Formulae Sheet – Statistics I

2nd Semester of 2019/2020

- Probability:

Let A, B and C be events of a sample space. Then:

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$

Let $\{A_1, A_2, \cdots\}$ be a partition of a sample space S and B an event of S. Then, if $P(A_j) > 0$, for all $j = 1, 2, \cdots$ and $P(B) > 0$,

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1} P(B|A_i)P(A_i)}.$$

- Expected values and other quantities (for a single RV):

Let X be a random variable. Then,

$$\begin{aligned}
E(g(X)) &= \int_{-\infty}^{+\infty} g(x) f_X(x) dx, \text{ if } X \text{ is a continuous RV} \\
E(g(X)) &= \sum_{x \in D_X} g(x) f_X(x), \text{ if } X \text{ is a discrete RV} \\
Var(X) &= E\left((X - \mu_X)^2\right) = E(X^2) - (E(X))^2 \\
mo(X) &= \arg\max_{x \in \mathbb{R}} f_X(x), \quad q_\alpha(X) = \min\{x \in \mathbb{R} : F_X(x) \geq \alpha\}, \quad M_X(t) = E(e^{Xt}).
\end{aligned}$$

- Expected values and other quantities (for functions of RV):

Let X and Y be two random variable. Then,

$$\begin{aligned}
E(g(X,Y)) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) f_{X,Y}(x,y) dx dy, \text{ if } X \text{ and } Y \text{ are continuous RV} \\
E(g(X,Y)) &= \sum_{(x,y) \in D_{X,Y}} g(x,y) f_{X,Y}(x,y), \text{ if } X \text{ and } Y \text{ are discrete RV} \\
Cov(X,Y) &= E\left((X - \mu_X)(Y - \mu_Y)\right) = E(XY) - E(X)E(Y) \\
E(g(X,Y)|Y=y) &= \int_{-\infty}^{+\infty} g(x,y) f_{X|Y=y}(x) dx, \text{ if } X \text{ and } Y \text{ are continuous RV} \\
E(g(X,Y)|Y=y) &= \sum_{x \in D_X} g(x,y) f_{X|Y=y}(x), \text{ if } X \text{ and } Y \text{ are discrete RV} \\
Var(X|Y=y) &= E\left((X - \mu_{X|Y=y})^2 | Y=y\right) = E(X^2|Y=y) - (E(X|Y=y))^2.
\end{aligned}$$

- The tower property (or Law of total expectation):

Let X and Y be two random variable. Then,

$$E(E(X|Y)) = E(X) \quad \text{and} \quad E(E(Y|X)) = E(Y).$$

- Theoretical distributions:
  - Discrete uniform distribution:

$$f_X(x_j) = \frac{1}{k},\ j = 1, 2, 3, ..., k$$
$$\text{if } x_j = j,\ \text{then } \mu_X = E(X) = \frac{k+1}{2}, \quad Var(X) = \frac{k^2-1}{12}$$

  - Binomial (and Bernoulli) distribution ($X \sim Bin(n,p)$):

$$f_X(x) = \binom{n}{x} \times p^x(1-p)^{n-x},\ x = 0, 1, 2, \cdots, n$$
$$E(X) = np, \quad Var(X) = np(1-p) \quad M_X(t) = \left[(1-p) + pe^t\right]^n$$

  - Hypergeometric distribution ($X \sim Hypergeometric(N, M, n)$):

$$P(X = k) = \frac{\binom{N-M}{n-k}\binom{M}{k}}{\binom{N}{n}}, \quad k = \max\{0, n-(N-M)\}, \cdots, \min n, M$$
$$E(X) = n \times \frac{M}{N}, \quad Var(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$$

  - Negative Binomial (and Geometric) distribution ($X \sim NB(k,p)$):

$$P(X = x) = \binom{x-1}{k-1}p^k(1-p)^{x-k}, \quad x = k, k+1, k+2, \cdots$$
$$E(X) = \frac{k}{p}, \quad Var(X) = \frac{k}{p}\left(\frac{1}{p} - 1\right), \quad M_X(t) = \left(\frac{pe^t}{1-e^t(1-p)}\right)^k$$

  - Poisson distribution ($X \sim Poi(\lambda)$):

$$P\left(X = x\right) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, 2, \cdots, \quad E(X) = Var(X) = \lambda, \quad M_X(X) = e^{\lambda\left(e^t - 1\right)}$$

  - Exponential distribution ($X \sim Exp(\lambda)$):

$$f_X(x) = \lambda e^{-\lambda x}, \quad x > 0, \quad E(X) = 1/\lambda, \quad Var(X) = 1/\lambda^2, \quad M_X(t) = \left(1 - t/\lambda\right)^{-1}, t < \lambda.$$

  - Gamma (and chi-squared) distribution ($X \sim Gamma(a,b)$):

$$E(X) = ab, \quad Var(X) = ab^2, \quad M_X(t) = \left(1 - bt\right)^{-a}, t < 1/b$$
$$\text{If } a = 1,\ b = \frac{1}{\lambda} \text{ then } X \sim Exp(\lambda); \quad \text{If } a = \frac{n}{2},\ b = 2 \text{ then } X \sim \chi^2_{(n)}.$$

  - Continuous uniform distribution ($X \sim U(a,b)$):

$$f_X(x) = \frac{1}{b-a}, \quad a < x < b, \quad E(X) = \left(a+b\right)/2, \quad Var(X) = (b-a)^2/12$$

  - Normal distribution ($X \sim N(\mu, \sigma^2)$):

$$f_X\left(x\right) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}, \quad E(X) = \mu, \quad Var(X) = \sigma^2, \quad M_X(t) = e^{\left(\mu t + 0.5\sigma^2 t^2\right)}$$

- Central Limit Theorem

Let X_i be a sequence of independent and identically distributed random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Then,

$$Z = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \overset{a}{\sim} N\left(0, 1\right)$$