# Mathematical Expectation (II): Function of a Random Variable

Le Wang

## Part 1: The Distribution of a function of a random variable

# The Distribution of a function of a random variable.

BH, Chapter 2.4

Data are often reported categorically. Education and Income are often reported in intervals.

$$
\begin{aligned}
Y &= \mu_1 \text{ if } -\infty < X \leq a \\
Y &= \mu_2 \text{ if } a < X \leq b \\
Y &= \mu_3 \text{ if } b < X
\end{aligned}
$$

## Examples in Economics

1. Decision Making Under Uncertainty: Maximization of Expected Utility of a random varible.

## Examples in Economics

1. Decision Making Under Uncertainty: Maximization of Expected Utility of a random varible.
2. (Paarsch and Hong, 2006) One of the most important features of **auctions**, when viewed through the lens of Harsanyi's noncooperative games of income information, is that

a bidder's equilibrium strategy is typically a function of his latent type – within the independent, private-value pradigm, his valuation.

Understanding how the dependence determines the distribution of the equilibrium bidding strategy is important.

# Road Map

1. The **distribution** and **expectation** of a function of a random variable (Continuous vs. Discrete)

# General Steps

1. Figure out the support points
2. Figure out the points of $X$ such that $g(x) \leq y$

$$B(y) = \{x \in \mathbb{R} : g(x) \leq y\}$$

3. $F(y) = \Pr[Y \leq y] = \Pr[X \in B(y)]$

# The Distribution and Expectation of a function of a random variable.

Two cases:

1. One-to-One: Simply move the support points from $X$ to $Y$. The probability is maintained.

$$\Pr[Y = y_i] = \Pr[X = x_i] \quad \text{where} \quad y_i = g(x_i)$$

# The Distribution and Expectation of a function of a random variable.

Two cases:

1. One-to-One: Simply move the support points from $X$ to $Y$. The probability is maintained.

$$\Pr[Y = y_i] = \Pr[X = x_i] \quad \text{where} \qquad y_i = g(x_i)$$

2. Many-to-One: Reduce the number of support points. For some $Y$, the probability is the sum of probabilities of several $X$.

$$\Pr[Y = y] = \Pr[\chi(y)] = \sum_{x_i \in \chi(y)} \Pr[X = x_i]$$

where

$$\chi(y) = \{x \in S_X : g(x) = y\}$$

Example 1

$Y = X^2$, $X \in \{-1, 0, 1\}$.

Then the support for $Y$ is $\{0, 1\}$

Example 1

$Y = X^2$, $X \in \{-1, 0, 1\}$.

Then the support for $Y$ is $\{0, 1\}$

$$\Pr[Y = 0] = \Pr[X = 0]$$
$$\Pr[Y = 1] = \Pr[X = -1] + \Pr[X = 1]$$

## Example 2

$$\Pr[Y = y] = \Pr[\chi(y)]$$

$$\Pr[Y = \mu_1] = \Pr[-\infty < X \leq a]$$

$$\Pr[Y = \mu_2] = \Pr[a < X \leq b]$$

$$\Pr[Y = \mu_3] = \Pr[b < X]$$

## Homework

Use the properties of the standard uniform variable, $U$. Can you think of a function such that $Y = f(U)$ so that $Y$ can take on six values $\{1, 2, 3, 4, 5, 6\}$ with the following distribution.

$$\begin{aligned}
\Pr[Y = 1] &= p_1 \\
\Pr[Y = 2] &= p_2 \\
\Pr[Y = 3] &= p_3 \\
\Pr[Y = 4] &= p_4 \\
\Pr[Y = 5] &= p_5 \\
\Pr[Y = 6] &= p_6
\end{aligned}$$

where $\sum p_i = 1$. This would be particularly useful for Monte Carlo Simulation for simulating a dice, fair or not.

# The Distribution and Expectation of a function of a random variable.

**Continuous Case** BH, Chapter 2.11

Let $X$ be a continuously distributed random variable with the CDF $F_X(x)$ and the PDF $f_X(x)$, where $g(x)$ is a montone increasing function . Then, $Y$ is continuously distributed with the CDF $F_X(g^{-1}(y))$ and the PDF

$$f_Y(y) = f_X(g^{-1}(y)) \frac{1}{g'(g^{-1}(y))}$$

One proof technique for such analysis is to write down the CDF first and then take the derivative.

**Advanced Example:** See the derivation of the distribution of **order statistics** in Casella and Berger's book, page 229.

# The Distribution and Expectation of a function of a random variable.

Prior to continuing, what are the properties of a monotone function?

1. the derivative, $g'(\cdot) = \frac{d}{dx} g(x)$, never changes sign.
2. there exists a unique inverse function, $g^{-1}(\cdot)$ such that

$$g^{-1}(y) = g^{-1}(g(x)) = x$$

# The Distribution and Expectation of a function of a random variable.

**Proof** (CDF of $Y$)

$$F_Y(y) = \Pr[Y \leq y]$$

# The Distribution and Expectation of a function of a random variable.

**Proof** (CDF of $Y$)

$$\begin{aligned} F_Y(y) &= \Pr[Y \leq y] \\ &= \Pr[g(X) \leq y] \end{aligned}$$

# The Distribution and Expectation of a function of a random variable.

**Proof** (CDF of $Y$)

$$\begin{aligned} F_Y(y) &= \Pr[Y \leq y] \\ &= \Pr[g(X) \leq y] \\ &= \Pr[X \leq g^{-1}(y)] \end{aligned}$$

## The Distribution and Expectation of a function of a random variable.

**Proof** (CDF of $Y$)

$$
\begin{aligned}
F_Y(y) &= \Pr[Y \leq y] \\
&= \Pr[g(X) \leq y] \\
&= \Pr[X \leq g^{-1}(y)] \\
&= F_X(g^{-1}(y))
\end{aligned}
$$

# The Distribution and Expectation of a function of a random variable.

**Proof** (PDF of $Y$)

$$\frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(g^{-1}(y))$$

# The Distribution and Expectation of a function of a random variable.

**Proof** (PDF of $Y$)

$$\begin{aligned}\frac{d}{dy}F_Y(y) &= \frac{d}{dy}F_X(g^{-1}(y))\\ &= f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y)\end{aligned}$$

# The Distribution and Expectation of a function of a random variable.

**Proof** (PDF of $Y$)

$$\begin{aligned}
\frac{d}{dy}F_Y(y) &= \frac{d}{dy}F_X(g^{-1}(y)) \\
&= f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) \\
&= f_X(g^{-1}(y))g^{-1'}(y)
\end{aligned}$$

Note that $g^{-1'}(y) = \frac{1}{g'(g^{-1}(y))}$. In other words, if the slope of $g()$ is $g'(\cdot)$, then the slope of the inverse function at the same point should be $\frac{1}{g'(\cdot)}$.

See BH for more concrete examples of how to apply this result to specific functions and parametric distributions.

# The Distribution and Expectation of a function of a random variable.

**Useful Expression** for the proof for the **expectation** of a function of a random variable later

$$f_Y(y)dy = f_X(x)dx$$

# The Distribution and Expectation of a function of a random variable.

**Proof:** (Skip in class. For your own reading)

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y)$$

# The Distribution and Expectation of a function of a random variable.

**Proof:** (Skip in class. For your own reading)

$$
\begin{aligned}
f_Y(y) = \frac{d}{dy} F_Y(y) &= f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) \\
&= f_X(x) \frac{d}{dy} x
\end{aligned}
$$

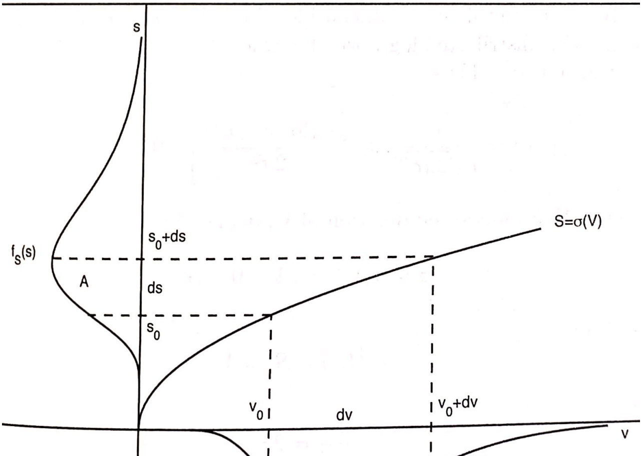
# The Distribution and Expectation of a function of a random variable.

**Proof:** (Skip in class. For your own reading)

$$
\begin{aligned}
f_Y(y) = \frac{d}{dy}F_Y(y) &= f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) \\
&= f_X(x)\frac{d}{dy}x \\
f_Y(y)dy &= f_X(x)dx
\end{aligned}
$$

# The Distribution and Expectation of a function of a random variable.

*Monotonic Nonlinear Transformations*



**Geometric Interpretation:** The area under the $f_S(s)$ (our case, $f_Y(y)$) curve between $s_0$ and $s_0 + ds$ ($s_0$ and $s_0 + dy$) must be equal to the area under the $f_V(v)$ (our case, $f_X(x)$) curve between $v_0$ and $v_0 + dv$ ($x_0$ and $x_0 + dx$).

## Homework

**Continuous Case** In your homework, you should also derive the formula for the case of monotnone **decreasing** functions. And combine the results together as one for **montone** functions, be it decreasing or increasing (in other words, Equation B-41 in Greene).

# The Distribution and Expectation of a function of a random variable.

These expressions would be particularly useful for writing down likelidhood functions for maximum likelihood estimation.

Homework: Consider a specific function of a random variable. The p-value, $p(z)$ from true nulls, is defined as

$$p(z) = \Pr[|Z| \geq |z|]$$

where $Z \sim N(0, 1)$ (the standard normal variable).

Show that this function is uniformly distributed. **Hint:** what is the relationship between $|z_1|$ and $|z_2|$ when we know $p(z_1) \geq p(z_2)$? And this should be a one-line proof.

Homework: Consider a specific function of a random variable. The p-value, $p(z)$ from true nulls, is defined as

$$p(z) = \Pr[|Z| \geq |z|]$$

where $Z \sim N(0, 1)$ (the standard normal variable).

Show that this function is uniformly distributed. **Hint:** what is the relationship between $|z_1|$ and $|z_2|$ when we know $p(z_1) \geq p(z_2)$? And this should be a one-line proof.

Application: This result is particularly useful for Hypothesis testing and machine learning on False Discovery Rate and Better design algorithm to reduce the false discovery rate. (see Matt Taddy, 2019, Business Data Science, pages 29-33)

Probability Integral Transformation later we will also derive the distribution of a particular function Monte Carlo Simulation

$$Y = F_X(X)$$

If you are interested, you can read the BH, Chapter 2.11 first!

## Example of Non-Monotonic Transformation (skip)

$$Y = g(X) = X^2$$

We know that $Y$ has the support $[0, \infty)$

Proof

$$\begin{aligned}
F_Y(y) &= \mathbb{P}[Y \leq y] \\
&= \mathbb{P}\left[X^2 \leq y\right] \\
&= \mathbb{P}[|X| \leq \sqrt{y}] \\
&= \mathbb{P}[-\sqrt{y} \leq X \leq \sqrt{y}] \\
&= \mathbb{P}[X \leq \sqrt{y}] - \mathbb{P}[X < -\sqrt{y}] \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y})
\end{aligned}$$

Proof

$$
\begin{aligned}
F_Y(y) &= \mathbb{P}[Y \le y] \\
&= \mathbb{P}\left[X^2 \le y\right] \\
&= \mathbb{P}[|X| \le \sqrt{y}] \\
&= \mathbb{P}[-\sqrt{y} \le X \le \sqrt{y}] \\
&= \mathbb{P}[X \le \sqrt{y}] - \mathbb{P}[X < -\sqrt{y}] \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y})
\end{aligned}
$$

$$
\begin{aligned}
f_Y(y) &= \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}} \\
&= \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}
\end{aligned}
$$

Homework:

How about the CDF and pdf for the following function?

$$Y = |X|$$

## Part 2: The Expectation of a function of a random variable

# The Expectation of a function of a random variable.

Let use state the results first. Let $Y = g(X)$.

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \sum g(x_i)p_i$$

where $p_i = \Pr[X = x_i]$

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

# The Expectation of a function of a random variable.

Note that the expectation of the function of a continuous variable is well defined if $\int_{-\infty}^{\infty} |g(x)|\, f(x)dx < \infty$. For a monotone function of a random variable, we can show the definition holds.

$$\begin{aligned} \mathbb{E}[Y] &= \int_{-\infty}^{\infty} y f_Y(y)dy \\ &= \int_{-\infty}^{\infty} g(x) f_X(x)dx \\ &= \mathbb{E}[g(X)] \end{aligned}$$

Later we will use this to prove a rather important result called **law of iterated expectation**, which is nothing but a simple application of our result here.

Specifically, we will consider a particular function of $X$ called **conditional expectation**

$$\mathbb{E}[Y \mid X] = m(X)$$

Let me highlight the significance of this result.

Example:

Derive the expectation of $Y = \lambda X$ where $X$ is distributed with $f(x) = \exp(-x)$ on $x \geq 0$.

How will you do it?

## Approach 1

$$\mathbb{E}[Y] = \int_0^\infty y \exp\left(-\frac{y}{\lambda}\right) \frac{1}{\lambda} dy = \lambda$$

Approach 1

$$\mathbb{E}[Y] = \int_0^\infty y \exp\left(-\frac{y}{\lambda}\right)\frac{1}{\lambda}dy = \lambda$$

Approach 2

$$\mathbb{E}[Y] = \mathbb{E}[\lambda X] = \int_0^\infty \lambda x f(x) dx = \lambda$$

## Approach 1

$$\mathbb{E}[Y] = \int_0^\infty y \exp\left(-\frac{y}{\lambda}\right) \frac{1}{\lambda} dy = \lambda$$

## Approach 2

$$\mathbb{E}[Y] = \mathbb{E}[\lambda X] = \int_0^\infty \lambda x f(x) dx = \lambda$$

## Approach 3

$$\mathbb{E}[Y] = \mathbb{E}[\lambda X] = \lambda \mathbb{E}[X] = \lambda$$

The significance of this definition is that we do not have to calculate the distribution of $Y$ first, and we can immediately derive the expectation based on the distribution of $X$ and the function, g(x), linking $X$ and $Y$.

# Homework

Show the definition for the case of discrete variable holds.

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \sum g(x_i)p_i$$

# Homework

1. **Theorem** Let $X$ be continuously distributed with the PDF $f_X(x)$. Let $a, b_1, b_2$ be some constants. Then,

$$\mathbb{E}[a + b_1 u_2(X) + b_2 u_2(X)] = a + b_1 \mathbb{E}[u_1(X)] + b_2 \mathbb{E}[u_2(X)]$$

2. **Fundamental Bridge Between Probability and Expectation**

$$F_X(x) = \Pr[X \leq x] = \mathbb{E}[\mathbb{I}(X \leq x)]$$

This result also lays the foundation for distributional regression.

# Homework

In machine learning, an important task is classification . Classification is about classifying an object into a particular group. For example, to identify whether or not an email is a spam or to predict whether or not someone will be elected into a public office. As you can immeditely recognize, this outcome of interest is actually a discrete variable, and classification is about predicting whether or not the outcome will be a particular value. Based on the distribution, one simplest possible classification algorithm is to classify or predict an outcome to be the most likely outcome (i.e., the value with the highest probability). This algorithm is intuitive, and can also be justified by minimizing the expected error.

Suppose that if you predict an outcome incorrectly, you receive an error of 1 and zero otherwise. What is the value that minimizing the expected error?

$$\min_{a} \quad \mathbb{E}[\mathbb{I}(X - a \neq 0)]$$

Note that we will skip Jesen's Inequality in this course, but that does not mean that it is not important. In fact, it is useful for deriving many econometric results, although not those in this course. I urge that you go through it and its applications whenever possible.

**Part : End**

# Answers to Homework Questions

The End (skip the slides below)

# The Expectation of a function of a random variable.

**Discrete Case** In your homework, you should also derive the formula for the case of discrete variables.

# The Expectation of a function of a random variable (A special Case)

**Theorem** Let $X$ be continuously distributed with the PDF $f_X(x)$. Let $a, b_1, b_2$ be some constants. Then,

$$\mathbb{E}[a + b_1 u_2(X) + b_2 u_2(X)] = a + b_1 \mathbb{E}[u_1(X)] + b_2 \mathbb{E}[u_2(X)]$$

# More on Expectation Operator

Show the following holds

$$F_X(x) = \Pr[X \leq x] = \mathbb{E}[\mathbb{I}(X \leq x)]$$

**Proof**

$$
\begin{aligned}
&\mathbb{E}[\mathbb{I}(X \leq x)] \\
&= \int \mathbb{I}(t \leq x) f(t) dt \\
&= \int_{-\infty}^{x} \mathbb{I}(t \leq x) f(t) dt + \int_{x}^{\infty} \mathbb{I}(t \leq x) f(t) dt \\
&= \int_{-\infty}^{x} 1 \cdot f(t) dt + \int_{x}^{\infty} 0 \cdot f(t) dt \\
&= \int_{-\infty}^{x} f(t) dt = F(x)
\end{aligned}
$$

# The Expectation of a function of a random variable.

Show that $\mathbb{E}[Y] = \mathbb{E}[g(X)] = \sum g(x_i)p_i$

**Discrete Case** Answer:

# The Expectation of a function of a random variable.

**Discrete Case** Answer:

$$\mathbb{E}[Y] = \sum_{y \in S_Y} y \cdot P_Y(y)$$

# The Expectation of a function of a random variable.

**Discrete Case** Answer:

$$
\begin{aligned}
\mathbb{E}[Y] &= \sum_{y \in S_Y} y \cdot P_Y(y) \\
&= \sum_{y \in S_Y} y \cdot \Pr[\chi(y)]
\end{aligned}
$$

# The Expectation of a function of a random variable.

**Discrete Case** Answer:

$$
\begin{aligned}
\mathbb{E}[Y] &= \sum_{y \in S_Y} y \cdot P_Y(y) \\
&= \sum_{y \in S_Y} y \cdot \Pr[\chi(y)] \\
&= \sum_{y \in S_Y} y \cdot \sum_{x_i \in \chi(y)} \Pr[X = x_i]
\end{aligned}
$$

# The Expectation of a function of a random variable.

**Discrete Case** Answer:

$$
\begin{aligned}
\mathbb{E}[Y] &= \sum_{y \in S_Y} y \cdot P_Y(y) \\
&= \sum_{y \in S_Y} y \cdot \Pr[\chi(y)] \\
&= \sum_{y \in S_Y} y \cdot \sum_{x_i \in \chi(y)} \Pr[X = x_i] \\
&= \sum_{y \in S_Y} \left( \sum_{x_i \in \chi(y)} y \cdot \Pr[X = x_i] \right)
\end{aligned}
$$

# The Expectation of a function of a random variable.

**Discrete Case** Answer:

$$
\begin{aligned}
\mathbb{E}[Y] &= \sum_{y \in S_Y} y \cdot P_Y(y) \\
&= \sum_{y \in S_Y} y \cdot \Pr[\chi(y)] \\
&= \sum_{y \in S_Y} y \cdot \sum_{x_i \in \chi(y)} \Pr[X = x_i] \\
&= \sum_{y \in S_Y} \left( \sum_{x_i \in \chi(y)} y \cdot \Pr[X = x_i] \right) \\
&= \sum_{y \in S_Y} \left( \sum_{x_i \in \chi(y)} g(x_i) \cdot \Pr[X = x_i] \right)
\end{aligned}
$$

## The Expectation of a function of a random variable.

**Discrete Case** Answer:

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{y \in S_Y} y \cdot P_Y(y) \\
&= \sum_{y \in S_Y} y \cdot \Pr[\chi(y)] \\
&= \sum_{y \in S_Y} y \cdot \sum_{x_i \in \chi(y)} \Pr[X = x_i] \\
&= \sum_{y \in S_Y} \left( \sum_{x_i \in \chi(y)} y \cdot \Pr[X = x_i] \right) \\
&= \sum_{y \in S_Y} \left( \sum_{x_i \in \chi(y)} g(x_i) \cdot \Pr[X = x_i] \right) \\
&= \sum_{x \in S_X} g(x_i) \cdot \Pr[X = x_i]
\end{aligned}$$

# The Expectation of a function of a random variable.

**Discrete Case** Answer Explained:

1. The second equality holds by $P_Y(y)$ derived above
2. The third equality holds by distributing $y$ with the terms $P_X(x)$ (holding $y$ constant in the inner sum).
3. The last equality holds because summing over all $x$'s in $\chi(y)$ and then summing over all $y$'s in $S_Y$ is equivalent to summing over all $x \in S_X$.

# The Expectation of a function of a random variable (A special Case)

**Theorem** Let $X$ be continuously distributed with the PDF $f_X(x)$. Let $a, b_1, b_2$ be some constants. Then,

$$\mathbb{E}[a + b_1 u_2(X) + b_2 u_2(X)] = a + b_1 \mathbb{E}[u_1(X)] + b_2 \mathbb{E}[u_2(X)]$$