ST2334 Midterm Cheat-sheet

Part 1 Statistical Description

1. Random variable: categorical (nominal, ordinal) & numerical (discrete, continuous).
2. Sampling: random (quota, convenience, judgment) & non-random (simple random, systematic, cluster – units in a cluster are like the population, stratified – units in a stratum are homogeneous).
3. Ways to display categorical variables: pie chart, bar chart, pareto chart.
4. Ways to display numerical variables: histogram, stem-and-leaf plot, dot-plot.
5. Measures of location: mean, median, mode (unimodal, bimodal, trimodal) & percentile (quantile).
6. Measures of spread: variance (standard deviation), coefficient of variation (CV), range, inter-quartile range (IQR), box plot (box-and-whisker plot), five-number summary.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$$

7. Shape of a distribution: symmetric & skewed (left, right). For a left-skewed distribution, its mean value is usually smaller than its median; for a right-skewed one, its mean is usually larger than its median.

Part 2 Probability

1. Terms: experiment, outcome, sample space, event.
2. Event: null, union, intersection, complement, mutually exclusive, independent.
3. Interpretation of probability: equally-likely outcomes, frequency interpretation, personal probability.
4. Conditional probability: $P(B|A) = \frac{P(A\cap B)}{P(A)}$
5. Two events are independent if and only if: 1) $P(A\cap B) = P(A)P(B)$; 2) $P(A|B) = P(A)$; 3) $P(B|A) = P(B)$.
6. Partition: If $B_1, B_2, \dots, B_n$ are mutually exclusive and $B_1 \cup B_2 \cup \dots \cup B_n = S$, we call them being a partition of S.
7. Total probability: for any partition of S, we have $P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i)$.
8. Bayes' Theorem: $P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}$.
9. Two non-trivial mutually-exclusive events must be dependent.

Part 3 Discrete Random Variable

1. Probability mass function (PMF): 1) $f(x_i) \geq 0$; 2) $\sum f(x_i) = 1$.
2. Variance: 1) $Var(a + bX) = b^2 \cdot Var(X)$; 2) $Var(X) = E(X^2) - (E(X))^2$.
3. Bernoulli distribution: If $W \sim B(n,p)$, then $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, we also have $E(x) = np$ & $Var(x) = np(1-p)$.
4. Geometric distribution: If $X \sim Geom(p)$, then $P(X = k) = (1-p)^{k-1}p$, we also have $E(x) = \frac{1}{p}$ & $Var(x) = \frac{1-p}{p^2}$, with CDF being $F(x) = 1 - (1-p)^x$.
4. Poisson distribution: If $X \sim Pois(\lambda)$, then $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$, we also have $E(x) = \lambda$ & $Var(x) = \lambda$.
5. When $n \geq 20, p \leq 0.05$ or $n \geq 100, np \leq 10$, we can use Poisson to approximate Binomial by $B(n, \lambda/n) \approx Pois(\lambda)$.
6. Cumulative distribution function (CDF): $F(x) = P(X \leq x)$,non-decreasing. CDF can be applied to discrete & continuous random variables.

Part 4 Continuous Random Variable

1. Probability density function (PDF): 1) $f(x_i) \geq 0$; 2) 1 $\int_{-\infty}^{\infty} f(x)\,dx = 1$.
2. Mean: $\mu = E(X) = \int_{-\infty}^{\infty} x\, f_x(x)\,dx$.
