EECS126 Spring 2021: Formulae Reference

Arushi Somani

University of California, Berkeley

1 Probability Basics

1. Conditional Probability
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \mathbb{P}(B) > 0$$

2. Total Probability Theorem
$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{P}(B|A_i)$$

3. Bayes Rules
$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\mathbb{P}(B)}$$

4. Union Bound
$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

5. Independence
$$\mathbb{P}(A|B) = P(A) \iff \text{A and B are independent}$$

6. Conditional Independence
$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C) \implies \text{A and B conditionally independent}$$

7. Independence of Several Events
$$\mathbb{P}(\cap_{i \in S} A_i) = \prod_{i \in S} \mathbb{P}(A_i)$$

8. Counting Permutations of Size k in n Objects
$$^{n}P_k = \frac{n!}{(n-k)!}$$

9. Counting Ways to Choose k Objects in n Objects

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

10. Counting Ways To Partition n Objects into n^i Groups

$$\binom{n}{n_1, n_2 ... n_k} = \frac{n!}{n_1! n_2! ... n_k!}$$

2 Discrete Random variables

1. Bernoulli Random Variable

$$\mathbb{P}(X = k) = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

$$\mathbb{E}(X) = p$$
$$var(X) = p(1-p)$$

2. Binomial Random Variable

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mathbb{E}(X) = np$$
$$var(X) = np(1-p)$$

3. Geometric Random Variable

$$\mathbb{P}(X = k) = (1-p)^{k-1} p$$

$$\mathbb{E}(X) = \frac{1}{p}$$
$$var(X) = \frac{1-p}{p^2}$$

4. Poisson Random Variable

$$\mathbb{P}(X_\lambda = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

$$\mathbb{E}(X) = \lambda$$
$$var(X) = \lambda$$

5. Linearity of a Poisson RV

$$Poisson(\lambda) + Poisson(\mu) \sim Poisson(\lambda + \mu)$$

6. Uniform Random Variable

$$\mathbb{P}(X = k) = \begin{cases} \frac{1}{b-a+1} & k \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

7. Joint PMFs

$$\mathbb{P}_{X,Y}(x, y) = \Pr(X = x, Y = y)$$

$$\mathbb{P}_X(x) = \sum_y \mathbb{P}_{X,Y}(x, y) \text{ and vice versa}$$

8. Conditional PMFs

$$\mathbb{P}_{X|A}(X = x|A) = \frac{\mathbb{P}(\{X = x\} \cap A)}{\mathbb{P}(A)}$$

$$\mathbb{P}_{X_Y}(x|y) = \frac{\mathbb{P}_{X,Y}(x, y)}{\mathbb{P}_Y(y)}$$

3 Expectation, Variance and Covariance

1. Expectation

$$\mathbb{E}(X) = \sum_x x\mathbb{P}(X = x)$$

2. Law of The Unconscious Statistician

$$\mathbb{E}(g(X)] = \sum_x g(x)\mathbb{P}(X = x)$$

3. Variance

$$var(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] \geq 0$$

4. Standard Deviation

$$\sigma = \sqrt{var}$$

5. Linearity of Expectation

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

6. Expectation of Joint Distribution

$$\mathbb{E}(g(X, Y)) = \sum_x \sum_y g(x, y)\mathbb{P}_{X,Y}(x, y)$$

7. Variance of a Sum of Random Variables

$$var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$$

8. Conditional Expectation

$$\mathbb{E}(X|Y = y) = \sum_x x \; \mathbb{P}_{X|Y}(x|y)$$

9. Total Expectation Theorem

$$\mathbb{E}(X) = \sum_y \mathbb{P}_Y(y)\mathbb{E}(X|Y = y)$$

10. Iterated Expectation

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))$$

11. Tower Property

$$\mathbb{E}[\mathbb{E}[X|Y]g(Y)] = \mathbb{E}[Xg(Y)]$$

12. Expectation of Independent Variables

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \text{ if X, Y independent}$$

13. Covariance

$$cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

14. Correlation Coefficient

$$\rho(X, Y) = \frac{cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$
$$|\rho| \leq 1$$

15. Variance of Two Independent Variables

$$Var[XY] = \mathbb{E}[X^2]\mathbb{E}[Y^2] - \mathbb{E}[X]^2\mathbb{E}[Y]^2$$

16. Law of Total Variance

$$var(X) = Var(\mathbb{E}(X|Y)) + \mathbb{E}(var(X|Y))$$

4 Continuous Random Variables

1. Probability Density Functions

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)dx$$

2. Cumulative Ditribution Function

$$F_X(x) = \int_{-\infty}^x f(t)dt$$

3. Uniform Distribution

$$f_X(x) = \frac{1}{b-a},\ a < x < b$$
$$\mathbb{E}(X) = \frac{a+b}{2}$$
$$var(X) = \frac{(b-a)^2}{12}$$

4. Exponential Distribution

$$f_X(x) = \lambda e^{-\lambda x},\ x > 0$$
$$F_X(x) = 1 - e^{-\lambda x}$$
$$\mathbb{E}(X) = \frac{1}{\lambda}$$
$$var(X) = \frac{1}{\lambda^2}$$

5. Gaussian Distribution

$$f_X(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

6. Sum of Two Gaussian Variables

$$aN(\mu_1, \sigma_1^2) + bN(\mu_2, \sigma_2^2) \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

7. Joint PDFs

$$f_{X|Y}(x|y) = \frac{f_{X,y}(x,y)}{f_Y(y)}$$

8. Independence of Continuous Variables

$$f_{X,Y}(x,y) = f_x(x) f_Y(y)$$

5 Order Statistics

1. Smallest RV in a set of RVs

$$\text{Let } Y = \min_{1 \le k \le n} X_k \text{ , iid with CDF } F_X$$
$$F_Y(y) = 1 - (1 - F_X(y))^n$$

2. Largest RV in a set of RVs

$$\text{Let } Y = \max_{1 \le k \le n} X_k \text{ , iid with CDF } F_X$$
$$F_Y(y) = (F_X(y))^n$$

6 Convolution

1. Discrete Convolution

$$p_Z(z) = \mathbb{P}(X + Y = z) = \sum_x \mathbb{P}(X = x, Y = z - x)$$
$$= \sum_x \mathbb{P}_x(x)\mathbb{P}_Y(z - x) \text{ if X, Y independent}$$

2. Continuous Convolution

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx$$

7 Moment Generating Function

1. MGF for a RV

$$M_x(s) = \mathbb{E}[e^{sx}]$$
$$= \int_{-\infty}^{\infty} e^{sx} f_X(x) dx$$

2. Derivative of an MGF

$$\frac{d^n M(s)}{ds^n}|_{s=0} = \int x^n f(x) dx = \mathbb{E}[X^n]$$

3. MGF of a Poisson RV

$$M(s) = e^{\lambda(e^s - 1)}$$

4. MGF of a Exponential RV

$$M(s) = \frac{\lambda}{\lambda - s} \;, \; s < \lambda$$

5. MGF of the Standard Normal Gaussian RV

$$M(s) = e^{s^2/2}$$

6. Moments of Standard Normal RV

$$\mathbb{E}(X^m) = \begin{cases} 0 & \text{, m odd} \\ 2^{-m/2} \frac{m!}{(m/2)!} & \text{, m even} \end{cases}$$

7. MGF of a Geometric RV

$$M(s) = \frac{pe^s}{1 - (1-p)e^s}$$

8. MGF of a Bernoulli RV

$$M(s) = 1 - p + pe^s$$

9. MGF of a Binomial RV

$$M(s) = (1 - p + pe^s)^n$$

10. MGF of a Uniform RV

$$M(s) = \begin{cases} \frac{e^{bs} - e^{as}}{s(b-a)} & s \neq 0 \\ 1 & s = 0 \end{cases}$$

11. MGF of a Sum of RVs

$$\text{Let } Z = \sum X_i$$
$$M_Z(s) = \prod M_{X_i}(s)$$

12. MGF of a $Y = a^T X$, X is Gaussian Vector

$$M_Y(s) = M_X(sa) = \exp\left(s(a^T \mu_x) + \frac{1}{2} s^2 a^T \Sigma a\right)$$

8 Bounds

1. Markov Inequality

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

2. Chebyshev's Inequality

$$\mathbb{P}(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

3. Chernoff Bound

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[e^{sx}]}{e^{sa}} \text{ , } s > 0$$

$$\mathbb{P}(X \leq a) \leq \frac{M(s)}{e^{sa}} \text{ , } s \leq 0$$

4. Jensen Inequality

$$f(\mathbb{E}(x)) \leq \mathbb{E}[f(x)] \text{ , f is convex, } f''(x) > 0$$

5. Weak Law of Large Numbers

$$\lim_{n \longrightarrow \infty} \mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]| \geq \epsilon) = 0$$

6. Strong Law of Large Numbers

$$\mathbb{P}(\lim_{n \longrightarrow \infty} M_n = \mu) = 1$$

7. Central Limit Theorem

$$\text{Define } Z = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

$$F_Z(z) \rightarrow \phi(z)$$

9 Convergences

1. Almost Sure Convergence

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

2. Convergence in Probability

$$\lim_{n \longrightarrow \infty} P(|X_n - X| \geq \epsilon) = 0$$

3. Convergence in Distribution

$$\lim_{n \longrightarrow \infty} F_{X_n}(x) = F_X(x) \quad \forall x$$

10 Entropy

1. Entropy

$$H(X) = -\sum_{i=1}^{n} p_i \ln(p_i)$$

2. Chain Rule of Entropy

$$H(X,Y) = H(Y) + H(X|Y) = H(X) + H(Y|X)$$

3. Convergence of Joint Entropy

$$-\frac{1}{n} \log p(x_1, x_2 ... x_n) \xrightarrow{p} H(X)$$

11 Information Theory

1. Source Coding Theorem

   As $n \to \infty$, consider N iid RVs with entropy $H(X)$. You can compress this into no more and no less than $NH(X)$ bits without sending over extra bits or losing information.

2. Channel Coding Theorem

   Define channel capacity as the $\frac{\text{\# of message input bits}}{\text{\# of bits transmitted}}$. Any sequence of codes with error probability $p \to 0$ has a rate R < capacity.

3. Capacity of a BEC

$$C = 1 - p$$

4. Capacity of a BSC

$$C = 1 - H(p)$$

5. Asymptotic Equipartition (AEP) Theorem

$$A_\epsilon^{(n)} = \{(x_1, x_2 ... x_n) : p(x_1, x_2 ... x_n) \geq 2^{-n(H(X)+\epsilon)}\}$$
$$p((x_1, x_2 ... x_n) \in A_\epsilon^{(n)}) \xrightarrow{n \to \infty} 1$$
$$|A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$$

6. Average Number of Bits Transmitted

$$\mathbb{E}[\# \; bits] \leq n(H(X) + \epsilon)$$

7. Mutual Information

$$I(X;Y) = \sum p_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x) P_y(Y)}$$

8. Mutual Information and Entropy

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

9. Capacity of A Channel

$$C = \max_{p_x} I(X;Y)$$

10. Upper Bound on Probability of Error in BEC

$$P(\text{error}) = 2^{-n(1-p)+L(n)}$$

where n = # bits of bits sent and L = # of bits in message

12 Discrete Time Markov Chains

1. Markov Property

$$P(X_{n+1}|X_n...X_1) = P(X_{n+1}|X_n)$$

2. Chapman Komogorov Equations

$$P_{ij}^n = [P^n]_{ij}$$

3. Periodicity

$$d(i) = gcd\{n \geq 1 : P_{ii}^n > 0\}$$

4. Stationary Distribution

$$\pi P = \pi$$

5. Hitting Time

$$\beta(i) = \begin{cases} 1 + \sum_j p_{ij}\beta_j & i \notin A \\ 0 & i \in A \end{cases}$$

6. Detailed Balance Equations

$$\pi_i P_{ji} = \pi_i P_{ij}, \quad i,j \in S$$

7. Stationary Distribution of an Undirected Graph

$$\pi(i) = \frac{d(i)}{\sum_j d(j)} = \frac{degree(i)}{2E}$$

13 Poisson Processes

1. Number of arrivals within t

$$\mathbb{P}(N_t = n) \sim Poisson(\lambda t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

2. Inter-arrival Time

$$S_i \sim Exp(\lambda)$$

3. Sum of Inter-arrival Times: Erlang Distribution

$$f_{T_n}(s) = \frac{\lambda e^{-\lambda s}(\lambda s)^{n-1}}{(n-1)!}$$

4. Memoryless Property

$$N_{T_i} - N_{T_{i-1}} \sim Poisson(\lambda(t_i - t_{i-1}))$$

5. Poisson Merging

$$PP(\lambda_1) + PP(\lambda_2) \sim PP(\lambda_1 + \lambda_2)$$

6. Poisson Splitting

$$\mathbb{P}(\min\{T_a, T_b\} = T_a) = \frac{\lambda_a}{\lambda_a + \lambda_b}$$

7. Random Incidence Paradox

$$L \sim Erlang(2, k)$$

14 Continuous Time Markov Chains

1. Temporal Homogeneity

$$\mathbb{P}(X_{t+\tau} \mid X_t = i, X_s = i_s \forall\ 0 \leq s < t) = \mathbb{P}(X_\tau = j | X_0 = i)$$

2. Rate of Self-Transition

$$Q(i,i) = -\sum_{j \neq i} Q(i,j)$$

3. Balance Equations

$$\sum_{i \neq j} \pi_i Q(i,j) = \pi_j \sum_{k \neq j} Q(j,k)$$

4. Uniformization (Simulated DTMC)

$$\text{Let } q = \sup\ q(i)\ \text{, strongest self-loop}$$
$$R = I + \frac{1}{q} Q$$

5. Hitting Time

$$\beta(i) = \begin{cases} \frac{1}{q(i)} + \sum_{j \neq i} \frac{Q(i,j)}{q(i)} \beta(j) & i \notin A \\ 0 & i \in A \end{cases}$$

15 Random Graph

1. Probability of a Random Graph Being Given Graph

$$\mathbb{P}(G = G_0) \sim Binomial(\binom{n}{2}, p)$$

2. Distribution of Degree of Vertex in Random Graph

$$\mathbb{P}(D = d) \sim Binomial(n-1, p) \xrightarrow{n \to \infty} Poisson((n-1)p)$$

3. Erdos Renyi Theorem

$$Let\ p(n) = \lambda \frac{\ln(n)}{n}$$
$$\mathbb{P}(G \text{ is connected}) \xrightarrow{n \to \infty} 0\ ,\ \lambda < 1$$
$$\mathbb{P}(G \text{ is connected}) \xrightarrow{n \to \infty} 1\ ,\ \lambda > 1$$

4. Combining Graphs

$$\mathbb{P}(e \in G = G_1 \cup G_2 | e \in G_1 \cup e \in G2) = p_1 + p_2 - p_1 p_2$$

16 Statistical Inference

1. Bayes Rule Redux
$$\mathbb{P}(X = x | Y = y) = \frac{\mathbb{P}_{Y|X}(y|x)\pi(x)}{\sum_i \mathbb{P}_{Y|X}(y|i)\pi(i)}$$

2. Maximum A-Posteriori Estimation (MAP)

$$\text{MAP}(X|Y = y) = \text{argmax}_x P_{X|Y}(x|y) = \text{argmax}_x P_{Y|X}(y|x)\pi(x)$$

3. Maximum Likelihood Estimation (MLE)

$$\text{MLE}(X|Y = y) = \text{argmax}_x P_{Y|X}(y|x)$$

4. Likelihood Ratio
$$L(y) = \frac{P_{Y|X}(y|1)}{P_{Y|X}(y|0)}$$

5. MLE of a BSC
$$MLE(X|Y = y) = \begin{cases} y & \text{if } p \leq 1/2 \\ 1 - y & \text{if } p > 1/2 \end{cases}$$
$$MLE(X|Y = y) = \begin{cases} 1 & \text{if } L(y) \geq 1 \\ 0 & \text{if } L(y) < 1 \end{cases}$$

6. MAP of a BSC
$$\text{MAP}(X|Y = y) = \begin{cases} 0 & \text{if } L(y) < \frac{\pi_0}{\pi_1} \\ 1 & \text{if } L(y) \geq \frac{\pi_0}{\pi_1} \end{cases}$$

7. Likelihood Ratio for $X \in \{0, 1\}$ with Gaussian Noise

$$L(y) = \exp\left[\frac{y}{\sigma^2} - \frac{1}{2\sigma^2}\right]$$

8. MAP for $X \in \{0,1\}$ with Gaussian Noise

$$\text{MAP}(X|Y=y) = \begin{cases} 0 & \text{if } L(y) < \frac{\pi_0}{\pi_1} = y \geq \frac{1}{2} + \sigma^2 log(\frac{\pi_0}{\pi_1}) \\ 1 & \text{if } L(y) \geq \frac{\pi_0}{\pi_1} \end{cases}$$

9. MLE for $X \in \{0,1\}$ with Gaussian Noise

$$MLE(X|Y=y) = \begin{cases} 1 & \text{if } L(y) \geq 1 = y \geq \frac{1}{2} \\ 0 & \text{if } L(y) < 1 \end{cases}$$

17 Binary Error Testing

1. Neyman-Pearson Lemma

$$\text{Minimizes P(false negatives) with P(false positive)} \leq \beta$$

$$\hat{X} = \begin{cases} 1 & L(y) > \lambda \\ 0 & L(y) < \lambda \\ Bern(\gamma) & L(y) = \lambda \end{cases}$$

$$\text{Setting } \mathbb{P}(\hat{X} = 1|X = 0) = \beta$$

18 Estimations

1. Mean Square Error (MSE)

$$\mathbb{E}[(X - \hat{X}(Y))^2]$$

2. Minimum Mean-Squared Estimation (MMSE)

$$\text{MMSE}(X|Y) = \text{argmin}_{\hat{X}} \mathbb{E}[(X - \hat{X}(Y))^2] = \mathbb{E}(X|Y)$$

3. MMSE Theorem

$$E[(X - g(Y))f(Y)] = 0 \ \forall f \implies g(Y) = \text{ MMSE}$$

4. Linear Least Squares Estimation

$$LL[X|Y] = \min_{a,b_1 \ldots b_n} \mathbb{E}[|X - \hat{X}(Y)|^2] = \min_{a,b_1 \ldots b_n} \mathbb{E}[|X - (a + \sum b_i Y_i)|^2]$$

$$\text{Let Y be a vector of all observations } Y_i$$

$$\text{Define } \sum_{XY} = \mathbb{E}[(X - \mu_x)(Y - \mu_y)^T]$$

$$\sum_{Y} = \mathbb{E}[(Y - \mu_y)(Y - \mu_y)^T]$$

$$LL[X|Y] = \mu_x + \sum_{XY} \sum_{Y} {}^{-1}(Y - \mu_y)$$

$$LL[X|Y] = \mu_x + \frac{cov(X,Y)}{var(Y)}(Y - \mu_y)$$

5. Linear Least Squared Error

$$LLSE = var(X) - \sum_{XY} \sum_{Y}{}^{-1} \sum_{YX}$$

19 Hilbert Spaces

1. Hilbert Projection Theorem

$$\begin{aligned} \forall v \in H, U \subseteq H,\ \exists \min_{u \in U} ||u - v|| \ :\ \text{u is unique} \\ < u - v, u' >= 0\ \forall\ u' \in U \end{aligned}$$

2. Hilbert Random Variable Theorem

$$< X, Y >= \mathbb{E}[XY]$$

3. LLSE in Hilbert Spaces

$$< LL[X|Y] - X, u >= \mathbb{E}[(LL[X|Y] - X)u] = 0\ \forall\ u$$

4. Orthogonality Principle

$$\begin{aligned} \mathbb{E}(LL[X|Y]) = \mathbb{E}[X] \\ \mathbb{E}[(LL[X|Y] - X)Y_i] = 0 \\ \mathbb{E}[(LL[X|Y] \cdot Y^T] = \mathbb{E}[XY^T] \end{aligned}$$

5. Magnitude

$$||X|| = \sqrt{< X, X >} = \sqrt{\mathbb{E}(|X|^2)}$$

6. Zero-Mean Multiple RVs

Let X, Y, Z zero-mean

$$L[X|Y, Z] = L[X|Y] - L[X|Z - L[Z|Y]]$$

$$L[X|Y, Z] = L[X|Y] - L[X|Z] \text{ if Y, Z uncorrelated}$$