ENGG2430A Probability and Statistics for Engineers

Chapter 4: Further Topics on Random Variables

Instructor: Shengyu Zhang

Content

- Derived Distributions
- Covariance and Correlation
- Conditional Expectation and Variance Revisited
- Transforms
- Sum of a Random Number of Independent Random Variables

more advanced topics

- We introduce methods that are useful in:
  - deriving the distribution of a function of one or multiple random variables;
  - dealing with the sum of independent random variables, including the case where the number of random variables is itself random;
  - quantifying the degree of dependence between two random variables.

- We’ll introduce a number of tools
  - transforms
  - convolutions,

- We’ll refine our understanding of the concept of conditional expectation.

Derived distributions

- Consider functions $Y = g(X)$ of a continuous random variable X.
- *Question*: Given the PDF of X, how to calculate the PDF of Y?
  - Also called a derived distribution.
- The principal method for doing so is the following two-step approach.

Calculation of PDF of $Y = g(X)$

1. Calculate the CDF F_Y of Y using the formula

$$F_Y(y) = P(g(X) \leq y) = \int_{\{x | g(x) \leq y\}} f_X(x) dx$$

2. Differentiate F_Y to obtain the PDF f_Y of Y:

$$f_Y(y) = \frac{dF_Y}{dy}(y).$$

Example 1

- Let X be uniform on $[0,1]$, and let $Y = \sqrt{X}$.
- Note that for every $y \in [0,1]$, we have

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P\left(\sqrt{X} \le y\right) \\
&= P(X \ \le y^2) \\
&= y^2
\end{aligned}
$$

Example 1

- We then differentiate and obtain the following (for $0 \leq y \leq 1$)

$$f_Y(y) = \frac{dF_Y}{dy}(y) = \frac{d(y^2)}{dy} = 2y,$$

- Outside the range $[0,1]$:
  - The CDF $F_Y(y)$ is constant: $F_Y(y) = \begin{cases} 0 & \text{for } y \leq 0 \\ 1 & \text{for } y \geq 1 \end{cases}$
  - Thus by differentiating, $f_Y(y) = 0$ for $y \notin [0,1]$.

Example 2

- John Slow is driving from Boston to NYC, a distance of 180 miles.
- The driving speed is constant, whose value is uniformly distributed between 30 and 60 miles per hour.
- *Question: What is the PDF of the duration of the trip?*

Example 2

- Let X be the speed and let $Y = g(X)$ be the trip duration:

$$g(X) = \frac{180}{X}.$$

- To find the CDF of Y, we must calculate

$$P(Y \leq y) = P\left(\frac{180}{X} \leq y\right) = P\left(\frac{180}{y} \leq X\right).$$

Example 2

- We use the given uniform PDF of X, which is

$$f_X(x) = \begin{cases} 1/30 & \text{if } 30 \leq x \leq 60, \\ 0 & \text{otherwise.} \end{cases}$$

- and the corresponding CDF, which is

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 30, \\ (x - 30)/30, & \text{if } 30 \leq x \leq 60, \\ 1 & \text{if } 60 \leq x. \end{cases}$$

Example 2

- Thus,

$$F_Y(y) = P\left(\frac{180}{y} \leq X\right) = 1 - F_X\left(\frac{180}{y}\right)$$

$$= \begin{cases} 0 & \text{if } x \leq \frac{180}{60}, \\ 1 - \frac{\frac{180}{y} - 30}{30} & \text{if } \frac{180}{60} \leq y \leq \frac{180}{30}, \\ 1 & \text{if } \frac{180}{30} \leq y. \end{cases}$$

Example 2

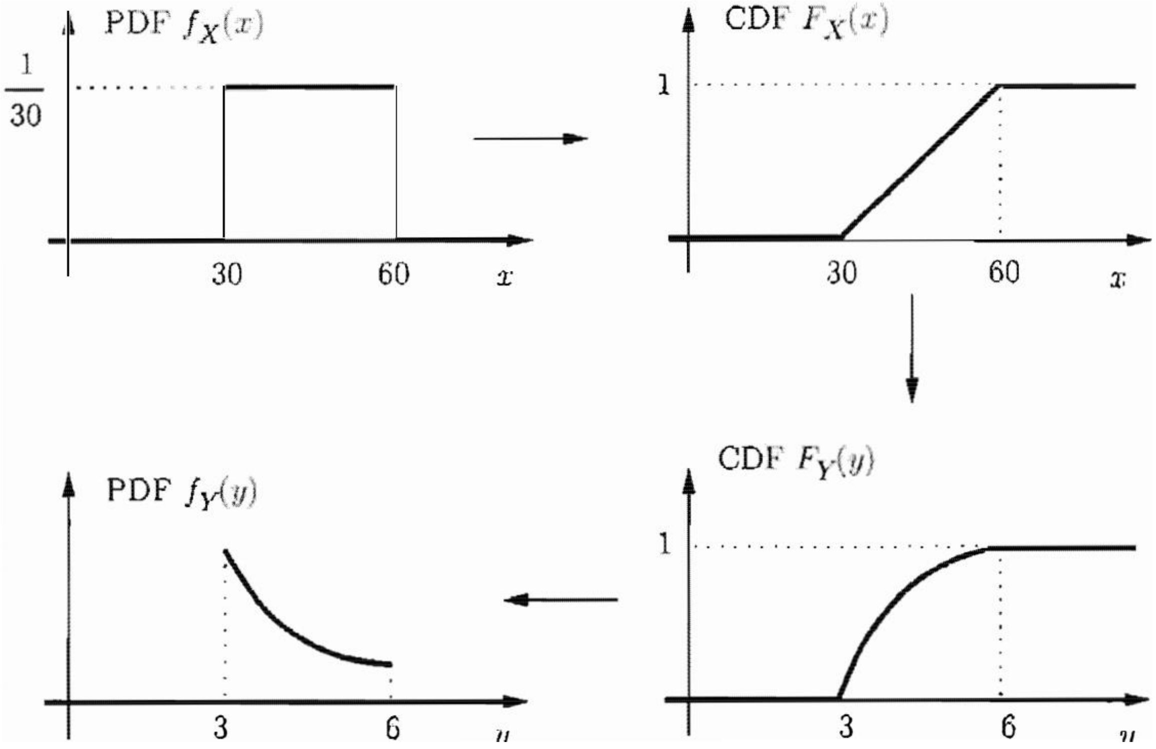
$$= \begin{cases} 0 & \text{if } y \leq 3, \\ 2 - (6/y) & \text{if } 3 \leq y \leq 6, \\ 1 & \text{if } 6 \leq y. \end{cases}$$

- Differentiating this expression, we obtain the PDF of Y:

$$f_Y(y) = \begin{cases} 0 & \text{if } y \leq 3, \\ 6/y^2 & \text{if } 3 \leq y \leq 6, \\ 0 & \text{if } 6 \leq y. \end{cases}$$

Illustration of the whole process

- The arrows indicate the flow of the calculation.



Example 3

- Let $Y = g(X) = X^2$, where X is a random variable with known PDF.
- For any $y \geq 0$, we have

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

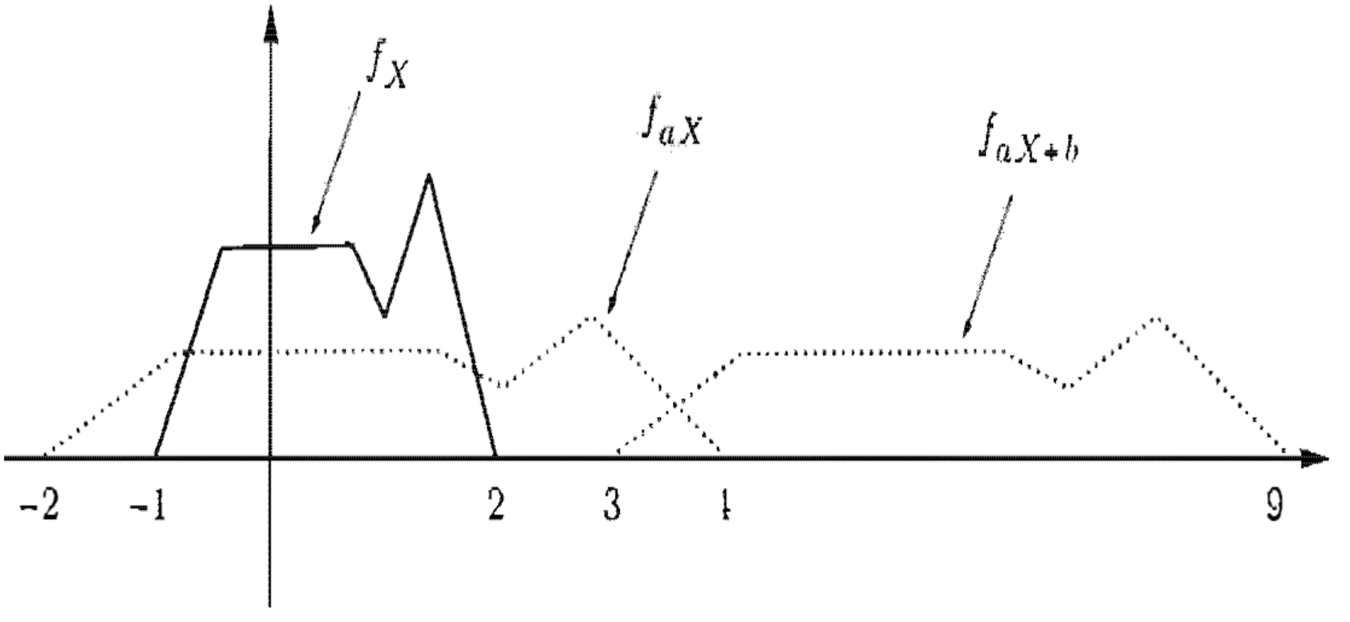
Example 3

- And therefore, by differentiating and using the chain rule,

$$f_Y = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}), y \geq 0.$$

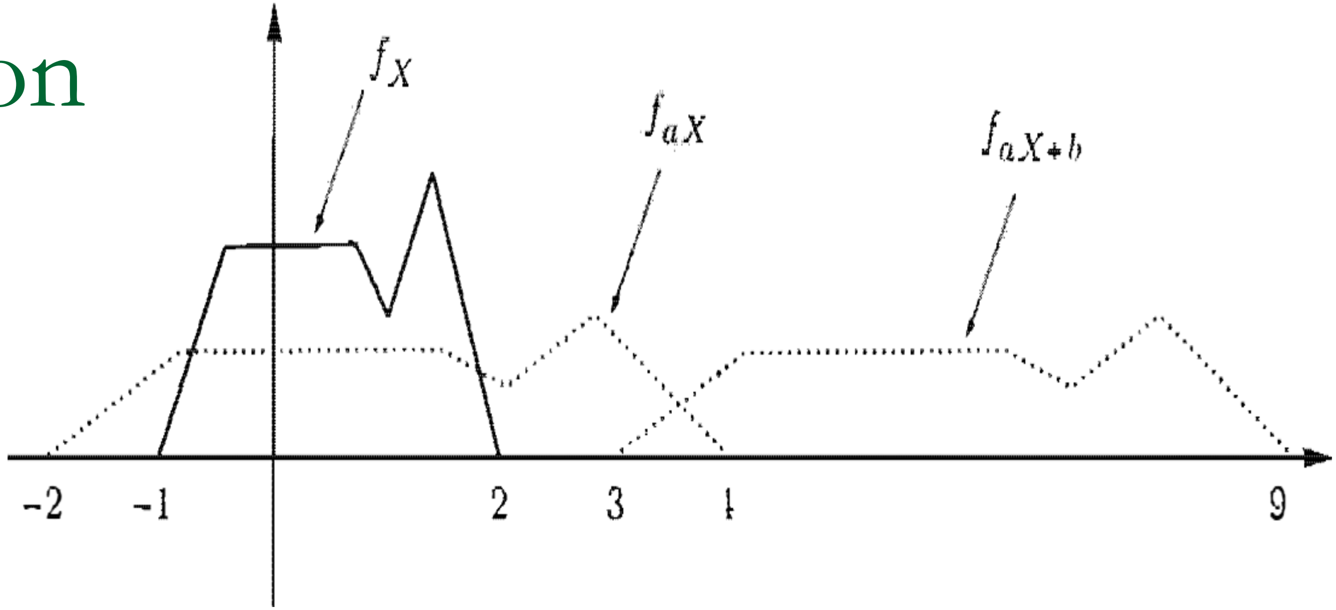
The Linear Case

- The PDF of $aX + b$ in terms of the PDF X.



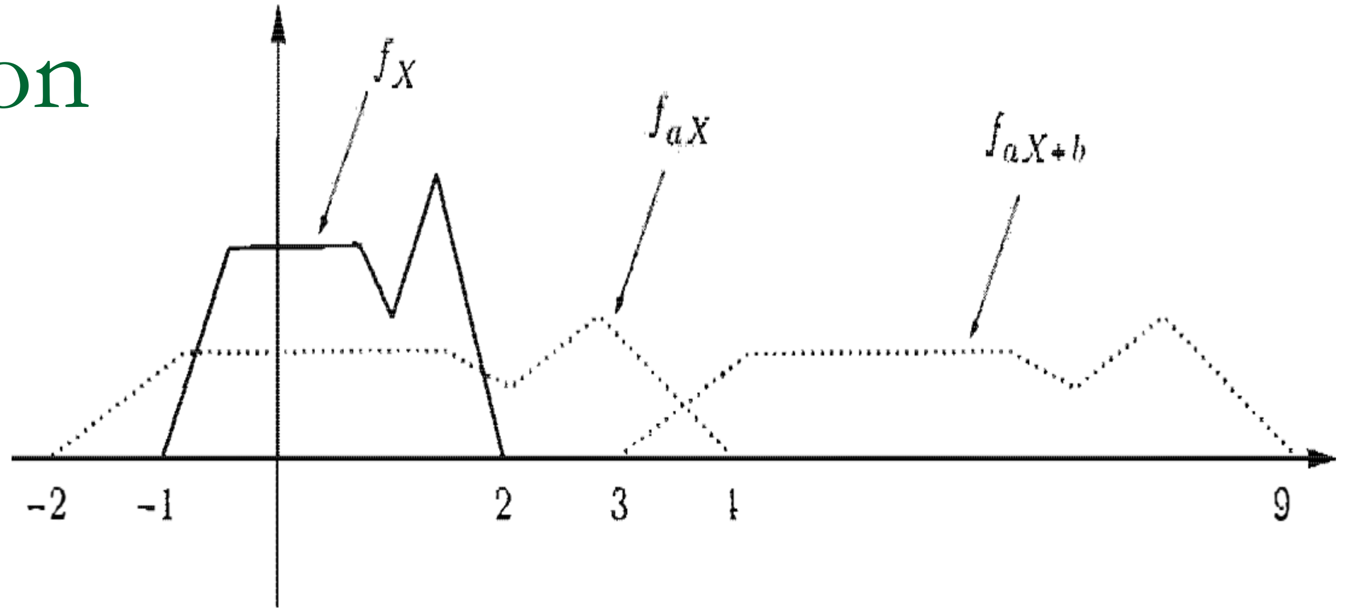
In this figure: $a = 2$ and $b = 5$.

Intuition



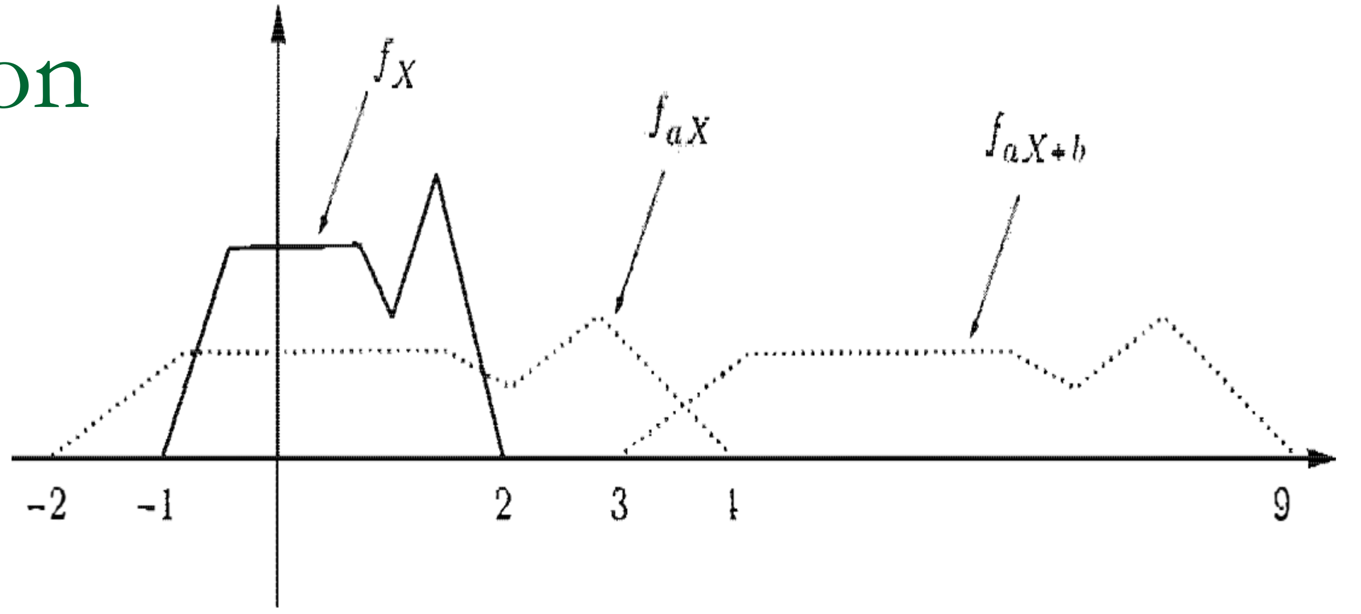
- As a first step, we obtain the PDF of aX. The range of Y is wider than the range of X, by a factor of a.
- Thus, the PDF must be stretched (scaled horizontally) by the factor.

Intuition



- But in order to keep the total area under the PDF (vertically) by the same factor a, we need to scale down the PDF (vertically) by the same factor a.

Intuition



- How about b?
- The random variable $aX + b$ is the same as aX except that its values are shifted by b.
- Accordingly, we take the PDF of aX and shift it (horizontally) by b.

Intuition

- The end result of these operations is the PDF of $Y = aX + b$ and is given mathematically by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

- Next: formal verification of this formula.

Formal calculation

- Let X be a continuous random variable with PDF f_X, and let
$$Y = aX + b,$$
where a and b are scalars, with $a \neq 0$.

- Then,
$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Formal calculation

- First calculate the CDF of Y, and differentiate.
- (1) Calculate the CDF , we have

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(aX + b \le y) \\
&= P\left(X \le \frac{y-b}{a}\right) \\
&= F_X\left(\frac{y-b}{a}\right)
\end{aligned}
$$

Formal calculation

- (2) Differentiate this equality and use the chain rule, to obtain

$$f_Y(y) = \frac{dF_Y}{dy}(y)$$
$$= \frac{1}{a} f_X\left(\frac{y-b}{a}\right).$$

- We only show the steps for the case where $a > 0$; the case $a < 0$ is similar.

Example 4. Linear of Exponential

- Suppose that X is an exponential random variable with PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

where λ is a positive parameter.

Example 4. Linear of Exponential

- Let $Y = aX + b$. Then,

$$f_Y(y) = \begin{cases} \frac{\lambda}{|a|} e^{-\lambda(y-b)/a}, & \text{if } (y-b)/a \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

- If $b = 0$ and $a > 0$, then Y is an exponential random variable with parameter λ/a.

Example 4. Linear of Exponential

- Let $Y = aX + b$. Then,

$$f_Y(y) = \begin{cases} \frac{\lambda}{|a|} e^{-\lambda(y-b)/a}, & \text{if } (y-b)/a \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

- In general, however, Y need not be exponential.
- For example, if $a < 0$ and $b = 0$, then the range of Y is the negative real axis.
  - Consider $Y = -X$.

Example 5. Linear of Normal

- Suppose that X is a normal random variable with mean μ and variance σ^2, and let $Y = aX + b$, where a and b are scalars, with $a \neq 0$.
- We have

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Example 5. Linear of Normal

- Therefore,

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

$$= \frac{1}{|a|}\frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}\right\}$$

$$= \frac{1}{\sqrt{2\pi}|a|\sigma} exp\left\{-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}\right\}$$

Example 5. Linear of Normal

$$f_Y(y) = \frac{1}{\sqrt{2\pi}|a|\sigma} exp\left\{-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}\right\}$$

- We recognize this as a normal PDF with mean $a\mu + b$ and variance $a^2\sigma^2$.
- In particular, Y is a normal random variable.

The Monotonic Case

- The calculation and the formula for the linear case can be generalized to the case where g is a monotonic function.

- Let X be a continuous random variable and suppose that its range is contained in a certain interval I,
  - $f_X(x) = 0, \forall x \in I$.

The Monotonic Case (cont.)

- We consider the random variable $Y = g(X)$, and assume that g is strictly monotonic over the interval I, so that either
  - (monotonically increasing case) $g(x) < g(x')$ for all $x, x' \in I$ satisfying $x < x'$, or
  - (monotonically decreasing case) $g(x) > g(x')$ for all $x, x' \in I$ satisfying $x < x'$.

The Monotonic Case (cont.)

- Furthermore, we assume that the function g is differentiable.

- Its derivative will necessarily be
  - nonnegative in the increasing case
  - nonpositive in the decreasing case

The Monotonic Case (cont.)

- An important fact is that a strictly monotonic function can be "inverted".
- There is some function h, called the inverse of g, such that for all $x \in I$, we have

$$y = g(x) \text{ if and only if } x = h(y).$$

The Monotonic Case (cont.)

- In Example 2, the inverse of the function $g(x) = 180/x$ is $h(y) = 180/y$, because we have $y = 180/x$ if and only if $x = 180/y$.
- Other such examples of pairs of inverse functions include

$$g(x) = \ ax + b, \qquad h(y) = \frac{y - b}{a}.$$

where a and b are scalars with $a \neq 0$

The Monotonic Case (cont.)

and

$$g(x) = e^{ax}, \qquad h(y) = \frac{\ln y}{a}.$$

where a is a nonzero scalar.

- For strictly monotonic functions g, the following is a convenient analytical formula for the PDF of the function $Y = g(X)$.

The monotonic case (cont.)

- Suppose that g is strictly monotonic and that for some function h and all x in the range of X we have

$$y = g(x) \text{ if and only if } x = h(y).$$

- Assume that h is differentiable.
- Then, the PDF of Y in the region where $f_Y(y) > 0$ is given by

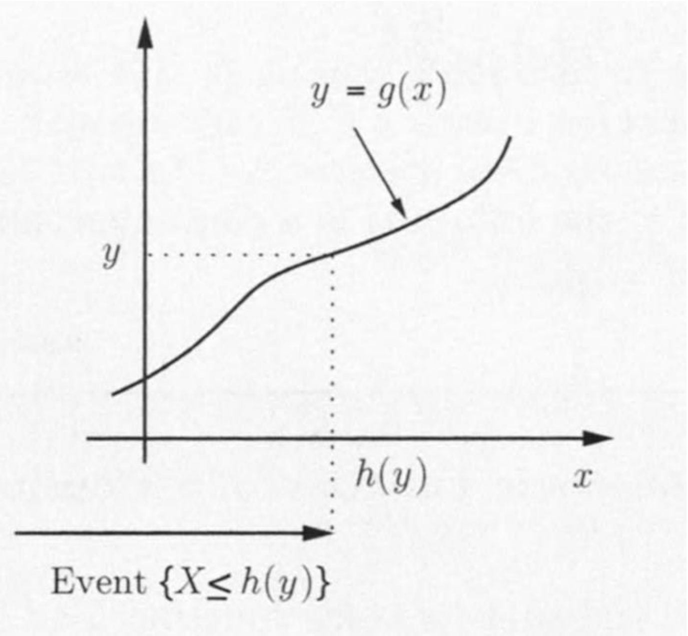
$$f_Y(y) = f_X(h(y)) \left| \frac{dh}{dy}(y) \right|$$

Verification

- Assume that g increasing.
- We have

$$F_Y(y) = P(g(X) \le y) = P(X \le h(y)) = F_X(h(y)),$$

  - The second equality can be justified the monotonically increasing property of g.



- Last slide: $F_Y(y) = F_X(h(y))$
- By differentiating this relation, using also the chain rule, we obtain

$$f_Y(y) = \frac{dF_Y}{dy}(y)$$

$$= f_X\left(h(y)\right) \frac{dh}{dy}(y)$$

- Because g is monotonically increasing, h is also monotonically increasing, so its derivative is nonnegative:

$$\frac{dh}{dy}(y) = \left|\frac{dh}{dy}(y)\right|$$

- This justifies the PDF formula for a monotonically increasing function g.

- The case of monotonically decreasing function is similar.
- We differentiate instead the relation

$$\begin{aligned} F_Y(y) &= P(g(X) \le y) \\ &= P(X \ge h(y)) \\ &= 1 - F_X(h(y)), \end{aligned}$$

- When g is decreasing, the event $\{g(X) \le y\}$ is the same as the event $\{X \ge h(y)\}$.

Example: quadratic function revisited

- Let $Y = g(X) = X^2$, where X is a continuous uniform random variable on $(0, 1]$.
- g is strictly monotonic within this interval.
- Its inverse is $h(y) = \sqrt{y}$.
- Thus, for any $y \in (0,1]$, we have

$$f_X(\sqrt{y}) = 1. \qquad \left|\frac{dh}{dy}(y)\right| = \frac{1}{2\sqrt{y}}.$$

Example: quadratic function revisited

- $f_X(\sqrt{y}) = 1.$ $\left|\frac{dh}{dy}(y)\right| = \frac{1}{2\sqrt{y}}$

- Thus

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}, & \text{if } y \in (0,1], \\ 0, & \text{otherwise.} \end{cases}$$

Functions of more random variables

- Consider now functions of 2 or more r.v.
- Recall the two-step procedure for one r.v.
  1. calculates the CDF
  2. differentiates to obtain the PDF.
- This applies to the case with ≥ 2 r.v. as well.

Example: archer shooting

- Two archers shoot at a target.
- The distance of each shot from the center of the target is uniformly distributed from 0 to 1, independent of the other shot.
- *Question: What is the PDF of the distance of the losing shot from the center?*

Example: archer shooting

- Let X and Y be the distances from the center of the first and second shots, respectively.
- Z: the distance of the losing shot:
$$Z = \max\{X, Y\}.$$
- Since X and Y are uniformly distributed over $[0,1]$,
- we have, for all $z \in [0, 1]$,
$$P(X \leq z) = P(Y \leq z) = z.$$

Example: archer shooting

- Thus, using the independence of X and Y, we have for all $z \in [0,1]$,

$$\begin{aligned} F_Z(z) &= P(\max\{X, Y\} \leq z) \\ &= P(X \leq z, Y \leq z) \\ &= P(X \leq z)P(Y \leq z) \\ &= z^2. \end{aligned}$$

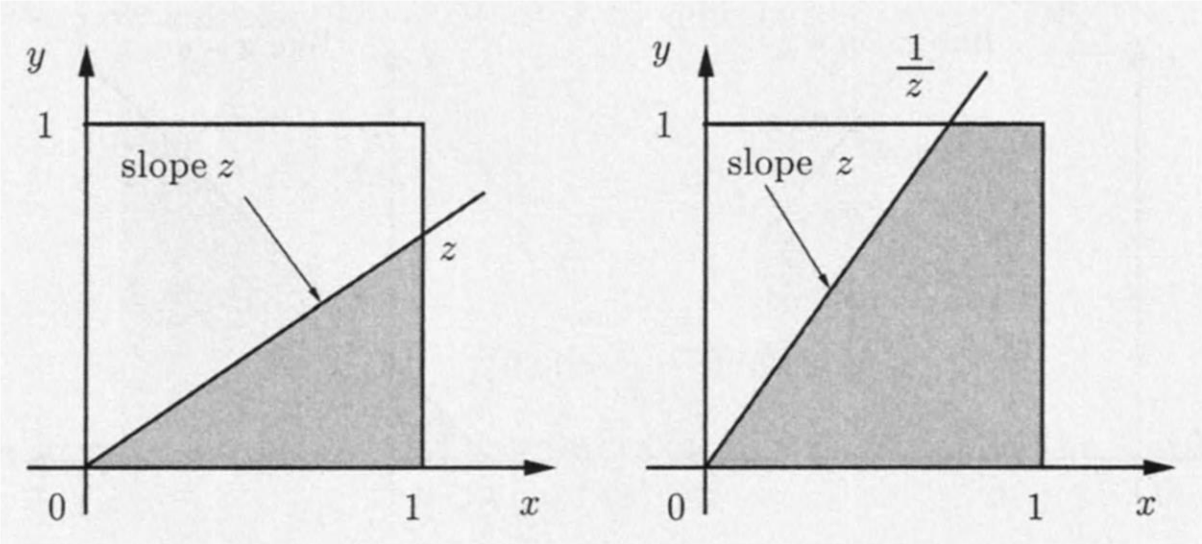
- Differentiating, we obtain

$$f_Z(z) = \begin{cases} 2z, & \text{if } 0 \leq z \leq 1. \\ 0, & \text{otherwise.} \end{cases}$$

Example: Y/X

- Let X and Y be independent random variables that are uniformly distributed on the interval $[0,1]$.

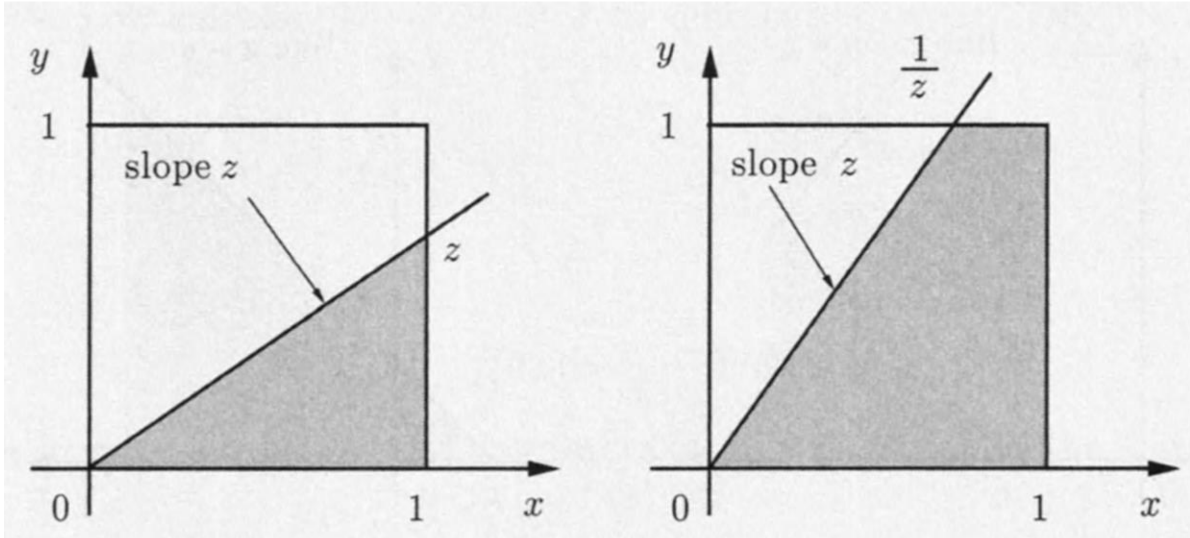
- *Question: What is the PDF of the random variable* $Z = Y/X$?



- The value $P(Y/X \leq z)$ is equal to shaded subarea of unit square.
  - The figure on the left deals with the case where $0 \leq z \leq 1$.
  - The figure on the right refers to the case where $z > 1$.

- We will find the PDF of Z by first finding its CDF and then differentiating.
- We consider separately the case $0 \leq z \leq 1$ and $z > 1$.

- We will find the PDF of Z by first finding its CDF and then differentiating.
- We consider separately the case $0 \leq z \leq 1$ and $z > 1$.



$$F_Z(z) = P(Y/X \le z)$$

$$= \begin{cases} z/2, & \text{if } 0 \le z \le 1, \\ 1 - 1/2z, & \text{if } z > 1, \\ 0, & \text{otherwise} \end{cases}$$

Example: Y/X

$$F_Z(z) = P(Y/X \leq z)$$
$$= \begin{cases} z/2, & \text{if } 0 \leq z \leq 1, \\ 1 - 1/2z, & \text{if } z > 1, \\ 0, & \text{otherwise} \end{cases}$$

- By differentiating, we obtain the pdf of Z:

$$f_Z(z) = \begin{cases} 1/2, & \text{if } 0 \leq z \leq 1, \\ 1/(2z^2), & \text{if } z > 1, \\ 0, & \text{otherwise} \end{cases}$$

Example: Romeo and Juliet

- Romeo and Juliet have a date at a given time, and each, independently, will be late by an amount of time that is exponentially distributed with parameter λ.

- *Question: What is the PDF of difference between their times of arrival?*

Example: $X - Y$

- We denote by X and Y the amounts by which Romeo and Juliet are late, respectively.
- We want to find the PDF of $Z = X - Y$, assuming that X and Y are independent and exponentially distributed with parameter λ.
- We will first calculate the CDF $F_Z(z)$ by considering separately the cases $z \geq 0$ and $z < 0$.

For $z \geq 0$

- $F_Z(z) = P(X - Y \leq z)$
$$= 1 - P(X - Y > z)$$
$$= 1 - \int_0^\infty \left(\int_{z+y}^\infty f_{X,Y}(x,y)\,dx\right)dy$$
$$= 1 - \int_0^\infty \lambda e^{-\lambda y}\left(\int_{z+y}^\infty \lambda e^{-\lambda x}\,dx\right)dy$$
$$= 1 - \int_0^\infty \lambda e^{-\lambda y} e^{-\lambda(z+y)}dy$$
$$= 1 - e^{-\lambda z}\int_0^\infty \lambda e^{-2\lambda y}dy$$
$$= 1 - \frac{1}{2}e^{-\lambda z} \qquad // \int_0^\infty 2\lambda e^{-2\lambda y}dy = 1.$$

$z < 0$

- For the case $z < 0$, we can use a similar calculation, but we can also argue using symmetry.
- Indeed, the symmetry of the situation implies that the random variables $Z = X - Y$ and $-Z = Y - X$ have the same distribution.

$z < 0$

- Thus we have $F_Z(z) = 1 - F_Z(-z)$.
- Recall when $z > 0$: $F_Z(z) = 1 - \frac{1}{2}e^{-\lambda z}$
- Thus for $z < 0$:

$$\begin{aligned} F_Z(z) &= 1 - F_Z(-z) \\ &= 1 - \left(1 - \frac{1}{2}e^{-\lambda(-z)}\right) \quad // \ -z > 0 \\ &= \frac{1}{2}e^{\lambda z} \end{aligned}$$

together

- Combining the two cases $z \geq 0$ and $z < 0$:

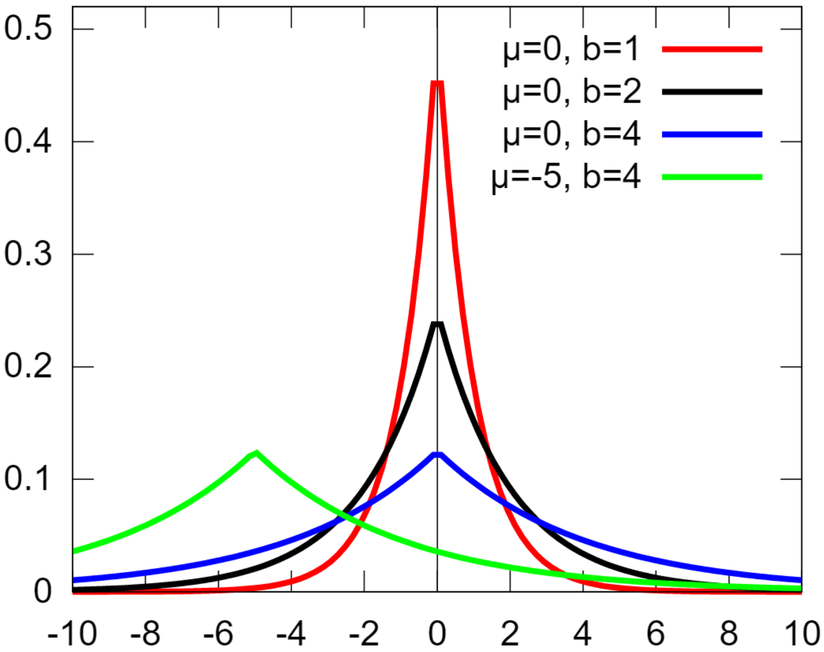
$$F_Z(z) = \begin{cases} 1 - \frac{1}{2}e^{-\lambda z}, & \text{if } z \geq 0, \\ \frac{1}{2}e^{\lambda z}, & \text{if } z < 0. \end{cases}$$

- Differentiating:

$$f_Z(z) = \begin{cases} (\lambda/2)e^{-\lambda z}, & \text{if } z \geq 0, \\ (\lambda/2)e^{\lambda z}, & \text{if } z < 0. \end{cases}$$

together

- This is known as a two-sided exponential PDF,
- Also called the Laplace PDF.
  - $f(x|\mu, b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$.



(see wiki page)

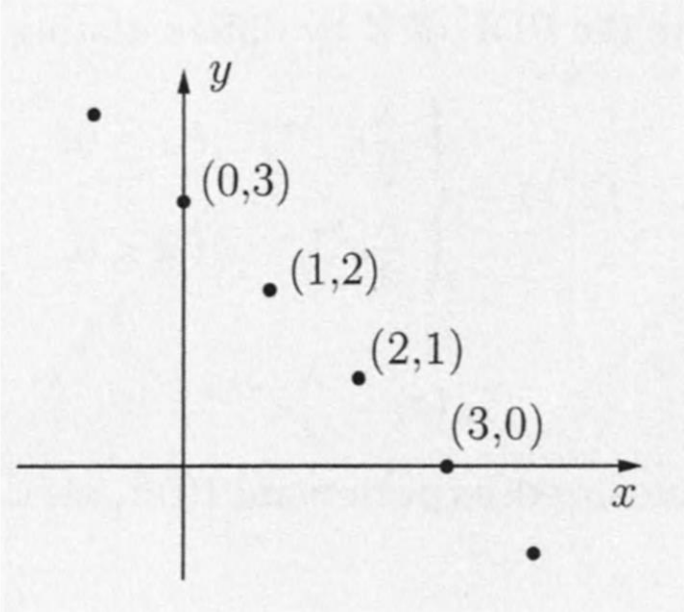
Convolution

- We now consider an important example of a function Z of two random variables, namely, the case where $Z = X + Y$, for independent X and Y.
- For some initial insight, we start by deriving a PMF formula for the case where X and Y are discrete.

- Let $Z = X + Y$, where X and Y are independent integer-valued random variables with PMFs P_X and P_Y, respectively.
- Then, for any integer z,

$$p_Z(z) = P(X + Y = z)$$
$$= \sum_{\{(x,y)|x+y=z\}} P(X = x, Y = y)$$
$$= \sum_x P(X = x, Y = z - x)$$
$$= \sum_x p_X(x) p_Y(z - x)$$

- The resulting PMF p_Z is called the convolution of the PMFs of X and Y.



- Suppose now that X and Y are independent continuous random variables with PDFs f_x and f_y, respectively.

- We wish to find the PDF of $Z = X + Y$.
- Two steps:
  - find the joint PDF of X and Z
  - integrate to find the PDF of Z.

- We first note that

$$P(Z \le z | X = x) = P(X + Y \le z | X = x)$$
$$= P(x + Y \le z | X = x)$$
$$= P(x + Y \le z)$$
$$= P(Y \le z - x)$$

  - The third equality follows from the independence of X and Y.

- $P(Z \leq z|X = x) = P(Y \leq z - x)$
- By differentiating both sides with respect to z, we see that $f_{Z|X}(z|x) = f_Y(z - x)$.
- Using the multiplication rule, we have

$$
\begin{aligned}
f_{X,Z}(x, z) &= f_X(x) f_{Z|X}(z|x) \\
&= f_X(x) f_Y(z - x)
\end{aligned}
$$

- $f_{X,Z}(x,z) = f_X(x) f_Y(z-x)$
- Thus $f_Z(z) = \int_{-\infty}^{\infty} f_{X,Z}(x,z)\, dx$

  $$= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$$
- The formula is entirely analogous to the one for the discrete case
  - $p_Z(z) = \sum_x p_X(x) p_Y(z-x)$
  - Except the summation is replaced by an integral and the PMFs are replaced by PDFs.

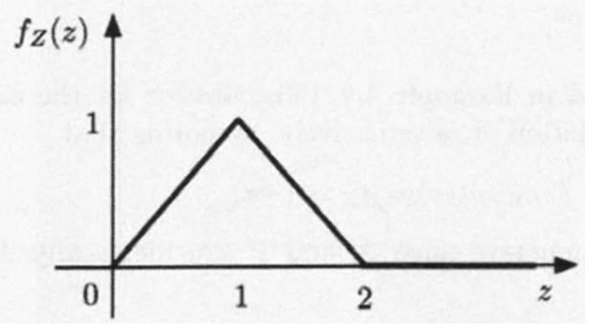
Example: convolution

- The random variables X and Y are independent and uniformly distributed in the interval $[0,1]$.
- The PDF of $Z = X + Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\, dx$$

Example: convolution

- The integrand $f_X(x)f_Y(z-x)$ is nonzero (and equal to 1) for $0 \leq x \leq 1$ and $0 \leq z - x \leq 1$.
- Combining these two inequalities, the integrand is 1 for $\max\{0, z-1\} \leq x \leq \min\{1, z\}$
  - and 0 otherwise.
- $f_Z(z) = \begin{cases} \min\{1, z\} - \max\{0, z-1\}, & 0 \leq z \leq 2, \\ 0, & \text{otherwise,} \end{cases}$



Example: sum of normals

- *Message: The Sum of Two Independent Normal Random Variables in Normal.*
- Let X and Y be independent normal random variables with means μ_x, μ_y, and variances σ_x^2, σ_y^2, respectively, and let $Z = X + Y$.
- $$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_x} exp\left\{-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right\} \frac{1}{\sqrt{2\pi}\,\sigma_y} exp\left\{-\frac{\left(z-x-\mu_y\right)^2}{2\sigma_y^2}\right\}$$

Example: sum of normals

- This integral can be evaluated in closed form, but the details are tedious and are omitted.
- Answer turns out to be

$$f_Z(z) = \frac{1}{\sqrt{2\pi(\sigma_x^2+\sigma_y^2)}} exp\left\{-\frac{\left(z-\mu_x-\mu_y\right)^2}{2(\sigma_x^2+\sigma_y^2)}\right\}$$

- It's a normal PDF with mean $\mu_x+\mu_y$ and variance $\sigma_x^2 + \sigma_y^2$.

Example: sum of normals

- We therefore reach the conclusion that the sum of two independent normal random variables is normal.

- Given that scalar multiples of normal random variables are also normal, it follows that $aX + bY$ is also normal, for any nonzero a and b.

Example: $X - Y$

- The convolution formula can also be used to find the PDF of $X - Y$, when X and Y are independent, by viewing $X - Y$ as the sum of X and $-Y$.
- Note: the PDF of $-Y$ is given by $f_{-Y}(y) = f_Y(-y)$.
- Thus $f_{X-Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_{-Y}(z - x) dx$

$$= \int_{-\infty}^{\infty} f_X(x) f_Y(x - z) dx$$

Example: $X - Y$

- When applying the convolution formula, often the most delicate step was to determine the correct limits for the integration.

- This is often tedious and error prone, but can be bypassed using a graphical method.

Graphical Calculation of Convolutions

- We use a dummy variable t as the argument of the different functions involved in this discussion.
- Consider two PDFs $f_X(t)$ and $f_Y(t)$. For a fixed value of z, the graphical evaluation of the convolution

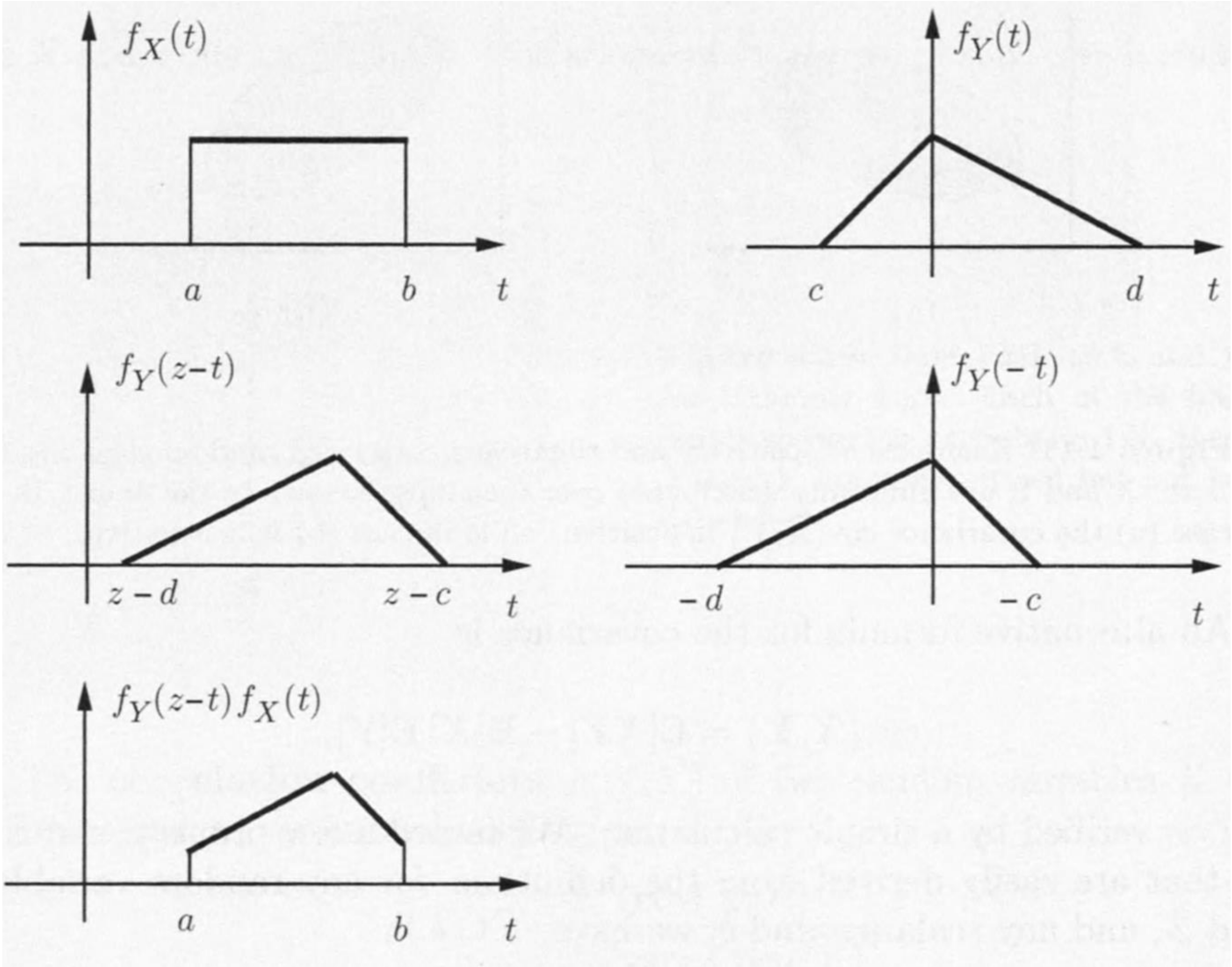
$$f_Z(z) = \int_{-\infty}^{\infty} f_X(t) f_Y(z-t)\,dt$$

consists of the following steps:

- We plot $f_Y(z-t)$ as a function of t.
  - This plot has the same shape as the plot of $f_Y(t)$ except that it is first “flipped” and then shifted by an amount z.
  - If $z > 0$, this is a shift to the right, if $z < 0$, this is a shift to the left.
- We place the plots of $f_X(t)$ and $f_Y(z-t)$ on top of each other, and form their product.

- Calculate the value of $f_Z(z)$ by calculating the integral of the product graph.

- By varying the amount z by which we are shifting, we obtain $f_Z(z)$ for any z.



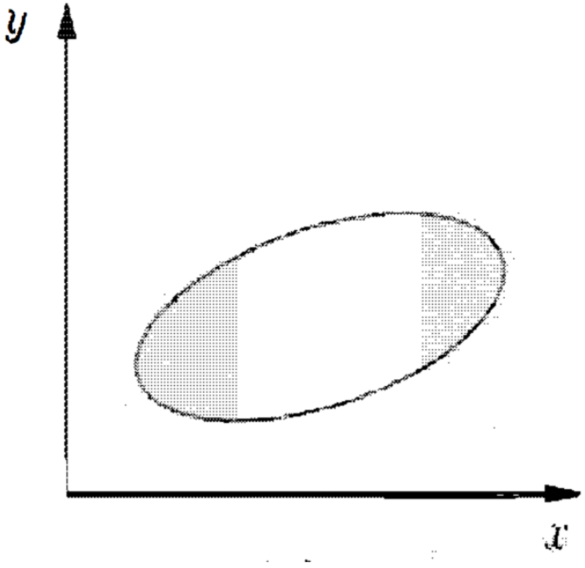
- $F(z)$ = integral of function shown in the last plot.

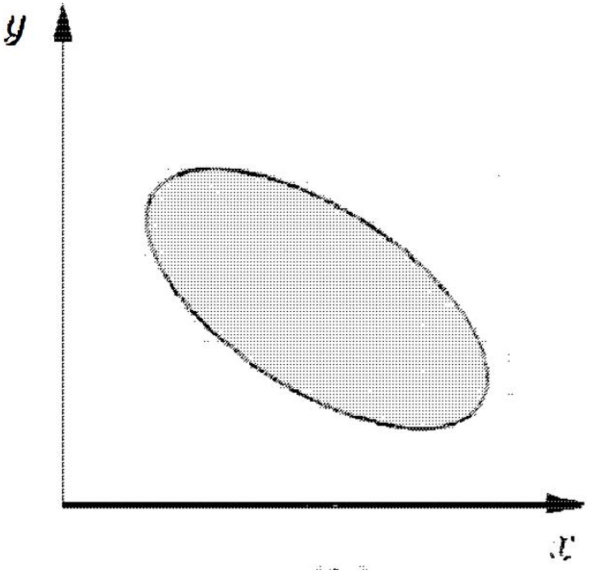
Content

- Derived Distributions
- Covariance and Correlation
- Conditional Expectation and Variance Revisited
- Transforms
- Sum of a Random Number of Independent Random Variables

Covariance and Correlation

- Covariance and correlation – the measurement of the strength and direction of the relation between 2 random variables.





Covariance

- The covariance of two random variables X and Y are defined as

$$\mathrm{cov}(X,Y) = \mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])].$$

- Alternatively

$$\mathrm{cov}(X,Y) = \mathbf{E}[XY]-\mathbf{E}[X]\mathbf{E}[Y].$$

  - Exercise: verify that the two definitions are equivalent.

Covariance - Properties

- For any random variable X, Y, and Z, and any scalars a and b:
- $\mathrm{cov}(X, X) = \mathrm{var}(X)$,
- $\mathrm{cov}(X, aY + b) = a \cdot \mathrm{cov}(X, Y)$,
- $\mathrm{cov}(X, Y + Z) = \mathrm{cov}(X, Y) + \mathrm{cov}(X, Z)$.

- Exercise: verify these.

Covariance - Properties

- Independent random variables are uncorrelated. In fact,

$$\mathrm{cov}(X,Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] = 0.$$

- But not vice versa, as illustrated by the next example.

Covariance - Example

- (X, Y) is uniformly distributed over $\{(1,0), (0,1), (-1,0), (0,-1)\}$, then

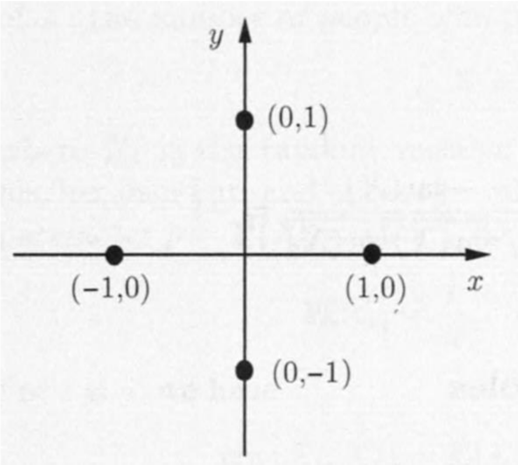
$$\mathrm{E}[XY] = 0$$

since either X or Y is 0.

- Also $\mathbf{E}[X] = \mathbf{E}[Y] = 0$. Thus

$$\mathrm{cov}(X, Y) = 0.$$

- But X and Y are not independent. For example, $X \neq 0 \Rightarrow Y = 0$.



Correlation Coefficient

- For any random variable X, Y with nonzero variances, the correlation coefficient $\rho(X, Y)$ of them is defined as

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}.$$

- It may be viewed as a normalized version of the covariance $\text{cov}(X, Y)$.
  - Recall $\text{cov}(X, X) = \text{var}(X)$.
  - It's easily verified that

$$-1 \leq \rho(X,Y) \leq 1$$

Correlation coefficient: Properties

- $\rho(X,Y) = 1$ iff $\exists$ a positive number c s.t.

$$X - \mathbf{E}[X] = c(Y - \mathbf{E}[Y]).$$

- If $\rho(X,Y) > 0$, then the values of $X - \mathbf{E}[X]$ and $Y - \mathbf{E}[Y]$ "tend" to have the same sign.
- $\rho(X,Y) = -1$ iff $\exists$ a negative number c s.t.

$$X - \mathbf{E}[X] = c(Y - \mathbf{E}[Y]).$$

- If $\rho(X,Y) < 0$, then the values of $X - \mathbf{E}[X]$ and $Y - \mathbf{E}[Y]$ "tend" to have the opposite sign.

Correlation coefficient: Examples

- Consider n independent tosses, with head probability p.
- $X =$ number of heads
- $Y =$ number of tails
- Then $X + Y = n$, and thus $\mathrm{var}(Y) = \mathrm{var}(X)$,
$$\mathbf{E}[X] + \mathbf{E}[Y] = \mathbf{E}[X + Y] = n = X + Y.$$
- Hence,
$$X - \mathbf{E}[X] = -(Y - \mathbf{E}[Y])$$

$\rho(X,Y)$ Examples

- Las slide: $X - \mathbf{E}[X] = -(Y - \mathbf{E}[Y])$
- Then,

$$\begin{aligned} \text{cov}(X,Y) &= \mathbf{E}[(X - \mathrm{E}[X])(Y - \mathbf{E}[Y])] \\ &= -\mathbf{E}[(X - \mathbf{E}[X])^2] \\ &= -\text{var}(X). \end{aligned}$$

- Hence,

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{-\text{var}(X)}{\sqrt{\text{var}(X)\text{var}(X)}} = -1.$$

Variance of Summations

- We know that in general

$$\mathrm{var}(X_1 + X_2) \neq \mathrm{var}(X_1) + \mathrm{var}(X_2)$$

- A more precise statement:

$$\mathrm{var}(X_1 + X_2) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + 2 \cdot \mathrm{cov}(X_1, X_2)$$

- In general, let $X_1, X_2, \cdots, X_n$ be random variables with finite variance, then we have

$$\mathrm{var}(\sum_i X_i) = \sum_i \mathrm{var}(X_i) + \sum_{i \neq j} \mathrm{cov}(X_i, X_j)$$

Variance of Summations

- Proof. Let $\tilde{X}_i = X_i - \mathbf{E}[X_i]$,

$$\operatorname{var}\left(\sum_i X_i\right) = \mathbf{E}\left[\left(\sum_i \tilde{X}_i\right)^2\right]$$

$$= \sum_i \sum_j \mathbf{E}[\tilde{X}_i \tilde{X}_j]$$

$$= \sum_i \mathbf{E}[\tilde{X}_i^2] + \sum_{i \neq j} \mathbf{E}[\tilde{X}_i \tilde{X}_j]$$

$$= \sum_i \operatorname{var}(X_i) + \sum_{i \neq j} \operatorname{cov}(X_i, X_i) .$$

Variance of Summations – Example

- Consider n people throwing their hats in a box and pick a hat at random.
- $X =$ number of people who pick their own hats.
- $X_i = \begin{cases} 1 & \text{if the } i\text{−th person picks its own hat,} \\ 0 & \text{otherwise.} \end{cases}$
- Then $X = X_1 + X_2 + \cdots + X_n$.
- And $\mathbf{E}[X_i] = \frac{1}{n}$.

Variance of Summations – Example

- For $i \neq j$, we have

$$\begin{aligned}\operatorname{cov}\left(X_i, X_j\right) &= \mathrm{E}\left[X_i X_j\right] - \mathrm{E}[X_i]\mathrm{E}\left[X_j\right] \\ &= \frac{1}{n(n-1)} - \frac{1}{n^2} \\ &= \frac{1}{n^2(n-1)}.\end{aligned}$$

- Also

$$\operatorname{var}(X_i) = \frac{1}{n}\left(1 - \frac{1}{n}\right).$$

Variance of Summations – Example

- Recall

$$\mathrm{var}\left(\sum_i X_i\right) = \sum_i \mathrm{var}(X_i) + \sum_{i \neq j} \mathrm{cov}(X_i, X_i)\,.$$

- We have

$$\mathrm{var}(X) = n\frac{1}{n}\left(1 - \frac{1}{n}\right) + \frac{n(n-1)}{n^2(n-1)} = 1.$$

Content

- Derived Distributions
- Covariance and Correlation
- Conditional Expectation and Variance Revisited
- Transforms
- Sum of a Random Number of Independent Random Variables

- We revisit the conditional expectation of a random variable X given another random variable Y.
- We introduced a random variable, denoted by $\mathbf{E}[X|Y]$, that takes value $\mathbf{E}[X|Y = y]$ when Y takes the value y.
- Since $\mathbf{E}[X|Y = y]$ is a function of y, $\mathbf{E}[X|Y]$ is a function of Y.

Example: coin

- A biased coin.
- Y = the probability of heads
- Y is itself random, with a known distribution over the interval [0,1].
- We toss the coin n times.
- X = the number of heads obtained.
- Then, for any $y \in [0,1]$, we have $\mathbf{E}[X|Y = y] = ny$
- so $\mathbf{E}[X|Y]$ is the random variable nY.

- Since $\mathbf{E}[X|Y]$ is a random variable, it has an expectation $\mathbf{E}\big[\mathbf{E}[X|Y]\big]$ of its own,
- which can be calculated using the expected value rule:

$$\mathbf{E}\big[\mathbf{E}[X|Y]\big] = \begin{cases} \sum_y \mathbf{E}[X|Y=y]p_Y(y) & Y \text{ discrete} \\ \int_{-\infty}^{\infty} \mathbf{E}[X|Y=y]f_Y(y)dy & Y \text{ continuous} \end{cases}$$

- By total probability theorem, RHS $= \mathbf{E}[X]$.
- Law of Iterated Expectations:

$$\mathbf{E}\big[\mathbf{E}[X|Y]\big] = \mathbf{E}[X].$$

Example: coin

- Y = the probability of heads for our coin
- Y is uniformly distributed over the interval $[0,1]$.
- Since $\mathbf{E}[X|Y] = nY$ and $\mathbf{E}[Y] = 1/2$.
- By the law of iterated expectations, we have

$$\mathbf{E}[X] = \mathbf{E}\big[\mathbf{E}[X|Y]\big] = \mathbf{E}[nY] = n\mathbf{E}[Y] = \frac{n}{2}$$

Example: stick breaking

- We start with a stick of length ℓ.
- Break it at a point which is chosen randomly and uniformly over its length,
- Keep the left piece.
- Repeat the same process on this piece.
- *Question: What is the expected length of the piece that we are left with after breaking twice?*

- $Y =$ the length of the piece after we break for the first time.
- $X =$ the length after we break for the second time.
- We have $\mathbf{E}[X|Y] = Y/2$,
  - since the breakpoint is chosen uniformly over a piece of length Y.
- For a similar reason, $\mathbf{E}[Y] = \ell/2$.
- $\therefore\ \mathbf{E}[X] = \mathbf{E}\big[\mathbf{E}[X|Y]\big] = \mathbf{E}[Y/2] = \mathbf{E}[Y]/2 = \ell/4$.

Content

- Derived Distributions
- Covariance and Correlation
- Conditional Expectation and Variance Revisited
- Transforms
- Sum of a Random Number of Independent Random Variables

Transforms

- We introduce the transform associated with a random variable.
- The transform provides us with an alternative representation of a probability law.
- It's not particularly intuitive, but it is often convenient for certain types of mathematical manipulations.

Transforms

- The transform associated with a random variable X, also referred to as the associated moment generating function, is a function $M_X(s)$ of a scalar parameter s, defined by

$$M_X(s) = \mathbf{E}[e^{sX}].$$

- The simpler notation $M(s)$ can also be used whenever the underlying random variable X is clear from the context.

Transforms

- For the defining formula $M_X(s) = \mathbf{E}[e^{sX}]$
- When X is a discrete random variable, the transform is given by

$$M(s) = \sum_x e^{sx}\, p_X(x).$$

- When X is a continuous random variable:

$$M(s) = \int_{-\infty}^{\infty} e^{sx}\, f_X(x) dx.$$

Example – a specific discrete r.v.

- Let

$$p_X(x) = \begin{cases} 1/2 & \text{if } x = 2, \\ 1/6 & \text{if } x = 3, \\ 1/3 & \text{if } x = 5. \end{cases}$$

- We have

$$M(s) = \mathbf{E}[e^{sX}] = \frac{1}{2}e^{2s} + \frac{1}{6}e^{3s} + \frac{1}{3}e^{5s}.$$

Example - Poisson

- Now consider the transform associated with a Poisson random variable.
- Let X be a Poisson random variable with parameter λ:

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0,1,2,\ldots$$

- The transform is

$$M(s) = \mathbf{E}[e^{sX}] = \sum_{x=0}^{\infty} e^{sx} \frac{\lambda^x e^{-\lambda}}{x!}$$

Example - Poisson

- We can simply this formula

$$M(s) = \sum_{x=0}^{\infty} e^{sx} \frac{\lambda^x e^{-\lambda}}{x!}$$

- Let $a = e^s \lambda$,

$$M(s) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{a^x}{x!} = e^{-\lambda} e^a = e^{\lambda(e^s - 1)}.$$

Example - Exponential

- Let X be an exponential random variable with parameter λ

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x \geq 0.$$

- Then, for $s < \lambda$,

$$M(s) = \mathbf{E}[e^{sX}] = \lambda \int_0^{\infty} e^{sx} e^{-\lambda x} dx$$

$$= \frac{\lambda e^{(s-\lambda)x}}{s - \lambda} \Big|_0^{\infty} \qquad (if\ s < \lambda)$$

$$= \frac{\lambda}{\lambda - s}$$

Transforms - Note

- It is important to realize that the transform is not a number but rather a function of a parameter s.
- Thus, we are dealing with a transformation that starts with a function, e.g., a PDF, and results in a new function.
- Strictly speaking, $M(s)$ is only defined for those values of s for which $\mathbf{E}[e^{sX}]$ is finite.
  - As in the preceding example.

Example - $aX + b$

- We consider the transform associated with a linear function of a random variable.
- Let $M_X(s)$ be the transform associated with a random variable X.
- Consider a new random variable

$$Y = aX + b.$$

Example - $aX + b$

- We then have

$$M_Y(s) = \mathbf{E}\left[e^{s(aX+b)}\right]$$
$$= e^{bs}\mathbf{E}[e^{saX}]$$
$$= e^{bs}M_X(sa).$$

- For example, if $\lambda = 1$, so that

$$M_X(s) = 1/(1-s)$$

and if $Y = 2X + 3$, then

$$M_Y(s) = e^{3s}/(1-2s)$$

Example - Normal

- Consider the transform associated with a normal random variable.
- Let X be a normal random variable with mean μ and variance σ^2.
- We first consider the special case of standard normal variable Y

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$

Example - Normal

- The associated transform is

$$M_Y(s) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} e^{sy} dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y^2/2)+sy} dy$$

$$= e^{s^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y-s)^2/2} dy$$

$$= e^{s^2/2}$$

Example - Normal

- For general normal random variable X with mean μ and variance σ^2

$$X = \sigma Y + \mu$$

- By applying transform of linear functions, we obtain

$$M_X(s) = e^{s\mu} M_Y(s\sigma) = e^{\frac{\sigma^2 s^2}{2} + \mu s}$$

From Transforms to Moments

- Why we gave transform an alternative name moment generating function?
- The moments of a random variable are easily computed from the associated transform.
- Consider the definition

$$M(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x)\,dx\,.$$

From Transforms to Moments

- Take derivative of both sides of

$$M(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x) dx \, .$$

- We obtain

$$\frac{d}{ds} M(s) = \frac{d}{ds} \int_{-\infty}^{\infty} e^{sx} f_X(x) dx$$

$$= \int_{-\infty}^{\infty} x e^{sx} f_X(x) dx$$

From Transforms to Moments

- Last slide: $\frac{d}{ds}M(s) = \int_{-\infty}^{\infty} xe^{sx} f_X(x)dx$
- Take $s = 0$

$$\left.\frac{d}{ds}M(s)\right|_{s=0} = \mathbf{E}[X]$$

- Generally, differentiating n times, we get

$$\left.\frac{d^n}{ds^n}M(s)\right|_{s=0} = \mathbf{E}[X^n]$$

Moments - Example

- Let

$$p_X(x) = \begin{cases} 1/2 & \text{if } x = 2, \\ 1/6 & \text{if } x = 3, \\ 1/3 & \text{if } x = 5. \end{cases}$$

- Recall that

$$M(s) = \frac{1}{2}e^{2s} + \frac{1}{6}e^{3s} + \frac{1}{3}e^{5s}.$$

Moments - Example

- $M(s) = \frac{1}{2}e^{2s} + \frac{1}{6}e^{3s} + \frac{1}{3}e^{5s}$.
- Then

$$\mathbf{E}[X] = \frac{d}{ds}M(s)\Big|_{s=0} = \frac{1}{2}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{3}\cdot 5 = \frac{19}{6}$$

- Also

$$\mathbf{E}[X^2] = \frac{d^2}{ds^2}M(s)\Big|_{s=0} = \frac{1}{2}\cdot 4 + \frac{1}{6}\cdot 9 + \frac{1}{3}\cdot 25 = \frac{71}{6}$$

Moments - Example

- For an exponential random variable with PDF
$$f_X(x) = \lambda e^{-\lambda x}, x \geq 0.$$
- We found earlier that
$$M(s) = \frac{\lambda}{\lambda - s}$$
- Thus $\frac{d}{ds}M(s) = \frac{\lambda}{(\lambda - s)^2}, \frac{d^2}{ds^2}M(s) = \frac{2\lambda}{(\lambda - s)^3}$.

Moments - Example

- Similarly

$$\mathbf{E}[X] = \frac{d}{ds}M(s)\bigg|_{s=0} = \frac{\lambda}{(\lambda - s)^2}\bigg|_{s=0} = \frac{1}{\lambda}$$

- Also

$$\mathbf{E}[X^2] = \frac{d^2}{ds^2}M(s)\bigg|_{s=0} = \frac{2\lambda}{(\lambda - s)^3}\bigg|_{s=0} = \frac{2}{\lambda^2}$$

Moments - Example

- We note two more useful and generic properties of transforms.
- For any random variable X, we have

$$M_X(0) = \mathbf{E}[e^{0X}] = 1$$

- And if X takes only nonnegative integer values, then

$$\lim_{s \to -\infty} M_X(s) = \mathrm{P}(X = 0)$$

Inversion of Transforms

- A very important property of the transform $M_X(s)$ is that it can be inverted,
- That is, it can be used to determine the probability law of the random variable X.
- To do this, some appropriate mathematical conditions are required, which are satisfied in all of the following examples.

Inversion of Transforms

- Formally, the transform $M_X(s)$ associated with a random variable X uniquely determines the CDF of X,
  - assuming that $M_X(s)$ is finite for all s in some interval $[-a, a]$, where a is a positive number.

Inversion of Transforms - Example

- We are told that the transform associated with a random variable X is

$$M_X(s) = \frac{1}{4}e^{-s} + \frac{1}{2} + \frac{1}{8}e^{4s} + \frac{1}{8}e^{5s}$$

- Then we can infer that X is a discrete random variable.
- The different values that X can take can be read from the corresponding exponents, and are -1, 0, 4, and 5.

Inversion of Transforms - Example

- $M_X(s) = \frac{1}{4}e^{-s} + \frac{1}{2} + \frac{1}{8}e^{4s} + \frac{1}{8}e^{5s}$
- The probability of each value x is given by the coefficient multiplying the corresponding e^{sx} term:

$$\mathrm{P}(X = -1) = 1/4$$
$$\mathrm{P}(X = 0) = 1/2$$
$$\mathrm{P}(X = 4) = 1/8$$
$$\mathrm{P}(X = 5) = 1/8$$

Inversion of Transforms - Example

- We are told that the transform associated with a random variable X is of the form
$$M(s) = \frac{pe^s}{1-(1-p)e^s}.$$
where $0 < p \leq 1$.
- Recall the formula for the geometric series valid (for $-1 < \alpha < 1$):
$$\frac{1}{1-\alpha} = 1 + \alpha + \alpha^2 + \cdots$$

Inversion of Transforms - Example

- We use the formula with $\alpha = (1-p)e^s$, and for s sufficiently close to zero so that $(1-p)e^s < 1$.
- We obtain

$$M(s) = pe^s(1 + (1-p)e^s + (1-p)^2 e^{2s} + (1-p)^3 e^{3s} + \cdots)$$

- We can infer that

$$P(X = k) = p(1-p)^{k-1}, \qquad k = 1, 2, \ldots$$

- which is the geometric distribution with parameter p.

Inversion of Transforms - Example

- We address the transform associated with a mixture of two distributions.
- Consider a neighborhood bank has three tellers, two of them fast, one slow.
- The time to assist a customer is exponentially distributed with parameter $\lambda = 6$ at the fast tellers, and $\lambda = 4$ at the slow teller.
- Alice enters the bank and chooses a teller at random, we try to find the PDF of the time X it takes.

Inversion of Transforms - Example

- We have

$$f_X(x) = \frac{2}{3}6e^{-6x} + \frac{1}{3}4e^{-4x}, x \geq 0$$

- Then

$$M(s) = \int_0^\infty e^{sx}\left(\frac{2}{3}6e^{-6x} + \frac{1}{3}4e^{-4x}\right)dx$$

$$= \frac{2}{3}\int_0^\infty e^{sx}6e^{-6x}dx + \frac{1}{3}\int_0^\infty e^{sx}4e^{-4x}dx$$

$$= \frac{2}{3}\frac{6}{6-s} + \frac{1}{3}\frac{4}{4-s}, \qquad \text{for } s < 4$$

Sums of Independent Variables

- Transform methods are particularly convenient when dealing with a sum of random variables.
- An important result is that addition of independent random variables corresponds to multiplication of transforms.

Sums of Independent Variables

- We X and Y be independent random variables, and let $Z = X + Y$. By definition we have
$$M_Z(s) = \mathrm{E}[e^{sZ}] = \mathrm{E}[e^{sX}e^{sY}]$$
- Since X and Y are independent, e^{sX} and e^{sY} are independent as well. Hence,
$$M_Z(s) = M_X(s)M_Y(s)$$

Sums of Independent Variables

- Generally, $X_1, \cdots, X_n$ is a collection of independent random variables, and

$$Z = X_1 + \cdots + X_n$$

- Then,

$$M_Z(s) = M_{X_1}(s) \cdots M_{X_n}(s)$$

Sums of Variables - Example

- We address the transform associated with the binomial.
- Let $X_1, \cdots, X_n$ be independent Bernoulli random variables with a common parameter p. Then $M_{X_i}(s) = (1-p)e^{0s} + pe^{1s} = 1 - p + pe^s$
- For $Z = X_1 + \cdots + X_n$,

$$M_Z(s) = (1 - p + pe^s)^n$$

Sums of Variables - Example

- We will show that the *sum of independent Poisson random variables is Poisson*.
- Let X and Y be independent Poisson random variables with means λ and μ, respectively.
- Let $Z = X + Y$

$$M_X(s) = e^{\lambda(e^s - 1)}$$
$$M_Y(s) = e^{\mu(e^s - 1)}$$

Sums of Variables - Example

- We have

$$M_Z(s) = M_X(s)M_Y(s) = e^{(\lambda+\mu)(e^s-1)}$$

- Thus, transform associated with X is the same as the transform associated with a Poisson random variable with mean $\lambda + \mu$.
- By the uniqueness property of transforms, Z is Poisson with mean $\lambda + \mu$.

Sums of Variables - Example

- We will show that the sum of independent normal random variables is normal.
- Let X and Y be independent normal random variables with means μ_x andμ_y, and variances σ_x^2, σ_y^2, respectively.
- Let $Z = X + Y$

Sums of Variables - Example

- Then

$$M_X(s) = e^{\frac{\sigma_x^2 s^2}{2} + \mu_x s}$$

$$M_Y(s) = e^{\frac{\sigma_y^2 s^2}{2} + \mu_y s}$$

- and

$$M_Z(s) = e^{\frac{(\sigma_x^2 + \sigma_y^2) s^2}{2} + (\mu_x + \mu_y) s}$$

- It corresponds to $N\left(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2\right)$.

Transforms Associated w/ Joint Dist.

- Consider random variables $X_1, \cdots, X_n$, the associated multivariate transform is a function with n parameters

$$M_{X_1,\cdots,X_n}(s_1, \cdots, s_n) = \mathbf{E}[e^{s_1 X_1 + \cdots + s_n X_n}]$$

- The inversion property of transforms can be extended to the multivariate case.

Transforms for Common Discrete r.v.

Bernoulli(p) $\quad (k = 0, 1)$

$$p_X(k) = \begin{cases} p, & \text{if } k = 1, \\ 1 - p, & \text{if } k = 0, \end{cases} \qquad M_X(s) = 1 - p + pe^s.$$

Binomial(n, p) $\quad (k = 0, 1, \ldots, n)$

$$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \qquad M_X(s) = (1 - p + pe^s)^n.$$

Geometric(p) $\quad (k = 1, 2, \ldots)$

$$p_X(k) = p(1 - p)^{k-1}, \qquad M_X(s) = \frac{pe^s}{1 - (1 - p)e^s}.$$

Poisson(λ) $\quad (k = 0, 1, \ldots)$

$$p_X(k) = \frac{e^{-\lambda} \lambda^k}{k!}, \qquad M_X(s) = e^{\lambda(e^s - 1)}.$$

Uniform(a, b) $\quad (k = a, a + 1, \ldots, b)$

$$p_X(k) = \frac{1}{b - a + 1}, \qquad M_X(s) = \frac{e^{sa}\left(e^{s(b-a+1)} - 1\right)}{(b - a + 1)(e^s - 1)}.$$

Transforms for Common Continuous Random Variables

Uniform(a, b) $\quad (a \le x \le b)$

$$f_X(x) = \frac{1}{b-a}, \qquad M_X(s) = \frac{e^{sb} - e^{sa}}{s(b-a)}.$$

Exponential(λ) $\quad (x \ge 0)$

$$f_X(x) = \lambda e^{-\lambda x}, \qquad M_X(s) = \frac{\lambda}{\lambda - s}, \qquad (s < \lambda).$$

Normal(μ, σ^2) $\quad (-\infty < x < \infty)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}, \qquad M_X(s) = e^{(\sigma^2 s^2/2)+\mu s}.$$

Content

- Derived Distributions
- Covariance and Correlation
- Conditional Expectation and Variance Revisited
- Transforms
- Sum of a Random Number of Independent Random Variables

Sums of A Random Number of Independent Random Variables

- So far we have always assumed that the number of variables in the sum is known and fixed.

- Now we will consider the case where the number of random variables being added is itself random.

Sums of A Random Number of Independent Random Variables

- That is, we consider

$$Y = X_1 + \cdots + X_N$$

- Where $X_1, \cdots, X_N$ are identical and independent random variables.
- And N is a random variable that takes nonnegative integer values.
  - Assume that its PMF is p_N.

Sums of A Random Number of Independent Random Variables

- Denote by $\mathbf{E}[X]$ and $\mathrm{var}(X)$ the common mean and variance, respectively, of the X_i.
- We wish to derive formulas for the mean, variance, and the transform of Y.
- We address this by first conditioning on event $N = n$.

Sums of A Random Number of Independent Random Variables

- Firstly,

$$\mathbf{E}[Y|N=n] = \mathbf{E}[X_1 + \cdots + X_n] = n\mathbf{E}[X]$$

- Hence,

$$\mathbf{E}[Y|N] = N \cdot \mathbf{E}[X]$$

  - Recall that $\mathbf{E}[Y|N]$ is a random variable, which takes value $\mathbf{E}[Y|N=n]$ when $N=n$.
- Then by the *law of iterated expectations*, we obtain

$$\mathbf{E}[Y] = \mathbf{E}\big[\mathbf{E}[Y|N]\big] = \mathbf{E}[N]\mathbf{E}[X]$$

Sums of A Random Number of Independent Random Variables

- Similarly,

$$\mathrm{var}(Y|N=n) = \mathrm{var}(X_1 + \cdots + X_n)$$
$$= n \cdot \mathrm{var}(X).$$

- *Law of total variance*: (proof omitted)

$$\mathrm{var}(X) = \mathrm{E}[\mathrm{var}(X|Y)] + \mathrm{var}(E[X|Y]).$$

- $\mathrm{var}(Y) = \mathbf{E}[\mathrm{var}(Y|N)] + \mathrm{var}(\mathbf{E}[Y|N])$

$$= \mathbf{E}[N \cdot \mathrm{var}(X)] + \mathrm{var}(N \cdot \mathbf{E}[X])$$
$$= \mathbf{E}[N]\mathrm{var}(X) + \mathbf{E}[X]^2\mathrm{var}(N)$$

Sums of A Random Number of Independent Random Variables

- Similarly, we can compute the transform.
- For each n,
  $\mathbf{E}[e^{sY}|N = n] = \mathbf{E}[e^{sX_1} \cdots e^{sX_n}] = M_X(s)^n$
- Then,
  $$M_Y(s) = \mathbf{E}[e^{sY}]$$
  $$= \mathbf{E}\big[\mathbf{E}[e^{sY}|N]\big] \quad \text{// iterated expectation}$$
  $$= \mathbf{E}[M_X(s)^N]$$
  $$= \sum_{n=0}^{\infty} M_X(s)^n p_N(n).$$

Sums of A Random Number of Independent Random Variables

- Observe that $M_X(s)^n = e^{n \cdot \log M_X(s)}$
- We have,

$$M_Y(s) = \sum_{n=0}^{\infty} e^{n \cdot \log M_X(s)} p_N(n).$$

- Recall

$$M_N(s) = \mathbf{E}[e^{sN}] = \sum_{n=0}^{\infty} e^{sn} p_N(n).$$

- Thus $M_Y(s) = M_N(\log M_X(s))$.
  - $M_Y(s)$ is obtained from the formula for $M_N(s)$, with s replaced with $\log M_X(s)$.

Summary of A Random Number of Independent Random Variables

- Expectation:

$$\mathbf{E}[Y|N] = N \cdot \mathbf{E}[X]$$

- Variance:

$$\mathrm{var}(Y) = \mathbf{E}[N]\mathrm{var}(X) + \mathbf{E}[X]^2\mathrm{var}(N)$$

- Transform:

$$M_Y(s) = M_N(\log M_X(s))$$

Examples

- A remote village has three gas stations.
- Each gas station is open on any given day with probability 1/2, independently of the others.
- The amount of gas X for each station is uniformly distributed between 0 and 1000.
- Let N be number of open gas stations and Y the total gas available.

Examples

- Firstly, N is binomial and

$$M_N(s) = (1 - p + pe^s)^3 = \frac{1}{8}(1 + e^s)^3$$

- The transform $M_X(s)$ for the amount X of gas in one gas station is $M_X(s) = \frac{e^{1000s}-1}{1000s}$.
  - Recall for uniform(a, b): $M_Z(s) = \frac{e^{sb}-e^{sa}}{s(b-a)}$.
- $M_Y(s) = M_N(\log M_X(s))$: Replace e^s in $M_N(s)$ with $M_X(s)$ and we get

$$M_Y(s) = \frac{1}{8}\left(1 + \frac{e^{1000s} - 1}{1000s}\right)^3$$

Examples

- Now we discuss the sum of a geometric number of independent exponential random variables.
- Suppose Alice visits a number of bookstores for a certain book.
- Any store carries the book with probability p.
- Alice spends an exponentially random amount of time X_i at store i, with mean λ.
  - Once she find it, she stops.

Examples

- Alice will keep visiting bookstores until she buys the book.
- The time spent in each is independent of everything else.
- We wish to find the mean, variance, and PDF of the total time spent in bookstores.

Examples

- N = the total number of stores she visits.
  - Geometric random variable with parameter p.
- Y = the total time spent in bookstores.
- $Y = X_1 + X_2 + \cdots + X_N$
  - Each X_i: exponential random variable with parameter λ.
- $\mathbf{E}[Y] = \mathbf{E}[N]\mathbf{E}[X] = \frac{1}{p\lambda}$
- $\mathrm{var}(Y) = \mathbf{E}[N]\mathrm{var}(X) + \mathbf{E}[X]^2\mathrm{var}(N)$

$$= \frac{1}{p}\frac{1}{\lambda^2} + \frac{1}{\lambda^2}\frac{1-p}{p^2} = \frac{1}{p^2\lambda^2}$$

Examples

- Recall

$$M_X(s) = \frac{\lambda}{\lambda - s}$$

$$M_N(s) = \frac{pe^s}{1 - (1-p)e^s}$$

- We obtain

$$M_Y(s) = \frac{pM_X(s)}{1 - (1-p)M_X(s)} = \frac{p\lambda}{p\lambda - s}$$

Examples

- Last slide: $M_Y(s) = \frac{p\lambda}{p\lambda - s}$

- We recognize this as the transform associated with an exponentially distributed r.v. with parameter $p\lambda$, thus

$$f_Y(y) = p\lambda e^{-p\lambda y}, \qquad y \geq 0$$