda}$
Uniform	$p(x) = \frac{1}{n}$	$\mu = \frac{n+1}{2}$	$\sigma^2 = \frac{n^2-1}{12}$	

$E_X(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$
$Var_X(x) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$
quantile: $Q_X(\alpha) = F_X^{-1}(\alpha)$

Uniform distribution: $U(a, b)$, $\frac{1}{b-a}$ only in (a, b), cdf $\frac{x-a}{b-a}$ in (a, b) $\mu = \frac{a+b}{2}, \sigma^2 = \frac{(b-a)^2}{12}$
Exponential distribution $Exp(\lambda)$, $f(x) = \lambda e^{-\lambda x}$, cdf $1 - e^{\lambda x}, x \geq 0, E(X) = \frac{1}{\lambda}, Var(X) = \frac{1}{\lambda^2}$
memoryless property $P(X > x + s | X > s) = \frac{P(X > x+s)}{P(X > s)} = \frac{e^{-\lambda(x+s)}}{e^{-\lambda x}} = e^{-\lambda x} = P(X > x)$

$$\begin{aligned} P(X > x + x_1 | X > x) = P(X > x_1) &\equiv P(N_x = 0) \\ &= \frac{(\lambda x)^0 e^{-\lambda x}}{0!} \\ &= e^{-\lambda x} \end{aligned}$$

Normal distribution $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}, F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\{-\frac{(t-\mu)^2}{2\sigma^2}\} dt$
N(0,1) standard normal distributon ($f(z) \equiv \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$)
some quantiles $\Phi(1.96) = 97.5\%, \Phi(2.58) = 99.5\%$w
mgf $M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$, or $\sum_{x_i \in supp(X)} e^{tx_i} p(x_i)$
moments and mgf $E[X^n] = \frac{d^n M_X(t)}{dt^n} \Big|_{t=0}$
$E[Z_1 Z_2] = E[Z_1]E[Z_2]$(this holds if independent) $M_{Z_1+Z_2} = M_{Z_1}(t) M_{Z_2}(t)$
joint cdf $F(x, y) = P_z(X \leq x, Y \leq y)$,$F_X(x) = F(x, \infty), F_Y(y) = F(\infty, y)$
must satisfy: 1. $0 \leq F(x, y) \leq 1, y \in \mathbb{R}$ 2. $x_1 < x_2 \Rightarrow F(x_1, y_1) \leq F(x_2, y_1)$ and $y_1 < y_2 \Rightarrow F(x_1, y_1) \leq F(x_1, y_2)$
$P_Z(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$
multinomial distribution $p(n_1, \ldots, n_r) = P_Z(X_1 = n_1, \ldots, X_r = n_r) = \frac{n!}{n_1! n_2! \ldots n_r!} q_1^{n_1} q_2^{n_2} \ldots q_r^{n_r}$
$F(x, y) = \int_{t=-\infty}^{y} \int_{s=-\infty}^{x} = f(s.t) ds dt$

$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$
independence random variables X and Y are independent if $F(x, y) = F_(x) F_Y(y)$
expectation of joint variables $E(g(X, Y)) = \sum_y \sum_x g(x, y) p(x, y)$, replace it with sum if discrete
Covariance $\sigma_{XY} = Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$
Correlation $\rho_{XY} = Cor(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$
multivariate normal distribution $f_X = \frac{1}{\sqrt{(2\pi)^n \det \sum}} \exp\Big(-\frac{1}{2}(x - \mu)^T \sum^{-1} (x - \mu)\Big)$ where x is a vector of random variables, they need not be independent
conditional pdf/pmf $p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x) p_X(x)}{p_Y(y)}$, $f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}$
bayes theorem in conditional context: $p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}, f_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}$
conditional cdfs: $F_{X|Y}(x|y) = P(X \leq x | Y = y) = \int_{u=-\infty}^{x} f_{X|Y}(u|y) dy$, $P(a < X \leq b | Y = y) = F_{X|Y}(b|y) - F_{X|Y}(a|y)$
conditional total probability law: $f_X(x) = \int_{y=-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) dy$ and $F_X(x) = \int_{y=-\infty}^{\infty} F_{X|Y}(x|y) f_Y(y) dy$
conditional expectation: $E_{Y|X}(Y|x) = \int_{y=-\infty}^{\infty} f_{Y|X}(y|x) dy$
$E_Y(Y) = E_X(E_{Y|X}(Y|X))$
Discrete Time Markov chains: $P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i) = (R)_{ij} = r_{ij}$
for example, if state 1 to state 2 is 0.5, then $r_{12} = 0.5$
π_0 denote the initial state $\pi_n = \pi_0 R^n, \pi_\infty = \pi_0 R^n, P(X_n = j) = \pi_\infty^*$, π_∞ can have multiple values, but there is only one π_∞^*
properties of DTMC 1. irreducible: if the matrix is storngly connected, for any state i and j, i can eventually reach j 2. periodic: if the time to return is a multiple of a fixed period, 3. if periodic and irreducible, then $\pi_\infty = \pi_\infty^*$ and the elements are strictly positive, $\pi_\infty R = \pi_\infty$ and $\sum_j \pi_{\infty,j} = 1$
sample and population: a sample is a subset of a population
bias-corrected variance: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$
efficiency: estimator T is more efficient than H if $\forall\theta, Var(T|\theta) \leq Var(H|\theta), \exists\theta, Var(T|\theta) < Var(H|\theta)$
T is efficient if T is more efficient than any other possible estimator
Consistency: T is a consistent estimator of the parameter θ if $\forall \epsilon > 0, P(|T(X) - \theta| > \epsilon) \to 0$ as $n \to \infty$
MLE (Maximum Likelihood Estimation): choosing a λ that maximizes the joint pdf: $L(\lambda) = f(X|\lambda) = f(x_1, \ldots, x_n|\lambda) = \prod_{i=1}^{10} f(x_i|\lambda)$ (independence needed), usually we use the log
CLT(Central Limit Theorem) $\lim_{n\to\infty} \frac{S_n - n\mu}{\sqrt{n}\sigma} \sim N(0, 1)$ or $\lim_{n\to\infty} \frac{X - \mu}{\sigma/\sqrt{n}}$
hypothesis testing: the null hypothesis is usually an equation, while the alternative $\neq$(two sided), $<, >$(one sided)
if we know the population mean, then we use the **z-test** otherwise we use the **t-test**
Confidence Intervals $[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$**Wrong assertion**: The probability that the true mean is contained in the confidence interval I created is 95%**Correct assertion**: approximately 95% of such intervals would cover μ
for (CI, still use t when population variance unknown)
Inverse Transform method: assume the $U = F(x)$ is something (usually U), get the distribution to sample X
Acceptance-Rejection method: find a g(x) easy to sample and $c = \max \frac{f(x)}{g(x)}$
convolution method: sample the individual distributions and sum the results
composition methods(discrete) $f(x) = \sum_{i=1}^{n} w_i f_i(x)$ and $w_i = P(Y = i), f_i(x) \equiv f(x|Y = i)$
pgf: $G(z) = E\{z^x\} = \sum_x z^x p(x)$
joint probability: $P_{XY}(X^{-1}(B_X) \cap Y^{-1} B_Y), B_X, B_Y \in \mathbb{R}$
exponential distribution P(X>Y) if X,Y has parameters λ, μ, then

$$\begin{aligned} P(X < Y) &= \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{y} f(x, y) dx dy \\ &= \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) dx dy \qquad \text{(by independence)} \\ &= \int_{y=-\infty}^{\infty} F_{X|Y}(y|y) f_Y(y) dy \\ &= \int_{0}^{\infty} (1 - e^{-\lambda y}) \mu e^{-\mu y} dy \\ &= 1 - \frac{\mu}{\lambda + \mu} = \frac{\lambda}{\lambda + \mu} \end{aligned}$$