Probability Summarization

Kelvin · Liang ziyoustep@gmail.com

September, 28, 2018

Contents

1 Set

1.1 Set Operations

Complement:	$\overline{A}$
Union:	$A \cup B$
Intersection:	$A \cap B$

1.2 Set Relations

Disjoint: In a collection of sets, no two of theem have a common element.
Partition: In a collection of sets which are disjoint and their union is Ω.

1.3 Set Algebra

Theorem:

Commutative laws:	$A \cup B = B \cup A$
Associative laws:	$(A \cup B) \cup C = A \cup (B \cup C)$
	$(A \cap B) \cap C = A \cap (B \cap C)$
Distributive laws:	$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$
	$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

2 Counting

2.1 The Counting Principle(Multiplication Principle)

Consider a process that consist of r stages, and there are n_i possible results at the ith stage. Then the total number of possible results of the r-stage process is:

Theorem:

$$n_1 n_2 \cdot \cdots \cdot n_r$$

2.2 Counting Results

Theorem:

Permutation(Distinguishable, without replacement, order matters)

$$n!$$

K-Permutation(Distinguishable, without replacement, order matters)

$$\frac{n!}{(n-k)!}$$

Combination(Distinguishable, without replacement, order no matters)

$$\tbinom{n}{k} = \frac{n!}{k!(n-k)!}$$

Partition(Distinguishable, without replacement, order no matters)

$$\tbinom{n}{n_1, n_2, \cdots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

Number of Subsets

$$2^n$$

3 Probability Axioms

3.1 Probability Laws

1. **Nonnegativity**

$$P(A) \geq 0$$

2. **Additivity**
   If A and B are two disjoint events then:

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

3. **Normalization**

$$P(\Omega) = 1$$

3.2 Some Properties of Probability Laws

1. If $A \subset B$, then $P(A) \leq P(B)$
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. $P(A \cup B) \leq P(A) + P(B)$

4 Conditional Probability

4.1 Definition

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

4.2 Multiplication Rule

Theorem:

$$P\left(\cap_{i=1}^{n} = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n| \cap_{i=1}^{n-1} A_i\right)$$

4.3 Total Probability Theorem

Theorem

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)$$

4.4 Bayes' Rule

Bayes' rule is often used for inference. There are a number of "cause" that may result in a certain "effect". We observe the effect and we wish to infer the cause. Let A_i associate with "cause" and B represents "effect". $P(A_i)$ is called *prior probability* and $P(A_i|B)$ is called *posterior probability*. What Bayes want to do is to find out posterior probability given prior probability.

Bayes' Rule
Let $A_1, A_2, \cdots A_n$ be disjoint events that form a partition of the sample space and assume that $P(A_i) > 0$, for all i. Then for any event B we have:

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i)P(B|A_i)}{P(B)} \\ &= \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n)} \end{aligned}$$

4.5 Independence

4.5.1 Definition

$$P(A|B) = P(A)$$

4.6 Conditioning Independence

4.6.1 Definition

$$P(A \cap B|C) = P(A|C)P(B|C)$$

4.7 Independence of Events

4.7.1 Definition

$$P\left(\underset{i \in B}{\cap} A_i\right) = \prod_{i \in S} P(A_i)$$

5 Discrerte Random Variables

5.1 Probability Mass Function

5.1.1 Definition

$$P_X(x) = P(X = x)$$

5.2 Cumulative Distribution Function

5.2.1 Definition

$$F_X(x) = P(X \leq x)$$

5.3 Expectation

5.3.1 Definition

$$E[X] = \sum_x x P_X(x)$$

5.4 Variance

5.4.1 Definition

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

5.4.2 Standard Deviation

Definition:

$$\sigma_X = \sqrt{Var[X]}$$

5.5 nth Moment

5.5.1 Definition

$$E[X^n]$$

5.6 Functions of Random Variables

Let $Y = g(X)$ then we have:
Theorem:

$$P_Y(y) = \sum_{\{x|g(x)=y\}} P_X(x)$$

5.6.1 Expectation for Functions of Random Variables

Theorem:

$$E[g(X)] = \sum_x g(x) P_X(x)$$

5.6.2 Linear Function of a Random Variable

Let $Y = aX + b$ then:
Theorem:

$$E[Y] = aE[X] + b$$
$$Var[Y] = a^2 Var[X]$$

5.7 Joint PMF

5.7.1 Definition

$$P_{X,Y}(x,y) = P(X = x, Y = y)$$

5.7.2 Marginal PMF

Theorem:

$$P_X(x) = \sum_y P_{X,y}(x,y)$$
$$P_Y(y) = \sum_x P_{X,y}(x,y)$$

5.7.3 Functions of Multiple Random Variable

Let $Z = g(X,Y)$ then:
Theorem:

$$
\begin{aligned}
P_Z(z) &= \sum_{\{(x,y)|g(x,y)=z\}} P_{X,Y}(x,y) \\
E[g(x,y)] &= \sum_x \sum_y g(x,y) P_{X,Y}(x,y) \\
E[aX + bY + c] &= aE[X] + bE[Y] + c
\end{aligned}
$$

5.8 Conditional PMF on Events

5.8.1 Definition

$$P_{X|A}(x) = P(X = x|A) = \frac{P(\{X = x\} \cap A)}{P(A)}$$

Theorem:

$$P_X(x) = \sum_{i=1}^{n} P(A_i) P_{X|A_i}(x)$$

5.9 Conditional PMF on Random Variable

5.9.1 Definition

$$
\begin{aligned}
P_{X|Y}(x|y) &= P(X = x|Y = y) \\
&= \frac{P(X = x, Y = y)}{P(Y = y)} \\
&= \frac{P_{X,Y}(x,y)}{P_Y(y)}
\end{aligned}
$$

Theorem:

$$
\begin{aligned}
P_{X,Y}(x,y) &= P_Y(y) P_{X|Y}(x|y) \\
P_X(x) &= \sum_y P_Y(y) P_{X|Y}(x|y)
\end{aligned}
$$

5.10 Conditional Expectation

5.10.1 Definition

$$E[X|A] = \sum_x x P_{X|A}(x)$$

Theorem:

$$E[g(x)|A] = \sum_x g(x) P_{X|A}(x)$$

If $A_1, A_2, \cdots, A_n$ are disjoint partition then:

$$E[X] = \sum_{i=1} P(A_i) E[X|A_i]$$

5.11 Independence of a Random Variable from an Event

5.11.1 Definition

$$P_{X|A}(x) = P_X(x)$$

5.12 Independence of a Random Variable from a Random Variable

5.12.1 Definition

$$P_{X|Y}(x|y) = P_X(x)$$

Theorem:

If X and Y are independent then:

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

5.13 Conditional Independence

5.13.1 Definition

$$P_{X,Y|A}(x,y) = P_{X|A}(x)P_{Y|A}(y)$$

6 Continuous Random Variable

6.1 Probability Density Function(PDF)

6.1.1 Definition

$$P(X \in B) = \int_B f_X(x)dx$$

$$f_X(x)\ isPDF$$

6.2 Expectation

6.2.1 Definition

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$$

Theorem:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

6.3 Variance

6.3.1 Definition

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

Theorem:

$$Var[X] = \int_{-\infty}^{\infty} (x - E[x])^2 f_X(x)dx$$
$$Var[X] = E[X^2] - (E[X])^2$$

If $Y = aX + b$ then we have:

$$E[Y] = aE[X] + b$$
$$Var[Y] = a^2 Var[X]$$

6.4 Cumulative Distribution Funciton(CDF)

6.4.1 Definition

$$F_X(x) = P(X \leq x) = \begin{cases} \sum\limits_{k \leq x} P_X(k) & \text{if X is discrete} \\ \int_{-\infty}^{x} f_X(t)dt & \text{if X is continuous} \end{cases}$$

6.5 Relation Between CDF and PDF

Theorem:

$$f_X(x) = \frac{dF_X(x)}{dx}$$

6.6 Joint PDF

6.6.1 Definition

$$P((X,Y) \in B) = \int \int_{(x,y) \in B} f_{X,Y}(x,y)dxdy$$
$$f_{X,Y}(x,y)\ isPDF$$

6.6.2 Marginal PDF

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

6.7 Joint CDF

6.7.1 Definition

$$\begin{aligned} F_{X,Y}(x,y) &= P(X \leq x, Y \leq y) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s,t)dtds \end{aligned}$$

6.8 Relation Betwween Joint CDF and Joint PDF

$$f_{X,Y}(x,y) = \frac{\delta^2 F_{X,Y}(x,y)}{\delta x \delta y}$$

6.9 Conditioning a Random Variable on an Event

6.9.1 Definition

$$f_{X|A} = \frac{f_X(x)}{P(X \in A)}$$

If $A_1, A_2, \cdots, A_n$ are disjoint partition, then
Theorem:

$$f_X(x) = \sum_{i=1}^{n} P(A_i) f_{X|A_i}(x)$$

6.10 Joint Conditional PDF

6.10.1 Definition

$$f_{X,Y|C}(x,y) = \frac{f_{X,Y}(x,y)}{P(C)}$$

Theorem:

$$f_{X|C}(x) = \int_{-\infty}^{\infty} f_{X,Y|C}(x,y)dy$$

6.11 Conditioning a Random Variable on another Random Variable

6.11.1 Definition

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Theorem:

$$P(X \in A|Y = y) = \int_A f_{X|Y}(x|y)dx$$

6.12 Condition Expectation

6.12.1 Definition

$$E[X|A] = \int_{-\infty}^{\infty} x f_{X|A}(x)dx$$
$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$

Theorem:

$$E[g(X)|A] = \int_{-\infty}^{\infty} g(x) f_{X|A}(x)dx$$
$$E[g(X)|Y = y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)dx$$

6.13 Total Expectation Theorem

$$E[X] = \sum_{i=1}^{n} P(A_i) E[X|A_i]$$
$$E[X|Y=y] = \int_{-\infty}^{\infty} E[X|Y=y] f_Y(y) dy$$

6.14 Independence

6.14.1 Definition

$$f_{X|Y}(x|y) = f_X(x)$$

If X, Y independent then:
Theorem:

$$f_{X,Y}(x,y) = f_X(x), f_Y(y)$$
$$F_{X,Y}(x,y) = F_X(x) F_Y(y)$$
$$E[XY] = E[X]E[Y]$$
$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$
$$Var[X+Y] = Var[X] + Var[Y]$$

6.15 Bayes' Rule for Continuous Random Variable

If X is *continuous*, then:

$$f_Y(y) f_{X|Y}(x|y) = f_X(x) f_{Y|X}(y|x)$$
$$and$$
$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)}$$
$$= \frac{f_X(x) f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(t) f_{Y|X}(y|t) dt}$$

If N is *discrete*, then:

$$f_Y(y) P(N=n|Y=y) = P_N(n) f_{Y|N}(y|n)$$
$$P(N=n|Y=y) = \frac{P_N(n) f_{Y|N}(y|n)}{f_Y(y)} = \frac{P_N(n) f_{Y|N}(y|n)}{\sum_i P_N(i) f_{Y|N}(y|i)}$$
$$f_{Y|N}(y|n) = \frac{f_Y(y) P(N=n|Y=y)}{P_N(n)} = \frac{f_Y(y) P(N=n|Y=y)}{\int_{-\infty}^{\infty} f_Y(t) P(N=n|Y=t) dt}$$

7 Further Topics on Random Variables

7.1 Calculating PDF of Y=g(X) from Continuous RV X

- **Step1.** Calculate CDF F_Y

$$F_Y(y) = P(g(x) \le y) = \int_{\{x|g(x)\le y\}} f_X(x) dx$$

- **Step2.** Differentiate to obtain PDF f_Y

$$f_Y(y) = \frac{dF_Y(y)}{dy}$$

7.2 Convelution

If $Z = X + Y$, and X, Y are independent.
When *discrete*:

$$P_Z(z) = \sum_x P_X(x) P_Y(z - x)$$

When *continuous*:

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

7.3 Covariance

7.3.1 Definition

$$\begin{aligned} Cov(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

Theorem:

$$\begin{gathered} Cov(X, X) = Var[X] \\ Cov(X, aY + b) = a \cdot Cov(X, Y) \\ Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z) \end{gathered}$$

7.4 Correlation Coefficient

7.4.1 Definition

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var[X]Var[Y]}} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

7.5 Variance of the Sum of Random Variables

Theorem:

$$\begin{gathered} Var[X_1 + X_2] = Var[X_1] + Var[X_2] + 2Cov(X_1, X_2) \\ Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] + \sum_{\{(i,j) \mid i \neq j\}} Cov(X_i, X_j) \end{gathered}$$

7.6 Conditional Expectation and Variance Revisited

7.6.1 Law of Iterated Expectation

$$E[E[X|Y]] = E[X]$$

7.6.2 Law of Total Variance

$$Var[X] = E[Var[X|Y]] + Var[E[X|Y]]$$

7.7 Moment Generating Function

7.7.1 Definition

$$M_X(s) = E[e^{sX}]$$

When *discrete*:

$$M_X(s) = \sum_x e^{sx} P_X(x)$$

When *continuous*:

$$M_X(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x)dx$$

Theorem:
If $Y = aX + b$ then we have:

$$M_Y(s) = E[e^{s(aX+b)}] = e^{sb} M_X(sa)$$

7.7.2 From MGF to Moments

$$\left.\frac{d^n M_X(s)}{d_{s^n}}\right|_{s=0} = \int_{-\infty}^{\infty} x^n f_X(x)dx = E[x^n]$$

7.7.3 MGF of Independent Random Variables

Theorem:
If X, Y independent , $Z = X + Y$ then:

$$M_Z(s) = M_X(s)M_Y(s)$$

Generalization: If $Z = X_1 + \cdots + X_n$ then:

$$M_Z(s) = M_{X_1}(s) \cdots M_{X_n}(s)$$

8 Limit Theorems

8.1 Markov Inequality

If a randowm variable X can only take nonnegative values, then:

$$P(X \geq a) \leq \frac{E[X]}{a}, \text{ for all } a > 0$$

8.2 Chebyshev's Inequality

If X is a random variable with mean u and variance σ^2, then:

$$P(|X-u| \geq c) \leq \frac{\sigma^2}{c^2}, \text{ for all } c > 0$$

Corollary:
Let $C = k\sigma$ we have:

$$P(|X-u| \geq k\sigma) \leq \frac{1}{k^2}$$

8.3 The weak law of large numbers

Let $X_1, X_2, \cdots$ be a sequence of independent identically distributed random variables with mean u, For every $\varepsilon > 0$, we have:

$$P\left(|M_n - u| \geq \varepsilon\right) = P\left(\left|\frac{X_1 + \cdots + X_n}{n} - u\right| \geq \varepsilon\right) \to 0, \ as\ n \to \infty$$

8.4 Convergence in Probability

Let $Y_1, Y_2, \cdots$ be a sequence of random variables(not necessary independent) and let a be a real number, we say that the sequence Y_n converges to a in probability, if for every $\varepsilon > 0$ we have:

$$\lim_{n\to\infty} P\left(|Y_n - a| \geq \varepsilon\right) = 0$$

8.5 Convergence with Probability 1

Let $Y_1, Y_2, \cdots$ be a sequence of random variables(not necessary independent) and let c be a real number, we say that Y_n convergeces to c **with probability 1** if:

$$P\left(\lim_{n\to\infty} Y_n = c\right) = 1$$

8.6 The Strong Law of Large Numbers

Let $X_1, X_2, \cdots$ be a sequence of independent identically distributed random variables with mean u. Then, the sequence of sample means $M_n = (X_1 + \cdots + X_n)/n$ converges to u **with probability 1**, that is

$$P\left(\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = u\right) = 1$$

8.7 Central Limit Theorem

Let $X_1, X_2, \cdots$ be a sequence of independent identically distributed random variables with common mean u and variance σ^2,
Define $S_n = \sum_{i=1}^{n} X_i$ then:

$$\lim_{n\to\infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right) = \Phi(x)\ ,\ -\infty < x < \infty$$

9 References

1. Introduction to Probability, 2nd Edition, by Dimitri P. Betsekas and John N. Tsitsiklis, 2008, ISBN 978-1-886529-23-6