MA3238 Midterm Cheatsheet

by Devansh Shah

Probability Review

Definition of Expectation

$$E[X] = \sum_{x \in R_X} xP(X = x)$$

Property of Expectation:

$$E[a + bX] = a + bE[X]$$

Linearity of Expectation does not require independence - it *always* holds true.

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

Minimization of Variance: $E[X]$ is the constant c that minimizes the squared loss $E[(X - c)^2]$.

Variance:

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

Properties of Variance:

$$Var(a + bX) = b^2 Var(X)$$

Moment Generating Function:

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

There is a 1-1 mapping between X and $M_X(t)$, i.e, the MGF completely describes the distribution of the random variable.

Usefulness of MGF:

$$E[X^k] = \frac{d^k}{dt^k} M_X(t)|_{t=0}$$

MGF of Linear Transformation of Variable:

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

Summary of Distributions

	Probability mass function, $p(x)$	Moment generating function, $M(t)$	Mean	Variance
Binomial with parameters n, p; $0 \le p \le 1$	$\binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$	$(pe^t + 1 - p)^n$	np	$np(1-p)$
Poisson with parameter $\lambda > 0$	$e^{-\lambda} \frac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$	$\exp\{\lambda(e^t - 1)\}$	λ	λ
Geometric with parameter $0 \le p \le 1$	$p(1-p)^{x-1}$, $x = 1, 2, \ldots$	$\frac{pe^t}{1 - (1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative binomial with parameters r, p; $0 \le p \le 1$	$\binom{n-1}{r-1} p^r (1-p)^{n-r}$, $n = r, r+1, \ldots$	$\left[\frac{pe^t}{1 - (1-p)e^t}\right]^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

Joint Distribution

$$p_{X,Y}(x, y) = P(X = x, Y = y) = P(\{\omega : X(\omega) = x \wedge Y(\omega) = y\})$$

Marginal Distribution

$$p_X(x) = P(X = x) = \sum_y P(X = x, Y = y) = P(\{\omega : X(\omega) = x\})$$

Covariance

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E[XY] - E[X]E[Y]$$

Correlation

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(x)}\sqrt{Var(Y)}}$$

Variance on linear combination of RVs:

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$$

When X_i's are independent, then $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$ since the pairwise covariance is zero.

Also, when the RVs are independent,

$$M_{\sum_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

That is, under independence of RVs, variance becomes additive and MGF becomes multiplicative.

Conditional Probability

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

Multiplication Law

$$p_{X,Y}(x, y) = p_{X|Y}(x|y) \times p_Y(y) = p_{Y|X}(y|x) \times p_X(x)$$

Law of Total Probability

$$p_X(x) = \sum_y p_Y(y) p_{X|Y}(x|y)$$

Bayes Theorem

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

Conditional Independence We say $X \perp Y$ given Z if for any x, y, z:

$$P(X = x, Y = y|Z = z) = P(X = x|Z = z)P(Y = y|Z = z)$$

Note: Independence and Conditional Independence are unrelated.

Law of Iterated Expectation

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

Law of Total Variance

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

Random Sum: $Y = \sum_{i=1}^{N} X_i$ where X_i's are i.i.d with mean μ and variance σ^2, and N is also random.

Expectation:

$$E[Y] = \mu E[N]$$

Variance:

$$Var(Y) = \sigma^2 E[N] + \mu Var(N)$$

Moment Generating Function:

$$M_Y(t) = M_N(ln(M_X(t)))$$

Markov Chain

Markovian Property: What happens afterwards $t > n$ is conditionally independent of what happened before $t < n$ given X_n.

Chapman-Kolmogorov Equations for higher order transition matrices:

$$P^{n,n+m+1} = P^{n,k} * P^{k,n+m+1} \quad \forall n < k < n + m + 1$$

Stationary MC: Transition Probability Matrix P does not depend on time n.

First Step Analysis

Express the quantity of interest as:

$$a_i = E[\sum_{n=0}^{T} g(X_n)|X_0 = i]$$

for every state i, and see what happens after one-step transitions.

General Solution to Gambler's Ruin

Case 1: When $p = 1/2$,

$$\text{P(broke)} = 1 - \frac{k}{N}$$

$$E[\text{games played}] = k(N - k)$$

Case 2: When $p \neq 1/2$,

$$\text{P(broke)} = 1 - \left(\frac{1 - (q/p)^k}{1 - (q/p)^N}\right)$$

$$E[\text{games played}] = \frac{1}{(p - q)} \left[\frac{N(1 - (q/p)^k)}{1 - (q/p)^N} - k\right]$$

A drunk man will find his home, but a drunk bird may get lost forever.

Classification of States

Accessible: $i \to j \implies \exists m > 0, P_{ij}^{(m)} > 0$

Communication: $i \longleftrightarrow j \implies i \to j \wedge j \to i$

Communication is an equivalence relation.

Irreducible MC: Only one communication class.

Reducible MC: Multiple communication classes.

Return Probability

$$P_{ii}^{(n)} = P(X_n = i | X_0 = i)$$

If $P_{ii}^{(n)} \to 0$ when $n \to \infty$, then the state i is transient.

First Return Probability

$$f_{ii}^{(n)} = P(X_1 \neq i, X_2 \neq i, \dots, X_{n-1} \neq i, X_n = i | X_0 = i)$$

Relation betewen return probability and first return probability:

$$P_{ii}^{(n)} = \sum_{k=0}^{n} f_{ii}^{(k)} P_{ii}^{(n-k)}$$

Define f_{ii} as the total probability of revisiting i in the future:

$$f_{ii} = \sum_{n=0}^{\infty} f_{ii}^{(n)} = \lim_{N \to \infty} \sum_{n=0}^{N} f_{ii}^{(n)}$$

A state i is said to be **recurrent** if $f_{ii} = 1$, and **transient** if $f_{ii} < 1$

Note: If i is a recurrent state, it does NOT imply that $P_{ii}^{(n)} \to 1$ as $n \to \infty$.

Number of Revisits

$$N_i = \sum_{i=0}^{\infty} I(X_n = i)$$

Theorem of Number of Revisits

- For transient state,

$$E[N_i | X_0 = i] = \frac{f_{ii}}{1 - f_{ii}}$$

- For recurrent state,

$$E[N_i | X_0 = i] = \infty$$

Number of Revisits and Return Probability:

$$E[N_i | X_0 = i] = \sum_{i=1}^{\infty} P_{ii}^{(n)}$$

Summary of Recurrent and Transient States

$$\text{R} \iff f_{ii} = 1 \iff \sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty \iff E[N_i | X_0 = i] = \infty$$

$$\text{T} \iff f_{ii} < 1 \iff \sum_{n=1}^{\infty} P_{ii}^{(n)} < \infty \iff E[N_i | X_0 = i] < \infty$$

State in the same communication class are either all recurrent or all transient.
An MC with finite states must have at least one recurrent class.

Long Run Performance

Period: For a state i, let $d(i)$ be the greatest common divisor of $\{n : n \geq 1, P_{ii}^{(n)} > 0\}$. If $\{n : n \geq 1, P_{ii}^{(n)} > 0\}$ is empty (starting from i, the chain will never revisit i), then we define $d(i) = 0$.
Aperiodic: State i is aperiodic $\iff d(i) = 1$
Periodicity Theorem For a MC, let $d(i)$ be the period of state i, then:

1. If i and j can communicate, $d(i) = d(j)$
2. There is an N such that $P_{ii}^{(N*d(i))} > 0$, and for any $n \geq N$, $P_{ii}^{(n*d(i))} > 0$
3. There is $m > 0$ such that $P_{ji}^{(m)} > 0$, and when n is sufficiently large, we have $P_{ji}^{(m+nd(i))} > 0$

Regular Markov Chain

A MC with transition probability matrix $\mathbf{P}$ is regular if $\exists k > 0, \forall i, j,\ P_{ij}^{k} > 0$.

If a MC is irreducible, aperiodic, with finite states, then it is a regular MC.

Main Theorem: Suppose P is a regular transition probability matrix with states $S = \{1, 2, \dots, N\}$. Then,

1. The limit $\lim_{n\to\infty} p_{ij}^{(n)}$ exists. Meaning, as $n \to \infty$, the marginal probability of $P(X_n = j | X_0 = i)$ will converge to a finite value.
2. The limit does not depend on the initial state, and we write:
$$\pi_j = \lim_{n \to \infty} P_{ij}^{(n)}$$
3. The distribution of all of the π_k is a probability distribution, i.e., $\sum_{k=1}^{N} \pi_k = 1$, and this is the **limiting distribution**
4. The limits $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ are the solution of the system of equations:
$$\pi_j = \sum_{k=1}^{N} \pi_k P_{kj}, \quad j = 1, 2, \dots, N$$
$$\sum_{k=1}^{N} \pi_k = 1$$
In matrix form,
$$\pi P = \pi, \quad \sum_{k=1}^{N} \pi_k = 1$$
5. The limiting distribution π is unique.

Interpretations of π

- π_j is the (marginal) probability that the MC is in state j for the long run (regardless of the actual instant of time, and the initial state, hence "marginal").
- π gives the limit of $\mathbf{P}^n$
- π can be seen as the long run proportion of time in every state. That is,
$$E\left[\frac{1}{m} \sum_{k=0}^{m-1} I(X_k = j) | X_0 = i\right] \to \pi_j \text{ as } m \to \infty$$
Until time m (for a large value of m), the chain visits state j around $m \times \pi_j$ times.

Irregular Markov Chain

2 possibilities:

1. $|S| = \infty$ and $\pi_i = 0$ for all i (which means that all the states are transient).
2. We find a solution π for $\pi P = \pi$ (the distribution doesn't "move")

Stationary Distribution A distribution $(p_1, p_2, \dots)$ on S is called a stationary distribution, if it satisfies for all $i = 1, 2, \dots$ that:

$$P(X_n = i) = p_i \implies P(X_{n+1} = i) = p_i$$

Note that if the initial distribution of X_0 is not π, we cannot claim any results.

For a regular MC, the stationary distribution is also a limiting distribution.

A key observation is that the stationary distribution must have $\pi_i = 0$ for all transient states i