ST2334 Cheatsheet

Chapter 1 Introduction

- $P(S) = 1$, $P(\emptyset) = 0$
- ${}^nC_r = n! / r!(n-r)!$
- Equally-likely, frequency

Axioms of Probability

- $0 \le P(A) \le 1$
- Addition principle: sum partitions
- $A \subseteq B$, $P(B) = P(A)\ P(BA^c)$
- Inclusion-Exclusion Principle
- Mutually-exclusive $\Leftrightarrow$ $P(A \cap B) = 0$
- Independence $\Leftrightarrow$ $P(A \cap B) = P(A)*P(B)$
  - S and $\emptyset$ independent of any event
  - Complements are pairwise independent
  - Mutual independence => pairwise independent
- Conditional Probability (reduced sample space)
  - $P(B|A) = P(B \cap A)/P(A)$, $P(S|A) = 1$
- Rule of Total Probability
  - B = partition of S, $P(A) = \Sigma_{1,n}\ P(B_i\ A)$
- Bayes' Theorem
  - $P(B_k|A) = P(B_k)*P(A|B_k) / \Sigma\ P(B_i)*P(A|B_i)$
- Examples: Monty Hall, Bertrand's Paradox, Birthday Problem, Inverse Birthday Problem

Chapter 2 Random Variables

- $X: S \rightarrow \mathbb{R}$ (assigns number to each outcome)
- Collection of $(x_i, f(x_i))$ → probability distribution

Discrete RV: (pmf/ pf), $\Sigma f(x_i) = 1$
Continuous RV: (pdf) $f(x) \ge 0$, $= \int_{-\infty,\infty} f(x)\,dx = 1$

Cumulative Distribution Function

- $F(x) = P(X \le x)$, non-decreasing
- [D] $F(x) = \Sigma_{t \le x} P(X = t)$
- [C] $F(x) = \int_{-\infty, x} f(t)\,dt$, $f(x) = d/dx(F(x))$

Mean (weighting avg)

- [D] $\mu_X = E(X) = \Sigma_{xi}\ x_i\ P(X = x_i)$
- [C] $\mu_X = \int_{-\infty,\infty} x\ f(x)\ dx$
- *Properties*
  - $E(a + bX) = \Sigma\ (a + bx)P(X=x) = a + bE(X)$
  - $E[g(x)] = \Sigma/\int g(x)\ f_x(x)$
  - k^{th} moment: $g(x) = x^k$, $E[g(x)] = E(X^k)$

Variance

- $g(x) = (x - \mu_X)^2$, $V(X) = E(X^2) = [E(X)]^2$
- $V(X) = 0 \Rightarrow P(X = \mu_X) = 1$
- $V(a + bX) = b^2\ V(X)$
- $\sigma_x = SD(X) = \sqrt{V(X)}$

Chebyshev's Inequality

- Estimating distribution using E(X) and V(X)
- $P(|X - \mu| > k\sigma) \le 1/k^2$, $P(|X - \mu| \le k\sigma) \ge 1 - 1/k^2$

Chapter 3 Joint Distributions

- (X, Y) is a 2D RV
- [D] (joint pmf) $f_{X,Y}(x, y) = P(X = x, Y = y)$
  - Range is finite/ countably inf (draw table)
- [C] (joint pdf) $P((X, Y) \in D) = \iint f_{X,Y}(x, y)\ dy\ dx$

Marginal Distribution: $f_X(x)$ or $f_Y(y)$ alone

- Σ or $\int$ over the other variable

Conditional Distributions

- $f_{X|Y}(x|y) = f_{X,Y}(x, y) / f_Y(y)$, $f_Y(y) > 0$ (non-zero)
- $f_Y(y) > 0$, $f_{X,Y}(x, y) = f_{X|Y}(x|y) * f_Y(y)$
- For fixed y, $f_{X|Y}(x|y) \ge 0$
- [D] $\Sigma_x f(x_i) = 1$, [C] $\int_{-\infty,\infty} f_{X|Y}(x|y)\ dx = 1$

Independent RVs

- $f_{X,Y}(x, y) = f_X(x) * f_Y(y)$, $\forall x, y$
- $f_{X|Y}(x|y) = f_X(x)$, $\forall x$ x has same distribution $\forall y$
- Dependence: $f_{X,Y}(x, y) = 0$, but $f_X(x) > 0 \wedge f_Y(y) > 0$

Expectation and Covariance

- $E[g(X, Y)] = \Sigma\Sigma/\iint$ all $g(x, y) f_{X,Y}(x, y)$
- *Covariance* $g(X, Y) = (X - \mu_X)*(Y - \mu_Y)$
  - $Cov(X, Y) = E[(X - \mu_X)*(Y - \mu_Y)] = E(XY) - \mu_X\mu_Y$
  - $Cov(X, X) = V(X)$
  - $Cov(aX + b, cX + d) = ac*Cov(X, Y)$
  - $V(aX + bY) = a^2V(X) + b^2V(Y) + 2ab*Cov(X, Y)$
  - Independent => Cov(X, Y) = 0
- *Correlation Coefficient*:
  - $\rho_{X,Y} = Cov(X, Y) / \sqrt{V(X)V(Y)}$, $-1 \le \rho_{X,Y} \le 1$

Chapter 4 Common Probability Distributions

Uniform

- [Continuous] $X \sim U(a, b)$
- $f_x(x) = 1/(b-a)$ within range
- $E(X) = (a+b)/2$, $V(X) = (b-a)^2/12$
- $F_X(x) = 0 \rightarrow (x-a)/(b-a) \rightarrow 1$

Binomial: $f_x(x) = {}^nC_x * p^x(1-p)^{1-x}$, $x = \mathbb{N}_0$

- $X \sim B(n, p)$, $E(X) = np$, $V(X) = np(1-p)$
- # of successful independent Bernoulli trials
- Bernoulli: $f_x(x) = p^x(1-p)^{1-x}$, x = 0 or 1
- Geometric: # trials for 1 success
  - *Memoryless*: $P(X > n + k|X > n) = P(X > k)$
- Negative Binomal: $f_x(x) = {}^{x-1}C_{k-1} * p^kq^{x-k}$, x = k, k+1…
  - # trials for k success, Geom(p) = NB(1, p)
  - $E(X) = k/p$, $V(X) = k(1-p)/p^2$

Poisson: # success in interval

- Intervals independent, p(success) $\propto$ interval size
- $f_x(x) = e^{-\lambda}\lambda^x / x!$, $x = \mathbb{N}_0$
- $X \sim Poisson(\lambda)$, $E(X) = V(X) = \lambda$
- Approximates Binomial as $n \rightarrow \infty$ and $p \rightarrow 0$
  - $n \ge 20$, $p \le 0.05$ or $n \ge 100$, $np \le 10$
  - For large p, let q = 1-p
  - $\lambda = np$, fairly constant as $n \rightarrow \infty$

Exponential: $f_X(x) = \lambda e^{-\lambda x}$, $x > 0$, $\lambda > 0$

- $X \sim Exp(\lambda)$, $E(X) = 1/\lambda$, $V(X) = 1/\lambda^2$
- *Memoryless*, models duration btwn successes

Normal: symmetric and asymptotic

- Max of $1/(\sigma \sqrt{2\pi})$ achieved at $x = \mu$
- Standardisation: $Z \sim N(0,1)$, $Y = aZ + b$, $Y \sim N(b, a^2)$
- Binomial approximation: $n \rightarrow \infty$, $p \rightarrow 0.5$
  - $np > 5$ and $n(1-p) > 5$
  - Apply continuity correction of 0.5

Chapter 5 Sampling Distributions

- Statistics from samples are RVs, no unknowns
- Sample Mean
  - $E(\bar{X}) = E(X)$, $V(\bar{X}) = V(X)/n$
  - LLN: $P(|\bar{X} - \mu| > \varepsilon) \rightarrow 0$ as $n \rightarrow \infty$
  - CLT: $\bar{X} \sim N(\mu, \sigma^2/n)$ approximately, for $n \geq 30$
    - Large n not needed for approx. normal X
- Difference of Sample Means of 2 Populations
  - Two independent samples, $n \geq 30$
  - $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$
- χ^2-Distribution
  - $Y \sim \chi^2(n)$, $E(Y) = 2$, $V(Y) = 2n$
  - $Y_1 \sim \chi^2(n_1)$, $Y_2 \sim \chi^2(n_2)$, $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$
  - $Z^2 \sim \chi^2(1)$
  - *Sample Variance* from $X \sim N(\mu, \sigma^2)$
    - $S^2 = \Sigma(X_i - \bar{X})^2 / (n-1)$, $E(S^2) = \sigma^2$
    - $U = (n-1)S^2/ \sigma^2 \sim \chi^2(n-1)$, used later in F
- t-Distribution
  - Unknown population σ, use S to estimate
  - $T = Z/ \sqrt{(U/n)} \sim t(n)$, where $U \sim \chi^2(n)$
  - $T \sim t(n)$, $E(T) = 0$, $V(T) = n/(n-2)$
  - $T = (\bar{X}-\mu)/S \sqrt{n} \sim t(n-1)$
- F-Distribution
  - Ratio of two estimates of variance
  - $F = (U/n_1) / (V/n_2) \sim F(n_1, n_2)$

Chapter 6 Estimation based on ~N

Point Estimation

- $\hat{\theta}$ estimates unknown parameter θ
- Unbiased: $E(\hat{\theta}) = \theta$

Interval Estimation

- Confidence Interval with (1- α) confidence:
  - $P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \alpha$
- Estimating μ_X
  - Known σ, X ~ N or large n: $\bar{X} \pm z_{\alpha/2}\, \sigma/ \sqrt{n}$
  - Given error margin e: $n \geq (z_{\alpha/2}\, \sigma/e)^2$
  - Unknown σ, X ~ N and small n: $\bar{X} \pm t_{n-1;\alpha/2}\, S/ \sqrt{n}$
  - Unknown σ, X ~ N and large n: $\bar{X} \pm z_{\alpha/2}\, S/ \sqrt{n}$
- Estimation $\mu_1 - \mu_2$
  - Known $\sigma_1 \neq \sigma_2$, X_1, X_2~N or large n
    - $(\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2} \sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)}$
  - Unknown $\sigma_1 \neq \sigma_2$, large n: $\sigma_1, \sigma_2 \rightarrow S_1, S_2$
  - Unknown $\sigma_1 = \sigma_2$, X_1, X_2~N and small n
    - $S_p = [(n_1-1)S_1^2 + (n_2-1)S_2^2]/(n_1 + n_2 - 2)$
    - $(\bar{X}_1 - \bar{X}_2) \pm t_{n1+n2-2;\alpha/2} S_p \sqrt{(1/n_1 + 1/n_2)}$
  - Unknown $\sigma_1 = \sigma_2$, large n: $t_{n1+n2-2;\alpha/2} \rightarrow z_{\alpha/2}$
  - Paired data, estimating μ_D
    - Find D, S_D^2, $T = (\bar{D}- \mu_D)/ S_D \sqrt{n} \sim t_{n-1}$
  - Small n: $\bar{D} \pm t_{n-1;\alpha/2}\, S_D/ \sqrt{n}$
  - Large n: $t_{n1+n2-2;\alpha/2} \rightarrow z_{\alpha/2}$
- Estimating σ
  - X ~ N, Known μ: $[(X_i - \mu)/ \sigma]^2 \sim \chi^2(1)$
    - $(\quad \Sigma(X_i - \mu)^2/ \chi^2_{n;\alpha/2}\,, \quad \Sigma(X_i - \mu)^2/ \chi^2_{n;1-\alpha/2} \quad)$
  - X ~ N, Unknown μ: $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$
    - $\Sigma(X_i - \mu)^2 \rightarrow (n-1)S^2$, $\chi^2_{n;\alpha/2} \rightarrow \chi^2_{n-1;\alpha/2}$
- Estimating $(\sigma_1/ \sigma_2)^2$, using F-distribution
  - $(\quad (S_1/S_2)^2/F_{n1-1;n2-1;\, \alpha/2}, \quad (S_1/S_2)^2/F_{n2-1;n1-1;\, \alpha/2} \quad)$

Chapter 7 Hypothesis Testing based on ~N

- Null hypothesis vs. alternative hypothesis
- One-tailed vs. two-tailed
  - > check right region, < check left region
- Level of significance α $P(\text{reject } H_0 | H_0 \text{ true})$
- Power 1-β $P(\text{reject } H_0 | H_0 \text{ false})$
- Critical value → critical region
- p-value: observed significance level

Hypothesis Test on Mean

- H_0: $\mu = k$
- Known σ, X~N or large n (CLT): $(\bar{X}- \mu)/\sigma \sqrt{n} \sim Z$
- Unknown σ, X~N; $(\bar{X}- \mu)/S \sqrt{n} \sim t_{n-1}$
- Using CI: $\bar{X}$ → CI for μ_X, see if H_0 falls in CI
  - Only for 2-sided tests

Hypothesis Test on Difference of Two Means

- Independent Samples
  - H_0: $\mu_1 - \mu_2 = \delta_0$
  - Known σ_1, σ_2, X_1, X_2~N or large n
    - $(\bar{X}_1 - \bar{X}_2) - \delta_0/ \sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)} \sim N(0, 1)$
  - Unknown σ_1, σ_2, large n: $\sigma_1, \sigma_2 \rightarrow S_1, S_2$
  - Unknown σ_1, σ_2, X_1, X_2~N, small n
    - $[(\bar{X}_1 - \bar{X}_2) - \delta_0]/ [S_p \sqrt{(1/n_1 + 1/n_2)}] \sim t_{n1+n2-2}$
- Paired samples (H_0: $\mu_D = \mu_{D,0}$)
  - D ~ N, small n: $(\bar{D} - \mu_{D,0})/S_D \sqrt{n} \sim t_{n-1}$
  - Large n: $(\bar{D} - \mu_{D,0})/S_D \sqrt{n} \sim N(0, 1)$

Hypothesis Test on Variance

- H_0: $\sigma^2 = \sigma_0^2$, $(n-1)S^2/ \sigma^2 \sim \chi^2(n-1)$

Hypothesis Test on Ratio of Variances

- H_0: $\sigma_1^2 = \sigma_2^2$, $(S_1/S_2)^2 \sim F(n_1-1, n_2-1)$