

MATH 320: Probability

Lecture 13: Functions of Random Variables

Chapter 5: Distributions of Functions of Random Variables (5.1)

Deductibles and caps: Expected value of a function of a random variable

Expected value of a loss or claim

- These examples are in an insurance applications, but are just expected value of a function of a random variable problems.
- Insurance loss.
 - Example: (a) The amount of a single loss X for an insurance policy is exponential, with density function

$$f(x) = 0.002e^{-0.002x}, \quad x \geq 0 \quad \implies \quad X \sim \text{Exp}(\lambda = 0.0002)$$

So the (base) expected value of a single loss is: $E(X) = \frac{1}{\lambda} = 500$

- Insurance with a deductible.
 - Continuing example: (b) Suppose now the insurance policy has a deductible of \$100 for each loss. Find the expected value of a single claim.

** Now loss amount \neq claim amount

★ – *STRATEGY*: We need to write a new function $g(X)$ that represents the new claim amount taking into account the deductible.

★ $g(X)$ will be a piecewise function. So think about the values $g(X)$ takes in cases based on the range of X .

NOTE: We are thinking about the values of the claim from the insurance company's perspective.

claim amount

$$g_1(x) = \begin{cases} 0 & 0 \leq x < 100 \\ x-100 & x \geq 100 \end{cases}$$

↓
policyholder pays first 100

$$\begin{aligned} E[g_1(X)] &= \int_0^{\infty} g_1(x) f(x) dx \\ &= \int_0^{100} 0 f(x) dx + \int_{100}^{\infty} (x-100)(0.002 e^{-0.002x}) dx \\ &= 0 + \text{< Integration by parts >} \\ &= 409.37 < 500 = E(X) \end{aligned}$$

- Insurance with a deductible and a cap.
 - Continuing example: (c) Now suppose the insurance policy has a deductible of \$100 per claim AND a restriction that the largest amount paid on any claim will be \$700.
 - Strategy: Use the same strategy as before for the first case, then just need to take into account the cap.

claim amount

$$g_2(x) = \begin{cases} 0 & 0 \leq x \leq 100 \\ x-100 & 100 \leq x \leq 800 \\ 700 & x > 800 \end{cases}$$

→ payments are capped at 700

⇒ loss > 800 receives a payment of 800-100 = 700

$$\begin{aligned} E[g_2(x)] &= \int_0^{\infty} g_2(x) \underbrace{f(x)}_{\text{Exp}(1) = 0.3679} dx \\ &= \int_0^{100} 0 f(x) dx + \int_{100}^{800} (x-100) f(x) dx + \int_{800}^{\infty} 700 f(x) dx \\ &= 0 + \text{Integration by parts} + 700 \int_{800}^{\infty} f(x) dx \\ &= 167.09 + 143.33 \\ &= 310.42 < 409.37 = E[g_1(x)] \end{aligned}$$

- Another example: The amount of a single loss X for an insurance policy has the density function $f(x)$ for $x \geq 0$ with deductible of \$150 and cap of \$900.
 - Find a function $g(X)$ for the amount paid (claim amount) for a loss x .
 - Write the integral to solve for the expected claim amount.

$$a) \quad g(x) = \begin{cases} 0 & 0 \leq x \leq 150 \\ x-150 & 150 \leq x \leq 1050 \\ 900 & x > 1050 \end{cases}$$

$$b) \quad E[g(x)] = \int_0^{150} 0 f(x) dx + \int_{150}^{1050} (x-150) f(x) dx + \int_{1050}^{\infty} 900 f(x) dx$$

- In general, if loss x with deductible d and cap c , we have

$$g(x) = \begin{cases} 0 & x < d \\ x-d & d \leq x \leq d+c \\ c & x > d+c \end{cases}$$

The distribution $Y = g(X)$

Transformations so far

- We have already seen simple methods for finding $E[g(X)]$ and $V[g(X)]$ for any type of variable.
- Example: The monthly maintenance cost for a machine $X \sim \text{Exponential}(\lambda = 0.01)$. Next year costs will be increased 5% due to inflation. Thus next year's monthly cost is $Y = g(X) = 1.05X$.

Find $E(Y)$.

$$E(X) = \frac{1}{\lambda} = 100 \rightarrow E(Y) = E(1.05X) = 1.05E(X) = 105$$

- Note we did not need to know the distribution of Y for this calculation.

However, the mean and variance alone are not sufficient to enable us to calculate probabilities for $Y = g(X)$, we need the actual distribution function $f(y)$.

- Discrete example: Same X with a new (discrete) model and inflation costs $Y = g(X) = 1.05X$:

(a) Find the distribution of $Y = g(X)$.

(b) Find $P(Y < 100)$.

x	$f(x)$	$y = 1.05x$	$f(y)$
0	0.28	0	0.28
50	0.43	52.5	0.43
100	0.20	105	0.20
150	0.09	157.5	0.09

$$P(Y < 100) = P(Y = 0) + P(Y = 52.5) \\ = 0.28 + 0.43 \\ = 0.71$$

- For the original continuous model, it is not as simple to find the new distribution.

Continuous transformations example

- Continuing example: Using the original $X \sim \text{Exponential}(\lambda = 0.01)$ model...
- Find $P(Y \leq 100)$.

★ ★ GOAL: Get the probability statement to be with respect to X .

★ ★ STRATEGY: "Indirectly" find the probability for Y based on the known cdf of X and using some simple algebra. Note that this is the same strategy we used to find lognormal probabilities based on the normal cdf.


$$\begin{aligned} \rightarrow \text{know } F_X(x) &= 1 - e^{-0.01x}, \quad x \geq 0 \\ \Rightarrow P(Y \leq 100) &= P(1.05X \leq 100) &< \text{substituting} > \\ &= P(X \leq \frac{100}{1.05}) &< \text{rearrange to get } x \text{ by itself} > \\ &= F_X\left(\frac{100}{1.05}\right) = 1 - e^{-0.01\left(\frac{100}{1.05}\right)} \\ &\approx 0.614 \end{aligned}$$

- Find the cdf $F_Y(y)$.

STRATEGY: Use the same reasoning as above, just for a general y :
 $P(Y \leq 100) = F_Y(100) \rightarrow P(Y \leq y) = F_Y(y)$ for any value $y \geq 0$.

$$\begin{aligned}
 F_Y(y) &= P_Y(Y \leq y) \\
 &= P_X(1.05X \leq y) \\
 &= P_X(X \leq \frac{y}{1.05}) \\
 &\downarrow \\
 &= F_X(\frac{y}{1.05}) = 1 - e^{-0.01(\frac{y}{1.05})}, \quad y \geq 0
 \end{aligned}$$

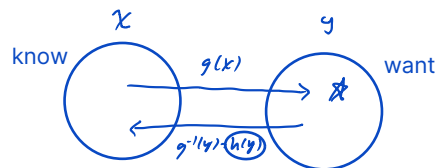
- Note that the range of X is the interval $[0, \infty)$. The range for $Y = 1.05X$ is the same interval. This will not always be the case for transformations $g(X)$.

 **STRATEGY:** Apply $g(x)$ to all pieces, ALWAYS check both sides.

$$\begin{aligned}
 \rightarrow \quad & X: \quad 0 \leq X < \infty \\
 \rightarrow Y = g(X) = 1.05X: \quad & 1.05(0) \leq 1.05X < 1.05(\infty) \\
 & 0 \leq Y < \infty
 \end{aligned}$$

Inverses

- Finding the distribution of $Y = g(X)$ like we did above is much simpler when the transformation function $g(X)$ has an inverse.
- Recall that the function $g(X)$ defines a mapping from the original range of X to a range of Y . That is,



** We do not know stuff (pdf, cdf, etc.); so we have to use the inverse function to go backwards. \mathcal{Y} is completely determined by \mathcal{X} .

- When do inverse functions exist?

If the function $g(x)$ is strictly **monotone** \implies one-to-one \iff inverse exists.

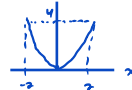
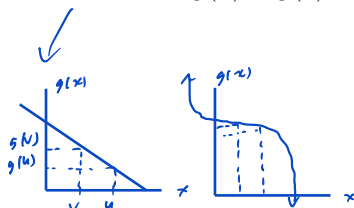
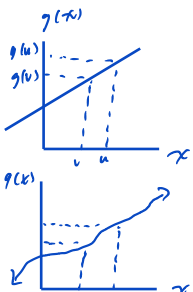
$u > v \implies g(u) > g(v) \implies$ strictly increasing

$u > v \implies g(u) < g(v) \implies$ strictly decreasing

uniquely pairs
 x 's & y 's

$$g^{-1}(y) = h(y) \quad \checkmark$$

ex) Not one-to-one
 $g(x) = x^2$



- For a function $g(x)$ that strictly increasing or strictly decreasing on the range of X , we can find an inverse function $h(y)$ defined on the range of Y . Thus we have:

$$h[g(x)] = x \quad \text{and} \quad g[h(y)] = y$$

A

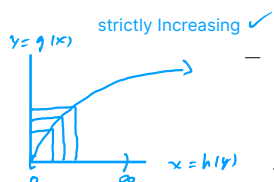
- If transformation is strictly increasing or strictly decreasing over \mathcal{X} , then use the methods described next.

- ★ 2. Check range of Y (i.e. ALSO transform range of X to range of Y).

- We will only generalize the methods for when $q(X)$ has an inverse. If this is true, then there are two cases.

- Let $h(y)$ be the inverse function of $g(x)$. The function $h(y)$ will also be strictly increasing. In this case, we can find $F_Y(y)$ as follows:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(h[g(X)] \leq h(y)) \\ &= P(X \leq h(y)) \\ &= F_X(h(y)) \quad \star \end{aligned}$$



- Example: Let $X \sim \text{Exponential}(\lambda = 3)$. Find the cdf of $Y = \sqrt{X}$.

$$\rightarrow F_X(x) = 1 - e^{-3x}, x \geq 0$$

There are two ways that we can solve this.

Long way

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\sqrt{X} \leq y) \\ &= P(X \leq y^2) \\ &= F_X(y^2) = 1 - e^{-y^2} \end{aligned}$$

Short way

$$\rightarrow y = \sqrt{x} \quad \xrightarrow{\text{solve for } x} \quad x = h(y) = y^2$$

Range:

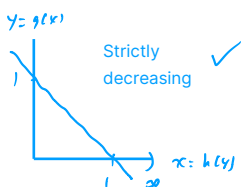
$$\begin{array}{lcl} x: & 0 & \leq x \leq \infty \\ y = g(x) = \sqrt{x}: & \sqrt{0} & \leq \sqrt{x} \leq \sqrt{\infty} \\ & 0 & \leq y \leq \infty \end{array}$$

• **Case 2: $g(x)$ is strictly decreasing on the range of X**

- Let $h(y)$ be the inverse function of $g(x)$. The function $h(y)$ will also be strictly decreasing. In this case, we can find $F_Y(y)$ as follows:

$$\begin{aligned}
 \rightarrow F_Y(y) &= P(Y \leq y) \\
 &= P(g(X) \leq y) \\
 &= P(h(g(X)) \geq h(y)) \\
 &= P(X \geq h(y)) \quad \rightarrow \text{survival function } S_X(h(y)) \\
 &= 1 - F_X(h(y)) \quad \star
 \end{aligned}$$

have to flip inequality because decreasing function \star



- Example: Let $X \sim \text{Exponential}(\lambda = 3)$. Find the cdf of $Y = 1 - X$.

Again, we can do the long (“derivation”) way or short way (skip to end result).

Long way

$$\begin{aligned}
 F_Y(y) &= P(Y \leq y) \\
 &= P(1 - X \leq y) \\
 &= P(X \geq 1 - y) \\
 &= 1 - F_X(1 - y) = 1 - [1 - e^{-3(1-y)}] \\
 &= e^{-3(1-y)}, \quad y \leq 1
 \end{aligned}$$

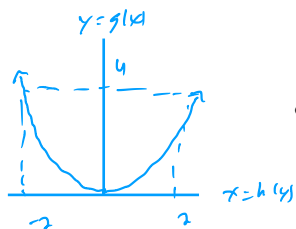
Short way \star

$$\begin{aligned}
 \rightarrow Y = 1 - X &\Rightarrow X = h(Y) = 1 - Y \\
 \rightarrow F_Y(y) &= 1 - F_X(1 - y) \\
 &= 1 - [1 - e^{-3(1-y)}] \\
 &= e^{-3(1-y)}, \quad y \leq 1
 \end{aligned}$$

Range:

$$\begin{aligned}
 X: & \quad 0 \leq X < \infty \\
 Y = g(X) = 1 - X: & \quad 1 - 0 \leq 1 - X < 1 - \infty \\
 & \quad 1 \leq Y < -\infty
 \end{aligned}$$

$$\Rightarrow -\infty < Y \leq 1 \quad \star \text{ bounds flip sides}$$



Not monotone (one-to-one)
 \Rightarrow no $h(y)$ over x

- If $g(x)$ does NOT have an inverse

– Example: Let $X \sim \text{Uniform}(a = -2, b = 2)$. Find the cdf of $Y = X^2$.

$$\begin{aligned} \rightarrow f_X(x) &= \frac{x - (-2)}{4} \\ &= \frac{x+2}{4} \end{aligned}$$

$$\rightarrow F_Y(y) = P(Y \leq y)$$

$$= P(X^2 \leq y)$$

$$= P(|X| \leq \sqrt{y})$$

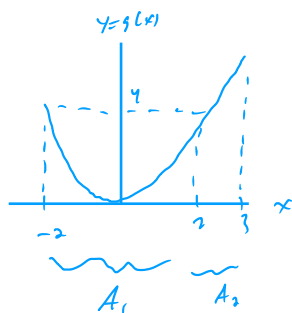
$$= P(-\sqrt{y} \leq X \leq \sqrt{y})$$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \frac{\sqrt{y}+2}{4} - \frac{-\sqrt{y}+2}{4}$$

$$= \frac{\sqrt{y}}{2}, \quad 0 \leq y \leq 4$$

– It can be even more complicated if there isn't a "balanced" range of Y .

Example: Let $X \sim \text{Uniform}(a = -2, b = 3)$. Find the cdf of $Y = X^2$.



\Rightarrow solve each A_i individually + piecewise

$$F_Y(y) = \begin{cases} \text{~~~~~} & x \in A_1 \\ \text{~~~~~} & x \in A_2 \end{cases}$$

– Both of these cases will be left for grad school :)

Finding the density function $f_Y(y)$ for $Y = g(X)$

- Finding $F_Y(y)$ gives us all the information that is needed to calculate probabilities for Y , as shown below:

$$P(Y \leq y) = F_Y(y) \quad P(Y \geq y) = 1 - F_Y(y) \quad P(a \leq Y \leq b) = F_Y(b) - F_Y(a)$$

Thus there is no real need to find the density function $f_Y(y)$. If the density function is required, it can be found by differentiating the cdf:

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

- If X is continuous, it is usually easier to find the cdf of Y and then the pdf of Y (rather than skipping straight to the pdf). But we will learn both methods, which we shall name:

Logic

1. Cdf method

$$\text{pdf } X \rightarrow \text{cdf } X \rightarrow \text{cdf } Y \rightarrow \text{pdf } Y$$

2. Pdf method

$$\text{pdf } X \rightarrow \text{pdf } Y$$

(aka change of variable technique)

- Again when working in situations when $g(x)$ has an inverse, there are two cases:

- Case 1: $g(x)$ is strictly increasing on the range of X**

- Setup: $h(y)$ is the inverse of $g(x)$ and $h(y)$ is strictly increasing.
- Previous results: $F_Y(y) = F_X(h(y))$
- We can find the pdf $f_Y(y)$ as follows:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(h(y)) = f_X'(h(y)) h'(y) = f_X(h(y)) h'(y)$$

Substitute chain rule simplify

- Case 2: $g(x)$ is strictly decreasing on the range of X**

- Setup: $h(y)$ is the inverse of $g(x)$ and $h(y)$ is strictly decreasing.
- Previous results: $F_Y(y) = 1 - F_X(h(y))$
- We can find the pdf $f_Y(y)$ as follows:

(Same steps)


$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} [1 - F_X(h(y))] = -f_X'(h(y)) h'(y) = -f_X(h(y)) h'(y)$$

- Since $h(y)$ is decreasing, its derivative is negative. Thus the final expression above is actually positive.

rewrite as $\rightarrow f_X(h(y)) \underbrace{(-h'(y))}_{>0} = f_X(h(y)) |h'(y)|$ * holds in both cases

Just simplify with absolute value

case 1 $f_Y(y)$ because $h'(y) < 0$ when increasing



- Theorem: Let X have cdf $F_X(x)$ with range \mathcal{X} , $Y = g(X)$ with range \mathcal{Y} and inverse $h(y)$.

- If $g(x)$ is strictly increasing on $\mathcal{X} \rightarrow F_Y(y) = F_X(h(y))$ for $y \in \mathcal{Y}$.
- If $g(x)$ is strictly decreasing on $\mathcal{X} \rightarrow F_Y(y) = 1 - F_X(h(y))$ for $y \in \mathcal{Y}$.
- If $g(x)$ is strictly increasing or strictly decreasing on \mathcal{X} , then

$$f_Y(y) = f_X(h(y)) \cdot |h'(y)| \quad \text{for } y \in \mathcal{Y}.$$

- Return to previous examples: Let $X \sim \text{Exponential}(\lambda = 3)$.

(a) Find the pdf of $Y = \sqrt{X}$.

Cdf method

$$\begin{aligned} \textcircled{1} & \text{ previously showed } F_Y(y) = 1 - e^{-3y^2} \\ \textcircled{2} & f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} [1 - e^{-3y^2}] \\ & \quad \downarrow \\ & = -e^{-3y^2} (-3(2)y) \\ & = 6ye^{-3y^2}, \quad y \geq 0 \end{aligned}$$

Pdf method

$$\begin{aligned} \rightarrow x = h(y) = y^2 & \rightarrow h'(y) = 2y \\ \rightarrow f_Y(y) = f_X(y^2) |h'(y)| \\ & \quad \downarrow \\ & = 3e^{-3(y^2)} |2y| \\ & = 6ye^{-3y^2}, \quad y \geq 0 \end{aligned}$$

(ranges shown previously)

(b) Find the pdf of $Y = 1 - X$.

Cdf method

$$\begin{aligned} \textcircled{1} & \text{ previously showed } F_Y(y) = e^{-3(1-y)} \\ \textcircled{2} & f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} [e^{-3(1-y)}] \\ & \quad \downarrow \\ & = e^{-3(1-y)} (3) \\ & = 3e^{-3(1-y)}, \quad y \leq 1 \end{aligned}$$

Pdf method

$$\begin{aligned} \rightarrow x = h(y) = 1 - y & \rightarrow h'(y) = -1 \\ \rightarrow f_Y(y) = f_X(1-y) |h'(y)| \\ & \quad \downarrow \\ & = 3e^{-3(1-y)} |-1| \\ & = 3e^{-3(1-y)}, \quad y \leq 1 \end{aligned}$$

(ranges shown previously)

More examples

1. Let X be the outcome when you roll a fair four sided die. If you get $Y = |X - 2|$ dollars based on your roll, find $f_Y(y)$.

★ Discrete Random variable
(don't need cdf/pdf method)

$$\rightarrow X = \{1, 2, 3, 4\} \rightarrow Y = \{0, 1, 2\}$$

$$\rightarrow f_Y(0) = P(X=2) = 1/4$$

$$f_Y(1) = P(X=1) + P(X=3) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$f_Y(2) = P(X=4) = 1/4$$

$$\rightarrow f_Y(y): \begin{cases} 1/4 & y=0 \\ 1/2 & y=1 \\ 1/4 & y=2 \end{cases}$$

2. Let $X \sim \text{Poisson}(\lambda = 4)$. If $Y = X^2$, find the ^{pmf} ~~pdf~~ of Y .

↳ Discrete

$$\rightarrow Y = g(X) = X^2 \text{ is one-to-one over } X = \{0, 1, 2, \dots\} \checkmark$$

$$\rightarrow P(Y=y) = P(X^2=y) = P(X=\sqrt{y})$$

$$= f_X(\sqrt{y}) = \frac{e^{-4} 4^{\sqrt{y}}}{(\sqrt{y})!}, \quad y = \{0^2, 1^2, 2^2, 3^2, \dots\}$$