Variance, Covariance

- Def (Var): For a r.v. X with $\mathbb{E}[X] = \mu$, then $\text{var}(X) = \mathbb{E}[(X-\mu)^2]$, and the standard deviation of X is $\sigma(X) = \sqrt{\text{var}(X)}$
- Th: For a r.v. X with $\mathbb{E}[X] = \mu$, then $\text{Var}(X) = \mathbb{E}[X^2] - \mu^2$
- HW: $\text{var}(cX) = c^2\ \text{var}(X)$, i.e. $\sigma(cX) = c \cdot \sigma(X)$
- For $X_n = I_1 + I_2 + \cdots + I_n$, then $\mathbb{E}[X_n^2] = \sum\limits_{i=1}^{n} \mathbb{E}[I_i^2] + 2\sum\limits_{i<j} \mathbb{E}[I_i I_j]$
- Th: For independent r.v. X, Y, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ and $\text{var}(X+Y) = \text{var}(X-Y) = \text{var}(X) + \text{var}(Y)$
- Def (Cov): $\text{Cov}(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
  - If X, Y are independent, then $\text{Cov}(X,Y) = 0$ (and so $\text{Corr}(X,Y) = 0$). However, the converse and negation are not true.
  - $\text{Cov}(X,X) = \text{var}(X)$
  - Cov is bilinear, for any collection of r.v. X, Y and fixed constants a, b, $\text{Cov}(\sum\limits_{i=1}^{n} a_i X_i, \sum\limits_{j=1}^{m} b_j Y_j) = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} a_i b_j \text{Cov}(X_i, Y_j)$
  - For general r.v. X, Y, $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$
- Def (Corr): Suppose X and Y are r.v. with $\sigma(X) > 0, \sigma(Y) > 0$. Then, $\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$
- Th: For any pair of r.v. X and Y with $\sigma(X) > 0, \sigma(Y) > 0$, then $-1 \leq \text{Corr}(X,Y) \leq +1$
- HW: If r.v. X has $\text{var}(X) = 0$, then X is a constant.
- HW: For r.v. X, Y, then $\mathbb{E}[\max(X,Y), \min(X,Y)] = \mathbb{E}[XY]$
- HW: If X, Y are independent r.v. with nonzero σ's, then it's not necessarily true that $\text{Corr}(\max(X,Y), \min(X,Y)) = \text{Corr}(X,Y)$ (Counter: $X, Y \sim \text{Bernoulli}(\frac{1}{2})$)
- Disc: If two r.v. are independent, they must have no relationship whatsoever, so they're uncorrelated. The converse, however, is not true: two uncorrelated variables may not be independent (Counter: Consider X and Y that follow a uniform joint distribution over the points $(1,0), (0,1), (-1,0), (0,-1)$).
- Disc: Two r.v. X, Y are independent $\iff$ The three criteria are equivalent and connected by Bayes rule: for all x, y such that $P(X=x), P(Y=y) > 0$, we need $P(X=x|Y=y) = P(X=x)$ or $P(Y=y|X=x) = P(Y=y)$ or $P(X=x, Y=y) = P(X=x)P(Y=y)$
- HW: Let $n \in \mathbb{Z}^+$ and $X_1, \ldots, X_n \sim \text{Uniform}[0,1]$ be i.i.d., then for $Y = \min\{X_1, \ldots, X_n\}$, $\text{var}(Y) = \frac{n}{(n+1)^2(n+2)}$
- Let $X_1, X_2, \ldots, X_n$ be n iid uniform r.v. on the interval $[0,1]$ (where $n \in \mathbb{Z}^+$). For $Y = \min\{X_1, \ldots, X_n\}, \mathbb{E}[Y] = \frac{1}{n+1}$; For $Z = \max\{X_1, \ldots, X_n\}, \mathbb{E}[Z] = \frac{n}{n+1}$
- Extra: Correlation is not necessarily transitive: if X correlates with Y, and Y correlates with Z, then this does not imply that X correlates with Z.

- Extra: Let X be a discrete r.v. that takes values in $[0,1]$, then $\text{var}(X) \leq \frac{1}{4}$ (Given the bounds, so $X^2 \leq X$, and so $\text{Var}(X) = \mathbb{E}[X^2] - \mu^2 \leq \mu - \mu^2 = \mu(1-\mu) \leq \frac{1}{4}$)
- Extra: Let X, Y be two r.v. and $Z = \min(X, Y)$. Then $E[Z] \leq \min(E[X], E[Y])$

Concentration Inequalities

- Th (Markov's Inequality): For a nonnegative r.v. X (i.e., $X(\omega) \geq 0 \; \forall \; \omega \in \Omega$) with finite mean and for any constant $c > 0$, then $\mathbb{P}[X \geq c] \;\leq\; \frac{\mathbb{E}[X]}{c}$
- Th (Generalized Markov's Inequality): Let Y be an arbitrary r.v. with finite mean, for any constants $c, r > 0$, then $\mathbb{P}[|Y| \geq c] \;\leq\; \frac{\mathbb{E}[|Y|^r]}{c^r}$
- Th (Chebyshev's Inequality): For any r.v. X with finite expectation $\mathbb{E}[X] = \mu$ and for any constant $c > 0$, then $\mathbb{P}[|X - \mu| \geq c] \;\leq\; \frac{\text{var}(X)}{c^2}$
  $\Longrightarrow$ Corollary: if X has finite $\sigma = \sqrt{\text{var}(X)}$, for any constant $k > 0$, then $\mathbb{P}[|X - \mu| \geq k\sigma] \;\leq\; \frac{1}{k^2}$
  $\Longrightarrow$ Chebyshev is the only one that could have a bound > 1 (when $c^2 \leq \sigma^2$)
  $\Longrightarrow$ For a probability p confidence interval for sampling n times with observed mean m, then the interval is: $m \pm \frac{\sigma}{\sqrt{n(1-p)}}$
- Th (Law of Large Numbers): Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v. with common finite expectation $\mathbb{E}[X_i] = \mu$ for all i. Then, for every $\epsilon > 0$, however small, their partial sums $S_n = X_1 + X_2 + \cdots + X_n$ satisfy

$$\mathbb{P}\big[|\frac{1}{n}S_n - \mu| < \epsilon\big] \longrightarrow 1 \text{ as } n \longrightarrow \infty$$

- Th (Central Limit Theorem, CLT): Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r,v, with common finite expectation $\mathbb{E}[X_i] = \mu$ and finite variance $\text{Var}(X_i) = \sigma^2$. Let $S_n = \sum_{i=1}^{n} X_i$, then the distribution of $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ converges to $N(0,1)$ as $n \to \infty$. In other words, for any constant $c \in \mathbb{R}$,

$$\mathbb{P}\Big[\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq c\Big] \to \Phi(c), \;\; i.e. \;\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} e^{-x^2/2}\, dx \text{ as } n \to \infty$$

  $\Longrightarrow$ For a probability p confidence interval for sampling n times with observed mean m, then we look for the k where p of the points are with $k\sigma$ (k standard deviations) of the mean, and so the confidence interval is: $m \pm \frac{k\sigma}{\sqrt{n}}$

  $\Longrightarrow$ For large n, we have: $\frac{S_n}{n} \sim N(\mu, \frac{\sigma^2}{n})$ approximately
  $\Longrightarrow$ Since CLT is an approximation not a bound, so points are on both sides of the curve
- To approximate a probability, use CLT; to bound, use Markov or Chebyshev.

Distributions

- For r.v. $X \sim$ Bernoulli(p), $\mathbb{P}[X = 1] = p = 1 - \mathbb{P}[X = 0]$ where $0 \leq p \leq 1$; $\mathbb{E}[X] = p$, $\text{var}(X) = p(1-p)$.
- For r.v. $X \sim$ Bin(n, p), $\mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, 1, \ldots, n\}$; $\mathbb{E}[X] = np$, $\text{var}(X) = np(1-p)$.
  Binomial distribution models the number of successes in a repeated experiment.
  - Th: Let $X \sim \text{Bin}(n, \frac{\lambda}{n})$ where $\lambda > 0$ is a fixed constant, then for every $i = 0, 1, \ldots, \mathbb{P}[X = i] \longrightarrow \frac{\lambda^i}{i!} e^{-\lambda}$ as $x \longrightarrow \infty$, i.e. the probability distribution of X converges to Poisson(λ).

- For r.v. $X \sim$ Geo(p), $\mathbb{P}[X = i] = p(1-p)^{i-1}$ for $i = 1, 2, \ldots$; $\mathbb{E}[X] = \frac{1}{p}$, $\text{var}(X) = \frac{1-p}{p^2}$.
  Geometric distrubution models how long we have to wait before a certain event happens. (Memoryless $\mathbb{P}(X = j + k \mid X > j) = \mathbb{P}(X = k)$)
  - $\mathbb{P}[X > k] = (1-p)^k$
- For r.v. $X \sim$ Poisson(λ), $\mathbb{P}[X = i] = \frac{\lambda^i}{i!} e^{-\lambda}$ for $i = 0, 1, \ldots$; $\mathbb{E}[X] = \lambda$, $\text{var}(X) = \lambda$.
  Models the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant rate and independently of the time since the last event.
  - Let $X \sim \text{Poisson}(\lambda), Y \sim \text{Poisson}(\mu)$ be independent r.v., then $X + Y \sim \text{Poisson}(\lambda + \mu)$ and $\mathbb{P}[X = k \mid X + Y = n] = \binom{n}{k}(\frac{\lambda}{\lambda+\mu})^k(1 - \frac{\lambda}{\lambda+\mu})^{n-k}$, i.e. $X|X + Y \sim \text{Bin}(X + Y, \frac{\lambda}{\lambda+\mu})$
- For continuous r.v. $X \sim$ Exp(λ) with $\lambda > 0$, it has CDF $\mathbb{P}[X \leq x] = 1 - e^{-\lambda x}$ and PDF $f(x) = \lambda e^{-\lambda x}$ if $x \geq 0$ and 0 otherwise. Then, $\mathbb{E}[X] = \int_0^\infty x \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}$, $\text{var}(X) = \frac{1}{\lambda^2}$ (Memoryless, i.e. $\mathbb{P}(X \geq s + t | X \geq s) = \mathbb{P}(X \geq t)$
  Models the time between events in a Poisson arrival process (continuous analogue of Geo).
  - For $X \sim \text{Exp}(\lambda)$ and any $t \geq 0$, then $\mathbb{P}[X > t] = \int_t^\infty \lambda e^{-\lambda x}\, dx = e^{-\lambda t}$
  - Let $U_1 \sim \text{Uniform}(0, 1)$, then $-\ln U_1 \sim \text{Expo}(1)$
  - Let $N_1, N_2 \sim \text{Normal}(0, 1)$ be iid, then $N_1^2 + N_2^2 \sim \text{Exp}(1/2)$
  - Disc: For $X_1 \sim \text{Exp}(\lambda_1), X_2 \sim \text{Exp}(\lambda_2)$, then $\mathbb{P}[\min(X_1, X_2) = X_1] = \mathbb{P}[X_2 \geq X_1] = \int_0^\infty \mathbb{P}[X_2 \geq X_1 \mid X_1 = x] f_{X_1}(x)\, dx = \frac{\lambda_1}{\lambda_1+\lambda_2}$
    (Extra: $Z = \min\{X_1, X_2\}$, then $Z \sim \text{Exp}(\lambda_1 + \lambda_2)$ Proof by calculating CDF; use induction for summing n Exp. distributions)
  - Extra: (Erlang) For $X_i \sim \text{Exp}(\lambda)$, let $Z = X_1 + X_2$, then $f_Z(z) = \lambda^2 z\, e^{-\lambda z}$; let $W = X_1 + \cdots + X_k$, then $f_W(w) = \frac{\lambda^k w^{k-1} e^{-\lambda w}}{(k-1)!}$
  - Extra: Let $X \sim \text{Exp}(\lambda), Y \sim \text{Exp}(\lambda)$ be iid, $M = \max(X, Y), L = \min(X, Y), Z = M - L$, i.e. $\max - \min$, then L, Z are independent, M, L and M, Z are dependent. (Proof for Z, L: Since X, Y independent, so $\mathbb{P}(M > m|L = l) = \mathbb{P}(X > m|X > l)$, and thus by memoryless property, $\mathbb{P}(Z > z|L = l) = \mathbb{P}(M > z + l|L = l) = \mathbb{P}(X > z + l|X > l) = \mathbb{P}(X > z)$
- For continuous r.v. $X \sim N(\mu, \sigma^2)$ with any $\mu \in \mathbb{R}$ and $\sigma > 0$, the normal distribution has PDF $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$. Then, $\mathbb{E}[X] = \mu$, $\text{var}(X) = \sigma^2$
  - Def: If $\mu = 0, \sigma = 1$, then X has the standard normal distribution.
  - Lemma: If $X \sim N(\mu, \sigma^2)$, then $Y = \frac{X-\mu}{\sigma} \sim N(0, 1)$. Equivalently, if $Y \sim N(0, 1)$, then $X = \sigma Y + \mu \sim N(\mu, \sigma^2)$
  - Corollary: Let $X \sim N(\mu_X, \sigma_X^2), Y \sim N(\mu_Y, \sigma_Y^2)$ be independent r.v., then for any constants $a, b \in \mathbb{R}$, the r.v. $Z = aX + bY$ is also normally distributed with $\mu = a\mu_X + b\mu_Y, \sigma^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$, i.e. $Z \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2)$
  - Th: Suppose X, Y are independent, then X_θ, Y_θ are independent $\iff$ X, Y are both Normal r.v. with the same variance σ^2 (Def the rotational/orthorgonal transformation as $\begin{bmatrix} X_\theta \\ Y_\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}$)
  - Normal distribution is symmetrical about the mean
  - From two uniform r.v. U_1, U_2, we can construct two standard normal distribution r.v. with $\sqrt{-2\ln(U_1)}\, \cos(2\pi U_2)$

- Variance for a uniform distribution in $[a, b]$ is $\frac{(b-a)^2}{12}$

- Th (Tail Sum Formula): Let X be a r.v. that takes values in $\{0, 1, 2, \dots\}$, then $\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}[X \geq i]$. For continous nonnegative r.v. Y, then $\mathbb{E}[Y] = \int_0^{\infty} \mathbb{P}[Y > y] \, dy$

- Poisson Arrival Process: Characterize a sequence of independent Bernoulli(p) trials
  - Bin(n, p) distribution governing the number of successes in n trials
  - Geo(p) distribution governing the waiting time to success
  - (Continuous analogous) Consider a random arrival process on $[0, \infty)$ satisfying the following property:
    1. For any fixed time interval $I \subset [0, \infty)$, the number $N(I)$ of arrivals in I is distributed as Poisson($\lambda \cdot |I|$), where $|I|$ denotes the length of interval I.
    2. If $I_1, I_2, \dots$ are disjoint time intervals, then $N(I_1), N(I_2), \dots$ are mutually independent.

    An equivalent description: For $i = 1, 2, \dots$, let W_i denote the waiting time to the i^{th} arrival, then
    1. $W_i \sim \text{Exp}(\lambda)$ for all $i = 1, 2, \dots$
    2. $W_1, W_2, \dots$ are mutually independent

- Application: For Coupon Collector/Safeway Monopoly, after i^{th} element, getting $(i+1)^{th}$ is like r.v. Geo($\frac{n-i}{n}$). Thus, $\mathbb{E}[X] = n \sum_{i=1}^{n} \frac{1}{i} \approx n \cdot (\ln n + \gamma_E)$ where $\gamma_E = 0.5772\dots$ is Euler's constant; and $\text{Var}(X) = n^2 (\sum_{i=1}^{n} \frac{1}{i^2}) - \mathbb{E}[X]$

- HW: Let $X_1, X_2 \sim \text{Exp}(\lambda)$ be independent, $\lambda > 0$. Then the density of $Y = X_1 + X_2$ is $f_Y(y) = \lambda^2 y \, e^{-\lambda y}$ for $y > 0$ and 0 otherwise.

- For r.v. $X \sim \text{HyperGeometric}(N, D, k)$, then $\mathbb{P}[X = b] = \frac{\binom{D}{b}\binom{N-D}{k-b}}{\binom{N}{k}}$ with mean $\mu = k\frac{D}{N}$, and $\text{var}(X) = k\frac{D}{N}\frac{N-D}{N}\frac{N-k}{N-1}$.
  Hypergeometric is a discrete probability distribution, modeling the probability of b successes in k draws, without replacement, from a finite population of size N that contains exactly D success states, wherein each draw is either a success or a failure. (Differs from Bin since Bin samples with replacement.)

Continuous Probability

- Def (Probability Density Function): A PDF for a real-valued r.v. X is a function $f : \mathbb{R} \to \mathbb{R}$ satisfying:
  1. f is non-negative: $f(x) \geq 0$ for all $x \in \mathbb{R}$.
  2. $\int_{-\infty}^{\infty} f(x) \, dx = 1$

  Then the distribution of X is given by: $\mathbb{P}[a \leq X \leq b] = \int_a^b f(x) \, dx$ for all $a < b$

- If $X \sim \text{Uniform}(0, k)$, then its density is $f(x) = 1/k$ for $0 \leq x \leq k$ and 0 otherwise.

- Def (Expectation): For continuous r.v. X, $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) \, dx$

- Def (Variance): For continuous r.v. X, $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_{-\infty}^{\infty} x^2 f(x) \, dx - \Big(\int_{-\infty}^{\infty} x f(x) \, dx \Big)^2$

- Def (Joint Density): A joint density function for two r.v. X, Y is a function $f : \mathbb{R}^2 \to \mathbb{R}$ satisfying:
  1. f is non-negative: $f(x, y) \geq 0$ for all $x, y \in \mathbb{R}$.
  2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dxdy = 1$

  Then the joint distribution of X, Y is given by: $\mathbb{P}[a \leq X \leq b, c \leq Y \leq d] = \int_c^d \int_a^b f(x, y)\, dxdy$ for all $a \leq b, c \leq d$

- Def (Independence for Continuous r.v.): Two continuous r.v. X, Y are independent if the events $a \leq X \leq b, c \leq Y \leq d$ are independent for all $a \leq b, c \leq d$:

$$\mathbb{P}[a \leq X \leq b, c \leq Y \leq d] = \mathbb{P}[a \leq X \leq b] \cdot \mathbb{P}[c \leq Y \leq d]$$

- Th: Joint density is $f(x, y) = f_X(x) f_Y(y)$ for all $x, y \in \mathbb{R}$; indiv. density can be calculated as $f(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$

- Application: Buffon's Needle - given a needle of length l, and a board ruled with horizontal lines at distance l apart
  - Since we assume a perfectly random throw, so the joint distribution has density $f(y, \theta)$, uniform over the rectangle $[0, l/2] \times [-\pi/2, \pi/2]$. Since this rectangle has area $\frac{\pi l}{2}$, so $f(y, \theta) = \frac{2}{\pi l}$ for $(y, \theta) \in [0, l/2] \times [-\pi/2, \pi/2]$ and 0 otherwise.
  - Let E denote the event that the needle crosses a line, then

$$\mathbb{P}[E] = \mathbb{P}[Y \leq \frac{l}{2} \cos\theta] = 2 \int_0^{\frac{\pi}{2}} \int_0^{l \cdot \cos\theta/2} f(y, \theta)\, dy d\theta = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \cos\theta\, d\theta = \frac{2}{\pi}$$

  - Equivalently, use indicator variable to form a circle of $d = 1$, with total length π and 2 intersections.

- HW: For a non-negative r.v. X with density f_X, one can extend the tail sum formula to give $\mathbb{E}[X^2] = \int_0^{\infty} 2s\, \mathbb{P}(X \geq s)\, ds$

- Extra: Conditional PDF for r.v. X on A is: $f_{X|A}(x) = \frac{f_X(x)}{\mathbb{P}[A]}$

Markov Chain

- For transition probability matrix $\mathbf{P}$, P_{ij} indicates the probability of from state i to state j

- $P_{ij} \geq 0\ \forall i, j \in \mathscr{S}$

- $\sum\limits_{j \in \mathscr{S}} P_{ij} = 1\ \forall i \in \mathscr{S}$, i.e. rows of P sum to 1

- The initial distribution is a row vector $\vec{\mu}^{(0)} = (\mu_i^{(0)} : i \in \mathscr{S})$ where $\mu_i^{(0)} \geq 0$ for all $i \in \mathscr{S}$ and $\sum_{i \in \mathscr{S}} \mu_i^{(0)} = 1$

- Th: For all $n \geq 0, \mu^{(n)} = \mu^{(0)} \mathbf{P}^n$. In particular, if $\mu_i^{(0)} = 1$ for some i, then $\mu_j^n = [\mathbf{P}^n]_{ij} = Pr[X_n = j | X_0 = i]$

- Hitting Time: Use first step equations (FSE) and solve them (possible due to recursion)
  - Let $\mathscr{A} \subset \mathscr{S}$ be a subset of states. For each $i \in \mathscr{S}$, let $\tau(i)$ be average number of steps until the Markov chain enters one of the states in $\mathscr{A}$, given that it starts in state i.

- FSE: $\tau(i) = 0$ if $i \in \mathscr{A}$, and $\tau(i) = 1 + \sum_{j \in \mathscr{S}} P_{ij} \cdot \tau(j)$ otherwise

- Def (Stationary or Invariant Distribution): A distribution $\vec{\pi} = (\pi_i : i \in \mathscr{S})$ is invariant (aka stationary) for the transition probability matrix $\mathbf{P}$ if it satisfies the following balance equations: $\vec{\pi} = \vec{\pi}\mathbf{P}$

- Th: $\vec{\mu}^{(n)} = \vec{\mu}^{(0)}$ for all $n \geq 0 \iff \mu^{(0)}$ is invariant (since by definition, $\vec{\mu}^{(n)} = \vec{\mu}^{(0)} P^n$)

- It is not necessarily true that X_{i+1} and X_{i-1} are uncorrelated. X_{i+1} only depends on X_i and is conditionally independent of X_{i-1}, but we do not know that X_{i+1} and X_{i-1} are independent.

- For 1-D random walk with absorbing boundaries (increase p, decrease $1-p$, stops at 0 and N)
  - Q1: $\alpha(i) = \mathbb{P}(\text{Hit 0 before } N \mid X_0 = i)$
  - Q2: $\tau(i) = \mathbb{E}(W \mid X_0 = i)$ where W = waiting time until either boundaries is hit
  - For $i \neq 0, N$: $\alpha(i) = (1-p)\,\alpha(i-1) + p\,\alpha(i+1)$ and $\tau(i) = 1 + (1-p)\,\tau(i-1) + p\,\tau(i+1)$ and also $\alpha(0) = 1, \alpha(N) = \tau(o) = \tau(N) = 0$
  - $\implies$ with $r = \frac{1-p}{p}$, then $\alpha(i) = 1 - \frac{i}{N}$ if $r = 1$ and $\frac{r^i - r^N}{1 - r^N}$ otherwise

Extra Sanity Checks

- When in doubt, always reduce to the very basic definitions, including $\text{Var}(X) = \mathbb{E}[X^2] - \mu^2$ and linearity of expectation

- When still in doubt, consider symmetry.

- $\mathbb{E}[g(X,Y)] = \sum_{x,y \in \mathbb{R}} g(x,y)\mathbb{P}[X = x, Y = y]$

- Extra: Let X, Y be r.v. The value of c that minimizes the variance of $X - cY$ is $\frac{cov(X,Y)}{var(Y)}$

- CLT provides a much smaller answer than Chebyshev. Because CLT is applied to a particular kind of r.v., namely the (scaled) sum of a bunch of rv. Chebyshev's inequality, however, holds for any r.v., and is therefore weaker. To approximate a probability, use CLT; to bound, use Markov or Chebyshev.

- For continuous r.v. with PDF $f(x)$, then $\int_{-\infty}^{\infty} f(x)\,dx = 1$

- To calculate the density (PDF) of a continuous r.v., first find its CDF $F(x)$, and then get $f(x) = \frac{dF(x)}{dx}$

- Extra: $\text{var}(X + Y) = \text{var}(X) + 2\,\text{Cov}(X,Y) + \text{var}(Y)$

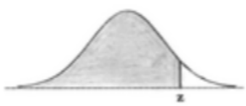
- X, Y are independent r.v. mod n. Don't know about X, but know Y is uniformly distributed. Then $Z = X + Y \pmod{N}$ is uniformly distributed.

- Def (Markov chain): A sequence of random variables $X_0, X_1, X_2, ...$ is a Markov chain if: $Pr[X_{t+1}|X_t] = Pr[X_{t+1}|X_t, X_{t-1}, ..., X_0]$ for all t and $Pr[X_{t'+1}|X_{t'}] = Pr[X_{t+1}|X_t]$ (i.e. the transition matrix P is always the same)

- Hashing, Load Balancing
  - Let A denote the event of having no collision. For m keys (balls) and n locations (bins), we want $\mathbb{P}[\overline{A}] \leq \frac{m^2}{2n} \leq \epsilon$, i.e. $m \leq \sqrt{2\epsilon n}$, using union bound.
  - A stricter bound would be: $\mathbb{P}[A] \approx e^{-\frac{m^2}{2n}} \geq 1 - \epsilon$, i.e. $m \leq \sqrt{2\ln(\frac{1}{1-\epsilon})} \cdot \sqrt{n}$, but still $O(\sqrt{n})$

- Load Balancing: Let A_k be the event that the load of some processor is at least k, then to find the smallest value k such that $\mathbb{P}[A_k] \leq p$, we can use the union bound to just inspect $\mathbb{P}[A_k(1)] \leq p \cdot \frac{1}{n}$ for identical bins.

- $\sum\limits_{k=0}^{\infty} \frac{x^k}{k!} = e^x$
- $\sum\limits_{k=1}^{n} \frac{1}{k} \approx \ln n + \gamma_E$ where $\gamma_E = 0.5772\ldots$ is known as Euler's constant.
- All sets of finite size objects are countable; there are countably infinite computer programs.
- Given a random variable $X \sim \text{Expo}(\lambda)$, consider the integer valued random variable $K = \lceil X \rceil$, then $\mathbb{P}[K = k] = e^{-\lambda(k-1)} \cdot (1 - e^{-\lambda}) \sim \text{Geo}(1 - e^{-\lambda})$
- A standard deck is 52 cards!
- Derangement (# of permutations of n elements with no fixed point): $D_n = n! \sum\limits_{k=1}^{n} \frac{(-1)^k}{k!} \approx \frac{n!}{e}$
- Extra: Suppose we draw cards from a standard 52-card deck without replacement, then the expected # of cards we must draw before we get all 13 spades (including the last spade) is $52 - 39 \cdot \frac{1}{14}$
- Extra: If $\mathbb{P}(A|B) = 1$, then $\mathbb{P}(B) \leq \mathbb{P}(A)$
- Extra: By Stirling's approximation, $\binom{2n}{n} 2^{-2n} \approx (\pi n)^{-1/2}$ for large n.
- Normal Cumulative Distribution Function: $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$. Also, the probability of a standard normal being within $\pm\epsilon$ is $\Phi(\epsilon) - \Phi(-\epsilon)$, and $\Phi(-x) = 1 - \Phi(x)$
- A random variable is a real-valued function of the outcome of a random experiment.
- For $Z \sim N(0,1)$, then $\mathbb{P}(|Z| \leq 1) = 0.68, \mathbb{P}(|Z| \leq 2) = 0.95, \mathbb{P}(|Z| \leq 3) = .997$



Probability Content from -oo to Z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990