Probability Cheatsheat

1 Random Variables

- Sample Space S: Set of all possible outcomes
  e.g {h,t} for a coin flip or {1,2,3,4,5,6} for a dice roll
- Event s: subset of the sample space
- Probability Measure P: Assigns every event a probability, $0 \leq P(s) \leq 1$ s.t the probabilities sum to 1
- Random Variable X: function from the sample space to $\mathbb{R}$
- Distribution of X: function from numbers to probabilities given by $x \to P(X^{-1}(x))$ ($X^{-1}(x)$ are elements of the sample space s.t X(s) = x)
  i.e Assigns every real number a probability based on the probability of the events that would cause X(s) to be that number
- Probability Mass Function $p_X(x)$: function from numbers to probabilities given by: $P(X = x) = p_X(x)$
- Probability Density Function $f_X(x)$: function integrated to convert number intervals to probabilities: $P(a \leq X \leq b) = \int_a^b f_X(x)dx$
- Cumulative Density Function $F_X(x)$: probability that X is less than x:
  $P(X \leq x) = F_X(x)$
  $$\text{Important Note: } \frac{d}{dx}F_X(x) = f_X(x)$$
- Change of Variables $X \to g(X) = Y$: if $g(X)$ is invertable on the support of X (x s.t $P(x = X) > 0$) with inverse $g^{-1}(x)$ then
  $$f_Y(y) = f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)|$$
  If g(X) is not invertable:
  $$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_y(y) \\ &= \frac{d}{dy}P(g(X) < y) \end{aligned}$$

2 Conditional Probability

- Conditional Probability: The probability that a occurs given that b occurred = $P(a|b) = \dfrac{P(a \cap b)}{P(b)}$
- Independent Events $a \perp b$: One event occurring does not effect the probability of the other occurring i.e $P(a|b) = P(a)$ or $P(a \cap b) = P(a) \cdot P(b)$
- Independant random variables $X \perp Y$: $f_{X,Y}(x, y) = f_X(x)f_Y(y)$
- Conditionally independant random variables $X \perp Y|Z$: $f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z)f_{Y|Z}(y|z)$
- Bayes Theorem: $P(A|B)\frac{P(B)}{P(A)} = P(B|A)$

- Chain Rule of Conditional Probabilities: $P(a, b, c) = P(a|b, c)P(b|c)P(c)$
- Conditional Distribution: $f_{X|Y}(x|y)\frac{f_{X,Y}(x,y)}{f_Y(y)}$

3 Marginal Probability

- Given a distrubution of a set of variables the distribution of a subset of the variables is called a marginal distribution
- To find a marginal distrubution: sum over the variables that you what to get rid of (Sum Rule)
- For discrete variables: $\Sigma_y p_{X,Y}(x, y) = p_X(x)$
- For continuous variables: $\int_{-\infty}^{\infty} f_{X,Y}(x, y)dy = f_X(x)$

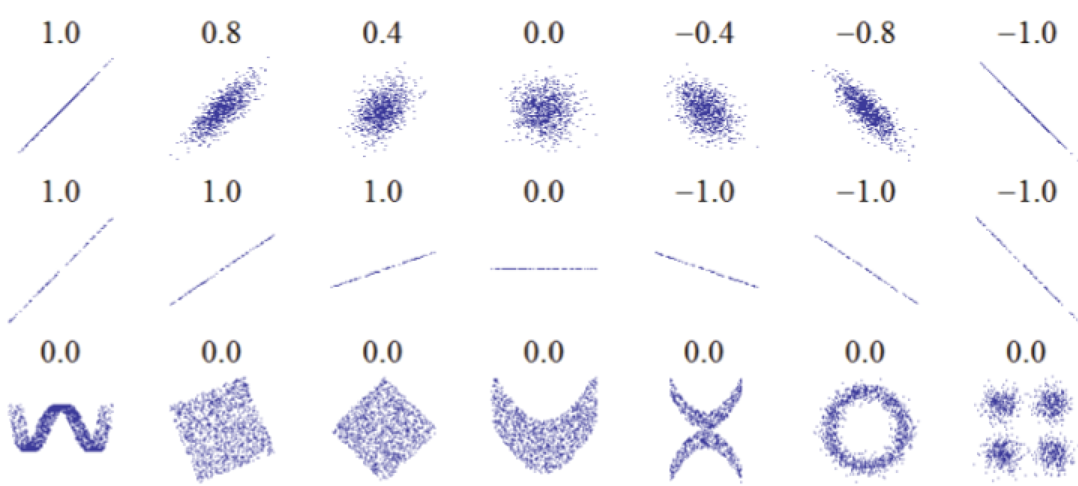
4 Functions of Distributions

- Expectation E(X): the average value of X = $\int x f_X(x)dx$
- Expectation of functions: E(f(X)) = $\int f(x)f_X(x)$
- Expection properties:
  - $E(aX) = aE(X)$
  - $E(X + Y) = E(X) + E(Y)$
  - $X \perp Y \rightarrow E(XY) = E(X)E(Y)$
- Varience Var(X): the average distance from the mean squared

$$Var(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$$

- Standard Deviation Std(X) = σ: the average distance from the mean = $\sqrt{Var(X)}$
- Covarience Cov(X,Y): how much two variable are linearly related to each other = $E((X - E(X))(Y - E(Y)) = E(XY) - E(X)E(Y)$
- Correlation Cor(X,Y): how much two variables are linearly related from -1 (high means low) to 1 (high means high)

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{var(X)var(Y)}}$$



Several sets of (x, y) points, with the correlation coefficient of x and y for each set.

5 Important Probability Mass Functions

- Bernoulli: number of heads in one coin flip with θ chance of head
  X $\sim$ Bernoulli(θ), $p_X(x) = (\theta^x(1-\theta)^{1-x}), x \in \{0,1\}$
- Binomial: number of heads in n coin flips with θ chance of head
  X $\sim$ Binomial(n,θ), $p_X(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, x \in \{0,1,\cdots,n\}$
- Geometric: sum of number of tails before head with θ chance of head
  X $\sim$ Geometric(θ), $p_X(x) = (1-\theta)^x\theta, x \in \{0,1,\cdots\}$
- Negative Binomial: sum of number of tails before r heads with θ chance of head
  X $\sim$ NegBin(r,θ), $p_X(x) = \binom{x+r-1}{x}(1-\theta)^x\theta^r, x \in \{0,1,\cdots\}$
- Poisson: number of events that occur over continuous interval with λ = expected number of events
  X $\sim$ Poisson(λ), $p_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}, x \in \{0,1,\cdots\}$

6 Important Probability Density Functions

- Uniform: Equal chance of occurring over the interval [L,R]
  X $\sim$ Uniform(L,R), $f_X(x) = \begin{cases} \frac{1}{R-L} & L \leq x \leq R \\ 0 & o/w \end{cases}$
- Exponential: Equal chance of occurring at any moment starting at 0
  X $\sim$ Exponential(λ), $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & 0 \leq x \\ 0 & o/w \end{cases}$
- Normal: limit of distribution when more and more random variables are summed together (log normal for variables multiplied together)
  X $\sim$ Normal(μ, σ^2), $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$ (standard normal if $\mu = 0, \sigma = 1$)
- Gamma: Used for integrating functions in the form $\int_0^\infty x^n e^{-mx}$
  X $\sim$ Gamma(α, β), $f_X(x) = \begin{cases} \dfrac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)} & 0 \leq x \text{ (where } \Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x} = (\alpha-1)!) \\ 0 & o/w \end{cases}$
- Beta: Used for integrating functions in the form $\int_0^1 x^n(1-x)^m$
  X $\sim$ Beta(α, β), $f_X(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} & 0 \leq x \leq 1 \\ 0 & o/w \end{cases}$

7 Convergence

- Convergence in distribution ($X_n \xrightarrow{\text{D}} X$): The sequence of random variables $X_1, X_2, \cdots$ converges in distribution to the random varible X iff:

$$\forall x \in \mathbb{R} \lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

- Convergence in probability ($X_n \xrightarrow{\text{P}} X$): The sequence of random variables $X_1, X_2, \cdots$ converges in probability to the random varible X iff:

$$\forall \epsilon > 0, \lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$$

- Convergence almost surely ($X_n \xrightarrow{\text{a.s.}} X$): The sequence of random variables $X_1, X_2, \cdots$ converges almost surely to the random varible X iff:

$$P(\lim_{n\to\infty} X_n = X) = 1$$

8 Important Theorems

- Weak law of large numbers: the sample mean converges in probability to the real mean (strong law is convegence almost surely)
  if $X_1, X_2, \cdots$ is a sequence of random variables and $\forall i, E(X_i) = \mu$ then:

$$\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{\text{P}} \mu$$

- Central limit theorem: the sample mean is roughly noramally distributed
  let $X_1, X_2, \cdots$ be i.i.d with $E(X) = \mu$ and $var(X) = \sigma^2$
  let $\bar{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$
  then: $\sqrt{n}(\bar{X_n} - \mu) \xrightarrow{\text{D}} N(0, \sigma^2)$
  or equivalently: $\frac{\sqrt{n}}{\sigma}(\bar{X_n} - \mu) \xrightarrow{\text{D}} N(0, 1)$