STA257: Probability and Statistics 1

Instructor: Katherine Daignault

Department of Statistical Sciences
University of Toronto

Week 6 and 7

Outline

Outline

Calculus Refresher: Definite Integrals

- Again you will need to be very comfortable with taking definite integrals of functions.
- Definition: $\int_a^b f(x)dx = F(x)|_a^b = F(b) - F(a)$
- Some useful properties:
  - Reversing limits: $\int_a^b f(x)dx = -\int_b^a f(x)dx$
  - Additivity: $\int_a^b [f(x) + g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx$
- Some useful results (written as indefinite integrals):
  - $\int adx = ax + C$
  - $\int x^a dx = \frac{x^{a+1}}{a+1} + C$
  - $\int (1/x)dx = ln(x) + C$
  - $\int e^{ax} dx = (1/a)e^{ax} + C$
  - $\int a^x dx = a^x / ln(a) + C$
  - $\int ln(x)dx = xln(x) - x + C$

Motivating Example

- A gambler is playing a dice rolling game.
- When he rolls a fair die, the following can happen:
  - if he rolls a 5 or 6, he gains \$15
  - if he rolls anything else, he loses \$9
- The gambler wants to know, if he repeats this game a large number of times, how much money he might make on average.
- We can think of this as trying to find his average or expected gains from playing this game.
- How might we go about finding this?

Motivating Example (cont.)

- We may think of this in terms of a random variable X, representing the amount of money earned/gained
  - X can take values $\{15, -9\}$
- We also know from counting what the long-run chances are of getting certain dice rolls:
  - The probability of rolling either a 5 or a 6 is $^2/_6$
  - The probability of rolling anything that isn't a 5 or a 6 is $^4/_6$
- Therefore, we can expect to gain \$15 with probability $^2/_6$ and we can expect to lose \$9 with probability $^4/_6$
- Putting these together, we find that our expected winnings are

$$15 \times \frac{2}{6} + (-9) \times \frac{4}{6} = -1$$

Expected Values

- What we have just found is the **expected value** of random variable X
- We interpret this as the value one would expect to see on average from a large number of repeated experiments
- This corresponds to the average or the mean of the probability distribution of X
- Therefore it is a value that has high probability of occurring, i.e. often the maximum of the PMF/PDF

Expected Value - Discrete Case

Expected Value for Discrete Random Variables

If X is a discrete random variable with frequency function/PMF $p(x)$, the expected value of X, denoted $E(X)$, is

$$E(X) = \sum_i x_i p(x_i)$$

provided that $\sum_i \left| x_i p(x_i) \right| < \infty$. If the sum diverges, the expectation is undefined.

- Expected values can be calculated/derived from any discrete distribution, where $E(X)$ is defined.

Example 1: Expectation of Geometric Random Variable

Suppose items produced in a plant are independently defective with probability p. Items are inspected one by one until a defective is found. On average, how many items must be inspected?

- This is clearly a Geometric random variable, as we are counting the number of inspections until the first defective item is found.
- My PMF is $p(k) = P(X = k) = q^{k-1}p$, where $q = 1 - p$
- I can find my expectation by using the definition:

$$E(X) = \sum_{k=1}^{\infty} kp(k) = \sum_{k=1}^{\infty} kpq^{k-1} = p \sum_{k=1}^{\infty} kq^{k-1}$$

- Where do we go from here?

Example 1: Expectation of Geometric RV (cont.)

- We can use a trick where we can express $kq^{k-1} = \frac{d}{dq}(q^k)$
- Then

$$E(X) = p \sum_{k=1}^{\infty} kq^{k-1} = p \sum_{k=1}^{\infty} \frac{d}{dq}(q^k)$$

- If we interchange the summation and differentiation operations, we can compute the sum:

$$\begin{aligned} E(X) &= p \sum_{k=1}^{\infty} \frac{d}{dq}(q^k) = p \frac{d}{dq} \left(\sum_{k=1}^{\infty} q^k \right) \\ &= p \frac{d}{dq} \left(\frac{q}{1-q} \right) \quad \text{by Geometric series} \\ &= \frac{p}{(1-q)^2} = \frac{1}{p} \end{aligned}$$

Example 1: Expectation of Geometric RV (cont.)

- So we find that the expected value of a Geometric random variable is
$$E(X) = \frac{1}{p}$$
- This means that, if 10% of items are defective, the average number of items that must be inspected in order to find one defective is $\frac{1}{0.1} = 10$
- It can be shown that interchanging the summation and differentiation is justified, but we won't go into it here.

Example 2: Expectation of Poisson RV

- We can similarly derive the form of the expected value of a Poisson random variable.
- First start from the definition of expected value:

$$E(X) = \sum_{k=0}^{\infty} k \times p(k) = \sum_{k=0}^{\infty} k \times \frac{\lambda^k e^{-\lambda}}{k!}$$

- If we cancel some terms, we get

$$E(X) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

where we need to adjust the summation limits because we cannot have a negative factorial

Example 2: Expectation of Poisson RV (cont.)

- Here, instead of a trick, we need to use the result of an infinite series:

$$\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{-\lambda}$$

- In order to use this, we need to adjust the limits of the sum on more time

$$E(X) = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$$

- Finally we get

$$E(X) = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

so the rate parameter λ is also the average/expected number of events.

Exercise - Give it a try!

A roulette wheel has numbers 1 to 36, as well as 0 and 00. If you bet \$1 that an odd number comes up, you win or lose \$1 depending on if that event occurs. If X denotes your net winnings, where $X = 1$ with probability $^{18}/_{38}$ and $X = -1$ with probability $^{20}/_{38}$. Find the expected winnings.

Expected Value - Continuous Case

Expected Value for Continuous Random Variables

If X is a continuous random variable with density function $f(x)$, then

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

provided that $\int |x| f(x)dx < \infty$. If the integral diverges, the expectation is undefined.

- Again, this represents the mean/average of the density of X
- It corresponds to the centre of mass of the density

Example: Expectation of Gamma Random Variable

- Recall that if X is a Gamma with parameters α and λ the density is
$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$$
- We can find the expected value of for the Gamma distribution using the definition:
$$E(X) = \int_0^\infty x \times \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha} e^{-\lambda x} dx$$
- Can we find the expected value without integrating?

Example: Expectation of Gamma RV (cont.)

- Yes we can!
- We just have to realize that if we could make

$$\frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha} e^{-\lambda x}$$

look like

$$\frac{\lambda^{\alpha^*}}{\Gamma(\alpha^*)} x^{\alpha^*-1} e^{-\lambda x}$$

then we would know that the integral

$$\int_0^\infty \frac{\lambda^{\alpha^*}}{\Gamma(\alpha^*)} x^{\alpha^*-1} e^{-\lambda x} dx = 1$$

because we are integrating the entire density function

Example: Expectation of Gamma RV (cont.)

- First notice that we can rewrite the superscript of x:

$$E(X) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{(\alpha+1)-1} e^{-\lambda x} dx$$

- Now consider what constants could be included in the integral which would allow us to create a new Gamma density

$$E(X) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \int_0^\infty \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^{(\alpha+1)-1} e^{-\lambda x} dx$$

- But the integral is now over a complete Gamma density, so it equals 1, giving

$$E(X) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} = \frac{\alpha}{\lambda}$$

Expectations of Functions of Random Variables

Expectations of $g(X)$

Suppose that $Y = g(X)$.

1. If X is discrete with PMF $p(x)$, then

$$E(Y) = \sum_x g(x)p(x)$$

   provided that $\sum_x \left|g(x)\right| p(x) < \infty$.

2. If X is continuous with PDF $f(x)$, then

$$E(Y) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

   provided that $\int \left|g(x)\right| f(x)dx < \infty$

Proof of Expectation of $g(X)$

Proof:

$\square$

Example: Rainfall in Midwest U.S.

Suppose we are interested in the square of total summer rainfall. Recall that summer rainfall totals $X \sim Gamma(2, 0.5)$ from Week 4. Find the expected value of square of total summer rainfall.

- We are interested in $Y = X^2$.
- We could of course use Week 5 content and find the distribution function of Y.
- However it is simpler when just asked for the expected value to use the previous result.
- Here, $g(X) = X^2$, so we use $E[g(X)] = \int g(x)f(x)dx$:

$$E[g(X)] = \int_0^\infty x^2 \times \frac{1}{4} x e^{-x/2} dx = \int_0^\infty \frac{1}{4} x^3 e^{-x/2} dx$$

Example: Rainfall in Midwest U.S. (cont.)

- To solve this integral, we need to use our handy trick of recognizing that we can turn this into a different Gamma distribution by multiplying/dividing in some specific constants.
  - Then the integral will just equal one, and leave us with some constants.
- We get that $x^3 e^{-x/2} = x^{4-1} e^{-x/2}$ which would mean we were dealing with a $Gamma(4, 0.5)$
- Therefore we multiply and divide by new constants involving this new $\alpha = 4$

$$E[X^2] = \frac{1}{4} \frac{\Gamma(4)}{0.5^4} \underbrace{\int_0^\infty \frac{0.5^4}{\Gamma(4)} x^{4-1} e^{-x/2} dx}_{=1} = \frac{3!(16)}{4} = 24$$

Exercise - Give it a try!

A roulette wheel has numbers 1 to 36, as well as 0 and 00. If you bet \$1 that an odd number comes up, you win or lose \$1 depending on if that event occurs. If X denotes your net winnings, where $X = 1$ with probability $^{18}/_{38}$ and $X = -1$ with probability $^{20}/_{38}$. Find the expected squared winnings.

Some Helpful Properties of Expectations

We list here some useful properties involving expectations, the proofs of which are left as exercises.

1. $E(aX + b) = aE(X) + b$ when both a and b are constants.
2. Generally, $E[g(X)] \neq g(E[X])$ for all g, with some exceptions to be discussed later.
3. If both $g : \mathbb{R} \Rightarrow \mathbb{R}$ and $h : \mathbb{R} \Rightarrow \mathbb{R}$, then $E[g(X) + h(X)] = E[g(X)] + E[h(X)]$

Outline

Variance of a Random Variable

- We saw that the expected value represents the centre of mass of the distribution (sometimes called the location parameter)
  - the median is another choice to measure the centre of the density
- In addition to where a distribution is centred, we want to talk about how spread out or variable the density is
- We can measure the variability through the **variance** of the random variable.
- Alternatively, we can talk about the standard deviation, which is a function of the variance
  - standard deviation measures how dispersed the density is around the mean.

Variance of a Random Variable

Definition of the Variance of a Random Variable

If X is a random variable with expected value $E(X)$, the variance of X is

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

provided that the expectation exists. The standard deviation of X is the square root of the variance,

$$SD(X) = \sqrt{Var(X)}$$

- The variance measures the average squared distance of the random variable from its expected value.
- We often denote the variance by σ^2 and the standard deviation by σ

Variance of a Random Variable

- Just like expectations, calculation of the variance of random variables differs depending on whether the random variable is discrete or continuous
- When the RV is discrete, and has PMF $p(x)$ and expected value $\mu = E(X)$, then

$$Var(X) = \sum_i (x_i - \mu)^2 p(x)$$

- When the RV is continuous with PDF $f(x)$ and expected value $\mu = E(X)$, then

$$Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

where the bounds of integration will depend on which density you have.

Example: Variance of Bernoulli Random Variable

- Recall that $X \sim Ber(p)$ where p is the probability of success and
$$p(x) = p^x(1-p)^{1-x},\ x \in \{0, 1\}$$
- We need to know $E(X)$ to continue:
$$E(X) = \sum_i x_i p(x_i) = 0 \times (1-p) + 1 \times p = p$$
- From the definition of variance, we can have:
$$\begin{aligned} Var(X) &= \sum_i (x_i - E(X))^2 p(x) \\ &= (0-p)^2(1-p) + (1-p)^2 p \\ &= p^2 - p^3 + p - 2p^2 + p^3 \\ &= p(1-p) \end{aligned}$$

Variances in Another Way

- The same procedure applies for finding the variance of a continuous random variable.
- However it can often be tedious to compute all squared differences
- We have an alternative way of expressing variances:

Alternative Variance Form

The variance of X, if it exists, may also be calculated as follows:

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

Proof of Alternative Form of Variance

Proof

$\square$

Example: Variance of Uniform RV

- Suppose $X \sim Unif(0,1)$, so $f(x) = \frac{1}{b-a} = 1$ for $x \in [0,1]$
- To find the variance of X, we need both $E(X)$ and $E(X^2)$.
- Let's start with $E(X)$:

$$E(X) = \int_0^1 x \times 1dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$$

- Now, we can find $E(X^2)$ in the same way as $E(X)$, just replace x with x^2 in the integral:

$$E(X^2) = \int_0^1 \color{red}{x^2} \times 1dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}$$

- Thus we have

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

A Useful Property of Variance

A Useful Property of Variance

If $Var(X)$ exists and $Y = a + bX$, then $Var(Y) = b^2 Var(X)$

Proof:

Since $E(Y) = a + bE(Y)$ then

$$\begin{aligned} E[(Y - E(Y))^2] &= E[(a + bX - a - bE(X))^2] \\ &= E[b^2(X - E(X))^2] \\ &= b^2 E[(X - E(X))^2] \\ &= b^2 Var(X) \end{aligned}$$

$\square$

Example: Normal Distributions

- This result makes intuitive sense because if I shift the distribution by a, I don't change the spread, but if I rescale the distribution (e.g. change the units), then I change the spread.
- It is most easily illustrated using the Normal distribution.
- Suppose I have $Z \sim N(0,1)$, a standard Normal RV.
  - mean $\mu = E(Z) = 0$
  - variance $\sigma^2 = Var(Z) = 1$
- Suppose we transform Z by $X = 4 + 2Z$
- We can find the mean and variance of X, and thus the distribution, by
  - $E(X) = E(4 + 2Z) = 4 + 2E(Z) = 4$
  - $Var(X) = Var(4 + 2Z) = 2^2 Var(Z) = 4(1) = 4$
  - so $X \sim N(4,4)$

Outline

Inequalities with Expectation and Variance

- In the previous section, we discussed ways to determine the centre of mass of a density/distribution function, as well as the dispersion of the density from the centre
- We can now use the expectation and variance to talk about the probability that
  - X could be take on a very large value
  - X could be very far away from the mean/expected value of its distribution
- We present two important inequalities that address these situations.

Markov's Inequality

Markov's Inequality

If X is a random variable with $P(X \geq 0) = 1$ (i.e. defined on non-negative values) and for which $E(X)$ exists, then

$$P(X \geq t) \leq \frac{E(X)}{t}$$

Proof:

$\square$

Intuition behind Markov's Inequality

- Suppose random variable X has expected value 3, and a distribution defined on only non-negative values.
- Let's look at how Markov's Inequality works for a few values of t.
- For $t = 10$, we have that $P(X \geq 10) \leq 3/10 = 30\%$
  - This says there is a 30% chance that X could be greater than 10 (which is already quite far from $E(X)$)
- For $t = 30$, we have that $P(X \geq 30) \leq 3/30 = 10\%$
  - Now there is a 10% chance that X could be greater than 30 (which is really far from $E(X)$)
- For $t = 3$, we have that $P(X \geq 3) \leq 3/3 = 100\%$
  - This says there is a 100% chance of observing an X that is larger than $E(X)$
  - Which makes sense because $E(X)$ is the centre of the density.

Chebyshev's Inequality

Chebyshev's Inequality

Let X be a random variable with mean μ and variance σ^2. Then, for any $t > 0$,

$$P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$$

Proof:

$\square$

Alternative Form of Chebyshev's Inequality

Alternative Chebyshev's Inequality

Let X be a random variable with mean μ and variance σ^2. Then for any $k > 0$,

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

- This version is sometimes easier to understand because it talks about the chances that X is within a certain number of standard deviations from the mean.
- All we have done here is to set $t = k\sigma$

Intuition behind Chebyshev's Inequality

- Using the alternative form of the inequality, we can see what this is telling us by trying different values of k.
- When $k = 2$, we have $1/2^2 = 25\%$ of values must be <u>at most</u> 2 standard deviations away from the mean
  - or alternatively, at least 75% must be <u>within</u> 2σ of the mean
- When $k = 3$, we have $1/3^2 = 11\%$ of values must be <u>at most</u> 3 standard deviations away from the mean
  - or alternatively, at least 89% must be <u>within</u> 3σ of the mean
- So Chebyshev's inequality tells us roughly what percentage of our distribution lies $k\sigma$ above and below the mean, i.e. percentage in the tails.

Example: Application of Chebyshev's Inequality

The number of customers per day at a sales counter X has been observed for a long period of time and has been found to have mean 20 and standard deviation 2. The probability distribution of X is not known. What can be said about the probability that, tomorrow, X will be greater than 16 but less than 24?

- Since we are dealing with $X =$ number of customers, X is therefore non-negative, so we are able to apply Chebyshev's here.
- We want $P(16 < X < 24)$, but to use the inequality, we need to rewrite these in terms of $k\sigma$ away from μ
  - $\mu + k\sigma = 20 + k(2) = 24 \Rightarrow k = 2$
  - Similarly, $\mu - k\sigma = 20 - k(2) = 16 \Rightarrow k = 2$
  - So we get

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = P(|X - \mu| < 2\sigma) \geq 1 - \frac{1}{2^2} = \frac{3}{4}$$

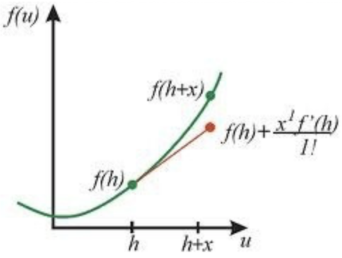
Outline

What are Moments?

Moments can be seen as analogous to terms of a Taylor series expansion:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2!} + \cdots$$

- Often a Taylor series can be used to obtain an approximation of a function $f(x)$
- The more derivative terms included, the better the approximation will be
- Moments of a RV can be used to approximate the density function



Moments of a Random Variable

- There are two types of moments that we can use to find the distribution of X.
- We say the r^{th} **moment** of X is $E(X^r)$
  - This implies that $E(X)$ is the first moment of X
  - We also used $E(X^2)$, the second moment of X, to find the variance of X.
- We also have the r^{th} **central moment** of X, defined as $E\left\{[X - E(X)]^r\right\}$
  - We call this a central moment because, by subtracting the mean, we are effectively centre the distribution of X at the mean.
  - Note that the variance $Var(X) = E[(X - \mu)^2]$ is the second central moment

Moment Generating Functions

- We have seen that it is possible to find moments and central moments by deriving each one directly using the distribution function.
- However, if we wanted to find all of the moments, this will become quite tiresome.
- It turns out that we can often find a function, called the **moment generating function (MGF)**, which allows us to find all the moments we want
- However, the MGF can only be found if the expectation for X is defined.

Moment Generating Functions

Definition of Moment Generating Function

The MGF of a random variable X is $M(t) = E\left(e^{tX}\right)$ if the expectation is defined. In the discrete case,

$$M(t) = E\left(e^{tX}\right) = \sum_{x} e^{tx} p(x)$$

and in the continuous case,

$$M(t) = E\left(e^{tX}\right) = \int_{\infty}^{\infty} e^{tx} f(x) dx$$

How Does the MGF Generate Moments?

- Let's consider the continuous case: $M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$
- Suppose that we can take the derivative of M and that we can switch the order of integration and differentiation:

$$M'(t) = \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f(x) dx = \int_{-\infty}^{\infty} x e^{tx} f(x) dx$$

- Now if we set $t = 0$, we have

$$M'(0) = \int_{-\infty}^{\infty} x e^{x(0)} f(x) dx = \int_{-\infty}^{\infty} x f(x) dx$$

- But this just gives us that $M'(0) = E(X)$.
- If we differentiate $M(t)$ r times, we get that $M^{(r)}(0) = E(X^r)$

Properties of MGFs

- Again, this only works if we can actually take the expectation of X to begin with,
- The results of the previous slide give us some helpful properties of MGFs:
  1. If the MGF exists for t in an open interval containing 0, it uniquely determines the probability distribution.
  2. If the MGF exists in an open interval containing 0, then $M^{(r)}(0) = E(X^r)$
- The appeal of using the MGF to find moments, rather than directly summing/integrating the PMF/PDF, is that differentiation can often be easier than working with series/integration

Example: MGF of a Poisson Random Variable

- We can find this MGF by working from the definition:

$$M(t) = \sum_x e^{tx} p(x) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k e^{-\lambda}}{k!}$$

- Group the terms that can be grouped with k:

$$M(t) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=0}^{\infty} \frac{\left(\lambda e^t\right)^k}{k!} e^{-\lambda}$$

- Now use the same trick as when finding the $E(X)$ for a Poisson:

$$M(t) = \sum_{k=0}^{\infty} \frac{\left(\lambda e^t\right)^k}{k!} e^{-\lambda} = e^{\lambda e^t} e^{-\lambda} = e^{\lambda(e^t - 1)}$$

Example: MGF of a Poisson RV (cont.)

- So the MGF for a Poisson is $M(t) = e^{\lambda(e^t - 1)}$
- To find the mean or first moment of the Poisson, I take the first derivative, evaluated at $t = 0$:

$$M'(t) = \frac{d}{dt} e^{\lambda(e^t - 1)} = \lambda e^t e^{\lambda(e^t - 1)} \Rightarrow M'(0) = \lambda$$

- To find the variance, or second central moment, I need the second moment:

$$M''(t) = \lambda e^t e^{\lambda(e^t - 1)} + \lambda^2 e^{2t} e^{\lambda(e^t - 1)} \Rightarrow M''(0) = \lambda^2 + \lambda$$

- So $Var(X) = M''(0) - [M'(0)]^2 = \lambda^2 + \lambda - \lambda = \lambda$

Example: MGF of a Gamma Random Variable

- Again, start with the definition of MGFs:

$$M(t) = \int_0^\infty e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{x(t-\lambda)} dx$$

- This integral only converges when $t < \lambda$.
- We can solve it with the same trick as before, by relating it to a $Gamma(\alpha, \lambda - t)$

$$M(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{x(t-\lambda)} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{\Gamma(\alpha)}{(\lambda - t)^\alpha} \right)$$

- So the MGF of a Gamma variable is $M(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha$ when $t < \lambda$

Exercise - Give it a try!

Use the MGF of the Gamma to determine the mean and variance of a Gamma random variable.

Property of MGFs

MGF of $Y = a + bX$
If X has the MGF $M_X(t)$ and $Y = a + bX$, then Y has the MGF

$$M_Y(t) = e^{at} M_X(bt).$$

Proof:

$\square$

Exercise - Give it a try!

What is the MGF for $Y = 4X$ when $X \sim \mathit{Gamma}(\alpha, \lambda)$?