STAT 3375Q: Introduction to Mathematical Statistics I

Lecture 11: Special Continuous Distributions: Normal (Gaussian)

Mary Lai Salvaña, Ph.D.

Department of Statistics
University of Connecticut

February 28, 2024

Outline

1. Quiz 3 Review Exercises Solutions
2. Previously...
   - Uniform Distribution
   - Normal (Gaussian) Distribution
   - Standard Normal Distribution
3. Normal (Gaussian) Distribution (cont'd)

Quiz 3 Review Exercises Solutions

Problem 1

Let X have a PDF defined by: $f(x) = \begin{cases} cxe^{-x}, & x \geq 2, \\ 0, & \text{elsewhere.} \end{cases}$

(a) Find the constant c.

Solution:

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x)dx \quad \text{def'n of PDF} \\ \Rightarrow 1 &= c\left(\int_{-\infty}^{2} 0dx + \int_{2}^{\infty} xe^{-x}dx\right) \quad \text{substitute the form of PDF above} \\ \Rightarrow 1 &= c\int_{2}^{\infty} xe^{-x}dx \quad \text{need to do integration by parts: } \int_a^b udv = uv|_a^b - \int_a^b vdu \end{aligned}$$

$$\int_2^\infty xe^{-x}dx$$

- $u = x \quad \Rightarrow du = dx$
- $dv = e^{-x}dx \quad \Rightarrow v = \int dv = \int e^{-x}dx = -e^{-x}$

$$\Rightarrow \int_2^\infty xe^{-x}dx = -xe^{-x}|_2^\infty - \int_2^\infty (-e^{-x})dx$$

(cont'd next slide)

Problem 1

Let X have a PDF defined by: $f(x) = \begin{cases} cxe^{-x}, & x \geq 2, \\ 0, & \text{elsewhere.} \end{cases}$

a. Find the constant c.

Solution:

$$\Rightarrow 1 = c\left(-xe^{-x}\Big|_2^{\infty} - \int_2^{\infty} (-e^{-x})dx\right) \quad \text{after integrating by parts}$$

$$\Rightarrow 1 = c\left\{\lim_{x\to\infty}(-xe^{-x}) - (-2e^{-2}) + \int_2^{\infty} e^{-x}dx\right\} \quad \text{evaluating bounds with } \infty$$

$$\Rightarrow 1 = c\left\{\lim_{x\to\infty}(-xe^{-x}) + 2e^{-2} - e^{-x}\Big|_2^{\infty}\right\} \quad \text{integrating the exponential}$$

$$\Rightarrow 1 = c\left\{\lim_{x\to\infty}(-xe^{-x}) + 2e^{-2} + \lim_{x\to\infty}(-e^{-x}) - (-e^{-2})\right\} \quad \text{evaluating bounds with } \infty$$

$$\Rightarrow 1 = c\left\{\lim_{x\to\infty}(-xe^{-x}) + 2e^{-2} + \lim_{x\to\infty}(-e^{-x}) + e^{-2}\right\}$$

$$\Rightarrow 1 = c\left\{\lim_{x\to\infty}(-xe^{-x}) + 2e^{-2} + \lim_{x\to\infty}\left(-\frac{1}{e^{x}}\right) + e^{-2}\right\}$$

$$\Rightarrow 1 = c\left\{\lim_{x\to\infty}(-xe^{-x}) + 2e^{-2} + 0 + e^{-2}\right\}$$

(cont'd next slide)

Problem 1

Let X have a PDF defined by: $f(x) = \begin{cases} cxe^{-x}, & x \geq 2, \\ 0, & \text{elsewhere.} \end{cases}$

(a) Find the constant c.

Solution:

$$\Rightarrow 1 \quad = \quad c\left\{\lim_{x\to\infty}(-xe^{-x}) + 2e^{-2} + e^{-2}\right\}$$

$$\lim_{x\to\infty}(-xe^{-x}) = \lim_{x\to\infty} -\frac{x}{e^x} = \frac{\infty}{\infty}$$ (indeterminate form)

Apply *L'Hospital's Rule*: differentiate the numerator, differentiate the denominator, then take the limit.

$$\Rightarrow \lim_{x\to\infty}(-xe^{-x}) = \lim_{x\to\infty} -\frac{x}{e^x} = \lim_{x\to\infty} -\frac{1}{e^x} = 0.$$

$$\begin{aligned} \Rightarrow 1 &= c(2e^{-2} + e^{-2}) \\ \Rightarrow 1 &= c3e^{-2} \\ \Rightarrow c &= \frac{e^2}{3}. \end{aligned}$$

Problem 1

Let X have a PDF defined by: $f(x) = \begin{cases} cxe^{-x}, & x \geq 2, \\ 0, & \text{elsewhere.} \end{cases}$

b Find $P(1 \leq X \leq 3)$.

Solution:

$$
\begin{aligned}
P(1 \leq X \leq 3) &= \int_1^3 f(x)dx \quad \text{probability = area under PDF} \\
&= c\left(\int_1^2 0dx + \int_2^3 xe^{-x}dx\right) \quad \text{substitute the form of PDF above} \\
&= \frac{e^2}{3}\int_2^3 xe^{-x}dx \quad \text{substitute value of } c \text{ from a)} \\
&= \frac{e^2}{3}\left(-xe^{-x}\Big|_2^3 - \int_2^3 (-e^{-x})dx\right) \quad \text{integration by parts} \\
&= \frac{e^2}{3}\left(-3e^{-3} + 2e^{-2} - e^{-x}\Big|_2^3\right) \\
&= \frac{e^2}{3}\left(-3e^{-3} + 2e^{-2} - e^{-3} + e^{-2}\right) \\
&= \frac{e^2}{3}\left(-4e^{-3} + 3e^{-2}\right) = -\frac{4}{3}e^{-1} + 1 = 0.51. \quad \square
\end{aligned}
$$

Problem 2

Let X denote a continuous random variable with PDF

$$f(x) = \begin{cases} \frac{x}{8}, & 0 < x < 4 \\ 0, & \text{otherwise.} \end{cases}$$

Define Y to be the integer that is closest X.

a. Explain why Y is a discrete random variable and give possible values for Y.

Solution:

Possible values of Y are 0, 1, 2, 3, and 4. Thus, Y is discrete.

Problem 2

Let X denote a continuous random variable with PDF

$$f(x) = \begin{cases} \frac{x}{8}, & 0 < x < 4 \\ 0, & \text{otherwise.} \end{cases}$$

Define Y to be the integer that is closest X.

ⓑ Compute the PMF of Y.

Solution:

To get the PMF, we need to compute the probabilities for each possible values of Y.

Approach 1: (Integrating the PDF of X)

$$\begin{aligned}
P(Y = 0) &= P(0 \leq X < 0.5) = \int_0^{0.5} f(x)dx = \frac{x^2}{16}\Big|_0^{0.5} = \frac{1}{64}. \\
P(Y = 1) &= P(0.5 \leq X < 1.5) = \int_{0.5}^{1.5} f(x)dx = \frac{x^2}{16}\Big|_{0.5}^{1.5} = \frac{9}{64} - \frac{1}{64} = \frac{1}{8}. \\
P(Y = 2) &= P(1.5 \leq X < 2.5) = \int_{1.5}^{2.5} f(x)dx = \frac{x^2}{16}\Big|_{1.5}^{2.5} = \frac{25}{64} - \frac{9}{64} = \frac{1}{4}. \\
P(Y = 3) &= P(2.5 \leq X < 3.5) = \int_{2.5}^{3.5} f(x)dx = \frac{x^2}{16}\Big|_{2.5}^{3.5} = \frac{49}{64} - \frac{25}{64} = \frac{24}{64} = \frac{3}{8}.
\end{aligned}$$

(cont'd next slide)

Problem 2

Let X denote a continuous random variable with PDF

$$f(x) = \begin{cases} \frac{x}{8}, & 0 < x < 4 \\ 0, & \text{otherwise.} \end{cases}$$

Define Y to be the integer that is closest X.

b Compute the PMF of Y.

Solution:

$$P(Y = 4) \quad = \quad P(3.5 \leq X < 4) = \int_{3.5}^{4} f(x)dx = \frac{x^2}{16}\Big|_{3.5}^{4} = 1 - \frac{49}{64} = \frac{15}{64}.$$

Therefore, the PMF of Y is: $p(y) = \begin{cases} \frac{1}{64}, & y = 0 \\ \frac{1}{8}, & y = 1 \\ \frac{1}{4}, & y = 2 \\ \frac{3}{8}, & y = 3 \\ \frac{15}{64}, & y = 4 \\ 0, & \text{elsewhere.} \end{cases}$

Problem 2

Let X denote a continuous random variable with PDF

$$f(x) = \begin{cases} \frac{x}{8}, & 0 < x < 4 \\ 0, & \text{otherwise.} \end{cases}$$

Define Y to be the integer that is closest X.

b Compute the PMF of Y.

Solution:

Approach 2: (Finding then evaluating the CDF of X)

When $0 < x < 4$,

$$\begin{aligned} F(x) &= \int_0^x \frac{t}{8} dt \\ &= \frac{t^2}{16}\Big|_0^x \\ &= \frac{x^2}{16}. \end{aligned}$$

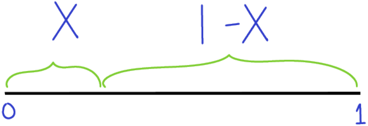
- $P(Y = 0) = P(0 \leq X < 0.5) = F(0.5) - F(0)$.
- $P(Y = 1) = P(0.5 \leq X < 1.5) = F(1.5) - F(0.5)$.
- $P(Y = 2) = P(1.5 \leq X < 2.5) = F(2.5) - F(1.5)$.
- $P(Y = 3) = P(2.5 \leq X < 3.5) = F(3.5) - F(2.5)$.
- $P(Y = 4) = P(3.5 \leq X < 4) = F(4) - F(3.5)$.

Problem 3

Divide a stick into two parts. Find the probability that the larger part of the stick is at least three times the shorter.

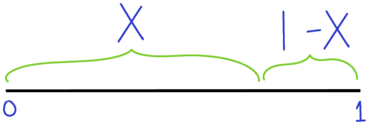
Solution:

- Assume the stick is the interval $(0, 1)$.
- Let X be a point in the interval $(0, 1)$ where we break the stick.
- Since every point on the stick has an equal chance of getting picked, $X \sim U(0, 1)$.
- Two things can happen:



Event 1: X is the shorter piece.

- $X \leq 1 - X$
- $\Rightarrow X \leq \frac{1}{2}$.



Event 2: X is the longer piece.

- $X > 1 - X$
- $\Rightarrow X > \frac{1}{2}$.

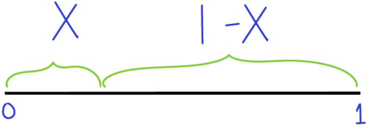
We want to find the probability that the longer segment is at least three times the shorter: $P(1 - X > 3X|E_1)P(E_1) + P\{X > 3(1 - X)|E_2\}P(E_2)$.

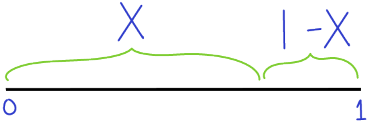
(cont'd next slide)

Problem 3

Divide a stick into two parts. Find the probability that the larger part of the stick is at least three times the shorter.

Solution:





Event 1: X is the shorter piece.

- $X \leq 1 - X$
- $\Rightarrow X \leq \frac{1}{2}$.

Event 2: X is the longer piece.

- $X > 1 - X$
- $\Rightarrow X > \frac{1}{2}$.

Compute: $P(1 - X > 3X | E_1)P(E_1) + P\{X > 3(1 - X)|E_2\}P(E_2)$

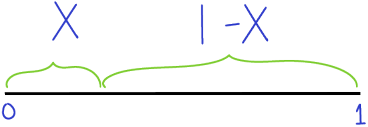
- $P(E_1) = P(X \leq \frac{1}{2}) = \frac{1}{2}$. Recall CDF of $U(a, b)$ is $F(x) = \frac{x-a}{b-a}$. Here, $x = \frac{1}{2}$, $a = 0$, and $b = 1$.
- $P(E_2) = P(X > \frac{1}{2}) = 1 - P(X \leq \frac{1}{2}) = 1 - \frac{1}{2} = \frac{1}{2}$.
- $P(1 - X > 3X | E_1) = \frac{P\{(X < \frac{1}{4}) \cap E_1\}}{P(E_1)} = \frac{P(X < \frac{1}{4})}{\frac{1}{2}} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$.
- $P\{X > 3(1 - X)|E_2\} = \frac{P\{(X > \frac{3}{4}) \cap E_2\}}{P(E_2)} = \frac{P(X > \frac{3}{4})}{\frac{1}{2}} = \frac{1 - P(X \leq \frac{3}{4})}{\frac{1}{2}} = \frac{1 - \frac{3}{4}}{\frac{1}{2}} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$.

(cont'd next slide)

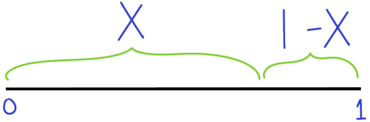
Problem 3

Divide a stick into two parts. Find the probability that the larger part of the stick is at least three times the shorter.

Solution:



Event 1: X is the shorter piece.



Event 2: X is the longer piece.

- $P(E_1) = \frac{1}{2}$.
- $P(E_2) = \frac{1}{2}$.
- $P(1 - X > 3X|E_1) = \frac{1}{2}$.
- $P\{X > 3(1 - X)|E_2\} = \frac{1}{2}$.

Hence,

$$P(1 - X > 3X|E_1)P(E_1) + P\{X > 3(1 - X)|E_2\}P(E_2) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{2}.$$

Problem 4

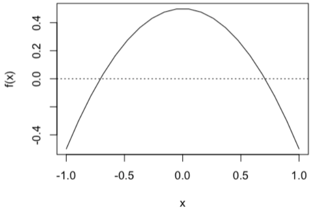
Suppose we have the following function:

$$f(y) = \begin{cases} c\left(\frac{1}{2} - y^2\right), & -1 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Is this a valid PDF? If not, is there a c for which this becomes a valid PDF?

Solution:

- When $c = 0$, then $f(y)$ does not integrate to 1.
- When $c \neq 0$, there is an interval in $-1 \leq y \leq 1$ over which $f(y)$ is negative.



- Therefore, the $f(y)$ above will NOT make a valid PDF. $\square$

Problem 5

Derive the PDF of $|X|$ where $X \sim U(-1, 1)$.

Solution:

- Since X is uniform on the interval $(-1, 1)$, its PDF is: $f(x) = \begin{cases} \frac{1}{2}, & -1 \leq x \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$
- The PDF of $|X|$ can be derived from its CDF, $F_{|X|}(x)$.
- We first need to find the CDF, $F_{|X|}(x)$. The CDF can be derived as follows: For $0 \leq x \leq 1$,

$$\begin{aligned} F_{|X|}(x) &= P(|X| \leq x) \quad \text{def'n of CDF} \\ &= P(-x \leq X \leq x) \quad \text{absolute value inequality} \\ &= \int_{-x}^{x} \frac{1}{2} dt \quad \text{probability = area under the PDF} \\ &= \frac{1}{2} t \Big|_{-x}^{x} = \frac{1}{2}(2x) = x. \end{aligned}$$

- We can now get the PDF of $|X|$ from the CDF as follows:

$$f_{|X|}(x) = \frac{d}{dx} F_{|X|}(x) = \begin{cases} 1, & 0 \leq x \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Previously...

Uniform Distribution

- Notation: $Y \sim U(\theta_1, \theta_2)$
- Parameters: θ_1 (minimum), θ_2 (maximum)
- PDF: $f(y) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \theta_1 \leq y \leq \theta_2, \\ 0, & \text{elsewhere} \end{cases}$
- CDF: $F(y) = \begin{cases} 0, & y < \theta_1 \\ \frac{y - \theta_1}{\theta_2 - \theta_1}, & \theta_1 \leq y \leq \theta_2 \\ 1, & y > \theta_2. \end{cases}$
- Mean or Expected Value: $\frac{\theta_1 + \theta_2}{2}$
- Variance: $\frac{(\theta_2 - \theta_1)^2}{12}$

Normal (Gaussian) Distribution

- Notation: $Y \sim \mathcal{N}(\mu, \sigma^2)$
- Parameters: μ (mean), σ (standard deviation)
- PDF: $f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad -\infty \leq y \leq \infty$
- CDF: $F(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$ (no explicit form)
- Mean or Expected Value: μ
- Variance: σ^2

Standard Normal Distribution

- Notation: $Z \sim \mathcal{N}(0, 1)$
- Parameters: $\mu = 0$ (mean), $\sigma = 1$ (standard deviation)
- PDF: $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty \leq z \leq \infty$
- CDF: $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ (no explicit form)
- Mean or Expected Value: 0
- Variance: 1

Computing $P(a \leq Z \leq b)$, $Z \sim \mathcal{N}(0,1)$

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691	.08534	.08379	.08226
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786

- Row headings: z-score up to the first decimal place.
- Column headings: second decimal place of the z-score.
- Cells: areas under the standard normal curve to the left of every z-score.

Example:

$$P(-1 \leq Z \leq 1) = \Phi(1) - \Phi(-1) = 0.84134 - 0.15866 = 0.68268$$

Computing $P(a \leq Y \leq b)$, $Y \sim \mathcal{N}(\mu, \sigma^2)$

$$
\begin{aligned}
P(a \leq Y \leq b) &= P(Y \leq b) - P(Y \leq a) && \text{probability = area under the PDF} \\
&= P\left(\frac{Y-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) - P\left(\frac{Y-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) && \text{standardization won't change the inequality} \\
&= P\left(Z \leq \frac{b-\mu}{\sigma}\right) - P\left(Z \leq \frac{a-\mu}{\sigma}\right) && \text{def'n of standard normal r.v.} \\
&= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). && \text{standard normal CDF}
\end{aligned}
$$

Example:

Let $Y \sim \mathcal{N}(3, 16)$, what is $P(2 < Y < 5)$?

$$
\begin{aligned}
P(2 < Y < 5) &= P\left(\frac{2-3}{\sqrt{16}} \leq \frac{Y-3}{\sqrt{16}} \leq \frac{5-3}{\sqrt{16}}\right) && \text{standardize } Y \\
&= P\left(-\frac{1}{4} \leq Z \leq \frac{2}{4}\right) \\
&= \Phi\left(\frac{1}{2}\right) - \Phi\left(-\frac{1}{4}\right) && \text{check corresponding values of CDF on the Z-table} \\
&= 0.69146 - 0.40129 = 0.29017.
\end{aligned}
$$

Normal (Gaussian) Distribution (cont'd)

Gaussian Distribution: Some Properties

If $Z \sim \mathcal{N}(0,1)$ with CDF $\Phi(z)$, then

1. $Y = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$.

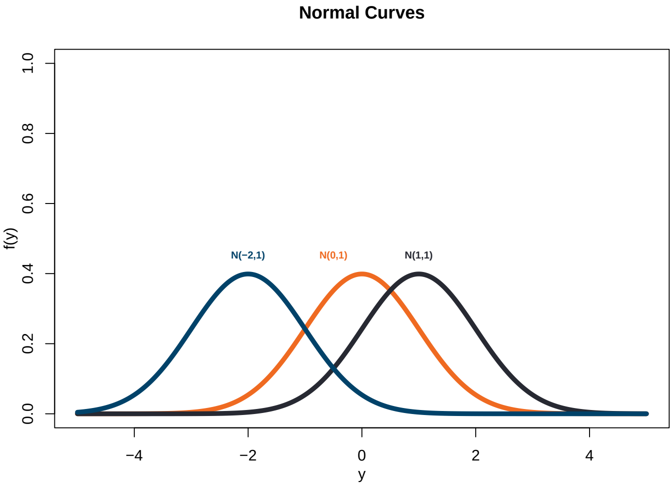
   Proof:

   Let $F_Y(y)$ be the CDF of Y and $f_Y(y)$ be the PDF of Y. Then,

$$
\begin{aligned}
F_Y(y) = P(Y \leq y) &= P\left(\mu + \sigma Z \leq y\right) \\
&= P\left(\sigma Z \leq y - \mu\right) \\
&= P\left(Z \leq \frac{y-\mu}{\sigma}\right) \\
&= F_Z\left(\frac{y-\mu}{\sigma}\right). \quad \text{standard normal CDF} \\
f_Y(y) = \frac{d}{dy} F_Y(y) &= \frac{d}{dy} F_Z\left(\frac{y-\mu}{\sigma}\right) \quad \text{def'n of PDF} \\
&= \frac{1}{\sigma} f_Z\left(\frac{y-\mu}{\sigma}\right) \quad f(z) = F'(z) \text{ and chain rule} \\
&= \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(\frac{y-\mu}{\sigma}\right)^2}{2}} \quad \text{PDF of standard normal} \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}. \quad \text{(This is the Gaussian PDF)}
\end{aligned}
$$

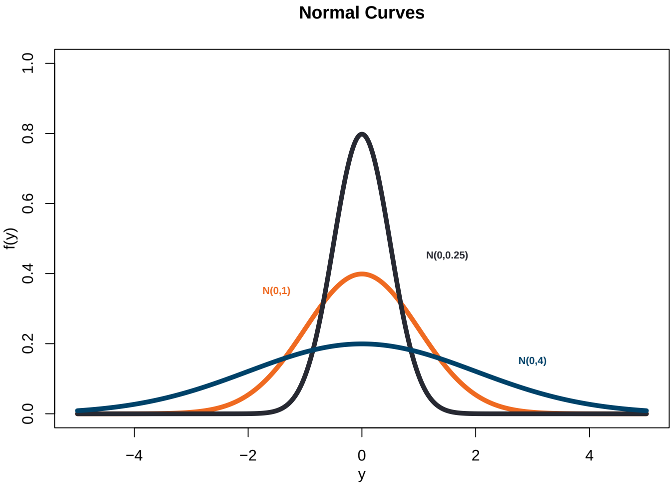
Gaussian Distribution: Mean, μ, Controls the Location

$Y = \mu + Z \sim \mathcal{N}(\mu, 1)$, $\mu = -2, 0, 1$.



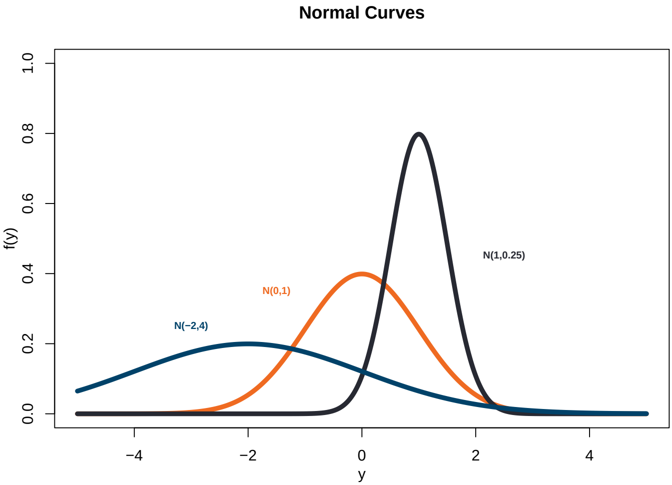
Gaussian Distribution: SD, σ, Controls the Width

$Y = \sigma Z \sim \mathcal{N}(0, \sigma^2)$, $\sigma = 1, 0.5, 2.$



Gaussian Distribution: (μ, σ) Shifts & Scales

$$Y = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2).$$



Gaussian Distribution: Some Properties

If $Z \sim \mathcal{N}(0,1)$ with CDF $\Phi(z)$, then

2. $\Phi(-z) = 1 - \Phi(z)$ (symmetry)

Proof:

$$
\begin{aligned}
\Phi(-z) &= P(Z \leq -z) = \int_{-\infty}^{-z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad \text{standard normal CDF} \\
&= -\int_{-z}^{-\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad \text{change the sign when flipping the bounds of integration}
\end{aligned}
$$

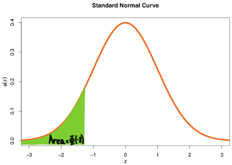
Next, do the change of variables: Let $u = -t \Rightarrow du = -dt$.

$$
\begin{aligned}
&= -\int_{z}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(-u)^2}{2}} (-du) \\
&= \int_{z}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \\
&= P(Z \geq z) \quad \text{probability} = \text{area under the curve} \\
&= 1 - P(Z < z) \quad \text{complement} \\
&= 1 - \Phi(z). \quad \text{standard normal CDF}
\end{aligned}
$$

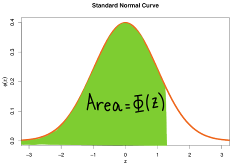
$\square$

If $Z \sim \mathcal{N}(0, 1)$ with CDF $\Phi(z)$, then

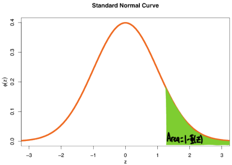
2. $\Phi(-z) = 1 - \Phi(z)$ (symmetry)



$= \quad 1 \quad -$



$=$



Gaussian Distribution: Some Properties

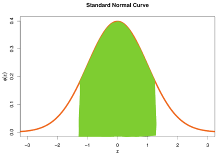
If $Z \sim \mathcal{N}(0,1)$ with CDF $\Phi(z)$, then

3. $P(-z < Z < z) = 2\Phi(z) - 1$

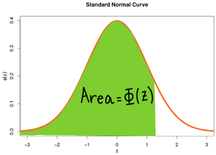
   Proof:

$$\begin{aligned}
P(-z < Z < z) &= P(Z \le z) - P(Z \le -z) && \text{probability = area under the curve} \\
&= \Phi(z) - \Phi(-z) && \text{standard normal CDF} \\
&= \Phi(z) - \{1 - \Phi(z)\} && \text{from previous slide: } \Phi(-z) = 1 - \Phi(z) \\
&= \Phi(z) - 1 + \Phi(z) \\
&= 2\Phi(z) - 1.
\end{aligned}$$

$\square$



$= \; 2\times$



$- \; 1$

Example 1:

Suppose X, the grade on a midterm exam, is normally distributed with mean 70 and standard deviation 10. The instructor wants to give 15% of the class an A. What cutoff should the instructor use to determine who gets an A?

Solution:

- We need to find z (cutoff "grade" in the std normal distribution) such that

$$\begin{aligned} \Phi(z) &= 0.85 \quad \text{85\% of the students have grades below z} \quad \text{(Eqn 1) OR} \\ 1 - \Phi(z) &= 0.15 \quad \text{15\% of the students have grades above z} \quad \text{(Eqn 2)} \end{aligned}$$

- In the Z-table, look for the row and column where you can find the cell value 0.85.
- From the Z-table, $\Phi(1.04) = 0.85$.
- Hence, z in (Eqn 1) should be 1.04.
- This means that the cutoff grade in the std normal distribution is $z = 1.04$.

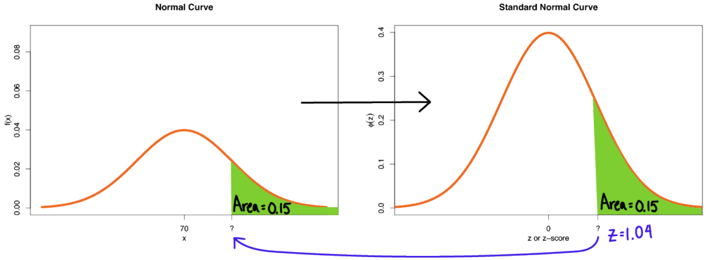
(cont'd next slide)

Gaussian Distribution

Example 1:

Suppose X, the grade on a midterm exam, is normally distributed with mean 70 and standard deviation 10. The instructor wants to give 15% of the class an A. What cutoff should the instructor use to determine who gets an A?

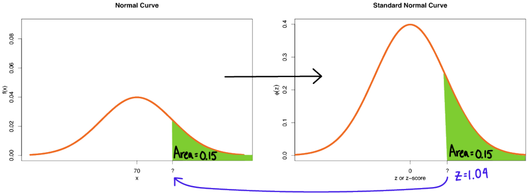
Solution:



We still need to compute the "real" cutoff score x.
(cont'd next slide)

Example 1:
Suppose X, the grade on a midterm exam, is normally distributed with mean 70 and standard deviation 10. The instructor wants to give 15% of the class an A. What cutoff should the instructor use to determine who gets an A?

Solution:



- To find the cutoff score x, in the original problem, we need to convert $z = 1.04$ back to the original magnitude/scale of the problem.
- Recall: $Z = \frac{X-\mu}{\sigma}$ and $X = \mu + \sigma Z$
- Applying the appropriate transformation, the cutoff score should be: $x = \mu + \sigma(1.04) = 70 + 10(1.04) = 80.4$
- This means that to ensure 15% of the students get an A, the cutoff grade for getting an A should be 80.4.

Example 2:

Let X be a Gaussian random variable with mean 5. If $P(X > 9) = 0.2$, approximately what is $V(X)$?

Solution: Let $\sigma^2 = V(X)$.

$$\begin{aligned} P(X > 9) &= 1 - P(X \leq 9) && \text{complement} \\ &= 1 - P\left(\frac{X-5}{\sigma} \leq \frac{9-5}{\sigma}\right) && \text{standardize } X \\ &= 1 - P\left(Z \leq \frac{4}{\sigma}\right) && \text{def'n of standard normal r.v.} \\ &= 1 - \Phi\left(\frac{4}{\sigma}\right). && \text{standard normal CDF} \end{aligned}$$

- We need $1 - \Phi\left(\frac{4}{\sigma}\right) = 0.2$ (given) $\Rightarrow \Phi\left(\frac{4}{\sigma}\right) = 0.8$
- In the Z-table, look for the row and column where you can find the cell value 0.8.
- From the Z-table, $\Phi(0.84) \approx 0.8$.
- Hence, $\frac{4}{\sigma}$ must be equal to 0.84. $\Rightarrow \frac{4}{\sigma} = 0.84$
- Therefore, $\sigma = 4.76$ and $V(X) = (4.76)^2 = 22.66$. □

Gaussian Distribution

Example 3:

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ such that $P(X > 20.6) = P(X \leq -18.6) = 0.025$? What are the values of μ and σ?

Solution:

$$\begin{aligned}
P(X > 20.6) &= 1 - P(X \leq 20.6) && \text{complement} \\
&= 1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{20.6-\mu}{\sigma}\right) && \text{standardize } X \\
&= 1 - P\left(Z \leq \frac{20.6-\mu}{\sigma}\right) && \text{def'n of standard normal r.v.} \\
&= 1 - \Phi\left(\frac{20.6-\mu}{\sigma}\right). && \text{standard normal CDF}
\end{aligned}$$

(cont'd next slide)

Example 3:
Let $X \sim \mathcal{N}(\mu, \sigma^2)$ such that $P(X > 20.6) = P(X \leq -18.6) = 0.025$? What are the values of μ and σ?

Solution:

$$
\begin{aligned}
P(X \leq -18.6) &= P\left(\frac{X-\mu}{\sigma} \leq \frac{-18.6-\mu}{\sigma}\right) && \text{standardize } X \\
&= P\left(Z \leq \frac{-18.6-\mu}{\sigma}\right) && \text{def'n of standard normal r.v.} \\
&= \Phi\left(\frac{-18.6-\mu}{\sigma}\right). && \text{standard normal CDF}
\end{aligned}
$$

(cont'd next slide)

Example 3:

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ such that $P(X > 20.6) = P(X \leq -18.6) = 0.025$? What are the values of μ and σ?

Solution:

- We need to satisfy the following:
  - $1 - \Phi\left(\frac{20.6-\mu}{\sigma}\right) = 0.025 \Rightarrow \Phi\left(\frac{20.6-\mu}{\sigma}\right) = 0.975$ and
  - $\Phi\left(\frac{-18.6-\mu}{\sigma}\right) = 0.025$
- From the Z-table, we can see
  - $\Phi(1.96) = 0.975$ and
  - $\Phi(-1.96) = 0.025$
- We have the following system of equations:

$$\begin{cases} \frac{20.6-\mu}{\sigma} = 1.96 \\ \frac{-18.6-\mu}{\sigma} = -1.96 \end{cases} \Rightarrow \begin{cases} 20.6 - \mu = 1.96\sigma & (1) \\ -18.6 - \mu = -1.96\sigma & (2) \end{cases}$$

- Subtracting Eqn (2) from Eqn (1), we have: $39.2 = 3.92\sigma$.
- Hence, $\sigma = 10$.
- Using Eqn (1) to solve for μ, we have: $\mu = 20.6 - 1.96(10) = 1$. $\square$

Example 4:

The annual rainfall (in inches) in a certain region is normally distributed with $\mu = 40$ and $\sigma = 4$. What is the probability that, starting with this year, it will take over 10 years before a year occurs having a rainfall of over 50 inches? What assumptions are you making?

Solution:

Let X denote the annual rainfall and E denote the event that it will take over 10 years starting from this year before a year occurs having a rain fall of over 50 inches.

$$\begin{aligned}
P(X > 50) &= 1 - P(X \leq 50) && \text{probability of rainfall over 50 inches} \\
&= 1 - P\left(\frac{X-40}{4} \leq \frac{50-40}{4}\right) && \text{standardize } X \\
&= 1 - P\left(Z \leq 2.5\right) && \text{def'n of standard normal r.v.} \\
&= 1 - \Phi\left(2.5\right) && \text{standard normal CDF} \\
&= 1 - 0.99379 = 0.006. && \text{Z-table value}
\end{aligned}$$

(cont'd next slide)

Gaussian Distribution

Example 4:

The annual rainfall (in inches) in a certain region is normally distributed with $\mu = 40$ and $\sigma = 4$. What is the probability that, starting with this year, it will take over 10 years before a year occurs having a rainfall of over 50 inches? What assumptions are you making?

Solution:

Let X denote the annual rainfall and E denote the event that it will take over 10 years starting from this year before a year occurs having a rain fall of over 50 inches.

We computed $P(X > 50) = 0.006$, the probability of getting more than 50 inches of rain in any given year.

The goal is to compute:

$P(E) = P(\text{None of the first 10 years have more than 50 inches of rain})$

$$= \overbrace{\{1 - P(X > 50)\}\{1 - P(X > 50)\} \cdots \{1 - P(X > 50)\}}^{10} \text{ independence}$$
$$= (1 - 0.006)^{10} = 0.94.$$

Here, we assumed that the annual rainfall is independent from year to year. □

Questions?

Homework Exercises: 4.61, 4.71, 4.73, 4.77, 4.81

Solutions will be discussed this Friday by the TA.