

STAT 501 Fall 2022 Course Notes

Chapter 2: Transformations and Expectations

When we are able to model phenomenon in terms of a random variable with a given CDF, we often want to also explore the behavior of functions of that random variable. This may include finding distributions of these functions and/or determining their average behavior.

2.1 Distributions of Functions of Random Variables

For a random variable X with known distribution function, $F_X(x)$ we are often interested in a function of that random variable $Y = g(X)$, which is also a random variable. Because this new random variable Y is a function of X , its distribution depends on the functions F_X and g . Formally, $y = g(x)$ defines a mapping from \mathcal{X} (the original sample space of X) to a new sample space \mathcal{Y} associated with Y . In general, for a random variable X with *pmf* or *pdf* $f_X(x)$, the supports (sample spaces) of X and $Y = g(X)$ are:

Example (Discrete) Consider a game in which a fair 6-sided die is thrown. The player pays \$5 to play, and earns \$2 times the number of dots shown on the die roll.

- Let $X =$
- Let $Y =$
- What are \mathcal{X}, \mathcal{Y} ?

Example (Continuous) Let X be a continuous RV with support, $\mathcal{X} = \{x : x \in [0, 1]\}$.

- Let $Y = X^2$. What is \mathcal{Y} ?
- Let $Y = -\ln(X)$. What is \mathcal{Y} ?

An inverse mapping g , denoted by g^{-1} , is a mapping from subsets of \mathcal{Y} to subsets of \mathcal{X} , and is defined by $g^{-1}(A) = \{x \in \mathcal{X} : g(x) \in A\}$,

Note: Ideally, g and g^{-1} are 1-1 functions where $y = g(x) \iff g^{-1}(y) = x$. However, this need not be the case.

Our goal is to find the *distribution* of Y . For the random variable $Y = g(X)$, we can write for any set $A \subset \mathcal{Y}$,

$$P_Y(Y \in A) =$$

If X is a discrete random variable, then $Y = g(X)$ is also discrete, because both \mathcal{X} and \mathcal{Y} are countable sets. From above, it follows that in the discrete cases, the *pmf* for Y is:

Example (Discrete) Let X be a discrete RV with the following *pmf*:

x	-2	-1	0	1	2
$f_X(x) = P(X = x)$	0.1	0.2	0.4	0.2	0.1

Let $Y = |X|$. What is the *pmf* of Y ?

Example (Discrete) Let X be a discrete RV that follows the Binomial distribution:

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \dots, n; p \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

where n is a positive integer. Let $Y = X/n$. What is the *pmf* of Y ?

If X and Y are continuous random variables, then in some cases it is possible to find formulas for the CDF and *pdf* of Y in terms of the CDF and *pdf* of X and the function g .

Examples (Continuous) Let X be a continuous RV with the following pdf: $f_X(x) = \frac{1}{10}I_{[0,10]}(x)$. Find the CDFs of the following RVs:

- $Y = X^2$

- $W = -\ln(X)$

THM 2.1.3 (CDF Method of Transformations for Monotone Functions) Let X be a continuous RV with CDF $F_X(x)$, let $Y = g(X)$, and let \mathcal{X} and \mathcal{Y} , be the supports of X and Y , respectively.

1. If g is an increasing function on \mathcal{X} , for $y \in \mathcal{Y}$, $F_Y(y) =$
2. If g is a decreasing function on \mathcal{X} , for $y \in \mathcal{Y}$, $F_Y(y) =$

Note: $F_Y(y)$ needs to be defined for all $y \in \mathbb{R}$ not just $y \in \mathcal{Y}$.

BUT what if we really want $f_Y(y)$? Assume X is a continuous random variable with CDF $F_X(x)$ and let $Y = g(X)$.

- **Increasing case:** If g is an increasing function on \mathcal{X} , $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.

- **Decreasing case:** If g is an decreasing function on \mathcal{X} , $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.

THM 2.1.5 (Jacobian Method of Transformations for Monotone Functions) Let X be a continuous RV with pdf $f_X(x)$ and let $Y = g(X)$, where g is a *monotone function*. Let \mathcal{X} and \mathcal{Y} be the supports of X and Y , respectively. Suppose that $f_X(x)$ is continuous on \mathcal{X} and that $g^{-1}(y)$ has a *continuous derivative* on \mathcal{Y} . Then the pdf of Y is given by:

Examples Let X be a random variable with pdf $f_X(x) = \frac{3}{2}x^2I_{(-1,1)}(x)$. Find the *pdf* for each of the functions of X below.

- $Y = 3X$

- $Y = X^2$

THM 2.1.8 (Jacobian Method of Transformations for Non-Monotone Functions):

Let X be a continuous RV with *pdf* $f_X(x)$ and support \mathcal{X} . Consider the transformation $Y = g(X)$, and suppose there exists a partition (A_0, A_1, \dots, A_k) of \mathcal{X} such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i . Further suppose there exist functions $g_1(X), g_2(X), \dots, g_k(X)$ defined on A_1, A_2, \dots, A_k , respectively, satisfying the following:

- i. $g(X) = g_i(X)$ for $x \in A_i$
- ii. $g_i(X)$ is monotone on each A_i
- iii. the set $\mathcal{Y} = \{y : y = g_i(x) \text{ for } x \in A_i\}$ is the same for each $i = 1, 2, \dots, k$, and
- iv. $g_i^{-1}(y)$ has a continuous derivative on \mathcal{Y} for each $i = 1, 2, \dots, k$.

Then,

Note: Each $g_i(x)$ is a 1–1 function from A_i onto \mathcal{Y} , and $g^{-1}(y)$ is a 1–1 function from \mathcal{Y} onto A_i . A_0 is a technical device for dealing with endpoints of intervals, etc., and we can ignore it since $P(X \in A_0) = 0$.

Example Let X have the standard normal distribution, $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}I_{(-\infty, \infty)}(x)$. Show that $Y = X^2 \sim \chi_1^2$ (see tables in back of the book for distributions).

Example Let X be a RV with pdf $f_X(x) = \frac{1}{4}I_{(-1,3)}(x)$. Find the pdf of $Y = X^2$.

THM 2.1.10 (Probability Integral Transformation) Let X have continuous CDF, $F_X(x)$ and define the random variable Y as $Y = F_X(x)$. Then Y is uniformly distributed on $(0,1)$. That is,

$$P(Y \leq y) = \begin{cases} 0 & y \leq 0 \\ y & 0 < y < 1 \\ 1 & y \geq 1 \end{cases}$$

Proof (assuming strictly increasing CDF):

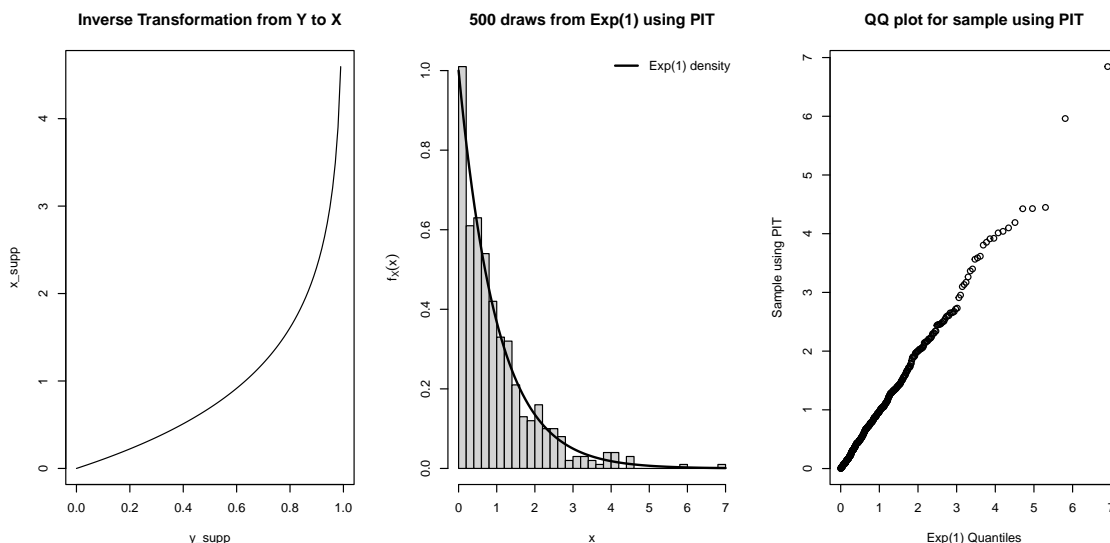
This theorem allows us to generate observations from a particular distribution. In fact, the real benefit of this theorem lies in its converse: If Y is uniformly distributed on $(0,1)$, then $X = F_X^{-1}(Y)$ has CDF $F_X(x)$.

Example Generate a random sample from $F_X(x) = 1 - e^{-x}I_{(0,\infty)}(x)$. Note: $X \sim \text{Exp}(\beta = 1)$

```
# Y = F_X(X) = 1 - exp(-X) <=> X = -ln(1-Y)
# THM 2.1.10 => Y ~ unif(0,1)
# plot the transformation on the support
y_supp <- seq(from = 0, to = 1, length = 100)
x_supp <- -log(1 - y_supp)
par(mfrow = c(1,3))
plot(x_supp ~ y_supp, type = "l",
     main = "Inverse Transformation from Y to X")
# generate random sample from uniform dist
y <- runif(n = 500, min = 0, max = 1)

# probability integral transformation (PIT)
x_samp <- -log(1 - y)

# compare samples to actual exp(1) dist
hist(x_samp, nclass = 25, freq = F,
     xlab = "x", ylab = expression(f[X](x)),
     main = "500 draws from Exp(1) using PIT")
curve(dexp(x, rate = 1/1), lwd = 2, add = T)
legend("topright", bty = "n", lwd = 2, lty = 1,
     legend = "Exp(1) density")
#qqplot to compare to theoretical exp(1) dist
qqplot(qexp(ppoints(500), 1), x_samp,
     xlab = "Exp(1) Quantiles", ylab = "Sample using PIT",
     main = "QQ plot for sample using PIT")
```



2.2 Expected Values

Ideally, we would know with certainty the outcome an experiment. However, this is often unrealistic, and we are instead left to deal with a random variable, i.e., a function from the sample space into the real numbers. By weighting the values of a random variable according to its probability distribution, we obtain what's known as the expected value, or expectation, of the random variable. This number is a measure of center that represents a "typical" value of the random variable, giving us an idea of what we might expect to happen in the long run.

DEF 2.2.1 (Expected Value) The *expected value* or mean of a random variable $g(X)$, denoted by $E(g(X))$, is

provided the sum or integral exists.

Note: $E(g(X))$ exists and is finite $\iff E(|g(X)|) < \infty$. That is, if $E(|g(X)|) = \infty$, then $E(g(X))$ does not exist.

Examples

- Meaghan, the president of the student chapter of the American Statistical Association at MSU has a fundraising idea for her club. She proposes a game in which a fair die is thrown. The player will pay \$5 to play and will win \$2 for each dot that appears on their roll. Suppose Emmanuel wanders by Meaghan's Catapalooza booth and sees the opportunity to play the game. Is it worth it for Emmanuel to pay \$5 to play?

- A machine produces copper wire, and occasionally there is a flaw at some point along the wire. The length of wire (in meters) produced between successive flaws is a continuous random variable X with *pdf* $f_X(x) = 2(1+x) - 3I_{(0,\infty)}(x)$. What is the expected length of the wire produced between successive flaws?
- Let $f_X(x) = e^{-x}I_{[0,\infty)}(x)$. Find the expected value of e^X .
- An insurance policyholder's loss, Y , follows a distribution with density function: $f_Y(y) = 2y^{-3}I_{(1,\infty)}$. The insurance company reimburses a loss up to a benefit limit of 20. What is the expected value of the benefit paid under the insurer's policy?

THM 2.2.5 (Properties of Expected Values) Let X be a random variable and let a , b , and c be constants. Then, for any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,

1. $E(ag_1(X) + bg_2(X) + c) = aE(g_1(x)) + bE(g_2(x)) + c.$
2. If $g_1(x) \geq 0$ for all x , then $E(g_1(x)) \geq 0.$
3. If $g_1(x) \geq g_2(x)$ for all x , then $E(g_1(x)) \geq E(g_2(x))$
4. If $a \leq g_1(x) \leq b$ for all x , then $a \leq E(g_1(x)) \leq b$

Proof: (of 1. for the continuous case)

Show $E(ag_1(X) + bg_2(X) + c) = aE(g_1(x)) + bE(g_2(x)) + c.$

Example (Revisited) Consider a game in which a fair die is thrown. The player pays \$5 to play and wins \$2 for each dot that appears on the roll. Is it worth the \$5 to enter?

2.3 Moments and Moment Generating Functions

The various moments of a distribution are an important class of expectations.

DEF 2.3.1 (Moments)

For each integer n , the n^{th} **moment** of X (or $F_X(x)$), μ'_n is $\mu'_n =$

The n^{th} **central moment** of X , μ_n , is $\mu_n =$ where $\mu = \mu'_1 = E(X)$.

DEF 2.3.2 (Variance/Standard Deviation)

The **variance** of a random variable X is its *second central moment* $Var(X) =$

The positive square root of $Var(X)$ is the *standard deviation* of X

More Facts About the Variance

- The first moment of a random variable, $\mu_X = \mu'_1 = E(X)$ must exist (i.e., be finite) in order for its second central moment, $Var(X) = \sigma_X^2$ to exist (i.e., be finite). In other words, $Var(X)$ can only exist if $E(X)$ exists. However, if $E(X)$ exists, $Var(X)$ does not necessarily exist. Therefore, $E(X)$ existing is a necessary, but not sufficient condition for $Var(X)$ existing.

THM 2.3.4 If X is a random variable with finite variance, then for any constants a and b ,

$$Var(aX + b) =$$

(see pg. 60 for proof)

- The following is an alternate formula for the variance of X :

Example (Revisited) Consider a game in which a fair die is thrown. The player pays \$5 to play and wins \$2 for each dot that appears on the roll. What is the variance of the net profit? The standard deviation? How would you interpret the standard deviation in context of the problem?

Example (Revisited) A machine produces copper wire, and occasionally there is a flaw at some point along the wire. The length of wire (in meters) produced between successive flaws is a continuous random variable X with *pdf*, $f_X(x) = 2(1+x) - 3I_{(0,\infty)}(x)$. What is the variance of the length of wire produced between successive flaws?

Example A Poisson random variable is a special discrete RV with *pmf*:

$$f_X(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & x = 0, 1, 2, 3, \dots \\ 0 & \text{else} \end{cases}$$

The Poisson distribution “counts” things in a fixed amount of time space, volume, etc. For example, the number of accidents at an intersection in a week, the number of hits to a website each minute, and the number of plants of a particular species found in a 1 m^2 area are all examples of random variables that may be modeled by a Poisson distribution. The constant, λ is a parameter of the distribution; it determines the numerical characteristics of the distribution. Find the expected value and variance of a Poisson random variable.

A special expected value that is quite useful is the moment generating function (mgf).

DEF 2.3.6 (Moment Generating Function) Suppose there is an $h > 0$ such that $E(e^{tX})$ exists for all $-h < t < h$. (That is, $E(e^{tX})$ exists for t in some neighborhood of zero.) For a random variable X with CDF $F_X(x)$, this expectation is called the *moment generating function (mgf)* of X , denoted by $M_X(t)$, and is found in the following manner:

Note: For $t = 0$, $M_X(0) = E(e^{0X}) = 1$ for any random variable X .

A moment generating function can be used to generate moments and to identify distributions.

THM 2.3.7 (Generating Moments) If X has mgf $M_X(t)$, then the n^{th} moment is equal to the n^{th} derivative of $M_X(t)$ evaluated at $t = 0$.

Proof: (for a continuous random variable)

Example Suppose X is a Poisson random variable with *pmf*: $f_X(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & x = 0, 1, 2, 3, \dots \\ 0 & \text{else} \end{cases}$.
Find the moment generating function of X and use it to find the mean and variance of X .

Example Suppose X is a Gamma random variable with parameters α and β . In this case,

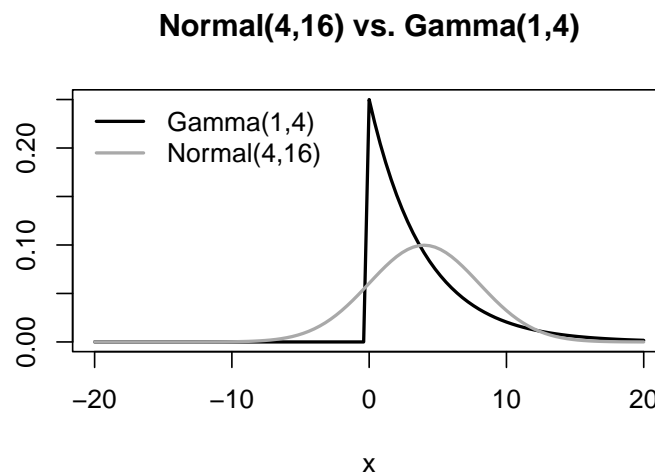
$$f_X(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}I_{(0,\infty)}(x).$$

Find the moment generating function of X and use it to find the mean and variance of X .

Note: The gamma function is defined as $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy$. Also, $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ and if α is an integer, $\Gamma(\alpha) = (\alpha - 1)!$.

Example Assume X , the lifetime of a randomly selected lightbulb, follows a Gamma distribution with a mean lifetime of 4 months and a standard deviation of 4 months. (You can verify that for this distribution, $\alpha = 1$ and $\beta = 4$). Assume Y , the low temperature on a randomly selected October day in Bozeman, follows a normal distribution with a mean of 4 degrees C and a standard deviation of 4 degrees C. How do the graphs of the pdfs compare to one another?

```
# plot gamma and normal on same fig
curve(dgamma(x, shape = 1, rate = 1/4),
      from = -20, to = 20, main = "Normal(4,16) vs. Gamma(1,4)",
      ylab = "", lwd = 2)
curve(dnorm(x, mean = 4, sd = 4), add = T,
      lwd = 2, col = "darkgray")
legend("topleft", lwd = c(2,2), col = c(1,"darkgray"),
      legend = c("Gamma(1,4)", "Normal(4,16)"),
      bty = "n")
```



Other Moments

In addition to the mean and variance of a random variable, there are other quantities that may be of interest. Two of these quantities are calculated from the 3rd and 4th central moments. Each of these central moments is normalized/standardized so that the quantities are dimensionless and represent the distribution independently of any linear change of scale.

Skewness: Measures lack of symmetry in a pdf and is defined by

Note: Random variables which are right-skewed have positive values for the skewness. Left-skewed random variables have negative values for the skewness. Symmetric random variables have a value of zero for the skewness.

Kurtosis: Measures the peakedness/flatness or the heaviness of the tails of a pdf and is defined by

Note: The normal distribution has a kurtosis of 3, and usually excess kurtosis (relative to the standard normal distribution) is calculated rather than kurtosis.

Using the MGF to Characterize Distributions

Although moment generating functions may be used to calculate moments, the main use of the moment generating function is to identify distributions. In fact, if the moment generating function, $M_X(t)$ exists in a neighborhood of zero, it uniquely determines the distribution of X and characterizes an infinite set of moments. However, an infinite set of moments does not uniquely determine a distribution, because there may be two distinct random variables that have the same moments (see Example 2.3.10).

This problem of uniqueness does not occur if the random variables have bounded support. If that is the case, then the infinite sequence of moments does uniquely determine the distribution. Furthermore, if the moment generating function exists in a neighborhood of zero, then the distribution is uniquely determined, no matter what its support.

THM 2.3.11 (Characterizing Distributions) Let $F_X(x)$ and $F_Y(y)$ be two CDFs, all of whose moments exist.

1. If X and Y have bounded support, then $F_X(u) = F_Y(u)$ for all u if and only if $E(X^r) = E(Y^r)$ for all integers $r = 0, 1, 2, \dots$
2. If the moment generating functions exist and $M_X(t) = M_Y(t)$ for all t in some neighborhood of zero, then $F_X(u) = F_Y(u)$ for all u .

THM 2.3.11 basically says that if two random variables have moment generating functions that exist and are equal, then they have the same distribution function. That is, when a moment generating function exists, it is unique. How can a moment generating function not exist? When the integral that must be evaluated to determine the moment generating function does not converge. In some cases, the moments will still exist, even when the moment generating function does not (e.g., lognormal distribution – Book Problems 2.35 and 2.36). There are also other distributions for which neither the moment generating function nor any moments exists (e.g., Cauchy distribution).

Example Suppose Y is a random variable with moment generating function, $M_Y(t) = (1 - 2t)^{-1}$, $t < 1/2$. What is the distribution of Y ?

THM 2.3.12 (Convergence of MGFs) Suppose $\{X_i, i = 1, 2, \dots\}$ is a sequence of random variables, each with moment generating function $M_{X_i}(t)$. Furthermore, suppose that $\lim_{i \rightarrow \infty} M_{X_i}(t) = M_X(t)$ for all t in a neighborhood of zero, and $M_X(t)$ is a moment generating function. Then there is a unique CDF, $F_X(x)$ whose moments are determined by $M_X(t)$ and, for all x where $F_X(x)$ is continuous, we have $\lim_{i \rightarrow \infty} F_{X_i}(x) = F_X(x)$ (convergence in distribution). That is, convergence, for $|t| < h$, of moment generating functions to a moment generating function implies convergence of CDFs.

Note: Example 2.3.13 illustrates how, as $n \rightarrow \infty$, the moment generating functions for $\text{Binomial}(n, p)$ random variables converge to the moment generating function of a Poisson random variable with $\lambda = np$. Therefore, as n gets large, a Poisson distribution may be used to approximate Binomial probabilities.

Moment generating functions can also be useful when dealing with transformations. A special case of this is the linear transformation of a random variable, but we'll see examples of other types of transformations (including those for functions of multiple random variables) later this semester.

THM 2.3.15 (MGF of Linear Function) For any constants a and b , the mgf of the random variable $Y = aX + b$ is given by

$$M_Y(t) =$$

Proof: (on your own)

Example Suppose Y is an Exponential random variable with parameter $\beta = 2$ and moment generating function, $M_Y(t) = (1 - 2t)^{-1}$, $t < 1/2$. What is the moment generating function of $U = 3Y + 10$?

2.4 Differentiating Under an Integral Sign (read on your own)

Casella and Berger discuss conditions under which it is legal to interchange the order of differentiation and integration, as well as the order of differentiation and summation. Both of these situations occur frequently in theoretical statistics. You may read this section on your own.

2.6.2 Other Generating Functions

In addition to the moment generating function, there are a number of other generating functions available (e.g., factorial moment generating function, cumulant generating function, probability generating function). The characteristic function is often the most useful of these.

Characteristic Function

Definition: The *characteristic function* of a random variable X is defined by

$$\phi_X(t) =$$

where i is the complex number $\sqrt{-1}$.

Notes:

- Moments, if they exist, can be found by differentiating $\phi_X(t)$.
- Although the moment generating function does not always exist, the characteristic function always does.
- Characteristic functions are unique and completely determine a distribution. Therefore, every cdf has a unique characteristic function.
- Finding the characteristic function requires complex integration.