CS 109 Study Notes

Stephen Koo

March 16, 2013

---

1 Fundamentals

1.1 DeMorgan's Laws

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c \qquad\qquad \left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$$

1.2 Axioms of Probability

Axiom 1: $0 \leq P(E) \leq 1$
Axiom 2: $P(S) = 1$
Axiom 3: For any sequence of mutually exclusive events $E_1, E_2, \ldots$

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

1.3 Inclusion-Exclusion Identity

$$P(E \cup F) = P(E) + P(F) - P(EF)$$
$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{r=1}^{n} (-1)^{(r+1)} \sum_{i_1 < \ldots < i_r} P(E_{i_1}, E_{i_2}, \ldots, E_{i_r})$$

1.4 Number of Integer Solutions of Equations

There are $\binom{n-1}{r-1}$ distinct positive integer-valued vectors $(x_1, x_2, \ldots, x_r)$ satisfying the equation

$$x_1 + x_2 + \cdots + x_r = n \qquad\qquad x_i > 0, i = 1, \ldots, r$$

There are $\binom{n+r-1}{r-1}$ distince nonnegative integer-valued vectors $(x_1, x_2, \ldots, x_r)$ satisfying the equation

$$x_1 + x_2 + \cdots + x_r = n$$

2 Conditional Probability

$$P(E|F) = \frac{P(EF)}{P(F)} \Leftrightarrow P(EF) = P(E|F)P(F)$$

2.1 Generalized Chain Rule

$$P(E_1E_2\dots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\dots P(E_n|E_1E_2\dots E_{n-1})$$

2.2 Bayes' Theorem

The many shapes and forms of Bayes' Theorem...

$$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)$$

$$P(F|E) = \frac{P(EF)}{P(E)} = \frac{P(E|F)P(F)}{P(E)}$$

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)}$$

Fully General Form:
If $F_1, F_2, \dots, F_n$ comprise a set of mutually exclusive and exhaustive events, then

$$P(F_j|E) = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

That's odd.
The odds of H given observed evidence E:

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)P(E|H)}{P(H^c)P(E|H^c)}$$

3 Independence

3.1 Definition

Two events are independent if $P(EF) = P(E)P(F)$. Otherwise they are dependent.
More generally, events $E_1, E_2, \dots, E_n$ are independent if for every subset $E_{1'}, E_{2'}, \dots, E_r$ where $r \le n$ it holds that

$$P(E_{1'}E_{2'}\dots E_r) = P(E_{1'})P(E_{2'})\cdots P(E_r)$$

3.2 Conditional Independence

Two events E and F are conditional independent given G if

$$P(EF|G) = P(E|G)P(F|G)$$

Dependent events can become independent, and vice-versa, by conditioning on additional information.

4 Random Distributions

4.1 Definitions and Properties

Probability Mass Function:

$$p(a) = P(X = a)$$

Probability Density Function:

$$P(a \le X \le b) = \int_a^b f(x)dx \qquad P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$$

Cumulative Distribution Function:

$$F(a) = F(X \le a) \text{ where } -\infty < a\infty$$

$$F(a) = \sum_{\text{all } x \le a} p(x) \qquad F(a) = \int_{-\infty}^{a} f(x)dx$$

Density f is the derivative of CDF F: $f(a) = \frac{d}{da}F(a)$

4.2 Joint distributions

Joint Probability Mass Function:

$$p_{X,Y}(a,b) = P(X = a, Y = b)$$

Marginal distributions:

$$p_X(a) = P(X = a) = \sum_y p_{X,Y}(a,y) \qquad p_Y(b) = P(Y = b) = \sum_x p_{X,Y}(x,b)$$

Joint Cumulative Probability Distribution (CDF):

$$F_{X,Y}(a,b) = F(a,b) = P(X \le a, Y \le b) \text{ where } -\infty < a, b < \infty$$

Marginal distributions:

$$F_X(a) = P(X \le a) = P(X \le a, Y < \infty) = F_{X,Y}(a,\infty)$$
$$F_Y(b) = P(Y \le b) = P(X < \infty, Y \le b) = F_{X,Y}(\infty,b)$$

Joint Probability Density Function:

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x,y)dydx$$

$$F_{X,Y}(a,b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x,y)dydx \qquad f_{X,Y}(a,b) = \frac{\partial^2}{\partial a \partial b} F_{X,Y}(a,b)$$

Marginal density functions:

$$f_x(a) = \int_{-\infty}^{\infty} f_{X,Y}(a,y)dy \qquad f_y(b) = \int_{-\infty}^{\infty} f_{X,Y}(x,b)dx$$

4.3 Independent Random Variables

n random variables $X_1, X_2, \ldots, X_n$ are called independent if

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i) \text{ for all } x_1, x_2, \ldots, x_n$$

or analogously for continuous random variables if

$$P(X_1 \le a_1, X_2 \le a_2, \ldots, X_n \le a_n) = \prod_{i=1}^{n} P(X_i \le a_i) \text{ for all } a_1, a_2, \ldots, a_n$$

4.4 Convolution

Let X and Y be independent random variables. The convolution of F_X and F_Y is F_{X+Y}:

$$F_{X+Y}(a) = P(X + Y \le a) = \int_{y=-\infty}^{\infty} F_X(a-y) f_Y(y) dy$$

$$f_{X+Y}(a) = \int_{y=-\infty}^{\infty} f_X(a-y) f_Y(y) dy$$

In discrete case, replace $\int_{y=-\infty}^{\infty}$ with $\sum_y$, and $f(y)$ with $p(y)$.

4.5 Conditional Distributions

Conditional PMF of X given Y:

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

Conditional PDF of X given Y:

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

Conditional CDF of X given Y:

$$\begin{aligned} F_{X|Y}(a|y) = P(X \le a, Y = y) &= \sum_{x \le a} p_{X|Y}(x|y) \\ &= \int_{-\infty}^{a} f_{X|Y}(x|y) dx \end{aligned}$$

n random variables $X_1, X_2, \ldots, X_n$ are conditionally independent given Y if

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | Y = y) = \prod_{i=1}^{n} P(X_i = x_i | Y = y) \text{ for all } x_1, x_2, \ldots, x_n, y$$

or analogously for continuous random variables if

$$P(X_1 \le a_1, X_2 \le a_2, \ldots, X_n \le a_n | Y = y) = \prod_{i=1}^{n} P(X_i \le a_i | Y = y) \text{ for all } a_1, a_2, \ldots, a_n, y$$

It is possible to mix continuous and discrete random variables in conditional distributions. For example let X be a continuous random variable and N be a discrete random variable. Then the conditional PDF of X given N and the conditional PMF of N given X are

$$f_{X|N}(x|n) = \frac{p_{N|X}(n|x) f_X(x)}{p_N(n)}$$
$$P_{N|X}(n|x) = \frac{f_{X|N}(x|n) p_N(n)}{f_X(x)}$$

5 Expectation

5.1 Definitions

The expected value for a discrete random variable X is defined as

$$E[X] = \sum_{x:p(x)>0} x p(x)$$

For a continuous random variable X, the expected value is

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

5.2 Properties

If I is an indicator variable for the event A, then

$$E[I] = P(A)$$

Let $g(X)$ be a real-valued function of X.

$$E[g(X)] = \sum_i g(x_i) p(x_i) \qquad\qquad E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Let $g(X,Y)$ be a real-valued function of two random variables.

$$E[g(X,Y)] = \sum_y \sum_x g(x,y) p_{X,Y}(x,y) \qquad\qquad E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$$

Linearity:

$$E[aX + b] = aE[X] + b$$

N-th Moment of X:

$$E[X^n] = \sum_{x:p(x)>0} x^n p(x)$$

Expected Values of Sums:

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Bounding Expectation:
If random variable $X \geq a$ then $E[X] \geq a$.
If $P(a \leq X < \infty) = 1$ then $a \leq E[X] < \infty$.
If random variables $X \geq Y$ then $E[X] \geq E[Y]$.

5.3 Conditional Expectation

Conditional Expectation of X given $Y = y$:

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

Expectation of conditional sum:

$$E\left[\sum_{i=1}^{n} X_i|Y = y\right] = \sum_{i=1}^{n} E[X_i|Y = y]$$

Expectation of conditional expectations:

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

6 Variance

6.1 Definition

If X is a random variable with mean μ then the variance of X, denoted $\mathrm{Var}(X)$, is:

$$\mathrm{Var}(X) = E[(X - \mu)^2] = E[X^2] - (E[X])^2$$

6.2 Properties

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

If $X_1, X_2, \ldots, X_n$ are independent random variables, then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

6.3 Covariance

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

If X and Y are independent, $\mathrm{Cov}(X, Y) = 0$ *Properties*:

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$$
$$\mathrm{Cov}(X, X) = \mathrm{Var}(X)$$
$$\mathrm{Cov}(aX + b, Y) = a\mathrm{Cov}(X, Y)$$

If $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ are random variables, then

$$\mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{Cov}(X_i, X_j)$$

6.4 Correlation

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

Note: $-1 \le \rho(X,Y) \le 1$.
Correlation measures linearity between X and Y.
If $\rho(X,Y) = 0$, X and Y are uncorrelated.

7 Moment Generating Functions

7.1 Definition

Moment Generating Function (MGF) of a random variable X, where $-\infty < t < \infty$, is

$$M(t) = E[e^{tX}]$$

$$\text{When } X \text{ is discrete:} \qquad = \sum_x e^{tx} p(x)$$

$$\text{When } X \text{ is continuous:} \qquad = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

For any n random variables $X_1, X_2, \ldots, X_n$

$$M(t_1, t_2, \ldots, t_n) = E[e^{t_1 X_1 + t_2 X_2 + \cdots + t_n X_n}]$$

The individual moment generating function is obtained:

$$M_{X_i}(t) = E[e^{tX}] = M(0, \ldots, 0, t, 0, \ldots, 0) \text{ where } t \text{ at } i\text{th place}$$

7.2 Properties

$$M^n(t) = \left(\frac{d^n}{dt^n}\right) M(t) = E[X^n e^{nX}]$$

$$M^n(0) = E[X^n]$$

$$M_X(t) = M_Y(t) \text{ iff } X \sim Y$$

$X_1, X_2, \ldots, X_n$ independent if and only if:

$$M(t_1, t_2, \ldots, t_n) = M_{X_1}(t_1) M_{X_2}(t_2) \ldots M_{X_n}(t_n)$$

8 Inequalities

8.1 Boole's Inequality

Let $E_1, E_2, \ldots, E_n$ be events with indicator random variables X_i.

$$\sum_{i=1}^{n} P(E_i) \ge P\left(\bigcup_{i=1}^{n} E_i\right)$$

8.2 Markov's Inequality

X is a nonnegative random variable.

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

8.3 Chebyshev's Inequality

X is a random variable with $E[X] = \mu$ and $\text{Var}(X) = \sigma^2$.

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

One-sided inequality:

$$P(X \geq E[X] + a) \leq \frac{\sigma^2}{\sigma^2 + a^2} \text{ for any } a > 0$$

$$P(X \leq E[X] - a) \leq \frac{\sigma^2}{\sigma^2 + a^2} \text{ for any } a > 0$$

8.4 Chernoff Bound

X is a random variable with MGF $M(t)$.

$$P(X \geq a) \leq e^{-ta} M(t) \text{ for all } t > 0$$

$$P(X \leq a) \leq e^{-ta} M(t) \text{ for all } t < 0$$

In practice, use the t that minimizes $e^{-ta} M(t)$.

8.5 Jensen's Inequality

If $f(x)$ is a convex function ($f''(x) \geq 0$ for all x) then

$$E[f(x)] \geq f(E[X])$$