STA 4321 REVIEW (Chapters 1 - 5)

1 Probability (Chapter 1)

1.2 Sample Spaces

- Experiment: any process of data collection or observations where the outcomes are subject to variation.

ex1) toss a coin three times.
ex2) observe # of jobs in a print queue of a mainframe computer.
ex3) observe the length of time between successive earthquakes in L.A.

- Sample space (Ω): the set of all possible outcomes

ex1) $\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.
ex2) $\Omega = \{0, 1, 2, 3, \ldots\}$.
ex3) $\Omega = \{t \in \mathcal{R} | t \geq 0\}$.

- Event: subset of Ω.

ex1) the first coin is Head, $A = \{HHH, HHT, HTH, HTT\}$.
ex2) fewer than five jobs, $A = \{0, 1, 2, 3, 4\}$.
ex3) the length of time is less than 10 seconds, $A = \{0 \leq t < 10\}$.

- Event relation

(i) Union: $C = A \cup B$, either A occurs or B occurs or both occur.
(ii) Intersection: $C = A \cap B$, both A and B occur.
(iii) Complement: A^c, A does not occur.
(iv) Disjoint: $A \cap B = \phi$.

- Laws of set theory

(i) Commutative laws: $A \cup B = B \cup A$, $A \cap B = B \cap A$.
(ii) Associative laws: $(A \cup B) \cup C = A \cup (B \cup C)$,
$(A \cap B) \cap C = A \cap (B \cap C)$.
(iii) Distributive laws: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$,
$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$.

1.3 Probability Measures

- $P : \Omega \rightarrow \mathcal{R}$ satisfies

(i) $P(\Omega) = 1$,

(ii) $0 \le P(A) \le 1$ for any $A \subset \Omega$,

(iii) $A_1 \cap A_2 = \phi \Rightarrow P(A_1 \cup A_2) = P(A_1) + P(A_2)$.
More generally, if $A_1, A_2, \ldots, A_n, \ldots$ are mutually disjoint, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

By three properties above,

P1. $P(A^c) = 1 - P(A)$
P2. $P(\phi) = 0$
P3. If $A \subset B$, then $P(A) \le P(B)$
P4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

1.4 Computing Probabilities: counting methods

- Uniform model: all the outcomes are equally likely.

$$P(A) = \frac{\#\text{ of ways } A \text{ can occur}}{\text{total }\#\text{ of outcomes}}$$

- More complex situations

  (i) Multiplication principle
  exp't 1: m outcomes, exp't 2: n outcomes $\Rightarrow$ mn outcomes
  exp't 1: n_1, exp't 2: n_2, exp't p: $n_p \Rightarrow n_1 n_2 \cdots n_p$ outcomes

  (ii) Sampling without replacement and with replacement
  For a set of size n and a sample size r,
  # of different ordered samples with replacement: n^r,
  without replacement: $n(n-1)\cdots(n-r+1)$,
  (if $n = r$, $n(n-1)\cdots 1 = n!$.)

  (iii) Binomial coefficients
  $\binom{n}{r}$: # of unordered samples without replacement

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

  In particular, $2^n = \sum_{k=0}^{n} \binom{n}{k}$.

  (iv) Multinomial coefficients

$$\binom{n}{n_1 n_2 \cdots n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

  : # of ways that n objects can be grouped into r classes with n_i in the ith class, $i = 1, \ldots, r$ and $\sum_{i=1}^{r} n_i = n$.

$$(x_1 + x_2 + \cdots + x_r)^n = \sum \binom{n}{n_1 n_2 \cdots n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}$$

1.5 Conditional Probability

- Conditional probability

  $$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0.$$

  : conditional probability of A given B.

- Multiplication law

  $$P(A \cap B) = P(A|B)P(B), \quad P(B) \neq 0.$$

- Law of total probability

  Let $B_1, B_2, \ldots, B_n$ be such that $\bigcup_{i=1}^n B_i = \Omega$ and $B_i \cap B_j = \phi$ for $i \neq j$, with $P(B_i) > 0$ for all i. Then, for any A,

  $$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i).$$

- Bayes rule

  Let A and $B_1, B_2, \ldots, B_n$ be events where the $\bigcup_{i=1}^n B_i = \Omega$ and $B_i \cap B_j = \phi$ for $i \neq j$, with $P(B_i) > 0$ for all i. Then,

  $$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(A|B_i)P(B_i)}.$$

1.6 Independence

A and B are independent events $\Longleftrightarrow P(A \cap B) = P(A)P(B)$.
mutually independent: $P(A_1 \cap \cdots \cap A_m) = P(A_1) \cdots P(A_m)$.

2 Random Variables (Chapter 2)

random variable (r.v.): a function from Ω to $\mathcal{R}$.

2.1 Discrete Random Variables

- Frequency function (probability mass function)

  $$p(x_i) = P(X = x_i), \quad \sum p(x_i) = 1.$$

- Cumulative distribution function (cdf)

  $$F(x) = P(X \le x), \quad -\infty < x < \infty.$$

  : non-decreasing and

  $$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1.$$

- Bernoulli: Success or Failure

  X: 0 or 1

  See the table.

- Binomial: n independent Bernoulli experiments

  X: total number of successes $\sim B(n, p)$.

  See the table.

- Bernoulli and Binomial

  $X_1, X_2, \ldots, X_n$: independent Bernoulli r.v. with $P(X_i = 1) = p$.

  $\Longrightarrow \quad Y = X_1 + X_2 + \cdots + X_n \sim B(n, p)$.

- Geometric

  X: under Bernoulli trials, # of trials up to and including the first success.

  See the table.

- Negative Binomial

  X: under Bernoulli trials, # of trials until r successes in all $\sim NB(r, p)$.

  See the table.

- Geometric and Negative Binomial

  $X_1, X_2, \ldots, X_r$: independent Geometric r.v. with p.

  $\Longrightarrow \quad Y = X_1 + X_2 + \cdots + X_r \sim NB(r, p)$.

- Hypergeometric

  Suppose an urn contains n balls, of which r are black and $n - r$ are white.

  X: # of black balls drawn when taking m balls without replacement $\sim H(m; r, n)$.

  See the table.

- Poisson

  X: # of times some event occurs in a given interval of time $\sim Poisson(\lambda)$.

  See the table.

- Binomial and Poisson

  - Setting: dividing an interval into a very large number of small subintervals of equal length.
  - Assumptions (of Poisson):
    (i) independent increment: what happens in one subinterval is independent of what happens in any other subinterval.
    (ii) stationarity: the probability of an event is the same in each subinterval.
    (iii) proportionality: the probability of an event is proportional to the length of the subinterval.
    (iv) rareness: events do not happen simultaneously.

  Binomial with large n and small $p \longrightarrow$ Poisson(λ) where $np \to \lambda$.

2.2 Continuous Random Variables

- Density function $f(x)$
  (i) $f(x) \geq 0$
  (ii) $\int_{-\infty}^{\infty} f(x)dx = 1$
  (iii) $P(a < X < b) = \int_a^b f(x)dx, \quad a < b.$

- Cdf and density function

$$f(x) = \frac{dF(x)}{dx}$$

- Uniform $[0,1] \sim U[0,1]$

  See the table.

- Exponential $\sim Exp(\lambda)$

  X: waiting time

  See the table.

  memoryless property – $P(T > t + s | T > s) = P(T > t)$

  : the probability that the unit will last t more time units does not depend on s.

- Gamma $\sim Gamma(\alpha, \lambda)$

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du,$$

  $\Gamma(\alpha) = (\alpha - 1)!$ if α is a natural number.

  See the table.

- Exponential and Gamma

  $X_1, X_2, \ldots, X_n$: independent Exponential r.v. with λ.

  $\Longrightarrow \quad Y = X_1 + X_2 + \cdots + X_n \sim Gamma(n, \lambda)$.

- Poisson, Exponential and Gamma
    - distribution of # of arrivals during $[0, t] \sim Poisson(\lambda t)$
    - distribution of waiting time until the first arrival $\sim Exp(\lambda)$
    - distribution of waiting time until the rth arrival $\sim Gamma(r, \lambda)$
- Normal $\sim N(\mu, \sigma^2)$

  See the table.

  Standard normal $\sim N(0, 1)$.
- Central Limit Theorem (CLT): if a r.v. is the sum of a large number of independent r.v.'s, it is approximately normally distributed.
- If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, then $Y \sim N(a\mu + b, a^2\sigma^2)$.

2.3 Functions of a Random Variable

- X: continuous r.v. with density f_X.

  $Y = g(X)$, g: differentiable, strictly monotonic on some interval I.

  $f(x) = 0$ if x is not in I.

  $\Longrightarrow Y$ has the density function

  $$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|,$$

  for y such that $y = g(x)$ for some x, and $f_Y(y) = 0$ if $y \neq g(x)$ for any x in I. Here g^{-1} is the inverse function of g.
- F: cdf of X and $Z = F(X)$, then Z has a uniform distribution on $[0, 1]$.
- $U \sim U[0, 1]$ and $X = F^{-1}(U)$. Then, the cdf of X is F.

3 Joint Distribution (Chapter 3)

If $X_1, X_2, \ldots, X_n$ are jointly distributed r.v.'s, their joint cdf is

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n)$$

3.2 Discrete Random Variables

- Joint frequency function (joint probability mass function)

  $$p(x_i, y_j) = P(X = x_i, Y = y_j)$$

– Marginal frequency function

$$\begin{aligned} p_X(x) &= \sum_j p(x, y_j) \\ p_Y(y) &= \sum_i p(x_i, y) \end{aligned}$$

– $X_1, \ldots, X_m$: discrete r.v.'s

$$p(x_1, \cdots, x_m) = P(X_1 = x_1, \cdots, X_m = x_m)$$

3.3 Continuous Random Variables

– Joint cdf: $F(x, y)$

– Joint density function: $f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$

– Marginal density function

$$\begin{aligned} f_X(x) &= \int f(x, y) dy \\ f_Y(y) &= \int f(x, y) dx \end{aligned}$$

– $P((X, Y) \in A) = \int\int_A f(x, y) dy dx$

– Bivariate normal

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\Big(- \frac{1}{2(1-\rho^2)} \Big[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \\ & \quad - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \Big] \Big) \\ & \quad -\infty < \mu_X, \mu_Y < \infty, \sigma_X, \sigma_Y > 0, -1 < \rho < 1 \end{aligned}$$

marginal distribution: $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$

3.4 Independent Random Variables

– $X_1, X_2, \ldots, X_n$: independent

$$\begin{aligned} \iff & F(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n) \\ \iff & f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n) \\ \iff & P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) \\ & = P(X_1 \in A_1) P(X_2 \in A_2) \cdots P(X_n \in A_n) \end{aligned}$$

for all $x_1, x_2, \ldots, x_n$ and $A_1, A_2, \ldots, A_n$.

3.5 Conditional Distribution

- Discrete case

$$\begin{aligned}
P(X = x_i|Y = y_j) &= \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} \\
&= \frac{p_{XY}(x_i, y_j)}{p_Y(y_j)} \quad \text{if } p_Y(y_j) > 0 \\
p_{XY}(x, y) &= p_{X|Y}(x|y)p_Y(y) \\
p_X(x) &= \sum_y p_{X|Y}(x|y)p_Y(y)
\end{aligned}$$

- Continuous case

$$\begin{aligned}
f_{Y|X}(y|x) &= \frac{f_{XY}(x, y)}{f_X(x)} \quad \text{if } f_X(x) > 0 \\
f_{XY}(x, y) &= f_{Y|X}(y|x)f_X(x) \\
f_Y(y) &= \int f_{Y|X}(y|x)f_X(x)dx
\end{aligned}$$

  Bivariate normal:

$$Y|X = x \sim N(\mu_Y + \rho(x - \mu_X)\sigma_Y/\sigma_X, \sigma_Y^2(1 - \rho^2))$$

3.6 Functions of Jointly Distributed Random Variables

- Sums: $Z = X + Y$ (convolution)

$$\begin{aligned}
p_Z(z) &= \sum_x p(x, z - x) \\
&= \sum_x p_X(x)p_Y(z - x) \quad \text{if } X\&Y: \text{ independent} \\
f_Z(z) &= \int f(x, z - x)dx \\
&= \int f_X(x)f_Y(z - x)dx \quad \text{if } X\&Y: \text{ independent}
\end{aligned}$$

- General case: change of variables (see p.95)

3.7 Functions of Jointly Distributed Random Variables

$X_1, X_2, \ldots, X_n \sim$ cdf F and density f (i.i.d.)

- $U = \max_i X_i$, $V = \min_i X_i$

$$\begin{aligned}
F_U(u) &= [F(u)]^n, & f_U(u) &= nf(u)[F(u)]^{n-1} \\
F_V(v) &= 1 - [1 - F(v)]^n, & f_V(v) &= nf(v)[1 - F(v)]^{n-1}
\end{aligned}$$

- Order statistic: $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$
  The density of $X_{(k)}$, the kth order statistic:

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} f(x)F^{k-1}(x)\left[1 - F(x)\right]^{n-k}$$

4 Expected Values (Chapter 4)

4.1 Expected Value of a Random Variable : measure of center

- Discrete r.v. with frequency function $p(x)$

  $$E(X) = \sum_i x_i p(x_i)$$

  provided that $\sum_i |x_i| p(x_i) < \infty$. If the sum diverges, the expectation is undefined.
- Continuous r.v. with density function $f(x)$

  $$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

  provided that $\int |x| f(x) dx < \infty$. If the integral diverges, the expectation is undefined.

4.1.1 Expectations of Functions of Random Variables

- Theorem A

  $Y = g(X)$.

  (a) X: discrete with frequency function $p(x)$

  $$\Longrightarrow E(Y) = \sum_x g(x) p(x)$$

  provided that $\sum |g(x)| p(x) < \infty$.

  (b) X: continuous with density function $f(x)$

  $$\Longrightarrow E(Y) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

  provided that $\int |g(x)| f(x) dx < \infty$.
- Theorem B

  $Y = g(X_1, \ldots, X_n)$.

  (a) X_i: discrete with joint frequency function $p(x_1, \ldots, x_n)$

  $$\Longrightarrow E(Y) = \sum_{x_1, \ldots, x_n} g(x_1, \ldots, x_n) p(x_1, \ldots, x_n)$$

  provided that $\sum_{x_1, \ldots, x_n} |g(x_1, \ldots, x_n)| p(x_1, \ldots, x_n) < \infty$.

  (b) X: continuous with joint density function $f(x_1, \ldots, x_n)$

  $$\Longrightarrow E(Y) = \int \cdots \int g(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \cdots dx_n$$

  provided that the integral with $|g|$ in place of g converges.
- Corollary A

  X, Y: independent r.v.'s

  g, h: fixed functions

  $$\Longrightarrow E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

4.2 Variance and Standard Deviation : measure of spreadness

- X: r.v. with $E(X)$

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

  provided that the expectation exist. The standard deviation of X is the square root of the variance.

  $$\begin{aligned} Var(X) &= \textstyle\sum_i (x_i - E(X))^2 p(x_i) && \text{(discrete)} \\ &= \textstyle\int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx && \text{(continuous)} \end{aligned}$$

- Theorem A
  $Var(X)$ exists and $Y = a + bX$.

  $$\begin{aligned} Var(Y) &= b^2 Var(X) \\ \sigma_Y &= |b|\sigma_X \end{aligned}$$

- Theorem B
  $Var(X)$ exists.

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

- Theorem C (Chebyshev's Inequality)
  X: r.v. with mean μ and variance σ^2. Then, for any $t > 0$,

  $$P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$$

- Corollary A
  If $Var(X) = 0$, then $P(X = \mu) = 1$.

- MSE (Mean Squared Error)

  $$\begin{aligned} MSE &= E[(X - x_0)^2] \\ &= Var(X) + [E(X - x_0)]^2 \\ &= \sigma^2 + \beta^2 \end{aligned}$$

4.3 Covariance and Correlation

- $X\&Y$: jointly distributed r.v.'s with μ_X and μ_Y respectively.

  $$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

  provided that the expectation exists.

- Theorem A
  $U = a + \sum_{i=1}^{n} b_i X_i$ and $V = c + \sum_{j=1}^{m} d_j Y_j$

  $$Cov(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_i d_j Cov(X_i, Y_j)$$

- Corollary A

$$Var(a + \sum_{i=1}^{n} b_i X_i) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_i b_j Cov(X_i, X_j)$$

- Corollary B

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i),$$

  if the X_i are independent.

- Correlation coefficient

  $X \& Y$: jointly distributed r.v.'s, variances and covariances exist, and the variances are nonzero.

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

- Theorem B

  $-1 \le \rho \le 1$. Furthermore, $\rho = \pm 1 \iff P(Y = a + bX) = 1$ for some a and b.

4.4 Conditional Expectation and Prediction

- Definitions and Examples
  - Conditional expectation of Y given $X = x$

$$\begin{aligned} E(Y|X=x) &= \textstyle\sum_y y p_{Y|X}(y|x) \quad \text{(discrete)} \\ &= \textstyle\int y f_{Y|X}(y|x) dy \quad \text{(continuous)} \end{aligned}$$

    More generally,

$$E[h(Y)|X=x] = \textstyle\int h(y) f_{Y|X}(y|x) dy$$

  - Law of total expectation: $E(Y) = E[E(Y|X)]$
  - $Var(Y) = Var[E(Y|X)] + E[Var(Y|X)]$
- Prediction

  Predicting Y by some function $h(X)$ in order to minimize MSE.

$$\begin{aligned} E[(Y-h(X))^2] &= E[E[(Y-h(X))^2|X]] \\ h(X) &= E[Y|X] \end{aligned}$$

4.5 Moment Generating Function (mgf)

$$\begin{aligned} M(t) &= E(e^{tx}) \\ &= \sum_x e^{tx} p(x) \quad \text{(discrete)} \\ &= \int_{-\infty}^{\infty} e^{tx} f(x) dx \quad \text{(continuous)} \end{aligned}$$

- PROPERTY A
  If the mgf exists for t in an open interval containing zero, it uniquely determines the probability distribution.
- PROPERTY B
  If the mgf exists for t in an open interval containing zero,
  $$M^{(r)}(0) = E(X^r)$$
- PROPERTY C
  If X has the mgf $M_X(t)$ and $Y = a + bX$, then Y has the mgf
  $$M_Y(t) = e^{at} M_X(bt)$$
- PROPERTY D
  $X \& Y$: independent with M_X and M_Y, and $Z = X + Y$.
  $$\Longrightarrow M_Z(t) = M_X(t) M_Y(t)$$
  on the common interval where both mgf's exist.
- Property D–1
  $X_1, X_2, \ldots, X_n$: independent
  $$\Longleftrightarrow \quad M_{X_1, X_2, \ldots, X_n}(t_1, t_2, \ldots, t_n) = M_{X_1}(t_1) M_{X_2}(t_2) \cdots M_{X_n}(t_n).$$

4.6 Approximate Methods

- Propagation error (δ method)

$$\begin{aligned} Y = g(X) &\approx g(\mu_X) + (X - \mu_X) g'(\mu_X) \text{ : first order} \\ \mu_Y &\approx g(\mu_X) \\ \sigma_Y^2 &\approx \sigma_X^2 [g'(\mu_X)]^2 \end{aligned}$$

$$\begin{aligned} Y = g(X) &\approx g(\mu_X) + (X - \mu_X) g'(\mu_X) + \frac{1}{2}(X - \mu_X)^2 g''(\mu_X) \text{ : second order} \\ \mu_Y &\approx g(\mu_X) + \frac{1}{2} \sigma_X^2 g''(\mu_X) \end{aligned}$$

5 Limit Theorems (Chapter 5)

- Law of Large Numbers
  $X_1, X_2, \ldots, X_i, \ldots$: a sequence of indep. r.v.'s, $E(X_i) = \mu$ & $Var(X_i) = \sigma^2$.
  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$
  Then, for any $\epsilon > 0$,
  $$P(|\bar{X}_n - \mu| > \epsilon) \longrightarrow 0 \quad \text{as } n \to \infty$$
  : convergence in probability

- Converge almost surely to α
  : if for every $\epsilon > 0$, $|Z_n - \alpha| > \epsilon$ only a finite number of times.
- Convergence in distribution
  $X_1, X_2, \ldots$: a sequence of r.v.'s with cdf $F_1, F_2, \ldots$.
  X: a r.v.'s with cdf F.
  X_n converges in distribution to X if
  $$\lim_{n\to\infty} F_n(x) = F(x)$$
  at every point at which is continuous.
- CONTINUITY THEOREM
  F_n: a sequence of cdfs with corresponding mgf M_n.
  F: a cdf with the mgf M.
  $M_n(t) \to M(t)$ for all t in an open interval containing zero.
  $$\Longrightarrow F_n(x) \longrightarrow F(x)$$
  at all continuity points of F.
- CENTRAL LIMIT THEOREM (CLT)
  $X_1, X_2, \ldots,$: a sequence of independent r.v.'s with mean 0 and variance σ^2, the cdf F and the mgf M defined in a neighborhood of zero.
  $S_n = \sum_{i=1}^n X_i$
  $$\Longrightarrow \lim_{n\to\infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \le x\right) = \Phi(x), \quad -\infty < x < \infty$$
  where $\Phi(\cdot)$ is the standard normal cdf.