MA3238 Finals 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:

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

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:

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

Moment Generating Function:

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

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

Accessibility: For a stationary MC $\{X_n, n = 0, 1, 2, \dots\}$ with transition probability matrix P, state j is said to be accessible from state i, denoted by $i \to j$, if $P_{ij}^{(m)} > 0$ for some $m \geq 0$.

Communication: If two states i and j are accessible from each other, i.e., $i \to j$ and $j \to i$, then they are said to communicate, denoted by $i \longleftrightarrow j$.

Reducibility: An MC is irreducible if ALL the states communicate with one another (i.e,. there is a single communication class). Otherwise, the chain is said to be reducible (more than one communication class).

Return Probability: For any state i, recall the probability that starting from state i and returns at i at the nth transition is that: $P_{ii}^{(n)} = P(X_n = i|X_0 = i)$.

By definition, $P_{ii}^{(0)} = 1$, $P_{ii}^{(1)} = P_{ii}$.

First Return Probability: For any state i, define the probability that starting from state i, the first return to i is at the nth transition:

$f_{ii}^{(n)} = P(X_1 \neq i, X_2 \neq i, \dots, X_{n-1} \neq i, X_n = i|X_0 = i)$. We set $f_{ii} = 0$.

Relationship between Return Probability and First Return Probability:

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

Note: Recurrency $\not\Longrightarrow$ $P_{ii}^{(n)} \to 1$.

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

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

Number of Revisits:

- If $f_{ii} < 1$ (i.e,. i is transient), there is $E[N_i|X_0 = i] = \dfrac{f_{ii}}{1 - f_{ii}}$
- If $f_{ii} = 1$ (i.e,. i is recurrent), there is $E[N_i|X_0 = i] = \infty$

We also have:

- $P(N_i \geq m|X_0 = i) = f_{ii}^m$ (probability of revisiting the state more than m times).
- $E[N_i|X_0 = i] = \sum_{n=1}^{\infty} P_{ii}^{(n)}$

Equivalent Definitions of Recurrence and Transience:

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

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

Note:

- If i and j are in the same communication class, then either they are both recurrent or they're both transient.
- Corollary: 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 $d(i) = 0$.

If $d(i) = 1$, we call the state i to be **aperiodic**.

Periodicity Theorem:

1. If i and j can communicate, $d(i) = d(j)$
2. There is a threshold 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 $P_{ji}^{(m+nd(i))} > 0$

If all the states in an MC have period $= 1$, then we say that the MC is aperiodic.

Regular MC: A Markov Chain with transition probability matrix P is called regular if there exists an integer $k > 0$ such that all the elements P^k are strictly positive (non-zero).

If a Markov Chain 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

Long Run Performance for Infinite MCs

First Return Time: $R_i = \min\{n \geq 1, X_n = i\}$. In words, it is the first time that the process X_n returns to i.

Relationship between first-return time, and first-return probability:

$f_{Ii}^{(n)} = P(R_i = n|X_0 = i)$.

Mean Duration Between Visits:

$$m_i = E[R_i|X_0 = i] = \sum_{n=1}^{\infty} nP(R_n = i|X_0 = i) = \sum_{n=1}^{\infty} nf_{ii}^{(n)}$$

Note that we can only define m_i when $f_{ii} = 1$. When we have $f_{ii} < 1$, then the probability that there are infinitely many steps between 2 visits is non-zero, and equal to $1 - f_{ii}$ so the expectation will be infinity (which is not very meaningful).

Limit Theorem

For any **recurrent irreducible MC**, define:

$$m_i = E[R_i|X_0 = i] = \sum_{n=1}^{\infty} nf_{ii}^{(n)}$$

Then,

1. For any $i, j \in S$,
$$\lim_{n\to\infty} \sum_{k=1}^{n} P_{ij}^{(k)}/n = 1/m_j$$
2. If $d = 1$, then
$$\lim_{n\to\infty} \sum_{n=1}^{\infty} P_{ij}^{(n)} = 1/m_j$$
3. If $d > 1$, then
$$\lim_{n\to\infty} \sum_{n=1}^{\infty} P_{jj}^{(nd)} = d/m_j$$

Note that the theorem applies for MCs with infinitely many states too! It also applies for periodic MCs.

Remarks:

- When $m_j = \infty$, the limiting probability at each state is 0, although it is recurrent. We call such a MC to be null recurrent. For example, consider the symmetric random walk with $p = 1/2$ and no absorbing state. Note that it is still recurrent (there's only one class so it must be recurrent).
- When $m_j < \infty$, the limiting probability at each state is $1/m_j$. In such a case, we call it a positive recurrent MC. e.g. Random walk with $p < 1/2$ (process eventually reaches 0) and "reflection" at 0, i.e., $P(X_n = 1|X_{n-1} = 0) = 1$
When $d > 1$, we can only consider the steps nd.
When $d = 1$, the limiting probability is positive, which means that it is a regular MC.

Basic Limit Theorem

For a positive recurrent ($m_j < \infty$), irreducible, and aperiodic MC,

- $\lim_{n\to\infty} P_{ij}^{(n)}$ exists for any i, j and is given by:
$$\lim_{n\to\infty} P_{ij}^{(n)} = \lim_{n\to\infty} P_{jj}^{(n)} = \frac{1}{m_j}$$
- If π is the solution to the equation $\pi P = \pi$, then we have:
$$\pi_j = \frac{1}{m_j}$$

A positive recurrent, irreducible, aperiodic MC is called an **ergodic** MC. Hence, the basic limit theorem applies to all ergodic MCs. We do NOT require the MC to have finite/infinite states for the theorem to hold.

Procedure for a General MC

1. Find all the classes C_k
2. Set up a new MC where every recurrent class is denoted by one state. Then, find $P(\text{absorbed in recurrent class } C_k | X_0 = i)$ denoted by $u_{k|i} \rightarrow$ this gives the probability of entering any recurrent class, given the initial distribution.
3. We can ignore all transient classes because the process will eventually leave them in the long-run, i.e., their long-term probability is zero.
4. For every recurrent class C_k, we find the period d.
   (a) Aperiodic (d =1): find the corresponding limiting distribution of state j in this class, denoted by $\pi_{j|k}$, by considering the sub-MC restricted on C_k
   (b) Periodic ($d > 1$): there is NO limiting distribution, but we can still check the long-run proportion of time in each state by finding m_j (i.e., we can still find π but the interpretation is different in this case)
5. Consider the initial state $X_0 = i$:
   (a) If j is transient, then $\pi_j = 0$
   (b) If $j \in C_k$ is recurrent, then:
$$\pi_{j|i} = u_{k|i}\pi_{j|k}$$
6. Finally, given the initial distribution $X_0 \sim \pi_0$, then:
$$\pi_{j|\pi_0} = \sum_{i\in S} \pi_{j|i}\pi_0(i)$$

Branching Process

Suppose initially there are X_0 individuals. In the n-th generation, the X_n individuals independently give rise to number of offsprings $\xi_1^{(n)}, \xi_2^{(n)}, \cdots, \xi_{X_n}^{(n)}$, which are i.i.d. random variables with the same distribution as:
$$P(\xi = k) = p_k, \quad k = 0, 1, 2, \cdots$$
The total number of individuals produced for the $(n+1)$-th generation is:
$$X_{n+1} = \xi_1^{(n)}, \xi_2^{(n)}, \cdots, \xi_{X_n}^{(n)}$$
Then, the process $\{X_n\}_{n=0}^{\infty}$ is a **branching process**.
An important (and strong) assumption of the branching process is that ξ is not dependent of X_n.

Partial Information

If we are only given the mean μ and variance σ^2 of ξ, and suppose $X_0 = k$:

- $$E[X_n|X_0 = k] = k\mu^n$$
- $$Var(X_n|X_0 = k) = k\mu^{n-1}\sigma^2 \times \begin{cases} \frac{1-\mu^n}{1-\mu}, & \mu \neq 1 \\ n, & \mu = 1 \end{cases}$$

In the derivation of the above, we use the law of total variance for a random sum: $Var(X_{n+1}) = \mu^2 Var(X_n) + \sigma^2 E[X_n]$

Complete Information

Probability Generating Function (PGF) For a discrete random variable X, the probability generating function is defined as:
$$\phi_X(t) = E[t^X] = \sum_{k=0}^{\infty} P(X = k)t^k$$
Note: If X and Y are independent, then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.
Distribution of X_n given $X_0 = k$:
$$\phi_{X_n}(t) = [\phi_\xi^{(n)}(t)]^k$$
Extinction Probability: Here, u_n is the probability of going extinct *by* the nth generation.
$$u_n^{(k)} = [\phi_\xi^{(n)}(0)]^k$$
Eventually Extinct: If $u_\infty = 1$, it means the population is guaranteed to go extinct eventually.
The value of u_∞ must be the solution of the equation:
$$x = \phi_\xi(x), \quad x \in [0, 1]$$
$\phi_\xi(x)$ is an increasing function on $(0, 1]$. The second derivative is also positive - hence, $\phi_\xi(x)$ will increase faster and faster. Note that:
$$\frac{d}{dx}\phi_\xi(x)|_{x=1} = \sum_{k=1}^{\infty} P(\xi = k) \cdot k \cdot 1^{k-1} = \sum_{k=0}^{\infty} kP(\xi = k) = E[\xi]$$
Consider a branching process with the distribution of ξ as F. The extinction probability u_∞ can be found as follows:

- If $P(\xi = 0)$, then $u_\infty = 0 \rightarrow$ no chance of extinction because every individual generates at least one offspring.
- If $P(\xi = 0) > 0$ and $E[\xi] < 1$, then the process is called subcritical, and $u_\infty = 1$ (the population eventually goes extinct)
- If $P(\xi = 0) > 0$ and $E[\xi] = 1$, then the process is called critical and $u_\infty = 1$ (still goes extinct)
- If $P(\xi = 0) > 0$ and $E[\xi] > 1$, then the process is called supercritical and $u_\infty < 1$, and it can be found by the equation: $x = \phi(x)$ where $\phi(x) = \sum_k P(\xi = k)x^k$

Page Rank Algorithm

- The state space S is the set of all webpages
- Index set $T = \{0, 1, 2, \cdots\}$
- Transition Probability Matrix:
$$P_{ij} = \begin{cases} \frac{1}{\text{\# of connected webpages}}, & \text{if there is an arrow from i to j} \\ 0, & \text{otherwise} \end{cases}$$

For an irreducible and positive recurrent MC induced, we order the webpages in the order:
$$(\pi_N)_{(1)} \geq (\pi_N)_{(2)} \geq \cdots \geq (\pi_N)_{(|S|)}$$
To handle absorbing states, we add perturbation to the MC at every step.
$$\pi_{n+1} = (1-\lambda)\pi_n P + \lambda\pi_0$$
where $0 < \lambda < 1$

MCMC Sampling

Global Balanced Equations:
$$\forall j, \ \pi(j) = \sum_{k\in S} \pi(k)P_{kj}$$
Local Balanced Equations:
$$\forall i \neq j, \ \pi(i)P_{ij} = \pi(j)P_{ji}$$
Local Balanced Equations in terms of Thinning Parameter:
$$\pi(i)Q_{ij}\alpha(i,j) = \pi_j Q_{ji}\alpha(j,i)$$
where $0 < \alpha \leq 1$

Hastings Metropolis Algorithm

1. Set up Q so that the MC with transition probability matrix Q is irreducible
2. Define $\alpha(i, j)$ as:
$$\alpha(i,j) = \min\left(\frac{\pi_j Q_{ji}}{\pi_i Q_{ij}}, 1\right)$$
3. Then, P is obtained as:
$$P_{ij} = Q_{ij}\alpha(i,j), \quad i \neq j \tag{1}$$
$$P_{ii} = Q_{ii} + \sum_{k\neq i} Q_{ik}(1 - \alpha(i,k)) \tag{2}$$

Simulation Algorithm

```
TOTAL_STEPS = 5000 # large enough to ensure convergence
process = [] # track the path of the process
x = 1 # initial state
for step in 1...TOTAL_STEPS
  obtain t from T ~ Binom(max(2 * x, 2), 1/2)
  calculate alpha(X_n, t)
  generate u from U ~ uniform(0, 1)
  if (u < alpha) {
    x = t # accept jump from X_n to y, i.e. X_{n+1} = t
  } else {
    x = x # no jump, thinning
  }
  process.add(x)
# cut of the first 1000 steps
process = process[1001:]
```

To use MCMC sampling, we only need the kernel function, not the normalising constant.

Poisson Process

Poisson Distribution

If $X \sim Poi(\lambda)$,
$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \ x = 0, 1, 2, \cdots$$

- Mean = λ, Variance = λ, PGF = $\exp[\lambda(t-1)]$
- When $n \rightarrow \infty$ and $p_n \rightarrow 0$, then $Poi(\lambda)$ is a good approximation for $Bin(n, p_n)$ where $\lambda = np_n$ is a constant.
- If $X \sim Poi(\lambda_1), Y \sim Poi(\lambda_2)$, then $X + Y \sim Poi(\lambda_1 + \lambda_2)$
- If $X \sim Po(\lambda)$ and $Z|X \sim Binomial(X, r)$, then $Z \sim Poi(\lambda r)$

Defining a Poisson Process

Definition 1: Using Poisson distribution.
X is a Poisson process with parameter λ if:

- $X(0) = 0$
- For any $t \geq 0$, $X(t) \sim Poi(\lambda t)$
- for any $s \geq 0, t > 0$, we have $X(s+t) - X(s) \sim Poi(\lambda t)$

Definition 2: Law of Rare Events Let $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ be independent Bernoulli random variables where $P(\epsilon_i = 1) = p_i$, and let $S_n = \epsilon_1 + \cdots + \epsilon_n$. The exact probability for S_n, and the Poisson probability with $\lambda = p_1 + \cdots + p_n$ differ by at most:
$$\left|P(S_n = k) - \frac{e^{-\lambda}\lambda^k}{k!}\right| \leq \sum_{i=1}^{n} p_i^2$$
Let $N((s,t])$ be a RV counting the number of events occurring in the interval $(s,t]$. Then, $N((s,t])$ is a Poisson process of intensity $\lambda > 0$ if:

- The process increments $N((t_0, t_1]), N((t_1, t_2]), \cdots, N((t_{n-1}, t_n])$ are independent random variables.
- $$P(N((t, t+h]) = k) = \begin{cases} 1 - \lambda h - o(h), & k = 0 \\ \lambda h, & k = 1 \\ o(h), & k \geq 2 \end{cases}$$

Definition 3: Using waiting times.

- We can completely specify a Poisson process by simply recording the waiting times (or the sojourn times).
- The waiting time W_1 has (exponential) PDF:
$$f_{W_1}(t) = \lambda e^{-\lambda t}, \ t \geq 0$$
- For $n \geq 2$, W_n follows a gamma distribution with PDF:
$$f_{W_n}(t) = e^{-\lambda t}\frac{\lambda^n t^{n-1}}{(n-1)!}, \ n = 1, 2, \cdots, \ t \geq 0$$
- Exponential distributions have a memorylessness property.
- Given that $X(t) = 1$, we have: $f_{W_1}(x) = \frac{1}{t}$ for all $x \leq t$ and 0 otherwise (uniform on the interval $(0, t]$.
- Given that $X(t) = n$, the joint distribution of n independent $Unif(0, t)$ random variables (followed by ordering in ascending order) gives the distribution of the waiting times to be:
$$f(w_1, w_2, \cdots, w_n | X(t) = n) = \frac{n!}{t^n}$$
- The PDF of the kth order statistic (i.e., the kth waiting time in this case) given that $X(t) = n$ is given by:
$$f_k(x) = \frac{n!}{(n-k)!(k-1)!}\frac{1}{t}\left(\frac{x}{t}\right)^{k-1}\left(\frac{t-x}{t}\right)^{n-k}$$

Appendix

Table 1: Common Discrete Distributions

Distribution	PMF	Mean	Variance	MGF	PGF
Bernoulli	$f(x;p) = p^x(1-p)^{1-x}$	p	$p(1-p)$	$M(t;p) = 1 - p + pe^t$	$G(z;p) = 1 - p + pz$
Binomial	$f(x;n,p) = \binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$	$M(t;n,p) = (1-p+pe^t)^n$	$G(z;n,p) = (1-p+pz)^n$
Poisson	$f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ	$M(t;\lambda) = e^{\lambda(e^t-1)}$	$G(z;\lambda) = e^{\lambda(z-1)}$
Geometric	$f(x;p) = (1-p)^{x-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$M(t;p) = \frac{pe^t}{1-(1-p)e^t}$	$G(z;p) = \frac{pz}{1-(1-p)z}, \lvert z\rvert < \frac{1}{1-p}$

Table 2: Common Continuous Distributions

Distribution	PDF	Mean	Variance	CDF	MGF
Uniform	$f(x;a,b)=\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$F(x;a,b)=\frac{x-a}{b-a}$	$M(t;a,b)=\frac{e^{tb}-e^{ta}}{t(b-a)}$
Normal	$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2	$\Phi(x;\mu,\sigma)=\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$	$M(t;\mu,\sigma)=e^{\mu t+\frac{1}{2}\sigma^2t^2}$
Exponential	$f(x;\lambda)=\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$F(x;\lambda)=1-e^{-\lambda x}$	$M(t;\lambda)=\frac{\lambda}{\lambda-t}$, $t<\lambda$
Gamma	$f(x;\alpha,\beta)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$\gamma(\alpha,\beta)=\frac{1}{\Gamma(\alpha)}\gamma(\alpha,\beta x)$	$M(t;\alpha,\beta)=\left(\frac{\beta}{\beta-t}\right)^\alpha$, $t<\beta$