STA347 Midterm Review

Axioms of Expectations

1. If $X \geq 0$, then $E[X] \geq 0$.
2. $E[X_1 + X_2] = E[X_1] + E[X_2]$
3. If c is a constant then $E[cX] = cE[X]$
4. $E[1] = 1$
5. If a seq of r.v.'s $X_i(\omega)$ changes monotonically to a limit $X(\omega)$ then $E[X] = \lim_{n\to\infty} E[X_n]$

Cor: $|E[X]| \leq E|X|$

- If $X \leq Y$ then $E[X] \leq E[Y]$
- X_i's are nonnegative r.v's s.t. Fatou's Lemma:
  $X_i(\omega) \rightarrow X(\omega)$ for some r.v. X.
  Then $\lim_{n\to\infty} \inf E[X_n] \geq E[X]$

? $E|X|$

- Cor of (Dominated Convergence Thm)
  If a seq $X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)$ and $|X_n(\omega)| \leq Y(\omega)$ with $E[Y] < \infty$
  then $E[X_n] \rightarrow E[X]$.
  $X(\omega)$
  (proof of this in lec 1 the end)

Discrete r.v.'s

$E[X] = \sum_{i=1}^{k} P_i X(\omega_i)$ with $P_i \geq 0$ & $\sum_{i=1}^{k} P_i = 1$

note $X(\omega) = \sum_{i=1}^{k} I\{\omega = \omega_i\} X(\omega_i)$

Continuous r.v.'s

$X: \Omega \rightarrow R$

$E[X] = \int_{-\infty}^{\infty} X(\omega) f(\omega) d\omega$

$f(\omega) \geq 0$, $\int_{-\infty}^{\infty} f(\omega) d\omega = 1$

Implications: Let $X = I\{\omega \in A\}$, A is a subset of R.

LHS $= E I\{\omega \in A\} = P\{\omega \in A\} = P(A)$

RHS $= \int_{-\infty}^{\infty} I\{\omega \in A\} f(\omega) d\omega = \int_A f(\omega) d\omega$

§2.5 Moments.

jth moment of a r.v. $\to \mu_j = E[X^j]$

cont. r.v.'s

$$E[(X-E(X))^2] = E(X^2) - (E(X))^2 = \mu_2 - \mu_1^2 = Var(X)$$

§2.7 Equiprobable Outcomes : Sample surveys.

$$E(X) = K^{-1} \sum_{k=1}^{K} X(\omega_k)$$

For a population of N.

$$E(X) = N^{-1} \sum_{k=1}^{N} X(\omega_k) = A(X)$$

$$A(X) = N^{-1} \sum_k x_k$$

$$A(X^2) = N^{-1} \sum_k x_k^2$$

population variance: $V(X) = A(X^2) - A(X)^2$

Let ~~$\xi_j = X(\zeta_j)$~~

$\xi_j = X(\zeta_j)$ be the value of X for the jth member of the sample.

Sample average: $\bar{\xi} = n^{-1} \sum_{j=1}^{n} \xi_j$

Thm 2.7.1. If sampling is without replacement then the Sample mean $\bar{\xi}$ has expectation.

$$E(\bar{\xi}) = A(x)$$

& variance

$$Var(\bar{\xi}) = E[(\bar{\xi} - A(X))^2] = \frac{1}{n} \frac{N-n}{N-1} V(X)$$

§2.8 LSE

$$\hat{X} = \sum_{j=1}^{m} a_j Y_j \to \text{Least square estimate (predictor)}$$

(linear)

$$E[(X-\hat{X})^2] = E(X^2) - 2\sum_j a_j E(XY_j) + \sum_j \sum_k a_j a_k E(Y_j Y_k)$$

In matrix form:

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \qquad \vec{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix}$$

$Cov(X, \vec{Y})$ — Covariance matrix

$$Cov(\vec{Y}) = \begin{pmatrix} Cov(Y_1, Y_1) & \cdots & Cov(Y_1, Y_m) \\ \vdots & & \vdots \\ Cov(Y_m, Y_1) & \cdots & \end{pmatrix}$$

$\hat{X} = a^T Y$ and

$$D(a) = E[(X - \hat{X})^2] = U_{xx} - 2a^T U_{Yx} + a^T U_{YY} a$$

U_{xx} is a scalar, a is column vector $\begin{bmatrix} a_1 \\ \vdots \\ a_j \end{bmatrix}$

U_{Yx} is a Column vector

U_{YY} has jth element $E(Y_j Y_k)$ (product moment matrix)

Thm 2.8.1 $\hat{X} = a^T Y$ yields an LLS est. of X in terms of Y iff $U_{YY} a = U_{Yx}$

$$\boxed{\begin{aligned} &Cov(X, Y_j) = E[(X - E(X))(Y_j - E(Y_j))] = E(XY_j) - E(X)E(Y_j) \\ &U_{xx} \to Var_{xx} = var(X) \\ &U_{Yx} \to V_{Yx} = (cov(X, Y_j)) \\ &U_{YY} \to \text{matrix } V_{YY} = (cov(Y_j, Y_k)) \end{aligned}}$$

~~Thm 2.8.2~~

$$\hat{X} = a_0 + a^T Y$$
$$E(X) = a_0 + a^T E(Y)$$
$$\hat{X} - E(X) = a^T (Y - E(Y))$$

Thm 2.8.2 yields an LLS est. of X in terms of Y allowing for a constant term iff

$$V_{YY} a = V_{Yx}$$

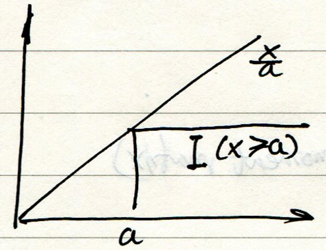
$$cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X) V(Y)}} \in [-1, 1]$$

§2.9. Some implications of the Axioms

① $I(X \geq a) \leq \frac{X}{a}$ $\quad (X \geq 0)$

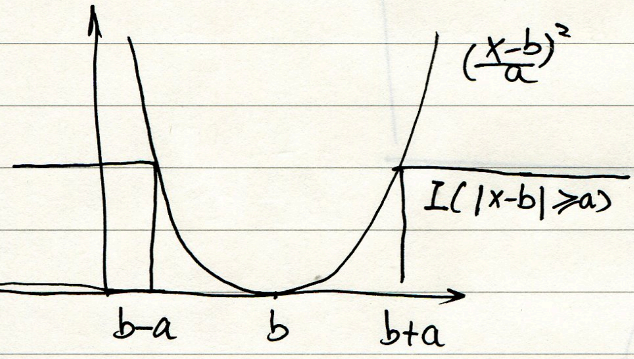
Take expectations on both sides

$P(X \geq a) \leq \frac{E(X)}{a}$ --- Markov inequality



② $I(|X-b| \geq a) \leq \left[\frac{X-b}{a}\right]^2$

$P(|X-b| \geq a) \leq E[(X-b)^2]/a^2$



The bound is minimal when $b = E(X)$

Then Chebyshev's inequality.

$P(|X-E(X)| \geq a) \leq Var(X)/a^2$

Sps $E(X) = \mu$, $var(X) = 0$.

So. $E[(X-\mu)^2] = 0$.

Say X is equal to μ in mean square.

$X \overset{m.s.}{=} \mu$.

Thm 2.9.1. A product matrix is symmetric & nonnegative definite. It's singular iff $C^T X \overset{m.s.}{=} 0$ holds for some nonzero constant vector c.

In particular $|U| \geq 0$ with equality iff holds and

$|E(X_1X_2)|^2 \leq E(X_1^2)E(X_2^2)$ (Cauchy's Ineq') iff $c_1X_1 + c_2X_2 \overset{m.s.}{=} 0$ holds.

Chapter 3 Probability

$A \subset \Omega$, $P(A) = E[I(A)]$

① $I(\bar{A}) = 1 - I(A)$

② A, B disjoint then $I(A \cup B) = I(A) + I(B)$.

③ $A_1, \cdots, A_n$ disjoint, $I(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} I(A_i) = \max_{1 \le i \le n} I(A_i)$

④ If $A \subseteq B$ then $I(A) \le I(B)$

⑤ $I(\bigcap_{i=1}^{m} A_i) = \min_{1 \le i \le m} I(A_i)$

Properties:

① $P(\Omega) = 1$

② $P(A) \geq 0$

③ A, B disjoint $\longrightarrow P(A \cup B) = P(A) + P(B)$

④ $A_1, \cdots, A_n$ a sequence of monotonic events

$P(A_\infty) = \lim_{i \to \infty} P(A_i)$

i. $A_\infty = \bigcup_{i=1}^{\infty} A_i$

ii. $A_\infty = \bigcap_{i=1}^{\infty} A_i$

say for ii.

$X_i = I(A_i)$

$A_{i+1} \subset A_i$

$I(A_i) \geq I(A_{i+1})$

X_i non-increasing

$\lim_{i \to \infty} E(X_i) = E(\lim_{i \to \infty} X_i)$

$\parallel$

$\lim_{i \to \infty} P(A_i)$

if $\lim_{i \to \infty} I(A_i) = 0$

then $\exists\, I(A_m) = 0$ in it.

So $I(A_\infty) = 0$

if $\lim_{i \to \infty} I(A_i) = 1$

$\cdots \forall\, I(A_m) = 1$ in it

So $I(A_\infty) = 1$

Therefore $P(A_\infty) = \lim_{i \to \infty} P(A_i)$ (Take expectation)

Cor:

① A, B not necessarily disjoint.

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

Pf: $I(A \cup B) = I(A) + I(B) - I(AB)$

Why?

b/c. $E(I(A \cup B)) = E(I(A)) + E(I(B)) - E(I(AB))$

$$\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

② $A_1, \cdots A_m$ are events

$$P(\bigcup_{i=1}^{m} A_i) = \sum P(A_i) - \sum P(A_i A_j) + \sum P(A_i A_j A_k) - \cdots + (-1)^r \sum P(A_1 \cdots A_r)$$
$$+ (-1)^m \sum P(A_1 \cdots A_m)$$

(inclusion - exclusion formula)

Pf by induction. (straightforward)

③ $A_1, \cdots, A_n$ disjoint events

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

Pf: Let $B_i = \bigcup_{j=1}^{i} A_j$ B_i is non-decreasing seq.

$$P(\lim_{i\to\infty} B_i) = \lim_{i\to\infty} P(B_i)$$

$$P(\bigcup_{i=1}^{\infty} A_i) = \lim_{i\to\infty} P(\bigcup_{j=1}^{\infty} A_j)$$

$\boxed{\text{Chapter 4 Some basic models}}$

A spatial Model.

N : moleculars in a spatial region

M : equal sized cells.

Z_i = the cell that ith molecular occupies, $i = 1, 2, \cdots, N$.

Let $X = X(Z_1, Z_2, \cdots, Z_n)$

Then $E(X) = M^{-N} \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_N=1}^{M} X(i_1, i_2, \cdots, i_N)$

* Thm If $\downarrow$ holds for every r.v. X, then $Z_1, \cdots, Z_i$ must be uniformly distributed and

$$\boxed{E\left[\prod_k H_k(Z_k)\right] = \prod_k E[H_k(Z_k)] \quad \text{for any functions } H_k \qquad \text{for any functions } H_k.}$$

Pf. $X = I\{Z_i = j\} \rightarrow P(Z_i = j) = \frac{1}{M}$

Let $X = \prod_k^N H_k(Z_k)$

$$E(X) = \frac{1}{M^N} \sum \cdots \sum X = \frac{1}{M} \sum H_1 \frac{1}{M} \sum H_2 \cdots \frac{1}{M} \sum H_N$$
$$= E(H_1(Z_1)) E(H_2(Z_2)) \cdots E(H_N(Z_N))$$

"independent"

5 IMPORTANT DISCRETE ~~RANDOM VARIABLES~~ Distributions

① Bernoulli distributions

Combinations: n choose p $\boxed{\binom{n}{p} = \frac{n!}{(n-p)!\,p!}}$

$X = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$

indicator r.v.'s are Bernoulli.

$X \sim Ber(p)$, $E(X) = 1 \times P(X=1) + 0 \times P(X=0) = p + 0 \cdot (1-p) = p$

$Var(X) = E(X^2) - (E(X))^2 = 1^2 \times P(X=1) + 0^2 \times P(X=0) - p^2 = p - p^2 = p(1-p)$

② Binomial distributions

n trials with p to succeed. X be total # of successes.

X is called a Binomial(n, p) r.v.

$X \sim Binomial(n, p)$

$x = 0, 1, \cdots, n$. $\qquad P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$

If $X \sim$ Binomial (n,p)

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = p + p + \cdots + p = np$$

each trial has an expectation of a

Bernoulli r.v.

- Note if X & Y indpdt, ~~Cov(X, Y) =~~
$$Cov(H(X), G(Y)) = 0 \text{ for any functions } H, G.$$

Pf: $Cov(H(X), G(Y)) = E[H(X)G(Y)] - E(H(X))E(G(Y)) = 0.$

- Note $X_1, \cdots, X_n$ indpdt, then

$$Var(a_1 x_1 + \cdots + a_n X_n)$$
$$= a_1^2 Var(X_1) + \cdots + a_n^2 Var(X_n)$$

Hence

$$Var(X) = V(X_1 + X_2 + \cdots + X_n) = V(X_1) + \cdots + V(X_n)$$
$$= n(1-p)p$$

③ Geometric distributions

X is Geo r.v. with sucess prob. p if X represents the number of indpdt trials prior to the first success.

$$P(X=x) = P(F \cdots FS) = P(F)P(F) \cdots P(S)$$
$$= (1-p)^x p$$

$$E(X) = \sum_{x=0}^{\infty} x p(X=x) = \sum_{x=0}^{\infty} x (1-p)^x p = p \sum_{x=0}^{\infty} x q^x = p \sum_{x=0}^{\infty} \frac{q}{(1-q)^2} = \frac{pq}{p^2} = \frac{q}{p}$$

$$\sum_{j=0}^{\infty} q^j = \frac{1}{1-q}$$

$$\sum_{j=0}^{\infty} j q^{j-1} = \frac{1}{(1-q)^2}$$

$$Var(X) = \frac{q}{p^2}$$

and $\cancel{Var(X)} = E(X^2) \cancel{- E(X)^2}$

$$= \sum x^2 P(X=x)$$

$$= \sum x^2 p q^x$$

$$= q \sum x^2 p q^{x-1}$$

$$= q\left(\frac{2}{p^3} - \frac{1}{p}\right)$$

$\cancel{\text{Then } Var(X)} = E(X^2) - E(X)^2$

$$E(X^2) = \sum x^2 P(X=x)$$

$$= \sum x^2 q^x p$$

$$= p \sum x^2 q^x = \frac{q+q^2}{p^2} = \frac{q(1+q)}{p^2}$$

$$E(X)^2 = \frac{q^2}{p^2}$$

$$VX = E(X^2) - E(X)^2 = \frac{q(1+q)}{p^2} - \frac{q^2}{p^2} = \frac{q}{p^2}$$

④ The negative binomial distributions

X is the number of fails prior to the rth success. indpdt. prob $= p$.

$$P(\underbrace{FFSF \cdots SF \cdots S \cdots S}_{r\text{'s } S})$$

$$P(X=x) = \binom{x+r-1}{r-1} p^{r-1} q^x p = \binom{x+r-1}{r-1} p^r q^x$$

why $x+r-1$ choose $r-1$?
cuz the last S is fixed.
don't need to choose.

- negative binomial : sum of geometrics

$$E(X) = E(Y_1) + E(Y_2) + \cdots + E(Y_r) = \frac{rq}{p}$$

$(X = Y_1 + Y_2 + \cdots + Y_r)$ Y_i is # of trials between ith & $(i-1)$th success

$$Var(X) = Var(Y_1) + \cdots + Var(Y_r) = \frac{rq}{p^2}$$

* if $\frac{V(X)}{E(X)} > 1$ consider negative binomial

⑤ Poisson distribution.

Divide time into n equal sized blocks.

n is large that there won't be two events happening in one block.

$$X = X_1 + X_2 + \cdots + X_n$$

X_i is # of events happened in the ith block.

$$E(X) = nP_n = \lambda$$

$$P(X=x) = \binom{n}{x} P_n^x (1-P_n)^{n-x}$$

$$= \frac{n!}{(n-x)!\,x!} P_n^x (1-P_n)^{n-x}$$

$$= \frac{1}{x!}[n(n-1)(n-2)\cdots(n-x+1)] P_n^x (1-P_n)^{n-x}$$

let $P_n = \lambda/n$, since $\lambda = nP$

$$P(X=x) = \frac{1}{x!}[n(n-1)(\cdots)\cdots(n-x+1)]\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{x}$$

$$= \frac{\lambda^x}{x!}\underbrace{\left[\frac{n(n-1)\cdots(n-x+1)}{n^x}\right]}_{A}\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{B}\Big/\underbrace{\left(1-\frac{\lambda}{n}\right)^x}_{C}$$

$$= \frac{\lambda^x}{x!}e^{-\lambda}$$

$n \to \infty$. $B \to e^{-\lambda}$

$C \to 1$

$A \to 1$

note $\lambda > 0$.

$X \sim \text{Poisson}(\lambda)$

$$E(X) = \sum^{\infty} xP(X=x)$$

$$= \sum_{x\geq 0} x \frac{1}{x!}\lambda^x e^{-\lambda}$$

$$= \lambda e^{-\lambda}\sum \frac{1}{(x-1)!}\lambda^{x-1}$$

$$= \lambda e^{-\lambda}\sum \frac{\lambda^j}{j!}$$

$$= \lambda e^{-\lambda}e^{\lambda}$$

$$= \lambda$$

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

$$= \sum x^2 P(X=x) - \lambda^2$$

$$= \lambda e^{-\lambda}\sum x \frac{1}{(x-1)!}\lambda^{x-1} - \lambda^2$$

$$= \lambda e^{-\lambda}\left(\sum (x-1)\frac{1}{(x-1)!}\lambda^{x-1} + \sum \frac{1}{(x-1)!}\lambda^{x-1}\right) - \lambda^2$$

$$= \lambda e^{-\lambda}\left(\lambda \sum_{x\geq 2}\frac{1}{(x-2)!}\lambda^{x-2} + \sum \frac{1}{(x-1)!}\lambda^{x-1}\right) - \lambda^2$$

$$= \lambda e^{-\lambda}(\lambda e^{\lambda} + e^{\lambda}) - \lambda^2$$

$$= \lambda^2 + \lambda - \lambda^2$$

$$= \lambda$$

Practice Midterm.

(2). $H(x)$ increasing non-negative function of x.

$$P(X \geq a) \leq E[H(X)]/H(a)$$

Markov's ineq.

$$P(X \geq a) = P(H(X) \geq H(a)) \leq \frac{E(H(X))}{H(a)}$$

(5). r.v. X s.t. $E(|X|^3) < \infty$, then

$$(E|X^2|)^2 \leq E|X| \times E(|X|^3)$$

By Cauchy's inequality.

$$(E|X^2|)^2 = E(|\,|X|^{0.5}|X|^{1.5})\big)^2$$
$$\leq E|X|\,E(|X|^3)$$

STA347

After class exercise (so called "HW" part

- corollary: $E[c_1X_1+c_2X_2+\cdots+c_nX_n] = c_1E[X_1]+c_2E[X_2]+\cdots+c_nE[X_n]$
- prove $I(A\cup B)=I(A)+I(B)-I(AB)$
  ~~Hints: take expectations~~
- proof of $Var(X)$ of geometric distribution
- Show if $X\sim Poisson(\lambda)$, then $E(X)=V(X)=\lambda$, and if X,Y iid $Pois(\lambda)$, then $X+Y\sim Pois(2\lambda)$
- In multinomial distr. $Cov(X_i,X_j)=-nP_iP_j$ for $i\neq j$.
- proof of independence, if $n=k$, "$\Leftarrow$" holds then "$\Leftarrow$" holds for $n=k+1$.
- MGF of ~~poison~~ poisson(λ)
- proof of 5 axioms of conditioning probabilities
- §8.2 最后
- $U_n \xrightarrow{a.s.} 0$? $(U_n=\min(X_1,\cdots,X_n))$

formula sheet candidates.

Axiom of expectations (⑤ axioms + ④ Cor

– ④ includes Fatou's Lemma

Dominated Covergence Theorem (proof needs

continuous r.v. expectation formula

Least Square (notations)

$Cor(x,y)=\dfrac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}$

$Cov(x,y)=E[(x-E(x))(y-E(y))]$

$Cov(x,x)=\cancel{\sigma^2}\ Var(x)$

Markov's inequality: X non-negative r.v. $a>0$. $P(X\geq a)\leq \dfrac{E(X)}{a}$

Chebyshev's ineq: $P(|X-E(X)|\geq a)\leq \dfrac{Var(X)}{a^2}$

(How? Let $X \Rightarrow |X-E(X)|$, done)

- Cauchy-Schwartz:
- 5 basis set theory statements (not A of E
- properties of Prob (4) proved by 手
- 3 Cor. ① $P(A\cup B)$ ② Inclu-exclu formula ③ $P(\cup A_i)\leq\sum P(A_i)$

spatial model:

- discrete r.v. { Bern., Bino., Geo, Negative binomial, Poisson, multinomial } (density, E, Var, Corollary
- independent (proof)
- Thm, cor. about MGF, PGF
- know St. Petersberg Paradox (the example?)
- Continuous r.v. { Unif., Exponential, Gamma, Normal }

(continued in next page)

condition $E[X|A] = \cdots$

· Law of Total Probability

· Baye's formula .

Conditional distribution as a r.v. (①②③)

5 properties

·continuous r.vector, CDF, thm

density (HW; · the proof), Cor

§8.2! 掌]

convergence a.s.

in prob.

property: ①②③④

two thms