ECE 302: Probability and Applications[1]

Week 5 Topics

- Popular Discrete Random Variables
  - Uniform, Bernoulli, Binomial, Geometric, Zipf
  - Poisson from Binomial
- Popular Continuous Random Variables
  - Uniform, Exponential, Pareto
  - Gamma
- Popular Continuous Random Variables
  - Uniform, Beta

1 Popular Discrete Random Variables

Some random variables arise in many diverse, unrelated applications. They are pervasive because they model fundamental mechanisms that underlie random behavior. In this section we present popular discrete random variables and discuss how they arise and how they are interrelated.

1.1 Summary of Discrete Random Variables

Bernoulli Random Variable

$$S_X = \{0, 1\}$$
$$p_0 = q = 1 - p \quad p_1 = p \quad 0 \le p \le 1$$
$$E[X] = p \quad \text{VAR}[X] = p(1-p) \quad G_X(z) = (q + pz)$$

Remarks: The Bernoulli random variable is the value of the indicator function I_A for some event $A; X = 1$ if A occurs and 0 otherwise.

Binomial Random Variable

$$S_X = \{0, 1, \ldots, n\}$$
$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \ldots, n$$
$$E[X] = np \quad \text{VAR}[X] = np(1-p) \quad G_X(z) = (q + pz)^n$$

Remarks: X is the number of successes in n Bernoulli trials and hence the sum of n independent, identically distributed Bernoulli random variables.

Geometric Random Variable
First Version:

$$S_X = \{0, 1, 2, \ldots\}$$
$$p_k = p(1-p)^k \quad k = 0, 1, \ldots$$
$$E[X] = \frac{1-p}{p} \quad \text{VAR}[X] = \frac{1-p}{p^2} \quad G_X(z) = \frac{p}{1 - qz}$$

Remarks: X is the number of failures before the first success in a sequence of independent Bernoulli trials. The geometric random variable is the only discrete random variable with the memoryless property.

Second Version:

$$S_{X'} = \{1, 2, \ldots\}$$
$$p_k = p(1-p)^{k-1} \quad k = 1, 2, \ldots$$
$$E[X'] = \frac{1}{p} \quad \text{VAR}[X'] = \frac{1-p}{p^2} \quad G_{X'}(z) = \frac{pz}{1-qz}$$

Remarks: $X' = X + 1$ is the number of trials until the first success in a sequence of independent Bernoulli trials.

Negative Binomial Random Variable

$S_X = \{r, r+1, \ldots\}$ where r is a positive integer
$$p_k = \binom{k-1}{r-1} p^r (1-p)^{k-r} \quad k = r, r+1, \ldots$$
$$E[X] = \frac{r}{p} \quad \text{VAR}[X] = \frac{r(1-p)}{p^2} \quad G_X(z) = \left(\frac{pz}{1-qz}\right)^r$$

Remarks: X is the number of trials until the r th success in a sequence of independent Bernoulli trials.

Poisson Random Variable

$$S_X = \{0, 1, 2, \ldots\}$$
$$p_k = \frac{\alpha^k}{k!} e^{-\alpha} \quad k = 0, 1, \ldots \quad \text{and } \alpha > 0$$
$$E[X] = \alpha \quad \text{VAR}[X] = \alpha \quad G_X(z) = e^{\alpha(z-1)}$$

Remarks: X is the number of events that occur in one time unit when the time between events is exponentially distributed with mean $1/\alpha$.

Uniform Random Variable

$$S_X = \{1, 2, \ldots, L\}$$
$$p_k = \frac{1}{L} \quad k = 1, 2, \ldots, L$$
$$E[X] = \frac{L+1}{2} \quad \text{VAR}[X] = \frac{L^2-1}{12} \quad G_X(z) = \frac{z}{L}\frac{1-z^L}{1-z}$$

Remarks: The uniform random variable occurs whenever outcomes are equally likely. It plays a key role in the generation of random numbers.

Zipf Random Variable

$S_X = \{1, 2, \ldots, L\}$ where L is a positive integer

$p_k = \frac{1}{c_L}\frac{1}{k} \quad k = 1, 2, \ldots, L$ where c_L is given by Eq. (3)

$E[X] = \frac{L}{c_L} \quad \text{VAR}[X] = \frac{L(L+1)}{2c_L} - \frac{L^2}{c_L^2}$

Remarks: The Zipf random variable has the property that a few outcomes occur frequently but most outcomes occur rarely.

Discrete random variables arise mostly in applications where counting is involved. We begin with the Bernoulli random variable as a model for a single coin toss. By counting the outcomes of multiple coin tosses we obtain the binomial, geometric, and Poisson random variables.

1.2 The Bernoulli Random Variable

Let A be an event related to the outcomes of some random experiment. The **Bernoulli random variable** I_A equals one if the event A occurs, and zero otherwise. I_A is a discrete random variable since it assigns a number to each outcome of S. It is a discrete random variable with range $= \{0, 1\}$, and its **pmf** is

$$p_I(0) = 1 - p \quad \text{and} \quad p_I(1) = p$$

where $P[A] = p$.

The **mean** of of I_A is

$$E[I_A] = 0p_1(0) + 1p_1(1) = p.$$

The sample mean in n independent Bernoulli trials is simply the relative frequency of successes and converges to p as n increases:

$$\langle I_A \rangle_n = \frac{0N_0(n) + 1N_1(n)}{n} = f_1(n) \to p$$

The **variance** of the Bernoulli random variable I_A is:

$$\text{VAR}[I_A] = E[I_A^2] - E[I_A]^2 = p - p^2 = p(1-p) = pq.$$

- The variance is quadratic in p, with value zero at $p = 0$ and $p = 1$ and maximum at $p = 1/2$. This agrees with intuition since values of p close to 0 or to 1 imply a preponderance of successes or failures and hence less variability in the observed values.

- The maximum variability occurs when $p = 1/2$ which corresponds to the case that is most difficult to predict.

- Every Bernoulli trial, regardless of the event A, is equivalent to the tossing of a biased coin with probability of heads p.

- In this sense, coin tossing can be viewed as representative of a fundamental mechanism for generating randomness, and the Bernoulli random variable is the model associated with it.

1.3 The Binomial Random Variable

The **Binomial random variable** X is the number of times a certain event A occurs in n independent trials of a random experiment. X is has range $S_X = \{0, 1, \ldots, n\}$.If we let I_j be the indicator function for the event A in the j th trial, then

$$X = I_1 + I_2 + \ldots + I_n$$

that is, X is the sum of the Bernoulli random variables associated with each of the n independent trials.

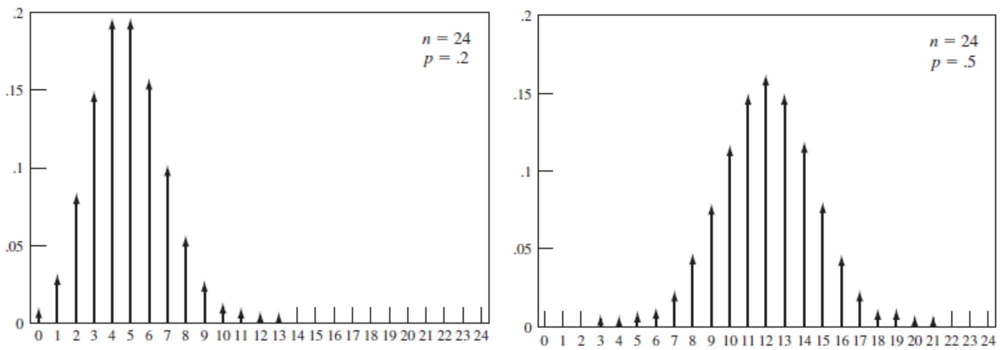
In our discussion on sequential experiments, we found that X has the Binomial **pmf** that depends on n and p:

$$P[X = k] = p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } k = 0, \ldots, n$$

- Figure 1 shows the pdf of X for $n = 24$ and $p = .2$ and $p = .5$.

- Note that $P[X = k]$ is maximum at $k_{\text{max}} = [(n+1)p]$, where $[x]$ denotes the largest integer that is smaller than or equal to x.

- When $(n+1)p$ is an integer, then the maximum is achieved at k_{max} and $k_{\text{max}} - 1$.



(a) Events A and B are independent. **(b)** Events A and C are not independent.

Figure 1: pmf's of binomial random variable (a) $\rho = 0.2$; (b) $\rho = 0.5$

The factorial terms grow large very quickly and cause overflow problems in the calculation of $\binom{n}{k}$. If we take the ratio of successive terms in the pmf allows, we can calculate $p_X(k+1)$ in terms of $p_X(k)$ and delay the onset of overflows:

$$\frac{p_X(k+1)}{p_X(k)} = \frac{n-k}{k+1}\frac{p}{1-p} \qquad \text{where } p_X(0) = (1-p)^n$$

- The binomial random variable arises in applications where there are two types of objects (i.e., heads/tails, correct/erroneous bits, good/defective items, active/silent speakers), and we are interested in the number of type 1 objects in a randomly selected batch of size n, where the type of each object is independent of the types of the other objects in the batch.

Example 1: Mean of a Binomial Random Variable

The **expected value or mean** of X is:

$$\begin{aligned} E[X] &= \sum_{k=0}^{n} k p_X(k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \end{aligned}$$

$$= np\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1}(1-p)^{n-k}$$

$$= np\sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} p^{j}(1-p)^{n-1-j} = np$$

where the first line uses the fact that the $k = 0$ term in the sum is zero, the second line cancels out the k and factors np outside the summation, and the last line uses the fact that the summation is equal to one since it adds all the terms in a binomial pmf with parameters $n-1$ and p.

The expected value $E[X] = np$ agrees with our intuition since we expect a fraction p of the outcomes to result in success.

Example 2: Variance of a Binomial Random Variable

To find $E\left[X^2\right]$ below, we remove the $k = 0$ term and then let $k' = k - 1$:

$$\begin{aligned} E\left[X^2\right] &= \sum_{k=0}^{n} k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} k \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\ &= np\{(n-1)p + 1\} = np(np+q) \end{aligned}$$

In the third line we see that the first sum is the mean of a binomial random variable with parameters $(n-1)$ and p, and hence equal to $(n-1)p$. The second sum is the sum of the binomial probabilities and hence equal to 1.

We obtain the variance as follows:

$$\sigma_X^2 = E\left[X^2\right] - E[X]^2 = np(np+q) - (np)^2 = npq = np(1-p)$$

- We see that the **variance** of the binomial is n times the variance of a Bernoulli random variable.
- We observe that values of p close to 0 or to 1 imply smaller variance, and that the maximum variability is when $p = 1/2$.

Example 3: Redundant Systems

A system uses triple redundancy for reliability: Three microprocessors are installed and the system is designed so that it operates as long as one microprocessor is still functional. Suppose that the probability that a microprocessor is still active after t seconds is $p = e^{-\lambda t}$. Find the probability that the system is still operating after t seconds.

Let X be the number of microprocessors that are functional at time t. X is a binomial random variable with parameter $n = 3$ and p. Therefore:

$$P[X \geq 1] = 1 - P[X = 0] = 1 - \left(1 - e^{-\lambda t}\right)^3$$

1.4 The Geometric Random Variable

The **geometric random variable** arises when we count the number M of independent Bernoulli trials until the first occurrence of a success. M is called the geometric random variable and it takes on values from the set $\{1, 2, \ldots\}$. In our discussion on sequential experiments, we found that the **pmf** of M is given by

$$P[M = k] = p_M(k) = (1-p)^{k-1}p \quad k = 1, 2, \ldots$$

where $p = P[A]$ is the probability of "success" in each Bernoulli trial. Figure 3(b) in Week 4 Notes shows the geometric pmf for $p = 1/2$. Note that $P[M = k]$ decays geometrically with k, and that the ratio of consecutive terms is $p_M(k+1)/p_M(k) = (1-p) = q$. As p increases, the pmf decays more rapidly.

The probability that $M \leq k$ can be written in closed form:

$$P[M \leq k] = \sum_{j=1}^{k} pq^{j-1} = p \sum_{j'=0}^{k-1} q^{j'} = p\frac{1-q^k}{1-q} = 1 - q^k$$

Sometimes we are interested in $M' = M - 1$, the number of failures before a success occurs. We also refer to M' as a geometric random variable. Its pmf is:

$$P\left[M' = k\right] = P[M = k+1] = (1-p)^k p \quad k = 0, 1, 2, \ldots$$

In Examples 13 and 17 in Week 4 Notes, we found the **mean and variance** of the geometric random variable:

$$m_M = E[M] = 1/p \quad \mathrm{VAR}[M] = \frac{1-p}{p^2}$$

We see that the mean and variance increase as p, the success probability, decreases.

The geometric random variable is the only discrete random variable that satisfies the memoryless property:

$$P[M \geq k+j \mid M > j] = P[M \geq k] \text{ for all } j, k > 1$$

- The above states that if a success has not occurred in the first j trials, then the probability of having to perform at least k more trials is the same as the probability of having to perform at least k trials if we were starting from scratch.
- Thus, each time a failure occurs, the system "forgets" and begins anew as if it were performing the first trial. For this reason we say that X is *memoryless*.
- The geometric random variable arises in applications where one is interested in the time (i.e., number of trials) that elapses between the occurrence of events in a sequence of independent experiments. Examples where the modified geometric random variable M' arises are: number of customers awaiting service in a queueing system; number of white dots between successive black dots in a scan of a black-and-white document.

1.5 The Poisson Random Variable

In many applications, we are interested in counting the number of occurrences of an event in a certain time period or in a certain region in space.

- The Poisson random variable arises in situations where the events occur "completely at random" in time or space.
- For example, the Poisson random variable arises in counts of emissions from radioactive substances, in counts of demands for phone connections, and in counts of defects in a semiconductor chip.

The **pmf** for the Poisson random variable is given by

$$P[N=k] = p_N(k) = \frac{\alpha^k}{k!}e^{-\alpha} \quad \text{for } k = 0, 1, 2, \ldots \tag{1}$$

where α is the average number of event occurrences in a specified time interval or region in space.

- Figure 2 shows the Poisson pmf for several values of α.
- For $\alpha < 1, P[N = k]$ is maximum at $k = 0$; for $\alpha > 1, P[N = k]$ is maximum at $[\alpha]$;
- if α is a positive integer, the $P[N = k]$ is maximum at $k = \alpha$ and at $k = \alpha - 1$.

The pmf of the Poisson random variable sums to one, since

$$\sum_{k=0}^{\infty} \frac{\alpha^k}{k!} e^{-\alpha} = e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} = e^{-\alpha} e^{\alpha} = 1$$

where we the second summation is the infinite series expansion for e^{α}.

It is easy to show that the **mean and variance** of a Poisson random variable is given by:

$$E[N] = \alpha \quad \text{and} \quad \sigma_N^2 = \text{VAR}[N] = \alpha$$

Example 4: Queries at a Call Center

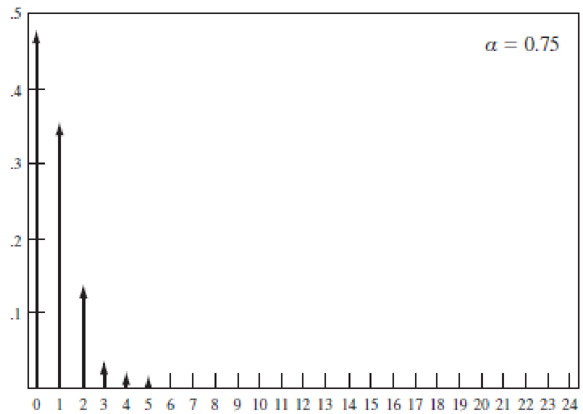
The number N of queries arriving in t seconds at a call center is a Poisson random variable with $\alpha = \lambda t$ where λ is the average **arrival rate** in queries/second. Assume that the arrival rate is four queries per minute. Find the probability of the following events: (a) more than 4 queries in 10 seconds; (b) fewer than 5 queries in 2 minutes.

The arrival rate in queries/second is $\lambda = 4$ queries /60sec $= 1/15$ queries/sec. In part a, the time interval is 10 seconds, so we have a Poisson random variable with $\alpha = (1/15$ queries /sec) * 10 seconds $= 10/15$ queries. The probability of interest is evaluated numerically:

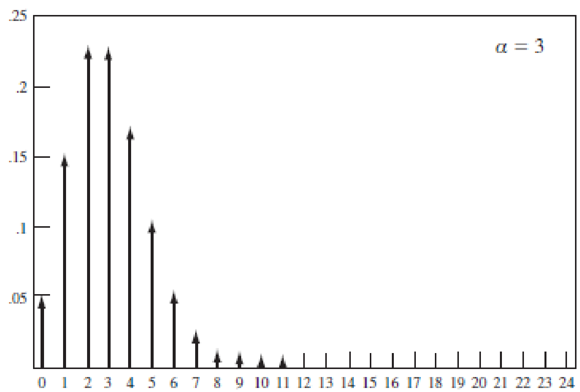
$$P[N > 4] = 1 - P[N \leq 4] = 1 - \sum_{k=0}^{4} \frac{(2/3)^k}{k!} e^{-2/3} = 6.33 \left(10^{-4}\right)$$

In part b, the time interval of interest is $t = 120$ seconds, so $\alpha = 1/15 * 120$ seconds $= 8$. The probability of interest is:

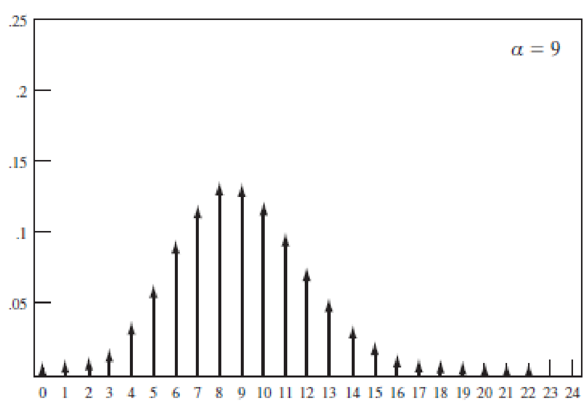
$$P[N \leq 5] = \sum_{k=0}^{5} \frac{(8)^k}{k!} e^{-8} = 0.10$$



(a) $\alpha = 0.75$



(b) $\alpha = 3$



(c) $\alpha = 9$

Figure 2: Probability mass functions of Poisson random variable

Example 5: Arrivals at a Packet Multiplexer

The number N of packet arrivals in t seconds at a multiplexer is a Poisson random variable with $\alpha = \lambda t$ where λ is the average arrival rate in packets/second. Find the probability that there are no packet arrivals in t seconds.

$$P[N = 0] = \frac{\alpha^0}{0!} e^{-\lambda t} = e^{-\lambda t}$$

- This equation has an interesting interpretation. Let Z be the time until the first packet arrival. Suppose we ask, "What is the probability that $X > t$, that is, the next arrival occurs t or more seconds later?" Note that $\{N = 0\}$ implies $\{Z > t\}$ and vice versa, so $P[Z > t] = e^{-\lambda t}$.
- The time until the next arrival has an exponential distribution with mean $1/\lambda$.

We can also express the probability of n or more arrivals in t seconds in terms of Poisson probabilities as follows:

$$P[N(t) \geq n] = 1 - P[N(t) < n] = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

The Poisson probabilities in Eq. (1) *were developed by Poisson to approximate the binomial probabilities in the case where p is very small and n is very larg*e, that is, where the event A of interest is very rare but the number of Bernoulli trials is very large. *We show that if $\alpha = np$ is fixed, then as n becomes large:*

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{\alpha^k}{k!} e^{-\alpha} \quad \text{for } k = 0, 1, \ldots \tag{2}$$

Equation (2) is obtained by taking the limit $n \to \infty$ in the expression for p_k, while keeping $\alpha = np$ fixed. First, consider the probability that no events occur in n trials:

$$p_0 = (1-p)^n = \left(1 - \frac{\alpha}{n}\right)^n \to e^{-\alpha} \quad \text{as } n \to \infty$$

where the limit in the last expression is a well known result from calculus. Consider the ratio of successive binomial probabilities:

$$\frac{p_{k+1}}{p_k} = \frac{(n-k)p}{(k+1)q} = \frac{(1-k/n)\alpha}{(k+1)(1-\alpha/n)}$$
$$\to \frac{\alpha}{k+1} \quad \text{as } n \to \infty$$

The limiting probabilities satisfy

$$p_{k+1} = \frac{\alpha}{k+1} p_k = \left(\frac{\alpha}{k+1}\right)\left(\frac{\alpha}{k}\right)\cdots\left(\frac{\alpha}{1}\right)p_0 = \frac{\alpha^k}{k!}e^{-\alpha}$$

Thus the Poisson pmf can be used to approximate the binomial pmf for large n and small p, using $\alpha = np$.

Example 6: Errors in Optical Transmission

An optical communication system transmits information at a rate of 10^9 bits/second. The probability of a bit error in the optical communication system is 10^{-9}. Find the probability of five or more errors in 1 second.

Each bit transmission corresponds to a Bernoulli trial with a "success" corresponding to a bit error in transmission. The probability of k errors in $n = 10^9$ transmissions (1 second) is then given by the binomial probability with $n = 10^9$ and $p = 10^{-9}$. The Poisson approximation uses $\alpha = np = 10^9\,(10^{-9}) = 1$. Thus

$$P[N \geq 5] = 1 - P[N < 5] = 1 - \sum_{k=0}^{4} \frac{\alpha^k}{k!}e^{-\alpha}$$
$$= 1 - e^{-1}\left\{1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!}\right\} = .00366$$

The Poisson random variable appears in numerous physical situations because many models are very large in scale and involve very rare events.

- For example, the Poisson pmf gives an accurate prediction for the relative frequencies of the number of particles emitted by a radioactive mass during a fixed time period. This correspondence can be explained as follows.

- A radioactive mass is composed of a large number of atoms n.
- In a fixed time interval each atom has a very small probability p of disintegrating and emitting a radioactive particle.
- If atoms disintegrate independently of other atoms, then the number of emissions in a time interval can be viewed as the number of successes in n trials.
- For example, one microgram of radium contains about $n = 10^{16}$ atoms, and the probability that a single atom will disintegrate during a one millisecond time interval is $p = 10^{-15}$.
- Clearly the conditions for the approximation in Eq. (2) hold: n is so large and p so small that one could argue that the limit $n \to \infty$ has been carried out and that the number of emissions is exactly a Poisson random variable.

The Poisson random variable also comes up in situations where we can imagine a sequence of Bernoulli trials taking place in time or space. Suppose we count the number of event occurrences in a T-second interval. Divide the time interval into a very large number, n, of subintervals as shown in Fig. 3. A pulse in a subinterval indicates the occurrence of an event. Each subinterval can be viewed as one in a sequence of independent Bernoulli trials if the following conditions hold:

1. At most one event can occur in a subinterval, that is, the probability of more than one event occurrence is negligible;

2. the outcomes in different subintervals are independent; and

3. the probability of an event in a subinterval is $p = \alpha/n$, where α is the average number of events observed in a 1-second interval.

The number N of events in 1 second is a binomial random variable with parameters n and $p = \alpha/n$. Thus as $n \to \infty$, N becomes a Poisson random variable with parameter α.

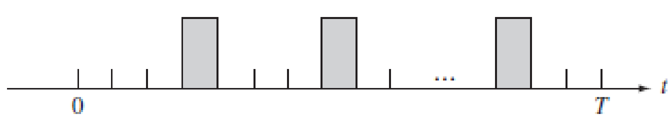


Figure 3: Event occurrences in n subintervals of $[0, T]$

1.6 The Uniform Random Variable

The **discrete uniform random variable** Y takes on values in a set of consecutive integers $S_Y = \{1, \ldots, L\}$ with equal probability:

$$p_Y(k) = \frac{1}{L} \quad \text{for} \quad k \in \{1, \ldots, L\}$$

This humble random variable occurs whenever outcomes are equally likely, e.g., toss of a fair coin or a fair die, spinning of an arrow in a wheel divided into equal segments, selection of numbers from an urn. It is easy to show that the mean and variance are:

$$E[Y] = \frac{L+1}{2} \quad \text{and} \quad \text{VAR}[Y] = \frac{L^2-1}{12}$$

Example 7: Discrete Uniform Random Variable in Unit Interval
Let X be a uniform random variable in $S_X = \{0, 1, \ldots, L-1\}$. We define the discrete uniform random variable in the unit interval by

$$U = \frac{X}{L} \quad \text{so} \quad S_U = \left\{0, \frac{1}{L}, \frac{2}{L}, \frac{3}{L}, \ldots, 1 - \frac{1}{L}\right\}$$

U has pmf:

$$p_U\left(\frac{k}{L}\right) = \frac{1}{L} \quad \text{for } k = 0, 2, \ldots, L-1$$

- The pmf of U puts equal probability mass $1/L$ on equally spaced points $x_k = k/L$ in the unit interval.
- The probability of a subinterval of the unit interval is equal to the number of points in the subinterval multiplied by $1/L$.
- As L becomes very large, this probability is essentially the length of the subinterval

1.7 The Zipf Random Variable

The **Zipf random variable** is named for George Zipf who observed that *the frequency of words in a large body of text is proportional to their rank.*

Suppose that words are ranked from most frequent, to next most frequent, and so on. Let X be the rank of a word, then $S_X = \{1, 2, \ldots, L\}$ where L is the number of distinct words. The **pmf** of X is:

$$p_X(k) = \frac{1}{c_L}\frac{1}{k} \quad \text{for } k = 1, 2, \ldots, L$$

where c_L is a normalization constant. The second word has 1/2 the frequency of occurrence as the first, the third word has 1/3 the frequency of the first, and so on. The **normalization constant** c_L is given by the sum:

$$c_L = \sum_{j=1}^{L} \frac{1}{j} = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{L} \tag{3}$$

The constant c_L occurs frequently in calculus and is called the L th harmonic mean and increases approximately as $\ln L$.

- For example, for $L = 100, c_L = 5.187378$ and $c_L - \ln(L) = 0.582207$.
- It can be shown that as $L \to \infty, c_L - \ln L \to 0.57721\ldots$

The **mean** of X is given by:

$$E[X] = \sum_{j=1}^{L} j p_X(j) = \sum_{j=1}^{L} j \frac{1}{c_L j} = \frac{L}{c_L}$$

The second moment and **variance** of X are:

$$E\left[X^2\right] = \sum_{j=1}^{L} j^2 \frac{1}{c_L j} = \frac{1}{c_L} \sum_{j=1}^{L} j = \frac{L(L+1)}{2c_L}$$

and

$$\mathrm{VAR}[X] = \frac{L(L+1)}{2c_L} - \frac{L^2}{c_L^2}$$

- The Zipf random variable has gained prominence with the growth of the Internet where it appears in measurement studies involving Web page sizes, Web access behavior, and Web page interconnectivity.

- These types of random variables had previously been found extensively in studies on the distribution of wealth and, not surprisingly, are now found in Internet video rentals and book sales.

Example 8: Rare Events and Long Tails

The Zipf random variable X has the property that a few outcomes (words) occur frequently but most outcomes occur rarely. Find the probability of words with rank higher than m.

$$P[X > m] = 1 - P[X \leq m] = 1 - \frac{1}{c_L}\sum_{j=1}^{m}\frac{1}{j} = 1 - \frac{c_m}{c_L} \quad \text{for } m \leq L$$

- $P[X > m]$ is the probability of the **tail of the distribution** of X.
- Figure 4 shows the $P[X > m]$ with $L = 100$ which has $E[X] = 100/c_{100} = 19.28$.
- Figure 4 also shows $P[Y > m]$ for a geometric random variable with the same mean, that is, $1/p = 19.28$.
- It can be seen that $P[Y > m]$ for the geometric random variable drops off much more quickly than $P[X > m]$.
- The Zipf distribution is said to have a "long tail" because rare events are more likely to occur than in traditional probability models.

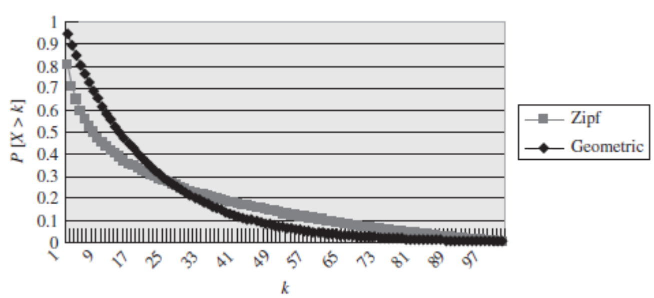


Figure 4: Zipf distribution and its long tail

If we try to let L approach infinity in Eq. (3), c_L grows without bound since the series does not converge. However, if we make the pmf proportional to $(1/k)^\alpha$ then the series converges as long as $\alpha > 1$. We define the ***Zipf or zeta random variable*** *with range* $\{1, 2, 3, \ldots\}$ *to have pmf*:

$$p_Z(k) = \frac{1}{z_\alpha}\frac{1}{k^\alpha} \quad \text{for } k = 1, 2, \ldots$$

where z_α is a normalization constant given by the zeta function which is defined by:

$$z_\alpha = \sum_{j=1}^{\infty} \frac{1}{j^\alpha} = 1 + \frac{1}{2^\alpha} + \frac{1}{3^\alpha} + \ldots \quad \text{for } \alpha > 1$$

The convergence of the above series is discussed in standard calculus books.

The **mean** of Z is given by:

$$E[Z] = \sum_{j=1}^{L} j p_Z(j) = \sum_{j=1}^{L} j \frac{1}{z_\alpha j^\alpha} = \frac{1}{z_\alpha}\sum_{j=1}^{L} \frac{1}{j^{\alpha-1}} = \frac{z_{\alpha-1}}{z_\alpha} \quad \text{for } \alpha > 2$$

where the sum of the sequence $1/j^{\alpha-1}$ converges only if $\alpha - 1 > 1$, that is, $\alpha > 2$. We can similarly show that the second moment (and hence the variance) exists only if $\alpha > 3$.

2 Popular Continuous Random Variables

We are always limited to measurements of some finite precision, so in effect, every random variable found in practice is a discrete random variable. Nevertheless, *there are several compelling reasons for using continuous random variable models.*

1. In general, continuous random variables are easier to handle analytically.

2. The limiting form of many discrete random variables yields continuous random variables.

3. There are a number of "families" of continuous random variables that can be used to model a wide variety of situations by adjusting a few parameters.

2.1 Summary of Continuous Random Variables

Uniform Random Variable

$$S_X = [a, b]$$

$$f_X(x) = \frac{1}{b-a} \quad a \le x \le b$$

$$E[X] = \frac{a+b}{2} \quad \text{VAR}[X] = \frac{(b-a)^2}{12} \quad \Phi_X(\omega) = \frac{e^{j\omega b} - e^{j\omega a}}{j\omega(b-a)}$$

Exponential Random Variable

$$S_X = [0, \infty)$$

$$f_X(x) = \lambda e^{-\lambda x} \quad x \ge 0 \text{ and } \lambda > 0$$

$$E[X] = \frac{1}{\lambda} \quad \text{VAR}[X] = \frac{1}{\lambda^2} \quad \Phi_X(\omega) = \frac{\lambda}{\lambda - j\omega}$$

Remarks: The exponential random variable is the only continuous random variable with the memoryless property.

Gaussian (Normal) Random Variable

$$S_X = (-\infty, +\infty)$$

$$f_X(x) = \frac{e^{-(x-m)^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \quad -\infty < x < +\infty \text{ and } \sigma > 0$$

$$E[X] = m \quad \text{VAR}[X] = \sigma^2 \quad \Phi_X(\omega) = e^{jm\omega - \sigma^2\omega^2/2}$$

Remarks: Under a wide range of conditions X can be used to approximate the sum of a large number of independent random variables.

Gamma Random Variable

$$S_X = (0, +\infty)$$

$$f_X(x) = \frac{\lambda(\lambda x)^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} \quad x > 0 \text{ and } \alpha > 0, \lambda > 0$$

where $\Gamma(z)$ is the gamma function **(Eq. 4.56)**.

$$E[X] = \alpha/\lambda \quad \text{VAR}[X] = \alpha/\lambda^2 \quad \Phi_X(\omega) = \frac{1}{(1 - j\omega/\lambda)^\alpha}$$

Special Cases of Gamma Random Variable:

- m-Erlang Random Variable: $\alpha = m$, a positive integer

$$f_X(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{m-2}}{(m-1)!} \quad x > 0 \quad \Phi_X(\omega) = \left(\frac{1}{1 - j\omega/\lambda}\right)^m$$

  Remarks: An m-Erlang random variable is obtained by adding m independent exponentially distributed random variables with parameter λ.

- Chi-Square Random Variable with k degrees of freedom: $\alpha = k/2, k$ a positive integer, and $\lambda = 1/2$

$$f_X(x) = \frac{x^{(k-2)/2}e^{-x/2}}{2^{k/2}\Gamma(k/2)} \quad x > 0 \quad \Phi_X(\omega) = \left(\frac{1}{1 - 2j\omega}\right)^{k/2}$$

  Remarks: The sum of k mutually independent, squared zero-mean, unit-variance Gaussian random variables is a chi-square random variable with k degrees of freedom.

Laplacian Random Variable

$$S_X = (-\infty, \infty)$$
$$f_X(x) = \frac{\alpha}{2}e^{-\alpha|x|} \quad -\infty < x < +\infty \text{ and } \alpha > 0$$
$$E[X] = 0 \quad \text{VAR}[X] = 2/\alpha^2 \quad \Phi_X(\omega) = \frac{\alpha^2}{\omega^2 + \alpha^2}$$

Rayleigh Random Variable

$$S_X = [0, \infty)$$
$$f_X(x) = \frac{x}{\alpha^2}e^{-x^2/2\alpha^2} \quad x \geq 0 \text{ and } \alpha > 0$$
$$E[X] = \alpha\sqrt{\pi/2} \quad \text{VAR}[X] = (2 - \pi/2)\alpha^2$$

Cauchy Random Variable

$$S_X = (-\infty, +\infty)$$
$$f_X(x) = \frac{\alpha/\pi}{x^2 + \alpha^2} \quad -\infty < x < +\infty \text{ and } \alpha > 0$$

Mean and variance do not exist. $\quad \Phi_X(\omega) = e^{-\alpha|\omega|}$

Pareto Random Variable

$S_X = [x_m, \infty)\, x_m > 0$

$$f_X(x) = \begin{cases} 0 & x < x_m \\ \alpha \frac{x_m^\alpha}{x^{\alpha+1}} & x \geq x_m \end{cases}$$

$$E[X] = \frac{\alpha x_m}{\alpha - 1} \text{ for } \alpha > 1 \quad \text{VAR}[X] = \frac{\alpha x_m^2}{(\alpha-2)(\alpha-1)^2} \text{ for } \alpha > 2$$

Remarks: The Pareto random variable is the most prominent example of random variables with "long tails," and can be viewed as a continuous version of the Zipf discrete random variable.

Beta Random Variable

$$f_X(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & 0 < x < 1 \text{ and } \alpha > 0, \beta > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = \frac{\alpha}{\alpha+\beta} \quad \text{VAR}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Remarks: The beta random variable is useful for modeling a variety of pdf shapes for random variables that range over finite intervals.

2.2 The Uniform Random Variable

The **uniform random variable** arises in situations where all values in an interval of the real line are equally likely to occur. The uniform random variable U in the interval $[a, b]$ has **pdf**:

$$f_U(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & x < a \text{ and } x > b \end{cases}$$

and **cdf**

$$F_U(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

See Figure 5 in Week 4 Notes. The **mean and variance** of U are given by:

$$E[U] = \frac{a+b}{2} \quad \text{and} \quad \text{VAR}[X] = \frac{(b-a)^2}{2}$$

- The uniform random variable appears in many situations that involve equally likely continuous random variables.
- Obviously U can only be defined over intervals that are finite in length.
- The uniform random variable plays a crucial role in generating random variables in computer simulation models. We do this by generating functions of the random variable U.

2.3 The Exponential Random Variable

The exponential random variable arises in the modeling of the time between occurrence of events (e.g., the time between customer demands for call connections), and in the modeling of the lifetime of devices and systems. The **exponential random variable** X with parameter λ has **pdf**

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

and cdf

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases} \tag{4}$$

The cdf and pdf of X are shown in Fig. 7 in Week 3 Notes.

The parameter λ is the **rate** at which events occur, so in Eq. (4) the probability of an event occurring by time x increases at the rate λ increases. Recall from Example 12 in Week 4 Notes that the interarrival times between events in a Poisson process Fig. 3 is an exponential random variable.

The **mean and variance** of X are given by:

$$E[X] = \frac{1}{\lambda} \quad \text{and} \quad \text{VAR}[X] = \frac{1}{\lambda^2} \% tag 4.45$$

In event interarrival situations, λ is in units of events/second and $1/\lambda$ is in units of seconds per event interarrival.

The exponential random variable satisfies the **memoryless** property:

$$P[X > t + h \mid X > t] = P[X > h]$$

- The expression on the left side is the probability of having to wait at least h additional seconds given that one has already been waiting t seconds.
- The expression on the right side is the probability of waiting at least h seconds when one first begins to wait.
- Thus the probability of waiting at least an additional h seconds is the same regardless of how long one has already been waiting!
- The memoryless property of the exponential random variable makes it the cornerstone for the theory of Markov processes.

We now prove the memoryless property:

$$\begin{aligned} P[X > t + h \mid X > t] &= \frac{P[\{X > t + h\} \cap \{X > t\}]}{P[X > t]} \quad \text{for } h > 0 \\ &= \frac{P[X > t + h]}{P[X > t]} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} \\ &= e^{-\lambda h} = P[X > h]. \end{aligned}$$

It can be shown that the exponential random variable is the only continuous random variable that satisfies the memoryless property.

2.4 The Gamma Random Variable

The **Gamma random variable** is a versatile random variable that appears in many applications. For example, it is used to model the time required to service customers in queueing systems, the lifetime of devices and systems in reliability studies, and the defect clustering behavior in VLSI chips.

The **pdf** of the gamma random variable has two parameters, $\alpha > 0$ and $\lambda > 0$, and is given by

$$f_X(x) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad 0 < x < \infty$$

where $\Gamma(z)$ is the **gamma function**, which is defined by the integral

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx \quad z > 0$$

The gamma function has the following properties:

$$\begin{aligned}
&\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi} && \\
&\Gamma(z+1) = z\Gamma(z) && \text{for } z > 0, \text{ and} \\
&\Gamma(m+1) = m! && \text{for } m \text{ a nonnegative integer.}
\end{aligned}$$

The versatility of the gamma random variable is due to the richness of the gamma function $\Gamma(z)$.

- The pdf of the gamma random variable can assume a variety of shapes as shown in Fig. 5.
- Many random variables are special cases of gamma random variable.
  - The *exponential random variable* is obtained by letting $\alpha = 1$.
  - By letting $\lambda = 1/2$ and $\alpha = k/2$, where k is a positive integer, we obtain the *chi-square random variable*, which appears in certain statistical problems.
  - The $\boldsymbol{m}$-Erlang random variable is obtained when $\alpha = m$, a positive integer. The m-Erlang random variable is used in the system reliability models and in queueing systems models.

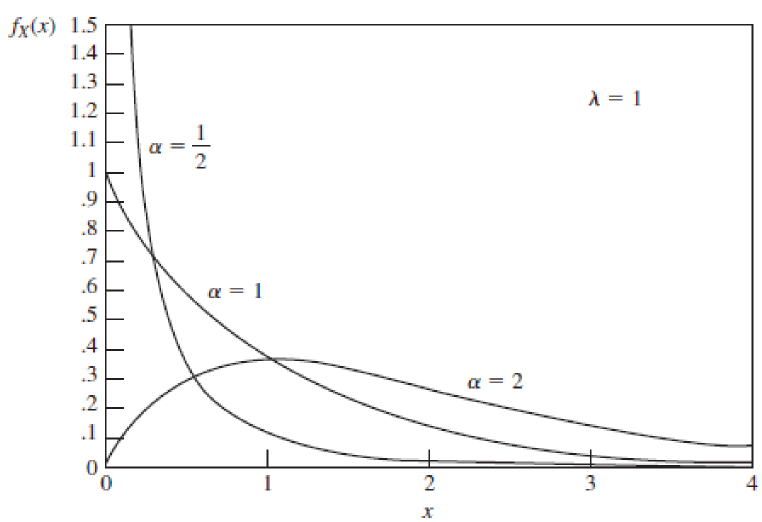


Figure 5: pdf of gamma random variable

Example 9: Integration of Gamma pdf

Show that the pdf of a gamma random variable integrates to one.

The integral of the pdf is

$$\begin{aligned}\int_0^\infty f_X(x)dx &= \int_0^\infty \frac{\lambda(\lambda x)^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha-1}e^{-\lambda x}dx\end{aligned}$$

Let $y = \lambda x$, then $dx = dy/\lambda$ and the integral becomes

$$\frac{\lambda^\alpha}{\Gamma(\alpha)\lambda^\alpha}\int_0^\infty y^{\alpha-1}e^{-y}dy = 1$$

where we used the fact that the integral equals $\Gamma(\alpha)$.

- In general, the cdf of the gamma random variable does not have a closed-form expression.
- The special case of the m-Erlang random variable does have a closed-form expression for the cdf by using its close interrelation with the exponential and Poisson random variables. The cdf can also be obtained by integration of the pdf.

Suppose that we observe the time S_m that elapses until the occurrence of the m th event. The times $X_1, X_2, \ldots, X_m$ between events are exponential random variables, so we must have

$$S_m = X_1 + X_2 + \cdots + X_m$$

We will show that S_m is an m**-Erlang random variable**. To find the cdf of S_m, let $N(t)$ be the Poisson random variable for the number of events in t seconds. Note that the m th event occurs before time t-that is, $S_m \leq t$-if and only if m or more events occur in t seconds, namely $N(t) \geq m$. The reasoning goes as follows. If the mth event has occurred before time t, then it follows that m or more events will occur in time t. On the other hand, if m or more events occur in time t, then it follows that the mth event occurred by time t. Thus

$$F_{S_m}(t) = P\left[S_m \leq t\right] = P[N(t) \geq m]$$

$$= 1 - \sum_{k=0}^{m-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{5}$$

where we have used the result of Example 12 in Week 4 Notes. If we take the derivative of the above cdf, we finally obtain the pdf of the m-Erlang random variable. Thus we have shown that S_m is an m-Erlang random variable.

Example 10: System Lifetime

A factory has two spares of a critical system component that has an average lifetime of $1/\lambda = 1$ month. Find the probability that the three components (the operating one and the two spares) will last more than 6 months. Assume the component lifetimes are exponential random variables.

The remaining lifetime of the component in service is an exponential random variable with rate λ by the memoryless property. Thus, the total lifetime X of the three components is the sum of three exponential random variables with parameter $\lambda = 1$. Thus X has a 3-Erlang distribution with $\lambda = 1$. From Eq. (5) the probability that X is greater than 6 is

$$\begin{aligned} P[X > 6] &= 1 - P[X \leq 6] \\ &= \sum_{k=0}^{2} \frac{6^k}{k!} e^{-6} = .06197. \end{aligned}$$

2.5 The Beta Random Variable

The **Beta random variable** X assumes values over a finite interval and provides a wide variety of pdf shapes that can cover a finite interval. It has **pdf**:

$$f_X(x) = cx^{a-1}(1-x)^{b-1} \quad \text{for } 0 < x < 1$$

where the normalization constant is the reciprocal of the beta function

$$\frac{1}{c} = B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx$$

and where the beta function is related to the gamma function by the following expression:

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

When $a = b = 1$, we have the uniform random variable.

- Other choices of a and b give pdfs over finite intervals that can differ markedly from the uniform.
  - If $a = b > 1$, then the pdf is symmetric about $x = 1/2$ and is concentrated about $x = 1/2$ as well.
  - When $a = b < 1$, then the pdf is symmetric but the density is concentrated at the edges of the interval.
  - When $a < b$ (or $a > b$) the pdf is skewed to the right (or left).

The mean and variance are given by:

$$E[X] = \frac{a}{a+b} \quad \text{and} \quad \text{VAR}[X] = \frac{ab}{(a+b)^2(a+b+1)}$$

- The versatility of the pdf of the beta random variable makes it useful to model a variety of behaviors for random variables that range over finite intervals.
- A classic application of the Beta random variable is to model the pdf of p in a Bernoulli trial experiment, that is, the probability of success p is itself be a random variable.

2.6 The Cauchy Random Variable

The **Cauchy random variable** X assumes values over the entire real line and has **pdf** with slowly decaying tails:

$$f_X(x) = \frac{1/\pi}{1+x^2}$$

It is easy to verify that this pdf integrates to 1. However, X *does not have any moments since the associated integrals do not converge.* The Cauchy random variable arises as the tangent of a uniform random variable in the unit interval.

2.7 The Pareto Random Variable

The **Pareto random variable** arises in the study of the distribution of wealth where Vilfredo Pareto used it to model the tendency for a small portion of the population to own a large portion of the wealth.

- The Pareto distribution has been found to capture the behavior of many quantities of interest in the study of Internet behavior, e.g., sizes of files, packet delays, audio and video title preferences, session times in peer-to-peer networks, etc.

- The Pareto random variable can be viewed as a continuous version of the Zipf discrete random variable.

The Pareto random variable X takes on values in the range $x > x_m$, where x_m is a positive real number. X has **complementary cdf** with shape parameter $\alpha > 0$ given by:

$$P[X > x] = \begin{cases} 1 & x < x_m \\ \frac{x_m^\alpha}{x^\alpha} & x \geq x_m \end{cases}$$

The tail of X decays algebraically with x which is rather slower in comparison to the exponential and Gaussian random variables. The Pareto random variable is the most prominent example of random variables with "**long tails**."

The **cdf and pd**f of X are:

$$F_X(x) = \begin{cases} 0 & x < x_m \\ 1 - \frac{x_m^\alpha}{x^\alpha} & x \geq x_m. \end{cases}$$

Because of its long tail, the cdf of X approaches 1 rather slowly as x increases.

$$f_X(x) = \begin{cases} 0 & x < x_m \\ \alpha \frac{x_m^\alpha}{x^{\alpha+1}} & x \geq x_m \end{cases}$$

Example 11: Mean and Variance of Pareto Random Variable

Find the **mean** and variance of the Pareto random variable.

$$E[X] = \int_{x_m}^{\infty} t\alpha \frac{x_m^\alpha}{t^{\alpha+1}} dt = \int_{x_m}^{\infty} \alpha \frac{x_m^\alpha}{t^\alpha} dt = \frac{\alpha}{\alpha - 1} \frac{x_m^\alpha}{x_m^{\alpha-1}} = \frac{\alpha x_m}{\alpha - 1} \quad \text{for } \alpha > 1$$

where the integral is defined for $\alpha > 1$, and

$$E\left[X^2\right] = \int_{x_m}^{\infty} t^2 \alpha \frac{x_m^\alpha}{t^{\alpha+1}} dt = \int_{x_m}^{\infty} \alpha \frac{x_m^\alpha}{t^{\alpha-1}} dt = \frac{\alpha}{\alpha - 2} \frac{x_m^\alpha}{x_m^{\alpha-2}} = \frac{\alpha x_m^2}{\alpha - 2} \quad \text{for } \alpha > 2$$

where the second moment is defined for $\alpha > 2$.

The **variance** of X is then:

$$\mathrm{VAR}[X] = \frac{\alpha x_m^2}{\alpha - 2} - \left(\frac{\alpha x_m^2}{\alpha - 1}\right)^2 = \frac{\alpha x_m^2}{(\alpha - 2)(\alpha - 1)^2} \quad \text{for } \alpha > 2$$