Math 324 Final Review

Part I

- Joint distributions

- Discrete

$$p(x,y) = P(X=x \text{ and } Y=y)$$

- Continuous

$$f(x,y)$$

- probabilities

Discrete: $$P[(X,Y)\in A] = \sum\sum_{(x,y)\in A} p(x,y)$$

Continuous: $$P[(X,Y)\in A) = \iint_A f(x,y)\,dy\,dx$$

- Marginal probabilities

$$p_X(x) = \sum_y p(x,y) \qquad p_Y(y) = \sum_x p(x,y)$$

$$f_X(x) = \int f(x,y)\,dy \qquad f_Y(x) = \int f(x,y)\,dx$$

- independence

$$X, Y \text{ independent} \iff \begin{matrix} p(x,y) = p(x)\,p(y) \\ f(x,y) = f(x)\,f(y) \end{matrix}$$

- Conditional Distributions

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

- Expected value

$$E[h(X,Y)] = \begin{cases} \sum_x \sum_y h(x,y)\,p(x,y) \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{\infty} h(x,y)\,f(x,y)\,dy\,dx \end{cases}$$

- Covariance

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

$$= \begin{cases} \sum_x \sum_y (x-\mu_x)(y-\mu_y)\,p(x,y) \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y)\,f(x,y)\,dy\,dx \end{cases}$$

$$Cov(X,Y) = E[XY] - E[X]E[Y]$$

- Correlation

$$\rho_{x,y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

X, Y independent $\Rightarrow \rho = 0$ (but reverse not true)

$-1 \leq \rho \leq 1$

- <u>Sampling Distributions</u>
  - statistic
    - quantity estimated from data
    - sampling variability
  - Random samples
    - i.i.d
  - Finding sampling distribution
    - Theory
    - Simulation

- distribution of $\overline{X}$ and T_n

$X_1, \dots, X_n$ iid $E[X_i] = \mu$
$Var(X_i) = \sigma^2$

- $E[\overline{X}] = \mu$
- $var[\overline{X}) = \dfrac{\sigma^2}{n}$
- $E[T_n] = n\mu$
- $Var(T_n) = n\sigma^2$
- If $X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$

  $\overline{X}$ exactly $N(\mu, \frac{\sigma^2}{n})$

  T_n exactly $N(n\mu, n\sigma^2)$
- CLT if $n \to \infty$ then

  $\overline{X}$ is $N(\mu, \sigma^2)$ approximately

  T_n is $N(n\mu, n\sigma^2)$ approximately

(rule of thumb $n > 30$)

- linear combinations

$$a_1X_1 + \cdots + a_nX_n = \sum a_iX_i$$

$$E\left[\sum a_iX_i\right] = \sum a_i E[X_i]$$

$$Var\left(\sum a_iX_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Cov(X_i, X_j)$$

Note this implies that if X, Y independent

$$Var(X_1 - X_2) = Var(X_1) + Var(X_2)$$

- if $X_1, \ldots X_n$ are all Normal then

$\sum a_iX_i$ is also normal

- Point estimation
  - ~~Parameter~~ ← population characteristic
  - Point estimate - single number $\hat{\theta}$
  - Point estimator - method or formula

eg μ $\bar{X}$

σ s

p $\hat{p}$

- Bias

$$\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$$

Unbiased $\Rightarrow E[\hat{\theta}] = \theta$

$$\left.\begin{aligned} E[\bar{X}] &= \mu \\ E[s^2] &= \sigma^2 \\ E[\hat{p}] &= p \end{aligned}\right\} \text{All unbiased}$$

- unbiased estimators are better
- Minimum Variance Unbiased Estimation
  - choose estimator with smallest variance
  - If $X_1, \ldots, X_n$ from $N(\mu, \sigma^2)$ then $\bar{X}$ is MVUE for μ.
- Standard error - (estimate of the standard deviation of an estimator

eg $SE[\bar{X}] = \frac{s}{\sqrt{n}}$ $\quad SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

– MME (method of Moments estimation)

equate

$$E[X^k] = \frac{\sum x^k}{n} \quad \text{for } k = 1, 2 \cdots$$

solve system of equations

– MLE (maximum likelihood estimation)

$$L(\underset{\sim}{\theta}) = f(x_1, x_2, \ldots x_n; \underset{\sim}{\theta}) \quad \underset{\sim}{\theta} = (\theta_1, \ldots \theta_p)$$

likelihood

log likelihood

$$l(\theta) = \log L(\underset{\sim}{\theta})$$

usually differentiate wrt to each of $\theta_1, \ldots, \theta_p$

Solve system of equations ie

$$\frac{\partial l}{\partial \theta_1} = 0$$

$$\vdots$$

$$\frac{\partial l}{\partial \theta_p} = 0$$

- If $\hat{\theta}_1, \ldots, \hat{\theta}_p$ is MLE for $\theta_1, \ldots \theta_p$ 8

then $h(\hat{\theta}_1, \ldots \hat{\theta}_p)$ is MLE of $h(\theta_1, \ldots \theta_p)$

- when n becomes very large

  $\hat{\underset{\sim}{\theta}}$ becomes unbiased

  $\hat{\underset{\sim}{\theta}}$ has minimum variance