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

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

    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

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

policyholder pays first 100

$$E[g_1(X)] = \int_0^\infty g_1(x) \underbrace{f(x)}_{\hookrightarrow \text{Exp}(\lambda = 0.002)} dx$$

$$= \int_0^{100} 0 \cancel{f(x)}\, dx + \int_{100}^\infty (x-100)\left(0.002 e^{-0.002x}\right) dx$$

$$0 \quad + \quad \text{< Integration by parts >}$$

$$= 409.37 < 500 = E(X)$$

- 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

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

$\rightarrow$ payments are capped at 700

$\Rightarrow$ loss > 800 recieves a payment of 800-100 = 700

$$E[g_2(x)] = \int_0^\infty g_2(x) f(x)\, dx \qquad \hookrightarrow \text{Exp}(\lambda = 0.007)$$

$$= \int_0^{100} 0\, f(x)\, dx + \int_{100}^{800} (x-100) f(x)\, dx + \int_{800}^{\infty} 700 f(x)\, dx$$

$$= 0 + \text{< Integration by parts >} + 700\, S_x(800)$$

$$= 167.09 + 143.33$$

$$= 308.42 < 404.37 = E[g_1(x)]$$

- Another example: The amount of a single loss X for an insurance policy has the density function $f(x)$ for $x \ge 0$ with deductible of \$150 and cap of \$900.

  (a) Find a function $g(X)$ for the amount paid (claim amount) for a loss x.
  (b) Write the integral to solve for the expected claim amount.

a)

$$g(X) = \begin{cases} 0 & 0 < x \le 150 \\ x-150 & 150 < x \le 1050 \\ 900 & x > 1050 \end{cases}$$

b) $$E[g(X)] = \int_0^{150} 0\, f(x)\, dx + \int_{150}^{1050} (x-150) f(x)\, dx + \int_{1050}^{\infty} 950\, f(x)\, dx$$

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

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

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 $X \sim$ 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 $E(Y)$.

  $E(x) = \frac{1}{\lambda} = 100 \longrightarrow E(y) = E(1.05x) = 1.05E(x) = 105$

- Note we did not need to 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 $f(x)$.
- 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 $X \sim$ Exponential $(\lambda = 0.01)$ model...
- Find $P(Y \leq 100)$.

GOAL: Get the probability statement to be with 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.

$\rightarrow$ know $F_X(x) = 1 - e^{-0.01x}, \quad x \geq 0$

$\Rightarrow P(Y \leq 100) = P(1.05X \leq 100)$ < substituting >
$= P(X \leq \frac{100}{1.05})$ < rearrange to get x by itself >
$= F_X(\frac{100}{1.05}) = 1 - e^{-0.01(\frac{100}{1.05})}$
≈ 0.614

- Find the cdf $F_Y(y)$.

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

$$F_Y(y) = P_Y(Y \le y) = P_X(1.05X \le y) = P_X\left(X \le \frac{y}{1.05}\right) = F_X\left(\frac{y}{1.05}\right) = 1 - e^{-0.01\left(\frac{y}{1.05}\right)}, \quad y \ge 0$$

- 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.

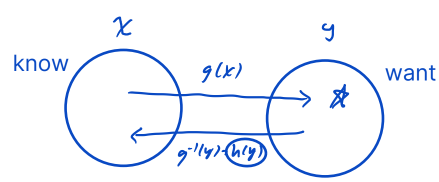
$\rightarrow \quad x: \quad 0 \le x < \infty$

$\rightarrow \quad y = g(x) = 1.05x: \quad 1.05(0) \le 1.05x < 1.05(\infty)$

$\qquad 0 \le y < \infty$

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 $\mathcal{X}$ 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?

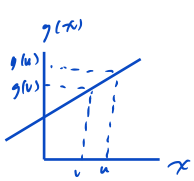
  If the function $g(x)$ is strictly **monotone** $\implies$ one-to-one $\iff$ inverse exists.

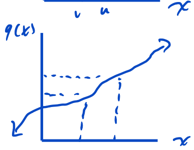
  $u > v \Rightarrow g(u) > g(v)$ $\rightarrow$ strictly increasing

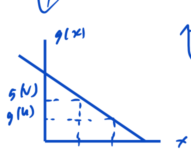
  $u > v \Rightarrow g(u) < g(v)$ $\rightarrow$ strictly decreasing

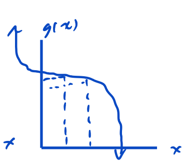
  uniquely pairs x's & y's

  $g^{-1}(y) = h(y)$ ✓



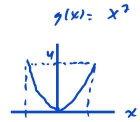






ex) Not one-to-one

$g(x) = x^2$



- Summary and results:

  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:

  $$h[g(x)] = x \quad \& \quad g[h(y)] = y$$

** Strategy when problem solving:

1. Draw a figure of the transformation.

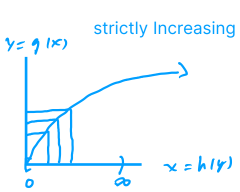
   If transformation is strictly increasing or strictly decreasing over $\mathcal{X}$, then use the methods described next.

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

Using $F_X(x)$ to find $F_Y(y)$ for $Y = g(X)$

- We will only generalize the methods for when $g(X)$ has an inverse. If this is true, then there are two cases.
- **Case 1: $g(x)$ is strictly increasing on the range of X**
  - 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:

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



  - Example: Let $X \sim \text{Exponential}(\lambda = 3)$. Find the cdf of $Y = \sqrt{X}$. $\rightarrow F_X(x) = 1 - e^{-3x}, \; x \ge 0$

    There are two ways that we can solve this.

    Long way

    $$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(\sqrt{X} \le y) \\ &= P(X \le y^2) \\ &= F_X(y^2) = 1 - e^{-3(y^2)} \end{aligned}$$

    Short way

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

    $\rightarrow F_Y(y) = F_X(h(y)) = F_X(y^2) = 1 - e^{-3y^2}$

rotate → $h(y)$ strictly increasing on $Y = [0, \infty)$

Range:

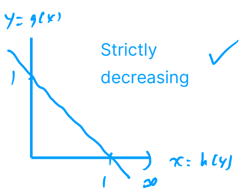
$$\begin{aligned} x: &\quad 0 \le x \le \infty \\ y = g(x) = \sqrt{x}: &\quad \sqrt{0} \le \sqrt{x} \le \sqrt{\infty} \\ &\quad 0 \le y \le \infty \end{aligned}$$

- **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:

$$\begin{aligned}
\to F_Y(y) &= P(Y \le y) \\
&= P(g(X) \le y) \\
&= P(h[g(X)] \ge h(y)) \\
&= P(X \ge h(y)) \to S_X(h(y)) \\
&= 1 - F_X(h(y))
\end{aligned}$$

have to flip inequality because decreasing function

survival function



Strictly decreasing ✓

  - 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

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

    Short way

$$\to Y = 1 - X \implies x = h(y) = 1 - y$$

$$\begin{aligned}
\to F_Y(y) &= 1 - F_X(1-y) \\
&= 1 - [1 - e^{-3(1-y)}] \\
&= e^{-3(1-y)}, \quad y \le 1
\end{aligned}$$

Range:

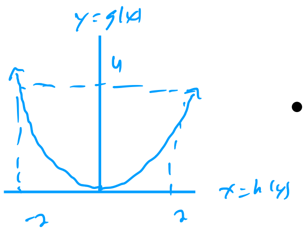
$x: \quad 0 \le x < \infty$

$y = g(x) = 1 - x: \quad 1 - 0 \ge 1 - x > 1 - \infty$

$1 \ge y > -\infty$

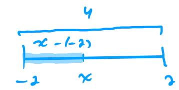
$\implies -\infty < y \le 1$

bounds flip sides



Not monotone (one-to-one)

$\Rightarrow$ no h(y) over x



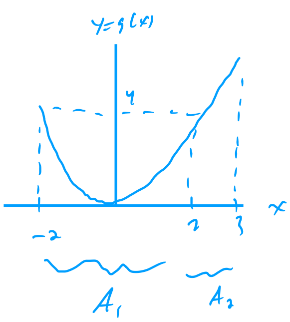
$\rightarrow F_X(x) = \frac{x-(-2)}{4} = \frac{x+2}{4}$

- **If $g(x)$ does NOT have an inverse**
  - Example: Let $X \sim \text{Uniform}\,(a = -2, b = 2)$. Find the cdf of $Y = X^2$.

$$\begin{aligned}
\rightarrow F_Y(y) &= P(Y \le y) \\
&= P(X^2 \le y) \\
&= P(|X| \le \sqrt{y}) \\
&= P(-\sqrt{y} \le X \le \sqrt{y}) \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \frac{\sqrt{y}+2}{4} - \frac{-\sqrt{y}+2}{4} \\
&= \frac{\sqrt{y}}{2}, \quad 0 \le y \le 4
\end{aligned}$$

  - It can be even more complicate 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

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

  - 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:

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

  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:

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

- 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 — pdf X $\longrightarrow$ cdf X $\longrightarrow$ cdf Y $\longrightarrow$ pdf Y
  2. Pdf method — pdf X $\longrightarrow$ 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)

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

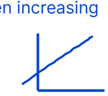
  - Since $h(y)$ is decreasing, its derivative is negative. Thus the final expression above is actually positive.

rewrite as $\rightarrow$ $f_X(h(y)) \underbrace{\underbrace{(-h'(y))}_{-(<0)}}_{>0} = f_X(h(y))\,|h'(y)| =$ case 1 $f_Y(y)$ because $h'(y) > 0$ when increasing

(negatives cancel)

Just simplify with absolute value

★ holds in both cases



- Theorem: Let X have cdf $F_X(x)$ with range $\mathcal{X}$, $Y = g(X)$ with and range $\mathcal{Y}$ and inverse $h(y)$.
  - If $g(x)$ is strictly increasing on $\mathcal{X} \longrightarrow F_Y(y) = F_X(h(y))$ for $y \in \mathcal{Y}$.
  - If $g(x)$ is strictly decreasing on $\mathcal{X} \longrightarrow F_Y(y) = 1 - F_X(h(y))$ for $y \in \mathcal{Y}$.
  - If $g(x)$ is strictly increasing or strictly decreasing on $\mathcal{X}$, then

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

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

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

Cdf method

1 previously showed $F_Y(y) = 1 - e^{-3y^2}$

2 $f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}\left[1 - e^{-3y^2}\right]$

$= -e^{-3y^2}(-3(2)y)$

$= 6ye^{-3y^2}, \quad y \geq 0$

Pdf method

$\rightarrow x = h(y) = y^2 \rightarrow h'(y) = 2y$

$\rightarrow f_Y(y) = f_X(y^2)\,|h'(y)|$

$= 3e^{-3(y^2)}\,|2y|$

$= 6ye^{-3y^2}, \quad y \geq 0$

(ranges shown previously)

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

Cdf method

1 previously showed $F_Y(y) = e^{-3(1-y)}$

2 $f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}\left[e^{-3(1-y)}\right]$

$= e^{-3(1-y)}(3)$

$= 3e^{-3(1-y)}, \quad y \leq 1$

Pdf method

$\rightarrow x = h(y) = 1 - y \rightarrow h'(y) = -1$

$\rightarrow f_Y(y) = f_X(1-y)\,|h'(y)|$

$= 3e^{-3(1-y)}\,|-1|$

$= 3e^{-3(1-y)}, \quad y \leq 1$

(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)

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

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

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

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

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

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

↳ Discrete

$\rightarrow Y = g(X) = X^2$ is one-to-one over $\mathcal{X} = \{0, 1, 2, \ldots\}$ ✓

$y = x^2$

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

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