STAT 330 Final Exam Summary
Mathematical Statistics

1 Moment Generating Functions

- If $Y = aX + b$ then

$$M_Y(t) = E\left[e^{Yt}\right] = E\left[e^{(aX+b)t}\right] = e^{bt}M_X(at)$$

- $M_X^{(n)}(0) = E[X^n]$
- If $Y = \sum_{i=1}^n X_i$, then $M_Y(t) = \prod_{i=1}^n M_{X_i}(t)$

2 Joint and Conditional Distributions

- Some basic properties are:
  - $f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)$
  - $F_{XY}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(s,t)\ ds\ dt$
  - $f_Y(y) = \int_{-\infty}^\infty f_{XY}(s,t)\quad ds, \quad f_X(x) = \int_{-\infty}^\infty f_{XY}(s,t)\ dt$
- If $X \perp Y$ then

$$\iint_A f_{XY}(x,y) d(x,y) = \iint_A f_X(x) f_Y(y) d(x,y)$$

- We define the **support** of a r.v. as $\{x : f_X(x) > 0\}$. If the support of a joint r.v. is **non-rectangular**, then the atomic components are not independent.
- If $X \perp Y$ then $g(X) \perp h(Y)$ for any functions g and h.
- The **double expectation formula** states

$$\begin{aligned} E[X] &= E\left(E[X|Y]\right) \\ Var[X] &= E[Var(X|Y)] + Var(E[X|Y) \end{aligned}$$

- If $\{X_k\}$ are a set of independent random variables, then

$$M_{\sum X_k}(t_1, ..., t_n) = \prod M_{X_k}(t_k)$$

3 Functions of Random Variables

- Given $X, Y = g(X)$, the **cdf method** involves finding the cdf of Y as

$$F_Y(y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y))$$

- Given X, Y, and $U = g(X), V = h(Y)$, the **Jacobian method** uses the fact that

$$\begin{aligned} f_{UV}(u,v) &= f_{XY}(x,y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right| \\ &= f_{XY}(g^{-1}(x), h^{-1}(y))\left|\frac{\partial(u,v)}{\partial(x,y)}\right|^{-1} \end{aligned}$$

- The **MGF method** is for the case where if $X_1, X_2, ..., X_n$ are independent and X_i has MGF $M_{X_i}(t)$ then if $Y = \sum_{i=1}^n X_i$ we have $M_Y(t) = \prod_{i=1}^n M_{X_i}(t)$ and if the $X_i's$ are i.i.d. then $M_Y(t) = M_{X_1}^n(t)$

4 Convergence of Random Variables

- The sequence $X_1, X_2, ..., X_n$ **converges in probability** to X if for any $\epsilon > 0$ we have

$$\lim_{n\to\infty} P(|X_n - X| \ge \epsilon) = 0 \iff \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$$

  We denote this by $X_n \xrightarrow{p} X$.
- We say that $\{X_n : \Omega \mapsto A_n\}_{n\in\mathbb{N}}$ **converges in distribution** to $X : \Omega \mapsto B$ if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

  at all parts where $F_X(x)$ is continuous. We then write $X_n \xrightarrow{d} X$.
- If $X_n \xrightarrow{p} X$ then $X_n \xrightarrow{d} X$.
- If $X_n \xrightarrow{d} b$ then $X_n \xrightarrow{p} b$.
- (**Markov**) For any $k \in \mathbb{N}$, $P(|X| > C) \le \frac{E[|X|^k]}{C^k}$
  - In the specific case of $k = 2$,

$$P(|X| > C) \le \frac{E[|X|^2]}{C^2} = \frac{Var(X) + (E[X])^2}{C^2}$$

- A property of the **mean of random variables** is $\bar{X} \xrightarrow{p} \mu$.
- (**Central Limit Theorem**) $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{d} N(0,1)$ where $\{X_n\}$ are i.i.d. r.v.s. with $X_n \sim (\mu, \sigma^2)$
- If $X_n \overset{p\ (D)}{\to} a$ then $g(X_n) \overset{p\ (D)}{\to} g(a)$. That is g is continuous at "a".
- (**Slutsky's Theorem**) Suppose that $X_n \xrightarrow{d} X$ and $Y \xrightarrow{p} b$. Then,
  - $X_n + Y_n \xrightarrow{d} X + b$
  - $X_n \cdot Y_n \xrightarrow{d} b \cdot X$
  - $X_n / Y_n \xrightarrow{d} X/b$, $b \ne 0$

- (**Delta Method**) Suppose that for $X_1, X_2, ..., X_n$ we have

$$\sqrt{n}(X_n - \theta) \xrightarrow{d} N(0, \sigma)$$

If $g(x)$ is differentiable at θ and $g'(\theta) \neq 0$ then

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{d} N(0, g'(\theta)^2\sigma^2)$$

5 Point Estimation

- In the **method of moments**, we want to set the sample/observed k^{th} moment equal to the theoretical moment:

$$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \longleftrightarrow E[X^l]$$

- Let's talk about the **MLE estimate**. Suppose that $X_1, ..., X_n$ are i.i.d. from $f(x, \theta)$. We call $L(\theta, X) = \prod_{i=1}^{n} f(x_i, \theta)$ the *likelihood* of θ and $l = \ln(L)$ the *log-likelihood* function. The MLE estimate is

$$\hat{\theta}_{ML} = \hat{\theta}_{MLE} = \text{argmax } L(\theta) = \text{argmax } l(\theta)$$

- The **score function** is $S(\theta) = \frac{\partial}{\partial\theta}\ln f(x, \theta)$, the **information function** is $I(\theta) = \frac{\partial}{\partial\theta}S(\theta) = \frac{\partial^2}{\partial\theta^2}\ln f(x, \theta)$, the **Fisher information matrix** is $J(\theta) = -E\left[I(\theta)\right]$; here are some properties:
  - $S(\hat{\theta}_{ML}) = 0$
  - $E\left[\frac{\partial}{\partial\theta}\ln f(x, \theta)\right] = 0$
  - $E\left[\frac{\partial^2}{\partial\theta^2}\ln f(x, \theta)\right] = E\left[\left(\frac{\partial}{\partial\theta}\ln f(x, \theta)\right)^2\right]$
  - If $X_1, ..., X_n$ are i.i.d. then $E\left[\left(\frac{\partial}{\partial\theta}\ln f(x, \theta)\right)^2\right] = nE\left[\left(\frac{\partial}{\partial\theta}\ln f(x_1, \theta)\right)^2\right]$

- (**Cramer-Rao Lower Bound**) Suppose that $T(X_1, ..., X_n)$ is an estimator for θ. Remark that if T is unbiased if $E[T(X)] = \theta$. If $E[T(X)] \neq \theta$ then $E[T(X)]$ is biased. Also, if $X_1, ..., X_n$ are samples from $f(x, \theta)$ then

$$Var(T) \geq \frac{\left(\frac{\partial}{\partial\theta}E[T]\right)^2}{E\left[\left(\frac{\partial}{\partial\theta}\ln f(x, \theta\right)^2\right]} \geq \frac{1}{E\left[\left(\frac{\partial}{\partial\theta}\ln f(x, \theta\right)^2\right]}$$

- For the maximum likelihood estimator, we have:
  - $\hat{\theta}_{ML} \xrightarrow{p} \theta$ (asymptotically)
  - $\sqrt{n}(\hat{\theta}_{ML} - \theta) \xrightarrow{d} N\left(0, \frac{1}{J(\theta)}\right)$ (asymptotically normal)
  - This will also imply that $\hat{\theta}_{ML} - \theta \xrightarrow{d} N\left(0, \frac{1}{nJ_1(\theta)}\right) = N\left(0, \frac{1}{J(\theta)}\right)$ and $\hat{\theta}_{ML} \rightarrow N\left(\theta, \frac{1}{J(\theta)}\right)$.