Refined Notes for Probability Theory

Yuzhe SHI

December 2019

Contents

Chapter 1

Probability

1.1 Set Operation

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

1.2 Prior Probability

$$P(B|A) = \frac{P(AB)}{P(A)}$$
$$\Rightarrow if\ A \subset B:\ P(B|A) = 1$$
$$\Rightarrow P(A_1 \ldots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\ldots P(A_n|A_1A_2\ldots A_{n-1})$$
$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

1.3 Posterior Probability (Bayesian Formula)

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^{n} P(B|A_k)P(A_k)}$$

1.4 1-Dimensional Probability Distribution

- $$F(x) = \int_{-\infty}^{x} f(x)dx$$
- $$P(X \leq x) = F(x) = \sum_{i=1}^{x} P(x_i)$$
- $$F(-\infty) = 0,\ F(+\infty) = 1$$
- $$\lim_{x \downarrow a} F(x) = F(a)$$
  $$\lim_{x \uparrow b} F(x) = F(b^-)$$
  $$p_i = P(X \leq x) = F(x_i) - F(x_i^-)$$
- $$\int_{-\infty}^{\infty} f(x)dx = 1$$

1.5 Probability Distribution and Density

- 2-Dimensional Random Variable:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$$

$$F(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v)dudv$$

$$F(-\infty,y) = F(x,-\infty) = F(-\infty,-\infty) = 0,\ F(+\infty,+\infty) = 1$$

$$X,\ Y\ are\ independent \Leftrightarrow f_X(x)f_Y(y) = f(x,y)$$

- Edged-Density:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

$$(X,Y) \sim U(D) \Rightarrow\ X \neq U(D_1),\ Y \neq U(D_2)$$

$$(X,Y) \sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho_{XY}) \Rightarrow\ X \sim N(\mu_1,\sigma_1^2),\ Y \sim N(\mu_2,\sigma_2^2)$$

- Prior Probability:

$$f(x|y) = \frac{f(x,y)}{f_Y(y)}$$

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

$$\Rightarrow f_X(x) = \int_{-\infty}^{\infty} f_Y(y)f(x|y)dy$$

$$\Rightarrow f_Y(y) = \int_{-\infty}^{\infty} f_X(x)f(y|x)dx$$

- Y=g(X):

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$$

$$if\ g(X)\ is\ monotonic:\ f_Y(y) = f_X(g^{-1}(y))|(g^{-1}(y)')|$$

- Z=X+Y:

$$f_Z(z) = \int_{-\infty}^{\infty} f(x,z-x)dx$$

and the domain of $z-x$ is restricted to the domain of y to maintain $f_X(x)f_Y(z-x) \neq 0$. Hence, the integration interval of z is splitted as

$$(y_l - x_l, y_h - x_l), (y_l - x_h, y_h - x_h)$$

the upper and lower bound of integration is

$$(y_l - x_l, z), (z - (y_l - x_h), y_h - x_h)$$

if X and Y are independent, we obtain **Convolution Operation**

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$$

$$e.g.\quad X_i \sim N(\mu_i,\sigma_i^2),\ \sum_{i=1}^{n} a_iX_i \sim N(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2)$$

- X is Discrete, Y is Continuous, Z=X+Y:

$$F(Z) = P(X+Y \leq z) = P(X=0, Y \leq z) + \cdots + P(X=k, Y \leq z-k)$$

- $Z = \frac{X}{Y}$

$$f_Z(z) = \int_{-\infty}^{\infty} |y| f(yz, y) dy$$

- $Z = \max\{X_1, \dots, X_n\}$

$$F_Z(z) = (F_X(z))^n$$
$$f_Z(z) = n(F(z))^{n-1} f(z)$$

- $Z = \min\{X_1, \dots, X_n\}$

$$F_Z(z) = 1 - (1 - F_X(z))^n$$
$$f_Z(z) = n(1 - F(z))^{n-1} f(z)$$

Chapter 2

Distribution, Expectation and Variance

2.1 Distributions

Distribution Type	Formula	EX	DX	Independent Additivity
$X \sim B(n,p)$	$C_n^k p^k (1-p)^{n-k}$	np	$np(1-p)$	$B(n_1+n_2,p)$
$X \sim P(\lambda)$	$\frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ	$P(\lambda_1+\lambda_2)$
$X \sim HyperGeometric$	$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$	nM/N	-	-
$X \sim Geometric$	$(1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	-
$X \sim U(a.b)$	$f = \frac{1}{a-b}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	-
$X \sim E(\lambda)$	$f = \lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	-
$X \sim N(\mu,\sigma^2)$	$f = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$

Table 2.1: Distributions

2.2 Properties of Expectation and Variance

- $$EX = \int_{-\infty}^{\infty} xf(x)dx$$
  $$EXY = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy$$
  $$EG(X) = \int_{-\infty}^{\infty} g(x)f(x)dx$$
  $$EG(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$$
- $$E(aX \pm bY) = aE(X) \pm bE(Y)$$
  $$E(\sum_i k_i X_i) = \sum_i k_i E(X_i)$$
- $$X,\ Y\ are\ independent \Rightarrow E(XY) = E(X)E(Y)$$
- $$X_1 \geq X_2 \Rightarrow E(X_1) \geq E(X_2)$$
  $$E|X| \geq |EX|$$

- Cauchy-Schwerz's Inequality:
$$E^2(XY) \leq E(X^2)E(Y^2)$$

-
$$DX = E(X - EX)^2 = E(X^2) - (EX)^2$$

-
$$D(cX) = c^2 D(X)$$

-
$$D(X \pm Y) = D(X) + D(Y) \pm 2E[(X - EX)(Y - EY)]$$
$$D(X \pm Y) = D(X) + D(Y) \pm 2Cov(X, Y)$$

-
$$X,\ Y\ are\ independent \Rightarrow D(XY) = D(X)D(Y) + (EX^2)D(Y) + (EY^2)D(X)$$

- Chebyshev's Inequality:
$$P(|X - EX| \geq \epsilon) \leq \frac{DX}{\epsilon^2}$$
$$P(|X - EX| \leq \epsilon) \geq 1 - \frac{DX}{\epsilon^2}$$

-
$$Cov(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - EXEY$$
$$\Rightarrow Cov(aX, bY) = abCov(X, Y)$$
$$\Rightarrow Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$$

-
$$\rho = \frac{Cov(X, Y)}{\sqrt{DX}\sqrt{DY}}$$

-
$$X,\ Y\ are\ non - correlative \Leftrightarrow Cov(X, Y) = 0 \Leftrightarrow \rho = 0$$

- Independency $\Rightarrow$ Non-Correlation.

2.3 Liev-Lindeberg's Law and De Moivre-Laplace's Law

$$X \sim B(n, p)$$
$$i.e.\ \eta_n \sim N(np, np(1 - p))$$
$$\Rightarrow \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \sim N(0, 1)$$
$$\Rightarrow \lim_{n \to \infty} P(\frac{\eta_n - np}{\sqrt{np(1 - p)}} \leq x) = \Phi(x)$$
$$\Rightarrow \lim_{n \to \infty} P(\frac{\eta_n - n\mu}{\sqrt{n}\sigma} \leq x) = \Phi(x)$$

2.4 Gamma Function

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$$
$$\Rightarrow \Gamma(n) = (n - 1)!,\ \Gamma(\frac{1}{2}) = \sqrt{\pi}$$

Chapter 3

Mathematical Statistics

3.1 Statistical Variable

- $$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \;\Rightarrow\; EX$$

- $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \Rightarrow\; DX$$
  $$\Rightarrow \tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

- $$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \;\Rightarrow\; E(X^k)$$

- $$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k \;\Rightarrow\; E(X - EX)^k = D(X^k)$$

3.2 χ^2 Distribution

$$X_i \sim N(0,1)$$
$$\chi^2 = \sum_{i=1}^{n} X_i^2$$
$$\Rightarrow \chi^2 \sim \chi^2(n)$$
$$\Rightarrow E(\chi^2) = n,\; D(\chi^2) = 2n$$
$$\Rightarrow \chi_1^2 \; and \; \chi_2^2 \; are \; independent, \; \chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$$

3.3 Student Distribution

$$X \sim N(0,1),\; Y \sim \chi^2(n)$$
$$\Rightarrow T = \frac{X}{\sqrt{Y/n}} \sim t(n)$$
$$\Rightarrow E(T) = 0,\; t(\infty) \sim N(0,1)$$

3.4 F Distribution

$$X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$$
$$\Rightarrow F = \frac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$$
$$\Rightarrow 1/F \sim F(n_2, n_1)$$
$$\Rightarrow F_{1-\alpha}(n_2, n_1) = \frac{1}{F_\alpha(n_1, n_2)}$$

3.5 Properties of Normal-distributed Samples

- $$E\overline{X} = EX,\ D\overline{X} = \frac{1}{N}DX$$
$$ES^2 = DX,\ DS^2 = (\frac{\sigma^2}{n-1})^2 D(\frac{(n-1)S^2}{\sigma^2}) = (\frac{\sigma^2}{n-1})^2 D(\chi^2(n-1)) = \frac{2\sigma^4}{n-1}$$

- $$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \Leftrightarrow \frac{\overline{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$$

- $$\overline{X}\ is\ independent\ from\ S^2$$

- $$\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$$
$$\Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2(n-1)$$
$$\Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - EX)^2 \sim \chi^2(n)$$

- $$\frac{\overline{X}-\mu}{S}\sqrt{n} \sim t(n-1)$$

- $$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$$

- $$\frac{\frac{1}{n_1}\sum_{i=1}^{n}(\frac{\overline{X_i}-\mu_1}{\sigma_1})^2}{\frac{1}{n_2}\sum_{i=1}^{n}(\frac{\overline{Y_i}-\mu_2}{\sigma_2})^2} \sim F(n_1, n_2)$$

- $$(X_1, \ldots, X_{n_1}) \leftarrow N(\mu_1, \sigma^2),\ (Y_1, \ldots, Y_{n_2}) \leftarrow N(\mu_2, \sigma^2),$$
$$T = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w}\sqrt{\frac{n_1 n_2}{n_1+n_2}} \sim t(n_1+n_2-2),$$
$$S_w^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

3.6 M-Estimation

$$EX = A_j$$
$$\Rightarrow \int_{-\infty}^{\infty} xf(x;\theta)dx = \frac{1}{n}\sum_{i=1}^{n} X_i^j$$

3.7 Likelihood Estimation

$$L(\theta) = \prod_{i=1}^{n} P(X_i = x_i; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

The target is to find a θ to maximize $L(\theta)$.

$$\Rightarrow \ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i; \theta)$$

$$if\ L\ is\ nonmonotonic: \ \frac{\partial L(\theta)}{\partial \theta} = 0$$

$$if\ L\ is\ monotonic: \begin{cases} \hat{\theta} = \min\{X_1, \dots, X_n\} = X_1^* & x \geq \theta \\ \hat{\theta} = \max\{X_1, \dots, X_n\} = X_n^* & x \leq \theta \end{cases} \tag{3.1}$$

3.8 Estimation Evaluation

$$Unbiased: \ E(\hat{\theta}) = \theta$$

Proof 1

$$E\hat{\theta_M} = E(g(\overline{X})) = \dots = h(\theta)$$

$$E\hat{\theta_L} = \int \hat{\theta_L} f_\theta(x) dx = \dots$$

$$e.g. \quad ES^2\ is\ unbiased\ estimation\ of\ DX,$$

$$E\tilde{S}^2\ is\ biased\ estimation\ of\ DX (= \frac{n-1}{n}\sigma^2),$$

$$\overline{X}\ is\ unbiased\ eatimation\ of\ EX.$$

$$Validate: \ D(\hat{\theta_0}) \leq D(\theta)$$

$$\Rightarrow D(\hat{\theta_0}) = \min_i D(\theta_i)$$

$$Consistency: \ P(|X - EX| \geq \epsilon) \leq \frac{DX}{\epsilon^2}$$

$$\Rightarrow P(|X - EX| < \epsilon) \geq 1 - \frac{DX}{\epsilon^2}$$

$$\Rightarrow P(|\hat{\theta} - \theta| < \epsilon) \geq 1 - \frac{D\hat{\theta}}{\epsilon^2}$$

$$\Rightarrow \lim_{n \to \infty} P(|\hat{\theta} - \theta| < \epsilon) = 1$$

3.9 Confidence Interval

- σ^2 known:

$$\frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N(0, 1)$$

$$\Rightarrow P(\frac{|\overline{X} - \mu|}{\sigma}\sqrt{n} < u_{\alpha/2}) = 1 - \alpha$$

$$\Rightarrow (\overline{X} - \frac{\sigma}{\sqrt{n}}u_{\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}}u_{\alpha/2})$$

- σ^2 unknown:

$$t = \frac{\overline{X} - \mu}{S}\sqrt{n} \sim t(n-1)$$

$$\Rightarrow P(|t| < u_{\alpha/2}) = 1 - \alpha$$

$$\Rightarrow (\overline{X} - \frac{S}{\sqrt{n}}t_{\alpha/2}(n-1), \overline{X} + \frac{S}{\sqrt{n}}t_{\alpha/2}(n-1))$$

- μ unknown:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\Rightarrow P(\chi^2_{1-\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2}) = 1-\alpha$$

$$\Rightarrow \sigma^2 \Rightarrow (\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)})$$

$$\sigma \Rightarrow (\frac{\sqrt{n-1}S}{\sqrt{\chi^2_{\alpha/2}(n-1)}}, \frac{\sqrt{n-1}S}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}})$$

- One-side Confidence Interval:

$$LowerBound \Rightarrow (\overline{X} - \frac{S}{\sqrt{n}} t_\alpha(n-1), +\infty)$$

$$UpperBound \Rightarrow (-\infty, \overline{X} + \frac{S}{\sqrt{n}} t_\alpha(n-1))$$