Measure Theory

1. Algebra Σ

Def. A collection Σ_0 of subsets of S is called an algebra on S if

① $S \in \Sigma_0$ ② $F \in \Sigma_0 \Rightarrow F^c \in \Sigma_0$ ③ $F, G \in \Sigma_0 \Rightarrow F \cup G \in \Sigma_0$

☆ $\phi \in \Sigma_0$, $F \cap G = F^c \cup G^c \in \Sigma_0$.

2. σ-Algebra

Def. A collection Σ of subsets of S is called a σ-algebra on S if Σ is an algebra on S s.t. $\{F_n\}_{n \in N} \subseteq \Sigma \Rightarrow \cup_{n\in N} F_n \in \Sigma$.

☆ $\cap_{n\in N} F_n = (\cup_{n\in N} F_n^c)^c \in \Sigma$.

3. (S, Σ) : measurable space

element of Σ : Σ-measurable subset of S [event]

4. Def. Let C be a class of subsets of S. Then $\sigma(C)$, the σ-algebra generated by C is the smallest σ-algebra Σ on S containing C.

5. Def. Borel σ-algebra on S, $B(S)$ is the σ-algebra generated by the family of open subsets of S.

6. Measure

*A function $\mu: \Sigma \to [0, \infty]$ is called a measure if

(1) $\mu(\phi) = 0$ (2) $\forall$ sequence of disjoint sets in Σ $\{F_n : n \in N\}$,

$\mu(\cup_{n=1}^{\infty} F_n) = \Sigma_{n=1}^{\infty} \mu(F_n)$ (S, Σ, μ) is called a measure space.

When $\mu(S) = 1$, μ is called a probability measure.

(measure $+ \mu(S) = 1$)

7. Independence

Def. Let F be σ-algebra on S. G is called a sub σ-algebra of F if G is itself a σ-algebra and $G \subseteq F$.

Def. Let $G_1, G_2, \cdots$ be sub σ-algebras of F. They are called independent if, whenever $G_{in} \in G_n$ and $P(G_{i_1} \cap G_{i_2} \cap \cdots G_{in}) = \Pi_{k=1}^{n} P(G_{ik})$.

☆ pairwise independent $\not\Rightarrow$ mutually independent.

8. Random variable

Def. Let (Ω, F), (S, Σ) be two measurable spaces. $X: \Omega \to S$ is said to be a measurable function from (Ω, F) to (S, Σ) if

$$X^{-1}(B) := \{w \in \Omega : X(w) \in B\} \in F, \forall B \in \Sigma$$

If $S = R$, $\Sigma = B(R)$, then X is called a random variable and if $S = R^k$, $\Sigma = B(R^k)$, X is called a random vector.

For BLP:

$V\hat{Y} = \rho^2 VY$

*$VU = (1-\rho^2)VY$

9. Checking a mapping is measurable or not ☆ (See notes)
10. We can convert any measure to be probability measure. $P(A) = \frac{\mu(A)}{\mu(S)}$ (S fin

Chapter 2 Probability

1. Bayes Theorem

Let $A_1, A_2, \cdots, A_n$ be mutually exclusive s.t. $P(A_1 \cup A_2 \cup \cdots \cup A_n) = 1$ and $P(A_i) > 0$ $\forall i$. Let E be an arbitrary event s.t. $P(E) > 0$. Then

$$P(A_i|E) = \frac{P(E|A_i)P(A_i)}{\sum_{j=1}^{n} P(E|A_j)P(A_j)} \quad i = 1, 2, \cdots, n.$$

Chapter 3 Random Variables and Probability Distributions

3.1

Def. A r.v. is a variable that takes values according to a certain probability distribution.

3.2 Discrete RV

1. Bivariate

① Marginal Probability. $P(X = x_i) = \sum_{j=1}^{m} P(X = x_i, Y = y_j)$, $i = 1, 2, \cdots, n$

② Conditional Probability. $P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$

2. Multivariate

3.3 Univariate Continuous RV

Density Function $P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)dx$

Conditional Density $f(x|S) = \frac{f(x)}{P(X \in S)}$ for $x \in S$

$= 0$ otherwise

3.4 Bivariate Continuous RV

1. $P(x_1 \le X \le x_2, y_1 \le Y \le y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} \underbrace{f(x,y)}_{\text{joint density}} dx dy$
2. $P[(X,Y) \in S] = \iint f(x,y) dx dy$ Example 3.4.4
3. Thm $f_X(x) = \int_{-\infty}^{\infty} f(x,y) dy$
4. Conditional Density. $f(x,y|S) = \frac{f(x,y)}{P[(X,Y) \in S]}$ for $(x,y) \in S$

$= 0$ otherwise

$$f(x | y_1 \le Y \le y_2) = \frac{\int_{y_1}^{y_2} f(x,y) dy}{\int_{-\infty}^{\infty} \int_{y_1}^{y_2} f(x,y) dy dx}$$

Thm 3.4.2 $f(x | Y = y_1 + cX) = \frac{f(x, y_1 + cx)}{\int_{-\infty}^{\infty} f(x, y_1 + cx) dx}$ ☆ Proof P37

Integral Mean-Value Theorem:

$$\int_a^b f(t)dt = f(c)(b-a)$$

$f(c)$: average value of $f(x)$ on $[a,b]$.

special case $(Y=y_i)$: $f(x|Y=y_i) = \frac{f(x,y_i)}{f(y_i)}$

3.5 Distribution Function

3.6 Change of variables

Thm. Let $f(x)$ be the density of X and let $Y=\phi(X)$, where ϕ is a monotonic differentiable function. Then the density $g(y)$ of Y is given by $g(y) = f[\phi^{-1}(y)] \cdot \left|\frac{d\phi^{-1}}{dy}\right| = \frac{f(x)}{|dy/dx|}$ (Proof P48)

For multivalued: $x_i = \psi_i(y)$ $\quad g(y) = \sum_{i=1}^{n_y} \frac{f(\psi_i(y))}{|\phi'[\psi_i(y)]|} = \sum_{i=1}^{n_y} \frac{f(x_i)}{|\phi'(x_i)|}$.

3.7 Joint Distribution of Discrete and continuous RVs

* $\varphi(x,y_i) = f(x|y_i)P(y_i)$ $\quad$ s.t. $P(a\le X\le b, Y\in S) = \int_a^b \sum_{y_i\in S}\varphi(x,y_i)dx$

or $= P(y_i|x)f(x)$ (Proof See Notes)

Chapter 4 Moments

1. Moment Generating Function

Let X be a r.v. s.t. for some $h>0$, $E(e^{tX})$ exists for all $t\in(-h,h)$. Then $M_X(t) := E(e^{tX})$.

① $M_X(0)=1$ ② if $F_X(z)=F_Y(z)\ \forall z$, then $M_X(t)=M_Y(t)$ in a neighborhood of.

Thm. Let X, Y be rvs, with moment generating functions M_X and M_Y. $M_X = M_Y$ existing in an open interval around 0. Then, $F_X(z)=F_Y(z), \forall z\in\mathbb{R}$.

$M_X'(0)=E(X)$, $M_X''(0)=E(X^2)$ ...

2. Variance

$V\phi(X,Y) = E_X V_{Y|X}\phi(X,Y) + V_X E_{Y|X}\phi(X,Y)$

$V\phi(X,Y) = E[V(\phi(X,Y)|X)] + V[E(\phi(X,Y)|X)]$

$V(X\pm Y) = VX + VY \pm 2Cov(X,Y)$

Chapter 5 Normal RVs

1. Covariance

Defs. $Cov(X,Y) = E[(Y-\mu_Y)(X-\mu_X)]$ $\quad Var(X) = Cov(X,X)$

$= E(YX) - E(X)E(Y)$

2. Conditional Expectation

(1) X,Y discrete. $E(X|Y=y_k) = \sum_{j=1}^{\infty} x_j \underbrace{P(X=x_j|Y=y_k)}_{\frac{P(X=x_j,Y=y_k)}{P(Y=y_k)}}$

(2) X,Y continuous. $E(X|Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$

(3) ILE: $E[E(X|\mathcal{H})|\mathcal{G}] = E(X|\mathcal{G})$ $\quad \mathcal{G}$ is sub σ-Algebra of $\mathcal{H}$.

① LIE

② $V Y = E(V(Y|X)) + V(E(Y|X))$

③ $Var(Y|X) = E(Y^2|X) - E^2(Y|X) = E[(Y-E(Y|X))^2|X]$

(4) X continuous, Y discrete $E(X|Y=y_i) = \int_{-\infty}^{\infty} f(x|y_i)dx$

(5) X discrete, Y continuous $E(X|Y=y) = \sum_{i=1}^{n} x_i P(X=x_i|Y=y)$

2. Binomial RVs

$X \sim B(n,p)$ $P(X=k) = C_n^k p^k q^{n-k}$ $EX = np$ $VX = npq$

3. Normal RVs

$X \sim N(\mu, \sigma^2)$

(1) $EX = \mu$, $VX = \sigma^2$

(2) Let $Y = \alpha + \beta X$, then we have $Y \sim N(\alpha + \beta\mu, \beta^2\sigma^2)$

Corollary: $Z \equiv (X-\mu)/\sigma \sim N(0,1)$

4. Bivariate Normal RVs

$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ 差—协方差矩阵

(1) Thm. Let (X,Y) be bivariate normal, then $E(Y|X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X) = E(Y) + \underbrace{\frac{Cov(X,Y)}{Var(X)}}_{\beta}(X-E(X))$

(Proof See Notes)

* $Var(Y|X) = \sigma_Y^2(1-\rho^2)$

(2) X, Y bivariate normal $\Rightarrow$ $\alpha X + \beta Y$ is normal

(3) $\{X_i\}$ pairwise independent, identically distributed as $N(\mu, \sigma^2) \Rightarrow \bar{X} \equiv \frac{1}{n}\sum X_i \sim N(\mu, \sigma^2/n)$

$E[X^p] = \begin{cases} 0 & \text{if } p \text{ odd} \\ \sigma^p(p-1)!! & \text{if } p \text{ even} \end{cases}$

(4) X, Y bivariate normal, $Cov(X,Y) = 0 \Rightarrow$ X, Y independent.

(5) BLP and BP coincide in the case of the normal distribution.

$9!! = 1 \times 3 \times 5 \times 7 \times 9 = 945$

5. Multivariate Normal RVs

(1) Thm. Let $\overset{n}{x} \sim N(\mu, \Sigma)$, $x' = (\overset{h}{y'}, \overset{k}{z'})$ [Partition] $h+k=n$

$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ $\Sigma_{11} = Vy$ $\Sigma_{12} = E(y-Ey)(z-Ez)'$ $\Sigma_{21} = (\Sigma_{12})'$ $\Sigma_{22} = Vz$.

$E(y|z) = Ey + \Sigma_{12}(\Sigma_{22})^{-1}(z-Ez)$ $V(y|z) = \Sigma_{11} - \Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{21}$

(2) $x \sim N(\mu, \Sigma)$, A be an m×n matrix of constants s.t. $m \le n$ and rows of A are linearly independent. Then $Ax \sim N(A\mu, A\Sigma A')$.

补充: X, Y are said to be jointly normal if they can be expressed as

$X = aU + bV$ $Y = cU + dV$ where U and V are independent normal.

Chapter 6 Large Sample Theory

$|\langle x, y\rangle| \le \|x\| \cdot \|y\|$

1. Inequalities

$VYVX \ge Cov^2(X,Y)$

* (1) Chebyshev: $Pr[g(X_n) \ge \varepsilon^2] \le \frac{E g(X_n)}{\varepsilon^2}$ *Proof

(2) Cauchy-Schwarz: $E\|X'Y\| \le (E\|X\|^2)^{\frac{1}{2}}(E\|Y\|^2)^{\frac{1}{2}}$

(3) $\|E(Y)\| \le E\|Y\|$ (4) $E\|X'Y\| \le (E\|X\|^p)^{\frac{1}{p}}(E\|Y\|^q)^{\frac{1}{q}}$

2. Convergence

(1) Conv a.s. $Pr(\lim_{n\to\infty} |Z_n - Z| \le \delta) = 1$

(2) Conv in Prob $\lim_{n\to\infty} Pr(|Z_n - Z| \le \delta) = 1$

(3) Conv in r^{th} moment $\lim_{n\to\infty} E[\|X_n - X\|^r] = 0$

AMEMIYA: 3: 2, 3, 4, 5, 6, 7, 8

4: ~~2, 3, 6~~ 1, 4, 11, 15, 19

5: 2, 3, 6

6: 9, 10

Specifically, when $r=2$, conv in mean square.

*(4) Conv in Distribution $F(t)=P(X\leq t)$ 称 $X_n \xrightarrow{d} X$ if $F_n(t) \to F(t)$ for each continuity point t of F.

3. Relationship

(1) $X_n \xrightarrow{M} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$ Proof

(2) $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X$, $\not\Leftarrow$ $\quad X_n \xrightarrow{a.s.} X \not\Rightarrow X_n \xrightarrow{d^r} X$, $\not\Leftarrow$

(3) Dominated Convergence Theorem: Suppose exists a r.v. Y s.t. $P(|X_n|\leq |Y|, \forall n)=1$ and $E|Y|^r<\infty$ then if $X_n \xrightarrow{P} X$ or $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{d^r} X$.

(4) $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$, $\not\Leftarrow$ Proof $\quad X_n \xrightarrow{d} c \Rightarrow X_n \xrightarrow{P} c$. Proof

4. CMT

- Let X_n be vector of rvs, g be a function continuous at a constant vector point a. Then $X_n \xrightarrow{P} a \Rightarrow g(X_n) \xrightarrow{P} g(a)$.

5. Slutsky

If $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{P} a$, then:

(i) $X_n + Y_n \xrightarrow{d} X+a$ (ii) $X_nY_n \xrightarrow{d} aX$ (iii) $(X_n/Y_n) \xrightarrow{d} X/a$ $(a\neq 0)$

6. LLN (WLLN Proof P105)

Assume $x_1, x_2, \cdots, x_n$ are i.i.d, $E\|x_i\|<\infty$. $\underbrace{\frac{1}{n}\sum x_i}_{\text{Sample avg}} \xrightarrow{a.s.} \underbrace{E(X)}_{\text{population avg}}$

7. CLT

$x_1, x_2, \cdots, x_n$ are i.i.d. $EX_i^2<\infty$, $Var(x_i)<\infty$, $EX^TX<\infty$ $(E\|X^2\|<\infty)$

$\frac{\bar{X}-E(\bar{X})}{\sqrt{Var(\bar{X})}} \xrightarrow{d} N(0,1)$ $\quad \frac{\frac{1}{n}\sum x_i-\mu}{\sigma_x/\sqrt{n}} \Leftrightarrow \frac{1}{\sqrt{n}}\left(\frac{\sum x_i - n\mu}{\sigma_x}\right) \xrightarrow{d} N(0,1)$

$\sqrt{n}(\bar{y}-\mu)=\frac{1}{\sqrt{n}}\sum(y_i-\mu) \xrightarrow{d} N(0,V)$.

8. ~~Delta~~ Method

If $\sqrt{n}(\hat{\mu}-\mu) \xrightarrow{d} \xi$, $g(u)$ is ~~dt~~ continuously dif in a nbh of μ, then as $n\to\infty$, $\sqrt{n}(g(\hat{\mu})-g(\mu)) \xrightarrow{d} G'\xi$, where $G(u)=\frac{\partial}{\partial u}g(u)'$ and $G=G(\mu)$. In particular, if $\xi \sim N(0,V)$, then $\sqrt{n}(g(\hat{\mu})-g(\mu)) \xrightarrow{d} N(0, G'VG)$. (Proof See Notes)

9. Stochastic Symbols (**)

$O_p(1)$

10. Vector and Matrix Norms

(1) $\|a\|=\sqrt{(a'a)}$ (2) $\|A\|=\sqrt{tr(A'A)}$ (3) $\|ab'\|=\|a\|\|b\|$ (4) $\|aa'\|=\|a\|^2$