# Functions of Several Random Variables (Ch. 5.5)

Dason Kurkiewicz

Iowa State University

June 19, 2013

# Outline

Functions of Several Random Variables

Expectations and variances of linear combinations

The Central Limit Theorem

Approximating the Mean and Variance of a Function

# Functions of several random variables

- We often consider functions of random variables of the form:

$$U = g(X, Y, \ldots, Z)$$

where $X$, $Y, \ldots, Z$ are random variables.
- $U$ is itself a random variable.

# Example: connecting steel parts

- Suppose that a steel plate with nominal thickness .15 in. is to rest in a groove of nominal width .155 in., machined on the surface of a steel block.

Relative Frequency Distribution of Plate Thicknesses

| Plate Thickness (in.) | Relative Frequency |
|---|---|
| .148 | .4 |
| .149 | .3 |
| .150 | .3 |

Relative Frequency Distribution of Slot Widths

| Slot Width (in.) | Relative Frequency |
|---|---|
| .153 | .2 |
| .154 | .2 |
| .155 | .4 |
| .156 | .2 |

- $X$ = plate thickness
- $Y$ = slot width
- $U = Y - X$, the "wiggle room" of the plate

# The distributions of $X$, $Y$, and $U$

Marginal and Joint Probabilities for $X$ and $Y$

| $y \setminus x$ | .148 | .149 | .150 | $f_Y(y)$ |
|---|---|---|---|---|
| .156 | .08 | .06 | .06 | .2 |
| .155 | .16 | .12 | .12 | .4 |
| .154 | .08 | .06 | .06 | .2 |
| .153 | .08 | .06 | .06 | .2 |
| $f_X(x)$ | .4 | .3 | .3 | |

The Probability Function for the Clearance $U = Y - X$

| $u$ | $f(u)$ |
|---|---|
| .003 | .06 |
| .004 | $.12 = .06 + .06$ |
| .005 | $.26 = .08 + .06 + .12$ |
| .006 | $.26 = .08 + .12 + .06$ |
| .007 | $.22 = .16 + .06$ |
| .008 | .08 |

- Determining the distribution of $U$ is difficult in the continuous case.

# Outline

# Expectations and variances of linear combinations

- $X_1, X_2, \ldots, X_n$ are independent random variables and

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

then:

$$\begin{aligned} E(Y) &= E(a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) \\ &= a_0 + a_1 E(X_1) + a_2 E(X_2) + \cdots + a_n E(X_n) \end{aligned}$$

$$\begin{aligned} Var(Y) &= Var(a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) \\ &= a_1^2 \cdot Var(X_1) + a_2^2 \cdot Var(X_2) + \cdots + a_n^2 \cdot Var(X_n) \end{aligned}$$

# Your turn: linear combinations

- Say we have two independent random variables $X$ and $Y$ with $E(X) = 3.3, Var(X) = 1.91, E(Y) = 25$, and $Var(Y) = 65$.
- Find:

$$\begin{aligned} E(3 + 2X - 3Y) \\ E(-4X + 3Y) \\ E(-4X - 6Y) \\ Var(3 + 2X - 3Y) \\ Var(2X - 5Y) \\ Var(-4X - 6Y) \end{aligned}$$

# Answers: linear combinations

$$
\begin{aligned}
E(3+2X-3Y) &= 3+2E(X)-3E(Y) \\
&= 3+2\cdot 3.3-3\cdot 25 \\
&= -65.4
\end{aligned}
$$

$$
\begin{aligned}
E(-4X+3Y) &= -4E(X)+3E(Y) \\
&= -4\cdot 3.3+3\cdot 25 \\
&= 61.8
\end{aligned}
$$

$$
\begin{aligned}
E(-4X-6Y) &= -4\cdot E(X)-6\cdot E(Y) \\
&= -4\cdot 3.3-6\cdot 25 \\
&= -163.2
\end{aligned}
$$

# Answers: linear combinations

$$
\begin{aligned}
Var(3+2X-3Y) &= 2^2 \cdot Var(X) + (-3)^2 Var(Y) \\
&= 4 \cdot 1.91 + 9 \cdot 65 \\
&= 592.64
\end{aligned}
$$

$$
\begin{aligned}
Var(2X-5Y) &= 2^2 \cdot Var(X) + (-5)^2 Var(Y) \\
&= 4 \cdot 1.91 + 25 \cdot 65 \\
&= 1632.64
\end{aligned}
$$

$$
\begin{aligned}
Var(-4X-6Y) &= (-4)^2 \cdot Var(X) + (-6)^2 Var(Y) \\
&= 16 \cdot 1.91 + 36 \cdot 65 \\
&= 2370.56
\end{aligned}
$$

# Your turn: more linear combinations

- Say $X \sim \text{Binomial}(n = 10,\ p = 0.5)$ and $Y \sim \text{Poisson}(\lambda = 3)$.
- Calculate:

$$E(5 + 2X - 7Y)$$
$$Var(5 + 2X - 7Y)$$

# Answer: more linear combinations

- First, note that:

$$\begin{aligned} E(X) &= np = 10 \cdot 0.5 = 5 \\ E(Y) &= \lambda = 3 \\ Var(X) &= np(1-p) = 10(0.5)(1-0.5) = 2.5 \\ Var(Y) &= \lambda = 3 \end{aligned}$$

Now, we can calculate:

$$\begin{aligned} E(5+2X-7Y) &= 5 + 2E(X) - 7E(Y) \\ &= 5 + 2 \cdot 5 - 7 \cdot 3 \\ &= -6 \end{aligned}$$

$$\begin{aligned} Var(5+2X-7Y) &= 2^2 \cdot Var(X) + (-7)^2 \cdot Var(Y) \\ &= 4 \cdot 2.5 + 49 \cdot 3 \\ &= 157 \end{aligned}$$

# iid random variables.

- **Identically Distributed**: Random variables $X_1, X_2, \ldots, X_n$ are identically distributed if they have the same probability distribution.
- **"iid"**: Random variables $X_1, X_2, \ldots, X_n$ are iid if they are **I**ndependent and **I**dentically **D**istributed.

# Your turn: averages of iid random variables

- $X_1, X_2, \ldots, X_n$ are iid with expectation $\mu$ and variance $\sigma^2$.
- Derive:

$$E(\overline{X})$$
$$Var(\overline{X})$$

where:

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

the mean of the $X_i$'s.

# Answers: averages of iid random variables

$$
\begin{aligned}
E(\overline{X}) &= E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \\
&= \frac{1}{n}E(X_1) + \frac{1}{n}E(X_2) + \cdots + \frac{1}{n}E(X_n) \\
&= \underbrace{\frac{1}{n}\mu + \frac{1}{n}\mu + \cdots + \frac{1}{n}\mu}_{n \text{ times}} \\
&= n \cdot \frac{1}{n}\mu \\
&= \boxed{\mu}
\end{aligned}
$$

- Remember $E(\overline{X}) = \mu$: it's an important result.

# Answers: averages of iid random variables



$$
\begin{aligned}
Var(\overline{X}) &= Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \\
&= \left(\frac{1}{n}\right)^2 Var(X_1) + \left(\frac{1}{n}\right)^2 Var(X_2) + \cdots + \left(\frac{1}{n}\right)^2 \cdot Var(X_n) \\
&= \underbrace{\frac{1}{n^2}\sigma^2 + \frac{1}{n^2}\sigma^2 + \cdots + \frac{1}{n^2}\sigma^2}_{n \text{ times}} \\
&= n \cdot \frac{1}{n^2}\sigma^2 \\
&= \boxed{\frac{\sigma^2}{n}}
\end{aligned}
$$

- Remember $Var(\overline{X}) = \frac{\sigma^2}{n}$: it's another important result.

# Example: length of seeds

- A botanist has collected a sample of 10 seeds and measures the length of each.
- The seed lengths $X_1, X_2, \ldots, X_{10}$ are supposed to be iid with mean $\mu = 5$ mm and variance $\sigma^2 = 2\ mm^2$.

$$E(\overline{X}) = \mu = 5$$
$$Var(\overline{X}) = \sigma^2/n = 2/10 = 0.2$$

# Outline

# The Central Limit Theorem

- If $X_1, X_2, \ldots, X_n$ are *any* iid random variables with mean $\mu$ and variance $\sigma^2 < \infty$ then as $n \to \infty$

$$\overline{X} \approx \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$$

- The Central Limit Theorem (CLT) one of the most important and useful results in statistics.

# Example: tool serial numbers

- $W_1$ = last digit of the serial number observed next Monday at 9 AM
- $W_2$ = last digit of the serial number the Monday after at 9 AM
- $W_1$ and $W_2$ are independent with pmf:

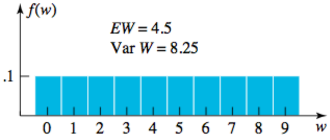
$$f(w) = \begin{cases} 0.1 & w = 0, 1, 2, \ldots, 9 \\ 0 & \text{otherwise} \end{cases}$$

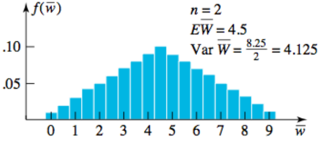
- $\overline{W} = \frac{1}{2}(W_1 + W_2)$ has the pmf:

The Probability Function for $\overline{W}$ for $n = 2$

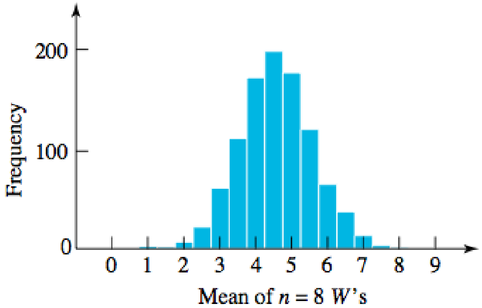
| $\bar{w}$ | $f(\bar{w})$ | $\bar{w}$ | $f(\bar{w})$ | $\bar{w}$ | $f(\bar{w})$ | $\bar{w}$ | $f(\bar{w})$ | $\bar{w}$ | $f(\bar{w})$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .01 | 2.0 | .05 | 4.0 | .09 | 6.0 | .07 | 8.0 | .03 |
| 0.5 | .02 | 2.5 | .06 | 4.5 | .10 | 6.5 | .06 | 8.5 | .02 |
| 1.0 | .03 | 3.0 | .07 | 5.0 | .09 | 7.0 | .05 | 9.0 | .01 |
| 1.5 | .04 | 3.5 | .08 | 5.5 | .08 | 7.5 | .04 | | |

# Example: tool serial numbers



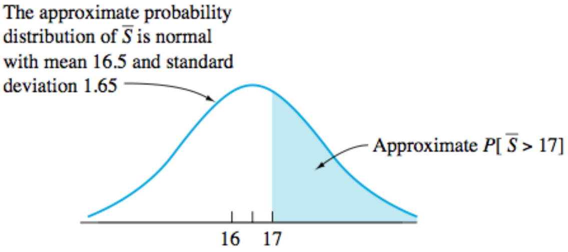




- What if $\overline{W} = \frac{1}{8}(W_1 + W_2 + \cdots + W_8)$, the average of 8 days of initial serial numbers?



# Example: excess sale time

- $\overline{S}$ = sample mean excess sale time (over a 7.5 s threshold) for 100 stamp sales.
- Each individual excess sale time should have an Exp($\alpha = 16.5$ s) distribution. That means:
  - $E(\overline{S}) = \alpha = 16.5$ s
  - $SD(\overline{S}) = \sqrt{\text{Var}(\overline{S})} = \sqrt{\frac{\alpha^2}{100}} = 1.65$ s
  - By the Central Limit Theorem, $\overline{S} \approx N(16.5, 1.65^2)$
- We want to approximate $P(\overline{S} > 17)$.



# Example: excess sale time

$$
\begin{aligned}
P(\overline{S} > 17) &= P\Big(\frac{\overline{S} - 16.5}{1.65} > \frac{17 - 16.5}{1.65}\Big) \\
&\approx P(Z > 0.303) \qquad (Z \sim N(0,1)) \\
&= 1 - P(Z \leq 0.303) \\
&= 1 - \Phi(0.303) \\
&= 1 - 0.62 \quad \text{from the standard normal table} \\
&= 0.38
\end{aligned}
$$

# Example: net weight of baby food jars

- Individual jar weights are iid with unknown mean $\mu$ and standard deviation $\sigma = 1.6$ g
- $\overline{V}$ = sample mean weight of n jars $\approx N\left(\mu, \frac{1.6^2}{n}\right)$.
- We want to find $\mu$. One way to hone in on $\mu$ is to find $n$ such that:

$$P(\mu - 0.3 < \overline{V} < \mu + 0.3) = 0.8$$

That way, our measured value of $\overline{V}$ is likely to be close to $\mu$.

# Example: net weight of baby food jars

$$
\begin{aligned}
0.8 &= P(\mu - 0.3 < \overline{V} < \mu + 0.3) \\
&= P\left(\frac{-0.3}{1.6/\sqrt{n}} < \frac{\overline{V} - \mu}{1.6/\sqrt{n}} < \frac{0.3}{1.6/\sqrt{n}}\right) \\
&\approx P(-0.19\sqrt{n} < Z < 0.19\sqrt{n}) \quad \text{(by CLT)} \\
&= 1 - 2\Phi(-0.19\sqrt{n}) \quad \text{(look at the N(0,1) pdf)} \\
\Phi^{-1}(0.1) &= -0.19\sqrt{n} \\
n &= \frac{\Phi^{-1}(0.1)^2}{(-0.19)^2} \\
&= \frac{(-1.28)^2}{(-0.19)^2} \quad \text{(standard normal table)} \\
&= 46.10
\end{aligned}
$$

- Hence, we'll need a sample size of $n = 47$.

# Example: cars

- Suppose a bunch of cars pass through certain stretch of road. Whenever a car comes, you look at your watch and record the time.
- Let $X_i$ be the time (in hours) between when the $i$'th car comes and the $(i+1)$'th car comes, $i = 1, \ldots, 44$. Suppose you know:

$$X_1, X_2, \ldots, X_{44} \sim \text{ iid } f(x) = e^{-x} \quad x \geq 0$$

- Find the probability that the average time gap between cars exceeds 1.05 hours.

# Example: cars

$$\mu = E(X_1)$$

$$= \int_{-\infty}^{\infty} x f(x) dx$$

$$= \int_{0}^{\infty} x e^{-x} dx$$

$$= -e^{-x}(x+1)|_0^{\infty} \quad \text{integration by parts}$$

$$= 1$$

# Example: cars

$$
\begin{aligned}
E(X_1^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx \\
&= \int_0^{\infty} x^2 e^{-x} dx \\
&= -e^{-x}(x^2 + 2x + 2)|_0^{\infty} \qquad \text{integration by parts} \\
&= 2 \\
\sigma^2 &= Var(X_1) \\
&= E(X_1^2) - E^2(X_1) \\
&= 2 - 1^2 \\
&= 1
\end{aligned}
$$

# Example: cars

$$\begin{aligned}\overline{X} &\sim \text{ approx. } N(\mu, \sigma^2/n) \\ &= N(1, 1/44)\end{aligned}$$

Thus:

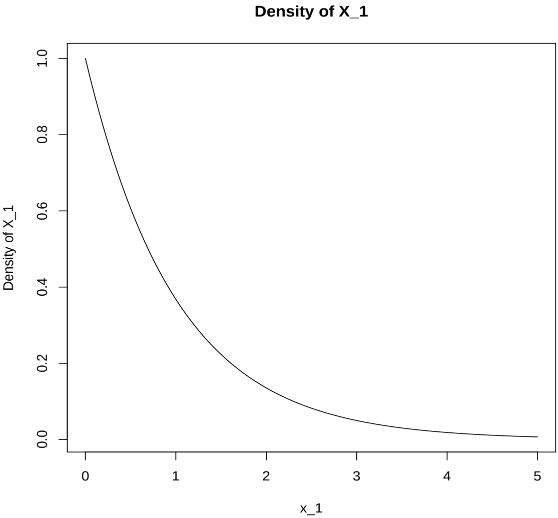
$$\frac{\overline{X} - 1}{\sqrt{1/44}} \sim N(0, 1)$$

# Example: cars

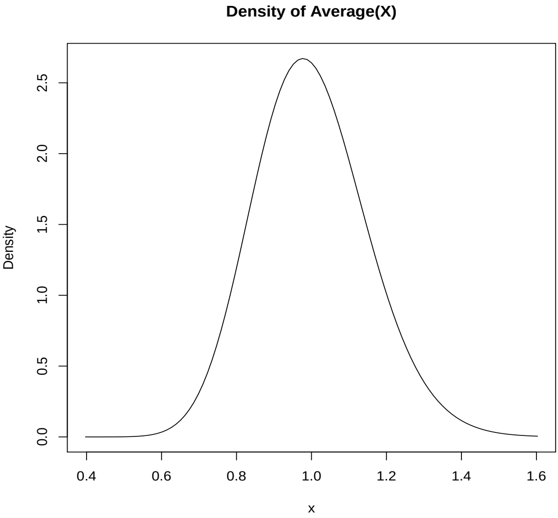
Now, we're ready to approximate:

$$\begin{aligned}
P(\overline{X} > 1.05) &= P\left(\frac{\overline{X} - 1}{\sqrt{1/44}} > \frac{1.05 - 1}{\sqrt{1/44}}\right) \\
&= P\left(\frac{\overline{X} - 1}{\sqrt{1/44}} > 0.332\right) \\
&\approx P(Z > 0.332) \\
&= 1 - P(Z \leq 0.332) \\
&= 1 - \Phi(0.332) \\
&= 1 - 0.630 = 0.370
\end{aligned}$$

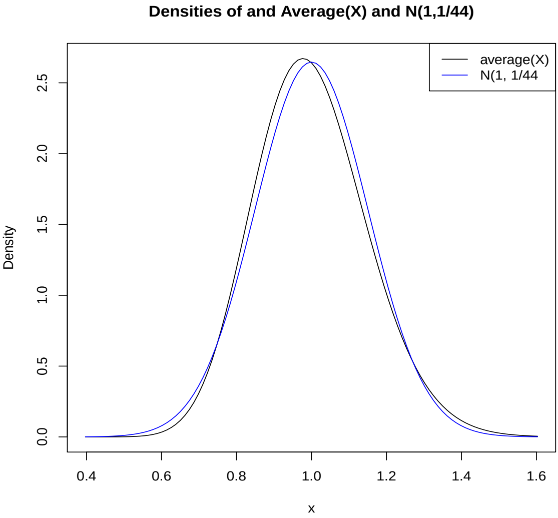
# Example: cars



# Example: cars



# Example: cars



# Outline

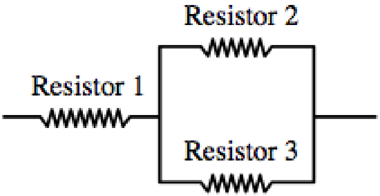
## Approximating $E(U)$ and $Var(U)$ when determining $f_U(u)$ is too hard

- If $X, Y, \ldots, Z$ are independent, $g$ is well-behaved, and the variances $\text{Var}(X)$, $\text{Var}(Y), \ldots,$ $\text{Var}(Z)$ are small enough, then $U = g(X, Y, \ldots Z)$ has:

$$E(U) \approx g(E(X), E(Y), \ldots, E(Z))$$
$$\text{Var}(U) \approx \left(\frac{\partial g}{\partial x}\right)^2 Var(X) + \left(\frac{\partial g}{\partial y}\right)^2 Var(Y) + \cdots + \left(\frac{\partial g}{\partial z}\right)^2 Var(Z)$$

- These formulas are often called the **propagation of error formulas**.

# Example: an electric circuit




- $R$ is the total resistance of the circuit.
- $R_1$, $R_2$, and $R_3$ are the resistances of resistors 1, 2, and 3, respectively.
- $E(R_i) = 100$, $\text{Var}(R_i) = 2$, $i = 1, 2, 3$.

$$R = g(R_1, R_2, R_3) = R_1 + \frac{R_2 R_3}{R_2 + R_3}$$

# Example: an electric circuit

$$E(R) \approx g(100, 100, 100) = 100 + \frac{(100)(100)}{100 + 100} = 150\Omega$$

$$\frac{\partial g}{\partial r_1} = 1$$

$$\frac{\partial g}{\partial r_2} = \frac{(r_2 + r_3)r_3 - r_2 r_3}{(r_2 + r_3)^2} = \frac{r_3^2}{(r_2 + r_3)^2}$$

$$\frac{\partial g}{\partial r_3} = \frac{(r_2 + r_3)r_2 - r_2 r_3}{(r_2 + r_3)^2} = \frac{r_2^2}{(r_2 + r_3)^2}$$

$$\text{Var(R)} \approx (1)^2(2)^2 + \left(\frac{(100)^2}{(100 + 100)^2}\right)^2 (2)^2 + \left(\frac{(100)^2}{(100 + 100)^2}\right)^2 (2)^2$$

$$= 4.5$$

$$\text{SD}(R)\sqrt{4.5} \approx 2.12\Omega$$