Inequalities and Limits Part A

Outline

- Inequalities: Markov, Chebyshev, Jensen, Hölder
- Limits
  - of Probability: Weak Law of Large Numbers
  - of Events: Strong Law of Large Numbers
  - of Cdfs: Central Limit Theorem
Inequalities: Preliminaries

- We are interested in moment-based inequalities that are informative about rvs.
- \( X(\omega) \leq Y(\omega) \) for \( \omega \in \Omega \) \( \Rightarrow \) \( E(X) \leq E(Y) \).
- For rvs \( X, Y \), \( P(X \leq Y) = 1 \) \( \Rightarrow \) \( E(X) \leq E(Y) \).
  - For rv \( Z \), \( P(Z \leq 0) = 1 \) \( \Rightarrow \) \( E(Z) \leq 0 \). Let \( Z = X - Y \) to obtain the above inequality.
  - For rv \( X \) and number \( y \), \( P(X \leq y) = 1 \) \( \Rightarrow \) \( E(X) \leq y \).

- Ex: A rv \( X \in [a, b] \) has the variance bound \( V(X) \leq \frac{(b-a)^2}{4} \).
  1. Consider the quadratic function \( g(\beta) : \mathbb{R} \to \mathbb{R} \) given as
     \[
     g(\beta) = E(X - \beta)^2 = E(X^2) - 2E(X)\beta + \beta^2 \quad \text{for all } \beta \in (-\infty, \infty).
     \]
  2. The function \( g(\beta) \) is minimized at \( \beta_0 = E(X) \).
  3. So \( V(X) = g(E(X)) \leq E(X - \beta)^2 \) for all \( \beta \in (-\infty, \infty) \).
  4. In particular, for \( \beta = \frac{a+b}{2} \), \( V(X) \leq g\left(\frac{a+b}{2}\right) = E\left(X - \frac{a+b}{2}\right)^2 \).
  5. Then \( V(X) \leq E\left(X - \frac{a+b}{2}\right)^2 = E\left(X^2 - aX - bX + \left(\frac{a+b}{2}\right)^2\right) = E\left((X - a)(X - b) + \left(\frac{a-b}{2}\right)^2\right) \).
  6. Since \( X \in [a, b] \), we have \( E((X - a)(X - b)) \leq 0 \).
  7. Hence, \( V(X) \leq \left(\frac{a-b}{2}\right)^2 \).

- The variance bound is tight because it is achieved by the rv \( X \) with \( P(X = a) = P(X = b) = \frac{1}{2} \).
- For this rv, \( E(X) = \frac{a}{2} + \frac{b}{2}, E(X^2) = \frac{a^2}{2} + \frac{b^2}{2} \) and \( V(X) = \frac{a^2}{2} + \frac{b^2}{2} - \left(\frac{a}{2} + \frac{b}{2}\right)^2 = \left(\frac{a-b}{2}\right)^2 \).
Markov’s and Chebyshev’s Inequalities

- For a nonnegative rv $Y$ and positive constant $\alpha$ and positive integer $k$,
  \[ E(Y^k) = \int_{y \geq 0} y^k f_Y(y) \, dy \geq \int_{y \geq \alpha} y^k f_Y(y) \, dy \geq \int_{y \geq \alpha} \alpha^k f_Y(y) \, dy = \alpha^k P(Y \geq \alpha) \]

- Inserting $Y = |X|$ into the above expression
  - Markov’s Inequality: $P(|X| \geq \alpha) \leq \frac{1}{\alpha^k} E(|X|^k)$ for $\alpha \geq 0$, $k \in \{1, 2, \ldots\}$.

- This can also be written as $P(|X| \geq \alpha) \leq \min_{k \in \{1, 2, \ldots\}} \left\{ \frac{1}{\alpha^k} E(|X|^k) \right\}$.

- In the special case of $k = 2$: $P(|X| \geq \alpha) \leq \frac{1}{\alpha^2} E(|X|^2)$ for $\alpha \geq 0$.

- Inserting $X = Z - E(Z)$ into the above inequality
  - Chebyshev’s Inequality: $P(|Z - E(Z)| \geq \alpha) \leq \frac{1}{\alpha^2} V(Z)$ for $\alpha \geq 0$.

- Ex: A rv $X$ is in the interval $\left( E(X) - m\sqrt{V(X)}, E(X) + m\sqrt{V(X)} \right)$ with probability
  \[ P\left( |X - E(X)| < m\sqrt{V(X)} \right) = 1 - P\left( |X - E(X)| \geq m\sqrt{V(X)} \right) \geq 1 - \frac{1}{m^2 V(X)} V(X) = 1 - \frac{1}{m^2} \]
  - When $m = 3$, $P\left( X \in \left( E(X) - 3\sqrt{V(X)}, E(X) + 3\sqrt{V(X)} \right) \right) \geq 1 - \frac{1}{9} \approx 0.88$.
  - When $m = 4$, $P\left( X \in \left( E(X) - 4\sqrt{V(X)}, E(X) + 4\sqrt{V(X)} \right) \right) \geq 1 - \frac{1}{16} \approx 0.9375$.
  - Hence, a random variable is likely to be around its mean.
Jensen’s Inequality

- **Jensen’s Inequality:** \( g(E(X)) \leq E(g(X)) \) for a convex function \( g \).
  - Convex function \( g(x) \) has a supporting hyperplane \( y = ax + b \) at each \( x_0 \) such that \( g(x_0) = ax_0 + b \) and \( g(x) \geq ax + b \) for every \( x \).
  - A hyperplane for a convex function is defined for the convex set \( \{(x, y): g(x) \leq y\} \)
    - The set is called the epigraph of the function and has one more dimension than its lower boundary defining the convex function. For \( g: \mathbb{R}^n \to \mathbb{R} \), the epigraph \( \{(x, y): g(x) \leq y\} \) is in \( \mathbb{R}^{n+1} \).
  - For \( g: \mathbb{R} \to \mathbb{R} \) and random variable \( X \), consider the supporting hyperplane at \( (E(X), g(E(X))) \) given by \( y = a_0 x + b_0 \) for particular values of \( a_0, b_0 \).
    \[ g(E(X)) = a_0 E(X) + b_0 \quad \text{and} \quad (a_0 X + b_0) \leq (g(X)) \]
  - Take the expectation in the inequality and combine:
    \[ g(E(X)) = a_0 E(X) + b_0 = E(a_0 X + b_0) \leq E(g(X)) \]

- **Ex:** \( g(x) = \exp(x) \) is convex as \( g''(x) = \exp(x) \geq 0 \) for \( x \in \mathbb{R} \).
  - Let \( x_0 = 0 \), the derivative at \( x_0 = 0 \) is \( \exp(0) = 1 \).
  - The supporting hyperplane at \( x_0 = 0 \) has the slope of \( a = 1 \), so it has the form \( y = x + b \).
  - To find \( b \), we set \( 1 = g(x_0 = 0) = 0 + b \). This gives \( b = 1 \) and in turn \( y = x + 1 \).
  - You can check visually \( g(x) = \exp(x) \geq x + 1 \).

- **Ex:** Use Jensen's inequality to prove \( E(X^{2k}) - E(X^k)^2 \geq 0 \) for integer \( k \geq 0 \).
  - Let \( g(u) = u^2 \) and note that it is convex.
  - \( (E(X^k))^2 = g\left( E(X^k) \right) \leq E\left( g(X^k) \right) = E(X^{2k}) \).
  - Setting \( k = 1 \) yields \( V(X) = E(X^2) - E(X)^2 \geq 0 \). Also obtain \( E(X^4) \geq (E(X^2))^2 \geq (E(X))^4 \).
Jensen’s Inequality

Ex: In inventory theory, inventory holding and backorder costs can be represented by \( c(y, x) \) for inventory level \( y \) and demand \( x \).
- E.g., \( c(y, x) = h(y - x)^+ + b(x - y)^+ \) with \( h \) and \( b \) respectively as the unit holding and backordering costs.
- \( c(y, x) \) is jointly convex.
- For every inventory level \( y \), we set \( g_y(\cdot) = c(y, \cdot) \), which is convex
- Use Jensen's inequality: \( c(y, E(X)) = g_y(E(X)) \leq E(g_y(X)) = E(c(y, X)) \) for every \( y \).
- The evaluation of the cost against the expected demand turns out to be lower than the true cost.
- This idea can be used to obtain lower bounds for the optimal cost.
**Hölder’s Inequality**

- **Young’s inequality** \( ab \leq \frac{a^p}{p} + \frac{b^q}{q} \) for \( a, b \geq 0; \ p, q > 1 \) and \( \frac{1}{p} + \frac{1}{q} = 1 \)
  - Since \( a, b \geq 0 \), we can set \( s = p \ln a \) and \( t = q \ln b \).
  - From the convexity of \( e^x \), the function \( e^{(p^{-1}s+q^{-1}t)} \leq p^{-1} e^s + q^{-1} e^t = \) chord connecting \((s, e^s)\) to \((t, e^t)\).
  - Inserting the values for \( s, t \) gives the desired inequality \( ab \leq p^{-1} a^p + q^{-1} b^q \).

- **Hölder’s Inequality**: \( E(|X||Y|) \leq (E(|X|^p))^{\frac{1}{p}} (E(|Y|^q))^{\frac{1}{q}} \) for \( \frac{1}{p} + \frac{1}{q} = 1 \).
  - Insert \( a = |X|/(E(|X|^p))^{\frac{1}{p}} \) and \( b = |Y|/(E(|Y|^q))^{\frac{1}{q}} \) into Young’s inequality
  \[
  \frac{|X|}{E(|X|^p)^{\frac{1}{p}}} \frac{|Y|}{E(|Y|^q)^{\frac{1}{q}}} \leq \frac{1}{p} \frac{|X|^p}{E(|X|^p)} + \frac{1}{q} \frac{|Y|^q}{E(|Y|^q)}
  \]
  - Take expected values on the both sides
  \[
  \frac{E(|X||Y|)}{E(|X|^p)^{\frac{1}{p}} E(|Y|^q)^{\frac{1}{q}}} \leq \frac{1}{p} \frac{E(|X|^p)}{E(|X|^p)} + \frac{1}{q} \frac{E(|Y|^q)}{E(|Y|^q)} = 1
  \]

- **Schwarz’s Inequality**: \( E(|X||Y|) \leq \sqrt{E(X^2)} \sqrt{E(Y^2)} \) by setting \( p = q = 2 \) in Hölder’s Inequality.

- Ex: Use Schwarz’s Inequality to get absolute value of correlation at most 1 for two rvs.
  - Shifting rvs by a constant does not affect their covariance or variance \( \Rightarrow \) WLOG, \( E(X) = E(Y) = 0 \).
    - \( \text{Cov}(X + a, Y + b) = E((X + a)(Y + b)) - E(X + a)E(Y + b) = E(XY) - E(X)E(Y) = \text{Cov}(X, Y) \) for \( a, b \in \mathbb{R} \)
  - \[ 1 \geq \frac{E(|X||Y|)}{\sqrt{E(X^2)} \sqrt{E(Y^2)}} = \frac{E(|X||Y|)}{\sqrt{V(X)V(Y)}} \geq \text{max} \left\{ \frac{E(XY)}{\sqrt{V(X)V(Y)}}, - \frac{E(XY)}{\sqrt{V(X)V(Y)}} \right\} = \text{max}\{\text{Cor}(X, Y), -\text{Cor}(X, Y)\} \]
  - Then \(-1 \leq \text{Cor}(X, Y) \leq 1\) for rvs \( X, Y \).
Summary

◆ Inequalities: Markov, Chebyshev, Jensen, Hölder

◆ Limits
  – of Probability: Weak Law of Large Numbers
  – of Events: Strong Law of Large Numbers
  – of Cdfs: Central Limit Theorem